



Università Mediterranea di Reggio Calabria



Teoria dell'informazione

Giuseppe Ruggeri

Università "Mediterranea" di Reggio Calabria, Italy





- Cos'è l'informazione?
- E' possibile misurare l'informazione?
- Limiti matematici nella rappresentazione dell'informazione.
- Limiti matematici nella trasmissione dell'informazione.
- Rappresentazione dei segnali Multimediali:
 - Voce
 - Audio
 - Video

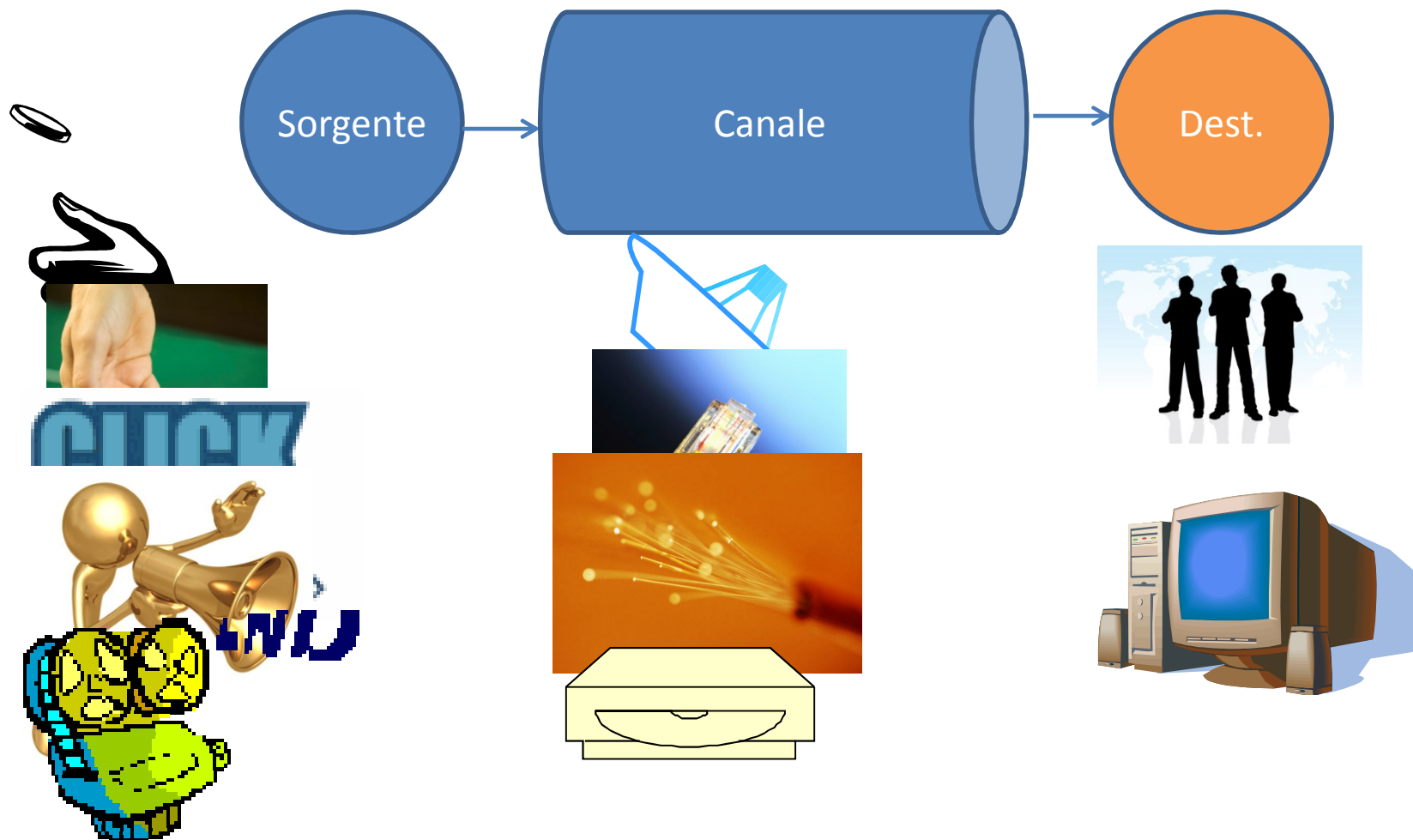




Università Mediterranea di Reggio Calabria

Modello di comunicazione

ARTS
Laboratory for Advanced Research
into Telecommunication Systems





La Teoria Matematica dell'Informazione è stata formulata da C. E. Shannon (1948), consente di:

- Misurare l'informazione emessa da una sorgente;
- Definire la quantità massima di informazione che un canale può trasportare;
- Definire il metodo migliore per codificare (rappresentare) l'informazione nel modo più "comprensibile" dal canale.





Ciascuna sorgente ha un proprio alfabeto S

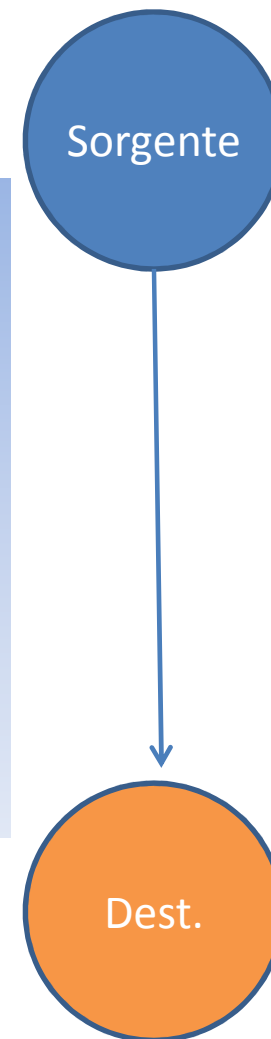
→ $S = \{\text{Testa, Croce}\}$

Una sorgente emette informazione ogni qual volta emette uno tra gli N possibili simboli del suo alfabeto

o anche

Si ottiene informazione quando si apprende del verificarsi di uno tra N possibili eventi
(emissione di uno tra N simboli)

→ $S = \{\text{Fonemi, i suoni basilari che costituiscono il parlato}\}$





Un evento porta con sé tanta più informazione quanto più improbabile è il suo verificarsi.



$$\rightarrow S = \{\text{Testa, Croce}\} \rightarrow P_s = \{P_T, P_C\} = \left\{ \frac{1}{2}, \frac{1}{2} \right\}$$



$$\rightarrow S = \{1, 2, 3, 4, 5, 6\} \rightarrow P_s = \{P_1, P_2, P_3, P_4, P_5, P_6\} = \left\{ \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right\}$$

L'informazione associata ad un simbolo dipende dalla probabilità P con cui viene emesso.

$$I_i = F(p_i)$$



$$F(\bullet)??$$



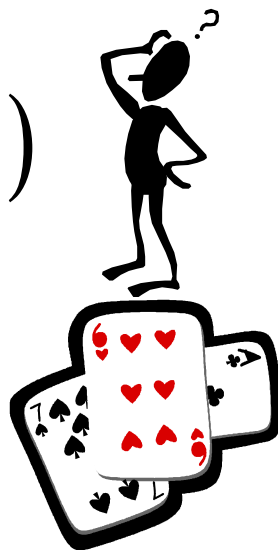


$$I_i = F(p_i)$$

$$F(\bullet)??$$



$$p_i = \frac{1}{40}$$



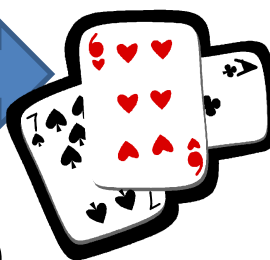
$$I_i = F\left(\frac{1}{40}\right)$$



$$p_c = \frac{1}{4}$$



$$p_{6/c} = \frac{1}{10}$$



$$I_i = F\left(\frac{1}{4}\right) + F\left(\frac{1}{10}\right)$$

$$I_i = F\left(\frac{1}{40}\right) = F\left(\frac{1}{4} * \frac{1}{10}\right) = F\left(\frac{1}{4}\right) + F\left(\frac{1}{10}\right)$$





$$I_i = F(p_i)$$



$$F(\bullet)??$$

$$I_i = F\left(\frac{1}{40}\right) = F\left(\frac{1}{4} * \frac{1}{10}\right) = F\left(\frac{1}{4}\right) + F\left(\frac{1}{10}\right)$$

$$I_i = -\log_x\left(\frac{1}{40}\right) = -\log_x\left(\frac{1}{4} * \frac{1}{10}\right) = -\log_x\left(\frac{1}{4}\right) - \log_x\left(\frac{1}{10}\right)$$



Vogliamo inoltre che l'informazione associata con l'evento elementare sia unitaria

$$I_{t/c} = -\log_x\left(\frac{1}{2}\right) = 1 \Rightarrow x = 2$$

$$I_i = \log_2\left(\frac{1}{p_i}\right) \text{ bits}$$



Quant'è l'informazione emessa mediamente da una sorgente?

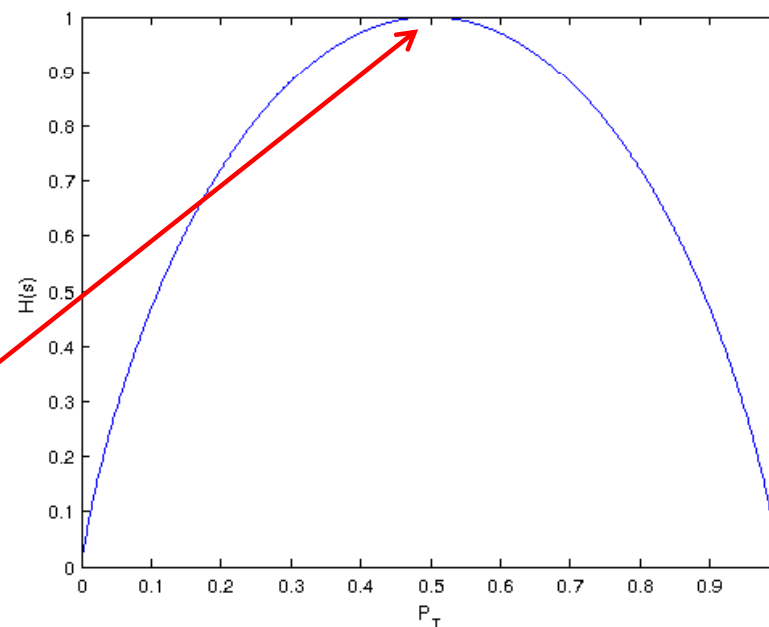


$$S = \{\text{Testa}, \text{Croce}\} \longrightarrow P_s = \{P_T, P_c\} = \{P_T, (1 - P_T)\}$$

$$H(s) = \sum_i P_i \log_2 \left(\frac{1}{P_i} \right)$$

$$H(s) = P_T \log_2 \frac{1}{P_T} + (1 - P_T) \log_2 \frac{1}{1 - P_T}$$

L'informazione emessa mediamente da una sorgente è massima quando i simboli sono equiprobabili





La rappresentazione dell'Informazione

È possibile inviare i simboli emessi dalla sorgente su un canale?

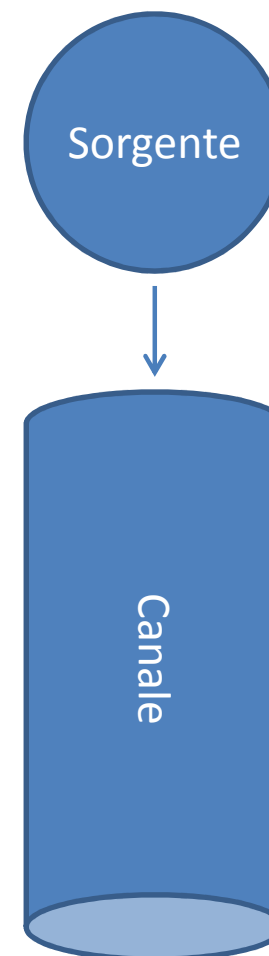
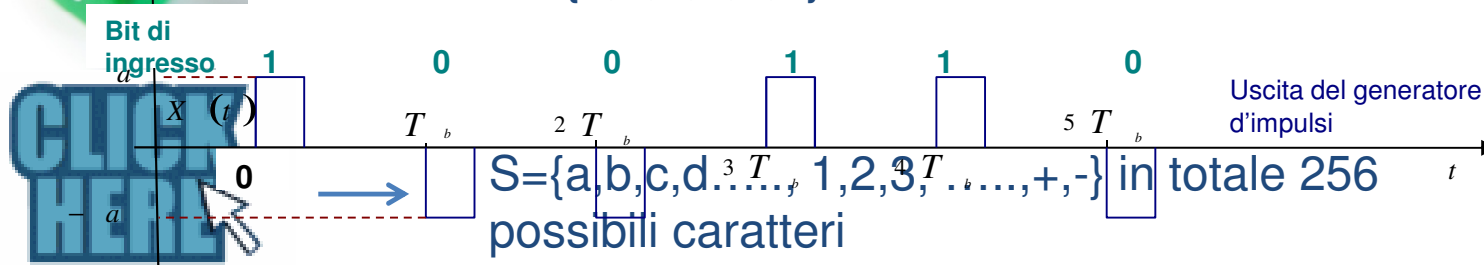


→ $S = \{\text{Testa, Croce}\}$

In generale NO: sorgenti e canali hanno alfabeti diversi



→ $S = \{1, 2, 3, 4, 5, 6\}$



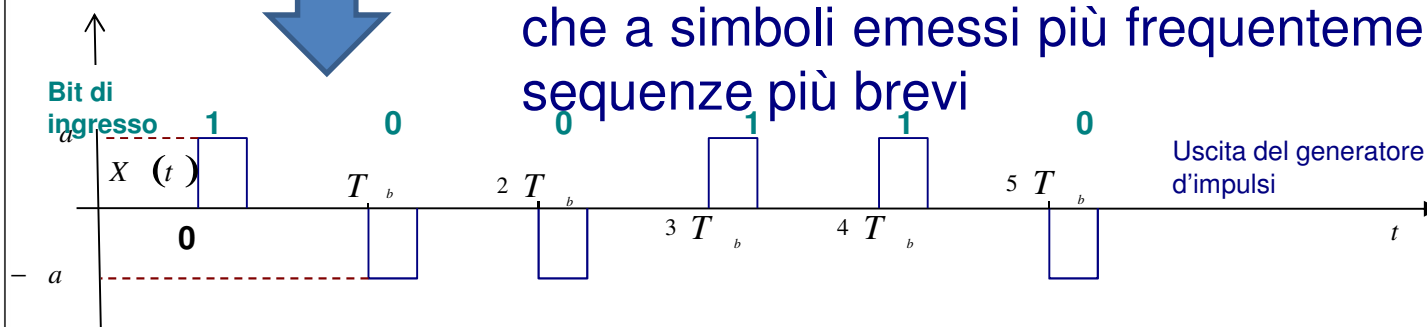
Nasce l'esigenza di rappresentare i Simboli emessi dalla Sorgente nell'alfabeto "compreso" dal Canale: **TRANSCODIFICA**



$$S = \{1, 2, 3, 4, 5, 6\} \quad P_s = \{P_1, P_2, P_3, P_4, P_5, P_6\}$$

• Obiettivo

cercare di assegnare a ciascun simbolo di sorgente una sequenza di simboli di canale (parole di codice) in modo che a simboli emessi più frequentemente corrispondano sequenze più brevi



	carattere	$P(a)$	Codice I	Codice II	Codice III
univocamente	a_1	1/2	1	0	0
•Regola del prefisso:	a_2	1/4	00	10	01
decodificabili	a_3	1/8	001	110	011
istantaneamente	a_4	1/8	10	111	111
decodificabili					





Codifica di Huffman (1952)

Passo 1

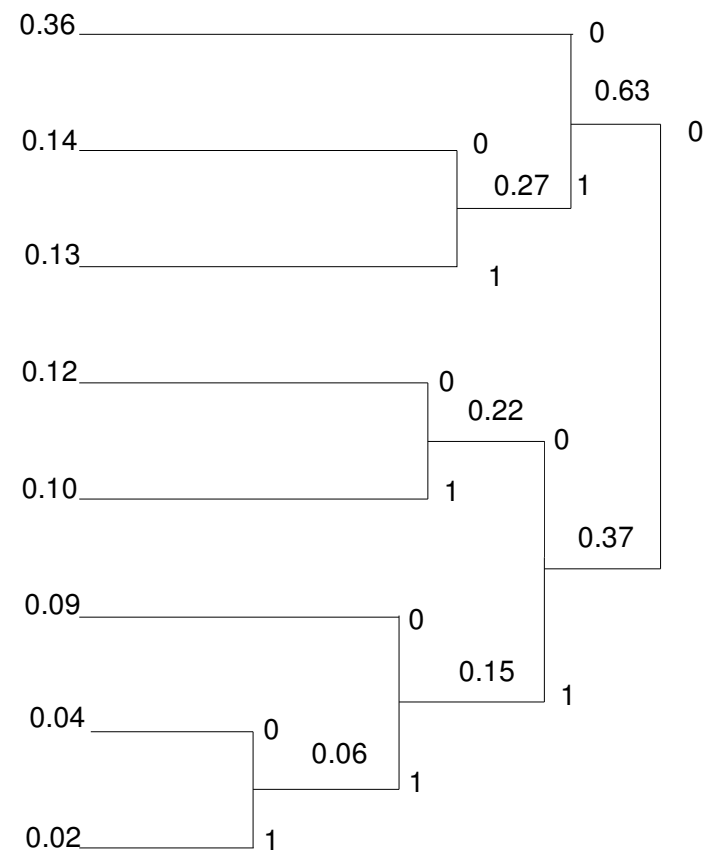
Si ordinino gli L^k simboli b_j secondo valori non crescenti delle loro probabilità e si indichi con $\{d^{(n)}\}$ la sequenza così riordinata di D simboli

Passo 2

Si raggruppino gli ultimi due simboli $d^{(D)}$ e $d^{(D-1)}$ in un simbolo equivalente con probabilità
 $P[d^{(D)}] + P[d^{(D-1)}]$

Passo 3

Posto $D=D-1$ si ripetano i passi 1 e 2 finché non rimanga un unico simbolo





Teorema fondamentale della codifica di sorgente

C. Shannon (1948)

« Data una sorgente S di entropia $H(S)$, N simboli emessi dalla sorgente possono essere compressi in più di $NH(s)$ simboli con una probabilità di perdita di informazione piccola a piacere, al tendere di N all'infinito; al contrario, se sono compresse in meno di $NH(s)$ simboli è virtualmente **certo che una parte dell'informazione andrà persa.** »



$\{S_1, S_2, S_3, S_4, \dots, S_N\}$
→

$$\bar{L} = \frac{1}{N} \sum_{i=1}^N \text{lunghe}(S_i)$$

Lunghezza media parole di
codice

Deve essere: $H(s) < \bar{L} \leq H(s) + \varepsilon$

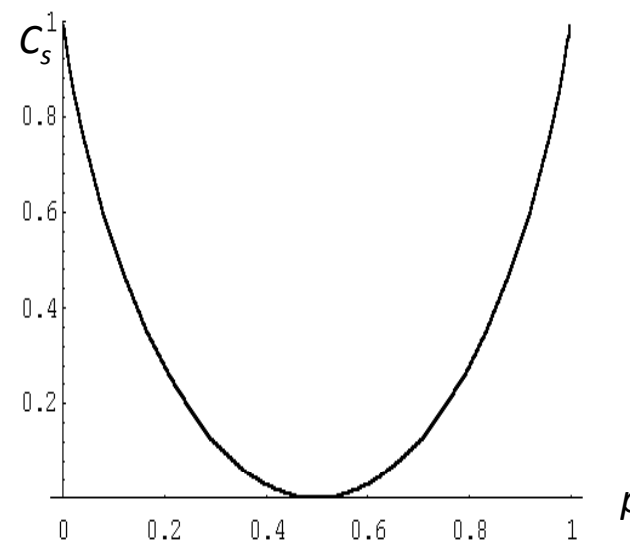
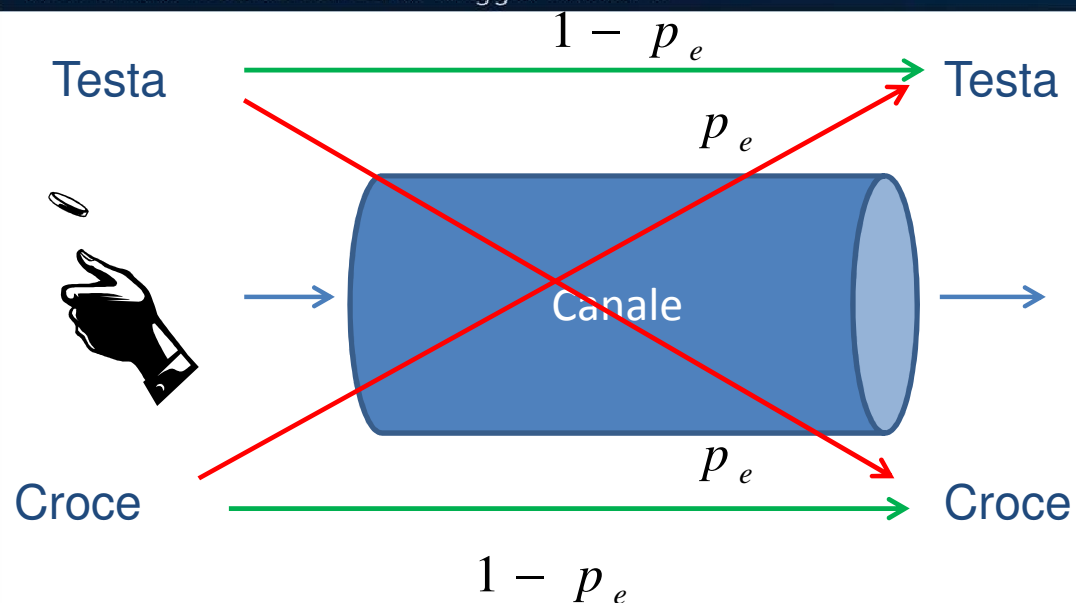
Per raggiungere i requisiti di molte applicazioni moderne (MP3, DTV, SatTV) è **necessario “perdere” una parte dell'informazione.**





Università Mediterranea di Reggio Calabria

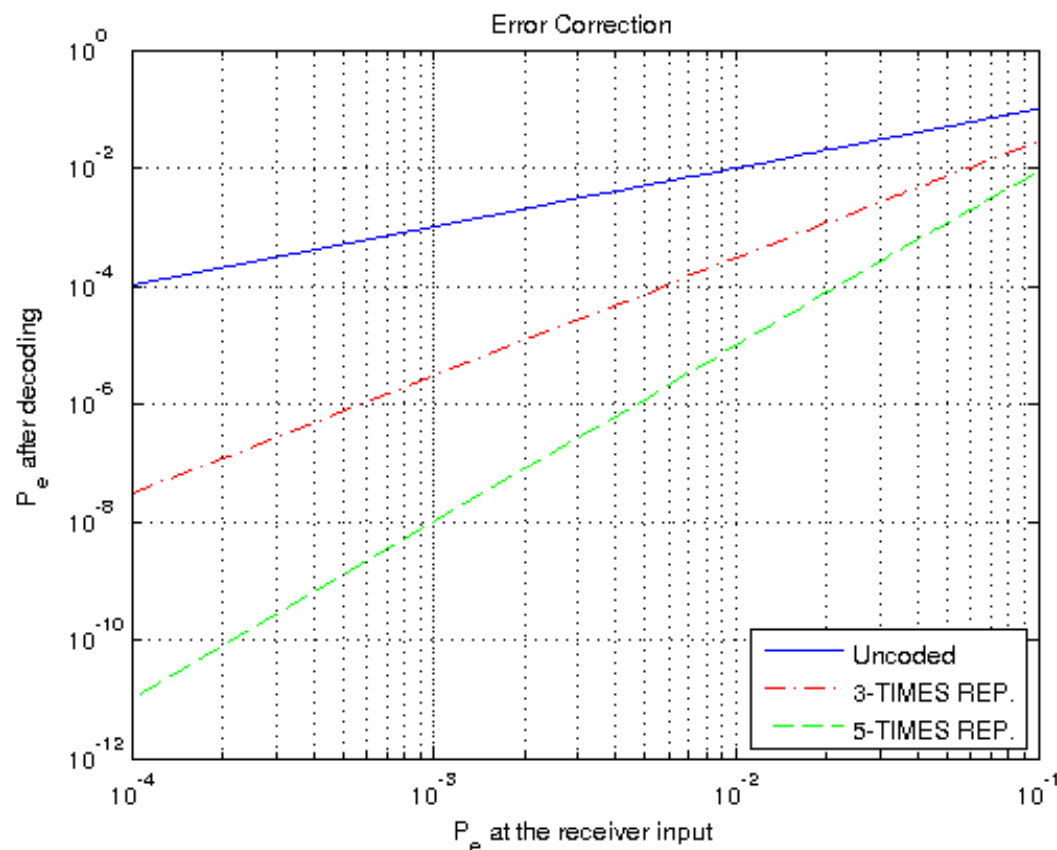
I canali commettono errori!



Il ricevitore rimane con il dubbio se il simbolo emesso dal canale sia effettivamente quello trasmesso dalla sorgente: **si perde informazione**.

Per ogni canale è possibile definire una capacità **C**: la quantità **massima di informazione** che può trasferire **per ciclo di segnalazione**.





p_e



Testa, Testa, Testa

Testa, Testa, Croce

Croce, Testa, Testa

Testa, Croce, Testa

Croce, Croce, Testa

Croce, Testa, Croce

Testa, Croce, Croce

Croce, Croce, Croce

E' possibile ridurre a piacere gli errori purchè l'informazione inviata sul canale non superi la sua Capacità

