Parallelizing a Particle Simulation – Homework 3 Report

The goal of this assignment was to parallelize a particle simulator using CUDA on the Bridges2 supercomputer and to compare its performance against a serial baseline implementation. The simulator models particle interactions under the assumption that the number of effective interactions per particle remains O(1) due to limited interaction range and sufficient sparsity. The parallelization aimed to improve runtime efficiency by offloading the simulation computations to GPU accelerators.

To evaluate correctness and performance, the simulation was run for three particle sizes: 5000, 10000, and 20000. For each configuration, the simulation was executed three times to obtain an average runtime. We verified correctness using the provided autocorrect tool, which checks that the minimum particle distance throughout the simulation remains above 0.4 units. For all GPU runs, the minimum distance was well within the expected range (typically between 0.7 and 0.8), indicating that the simulation maintained valid particle interactions and passed the correctness criterion.

The compilation and execution of the serial code were straightforward. We used the command make serial to compile and executed the binary with /usr/bin/time -f "Time: %e seconds" ./serial -n <particles> -o serial_<particles>_<trial>.txt, replacing <particles> and <trial> with the appropriate values. The GPU version was compiled with make gpu after loading CUDA modules using module load cuda. To execute GPU jobs, we used the command srun --partition=GPU-small --gpus=1 --time=00:10:00 --pty bash, followed by ./gpu -n <particles> -o gpu_<particles>.txt. We verified correctness for each output using ./autocorrect -s gpu_<particles>.txt.

The results demonstrated a strong performance advantage of the GPU implementation over the serial version. For 5000 particles, the average serial runtime was 54.07 seconds, while the GPU runtime averaged just 0.465 seconds, yielding a speedup of approximately 116.31x. At 10000 particles, the serial runtime averaged 219.01 seconds and the GPU runtime was 1.887 seconds, maintaining a similar speedup of 116.06x. For the largest test size of 20000 particles, the serial version averaged 879.51 seconds, whereas the GPU implementation completed in 7.552 seconds on average, for a speedup of approximately 116.46x. These consistent gains confirm that the CUDA implementation scales effectively and achieves the expected O(n) behavior, as required.

In conclusion, the assignment successfully demonstrates the power of GPU acceleration for computational simulations with regular interaction patterns. The CUDA implementation passed all correctness checks and exhibited excellent scalability and performance gains across increasing input sizes. The assignment directory contains all source files, output logs, and verification results, and has been packaged for submission using the command tar -cvf hw3.tar *.

This report summarizes the methods, commands used, results obtained, and conclusions drawn from the assignment, meeting all stated requirements and performance goals.