COMPUTER SCIENCE

Computer Organization and Architecture

Floating Point Representation

Lecture_02

Vijay Agarwal sir

TOPICS TO BE COVERED

01 Floating Point Representation

Signed & Unsigned

1's & 2's Compleement

Why 2's Complement are Used.

Floating Point Representation

↳ Why its USED ?

Floating point Representation

1 bit

S	E	M

S: Sign → 0 (+ve), 1 (−ve)

$E = e + bias$

or

$BE = AE + bias$

M: Mantissa.

$$+/- \;\_ . \underbrace{\_\_\_\_\_\_\_}_{M} \times 2^{e}$$

M: Mantissa

e: exponent / Actual Exponent

Why Bias Exponent is Need ?

2's Complement = $\boxed{-2^{n-1}}$ to $+2^{n-1}-1$

4bit $= -2^{4-1}$ to $+2^{4-1}-1$

$\Rightarrow -8$ to $+7$

E = 4bit

$e = -2$

$\Downarrow$ 2's

0010

1's 1101

$\Downarrow$

1110

$e = 14$

OR

$$\boxed{\text{Bias} = 2^{n-1}}$$

if Exponent is k bit

then $\boxed{\text{bias} = 2^{k-1}}$

Floating-Point Representation

16 bit fixed point data format then

Range = -2^{16-1} to $+(2^{16-1} - 1)$

$\Rightarrow$ $-(2^{15})$ to $+(2^{15} - 1)$

If we want to store 61,000 then we cannot store

Because range [–32k to + 32 k –1]

So floating point representation is to represent very large data and very small fraction and consume less memory

Floating point used to represent $\begin{cases} +8.56410000000000.... [\Rightarrow \infty] \\ +0.0000000007892 \quad \Rightarrow [\Rightarrow 0] \end{cases}$

Floating-Point Representation

S	E	M

S: sign bit — 0 +ve, 1 –ve

E: exponent

M: Mantissa

$$\overset{S}{+/-}\; x.\underbrace{\ldots\ldots}_{M} \times 2^{e}$$

$. \ldots\ldots \times 2^{e}$

6.5 in Binary $\Rightarrow$ 110.1

Q. 1 $+6.5$

$$6.5 = (110.1)_2$$

$$\underset{S}{\underline{0.}}\,\underset{M}{\underline{1101}} \times \underset{2^e}{\underline{2^3}}$$

$$S = 0\ (+)$$

$$M = 1101$$

$$e = 3 = (11)_2$$

S	e	M
0	11	1101

Very. Imp

$$6.5 = 110.1$$

$$= .1101 \times 2^3$$

$$= [.2^{-1} + 2^{-2} + 2^{-4}] \times 2^3$$

$$= [2^2 + 2^1 + 2^{-1}]$$

$$= 6.5$$

Q. 2

$+4.5$

100.1

0.1001×2^3

S =0 (+ve)

$M = 1001$

$e = 3$ [11]

S	e	M
0	11	1101

Q. 3 $+4.75$

100.11

$.10011 \times 2^3$

$S = 0$

M: 10011

$e = 3 \Rightarrow (11)_2$

Most Negitive Number $= -2^{k-1}$

So bias $= +2^{k-1}$

S	e	M
0	11	10011

NOTE:

Mantissa alignment process is used to adjust the decimal point; in this process right alignment increments the exponent and left alignment decrements the exponent.

$2^{+\text{shift}}$ power(+) = Right alignment $\Rightarrow$ Increment the exponent

$2^{-\text{shift}}$ power (-) = Left alignment $\Rightarrow$ Decrease the exponent

Right Alignment

6.5

110.1 $\quad 110.1$

$\Rightarrow .1101 \times 2^{3}$ $\quad \boxed{.1101 \times 2^{+3}}$

$\Rightarrow [.2^{-1} + 2^{-2} + 2^{-4}] \times 2^{3}$

$\Rightarrow 2^{2} + 2^{1} + 2^{-1}$

$\Rightarrow 4 + 2 + 0.5$

$\Rightarrow$ 6.5 Ans

Left Alignment

Data: $0.0000000101 \times 2^{+5}$ $\quad 2^{-8}$

$[1.01 \times 2^{+5-8}]$

$+1.01 \times 2^{-3}$

(Align to use upto 8 times)

Q. 4 $+0.00101$

0.101×2^{-2}

$M = 101$

$E = -2$

$S = 0$

S	E(4bit)	M(5 bit)
0	1110	10100
	E	M

$\boxed{E = -2} = (1110)_2$ 2's complement

& all Numbers.

Biasing: is method in which we convert the negative number into the positive number

Bit	Bit	Bit
S	E	M

S = Sign

E/BE = Exponent or

BE = bias exponent

M = Mantissa

E = e + bias

$\boxed{\text{Bias} = 2^{K-1}}$ where K is exponent bits

Example

If K = 4 bits

Exponent = 4 bit then

bias = $2^{K-1} = 2^{4-1} = 8$

1 Bit	4 Bit	
S	E	M

x bit

Bias = $2^{K-1} = 2^{4-1}$

bias = 8

E = e + bias

E = e + 8

E = 4 bit

or

Excess 8 code

$2^{K-1} = 8$

$2^{K-1} = 2^3$

K − 1 = 3

K = 4

E = 4 bit

e [original exponent]	Stored exponent [BE] E (E)
-8 e+bias (+8)	0
-7	1
-6	2
-5	3
-4	4
-3	5
-2	6
-1	7
0	8
1	9
2	10
3	11
4	12
5	13
6	14
7 +7+8	15

$$\boxed{bias = 2^{k-1}}$$

$$.\,\dots\dots \times 2^{e}$$

e: Actual exponent

$$\boxed{E = e + bias}$$

$$\boxed{BE = AE + bias}$$

Q.

From previous question

0.00101

0.101×2^{-2}

$M = 101$

$\text{Bias} = 2^{5-1}$

$\text{Bias} = 16$

$e = -2$

$E = e + \text{bias}$

$E = -2 + 16$

$E = 14$

$E = (01110)_2$

$\boxed{E = e + \text{bias}}$

$BE = AE + \text{bias}.$

$\boxed{\text{bias} = 2^{k-1}}$

if exponent is 'k' bit

-2^{k-1} to $+2^{k-1}-1$

-8 to $+7$ (if E = 4 bit)

OR

$\boxed{\text{bias} = \frac{2^k}{2}}$

1 bit	5 bit	4 bit
S	E	M

1 bit	5 bit	4 bit
0	01110	1010

Ans

Formula: $(-1)^S \times 0.M \times 2^e$

$(-1)^0 \times 0.101 \times 2^{E\text{-bias}}$

$0.101 \times 2^{14-16} = 0.101 \times 2^{-2}$

0.000101 Ans

Mantissa.

(1) 0.0001010

(2) 101011.0110101

(3) 100.11101

(4) 0.11010011101

⋮

Solution is Normalized the Mantissa

BE = AE + bias.

$\boxed{E = e + bias}$

① Implict → 1. something.

② Exblicit → 0.1 . . . . .

Exblicit

$\boxed{0.\underbrace{1 . . . . . .}_{M} \times 2^{e}}$

OR

$\boxed{(-1)^{S}\, 0.M \times 2^{e}}$ ⇒ $\boxed{(-1)^{S}\, 0.M \times 2^{E-bias}}$

Implict

$\boxed{1. \_ \_ \_ \_ \_ \_ \times 2^{e}}$

$(-1)^{S}\, 1.M \times 2^{e}$

or

$(-1)^{S}\, 1.M \times 2^{E-bias}$

S	E	M
1 bit	x bit	y bit

Normalized Mantissa

1 Explicit Normalized

Syntax

$$\underbrace{0.1\ldots\ldots}_{M} \times 2^{e}$$

Formula to get number

[value formula]

$(-1)^S \times 0.M \times 2^e$

$$\boxed{(-1)^S \times 0.M \times 2^{E-bias}}$$

2 Implicit Normalized

Syntax

$$1.\underbrace{\ldots\ldots}_{M} \times 2^{e}$$

Formula to get number

[value formula]

$(-1)^S \times 1.M \times 2^e$

$$\boxed{(-1)^S \times 1.M \times 2^{E-bias}}$$

Explicit

$0 \cdot 1$

0.1 After the point,

Immediate first bit should be 1

Example Explicit $(\underline{0 \cdot 1} \ldots)$

(101.11)

$0.\underbrace{10111}_{M} \times 2^{+3}$ $\boxed{e = +3}$

$M = 10111,$

$\boxed{e = 3}$

$\boxed{E = e + \text{bias}}$

Implicit

Before the point 1, means 1. ……

Example Implicit (1 · Something)

(101.11)

1.0111×2^2 $\boxed{e = +2}$

$\boxed{M = 0111,}$

$\boxed{e = 2}$

$\boxed{E = e + \text{bias}}$

Floating-Point Representation

S	E	M

S: sign bit
- 0 +ve
- 1 –ve

E:Biased exponent

M: Mantissa

E = e + bias

or

BE = AE + bias

$$\overset{S}{+/-}\ x\bullet\underbrace{\ldots\ldots\ldots}_{M}\ X\ 2^{e}$$

. ... $X2^{e}$

Explicit

Implicit

O • L x x x x

L • x x x

Q. 1 +(6.75) format

Then do explicit and implicit normalization

if exponent is k bit

$\boxed{bias = 2^{k-1}}$

1 bit	4 bit	5 bit
S	E	M

$+6.75$

$\underset{4}{1}\underset{2}{1}\underset{1}{0}.\underset{.5}{1}\underset{.25}{1}$

2^5	2^4	2^3	2^2	2^1	2^0	.	2^{-1}	2^{-2}	2^{-3}	2^{-4}
32	16	8	4	2	1	.	$1/2$	$1/4$	$1/8$	$1/16$
						.	0.5	0.25	0.125	0.0625

$bias = 2^{4-1}$

$= 2^3$

$\boxed{bias = 8}$

$E = e + bias$ (or) $BE = AE + bias$

$\boxed{E = e + 8.}$

Q (+6.75)

S	E	M
1bit	4bit	5bit

$\boxed{\text{bias} = 8}$

Explicit

(110.11)

$$\boxed{+0.\underbrace{11011}_{M} \times 2^{+3}}$$

Sign[S] = 0 (+ve)

M = 11011

E = 11 [1011]

bias = 8

$e = +3$

$\boxed{E = e + \text{bias} \text{ OR } BE = AE + \text{bias}}$

$E = 3 + 8 \Rightarrow E = 11$ in 4bit

S(1bit)	E(4bit)	M(5bit)
0	1011	11011

Ans

1 7 B

$(17B)_{16}$ Ans

(+6.75)

S	E	M
1bit	4bit	5bit

$\boxed{\text{bias} = 8}$

Implicit

$(1.10.11)$

$$+\boxed{1.1011 \times 2^{+2}}$$

S = 0

M = 1011

$e = +2$

bias = 8

$\boxed{E = e + \text{bias} \text{ or } BE = AE + \text{bias}}$

$E = 2 + 8 \Rightarrow E = 10 \Rightarrow \boxed{1010}$

S(1bit)	E(4bit)	M(5bit)
0	1010	10110

Ans

1 5 6

$(156)_{16}$

+6.75

S	E	M
1bit	4bit	5bit

$\boxed{\text{bias} = 8}$

Explicit

Q

S(1bit)	E(4bit)	M(5bit)
0	1011	11011

Sol

Value formula

$E = e + \text{bias}$

$$\boxed{(-1)^S \ 0.M \times 2^{e}}$$

$$\boxed{e = E - \text{bias}}$$

bias = 8

$(-1)^S \ 0.M \times 2^{E-\text{bias}}$

$E = 1011 \Rightarrow \boxed{E = 11}$

$S = 0, \ M = 11011$

$(-1)^0 \ 0.11011 \times 2^{11-8}$

$+0.11011 \times 2^{+3}$

$\Rightarrow 110.11$

$\Rightarrow +(6.75)$ Ans

+6.75

S	E	M
1bit	4bit	5bit

$\boxed{\text{bias} = 8}$

Implicit

Q

S(1bit)	E(4bit)	M(5bit)
0	1010	10110

Sol

Value formula

bias = 8

$$\boxed{(-1)^S \ 1.M \times 2^{e}}$$

$E = 1010 \Rightarrow 10$

$\boxed{E = 10}$

$(-1)^S \ 1.M \times 2^{E-\text{bias}}$

S = 0

$\boxed{M = 10110}$

$(-1)^0 \ 1.10110 \times 2^{10-8}$

$+1.10110 \times 2^{+2}$

110.110

+(6.75) Ans.

1bit 4bit 5bit

S	E	M

bias = 8

$+5.5$ Explicit

$+(101.1)$

$0.1011 \times 2^{+3}$ $e = +3$

$E = 3 + 8 = 11$

$m = 10110$ $E = 1011$

S	E(4bit)	M(5bit)
0	1011	10110

$+(5.5)$ Implicit

(101.1) $E = 2 + 8 = 10$

$1.011 \times 2^{+2}$ $E = 1010$

$\boxed{M : 01100}$

S(1bit)	E(4bit)	M(5bit)
0	1010	01100

$(-1)^S \, 0.M \times 2^{E-\text{bias}}$ $\boxed{E = e + \text{bias}}$

$E = 11$

$(-1)^0 \, 0.10110 \times 2^{11-8}$

$0.10110 \times 2^{+3}$

101.10

$\Rightarrow (5.5)$ Ans

$(-1)^S \, 1.M \times 2^{E-\text{bias}}$

$(-1)^0 \, 1.01100 \times 2^{10-8}$

$+1.01100 \times 2^{+2}$

$101.100 \Rightarrow +(5.5)$ Ans

S	E	M

$+/-\ 1.xxxx \times 2^{e}$

$$S \begin{cases} 0\ (+ve) \\ 1\ (-ve) \end{cases}$$

Implicit

$$\boxed{E = e + bias}$$

or

$$\boxed{BE = AE + bias}$$

$$(-1)^{S}\ 1.M \times 2^{e}$$

$$\boxed{(-1)^{S}\ 1.M \times 2^{E-bias}}$$

4.875

100.111

S	E	M
1bit	4bit	5bit

$$bias = 2^{4-1} = 8$$

$$\boxed{bias = 8}$$

$$\boxed{E = e + 8}$$

Q. 2 +(4.875) format

Then do explicit and implicit normalization

Explicit

(+4.875)

100.111

0.100111×2^{3}

M = 100111

e = 3, bias = 2^{4-1}

E = 3 + 8

E = 11

E = 1011

1 bit	4 bit	5 bit
0	1011	10011

Value Formula: $(-1)^{S} \times 0.M \times 2^{e}$

$(-1)^{0} \times 0.10011 \times 2^{11-8}$

0.10011×2^{3}

100.11

4.75

(Not getting very accurate)

Implicit

$(+4.875)$

100.111

1.00111×2^2

$M = 00111$

$e = 2$, bias $= 2^{4-1}$

$E = 2 + 8$

$E = 10$

$E = 1010$

1 bit	4 bit	5 bit
0	1010	00111

Value Formula: $(-1)^S \times 1.M \times 2^e$

$(-1)^0 \times 1.00111 \times 2^{10-8}$

1.00111×2^2

100.111

4.875

(Getting very accurate)

either Increase the bits in Mantissa

OR

Using Implicit Normalization.

Note

Mantissa : Giving Precision. (More & More bit in Mantissa. getting very accurate for very small Number.

Exponent : gives the Range.
(More bit in Exponent means large larger Number).

Q. Consider a 16 bit register used to store floating point number. Mantissa is normalized signed fraction number. Exponent is in Excess-32 form then what is 16-bit for $+(13.5)_{10}$ in the register?(Using Explicit & Implicit)

Assume E = k bit

← 16 bit →

S	E	M
1 bit	k bit	(15 − k)
	6 bit	9 bit

2's Complement Range = $-(2^{k-1})$ to $+2^{k-1}-1$

most −ve No.

So bias = $+2^{k-1}$

$\Rightarrow 2^{k-1} = 32$ Excess

$2^{k-1} = 2^5$

$k-1 = 5$

$\boxed{k=6}$

1 bit	6 bit	9 bit
S	E	M

$\text{bias} = 2^{6-1}$

$\boxed{\text{bias} = 32}$

OR

if Excess-32 ⇒ $\boxed{\text{bias} = 32}$

Explicit $+(13.5)$

1101.1

$0.11011 \times 2^{+4}$

bias $= 32$

$e = +4$

$\boxed{M : 11011}$

$E = e + \text{bias}$

$E = 4 + 32 = 36$

$\boxed{E = 100100} \Rightarrow E = 36$

S	E (6 bit)	M (9 bit)
0	100100	11011 0000

$(4\ 9\ B\ 0)_{16}$ Ans

S	E	M
1 bit	6 bit	9 bit

bias $= 32$

Implicit

1101.1

$1.1011 \times 2^{+3}$

M : 1011

$e = +3$

bias $= 32$

$E = 3 + 32 = E = 35$

$\boxed{M : 1011}$

$\boxed{E = 100011}$

	E (6 bit)	M (9 bit)
0	100011	1011 00000

$(4\ 7\ 6\ 0)_{16}$ Ans

Explicit

$E = 36$ M(9bit)   bias=32

0	100100	110110000

$(-1)^S \; 0.M \times 2^{E-\text{bias}}$

$(-1)^0 \; 0.110110000 \times 2^{36-32}$

$0.11011\underline{0000} \times 2^{+4}$

1101.10000

(13.5) Ans

Implicit

$E = 35$ M(9bit)   bias=32

0	100011	101100000

$(-1)^S \; 1.M \times 2^{E-\text{bias}}$

$(-1)^0 \; 1.101100000 \times 2^{35-32}$

$1.101100000 \times 2^{+3}$

1101.100000

(13.5) Ans

Q. $+21.75$

Implicit ?

1 bit	7 bit	8 bit
S	E	M

10101.11

1.010111×2^4

$M = 010111$

$e = 4$, bias $= 2^{7-1}$

$E = 4 + 64$

$E = 68 = (1000100)_2$

Value Formula:

$(-1)^S \times 1.M \times 2^e$

$(-1)^0 \times 1.010111 \times 2^{68-64}$

1.010111×2^4

$10101.11 = (21.75)_{10}$

Ans

S(1bit)	E(7bit)	M(8 bit)
0	1000100	01011100

Hexadecimal $= (445C)_{16}$

Q. Consider a 16 bit register used to store floating point number. Mantissa is Implicit normalized signed fraction number. Exponent is in Excess-64 form then

(i) what is the First Smallest Positive number?

(ii) what is the Second Smallest Positive number?

(iii) what is the Difference between First Smallest & Second Smallest Positive number?

$$2^{k-1} = 2^6 \,[64]$$

$$k-1 = 64$$

$$\boxed{(\text{Exponent})\ k = 7\text{bit}}$$

1bit	7bit	8bit
S	E	M

$$\text{bias} = 2^{7-1} = 2^6$$

$$\boxed{\text{bias} = 64}\ (\text{Excess } 64).$$

We Can not Represent '0' in either.

(i) <u>Implicit</u> $1.xxx) \Rightarrow 1.0$ (Not zero)

OR

(ii) <u>Explicit</u> $0.1xx\ldots \Rightarrow 0.1$ (Not Zero)

1bit	7bit	8bit
S	E	M

Sign = 0 +ve

(i) First Smallest +ve Number.

S(1bit)	E(7bit)	M(8bit)
0	0000000	00000000

Note In the explicit Representation we Cannot Represent '0'

Because 0.1 . . . . .

Note In the Implicit we can not write/Represent Only 0 Because 1.0 Required

"So we Use Single Precision & Double Presion IEEE 754 Format".

But Here for first Smallest we put all 0's in E & Mantissa.

Here exactly value is Not zero (1. . . . . .)

① I^{st} Smallest

S(1bit)	E(7bit)	M(8 bit)
0	0000000	00000000

E : 0000000 ⇒ E = 0

$bias = 2^{7-1} \Rightarrow \boxed{bias = 64}$

$(-1)^S \; 1.M \times 2^e$

$\Rightarrow (-1)^0 \; 1.00000000 \times 2^{0-64}$

$\Rightarrow (-1)^S \; 1.M \times 2^{E-bias}$

I^{st} Smallest +ve No. $\boxed{1.00000000 \times 2^{-64}}$ Ans

---

② II^{nd} Smallest +ve Number :

S(1bit)	E(7bit)	M(8 bit)
0	0000000	00000001

$(-1)^S \; 1.M \times 2^e$

$\Rightarrow (-1)^0 \; 1.00000001 \times 2^{0-64}$

$(-1)^S \; 1.M \times 2^{E-bias}$

E : 0000000 = 0

bias = 64

2^{nd} Smallest +ve Number. $\boxed{1.00000001 \times 2^{-64}}$ Ans

(3) Difference b/w 1^{st} smallest & 2^{nd} smallest.

$$1.00000001 \times 2^{-64} - 1.00000000 \times 2^{-64}$$

$$[1.00000001 - 1.00000000] \times 2^{-64}$$

$$0.\underbrace{00000001}_{2^{-8}} \times 2^{-64}$$

$$2^{-8} \times 2^{-64}$$

Positive Diff.

$$= 2^{-72} \text{ Ans}$$

16bit

Possibility

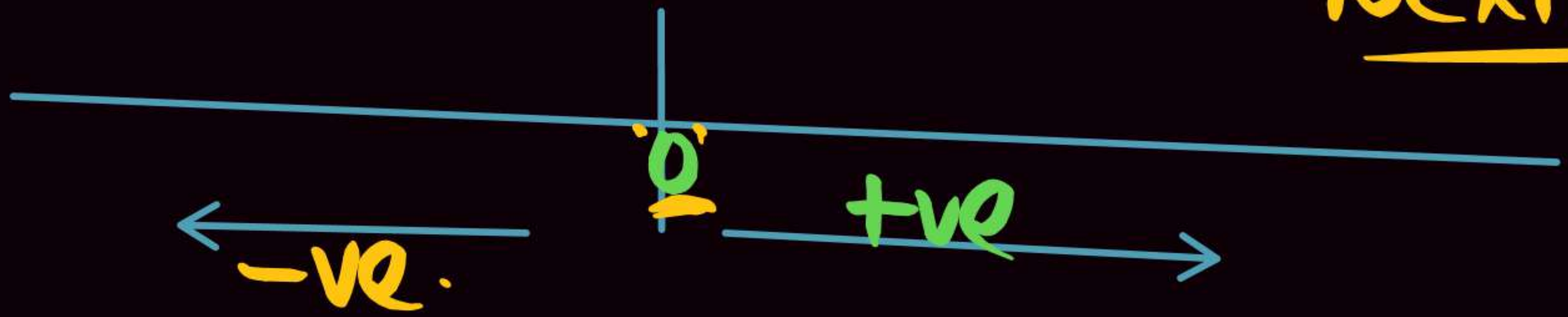
2^{16}

1bit	7bit	8bit
S	E	M

If we want the Answer for Negitive Number

then PLz P.T.O

Next Page



(3) Difference b/w I^{st} smallest & 2^{nd} smallest.
Negative Number

$$-\left[1.00000001 \times 2^{-64} - 1.00000000 \times 2^{-64}\right]$$

$$-\left(\left[1.00000001 - 1.00000000\right] \times 2^{-64}\right)$$

$$-\left[0.\underline{00000001} \times 2^{-64}\right] \quad 2^{-8}$$

$$-\left[2^{-8} \times 2^{-64}\right]$$

$$-2^{-72} \text{ Ans}$$

Difference b/w I^{st} & 2^{nd} smallest -ve Number.

16bit
Possibility 2^{16}

1bit	7bit	8bit
S	E	M

Q. Consider a 16 bit register used to store floating point number. Mantissa is Implicit normalized signed fraction number. Exponent is in Excess-64 form then

(i)what is the First Highest Positive number?

(ii)what is the Second Highest Positive number?

(iii)what is the Difference between First Highest & Second Highest Positive number?

1st Highest

(1bit)	E(7bit)	m(8bit)
0	1111111	11111111

2nd Highest

S(1bit)	E(7bit)	M(8bit)
0	1111111	11111110

Home work

V.Imp Concept How to Deals with bits

eg (i) $0.111 \Rightarrow 1 - \frac{1}{2^3}$ Ans

eg (ii) $0.1111111 \Rightarrow 1 - \frac{1}{2^7}$ Ans

eg (iii) $0.11111111 \Rightarrow 1 - \frac{1}{2^8}$

or $(1 - 2^{-8})$

OR

Proof

$0.111 \Rightarrow \underline{111} \times 2^{-3}$

$111 \rightarrow 7$

$[2^3 - 1] \times 2^{-3}$

$\Rightarrow 2^{+3-3} - 2^{-3}$

$2^0 \rightarrow 1$

$\Rightarrow 1 - 2^{-3}$

$\Rightarrow 1 - \frac{1}{2^3}$ Ans

eg (0.1111)

$0.\ \frac{1}{2}\ \frac{1}{2^2}\ \frac{1}{2^3}\ \frac{1}{2^4}$

$\Downarrow$

$\boxed{1-\frac{1}{2^4}}$ Ans

G.P Series

$a = \frac{1}{2}$

$r = \frac{1}{2}$

$$\frac{a(1-r^n)}{1-r}$$

$$\frac{\frac{1}{2}\left(1-\frac{1}{2^4}\right)}{\left(1-\frac{1}{2}\right)} \Rightarrow \frac{\cancel{\frac{1}{2}} \times \left(1-\frac{1}{2^4}\right)}{\cancel{\frac{1}{2}}}$$

$\Downarrow$

$1-\frac{1}{2^4}$ Ans

$15 \Rightarrow (2^4 - 1$

Alternate Approach

eg $\underline{(0.1111)}$ $\times 2^x$

$\underline{1111}.0 \times 2^{-4}$

$(2^4 - 1) \times 2^{-4}$

$\cancel{2^4} \times \cancel{2^{-4}} - 2^{-4}$

$(1 - 2^{-4})$

$\left(1 - \frac{1}{2^4}\right) \underline{\cancel{A}} \, 2^x$

$\frac{1}{2} \quad \frac{1}{2^2} \quad \frac{1}{2^3} \quad \frac{1}{2^4}$

G.P Series $\quad a = \frac{1}{2}$

$r = \frac{1}{2}$

$$\frac{a(1-r^n)}{1-r}$$

$$\frac{\frac{1}{2}\left(1 - \frac{1}{2^4}\right)}{\left(1 - \frac{1}{2}\right)} \Rightarrow \frac{\cancel{\frac{1}{2}} \times \left(1 - \frac{1}{2^4}\right)}{\cancel{\frac{1}{2}}}$$

$\Downarrow$

$1 - \frac{1}{2^4}$

Q. Consider a 16 bit register used to store floating point number. Mantissa is Explicit normalized signed fraction number. Exponent is in Excess-32 form then what is 16-bit for $-(29.75)_{10}$ in the register?

Solution

1 bit	6 bit	9 bit
S	E	M

-29.75

-11101.11

0.1110111×2^5

M: 1110111

$e = 5$

$\text{bias} = 2^{6-1}$

$\text{bias} = 32$

$E = 5 + 32 = 37 = (100101)_2$

S(1 bit)	E(6 bit)	M(9 bit)
1	100101	111011100

Q. $+21.75$

1 bit	7 bit	8 bit
S	E	M

Implicit ?

10101.11

1.010111×2^4

$M = 010111$

$e = 4$, bias $= 2^{7-1}$

$E = 4 + 64$

$E = 68 = (1000100)_2$

Value Formula:

$(-1)^S \times 1.M \times 2^e$

$(-1)^0 \times 1.010111 \times 2^{68-64}$

1.010111×2^4

$10101.11 = (21.75)_{10}$

Ans

S(1bit)	E(7bit)	M(8 bit)
0	1000100	01011100

Hexadecimal $= (445C)_{16}$

THANK YOU!