UNIVERSITY OF DUBLIN

TRINITY COLLEGE

Faculty of Engineering and Systems Sciences

Department of Computer Science

B.A (Mod.) Computer Science — Trinity Term 2000

Junior Sophister Examination

3BA4 - Computer Architecture II

Tuesday 30th May — Exam Hall — 9.30 - 12.30

Dr. J.O. Jones, Dr. Andrew Butterfield

Attempt **FIVE** questions at least two from each section.
Use separate answer books for each section.

SECTION A

Q1. What is a pipelined processor? What are the benefits of pipelining? Explain the organization & operation of the DLX five stage execution pipeline.

What effect would a "simple" implementation of branch instructions have on the DLX pipeline. What is branch prediction? Explain the operation and advantage(s) of a "two bit" branch prediction scheme.

Show how branch prediction is implemented in the DLX processor and illustrate its operation by showing what happens under the following circumstances (i) a particular branch instruction is executed for the first time (ii) the particular branch instruction is executed again & is predicted correctly and (iii) the particular branch instruction is executed again but predicted incorrectly.

Q2. Explain the meaning of the three constants L, K & N in terms of cache organisation. Describe the operation of a LKN cache.

Write two functions (eg. in C, C++) to model a LKN cache with an LRU replacement policy.

```
Cache *cache (l, k, n)
        // returns a pointer to a suitably initialised Cache data structure

int hit (c, a)
        // returns 1 if address a is in cache c, otherwise 0
```

Clearly state any assumptions made.

Q3. What is the cache coherency problem? Under what conditions are the caches in a system considered to be coherent?

Explain (i) the meaning of the 4 cache line states used in the write-once cache coherency protocol and (ii) the basic operation of the protocol. Given a 2 CPU+cache multiprocessor system, illustrate the bus traffic and cache state transitions that would occur if the following CPU memory requests are issued:

CPU 0: read a0
CPU 1: read a0
CPU 0: write a0
CPU 0: write a0
CPU 0: write a0
CPU 1: write a0
CPU 1: write a0
CPU 0: read a0

What advantage does the write-once protocol have over the simpler write-through scheme.

Q4. What is a spin lock? Two tasks executing on separate processors in a multiprocessor update a shared variable. Why should a spin lock be used to protect access to the shared variable? What bearing does any underlying cache coherency protocol have on updating the shared variable?

What is an atomic instruction? What action does an atomic "fetch & increment" instruction perform? What steps are needed to make an instruction atomic in a multiprocessor environment?

Code (eg. C or C++) the algorithm for a ticket lock with proportional backoff.
Explain the theory and advantages of the ticket lock with reference to your code.

SECTION B

Q5 Given the following logic function: Y = NOT((P+Q)*(R+S))

(i) Design a CMOS switch circuit that implements this function. Your design should include a determination of the widths of the transistors involved (assuming a minimum width of 0.2μm, and a μ_n/μ_p ratio of **3.5**).

(ii) Express the CMOS layout topology of your circuit using a Stick Diagram, subject to the following constraints: Power and Ground run horizontally across the top and bottom respectively, of the circuit, *all inputs enter on the top*, on Polysilicon, and the *output emerges on both the left and right*, also on Polysilicon.

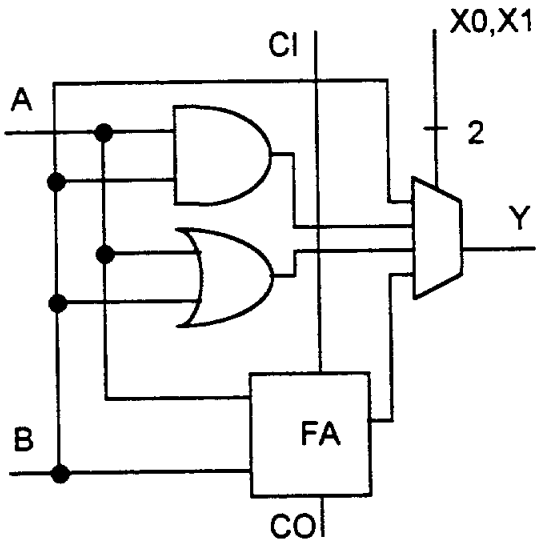
Q6 (i) For the following types of design rules:separation and extension, describe the nature of the rules and explain what manufacturing problems they are designed to solve.
(ii) Assume a manufacturing process has the following layout and electrical design rules:

- min. overlap of Metal1 over Ccut: 1μm
- min width of Ccut: 2μm
- min sep of Ccut: 2μm
- max. current density of Metal1 wire: 5mA/μm
- max. current density of Ccut perimeter wire: 0.5mA/μm

Design a contact-region to connect Metal1 to nDiff, capable of carrying a load of 100mA with minimum overall area.

Q7 Consider a process technology where the effective resistance and load capacitance of a minimum-sized inverter are 20kΩ and 5fF respectively.
(i) Design a chain of inverters to drive a load of 1000fF as fast as possible.
(ii) Design a 2000μm long, 1μm wide line with inverters added at intervals, to be as fast as possible, given that the line has the following characteristics: sheet resistance: 20Ω/square, Capacitance/Unit Area: 0.05fF/μm^2.

Q8 Context is important in IC Design - a given logic function may have different forms in various parts of an integrated circuit according to the differing relationships it has with its neighbours. Use the design of one bit of an *n*-bit ALU shown below as an example to show how some of its subparts with the *same* logic function end up having *different* physical implementations.



A, B, Y are corresponding bits of the *n*-bit inputs and outputs of the ALU.
CI is the carry in to the 1-bit ALU, while CO is the carry out.
X0, X1 are the control inputs determining the ALU function.