



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**

*Test questions*

Nizhni Novgorod

2014

## 01\_LECTURE. GENERAL PURPOSE COMPUTING ON GPU

S,N,1,2.0.0.

What is the difference in peak performance of top CPUs and GPUs in single precision?

-3,Y

Same order of magnitude

9,Y

GPUs are superior by an order of magnitude

-3,N

CPUs are superior by an order of magnitude

-3,N

It is impossible to compare

---

S,N,1,2.0.0.

What is historically first way of GPGPU programming?

-3,N

CUDA

-3,N

OpenCL

-3,N

AMD Stream Computing

9,Y

Shading languages

---

M,N,1,2.0.0.

Why is load balancing important for heterogeneous computing?

5,N

Different performance

-5,N

Different programming languages

5,Y

Different overheads on data transfers

-5,N

None of the above

---

## 02\_LECTURE. CUDA C

S,N,1,2.0.0.

How many cores do modern discrete GPUs have?

-3,N

1-10

-3,N

10-100

9,Y

hundreds or thousands

-3,N

millions

---

S,N,1,2.0.0.

Which keyword is used in CUDA C for function called from CPU and executed on GPU?

-3,N

\_\_host\_\_

-3,N

\_\_device\_\_

9,Y

\_\_global\_\_

-3,N

\_\_kernel\_\_

---

S,N,1,2.0.0.

Which keyword is used in CUDA C for function called from GPU and executed on GPU?

-3,N

\_\_host\_\_

9,Y

\_\_device\_\_

-3,N

\_\_global\_\_

-3,N

\_\_kernel\_\_

---

### 03\_PRACTICE. VECTOR ADDITION

S,N,1,2.0.0.

Which inline variable is used to obtain index of current thread block inside CUDA kernel?

-3,N

threadId

-3,N

threadIdx

-3,N

blockId

9,Y

blockIdx

---

S,N,1,2.0.0.

Which inline variable is used to obtain index of current thread inside thread block inside CUDA kernel?

-3,N

threadId

9,Y

threadIdx

-3,N

blockId

-3,N

blockIdx

---

S,N,1,2.0.0.

Which inline variable is used to obtain index of current thread among all threads inside CUDA kernel?

-3,N

threadId

-3,N

threadIdx

-3,N  
globalThreadId

9,Y  
There is no such variable, but it can be computed using other inline variables  
---

## 04\_PRACTICE. NUMERICAL INTEGRATION OF HEAT EQUATION

S,N,1,2.0.0.

Function declared with \_\_global\_\_ is called and executed by:

-3,N

called from host, executed on host

9,Y

called from host, executed on device

-3,N

called from device, executed on host

-3,N

called from device, executed on device

---

S,N,1,2.0.0.

Function declared with \_\_device\_\_ is called and executed by:

-3,N

called from host, executed on host

-3,N

called from host, executed on device

-3,N

called from device, executed on host

9,Y

called from device, executed on device

---

S,N,1,2.0.0.

Function declared with \_\_host\_\_ is called and executed by:

9,Y

called from host, executed on host

-3,N

called from host, executed on device

-3,N

called from device, executed on host

-3,N



called from device, executed on device

---

## 05\_LECTURE. CUDA THREAD EXECUTION AND MEMORY HIERARCHY

S,N,1,2.0.0.

How are 2D indexes of threads used?

-3,N

threadId[0], threadId[1]

-3,N

threadIdx[0], threadIdx[1]

-3,N

threadIdx, threadIdx

9,Y

threadIdx.x, threadIdx.y

---

S,N,1,2.0.0.

How are 2D indexes of blocks used?

-3,N

blockId[0], blockId[1]

-3,N

blockIdx[0], blockIdx[1]

-3,N

blockIdx, blockIdx

9,Y

blockIdx.x, blockIdx.y

---

S,N,1,2.0.0.

Which data type is used to set 2D indexes in kernel call?

9,Y

dim3

-3,N

int3

-3,N  
dim2

-3,N  
int2

---

## 06\_LECTURE. OPTIMIZATION OF CUDA APPLICATIONS

S,N,1,2.0.0.

Can every CUDA kernel with 2D indexes be equivalently rewritten to use 1D indexes?

5,Y

Yes

-5,N

No

---

S,N,1,2.0.0.

Choose the correct statement:

-3,N

Each multiprocessor always executes one thread block

-3,N

Each multiprocessor always executes several thread blocks

9,Y

Each multiprocessor executes one or several thread blocks

-3,N

None of the above

---

S,N,1,2.0.0.

Choose the correct statement:

9,Y

All threads of a block are always executed on the same multiprocessor

-3,N

All threads of a block are always executed on different multiprocessors

-3,N

Threads of a block can be executed on one or several different multiprocessors

-3,N

None of the above

---

## 07\_PRACTICE. MATRIX MULTIPLICATION

S,N,1,2.0.0.

Shared memory is accessible to:

-3,N

All threads and CPU

-3,N

Threads of a block and CPU

-3,N

All threads, not CPU

9,Y

Threads of a block, not CPU

---

S,N,1,2.0.0.

Global memory is accessible to:

9,Y

All threads and CPU

-3,N

Threads of a block and CPU

-3,N

All threads, not CPU

-3,N

Threads of a block, not CPU

---

M,N,1,2.0.0.

Which of the following problem decomposition schemes are most frequently used on GPU?

5,Y

Decomposition into thousands or more independent subproblems

-5,N

No decomposition, sequential solving

5,Y

Decomposition into tens/hundreds of independent subproblems, each is decomposed into smaller problems

-5,N

Decomposition into tens independent subproblems

---

## 08\_LECTURE. CUDA LIBRARIES

S,N,1,2.0.0.

Asynchronous data transfer between CPU and GPU is

-3,N

Never possible

9,Y

Possible if CUDA streams and asynchronous routines are used

-3,N

Always possible on Tesla GPUs, never possible on GeForce GPUs

-3,N

Always possible on all GPUs

---

M,N,1,2.0.0.

Which of the following conditions might affect coalescing?

3,Y

Locality of thread accesses'

3,Y

Alignment

-3,N

Conditional statements

-3,N

Cache miss rate

---

S,N,1,2.0.0.

Which of the following conditions might affect bank conflicts for shared memory?

-3,N

Locality of thread accesses'

-3,N

Alignment

-3,N

Conditional statements

9,Y

None of the above

---



## 9\_PRACTICE. CUDA LIBRARIES. MINIMAL RESIDUAL METHOD. CONVOLUTION

S,N,1,2.0.0.

BLAS level 1 routines perform

9,Y

Vector-vector operations

-3,N

Matrix-vector operations

-3,N

Matrix-matrix operations

-3,N

All of the above

---

S,N,1,2.0.0.

BLAS level 2 routines perform

-3,N

Vector-vector operations

9,Y

Matrix-vector operations

-3,N

Matrix-matrix operations

-3,N

All of the above

---

S,N,1,2.0.0.

BLAS level 3 routines perform

-3,N

Vector-vector operations

-3,N

Matrix-vector operations

9,Y

Matrix-matrix operations

-3,N

All of the above

---

## 10\_PRACTICE. MONTE CARLO INTEGRATION AND OPTION PRICING

M,N,1,2.0.0.

Which interfaces does CURAND have?

-5,N

MKL

-5,N

rand()

5,Y

Host API

5,Y

Device API

---

M,N,1,2.0.0.

Which kinds of pseudo-random numbers can be generated from kernels?

3,Y

Uniformly distributed

3,Y

Normally distributed

-9,N

Poisson distributed

3,Y

Random bits

---

S,N,1,2.0.0.

What is convergence order of Monte Carlo integration method?

9,Y

1/2

-3,N

1

$$-3,N$$

$$3/2$$

$$-3,N$$

$$2$$

$$---$$