SC3050
Advanced Computer Architecture

(Course Information)

College of Computing and Data Science
Nanyang Technological University

Instructors

Smitha K G
Lecture Module 1 - 4 (Week 1 - 7)

Email: Smitha@ntu.edu.sg
Office: N4-2c-75a
https://personal.ntu.edu.sg/smitha/

Weichen Liu
Lecture Module 5 - 7 (Week 8 - 13)

Email: liu@ntu.edu.sg
Office: N4-02b-61
https://personal.ntu.edu.sg/liu/

(course coordinator)

- **Please refer to NTULearn/Information page for info on tutors and lab TAs.**
- Please direct your questions related to lectures, tutorials and labs to your correct instructor, tutor and TA to get timely and effective feedback.

Learning Objectives

- Interpret the performance of a processor based on metrics such as execution speed and power.
- Design processors that achieve the desired performance in a step-by-step manner.
- Design and describe pipeline data-path for performance enhancement.
- Predict the challenges of realizing different kinds of parallelism (such as instruction, data and thread levels) and leverage them for performance advancement for the present, and in the future.

- Design cache that takes into account the importance of efficient memory design to overcome memory wall.
- Develop high performance programs by taking into consideration data-path, memory design and parallelism at instruction, data and thread levels.
- GPU architecture.

Course Outline

- **Module-1: Introduction and Background:** Review of basic computer architecture, Technology trend and design goals, Performance metrics and performance enhancement techniques, Power dissipation in processors, power metrics, and low-power design techniques.
- **Module-2: Instruction Set Architecture Design:** Instruction set design, implementation and performance perspectives, relative advantages of RISC and CISC instruction sets.
- **Module-3: Micro-architecture Design:** Single-cycle data path design, Pipeline data-path design.
- **Module-4: Instruction-Level Parallelism:** Concept and examples of data-dependence, Challenges in ILP realization, Pipeline hazards and their solutions, Data forwarding, Register renaming, Reordering of instructions, Out-of-order execution, Branch prediction, dynamic scheduling, Limitations of scalar pipelines, VLIW and superscalar processors, Instruction, data and memory-flow challenges in superscalar and out-of-order processors.

Course Outline

- **Module-5: Memory Systems:** Memory hierarchy, Cache design considerations, instruction vs. data caches, write-policy and replacement policy, analysis of cache performance, and cache design for performance enhancement.
- **Module-6: GPU Architecture and Internal:** Motivation for GPU, Evolution of GPU architecture, GPGPU and GPU Programming with CUDA; Streaming Multiprocessor, Streaming Processor, Warp, and Warp scheduler operation, Resource allocation.
- **Module-7: Data-Level and Thread-Level Parallelism:** Introduction to vector architecture, SIMD instruction set extensions; Motivation for multicore and many-core systems, Amdahl's law under power constraint, Challenges in efficient multi-core system design, Cache coherence problem.

- **Module-8 (e-learning module, non-examinable): Other Emerging Computing Trends:** Application specific architectures: ASIP. FPGA and ASIC, Heterogeneous multicore platform, Introduction to domain-specific computing, Comparison of performance and power consumption of general-purpose processors, DSP, GPU, FPGA and ASIC.

Course Highlights

- More depth over basic architecture topics
  - Performance measurement and performance limiting factors
  - Power consumption
  - Instruction set architecture design
  - Data-path design
  - Cache
- Different types of parallelism
  - Instruction level parallelism
  - Data level parallelism
  - Thread/Task level parallelism
- GPU architecture and CUDA

Lectures and Tutorials

Lecture Information

- Number of Lectures: 22 over Week 1-6 and Week 8-12
- Lectures are to be delivered **in person** except Week 3 (home-based learning)
- Lecture Commencement: **Week 1**
- E-learning lecture: Week 3 (online lecture, home-based learning, examinable)
  Week 7 (self study without online lecture, non-examinable)

Tutorial Information

- Number of Tutorials: 8 over Week 3-6 and Week 8-11
- Tutorials are to be delivered **in person** except Week 3 (home-based learning)
- Tutorial Commencement: **Week 3**
- According to the college policy to encourage participation
  - Tutorial solutions will NOT be provided
  - We will take attendance for all tutorial sessions. Up to 5 bonus marks will be awarded for high participation rates
  - No make-up tutorial will be granted as we do not require 100% participation.
- E-learning tutorial: Week 3 (online tutorial, home-based learning, examinable)

Note: there may be ad-hoc arrangement of schedules for individual tutorial or lab groups due to conflicts with public holidays or university events. Please follow your tutor or TA's instructions.

Labs

Lab Information

- Number of Labs: 5 over Week 3-12 (bi-weekly)
- All labs are to be delivered **in person**
- Lab Commencement: **Week 3**
- Location: HWLAB1 and HWLAB3
- Lab attendance is mandatory. Late arrivals after 20 minutes will be labelled as "Late" and may lead to penalties

- Labs 1-4 are based on Verilog
- Software: **AMD/Xilinx Vivado**, etc.
- Code will be provided. You need to understand the code to understand the architecture design.

- Lab 5 is based on **CUDA and Nvidia Jetson TX2** embedded GPU platform
- https://developer.nvidia.com/embedded/jetson-tx2

Note: there may be ad-hoc arrangement of schedules for individual tutorial or lab groups due to conflicts with public holidays or university events. Please follow your tutor or TA's instructions.

Assessment

Assessment Information

- **Everything taught will be examinable excluding e-learning lecture for Week 7**

- **Lab participation and report (10%)**
  - Attendance taking and report submission will be done for **all the 5 labs**
  - Late arrivals after 20 minutes will be labelled as "Late" and lead to penalties
  - Late submission of lab report after lab session will NOT be accepted
  - Lab attendance taking and report submission (**10%**)

- **Lab quizzes (40%)**
  - Number of quizzes: **4 (each taking 10%)**
  - Quizzes are to be held at the end of Lab 2, Lab 3, Lab 4 and Lab 5. No quiz for Lab 1
  - Each quiz takes 15 minutes with 3-5 questions based on learned materials
  - Use Respondus Lockdown Browser in lab computer. No remote access is allowed

- **Close-book final exam (50%)**

Lab Quiz: Rules and Guidelines

1. You must sign attendance for every lab session.
2. The quizzes must be taken using the lockdown browser on lab computers during the lab sessions. You can only access the quiz page on the lab computer but not anything else during the quiz.
3. You are NOT allowed to use any other electronic device, like smartphones or tablets. Your devices must be kept inside your bag while taking a quiz. You can bring reference materials in printouts.
4. You will be given draft paper for rough work. You are NOT allowed to write on anything else. The paper will be collected at the end of a quiz and cannot be taken away from the site.

5. Use the Respondus Lockdown Browser to access course site at NTULearn.
6. Quiz at "SC3050 -> Labs -> Lab X -> Lab X Online Quiz".
7. You are NOT allowed to click the "Begin" button until you are told to do so.
8. After clicking "Begin" and entering the quiz password, you have 15 minutes sharp to finish the quiz. Please start immediately.
9. NTULearn will automatically count the timer and end the quiz. Please avoid submitting your answer in the last second as we are not able to extend the quiz timer for anyone.
10. The system will automatically record your answers.
11. You can always resume the quiz if there is any technical issue during the quiz. No data will be lost but the timer will keep counting even when you leave the test.

Textbooks and Expected Reading

Textbooks

1. David A. Patterson, John L. Hennessy, Computer Architecture: A Quantitative Approach, Morgan Kaufmann, 6th edition, 2017, ISBN: 978-0128119051.
2. David B. Kirk, Wen Mei W. Hwu, Programming Massively Parallel Processors, A Hands-on Approach, Morgan Kaufmann, 2nd edition, 2013, ISBN: 978-0124159921.

Reference materials

1. Ian McLoughlin, Computer Architecture: An Embedded Approach, McGraw-Hill Education (Asia), 2011, ISBN: 978-0071311182.
2. David A. Patterson, John L. Hennessy, Computer Organization and Design: The Hardware/ Software Interface, Morgan Kaufmann, 5th edition, 2013, ISBN: 978-0124078864