The Ministry of Education and Science of the Russian Federation

Lobachevsky State University of Nizhni Novgorod

Computing Mathematics and Cybernetics faculty

The competitiveness enhancement program
of the Lobachevsky State University of Nizhni Novgorod
among the world's research and education centers

Strategic initiative
“Achieving leading positions in the field of supercomputer technology
and high-performance computing”

Introduction to GPU programming

Training course program

Nizhni Novgorod

2014

INTRODUCTION

This course is an introduction to GPU programming. It considers most popular technologies of GPU programming, GPU architecture, thread execution and memory hierarchy, basic optimization principles. We focus on CUDA and present basic features of CUDA C programming language as well as optimized CUDA libraries CUBLAS, CUFFT and CURAND. Theoretical materials are accompanied by a wide set of examples and practical problems.

The course materials were developed and translated to English in 2014 as part of the "5-100 project" by Sergey Bastrakov.

COURSE OBJECTIVES

The main objective of the course is to introduce basic principles and acquire skills of GPU programming.

It involves solving the following **problems**:

1. Overview of GPU programming technologies.
2. Study of GPU architecture.
3. Getting familiar with CUDA C programming languages. Mastering writing kernels and device functions, workload distribution, and data exchanges between host and device memory.
4. Getting acquainted with examples of implementation and optimization of applications on GPU.
5. Using optimized CUDA libraries: CUBLAS, CUFFT, CURAND.

The course is oriented on engineers, teachers, scientists, and students.

PREREQUISITES

The course is oriented on students with basic skills of C/C++ program development and familiar with basic concepts of parallel programming. Experience with OpenMP, TBB and MPI is not required, but is beneficial. All necessary information of GPU architecture and programming techniques is included in the course materials.

The students are supposed to have basic mathematical knowledge corresponding to 2-3 years of university.

COURSE OUTLINE

The course consists of 32 hours, including 10 lecture hours and 10 practice hours. Practice classes can be held as lab works (students carry out assignments step-by-step under supervision) or master class (supervisor demonstrates and explains step-by-step solutions). 12 hours are allocated for individual work. The authors encourage additional work.

Course outline is as follows:

#	Type of class	Name	Duration	Individual work
1	Lecture	General purpose computing on GPU	2 hours	2 hours
2	Lecture	CUDA C	2 hours	
3	Practice	Vector addition	2 hours	
4	Practice	Numerical integration of heat equation	2 hours	2 hours
5	Lecture	CUDA thread execution and memory hierarchy	2 hours	
6	Lecture	Optimization of CUDA applications	2 hours	2 hours
7	Practice	Matrix multiplication	2 hours	2 hours
8	Lecture	CUDA libraries	2 hours	2 hours
9	Lecture	CUDA Libraries. Minimal residual method. Convolution	2 hours	2 hours
10	Practice	Monte Carlo integration and option pricing	2 hours	
		OVERALL:	**20 hours**	**12 hours**

LEARNING OUTCOMES

Graduates from the course will have knowledge of architecture and programming techniques of GPUs and be able to use it for applied applications.

REFERENCES

1. Sanders J., Kandrot E. CUDA by Example: An Introduction to General-Purpose GPU Programming. – Addison-Wesley Professional, 2010. – 312 p.
2. Farber R. CUDA Application Design and Development. – Morgan Kaufmann, 2011. – 336 p.
3. NVIDIA CUDA C Programming Guide. [http://docs.nvidia.com/cuda/cuda-c-programming-guide/].

FURTHER READING

4. GPU Computing Gems Emerald Edition, ed. Wen-mei W. Hwu. – Morgan Kaufmann, 2011. – 886 p.
5. NVIDIA CUDA C Best Practices Guide [http://docs.nvidia.com/cuda/cuda-c-best-practices-guide#axzz3JRcPurfI]
6. NVIDIA CUBLAS Documentation [http://docs.nvidia.com/cublas/index.html#axzz3JRcPurfI]
7. NVIDIA CUFFT Documentation [http://docs.nvidia.com/cufft/index.html#axzz3JRcPurfI]
8. NVIDIA CURAND Documentation [http://docs.nvidia.com/curand/index.html#axzz3JRcPurfI]

AUTHORS

Sergey Bastrakov

Postgraduate student, assistant of the Software department of CMC faculty. Lines of research: high performance computing, computational geometry. Co-authored 2 textbooks, 15 papers.

email: bastrakov@vmk.unn.ru