A. Let's take, $\beta = 2, m = 4$ and $-2 \leqslant e \leqslant 2$. Based on these, answer the following questions:

1. (5 marks) For last two forms (normalized and denormalized form), find the smallest positive and negative numbers with the largest number representable by the system. Express these numbers in both binary and decimal formats for all two forms.
2. (2 marks) How many numbers can this system represent or store in all these three forms?
3. (3 marks) Using standard form, find all the decimal numbers for e = -2 and e = 2 without negative support, plot them on a real line, and show if the number line is equally spaced or not.

A. Given that $\beta = 2,\ m = 4,\ e_{min} = -2, e_{max} = 2$

1.

[Normalized Form]

(Smallest positive)

$$= (0.10000)_2 \times 2^{-2}$$
$$= 1 \times 2^{-1} \times 2^{-2}$$
$$= \left(\frac{1}{8}\right)_{10}$$

(Largest positive)

$$= (0.11111)_2 \times 2^{2}$$
$$= \left(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5}\right) \times 2^{2}$$
$$= \frac{31}{32} \times 2^{2}$$
$$= \left(\frac{31}{8}\right)_{10}$$

(Smallest negative)

$$= -(0.11111)_2 \times 2^{2}$$
$$= -\left(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5}\right) \times 2^{2}$$
$$= -\frac{31}{32} \times 2^{2}$$
$$= -\left(\frac{31}{8}\right)_{10}$$

[Denormalized Form]

(Smallest positive)

$= (1.0000)_2 \times 2^{-2}$

$= 2^{-2}$

$= \left(\frac{1}{4}\right)_{10}$

(Largest positive)

$= (1.1111)_2 \times 2^2$

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

$= \frac{31}{16} \times 2^2$

$= \left(\frac{31}{4}\right)_{10}$

(Smallest negative)

$= -(1.1111)_2 \times 2^2$

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

$= -\frac{31}{16} \times 2^2$

$= -\left(\frac{31}{4}\right)_{10}$

2.

[Standard Form]

Total numbers $= 2 \times 2^3 \times 5 = 80$ (Considering + & - numbers)

[Normalized Form]

Total numbers $= 2 \times 2^4 \times 5 = 160$ (Considering + & - numbers)

[Denormalized Form]

Total numbers $= 2 \times 2^4 \times 5 = 160$ (Considering + & - numbers)

3.

[Standard Form]

For $e = -2$,

$= (0.1000)_2 \times 2^{-2} = \frac{1}{8}$

$= (0.1001)_2 \times 2^{-2} = \frac{9}{64}$

$= (0.1010)_2 \times 2^{-2} = \frac{5}{32}$

$= (0.1011)_2 \times 2^{-2} = \frac{11}{64}$

$= (0.1100)_2 \times 2^{-2} = \frac{3}{16}$

$= (0.1101)_2 \times 2^{-2} = \frac{13}{64}$

$= (0.1110)_2 \times 2^{-2} = \frac{7}{32}$

$= (0.1111)_2 \times 2^{-2} = \frac{15}{64}$

For $e = 2$,

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

$= (0.1001)_2 \times 2^{2} = \frac{9}{4}$

$= (0.1010)_2 \times 2^{2} = \frac{5}{2}$

$= (0.1011)_2 \times 2^{2} = \frac{11}{4}$

$= (0.1100)_2 \times 2^{2} = 3$

$= (0.1101)_2 \times 2^{2} = \frac{13}{4}$

$= (0.1110)_2 \times 2^{2} = \frac{7}{2}$

$= (0.1111)_2 \times 2^{2} = \frac{15}{4}$

Now, if you plot all these numbers in a real line, you will find that the numbers are equally spaced for a specific exponent. However, considering e = -2 and e = 2, all numbers are not equally spaced. Therefore, numbers are not equally spaced.

B. Let's take, $\beta = 2, m = 4$ and $-2 \leqslant e \leqslant 3$. Based on these, answer the following questions:

1. (4 marks) Compute the minimum of $|x|$ for normalized and denormalized form.
2. (4 marks) Compute the machine epsilon or unit roundoff value for the normalized and denormalized form.
3. (2 marks) Compute the maximum delta value for the standard form.

B. Given that $\beta = 2,\ m = 4,\ e_{min} = -2, e_{max} = 3$

1.

[Normalized Form]

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

$$= 2^{-1} \times 2^{-2}$$
$$= \frac{1}{8}$$

[Denormalized Form]

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

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

2.

[Normalized Form]

$$\epsilon_{m} = \frac{1}{2}\beta^{-m}$$
$$= \frac{1}{2} \times 2^{-4}$$
$$= \frac{1}{32}$$

[Denormalized Form]

$$\epsilon_{m} = \frac{1}{2}\beta^{-m}$$
$$= \frac{1}{2} \times 2^{-4}$$
$$= \frac{1}{32}$$

3.

[Standard Form]

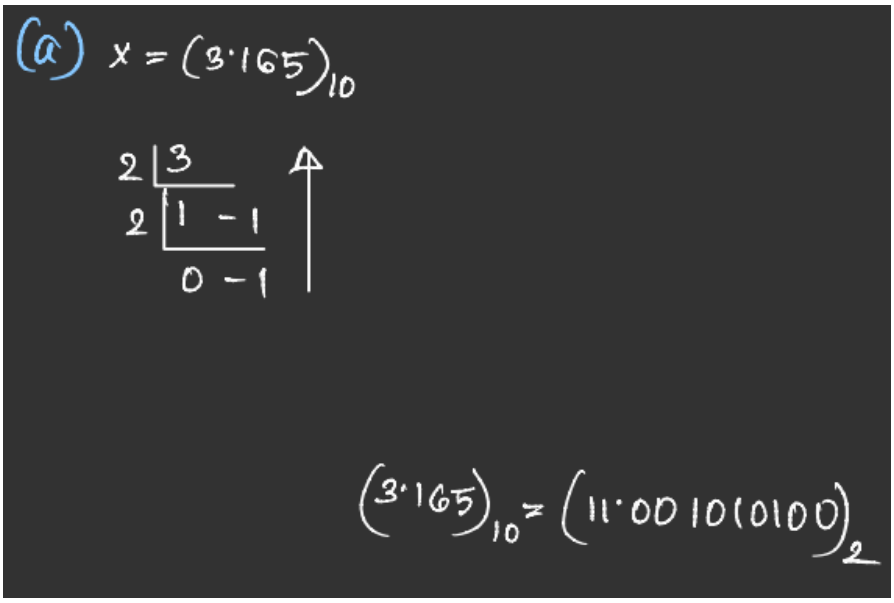
$$\delta_{max} = \frac{|fl(x)-x|_{max}}{|x|_{min}}$$
$$= \frac{\frac{1}{2}\beta^{-m}\beta^{e}}{\beta^{-1}\beta^{e}}$$
$$= \frac{\frac{1}{2}2^{-4}2^{3}}{2^{-1}2^{-2}}$$
$$= 2$$

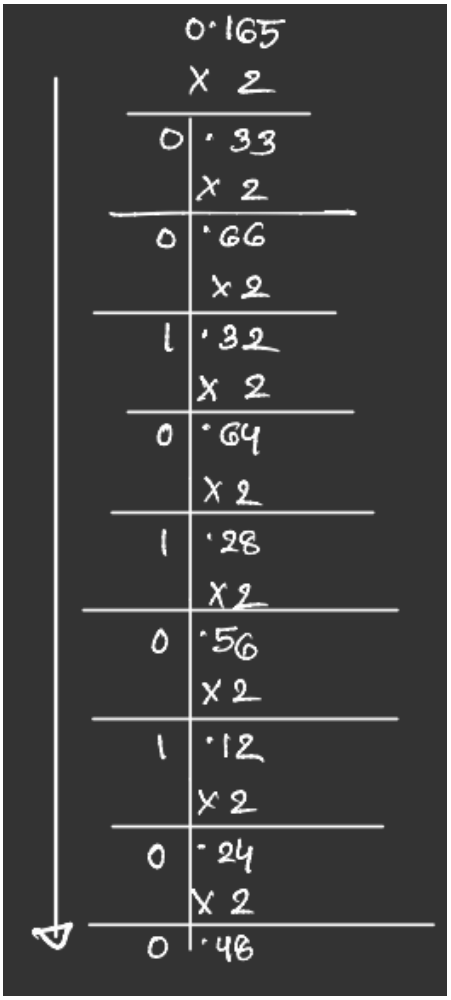
C. Consider the real number $x = (3.165)_{10}$

1. (3 marks) Convert the decimal number x in binary format at least up to 9 decimal/binary places.
2. (4 marks) What will be the binary value of x [Find $fl(x)$] if you store it in a system with $m = 4$ and $m = 6$ using the standard form.
3. (3 marks) Now convert back to the decimal form the stored values you obtained in the previous part, and calculate the rounding error of both numbers.

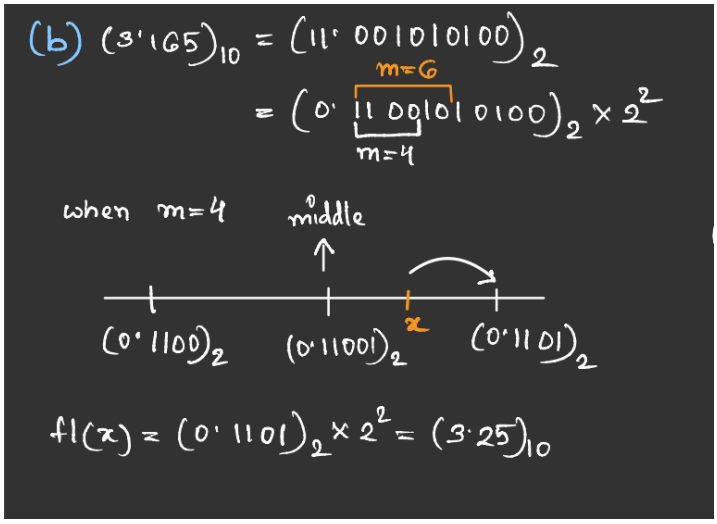
C. Given that $real\ number\ x = 3.165_{10}$

1.

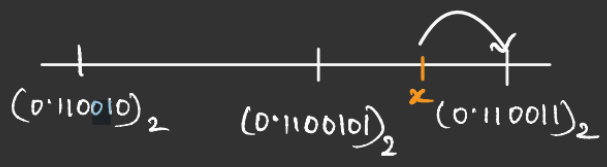



0·165
X 2
0 | ·33
X 2
0 | ·66
X 2
1 | ·32
X 2
0 | ·64
X 2
1 | ·28
X 2
0 | ·56
X 2
1 | ·12
X 2
0 | ·24
X 2
0 | ·48

2.


(b) $(3.165)_{10} = (11.00101010100)_2$
m=6
$= (0.\underbrace{1100101}_{m=4}0100)_2 \times 2^2$
when m=4
middle
$(0.1100)_2$
$(0.11001)_2$
x
$(0.1101)_2$
$fl(x) = (0.1101)_2 \times 2^2 = (3.25)_{10}$

when $m = 6$



$$fl(x) = (0.1100011)_2 \times 2^2 = (3.1875)_{10}$$

3.

(c) For $m = 4$,

$$\delta = \frac{|fl(x) - x|}{|x|} = \frac{|3.25 - 3.165|}{|3.165|} = 0.02685 \text{ (approx.)}$$

For $m = 6$,

$$\delta = \frac{|fl(x) - x|}{|x|} = \frac{|3.1875 - 3.165|}{|3.165|} = 7.109 \times 10^{-3} \text{ (approx.)}$$