ECE 4100/6100
Interconnection Networks
Reading: Chapter 8
Interconnection Networks

- Direct vs. indirect networks
  - direct networks
    - fixed degree
    - modular
  - indirect networks
    - uniform base latency
    - centralized or distributed control
- Engineering approximations to direct networks
Network Topologies

- Indirect Networks
  - Multistage networks
  - Fat Trees
- Direct networks
  - Tori
  - Hypercubes
- Evaluation Metrics
  - Bisection bandwidth
  - Node degree
  - latency
The Message Path

- Wire bandwidth is not the bottleneck!
- Operating system and/or user level software limits performance
Messaging Functions

- The intra-processor bus limits latency and overall message bandwidth
**Messaging Hierarchy**

- **Routing Layer**: *Where?:* Destination decisions, i.e., which output port
- **Switching Layer**: *When?:* When is data forwarded
- **Physical Layer**: *How?:* Synchronization of data transfer

- This organization is distinct from traditional networking implementations
- Emphasis is on low latency communication
  - Only recently has standards been evolving
• Data is transmitted based on a hierarchical data structuring mechanism
  - Messages $\rightarrow$ packets $\rightarrow$ flits $\rightarrow$ phits
Flow Control

- Flow control: synchronized transfer of a **unit** of information
- Buffering requirements
- Asynchronous vs. synchronous flow control
- Flow control occurs at multiple levels
  - message flow control
  - physical flow control
- Mechanisms
  - Credit based flow control

![Diagram of Flow Control](attachment://flow_control_diagram.png)
Switching Layer

- Comprised of three sets of techniques
  - switching techniques
  - flow control
  - buffer management
- Organization and operation of routers are largely determined by the switching layer
- Connection Oriented vs. Connectionless communication
Circuit Switching

- Hardware path setup by a routing header or probe
- End-to-end acknowledgment initiates transfer at full hardware bandwidth
- Source routing vs. distributed routing
- System is limited by signaling rate along the circuits --> wave pipelining
Packet Switching

- Blocking delays in circuit switching avoided in packet switched networks --> full link utilization in the presence of data
- Increased storage requirements at the nodes
- Packetization and in-order delivery requirements
- Buffering
  - use of local processor memory
  - central queues
Virtual Cut-Through

- Messages cut-through to the next router when feasible
- In the absence of blocking, messages are pipelined
  - pipeline cycle time is the larger of intra-router and inter-router flow control delays
- When the header is blocked, the complete message is buffered
- High load behavior approaches that of packet switching
Wormhole Switching

- Messages are pipelined, but buffer space is on the order of a few flits
- Small buffers + message pipelining --> small compact buffers
- Supports variable sized messages
- Messages cannot be interleaved over a channel: routing information is only associated with the header
- Base Latency is equivalent to that of virtual cut-through
Virtual Channels

- Each virtual channel is a pair of unidirectional channels
- Independently managed buffers multiplexed over the physical channel
- De-couples buffers from physical channels
- Originally introduced to break cyclic dependencies
- Improves performance through reduction of blocking delay
- Virtual lanes vs. virtual channels
- As the number of virtual channels increase, the increased channel multiplexing has two effects
  - decrease in header delay
  - increase in average data flit delay
- Impact on router performance
  - switch complexity
Pipelined Circuit Switching

- Combines circuit switching and wormhole switching
- Paths are setup using virtual channels: reserved channels do not consume physical channel bandwidth
- Increased flexibility in routing the header flit
- Motivated by reliability concerns
- The use of multiple flit types lead to the use of control channels
- Each virtual channel is augmented with a virtual control channel
- Header flits, acknowledgment flits and backtracking flits use the control channels
- Control channels provide a separate control network for distributed routing and recovery in the presence of faults
Comparison of Switching Techniques

- Packet switching and virtual cut-through
  - consume network bandwidth proportional to network load
  - predictable demands
  - VCT behaves like wormhole at low loads and like packet switching at high loads
  - link level error control for packet switching

- Wormhole switching
  - provides low latency
  - lower saturation point
  - higher variance of message latency than packet or VCT switching

- Virtual channels
  - blocking delay vs. data delay
  - router flow control latency

- Optimistic vs. conservative flow control
Scalable Switching Fabrics for Internet Routers

- Internet bandwidth growth → routers with
  - large numbers of ports
  - high bisection bandwidth
- Historically these solutions have used
  - Backplanes
  - Crossbar switches
Requirements

- **Scalable**
  - Incremental
  - Economical → cost linear in the number of nodes
- **Robust**
  - Fault tolerant → path diversity + reconfiguration
  - Non-blocking features
- **Performance**
  - High bisection bandwidth
  - Quality of Service (QoS)
    - Bounded delay
Switching Fabric

- Three components
  - Topology → 3D torus
  - Routing → source routing with randomization
  - Flow control → virtual channels and virtual networks
- Maximum configuration: $14 \times 8 \times 5 = 560$
- Channel speed is 10 Gbps
Packaging

- Uniformly short wires between adjacent nodes
  - Can be built in passive backplanes
  - Run at high speed
    - Bandwidth inversely proportional to square of wirelength
  - Cabling costs
  - Power costs

Figures are from *Scalable Switching Fabrics for Internet Routers*, by W. J. Dally (can be found at www.avici.com)
Available Bandwidth

- Distinguish between capacity and I/O bandwidth
  - Capacity: Traffic that will load a link to 100%
  - I/O bandwidth: bit rate in or out

- Discontinuities
Properties

- Path diversity
  - Avoids tree saturation
  - Edge disjoint paths for fault tolerance
    - Heart beat checks (100 microsecs) + deflecting while tables are updated

Figures are from Scalable Switching Fabrics for Internet Routers, by W. J. Dally (can be found at www.avici.com)
## Properties

<table>
<thead>
<tr>
<th></th>
<th>Path Diversity</th>
<th>Bandwidth Scaling</th>
<th>Extensibility</th>
<th>Channel Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-D Torus</td>
<td>High</td>
<td>$\sqrt{N}$</td>
<td>1 Node</td>
<td>All Short</td>
</tr>
<tr>
<td>3-D Torus</td>
<td>High</td>
<td>$N^{\frac{1}{2}}$</td>
<td>1 Node</td>
<td>All Short</td>
</tr>
<tr>
<td>Hypercube</td>
<td>High</td>
<td>N</td>
<td>1 Node (limited by degree)</td>
<td>Some Long</td>
</tr>
<tr>
<td>Butterfly</td>
<td>None</td>
<td>N</td>
<td>Doubling</td>
<td>Some Long</td>
</tr>
<tr>
<td>Buses</td>
<td>High</td>
<td>N</td>
<td>Doubling</td>
<td>Some Long</td>
</tr>
<tr>
<td>Crossbar</td>
<td>N/A</td>
<td>N</td>
<td>None</td>
<td>Some Long</td>
</tr>
</tbody>
</table>
Use of Virtual Channels

- Virtual channels aggregated into virtual networks
  - Two networks for each output port

- Distinct networks prevent undesirable coupling
  - Only bandwidth on a link is shared
  - Fair arbitration mechanisms

- Distinct networks enable QoS constraints to be met
  - Separate best effort and constant bit rate traffic