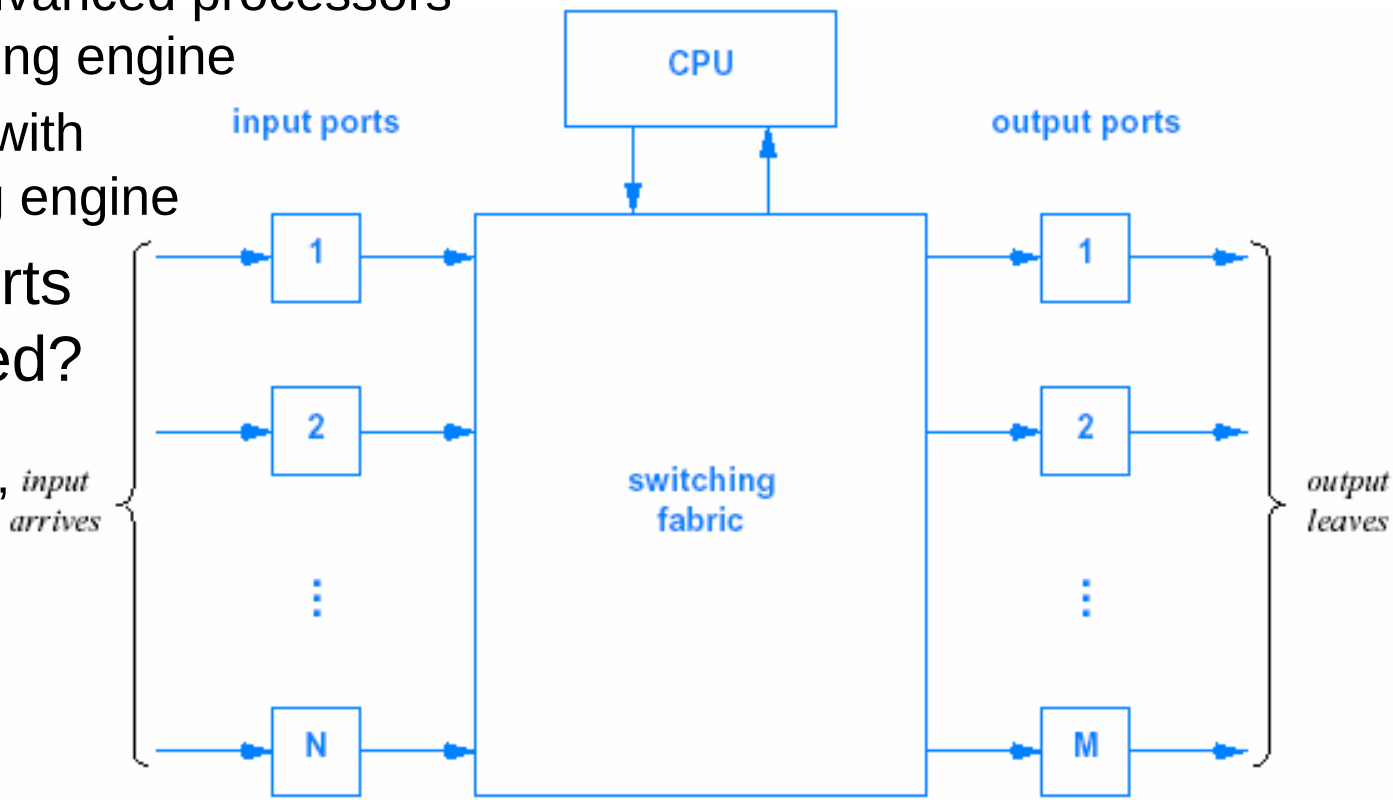
ECE 697J – Advanced Topics in Computer Networks

Switching Fabrics

10/02/03

Router Data Path

- Last class:
  - Single CPU is not fast enough for processing packets
  - Multiple advanced processors in processing engine
  - Each port with processing engine
- How can ports be connected?
  - "System backplane"



Switching Fabric

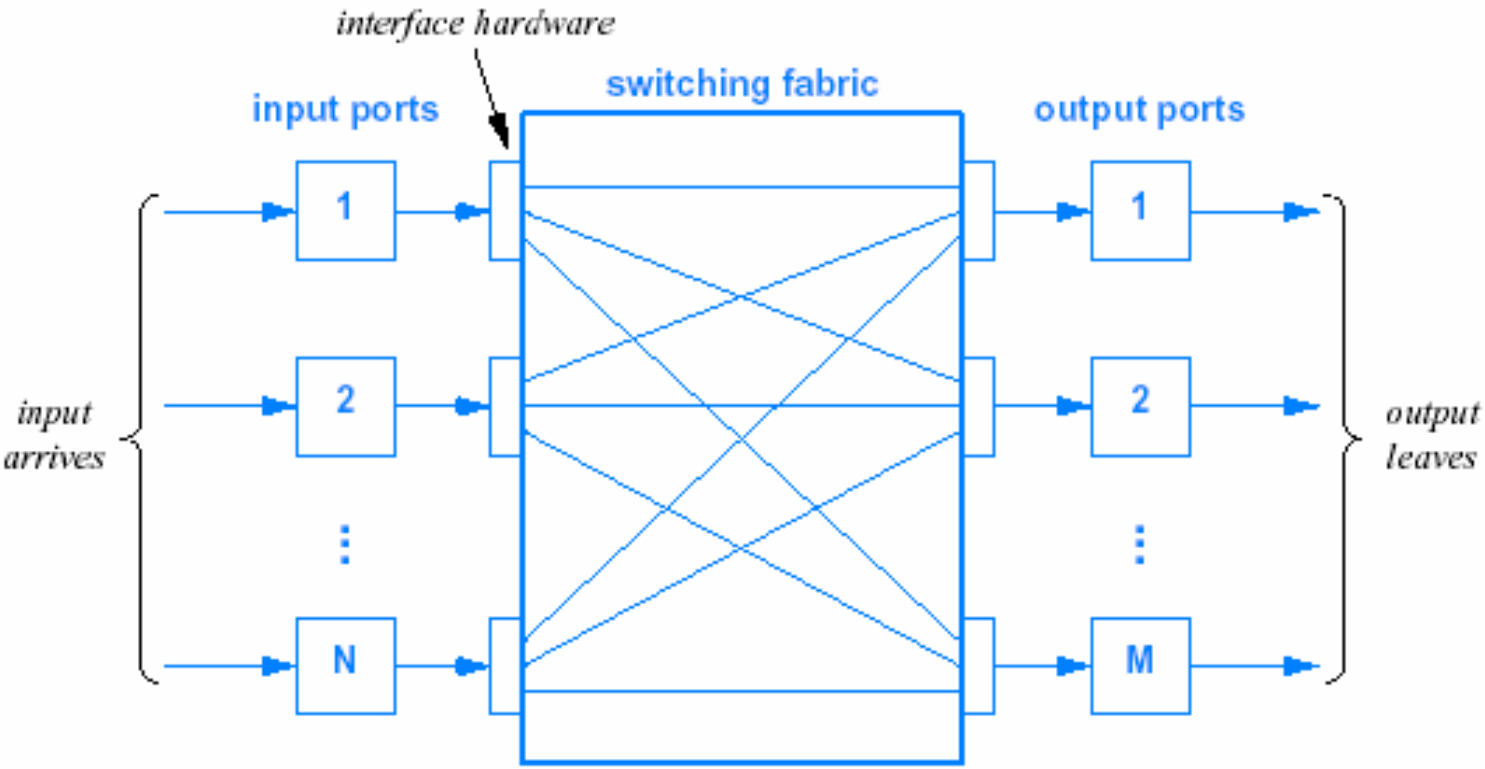
- Requirements:
  - Designed for use inside a single network system
  - Provide interconnection between ports and control processor
  - Support transfer of unicast, multicast, and broadcast packets
  - Scale with data rate on input/output
  - Scale with packet rate on input/output
  - Scale with number of input/output ports
  - Low overhead (e.g., “headers” inside fabric)
  - Low cost
- Not all requirements can be achieved
- Tradeoffs necessary (e.g., cost vs. bandwidth)

Switching Fabric Types

- Synchronous vs. asynchronous
  - Sync: regular traffic (e.g., fixed cells), async: variable size (e.g., packets)
  - In practice synchronous fabrics are used
  - Packets are divided into fixed-sized cells
- Multiplexing:
  - Time Division: single or few paths shared among many ports
  - Space Division: many paths used for less delay
- We'll look at:
  - SDM: fully connected, crossbar
  - TDM: shared bus, shared memory
  - Multistage: Banyan

Fully Interconnected Fabric

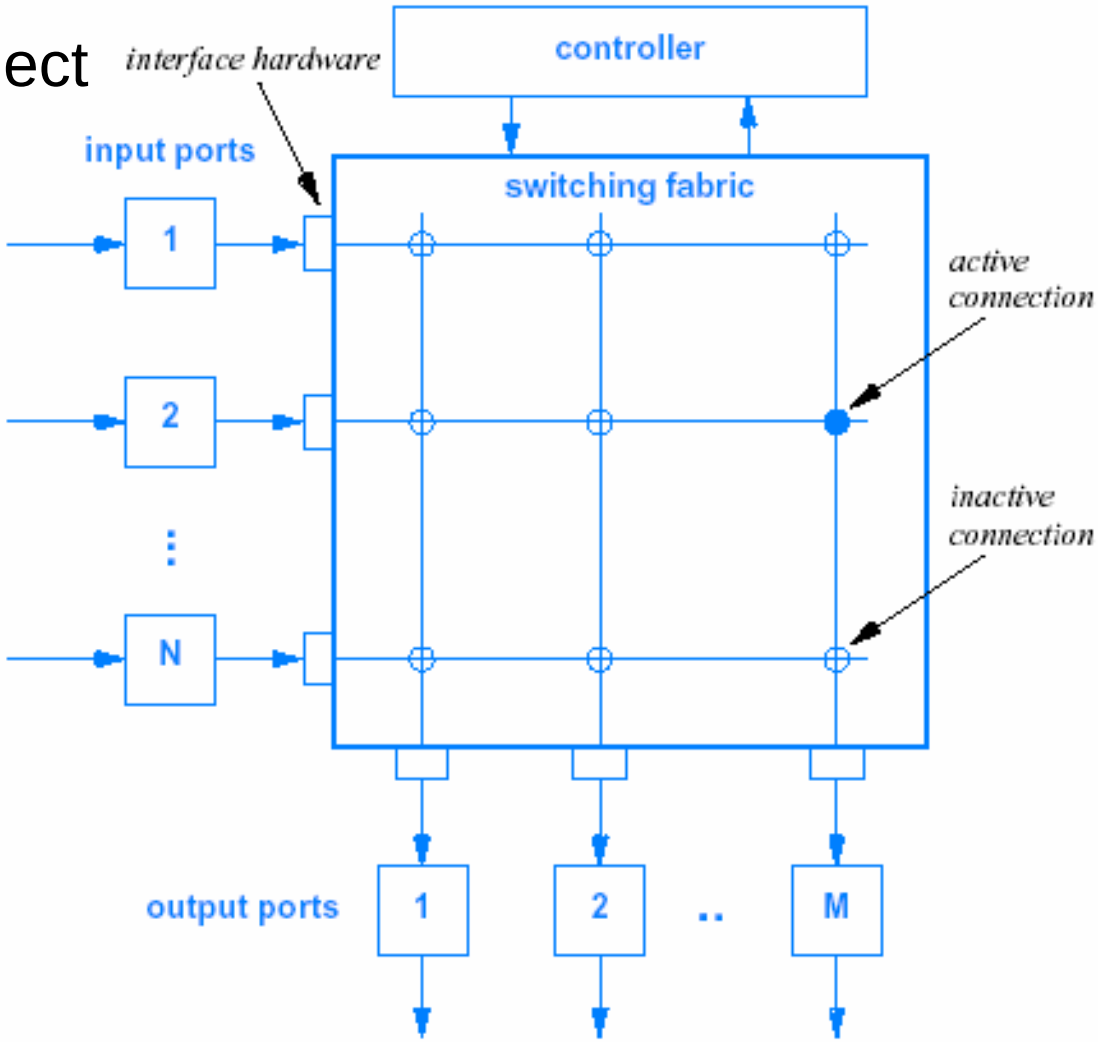
- Separate physical data path between each port pair



- Problem?

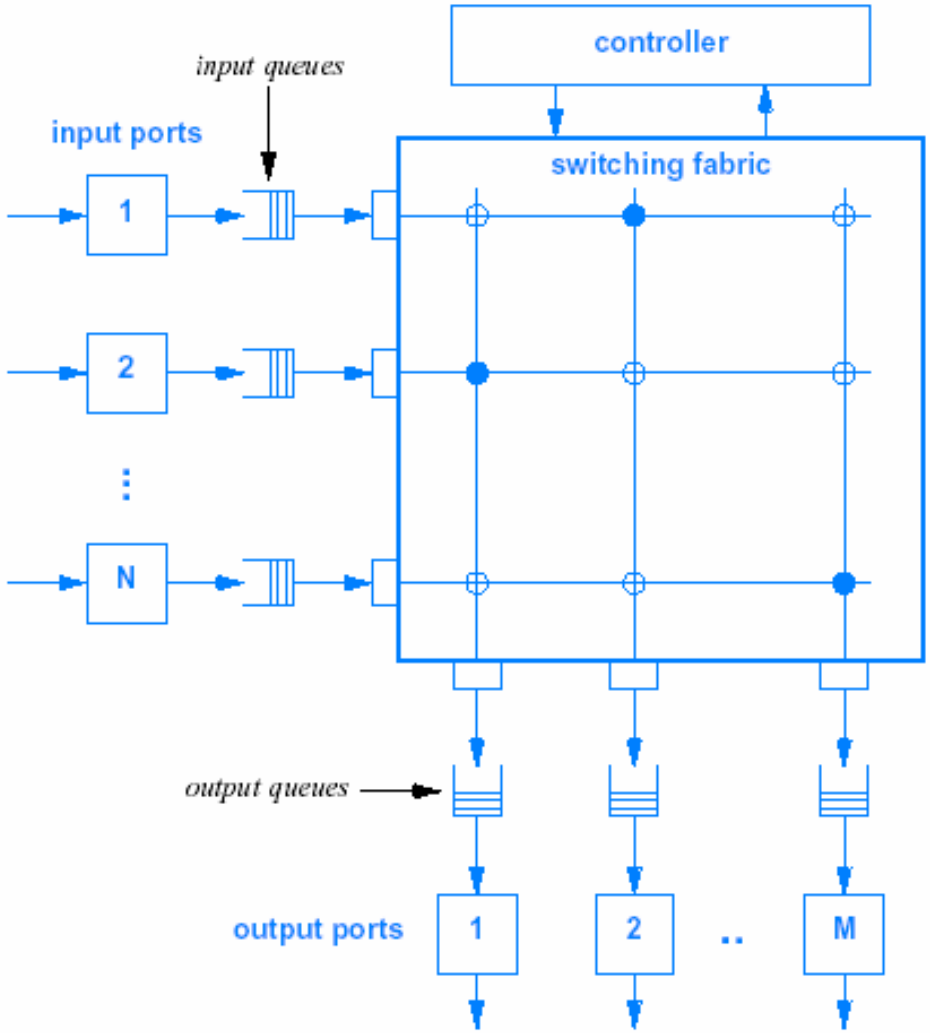
Crossbar Fabric

- Switched interconnect
- Allows multiple simultaneous connections
- Controller decides on assignment of active connections
- Problems?



Contention

- What happens if multiple input ports send to the same output port?
  - Causes “port contention”
  - Is this a problem of the fabric architecture?
- How can contention be addressed?
  - Queues that buffer packets

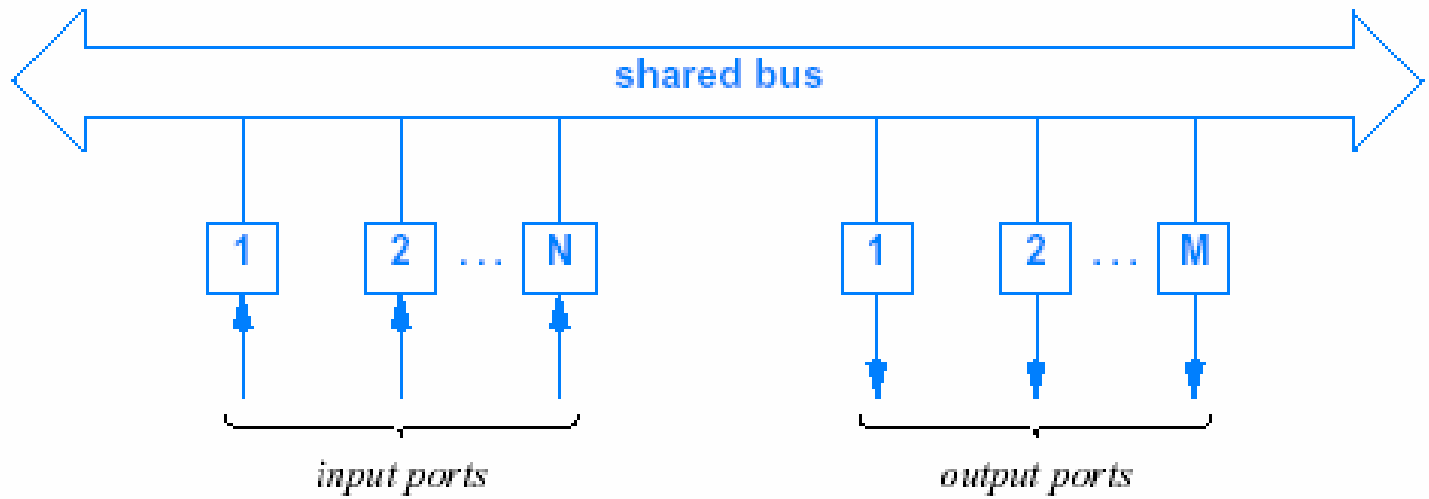


Queuing

- Variety of queuing designs possible
- Input queuing:
  - Packets are buffered on input side until output is available
  - Problem: head of line blocking
- Output queuing:
  - Bursts from inputs to one output are buffered on output side
  - Problem: queue size, switch fabric speedup
- Virtual queuing:
  - Partial output queues are maintained on input size
  - Each port gets rate or virtual time assigned for sending
  - Problem: requires complex coordination

Shared Bus Fabric

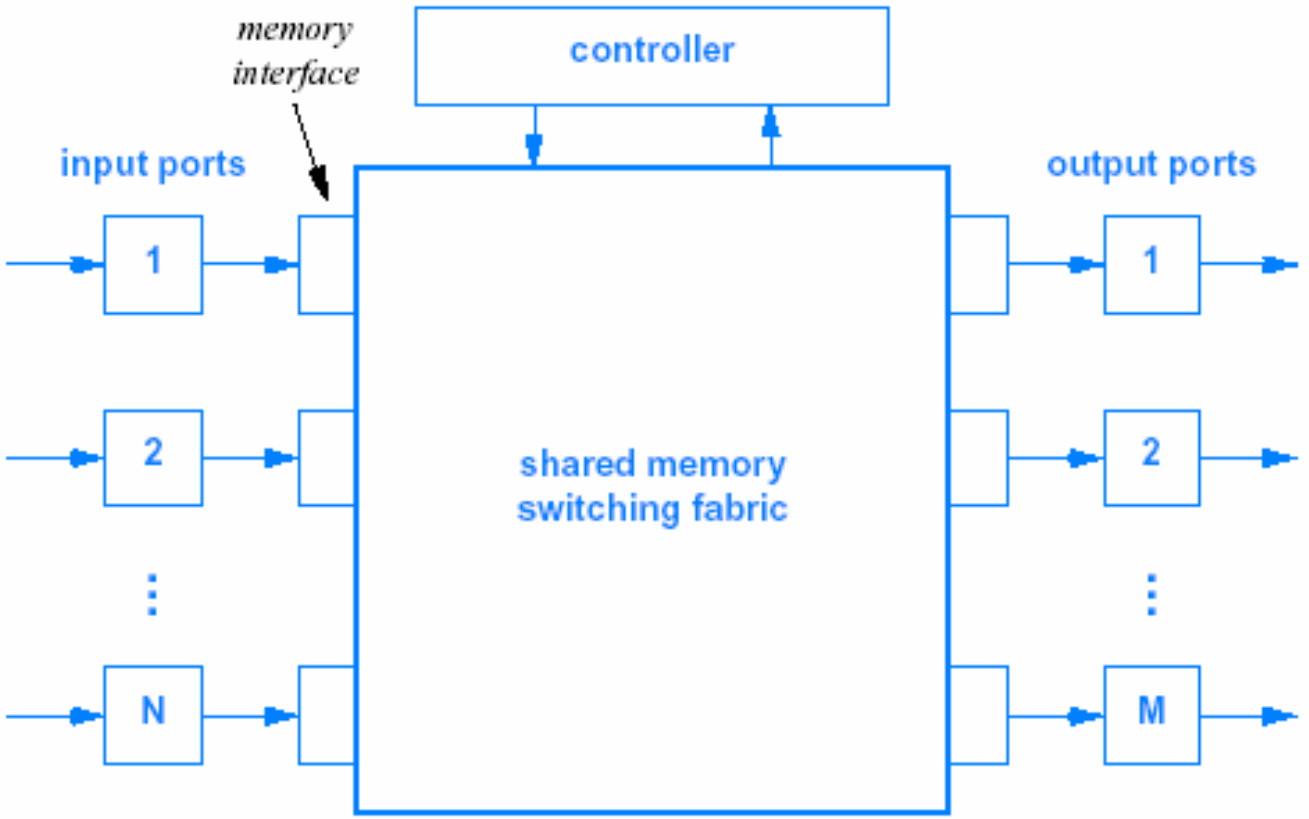
- Full interconnect and crossbar can be costly (N^2)
- Shared bus uses one path for all ports (cost is 2N)



- Disadvantage: lower aggregate data rate
  - Hardware needs to operate at N times faster than ports
- What granularity is best for access: packet, cells?

Shared Memory Fabric

- Input ports deposit packets in memory for output ports
- Problem: cost of memory interfaces

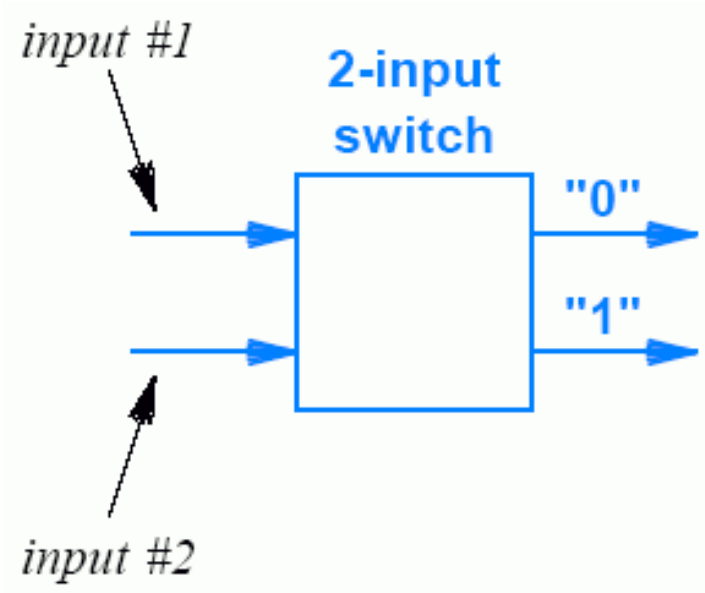


Multistage Fabrics

- None of the fabrics so far really scales
  - Full interconnect: squared cost with port number
  - Crossbar: limitations in controller
  - Bus: interconnect needs to be N times faster than port
  - Shared Memory: memory interfaces are costly and not scalable
- Multistage fabrics:
  - Multiple “steps” between input and output
  - Provides multiple data paths, but not as many as fully connected
  - Packets can be queued or not on each stage
- Nice properties:
  - Scalability
  - Self routing

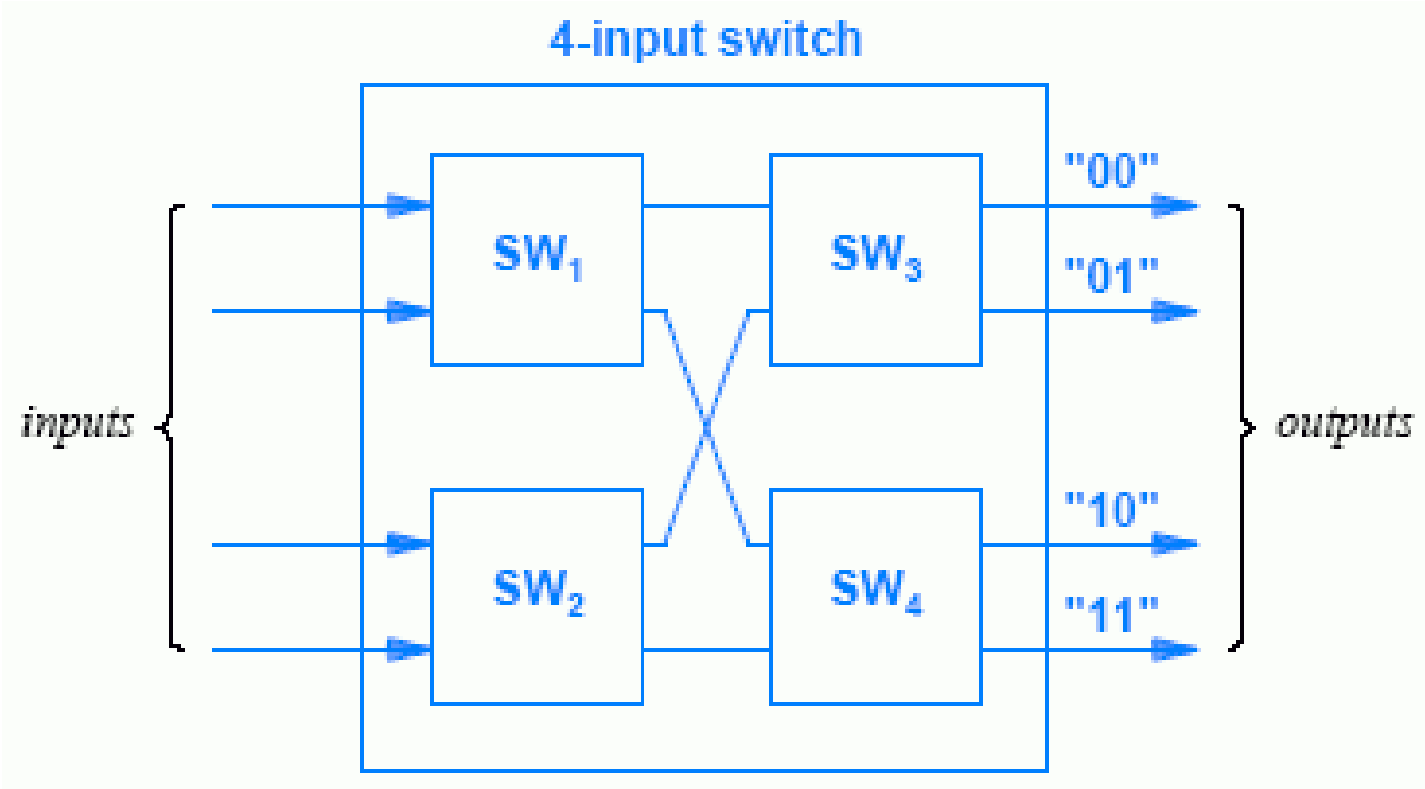
Switching Elements

- Switching element is basic component:
  - Two inputs, two outputs
  - Two different settings:
    - Straight through
    - Crossover
  - Setting is determined by “label”
- Multiple switching elements make up larger fabric
- Various topologies possible:
  - Banyan
  - Benes
  - Delta
  - Etc.



Banyan Switch

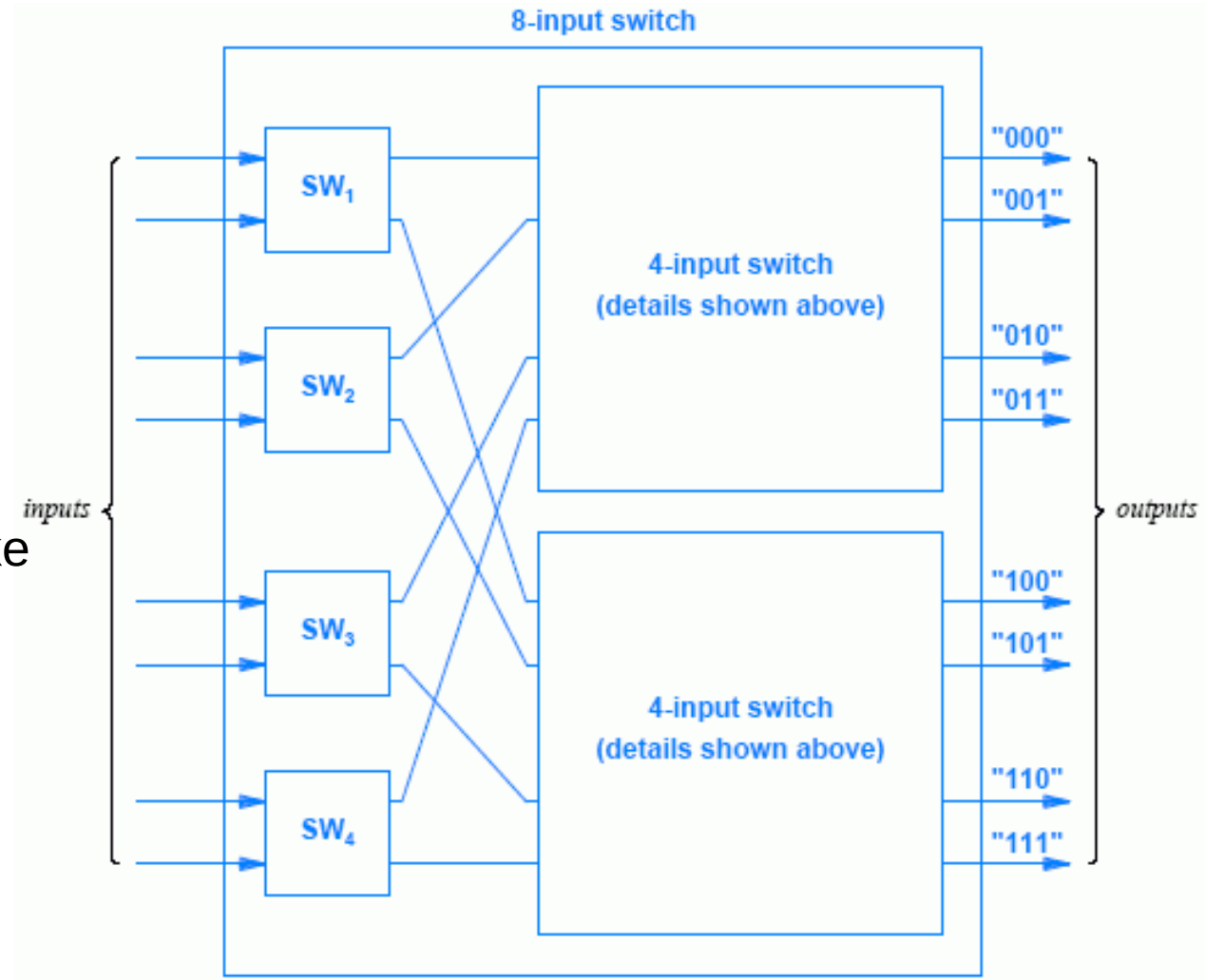
- From 2-port switch to 4-port switch:



- Can be applied recursively

Banyan Switch

- 8-port:
  - Recursive extension
- Self-routing:
  - Based on output port "label" each stage can make local decision
  - Recursive structure will lead to correct destination



Banyan Switch

- Scalability due to recursive structure
  - More ports will require more stages (grows with log n)
- Contention can occur
  - How?
  - Usually due to overloading of one output
- Can there be a traffic configuration that causes blocking without overloading one output?
- Other multistage switch fabrics:
  - Different recursive rules
  - Additional stages: distribution to avoid blocking
- Practical note:
  - Standards define Switch Fabric Interfaces

Next Class

- This completes “traditional” network systems
- Next class:
  - New “applications” for networks
  - Basically advanced functionality pushed into routers
  - First step towards “programmable” network systems
- Read:
  - Application papers (two)