CSARCH Lecture Series: IEEE-754 Floating-Point Rounding

Sensei RL Uy
College of Computer Studies
De La Salle University
Manila, Philippines

Copyright Notice

Overview

Reflect on the following questions:

- How many significant digits are there in decimal 1300?
- How many rounding methods do you know?

Overview

- This sub-module introduces various IEEE-754 rounding methods
- The objectives are as follows:
  - ✓Describe the process of determining significant figures in a number
  - ✓Describe the process of performing different types of IEEE-754 rounding methods

Significant figures

- All non-zero are significant.
  - 91: 2 significant digits, 123.45: 5 significant digits
  - 11_2: 2 significant bits; 111.11_2: 5 significant bits
- Zeroes appearing between 2 non-zero digits are significant.
  - 607.1203: 7 significant digits
  - 101.101_2: 6 significant bits

Significant figures

- Leading zeroes are NOT significant
  - 0.0053: 2 significant digits
  - 0.0011_2 : 2 significant bits
- Trailing zeroes in a number containing a radix point are significant
  - 12.2300: 6 significant digits; 0.000122300: 6 significant digits
  - 101.1100_2: 7 significant bits; 0.0011100_2: 5 significant bits

Significant figures

- The significance of trailing zeros in a number not containing a radix point can be ambiguous. Thus 1300 can mean 2 or 4 significant figures.
  - A decimal point maybe placed after the number to denote significance
    - 1300. means 4 significant digits
  - Use scientific notation
    - 1.3×10^2 denotes 2 significant digits while 1.300×10^2 denotes 4 significant digits
  - A bar or underline maybe placed over the last significant figure
    - 1$\underline{3}$00 denotes 2 significant digits

Significant figures

- The significance of trailing zeros in a number not containing a radix point can be ambiguous. Thus 1100_2 can mean 2 or 4 significant figures.
  - A decimal point maybe placed after the number to denote significance
    - $1100._2$ means 4 significant bits
  - Use scientific notation
    - $1.1_2 \times 2^2$ denotes 2 significant bits while $1.100_2 \times 2^2$ denotes 4 significant bits
  - A bar or underline maybe placed over the last significant figure
    - $1\underline{1}00_2$ denotes 2 significant bits

Try →Pause your video first before answering

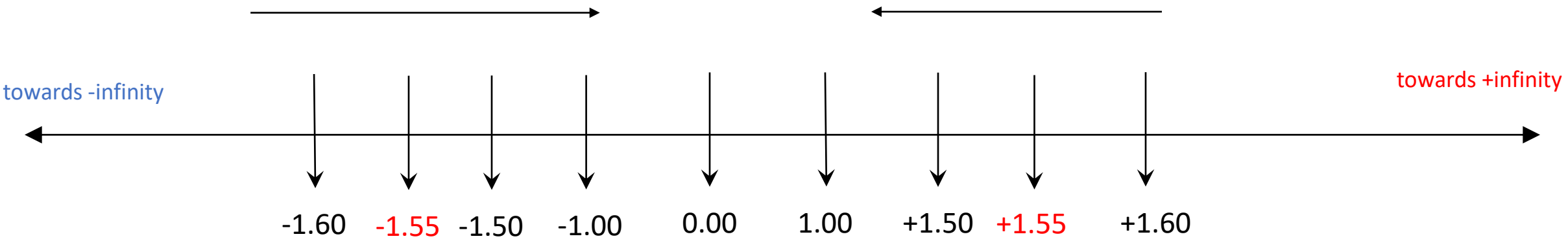
Decimal	Significant digits
59.75	4
23.0570	6
530.000	6
0.000045	2

Binary	Significant bits
1.000000	7
111	3
101.0110	7
0.00001	1

IEEE-754 Rounding rules

Round towards 0	2 digits
1.55	1.5
-1.55	-1.5

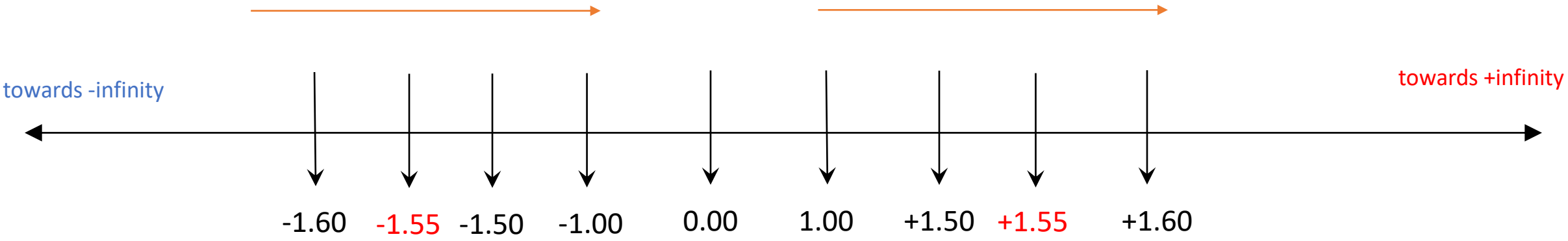
- Directed Rounding
  - Round towards 0: also known as *truncation*



IEEE-754 Rounding rules

Round towards +infinity	2 digits
1.55	1.6
-1.55	-1.5

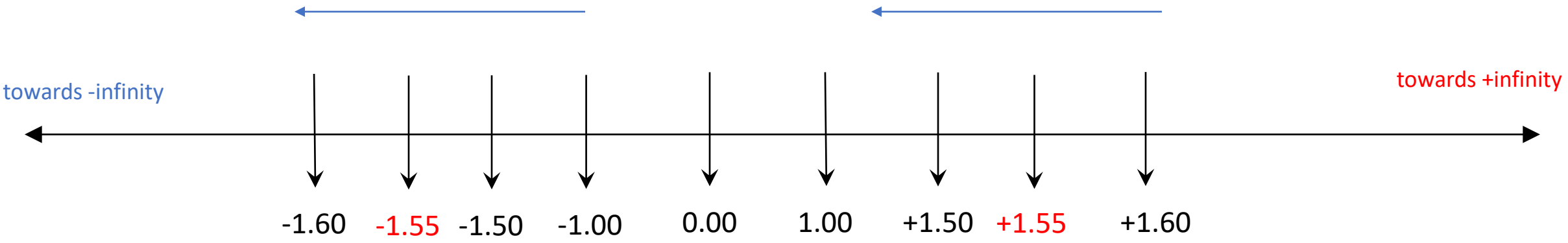
- Directed Rounding
  - Round towards +infinity: also known as *rounding up* or *ceiling*



IEEE-754 Rounding rules

Round towards -infinity	2 digits
1.55	1.5
-1.55	-1.6

- Directed Rounding
  - Round towards –infinity: also known as *round down* or *floor*

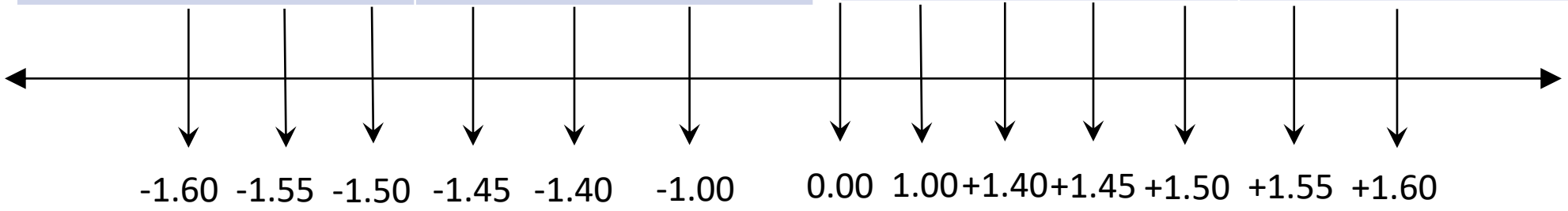


IEEE-754 Rounding rules

- Rounding to nearest
  - Round to nearest, ties away from zero

Ties away from zero	2 digits
-1.54	-1.5
-1.55	-1.6
-1.56	-1.6

Ties away from zero	2 digits
1.54	1.5
1.55	1.6
1.56	1.6



Ties away from zero	2 digits
-1.44	-1.4
-1.45	-1.5
-1.46	-1.5

Ties away from zero	2 digits
1.44	1.4
1.45	1.5
1.46	1.5

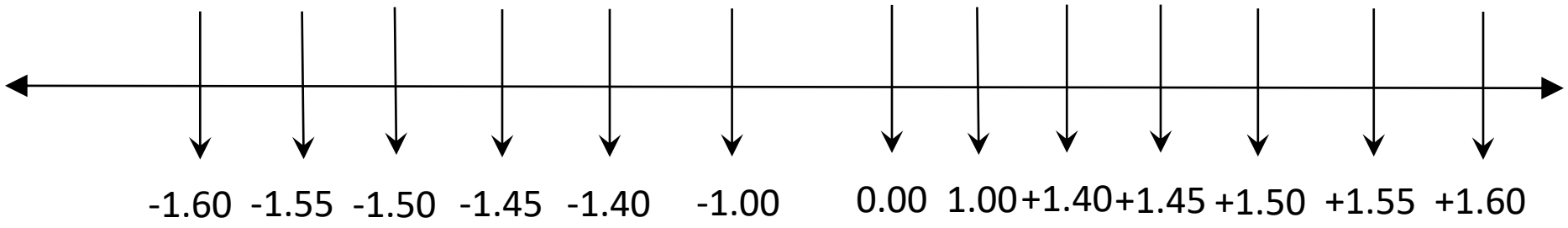
1.5
1.51
1.52
1.53
1.54
1.55 (tie)
1.56
1.57
1.58
1.59
1.6

Round to nearest, ties to even

- Round to the nearest value
- If the number falls midway it is rounded to the nearest value with an even least significant bit
- Default for binary, recommended default for decimal
- Also known as banker’s rounding, unbiased rounding, convergent rounding, statistician’s rounding, Dutch rounding, Gaussian rounding, odd-even rounding, broken rounding

IEEE-754 Rounding rules

- Rounding to nearest
  - Round to nearest, ties to even



-1.6
-1.59
-1.58
-1.57
-1.56
-1.55 (tie)
-1.54
-1.53
-1.52
-1.51
-1.5

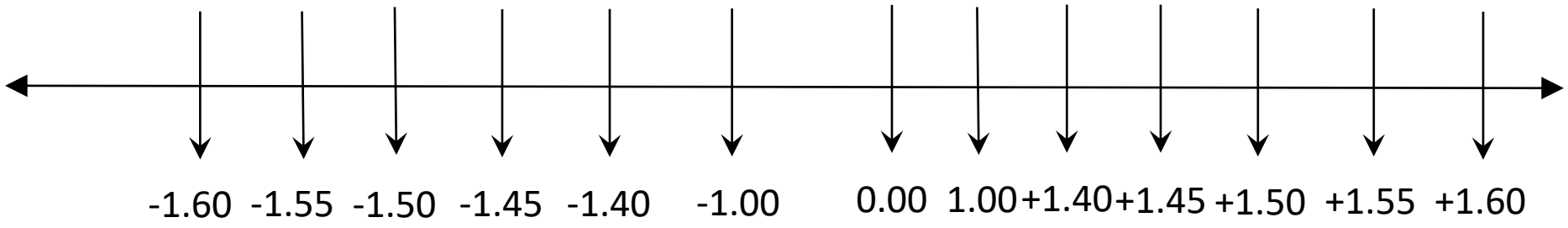
1.5
1.51
1.52
1.53
1.54
1.55 (tie)
1.56
1.57
1.58
1.59
1.6

Ties to even	2 digits
-1.54	-1.5
-1.55	-1.6
-1.56	-1.6

Ties to even	2 digits
1.54	1.5
1.55	1.6
1.56	1.6

IEEE-754 Rounding rules

- Rounding to nearest
  - Round to nearest, ties to even



-1.5
-1.49
-1.48
-1.47
-1.46
-1.45 (tie)
-1.44
-1.43
-1.42
-1.41
-1.4

1.4
1.41
1.42
1.43
1.44
1.45 (tie)
1.46
1.47
1.48
1.49
1.5

Ties to even	2 digits
-1.44	-1.4
-1.45	-1.4
-1.46	-1.5

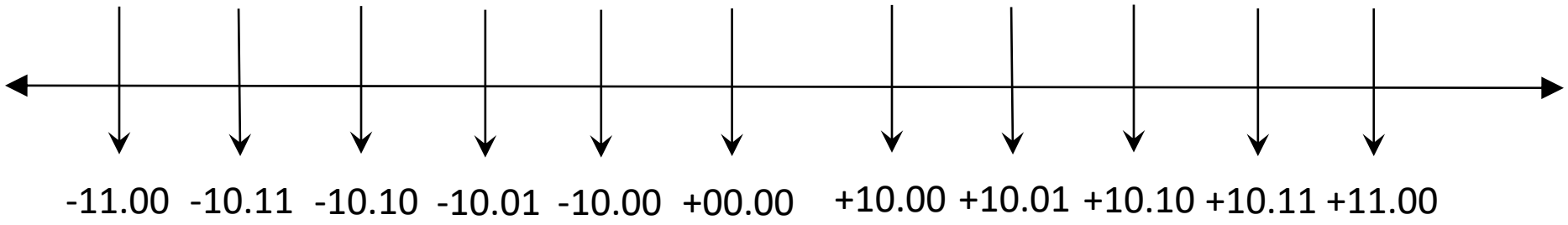
Ties to even	2 digits
1.44	1.4
1.45	1.4
1.46	1.5

Round to nearest, ties to even (binary)

- “Even” when least significant bit is 0
- “Half way” when bits to the right of rounding position = $100....0_2$

IEEE-754 Rounding rules

- Rounding to nearest
  - Round to nearest, ties to even



-11.00
-10.11
-10.10 (tie)
-10.01
-10.00

10.00
10.01
10.10 (tie)
10.11
11.00

Ties away to even	2 bits
-10.01	-10.
-10.10	-10.
-10.11	-11.

Ties away to even	2 bits
10.01	10.
10.10	10.
10.11	11.

Try

0.7783	4 digits	3 digits	2 digits
Truncate			
Round down			
Round up			
Round to Nearest (ties to even)			

Try

-0.7783	4 digits	3 digits	2 digits
Truncate			
Round down			
Round up			
Round to Nearest (ties to even)			

Try

Number (7 bits)	Round down	Round up	Truncate	Round to nearest ties to even
+0.100101*110*				
+0.100101*010*				
+0.100101*100*				

To recall …

- What have we learned:
  - ✓Describe the process of determining the number of significant figures in a number
  - ✓Describe the process of performing different types of IEEE-754 rounding methods