

* IEEE 754 Notations

Lecture 13

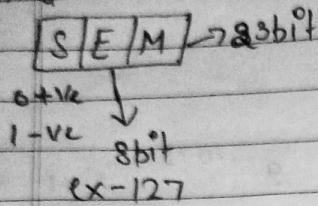
Part 1

Single Precision

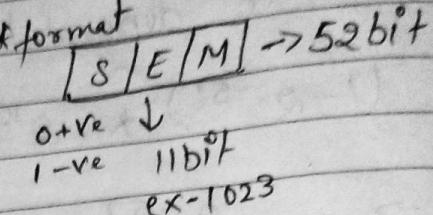
32bit (4byte)

double precision
64bit (8byte)

* format



* format



* IEEE Normalized

1.0 xxx - - -

do not store
(implicit)

Ex : $(3.5)_{10} \rightarrow$ Convert it into IEEE Single precision normalized number

$$\text{Step 1} \Rightarrow \text{Binary} = (11.1)_2$$

$$\text{Step 2} \Rightarrow \text{Normalized} = (11.1)_2 \Rightarrow 1.11 \times 2^1$$

01000000011000 - - -
 sign exponent mantissa
 40600000H

Any

* $(0.5)_{10} \Rightarrow (0.1)_2 \Rightarrow 1.0 \times 2^{-1}$

00111100...
3F000000H Ans

* $(6.25)_{10} \Rightarrow (110.01)_2 \Rightarrow 1.1001 \times 2^3$

01000001001...
40C8000H Ans

* $3FC0000H \rightarrow$ IEEE Single Precision, decimal?

00111111000000...
S E M

$+ 1.0 \times 2^{127-127} \Rightarrow 1.0 \times 2^0 \Rightarrow (1.0)_2$

$(1.5)_{10}$ Ans

* $40C8000H \rightarrow$ IEEE Single Precision, decimal?

010000001001000...
E M

$+ 1.1001 \times 2^{129-127} \Rightarrow 1.1001 \times 2^2 \Rightarrow (110.01)_2$

$(6.25)_{10}$ Ans

* Denormals \Rightarrow exponent all 0's

if exponent is all 0 means
it is normalized

"1" is not implicit

* Ex: $\begin{array}{c} s \quad e \\ \underline{0} \quad \underline{00000000} \quad \underline{0000} \end{array}$

- all exponent 0 means it is denormalized and "1" is not implicit.

$\begin{array}{c} \underline{0} \quad \underline{00000000} \quad \underline{000} \end{array}$

$$+ 0 * 2^{0-127} \Rightarrow 0$$

$\begin{array}{c} \downarrow \\ \underline{1} \quad \underline{00000000} \quad \underline{000} \end{array}$

$$- 0 * 2^{0-127} = -0 \quad \underline{\text{Ans}}$$

$$0 \quad 00000000H \Rightarrow +0$$

$$80000000H \Rightarrow -0$$

* These two are special number

S E M
0111111 000

+ ∞

7F800000H $\Rightarrow +\infty \Rightarrow +ve \text{ infinity}$

S E M
1111111 000

FF800000H $\Rightarrow -\infty \Rightarrow -ve \text{ infinity}$

both are special numbers

• 01111111 ≠ 0 — Not a number

11111111 ≠ 0 — Not a number

* S.P.N	S	E	M
+0	0	all 0's	all 0's
-0	1	all 0's	all 0's
+∞	0	all 1's	all 0's
-∞	1	all 1's	all 0's
NAN	0/1	all 1's	≠ 0
Denormals	0/1	all 0's	≠ 0

Largest * normalized number

- Largest +ve

S E M
0/111111101111 --- 1

$$7 F7 FFFF H \quad \underline{\text{Ans}}$$

$$1.01111... \times 2^{234-127} \Rightarrow 2 - 2^{-2^3} * 2^{127}$$

- Smallest +ve

S E M
0 00000001 0000 ---

0 0800000 H Ans

$$1.0 * 2^{1-127} \Rightarrow 1.0 * 2^{-126} \Rightarrow 2^{-126} A$$

- Smallest -ve

$$-(2 - 2^{-2^3}) * 2^{127}$$

FF7FFFFF H Ans

- Largest -ve

$$-1.0 * 2^{-126} \Rightarrow$$

$$80800000 H \underline{\text{Ans}}$$

* Denormalized number

- Largest +ve

S E
0 0000 0000 1111---1

007FFFFFFH An exception

$$+ 1 - 2^{-23} * 2^{1-127} \Rightarrow 1 - 2^{-23} * 2^{-126}$$

o

Exception: when all the exponent are 0's
 we'll consider it as "1"

- Smallest ~~toe~~ -ve

1 0000 0000 1111 ---

807FFFFFFH

- Smallest +ve

S E M
0 0000 0000 0000 ---1

$$00000001H \underline{\text{An}} \Rightarrow 2^{-23} * 2^{1-127}$$

- Largest -ve

1 0000 000000----1

80000001H

$$-2^{-23} * 2^{-126} \Rightarrow -2^{-149} \underline{\text{An}}$$

$$\Rightarrow 2^{-23} * 2^{-126}$$

$$\Rightarrow 2^{-149} \underline{\text{An}}$$

Background:-

- * Find the smallest diff b/w two consecutive normalized +ve number

SPN

$$0 \ 00000001\ 0000 \dots - \\ 00800000H \Rightarrow 1.0 * 2^{-127} = 2^{-126}$$

\downarrow Next consecutive number

$$0 \ 00000001\ 0000 \dots - 1 \\ 1 + 2^{-23} * 2^{-126}$$

Gap in Consecutive no. $\Rightarrow (1 + 2^{-23}) * 2^{-126} - 1 * 2^{-126}$

$$\cancel{2^{-126}} + 2^{-149} - \cancel{2^{-126}}$$

* Minimum gap : 2^{-149}

* Max gap :

$$0 \underline{11111110} 1111 \dots - 1 \\ 7F7FFFFFH \Rightarrow 2^{-2-23} * 2^{127}$$

$$0 \underline{11111110} 0000 \dots 0 \\ 1.0 * 2^{254-127} = 2^{127}$$

$$0 \underline{11111110} 0000 \dots 1$$

$$(1 + 2^{-23}) * 2^{127} \Rightarrow 2^{104}$$

gap $\Rightarrow \cancel{2^{127}} + 2^{104} - \cancel{2^{127}} \Rightarrow 2^{104}$ Ans

Date _____
Page No. _____

$2^{-149} \rightarrow$ Min gap b/w normalized number

$2^{104} \rightarrow$ Max — a — 11

IEEE \rightarrow non uniform distribution

* gap in denormals

$0\overline{0000000}/000 \dots 1$ $\xrightarrow{\text{to maintain the minimum}}$ gap

$$2^{-23} \times 2^{1-127} \Rightarrow 2^{-149}$$