CS 610: Programming for Performance

Swarnendu Biswas

Semester 2020-2021-I

CSE, IIT Kanpur

Copyright Information

- "The instructor of this course owns the copyright of all the course materials. This lecture material was distributed only to the students attending the course CS 610: Programming for Performance of IIT Kanpur, and should not be distributed in print or through electronic media without the consent of the instructor. Students can make their own copies of the course materials for their use."

Course Details

- CS 610: Programming for Performance

- Semester 2020-2021-I

- Class hours: WF 10:35-11:50 AM (online, asynchronous)

- Discussion hours: W 10:30-11:30 AM (online, synchronous)

- Webpage: https://www.cse.iitk.ac.in/~swarnendu/courses/cs610/

- LMS: **REGISTER** for CS 610 on mooKIT

Swarnendu Biswas

Instructor Details

- Name: Swarnendu Biswas

- Office: KD 302

- Webpage: https://www.cse.iitk.ac.in/~swarnendu

- Email: swarnendu@cse.iitk.ac.in

Swarnendu Biswas

TA Details

Name	Email
Fahad Mohmedisuf Shaikh	fahad
Samvid Mistry	samvid
Sharwari Samdekar	sharwari
Prafulla Saxena	prafulla
Vipin Patel	vipinpat

Swarnendu Biswas

Course Outline

- Introduction: Challenges in parallel programming, correctness and performance errors, understanding performance, performance models
- Exploiting spatial and temporal locality with caches, analytical cache miss analysis
- Compiler transformations: Dependence analysis, Loop Transformations
- Shared-memory programming and Pthreads
- Compiler vectorization: vector ISA, auto-vectorizing compiler, vector intrinsics, assembly

Swarnendu Biswas

Course Outline

- OpenMP: Core OpenMP and Advanced OpenMP
- Parallel Programming Models and Patterns
- Intel Threading Building Blocks
- GPGPU programming: GPU architecture and CUDA Programming
- Performance bottleneck analysis: PAPI counters, Using performance analysis tools

Swarnendu Biswas

Course Outline

- Optional topics
  - Heterogeneous Programming with OpenMP
  - Java Fork-Join Parallelism
  - Concurrent data structures
  - Shared-memory synchronization
  - Memory consistency models
  - Transactional memory

Swarnendu Biswas

Prerequisites

- Good background in OS and Computer Architecture will help
  - CS 210 (Computer Organization)
  - CS 330 (Operating Systems)
  - CS 335 (Compiler Design)
  - CS 422 (Computer Architecture)
  - Equivalent for non-IITK courses

- Programming maturity is desirable
  - We will have a few programming assignments
  - Primarily C/C++/Java, with OpenMP, TBB, and CUDA extensions

Swarnendu Biswas

Guest Lectures

- We will have around 4-6 lectures from Intel researcher, Dr. Nitya Hariharan

- Planned topics that Nitya will cover
  - OpenMP
  - Performance Analysis

Course Policies

- Online Discussion
  - Be **ON TIME**!
  - Try to FOCUS and PARTICIPATE!
  - Avoid DISTRACTIONS!
  - Keep your mobile phones SILENT and AWAY from you

- Email subjects **SHOULD** start with [CS610]

- Submitting your assignments late will mean losing points automatically. You will lose 25% for each day that you miss, for up to two days.

Swarnendu Biswas

Evaluation

Class participation/quizzes/paper critiques	10%
Assignments	40%
Mid-sem	20%
End-sem	30%

Swarnendu Biswas

Evaluation ✂

Class participation/quizzes/paper	10%

- This is a tentative allocation
  - Might change allocations slightly depending on the strength of the class
- Grading will be relative

Swarnendu Biswas

Academic Integrity

- You **MAY** discuss concepts with classmates

- All assignments **MUST** be your own or your team's work when teamwork is permitted

- You **MAY NOT** search online for existing solutions related to the assignments, even as a reference

- Students caught **CHEATING/PLAGIARIZING** will be punished

Swarnendu Biswas

Teaching Philosophy

- I am happy to clarify **EVERY REASONABLE** question

- Slides will primarily be pointers to concepts and materials
    - I will post optional reading material

- I encourage you to make use of the discussion hours and the LMS

Swarnendu Biswas

Resources

- Optimizing Compilers for Modern Architectures – R. Allen and K. Kennedy

- An Introduction to Parallel Programming – P. Pacheco

- Programming Massively Parallel Processors: A Hands-on Approach – David Kirk Wen-mei W. Hwu

- Intel Threading Building Blocks – James Reinders

- Other handouts and research papers

Swarnendu Biswas

What this course is and is not?

- This is not a programming "tips and tricks" course
  - We will discuss more generic abstract concepts

- We will focus on a wide variety of topics
  - Focus will be on shared memory systems

- This is not an introductory course to any specific tool
  - However, we will probably make use of a few

Swarnendu Biswas

Collaborative Learning

- Make use of the discussion hours

> Discussion hours: W 10:30-11:30 AM (online)

- You are welcome to **PROVIDE** feedback anytime

Questions?