Homework Assignment #2

- DUE 5PM Thursday, February 19
- Objective:
  - Establish some concepts in correctness and memory hierarchy optimization
  - For CUDA examples, you do not need to execute the code.
- Turning in assignment:
  - Use the "handin" program on the CADE machines
  - Use the following command:

    "handin cs6963 hw1 <hwfile>"
  - The file <hwfile> should be a raw text file, or a PDF file.

Homework Assignment #2

Problem 1: Dependence Analysis, Distance/Direction Vectors

Consider the following loop nest:

```
for (j=0; j<n; j++)
    for (i=0; i<n; i++) {
        A[i][j] = A[i-1][j] + A[i+1][j-1];
        A[i-1][j+1] = A[i][j];
    }
```

Identify true, anti and output dependences for each pair of references and each loop in the nest.

Homework Assignment #2

Problem 2: Dependence Analysis, Distance/Direction Vectors

Construct all direction vectors for the following loop and indicate the type of dependence (true, anti or output) associated with each.

```
for (k=0; k<100; k++)
  for (j=0; j<100; j++)
    for (i=0; i<100; i++)
      A[i+1][j][k] = A[i][j][5] + c;
```

Provide a CUDA kernel function and its invocation that represents a reordering transformation that preserves the meaning of this sequential code.

Homework Assignment #2

Problem 3: Safety of loop reordering transformations and tiling

For the following code:

```
for (k=1; k<p; k++)
  for (j=1; j<m+1; j++)
    for (i=1; i<n+1; i++)  {
        A[i][j][k] = A[i][j-1][k] + A[i-1][j][k];
        B[i][j][k+1] = B[i][j][k] + A[i][j][k]
    }
```

a) Specify the dependences and their distances.

b) Is it legal to interchange I and K? If yes, show the new dependences; otherwise, show which dependences are violated.

c) For large values of p, m and n, show how you can use tiling to improve the memory behavior of this code. Provide CUDA code that employs tiling for the shared memory of an SM.

Homework Assignment #2

Problem 4: Utilizing the memory hierarchy

For the following sequential code, derive a CUDA program that preserves the sequential program's meaning. Assume a float variable is 4 bytes. Sizes for different portions of the memory hierarchy are as follows: 16KB shared memory,8KB constant cache, 256MB global memory, and use the latencies to these memories from Lecture 6. Your CUDA code should utilize the memory hierarchy as effectively as possible.

```
float result[64][64], input1[64], input2[64][64], temp[64];
for (i=0; i<64; i++) {
    temp[i] = 0.0;
    for (j=0; j<64; j++) {
      temp[i] += input2[i,j];
      result[i,j] = temp[i];
      for (k=0; k<64; k++)
         result[i,j] += input1[j] * input1[k];
```

Homework Assignment #2

Extra credit: Omega calculator

For the example in problem #2, use the Omega calculator to set up the dependence equations. What are the relations you provide to Omega.