Birou Rares
Group: 911

Subject 3 - Student 1

- $x = 84230,12$
- SP = 32 bits , mantissa > 1

SP - single precision => n = 32 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

84230 | 8
8 | 10528
04
0
42
40
23
16
70
64
(6)

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

10528 | 8
1316

25
12
48
0

it 1 10 : 8 = 1, r 2
it 2 25 : 8 = 3, r 1
it 3 12 : 8 = 1, r 4
it 4 48 : 8 = 6, r 0

1316 | 8
164

51
36
4

it 1 13 : 8 = 1, r 5
it 2 51 : 8 = 6, r 3
it 3 36 : 8 = 4, r 4

164 | 8
20

04
4

it 1 16 : 8 = 2, r 0
it 2 04 : 8 = 0, r 4

20 | 8
2

4

it 1 20 : 8 = 2, r 4

2 | 8
0

2

it 1 2 : 8 = 0, r 2.

Now we take the remainders from the last one

$\Rightarrow 84230_{(10)} = 244406_{(8)}$

Furthermore we will use rapid conversion.

$244406_{(8)} = 010\ 100\ 100\ 100\ 000\ 110_{(2)}$ (Rapid conversion)

- Conversion of the fractional part: successive multiplication by 8

$0,12 \times 8 = \textcircled{0},96$ $\quad$ $0,96 \times 8 = \textcircled{7},68$ $\quad$ $0,68 \times 8 = \textcircled{5},44$

So we have $0,12 = 0,075_{(8)} = 0,000\ 111\ 101_{(2)}$ (Rapid conversion)

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

- $X = 84230,12_{(10)} = 10\ 100\ 100\ 100\ 000\ 110,000\ 1111_{(2)}$

$X = 1,0100\ 100\ 100\ 000\ 110\ 000\ 1111_{(2)} * 2^{16}$ $\quad$ $16 = e$

1 → hidden bit

$0100\ 100\ 100\ 000\ 110\ 000\ 1111$ → mantissa

- $C = e + 127$

$= 16 + 127$

$= 143$

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

$= 1000\ 1111_{(2)}$

S | C (8 bits) | m (23 bits)

0 | 1 0 0 0 1 1 1 1 | 0 1 0 0 1 0 0 1 0 0 0 0 0 1 1 0 0 0 0 1 1 1 1

4 7 A 4 8 3 0 F

$\Rightarrow 47A4830F_{(16)}$