Floating Point Numbers (FPN)

Wednesday, 16 November 2022 4:17 PM

.... 32 16 8 4 2 1 . .5 .25 .125 .0625 .03125 .015625

1 1 0 0 . 1 1

1 0 1 . 0 1

$(12.75)_{10} = (0110011000\ \ 000100)_2$

=> 1100.11

=> 0.110011 × 2^4

=> 0110011 0100

Mantissa Exponent

10b M 6b Exp

0110011000 000100 (FPN) Ans ?.

-9.25

+9.25

1001.01

0.100101 × 2^4

0100101 0100 10b M

+ve 0100101000 000100 6b E

-ve 1011011000 000100

+ve

0101010000 000011

0.10101 × 2^3

101.01

+5.25

-ve

Fraction	Int.
1011010000	000100
0100110000	000100

0.10011 × 2^4

1001.1

(-9.5) ✓.

2^0 2^{-1} 2^{-2} 2^{-3} 2^{-4} | -2^2 2^1 2^0

-1 .5 .25 .125 .0625 | -4 2 1

+3

-4

MANTISSA | EXPONENT

MANTISSA	EXPONENT
- Fraction	- Int.
- 2's	- 2's
- fill left to Right	- fill Right to left.
- Accuracy	- Size/Range.
- drop from right (underflow)	- drop from left (overflow)

-ve

1111111011 01001

0000000101 01001

0.00000101 × 2^9

1010.00 -10.0

11-4=7

0000010101 001011

Non-Normalised

0101010100 00111

M E

...32 16 8 4 2 1 . .5 .25 .125 .0625 .03125 .015625

1 1 1 . 0 1

$(7.25)_{10}$

=> 111.01

=> 0.11101 × 2^3

=> 011101 011 8bits M

=> 01110100 00011 5bits E

✓ -11.0625

11.0625

1011.0001

0.10110001 × 2^4 10bits M

010110001 0100 6bits E

+ve 0101100010 000100

-ve 1010011110 000100 Ans ?.

+ve

0100111000 000100

0.100111 × 2^4

1001.11

+9.75

-ve

1011001000 000100

0100111000 000100

0.100111 × 2^4

1001.11

-9.75

0000010100 001001 9-4=5

0101000000 000101

0.101 × 2^5

10100 Normalised

+20.0

0000010100 001001

0.00001010 × 2^9

10100

+20.0

1111101010 001000

1010100000 000100

01011 0100

0.1011 × 2^4

1011.0

$(-11.0)_{10}$

	Mantissa	Exponent
L+	L+ 011→	L+ 011→
S+	S+ 010→ *	S- 1000→
L-	L- 1011 *	S- 1000
S-	S- 1000	L+ 0111

* Roughly Normalised

1/∞

FP | Mantissa | Exponent |

Accuracy < > | Size/Range > <

Smallest Magnitude +ve Num.

S-					L-	S+						L+
-32	-16	-8	-4	-2	-1	0	+1	+2	+4	+8	+16	+32

Largest Magnitude -ve Num ✗

Smallest Magnitude -ve Num. ✗

0 ↓ Fictions Doesn't exist.

Largest Magnitude +ve Num.

4bits Mantissa

4bits Exponent.

L+:

0111 0111

0.111 × 2^7

1110000

+112.0

S+:

0100 1000

0.1 × 2^{-8}

0.000000001

2^{-9}

L-:

1011 1000

0101 1000

0.101 × 2^{-8}

0.00000000101

$-(2^{-9}+2^{-11})$

S-:

1000 0111

-1×2^7

-2^7

-128