



Midterm Review

Parallel Computing - COSC 3P93

Introduction

- Reasoning for parallel computing
- Definition
 - Parallel, concurrent, and distributed
- von Neumann Architecture
 - **Extensions**
 - Multitasking and threads
 - Caches → mapping
 - Virtual memory
 - Instruction-level Parallelism (ILP)
 - Pipelining and Multiple issue
 - Multithreading
 - Hardware and SMT
- **Terminology**

Parallel Hardware

- **Flynn's Classical Taxonomy**
 - SISD, SIMD, MISD, MIMD
- **Shared memory systems** → multiprocessor
 - Definition and characteristics
 - Advantages and disadvantages
 - Classes of multiprocessor architectures
 - Process-to-process mapping
 - Anonymous vs dedicated processors
 - Organization
 - UMA → SMP
 - Symmetry
 - NUMA
 - Issues in Shared memory
 - Access latency and memory conflicts
 - Solutions → drawbacks

Parallel Hardware

- **Distributed memory systems**
 - Advantages and disadvantages
 - Communication lines → topologies
- **Interconnect networks**
 - Shared memory → contention
 - Bus and crossbar
 - Distributed memory
 - Elements → mode of operation, control strategy, switching techniques and topology
 - Direct interconnect
 - Bisection width
 - Bandwidth
 - Indirect interconnects
 - Latency and bandwidth
 - Cache coherence → snooping and directory
 - False sharing

Parallel Software

- **Data parallel and task parallel**
- **Shared memory**
 - Threads → dynamic vs static threads
 - Nondeterminism
 - Coordinating Processes/threads
 - Race conditions → critical regions
 - Locks, mutex, semaphores, monitors
 - Transactions
- **Distributed memory**
 - MPI → blocking, nonblocking, broadcast, reduction
- **One-sided Communication**
- **Partitioned Global Address Space Languages**
- **I/O**

Performance

- **Performance** → definition
- Embarrassingly parallel computations
- **Scalability** → definition
 - Evaluations → sequential vs parallel
 - Types
 - **Strongly scalable** and **weakly scalable**
- Performance metrics
 - **Speedup**, **efficiency**, and **cost**
- Limits and costs
 - **Amdahl's law** → definition
 - Major limiting factors
 - Overhead
 - **Gustafson-Barsis' Law** → definition

Performance

- **Workflow** → DAG
 - **Work-span**
 - **Lower and upper bounds**
- Asymptotic complexity analysis (O , Ω , Θ)
- Why applications not scaling
 - **Reasons**
 - Seq performance, critical paths, bottlenecks, algorithmic overhead, communication overhead, load imbalance, speculative loss
- **Isoefficiency** → definition