

Arquitecturas de Alto Desempenho

General Description

António Rui Borges

Bolonha Model

- It promotes a student-centred teaching by
 - stimulating an autonomous learning
 - proposing the *problem solving* paradigm as the main methodological strategy for teaching
 - stressing the development of specific skills vs. a more or less automatic building up of general knowledge.
- It establishes very precise metrics on the work being carried out
 - the academic week is defined to be 40 hours of effective work, meaning a total of 30 ECTS credits
 - 1 ECTS = 4/3 h of weekly study
 - each course of the curriculum is assigned a very definite work load
 - AAD: 6 ECTS \Rightarrow 8 h of weekly study (attending classes + home work).

Main Objectives

- to introduce the most relevant design concepts present in recent generations of processors and how they affect the performance of a computer system
- to describe the organization of the memory hierarchy, in particular cache and virtual memory
- to get acquainted with computer architecture simulation tools and how they can be used to assess performance.

Learning Outcomes

- to understand the architecture and the most important decisions taken on the design of modern processors and how they affect program execution
- to be able to plan and carry out a set of simulations for testing different processor configurations.

Prerequisites

- good operating knowledge on digital circuit design
- basic notions on computer architecture and on communication protocols with input-output devices (pooled I/O, interrupt driven I/O and DMA based I/O)
- programming skills in C Language and VHDL at a fair to good level.

Syllabus

- 1. Fundamentals of quantitative design and analysis of processors.
- 2. Pipelined processors organization and limitations to their operation.
- 3. Memory hierarchy: cache memory and virtual memory.
- 4. Exploitation of parallelism at the instruction level: advanced techniques for branch prediction, static and dynamic scheduling, speculative processing, multi-instruction issue.
- 5. Exploitation of parallelism at the task level: multithreading and multiprocessors.
- 6. Exploitation of parallelism at the data level: vector and graphic (GPUs) processors.

Main bibliography

- J. Hennessy, D. A. Patterson, "Computer Architecture a Quantitative Approach", 6th Edition, Morgan Kaufmann, 2017
- W. Stallings, "Computer Organization and Architecture Designing for Performance", 10th Edition, Pearson Education, 2016
- D. B. Kirk, W. W. Hwu, "Programming Massively Parallel Processors: A Hands-on Approach", Morgan Kaufmann, 2017

Important note – The sections of these books referred in *suggested reading* of each chapter should be really read!

Lectures

Lectures present specific topics of the syllabus. The adopted approach tries to entice the students to participate actively in the discussion and to help them to develop skills of critical reasoning and to learn general techniques of problem solving.

Lab classes - 1

Labs follow the motto "you learn by doing" and aim the completion of small tasks to prepare the students for the work assignments.

Work assignment 1 – Digital circuitry simulation

Implementing and evaluating specific features in a pipelined architecture.

Work assignment 2 – CUDA / OpenCl programming

Solving a computer intensive task in a GPU architecture.

Students are organized in working groups composed of two elements. Each group must present and defend their own solution to the proposed problems.

Lab classes - 2

Labs take place in site at department room 101.

In the week of September 19 to 25, labs classes P1 and P2 will be moved to the afternoon of September 21 (16h - 18h) and 18h - 20h, respectively) due to the welcome of new students.

In the week of October 31 to November 6, labs classes P3 and P4 will be moved to the afternoon of November 2 (16h - 18h and 18h - 20h, respectively) due to the holiday of November 1.

Tutorials

Tutorials take place every week on Mondays, at 18h, in Anf. V.

Some of them will have an expositive character and aim to help the students to overcome deficiencies in background knowledge as well as to provide a space for the discussion of specific aspects of the course.

Themes to be treated include

- Digital circuitry modeling
- CUDA / Open Cl programming.

Grading - 1

course grade =
$$\frac{5 \text{ x theoretical mark } + 5 \text{ x lab mark}}{10}$$

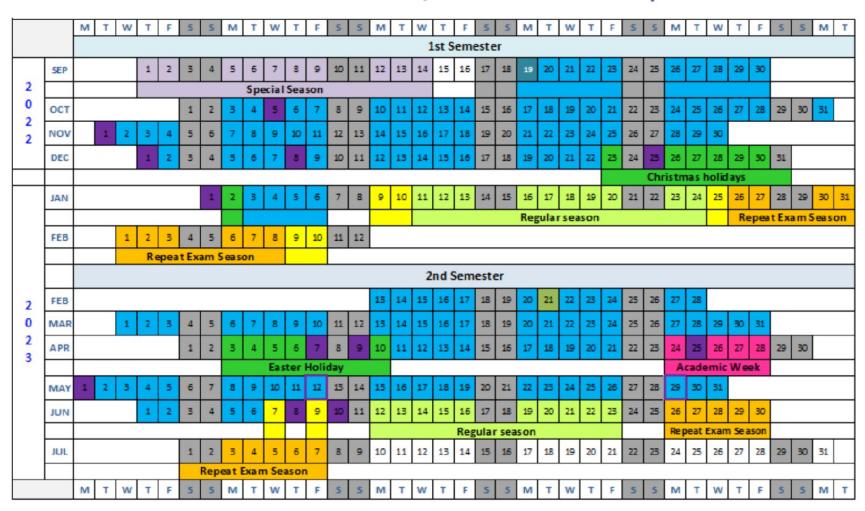
- rounding is always carried out *half up* to unities, except when the lab mark is higher than the theoretical mark by more than three units; in this case, rounding is carried out *half down*
- theoretical grading
 - written examination (época normal ou época de recurso)
 - challenge placed during lectures (optional)
- lab grading
 - composed of work assignment 1 and work assignment 2, each having equal weight
 - the mark is limited to 17 units: a higher grade requires an additional assignment.

Grading - 2

- Pass
 - course grade higher or equal to 10 units
- Fail
 - final grade lower than 10 units

Calendar

Academic Calendar for the 2022/2023 Academic Year: update



Final remarks

- special dates
 - deadline for delivering work assignment 1: November 13, 2022
 - deadline for delivering work assignment 2: January 2, 2023
- all documentation about the course can be found in the *e-learning* site (moodle)
- any further questions may be answered by the course operational document or by myself.