CS5173 – High Performance Computing

Teaching plan:

# of Hours	Topic
1	Implicit parallelism, Limitations of memory system performance
1	Dichotomy of parallel computing platforms, PRAM Model
2	Communication model, physical organization (Interconnection networks)
1	Cache coherence in multiprocessor systems
1	Routing mechanisms and communication costs of parallel platforms Mapping techniques
2	Pthread library introduction, attributes of threads, mutexes and condition objects, Implementation of barrier, reader-writer lock, computing π value
2	OpenMP introduction, parallel and synchronization directives, global functions and environment variables, computing π value
2	Introduction to parallel algorithm design, Decomposition, Tasks, Dependency Graphs, Interaction Graphs
1	Characteristics of interactions and tasks, mapping techniques
1	Methods for reducing interaction overheads, Parallel algorithm models
5	Basic communication algorithms- One-to-all broadcast and all-to-one reduction, All-to-all broadcast and reduction operations, All reduce and prefix sums operations, Scatter and gather operations and all-to-all personalized communication, Circular shift operation and improving the communication operations
1	Analytical modeling of parallel systems- Sources of overhead, performance metrics
3	Scalability of parallel systems, iso-efficiency function, Minimum execution time and minimum cost optimal execution time, Other scalability metrics – scaled speedup
1	Programming using message passing paradigm: Principles, building blocks, essential MPI functions
2	Topologies and embedding, Overlapping communication and computation, collective communication operations, Groups and communicators
4	Sequential and parallel algorithms for matrix - vector multiplication, matrix-matrix multiplication, solving system of linear equations
4	Sorting networks, Bubble sort, Quick sort, Bucket and sample sort, Other sorting algorithms – enumeration sort and radix sort
1	Graph algorithms: Minimum spanning tree, single source shortest paths
3	All-pairs shortest paths, Transitive closure, Connected components, algorithms for sparse graphs
4	*An Overview of Brief History of GPUs; An Overview of GPU Programming; An Overview of GPU Memory Hierarchy Features; Introduction to Heterogeneous Computing – OpenCL; The OpenCL Kernel, The OpenCL Memory Model, The OpenCL Execution Model; OpenCL Platform and Devices; OpenCL Execution Environment, An Overview of OpenCL API; Heterogeneous Programming in OpenCL An Overview of CUDA enabled NVIDIA GPUs, Introduction to CUDA C, Parallel Programming in CUDA C*

Assignment:

S.No.	Programming Problems
1	Matrix-Matrix multiplication – simple, Cannon's, DNS algorithm
2	Gaussian elimination – simple, partial pivoting, Back-substitution (triangular system of equations)
3	Soting – Bitonic, Odd-even transposition, Shellsort, Quicksort, Bucket, Radix
4	Minimum spanning tree, Single source shortest paths, All-pairs shortest paths – Dijkstra's algorithm, Floyd's algorithm
5	Transitive closure, Connected components, Maximal independent set for sparse graphs

Reading:

1. Ananth Grama, Anshul Gupta, George Karypis, Vipin Kumar, *Introduction to Parallel Computing*, Second Edition, Pearson Education, 2007
2. Michael J. Quinn, *Parallel Computing-Theory and Practice*, McGraw-Hill International Editions, Computer Science Series, 2004
3. Benedict R Gaster, Lee Howes, David R Kaeli Perhaad Mistry Dana Schaa, *Heterogeneous Computing with OpenCL*, McGraw-Hill, Inc. Newyork , 2011
4. Jason Sanders, Edward Kandrot, *CUDA By Example – An Introduction to General-Purpose GPU Programming*, Addison Wesley, 2011.

Evaluation:

a. Programming Assignments - 30 Marks
b. Minor Tests – Two Tests 30 Marks
c. End Semester Examination – 40 Marks

TOTAL 100 Marks