Appunti di Trasmissione Numerica I
Parte I - Concetti Introduttivi ed Elementi di Teoria dell' Informazione

Stefano Buzzi[1]

Maggio 2004

[1]Università degli Studi di Cassino, Dipartimento di Automazione, Elettromagnetismo, Matematica Industriale e Ingegneria dell'Informazione (DAEIMI), Via G. Di Biasio, 43, I-03043 Cassino (FR), Italia. E-mail: buzzi@unicas.it.

Premessa

Questi appunti rappresentano un primo tentativo di creazione di materiale di supporto allo studio del corso di "Trasmissione Numerica I", insegnamento impartito presso l'Università degli Studi di Cassino nell'ambito del percorso formativo previsto per la Laurea in Ingegneria delle Telecomunicazioni. Le segnalazioni degli studenti, preferibilmente via e-mail all'indirizzo `buzzi@unicas.it`, su sviste, inesattezze ed errori tipografici, come pure qualsiasi critica e/o suggerimento, sono ovviamente benvenute.

Concetti Introduttivi

1 Introduzione

Cominciamo subito col dire cosa si intende per "Trasmissione Numerica" al fine di capire di cosa parleranno le prossime pagine. Questo corso rappresenta un'introduzione alle strategie e metodologie per la trasmissione dell'informazione in forma numerica. Per meglio capirci, è opportuno definire cosa si intende per informazione e cos'è l'informazione in forma numerica. Bene, rimandando alle prossime pagine una definizione più strettamente matematica del termine informazione, possiamo genericamente affermare che informazione è ogni espressione o manifestazione che sia di interesse per uno o più soggetti. Il concetto di trasmissione è invece legato alla possibilità di poter rendere partecipi dell'informazione soggetti che si trovano in luoghi remoti rispetto al punto in cui l'informazione è stata generata o è presente in un dato momento. In maniera più generale, si parla di trasmissione anche tutte le volte che una certa informazione debba poter essere resa disponibile nello stesso luogo ma più avanti nel tempo, anche se in tal caso forse il termine più appropriato sarebbe quello di memorizzazione piuttosto che trasmissione dell'informazione. Un esempio può essere utile per chiarire tale concetto. Se facciamo riferimento ad un evento sportivo (ad esempio una partita di calcio), l'informazione può essere rappresentato dalla successione di immagini riprese da una o più telecamere. E' evidente che tale informazione ha valore se è possibile trasmetterla a utenti remoti, ovvero non presenti sul luogo dell'evento in quel momento, e, anche, se è possibile memorizzarla o archiviarla in modo da poterla rendere disponibile in momenti successivi. Allo stesso modo, in una conversazione telefonica, l'informazione è rappresentata dalla

voce emessa dai due utenti impegnati in conversazione, e, ovviamente, tale voce deve essere trasmessa attraverso la rete telefonica e giungere all'orecchio dell'altro utente affinché l'informazione stessa abbia un valore. Questi esempi aiutano a chiarire che cosa è l'informazione, e perchè sia necessario trasmetterla e/o memorizzarla, ma fanno anche capire che, nel nostro ambito, è necessario giungere ad un livello di astrazione superiore al fine di poter arrivare alla progettazione di adeguati sistemi di trasmissione. Per quanto ci riguarda, quindi, l'informazione è rappresentata da un segnale elettrico (corrente e/o tensione) variabile nel tempo e rappresentabile mediante una funzione del tempo $s(t)$. E' opportuno sottolineare come al giorno d'oggi sia possibile trasformare una qualsiasi informazione in un segnale elettrico, per cui il considerare come "informazione" da trasmettere un segnale elettrico non è per niente riduttivo e/o limitativo. Ad esempio, se pensiamo a una telefonata, è ben noto che la voce, come qualsiasi altro suono, può essere convertita in un segnale elettrico mediante un comune microfono, e riconvertita poi in forma sonora da un altoparlante, ragion per cui una telefonata comporta la trasmissione bidirezionale di segnali elettrici tra luoghi remoti. Allo stesso modo, una telecamera (analogica) altro non è che un dispositivo che converte una sequenza di immagini in un segnale elettrico. Andando avanti col nostro ragionamento, è il momento di fare un'ulteriore importante distinzione. L'informazione può essere di tipo *analogico* o di tipo *digitale* (o *numerico*). Diremo che un'informazione è di tipo analogico quando il segnale elettrico che la rappresenta, $s(t)$, assume valori in un insieme continuo. Ad esempio, il segnale in uscita ad un microfono è analogico in quanto esso può assumere tutti i valori appartenenti ad un determinato intervallo contenenente l'origine. Allo stesso modo, il segnale radiofonico è un segnale analogico, e l'insieme degli apparati trasmittenti di una stazione radio tradizionale e degli apparati riceventi dei vari ascoltatori formano un sistema di comunicazione analogico di tipo punto-multipunto. L'informazione è invece numerica (o digitale) quando è rappresentata da una successione di elementi che possono assumere solo un numero finito di valori, ovvero appartengono ad un insieme discreto. Ad esempio, se consideriamo come informazione l'estrazione del lotto della ruota di Milano, l'informazione sarà costituita da una successione di 5 numeri appartenenti all'insieme $\{1, 2, \ldots, 89, 90\}$, e tale informazione potrà essere ad esempio rappresentata da un segnale $s(t)$ costante a tratti il cui codominio rappresenta i 5 numeri estratti. Allo stesso modo, un file mp3 di 4Megabyte non è altro che una successione di 4 milioni di byte, che sappiamo essere ottetti di bit, e che quindi possono assumere solo 2^8 possibili valori; ne consegue quindi che un file mp3, o, in generale, un qualsiasi file, rappresenta un'informazione di tipo digitale.

Ancora, questa pagina dattiloscritta rappresenta informazione in forma numerica, in quanto costituita da una successione di simboli discreti, ovvero le lettere dell'alfabeto italiano. Bene, la trasmissione numerica si occupa della definizione e del progetto delle strategie che permettono di realizzare sistemi di trasmissione (e memorizzazione) dell'informazione in forma numerica. Abbiamo già fatto l'esempio di un sistema di comunicazione analogico (la radio); un esempio di sistema di trasmissione numerico, alquanto complesso, è costituito dalla rete Internet. Tutti oggi sanno bene cosa sia la rete Internet; per quel che ci riguarda, possiamo dire che Internet è una ragnatela di PC e server interconnessi tra loro e che si scambiano file di varia natura, ovvero informazioni numeriche. E' bene sottolineare come al giorno d'oggi i sistemi di trasmissione numerica abbiano preso il sopravvento sui sistemi di trasmissione analogici, che sono in netta fase di decadenza rispetto ai sistemi di trasmissione numerica. La ragione di tutto ciò è legata essenzialmente a vari fattori, nel seguito brevemente illustrati.

a.) L'informazione analogica può essere convertita, mediante opportuni procedimenti, in forma numerica. Banalmente, ogni segnale analogico $s(t)$, sotto opportune condizioni, può essere trasformato, mediante operazioni di campionamento e quantizzazione, in un segnale tempo-discreto ed ampiezza-discreto, ovvero in informazione numerica. E' ben noto che tale processo di conversione comporta una perdita irreversibile di informazione, nel senso che non è in generale possibile ricostruire il segnale originario $s(t)$ a partire dalla sua versione numerizzata. Pur tuttavia, il processo di numerizzazione, se opportunamente eseguito, introduce degradazioni sull'informazione che sono impercettibili al fruitore del sistema di comunicazione. Possiamo quindi affermare che al giorno d'oggi sono disponibili metodologie per convertire una qualsiasi sorgente di informazione analogica in informazione di tipo digitale, o, più in particolare, in un flusso di simboli binari emessi con una certa velocità (bit-rate).

b.) Un inevitabile corollario delle affermazioni precedenti è che quindi sistemi di trasmissione numerica possono essere utilizzati per la trasmissione di informazioni originariamente analogiche che sono state assoggettate ad un processo di numerizzazione. Oggi si fa ampio ricorso a tale tecnica. A titolo di esempio, la rete di telefonia fissa è stata inizialmente progettata (oltre un secolo fa) e realizzata come un sistema di comunicazione analogico. Oggi, nella quasi totalità del mondo industrializzato la rete telefonica è un sistema di trasmissione numerico in cui la voce viaggia in

forma numerizzata; in particolare, il segnale vocale in tale rete è convertito in un flusso dati numerico avente un bit-rate di 64Kbit/s. Analogamente, la rete di telefonia cellulare GSM si basa sull'utilizzo di una modulazione numerica che trasmette voce umana numerizzata a circa 13Kbit/s. Ancora, filmati video sono convertiti mediante gli standard di compressione mpeg2 e similari in flussi dati a una velocità di alcuni megabit/s.

c.) Il fatto che sorgenti analogiche di natura diverse (voce, filmati, brani musicali, etc.) possano essere indistintamente convertiti in flussi di bit ad un certo bit-rate introduce una modularità nel sistema di trasmissione, dal momento che questo può essere progettato avendo come obiettivo la trasmissione di simboli binari, indipendentemente dalla loro natura. Altrimenti detto, non bisogna preoccuparsi della natura del segnale analogico originario che ha generato i bit, e si può progettare il sistema di trasmissione in maniera indipendente.

d.) Il fatto che ogni sorgente analogica possa essere convertita in flussi di simboli binari, unito al fatto che i simboli binari sono le unità di informazione elementari che i calcolatori elettronici utilizzano per le loro elaborazioni, fa sì che molte delle operazioni realizzate in una catena di trasmissione e ricezione di un sistema di trasmissione numerico possano essere effettuate ricorrendo a dei calcolatori elettronici, e tutto ciò semplifica enormemente i processi di trasmissione, elaborazione e memorizzazione dell'informazione. In un cero senso, i bit possono essere assunti rappresentare una sorta di "moneta unica" del mondo delle tecnologie dell'informazione, attorno a cui ruota la progettazione, implementazione ed il funzionamento dei moderni calcolatori e sistemi di trasmissione e memorizzazione delle informazioni.

e.) *Last, but not least*, i sistemi di trasmissione numerica possono essere resi di gran lunga più affidabili dei sistemi di comunicazione analogica, ed è quindi preferibile ricorrere alla trasmissione dell'informazione in forma numerica piuttosto che analogica.

L'affermazione del punto *e.)* potrebbe in questo momento apparire alquanto dogmatica e oscura. Di fatto, questo corso è dedicato ai sistemi di trasmissione numerica, ragion per cui nelle pagine e nei capitoli seguenti sarà fatta maggiore chiarezza, e sarà spiegato in che senso e perché i sistemi di trasmissione numerica possono essere resi affidabili, e, quindi, i motivi per cui questi hanno soppiantato i sistemi di

tipo analogico.

2 Modello di un sistema di trasmissione numerico

Cominciamo ora ad illustrare lo schema a blocchi di principio di un sistema di trasmissione numerico, come rappresentato in Fig. 1. Avremo modo di entrare nel dettaglio dei singoli blocchi presenti in figura, per cui per il momento ne daremo una descrizione sommaria che possa dare un'idea d'insieme del sistema. Il blocco "Sorgente discreta" è ovviamente la fonte dell'informazione che si vuole trasmettere verso un luogo remoto o anche che si vuole memorizzare su un supporto ad esempio magnetico. I simboli emessi da tale sorgente possono rappresentare un'informazione che era già di per se in forma numerica (ad esempio, un file di dati), o un'informazione analogica che è stata convertita in forma numerica mediante le usuali operazioni di campionamento, quantizzazione ed, eventualmente, codifica. Assumeremo che tale sorgente emetta simboli appartenenti ad un'alfabeto $\mathcal{X}$ a cardinalità $|\mathcal{X}| = C$ con cadenza uniforme; possiamo ad esempio indicare con T_s il tempo - usualmente misurato in secondi - che intercorre tra l'emissione di un simbolo ed il successivo, e che possiamo definire *intervallo di simbolo*. L'inverso di T_s definisce il tasso o frequenza di simbolo o anche *symbol-rate*, usualmente indicato col simbolo R_s, e che misura il numero di simboli che la sorgente emette nell'unità di tempo; usualmente R_s si misura in simboli al secondo o anche *baud*. Un caso notevole di sorgente si ha per $C = 2$, ovvero quando l'alfabeto di sorgente $\mathcal{X}$ ha cardinalità pari a 2. In tal caso la sorgente è detta *binaria*, emette simboli binari, o bit (il termine bit proviene dalla contrazione del termine inglese *binary digit*). Il tempo intercorrente tra l'emissione di due bit è detto *intervallo di bit* e si indica con T_b, mentre il suo inverso è comunemente detto *bit-rate*, si indica con R_b e si misura in bit al secondo (bit/s). Ad esempio, per un segnale vocale numerizzato per la trasmissione sulla rete telefonica fissa si è già visto che $R_b = 64000$bit/s, ragion per cui è $T_b = 1/R_b = 1.5625 \cdot 10^{-5}$ s. Ovviamente, una qualsiasi sorgente non binaria può essere convertita in una sorgente binaria mediante una semplice operazione di codifica che assegni a ciascun simbolo di $\mathcal{X}$ una stringa di bit di opportuna lunghezza. E' immediato rendersi conto che se ad esempio $C = 8$, è possibile assegnare a ciascuno dei possibili simboli emessi da una sorgente una stringa di $\log_2 8 = 3$ bit e ottenere una sorgente binaria. Si noti inoltre che l'intervallo di bit di tale sorgente sarà un terzo dell'intervallo di simbolo della sorgente originaria, mentre il bit-rate sarà il triplo della frequenza

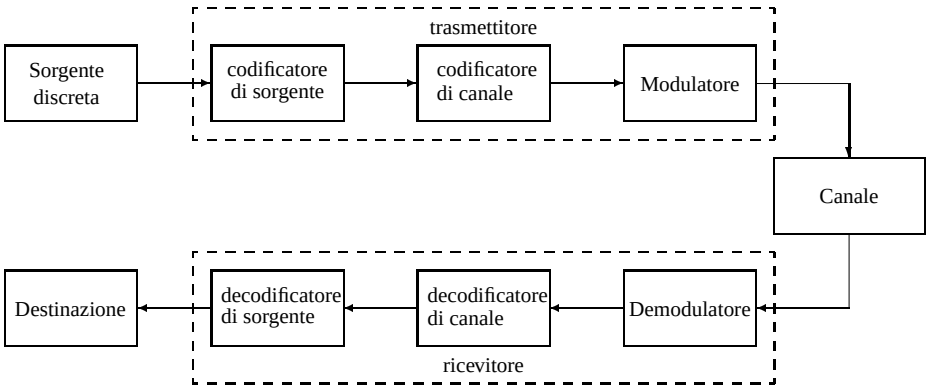

Figure 1: Sistema di comunicazione numerico.

di simbolo R_s della sorgente originaria. In generale, una sorgente a cardinalità C arbitraria può essere convertita in una sorgente binaria associando a ciascun simbolo di sorgente una stringa di $k = \lceil \log_2 C \rceil$ bit. La sorgente binaria equivalente così ottenuta sarà caratterizzata da un intervallo di bit $T_b = T_s/k$ e un bit-rate $R_b = kR_s$.

Occupiamoci ora del "codificatore di sorgente". Tale blocco ha il compito di effettuare la cosiddetta codifica di sorgente, ovvero di rappresentare i simboli all'uscita della sorgente mediante stringhe di bit *della minima lunghezza possibile*. In altre parole, il codificatore di sorgente ha il compito di rimuovere la cosiddetta ridondanza dell'informazione, in maniera tale che questa possa essere trasmessa sprecando meno risorse possibili. Nel seguito ci soffermeremo su tale blocco. Per il momento, è utile osservare che la codifica di sorgente è in realtà un'operazione familiare ai più: infatti, ogni volta che procediamo allo zippaggio di un file mediante gli usuali software (ad esempio il winzip), effettuiamo una codifica di sorgente, che ci permette poi di conservare e/o inviare i file con minor spreco di risorse.

Mentre il codificatore di sorgente si occupa della rimozione della ridondanza, il codificatore di canale aggiunge invece ridondanza al fine di rendere l'informazione meno vulnerabile agli errori che eventualmente potranno verificarsi durante la fase di ricostruzione dell'informazione che si opera a destinazione. Un esempio alquanto semplice, ma tuttavia comune, di codifica di canale è l'aggiunta di bit di parità. Aggiungendo dei bit di parità in modo tale che ad esempio stringhe di bit di opportuna lunghezza abbiano un numero pari di uni può aiutare ad individuare in ricezione quelle stringhe di bit che possono essere

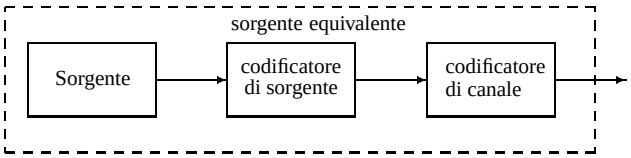

Figure 2: Sorgente numerica equivalente.

affette da errori. A qualcuno potrebbe sembrare una sorta di nonsenso rimuovere la ridondanza tramite la codifica di sorgente e poi aggiungervela tramite la codifica di canale. Il punto è che il codificatore di sorgente rimuove la ridondanza inutile e non strutturata, permettendo una rappresentazione efficiente e compatta dell'informazione da trasmettere. Il codificatore di canale, poi, aggiunge ridondanza "intelligente", ovvero con una data struttura, nota al ricevitore, che può quindi avvalersi di tale conoscenza per rivelare gli errori e in molti casi correggerli. Si noti infine che l'insieme dei blocchi "sorgente", "codificatore di sorgente" e "codificatore di canale" costituiscono una sorgente numerica equivalente, rappresentata in Fig. 2 che è usualmente assunta essere binaria.

Passiamo quindi al Modulatore, uno dei blocchi più importanti del sistema di trasmissione numerico. Nella sua forma più semplice, il modulatore può essere riguardato come un dispositivo che associa, a simboli o stringhe di simboli che si presentano in ingresso, una data forma d'onda, o, se vogliamo, un segnale elettrico $s(t)$ che viene trasmesso sul canale. Assumendo di trattare con modulatori senza memoria (le modulazioni con memoria sono essenzialmente oggetto del corso di Trasmissione Numerica II) un modulatore è quindi specificato generalmente dai seguenti parametri:

- La cardinalità M del modulatore, che è un numero intero, usualmente una potenza di 2. Se $M = 2$ il modulatore si dice binario, altrimenti in generale si parlerà di modulatore M-ario.

- Un insieme di M forme d'onda $\{s_1(t), s_2(t), \ldots, s_M(t)\}$. Usualmente si assume che tali forme d'onda abbiano durata limitate, ovvero siano non nulle solo per t appartenente ad un dato intervallo $[0, T]$, e che siano al quadrato sommabili, ovvero siano segnali di energia. In particolare, indicheremo con

$$\mathcal{E}_i = \int_{-\infty}^{+\infty} s_i^2(t)dt = \int_0^T s_i^2(t)dt$$

l'energia associata all'i-esima forma d'onda $s_i(t)$ del modulatore.

- Una funzione biunivoca che associa a stringhe binarie di lunghezza $k = \lfloor \log_2 M \rfloor$ una tra le M forme d'onda a disposizione del modulatore.

Il modulatore, quindi, sostituisce a stringhe di k bit delle forme d'onda che possano essere trasmesse sul canale. Ad esempio, se $M = 4$, il modulatore ha a disposizione le 4 forme d'onda $s_1(t), \ldots, s_4(t)$, e può compiere la seguente associazione

$$00 \ \rightarrow s_1(t) \ , \quad 01 \ \rightarrow s_2(t) \ , \quad 10 \ \rightarrow s_3(t) \ , \quad 11 \ \rightarrow s_4(t) \ .$$

Ovviamente, affinché il sistema di trasmissione possa operare con continuità, è necessario che la durata T delle forme d'onda del modulatore (usualmente indicata col termine di intervallo di segnalazione) sia non maggiore del tempo che il codificatore di canale impiega a produrre coppie di simboli binari. In tal modo, quando in uscita al codificatore di canale si presenta una coppia di simboli binari, il modulatore invia sul canale una delle forme d'onda a sua disposizione, e tale trasmissione termina *prima* che una nuova coppia di simboli binari sia emessa dal codificatore di canale. Se, invece, T fosse maggiore del tempo di emissione di una coppia di simboli, nuove coppie si renderebbero disponibili all'ingresso del modulatore prima che la trasmissione delle forme d'onda associate alle coppie precedenti possa essere ultimata. Questa situazione è chiaramente da evitare quando si vuole conseguire una trasmissione in tempo reale dell'informazione e si vuole evitare l'uso di memorie tampone di elevata capacità. Ritornando al nostro esempio, se il codificatore di canale produce la sequenza di bit 011100, il modulatore trasmette sul canale il segnale

$$s_2(t) + s_4(t - T) + s_1(t - 2T) \ .$$

E' opportuno notare che tra modulatore e canale sono presenti usualmente dei blocchi aggiuntivi, quali ad esempio amplificatori, convertitori a radio frequenza, e, nel caso di trasmissione sul canale radio, un'antenna trasmittente. Poiché tuttavia tali blocchi sono presenti in un qualsiasi sistema di trasmissione, anche analogico, daremo per scontata la loro presenza e non vi presteremo attenzione.

Il blocco "Canale" schematizza tutto ciò che è presente tra il trasmettitore ed il ricevitore. Esso può essere fisicamente rappresentato da un filamento di materiale conduttore, se ci riferiamo a comunicazioni cosiddette *wired*, come ad esempio le comunicazioni di telefonia fissa, ove il canale, almeno

relativamente al tratto urbano della chiamata è costituito da una coppia di fili di rame (doppino telefonico). Oppure, può essere costituito dall'atmosfera e da tutti gli oggetti che ci circondano se facciamo riferimento a comunicazioni *wireless*, ovvero senza fili, come ad esempio le comunicazioni di telefonia cellulare. In aggiunta, per comunicazioni ottiche, il canale è usualmente rappresentato da un supporto di materiale dielettrico denominato fibra ottica. E' facile rendersi conto che ciascuno dei citati canali di comunicazione si connota per caratteristiche peculiari che rendono necessaria l'adozione sia di modelli matematici che di tecniche e strategie di trasmissione diverse per ciascuno di essi. Per quanto ci riguarda, faremo essenzialmente riferimento a due semplici modelli di canale. Il primo canale è il cosiddetto canale additivo Gaussiano bianco, detto anche canale AWGN (Additive White Gaussian Noise), schematizzato in Fig. 3. Tale modello prevede che il segnale ricevuto $r(t)$ possa esprimersi come la somma del segnale trasmesso $s(t)$ e di un disturbo aggiuntivo $n(t)$, assunto essere una realizzazione di un processo aleatorio Gaussiano bianco con densità spettrale di potenza (PSD) pari ad $\mathcal{N}_0/2$. E' opportuno osservare che nel nostro modello abbiamo assunto che il canale non introduca alcuna attenuazione, ovvero si è supposto che il segnale ricevuto $r(t)$ contenga una versione non scalata del segnale trasmesso. Di fatto, avremmo anche potuto assumere valida la relazione

$$r(t) = \alpha s(t) + n(t) \, ,$$

ovvero assumere che il canale introduca un coefficiente di attenuazione $\alpha < 1$ sul segnale trasmesso. Per semplicità, assumeremo che sia $\alpha = 1$, ovvero supporremo che l'energia del segnale utile sia tale da inglobare l'effetto dell'attenuazione presente sul canale. Si osservi anche che, in accordo allo schema in Fig. 3, si vede che il canale introduce sul segnale ricevuto un disturbo di tipo additivo. Nella maggior parte dei casi, in realtà, il rumore non è introdotto dal canale ma dagli stadi di ricezione del ricevitore, che sono caratterizzati da una temperatura equivalente di rumore non nulla. Questo per dire che $n(t)$ può essere assunto pari alla somma del rumore introdotto al ricevitore e di un eventuale rumore additivo introdotto dal canale, magari sotto forma di interferenze esterne che occupano la stessa banda del segnale utile $s(t)$. Il canale AWGN è un modello classico della teoria delle comunicazioni, ed il progetto e l'analisi dei sistemi di trasmissione, sia analogici che numerici è stata classicamente condotta con riferimento a tale modello. Sebbene sia facilmente intuibile che molti canali reali debbano essere schematizzati ricorrendo a modelli più complessi, il canale AWGN modella in maniera accurata molti

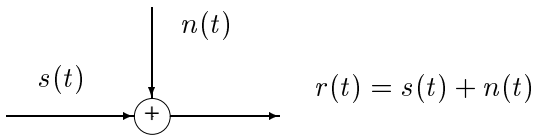

Figure 3: Canale AWGN.

canali di comunicazione di tipo wired, e alcuni tipi di canale wireless tra cui alcuni canali satellitari e il canale spaziale che si frappone nelle comunicazioni tra la terra e sonde presenti nello spazio profondo. L'uso del semplice modello AWGN, d'altra parte, è utile alla comprensione e allo sviluppo dei risultati fondamentali della trasmissione numerica. Inoltre, dal momento che i canali reali sono usualmente "peggiori" del canale AWGN, valutare le prestazioni delle modulazioni numeriche permette di stabilire una sorta di *benchmark* con cui andare a confrontare le prestazioni realmente ottenibili con modelli di canale maggiormente aderenti alla realtà. Infine, è possibile dimostrare che mediante l'utilizzo di opportune tecniche, quali ad esempio la tecnica della diversità[1], è possibile far sì che canali reali tendano a comportarsi come il canale AWGN. Queste affermazioni ci fanno capire l'importanza del modello di canale AWGN, ed infatti gran parte delle pagine seguenti saranno dedicate allo studio delle trasmissioni numeriche su tale canale.

Un altro modello di canale ugualmente importante e che sarà preso dettagliatamente in esame nel seguito è il cosiddetto modello di *canale AWGN distorcente*. In tal caso, il segnale ricevuto contiene una versione distorta del segnale trasmesso. In particolare, si assume che il canale di trasmissione possa essere schematizzato mediante un filtro lineare tempo-invariante (LTI) di risposta impulsiva $c(t)$, ragion per cui il segnale ricevuto è espresso come

$$r(t) = s(t) * c(t) + n(t) = \int s(\tau)c(t-\tau)d\tau + n(t)\ . \tag{1}$$

Tale modello di canale è utile per rappresentare quelle situazioni in cui il canale produce distorsioni di tipo LTI sul segnale trasmesso. Un caso pratico è costituito da quei sistemi di trasmissione che possono operare solo all'interno di una data banda di frequenze, ad esempio la banda $[f_L, f_U]$, perché le bande adiacenti sono già occupate da altri segnali di comunicazione. Bene, in tal caso, ammesso che non vi

[1] Lo studio di tale strategia è argomento del corso di Trasmissione Numerica II.

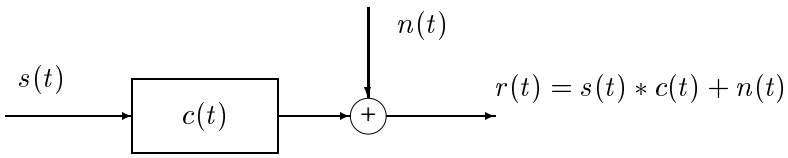

Figure 4: Canale AWGN distorcente.

siano distorsioni di altra natura, $c(t)$ è assunto essere pari ad un filtro passa-banda ideale con frequenza di taglio inferiore f_L e frequenza di taglio superiore f_U. Si parla in tal caso anche di "modello di canale AWGN a banda limitata." Si osservi inoltre che il modello AWGN distorcente ingloba come caso particolare quello di canale AWGN, cui si riconduce scegliendo $c(t) = \delta(t)$.

Occupiamoci ora del blocco "Demodulatore". Tale blocco, come si può intuire, esegue un'operazione inversa a quella del modulatore, in quanto, sulla base del segnale ricevuto $r(t)$ deve individuare quale tra le M forme d'onda a disposizione del modulatore è stata inviata sul canale, al fine di poter poi risalire alla k-upla di bit che era stata prodotta dal codificatore di canale. E' ovvio che il processo di demodulazione non è in generale banale, dal momento che, anche nel caso del semplice canale AWGN, l'aggiunta del rumore termico rende il segnale ricevuto un processo aleatorio, e non è garantito che il demodulatore riesca a decidere correttamente su quale forma d'onda sia stata trasmessa. E' quindi naturale aspettarsi che il demodulatore sarà affetto da una certa *probabilità di errore*, legata alla PSD del rumore additivo introdotto dal canale, ed è quindi naturale porsi il problema di individuare il "demodulatore ottimo", ove il criterio di ottimalità è la minimizzazione della probabilità di errore. Altrimenti detto, uno degli argomenti cardine della trasmissione numerica consiste nell'individuazione delle regole di decisione ottime, ossia delle strategie di elaborazione del segnale ricevuto $r(t)$ che permettano di decidere sulla forma d'onda trasmessa con la minima probabilità di errore. Nel seguito, ci soffermeremo quindi sull'esposizione della struttura del demodulatore ottimo per entrambi i modelli di canale discussi.

A causa degli errori che inevitabilmente saranno introdotti nel processo di demodulazione, la stringa di bit in ingresso al decodificatore di canale non sarà in generale perfettamente coincidente con quella osservata in uscita al codificatore di canale. Compito del decodificatore di canale è quindi quello di

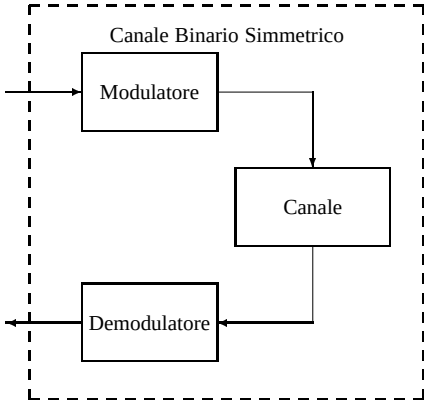

Figure 5: Modello esploso di un canale binario simmetrico.

sfruttare la ridondanza intelligente introdotta dal codificatore di canale al fine di individuare e correggere gli errori presenti al suo ingresso. Si capisce che il decodificatore di canale non è in grado di compiere la "magia" di azzerare la probabilità di errore, ma può solo limitarsi a ridurla. Le capacità di rivelazione e correzione del decodificatore dipendono dalla struttura e complessità del codice di canale utilizzato, e, anche, dalla quantità di ridondanza introdotta. Maggiore sarà la ridondanza, maggiori saranno le capacità di correzione del codice. D'altra parte, si è già discusso come la ridondanza porti ad uno spreco di risorse, per cui, da un lato, conviene introdurre ridondanza per aumentare le capacità di correzione degli errori del codice di canale, e, dall'altro lato, conviene ridurre la ridondanza per risparmiare risorse[2]. Vi sono quindi due esigenze contrastanti, tra le quali bisognerà necessariamente mediare per giungere a un compromesso.

Il decodificatore di sorgente, come è naturale aspettarsi, esegue infine l'operazione inversa a quella eseguita dal codificatore di sorgente. Esso quindi, a partire dai dati compressi, reintroduce la ridondanza statistica al fine di restituire ai dati il loro significato semantico e renderli quindi fruibili al destinatario del processo di trasmissione. L'operazione di decodifica di sorgente equivale alla comune operazione di *unzipping* che viene eseguita su dati zippati al fine di ottenere dei file utilizzabili e maneggiabili dagli utenti finali.

[2]Il significato del termine risorse sarà più chiaro nel seguito. Al momento, possiamo dire che all'aumentare della ridondanza aumentano sia la complessità del modulatore e del demodulatore, sia, eventualmente, la banda impegnata dal segnale trasmesso.

Con questo, abbiamo descritto per sommi capi tutti i blocchi costituenti lo schema di principio di un sistema di trasmissione numerico, come rappresentato in Fig. 1. Scopo del corso di Trasmissione Numerica I è quello di studiare in dettaglio tale schema di trasmissione al fine di familiarizzare gli studenti con i principi e i concetti di base propri della teoria della trasmissione numerica. Concludiamo questo capitolo con due precisazioni. Anzitutto, la cascata dei blocchi modulatore, canale e demodulatore, come rappresentata in Fig. 5, è un sistema che accetta in ingresso simboli binari e produce in uscita simboli binari che, con una certa probabilità di errore p_e, possono essere diversi dai corrispondenti simboli in ingresso. Se assumiamo (e questa è un'ipotesi del tutto ragionevole e soddisfatta nella pratica) che la probabilità di confondere un uno con uno zero sia uguale alla probabilità di confondere uno zero con un uno, il blocco di Fig. 5 non è altro che il canale binario simmetrico (BSC), noto dal corso di Teoria dei Fenomeni Aleatori.

Infine, è opportuno ancora una volta sottolineare che lo schema di Fig. 1 è utile anche per schematizzare un procedimento di memorizzazione dati su un supporto di massa. Di fatto, anche nei dispositivi di memorizzazione sono presenti le operazioni di codifica di sorgente e di canale, e le loro operazioni inverse di decodifica. L'operazione di modulazione consiste invece nel processo di scrittura sulla memoria di massa. Il canale di comunicazione è rappresentato dal supporto fisico di memorizzazione, mentre l'operazione di demodulazione consiste nel processo di lettura dai dati. Gli errori in tal caso possono essere introdotti sia a causa di imperfezioni nel processo di scrittura, sia a causa di imperfezioni nel processo di lettura, sia a causa di danni al supporto fisico che possono accadere tra le operazioni di lettura e scrittura (si pensi ad esempio ad un CD dati graffiato). A causa di tali errori, quindi, il demodulatore dovrà compiere opportune elaborazioni sul segnale ricevuto (o, con maggiore esattezza, letto) al fine di ricostruire con minima probabilità di errore i dati precedentemente scritti sul supporto. E' chiaro inoltre che, nel caso di sistemi per la memorizzazione dati, il modello di canale adottabile non coincide col modello AWGN, in quanto c' bisogno di un modello che porti in conto il meccanismo di generazione degli errori. La trattazione di tali sistemi esula ad ogni modo dai nostri scopi e non sarà ulteriormente approfondita.

Elementi di Teoria dell'Informazione

In questo capitolo saranno dati alcuni cenni di teoria dell'informazione, al fine di introdurre concetti teorici utili alla comprensione dei sistemi di trasmissione numerica. Ci soffermeremo in particolare sulla caratterizzazione matematica e quantitativa del contenuto informativo di una sorgente, sulla codifica di sorgente, e sul modello di canale AWGN. Inoltre, introdurremo il concetto, di primaria importanza, di capacità di canale ed enunceremo il teorema di Shannon sulla codifica di canale, senza dubbio una delle pietre miliari della teoria dell'informazione e della teoria dei sistemi di trasmissione numerica.

1 Codifica di sorgenti binarie senza memoria

1.1 Sorgenti senza memoria

Si consideri un insieme di oggetti (o simboli) $\mathcal{X} = \{\xi_1, \xi_2, \ldots, \xi_C\}$ a cardinalità C, ovvero contenente C elementi. Sia $x(0), x(1), x(2), \ldots, x(n), x(n+1), \ldots$ una successione di variabili aleatorie, aventi valore in $\mathcal{X}$; tale successione definisce un messaggio o sequenza di simboli presi dall'alfabeto $\mathcal{X}$. Si indichi inoltre con $\{p_i(\ell)\}_{i=1}^{C}$ la distribuzione di probabilità della v.a. $x(\ell)$, ovvero si assuma che

$$\text{Prob}\{x(\ell) = \xi_i\} = p_i(\ell)\ . \tag{2}$$

Si definisce *sorgente di informazione discreta* un dispositivo capace di generare sequenze di simboli appartenenti all'insieme $\mathcal{X}$ secondo la legge di probabilità (2). Al riguardo valgono le seguenti definizioni:

- L'insieme $\mathcal{X}$ è detto alfabeto della sorgente e C è detta cardinalità della sorgente o dell'alfabeto di sorgente.

- La sorgente è detta stazionaria se vale la seguente proprietà:

$$\begin{aligned}&\text{Prob}\{x(i_1) = \xi_{k_1}, x(i_2) = \xi_{k_2}, \ldots, x(i_m) = \xi_{k_m}\} = \\ &\text{Prob}\{x(i_1 + h) = \xi_{k_1}, x(i_2 + h) = \xi_{k_2}, \ldots, x(i_m + h) = \xi_{k_m}\}\end{aligned} \tag{3}$$

  per tutti gli interi non negativi $i_1, \ldots, i_m$, e h, e per tutti i simboli $\xi_{k_1}, \ldots, \xi_{k_m}$ appartenenti all'alfabeto $\mathcal{X}$. Si noti che la stazionarietà della sorgente implica che le probabilità marginali

$p_i(\ell)$ nella (2) sono di fatto indipendenti dall'indice temporale ℓ. Poiché nel seguito ci occuperemo esclusivamente di sorgenti stazionarie, per semplicità useremo quindi la notazione p_i invece che $p_i(\ell)$.

- La sorgente è detta *senza memoria* se i simboli da essa emessi sono schematizzabili come una sequenza di variabili aleatorie indipendenti. In tal caso, la probabilità di osservare un certo simbolo all'uscita della sorgente non dipende dalla cosiddetta storia passata della sorgente, ovvero non dipende da quali particolari simboli siano stati emessi fino a quel momento.

- Spesso, è usuale associare ad ogni sorgente un tempo di simbolo T_s, che equivale all'intervallo temporale che intercorre tra l'emissione di due simboli consecutivi. Considereremo sorgenti con tempo di simbolo costante, ovvero trascureremo il caso in cui la sorgente emetta simboli con cadenza irregolare.

1.1.1 Esempio

Un semplice esempio di sorgente è la sorgente binaria, per la quale l'alfabeto $\mathcal{X}$ contiene solo 2 elementi, che possiamo per comodità denotare con "0" e "1", ossia $\mathcal{X} = \{0, 1\}$. Dal momento che i messaggi saranno in tal caso costituiti da sequenze di simboli binari, si intuisce che la sorgente binaria modella la sorgente d'informazione tutte le volte che si vuole trasmettere un file dati (che è intrinsecamente in forma binaria), o informazione analogica che è stata sottoposta ad un processo di numerizzazione (si pensi ad esempio alla voce umana, convertita nel sistema cellulare GSM in un flusso binario di circa 13000 bit/s.). A seconda poi delle caratteristiche dei messaggi che la sorgente genera, questa potrà essere con memoria o senza memoria. Ad esempio, se il flusso binario generato contiene bit di parità il modello di sorgente senza memoria non è applicabile.

1.2 Entropia dell'alfabeto sorgente

Dato l'alfabeto di sorgente $\mathcal{X}$, ci poniamo ora il problema di definire da un punto di vista matematico il contenuto di informazione che l'osservazione di un elemento di $\mathcal{X}$ all'uscita della sorgente può fornire. In altri termini, andiamo alla ricerca di una funzione $I(\xi_i)$ che possa indicare il contenuto di informazione associato all'osservazione del simbolo ξ_i. Bene, tale funzione deve possedere le seguenti due proprietà

dettate dal buon senso e dalla comune intuizione, ossia:

- $I(\xi_i)$ deve essere una funzione decrescente della probabilità p_i del simbolo ξ_i. Invero, se per ipotesi fosse $p_i = 1$, la sorgente emetterebbe deterministicamente sempre il simbolo ξ_i e quindi l'osservazione della sua uscita non porterebbe alcuna informazione. Viceversa, se p_i è prossima allo zero, allora l'osservazione del simbolo ξ_i è un evento raro, che porta una gran quantità di informazione.

- La quantità di informazione associata all'emissione di due simboli indipendenti deve essere pari alla somma delle quantità di informazione legate all'osservazione dei due simboli separatamente. In altri termini:

$$\text{Prob}(\xi_i, \xi_j) = \text{Prob}(\xi_i)\text{Prob}(\xi_j) \;, \qquad \Rightarrow \qquad I(\xi_i, \xi_j) = I(\xi_i) + I(\xi_j) \tag{4}$$

  E' possibile dimostrare[3] che l'unica funzione che soddisfa tali proprietà è la seguente:

$$I(\xi_i) = -\log_a(p_i) = \log_a(1/p_i) \;, \tag{5}$$

  Usualmente si assume come base del logaritmo $a = 2$; in tal caso l'unità di misura dell'informazione è detta *bit*. Infatti, se si ha a che fare con una sorgente binaria che emette simboli equiprobabili, il contenuto di informazione associato ai simboli di sorgente è $I(\xi_1) = I(\xi_2) = \log_2 2 = 1$ bit. Se, invece, si considera come base l'usuale numero di nepero e, allora l'informazione si misura in *nat*.

  Una volta definita la quantità di informazione associata ad ogni simbolo dell'alfabeto sorgente, è possibile definire il contenuto di informazione associato all'osservazione di un generico simbolo dell'alfabeto sorgente. Ricordando che i vari simboli dell'alfabeto sono emessi dalla sorgente in accordo alle distribuzione di probabilità $\{p_i\}_{i=1}^{C}$, si può definire il contenuto di informazione medio degli elementi di $\mathcal{X}$ nel modo seguente:

$$H(\mathcal{X}) = \sum_{i=1}^{C} p_i I(\xi_i) = \sum_{i=1}^{C} p_i \log_2(1/p_i) \;. \tag{6}$$

  La quantità $H(\mathcal{X})$ è detta *entropia* dell'alfabeto di sorgente e si misura in bit/simbolo.

[3] La dimostrazione viene omessa per motivi di semplicità e brevità.

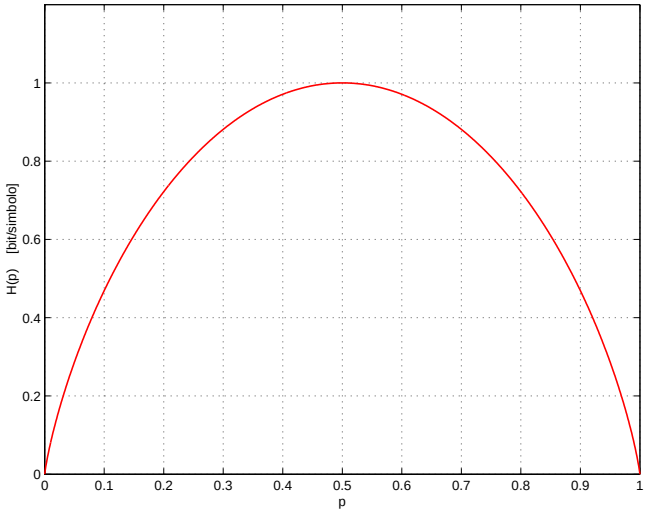

Figure 6: Entropia di una sorgente binaria di parametro p.

1.2.1 Esempi

Si consideri un alfabeto a cardinalità 4 per cui è $p_1 = 1/2$, $p_2 = 1/4$, e $p_3 = p_4 = 1/8$. L'entropia di tale alfabeto è data da

$$H(\mathcal{X}) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + 2\frac{1}{8}\log_2 8 = 1.75 \text{ bit/simbolo} . \tag{7}$$

Un alfabeto a cardinalità C con simboli equiprobabili, ovvero con $p_i = 1/C$, $\forall i$, ha entropia pari a $H(\mathcal{X}) = \log_2 C$ bit/simbolo.

Una sorgente binaria caratterizzata da un alfabeto i cui due simboli hanno probabilità p e $q = 1 - p$ ha entropia

$$H(\mathcal{X}) = p\log_2(1/p) + (1-p)\log_2\left(\frac{1}{1-p}\right) . \tag{8}$$

La quantità definita in (8) è spesso indicata col simbolo $H(p)$; il grafico di tale funzione è riportato in Fig. 1.2.1. Come c'era da aspettarsi, tale funzione assume il suo massimo per $p = 1/2$, e si annulla nei casi estremi in cui $p = 0$ o $p = 1$. Infatti, in tali casi non vi è alcuna incertezza sull'uscita della sorgente, che è di fatto deterministica, mentre nel caso di simboli equiprobabili l'informazione associata all'osservazione dei simboli della sorgente è massima.

In generale, è possibile dimostrare che per un alfabeto a cardinalità C, l'entropia è non maggiore di $\log_2 C$, con eguaglianza quando i simboli dell'alfabeto sono equiprobabili. In altri termini, vale la disuguaglianza

$$H(\mathcal{X}) \leq \log_2 C \; . \tag{9}$$

Tale relazione può essere dimostrata con semplici passaggi. Si ha infatti:

$$H(\mathcal{X}) - \log_2 C = \sum_{i=1}^{C} p_i \log_2 (1/p_i) - \sum_{i=1}^{C} p_i \log_2 C = \sum_{i=1}^{C} p_i \log_2 \left(\frac{1}{M p_i} \right) \tag{10}$$

Ora, si osservi che, per ogni $y > 0$, vale la relazione

$$\ln y \leq y - 1 \; ,$$

ove $\ln(\cdot)$ indica il logaritmo naturale. Ne discende quindi che $\log_2 y = \log_2 e \ln y \leq (y-1) \log_2 e$. Dalla (10) si ottiene quindi

$$H(\mathcal{X}) - \log_2 C \leq \log_2 e \sum_{i=1}^{C} p_i \left(\frac{1}{M p_i} - 1 \right) = \ldots = 0 \; . \tag{11}$$

Infine, la prova che la (9) vale col segno di uguaglianza nel caso di distribuzione di probabilità uniforme è banale ed è lasciata al lettore per esercizio.

1.3 Cenni sulla codifica di sorgente: codici a prefisso

Si consideri ora una sorgente stazionaria senza memoria e a cardinalità C. Ci poniamo il problema di rappresentare i simboli emessi da tale sorgente con stringhe binarie, ovvero di associare ad ogni possibile simbolo ξ_i emesso dalla sorgente una stringa di bit, detta *parola codice*, di lunghezza n_i. Se tutti i simboli ξ_i sono codificati con parole codice della stessa lunghezza allora diremo che il codice è a *lunghezza fissa*. Se invece non tutte le parole codice hanno la stessa lunghezza il codice è detto a *lunghezza variabile*. Tralasciando, per il momento, la possibilità che sul canale vengano introdotti degli errori, è ragionevole presupporre che obiettivi di efficienza e sfruttamento ottimale delle risorse trasmissive richiedano che ogni simbolo di sorgente sia rappresentato da stringhe il più possibile corte. Al riguardo, se indichiamo con n la variabile aleatoria che rappresenta la lunghezza della generica parola codice, è utile definire la

lunghezza media del codice nel modo seguente[4]:

$$\overline{n} = E\{n\} = \sum_{i=1}^{C} p_i n_i \ . \tag{12}$$

Ne consegue quindi che è di interesse andare alla ricerca di codici che rendano minima la lunghezza media. Ovviamente, nello scegliere un buon codice di sorgente la minimizzazione della lunghezza media non è l'unico parametro da portare in conto. C'è bisogno infatti che il codice sia anche *univocamente decifrabile*, ovvero che, a partire dalla sequenza binaria che codifica un dato messaggio di sorgente, sia possibile, in assenza di errori sul canale, risalire al messaggio originario in maniera univoca. Ad esempio, si consideri, per una sorgente a cardinalità 4, il seguente codice:

$$\xi_1 \to 1 \ , \qquad \xi_2 \to 10 \ , \qquad \xi_3 \to 01 \ , \qquad \xi_4 \to 011 \ . \tag{13}$$

Bene, la sequenza binaria 101101 potrebbe essere interpretata come $\xi_1\,\xi_3\,\xi_2\,\xi_1$, ma anche come $\xi_1\,\xi_4\,\xi_3$, ragion per cui il codice (13) è in realtà ambiguo, ovvero non univocamente decifrabile. Un modo per ottenere l'univoca decifrabilità è assicurarsi che venga rispettata la cosiddetta *condizione del prefisso*, ovvero richiedere che nessuna parola codice possa essere prefisso di un'altra parola codice. Infatti, se nessuna parola codice è prefisso di un'altra parola, non vi è ambiguità nella decodifica del messaggio. E' immediato verificare che il codice (13) non rispetta la condizione del prefisso. Un modo molto semplice e diretto per ottenere codici che rispettano la condizione del prefisso è quello di rappresentare graficamente le parole codice come nodi terminali di un albero con diramazioni binarie; partendo dalla radice dell'albero, i due rami che portano ai nodi del primo ordine corrispondono alla scelta tra 0 e 1 per la prima cifra binaria del codice; andando avanti, per ciscuno dei due nodi del primo ordine, i due rami che portano al nodo del secondo ordine corrispondono alla scelta tra 0 e 1 per la seconda cifra binaria della parola codice; il ragionamento ovviamente si ripete per nodi di ordine qualsiasi. Un albero di ordine h contiene nodi di ordine fino ad h. In Fig. 7 è rappresentato un albero; associando le stringhe rappresentative dei nodi terminali ai simboli di sorgente è possibile ottenere il seguente codice:

$$\xi_1 \to 0 \ , \qquad \xi_2 \to 10 \ , \qquad \xi_3 \to 110 \ , \qquad \xi_4 \to 111 \ . \tag{14}$$

E' immediato verificare che tale codice è univocamente decifrabile e rispetta la condizione del prefisso.

[4] $E\{\cdot\}$ denota l'usuale operatore di media statistica.

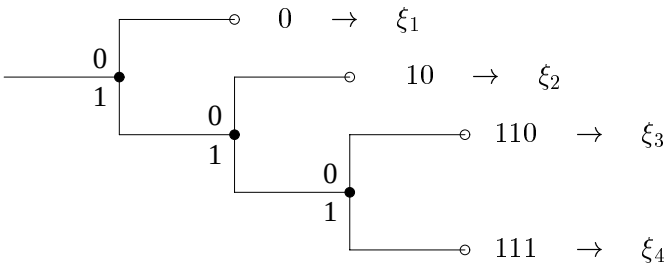

Figure 7: Rappresentazione di un codice a prefisso tramite un albero.

In generale, ci si chiede se, data una sorgente a cardinalità C, ed un set di C interi positivi $n_1, \ldots, n_C$, sia possibile individuare un codice univocamente decifrabile le cui parole codice abbiano lunghezza $n_1, \ldots, n_C$. Una risposta a tale quesito ci è data dal seguente teorema.

Disuguaglianza di Kraft: Condizione necessaria e sufficiente per l'esistenza di un codice binario a prefisso con lunghezze $\{n_1, n_2, \ldots, n_C\}$ è che valga la condizione

$$\sum_{i=1}^{C} 2^{-n_i} \leq 1 \; . \tag{15}$$

Dimostrazione

Cominciamo col dimostrare la necessarietà, ovvero che l'esistenza di un codice binario a prefisso con lunghezze $n_1, \ldots, n_C$ implica che la (15) sia soddisfatta. All'uopo, sia $n = \max\{n_1, \ldots, n_C\}$ la massima lunghezza del codice e si consideri un albero di ordine n. Un albero di questo tipo ha 2^n nodi. Se ad ogni parola codice corrisponde un nodo terminale allora prendere un simbolo di lunghezza n_i porta ad eliminare dall'albero originario tutti i nodi del sottoalbero che partono dal nodo associato alla parola codice; in particolare, tali nodi, che, se riferiti all'albero originario sono nodi di ordine n, sono in numero di 2^{n-n_i}. Dal momento che per ipotesi esiste un codice di lunghezze $n_1, \ldots, n_C$, la somma di tutti i nodi di ordine n eliminati non deve superare il numero totale 2^n di nodi di ordine n; in altre parole deve essere

$$\sum_{i=1}^{C} 2^{n-n_i} \leq 2^n \qquad \Rightarrow \qquad \sum_{i=1}^{C} 2^{-n_i} \leq 1; \, . \tag{16}$$

Dimostriamo ora la sufficienza, ovvero che la validità della (15) implica l'esistenza di un codice univocamente decifrabile con lunghezze $n_1, \ldots, n_C$. Bene, Nell'ipotesi che valga la (15) si possono

eliminare da un albero di ordine n un numero di nodi che è minore del numero totale di nodi di ordine n dell'albero di origine. Per cui in questo modo il codice può essere posizionato sull'albero in modo che sia univocamente decifrabile. In altri termini, supponiamo che le lunghezze siano ordinate in senso crescente, ovvero $n_1 \leq n_2 \leq \ldots \leq n_C$. Si consideri un albero di ordine n e si scelga un nodo di ordine n_1, cancellando i 2^{n-n_1} nodi di ordine n che si dipartono da tale nodo di ordine n_1. Poi, si scelga, tra i nodi disponibili un nodo di ordine n_2, e si cancellino i 2^{n-n_2} nodi di ordine n che si dipartono da tale nodo di ordine n_2. Tale procedura viene portata avanti fino a considerare l'intero n_C, che comporta la scelta di un nodo di ordine n_C e l'eliminazione dei 2^{n-n_C} nodi di ordine n che si dipartono da tale nodo. Dal momento che vale la (15), siamo garantiti che tale procedura può essere portata a termine, in quanto il numero di nodi cancellati, pari a $\sum_{i=1}^{C} 2^{n-n_i}$ è non maggiore del numero 2^n di nodi disponibili sull'albero. Di conseguenza, il fatto che è possibile individuare su tale albero dei nodi di ordine $n_1, \ldots, n_C$ porta automaticamente alla costruzione di un codice con tali lunghezze e univocamente decifrabile. ■

Si consideri ora una sorgente discreta stazionaria e senza memoria, a cardinalità C, e si indichi con $\{p_i\}_{i=1}^{C}$ la distribuzione di probabilità dei suoi simboli. Si assuma di voler codificare ciascun simbolo ξ_i con una parola codice la cui lunghezza n_i sia il più piccolo intero maggiore o uguale di $I(\xi_i)$, ovvero

$$I(\xi_i) \leq n_i < I(\xi_i) + 1 \; . \tag{17}$$

Si osservi anzitutto che l'esistenza di un codice univocamente decifrabile con tali lunghezze è garantita dalla disuguaglianza di Kraft, in quanto dalla (17) si ha:

$$I(\xi_i) \leq n_i \; , \quad \Rightarrow \quad -\log_2 p_i \leq n_i \; , \quad \Rightarrow \quad p_i \geq 2^{-n_i} \; , \quad \Rightarrow \quad \sum_{i=1}^{C} 2^{-n_i} \leq \sum_{i=1}^{C} p_i = 1 \; . \tag{18}$$

Inoltre, moltiplicando ambo i membri della (17) per p_i e sommando rispetto all'indice i si ottiene

$$\sum_{i=1}^{C} p_i I(\xi_i) \leq \overline{n} < \sum_{i=1}^{C} p_i I(\xi_i) + \underbrace{\sum_{i=1}^{C} p_i}_{=1} \; , \tag{19}$$

ovvero

$$H(\mathcal{X}) \leq \overline{n} < H(\mathcal{X}) + 1 \; . \tag{20}$$

Si è quindi dimostrato che, data una sorgente stazionaria e senza memoria, è possibile individuare un codice univocamente decifrabile la cui lunghezza media rispetta la condizione (20). Il limite inferiore alla lunghezza media, ovvero la disuguaglianza $\overline{n} \geq H(\mathcal{X})$ è suscettibile anche di un'interpretazione intuitiva. Infatti, dal momento che $H(\mathcal{X})$ denota il contenuto di informazione medio, espresso in bit, per ogni simbolo emesso dalla sorgente, è naturale aspettarsi che la lunghezza media in bit necessaria a rappresentare i simboli dell'alfabeto $\mathcal{X}$ debba essere non inferiore a $H(\mathcal{X})$. Dunque, non è possibile costruire codici univocamente decifrabili che abbiano una lunghezza media inferiore ad $H(\mathcal{X})$.

1.3.1 Algoritmo di Huffman

Nel seguito illustriamo una procedura che permette di costruire un codice univocamente decifrabile a partire dalla distribuzione di probabilità $\{p_i\}_{i=1}^{C}$ dei simboli dell'alfabeto sorgente. La procedura descritta è ottima, nel senso che minimizza la lunghezza media del codice, ed è stata ideata da Huffman. Ometteremo per semplicità e brevità la dimostrazione dell'ottimalità dell'algoritmo di Huffman. La procedura si basa sui seguenti passi.

Passo 1. Ordina i C simboli in modo che abbiano probabilità decrescenti.

Passo 2 Raggruppa insieme i due ultimi simboli (quelli con le probabilità minori) in un macrosimbolo con probabilità pari alla somma delle probabilità dei simboli raggruppati.

Passo 3 Ripeti i passi 1 e 2 finché non si giunge ad un unico macrosimbolo con probabilità 1.

Passo 4 Sistema i simboli su un albero in accordo a quanto fatto in Fig. 8 con riferimento ad un esempio particolare.

La Fig. 8 si riferisce al caso che si debba codificare una sorgente con $C = 4$ e con $p_1 = 1/2$, $p_2 = 1/4$ e $p_3 = p_4 = 1/8$. E' facile rendersi conto che applicando la procedura descritta si arriva al codice (14). Tenendo conto della distribuzione di probabilità, è facile verificare che in tal caso si ha

$$H(\mathcal{X}) = \overline{n} = 1.75 \text{ bit} , \tag{21}$$

ovvero la procedura di Huffman ci ha condotto ad un codice con una lunghezza media coincidente con l'entropia dell'alfabeto. Si noti che questo è un caso particolare, legato al fatto che le probabilità p_i

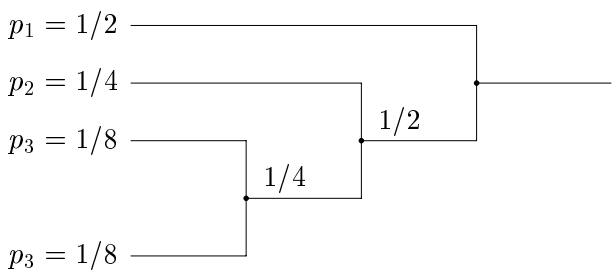

Figure 8: Rappresentazione grafica dell'algoritmo di Huffman.

sono potenze negative di due[5]. In generale, si giungerà sempre ad una lunghezza media strettamente maggiore di $H(\mathcal{X})$. Ad esempio, il lettore è invitato a verificare che, per un alfabeto con $C = 6$ e con probabilità $\{0.5,\ 0.15,\ 0.15,\ 0.1,\ 0.05,\ 0.05\}$, si ha $H(\mathcal{X}) = 2.086$bit/simbolo, mentre l'applicazione dell'algoritmo di Huffman porta a $\overline{n} = 2.1$.

E' importante osservare che la procedura di Huffman permette di ottenere un codice di sorgente a minima lunghezza media e, quindi, ottimo. Tale codice è a lunghezza variabile e univocamente decifrabile, per cui da una stringa di bit è possibile risalire senza equivoci alla sequenza di simboli dell'alfabeto di sorgente che l'ha generata. D'altra parte è possibile verificare che l'occorrenza anche di un solo errore nella sequenza binaria rivelata in ricezione comporta un catastrofico effetto di propagazione dell'errore che rende inutilizzabile l'intera stringa di bit. Quindi, il codice di Huffman, seppur ottimo nel senso della riduzione della lunghezza media, è, come tutti i codici a lunghezza variabile, estremamente vulnerabile agli errori, ragion per cui bisognerà prevedere opportune forme di protezione in fase di trasmissione. Dall'introduzione, sappiamo già che nel sistema di trasmissione è usualmente implementata la funzione di codifica di canale, che ha il compito di contrastare gli effetti del rumore e (cercare di) correggere in ricezione i simboli pervenuti errati. Di come ciò possa essere realizzato se ne parlerà in seguito.

1.4 Codifica di sorgente a blocchi

Sino a questo momento, abbiamo illustrato come codificare singoli simboli emessi da una sorgente, e, inoltre, si è visto come, tramite l'algoritmo di Huffman, si riesce ad ottenere un codice con lunghezza

[5]Invero è possibile dimostrare che la condizione $\overline{n} \geq H(\mathcal{X})$ può essere soddisfatta con l'eguaglianza se e solo se le probabilità p_i si esprimono come potenze negative di 2.

media $\overline{n} \in [H(\mathcal{X}), H(\mathcal{X})+1[$, e che il limite inferiore, ovvero l'eguaglianza $\overline{n} = H(\mathcal{X})$, è raggiungibile solo in alcuni casi particolari. Un risultato importante è rappresentato dal fatto che ci si può avvicinare al limite inferiore $H(\mathcal{X})$ effettuando una codifica di blocchi di simboli di sorgente. In particolare, supponiamo di considerare blocchi di ν simboli di una sorgente stazionaria e senza memoria. Ovviamente, tali blocchi possono assumere valori nell'insieme $\mathcal{X}^\nu$ costituito dalle C^ν possibili sequenze di elementi di $\mathcal{X}$ e lunghe ν. La considerazione di blocchi di simboli di sorgente lunghi ν porta quindi alla definizione di una nuova sorgente virtuale, il cui alfabeto è $\mathcal{Y} = \mathcal{X}^\nu$, e a cardinalità C^ν. Applicando l'algoritmo di Huffman a tale sorgente virtuale si ottiene un codice la cui lunghezza media, che indichiamo con $\overline{n}_\nu$, soddisfa la relazione

$$H(\mathcal{Y}) \leq \overline{n}_\nu < H(\mathcal{Y}) + 1 \; . \tag{22}$$

D'altro canto, è facile dimostrare che, dal momento che i simboli emessi dalla sorgente sono indipendenti, l'entropia associata a blocchi di ν simboli di sorgente coincide con ν volte l'entropia associata al singolo simbolo dell'alfabeto, ovvero $H(\mathcal{Y}) = \nu H(\mathcal{X})$; si ottiene quindi

$$H(\mathcal{X}) \leq \frac{\overline{n}_\nu}{\nu} < H(\mathcal{X}) + \frac{1}{\nu} \; . \tag{23}$$

Considerato che il rapporto $\overline{n}_\nu / \nu$ non è altro che il numero medio di bit utilizzati per rappresentare il singolo simbolo di sorgente, dalla (23) si deduce che tale numero, per ν sufficientemente grande, può essere reso arbitrariamente prossimo all'entropia di sorgente $H(\mathcal{X})$. In definitiva, anche se la distribuzione di probabilità dei simboli di sorgente è tale che, codificando singolarmente i simboli dell'alfabeto, si giunge ad un codice con lunghezza media molto diversa dall'entropia dell'alfabeto, è possibile ridurre tale lunghezza media ricorrendo ad una codifica a blocchi. In tal modo, si ottiene una progressiva riduzione della lunghezza media del codice, al prezzo, tuttavia, di un aumento della complessità dell'algoritmo di codifica e decodifica, come pure del ritardo di codifica e decodifica.

1.4.1 Esempio

Dato l'alfabeto di sorgente $\mathcal{X}$ con $C = 3$ e distribuzione di probabilità $p_1 = 0.5$, $p_2 = 0.3$ e $p_3 = 0.2$, proviamo a costruire la distribuzione della sorgente virtuale che si ottiene considerando coppie di simboli di sorgente. Il nuovo alfabeto $\mathcal{Y} = \mathcal{X}^2$ ha $C^2 = 9$ elementi con la seguente distribuzione di probabilità

$$\text{Prob}(\xi_i \; \xi_j) = p_i p_j \, , \qquad \forall i = 1, 2, 3 \, , \;\; \text{e} \;\; \forall j = 1, 2, 3 \; .$$

E' facile verificare che codificando i simboli di $\mathcal{Y}$ con l'algoritmo di Huffman si giunge ad un codice con lunghezza media minore di quella ottenibile codificando singolarmente i simboli di $\mathcal{X}$.

2 Sorgenti Stazionarie con memoria

I risultati del paragrafo precedente facevano riferimento ad una sorgente senza memoria, ovvero che emettesse simboli indipendenti. In virtù di tale ipotesi, i risultati stabiliti con riferimento ai simboli dell'alfabeto di sorgente possono essere ritenuti validi anche per i messaggi emessi dalla sorgente. Ciò ci ha permesso di identificare l'entropia dell'insieme $\mathcal{Y}$ con il contenuto informativo associato a messaggi di ν simboli emessi dalla sorgente. Di fatto, tuttavia, nella maggior parte dei casi pratici si ha a che fare con sorgenti che emettono simboli statisticamente dipendenti, per cui i risultati relativi all'alfabeto sorgente non possono essere ritenuti validi anche per messaggi emessi dalla sorgente. Sorge quindi l'esigenza di definire il contenuto di informazione medio associato ad un simbolo emesso da una sorgente con memoria. Si consideri quindi una sorgente che emetta una sequenza $x(0), x(1), \ldots$, ove la generica componente $x(\ell)$ è una variabile aleatoria a valori in $\mathcal{X}$. Il contenuto di informazione del primo simbolo $x(0)$ coincide ovviamente con l'entropia dell'alfabeto $\mathcal{X}$, ovvero

$$H(x(0)) = H(\mathcal{X}) = \sum_{i=1}^{C} p_i \log_2 \left(\frac{1}{p_i}\right) . \tag{24}$$

Una volta osservato $x(0)$, il contenuto informativo associato all'osservazione del secondo simbolo $x(1)$ è rappresentato dalla cosiddetta *entropia condizionale*, che altro non è che la media statistica dell'informazione condizionale $I(x(1)|x(0)) = -\log_2(\text{Prob}(x(1)|x(0)))$. In particolare, si ha:

$$H(x(1)|x(0)) = \sum_{i=1}^{C}\sum_{i=1}^{C} \text{Prob}\left(x(0) = \xi_i,\, x(1) = \xi_j\right) \log_2 \left(\frac{1}{\text{Prob}\left(x(1) = \xi_j \,|\, x(0) = \xi_i\right)}\right) . \tag{25}$$

Per semplicità di notazione, la (25) si può anche esprimere in forma più compatta come

$$H(x(1)|x(0)) = \sum_{x(0)}\sum_{x(1)} \text{Prob}\left(x(0),\, x(1)\right) \log_2 \left(\frac{1}{\text{Prob}\left(x(1) \,|\, x(0)\right)}\right) . \tag{26}$$

Generalizzando il precedente ragionamento, è facile intuire che il contenuto informativo associato al simbolo $x(\ell)$, una volta che siano stati osservati i precedenti simboli $x(0), \ldots, x(\ell-1)$, è dato da

$$H\left(x(\ell) \,|\, x(0), \ldots x(\ell)-1\right) = \sum_{x(0)} \ldots \sum_{x(\ell)} \text{Prob}(x(0), x(1), \ldots, x(\ell)) \log_2 \left(\frac{1}{\text{Prob}(x(\ell) \,|\, x(0) \ldots x(\ell-1))}\right) . \tag{27}$$

Di conseguenza, assumendo di poter osservare una sequenza di simboli infinitamente lunga, si definisce entropia di una sorgente stazionaria con memoria la seguente quentità:

$$H_\infty(\mathcal{X}) = \lim_{\ell\to\infty} H(x(\ell)\,|\,x(0),\ldots x(\ell-1))\ . \tag{28}$$

Naturalmente, affinché tale definizione abbia senso bisogna assicurarsi che il limite in (28) esiste. Al riguardo, cominciamo innanzitutto con provare il seguente teorema.

Teorema: L'entropia condizionale $H(x(1)|x(0))$ è tale che

$$H(x(1)\,|\,x(0)) \le H(x(1))\ . \tag{29}$$

Dimostrazione

Si consideri la quantità

$$\begin{gathered} H(x(1)\,|\,x(0)) - H(x(1)) = \\ \sum_{x(0)}\sum_{x(1)} \text{Prob}(x(0),x(1)) \log_2\left(\frac{1}{\text{Prob}(x(1)|x(0))}\right) - \sum_{x(1)} \text{Prob}(x(1)) \log_2\left(\frac{1}{\text{Prob}(x(1))}\right)\ . \end{gathered} \tag{30}$$

Tuttavia, dal momento che $\sum_{x(0)} \text{Prob}(x(0),x(1)) = \text{Prob}(x(1))$, si ha che

$$\begin{aligned} H(x(1)\,|\,x(0)) - H(x(1)) &= \sum_{x(0)}\sum_{x(1)} \text{Prob}(x(0),x(1)) \log_2\left(\frac{\text{Prob}(x(1))}{\text{Prob}(x(1)|x(0))}\right) \le \\ &\le \log_2 e \sum_{x(0)}\sum_{x(1)} \text{Prob}(x(0),x(1)) \left[\frac{\text{Prob}(x(1))}{\text{Prob}(x(1)|x(0))} - 1\right] = \\ &= \log_2 e \sum_{x(0)}\sum_{x(1)} [\text{Prob}(x(1))\text{Prob}(x(0)) - \text{Prob}(x(0),x(1))] = 0 \end{aligned} \tag{31}$$

■

Tale teorema prova rigorosamente che un condizionamento non può in nessun caso portare ad un aumento del contenuto informativo di un simbolo, ma può invece diminuirlo. Tale teorema è pienamente giustificabile sulla base del buon senso. Utilizzando tale teorema è possibile dimostrare l'esistenza del limite in (28) e quindi la congruenza della definizione di $H_\infty(\mathcal{X})$. Si ha infatti:

$$H(x(\ell)|x(\ell-1),\ldots,x(0)) \le H(x(\ell)|x(\ell-1),\ldots,x(1)) = H(x(\ell-1)|x(\ell-2),\ldots,x(0))\ , \tag{32}$$

ove l'ultima uguaglianza vale per la stazionarietà della sorgente. La (32) dimostra che la successione $H(x(\ell)|x(\ell-1),\ldots,x(0))$ è una funzione decrescente di ℓ. Ciò, unito al fatto che l'entropia condizionale assume valori non negativi, prova l'esistenza del limite in (28).

E' opportuno notare come, per sorgenti senza memoria, la definizione (28) restituisca l'entropia dell'alfabeto di sorgente $H(\mathcal{X})$, ovvero si ha $H_\infty(\mathcal{X}) = H(\mathcal{X})$.

In generale, il calcolo dell'entropia di sorgente è alquanto complicato in molti casi. Fra le eccezioni ricordiamo le sorgenti senza memoria, di cui abbiamo già trattato, e le cosiddette sorgenti di Markov, su cui tuttavia non ci soffermeremo per brevità.

3 I canali di comunicazione

Tralasciando, per ora, il blocco di codifica di canale, occupiamoci della caratterizzazione informazionale dei canali di comunicazione. Anzitutto, è opportuno osservare che possono definirsi diversi tipi di canale, a seconda delle sezioni di ingresso e uscita cui si fa riferimento. Ad esempio, il blocco canale rappresentato in Fig. 1 è un *canale continuo*, o, anche, *canale a forma d'onda*, e possibili modelli per un tale canale sono quelli rappresentati in Fig. 3 and 4. Se invece consideriamo come sezione di ingresso al canale l'ingresso del demodulatore e come uscita del canale l'uscita del demodulatore, otteniamo un canale discreto che è noto come canale binario simmetrico o BSC (*binary symmetric channel*). Tale canale è detto discreto perché i simboli in ingresso e in uscita appartengono a insiemi a cardinalità discreta, e, inoltre, il canale è anche a tempo discreto in quanto i simboli si presentano in ingresso ed in uscita in istanti discreti di tempo.

Nel seguito, ci soffermeremo dapprima sui canali discreti, e poi tratteremo i canali a forma d'onda, ed in particolare il canale AWGN.

3.1 Canale discreto senza memoria

Formalmente, un canale discreto è un dispositivo caratterizzato da un alfabeto di simboli di ingresso $\mathcal{X} = \{\xi_1, \ldots, \xi_{N_\mathcal{X}}\}$ con cardinalità $N_\mathcal{X}$, un alfabeto di simboli di uscita $\mathcal{Y} = \{\zeta_1, \ldots, \zeta_{N_Y}\}$ con cardinalità $N_\mathcal{Y}$, e da un insieme di probabilità di transizione

$$\text{Prob}\,(\zeta_j|\xi_i) = p_{j,i}\,, \qquad \forall i = 1, \ldots, N_\mathcal{X}\,, \quad \forall j = 1, \ldots, N_\mathcal{Y}\,, \tag{33}$$

che rappresentano la probabilità di osservare in uscita il simbolo ζ_j quando in ingresso è presente il simbolo ξ_i. Si noti che tali probabilità possono essere disposte a formare una matrice di dimensione

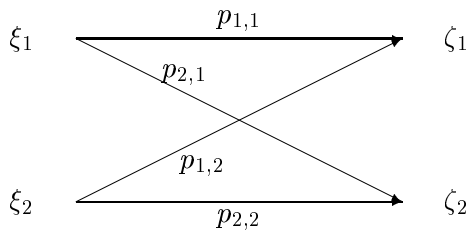

Figure 9: Il Canale Binario Simmetrico.

$N_{\mathcal{X}} \times N_Y$, detta anche *matrice di transizione* del canale:

$$\boldsymbol{P} = \begin{bmatrix} p_{1,1} & p_{2,1} & \cdots & p_{N_{\mathcal{Y}},1} \\ p_{1,2} & p_{2,2} & \cdots & p_{N_{\mathcal{Y}},2} \\ . & \cdots & \cdots & . \\ p_{1,N_{\mathcal{X}}} & p_{2,N_{\mathcal{X}}} & \cdots & p_{N_{\mathcal{Y}},N_{\mathcal{X}}} \end{bmatrix} . \tag{34}$$

Si noti che la somma degli elementi di ogni riga di tale matrice è pari a 1. Il canale si definisce senza memoria quando la probabilità di osservare un dato simbolo in uscita dipende solo dal corrispondente simbolo in ingresso e non dai simboli precedentemente prodotti dal canale o messi in ingresso. Dal punto di vista matematico, il canale si definisce senza memoria quando, definita una sequenza di simboli di ingresso $x(0), x(1), \ldots$, e una corrispondente sequenza di simboli di uscita $y(0), y(1), \ldots$, si ha che[6]

$$\text{Prob}\left(y(1), \ldots, y(\ell) | x(1), \ldots, x(\ell)\right) = \prod_{i=1}^{\ell} \text{Prob}\left(y(i)|x(i)\right) \ . \tag{35}$$

In tal caso, il canale è completamente caratterizzato dalla sua matrice di transizione $\boldsymbol{P}$.

Esempio: Il canale binario

Un caso speciale di canale discreto lo si ha quando $N_{\mathcal{X}} = N_Y = 2$; in tal caso il canale è detto binario. Una rappresentazione grafica di tale canale è riportata in Fig. 9. Si noti che, se $p_{1,1} = p_{2,2}$ e $p_{1,2} = p_{2,1}$ il canale binario si particolarizza nel ben noto BSC. Possiamo per tale canale definire l'evento errore come l'osservazione di un simbolo ζ_2 quando è stato trasmesso ξ_1 o l'osservazione di ζ_1 quando è stato

[6] Al solito, $x(\ell)$ è una variabile aleatoria che assume valori in $\mathcal{X}$, mentre $y(\ell)$ è una variabile aleatoria che assume valori in $\mathcal{Y}$.

trasmesso ξ_2. La probabilità dell'evento errore è espressa come

$$P(e) = \text{Prob}(\xi_1, \zeta_2) + \text{Prob}(\xi_2, \zeta_1) = \text{Prob}(\xi_1) p_{2,1} + \text{Prob}(\xi_2) p_{1,2} \ .$$

La probabilità di corretta ricezione (o decisione) è invece data dalla probabilità di osservare le coppie (ξ_1, ζ_1) o (ξ_2, ζ_2). Banalmente di dimostra che la probabilità di corretta ricezione è il complemento a uno della probabilità dell'evento errore.

In generale, per un canale non binario per cui sia $N_X = N_\mathcal{Y} = N$ la probabilità di errore si esprime come:

$$\begin{aligned} P(e) = \text{Prob}(x(\ell) \neq y(\ell)) = \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \text{Prob}(\xi_i, \zeta_j) = \\ \sum_{i=1}^{N} \text{Prob}(\xi_i) \sum_{j=1, j\neq i}^{N} p_{j,i} = \sum_{i=1}^{N} \text{Prob}(\xi_i)(1 - p_{i,i}) \ , \end{aligned} \tag{36}$$

ove, nell'ultima eguaglianza, si è sfruttato il fatto che la somma rispetto all'indice j delle $p_{j,i}$ è pari a 1 qualsiasi sia i. La probabilità di corretta ricezione è inceve data da

$$P(c) = 1 - P(e) = \sum_{i=1}^{N} \text{Prob}(\xi_i, \zeta_i) = \sum_{i=1}^{N} \text{Prob}(\xi_i) p_{i,i} \tag{37}$$

Facciamo ora qualche altro esempio di canale notevole.

Canale ideale.

Il canale ideale, o anche non rumoroso (noiseless), è tale che l'uscita permette di identificare univocamente l'ingresso. Ciò accade ad esempio se $N_X = N_\mathcal{Y}$ e la matrice $\boldsymbol{P}$ è diagonale. E' facile verificare che per tale canale la probabilità di errore è nulla.

Canale inutile.

Un canale si dice inutile se $N_\mathcal{X} = N_\mathcal{Y}$ ed è tale che

$$\text{Prob}(\zeta_j | \xi_i) = \text{Prob}(\zeta_j) \ , \qquad \forall i = 1, \ldots, N_\mathcal{X} \ , \qquad \forall j = 1, \ldots, N_\mathcal{Y} \ . \tag{38}$$

Si noti che tale relazione implica che $\text{Prob}(\xi_i | \zeta_j) = \text{Prob}(\xi_i)$, ovvero l'osservazione di un dato simbolo in uscita non da alcuna informazione sui simboli di ingresso. Per tale canale ingresso e uscita sono statisticamente indipendenti. E' facile verificare che per tale canale la matrice $\boldsymbol{P}$ ha le righe tutte uguali.

Analizziamo ora il canale in termini di entropia. Indicando con X e Y un generico simbolo in ingresso e in uscita al canale, già sappiamo che vale la relazione

$$H(X|Y) \leq H(X) \ ,$$

in quanto un condizionamento non può aumentare l'entropia. La quantità $H(X|Y)$ è detta *equivocazione,* e rappresenta la quantità di informazione mediamente persa sul canale. Tale quantità misura infatti l'incertezza che c'è sul simbolo di ingresso quando è nota l'uscita Y del canale. Il termine equivocazione sembra appropriato in quanto per il canale ideale è $H(X|Y) = 0$, mentre per il canale inutile $H(X|Y) = H(X)$, ovvero nel canale si perde tutta l'informazione relativa al simbolo in ingresso. Definita l'entropia congiunta tra ingresso e uscita

$$H(X,Y) = \sum_{i=1}^{N_{\mathcal{X}}} \sum_{j=1}^{N_{\mathcal{Y}}} \text{Prob}(\xi_i, \zeta_j) \log_2 \left(\frac{1}{\text{Prob}(\xi_i, \zeta_j)} \right) \ , \tag{39}$$

che si misura in bit per coppia di simboli, vale la relazione

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \ . \tag{40}$$

Dimostrare la (40) è alquanto immediato. Si ha infatti

$$\begin{gathered} H(X,Y) = \sum_{i=1}^{N_{\mathcal{X}}} \sum_{j=1}^{N_{\mathcal{Y}}} \text{Prob}(\xi_i, \zeta_j) \log_2 \left(\frac{1}{\text{Prob}(\xi_i, \zeta_j)} \right) = \\ \sum_{i=1}^{N_{\mathcal{X}}} \sum_{j=1}^{N_{\mathcal{Y}}} \text{Prob}(\xi_i, \zeta_j) \log_2 \left(\frac{1}{\text{Prob}(\xi_i|\zeta_j)\text{Prob}(\zeta_j)} \right) = \\ \sum_{i=1}^{N_{\mathcal{X}}} \sum_{j=1}^{N_{\mathcal{Y}}} \text{Prob}(\xi_i, \zeta_j) \log_2 \left(\frac{1}{\text{Prob}(\xi_i|\zeta_j)} \right) + \\ \sum_{i=1}^{N_{\mathcal{X}}} \sum_{j=1}^{N_{\mathcal{Y}}} \text{Prob}(\xi_i, \zeta_j) \log_2 \left(\frac{1}{\text{Prob}(\zeta_j)} \right) = H(X|Y) + H(Y) \ . \end{gathered} \tag{41}$$

In maniera del tutto analoga si dimostra che $H(X,Y) = H(Y|X) + H(X)$.

Si è visto che parte dell'informazione presente in ingresso al canale viene disperso in esso; in particolare, la quantità di informazione mediamente persa viene misurata dall'equivocazione $H(X|Y)$. Ma allora, la differenza tra l'informazione presente in ingresso al canale e l'equivocazione è una misura

della quantità di informazione che fluisce attraverso il canale e giunge quindi a destinazione. Si definisce quindi *flusso informativo* o anche *informazione mutua* tra ingresso e uscita dal canale la quantità

$$I(X;Y) = H(X) - H(X|Y) \ . \tag{42}$$

Tale termine si misura in bit/simbolo; si noti che vale anche la relazione $I(X;Y) = H(Y) - H(Y|X)$. Per il canale inutile è $I(X;Y) = 0$, mentre per il canale ideale è $I(X;Y) = H(X)$, ovvero tutta l'informazione presente in ingresso si ritrova in anche in uscita.

Esempio

Consideriamo un BSC con $p_{1,2} = p_{2,1} = p = 0.1$ e calcoliamo l'informazione mutua nel caso che i simboli di ingresso siano equiprobabili. Dal momento che il canale è simmetrico è facile verificare che anche i simboli di uscita sono equiprobabili, quindi è

$$\text{Prob}(\xi_1) = \text{Prob}(\xi_2) = \text{Prob}(\zeta_1) = \text{Prob}(\zeta_2) = 1/2 \ .$$

Le probabilità congiunte sono invece date da

$$\begin{aligned}
\text{Prob}(\xi_1, \zeta_1) &= \text{Prob}(\xi_1)\text{Prob}(\zeta_1|\xi_1) = 0.45 \ , \\
\text{Prob}(\xi_1, \zeta_2) &= \text{Prob}(\xi_1)\text{Prob}(\zeta_2|\xi_1) = 0.05 \ , \\
\text{Prob}(\xi_2, \zeta_1) &= \text{Prob}(\xi_2)\text{Prob}(\zeta_1|\xi_2) = 0.05 \ , \\
\text{Prob}(\xi_2, \zeta_2) &= \text{Prob}(\xi_2)\text{Prob}(\zeta_2|\xi_2) = 0.45 \ ,
\end{aligned}$$

da cui è facile verificare che $H(X,Y) = 1.469$. L'informazione mutua pertanto è:

$$I(X;Y) = H(X) - H(X|Y) = H(X) - (H(X,Y) - H(X)) = 1 - 1.469 + 1 = 0.531 \text{bit/simbolo}$$

In generale, se assumiamo che il BSC abbia parametro p, e poniamo $\text{Prob}(\xi_1) = \alpha$, è possibile dimostrare che l'entropia condizionale $H(Y|X)$ è indipendente da α e si esprime come

$$H(Y|X) = (1-p)\log_2 \frac{1}{1-p} + p \log_2 \frac{1}{p} = H(p) \ . \tag{43}$$

Inoltre, essendo

$$\text{Prob}(\zeta_1) = (1-p)\alpha + p(1-\alpha) \ , \qquad \text{Prob}(\zeta_2) = (1-p)(1-\alpha) + p\alpha \ ,$$

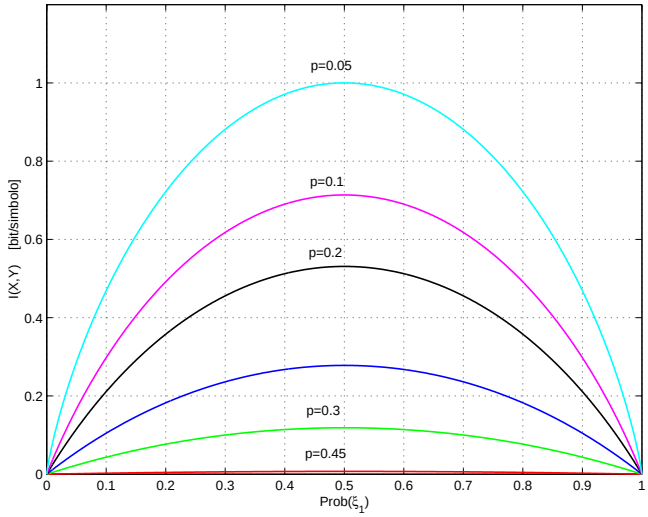

Figure 10: Informazione mutua di un BSC in funzione della probabilità del simbolo ξ_1 e per vari valori del parametro del BSC p.

si trova

$$H(Y) = [(1-p)\alpha + p(1-\alpha)] \log_2 \frac{1}{(1-p)\alpha + p(1-\alpha)} + [(1-p)(1-\alpha) + p\alpha] \log_2 \frac{1}{(1-p)(1-\alpha) + p\alpha} \,. \tag{44}$$

Sottraendo la (43) alla (44) si ottiene quindi un'espressione generale per l'informazione mutua. In Fig. 3.1 è riportata $I(X;Y)$ in funzione della probabilità α del simbolo ξ_1 e per vari valori del parametro p del BSC. Si vede che al tendere di p a 0.5 l'informazione mutua tende ovviamente a zero[7], mentre, viceversa, al diminuire di p l'informazione mutua tende a crescere. Per $p \to 0$ l'informazione mutua coincide con l'entropia di una sorgente binaria rappresentata in Fig. 1.2.1. E' importante notare come, indipendentemente dal parametro del BSC p, il massimo dell'informazione mutua si ottiene quando i simboli in ingresso al canale sono equiprobabili. Di fatto, in tali condizioni è massimo il flusso informativo che fluisce attraverso il canale. Tale osservazione ci guida naturalmente all'introduzione di un parametro fondamentale per ogni canale di comunicazione, ovvero la *capacità informazionale* del canale. Formalmente, la capacità di un canale discreto $\mathcal{C}$ è definita come il massimo dell'informazione mutua rispetto

[7] Si ricordi che per $p = 0.5$ il BSC diventa un canale inutile.

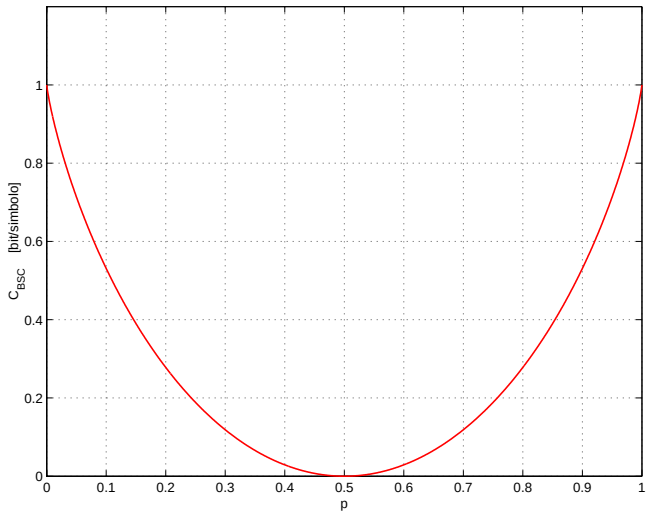

Figure 11: Capacità del canale BSC in funzione del parametro p.

alla distribuzione di probabilità dei simboli di ingresso, ossia

$$\mathcal{C} = \max_{P(\mathcal{X})} I(X;Y) \ , \tag{45}$$

ove si è indicato con $P(\mathcal{X})$ l'insieme di tutte le possibili distribuzioni di probabilità degli elementi dell'alfabeto di ingresso al canale $\mathcal{X}$.

Per un BSC, si è già visto che la distribuzione di probabilità che massimizza il flusso informativo è quella uniforme. Ponendo $\alpha = 0.5$ nella (44) si ottiene $H(Y) = 1$, ragion per cui la capacità del BSC si esprime come

$$\mathcal{C}_{\mathrm{BSC}} = 1 - H(p) = 1 + p\log_2 p + (1-p)\log_2(1-p) \ , \qquad [\mathrm{bit/simbolo}] \ . \tag{46}$$

Il grafico di $\mathcal{C}_{BSC}$ è riportato in Fig. 3.1. Come è lecito aspettarsi, la capacità è massima nei pressi di $p = 0$ e $p = 1$, dal momento che in tali condizioni il canale tende a comportarsi come un canale ideale con probabilità di errore nulla, mentre invece essa è nulla per $p = 0.5$. In tal caso, infatti il BSC è un canale inutile e non vi è modo di realizzare un processo di comunicazione attraverso di esso. Ritorneremo più volte sul concetto di capacità informazionale di un canale, in quanto tale parametro riveste un ruolo fondamentale nella teoria della trasmissione numerica. In generale, il calcolo della capacità di un canale discreto non è semplice nella maggior parte dei casi. Al riguardo, sono stati

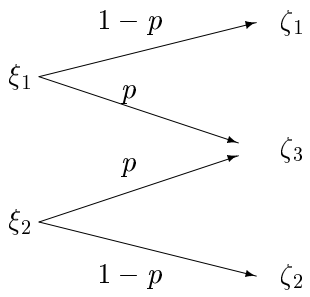

Figure 12: Il canale binario a cancellazione.

ideati degli algoritmi che permettono di calcolare numericamente la capacità mediante l'ausilio di un calcolatore elettronico. Nel seguito sono riportati ulteriori esempi di canali discreti per cui invece il calcolo della capacità è immediato.

Canale ideale

Dal momento che per tale canale l'equivocazione è nulla, si ha

$$\mathcal{C} = \max_{P(\mathcal{X})} [H(X) - H(X|Y)] = \max_{P(\mathcal{X})} H(X) = \log_2 N_{\mathcal{X}} \ . \tag{47}$$

Canale inutile

Essendo ora $H(X|Y) = H(X)$, è immediato verificare che per tale canale è $\mathcal{C} = 0$.

Canale binario a cancellazione

Il canale binario a cancellazione o BEC (*binary erasure channel*) è rappresentato in Fig. 12. Le uscite ζ_1 e ζ_2 corrispondono agli ingressi ξ_1 e ξ_2, mentre ζ_3 corrisponde ad un'uscita ambigua in corrispondenza della quale non viene presa alcuna decisione sul simbolo trasmesso. Tale canale modella quei sistemi in cui, in corrispondenza di simboli fortemente corrotti dal rumore, si preferisce evitare di prendere una decisione anziché rischiare di commettere un errore. Il calcolo della capacità per tale canale è alquanto immediato. Infatti, è immediato verificare che l'equivocazione può essere espressa come

$$H(X|Y) = H(X|Y = \zeta_3)\text{Prob}(Y = \zeta_3) + H(X|Y \neq \zeta_3)\text{Prob}(Y \neq \zeta_3) \ . \tag{48}$$

D'altra parte, è $H(X|Y \neq \zeta_3) = 0$, dal momento che l'incertezza sul simbolo di ingresso, osservando un simbolo di uscita che non sia ζ_3, è nulla. Inoltre, è anche $H(X|Y = \zeta_3) = H(X)$, in quanto

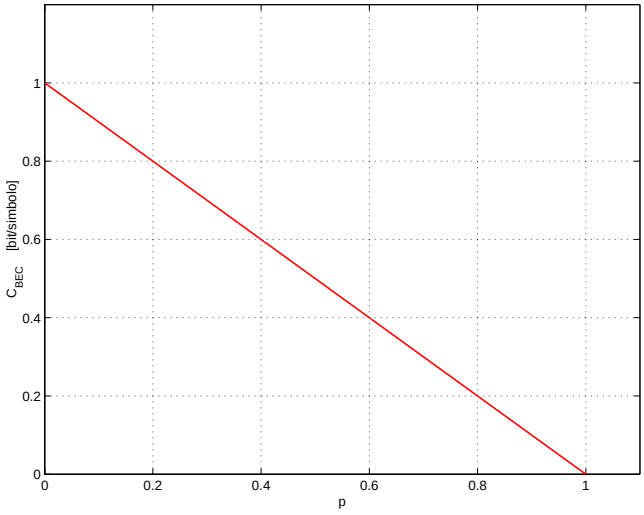

Figure 13: Capacità del canale BEC in funzione del parametro p.

l'osservazione del simbolo ζ_3 in uscita non diminuisce in alcun modo l'incertezza su quale sia stato il simbolo trasmesso. Di conseguenza, l'equivocazione si esprime come

$$H(X|Y) = pH(X) \,, \tag{49}$$

e l'informazione mutua è

$$I(X;Y) = H(X) - H(X|Y) = H(X)(1-p) \,. \tag{50}$$

Il massimo di $I(X;Y)$ si ottiene banalmente per simboli di ingresso equiprobabili, il che porta alla conclusione

$$\mathcal{C}_{\mathrm{BEC}} = 1 - p \,. \tag{51}$$

Il grafico della capacità del BEC è riportata in Fig. 3.1, in funzione del parametro p. Confrontando tale grafico con quello in Fig. 3.1 si osserva che la strategia a cancellazione permette di conseguire prestazioni migliori quando il canale è poco affidabile (ossia per p intorno a 0.5).

3.1.1 La disuguaglianza di Fano

La disuguaglianza di Fano è una relazione matematica che lega la probabilità di errore di un canale alla sua equivocazione. Da un punto di vista intuitivo è infatti evidente che tali quantità, in quanto en-

trambe misure della bontà di un canale, debbano poter essere messe in relazione. Bene, se introduciamo l'entropia

$$H(e) = -P(e)\log_2 P(e) - (1 - P(e))\log_2(1 - P(e)) \ ,$$

come il contenuto medio di informazione, espresso in bit, necessario a specificare se si è verificato un errore su un canale con probabilità di errore $P(e)$, per un canale discreto con $N_{\mathcal{X}} = N_{\mathcal{Y}}$ e caratterizzato da una probabilità di errore $P(e)$, vale la relazione

$$H(X|Y) \leq H(e) + P(e)\log_2(N_{\mathcal{X}} - 1) \ . \tag{52}$$

Per semplicità, ci limiteremo a dare una semplice giustificazione intuitiva di tale relazione. L'equivocazione $H(X|Y)$, e quindi l'informazione che è stata persa durante la trasmissione sul canale, è maggiorata dalla somma dell'incertezza $H(e)$ sul fatto che si sia o meno verificato un errore, e dall'incertezza sul simbolo effettivamente trasmesso (che è maggiorata da $\log_2(N_{\mathcal{X}} - 1)$) nel caso che si è verificato l'errore, moltiplicato per la probabilità di errore $P(e)$. ■

La disuguaglianza di Fano permette di giungere anche ad un'interessante rappresentazione grafica di un risultato fondamentale della teoria dell'informazione. Infatti, dal momento che vale la relazione

$$I(X;Y) = H(X) - H(X|Y) \leq \mathcal{C} \ ,$$

isolando al primo membro il termine $H(X)$ ed applicando la disuguaglianza di Fano si ha

$$H(X) \leq \mathcal{C} + H(X|Y) \leq C + H(e) + P(e)\log_2(N_{\mathcal{X}} - 1) \ . \tag{53}$$

In Fig. 3.1.1 è rappresentata graficamente tale disuguaglianza. Come si vede, la relazione (53) separa il piano $(P(e), H(X))$ in due zone. La zona sottostante la curva è la regione delle coppie ammissibili di valori di $P(e)$ e $H(\mathcal{X})$, mentre nella regione sovrastante la curva vi sono coppie di valori non raggiungibili. Da tale grafico si evince che se ci si accontenta di una probabilità di errore non nulla allora si possono inviare su un canale di capacità $\mathcal{C}$ anche simboli con entropia $H(\mathcal{X}) > \mathcal{C}$. Se, invece, si vogliono realizzare sistemi affidabili, ovvero con probabilità di errore $P(e)$ nulla o molto prossima allo zero, allora si deve operare con $H(\mathcal{X}) < C$. In altri termini, si può osservare che *se l'entropia dell'alfabeto di ingresso è maggiore della capacità di canale, è impossibile trasmettere l'informazione sul canale con una probabilità di errore piccola a piacere*. Tale affermazione rappresenta una versione semplificata

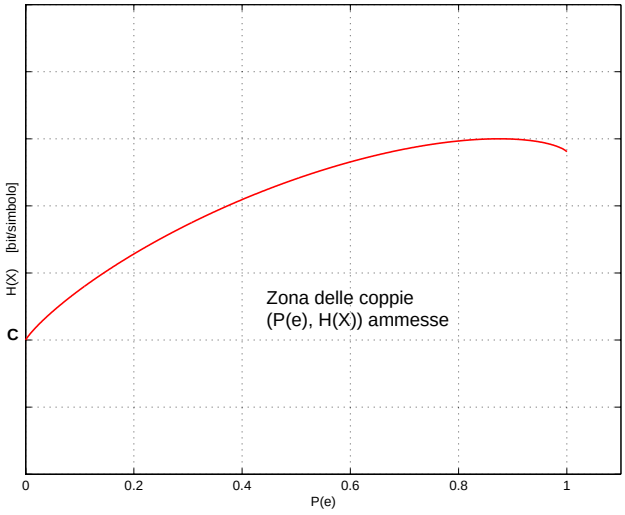

Figure 14: Grafico della funzione $\mathcal{C} + H(e) + P(e)\log_2(N_{\mathcal{X}} - 1)$ in funzione della probabilità di errore.

della proposizione inversa del teorema fondamentale della teoria dell'informazione e della trasmissione numerica; tale teorema va sotto il nome di *Teorema della codifica di canale*, e sarà enunciato nel seguito ma non dimostrato. Esso fu dimostrato nel 1948 da Claude E. Shannon. Tale teorema presuppone che tra il codificatore di sorgente ed in canale sia interposto un codificatore di canale che sostituisce blocchi di k bit con blocchi di n bit (ove, naturalmente $n > k$ in quanto il codificatore di linea introduce ridondanza).

Teorema sulla codifica di canale

Data una sorgente binaria d'informazione con entropia $H_\infty(\mathcal{X})$ bit/simbolo, ed un canale discreto senza memoria con capacità $\mathcal{C}$, esiste un codice di linea di tasso $R_c = k/n$ per il quale la probabilità di ricevere una parola codice errata $P_c(e)$ è limitata superiormente come segue

$$P_c(e) < 2^{-nE(R)} , \qquad R = R_c H_\infty(\mathcal{X}) , \tag{54}$$

ove $E(R)$ è una funzione U-convessa, non negativa, e decrescente di R nell'intervallo $[0, \mathcal{C}]$.

Sulla base della relazione (54) tre differenti strategie possono essere adottate per migliorare le prestazioni di un sistema di trasmissione numerico.

1. La prima possibilità è diminuire R facendo diminuire il rapporto k/n. Diminuire R porta ad un aumento della funzione $E(R)$ e quindi ad un decremento del limite sulla probabilità di parole codice

errata in ricezione. Tale strategia consiste essenzialmente nel ricorrere a codici che introducono maggior ridondanza, e, quindi, fissato il tasso di emissione dei simboli da parte del codificatore di sorgente, essa richiede un utilizzo più frequente del canale. Vedremo che ciò comporta usualmente l'utilizzo di una banda di frequenze maggiore da parte del sistema di trasmissione.

2. Altra possibilità è aumentare la capacità del canale ad esempio aumentando la potenza del segnale trasmesso. In tal modo, infatti, si ottiene un innalzamento della curva $E(R)$ e quindi, una più bassa $P_c(e)$. Il termine "aumentare la potenza del segnale trasmesso" può sembrare in un certo senso incomprensibile in questo momento, dal momento che stiamo trattando con simboli discreti e non con segnali. Si ricordi in ogni caso che un canale discreto è la schematizzazione della cascata di un modulatore, un canale fisico e un demodulatore.

3. Infine, l'ultima possibilità è aumentare il termine n, tenendo però fisso il rapporto k/n. Tale strategia non richiede nè un aumento della banda impegnata, nè un aumento della potenza trasmessa, e permette quindi di aumentare le prestazioni semplicemente aumentando la dimensione delle parole codice utilizzate. Ciò ovviamente comporta una maggiore complessità implementativa del codificatore e del decodificatore, come pure ritardi maggiori di decodifica.

Mentre le prime due strategie erano ben note, la scoperta della terza tecnica rappresenta uno dei risultati fondamentali della teoria sviluppata da Shannon.

3.2 Il Canale Gaussiano additivo a tempo discreto

Abbandoniamo per ora lo studio dei canali discreti e consideriamo un nuovo tipo di canali. In particolare, in questo paragrafo ci occuperemo dei canali ad ampiezza continua e a tempo discreto. Altrimenti detto, tali canali accettano simboli di ingresso appartenenti ad insiemi continui.

Ci occuperemo essenzialmente del canale Gaussiano additivo, il cui modello è rappresentato in Fig. 15. Si tenga presente che tale canale è diverso da quello riportato in Fig. 3, dal momento che il canale Gaussiano additivo in Fig. 15 accetta in ingresso simboli continui, e non forme d'onda come il canale in Fig. 3. Si assume quindi che ogni T_s secondi la sorgente invii sul canale un simbolo scelto da un insieme continuo, e che il canale riproduca in uscita il simbolo in ingresso con l'aggiunta di una variabile aleatoria gaussiana a media nulla e varianza σ_ν^2. L'ipotesi che il rumore additivo sia Gaussiano è, da

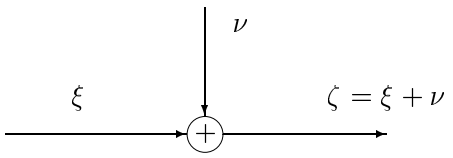

Figure 15: Il canale gaussiano additivo.

un lato, molto conveniente dal punto di vista della trattabilità analitica del modello, e, dall'altro lato, giustificata in diverse situazioni di interesse pratico. In questa sezione, indicheremo con $\mathcal{X}$ l'insieme dei simboli di ingresso al canale (ovviamente, $\mathcal{X}$ contiene ora infiniti elementi), e con $\mathcal{Y}$ l'insieme dei possibili simboli in uscita. Essendo $\zeta = \xi + \nu$, è immediato rendersi conto che, dato il simbolo in ingresso ξ, l'uscita ζ è una variabile aleatoria gaussiana di media ξ e varianza σ_ν^2.

Per poter calcolare la capacità del canale Gaussiano additivo, è opportuno giungere a una definizione del concetto di entropia anche per alfabeti continui. Si definisce entropia di una variabile aleatoria ξ che assume valori nell'insieme $\mathcal{X}$ e avente pdf $f_\xi(x)$ la quantità

$$H(\mathcal{X}) = -\int_{-\infty}^{+\infty} f_\xi(x) \log_2 f_\xi(x) dx \ . \tag{55}$$

Tale definizione sembra una diretta estensione della definizione (6) di entropia di una sorgente discreta, ove, tenendo conto della continuità dei valori assunti dai simboli dell'alfabeto, si è sostituita la sommatoria con un integrale. Tuttavia, l'entropia in (55) è profondamente diversa dall'entropia (6). Innanzitutto, l'entropia di una variabile aleatoria continua non è sempre non negativa come l'entropia di un alfabeto discreto. In particulare, l'entropia può assumere valori nell'insieme dei numeri reali. Ad esempio, è immediato verificare che se la variabile aleatoria ξ ha una distribuzione uniforme in $[-a, a]$, la sua entropia è

$$H(\mathcal{X}) = \int_{-a}^{a} \frac{1}{2a} \log_2(2a) dx = \log_2(2a) < 0 \quad \text{per } \ a < 0.5 \ . \tag{56}$$

Di conseguenza, per variabili continue, viene meno l'interpretazione dell'entropia come il contenuto informazionale medio, espresso in bit, associato al generico simbolo dell'alfabeto $\mathcal{X}$. D'altra parte, un simbolo appartenente ad un insieme continuo costituito da infiniti elementi può essere specificato ricorrendo ad un numero infinito di bit, ragion per cui il contenuto informazionale associato all'emissione di

un simbolo continuo è in generale infinito. Sulla base di tali ragioni, l'entropia in (55) viene usualmente chiamata *entropia differenziale*.

Analogamente alla definizione (55), è possibile poi definire anche l'entropia associata all'osservazione di una coppia di simboli. Ad esempio, se facciamo riferimento alla coppia di variabili aleatorie (ξ, ζ), l'entropia differenziale ad essa associata è

$$H(\mathcal{X}, \mathcal{Y}) = -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{\xi\zeta}(x, y) \log_2 f_{\xi\zeta}(x, y) dx dy \ . \tag{57}$$

L'entropia condizionale è invece data da

$$H(\mathcal{X}|\mathcal{Y}) = -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{\xi\zeta}(x, y) \log_2 f_{\xi|\zeta}(x|y) dx dy \ . \tag{58}$$

e

$$H(\mathcal{Y}|\mathcal{X}) = -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{\xi\zeta}(x, y) \log_2 f_{\zeta|\xi}(y|x) dx dy \ . \tag{59}$$

Assumendo poi che sia $H(\mathcal{X})$ che $H(\mathcal{Y})$ siano finite, è possibile dimostrare, analogamente a quanto fatto per il caso discreto, che valgono le seguenti relazioni.

$$\begin{aligned} &H(\mathcal{X}, \mathcal{Y}) \leq H(\mathcal{X}) + H(\mathcal{Y}) \ , \\ &H(\mathcal{X}, \mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}) = H(\mathcal{Y}) + H(\mathcal{X}|\mathcal{Y}) \ , \\ &H(\mathcal{X}|\mathcal{Y}) \leq H(\mathcal{X}) \ , \\ &H(\mathcal{Y}|\mathcal{X}) \leq H(\mathcal{Y}) \ . \end{aligned} \tag{60}$$

In tali relazioni, le disuguaglianze si riducono a delle uguaglianze nel caso che le variabili aleatorie ξ e ζ siano statisticamente indipendenti.

E' stato visto che, per variabili aleatorie discrete, l'entropia è massima quando i simboli della sorgente sono equiprobabili. Nel caso continuo si ha invece il seguente risultato

Teorema

Sia ξ una variabile aleatoria continua con pdf $f_\xi(x)$. Se la varianza σ_ξ^2 di ξ è finita, allora l'entropia differenziale ad essa associata $H(\xi)$ è finita e soddisfa la relazione

$$H(\xi) \leq \frac{1}{2} \log_2(2\pi e \sigma_\xi^2) \ , \tag{61}$$

ed il vincolo si riduce ad un'eguaglianza se e solo se ξ è una variabile aleatoria Gaussiana.

Prova: La dimostrazione di tale risultato è omessa per motivi di semplicità e brevità.

In altre parole, tra tutte le variabili aleatorie aventi varianza finita, l'entropia è massima quando la pdf è gaussiana. Tale risultato è di notevole importanza e sarà utilizzato per ricavare l'espressione della capacità del canale Gaussiano additivo.

In analogia a quanto fatto per canali discreti, è possibile definire, per un canale continuo avente alfabeto di ingresso $\mathcal{X}$ e alfabeto di uscita $\mathcal{Y}$ l'informazione mutua, che si definisce come

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) \ . \tag{62}$$

La definizione è del tutto analoga a quella data per canali discreti. Inoltre, è opportuno osservare come, benché l'entropia differenziale non può essere interpretata, per i motivi già discussi, come la quantità di informazione associata all'emissione del generico simbolo dell'alfabeto, l'informazione mutua per variabili aleatorie continue conserva invece il significato di flusso di informazione che si propaga attraverso il canale. Analogamente al caso di canale discreto, è poi possibile definire la capacità di canale come il massimo, rispetto alla distribuzione del simbolo di ingresso, dell'informazione mutua. Altrimenti detto, si ha

$$\mathcal{C} = \max_{f_\xi(x)} I(\mathcal{X};\mathcal{Y}) \qquad \text{bit/simbolo} \ . \tag{63}$$

Il calcolo della capacità del canale Gaussiano additivo è alquanto semplice. Infatti, dal momento che

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X})$$

si ha che massimizzare $I(\mathcal{X};\mathcal{Y})$ rispetto alla distribuzione dell'alfabeto di ingresso equivale a massimizzare esclusivamente $H(\mathcal{Y})$. Infatti, il termine $H(\mathcal{Y}|\mathcal{X})$ per via del condizionamento è indipendente dalla distribuzione del simboli di ingresso[8]. Dunque, $I(\mathcal{X};\mathcal{Y})$ è massima quando è massima l'entropia del simbolo di uscita $H(\mathcal{Y})$ è massima. Per il teorema precedentemente enunciato $H(\mathcal{Y})$ è massimizzata quando l'uscita del canale è gaussiana; ma, poiché l'uscita del canale è espressa come $\zeta = \xi + \nu$, e il rumore ν è gaussiano, l'uscita ζ è gaussiana se anche il simbolo di ingresso ξ è gaussiano. In tali condizioni, indicando con σ_ξ^2 la varianza dell'ingresso ξ, l'entropia dell'uscita è espressa come

[8]Ricorda che condizionatamente all'ingresso ξ, l'uscita $\zeta = \xi + \nu$ è gaussiana a media nulla e varianza σ_ν^2, per cui è $H(\mathcal{Y}|\mathcal{X}) = \frac{1}{2}\log_2(2\pi e\sigma_\nu^2)$.

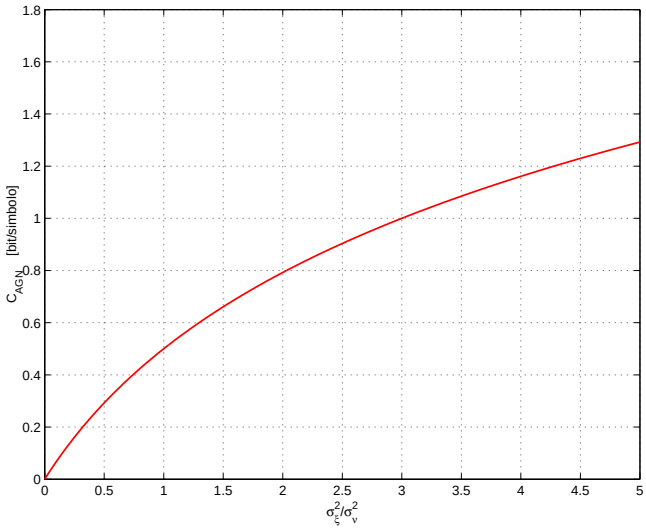

Figure 16: Capacità del canale Gaussiano additivo in funzione del rapporto tra la varianza dei simboli di ingresso e la varianza del rumore additivo.

$H(\mathcal{Y}) = \frac{1}{2}\log_2(2\pi e(\sigma_\xi^2 + \sigma_\nu^2))$. Ne consegue quindi che la capacità del canale gaussiano additivo è:

$$\begin{aligned} \mathcal{C}_{\mathrm{AGN}} = \max_{f_\xi(x)} H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) &= \frac{1}{2}\log_2(2\pi e(\sigma_\xi^2 + \sigma_\nu^2)) - \frac{1}{2}\log_2(2\pi e(\sigma_\nu^2)) \\ &= \frac{1}{2}\log_2\left(1 + \frac{\sigma_\xi^2}{\sigma_\nu^2}\right) \qquad \text{bit/simbolo .} \end{aligned} \tag{64}$$

L'unità di misura di $\mathcal{C}_{\mathrm{AGN}}$ è il bit per simbolo, o, equivalentemente, il bit per uso del canale. Riassumendo, si è calcolata l'espressione della capacità del canale Gaussiano additivo e, inoltre, si è dimostrato che il flusso informativo attraverso tale canale è massimizzato quando i simboli di ingresso sono realizzazioni di una variabile aleatoria Gaussiana. In Fig. 3.2 è rappresentata la capacità (64) in funzione del rapporto $\sigma_\xi^2/\sigma_\nu^2$.

3.3 Capacità del canale AWGN a banda limitata

Sino ad ora ci si è occupati esclusivamente di canali a tempo discreto. Cominciamo ora a considerare il caso di canali tempo continui, ovvero canali di comunicazione che accettano al loro ingresso forme d'onda. E' opportno osservare come tali canali siano di grandissimo interesse, dal momento che la maggior parte dei canali reali di comunicazione sono a tempo continuo. Consideriamo quindi un canale

AWGN a banda limitata, che sia capace di trasportare segnali il cui spettro è non nullo all'interno della banda $(-W, W)$. Si dirà quindi che W è la banda monolatera del canale di trasmissione, e che il canale è un canale in banda base, dal momento che la sua banda passante è concentrata nell'intorno della frequenza nulla. I risultati che deriveremo sono tuttavia validi anche per un canale la cui banda passante coincide con l'intervallo $(f_0 - 1/(2W), f_0 + 1/(2W)$. Indicato quindi con $s(t)$ il segnale trasmesso su tale canale (e ovviamente si assume che lo spettro di $s(t)$ sia nullo all'esterno delle frequenze $(-W, W)$), tale canale produce in uscita il segnale

$$r(t) = s(t) + n(t) \ ,$$

ove $n(t)$ è assunto essere un processo aleatorio Gaussiano con PSD $\frac{\mathcal{N}_0}{2} \prod \left(\frac{f}{2W} \right)$. La capacità di tale canale può essere calcolata mediante il ragionamento che segue. Dal momento che tale canale tratta segnali a banda limitata, a norma del teorema del campionamento è possibile sostituire ai segnali tempo-continui la successione dei loro campioni presa a frequenza $2W$. Ne consegue quindi che la trasmissione sul canale AWGN a banda limitata è del tutto equivalente alla trasmissione, sul canale Gaussiano additivo tempo-discreto, dei campioni del segnale da trasmettere prelevati alla frequenza di Nyquist $2W$. Per il teorema del campionamento, ciascun campione di ingresso avrà una varianza σ_ξ^2 pari alla potenza P del segnale $s(t)$, mentre i campioni del rumore avranno varianza $\sigma_\nu^2 = N_0 W$. Di conseguenza, la capacità del canale AWGN a banda limitata, espressa in bit al secondo è pari a $2W$ volte la capacità del canale Gaussiano additivo con simboli di ingresso a varianza P e disturbo con varianza $\mathcal{N}_0 W$. In definitiva, si ha quindi

$$\mathcal{C}_{\mathrm{AWGN-BL}} = W \log_2 \left(1 + \frac{P}{\mathcal{N}_0 W} \right) \qquad \text{bit/secondo} \ . \tag{65}$$

L'equazione (65) esprime, a norma della generalizzazione del teorema di Shannon a canali continui, la massima velocità con cui è possibile trasmettere informazione in modo affidabile sul canale AWGN a banda limitata. Possiamo infatti affermare che, *dato uno schema di trasmissione numerica che trasmette su un canale con data capacità $\mathcal{C}$ dati ad una velocità pari ad R_b bit/secondo, se $R_b > \mathcal{C}$ non esiste alcuno schema di trasmissione che permette di ottenere prestazioni affidabili, ovvero capace di rendere la probabilità di errore sulla sequenza ricevuta piccola a piacere. Se, invece, $R_b < \mathcal{C}$, è possibile individuare schemi di trasmissione e ricezione capaci di rendere la probabilità che la sequenza ricevuta sia errata arbitrariamente piccola*. La capacità di canale fissa quindi un limite alla massima velocità

con cui è possibile trasmettere su un dato canale. Sfortunatamente, la dimostrazione del teorema di Shannon non è costruttiva, ragion per cui essa dimostra l'esistenza di schemi di trasmissione affidabili quando $R_b < \mathcal{C}$, ma non fornisce alcuna indicazione su come questi schemi siano fatti. La teoria della trasmissione numerica si focalizza quindi sullo studio e la ricerca di sistemi di trasmissione che siano capaci di offrire prestazioni affidabili con velocità di comunicazione quanto più prossime possibile alla capacità di canale. E' d'interesse notare che la capacità del canale AWGN a banda limitata (65) dipende dal rapporto $P/\mathcal{N}_0$, ed in particolare è una funzione crescente della potenza del segnale trasmesso e decrescente della PSD del disturbo additivo. In particolare, su canali non rumorosi la capacità è infinita. Si noti inoltre che, nel caso si utilizzi un modulatore numerico M-ario con intervallo di segnalazione T, il bit-rate R_b col quale si trasmette è pari a $\log_2 M/T$. D'altra parte, indicando con $\mathcal{E}$ l'energia media delle forme d'onda utilizzate dal modulatore numerico, si ha:

$$P = \frac{\mathcal{E}}{T} = \frac{\mathcal{E}_b \log_2 M}{T} = \mathcal{E}_b R_b \ , \tag{66}$$

ove si è indicato con $\mathcal{E}_b$ l'energia mediamente spesa per ogni bit trasmesso sul canale. Sostituendo la (66) nella (65) si ottiene

$$\mathcal{C}_{\mathrm{AWGN-BL}} = W \log_2 \left(1 + \frac{\mathcal{E}_b}{\mathcal{N}_0} \frac{R_b}{W} \right) \qquad \text{bit/secondo} \ . \tag{67}$$

Si è quindi giunti ad un'espressione diversa per la capacità del canale AWGN, che dipende esplicitamente dai rapporti $\mathcal{E}_b/\mathcal{N}_0$ ed R_b/W. Il primo termine, ovvero $E_b/\mathcal{N}_0$, è usualmente detto *contrasto di energia*, ed esprime il rapporto tra l'energia mediamente spesa per ogni bit trasmesso ed il doppio della PSD del rumore additivo. Tale termine è usualmente riportato sull'ascissa dei grafici che rappresentano l'andamento della probabilità di errore per le varie modulazioni numeriche. Spesso tale termine viene anche indicato, o meglio confuso, col rapporto segnale rumore, dal momento che, per $R_b/W = 1$, esso coicide numericamente col rapporto tra la potenza di segnale utile e la potenza di rumore. Il termine R_b/W è invece usualmente denominato *efficienza spettrale*. Tale termine si misura in bit al secondo per hertz e indica quanti bit al secondo si trasmettono per ogni hertz della banda impegnata. Ovviamente, conseguire elevati valori dell'efficienza spettrale porta alla favorevole condizione di trasmettere molti bit al secondo (ovvero molta informazione) in poca banda. Viceversa bassi valori dell'efficienza spettrale implicano che la banda a disposizione è utilizzata in maniera inefficiente. Di fatto, nel progetto dei sistemi di trasmissione numerica è di interesse riuscire a conseguire prestazioni affidabili con bassi valori

del contrasto di energia (al fine di risparmiare potenza) e alti valori di efficienza spettrale. Nella realtà, tuttavia, vedremo che tale obiettivo non è raggiungibile, e ci si deve accontentare o di valori elevati di $\mathcal{E}_b/\mathcal{N}_0$ e R_b/W (e si parlerà in tal caso di sistemi efficienti in banda e inefficienti in potenza) o di valori bassi per entrambi i parametri (e si parlerà quindi di sistemi efficienti in potenza ma inefficienti in banda). Al fine di chiarire tale affermazione, si osservi che il confine tra i sistemi di trasmissione affidabili e quelli non affidabili si ha quando $R_b = \mathcal{C}$; in tali condizioni, si ha

$$\begin{aligned} R_b = \mathcal{C} = W \log_2 \left(1 + \tfrac{\mathcal{E}_b}{\mathcal{N}_0}\tfrac{R_b}{W}\right) \qquad &\Rightarrow \qquad 2^{R_b/W} = 1 + \tfrac{\mathcal{E}_b}{\mathcal{N}_0}\tfrac{R_b}{W} \\ \Rightarrow \qquad \frac{\mathcal{E}_b}{\mathcal{N}_0} &= \frac{2^{R_b/W} - 1}{R_b/W} \end{aligned} \tag{68}$$

In Fig. 3.3 è rappresentata il grafico dell'ultima relazione in (68). Tale figura è usualmente nota come *Piano di Shannon*. La zona di spazio sottostante la curva è la regione ove $R_b < \mathcal{C}$; in tale regione sono situati tutti i sistemi di trasmissione numerica realmente esistenti. La zona superiore alla curva corrisponde invece al caso che $R_b > \mathcal{C}$; in tale regione, a norma del teorema di Shannon, non è possibile realizzare sistemi di trasmissione affidabili, e quindi nessun sistema reale si troverà ad operare con coppie di valori di efficienza spettrale e contrasto di energia appartenenti a tale zona. In tale grafico è anche evidenziata la retta orizzontale $R_b/W = 1$. Tale retta divide la zona di spazio efficiente in banda da quella non efficiente in banda, che corrisponde al caso che l'efficienza spettrale sia minore dell'unità. Da tale grafico si evince come i punti caratterizzati da elevata efficienza spettrale e basso contrasto di energia si trovino nella zona "inaffidabile", e quindi non è possibile realizzare sistemi di trasmissione operanti con elevata efficienza spettrale e basso contrasto di energia. Nel seguito avremo modo di ritornare al piano di Shannon e di fornire ulteriori dettagli.

3.4 Capacità del canale AWGN

Consideriamo, infine, il canale AWGN a banda illimitata. Il calcolo della capacità di tale canale è immediato se si osserva che tale canale può ottenersi come limite del canale AWGN a banda limitata quando $W \to \infty$. La capacità di tale canale è quindi espressa come

$$\mathcal{C}_{\mathrm{AWGN}} = \lim_{W \to +\infty} \mathcal{C}_{\mathrm{AWGN-BL}} = \frac{P}{\mathcal{N}_0} \log_2 e = \frac{\mathcal{E}_b}{\mathcal{N}_0} R_b \log_2 e \ . \tag{69}$$

Si noti che, imponendo che sia $R_b < \mathcal{C}$, si ottiene $\mathcal{E}_b/\mathcal{N}_0 > \log 2$. Ovvero, non vi è modo di ottenere prestazioni affidabili sul canale AWGN a banda infinita se il contrasto di energia è inferiore al logaritmo

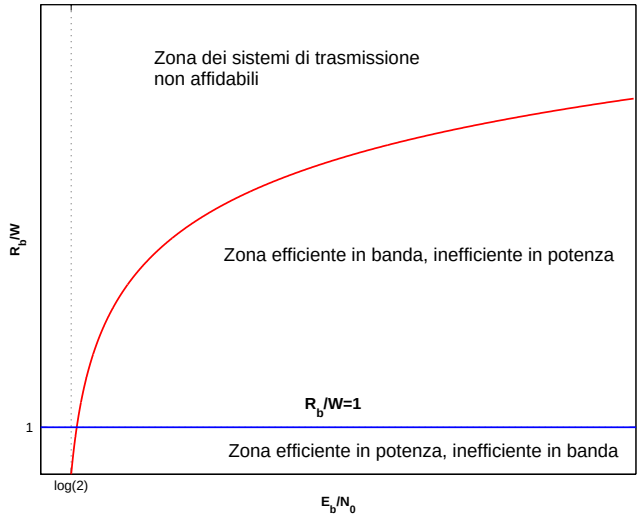

Figure 17: Il piano di Shannon.

naturale di 2. Espresso in dB, tale limite è pari a -1.59 dB, ed è usualmente approssimato con -1.6 dB. Il limite $\mathcal{E}_b/\mathcal{N}_0 > -1.6$dB rappresenta il limite di Shannon per canali a banda infinita. Vedremo nel seguito che tale limite è raggiungibile con una semplice segnalazione non codificata.