CSE 330 Assignment - 01

Solution :

① $\beta = 2$, $m = 3$, $e_{min} = -2$, $e_{max} = 2$ $e \in \{-2, -1, 0, 1, 2\}$

ⓐ Total no. of combinations for Standard form :

$$F = \pm (0.d_1d_2d_3)_2 \times 2^e$$
$$= \pm (0.1\, d_2d_3)_2 \times 2^e$$

$$\begin{bmatrix} 0\,0 \\ 0\,1 \\ 1\,0 \\ 1\,1 \end{bmatrix} \rightarrow \text{4 possible combinations.}$$

$4 \times 5 = 20$ possible combinations. (Ans)

Total no. of combinations for Normalized form :

$$F = \pm (1.d_1d_2d_3)_2 \times 2^e$$

$$\begin{bmatrix} 0\,0\,0 \\ 0\,0\,1 \\ 0\,1\,0 \\ 0\,1\,1 \\ 1\,0\,0 \\ 1\,0\,1 \\ 1\,1\,0 \\ 1\,1\,1 \end{bmatrix} \rightarrow \text{8 possible combinations}$$

$8 \times 5 = 40$ possible combinations. (Ans)

CS CamScanner

Total no. of combinations for Denormalized form:

$$F = \pm(\underbrace{0.1}\, d_1 d_2 d_3)_2 \times 2^e$$

$$= \pm(0.1\, \underbrace{d_1 d_2 d_3}_{})_2 \times 2^e$$

$$\begin{bmatrix} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{bmatrix} \longrightarrow \text{8 possible combinations}$$

$8 \times 5 = 40$ possible combinations (Ans)

(b) Maximum/largest number for Standard form:

$$F = \pm(0.111)_2 \times 2^2$$

$$= (2^{-1} + 2^{-2} + 2^{-3}) \times 2^2$$

$$= \frac{7}{2} \text{ (Ans)}$$

Maximum/largest number for Normalized form:

$$F = (1.111)_2 \times 2^2$$

$$= (2^0 + 2^{-1} + 2^{-2} + 2^{-3}) \times 2^2$$

$$= \frac{15}{2} \text{ (Ans)}$$

Maximum/largest number for Denormalized form:

$$F = (0.1111)_2 \times 2^2$$

$= (2^{-1} + 2^{-2} + 2^{-3} + 2^{-4}) \times 2^{2}$

$= \frac{15}{4}$ (Ans).

(c) Smallest number for Standard Form (non-negative):

$F = (0.100)_2 \times 2^{-2}$

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

$= \frac{1}{8}$ (Ans)

Smallest number for Normalized Form (non-negative):

$F = (1.000)_2 \times 2^{-2}$

$= (2^{0} \times 2^{-2})$

$= \frac{1}{4}$. (Ans)

Smallest number for Denormalized Form (non-negative)

$F = (0.1000)_2 \times 2^{-2}$

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

$= \frac{1}{8}$. (Ans)

(d) Maximum/largest numbers will be same as the answers in (b).

Minimum/smallest number for Standard form (negative support)

$F = -(0.111)_2 \times 2^{2}$

$= -\frac{7}{2}$. (Ans)

Minimum/smallest number for Normalized form (negative support)

$$F = -(1.111)_2 \times 2^2$$
$$= -\frac{15}{2} \text{ (Ans)}$$

Minimum/smallest number for Denormalized form (negative support)

$$F = -(0.1111)_2 \times 2^2$$
$$= -\frac{15}{4} \text{ (Ans)}$$

(2) $x = (6.235)_{10}$

(a)

$$\begin{array}{r|l|l}
2 & 6 & \\
\hline
2 & 3 & 0 \\
\hline
2 & 1 & 1 \\
\hline
 & 0 & 1
\end{array} \uparrow \text{LSB}$$

$(6)_{10} = (110)_2$

$$\begin{array}{r|l}
 & .235 \\
 & \times\ 2 \\
\hline
 & .47 \\
0 & \times\ 2 \\
\hline
0 & .94 \\
 & \times\ 2 \\
\hline
1 & .88 \\
 & \times\ 2 \\
\hline
1 & .76 \\
 & \times 2 \\
\hline
1 & .52 \\
 & \times\ 2 \\
\hline
1 & .04
\end{array}$$

MSB ↓

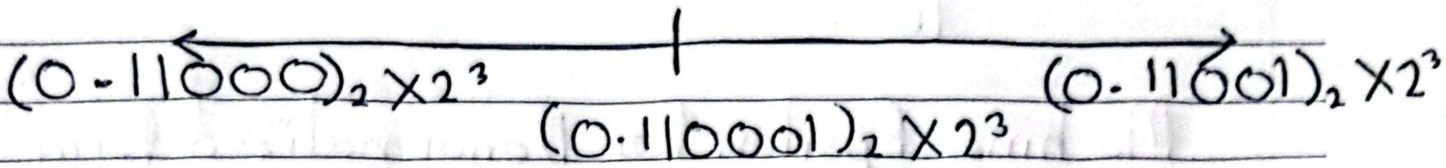
$(.235)_{10} = (001111)_2$

$(6.235)_{10} = (110.001111)_2$

(b) $m = 5$ [Using General/Standard Form]

$$(6.235)_{10} = \underbrace{(0.110001111)_2 \times 2^3}_{\text{Standard Form}}$$

Considering $m = 5$ then,



$fl(x) = (0.11001)_2 \times 2^3$

(c) $m = 5$ then,

$$fl(x) = (0.11001)_2 \times 2^3$$
$$= (1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-5})_2 \times 2^3$$
$$= \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{32}\right) \times 2^3$$
$$= 6.25$$

$$\delta = \left|\frac{fl(x) - x}{x}\right|$$

$$\delta = \left|\frac{6.25 - 6.235}{6.235}\right|$$

$$\delta = 2.406 \times 10^{-3}$$

③ $\beta = 2$, $m = 3$, exponent $= 2$.

ⓐ Machine Epsilon for Normalized form:

$$\epsilon = \frac{1}{2}\beta^{-m}$$

$$\epsilon = \frac{1}{2} \times (2)^{-3}$$

$$\epsilon = \frac{1}{16} \text{ (Ans)}$$

Machine Epsilon for Denormalized form:

$$\epsilon = \frac{1}{2}\beta^{-m}$$

$$\epsilon = \frac{1}{2} \times (2)^{-3}$$

$$\epsilon = \frac{1}{16} \text{ (Ans)}$$

ⓑ $|x|_{min}$ for Normalized Form:

$$|x|_{min} = \beta^{e}$$

$$|x|_{min} = 2^{2}$$

$$|x|_{min} = 4 \text{ . (Ans)}$$

$|x|_{min}$ for Denormalized Form:

$$|x|_{min} = \beta^{-1}\beta^{e}$$

$$|x|_{min} = (2)^{-1} \times (2)^{2}$$

$$|x|_{min} = 2 \text{ (Ans)}$$

(c) Maximum Delta Value means Machine Epsilon value

$$\delta_{max} = \epsilon = \frac{1}{2} \beta^{1-m}$$

$$\delta_{max} = \frac{1}{2} \times (2)^{1-3}$$

$$\delta_{max} = \frac{1}{8} \text{ (Ans)}$$