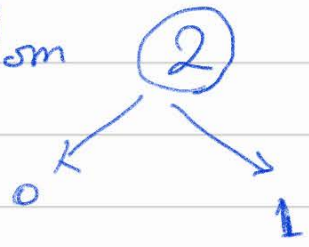
بسم الله الرحمن الرحيم

جمعة ٦/٧/١٤٠٠

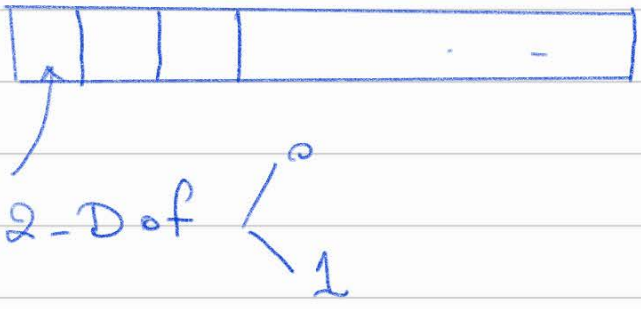
Number Representation

* Magnetic properties → Magnetic Moment

Degree of freedom (2)



N - Bits



* IEEE

type of Number Representation

- class (A) : Fixed Point Representation
- class (B) Floating Representation

Ex F.P.R.

$$R \equiv \underbrace{Sign}_{\text{علامة}} \times \left(\alpha_n 2^n + \alpha_{n-1} 2^{n-1} + \alpha_{n-2} 2^{n-2} + \cdots + \alpha_0 2^0 + \alpha_{-1} 2^{-1} \cdots \right)$$

$+ \alpha_{-m} 2^{-m})$

$$n + m = N - 1 - 1$$

علامت Sign

Ex: F. R. → $R = (1-2S)\, 2^{e - \text{Bias}} (1.f)$

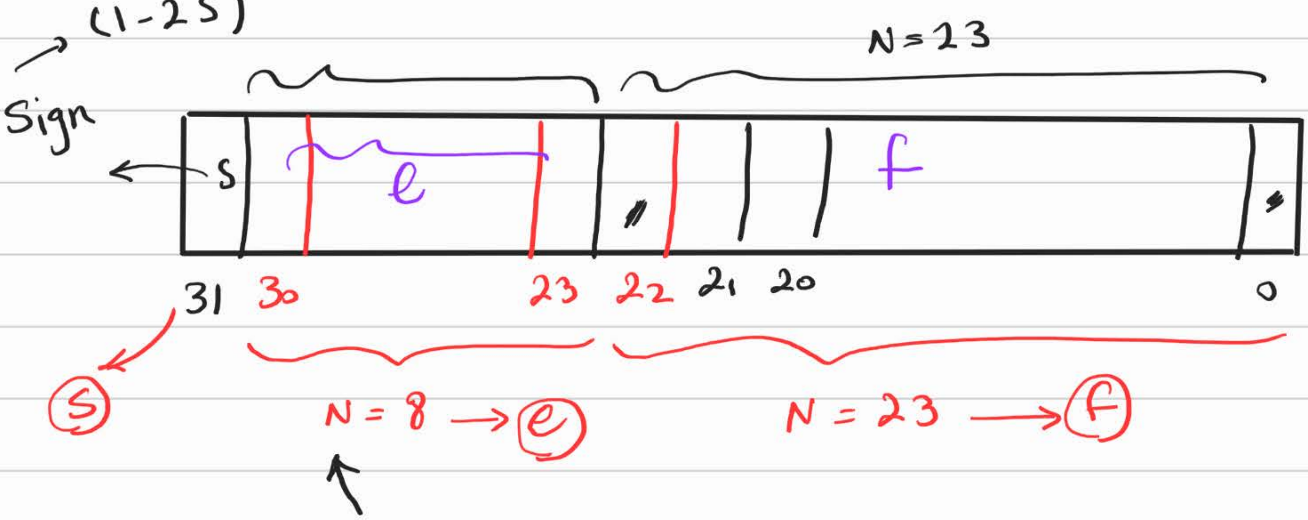
(0,1) — (+,-) علامت — نمای علمی — اعشار

Precision Classification

→ (A) Single Precision → N = 32 bits

→ (B) Double Precision → N = 64 bits

(C) Extended precision → $[\, 2^{-32766} - 2^{+3767} \,]$



بخش اعشاری

e - Bias

$$R = (1-2S)\, 2^{(e-\text{Bias})}\, (1.f)$$

2^8 — $[0, 255]$ $\xrightarrow{\text{IEEE}}$ $[1, 254]$

$1 \leq e \leq 254$ — 2^e

$e-127$

$\underline{\text{Bias} \equiv 127}$ $\longrightarrow$ 2 $\longrightarrow$ $\left[2^{-126} - 2^{+127}\right]$

اعداد خیلی کوچک را نیز می‌توان ضبط کرد.

بازه محدودی برای اعداد قابل ضبط در کامپیوتر

$$1.f \equiv 1 + f = 1 + \left[m_{22}\, 2^{-1} + m_{21}\, 2^{-2} + \dots + m_0\, 2^{-23}\right]$$

e - Bias

$$R_{max} = ? = (1-2S)\, 2^{(e-\text{Bias})}\, (1+f)$$

$$R_{max} = \begin{cases} S = 0 \longrightarrow +1 \\ e = 254 \longrightarrow 2^{+127} \\ f = \{1, 1, \dots, 1\} \end{cases}$$

$N = 23$

$$R_{max} = (+1)\, 2^{+127} \times \Big[1 + \underbrace{\sum_{j=0}^{22} 2^{-(j+1)}}_{1} \Big]$$

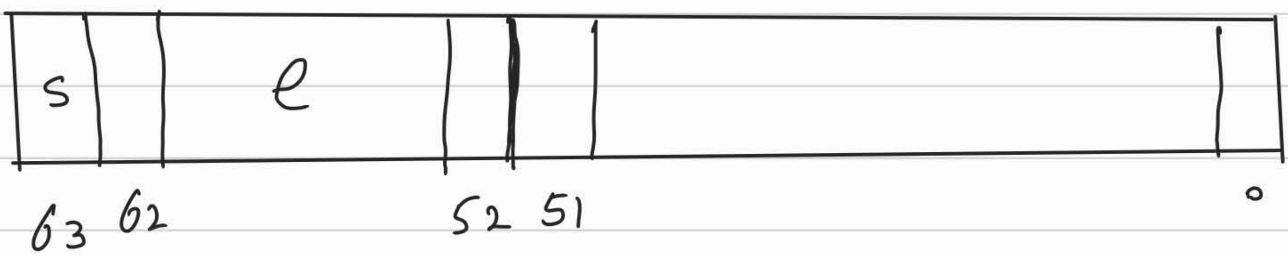
$$\boxed{R_{max} = +2^{127} \times 2 = +2^{128}}$$

قرارداد

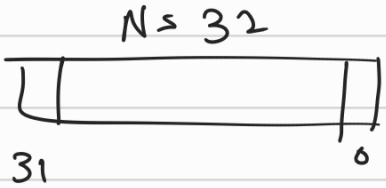
$$R_{min}^{(+)} = \begin{cases} S = 0 \\ e = 1 \\ f : \{\underbrace{0, 0 \ldots, 1}_{23}\} \end{cases} \Rightarrow +1 \times 2^{-126} \times (0 + 2^{-23})$$

$$\boxed{R_{min}^{(+)} = 2^{-149}}$$

For Double Precision $N = 64$



Ex $639.6875 \rightarrow$



$+639.6875$

↑

S = 0 , Bias = 127

e = ?

f = ?

$$639 = 1\times2^9 + 0\times2^8 + 0\times2^7 + 1\times2^6 + 1\times2^5 + 1\times2^4 + 1\times2^3 + 1\times2^2 + 1\times2^1 + 1\times2^0$$

$639 = 1\,00\,1111111$

$0.6875 \mid 2^{-1} = 0.5$ → 1

$0.1875 \mid 2^{-2}$ → 0

$0.1875 \mid 2^{-3} = 0.125$ → 1

0.1250

$0.0625 \mid 2^{-4} = 0.0625$ → 1

0.0000

$0.6875 = 1011$

2^{-1} 2^{-2} 2^{-3} 2^{-4}

$639.6875 = |00|||||||.|0||$

$2^7\ 2^5\ 2^3\ 2^1\ 2^{-1}\ 2^{-3}$

$2^9\ 2^8\ 2^6\ 2^4\ 2^2\ 2^0\ 2^{-2}\ 2^{-4}$

$(1-2S)\,2^{e-Bias}\,(1.f) = |00|||||||.|0||$

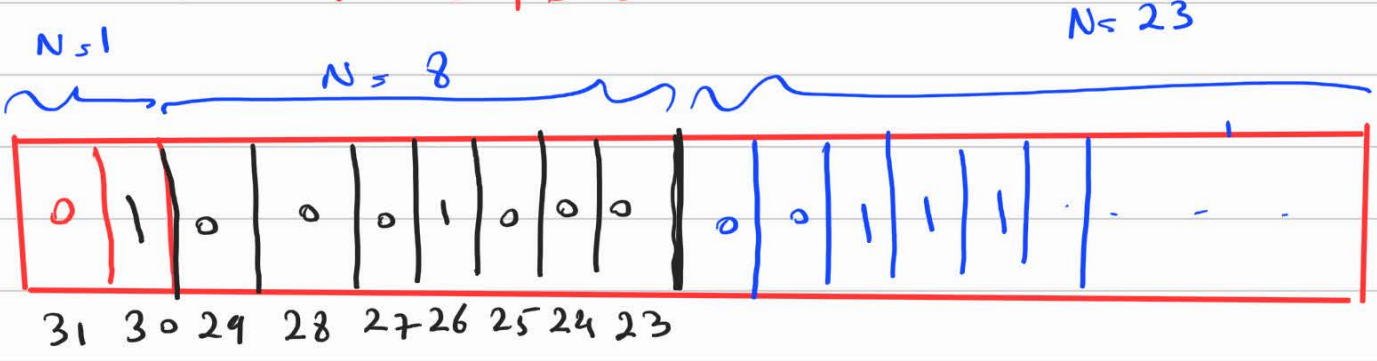
$(1-2S)\,2^{e-Bias}\,\boxed{(1.f)} = (+)\boxed{2^9} \times 1.00|||||||01|000000$

$S = 0$

$e - Bias = 9 \longrightarrow \boxed{e = 136}$

127

$136 = |000|000$



$N = 32$

Exercise $7 + \overline{10}^{7} = 9$

$= 7$

$\boxed{N = 32}$

Rounding Error Source

$$237 = 2.37 \times 10^{2}$$

$$111 = 1.11 \times 2^{2}$$