COMPUTER SCIENCE TRIPOS Part II

Wednesday 2 June 2010 1.30 to 4.30

COMPUTER SCIENCE 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

STATIONERY REQUIREMENTS
Script paper
Blue cover sheets
Tags

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Advanced Graphics

(*a*) Discrete curvature

(*i*) Give each of the following: the *Gaussian curvature* at the exact centre of any face of a cube; the *Gaussian curvature* at any corner of a cube; and the *angle deficit* at any corner of a cube. [1 mark]

(*ii*) Sketch a picture of a closed manifold surface with total angle deficit -4π. The picture must be intelligible but you will not otherwise be marked on artistic skill. [2 marks]

(*iii*) If your hypothetical surface had 20 vertices and 20 faces then how many edges must it have? [2 marks]

(*b*) The convex hull

(*i*) In no more than ten sentences and/or half a page of pseudocode, describe a method for finding the convex hull of a set of n points in 3D. For full marks, give an algorithm that runs in $O(n^2)$ time or faster; partial marks will be given to any slower solution. You must give enough detail that a programmer with no knowledge of computational geometry could implement your algorithm. [4 marks]

(*ii*) Give the running time of your algorithm in big-O notation. [1 mark]

(*c*) Global illumination

(*i*) In no more than six sentences, describe **either** radiosity rendering **or** photon mapping. [3 marks]

(*ii*) In no more than six sentences, compare your chosen method with the other one. [2 marks]

(*iii*) Which of these two is an example of a *Monte Carlo* algorithm? [1 mark]

(*d*) Ray tracing

A perfectly reflective mirrored sphere, S, is centred at the origin $(0, 0, 0)$. Directly above it is a bright red $2 \times 2 \times 2$ cube, C, centred at $(0, 5, 0)$. The default background colour of the scene is blue. A ray-tracing ray R is fired from $(0, 1, 10)$ with direction $(0, 0, -1)$. The scene is lit by an ambient light source and there are no other objects in the scene. What is the maximum radius of S such that the colour calculated for R is red? Full marks for the correct answer; partial marks if you answer incorrectly but your work justifies your response. [4 marks]

2 Artificial Intelligence II

Consider the following learning problem in which we wish to classify inputs, each consisting of a single real number, into one of two possible classes C_1 and C_2. There are three potential hypotheses where $\Pr(h_1) = 3/10$, $\Pr(h_2) = 5/10$ and $\Pr(h_3) = 2/10$. The hypotheses are the following functions

$$h_i(x) = x - \frac{i-1}{5}$$

and the likelihood for any hypothesis h_i is

$$\Pr(x \in C_1 | h_i, x) = \sigma(h_i(x))$$

where $\sigma(y) = 1/(1 + \exp(-y))$. You have seen three examples: $(0.9, C_1)$, $(0.95, C_2)$ and $(1.3, C_2)$, and you now wish to classify the new point $x = 1.1$.

(a) Explain how in general the *maximum a posteriori (MAP)* classifier works. [3 marks]

(b) Compute the class that the MAP classifier would predict in this case. [10 marks]

(c) The preferred alternative to the MAP classifier is the Bayesian classifier, computing $\Pr(x \in C_1 | x, \mathbf{s})$. where $\mathbf{s}$ is the vector of examples. Show that

$$\Pr(x \in C_1 | x, \mathbf{s}) = \sum_{h_i} \Pr(x \in C_1 | h_i, x) \Pr(h_i | \mathbf{s})$$

What are you assuming about independence in deriving this result? [3 marks]

(d) Compute the class that the Bayesian classifier would predict in this case. [4 marks]

3 Digital Communication II

(*a*) Work conserving packet switches follow the law that for a given packet arrival rate λ_n, and packet service rate μ_n, the sum

$$\sum_{n=1}^{N} \rho_n q_n = C$$

where C is a constant and $\rho_n = \lambda_n \mu_n$.

(*i*) Explain what trade-off is represented in this equation, perhaps using a simple numerical example. [8 marks]

(*ii*) What is the key rationale for work conserving packet switches? [2 marks]

(*b*) The Transmission Control Protocol (TCP) uses a congestion control scheme which adjusts a congestion window in response to congestion feedback from the network. Explain how this can lead to a long-term steady-state rate of approximately:

$$\text{rate} \propto \text{MSS}/(\text{RTT} \times \sqrt{p})$$

where MSS is the Maximum Segment Size, and RTT is the Round Trip Time, and p is the mean packet loss rate. You may find it helpful to include a diagram of the evolution of the congestion window over time. [8 marks]

(*c*) Explicit Congestion Notification (ECN) can be used to replace *packet loss* as a signal of congestion. What is the key advantage of doing this? [2 marks]

4 Computer Vision

(*a*) Explain how each of the following equations or expressions can be used for detecting and estimating visual motion in a spatio-temporal image sequence $I(x, y, t)$. Include in each answer also the name used to describe each of these general classes of motion extraction models:

(*i*) $$-\frac{\partial I(x,y,t)}{\partial t} = \vec{v} \cdot \vec{\nabla} I(x,y,t)$$ [3 marks]

(*ii*) $$-\frac{\partial}{\partial t}\left[\nabla^2 G_\sigma(x,y) * I(x,y,t)\right]$$ [3 marks]

(*iii*) $$\text{argmax} \int_x \int_y \int_t I(x,y,t) \cdot I(x - v_x\tau,\ y - v_y\tau,\ t - \tau)\ dx\ dy\ dt$$ [3 marks]

(*iv*) $$F(\omega_x, \omega_y, \omega_t) = e^{-i(\omega_x v_x \tau\ +\ \omega_y v_y \tau\ +\ \omega_t \tau)} F(\omega_x, \omega_y, \omega_t)$$
where $$F(\omega_x, \omega_y, \omega_t) = \int_x \int_y \int_t I(x,y,t) e^{-i(\omega_x x + \omega_y y + \omega_t t)}\ dx\ dy\ dt$$ [3 marks]

(*b*) Explain the key principles underlying the Scale-Invariant Feature Transform (SIFT). What is it used for, what goals does it achieve, and how does it achieve them? [5 marks]

(*c*) Suppose you were trying to design a machine vision system based as closely as possible upon human vision; for example, perhaps a visual prosthesis for the blind. Would you aim to include, as design goals, the standard geometric visual illusions as well? If such errors of vision emerged as unintended consequences of your design, would you consider them to be features, or bugs, and why? [3 marks]

5 Comparative Architectures

(*a*) Why is it that clock frequency improvements achieved from increasing the length of a processor's pipeline do not directly improve overall performance? [4 marks]

(*b*) Why might it be advantageous to design a scalar processor with multiple execution pipelines? [5 marks]

(*c*) What advantages does dynamic scheduling offer when compared with static scheduling in a superscalar processor? [5 marks]

(*d*) What has limited our ability to improve the performance of superscalar processors even as fabrication technologies have scaled? [6 marks]

6 Digital Signal Processing

The *Purpletoe* standard for trouser-area networking uses a radio signal with a bandwidth of less than 1 MHz. The carrier frequency is $f_c(k) = (2400 + 2k)$ MHz, where $k \in \{1, 2, 3, \ldots, 40\}$ is the channel number. Consider a receiver design in which the antenna signal is first multiplied with a sine wave of *fixed* frequency f_m, is then band-pass filtered to eliminate frequencies outside the range 1 MHz to 100 MHz, and is finally sampled by an analogue-to-digital converter with sampling frequency f_s for further digital processing.

(*a*) What is the largest set of frequencies from which f_m can be chosen such that no information is lost from any of the 40 channels? [4 marks]

(*b*) Which of the combinations of f_m and f_s that preserve all information from all 40 channels in the sampled output has the lowest sampling frequency f_s, assuming there is no signal outside these channels? [4 marks]

(*c*) To make eavesdropping more difficult, *Purpletoe* transmitters hop several times each second from one channel to another, in a secret pseudo-random order that is cryptographically pre-agreed and shared only with intended receivers. Consider for your receiver a special eavesdropping mode that exploits aliasing such that transmissions of a data packet using different channel numbers k all look the same after sampling (assuming that there is only a single transmitter in range). Which combination of f_s and f_m achieves that, and how? [8 marks]

(*d*) Cost pressures force you to use a cheaper circuit that multiplies the radio signal with a square wave of frequency f_m, instead of a sine wave. How does this affect the design of your receiver? [4 marks]

7 Distributed Systems

(*a*) The goal is to design a naming system that, along with a storage service, allows authorisation software to retrieve role-based access control data.

(*i*) Suppose that the authorisation software is presented with an authenticated principal. To make its authorisation decision, what information will it need to retrieve from storage and therefore what sorts of objects should be in the naming system? [3 marks]

(*ii*) Should names be pure or impure? Justify your claim. [2 marks]

(*iii*) How might impure names be structured? [2 marks]

(*iv*) It is likely that the authorisation software will need to understand principals from several organisations. How should names reflect this? [2 marks]

(*b*) Suppose that it is early 2005 and you are designing a web site allowing registered users to upload videos and anybody to download and watch them. You expect and hope that the site will be popular, so you use replication to handle the anticipated request rate. (Assume that, unlike a certain mainstream site whose name you can probably guess, no comments are associated with the videos.)

(*i*) What are strong and weak consistency? Describe the advantages and disadvantages of using each for this application, paying particular attention to any assumptions you must make about the users' behaviour. [4 marks]

(*ii*) You are asked to modify your site so that it can be used to store and annotate video evidence for law enforcement. This requires using strong consistency; why? [2 marks]

(*iii*) Suppose that you use nine replicas to meet demand and quorums to ensure strong consistency. What are the constraints on the sizes of read and write quorums? Give an example of quorum processing involving a write, two reads, then a write. Use the notation v_i to indicate version i of the data and suppose that at the start of your example, each replica contains version 3. Make sure that your diagram is easy to understand and that you explain any notation that you use. [5 marks]

8 E-Commerce

(*a*) Discuss possible business models for a social networking site. [5 marks]

(*b*) Explain what is meant by "network externality" for such a site. [5 marks]

(*c*) A music publisher complains to you as owner of a social networking site that a user has posted copyright material to your site, and threatens to take an action against you as effective publisher unless you reveal the name and address of the user. What is your defence, if any? [5 marks]

(*d*) The police contact you concerning a post on your site and demand the name and address of the originator. Under what conditions must you reveal the information? [5 marks]

9 Information Retrieval

(*a*) The vector space model of document retrieval uses the notion of an *information space*.

(*i*) What are the typical basis vectors for this space? What are the vectors in the space intended to represent? [2 marks]

(*ii*) How is *term weighting* used to position objects in the space? Why is term weighting important for effective document retrieval? [3 marks]

(*iii*) What is the *orthogonality assumption* typically employed in vector space models of document retrieval? Why might this assumption be false, and why might it lead to errors in retrieval? [3 marks]

(*iv*) Suggest one way in which the orthogonality assumption could be relaxed. (Just a short high-level description of a possible method is required.) [2 marks]

(*b*) The PageRank algorithm uses a model of a "random surfer" to calculate the validity of a page.

(*i*) Explain how the random surfer moves about the web. [2 marks]

(*ii*) Describe how the random surfer can be modelled as an ergodic Markov chain, and how this leads to the PageRank values being calculated as the principal left eigenvector of the transition probability matrix. (You are not required to give a formal definition of an ergodic Markov chain; an informal description will suffice.) [8 marks]

10 Natural Language Processing

The following shows a simple context-free grammar:

```
S -> NP VP     N -> N PP     Det -> the
PP -> P NP     N -> cat      P -> in
NP -> Det N    N -> box      VP -> snored
N -> Adj N     Adj -> big
```

(*a*) Describe the concept of a *chart* as used in natural language parsing. [7 marks]

(*b*) Assuming that the sample grammar is applied by a bottom-up passive parsing algorithm without packing, show the chart that would be produced from a complete analysis of the following sentence:

the big cat in the box snored

[8 marks]

(*c*) What is packing and what is its effect on the computational complexity of parsing with a context-free grammar? [3 marks]

(*d*) Show what the differences would be between the packed chart and the chart you gave in answer to part (*b*). [2 marks]

11 Security

(*a*) Social networking sites are becoming ever more popular, and many other sites now let users add each other as friends. Discuss the effect that social context has on

(*i*) phishing; [4 marks]

(*ii*) inference control; [4 marks]

(*iii*) the market for privacy; [4 marks]

(*iv*) community detection. [4 marks]

(*b*) In what ways might social context be used to protect against harm online? [4 marks]

12 Specification and Verification I

Consider the following formal specification of a program that is intended to reverse an array `A` from `A`[0] up to `A`[`N`].

```
{0 ≤ N ∧ ∀i.  0 ≤ i ∧ i < N ⇒ A[i] = a[i]}
 I:=0; J:=N;
 WHILE I<J DO
  BEGIN
   VAR TEMP;
   TEMP:=A[I]; A[I]:=A[J]; A[J]:=TEMP;
   I:=I+1; J:=J-1
  END
{∀i. 0 ≤ i ∧ i < N ⇒ A[i] = a[N-i]}
```

(*a*) What is the purpose of the array variable `a` occurring in the precondition and postcondition? [2 marks]

(*b*) How does the program work? Explain your answer in English, using diagrams or example runs to make your description as clear as possible (marks will be given for clarity). [6 marks]

(*c*) Write down and carefully explain an invariant for the `WHILE`-loop that could be used to verify that the program meets its specification (marks will be given for clarity). [8 marks]

(*d*) Does the program always terminate? Justify your answer. [4 marks]

13 System-on-Chip Design

(*a*) Briefly describe polling, interrupt-driven I/O, and DMA in the context of a system-on-chip where the I/O device is connected to the same local bus as its controlling processor. [3 marks]

(*b*) What complexity and performance issues arise for each of the three techniques of part (*a*) when the I/O device is attached to a remote bus? [4 marks]

(*c*) Two processors in a system-on-chip communicate using a hardware packet channel that conveys fixed-length messages of 48 bytes.

(*i*) Draw or describe a suitable programming model for an end point that uses interrupts but no DMA. [4 marks]

(*ii*) Sketch assembly or C-like code for the transmit side of a suitable interrupt service routine that uses a circular buffer in local RAM or similar. [4 marks]

(*iii*) A high-level model of the channel is needed as part of a system simulation that uses blocking TLM (transactional level modelling). Sketch the core code of a suitable model using constructs from SystemC and the TLM 2.0 standard or some similar library. [3 marks]

(*iv*) Explain the relationship between the firmware fragment of your assembly code in part (*ii*) and your high-level model in part (*iii*). When and how would they ever interact? [2 marks]

14 Topics in Concurrency

(*a*) Describe the syntax and transition semantics of CCS (without value-passing). [6 marks]

(*b*) Define the relation of strong bisimilarity $\sim$ between states of a transition system. [2 marks]

(*c*) Show, for CCS processes p, p', q that if $p \sim p'$ then $p \parallel q \sim p' \parallel q$. [The expression $p \parallel q$ denotes the CCS parallel composition of p and q.] [5 marks]

(*d*) Describe diagrammatically a Petri net semantics of the CCS operations of prefixing, sum and parallel composition. Your description should include an explanation of the "token game" you assume on Petri nets. State, without proof, the mathematical relationship you expect to hold between your Petri net semantics and the transition semantics of CCS. [7 marks]

15 Types

(a) Give brief explanations of the following concepts.

(i) The relation *specialization* between Mini-ML type schemes and types. [2 marks]

(ii) A *principal type scheme* for a closed Mini-ML expression. [2 marks]

(iii) The *most general unifier* of two Mini-ML types. [2 marks]

(b) State the Hindley–Damas–Milner theorem for the Mini-ML typeability problem. [3 marks]

(c) What aspect of the Mini-ML type system facilitates proofs that expressions are *not* typeable? Illustrate your answer by explaining why $\lambda x((x\,x)\,x)$ is not typeable. [5 marks]

(d) Consider adding "conjunction" types $\tau_1 \,\&\, \tau_2$ to Mini-ML with the following three typing rules:

$$\frac{\Gamma \vdash M : \tau_1 \quad \Gamma \vdash M : \tau_2}{\Gamma \vdash M : \tau_1 \,\&\, \tau_2}(\text{and}) \qquad \frac{\Gamma \vdash M : \tau_1 \,\&\, \tau_2}{\Gamma \vdash M : \tau_1}(\text{pr}_1) \qquad \frac{\Gamma \vdash M : \tau_1 \,\&\, \tau_2}{\Gamma \vdash M : \tau_2}(\text{pr}_2)$$

(i) Show that $\lambda x((x\,x)\,x)$ is typeable in this extended system. [4 marks]

(ii) Suggest an expression that is not typeable in the extended system. What, if anything, makes it difficult to prove that this is the case? [2 marks]

END OF PAPER