* IEEE 754 Notations Lecture 13 Part 1

- Single Precision 32bit (4byte)
  - * format: S | E | M → 32bit
    - S: 0 +ve, 1 -ve
    - E: 8bit, ex-127
- double precision 64bit (8byte)
  - * format: S | E | M → 52bit
    - S: 0 +ve, 1 -ve
    - E: 11bit, ex-1023

* IEEE Normalized

1.xxx.....

do not store (implicit)

Ex: $(3.5)_{10}$ → Convert it into ieee Single precision normalized number

Step1 => Binary = $(11.1)_2$

Step2 => Normalized = $(11.1)_2$ => $1.11 * 2^1$

0 10000000 11000.....

- sign
- exponent
- mantissa

40600000H Ans

* $(.5)_{10} \Rightarrow (.1)_2 \Rightarrow 1.0 * 2^{-1}$

0 01111110 0...

3F000000H Ans

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

0 10000001 1001.....

40C8000H Ans

* 3FC00000H → IEEE Single Precision, decimal?

0 01111111 10000000....

S E M

$+ 1.1 * 2^{127-127} \Rightarrow 1.1 * 2^0 \Rightarrow (1.1)_2$

$(1.5)_{10}$ Ans

* 40C80000H → IEEE Single Precision, decimal?

0 10000001 10010000.....

E M

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

$(6.25)_{10}$ Ans

* Denormals => exponent all 0's

if exponent all !=0 means it is normalized

"1" is not implicit

* Ex: S e m
0 00000000 0000...

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

0 00000000 000...

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

- 1 00000000 000...

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

- 00000000H => +0
80000000H => -0

* These two are special number

- S E M
  0 11111111 000 ---

  $+\infty$

  7F800000H = $+\infty$ => +ve infinity

- S E M
  1 11111111 000 ---

  FF800000H => $-\infty$ => -ve infinity

both are special number

- 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
$+\infty$	0	all 1's	all 0's
$-\infty$	1	all 1's	all 0's
NAN	0/1	all 1's	≠0
Denormels	0/1	all 0's	≠0

Part 2

Largest * normalized number

- Largest +ve

S | E | M

0 | 1111 1110 | 1111 ---1

7 F7FFFFFF H Ans

$1.1111\ldots * 2^{254-127} \Rightarrow 2-2^{-23} * 2^{127}$

- Smallest +Ve

S | E | M

0 | 0000 0001 | 0000 ----

0 0800000 H Ans

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

- Smallest -Ve

$-(2-2^{-23}) * 2^{127}$

FF7FFFFF H Ans

- Largest -Ve

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

80800000 H Ans

* Denormalized number

- Largest +ve

S E

0 00000000 1111---1

007FFFFF H Ans

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

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

- Smallest ~~+ve~~ -ve

1 00000000 1111---

807FFFFF H

- Smallest +ve

S E M

0 00000000 0000---1

00000001H Ans $\Rightarrow 2^{-23} * 2^{1-127}$

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

$\Rightarrow 2^{-149}$ Ans

- Largest -ve

1 00000000000---1

80000001H

$-2^{-23} * 2^{-126} \Rightarrow -2^{-149}$ Ans

Background :-

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

SPN 0 00000001 0000---
00800000H ⇒ $1.0 * 2^{1-127} = 2^{-126}$

↓ Next Consecutive number

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

gap in Consecutive no. ⇒ $(1 + 2^{-23}) * 2^{-126} - 1 * 2^{-126}$

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

* Minimum gap : 2^{-149}

* Max gap :

0 1111 1110 1111---1
7F7FFFFFH ⇒ $2 - 2^{-23} * 2^{127}$

0 1111 1110 0000---0
$1.0 * 2^{254-127} = 2^{127}$

0 1111 1110 000---1
$(1 + 2^{-23}) * 2^{127} \Rightarrow 2^{104}$

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

2^{-149} → Min gap b/w normalized number

2^{104} → Max ——— a ——— <1

IEEE → Non uniform distribution

* gap in denormals

→ to maintain the minimum gap

0 00000000 000 --- 1

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