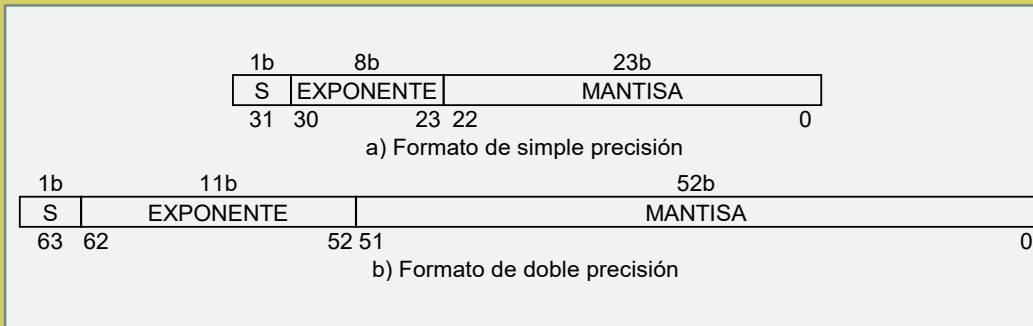


# FORMATO IEEE-754 PARA NÚMEROS EN COMA FLOTANTE

Representación y operaciones básicas

# FORMATO Y REPRESENTACIONES



Rangos de representación:

SIMPLE PRECISIÓN:  $(-1)^S \times 1,M \times 2^{\text{Exp}-127}$

DOBLE PRECISIÓN:  $(-1)^S \times 1,M \times 2^{\text{Exp}-1023}$

- La mantisa se representa en valor absoluto; es fraccionaria y normalizada con un uno implícito a la izquierda de la coma decimal.
- Serán de la forma: 1,XXXXXX...
- Al estar la mantisa normalizada los valores posibles se encontraran entre 1,0000... y 1,1111...
- La base del exponente es siempre 2, y el exponente está representado en exceso a  $2^{q-1}-1$  donde q representa el número de bits que se destinan al campo de exponente (8 u 11).
- Exceso a 127 en el formato de simple precisión
- Exceso a 1023 en el de doble precisión

## VALORES ESPECIALES

E	M	Valores
$2^q - 1$	$\neq 0$	NaN (No un Número)
$2^q - 1$	0	$+\infty$ ó $-\infty$ según el signo S
0	0	Cero
0	$\neq 0$	Números desnormalizados

$\pm 0$  S | 000...00 | 000...00

$\pm \infty$  S | 111...11 | 000...00

NaN X | 111...11 | M  $\neq$  0

S | 000..00 | M  $\neq$  0  $(-1)^S \cdot 0.M \cdot 2^{-126}$   
 $(-1)^S \cdot 0.M \cdot 2^{-1022}$

- Cero e infinito:

- se entienden como límites matemáticos; por eso:

$$+\infty + +\infty = +\infty; -\infty + -\infty = -\infty; \text{etc.}$$

$$+\infty \times \text{positivo} = +\infty; +\infty \times \text{negativo} = -\infty; \text{etc.}$$

$$\text{positivo DIV } +0 = +\infty; \text{positivo DIV } -0 = -\infty; \text{etc.}$$

- atención a las comparaciones:  $+0$  y  $-0$  son iguales

- Not a Number:*

- propagación: cualquier operación donde un operando es NaN dará como resultado NaN

- generación: NaN es el resultado de  $(+\infty) + (-\infty)$ ,  $\pm 0 \times \pm \infty$ ,  $\pm 0 / \pm 0$ ,  $\pm \infty / \pm \infty$  y otras

- una comparación ( $=$ ,  $<$ ,  $\geq$ , etc) entre NaN y otro número resulta siempre falsa

- formato desnormalizado con valores de la forma  $0.M \times 2^{-126}$  para el formato de simple precisión ó  $0.M \times 2^{-1022}$  para el formato de doble precisión.

## REDONDEOS

La norma admite 4 tipos de redondeo:

- Hacia  $+\infty$
- Hacia  $-\infty$
- Hacia 0
- Hacia el más próximo de los dos (este es el modo por omisión).

## Redondeo al más próximo:

En este método hay que resolver de alguna forma una situación de empate.

En el MIPS se redondea hacia la mantisa par, es decir la que acabe en 0.

M	se elige	M resultante
010000	(exacta)	0100
010001	M <sub>-</sub> (más próxima)	0100
010010	M <sub>-</sub> (par)	0100
010011	M <sub>+</sub> (más próxima)	0101
010100	(exacta)	0101
010101	M <sub>-</sub> (más próxima)	0101
010110	M <sub>+</sub> (par)	0110
010111	M <sub>+</sub> (más próxima)	0110
011000	(exacta)	0110

## COMPARACIONES

Las comparaciones pueden ser llevadas a cabo utilizando técnicas de aritmética entera sin signo.

Los signos se procesan por separado, los números negativos ( $S = 1$ ) son menores que los positivos ( $S = 0$ ).

El exponente y la mantisa se pueden procesar concatenados como si se tratase de números enteros positivos, ya que el formato de representación del exponente (exceso a  $2^{q-1}-1$ ) así lo permite.

La representación en exceso a  $2^{q-1}-1$ :

En simple precisión:

- Exponente real:  $-126 \leq \text{exp} \leq 127$
- Exponente representado:  $0 < \text{EXP} < 255$

EXPONENTE REPRESENTADO	VALOR REAL
00000000	Reservado (-127)
00000001	-126
00000010	-125
00000011	-124
.....	.....
01111111	0
10000000	+ 1
10000001	+2
10000010	+3
.....	.....
11111110	+127
11111111	Reservado (+128 )

# OPERACIONES ADITIVAS:

## **A+B ó A-B**

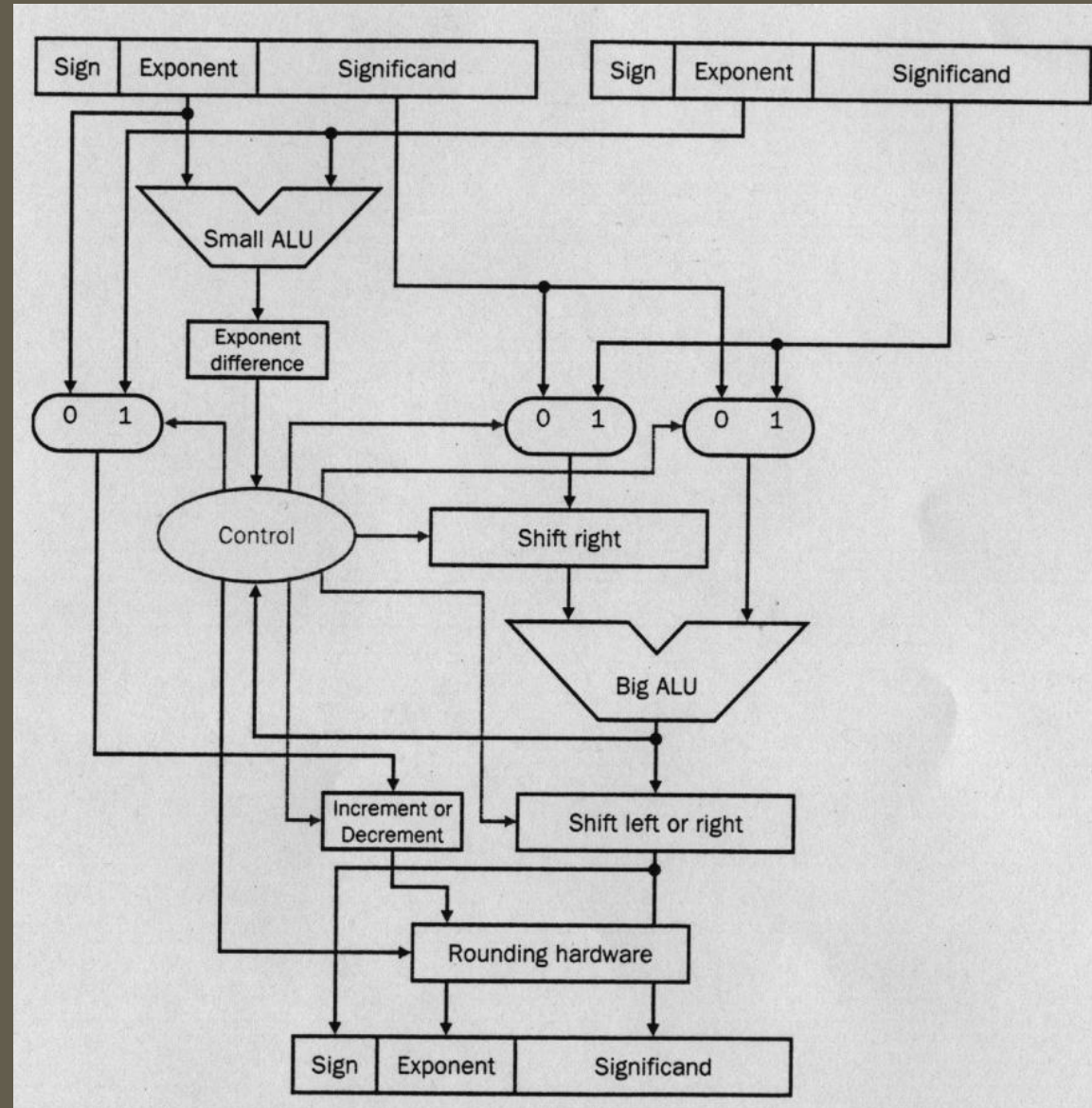
### Pasos:

- Seleccionar el número de menor exponente (requiere comparar exponentes).
- Desplazar su mantisa a la derecha (núm. de posiciones = diferencia absoluta de exponentes)
- El exponente del resultado es el mayor
- Operar las mantisas (en función de la operación a realizar y los signos de los operandos). El resultado se obtiene en formato signo y magnitud.
- Normalizar el resultado i redondear la mantisa

- Selección de la operación a realizar

<i>Signo<sub>A</sub></i>	<i>Signo<sub>B</sub></i>	<i>Operación</i>	<i>Operación Real</i>	<i>Signo Resultado</i>
0	0	Suma	A+B	0
0	1	Suma	A-B	Según Resultado
1	0	Suma	B-A	Según Resultado
1	1	Suma	A+B	1
0	0	Resta	A-B	Según Resultado
0	1	Resta	A+B	0
1	0	Resta	A+B	1
1	1	Resta	B-A	Según Resultado

# SUMADOR/RESTADOR



# OPERACIONES DE MULTIPLICACIÓN Y DIVISIÓN

**$A * B$  ó  $A / B$**

Pasos:

- Sumar o restar los exponentes según la operación a realizar: multiplicación suma exponentes, división resta exponentes
- Sumar o restar el exceso al exponente del resultado
- Multiplicar o dividir las mantisa según la operación a realizar
- Normalizar el resultado y redondear la mantisa

