THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

UNIVERSITY OF LONDON **CO3325 ZB**

BSc Examination

COMPUTING AND INFORMATION SYSTEMS, CREATIVE COMPUTING, AND COMBINED DEGREE SCHEME

Data Compression

Date and Time: Friday 20 May 2016: 10.00–12.15

Duration: 2 hours 15 minutes

There are THREE questions in this paper. Candidates should answer all **THREE** questions. All questions carry equal marks and full marks can be obtained for complete answers to **THREE** questions. The marks for each part of a question are indicated at the end of the part in [.] brackets.

There are 75 marks available on this paper.

A hand held calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

Question 1

(a) Briefly describe the static Huffman encoding algorithm and the Shannon-Fano encoding algorithm. Demonstrate each algorithm using the text example 'BAGHABGHGGGAAGH', focusing on the differences between the two algorithms. [10]

(b) Explain why the Reflected Grey Code is a better representation than normal binary codes for coding the colours of greyscale images. Derive the *Reflected Grey Code* for the colour codes in decimal below. [5]

11	10
10	9

(c) Consider two commonly used colour representations RGB and LC, and the transform functions for mapping $RGB \to LC$: [10]

$$\begin{cases} Y & \approx & 0.3R + 0.6G + 0.1B \\ C_b & = & B - Y \\ C_r & = & R - Y \end{cases}$$

i. Explain what is meant by *transform* in the context of Data Compression.
ii. Given $(R, G, B) = (1, 2, 3)$, what are the corresponding values for (Y, C_b, C_r)?
iii. Given $(Y, C_b, C_r) = (1, 2, 3)$, what are the corresponding RGB values after the detransform $LC \to RGB$?

Question 2

(a) Explain and demonstrate in a diagram, using the example code (0, 110, 111, 1), how a tree (data structure) can be used both to represent a binary code, and to tell whether or not the code is a prefix code. [3]

(b) Consider the context of text compression. Let the probabilities of the n symbols of a source alphabet be $p_1, p_2, \cdots, p_n$, and the codeword lengths be $l_1, l_2, \cdots, l_n$ respectively.

i. Explain what an *optimal code* is. [2]

ii. Discuss the truth of each of the following statements. "1. Huffman codes are optimal in general. 2. Shannon-Fano codes are not optimal in general."
Justify your answers and show all your work. [9]

iii. Give an example of an optimal code for $n = 4$ with its codeword probabilities. [3]

(c) Explain each of the following terms in the context of Data Compression. Provide an example of the entity described by each term. [8]

i. variable-to-variable model

ii. bi-level image

Question 3

(a) Consider a binary source file consisting of characters A and B with probability of 0.2 for B. Discuss the cause of inefficiency of applying the static Huffman compression algorithm to this source, and compute how much less than optimal it is. Explain and demonstrate how the efficiency of encoding may be improved. [10]

(b) Consider part of a grayscale image with 16 shades of gray that is represented by the array `A` below:

```
   0011 0010 1100 0111
A: 0011 0001 1100 0110
   0111 1100 1101 1011
```

Demonstrate how the array A can be separated into (four) bitplanes as a preprocess. Explain how the bitplanes can be used to achieve a better compression ratio in the main compression process. [8]

(c) Explain what is meant by *sampling* in the context of sound compression. Consider a sine wave signal $s(t) = 2sin(880\pi t)$. Draw the *frequency spectrum diagram* for the signal. What is the minimum sample rate which would allow the reconstruction of the signal from the samples? [7]

END OF PAPER