CUDA: GPU basics

1   Preliminary

To use CUDA, add `module load cuda` in your `.bashrc` or `.bash_profile`. Also remove the `module load gcc`.

To submit a CUDA job, use -lnodes=1:ppn=7:gpus=1

Processing on the GPU can be asynchronous, which can lead to time measurement of 0 seconds. Use `cudaDeviceSynchronize()` to ensure previously submitted task are complete.

2   Polynomial expansion

The problem is simple. It is to compute a polynomial function F of degree d:

$$F(x) = \sum_{i=0}^{d} a_i x^i$$

for an array of n values of x.

(The code for polynomial expansion on the CPU is provided.)

Question: Write a simple CUDA code that allocates and fill an array on the CPU and transfer it to the GPU. (Take array size as a parameter)

Question: Compute the polynomial expansion of each element of the array on the GPU. (Take block size and degree of the polynomial as a parameter.)

Question: Bring the results back on the CPU and confirm the GPU code is correct.

3   Measurements

Question: Measure the speedup of the CUDA code for polynomial function of degree 1, 10, 100, 1000, and for $n = 10^9$.

Question: Compare to the performance of the CPU implementation. Which is faster in which configuration. Why do you think?