GPU Teaching Kit

Accelerated Computing

Module 20 – Related Programming Models: OpenCL
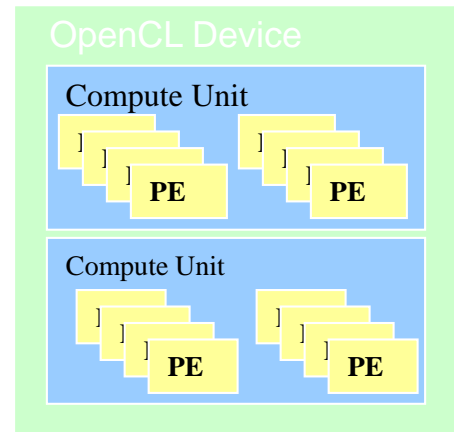
Lecture 20.2 - OpenCL Device Architecture

Objective

– To Understand the OpenCL device architecture
  – Foundation to terminology used in the host code
  – Also needed to understand the memory model for kernels

OpenCL Hardware Abstraction

– OpenCL exposes CPUs, GPUs, and other Accelerators as "devices"

– Each device contains one or more "compute units", i.e. cores, Streaming Multicprocessors, etc...

– Each compute unit contains one or more SIMD "processing elements", (i.e. SP in CUDA)
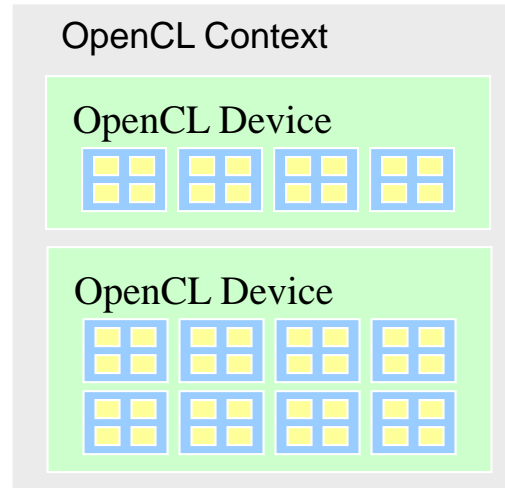
OpenCL Device Architecture
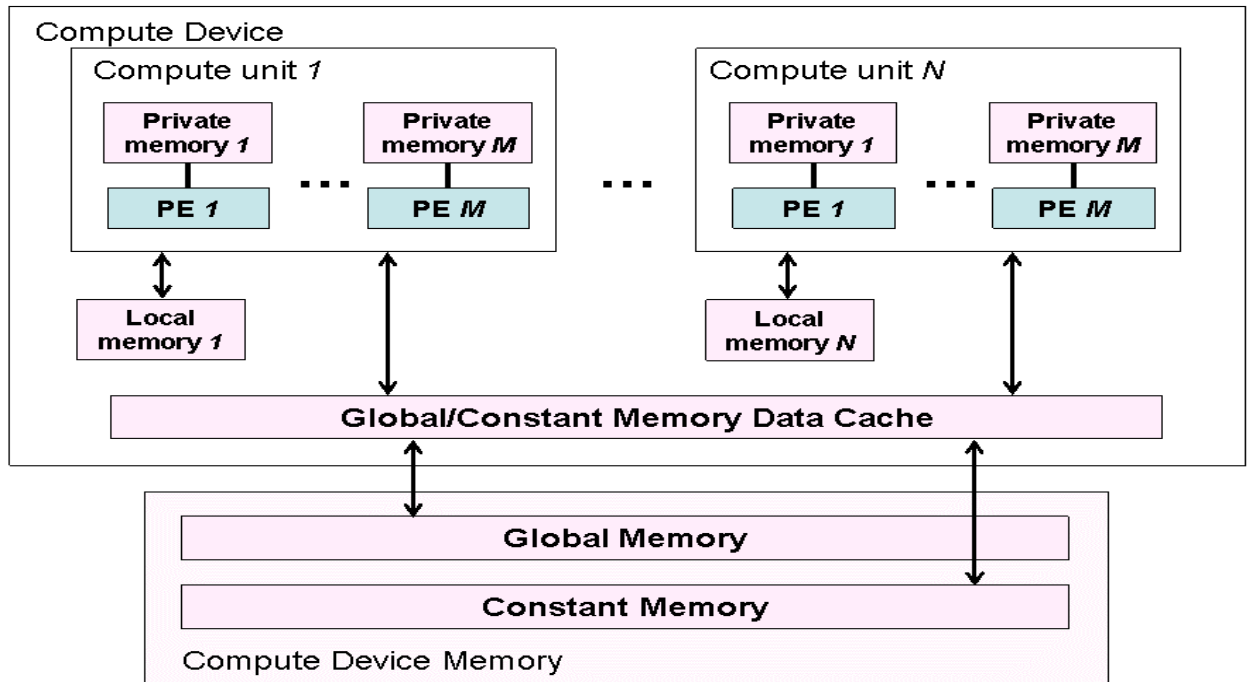
OpenCL Device Memory Types

Memory Type	Host access	Device access	CUDA Equivalent
global memory	Dynamic allocation; Read/write access	No allocation; Read/write access by all work items in all work groups, large and slow but may be cached in some devices.	global memory
constant memory	Dynamic allocation; read/write access	Static allocation; read-only access by all work items.	constant memory
local memory	Dynamic allocation; no access	Static allocation; shared read-write access by all work items in a work group.	shared memory
private memory	No allocation; no access	Static allocation; Read/write access by a single work item.	registers and local memory

OpenCL Context

– Contains one or more devices
– OpenCL device memory objects are associated with a context, not a specific device

GPU Teaching Kit

Accelerated Computing