4.5.4 Binary Number Systems

1) A particular computer uses a **normalised** floating point representation with an 8-bit mantissa and a 4-bit exponent, both stored using **two's complement**.

Four bit patterns that are stored in this computer's memory are listed in **Figure 1** and are labelled **A**, **B**, **C** and **D**. Some of the bit patterns are valid normalised floating point numbers.

Figure 1

	Mantissa								Exponent			
A	0	0	0	0	0	0	1	1	1	1	0	0
B	0	1	0	0	0	0	0	0	0	0	0	0
C	1	0	0	0	0	0	0	0	0	0	1	0
D	1	1	0	0	1	1	1	0	1	0	0	0

a) State which bit pattern represents a negative normalised value.

[1 mark]

- C

b) State which bit pattern represents the smallest positive normalised value.

[1 mark]

- B

c) The following is a floating point representation of a number:

Mantissa								Exponent			
1	0	1	0	1	1	0	0	0	1	0	1

Calculate the decimal equivalent of the number. You **must** show your working.

[2 marks]

- Mantissa $= -\frac{21}{32}$ Exponent $= 5$
- $-\frac{21}{32} \times 2^5 = -21$

d) Write the normalised floating point representation of the decimal value 23.75. You **must** show your working.

[3 marks]

- **23.75 in binary is 010111.11**
- **Shift point 5 places to the left // +5 to exponent**
- **Mantissa 0.1011111 Exponent 0101**

There can be a loss of precision when a decimal number is stored when using a floating point system.

The closest possible representation of the decimal number 13.8 is shown below.

Mantissa

0	1	1	0	1	1	1	0

Exponent

0	1	0	0

By converting this bit pattern back into denary the actual number stored is 13.75, not 13.8.

e) Calculate the absolute error that has occurred.

[1 mark]

- $\mathbf{13.8 - 13.75 = 0.05}$

f) Calculate the relative error that has occurred. Express your answer as a percentage to two decimal places.

[1 mark]

- $\mathbf{\frac{0.05}{13.8} \times 100 = 0.36\%}$

2) Place a tick in the row to indicate which one of the unsigned numbers in **Table 1** has the largest value.

a)

Table 1

Number base	Number	Tick ✓
Binary	101111001	
Hexadecimal	20A	✓
Decimal	39619	

[1 mark]

b) The following is a number in **normalised** floating point representation with a 7-bit mantissa and a 5-bit exponent, both stored in **two's complement.**

Mantissa

0	1	0	1	1	0	0

Exponent

1	1	1	0	1

Calculate the decimal equivalent of the number. You **must** show your working.

[2 marks]

- **Mantissa** $= \frac{11}{16}$ **Exponent** $= -3$
- $\frac{11}{16} \times 2^{-3} = \frac{11}{128}$

c) Write the normalised floating point representation of the decimal value -608. You **must** show your working.

[3 marks]

- **608 in binary is 01001100000.**
- **-608 in binary is 10110100000.**
- **Shift point 10 places to the left // +10 to exponent**
- **Mantissa 1.011010 Exponent 01010**

3) A student has attempted to add together the binary numbers but has made a mistake

	A	B	C	D	E	F	G	H
	0	0	1	1	0	0	1	1
+	1	0	1	1	0	1	1	0
Carry	0	1	1	0	1	1	0	
Result	1	1	0	0	1	0	0	1

Explain what mistake the student has made. The columns have been labelled **A** to **H** to aid in your explanation.

[1 mark]

- **In column C, Result should be 1**

---

4) The following questions use a **normalised** floating point representation with an **8-bit** mantissa and a **4-bit** exponent, both stored using **two's complement**.

a) Write the smallest positive number that can be represented by the floating point system in the boxes below.

Mantissa

0	1	0	0	0	0	0	0

Exponent

1	0	0	0

[2 marks]

b) The following is a floating point representation of a number:

Mantissa

1	0	1	1	0	0	1	0

Exponent

0	0	1	0

Calculate the decimal equivalent of the number

You **must** show your working.

[2 marks]

- **Mantissa = -39/64 Exponent = 2**
- $-\frac{39}{64} \times 2^2 = -\frac{39}{16}$

c) Write the normalised floating point representation of the decimal value 0.15625 (5/32 as a fraction).

You **must** show your working.

- **0.15625 in binary is 0.00101**
- **Shift point two places right // -2 to exponent**

Mantissa								Exponent			
0	1	0	1	0	0	0	0	1	1	1	0

[3 marks]

d) The two floating point numbers below are multiplied together:

Mantissa								Exponent			
0	1	0	1	0	0	0	0	0	1	1	1

Mantissa								Exponent			
0	1	1	0	0	1	0	0	0	0	1	1

A problem occurs as a result of the multiplication operation.

Explain what problem has occurred and how the floating point representation could be redesigned to avoid it.

[3 marks]

- **Overflow will occur // it will not be possible to store the result/the exponent in the available number of bits**
- **More bits should be added // bits could be reallocated from the mantissa to the exponent**

5) The following questions use a **normalised** floating point representation with a **7-bit** mantissa and a **5-bit** exponent, both stored using **two's complement**.

a) Non-integer values such as -2.35 and 7/117 can be represented by a computer using a fixed point or a floating point system.

State **one** advantage of using a floating point system over a fixed point system and **one** advantage of using a fixed point system over a floating point system.

You should assume that the two systems use the same number of bits to store a value.

[2 marks]

Floating point
- **Can represent numbers with a greater range**

Fixed point
- **Can represent (some) numbers more precisely**
- **Calculations can be performed more quickly**
- **Represents all numbers to a constant level of accuracy**

b) Calculate the decimal equivalent of the number below:

Mantissa

1	0	1	0	1	0	0

Exponent

1	1	1	0	1

Express your answer to at least four decimal places or as a fraction. You should show your working.

[2 marks]

- **Mantissa =** $-\frac{11}{16}$ **Exponent = -3**
- $-\frac{11}{16} \times 2^{-3} = -\frac{11}{128}$

c) Write the normalised floating point representation of the decimal value 1632 in the boxes below.

You should show your working.

- **1632 in binary is 011001100000.**
- **Shift point 11 places to the left // +11 to the exponent**

Mantissa

0	1	1	0	0	1	1

Exponent

0	1	0	1	1

[3 marks]

d) State, in decimal, the highest (most positive) and lowest (most negative) values that could be represented by this floating point system.

You should show your working.

[3 marks]

- **Highest value: 32,256**
- **Lowest value: -32,768**

e) When the decimal value 28.25 is converted into binary using this floating point system, a rounding error occurs.

Explain:

- Why a rounding error has occurred
- What the system might do when the value 28.25 is converted into binary.

[2 marks]

- **There are not enough bits in the mantissa to represent 28.25 exactly**
- **It could be rounded/truncated // represented as 28/28.5**

---

6) Use binary addition in 8-bit two's complement to perform the subtraction:

18 – 72

You **must** show both your working and your final answer in binary.

[2 marks]

- **18: 00010010   -72: 10111000**
- **Answer: 11001010**

7) The following questions use a **normalised** floating point representation with an **8-bit** mantissa and a **4-bit** exponent, both stored using **two's complement**.

a) Four-bit patterns (**A**-**D**) are shown below. Three of the patterns are valid normalised floating point numbers and one is not.

	Mantissa								Exponent			
A	1	0	0	0	0	0	0	0	0	1	1	1
B	0	1	0	0	0	0	0	0	1	1	1	1
C	1	1	1	0	1	0	0	0	1	0	1	0
D	0	1	1	1	1	1	1	1	0	1	1	1

Complete the table below by writing the appropriate letter from **A** to **D** in indicate which bit pattern above is an example of the type of value described in the **Value description** column.

Value description	Correct letter (A-D)
A negative value that is valid in the representation.	A
The largest positive value that can be represented in the system.	D
A value that is not valid in the representation because it is not normalised	C

[3 marks]

b) The following is a floating point representation of a number:

Mantissa

0	1	1	0	1	0	0	0

Exponent

0	1	1	0

Calculate the decimal equivalent of the number. You **must** show your working.

[2 marks]

- **Mantissa = 13/16 exponent = 6**
- **Shift point 6 places to the right 0110100.0 = 52**

c) In this floating point system, the closest possible representation of the decimal number 104.7 is shown below.

Mantissa

0	1	1	0	1	0	0	0

Exponent

0	1	1	0

By converting this number to decimal, the actual value stored is 105.

Calculate the **absolute error** that has occurred when representing 104.7.

[1 mark]

- **0.3**

d) Calculate the **relative error** that has occurred when representing 104.7.

[1 mark]

- **0.29%**

e) Explain why the relative error is usually considered to be a more important measure of error than the absolute error.

[1 mark]

- **The effect of the error depends on its size relative to the number that is should be represented**

8) The number below is stored using a **fixed point** representation and **two's complement**, with six bits before and four bits after the binary point.

Figure 2

1	0	1	1	0	● 0	1	0	1	1

Convert the number in **Figure 2** to decimal.

Show should show your working.

[2 marks]

- -309/16

---

9) State **two** reasons why values stored using a **floating point** representation are usually stored in normalised form.

[2 marks]

- Maximises precision / accuracy for the given number of bits
- Gives a unique representation of each number // simpler to test for equality of numbers

---

10) The following questions use a **normalised** floating point representation with an **8-bit** mantissa and a **4-bit** exponent, both stored using **two's complement**.

a) The following is a floating point representation of a number:

Mantissa

0	1	1	0	1	0	0	0

Exponent

1	0	0	1

Calculate the decimal equivalent of the number.

Express your answer to at least four decimal places or as a fraction. You should show your working.

[2 marks]

- Mantissa = 27/32   exponent = -7
- $\frac{27}{32} \times 2^{-7} = \frac{27}{4096}$

11) Write the normalised floating point representation of the decimal value -23.25 in the boxes below.

- **23.25 = 010111.01**
- **-23.25 = 101000.11**
- **Shift point 5 to the left // +5 to exponent**

Mantissa

1	**0**	**1**	**0**	**0**	**0**	**1**	**1**

Exponent

0	**1**	**0**	**1**

You should show your working.

[3 marks]

---

12) Figure 3 shows the closest possible representation of the decimal number -0.22558594 in the floating point system.

Figure 3

Mantissa

0	**1**	**1**	**0**	**1**	**0**	**0**	**0**

Exponent

0	**1**	**1**	**0**

By converting this number back to decimal, the actual value stored is -0.2265625

Calculate the **relative error** that has occurred as a percentage to 2 decimal places. You should show your working.

[2 marks]

- **0.43%**

---

13) The following questions use a **normalised** floating point representation with an **8-bit** mantissa and a **4-bit** exponent, both stored using **two's complement**.

a) The following is a floating point representation of a number:

Mantissa

0	1	1	0	1	0	0	0

Exponent

1	1	0	1

Calculate the decimal equivalent of the number.

Express your answer to at least four decimal places or as a fraction. You should show your working.

[2 marks]

- **Mantissa = 13/16 exponent = -3**
- $\frac{13}{16} \times 2^{-3} = \frac{13}{128}$

b) On each row of **Table 1**, state the name of the **Type of error** that has occurred in the **Situation** that is described.

Table 1

Situation	Type of error
A calculation is performed, and the result of the calculation is so close to zero that the number that is stored is zero.	**Underflow**
A calculation is performed, and the result of the calculation is too large to fit in the available number of bits.	**Overflow**
A decimal value is converted to floating point, but it cannot be represented exactly in the available number of bits.	**Rounding**

[2 marks]

c) Explain how the floating point representation could be modified to represent numbers more precisely, without changing the total number of bits used to represent a number.

[1 mark]

- **Move a/some bit(s) from the exponent to the mantissa**

14) The following questions use a **normalised** floating point representation with an **8-bit** mantissa and a **4-bit** exponent, both stored using **two's complement**.

a) Five statements about representing numbers using fixed and floating point representations are shown below.

A	A processor can usually carry out calculations on fixed point number more quickly than calculations on floating point numbers.
B	Fixed point numbers represent data using a mantissa and an exponent.
C	In a given number of bits, a fixed point system can represent positive numbers that are closer to zero than a floating point system.
D	In a given number of bits, a fixed point system can represent some numbers more precisely than a floating point system.
E	In a given number of bits, a floating point system can represent a bigger range of numbers than a fixed point system.

Which **three** of the statements (**A**-**E**) are **true**.

[2 marks]

- **A, D and E**

b) The following is a floating point representation of a number:

Mantissa

1	0	1	1	1	0	0	0

Exponent

1	0	1	1

Calculate the decimal equivalent of the number.

[2 marks]

- **Mantissa** $= -\frac{9}{16}$ **exponent = -5**
- $-\frac{9}{16} \times 2^{-5} = -\frac{9}{512}$

c) The decimal number 12.765625 (which can also be expressed as $12\frac{49}{64}$) cannot be represented exactly in this floating point system.

Write the closest possible normalised representation of this number in the boxes below.

You should show your working.

Mantissa

0	1	1	0	0	1	1	0

Exponent

0	1	0	0

[3 marks]

d) What is the smallest number of bits that would need to be added to the mantissa so that the decimal number 12.765625 could be represented exactly?

[1 mark]

- **3 bits**

e) A different system uses a **normalised** floating point representation with a **10-bit mantissa** and a **6-bit exponent**, both stored using **two's complement**.

Mantissa

Exponent

In decimal, what is the most negative number that this system could represent? You should show your working.

[2 marks]

- **1.000000000 011111**
- -2^{31} **// -2147483648**

15) The following questions use a **normalised** floating point representation with an **8-bit** mantissa and a **4-bit** exponent, both stored using **two's complement**.

a) The following is a floating point representation of a number:

Mantissa

0	1	0	1	1	0	0	0

Exponent

0	1	0	1

Calculate the decimal equivalent of the number.

Express your answer to at least four decimal places or as a fraction. You should show your working.

[2 marks]

- **Mantissa = 0.6875** **Exponent = 5**
- $0.6875 \times 2^5 = 22$

b) Write the normalised floating point representation of the decimal value -6.75 in the boxes below.

- **-6.75 is 1001.11**
- **Shift point 3 to the left // +3 to the exponent**

Mantissa

1	0	0	1	0	1	0	0

Exponent

0	0	1	1

You should show your working.

[3 marks]

c) An alternative two's complement format representation is proposed. In the alternative representation 6 bits will be used to store the mantissa, and 6 bits will be used to store the exponent.

Explain the effects of using the proposed alternative representation instead of the existing representation

[2 marks]

- **Reduced precision**
- **Increased range**

16) Figure 3 contains a bit pattern.

Figure 3

0	0	1	1	1	0	0	1

a) What is the decimal equivalent of the bit pattern shown in **Figure 3** if it represents a **two's complement binary integer**?.

[1 mark]

- **57**

b) What is the decimal equivalent of the bit pattern show in **Figure 3** if it represents an **unsigned fixed-point binary** value with four bits before the binary point and four bits after the binary point?

[2 marks]

- $\frac{57}{16}$ **// 3.5625**