Advanced Computer Architecture
ELEC3219 (2017/18)

Networks-on-Chip
(Answers)

Dr. Terrence Mak
tmak@ecs.soton.ac.uk
Torus (1) (Direct? Or Indirect?)

• K-ary n-cube: $k^n$ network nodes
• n-dimensional grid with k nodes in each dimension
Torus (2)

- Topologies in Torus Family
  - Ring k-ary 1-cube
  - Hypercubes 2-ary n-cube
- Edge Symmetric
  - Good for load balancing
  - Removing wrap-around links for mesh loses edge symmetry
    - More traffic concentrated on center channels
- Good path diversity
- Exploit locality for near-neighbor traffic
Channel Load for Torus

- Even number of k-ary (n-1)-cubes in outer dimension
- Dividing these k-ary (n-1)-cubes gives 2 sets of $k^{n-1}$ bidirectional channels or $4k^{n-1}$
- 1/2 (half) Traffic from each node cross bisection

$$\text{channel load} = \frac{N}{2} \cdot \frac{k}{4N} = \frac{k}{8}$$

- Mesh has 1/2 of the bisection bandwidth of torus
Butterfly (Direct? Or Indirect?)

• K-ary n-fly: $k^n$ network nodes
• Example: 2-ary 3-fly
• Routing from 000 to 010
  – Dest address used to directly route packet
  – Bit $n$ used to select output port at stage $n$
Butterfly (2)

• No path diversity \( |R_{xy}| = 1 \)
• Hop Count
  – \( \log_k n + 1 \)
  – Does not exploit locality
    • Hop count same regardless of location
• Switch Degree = 2k
• Channel Load \( \rightarrow \) uniform traffic

\[ \frac{NH_{\min}}{C} = \frac{k^n(n+1)}{k^n(n+1)} = 1 \]
  – Increases for adversarial traffic
Flattened Butterfly

• Proposed by Kim et al (ISCA 2007)
  – Adapted for on-chip (MICRO 2007)

• Advantages
  – Max distance between nodes = 2 hops
  – Lower latency and improved throughput compared to mesh

• Disadvantages
  – Requires higher port count on switches (than mesh, torus)
  – Long global wires
  – Need non-minimal routing to balance load
Flattened Butterfly

• Path diversity through non-minimal routes
Clos Network (Direct? Or Indirect?)

University of Southampton
Clos Network

• 3-stage indirect network
• Characterized by triple \( (m, n, r) \)
  – M: # of middle stage switches
  – N: # of input/output ports on input/output switches
  – R: # of input/output switching
• Hop Count = 4
Folded Clos (Fat Tree) (Direct? Or Indirect?)

- Bandwidth remains constant at each level
- Regular Tree: Bandwidth decreases closer to root
Fat Tree (2)

• Provides path diversity
Common On-Chip Topologies

• Torus family: mesh, concentrated mesh, ring
  – Extending to 3D stacked architectures
  – Favored for low port count switches
• Butterfly family: Flattened butterfly
Topology Summary

• First network design decision
• Critical impact on network latency and throughput
  – Hop count provides first order approximation of message latency
  – Bottleneck channels determine saturation throughput
• Question?