COMPUTER SCIENCE TRIPOS Part Ib

Thursday 4 June 1998 1.30 to 4.30

Paper 6

Answer **five** *questions.*
No more than **two** *questions from any one section are to be answered.*
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.*
Write on **one** *side of the paper only.*

SECTION A

1 Structured Hardware Design

A child's toy is battery operated and has four large illuminated push buttons, a number of smaller buttons and a loudspeaker that is used to play tunes. Production will be 50 000 units per month for at least six months.

(*a*) Give a block diagram of the product, showing the principal partition decisions. Describe the function of each block, explaining why it exists and how it should be made. [6 marks]

(*b*) You may (or may not) have used a microprocessor – explain why your design strategy is a good idea and discuss the alternative approach. [6 marks]

(*c*) How would you test the product concept before production? [4 marks]

(*d*) How would you test the product in production before shipping? [4 marks]

2 Computer Design

Computer memory is usually organised as a *memory hierarchy*. Why is this the case? [4 marks]

What are the relative latency and bandwidth characteristics of each level of a typical memory hierarchy? [4 marks]

How does a direct-mapped cache work and what might the data replacement policy be? [6 marks]

How does a set-associative cache work and what might the data replacement policy be? [6 marks]

3 Digital Communication I

You are required to design a topology discovery protocol for a network of switching nodes interconnected by links. There are n nodes, l links, the maximum degree of any node is k and there is a path between any two nodes of not more than d hops. All links are bi-directional.

Each node has a unique identifier of four bytes which it knows.

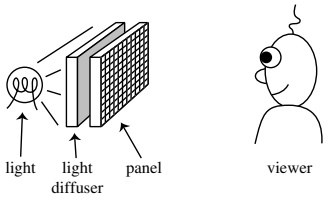
(a) Design a protocol (including message formats) for a node to learn about its immediate neighbours. [5 marks]

(b) Design a protocol (including message formats) for distributing this information across the network. [10 marks]

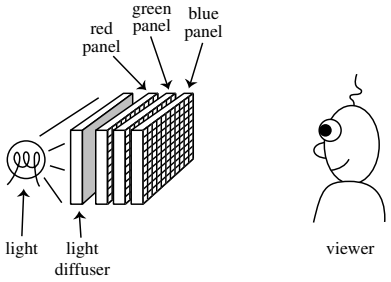
(c) Give a bound on the total amount of information which is transmitted to ensure that every node acquires complete topology information. [5 marks]

4 Computer Graphics and Image Processing

An inventor has recently developed a new display device: it is a transparent panel with a rectangular array of square pixels. The panel is tinted with a special ink which allows each pixel to range from totally transparent to transmitting only the colour of the ink. Each pixel has an 8-bit value. For example, if the ink is blue then a pixel value of 0 would be totally transparent, 255 totally blue (only blue light transmitted) and 100 a light blue.



The inventor has recently found that he can make the special ink in *any* colour he likes, but that each panel can be tinted with only one of these colours. He proposes to use three inks in three panels to make a 24-bit colour display: a red-tinted panel, a green-tinted panel and a blue-tinted panel will be stacked up to make a full-colour display (see picture). A value of $(0, 0, 0)$ will thus be white (transparent), $(255, 0, 0)$ red and $(255, 255, 255)$ black.



Explain why this will not work. [4 marks]

Modify the three-panel design so that it will work. [3 marks]

In common with other 24-bit "full-colour" displays (for example CRT, LCD), your display *cannot* display *every* colour which a human can perceive. Why not? [3 marks]

In image compression we utilise three different mechanisms to compress pixel data:

(a) mapping the pixel values to some other set of values

(b) quantising those values

(c) symbol encoding the resulting values

Explain each mechanism, why it helps us to compress the image, and whether (giving reasons) the resulting image noticeably differs. [10 marks]

SECTION B

5 Compiler Construction

You have been given a new programming language with a C-like syntax, with integer variables and functions and with static binding of free variables. Your manager can parameterise certain aspects of the language, including the following three options:

- For "`int x = e;`" whether the variable `x` has the same l-value of `e` or whether a new l-value is created and initialised to the r-value of `e`. If `e` is only an r-value then a new l-value is created in both circumstances.

- For "`int f(int x) { ... }`" whether the variable `x` is passed by l-value ("by reference") or by r-value ("by value"). If the switch is set to "l-value" and the value passed is only an r-value then a new l-value is created, initialised and passed.

- For "`int f(int x) { ... y ... }`" (where the variable `y` is free to `f`) whether the value of `y` is calculated at the times of its uses (association by l-value) or at the time of the definition of `f` (association by r-value).

As a test of your programming skills your manager asks you to write a program which tells how the language has been parameterised. Do so by printing a 3-digit decimal number where the "hundreds" digit is one or two according to whether the first option is by l-value or r-value respectively, similarly with the "tens" digit for the second and with the "units" digit for the third option.

[10 marks]

Explain the structure of an object module which an assembler or compiler might produce to be processed by a linker. Your answer should include discussion of the various object module features needed to represent the compiled form of the C program:

```
int a[10] = { 2,3,5,7,11,13,17,19,23,29 };
extern int b[10];
extern int g(int);
int f(int y)
{   return g(y) + b[5] + a[6];
}
```

[10 marks]

6 Comparative Programming Languages

Many languages either forbid explicit pointer arithmetic or restrict its use. What kinds of problem are they seeking to avoid? Why is it allowed (with some restrictions) in C and C++, and commonly used by programmers? [8 marks]

What are the dangers inherent in allowing memory deallocation to be under the direct control of the programmer? Given these dangers, why has memory deallocation not been automated in standardised C or C++? [12 marks]

7 Prolog for Artificial Intelligence

According to the rules of the Billy Badger Fan Club, an applicant is acceptable for membership provided that:

- The applicant must have two proposers who are members of the club.
- The applicant must be aged between 18 and 30 years of age (inclusive).
- Each proposer must have been a member for at least two years.
- Each proposer must not be a parent of the applicant.

Write a Prolog program that includes a rule for deciding whether an applicant is acceptable for membership, illustrating with a sample database. [20 marks]

8 Databases

Explain what is meant by a *referential integrity constraint* in a relational data model. [4 marks]

The University of Cambridge is determined to maintain its standards under increasing financial pressure. The government maintains league tables of various kinds: teaching quality, research rating, unit cost of each student place. The university still enjoys a high reputation worldwide, but it is in competition with institutions such as MIT and Stanford whose unit costs are much greater. The only way to provide facilities such as new research laboratories, graduate accommodation, a much-needed swimming pool, is by public appeal. It is vital that there is close liaison between the university development offices and colleges so that prospective donors do not become alienated by simultaneous demands.

You have been invited by the university to assist in recording details of its fund-raising. The main purpose is to coordinate the activities of *agencies* such as the university and college development offices in their dealings with *prospects*. The latter may be alumni of some college, charitable foundations or major companies with an educational commitment; it is important to record their *interests* (bioscience, student welfare, sport) so that they can be approached in a favourable context; also their potential resources, in order to maximise the possible benefit to the university. At any time a number of *projects* need funding: each will be developed by a single agency, requiring that a target sum be raised by a given date; projects will have one or more *purposes* which may be linked with the interests of prospects. The other main use of the database is to keep a *diary* of interaction with prospects; in order to retain their goodwill it is essential to know who has been invited where and when, and in what context.

Design the schema for a relational database that will record this information. State clearly any assumptions that you need to make in order to complete the design, and indicate any difficulties that you foresee in maintaining the database.

[16 marks]

SECTION C

9 Foundations of Functional Programming

The binary trees, denoted by B, whose branch nodes contain natural numbers, are generated by the grammar

$$B ::= \mathsf{Leaf} \mid \mathsf{Br}\ (n, B, B)$$

where n ranges over natural numbers. Although this question concerns the encoding of binary trees as λ-terms, you may use the encodings of other well-known data structures, such as booleans and pairs, provided you state the properties assumed.

Give an encoding of binary trees as λ-terms by defining as λ-terms

(a) **Leaf** and **Br**, used to construct the λ-terms corresponding to binary trees;

(b) **isLeaf**, which tests whether a λ-term corresponds to a leaf or a branch node;

(c) **value**, **fstsubtree** and **sndsubtree**, used to identify respectively the natural number and the two subtrees at a branch node.

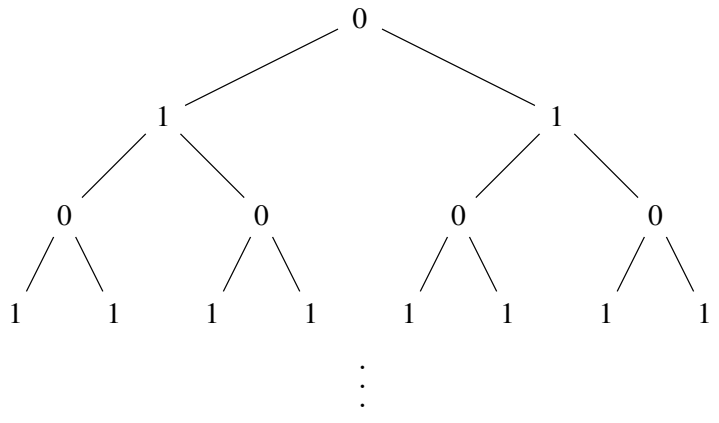
Justify your answer by describing the behaviour of **isLeaf**, **value**, **fstsubtree** and **sndsubtree**: for example, the reduction **isLeaf**(**Leaf**) $\twoheadrightarrow$ **true** describes part of the behaviour of **isLeaf**. [8 marks]

Consider the function $\mathsf{treeadd}$ defined inductively on the structure of binary trees by

$$\begin{aligned} \mathsf{treeadd}\ (m, \mathsf{Leaf}) &= \mathsf{Leaf} \\ \mathsf{treeadd}\ (m, \mathsf{Br}\ (n, B_1, B_2)) &= \mathsf{Br}\ (m+n, \mathsf{treeadd}\ (m, B_1), \mathsf{treeadd}\ (m, B_2)) \end{aligned}$$

Give and justify a λ-term which encodes $\mathsf{treeadd}$, using the λ-term $Y \equiv \lambda f.(\lambda x.f(xx))(\lambda x.f(xx))$. [6 marks]

Give the λ-term for the infinite binary tree whose branch nodes consist of zeros at even depths and ones at odd depths, as pictured below:



[6 marks]

10 Logic and Proof

Outline the concepts behind modal logic. Illustrate your answer by explaining the meaning of the axiom $A \rightarrow \Box\Diamond A$. [6 marks]

Attempt to prove, using rules for S4, the following sequents:

$$\Box(A \vee B) \Rightarrow (\Box A) \vee (\Box B)$$
$$(\Box A) \vee (\Box B) \Rightarrow \Box(A \vee B)$$

For each sequent, briefly explain why it is valid (or is not, as the case may be). [4 + 4 marks]

Use resolution to derive a contradiction from this set of clauses:

$$\{\neg P(x,x), P(x,a)\} \quad \{P(x,x), P(x,f(a))\} \quad \{\neg P(y,f(x)), \neg P(y,x)\}$$

[6 marks]

11 Complexity Theory

Here are some informally expressed opinions about computational complexity. They may be correct, incorrect, misleading or meaningless. In some cases the truth or otherwise of the statement might not be known, either in the sense of it not having been covered in the course or by the answer not being known by anybody anywhere. For each statement comment on its validity and in cases where that is both necessary and straightforward produce an adjusted version of the observation that is properly valid. You are not expected to include proofs to support your claims.

(a) Problems that are not NP-complete are easy to solve. [3 marks]

(b) Problems that are NP-complete will never be solved in reasonable amounts of time even though computers continue to get faster and faster. [3 marks]

(c) To test a number N to see whether it is prime you just have to do a test-division by each of the numbers from 2 to $N-1$, and since there are only $N-2$ of these and division can be done in time $O(n^2)$ this is polynomial time. Thus primality testing is in the class P. [4 marks]

(d) There is a polynomial-time reduction from the k-clique problem to 3-SAT. [3 marks]

(e) There is a polynomial-time reduction from 3-SAT to the k-clique problem. [3 marks]

(f) There have been proposals that biological computers based on DNA might use the massive parallelism of their biochemical activity to solve NP problems rapidly. If such systems could be made to work reliably this would solve the theoretical challenge posed by the concept of NP-completeness. [4 marks]

12 Semantics of Programming Languages

What is meant by a *labelled transition system*? [2 marks]

A language of commands, C, for interactive input/output is given by

$$\begin{aligned} C ::= {} & \mathbf{skip} \mid \ell := \ell \mid \mathbf{getc}(\ell) \mid \mathbf{putc}(\ell) \mid C \,;\, C \\ \mid {} & \mathbf{if}\ \ell = \ell\ \mathbf{then}\ C\ \mathbf{else}\ C \mid \mathbf{while}\ \ell = \ell\ \mathbf{do}\ C \end{aligned}$$

where ℓ ranges over some fixed set of locations for storing characters. The command $\ell_1 := \ell_2$ copies the contents of ℓ_2 to ℓ_1. The command $\mathbf{getc}(\ell)$ reads the next character from the standard input stream into ℓ. The command $\mathbf{putc}(\ell)$ writes the contents of ℓ to the standard output stream. The conditional and while-loop commands involve testing whether or not the contents of two locations are equal. The commands **skip** and $C_1 \,;\, C_2$ have their usual meanings. Define an operational semantics for this language as a labelled transition system whose configurations are (command, state)-pairs, $\langle C, s\rangle$, and whose actions are of the form $get(c)$ (for reading a character c), $put(c)$ (for writing a character c), and τ (for transitions not involving input/output). [9 marks]

Explain, without proof, in what sense this labelled transition system is deterministic. [2 marks]

A finite list t of non-τ actions is a *trace* of $\langle C, s\rangle$ if there is a sequence of labelled transitions starting from $\langle C, s\rangle$ whose corresponding list of actions is equal to t once any τ-actions have been erased from it. Write $C \approx_{tr} C'$ to mean that for any state s, the configurations $\langle C, s\rangle$ and $\langle C', s\rangle$ have the same traces. Show that $C_1 \approx_{tr} C_2$, where

$$C_1 \stackrel{\text{def}}{=} \mathbf{getc}(\ell_1) \,;\, \mathbf{putc}(\ell_1) \quad \text{and} \quad C_2 \stackrel{\text{def}}{=} \mathbf{getc}(\ell_1) \,;\, \ell_2 := \ell_1 \,;\, \mathbf{putc}(\ell_2).$$

[3 marks]

By considering $C_1 \,;\, C$ and $C_2 \,;\, C$ for a suitable C, or otherwise, show that $\approx_{tr}$ does not have the congruence property for this language. [4 marks]