COMPUTER SCIENCE TRIPOS Part II (General)
DIPLOMA IN COMPUTER SCIENCE

Thursday 7 June 2007 1.30 to 4.30

PAPER 13 (PAPER 4 OF DIPLOMA IN COMPUTER SCIENCE)

*Answer **five** questions.*

*Submit the answers in five **separate** bundles, each with its own cover sheet. On each cover sheet, write the numbers of **all** attempted questions, and circle the number of the question attached.*

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

STATIONERY REQUIREMENTS
Script paper
Blue cover sheets
Tags

SPECIAL REQUIREMENTS
None

1 Natural Language Processing

(*a*) There are several different types of ambiguity in natural language. For *each* of the following cases, briefly describe how ambiguity may arise, illustrating your answer with examples:

(*i*) morphological ambiguity;

(*ii*) lexical ambiguity;

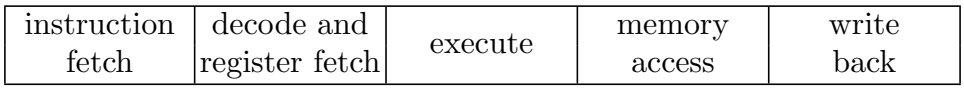
(*iii*) syntactic ambiguity;

(*iv*) ambiguity in rhetorical relations (discourse relations).

[12 marks total]

(*b*) Explain what is meant by "packing" in parsing and discuss its relevance to the treatment of ambiguity. [8 marks]

2 Computer Design



With reference to the pipeline above:

(*a*) What is a *control hazard* and how can it be dealt with? [4 marks]

(*b*) What are *data hazards* and how can they be eliminated? [4 marks]

(*c*) A branch could be executed in either the decode or the execute stages. Assuming that branch prediction and branch delay slots are not used, how many bubbles would be introduced into the pipeline in either case? [4 marks]

(*d*) If the memory access results in a cache miss, what happens to the pipeline? [4 marks]

(*e*) For arithmetic operations the result is available after the execute stage. These results could be written directly to the register file during the memory access stage, but what would be the disadvantages of doing so? [4 marks]

3 Digital Communication I

(*a*) Define the term *flow control* as used in communication networks. [4 marks]

(*b*) Describe on–off flow control, window-based flow control, and flow control used in circuit switching. [9 marks]

(*c*) Consider a channel of capacity b and delay τ, over which packets of size p are sent. Compare the performance of window-based flow control protocols having:

(*i*) a window size of one packet;

(*ii*) a window size of two packets; and

(*iii*) a window size of one packet, but with a packet size of $2p$.

[7 marks]

4 Computer Graphics and Image Processing

(*a*) Explain what a *MIPmap* is, how to create one, why one would want to use one, where one would be used, and how one is used. [8 marks]

(*b*) Describe an algorithm that converts a greyscale image into a black and white image using halftoning. Assume that the black and white image has eight times the resolution of the greyscale image in each dimension. [6 marks]

(*c*) Various types of visual artifact ("aliasing") occur if images are rendered using only one sample per pixel.

(*i*) Describe at least *three* different artifacts that occur. [3 marks]

(*ii*) Describe a straightforward method to ameliorate these artifacts. [3 marks]

5 Distributed Systems

(*a*) When distributed systems are designed and engineered, certain fundamental properties have to be taken into account, including:

1. concurrent execution of components
2. independent failure modes
3. communication delay
4. no global time

Give *three* examples of the implications of these properties (separately or in combination) on the engineering of large-scale, widely distributed systems. [9 marks]

(*b*) (*i*) Define role-based access control (RBAC).

(*ii*) Outline how RBAC could be used for a national healthcare system comprising many administration domains such as primary care practices, hospitals, specialist clinics, etc. Principals may, from time to time, work in domains other than their home domain, and must be authorised to do so.

(*iii*) A national Electronic Health Record (EHR) service must be accessible from all domains. It is required by law that access control policy should be able to capture exclusion of principals and relationships between them. How could this requirement be met in an RBAC design?

[11 marks]

6 Compiler Construction

(*a*) Garbage Collection.

(*i*) Explain how it is possible to "leak memory" using a reference counting garbage collector. [3 marks]

(*ii*) Describe any technique that might be used to address this problem. [3 marks]

(*b*) Explain in detail how we might translate code generated for a stack-only machine (such as the JVM) to a register-based machine (such as ARM). [6 marks]

(*c*) Describe in detail how static and dynamic methods are compiled differently for object-oriented languages with single inheritance. [8 marks]

7 Concepts in Programming Languages

(a) Give an overview of the LISP abstract machine (or execution model) and comment on its merits and drawbacks from the viewpoints of programming, compilation, execution, etc. [5 marks]

(b) Define the following parameter-passing mechanisms: pass-by-value, pass-by-reference, pass-by-value/result, and pass-by-name. Briefly comment on their merits and drawbacks. [5 marks]

(c) What is *aliasing* in the context of programming languages? Explain the contexts in which it arises and provide examples of the phenomenon. [5 marks]

(d) Consider the Simula declarations

```
CLASS A; A CLASS B;
```

which have the effect of producing the subtype relation `B<:A`, and

```
REF(A) a; REF(B) b;
```

Recall that Simula uses the semantically incorrect principle that

if `B<:A` then `REF(B)<:REF(A)`

and consider now the following Simula code

```
PROCEDURE ASSIGNa( REF(A) x )
    BEGIN x :- a END ;

ASSIGNa(b);
```

Does it statically type check? If so, will it cause a run-time type error?

Justify your answers. [5 marks]

8 Databases

(a) Define the notion of a *functional dependency.* [2 marks]

(b) Consider the following "rule" for functional dependencies.

$$\text{if } A \to B \text{ and } B, C \to D, \text{ then } A, C \to D.$$

Either prove this rule is correct, or present a counter-example showing that the rule is false. [4 marks]

(c) The *union rule for functional dependencies* states that if $F \models X \to Y$ and $F \models X \to Z$, then $F \models X \to Y \cup Z$ (this can also be written as $F \models X \to Y, Z$).

Prove this rule using only Armstrong's axioms. [4 marks]

(d) Suppose that $R(A, B, C)$ is a relational schema. Write a relational algebra query that evaluates to the empty set exactly when the functional dependency $B \to C$ holds on R. [4 marks]

(e) The schema $R(A,\ B,\ C,\ D,\ E)$ has the following functional dependencies.

$$A \to B, C$$
$$C, D \to E$$
$$B \to D$$
$$E \to A$$

Is D, E a candidate key for R? Explain your answer. [6 marks]

9 Quantum Computing

(a) Consider a quantum finite automaton with two basis states, $|0\rangle$ being the start state and $|1\rangle$ the only accepting state. The automaton operates on a two-letter alphabet, with matrices $M_a = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$ and $M_b = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

What are the probabilities that the automaton accepts each of the following input strings?

(i) a [3 marks]

(ii) aba [3 marks]

(iii) abb [3 marks]

(b) Give a complete description of the probabilities of acceptance associated with various possible input strings. [5 marks]

(c) Prove that there is no *two-state* probabilistic automaton with the same behaviour as the automaton described in part (a). [6 marks]

10 Bioinformatics

(a) Describe a bioinformatics application of hidden Markov models. [6 marks]

(b) Discuss the properties of the Markov clustering algorithm and the differences with respect to the k-means and hierarchical clustering algorithms. [8 marks]

(c) Describe the Gillespie algorithm and discuss its relationship with genetic or biochemical networks (give one example). [6 marks]

11 Software Engineering and Design

(*a*) How can understanding of human memory help us to design interactive systems that assist memory? Explain how you might design a web-based reminder service that exploits this understanding. [4 marks]

(*b*) How would you go about a user-centred design process for such a reminder service? Describe

(*i*) the overall process model; [4 marks]

(*ii*) specific actions to be taken at *three* different points in this process. [4 marks each]

12 Complexity Theory

Let **Bounded Factor** denote the following decision problem:

> Given positive integers n and k, decide whether n has a proper factor that is less than k.

For each of the following questions, give a detailed justification of your answer. Note that, in some cases, the answer may not be a simple "yes" or "no" but may instead be linked to open problems in complexity theory such as whether P=NP or the existence of one-way functions. In such cases, you should clearly explain the links and state what the consequences might be of both a positive and a negative answer to the question.

(*a*) Is **Bounded Factor** in NP? [4 marks]

(*b*) Is **Bounded Factor** in co-NP? [4 marks]

(*c*) Is **Bounded Factor** NP-hard? [6 marks]

(*d*) Is **Bounded Factor** in P? [6 marks]

END OF PAPER