Capitolo 2 TRASMISSIONE

Modelli e trasmissione in banda base

2.1 MODELLO DI UN SISTEMA DI TRASMISSIONE DELL'INFORMAZIONE

Un sistema di trasmissione dell'informazione può essere schematizzato attraverso un modello composto dai seguenti sottosistemi:

— una sorgente d'informazione, di natura continua (analogica) o discreta;

— un codificatore dell'informazione emessa dalla sorgente, ai fini di una trasmissione efficiente sul canale disponibile.
Si noti che informazione di natura originariamente analogica potrà eventualmente essere trasformata in forma numerica;

— un canale sul quale l'informazione codificata è trasmessa.
La trasmissione sul canale è generalmente perturbata dalla presenza di disturbi di varia natura (il "rumore") che si aggiungono al segnale trasmesso, nonché eventualmente da distorsioni prodotte sui segnali trasmessi da caratteristiche non ideali del canale;

— un sistema di ricezione che ricostruisce l'informazione originaria a partire dal segnale ricevuto dal canale.

Poiché il segnale ricevuto è contaminato da rumore, il segnale trasmesso potrà essere ricostruito in ricezione solo imperfettamente e sarà quindi affetto da errore. Nel caso di segnali analogici si potrà misurare l'errore per esempio tramite il valore quadratico medio della differenza tra segnale trasmesso e segnale ricostruito in ricezione (ossia la potenza di rumore in uscita). Per il caso numerico, invece, i segnali trasmessi possibili sono in numero finito e in ricezione si deciderà, in maniera corretta oppure errata; le prestazioni del sistema di trasmissione potranno quindi essere misurate con la probabilità di errore che, per una sequenza stazionaria di simboli trasmessi, dà il tasso medio di errore.

Si noti che, nei sistemi reali, a ciascun sottosistema di questo modello possono corrispondere più funzioni. Per esempio ciò che indichiamo qui col termine "codificatore" può corrispondere alla sequenza di tre operazioni: codificazione di sorgente, per la rappresentazione efficiente, per esempio in forma binaria, dei messaggi emessi dalla sorgente; codificazione di canale, per la protezione dell'informazione dagli errori che i disturbi presenti sul canale di trasmissione potranno causare; modulazione, cioè la rappresentazione dei messaggi codificati mediante opportuni segnali che siano adatti alle caratteristiche trasmissive del canale disponibile.

La *fig. 2.1* mostra uno schema a blocchi delle principali funzioni che normalmente si ritrovano in un sistema trasmissivo. A volte, naturalmente, alcune di queste funzioni (codificazione di sorgente, codificazione di canale ecc.) non sono presenti, se le caratteristiche del sistema non le richiedono. D'altro canto possono essere presenti a seconda dei casi anche altre funzioni non menzionate nella figura. Citiamo per esempio: l'utilizzazione simultanea multipla ("multiplex", accesso multiplo) di uno stesso canale trasmissivo da parte di più segnali distinti; la crittografia, quando si voglia rendere l'informazione in ricezione interpretabile solo da parte di destinatari autorizzati, e perciò in possesso di opportune "chiavi" per la decifrazione; la sincronizzazione, in pratica sempre presente e di fondamentale importanza nei sistemi di trasmissione numerica, poiché la ricezione corretta di simboli numerici trasmessi in sequenza richiede in ricezione la conoscenza della posizione temporale di tali simboli.

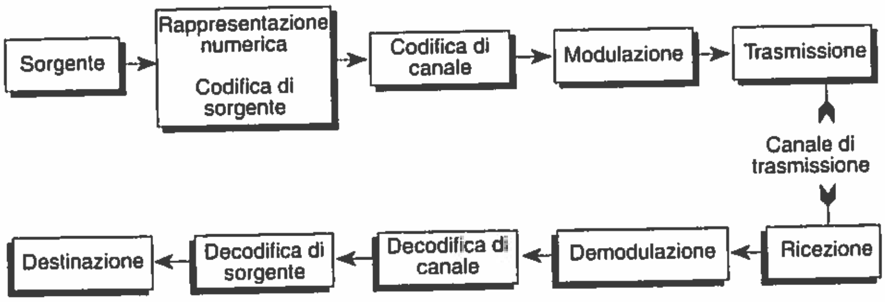

Figura 2.1 - Schema delle principali funzioni di un generico sistema di trasmissione numerico

Si osservi che in realtà il modello che stiamo descrivendo è rappresentativo non solo di un sistema di trasmissione a distanza, ma anche per esempio di sistemi di memorizzazione, con relative operazioni di "scrittura" e "lettura" sulla memoria (nastri o dischi magnetici, o ottici ecc.). In questi l'operazione di scrittura corrisponde alla trasmissione, quella di lettura alla ricezione, e tali operazioni saranno disturbate da "rumore" o errori analogamente a quanto avviene nella trasmissione a distanza.

Svariati sono i tipi di informazione emessi dalle diverse sorgenti. In natura l'informazione si presenta in forma sia discreta che continua. Anche l'informazione originariamente in forma continua è spesso ridotta in forma numerica. Tra le ragioni più importanti per tale riduzione citiamo:

— maggior convenienza per la trasmissione efficiente in presenza di rumore;
— maggior adattabilità dei moderni circuiti elettronici integrati alla elaborazione e alla memorizzazione dei segnali numerici;
— omogeneizzazione dei vari tipi di informazione distribuiti da una rete di comunicazione onde ottenere un unico tipo di segnale, appunto un segnale numerico, circolante nella rete.

Di grande importanza è il caso della rappresentazione numerica dell'informazione in forma binaria, cioè mediante due soli simboli, 1 e 0. A questo caso faremo spesso riferimento per la sua semplicità e importanza pratica, anche per illustrare esemplificativamente temi riguardanti in generale i vari tipi di sistema. Il singolo messaggio è detto in questo caso "bit" (binary digit, cifra binaria). La sorgente d'informazione emette simboli, scelti nel suo alfabeto, in questo caso binario; il particolare simbolo emesso in un certo istante non è ovviamente noto a priori, in quanto appunto è portatore d'informazione. Pertanto si può dare una descrizione dell'emissione della sorgente solo in termini statistici. Nel caso binario citato, si potrà dare la distribuzione probabilistica dei simboli 1 e 0, cioè la loro probabilità, e , se esistono, le correlazioni tra i simboli successivi emessi in sequenza.

Più in generale la sorgente d'informazione potrà essere descritta come processo casuale, discreto o continuo.

2.2 ALCUNI ESEMPI DI SORGENTI E SEGNALI

Testi

Intendiamo con "testo" una sequenza di caratteri alfanumerici. Il classico codice di Morse rappresenta storicamente il primo esempio di codificazione di testi per le comunicazioni elettriche. Esso fu usato nel 1843 per la prima trasmissione elettrica a distanza, dando così origine alla telegrafia tradizionale, ed è basato su un alfabeto di simboli ternario e su una codifica a lunghezza variabile. I simboli elementari usati sono "punto", "linea" e "spazio", che vengono usati in combinazione per rappresentare i segni dell'alfabeto ordinario. Fisicamente ai tre simboli si fa corrispondere rispettivamente un "segnale impulsivo breve", un "segnale impulsivo lungo" e "assenza di segnale". Tale codice costituisce un buon esempio di codificazione efficiente della sorgente. Per codificazione efficiente intendiamo la riduzione al minimo della lunghezza *media* dei caratteri, ottenuta a partire dalle caratteristiche statistiche della sorgente, per esempio le probabilità delle varie lettere dell'alfabeto in un testo italiano. Si rappresentano cioè le lettere più frequenti con i caratteri più corti (per esempio la lettera "e", che è molto frequente, con "punto") e quelle meno frequenti con i caratteri più lunghi (per esempio la lettera "q" con "linea" "linea" "punto" "linea").

Vari tipi di codici di rappresentazione si sono poi susseguiti nell'uso, seguendo l'evolversi degli standard internazionali. Normalmente è conveniente usare codici binari con parole di lunghezza costante, come per esempio il codice internazionale n. 5 (ASCII) che usa parole di 7 bit per ciascun carattere, consentendo quindi di rappresentare 128 caratteri (lettere dell'alfabeto ordinario maiuscole e minuscole, numeri decimali, segni d'interpunzione e messaggi speciali quale per esempio il messaggio "fine della trasmissione").

Nella maggior parte dei casi l'alfabeto usato per la codificazione di sorgenti è binario. In alcuni casi l'alfabeto consiste di N simboli, per esempio N livelli di ampiezza di un impulso di forma prefissata; un tale alfabeto può anche essere usato per rappresentare, quando ciò risulti conveniente, sequenze binarie cioè gruppi di $\log_2 N$ bit, le cui possibili configurazioni sono appunto N.

Il segnale fonico

Il segnale telefonico si presenta sulla scena delle telecomunicazioni dopo il segnale telegrafico per la trasmissione di testi, ma viene a costituire, e tuttora costituisce, il segnale più importante e diffuso nelle reti di comunicazione. Naturalmente le caratteristiche del segnale elettrico corrispondente ad un segnale acustico (voce, musica) dipendono dalla qualità di riproduzione desiderata e quindi dalla applicazione specifica (telefonia ovvero musica ad alta fedeltà ecc.). Per esempio nel caso della telefonia possiamo dire che è sufficiente trasmettere le componenti spettrali di frequenza inferiore a circa 4 kHz per avere una qualità accettabile per l'applicazione in questione. Se si vuole invece trasmettere un segnale musicale di qualità, si dovranno trasmettere anche le componenti a frequenza più elevata, per esempio fino a 15 kHz. È possibile rappresentare questi segnali, di natura analogica all'origine, in forma numerica, per esempio tipicamente con sequenze di bit: ciò si ottiene prelevando, dal segnale analogico, campioni con un ritmo pari almeno al doppio della massima frequenza contenuta nel segnale analogico, discretizzando quindi l'ampiezza di tali campioni con un numero M di livelli sufficientemente grande, e codificando infine i livelli con parole binarie di $\log_2 M$ bit (si vedano più avanti i sistemi a modulazione impulsiva codificata PCM).

I segnali visivi

Anche per i segnali che rappresentano immagini si ha una grande varietà di casi. Le immagini da trasmettere possono essere fisse o in movimento, oltre che in bianco e nero o a colori.

Possiamo assimilare ai sistemi di trasmissione di immagini anche i sistemi "facsimile", largamente usati per la trasmissione di documenti, foto ecc. In essi la scansione riga per riga della pagina o dell'immagine genera un segnale analogico o numerico a seconda dei casi. Considerando il caso numerico, la velocità in bit/s del segnale numerico dipenderà dalla risoluzione desiderata, oltre che naturalmente dal tempo impiegato per la trasmissione. Se si considera per esempio una risoluzione orizzontale di 1800 punti per riga e una risoluzione verticale di 4 linee/mm, una normale pagina (20×30 cm) presenta circa 2.10^6 punti, e quindi la trasmissione comporta un flusso di circa 2 Mbit per pagina, rappresentando ogni punto (bianco o nero) con un bit. Maggiore efficienza si può ottenere con opportuni metodi di codificazione che tengano conto delle differenti probabilità delle varie possibili sequenze di punti bianchi e neri.

Analogamente se si considera, passando alle immagini in movimento, il caso dell'ordinario segnale televisivo, si ha che il segnale analogico che rappresenta la luminosità di ciascun punto (*fig. 2.2*) ha uno speto che si estende fino a circa 5MHz. Tale segnale si ottiene, per esempio

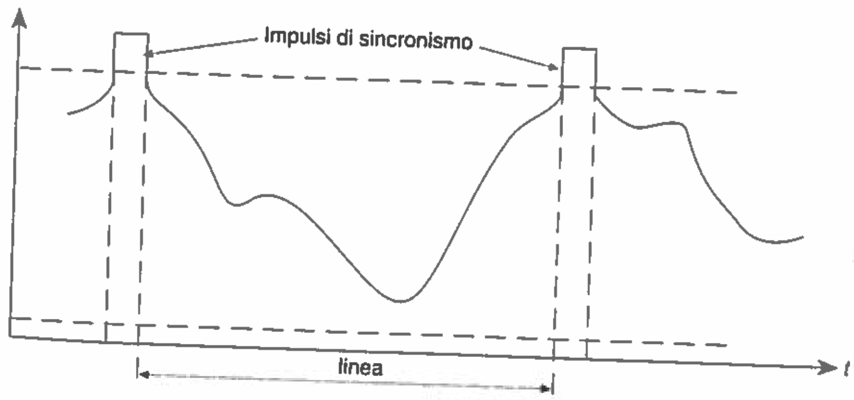

Figura 2.2 - Segnale televisivo: andamento della luminosità lungo una generica linea del quadro.

secondo lo standard europeo, dalla scansione a 625 linee di ogni quadro, al ritmo di 25 quadri al secondo. Pensando ad una rappresentazione numerica di tale segnale, possiamo considerare una risoluzione di 625 righe e circa 800 punti per riga, e quindi circa 5.10^5 punti per immagine. La trasmissione di 25 immagini (quadri) al secondo è sufficiente. In pratica si ottiene per l'occhio un effetto migliore con una scansione alterna, anziché sequenziale, delle righe: il quadro di 625 righe è scomposto in due semiquadri interallacciati, uno costituito dalle righe di ordine pari e l'altro da quelle di ordine dispari, semiquadri che vengono trasmessi successivamente. Per 25 immagini trasmesse al secondo si ha un flusso di $12{,}5 \cdot 10^6$ punti/s. Consideriamo immagini in bianco e nero, cioè una rappresentazione della luminosità dei punti dell'immagine attraverso vari livelli di grigio. Quantizzando la luminosità con un numero M di livelli, possiamo rappresentare tali livelli con caratteri di $\log_2 M$ bit, e quindi in definitiva si avrà un flusso di dati binari pari a $12{,}5 \cdot 10^6 \log_2 M$ bit/s. Per $M = 512$, occorrono caratteri di 9 bit per rappresentare la luminanza di ciascun punto e ciò corrisponde ad un flusso totale superiore ai 100 Mbit/s. Se si desidera una maggior risoluzione, per esempio doppia sia in orizzontale che in verticale (sistemi ad alta definizione, a 1250 righe), il flusso d'informazione da trasmettere sarà quadruplo. Viceversa per sistemi in cui si trasmettono a ritmo lento immagini fisse, è sufficiente una trasmissione a velocità molto inferiore, che dipenderà naturalmente dal ritmo desiderato.

2.3 IL "RUMORE"

Nel passaggio del segnale che porta l'informazione attraverso i vari circuiti o mezzi trasmissivi che costituiscono un sistema di comunicazione, avverrà inevitabilmente che altri segnali spuri indesiderati vadano ad aggiungersi al segnale utile disturbando così la ricezione. Tali disturbi possono essere interferenze provenienti dall'esterno; ma anche se queste non ci sono, ci saranno comunque inevitabilmente segnali spuri legati alla natura fisica stessa di circuiti e dispositivi, e della corrente elettrica o della luce (rumore termico, elettronico, fotonico). Rimandando la descrizione più approfondita di questi fenomeni (si veda l'Appendice A), diamo per ora qualche breve cenno sul rumore termico, a titolo esemplificativo per un aiuto alla comprensione della trattazione teorica che segue sulla trasmissione di segnali in presenza di rumore.

Si consideri per esempio un resistore (resistenza R). L'agitazione termica casuale degli elettroni genera ai capi del resistore stesso una fluttuazione casuale di tensione, e quindi un rumore; se nel resistore transita una data corrente continua, a questa (che è il valor medio) risulta sovrapposta una componente casuale dovuta alle fluttuazioni degli elettroni.

Con argomenti termodinamici si può mostrare che il processo casuale che rappresenta la tensione ai capi di un resistore in circuito aperto ha un valore quadratico medio dato da $4KTRB$ essendo K la costante di Boltzmann pari a $1{,}38.10^{-23}$, T la temperatura assoluta in gradi Kelvin, R il valore della resistenza e B la banda di frequenza in cui la tensione è misurata. Tale rumore ha componenti spettrali distribuite uniformemente lungo tutto l'asse delle frequenze, cioè, come si dice, è "bianco" (per analogia con la luce bianca che contiene i vari colori, cioè varie frequenze). Questo vale nell'intervallo di frequenze in cui si possono trascurare gli effetti quantistici, cioè fino a frequenze dell'ordine di 10^{13}Hz.

Il rumore termico è il tipico fenomeno, risultante da combinazioni di innumerevoli eventi microscopici, cui può applicarsi il teorema del limite centrale; pertanto esso può essere considerato un processo casuale gaussiano. Per una sua descrizione completa, basta allora conoscere, nel caso di stazionarietà del processo, lo spettro di potenza, ovvero la funzione di autocorrelazione (che ne è la trasformata di Fourier).

Nel caso limite di rumore bianco (densità spettrale di potenza $S(f)=\frac{N_0}{2}$ costante lungo tutto l'asse delle frequenze), si ha che la funzione di autocorrelazione è $R(\tau)=\frac{N_0}{2}\,\delta(\tau)$.

I valori presi in istanti distinti sono dunque variabili casuali incorrelate (e statisticamente indipendenti se il processo è anche gaussiano). Per la tensione di rumore termico di un resistore R alla temperatura assoluta T, si ha $S(f)=2KTR$ essendo K la costante di Boltzmann.

2.4 DISTORSIONE DEI SEGNALI NEL CANALE DI TRASMISSIONE

La trasmissione è perturbata non solo da segnali spuri che si aggiungono al segnale utile desiderato, ma anche dal fatto che i canali e circuiti di trasmissione possono distorcere i segnali in essi transitanti.

È immediato constatare che la funzione di trasferimento $H(f)=|H|e^{j\theta(f)}$ di un canale o circuito lineare non distorcente, cioè che genera al massimo un ritardo t_0 o una attenuazione senza che venga alterata la forma dei segnali, deve essere del tipo

$$H(f)=Ce^{-j\omega t_0}$$

con C costante. La caratteristica di ampiezza deve cioè essere costante, e la caratteristica di fase lineare con la frequenza, nella banda che interessa. Infatti in questo caso, se il segnale d'ingresso al canale o circuito è $s(t)$, l'uscita corrispondente risulta essere $Cs(t-t_0)$, in base alle ben note proprietà della trasformata di Fourier.

La condizione di linearità della fase equivale alla condizione di invarianza con la frequenza, nella banda del segnale, della derivata della fase $\frac{d\vartheta}{d\omega}$ cioè del cosiddetto ritardo di gruppo, in questo caso pari a t_0 (si veda l'Appendice B).

Oltre a questo tipo di distorsioni presenti nei sistemi lineari, si possono incontrare anche distorsioni prodotte da nonlinearità delle caratteristiche dei circuiti. Consideriamo il caso più semplice di circuiti con caratteristica istantanea ingresso-uscita $y=g(x)$ non lineare: un esempio tipico è dato dagli amplificatori (*fig. 2.3*), le cui caratteristiche tensione di uscita/tensione d'ingresso, entro certi limiti d'ampiezza per il segnale d'ingresso, possono essere considerate perfettamente lineari ma al di là di questi limiti diventano non lineari (saturazione). In questo caso la distorsione non lineare dell'ampiezza produce non solo una deformazione delle caratteristiche di ampiezza e fase dello spettro del segnale, ma anche l'apparizione di nuove frequenze e quindi un allargamento dello spettro. Prendiamo per esempio una semplice nonlinearità corrispondente ad una caratteristica quadratica: se l'ingresso è costituito da una somma di sinusoidi a diverse frequenze, troveremo all'uscita, come si verifica subito, oltre alle frequenze d'ingresso anche componenti a frequenze date da somme e differenze delle frequenze d'ingresso. Sinteticamente, se calcoliamo lo spettro di Fourier $Y(f)$ all'uscita di un quadratore in corrispondenza di un ingresso $x(t)$ con trasformata di Fourier $X(f)$, si ha che lo spettro all'uscita è dato dalla convoluzione dello spettro $X(f)$ con se stesso (si ricordi che l'operazio-

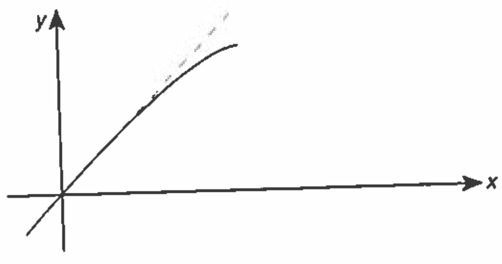

Figura 2.3 - Caratteristica ingresso x / uscita y di un amplificatore.

ne di moltiplicazione n
delle frequenze, ad un

E così via analogam
caratteristica non linea
con uno sviluppo in se

si ha per lo spettro del

$Y(f)=aX(f)+$

Alla componente di
za generati dall'interazi
so la non linearità ("in

2.5 LA TRASMISSION

Premessa

Per i segnali analogic
mente emessi dalla sorg
in ricezione di sottoporre
in modo da minimizzare
storsioni di canale e dei

Il caso della trasmissi
ad esso dedicheremo es
smissione numerica, si è
impulsi rappresentanti i s
vallo tra impulsi successi
sincrona. In ricezione oc
sta identificazione è osta
nale al segnale trasmesso,
dal canale possono gene
(interferenza intersimboli
identificare i simboli tras
tiva alla temporizzazione
mente la posizione temp
simbolo).

La *fig. 2.4* mostra lo sc
smissione, di forma $g(t)$,

ne di moltiplicazione nel dominio del tempo corrisponde, nel dominio delle frequenze, ad una operazione di convoluzione):

$$Y(f) = X(f) * X(f)$$

E così via analogamente per nonlinearità di ordine superiore. Se la caratteristica non lineare $y = g(x)$ può essere espressa o approssimata con uno sviluppo in serie del tipo

$$y = ax + bx^2 + cx^3 + \dots$$

si ha per lo spettro del segnale $y(t)$

$$Y(f) = \mathrm{a}X(f) + \mathrm{b}X(f)*X(f) + \mathrm{c}X(f)*X(f)*X(f) + \dots$$

Alla componente di segnale $x(t)$ si aggiungono termini d'interferenza generati dall'interazione delle componenti spettrali di $X(f)$ attraverso la non linearità ("intermodulazione").

2.5 LA TRASMISSIONE IN BANDA BASE

Premessa

Per i segnali analogici, inviati sul canale così come sono originariamente emessi dalla sorgente, ci limitiamo ad osservare che si tratterà in ricezione di sottoporre il segnale ricevuto ad un opportuno filtraggio in modo da minimizzare gli effetti (l'errore quadratico medio) delle distorsioni di canale e dei disturbi.

Il caso della trasmissione numerica richiede una analisi più estesa e ad esso dedicheremo essenzialmente la nostra attenzione. Nella trasmissione numerica, si è in presenza normalmente di una sequenza di impulsi rappresentanti i simboli trasmessi al ritmo di $R = \frac{1}{T}$. T è l'intervallo tra impulsi successivi, che è costante nella trasmissione cosiddetta sincrona. In ricezione occorre identificare il simbolo trasmesso, e questa identificazione è ostacolata sia dal rumore che si è aggiunto sul canale al segnale trasmesso, sia anche dal fatto che le distorsioni prodotte dal canale possono generare interferenze mutue tra impulsi adiacenti (interferenza intersimbolica). Si noti inoltre che il ricevitore, per poter identificare i simboli trasmessi, dovrà disporre dell'informazione relativa alla temporizzazione della sequenza onde poter conoscere esattamente la posizione temporale di ciascun impulso (sincronizzazione di simbolo).

La *fig. 2.4* mostra lo schema che si vuole studiare. Gli impulsi in trasmissione, di forma $g(t)$, sono "formati" da un filtro di funzione di

Figura 2.4 - Sistema di trasmissione impulsiva in banda base.

trasferimento $H_T(f)$, cioè generati da impulsi ideali inviati con frequenza $\frac{1}{T}$ in ingresso al filtro. $g(t)$ rappresenta quindi la risposta impulsiva, ossia la trasformata di Fourier di $H_T(f)$, mentre con a_k indichiamo l'ampiezza attribuita all'impulso k-esimo, determinata dall'informazione da trasmettere. Se M sono i possibili valori dell'ampiezza a_k, ogni impulso può trasportare $\log_2 M$ bit d'informazione. Il segnale inviato sul canale potrà essere rappresentato come

$$s(t) = \sum_k a_k g(t - kT)$$

In ricezione sarà necessario un opportuno filtraggio H_R per ridurre al minimo l'effetto del rumore che si aggiungerà, sul canale, al segnale; quindi si campionerà l'uscita nell'istante più opportuno, quello in cui il segnale assume il valore massimo, per poter stimare il valore più probabile del livello trasmesso a_k. Il campionatore è controllato da un opportuno sistema di sincronizzazione di simbolo. Il valore $r(kT)$ letto dal campionatore in corrispondenza del k-esimo simbolo può essere espresso come

$$r(kT) = a_k + \sum_{i \neq k} a_i p[(i-k)T] + n_k$$

$p(t)$ rappresenta la forma d'onda impulsiva in uscita, che dipenderà naturalmente, oltre che da H_T, anche dalla funzione di trasferimento del canale H_c e del filtro H_R. Lo spettro di Fourier di $p(t)$ è dato da $H_T H_c H_R$. In generale oltre al valore dell'ampiezza a_k dell'impulso k-esimo, si avranno all'istante kT le "code" (*fig. 2.5*) degli impulsi tra-

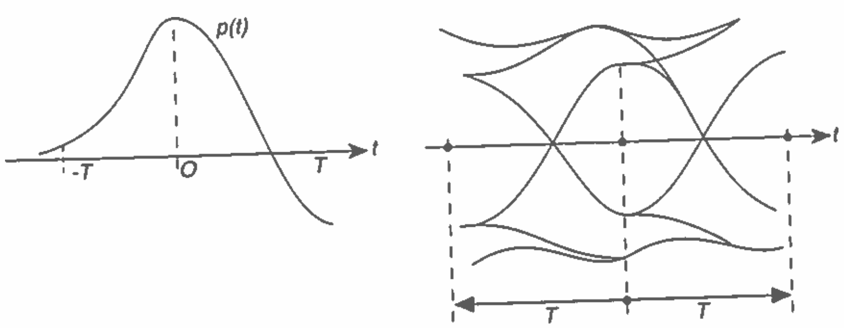

Figura 2.5 - Interferenza intersimbolica e diagramma a occhio.

smessi in intervalli adiacenti cioè un termine $\sum a_i p[(i-k)T]$ che rappresenta l'interferenza intersimbolica. Il termine n_k rappresenta gli effetti del rumore che si aggiunge al segnale trasmesso, sul canale e nei circuiti del ricevitore. Un diagramma del tipo illustrato in fig. 2.5 (indicato spesso come "diagramma ad occhio" a causa della sua forma tipica) può essere costruito, data una generica sequenza casuale di simboli in trasmissione, sovrapponendo gli andamenti del segnale in uscita dal filtro di ricezione, nei vari intervalli di simbolo di durata T. Esso dà un'idea approssimativa delle prestazioni del sistema in quanto mostra l'effetto dell'interferenza intersimbolica sulla distanza tra i possibili livelli di segnale negli istanti di lettura degli impulsi: quanto maggiore è l'apertura dell'"occhio" tanto migliori saranno le prestazioni in presenza di rumore.

2.5.1 Trasmissione impulsiva in banda base

I fattori che limitano le possibilità di trasmissione su un dato canale trasmissivo sono le distorsioni di segnale generate dal canale ed il rumore che si aggiunge al segnale trasmesso. In molti casi le caratteristiche trasmissive di un canale sono racchiuse e descritte con un unico parametro, la sua banda B; ciò corrisponde ad usare un modello di canale dato da un filtro passabasso ideale di banda B. In questo canale ogni segnale con spettro interamente contenuto in B viene trasmesso senza alcuna distorsione (a parte inessenziali ritardi), mentre tutte le componenti spettrali (rumore o altro) a frequenza superiore a B vengono cancellate. In alcuni casi questo modello costituisce una approssimazione accettabile, in altri la funzione di trasferimento del canale H_c si discosta alquanto da quella di un filtro passabasso ideale.

Banda e velocità di trasmissione. Criterio di Nyquist

Si è gia visto quali caratteristiche deve avere la funzione di trasferimento di un canale affinché una forma d'onda vi possa transitare senza distorsioni. Possiamo definire la banda B di trasmissione di un canale come l'intervallo di frequenze in cui esso presenta caratteristiche ideali non distorcenti. In pratica tale definizione tende ad essere astratta e poco utile in quanto raramente s'incontrano caratteristiche assolutamente ideali, ed inoltre spesso si potrà porre rimedio a distorsioni lineari prodotte dal canale trasmissivo mediante opportune trasformazioni in ricezione (filtraggio per l'equalizzazione del canale) atte a compensare la distorsione introdotta dal canale. La banda di un canale spesso può essere definita solo in modo convenzionale e ciò è comprensibile data l'impossibilità di rappresentare con un unico parametro B la funzione di trasferimento del canale stesso. Diciamo allora approssimativamente che la banda di un canale è quell'intervallo di frequenze che il canale fa transitare senza drastiche attenuazioni, rendendole quindi utilizzabili in ricezione. Del resto una simile ambiguità s'incontra nel definire la banda occupata da un dato segnale, dal momento che lo spettro può estendersi all'infinito anche se il grosso delle componenti spettrali è concentrato in una banda limitata: potremo dire per esempio che banda di un segnale è quell'intervallo di frequenze in cui è contenuto il 99% della potenza o energia (a seconda che si tratti di segnali continui o impulsivi).

Ci proponiamo anzitutto di analizzare un canale ideale di banda B, cioè un filtro passabasso ideale di banda B, determinando a quale velocità (misurata in impulsi al secondo, ovverosia in "baud" secondo la terminologia standard) è possibile trasmettere impulsi, senza che questi interferiscano tra loro in ricezione ostacolando l'identificazione di ciascuno. Si faccia riferimento ad una trasmissione di impulsi di forma $Ag(t)$ al ritmo di $R=\frac{1}{T}$ baud. L'informazione sarà contenuta nell'ampiezza A: per esempio nel caso di trasmissione binaria essa potrà assumere i valori 1 e -1.

Si ha che, su un canale ideale di banda B, è possibile trasmettere fino alla velocità di $2B$ impulsi/s (detta velocità di Nyquist) senza interferenza intersimbolica. Si considerino infatti impulsi con spettro perfettamente rettangolare, di banda B, e quindi transitanti nel canale senza distorsioni. La forma d'onda impulsiva $g(t)$ relativa ad uno spettro rettangolare di banda B è data da (*fig. 2.6*):

$$g(t)=\frac{\sin 2\pi Bt}{2\pi Bt}$$

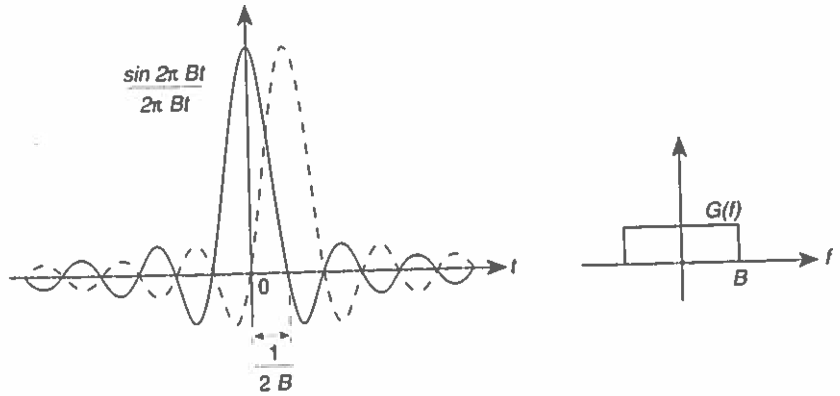

Figura 2.6 - Impulsi ideali a spettro rettangolare.

Come si vede, la forma d'onda, che presenta zeri equidistanziati di $\frac{1}{2B}$ sec, si estende all'infinito e quindi interferisce con altri impulsi trasmessi in istanti diversi. Si noti però che in trasmissione numerica, come si vedrà ampiamente più avanti, si tratterà in ricezione di identificare il livello d'ampiezza dell'impulso ricevuto, che è quello che porta l'informazione; tale livello lo si andrà ovviamente a determinare campionando il segnale ricevuto nell'istante in cui la forma d'onda impulsiva $g(t)$ presenta il suo massimo. Se ora gli impulsi trasmessi in sequenza sono distanziati di $\frac{1}{2B}$, si ha che negli istanti utili di campionamento $t_n = \frac{n}{2B}$ non si ha interferenza dovuta a simboli adiacenti in quanto questi danno contributo nullo in quei particolari istanti.

In pratica si hanno difficoltà a realizzare una caratteristica spettrale di tipo rettangolare, ed inoltre la condizione di interferenza intersimbolica nulla si verifica solo negli istanti di campionamento $\frac{n}{2B}$; a causa di un'imperfetta sincronizzazione di simbolo in ricezione per cui il campionamento non avvenga esattamente negli istanti teoricamente ottimali, l'interferenza può manifestarsi in modo rilevante data l'entità delle code dell'impulso a spettro rettangolare. Però, adottando impulsi con spettri opportunamente "arrotondati" e praticamente realizzabili, si possono ottenere forme d'onda impulsive ancora con interferenza intersimbolica nulla negli istanti di campionamento, ma con code molto attenuate rispetto al caso di spettro rettangolare e quindi con minor sensibilità agli errori di sincronizzazione.

Si consideri ad esempio uno spettro arrotondato del tipo illustrato in *fig. 2.7*, che presenta una simmetria di tipo dispari rispetto alla frequenza $\frac{1}{2T} \equiv B_0$:

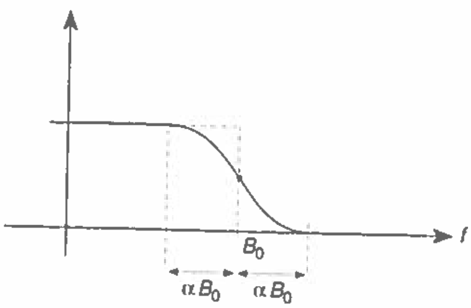

Figura 2.7 - Impulso a spettro "arrotondato" (a simmetria dispari).

$$G(f)=\begin{cases} 1 & 0\leq f<(1-\alpha)B_0 \\ \dfrac{1}{2}-\dfrac{1}{2}\sin\left(\dfrac{\pi}{2}\,\dfrac{f-B_0}{\alpha B_0}\right) & (1-\alpha)\,B_0\leq f\leq(1+\alpha)\,B_0 \\ 0 & f>(1+\alpha)\,B_0 \end{cases}$$

α rappresenta il fattore di arrotondamento ("roll-off"). A questo spettro corrisponde la forma d'onda impulsiva

$$g(t)=\frac{\sin 2\pi B_0 t}{2\pi B_0 t}\;\frac{\cos 2\pi\alpha B_0 t}{1-16\alpha^2 B_0^2 t^2} \qquad [2.1]$$

Si aumenta così la banda occupata rispetto alla spettro rettangolare (da B_0 a $(1+\alpha)B_0$), ma si mantengono gli zeri della forma d'onda negli istanti $\frac{n}{2B_0}$, con code molto ridotte rispetto a quelle dell'impulso $\frac{\sin 2\pi B_0 t}{2\pi B_0 t}$ (dalla [2.1] si vede infatti che le code decrescono nel tempo con la terza potenza del tempo). Per il caso $\alpha=1$, lo spettro ha la forma di un coseno rialzato:

$$G(f)=\cos^2\left(\frac{\pi f}{4B_0}\right)=\frac{1}{2}+\frac{1}{2}\cos\left(\frac{\pi f}{2B_0}\right);\quad B=2B_0$$

$$g(t)=\frac{\sin 2\pi B_0 t}{2\pi B_0 t}\;\frac{1}{1-16B_0^2 t^2}$$

Più in generale, qualsiasi spettro ottenuto dallo spettro ideale rettangolare aggiungendovi una funzione dispari rispetto a $B_0=\frac{1}{2T}$ (il caso del coseno rialzato è un caso particolare) preserva la proprietà degli zeri equidistanti con intervallo pari a T. È questo il cosiddetto criterio di Nyquist che si può dimostrare osservando che la forma d'onda desiderata $g(t)$ deve soddisfare la condizione $g(kT)=0$ per $k\neq 0$ ossia

$$= g(t)\sum\delta(t-kT)=g(0)\;\delta(t)$$

Trasformando secondo Fourier tale relazione si ottiene

$$G(f)*\frac{1}{T}\sum_{-\infty}^{\infty}{}_m\,\delta\left(f-\frac{m}{T}\right)=\frac{1}{T}\sum_{-\infty}^{\infty}{}_m\,G\left(f-\frac{m}{T}\right)=g(0)$$

Si ha cioè che lo spettro $G(f)$ dell'impulso $g(t)$, se periodizzato con periodo T, deve dar luogo ad una costante sull'asse delle frequenze. Uno spettro con le caratteristiche dispari di cui sopra soddisfa a tale condizione (*fig. 2.8*).

In definitiva si possono dunque teoricamente trasmettere, su un canale di banda B, $2B$ impulsi/s, senza interferenza intersimbolica, usando impulsi ideali a spettro rettangolare.

Figura 2.8 - Spettro di im

Tali impulsi po
pendente dal numer
assumere. Usando
bit/s. Ma, come vec
bili trova un limite

2.5.2 Progetto ottir

Se assumiamo cor
re e le caratteristiche
ne il problema del pro
to il problema per un
terferenza causata, s
casi, per esempio nei
impulsi isolati. Il tras
in parte dal bersaglio
viene rivelato dal rice
sione e ricezione dell'
ne costituisce quindi u
zione si ha invece a c
quanto si otterrà vari
quenza solo quando
Le prestazioni de
rapporto segnale-rur

[1] Quando nei sistemi di c
porti di potenze ecc., è u
bel" (abbreviato in "dB"
ze $\frac{P_1}{P_2}$ viene misurato in

A volte si usa anche espri
ad un valore di potenza d
potenza misurata in dBm

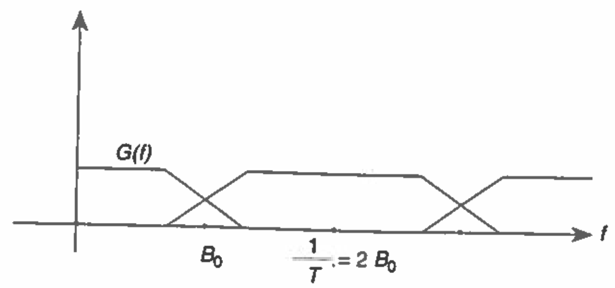

Figura 2.8 - Spettro di impulso a interferenza intersimbolica nulla (Criterio di Nyquist).

Tali impulsi possono trasportare una quantità d'informazione dipendente dal numero di livelli d'ampiezza che gli impulsi stessi possono assumere. Usando M livelli si possono quindi trasmettere $2B\log_2 M$ bit/s. Ma, come vedremo nel seguito, il numero di livelli in pratica usabili trova un limite nel rapporto segnale/rumore disponibile.

2.5.2 Progetto ottimale del ricevitore (filtro): il filtro "adattato"

Se assumiamo come dati la forma d'onda impulsiva usata al trasmettitore e le caratteristiche del canale (funzione di trasferimento e rumore), si pone il problema del progetto ottimale del ricevitore. Considereremo anzitutto il problema per un impulso singolo, senza considerare cioè l'eventuale interferenza causata, sull'impulso in esame, da impulsi adiacenti. In alcuni casi, per esempio nei sistemi radar, ci si trova proprio nella situazione di impulsi isolati. Il trasmettitore radar emette un impulso radio che, riflesso in parte dal bersaglio (oggetto che si vuol rivelare), ritorna verso il radar e viene rivelato dal ricevitore. L'intervallo di tempo che intercorre tra emissione e ricezione dell'impulso è proporzionale alla distanza del bersaglio e ne costituisce quindi una misura. Normalmente nei sistemi di telecomunicazione si ha invece a che fare con sequenze continue di impulsi, e quindi quanto si otterrà varrà anche per il caso di trasmissione di impulsi in sequenza solo quando non ci sia interferenza intersimbolica tra di essi.

Le prestazioni del ricevitore in presenza di rumore dipendono dal rapporto segnale-rumore[1] all'uscita del filtro H_R che sarà necessario

---

[1] Quando nei sistemi di comunicazione si considerano rapporti segnale/rumore, rapporti di potenze ecc., è usuale adoperare unità di misura logaritmiche quale il "decibel" (abbreviato in "dB") poiché possono risultare più comode: un rapporto di potenze $\frac{P_1}{P_2}$ viene misurato in dB come

$$10\log_{10}\frac{P_1}{P_2}\ \text{dB}$$

A volte si usa anche esprimere valori assoluti di potenza come rapporto in dB rispetto ad un valore di potenza di riferimento, per esempio 1 mW (in questo caso si parla di potenza misurata in dBm).

usare in ricezione per ridurre il più possibile gli effetti del rumore. Intuitivamente si comprende che, restringendo la banda di tale filtro, si ridurrà la potenza di rumore in uscita dal filtro di ricezione; ma andando oltre certi limiti, che dipendono dalla banda del filtro in relazione alla larghezza dello spettro del segnale, si produrrà anche una diminuzione della componente utile di segnale all'uscita del filtro di ricezione. Se si considera per esempio l'uscita di un filtro passabasso *RC* per un impulso rettangolare in ingresso, si ha che, al crescere della costante di tempo *RC*, la banda diminuisce, il rumore in uscita diminuisce ma l'impulso utile in uscita si distorce sempre più (*fig. 2.9*); esiste quindi un valore ottimale per la costante di tempo *RC* che rende massimo il rapporto segnale-rumore all'uscita del filtro (misurato naturalmente nell'istante in cui il segnale in uscita assume il suo valore massimo).

Più in generale, mettendo per il momento tra parentesi il problema dell'interferenza intersimbolica, esaminiamo quale sia la caratteristica più conveniente del filtro di ricezione H_R allo scopo di rendere massimo, all'uscita del filtro, il rapporto segnale-rumore.

Si consideri un impulso $g(t)$ di durata T e di energia $E=\int g^2(t)dt=\int|G(f)|^2df$, in arrivo al ricevitore in presenza di rumore bianco $n(t)$ di densità (bilaterale) $N_0/2$, e il suo passaggio attraverso un filtro di ricezione H_R. La componente di segnale utile all'uscita del filtro sarà un impulso $u_s(t)$ di cui consideriamo il valore di picco $u_s(t_0)$, assunto ad un determinato istante t_0.

Tale valore di picco può essere espresso, utilizzando l'analisi nel dominio delle frequenze, secondo la relazione $u_s(t_0)=\int G(f)H_R(f)e^{j\omega t_0}df$.

D'altro canto, il valore quadratico medio (potenza) della componente di rumore all'uscita del filtro sarà dato da

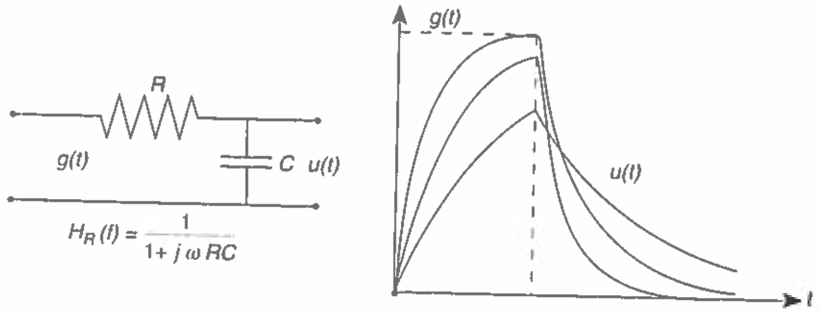

Figura 2.9 - Risposta di un filtro RC ad un impulso rettangolare, per vari valori della sua banda (funzione di RC).

$$\sigma_n^2 = \frac{N_0}{2}\int |H_R|^2 df$$

Converrà usare un filtro con funzione di trasferimento H_R tale da rendere massimo il rapporto segnale-rumore all'uscita, esprimibile come

$$\frac{u_s^2(t_0)}{\sigma_n^2} = \frac{|\int G(f)H_R(f)e^{j\omega t_0}df|^2}{\frac{N_0}{2}\int |H_R(f)|^2 df} \qquad [2.2]$$

Per determinare il massimo di questa espressione, si può per esempio applicare la nota diseguaglianza di Schwartz: si ha che

$$|\int G H_R e^{j\omega t_0} df|^2 \leq \int |G|^2 df \int |H_R|^2 df$$

valendo il segno di uguaglianza quando $H_R = [Ge^{j\omega t_0}]^* = G^*(f)e^{-j\omega t_0}$. G^* è il complesso coniugato dello spettro del segnale trasmesso. Risulta quindi che

$$\frac{u_s^2(t_0)}{\sigma_n^2} \leq \frac{2}{N_0}\int |G|^2 df = \frac{2E}{N_0} \qquad [2.3]$$

Il rapporto segnale/rumore in uscita raggiunge il suo massimo quando nella [2.3] vale il segno d'eguaglianza, cioè quando $H_R = G^* e^{-j\omega t_0}$, che rappresenta pertanto la funzione di trasferimento ottimale. Si noti che la caratteristica di ampiezza $|H_R|$ di tale filtro, denominato filtro "adattato", coincide con $|G|$, cioè ricopia esattamente lo spettro dell'impulso d'ingresso. Il termine $e^{-j\omega t_0}$ corrisponde semplicemente al ritardo t_0 connesso all'istante di campionamento.

Il valore del rapporto segnale-rumore che si ottiene con il filtro adattato è pari a $2E/N_0$ e dipende esclusivamente dal rapporto tra l'energia impiegata E e la densità di rumore N_0, ma non della forma $g(t)$ dell'impulso. Tale forma assume invece importanza nella trasmissione di impulsi in sequenza a causa dell'interferenza intersimbolica. Si osservi infine che si è inteso come segnale $g(t)$ quello in arrivo al ricevitore dal canale. Pertanto nel caso di canale distorcente, $g(t)$ andrà determinato in base allo spettro del segnale propriamente trasmesso e alla funzione di trasferimento del canale H_c.

ESERCIZIO

Si consideri il caso della trasmissione di un impulso $g(t)$ rettangolare e di durata T, in presenza di rumore bianco, e si usi in ricezione, anziché un filtro adattato, un semplice filtro RC (fig. 2.9), campionato all'istante $t = T$. Si determini il rapporto segnale/rumore all'uscita e lo si confronti con il valore ottenibile con il filtro ottimo.

La risposta impulsiva del filtro adattato si ottiene con la trasformata di Fourier della H_R, ed è quindi data da

$$h(t) = g(t_0 - t)$$

Essa coincide cioè con l'impulso $g(t)$ rovesciato nel tempo e ritardato di t_0. Per la fisica realizzabilità del filtro, quando debba operare in tempo reale, occorre che $h(t)$ sia nulla per $t<0$ e quindi t_0 sia maggiore o uguale alla durata dell'impulso $g(t)$.

Ne segue che l'uscita $u_s(t)$ di segnale del filtro adattato risulta essere la funzione di autocorrelazione $R_g(t-t_0)$ dell'impulso $g(t)$; si ha nel dominio del tempo e delle frequenze

$$u_s(t) = g(t) * h(t) = \int g(\tau) g(t_0 - t + \tau) d\tau$$
$$U_s(f) = GG^* e^{-j\omega t_0} = |G|^2 e^{-j\omega t_0}$$

Nella *fig. 2.10* è riportato l'esempio di impulso $g(t)$ rettangolare. Quanto sopra fa riferimento alla ricezione di un singolo impulso, di forma prefissata. Considerando il caso della trasmissione di impulsi in sequenza, c'è da tenere conto anche della interferenza intersimbolica, e si pone il problema di scegliere nel modo migliore sia la forma d'onda in trasmissione (cioè il filtraggio H_T) sia il filtro H_R in ricezione. Questo vale naturalmente nel caso in cui la forma d'onda in trasmissione non sia per qualche ragione predeterminata. Sia $H_N(f)$ uno spettro soddisfacente il criterio di Nyquist, più sopra illustrato. Per avere interferenza intersimbolica nulla, dovrà essere $H_T H_R = H_N$ (si noti che è lo spettro all'uscita del filtro di ricezione che deve soddisfare il criterio di Nyquist!). Se si ha $|H_T| = |H_R| = \sqrt{H_N}$, e si scelgono le caratteristiche di fase di H_T e H_R arbitrariamente purché si compensino, la condizione di Nyquist è realizzata, e si ha anche l'adattamento del filtro di ricezione alla forma d'onda trasmessa sul canale poiché $H_R = H_T^*$. La scelta di una caratteristica di Nyquist e la sua "suddivisione" a metà tra trasmissione e ricezione è appunto il criterio usato in molti sistemi di trasmissione dati.

Nella trattazione precedente del filtro adattato si è assunto per semplicità che il rumore fosse bianco. Se il rumore non è bianco ma

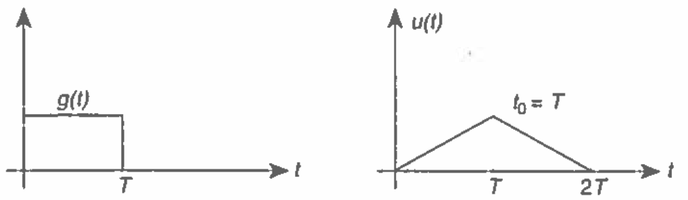

Figura 2.10 - Ingresso e uscita del filtro adattato nel caso di impulso rettangolare.

ha spettro di potenza $N(f)$, basta considerare nel numeratore della [2.2] in luogo dell'espressione $GH_Re^{-j\omega t_0}$ l'espressione equivalente $\frac{G}{\sqrt{N}}H_R\sqrt{N}e^{-j\omega t_0}=\tilde{G}\tilde{H}_Re^{-j\omega t_0}$, con $\tilde{H}_R=H_R\sqrt{N}$ e $\tilde{G}=\frac{G}{\sqrt{N}}$. Naturalmente al denominatore comparirà la potenza di rumore $\sigma_n^2=\int N|H_R|^2df=\int|\tilde{H}_R|^2df$. La massimizzazione del rapporto segnale/rumore porta, con lo stesso procedimento adottato nel caso di rumore bianco, al risultato $\tilde{H}_R=\tilde{G}^*e^{-j\omega t_0}$, cioè

$$H_R=\frac{G^*}{N}\,e^{-j\omega t_0} \qquad [2.4]$$

È interessante considerare questo risultato anche da un altro punto di vista. Possiamo anzitutto, all'ingresso del ricevitore, operare un pre-filtraggio di "imbiancamento" del rumore, cioè far passare ciò che arriva dal canale in un filtro $\frac{1}{\sqrt{N(f)}}$ che trasforma lo spettro del rumore in uno spettro costante (e naturalmente trasforma anche lo spettro di segnale in $\frac{G(f)}{\sqrt{N(f)}}$). Ci ritroviamo a questo punto nella situazione di rumore bianco già studiata, e possiamo quindi ricavare la funzione di trasferimento del filtro adattato allo spettro di segnale $\frac{G(f)}{\sqrt{N(f)}}$. La cascata di questo e del filtro di imbiancamento dà poi la funzione di trasferimento complessiva [2.4] del filtro ottimale. Se si esamina il modulo della funzione di trasferimento del filtro adattato, si può notare che esso è proporzionale, frequenza per frequenza, al rapporto tra spettro del segnale e densità spettrale del rumore.

Equivalenza di filtro adattato e ricevitore a correlazione

Il funzionamento del filtro adattato può anche essere interpretato in un altro modo, che aiuta a meglio comprendere le ragioni della sua ottimalità. Infatti, poiché la risposta impulsiva del filtro è data da $h(t)=g(t_0-t)$, l'uscita del filtro nell'istante t_0 può essere espressa come

$$u(t_0)=\int_{-\infty}^{t_0}r(\tau)h(t_0-\tau)d\tau=\int_{-\infty}^{t_0}r(\tau)g(\tau)d\tau$$

essendo $r(t)=g(t)+n(t)$ il segnale in arrivo al ricevitore. Si ha quindi che l'operazione del filtro adattato equivale ad una correlazione del segnale ricevuto con una copia della forma d'onda $g(t)$ usata in trasmissione (*fig. 2.11a*).

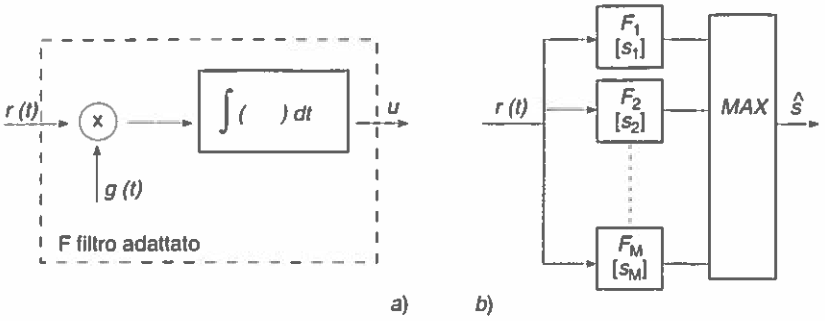

Figura 2.11 - Ricevitore a correlazione.

Il valore della correlazione tra due forme d'onda costituisce una misura della similitudine tra esse. Abbiamo visto che nella trasmissione, binaria o a più livelli, di una data forma d'onda, il ricevitore ottimale può essere visto sostanzialmente come un correlatore. Lasciando a trattazioni più specifiche lo studio della teoria generale della ricezione ottimale in presenza di rumore, ci limitiamo qui a suggerire intuitivamente l'estensione del concetto a casi più generali. In un sistema discreto più generale della trasmissione impulsiva a più livelli finora considerata, si avrà a che fare con più forme d'onda diverse in trasmissione. È naturale pensare allora che uno schema di ricevitore efficiente può essere costituito da un insieme di correlatori, uno per ciascuna forma d'onda possibile in trasmissione. Le uscite dei correlatori danno una misura della similitudine tra segnale ricevuto e ciascuno dei possibili segnali trasmessi, e quindi dal loro confronto si può determinare quale segnale sia stato più probabilmente trasmesso. Si consideri il caso in cui si abbiano in trasmissione M possibili simboli o messaggi rappresentati da M segnali $s_1(t)$, $s_2(t)$, ... $s_M(t)$, tutti con la stessa energia e la stessa probabilità a priori $\frac{1}{M}$. Il ricevitore ottimale che rende minima la probabilità d'errore è costituito da un banco di M correlatori tra $r(t)$ e $s_i(t)$. I valori delle M correlazioni

$$u_i = \int r(\tau) s_i(\tau) d\tau$$

vengono confrontati e si sceglie, quale messaggio trasmesso, il messaggio, tra gli s_i, corrispondente al massimo tra i valori delle M correlazioni u_i (*fig. 2.11b*).

Se le energie e le probabilità a priori dei segnali $s_i(t)$ non sono eguali, se ne tiene conto pesando opportunamente le uscite u_i dei correlatori. Si può vedere che, con certe ipotesi sulla statistica del rumore, le operazioni di questo tipo di ricevitore equivalgono al calcolo, in base al segnale ricevuto $r(t)$, della probabilità a posteriori $P[s_i/r]$ per i vari segnali $s_i(t)$ possibili in trasmissione e alla determinazione del valore massimo tra di esse. Ciò permette di ottenere in ricezione la minima probabilità d'errore.

2.5.3 Gli effetti del rumore sulle prestazioni del sistema di trasmissione

Consideremo come esempio introduttivo un caso tipico e particolarmente importante, la trasmissione binaria in presenza di rumore termico. La trattazione è facilmente estendibile al caso di impulsi a più livelli. I simboli binari 1 e 0 siano rappresentati rispettivamente da un impulso $g(t)$ e dal suo opposto $-g(t)$, cioè da due segnali antipodali. In ricezione si userà un filtro passabasso, per esempio un filtro adattato, per l'eliminazione delle componenti di rumore esterne alla banda occupata dallo spettro del segnale $g(t)$. L'uscita del filtro viene campionata nell'istante in cui il segnale utile assume il suo massimo valore A, e il campione prelevato sarà $u = \pm A + n$, essendo n la componente di rumore. La componente di rumore n è una variabile casuale gaussiana con valor medio nullo e varianza data da

$$\sigma_n^2 = \int N(f)\,|H_R(f)|^2 df$$

essendo H_R la funzione di trasferimento del filtro e $N(f)$ lo spettro di potenza del rumore sul canale.

La decisione sul valore del simbolo ricevuto viene presa mediante un circuito di soglia che darà in uscita 1 o 0 a seconda che l'uscita u del filtro sia positiva o negativa. Infatti se si esaminano le densità di probabilità dell'uscita u condizionate rispettivamente dalla trasmissione del simbolo 1 oppure 0, cioè $p(u/1)$ e $p(u/0)$ (*fig. 2.12*), si vede subito che per $u>0$, $p(u/1)$ è maggiore di $p(u/0)$ e viceversa. La regola di decisione suddetta equivale quindi a scegliere come segnale trasmesso quello più "verosimile". La densità di probabilità dell'uscita condizionata al segnale trasmesso viene indicata come funzione di verosimiglianza.

Se si conoscono le probabilità a priori dei simboli trasmessi, in questo caso $P(1)$ e $P(0)$, si può raffinare la regola di decisione tenendo conto di esse. Si sceglierà il simbolo che, in base all'uscita u, risulta a posteriori il più probabile, cioè nel nostro caso si confronterà la probabilità a posteriori $p(1/u) = \dfrac{P(1)p(u/1)}{p(u)}$ con la probabilità a posteriori $p(0/u) = \dfrac{P(0)p(u/0)}{p(u)}$, decidendo in conseguenza a favore di 1 o di 0.

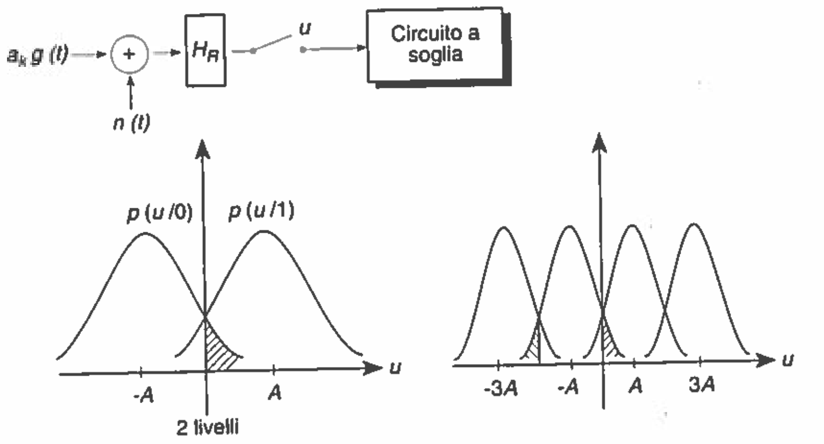

Figura 2.12 - Ricezione di impulso a più livelli in presenza di rumore gaussiano.

ESERCIZIO
Per un ricevitore binario si determini la regola di decisione in base al criterio della massima probabilità a posteriori, cioè in pratica il valore della soglia da usare nel circuito di soglia finale in funzione di $P(1)$ e $P(0)$.

Considerando ora la probabilità con cui si commette errore nella decisione a massima verosimiglianza, si vede chiaramente che, nel caso si trasmetta il simbolo 0, si sbaglia quando $u>0$; e ciò avviene con probabilità

$$P[E/0]=\int_0^\infty \frac{1}{\sqrt{2\pi\sigma_n^2}}\, e^{-\frac{(u+A)^2}{2\sigma_n^2}}du=\int_{\frac{A}{\sigma_n}}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}dx\equiv Q\left(\frac{A}{\sigma_n}\right)$$

NOTA
La funzione

$$Q(x)\equiv\int_x^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}dy$$

rappresenta il complemento all'unità dell'integrale della funzione di Gauss $\frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$ cioè l'area della "coda" della funzione. Quindi è la probabilità che una variabile gaussiana di valor medio nullo e varianza unitaria assuma valori maggiori di x.
La funzione Q decresce molto rapidamente al crescere di x. Per valori non piccoli di x, un'ottima approssimazione è data dalla funzione:

$$Q(x)\cong \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}x} \qquad [2.5]$$

La *fig. 2.13* mostra l'andamento di $Q(x)$ e della sua approssimazione [2.5].

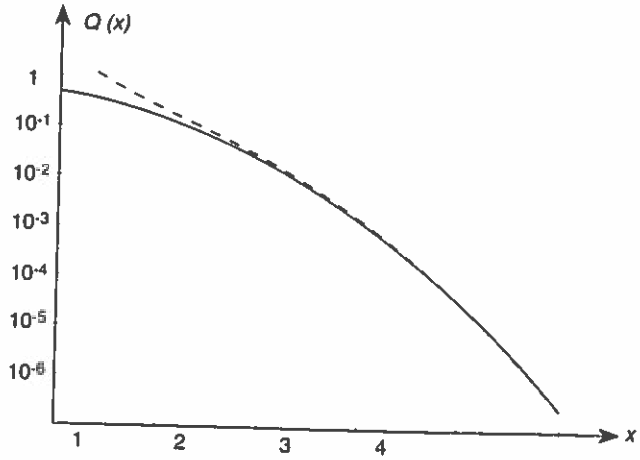

Figura 2.13 - La funzione Q (x) e la sua approssimazione $\dfrac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}\;x}$

Un risultato analogo vale quando si trasmette 1, e quindi la probabilità d'errore totale sarà nel caso binario esaminato (soglia di decisione $u=0$)

$$P(E)=P(1)P(E/1)+P(0)P(E/0)=Q\left(\frac{A}{\sigma_n}\right)$$

Se il filtro usato in ricezione è quello ottimo, il rapporto segnale/rumore in uscita, in presenza di rumore bianco con densità $\frac{N_0}{2}$, è dato come si è visto da

$$\frac{u_s^2}{\sigma_n^2}=\frac{A^2}{\sigma_n^2}=\frac{2E_s}{N_0}$$

essendo E_S l'energia dell'impulso $g(t)$.

Per esempio si ha $P(E)=10^{-5}$ per $\frac{A}{\sigma_n}=4.27$, cioè per $\frac{E_S}{N_0}=\frac{A^2}{2\sigma_n^2}=$ $=9.12$ (9,6 dB).

Può essere utile esprimere le prestazioni in funzione della "distanza" d tra i livelli possibili; in questo caso $d=2A$. In generale, per i vari sistemi di trasmissione numerica, è ragionevole pensare che le prestazioni saranno tanto migliori quanto maggiori saranno, per un dato rumore, le "distanze" tra i possibili segnali.

ESERCIZIO

Si consideri un sistema di trasmissione binario in cui i simboli 1 e 0 siano rispettivamente rappresentati da un impulso $g(t)$ e da un segnale nullo. Si confrontino le prestazioni con quelle relative ad un sistema antipodale, in cui i simboli binari sono rappresentati da $g(t)$ e dal suo opposto $-g(t)$.

L'analisi fin qui condotta si può facilmente estendere a casi più generali. Se si considera la modulazione in ampiezza a più livelli dell'impulso $g(t)$, cioè la trasmissione di un impulso $a_k g(t)$ che possa assumere un numero di livelli $M>2$, $a_k = \pm 1, \pm 3, \ldots$ (fig. 2.12), il ricevitore sarà simile a quello di fig. 2.12, solo che nel circuito di decisione avremo più valori di soglia, pari a $0, \pm 2, \pm 4, \ldots$ per poter discriminare le M possibili ampiezze dell'impulso trasmesso. Si vede subito che la probabilità d'errore vale, nel caso di livelli a_k a priori egualmente probabili $\left(P(a_k) = \frac{1}{M}\right)$

$$P(E) = \frac{1}{M}\left\{(M-2)2Q\left(\frac{A}{\sigma_n}\right) + 2Q\left(\frac{A}{\sigma_n}\right)\right\} = 2\,\frac{M-1}{M}\,Q\left(\frac{d}{2\sigma_n}\right) \qquad [2.6]$$

NOTA
Equalizzazione di canale

Nell'analizzare la struttura e le prestazioni di un sistema di trasmissione numerico, non si è tenuto conto dell'interferenza intersimbolica, ipotizzando quindi che questa sia nulla o trascurabile. In molti casi ciò non è, e risulta allora necessario compensare la distorsione introdotta dal canale onde annullarla o almeno ridurla entro limiti accettabili. Se la distorsione è nota a priori, possiamo, con un opportuno filtro fisso di equalizzazione, compensare la distorsione rendendo ottimale la funzione di trasferimento globale del sistema. Ma a volte la distorsione di canale non è nota e varia anche nel tempo. In questo caso si può far in modo che l'equalizzatore si adatti automaticamente alle condizioni del canale.

Tipicamente per ridurre l'interferenza intersimbolica presente nei campioni $r(kT)$ in uscita dal filtro adattato si può effettuare un filtraggio lineare di tali campioni: utilizzando degli elementi di ritardo (*fig. 2.14*), si effettua una combinazione lineare $\sum w_i r[(k-i)T]$ scegliendo opportunamente i coefficienti w_i, per rendere minimo per esempio l'errore quadratico medio causato sia dall'interferenza intersimbolica sia dal rumore presente sul canale (filtro trasversale). Se le condizioni del canale variano nel tempo, si può realizzare un controllo adattativo dei pesi w_i, che vengono via via aggiornati sulla base di una stima delle condizioni del canale.

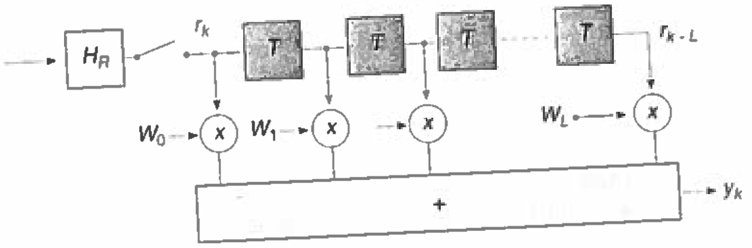

Figura 2.14 - Schema di equalizzatore lineare.

Se indichiamo con a_k i
ideali dovrebbe aversi $r($

e i pesi w_i saranno scel
$\Lambda(k) = E[\epsilon_k^2]$. Nel caso di e
trollo nel tempo dei w_i.
te": il gradiente $\nabla_i(k)$ de
spetto al coefficiente i-es

L'algoritmo consiste
qualizzatore mediante u
del gradiente $\nabla_i(k)$. In t
minimo della funzione Λ
nale alla correlazione m
possibile stima istantane
dotto $\epsilon_k r_{k-i}$, per cui l'algo

w_i

essendo μ una costante
te verso il minimo, senz
non essendo il valore d
trasmettere ogni tanto p
a priori, sulla base dei
quindi in funzionamento
presume che i dati in u
quindi che l'errore $y_k - \hat{}$
cidente con l'errore verc
zioni nel tempo della fu
Si possono utilizzare
non lineare: la soluzione
ha la maggior probabilit
l'insieme degli $r(kT)$.

NOTA
Sincronizzazione di sim

La ricezione di un fluss
ne di simbolo per l'iden
Sia T la durata di simbo

Se nello spettro del seg
di simbolo $1/T$, con un
$1/T$ è possibile ricupera
rizzazione alla frequenz
può creare mediante u
per esempio far passa
$u = |s|$; in molti casi nel

Se indichiamo con a_k i valori di ampiezza degli impulsi di segnale (in condizioni ideali dovrebbe aversi $r(kT)=a_k$), avremo all'uscita dell'equalizzatore un errore

$$\epsilon_k = y_k - a_k = \sum_{0}^{L}{}_i w_i r_{r-i} - a_k$$

e i pesi w_i saranno scelti in modo da rendere minimo l'errore quadratico medio $\Lambda(k)=E[\epsilon_k^2]$. Nel caso di equalizzatore adattativo, occorrerà un algoritmo per il controllo nel tempo dei w_i. Una soluzione naturale è usare l'algoritmo del "gradiente": il gradiente $\nabla_i(k)$ dell'errore quadratico medio $\Lambda(k)$ rappresenta la derivata rispetto al coefficiente i-esimo w_i:

$$\nabla_i(k) = \frac{\partial\Lambda(k)}{\partial w_i(k)} \qquad i=0,1,2,\dots L$$

L'algoritmo consiste nel controllare ad ogni passo kT il peso i-esimo w_i dell'equalizzatore mediante una correzione proporzionale e di segno opposto a quella del gradiente $\nabla_i(k)$. In tal modo si tende a raggiungere iterativamente il punto di minimo della funzione $\Lambda(k)$. Dal calcolo del gradiente risulta che esso è proporzionale alla correlazione mutua tra l'errore ϵ_k ed il campione r_{k-i} cioè a $E[\epsilon_k r_{k-i}]$. Una possibile stima istantanea di tale correlazione è data dal valore istantaneo del prodotto $\epsilon_k r_{k-i}$, per cui l'algoritmo iterativo di aggiornamento dei pesi w_i è dato da

$$w_i(k+1) = w_i(k) + \mu\epsilon_k r_{k-i} \qquad i=0,1,\dots.L$$

essendo μ una costante scelta opportunamente per rendere l'algoritmo convergente verso il minimo, senza instabilità. C'è il problema di valutare l'errore $\epsilon_k = y_k - a_k$, non essendo il valore dell'informazione a_k disponibile in ricezione. Si può allora trasmettere ogni tanto per un certo intervallo di tempo una sequenza di simboli noti a priori, sulla base dei quali il ricevitore può mettere a posto i pesi w_i; passando quindi in funzionamento normale alla trasmissione dei simboli d'informazione, si presume che i dati in uscita dal ricevitore siano non errati con alta probabilità e quindi che l'errore $y_k - \hat{a}_k$ determinato dal dato in uscita $\hat{a}_k$ sia quasi sempre coincidente con l'errore vero $\epsilon_k = y_k - a_k$. In tal modo l'equalizzatore "insegue" le variazioni nel tempo della funzione di trasferimento del canale.

Si possono utilizzare per l'equalizzazione anche altre strutture circuitali di tipo non lineare: la soluzione ottimale è basata sulla stima della sequenza di simboli che ha la maggior probabilità di essere stata trasmessa, stima che viene ricavata dall'insieme degli $r(kT)$.

NOTA

Sincronizzazione di simbolo

La ricezione di un flusso sincronizzato di dati numerici comporta la sincronizzazione di simbolo per l'identificazione in ricezione degli istanti di lettura e di decisione. Sia T la durata di simbolo nella trasmissione sincrona di impulsi a due o più livelli:

$$s(t) = \sum_k a_k g(t-kT).$$

Se nello spettro del segnale trasmesso compare una riga spettrale alla frequenza di simbolo $1/T$, con un filtro passabanda a banda stretta centrato sulla frequenza $1/T$ è possibile ricuperare un'onda sinusoidale da cui ricavare gli impulsi di temporizzazione alla frequenza di simbolo. Se invece tale riga spettrale è assente, la si può creare mediante un'opportuna trasformazione *nonlineare* del segnale. Si può per esempio far passare il segnale in un quadratore $u=s^2$ o in un raddrizzatore $u=|s|$; in molti casi nel segnale $u(t)$ è contenuta una componente periodica di pe-

riodo T, ed è quindi possibile estrarne la frequenza di simbolo. Vi sono inoltre vari altri metodi di sincronizzazione basati su principi diversi dalla generazione di righe spettrali tramite nonlinearità.

2.6. CAPACITÀ DI UN CANALE DI TRASMISSIONE

La capacità di un canale di trasmettere informazione è determinata dalla banda e dal rumore in esso presente.

Si è visto che, su un canale di banda B, teoricamente possiamo inviare $2B$ impulsi/s, senza interferenza intersimbolica; se questi possono assumere M livelli, a ciascun impulso si possono associare $\log_2 M$ bit d'informazione e pertanto l'informazione trasmessa sarà

$$C = 2B\log_2 M \text{ bit/s}$$

Volendo far giungere tale quantità d'informazione al ricevitore, s'incontra però un ostacolo nel rumore che limita il numero di livelli M distinguibili in ricezione con una probabilità d'errore accettabile. Usando M livelli di ampiezza equidistanti di δ, i valori d'ampiezza degli impulsi trasmessi saranno $\pm\frac{\delta}{2}, \pm\frac{3}{2}\delta, \ldots, \pm\frac{M-1}{2}\delta$. La potenza (valor quadratico medio) corrispondente sarà

$$P_S = \frac{1}{M}\left\{2\left(\frac{\delta}{2}\right)^2 + 2\left(\frac{3\delta}{2}\right)^2 + \ldots\right\} = (M^2 - 1)\,\frac{\delta^2}{12}$$

Si ha quindi

$$C = B\log_2\left[1 + \frac{12P_S}{\delta^2}\right]$$

Si scelga δ, a pari valore di picco del segnale, in modo da avere in ricezione un numero tollerabile di errori in presenza di rumore di potenza σ_n^2. Per esempio si ha $P(E) = 10^{-5}$ per $\delta = 8.84\sigma_n$, e la capacità diventa

$$C = B\log_2\left[1 + \frac{P_S}{6.5\sigma_n^2}\right]$$

Questa relazione si riferisce ad un semplice tipo di trasmissione, cioè la trasmissione impulsiva modulata in ampiezza, senza alcuna particolare codificazione. Il tipo di dipendenza funzionale della capacità C dai parametri fondamentali del canale di trasmissione, e cioè banda e rapporto segnale/rumore, ha però un rilievo che trascende il particolare caso considerato. Ciò a causa di uno dei risultati fondamentali della teoria dell'informazione, il teorema di Shannon (1948) sulla capacità di canale. In teoria, si può dimostrare infatti che, su un canale ideale di banda B con rumore gaussiano a densità spettrale di potenza N_0 (unilaterale) costante e con potenza di segnale P_S, è possibile trasmettere informazione ad un ritmo che al limite può raggiungere il valore

$$C = B\log_2\left[1 + \frac{P_s}{\sigma_n^2}\right]; \; \sigma_n^2 = N_0 B \qquad [2.7]$$

con probabilità di errore nulla (!).

Ciò comporta però l'uso di metodi di codificazione dell'informazione di complessità al limite infinita con conseguenti tempi di decodificazione infiniti.

Il risultato [2.7] traduce per il caso particolare qui considerato di canale di banda B con rumore gaussiano a spettro costante il teorema di Shannon riguardante in generale la capacità di trasmissione dell'informazione per canali qualsiasi. Il teorema in sostanza dice che se si trasmette informazione, su un dato canale, ad una velocità (bit/s) inferiore ad un limite proprio di ciascun canale (appunto la capacità di canale), è possibile eliminare completamente l'effetto dei disturbi ottenendo una probabilità d'errore nulla. Ciò può essere ottenuto con l'uso di una opportuna codificazione dell'informazione emessa dalla sorgente. Non è invece possibile ottenere tale risultato per velocità superiori alla capacità. La capacità, secondo il teorema di Shannon, rappresenta un termine ideale di riferimento, con il quale confrontare le prestazioni dei sistemi pratici di codificazione di canale. Al limite posto da questo teorema si può tendere usando sistemi di modulazione e codifica sempre più complessi.

Confrontando il teorema di Shannon con la semplice espressione prima derivata per una trasmissione impulsiva a più livelli senza codificazione, si riscontra una perfetta analogia nella dipendenza funzionale dai fattori fondamentali che determinano le prestazioni trasmissive di un canale, cioè la banda e il rapporto segnale-rumore. La trasmissione senza codificazione comporta, a pari banda e rumore, un incremento della potenza di segnale necessaria pari ad un fattore 6,5, rispetto al limite di Shannon (per una probabilità d'errore pari a 10^{-5} anziché nulla). Tale fattore esprime il guadagno massimo che si può ottenere (in termini di minor potenza necessaria o maggior rumore tollerabile) usando codici efficienti per trasmettere l'informazione sul canale.

ESEMPIO

Si consideri un canale telefonico della rete ordinaria, per il quale si può dire che la banda utilizzabile è circa 3 kHz. Se il rapporto segnale-rumore sul canale è per esempio pari a 200 (23 dB), la formula di Shannon dà per tale canale una capacità di circa 23 kbit/s.

La *fig. 2.15* mostra l'andamento della capacità in funzione della banda B: C è funzione crescente con B e tende per $B \rightarrow \infty$ al limite $\frac{P_S}{N_0 \ln 2}$. Questo valore asintotico dà il limite estremo a cui si può arrivare per la capacità di trasmissione quando l'unico vincolo sia posto dal rumore bianco presente sul canale.

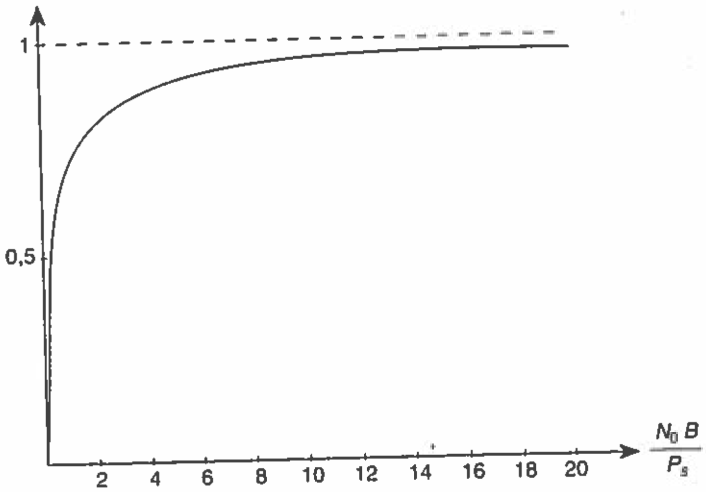

Figura 2.15 - Capacità di trasmissione di un canale di banda B in presenza di rumore bianco gaussiano.

Conclusione

La scelta delle forme d'onda impulsive per la trasmissione numerica è legata a vari fattori. Essi sono principalmente i seguenti:
— compatibilità con la banda e le caratteristiche trasmissive del canale disponibile. Come si è già visto, la banda di canale limita la velocità di trasmissione (espressa in impulsi/s), e lo spettro delle forme d'onda impulsive dovrà essere tale da generare una interferenza intersimbolica nulla o sufficientemente piccola;
— quantità d'informazione trasmessa per impulso. Questa dipende dal numero di livelli che l'ampiezza dell'impulso può assumere. Il numero M di livelli è legato quindi al rapporto tra velocità possibile sul canale e ritmo di emissione dell'informazione da parte della sorgente;
— opportuna codificazione di canale, se è necessario aumentare la protezione contro i disturbi.

Vi sono inoltr
quisiti spettrali,
specifico, che por
ta codificazione c

2.7 CENNI SUL

La scelta del ti
merica può essere
già trattata del cc
genze dipendono
canale in gioco. U
ti canali spettri di
spettrali trascurab
nea di trasmission
sformatore, per ot
trasformatore fa s
li a frequenze mol
in modulazione d'
senza di componer
base rende più se
bande laterali. C'è
mediante una con
cronizzazione di si
Di solito ci si ri
zioni di linea. Data
alcuni semplici ese
Consideriamo c
di durata T, usand
golo bit, cioè un in
la sigla NRZ, "non
conda del valore de
di impulsi indipenc

essendo $\frac{\sin \pi f T}{\pi f T}$ lo s
rata T; esso è quinc
ce associamo ai sim
come quella di *fig.*
lore del bit), si ha
dell'ampiezza dello

Vi sono inoltre esigenze e considerazioni particolari (particolari requisiti spettrali, facilità di sincronizzazione ecc.) dipendenti dal caso specifico, che portano a scegliere particolari forme d'onda (la cosiddetta codificazione di linea, trattata brevemente nel seguito).

2.7 CENNI SULLA CODIFICAZIONE DI LINEA

La scelta del tipo di forme d'onda da usare per la trasmissione numerica può essere condizionata anche da altre esigenze, oltre a quella già trattata del controllo dell'interferenza intersimbolica. Queste esigenze dipendono di solito da caratteristiche del particolare sistema o canale in gioco. Un caso tipico è dato dalla opportunità di avere su certi canali spettri di una certa forma, per esempio spettri con componenti spettrali trascurabili sulle frequenze vicine allo zero. Si pensi ad una linea di trasmissione che si chiude sul carico in uscita attraverso un trasformatore, per ottenere un adattamento di impedenza; la presenza del trasformatore fa sì che sia conveniente l'assenza di componenti spettrali a frequenze molto basse. Un altro esempio è dato dalla trasmissione in modulazione d'ampiezza a banda laterale unica (Cap. 4) dove l'assenza di componenti a bassa frequenza nel segnale modulante in banda base rende più semplice la separazione mediante filtraggio delle due bande laterali. C'è da tener presente anche l'opportunità di facilitare mediante una conveniente scelta delle forme d'onda trasmesse la sincronizzazione di simbolo in ricezione.

Di solito ci si riferisce a tali scelte di forme d'onda come a *codificazioni di linea*. Data la grande varietà di codici di linea, riportiamo solo alcuni semplici esempi.

Consideriamo come riferimento il caso di trasmissione di dati binari di durata T, usando la forma d'onda più semplice per trasmettere il singolo bit, cioè un impulso rettangolare di durata T (spesso indicato con la sigla NRZ, "non ritorno a zero"), di segno positivo o negativo a seconda del valore del bit. Lo spettro di potenza di una sequenza casuale di impulsi indipendenti di questo tipo è proporzionale a

$$\left[\frac{\sin \pi f T}{\pi f T}\right]^2$$

essendo $\dfrac{\sin \pi f T}{\pi f T}$ lo spettro di Fourier di un impulso rettangolare di durata T; esso è quindi concentrato attorno alla frequenza nulla. Se invece associamo ai simboli binari una forma d'onda di valor medio nullo come quella di *fig. 2.16* (di segno positivo o negativo a seconda del valore del bit), si ha uno spettro di potenza proporzionale al quadrato dell'ampiezza dello spettro di Fourier della forma d'onda cioè