Floating Point Adder

EE18ACMTECH11006 Bhavani Machnoori
EE18MTECH11005 Ajay Hase

March 8, 2019

Sign bit-1	Exponent-8	Mantissa-23

Fig (1): Single Precision Format

Steps for converting decimal to floating point number:

1. Convert a Decimal number to Binary number $(975.75)_10 = (1111001111.11)_2$
2. Normalize the number $1.11100111111 * 2^9$
3. From this normalized number we can fill all 32-bits of floating point number Sign bit = 0 (number is positive)
4. Exponent = Bias + 9 = 127 + 9 = $(136)_10 = (10001000)_2$
5. Fraction part will contain all the bits after decimal point.
6. $(975.75)_10$ is expressed as shown below in single precision floating point format.

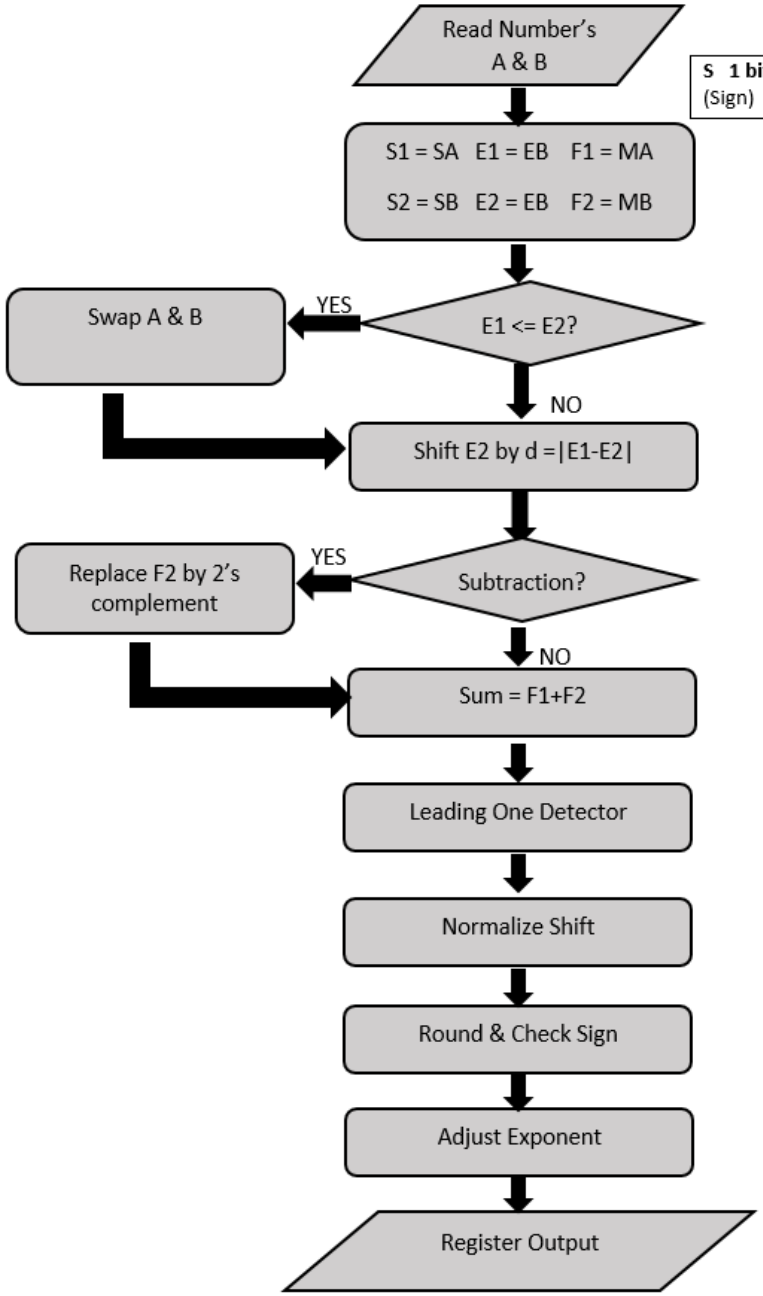
0	1000 1000	11100111111000000000000

Fig (2): Single Precision Format of $(975.75)_{10}$

Steps for floating point adder:
1.Sort:Find the largest number.
2.Align:Make the exponent equal.
3.Add/Sub:Perform addition or subtraction
4:Normalize: set MSB of mantissa

Examples:

		sort	align	add/sub	normalize
eg.1	+0.54E3	-0.87E4	-0.87E4	-0.87E4	-0.87E4
	-0.87E4	+0.54E3	+0.05E4	+0.05E4	+0.05E4
				-0.82E4	-0.82E4
eg.2	+0.54E3	-0.55E3	-0.55E3	-0.55E3	-0.55E3
	-0.55E3	+0.54E3	+0.54E3	+0.54E3	+0.54E3
				-0.01E3	-0.10E2
eg.3	+0.54E0	-0.55E0	-0.55E0	-0.55E0	-0.55E0
	-0.55E0	+0.54E0	+0.54E0	+0.54E0	+0.54E0
				-0.01E0	-0.00E0
eg.4	+0.56E3	+0.56E3	+0.56E3	+0.56E3	+0.56E3
	+0.52E3	+0.52E3	+0.52E3	+0.52E3	+0.52E3
				+1.07E3	+0.10E4



S 1 bit (Sign)	E 8 bits (Exponents)	F 23 bits (Fraction / Mantissa)

Flow chart :

Applications:

Floating point numbers are used in various applications such as medical imaging, radar, telecommunications:

1. CPU
2. Calculators