Contents

[1. Objective 2](#_Toc94192908)

[2. Purpose 2](#_Toc94192909)

[3. Summary 2](#_Toc94192910)

[4. Introduction 2](#_Toc94192911)

[4.1. Base 10 2](#_Toc94192912)

[4.2. Base 2 (Binary) 2](#_Toc94192913)

[4.3. Base 8 (Octal) 2](#_Toc94192914)

[4.4. Base 16 (Hexadecimal) 2](#_Toc94192915)

[5. Convert Base N to Base 10 (Multiply method) 3](#_Toc94192916)

[5.1. Example 1 - convert base 2 to 10 3](#_Toc94192917)

[5.2. Example 2 - convert base 3 to 10 3](#_Toc94192918)

[5.3. Example 3 - convert base 16 to 10 3](#_Toc94192919)

[6. Convert Base 10 to Base N (Division Method) 4](#_Toc94192920)

[6.1. Example 1 – convert base 10 to base 2 4](#_Toc94192921)

[6.2. Example 2 – convert base 10 to base 16 4](#_Toc94192922)

[7. Tips 4](#_Toc94192923)

[7.1. HEX to BIN (“4-bit partition method”) 4](#_Toc94192924)

[7.1.1. Convert 0xAD3EF23750F 4](#_Toc94192925)

[7.2. BIN to HEX (“4-bit partition method”) 4](#_Toc94192926)

[7.2.1. Convert 1010\_1101\_0011\_1110\_1111\_0010\_0011\_0111\_0101\_0000\_1111 4](#_Toc94192927)

[7.3. OCT to BIN (“3-bit partition method”) 4](#_Toc94192928)

[7.3.1. Convert 075643742153 4](#_Toc94192929)

[7.3.2. Convert 111\_101\_110\_100\_011\_111\_100\_010\_001\_101\_011 4](#_Toc94192930)

[8. Bit Conversion - Questions 5](#_Toc94192931)

[9. Bit Conversion - Answers 6](#_Toc94192932)

# Objective

* Learn how floating numbers are stored in most modern computers, IEEE754 format

# Introduction

## Step summary

* + 1. Convert Base 10 number to Base 2
    2. Convert the whole number part
    3. Convert the decimal part
    4. Shift the floating point to an exponent format
    5. Now consist of 4 parts, sign bit, exponent part, whole part, and the mantissa part
    6. Only the exponent part needs to convert
    7. The rest is stored normally as binary

# History

# Fixed floating point

# Number TO Floating point - IEEE 754 Single precision 32-bit

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

* 1-bit (sign bit)
* 8-bit (exponent)
* 23-bit (mantissa)

Example: Convert -66.66

### Store sign bit

-66.65 is a negative, therefore the sign bit is 1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

### Convert the whole number part

### Convert the decimal part

Check:

When to stop?

For 32-bit floating point, we just need to repeat the pattern and fill up the whole 23-bit

### Combine back the Base 2

### Convert to Exponent format

### Convert exponent

### Convert exponent to base 2

### Slot in the base 2 exponent

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

### Slot in the mantissa

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

# Floating point TO Number - IEEE 754 Single precision 32-bit

The reverse order of the convertion