

Sistemas de Computação

Ponto Flutuante

Universidade Federal de Ouro Preto - UFOP

Haroldo Gambini Santos

Sistemas de Computação

[illegible]

Seção

[illegible][illegible]

100111100010001011000000111111010111110001111010100

0.0000 0000 00000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0100

Haroldo Gambini Santos

Sistemas de Computação

2/1

[illegible]

Ponto Flutuante

Aplicações

- Engenharia
- Matemática
- Física
- Processamento de Imagens
- Computação Gráfica 3D

Haroldo Gambini Santos

Sistemas de Computação

3 / 1

Notas

Ponto Flutuante

FLOP

Floating Point Operations per Second
Medida importante para computação de alto desempenho.
Supercomputadores atuais: escala de petaflops (10^{15}).

Haroldo Gambini Santos

Sistemas de Computação

4 / 1

Notas

Notação Científica

Útil para separar a faixa da precisão de um número.

$$n = \mathbf{f} \times 10^{\mathbf{e}}$$

f **fração** ou **mantissa**;
e **expoente**: um inteiro positivo ou negativo.

Haroldo Gambini Santos

Sistemas de Computação

5 / 1

Notas

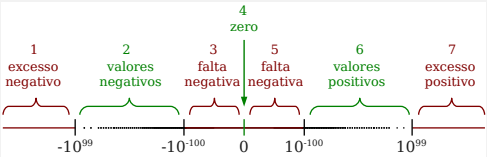
Notação Científica - Exemplo

Representação Decimal com Nr. de Dígitos Limitado

- Fração
 - $0,1 \leq |f| < 1$
 - 3 dígitos
 - 0,1 , ... , 0,999
 - + ou -
- Expoente
 - 2 dígitos
 - 0,...,99
 - + ou -

Notas

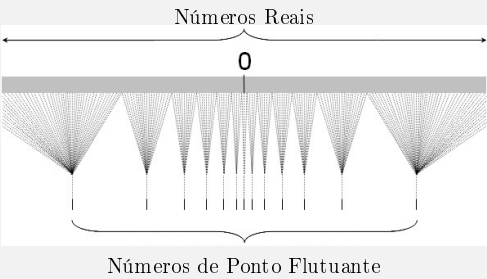
Notação Científica - Exemplo



- 1 números negativos grandes menores do que $-0,999 \times 10^{99}$
- 2 números negativos entre $-0,999 \times 10^{99}$ e $-0,100 \times 10^{-99}$
- 3 números negativos pequenos com grandezas menores do que $0,1 \times 10^{-99}$
- 4 0
- 5 números positivos pequenos com grandezas menores do que $0,1 \times 10^{-99}$
- 6 números positivos entre $0,1 \times 10^{-99}$ e $0,999 \times 10^{99}$
- 7 números positivos grandes maiores do que $0,999 \times 10^{99}$

Notas

De Reais para Números de Ponto Flutuante



Notas

Notas

Notas

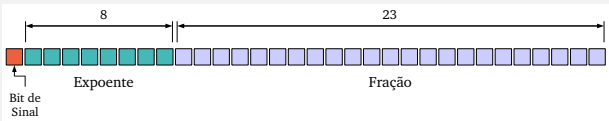
IEEE 754

- Padrão amplamente usado em praticamente todas as CPUs modernas
- Especifica a representação de ponto flutuante no formato:

$$mantissa \times 2^{expoente}$$

Notas

Precisão Simples: 32 bits



O Padrão IEEE

Precisões

	Expoente	Bits		Declaração em C
		Fração	Total ¹	
Simples	8	23	32	<code>float</code>
Dupla	11	52	64	<code>double</code>

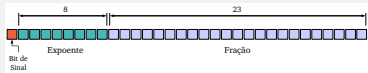
¹inclui sinal

Notas

Seção

Notas

Expoente

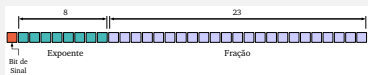


O Campo Expoente

- representado na notação *por excesso* (*biased*)
 - valor fixo é adicionado ao expoente
 - em precisão simples, excesso = 127
 - expoente = valor do campo expoente - 127

Notas

Notas



Notas

O Campo Fração

- Codifica a parte fracionária de um número com parte inteira = 1
- Utiliza a representação de sinal e magnitude

Mantissa no padrão IEEE 754:

$1.fração$

Notas

Exemplo

Qual é o campo de fração da representação de 6,25 em ponto flutuante de representação simples ?

Resolução

- para a parte inteira: $110 = 2^2 + 2^1 = 6$
- para a parte fracionária, o que equivale a 0,25 ?

Fração

- De frações decimais para frações binárias:
- Vamos preencher os dígitos binários da fração da direita (-1) para a esquerda $(-n)$:
 - $\dots, d_{-1} d_{-2} \dots d_{-n}$
 - Preencha os dígitos do seguinte modo, considerando a fração decimal **fração**
 - 1 **fração** = **fração** $\times$ 2
 - 2 se **fração** ≥ 1 , então **dígito** = 1 e diminua 1 de **fração**
 - 3 senão, **dígito** = 0
 - 4 se **fração** > 0 volte para 1

Notas

Fração

De frações decimais para frações binárias:

Convertendo 0,40625:

Fração		Dígito
0,40625	$\times 2 = 0,8125$	0
0,8125	$\times 2 = 1,625$	1
0,625	$\times 2 = 1,25$	1
0,25	$\times 2 = 0,5$	0
0,5	$\times 2 = 1,0$	1

Notas

Fração

Representação

Então, como 6,25 é armazenado no campo de fração ?

$6,25 = 0b110,01$
 $0b110,01 = 0b1,1001 \times 2^2$

No padrão para precisão simples, considera-se uma parte inteira fixa de 1. Desse modo, a parte fracionária armazenada é:

10010000000000000000

Notas

Interpretação dos Números

Campo Expoente	Campo Fração	Representa
0	0	0
0	não 0	$\pm (0, \text{fração}) \times 2^{1-\text{excesso}}$
0s e 1s	qualquer	$\pm (1, \text{fração}) \times 2^{\text{expoente}-\text{excesso}}$
todos 1	0	$\pm$ infinito
todos 1	não zero	NaN

Notas

Notas

Notas
