

UNIVERSITY OF TURKU  
Department of Information Technology

JUHO SUKONEN: General-purpose Programming on Graphics Processing Units

Master of Science in Technology Thesis, 56 p., 26 app. p.  
Embedded Systems  
December 2010

---

Nowadays graphics processing units (GPUs) can perform general-purpose computations besides only processing graphics. Software developers can write C or C++ programs and choose to run them on the GPU instead of the central processing unit (CPU). In contrast to CPUs sequential execution, GPU executes programs in parallel utilizing hundreds of concurrent threads. Due to the parallel nature of GPUs, they offer great computational power and efficient performance, when dealing with algorithms that can be run in parallel.

Both NVIDIA and ATI have released their own GPU architectures that support general-purpose computing. These architectures apply so called stream programming model, which is generally used with GPU computing. Furthermore, programming languages have been developed for writing programs for GPUs. This thesis introduces both the CUDA architecture of NVIDIA and the ATI Stream Computing technology of ATI. The stream programming model and Brook programming language for GPUs are also introduced.

In the empirical part of this thesis, general-purpose computation on a GPU is tried out in practise. Matrix multiplication and bitonic sorting algorithms are implemented both for GPU and CPU execution. Both implementations are tested with identical input data and execution times are compared in order to find out, how much the performance of GPU differs from CPU.

Keywords: GPU, CUDA, Stream programming, ATI Stream Computing, BrookGPU

TURUN YLIOPISTO  
Informaatioteknologian laitos

JUHO SUKONEN: General-purpose Programming on Graphics Processing Units

Diplomityö, 56 s., 26 liites.

Sulautetut järjestelmät

Joulukuu 2010

---

Nykyään näytönohjaimien grafiikkaprosessoreilla voidaan suorittaa normaalia ohjelmakoodia ja laskentaa grafiikan prosessoinnin lisäksi. Ohjelmoijat voivat kirjoittaa C- tai C++ -ohjelmia ja ajaa niitä grafiikkaprosessorilla tietokoneen varsinaisen prosessorin sijaan. Grafiikkaprosessori kykenee hyödyntämään ohjelmaa suorittaessaan satoja rinnakkaisia säikeitä. Tämän ansiosta grafiikkaprosessorin tarjoama laskentateho ja suorituskyky on parempi verrattuna tietokoneen suorittimeen, kun suoritetaan algoritmeja, joita voidaan ajaa rinnakkain useissa säikeissä.

Sekä NVIDIA että ATI ovat julkaisseet omat grafiikkaprosessoriarkkitehtuurinsa, jotka tukevat ohjelmakoodin suorittamista. Nämä molemmat arkkitehtuurit hyödyntävät niin sanottua rinnakkaisen tietovirtalaskennan mallia. Arkkitehtuurien lisäksi rinnakkaista tietovirtalaskentaa varten on myös kehitetty omia ohjelmointikieliä. Tässä opinnäytetyössä esitellään sekä NVIDIA:n CUDA-arkkitehtuuri että ATI:n ATI Stream Computing -teknologia. Tämän lisäksi esitellään Brook-ohjelmointikieli, jolla voidaan kirjoittaa grafiikkaprosessoreilla suoritettavia ohjelmia.

Tämän opinnäytetyön kokeellisessa osuudessa ohjelmoidaan algoritmit sekä matriisin kertolaskulle että bitoniselle lajittelulle. Molemmat algoritmit ohjelmoidaan sekä grafiikkaprosessorille että suorittimelle. Tämän jälkeen algoritmeja ajetaan identtisillä syötteillä. Algoritmien suoritusajoina vertaillaan, jotta voidaan nähdä, millaisesta suorituskykyerosta grafiikkaprosessoriin ja suorittimen välillä on kyse.

Asiasanat: GPU, CUDA, Rinnakkainen tietovirtaohjelmointi, ATI Stream Computing,  
BrookGPU