UNIVERSITY OF CAMBRIDGE

COMPUTER SCIENCE TRIPOS Part II

Tuesday 2 June 2015 1.30 to 4.30 pm

COMPUTER SCIENCE Paper 7

*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
Rough work pad

SPECIAL REQUIREMENTS
Approved calculator permitted

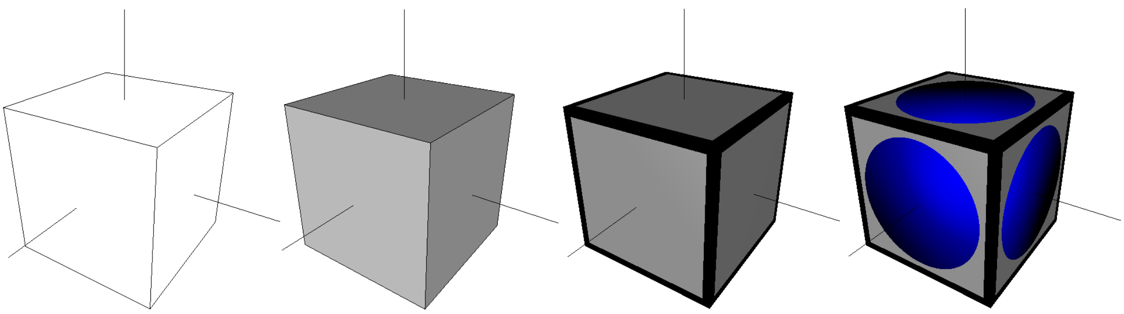
1 Advanced Graphics

In this question, you are asked to write OpenGL GLSL code. Correct GLSL syntax is not required and is not part of the marking scheme. If you are unsure of a method name or syntax, use a reasonable approximation. Full marks will be awarded if you make your intention clear with the use of suitable comments.

An OpenGL program is using the following GLSL shader fragment program:

```
#version 330

const vec3 BLACK = vec3(0, 0, 0);
const vec3 WHITE = vec3(1, 1, 1);
const vec3 BLUE = vec3(0, 0, 1);

uniform vec3 eyePosition;        // Camera position
uniform vec3 lightPosition;      // Light source position

in vec3 position;                // Fragment position
in vec3 normal;                  // Fragment normal

out vec4 fragmentColor;          // GLSL fragment output color

void main() {
  vec3 color = WHITE;

  // [YOUR CODE HERE]

  fragmentColor = vec4(color, 1);
}
```

The shader is being used to render a cube. The cube is centered at the origin, axis-aligned, and has edge length 2; its corners are from $(-1, -1, -1)$ to $(1, 1, 1)$.



(*a*) **Diffuse shading:** Fill in the body of the shader fragment program to implement diffuse shading. [4 marks]

(*b*) **Specular shading:** Extend your solution to implement specular shading as well as diffuse shading. Shade the surface with a 75% diffuse, 25% specular balance between the two methods. [4 marks]

(*c*) **Procedural texturing:** Extend your solution to render thick black edges on all faces of the cube (see illustration). The edges should be approximately 0.1 units thick. [6 marks]

(*d*) **Procedural texture effects:** Extend your solution to create the illusion of a dark blue sphere of radius 1.3 units emerging through the sides of the cube. [6 marks]

2 Advanced Algorithms

(*a*) State the fundamental theorem of linear programming. [3 marks]

(*b*) Consider the following linear program:

$$\text{Minimize} \quad -3x_1 - 2x_2$$

subject to:

$$\begin{aligned} 3x_1 + x_2 &\leq 5 \\ -2x_1 &\geq -10 + 4x_2 \\ x_1, x_2 &\geq 0. \end{aligned}$$

(*i*) Convert this LP into standard and slack form, and specify the initial basic solution. [4 marks]

(*ii*) Solve this LP using the simplex algorithm. Specify the associated basic solution after each iteration. [4 marks]

(*c*) We consider the *Steiner Tree Problem* defined as follows. We are given an undirected, connected graph $G = (V, E)$ with a non-negative cost-function $c : E \rightarrow \mathbb{R}_+$. Further, we are given a set $S \subseteq V$ of terminals. The goal is to find a minimum-cost subgraph of G that connects all terminals, where the cost of a subgraph is the sum of the costs of its edges.

Consider the following algorithm:

- Let $H = (V, E')$ be the *metric completion* of G, where $E' = \{\{u, v\} \colon u, v \in V\}$ and $c(\{u, v\})$ is the cost of the shortest path from u to v in G.
- Compute a Minimum Spanning Tree T on the subgraph $H[S]$ induced by the set of terminals S.
- Replace every edge $\{u, v\}$ in T by the edges of a shortest path from u to v in G, and return the solution.

(*i*) Prove an upper bound of $2(1 - \frac{1}{|S|})$ on the approximation ratio of this algorithm.

[*Hint:* You can use an approach similar to the analysis of Approx-TSP-Tour.] [6 marks]

(*ii*) Construct an example which provides a matching lower bound on the approximation ratio. [3 marks]

3 Bioinformatics

(*a*) Compute the global alignment and the best score of the following sequences CGTGAA, GACTTAC with the following parameters: match score = +5, mismatch score = −3, gap penalty = −4. [4 marks]

(*b*) If the sequences have different base composition or length, what parameter values would you choose in order to determine multiple alignment of the sequences? Justify your answer. [4 marks]

(*c*) Discuss the main features of the Burrows-Wheeler transform (BWT) using the following example: T = GATTACA$. Also explain the reversibility of BWT. [4 marks]

(*d*) Discuss the use of clustering in the analysis of gene expression microarray data. [4 marks]

(*e*) Discuss the UPGMA algorithm as applied to the following distance matrix of the species A, B, C and D.

$$\begin{array}{l} \textit{species} \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 2 & & \\ 4 & 4 & \\ 6 & 6 & 6 \end{pmatrix} \end{array}$$

[4 marks]

4 Business Studies

(*a*) Explain the concept of Net Present Value (NPV) and include a description of how it is calculated. [5 marks]

(*b*) You are asked to help prepare a bid, under the Private Finance Initiative (PFI) to develop, install and maintain for 4 years, a patient information system for a hospital. The hospital Trust will pay the bid amount in 5 equal annual payments at the beginning of each year. Your development team estimates development and installation will cost £1M and maintenance £100k per annum. Which of the following bids: £1500k, £1600k or £1700k should your team submit, assuming a discount rate of 10% per annum? Show your working. [10 marks]

(*c*) Draw up a cash flow for this project. How much funding is needed assuming a loan interest rate of 10% per annum? [5 marks]

5 Comparative Architectures

(*a*) Why do modern processors typically exploit multiple cores rather than a single multithreaded processor? [5 marks]

(*b*) What does support for vector chaining and tailgating allow in a vector processor? [5 marks]

(*c*) How might recent advances in die stacking help to improve microprocessor performance and reduce costs? [5 marks]

(*d*) In what ways might parallelism be exploited to reduce the power consumption of a microprocessor? [5 marks]

6 Denotational Semantics

(*a*) For monotone functions $f, f' : P \to Q$ between posets $(P, \sqsubseteq_P)$ and $(Q, \sqsubseteq_Q)$, let $f \sqsubseteq f' \overset{\text{def}}{\iff} \forall x \in P.\ f(x) \sqsubseteq_Q f'(x)$.

(*i*) Prove that the binary relation $\sqsubseteq$ is a partial order. [3 marks]

(*ii*) For monotone functions between posets $p : P' \to P$, $f, f' : P \to Q$, and $q : Q \to Q'$, prove that $f \sqsubseteq f' \implies q \circ f \circ p \sqsubseteq q \circ f' \circ p$. [1 mark]

(*b*) An *adjoint pair* $\big(f : P \to Q, g : Q \to P\big)$ is a pair of monotone functions between posets such that $\mathrm{id}_P \sqsubseteq g \circ f$ and $f \circ g \sqsubseteq \mathrm{id}_Q$.

(*i*) Let $f_1, f_2 : P \to Q$ and $g_1, g_2 : Q \to P$ be monotone functions between posets such that (f_1, g_1) and (f_2, g_2) are adjoint pairs. Prove that:

(A) $f_1 \sqsubseteq f_2 \iff g_2 \sqsubseteq g_1$ [4 marks]

(B) $f_1 = f_2 \iff g_1 = g_2$ [2 marks]

(*ii*) Let $\big(f : P \to Q, g : Q \to P\big)$ be an adjoint pair where the posets P and Q have least elements. Prove that the monotone function f is strict. [2 marks]

(*c*) (*i*) Define the notion of lub (least upper bound) of a countable increasing chain in a poset. [2 marks]

(*ii*) Let $\big(f : D \to E, g : E \to D\big)$ be an adjoint pair where each of the posets D and E is a cpo (chain complete poset). Prove that the monotone function f is continuous. [6 marks]

7 Hoare Logic

(*a*) Explain informally the difference between Gödel's completeness theorem and his first incompleteness theorem. [8 marks]

(*b*) State the meaning of Hoare triples $\{P\}\ C\ \{Q\}$ in separation logic. [3 marks]

(*c*) For each Hoare triple (*i*), (*ii*) and (*iii*) below state whether the triple is true. Explain your answer and if the triple is true give a proof using separation logic. You may assume $\vdash P$ if P is a true separation statement (e.g. $\vdash \mathtt{X} \mapsto 0 \ \Rightarrow\ \mathtt{X} \mapsto \_$). Such assumed statements should be stated explicitly and informally justified.

(*i*) $\{\mathtt{X} = 0\}\ [\mathtt{X}] := 1\ \{\mathtt{0} \mapsto 1\}$

(*ii*) $\{\mathtt{X} \doteq 0\}\ [\mathtt{X}] := 1\ \{\mathtt{0} \mapsto 1\}$

(*iii*) $\{\mathtt{X} \doteq 0\ \wedge\ \mathtt{X} \mapsto 0\}\ [\mathtt{X}] := 1\ \{\mathtt{0} \mapsto 1\}$

[9 marks]

8 Human–Computer Interaction

This question asks you to analyse the design of a familiar website, but applying fundamental principles of human-computer interaction rather than simply repeating web design conventions. Start by choosing the website that you will use for the remainder of the question. This can be any site that has sufficient complexity for the user to manipulate structured information, and should be a site that you yourself use regularly enough to be familiar with its interface. You may assume that the examiner is familiar with current versions of Facebook, Wikipedia, Gmail, Hermes webmail, YouTube and Amazon. If you wish to use a site other than one of these, please provide a pictorial sketch that you can refer to in your answer.

(*a*) Name the website that you will refer to in the rest of this question, and explain the nature of the information structure that the user creates and interacts with when using this site. Provide a pictorial sketch if necessary, as described above. [1 mark]

(*b*) Describe *two* aspects of the visual language (marks, symbols, regions, surfaces) used in the design of this site. For *each* aspect, explain the nature of the correspondence between the visual appearance and its meaning or purpose within the interaction design. [6 marks]

(*c*) Describe a typical activity in which the user interacts with the information structure presented by the site. You should refer to *two* different Cognitive Dimensions that are particularly salient in this activity, and explain what effect *each* of these has on the user's experience. [6 marks]

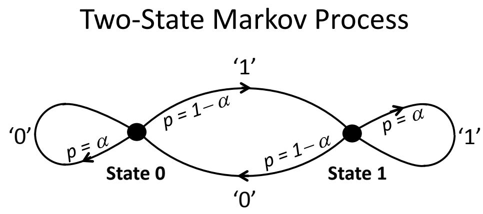
(*d*) Propose a way in which the visual design of the site might be modified, that would have an effect on *one* of the Cognitive Dimensions described in part (*c*). Consider any trade-offs that might result. [3 marks]

(*e*) Describe how you could carry out an investigation to evaluate the effects predicted in part (*d*). [3 marks]

(*f*) How would you classify the method you have described in part (*e*), using the distinctions between qualitative/quantitative, empirical/analytic, and summative/formative methods? [1 mark]

9 Information Theory and Coding

(*a*) A two-state Markov process may emit a '0' in State 0 or a '1' in State 1, each with probability α, and return to the same state; or with probability $1-\alpha$ it emits the other symbol and switches to the other state. Thus it tends to be "sticky" or oscillatory, two forms of predictability, depending on α.



(*i*) What are the state occupancy probabilities for $0 < \alpha < 1$? [2 marks]

(*ii*) What are the entropy of State 0, the entropy of State 1, and the overall entropy of this source? Express your answers in terms of α. [2 marks]

(*iii*) For what value(s) of α do both forms of predictability disappear? What then is the entropy of this source, in bits per emitted bit? [2 marks]

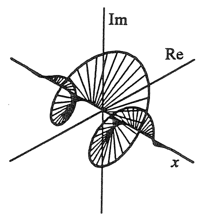
(*b*) Consider a binary symmetric channel with error probability p that any bit may be flipped. Two possible error-correcting coding schemes are available.

(*i*) Without any error-correcting coding scheme in place, state all the conditions that would maximise the channel capacity. Include conditions on the error probability p and also on the probability distribution of the binary source input symbols. [2 marks]

(*ii*) If a (7/4) Hamming code is used to deliver error correction for up to one flipped bit in any block of seven bits, provide an expression for the residual error probability P_e that such a scheme would fail. [2 marks]

(*iii*) If instead repetition were used to try to achieve error correction by repeating every message an odd number of times $N = 2m + 1$, for a positive integer m followed by majority voting, provide an expression for the residual error probability P_e that the repetition scheme would fail. [2 marks]

(*c*) Gabor wavelets are an important class of complex-valued functions for encoding information with maximal resolution simultaneously in the frequency domain and the signal domain. Using an expression for their functional form, explain:

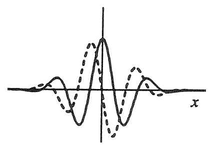


(*i*) their spiral helical trajectory as phasors, shown here with projections of their real and imaginary parts;

(*ii*) the Uncertainty Principle under which they are optimal;

(*iii*) the spaces they occupy in the Information Diagram;

(*iv*) some of their uses in pattern encoding and recognition.



[8 marks]

10 Natural Language Processing

The following grammar fragment associates context free syntax rules with lambda calculus rules for semantic composition:

```
S -> NP VP
```
$\text{VP}'(\text{NP}')$
```
VP -> Vditrans NP PPto
```
$(\text{Vditrans}'(\text{NP}'))(\text{PPto}')$
```
VP -> Vintrans
```
$\text{Vintrans}'$
```
PPto -> to NP
```
NP'
```
Vditrans -> gives
```
$\lambda y \lambda x \lambda z[\text{give}'(z, y, x)]$
```
Vintrans -> sleeps
```
$\lambda w[\text{sleep}'(w)]$
```
NP -> Sandy
```
s
```
NP -> Kim
```
k
```
NP -> Fido
```
f

(*a*) Show the semantics that the grammar above would assign to:
Kim sleeps
Kim gives Fido to Sandy [2 marks]

(*b*) Extend the grammar to also allow:
Kim gives Sandy Fido
Show the semantics your grammar gives for this sentence. [2 marks]

(*c*) Extend the grammar so the following sentences are all given the semantics $\text{reject}'(k, f)$, explaining the reasons for the choices you have made.
Kim rejects Fido
Kim turns Fido down [7 marks]
Kim turns down Fido

(*d*) Most native English speakers accept sentences like (*d*)(*i*) below as grammatical but find sentences like (*d*)(*ii*) ungrammatical. Outline how you would augment the grammar you gave in answer to (*c*) so that (*d*)(*i*) was accepted but (*d*)(*ii*) was not.

(*i*) Kim turns it down

(*ii*) *Kim turns down it

[3 marks]

(*e*) What alternative strategy might you use in a large-scale language generation system to ensure that (*d*)(*i*) was generated in preference to (*d*)(*ii*)? In general, what are the advantages and disadvantages of this approach in comparison with constraining generation by the grammar? [6 marks]

11 Optimising Compilers

This question concerns intraprocedural analysis of a flowgraph-style program P whose instructions are in three-address form and labelled $S_1, S_2, \ldots$, with S_1 labelling the first instruction to be executed. Input parameters to the program are simulated by the first few instructions of P being of the form `x = read()`, and the result given by a `return` instruction.

An instruction T is a *semantic reaching definition* at instruction U if, for some execution starting at S_1, instruction T writes to a variable x which does not suffer an intervening assignment when execution reaches instruction U. We write $RD(i)$ for the set of instructions S_j whose definitions reach instruction S_i.

(*a*) By analogy with live variable analysis or available expression analysis, derive dataflow equations for RD and give an algorithm for solving these. Explain any approximation you make, carefully justifying the form of this approximation. [*Hint*: you may find it useful to define *gen* and *kill* for instructions.] [8 marks]

(*b*) Is your analysis for reaching definitions flow-sensitive or flow-insensitive? Give a one-sentence justification of your answer. [2 marks]

(*c*) One use of reaching definitions is for constant propagation: when we know that reading a variable in an operand in a given instruction will always result in the same value k, we may replace the operand with k. Carefully explain how we can use the result of reaching-definitions analysis to perform constant propagation. [*Hint*: you may find it useful to consider the instruction form `z:=x+y`.] [3 marks]

(*d*) Explain how your constant-propagation algorithm would react to the following flowgraph expressed as C code:

```
int t,r,x;
x = read();
if (x>91) t=7; else t=6;
r = t/2;
return r+39;
```

Either explain why your resulting code is optimal, or indicate the source of any information loss which precludes it being optimal. [3 marks]

(*e*) Suppose now the 3-address code were in SSA (single static assignment) form. How would this affect the result of reaching-definitions analysis? [4 marks]

12 Security II

(*a*) Clearly explain the Clark-Wilson security policy model and what it tries to achieve, defining technical terms such as CDI, UDI, CW triples, IVP, TP and auditing. [6 marks]

(*b*) (*i*) What is a master key system? What is its purpose? How can we turn a normal pin-tumbler lock into one supporting a master key? [5 marks]

(*ii*) Describe in detail the Blaze Privilege Escalation attack on master key systems. What resources does an attacker need and what can be achieved? Compare the effort required to that of a brute-force attack. [5 marks]

(*c*) Discuss the security, privacy and economic aspects of the iPhone's "App Store" model, as opposed to the traditional desktop software model. [4 marks]

END OF PAPER