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

Thursday 8 June 1995 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.*
Write on **one** *side of the paper only.*

1 Processor Architecture

What is a *data cache* and what properties of data access does it exploit? [10 marks]

Why are caches necessary for RISC processors? [5 marks]

A translation lookaside buffer (TLB) is a specialised cache. What does it typically store? [5 marks]

2 Computer Architecture

Write short notes on the following:

(*a*) static memory bit [6 marks]

(*b*) dynamic memory bit [6 marks]

(*c*) memory chip organisation [4 marks]

(*d*) memory module organisation [4 marks]

3 Digital Communication I

What is the purpose of a signalling system in a digital telephone network? Describe the operations that the signalling system should perform. [6 marks]

Describe a possible implementation of a signalling system, including a discussion of how signalling information might be carried over the network. [8 marks]

What are the advantages/disadvantages of having a separate network for signalling traffic? [3 marks]

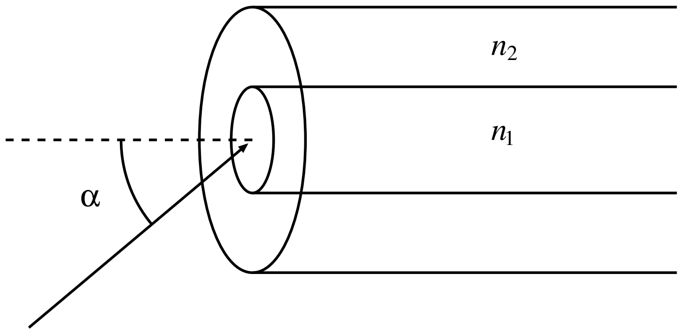
To what extent is a signalling system different from a general-purpose distributed computation? [3 marks]

4 Graphics

In ray tracing a large computational cost is associated with determining ray–object intersections. Explain how the use of bounding volumes and space subdivision methods may reduce this cost. [20 marks]

5 Developments in Technology

Light is incident from air on the end face of a multimode optical fibre at angle of incidence α as shown below.



The refractive indices of the core and cladding are n_1 and n_2 respectively, where $\mid n_1 - n_2 \mid \ll 1$. Prove the following condition for the incident light to be guided by the fibre

$$\sin \alpha \leqslant \sqrt{n_1^2 - n_2^2}$$

[4 marks]

What is the main cause of pulse spreading in a step index multimode fibre?

[3 marks]

Show, for the same fibre parameters as given above, that the bandwidth × length product for the fibre is given approximately by

$$B.L = \frac{nc}{(n_1^2 - n_2^2)}$$

where c is the speed of light in free space, and the approximation $n_1 \approx n_2 \approx n$ has been made. [5 marks]

Explain carefully how, by appropriate design of the multimode fibre, the bandwidth might be increased. [3 marks]

Explain carefully what is meant by the term *material dispersion* in an optical fibre, explaining its relative importance for single and multimode fibres. How might it be minimised? [5 marks]

6 Programming in C and C++

Write brief notes on *four* of the following aspects of the language C++. In some of the cases it may be appropriate to compare what C++ does with the situation in other programming languages such as C, ML or Modula-3.

(*a*) Overloaded functions and operators.

(*b*) Templates as a way of achieving operations on lists where the types of the items in the lists must be kept flexible.

(*c*) Consistency checks when a program is kept in several files and these are compiled at different times.

(*d*) Inheritance and virtual functions.

(*e*) The problems of writing code that is portable from one host machine to another.

[5 marks each]

7 Compiler Construction

Compare two possible mechanisms that can be used to allow efficient execution time access to free variables of functions in a programming language that permits functions to be defined within other functions. [12 marks]

Discuss whether more efficient implementations are possible if the language permits no recursive function calls, while still allowing functions to be defined within other functions. [8 marks]

8 Prolog for Artificial Intelligence

Consider the task of normalising sum expressions. For example, the sums $(a + b) + (c + d)$ and $(a + (b + (c + d)))$ may be normalised into a standard form that is left associative: $a+b+c+d$ or equivalently $((a+b)+c)+d$. Write a Prolog procedure to define predicate `normsum` such that the goal `normsum(X,Y)` succeeds when the sum expression `X` normalises to `Y`. Procedures not using the technique of difference structures will not receive full marks. [20 marks]

9 Databases

Following a directive issued by the Inquisition, the University is to set up a database that will keep track of the way in which undergraduates are supervised.

At the end of each year this database will be consulted so that reports can be prepared to show the percentage of supervisions skipped (broken down by excuse, if any, proffered).

Students' exam grades will be correlated against their previous academic record, the number of supervisions they attended and the identity of their supervisors, so that pressure can be applied to both classes of participant.

The scheme will cover supervisions individually and in groups of two, three or four. The authorities will want to be able to determine whether group size, continuity of contact with one particular supervisor or College affiliation influence the outcome.

The same database is also to be used to help coordinate and regularise the way in which research students are used as supervisors, so they will be invited to record what subjects or year-groups they feel able to cope with and how much supervising they will undertake (which amount may vary from term to term). Note that only around 50% of all supervisions are given by research students — others are given by post-doctoral research workers, the teaching staff or people from outside the department.

Design a schema for a relational database that is to record the relevant information. You should state any assumptions that lie behind the schema design that you present.

[20 marks]

10 Natural Language Processing

Imagine you were to construct a natural language processing system for solving syllogisms expressed in English, behaving approximately as indicated:

User:	All men are mortal.	System:	OK
User:	Socrates is a man.	System:	OK
User:	Is Socrates mortal?	System:	Yes

Describe the components of such a system and how they would fit together.

[20 marks]

11 Computer Systems Modelling

In queueing networks, what is meant by a *closed* system? [4 marks]

Consider two closed systems. One has two devices, A and B, and three customers, the other three devices, A, B and C, and two customers. Both have exponentially distributed service times which are device dependent but customer independent. In the first system a customer completing service at a device always moves to the other device. In the second system a customer completing service moves to one of the other two devices with equal probability.

Draw state diagrams for the Markov chains representing these systems. Choose one system to solve for device utilisation in terms of service rates. [10 marks]

For the chosen system, when the service rates are equal does the utilisation of each device correspond to that for a balanced system ($U = \frac{N}{N+K-1}$ where N is the number of customers and K the number of devices)? [3 marks]

Describe the state space for a Markov chain for one of the systems if the service rates were both customer and server dependent. [3 marks]

12 Introduction to Functional Programming

Here is the specification of a dictionary package, based on a type `T`, whose values are data structures that represent finite maps from keys of type `string` to values of type `int`.

```
exception Dict
empty : T
lookup : T * string -> int
update : T * string * int -> T
delete : T * string -> T
merge : T * T -> T
```

Value `empty` represents the empty map. Expression `lookup`(t, a) returns the value bound to the key a in t, if one exists; otherwise it raises the exception `Dict`. Expression `update`(t, a, i) returns a map that binds key a to the value i, and on other keys acts the same as t. Expression `delete`(t, a) returns the map that is the same as t except that key a is unbound. Expression `merge`(t_1, t_2) is a map that binds a key a if it is bound in t_1 or t_2. If it binds a, it binds a to the value given by t_1, if it exists, otherwise to the value given by t_2.

We may represent dictionaries using lists, where `T` is `(string * int) list`. Using this representation, write ML definitions for the values `empty`, `lookup`, `update`, `delete` and `merge`. [8 marks]

Repeat using a functional representation where `T` is `string -> int`. [8 marks]

Discuss the performance of `lookup` in your two implementations. Briefly explain how to obtain a better performance by changing the definition of `T`. [4 marks]

13 Complexity Theory

It turns out that the following program, when run using floating-point arithmetic that remains accurate to N decimal places, will compute and print the value of π correct to almost N places. The loop (which has as its main part a step which replaces the values in `a` and `b` by their arithmetic and geometric means, respectively) will be traversed about $\log(N)$ times.

```
a := 1;
b := 1/sqrt(2);
u := 1/4;
x := 1;
pn := 4;
do { p := pn;
     y := a; a := (a+b)/2; b := sqrt(y*b);
     u := u-x*(a-y)*(a-y);
     x := 2*x;
     pn := a**2/u; } while (pn<p);
print(p);
```

You are provided with procedures that can compute Fourier Transforms and their inverses with a transform on k points (using floating-point arithmetic), taking time proportional to $k \log k$. Explain how you could implement the high-precision arithmetic needed to make this program run fast. Do not discuss how the Fourier transform will be implemented — just how it is used, and assume that the floating-point accuracy achieved by the transform will be adequate for your purposes.

[14 marks]

Overall how long (as a function of N) would you expect the complete program to take to run? [6 marks]

You do not need to understand how or why this particular calculation arrives at a value for π, or why the loop is executed only $\log(N)$ times.

14 Numerical Analysis II

With reference to solution of the differential equation $y' = f(x, y)$, explain the conventional notation x_n, $y(x_n)$, y_n, f_n. [4 marks]

Derive Euler's method

$$y_{n+1} = y_n + hf(x_n, y_n). \tag{1}$$

[3 marks]

Euler's method has *local error*

$$\frac{h^2}{2} y''(\xi).$$

Explain the terms *local error*, *global error*. [2 marks]

The multistep formula

$$y_{n+1} = y_{n-3} + \frac{4h}{3}(2f_n - f_{n-1} + 2f_{n-2}) \tag{2}$$

has *local error*

$$\frac{14}{45} h^5 y^{(5)}(\xi).$$

Outline the technique for deriving multistep formulae such as (2). (Omit algebraic details.) [2 marks]

Suppose Euler's formula is used as a starting procedure for formula (2). How many initial steps of formula (2) need to be evaluated using Euler? [2 marks]

Estimate very roughly the number N of Euler steps needed to approximate f_1. (Assume that $| y^{(5)}(x) | \simeq 30$, and Euler's method has *global error* h/N.) [5 marks]

What is the most important requirement for a starting procedure? Suggest a more suitable starting procedure than Euler. [2 marks]