University of Pittsburgh
Department of Computer Science

CS2410 – Computer Architecture

Assignment #3
(Due on Monday November 7)

NOTE: (1) This is an individual assignment. (2) No late submission is accepted.

PART I [60]

1. [20]

Consider the following four data cache configurations, labeled A, B, C, and D. Determine (1) which address bits are used for indexing and tagging and (2) the size of the tag and data array(s) for each configuration. Fill the blanks of the following table. Assume a 64-bit (byte) address and no virtual memory. Use the address notation A[N:M], where N>=M and 0<=N,M<=63, to describe which address bits are used for indexing or tagging. For example, A[63:0] denotes the whole 64-bit address. A[12:4] shows the address bits from the bit position 4 (LSB) to 12 (MSB), 9 bits total. All the cache configurations use the write-back policy and thus each cache block has a "D" (dirty) bit, as well as a "V" (valid) bit. Include these two bits when computing # of bits in the tag portion of the cache.

Configuration	A	B	C	D
Cache Size	32kB	64kB	16kB	8kB
Block size	16B	128B	32B	64B
Associativity	Direct mapped	8-way associative	16-way associative	Fully associative
Address bits used for indexing				
Address bits used for tagging				
total # of bits in all tag arrays				
total # of bits in all data arrays				

2. [20] Suppose that you have a 4-way associative cache. In one of the sets of the cache, you see a stream of accesses to the following block addresses: A B C D E A G H F G G C D F G E B A. Determine whether each of the accesses would result in a cache hit, a cache miss, or a cache miss and replacement. Show what block is replaced for each of the following replacement policies: LRU, OPT (MRU–most remotely used), and FIFO.

In each table entry, use the following notation: "H" for hit, "M" for miss, "MR-x" for miss and replace block x.

Accessed Blk	LRU	OPT	FIFO
A			
B			
C			
D			
E			
A			
G			
H			
F			
G			
G			
C			
D			
F			
G			
E			
B			
A			

3. [20] Let's try to show how you can make "unfair" benchmarks. Here are two machines with the same processor and main memory but different cache organizations. Assume the miss time is 10 times a cache hit time for both machines. Assume writing a 32-bit word takes 5 times as long as a cache hit (for the write-through cache), and that writing a whole 16-byte block takes 10 times as long as a cache read hit (for the write-back cache). The caches are unified; that is, they contain both instructions and data.

Cache A: 64 sets, 2 blocks per set, each block is 16 bytes, and it uses write through.

Cache B: 128 sets, 1 block per set, each block is 16 bytes, and it uses write back.

(a) [7] Describe a program that makes machine A run as much faster as possible than machine B. (Be sure to state any further assumptions you need, if any.)

(b) [7] Describe a program that makes machine B run as much faster as possible than machine A. (Be sure to state any further assumptions you need, if any.)

(c) [6] Approximately how much faster is the program in Part (a) on machine A than machine B, and approximately how much faster is the program in Part (b) on machine B than machine A?

PART II [40]

Read the following two papers (posted on the course web page): (1) Jouppi and (2) Jin and Cho.

Briefly summarize the papers first. When summarizing a paper, be specific about the points made by the authors. For example, rather than saying "I find the paper interesting and it has lots of information", say "The authors compare two branch prediction schemes: single-bit and two-bit. The single-bit scheme ... The two-bit scheme differs from the single-bit scheme in that ...".

Second, discuss new (interesting) things that you learned.

Finally, describe complexities that may arise to the victim cache and macro data load scheme with the introduction of simultaneous multithreading.

Submit your work at the class or directly to the mailbox of the instructor (box #276), located in the mail room on 5th floor, SENSQ.