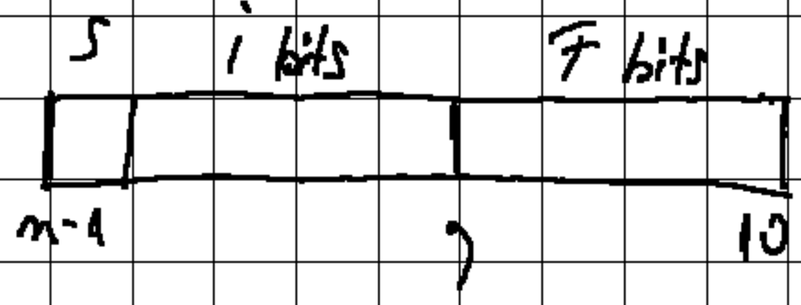
Fixed point repres. of real numbers



$n = 1 + i + F$ bits

$n = 32$ bits , $i = 15$ bits $F = 16$ bits

$X = 1234,56$

$1234 = 2322_{(8)} = 010\,011\,010\,010_{(2)}$

1234	8			
1	154	8		
43	1	19	8	
1	74	1	2	8
34	1	3	1	0
1	2		2	
2				

$0,56 \cdot 8 = (4),48$

$0,48 \cdot 8 = (3),84$

$0,84 \cdot 8 = (6),72$

$0,72 \cdot 8 = (5),76$

$0,76 \cdot 8 = (6),08$

$0,08 \cdot 8 = (0),64$ V

$0,56 = 0,436560_{(8)} =$

$= 0,100011110101110000_{(2)}$

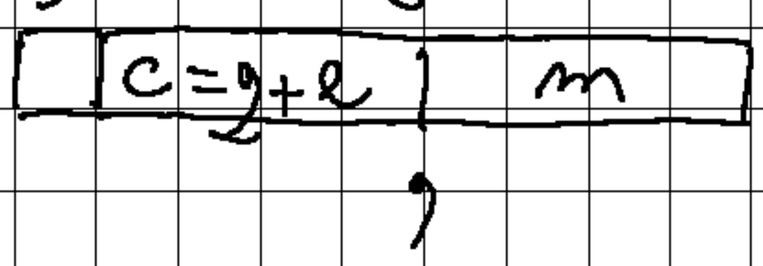
S | $i = 15$ bits | $F = 16$ bits

0	000010011010010	1000111101011100

0 4 D 2 8 F 5 C

Homework: 10.5.

Ex2: Floating-point representation



$1234,56 = 0,123456 \cdot 10^{4}$ → exp

$0,013 = 0,13 \cdot 10^{-1}$

Single precision: 32 bits, e = 127, c on 8 bits, m on 23 bits

Double precision: 64 bits, e = 1023, c on 11 bits, m on 52 bits

$x = -465,78$, SP, $m > 1$

465	8	8	8
1	58	7	0
65	1	1	
1	2	7	
1			

$456 = 721_{(8)}$

$0,78 = 0,61727_{(8)}$

$0,78 \cdot 8 = 6,24$

$0,24 \cdot 8 = 1,92$

$0,92 \cdot 8 = 7,36$

$0,36 \cdot 8 = 2,88$

$0,88 \cdot 8 = 7,04$

$x = -465,78 = -721,61727$

$= -111010001,110001111010111$

$= -1,11010001110001111010111$

hidden bit

m

C 3 E 8 E 3 D 7

1	1 0 0 0 0 1 1 1	1 1 0 1 0 0 0 1 1 1 0 0 0 1 1 1 1 0 1 0 1 1 1
S	8 bits	m, 23 bits

$C = 127 + 8 = 135$

$135 = 128 + 4 + 2 + 1 = 2^7 + 2^2 + 2^1 + 2^0 = 10000111_{(2)}$

EX3: $X = 0,17$, SP, $m < 1$

$0,17 \cdot 8 = 1,36$

$0,36 \cdot 8 = 2,88$

$0,88 \cdot 8 = 7,64$

$0,04 \cdot 8 = 0,32$

$0,32 \cdot 8 = 2,56$

$0,56 \cdot 8 = 4,48$

$0,48 \cdot 8 = 3,84$

$0,84 \cdot 8 = 6,72$

$0,72 \cdot 8 = 5,76$

$0,17 = 0,12702436574_{(8)}$

$= 0,00101011100001010001111010101_{(2)}$

$= 0,10101111000010100011110101_{(2)} \cdot 10^{-2}$ → exp.

$C = e+2 = -2+127 = 125 = 64+32+16+8+4+1 =$

$= 2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^0 = 01111101_{(2)}$

S C, 8 bits ; m, 23 bits

0	0 1 1 1 1 1 0 1	1 0 1 0 1 1 1 0 0 0 0 1 0 1 0 0 0 1 1 1 1 0 1

3 E D 7 0 A 3 D

Ex4: Find the real number X having $C1DE0000_{(16)}$ as it's fixed-point repres on 32 bits with $i = 12$ bits and $f = 19$ bits.

C 1 D E 0 0 0 0

1	1 0 0 0 0 0 1 1 1 0 1 1	1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

S $i = 12$ bits ; $f = 19$ bits

$X = -100000111011,11_{(2)} =$

$= -\left(2^{11} + 2^5 + 2^4 + 2^3 + 2^1 + 2^0 + 2^{-1} + 2^{-2}\right) =$

$= -\left(2048 + 32 + 16 + 8 + 2 + 1 + \frac{1}{2} + \frac{1}{4}\right) = -2107,75$

Ex: Find the real number x having $C1DE0000_{(16)}$ as IEEE floating point repres. SP, m > 1

$C = 10000011_{(2)} = 2^7 + 2 + 1 = 128 + 3 = 131$

$l = 131 - 127 = 4$

$X = -1,101111 \cdot 10^4 = -11011,11_{(2)}$

hidden bit

$-11011,11_{(2)} = -(2^4 + 2^3 + 2 + 1 + 0,5 + 0,25) =$

$= -27,75_{(10)}$