Heidelberg University
Institute for Computer Engineering (ZITI)

Master of Science Computer Engineering

GPU Computing

Exercise 7

gpucomp03

Benjamin Maier
Daniel Barley
Laura Nell

Due date January 26th, 09:00

7.1 Naive GPU implementation

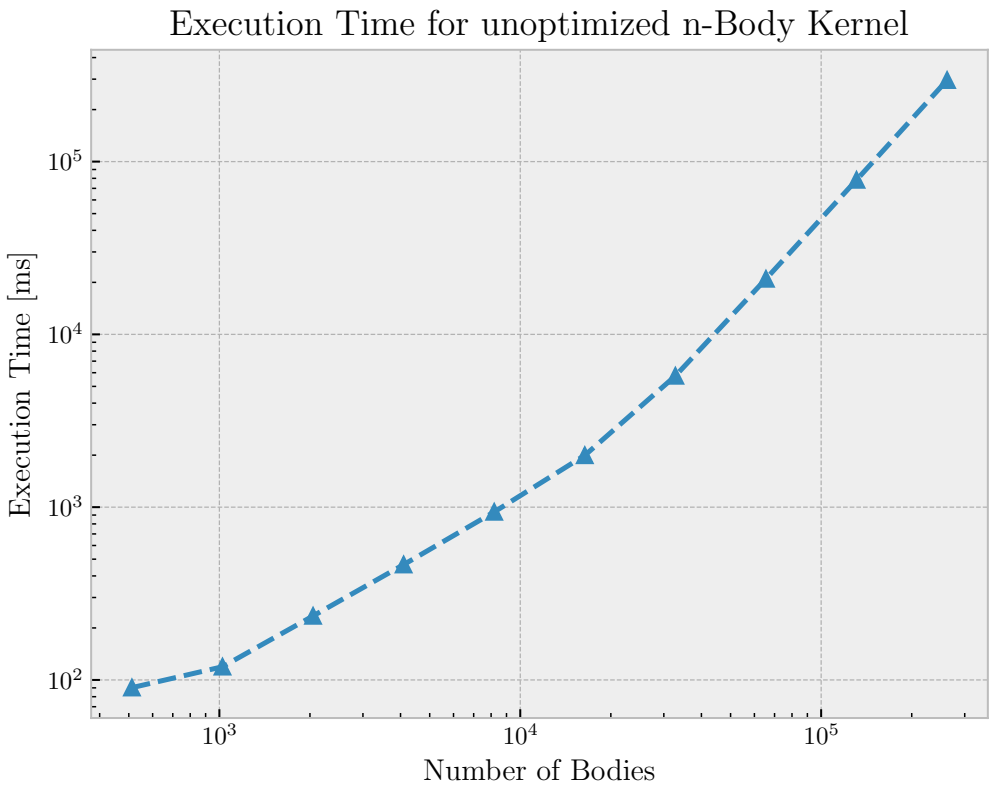

Figure 1: Execution times for unoptimized n-body kernel

For the first exercise we implemented a naive kernel for a n-Body simulation. The measurements are shown in Figure 1. We can definitely see the $\mathcal{O}(N^2)$ scaling behavior of the algorithm.

7.2 Optimized GPU implementation

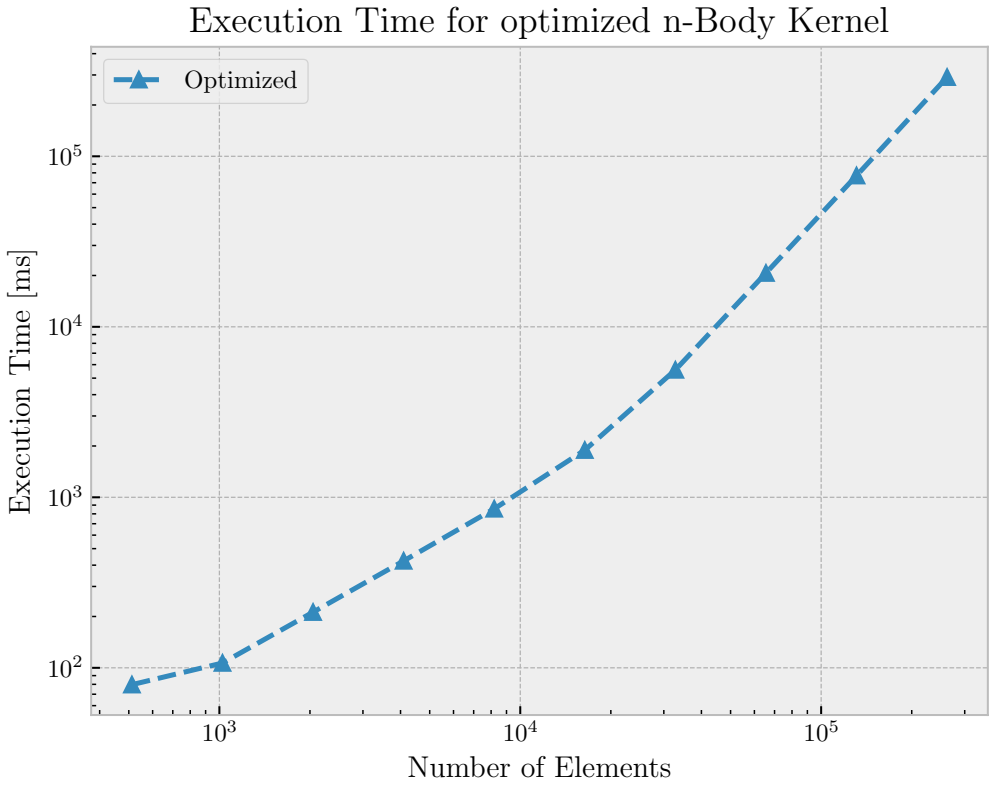

Figure 2: Execution times for unoptimized n-body kernel

For the optimized implementation we decided to use structure of arrays for the better access pattern and tried to use tiling to optimize for shared memory use. We also were trying to use loop unrolling, but this always decreased our performance. The results are shown in Figure 2 and 3. We can only observe an increasing performance for smaller number of bodies. Maybe the additional writing in shared memory and synchronization steps are limiting the performance.

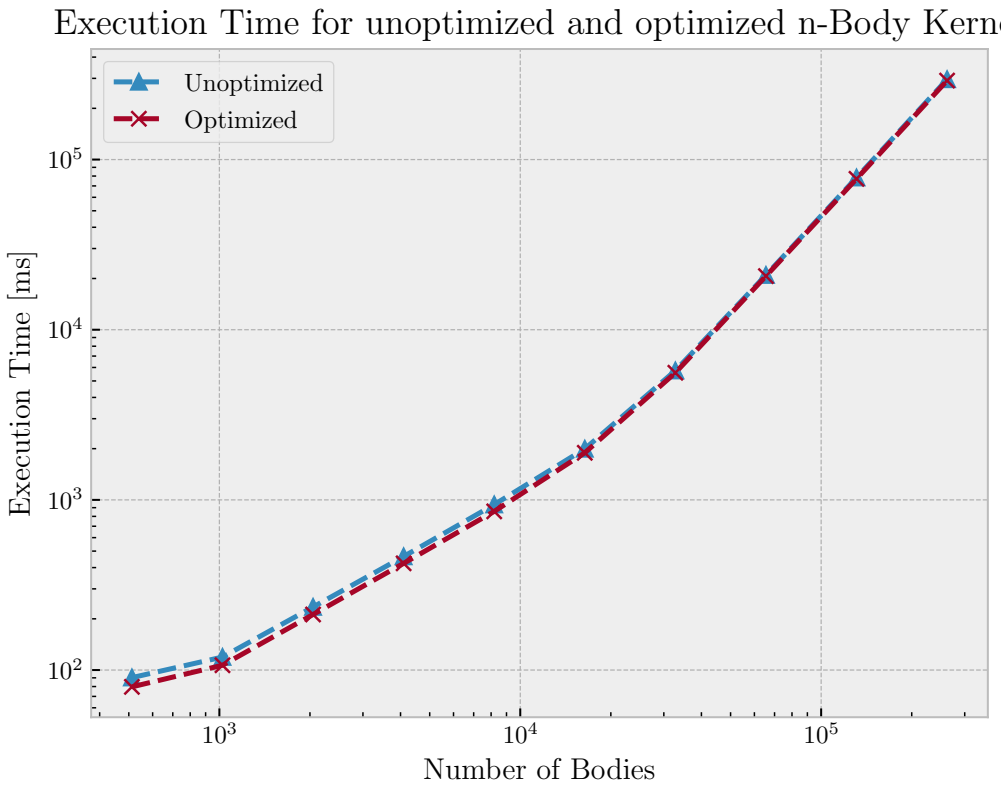

Figure 3: Comparison of execution times for optimized and unoptimized versions

7.3 n-Body GPU computations and streaming

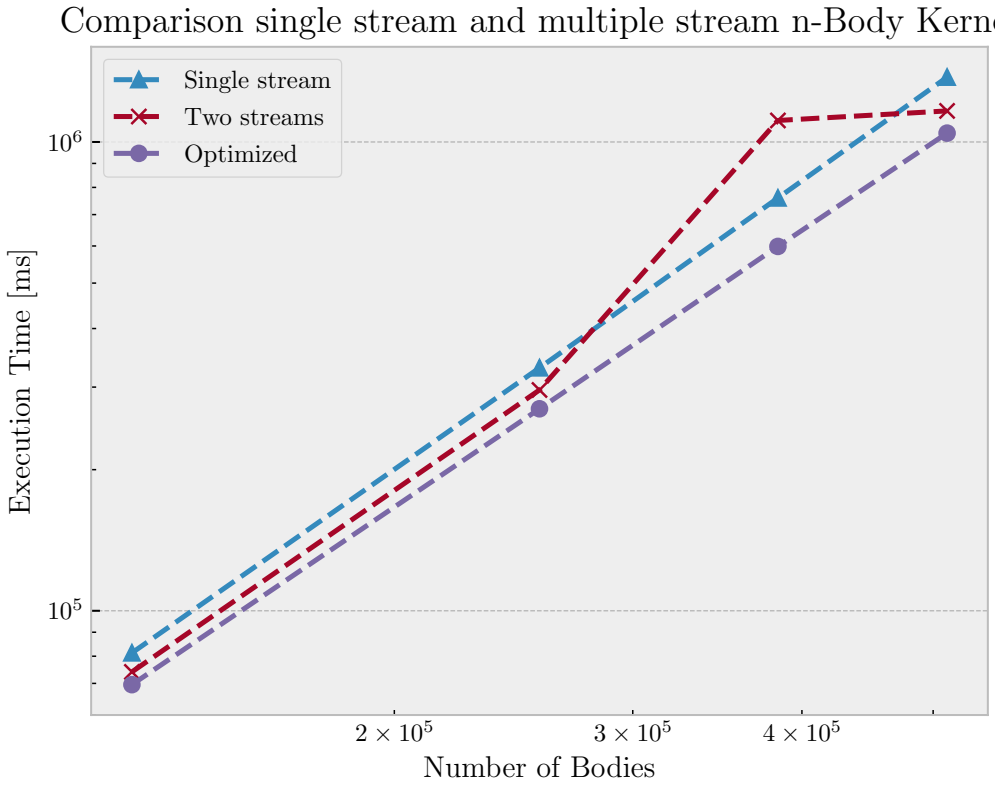

Figure 4: Execution times for n-body kernel using one and two streams

For the last part we artificially limited the memory of the GPU to 4MB or to be more precise to hold a total amount of 128k bodies. To solve larger problems the host memory is acting as a swap space. To overlap the kernel execution with the memcpy operation we were using two streams, as more streams would split up the memory in even smaller parts. We were varying the segment size of our device memory between the persistent data and the swapped data but did not observe any big difference. The results are shown in Figure 4. As one can see the optimized version from exercise 7.2 is faster since there is no memory limit. However, even the single streamed version is faster than our two stream implementation, which is bit strange, but for a large amount of bodies the two stream version gets faster.

7.4 Willingness to present

Hereby, we declare our will to present the results shown in the former sections.