

## O Padrão IEEE754

1. O padrão IEEE754, o qual foi definido pelo Instituto de Engenheiros Eletricistas e Eletrônicos (IEE), foi adotado para padronizar regras a serem seguidas em relação às operações e representações de números binários com ponto flutuante.

Antes de ser adotado, cada empresa utilizava um padrão diferente, ou seja, alguns números eram menos precisos do que outros. A universalização do padrão IEEE754 permitiu a otimização da aritmética de ponto flutuante e melhorias na FPU (Float Point Unit).

2. Esse padrão trata dos seguintes tópicos: formatos para aritmética, formatos para troca de dados, regras para arredondamento, operações aritméticas e tratamento de exceções.

A normalização ocorre de acordo com a seguinte fórmula:

$$S M \cdot 2^E$$

S = Sinal. Sempre representado por um bit (negativo (1) ou positivo(0)).

M = Mantissa (parte fracionária).

2 = base binária

E = Expoente.

Além disso, padroniza o arredondamento (em direção ao zero, em direção ao infinito positivo, em direção ao infinito negativo ou para o mais próximo) e a parte inteira (que sempre será 1, ou seja, ficará implícita e liberará um bit).

Exemplos:

Conversão do número 7.0 para o padrão IEEE754 =

O sinal (+) será representado por 0, o expoente será 2 (pois  $2^2 < 7.0 < 2^3$ ) e a parte fracionária será 1.75 (pois  $7.0 / 4 = 1.75$ ). 7.0 será representado por 01000000111000000000000000000000.

Conversão do número -12.0 para o padrão IEEE754 =

O sinal (-) será representado por 1, o expoente será 3 (pois  $2^3 < 12.0 < 2^4$ ) e a parte fracionária será 1.5 (pois  $12.0 / 8 = 1.5$ ). -12.0 será representado por 11000001010000000000000000000000.

3. A aritmética do padrão IEEE754 é realizada da seguinte forma:

Primeiro, o sinal é analisado. Se for negativo, será colocado 1; se for positivo, será colocado 0. Depois, o número é convertido para binário e sua vírgula é deslocada até a primeira casa. O binário é então multiplicado por 2 elevado à potência mais próxima (e menor). Essa potência será normalizada (somada ao número 127) e convertida também para binário, assim como o 127. Ambos serão somados e teremos o resultado, sendo os espaços restantes preenchidos com zero.

### Bibliografia:

Tópico 1

[https://pt.wikipedia.org/wiki/V%C3%ADrgula\\_flutuante](https://pt.wikipedia.org/wiki/V%C3%ADrgula_flutuante)

[https://pt.wikipedia.org/wiki/Unidade\\_de\\_ponto\\_flutuante](https://pt.wikipedia.org/wiki/Unidade_de_ponto_flutuante)

Tópicos 1 e 2

<https://ufsj.edu.br/portal2-repositorio/File/nepomuceno/ca/03a-compieeee.pdf>

Tópico 2

<https://www.h-schmidt.net/FloatConverter/IEEE754.html>

Tópicos 1, 2 e 3

[https://pt.wikipedia.org/wiki/IEEE\\_754](https://pt.wikipedia.org/wiki/IEEE_754)

**Gabriela Satie Faria Nishimi, aluna de Ciências de Computação na USP de São Carlos**  
**Número USP = 11892820**