CUDA Programming on NVIDIA GPUs
Mike Giles

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.

- Make and run the application `scan`.

  This 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.

- 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.

- Following the mathematical description in lecture 4, modify the scan routine to handle recurrence equations, given input data s_n, u_n.

- If you have time, you could modify the implementation to handle longer scans (or recurrence equations) within a single thread block, by first doing a scan on the first 128 elements (for example) then the next 128 elements, and so on. This might be the best approach if there are lots of scans to be performed (or recurrence equations to be solved) so each thread block is responsible for one of them.