An Introduction to HPC and Scientific Computing

CWM, Department of Engineering Science

University of Oxford

Karel Adámek

Practical 6: Introduction to the CUDA programming language

This practical will review the examples discussed in lectures 7 and 8. You should begin by reviewing the 'Hello world' example and then expand this so that you can study the scheduling of threads, warps and blocks on the GPU.

The second example to complete is based on vector addition. Here you will study different memory access patterns with the aim of understanding the importance of coalescing memory accesses.

Finally, you are to write a code that calculates the decimation of a series of integer numbers.

The learning outcomes of this practical are:

- To be become familiar with code structure when using GPUs.
- To know how to write GPU kernel code and how to configure GPU kernels.
- To understand the scheduling of threads, warps and blocks on GPUs.

All practical sessions for this course will be carried out on the University's ARCUS-HTC computer. To understand how to use ARCUS-HTC see the introduction video and for more details watch the video demonstrations in lecture 3. As a reminder:

1. Connect to the university VPN (instructions can be found on canvas).
2. For windows: open MobaXterm and open a ssh session.
3. Connect to "oscgate" by:
    a. Setting "*remote host*" to be oscgate.arc.ox.ac.uk
    b. Tick "*specify username*" and set it to teachingXY (where this is the username we have issued you with)
    c. This will open a shell, from here you can connect to "arcus-htc" using ssh as follows:

        ssh –CX teachingXY@arcus-htc

4. Or for mac: ssh -CY teachingxx@oscgate.arc.ox.ac.uk; ssh -CY teachingxx@arcus-htc

Instructions for this practical

1. If you have not done so clone the github repo for this CWM. To do this, at the command prompt type:

$ module load git
$ git clone https://github.com/wesarmour/CWM-in-HPC-and-Scientific-Computing-2020.git

If you have already cloned the repo, pull it again to ensure you are working with the most up-to-date codes and notes:

$ module load git
$ git pull

2. Next you should navigate to the examples directory in the prac6 folder:

   $ cd CWM-in-HPC-and-Scientific-Computing-2020/practicals/prac6/code

3. When logged into the Arcus-HTC head node, you will need to use the command

   module load gpu/cuda

The first part of this practical is to review the examples given in the lecture. These are in the 'code' directory of this practical. These examples are

1. helloworld,
2. helloworld_scheduling,
3. vector_addition,
4. vector_addition_memory.

Please follow instructions contained in the code files themselves.

Decimation

The second part of this practical is to write a code which calculates the decimation of a series of integer numbers.

The process of decimation adds neighbouring numbers in a series of length n to produce a new series of length $\frac{n}{2}$, An example of the algorithm is given in figure 1.

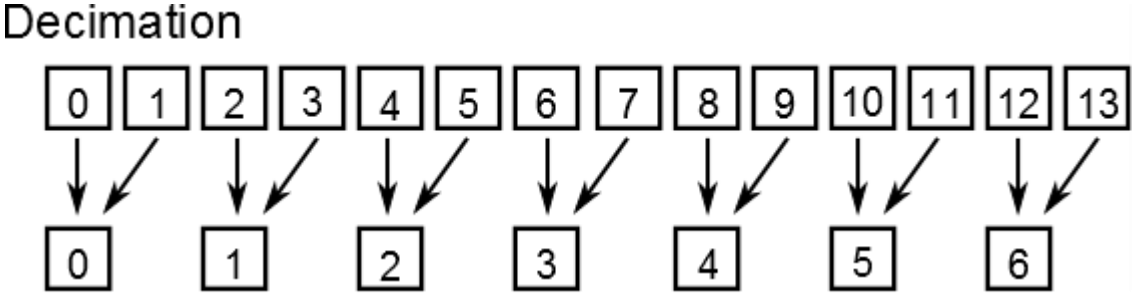

Figure 1: Decimation, two neighbouring numbers are added together to create new series of numbers (the numbers in this figure are the indices of the arrays).

For example, if we have a series of numbers {1,2,5,7,3,8} the decimation will perform {1+2, 5+7, 3+8}

$$\{1+2, 5+7, 3+8\} = \{3,12,11\}.$$

The code (with a description) and the specific tasks that need to be completed (detailed in the code itself) are contained in the 'code' directory for this practical (prac6).