

Galen Turoci

gturoci@ucsc.edu

Lab 7: Floating Point Calculations

Section 01A: Michael Powell

Due 03/16/2018 (Extended to 03/17/2018)

Lab 7

$$\begin{matrix} & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & = 128 - 127 = 1 \\ 128 & 64 & 32 & 16 & 8 & 4 & 2 & 1 & & \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & & \\ 123 & = & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \end{matrix}$$

IEEE 754 SP	0x40c80000	0xb4cccc	0x41233333 *	0xb4cccc *
Sign Bit (or I)	0	1	0	1
Sign (+/-)	+1	-1	+1	-1
Exponent bits	10000001	01111000	10000010	01111011
Real Exponent Factor	2^2	2^{-7}	2^3	2^{-4}
Significant Bits	10010000	10011001	010001100110	0000
Significant	1.5625	1.6	1.2749	1.0

0x40c80000

0100 0000 1000 0000

0000 0000 0000 0000

$$.5625 \times 2 = 1.125$$

$$.125 \times 2 = 0.25$$

$$.25 \times 2 = 0.5$$

$$.5 \times 2 = 1.0$$

0xb4cccc

0111 1000 0100 1100

1100 1100 1100 1100

$$\begin{matrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{1}{32} & \frac{1}{64} & \frac{1}{128} & \frac{1}{256} & \frac{1}{512} & \frac{1}{1024} & \frac{1}{2048} & \frac{1}{4096} & \frac{1}{8192} & \frac{1}{16384} \end{matrix}$$

$$\frac{1}{2} + \frac{1}{16} + \frac{1}{32} + \frac{1}{256} + \frac{1}{512} + \frac{1}{4096} =$$

$$.5 + .0625 + .03125 + .00390625 + .0009765625 + .000244140625 =$$

$$.5999999999 \approx 1.6$$

0100 0001 0000 0011

0011 0011 0011 0011

4 1 2 3

3 3 3 3

0100 0110 0110

$\frac{1}{4}$ $\frac{1}{64}$ $\frac{1}{128}$ $\frac{1}{1024}$ $\frac{1}{2048}$

$$.25 + .015625 + .0078125 + .0001171875 + .000029296875 =$$

$$.2749$$

$$x - 127 = -4 \quad x = 123$$

$$+127 \quad +127$$

0111 1101 1000 0000

0000 0000 0000 0000

b d 8 c

0 0 0 0

Lab 7 part 2

1 a) 4 0 8 6 6 6 6 6 10000001 = 129 - 127 = 2

0100 0000 0000 0110 0110 0110 0110 0110

sign: 0 exponent: 10000001 significand: 0000 1100 1100 1100 1100 1100

$$\frac{1}{2^{32}} + \frac{1}{2^{31}} + \frac{1}{2^{30}} + \frac{1}{2^{29}} + \frac{1}{2^{28}} + \frac{1}{2^{27}} =$$

$$\Rightarrow .0313 + .0156 + .0078 + .0039 + .0019 + .0009 = .0713$$

$$+ .0156$$

$$+ .0078$$

$$+ .0039$$

$$+ .0019$$

$$+ .0009 = .0999$$

$$1.09 \times 2^2 = 1.09 \times 4 = 4.36 \times 10^0$$

b) 6 6 f f 0 0 0 0 11001101 = 125 + 64 + 8 + 4 + 1 = 208

0110 0110 1111 0000 0000 0000 0000 0000

$$\frac{-127}{79}$$

sign: 0 exponent: 11001101 significand: 1111 1100 0000 0000 0000 0000

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} + \frac{1}{128} = .5 + .25 + .125 + .0625 + .0313 + .0156 + .0078 = .75 + .125 = .875 +$$

$$.0625 = .9375 + .0313 = .9688 + .0156 = .9844 +$$

$$.0078 = .9922 \times 2^{78} = 6.0211 \times 10^{23}$$

$$6.0211 \times 10^{23}$$

2 a) -1600,6666 1024 512 256 128 64 32 16 8 4 2 1

$$S = 1$$

$$1.1 0 0 0 0 0 0 0 0 2^{10}$$

$$10 + 127 = 137$$

$$.666 \times 2 = 1.332 \quad 1100 \quad 0100 \quad 1100 \quad 1000 \quad 0001 \quad 0101 \quad 0101 \quad 0101 \quad 137 = 10001011$$

$$.332 \times 2 = 0.664 \quad c \quad 4 \quad c \quad 8 \quad 1 \quad 5 \quad 5 \quad 5$$

$$.664 \times 2 = 1.328 \quad [1 \times c \quad 4 \quad c \quad 8 \quad 1 \quad 5 \quad 5 \quad 5]$$

$$b) -1.6 \times 10^{-19}$$

$$2^{-63}$$

$$-63 + 127 = 64$$

$$.6 \times 2 = 1.2$$

$$S = 1$$

$$1010 \quad 0000 \quad 0100 \quad 1001 \quad 1001 \quad 1001 \quad 1001 \quad 1001 \quad 12 \times 2 = 0.4$$

$$c \quad 0 \quad 4 \quad c \quad 8 \quad 1 \quad 5 \quad 5 \quad 5 \quad 12 \times 2 = 0.8$$

$$[1 \times c \quad 0 \quad 4 \quad c \quad 8 \quad 1 \quad 5 \quad 5 \quad 5]$$

$$12 \times 2 = 1.6$$

$$\begin{array}{r} 3a \quad 4480000 \times 0100 \ 0100 \ 1000 \ 0000 \\ + 35000000 + 1 \ 0011 \ 1111 \ 0000 \end{array}$$

$$x: S=0 \quad e=10001001=128+8+1=137-127=10 \quad m=1,0 \times 2^{10}$$

$$y: S=0 \quad e=01111101=64+32+16+8+4+2=126-127=-1 \quad m=1,0 \times 2^{-1}$$

$$\begin{array}{r} 1,00000000 \\ + 0,00000001 \\ \hline 1,00000001 \end{array}$$

$$0,00000001 \times 2^0$$

$$1,00000001 \times 2^{10} \quad 10+127=137=10001001$$

$$0100 \ 0100 \ 1000 \ 0000 \ 0001 \ 0001 \ 0000 \ 0000$$

$$\boxed{4 \ 4 \ 8 \ 0 \ 1 \ 0 \ 0 \ 0} \rightarrow \boxed{0x44801000}$$

$$3b \quad 3c066666 \ 0011 \ 1100 \ 0000 \ 1011 \ 1011 \ 1011 \ 1011$$

$$+ 3c111111 \ 0011 \ 1100 \ 0001 \ 0001 \ 0001 \ 0001 \ 0001$$

$$x: S=0 \quad e=01111001=64+32+16+8=120-127=-7 \quad m=1,00010111 \times 2^{-7}$$

$$y: S=0 \quad e=01111001=64+32+16+8=120-127=-7 \quad m=1,00010001 \times 2^{-7}$$

$$1,00010111011011011011$$

$$+ 1,00010001000100010001$$

$$1,00010111011011011011 \times 2^{-7} \quad -7+127=120=01111000$$

$$0011 \ 1100 \ 0001 \ 1100$$

$$\boxed{3 \ c \ 1 \ c \ c \ c \ c} \rightarrow \boxed{0x3c1cccc}$$

$$3c \quad 4280000 \ 0100 \ 0011 \ 1000 \ 1000 \ 0000$$

$$+ 6f80000 \ 1100 \ 0001 \ 1111 \ 1000 \ 0000$$

$$\boxed{0x42860000}$$

$$x: S=0 \quad e=10000101=128+4+1=133-127=6 \quad m=1,100100000000 \times 2^6$$

$$y: S=1 \quad e=10000011=128+2+1=131-127=4 \quad m=1,111100000000 \times 2^4$$

$$1,10010000$$

$$010000101000 \ 0110$$

$$+ 0,01111100 \quad 6+127=133=10000101$$

$$1,10010001 \times 2^6$$

$$\begin{array}{r} 4 \ 2 \ 8 \ 0 \\ 0 \ 0 \ 0 \ 0 \end{array}$$

bed c f f f e

$\frac{-127}{D_9 = 01111100}$

[illegible]