

Sistema de Numeración en Punto Flotante

Objetivos de la práctica: que el alumno domine los tópicos de sistemas de numeración referidos a las representaciones en punto flotante, tales como:

- Representación e interpretación.
- Operaciones aritméticas.
- IEEE 754.

Bibliografía:

- "Organización y Arquitectura de Computadores" de W. Stalling, capítulo 8.
- Apunte de la Cátedra, "Sistemas de numeración: Punto flotante".

1. Considerando el sistema de Punto Flotante cuya mantisa es fraccionaria, con 6 bits , está expresada en BSS (en el inciso a) o BCS (en el inciso b) y su exponente en BCS con 4 bits, escriba el significado de las siguientes cadenas de bits (mantisa a la izquierda):

Cadena	a) Mantisa en BSS	b) Mantisa en BCS
0101110110		
0000010000		
0000111001		
1111111111		
0000000000		
0000001111		
1111110000		
1000000000		
0000011111		

a) Mantisa fraccionaria, 6 bits, BSS, Exponente BCS 4 bits:

0101110110

Los 6 bits de más a la izquierda componen la mantisa, al ser fraccionaria, su forma será 0,010111, usar teorema fundamental de la numeración:
 $0,010111 = (2^{-2} + 2^{-4} + 2^{-5} + 2^{-6})$

El exponente es en BCS 0110 \Rightarrow 6 en decimal; entonces:

$$0,010111 \times 2^{0110} = (2^{-2} + 2^{-4} + 2^{-5} + 2^{-6}) \times 2^6$$

Producto de potencias de igual base, se suman los exponentes

$$= 2^4 + 2^2 + 2^1 + 2^0 = 16 + 4 + 2 + 1 = 23$$

0000010000

Mantisa:

$$0,000001 = (2^{-6})$$

El exponente es en BCS 0000 \Rightarrow 0 en decimal; entonces:

$$0,000001 \times 2^{0000} = (2^{-6}) \times 2^0 = 2^{-6}$$

0000111001

Mantisa:

$$0,000011 = (2^{-5} + 2^{-6})$$

El exponente es en BCS 1001 \Rightarrow -1 en decimal; entonces:

$$0,000011 \times 2^{1001} = (2^{-5} + 2^{-6}) \times 2^{-1} = 2^{-6} + 2^{-7} = 1/2^6 + 1/2^7 = 1/64 + 1/128 = (2+1)/128 = 3/128$$

Organización de Computadoras 2020

1111111111

Mantisa:

$$0,111111 = (2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5} + 2^{-6}) = 1 - 2^{-6} = (64-1)/64 = 63/64$$

El exponente es en BCS 1111 \Rightarrow -7 en decimal; entonces:

$$0,111111 \times 2^{1111} = (2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5} + 2^{-6}) \times 2^{-7} = 2^{-8} + 2^{-9} + 2^{-10} + 2^{-11} + 2^{-12} + 2^{-13}$$

0000000000

$$000000 \times 2^{0000} = 0 \times 2^0 = 0$$

b) Mantisa fraccionaria, 6 bits, BCS, Exponente BCS 4 bits:

0101110110

Los 6 bits de más a la izquierda componen la mantisa, al ser fraccionaria, el bit de más a la izquierda es el bit de signo de la mantisa lo separo, entonces su forma será 0 0,10111, usar teorema fundamental de la numeración:

$$0 \ 0,10111 = + (2^{-1} + 2^{-3} + 2^{-4} + 2^{-5})$$

El exponente es en BCS 0110 \Rightarrow 6 en decimal; entonces:

$$0 \ 10111 \ 0110 = 0 \ 0,10111 \times 2^{0110} = + (2^{-1} + 2^{-3} + 2^{-4} + 2^{-5}) \times 2^6 = + (2^5 + 2^3 + 2^2 + 2^1) = 32 + 8 + 4 + 2 = 46$$

0000010000

Los 6 bits de más a la izquierda componen la mantisa, al ser fraccionaria, el bit de más a la izquierda es el bit de signo de la mantisa lo separo, entonces su forma será 0 0,00001, usar teorema fundamental de la numeración:

$$0 \ 0,00001 = + (2^{-5})$$

El exponente es en BCS 0000 \Rightarrow 0 en decimal; entonces:

$$0 \ 00001 \ 0000 \Rightarrow 0 \ 0,00001 \times 2^{0000} = + (2^{-5}) \times 2^0 = + 2^{-5}$$

0000111001

El exponente es en BCS 1001 \Rightarrow -1 en decimal; entonces:

$$0 \ 00011 \ 1001 \Rightarrow 0 \ 0,00011 \times 2^{1001} = + (2^{-4} + 2^{-5}) \times 2^{-1} = + (2^{-5} + 2^{-6})$$

1111111111

Mantisa:

$$1 \ 0,11111 = - (2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5})$$

El exponente es en BCS 1111 \Rightarrow -7 en decimal; entonces:

$$0,111111 \times 2^{1111} = - (2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5}) \times 2^{-7} = - (2^{-8} + 2^{-9} + 2^{-10} + 2^{-11} + 2^{-12})$$

Organización de Computadoras 2020

0000000000

$$0\ 0,00000 \times 2^{0000} = 0 \times 2^0 = 0$$

Enlaces asociados :

[Introducción a Punto Flotante](#) [parte 2](#) [parte 3](#) [parte 4](#)

[Ejemplo](#) [Ejemplo \(parte 2\)](#) [Ejemplo \(parte 3\)](#)

2. Dado un sistema de Punto Flotante cuya mantisa es fraccionaria , está expresada en BCS con 5 bits y su exponente en BSS con 3 bits, interprete las siguientes cadenas del considerando que la mantisa está sin normalizar, normalizada, o normalizada con bit implícito Identifique aquellas cadenas que no pueden ser interpretadas y mencione por que.

Cadena	Sin normalizar	Normalizada	Normalizada con Bit Implícito
01000111			
11000011			
00000000			
11111111			

a) Mantisa fraccionaria, 5 bits, BCS, Exponente BSS 3 bits:

01000111

. Mantisa sin Normalizar:

Mantisa: 01000 $\Rightarrow 0\ 0,1000 = + (2^{-1})$

Exponente (BSS): 111 = 7 en decimal;

Entonces

$$01000\ 111 \Rightarrow 0\ 0,1000 \times 2^{111} = + (2^{-1}) \times 2^7 = + (2^6) = 64$$

. Mantisa Normalizada: ‘1’ después de la coma.

Mantisa: 01000 $\Rightarrow 0\ 0,1000 = + (2^{-1})$

Exponente (BSS): 111 = 7 en decimal;

Entonces

$$01000\ 111 \Rightarrow 0\ 0,1000 \times 2^{111} = + (2^{-1}) \times 2^7 = + (2^6) = 64$$

Mantisa Normalizada con bit implícito:

El ‘1’ después de la coma no se “ve” en la representación, pero se debe utilizar para el cálculo, lo agrego:

$$\text{Mantisa: } 01000 \Rightarrow 0\ 0,11000 = + (2^{-1} + 2^{-2})$$

Exponente (BSS): 111 = 7 en decimal;

Entonces

$$01000\ 111 \Rightarrow 0\ 0,11000 \times 2^{111} = + (2^{-1} + 2^{-2}) \times 2^7 = + (2^6 + 2^5) = 64 + 32 = 96$$

11000011

. Mantisa sin Normalizar:

Mantisa: 11000 $\Rightarrow 1\ 0,1000 = - (2^{-1})$

Organización de Computadoras 2020

Exponente (BSS): 011 = 3 en decimal;

Entonces

$$11000\ 011 \Rightarrow 1\ 0,1000 \times 2^{011} = -(2^{-1}) \times 2^3 = -(2^2) = -4$$

. Mantisa Normalizada:

Mantisa: 11000 $\Rightarrow 1\ 0,1000 = -(2^{-1})$

Exponente (BSS): 011 = 3 en decimal;

Entonces

$$11000\ 011 \Rightarrow 1\ 0,1000 \times 2^{011} = -(2^{-1}) \times 2^3 = -(2^2) = -4$$

Mantisa Normalizada con bit implícito:

El '1' después de la coma no se "ve" en la representación, pero se debe utilizar para el cálculo, lo agrego:

$$\text{Mantisa: } 11000 \Rightarrow 1\ 0,11000 = -(2^{-1} + 2^{-2})$$

Exponente (BSS): 011 = 3 en decimal;

Entonces

$$11000\ 011 \Rightarrow 1\ 0,11000 \times 2^{011} = -(2^{-1} + 2^{-2}) \times 2^3 = -(2^2 + 2^1) = -(4 + 2) = -6$$

00000000

. Mantisa sin Normalizar:

Mantisa: 00000 $\Rightarrow 0\ 0,1000 = +0$

Exponente (BSS): 000 = 0 en decimal;

Entonces

$$00000\ 000 \Rightarrow 0\ 0,0000 \times 2^{000} = +0$$

. Mantisa Normalizada:

No se puede calcular, pues la mantisa no empieza con 0,1...

Mantisa Normalizada con bit implícito:

El '1' después de la coma no se "ve" en la representación, pero se debe utilizar para el cálculo, lo agrego:

$$\text{Mantisa: } 00000 \Rightarrow 0\ 0,10000 = (2^{-1})$$

Exponente (BSS): 000 = 03 en decimal;

Entonces

$$00000\ 000 \Rightarrow 0\ 0,10000 \times 2^{000} = +(2^{-1}) \times 2^0 = +(2^{-1}) \times 1 = +0,5$$

11111111

. Mantisa sin Normalizar:

$$\text{Mantisa: } 11111 \Rightarrow 1\ 0,1111 = -(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4})$$

Exponente (BSS): 111 = 7 en decimal;

Organización de Computadoras 2020

Entonces

$$11111\ 111 \Rightarrow 1\ 0,1111 \times 2^{111} = -(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4}) \times 2^7 = -(2^6 + 2^5 + 2^4 + 2^3) = -(64 + 32 + 16 + 8) = -120$$

. Mantisa Normalizada:

$$\text{Mantisa: } 11111 \Rightarrow 1\ 0,1111 = -(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4})$$

Exponente (BSS): 111 = 7 en decimal;

Entonces

$$11111\ 111 \Rightarrow 1\ 0,1111 \times 2^{111} = -(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4}) \times 2^7 = -(2^6 + 2^5 + 2^4 + 2^3) = -(64 + 32 + 16 + 8) = -120$$

Mantisa Normalizada con bit implícito:

El '1' después de la coma no se "ve" en la representación, pero se debe utilizar para el cálculo, lo agrego:

$$\text{Mantisa: } 11111 \Rightarrow 1\ 0,11111 = -(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5})$$

Exponente (BSS): 111 = 7 en decimal;

Entonces

$$11111\ 111 \Rightarrow 1\ 0,11111 \times 2^{111} = -(2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5}) \times 2^7 = -(2^6 + 2^5 + 2^4 + 2^3 + 2^2) = -(64 + 32 + 16 + 8 + 4) = -124$$

Enlaces :

[Mantisa normalizada](#) [Normalización](#) [Normalización \(parte 2\)](#)

[Normalizada con bit implícito](#) [Normalizada con bit implícito BCS](#)

3. Calcule rango y resolución en extremos inferior negativo, superior negativo, inferior positivo y superior positivo para los siguientes sistemas de representación en punto flotante:
- Mantisa fraccionaria en BSS de 8 bits y exponente en BSS 4 bits
 - Mantisa fraccionaria normalizada en BSS de 15 bits y exponente en CA1 10 bits
 - Mantisa fraccionaria normalizada con bit implícito en BCS de 15 bits y exponente en Exceso 5 bits
 - Mantisa fraccionaria normalizada con bit implícito en BCS de N bits y exponente en CA2 de M bits

Observe que:

- En las mantisas BSS no se puede expresar números negativos, con lo que aun con exponente negativo expresaremos un número positivo por un factor de escala menor a 1, pero también positivo. Ejemplo: $2 \times 2^{-4} = 0,125$.
- Las mantisas fraccionarias suponen el punto al principio de la mantisa.
- Los exponentes negativos indican factores de escala menores a 1 que mejoran la resolución.
- Mantisa normalizada implica que empieza con 1, o sea mantisa mínima 0,1 para la fraccionaria, igual a 0,5 en decimal. Esto hace que no se pueda representar el 0.
- Mantisa normalizada con bit implícito, significa agregar un 1 al principio de la misma al interpretarla. Ejemplo: 00000 se interpreta 0,100000, o 0,5 en base 10.

a. Mantisa fraccionaria en BSS de 8 bits y exponente en BSS 4 bits

$N_1 = 0$ (el número más chico en este sistema, minimizando la mantisa y exponente)

$N_2 = 0,11111111 \times 2^{1111} = (1 - 2^{-8}) \times 2^{15} = (2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5} + 2^{-6} + 2^{-7} + 2^{-8}) \times 2^{15}$ (el más grande)

Porqué $0,11111111 = (1 - 2^{-8})$?

Pues $(1 - 2^{-8}) = 1,00000000 - 0,00000001 = 0,11111111$

Entonces el rango es: $= [N_1; N_2] = [0; (1 - 2^{-8}) \times 2^{15}]$

La resolución (recordar, diferencia entre 2 números consecutivos), ya no es fija en todo el intervalo, debemos calcularla en cada sector del intervalo de representación:

Organización de Computadoras 2020

R_1 = Resolución extremo inferior: debemos obtener el siguiente número representable en este sistema que sigue al cero, lo llamaremos N_3 . Para ello, debemos minimizar mantisa al número posterior al cero y minimizar el exponente.

$$N_3 = 0,00000001 \times 2^{0000} = 2^{-8} \text{ (el más cercano a } N_1)$$

Entonces para obtener la resolución en el extremo inferior positivo debo restarle a N_3 el valor de N_1

$$R_1 = \text{Resolución extremo inferior} = N_3 - N_1 = (0,00000001 - 0) \times 2^{0000} = 2^{-8}$$

R_2 = Resolución extremo superior: debemos obtener el número representable en este sistema anterior al más grande, lo llamaremos N_4 . Para ello, debemos maximizar mantisa al número anterior a la mantisa máxima y maximizar el exponente.

$$N_4 = 0,11111110 \times 2^{1111} \text{ (el más cercano a } N_2)$$

Entonces para obtener la resolución en el extremo superior positivo debo restarle a N_2 el valor de N_4

$$R_2 = \text{Resolución extremo superior} = N_2 - N_4 = (0,11111111 - 0,11111110) \times 2^{15} = 2^{-8} \cdot 2^{15} = 2^7$$

b. Mantisa fraccionaria normalizada en BSS de 15 bits y exponente en Ca1 10 bits

Como la mantisa es BSS, los números representables son todos positivos. Para obtener el mínimo, minimizo mantisa (normalizada) y exponente, el cual debe ser el máximo negativo en ca1.

$$N_1 = 0,1000000000000000 \times 2^{1000000000} = 2^{-1} \times 2^{-511} = 0,5 \times 2^{-511}$$

Se observa que el cero no es representable en este sistema, por estar la mantisa normalizada.

Para obtener el máximo, maximizo la mantisa (en BSS) y el exponente (en Ca1),

$$N_2 = 0,1111111111111111 \times 2^{0111111111} = (1 - 2^{-15}) \times 2^{+511}$$

$$\text{Entonces el rango es: } = [N_1, N_2] = [0,5 \times 2^{-511}; (1 - 2^{-15}) \times 2^{+511}]$$

Para la resolución en el extremo inferior, debo obtener el número siguiente al mínimo, lo llamamos N_3 ; para ello tomo la siguiente mantisa y utilizo el mismo exponente que para representar ese número

$$N_3 = 0,1000000000000001 \times 2^{1000000000} = (2^{-1} + 2^{-15}) \times 2^{-511}$$

$$R_1 = \text{Resolución extremo inferior} = N_3 - N_1 = (0,1000000000000001 - 0,1000000000000000) \times 2^{1000000000}$$

$$= 0,0000000000000001 \times 2^{1000000000} = 2^{-15} \cdot 2^{-511}$$

Para la resolución extremo superior debemos obtener el número representable en este sistema anterior al más grande, lo llamaremos N_4 . Para ello, debemos maximizar mantisa al número anterior a la mantisa máxima y maximizar el exponente:

$$N_4 = 0,111111111111110 \times 2^{0111111111}$$

$$R_2 = \text{Resolución extremo superior} = N_2 - N_4 = (0,1111111111111111 - 0,111111111111110) \times 2^{0111111111}$$

$$= 0,0000000000000001 \times 2^{0111111111} = 2^{-15} \times 2^{+511}$$

c. Mantisa fraccionaria normalizada con bit implícito en BCS de 15 bits y exponente en Exceso 5 bits

En este sistema tenemos valores positivos y negativos.

Para obtener el mínimo positivo (N_1), colocamos la mantisa mínima positiva (primer bit en cero por ser BCS), pero asumiendo un dígito en 1 de más después de la coma por estar normalizada con bit implícito:

Organización de Computadoras 2020

Exponente en exceso = $10000 + 10000 = 00000$ (-16 en exceso)

$$N_1 = 0,1000000000000000 \times 2^{00000} = + (2^{-1}) \times 2^{-16} = + 0,5 \times 2^{-16} \text{ (el más chico +)}$$

Para obtener el máximo positivo (N2), maximizo mantisa (recordando bit implícito) y exponente

$$N_2 = 0,1111111111111111 \times 2^{11111} = + (1 - 2^{-15}) \times 2^{+15} \text{ (el más grande +)}$$

Para la resolución en el extremo inferior obtengo el número positivo más cercano a N1

$$N_3 = 0,1000000000000001 \times 2^{00000} = + (2^{-1} + 2^{-15}) \times 2^{-16} \text{ (el más cercano a } N_1 \text{)}$$

$$R_1 = N_3 - N_1 = (2^{-1} + 2^{-15} - 2^{-1}) \times 2^{-16} = (2^{-15}) \times 2^{-16} = 2^{-31}$$

Para la resolución en el extremo superior obtengo el número positivo más cercano a N2

$$N_4 = 0,111111111111110 \times 2^{11111} \text{ (el más cercano a } N_2 \text{)}$$

$$R_2 = N_2 - N_4 = (0,111111111111111 - 0,111111111111110) \times 2^{+15} = 2^{-15} \times 2^{+15} = 2^0 = 1$$

Como tenemos números negativos debemos calcular los valores negativos mínimo y máximo:

$$N_5 = 1,1000000000000000 \times 2^{00000} = - (2^{-1}) \times 2^{-16} = - 0,5 \times 2^{-16} \quad (\text{mínimo negativo en módulo})$$

$$N_6 = 1,1111111111111111 \times 2^{11111} = - (1 - 2^{-15}) \times 2^{+15} \quad (\text{máximo negativo en módulo})$$

En este caso, los valores que se obtienen para las resoluciones en los extremos mínimo y máximo negativo son los mismos que los positivos, en módulo.

El Rango es $[N_6, N_5] \cup [N_1, N_2]$ U: Unión

$$\text{Rango : } [- (1 - 2^{-15}) \times 2^{+15}; - 0,5 \times 2^{-16}] \cup [+ 0,5 \times 2^{-16}; + (1 - 2^{-15}) \times 2^{+15}]$$

Enlaces :

[Rango y Resolución](#) [Parte 2](#) [Parte 3](#) [Parte 4](#) [Anexo](#)

[Rango y Resolución \(normalizada\)](#) [Rango y Resolución \(con bit implícito\)](#)

4. Dado un sistema de Punto Flotante cuya mantisa es fraccionaria , está expresada en BCS con 10 bits y su exponente en CA2 con 5 bits, obtenga la representación de los siguientes números, considerando que la mantisa está sin normalizar, normalizada, o normalizada con bit implícito

Cadena	Sin normalizar	Normalizada	Normalizada con Bit Implícito
0			
1			
9			
-5,0625			
34000,5			
0,015625			
N° máximo			
N° mínimo			

0

Sin normalizar

$$0 = 0,000000000 \times 2^E,$$

Organización de Computadoras 2020

(Exponente cualquiera)

Normalizada

0 = **No se puede representar**. El cero no está en el rango.

Normalizada con bit implícito

0 = **No se puede representar**. El cero no está en el rango.

+1

Sin normalizar

Tenemos distintas representaciones del valor 1

Por cada decremento la mantisa (es decir por cada corrimiento de la coma a la izquierda de la cadena) debo incrementar en 1 el exponente:

$$1 = 0,100000000 \times 2^{00001} = 0,010000000 \times 2^{00010} = 0,001000000 \times 2^{00011}$$

Represento el signo (que lo separo de los bits de la mantisa), luego los 5 bits del exponente y finalmente los 9 bits restantes de la mantisa:

0 00011 100000000

Normalizada

$$1 = 0,100000000 \times 2^{00001}$$

0 00001 100000000

Normalizada con bit implícito

$$1 = 0,100000000 \times 2^{00001}$$

Para el cálculo de la mantisa tengo 10 bits (por el bit implícito, que al representar la cadena no muestro), más el bit de signo.

0 00001 000000000

+9

Sin normalizar

Varias representaciones:

$$9 = 0,100100000 \times 2^{00100} = + (2^{-1} + 2^{-4}) \times 2^4 = + (0,5 + 0,0625) \times 16 = 9$$

$$9 = 0,010010000 \times 2^{00101} = + (2^{-2} + 2^{-5}) \times 2^5 = + (0,25 + 0,03125) \times 32 = 9$$

$$9 = 0,001001000 \times 2^{00110} = + (2^{-3} + 2^{-6}) \times 2^6 = + (0,125 + 0,015625) \times 64 = 9, \text{ represento este último:}$$

0 00110 001001000

Normalizada

$$9 = 0,100100000 \times 2^{00100} = + (2^{-1} + 2^{-4}) \times 2^4 = + (0,5 + 0,0625) \times 16 = 9$$

0 00100 100100000

Normalizada con bit implícito (recordar que el bit implícito se usa para el cálculo pero no para la representación de la cadena)

$$9 = 0,100100000 \times 2^{00100} = + (2^{-1} + 2^{-4}) \times 2^4 = + (0,5 + 0,0625) \times 16 = 9$$

0 00100 001000000

-5,0625

Sin normalizar

$$-5,0625 = 1,101,000100 \times 2^0 = 1,0101000100 \times 2^{00011} = - (2^{-1} + 2^{-3} + 2^{-7}) \times 2^3 = -0,6328125 \times 8 = -5,0625$$

Organización de Computadoras 2020

$$-5,0625 = 1\ 101,000100 \times 2^0 = 1\ 0,010100010 \times 2^{00100} = -(2^{-2} + 2^{-4} + 2^{-8}) \times 2^4 = -0,31640625 \times 16 = -5,0625$$

$$-5,0625 = 1\ 101,000100 \times 2^0 = 1\ 0,001010001 \times 2^{00101} = -(2^{-3} + 2^{-5} + 2^{-9}) \times 2^5 = -0,158203125 \times 32 = -5,0625 \text{ represento este último:}$$

1 00101 001010001

Normalizada

$$-5,0625 = 1\ 101,000100 \times 2^0 = 1\ 0,101000100 \times 2^{00011} = -(2^{-1} + 2^{-3} + 2^{-7}) \times 2^3 = -0,6328125 \times 8 = -5,0625$$

1 00011 101000100

Normalizada con bit implícito

$$-5,0625 = 1\ 0,1010001000 \times 2^{00011} = -(2^{-1} + 2^{-3} + 2^{-7}) \times 2^3 = -0,6328125 \times 8 = -5,0625$$

1 00011 010001000

34000,5

Sin normalizar

$$\text{El más grande} = 0\ 0,111111111 \times 2^{01111} = +(1 - 2^{-9}) \times 2^{15} = +(0,998046875) \times 32768 = +32704$$

Por lo tanto 34000,5 está fuera de rango.

Normalizada

$$\text{El más grande} = 0\ 0,111111111 \times 2^{01111} = +(1 - 2^{-9}) \times 2^{15} = +(0,998046875) \times 32768 = +32704$$

Por lo tanto 34000,5 está fuera de rango.

Normalizada con bit implícito

$$\text{El más grande} = 0\ 0,111111111 \times 2^{01111} = +(1 - 2^{-10}) \times 2^{15} = +(0,9990234375) \times 32768 = +32736$$

Por lo tanto 34000,5 está fuera de rango.

0,015625

Aplicando el cálculo para pasar de decimal a binario:

$$0,015625 \times 2 = 0,03125$$

$$0,03125 \times 2 = 0,0625$$

$$0,0625 \times 2 = 0,125$$

$$0,125 \times 2 = 0,25$$

$$0,25 \times 2 = 0,5$$

$$0,5 \times 2 = 1,0$$

$$0,0 \times 2 = 0,0 \quad \text{Me quedo con los valores enteros de los resultados } 0000010$$

Sin normalizar

$$+0,015625 = 0\ 0,000001000 \times 2^0 = +(2^{-6}) \times 2^0 = +0,015625 \times 1 = +0,015625$$

$$+0,015625 = 0\ 0,000010000 \times 2^{-1} = +(2^{-5}) \times 2^{-1} = +0,03125 \times 0,5 = +0,015625$$

$$+0,015625 = 0\ 0,000100000 \times 2^{-2} = +(2^{-4}) \times 2^{-2} = +0,0625 \times 0,25 = +0,015625$$

0 11110 000100000

Normalizada

$$+0,015625 = 0\ 0,100000000 \times 2^{-5} = 0\ 0,100000000 \times 2^{11011} = +0,5 \times 0,03125 = +0,015625$$

0 11011 100000000

Organización de Computadoras 2020

Normalizada con bit implícito

$$+ 0,015625 = 0\,0,1000000000 \times 2^{-5} = 0\,0,1000000000 \times 2^{11011} = + 0,5 \times 0,03125 = + 0,015625$$

0 11011 000000000

Nº máximo

Sin normalizar

$$\text{El más grande} = 0\,0,111111111 \times 2^{01111} = + (1 - 2^{-9}) \times 2^{15} = + (0,998046875) \times 32768 = + 32704$$

0 01111 111111111

Normalizada

$$\text{El más grande} = 0\,0,111111111 \times 2^{01111} = + (1 - 2^{-9}) \times 2^{15} = + (0,998046875) \times 32768 = + 32704$$

0 01111 111111111

Normalizada con bit implícito

$$\text{El más grande} = 0\,0,111111111 \times 2^{01111} = + (1 - 2^{-10}) \times 2^{15} = + (0,9990234375) \times 32768 = + 32736$$

0 01111 111111111

Nº mínimo

Sin normalizar

$$\text{El más chico ("negativo más grande")} = 1\,0,111111111 \times 2^{01111} = - (1 - 2^{-9}) \times 2^{15} = - (0,998046875) \times 32768 = - 32704$$

1 01111 111111111

Normalizada

$$\text{El más chico ("negativo más grande")} = 1\,0,111111111 \times 2^{01111} = - (1 - 2^{-9}) \times 2^{15} = - (0,998046875) \times 32768 = - 32704$$

1 01111 111111111

Normalizada con bit implícito

$$\text{El más chico ("negativo más grande")} = 1\,0,111111111 \times 2^{01111} = + (1 - 2^{-10}) \times 2^{15} = + (0,9990234375) \times 32768 = + 32736$$

Enlaces :

[Decimal a IEEE 754](#) [Parte 2](#) [Parte 3](#)

5. Diga cómo influyen las siguientes variantes en el rango y resolución:

- Mantisa con signo y sin signo.
- Exponente con signo y sin signo.
- Tamaño de mantisa.
- Tamaño de exponente.
- Mantisa fraccionaria, fraccionaria normalizada y fraccionaria normalizada con bit implícito.

Organización de Computadoras 2020

Enlaces :

[Mantisa normalizada](#) [Rango y Resolución \(normalizada\)](#)

[Normalizada con bit implícito](#) [Rango y Resolución \(con bit implícito\)](#) [Normalizada con bit implícito BCS](#)

6. Efectúe las siguientes sumas para un sistema de punto flotante con mantisa BSS de 8 bits y exponente en BCS 8 bits.

$$00001111 \ 00000011 + 00001000 \ 00000010 = ?$$

$$01111111 \ 00000000 + 11111100 \ 10000001 = ?$$

$$00000001 \ 00000111 + 00011100 \ 00000000 = ?$$

Observe que los factores de escala deben ser los mismos, sino sumáramos dos mantisas con pesos distintos (recordar que se puede correr los unos y sumar o restar este corrimiento al exponente para obtener una cadena equivalente).

$$00001111 \ 00000011 + 00001000 \ 00000010 = 15 \times 2^3 + 8 \times 2^2 = 120 + 32 = 152$$

Para hacer la suma, debemos igualar los exponentes y correr la coma de la mantisa hacia la derecha o izquierda según el caso:

Si incremento exponente => decremento mantisa, es decir corro coma a la izquierda

Si decremento exponente => incremento mantisa, es decir corro coma a la derecha

Tener en cuenta que al hacer el corrimiento de mantisa no se pierda la representación exacta del número dado.

En este caso, vamos a incrementar el exponente de la segunda cadena:

$$00001000 \ 00000010 = 8 \times 2^2 = 4 \times 2^3 \quad 00000100 \ 00000011$$

$$\begin{array}{r} 00001111 \times 2^3 \\ \pm \quad 00000100 \times 2^3 \\ \hline 00010011 \times 2^3 = 19 \times 2^3 = 19 \times 8 = 152 \end{array}$$

$$01111111 \ 00000000 + 11111100 \ 10000001 = 127 \times 2^0 + 252 \times 2^{-1} = 127 + 126 = 253$$

Incrementamos exponente 2da cadena, decrementamos mantisa:

$$\begin{array}{r} 111111 \\ 01111111 \times 2^0 \\ \pm \quad 01111110 \times 2^0 \\ \hline 11111101 \times 2^0 = 253 \times 2^0 = 253 \end{array}$$

$$00000001 \ 00000111 + 00011100 \ 00000000 = 1 \times 2^7 + 28 \times 2^0 = 128 + 28 = 156$$

Decrementamos exponente 1ra cadena en 7, incrementamos mantisa con 7 corrimientos:

$$\begin{array}{r} 10000000 \times 2^0 \\ \pm \quad 00011100 \times 2^0 \\ \hline 10011100 \times 2^0 = 156 \times 2^0 = 156 \end{array}$$

7. Suponiendo que los números que no son representables se aproximan al más próximo, obtenga las representaciones o aproximaciones de los números 8,625; 0,4 y 2,5 en los sistemas:

- Mantisa fraccionaria normalizada de 5 bits BSS exponente 4 bits CA2
- Mantisa fraccionaria normalizada de 10 bits BCS exponente 3 bits CA2

8,625

- Mantisa fraccionaria normalizada de 5 bits BSS exponente 4 bits CA2

$8,625 = 1000,101 \times 2^0 = 0,1000101 \times 2^4$ Pero la mantisa tiene sólo 5 bits, no se puede representar completa; calculamos el número más aproximado:

Organización de Computadoras 2020

$$0,10001 \times 2^4 = (2^{-1} + 2^{-5}) \times 2^4 = \mathbf{8,5} \quad 0100 \ 10001$$

$$\text{El N}^\circ \text{ que le sigue} = (0,10001 + 0,00001) \times 2^4 = 0,10010 \times 2^4 = (2^{-1} + 2^{-4}) \times 2^4 = \mathbf{9}$$

$$\text{Error1} = 8,625 - 8,5 = 0,125$$

$$\text{Error2} = 9 - 8,5 = 0,5 \quad \Rightarrow \text{Menor error (valor más cercano) es } 8,5$$

8,625 no tiene una representación exacta en este sistema. 8,5 está más cerca que 9.

b. Mantisa fraccionaria normalizada de 10 bits BCS exponente 3 bits CA2

$$8,625 = 0 \ 1000,101 \times 2^0 = 0 \ 0,100010100 \times 2^4$$

El exponente 4 no se puede expresar con 3 bits en CA2

$$\text{El número más grande} = 0 \ 0,111111111 \times 2^{+3} = + (1 - 2^{-9}) \times 8 = (0,998046875) \times 8 = 7,984375$$

Éste sería el más cercano a 8,625.

0 011 11111111

0,4

a. Mantisa fraccionaria normalizada de 5 bits BSS exponente 4 bits CA2

$$0,4 \times 2 = 0,8$$

$$0,8 \times 2 = 1,6$$

$$0,6 \times 2 = 1,2$$

$$0,2 \times 2 = 0,4$$

$$0,4 \times 2 = 0,8$$

$$0,8 \times 2 = 1,6$$

$0,4 = 0,01100 \times 2^0$ corriendo la coma a derecha (es decir incrementando mantisa) entra un dígito más (decremento exponente)

$$0,11001 \times 2^{-1} = 0,11001 \times 2^{1111} = 0,390625$$

$$\text{El que sigue} = 0,11010 \times 2^{-1} = (0,5 + 0,25 + 0,0625) \times 0,5 = 0,40625$$

$$\text{Error1} = 0,4 - 0,390625 = 0,009375$$

$$\text{Error2} = 0,40625 - 0,4 = 0,00625 \quad \Rightarrow \text{Esta representación es más cercana a } 0,4.$$

b. Mantisa fraccionaria normalizada de 10 bits BCS exponente 3 bits CA2

$$0,4 \times 2 = 0,8$$

$$0,8 \times 2 = 1,6$$

$$0,6 \times 2 = 1,2$$

$$0,2 \times 2 = 0,4$$

$$0,4 \times 2 = 0,8$$

$$0,8 \times 2 = 1,6$$

$$0,6 \times 2 = 1,2$$

$$0,2 \times 2 = 0,4$$

$$0,4 \times 2 = 0,8$$

$$0,4 = 0 \ 0,011001100 \times 2^0$$

$$\text{Normalizada} = 0 \ 0,110011001 \times 2^{-1} = (2^{-1} + 2^{-2} + 2^{-5} + 2^{-6} + 2^{-9}) \times 0,5 = 0,399414062$$

$$\text{El que sigue} = 0 \ 0,110011010 \times 2^{-1} = (2^{-1} + 2^{-2} + 2^{-5} + 2^{-6} + 2^{-8}) \times 0,5 = 0,400390625$$

$$\text{Error1} = 0,4 - 0,399414062 = 0,000585938$$

$$\text{Error2} = 0,400390625 - 0,4 = 0,000390625 \quad \Rightarrow 0,400390625 \text{ está más cerca de } 0,4 \text{ (menor error)}$$

2,5

a. Mantisa fraccionaria normalizada de 5 bits BSS exponente 4 bits CA2

$$2,5 = 10,1 \times 2^0 = 0,101 \times 2^2 = 0,10100 \times 2^{010} = (2^{-1} + 2^{-3}) \times 2^2 = (0,5 + 0,125) \times 4 = 2,5$$

0010 10100

b. Mantisa fraccionaria normalizada de 10 bits BCS exponente 3 bits CA2

$$2,5 = 0,101000000 \times 2^{010} = (2^{-1} + 2^{-3}) \times 2^2 = (0,5 + 0,125) \times 4 = 2,5$$

0 010 101000000

8. Definimos Error Absoluto y Error Relativo de un número x en un sistema de la siguiente forma:

$EA(x) = |x' - x|$ y $ER(x) = EA(x) / x$; donde x' es el número representable del sistema más próximo a x .

Calcule los errores absolutos y relativos para los casos del ejercicio anterior.

8,625

a) Mantisa fraccionaria normalizada de 5 bits BSS exponente 4 bits CA2

Como vimos anteriormente, los valores más aproximados representables en este sistema eran 8,5 y 9.

$$8,5 < 8,625 < (2^{-1} + 2^{-5}) \times 2^4 = 9 \quad 8,5 \text{ es el más próximo;}$$

$$E_A = 8,625 - 8,5 = 0,125 \quad (\text{Menor error})$$

$$E_R = E_A / N^\circ \text{ a representar} = 0,125 / 8,625 \sim 0,0145$$

b) Mantisa fraccionaria normalizada de 10 bits BCS exponente 3 bits CA2

En este caso se obtuvo que el número más cercano representable es el 7,984375

$$E_A = 8,625 - 7,984375 = 0,640625$$

$$E_R = E_A / N^\circ \text{ a representar} = 0,640625 / 8,625 \sim 0,0742753623$$

0,4

a) $0,390625 < 0,4 < 0,40625$

$$E_A = 0,40625 - 0,4 = 0,00625 \quad (\text{menor error})$$

$$E_R = 0,00625 / 0,4 = 0,015625$$

b) $0,399414062 < 0,4 < 0,400390625$

$$E_A = 0,400390625 - 0,4 = 0,000390625 \quad \text{Menor error}$$

$$E_R = 0,000390625 / 0,4$$

2,5

a) $E_A = 0$ Representación exacta.

b) $E_A = 0$ Representación exacta.

Enlaces :

Error absoluto y error absoluto máximo **Error relativo**

9. Considerando que en los procesos de truncamiento o redondeo la elección se basa en la representación más cercana, estime el Error Absoluto Máximo cometido en las representaciones del ejercicio 1. Recuerde que la distancia entre 2 representaciones sucesivas se conoce como resolución (R), por lo que $EA_{\max} \leq R / 2$.

Enlaces :

Error absoluto y error absoluto máximo

10. Tome un sistema de punto flotante cualquiera y dibuje la forma del gráfico de cada tipo de error en función del número que se quiere representar.

11. Detalle las características del estándar IEEE 754 para simple precisión y doble precisión

El IEEE 754 es un estándar de aritmética en coma flotante. Este estándar especifica como deben representarse los números en coma flotante con simple precisión (32 bits) o doble precisión (64 bits), y también cómo deben realizarse las operaciones aritméticas con ellos.

Emplea mantisa fraccionaria, normalizada y en representación signo magnitud (M y S), sin almacenar el primer dígito, que es igual a 1. El exponente se representa en exceso, que en este caso no se toma como $2n-1$, sino como $2n-1 - 1$

Simple Precisión

El estándar IEEE-754 para la representación en simple precisión de números en coma flotante exige una cadena de 32 bits. El primer bit es el bit de signo (S), los siguientes 8 son los bits del exponente (E) y los restantes 23 son la mantisa (M):

La mantisa es fraccionaria normalizada, con la coma después del primer bit que es siempre uno (1.), en M y S

El exponente se representa en exceso.

Doble precisión

El estándar IEEE-754 para la representación en doble precisión de números en coma flotante exige una cadena de 64 bits. El primer bit es el bit de signo (S), los siguientes 11 son los bits del exponente (E) y los restantes 52 son la mantisa (M)

Enlaces :

IEEE 754	Interpretación	Interpretación (nº normalizado)	Interpretación (nº denormalizado)
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50
51	51	51	51
52	52	52	52
53	53	53	53
54	54	54	54
55	55	55	55
56	56	56	56
57	57	57	57
58	58	58	58
59	59	59	59
60	60	60	60
61	61	61	61
62	62	62	62
63	63	63	63
64	64	64	64
65	65	65	65
66	66	66	66
67	67	67	67
68	68	68	68
69	69	69	69
70	70	70	70
71	71	71	71
72	72	72	72
73	73	73	73
74	74	74	74
75	75	75	75
76	76	76	76
77	77	77	77
78	78	78	78
79	79	79	79
80	80	80	80
81	81	81	81
82	82	82	82
83	83	83	83
84	84	84	84
85	85	85	85
86	86	86	86
87	87	87	87
88	88	88	88
89	89	89	89
90	90	90	90
91	91	91	91
92	92	92	92
93	93	93	93
94	94	94	94
95	95	95	95
96	96	96	96
97	97	97	97
98	98	98	98
99	99	99	99
100	100	100	100
101	101	101	101
102	102	102	102
103	103	103	103

12. ¿Qué valores están representados por las siguientes cadenas si responden al estándar IEEE 754?

[illegible]

0 11000100 000000000000000000000000

Exponente: en 8 bits, en exceso $128 - 1 = 127$

Al valor le restamos el exceso:

Organización de Computadoras 2020

$$11000100 \Rightarrow (128 + 64 + 4) - 127 = 196 - 127 = 69$$

Mantisa: recordar que se antepone 1, y valor, en este caso 1,0

Entonces la cadena representa el valor + **$1,0 \cdot 2^{+69}$**

1 11111110 1010000000000000000000

Exponente

$$11111110 \Rightarrow 254 - 127 = 127$$

Mantisa: $1 + 0,5 + 0,125 = 1,625$

Signo negativo,

Entonces la cadena representa el valor **$-1,625 \cdot 2^{+127}$**

$$0 \text{ 00000000 000000000000000000000001} = +(2^{-23}) \times 2^{-126}$$

$$1 \text{ 00000000 000000000000000000000000} = -0$$

$$1 \text{ 11111111 000000000000000000000000} = +\infty$$

$$1 \text{ 11111111 000001000000000000000000} = \text{NaN}$$

Enlaces :

[IEEE 754](#) [Interpretación](#) [Interpretación \(n° normalizado\)](#) [Interpretación \(n° denormalizado\)](#)

13. Hallar la representación en simple precisión del estándar IEEE 754 de los siguientes números

1; 13; 257; -40000; 0,0625

Escribo la mantisa en binario y luego desplazo la coma hasta el primer 1 de la izquierda (achico mantisa y elevo exponente)

$$-40000 = 1 \text{ 1001110001000000} \times 2^0 = 1 \text{ 1,001110001000000000000000} \times 2^{15} =$$

$$\text{Exponente} = 15 + 127 = 142 \Rightarrow \text{exponente } 10001110$$

$$= 1 \text{ 1,001110001000000000000000} \times 2^{10001110}$$

Entonces la cadena es **1 10001110 001110001000000000000000**

$$0,0625 = 0 \text{ 0,0001000000.....00} \times 2^0 = 0 \text{ 1,00000.....000} \times 2^4 = 0 \text{ 1,000000000000000000000000} \times 2^{10000011}$$

$$\text{Exponente} = 4 + 127 = 131 = 10000011 = (+4 \text{ en exceso } 127)$$

Entonces la cadena es **0 10000011 000000000000000000000000**

$$+1 = 0 \text{ 1,00000.....000} \times 2^0 = 0 \text{ 1,00000....000} \times 2^{01111111}$$

$$\text{Exponente} = 0 + 127 = 127 = 01111111 = (0 \text{ en exceso } 127)$$

Entonces la cadena es **0 01111111 000000000000000000000000**

$$+13 = 0 \text{ 1101,00000...000} \times 2^0 = 0 \text{ 1,101000....000} \times 2^{+3} = 0 \text{ 1,101000....000} \times 2^{10000010}$$

Organización de Computadoras 2020

Exponente = $3 + 127 = 130 = 10000010 = (+3 \text{ en exceso } 127)$

Entonces la cadena es **0 10000010 101000000000000000000000**

Enlaces :

[Decimal a IEEE 754](#) [Parte 2](#) [Parte 3](#)

14. Calcule rango y resolución en extremos inferior negativo y superior positivo para los sistemas de simple precisión y doble precisión del estándar IEEE 754. ¿Cuál es el menor número positivo distinto de '0' que se puede representar?

Enlace:

[Rango](#)

15. Efectúe las siguientes sumas (las cadenas son representaciones en el estándar IEEE 754)

$00001111\ 010000000000000000000000 + 00010000\ 010000000000000000000000 = ?$
 $11111111\ 101010101010101010101010 + 11111100\ 100000011111000001101010 = ?$

16. En el estándar IEEE 754, ¿para qué sirve, cuando el exponente es 0 y la mantisa no es nula, que la mantisa no esté normalizada?

Enlace:

[Interpretación \(nº denormalizado\)](#)