RTU Course "Introduction to High Performance Computing Technology CUDA"

33000 Faculty of Computer Science, Information Technology and Energy

General data

Code	DMI741
Course title	Introduction to High Performance Computing Technology CUDA
Course status in the programme	Compulsory/Courses of Limited Choice
Responsible instructor	Arnis Lektauers
Volume of the course: parts and credits points	1 part, 4.5 credits
Language of instruction	LV
Annotation	This study course covers the theoretical and practical principles of high performance computing that are implemented using graphics processing hardware and specialized software. The study course includes an overview of CUDA parallel computing platform architecture based on graphics processors, parallel computing algorithms, application libraries and tools. An in-depth focus is put on the interdisciplinary application of CUDA, for example, in the areas of big data analysis, interoperability with computer graphics, image processing, computational modelling and machine learning. In addition to the theoretical lectures, in the laboratory classes there are provided the opportunities to gain practical skills in the development of information technology solutions using the CUDA technology.
Structure and tasks of independent studies	The independent work is characterized by the following activities: execution of laboratory tasks and individual research, analysis of obtained results, analytical work with scientific literature and other sources of information for individual research on CUDA high performance computing technology.
Recommended literature	Obligātā. / Obligatory: Cheng, John, Grossman, Max and McKercher, Ty. Professional CUDA C Programming. Wrox, 2014, p. 528. Kirk, David B. and Hwu, Wen-mei W. Programming Massively Parallel Processors: A Hands-on Approach, Morgan Kaufmann; 3rd edition, 2016, p. 576, Sanders, Jason and Kandrot, Edward. CUDA by Example: An Introduction to General Purpose GPU Programming. Addison-Wesley, 2011, p. 312. Wilt, Nicholas. The CUDA Handbook: A Comprehensive Guide to GPU Programming. Addison-Wesley Professional, 2013, p. 528. Papildu. / Additional: Soyata, Tolga. GPU Parallel Program Development Using CUDA. Chapman and Hall/CRC, 2018, p. 440. Storti, Duane and Yurtoglu, Mete. CUDA for Engineers: An Introduction to High-Performance Parallel Computing Addison-Wesley Professional, 2015, p. 352.
Course prerequisites	Basic knowledge of C / C ++ programming language

Course contents

Content	Full- and part-time intramural studies		Part time extramural studies	
	Contact Hours	Indep. work	Contact Hours	Indep. work
Architecture and programming foundations of parallel computing platform CUDA	6	6	0	0
Parallel algorithms in CUDA environment	4	4	0	0
CUDA application libraries and tools	6	6	0	0
CUDA performance optimization methods	2	2	0	0
CUDA interaction with computer graphics	2	2	0	0
CUDA business applications: image processing	2	2	0	0
CUDA business application: numerical modeling	2	2	0	0
CUDA business application: machine learning	4	4	0	0
Implementation of CUDA in a multiprocessor environment	2	2	0	0
CUDA integration in high-level programming languages	2	2	0	0
Lab. task	4	8	0	0
Lab. task “Linear algebra solutions in CUDA environment”	4	8	0	0
Lab. task "Practical application of CUDA based software development technologies"	4	8	0	0
Lab. task "Implementation of image processing and analysis tasks in CUDA environment"	4	16	0	0
Total:	48	72	0	0

Learning outcomes and assessment

Learning outcomes	Assessment methods
Able to define, interpret and use professional terminology in the field of graphics processor-based high-performance computing.	Test successfully passed.

Able to develop a software solution based on CUDA technology.	Is able to explain the nature, possibilities, limitations and importance of the use of high-performance computing technologies in certain fields of science and practice.
Able to evaluate the basic ways of developing the proposed high-performance software solution, as well as the limitations of use and optimization options.	During the laboratory work and individual research, the student is able to identify possible solutions and limitations of the given task and offer alternative solutions.
Able to explain the nature, possibilities, limitations and importance of the use of high-performance computing technologies in certain fields of science and practice.	During the exam, the ability to recognize the essence of the formulated thematic questions, as well as to give a concise and correct explanation of the given topics is demonstrated.

Evaluation criteria of study results

Criterion	%
Laboratory tasks	25
Research task	25
Test	20
Exam	30
Total:	100

Study subject structure

Part	CP	Hours			Tests		
		Lectures	Practical	Lab.	Test	Exam	Work
1.	4.5	2.0	0.0	1.0		*	