

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY
(Autonomous) Kanuru, Vijayawada-520007
Department of Computer Science & Engineering (AI&ML)

IV B. Tech – I Semester

High Performance Computing

Course Code	23AM4702C	Year	IV	Semester	I
Course Category	PEC	Branch	CSE (AI&ML)	Course Type	Theory
Credits	3	L-T-P	3-0-0	Prerequisites	Computer Fundamentals
Continuous Internal Evaluation	30	Semester End Evaluation	70	Total Marks	100

Course Outcomes		
Upon Successful completion of course, the student will be able to		
CO1	Describe fundamentals of parallel computing, interactions, programming models, communication operations, performance analysis, and parallel algorithms to learn parallel system design principles.	L2
CO2	Apply communication and analytical methods to measure their efficiency in matrix-based computations.	L3
CO3	Use parallel algorithm design techniques and programming methods, GPU computing and CUDA to solve computational problems.	L3
CO4	Analyze modern parallel computing systems, programming models, communication mechanisms, and GPU-based architectures, to assess their performance and efficiency.	L4

[illegible]

Syllabus		
Unit No	Contents	Map ped CO
I	Introduction: Motivating Parallelism, Scope of Parallel Computing, Parallel Programming Platforms: Implicit Parallelism, Trends in Microprocessor and Architectures, Limitations of Memory, System Performance, Dichotomy of Parallel Computing Platforms, Physical Organization of Parallel Platforms, Communication Costs in Parallel Machines, Scalable design principles	CO1
II	Parallel Programming: Principles of Parallel Algorithm Design: Preliminaries, Decomposition Techniques, Characteristics of Tasks and Interactions, Mapping Techniques for Load Balancing, Methods for Containing Interaction Overheads, Parallel Algorithm Models, The Age of Parallel Processing, the Rise of GPU Computing, A Brief History of GPUs, Early GPU.	CO1, CO3, CO4
III	Basic Communication: Operations- One-to-All Broadcast and All-to-One Reduction, All-to-All Broadcast and Reduction, All-Reduce and Prefix-Sum Operations, Scatter and Gather, All-to-All Personalized Communication, Circular Shift, Improving the Speed of Some Communication Operations. Programming shared address space platforms: threads- basics, synchronization, OpenMP programming	CO1, CO2, CO4
IV	Analytical Models: Sources of overhead in Parallel Programs, Performance Metrics for Parallel Systems, and The effect of Granularity on Performance, Scalability of Parallel Systems, Minimum execution time and minimum cost, optimal execution time. Dense Matrix Algorithms: MatrixVector Multiplication, Matrix-Matrix Multiplication.	CO1, CO2
V	Parallel algorithms- sorting and graph: Issues in Sorting on Parallel Computers, Bubble Sort and its Variants, Parallelizing Quick sort, All-Pairs Shortest Paths, Algorithm for sparse graph, Parallel Depth-First Search, Parallel BestFirst Search. CUDA Architecture: CUDA Architecture, Using the CUDA Architecture, Applications of CUDA Introduction to CUDA C-Write and launch CUDA C kernels, Manage GPU memory, Manage communication and synchronization, Parallel programming in CUDA- C.	CO1, CO3, CO4

Learning Resources
Text Books
1. Introduction to Parallel Computing, Ananth Grama, Anshul Gupta, George Karypis, and Vipin Kumar, 2nd edition, 2003, Addison-Wesley.
2. CUDA by Example, Jason sanders, Edward Kandrot, 1st Edition, 2010, Addison-Wesley.
References
1. Scalable Parallel Computing, Kai Hwang, 1st Edition, 1998, McGraw Hill
2. CUDA Programming: A Developer's Guide to Parallel Computing with GPUs, Shane Cook, 1st Edition, 2013, Morgan Kaufmann.
3. Parallel Computer Architecture: A Hardware/ Software Approach, David Culler Jaswinder Pal Singh, 1st Edition, 1999, Morgan Kaufmann.
4. Essential Algorithms, Rod Stephens, 2nd Edition, 2013, Wiley.
E-Recourses and other Digital Material
1. https://onlinecourses.nptel.ac.in/noc20_me61/preview
2. https://nptel.ac.in/courses/106104025
3. https://developer.nvidia.com/cuda
4. https://www.youtube.com/results?search_query=cuda+programming+tutorial