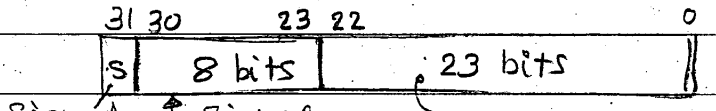
U3A Maths (No class on 30-Apr-24)

9-Apr-2024 · Term 1 / Week 11

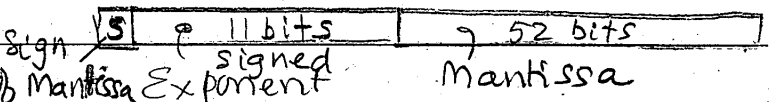
Floating Point Number
IEEE Standard 754

- In the early days, computer manufacturers had their own way of representing floating point numbers
- In 1985, Institution of Electrical & Electronic Engineers (USA) introduced the standard IEEE 754 to represent floating point numbers.

- The standard is a bit quirky, but it is aimed at the efficiency of the hardware design for floating point processors.
- The standard provides for:
  - 32 bit (Single precision) numbers



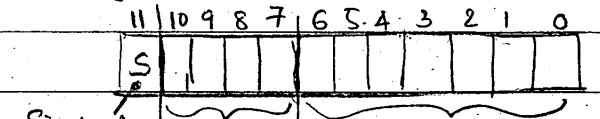
  - This provides a number range from $\approx 10^{-38}$ to 10^{+38}
  - 64 bit (Double Precision) numbers



This provides a range of $\approx 10^{-308}$ to 10^{+308}

- The quirky features of IEEE Standard 754 are:
  - Most significant bit of the mantissa is not explicitly stored, since it is always binary 1. Hence, we effectively get a 24 bit mantissa with only 23 bits!
  - Due to explicit sign bit for mantissa; +0 & -0 are are treated as equal for comparision.

  - Most confusing part is the signed exponent; which uses a "bias" system, instead of 2's complement! There is of course, a reason for this madness!!
- To understand biased exponent, let us go back to our hypothetical 12 bit computer



- Let us also consider the same Example of last week.

Example

Express $(25.625)_{10}$ in (IEEE 754) floating point format

We have;

$(25.625)_{10} = (11001.101)_2$

IEEE 754 standard form is

1.1001101×2^{4} ← Exponent

1 — This bit is always 1 and it is not stored

1001101 — Stored Mantissa

11	10	9	8	7	6	5	4	3	2	1	0
0	?	?	?	?	1	0	0	1	1	0	1

Sign (+) — Mantissa.

Range of Exponent (4 binary digits)

Dec	Binary	Bias 7	2's Compl.
0	0000	−7 Reserved by IEEE 754	0
1	0001	−6	1
2	0010	−5	2
3	0011	−4	3
4	0100	−3	4
5	0101	−2	5
6	0110	−1	6
7	0111	0	7
8	1000	1	−8
9	1001	2	−7
10	1010	3	−6
11	1011	4	−5
12	1100	5	−4
13	1101	6	−3
14	1110	7	−2
15	1111	8 Reserved	−1

Ascending ↓

Ascending ↑

??

- For our example, the Biased exponent is obtained by adding the middle number (7) to the actual exponent
- We have 2^4, that is Exponent = 4

∴ Biased Exponent $= 4 + 7 = (11)_{10}$

$(11)_{10} = (1011)_2$

Decimal ↑ Binary.

Hence the $(25.625)_{10}$ is stored as

11	10	9	8	7	6	5	4	3	2	1	0
0	1	0	1	1	1	0	0	1	1	0	1

Exponent (2^4) — Mantissa

Sign of the Mantissa (+)

In Biased exponent form, the exponents are in an ascending order. In 2's complement form, the exponent values are topsy turvy! (goes from +7 to −8)

- Finally, IEEE 754 reserves the first (0) & last (15) numbers for other uses. Hence, the exponent range is 2^{-6} to 2^{+7} for our Example.
- For IEEE 754 single precision, the exponent range is 8 bits i.e., 0 to 255. Bias is 127. The range 1 to 254, i.e 2^{-126} to 2^{+127}.

Home work Express in IEEE 32 bit form

① 25.625 ② 6.0 ③ 0.075

④ 2.5 ⑤ −2 ⑥ 1 ⑦ 0