Data Representation

Prof. Dr. Selim Akyokuş
sakyokus@medipol.edu.tr

Layers of a Computing System

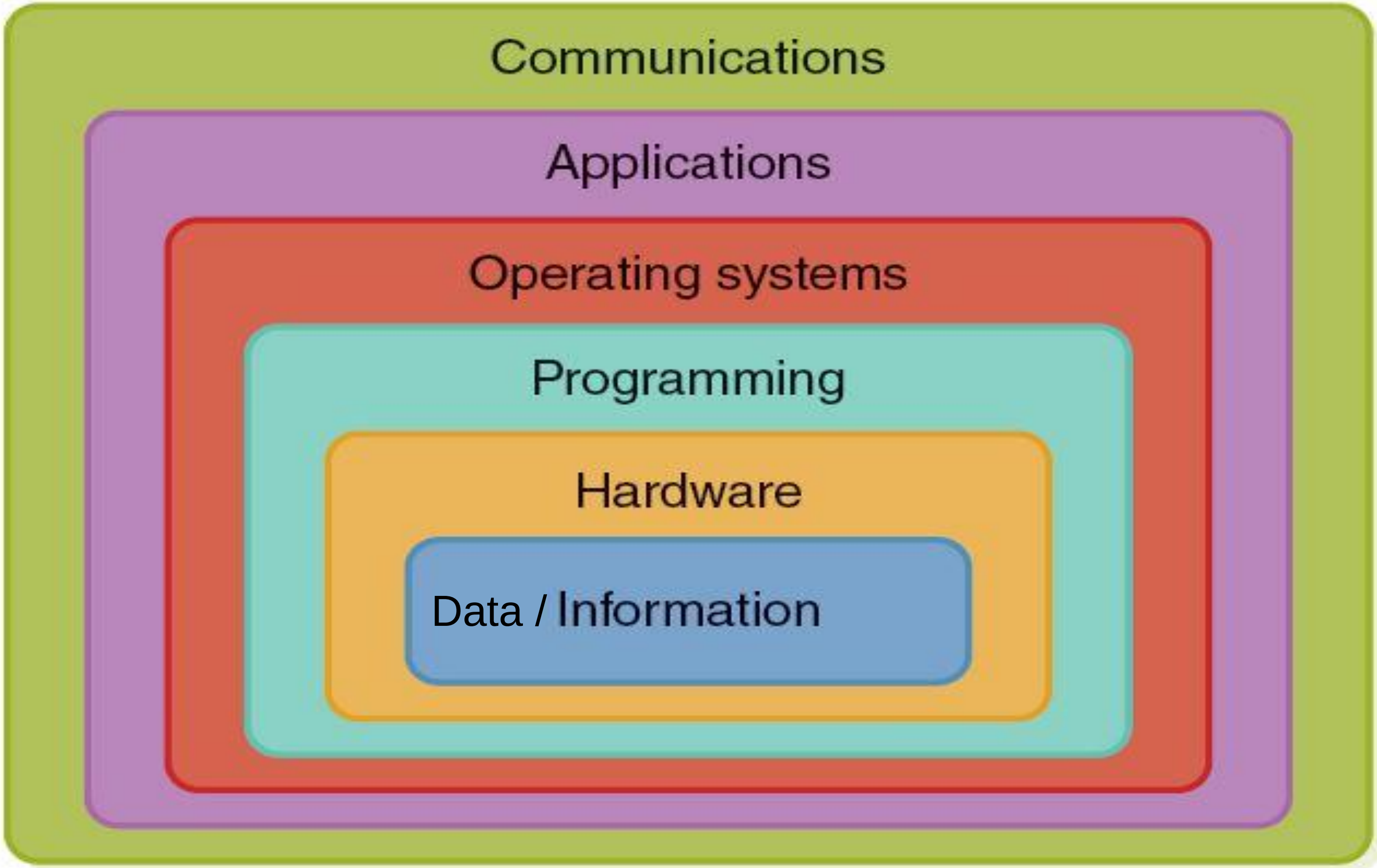


FIGURE 1.1 The layers of a computing system

Data Representation

- Computers can represent any INFORMATION that can be discretized and digitized
- Algorithms (Programs) process and transform data
  - Search for patterns
  - Create simulations
  - Generate knowledge and insight
- How do computers store or represent data about common digital artifacts?
  - Numbers, text, images, sounds, and video

Analog and Digital Information

Information can be represented in one of two ways: **analog** or **digital**

Analog data

A continuous representation, analogous to the actual information it represents

Digital data

A discrete representation, breaking the information up into separate elements

Analog and Digital Information

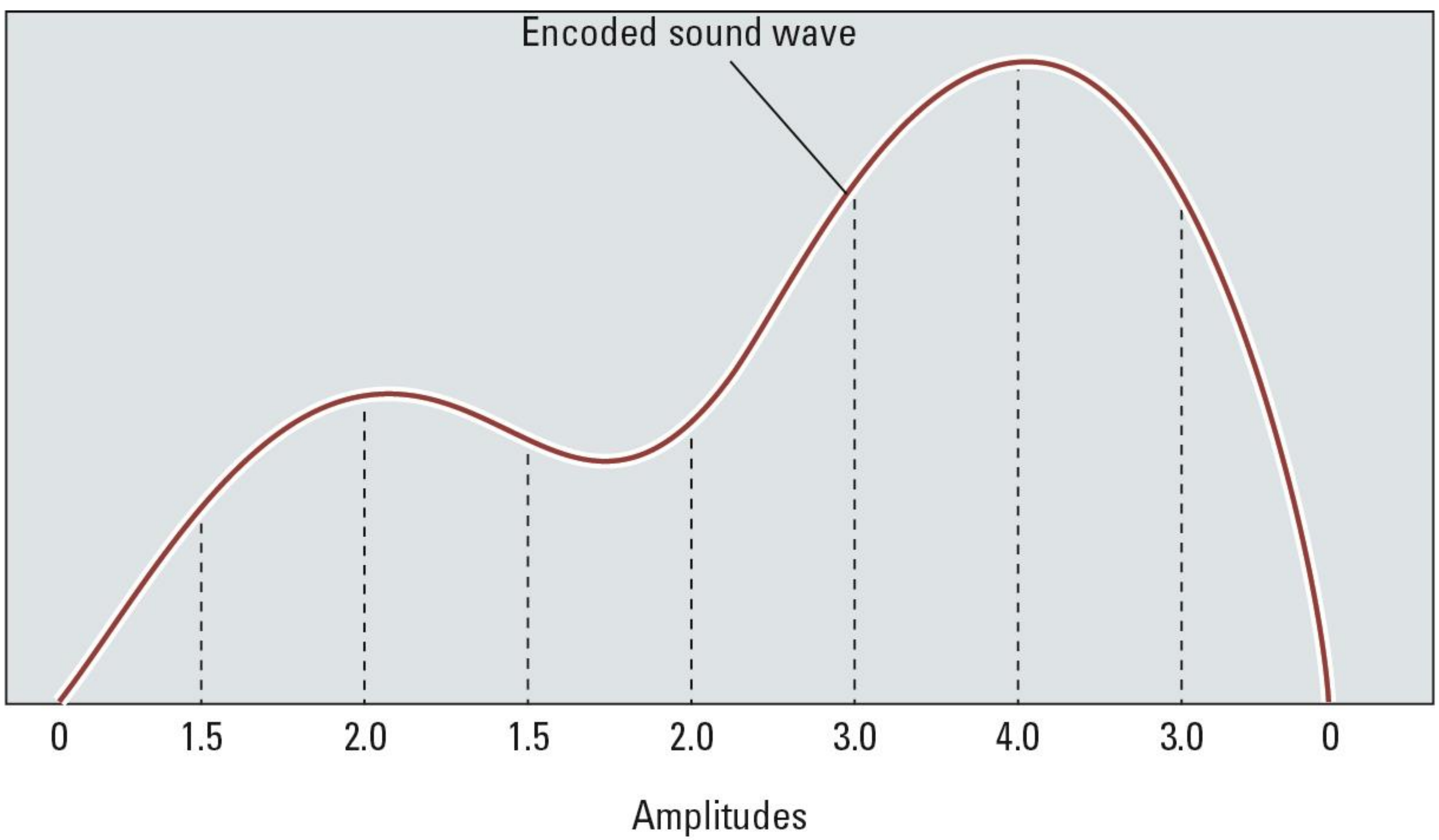
Computers cannot work well with analog data, so we digitize the data

Digitize

Breaking data into pieces and representing those pieces separately

Why do we use binary to represent digitized data?

The sound wave represented by the sequence 0, 1.5, 2.0, 1.5, 2.0, 3.0, 4.0, 3.0, 0



Electronic Signals

Important facts about electronic signals

- An analog signal continually fluctuates in voltage up and down
- A digital signal has only a high or low state, corresponding to the two binary digits
- All electronic signals (both analog and digital) degrade as they move down a line
- The voltage of the signal fluctuates due to environmental effects

FIGURE 3.2 An analog signal and a digital signal

Binary Numbers and Computers

Computers have storage units called binary digits or bits

Low Voltage = 0
High Voltage = 1 all bits have 0 or 1

… or the other way around, but we don't need to worry about that

Number Systems

Natural Numbers

Zero and any number obtained by repeatedly adding one to it.

Examples: 100, 0, 45645, 32

Negative Numbers

A value less than 0, with a – sign

Examples: -24, -1, -45645, -32

Numbers

Integers

A natural number, a negative number

Examples: 249, 0, - 45645, - 32

Rational Numbers

An integer or the quotient of two integers

Examples: -249, -1, 0, 3/7, -2/5

Positional Notation

How many ones are there in 642?

600 + 40 + 2 ?

Or is it

384 + 32 + 2 ?

Or maybe…

1536 + 64 + 2 ?

Positional Notation

Aha!

642 is 600 + 40 + 2 in **BASE 10**

The **base** of a number determines the number of different digit symbols (numerals) and the values of digit positions

Positional Notation

Continuing with our example…

642 in base 10 *positional notation* is:

$$
\begin{aligned}
& 6 \times 10^2 = 6 \times 100 = 600 \\
+ & \ 4 \times 10^1 = 4 \times 10 = 40 \\
+ & \ 2 \times 10^0 = 2 \times 1 = 2 \quad = 642 \text{ in base 10}
\end{aligned}
$$

This number is in base 10

The power indicates the position of the number

Positional Notation

R is the base of the number

As a formula:

$$d_n * R^{n-1} + d_{n-1} * R^{n-2} + \ldots + d_2 * R^1 + d_1 * R^0$$

n is the number of digits in the number

d is the digit in the i^{th} position in the number

642 is $6 * 10^2 + 4 * 10^{+} 2 * 1$

Positional Notation

What if 642 has the base of 13?

+ 6 x 13^2 = 6 x 169 = 1014
+ 4 x 13^1 = 4 x 13 = 52
+ 2 x 13^0 = 2 x 1 = 2
= 1068 in base 10

642 in base 13 is equivalent to 1068 in base 10

Binary

Decimal is base 10 and has 10 digit symbols:

0,1,2,3,4,5,6,7,8,9

Binary is base 2 and has 2 digit symbols:

0,1

For a number to exist in a given base, it can only contain the digits in that base, which range from 0 up to (but not including) the base.

What bases can these numbers be in? 122, 198, 178, G1A4

Bases Higher than 10

How are digits in bases higher than 10 represented?

With distinct symbols for 10 and above.

Base 16 has 16 digits:

0,1,2,3,4,5,6,7,8,9,A,B,C,D,E, and F

The hexadecimal coding system

Bit pattern	Hexadecimal representation
0000	0x0
0001	0x1
0010	0x2
0011	0x3
0100	0x4
0101	0x5
0110	0x6
0111	0x7
1000	0x8
1001	0x9
1010	0xA
1011	0xB
1100	0xC
1101	0xD
1110	0xE
1111	0xF

Converting Octal to Decimal

- Octal means Base 8

- What is the decimal equivalent of the octal number 642?

$$
\begin{aligned}
& 6 \times 8^2 = 6 \times 64 = 384 \\
+ & 4 \times 8^1 = 4 \times 8 = 32 \\
+ & 2 \times 8^0 = 2 \times 1 = 2 \\
& = 418 \text{ in base } 10
\end{aligned}
$$

Converting Hexadecimal to Decimal

What is the decimal equivalent of the hexadecimal number DEF?

$$
\begin{aligned}
&\mathbf{D \times 16^2 = 13 \times 256 = 3328}\\
&\mathbf{+\ E \times 16^1 = 14 \times 16 = 224}\\
&\mathbf{+\ F \times 16^0 = 15 \times 1 = 15}\\
&\mathbf{= 3567 \text{ in base } 10}
\end{aligned}
$$

Remember, the digit symbols in base 16 are 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F

Converting Binary to Decimal

What is the decimal equivalent of the binary number 1101110?

$$
\begin{aligned}
& 1 \times 2^6 = 1 \times 64 = 64 \\
+ & 1 \times 2^5 = 1 \times 32 = 32 \\
+ & 0 \times 2^4 = 0 \times 16 = 0 \\
+ & 1 \times 2^3 = 1 \times 8 = 8 \\
+ & 1 \times 2^2 = 1 \times 4 = 4 \\
+ & 1 \times 2^1 = 1 \times 2 = 2 \\
+ & 0 \times 2^o = 0 \times 1 = 0 \\
& = 110 \text{ in base } 10
\end{aligned}
$$

The binary addition

$$\begin{array}{r} 0 \\ +\,0 \\ \hline 0 \end{array} \qquad \begin{array}{r} 1 \\ +\,0 \\ \hline 1 \end{array} \qquad \begin{array}{r} 0 \\ +\,1 \\ \hline 1 \end{array} \qquad \begin{array}{r} 1 \\ +\,1 \\ \hline 10 \end{array}$$

Arithmetic in Binary

Remember that there are only 2 digit symbols in binary, 0 and 1

1 + 1 is 0 with a carry

```
1 0 1 1 1 1 1        <- Carry Values
  1 0 1 0 1 1 1
 +1 0 0 1 0 1 1
 ---------------
1 0 1 0 0 0 1 0
```

Converting Binary to Octal

- Mark groups of *three* (from right)
- Convert each group

10101011	**<u>10</u>**	**<u>101</u>**	**<u>011</u>**
	2	**5**	**3**

10101011 is 253 in base 8

Converting Binary to Hexadecimal

- Mark groups of *four* (from right)
- Convert each group

10101011   **<u>1010</u> <u>1011</u>**

A B

10101011 is AB in base 16

Converting Decimal to Other Bases

Algorithm for converting number in base 10 to other bases

```
While (the quotient is not zero)
    Divide the decimal number by the new base
    Make the remainder the next digit to the left in the answer
    Replace the original decimal number with the quotient
```

Converting Decimal to Octal

What is 1988 (base 10) in base 8?

Try it!

Converting Decimal to Octal

248	31	3	0
8 ⟌ 1988	8 ⟌ 248	8 ⟌ 31	8 ⟌ 3
16	24	24	0
38	08	7	3
32	8		
68	0		
64			
4			

Answer is : **3 7 0 4**

Converting Decimal to Hexadecimal

What is 3567 (base 10) in base 16?

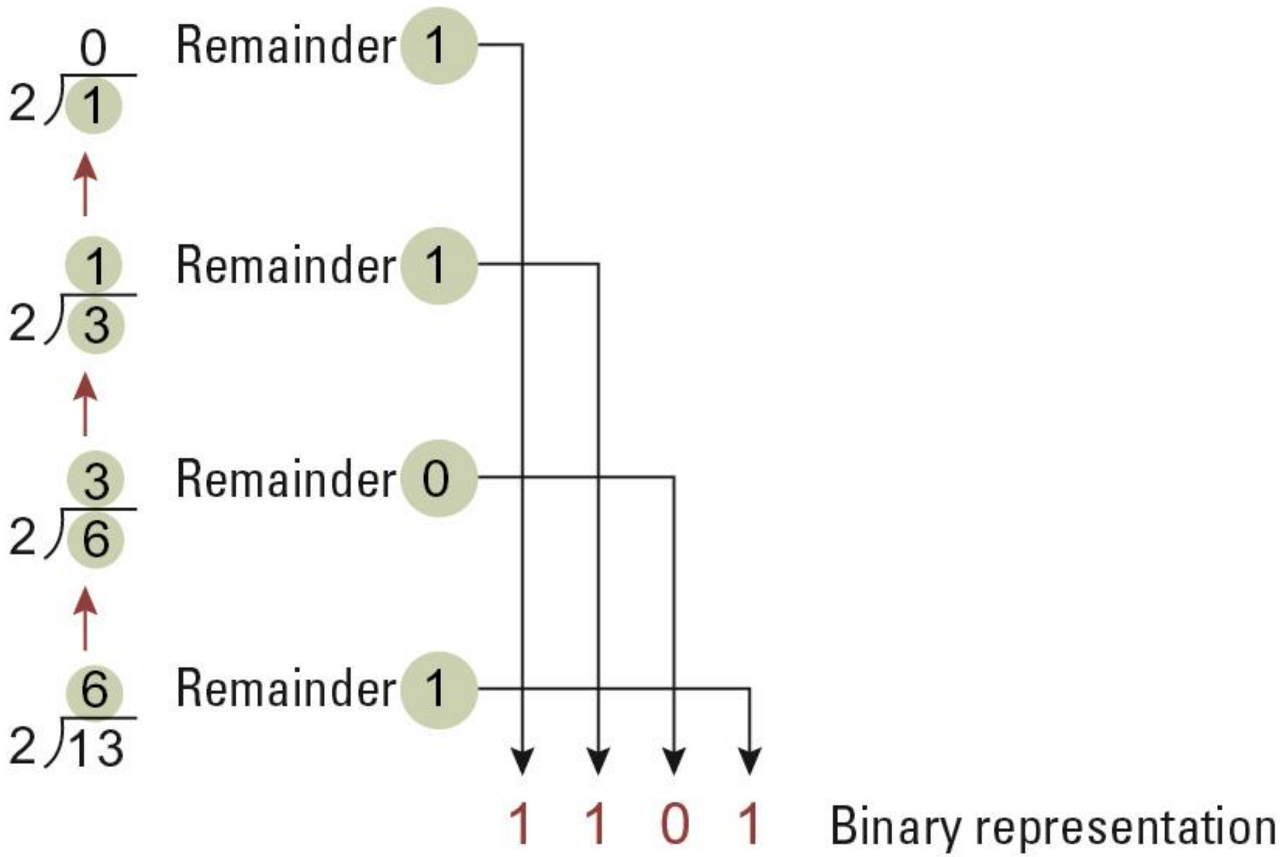
Try it!

Converting Decimal to Hexadecimal

222	13	0
16 ⟌ 3567	16 ⟌ 222	16 ⟌ 13
<u>32</u>	<u>16</u>	<u>0</u>
36	62	13
<u>32</u>	<u>48</u>	
47	14	
<u>32</u>		
15		

D E F

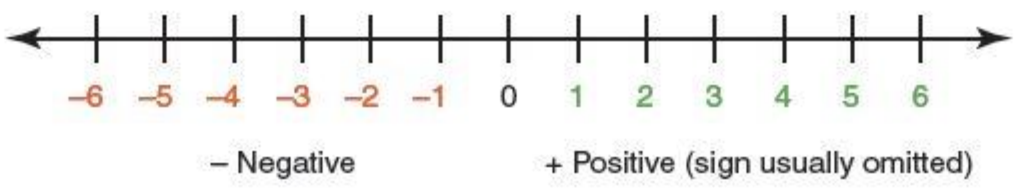
Applying the algorithm to obtain the binary representation of thirteen



Representing Negative Values

Signed-magnitude number representation

- Used by humans
- The sign represents the ordering (the negatives come before the positives in ascending order)
- The digits represent the magnitude (the distance from zero)



Representing Negative Values

Problem: Two zeroes (positive and negative)

No problem for humans, but would cause unnecessary complexity in computers

Solution: Represent integers by associating them with natural numbers

Half the natural numbers will represent themselves

The other half will represent negative integers

How to Represent Integer Numbers ?

- **Two's complement notation:** The most popular means of representing integer values with negative values
- Allows the use of the same circuit for both addition and subtraction

Two's complement notation systems

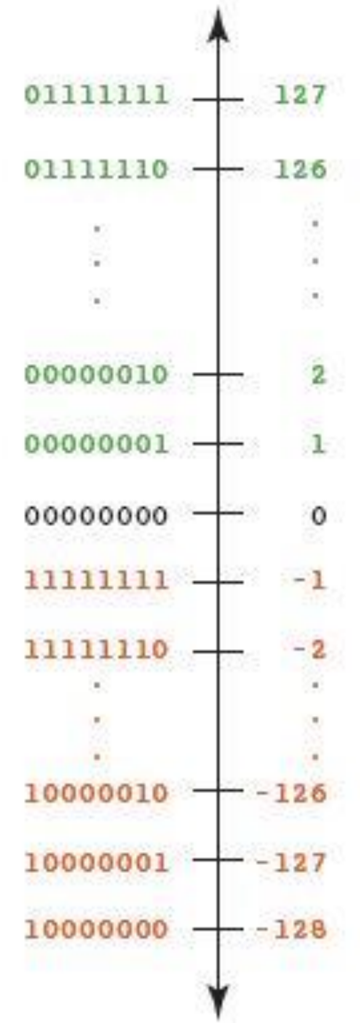
a. Using patterns of length three

Bit pattern	Value represented
011	3
010	2
001	1
000	0
111	–1
110	–2
101	–3
100	–4

b. Using patterns of length four

Bit pattern	Value represented
0111	7
0110	6
0101	5
0100	4
0011	3
0010	2
0001	1
0000	0
1111	–1
1110	–2
1101	–3
1100	–4
1011	–5
1010	–6
1001	–7
1000	–8

Representing Negative Values



Two's Complement

(The binary number line is easier to read when written vertically)

Remember our table showing how to represent natural numbers?

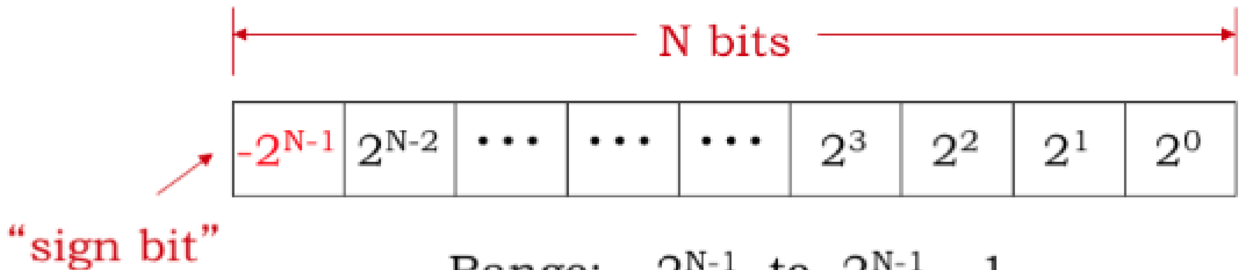
Do you notice something interesting about the left-most bit?

Two's Complement Conversion

- Negate numbers (Change 1s to 0s and 0s to 1s, and add 1)
- $13_{10} = 1101_2$
- $- 13_{10} = 0011_2$
  - Two's Complement Representation: 0010 + 1 = 0011

Signed Two's Complement Representation

In a two's complement encoding, the high-order bit of the N-bit representation has negative weight:



Range: -2^{N-1} to $2^{N-1}-1$

- Negative numbers have "1" in the high-order bit
- Most negative number: 10…0000 -2^{N-1}
- Most positive number: 01…1111 $+2^{N-1} - 1$
- If all bits are 1: 11…1111 -1
- If all bits are 0: 00…0000 0

Representing Negative Values

Addition and subtraction are the same as in ten's complement arithmetic

```
 -127     10000001
+   1     00000001
 -126     10000010
```

What if the computed value won't fit?

Number Overflow

If each value is stored using 8 bits, then 127 + 3 overflows:

```
    01111111
+   00000011
    10000010
```

Apparently, 127 + 3 is -126. Remember when we said we would always fail in our attempt to map an infinite world onto a finite machine?

Most computers use 32 or 64 bits for integers, but there are always infinitely many that aren’t represented

Representing Real Numbers

Real numbers are numbers with a whole part and a fractional part (either of which may be zero)

104.32
0.999999
357.0
3.14159

In decimal, positions to the right of the decimal point are the tenths, hundredths, thousandths, etc.:

10^{-1}, 10^{-2} , 10^{-3} …

Representing Real Numbers

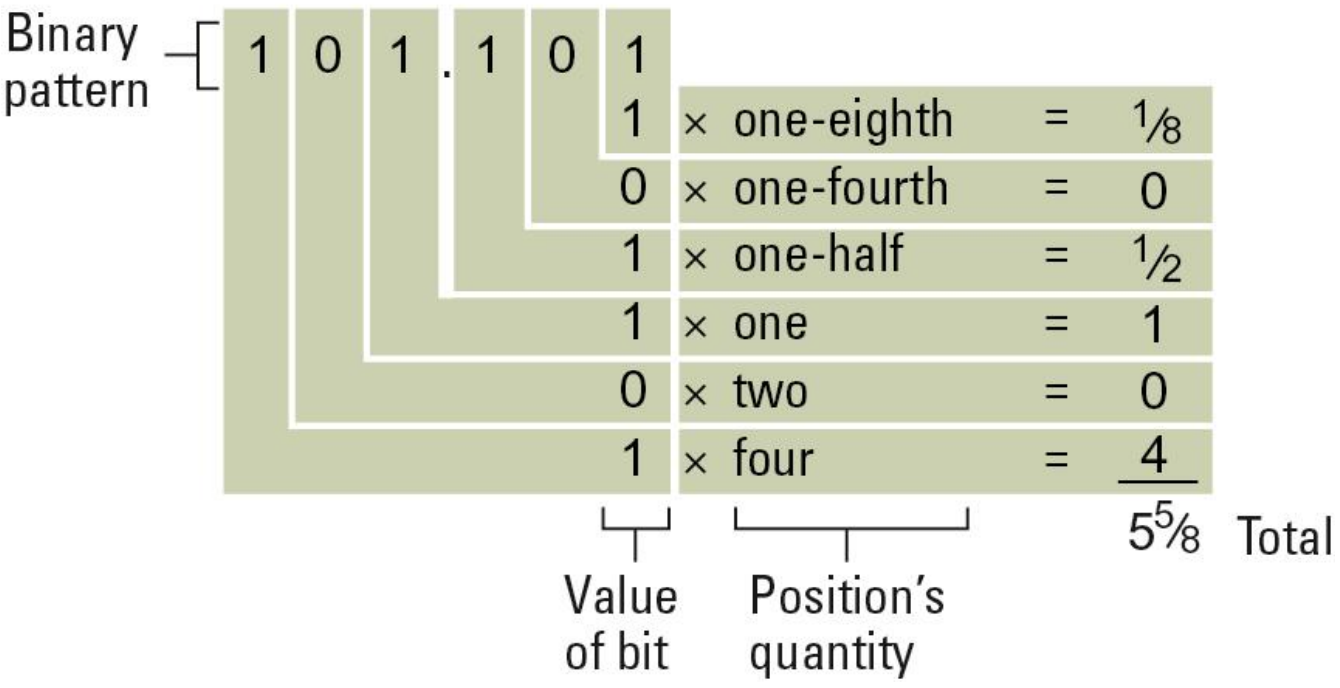
Same rules apply in binary as in decimal

Radix point is general term for “decimal point”

Positions to the right of the radix point in binary:

2^{-1} (halves position),
2^{-2} (quarters position),
2^{-3} (eighths position)
…

Decoding the binary representation 101.101



Representing Real Numbers

A real value in base 10 can be defined by the following formula where the mantissa is an integer

$$\text{sign} * \text{mantissa} * 10^{\text{exp}}$$

This representation is called **floating point** because the radix point "floats"

In analogy to the fixed number of bits that computers use to represent integers, we'll treat the mantissa as having a fixed number of digits

Representing Real Numbers

TABLE 3.1 Values in decimal notation and floating-point notation (five digits)

Real Value	Floating-Point Value
12001.00	$12001 * 10^0$
–120.01	$-12001 * 10^{-2}$
0.12000	$12000 * 10^{-5}$
–123.10	$-12310 * 10^{-2}$
155555000.00	$15555 * 10^4$

Floating-point in binary:

sign * mantissa * 2^{exp}

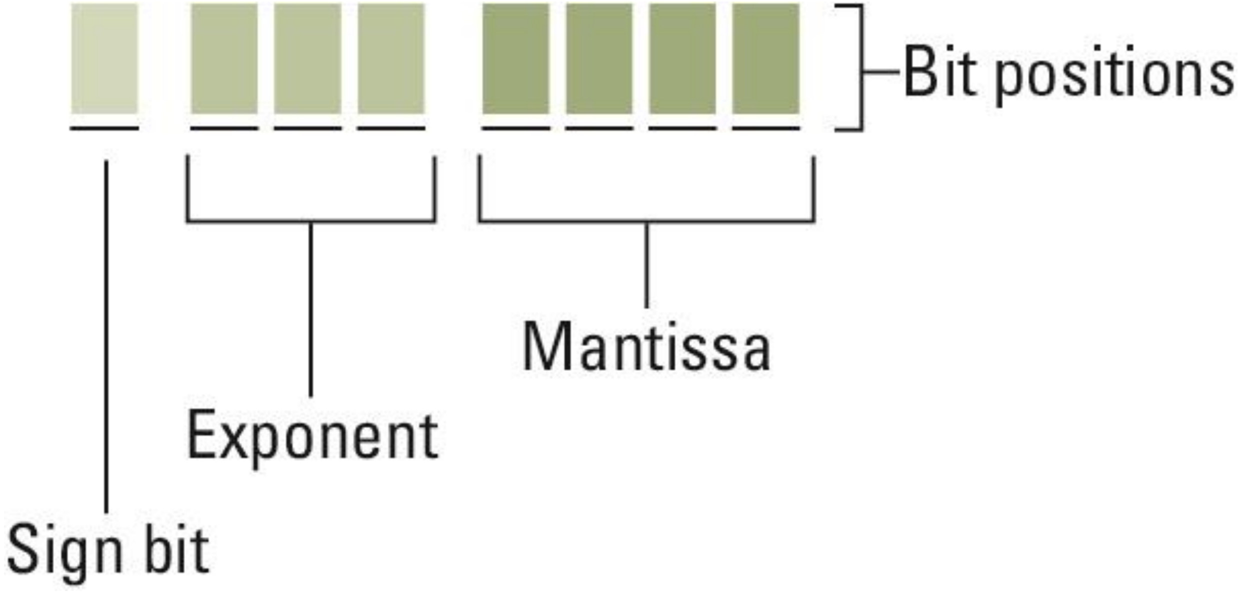
Only the base value is different from decimal

Fundamentally, the floating-point used by computers is very similar, but uses complicated tricks to represent more numbers and improve efficiency

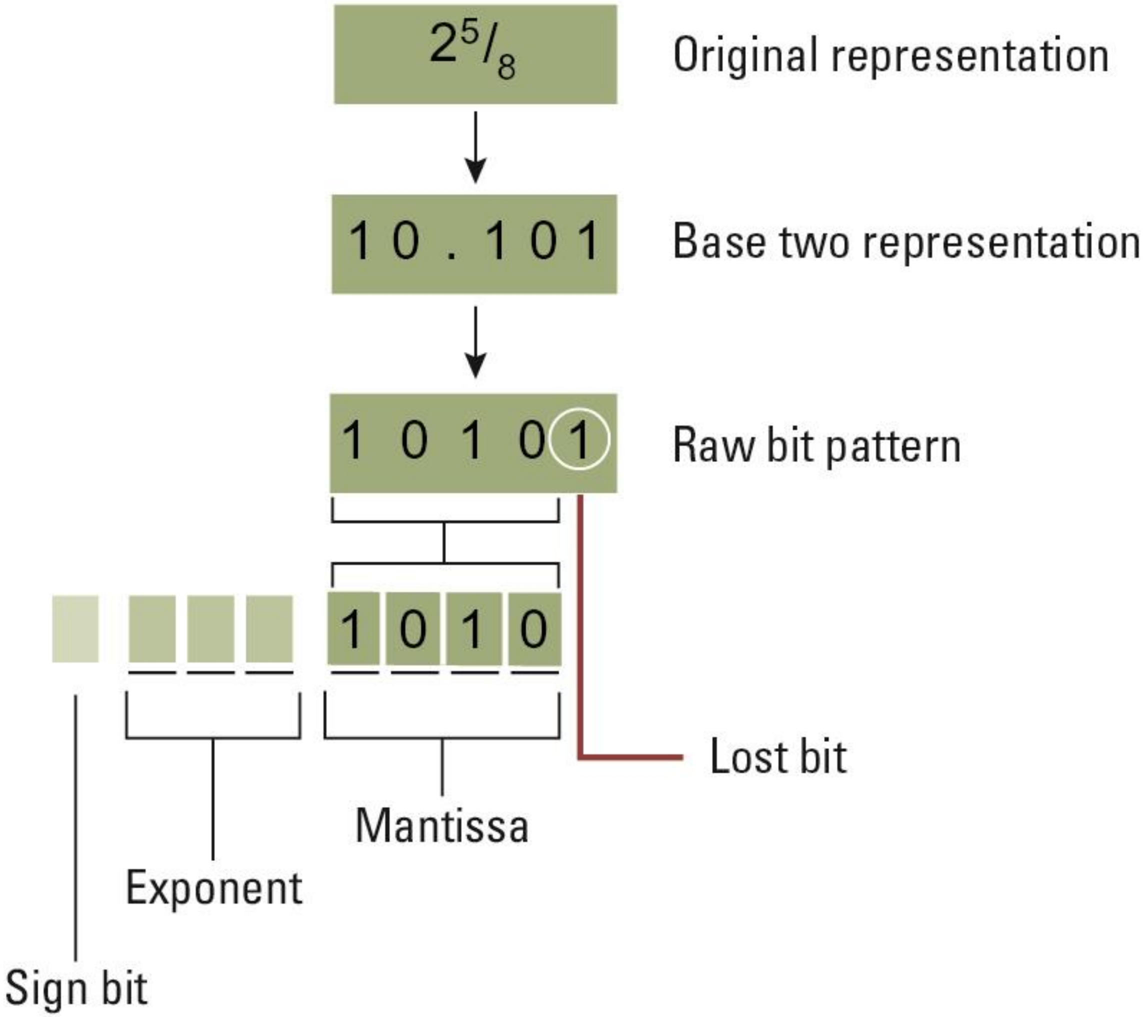
Storing Fractions in Binary

- **Floating-point Notation:** Consists of a sign bit, a mantissa field, and an exponent field.
  - Normalized form: fill the mantissa starting with the left-most 1

Floating-point notation components



Encoding the value 2 ⅝



Truncation (Round-off) Errors

- Occur when part of the value being stored is lost because the mantissa is not large enough
- Non-terminating expansions of fractions
  - This happens more often with binary notation
  - The value of one-tenth cannot be stored exactly binary notation
  - Often these values are converted to integers

Representing Text

What must be provided to represent text?

The number of characters to represent is finite (whew!), so list them all and assign each a binary string

Character set

A list of characters and the codes used to represent each one

Computer manufacturers agreed to standardize

The ASCII Character Set

ASCII stands for American Standard Code for Information Interchange

ASCII originally used seven bits to represent each character, allowing for 128 unique characters

Later extended ASCII evolved so that all eight bits were used

How many characters could be represented?

ASCII Character Set Mapping

Left Digit(s) \ Right Digit	ASCII 0	1	2	3	4	5	6	7	8	9
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT
1	LF	VT	FF	CR	SO	SI	DLE	DC1	DC2	DC3
2	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS
3	RS	US	□	!	“	#	$	%	&	'
4	(	)	*	+	,	–	.	/	0	1
5	2	3	4	5	6	7	8	9	:	;
6	<	=	>	?	@	A	B	C	D	E
7	F	G	H	I	J	K	L	M	N	O
8	P	Q	R	S	T	U	V	W	X	Y
9	Z	[	\	]	^	_	`	a	b	c
10	d	e	f	g	h	i	j	k	l	m
11	n	o	p	q	r	s	t	u	v	w
12	x	y	z	{	|	}	~	DEL		

FIGURE 3.5 The ASCII character set

The ASCII Character Set

The first 32 characters in the ASCII character chart do not have a simple character representation to print to the screen

What do you think they are used for?

The Unicode Character Set

- Extended ASCII is not enough for international use
- One Unicode mapping usually uses 16 bits per character
  - *How many characters can this mapping represent?*
- The first 256 characters correspond exactly to the extended ASCII character set

The message “Hello.” in ASCII or UTF-8 encoding

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

The Unicode Character Set

Code (Hex)	Character	Source
0041	A	English (Latin)
042F	Я	Russian (Cyrillic)
0E09	ฉ	Thai
13EA	Ꮺ	Cherokee
211E	℞	Letterlike symbols
21CC	⇌	Arrows
282F	⠯	Braille
345F	㑟	Chinese/Japanese/Korean (common)

FIGURE 3.6 A few characters in the Unicode character set

Summary of Text Coding Standards

- **Each character (letter, punctuation, etc.) is assigned a unique bit pattern.**
  - **ASCII**: Uses patterns of 7-bits to represent most symbols used in written English text
  - **Extended ASCII**: uses 8-bits
  - **ISO** developed a number of 8 bit extensions to ASCII, each designed to accommodate a major language group
  - **Unicode**: Uses patterns up to 21-bits to represent the symbols used in languages world wide, 16-bits for world’s commonly used languages

Representing Audio Information

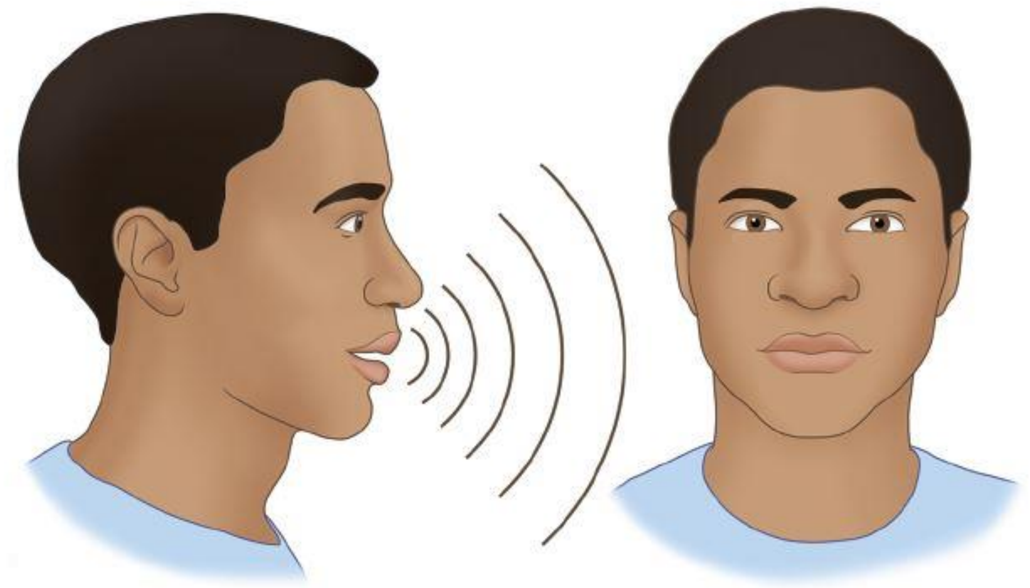


FIGURE 3.7 A sound wave vibrates our eardrums

We perceive sound when a series of air pressure waves vibrate a membrane in our ear, which sends signals to our brain

Representing Audio Information

Your parents may use a "stereo" to listen to music at home. It sends an electrical signal to each speaker, which then vibrates to produce sound. Your MP3 player and ear buds do the same thing.

The signal controls the motion of a membrane in the speaker, which in turn creates the pressure waves that reach our ears

Thus, the signal is an analog representation of the sound wave

Representing Audio Information

Digitize the signal by

- Sampling: periodically measure the voltage
- Quantization: represent the voltage as a number using a finite number of bits

How often should we sample?

A sampling rate of about 40,000 times per second is enough to create a reasonable sound reproduction

Representing Audio Information

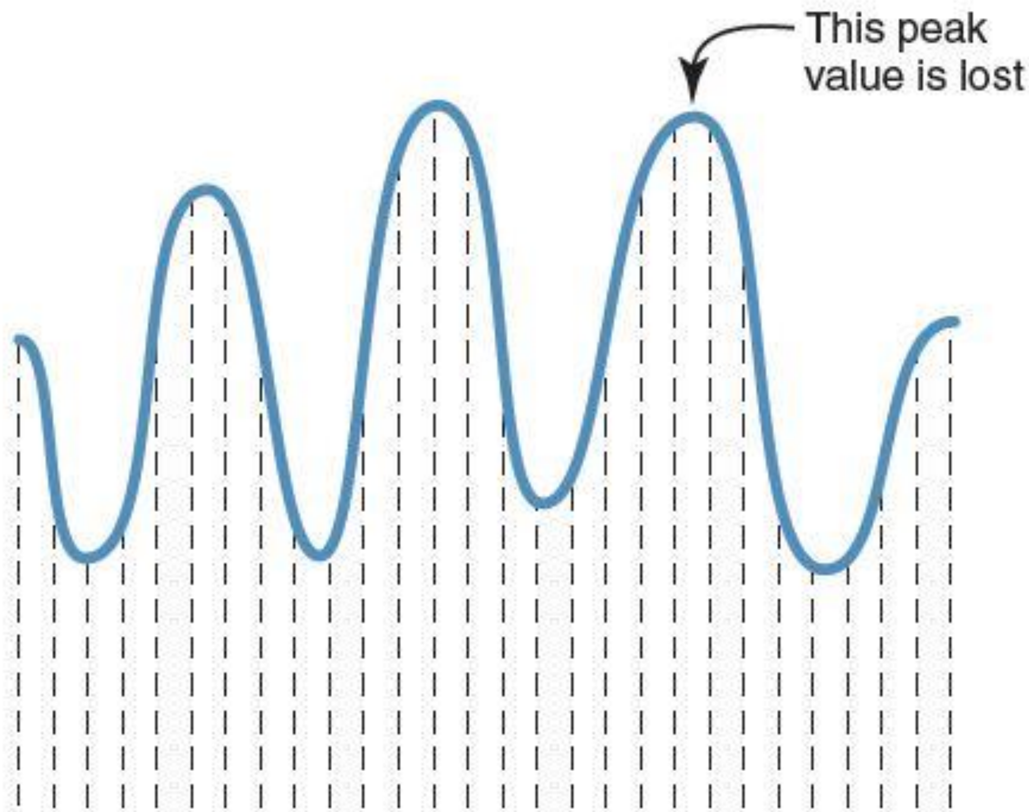


FIGURE 3.8 Sampling an audio signal

Some data is lost, but a reasonable sound is reproduced

Audio Formats

Audio Formats

- WAV, AU, AIFF, VQF, and MP3
- Use various compression techniques

MP3 is dominant

- MPEG-2, audio layer 3 file
- MPEG = Motion Picture Experts Group
- Based on studies of interrelation between ear and brain, discards frequency information that isn’t perceived by humans (science!)
- Additional compression by a form of Huffman encoding

Is this a lossy or lossless compression (or both)?

Representing Images and Graphics

Color

- We take it for granted, but what is it really?

Retinas of our eyes have three types of photoreceptor cone cells

- Each type responds to a different set of frequencies of light
- Our brain translates that response into a perception of red, green, or blue

Representing Images and Graphics

Color is expressed as an RGB (red-green-blue) value – three numbers that indicate the relative contribution of each of these three primary colors

An RGB value of (255, 255, 0) maximizes the contribution of red and green, and minimizes the contribution of blue, which results in a bright yellow

Representing Images and Graphics

Color depth

The amount of data that is used to represent a color

HiColor

A 16-bit color depth: five bits used for each number in an RGB value with the extra bit sometimes used to represent transparency

TrueColor

A 24-bit color depth: eight bits used for each number in an RGB value

Representing Images and Graphics

RGB VALUE			
Red	Green	Blue	Color
0	0	0	black
255	255	255	white
255	255	0	yellow
255	130	255	pink
146	81	0	brown
157	95	82	purple
140	0	0	maroon

A few TrueColor RGB values and the colors they represent

Representing Images and Graphics

A color palette is a set of colors, for example

- Colors supported by a monitor
- Web-safe colors for use with Internet browsers
- Colors from which user can choose
- Colors used in an image

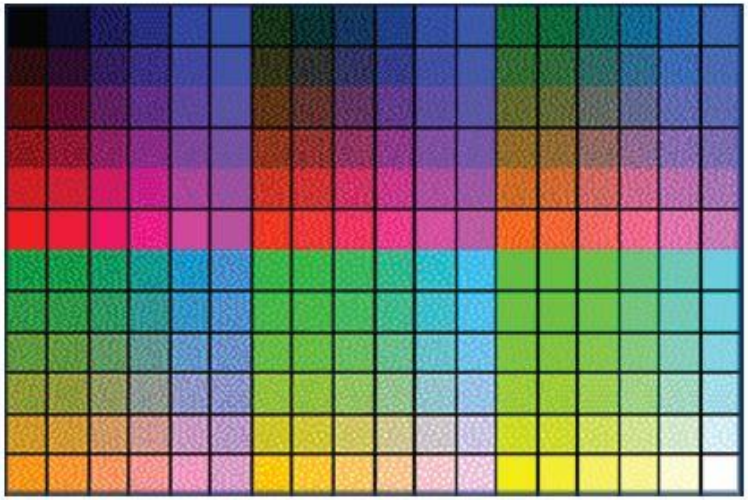


FIGURE 3.11 A restricted color palette

Digitized Images and Graphics

- Pixels (picture elements)
  - – Dots of color in image (or display device)
- Resolution
  - – Number of pixels in image (or device)
- Raster Graphics
  - – Treat image as collection of pixels
  - – Most common formats: BMP, GIF, PNG, and JPEG
- Vector Graphics
  - – Treat image as collection of geometric objects
  - – Most important formats: Flash and SVG

Digitized Images and Graphics

- BMP (bitmap)
  - TrueColor color depth, or less to reduce file size
  - Well suited for compression by run-length encoding

- GIF (indexed color)
  - File explicitly includes palette of 256 or fewer colors
  - Each pixel thus requires only 8 or fewer bits
  - Animated GIFs are short sequences of images

- PNG (Portable Network Graphics)
  - Intended to replace GIFs
  - Greater compression with wider range of color depths
  - No animation

Digitized Images and Graphics

- JPEG (Joint Photographic Experts Group)
  - Averages hues over short distances
    - Why? Human vision tends to blur colors together within small areas (science!)
    - How? Transform from the spatial domain to the frequency domain, then discard high frequency components (math!)
    - Sound familiar? Essentially the same idea used in MP3
  - Adjustable degree of compression

Raster graphics recap: BMP, GIF, PNG, and JPEG

Which use lossless compression? Lossy?

Which would you use for line art? For a color photograph?

Representing Video

Video codec COmpressor/DECompressor

Methods used to shrink the size of a movie to allow it to be played on a computer or over a network

Almost all video codecs use lossy compression to minimize the huge amounts of data associated with video

Representing Video

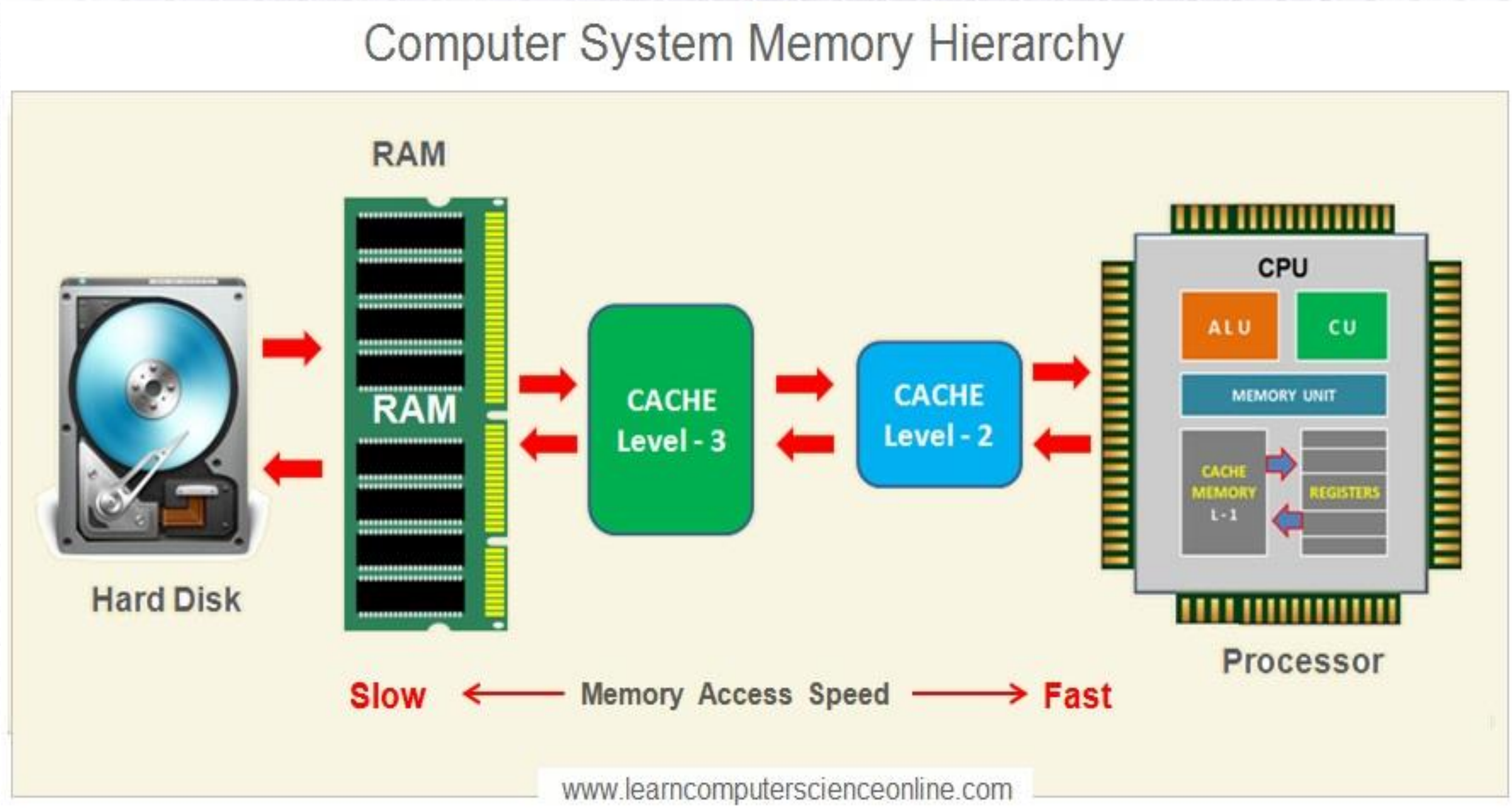
Temporal compression

A technique based on differences between consecutive frames: If most of an image in two frames has not changed, why should we waste space duplicating information?

Spatial compression

A technique based on removing repetitive information within a frame: This problem is essentially the same as that faced when compressing still images

Data Storage Devices



Bit, Byte, and Word

Byte

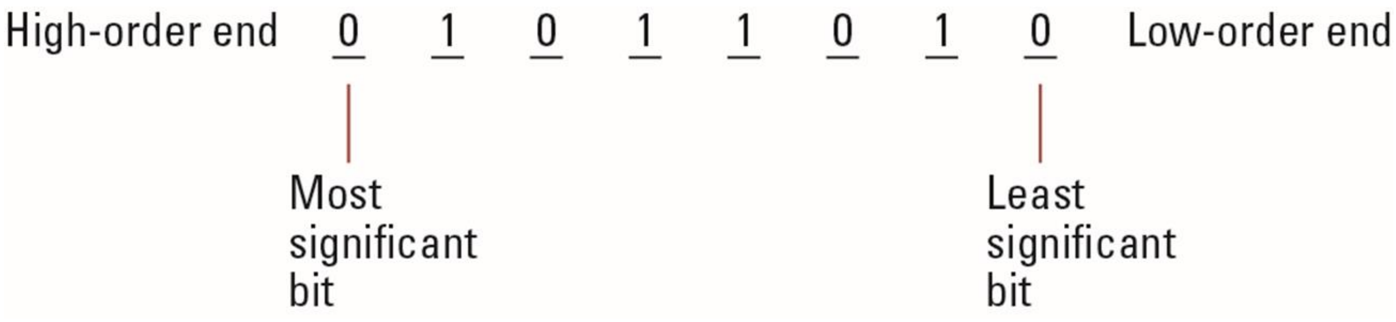
8 bits

The number of bits in a word determines the word length of the computer, which is usually a multiple of 8

- 32-bit machines
- 64-bit machines etc.

Main Memory

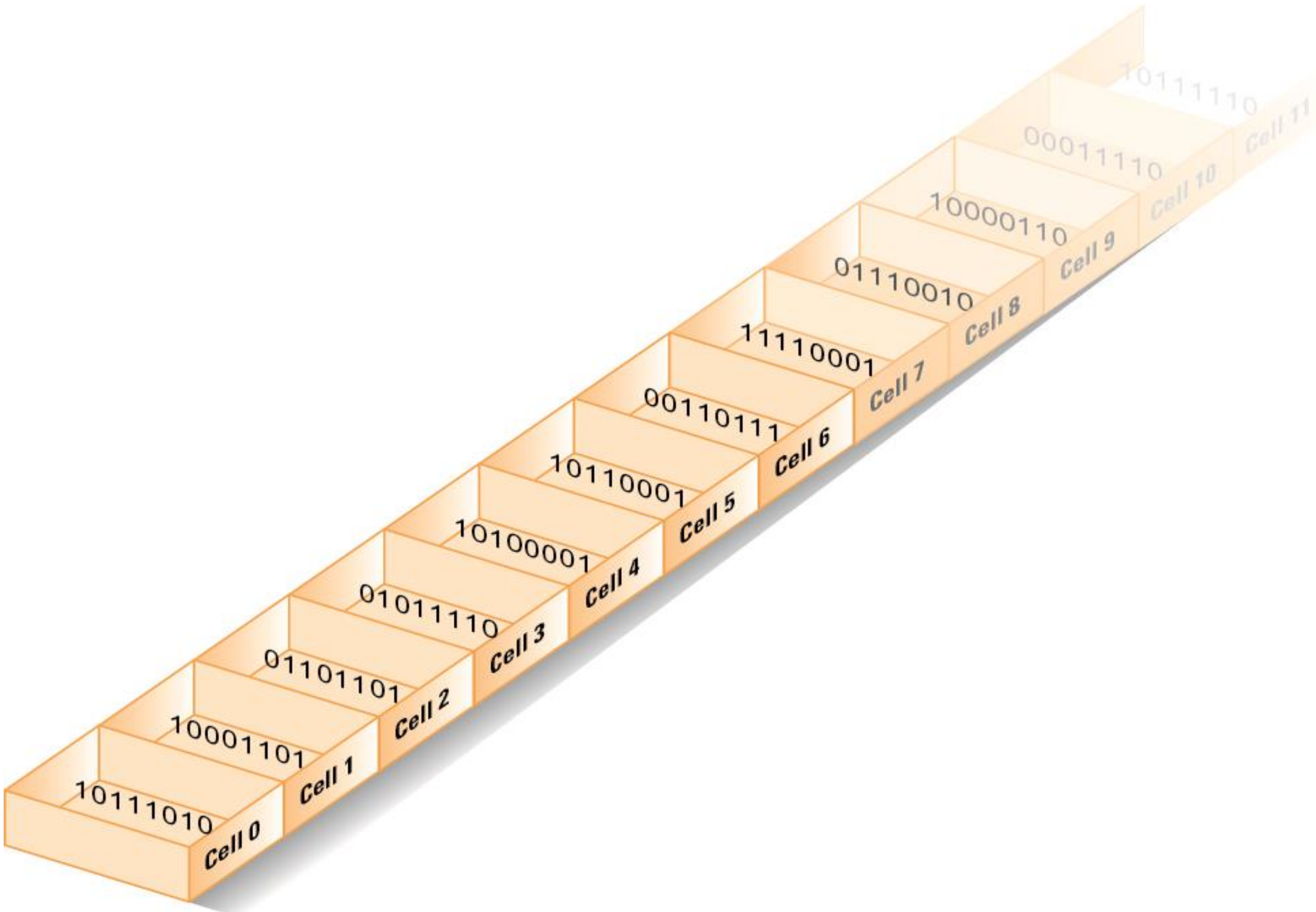
- **Cell:** A unit of main memory (typically 8 bits which is one **byte**)
  - **Most significant bit:** the bit at the left (high-order) end
  - **Least significant bit:** the bit at the right (low-order) end



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 cells arranged by address



Memory Terminology

- **Random Access Memory (RAM):** Memory in which individual cells can be easily accessed in any order
- **Dynamic Memory (DRAM):** RAM composed of volatile memory

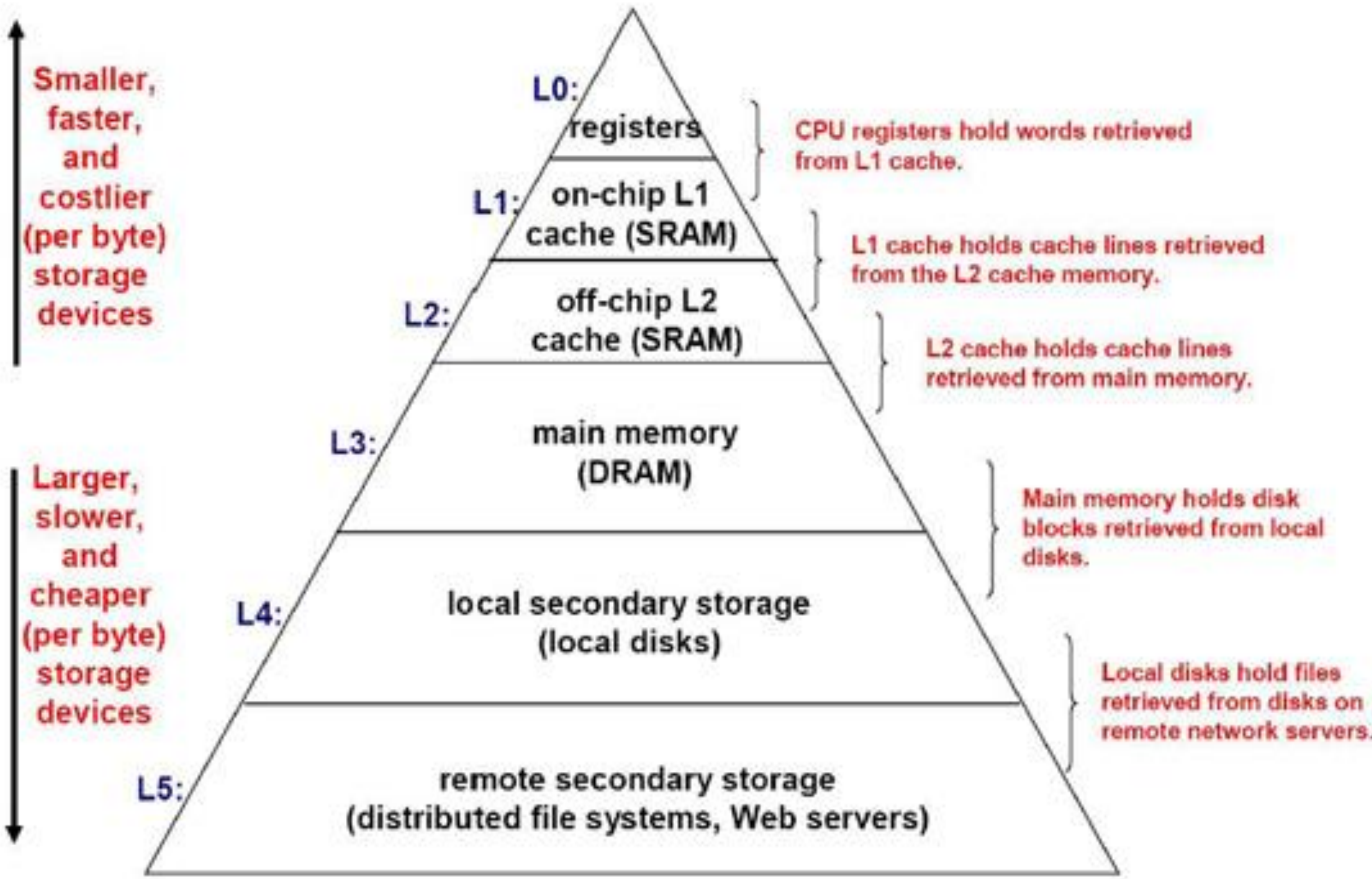
Measuring Memory Capacity

- **Kilobyte:** 2^{10} bytes = 1024 bytes
  - Example: 3 KB = 3 times1024 bytes
- **Megabyte:** 2^{20} bytes = 1,048,576 bytes
  - Example: 3 MB = 3 times 1,048,576 bytes
- **Gigabyte:** 2^{30} bytes = 1,073,741,824 bytes
  - Example: 3 GB = 3 times 1,073,741,824 bytes

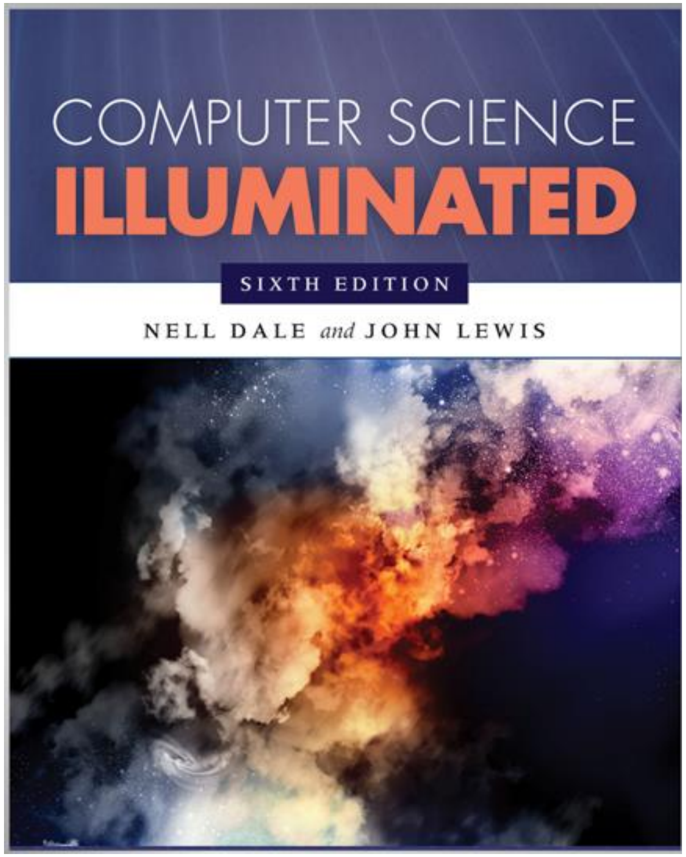
Mass Storage

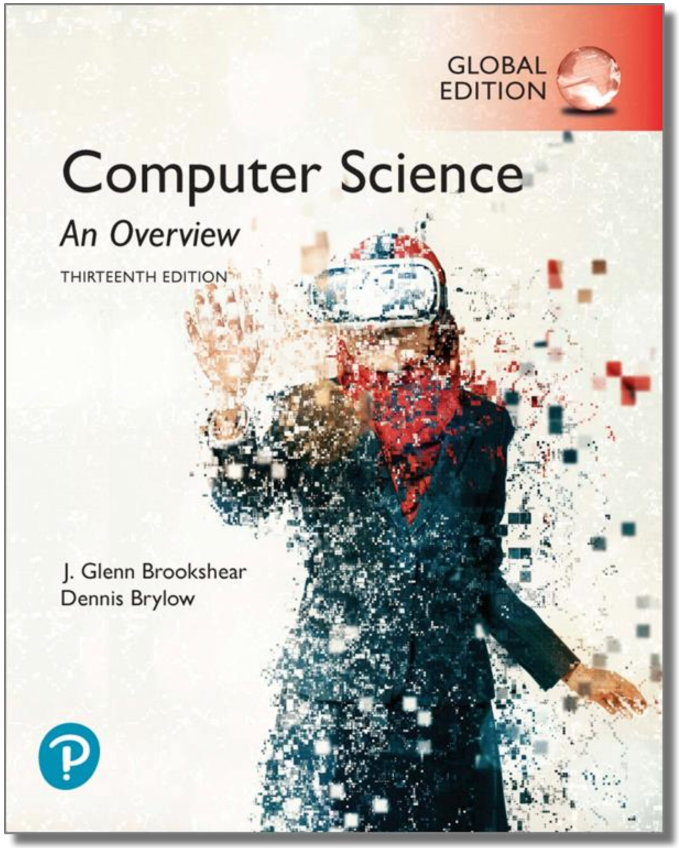
- Additional devices:
  - Magnetic disks
  - CDs
  - DVDs
  - **Magnetic tapes**
  - **Flash drives**
  - **Solid-state drives**
- Advantages over main memory
  - Less volatility
  - Larger storage capacities
  - Low cost
  - In many cases can be removed

Memory Hierarchy



References


COMPUTER SCIENCE
ILLUMINATED
SIXTH EDITION
NELL DALE and JOHN LEWIS


GLOBAL EDITION
Computer Science
An Overview
THIRTEENTH EDITION
J. Glenn Brookshear
Dennis Brylow