

CS704 – Advanced Computer Architecture-II

Due Date: 17th July, 2012

Assignment 3

Instructions to Solve Assignments

The purpose of assignments is to give you hands on practice. It is expected that students will solve the assignments themselves. Following rules will apply during the evaluation of assignment.

- Cheating from any source will result in zero marks in the assignment.
- Any student found cheating in any two of the assignments submitted will be awarded "F" grade in the course.
- No assignment after due date will be accepted.

Question 1: Total Points (10+10=20)

Consider the speculative processor discussed in lectures. Since the reorder buffer contains a value field, you might think that the value field of the reservation stations could be eliminated.

- (a) Show an example where this is the case and an example where the value field of the reservation stations is still needed. Use the speculative processor shown in Figure 1.1. Show MIPS code for both examples. How many value fields are needed in each reservation station?
- (b) Find a modification to the rules for instruction commit that allows elimination of the value fields in the reservation station. What are the negative side effects of such a change?

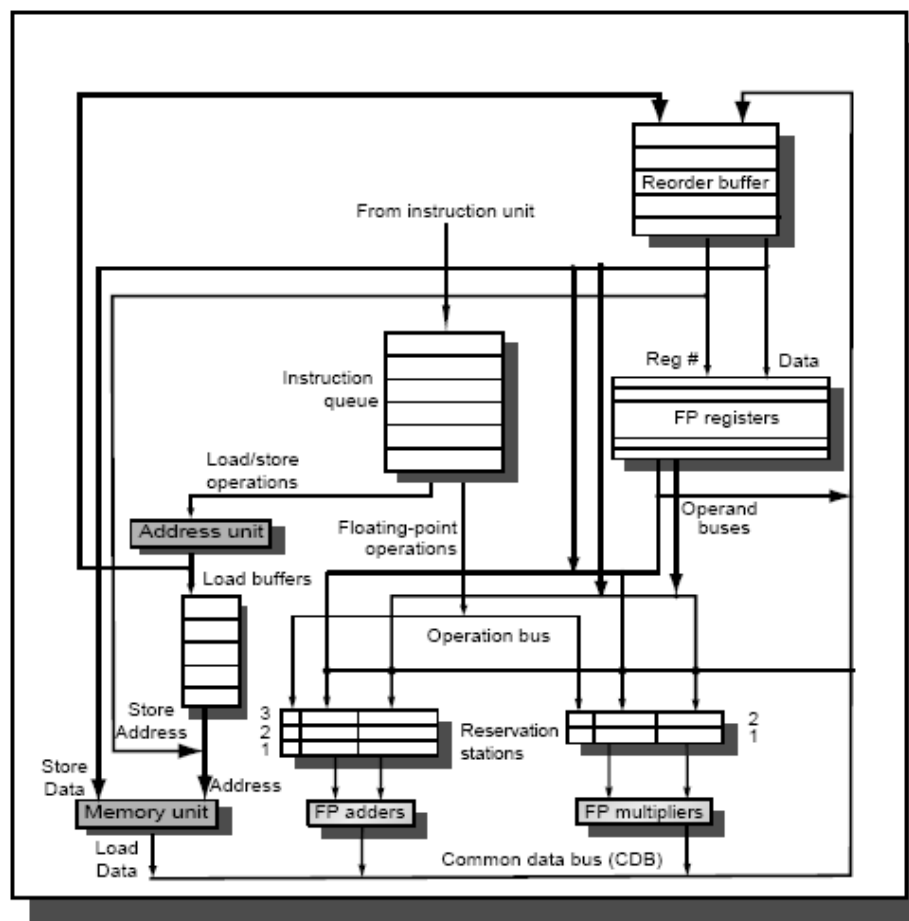


FIGURE 1.1 The basic structure of a MIPS FP unit using Tomasulo's algorithm and extended to handle speculation.

Question 2: Total Points (5+5+5+5=20)

For this problem, assume that you have a processor with a cache connected to main memory via a bus. A cache access takes 1 cycle. A successful cache access (a hit) finishes within that cycle. On an unsuccessful access (a miss) additional work must

be performed to fetch a block from main memory over the bus. A bus transaction consists of one cycle to send the address to memory, four cycles of idle time for main-memory access, and then one cycle to transfer each word in the block to the cache. (Assume that the processor continues execution only after the last word of the block has arrived.) In other words, if the block size is B words (at 32 bits/word), a cache miss will cost $1 + 4 + B$ cycles.

Following table gives the average cache miss rates of a 1 megabyte cache for various block sizes.

Block size(B)	Miss ratio(m) %
1 word	3.40%
4 words	1.00%
16 words	0.40%
64 words	0.25%
256 words	0.19%

- Write an expression for the average memory access time for a 1-MByte cache and a B -word block size (in terms of m and B).
- What block size yields the best average memory access time?
- If bus contention adds three cycles to the main-memory access time, which block size yields the best average memory access time?
- Ignoring bus contention, if the bus width is doubled to 64 bits, what is the optimal block size?

Question 3: Total Points (10+5+10=25)

You have a computer with two levels of cache memory and the following specifications:

- Processor: 2GHz, 64-bit RISC CPU
- On-chip L1 caches
 - split instruction & data cache, blocking, single-ported
 - write-through & non-write allocate
 - 1 CPU cycle access time
 - Block size = 32 bytes
- Off-chip L2 cache off-chip
 - unified single-ported cache, blocking
 - write-back
 - 10 CPU cycles access time (L1 miss penalty) for both reads and writes
 - Block size = 32 bytes
- Main memory:
 - Bus: 64-bit data transfers at 400 MHz
 - Latency: 15+5+5+5 CPU cycles access time for 32 bytes (L2 miss penalty – includes latency of both DRAM and memory bus)

(a) What is the:

- Peak L1 data cache bandwidth available to CPU (assuming 0% L1 misses)?
- Peak L2 cache bandwidth available to L1 cache (assuming 0% L2 misses)?
- Main memory bandwidth available to L2 cache?

[Calculate the bandwidths in Gbytes/sec, i.e. 2^{30} bytes/sec.]

(b) You are given the following L1 cache statistics for a program executing on this system

Metrics	Access Type:				
	Total	Instrn	Data	Loads	Stores
Accesses	10000000	7362210	2637790	1870945	766845
Misses	52206	8466	43740	36764	6976

Words Read From Lower-levels 180920 (i.e. 45230 cache lines)
 Words Written-back to Lower-levels 766845
 Total Traffic 947765

How long does an average instruction take to execute (in ns), assuming 1 clock cycle per instruction in the absence of memory hierarchy stalls, no write buffering at the L1 cache level, and 0% L2 miss rate? Ignore register dependencies between instructions.

(c) You are considering replacing the L2 cache with a victim cache. Given the information provided to you, compute a measure of “speed” for each alternative and indicate which is the faster solution. Justify the metric you choose to compare the two alternatives and state your assumptions. Assume the performance statistics are:

- L2 cache local miss ratio = 0.18
- Victim cache miss ratio = 0.23
- Victim cache transport time from L1 miss = 2 CPU clock

Question 4: Total Points (10)

A large amount (more than a third) of DRAM power can be due to page activation (see <http://download.micron.com/pdf/technotes/DDR2/TN4704.pdf> and <http://www.micron.com/systemcalc>). Assume you are building a system with 1 GB of memory using either 4-bank 512 Mbit × 4 DDR2 DRAMs or 8-bank 1 Gbit × 8

DRAMs, both with the same speed grade. Both use a page size of 1 KB. Assume DRAMs that are not active are in precharged standby and dissipate negligible power. Assume the time to transition from standby to active is not significant. Which type of DRAM would be expected to result in lower power? Explain why.

Question 5: Total Points (5+10=15)

Read the paper "**Future Cache Design using STT MRAMs for Improved Energy Efficiency: Devices, Circuits and Architecture**" and answer the following questions:

(a) What is STT MRAM and how is it different from SRAM?

(b) Evaluate STT MRAM and SRAM caches based on the following parameters:

- Cache utilization
- Energy consumption

Question 6: Total Points (10)

Read the paper "**Energy Reduction for STT-RAM Using Early Write Termination**" and elaborate how Early Write Termination (EWT) improves energy efficiency of STT-RAM cache?