

# 710193M Arquitectura de computadores II

## Aritmética del computador: Representación punto flotante

Facultad de Ingeniería. Universidad del Valle

Marzo de 2022

- 1 Representación en punto flotante
- 2 Estándar del IEEE para punto flotante
- 3 Aritmética en punto flotante

# Contenido

- 1 Representación en punto flotante
- 2 Estándar del IEEE para punto flotante
- 3 Aritmética en punto flotante

# Representación en punto flotante

## Definiciones

- En la representación complemento a dos es posible representar enteros positivos o negativos
- Sin embargo, no es posible representar números fraccionarios
- Se requiere un gran número de bits para representar números grandes

# Representación en punto flotante

## Definiciones

- En el sistema decimal las anteriores limitaciones se superan usando notación científica:
  - 1  $10000000000 = 1 \times 10^{10}$
  - 2  $0,00000254 = 2,54 \times 10^{-6}$
- Esta representación se utiliza para correr la coma decimal, tantas potencias de 10 se le indique. Si se corre la izquierda la potencia es positiva y si se corre a la derecha es potencia negativa.

# Representación en complemento a dos

## Ejercicio en clase

Transforme las siguientes expresiones en notación científica:

■ 15200000000.

$1,52 \times 10^{10}$

■ 0,0000012445

$1,2445 \times 10^{-6}$

■ 0,155457878454

$1,55457878454 \times 10^{-1}$

$1,554578 \times 10^{-1}$

# Representación en complemento a dos

## Ejercicio en clase

Respuesta:

- $1,52 \times 10^{10}$
- $1,2445 \times 10^{-6}$
- $0,155457878454 \times 10^0$

## Limitaciones

Observe que no aplica para todos los casos, como es el caso de  $0,155457878454$ , no se puede acortar la representación utilizando notación científica de lo contrario se podría perder información.

# Representación en punto flotante

## Definiciones

La técnica de notación científica se puede aplicar a los números binarios.

## Notación científica binarios

$$\pm S \times B^{\pm E} \quad (1)$$

- 1  $\pm$  Signo
- 2  $S$  Mantisa: Parte significativa
- 3  $E$  Exponente
- 4  $B$  Base, en binario es 2



# Representación en punto flotante

## Definiciones

- 1 Si el bit MSB (más significativo) es 0, el número es positivo y si es 1, el número es negativo
- 2 El exponente consta de 8 bits.
- 3 Se utiliza **representación sesgada**


# Representación en punto flotante

## Representación sesgada

- 1 Es un valor que se le resta al exponente
- 2 Es denominado como **sesgo**
- 3 Tiene un valor de  $2^{k-1} - 1$   $k$  es el número de bits del exponente
- 4 Por lo que en un campo de 8 bits, este comprende entre un valor en el rango -127 a +128

$$\begin{array}{r} 00000000 \\ 22111111 \\ \hline 120411168928 \end{array} = 255$$

Sesgo  $2^7 - 1 = 127$



# Representación en punto flotante

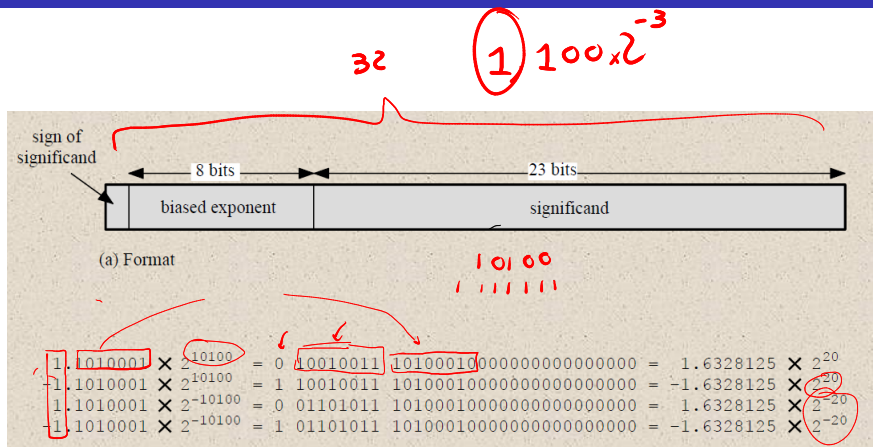


Figura 1: Formato típico de 32 bits en coma flotante

# Representación en punto flotante

## Observaciones

- 1 Las siguientes representaciones son equivalentes:

$$0,110 * 2^5$$

$$110 * 2^2$$

$$0,0110 * 2^6$$

Para simplificarlos cálculos se utiliza la siguiente representación:

$$\pm 1,bbbb * 2^{\pm E} - \epsilon$$

- 2 Como se puede observar siempre tiene un 1 el el bit más a la izquierda
- 3 Este bit se puede quitar y cuando se realizan los cálculos se vuelve a colocar

# Representación en punto flotante

$$\textcircled{1}, 666 \times 2^{E-3}$$

## Observaciones

Por lo que este formato:

- 1 El signo se almacena en el bit más a la izquierda
- 2 El primer bit de la parte significativa siempre es 1, por lo que puede quitarse
- 3 Se suma 127 al exponente original
- 4 La base es 2

# Complemento a dos vs representación en punto flotante

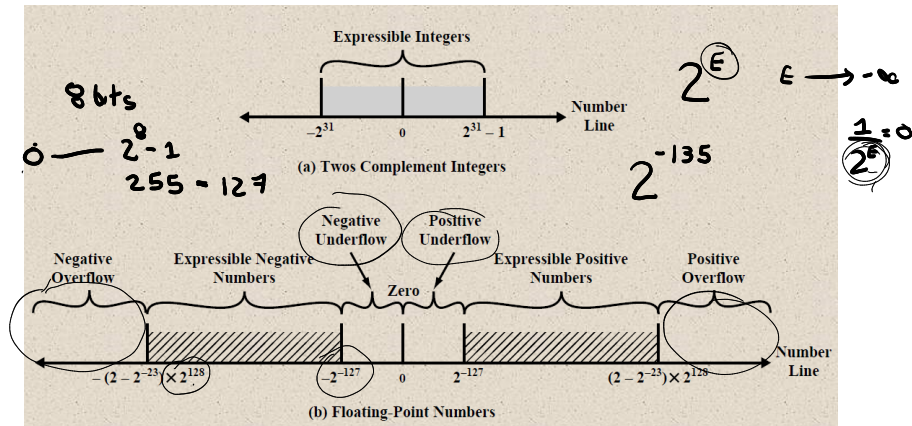


Figura 2: Comparación representación a dos vs punto flotante

# Representación en punto flotante

## Rango

Para el formato de representación en punto flotante se tienen los siguientes rangos:

- 1 Números negativos entre  $-(2 - 2^{23}) * 2^{128}$  y  $-2^{-127}$
- 2 Números positivos entre  $-2^{-127}$  y  $(2 - 2^{23}) * 2^{128}$

# Representación en punto flotante

## Regiones

En la recta se pueden observar las siguientes regiones se encuentran excluidas:

- 1 **Desbordamiento negativo:** números negativos menores que:  
 $-(2 - 2^{23}) \times 2^{128}$
- 2 **Agotamiento negativo:** números negativos mayores que:  
 $-2^{-127}$
- 3 **Agotamiento positivo:** números positivos menores  $2^{-127}$
- 4 **Desbordamiento positivo:** números positivos mayores que  
 $(2 - 2^{23}) \times 2^{128}$
- 5 El zero





# Representación en punto flotante

## Densidad

Se puede observar que el espaciamiento entre los números en coma flotante no es uniforme:

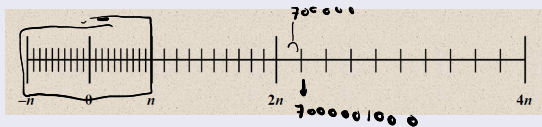


Figura 3: Densidad de los números en coma flotante

Se puede observar que existe una gran población de números cerca al origen

# Contenido

big Integer  
big Decimal

1 Representación en punto flotante

2 Estándar del IEEE para punto flotante

754

3 Aritmética en punto flotante

# Estándar del IEEE para punto flotante

## Formatos

- 1 Formato simple: 32 bits *Float*
- 2 Formato doble: 64 bits *double ←*
- 3 Formato de 128 bits. *quadruple ~ - - -*

# Estándar del IEEE para punto flotante

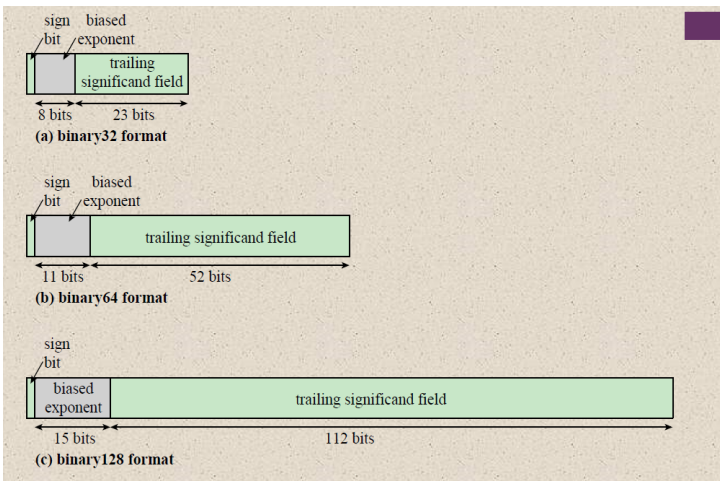


Figura 4: Formatos IEEE 754

# Estándar del IEEE para punto flotante

Parameter	Format		
	binary32	binary64	binary128
Storage width (bits)	32	64	128
Exponent width (bits)	8	11	15
Exponent bias / <i>5 es go</i>	127	1023	16383
Maximum exponent	128	1024	16385
Minimum exponent	-127	-1023	-16383
Approx normal number range (base 10)	$10_{-38}, 10_{+38}$	$10_{-308}, 10_{+308}$	$10_{-4932}, 10_{+4932}$
Trailing significand width (bits)*	23	52	112
Number of exponents	254	2046	32766
Number of fractions	$2_{23}$	$2_{52}$	$2_{112}$
Number of values	$1.98 \leftrightarrow 2_{31}$	$1.99 \leftrightarrow 2_{63}$	$1.99 \leftrightarrow 2_{128}$
Smallest positive normal number	$2_{-126}$	$2_{-1022}$	$2_{-16362}$
Largest positive normal number	$2_{128} - 2_{104}$	$2_{1024} - 2_{971}$	$2_{16384} - 2_{16271}$
Smallest subnormal magnitude	$2_{-149}$	$2_{-1074}$	$2_{-16494}$

$2^{15} = 16383$

Figura 5: Interpretación números en coma flotante

# Estándar del IEEE para punto flotante

## Ejemplos

Transforme al estándar IEEE 754 de 32 bits los siguientes números:

1  $10,5_{10}$

2  $17236_{10}$

3  $-127,125_{10}$

$$\begin{array}{c} \boxed{X} \overbrace{1'0'1'0,1} \\ \underbrace{2^2 \ 2^1 \ 2^0 \ 2^{-1}} \end{array}$$

$$3 + 127 = 130$$

$$\underline{1} \ \underline{0} \ \underline{0} \ \underline{0} \ \underline{0} \ \underline{0} \ \underline{1} \ \underline{1}$$

$$\begin{array}{ccccccc} 0 & | & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & | & 0 & 1 & 0 & 1 & \dots \\ & & S & & & & & & E & & & & & & M & \end{array}$$

$$17326 - 2^{14} = 942 - 512 = 430 - 256 = 174 - 128 = 46 - 32 = 14$$

$$\begin{array}{cccccccccccccccc} \textcircled{1} & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ \hline 14 & 13 & 12 & 11 & 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 \end{array}$$

Punto  
Five

$$1 \overline{00001110101110}$$

$$127 + 14 = 141 - 128 = 13$$

$$\underline{1} \quad \underline{0} \quad \underline{0} \quad \underline{0} \quad \underline{1} \quad \underline{1} \quad \underline{0} \quad \underline{1}$$

$$0 | 10001101 | 00001110101110 \text{ —————}$$

-127, 125

$$\boxed{1111111, 001}$$

$\frac{1}{2} \quad \frac{1}{2^2} \quad \frac{1}{2^3}$

$$6+127=133$$

$$\underline{1} \quad \underline{0} \quad \underline{0} \quad \underline{0} \quad \underline{0} \quad \underline{1} \quad \underline{0} \quad \underline{1}$$

$$1 \mid 10000101 \mid 111111001$$



# Estándar del IEEE para punto flotante

## Ejemplos

**Paso 1:** Transforme el número en binario

1  $1010,1_2$

2  $100001101010100_2$

3 (Signo negativo)  $1111111,001_2$

# Estándar del IEEE para punto flotante

## Ejemplos

**Paso 2:** Normalice el número

1  $1,0101_2 * (10^3)_2$

2  $1,00001101010100_2 * (10^{14})_2$

3 (Signo negativo)  $1,111111001_2 * (10^6)_2$

# Estándar del IEEE para punto flotante

## Ejemplos

**Paso 3:** Calcula el exponente, recuerda que para el caso de 32 bits, el sesgo es  $2^8 - 1$  entonces es  $127_{10} = 1111111_2$

1  $127_{10} + 3_{10} = 130_{10} = 10000010_2$

2  $127_{10} + 14_{10} = 141_{10} = 10001101_2$

3 (Signo negativo)  $127_{10} + 6_{10} = 133_{10} = 10000101_2$

# Estándar del IEEE para punto flotante

## Ejemplos

**Paso 4:** Se obtiene el número en estándar IEEE 754 de 32 bits:

Signo	Exponente	Mantisa
0	10000010	<b>0101</b> 00000000000000000000000000000000
0	10001101	<b>00001101010100</b> 000000000000000000000000
1	10000101	<b>111111001</b> 000000000000000000000000000000

Verifiquemos:

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

Consultado Feb 2016

# Estándar del IEEE para punto flotante

## Ejercicio en clase

Transforme al estándar IEEE 754 de 32 bits los siguientes números:

1  $18,75_{10} = 10010,11_2$

2  $14578,5_{10} = 11100011110010,1_2$

3  $-0,625_{10} = 0,101_2$

$0,0000,1_2$   
 $-5$

1)  $4 + 127 = 131 = 10000011$

2)  $13 + 127 = 140 = 10001100$

3)  $-1 + 127 = 126 = 01111110$

# Estándar del IEEE para punto flotante

## Ejercicio en clase

Respuestas:

- 1 01000001100101100000000000000000<sub>2</sub>
- 2 01000110011000111100101000000000<sub>2</sub>
- 3 10111111001000000000000000000000<sub>2</sub>

# Estándar del IEEE para punto flotante

## Ejercicio en clase

Ahora intenta para 64 bits, recuerda: exponente 11 bits y mantisa 53 bits.

- 1  $18,75_{10} = 10010,11_2$   $1023 + 4 = 1027$   $\frac{1}{2} \frac{0}{2} \frac{0}{2} \frac{0}{2} \frac{0}{2} \frac{0}{2} \frac{0}{2} \frac{1}{2} \frac{1}{2}$
- 2  $14578,5_{10} = 11100011110010,1_2$   $1023 + 3 =$
- 3  $-0,625_{10} = 0,101_2$   $1023 - 1 = 1022$   $1 \frac{0}{2} \frac{0}{2} \frac{0}{2} \frac{0}{2} \frac{0}{2} \frac{0}{2} \frac{0}{2} \frac{1}{2} \frac{1}{2} \frac{0}{2} \frac{0}{2}$

0111111110

¿Cuanto es el sesgo?

Recuerda es:  $2^{k-1} - 1$ , donde  $k$  es el número de bits del exponente

$$2^{10} - 1 = 1023$$

# Estándar del IEEE para punto flotante

## Ejercicio en clase

Respuestas:

**1**   01000000001100101100<sub>2</sub>

**2**    010000001100110001111001010000000000000000000000000000000000<sub>2</sub>

[illegible]



# Estándar del IEEE para punto flotante

## Ejemplos

Transforme a decimal los siguientes números en el estándar IEEE 754 de 32 bits:

- 1)  $01000001001100000000000000000000_2$   $1) 130 - 127 = 3$   $1,5 \times 10^3$   
 $1500$
- 2)  $11000000110100000000000000000000_2$   $1,011 \times 2^3$
- 3)  $01000010010010000000000000000000_2$   $1011_2 = 11_{10}$

$$\begin{aligned} 2) \quad 129 - 127 &= 2 \\ &= 1,101 \times 2^2 = 110,1 \\ &= 6.5 \end{aligned}$$

$$\begin{aligned} 3) \quad 132 - 127 &= 5 \\ &= 11001_2 \times 2^5 = 110010 \\ &32 + 16 + 2 = 50 \end{aligned}$$

# Estándar del IEEE para punto flotante

## Ejemplos

**Paso 1:** Reste el sesgo al exponente

1  $10000010_2 - 1111111_2$

- $130_{10} - 127_{10}$

- $3_{10}$

2  $10000001_2 - 1111111_2$

- $129_{10} - 127_{10}$

- $2_{10}$

3  $10000100_2 - 1111111_2$

- $132_{10} - 127_{10}$

- $5_{10}$

# Estándar del IEEE para punto flotante

## Ejemplos

**Paso 2:** Agregue 1. a la mantisa

1 1,011000000000000000000000<sub>2</sub>

2 1,101000000000000000000000<sub>2</sub>

3 1,100100000000000000000000<sub>2</sub>

# Estándar del IEEE para punto flotante

## Ejemplos

**Paso 3:** Multiplique por 2 elevado al exponente al que se le ha restado el sesgo

1  $1,01100000000000000000000_2 * 2^3$

2  $1,10100000000000000000000_2 * 2^2$

3  $1,10010000000000000000000_2 * 2^5$

En binario, multiplicar por 2 es análogo a multiplicar por 10 en decimal, **sólo debes correr la coma tantas posiciones a la derecha o izquierda**

# Estándar del IEEE para punto flotante

## Ejemplos

**Paso 4:** Convierta en decimal, tome en cuenta el bit del signo

1  $1011,00000000000000000000_2 = 11_{10}$

2  $-110,10000000000000000000_2 = -6,5_{10}$

3  $110010,000000000000000000_2 = +50_{10}$

# Estándar del IEEE para punto flotante

## Ejercicio

Transforme a decimal los siguientes números en el estándar IEEE 754 de 32 bits:

\* 1  $01000010000001010000000000000000_2$

$\leftarrow 33,25$

2  $11000001011110000000000000000000_2$

$10000100$

3  $01000010110010001000000000000000_2$

$102 - 127 = 5$

$1,0000101 \times 2^5$

$100001,01$

$33,25_{10}$

# Estándar del IEEE para punto flotante

## Ejercicio

Respuestas:

- 1  $33,25_{10}$
- 2  $-15,5_{10}$
- 3  $100,25_{10}$

# Contenido

- 1 Representación en punto flotante
- 2 Estándar del IEEE para punto flotante
- 3 Aritmética en punto flotante**



# Aritmética en punto flotante

## Definiciones

- Sumas y restas más complicadas que la multiplicación y división, imagine que tiene que sumar  $3 * 10^5 + 4 * 10^3$ .
- Por lo tanto, para sumas y restas es necesario ajustar el exponente.

# Aritmética en punto flotante

## Sumas y restas

Se deben realizar 4 pasos:

- 1 **Comprobación de cero:** Debido a que la suma y la resta son idénticas, excepto por el cambio de signo, en el caso de restas se cambia el signo del **substraendo**
- 2 **Ajuste cifras significativas:** Ajuste el menor exponente igual al mayor exponente, realizando los corrimientos de coma necesarios:  
$$123 * 10^0 + 456 * 10^{-2} = 123 * 10^0 + 4,56 * 10^0 = 127,56 * 10^0$$
- 3 **Realice la suma respectiva, tomando en cuenta el signo de los operandos**
- 4 Normalización, recuerde que la mantisa tiene forma 1.*bbbb*

# Aritmética en punto flotante

## Ejemplo 1

Realice la siguiente operación:

1  $01000010100101010000000000000000_2 +$   
 $0100000000111000000000000000000000_2$   
 $74,5_{10} + 3,75_{10}$

# Aritmética en punto flotante

## Ejemplo 1

### Paso 1: Comprobación de cero:

$$\begin{array}{l} \text{1 } 01000010100101010000000000000000_2 + \\ 01000000011100000000000000000000_2 \end{array}$$

En este caso ya que los dos son positivos y se trata de una suma, no se realiza ningún cambio de signo

# Aritmética en punto flotante

## Ejemplo 1

### Paso 2: Ajuste cifras significativas:

■ Exponentes:  $10000101_2$ ,  $10000000_2$

■ Mantisas:

$1,001010100000000000000000_2$ ,  $1,111000000000000000000000_2$

■ La diferencia entre los exponentes es:  $101_2 = 5_{10}$  por lo que se deben correr 5 posiciones a la izquierda la segunda mantisa así:

Mantisas:  $1,001010100000000000000000_2$ ,

$0,0000111100000000000000000000_2$

# Aritmética en punto flotante

## Ejemplo 1

**Paso 3: Suma:** Mantisas normalizadas:

$1,00101010000000000000000_2,$

$0,000011110000000000000000000_2$

- Se realiza la suma de las mantisas para obtener:  
 $1,00111001000000000000000_2$
- Como en este caso ha quedado de la forma 1,0 no es necesario cambiar el exponente.
- Por lo que el resultado es:

$$0,1000010100111001000000000000000_2 = 78,25_{10}.$$

# Aritmética en punto flotante

## Ejemplo 2

Realice la siguiente operación:

$$\begin{array}{r} 1 \quad 11000001110011000000000000000000_2 - \\ \quad 01000001100011100000000000000000_2 \\ -25,5_{10} - 17,75_{10} \end{array}$$

# Aritmética en punto flotante

## Ejemplo 2

### Paso 1: Comprobación de cero:

$$\begin{array}{l} 1 \quad 11000001110011000000000000000000_2 + \\ 11000001100011100000000000000000_2 \end{array}$$

En este caso se tiene una resta por lo que el **substraendo** se cambia de signo.



# Aritmética en punto flotante

## Ejemplo 2

### Paso 2: Ajuste cifras significativas:

- Exponentes:  $10000011_2, 10000011_2$
- Mantisas:  $1,1001100000000000000000_2,$   
 $1,000111000000000000000000_2$
- Como se puede observar ambos exponentes son iguales, por lo que no se realizan cambios en las mantisas.

# Aritmética en punto flotante

## Ejemplo 2

~~textbf~~Paso 3: Suma: Mantisas:  $1,1001100000000000000000_2$ ,  
 $1,000111000000000000000000_2$

- Como son ambos negativos, se suma sin problema y se toma en cuenta que el resultado es negativo
- El resultado de la suma es:  $10,101101000000000000000000_2$
- Como en este caso no ha quedado de la forma 1,0 es necesario correr la coma una posición a la izquierda, **que es equivalente a sumar 1 al exponente** y la mantisa ahora es:  $1,010110100000000000000000_2$
- Por lo que el nuevo exponente es:  $10000100_2$
- El resultado total es:  
 $11000010001011010000000000000000_2 = -43,25_{10}$ .

34.25

1,5

 $10.0.0.10,01 \quad \} \quad 35,75$ 

(1,1)

$$\left\{ \begin{array}{l} 127+5=132 \\ 127 \end{array} \right. \quad \begin{array}{l} 010000100 \\ 001111111 \end{array} \quad \begin{array}{l} |0001001 \\ |1000 \end{array} \quad +$$

 $01000100 | 0,0,0,0,1 \quad 1$ 
 $1,000.1001$ 
 $0,000011$ 
 $\hline 1,0001111$ 

(1,0)

 $010000100 | 0001111$ 

35,75

## Signo del resultado

Si ambos son positivos da positivo

Si ambos son negativos da negativo

Si tienen signos diferentes el resultado es el que sea mayor (mayor exponente)

34

-120

---

-86

---

# Aritmética en punto flotante

## Multiplicación y división

- Para el caso de multiplicación, se suman los exponentes, pero a uno de ellos se le debe restar el sesgo, de lo contrario lo está sumando 2 veces. En el caso de la división se realiza la resta y se le suma el sesgo.
- Las mantisas se multiplican o dividen segundo el caso
- Considere los signos y aplique la **ley de los signos** según el caso

# Aritmética en punto flotante

## Ejemplo

Realice las siguientes multiplicaciones y divisiones:

1  $01000001100000000000000000000000_2 +$   
 $01000000001000000000000000000000_2$   
 $16_{10} * 2,5_{10}$

2  $01000001100000000000000000000000_2 -$   
 $01000000100000000000000000000000_2$   
 $16_{10}/4_{10}$

# Aritmética en punto flotante

## Ejemplo

Ajuste exponentes:

$$\begin{aligned} & \text{1} \quad 01000001100000000000000000000000_2 + \\ & \quad 01000000001000000000000000000000_2 \\ & \quad 10000011_2 + 10000000_2 - 1111111_2 = 10000100_2 \end{aligned}$$

$$\begin{aligned} & \text{2} \quad 01000001100000000000000000000000_2 - \\ & \quad 01000000100000000000000000000000_2 \\ & \quad 10000011_2 - 10000001_2 + 1111111_2 = 10000001_2 \end{aligned}$$



# Aritmética en punto flotante

## Ejemplo

Ajuste mantisas:

$$\begin{aligned} 1 \quad & 01000001100000000000000000000000_2 + \\ & 01000000001000000000000000000000_2 \\ & 1,000000000000000000000000_2 * \\ & 1,010000000000000000000000 = \\ & 1,010000000000000000000000_2 \end{aligned}$$

$$\begin{aligned} 2 \quad & 01000001100000000000000000000000_2 - \\ & 01000000100000000000000000000000_2 \\ & 1,000000000000000000000000_2 / 1,000000000000000000000000 = \\ & 1,000000000000000000000000_2 \end{aligned}$$

# Aritmética en punto flotante

## Ejemplo

Uniando resultados:

1  $01000010001000000000000000000000_2 = 40_{10}$

2  $01000000100000000000000000000000_2 = 4_{10}$

¿Preguntas?

Siguiente clase:  
Repertorio de Instrucciones:  
Características y funciones