

Scalability and Classifications

1 Types of Parallel Computers

- MIMD and SIMD classifications
- shared and distributed memory multicomputers
- distributed shared memory computers

2 Network Topologies

- static connections
- dynamic network topologies by switches
- ethernet connections

MCS 572 Lecture 2
Introduction to Supercomputing
Jan Verschelde, 11 January 2012

Scalability and Classifications

1 Types of Parallel Computers

- **MIMD and SIMD classifications**
- shared and distributed memory multicomputers
- distributed shared memory computers

2 Network Topologies

- static connections
- dynamic network topologies by switches
- ethernet connections

the classification of Flynn

In 1966, Flynn introduced the following classification:

- SISD** = Single Instruction Single Data stream
one single processor handles data sequentially
use pipelining (e.g.: car assembly) to achieve parallelism
- MISD** = Multiple Instruction Single Data stream
called systolic arrays, has been of little interest
- SIMD** = Single Instruction Multiple Data stream
graphics computing, issue same command for pixel matrix
vector and arrays processors for regular data structures
- MIMD** = Multiple Instruction Multiple Data stream
this is the general purpose multiprocessor computer

Single Program Multiple Data Stream

One model is SPMD: Single Program Multiple Data stream:

- 1 All processors execute the same program.
- 2 Branching in the code depends on the identification number of the processing node.

Manager worker paradigm:

- manager (also called root) has identification zero,
- workers are labeled $1, 2, \dots, p - 1$.

Scalability and Classifications

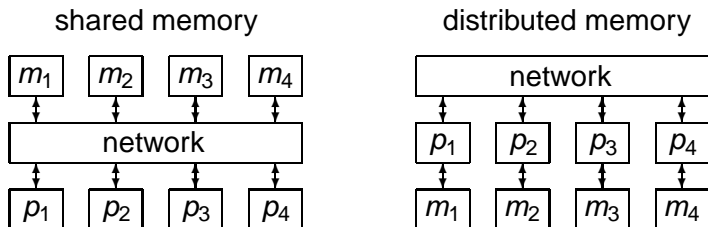
1 Types of Parallel Computers

- MIMD and SIMD classifications
- **shared and distributed memory multicomputers**
- distributed shared memory computers

2 Network Topologies

- static connections
- dynamic network topologies by switches
- ethernet connections

types of parallel computers



One crude distinction concerns memory, shared or distributed:

- A shared memory multicomputer has one single address space, accessible to every processor.
- In a distributed memory multicomputer, every processor has its own memory accessible via messages through that processor.

Many parallel computers consist of multicore nodes.

clusters

Definition

A cluster is an independent set of computers combined into a unified system through software and networking.

Beowulf clusters are scalable performance clusters based on commodity hardware, on a private network, with open source software.

What drove the clustering revolution in computing?

- 1 commodity hardware: choice of many vendors for processors, memory, hard drives, etc...
- 2 networking: Ethernet is dominating commodity networking technology, supercomputers have specialized networks;
- 3 open source software infrastructure: Linux and MPI.

Message Passing and Scalability

$$\text{total time} = \text{computation time} + \underbrace{\text{communication time}}_{\text{o v e r h e a d}}$$

Because we want to reduce the overhead, the

$$\text{computation/communication ratio} = \frac{\text{computation time}}{\text{communication time}}$$

determines the *scalability* of a problem:

*How well can we increase the problem size n ,
keeping p , the number of processors fixed?*

Desired: order of overhead \ll order of computation, so ratio $\rightarrow \infty$,
examples: $O(\log_2(n)) < O(n) < O(n^2)$.

Remedy: overlapping communication with computation.

Scalability and Classifications

1 Types of Parallel Computers

- MIMD and SIMD classifications
- shared and distributed memory multicomputers
- **distributed shared memory computers**

2 Network Topologies

- static connections
- dynamic network topologies by switches
- ethernet connections

distributed shared memory computers

In a distributed shared memory computer:

- memory is physically distributed with each processor, and
- each processor has access to all memory in single address space.

Benefits:

- 1 message passing often not attractive to programmers,
- 2 shared memory computers allow limited number of processors, whereas distributed memory computers scale well

Disadvantage: access to remote memory location causes delays and the programmer does not have control to remedy the delays.

Scalability and Classifications

1 Types of Parallel Computers

- MIMD and SIMD classifications
- shared and distributed memory multicomputers
- distributed shared memory computers

2 Network Topologies

- **static connections**
- dynamic network topologies by switches
- ethernet connections

some terminology

- bandwidth: number of bits transmitted per second
- on latency, we distinguish tree types:
 - message latency: time to send zero length message (or startup time),
 - network latency: time to make a message transfer the network,
 - communication latency: total time to send a message including software overhead and interface delays.
- diameter of network: minimum number of links between nodes that are farthest apart
- on bisecting the network:
 - bisection width : number of links needed to cut network in two equal parts,
 - bisection bandwidth: number of bits per second which can be sent from one half of network to the other half.

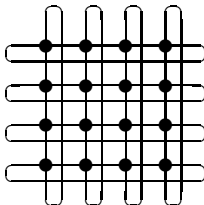
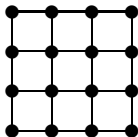
arrays, rings, meshes, and tori

Connecting p nodes in complete graph is too expensive.

An array and ring of 4 nodes:

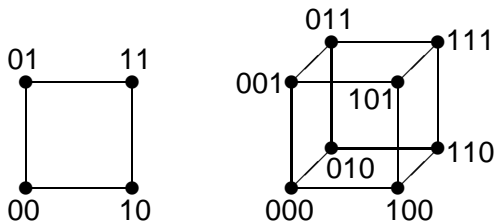


A matrix and torus of 16 nodes:



hypercube network

Two nodes are connected \Leftrightarrow their labels differ in exactly one bit.



e-cube or left-to-right routing: flip bits from left to right, e.g.: going from node 000 to 101 passes through 100.

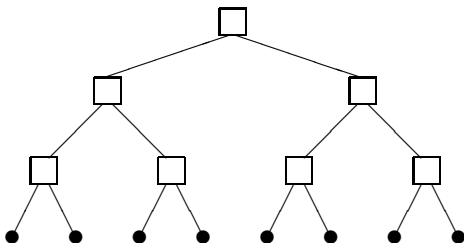
In a hypercube network with p nodes:

- maximum number of flips is $\log_2(p)$,
- number of connections is ...?

a tree network

Consider a binary tree:

- The leaves in the tree are processors.
- The interior nodes in the tree are switches.



Often the tree is *fat*:

with an increasing number of links towards the root of the tree.

Scalability and Classifications

1 Types of Parallel Computers

- MIMD and SIMD classifications
- shared and distributed memory multicomputers
- distributed shared memory computers

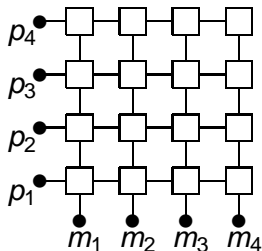
2 Network Topologies

- static connections
- **dynamic network topologies by switches**
- ethernet connections

a crossbar switch

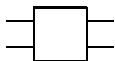
In a shared memory multicomputer, processors are usually connected to memory modules by a crossbar switch.

For example, for $p = 4$:



A p -processor shared memory computer requires p^2 switches.

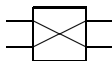
2-by-2 switches



a switch

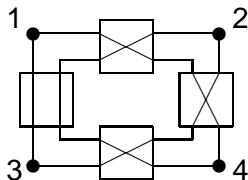
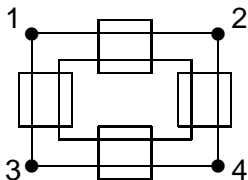


pass through



cross over

Changing from *pass through* to *cross over* configuration:

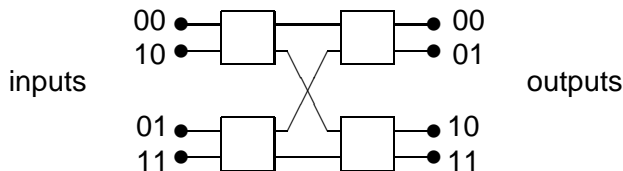


a multistage network

Rules in the routing algorithm:

- 1 bit is zero: select upper output of switch
- 2 bit is one: select lower output of switch

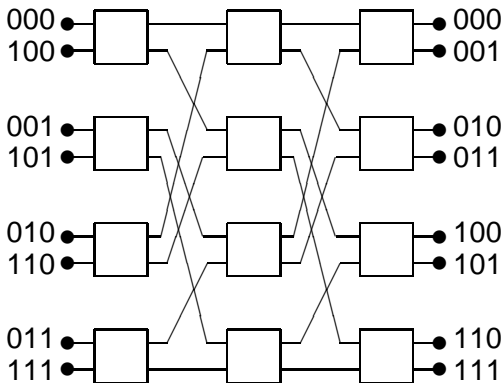
The first bit in the input determines the output of the first switch, the second bit in the input determines the output of the second switch.



The communication between 2 nodes using 2-by-2 switches causes *blocking*: other nodes are prevented from communicating.

a 3-stage Omega interconnection network

circuit switching:



number of switches for p processors: $\log_2(p) \times \frac{p}{2}$

circuit and packet switching

If all circuits are occupied, communication is blocked.

Alternative solution: packet switching:
message is broken in packets and sent through network.

Problems to avoid:

- deadlock:** Packets are blocked by other packets waiting to be forwarded.
This occurs when the buffers are full with packets.
Solution: avoid cycles using e-cube routing algorithm.
- livelock:** a packet keeps circling the network and fails to find its destination.

Scalability and Classifications

1 Types of Parallel Computers

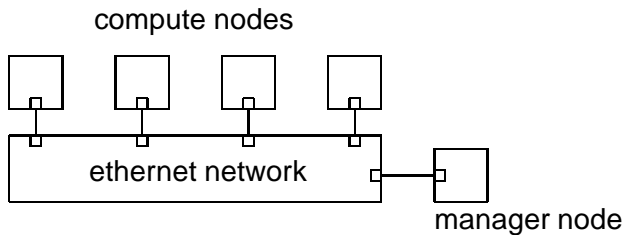
- MIMD and SIMD classifications
- shared and distributed memory multicomputers
- distributed shared memory computers

2 Network Topologies

- static connections
- dynamic network topologies by switches
- **ethernet connections**

ethernet connection network

Computers in a typical cluster are connected via ethernet.



personal supercomputing

The SiCortex SC072 is a desktop supercomputer:

- 72 processors, 12 nodes of 6 cores each,
- operates at 1GFlops,
- 48GB of memory,
- green supercomputing: low energy usage.

HP workstation Z800 RedHat Linux:

- two 6-core Intel Xeon at 3.47Ghz,
- 24GB of internal memory,
- 2 NVIDIA Tesla C2050 general purpose graphic processing units.

the ACCC cluster argo

<http://www.uic.edu/depts/accc/hardware/argo>

Beowulf cluster with 57 PCs running RedHat Linux 4.2, with one master node and 56 compute nodes (heterogeneous), and approximately 160GB of total memory.

The master node is used

- to create, edit, and compile programs,
- to submit jobs for execution on compute nodes.

User programs ***must not*** run on the master node.

Cluster monitoring system at <https://argo.cc.uic.edu/> (requires only UIC netid) allows to see job list & usage stats.

summary and recommended reading

We classified computers and introduced network terminology, we have covered chapter 1 of Wilkinson-Allen almost entirely.

Available to UIC via the ACM digital library:

- Michael J. Flynn and Kevin W. Rudd: **Parallel Architectures.**
ACM Computing Surveys 28(1): 67-69, 1996.

Available to UIC via IEEE *Xplore*:

- George K. Thiruvathukal: **Cluster Computing.**
Guest Editor's Introduction.
Computing in Science & Engineering 7(2): 11-13, 2005.

Visit www.beowulf.org.

Exercises

Homework will be collected at a to be announced date.

Exercises:

- 1 Derive a formula for the number of links in a hypercube with $p = 2^k$ processors for some positive number k .
- 2 Consider a network of 16 nodes, organized in a 4-by-4 mesh with connecting loops to give it the topology of a torus (or doughnut). Can you find a mapping of the nodes which give it the topology of a hypercube? If so, use 4 bits to assign labels to the nodes. If not, explain why.
- 3 We derived an Omega network for eight processors. Give an example of a configuration of the switches which is blocking, i.e.: a case for which the switch configurations prevent some nodes from communicating with each other.
- 4 Draw a multistage Omega interconnection network for $p = 16$.