EEL308 Major Exam — Semester 1302 — 6 May 2014

(Permitted references: Textbook (Patterson, Hennessy) and lecture notes of this course only.) Absolutely no Internet access, use of Wikipedia, tutorial problem solutions, etc. This and/or any other malpractice will earn you an 'F' grade in the entire course. No explanations will be entertained.

General instructions

A. *Attempt all 9 questions. Marks distribution: 2+4+6+7+6+2+4+3+1=35*
B. *Suggested times needed for answering each question are provided. They add up to 109 minutes, which gives you 11 minutes of slack.*
C. *Please begin your answer to each question on a new page. All parts of a question must be answered together.*
D. *Read each question carefully. All of the data you need are provided in the question. Make no random assumptions. Questions 5 (6 marks) and 7 (4 marks) require greater attention.*
E. *If the answer is not an integer, write at least three non-zero digits after the decimal point.*

Good luck! ☺

Time: 2 hours **Full marks: 35**

1. (a) What is a general purpose register architecture? (b) Give two examples of architectures that DO NOT belong to this category.

[1 + 1 = 2 marks]

(Suggested time to be spent on answering this question: 2 minutes)

2. A data centre using 500 hard disks reports an AFR of 0.876%. (a) What is the MTTF of each hard disk? (b) If the manufacturer also quotes the same MTTF under the assumption that the failure rate of each disk follows an exponential distribution, what is the theoretical time (in years) after which the probability of failure of a disk is 50%?

[2 + 2 = 4 marks]

(Suggested time to be spent on answering this question: 12 minutes)

3. In a particular single-cache memory organisation, using a standard suite of benchmarks, we have determined that the cache miss rate is 5%. Cache hit determination takes 2 CPU clock cycles. The main memory has an interleaved organisation (where each bank is a word wide) and the block size is 4 words. (a) If the CPU is clocked (cpu_clk) at 1 GHz, the system bus (mem_clk) to the main memory is clocked at 250 MHz, and the data access time within the DRAM is 20 mem_clk cycles, find out AMAT (in standard units of time) for this benchmark suite when no optimisation is used for servicing misses. Then, compare AMAT using two separate modifications: (b) using critical word first (where cache miss rate remains the same and there is no extra time penalty), and (c) changing the block size to 1 word, but the miss rate shoots up to 10%.

[2 + 2 + 2 = 6 marks]

(Suggested time to be spent on answering this question: 18 minutes)

4. (a) In a 5-stage MIPS pipeline, will there always be a stall after a lw instruction that is followed by another instruction that uses the loaded value? If not, provide an example (write the instructions). (b) To enable this, write the condition-checking statements for the hazard detection unit. (c) How will the forwarding unit and associated datapath be modified? (d) What will be the new condition-checking statements in the forwarding unit? (Hint: Use the format of writing condition-checking statements provided in the textbook and lecture notes.)

[1 + 2 + 2 + 2 = 7 marks]

(Suggested time to be spent on answering this question: 18 minutes)

5. In an ARM-based architecture, a very simple IO device has the following memory-mapped registers, whose details are provided in the table below.

Name of the register	Size	Purpose	Values and their meaning
CMD	8-bit	Provide commands to the IO device	0x01: Read 0x10: Write data (non-bursty) 0x11: Provide address (for read/write) 0x24: Write data (bursty) 4-words 0x28: Write data (bursty) 8-words
STS	8-bit	Reflect the IO device status or error messages	0x00: Idle 0x01: Busy 0xff: Error
DATA0, DATA1, ..., DATA7	32-bit (word) each	Data to be read/written from/to device. Address to be provided to device.	Address (in the IO device to where data gets written to/read from) is provided through DATA0. The IO device has its own address range (16 bit addressing in the device) from 0x0000 to 0xffff, where each address denotes a location where a word can be written to or read from. Non-bursty data payload is provided through DATA0. Bursty data payload is provided through DATA0...DATA3 (4 words) or DATA0...DATA7 (8 words).

All the memory-mapped registers (irrespective of their size) are word-aligned in little-endian fashion. The register r0 contains the base address in main memory where the memory mapping for the registers of this device begins, and continues with increasing address for CMD, STS, DATA0, ..., DATA7 respectively. For example, CMD is mapped in memory at the address contained in r0. STS is mapped to the next valid address, and so on. When some program wants to write a 4-word burst to the IO device at address 0x1010, it needs to undertake the following procedure:
Keep checking STS till it becomes idle or throws up an error. If there is an error, exit. If it becomes idle, first provide the address to the device, then provide the requisite data words (0x000023ee; 0x89019900; 0x89019090; 0x000046ff). In each case, the device has to be informed (through CMD) whether it is being provided with address or data. Write ARM assembly code to accomplish this. **[6 marks]**

(Suggested time to be spent on answering this question: 24 minutes)

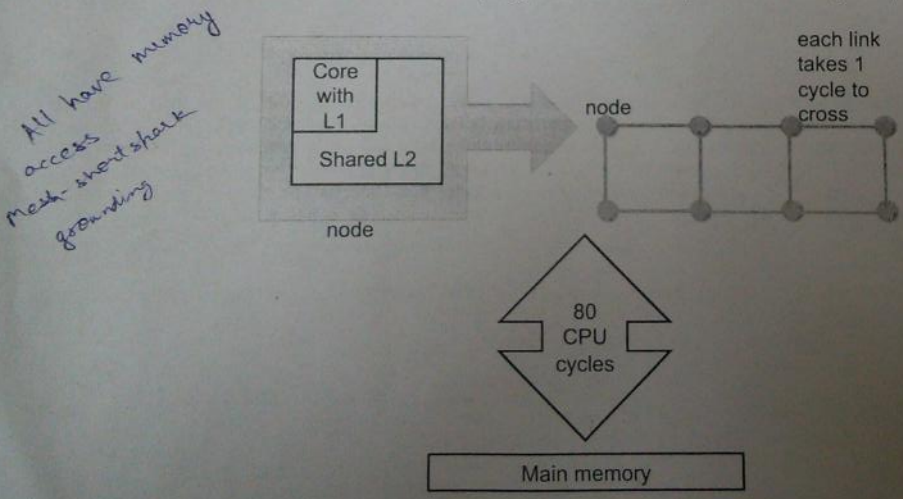
6. Explain how a write-buffer would help in the case of (a) write-through cache and (b) write-back cache. Be very specific about what benefit is achieved under what circumstances. **[1 + 1 = 2 marks]**

(Suggested time to be spent on answering this question: 7 minutes)

7. A shared-memory multiprocessor architecture containing 8 cores arranged in a rectangular 4x2 mesh (see figure below) has private L1 caches per core and a logically shared but physically distributed L2 cache. In terms of CPU clock cycles, time taken for comparing tags and declaring a hit in L2 is 2 cycles and overall L2 miss penalty is 80 cycles (uniform from any node). Since L2 is physically distributed, the total access time for L2 also needs to include the time taken to communicate with the node where that segment of the shared L2 cache is physically located. If the segment is located at the same node where L1 miss occurred, no extra cycles are spent. If the segment is located at a different node, communication time between the processing node and the node containing the target L2 segment needs to be added while computing the overall L2 access latency. By considering ideal caches on this octa-core system, we evaluated an **IPC** of 4.0. Now, considering real caches with each L1 having a miss rate of 5% and the shared L2 cache having a miss rate of 1%, find out the best and worst case **IPC** given the physical arrangement of the L2 cache. Time taken to go from one node in the 4x2 mesh to any of its neighbours is one CPU clock cycle. All links are bidirectional and have enough bandwidth. Assume L2 access begins after L1 miss, and main memory access begins after L2 miss is known. Also assume load is evenly distributed across all 8 cores and there is no congestion during communication.

[4 marks]

(Suggested time to be spent on answering this question: 16 minutes)



8. Explain the difference between VLIW and superscalar multiple-issue architectures. What is the advantage each could offer over the other? **[1 + 1 + 1 = 3 marks]**

(Suggested time to be spent on answering this question: 10 minutes)

9. What is the significance of (a) bits 31:28 and (b) bit 20 in an ARM data processing instruction? **[1 mark]**

(Suggested time to be spent on answering this question: 2 minutes)