UNIVERSITY OF DUBLIN

TRINITY COLLEGE

Faculty of Engineering and Systems Sciences

Department of Computer Science

B.A (Mod.) Computer Science
Junior Sophister Examination

Trinity Term 2004

3BA4 - Computer Architecture 2

Monday 31st May Sam. Beckett Rooms 09:30 - 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. How are virtual addresses converted to physical addresses by a two level memory management unit? Illustrate the advantage of using a 2-level page table structure by comparing the amount of physical memory needed for the page tables of a small and a maximum sized process with that of using a single level page table structure. Assume 4GB virtual & physical address spaces with a 10-10-12 address configuration like the Intel 486.

What is demand paged memory management? Given the 2–level page table structure, show the organisation of the initial page table for a user process which has 0x2000 bytes of code, 0x3000bytes of initialised data, 0x2000 bytes of uninitialised data and 0x0800 bytes of stack data copied from its parent. How many pages of memory need be allocated to the process initially? Explain what happens when code is executed and data accessed in the different memory regions? How are illegal memory accesses detected?

Q2. What is a cache? How does a cache reduce the effective memory access time of a CPU?

Explain how the organization of a cache can be characterised by the three constants L, K & N. What special names are given to cache organisations where (i) N=1 (ii) K=1 and (iii) K=4. Explain in detail how a LKN cache is "searched" for the contents of a memory address. What actions take place on a cache hit and a cache miss?

Cache misses can be classified into 3 types - compulsory, capacity and conflict. Explain the meaning of these terms and explain how the number of compulsory, capacity and conflict misses could be calculated for a given cache.

What is meant by the replacement policy? Explain how Intel's pseudo-LRU and Maruyama's LRU replacement policy algorithms operate. Comment on the number of "extra" bits needed to implement each algorithm.

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

Given a three CPU multiprocessor system where each CPU has a local cache and the following memory requests:

CPU 0:read a2
CPU 0:write a2
CPU 0:write a2
CPU 1:read a2
CPU 1:read a0
CPU 0:write a2
CPU 0:write a2

Explain in detail what happens on each memory request (e.g. bus traffic and cache line state transitions) if (i) a MESI and (ii) a Firefly cache coherency protocol is used.

Assume (i) each cache is direct mapped with 2 cache lines (ii) even addresses map to line 0 & odd addresses to line 1 and (iii) the caches initially contain addresses a0 & a1.

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

Give the algorithm (in C or pseudo code) for a ticket lock with proportional back off. Explain the theory and advantages of the ticket lock with reference to your code.

Section B

Q5 Given the following logic function: Y = NOT(A*B)*NOT(C*D)

(i) Design a CMOS switch circuit that implements this function (you may need to re-write it in a different form first). Your design should include a determination of the widths of the transistors involved
(assuming a minimum width of 0.1μm, and a μ_n/μ_p ratio of **3**).

(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, *inputs A and C enter on the left, while inputs B and D enter on the right*, on Polysilicon, and the *output emerges on the bottom*, also on Polysilicon.

Q6 Describe how the cost of manufacturing a integrated circuit is affected by the physical size of the device, and how this is linked to the minimum feature size allowed by the manufacturing technology. Derive any formulas you use to defend your arguments, stating clearly any assumptions made.

Q7 (i) Discuss all the factors that affect the speed performance of a logic gate.
(ii) Explain why making transistor sizes larger in dense logic circuits does not result in an improvement in speed.
(iii) Given the need to drive large external loads, explain the optimal way to connect the minimum sized transistors of internal logic to the high capacitance external devices.

Q8 (i) Show how CMOS technology, in an appropriate form, can be used to produce regular programmable logic arrays (PLAs). You should describe the floorplan scheme, and give floorplan diagrams for each of the basic tiles, and stick diagrams showing two possibilities for the array core tiles.

(ii) Sketch out a PLA to implement the following logic, using only four product terms:
V = /A*/B
W = /A*B
X = A+B
Y= B*A
Z = A*/B
where A and B are inputs and V, W, X, Y and Z are outputs. Your answer should explain how you reduce the product terms to four and the programming of the AND- and OR-arrays.