

Course code								
Type and description	EC							
ECTS credit	1							
Course name	General-purpose computing on graphics processing units							
Course name in Polish	Obliczenia na procesorach graficznych							
Language of instruction	English							
Course level	8 PRK							
Course coordinator	Piotr Napieralski							
Course instructors	Dariusz Puchała							
Delivery methods and course duration		Lecture	Tutorials	Laboratory	Project	Seminar	Other	Total of teaching hours during semester
	Contact hours				15			15
	E-learning	No	No	No	No	No	No	
	Assessment criteria (weightage)				100%			
Course objective	Understanding and ability to use general-purpose computing on graphics processing units							
Learning outcomes	Knowledge and ability to use general-purpose computing on graphics processing units W4, U4, K1							
Assessment methods	Presentation of the progress of work during the realization of the entire semester project and the assessment of partial and summary reports.							
Prerequisites	Knowledge of C++ programming language, knowledge of data structures.							
Course content with delivery methods	The aim of the course is to present and explain the most important issues related to the physical organization and programming of modern graphics processing units (GPUs) with application of CUDA and OpenCL libraries. The presented knowledge is aimed at gaining practical skills of programming highly efficient general purpose algorithms using available and popular programming tools. It should be emphasized that the contents of the subject							

	<p>fully correspond to the contemporary trend of removing time consuming computational tasks from the CPU and transferring them to graphics processing units. During the course, students propose and implement their own solution to the problem that is strictly related to their research topics.</p> <p>In particular, the presented theoretical content is related to:</p> <ol style="list-style-type: none"> 1. introduction to architecture and organization of graphics processing units by taking into account the hierarchy of resource availability from the level of threads, blocks and grids of blocks, 2. discussion of the mechanism of hiding memory latency, and the practical consequences of branch divergence, 3. basics of programming massively-parallel applications for graphics processing units with use of CUDA and OpenCL libraries, 4. introduction to two-dimensional grids of threads on the example of image processing algorithms, 5. the concepts of local and global synchronization with multi-stage algorithms (fast linear transformation algorithms, calculating increasing sums, neural networks, etc.) as an example of a class of algorithms requiring inter-stage synchronization, 6. utilization of shared memory, 7. constant memory and atomic operations, 8. measurement of the execution time of algorithms dedicated for graphics processing units.
Basic reference materials	<ol style="list-style-type: none"> 1. Course materials in the form of pdf files. 2. CUDA library documentation (available online). 3. J. Sanders, E. Kandrot, <i>CUDA by Example: An Introduction to General-Purpose GPU Programming</i>, Addison Wesley Pub Co Inc, 2010. 4. J. Han, B. Sharma, <i>Learn CUDA Programming: A beginner's guide to GPU programming and parallel computing with CUDA 10.x and C/C++</i>, Packt Publishing Limited, 2019.
Other reference materials	<p>W. W. Hwu, D. B. Kirk, I. E. Hajj, <i>Programming Massively Parallel Processors</i>, Elsevier Science & Technology, 2022.</p>
Average student workload outside classroom	10 h
Comments	
Last update	26.04.2023