Part II Homework Problems: 50 points

1. Problem 5.3 from the book (15 points)

At power on the cache is completely empty.

5.3 For a direct-mapped cache design with a 32-bit address, the following bits of the address are used to access the cache.

Tag	Index	Offset
31–10	9–5	4–0

5.3.1 [5] <§5.3> What is the cache block size (in words)?

5.3.2 [5] <§5.3> How many entries does the cache have?

5.3.3 [5] <§5.3> What is the ratio between total bits required for such a cache implementation over the data storage bits?

Starting from power on, the following byte-addressed cache references are recorded.

Address											
0	4	16	132	232	160	1024	30	140	3100	180	2180

5.3.4 [10] <§5.3> How many blocks are replaced?

5.3.5 [10] <§5.3> What is the hit ratio?

5.3.6 [20] <§5.3> List the final state of the cache, with each valid entry represented as a record of <index, tag, data>.

Answers:

1. 8 words
2. DM cache has 32 entries
3. 22 bits for the tag + 1 dirty bit + 1 valid bit = 24 bits per entry. The directory part of the cache takes 32*24 bits. The data part of the cache size is 32*32 = 1K bytes.
   Ratio = (1 K byte * 8 bits/byte + 32*24)/ (1 K byte * 8 bits/byte) = 1.09375
4. And 5 below in table

Dec. Addr	Binary Address	Offset	Index	Index (Dec)	Tag	Hit/ Miss	Replace(True/ False)
0	0000000000000000000000 00000 00000	00000	00000	0	0000000000000000000000	M	0

4	0000000000000000000000 00000 00100	00100	00000	0	0000000000000000000000	H	0
16	0000000000000000000000 00000 10000	10000	00000	0	0000000000000000000000	H	0
132	0000000000000000000000 00100 00100	00100	00100	4	0000000000000000000000	M	0
232	0000000000000000000000 00111 01000	01000	00111	7	0000000000000000000000	M	0
160	0000000000000000000000 00101 00000	00000	00101	5	0000000000000000000000	M	0
1024	0000000000000000000001 00000 00000	00000	00000	0	0000000000000000000001	M	1
30	0000000000000000000000 00000 11110	11110	00000	0	0000000000000000000000	M	1
140	0000000000000000000000 00100 01100	01100	00100	4	0000000000000000000000	H	0
3100	0000000000000000000011 00000 11100	11100	00000	0	0000000000000000000011	M	1
180	0000000000000000000000 00101 10100	10100	00101	5	0000000000000000000000	H	0
2180	0000000000000000000010 00100 00100	00100	00100	4	0000000000000000000010	M	1

4 hits/12 reads => 0.33 hit rate.

4 Replaces

2. Cache and Virtual Memory (35 points)

Consider a processor with

- 2 Mbyte virtual address space
- 128 Kbyte pages
- 512 Kbytes of memory
- 8 Kbyte direct mapped cache with 512 bytes/ line and a 4 entry fully associative TLB
- 32 Kbyte set associative L2 cache with 2048 bytes/ line, and associativity of 2
- LRU bit is set if the lower line in the set is the least recently used

The state of the memory hierarchy is as shown. Answer the questions that follow with respect to this memory system.

TLB

Valid	Dirty	Virtual Page No.	Physical Page No.
1	1	C	01
1	1	0	11
1	0	1	00
1	1	D	10

TLB LRU Stack
C
0
D
1

Page Table

Valid	Dirty	Physical Page Address
1	1	11
1	0	00
1	1	01

1	1	10

	L1 Cache
Line 0	22
	29
	0a
	16
	10
	20
	30
	3c
	1a
	18
	26
	30
	11
	2a
	10
Line F	33

	L2 Cache
Set 0	13
	09
Set 1	0c
	0a
Set 2	12
	15
Set 3	05
	00
Set 4	14
	07
Set 5	
	1d
Set 6	1e
	08
Set 7	06
	08

L2 LRU Bit
0
0
1
1
0
1
0
1

Main Memory
1
C
D
0

Mem LRU Stack
C
0
D
1

a) **(5 points) Show the breakdown of the (i) virtual address, (ii) physical address, (iii) physical address as interpreted by the L1 cache, and (iv) the physical address as interpreted by the L2 cache**

Virtual Address (21 bits wide)

4 bits	17 bits

Physical Address (19 bits wide)

2 bits	17 bits

L1 (19 bits wide)

6 bits	4 bits	9 bits

L2 (19 bits wide)

5 bits	3 bits	11 bits

b) (5 points) Treat the L2 cache as a main memory for the L1 cache, and consider the wide memory organization for L2 where the width of the bus is now 8 banks. L2 access time is 4 cycles for a bank. Compute the miss penalty from L2 to L1.

One cycle for address transfer, four cycles for memory access and one cycle for transfer of eight words back. The cache line is 128 words. Therefore we have Penalty = (1 + 4 + 1) * 16 = 96 cycles

c) (5 points) If virtual page 1 is replaced in memory, which lines in the L2 and L1 cache must be flushed to avoid false hits?

Virtual page 1 is in main memory page 00. Therefore all lines from this page will have the upper two bits of the physical address set to 00. The remaining bits can have any value. For L1 this leaves 4 bits in the tag. Thus any line whose tag matches 00XXXX in the L1 cache is from main memory page 00 and therefore from virtual page 1. Similar arguments show that the lines from 00 in the L2 cache correspond to those lines with tags 00XXX

d) (5 points) Compute the miss penalty in cycles for L2 assuming the main memory is organized with a single word wide bus and interleaved across 32 banks and it takes 20 cycles to access a word.

There are 512 words/cache line.

- One cycle for address
- 20 cycles for memory access time
- 32 cycles to transfer the 32 words to the L2

This is repeated 16 times. Penalty = (1 + 20 + 32) * 16 = 948 cycles

e) (5 points) Now suppose we have a page fault. The disk has an average seek time of 9 ms and rotates at 7200 rpm. The transfer rate out of the disk is 8 Mbytes/sec and the disk controller overhead is 1 ms. How long does it take to service a page fault?

Servicing implies transferring 1 page = 128 Kbytes from HD to memory

Access Time = Controller overhead +
Seek time +
Rotational delay (average ½ rotation) +
Transfer time

Access time = 1*10-3 + 9 *10-3 + (0.5 / (7200/60)) + 128 Kbytes/8192 Kbyes/sec

f) (5 points) Suppose we have a 32 x 32 matrix stored in row major form in memory. Consider the initial set of accesses to this matrix which are all of the elements along the main diagonal. How many compulsory misses are there in the L1 cache? How many compulsory misses are there in the L2 cache? Assume that the caches are initially empty. Show your work.

The L1 cache has 512 byte lines which will contain four rows (32 elements in each row at 4 bytes per element storage). Thus a sequential access of elements along the diagonal will produce a miss every four rows = 8 compulsory misses. The L2 cache has 2048 byte lines each of which can contain 16 lines of the matrix. Thus we will have 2 compulsory misses.

g) (5 points) If the local miss rates for the L1 and L2 caches are 1.8% and 36% respectively, and 34% of the instructions are load/store instructions, what is the increase in CPI due to cache misses. L1 miss penalty is 10 cycles. L2 miss penalty is 20 cycles.

The number of memory references/instruction is 1.34.

Each of these references will potentially add cycles as follows.

1.34 * (0.018 * L1 miss penalty + (0.018 * 0.036) * L2 miss penalty))