Homework #5 CS 1645/2045
Due: February 19, 2018 Spring 2017

Question 1:
Consider the summation of the elements of a very large vector on a 2 GHz processor. Each cache block in that processor has 4 words (16 bytes) and the memory latency in the system is equivalent to 80 processor clock cycles. Assume each element in the array is 4 bytes long.

a. Assume that the processor can execute one addition every clock cycle, what is the minimum memory bandwidth that will allow the processor to achieve peak performance in the case of perfect operand pre-fetching or unlimited multithreading?

b. If no pre-fetching or multithreading is performed, what is the maximum achievable performance (as a fraction of the peak performance the processor could achieve) even if we assume an infinite memory bandwidth?

Question 2:
Consider a system in which the CPU can execute at a peak performance of 100 MFLOPS. Assume a relatively small cache (say 100Kbytes) with 4-words cache lines (16 bytes), no pre-fetching capabilities and a delay of 240 nanoseconds to fetch a cache line from memory. Assume each element in the array is 4 bytes long.

a) What is the actual FLOPS resulting from the execution of the following code segment assuming that the array b[][] is stored row-wise in memory?

```
sum = 0;
  for (i = 0; i < 10000; i++)
    for (j = 1; j < 10000; j++)
      sum = sum + b[i][j] + b[i][j-1];
```

Note: that the size of the array b[][] is relatively large (~400 Mbytes). Count each addition as one operation (2 operations per cycle).

b) What would be the actual FLOPS if the i and j loops are interchanged?

Question 3:
The X-Y deterministic routing algorithm given in the slides is for routing on an NxN 2-dimensional mesh <u>without</u> wrap-around links. Guided by this algorithm, give an X-Y-Z deterministic routing algorithm for an NxNxN 3-dimensional mesh <u>with</u> wrap-around links.
Hint: Start by developing a routing algorithm in the X-Y-Z space, then think about how best to use the wrap-around links.

Question 4:

Suppose we have a traditional 2d mesh network of size NxN (also known as planar mesh).

a. Describe a mapping of an NxN 2d torus into this network with a load of 1.
    i. What is the congestion?
    ii. What is the dilation?
b. Describe the mapping of a $^{N}/_{2}$ x $^{N}/_{2}$ 2d torus onto the NxN 2d mesh, can you achieve better mapping than that found in part (a)? Justify your answer.