CMPS1134
Fundamentals of Computing

Data Storage 1

Computer Science: An Overview
Eleventh Edition
J. Glenn Brookshear
Chapter 1

Chapter 1: Data Storage

- **Bits and Their Storage**
- **Main Memory**
- **Mass Storage**
- **Representing Information as Bit Patterns**
- The Binary System
- Storing Integers
- Storing Fractions
- Data Compression
- Communications Errors

Bits and Bit Patterns

Inside today's computers information is encoded as patterns of 0s and 1s.

- **Bit:** Binary Digit (0 or 1)
- **Bit Patterns** are used to represent information.
  - Numbers
  - Text characters
  - Images
  - Sound
  - And others

Boolean Operations

- **Boolean Operation:** An operation that manipulates one or more true/false values
- Specific operations
  - AND
  - OR
  - XOR (exclusive or)
  - NOT

Figure 1.1 **The Boolean operations AND, OR, and XOR (exclusive or)**

The AND operation

0	0	1	1
AND 0	AND 1	AND 0	AND 1
0	0	0	1

The OR operation

0	0	1	1
OR 0	OR 1	OR 0	OR 1
0	1	1	1

The XOR operation

0	0	1	1
XOR 0	XOR 1	XOR 0	XOR 1
0	1	1	0

Gates

- **Gate:** A device that computes a Boolean operation
  - Often implemented as (small) electronic circuits
  - Provide the building blocks from which computers are constructed

Figure 1.2 **AND, OR, XOR, and NOT gates & their input/output values (truth table)**

AND

Inputs	Output
0 0	0
0 1	0
1 0	0
1 1	1

OR

Inputs	Output
0 0	0
0 1	1
1 0	1
1 1	1

XOR

Inputs	Output
0 0	0
0 1	1
1 0	1
1 1	0

NOT

Inputs	Output
0	1
1	0

Flip-flops

- **Flip-flop:** A circuit built from gates that can store one bit.
  - One input line is used to set its stored value to 1
  - One input line is used to set its stored value to 0
  - While both input lines are 0, the most recently stored value is preserved
- **VLSI** (Very Large Scale Integration)
  - Allows millions of electrical components to be constructed on a wafer (called a chip).
  - Used to create miniature devices that contains millions of flip-flops along with their controlling circuitry
  - Chips used as abstract tools in the construction of computer systems
  - In some cases VLSI is used to create entire computer systems on a single chip

Figure 1.3 **A simple flip-flop circuit**

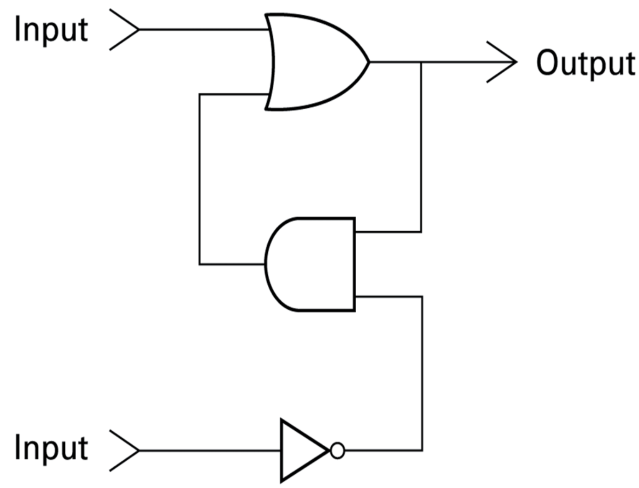


Figure 1.4 **Setting the output of a flip-flop to 1. And 0?**

a. 1 is placed on the upper input	b. This causes the output of the OR gate to be 1 and, in turn, the output of the AND gate to be 1.	c. The 1 from the AND gate keeps the OR gate from changing after the upper input returns to zero.

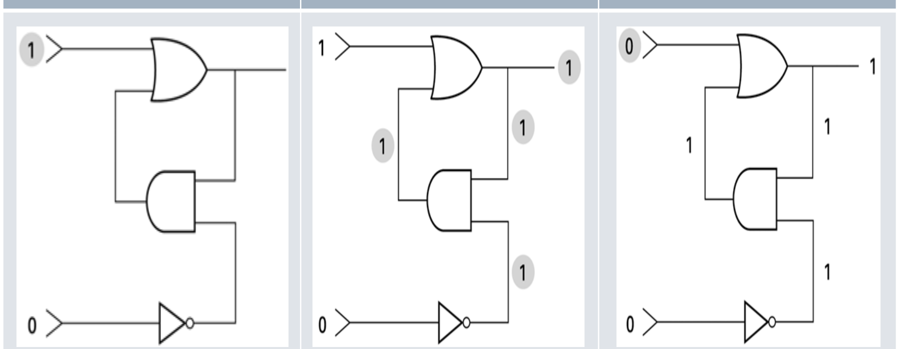
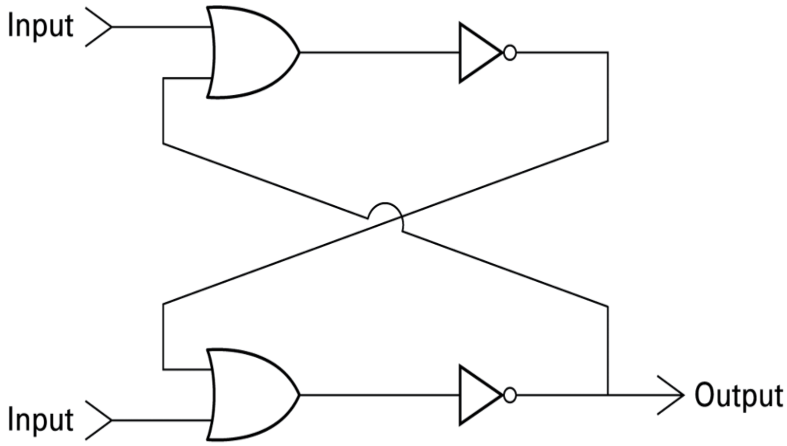


Figure 1.5 **Another way of constructing a flip-flop**



Hexadecimal Notation

A long string of bits (often called a **stream**) are difficult for the human mind to comprehend.

- **Hexadecimal notation:** A shorthand notation for long bit patterns
  - Also called base 16, or hex
  - Divides a pattern into groups of four bits each
  - Represents each group by a single symbol (0-9, A-F)
- Example:
  1010<u>0011</u> becomes A3

Bit pattern	Hexadecimal representation
0000	0
0001	1
0010	2
0011	3
0100	4
0101	5
0110	6
0111	7
1000	8
1001	9
1010	A
1011	B
1100	C
1101	D
1110	E
1111	F

Main Memory Cells

Computers contain large collections of circuits (such as flip-flops) for the purpose of storing data. This bit reservoir is known as the machine's **main memory**.

- **Cell:** A unit of main memory (typically 8 bits which is one **byte**)
  - **Most significant bit:** the bit at the left (**high-order**) end of the conceptual row of bits in a memory cell
  - **Least significant bit:** the bit at the right (**low-order**) end of the conceptual row of bits in a memory cell

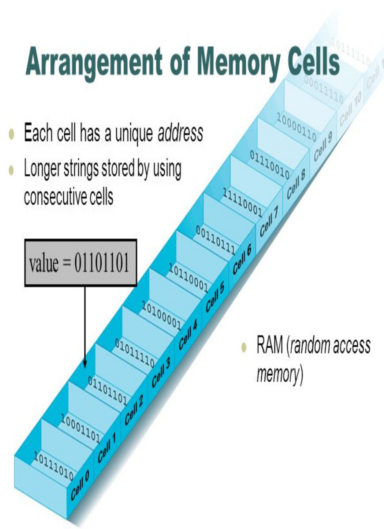
High-order end 0 1 0 1 1 0 1 0 Low-order end

Most significant bit

Least significant bit

Main Memory Addresses

- **Address:** A "name" that uniquely identifies one cell in the computer's main memory
  - The names are actually numbers.
  - These numbers are assigned consecutively starting at zero.
  - Numbering the cells in this manner associates an order with the memory cells.



Memory Terminology

- **Random Access Memory (RAM):** Memory in which individual cells can be easily accessed in any order
- Other than flip-flops, RAM in modern computers is constructed with technologies that provide greater miniaturization and faster response time.
  - They store bits as tiny electrical charges that dissipate quickly
  - They require a **refresh circuit** to repeatedly replenish the charges many times a second
  - This volatile memory is often called **dynamic memory**
  - **Dynamic Memory (DRAM)**: RAM composed of volatile memory
  - **Synchronous DRAM (SDRAM)**: DRAM that applies additional technologies to decrease the time needed to retrieve the contents from its memory cells

Measuring Memory Capacity

- **Kilobyte:**
  - 1 Kilobyte = 1024 bytes
  - Example:
    3 KB = 3 x 1024 bytes
- **Megabyte:**
  - 1 Megabyte = 1,048,576 bytes
  - Example:
    3 MB = 3 x 1,048,576 bytes
- **Gigabyte:**
  - 1 Gigabyte = 1,073,741,824 bytes
  - Example:
    3 GB = 3 x 1,073,741,824 bytes
- **Terra**, **Peta**, **Exa**, **Zetta**, **Yotta**, **Bronto**, …

Mass Storage

- On-line versus off-line
- Mass Storage Systems:
  - **Magnetic Systems** - e.g. Disk, Tape
  - **Optical Systems** – e.g. CD, DVD
  - **Solid-state** - e.g. flash drives, Secure Digital (SD) memory card, Solid-State Disks (SSD)
- Advantages over main memory
  - Less volatility
  - Larger storage capacities
  - Low cost
  - In many cases can be removed
- Major Disadvantage
  - Typically require mechanical motion which causes significantly slower storage and retrieval time than main memory where all activities are electronic

Figure 1.10 **Magnetic tape storage**

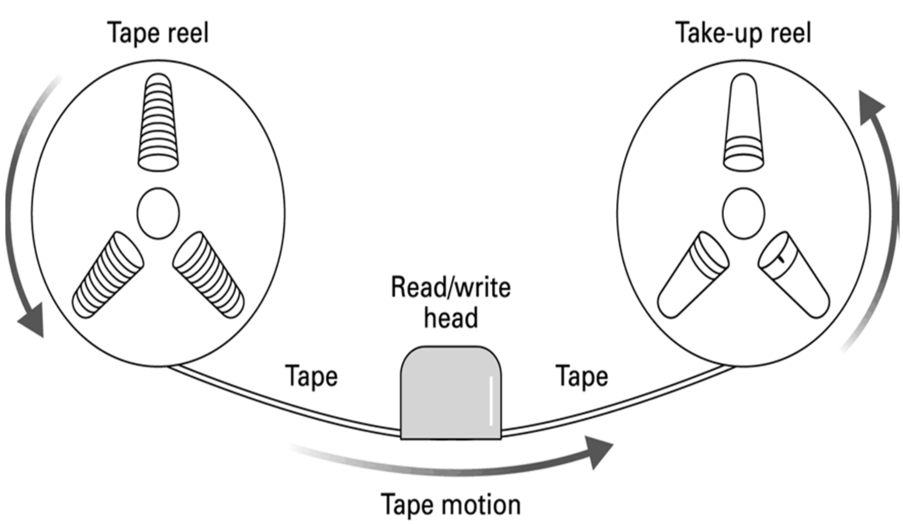


Figure 1.9 **Magnetic disk storage**

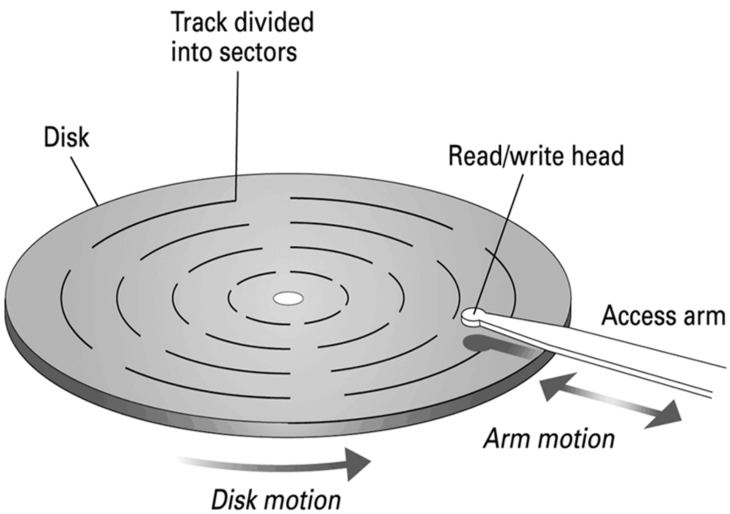
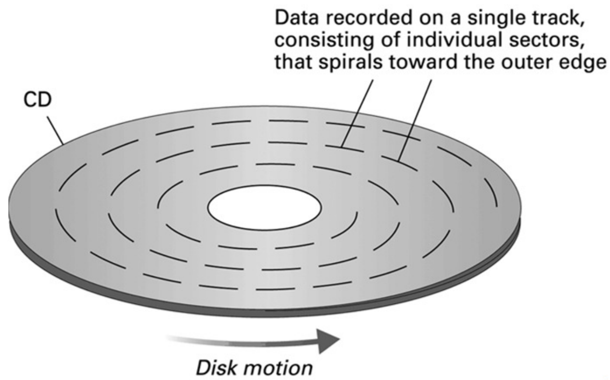
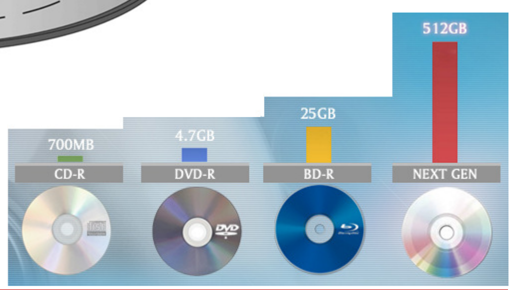


Figure 1.11 **Optical CD storage**





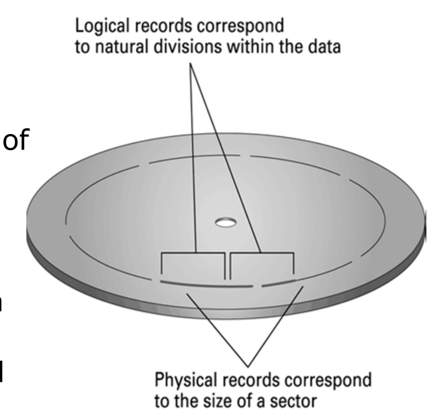
Solid State Storage

- **Flash Memory** – circuits that traps electrons in tiny silicon dioxide chambers
- Repeated erasing slowly damages the media
- Mass storage of choice for:
  - Digital cameras
  - Smartphones
- **Flash Drives** – portable units with capacities of up to a few hundred GBs, are typically connected via USB ports
- **Secure Digital (SD) Cards** provide GBs of storage
- **Solid State Disks (SSD)**

	Full SD	miniSD	microSD	SD Card Capacity
SD				up to 2GB
SDHC				High Capacity 4GB to 32GB
SDXC				Extended Capacity Over 32GB up to 2TB

Files

- **File:** A unit of data stored in mass storage system
- **Physical record** versus **Logical record**
- Logical records often consist of smaller units called **fields** (e.g. name, address, etc.)
- Logical records may be uniquely identified by a **key field**, whose value is called a **key**
- **Buffer:** A memory area used for the temporary storage of data (usually as a step in transferring the data)



Representing Text

- **Each character (letter, punctuation, etc.) is assigned a unique bit pattern.**
  - **ASCII**: The original 128 character codes uses patterns of 7-bits to represent most symbols used in written English text. DOS extended this to 256 characters (8-bit).
  - **International Organization for Standardization (ISO)** developed a number of 8 bit extensions to ASCII, each designed to accommodate a major language group
  - **Unicode**: Uses patterns of 16-bits to represent the major symbols used in languages world wide

Figure 1.13 **The message "Hello." in ASCII or UTF-8 encoding**

Decimal:	ArtNet:	Hex:	ASCII:
0	0:0	00	NUL
1	0:1	01	SOH
2	0:2	02	STX
3	0:3	03	ETX
4	0:4	04	EOT
5	0:5	05	ENQ
6	0:6	06	ACK
7	0:7	07	BEL
8	0:8	08	BS
9	0:9	09	TAB
10	0:A	0A	LF
11	0:B	0B	VT
12	0:C	0C	FF
13	0:D	0D	CR
14	0:E	0E	SO
15	0:F	0F	SI
16	1:0	10	DLE
17	1:1	11	DC1
18	1:2	12	DC2
19	1:3	13	DC3
20	1:4	14	DC4
21	1:5	15	NAK
22	1:6	16	SYN
23	1:7	17	ETB
24	1:8	18	CAN
25	1:9	19	EM
26	1:A	1A	SUB
27	1:B	1B	ESC
28	1:C	1C	FS
29	1:D	1D	GS
30	1:E	1E	RS
31	1:F	1F	US

Decimal:	ArtNet:	Hex:	ASCII:
32	2:0	20	(space)
33	2:1	21	!
34	2:2	22	"
35	2:3	23	#
36	2:4	24	$
37	2:5	25	%
38	2:6	26	&
39	2:7	27	'
40	2:8	28	(
41	2:9	29	)
42	2:A	2A	*
43	2:B	2B	+
44	2:C	2C	,
45	2:D	2D	-
46	2:E	2E	.
47	2:F	2F	/
48	3:0	30	0
49	3:1	31	1
50	3:2	32	2
51	3:3	33	3
52	3:4	34	4
53	3:5	35	5
54	3:6	36	6
55	3:7	37	7
56	3:8	38	8
57	3:9	39	9
58	3:A	3A	:
59	3:B	3B	;
60	3:C	3C	<
61	3:D	3D	=
62	3:E	3E	>
63	3:F	3F	?

Decimal:	ArtNet:	Hex:	ASCII:
64	4:0	40	@
65	4:1	41	A
66	4:2	42	B
67	4:3	43	C
68	4:4	44	D
69	4:5	45	E
70	4:6	46	F
71	4:7	47	G
72	4:8	48	H
73	4:9	49	I
74	4:A	4A	J
75	4:B	4B	K
76	4:C	4C	L
77	4:D	4D	M
78	4:E	4E	N
79	4:F	4F	O
80	5:0	50	P
81	5:1	51	Q
82	5:2	52	R
83	5:3	53	S
84	5:4	54	T
85	5:5	55	U
86	5:6	56	V
87	5:7	57	W
88	5:8	58	X
89	5:9	59	Y
90	5:A	5A	Z
91	5:B	5B	[
92	5:C	5C	\
93	5:D	5D	]
94	5:E	5E	^
95	5:F	5F	_

Decimal:	ArtNet:	Hex:	ASCII:
96	6:0	60	`
97	6:1	61	a
98	6:2	62	b
99	6:3	63	c
100	6:4	64	d
101	6:5	65	e
102	6:6	66	f
103	6:7	67	g
104	6:8	68	h
105	6:9	69	i
106	6:A	6A	j
107	6:B	6B	k
108	6:C	6C	l
109	6:D	6D	m
110	6:E	6E	n
111	6:F	6F	o
112	7:0	70	p
113	7:1	71	q
114	7:2	72	r
115	7:3	73	s
116	7:4	74	t
117	7:5	75	u
118	7:6	76	v
119	7:7	77	w
120	7:8	78	x
121	7:9	79	y
122	7:A	7A	z
123	7:B	7B	{
124	7:C	7C	|
125	7:D	7D	}
126	7:E	7E	~
127	7:F	7F	DEL

Decimal:	ArtNet:	Hex:	ASCII:
128	8:0	80	€
129	8:1	81	
130	8:2	82	‚
131	8:3	83	ƒ
132	8:4	84	„
133	8:5	85	…
134	8:6	86	†
135	8:7	87	‡
136	8:8	88	ˆ
137	8:9	89	‰
138	8:A	8A	Š
139	8:B	8B	‹
140	8:C	8C	Œ
141	8:D	8D	
142	8:E	8E	Ž
143	8:F	8F	
144	9:0	90	
145	9:1	91	‘
146	9:2	92	’
147	9:3	93	“
148	9:4	94	”
149	9:5	95	•
150	9:6	96	–
151	9:7	97	—
152	9:8	98	˜
153	9:9	99	™
154	9:A	9A	š
155	9:B	9B	›
156	9:C	9C	œ
157	9:D	9D	
158	9:E	9E	ž
159	9:F	9F	

Decimal:	ArtNet:	Hex:	ASCII:
160	A:0	A0	
161	A:1	A1	¡
162	A:2	A2	¢
163	A:3	A3	£
164	A:4	A4	¤
165	A:5	A5	¥
166	A:6	A6	¦
167	A:7	A7	§
168	A:8	A8	¨
169	A:9	A9	©
170	A:A	AA	ª
171	A:B	AB	«
172	A:C	AC	¬
173	A:D	AD	-
174	A:E	AE	®
175	A:F	AF	¯
176	B:0	B0	°
177	B:1	B1	±
178	B:2	B2	²
179	B:3	B3	³
180	B:4	B4	´
181	B:5	B5	µ
182	B:6	B6	¶
183	B:7	B7	·
184	B:8	B8	¸
185	B:9	B9	¹
186	B:A	BA	º
187	B:B	BB	»
188	B:C	BC	¼
189	B:D	BD	½
190	B:E	BE	¾
191	B:F	BF	

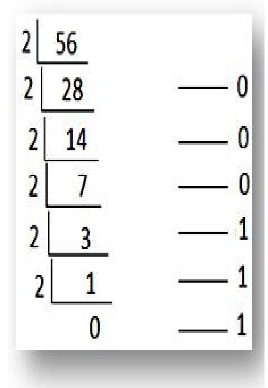
Decimal:	ArtNet:	Hex:	ASCII:
192	C:0	C0	À
193	C:1	C1	Á
194	C:2	C2	Â
195	C:3	C3	Ã
196	C:4	C4	Ä
197	C:5	C5	Å
198	C:6	C6	Æ
199	C:7	C7	Ç
200	C:8	C8	È
201	C:9	C9	É
202	C:A	CA	Ê
203	C:B	CB	Ë
204	C:C	CC	Ì
205	C:D	CD	Í
206	C:E	CE	Î
207	C:F	CF	Ï
208	D:0	D0	Ð
209	D:1	D1	Ñ
210	D:2	D2	Ò
211	D:3	D3	Ó
212	D:4	D4	Ô
213	D:5	D5	Õ
214	D:6	D6	Ö
215	D:7	D7	×
216	D:8	D8	Ø
217	D:9	D9	Ù
218	D:A	DA	Ú
219	D:B	DB	Û
220	D:C	DC	Ü
221	D:D	DD	Ý
222	D:E	DE	Þ
223	D:F	DF	

Decimal:	ArtNet:	Hex:	ASCII:
224	E:0	E0	à
225	E:1	E1	á
226	E:2	E2	â
227	E:3	E3	ã
228	E:4	E4	ä
229	E:5	E5	å
230	E:6	E6	æ
231	E:7	E7	ç
232	E:8	E8	è
233	E:9	E9	é
234	E:A	EA	ê
235	E:B	EB	ë
236	E:C	EC	ì
237	E:D	ED	í
238	E:E	EE	î
239	E:F	EF	ï
240	F:0	F0	ð
241	F:1	F1	ñ
242	F:2	F2	ò
243	F:3	F3	ó
244	F:4	F4	ô
245	F:5	F5	õ
246	F:6	F6	ö
247	F:7	F7	÷
248	F:8	F8	ø
249	F:9	F9	ù
250	F:A	FA	ú
251	F:B	FB	û
252	F:C	FC	ü
253	F:D	FD	ý
254	F:E	FE	þ
255	F:F	FF	

01001000	01100101	01101100	01101100	01101111	00101110
H	**e**	**l**	**l**	**o**	**.**

Representing Numeric Values

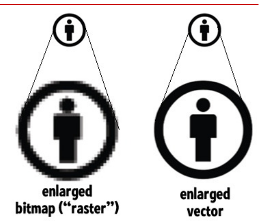
- **Binary notation**: Uses bits to represent a number in base two
- Limitations of computer representations of numeric values
  - Overflow: occurs when a value is too big to be represented
  - Truncation: occurs when a value cannot be represented accurately

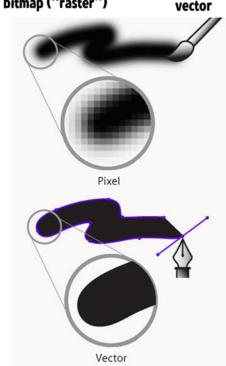


56 =111000

Representing Images

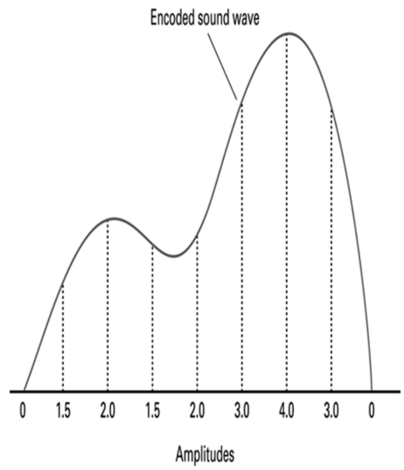
- **Pixel**: short for "picture element"
- **Bit map** techniques
  - The image is encoded as a collection of encoded pixels
  - RGB (3 byte color encoding)
  - Luminance and chrominance
    - Luminance - Brightness (white light: R+G+B)
    - Chrominance – Two color component
      - Blue chrominance (Diff of Luminance and B light)
      - Red chrominance (Diff of Luminance and R light)
  - Images cannot be rescaled easily (enlargement results in "pixelation")
- **Vector** techniques
  - Image described as collection of geometric structures
  - Scalable
  - TrueType, PostScript, CAD, …





Representing Sound

- Sampling techniques
  - Used for high quality recordings
  - Records actual audio
  - Telephone communication: 8000 samples/second
  - High-fidelity music recordings: 44,100 samples/ second
- **MIDI**
  - Used in music synthesizers
  - Records "musical score"
  - What instrument for what time duration
  - Clarinet playing D note for 2 seconds:
    - Sampling: >2 million bits
    - MIDI: 3 bytes (24 bits)



Sound wave represented by sequence:
0, 1.5, 2.0, 1.5, 2.0, 3.0, 4.0, 3.0, 0