Problem 7

Even Flørenæs

Fall 2016

TDT4200 Parallel Computing
Department of Computer Science

Contents

1 Theory **3**
1.1 Key differences between CUDA and OpenCL 3
1.1.1 Naming differences 3

1 Theory

1.1 Key differences between CUDA and OpenCL

Programming in CUDA is restricted to building application for NVIDIAs GPUs. OpenCL can be used to program any heterogenous platform e.g GPUs or FPGAs. OpenCL is much more verbose than CUDA.

1.1.1 Naming differences

CUDA	**OpenCL**
Thread block	Work-group
Thread	Work item
Shared memory	Local memory
Local memory	Private memory
Multiprocessor	Compute unit

It isn't defined in the OpenCL standard. A warp is a thread as executed by the hardware (CUDA threads are not really threads and map onto a warp as separate SIMD elements with some clever hardware/software mapping). It is a collection of work-items and there can be multiple warps in a work-group.

An OpenCL subgroup was designed to be compatible with a hardware thread, and hence is able to represent a warp in the OpenCL kernel, but it is entirely up to NVIDIA to decide to implement subgroups or not and of course an OpenCL subgroup cannot expose every feature that NVIDIA can expose for warps because it is a standard, while NVIDIA can do anything they like on their own devices.