Escuela de
Ingeniería y Arquitectura
Universidad Zaragoza

Departamento de
Informática e Ingeniería
de Sistemas
Universidad Zaragoza

Tema 11 – Redes Directas

Multiprocesadores

3er curso, Grado Ingeniería Informática
Especialidad Ingeniería Computadores

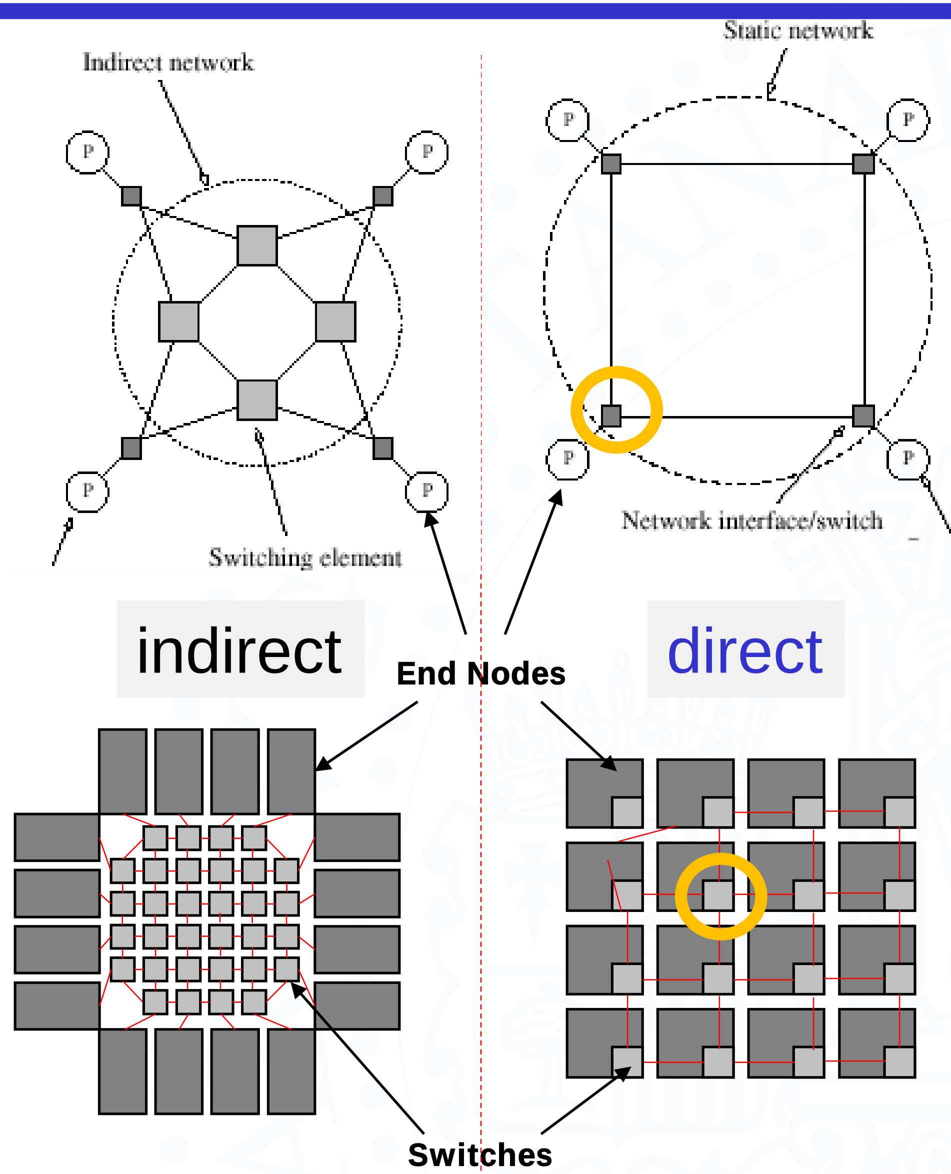
V. Viñals y J. Alastruey
Arquitectura y Tecnología de Computadores
Departamento de Informática e Ingeniería de Sistemas

Guión del tema

- Intro
- Topologías extremas:
  - conexión completa, estrella, línea, anillo
- Malla y toro
- Árbol (directa o indirecta)
- Hipercubo
- Métricas

Indirect vs direct: recap

- Indirect, dynamic networks
  - built out of links and switches
  - Two nodes of the "graph" only get and edge after a switching action
  - examples:
    - bus, crossbar, multistage, tree

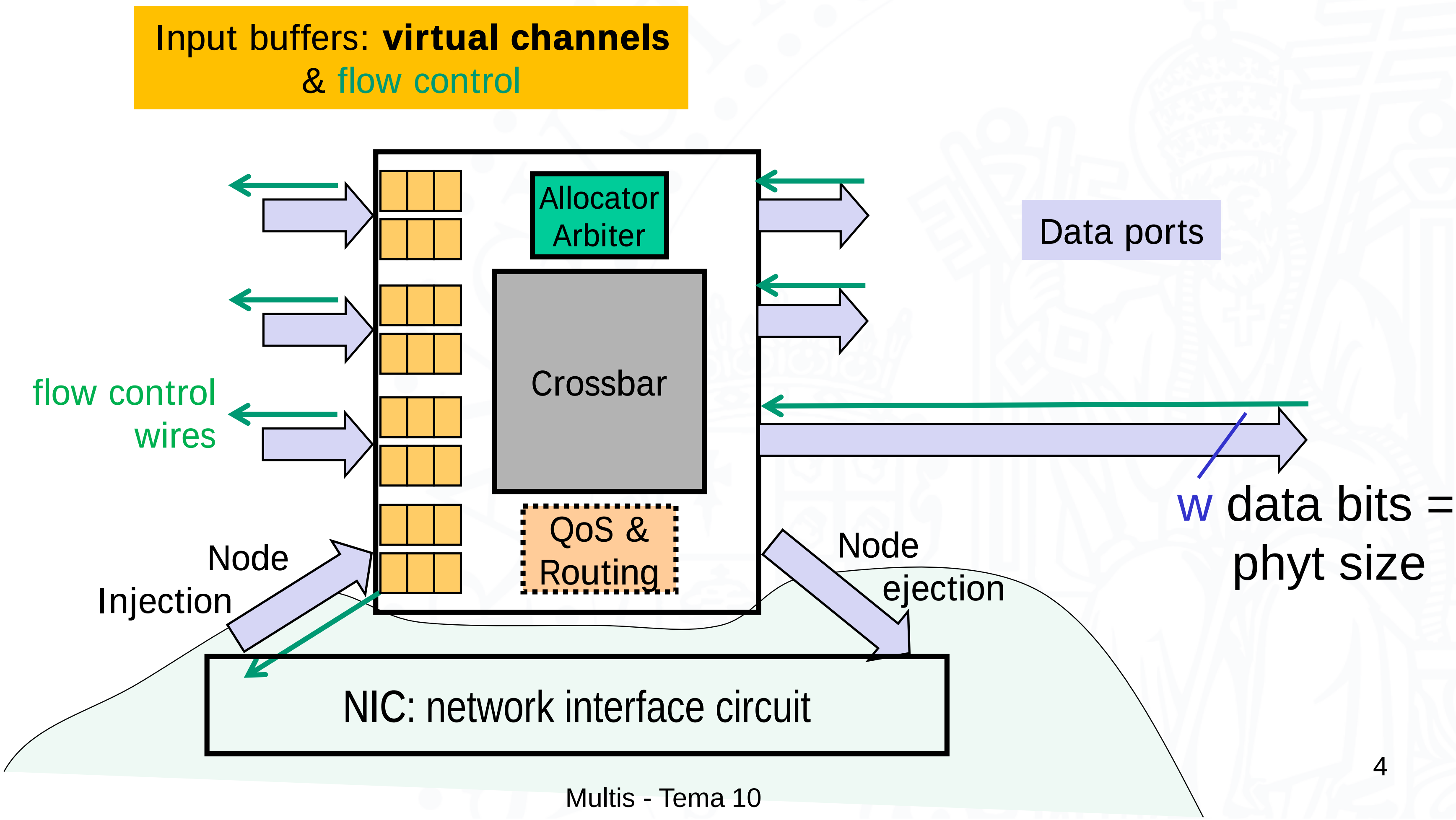


- Direct, static networks
  - built out of point-to-point links between nodes: cpu+ mem+ switch
  - examples:
    - completely or star- connected, linear array, ring, mesh, hypercube, …

Both are used in *SIMD machines*, *shared memory multiprocessors* and *multicomputers*

Network interface and router

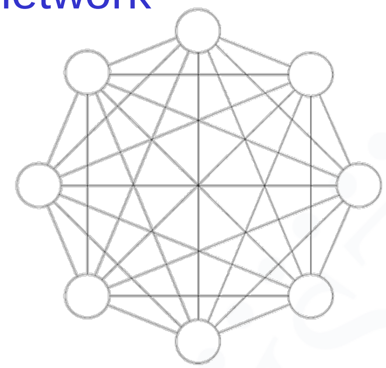
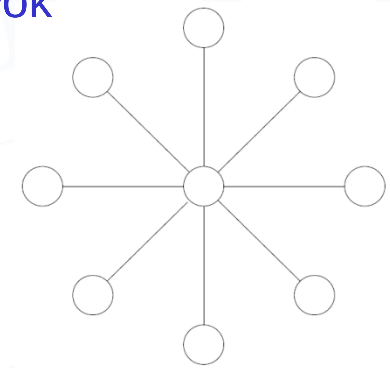
- Router: receives and forwards packets
- Buffers have dual function → synchronization & queueing



Examples of direct interconnection networks

- Completely connected
- Star connected
- Linear Array and Ring
- Mesh and Torus
- Tree and Fat Tree (indirect or direct)
- Hypercube
- …

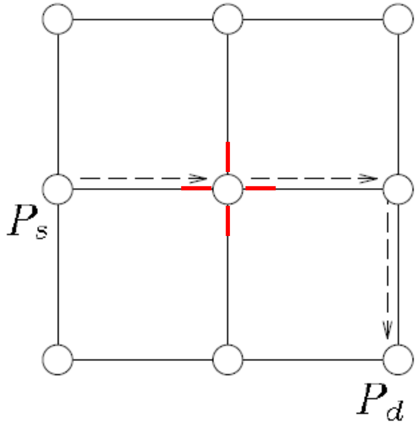
Four extreme topologies

Completely connected network • N nodes • N(N-1)/2 edges  • static analogous of **crossbar**	Star connected netwok  • static analogous of **bus**
Linear array 	Ring 

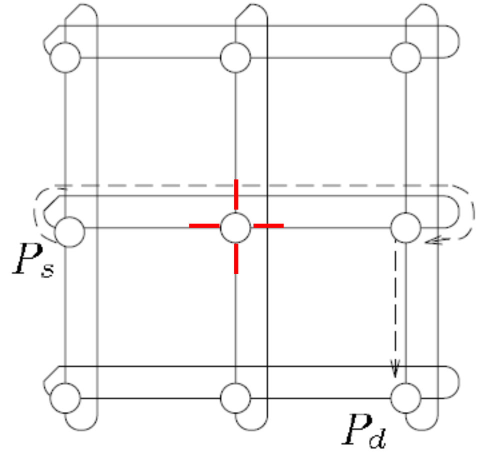
- Some graph-related *quality indices* of a network:
  (graph = set of *vertices* or nodes, and *edges* or links)
  - Network degree: - ∀ node, max # of links
  - Distance betw. nodes A-B: - min # links A↔B
  - Network diameter: - ∀ node pairs, max distance
  - Network average distance: - arithmetic mean of all distances

Mesh and Torus

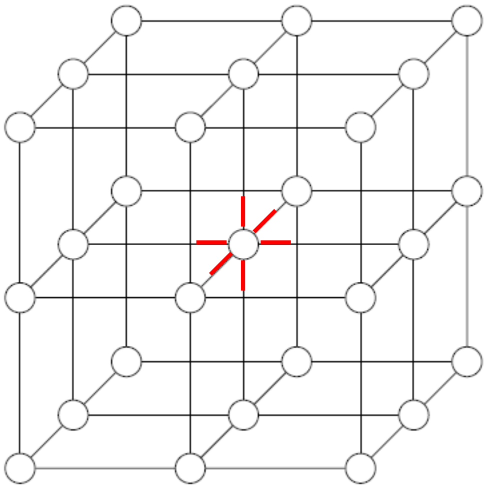
- A *n-dimensional mesh* is an extension of the linear array [*torus* ← *ring*]



2D mesh



Torus: 2D mesh with wraparound



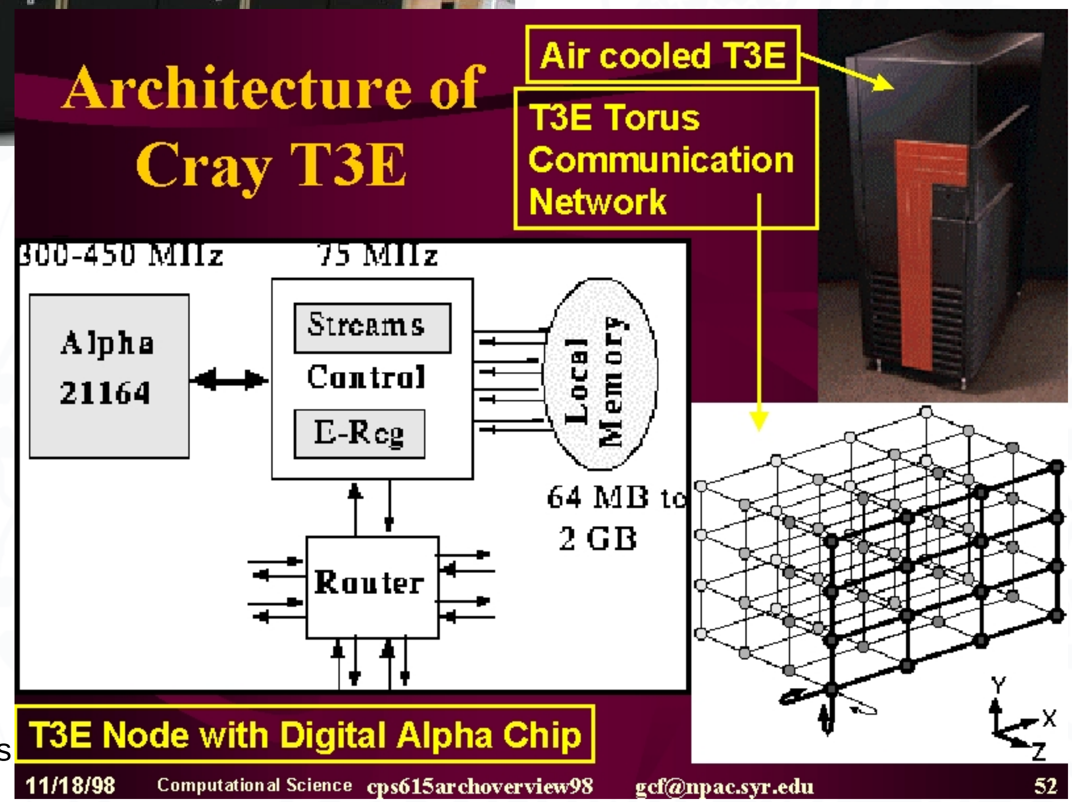
3D mesh

- Ps → Pd: Dimension Order Routing, DOR (deterministic)
- Some old message-passing computers:
  - 1993, Intel Paragon, 2D mesh, 32-2048 nodes, Intel i860 RISC,
  - 1993, Cray T3D, 3D torus, 16-1024 nodes, 2 DEC Alpha 21064@150Mhz / node
  - 1995, **Cray T3E**, 3D torus, 4-1038 nodes, 2 DEC Alpha 21164@300Mhz / node
    1st to achieve 1 TFLOPS in 1998 with 1480 cpus, simulating “metallic magnetism”

Cray T3E

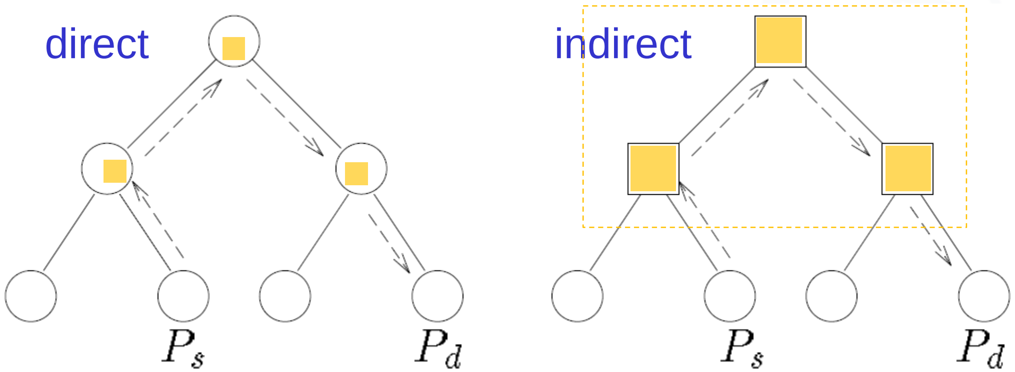
On the left: motor generator providing 400Hz 3~ power for Control Data Cyber machines.
Front: three Control data disk drives. Not connected anymore, they're just there for demonstration and exhibition purpose.

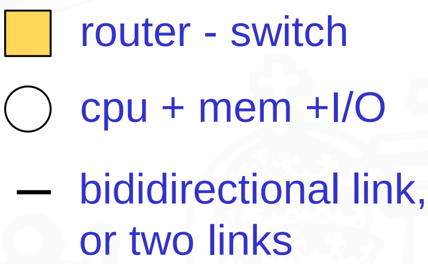
Back: **one row of T3E**, the disk cabinets for the T3E (two black frames) and a Cray YMP-EL (rightmost). By Stiefkind http://www.flickriver.com/photos/stiefkind/4018953017/



Tree networks

- Simple, binary trees

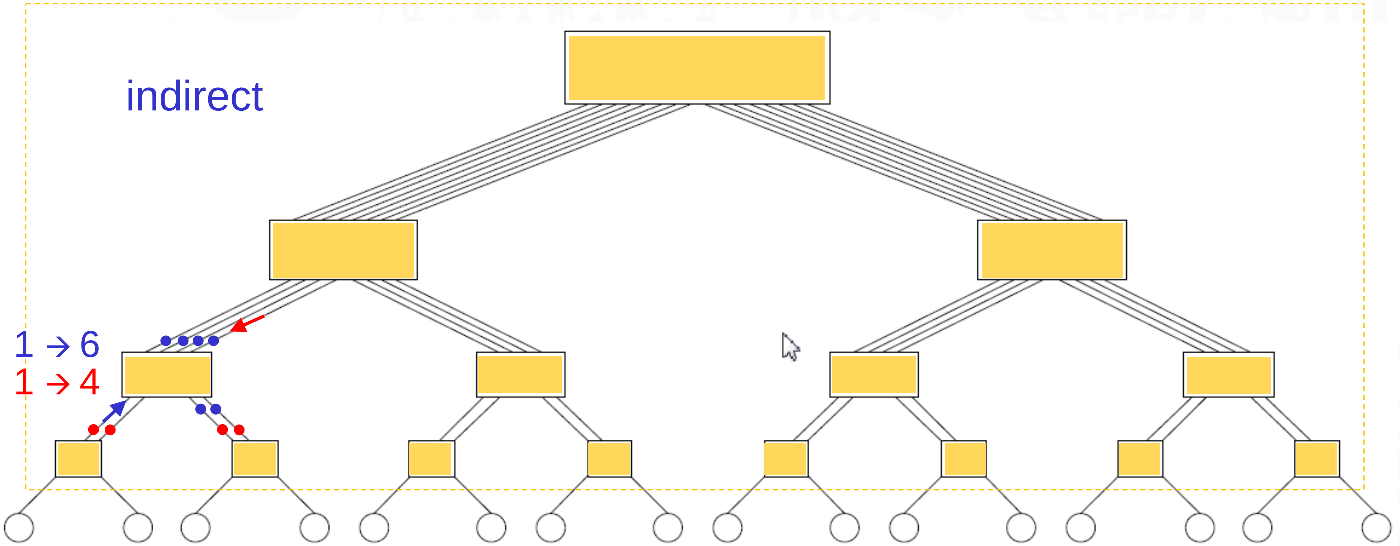




  - Note that the direct version offers a set of different distances to every node

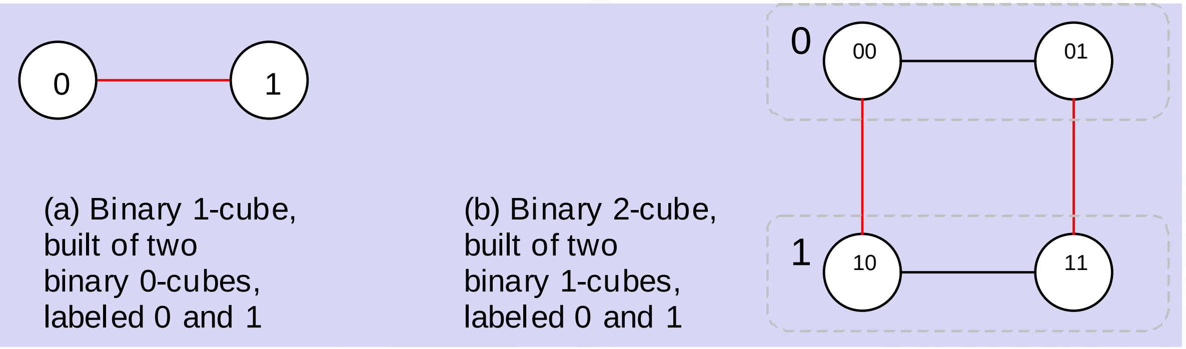
- 4-level fat tree
  - x2 BW as we go up the tree

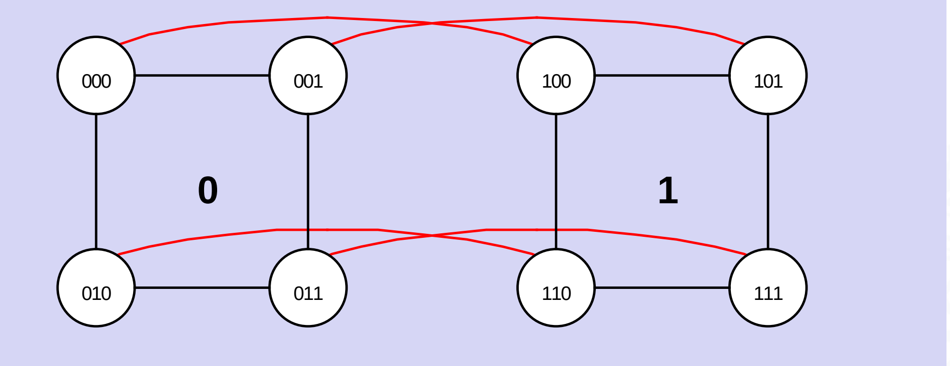
dist.	# nodes
1	-
2	1
3	-
4	2
6	4
8	8



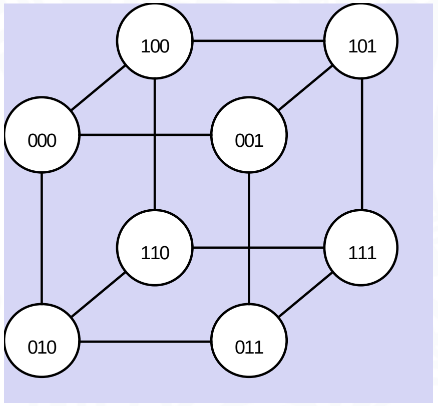
Binary hypercube

- Multi-dimensional mesh that doubles the number of nodes each time a dimension is added





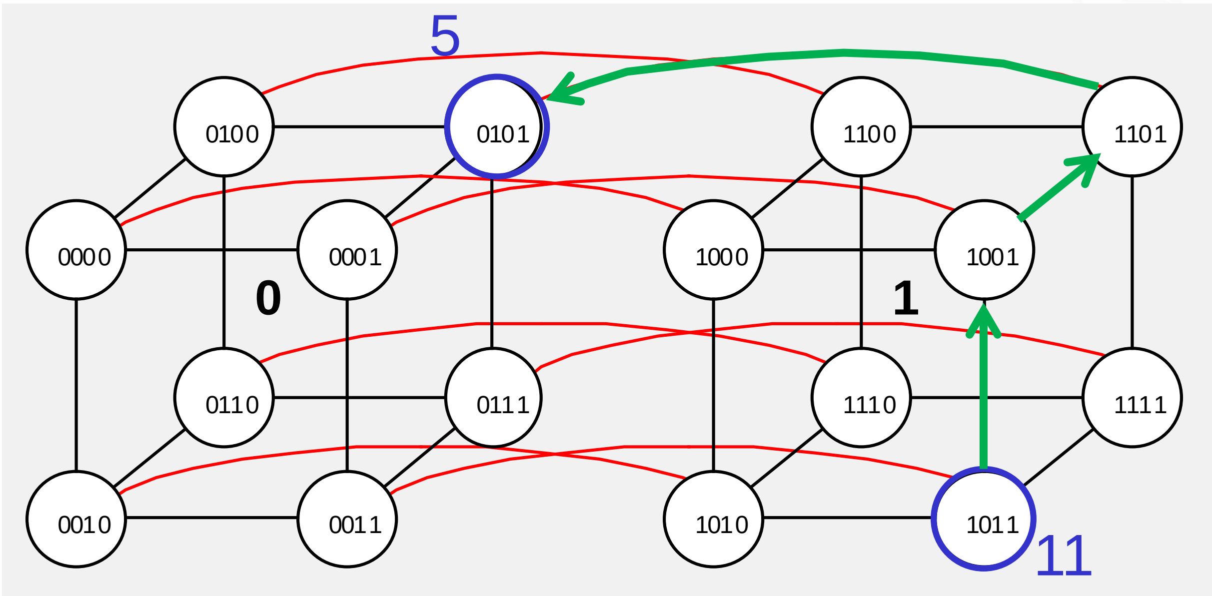
(c) Binary 3-cube, built of two binary 2-cubes, labeled 0 and 1



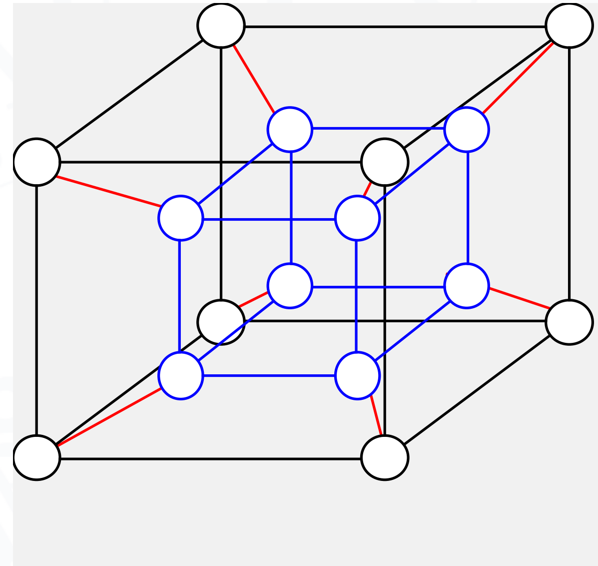
- Some old message-passing computers:
  - Caltech Cosmic Cube (Seitz & Fox 1981), nCube 10 (1024 nodes, 1985), Intel iPSC (32-128 nodes, 1985), Thinking Machines CM-1 (Danny Hillis-MIT, 65536 1-bit processors, 1985)
  - SGI Origin 2000 (2-512 nodes, shared memory, 1996)

Binary hypercube

- 4D hypercube



(d) Binary 4-cube, built of two binary 3-cubes, labeled 0 and 1



	t	z	y	x
11 =	1	0	1	1
	1	0	0	1
	1	1	0	1
5 =	0	1	0	1

- node 11 → node 5 ?
  Routing DOR: t z y x **1ºx, 2ºy, …** (max 4 hops)
  y la vuelta ?

Binary *d*-hypercube properties

- One node connected to *d* others (*d* = network degree)
- One bit difference in labels $\rightarrow$ direct link
- One hyper can be partitioned in two (*d*-*1*) hypers
- The Hamming distance = shortest path length
  - Hamming distance = # of bits different in *source* and *dest*

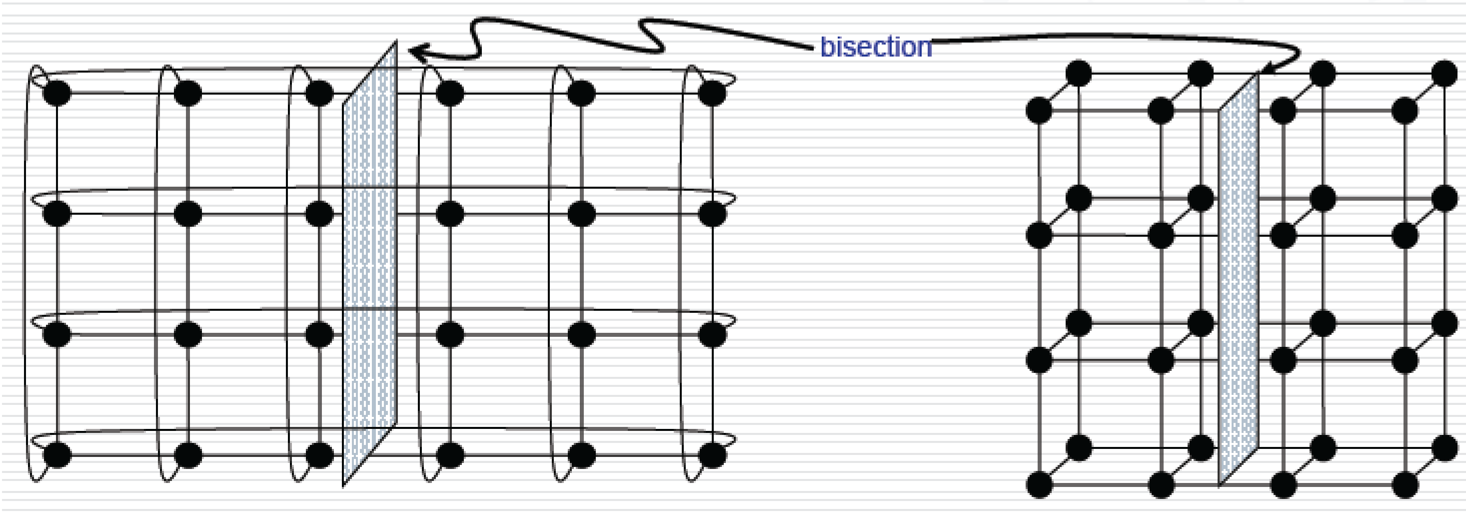
    = # of ones in (*source* $\oplus$ *dest*)

    = Population Count (*source* $\oplus$ *dest*)

Some metrics for direct networks

- Network degree: - ∀ nodes, max # of links
- Distance betw. nodes A-B: - min # links A↔B
- Network diameter: - ∀ node pairs, max distance
- Network average distance: - arithmetic mean of all distances
- Network Connectivity is a way of measuring the multiplicity of paths between any two nodes
  - arc connectivity - min number of links to remove for breaking the network in two
    - 1 for linear arrays, star and tree networks, 2 for rings and 2D meshes, 4 for 2D wraparound meshes
- Network Bisection Width - min number of links to remove to partition the network in two equal halves
  - $p^2/4$ for a compl. connected network, 2 for rings, $\sqrt{p}$ for 2D mesh, $2\sqrt{p}$ for 2D wraparound mesh, $p/2$ for a hypercube
- Network Bisection Bandwidth - min bandwidth between the two halves resulting from bisection
  - equals to Bisection Width times link bandwidth

Bisection



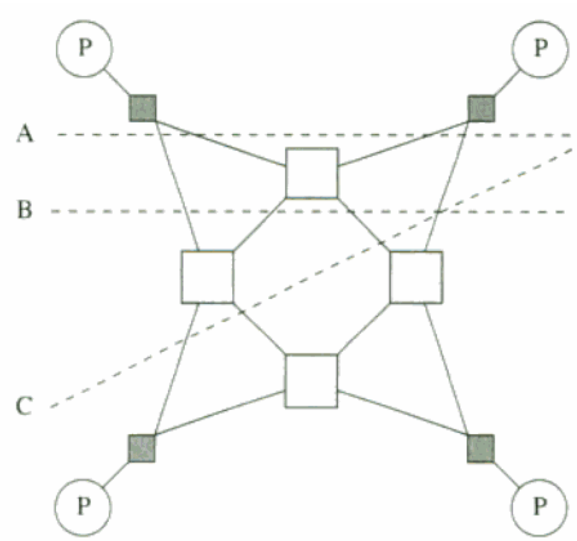


Figure 2.20 Bisection width of a dynamic network is computed by examining various equipartitions of the processing nodes and selecting the minimum number of edges crossing the partition. In this case, each partition yields an edge cut of four. Therefore, the bisection width of this graph is four.

Interconnect comparison

Network	Diameter	Bisection Width	Arc Connectivity	Cost (No. of links)
Completely-connected	1	$p^2/4$	$p-1$	$p(p-1)/2$
Star	2	1	1	$p-1$
Complete binary tree	$2\log((p+1)/2)$	1	1	$p-1$
Linear array	$p-1$	1	1	$p-1$
Ring	$\lfloor p/2 \rfloor$	2	2	p
2-D mesh without wraparound	$2(\sqrt{p}-1)$	$\sqrt{p}$	2	$2(p-\sqrt{p})$
2-D wraparound mesh	$2\lfloor \sqrt{p}/2 \rfloor$	$2\sqrt{p}$	4	$2p$
Hypercube	$\log p$	$p/2$	$\log p$	$(p\log p)/2$
Wraparound k-ary d-cube	$d\lfloor k/2 \rfloor$	$2k^{d-1}$	$2d$	dp

Referencias

- W. DALLY and B. TOWLES, Principles and Practices of Interconnection Networks, Morgan Kaufmann, 2004.
- Highly Parallel Computing. The Benjamin/Cummings Series in Computer Science and Engineering. George S. Almasi & Allan Gottlieb, 1993.
- Interconnection Networks: An Engineering Approach, J. Duato, S. Yalamanchili and L. Ni, Morgan Kaufmann (pubs.), 2003
- Interconnection Networks Computer Architecture: A Quantitative Approach, 4th Edition, Appendix E. T. Pinkston & J. Duato (*…with major presentation contribution from José Flich, UPV)* http://ceng.usc.edu/smart/slides/appendixE.html