EEL 4713: Computer Architecture - Spring 2012
Midterm 2 - 120 points possible

15

1. **(30 points total)** Architectures:

Consider two pipelined implementations of the MIPS: P1 and P2. Assume there is a split level-one cache, with perfect (100%) hit rates.
- In P1, there are five stages (IF, ID, EX, MEM, WB) as discussed in class and in your lab. The clock rate is 2GHz, and the hit latency for both the instruction and the data caches is 1 cycle.
- In P2, the clock rate is increased to 3.0GHz but the number of pipeline stages is increased to 6: IF, ID, EX, MEM1, MEM2, WB. The MEM stage is divided into two separate stages: in the first stage (MEM1) the address for the memory access is used to index the cache and compare tags; in the second stage (MEM2), the data is retrieved from the cache (for a load) or stored in the cache (for a store).

A) **(5 points)** What is the best-case throughput (in instructions per second) that each pipeline can achieve?

5

P1: $2 \cdot 10^9$ instructions/sec

P2: $3 \cdot 10^9$ instructions/sec

B) **(5 points)** What is the worst-case throughput (in instructions per second) for each pipeline? Assume there are no stalls due to branches.

0

It's the same. The extra cycle does not cause a decrease in rate; it just increase the time needed for a single instruction to get all the way through. 100% hit rate on both implies no penalty on miss in either case.

P1: $2 \cdot 10^9$ instructions/sec

P2: $3 \cdot 10^9$ instructions/sec

C) **(20 points)** Describe what additional hazards need to be detected in the P2 pipeline (with respect to P1). Describe which ones can be dealt with by forwarding, and which ones cause the pipeline to stall. (Consider only register-register operations, loads, and stores – don't worry about branches).

For each hazard you specify, write down the comparison that is needed to detect the hazard condition (as done in the book (e.g., IF/ID.rs == ID/EX.rd) and explain which pipeline registers need to be stalled under this condition, or what is the source and destination in the datapath for forwarding.

~~There are 2 hazards involving cache misses – one for a read, one for a write. On a read, there will need to be logic in the MEM1 stage that detects whether the cache hit. If it did not, then the IF/ID and ID/EX registers will need their delay increased by the number of cycles corresponding to a cache miss if IF/ID.rs, IF/ID.rt, ID/EX.rs, or ID/EX.rt equals EX/MEM1.rt on subsequent instructions.~~

~~A similar delay will need to be implemented if the data written on a write instruction is subsequently fetched after a cache miss.~~

From P1 to P2, the time before data loaded becomes available is increased, because there is an extra cycle before the data reaches the pipeline from the cache. Thus, the former 1-cycle delay needed in P1 when a register instruction followed a load by 1 cycle is now needed in P2 when a register instruction follows a load by 2 cycles. Similarly, a 2-cycle delay is now needed when a register instruction follows a load by 1 cycle. (Delays on IF/ID and ID/EX)

$$\left. \begin{array}{l} \text{IF/ID.rs == EX/MEM1.rs} \\ \text{IF/ID.rt == EX/MEM1.rs} \end{array} \right\} \rightarrow \text{Delay 1 cycle, forward cache output to ALU}$$

$$\left. \begin{array}{l} \text{ID/EX.rs == EX/MEM1.rs} \\ \text{ID/EX.rt == EX/MEM1.rs} \end{array} \right\} \rightarrow \text{Delay 2 cycles, forward cache output to ALU}$$

88/120

25

2. **(30 points total)** Caches:

A) **(6 points)** List the 3 Cs for cache miss classification.

0

cold, conflict, capacity

B) **(4 points)** A write buffer improves the performance of which cache write policy?

0

C) **(10 points)** Given a two level cache hierarchy where the level one cache has a 1 cycle hit latency and 10 cycle miss penalty and the level two cache has a 50 cycle miss penalty, calculate the average memory access time if the level one cache hit rate is 10% and the level two cache hit rate is 50%.

$1 \cdot 0.1 + 0.9 \cdot (0.8 \cdot 10 + 0.5 \cdot 50) = 27.1$ cycles

D) **(10 points)** Consider two level-one data caches; both are 2 bytes in size, 1 byte per block. Cache C1 is direct-mapped, cache C2 is 2-way set associative with LRU replacement. Suppose a program issues load instructions that access bytes of memory in the following sequence: 0, 1, 2, 0, 1, 2, 0, 1, 2, 0, ... for a large number of iterations. What is the approximate hit rate for C1 and C2?

10

C1: 33%

C2: 0%

3. **(25 points total)** Queuing Theory:

A) **(5 points)** What is the disk (server) utilization if a disk gets 80 I/O requests per second and the average disk service time is 10 ms?

$80 \cdot 0.010 = 0.8 = 80\%$

B) **(10 points)** What is the average system response time for the disk in part A)

$\frac{0.010 \cdot 0.8}{1 - 0.8} = 0.04 = 40$ ms $+ 10$

C) **(10 points)** What arrival rate in I/O requests per second would cause the system to saturate (server utilization > 100%)?

$X \cdot 0.010 > 1 \longrightarrow X > 100$

4. **(20 points total)** Processor Evaluation:

Assume processors A and B have the following instruction mixes and cycles per instruction type:

Instruction type	Processor A		Processor B	
	Mix	Cycles	Mix	Cycles
Load/store	10%	5	40%	5
Branch	10%	3	10%	3
ALU	80%	1	60%	1

a) **(6 points)** Calculate the CPI for each processor:

$$PA: 1\cdot0.8 + 3\cdot0.1 + 5\cdot0.1 = 1.6$$
$$PB: 1\cdot0.6 + 3\cdot0.1 + 5\cdot0.4 = 2.9$$

b) **(6 points)** How much faster is processor A as compared to processor B?

$$\frac{2.9}{1.6} = 1.81 \rightarrow 81\% \text{ faster, assuming same clock speed.}$$

c) **(8 points)** Suppose you want to make processor B faster than processor A and you choose to optimize the load/store operation. The optimization would reduce the number of cycles required by load/store operations by a whole number amount (i.e., 1, 2, 3, or 4 cycles instead of 5). If you only want to optimize load/store until the point that processor B is faster than processor A, is it possible to make processor B faster than processor A? If so, how many cycles will the optimized load/store require? Show your work to support your answer.

$$1\cdot0.6 + 3\cdot0.1 + x\cdot0.5 = 1.6$$

$$0.9 + 0.5x = 1.6$$

$$0.5x = 0.7 \rightarrow x = 1.4$$

So load/store would need to take 1 cycle, assuming same clock speed.

q

5. **(15 points total)** Virtual memory:

A) **(5 points)** List two advantages that virtual memory provides.

5

Increases overall memory capacity

Allows programs their own memory areas to avoid collisions

B) **(5 points)** Page faults are typically handled by software routines instead of using dedicated hardware, however software is orders of magnitude slower than hardware. Explain why a software routine is still appropriate to handle page faults.

O

Because one of the jobs of the O.S. is to manage memory utilization between programs

Doesn't have to be

C) **(5 points)** Since virtual addresses must be translated into physical addresses before indexing into a physically-indexed physically-tagged cache, this translation is on the critical path. What fundamental structure reduces the average translation time?

4

Lookup tables for the virtual indexing

BONUS QUESTION (5 EXTRA CREDIT POINTS):

What is the 4ᵗʰ cache miss C? (I said I wouldn't "test" you on this, but if you were paying attention in the review session you should get this)

Compulsory

HAVE A NICE SUMMER!