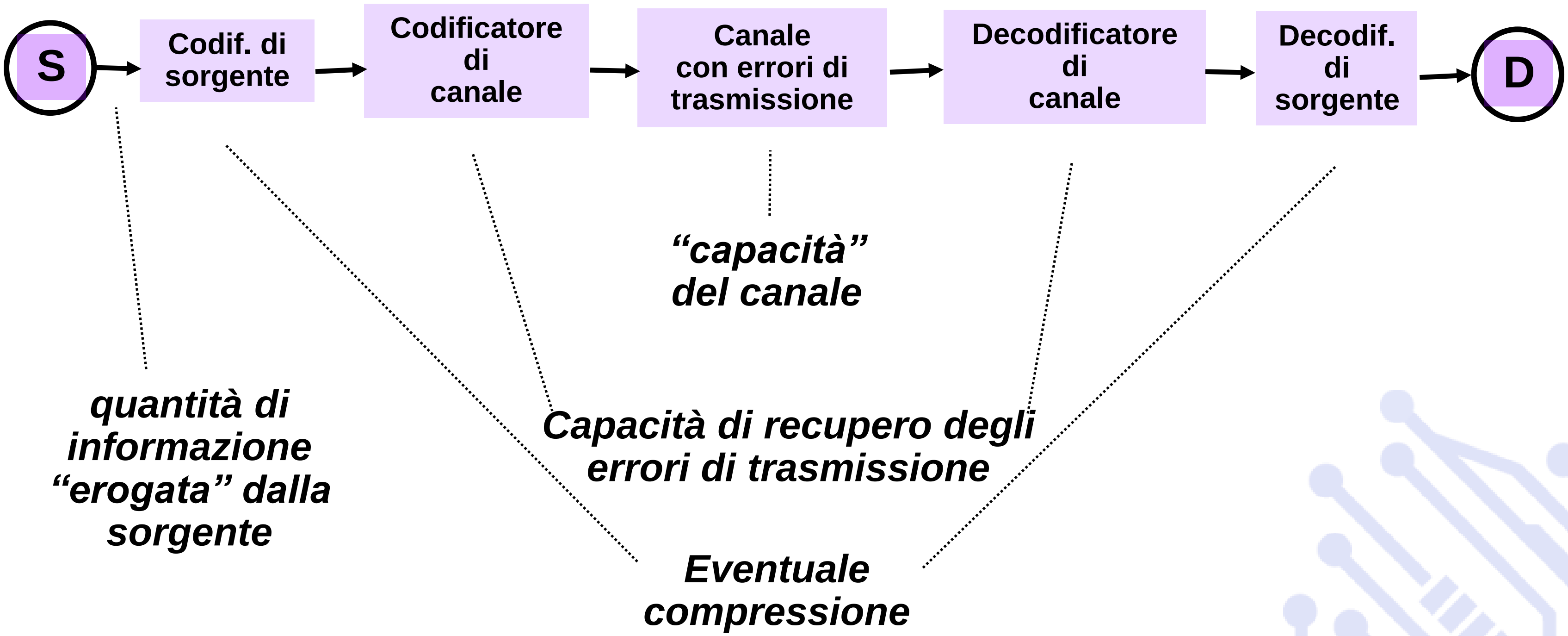
Teoria dell’Informazione e della Trasmissione
TIT

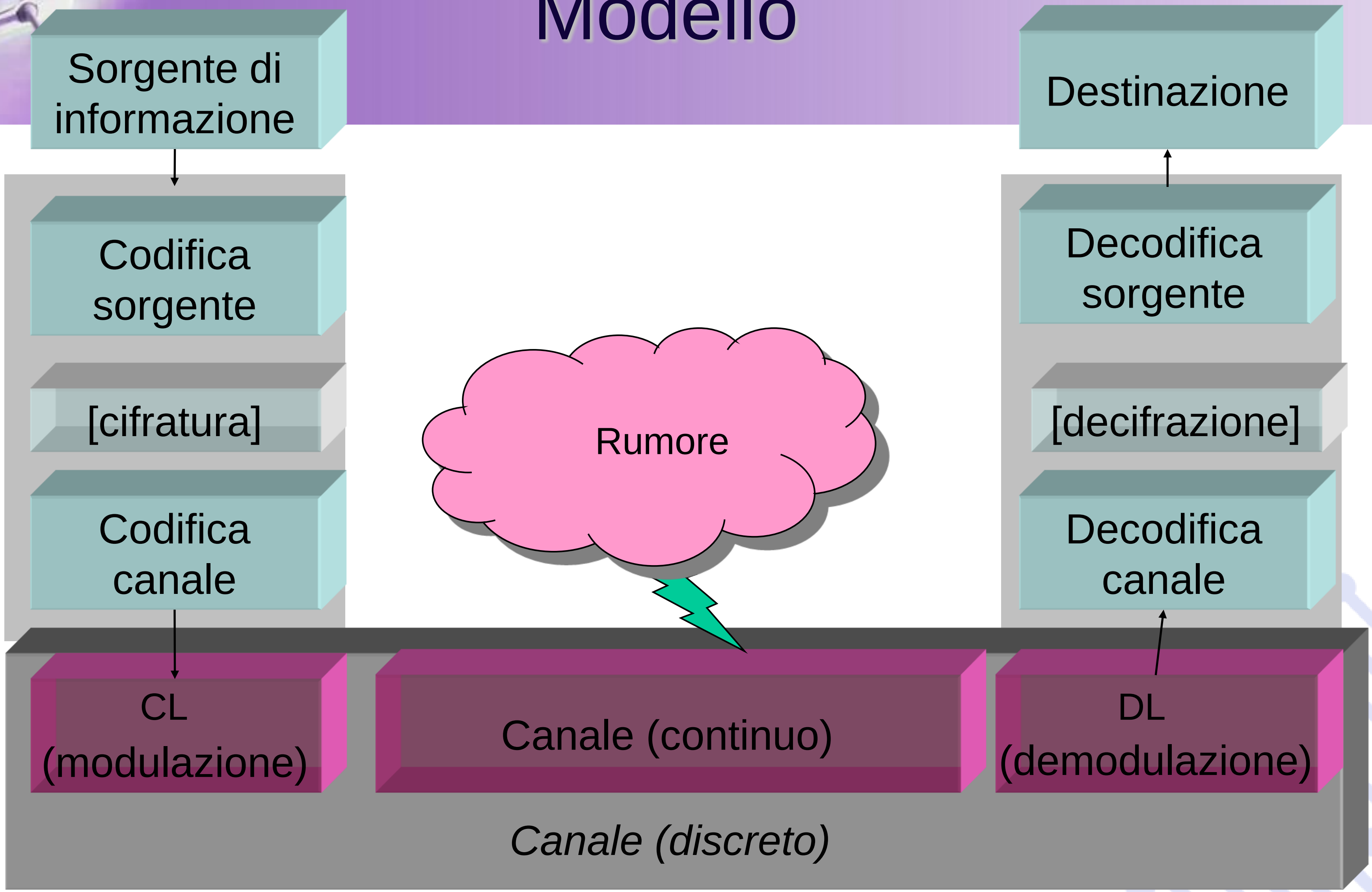
Che cos'è la teoria della informazione?

- ✓ E' una teoria matematica che tratta l'aspetto "simbolico" dell'informazione
- ✓ La teoria dell'informazione ha origine nel 1948 con la pubblicazione da parte di C. E. Shannon di "A Mathematical theory of Communication"
- ✓ Inizia un approccio quantitativo alla nozione di informazione
- ✓ Viene introdotto un modello di un sistema di comunicazione in cui compare una sorgente di informazione ed una di rumore

Modello generale di un sistema di comunicazione

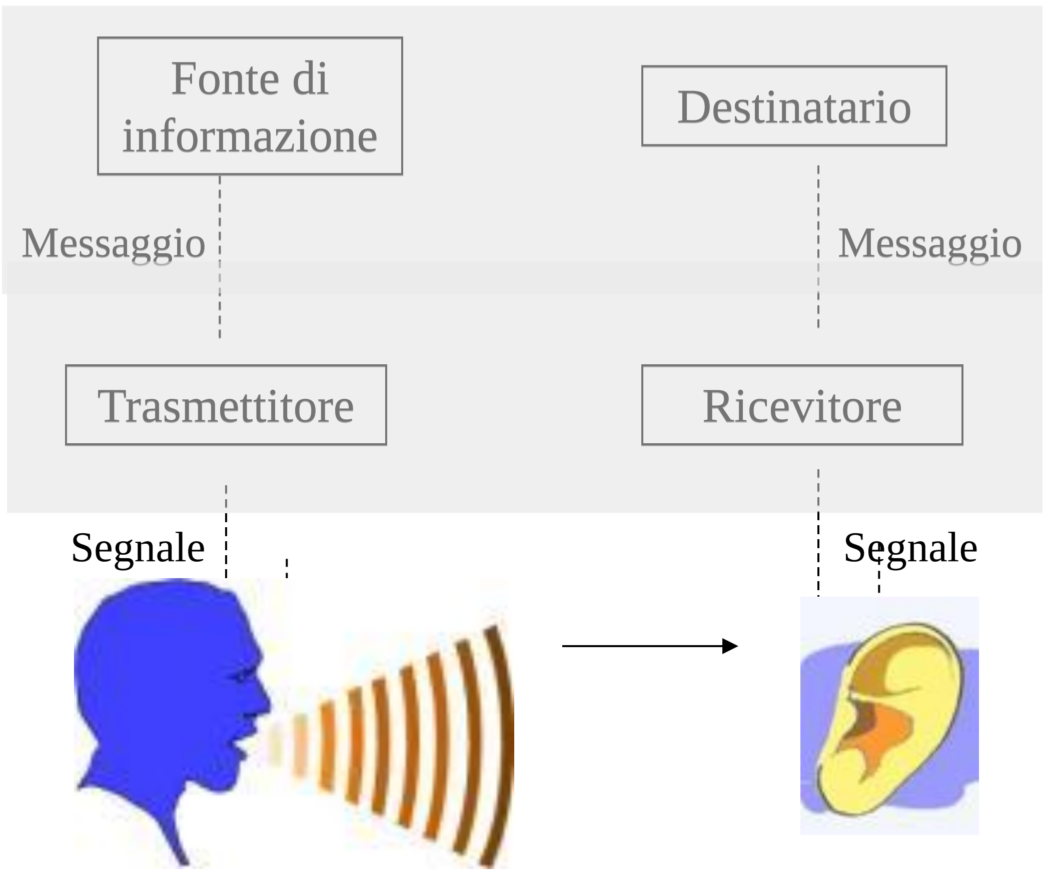


Modello

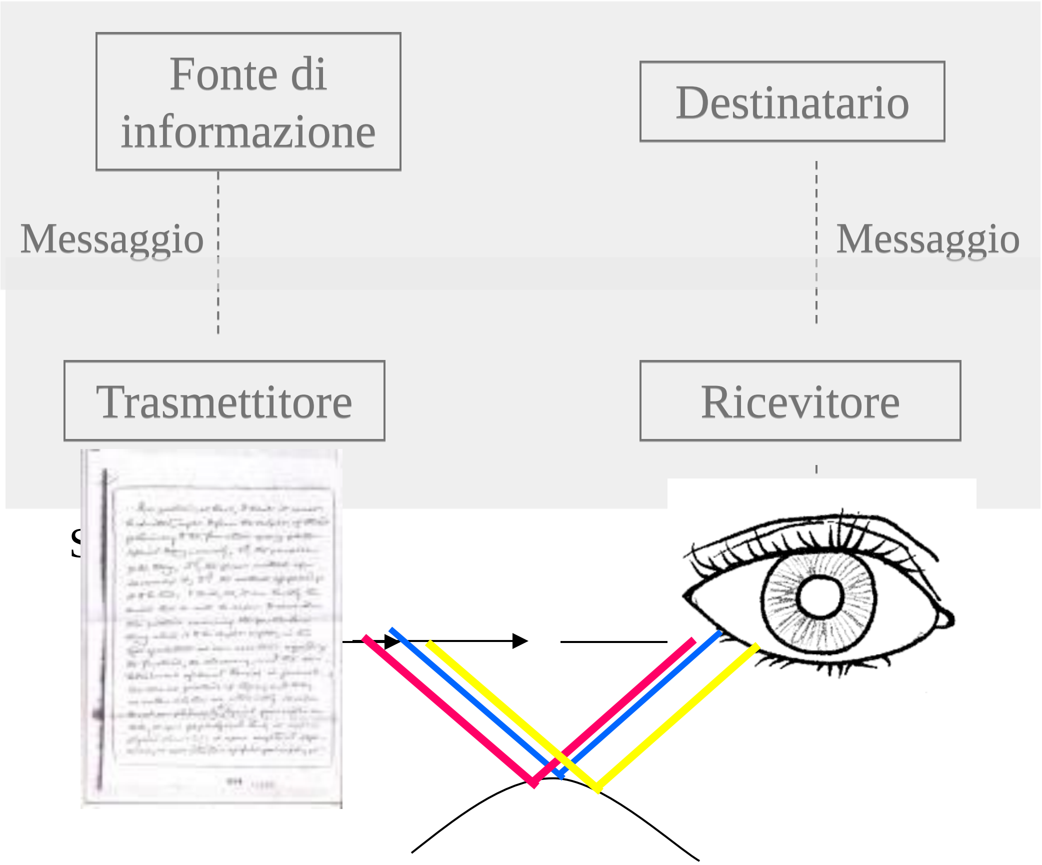


Legenda | [ ] = opzionale
( ) = esempio

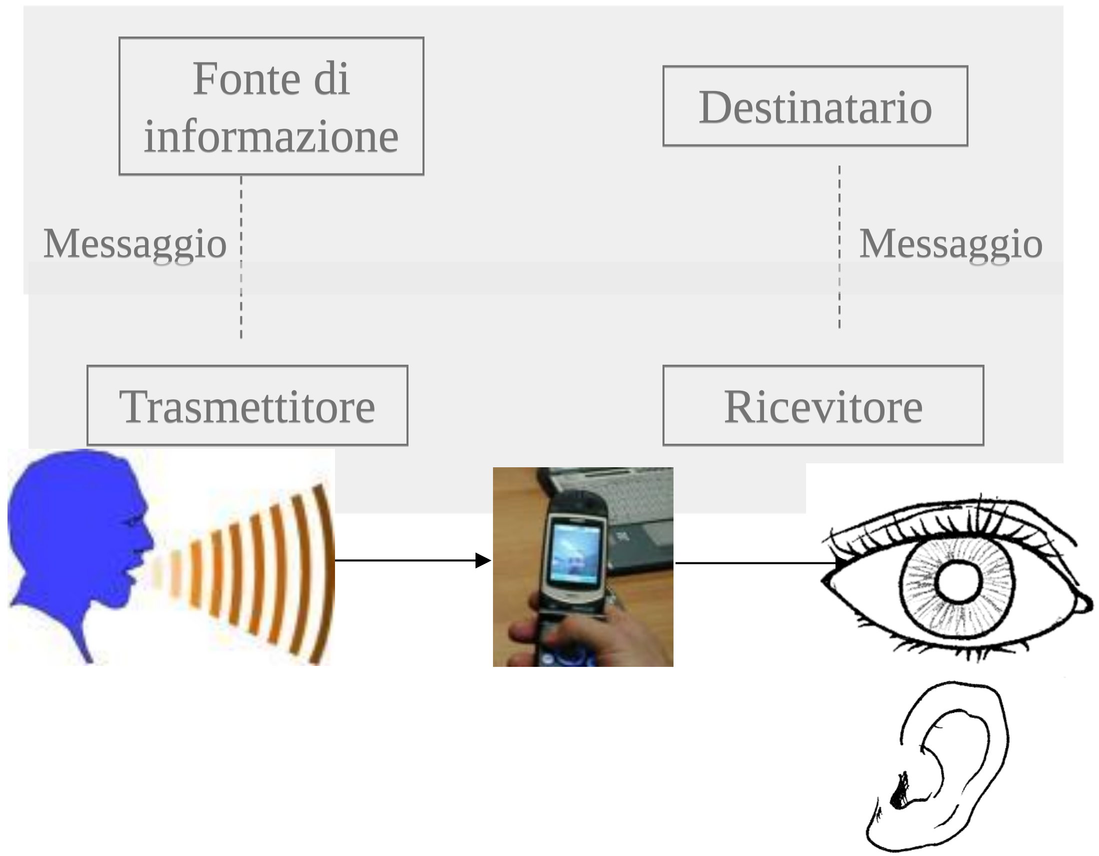
Il modello standard della comunicazione



Il modello standard della comunicazione



Il modello standard della comunicazione



Sorgenti discrete

Sorgenti e messaggi informativi

- Come misurare la "quantità d'informazione" emessa da una sorgente?
- L'uscita di una sorgente (il messaggio) consiste in una sequenza di simboli scelti in un insieme finito (l'alfabeto di sorgente)
- Un meccanismo probabilistico regola l'emissione di simboli consecutivi nel messaggio

Frequenza relativa di una parola

Si tratta del numero totale di volte che una data parola compare in un contesto diviso il numero complessivo delle occorrenze di tutte le parole.

Come sinonimo di *frequenza relativa* si utilizza **probabilità**

Quantità di informazione

Secondo Shannon **la quantità di informazione trasmessa da un simbolo dipende dalla probabilità di occorrenza del simbolo**.

Maggiore è questa probabilità minore è la quantità di informazione trasmessa dal simbolo;

Un simbolo la cui probabilità di occorrenza è pari ad 1 (cioè del 100%) introduce una quantità di informazione uguale a zero.

Un simbolo con bassa probabilità convoglia invece una grande quantità di informazione

Quantità di informazione

Se i simboli emessi da una sorgente hanno tutti la stessa probabilità di emissione, cioè:

$$P = 1/N$$

Maggiore è il numero dei simboli N, più bassa è la loro probabilità e maggiore è l'informazione trasportata da ciascuno di essi.

La definizione di Shannon

L'unità di misura dell'informazione è il bit, che corrisponde all'informazione associata all'emissione di uno tra i due simboli equiprobabili

Infatti questo è considerato il caso più semplice di sorgente

Shannon ne diede una formulazione matematica

Misura dell'informazione: ENTROPIA dell'alfabeto di sorgente

- ✓ La quantità di informazione $I(x_i)$ associata al simbolo i-esimo dell'alfabeto di sorgente è legata alla incertezza riguardo la sua emissione

- ✓ maggiore incertezza i.e. minore probabilità => maggiore informazione:

$$I(x_j) > I(x_i) \quad \text{se} \quad p_j < p_i \tag{1}$$

- ✓ La quantità ed il contenuto informativo associato all'emissione di due simboli indipendenti è la somma delle due informazioni:

$$\text{Se } P(x_i, x_j) = P(x_i)P(x_j) \quad \text{allora} \quad I(x_i, x_j) = I(x_i) + I(x_j) \tag{2}$$

Misura dell'informazione

Una formulazione che soddisfa i requisiti (1) e (2) è la seguente:

$$I(x_i) = \log_a\left(\frac{1}{p_i}\right)$$

a = base del logaritmo, determina l'unità di misura di $I(x_i)$

Se *a=2* allora $I(x_i)$ si misura in bit (binary digit)

In tal caso si ha: $I(x_1) = I(x_2) = \log_2 2 = 1$

e la corretta identificazione di uno tra due simboli **equiprobabili** porta esattamente una quantità d'informazione pari ad 1bit

Misura dell’informazione

- ✓ Messaggi diversi in genere trasportano quantità diverse di informazione

Si deve allora definire per una sorgente una quantità d’informazione media (ENTROPIA)

- ✓ l’entropia della sorgente rappresenta il numero medio di simboli binari (digit) <u>necessario</u> per rappresentare ogni simbolo del messaggio

Misura dell'informazione: ENTROPIA dell'alfabeto di sorgente

- ✓ Nel seguito si assume $a = 2$
- ✓ Partendo dalla definizione di $I(x_i)$ si può caratterizzare l'intero alfabeto S della sorgente tramite la quantità:

$$H(\mathrm{S}) = \sum_{i=1}^{M} p_i I(x_i) = \sum_{i=1}^{M} p_i \log\left(\frac{1}{p_i}\right)$$

- ✓ $H(S)$ è chiamata ENTROPIA dell'alfabeto di sorgente, ed indica il contenuto informativo medio della sorgente S.
- ✓ $H(S)$ si misura in bit/simbolo

Esempio

- Entropia di un alfabeto di sorgente composto di $M=4$ simboli aventi le probabilità:

$$p_1 = 1/2,\ \ p_2 = 1/4,\ \ p_3 = p_4 = 1/8$$

si calcola così:

$$H(\mathrm{S}) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + 2\cdot\frac{1}{8}\log 8 = 1.75 \ \text{bit/simbolo}$$

- Entropia nel caso di alfabeto di sorgente con M simboli equiprobabili, per M generico

si calcola così: $$H(\mathrm{S}) = \sum_{i=1}^{M}\frac{1}{M}\log M = \log M$$

Misura dell'informazione: ENTROPIA dell'alfabeto di sorgente

"*L'entropia* $H(S)$ *di un alfabeto di sorgente con* M *simboli soddisfa la:*

$$H(x) \leq \log M$$

con l'eguaglianza nel caso di simboli equiprobabili"

Codifica binaria dell’informazione

- In base alle proprietà dell’informazione così formulata, l’uscita di una sorgente può essere “sostituita” da una stringa di simboli binari che trasportano la stessa quantità di informazione ed hanno un numero medio di digit per simbolo emesso della sorgente originaria vicino quanto desiderato all’entropia della sorgente

- Il canale di comunicazione è il mezzo fisico usato per connettere la sorgente d’informazione con il suo utente

Sorgenti e messaggi informativi

- Tuttavia, qualunque sia la distanza tra sorgente e destinatario, il canale di comunicazione è affetto da vari disturbi: RUMORE e DISTORSIONI
- La presenza di un canale disturbato può rendere difficile l'esatta riproduzione del messaggio emesso presso l'utente.
- Il progettista di un sistema di comunicazione cercherà comunque di fornire all'utente una replica *il più vicino possibile* al messaggio originale
- A tale scopo si farà ricorso ad un'idonea codifica di canale…

Esempio 1

Da una sorgente binaria con simboli equiprobabili ed indipendenti viene emesso il seguente messaggio

M = 1111000.

Determinare la quantità di informazione associata al messaggio.

P(s) = 1/2

I(s) = log [1/P(S)]= log [1/1/2] = log 2 = 1

I(M) = 7 * 1 = 7 bit

Esempio 2

Una sorgente binaria con simboli non equiprobabili ed indipendenti emette con probabilità:

P(0) = 1/4 e P(1) =   3/4

Determinare la quantità di informazione associata al messaggio M=11100

Le probabilità sono diverse, quindi:

I(0) = log [1/P(0)]= log [1/1/4]= log 4 = 2

I(1) = log 1/P(1) = log [1/3/4] = log 4/3 = log 4 – log 3 = 2 – 1,58 = 0,42

I(M) = 3* 0,42 + 2*2 = 5,26

Nota: per calcolare il logaritmo in base 2 di un numero che non sia potenza di 2, occorre applicare la regola:

$$\mathrm{Log}_b\, N = \frac{\log N}{\log b}$$

Esercizio 3

Calcolare l'entropia di una sorgente caratterizzata dalle seguente distribuzione di probabilità:

$P(s_1)= 1/8$

$P(s_2)= 1/8$

$P(s_3)= 1/4$

$P(s_4)= 1/2$

Calcolare la quantità di informazione associata al messaggio:

$s_1 s_2 s_1 s_3 s_4$