

ECE 2050 Digital Logic and Systems

# Chapter 2 : Number Systems

Instructor: Tinghuan CHEN



# Announcements

- There are tutorials for the second week;



# Last Week

- ❑ Analog versus Digital
- ❑ Bits (Binary digits), Logic Levels and Digital Waveforms
- ❑ Basic logic functions: NOT, AND and OR
- ❑ Combinational & sequential logic functions:  
comparator, adder, encoder/decoder, (de)multiplexer,  
flip-flops, registers, counter
- ❑ Integrated circuit (IC): Programmable versus Fixed-function
  - ◆ Package: Surface-mounted and Through-hole
  - ◆ Programmable: PLD (SPLD and CPLD) and FPGA
  - ◆ Fixed-function : SSI/MSI/VLSI/ULSI



# Binary Numbers



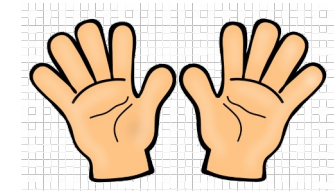
# Number Systems

- A system of writing to express numbers
- Daily life: 10, 12, 60, ...
- Computer systems: 2, 8, 16, ...



# Decimal Numbers

1's column  
10's column  
100's column  
1000's column



$$5374_{10} = 5 \times 10^3 + 3 \times 10^2 + 7 \times 10^1 + 4 \times 10^0$$

five thousands      three hundreds      seven tens      four ones

Base-10 Numbers

# Base-R Number System

- **Base:**  $R$  ( $R \geq 2$ )
- **Digit:**  $[0, R-1]$
- **Weight:**  $R^x$

$$\begin{aligned}(N)_R &= a_{n-1} a_{n-2} \dots a_1 a_0 a_{-1} \dots a_{-m} \\&= a_{n-1} \times R^{n-1} + a_{n-2} \times R^{n-2} + \dots + a_1 \times R^1 + a_0 \times R^0 + a_{-1} \times R^{-1} + \dots + a_{-m} \times R^{-m} \\&= \sum_{i=-m}^{n-1} a_i \times R^i\end{aligned}$$

- **Decimal numbers:** base-10
- **Binary numbers:** base-2
- **Octal numbers:** base-8
- **Hexadecimal numbers:** base-16



# Decimal Number System

Base: 10

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

Weight: ... $10^2$   $10^1$   $10^0$   $10^{-1}$   $10^{-2}$   $10^{-3}$ ...

Fractional numbers

Decimal point

Decimal numbers in digital systems mean any base 10 numbers, not just those with a decimal point.

$$\begin{aligned} 568.23 &= (5 \times 10^2) + (6 \times 10^1) + (8 \times 10^0) + (2 \times 10^{-1}) + (3 \times 10^{-2}) \\ &= (5 \times 100) + (6 \times 10) + (8 \times 1) + (2 \times 0.1) + (3 \times 0.01) \\ &= 500 + 60 + 8 + 0.2 + 0.03 \end{aligned}$$





# Binary Number System

Base: 2

Digit: 0, 1

Weight: ... $2^2$   $2^1$   $2^0$   $2^{-1}$   $2^{-2}$   $2^{-3}$  ...

↑  
binary point

Binary-to-Decimal Conversion:

$$\begin{aligned}(10.101)_2 &= (1 \times 2^1) + (0 \times 2^0) + (1 \times 2^{-1}) + (0 \times 2^{-2}) + (1 \times 2^{-3}) \\&= (1 \times 2) + (0 \times 1) + (1 \times 0.5) + (0 \times 0.25) + (1 \times 0.125) \\&= 2 + 0 + 0.5 + 0 + 0.125 \\&= (2.625)_{10}\end{aligned}$$



# Powers of Two

- $2^0 = 1$
- $2^1 = 2$
- $2^2 = 4$
- $2^3 = 8$
- $2^4 = 16$
- $2^5 = 32$
- $2^6 = 64$
- $2^7 = 128$

Handy to  
memorize



# Counting in Binary

**Binary**

0	0
1	1
10	2
11	3
100	4
101	5
110	6
111	7

**Decimal**

1-Bit Binary Numbers	2-Bit Binary Numbers	3-Bit Binary Numbers	4-Bit Binary Numbers	Decimal Equivalents
0	00	000	0000	0
1	01	001	0001	1
	10	010	0010	2
	11	011	0011	3
		100	0100	4
		101	0101	5
		110	0110	6
		111	0111	7
			1000	8
			1001	9
			1010	10
			1011	11
			1100	12
			1101	13
			1110	14
			1111	15



# Number Conversion

- Binary to decimal conversion:
  - Convert  $10011_2$  to decimal
  - $1 \times 16 + 0 \times 8 + 0 \times 4 + 1 \times 2 + 1 \times 1 = 19_{10}$
- Decimal to binary conversion:
  - Convert  $53_{10}$  to binary



# Decimal to Binary Conversion

- Two methods:
  - **Method 1:** Find the largest power of 2 that fits, subtract and repeat
  - **Method 2:** Repeatedly divide by 2, remainder goes in next most significant bit



# Decimal to Binary Conversion

- $53_{10}$
- **Method 1:** Find the largest power of 2 that fits, subtract and repeat
  - $53_{10}$   $32 \times 1$
  - $53 - 32 = 21$   $16 \times 1$
  - $21 - 16 = 5$   $4 \times 1$
  - $5 - 4 = 1$   $1 \times 1$   $= 110101_2$
- **Method 2:** Repeatedly divide by 2, remainder goes in next most significant bit
  - $53_{10} = 53/2 = 26 \text{ R}1$
  - $26/2 = 13 \text{ R}0$
  - $13/2 = 6 \text{ R}1$
  - $6/2 = 3 \text{ R}0$
  - $3/2 = 1 \text{ R}1$
  - $1/2 = 0 \text{ R}1$   $= 110101_2$



# Binary Values and Range

- **$N$ -digit decimal number**
  - How many values?  $10^N$
  - Range?  $[0, 10^N - 1]$
  - Example: 3-digit decimal number:
    - $10^3 = 1000$  possible values
    - Range:  $[0, 999]$
- **$N$ -bit binary number**
  - How many values?  $2^N$
  - Range:  $[0, 2^N - 1]$
  - Example: 3-digit binary number:
    - $2^3 = 8$  possible values
    - Range:  $[0, 7] = [000_2 \text{ to } 111_2]$



# Octal & Hexadecimal Numbers





# Octal Number System

Base: 8

Digit: 0, 1, 2, 3, 4, 5, 6, 7

Weight: ... $8^2$   $8^1$   $8^0$   $8^{-1}$   $8^{-2}$   $8^{-3}$ ...

Octal-to-Decimal Conversion:

$$\begin{aligned}(2374)_8 &= (2 \times 8^3) + (3 \times 8^2) + (7 \times 8^1) + (4 \times 8^0) \\ &= (2 \times 512) + (3 \times 64) + (7 \times 8) + (4 \times 1) \\ &= 1024 + 192 + 56 + 4 \\ &= (1276)_{10}\end{aligned}$$



# Hexadecimal Number System

Base: 16

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

Weight: ... $16^2$   $16^1$   $16^0$   $16^{-1}$   $16^{-2}$   $16^{-3}$ ...

Hex-to-Decimal Conversion:

$$\begin{aligned}(E5)_{16} &= (E \times 16^1) + (5 \times 16^0) \\ &= (14 \times 16) + (5 \times 1) \\ &= 224 + 5 \\ &= (229)_{10}\end{aligned}$$



# Decimal-to-Octal Conversion

Repeated division by 8 with the 1<sup>st</sup> remainder being the least significant digit (LSD)

8	359	remainder	
8	44	7	LSD
8	5	4	
	0	5	MSD

$$(359)_{10} = (547)_8$$



# Decimal-to-Hexadecimal Conversion

Repeated division by 16 with the 1<sup>st</sup> remainder being the least significant digit (LSD)

16	650	remainder	
16	40	10 (A)	LSD
16	2	8	
	0	2	MSD

$$(650)_{10} = (28A)_{16}$$



# Conversion btw. Binary, Octal, & Hex. Numbers

- Octal/Hex  $\rightarrow$  Binary
  - Replace each octal (hexadecimal) digit with 3 (4) bits

## □ Octal-to-Binary

7 5 2 6  
↓ ↓ ↓ ↓  
111101010110

$$7526_8 = 111101010110_2$$

## □ Hexadecimal-to-Binary

C F 8 E  
↓ ↓ ↓ ↓  
1100111110001110

$$CF8E_{16} = 1100111110001110_2$$

# Conversion btw. Binary, Octal, & Hex. Numbers

- Binary  $\rightarrow$  Octal/Hex
  - Convert each 3-bit (4-bit) group to the equivalent octal (hexadecimal) digit
  - Start from right-most, move from right to left, leading 0s

## □ Binary-to-Octal

$$\begin{array}{ccccccc} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ \hline \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & 6 & 3 & 2 & & & & & & & & \end{array} = 4632_8$$

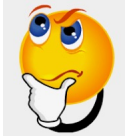
$$100110011010_2 = 4632_8$$

## □ Binary-to-Hexadecimal

$$\begin{array}{ccccccc} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ \hline \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ C & A & 5 & 7 & & & & & & & & & & & & \end{array} = CA57_{16}$$

$$1100101001010111_2 = CA57_{16}$$

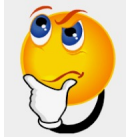
# Why do computers use Binary?



Boolean Logic &  
Logic Gates: Crash  
Course Computer  
Science #3

<https://thecrashcourse.com/topic/computerscience/>





# Why Octal and Hexdecimal?

- A more convenient/compact way to represent large binary numbers
- Easy Conversion from/to Binary
- Examples:
  - Octal: file permission in Linux
  - Hex: memory address





# Bytes, Nibbles, & All that Jazz



# Bits, Bytes, Nibbles ...

- Byte: 8 bits
  - Represents one of \_\_\_\_\_ values
  - [\_\_, \_\_]
- Nibble: 4 bits
  - Represents one of \_\_\_\_\_ values
  - [\_\_, \_\_]

One binary digit is \_\_ bit

One hex digit is \_\_\_\_ bits or \_\_\_\_ nibble

Two hex digits make \_\_\_\_ byte

Most significant on left

Least significant on right

10010110

most significant bit      least significant bit

byte

10010110

nibble

CEBF9AD7

most significant byte      least significant byte



# Large Powers of Two

- $2^{10} = 1$  kilo  $\approx 10^3$  (1024)
- $2^{20} = 1$  mega  $\approx 10^6$  (1,048,576)
- $2^{30} = 1$  giga  $\approx 10^9$  (1,073,741,824)
- $2^{40} = 1$  tera  $\approx 10^{12}$
- $2^{50} = 1$  peta  $\approx 10^{15}$
- $2^{60} = 1$  exa  $\approx 10^{18}$



# Estimating Powers of Two

- What is the value of  $2^{24}$ ?

$$2^4 \times 2^{20} \approx 16 \text{ million}$$

- How large of a value can a 32-bit integer variable represent?

$$2^2 \times 2^{30} \approx 4 \text{ billion}$$



# Binary Arithmetic: $+$ $-$ $\times$ $\div$



# Addition

- Decimal

$$\begin{array}{r} 11 \leftarrow \text{carries} \\ 3734 \\ + 5168 \\ \hline 8902 \end{array}$$

- Binary

$$\begin{array}{r} 11 \leftarrow \text{carries} \\ 1011 \\ + 0011 \\ \hline 1110 \end{array}$$

## Addition rules

$0 + 0 = 0$	Sum of 0 with a carry of 0
$0 + 1 = 1$	Sum of 1 with a carry of 0
$1 + 0 = 1$	Sum of 1 with a carry of 0
$1 + 1 = 10$	Sum of 0 with a carry of 1



# Binary Addition Examples

- Add the following 4-bit binary numbers

$$\begin{array}{r} 1001 \\ + 0101 \\ \hline \end{array}$$

- Add the following 4-bit binary numbers

$$\begin{array}{r} 1011 \\ + 0110 \\ \hline \end{array}$$



# Overflow

- Digital systems operate on a **fixed number of bits**
- Overflow: when result is too big to fit in the available number of bits
- See previous example of  $11 + 6$





# Subtraction

$$\begin{array}{r} 101 \\ -011 \\ \hline 010 \end{array}$$

## Subtraction rules

$$0 - 0 = 0$$

$$1 - 1 = 0$$

$$1 - 0 = 1$$

$$10 - 1 = 1 \quad 0 - 1 \text{ with a borrow of } 1$$



# Multiplication

$$\begin{array}{r} 111 \\ \times 101 \\ \hline 111 \\ 000 \\ +111 \\ \hline 10011 \end{array}$$

Partial products

## Multiplication Rules

$$0 \times 0 = 0$$

$$0 \times 1 = 0$$

$$1 \times 0 = 0$$

$$1 \times 1 = 1$$



# Division

- Decimal

$$\begin{array}{r} 25 \\ 5 \overline{)125} \\ \underline{10} \phantom{0} \\ 25 \\ \underline{25} \\ 0 \end{array}$$

- Binary

$$\begin{array}{r} 11 \\ 10 \overline{)110} \\ \underline{10} \phantom{0} \\ 10 \\ \underline{10} \\ 00 \end{array}$$

$$\begin{array}{r} 11 \\ 11 \overline{)1001} \\ \underline{11} \phantom{00} \\ 11 \\ \underline{11} \\ 00 \end{array}$$



# Signed Numbers



# Signed Binary Number

- Sign/Magnitude Numbers
- Two's Complement Numbers



# Sign/Magnitude Numbers

- 1 sign bit,  $N-1$  magnitude bits
- Sign bit is the most significant (left-most) bit
  - Positive number: sign bit = 0
  - Negative number: sign bit = 1
- Example, 4-bit sign/mag representations of  $\pm 6$ :
  - +6 = **0110**
  - 6 = **1110**
- Range of an  $N$ -bit sign/magnitude number:  
 **$[-(2^{N-1}-1), 2^{N-1}-1]$**

$$A: \{a_{N-1}, a_{N-2}, \dots, a_2, a_1, a_0\}$$

$$A = (-1)^{a_{N-1}} \sum_{i=0}^{N-2} a_i 2^i$$



# Sign/Magnitude Numbers

## Problems:

- Addition doesn't work, for example  $-6 + 6$ :

$$\begin{array}{r} 1110 \\ + 0110 \\ \hline 10100 \text{ (wrong!)} \end{array}$$

- Two representations of 0 ( $\pm 0$ ):

1000  
0000



# Two's Complement Numbers

- Don't have same problems as sign/magnitude numbers:
  - **Addition works**
  - **Single representation for 0**





# Two's Complement Numbers

- MSB has weight of  $-2^{N-1}$

$$A = a_{N-1}(-2^{N-1}) + \sum_{i=0}^{N-2} a_i 2^i$$

- Most positive 4-bit number: **0111**
- Most negative 4-bit number: **1000**
- The most significant bit still indicates the sign (1 = negative, 0 = positive)
- Range of an  $N$ -bit two's complement number:

$$[-(2^{N-1}), 2^{N-1}-1]$$

Decimal	2's Complement
-1	11111111
-2	11111110
-3	11111101
-4	11111100
...	...
-128	10000000



# Reverse the Sign

- **Reverse the sign** of a two's complement number
- **Method:**
  1. Invert the bits
  2. Add 1
- **Example:** Reverse the sign of  $3_{10} = 0011_2$ 
  1. 1100
  2. + 1

1101 =  $-3_{10}$

Historically, this reversing the sign method has been called: “Taking the Two's complement”. But this terminology can be confusing, so we instead we call it “reversing the sign”.



# Two's Complement Examples

- Reverse the sign of  $6_{10} = 0110_2$

1. 1001

2.  $\begin{array}{r} + \quad 1 \\ \hline \end{array}$

$$1010_2 = -6_{10}$$

- What is the decimal value of the two's complement number  $1001_2$ ?

1. 0110

2.  $\begin{array}{r} + \quad 1 \\ \hline \end{array}$

$$0111_2 = 7_{10}, \text{ so } 1001_2 = -7_{10}$$



# Two's Complement Addition

## □ 3-step procedures:

- 1 Convert **the decimal numbers** into the binary form with negative numbers expressed in the 2's complement form
- 2 Perform binary addition
- 3 Sign: If both numbers positive (negative) → the result positive (negative).  
If the sign is incorrect, **overflow** has occurred!

**Both numbers positive**

00000111	← 1	7
+ 00000100	← 2	+ 4
<hr/>		
00001011	3	11

**Both numbers negative**

$5_{10} = 00000101_2$	1	11111011	<table><tr><td>-5</td></tr><tr><td>+ -9</td></tr><tr><td><hr/></td></tr><tr><td>-14</td></tr></table>	-5	+ -9	<hr/>	-14
-5							
+ -9							
<hr/>							
-14							
$9_{10} = 00001001_2$	2's complement	+ 11110111					
<hr/>							
Discard carry → 1		11110010	2				
		3					

↓

$00001110_2 = 14_{10}$



# Two's Complement Addition

## □ 3-step procedures:

- 1 Convert **the decimal numbers** into the binary form with negative numbers expressed in the 2's complement form
- 2 Perform binary addition
- 3 Sign: If both numbers positive (negative) → the result positive (negative).  
If the sign is incorrect, **overflow** has occurred!

8-bit numbers [-128, +127]

$$\begin{array}{r} 01111101 \\ + 00111010 \\ \hline 10110111 \end{array} \quad \begin{array}{r} 125 \\ + 58 \\ \hline 183 \end{array}$$

↑  
incorrect

$$\begin{array}{r} 10000011 \\ + 11000110 \\ \hline 101001001 \end{array} \quad \begin{array}{r} -125 \\ + -58 \\ \hline -183 \end{array}$$

↑  
incorrect



# Two's Complement Addition: one negative number

$$\begin{array}{r} 00010000 \\ \textcircled{2} + 11101000 \\ \hline 11111000 \end{array}$$

↓ 2's complement

$$\textcircled{3} \quad 00001000_2 = 8_{10} \rightarrow -8$$

$$\begin{array}{r} 16 \\ + -24 \\ \hline -8 \end{array}$$

$$16_{10} = 00010000_2$$

$$24_{10} = 00011000_2$$

$$-24_{10} = 11101000_2$$

No carry indicates the result is negative

$$\begin{array}{r} 00001111 \\ \textcircled{2} + 11111010 \\ \hline \text{Carry} \rightarrow 1 \quad 00001001 \end{array}$$

$$\textcircled{3} \quad 00001001_2 = 9_{10}$$

$$\begin{array}{r} 15 \\ + -6 \\ \hline 9 \end{array}$$

$$15_{10} = 00001111_2$$

$$6_{10} = 00000110_2$$

$$-6_{10} = 11111010_2$$

Carry indicates the result is positive



# Subtraction

- Subtract a 2's complement number by reversing the sign and adding.
- Reverse sign by taking 2's complement

- Ex:  $3 - 5 = 3 + (-5)$

$$\begin{array}{r} 0011 \quad 3 \\ + 1011 \quad -5 \\ \hline 1110 \quad -2 \end{array}$$

- Ex:  $7 - 3 = 7 + (-3)$

$$\begin{array}{r} 0111 \quad 7 \\ + 1101 \quad -3 \\ \hline 10100 \quad 4 \end{array}$$

Discard the carry



# Multiplication

## ❑ Direct addition :

- ◆ Very lengthy if the multiplier is a large number
- ◆ both numbers must be in true (uncomplemented) form.

## ❑ Partial product

- ◆ Same sign → Positive;
- ◆ Different signs → Negative

A Decimal Example

$$\begin{array}{r}
 239 \\
 \times 123 \\
 \hline
 717 \\
 478 \phantom{0} \\
 + 239 \phantom{00} \\
 \hline
 29,397
 \end{array}$$

## EXAMPLE 2-22

01010011 (multiplicand) and 11000101 (multiplier).

83

-59

1010011	Multiplicand
× 0111011	Multiplier
1010011	1st partial product
+ 1010011	2nd partial product
11111001	Sum of 1st and 2nd
+ 0000000	3rd partial product
011111001	Sum
+ 1010011	4th partial product
1110010001	Sum
+ 1010011	5th partial product
100011000001	Sum
+ 1010011	6th partial product
1001100100001	Sum
+ 0000000	7th partial product
01001100100001	Final product
10110011011111	2's complement
-4897	

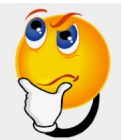
4897





# Division

- ❑ Division operation in computers is accomplished using subtraction
- ❑ Quotient and Remainder
  - ◆ When to stop the subtraction: the remainder is a zero or a negative number
  - ◆ Quotient: # of times of subtraction performed
- ❑ Sign of the quotient
  - ◆ Same sign → Positive;
  - ◆ Different signs → Negative



*What about dividing 100 by -25?*

**EXAMPLE 2-23** Divide 01100100 by 00011001

100

25

Step 1: 00011001  $\xrightarrow{\text{2's complement}}$  11100111

Step 2: 
$$\begin{array}{r} 01100100 \\ + 11100111 \\ \hline 01001011 \end{array}$$
 Add 1 to quotient

Step 3: 
$$\begin{array}{r} 01001011 \\ + 11100111 \\ \hline 00110010 \end{array}$$
 Add 1 to quotient

Step 4: 
$$\begin{array}{r} 00110010 \\ + 11100111 \\ \hline 00011001 \end{array}$$
 Add 1 to quotient

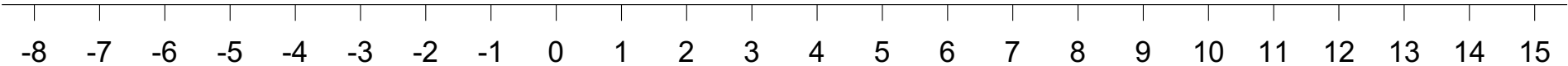
Step 5: 
$$\begin{array}{r} 00011001 \\ + 11100111 \\ \hline 00000000 \end{array}$$
 Add 1 to quotient  
Quotient = 4<sub>10</sub> = 0100



# Number System Comparison

Number System	Range
Unsigned	$[0, 2^N-1]$
Sign/Magnitude	$[-(2^{N-1}-1), 2^{N-1}-1]$
Two's Complement	$[-2^{N-1}, 2^{N-1}-1]$

For example, 4-bit representation:



Unsigned

0000 0001 0010 0011 0100 0101 0110 0111 1000 1001 1010 1011 1100 1101 1110 1111

1000 1001 1010 1011 1100 1101 1110 1111 0000 0001 0010 0011 0100 0101 0110 0111

Two's Complement

1111 1110 1101 1100 1011 1010 1001 0000  
1000 0001 0010 0011 0100 0101 0110 0111

Sign/Magnitude

# Fixed-Point Numbers



# Number Systems

- Numbers we can represent using binary representations
  - **Positive numbers**
    - Unsigned binary
  - **Negative numbers**
    - Two's complement
    - Sign/magnitude numbers
- What about **fractions**?



# Numbers with Fractions

- Two common notations:
- **Fixed-point:** binary point fixed
- **Floating-point:** binary point floats to the right of the most significant 1



# Fixed-Point Numbers

- 6.75 using 4 integer bits and 4 fraction bits:

01101100

0110.1100

$$2^2 + 2^1 + 2^{-1} + 2^{-2} = 6.75$$

- Binary point is implied
- The number of integer and fraction bits must be agreed upon beforehand



# Unsigned Fixed Point Formats

- **Ua.b**: unsigned number with
  - **a** integer bits
  - **b** fractional bits.
- **Example: 6.75** is
  - **U4.4**: 01101100
  - **U3.5**: 11011000
  - **U6.2**: 00011011
- 8, 16, and 32-bit fixed point numbers are common
  - **U8.8** often represents sensor data, audio, pixels
  - **U16.16** used for higher precision signal processing



# Signed Fixed Point Formats

- **Qa.b:** signed 2's complement number with
  - **a** integer bits (including the sign bit)
  - **b** fractional bits
- **To negate a Q fixed point number:**
  - Invert the bits
  - Add one to the LSB
- **Example:** write **-6.75** in Q4.4
  - 6.75 = 01101100
  - Invert: 10010011
  - Add 1 LSB: 10010100
- **Q1.15** (aka Q15) is common for signal processing (1, -1]



# Saturating Arithmetic

- Fixed point **overflow** is usually bad
  - Produces undesired artifacts:
    - Video: dark pixel in middle of bright pixels
    - Audio: clicking sounds
- **Saturating arithmetic**
  - Instead of overflowing, use largest value
  - In U4.4:  $11000000 + 01111000 = 11111111$ 
    - $12 + 7.5 = 15.9375$



# Floating-Point Numbers



# Floating-Point Numbers

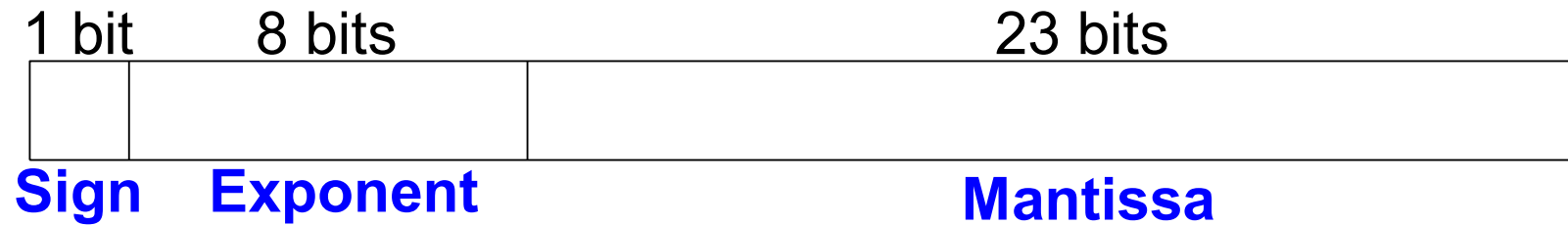
- Binary point floats to the right of the most significant 1
- Similar to decimal scientific notation
- For example, write  $273_{10}$  in scientific notation:
  - $273 = 2.73 \times 10^2$
- In general, a number is written in scientific notation as:
$$\pm M \times B^E$$
  - $M$  = mantissa
  - $B$  = base
  - $E$  = exponent
  - In the example,  $M = 2.73$ ,  $B = 10$ , and  $E = 2$

# Floating vs. Fixed Point Numbers

- Floating point numbers are like **scientific notation**
  - Allow a **greater dynamic range** of smallest to largest
  - **Arithmetic is harder**
    - Mantissa must be aligned before adding
    - This costs performance and power
- Fixed point numbers are **harder for the programmer**
  - **Smaller dynamic range**
  - Take care of **overflow**
- **Floating Point** is preferred for general-purpose computing where programming time is most important
- **Fixed Point** is preferred for signal processing performance, power, and hardware cost matter most
  - Machine learning, video



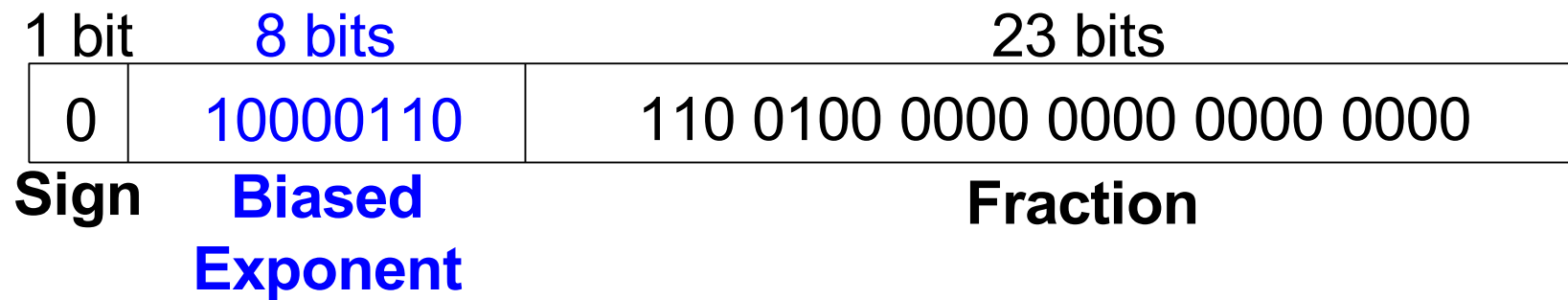
# Floating-Point Numbers



- **Example:** represent the value  $228_{10}$  using a 32-bit floating point representation

# IEEE 754 floating-point standard

- First bit of the mantissa is always 1:  $228_{10} = 11100100_2 = \mathbf{1.11001} \times 2^7$ 
  - So, no need to store it: *implicit leading 1*
- Store just **fraction bits** in 23-bit field
- *Biased exponent*: bias = 127 ( $01111111_2$ )
  - Biased exponent = bias + exponent
  - Exponent of 7 is stored as:  $127 + 7 = 134 = 0x10000110_2$



in hexadecimal: **0x43640000**



# Why Biased Exponent?

- The range of the biased exponent:  $[-127, +128]$ 
  - Extremely large or Small numbers can be represented in this way
- 



# Floating-Point Example

- Write  $-58.25_{10}$  in floating point (IEEE 754)

1. Convert magnitude of decimal to binary:

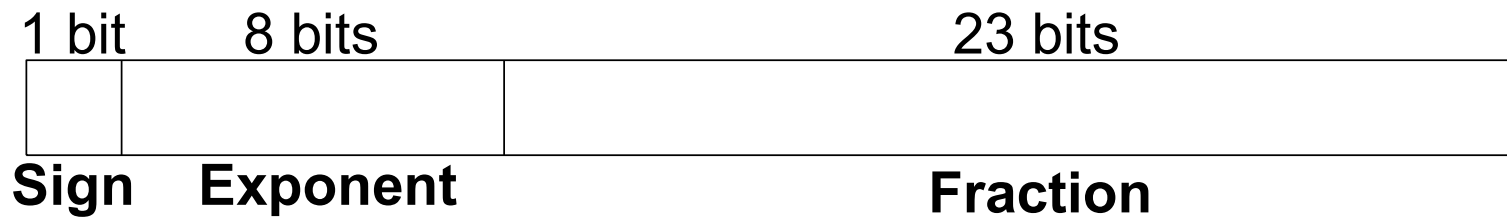
- $58.25_{10} = 111010.01_2$

2. Write in binary scientific notation:

- $1.1101001 \times 2^5$

3. Fill in fields:

- **Sign bit:** 1 (negative)
- **8 exponent bits:**  $(127 + 5) = 132 = 10000100_2$
- **23 fraction bits:** 110 1001 0000 0000 0000 0000



- 
- in hexadecimal: **0xC2690000**



# Floating-Point Special Cases

Number	Sign	Exponent	Fraction
0	X	00000000	000000000000000000000000
$\infty$	0	11111111	000000000000000000000000
$-\infty$	1	11111111	000000000000000000000000
NaN	X	11111111	Non-zero

# Non-positional/Non-weighted Number Systems



# Positional/Weighted Number System

## Positional/Weighted number system

The value of a symbol is determined by its position

Examples: decimal / binary / octal / hexadecimal number systems ...

## Non-positional/Non-weighted number system

The value of a symbol is NOT determined by its position

Examples: gray code, cyclic code ...



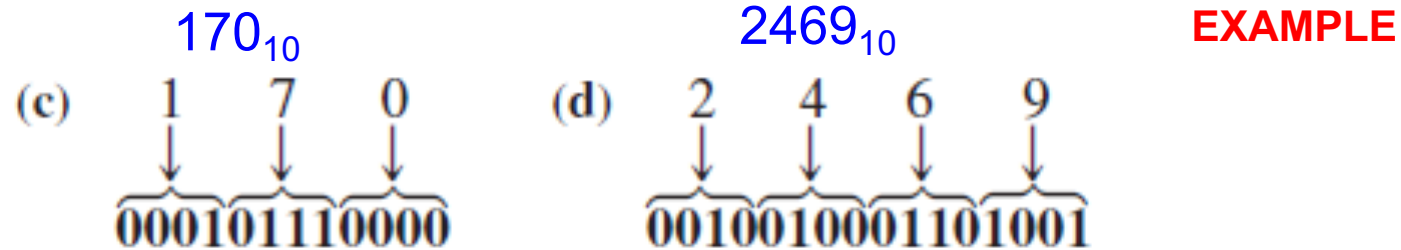
# Binary Coded Decimal (BCD)

- ❑ Express each of the decimal digits with a binary code.
- ❑ The 8421 code

**TABLE 2-5**

Decimal/BCD conversion.

Decimal Digit	0	1	2	3	4	5	6	7	8	9
BCD	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001



## ❑ BCD-to-Decimal Conversion

- ◆ Starting from the right-most bit and divide the code into groups of **four** bits



# BCD Addition

1. Add the two BCD numbers, using the rules for binary addition
2. If a 4-bit sum  $\leq 9 \rightarrow$  Valid BCD number
3. Otherwise, add 6 (0110) to the 4-bit sum

Decimal/BCD conversion.

Decimal Digit	0	1	2	3	4	5	6	7	8	9
BCD	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001

**EXAMPLE**

$$\begin{array}{r} 1001 \\ + 0100 \\ \hline 1101 \end{array} \quad \begin{array}{r} 9 \\ + 4 \\ \hline 13 \end{array}$$

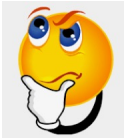
Invalid BCD number ( $> 9$ )

Add 6

Valid BCD number

$$\begin{array}{r} 0001 \\ + 0110 \\ \hline 0011 \end{array}$$

1 3

 Why adding 6?



# The Gray Code (I)

- ❑ Only a single bit change from one code word to the next [in sequence](#)

Four-bit Gray code.

Decimal	Binary	Gray Code	Decimal	Binary	Gray Code
0	0000	0000	8	1000	1100
1	0001	0001	9	1001	1101
2	0010	0011	10	1010	1111
3	0011	0010	11	1011	1110
4	0100	0110	12	1100	1010
5	0101	0111	13	1101	1011
6	0110	0101	14	1110	1001
7	0111	0100	15	1111	1000



# Gray Code: How to Transfer (II)

## EXAMPLE

(a) Convert the binary number 11000110 to Gray code.

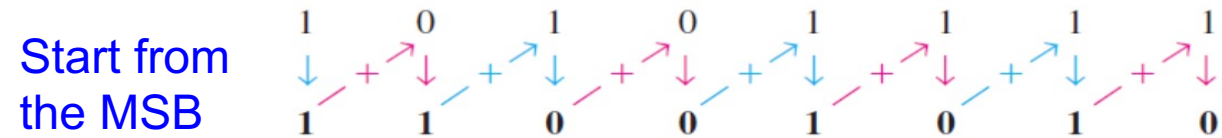
(b) Convert the Gray code 10101111 to binary.

### Solution

(a) Binary to Gray code:

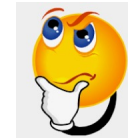


(b) Gray code to binary:



Encoding transmit symbols with the Gray Code

- ◆ Transmitter sends one of the four symbols;
- ◆ Transmitted symbols are distorted by noise;
- ◆ The receiver tries to decode the distorted symbols;
- ◆ The probability of making a symbol detection error is **inversely** proportional to the symbol distance

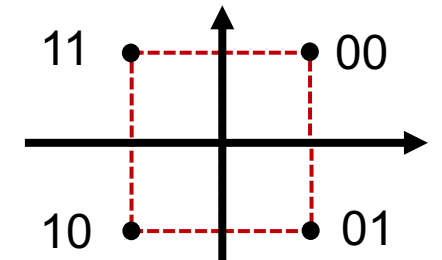


011 → 100,  
intermediate states:  
010, 001, 101, 110, 111

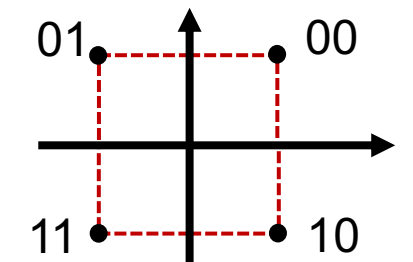
*Why Gray Code has an advantage?*

Constellation Diagram

Without gray code



With gray code



# Alphanumeric Codes

□ ASCII :

American

Standard

Code for

Information

Interchange

Control Characters				Graphic Symbols											
Name	Dec	Binary	Hex	Symbol	Dec	Binary	Hex	Symbol	Dec	Binary	Hex	Symbol	Dec	Binary	Hex
NUL	0	0000000	00	space	32	0100000	20	@	64	1000000	40	'	96	1100000	60
SOH	1	0000001	01	!	33	0100001	21	A	65	1000001	41	a	97	1100001	61
STX	2	0000010	02	”	34	0100010	22	B	66	1000010	42	b	98	1100010	62
ETX	3	0000011	03	#	35	0100011	23	C	67	1000011	43	c	99	1100011	63
EOT	4	0000100	04	\$	36	0100100	24	D	68	1000100	44	d	100	1100100	64
ENQ	5	0000101	05	%	37	0100101	25	E	69	1000101	45	e	101	1100101	65
ACK	6	0000110	06	&	38	0100110	26	F	70	1000110	46	f	102	1100110	66
BEL	7	0000111	07	,	39	0100111	27	G	71	1000111	47	g	103	1100111	67
BS	8	0001000	08	(	40	0101000	28	H	72	1001000	48	h	104	1101000	68
HT	9	0001001	09	)	41	0101001	29	I	73	1001001	49	i	105	1101001	69
LF	10	0001010	0A	*	42	0101010	2A	J	74	1001010	4A	j	106	1101010	6A
VT	11	0001011	0B	+	43	0101011	2B	K	75	1001011	4B	k	107	1101011	6B
FF	12	0001100	0C	,	44	0101100	2C	L	76	1001100	4C	l	108	1101100	6C
CR	13	0001101	0D	—	45	0101101	2D	M	77	1001101	4D	m	109	1101101	6D
SO	14	0001110	0E	.	46	0101110	2E	N	78	1001110	4E	n	110	1101110	6E
SI	15	0001111	0F	/	47	0101111	2F	O	79	1001111	4F	o	111	1101111	6F
DLE	16	0010000	10	0	48	0110000	30	P	80	1010000	50	p	112	1110000	70
DC1	17	0010001	11	1	49	0110001	31	Q	81	1010001	51	q	113	1110001	71
DC2	18	0010010	12	2	50	0110010	32	R	82	1010010	52	r	114	1110010	72
DC3	19	0010011	13	3	51	0110011	33	S	83	1010011	53	s	115	1110011	73
DC4	20	0010100	14	4	52	0110100	34	T	84	1010100	54	t	116	1110100	74
NAK	21	0010101	15	5	53	0110101	35	U	85	1010101	55	u	117	1110101	75
SYN	22	0010110	16	6	54	0110110	36	V	86	1010110	56	v	118	1110110	76
ETB	23	0010111	17	7	55	0110111	37	W	87	1010111	57	w	119	1110111	77
CAN	24	0011000	18	8	56	0111000	38	X	88	1011000	58	x	120	1111000	78
EM	25	0011001	19	9	57	0111001	39	Y	89	1011001	59	y	121	1111001	79
SUB	26	0011010	1A	:	58	0111010	3A	Z	90	1011010	5A	z	122	1111010	7A
ESC	27	0011011	1B	;	59	0111011	3B	[	91	1011011	5B	{	123	1111011	7B
FS	28	0011100	1C	<	60	0111100	3C	\	92	1011100	5C		124	1111100	7C
GS	29	0011101	1D	=	61	0111101	3D	]	93	1011101	5D	}	125	1111101	7D
RS	30	0011110	1E	>	62	0111110	3E	^	94	1011110	5E	~	126	1111110	7E
US	31	0011111	1F	?	63	0111111	3F	_	95	1011111	5F	Del	127	1111111	7F





# Error Codes: Parity Bit for Error Detection

- ❑ A parity bit is attached to a group of bits to make the total number of 1's in a group always even or always odd.

**TABLE 2-8**

The BCD code with parity bits.

Even Parity		Odd Parity	
<i>P</i>	BCD	<i>P</i>	BCD
0	0000	1	0000
1	0001	0	0001
1	0010	0	0010
0	0011	1	0011
1	0100	0	0100
0	0101	1	0101
0	0110	1	0110
1	0111	0	0111
1	1000	0	1000
0	1001	1	1001

## EXAMPLE

An **odd** parity system receives the following code groups: 10110, 11010, 110011, 110101110100, and 1100010101010. Determine which groups, if any, are in error.

	# of 1's
10110	3
11010	3
<b>110011</b>	4
110101110100	7
<b>1100010101010</b>	6

Since odd parity is required, any group with an even number of 1s is incorrect. The following groups are in error: 110011 and 1100010101010.



# Error Codes: Cyclic Redundancy Check

- ❑ Cyclic Redundancy Check (CRC) : an error detection method that can detect multiple errors in data blocks
- ❑ At the sending end, a **checksum** is appended to a block of data.
- ❑ At the receiving end, the check sum is generated and compared to the sent checksum. **If the check sums are zeros, no error is detected.**

D: 11010011 Data  
G: 1010 Generator code

**Step 1:** Append 000 to the data

D' = 11010011000

**Step 2:**  $\frac{D'}{G} = \frac{11010011000}{1010}$

11010011000  
1010  
-----  
1110  
1010  
-----  
1000  
1010  
-----  
1011  
1010  
-----  
1000  
1010  
-----  
10

Modulo-2  
operation  
(XOR)

Non-zero  
remainder  
Error detected!

One-bit error at  
the receiving end

10010011010  
1010  
-----  
1100  
1010  
-----  
1101  
1010  
-----  
1111  
1010  
-----  
1010  
1010  
-----  
010

Remainder (Checksum) : 010

**Step 3:** The transmitted CRC is 11010011010



# Chapter Review

- ☐ Base-R Number System
  - ☐ Decimal Number System
  - ☐ Binary Number System
  - ☐ Number Conversion
- ☐ Octal & Hexadecimal Numbers
  - ☐ Conversion btw. Binary, Octal, & Hex. Numbers
- ☐ Binary Arithmetic
- ☐ Signed Numbers and Two's Complement Numbers
- ☐ Fixed-Point Numbers & Floating-Point Numbers
- ☐ Binary Coded Decimal and Gray Code
- ☐ Error Codes



# True/False Quiz



The octal number system is a weighted system with eight digits.



The binary number system is a weighted system with two digits.



MSB stands for most significant bit.



In hexadecimal,  $9 + 1 = 10$ .



The 2's complement of the binary number 1111 is 0000.



The right-most bit in a signed binary number is the sign bit.



The hexadecimal number system has 16 characters, six of which are alphabetic characters.



BCD stands for binary coded decimal.



An error in a given code can be detected by verifying the parity bit.



CRC stands for cyclic redundancy check.