COMPUTER SCIENCE TRIPOS Part II

Thursday 7 June 2012 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 Bioinformatics

(*a*) Considerable recent research has focused on *alignment of sequences.*

(*i*) Why do we use dynamic programming algorithms for sequence alignment problems? [3 marks]

(*ii*) Describe what needs to be taken into account for gaps in DNA sequence alignment. [3 marks]

(*b*) Considerable recent research has focused on sequence database search methods.

(*i*) What are the most important differences between PatternHunter, BLAST, Smith-Waterman and Needleman-Wunsch algorithms? [9 marks]

(*ii*) Compare the heuristic used by Clustal with a dynamic programming algorithm for multiple alignment. [5 marks]

2 Computer Systems Modelling

(*a*) Consider a general birth-death process with birth rates λ_i and death rates μ_i in state i ($i = 0, 1, 2, \ldots$). What are the detailed balance equations for this process? [2 marks]

(*b*) Derive the steady-state distribution for the general birth-death process considered in part (*a*). What are the conditions for the steady-state distribution to exist? [4 marks]

(*c*) Describe the $M/M/1$ queue and give a stochastic model for the number N of customers present. Find the steady-state distribution for N and state the conditions for it to exist. [4 marks]

(*d*) Derive the mean and variance of N. [4 marks]

(*e*) State Little's law and use it to derive the mean time spent in the $M/M/1$ queue under steady state conditions. [2 marks]

(*f*) Discuss what is meant by the traffic intensity for an $M/M/1$ queue and explain what happens to the distribution of the number of customers present as the traffic intensity increases towards one. [4 marks]

3 Computer Vision

(*a*) Briefly define each of the following concepts as it relates to vision:

(*i*) active contours and energy-minimising snakes [2 marks]

(*ii*) Hadamard's criteria for well-posed problems [2 marks]

(*iii*) the hermeneutical cycle [2 marks]

(*iv*) reflectance map [2 marks]

(*v*) Bayesian prior and its role in visual inference [2 marks]

(*b*) Detecting, classifying, and recognising human faces is a longstanding goal in computer vision. Yet because the face is an expressive social organ, as well as an object whose image depends on identity, age, pose and viewing angle, and illumination geometry, many forms of variability are all confounded together, and the performance of algorithms on these problems remains very poor. Discuss how the different kinds and states of variability (*e.g.* same face, different expressions; or same identity and expression but different lighting geometry) might best be handled in a statistical framework for generating categories, making classification decisions, and recognising identity. In such a framework, what are some of the advantages and disadvantages of wavelet codes (Haar or Gabor) for facial structure and its variability? [10 marks]

4 Denotational Semantics

Given a closed PCF term F of type $nat \to nat$ and a function $f : \mathbb{N} \to \mathbb{N}$, say that F *represents* f if $F(\mathbf{succ}^n(\mathbf{0})) \Downarrow_{nat} \mathbf{succ}^{f(n)}(\mathbf{0})$ holds for all $n \in \mathbb{N}$.

(*a*) What is the *soundness* property of the denotational semantics of PCF? Use it to show that if f is not a constant function (that is, $f(m) \neq f(n)$ for some $m \neq n$), then the denotation $[\![F]\!] : \mathbb{N}_\perp \to \mathbb{N}_\perp$ of any F that represents f is the strict function that equals f when restricted to $\mathbb{N}$. [4 marks]

(*b*) If f is a constant function ($f(n) = c$ for all n, say), give, with justification, two PCF terms that represent it and that are not contextually equivalent. [5 marks]

(*c*) Consider the PCF term

$$G \stackrel{\text{def}}{=} \mathbf{fix}(\mathbf{fn}\, x : nat \to nat \,.\, \mathbf{fn}\, y : nat \,.\, \mathbf{ifzero}(F\, y)\ \mathbf{then}\ y\ \mathbf{else}\ x(\mathbf{succ}(y)))$$

where F represents a function $f : \mathbb{N} \to \mathbb{N}$ with the property that $f(n) = 0$ holds for infinitely many $n \in \mathbb{N}$. Let $\Phi : (\mathbb{N}_\perp \to \mathbb{N}_\perp) \to (\mathbb{N}_\perp \to \mathbb{N}_\perp)$ be the continuous function whose least fixed point is $[\![G]\!]$. Show by induction on k that for all $k, n \in \mathbb{N}$

$$\Phi^k(\perp)(n) = \begin{cases} \text{least } m \text{ such that } n \leq m < n + k \text{ and } f(m) = 0 \\ \perp \quad \text{if no such } m \text{ exists.} \end{cases}$$

[4 marks]

(*d*) State the *adequacy* property of the denotational semantics of PCF and *Tarski's Fixed Point Theorem* for continuous functions on a domain. Use them to deduce that the term G in part (c) represents the function $\mu_f : \mathbb{N} \to \mathbb{N}$ that maps each $n \in \mathbb{N}$ to the least $m \geq n$ such that $f(m) = 0$. [7 marks]

5 Digital Signal Processing

(*a*) Make the following statements correct by changing one word or number. (Negating the sentence is not sufficient.)

(*i*) The z-transform of a sequence shows on the unit circle its discrete-time cosine transform. [1 mark]

(*ii*) Delaying a sequence by two samples corresponds in the z-domain to multiplication with z^2. [1 mark]

(*b*) Consider a causal digital IIR filter of order 2, operated at a sampling frequency of 48 kHz, where the impulse response $\{h_n\}$ has (for $n > 2$) the shape of a sine wave of frequency 8 kHz (amplitude and phase do not matter).

(*i*) Where in the z domain can you place two zeros and two poles to achieve such an impulse response $\{h_n\}$ in the time domain? [4 marks]

(*ii*) Write down the z transform of $\{h_n\}$ as a rational function (with those zeros and poles). [6 marks]

(*iii*) Provide the constant-coefficient difference equation that describes the time-domain behaviour of that filter. [4 marks]

(*iv*) How can you use such a filter design to digitally generate an 8 kHz sinewave sampled at 48 kHz with very little computational effort? [4 marks]

6 Information Theory and Coding

(*a*) Prove that the information measure is additive: that the information gained from observing the combination of N independent events, whose probabilities are p_i for $i = 1 \ldots N$, is the *sum* of the information gained from observing each one of these events separately and in any order. [3 marks]

(*b*) What is the shortest possible decodable code length, in bits per average symbol, that could be achieved for a six-letter alphabet whose symbols have probabilities $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{32}\right)$? [3 marks]

(*c*) Consider Shannon's third theorem, the *Noisy Channel Coding Theorem,* for a continuous communication channel having bandwidth W Hertz, perturbed by additive white Gaussian noise of power spectral density N_0, and average transmitted power P.

(*i*) Is there any limit to the capacity of such a channel if you increase its signal-to-noise ratio $\dfrac{P}{N_0 W}$ without limit? If so, what is that limit? [4 marks]

(*ii*) Is there any limit to the capacity of such a channel if you can increase its bandwidth W in Hertz without limit, but while not changing N_0 or P? If so, what is that limit? [3 marks]

(*d*) For each of the four classes of signals in the left table below, identify its characteristic spectrum from the right table. ("Continuous" here means supported on the reals, *i.e.* at least piecewise continuous but not necessarily everywhere differentiable. "Periodic" means that under multiples of some finite shift the function remains unchanged.) Give your answer just in the form 1-A, 2-B, etc. Note that you have 24 different possibilities. [4 marks]

Class	*Signal Type*
1.	continuous, aperiodic
2.	continuous, periodic
3.	discrete, aperiodic
4.	discrete, periodic

Class	*Spectral Characteristic*
A.	continuous, aperiodic
B.	continuous, periodic
C.	discrete, aperiodic
D.	discrete, periodic

(*e*) Define the Kolmogorov algorithmic complexity K of a string of data. What relationship is to be expected between the Kolmogorov complexity K and the Shannon entropy H for a given set of data? Give a reasonable estimate of the Kolmogorov complexity K of a fractal, and explain why it is reasonable. [3 marks]

7 Mobile and Sensor Systems

Pelagus Sensing Ltd is consulting you about the design of a system to monitor sea water content along the coasts of Britain.

(*a*) Initially, imagine that fixed sensors need to be installed on buoys in the sea. The system needs to be able to report data on sea water, such as temperature and saline composition, ten times a day for more than a year over a coastal area of one kilometre.

(*i*) List the challenges in defining the above defined architecture and suggest a physical architecture of the system. [2 marks]

(*ii*) Describe the *Low-Energy Adaptive Clustering Hierarchy (LEACH)* approach, presented in the lectures. [2 marks]

(*iii*) Illustrate how *LEACH* can be employed to design a solution to the request of *Pelagus Sensing Ltd*, indicating advantages and disadvantages. [4 marks]

(*b*) Now imagine that *Pelagus Sensing Ltd* coordinates with zoologists monitoring seals through sensors attached to the animals in the same coastal area. The zoologists offer to put additional water temperature and saline composition sensors on their seal tags to help *Pelagus Sensing Ltd* with sea water monitoring.

(*i*) Describe an effective physical architecture (specifying which radio the tags would use and which other hardware devices should be present in the network, if any) which would allow sensing data to be collected through the animal tags and illustrate advantages and limitations. [2 marks]

(*ii*) Illustrate the MAC and routing protocols that you would use in such an architecture explaining advantages over other protocols described in the lectures. [4 marks]

(*iii*) Indicate how data aggregation could be used in this setting, highlighting advantages and disadvantages. [3 marks]

(*iv*) If the mobile sensors attached to the seals were to be eventually reprogrammed, which protocol would you suggest to be installed on them and why? [3 marks]

8 Natural Language Processing

The following shows a simple context free grammar (CFG) for a fragment of English:

```
S -> NP VP            NP -> Sandy
VP -> V NP            Det -> a
VP -> V NP NP         N -> puppy
VP -> V NP PP         N -> garden
VP -> VP PP           V -> walks
NP -> Det N           V -> gives
PP -> P NP            P -> to
NP -> Kim             P -> in
```

(*a*) Give all analyses (parse trees) this grammar assigns to the following sentences:

Kim gives a puppy to Sandy.
Kim walks a puppy in a garden.

Which of the analyses is correct? [6 marks]

(*b*) The example sentences in (*a*) illustrate a type of ambiguity common in natural languages. Characterise it, and explain how it arises. [4 marks]

(*c*) Indicate at least two (other) sources of incorrect analyses in the above grammar, illustrating them with the corresponding ungrammatical sentences. Explain what causes the incorrect analysis. [4 marks]

(*d*) The incorrect analyses in (*a*) could be avoided using an appropriate feature structure based grammar. Describe the exact mechanism for how this can be done. Does this treatment also deal with your incorrect analyses in (*c*)? [6 marks]

9 Optimising Compilers

(*a*) Explain the core ideas of strictness analysis, including the abstract values used for abstracting non-function values and what concrete values they represent. Briefly explain how program functions f are abstracted to strictness functions $f^\sharp$. Give the abstractions of `λ(x,y).x+y` and `λ(x,y). if random() then x else y`. [5 marks]

(*b*) Justify or correct the following statements: (*i*) "since abstract interpretation replaces real-world computation with a directly corresponding abstract computation then strictness analysis fails to terminate on non-terminating programs"; and (*ii*) "when a strict function is applied to an expression e then e is necessarily evaluated during the call". [4 marks]

(*c*) We now wish to extend strictness analysis from simple `int` expressions to allow also (lazy) `int list` expressions. These represent lists whose head and tail components are only evaluated when required. Wadler suggested capturing strictness-like properties on lazy lists using an abstract interpretation with four abstract values for `int list` concrete values:

- 0: non-termination
- ∞: a chain of cons cells, either infinite or having some tail component which does not terminate
- $0\in$: a chain of cons cells ending in `nil` but having at least one member which does not terminate
- $1\in$: a possibly empty chain of cons cells ending in `nil` every member of which terminates

By analogy with ordinary strictness functions, give abstract interpretations in truth-table form (noting that values of type `int list` have four values rather than the standard two) for the following functions involving lazy list values:

(*i*) `λ(x:int list). nil` [1 mark]

(*ii*) `λ(x:int list). cons(42,x)` [1 mark]

(*iii*) `λ(x,y:int list). if random() then x else y`
Explain how you resolved any choice which arose. [3 marks]

(*iv*) `hd` [2 marks]

(*v*) `tl` [1 mark]

(*vi*) `append` [2 marks]

(*vii*) `reverse` [1 mark]

10 Principles of Communications

(*a*) Describe the operation of Distance Vector (D-V) and Link State (L-S) routing algorithms, with particular reference to their complexity in terms of number of messages given the number of nodes and edges in the network. [5 marks each]

(*b*) Small world networks are a particular form of random graph which exhibit clustering.

(*i*) Describe two common ways this clustering can occur in the way the graph is created. [3 marks each]

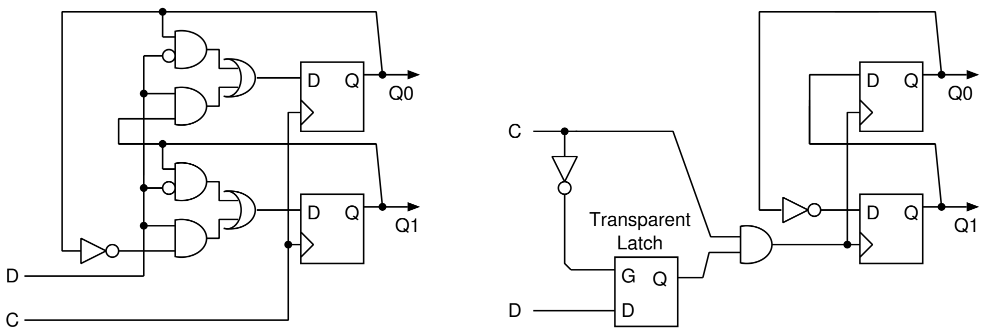
(*ii*) Discuss the design of a routing algorithm for a small world network. Would you stick with existing schemes like D-V or L-S? [4 marks]

11 System-on-Chip Design

(*a*) Give a basic formula for modelling delay in logic circuits once wiring length is known. [2 marks]

(*b*) Give a basic formula for modelling power consumption in logic circuits once wiring length is known. [2 marks]

(*c*) The figure below gives two versions of a finite-state machine before and after a modification.



Give the name of this modification, say why is it performed and fully explain the role of the additional latch. [4 marks]

(*d*) Three power saving techniques used in current SoCs are dynamic voltage and frequency scaling (DVS), clock gating and power supply gating. Compare and contrast these three techniques in terms of

(*i*) physical granularity [3 marks]

(*ii*) temporal granularity [3 marks]

(*iii*) the manual complexity they add to the design flow [3 marks]

(*e*) Including power consumption, discuss what design considerations result in a product using an FPGA (field-programmable gate array) rather than an application-specific integrated circuit (ASIC)? [3 marks]

[*Note:* In all sections, marks will be awarded for sensible argument even if assumptions or results are incorrect.]

12 Temporal Logic and Model Checking

If p and q are atomic properties of states and R is a transition relation, then the property q *is an immediate successor to* p holds if and only if for every state s such that $p\ s$ is true, there is at least one state s' such that $R\ s\ s'$ and $q\ s'$ are true.

(*a*) Is it possible to express the property that q is an immediate successor to property p in LTL? If so, write down the LTL formula that expresses it, if not give an argument why it cannot be expressed. [10 marks]

(*b*) Is it possible to express the property that q is an immediate successor to property p in CTL? If so, write down the CTL formula that expresses it, if not give an argument why it cannot be expressed. [10 marks]

[*Hint:* Recall the discussion of the property "can always reach a state satisfying P" in the lectures.]

13 Topical Issues

(*a*) (*i*) Discuss the resilience of UWB impulse radio systems to multipath fading. Explain how they avoid causing interference to traditional narrowband communications systems. [5 marks]

(*ii*) The Wireless USB standard describes a high bandwidth radio protocol intended to replace wired USB systems. In the course of its development, both the 3.1–10.6 GHz UWB band and the unlicensed EHF radio band (57–64 GHz) were considered for the physical layer. Suggest **two** advantages of UWB over EHF in this context. Explain your answers. [4 marks]

(*b*) Describe how UWB radio can be used for pulse radar. Include in your answer the two approaches for handling slow and fast-moving objects and explain why there is a well-defined maximum operating distance for such systems. Explain what effect separating the transmitters and receivers spatially would have. [11 marks]

14 Types

(*a*) For each type variable α, the *option type* O_α is defined to be the PLC type

$$O_\alpha \stackrel{\text{def}}{=} \forall\beta\,(\beta \to (\alpha \to \beta) \to \beta).$$

Prove that there are closed PLC expressions *None*, *Some* and *Case* with the following typing and beta-conversion properties.

(*i*) $\vdash None : \forall\alpha\,(O_\alpha)$

(*ii*) $\vdash Some : \forall\alpha\,(\alpha \to O_\alpha)$

(*iii*) $\vdash Case : \forall\alpha, \beta\,(\beta \to (\alpha \to \beta) \to O_\alpha \to \beta)$

(*iv*) $Case\,\alpha\,\beta\,y\,f\,(None\,\alpha) =_\beta y$

(*v*) $Case\,\alpha\,\beta\,y\,f\,(Some\,\alpha\,x) =_\beta f\,x.$

[10 marks]

(*b*) Use *Case* and *None* to define a closed PLC expression *Lift* of type $\forall\alpha_1, \alpha_2\,((\alpha_1 \to O_{\alpha_2}) \to O_{\alpha_1} \to O_{\alpha_2})$ with the property that for all closed types τ and all closed expressions M of type $O_\alpha[\tau/\alpha]$, $Lift\,\tau\,\tau(Some\,\tau)\,M =_\beta M$. You may assume that any closed beta-normal form of type $O_\alpha[\tau/\alpha]$ is beta-convertible either to $None\,\tau$, or to $Some\,\tau\,N$ where N is a beta-normal form of type τ. Any standard results about PLC that you use should be carefully stated. [10 marks]

END OF PAPER