

CP631 Parallel Programming

ASSIGNMENT 3

Due date: after the end of week 12, see Dropbox for exact date

Question 1

For this question submit a file **q1.cu** with the source code to Dropbox for question 1. Also, for discussion submit a PDF file **q1.pdf**

Looking for minimum distance between particles

In this question you will look for the minimum distance between two particles in a set of particles scattered randomly in 2D space.

First generate a set of random co-ordinates (x,y) for N particles, where x and y are between 0 and 1. Next:

Part 1: write a CPU serial function which tries all the possible distances between pairs of particles and finds the minimum distance.

Part 2: similarly write a GPU CUDA function which finds the minimum distance, using one thread per particle. Choose a reduction method for finding the minimum which is efficient.

Part 3: similarly write a GPU CUDA function which finds the minimum distance, using one thread per each pair of particles. Choose a reduction method for finding the minimum which is efficient.

Compare and comment on the efficiency of approaches in parts 1, 2, and 3.

Part 4: Discuss what improvements could be made to the simple algorithm to improve the speed of the calculation.

Note: For each of the approaches in part 1,2,3, increase N until your code for each part takes at most 30 seconds to run. Because code in each part will run at different speeds, the value of N will be different. You don't necessarily have to run all the parts for a given value of N, as for example for some large value the serial function would take too long.

Keep in mind that some atomic reduction function only work for integers, so you might have to convert your data from float to integer and back as needed.

Your code must be efficient to get full marks. So, for example, you should not do atomic reduction only as that will be very slow.

Question 2

Submit a single Jupyter notebook file to Dropbox with all the Python Spark source code, figures and discussion for this question, call it **q2.ipynb**

Do not submit your dataset to Dropbox, instead upload it to the course server and then mention the location in your notebook file.

Working with Spark

This is the last assignment question of the course and it is more loosely formulated. If you need more guidance about what you to do here, please use the discussion forum to ask the instructor.

Find some text data online that is of interest to you and download it. This can be a ready-made dataset, or you can use your own scraper program to collect the data. Try to keep the total data set size limited to less than 0.5 Gb by truncating as needed.

Once you have the data, use Spark read it in as an RDD and perform some analysis similar to what you have seen in Lesson 11 and 12.