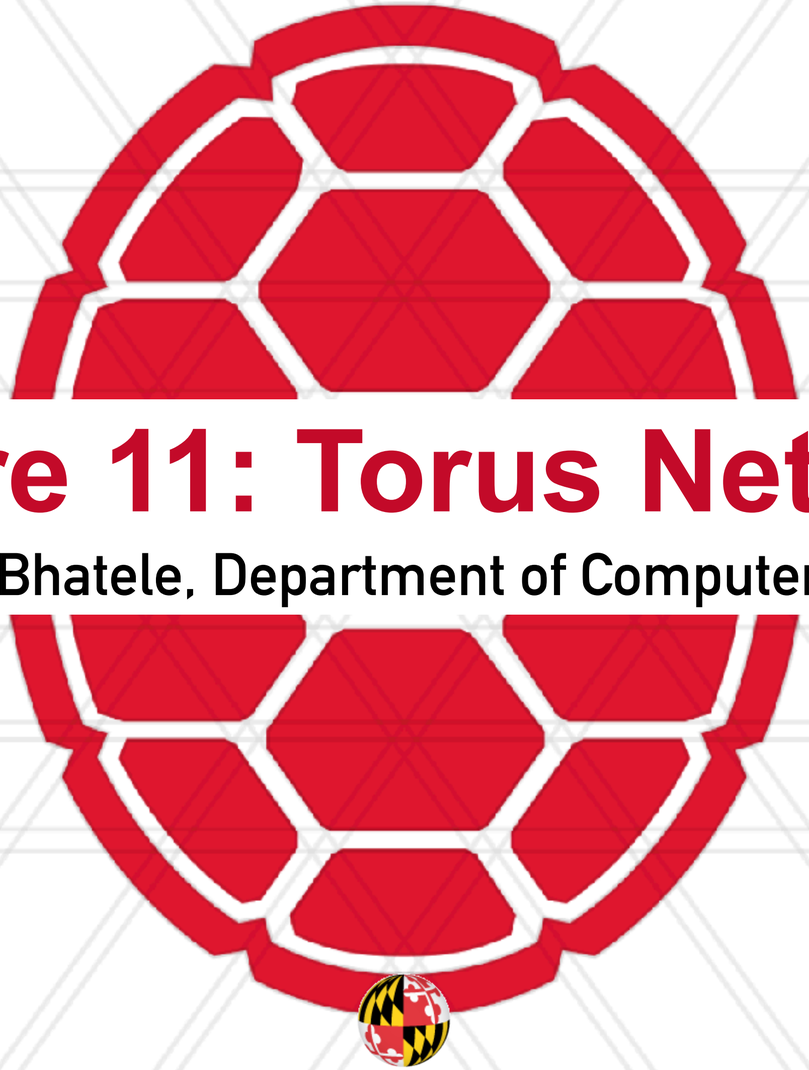
High Performance Computing Systems (CMSC714)



Lecture 11: Torus Networks

Abhinav Bhatele, Department of Computer Science

UNIVERSITY OF MARYLAND

Summary of last lecture

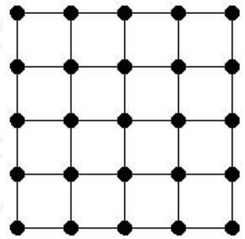
- Shared memory architectures
  - Distributed globally addressable memory
- SGI Altix series
- Directory-based protocol for cache coherence
  - Used hypercube and fat-tree networks
- PGAS languages: Global Arrays is an example

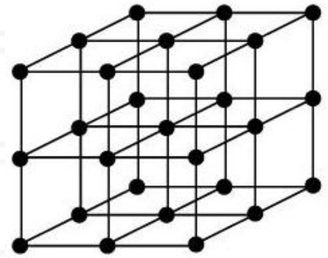
High-speed interconnection networks

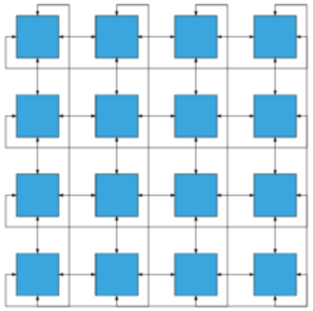
- Key requirements: extremely low latency, high bandwidth
- Scalable: Adding more nodes shouldn't degrade network properties dramatically
  - Low network diameter, high bisection bandwidth
- Compute nodes connected together in many different logical topologies
  - Fat-tree: Charles Leiserson in 1985
  - Mesh and torus networks
  - Dragonfly networks

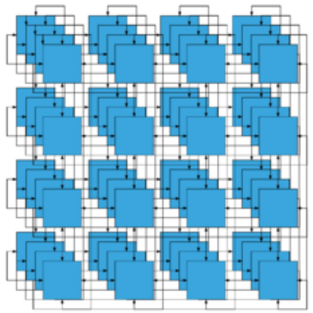
n-dimensional torus networks

- Derived from *k*-ary *n*-cube networks
  - k = number of nodes in each dimension, n = number of dimensions
- 2-dimensional mesh: k-ary 2-cube
- 3-dimensional mesh: k-ary 3-cube
- Torus networks: add wraparound links to the corresponding mesh network









https://en.wikipedia.org/wiki/Torus_interconnect

Routing protocols

- Minimal hop / shortest-path routing
- Static (dimension-ordered) or dynamic (follow path of least congestion)
- Switching techniques
  - Virtual cut-through, wormhole

Switching techniques: http://pages.cs.wisc.edu/~tvrdik/7/html/Section7.html#AAAAABasic%20switching%20techniques

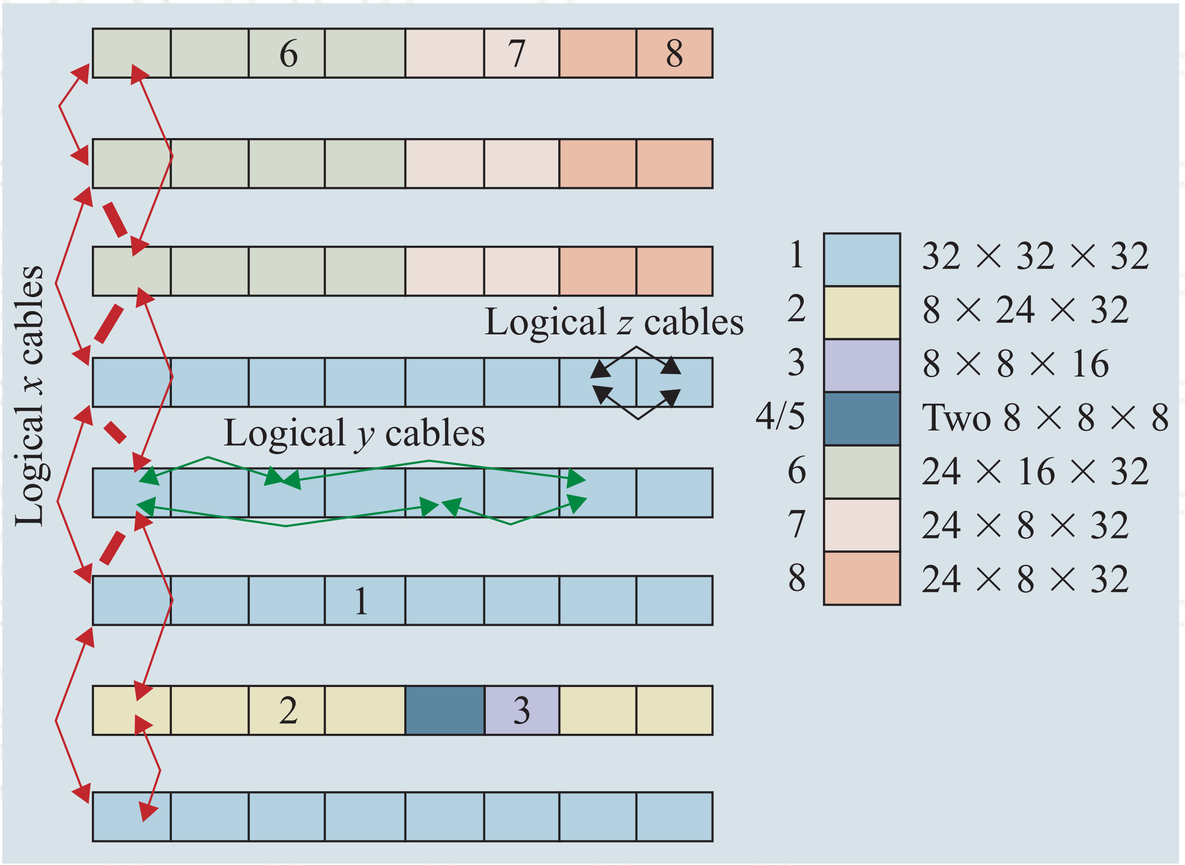
History of torus computers

- Cray T3D was launched in 1993
  - 300 MB/s of bandwidth in each direction
- Cray T3E, XT3/4/5 (SeaStar), XE6/XK7 (Gemini) - 3D tori
- IBM Blue Gene/L/P (3D torus)
- IBM Blue Gene/Q (5D torus with E dimension of size 2)
- Fujitsu Tofu interconnect (6D torus)

History: https://www.extremetech.com/extreme/125271-the-history-of-supercomputers

Blue Gene/L: Five networks

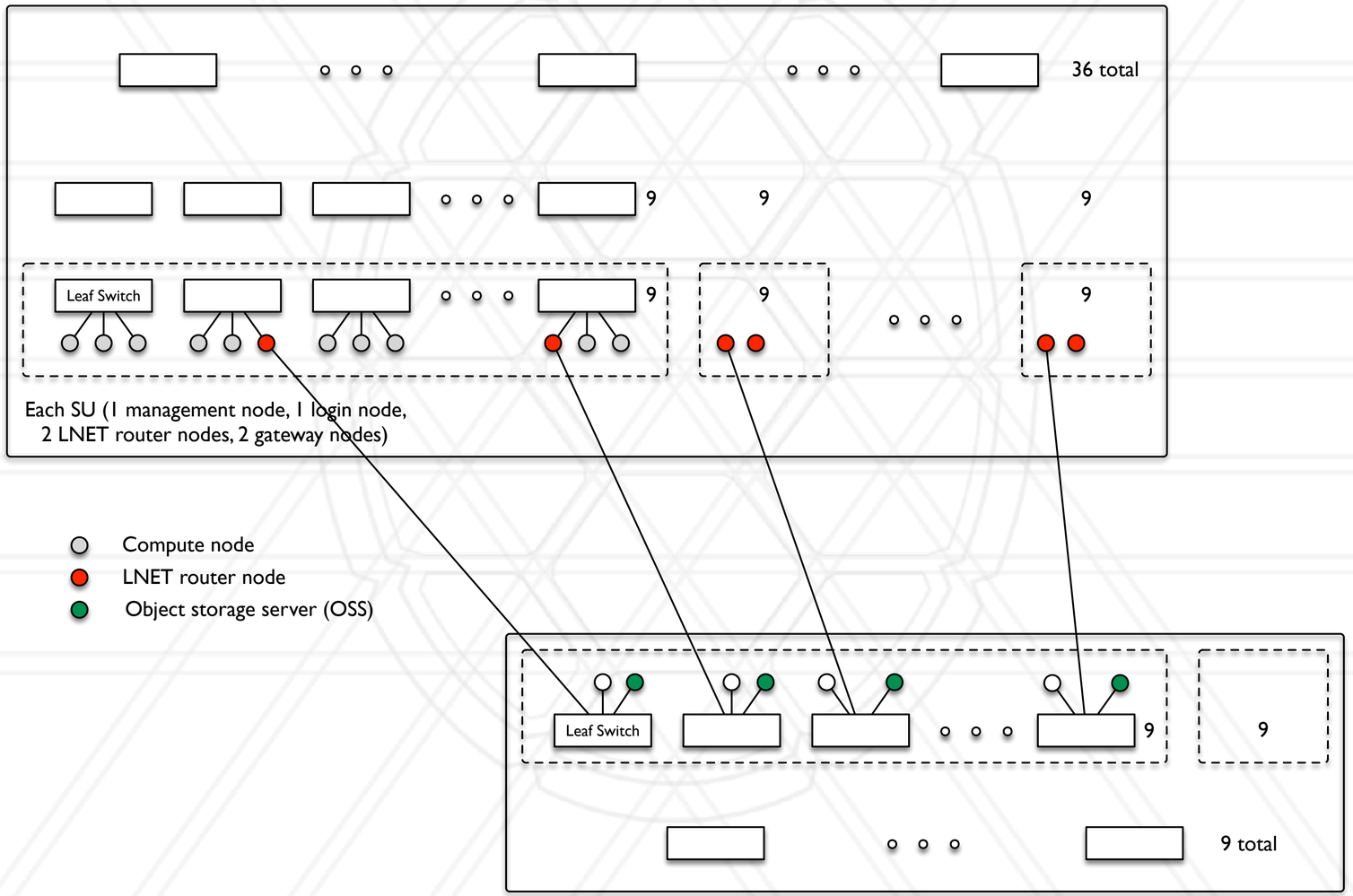
- 3-dimensional torus: 64 x 32 x 32 = 65,536 nodes
  - Build block: 1 mid plane of 8 x 8 x 8 nodes
- Collective network
  - Integer reductions, broadcast
- Barrier network
- Gigabit Ethernet
  - Parallel I/O
- Control system network (Ethernet)



Blue Gene/Q's 5D torus network

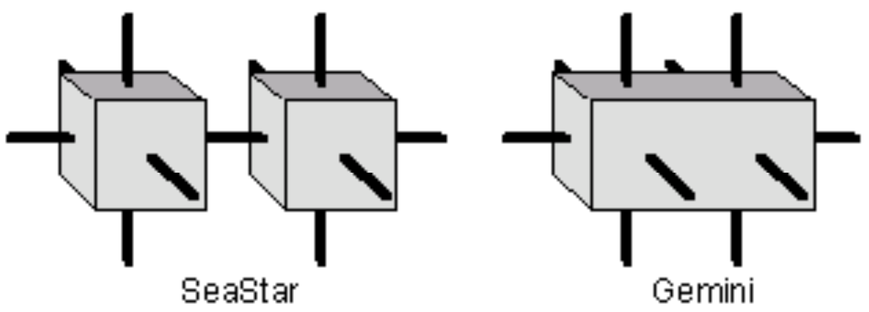
- What happens when we move from 3D to 5D?
- Network bandwidth: 2 GB/s vs. 175 MB/s
- No separate collective and barrier networks
- Still had a separate port for connecting to I/O nodes

Links between cluster and filesystem



Cray Gemini network

- Each Gemini router has 2 nodes attached to it
- 2 pairs of links in the X and Z dimensions, one in the Y dimension (10 total)
- Router has 48 ports
  - 8 are dedicated to the NICs
  - Remaining 40 are used in groups of 4
- Link bandwidths: 4.68 to 9.375 GB/s



Questions?

UNIVERSITY OF
MARYLAND

Abhinav Bhatele

5218 Brendan Iribe Center (IRB) / College Park, MD 20742

phone: 301.405.4507 / e-mail: bhatele@cs.umd.edu