

Birou Hoes

Group: 911

Subject 3 - Student 1

• $x = 84230,12$

• $SP = 32 \text{ bits}$, mantissa > 1

SP - single precision $\Rightarrow m = 32 \text{ bits}$

- e on 8 bits

- $q = 127$

- m on 23 bits

Floating point representation of real numbers

- used to represent very large and very small numbers with a high precision
- if there is an overflow, then the least significant digits are lost.

• Conversion of the integer part: successive divisions by 8

$$\begin{array}{r} 84230 \overline{) 8} \\ \underline{8} \\ 04 \\ \underline{0} \\ 42 \\ \underline{40} \\ 23 \\ \underline{20} \\ 70 \\ \underline{64} \\ 6 \end{array}$$

it 1 $8:8 = 1, r 0$

it 2 $4:8 = 0, r 4$

it 3 $42:8 = 5, r 2$

it 4 $23:8 = 2, r 7$

it 5 $70:8 = 8, r 6$

$$\begin{array}{r|l} 10528 & 8 \\ \hline 1 & 1316 \\ \hline 25 & \\ \hline 12 & \\ \hline 48 & \\ \hline 0 & \end{array}$$

$$\begin{aligned} \text{st 1 } 10:8 &= 1, \wedge 2 \\ \text{st 2 } 25:8 &= 3, \wedge 1 \\ \text{st 3 } 12:8 &= 1, \wedge 4 \\ \text{st 4 } 48:8 &= 6, \wedge 0 \end{aligned}$$

$$\begin{array}{r|l} 1316 & 8 \\ \hline 1 & 164 \\ \hline 51 & \\ \hline 36 & \\ \hline 9 & \end{array}$$

$$\begin{aligned} \text{st 1 } 13:8 &= 1, \wedge 5 \\ \text{st 2 } 51:8 &= 6, \wedge 3 \\ \text{st 3 } 36:8 &= 4, \wedge 4 \end{aligned}$$

$$\begin{array}{r|l} 164 & 8 \\ \hline 1 & 20 \\ \hline 04 & \\ \hline 9 & \end{array}$$

$$\begin{aligned} \text{st 1 } 16:8 &= 2, \wedge 0 \\ \text{st 2 } 04:8 &= 0, \wedge 4 \end{aligned}$$

$$\begin{array}{r|l} 20 & 8 \\ \hline 1 & 2 \\ \hline 9 & \end{array}$$

$$\text{st 1 } 20:8 = 2, \wedge 4$$

$$\begin{array}{r|l} 2 & 8 \\ \hline 1 & 0 \\ \hline 2 & \end{array}$$

$$\text{st 1 } 2:8 = 0, \wedge 2$$

Now we take the remainders from the last one

$$\Rightarrow 84230_{(10)} = 244406_8$$

Furthermore we will use rapid conversion.

$$244406_8 = 010\ 100\ 100\ 100\ 000\ 110_{(2)} \quad (\text{rapid conversion})$$

- Conversion of the fractional part: successive multiplication by 8

$$\begin{array}{r} 0,12_x \\ \times 8 \\ \hline 0,96 \end{array}$$

$$\begin{array}{r} 0,96_x \\ \times 8 \\ \hline 7,68 \end{array}$$

$$\begin{array}{r} 0,68_x \\ \times 8 \\ \hline 5,44 \end{array}$$

So we have $0,12 = 0,075_{(8)} = 0,000111101_{(2)}$ (rapid conversion)

From the number above we will only take 7 digits from fractional part, as we have 16 bits in total part, so $23 - 16 = 7$

- $X = 84230,12_{(10)} = 10100100100000110,0001111_{(2)}$

$$X = 1, \underbrace{01001001000001100001111}_{\text{mantissa}}_{(2)} \times 2^{\textcircled{16}=2}$$

↓
hidden bit

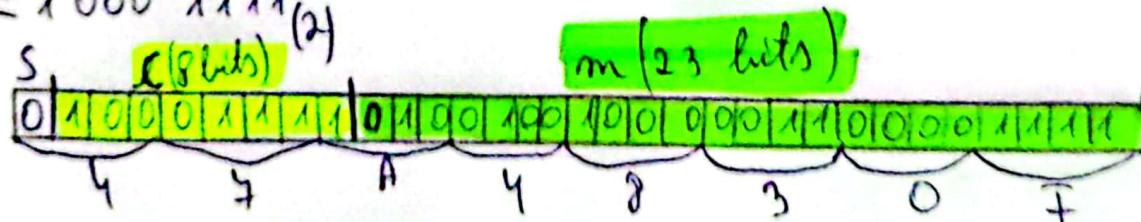
- $C = e + 127$

$$= 16 + 127$$

$$= 143$$

$$= 2^7 + 2^3 + 2^2 + 2^1 + 2^0$$

$$= 10001111_{(2)}$$



$$\Rightarrow 47A48307_{(16)}$$