Accelerators at Virginia Tech
Invent the Future

Courses

CS@Virginia Tech

- **CS 4234:** Parallel Computation (undergraduate course)
- **CS 4984:** Accelerator-Based Parallel Computing (undergraduate course)
- **CS 5234:** Advanced Parallel Computation (graduate course) (Same website as CS 4234)
- **CS 5984:** Accelerator-Based Parallel Computing (graduate course)
- **CS 6504:** Heterogenous Multicore Computing

External Resources

- **University of California at Davis** EE 171: Parallel Computer Architecture
- **University of Illinois at Urbana-Champaign** ECE 498AL: Programming Massively Parallel Processors
- **University of Pennsylvania** CIS 665 GPU Programming and Architecture
- **University of North Carolina at Chapel Hill** COMP 790-058: GPGPU: General Purpose Computation using Graphics Processors
- **AMD (Third-Party Courses)** AMD Opteron Processor Architecture Course
- **AMD (Third-Party Courses)** AMD64 Software Architecture Course
- **NVIDIA** Introduction to CUDA Technical Training Courses

Lectures

- **Course Overview and Introduction for Parallel Computing**
- **Multicore Architecture and Why Memory Matters**
- **Memory Architecture**
- **Why network matters?**
- **Threads**

- **Threading: Correctness and Performance**
- **GPGPU Computing**
- **CUDA Programming Model**
- **Computation to Core Mapping**
- **GPU Memory (Shared Memory Example)**
- **GPU Memory II (Memory Hardware and Bank Conflict)**
- **Optimization Strategies (Global Memory Access Pattern and Control Flow)**

Programming Assignments

- **Programming Assignment I**
- **Programming Assignment II**

Projects

Introduction

Final project is required for all graduate students in CS5984. Each project team can include two members. We encourage discussion within each project team. Each project team is required to submit a project proposal and a final project report. We also ask each team to give a final project presentation as the end of the semester.

The research topic of the project should be related to GPGPU research, which can be performance analysis, optimization techniques, computation-to-core mapping for a specific algorithm and applications with GPU implementation. If you would like to propose a new research topic for your project, please contact the instructors for approval.

Project Proposal

Project proposal should follow ACM SIG Processings Templates, which is two-column and single-space. You can use either Word or Latex template. Your project proposal should be at least 4 pages long (including references) and include the following sections:

- Introduction / Motivation
- Problem Statement (If multiple problems are addressed, bullet points format is recommened.)
- Related Works
- Approaches
- Anticipated Results and Preliminar Results (if any)
- Project Plan and Potential Risks
- References

Final Project Report

Final project report should also follow ACM SIG Processings Templates, and be at least 6 pages long. The sections for final project are:

- Introduction (Motivation and novel contributions. If multiple contributions are presented, bullet points format is recommended)
- Problem Statement
- Related Works
- Approaches
- Results
- References

The structure of the final report is not restricted to previous sections. Please feel free to use any paper structure in your research area as long as the research contributions are clearly presented.

Note: Please clearly state the persentage of contribution and responsibility from each team member, if you project team includes more than one member.

Resources

Performance

- S. Che, M. Boyer, J. Meng, D. Tarjan, J. W. Sheaffer, and K. Skadron. "A Performance Study of General Purpose Applications on Graphics Processors using CUDA." Journal of Parallel and Distributed Computing 2008.
- S. Ryoo, C. Rodrigues, S. Stone, S. Baghsorkhi, S.-Z. Ueng and W. Hwu, "Program Optimization Study on a 128-Core GPU", Workshop on General Purpose Processing on Graphics Processing Units 2007

Optimization

- Daniel Cederman, Philippas Tsigas, "On Dynamic Load-Balancing on Graphics Processors", Graphics Hardware 2008.
- Shane Ryoo and Christopher I. Rodrigues and Sara S. Baghsorkhi and Sam S. Stone and David B. Kirk and Wen-mei W. Hwu, "Optimization principles and application performance evaluation of a multithreaded GPU using CUDA", 13th ACM SIGPLAN Symposium on Principles and practice of parallel programming, 2008.

Optimization II

- Silberstein, Mark and Schuster, Assaf and Geiger, Dan and Patney, Anjul and Owens, John D, "Efficient Computation of Sum-products on GPUs Through Software-Managed Cache." ICS '08: Proceedings of the 22nd annual international conference on Supercomputing, 2008.
- Phuong Hoai Ha and Tsigas, P. and Anshus, O.J, "Wait-free Programming for General Purpose Computations on Graphics Processors", IEEE International Symposium on Parallel and Distributed Processing, 2008. IPDPS 2008.

Computational Models

- Bingsheng He, Naga K. Govindaraju, Qiong Luo, Burton Smith. Efficient Gather and Scatter Operations on Graphics Processors. SC2007: ACM/IEEE SuperComputing 2007.
- Shubhabrata Sengupta, Mark Harris, Yao Zhang, John D. Owens, "Scan Primitives for GPU Computing", Graphics Hardware 2007.

Map Reduce Model

- Bryan Catanzaro, Narayanan Sundaram and Kurt Keutzer, "A Map Reduce Framework for Programming Graphics Processors", 25th International Symposium on Theoretical Aspects of Computer Science 2008.
- Bingsheng He, Wenbin Fang, Qiong Luo, Naga K. Govindaraju, and Tuyong Wang,"Mars: A MapReduce Framework on Graphics Processors", Parallel Architectures and Compilation Techniques (PACT) 2008.

Application: Database

- Bingsheng He and Ke Yang and Rui Fang and Mian Lu and Naga Govindaraju and Qiong Luo and Pedro Sander, "Relational joins on graphics processors", SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data.
- Michael D. Lieberman, Jagan Sankaranarayanan and Hanan Samet, "A Fast similarity Join Algorithm Using Graphics Processing Units", 24th IEEE International Conference on Data Engineering 2008.

Application: Data Mining

- Wenbin Fang, Ka Keung Lau, Mian Lu, Xiangye Xiao, Chi Kit Lam, Yang, Philip Yang Yang, Bingsheng He, Qiong Luo, Pedro V. Sander and Ke Yang, "Parallel Data Mining on Graphics Processors", Technical Report HKUST-CS08-07, Oct 2008.
- Jeremy. Archuleta, Yong Cao, Wuchun Feng and Tom Scogland, "Multi-Dimensional Characterization of Temporal Data Mining on Graphics Processors", IEEE International Parallel & Distributed Processing Symposium (IPDPS 2009).

Algorithm: Sorting

- N. Satish, M. Harris, and M. Garland, "Designing efficient sorting algorithms for manycore GPUs", Proc. 23rd IEEE Int'l Parallel & Distributed Processing Symposium 2009.
- Erik Sintorn and Ulf Assarsson, "Fast parallel GPU-sorting using a hybrid algorithm", Journal of Parallel and Distributed Computing 2008.

Algorithm: Graph Search

- Pawan Harish, P J Narayanan , "Accelerating large graph algorithms on the GPU using CUDA", Proc of IEEE International Conference on High Performance Computing (HiPC 2007) Goa, December, 2007.

- Gary J. Katz and Joseph T. Kider, Jr, "All-pairs shortest-paths for large graphs on the GPU", GH '08: Proceedings of the 23rd ACM SIGGRAPH/EUROGRAPHICS symposium on Graphics hardware.

Algorithm: Graphics Cut

- Mohamed Hussein, Amitabh Varshney, and Larry Davis, "On Implementing Graph Cuts on CUDA," First Workshop on General Purpose Processing on Graphics Processing Units 2007.
- Vineet, V. and Narayanan, P.J. , "CUDA cuts: Fast graph cuts on the GPU", CVPRW '08: Computer Vision and Pattern Recognition Workshops, 2008.

Algorithm: Hierachical Data Structure

- Zhou, Kun and Hou, Qiming and Wang, Rui and Guo, Baining, "Real-time KD-tree Construction on Graphics Hardware", ACM Transaction on Graphics 2008.
- Christian Lauterbach, Michael Garland, Shubhabrata Sengupta, David Luebke, and Dinesh Manocha1, "Fast BVH Construction on GPUs", EuroGraphics 2009.

Others: Distance Map and Language

- Hou, Qiming and Zhou, Kun and Guo, Baining, "BSGP: bulk-synchronous GPU programming", SIGGRAPH '08: ACM SIGGRAPH 2008.
- Weber, Ofir and Devir, Yohai S. and Bronstein, Alexander M. and Bronstein, Michael M. and Kimmel, Ron, "Parallel algorithms for approximation of distance maps on parametric surfaces", ACM Transaction on Graphics, 2008.

Our research is supported by NSF, RNet Technologies, IBM, AMD, NVIDIA, Eli Lilly & Company, SURA, and Virgnia Tech Foundation