ASSIGNMENT 3 Dsicussion

# Question 1

Part 4: Compare and comment on the efficiency of approaches 1, 2, and 3. Measure the runtime of the three functions for different numbers of particles. In your tests make N large enough so that the CPU version runs for up to 30 seconds.

*As can be obviously observed ion following char, as the NMAX doubles, the CPU times increase by n^2, which follows the time complexity estimation of serial code O(N^2). However, GPU codes running time keeps steady which shows powerful speedup effect of GPU in situation of large-scale calculation!*

|  |  |  |  |
| --- | --- | --- | --- |
| NMAX | CPU time(s) | GPU k1 time(s) | GPU k2 time(s) |
| 16384 | ~1.2 | 5\*10^-6 | 4\*10^-6 |
| 32768 | ~4.2 | 7\*10^-6 | 5\*10^-6 |
| 65536 | ~17.4 | 5\*10^-6 | 5\*10^-6 |
| 92682 | ~33.8 | 5\*10^-6 | 4\*10^-6 \*\* |

A computer screen shot of a black screen

Description automatically generated

A screenshot of a computer program

Description automatically generated

A computer screen with white text

Description automatically generated

A computer screen with white text

Description automatically generated

\*\* k2 code failed to get correct result for some unknown reason.

Part 5: Discuss what improvements could be made to the algorithm to improve the speed of the calculation.  How does the runtime of algorithms in 1,2 and 3 scale with the number of particles?  Could the algorithm be redesigned to obtain a different, better scaling?

*Possible improvement will be discussed in different aspects:*

* *Portion of parallelized code: Since the codes utilize the fastest algorithm discussed in Exercises week11, namely, atomic algorithm which maximize the chance of parallel execution of code. And the performance monitor above reflects the efficiency of GPU code.*
* *More efficient execution of code: as hinted, maybe we can use integer calculation to replace float-point calculation.*

*To sum up, there may be some space to improve the efficiency further, but the extent is limited because the performance of existing code is very good already!*

# Question 2

Since the limit of time, there’s more direction can be analyzed for sampling method:

I will keep updating on:

<https://github.com/daiqing2009/AdvParallelProgramming/blob/main/assignment3/question2/sampling_movietv.ipynb>

I use this as preprocessing of my Data Analysis project. Spark is really a powerful tool!