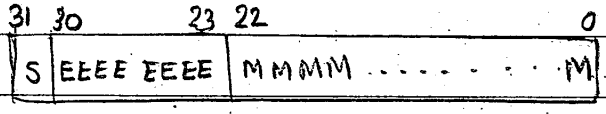
U3A Maths

7-May-2024 Term 2 / Week 2

IEEE Standard 754 Single Precision Floating Point Number

- Single precision floating point number as per IEEE Standard 754 is defined as below.

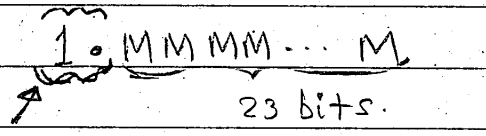


Bit 31 ⇒ Sign bit (0 ⇒ +; 1 ⇒ −)

Bits 30-23 ⇒ 8 bit Exponent.

Bits 22-0 ⇒ 23 bit Mantissa.

- IEEE 754 single precision number has a total of 32 binary digits (bits)
- The most significant bit of Mantissa is always 1, and hence it is not stored! For Ex:



Always assumed as 1, and not stored!

- Hence we effectively get a 24 bit number!

- The Exponent part is stored as a "biased number", with a bias of a bias value of 127 − the mid-way value.
- for Ex:

2^4 is stored as $127+4=131$

2^{-3} is stored as $127-3=124$

- For clarity a table for exponent representation is as given below:

- 8 bit binary numbers have values ranging from 0 to 255

Dec	Binary	Biased Exponent
0	0000 0000	Special use (0)
1	0000 0001	−126 (−126+127) =1
2	0000 0010	−125 (−125+127) =2
⋮	⋮	⋮
125	0111 1101	−2 (−2+127) =125
126	0111 1110	−1 (−1+127) =126
127	0111 1111	0 (0+127)
128	1000 0000	+1 (1+127)
129	1000 0001	+2 (2+127)
⋮		
253	1111 1101	+126 (126+127)
254	1111 1110	+127 (127+127)
255	1111 1111	Special use (∞)

Home work Problems

Express the following decimal values in IEEE 754 Single precision form.

Ex.1 26.625

converting to binary form we have

11001.101

$\therefore$ IEEE 754 form is

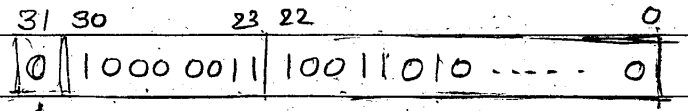
1.1001101×2^4 ← Exponent

Stored Mantissa.

$\therefore$ Biased exponent is

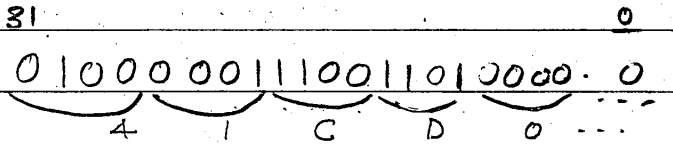
$4 + 127 = 131$

$(131)_{10} = (1000\ 0011)_2$



Sign(+)

The above value is normally written as hexa decimal value including the sign bit!



$(41CD\ 0000)_{hex}$

Notes

1000 - 8	1010 - A	1100 - C	1110 - E
1001 - 9	1011 - B	1101 - D	1111 - F

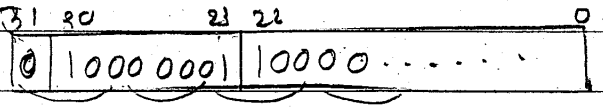
Ex.2 6.0 $\Rightarrow$ 110.0

Std form is

1.100×2^2

Mantissa

Exponent: $2 + 127 = 129 = (1000\ 0001)_2$



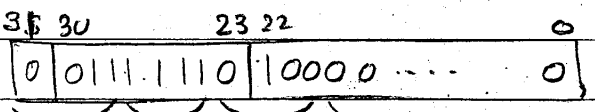
$(40C0\ 0000)_{hex}$

Ex.3

$0.75 = 0.11$

Std form: 1.1×2^{-1}

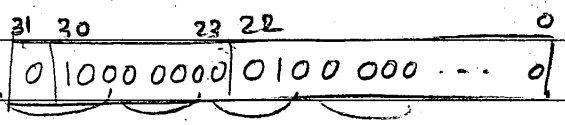
Exponent $-1 + 127 = 126 = (0111\ 1110)_2$



$(3F40\ 0000)_{hex}$

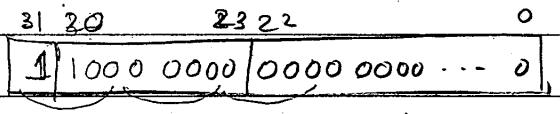
Ex.4 $2.5 = (10.1)_2 = 1.01 \times 2^1$

Exponent $1 + 127 = 128 = (1000\ 0000)_2$



$(4020\ 0000)_{hex}$

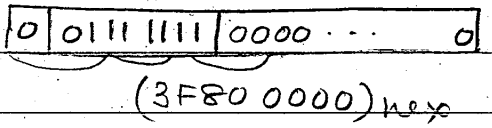
Ex.5 $-2 = -10 \Rightarrow -1.0 \times 2^1$



$(C000\ 0000)_{hex}$

Ex.6 $1 \Rightarrow 1.0 \times 2^0$

Exponent $0 + 127 = 127 = (0111\ 1111)_2$



$(3F80\ 0000)_{hex}$

Ex.7 0 is a special case!