10M11CI114: High Performance Computer Architecture

Course Credit: 3
Semester: M.Tech, I

Introduction

High Performance Computer Architecture is a masters-level course which covers the design of advanced modern computing systems. In particular, the design of modern microprocessors, characteristics of the memory hierarchy, and issues involved in multi-threading and multi-processing are discussed. The main objective of this course is to provide students with an understanding and appreciation of the fundamental issues and tradeoffs involved in the design and evaluation of modern computers. Topics will include cost/performance analysis, design and evaluation of instruction set architectures, pipelining techniques, multi-level memory hierarchies, superscalar processor design, multi-threading and multi-processing. Through programming and analysis assignments students will build, in stages, a timing simulator for a simplified out-of-order multiple-issue microprocessor in order to examine the impact of various architectural techniques.
The format of the course will be lecture-discussions, assignments. Students are strongly encouraged to participate actively in class discussions.

Course Objectives (Post-conditions)
Knowledge objectives:

At the conclusion of the course, following learning objectives are expected to be achieved:

1. You will gain basic knowledge required to design an anlyze high performance computer systems.
2. You will become aware of parallel architecture in modern intel i3, i5 and ARM based computer systems.
3. You will increase your knowledge about variour parallel computing model in modern systems.
4. You will know how parallelism is achieved using various pipelining techniques in ARM and Intel high performance systems.
5. You will learn how to evaluate and analyze cost and performance of multi processor systems.
6. You will learn various type of interconnection networks used to achive high performance in modern sysetms.
7. You will learn how various type of memories are used in parallel architecture to achieve data parallelism.
8. You will acquire the background for understanding next-generation GPUs using CUDA.
9. You will learn how to achive parallelism using pipelining and memory arganization in CUDA architecture using GPUs.

Application objectives:

The project and homework portions of the course are intended to help you apply your understanding,

1. To develop and simulate assembly language programs on pipelined and serial architecture.
2. To develop, implement, and demonstrate the learning through a project that meet stated specifications.
3. To develop, implement and debug CUDA language program that meet stated specifications.
4. To understand and be able to explain different parallel architectures, interconnections and various memory organization in modern high performance architectures.

to lay a foundation for pursuing some additional career options.

Expected Student Background (Preconditions)

Students are expected to have a solid grasp of the fundamentals of computer system, including a basic understanding of computer organization and microcontrollers. In addition, students are expected to know application development environment and programming concepts. Assembly programming ability will be helpful, as we will be looking at implementations of microprocessor instruction set operations.

Topics Outline:

S NO	Topics	Hrs
1	Introduction to High performance computing. Overview of the modern computer performance improvements, RISC processors, RISC Vs CISC, Levels of parallelism(instruction, transaction, task, thread, memory, function), Models (SIMD, MIMD, SIMT, SPMD, Dataflow Models, Demand-driven Computation etc), Architectures: N-wide superscalar architectures, multi-core, multi-threaded	6
2	Program and Network Properties ,Conditions of parallelism, program partition and scheduling ,Program Flow Mechanisms, System Interconnect Architecture	6
3	Principles of scalable Performance, performance Metrics and Measures, Parallel processing Applications , Speedup Performances Laws, Scalability and approaches	6
4	Basic concepts of pipelining, Arithmetic pipelines,	6

	Instruction pipelines Hazards in a pipeline: structural, data, and control hazards Overview of hazard resolution techniques, Dynamic instruction scheduling Branch prediction techniques, Instruction-level parallelism using software approaches ,Superscalar techniques Speculative execution ,Review of modern processors. a. Intel Processor ,b. ARM Processor	
5	Memory Hierarchies, Basic concept of hierarchical memory organization Main memories, Cache memory design and implementation, Virtual memory design and implementation, Secondary memory technology, RAID	6
6	Centralized vs. distributed shared memory ,Interconnection topologies, Multiprocessor architecture, Multiprocessor system interconnects, cache coherence and synchronization Mechanism, Three generations of Multicomputer, Message Passage Mechanism	6
7	Parallel Processing with GPU (CUDA), Processor Architecture, Interconnect, Communication, Memory Organization, and Programming Models in high performance computing architectures: (Examples: IBM CELL BE, Nvidia Tesla GPU), Memory hierarchy and transaction specific memory design, Thread Organization	6
	Total	42

References

1. Computer Architecture: A Quantitative Approach, 4th Edition by John L. Hennessy and David A. Patterson. ISBN-10: 0123704901.
2. Advanced Computer Architecture, Parallelism, Scalability, Programmability, Kai Hwang, Mc Graw Hill International Editions, Computer Science Series.
3. Parallel Computer Architecture A Hardware / Software Approach David E. Culler, Jaswinder Pal Singh with Anoop Gupta, Morgan Kaufmann.

Evaluation Scheme:

S.No	Examination	Marks
1	T-1	15
2	T-2	25
3	T-3	35
4	*Internal Marks	25

*Internal Marks Breakdown:

Assignments 9 marks (3x3)

Quizzes 12 marks (3x4)

Regularity 4 Marks