COURSE DESCRIPTION
HIGH PERFORMANCE AND QUANTUM COMPUTING

SSD: SISTEMI DI ELABORAZIONE DELLE INFORMAZIONI (ING-INF/05)

DEGREE PROGRAMME: INGEGNERIA INFORMATICA (M63)
ACADEMIC YEAR 2022/2023

COURSE DESCRIPTION

TEACHER: CILARDO ALESSANDRO
PHONE: 081-7683852
EMAIL: alessandro.cilardo@unina.it

GENERAL INFORMATION ABOUT THE COURSE

INTEGRATED COURSE: NOT APPLICABLE
MODULE: NOT APPLICABLE
CHANNEL: FG A-Z
YEAR OF THE DEGREE PROGRAMME: I
PERIOD IN WHICH THE COURSE IS DELIVERED: SEMESTER I
CFU: 6

REQUIRED PRELIMINARY COURSES

None.

PREREQUISITES

Medium-level knowledge of computer architecture and C programming.

LEARNING GOALS

The course aims at providing students with advanced knowledge covering modern computer architectures employed for high-performance computing, going into detail of the internal organization of superscalar processors, then moving to multi- and many-core computers. In that respect, the course also covers heterogeneous computer architectures, particularly GPUs, and related programming models. The presentation of advanced computer architectures is then extended to emerging Quantum Computing technologies, introducing students to topics related to system design and management as well as programming aspects.

EXPECTED LEARNING OUTCOMES (DUBLIN DESCRIPTORS)

Knowledge and understanding

The course approach, based on theoretical and practical lessons, aims to provide students with the knowledge related to the architecture of today's computers used for high performance computing (HPC), covering the different forms of parallelism exposed to applications. Therefore, the course provides the understanding of the internal organization of superscalar processors, then extending the discussion to multi- and many-core computers. The first part of the program is also aimed at the presentation of heterogeneous computing architectures, in particular based on GPUs, an established paradigm for the development of high-performance parallel applications. The second part of the course, on the other hand, is aimed at investigating the opportunities offered by emerging Quantum Computing (QC) technologies in the perspective of high-performance computing, addressing topics related to project and system management as well as programming aspects.

Applying knowledge and understanding

The course approach aims at building the skills and methodological/operational tools needed to apply the acquired knowledge in terms of development of parallel applications, programming of advanced and emerging computer architectures, and performance evaluation. Specifically, the practical part of the course is devoted to the presentation of programming models for GPUs, particularly CUDA and OpenCL. Furthermore, it provides insights into emerging Quantum Computing technologies, focusing on experimental QC platforms currently made available by international companies. In that respect, the course program includes the presentation of real case-studies, partly developed in an interactive form with students, addressing both conventional parallel architectures and emerging QC scenarios.

COURSE CONTENT/SYLLABUS

The course is organized in two groups of contents, a theoretical part and an applicative part.

Theoretical part

- Superscalar architectures of modern processors, out-of-order execution, hardware multi-threading, etc.
- Memory organization in parallel systems: coherence and consistency issues
- System-level interconnections and on-chip networks
- Vector Extensions: general concepts and case-studies related to real processors
- Graphics Processing Unit: architectural aspects and real case-studies
- GPU programming models: insights into advanced programming concepts
- Quantum Computing: technological aspects and computational models

Applicative part

For the applicative part, students will individually develop an assignment that will either include the in-depth study of a selected technological aspect on advanced or emerging computing architectures (QC), or alternatively the optimized development and presentation of a program on a GPU architecture or a vector extension.

READINGS/BIBLIOGRAPHY

- J. Hennessy, D. A. Patterson, Computer Architecture: A Quantitative Approach, 6th Edition, Morgan Kaufmann, 2019
- R.S. Sutor, Dancing with Qubits: How quantum computing works and how it can change the world, Packt Publishing, 2019
- NVIDIA, CUDA C Programming Guide, v. 11.1.0, online, NVIDIA 2020
- V. Silva, Practical Quantum Computing for Developers, Apress, 2018
- D. Kaeli, P. Mistry, Heterogeneous Computing with OpenCL 2.0, Morgan Kaufmann, 3rd edition, 2015
- D. J. Sorin, M. D. Hill, and D. A. Wood, A Primer on Memory Consistency and Cache Coherence, Morgan Claypool 2011
- N. E. Jerger, T. Krishna, and L.-S. Peh, On-Chip Networks, Morgan Claypool, 2nd edition, 2017
- Manuals and handouts provided during the course.

TEACHING METHODS OF THE COURSE (OR MODULE)

The teacher will rely on: a) lectures for approximately 65% of the total hours, b) exercises to practically apply the theoretical knowledge for 8 hours, c) laboratories to further elaborate on the applied knowledge for 6 hours, d) seminars to elaborate on specific topics for 2 hours.

EXAMINATION/EVALUATION CRITERIA

a) Exam type

- ☐ Written
- ☑ Oral
- ☑ Project discussion
- ☐ Other

In case of a written exam, questions refer to

- ☐ Multiple choice answers
- ☐ Open answers
- ☐ Numerical exercises

b) Evaluation pattern

Combined evaluation of oral exam and project discussion.