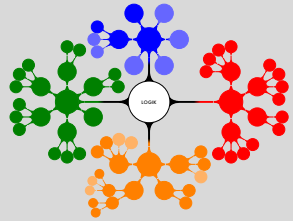
Lecture 2

Introduction I

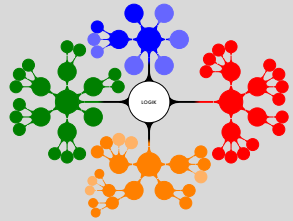
View of the Field

IKC-MH.57 *Introduction to High Performance and Parallel Computing* at October 20, 2023

Dr. Cem Özdoğan
Engineering Sciences Department
İzmir Kâtip Çelebi University

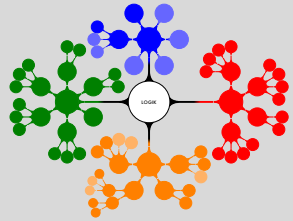


Contents



Field I

- Data-intensive applications;
  - transaction processing,
  - information retrieval,
  - data mining and analysis,
  - multimedia services,
  - computational physics/chemistry/biology and nanotechnology.
- High performance may come from
  - fast dense circuitry,
  - parallelism.
- Parallel processors are computer systems consisting of
  - multiple *processing units*
  - connected via some *interconnection network*
  - plus the software needed to make the processing units work together.



Field II

- *Uniprocessor* – Single processor supercomputers have achieved great speeds and have been pushing hardware technology to the physical limit of chip manufacturing.
  - Physical and architectural bounds (Lithography, μm size, destructive quantum effects.
  - Proposed solutions are maskless lithography process and nanoimprint lithography for the semiconductor).
  - Uniprocessor systems can achieve to a limited computational power and not capable of delivering solutions to some problems in reasonable time.
- *Multiprocessor* – Multiple processors cooperate to jointly execute a single computational task in order to speed up its execution.

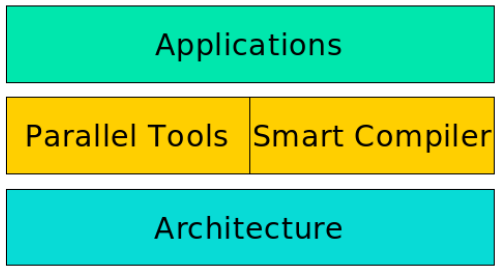
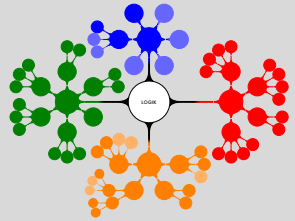


Figure: Abstraction Layers



Field III

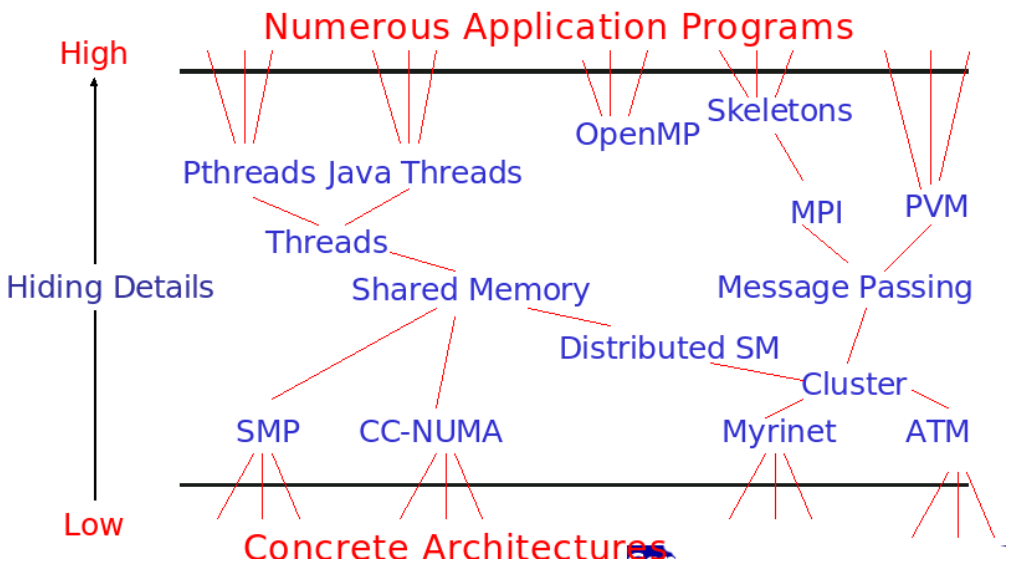
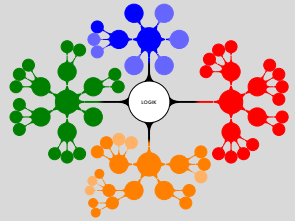


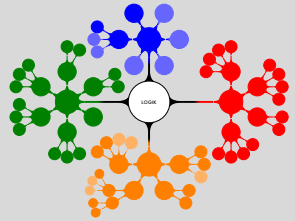
Figure: View of the Field

- New issues arise;
  - Multiple threads of control vs. single thread of control
  - Partitioning for concurrent execution
  - Task Scheduling
  - Synchronization
  - Performance



Trends

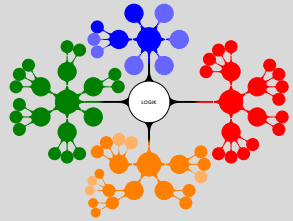
- Past Trends in Parallel Architecture (inside the box)
  - Completely custom designed components; *processors, memory, interconnects, I/O*.
  - The first three are the major components for the aspects of the parallel computation.
    - Longer R&D time (2-3 years).
    - Expensive systems.
    - Quickly becoming outdated.
  - In the form of internally linked processors was the main form of parallelism.
  - Advances in computer networks $\Rightarrow$ in the form of networked autonomous computers.
- New Trends in Parallel Architecture (outside the box)
  - Instead of putting everything in a single box and *tightly couple* processors to memory, the Internet achieved a kind of parallelism by *loosely* connecting everything outside of the box.
  - Network of PCs and workstations connected via LAN or WAN forms a Parallel System.
  - Compete favourably (cost/performance).
  - Utilize unused cycles of systems sitting idle.



Four Decades of Computing

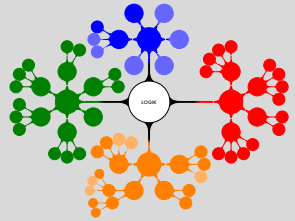
Most computer scientists agree that there have been four distinct paradigms or eras of computing. These are: batch, time-sharing, desktop, and network.

1. Batch Era
2. Time-Sharing Era
3. Desktop Era
4. Network Era. They can generally be classified into two main categories:
   1. shared memory,
   2. distributed memory systems.
   - The number of processors in a single machine ranged from several in a shared memory computer to hundreds of thousands in a massively parallel system.
   - Examples of parallel computers during this era include Sequent Symmetry, Intel iPSC, nCUBE, Intel Paragon, Thinking Machines (CM-2, CM-5), MsPar (MP), Fujitsu (VPP500), and others.
5. Current Trends: Clusters, Grids.



Flynn's Taxonomy of Computer Architecture I

- The most popular taxonomy of computer architecture was defined by Flynn in 1966.
- Flynn's classification scheme is based on the notion of a stream of information.
  - Two types of information flow into a processor:
    1. **Instruction**. The instruction stream is defined as the sequence of instructions performed by the processing unit.
    2. **Data**. The data stream is defined as the data traffic exchanged between the memory and the processing unit.
- According to Flynn's classification, either of the instruction or data streams can be **single** or **multiple**.
- Computer architecture can be classified into the following four distinct categories:
  1. single instruction single data streams (SISD)
  2. single instruction multiple data streams (SIMD)
  3. multiple instruction single data streams (MISD)
  4. multiple instruction multiple data streams (MIMD).



Flynn's Taxonomy of Computer Architecture II

- SISD;

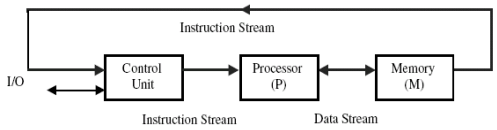


Figure: SISD Architecture.

- SIMD;

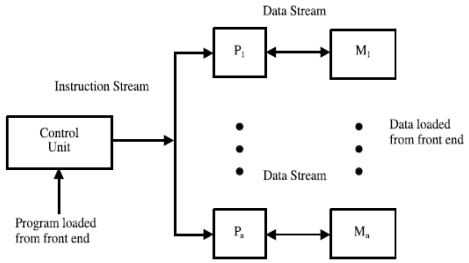


Figure: SIMD Architecture.

- MIMD;

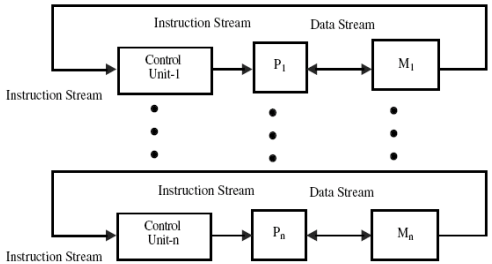
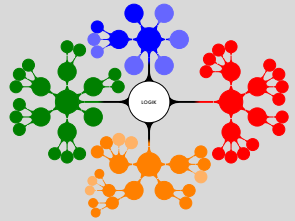


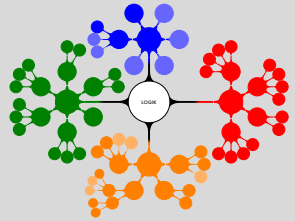
Figure: MIMD Architecture.



Flynn's Taxonomy of Computer Architecture III

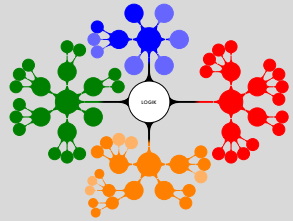
Parallel computers are either SIMD or MIMD.

- When there is only one control unit and all processors execute the same instruction in a synchronized fashion, the parallel machine is classified as SIMD.
- In a MIMD machine, each processor has its own control unit and can execute different instructions on different data.
- In the MISD category, the same stream of data flows through a linear array of processors executing different instruction streams. In practice, there is no viable MISD machine; however, some authors have considered *pipelined machines* as examples for MISD.



Parallel and Distributed Computers I

- The processing units can communicate and interact with each other using either
  - shared memory
  - or message passing methods.
- The interconnection network for shared memory systems can be classified as
  - bus-based
  - switch-based.
- SIMD Computers
- MIMD Shared Memory, MIMD Distributed Memory
- Bus based, Switch based
- CC-NUMA
- Clusters, Grid Computing
  - Grids are geographically distributed platforms for computation.
  - They provide dependable, consistent, general, and inexpensive access to high end computational capabilities.



Parallel and Distributed Computers II

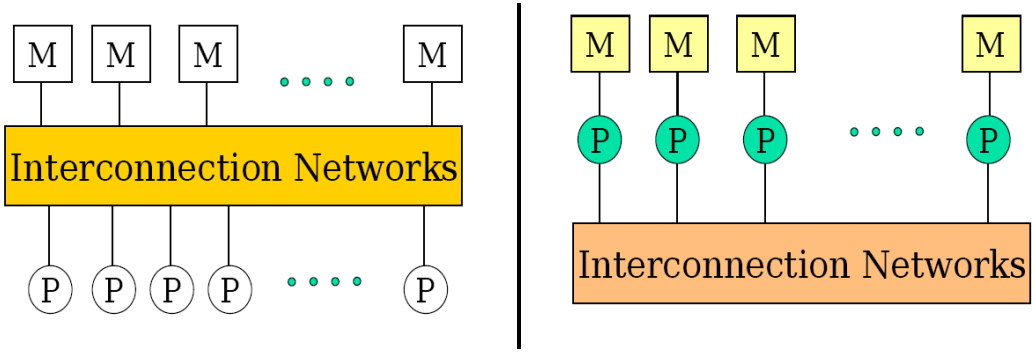


Figure: (a) MIMD Shared Memory, (b) MIMD Distributed Memory.

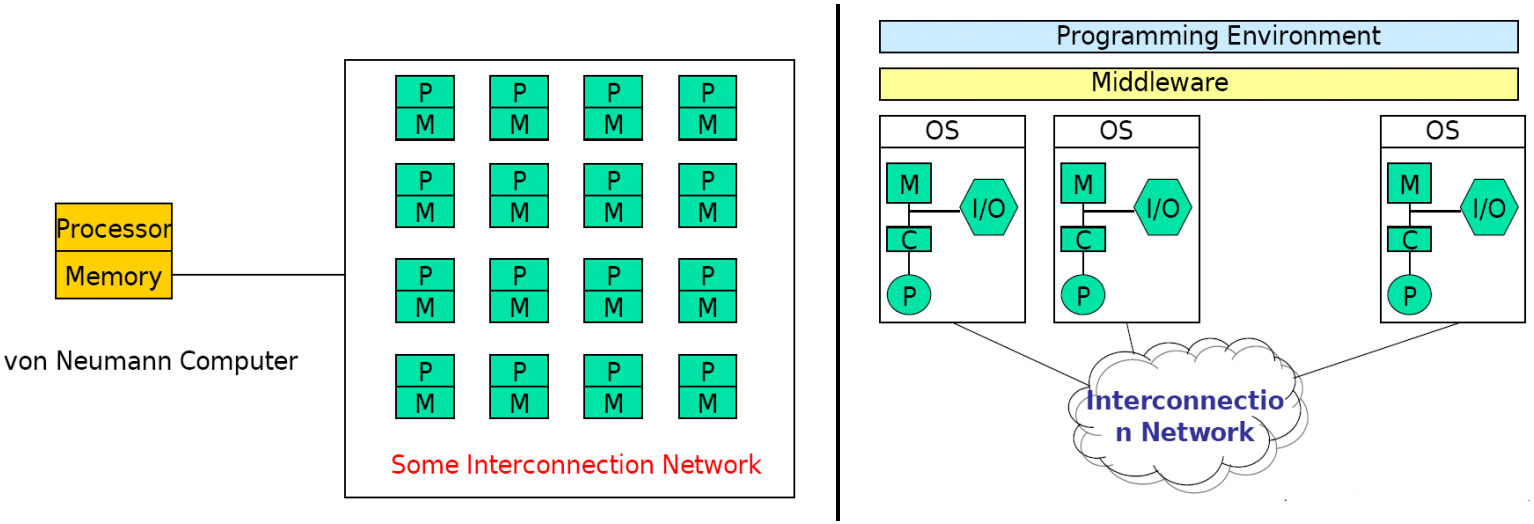


Figure: (a) SIMD Distributed Computers, (b) Clusters.