COMPUTER SCIENCE TRIPOS Part II

Thursday 7 June 2018 1.30 to 4.30

COMPUTER SCIENCE Paper 9

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 Algorithms

(*a*) Given an algorithm for the SET-COVER problem as a black box, how could you use this to solve the unweighted VERTEX-COVER problem? [4 marks]

(*b*) Following the approach in Part (*a*), which approximation ratio for the VERTEX-COVER problem do you achieve by applying the greedy algorithm for the SET-COVER problem? What happens if every vertex in the graph has at most 4 neighbours? [6 marks]

(*c*) Consider the following greedy algorithm for the unweighted VERTEX-COVER problem:

Compute a directed Depth-First-Search tree (DFS-tree) from every connected component in the graph, and output all nodes which are not leaves in the DFS-tree (a vertex is a leaf if it has no outgoing edges in the DFS-tree).

(*i*) What is the running time of this algorithm? [2 marks]

(*ii*) Why is the returned solution a valid vertex cover? [4 marks]

(*iii*) Derive a bound, as good as possible, on the approximation ratio of this algorithm.
Hint: You may use the fact that in any undirected graph $G = (V, E)$, $\sum_{u \in V} \deg(u) = 2|E|$, where $\deg(u)$ denotes the number of neighbours of u. [4 marks]

2 Bioinformatics

(*a*) Discuss how to efficiently cluster a set of gene expression data. [5 marks]

(*b*) Explain with one example how you would detect CG islands in a genome. [5 marks]

(*c*) Discuss the use of the Burrows-Wheeler transform in genome assembly and its algorithmic complexity. [6 marks]

(*d*) Compare the advantages and disadvantages of having long versus short k-mers in genome assembly. [4 marks]

3 Computer Systems Modelling

(*a*) Consider a birth death process $X(t)$ with state space $0, 1, 2, \ldots$ and state dependent birth and death rates given by λ_i $(i \geq 0)$ and μ_i $(i > 0)$, respectively.

(*i*) Draw a state space diagram for the birth death process labelled with the transition rates. [2 marks]

(*ii*) Derive the Chapman-Kolomogorov equations for $P_i(t)$ $(i = 0, 1, \ldots)$ where

$$P_i(t) = \mathbb{P}(X(t) = i)\,.$$

[4 marks]

(*iii*) Use the detailed balance method to derive the stationary distribution $p_i = \lim_{t\to\infty} P_i(t)$ for $i = 0, 1, \ldots$. You should state any conditions required for the existence of the stationary distribution. [4 marks]

(*b*) Consider a data centre comprising N nodes forming a computer cluster. Suppose that the individual nodes are unreliable where the time a node runs before breaking down and needing repair is exponential with rate λ independent of other nodes. A single repairer is able to return a broken-down node to service in a time that is exponentally distributed with rate μ independent of other nodes. Model the number of broken down nodes by a birth death process, draw the state space diagram, state the birth and death rates of the model and determine its stationary distribution. Comment on whether the stationary distribution always exists. Given the stationary distribution derive an expression for the mean number of broken down nodes and the stationary long run probability that a given node is working. [10 marks]

4 Computer Vision

(*a*) Consider an object's surface reflectance map $\phi(i, e, g)$ specifying the amount of incident light reflected towards a camera from each point on the surface, where the angle of the illuminant (a point source) relative to the local surface normal N is i, the angle relative to N of a ray of light re-emitted from the surface is e, and the angle between the emitted ray and the illuminant is g.

(*i*) For what kind of surface is the reflectance map simply $\phi(i, e, g) = \cos(i)$? Name this type of surface and describe its key properties. [4 marks]

(*ii*) For what kind of surface does the reflectance map simplify to $\phi(i, e, g) = 1$ if $i = e$ and both i and e are coplanar with the surface normal N, and $\phi(i, e, g) = 0$ otherwise? Name this type of surface and describe its key properties. [4 marks]

(*iii*) For what kind of surface does the reflectance map depend only on the ratio of the cosines of the angles of incidence and emission, $\cos(i)/\cos(e)$, but not upon their relative angle g nor upon the surface normal N? Give an example of such an object, and explain the consequence of this special reflectance map for the object's appearance. [4 marks]

(*b*) The binary pixel array on the left below was convolved with what operator [?] to produce the result on the right? Specify the operator by numbers within an array, and identify what task this convolution accomplishes in computer vision.

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	**1**	**1**	**1**	**1**	0	0	0
0	0	0	**1**	**1**	**1**	**1**	0	0	0
0	0	0	**1**	**1**	**1**	**1**	0	0	0
0	0	0	**1**	**1**	**1**	**1**	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$*$ [?] $\Rightarrow$

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	**-1**	**1**	0	0	**1**	**-1**	0
0	**-1**	**1**	0	0	**1**	**-1**	0
0	**-1**	**1**	0	0	**1**	**-1**	0
0	**-1**	**1**	0	0	**1**	**-1**	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

[4 marks]

(*c*) When visually inferring a 3D representation of a face, it is useful to extract separately both a shape model, and a texture model. Explain the purposes of these steps, their use in morphable models for pose-invariant face recognition, and how the shape and texture models are extracted and later re-combined.

[4 marks]

5 Denotational Semantics

(*a*) Give the definitions of poset (partially ordered set), cpo (chain complete poset), and domain. [2 marks]

(*b*) An *ideal* I of a poset $\mathsf{P} = (P, \leq)$ is a subset I of the set P such that:

- I is non-empty,
- for every $x \in I$ and $y \in P$, $y \leq x$ implies $y \in I$, and
- for every $x, y \in I$, there is $u \in I$ such that $x \leq u$ and $y \leq u$.

We write $\mathit{Idl}(\mathsf{P})$, where P is a poset, for the set of all the ideals of P.

(*i*) Show that $(\mathit{Idl}(\mathsf{P}), \subseteq)$, where $\subseteq$ denotes the subset-inclusion relation, is a cpo. [3 marks]

(*ii*) Show that $\mathit{Idl}(P, \leq)$ is non-empty iff P is non-empty. [2 marks]

(*iii*) Give a necessary and sufficient condition on a poset P for the cpo $(\mathit{Idl}(\mathsf{P}), \subseteq)$ to be a domain. Justify your answer. [2 marks]

(*c*) Give the definitions of monotone function between posets and of continuous function between cpos. [2 marks]

(*d*) (*i*) For a monotone function $f : \mathsf{P} \to \mathsf{Q}$ between posets P and Q, define a continuous function $f^{\#} : (\mathit{Idl}(\mathsf{P}), \subseteq) \to (\mathit{Idl}(\mathsf{Q}), \subseteq)$ between the cpos $(\mathit{Idl}(\mathsf{P}), \subseteq)$ and $(\mathit{Idl}(\mathsf{Q}), \subseteq)$. Prove that your definition is as requested. [4 marks]

(*ii*) For the identity function id_{P} on a poset P, show that $(id_{\mathsf{P}})^{\#}$ is the identity function on the cpo $(\mathit{Idl}(\mathsf{P}), \subseteq)$. [1 mark]

(*iii*) For monotone functions $f : \mathsf{P} \to \mathsf{Q}$ and $g : \mathsf{O} \to \mathsf{P}$ between posets O, P, and Q, show that $f^{\#} \circ g^{\#} = (f \circ g)^{\#} : (\mathit{Idl}(\mathsf{O}), \subseteq) \to (\mathit{Idl}(\mathsf{Q}), \subseteq)$. [4 marks]

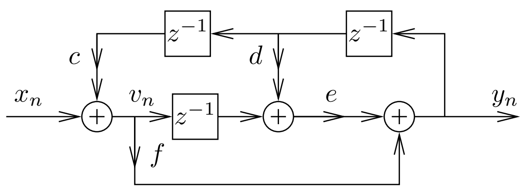
6 Digital Signal Processing

(*a*) When converting a digital audio signal from one sampling frequency to another, it is common practice to use a low-pass filter. What is the purpose of this low-pass filter, and what cut-off frequency should it have if the change of sampling frequency is

(*i*) from 12 kHz to 48 kHz;

(*ii*) from 48 kHz to 12 kHz. [4 marks]

(*b*) You are working on the firmware of a quadcopter drone. Your colleague, through trial and error, found that the following recursive filter nicely avoids unwanted oscillations in the control system:



(*i*) What are the first three samples h_0, h_1, h_2 of the impulse response of this filter? [*Note:* All delay elements have been initialized to zero.] [6 marks]

(*ii*) What is the z-transform $H(z) = Y(z)/X(z)$ of the impulse response of this digital filter? [5 marks]

(*iii*) The software development kit of your flight controller can only implement digital filters of the form

$$y_n = \sum_{k=0}^{3} b_k \cdot x_{n-k} - \sum_{l=1}^{3} a_l \cdot y_{n-l}.$$

What coefficient values a_l and b_k ($0 \le k \le 3$, $1 \le l \le 3$) will implement the same impulse response as your colleague's filter? [5 marks]

7 Hoare Logic and Model Checking

Consider a programming language that consists of commands `C` composed from assignments `V := E` (where `V` is a program variable and `E` is an expression), the no-op `skip`, sequencing `C1;C2`, conditionals `if B then C1 else C2` (where `B` is a boolean expression), and loops `while B do C`.

(*a*) Explain informally what it means for a partial correctness triple $\{P\}$ `C` $\{Q\}$ to be valid. [2 marks]

(*b*) Consider the partial correctness triple $\{\top\}$ `C` $\{\bot\}$ (where $\top$ is the true assertion, and $\bot$ is the false assertion). Give a command `C` that makes the triple valid or explain why no such command exists. [2 marks]

(*c*) Consider a new primitive command `either C1 C2` which non deterministically executes either one of its arguments: `C1` or else `C2`. Give a partial correctness logic rule for such a command, maintaining soundness and relative completeness. Give an alternative partial correctness logic rule for such a command, maintaining soundness but *not* relative completeness. [2 marks]

(*d*) Consider a new command `flip V` which randomly assigns either 0 or 1 to the variable `V`. Give a logic rule for partial correctness for such a command, maintaining soundness and relative completeness. Define `flip` using `either` from Part (*c*). [2 marks]

(*e*) Consider a new primitive command `havoc V` which assigns a random integer to the variable `V`. Give a logic rule for partial correctness for such a command, maintaining soundness and relative completeness. [2 marks]

(*f*) Consider the program `Z:=0; while (Z` $\neq$ `X` $\wedge$ `Z` $\neq$ `Y) do Z := Z+1`. Give a reasonable pre-condition so that the program terminates with `Z` equal to the minimum of `X` and `Y`. Propose an invariant for the while loop, and use it to prove that the program satisfies its partial correctness specification. [5 marks]

(*g*) Consider an extension of our programming language above with heap assignment `[E1] := E2`, heap dereference `X := [E2]`, and disposal of heap locations `dispose(E)`. Recall the list representation predicate

$$list(t, []) = (t = \texttt{null})$$
$$list(t, h :: \alpha) = (\exists y.\, t \mapsto h * (t+1) \mapsto y * list(y, \alpha))$$

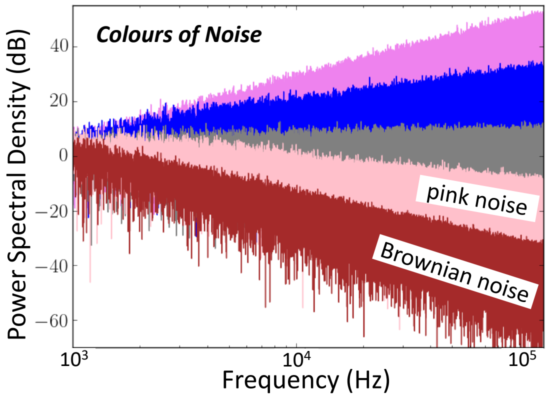
Consider the following program that deallocates a list, and counts how many list elements it deallocated:

```
while (X≠null) do (N:=N+1; Y:=[X+1]; dispose(X); dispose(X+1); X:=Y)
```

Propose an invariant for the loop that, given precondition $N = 0 \wedge list(\texttt{X}, \alpha)$, is sufficient to establish the postcondition $\texttt{N} = length(\alpha) \wedge list(\texttt{X}, [])$. [5 marks]

8 Information Theory

(*a*) Shannon's *Noisy Channel Coding Theorem* showed how the capacity C of a continuous communication channel is limited by added white Gaussian noise; but other colours of noise are available. Among the "power-law" noise profiles shown in the figure as a function of frequency ω, Brownian noise has power that attenuates as $(\frac{\omega}{\omega_0})^{-2}$, and pink noise as $(\frac{\omega}{\omega_0})^{-1}$, above some minimum ω_0.



Consider three channels suffering from either white, pink, or Brownian noise. At frequency $\omega = \omega_0$ all three channels have the same signal-to-noise ratio $\text{SNR}(\omega_0)$ and it remains at this level for the white channel, but at higher frequencies ω it improves as $(\frac{\omega}{\omega_0})$ for the pink channel and as $(\frac{\omega}{\omega_0})^2$ for the Brownian channel. Show that across any frequency band $[\omega_1, \omega_2]$ $(\omega_0 < \omega_1 < \omega_2)$ the Brownian and the pink noise channels have higher capacity than the white noise channel, and show that as frequency grows large the Brownian channel capacity approaches twice that of the pink channel. [7 marks]

(*b*) Random variable X can be any of $N = 64$ possible events, about which you have some imperfect knowledge by observing random variable Y. The average uncertainty remaining about X, given observations Y, is $H(X|Y) = 3$ bits. Use Fano's Inequality to estimate a lower bound on the probability of error P_e when guessing X from observations Y. [5 marks]

(*c*) Consider three variable-length codes for a four-symbol alphabet {A,B,C,D} having probabilities $p(x)$ as shown:

x	$p(x)$	Code 1	Code 2	Code 3
A	1/4	00	10	01
B	1/2	1	0	0
C	1/8	01	110	011
D	1/8	10	111	111

Compare the average codeword length of each code to the entropy of the alphabet, and for each code give all possible decodings of the bit sequence '1001' as a complete message. Which codes are uniquely decodable; which have the prefix (instantaneous) property; which code is the best, and why? [8 marks]

9 Mobile and Sensor Systems

In some hazardous commercial and industrial environments it is a safety risk for employees to work alone. A company wishes to create smartphone apps to support such scenarios. You should assume employees are motivated to carry their personal smartphone running the app at all times for their own safety.

(*a*) For some environments, lone working is permissible if the worker's absolute position is known at all times. On the assumption that WiFi has been deployed for communications, the company proposes to use fingerprint-based positioning. Describe how this works and discuss the practical challenges in deploying such a system. [8 marks]

(*b*) For other environments, two people must always be present but absolute location is not necessary. The company proposes two different approaches to designing an app that can raise an alarm if an employee is not close enough to another:

(*i*) to use the smartphone Bluetooth radios to detect the proximity of employee smartphones;

(*ii*) to use the smartphone speakers and microphones to estimate the distance between pairs of smartphones using ultrasonic broadcasts.

Describe how each approach would work and compare and contrast them in the context of the application. [8 marks]

(*c*) Describe how the company could combine the techniques described in Parts (*a*) and (*b*) to form an improved absolute positioning system. Discuss any disadvantages of this collaborative approach. [4 marks]

10 Natural Language Processing

(*a*) Use the following text to derive distributions for *rat* and *chased*. Use a five-word window, including open- and closed- class words, ignore case, punctuation and sentence boundaries and weight contexts by frequency.

> The cat chased a rat. A big rat chased the big dog.

[4 marks]

(*b*) Show unlabelled dependency links for each sentence and give distributions for *rat* and *chased* using contexts derived from single dependency links. [5 marks]

(*c*) In general, what type of differences will arise in distributions if dependency links are used instead of word windows as context? [4 marks]

(*d*) Outline one or more experiments to determine whether distributional information could be used in conjunction with a syntactic parser to help disambiguate prepositional phrase attachment ambiguities. [7 marks]

11 Optimising Compilers

(*a*) Describe abstract interpretation, making reference to the required domains and functions and how safety is maintained. Illustrate your answer through abstract interpretation of the rule-of-signs. [5 marks]

(*b*) Create an abstraction with three values for deciding whether the result of a calculation definitely divides by 6, definitely does not divide by 6, or may divide by 6. You need only consider multiplication and addition on integers. Show that your abstraction produces the following:

(*i*) $(4 + 3) * 6$ does divide by 6.

(*ii*) $(2 * 6) + 3$ does not divide by 6.

(*iii*) $(5 * 6) + (2 * 3)$ may divide by 6.

[10 marks]

(*c*) Explain why Part (*b*)(*iii*) cannot be computed precisely and then give an alternative abstraction that can correctly identify Part (*b*)(*iii*) as divisible by 6. [5 marks]

12 Principles of Communications

(*a*) Describe the alpha and beta models of graphs that allow for small-world and clustering characteristics. [10 marks]

(*b*) Researchers have proposed using coding of multiple blocks of data together to provide protection against packet loss, rather than relying on timeout and subsequent retransmissions. Describe how this idea can be integrated into TCP with codes instead of retransmits. Describe how the redundancy level is chosen, and also, explain role of acknowledgements in the modified TCP.

[10 marks]

13 System-on-Chip Design

(*a*) The police services of all European countries together decide to implement a new datacentre to host a shared database that holds human DNA fingerprints. The main operation that requires high-performance will be lookup by string matching.

For each of the following design considerations, write notes on the suitability of custom hardware accelerators in general and for the specific police DNA application.

(*i*) The type of arithmetic and logic likely to be needed. [2 marks]

(*ii*) The memory bandwidth and overall topology. [2 marks]

(*iii*) The relative advantages of ASIC versus FPGA implementation. [3 marks]

(*iv*) The likely energy and performance benefits. [2 marks]

(*b*) An SSRAM is a static, synchronous random-access memory.

(*i*) Draw a schematic symbol and/or describe the net-level connections to a typical SSRAM. [1 mark]

(*ii*) Describe two ways in which an SSRAM may get instantiated in an RTL (Verilog or VHDL) design. [2 marks]

(*iii*) What constraints does SSRAM have on its use in RTL? Give a conforming RTL fragment that increments one SSRAM location while obeying the constraints. [5 marks]

(*iv*) In what stage of HLS do tools instantiate RAMs and how are constraints over port use overcome? [3 marks]

14 Topical Issues

(*a*) Explain the meaning of *spread-spectrum* as it applies to radio communications. Describe the principles of Frequency-hopping and Code-Division Multiple Access spread-spectrum techniques. Give an example consumer technology for each. [5 marks]

(*b*) Explain what is meant by an *underlay system* in the context of radio systems. [3 marks]

(*c*) Consider an automotive UWB pulsed radar system used to detect objects around a vehicle.

(*i*) Describe the operating principles of UWB pulsed radar in this context. Your description should discuss range ambiguities, the effect of the Pulse Repetition Frequency (PRF), and how the distance to multiple objects can be determined. [6 marks]

(*ii*) When multiple vehicles are equipped with UWB pulsed radars, the pulses can interfere. Discuss the effects this will produce for different PRFs. [4 marks]

(*iii*) Another form of radar involves sending a continuous radio wave and measuring doppler shift. Suggest what could happen if such a signal interfered with a pulsed radar. [2 marks]

15 Types

(a) Explain what logical connective the product type corresponds to, and what patterns of proof the introduction and elimination forms for products correspond to. [2 marks]

(b) Explain what a solution for a typing problem $\Gamma \vdash M : ?$ is, and when a solution is principal. [4 marks]

(c) Give the typing rule for the let-binding form $\mathsf{let}\ x = M_1\ \mathsf{in}\ M_2$ in mini-ML with references. [6 marks]

(d) Do the following programs (for mini-ML with references) have solutions? If so, what is the type of the expression? Justify your answer in each case:

(i) $\mathsf{let}\ f = \lambda x\ (x)\ \mathsf{in}\ (f\ \mathsf{true}) :: (f\ \mathsf{nil})$ [4 marks]

(ii) $\mathsf{let}\ f = \lambda x\ (x)\ \mathsf{in}\ \mathsf{let}\ g = f\ f\ \mathsf{in}\ (g\ true) :: (g\ \mathsf{nil})$ [4 marks]

END OF PAPER