# Rappresentazione dell'informazione

Notazione posizionale: valore cifra dipendente dalla posizione (sistema decimale)

#### Numeri naturali

b (base)	2	8	10	16
insieme S	0, 1	0,1,2,3,4,5,6,7	0,1,2,3,4,	0,1,2,3,4,5,6,7,8,9,
$c_{\dot{1}} \in [0, b-1]$			5,6,7,8,9	A,B,C,D,E,F

numero N (posizionale) =  $c_{n-1} c_{n-2} \dots c_0$  con  $c_i \in S$ 

# Interpretazione in base decimale di un numero in base b

$$N_{10} = \sum_{i=0}^{n-1} c_i b^i$$
, dove n= numero cifre

Es. 
$$N_2 = 101_2 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 5_{10}$$
  
 $N_8 = 101_8 = 1 \cdot 8^2 + 0 \cdot 8^1 + 1 \cdot 8^0 = 65_{10}$   
 $N_{10} = 101_{10} = 1 \cdot 10^2 + 0 \cdot 10^1 + 1 \cdot 10^0 = 101_{10}$ 

#### Conversione numero naturale da base 10 a base 2

- 1. Numero iniziale diviso (divisione intera) ripetutamente per la base 2 fino a quando il risultato è 0;
- 2. i resti prodotti dalle divisioni intere presi in ordine inverso costituiscono la codifica cercata.

Osservazione: l'algoritmo converge sempre





Esempio:  $N_{10} = 26$ resti base base 8 0 - significativo 26 |  $26 \mid 2$ 13 |  $3 \mid 3$ fine 6 ()3 1+ significativo ←fine  $26_{10} = 11010_2 = 32_8 = 1A_{16}$ 

Conversioni utili

Binario 
$$\leftarrow \rightarrow$$
 ottale  $26_{10} = (0)11.010 = 32$ 

Binario 
$$\longleftrightarrow$$
 esadecimale  $26_{10}$ = (000)1.1010 = 1 A

# Rappresentazione dei numeri naturali nel calcolatore

L'unità atomica di rappresentazione è la cifra binaria (base 2) chiamata bit (BInary digiT).

Un numero naturale binario è memorizzato in una parola di memoria composta da N bit (cifre).

 $\downarrow \downarrow$ 

Impossibile rappresentare tutti i numeri naturali:  $\infty$  richiede  $\infty$  bit

La rappresentazione finita riduce quindi l'intervallo rappresentabile





#### Intervallo di rappresentabilità

Dati N bit quanti e quali numeri naturali si possono rappresentare?

2<sup>N</sup> combinazioni, ossia tutti i numeri nell'intervallo [0, 2<sup>N</sup>-1]

N	2 <sup>N</sup>
10	1024 (1K)
16	65536(64K)
20	≅1.000.000(M)
32	≅ 4 miliardi (G)

Corrispondenza tra combinazione binaria e numero per N=16

 $0000\ 0000\ 0000\ 0000\ = 0$ 

 $0000\ 0000\ 0000\ 0001\ = 1$ 

 $0000\ 0000\ 0000\ 0010\ = 2$ 

...

 $0000\ 0000\ 0001\ 1010\ = 26$ 

 $1111\ 1111\ 1111\ 1111\ = 65535$ 

#### Osservazioni:

- l'intervallo di rappresentabilità è finito
- tutti i numeri all'interno dell'intervallo sono rappresentabili

		Aritmetica	binaria		
Somma		Sottrazione		Prodott	0
00001 ←	riporti	00100 ←	prestiti	010 x	2 x
011001 +	25+	101001 -	41 -	101 =	5 =
100001 =	33=	100101 =	37 =		
				010	
111010	58	000100	4	000	
				010	
Αι	ımenta il	num. di bit nec	essari ->	01010	10





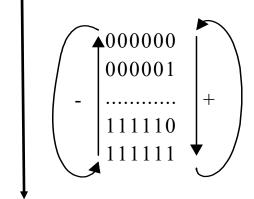
#### Osservazioni

- possibile produzione di valori non rappresentabili (es. 65535+1)
- Ogni operazione produce sempre un risultato

(1) 0000 0000 0000 0000

Riporto (carry) -> overflow

Circolarità nel calcolo sommatore



Registro di stato della CPUxx (SR)

ZNCV

Z = 1 se addizione = 0

N=1 se risultato addizione è negativo

C=1 se riporto sul bit + significativo in addizione (carry)

V=1 se overflow

Dimostrazione In limits.h unsigned short int (max=65535)

unsigned short int v=0;

• • • • •

while (1) {v++; if (v==0){printf("\n v=0");break; } else printf("\n%d", v); }





```
Exec

1

2

3

...

65534

65535

v=0

Es2. In limits.h short int (min=-32768, max=32767)

short int v=0;
while (1)
{v++; printf("\n%d", v);
if(v<=0) { sleep (1);}
if (v==0){printf("\n v=0");break; }
}
```

#### Exec

1	0000 0000 0000 0001
2	0000 0000 0000 0010
•••	
32767	0111 1111 1111
-32768	1000 0000 0000 0000
-32767	1000 0000 0000 0001
•••	
-1	1111 1111 1111

Errore assoluto è costante < 1 Errore relativo che diminuisce al crescere dei numeri.





## Rappresentazione dei numeri interi

L'intervallo di rappresentabilità viene suddiviso tra numeri positivi e negativi

Esempi di intervalli f(hardware, linguaggio) Ogni linguaggio, compilatore decide come gestire questi limiti.

Tipi C	Bit	Intervallo possibile (limits.h)
short integer	16	-32768 a 32767
		SHRT_MIN, SHRT_MAX
unsigned short integer	16	0 a 65535 USHRT_MAX
Integer	32	-2147483648 a 2147483647
		INT_MIN, INT_MAX
unsigned integer	32	0 a 4294967295 UINT_MAX
long integer	32	-2147483648 a 2147483647
		LONG_MIN, LONG_MAX
unsigned long integer	32	0 a 4294967295 ULONG_MAX

Convenzione codifica in modulo e segno (N bits)

- bit + significativo: 1 (negativi), 0 (positivi)
- valori rappresentabili:  $-(2^{N-1}) \longleftrightarrow (2^{N-1})$
- doppia rappresentazione dello 0

#### Esempio:

111(-3), 110(-2), 101(-1), 100(-0), 000(+0), 001(+1), 010(+2), 011(+3)

Codifica in complemento a 2

- si abbandona la regola del cambio base
- valori rappresentabili:  $-(2^{N-1}) \longleftrightarrow (2^{N-1}-1)$
- primo bit mantiene il segno

# Esempio:

100(-4), 101(-3), 110(-2), 111(-1), 000(+0), 001(+1), 010(+2), 011(+3)





Come si calcola

Numeri positivi: codifica binaria del valore assoluto

Numeri negativi:  $|N_2| + (N_2) = 0$ 

ES. -3

011 (3)  
101 (-3) 
$$\rightarrow N_2 = 2^N - |N_2|$$
  
-----  
(1)000 011  
-----  
101

Regole immediate per numeri negativi

- 1. complemento a 1 di  $+N_2$  +1 (011  $\rightarrow$ 100 +1  $\rightarrow$  101)
- 2. ricopiare bit significativi di N<sub>2</sub> sino a primo 1 incluso e poi complemento a 1 dei rimanenti.

#### Osservazioni:

- unica rappresentazione dello 0 (complemento a 2)
- sottrazione realizzata come somma

#### Torniamo all'overflow di A + B

Come vengono caricati i bit C e V

- se risultato ha primo bit a  $0 \Rightarrow$  no overflow altrimenti V=1





A>0, B<0 (o viceversa) 
$$A + (2^{N} - |B|) = 2^{N} - (|B| - A)$$
  
- V=0  
- |B| > A \rightarrow C=0  
010 + 2  
101 = -3  
----  
111 -1

- 
$$|B| \le A \to C=1$$
 corretto dopo eliminazione riporto  
 $011 + (3)$   $011 + (3)$   
 $101 = (-3)$   $110 = (-2)$   
----- (1)000 (1)001

$$A < 0, B < 0$$
  $(2^N - |A|) + (2^N - |B|) = 2^N + 2^N - (|B| + |A|)$ 

- C=1 riporto è strutturale e quindi non significativo
- se risultato ha primo bit a  $1 \Rightarrow$  no overflow altrimenti V=1

$$V=0 111 + -1 101 = -3 --- (1)100 -4$$

V=1 
$$101+(-3)$$
  
 $101=(-3)$   
 $----$   
 $010 \rightarrow +2$ 





#### Rappresentazione dei caratteri

- lettere maiuscole/minuscole A a .. 7 z
- spazio
- cifre 0 .... 9
- segni di interpunzione , ; ; .
- simboli ! " # % @ ) < =
- caratteri di controllo per gestire la visualizzazione, la stampa, la trasmissione dei caratteri (inizio riga, salto di riga, salto pagina ....)

Codifica ASCII (American Standard Code for Information Interchange) basata su 8 bit

- da 0 a 32 caratteri speciali non stampabili
- 65 A, 66 B, 67 C, ..... 97 a, 98 b ordinamento per concetto di precede/segue e distanza tra due lettere e tra la stessa lettera minuscola e maiuscola (costante)
- lettere senza accento (alfabeto inglese)
- 128 configurazioni perse codifiche nazionali nell'intervallo 128-255 ⇒ portabilità

La dualità carattere e intero

char b='z'; printf("%c %d",b,b); z 122





# **ASCII TABLE**

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Hex Char
0	0	[NOLL]	32	20	(SPACE)	64	40	0	96	09	
1	1	(START OF HEADING)	33	21	_	65	41	A	26	61	в
2	2	(START OF TEXT)	34	22		99	42	8	86	62	p
3	e	(END OF TEXT)	35	23	#	29	43	v	66	63	v
4	4	[END OF TRANSMISSION]	36	24	\$	89	44	۵	100	64	p
5	2	(ENQUIRY)	37	25	%	69	45	ш	101	65	9
9	9	[ACKNOWLEDGE]	38	56	Š	70	46	ш	102	99	+
7	7	(BELL)	39	27	-	71	47	U	103	29	6
8	80	[BACKSPACE]	40	28	_	72	48	I	104	89	, u
6	6	[HORIZONTAL TAB]	41	29	_	73	49	_	105	69	
10	A	(LINE FEED)	42	2A	*	74	4A	_	106	eA	
11	В	(VERTICAL TAB)	43	28	+	75	4B	¥	107	68	~
12	U	(FORM FEED)	44	2C		9/	4C	_	108	29	_
13	٥	[CARRIAGE RETURN]	45	2D		77	4D	Σ	109	<b>Q9</b>	E
14	ш	(SHIFT OUT)	46	2E		78	4E	z	110	99	2
15	u.	(SHIFT IN)	47	2F	_	79	4F	0	111	6F	0
16	10	(DATA LINK ESCAPE)	48	30	0	80	20	۵	112	70	d
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	0	113	71	. 6
18	12	[DEVICE CONTROL 2]	20	32	2	82	52	~	114	72	
19	13	[DEVICE CONTROL 3]	51	33	m	83	23	s	115	73	S
. 02	14	[DEVICE CONTROL 4]	52	34	4	84	54	-	116	74	t
21	15	(NEGATIVE ACKNOWLEDGE)	53	35	2	85	55	_	117	75	,
22	16	(SYNCHRONOUS IDLE)	54	36	9	98	99	>	118	9/	>
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	22	8	119	17	W
24	18	[CANCEL]	26	38	8	88	28	×	120	78	×
25	19	(END OF MEDIUM)	57	39	6	89	59	<b>&gt;</b>	121	79	Α
26	IA	(SUBSTITUTE)	58	34		06	5A	Z	122	7A	N
27	18	(ESCAPE)	59	38		91	5B	_	123	78	÷
28	10	[FILE SEPARATOR]	09	30	v	92	20	_	124	70	_
29	10	[GROUP SEPARATOR]	61	30	П	93	2D	_	125	7D	_
30	1E	(RECORD SEPARATOR)	62	3E	٨	94	SE.	<	126	7E	1
31	1F	[UNIT SEPARATOR]	63	3F	2	95	5F	ı	127	7F	[DEL]
	΄, μ							_			

Prof. Mauro Negri

Da uno standard per l'inglese e per la stampa ad uno per l'interoperabilità estesa ad altri alfabeti

ISO 8859 standard per usare ottavo bit, ma 128 carattere sono pochi se si esce dall'inglese (8859-1 sino a 8859-16 di cui 1 abbandonata) ISO 8859-1 (latin 1) Europa Occidentale

- Lingue anche parzialmente coperte: <u>danese</u>, <u>finlandese</u>, <u>francese</u>, <u>gaelico scozzese</u>, <u>inglese</u>, <u>irlandese</u>, <u>islandese</u>, <u>italiano</u>, <u>norvegese</u>, <u>olandese</u>, <u>portoghese</u>, <u>romancio</u>, <u>spagnolo</u>, <u>svedese</u> e tedesco, <u>albanese</u>, .....
- € in revisione ISO 8859-15 versione riveduta. Il corrispondente repertorio di caratteri approvato dall'IANA ISO-8859-1
- è la codifica predefinita per documenti <u>HTML</u> e per documenti trasmessi tramite messaggi <u>MIME</u>, come le risposte <u>HTTP</u> quando il tipo documento è "text" (come in "text/html").

Unicode (1991) Obiettivi

- consistentza con ASCII
- creare l'universal character set
- raggruppiamo alfabeti in european (African,...) scripts
- European scripts sono:

Armenian, Cyrillic, Greek, Latin-1 (Basic latin o ASCII) Latin 1 supplemented,.....





	000	001	002	003	004	005	006	007
0	NUL	DLE 0010	S P 0020	0030	<u>@</u>	P	0060	p
1	SOH	DC1	0021	1	A 0041	Q	a ************************************	q
2	STX 0002	DC2 0012	0022	2	B 0042	R 0052	b	r 0072
3	ETX	DC3	#	3	C	S 0053	C	S 0073
4	EOT 0004	DC4	\$	4	D 0044	T 0054	d	t 0074
5	ENQ 0005	NAK 0015	% 0025	5	E 0045	U 0055	e ************************************	u 0075
6	ACK 0006	SYN 0016	&	6	F	V	f	V 0076
7	BEL 0007	ETB 0017	0027	7	G 0047	W 0057	g	W 0077
8	BS	CAN 0018	0028	8	H 0048	X 0058	h	X 0078
9	HT 0009	E M 0019	0029	9	I 0049	Y	i	y 0079
Α	L F	SUB 001A	** 002A	003A	J 004A	Z	j	<b>Z</b>
В	V T	ESC 0018	+ 0028	• • • •	K	0058	k ************************************	{ 0078
С	FF 0000	FS	• 002C	< 0030	L	) 005C	1	0070
D	CR 880	GS 001D	0020	0030	M	] 005D	m	} ∞70
E	SO	R S 001E	002E	> 003E	N OG4E	<b>∧</b>	n	~ 007E
F	SI	U S 001F		?	O 004F	005F	O 006F	DEL 007F





	800	009	00A	00B	00C	00D	00E	00F
0	X X X	DCS 0090	NB SP 00A0	O 0080	À	Đ	à	ð
1	X X X	PU1 0091	00A1	<u>+</u>	Á	$\widetilde{\overset{\sim}{N}}_{_{0001}}$	á	$\widetilde{\mathbf{n}}_{_{\mathtt{ODF1}}}$
2	BPH 0082	PU2	¢	2 0082	Â	Ò	â	Ò 00F2
3	NBH 0083	STS 0093	£	3 0083	${\displaystyle \mathop{\widetilde{A}}_{{\scriptscriptstyle{00C3}}}}$	Ó	ã 	Ó 00F3
4	I ND 0084	CCH 0094	Ø 8084	0084	Ä	Ô	ä	Ô 00F4
5	NEL 0085	MW 0095	¥	μ	Å	Õ	å	Õ oofs
6	SSA 0086	SPA 0096	 00A6	¶ 0086	Æ	Ö	æ	Ö
7	ESA 0087	EPA 0097	§ 00.47	0087	Ç	X 00D7	Ç	• • 00F7
8	HTS 0088	SOS	•• 00A8	.5 0088	È	Ø	è	Ø oofs
9	HTJ 0089	XXX	© 8888	1	É	Ù	é	ù
Α	VTS 008A	SCI	<u>a</u>	<u>Q</u> 008A	Ê	$ m \acute{U}$	ê	Ú
В	PLD 0088	CSI 0098	≪ 00AB	>>>	Ë	$\hat{\mathbf{U}}_{\infty}$	ë	û
С	PLU 0080	ST	ODAC	1/4	Ì	$\ddot{\mathbf{U}}_{\infty}$	ì	ü
D	RI	OSC	SHY	1/2	Í	Ý	í med	ý
E	SS2 008E	PM cose	R	3/4 008E	Î	Þ	î	b
F	SS3	APC 009F	00AF	COBF	Ï oocf	ß	ï OGEF	$\mathop{\mathbf{\dot{y}}}_{\text{\tiny ODFF}}$





	010	011	012	013	014	015	016	017
0	Ā	Đ 0110	Ġ	İ 0130	1.	Ő 0150	Š	Ű
1	ā	đ	ġ	1	Ł	Ő	Š 0161	ű
2	Ă 0102	Ē	Ģ 0122	IJ 0132	ł 0142	Œ 0152	Ţ 0102	Ų 0172
3	ă 0103	ē	g 0123	ij 0133	Ń 0143	œ	ţ	ų 0173
4	<b>A</b>	Ĕ	Ĥ	Ĵ 0134	ń	Ŕ	Ť	$\hat{W}_{_{0174}}$
5	ą 0105	ĕ 0115	ĥ	ĵ 0135	Ņ 0145	ŕ	ť'	ŵ
6	Ć	Ė	Ħ 0126	Ķ 0136	ņ 0146	Ŗ	Ŧ	$\hat{Y}_{_{0176}}$
7	ć	ė 0117	ħ 0127	ķ 0137	$\check{N}_{_{0147}}$	<b>ŗ</b> 0157	ŧ	ŷ
8	Ĉ	Ę	$\displaystyle \widetilde{I}_{_{\text{0128}}}$	K 0138	ň 0148	Ř	$ ilde{ ilde{U}}_{ ext{ones}}$	Ÿ
9	ĉ	ę 0119	ĩ 0129	Ĺ	'n	ř	ũ	0178 Ź 0179
Α	Ċ	Ě	Ī 012A	Í	N 014A	Ś	$ar{U}_{_{\text{O16A}}}$	ź
В	Ċ 0108	ě	1 0128	Ļ 0138	Ŋ 0148	Ś 0158	ū 0168	017A Ż
С	Č	Ĝ	Ĭ 0120	ļ 0130	Ō 0140	Ŝ	Ŭ	
D	č	ĝ	Ĭ 0120	Ľ	ō	Ŝ	ŭ	Ż 017C Ž 017D Ž
E	Ď 010E	Ğ	Į O12E	1' 013E	014D Ŏ 014E	Ş	0160 Ů 016E	Ž
F	ď	ğ	į 012F	Ŀ	Ŏ 014F	S 015F	ů	f 017F





Rappresentazione logica del carattere

- Ad ogni carattere è associato un codepoint
- Un codepoint =<uniqueNumberValue, name>
   where

numbervalue espresso in esadecimale utilizzando prima 16 e ora 21 bits

Es., U+0041 -65dec -A

Name descrizione

Es., Latin Capital Letter A

Implementazione fisica del codepoint

Codepoint

Unicode Transformation Format (UTF biunivoca)

UTF8bits

UTF16bits

UTF32bits

#### UTF16 bit distribution

Scalar Value	UTF-16
XXXXXXXXXXXXX	XXXXXXXXXXXXX
000uuuuuxxxxxxxxxxxxxx	110110wwwwxxxxxx 110111xxxxxxxxxxx

Note: www = uuuuu - 1

#### UTF8 bit distribution

Scalar Value	First Byte	Second Byte	Third Byte	Fourth Byte
00000000 0xxxxxxx	0xxxxxxx			
00000yyy yyxxxxxx	110ууууу	10xxxxxx		
zzzzyyyy yyxxxxxx	1110zzzz	10уууууу	10xxxxxx	
000uuuuu zzzzyyyy yyxxxxxx	11110uuu	10uuzzzz	10уууууу	10xxxxxx





**UTF8** Examples

codepoint	Carattere	Codeunits
U+004D	M	4D
U+0431 (1)	б	D0 B1
U+4E8B (2)	事	E4 BA 8B

- (2)  $4E8B = 0100 \ 1110 \ 1000 \ 1011$ Schema zzzzyyyy yyxxxxxx  $\downarrow \downarrow$   $1110zzzz \ 10yyyyyy \ 10xxxxxx$   $\downarrow \downarrow$   $1110.0100 \ .1011.1010 \ 1000.1011$   $\downarrow \downarrow$   $E \ 4 \ B \ A \ 8 \ B$





# Rappresentazioni di immagini come bitmap

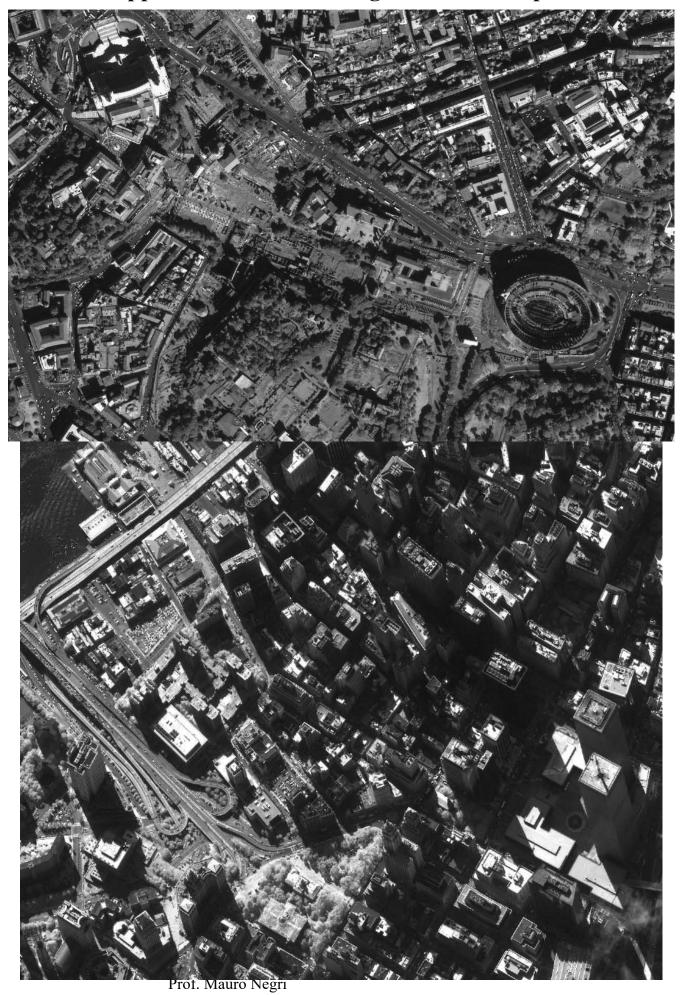


Immagine → digitalizzazione (cm x cm→ punti x punti)

↑
risoluzione dpi(dot per inch - es. 1200 per fotogrammetria)

Numero bit x punto (pixel):

- 1. bianco/nero 1 bit x pixel
- 2. scala grigi (256 toni) 8 bit x pixel
- 3. colori a video: sistema RGB (256 toni per colore) 8 bit x 3 colori (24 bit)

Dimensioni: area(pixel \* pixel)\*bit x punto.

Es

schermo colori: (1400\*1050)\*3\*8=35.280.000 bits ~ 4,4Mbytes

# Distanza di distinguibilità

2,56 cm (pollice) \* 1/dpi (ad es., 1200)  $\cong$  21 µm

Fotogrammetria: risoluzione al suolo

scala	scala media fotogrami	na risoluzione al suolo
1000	4500	21μm *4500=94500μm≅ 9,5cm
2000	7000	$21\mu m *7000=147000\mu m \cong 14,7cm$
5000	13000	$21\mu m*13000=273000\mu m\cong 27,3cm$
10000	22000	$21\mu m *22000=462000\mu m \cong 46,2cm$

#### Tecniche di compressione

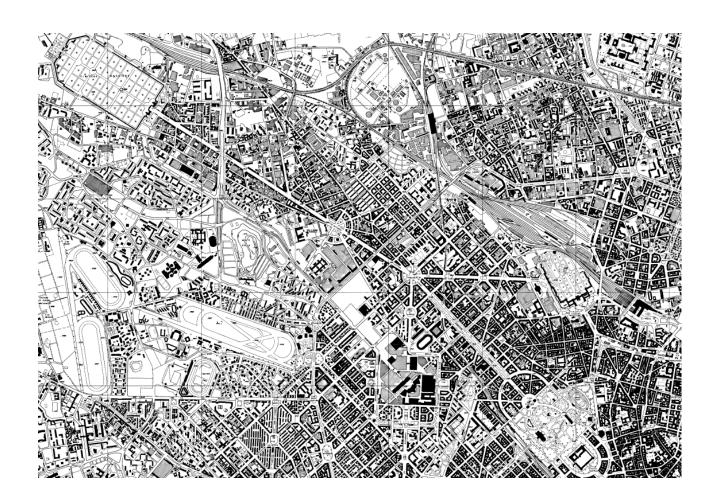
senza TIFF con perdita di informazione JPEG





# Dalle immagini alla sintesi vettoriale (cartografia)





# Dalle immagini alla sintesi vettoriale (DB spaziali)







## Rappresentazione dei video

Video: sequenza di immagini

Sistema PAL: 720x576 pixel - 25 frame/sec.

$$(720*576*3*8)*25 = (1,244MB)*25 = \sim 31MB/sec$$

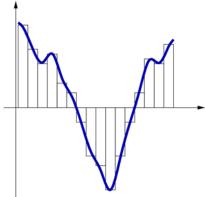
Film 133 minuti = 7980" =  $\sim 247$ GB (29 DVD SSDL-8,5GB)

Tecniche di compressione variabile MPEG-2

- variazione tra fotogramma i-esimo e (i-1)- esimo
- velocità media 3,5Mb/sec

Film 133 minuti = 7980" \*3,5=  $\sim$ 28Gb =  $\sim$ 3,5GB (1,2GB restanti per sottotitoli, lingue)  $\Rightarrow$  1 DVD SSSL-4,7GB)

#### Rappresentazione del suono



Frequenza campionamento = 2\*22.000Hz = ~44.100campioni/sec Numero bit per campione per canale: 16 Dimensione= 44.100\*16\*2= 1,4Mb/sec = 172KB/sec

Voce (freq. Camp.2\*4.000Hz)



