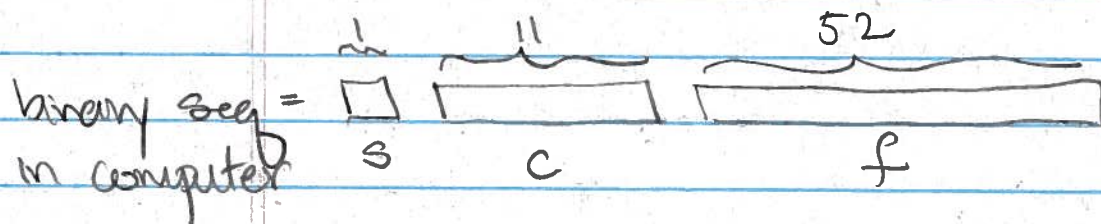


Lecture 2 Computer Arithmetic & Round-off error



number
that binary
seq represents

$$(-1)^s 2^{c-1023} (1+f)$$

sets the "scale" sets "precision"

Counting in binary:

10	9	8	...	3	2	1	0		
0	0	0	...	0	0	0	0	= 0 =	$0 \cdot 2^0$
0	0	0	...	0	0	0	1	= 1	$1 \cdot 2^0$
0	0	0	...	0	0	1	0	= 2 =	$2^1 + 0 \cdot 2^0$
0	0	0	...	0	0	1	1	= 3	$2^1 + 1 \cdot 2^0$
0	0	0	...	0	1	0	0	= 4 =	$2^2 + 0 \cdot 2^1 + 0 \cdot 2^0$
0	0	0	...	0	1	0	1	= 5	$2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$
0	0	0	...	0	1	1	0	= 6	$2^2 + 1 \cdot 2^1 + 0 \cdot 2^0$
0	0	0	...	0	1	1	1	= 7	$2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$
0	0	0	...	1	0	0	0	= 8 =	$2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0$

$$\Rightarrow \max(c) = 2^{11} - 1 = 2047$$

$$\Rightarrow \max(c-1023) = 1024 \Rightarrow \max(2^{c-1023}) \gg 1$$

$$\min(c-1023) = -1023 \Rightarrow \min(2^{c-1023}) \ll 1$$

$$\min(f) = 0\left(\frac{1}{2}\right)^1 + 0\left(\frac{1}{2}\right)^2 + \dots + 0\left(\frac{1}{2}\right)^{52} = 0.$$

$$\max(f) = 1\left(\frac{1}{2}\right)^1 + 1\left(\frac{1}{2}\right)^2 + \dots + 1\left(\frac{1}{2}\right)^{52} = 1 - \left(\frac{1}{2}\right)^{52}$$

Example:

$$\frac{1}{10} = \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^8 + \dots$$

ie. $0.1 \approx 0.00011001 \dots$

\uparrow
base
10

\uparrow
base
2
