Interconnection Networks

Ned Nedialkov

McMaster University
Canada

Outline

Shared-memory interconnects

- Most widely used are buses and crossbars
- Bus
  - Devices are connected to a common bus
  - Low cost and flexibility
  - Increasing the number of processes decreases performance, as wires are shared
  - As the number of processors increases, buses are replaced by switched interconnects
- Switched interconnects use switches to control the routing

Crossbar
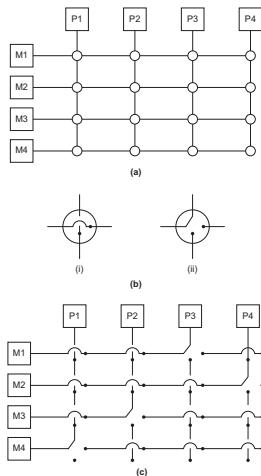
- ▶ Lines: bidirectional links, squares: cores or memory, circles: switches
- ▶ Allows simultaneous communication between nodes
- ▶ Faster than buses
- ▶ Cost of switches and links is relatively high

Figure from P. Pacheco, An Introduction to Parallel Programming, 2011

Distributed-memory interconnects

- ▶ Direct interconnects: each switch is connected to processor-memory
- ▶ (a) Ring: each switch has 3 links; p processors $2p$ links
- ▶ (b) Toroidal mesh: each switch has 5 links
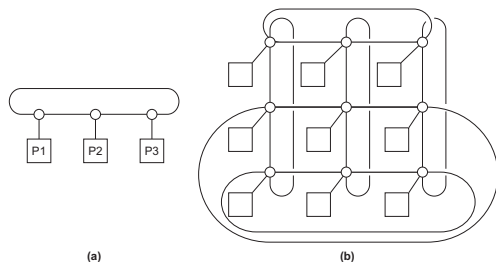 p processors: $2p$ links

(a)   (b)

Figure from P. Pacheco, An Introduction to Parallel Programming, 2011

- IBM's Blue Gene/L and Blue Gene/P, Cray XT3: three dimensional torus
- IBM's Blue Gene/Q: five dimensional torus
- Fujitsu K: six dimensional torus interconnect called Tofu

Bisection width

- ▶ The minimum number of links that can be removed to split the nodes into two halves
- ▶ The bisection width is a measure of how many communications can be done simultaneously
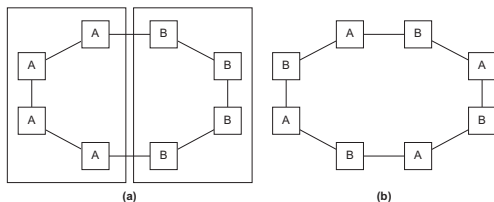- ▶ (a) 2, (b) 2

Figure from P. Pacheco, An Introduction to Parallel Programming, 2011

Example

- ▶ Two-dimensional toroidal mesh with $p = q^2$ nodes
- ▶ Remove middle horizontal and middle wrap around links
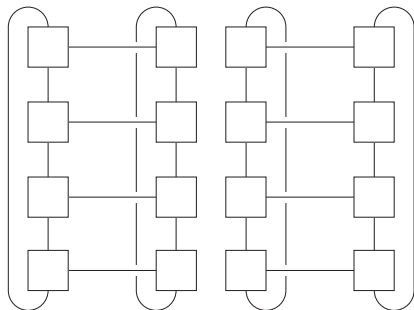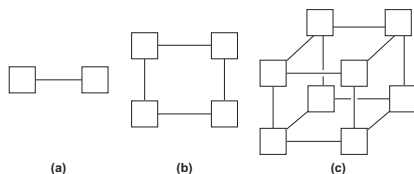- ▶ $2q = 2\sqrt{p}$ bisection width

Figure from P. Pacheco, An Introduction to Parallel Programming, 2011

Bandwidth

- Bandwidth is the rate at which data is transmitted
- Bisection bandwidth:
  (bisection width)$\times$ (bandwidth of a link)

Hypercube

- ▶ Hypercubes of dimension (a) 1, (b) 2, (c) 3
- ▶ Hypercube of dimension d has $p = 2^d$ nodes
- ▶ Bisection width is $p/2$

Figure from P. Pacheco, An Introduction to Parallel Programming, 2011

© 2013–16 Ned Nedialkov

Indirect interconnects

- The switches may not be connected directly to processors
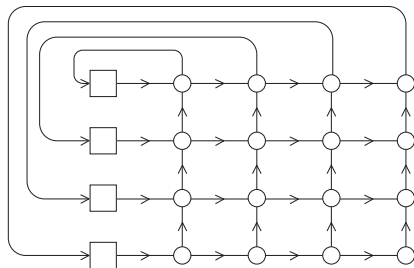- Crossbar with unidirectional links:

Figure from P. Pacheco, An Introduction to Parallel Programming, 2011

Omega network

- The switches are 2×2 crossbars
- $2p \log_2 p$ switches versus p^2 switches in a crossbar
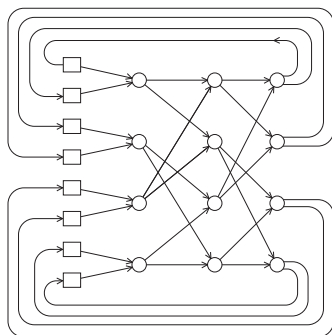
Figure from P. Pacheco, An Introduction to Parallel Programming, 2011

Blue Gene: 3D Torus

From TOP 500
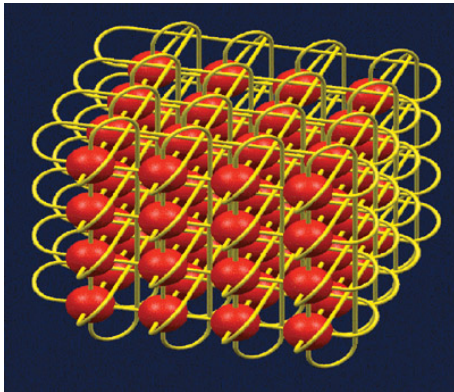
- http://www.top500.org/list/2013/11/, November 2013
- Performance is measured on the Linpack Benchmark
- Petaflops (Pflops), 10^{15} flops per second

no.	super computer	Pflops	cores
1	**Tianhe-2**, National University of Defense Technology, China	33.86	3,120,000
2	**Titan, Cray XK7**, Department of Energy (DOE) Oak Ridge National Laboratory, USA	17.59	560,640
3	**Sequoia, IBM BlueGene/Q**, Lawrence Livermore National Laboratory, USA	17.17	1,572,864
4	**Fujitsu K**, RIKEN Advanced Institute for Computational Science (AICS), Japan	10.51	705,024
5	**Mira, BlueGene/Q**, Argonne National Laboratory, USA	8.59	786,432

no.	super computer	Pflops	cores
6	**Piz Daint, Cray XC30**, Swiss National Supercomputing Centre (CSCS), Switzerland	6.27	115,984
89	**BGQ - BlueGene/Q**, Southern Ontario Smart Computing Innovation Consortium/University of Toronto	0.36	32,768

Highlights

- Tianhe-2 uses Intel Xeon Phi processors to speed up computational rate
- No. 2 Titan and No. 6 Piz Daint use NVIDIA GPUs to accelerate computation
- Total of 53 systems on the list use accelerator/co-processor technology
  - 38 use NVIDIA chips
  - 3 use ATI Radeon
  - 13 use use Intel MIC technology (Xeon Phi)

- ▶ Intel provides 82.4% of TOP500 systems processors
- ▶ 94% use processors with 6 or more cores
- ▶ 75% use processors with 8 or more cores
- ▶ Biggest users of HPC
    1. USA
    2. China
    3. Japan

	out of 500
USA	265
Asia	115
Europe	102
UK	23
France	22
Germany	20

TOP500 started in June 1993

Number 1:

year	super computer	Gflops	cores
Nov/1993	CM-5/1024, Thinking Machines Corp. Los Alamos, USA	59.7	1,024
Nov/2003	Earth-Simulator NEC, Japan Agency for Marine-Earth Science and Technology	35,860.0	5,120