Memory hierarchy exercise problems – Tutorial (13th and 14th April 2018)

1) Principle of locality

Consider the following C code involving matrix computations where elements within the same row are stored contiguously.

```
for(i=0;i<4000;i++)
  for(j=0;j<8;j++)
    A[i][j] = B[j][0]+A[j][i];
```

a) How many 32-bit integers can be stored in a 16-byte cache line?
b) Which variables exhibit temporal locality?
c) Which variables exhibit spatial locality?

2) Direct-mapped and Associative caches

Consider the following sequence of 32-bit memory address references given as word addresses. 1, 134, 212, 1, 135, 213, 162,161, 2, 44, 41, 221

a) For each of these references, identify the binary address, the tag and the index given a direct-mapped cache with 16 one-word blocks. Also list if each reference is a hit or a miss assuming the cache is initially empty.
b) For each of these references, identify the binary address, the tag and the index given a direct-mapped cache with two-word blocks and a total size of eight blocks. Also list if each reference is a hit or a miss assuming the cache is initially empty.
c) Using these address references, show the final cache contents for a three-way set-associative cache with two-word blocks and a total size of 24 words. Use LRU replacement. For each reference identify the index bits, the tag bits, the block offset bits and if it is a hit or a miss
d) Using these address references, show the final cache contents for a fully associative cache with two word-blocks and total size of eight words, using LRU replacement. For each reference identify the index bits, the tag bits, and if it is a hit or a miss

3) Processor and Cache performance

Assume that main memory accesses take 70ns and that memory accesses are 36% of all instructions. 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
a)	P1	1KB	11.4%	0.62ns
	P2	2KB	8.0%	0.66ns
b)	P1	8KB	4.3%	0.96ns
	P2	16KB	3.4%	1.08ns

(i) Assuming that the L1 hit time determines the cycle times for P1 and P2 what are their respective clock rates
(ii) What is the AMAT for each of P1 and P2?

(iii) Assuming a base CPI of 1.0, what is the total CPI for each of P1 and P2? Which processor is faster?

Consider an addition of an L2 cache to P1. Use the L1 cache capacities and hit times from the above table for the below parts of this question.

	L2 size	**L2 miss rate**	**L2 hit time**
a.	512KB	98%	3.22ns
b.	4MB	73%	11.48ns

(iv) What is the AMAT for P1 with the addition of an L2 cache? Is the AMAT better or worse with the L2 cache?

(v) Assuming a base CPI of 1.0 what is the total CPU for P1 with the addition of an L2 cache?

(vi) Which processor is faster, now that P1 has an L2 cache? If P1 is faster, what miss rate would P2 need in its L1 cache to match P1's performance? If P2 is faster, what miss rate would P1 need in its L1 cache to match P2's performance?