

University of California at Berkeley
College of Engineering
Department of Electrical Engineering and Computer Sciences
Computer Science Division

CS 152
Spring 1994

J. Wawrzynek
Handout #22

Computer Architecture and Engineering
Midterm II

Your Name: _____

ID Number: _____

This is a *closed-book, closed-note* exam. You have 90 minutes. You might not have time to complete all the questions, so read them all first. Each question is marked with its number of points (one point per expected minute of time).

Show your work. Write neatly and be well organized.
Good luck!

problem	maximum	score
1	10pts	
2	15pts	
3	30pts	
4	15pts	
5	20pts	
total	90pts	

1. [10 points]

Consider the design of three separate caches, a directed mapped version (DM), a two-way set associative version (SA), and a fully associative version (FA). The total cache size in data words for each is 1K words and cache blocks (lines) contain four words of data. Assuming a 16-bit memory address, **fill in the table below** to indicate how many bits of the address to assign to each field.

	DM	SA	FA
byte in word (byte)			
(tag)			
set identifier (set)			
word in block (word)			

For each cache version, illustrate the breakdown of the address into fields. Label each field using the names in parenthesis in from the table above.

DM:

SA:

FA:

2. [15 points] Sketch a block diagram for a memory system with a *parallel* arrangement of the cache and TLB. Your diagram must include a block for each of the CPU, the main memory, TLB, and cache. Assume the cache is direct mapped and that virtual memory pages are 1K Bytes. Label all necessary connections.

In the above arrangement, what is the maximum size of a direct mapped cache?

What could you change to increase the cache capacity?

3. [30 points] Short answers. Please provide *short* answers to the following questions. If you do not know the answer, leave it blank. Correct answers are worth the number of points indicated by each question, blanks are worth zero, and incorrect answers minus the number of points indicated.
- (a) [1 points] What is the maximum theoretical speed-up for a n-stage pipeline processor over a “single-cycle” processor implementation?
 - (b) [2 points] List two reasons why in (a) the speedup will actually be less in practice.
 - (c) [1 points] For a standard MIPS implementation, approximately what percentage of branch delay slots would you expect could be filled?
 - (d) [2 points] Data forwarding can solve some data hazards in processor pipelines. Briefly, how does forwarding effect the processor cost?
 - (e) [1 points] True or False. A two-bit branch prediction scheme will always miss-predict less than a one-bit scheme.
 - (f) [1 points] Rank order the three cache organizations according to *miss rate*.
 - (g) [1 points] Rank order the three cache organizations according to *implementation cost*.
 - (h) [2 points] In what way does a system with a write-back cache achieve a performance advantage over a write-through cache?
 - (i) [1 points] In what way is a write-back cache more complex than a write-through cache?
 - (j) [1 points] True or False. Because of its simplicity, one-transistor Dynamic RAM cells are usually used for register file implementations.

- (k) [3 points] What is the approximate data bandwidth out of a frame buffer to feed a 1K X 1K pixel color display device?
- (l) [2 points] What are three factors that contribute to the time needed to access a block from a magnetic disk?
- (m) [1 points] True or False. Unlike ethernet, token-ring networks use no arbitration scheme.
- (n) [2 points] What is the primary advantage of asynchronous bus signalling over synchronous?
- (o) [1 points] True or False. The primary disadvantage of a vector processor over a super-pipelined processor is that the vector processor requires multiple execution units.
- (p) [2 points] List two limiters of super-scalar processor performance.
- (q) [2 points] List two limiters of super-pipelined processor performance.
- (r) [1 points] Name one commercial SIMD parallel computer system.
- (s) [2 points] True or False. The theoretically lowest diameter network for multiprocessor is a hypercube topology.
- (t) [1 points] True or False. All massively parallel processors use a *message passing* model for inter-processor communication.

4. [15 points] Consider a 100-MHz processor with a write-back/write-allocate cache connected to the memory system. Suppose we have a memory system that transmits or accepts 8-word read and write requests at the rate of one word per cycle. For reads or writes the accesses occur as follows:
- (a) 1 cycle to accept the address,
 - (b) 1 cycle of latency, and
 - (c) 8 cycles to transmit or accept the 8 words.

Assume the system uses a unified instruction and data cache and the following measurements have been made:

- The cache miss rate is 10% misses per instruction.
- 50% of the misses involve dirty cache lines.

Assuming the processor is stalled for the duration of a miss, find the number of cycles per instruction spent handling cache misses:

Assuming an average CPI of 4.0, what is the minimum bus bandwidth in MBytes/second needed to prevent additional processor stalls?

- (d) Would you expect that the average CPI would go up or down with this change to the pipeline? **Justify your answer.**