CEG 4136 Computer Architecture III (3 units)
Fall 2021

Laboratory #2

Due: November 2, 2021
Goals: The goal of the lab is the design and implementation of the sum of N elements using GPU parallel reduction using the oneAPI DPC++ platform.

1. Details:

A reduction is an operation that is provided in pairs to each of the elements. A typical operation is the parallel sum. Now suppose you wanted to calculate the sum of all the elements in the array. If you, did it sequentially, you would write the sum as, you know, the sum of the first two elements, plus another element, plus another element, and so on. You could present it as Figure 1.

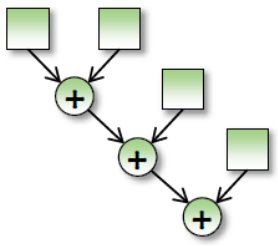

Figure 1

This is not the right method for parallelization because each subsequent operation depends on the previous operation. So, it looks like they are all related, but in this case, we are lucky because the addition operator is associative so we can continue to subdivide the problem space. So initially, if instead of calculating the sum of 1 to N, we can calculate the sum of 1 to N/2 and add it to the sum of all the elements from N/2 to N. We can kind of go on and subdivide it until we get something that looks a lot like a binary tree. So, something where all the operations happen in pairs and then they are matched higher up the tree to the root, to the final addition operation. So, in practice, the parallel reduction of the GPU would look like figure 2 shows us.

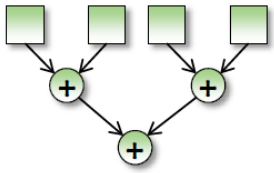

Figure 2

Figure 3 shows an example of how to calculate the sum of 8 elements using parallel reduction.

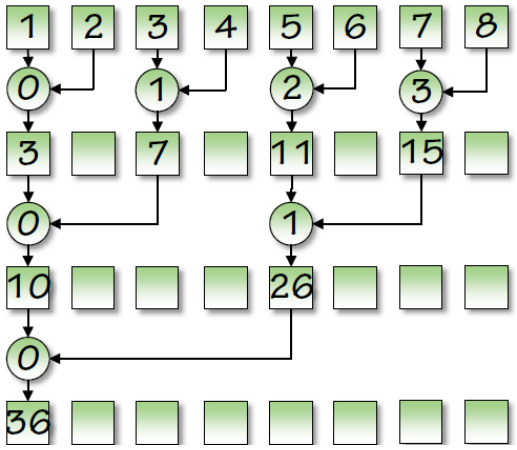

Figure 3

3. Deliverables

a. Brief description of the purpose and theory of the problem.
b. Brief explanation of your solution algorithm.
c. Amdahl's law defines the speed gain that can be obtained by using the GPU, multiple processors, or other execution accelerators. What is the speed gain compared to a single processor?
d. Design document using UML diagrams.
e. Screenshots of the demonstration for each application
f. Discussion
g. Conclusion

Evaluation

- The proper functioning of the executable code: 40%.
- The report: 60%.