Embedded Systems with ARM Cortex-M Microcontrollers in Assembly Language and C

Chapter 12 & 13
Fixed-point and Floating-point Arithmetic Exercises ANS

Z. Gu

Fall 2025

Acknowledgement: Lecture slides based on Embedded Systems with ARM Cortex-M Microcontrollers in Assembly Language and C, University of Maine https://web.eece.maine.edu/~zhu/book/

Encoding 5.25 into IEEE Std 754 Single-Precision

- Normalization:
- $2^2 < 5.25 < 2^3, \frac{5.25}{2^2} = 1.3125$
- Hence $5.25 = 1.3125 \times 2^2 = (1 + 0.3125) \times 2^2$
- Conversion:
- $Sign = 0$
- $Exponent = 2 + 127 = 129 = 1000001_2$
- $Fraction = 0.0101_2$ (multiply by 2 repeatedly)
- Assume $Fraction = b1 \times 2^{-1} + b2 \times 2^{-2} + b3 \times 2^{-3} + b4 \times 2^{-4} + \cdots$
  - $0.3125 \times 2 = 0.625 = 0 + 0.625$ => b1 = 0
  - $0.625 \times 2 = 1.25 = 1 + 0.25$ => b2 = 1
  - $0.25 \times 2 = 0.5 = 0 + 0.5$ => b3 = 0
  - $0.5 \times 2 = 1$ => b4 = 1
- 14.5 = 01000000101010000000000000000000 in binary or 0x41680000 in hex

Decoding 01000000101010000000000000000000

- Binary 01000000101010000000000000000000
- Sign = 0
- Exponent = $10000001_2 = 129$
- Fraction = 2^-2 + 2^-4 $= 0.3125$
- $f = (-1)^S \times (1 + Fraction) \times 2^{Exponent-127}$

$$= (-1)^0 \times (1 + 0.3125) \times 2^{129-127}$$
$$= 1.3125 \times 2^2$$
$$= 5.25$$

Decoding 0x42F6E979 into a floating-point number

- Binary 01000010111101101110100101111001
- Sign = 0
- Exponent = $10000101_2 = 133$
- Fraction = $0.92712497711181 64$
- $f = (-1)^S \times (1 + Fraction) \times 2^{Exponent-127}$

$$= (-1)^0 \times (1 + 0.92712497711181 64) \times 2^{133-127}$$
$$= -1 \times 1.92712497711181 64 \times 2^6$$
$$= 123.45600128173828 \quad \text{(Last step not required)}$$

- (I will not give overly complete numbers for exams)

Decoding 0x88888000 into a floating-point number

- Binary 10001000100010001000000000000000
- Sign = 1
- Exponent = $00010001_2 = 17$
- Fraction = 2^-3 + 2^-7 $= 0.1328125$
- $f = (-1)^S \times (1 + Fraction) \times 2^{Exponent-127}$

$$= -1 \times (1 + 0.06640625) \times 2^{17-127}$$

$$= -1.06640625 \times 2^{-110}$$