

IN420 – Teoria dell'Informazione

Presentazione del corso

Prof. Vincenzo Bonifaci
Università degli Studi Roma Tre

Contatti ed orari

- Email: vincenzo.bonifaci@uniroma3.it
- Web: <http://ricerca.mat.uniroma3.it/users/vbonifaci/>
- Ricevimento: in remoto, su appuntamento
- Lezioni:
 - Mercoledì, ~~8.00-10.00~~ 14.00-16.00 ?
 - Giovedì, ~~12.00-14.00~~ 14.00-16.00 ?
- Esercitazioni e complementi:
 - Martedì, 14.00-16.00

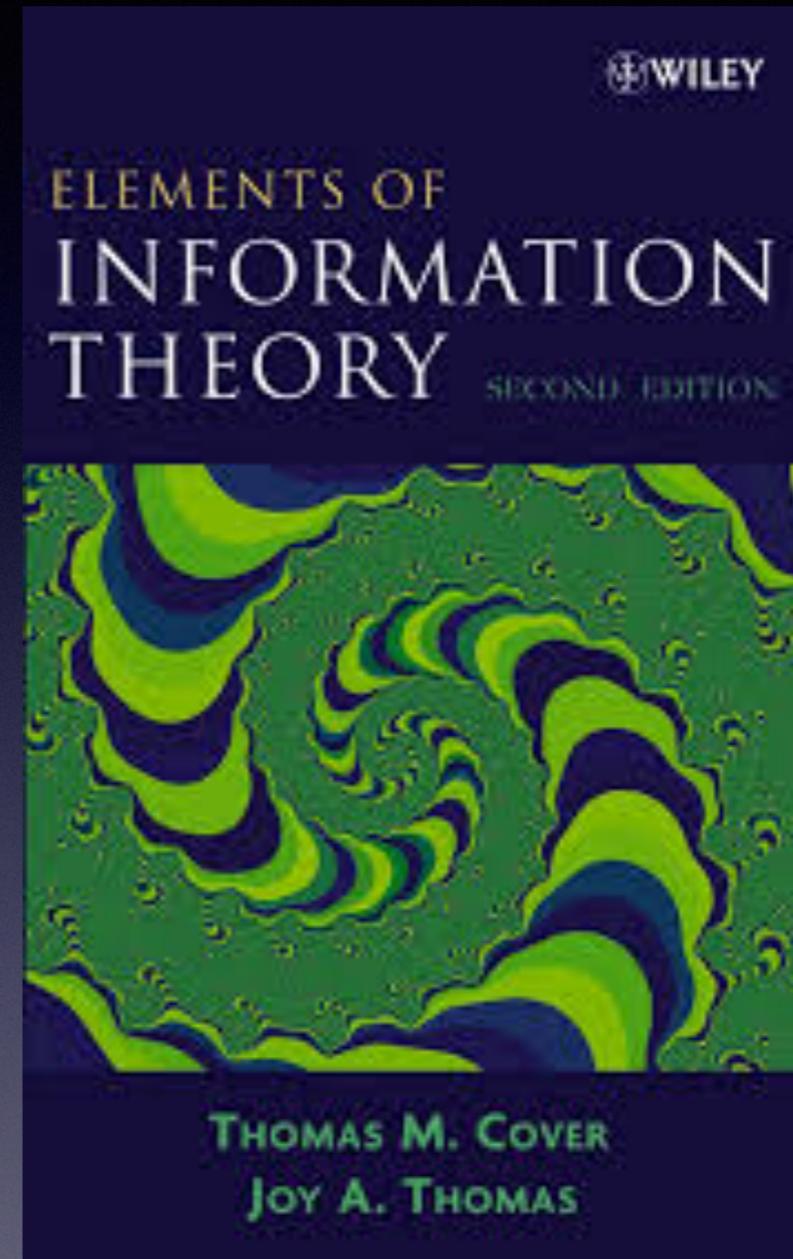
Libro di testo

- Francesco Fabris
*Teoria dell'informazione,
codici, cifrari*
Bollati Boringhieri,
2001



Ulteriori testi utili

- Thomas M. Cover,
Joy A. Thomas,
*Elements of Information
Theory*
Wiley, 1991
- Altri testi di
riferimento sono
indicati sulla scheda
GOMP del corso



Prerequisiti

- **Necessario:**
 - Concetti elementari di probabilità discreta (distribuzione di probabilità, varianza, valore atteso) [CP210]
- **Consigliato:**
 - Ulteriori concetti di probabilità (condizionamento, legge dei grandi numeri) [CP410]
- **Utile, ma del tutto opzionale:**
 - Qualche rudimento di statistica (stimatori, stima a massima verosimiglianza) [ST410]

Modalità di esame

- Esame orale (tipicamente alla lavagna)

Introduzione alla Teoria dell'Informazione

Una telefonata del 25 Aprile 1945

- Winston Churchill chiama Harry S. Truman per discutere l'offerta di resa di Heinrich Himmler
- La conversazione avviene tramite il sistema di cifratura vocale SIGSALY
- *Il sistema è veramente al sicuro da intercettazioni?*



II SIGSALY



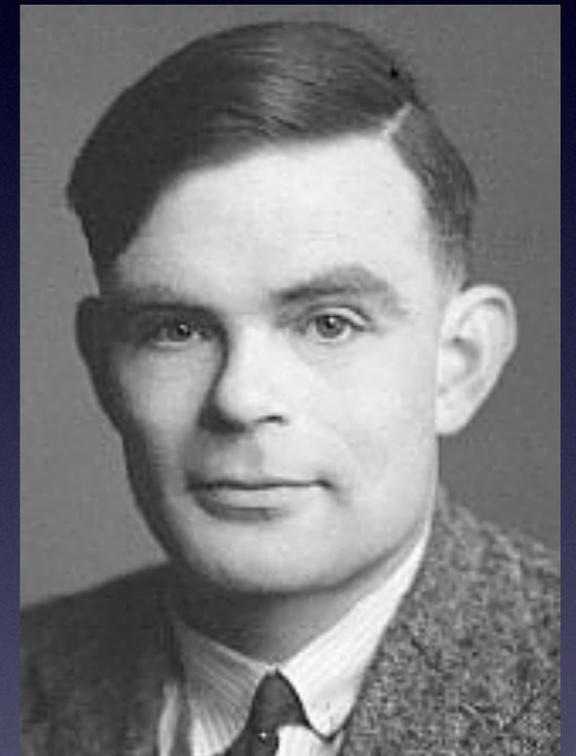
II SIGSALY



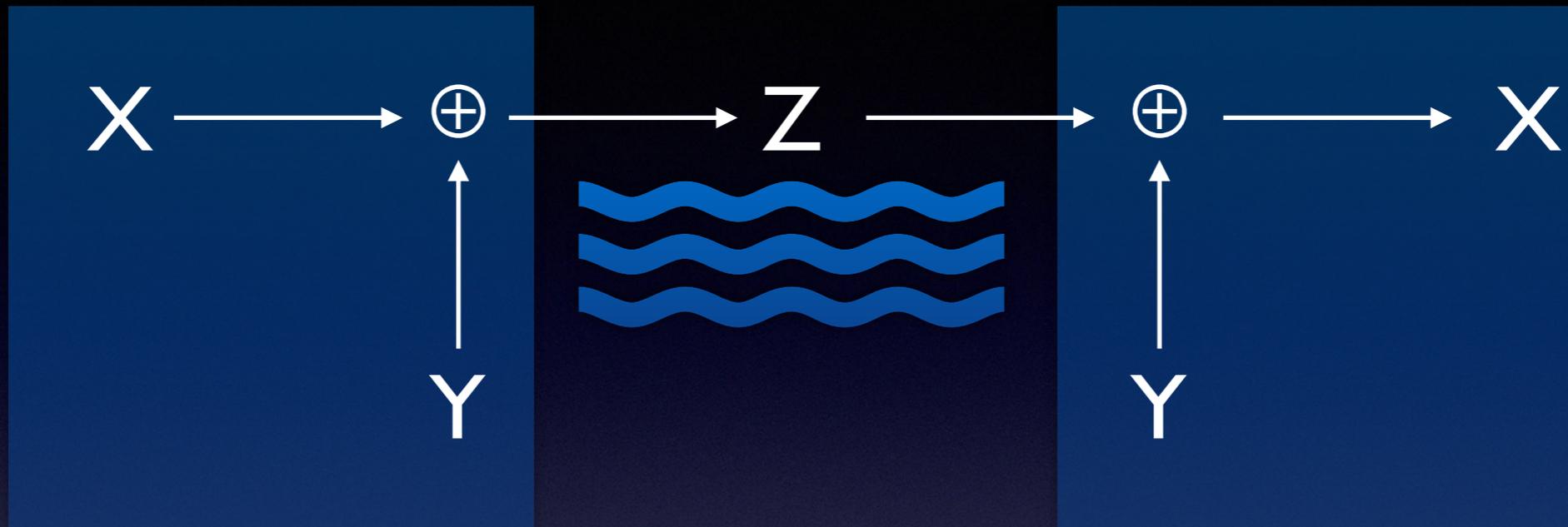
Sviluppo del SIGSALY ai Bell Telephone Labs

- A. B. Clark (supervisione)
- Claude Shannon
- Harry Nyquist
- Alan Turing (in visita)

e vari altri ingegneri e matematici



Schema astratto del SIGSALY

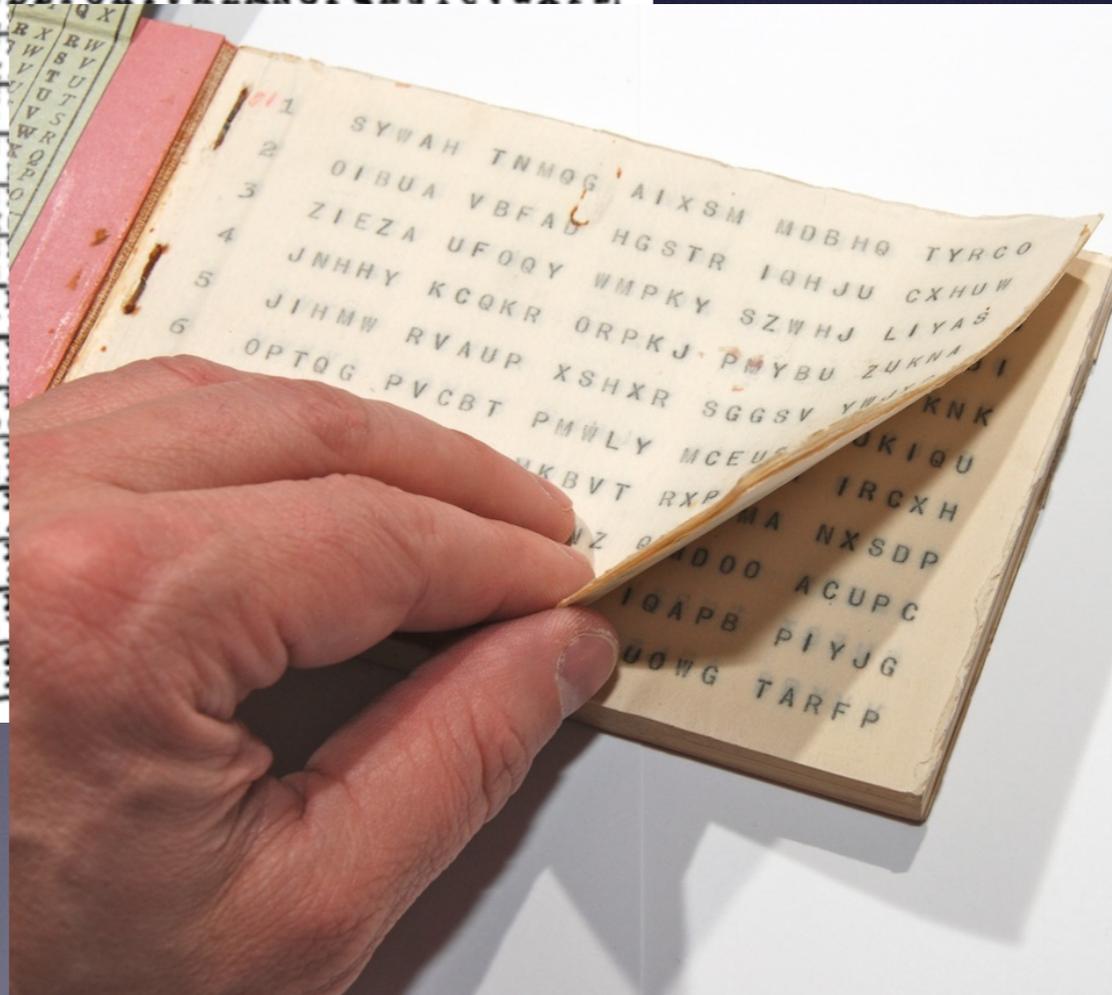


- X : un bit di informazione da trasmettere
- Y : un bit di rumore, generato casualmente
- $Z = X \oplus Y$: il bit trasmesso sul canale di telecomunicazione
- Se Y è una variabile aleatoria distribuita uniformemente su $\{0, 1\}$, Z non fornisce *nessuna informazione* su X

Cifrario monouso (One-Time Pad)

LFHNY ZAHBB JRNXX BYMFW K0ZAT
 VRETH JPCBU RUSYQ JUKMN ELQEL
 PODYF JJLVJ XFEKL HPLGA ZXVZY
 TSUIO XBNKI NBSND HPNPI OZV0Z
 EYJWF OBKKR PNTVY YTK&K ATOPR
 NHCJK FPNBV BRZZN QQZYN CYSDB
 YIIUJ TURRZ QHRDE YOVRJ HOCBY
 HALOK NHIIM CAIDY RDTEH ZDZMP
 OINDS CHDFE XGBVJ CAYSO ISBHU
 KLSZX OZJIM DBRCY BNUVZ LFRKT
 TIKTI BWIFH IHNSF RUVVC UITRM
 NQQNG ZUBZB EPVJI NCZZY FBTEX
 VEIOE HDVTN GSSNG LRZVG UKUGK
 POFRI QCFAA NLTKE DXMOA QAIMU
 HEINQ LDTWP NVBNX MNUUK ACPKA
 AYGFB ZNF0U SYNVX IYIPD RJCEK
 PROPO JFBIO NYLIX GETNC Q0XXH
 FSGNA UDTLB UNKAN HARKG TZVXH
 UGBOA JXMFY HTUNH WCTXH OFLSY

A	ABCDEFGHIJKLMN OPQRSTUVWXYZ
B	ZYXWVUTSRQPONMLKJIHGFEDCBA
C	ABCDEFGHIJKLMN OPQRSTUVWXYZ
D	XWVUTSRQPONMLKJIHGFEDCBAZY
E	ABCDEFGHIJKLMN OPQRSTUVWXYZ
F	WVUTSRQPONMLKJIHGFEDCBAZYX
G	ABCDEFGHIJKLMN OPQRSTUVWXYZ
H	VUTSRQPONMLKJIHGFEDCBAZYXW
I	ABCDEFGHIJKLMN OPQRSTUVWXYZ
J	UTSRQPONMLKJIHGFEDCBAZYXWV
K	ABCDEFGHIJKLMN OPQRSTUVWXYZ
L	TSRQPONMLKJIHGFEDCBAZYXWVU
M	ABCDEFGHIJKLMN OPQRSTUVWXYZ
N	SRQPONMLKJIHGFEDCBAZYXWVUT
O	ABCDEFGHIJKLMN OPQRSTUVWXYZ
P	RQPONMLKJIHGFEDCBAZYXWVUTS
Q	ABCDEFGHIJKLMN OPQRSTUVWXYZ
R	QPONMLKJIHGFEDCBAZYXWVUTSR
S	ABCDEFGHIJKLMN OPQRSTUVWXYZ
T	PONMLKJIHGFEDCBAZYXWVUTSRQ
U	ABCDEFGHIJKLMN OPQRSTUVWXYZ
V	ONMLKJIHGFEDCBAZYXWVUTSRQP
W	ABCDEFGHIJKLMN OPQRSTUVWXYZ
X	NMLKJIHGFEDCBAZYXWVUTSRQP
Y	ABCDEFGHIJKLMN OPQRSTUVWXYZ
Z	MLKJIHGFEDCBAZYXWVUTSRQP



Gli articoli di Shannon (1945, 1948)

- *A Mathematical Theory of Cryptography* (1945), memorandum segreto per i Bell Telephone Labs
- *A Mathematical Theory of Communication* (1948), Bell System Technical Journal

Sono considerati i contributi fondamentali alla Teoria dell'Informazione

Cosa studia la Teoria dell'Informazione

Studio della

- misurazione,
- trasmissione,
- preservazione

dell'informazione

Non si focalizza sull'*elaborazione* dell'informazione

Ambiti rilevanti

- Memorizzazione dell'informazione (es.: CD/DVD)
- Comunicazioni remote (es.: telefonia, satelliti)
- Meccanica statistica
- Inferenza statistica
- Linguistica e semiotica

TI è propedeutica alla *Crittografia* e al *Machine Learning*

Ambiti rilevanti

- Memorizzazione dell'informazione (es.: CD/DVD)
- Comunicazioni remote (es.: telefonia, satelliti)
- Meccanica statistica
- Inferenza statistica
- Linguistica e semiotica



TI è propedeutica alla *Crittografia* e al *Machine Learning*

Ambiti rilevanti

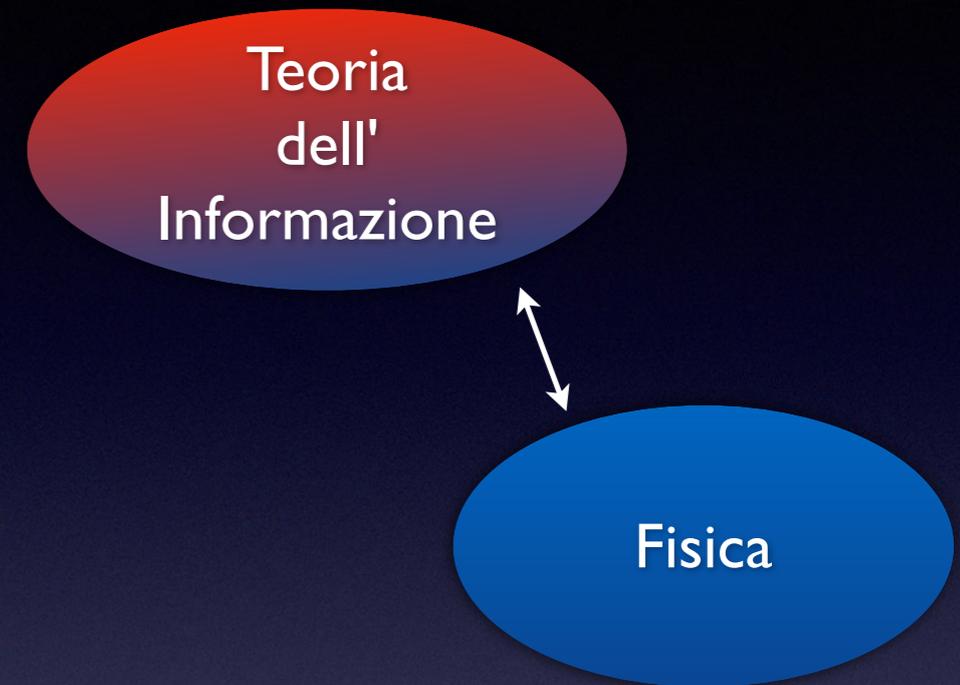
- Memorizzazione dell'informazione (es.: CD/DVD)
- Comunicazioni remote (es.: telefonia, satelliti)
- Meccanica statistica
- Inferenza statistica
- Linguistica e semiotica



TI è propedeutica alla *Crittografia* e al *Machine Learning*

Ambiti rilevanti

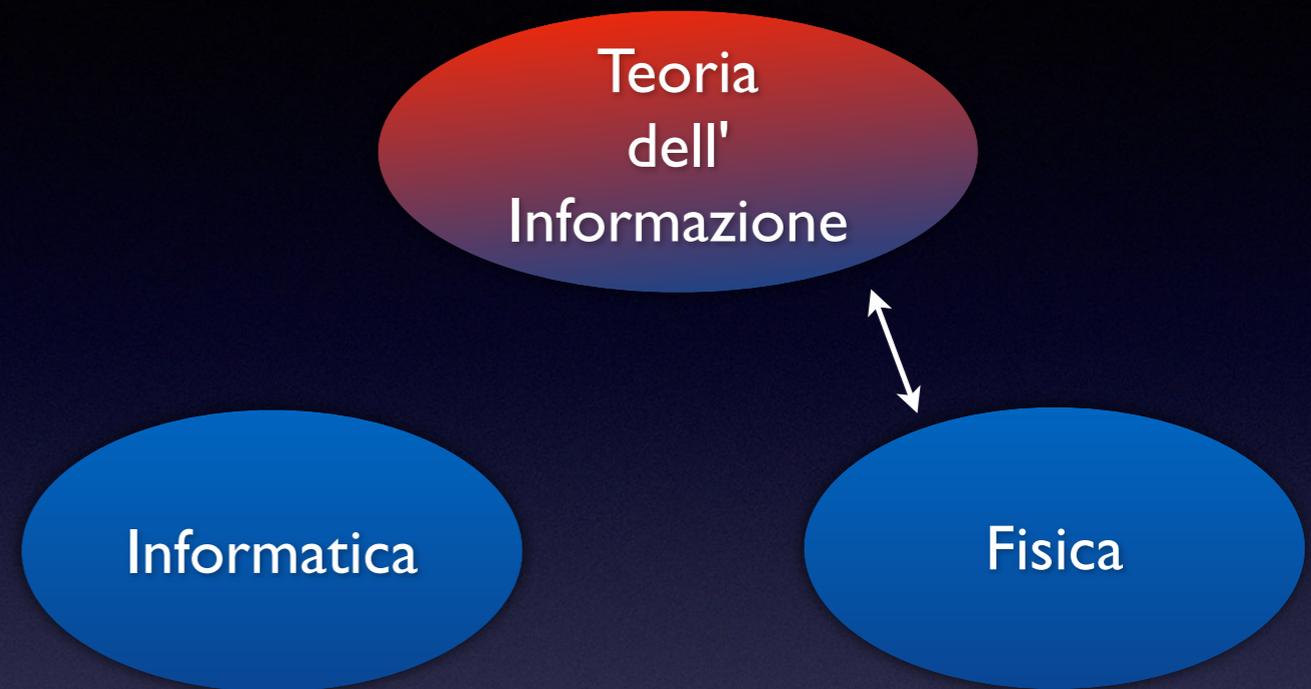
- Memorizzazione dell'informazione (es.: CD/DVD)
- Comunicazioni remote (es.: telefonia, satelliti)
- Meccanica statistica
- Inferenza statistica
- Linguistica e semiotica



TI è propedeutica alla *Crittografia* e al *Machine Learning*

Ambiti rilevanti

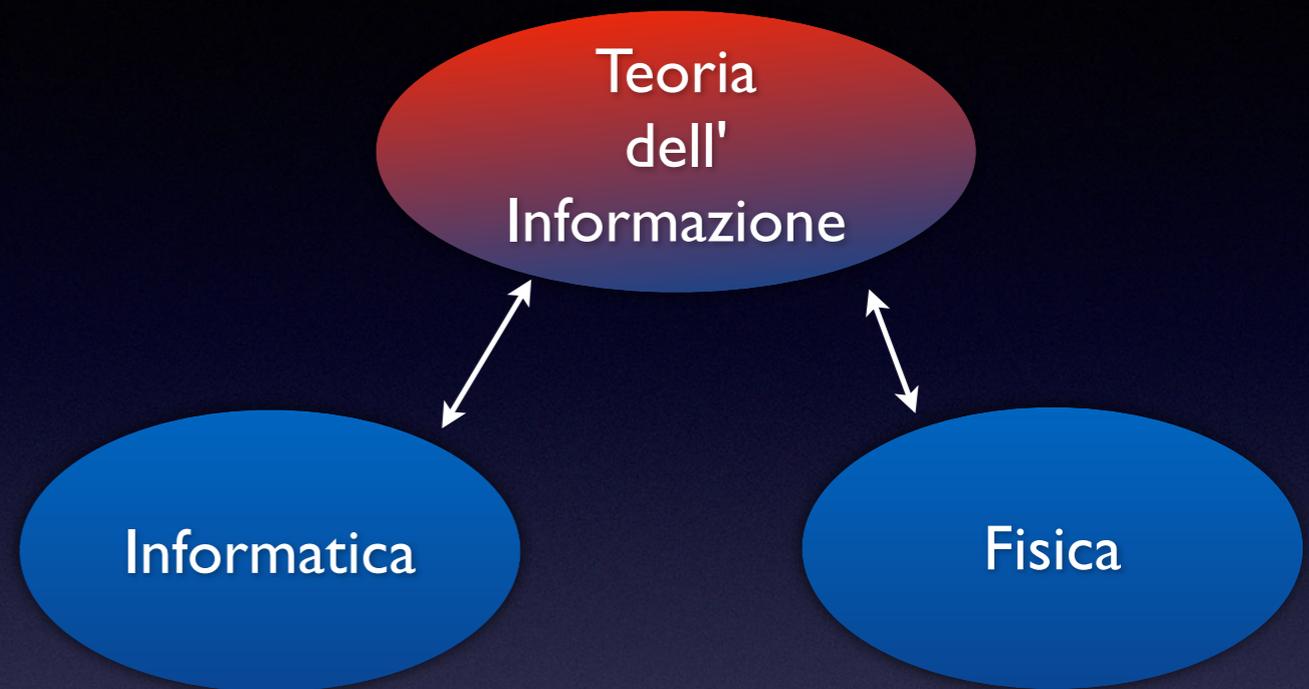
- Memorizzazione dell'informazione (es.: CD/DVD)
- Comunicazioni remote (es.: telefonia, satelliti)
- Meccanica statistica
- Inferenza statistica
- Linguistica e semiotica



TI è propedeutica alla *Crittografia* e al *Machine Learning*

Ambiti rilevanti

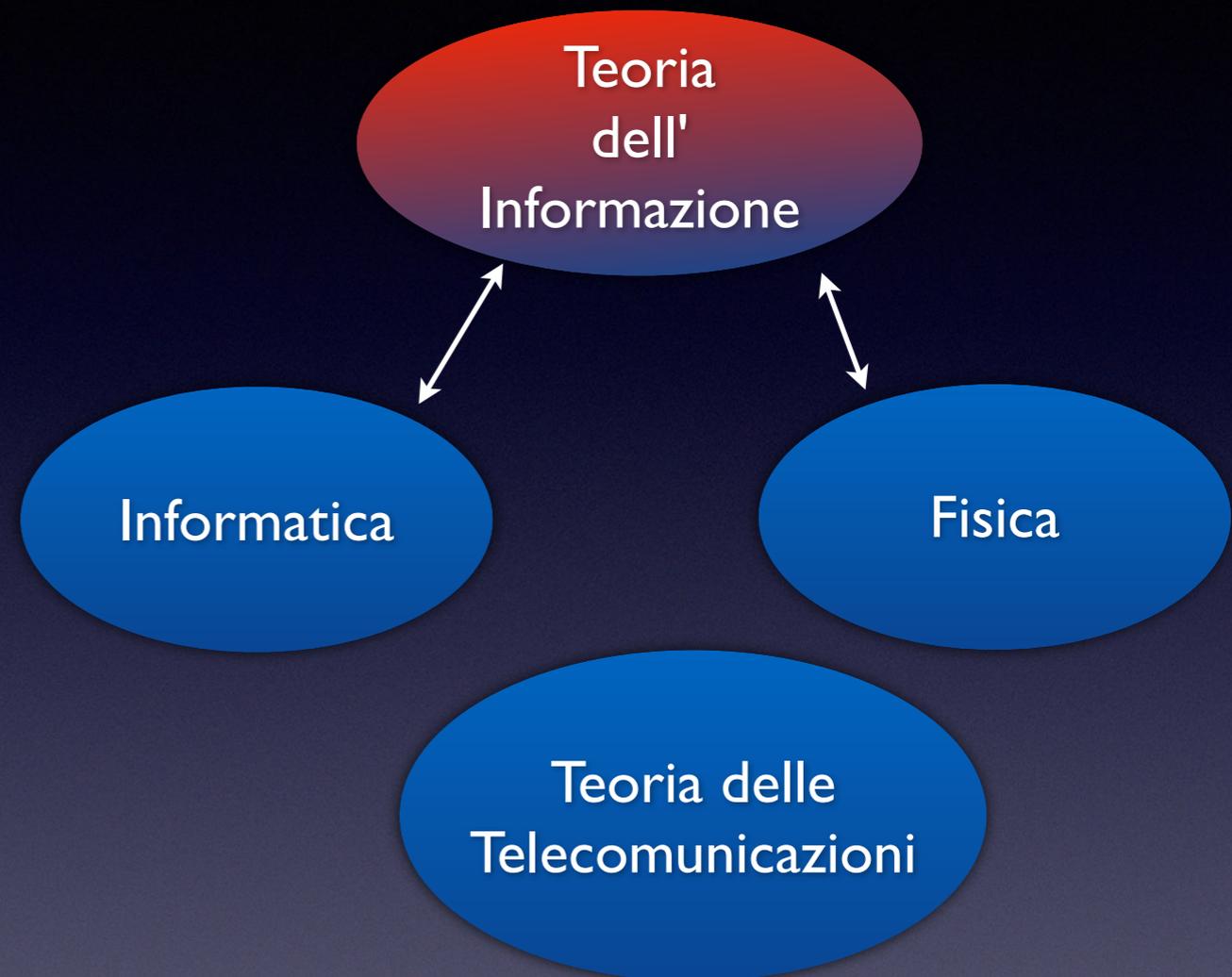
- Memorizzazione dell'informazione (es.: CD/DVD)
- Comunicazioni remote (es.: telefonia, satelliti)
- Meccanica statistica
- Inferenza statistica
- Linguistica e semiotica



TI è propedeutica alla *Crittografia* e al *Machine Learning*

Ambiti rilevanti

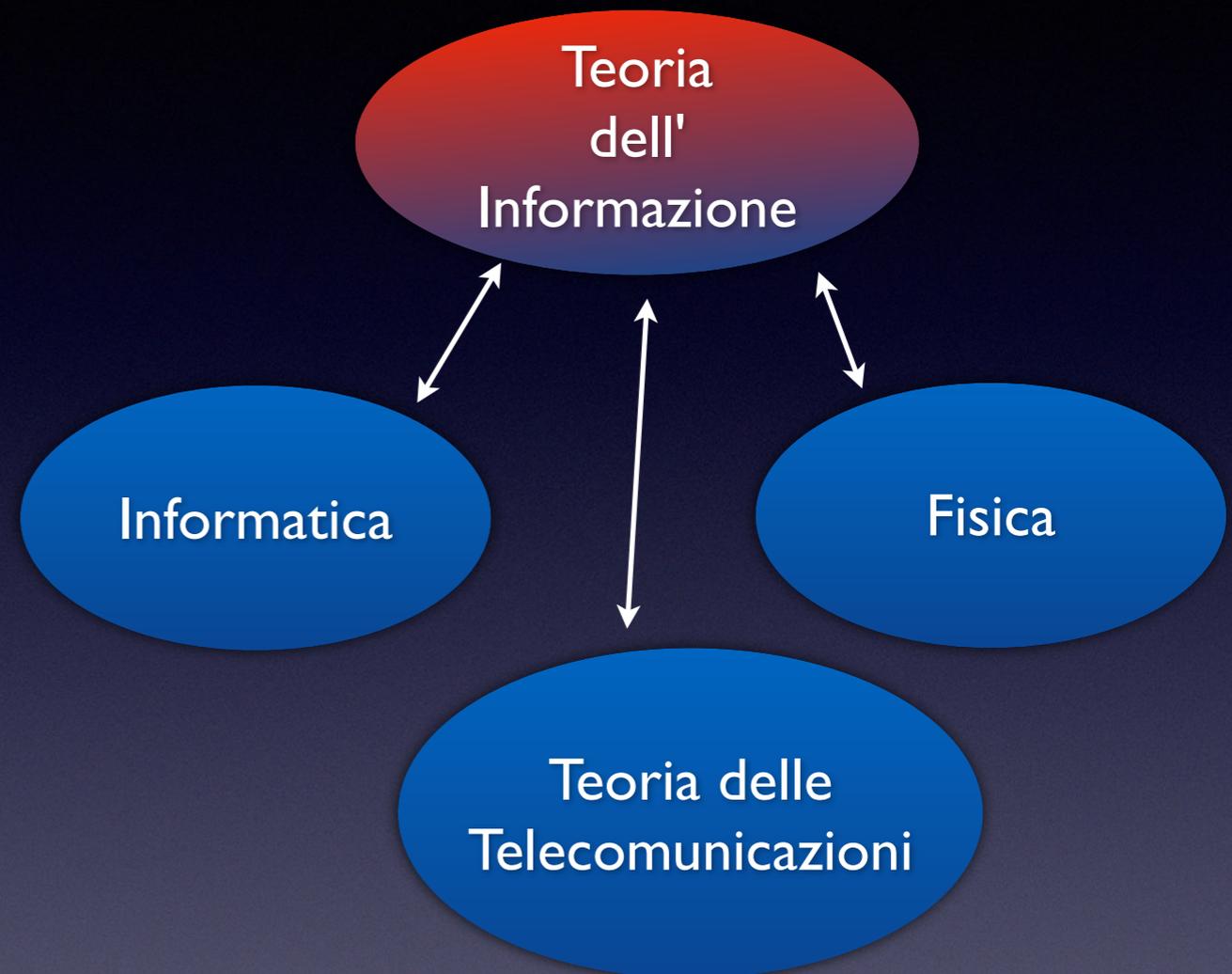
- Memorizzazione dell'informazione (es.: CD/DVD)
- Comunicazioni remote (es.: telefonia, satelliti)
- Meccanica statistica
- Inferenza statistica
- Linguistica e semiotica



TI è propedeutica alla *Crittografia* e al *Machine Learning*

Ambiti rilevanti

- Memorizzazione dell'informazione (es.: CD/DVD)
- Comunicazioni remote (es.: telefonia, satelliti)
- Meccanica statistica
- Inferenza statistica
- Linguistica e semiotica



TI è propedeutica alla *Crittografia* e al *Machine Learning*

TI e meccanica statistica

- L'*entropia*, concetto fondamentale della TI, è ispirato dalla meccanica statistica
- Entropia di Boltzmann–Gibbs:

$$S = -k_B \sum_i p_i \log p_i$$

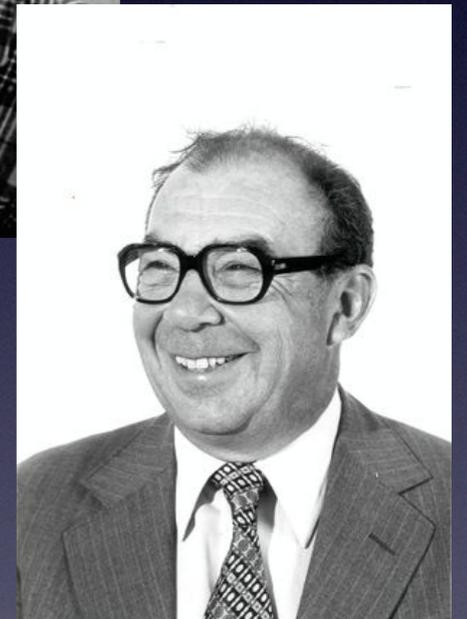
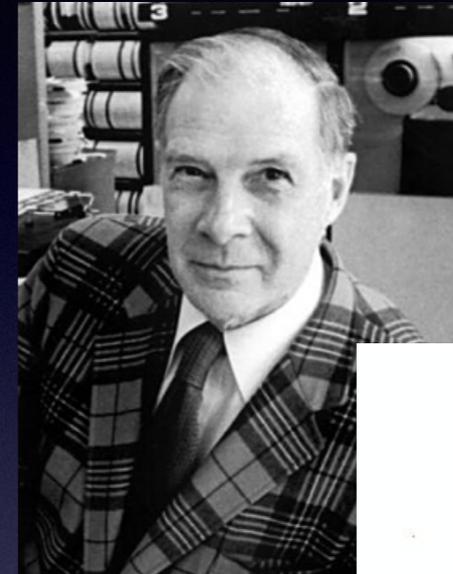
- Entropia di Shannon:

$$H = -\sum_i p_i \log_2 p_i$$

«La mia più grande preoccupazione era come chiamarla. Pensavo di chiamarla *informazione*, ma la parola era fin troppo usata, così decisi di chiamarla *incertezza*. Quando discussi della cosa con John Von Neumann, lui ebbe un'idea migliore. Mi disse che avrei dovuto chiamarla *entropia*, per due motivi: "Innanzitutto, la tua funzione d'incertezza è già nota nella meccanica statistica con quel nome. In secondo luogo, e più significativamente, nessuno sa cosa sia con certezza l'*entropia*, così in una discussione sarai sempre in vantaggio"» –Claude Shannon

La TI dopo Shannon

- Richard Hamming (1947)
- Roberto M. Fano (1949)
- David A. Huffman (1951)
- Irving Reed, Gustave Solomon (1960)
- Jacob Ziv, Abraham Lempel (1977)
- Teoria dei codici a correzione d'errore
 - Applicazioni: CD, sonde spaziali, ...
- Teoria della compressione dati
 - Applicazioni: formati file .zip, .bzip, .jpg, .mp3, ...



Programma sintetico

1. Introduzione alla teoria dell'informazione

Trasmissione affidabile dell'informazione. Contenuto informativo secondo Shannon. Misure di informazione. Entropia, mutua informazione, divergenza informazionale. Compressione dati. Correzione d'errore. Teoremi di elaborazione dei dati. Disuguaglianze fondamentali. Diagrammi d'informazione. Divergenza informazionale e massima verosimiglianza.

2. Codifica di sorgente e compressione dati

Sequenze tipiche. Tipicità in probabilità. Proprietà di equipartizione asintotica. Codifica a blocco e a lunghezza variabile. Tasso di codifica. Teorema della codifica di sorgente. Compressione dati senza perdita. Codice di Huffman. Codici universali. Compressione Ziv-Lempel.

3. Codifica di canale

Capacità di canale. Canali discreti senza memoria. Informazione trasportata da un canale. Criteri di decodifica. Teorema della codifica di canale con rumore.

4. Codici correttori ed applicazioni

Spazio di Hamming. Codici lineari. Matrice generatrice e matrice di controllo. Codici di Hamming. Codici hash.

Trasmissione dell'informazione



- *Sorgente e destinazione* della trasmissione
- *Canale* di trasmissione (affetto da possibili errori)
- *Codifica di sorgente*: elimina ridondanza nei dati (per rendere più *efficiente* la trasmissione)
- *Codifica di canale*: aggiunge ridondanza utile a proteggere l'informazione da errori di trasmissione (per rendere più *affidabile* la trasmissione)

Trasmissione dell'informazione



- *Sorgente e destinazione* della trasmissione
- *Canale di trasmissione* (affetto da possibili errori)
- *Codifica di sorgente*: elimina ridondanza nei dati (per rendere più *efficiente* la trasmissione)
- *Codifica di canale*: aggiunge ridondanza utile a proteggere l'informazione da errori di trasmissione (per rendere più *affidabile* la trasmissione)

Trasmissione dell'informazione



- *Sorgente e destinazione* della trasmissione
- *Canale* di trasmissione (affetto da possibili errori)
- *Codifica di sorgente*: elimina ridondanza nei dati (per rendere più *efficiente* la trasmissione)
- *Codifica di canale*: aggiunge ridondanza utile a proteggere l'informazione da errori di trasmissione (per rendere più *affidabile* la trasmissione)

Trasmissione dell'informazione



- *Sorgente e destinazione* della trasmissione
- *Canale* di trasmissione (affetto da possibili errori)
- *Codifica di sorgente*: elimina ridondanza nei dati (per rendere più *efficiente* la trasmissione)
- *Codifica di canale*: aggiunge ridondanza utile a proteggere l'informazione da errori di trasmissione (per rendere più *affidabile* la trasmissione)

Trasmissione dell'informazione



- *Sorgente e destinazione* della trasmissione
- *Canale* di trasmissione (affetto da possibili errori)
- *Codifica di sorgente*: elimina ridondanza nei dati (per rendere più *efficiente* la trasmissione)
- *Codifica di canale*: aggiunge ridondanza utile a proteggere l'informazione da errori di trasmissione (per rendere più *affidabile* la trasmissione)

Un esempio di codifica di sorgente

Run-Length Encoding: (es.: usata nelle trasmissioni fax)

- Rimpiazza ripetizioni dello stesso simbolo con la coppia *(lunghezza della ripetizione, simbolo)*
- Esempio: N = pixel nero, B = pixel bianco
- NNNNNBNNNNNNNBBBNN
diventa
5N1B6N3B2N

Nota: non sempre la lunghezza del messaggio è ridotta:

- NBNBNBNBNB
diventa
1N1B1N1B1N1B1N1B1N1B

Un esempio di codifica di canale

Codice binario a ripetizione di lunghezza 3:

- **Codifica:** rimpiazza 0 con 000, 1 con 111
- **Decodifica:** restituisci il simbolo che appare *più volte* in ogni terzetto ricevuto
- **Esempio:** 100 è decodificato come 0, 101 come 1
- Riesce a *correggere* fino ad 1 errore ogni 3 bit trasmessi
- Riesce a *rilevare* fino a 2 errori ogni 3 bit trasmessi