

# L'informazione e la sua codifica

**Informatica e Informazione**

**La codifica dell'informazione**

**Analogico vs digitale**

# Informazione e Informatica

# Informatica e telecomunicazione

## ➤ Cos'è l'informatica?

- lo studio sistematico degli algoritmi che descrivono e trasformano l'informazione: la loro teoria, analisi, progetto, efficienza, realizzazione e applicazione [ACM - Association for Computing Machinery]
- la scienza della rappresentazione e dell'elaborazione dell'informazione

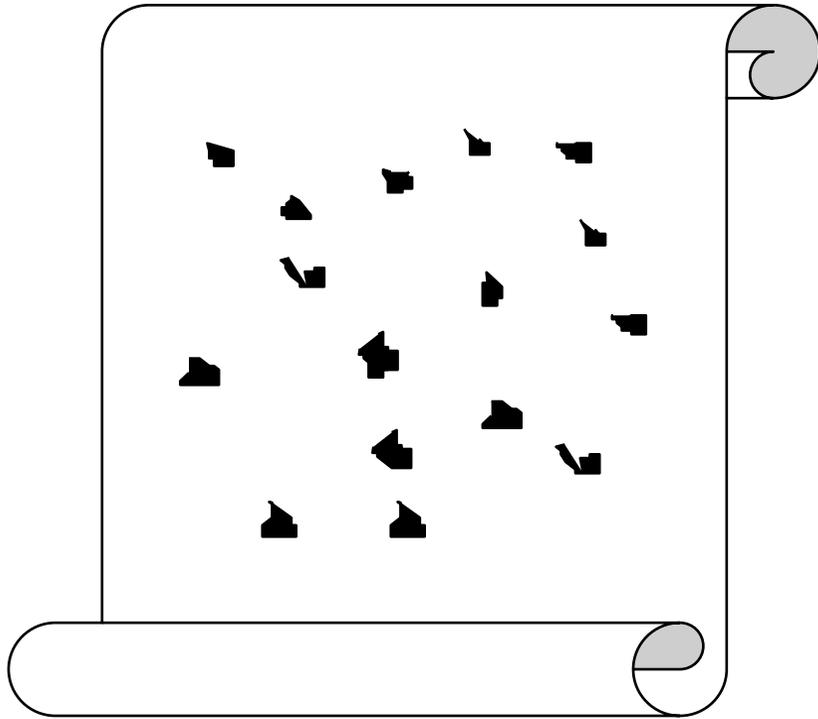
## ➤ Cos'è la telecomunicazione?

- la trasmissione **rapida** a **distanza** dell'informazione

## ➤ Attenzione:

- Non si parla di tecnologia dei calcolatori !
- Si attribuisce ruolo centrale al concetto di informazione !

# Il concetto di informazione



Configurazione 1



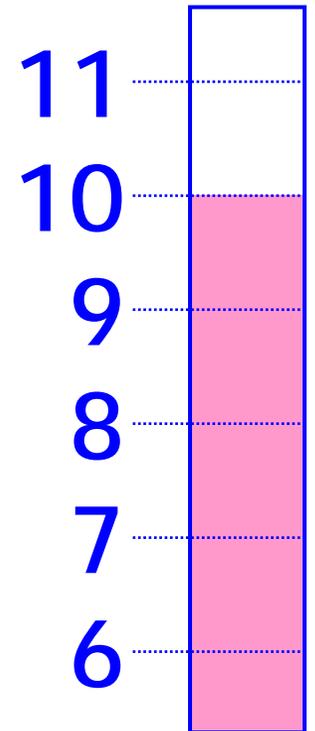
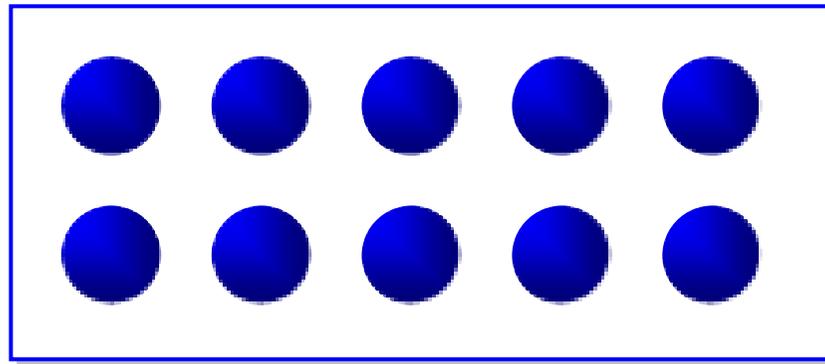
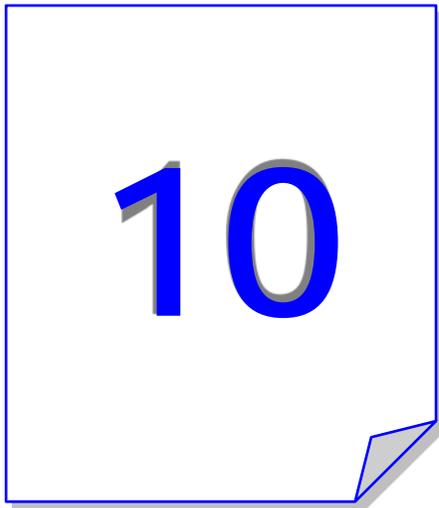
Configurazione 2

# Informazione e supporto

- L'informazione è “portata da”, o “trasmessa su”, o “memorizzata in”, o “contenuta in” qualcosa; questo “qualcosa” però non è l'informazione stessa.
- Ogni supporto ha le sue caratteristiche in quanto mezzo su cui può essere scritta dell'informazione.

# Informazione e supporti (1)

La stessa informazione può essere scritta su **supporti differenti**.



# Informazione e supporti (2)

Lo **stesso supporto** può portare **informazioni differenti**.



# Informazione e supporto (3)

- Distinguere informazione e supporto fisico è distinguere tra “**entità logiche**” ed “**entità fisiche**”:
  - l'informazione **richiede un supporto fisico**, ma non coincide con esso;
  - l'informazione è un'entità **extra-fisica**, non interpretabile in termini di materia-energia e sottoposta alle leggi della fisica solo perché basata su un supporto fisico.
- L'informazione si può **creare** e **distruggere**.

## Quali caratteristiche deve avere un sistema fisico per supportare informazioni?

- Si ottiene informazione quando, dato un insieme di alternative possibili, la lettura del supporto ne elimina alcune e ne seleziona altre.
- **Condizione necessaria** perché un supporto possa portare informazione è che possa assumere **configurazioni differenti**, a ognuna delle quali venga associata una differente **entità di informazione**.

# Supporto fisico: 1<sup>a</sup> condizione



- Deve consentire di potere identificare delle differenze
  - Es: voglio rappresentare 2 alternative



- Cosa **rappresenta** ciascuna configurazione?

# Configurazioni e codici

- A ogni configurazione del supporto deve essere associata un'entità di informazione:
  - Prima Configurazione = interruttore "ON" = "Divina Commedia";
  - Seconda Configurazione = interruttore "OFF" = "I Promessi Sposi".
- Per interpretare le differenti configurazioni del supporto in termini di informazione è necessario conoscere il codice (cioè la regola) che a ogni configurazione ammessa del supporto associa un'entità di informazione.
- La definizione di un codice comporta che siano identificati in modo non ambiguo l'insieme delle possibili configurazioni del supporto e l'insieme delle possibili entità di informazione a cui ci si vuole riferire.
- Variando il codice è possibile riferirsi a entità di informazione differenti utilizzando uno stesso supporto fisico.

# Supporto fisico: 2a condizione

Deve essere **condivisa** una regola per attribuire un **significato** a ciascuna configurazione

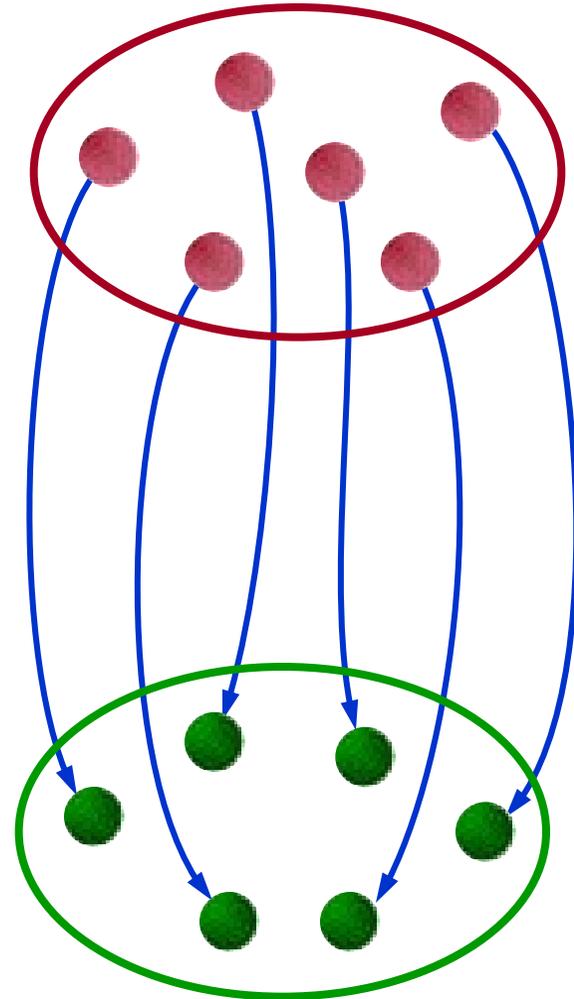


# Definire un codice

## ➤ Identificare

- { Configurazioni }
- { Entità informazione }

## ➤ Associare gli elementi dei 2 insiemi



# Livelli di informazione: il telegrafo

- **Supporto fisico**: conduttore in cui transita corrente continua
- **Entità di informazione**: “punti” e “linee”, che rappresentano le lettere dell’alfabeto
- **Codice** = regola che specifica la corrispondenza:
  - passa corrente per 1 s  $\Rightarrow$  punto
  - passa corrente per 2 s  $\Rightarrow$  linea
  - non passa corrente per 1 s  $\Rightarrow$  separazione tra punti e linee della stessa lettera
  - non passa corrente per 2 s  $\Rightarrow$  separazione tra due lettere
  
  - “passa corrente per 1 s, non passa per 1 s, passa per 2 s” = “punto linea”.
- **Codice Morse** = relazione tra lettere e sequenze di punti linee:
  - punto linea  $\Rightarrow$  lettera “A”
  - linea punto punto punto  $\Rightarrow$  lettera “B”
  - linea punto linea punto  $\Rightarrow$  lettera “C”
  - ecc.
- **Abbiamo associato simboli ad altri simboli!!**

# Messaggi vs configurazione del supporto

- **Configurazione del supporto fisico (livello fisico, LF):**
  - passa corrente per 2s, non passa per 1s, passa per 1s, non passa per 2s, passa per 2s, non passa per 1s, passa per 2s, non passa per 1s, passa per 2s
- **messaggio, al livello logico 1 (LL1):**
  - linea punto, separazione di lettera, linea linea linea
- **messaggio, al livello Logico 2 (LL2):**
  - "N" "O"
- **messaggio, al livello Logico 3 (LL3):**
  - "NO", con il significato (per esempio) di "incontro non confermato"
- **Le relazioni tra questi livelli definiscono ognuna un codice per interpretare il significato delle entità che compaiono al livello precedente in termini delle entità del livello successivo: sono relazioni di significazione.**

# Tre diversi livelli di informazione

- Informazione **sintattica**
- Informazione **semantica**
- Informazione **pragmatica**

# Teoria dell'informazione

- Quando si parla di “teoria dell'informazione” si fa riferimento al solo livello “sintattico”
- **Ambito di applicazione:** caratterizzare le condizioni per la trasmissione di segnali in termini di
  - **adeguatezza** del supporto adottato per la trasmissione
  - **accuratezza** della trasmissione stessa.
- **Quali problemi si pone:**
  - un certo supporto può essere utilizzato per la memorizzazione di una certa quantità di informazione?
  - con quale velocità una certa quantità di informazione può essere trasferita a distanza mediante un certo supporto?
  - con quale grado di accuratezza un certo messaggio è stato trasmesso?

# Informazione e incertezza

- La presenza di informazione è condizionata dal fatto che il supporto sia in grado di assumere **diverse configurazioni**.
- Se la nostra ignoranza, o più formalmente la nostra incertezza, circa l'effettiva configurazione del supporto viene ridotta dall'accesso al supporto, allora sembra del tutto ragionevole assumere che si possa affermare che tale atto ci ha **portato dell'informazione**.
- Se fossimo in grado di misurare il **grado di incertezza** in cui ci trovavamo **prima della lettura e quello successivo a essa**, la **quantità di informazione** portata dalla configurazione che abbiamo letto sul supporto potrebbe essere definita proprio dalla **differenza tra tali gradi di incertezza**.

# La codifica dell'informazione

# Codifica dati e istruzioni

## ➤ Algoritmo

- **descrizione** della **soluzione di problema** scritta in modo da poter essere eseguita da un **esecutore** (eventualmente diverso dall'autore dell'algoritmo)
- sequenza di **istruzioni** che operano su **dati**.

## ➤ Programma

- **algoritmo** scritto in modo da poter essere eseguito da un **calcolatore** (esecutore automatico)

- ## ➤ Per scrivere un **programma** è necessario rappresentare **istruzioni** e **dati** in un formato tale che **l'esecutore automatico** sia capace di **memorizzare** e **manipolare**.

# Codifica dati e istruzioni

## ➤ Alfabeto dei simboli

- cifre "0", "1", ..., "9", separatore decimale ("."), separatore delle migliaia (",") e segni positivo ("+") o negativo ("-").

## ➤ Regole di composizione (sintassi), che definiscono le successioni "ben formate"

- "1.234,5" è la rappresentazione di un numero;
- "1,23,45" non lo è.

## ➤ Codice (semantica)

- "1.234,5" =  $1 \times 10^3 + 2 \times 10^2 + 3 \times 10^1 + 4 \times 10^0 + 5 \times 10^{-1}$
- "1,23,45" = ??

## ➤ Lo stesso alfabeto può essere utilizzato con codici diversi:

- "123,456" =  $1 \times 10^2 + 2 \times 10^1 + 3 \times 10^0 + 4 \times 10^{-1} + 5 \times 10^{-2} + 6 \times 10^{-3}$ , [IT]
- "123,456" =  $1 \times 10^5 + 2 \times 10^4 + 3 \times 10^3 + 4 \times 10^2 + 5 \times 10^1 + 6 \times 10^0$ , [UK]

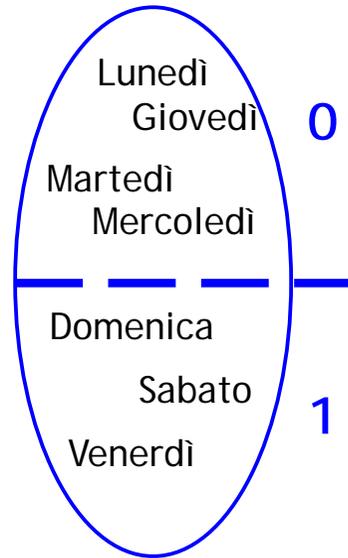
# Codifica Binaria

- **Alfabeto binario**: usiamo dispositivi con solo due stati
- **Problema**: assegnare un **codice univoco** a tutti gli oggetti compresi in un insieme predefinito (e.g. studenti)
- **Quanti oggetti** posso codificare con **k bit**:
  - 1 bit  $\Rightarrow$  2 stati (0, 1)  $\Rightarrow$  2 oggetti (e.g. Vero/Falso)
  - 2 bit  $\Rightarrow$  4 stati (00, 01, 10, 11)  $\Rightarrow$  4 oggetti
  - 3 bit  $\Rightarrow$  8 stati (000, 001, ..., 111)  $\Rightarrow$  8 oggetti
  - ...
  - **k bit  $\Rightarrow$   $2^k$  stati  $\Rightarrow$   $2^k$  oggetti**
- **Quanti bit** mi servono per codificare **N oggetti**:
  - $N \leq 2^k \Rightarrow k \geq \log_2 N \Rightarrow$   **$k = \lceil \log_2 N \rceil$**  (intero superiore)
- **Attenzione**:  
ipotesi implicita che i codici abbiano tutti la **stessa lunghezza**

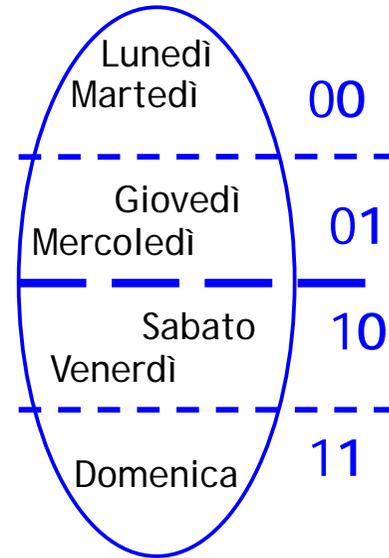
# Esempio di codifica binaria

- **Problema:**  
assegnare un codice binario univoco a tutti i giorni della settimana
- **Giorni della settimana:**  $N = 7 \Rightarrow k \geq \log_2 7 \Rightarrow k = 3$
- **Con 3 bit possiamo ottenere 8 diverse configurazioni:**
  - Ne servono 7, quali utilizziamo?
  - Quale configurazione associamo a quale giorno?
- **Attenzione:**  
ipotesi che i codici abbiano tutti la stessa lunghezza

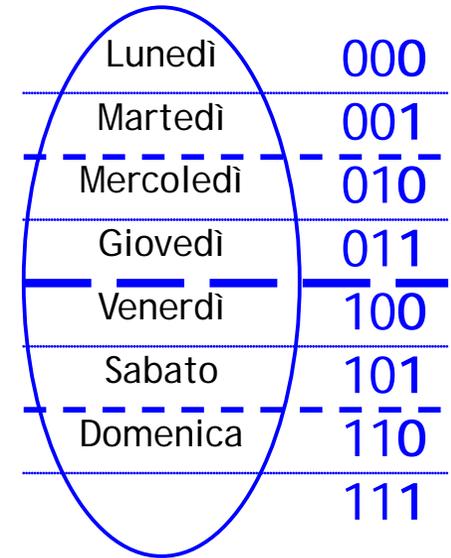
# I giorni della settimana in binario (1)



**1 bit**  
**2 "gruppi"**

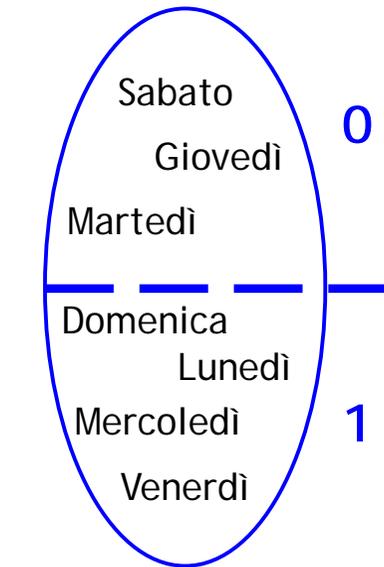


**2 bit**  
**4 "gruppi"**

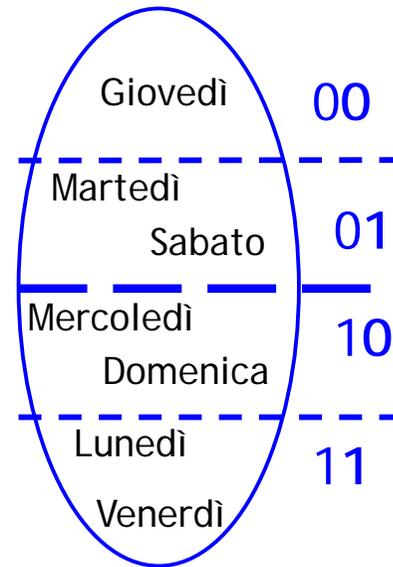


**3 bit**  
**8 "gruppi"**

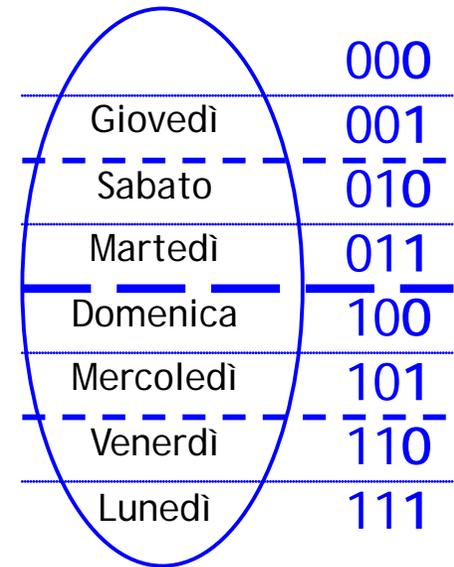
# I giorni della settimana in binario (2)



**1 bit**  
**2 "gruppi"**



**2 bit**  
**4 "gruppi"**



**3 bit**  
**8 "gruppi"**

# Codifica binaria dei caratteri

## ➤ Quanti sono gli oggetti compresi nell'insieme?

- 26 lettere maiuscole + 26 minuscole  $\Rightarrow$  52
- 10 cifre
- Circa 30 segni d'interpunzione
- Circa 30 caratteri di controllo (EOF, CR, LF, ...)

**circa 120 oggetti complessivi  $\Rightarrow k = \lceil \log_2 120 \rceil = 7$**

## ➤ Codice ASCII: utilizza 7 bit e quindi può rappresentare al massimo $2^7=128$ caratteri

- Con 8 bit (= byte) rappresento 256 caratteri (ASCII esteso)
- Si stanno diffondendo codici più estesi (e.g. UNICODE) per rappresentare anche i caratteri delle lingue orientali

# ASCII su 7 bit

	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
010	sp	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
011	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
100	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
101	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
110	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
111	p	q	r	s	t	u	v	w	x	Y	z	{		}	~	canc

# bit, Byte, KiloByte, MegaByte, ...

**bit** = solo due stati, "0" oppure "1".

**Byte** = 8 bit, quindi  $2^8 = 256$  stati

**KiloByte [KB]** =  $2^{10}$  Byte = 1024 Byte  $\sim 10^3$  Byte

**MegaByte [MB]** =  $2^{20}$  Byte = 1'048'576 Byte  $\sim 10^6$  Byte

**GigaByte [GB]** =  $2^{30}$  Byte  $\sim 10^9$  Byte

**TeraByte [TB]** =  $2^{40}$  Byte  $\sim 10^{12}$  Byte

**PetaByte [PB]** =  $2^{50}$  Byte  $\sim 10^{15}$  Byte

**ExaByte [EB]** =  $2^{60}$  Byte  $\sim 10^{18}$  Byte

# Lo standard IEC per i prefissi binari

Grandezza	Nome	Simbolo		Dimensione	SI	Diff. %
Kilo binario	Kibi	Ki	$2^{10}$	1'024	$10^3$	2.40%
Mega binario	Mebi	Mi	$(2^{10})^2$	1'048'576	$(10^3)^2$	4.86%
Giga binario	Gibi	Gi	$(2^{10})^3$	1'073'741'824	$(10^3)^3$	7.37%
Tera binario	Tebi	Ti	$(2^{10})^4$	1'099'511'627'776	$(10^3)^4$	9.95%
Peta binario	Pebi	Pi	$(2^{10})^5$	1'125'899'906'842'624	$(10^3)^5$	12.59%
Exa binario	Exbi	Ei	$(2^{10})^6$	1'152'921'504'606'846'976	$(10^3)^6$	15.29%
Zetta binario	Zebi	Zi	$(2^{10})^7$	1'180'591'620'717'411'303'424	$(10^3)^7$	18.06%
Yotta binario	Yobi	Yi	$(2^{10})^8$	1'208'925'819'614'629'174'706'176	$(10^3)^8$	20.89%

➤ Usando questi prefissi si può risolvere l'ambiguità

- capacità di un disco fisso: 120 GB = 111.76 GiB,
- capacità di un floppy: 1.406 MiB = 1.475 MB

# La codifica delle istruzioni

- Si segue lo schema presentato per i caratteri alfanumerici:
- **quali e quante** sono le istruzioni da codificare?
  - qual è la **lunghezza** delle successioni di bit da utilizzare ?
  - qual è la **corrispondenza** tra istruzioni e successioni di bit ?

Istruzioni aritmetico-logiche	
Codice	Istruzione
0111 1100	<b>ADD</b>
0111 1101	<b>SUB</b>
0111 1110	<b>AND</b>
... ..	... ..

Istruzioni per il trasferimento dati	
Codice	Istruzione
1110 1000	<b>LOAD</b>
1111 1000	<b>STORE</b>
... ..	... ..
... ..	... ..

Istruzioni di controllo	
Codice	Istruzione
0100 1001	<b>IF_EQ</b>
0100 1000	<b>GOTO</b>
0100 1100	<b>RETURN</b>
... ..	... ..

# Oltre al codice operativo

- ... è necessario far riferimento ai **dati** necessari per completare l'esecuzione dell'istruzione,
  - e.g. addizione: è necessario che sia specificato (anche **implicitamente**) dove leggere i due operandi da sommare e dove scrivere il risultato;
- il **numero** dei dati da specificare è variabile, in funzione delle istruzioni.

Codice Operativo	Destinazione	Sorgente 1	Sorgente 2	Estensione del codice operativo
------------------	--------------	------------	------------	---------------------------------

Codice Operativo	Destinazione	Sorgente 1	Operando (immediato)
------------------	--------------	------------	----------------------

Codice Operativo	Operando (immediato)
------------------	----------------------

# Analogico vs digitale

# Informazione “classificatoria” e più che “classificatoria”

- Informazione **classificatoria**:  
“è questo, ma avrebbe potuto essere quest’altro”.
- Informazione **più che classificatoria**:
  - riconoscere distinzioni;
  - stabilire una relazione d’ordine (“questo è maggiore di quest’altro”);
  - stabilire una metrica (“questo è distante un certo valore da quest’altro”).
- L’insieme delle entità di informazione ha una **struttura**.
  - La struttura dice “cosa si può fare” con le entità di informazione dell’insieme, in termini di operazioni di combinazione e di confronto.
  - Si tratta di informazione su informazione (meta-informazione) in presenza della quale l’insieme delle entità di informazione diventa un sistema, cioè appunto un “insieme con struttura”.

# Due alternative

- **Meta-informazione *esplicita* nel supporto:**
  - il supporto ha una struttura corrispondente a quella presente tra entità di informazione.
- **Meta-informazione *implicita* nella regola di codifica:**
  - al supporto si richiede solo di avere configurazioni molteplici e distinguibili l'una dall'altra,
  - la meta-informazione è definita in modo estensionale nella regola di codifica.
- **Cosa succede nei due casi se si aggiungono i “mezzi punti”?**

# Analogico e digitale



➤ Meta-informazione esplicita nel supporto

➤ **Codifica ANALOGICA**



➤ Meta-informazione implicita nella codifica

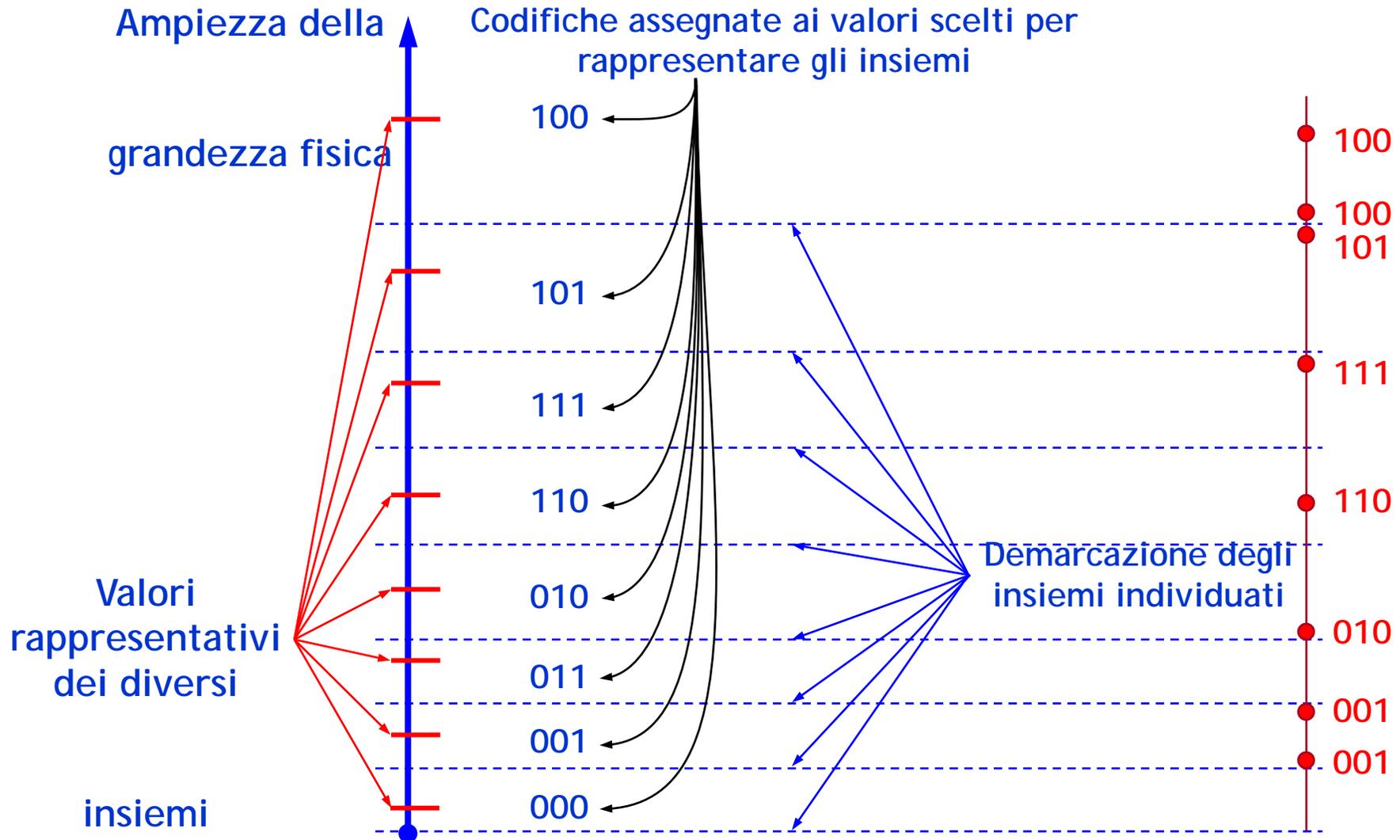
➤ **Codifica DIGITALE**

# Analogico vs digitale

	Analogico	Digitale
Definizione della codifica	Intensionale	Estensionale
Ridefinizione codifica vs variazione entità informazione	Nessuna ridefinizione	Estendere il processo definito in precedenza
Applicabilità vs cardinalità entità informazione	OK anche per cardinalità non finita	Solo per cardinalità finita e con entità note a priori
Applicabilità vs struttura entità informazione	Solo se c'è una struttura	OK in ogni caso
Condizioni sul supporto	Solo se il supporto ha una struttura	OK in ogni caso

# Da analogico a digitale:

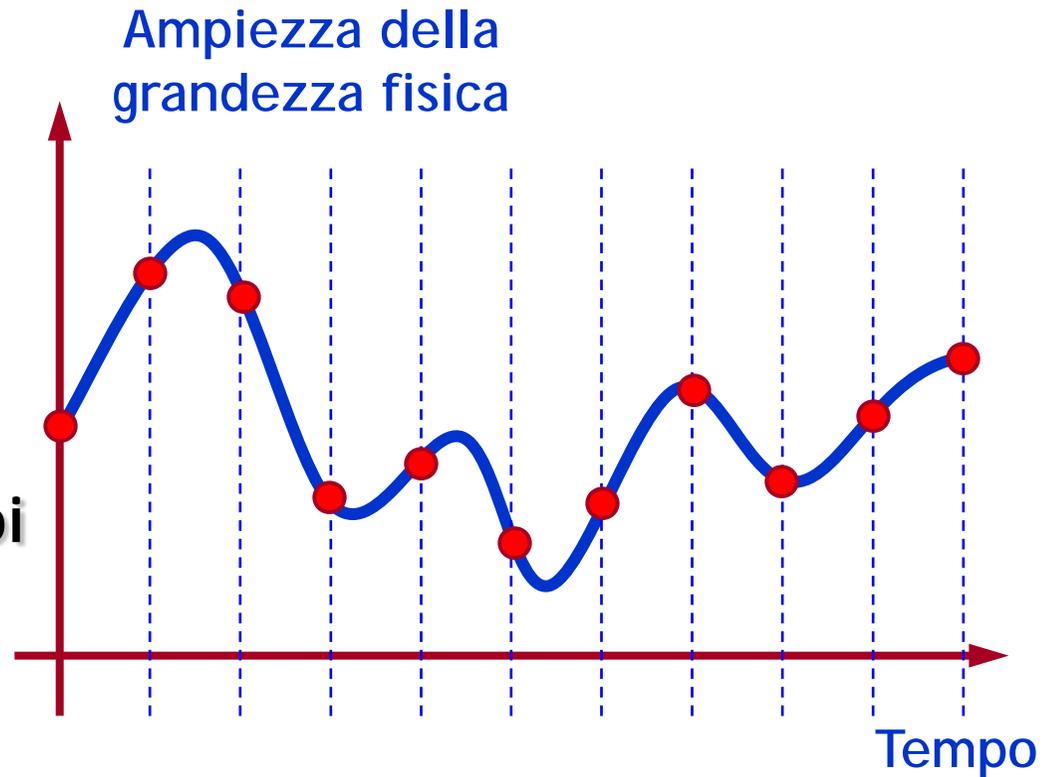
## 1. la quantizzazione



# Da analogico a digitale:

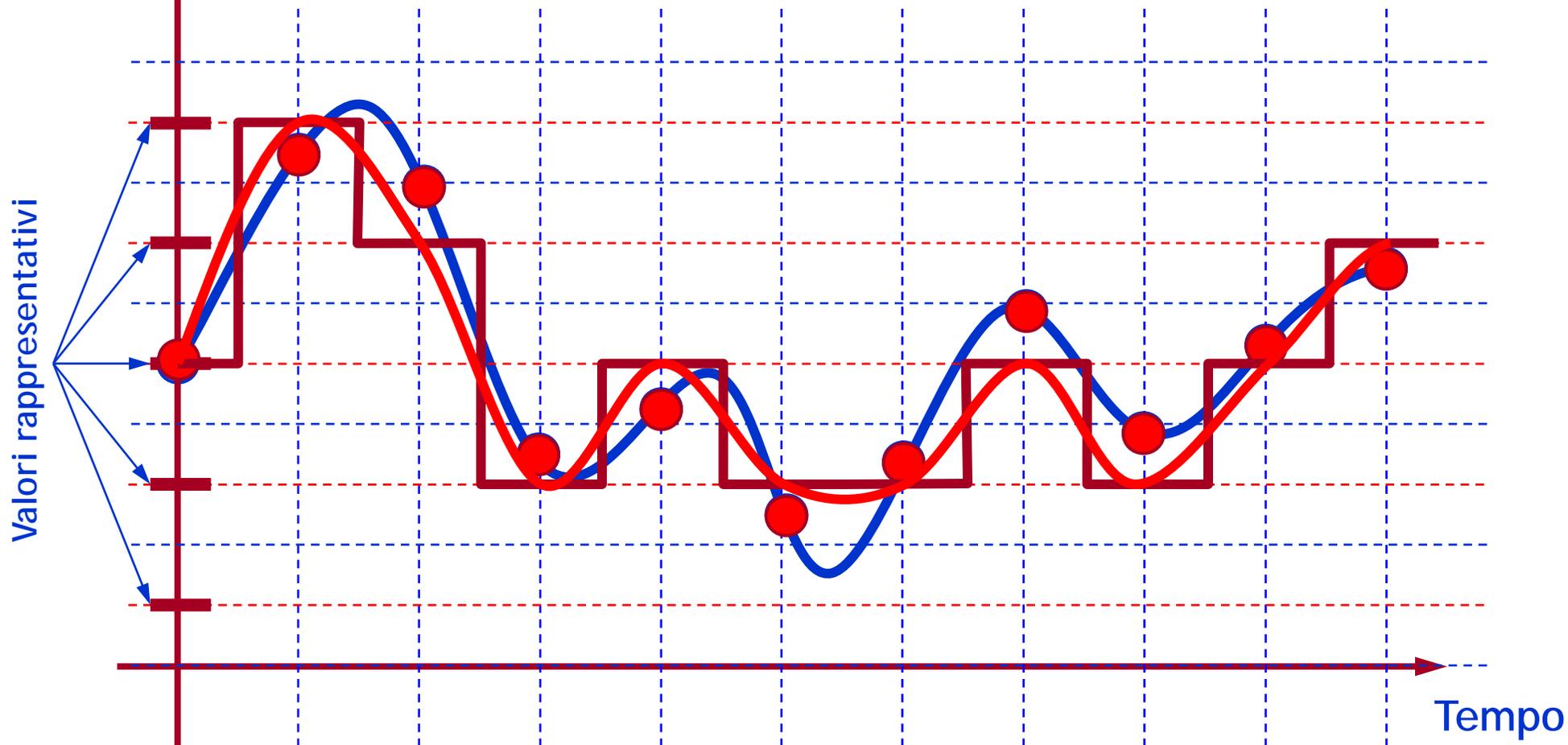
## 2. il campionamento

- La grandezza varia nel tempo e non può essere rappresentata da un solo valore.
- I valori di riferimento debbono essere rilevati in diversi istanti di tempo (frequenza di campionamento).
- La quantizzazione deve poi essere ripetuta per ogni valore campionato.

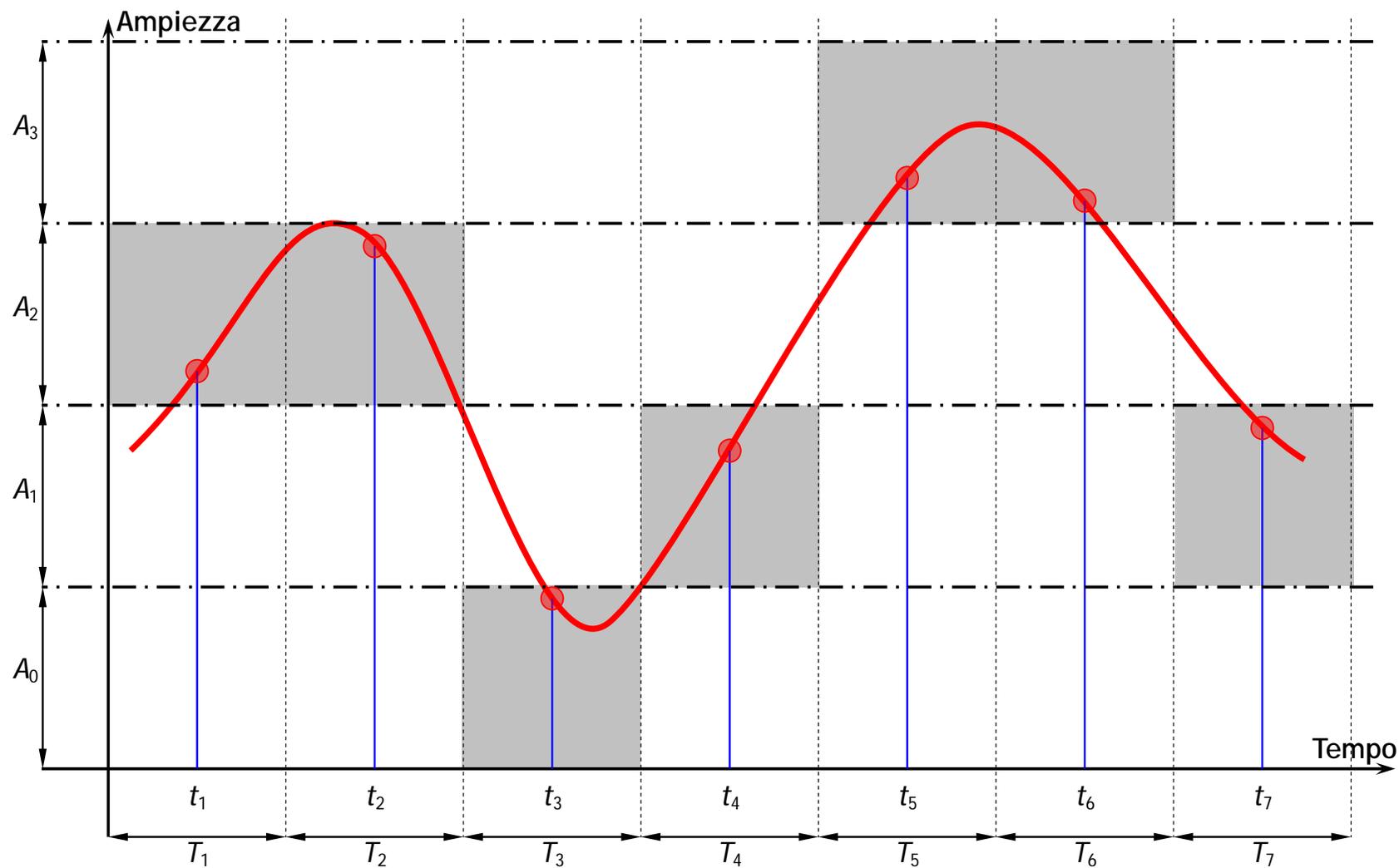


# Campionamento e quantizzazione

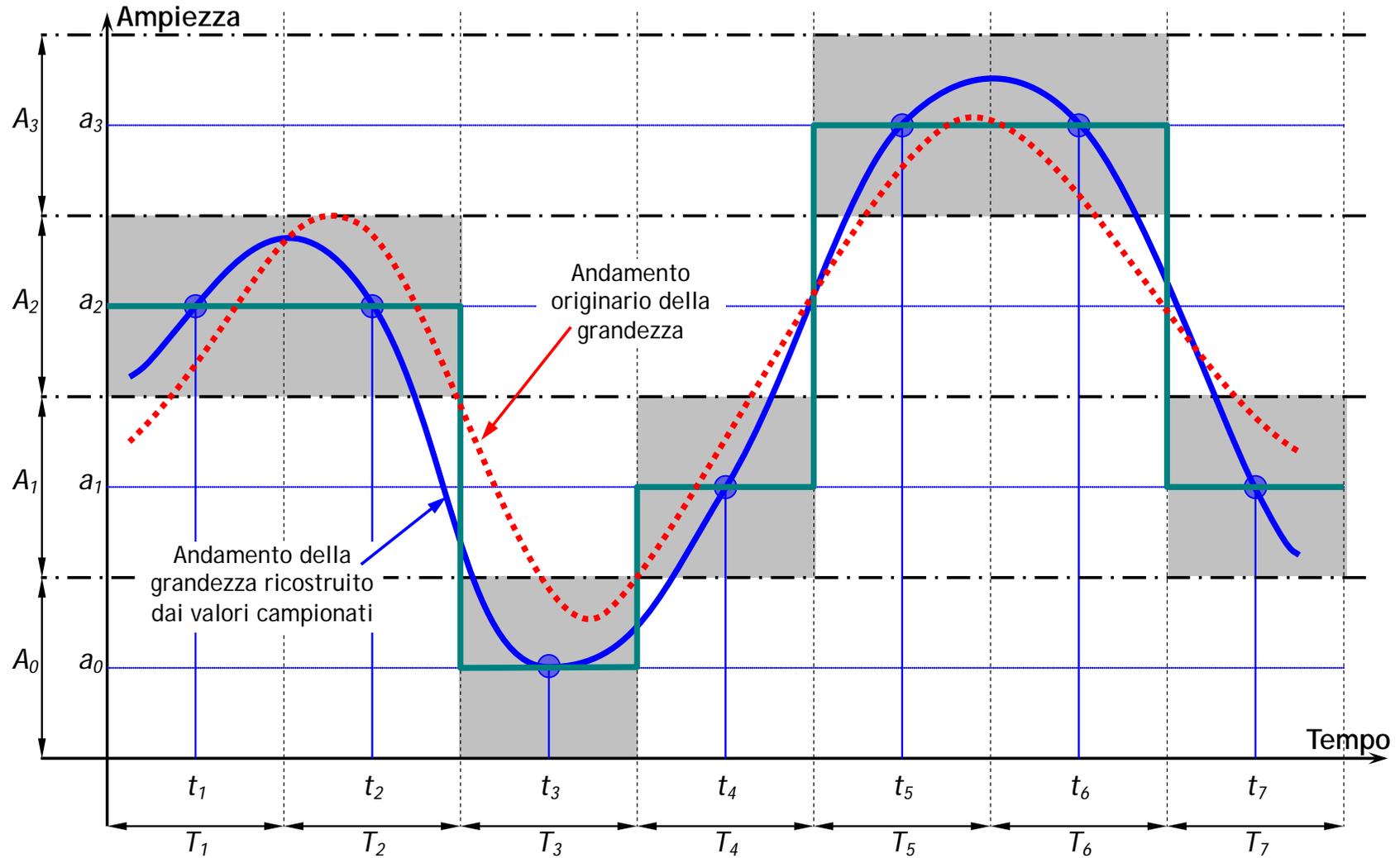
Ampiezza della  
grandezza fisica



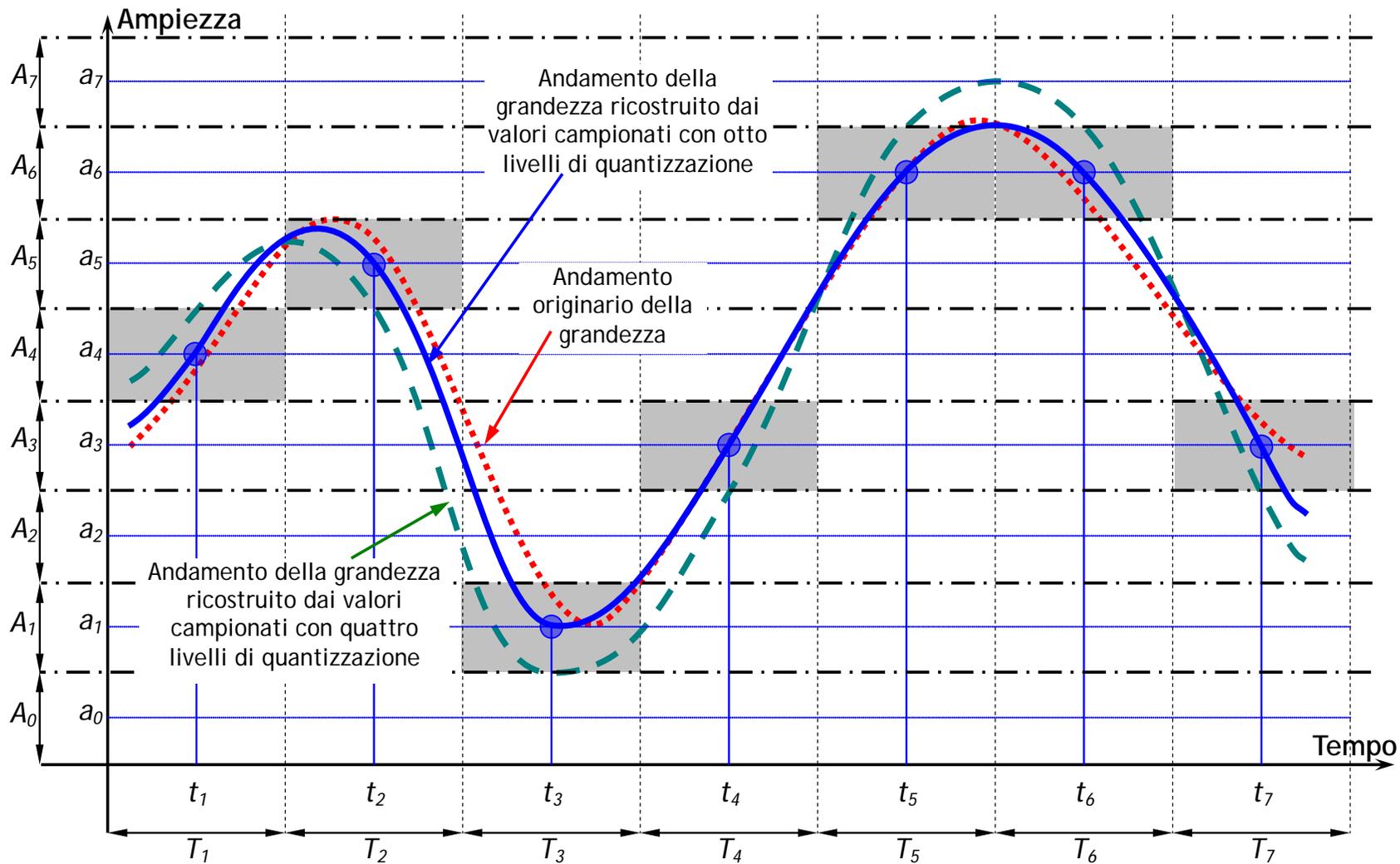
# Campionamento e quantizzazione



# Ricostruzione



# Quantizzazione su più livelli



# Suono digitale

- **Formato standard per i CD audio**
  - frequenza di campionamento di 44'100 Hz
  - quantizzazione su 65'536 livelli  
(un campione viene codificato su 16 bit)
- **Un secondo di musica stereo richiede 44'100 campioni di 16 bit (2 byte) ciascuno per due canali, quindi 176'400 byte.**
- **L'errore che si commette nella ricostruzione del segnale sonoro è difficilmente rilevabile da parte di un orecchio umano.**



# Foto digitali

- Per la codifica digitale delle immagini le operazioni di campionamento e quantizzazione si applicano nello **spazio** invece che nel **tempo**.
- Il **campionamento** consiste nel dividere l'immagine in sottoinsiemi (**pixel**, cioè **picture element**), per ognuno dei quali si dovrà prelevare un campione che si considera rappresentativo del colore di tutto il sottoinsieme.
- La **quantizzazione** è la codifica del colore associato a ogni pixel: i più recenti formati utilizzano **32 bit** (4 byte) per pixel: 8 bit per ognuna delle tre componenti fondamentali (**RGB: red, green, blue**) e altri 8 per gestire le  **trasparenze**.
- **Memoria necessaria per immagini non compresse (bitmap)**
  - per un'immagine di 640×480 pixel servono 1'228'800 byte;
  - per un'immagine di 800×600 pixel servono 1'920'000 byte;
  - per un'immagine di 1024×768 pixel servono 3'145'728 byte;
  - per un'immagine di 1280×1024 pixel servono 5'242'880 byte;
  - per un'immagine di 1600×1200 pixel servono 7'680'000 byte;
  - ...

# Il successo del digitale

- Rumore: effetto dell'ambiente sul supporto.
- Quanto un supporto è "immune" al rumore?
  - **Codifica analogica**: ogni configurazione è lecita dal punto di vista informazionale e quindi risulta impossibile distinguere il rumore dal segnale.
  - **Codifica digitale**: un valore binario è associato a un insieme di configurazioni valide quindi si può
    - riconoscere il rumore che porta in configurazioni non lecite
    - trascurare il rumore che non fa uscire il segnale dall'insieme associato alla stessa configurazione



# Come ridurre il numero di bit

## ➤ Esempio

- successione di un milione di caratteri, ognuno scelto dall'insieme {A, C, G, T};
- la frequenza dei quattro caratteri all'interno della successione non è uguale: A si presenta nel 50% dei casi, C nel 25%, G e T solo nel 12.5% dei casi.

## ➤ Codifica digitale a lunghezza costante

- due bit per ciascuno dei simboli, per esempio: A = 00, C = 01, G = 10 e T = 11;
- la lunghezza complessiva della successione è quindi pari a 2 milioni di bit.

## ➤ Codifica a lunghezza variabile (che tenga conto della distribuzione)

- A = 0, C = 10, G = 110 e T = 111;
- la lunghezza complessiva della successione è di 1.75 milioni di bit  
( $1 \times 50\% + 2 \times 25\% + 3 \times 12.5\% + 3 \times 12.5\%$ ) bit/carattere  $\times$  1 milione di caratteri

## ➤ Il cambiamento di codifica permette di ridurre il numero di bit utilizzato senza perdere informazione.

# La compressione dei dati

- **Gli algoritmi di compressione dei dati possono essere distinti in due categorie fondamentali**
  - Compressione **lossless**, se **non provoca** la perdita di informazione
  - Compressione **lossy**, se **provoca** la perdita di informazione
- **Gli algoritmi di compressione lossless sfruttano le regolarità nei dati**
  - RLE (Run Length Encoding): successioni di n simboli uguali vengono rappresentati con una coppia <simbolo, n>
  - Tecniche basate su dizionario: sequenze di simboli ripetute trovate analizzando i dati sono sostituite con simboli elementari memorizzando le corrispondenze in un dizionario
- **Gli algoritmi di compressione lossy sono specifici per i diversi domini applicativi.**

# Un semplice esempio con dizionario

## ➤ Compressione lossless con tecnica basata su un dizionario

- Testo di esempio:

*"I re di Francia della dinastia Carolingia sono: Carlo II, Luigi II di Francia, Luigi III di Francia, Carlomanno di Francia, Carlo III detto il grosso, Odo, Carlo III detto il semplice, Roberto I di Francia, Rodolfo Duca di Borgogna, Luigi IV di Francia, Lotario di Francia, Luigi V di Francia"* (lunghezza: 292 caratteri)

## ➤ Analisi delle regolarità presenti nel testo:

- Si individuano le sequenze ripetute (parole) contando le ripetizioni e si compila il dizionario (vedere tabella)
- Si assegna un simbolo che la sostituisce ad ogni parola
- Il testo diventa:  
*"I re 1 2 della 1nastia Carolingia sono: 5 3, 4 3 1 2, 4 3I 1 2, 5manno 1 2, 5 3I 6 7 grosso, Odo, 5 3I 6 7 semplice, Roberto I 1 2, Rodolfo Duca 1 Borgogna, 4 IV 1 2, Lotario 1 2, 4 V 1 2"* (lunghezza: 187 caratteri + 35 caratteri per il dizionario = 222 caratteri - 76% della lunghezza originaria)
- Un testo più lungo permette una efficienza maggiore!

Indice	Parola	N
1	di	10
2	Francia	8
3	Il	5
4	Luigi	4
5	Carlo	4
6	detto	2
7	il	2