ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA EN COMPUTACION

PERÍODO ACADÉMICO: 2025-A
ASIGNATURA: ICCD412 Métodos Numéricos **GRUPO:** GR2
TIPO DE INSTRUMENTO: Deber N°3
FECHA DE ENTREGA LÍMITE: 04/05/2025
ALUMNO: Lema Luis

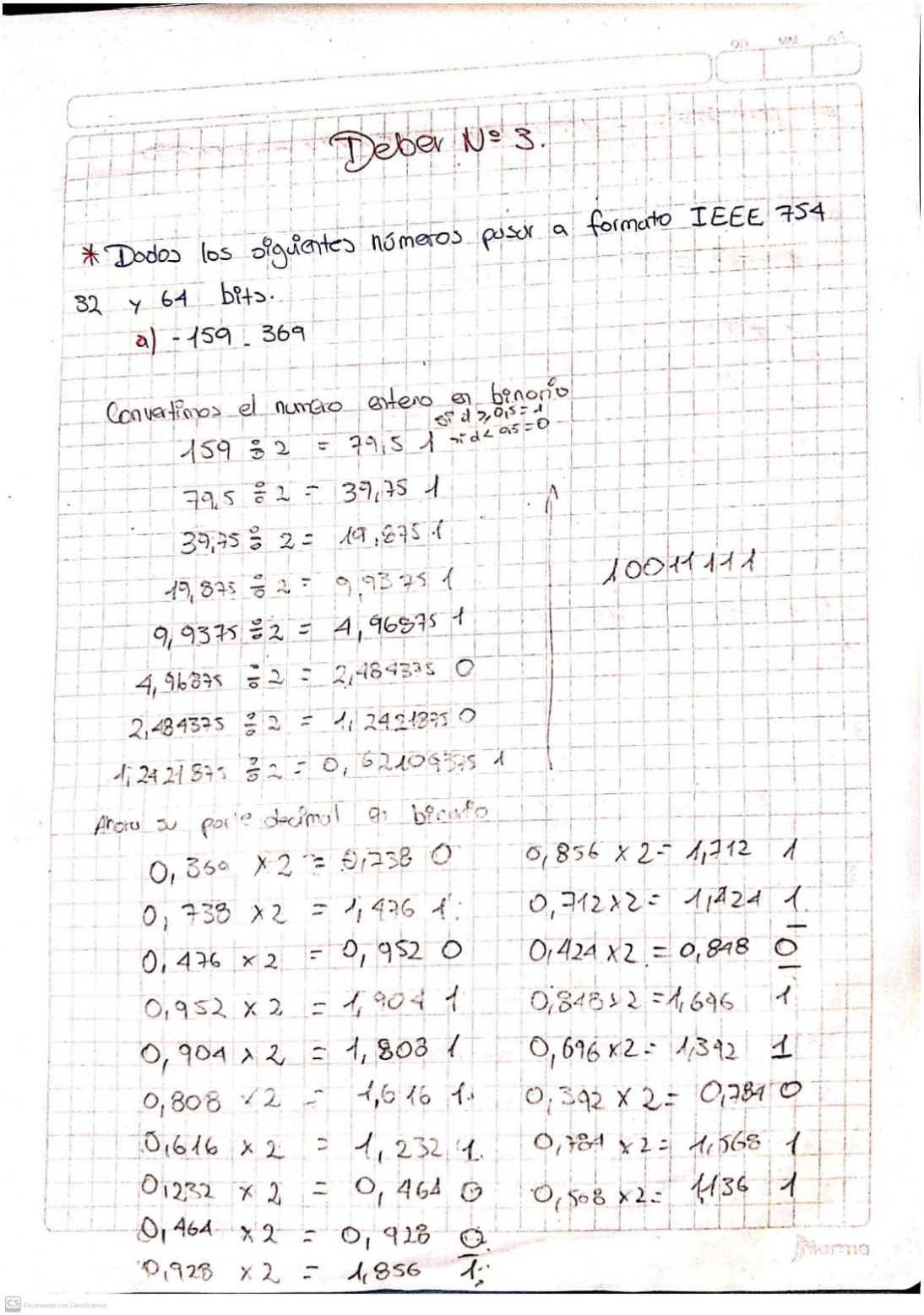
TEMA

Calculo de Errores

OBJETIVOS

- Aplicar los conocimientos vistos en clase sobre la representación en punto flotante según el estándar IEEE 754

DESARROLLO



Deber N° 3.

* Dados los siguientes números pasar a formato IEEE 754 32 y 64 bits.

a) -159 . 369

Convertimos el numero entero en binario

159 ÷ 2 = 79,5 1 si d ≥ 0,5 = 1 si d < 0,5 = 0

79,5 ÷ 2 = 39,75 1

39,75 ÷ 2 = 19,875 1

19,875 ÷ 2 = 9,9375 1

9,9375 ÷ 2 = 4,96875 1

4,96875 ÷ 2 = 2,484375 0

2,484375 ÷ 2 = 1,2421875 0

1,2421875 ÷ 2 = 0,62109375 1

10011111

Ahora su parte decimal en binario

0,369 x 2 = 0,738 0

0,738 x 2 = 1,476 1

0,476 x 2 = 0,952 0

0,952 x 2 = 1,904 1

0,904 x 2 = 1,808 1

0,808 x 2 = 1,616 1

0,616 x 2 = 1,232 1

0,232 x 2 = 0,464 0

0,464 x 2 = 0,928 0

0,928 x 2 = 1,856 1

0,856 x 2 = 1,712 1

0,712 x 2 = 1,424 1

0,424 x 2 = 0,848 0

0,848 x 2 = 1,696 1

0,696 x 2 = 1,392 1

0,392 x 2 = 0,784 0

0,784 x 2 = 1,568 1

0,568 x 2 = 1,136 1

* $-159,369 = -10011111.0101111001111\overline{0111}$

$= -1.0011111010111100111\overline{0111} \times 2^7$

$127 + 7 = 134$

$134 \div 2 =$	$67,0$	0
$67 \div 2 =$	$33,5$	1
$33,5 \div 2 =$	$16,75$	1
$16,75 \div 2 =$	$8,375$	0
$8,375 \div 2 =$	$4,1875$	0
$4,1875 \div 2 =$	$2,09375$	0
$2,09375 \div 2 =$	$1,046875$	0
$1,046875 \div 2 =$	$0,5234375$	1

$131_2 = 10000110.$

* 32 bits

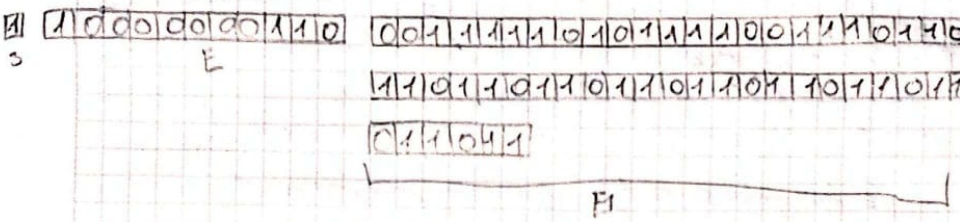
1	10000110	00111111011100111
S	E	M

* 64 bits

$1023 + 7 = 1030$

$1030 \div 2 =$	$515,0$	0
$515 \div 2 =$	$257,5$	1
$257,5 \div 2 =$	$128,75$	1
$128,75 \div 2 =$	$64,375$	0
$64,375 \div 2 =$	$32,1875$	0
$32,1875 \div 2 =$	$16,09375$	0
$16,09375 \div 2 =$	$8,046875$	0
$8,046875 \div 2 =$	$4,0234375$	0
$4,0234375 \div 2 =$	$2,01171875$	0
$2,00585937 \div 2 =$	$1,00585937$	0
$1,00585937 \div 2 =$	$0,50292968$	1

1000000001



b) 3A.28F5C28F5C28

* DE Num_{10} a Num_2

* PARTE ENTERA

3 -> 0011

A -> 1010

0011.1010_2

* Parte Decimal

2 -> 0010

8 -> 1000

F -> 1111

5 -> 0101

C -> 1100

$,0010\,1000\,1111\,0101\,1100_2$

$3A.\overline{28F5C}_{16} = 0011.1010,0010\,1000\,1111\,0101\,1100$

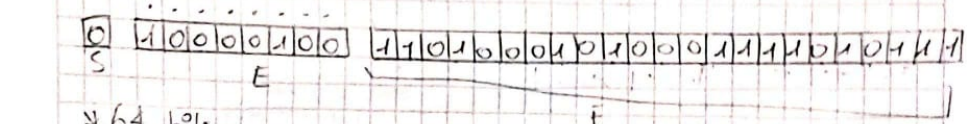
$= 1,1101000101000111101011100 \times 2^5$

$127 + 5 = 132$

$132 \div 2 = 66,0\ \ 0$
$66 \div 2 = 33,0\ \ 0$
$33 \div 2 = 16,5\ \ 1$
$16,5 \div 2 = 8,25\ \ 0$
$8,25 \div 2 = 4,125\ \ 0$
$4,125 \div 2 = 2,0625\ \ 0$

$2,0625 \div 2 = 1,03125\ \ 0$
$1,03125 \div 2 = 0,515625\ \ 1$

10000100



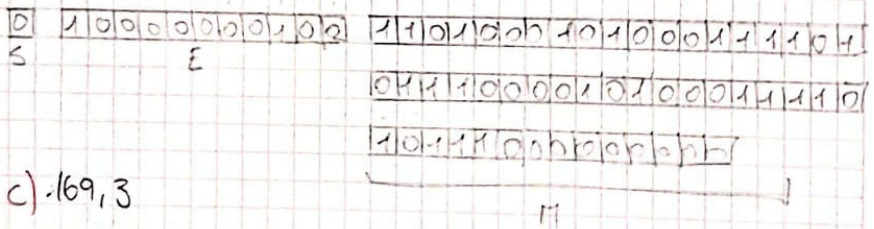
* 64 bits

1023 + 5 = 1028

1028 ÷ 2 = 514,0 0
514 ÷ 2 = 257,0 0
257 ÷ 2 = 128,5 1
128,5 ÷ 2 = 64,25 0
64,25 ÷ 2 = 32,125 0
32,125 ÷ 2 = 16,0625 0

16,0625 ÷ 2 = 8,03125 0
8,03125 ÷ 2 = 4,015625 0
4,015625 ÷ 2 = 2,0078125 0
2,0078 ÷ 2 = 1,00390625 0
1,0039 ÷ 2 = 0,5019531251

10000000100



c) -169,3

* Convertimos en bin_2

+ Parte entera

169 ÷ 2 = 84,5 1
84,5 ÷ 2 = 42,25 0
42,25 ÷ 2 = 21,125 0
21,125 ÷ 2 = 10,5625 1
10,5625 ÷ 2 = 5,28125 0
5,28125 ÷ 2 = 2,640625 1
2,640625 ÷ 2 = 1,3203125 0

1,3203125 ÷ 2 = 0,66015625 1

bin_2 = 10101001.

* Parte decimal

0,30 x 2 = 0,60 0
0,60 x 2 = 1,20 1
0,20 x 2 = 0,40 0
0,40 x 2 = 0,80 0
0,80 x 2 = 1,60 1
0,60 x 2 = 1,20 1
0,20 x 2 = 0,40 0
0,40 x 2 = 0,80 0
0,80 x 2 = 1,60 1
0,60 x 2 = 1,20 1

$0,10\overline{0011}$

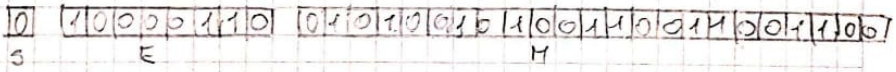
$127+7 = 134$ en código exceso

10000110_2

* Notación Científica

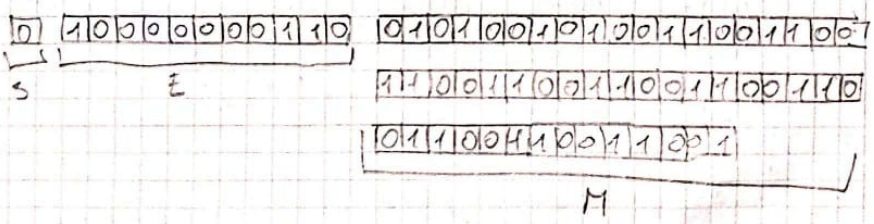
$169,3_{10} = 10101001,010011_2$

$= 1,0101001010\overline{0011} \times 2^7$



S E M

* Arrow 64 bits. $1023+7 = 1030_{10} \rightarrow 10000000110_2$



S E

M

DD MM AA

* Dados el siguiente numero binario en IEEE 754 de 64 bits pasar a decimal.

0 1 0 0 0 0 0 0 0 1 1 1 1 0 0 0 1 1 1 0 1 0 1 0 0 0 1 1 1 1 0 1 0 1 1 1 0 0 0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

1 0 1 0 0 0 1 1 1 1 0 1 0 1 1 1 0 0 0 0 1 0 1 0
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52

* Separamos en Partes.

Signo 0 → Positivo

Exponente → $1000000011_2 = (1\times2^{10}) + (1\times2^1) + (1\times2^0) = 1031_{10}$

Mantiza → 52 bits.

* Calculamos e

$e = c - bias = 1031 - 1023 = 8$

* Calculamos f

$$f = \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^6 + \left(\frac{1}{2}\right)^7 + \left(\frac{1}{2}\right)^9 + \left(\frac{1}{2}\right)^{11} + \left(\frac{1}{2}\right)^{15} + \left(\frac{1}{2}\right)^{16} + \left(\frac{1}{2}\right)^{17} + \left(\frac{1}{2}\right)^{18} + \left(\frac{1}{2}\right)^{20} + \left(\frac{1}{2}\right)^{22} + \left(\frac{1}{2}\right)^{23} + \left(\frac{1}{2}\right)^{24} + \left(\frac{1}{2}\right)^{29} + \left(\frac{1}{2}\right)^{31} + \left(\frac{1}{2}\right)^{35} + \left(\frac{1}{2}\right)^{36} + \left(\frac{1}{2}\right)^{39} + \left(\frac{1}{2}\right)^{38} + \left(\frac{1}{2}\right)^{40} + \left(\frac{1}{2}\right)^{42} + \left(\frac{1}{2}\right)^{43} + \left(\frac{1}{2}\right)^{44} + \left(\frac{1}{2}\right)^{49} + \left(\frac{1}{2}\right)^{51}$$

$f = 0,55719 \rightarrow$ 4 C.S y Redondeo

$f+1 = 1,55719_{10}$

$X = (-1)^0 \cdot 2^8 \cdot 1,55719$

$X = 398,6 \rightarrow$ 4 Sifras.

* Dado el siguiente numero binario en IEEE 754 de 32 bits pasar a decimal.

1 10000110 1 00 1 1 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0 1 1 0 1

1 5 12 13 16 17 20 21 23

* Signo = 1 = negativo.

* Exponente = $10000110 = (1 \times 2^7) + (1 \times 2^2) + (1 \times 2^1)$

$= 134_{10}$

$e = 134 - 127 = 7$

* Mantisa = $f = \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^{12} + \left(\frac{1}{2}\right)^{13} + \left(\frac{1}{2}\right)^{16} + \left(\frac{1}{2}\right)^{17} + \left(\frac{1}{2}\right)^{20} + \left(\frac{1}{2}\right)^{21} + \left(\frac{1}{2}\right)^{23}$

$f = 0,5941 \rightarrow$ 4 Cifras.

$f + 1 = 1,5941.$

$X = (-1)^1 \, 2^7 \, (1,5941)$

$X = -204,045_{11}$

CONCLUSIONES

- Realizar los ejercicios propuestos permitió tener un conocimiento mas profundo sobre los números en los sistemas digitales usando el estándar IEEE 754 tanto en 32 bits como en 64 bits.