



Università degli Studi di Foggia

Corso di aggiornamento INPS “Valore P.A.”

Gestione del Documento Informatico

Anno 2022

Modulo 2

Dematerializzazione

Prof. Crescenzo Gallo

crescenzo.gallo@unifg.it

<http://www.crescenziogallo.it/unifg/>

Professore Aggregato di Sistemi di Elaborazione delle Informazioni

Università degli Studi di Foggia

Dipartimento di Medicina Clinica e Sperimentale



Comprendere la differenza tra documento cartaceo e documento informatico e la valenza giuridica del documento informatico



- Dal documento cartaceo al documento informatico
- Equivalenza giuridica di documenti analogici e digitali
- Regole e tecniche per la formazione del documento informatico
- Tipologie di copie del documento informatico
- Copie informatiche di documenti analogici
- Estratti e duplicati informatici



Dal documento cartaceo al documento informatico



Dematerializzazione: *processo che ha come obiettivo ultimo la creazione di un flusso di documenti digitali aventi pieno valore giuridico, che vada prima ad affiancare e poi, sul lungo periodo, a sostituire la normale documentazione cartacea presente negli archivi di qualunque attività pubblica o privata.*

Il documento informatico è infatti definito come: *“la rappresentazione informatica di atti, fatti o dati giuridicamente rilevanti”*.



La dematerializzazione porta con sé una serie di vantaggi pratici rappresentati in prima istanza dall'**incremento di efficienza e la riduzione dei costi.**

La gestione del tradizionale documento cartaceo è infatti particolarmente onerosa e, se vogliamo, carente da diversi punti di vista: difficoltà di condivisione, facilità di smarrimenti, elevati tempi di ricerca e via discorrendo.



Alcune fasi temporali fondamentali:

- *Dematerializzazione degli strumenti finanziari, disciplinata in Italia dal D.Lgs. n. 213/1998 (Introduzione dell'Euro)*
- *DPR 445/2000 (Autocertificazione, protocollo informatico, ...)*
- *Direttiva CE n. 94/820/CE del 19/10/1994 (Electronic Data Interchange)*
- *Codice Amministrazione Digitale (D.Lgs 82/2005) (Definizione di documento informatico)*
- *Legge finanziaria 2008 (Obbligo fatturazione elettronica)*
- *Piano e-gov 2012 (Informatizzazione processi PA)*
- *Nuovo CAD (D.Lgs. 9/2018) (Digital first)*

Gli strumenti per la dematerializzazione



La piena riuscita del **processo di dematerializzazione** è garantita anche dall'applicazione diffusa e sistematica di tutti quegli **strumenti** disponibili a garantire l'autenticità dei documenti e all'adozione di sistemi di classificazione univoci e dettagliati che includano procedure per la conservazione e la selezione dei documenti stessi.

Gli strumenti per la dematerializzazione

Protocollazione e gestione documentale (DPR 445/2000)

L'attività di *protocollazione* è quella fase del processo amministrativo che *certifica provenienza e data di acquisizione del documento, identificandolo in maniera univoca per mezzo dell'apposizione di informazioni numeriche e temporali.*

Nella maggior parte delle amministrazioni, un sistema di protocollo informatico rappresenta quindi il primo passo nell'automazione dei procedimenti amministrativi o, più in generale, nel supporto all'informatizzazione dei processi o flussi di lavoro.

Gli strumenti per la dematerializzazione

Classificazione e Fascicolazione

La classificazione è un'attività di organizzazione logica dei documenti, protocollati e non, che nel caso della Pubblica Amministrazione dipende da una Area Organizzativa Omogenea, secondo uno schema articolato di voci che identificano funzioni, attività e materie specifiche della AOO stessa.

Mediante le operazioni di classificazione e registrazione di protocollo vengono attribuiti a ciascun documento dei codici di riferimento che lo identificano e lo associano agli altri documenti che formano la stessa pratica, nell'ambito di una delle serie di un determinato archivio.

Tale ordine reciproco prende la forma del fascicolo ossia di un *contenitore logico di documenti riferiti ad uno stesso procedimento amministrativo* e ad una stessa materia, utili allo svolgimento di una determinata attività. In un fascicolo trovano posto documenti diversi per formato, natura e contenuto.

Gli strumenti per la dematerializzazione

Firma digitale

La firma digitale è il risultato di una procedura informatica che garantisce l'autenticità e integrità dei messaggi e documenti scambiati e archiviati con mezzi informatici, al pari della firma autografa per i documenti tradizionali.

In sostanza i requisiti assolti sono l'autenticità (con un documento firmato digitalmente si può essere certi dell'identità del sottoscrittore) e l'integrità (sicurezza che il documento informatico non sia stato modificato dopo la sua sottoscrizione).

Gli strumenti per la dematerializzazione

Posta Elettronica Certificata (PEC)

La posta elettronica certificata è una e-mail che *garantisce ora e data di spedizione e di ricezione, provenienza ed integrità del contenuto.*

La PEC consente di inviare e ricevere messaggi con lo stesso valore legale di una raccomandata con avviso di ricevimento. È pertanto, un sistema di posta elettronica in grado di poter essere utilizzato in qualsiasi contesto nel quale sia necessario avere prova opponibile dell'invio e della consegna di un determinato documento.

Gli strumenti per la dematerializzazione

Conservazione delle risorse digitali

Un sistema di conservazione digitale è un *insieme di procedure, attività e strumenti tramite i quali ci si propone di salvaguardare nel tempo le memorie digitali.*

Insieme ai contenuti, si devono conservare anche le informazioni necessarie a garantire la riproducibilità della loro forma esteriore, nonché i collegamenti con le varie parti componenti l'intero sistema informativo.

Pertanto il processo di conservazione digitale è finalizzato al mantenimento delle caratteristiche di autenticità, integrità, leggibilità, accessibilità e riservatezza dei documenti informatici e deve prendere avvio subito dopo il loro ingresso nell'archivio digitale.



Equivalenza giuridica di documenti analogici e digitali

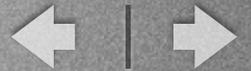


Il principio generale (espresso nel comma 1-bis dell'art. 20 del CAD, aggiornato dall'art. 17 del D.Lgs. 179/2016) prevede che **l'idoneità del documento informatico a soddisfare il requisito della forma scritta e il suo valore probatorio siano liberamente valutabili in giudizio**, tenuto conto delle sue caratteristiche oggettive di qualità, sicurezza, integrità ed immodificabilità.



La normativa di dettaglio, inoltre, definisce le misure tecniche, organizzative e gestionali volte a garantire l'integrità, la disponibilità e la riservatezza delle informazioni contenute nel documento informatico.

È inoltre previsto che l'intera attività di **dematerializzazione** dei documenti si svolga in maniera tale da garantire, comunque, il rispetto delle norme in materia di protezione dei dati personali.



Regole e tecniche per la formazione del documento informatico

Riferimenti normativi

- D. Lgs. n. 82/2005 (*CAD*): artt. 1, 20, 21, 22, 23, 23-bis, 23-ter, 23-quater, 40, 41, 71 (e successive modificazioni introdotte dal D.Lgs. 179/2016)
- Legge n. 241/1990 (*Trasparenza atti amministrativi*): art. 22
- Codice Civile: artt. 2702 (*efficacia scrittura privata*), 2712 (*riproduzione meccanica*)

Regole tecniche e provvedimenti attuativi

- Deliberazione CNIPA n. 45/2009 e Determinazione DigitPA 69/2010 (*Regole riconoscimento/verifica documento informatico*)
- DPCM 22/02/2013 e DPCM 03/12/2013 (*Regole tecniche firma digitale*)
- DPCM 13/11/2014 (*Regole tecniche in materia di formazione/gestione/conservazione dei documenti informatici delle pubbliche amministrazioni*)



La dematerializzazione permette di produrre documenti digitali che abbiano pieno valore giuridico.

Ciò significa, tra le varie cose, che anche con i documenti elettronici è necessario adottare un sistema che consenta di accertare in maniera chiara ed univoca il sottoscrittore di un documento.

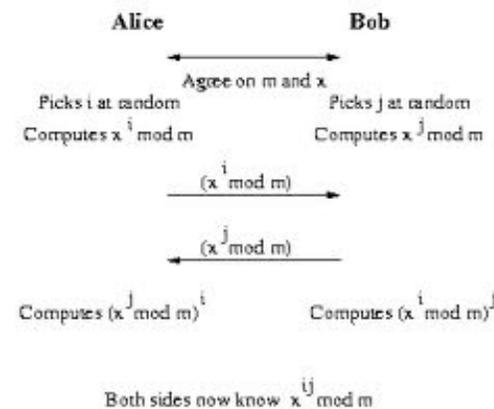


Nella tradizionale gestione della documentazione cartacea ciò è rappresentato dalla **firma autografa**: la firma apposta di pugno da chiunque sottoscriva un documento è considerato un elemento distintivo avente caratteristiche uniche e personali.



Perché, quindi, non applicare un concetto simile anche al documento elettronico?

Si tratta della **firma digitale**, un concetto già in uso da qualche anno e che sta ultimamente assumendo un'importanza via via maggiore proprio grazie alla spinta verso una gestione documentale sempre più digitalizzata.



È dalla seconda metà degli anni '70 che è iniziato a circolare il concetto di firma digitale con la descrizione da parte di Whitfield Diffie e Martin Hellman del primo schema di firma digitale, seppur fermo ad una fase concettuale.



Key Generation	
Select p, q	p, q both prime, p≠q
Calculate n = p×q	
Calculate φ(n) = (p-1)×(q-1)	
Select integer e	gcd(φ(n),e) = 1; 1<e<φ(n)
Calculate d	
Public key	KU = {e, n}
Private key	KR = {d, n}

Encryption	
Plaintext:	M < n
Ciphertext:	C = M ^e (mod n)

Decryption	
Ciphertext:	C
Plaintext:	M = C ^d (mod n)

Solamente qualche anno dopo, nel 1977, venne inventato l'**algoritmo RSA** da Ronald Rivest, Adi Shamir e Len Adleman che consentì di gettare le basi per realizzazione dei primi schemi di firma digitale a doppia chiave pubblica/privata.



- Nel 1989 è stato Lotus Notes 1.0 il primo software largamente disponibile in grado di usare l'algoritmo RSA.
- Da allora gli studi e le tecnologie legati alla firma digitale sono proseguiti su un cammino evolutivo che ha portato, per varie tappe, alle tecnologie di firma digitale di cui possiamo disporre oggi.



Se è dal 1979 che si parla, almeno concettualmente di firma digitale, è solamente 20 anni dopo nel 1999 che entrano in vigore le prime regole tecniche in materia di firma digitale, assieme alla Direttiva Comunitaria 1999/93/CE, dove si parla di firma digitale con riferimento a ciò che oggi viene definito in maniera molto più rigorosa come **Firma Elettronica Qualificata**^(*).

^(*) La firma elettronica qualificata (FEQ) - o digitale - è il risultato di una procedura informatica, detta validazione, che garantisce l'autenticità, l'integrità e il non ripudio dei [documenti informatici](#).

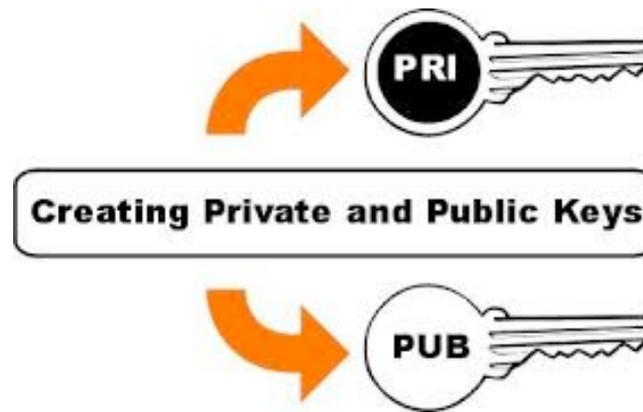


La firma digitale è lo strumento tecnologico che permette di sottoscrivere documenti digitali al massimo livello di **sicurezza, affidabilità, valore probatorio**.

Nel nostro paese, basato sulla *civil law*, ovunque sia richiesta la forma scritta e si desideri liberarsi della forma cartacea del documento, la firma digitale è la soluzione più naturale.



- Cerchiamo di capire quali sono i presupposti necessari affinché sia possibile utilizzare la firma digitale.
- Anzitutto è bene chiarire che il principio di funzionamento della Firma Elettronica Qualificata si basa sul concetto di *crittografia asimmetrica*: dal documento che viene sottoposto alla firma si ricava un'impronta, che viene successivamente criptata con una chiave privata.



- Per mezzo di una **chiave pubblica** si potrà accertare se la firma è stata cifrata con la corretta **chiave privata**.
- Si tratta di operazioni comunque automatiche e trasparenti all'utente, che non ha bisogno di conoscere il meccanismo di funzionamento.



La Firma Elettronica Qualificata necessita di tre elementi:

1. un **dispositivo di firma** ad elevata sicurezza, che contiene la chiave privata e che sia in possesso esclusivo del titolare. Si tratta, per entrare nel concreto, di una smartcard o di una chiavetta USB, cui vanno aggiungendosi recentemente servizi di *firma remota* basati su server centralizzati che ospitano le chiavi di molti utenti.



2. In secondo luogo è necessario un **certificato qualificato**: si tratta di un attestato elettronico mediante il quale il fornitore dei servizi di certificazione dichiara di avere identificato il titolare del dispositivo di firma o degli strumenti per l'accesso al servizio di firma remota e di averglieli consegnati.



3. Infine è necessario un software che sia in grado di operare e gestire il dispositivo di firma e di produrre documenti firmati mediante la chiave privata e completi di una copia del certificato qualificato, il quale consentirà di verificare la firma e accertare così il sottoscrittore del documento.



- Sul fronte della sicurezza uno dei vantaggi della firma digitale è rappresentato non solo dall'identificazione immediata del sottoscrittore, ma soprattutto dall'**impossibilità di modificare il contenuto** del documento una volta firmato: abbiamo infatti visto che la firma digitale viene collegata all'impronta del documento e cifrata, per questo motivo anche l'alterazione di un solo bit produrrebbe un'anomalia immediatamente riscontrabile.
- Ciò, se ci pensiamo, non è possibile con le normali firme su carta, in quanto il contenuto di un documento potrebbe essere facilmente alterato senza lasciare alcuna traccia.



Il processo di rilascio del dispositivo di firma e del certificato qualificato (con importanti responsabilità in capo al Certificatore) e lo strumento della revoca del certificato in caso di perdita o furto del dispositivo, rendono praticamente impossibile impersonare un altro individuo, a meno di falsificare la documentazione in fase di rilascio (di cui però rimane traccia per 20 anni).



- Inoltre la verifica dell'**autenticità** di una firma digitale e dell'**identità** del sottoscrittore è un processo automatico, istantaneo, molto affidabile.
- Basta pensare ad un documento firmato su carta per comprendere la differenza: solo se conosciamo personalmente il sottoscrittore e il suo modo di firmare possiamo capire a prima vista se la firma autografa è autentica o meno.
- La falsificazione di una firma autografa può essere scoperta solo da un perito calligrafo.



L'unico caso in cui sia possibile compromettere questo processo di sicurezza è, ancora una volta, ad opera dell'*anello più debole della catena*: l'utente, a causa di un suo comportamento superficiale (come ad esempio lasciare incustoditi i dispositivi come smartcard o chiave USB, assieme magari al relativo PIN di controllo!).



Tipologie di copie del documento informatico



Nel discorso della dematerializzazione il vero problema sta nella *legislazione* che regola i vari ambiti normativi in tema di documentazione; e va ricercato nella UNIFORMITÀ OPERATIVA e nell'UTILIZZO DI UNO STANDARD anch'esso uniforme per la digitalizzazione dei documenti.



Le copie informatiche di documenti analogici (che hanno contenuto giuridico identico a quelli analogici da cui sono tratti), provenienti da depositari pubblici autorizzati o da pubblici ufficiali, hanno piena efficacia se colui che le rilascia vi appone o vi associa una firma digitale o altra firma elettronica qualificata.

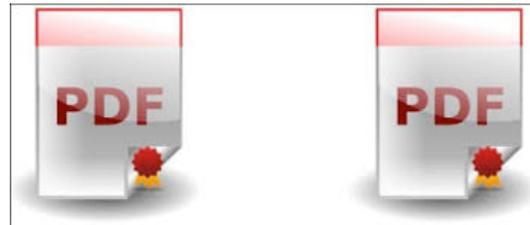
I documenti conformi a queste prescrizioni sostituiscono, a tutti gli effetti, gli originali.

(riferimenti normativi: Codice Civile artt. 2714 e segg.; D.P.R. 445/2000 artt. 18 e 19; D.Lgs. 82/2005)

Tipologie di copie

- le **copie per immagine** su supporto informatico di documenti originali formati in origine su supporto analogico (es. scansione di un documento originale cartaceo): analogico → digitale;
- le **copie su supporto analogico** di documenti informatici (ad es. la stampa di un documento), anche sottoscritti con firma elettronica avanzata, qualificata o digitale: digitale → analogico.





Con l'espressione **duplicato informatico** (definito come *documento informatico ottenuto mediante la memorizzazione, sullo stesso dispositivo o su dispositivi diversi, della medesima sequenza di valori binari del documento originario*) la normativa individua i documenti informatici che, oltre a rappresentare gli stessi atti o fatti giuridicamente rilevanti (quindi i contenuti del documento), mantengono il medesimo formato del file originale.

I duplicati informatici, conformi alle regole tecniche, hanno il medesimo valore giuridico del documento informatico da cui sono tratti.



Al contrario, nella **copia informatica** di un documento informatico (definita come il *documento informatico avente contenuto identico a quello del documento da cui è tratto con diversa sequenza di valori binari*) invece, pur se restano invariati i fatti giuridicamente rilevanti rappresentati nel documento, varia il formato del file.

Le copie e gli estratti informatici del documento informatico, hanno la stessa efficacia probatoria dell'originale da cui sono tratte se la loro conformità all'originale, in tutti le sue componenti, è attestata da un pubblico ufficiale a ciò autorizzato o se non è espressamente disconosciuta.



La rappresentazione delle informazioni

Il bit

Si consideri un alfabeto di 2 simboli: '0' e '1'



Che tipo di informazione si può rappresentare con un bit?

Il bit

Con un solo bit è possibile gestire un'informazione binaria, cioè un'informazione che può specificare uno tra due valori possibili (es. un punto di un'immagine bianco o nero).

Quanti stati possibili può assumere un insieme di bit ?

00	000	0000
01	001	0001
10	010	0010
11	011	0011
	100	0100
	101	0101
	110	0110
	111	0111
		1000
		1001
		1010
		1011
		1100
		1101
		1110
		1111

2 bit → 4 stati

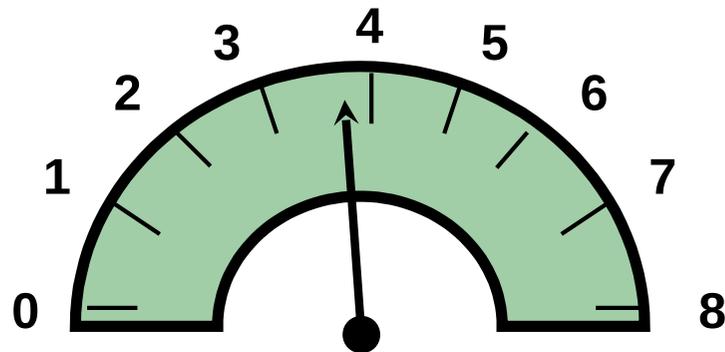
3 bit → 8 stati

4 bit → 16 stati

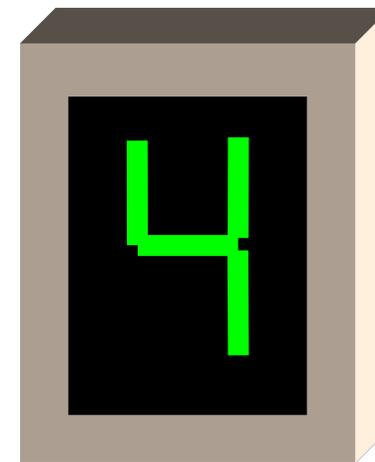
...

Con N bit: 2^N
possibilità diverse

Digitale vs Analogico



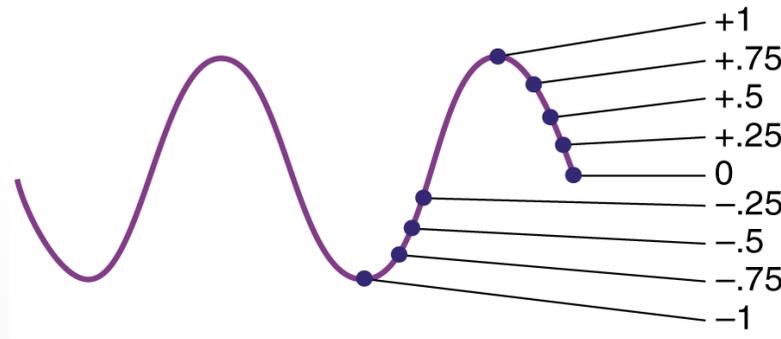
Nei sistemi **analogici**, le quantità vengono rappresentate in maniera **continua**.



Nei sistemi **digitali** le quantità vengono rappresentate in maniera **discreta**.

Digitale vs Analogico

I **segnali analogici** sono molto sensibili alle interferenze (rumore)



I **segnali digitali** possono assumere solo due stati

— [Per un dispositivo è semplice distinguere questi due stati, per cui vi è una maggiore immunità alle interferenze



Digitale

Una buona foto in bianco e nero presa da un giornale avrà circa 256 sfumature di grigio.

Rappresentazione analogica

256 gradi di luminosità con una lampada

Attenzione alle interferenze provocate dalla nebbia!

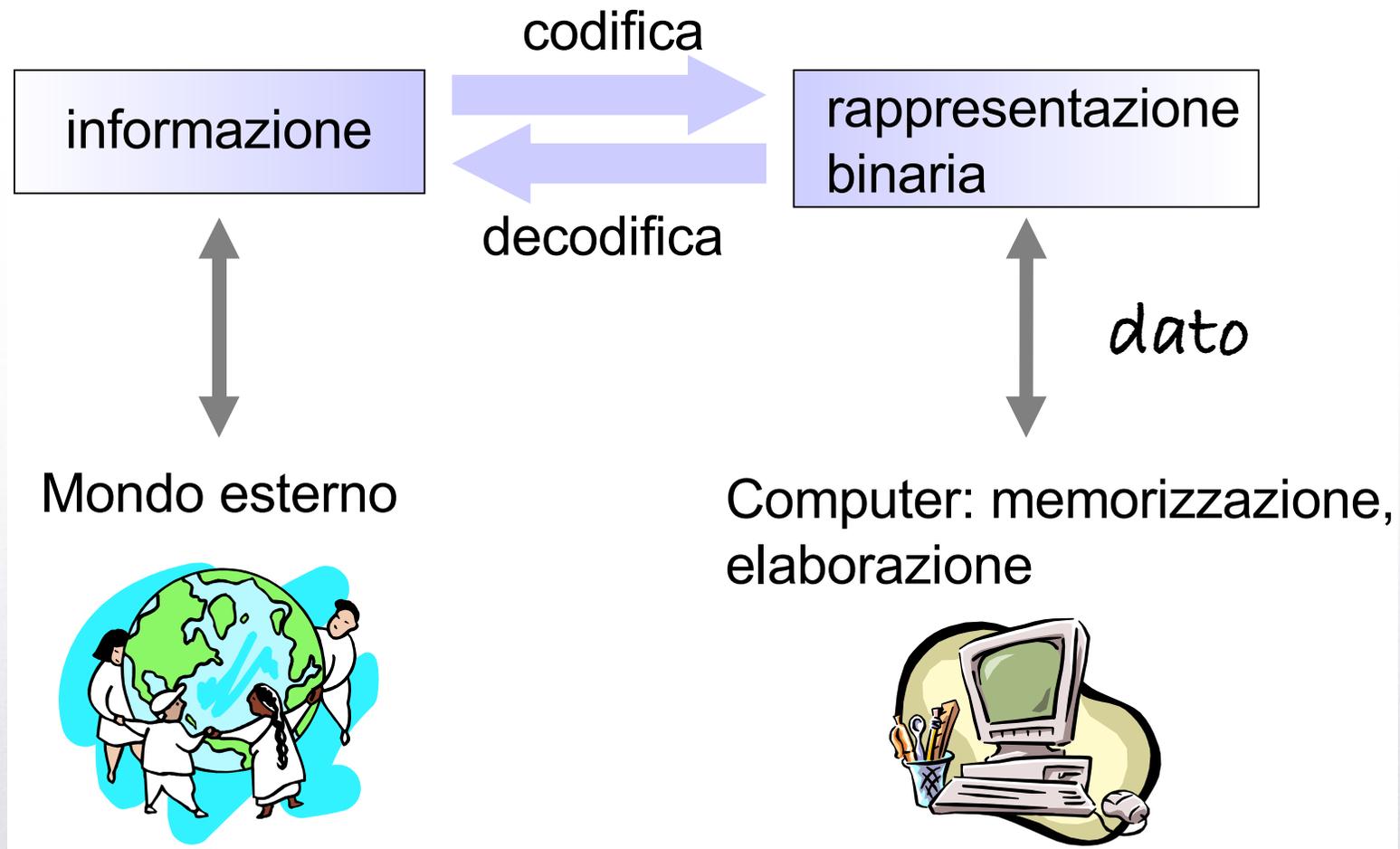
Rappresentazione digitale

8 lampade (bit; 256 configurazioni diverse)

Ciascuna configurazione sarebbe più **sicura** anche in caso di nebbia!

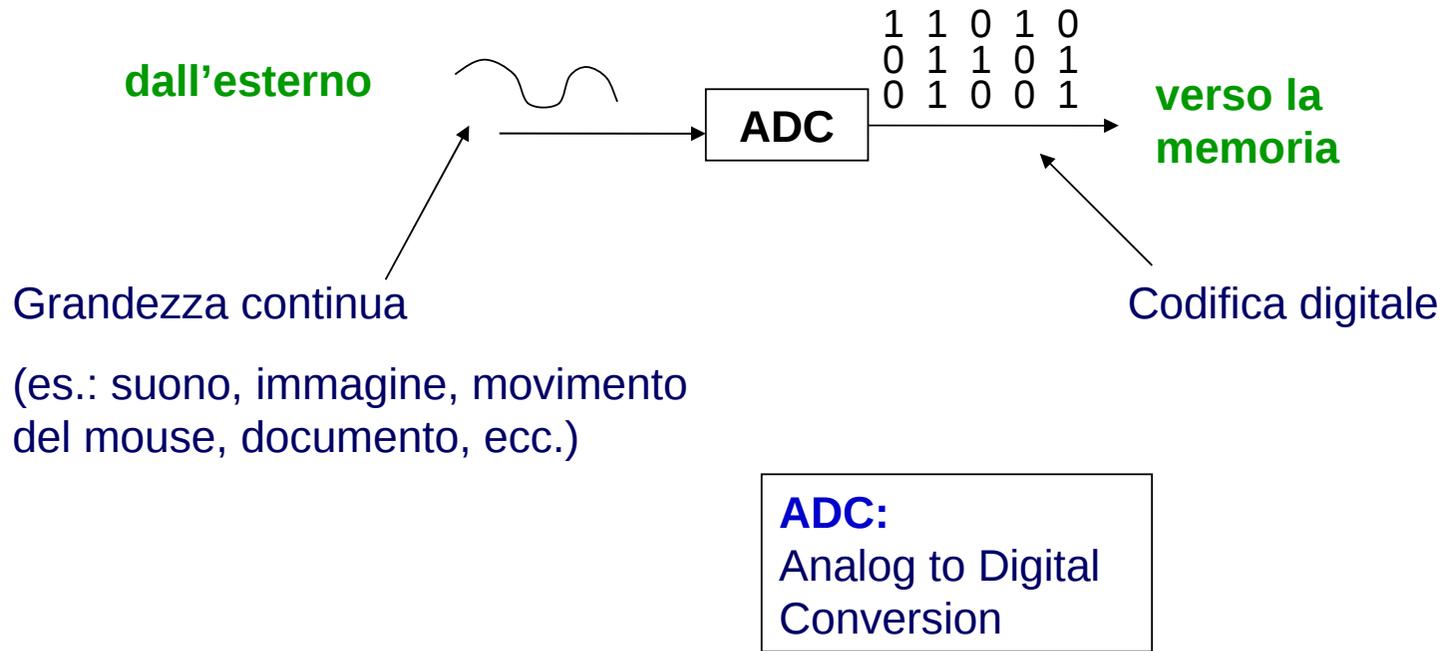


Rappresentazione binaria



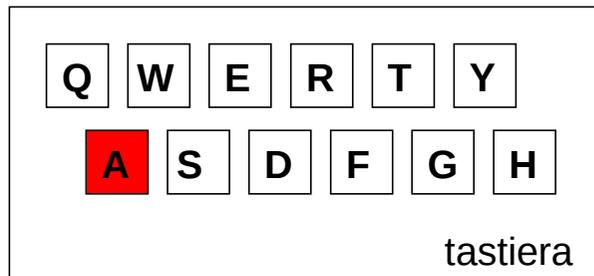
Analogico → Digitale

Compito delle periferiche di ingresso è quello di codificare una grandezza continua in ingresso tramite una rappresentazione digitale utilizzabile dal calcolatore.



Analogico → Digitale

Il tipo di informazione rappresentata dipende dalla periferica impiegata



10000001

the Algorithm
Stochastic

Eugen

scrimination is a general methodolog
rary numbers of very weak compone
very complex and accurate classifie
neralize to new data. In fact, it is ofte
ments are added, ever after an

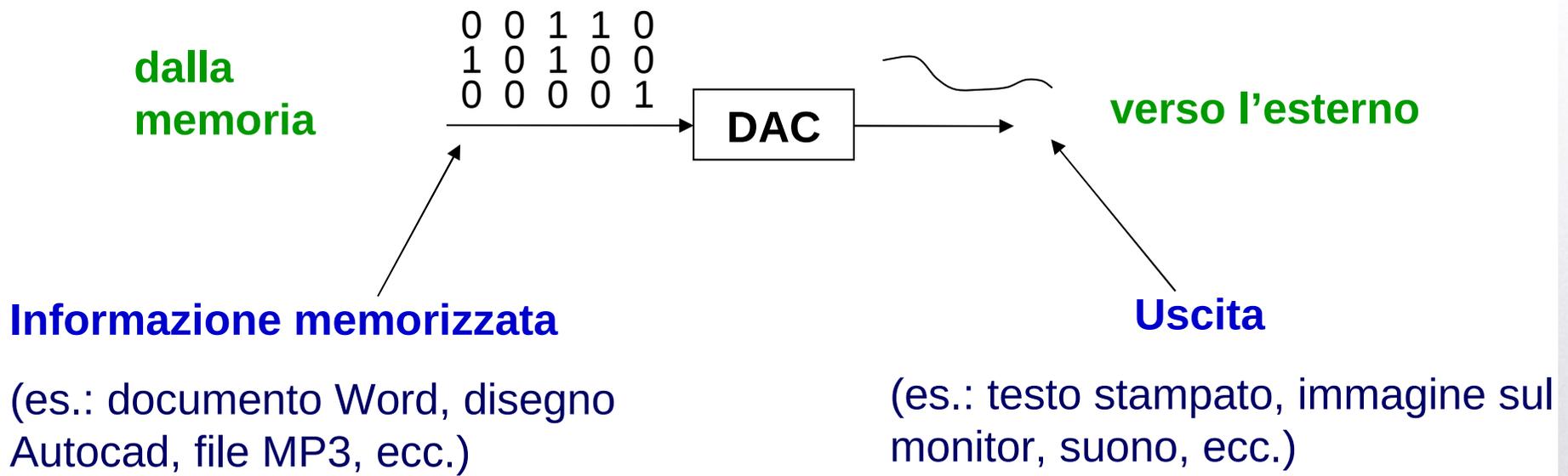


bit map

0	0	1	0	0
0	1	0	1	0
0	1	0	1	0
1	1	1	1	1
1	0	0	0	1
1	0	0	0	1

Digitale → Analogico

Compito delle periferiche di uscita è quello di creare, delle informazioni codificate in digitale nella memoria del calcolatore, una rappresentazione direttamente comprensibile dall'utente umano.





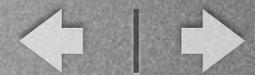
Sistemi di numerazione



Generalità

Per determinare un sistema di numerazione occorre:

- un insieme limitato di simboli (le **cifre**), che rappresentano quantità prestabilite (0,1,2,V,X,M)
- le **regole** per costruire i numeri
 - sistemi di numerazione **posizionali**
 - sistemi numerici **non posizionali**



Sistemi di numerazione

Sistemi numerici **non posizionali**

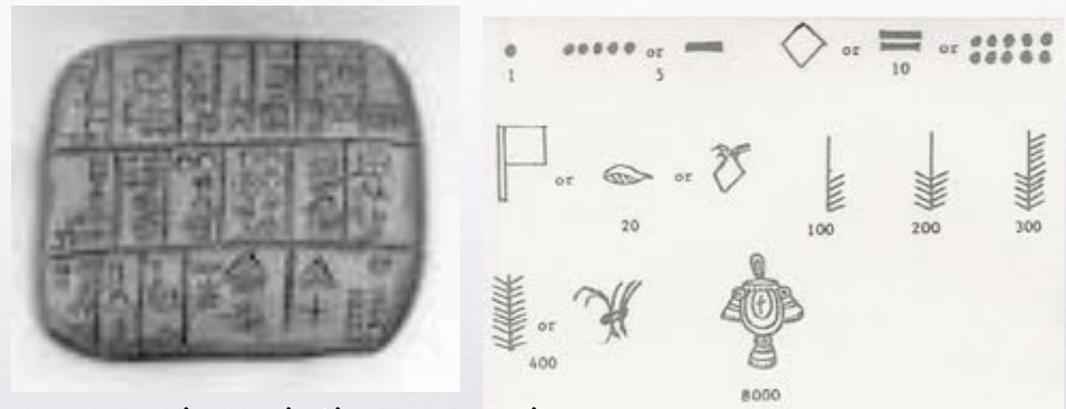
Il valore delle cifre è **indipendente** dalla posizione

- Es. Numeri romani

Sistemi numerici **posizionali**

Il valore delle cifre **dipende** dalla loro posizione all'interno del numero

- Ogni posizione ha un **peso**



sistemi di numerazione sumero e azteco



Sistemi posizionali

Esempio

$$N = c_3c_2c_1c_0$$

$$V(N) = c_3 \times p_3 + c_2 \times p_2 + c_1 \times p_1 + c_0 \times p_0$$

N = Rappresentazione del numero

$V(N)$ = Valore del numero

Sistemi a base fissa

$$p_i = r^i$$

r è la base del sistema



Sistema Decimale

È un sistema numerico posizionale a base fissa.

Il sistema decimale utilizza:

- $r = 10$ (*base*)
- $c = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ (*cifre*)



Sistema Decimale

Cifra più
significativa

Cifra meno
significativa

8427

=

$$8 \times 10^3 + 4 \times 10^2 + 2 \times 10^1 + 7 \times 10^0$$



Sistema Binario

Anche il sistema binario è un sistema numerico posizionale a base fissa.

Il sistema binario utilizza:

- $r = 2$ (*base*)
- $c = \{0,1\}$ (*cifre*)

Ogni cifra è detta **bit** (da **BI**nary digi**T**)



Sistema Binario

Bit più
significativo (MSB)

Bit meno
significativo (LSB)

1011₂

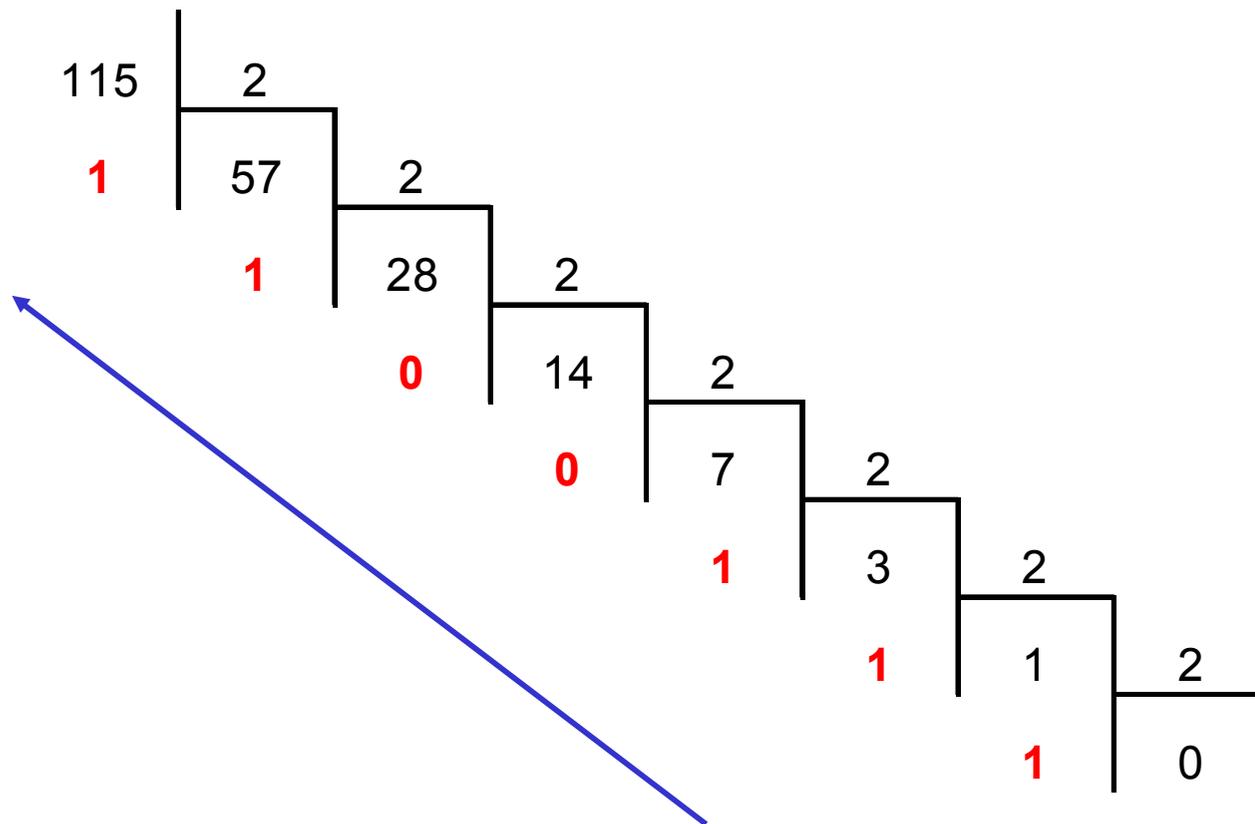
=

$$\mathbf{1} \times 2^3 + \mathbf{0} \times 2^2 + \mathbf{1} \times 2^1 + \mathbf{1} \times 2^0 = \mathbf{11}_{10}$$



Da Decimale a Binario

Es.: $115_{10} = 1110011_2$





Altre basi di numerazione

Sistema **ottale**

- $r = 8$ (*base*)
- $c = \{0, 1, 2, 3, 4, 5, 6, 7\}$ (*cifre*)

Sistema **esadecimale**

- $r = 16$ (*base*)
- $c = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F\}$ (*cifre*)



Base 2, 8 e 16

Esiste una corrispondenza diretta tra cifre ottali, esadecimali e il corrispondente binario.

Ottale: 8 cifre

- 3 bit per rappresentare una cifra ottale

11	000	110	₂
3	0	6	

Esadecimale: 16 cifre

- 4 bit per rappresentare una cifra esadecimale

1100	0110	₂
C	6	

$$198_{10} = 11000110_2 = 306_8 = C6_{16}$$



Bit necessari

Le macchine hanno vincoli spaziali

- E' necessario conoscere il massimo valore rappresentabile
 - Con n bit si può rappresentare al massimo il numero $2^n - 1$
- E' facile determinare che per poter rappresentare al massimo il valore X , sono necessari un numero n di bit pari a:

$$\lceil \log_2 X \rceil$$

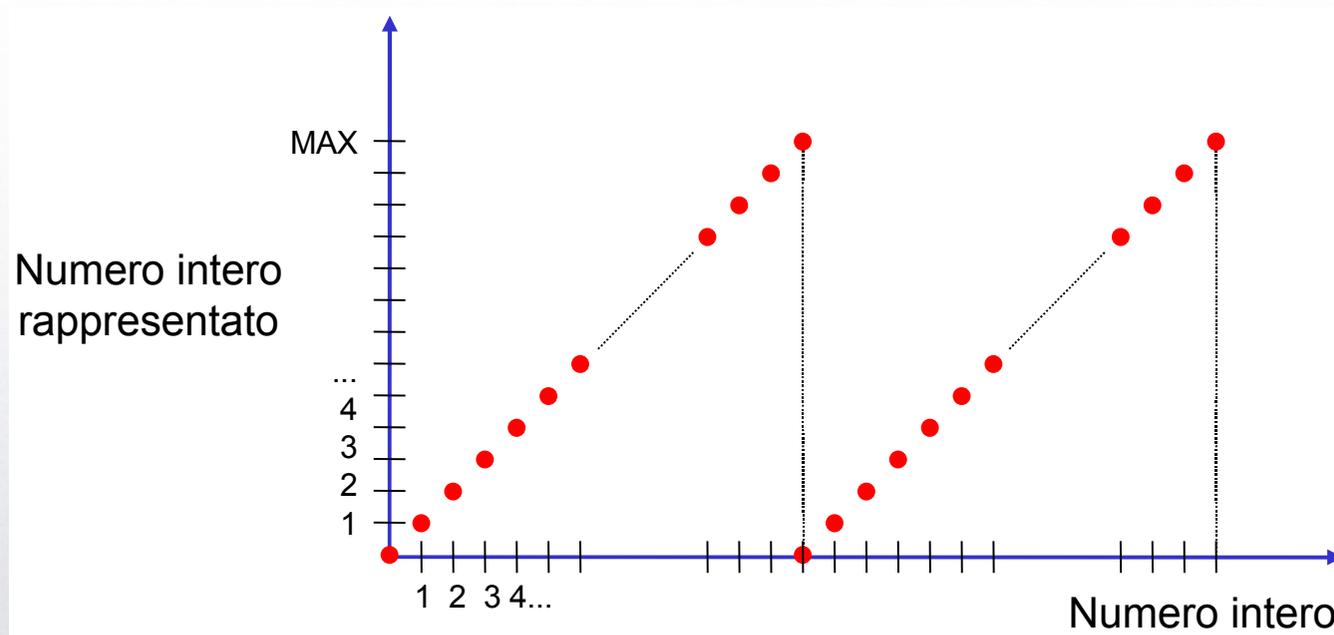
dove $\lceil y \rceil$ restituisce il più piccolo numero intero maggiore o uguale a y .



Overflow

Esiste un limite al numero di bit impiegati per rappresentare un numero.

Dato che la rappresentazione è formata da un numero finito di bit, se si supera tale limite si ha errore (**overflow**).





Codifica delle informazioni



Informazioni

Numeri

- *Naturali*
- *Relativi*
- *Reali*

Testi



*Informazioni
tradizionali*

Immagini

- *Bitmap*
- *Vettoriali*

Audio

Video



*Informazioni
multimediali*



Bit, Byte e Word

- L'unità atomica è il **bit** (**B**inary **D**igi**T**)
- L'insieme di 8 bit è detto **byte**
- **Word**
 - ✓ Tipicamente 16, 32 o 64 bit
 - ✓ Insieme di bit la cui dimensione è una importante caratteristica del calcolatore considerato. Essa influenza:
 - La larghezza degli indirizzi
 - La dimensione dei registri del processore
 - Larghezza dei bus (word o multipli di essa)



Il problema della codifica

- Un calcolatore può trattare **diversi tipi di dati**: numeri (interi, reali), testo, immagini, suoni, etc. che vanno comunque memorizzati in registri di memoria.
- È quindi necessario adottare una **codifica** del tipo di dato considerato: occorre, cioè, mettere in corrispondenza biunivoca i valori del tipo con gli stati che può assumere il registro.

Registro da un byte $\Rightarrow 2^8 = 256$ stati possibili. Che cosa è possibile codificare ?

Numeri naturali [0,255]

0	↔	00000000
1	↔	00000001
...		
255	↔	11111111

Numeri interi [-128,127]

-128	↔	00000000
-127	↔	00000001
0	↔	10000000
+127	↔	11111111

Numeri reali [0,1[

0.0000	↔	00000000
0.0039	↔	00000001
0.0078	↔	00000010
...		
0.9961	↔	11111111

Caratteri

A	↔	01000001
a	↔	01100001
0	↔	00110000
1	↔	00110001

La codifica implica una rappresentazione dei dati limitata e discreta



Codifica binaria

- Esiste una particolare aggregazione di bit che è costituita da **8 bit** ($2^8 = 256$ informazioni) e prende il nome di **byte (B)**
- Di solito si usano i multipli del byte

Kilo	KB	2^{10} (~ un migliaio, 1024 byte)
Mega	MB	2^{20} (~ un milione, 1KB x 1024)
Giga	GB	2^{30} (~ un miliardo, 1MB x 1024)
Tera	TB	2^{40} (~ mille miliardi, 1GB x 1024)
Peta	PB	2^{50} (~ miliardo miliardi, 1TB x 1024)



Codifica dei testi

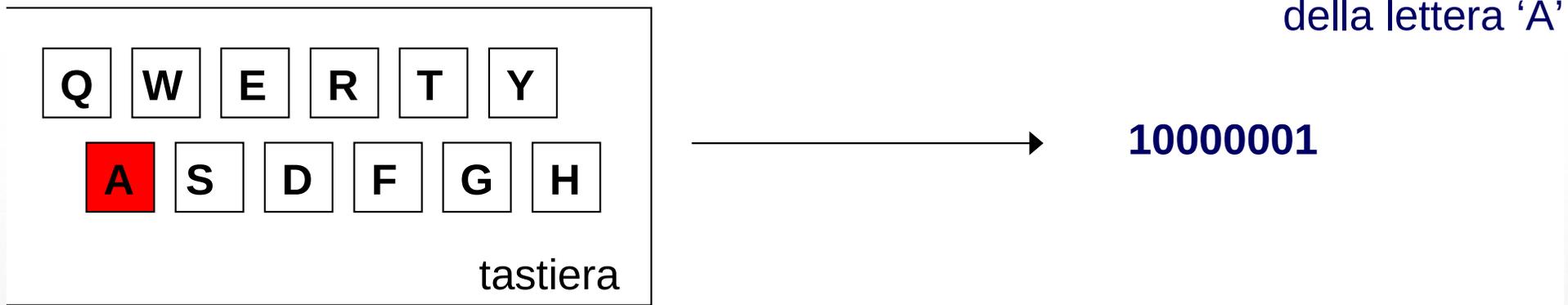
- ➔ Si utilizza una tabella (arbitraria)
- ➔ Standard oggi (quasi) universalmente riconosciuto il codice ASCII (**A**merican **S**tandard **C**ode for **I**nformation **I**nterchange)

Nella versione base ogni carattere (simbolo) è codificato con 7 bit = 128 simboli diversi

Versione estesa di 8 bit = 256 simboli diversi



La codifica ASCII



La Codifica ASCII serve a codificare i caratteri alfanumerici.

Il Formato RTF (Rich Text Format) memorizza alcune caratteristiche aggiuntive dei caratteri.



Tabella codici ASCII base

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040	 	Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	\$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	((72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051))	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[[123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174	|	
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135]]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177		DEL



Codice ASCII esteso

0		32		64	@	96	·	128	Ç	160	á	192	L	224	ó
1	☺	33	!	65	A	97	a	129	ü	161	í	193	↓	225	ß
2	☹	34	"	66	B	98	b	130	é	162	ó	194	↑	226	Ô
3	♥	35	#	67	C	99	c	131	â	163	ú	195	†	227	Õ
4	♦	36	\$	68	D	100	d	132	ä	164	ñ	196	—	228	ô
5	♣	37	%	69	E	101	e	133	à	165	Ñ	197	+	229	Ö
6	♠	38	&	70	F	102	f	134	â	166	ª	198	ã	230	µ
7	·	39	'	71	G	103	g	135	ç	167	º	199	Ä	231	þ
8	☐	40	(72	H	104	h	136	ê	168	¿	200	Å	232	ƒ
9	○	41)	73	I	105	i	137	ë	169	®	201	Æ	233	Ú
10	☒	42	*	74	J	106	j	138	è	170	™	202	⊥	234	Û
11	♂	43	+	75	K	107	k	139	ÿ	171	½	203	⌢	235	Ü
12	♀	44	,	76	L	108	l	140	î	172	¼	204	⌣	236	ý
13	♪	45	-	77	M	109	m	141	ï	173	⅓	205	≡	237	ÿ
14	♫	46	.	78	N	110	n	142	ÿ	174	«	206	≠	238	—
15	☼	47	/	79	O	111	o	143	ÿ	175	»	207	α	239	˙
16	▶	48	0	80	P	112	p	144	É	176		208	ø	240	-
17	◀	49	1	81	Q	113	q	145	æ	177		209	Ð	241	±
18	↑	50	2	82	R	114	r	146	Æ	178		210	È	242	—
19	!!	51	3	83	S	115	s	147	ø	179		211	Ë	243	¾
20	¶	52	4	84	T	116	t	148	ö	180	↓	212	Ë	244	¶
21	§	53	5	85	U	117	u	149	ò	181	À	213	ì	245	§
22	—	54	6	86	V	118	v	150	û	182	Ã	214	í	246	÷
23	↑	55	7	87	W	119	w	151	ù	183	Ä	215	î	247	,
24	↑	56	8	88	X	120	x	152	ÿ	184	©	216	ï	248	°
25	↓	57	9	89	Y	121	y	153	ÿ	185	®	217	ÿ	249	ˆ
26	→	58	:	90	Z	122	z	154	ÿ	186		218	ÿ	250	˙
27	←	59	;	91	[123	{	155	ø	187	¶	219	☐	251	ˆ
28	↔	60	<	92	\	124		156	£	188	¶	220	☐	252	³
29	↔	61	=	93]	125	}	157	Ø	189	ø	221	ÿ	253	²
30	▲	62	>	94	^	126	~	158	x	190	¥	222	ÿ	254	☐
31	▼	63	?	95	_	127		159	f	191	∟	223	☐	255	



Codice EBCDIC

In informatica, la sigla **EBCDIC** (dall'inglese Extended Binary Coded Decimal Interchange Code) indica un sistema di codifica dell'informazione a 8 bit usato in numerosi sistemi operativi di produzione IBM, sia per elaboratori di classe mainframe che per minicomputer.

Viene inoltre utilizzato da varie piattaforme di altri produttori.

Deriva dalla codifica a 6 bit binary-coded decimal, utilizzata nelle schede perforate e nella maggior parte delle periferiche IBM della fine degli anni 1950 e dell'inizio degli anni 1960.

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F
0-	NUL 00	SOH 01	STX 02	ETX 03	SEL 04	HT 09	RNL 08	DEL 7F	GE 0E	SPS 09	RPT 0A	VT 0B	FF 0C	CR 0D	SO 0E	SI 0F
1-	DLE 10	DC1 11	DC2 12	DC3 13	RES ENP	NL 0A	BS 08	POC 09	CAN 18	EM 19	UBS 0A	CU1 1C	IFS 1C	IGS 1D	IRS 1E	IUS ITB 1F
2-	DS 0A	SOS 0B	FS 0C	WUS 0D	BYP INP	LF 0A	ETB 17	ESC 1B	SA 0A	SFE 0A	SM SW	CSP 0A	MFA 0A	ENQ 05	ACK 06	BEL 07
3-			SYN 16	IR 0A	PP 0A	TRN 0A	NBS 0A	EOT 04	SBS 0A	IT 0A	RFF 0A	CU3 0A	DC4 14	NAK 15		SUB 1A
4-	SP 20	RSP A0	á E2	â E4	ã E0	ä E1	å E3	æ E5	ç E7	ñ F1	[5B	. 2E	< 3C	(28	+ 2B	! 21
5-	& 26	ê E9	ë EA	è EB	é E8	í E0	î E2	ï E4	ì E6	í E8] 5D	\$ 24	* 2A) 29	; 3B	^ 5E
6-	- 2D	/ 2F	Ã C2	Ä C4	Å C0	Á C1	Ã C3	Ä C5	Ç C7	Ñ D1	: A6	' 2C	% 25	_ 5F	> 3E	? 3F
7-	ø F8	É C9	Ê CA	Ë CB	È C8	Ì CD	Í CE	Î CF	Ï CC	· 60	: 3A	# 23	@ 40	' 27	= 3D	" 22
8-	Ø D8	a 61	b 62	c 63	d 64	e 65	f 66	g 67	h 68	i 69	« AB	» BB	ð F0	ý FD	þ FE	± B1
9-	° B0	j 6A	k 6B	l 6C	m 6D	n 6E	o 6F	p 70	q 71	r 72	ª AA	º BA	æ E6	· B8	Æ C6	¤ A4
A-	µ B5	~ 7E	s 73	t 74	u 75	v 76	w 77	x 78	y 79	z 7A	ı A1	ı B2	Đ D0	Ÿ DD	ƒ DE	® AE
B-	¢ A2	£ A3	¥ A5	· B7	© A9	§ A7	¶ B6	¼ BC	½ BD	¾ BE	¬ AC	ı 7C	- A2	- A8	- B4	x D7
C-	{ 7B	A 41	B 42	C 43	D 44	E 45	F 46	G 47	H 48	I 49	SHY AD	ô F4	õ F6	ò F2	ó F3	ô F5
D-	} 7D	J 4A	K 4B	L 4C	M 4D	N 4E	O 4F	P 50	Q 51	R 52	ı B9	û FB	ü FC	ù F9	ú FA	ÿ FF
E-	\ 5C	÷ F7	S 53	T 54	U 55	V 56	W 57	X 58	Y 59	Z 5A	² 82	Ó D4	Ô D6	Ò D2	Ó D3	Ô D5
F-	0 30	1 31	2 32	3 33	4 34	5 35	6 36	7 37	8 38	9 39	³ 83	Û D8	Ü DC	Ù D9	Ú DA	EO



UNICODE

Un altro codice molto diffuso oggi è lo UNICODE, che impiega 16 bit per carattere (Extended ASCII + caratteri etnici):

$2^{16} = 65.536$ simboli diversi

Derivazioni: UTF-8,16,32 per il web

Da Wikipedia

Unicode era stato originariamente pensato come una codifica a 16 bit (quattro cifre esadecimali) che dava la possibilità di codificare 65.536 caratteri. Tanto si riteneva essere sufficiente per rappresentare i caratteri impiegati in tutte le lingue scritte del mondo.

Ora invece lo standard Unicode, che tendenzialmente è perfettamente allineato con la norma [ISO/IEC 10646](#), prevede una codifica fino 21 bit e supporta un repertorio di codici numerici che possono rappresentare circa un milione di caratteri. Ciò appare sufficiente a coprire anche i fabbisogni di codifica di scritti del patrimonio storico dell'umanità, nelle diverse lingue e negli svariati sistemi di segni utilizzati.



Informazioni multimediali (immagini, suoni, video)



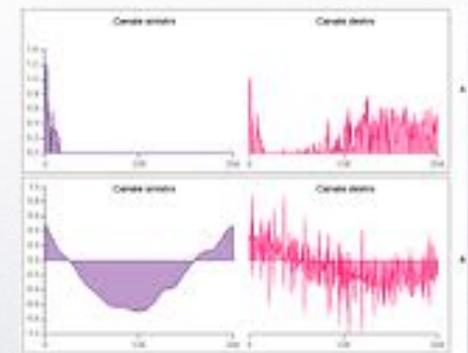
Lettere e numeri non costituiscono le uniche informazioni utilizzate dagli elaboratori, ma ci sono sempre più applicazioni che utilizzano ed elaborano anche altri tipi di informazione:

→ **diagrammi, immagini, suoni.**

In questi casi si parla di applicazioni di tipo *multimediale*.

Abbiamo bisogno di un sistema di **CODIFICA**.

In generale si parla di **DIGITALIZZAZIONE**.





La digitalizzazione

Il termine digitale deriva dall'inglese **digit** (cifra, numero).

Per *digitalizzazione* si intende la rappresentazione (codifica) di un qualunque fenomeno o oggetto fisico attraverso una sequenza di numeri.

La digitalizzazione consente di rappresentare un qualunque oggetto o fenomeno all'interno della memoria del computer.

Consente ad un calcolatore di interagire, analizzare, modificare le rappresentazioni di tali oggetti.





Esempi di digitalizzazione

Immagine



Codifica

Codifica jpeg

```

0101000101010111101
0101110101000000101
0101000100100010000
0011110101111100010
1010110101010100110
1101010100000110101

```

Testo

Il termine *word* indica, invece, una serie di bit (in numero tale da essere potenza di 2) che hanno un particolare significato; si parla quindi di *word* di 4, 16, 32 o 64 bit. Normalmente, una *word* è la dimensione minima di bit su cui un calcolatore può eseguire operazioni elementari; i vecchi PC lavoravano con *word* di 8 o 16 bit, mentre gli attuali elaboratori hanno *word* di 64 bit.

In definitiva, se per noi è semplice eseguire calcoli nella forma $12+15$ o ricercare la parola "ciao" in un testo, per un elaboratore elettronico è molto più facile sommare 1100 a 1110 (intesi come numeri binari) o cercare la sequenza di bit "01110101011101100101011011001101" (possibile codifica della parola "ciao") in una serie di un milione di cifre binarie.

4.1.2 Multipli utilizzati

La memoria utilizzata per codificare una pagina di testo è di qualche *migliaio* di byte, quella usata per una immagine può raggiungere il *milione* di byte, mentre un lungo filmato può richiedere *miliardi* di byte per essere memorizzato.

Si avverte la necessità, come nel *sistema matricale decimale*, di utilizzare dei simboli per rappresentare i multipli delle grandezze elementari; nella terminologia informatica sono stati quindi adottati gli stessi simboli del sistema decimale, ma visto che la misurazione della memoria ha come sua base principale il 2, il loro significato è *leggermente diverso*.

La seguente tabella riassume i simboli e i valori dei multipli più usati in campo informatico:

Codifica

Codifica ASCII

```

0101000101010111101
0101110101000000101
0101000100100010000
0011110101111100010
1010110101010100110
1101010100000110101

```



Conversione analogico-digitale



La conversione analogico-digitale trasforma un segnale analogico (valori continui in un tempo continuo) in segnale numerico (valori discreti in tempo discreto).

Questa operazione è costituita da tre fasi:

- ▶ **Campionamento**
- ▶ **Quantizzazione**
- ▶ **Codifica**

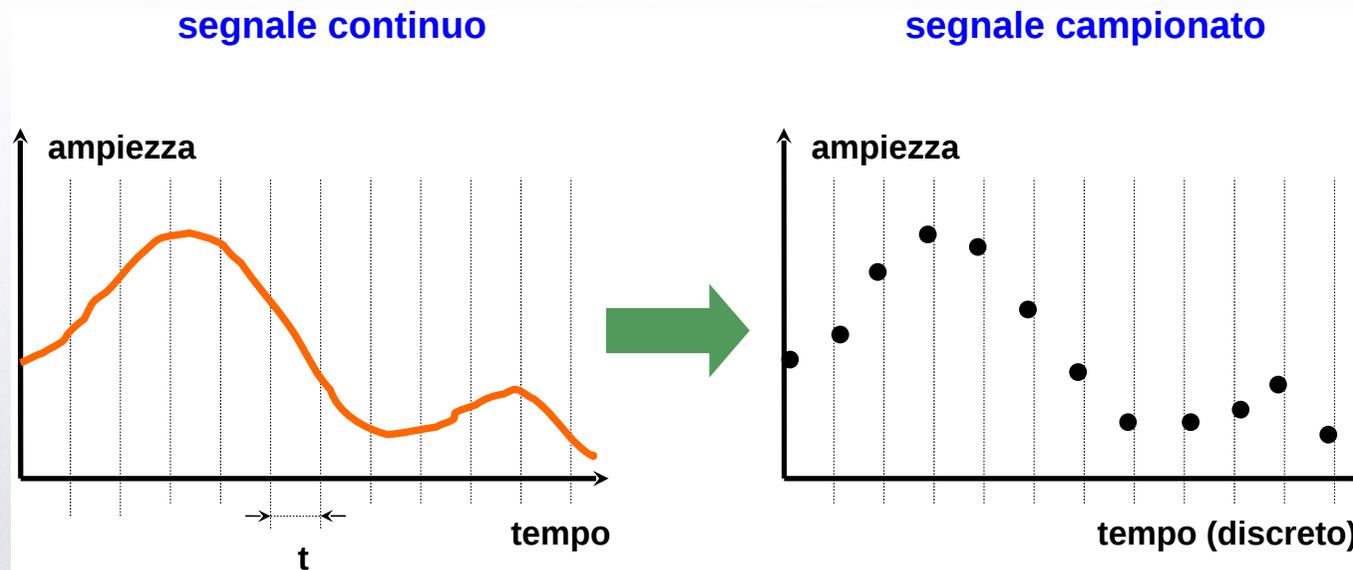


Campionamento

Il **campionamento** è una tecnica di rappresentazione del segnale che consiste nella valutazione dell'ampiezza dello stesso ad intervalli di tempo regolari.

L'accuratezza di un campionamento dipende essenzialmente dalla **frequenza** di campionamento che è il numero di campioni rilevato nell'unità di tempo.

Esempio:

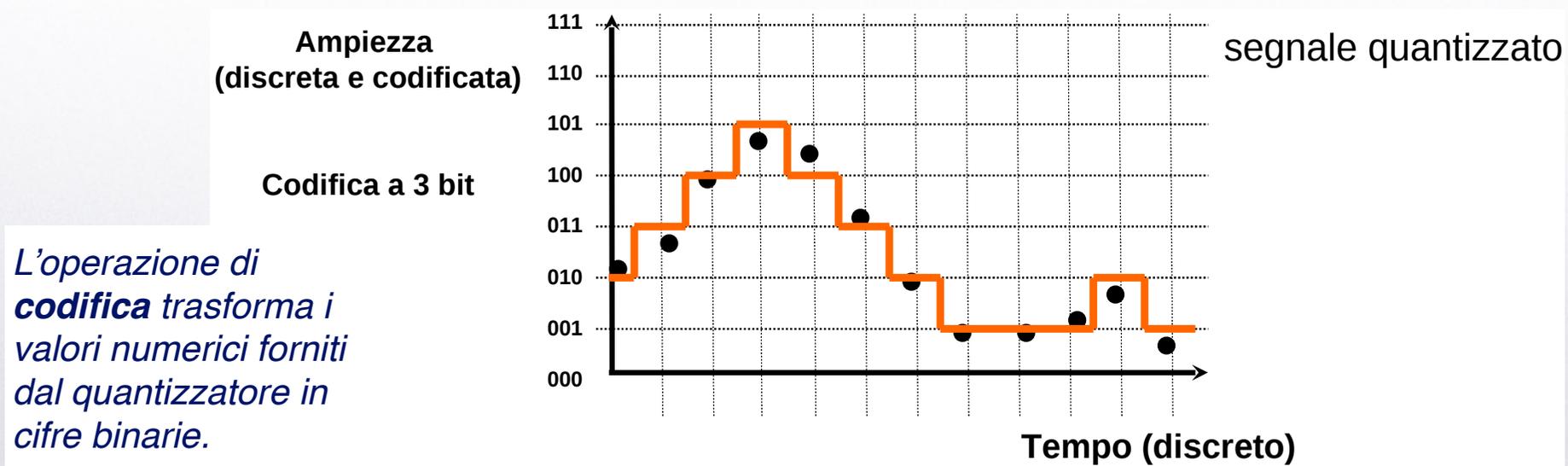




Quantizzazione/Codifica

- ▶ In natura, la maggior parte delle grandezze possono assumere un insieme infinito e continuo di valori.
- ▶ Affinché una grandezza sia trasmissibile e codificabile con un numero finito di bit, è però necessario far sì che possa assumere solo un numero finito di valori discreti; ciò avviene tramite la **quantizzazione**.
- ▶ Un **quantizzatore** associa ad ognuno dei valori continui in ingresso il più prossimo tra quelli definiti.
- ▶ L'**errore di quantizzazione** è definito come la differenza tra due valori numerici successivi.

Esempio:





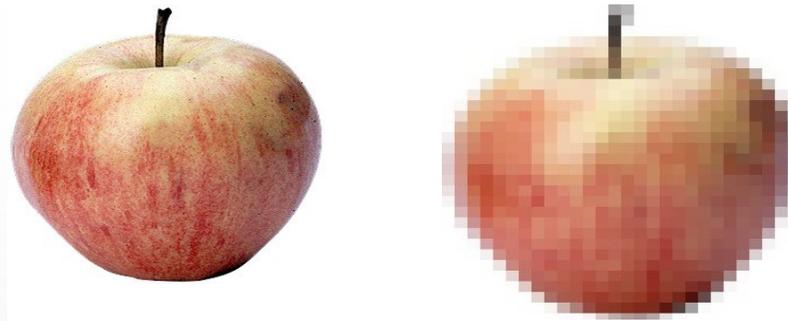
Rappresentazione delle immagini

Le forme e i colori presenti nelle immagini sono grandezze di tipo continuo.

È necessario rappresentare tale grandezze in maniera discreta.

Nel caso delle immagini non è presente la dimensione temporale (come nel suono).

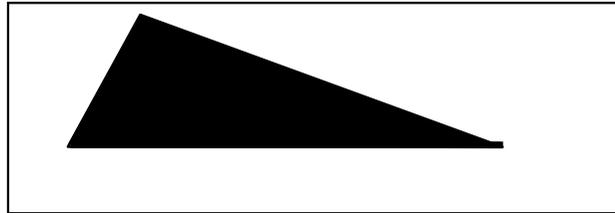
- ▶ Una maniera immediata per rappresentare un'immagine è quella di suddividerla in un insieme di piccoli quadratini (**pixel**) e di memorizzare l'informazione relativa al colore presente nel quadratino.
- ▶ Ogni quadratino rappresenta appunto un pezzettino dell'immagine.



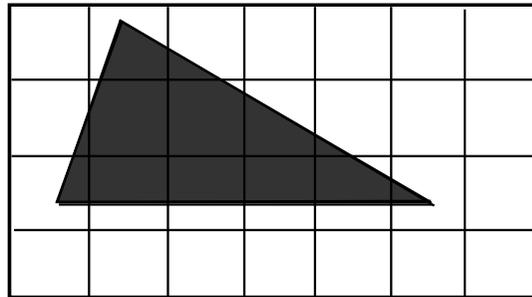
- ▶ Questo tipo di codifica delle immagini viene detto **bitmap** o **raster** (dal latino *rastrum*, rastrello) ad indicare il campionamento sistematico dell'immagine per mezzo dell'insieme di pixel che la rappresentano.

Un esempio di codifica

Caso più semplice: immagini in bianco e nero senza livelli di grigio



Suddividiamo l'immagine mediante un insieme di quadratini di dimensioni costante (campionamento dell'immagine)



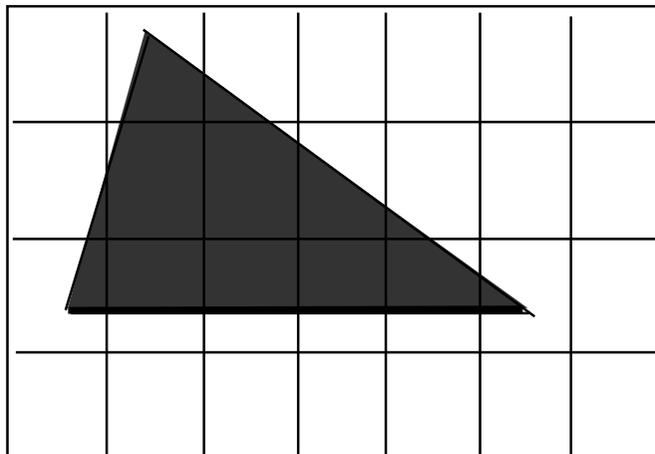
Assumiamo che un pixel sia codificato con un singolo bit che vale

0 se nel pixel il *bianco* è predominante

1 se nel pixel il *nero* è predominante

Un esempio di codifica

Poiché una sequenza di bit è lineare, è necessario definire delle convenzioni per ordinare la griglia dei pixel in una sequenza. Assumiamo che i pixel siano ordinati dal basso verso l'alto e da sinistra verso destra



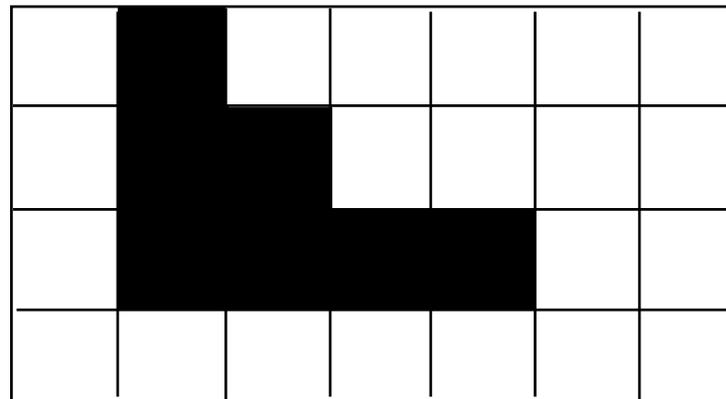
0	1	0	0	0	0	0
22	23	24	25	26	27	28
0	1	1	0	0	0	0
15	16	17	18	19	20	21
0	1	1	1	1	0	0
8	9	10	11	12	13	14
0	0	0	0	0	0	0
1	2	3	4	5	6	7

Con questa convenzione la rappresentazione della figura sarà data dalla stringa binaria

000000 0111100 0110000 0100000

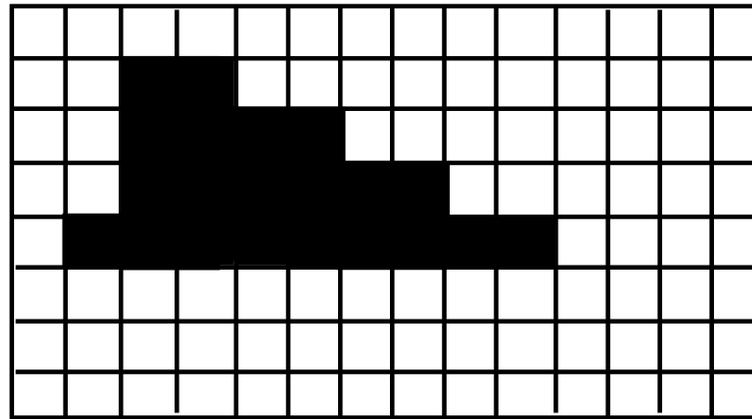
Risoluzione

- Non sempre il contorno della figura coincide con le linee della griglia. Quella che si ottiene nella codifica è un'approssimazione della figura originaria
- Se riconvertiamo la stringa `000000011110001100000100000` in immagine otteniamo



Risoluzione

La rappresentazione sarà più fedele all'aumentare del numero di pixel, ossia al diminuire delle dimensioni dei quadratini della griglia in cui è suddivisa l'immagine:



Il numero di pixel in cui è suddivisa un'immagine si chiama **risoluzione** e si esprime con una coppia di numeri ad es. 640 × 480 pixel (orizzontali × per verticali)

Codifica dei livelli di grigio

Per ogni pixel viene misurato il livello medio di intensità luminosa (il **livello di grigio**): ogni pixel è codificato con un numero di bit > 1 .

Ad esempio:

- se utilizziamo quattro bit possiamo rappresentare $2^4=16$ livelli di grigio
- se utilizziamo otto bit ne possiamo distinguere $2^8=256$, etc.



Codifica di immagini a colori

- ▶ Analogamente possiamo codificare le immagini a colori. In questo caso si tratta di individuare un certo numero di sfumature di colore differenti e di codificare ogni sfumatura mediante un'opportuna sequenza di bit.
- ▶ Ad esempio, i monitor utilizzano **risoluzioni** di 640×480 (VGA), 1024×768 (XGA), oppure 1280×1024 (SXGA) ed un numero di colori per pixel che va da 256 (8 bit) fino a sedici milioni (24 bit).
- ▶ Il numero di bit usato per codificare i colori è detto **profondità** dell'immagine.



Codifica di immagini a colori

La rappresentazione di un'immagine mediante la codifica dei pixel, viene chiamata **codifica bitmap**.

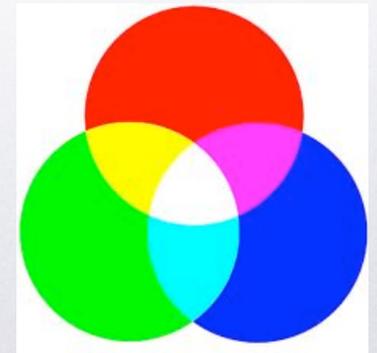
Il numero di byte richiesti dipende dalla risoluzione e dal numero di colori che ogni pixel può assumere.

Per distinguere 256 colori sono necessari otto bit per la codifica di ciascun pixel: la codifica di un'immagine formata da 640×480 pixel richiederà 2.457.600 bit (307.200 byte).



La codifica RGB

- ▶ Come è possibile rappresentare l'infinità di colori presenti in natura?
- ▶ Un possibile modello di rappresentazione è noto con il nome di **RGB** (Red, Green, Blue), il quale usa questi tre colori primari per rappresentare tutti i possibili colori.
- ▶ Nella codifica RGB ogni pixel è rappresentato da una combinazione di tre numeri, ognuno rappresentante una diversa gradazione di uno dei colori primari.
- ▶ Con 8 bit per colore otteniamo: $256 \times 256 \times 256 = 16.777.216$ colori diversi.
- ▶ Per ogni pixel sono quindi richiesti 3 byte (24 bit di *profondità*).



Esempio di codifica RGB

La sfumatura di azzurro è determinata da una certa combinazione di RGB.

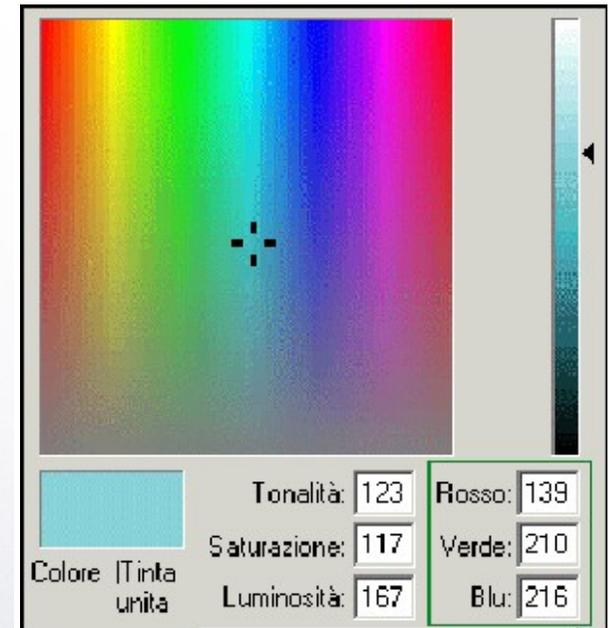
Le sequenze di bit sono espresse in base decimale.

Ogni sfumatura di colore primario è rappresentabile da 1 byte:

$$139_{10} \Rightarrow 10001011_2$$

$$210_{10} \Rightarrow 11010010_2$$

$$216_{10} \Rightarrow 10100111_2$$



Altri esempi di codifica RGB



R	G	B
0	132	209



R	G	B
102	0	102



R	G	B
125	166	71

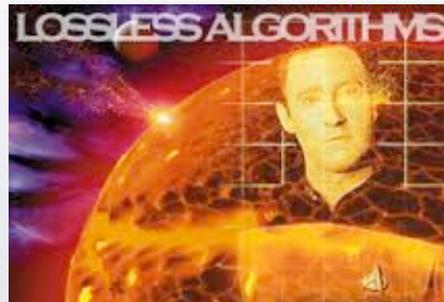


Compressione delle immagini

Tecniche di compressione

Esistono tecniche di compressione per ridurre lo spazio occupato dalle immagini. Queste tecniche sfruttano le regolarità delle immagini.

- ❖ **Compressione senza perdita di informazione** (sono dette **loss-less**): si memorizzano pixel vicini identici una volta sola e si ricorda quante volte occorrono nell'immagine.
- ❖ **Compressione con perdita di informazione** (sono dette **lossy**): non si memorizzano tutti i pixel, ma solo una frazione di essi. Si usano funzioni matematiche di interpolazione per ricostruire i pixel mancanti.





La compressione loss-less

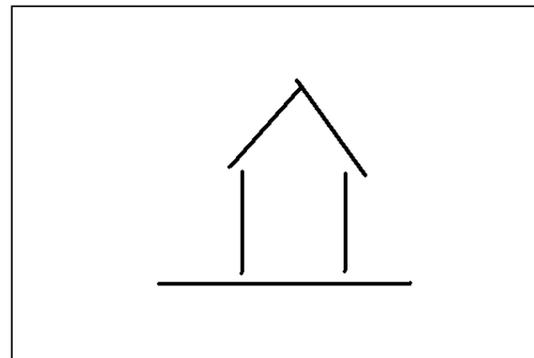
Nella compressione dei dati senza perdita si sfruttano le ripetizioni dell'informazione. Ad esempio la stringa **111111111111111111111111111111111111** può essere codificata con la stringa **1:36**.

Esempio di compressione

Immagine1.bmp (2.63 MB)



Immagine2.bmp (2.63 MB)



DOMANDA

Qual è l'immagine che può essere compressa di più?

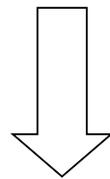


La compressione loss-less

Immagine1.bmp (2.63 MB)



Il fattore di compressione
è pari al 71%



compressione

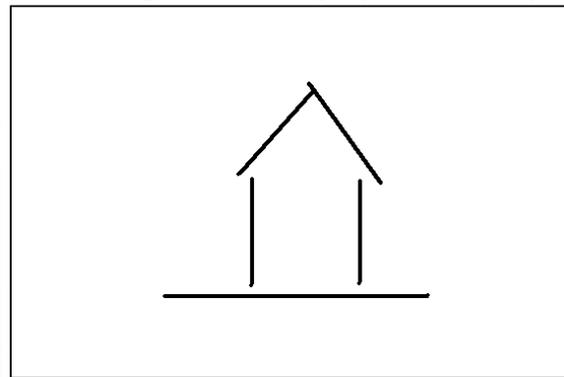
Immagine1.zip (1.876 MB)



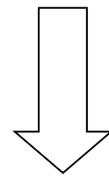


La compressione loss-less

Immagine2.bmp (2.63 MB)



Il fattore di compressione
è pari al 0.3%!



compressione

Immagine2.zip (5 KB)





Formati standard

GIF (*Graphic Interchange Format*, brevettato da Unisys) utilizza 8 bit per pixel e quindi distingue 256 colori. Usa una tecnica di compressione senza perdita (algoritmo LZW).

JPEG (*Joint Photographic Expert Group*) utilizza 24 bit, quindi 16,8 milioni di colori. Usa una tecnica sofisticata di compressione con perdita.

Altri formati senza perdita sono **PNG** (*Portable Network Graphics*, algoritmo zlib) e **TIFF** (*Tagged Image File Format*).





Formato JPEG

- ▶ Lo standard JPEG (Joint Photographic Expert Group) è aperto ed è stato sviluppato da un gruppo di esperti di fotografia ed utilizza 8 bit per pixel.
- ▶ Lo scopo dello standard è quello di comprimere immagini di tipo fotografico minimizzando la perdita di informazione al quale l'occhio umano è più sensibile.
- ▶ Quando un immagine BMP viene trasformata in una JPEG, il reticolo di pixel dell'immagine JPEG viene suddiviso in blocchi di dimensione 8x8.
- ▶ Ogni blocco 8x8 viene poi rappresentato con dei valori medi.
- ▶ Lo standard prevede diversi livelli di compressione.

Maggiore è la compressione, minore sarà la qualità dell'immagine



Formato GIF

- ▶ Il formato GIF (Graphic Interchange Format) riduce l'occupazione su disco di un'immagine limitando il numero di colori che compaiono in essa.
- ▶ Vengono scelti quelli più frequenti, alcune sfumature vengono perse e sostituite dalle sfumature più vicine fra quelle mantenute.
- ▶ Più si limita il numero di colori più l'immagine sarà piccola; il numero può andare da un minimo di 2 ad un massimo di 256.
- ▶ L'insieme dei colori utilizzati viene salvato insieme all'immagine come *palette* di colori.
- ▶ Il formato GIF è adatto ad immagini geometriche, possibilmente con un numero di colori non elevato.



La tavolozza dei colori

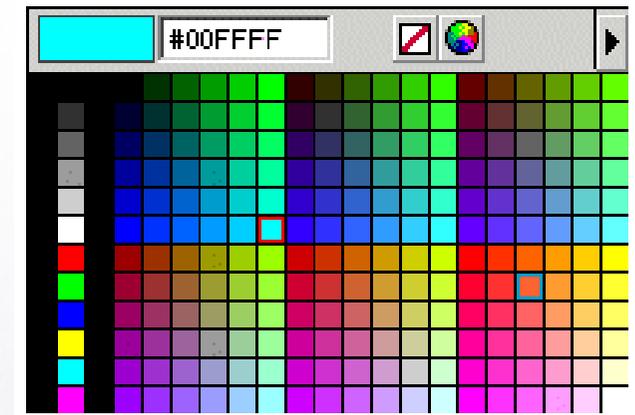
Nella maggior parte delle immagini sono presenti un numero ridotto di colori.

Questo fatto può essere sfruttato per costruire una tavolozza dei colori (*colour palette*).

Questa tavolozza non è altro che un elenco dei colori presenti nell'immagine.

Questo può portare ad un notevole risparmio di spazio.

Ad esempio: se la tavolozza contiene 256 colori, posso utilizzare un byte per codificare i colori, ottenendo così un fattore di compressione pari a 3.





Formati standard: un esempio



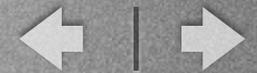
Formato gif 93K



Formato jpeg 30K



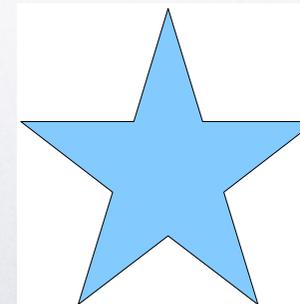
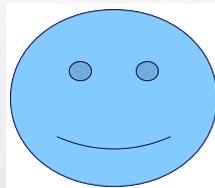
Grafica vettoriale



- ▶ Nella grafica vettoriale le immagini sono rappresentate per mezzo di primitive geometriche.
- ▶ Esempi di queste primitive sono: segmenti, curve, poligoni, cerchi, ellissi.
- ▶ Queste primitive sono a loro volta rappresentate per mezzo di equazioni matematiche.

Esempio

- Un oggetto di tipo circolare può essere rappresentato specificando le coordinate del centro e il suo raggio.
- Ogni oggetto possiede inoltre un insieme di attributi che ne specificano il colore, il riempimento, il bordo etc.



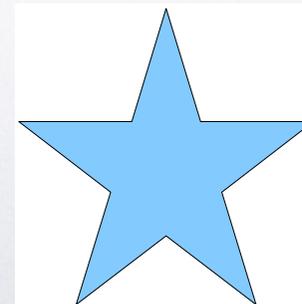
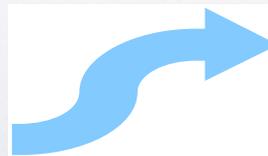
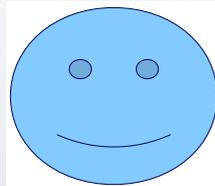


Vantaggi

- Gli oggetti presenti in un'immagine possono essere modificati (ruotati, allungati, etc.) in maniera immediata per mezzo di semplici operatori matematici.
- Un immagine può essere rappresentata in maniera compatta specificando solo le proprietà degli oggetti presenti.
- È indipendente dalla risoluzione.

Svantaggi

- La visualizzazione delle immagini richiede l'uso di software complessi in grado di risolvere le equazioni associate alle primitive per tracciare i punti (pixel) che le soddisfano.
- Per immagini molto complesse può essere necessario l'uso di computer molto potenti.





Formati vettoriali

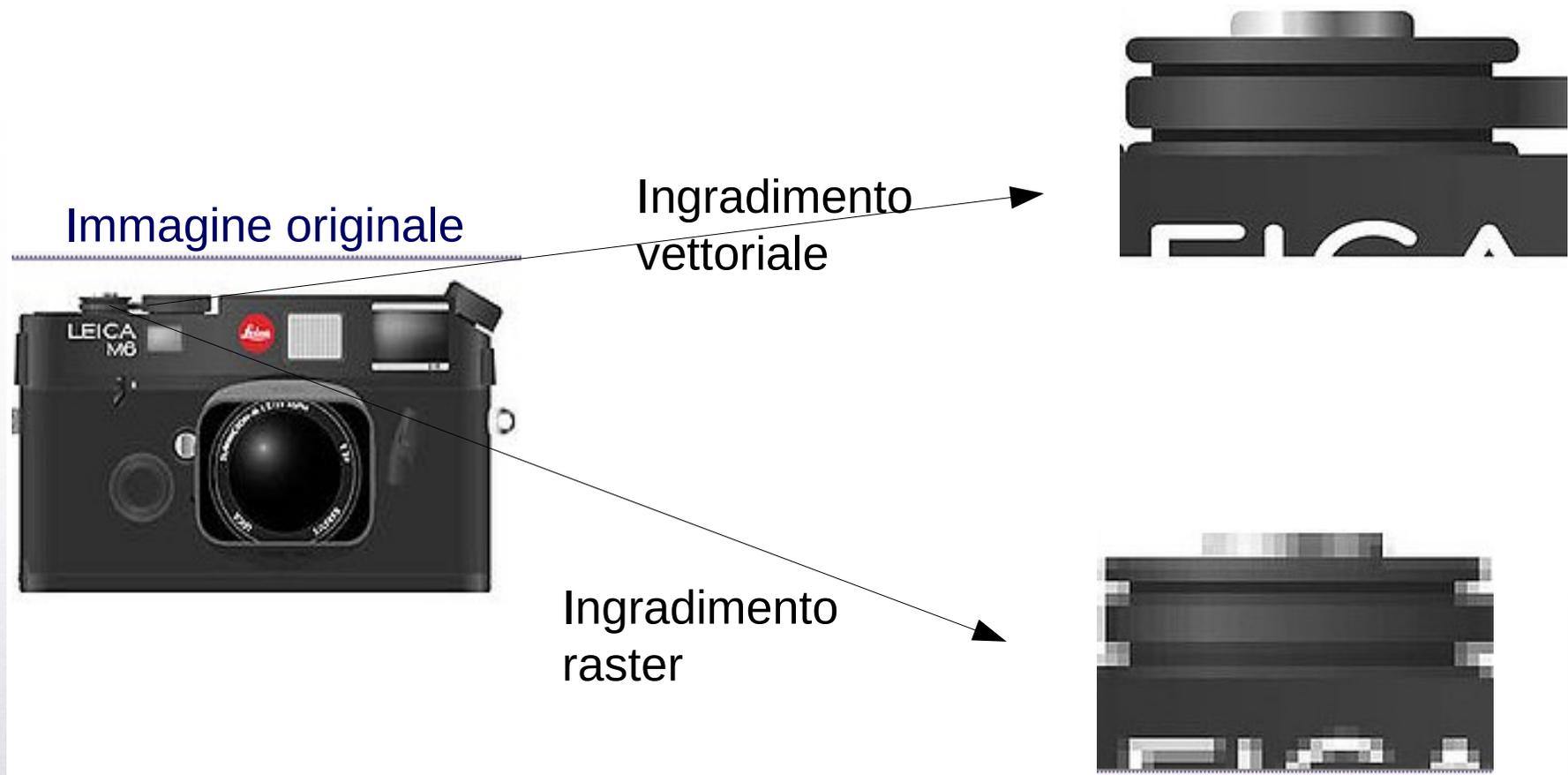
- ▶ **CDR** (il formato di CorelDraw)
- ▶ **SWF** (ShockWave Flash, il formato di Macromedia Flash, utilizzato per la creazione di animazioni destinate al web. Richiede il plug-in Flash Player; in via di sostituzione con lo standard HTML5)
- ▶ **SVG** (Scalable Vector Graphics, formato standard del W3 Consortium, creato nel 1999, è visualizzabile dai browsers ma necessita del plug-in Adobe SVG Viewer. Come il formato di Flash, permette di creare delle animazioni. Si basa sul linguaggio XML)
- ▶ **DWG** (Drawing, formato Autocad di Autodesk)
- ▶ **WMF** (Windows Meta File, il formato degli oggetti Clipart di Office)

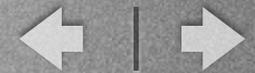
Formati ibridi (possono essere usati sia per le immagini vettoriali che raster):

- ▶ **EPS** (Encapsulated Postscript, utilizzato nel campo della stampa professionale)
- ▶ **PDF** (Portable Document Format. E' il formato visualizzabile con Adobe Acrobat Reader)
- ▶ **PSD** (il formato di Adobe Photoshop)



Confronto vettoriale-raster



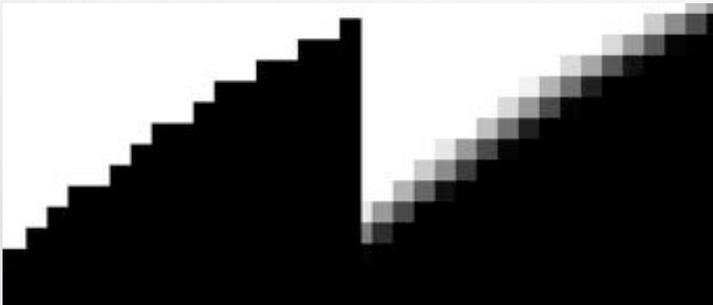


Effetti



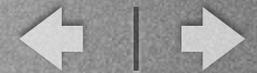
► Il **ray tracing** è una particolare tecnica di ombreggiatura.

► Il **texture-mapping** è un processo che consiste nell'applicare ad una superficie vettoriale un'immagine bitmap.



► L'**anti-aliasing** ammorbidisce i contorni di un oggetto (aliasing = scalettatura).





Font scalabili

Tipi di carattere e font

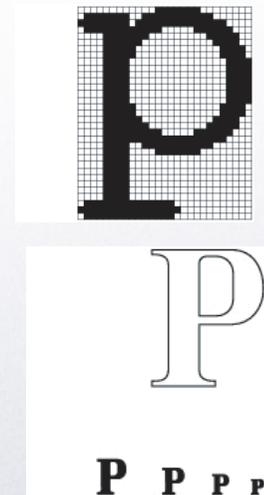
Un font è un insieme di caratteri e simboli creati in base a un particolare modello (tipo di carattere). La scelta dei tipi di carattere più consoni alle varie parti di un documento facilita la lettura di quest'ultimo.

Font bitmap e font scalabili

Le stampanti utilizzano sia font scalabili che font bitmap.

I font bitmap vengono memorizzati come modelli di bit predefiniti che rappresentano un tipo di carattere con dimensione, stile e risoluzione specifici.

I font scalabili (chiamati anche font outline) sono gestiti da un programma detto "rasterizzatore" e sono memorizzati come programmi (serie di formule) che definiscono i profili dei caratteri di un font. Ogni volta che si stampano dei caratteri di un font scalabile, la stampante crea una bitmap dei caratteri con la dimensione in punti selezionata e la salva temporaneamente nella memoria.



Optical Character Recognition (OCR)

- I sistemi di riconoscimento ottico dei caratteri, detti anche OCR (Optical Character Recognition) sono programmi dedicati alla conversione di un'immagine contenente testo, solitamente acquisite tramite scanner, in testo digitale modificabile con un normale editor.
- Il testo può essere convertito in formato ASCII semplice, Unicode o, nel caso dei sistemi più avanzati, in un formato contenente anche l'impaginazione del documento.
- L'OCR è un campo di ricerca dell'intelligenza artificiale, della visione artificiale e del pattern recognition, legati al riconoscimento delle immagini.





Formati audio



Digitalizzazione audio

- ▶ Un suono digitalizzato con qualità **CD-DA** viene campionato a 44,1 KHz su 16 bit (2 canali stereo).
- ▶ I formati audio si dividono in **lossy** (con perdita) e **no lossy** (o **lossless**, senza perdita di qualità).
- ▶ L'idea dei formati lossy (i più diffusi tra gli utenti comuni) nasce dall'idea che non tutti i suoni presenti nei 44.100 campioni al secondo contenuti in un file WAV standard vengono correttamente percepiti dall'orecchio umano.





Digitalizzazione audio

- ▶ In questo modo (così come avviene per il JPEG delle immagini) si vanno a tagliare le alte frequenze, che si ritiene siano meno distinte dal nostro orecchio.
- ▶ Più si tagliano queste frequenze, più lo spazio occupato dalla nostra traccia diminuisce ma diminuisce anche la qualità del risultato.
- ▶ I formati senza perdita (**lossless**, come ad es. WAV) invece cercano di diminuire lo spazio occupato dalla traccia senza andare a toccare il suono; la percentuale di compressione sarà decisamente inferiore rispetto ai lossy, ma non ci sarà perdita di qualità.
- ▶ Se riconvertito da WAV (magari per essere elaborato), il suono sarà **identico** all'originale.



Digitalizzazione audio

- Lo standard **MIDI** (Musical Instrument Digital Interface) viene utilizzato per uniformare le comunicazioni tra gli elaboratori e gli strumenti musicali elettronici.
- **WAV** (contrazione di WAVEform) è un formato audio sviluppato da Microsoft e IBM per personal computer IBM compatibili:
 - È una variante del formato **RIFF**^(*) di memorizzazione dei dati.
 - I dati vengono salvati in "chunk" (blocchi).
 - È simile anche al formato **AIFF**⁽⁺⁾ utilizzato dai computer Apple Mac.



(*) RIFF = Resource Image File Format, una specifica Microsoft per la memorizzazione di file multimediali.

(+) AIFF = Audio Interchange File Format

Formato MP3

- Il formato **MP3** usa una compressione con perdita di informazioni.
- **MP3** (noto anche come *MPEG Audio Layer III*) è un algoritmo di compressione audio di tipo lossy, sviluppato dal gruppo MPEG (Motion Picture Expert Group), in grado di ridurre drasticamente la quantità di dati richiesti per memorizzare un suono, rimanendo comunque una riproduzione accettabilmente fedele del file originale non compresso.
- Il *bitrate* per un file in formato MP3 indica il numero medio di bit per un secondo di dati audio. Tipicamente minimo 128 Kbps per una qualità accettabile.





Formato AAC

- Il formato **Advanced Audio Coding (AAC)** è un formato di compressione audio creato dal consorzio MPEG e incluso ufficialmente nell'MPEG-4. L'AAC fornisce una qualità audio superiore al formato MP3 mantenendo la stessa dimensione di compressione.
- Attualmente viene utilizzato principalmente da **Apple** nei suoi prodotti dedicati all'audio (*iTunes*); difatti Apple usa sia una variante dell'AAC che gestisce i diritti d'autore DRM (AAC Protected), con compressione a 128 Kbps, sia una versione senza protezione (AAC Plus), con compressione a 256 Kbps.
- L'AAC è diventato recentemente il formato standard di audio per le console Sony PlayStation, Nintendo DSi e Wii.





Formati AC3, Real

AC3. E' il formato audio usato dai DVD. In genere lo troviamo a 384 kbps (e 6 canali), ma è possibile anche averlo in soli 2 canali e a bitrate inferiori. Uno dei programmi gratuiti che lo supporta (e che consente per esempio di ridurre il bitrate) è *BeLight*.

Real Player (.ra, .rm e altre estensioni simili). Codec molto usato in certi video e in certi audio. RealPlayer è un programma gratuito (con certi limiti...) che consente di eseguire tracce audio e video create con il suo codec.

Una valida alternativa, *davvero* gratuita, a Real Player è rappresentata dal programma *Real Alternative* (solo Windows) o - molto meglio - **VLC** (Video Lan Controller), un player ed un framework multimediale multi-piattaforma gratuito e open-source che riproduce moltissimi file multimediali così come DVD, CD audio, VCD e diversi protocolli di trasmissione.





Immagini in movimento



- Memorizzazione mediante sequenze di **fotogrammi**.
- La qualità della memorizzazione dipende dal **numero di fotogrammi al secondo**.

Esempio:

- le immagini televisive vengono trasmesse con 25 fotogrammi al secondo
- poniamo di avere una risoluzione di 1024×768, in formato JPEG
- se ogni immagine è 200 Kbyte, dieci minuti di filmato occupano 3 Gbyte





Compressione immagini in movimento

Problema dell'occupazione di spazio: per ottimizzare lo spazio non si memorizzano tutti i fotogrammi.

I fotogrammi variano in modo continuo: si memorizza un primo fotogramma in modo completo, e per i successivi N solo le differenze con il primo (compressione **inter-fotogramma**).

Anche il singolo fotogramma completo viene compresso (come un'immagine) per ridurre l'occupazione (compressione **intra-fotogramma**).

I film su DVD usano la compressione MPEG-2.





Codec video

Il processo di compressione consiste nell'applicare un algoritmo al video di origine allo scopo di creare un file compresso pronto per la trasmissione o la memorizzazione.

Al momento della riproduzione del file compresso, viene invece applicato un algoritmo inverso che genera un video contenente praticamente lo stesso contenuto del video originale.

Il tempo richiesto per comprimere, trasmettere, decomprimere e visualizzare un file rappresenta la cosiddetta **latenza**. Più avanzato è l'algoritmo di compressione, più alta è la latenza.

Una coppia di algoritmi utilizzata contemporaneamente rappresenta un **codec** (codificatore/decodificatore) **video**.

I codec video di standard diversi non sono generalmente compatibili tra loro, il che significa che il video compresso con uno standard non può essere normalmente decompresso usando un altro standard.



Ad esempio, un decodificatore che supporta lo standard MPEG-4 non può essere utilizzato con un codificatore che supporta lo standard H.264, semplicemente perché uno dei due algoritmi non è in grado di decodificare l'output dell'altro algoritmo correttamente; tuttavia, è possibile implementare più algoritmi nello stesso software o hardware e consentire la coesistenza di più formati.



Confronto tra compressione delle immagini e compressione video

- ▶ I vari standard di compressione utilizzano metodi diversi per ridurre i dati e offrono, quindi, velocità di trasmissione in bit, qualità e latenze diverse.
- ▶ Gli algoritmi di compressione si suddividono in due tipi: compressione delle immagini e compressione video.
- ▶ La compressione delle immagini utilizza la tecnologia di codifica intra-fotogramma. I dati vengono ridotti all'interno di un fotogramma immagine semplicemente rimuovendo le informazioni non necessarie che potrebbero essere non visibili all'occhio umano.
- ▶ Motion JPEG (**MPEG**) è un tipico esempio di standard di compressione di questo tipo. In una sequenza Motion JPEG le immagini sono codificate o compresse come singole immagini JPEG.

Nel formato Motion JPEG le tre immagini della sequenza mostrata vengono codificate e trasmesse come immagini univoche distinte (fotogrammi di tipo I) senza dipendenze tra loro.





Compressione inter-fotogramma

Gli algoritmi di compressione di video, quali MPEG-4 e H.264, utilizzano la predizione inter-fotogramma per ridurre i dati video tra una serie di fotogrammi.

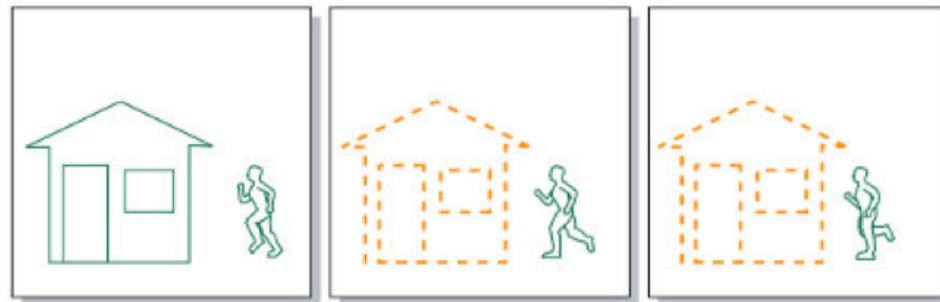
Ciò implica tecniche come ad esempio la codifica differenziale, dove ciascun fotogramma viene confrontato con quello di riferimento e vengono codificati solo i pixel modificati rispetto al fotogramma di riferimento.

Pertanto, il numero dei valori dei pixel codificati e trasmessi risulta significativamente ridotto. Quando si visualizza una sequenza codificata di questo tipo, le immagini vengono riprodotte come nella sequenza video originale.

Nella codifica differenziale, viene codificata interamente solo la prima immagine (fotogramma di tipo I). Per le due immagini successive (fotogrammi di tipo P), vengono stabiliti dei collegamenti agli elementi statici della prima immagine, come la casa. Solo gli oggetti in movimento, come l'uomo che corre, vengono codificati usando vettori di movimento, in modo da ridurre la quantità di informazioni trasmesse e memorizzate.

— Trasmesso

- - Non trasmesso





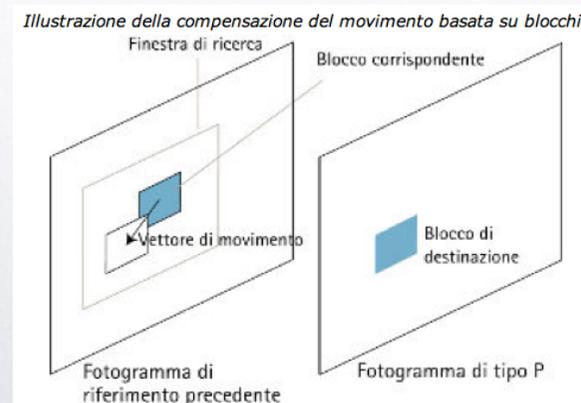
Compensazione del movimento basata su blocchi

Per ridurre ulteriormente i dati, è possibile utilizzare altre tecniche come la **compensazione del movimento basata su blocchi**.

Questa tecnica individua la parte di un nuovo fotogramma di una sequenza video che corrisponde a quella di un fotogramma precedente, anche se riferito a una posizione diversa, e divide il fotogramma in una serie di macroblocchi (blocchi di pixel).

Ciò consente di creare o “prevedere”, blocco dopo blocco, un nuovo fotogramma, ricercando un blocco corrispondente in un fotogramma di riferimento.

Se si rileva una corrispondenza, il codificatore codifica la posizione in cui si trova il blocco corrispondente all'interno del fotogramma di riferimento. La cosiddetta codifica del vettore di movimento richiede una quantità inferiore di bit rispetto alla codifica del contenuto effettivo del blocco.





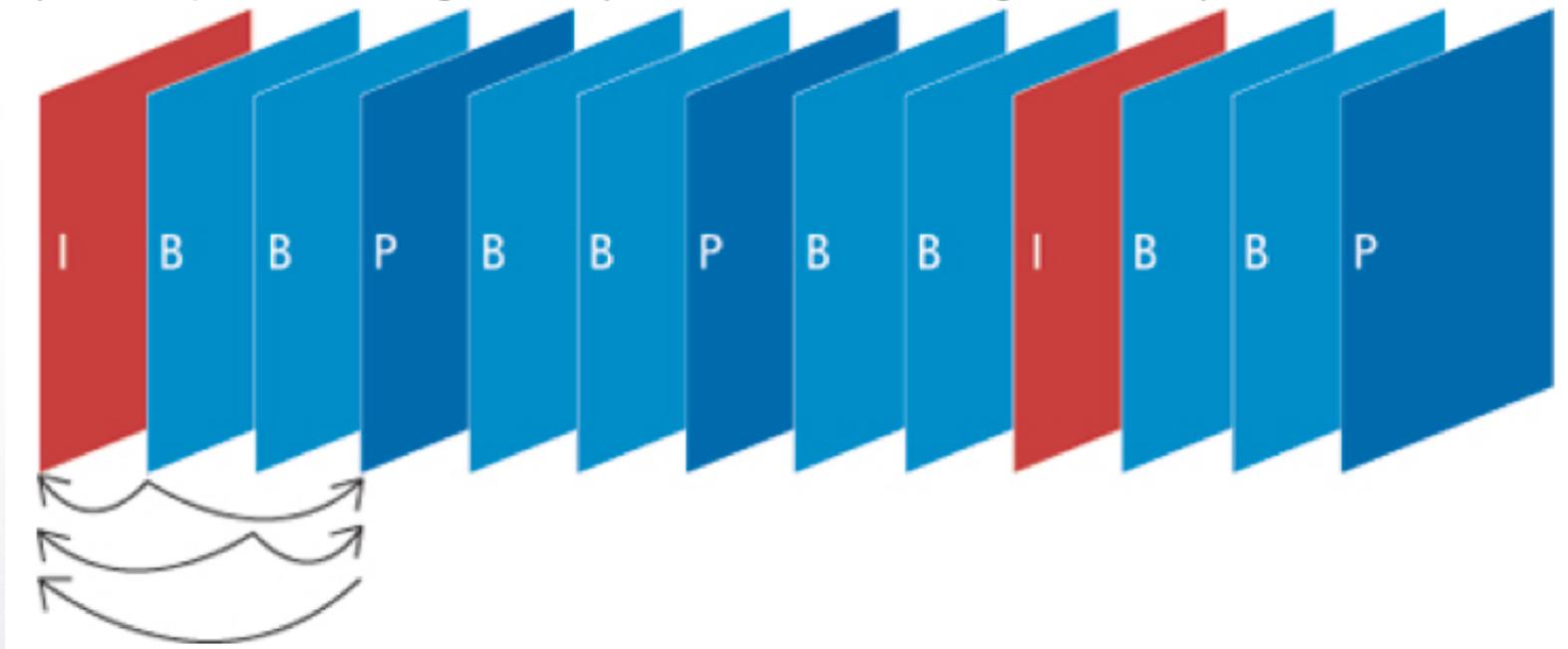
Fotogrammi di tipo I, P o B

- Con la predizione inter-fotogramma, ciascun fotogramma di una sequenza di immagini viene classificato come un determinato tipo di fotogramma, ad esempio I, P o B.
- Un **fotogramma di tipo I** o intra-fotogramma è un fotogramma a sé stante che può essere decodificato in modo indipendente senza fare riferimento ad altre immagini. La prima immagine di una sequenza video è sempre rappresentata da un fotogramma di tipo I. Questo tipo di fotogrammi viene usato come punto iniziale per i nuovi visualizzatori o come punto di risincronizzazione nel caso in cui il flusso in bit trasmesso risulti danneggiato. I fotogrammi di tipo I possono essere usati anche per le funzioni di avanzamento veloce, il riavvolgimento e altre funzioni di accesso. I codificatori inseriscono automaticamente fotogrammi di tipo I a intervalli regolari o su richiesta, se il flusso video deve essere visualizzato su nuovi client. Lo svantaggio di questo tipo di fotogrammi è rappresentato dal fatto che richiedono una maggiore quantità di bit, ma non producono molti artefatti, causati da dati mancanti.
- I **fotogrammi di tipo P**, che sta per inter-fotogramma “predittivo”, fa riferimento al cambiamento a sezioni di fotogrammi I e/o P precedenti per codificare il fotogramma. Questo tipo di fotogramma richiede generalmente una quantità inferiore di bit rispetto ai fotogrammi di tipo I, ma in compenso è molto sensibile agli errori di trasmissione a causa della stretta dipendenza dai fotogrammi P e/o I precedenti.
- Un **fotogramma di tipo B**, o inter-fotogramma “bi-predittivo”, fa riferimento sia a un fotogramma precedente che a un fotogramma successivo. L’uso di fotogrammi di tipo B aumenta la latenza.



Fotogrammi di tipo I, P o B

Sequenza tipica con fotogrammi I, B e P. Un fotogramma P può fare riferimento solo a fotogrammi I o P precedenti, mentre un fotogramma B può fare riferimento a fotogrammi I o P precedenti e successivi.



Quando un decodificatore video ripristina il video decodificando il flusso di bit un fotogramma alla volta, la **decodifica** deve sempre iniziare con un fotogramma di tipo I.

I fotogrammi di tipo P e B, se utilizzati, devono essere decodificati insieme ai fotogrammi di riferimento.



Grazie per l'attenzione!