Exercises for Week 10
Textbook Sections 5.4 and 5.10

Question 1

Textbook Exercise 5.10 <§5.4>

In this exercise, we will look at the different ways capacity affects overall performance. In general, cache access time is proportional to capacity. Assume that main memory accesses take 70 ns and that 36% of all instructions access data memory. The following table shows data for L1 caches attached to each of two processors, P1 and P2.

	L1 size	L1 miss rate	L1 hit time
P1	2 KB	8.0%	0.66 ns
P2	4 KB	6.0%	0.90 ns

a. Assuming that the L1 hit time determines the cycle times for P1 and P2, what are their respective clock rates?

b. What is the Average Memory Access Time for each of P1 and P2 (in cycles)?

c. Assuming a base CPI of 1.0, what is the total CPI for each of P1 and P2? Which processor is faster? (When we say a "base CPI of 1.0", we mean that instructions complete in one cycle, unless either the instruction access or the data access causes a cache miss.)

For the next three problems, we will consider the addition of an L2 cache to P1 (to presumably make up for its limited L1 cache capacity). Use the L1 cache capacities and hit times from the previous table when solving these problems. The L2 miss rate indicated is its local miss rate. [Note that the L2 miss rate here is different from that in the exercise in the textbook. The value there is unrealistically high, and is probably an error.)]

L2 size	L2 miss rate	L2 hit time
1 MB	5%	5.62 ns

d. What is the AMAT for P1 with the addition of an L2 cache? Is the AMAT better or worse with the L2 cache?

e. Assuming a base CPI of 1.0 without any memory stalls, what is the total CPI for P1 with the addition of an L2 cache?

f. What would the L2 miss rate need to be in order for P1 with an L2 cache to be faster than P1 without an L2 cache?

g. What would the L2 miss rate need to be in order for P1 with an L2 cache to be faster than P2 without an L2 cache?

Question 2

Consider a 2-way set-associative data cache of size 128Kbytes with block size of 16 bytes. The cache uses LRU replacement strategy (assuming initially that entry 0 in a set is least recently used) and write-back/write-allocate write strategy.

a. Show how a 32-bit data address is subdivided into tag, index and offset fields, giving the number of bits in each field.

b. Starting from power on, the following byte-addressed cache references by the CPU are recorded:

Read	0x20080128
Read	0x2008012C
Read	0x20080130
Write	0xA1150120
Read	0xA1150128
Write	0x20080130
Read	0x07F00124
Write	0x96880120

Trace the accesses in the cache, showing for each access:

- the address offset, index and tag value
- whether the access is a hit or a miss
- any memory operations required

Question 3

Suppose a system has two processors P1 and P2, each with a write-back cache using an invalidation coherence protocol, as described in Figure 5.42. When a cache intervenes on another cache's memory read, the intervening cache also updates memory. Variables X[0] and X[1], both initially 0 in memory, are contained within the same block. The processors access the variables in the following order:

P1 reads X[0]
P2 reads X[1]
P1 writes X[0] = 9
P1 reads X[0]
P2 reads X[1]
P1 reads X[0]

Trace the bus actions performed in each cache to maintain coherence, showing the contents of the block in each cache and in memory after each access.