Lab12 – CUDA environment

Compile a CUDA program.

```
Nvcc –o prog program.cu // ignore the warning error
```

```
Login ID csxx(xx:even): cudaSetDevice(0);
                (xx:odd): cudaSetDevice(1);
```

e.g.1 Check CUDA 1-d indices inside a CUDA kernel.

```
#include <stdio.h>

#define X 12
#define THREAD_X 4

__global__ void index(int *A)
{
    int i = blockDim.x*blockIdx.x+threadIdx.x;

    (1) A[i] = 123;
    (2) A[i] = i;
    (3) A[i] = gridDim.x;
    (4) A[i] = blockDim.x;
    (5) A[i] = threadIdx.x;
}

int main()
{
    int A[X], *A_d, i;

    dim3 dimBlock(THREAD_X);
    dim3 dimGrid(X/dimBlock.x);

    cudaSetDevice(0); // or 1
    cudaMalloc((void **)&A_d, sizeof(int)*X);

    for (i=0; i<X; i++)
        A[i] = -1;

    cudaMemcpy(A_d, A, sizeof(int)*X, cudaMemcpyHostToDevice);

    index<<<dimGrid, dimBlock>>>(A_d);

    cudaMemcpy(A, A_d, sizeof(int)*X, cudaMemcpyDeviceToHost);

    for (i=0; i<X; i++)
        printf("%d ", A[i]);
    printf("₩n");

    cudaFree(A_d);
}
```

e.g.2 Check CUDA 2-d indices inside a CUDA kernel.

```
#include <stdio.h>

#define X 9
#define Y 8

#define THREAD_X 3
#define THREAD_Y 2

#define A(i,j) A[i*Y+j]

__global__ void index(int *A)
{
    int i = blockDim.x*blockIdx.x+threadIdx.x;
    int j = blockDim.y*blockIdx.y+threadIdx.y;

    (1)  A(i,j) = threadIdx.x;
    (2)  A(i,j) = threadIdx.y;
    (3)  A(i,j) = blockIdx.y;
    (4)  A(i,j) = blockIdx.y;
    (5)  A(i,j) = gridDim.x;
    (6)  A(i,j) = gridDim.y;
    (7)  A(i,j) = blockDim.x;
    (8)  A(i,j) = blockDim.y;
    (9)  A(i,j) = i;
    (10) A(i,j) = j;
    (11) A(i,j) = i*Y+j;
}
int main()
{
    int A[X][Y], *A_d;
    int i, j;

    dim3 dimBlock(THREAD_X,THREAD_Y);
    dim3 dimGrid(X/dimBlock.x,Y/dimBlock.y);
    cudaSetDevice(0); // or 1

    cudaMalloc((void **)&A_d, sizeof(int)*X*Y);

    for (i=0; i<X; i++)
        for (j=0; j<Y; j++)
            A[i][j] = -1;

    cudaMemcpy(A_d, A, sizeof(int)*X*Y, cudaMemcpyHostToDevice);

    index<<<dimGrid, dimBlock>>>(A_d);

    cudaMemcpy(A, A_d, sizeof(int)*X*Y, cudaMemcpyDeviceToHost);

    for (i=0; i<X; i++) {
        for (j=0; j<Y; j++)
            printf("%d ", A[i][j]);
        printf("₩n");
    }

    cudaFree(A_d);
}
```

Ex. Complete the following CUDA program(vecadd.cu) that adds two vectors.

```
#include <stdio.h>

#define THREADS 10
#define N 10000

__global__ void vecAdd(float *A, float*B, float *C)
{
      // CUDA kernel
}

int main()
{
    float A[N], B[N], C[N], C2[N], *A_d, *B_d, *C_d;
    int i;

    dim3 dimBlock(THREADS);
    dim3 dimGrid((N+dimBlock.x-1)/dimBlock.x);
    cudaSetDevice(0); // or 1

    for (i=0; i<N; i++) {
        A[i] = i*2;
        B[i] = N-i;
        C2[i] = A[i] + B[i]; // C2 is used to check the results
    }
    // allocate array on device
    …
    // copy array from host memory to device memory
    …
    // call GPU function(vecAdd)
    …
    // Copy the result back
    …
    // Check the results
    for (i=0; i<N; i++)
        if (C[i] != C2[i]) printf("%d %f %f₩n", i, C[i], C2[i]);

    // free memory on the device
    …
}
```

Submit the program when you are done.