

***ONLINE LESSION***

***-CS-UNIT-2-***

***Fractional Numbers Representation***

## 1. Fractional Converting.

### a. Decimal fractions to Binary Conversions.

**21.25, relevant** binary number?

### b. Binary fraction.

2	1	0	.	-1	-2	-3	-4
$2^2$	$2^1$	$2^0$		$2^{-1}$	$2^{-2}$	$2^{-3}$	$2^{-4}$
4	2	1		$1/2^1$	$1/2^2$	$1/2^3$	$1/2^4$
				$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$
				0.5	0.25	0.125	0.0625
				↓	↓	↓	↓
				1	2	3	4

$$\begin{aligned}
 0.1_2 &= 0.5_{10} \\
 0.01_2 &= 0.25_{10} \\
 0.11_2 &= 0.75_{10} \\
 0.001_2 &= 0.125_{10} \\
 0.101_2 &= 0.625_{10} \\
 0.111_2 &= 0.875_{10}
 \end{aligned}$$

Ex:  $101.11_2$

$$\begin{array}{rcl}
 \begin{array}{ccc} 1 & 0 & 1 \\ \hline 4 & 2 & 1 \end{array} & . & \begin{array}{cc} 1 & 1 \\ \hline 0.5 & 0.25 \end{array} \\
 \hline
 4 + 0 + 1 & & 0.5 + 0.25 \\
 \hline
 5 & & 0.75 \\
 \hline
 5.75_{10}
 \end{array}$$

IMPORTANT.

Method 7, 8

2018-07

7) Convert the decimal number 0.171875 to binary

(a) 0.010111

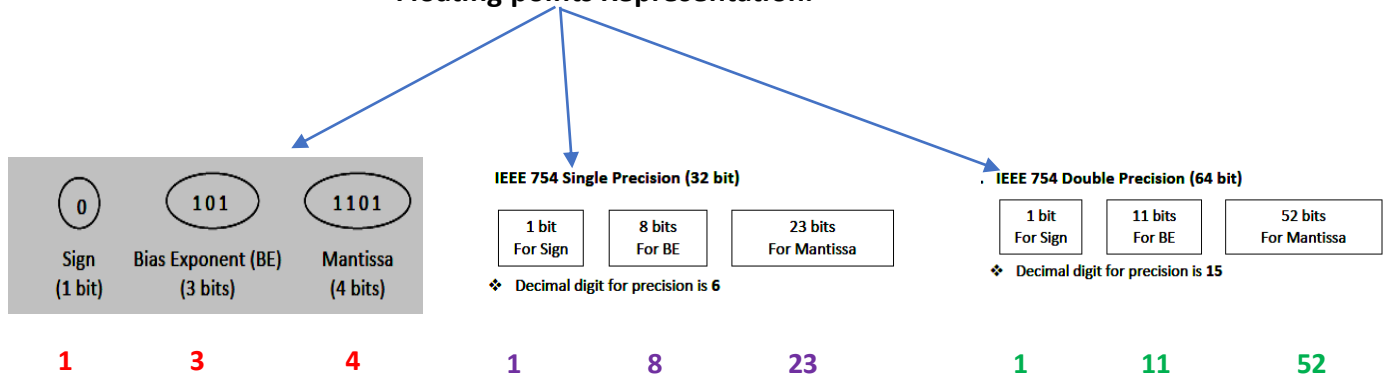
(b) 0.001101

(c) 0.001011

(d) 0.001110

(e) 0.000111

### Floating points Representation.



## Samples → 3.125 using Simple model (8 bit).

Step 0 → Write the given format.

0

1 (Sign bit)

3 (BE)

1 0 0 1

4(Mantissa)

Step 0 → Apply sign bit value {Positive number → 0, Negative number → 1}

Step 1 → Get Binary.

3.125  
↓  
11.001

Step 2 → convert binary pattern in to scientific notation and get mantissa.

1.1001 × 2<sup>1</sup>  
↑  
Significant / Mantissa

Step 3 → Find the BE using followed

$$BE = E + k$$

↑  
Real Exponent

Constant based on number of digits of bias exponent

$$BE = 1 + 3$$

$$= 4$$

equations,

$$k = 2^{n-1} - 1$$

$$\begin{aligned} k &= 2^{3-1} - 1 \\ &= 2^{3-1} - 1 \\ &= 3 \end{aligned}$$

Step 4 → convert BE decimal number into binary and complete the pattern.

4 → 011

Step 5 → Complete pattern.

0

1 (Sign bit)

0 1 1

3 (BE)

1 0 0 1

4(Mantissa)

**FORWORD QUESTIONS.****2014-11**

- 11) The IEEE standard 32-bit floating point representation of the binary number **32.5** is

- (a) 0 01111111 110000000000000000000000
- (b) 1 10000011 000011000000000000000000
- (c) 0 10000100 000001000000000000000000
- (d) 0 10000011 000001000000000000000000
- (e) 0 11000001 111000000000000000000000

**2015-11**

11) The IEEE standard 32-bit floating point representation of the binary number +42.625 is

- (a) 0 01111111 110000000000000000000000
- (b) 1 10000011 010101010000000000000000
- (c) 0 10000100010101010000000000000000
- (d) 0 10000011 101010100000000000000000
- (e) 0 10000010010111100000000000000000

## BACKWORD QUESTIONS.

2012-8

- 8) The equivalent in decimal number to the IEEE standard 32-bit floating point representation of **1 01111111 11000000000000000000** is

- |          |           |           |
|----------|-----------|-----------|
| (a) +1.1 | (b) -1.0  | (c) +1.11 |
| (d) -0.1 | (e) -1.11 |           |

**Step 1**  
 Sign bit  $\rightarrow 1 \therefore$  value negative  
 $\therefore$  Ans **(a) X**

**Step 2**  
 calculate the B/E - decimal value

0 1 1 1 1 1 1 1  
 $\uparrow \uparrow \uparrow \uparrow \uparrow \uparrow \uparrow \uparrow$   
 $64 + 32 + 16 + 8 + 4 + 2 + 1 = 127$   
**BE = 127**

**(iii) Step 3 - calculate the value of K**

$$K = 2^{(E-1)} - 1$$

$$= 2^{(127-1)} - 1$$

$$= 2^{126} - 1$$

$$= 128 - 1$$

$$= 127$$

**BE = E + K**  
 $E = BE - K$   
 $= 127 - 127$   
 $= 0 //$

**(iv) Step 4 - write mantissa,**  
 \*\*\* **1.** mantissa  $\times 2^E$

$$1.11 \times 2^E$$

$$1.11 \times 2^0$$

**Answer (e)**

$\frac{1}{2^0} = 1$

$\frac{0.1}{0.01} = 0.5$   
 $\frac{0.1}{0.01} = 0.25$   
 $\frac{0.1}{0.01} = 0.75$   
**total  $\rightarrow -1.75 //$**

**2017-10**

- 10) Which of the following is the correct decimal number of the 16-bit floating point representation 010101 0101011010 with a sign bit, 5-bit exponent and 10-bit mantissa?

(a) +87.625	(b) +85.625	(c) +89.625
(d) +43.625	(e) +47.625	



## Floating points -Round-Off-Error

2014-09

- 9) What is the loss of accuracy (round-off-error) when converting the decimal value +255.9375 to 16-bit floating point representation with a sign bit, 5-bit exponent and a 10-bit mantissa?

(a) 0.0625

(b) 0.125

(c) 0.1875

(d) 0.25

(e) 0.5

**Handwritten Solution:**

**Formula:**  $\text{R.O.F. Error} = A - B$

**Conversion of 255.9375 to 16-bit floating point:**

255.9375 is converted to binary. The integer part 255 is  $11111111_2$  and the fractional part 0.9375 is  $0.1111_2$ . The combined binary is  $11111111.1111_2$ .

**Ex: Conversion of 0.9375 to binary:**

$0.9375 \times 2$	$1.875$	1
$0.875 \times 2$	$1.75$	1
$0.75 \times 2$	$1.5$	1
$0.5 \times 2$	$1.0$	1
	$0.0$	

Result:  $0.1111_2$

**Exponent Calculation:**

$BE = E + K$   
 $= 7 + 15$   
 $= 22$   
 Binary:  $10110$

**Mantissa:** The mantissa is the fractional part of the binary number, which is  $1111$ . It is stored in the mantissa field.

**Final 16-bit floating point representation:**

Sign: 0  
 Exponent: 10110  
 Mantissa: 1111

**Loss of Accuracy (Round-off-error):**

$\text{Error} = A - B$   
 $255.9375 - 255.875 = 0.0625$

**Verification:**

255.875 is converted to binary:  $11111111.111_2$ . The mantissa is  $111$ .

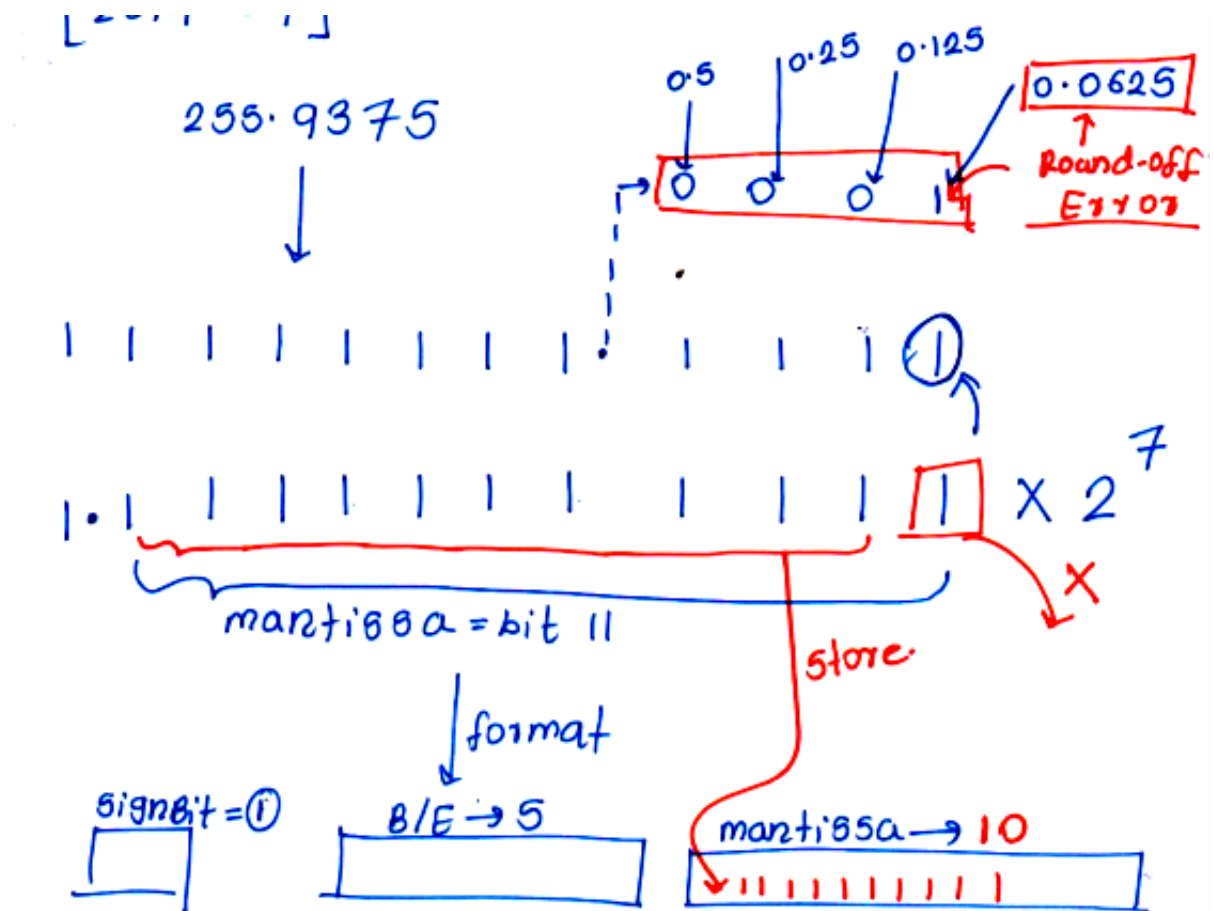
**Final Answer:** (a) 0.0625

## Round-Off-Error -Short Method

2014-09

- 9) What is the loss of accuracy (round-off-error) when converting the decimal value +255.9375 to 16-bit floating point representation with a sign bit, 5-bit exponent and a 10-bit mantissa?

(a) 0.0625	(b) 0.125	(c) 0.1875
(d) 0.25	(e) 0.5	



**2015-09**

- 9) What is the loss of accuracy (round-off-error) when converting the decimal value +511.875 to 16-bit floating point representation with a sign bit, 5-bit exponent and a 10-bit mantissa?

- |            |           |            |
|------------|-----------|------------|
| (a) 0.0625 | (b) 0.125 | (c) 0.1875 |
| (d) 0.25   | (e) 0.5   |            |

**2016-10**

- 10) What is the loss of accuracy (round-off-error) when converting the decimal value +1000.875 to 16-bit floating point representation with a sign bit, 5-bit exponent and a 10-bit mantissa?

- |           |          |           |
|-----------|----------|-----------|
| (a) 0.125 | (b) 0.25 | (c) 0.375 |
| (d) 0.625 | (e) 0.75 |           |