

Practical 8: scan operation and recurrence equations

Lecture 4 explains how to perform scan operations, and also outlines how to extend the implementation to solve long recurrence equations. This practical starts by providing you with an implementation of the scan routine for a single thread block. You are then to extend it to multiple thread blocks, and also the parallel solution of a long recurrence equation.

1. Click on the link in the course webpage to the Google Colab notebook.
2. Carefully follow the instructions in the notebook.
3. The application `scan` performs an addition scan operation using a single thread block, reading in the input data from device memory, and putting the output (which is the sum of the preceding input elements) back into device memory.

Read through the code and understand what it is doing.

4. Extend the implementation to multiple thread blocks using either of the approaches described in the lecture. If you have time, perhaps you can do both and compare the execution times.
5. Modify your code to use shuffles instead for the scan within each block.
6. Following the mathematical description in lecture 4, modify the scan routine to handle recurrence equations, given input data s_n, u_n .
7. If you have time and interest, you could read about [parallel radix sort](#) (which needs a prefix-scan) and perhaps implement it.

Suggestion: do 1 bit at a time to simplify the programming. For better performance, I would probably do 4 bits at a time, which requires $2^4 - 1 = 15$ scans to be performed at each stage. Discuss with me if you would like to understand this better.