Computação em Larga Escala

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
    - CLE: 6 ECTS $\Rightarrow$ 8 h of weekly study (attending classes + home work).

Main Objectives

- to acquaint the students to the principles of high performance computing
- to introduce the most important paradigms of algorithmic design, communication and synchronization in parallel programming.

Learning Outcomes

- to gain a good understanding on the main issues related to programming at different levels of parallelism (coarse, medium and fine)
- to develop skills for the design and the implementation of simple parallel applications on multiprocessing architectures
- to acquaint the students with the functionality of C-based parallel programming environments using the pthread library, MPI and CUDA.

Prerequisites

- basic knowledge of computer architecture
- basic knowledge about operating systems and multiprogrammed environments
- fair to advanced working knowledge of C language and some knowledge of the principles of concurrent programming.

Syllabus

High performance computing

- Architectural basics of a parallel machine
- Decomposition techniques
- Law of Amdahl
- Tools used in programming parallel applications

Medium-grain parallelism

- Revision of basic notions about computer architecture
- General principles of concurrency
- Synchronization devices
- Decomposition model
- Library pthead

Coarse-grain parallelism

- Concept of message exchange
- Decomposition model
- Synchronization devices
- Computational model provided by MPI

Fine-grain parallelism

- Heterogeneous computing using GPUs (Graphical Processing Units)
- Architecture of a GPU
- Language CUDA C
- Computational model
- Decomposition techniques.

Main bibliography

- *Introduction to HPC with MPI for Data Science*, F. Nielsen, Springer, 2016
- *Parallel Programming in C with MPI and OPenMP*, M. Quinn, McGraw-Hill, 2003
- *Professional Cuda C Programming*, J. Cheng, M.Grossman, T. McKercher, John Wiley & Sons, 2014
- *Programming Massively Parallel Processors*, D. Kirk, W. Hwu, Morgan Kaufmann, 2017

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

Labs follow the motto "*you learn by doing*" and are mostly devoted to discuss implementation issues about the solution of a set of problems.

***Work assignment* 1**

Purely concurrent (multithreaded) solution of the problems.

***Work assignment* 2**

Parallel solution of the problems based on message passing using MPI.

***Work assignment* 3**

Parallel solution of the problems based on shared variables using CUDA.

Students are organized in working groups composed of two elements. Each group must present and defend its approach to the solution and its implementation during a query session.

Tutorials

Tutorials take place every week on Mondays, at 14h.

Tutorials have for the most part an exposition character and aim to help the students to overcome deficiencies on the background knowledge some of them may have, as well as to provide a space for the discussion of specific aspects of the course.

Grading - 1

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

- 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)
- *lab grading*
  - composed of *work assignments* 1 through 3, each having equal weight

Grading - 2

- *Pass*
  - both theoretical and lab marks higher or equal to 8,5 units ***and*** course grade higher or equal to 10 units
- *Fail*
  - theoretical mark lower than minimum mark ***or*** lab mark lower than minimum mark ***or*** final grade lower than 10 units
- *Fail by minimum mark*
  - lab mark lower than minimum mark
- *Fail by absence (regular student)*
  - missing more than three lab classes

Final remarks

- special dates
  - deadline for delivering work assignment 1: 31 de Março de 2024
  - deadline for delivering work assignment 2: 4 de Maio de 2024
  - deadline for delivering work assignment 3: 5 de Junho de 2024
- all documentation about the course can be found in the *elearning* site (moodle)
- any further questions may be answered by the course operational document or by myself.

elearning site