COMPUTER SCIENCE TRIPOS Part II

Wednesday 4 June 2003 1.30 to 4.30

Paper 8

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

1 Comparative Architectures

The data sheet for a new workstation CPU tells you simply that it has a 8KB L1 data cache and 512KB L2 cache. As an inquisitive computer scientist, you wish to learn more about the system's memory hierarchy.

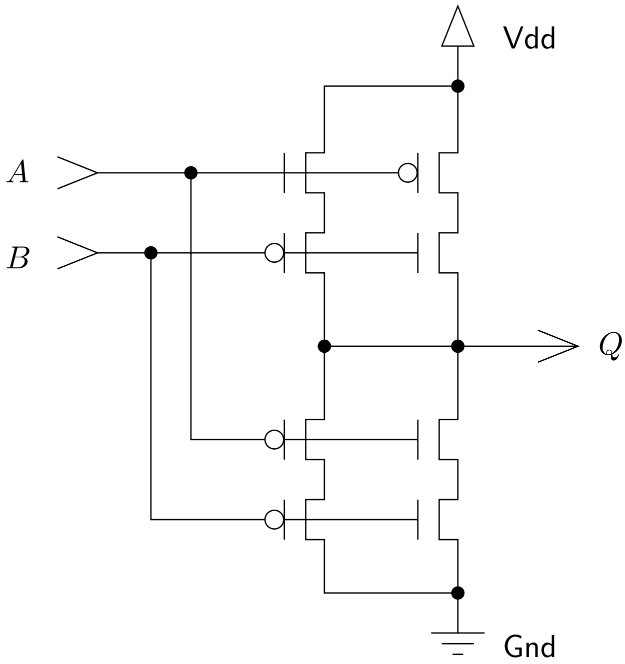
(*a*) Devise a method for determining the cache line size. You may assume that the L1 and L2 caches use the same line size, and that the operating system provides a microsecond accurate time function. Provide high-level language pseudo code of any test programs you would use, and describe how you would interpret their output. [6 marks]

(*b*) Describe how you produce an accurate estimate for the load latency incurred by accesses to the two caches and main memory. [6 marks]

(*c*) Outline a method you could use to determine the associativity of the caches. [8 marks]

2 VLSI Design

The designer of a CMOS standard cell library has proposed the following circuit for an XOR gate:



Explain how the circuit operates. [2 marks]

Although the circuit appears digitally correct, it operates slowly. Moreover, when two of the XOR gates are cascaded to calculate the XOR of three signals, the results are unreliable. Suggest explanations for these two problems. [4 marks]

Propose a modification to the circuit which retains the general design but introduces two pairs of additional inverters to address the problems. [4 marks]

Another designer cascades two of the revised gates as part of a full adder, and observes that the resulting circuit is large and slow. Extend the design to provide the XOR of three inputs directly. Compare its size and speed with the cascaded version. [5 marks]

Yet another designer is working on an asynchronous design and needs a two-input XOR gate for dual-rail logic. Sketch a design which ensures that all the input signals become valid before any output does so, and that all the inputs become clear before any outputs do so. [5 marks]

3 Digital Communication II

Two hosts are connected to two separate 100 Mbps local area networks. One is a user's machine with a web browser, the other is a web server. The networks are connected via a pair of routers and a long haul, point-to-point link with the following characteristics: 640 kbps, 50 ms delay, and a Maximum Packet Size of 1 k bytes. The routers have 8 kbytes of packet buffer memory.

(*a*) Assuming standard TCP behaviour (slow start, congestion control, and fast retransmit/recovery), approximately how long does it take to download a file of 8 Kbytes of data? You may assume that packet headers and store-and-forward times are insignificant. You may find it easier to illustrate your answer with a time-sequence diagram of the packet exchanges. [5 marks]

(*b*) How long, approximately, does it take to download a file of 8 Mbytes from one computer to the other using TCP? [5 marks]

(*c*) Illustrate your answer to the previous section with a diagram showing TCP's behaviour over time, indicating the two important stages and transitions between them. [5 marks]

(*d*) What would happen qualitatively if the routers had only 4 Kbytes of packet buffer memory? [5 marks]

4 Distributed Systems

(*a*) You are to design a component of a distributed system which takes action on the arrival of an alarm event from another component.

Discuss the design issues, relating to the characteristics of distributed systems, of the component and its communication. [4 marks]

(*b*) You are to design a service that takes in streams of messages from distributed sources and notifies its clients when certain specified patterns of messages occur.

Discuss the design issues associated with supporting the following operators for constructing message patterns, where A and B represent messages:

(*i*) A OR B;

(*ii*) A AND B (unordered pairs);

(*iii*) A BEFORE B (ordered pairs).

[16 marks]

5 Advanced Systems Topics

(*a*) Considerable recent operating system research has focused on *extensibility.*

(*i*) What is the motivation for this work? [1 mark]

(*ii*) Describe with the aid of examples *three* different techniques for providing extensibility. In each case comment on the granularity of extensibility provided, and discuss the issues of performance and safety.

[5 marks each]

(*b*) Some researchers claim that programming language virtual machines such as the JVM should be extensible.

(*i*) Why might this be desirable? [1 mark]

(*ii*) How could this be safely achieved? Comment on how your solution might differ from those used in the case of operating system extensibility.

[3 marks]

6 Security

(*a*) The Digital Signature Standard is computed using the following equations:

$$r = (g^k \bmod p) \pmod q$$
$$s = (h(M) - xr)/k \pmod q$$

Describe what the various symbols represent. [4 marks]

(*b*) Write down the equation(s) used to verify a signature. [4 marks]

(*c*) The standard specifies that r must lie strictly between 0 and q. What might go wrong if an implementation does not check this? [4 marks]

(*d*) A designer decides to economise on code size by omitting the hash function computation, that is, replacing $h(M)$ by M. What are the consequences of this optimisation? [8 marks]

7 Optimising Compilers

(*a*) Explain in one sentence what is meant by the "phase order problem". [2 marks]

For the rest of the question, it is recommended that you restrict attention to a single basic block containing only unary and binary arithmetic instructions, e.g. `add#4 dst,src` or `mul dst,src1,src2` and with no variable written to after being read (SSA form). Consider all such instructions to execute in one cycle.

(*b*) Describe how a directed acyclic graph expressing instruction dependencies suitable for instruction scheduling can be obtained from such a basic block. [4 marks]

(*c*) Briefly describe how a (register interference) graph can be obtained from such a basic block. Also state a requirement for colouring this graph with registers. [4 marks]

(*d*) Consider the two programs (where the first four lines are the same):

```
add   t1,a,b
add#2 t2,c
add#3 t3,c
mul   t4,t2,t3
sub   z,t1,t4
```

```
add   t1,a,b
add#2 t2,c
add#3 t3,c
mul   t4,t2,t3
sub   t5,t1,t4
xor   t6,t5,a
xor   z,t6,b
```

In each case consider scheduling the `add t1,a,b` to appear as early as possible or as late as possible. Determine the number of registers required to colour the program in all four resulting cases. (Assume that only `z`, allocated to `r1`, is live on exit and that `a,b,c` are allocated registers `r1,r2,r3`.) [10 marks]

[Remark: you might note that doing `add t1,a,b` in the left-hand program decreases the number of live registers by one, while it increases the number by one on the right-hand program.]

8 Artificial Intelligence II

(*a*) Describe the *STRIPS* language for representing states, goals, and operators within a planning system. [5 marks]

(*b*) Give a definition of a *plan*, a *consistent plan*, and a *complete plan*. [5 marks]

(*c*) Describe the *initial plan* used as a starting point by the partial-order planning algorithm. [5 marks]

(*d*) Outline the way in which the partial-order planning algorithm constructs a plan beginning with the initial plan. Include in your answer a description of a *threat* along with an explanation of how the algorithm can attempt to remove threats by promotion or demotion. [5 marks]

9 Database Theory

(*a*) Describe the key features of the ODMG object model. [6 marks]

(*b*) Describe the key features of the semi-structured data model. [6 marks]

(*c*) Consider the following DTD.

```
<!ELEMENT class   (student)*>
<!ELEMENT student (name, college, course*, (phone|email))>
<!ELEMENT name    (#PCDATA)>
<!ELEMENT college (#PCDATA)>
<!ELEMENT course  (title, year)>
<!ELEMENT title   (#PCDATA)>
<!ELEMENT year    (#PCDATA)>
<!ELEMENT phone   (#PCDATA)>
<!ELEMENT email   (#PCDATA)>
```

(*i*) Detail how XML data satisfying such a DTD could be stored in a traditional (non-complex value) relational DBMS. Comment on the efficiency of your strategy. [5 marks]

(*ii*) Detail how XML data satisfying such a DTD could be stored in an ODMG-compliant DBMS. [3 marks]

10 Information Theory and Coding

(*a*) An alphabet has six symbols with the following probabilities of occurrence:

a	b	c	d	e	f
0.31	0.18	0.21	0.08	0.17	0.05

(*i*) Construct an optimal prefix code (an Huffman code) for this alphabet. [5 marks]

(*ii*) What is the average number of bits per symbol in your code? Is this greater than, less than, or equal to the entropy of the alphabet? [3 marks]

(*b*) Jane's Furniture Store sells 2-, 3- and 4-seater sofas in two styles: classic and modern. At the end of May, Jane tallies the number of each type that has been sold in each style during May.

Style	Type (no. of seats) 2	3	4
Classic	8	24	4
Modern	8	8	12

(*i*) Let X be the type of sofa (2-, 3-, or 4-seats) and Y be the style (classic or modern). Calculate the values of $H(X)$, $H(Y)$, $H(X,Y)$, $H(X|Y)$, $H(Y|X)$, and $I(X;Y)$. [9 marks]

(*ii*) In part (*b*)(*i*), you should have found that $I(X;Y) \neq 0$. For the month of June, Jane wants to have $I(X;Y) = 0$ while still selling the same number of sofas of each type as in May. Change the numbers in the table so that the same number of sofas of each type are sold (i.e. the column totals remain unchanged) but so that $I(X;Y) = 0$. What is $H(X,Y)$ in this revised table? [3 marks]

11 Numerical Analysis II

(a) Let n_+ be the number of positive real roots of a polynomial $p_n(x)$. Let c be the number of changes of sign when the coefficients are taken in order. State *Descartes' rule of signs.* [2 marks]

(b) If $p_3(x) = x^3 + 13x^2 + 54x + 72$ what does the rule say about the polynomials $p_3(x)$, $p_3(-x)$? [2 marks]

(c) Using Descartes' rule, a simple search technique and factorisation, find all the real and complex roots of

$$q_5(x) = x^5 + 5x^4 + 32x^3 + 160x^2 + 256x + 1280.$$

[7 marks]

(d) The Newton–Raphson formula $\tilde{x} = x - f(x)/f'(x)$ can often be used to find real roots of a polynomial. However, a user of numerical software reports failure to find an accurate root of

$$f(x) = 3x^4 - 28x^3 + 24x^2 + 144x + 432$$

for $x > 3$ using Newton–Raphson. The user has tried different starting values but gets either a floating-point exception or failure to converge. Using Descartes' rule on $f(x)$ and $f'(x)$ investigate the cause of these numerical problems. [9 marks]

12 Specification and Verification I

(a) Define the specification $[P, Q]$ as used in program refinement. [2 marks]

(b) Devise refinement rules for `FOR`-commands. [8 marks]

(c) Show how your rule can be justified using Floyd–Hoare logic. [4 marks]

(d) Use your rule to show that

`[SUM=0 ∧ 1≤M, SUM = M×N] ⊇ FOR I := 1 UNTIL M DO SUM := SUM+N`

[6 marks]

13 Natural Language Processing

(*a*) Define the following terms in morphology:

(*i*) morpheme

(*ii*) affix

[4 marks]

(*b*) In English morphology, 'y' maps to 'ie' when preceded by a consonant and followed by the affix 's'. Give a finite state transducer that implements this spelling rule, explaining the notation that you use. Your transducer should accept the following pairings:
party/party, parties/party ˆ s, partying/party ˆ ing

It should reject:
partys/party ˆ s, toies/toy ˆ s

[12 marks]

(*c*) The 'y' to 'ie' mapping also applies when the affix is 'ed'. Briefly discuss how this might be handled, taking into account that the morphology system should accept partied/party ˆ ed and not partieed/party ˆ ed. [4 marks]

14 Denotational Semantics

(*a*) Explain the meanings of *soundness*, *adequacy* and *full abstraction* for the operational and denotational semantics of PCF. Why does the denotational semantics of PCF fail to be fully abstract? [10 marks]

(*b*) State Scott's principle of fixed-point induction. [2 marks]

(*c*) Let D and E be domains, and $h : D \to E$ a strict continuous function, i.e. such that $h(\bot) = \bot$.

Show that the property $h(x) = y$, in $x \in D$ and $y \in E$, is admissible.

[3 marks]

(*d*) Let $f : D \to D$ and $g : E \to E$ be continuous functions such that $g \circ h = h \circ f$. Prove

$$h(\mathit{fix}(f)) = \mathit{fix}(g)$$

by fixed-point induction. [5 marks]

15 Computer Systems Modelling

(*a*) Describe the congruential class of methods for generating pseudo-random numbers from a uniform (0,1) distribution. [3 marks]

(*b*) Let U be a uniform $(0,1)$ random variable. Show that for any continuous probability distribution function $F(x)$ the random variable, X, defined by

$$X = F^{-1}(U)$$

has the probability distribution function $F(x)$. [3 marks]

(*c*) Suppose that $X_1, X_2, \ldots, X_n$ are independent identically distributed random variables with mean μ and variance σ^2. Define the sample mean, $\overline{X}$, and sample variance, S^2. [2 marks]

(*d*) Calculate the expectation and variance of $\overline{X}$. [4 marks]

(*e*) Use the central limit theorem to derive an approximate $100(1-\alpha)$ percent confidence interval for estimating μ. [6 marks]

(*f*) Describe an algorithm to control the length of a simulation run such that the estimate obtained has an approximate $100(1-\alpha)$ confidence interval of length at most ℓ. [2 marks]

END OF PAPER