

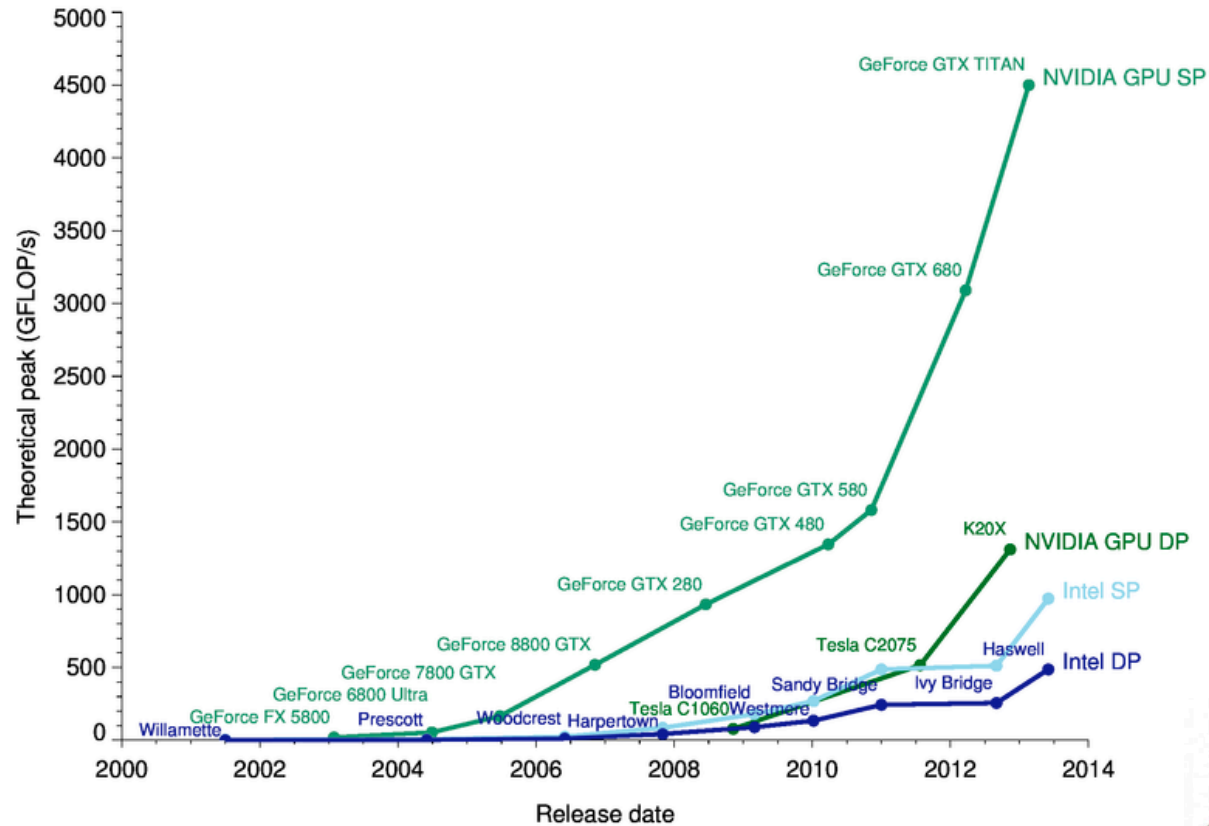
Uporaba GPU

Alen Orbanić,
Abelium d.o.o.

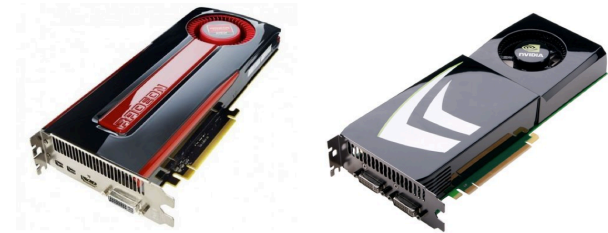
Abelium d.o.o.

- Ustanovljeno 2009
- Jedro tvorijo matematiki in računalničarji
- 15+ zaposlenih, 7+ doktorjev znanosti
- Vpetost v industrijo in v akademsko sfero
- Naslavljamo zahtevnejše probleme v industriji na področju algoritmov, optimizacij, računalniškega vida, podatkovne analitike, pametih informacijskih sistemov, ...
- Več nagrad za inovacije, mednarodno sodelovanje, ...
- <http://www.abelium.eu>

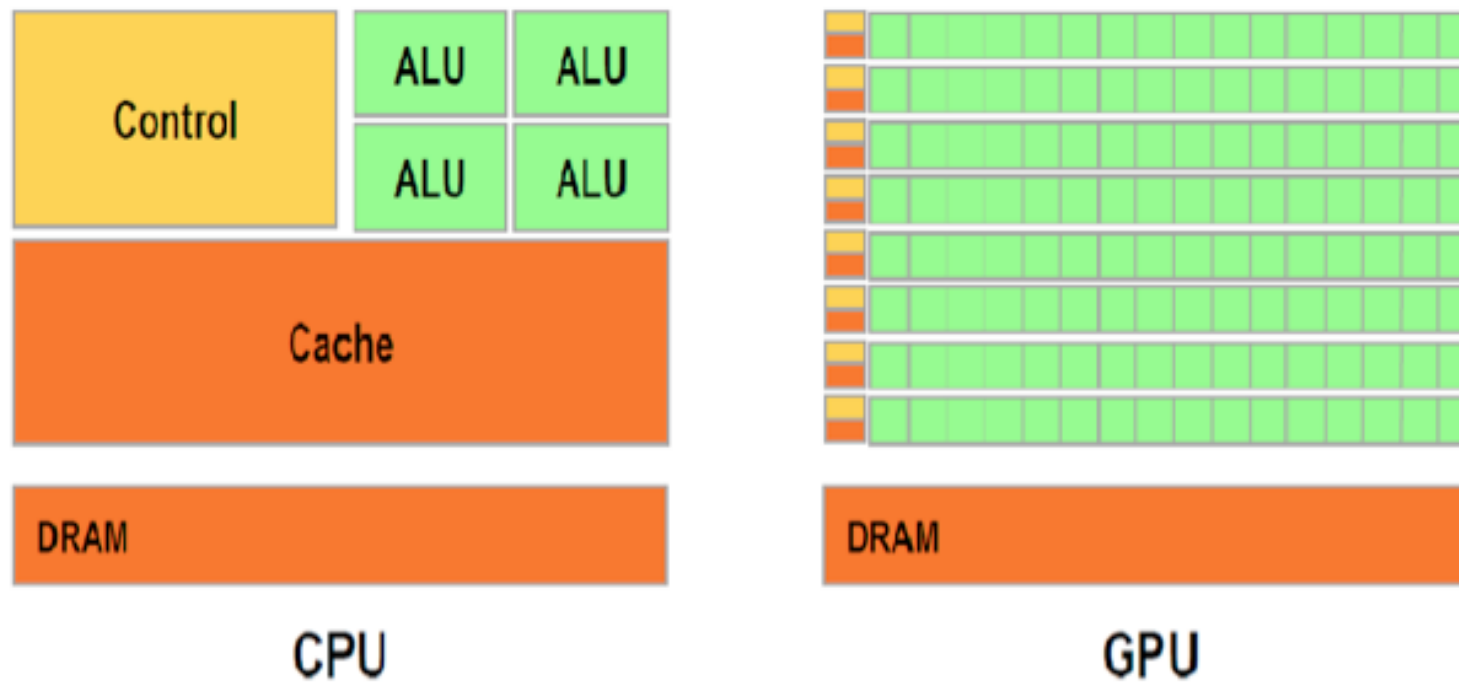
Motivacija



GPU – Graphics Processing Unit (grafični procesor)



CPU vs. GPU



Vir: http://www.training.prace-ri.eu/uploads/tx_pracetmo/gpu-talk-pdc.pdf

GPU

- Specializiran za masovno paralelne aritmetične operacije na vektorskih podatkovjih (SIMD – Single Instruction Multiple Data)
- Visoka aritmetična zmogljivost, podatkovna pretočnost, energijska učinkovitost
- Zahteva “lahke” niti
- Težje programiranje

Programske platforme

- CUDA - Compute Unified Device Architecture
 - Platforma in programski model
 - Samo na karticah podjetja NVIDIA
 - Nekoliko lažje programiranje
- OpenCL
 - Odprt standard
 - Na večih platformah (AMD, NVIDIA, Intel, ...)
 - Khronos Group
 - Podpora širokega nabora podjetij in naprav, tudi na mobilnih napravah

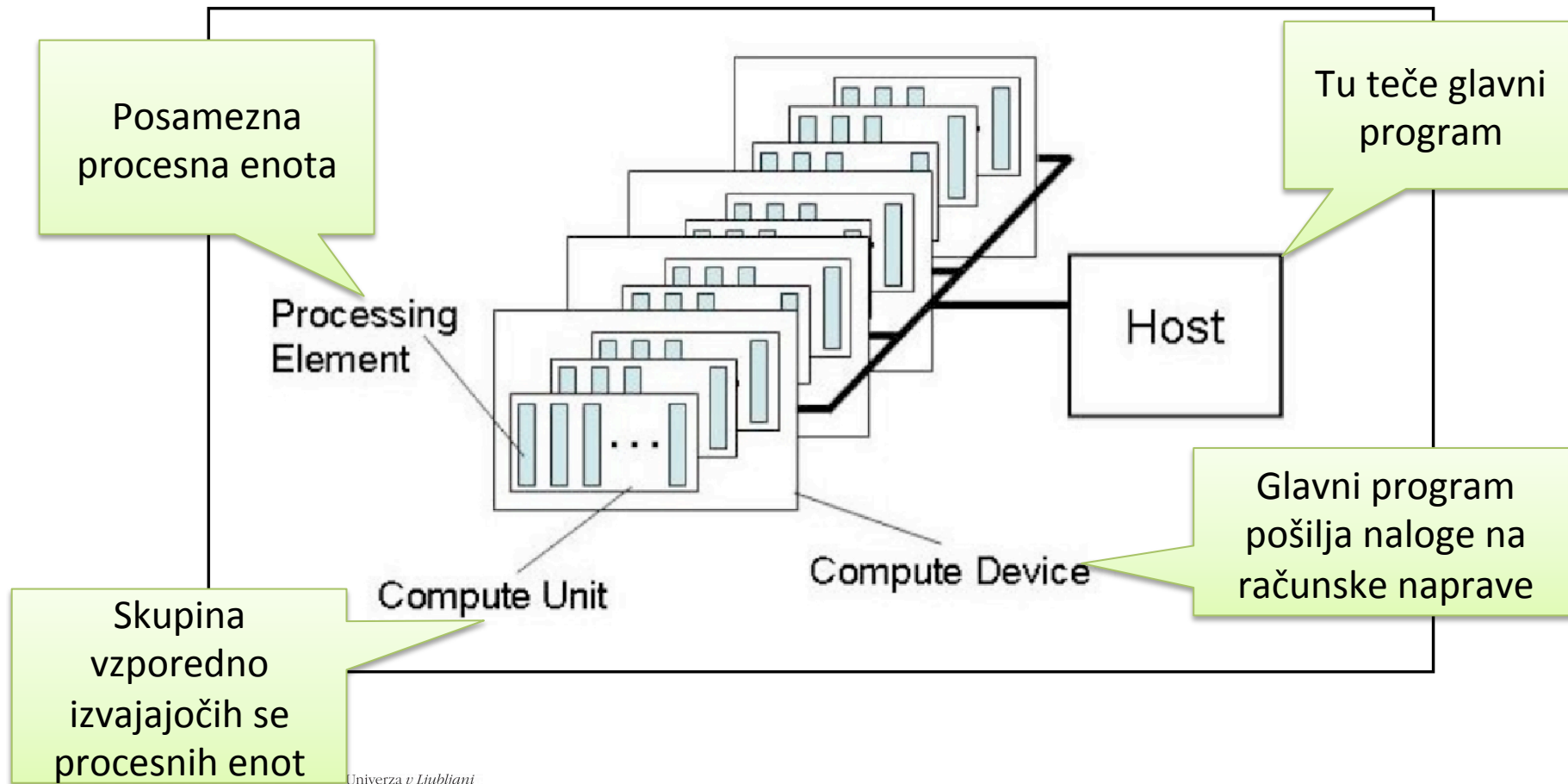


OpenCL

- Enoten pogled na vse računske vire (CPU + GPU)
- Podpora na različnih napravah
- Učinkovit programski model
 - Programski jezik temelji na okrnjenem C99
 - Podatkovno- in funkcijsko- paralelizabilen model
 - Abstraktizacija strojne opreme
 - Nastavljiva numerična natančnost (single, double)

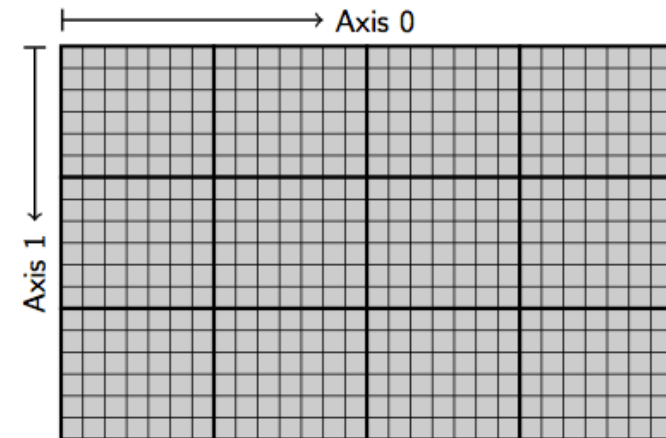


Računski model



Delitev dela

- Več dimenzionalno indeksiranje delovnih niti (work-items)
- Grupiranje v hkratno vzporedne enote – delovne skupine (work-groups)

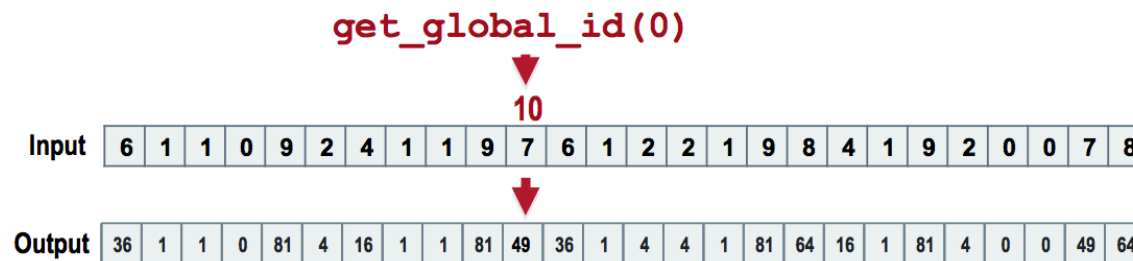


Software representation

Kernel

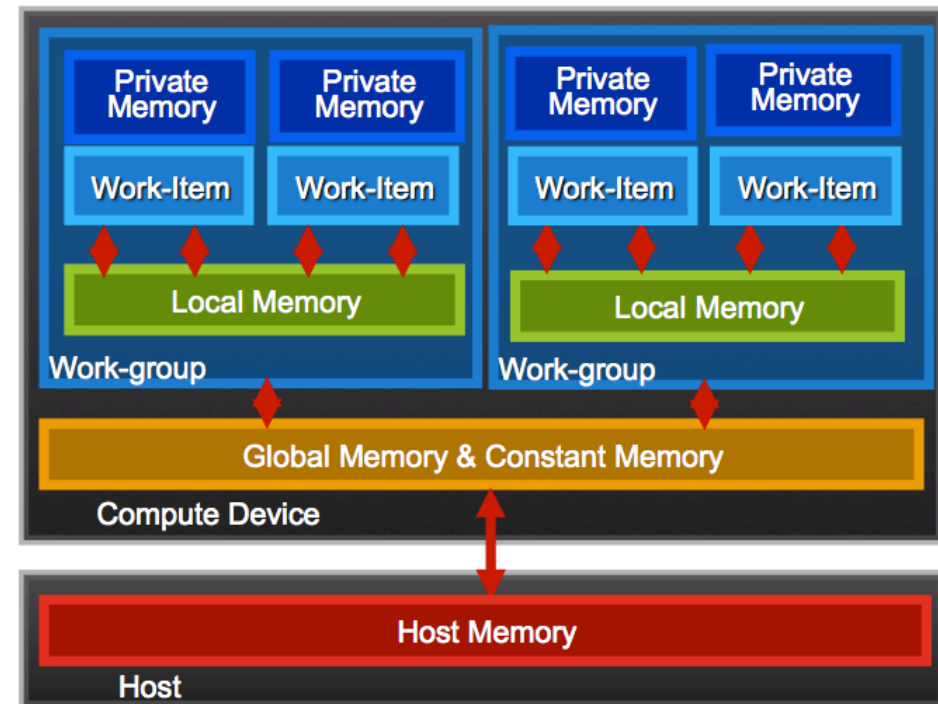
- Osnovna ideja: indeksirane (gnezdene) zanke izvajamo vzporedno, notranjost zanke opišemo s funkcijo (kernel)

```
kernel void square(
    global float* input,
    global float* output)
{
    int i = get_global_id(0);
    output[i] = input[i] * input[i];
}
```



Model spomina

- **Host** - spomin pri gostitelju
- **Global** – viden vsem delovnim skupinam
- **Local** – viden znotraj delovne skupine
- **Private** – lokalne spremenljivke posameznih niti



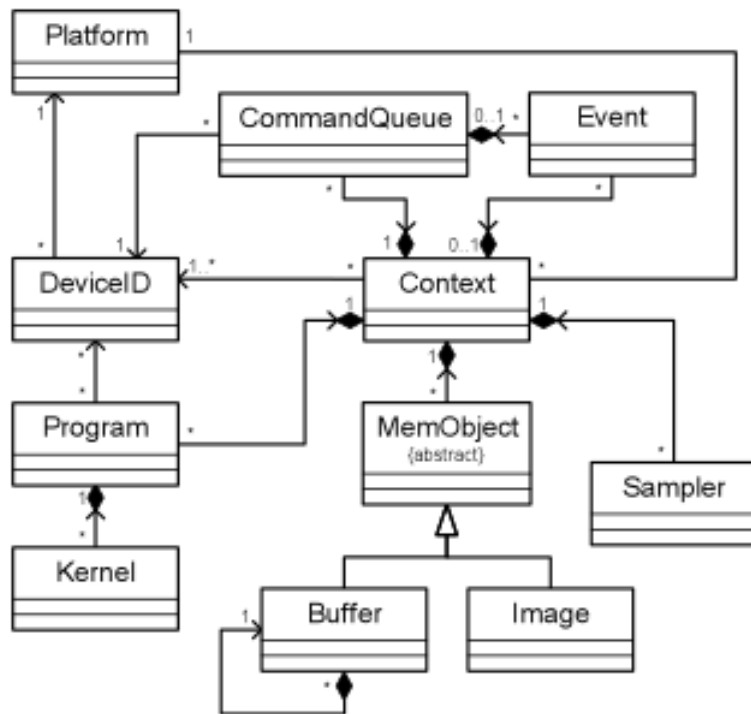
Eksplicitno moramo prenašati podatke: host -> global -> local in potem v istem zaporedju nazaj

Posebnosti

- Na nivoju delovne niti se spomni obnaša konsistentno
- Na nivoju delovne enote se spomin lokalni spomin (local) sinhronizira na točki sinhronizacije (barrier).
- Globalni spomin (global) je konsistenten na nivoju delovne enote ni pa med posameznimi delovnimi enotami
- Po vsakem klicu jedra je vsiljena sinhronizacija



Pomembni pojmi



Platform: zbirka računskih naprav (compute device)

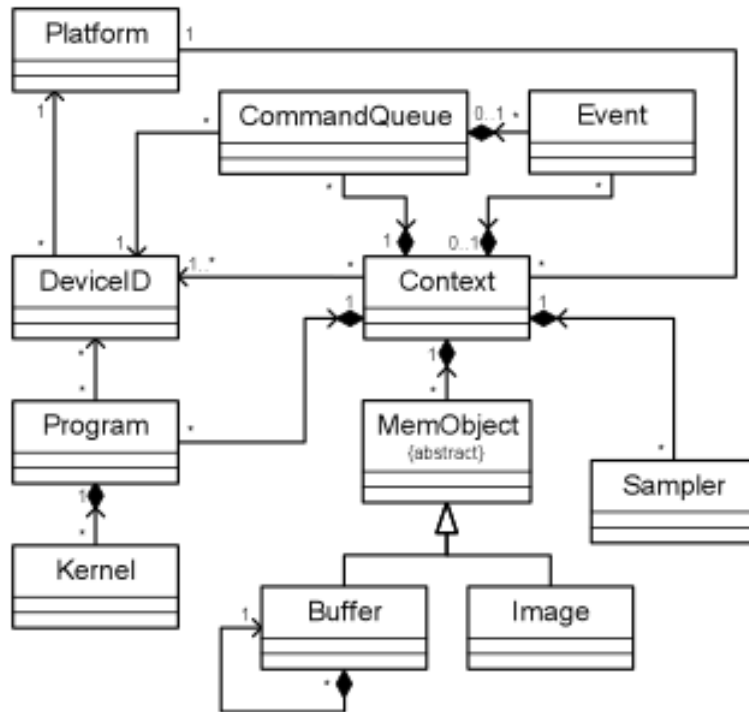
Compute device: CPU ali GPU, vse naprave delujejo z isto verzijo OpenCL

Context: stanje in upravljanje s stanjem ene ali večih naprav (alociranje spomina, prevajanje programov, vodenje ukaznih vrst, izvajanje na indeksnih mrežah)

Kernel: predstavlja prevedeno funkcijo, ki jo lahko konfiguriramo izvedbo

Program: prevedena jedra iz istega vira (niza, datoteke)

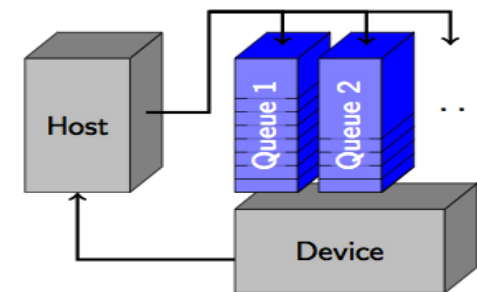
Pomembni pojmi



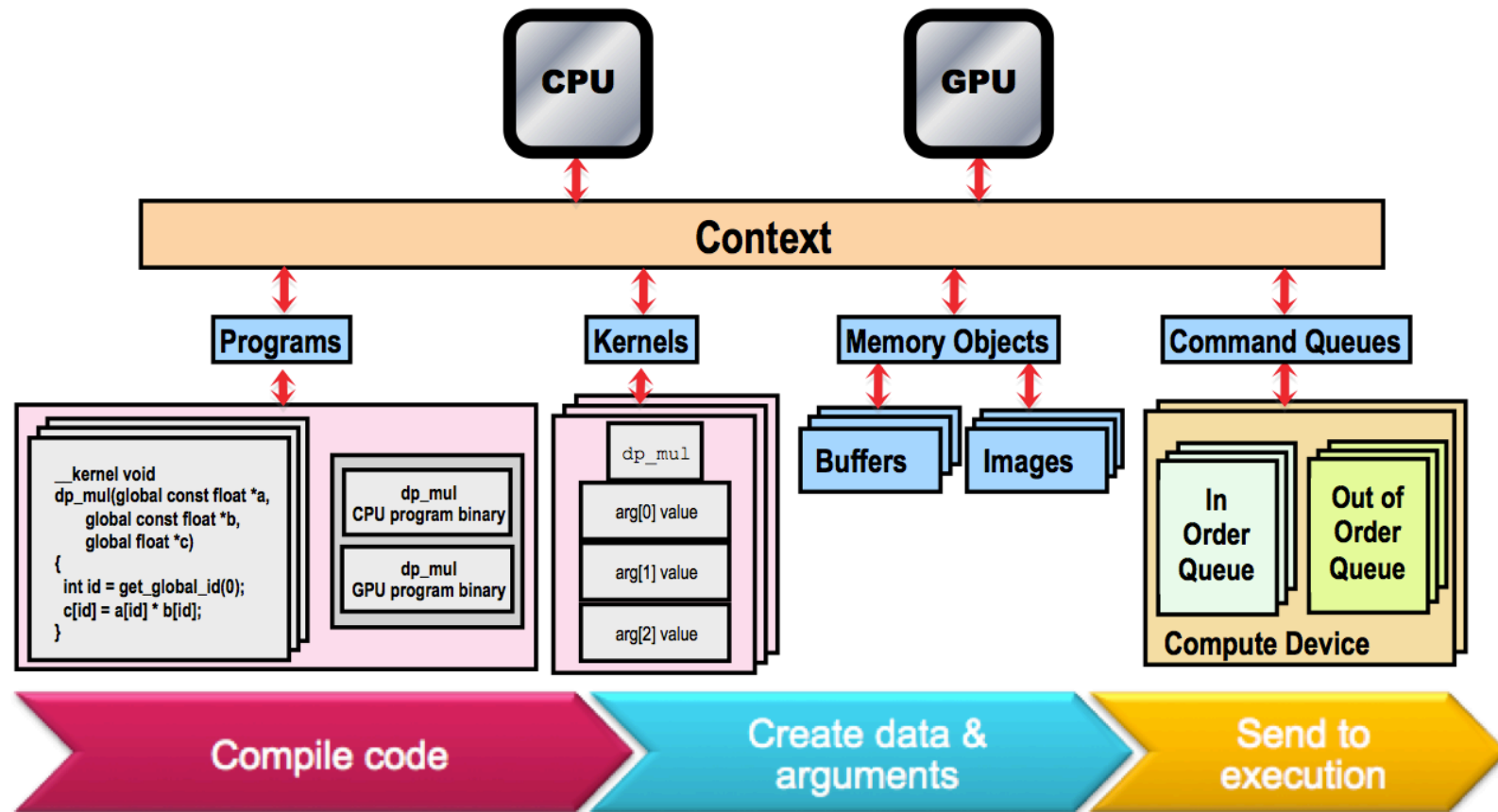
Buffer: reprezentacija dela spomina (tabela, zaporedje bytov), ki ni nujno vezana na konkreten naslov.

Command queue: predstavlja urejeno ali neurejeno vrsto v kontekstu za računsko napravo, ki lahko vsebuje:

- Klice funkcij (kernel)
- Prenose spomina
- Ukaze za sinhronizacijo



Povzetek



© Simon McIntosh-Smith 2011

- Page 35



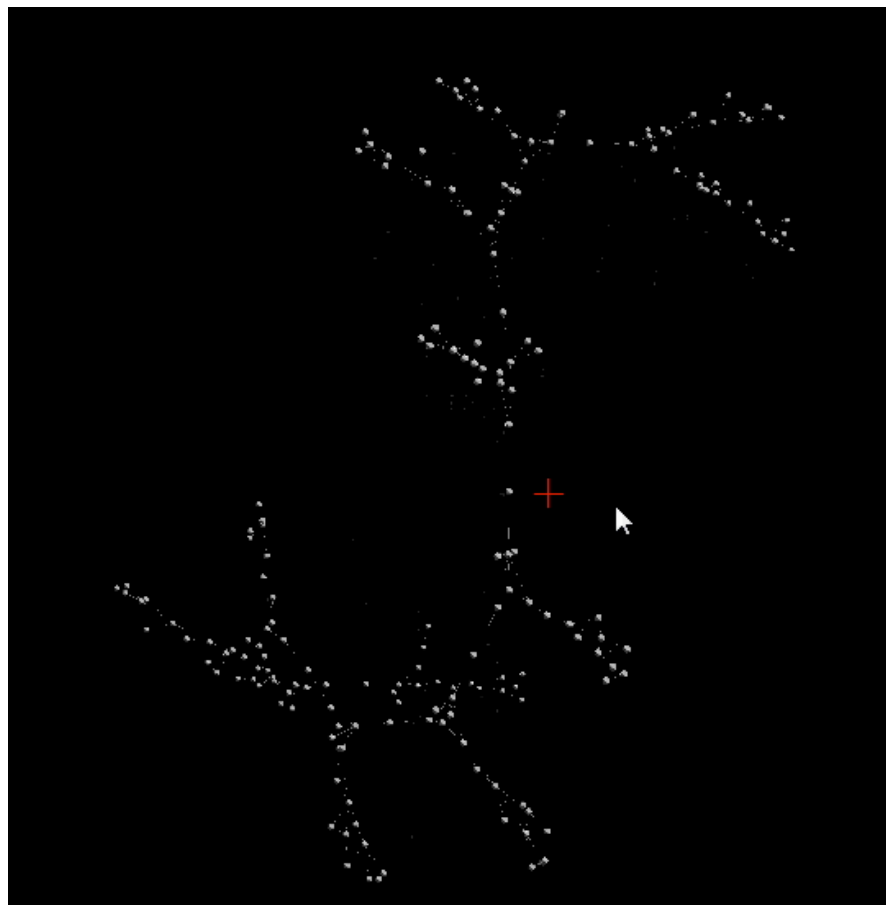
Uporaba OpenCL

- Vsak resnejši programski jezik ima knjižnice za upravljanje infrastrukture OpenCL (C++, C#, Java, Python, ...)
- pyopencl:
 - OpenCL podpora na Pythonu (2 in 3)
 - Integracija z OpenGL
 - Preslikava buffer-jev na numpy tabele
 - Primer: vsota kvadratov

Aplikacije pri Abeliumu

- Net.Plexor
 - orodje za analizo in vizualizacijo velikih omrežij
 - Izrisovalnik grafov (spring embedder)
 - FMM metoda za izračun odbojnih sil
 - Multi-grid metoda za piramidni izračun približkov slike
 - Uporaba posebnega zapisa k-d drevesa

Risalnik velikih grafov

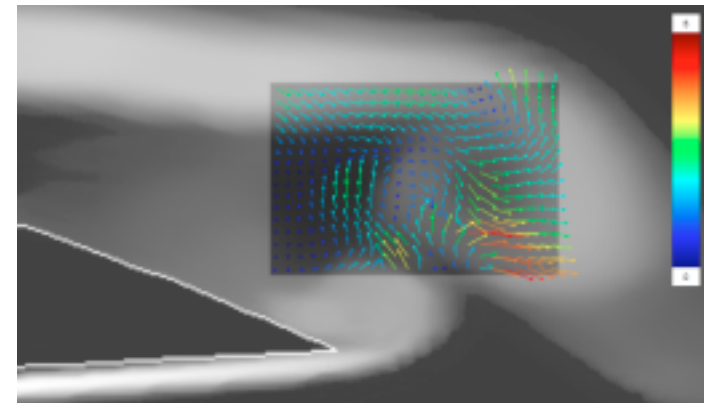
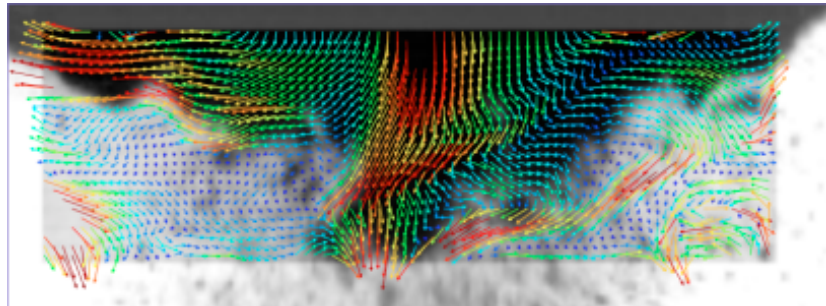


- [Demo1](#)
- [Demo2](#)



Kaj še počnemo?

- ADMflow – vizualizacija fluidnih tokov (skupaj s FS) – poteka optimizacija kode v GPU
- Rudarjene kripto valut (projekt v teku)



Viri

- http://education.ivec.org/training/September2013/gpu_computing_overview.pdf (splošno o GPU)
- http://www.cs.bris.ac.uk/home/simonm/workshops/OpenCL_lecture1.pdf (lecture 1, 2, 3 in 4)
- <http://fisica.cab.cnea.gov.ar/gpgpu/images/charlas/lab-slides.pdf> (pyopencl)

Razno

Operacijo delno financira Evropska unija in sicer iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa krepitev regionalnih razvojnih potencialov za obdobje 2007-2013, 1. razvojne prioritete: Konkurenčnost podjetij in raziskovalna odličnost, prednostne usmeritve 1.1: Izboljšanje konkurenčnih sposobnosti podjetij in raziskovalna odličnost.