COMPUTER SCIENCE TRIPOS Part II

Wednesday 8 June 2011 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*) Define the following terms:

(*i*) the *Voronoi diagram* of a set of points P_i; [1 mark]

(*ii*) the *Delaunay triangulation* of a set of points P_i; [1 mark]

(*iii*) the *equiangularity* of a triangulation of a set of points P_i; [1 mark]

(*iv*) the *empty circle* property of a Voronoi diagram. [1 mark]

(*b*) Photon mapping is a two-phase algorithm consisting of *photon scattering* followed by *rendering*.

(*i*) Describe photon mapping's *scattering* phase. [4 marks]

(*ii*) Describe photon mapping's *rendering* phase. [4 marks]

(*iii*) Explain why photon mapping is considered to be a *Monte Carlo* algorithm. [2 marks]

(*c*) Compare and contrast *vertex* and *fragment* shaders. Explain where each fits in the modern graphics pipeline. [6 marks]

2 Hoare Logic

The programming language **L** consists of commands C composed from assignments $V{:=}E$ (where E is an expression) using sequences $C_1;C_2$, conditionals `IF` S `THEN` C_1 `ELSE` C_2 (where S is statement) and while-loops `WHILE` S `DO` C.

(*a*) Devise a command `SKIP` in **L** that has no effect and, for arbitrary P, prove using the Hoare logic axioms and rules for the constructs of **L** that $\vdash \{P\}\texttt{SKIP}\{P\}$. [4 marks]

(*b*) Devise a one-armed conditional `IF` S `THEN` C built only from S, C and constructs of **L** and show using the Hoare logic for **L** that if $\vdash \{P \wedge S\}C\{Q\}$ and $\vdash P \wedge \neg S \Rightarrow Q$ then $\vdash \{P\}\texttt{IF}\ S\ \texttt{THEN}\ C\{Q\}$. [6 marks]

(*c*) Define a command `MAGIC` in **L** that has the property $\vdash \{P\}\texttt{MAGIC}\{Q\}$ for any precondition P and postcondition Q. Prove that your definition of `MAGIC` has this property using the Hoare logic for **L**. [10 marks]

3 Comparative Architectures

(*a*) Why might a branch target buffer provide a poor prediction of procedure return addresses and what hardware solution may be employed to improve the accuracy of such predictions? [4 marks]

(*b*) What challenges must be overcome in order to achieve high instruction fetch rates for wide-issue superscalar processors? [6 marks]

(*c*) Embedded processors often allow both 16-bit and 32-bit instructions to be used in the same program. Why might this be advantageous? [4 marks]

(*d*) Branch prediction and speculative execution are often used to expose greater amounts of instruction-level parallelism in superscalar processors. A *reorder buffer* or *unified register file* may be used to help recover after mispredicted branches are detected.

(*i*) How are an instruction's operands located when a reorder buffer is used? [3 marks]

(*ii*) What actions are taken to recover from a mispredicted branch when a unified register file is used? [3 marks]

4 System-on-Chip Design

(*a*) Define the terms *interface*, *protocol* and *flow control* with respect to the electrical connections between sub-circuits or instantiated components in a SoC (system on chip). [2 marks each]

(*b*) Why is it critical that a protocol embodies the concept of being idle when an interface joins two different clock domains? [2 marks]

(*c*) When a pair of components are modelled using separate classes in an object-oriented language, describe **two** techniques for modelling the data transferred between them and emphasise how each technique incorporates flow control. One technique should use shared variables to model wires. [3 marks each]

(*d*) Describe and compare **two** methods for modelling the delays experienced when a pair of components communicate over a resource that may become congested (such as a SoC bus or network on chip). [3 marks each]

5 Computer Vision

(*a*) Define the notion of the "semantic gap" in the context of systems for content-based image retrieval. [2 marks]

(*b*) Many systems for optical character recognition make use of *convolutional neural networks*.

(*i*) In what sense are such networks "convolutional", and to what extent do they recognise features independent of position? [3 marks]

(*ii*) Outline how such a network could be used to recognise the characters in a high resolution digital image of this examination question, and highlight which aspects of convolutional neural networks allow for efficient detection and recognition. Assume that the network was trained to recognise only isolated instances of each of the characters. [4 marks]

(*iii*) Consider a convolutional neural network with input image size 29×29 pixels where the first stage is a convolutional layer whose feature maps each have 37 weights. The second stage of the network implements spatial subsampling by a factor of 2 in each dimension, and this is followed by another convolutional layer (third stage) whose feature maps have 26 weights each. How many neurons are there in each of the feature maps of the third stage? [3 marks]

(*iv*) Why is handwriting recognition a more difficult problem than the recognition of printed text? [1 mark]

(*c*) A template-based face detector with a basic detector size of 20×20 pixels is to be applied to an image using a multi-scale sliding window approach. The detector has a hit rate of 99.99% and a false positive rate of 0.1%.

(*i*) Explain whether this detector is likely to yield good recognition performance and give a lower bound estimate of the likely number of false positives if the image has a resolution of 4 megapixels. [2 marks]

(*ii*) Briefly describe the detection mechanism of the Viola–Jones approach to face detection, and highlight **two** aspects of this approach that make it efficient as a sliding-window detector. [5 marks]

6 Digital Signal Processing

(a) What can you say about the Fourier transform $X(f)$ if

(i) $x(t)$ is real; [2 marks]

(ii) $x(t) = -x(-t)$? [2 marks]

(b) Give the result of the Fourier transform $X(f) = \int_{-\infty}^{\infty} x(t)\,\mathrm{e}^{-2\pi \mathrm{j} f t}\,\mathrm{d}t$, using Dirac's delta where appropriate, of

(i) $x(t) = 1$; [1 mark]

(ii) $x(t) = \cos(2\pi t)$; [2 marks]

(iii) $x(t) = \mathrm{rect}(t)$; [2 marks]

(iv) $x(t) = [\frac{1}{2} + \frac{1}{2} \cdot \cos(2\pi t)] \cdot \mathrm{rect}(t)$. [3 marks]

(c) When is a random sequence $\{x_n\}$ called a "white noise" signal? [2 marks]

(d) Consider an n-dimensional random vector variable $\mathbf{X}$.

(i) How is its covariance matrix defined? [2 marks]

(ii) How can you change its representation without loss of information into a random vector of equal dimensionality in which all elements are mutually uncorrelated? [4 marks]

7 E-Commerce

(*a*) What is meant by the abbreviations CPC, CTR, CPA, ARPU, CLV in relation to online marketing? [5 marks]

(*b*) You decide to offer an online ecommerce course. The target sales price is £995.

(*i*) How will you market and promote the course online? [5 marks]

(*ii*) How will you monitor your marketing campaigns? [5 marks]

(*iii*) The table below gives Google estimates of typical keywords. Should you bid for any of these? Justify your answer. [5 marks]

Keyword	Global Monthly Searches	Local Monthly Searches	Estimated Avg. CPC	Estimated Ad Position	Estimated Daily Clicks	Estimated Daily Cost
ecommerce	1,500,000	201,000	£3.02	1.64	62	£191.94
business course	110,000	18,100	£2.63	1.32	8	£21.95
marketing course	60,500	8,100	£2.92	1.3	2	£8.81

8 Artificial Intelligence II

(*a*) Give a definition of *expected utility* and explain why the concept is useful in the context of decision-making. [2 marks]

(*b*) Give a definition of the *value of perfect information* and explain why the concept is useful in the context of decision-making. [4 marks]

(*c*) A talented, but nervous, student has to sit a difficult and important examination. There are only two possible outcomes: `pass` or `fail` and the student attaches to these utilities of $U(\texttt{pass}) = 10^6$ and $U(\texttt{fail}) = -10^8$. Lacking in confidence, his beliefs are that $\Pr(\texttt{pass}|\texttt{revise}) = 0.55$ and $\Pr(\texttt{pass}|\neg\texttt{revise}) = 0.2$. Calculate the expected utility of the situation described. [4 marks]

(*d*) The student finds what he believes might be a copy of this year's examination paper, discarded by a careless examiner. He believes that

$$\Pr(\texttt{pass}|\texttt{revise}, \texttt{thisYearsPaper}) = 0.75$$

However, should he be wrong then

$$\Pr(\texttt{pass}|\texttt{revise}, \neg\texttt{thisYearsPaper}) = 0.1$$

as he will waste time learning to answer the wrong questions, because he will revise from the wrong paper. Not revising implies

$$\Pr(\texttt{pass}|\neg\texttt{revise}, \texttt{thisYearsPaper}) = 0.7$$

However, should he be wrong then

$$\Pr(\texttt{pass}|\neg\texttt{revise}, \neg\texttt{thisYearsPaper}) = 0.08$$

He considers bribing somebody to tell him whether he has this year's paper or not; however, he thinks it is unlikely that he in fact has this year's paper, and therefore believes that

$$\Pr(\texttt{thisYearsPaper}) = 0.7$$

Compute the value of perfect information associated with finding out whether the paper is the right one. [10 marks]

9 Information Retrieval

The figure below shows the output of two information retrieval systems on the same two queries in a competitive evaluation. The top 15 ranks are shown. Crosses correspond to a document which has been judged relevant by a human judge; dashes correspond to irrelevant documents. There are no relevant documents in lower ranks.

System 1		
Rank	Q1	Q2
1	–	X
2	X	–
3	X	–
4	X	–
5	–	–
6	–	–
7	–	–
8	X	–
9	X	–
10	X	–
11	X	–
12	–	–
13	–	X
14	–	X
15	X	–

System 2		
Rank	Q1	Q2
1	X	X
2	X	–
3	X	–
4	–	X
5	X	X
6	X	–
7	–	–
8	–	–
9	–	–
10	–	–
11	X	–
12	X	–
13	–	–
14	–	–
15	X	–

(*a*) Explain the following evaluation metrics and give results for query Q1 for both systems.

(*i*) Precision at rank 10. [2 marks]

(*ii*) Recall at precision 0.5. [2 marks]

(*b*) The metrics in part (*a*) above are not adequate measures of system performance for arbitrary queries. Why not? What other disadvantages do these metrics have? [3 marks]

(*c*) Give the formula for mean average precision (MAP), and illustrate the metric by calculating System 1's MAP. [6 marks]

(*d*) For each system, draw a precision-recall curve. Explain how you arrived at your result. How could one create more informative curves? [7 marks]

10 Principles of Communications

(*a*) Draw the transition rate diagram for an $M/M/1$ queueing system, explaining each type of label, symbol and transition you have used. [5 marks]

(*b*) Suppose an $M/M/1$ server is being designed with the following business in mind:

Each customer arrival earns the service 5 Euros. However, for each unit of time the customer waits in the system, there is a refund of 1 Euro.

(*i*) What is the range of arrival rates for which the system makes a net profit? [10 marks]

[Hint: You may wish to use the result that $P_j = \rho^j(1-\rho)$, and (hence) $P_0 = 1-\rho$, where P_i is the probability of i customers being in the service and $\rho = \lambda/\mu$ is the *utilisation* of the service, for the mean arrival rate of a Poisson process, λ and a mean service time μ.]

(*ii*) Imagine that a priority queue is made available for an additional fee. Discuss qualitatively the relationship between the customers' willingness to pay, and the appropriate setting of the fee to continue to maximise profit. [5 marks]

11 Security II

You are consulting for a large online services company which stores personal information on millions of customers. Your client's directors are alarmed by the Wikileaks saga and are concerned about damage to their company's reputation should a disaffected member of staff steal and publish personal information on a large number of customers.

Discuss the security policy options available to your client to minimise the damage that a member of staff could do. [20 marks]

12 Computer Systems Modelling

(*a*) Consider an open Jackson queueing network.

(*i*) Give a description of an open Jackson network. Explain the parameters that specify the network and the state space that you would use to model its behaviour. [2 marks]

(*ii*) Derive the traffic equations for the arrival rates λ_i at each node i in the network. [2 marks]

(*iii*) What is the condition for the existence of an equilibrium distribution? [2 marks]

(*iv*) State Jackson's Theorem for an open Jackson network. [2 marks]

(*b*) Now consider the $M/M/m/m$ loss system with traffic intensity ρ.

(*i*) Show that the steady state loss probability, $E(\rho, m)$, that all servers are occupied is given by

$$E(\rho, m) = \frac{\rho^m/m!}{\sum_{i=0}^{m} \rho^i/i!}$$

[6 marks]

(*ii*) Show that $E(\rho, m)$ solves the recurrence relation

$$E(\rho, m) = \frac{\rho E(\rho, m-1)}{m + \rho E(\rho, m-1)}$$

with the boundary condition $E(\rho, 0) = 1$ and comment on why the recurrence relation is useful in practice. [6 marks]

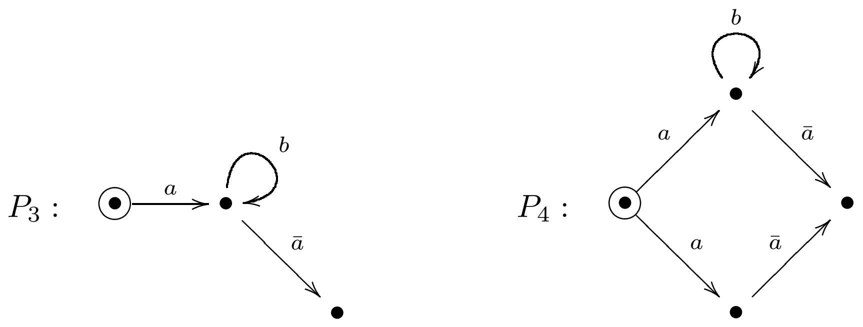
13 Topics in Concurrency

(*a*) Draw the transition systems of the following two pure CCS terms:

$$P_1 \stackrel{\text{def}}{=} (a.(b+c) \parallel \bar{b}) \setminus \{b\} \qquad\qquad P_2 \stackrel{\text{def}}{=} a.(c+\tau)$$

[3 marks]

(*b*) Write down pure CCS terms for the following two transition systems:



[3 marks]

(*c*) Carefully justify your answers to the following two questions either by exhibiting a bisimulation or by providing a Hennessy–Milner logic formula satisfied by one process and not by the other:

(*i*) Are P_1 and P_2 bisimilar? [3 marks]

(*ii*) Are P_3 and P_4 bisimilar? [3 marks]

(*d*) A *trace* of a process p_0 is a finite sequence of action labels

$$\pi = (\lambda_1, \ldots, \lambda_k)$$

for which, if π is nonempty, there exist $p_1, \ldots, p_k$ such that $p_{i-1} \xrightarrow{\lambda_i} p_i$ for all $0 < i \leq k$. Two processes p and p' are said to be *trace-equivalent* if, for all sequences of action labels π,

$$\pi \text{ is a trace of } p \text{ if, and only if, } \pi \text{ is a trace of } p'$$

(*i*) Are trace-equivalent processes always bisimilar?

(*ii*) Are bisimilar processes always trace-equivalent?

In each case, provide either a proof or a counterexample. [8 marks]

14 Types

Terms in the polymorphic lambda calculus (PLC) are given by the grammar

$$M ::= x \quad | \quad \lambda x : \tau(M) \quad | \quad M\ M \quad | \quad \Lambda\alpha(M) \quad | \quad M\ \tau$$

where τ is a type, α a type variable and x a variable.

(*a*) Give the rules for the type system in PLC. [2 marks]

(*b*) Give the rules for the relation, $\to_\beta$, of beta-reduction in PLC. Explain what it means for a term to be beta-normal. [3 marks]

(*c*) Terms in *head-normal form*, H, can be described in PLC by the grammar

$$\begin{aligned} A &::= x \quad | \quad A\ H \quad | \quad A\ \tau \\ H &::= A \quad | \quad \lambda x : \tau(H) \quad | \quad \Lambda\alpha(H) \end{aligned}$$

Arguing by induction on the structure of terms, or otherwise, prove that every term in PLC that is both typable and beta-normal is of head-normal form. [8 marks]

(*d*) Natural numbers can be encoded into PLC using the type *nat* defined as

$$nat \stackrel{\text{def}}{=} \forall\alpha(\alpha \to (\alpha \to \alpha) \to \alpha)$$

The encoding uses the Church-numerals defined as

$$\overline{n} \stackrel{\text{def}}{=} \Lambda\alpha(\lambda x{:}\alpha(\lambda y{:}\alpha \to \alpha(\underbrace{y(y(y \ldots (y}_{n \text{ occurrences}}\ x)\ldots)))))$$

Prove that the Church-numerals are the only closed, beta-normal terms of type *nat*.

Hint: Use the result proved in part (*c*) and do a case analysis over the form of terms in head-normal form. You may assume without proof that if $\Gamma \vdash M : \tau$ is provable in the PLC type system, then the free variables of the term M are contained in the domain of the typing environment Γ.

[7 marks]

END OF PAPER