N.B. Assignments are individual work, due as pdf format. Submit through avenue, by 3rd December at 11:59 PM. Your assignment should contain you name and student number and the name of the file should follow the naming convention, i.e., fistname_lastname_studentNumber_A4.pdf.

Assignment Question 1 (4 + 3 Marks): To 6.4

a) A thread switch should be much more efficient than a process switch, which typically requires hundreds to thousands of processor cycles while a thread switch can be instantaneous. Why?
b) Suppose you want to perform two sums: one is a sum of 20 scalar variables (assume can not parallelize), and one is a matrix sum of a pair of two-dimensional arrays, with dimensions 20 by 20. What speed-up do you get with 100 processors? Show the impact on speed-up if one processor's load is twice than all the rest.

Assignment Question 2 (2 + 2 Marks): To 5.1

The following code is written in C, where elements within the same row are stored contiguously.

```
int size = 10;
int outer, inner;
for(outer =1; outer< size; outer++)
        for(inner=1;inner<size; inner++)
                rMat[outer][inner]=lMat[outer][1] + rMat[inner][outer]
```

a) Which variables exhibit temporal locality?
b) Which variables exhibit spatial locality?

Assignment Question 3 (4 + 4 + 4 Marks): To 5.3

Below is a list of 32-bit memory address references, given as word addresses. 3, 180, 43, 2, 191, 88, 190, 14, 181, 44, 186, 253

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 8 blocks. Also list if each reference is a hit or a miss, assuming the cache is initially empty.
c) You are asked to optimize a cache design for the given references. There are three direct-mapped cache designs possible, all with a total of 8 words of data: C1 has 1-word blocks, C2 has 2-word blocks, and C3 has 4-word blocks. In terms of miss rate, which cache design is the best? If the miss stall time is 25 cycles, and C1 has an access time of 2 cycles, C2 takes 3 cycles, and C3 takes 5 cycles, which is the best cache design (**show the details calculation**)?

Assignment Question 4 (4 + 4 + 4 Marks): To 5.4

Below is a list of 32-bit memory address references, given as word addresses. 3, 180, 43, 2, 191, 88, 190, 14, 181, 44, 186, 253

a) 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 off set bits, and if it is a hit or a miss.
b) Show the final cache contents for a fully associative cache with one-word blocks and a total size of 8 words. Use LRU replacement. For each reference identify the index bits, the tag bits, and if it is a hit or a miss.
c) What is the miss rate for a fully associative cache with two-word blocks and a total size of 8 words, using LRU replacement? What is the miss rate using MRU (most recently used) replacement? Finally what is the best possible miss rate for this cache, given any replacement policy?