COMPUTER SCIENCE TRIPOS Part II

Wednesday 7 June 2017 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*) Two pigments used to dye fabric produce identical colour under flourescent lighting but they differ when seen in sunlight. Explain how this is possible and what mechanism in the visual system gives the sensation of matching colours. Describe how light interacts with the pigments and produces the sensation of colour under different illumination. [8 marks]

(*b*) Consider the design of a tone-mapping operator for a rendering engine used in a realistic cycling simulator. The simulator will be used to study how cyclists ride at night in low light conditions and in particular how well they can spot hazards. A typical, non-HDR LCD display will be used for the simulator.

(*i*) What perceptual effects would you need to simulate in the tone-mapping operator? [4 marks]

(*ii*) Illustrate a high-level design of such an operator on a diagram showing major processing blocks. Use a similar notation to that used in the lectures: blocks for processing, arrows for data. Describe the processing pipeline succinctly, stating the purpose and output of each processing block. Note that the problem can be solved in many ways and there is no single correct solution. [8 marks]

2 Machine Learning and Bayesian Inference

(*a*) Give a detailed definition of a *hidden Markov model* (HMM). Include in your answer a description of the assumptions made by this model. [4 marks]

(*b*) Derive the *Viterbi algorithm* for computing, on the basis of some HMM, the most likely sequence of states to have produced a given sequence of observations. [6 marks]

(*c*) An HMM has three states $\{s_1, s_2, s_3\}$ with the prior

$$\Pr(S_0 = s_1) = 0.3, \Pr(S_0 = s_2) = 0.3, \Pr(S_0 = s_3) = 0.4.$$

The transition model is

$$\mathbf{T} = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.1 & 0.4 & 0.5 \\ 0.6 & 0.1 & 0.3 \end{pmatrix}$$

where $\mathbf{T}_{ij} = \Pr(S_{t+1} = s_j | S_t = s_i)$. Observations have values $\{e_1, e_2\}$, and the sensor model is

$$\Pr(E_t = e_1 | S_t = s_1) = 0.6, \Pr(E_t = e_1 | S_t = s_2) = 0.5, \Pr(E_t = e_1 | S_t = s_3) = 0.1.$$

You observe that at times 0, 1 and 2 the three corresponding observations are (e_1, e_2, e_1). Use the Viterbi algorithm to compute the most likely sequence (S_0, S_1, S_2) of states. [10 marks]

3 Comparative Architectures

(*a*) What features of a bus does a snoopy cache coherence protocol depend on? [5 marks]

(*b*) It is now possible to integrate general-purpose processor cores and a GPU on a single die and for them to share the same address space. Why might supporting hardware coherence between multiple general-purpose cores and a GPU be problematic? [5 marks]

(*c*) Consider a chip multiprocessor constructed from identical building blocks or tiles. Each tile contains a processor and private L1 caches, an on-chip network router and a slice (or fraction) of the shared L2 cache. Addresses are interleaved across these L2 slices at the granularity of cache lines. Coherence is maintained in the shared L2 by adding directory bits to each L2 line to track sharers (i.e. those L1 caches that hold a copy of the same block).

(*i*) What would be the advantages and disadvantages of adopting private L2 caches rather than a single shared L2 cache? [5 marks]

(*ii*) It is suggested that the performance of the shared cache could be improved by also keeping copies of evicted L1 cache lines in the tile's local L2 slice. Copies are made when L1 cache lines are evicted because of conflict or capacity misses. What extra actions are now required on a L1 cache miss and when a tile receives an invalidation request? [5 marks]

4 Computer Systems Modelling

A system is to be built that can serve customers at a rate μ. Three different configurations, A, B, and C, are to be considered:

A has two servers each serving at a rate $\mu/2$, each with its own queue,

B has two servers each serving at a rate $\mu/2$, but with a shared queue,

C has one server serving at a rate μ and has a single queue,

Customers arrive into the system at a rate λ. In the case of configuration A, customers join the shortest queue and cannot change queues subsequently. If the queues are the same length, the choice is random with equal probability. The arrival process is Poisson and service times are exponentially distributed.

(a) Which two of these can be modelled as *birth-death* processes? [2 marks]

(b) Draw state transition diagrams representing corresponding Markov processes for each option. Your diagrams need not go beyond four customers in the system. [4 marks]

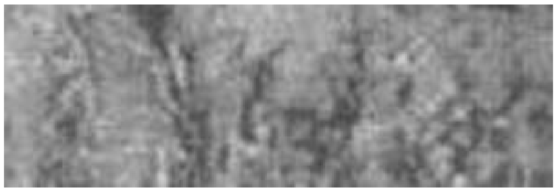
(c) For each of the birth-death processes, derive the steady state probabilities, expected utilisation and average queue length. [10 marks]

(d) Which of the birth-death configurations gives the shorter expected time in the system for a customer? Under what circumstances is this most pronounced? [3 marks]

(e) Consider the configuration that is not a birth-death process. What small change could be made that would make it easier to analyse? [1 mark]

5 Computer Vision

(*a*) Discuss the problem of figure-ground segmentation for the natural image below, and evaluate the prospects that a data-driven ("bottom-up") edge detection algorithm can correctly find the outline of the fox in the lower left. (See extract.) Explain how global, Gestalt information about the entire scene can help bring high-level knowledge and low-level data together, and consider the relevance of the massive feedback projection from the brain's visual cortex back down to the thalamus where it meets afferent retinal data. [5 marks]



(*b*) Define two-dimensional convolution and list some of its uses in computer vision. Derive an estimate for the size $(N \times N)$ of a convolution kernel to be used on a (512×512) image, at which it becomes more efficient to perform the convolution using a Fast Fourier Transform algorithm rather than explicitly computing the 2D convolution integral. [5 marks]

(*c*) Explain what is meant by the phrase "to estimate a homography". Discuss how this can be achieved, and for what purposes it is useful. [5 marks]

(*d*) For two-state decision problems, define and draw a *receiver operating characteristic* (ROC) curve. Explain why the total area under an ROC curve is a metric for the discriminability in a "same versus different" classification problem. Does the area-under-ROC metric depend on a specific choice for decision threshold? [5 marks]

6 Digital Signal Processing

A zoologist wants to record the echo-location sounds emitted by a bat. The species of bat to be recorded emits only sounds in the frequency range 40 kHz to 80 kHz and the microphone used includes an analog filter with that passband.

(*a*) Explain for each of the following sampling techniques how it can be used to convert a continuous ultrasonic microphone signal $x(t)$ into a discrete-time sequence $\{x_n\}$ and state for each technique the lowest sampling frequency f_s that enables the exact reconstruction of $x(t)$ from $\{x_n\}$:

(*i*) Passband sampling [3 marks]

(*ii*) IQ downconversion [5 marks]

(*b*) Using a 32-bit floating-point data type, how many bytes per second are required to store each of the two resulting discrete sequences from part (*a*)? [2 marks]

(*c*) Compare your answers to part (*b*) with the memory required for storing $x(t)$ sampled at the Nyquist rate of 160 kHz and explain the difference in terms of redundancy in the acquired spectrum. [2 marks]

(*d*) If the sampling techniques from part (*a*) are applied to a test signal $x(t) = \cos(2\pi f t)$ with $f = 45$ kHz, what does the discrete-time Fourier transform of the resulting discrete sequence $\{x_n\}$ look like (over the normalized frequency range $-\pi < \dot{\omega} \leq \pi$) for each technique? [4 marks]

(*e*) For both sampling techniques described in part (*a*), briefly outline the steps needed to reconstruct the original continuous waveform from the discrete sequence. [4 marks]

7 E-Commerce

(*a*) Describe five E-Commerce business models. [5 marks]

(*b*) Describe five things that need to be taken into consideration when internationalising a product or service. [5 marks]

(*c*) Describe five requirements of the Electronic Commerce Regulations 2002. [5 marks]

(*d*) How will Brexit affect the requirements in part (*c*)? [5 marks]

8 Information Retrieval

(*a*) (*i*) Given the query **"indiana jones film"** and the following term-frequencies for the two documents doc_1 and doc_2:

	indiana	**jones**	**archaeologist**	**grail**	**film**	**crusade**
doc_1	5	4	3	3	0	5
doc_2	2	2	0	2	1	3

calculate the unsmoothed query-likelihoods for both documents. [2 marks]

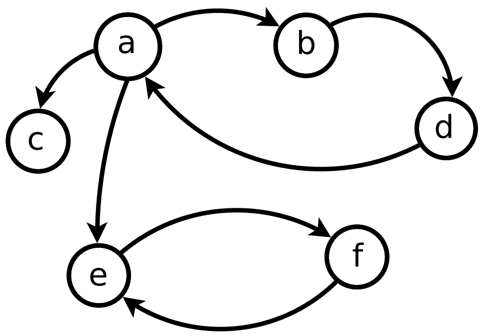
(*ii*) Describe two ways in which smoothing affects the retrieval of these documents. [2 marks]

(*iii*) Is smoothing more important for long or short queries? Justify your answer. [2 marks]

(*b*) (*i*) PageRank calculates a measure of *importance* for webpages. Give one high-level interpretation of this measure. [3 marks]

(*ii*) PageRank can be modelled as a Markov chain. What practical considerations must be addressed to ensure that the Markov chain has a stationary distribution? [3 marks]

(*iii*) Give the Markov transition matrix for the following graph assuming a teleportation probability of $\alpha = 0.5$. Discuss the suitability of this level (i.e. $\alpha = 0.5$) of teleportation for this graph. [5 marks]



(*iv*) Given the transition matrix from part (*b*)(*iii*), describe in detail how you would calculate the PageRank of each page. [3 marks]

9 Principles of Communications

(*a*) How might the small world graph properties of the Internet topology impact the design of multicast routing? [10 marks]

(*b*) The Border Gateway Protocol (BGP) provides policy control over paths taken between autonomous systems (AS). Outline the decision process at the heart of the scheme. Give illustrative examples. [10 marks]

10 Quantum Computing

Recall that the four states of the *Bell basis* are:

$$\beta_{00} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \qquad \beta_{01} = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$$
$$\beta_{10} = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \qquad \beta_{11} = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

(*a*) Give, in matrix form, a unitary operator B such that for any computational basis state $|ij\rangle(i, j \in \{0, 1\})$, we have $B|ij\rangle = |\beta_{ij}\rangle$. [2 marks]

(*b*) For the operator B defined in part (*a*), give a description of the six states: $B^2|00\rangle$, $B^2|01\rangle$, $B^3|01\rangle$, $B^3|10\rangle$, $B^4|11\rangle$, $B^4|10\rangle$. [6 marks]

(*c*) Consider a *quantum finite automaton* in a two letter alphabet $\{a, b\}$. The automaton has four states: $|0\rangle, |1\rangle, |2\rangle, |3\rangle$. The transitions on input a are given by the rule:

$$M_a : |i\rangle \mapsto |(i + 1) \bmod 4\rangle$$

and the transition matrix for input b is given by the operator B from part (*a*).

For each of the strings w below, give the probability that the automaton, when started in state $|0\rangle$, after reading w is measured to be in state $|0\rangle$.

(*i*) aaa

(*ii*) a^8

(*iii*) a^3b^4

(*iv*) $ababab$

[2 marks each]

(*d*) For the automaton defined in part (*c*), give a complete description of all strings w such that when the automaton is started in state $|0\rangle$ it reaches state $|0\rangle$ with probability 1 after reading w. [4 marks]

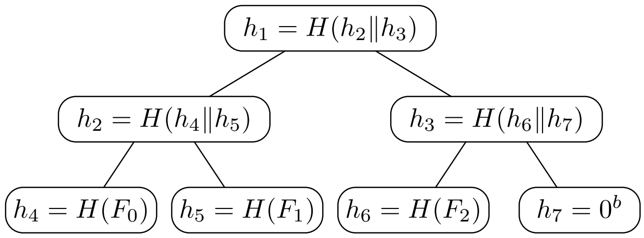
11 Security II

(*a*) An RSA encryption routine calculates the value $m^e \bmod n$ using a square-and-multiply algorithm. During the execution of that algorithm, you can briefly hear a buzzing sound (through radio-frequency interference) on an AM radio receiver located near the computer. You record that sound, and discover that it is actually the following sequence of two different sounds A and B: $BABAABABAAB$. What is the value of e? [6 marks]

(*b*) MHash implements a hash function over file sequences $F_0, F_1, \ldots, F_{n-1} \in \{0,1\}^*$ with $n > 0$, using a collision-resistant hash function $H : \{0,1\}^* \to \{0,1\}^b$:

MHash$(n, F_0, F_1, \ldots, F_{n-1})$:
  $d := \lceil \log_2 n \rceil$
  for $i := 0$ **to** $n-1$
    $h_{2^d+i} := H(F_i)$
  for $i := n$ **to** $2^d - 1$
    $h_{2^d+i} := 0^b$
  for $i := 2^d - 1$ **downto** 1
    $h_i := H(h_{2i} \| h_{2i+1})$
  return h_1

Example calculation for $n = 3$:



(*i*) Show that MHash is not collision resistant if n is not fixed, by constructing two different sequences of files that result in the same output h_1. [8 marks]

(*ii*) Suggest an improvement to MHash to make it collision resistant. [6 marks]

12 System-on-Chip Design

A hardware FIFO is a type of memory component. It has two ports, one for reading and one for writing, and implements the first-in/first-out queuing discipline.

(*a*) Why is flow control important when using FIFOs and what net-level signals must a hardware FIFO accept and generate for proper handshaking? [3 marks]

(*b*) Give a suitable signature of such a FIFO component in either a Register Transfer or Transactional Level Modelling language (RTL or TLM). Explain in detail how the handshaking works with your signature. Comment on how it would differ for the other signature.
Note: the signature of a component is its interface type, where method bodies or implementation logic are missing, such as `(input [6:0] foo, output bar)` in Verilog. [6 marks]

(*c*) In an embedded system, a FIFO is to be used for passing messages, in one direction, between a pair of processors that share no memory but which each have memory-mapped access to one of the FIFO's ports. An interrupt-driven device driver would offer no advantage in the target application. Using assembler, C or pseudocode, give either the read or write methods of a polling device driver and explain how the other would differ. [5 marks]

(*d*) An alternative approach for passing data between the pair of processors is to use shared memory. You are asked to predict the performance and energy use of the two designs in advance of making the hardware. Discuss the role a TLM model might have and whether there is a better alternative model or technique. [6 marks]

13 Topical Issues

A key challenge for autonomous vehicles is the classification of nearby objects, particularly in urban environments. For example, understanding whether an object is a static litter bin or a child about to cross the road should alter a vehicle's behaviour.

(*a*) As a first step many autonomous vehicles form a point cloud representing their local environment. Explain how this is done using lidar sensors. Your explanation should cover the pulsed and continuous-wave variants and the resolution of range ambiguities. [6 marks]

(*b*) One approach to classifying objects is for the vehicle to match against a detailed world model containing the positions of all street furniture. Discuss the advantages and disadvantages of this approach and describe how such a large model could be built. [8 marks]

(*c*) An alternative IoT-based approach has all street furniture tagged using Bluetooth Low Energy beacons. Suggest how a vehicle could reliably associate beacon broadcasts with point-cloud objects. How could your solution be improved if a cluster of nearby vehicles were able to communicate? [6 marks]

14 Topics in Concurrency

Recall that *basic* Petri nets can be considered to be general Petri nets where all multiplicities (including the capacity of conditions) are less than or equal to 1, so the initial marking and pre- and post-conditions form sets.

(*a*) Give the *token game* for basic Petri nets, specifiying when $\mathcal{M} \xrightarrow{e} \mathcal{M}'$. [3 marks]

(*b*) Draw basic Petri nets to illustrate the concepts of forwards conflict, backwards conflict and independence (also called concurrency). [4 marks]

(*c*) Draw a basic Petri net for which the token game gives the following transition system:

$$\begin{tikzcd}
 & \mathcal{M}_1 \arrow[dr, "b"] & \\
\mathcal{M}_0 \arrow[ur, "a"] \arrow[rr, "c"] \arrow[dr, "b"'] & & \mathcal{M}_3 \arrow[dl, bend right, "d"'] \\
 & \mathcal{M}_2 \arrow[ur, bend right, "a"'] &
\end{tikzcd}$$

The initial marking should be $\mathcal{M}_0$ and your answer should have four events, a, b, c and d. [5 marks]

(*d*) Show how conditions can be added to an arbitrary basic net to create a *safe* net with the same behaviour.

Your answer should describe what a safe net is, give the construction generally and prove briefly why the net obtained is safe.

Hint: systematically add conditions that represent the vacancy of conditions in the original net. [8 marks]

15 Types

(*a*) (*i*) Explain the difference with respect to polymorphic occurrences of variables between the way let-bound and lambda-bound variables are treated in languages featuring ML-style polymorphism. Illustrate your answer by explaining why in the Mini-ML language one of the two expressions

$$\lambda g\,(\texttt{let}\,f = \lambda x\,(\texttt{let}\,u = g\,x\,\texttt{in}\,x)\,\texttt{in}\,f\,f)$$
$$\texttt{let}\,f = \lambda x\,(\texttt{let}\,u = x\,\texttt{in}\,x)\,\texttt{in}\,f\,f$$

is typeable and the other is not. As part of your answer you should define what it means for a type to be a specialisation of a type scheme and give the Mini-ML typing rules for variables, function abstractions, function application and let-expressions. [14 marks]

(*ii*) What property of the type system is lost if one attempts to treat polymorphic occurrences of lambda-bound variables in the same way as let-bound ones? [1 mark]

(*b*) Give the typing rules for expressions associated with ML-style reference types, namely reference creation $\texttt{ref}\,M$, reference look-up $!M$ and assignment $M := M'$ expressions. What is the *value-restriction* that is imposed on the typing rule for let-expressions in the presence of these forms of expression and what is its purpose? [5 marks]

END OF PAPER