University of Nizhni Novgorod
Faculty of Computational Mathematics & Cybernetics

Introduction to Parallel Programming

Section 1.

Overview of Parallel Computer Systems

Gergel V.P., Professor, D.Sc.,
Software Department

Contents

- Preconditions of Parallel Computing
- Types of Parallel Computer Systems
  - Supercomputers
  - Clusters
- Taxonomy of Parallel Computer Systems
  - Multiprocessors
  - Multicomputers
- Overview of Interconnection Networks
- Software System Platforms for High-Performance Clusters
- Summary

Preconditions of Parallel Computing…

The term *"parallel computation"* is generally applied to any data processing, in which several computer instructions can be executed simultaneously

Preconditions of Parallel Computing…

- Achieving parallelism is only possible if the following requirements are met:
  - **independent functioning of separate computer devices** (input/output devices, processors, storage devices,…),
  - **redundancy of computing system elements:**
    - *use of specialized devices* (separate processors for integer and float valued arithmetic, multilevel memory devices,…),
    - *duplication of computer devices* (separate processors of the same type or several RAM devices,…),
  - Processor **pipeline** implementation may be an additional form of achieving parallelism

Preconditions of Parallel Computing…

- **Modes** of independent program parts execution:
  - *Multitasking mode (time sharing mode)*, when a single processor is used for carrying out processes (this mode is pseudo-parallel as only one process can be active),
  - *Parallel execution*, when several instructions of data processing can be carried out simultaneously (can be provided if several processors are available and by means of pipeline and vector processing devices),
  - *Distributed computations,* which involves the use of several processing devices located at a distance from each other, the data transmission through communication lines among the processing devices leads to considerable time delays

Preconditions of Parallel Computing

Here we will discuss the second type of parallel computing in multiprocessor computer systems

Types of parallel computer systems…

❑ Supercomputers

Supercomputer is a computational system, whose processing power is the best of all systems power at the current moment

Types of parallel computer systems…

- **Supercomputers. ASCI** (*Accelerated Strategic Computing Initiative*)
  - **1996**, **ASCI Red** system, developed by Intel Corp., with the performance of 1 TFlops,
  - **1999**, **ASCI Blue Pacific** by IBM and **ASCI Blue Mountain** by SGI, with the performance of 3 TFlops,
  - **2000**, **ASCI White**, the peak performance was higher than 12 TFlops (the computing power which was actually demonstrated in LINPACK test was 4938 GFlops)

Types of parallel computer systems…

- **Supercomputers. ASCI White…**
  - ASCI White hardware is IBM RS/6000 SP system with 512 symmetric multiprocessor (SMP) nodes, each node has 16 processors,
  - All nodes are IBM RS/6000 POWER3 symmetric multiprocessors with 64 –bit architecture, processors are superscalar 64-bit pipeline chips with two devices processing floating point instructions and three integer instruction processing devices. They are able to execute up to eight integer instructions per clock cycle and up to four floating point instructions per clock cycle. The clock cycle of each processor is 375 MHz,
  - The total RAM is 4 TBytes,
  - The capacity of the disk memory is 180 TBytes.

Types of parallel computer systems…

- **Supercomputers. ASCI White:**
  - The operating system is a UNIX – IBM AIX version,
  - ASCI White software supports a mixed programming model which means message transmission among the nodes and multi-treading among an SMP node,
  - MPI, OpenMP libraries, POSIX threads and a translator of IBM directives are supported. Moreover, there is also an IBM parallel debugger.

Types of parallel computer systems…

- **Supercomputers. BlueGene:**
  - It is still being developed, at present the current name of the system is "**BlueGene/L DD2 beta-System**", this is the first phase of the complete computer system,
  - Peak performance is forecasted to reach 360 TFlops by the time the system is put into the final configuration,
  - The features of the current variant of the system:
    - 32 racks with 1024 dual-kernel 32-bit PowerPC 440 0.7 GHz processors in each;
    - Peak performance is approximately 180 TFlops;
    - The maximum processing power demonstrated by LINPACK test is 135 TFlops.

Types of parallel computer systems…

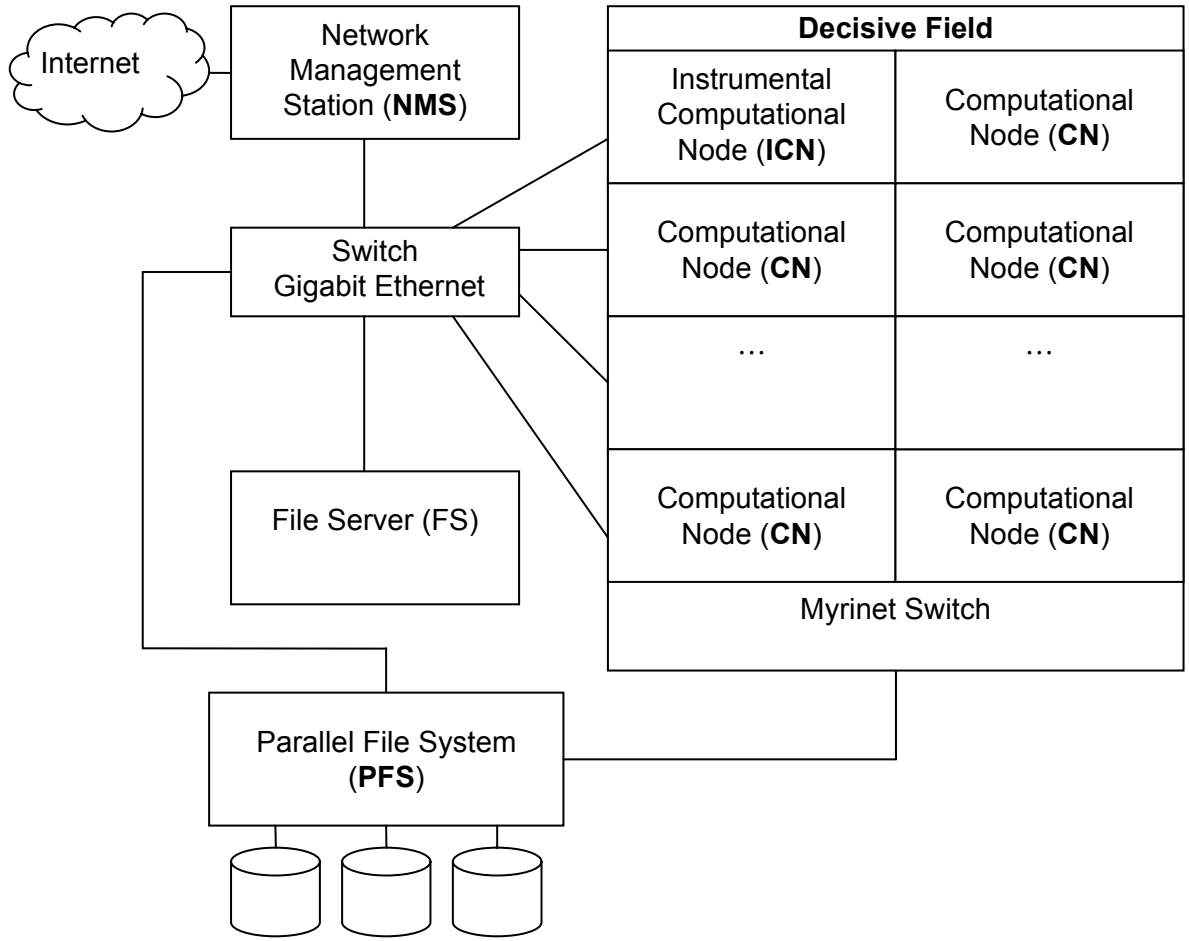
- **Supercomputers. MBC-15000…**

  (*Interdepartmental Supercomputer Center of Russian Academy of Science*)

  – The total number of nodes is 276 (552 processors), each computational nodes includes:
    - 2 IBM PowerPC 970 processors with 2.2 GHz, cache L1 96 Kb and cache L2 512 Kb,
    - 4 Gb RAM per node,
    - 40 Gb IDE hard disc
  – SuSe Linux Enterprise Server operating systems, version 8 for the platforms x86 and PowerPC,
  – Peak performance is 4857.6 GFlops and the maximum processing power demonstrated in LINPACK test is 3052 GFlops.

Types of parallel computer systems…

❑ Supercomputers. MBC-15000



Types of parallel computer systems…

❑ Clusters

A cluster is group of computers connected in a local area network (LAN). A cluster is able to function as a unified computational resource.

It implies higher *reliability* and *efficiency* than an LAN as well as a considerably *lower cost* in comparison to the other parallel computer system types (due to the use of standard hardware and software solutions).

Types of parallel computer systems…

- **Clusters. Beowulf…**
  - Nowadays a "*Beowulf*" type cluster is a system which consists of a server node and one or more client nodes which are connected with the means of Ethernet or some other network. The system is made of commodity off-the-shelf components able to operate under Linux, standard Ethernet adaptors and switches. It does not contain any specific hardware and can be easily reproduced.

Types of parallel computer systems…

- **Clusters. Beowulf:**
  - **1994**, NASA Goddard Space Flight Research Center, the cluster was created under Thomas Sterling and Don Becker's supervision:
    - 16 computers based on 486DX4 100 MHz processors,
    - Each node had 16 Mb RAM,
    - The connection of the nodes was provided by three 10Mbit/s network adaptors,
    - Linux operating system, GNU compiler and MPI library.

Types of parallel computer systems…

- **Clusters. Avalon**
  - **1998**, Avalon system, Los Alamos National Laboratory (USA), the supervisor was astrophysicist Michael Warren:
    - 68 processors (later it was expanded up to 140) Alpha21164A with the clock frequency 533 MHz,
    - 256 Mb RAM, 3 Gb HDD, Fast Ethernet card on the each node,
    - Linux operating system,
    - The peak performance was 149 GFlops and the computing power of the 48.6 GFlops demonstrated in LINPACK test.

Types of parallel computer systems…

- **Clusters. AC3 Velocity Cluster**
  - **2000**, Cornell University (USA), AC3 Velocity Cluster was the result of the university collaboration with AC3 (Advanced Cluster Computing Consortium) established by Dell, Intel, Microsoft, Giganet and 15 more software manufacturers:
    - 64 four-way servers Dell PowerEdge 6350 on the base of Intel Pentium III Xeon 500 MHz, 4 GB RAM, 54 GB HDD, 100 Mbit Ethernet card,
    - 1 eight-way server Dell PowerEdge 6350 based on Intel Pentium III Xeon 550 MHz, 8 GB RAM, 36 GB HDD, 100 Mbit Ethernet card,
    - Microsoft Windows NT 4.0 Server Enterprise Edition operating system,
    - Peak performance is 122 GFlops, processing power of 47 GFlops demonstrated in LINPACK test.

Types of parallel computer systems…

- **Clusters. NCSA NT Supercluster**
  - **2000**, National Center for Supercomputing Applications (USA):
    - 38 two-way systems Hewlett-Packard Kayak XU PC workstation on the base of Intel Pentium III Xeon 550 MHz, 1 Gb RAM, 7.5 Gb HDD, 100 Mbit Ethernet card,
    - Microsoft Windows operating system,
    - Peak performance is 140 GFlops and the processing power of 62 GFlops demonstrated in LINPACK test.

Types of parallel computer systems…

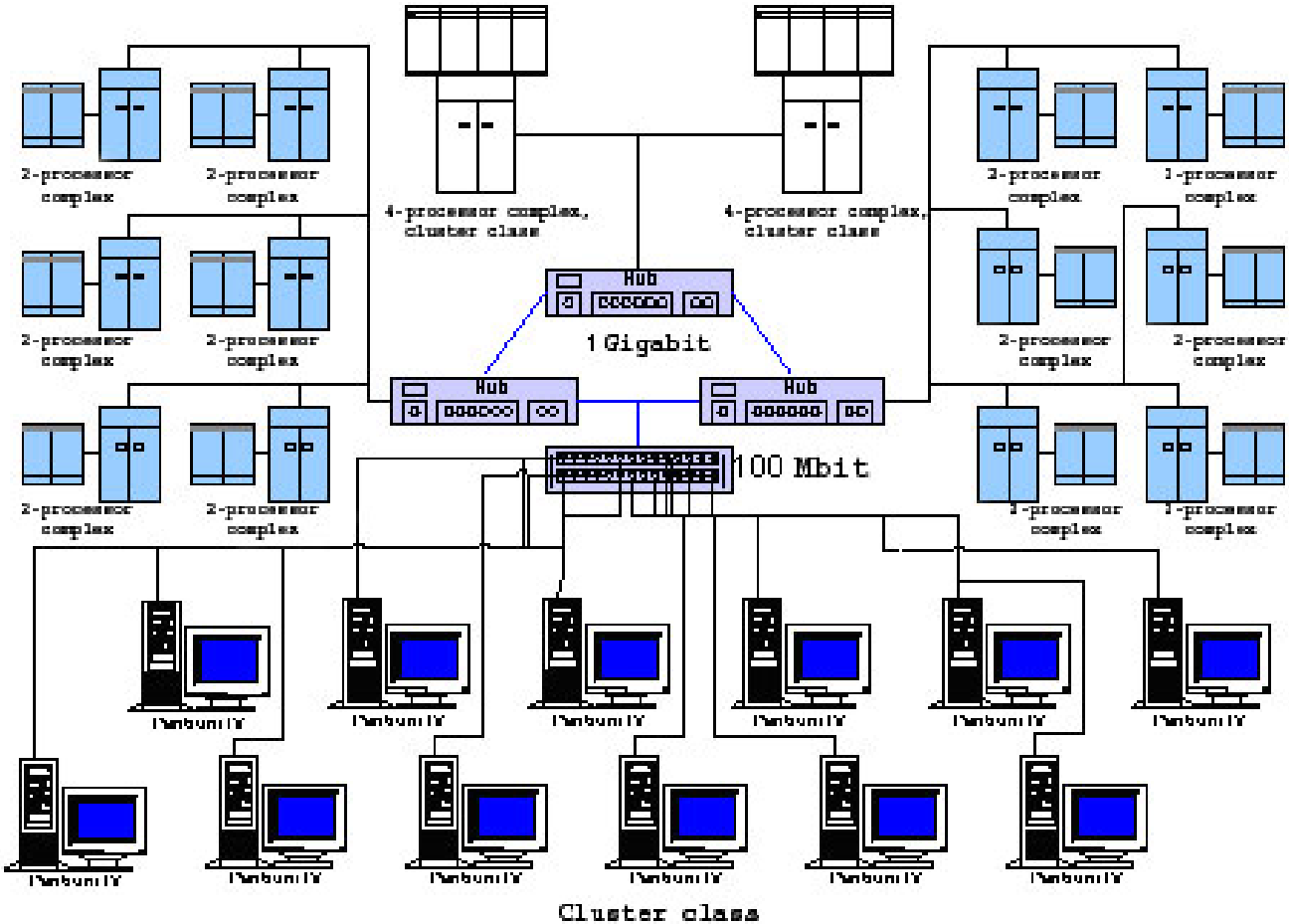
- **Clusters. Thunder**
  - **2004**, Livermore National Laboratory (USA) :
    - 1024 servers with 4 Intel Itanium 1.4 GHz processors in each,
    - 8 Gb RAM per node,
    - Total disc capacity 150 Tb,
    - Operating system CHAOS 2.0,
    - At present Thunder Cluster with its performance 22938 GFlops and the maximum one shown in LINPACK test 19940 GFlops takes the 5th position of the Top500 (in the summer of 2004 it occupied the 2nd position)

Types of parallel computer systems…

- **Clusters. NNSU Computational Cluster…**
  - **2001**, University of Nizhni Novgorod, the equipment was donated by Intel:
    - 2 computational servers, each has 4 processors Intel Pentium III 700 Mhz, 512 MB RAM, 10 GB HDD, 1 Gbit Ethernet card,
    - 12 computational servers, each has 2 processors Intel Pentium III 1000 Mhz, 256 MB RAM, 10 GB HDD, 1 Gbit Ethernet card,
    - 12 workstations based on Intel Pentium 4 1300 Mhz, 256 MB RAM, 10 GB HDD, 10/100 Fast Ethernet card,
    - Microsoft Windows operating system.

Types of parallel computer systems

- **Clusters. NNSU Computational Cluster**



Taxonomy of Parallel Computer Systems…

❑ Flynn's taxonomy

– Flynn's taxonomy is the best-known classification scheme for computer systems. It provide to specify the multiplicity of hardware used to operate instruction and data streams:

- **SISD** (Single Instruction, Single Data)
- **SIMD** (Single Instruction, Multiple Data)
- **MISD** (Multiple Instruction, Single Data)
- **MIMD** (Multiple Instruction, Multiple Data)

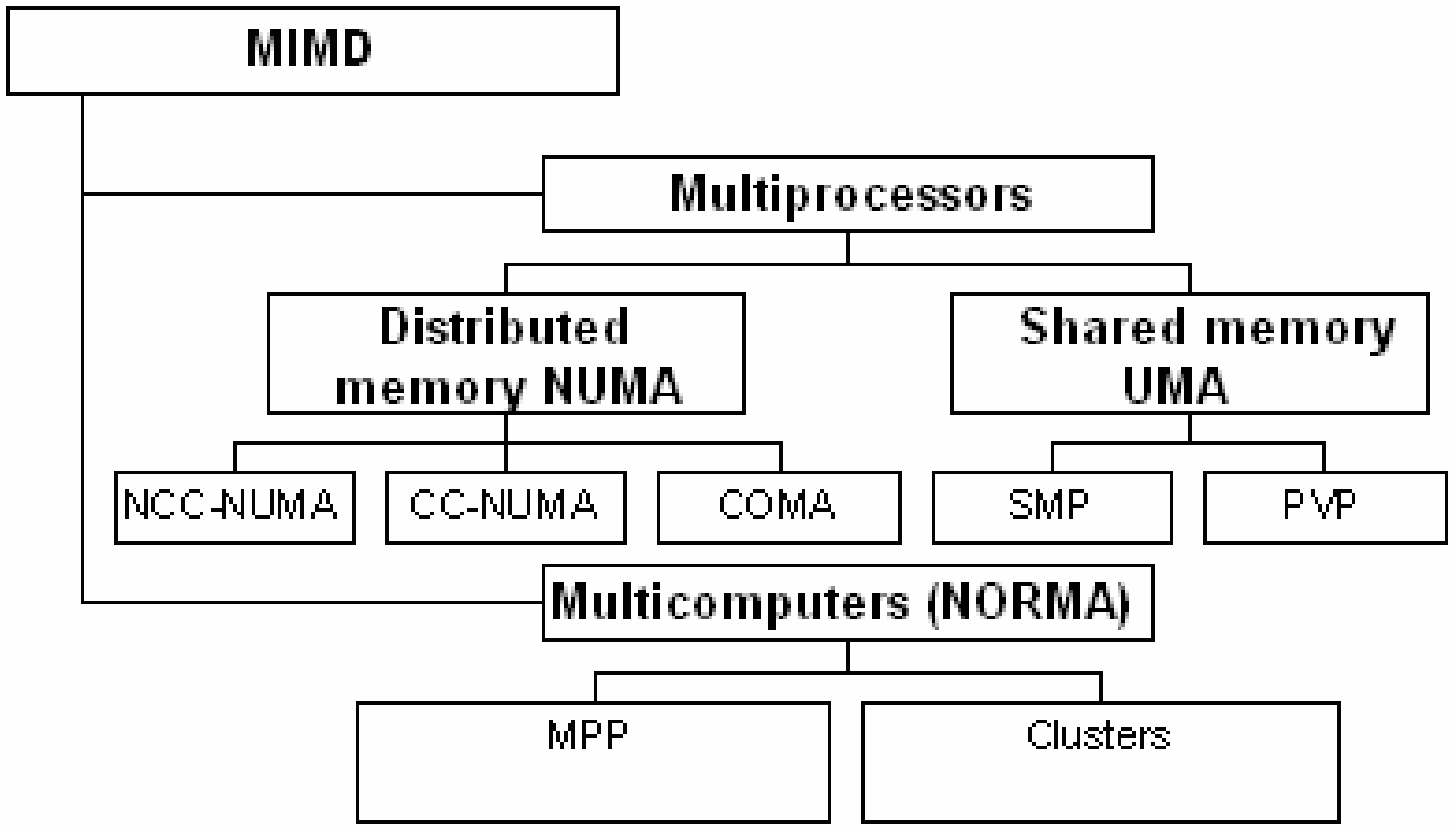
*All the parallel systems despite their considerable heterogeneity belong to the same group **MIMD**.*

Taxonomy of Parallel Computer Systems…

- **Flynn's taxonomy, further MIMD classification…**
  - Is focused on ability of the processor to access to all memory of computer system,
  - Allows differentiating between the two important multiprocessor system types:
    - ***multiprocessors*** or multiprocessor systems with shared memory,
    - ***multicomputers*** or multiprocessor systems with distributed memory.

Taxonomy of Parallel Computer Systems…

- Flynn's taxonomy, further MIMD classification

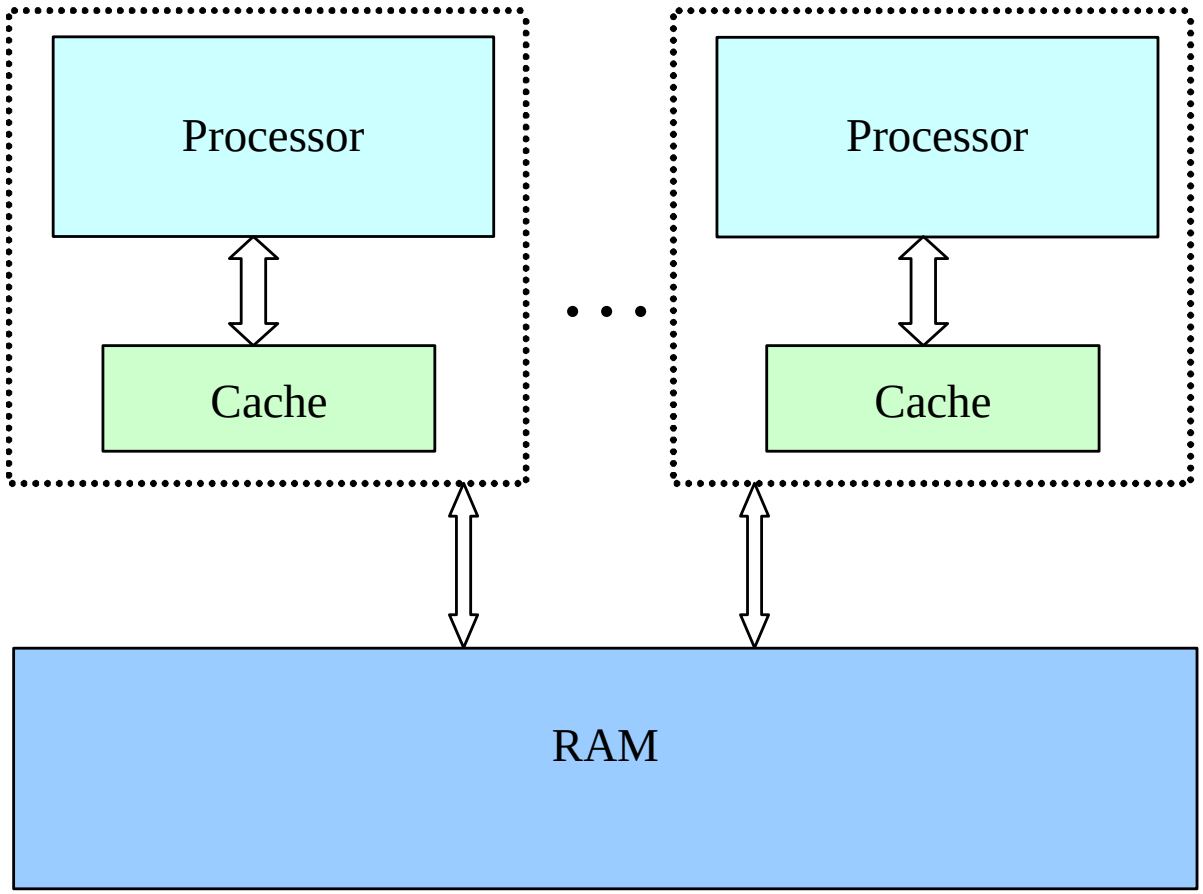


Taxonomy of Parallel Computer Systems…

- ***Multiprocessors*** (systems with shared memory)…
  - ensure uniform memory access (*UMA*),
  - serve as the basis for designing:
    - *parallel vector processors (PVP)*, e.g.: Cray T90,
    - *symmetric multiprocessor (SMP),* e.g.: IBM eServer, Sun StarFire, HP Superdome, SGI Origin.

Taxonomy of Parallel Computer Systems…

- **Multiprocessors** (case of single centralized shared memory)…



Taxonomy of Parallel Computer Systems…

- **Multiprocessors** (case of single centralized shared memory)

Problems:

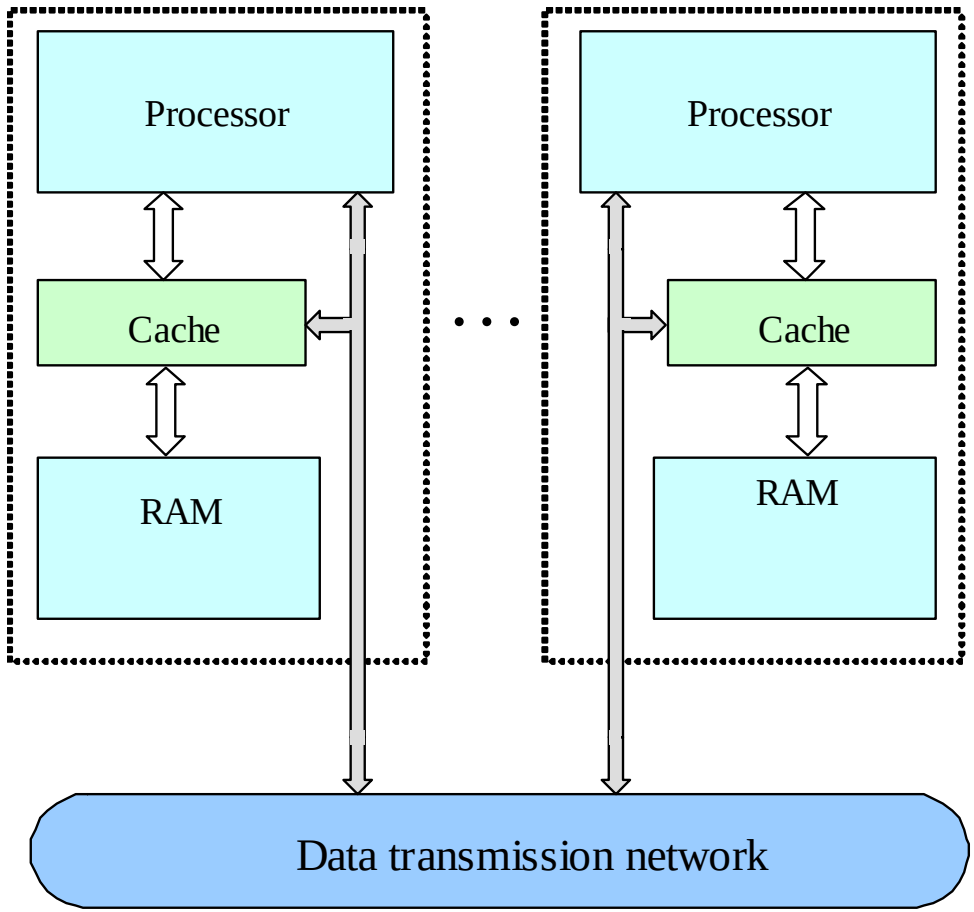
- access to the shared data from different processors and providing in this relation the coherence of different cache contents (*cache coherence problem*),
- the necessity to synchronize the interactions of simultaneously carried out instruction streams.

Taxonomy of Parallel Computer Systems…

- **Multiprocessors** (case of *distributed shared memory* or *DSM*)…
  - *non-uniform memory access or NUMA*,
  - The systems with such memory type fall into the following groups:
    - *Cache-only memory architecture or COMA* (e.g.: KSR-1 and DDM systems),
    - *cache-coherent NUMA or CC-NUMA* (e.g.: SGI Origin 2000, Sun HPC 10000, IBM/Sequent NUMA-Q 2000),
    - *non-cache coherent NUMA or NCC-NUMA* (e.g.: Cray T3E).

Taxonomy of Parallel Computer Systems…

- **Multiprocessors** (case of *distributed shared memory*)…



Taxonomy of Parallel Computer Systems…

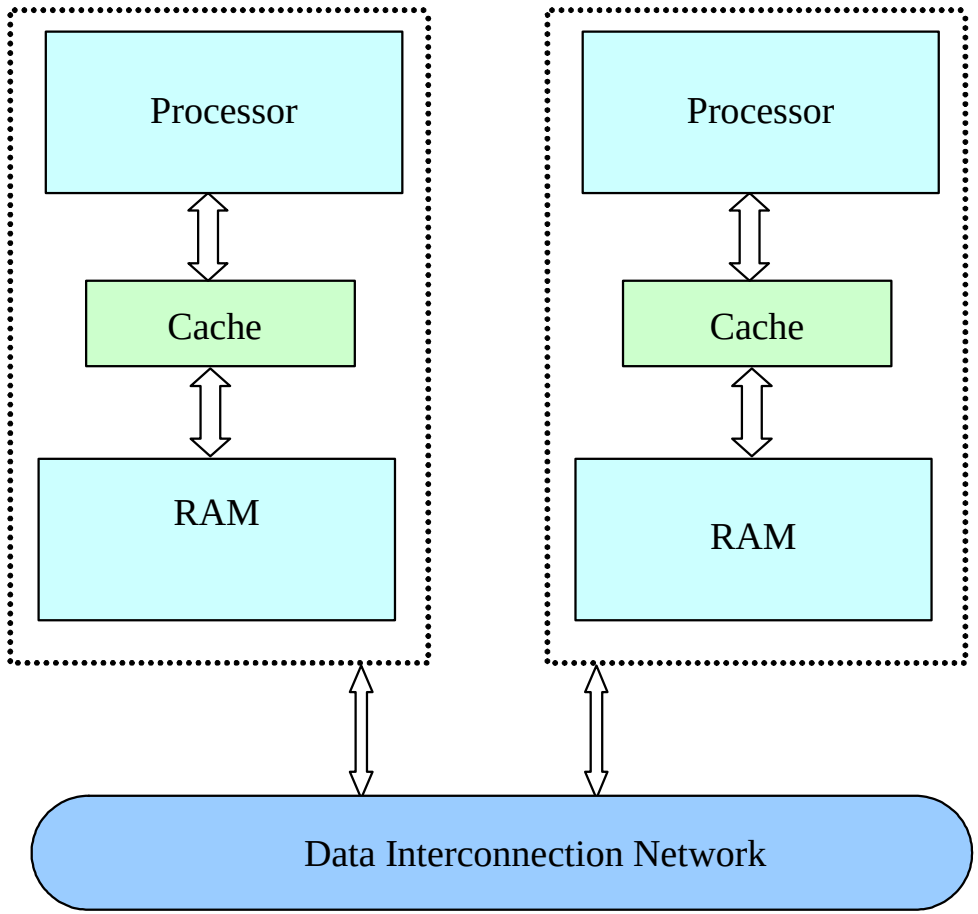
- **Multiprocessors** (case of *distributed shared memory*):
  - simplify the problems of large multiprocessor system design (nowadays NUMA systems can come across computers with several thousands processors),
  - the rising problems of efficient use of distributed shared memory (time to access to local and remote memory may be several orders different) causes a significant increase of parallel programming complexity

Taxonomy of Parallel Computer Systems…

- **Multicomputers…**
  - *no-remote memory access or NORMA*,
  - each system processor is able to use only its local memory,
  - getting access to the data available on other processors requires explicit execution of *message passing operations*.

Taxonomy of Parallel Computer Systems…

❑ Multicomputers…



Taxonomy of Parallel Computer Systems…

- **Multicomputers…**

  This approach is used in developing two important types of multiprocessor computing systems:

  – *massively parallel processor or MPP*, e.g.: IBM RS/6000 SP2, Intel PARAGON, ASCI Red, Parsytec transputer system,

  – *clusters*, e.g.: AC3 Velocity and NCSA NT Supercluster.

Taxonomy of Parallel Computer Systems…

- **Multicomputers. Clusters…**

*The **cluster** is usually defined as a set of separate computers connected into a network. Single system image, availability of reliable functioning and efficient performance for these computers are provided by special software and hardware*

Taxonomy of Parallel Computer Systems…

- **Multicomputers. Clusters…**

  Advantages:

  – Clusters can be either created on the basis of separate computers available for consumers or constructed of standard computer units, this allows to cut down on costs,

  – The increase of computational power of separate processors makes possible to create clusters using a relatively small number (several tens) of separate processors (*lowly parallel processing*),

  – It allows to subdivide into only large independent parts (*coarse granularity*) in the computational algorithm for parallel execution.

Taxonomy of Parallel Computer Systems

- **Multicomputers. Clusters**

Problems:

- Arranging the interaction among computational cluster nodes with the use of data transmission usually leads to considerable time delays,
- Additional restrictions for the type of parallel algorithms and programs being developed (*low intensity of streams of data transmission*).

Overview of Interconnection Networks…

Data transmission among the processors of computer system is used to provide interaction, synchronization and mutual exclusion of parallel processes executed at the time of parallel computations.

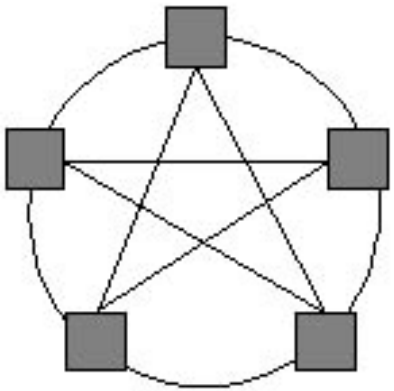
*Data interconnection network **topology** is the structure of communication links among the processors of computer system*

Overview of Interconnection Networks…

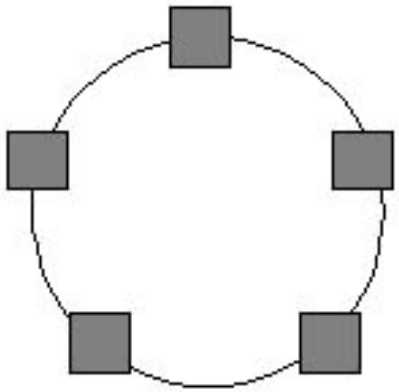
- The following processor communication schemes are usually referred to the basic topologies:
  - **Completely-Connected Graph or Clique** is a system where each pair of processors is connected by means of a direct communication link,
  - **Linear Array** or **Farm** where all the processors are enumerated in order and each processor except the first and the last ones has communication links only with the neighboring processors,
  - **Ring** can be derived from a linear array if the first processor of the array is connected to the last one,
  - **Star,** where all the processors are connected by means of communication links to some managing processor,
  - **Mesh,** where the graph of the communication links creates a rectangular mesh,
  - **Hypercube** is a particular case of mesh topology, where there are only two processors on each mesh dimention.

Overview of Interconnection Networks…

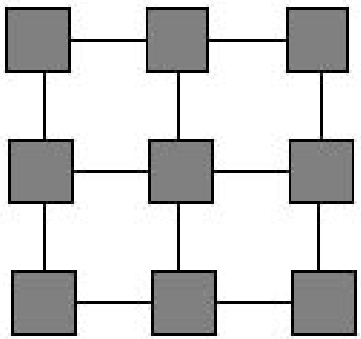
- **Topologies of Multiprocessor Interconnection Networks**



Completely-Connected Graph or *Clique*

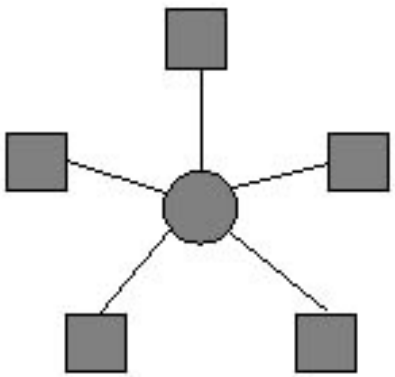


Ring

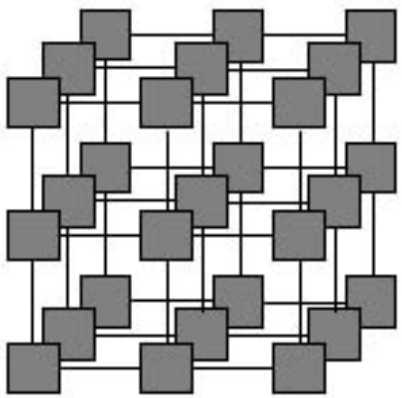


Mesh

Linear Array or Farm



Star



Overview of Interconnection Networks…

- **Network Topology of Computational Cluster:**
  - In many cases a switch through which all cluster processors are connected with each other is used to build a cluster system,
  - The simultaneous execution of several transmission operations is limited.

At any given moment of time each processor can participate only in one data transmission operation

Overview of Interconnection Networks…

- **Topology Network Characteristics…**
  - ***diameter*** determines the maximum distance between two network processors; this value can characterize the maximum time necessary to transmit the data between processors,
  - ***connectivity*** is the minimum number of edges which have to be necessary removed for partitioning the data interconnection network into two disconnected parts,
  - ***bisection width*** is the minimum number of edges which have to be obligatory eliminated for partitioning the data interconnection network into two disconnected parts of the same size,
  - ***cost*** is the total number of data transmission links in a multiprocessor computer system.

Overview of Interconnection Networks…

❑ Topology Network Characteristics

Topology	Diameter	Bisection width	Connectivity	Cost
Complete Graph	1	$p^2/4$	(p-1)	p(p-1)/2
Star	2	1	1	(p-1)
Farm	p-1	1	1	(p-1)
Ring	$\lfloor p/2 \rfloor$	2	2	p
Hypercube	$Log_2(p)$	p/2	$Log_2(p)$	$pLog_2(p)/2$
Mesh (N=2)	$2\lfloor \sqrt{p}/2 \rfloor$	$2\sqrt{p}$	4	2p

Software System Platforms for High-Performance Clusters

- To be added

Summary

- In the beginning the requirements to hardware for providing parallel calculations are discussed
- The difference between multitask, parallel and distributed modes of program execution is considered
- A number of parallel computer systems are given
- The best-known classification of computer systems – *Flynn's taxonomy* – is presented
- Two important groups of parallel computer systems are studied - the systems with shared and distributed memory (*multiprocessors* and *multicomputers*)
- Finally overview of interconnection networks is given

Discussions

- What are the main ways to achieve parallelism?
- What differences between parallel computer systems exist?
- What is Flynn's taxonomy based on?
- What is the principle of multiprocessor systems subdivision into multiprocessors and multicomputers?
- What are the advantages and disadvantages of cluster systems?
- What topologies are widely used in interconnection networks of multiprocessor systems?
- What are the features of data transmission networks for clusters?
- What are the basic characteristics of interconnection networks?
- What software system platforms for high-performance clusters can be used?

Exercises

- Give some additional examples of parallel computer systems
- Consider some additional methods of computer systems classification
- Consider the ways of cache coherence provision in the systems with shared memory
- Make a review of the program libraries which provide carrying out data transmission operations for the systems with distributed memory
- Consider the binary tree topology of interconnection network
- Give examples of efficiently realized computational problems for each type of interconnection network topologies

References

- **Barker, M.** (Ed.) (2000). Cluster Computing Whitepaper at http://www.dcs.port.ac.uk/~mab/tfcc/WhitePaper/.
- **Buyya, R.** (Ed.) (1999). High Performance Cluster Computing. Volume1: Architectures and Systems. Volume 2: Programming and Applications. - Prentice Hall PTR, Prentice-Hall Inc.
- **Culler, D., Singh, J.P., Gupta, A.** (1998) Parallel Computer Architecture: A Hardware/Software Approach. - Morgan Kaufmann.
- **Dally**, W.J., **Towles, B.P.** (2003). Principles and Practices of Interconnection Networks. - Morgan Kaufmann.
- **Flynn, M.J.** (1966) Very high-speed computing systems. Proceedings of the IEEE 54(12): P. 1901-1909.

References

- **Hockney, R. W., Jesshope, C.R.** (1988). Parallel Computers 2. Architecture, Programming and Algorithms. - Adam Hilger, Bristol and Philadelphia. (русский перевод 1 издания: Хокни Р., Джессхоуп К. Параллельные ЭВМ. Архитектура, программирование и алгоритмы. - М.: Радио и связь, 1986)
- **Kumar V., Grama A., Gupta A., Karypis G.** (1994). Introduction to Parallel Computing. - The Benjamin/Cummings Publishing Company, Inc. (2nd edn., 2003)
- **Kung, H.T.** (1982). Why Systolic Architecture? Computer 15 № 1. P. 37-46.
- **Patterson, D.A., Hennessy J.L.** (1996). Computer Architecture: A Quantitative Approach. 2d ed. - San Francisco: Morgan Kaufmann.

References

- **Pfister, G. P.** (1995). In Search of Clusters. - Prentice Hall PTR, Upper Saddle River, NJ (2nd edn., 1998).
- **Sterling, T.** (ed.) (2001). Beowulf Cluster Computing with Windows. - Cambridge, MA: The MIT Press.
- **Sterling, T.** (ed.) (2002). Beowulf Cluster Computing with Linux. - Cambridge, MA: The MIT Press
- **Tanenbaum, A.** (2001). Modern Operating System. 2nd edn. – Prentice Hall
- **Xu, Z., Hwang, K.** (1998). Scalable Parallel Computing Technology, Architecture, Programming. – Boston: McGraw-Hill.

Next Section

- Modeling and Analysis of Parallel Computations

Author's Team

Gergel V.P., Professor, Doctor of Science in Engineering, Course Author

Grishagin V.A., Associate Professor, Candidate of Science in Mathematics

Abrosimova O.N., Assistant Professor (chapter 10)

Kurylev A.L., Assistant Professor (learning labs 4,5)

Labutin D.Y., Assistant Professor (ParaLab system)

Sysoev A.V., Assistant Professor (chapter 1)

Gergel A.V., Post-Graduate Student (chapter 12, learning lab 6)

Labutina A.A., Post-Graduate Student (chapters 7,8,9, learning labs 1,2,3, ParaLab system)

Senin A.V., Post-Graduate Student (chapter 11, learning labs on Microsoft Compute Cluster)

Liverko S.V., Student (ParaLab system)

About the project

The purpose of the project is to develop the set of educational materials for the teaching course “Multiprocessor computational systems and parallel programming”. This course is designed for the consideration of the parallel computation problems, which are stipulated in the recommendations of IEEE-CS and ACM Computing Curricula 2001. The educational materials can be used for teaching/training specialists in the fields of informatics, computer engineering and information technologies. The curriculum consists of **the training course “Introduction to the methods of parallel programming”** and **the computer laboratory training “The methods and technologies of parallel program development”**. Such educational materials makes possible to seamlessly combine both the fundamental education in computer science and the practical training in the methods of developing the software for solving complicated time-consuming computational problems using the high performance computational systems.

The project was carried out in Nizhny Novgorod State University, the Software Department of the Computing Mathematics and Cybernetics Faculty (http://www.software.unn.ac.ru). The project was implemented with the support of Microsoft Corporation.