Lecture 7: Switches

CS/ECE 438: Communication Networks

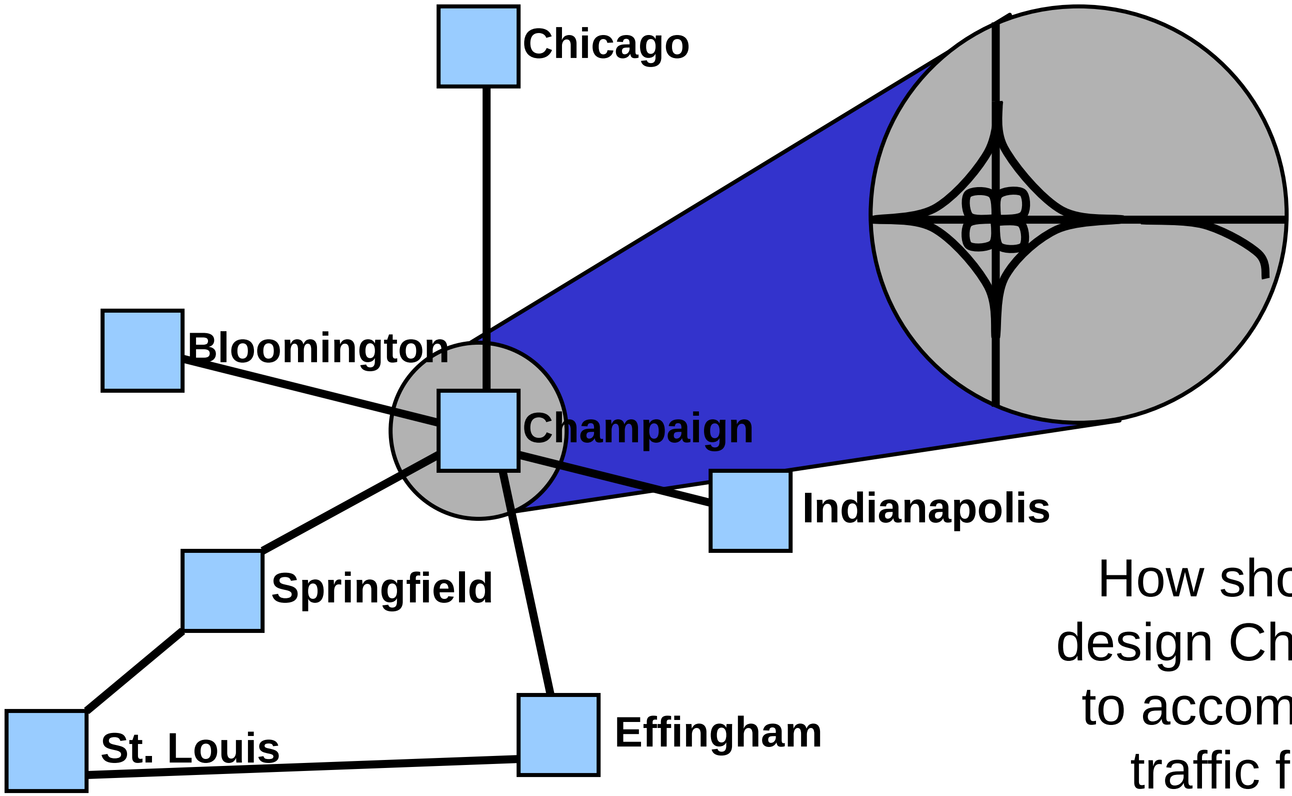
Prof. Matthew Caesar

Feb 26, 2010

Where are we?

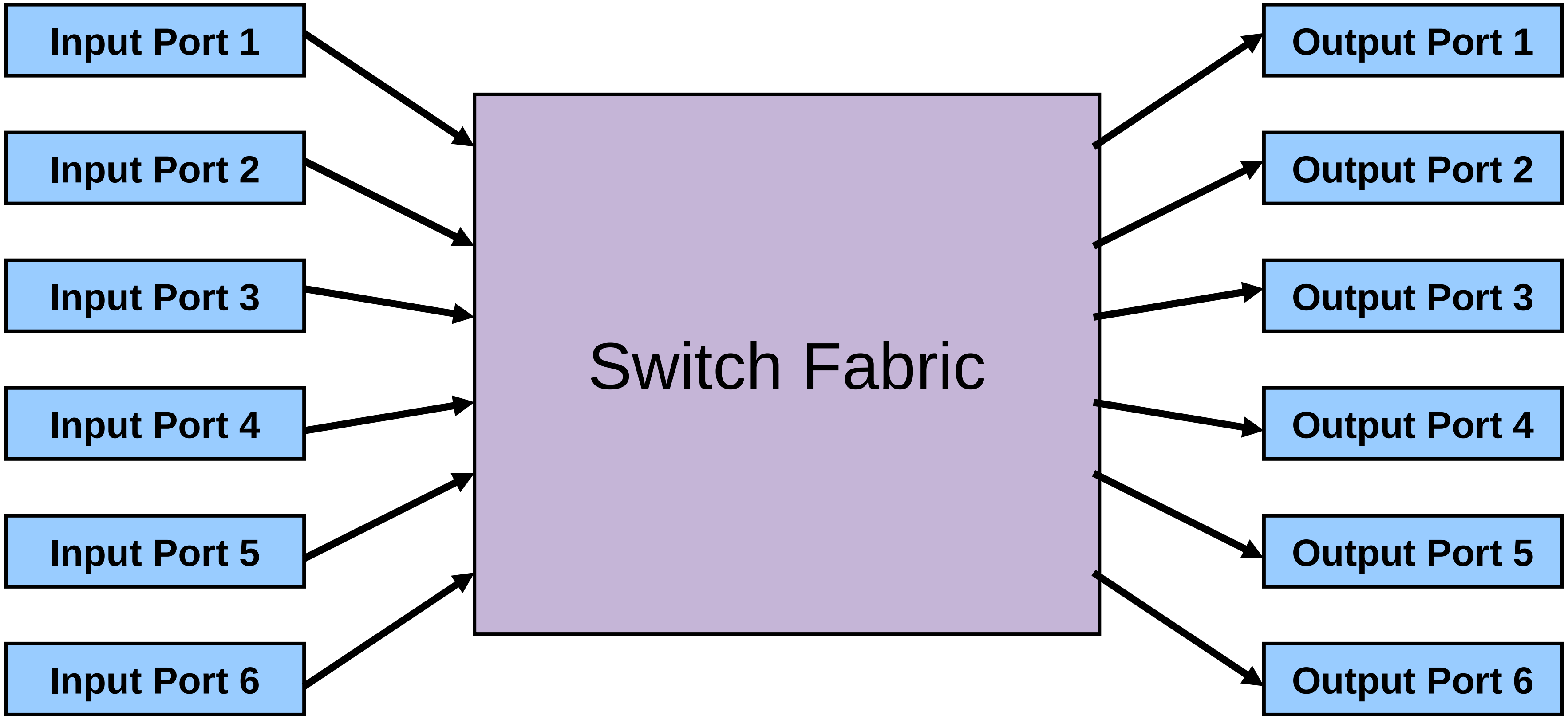
- Understand
  - Different ways to move through a network (forwarding)
    - Read signs at each switch (datagram)
    - Follow a known path (virtual circuit)
    - Carry instructions (source routing)
  - Bridge approach to extending LAN concept
- Next: how switches are built and contention within switches

Switch Design



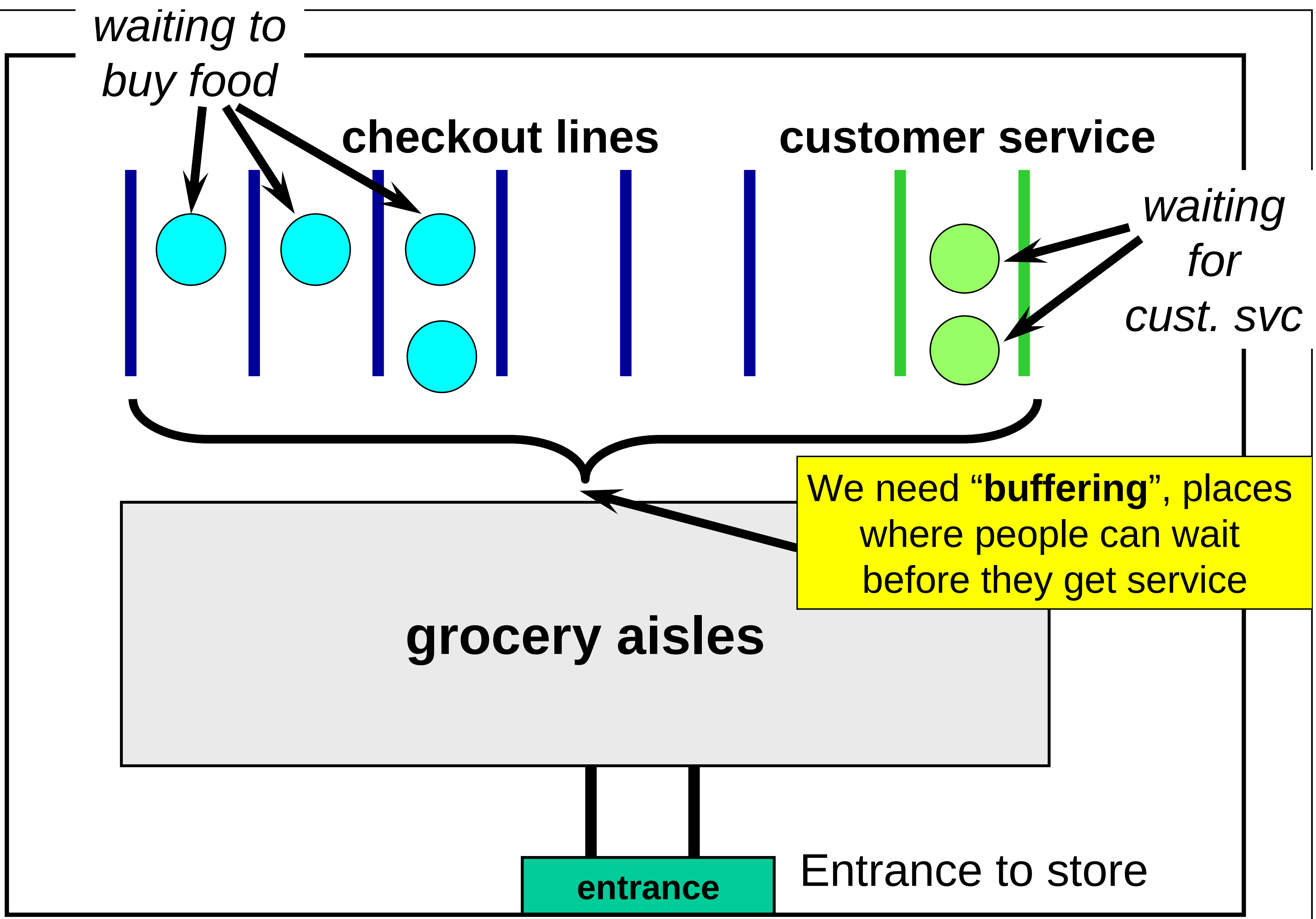
How should we design Champaign to accommodate traffic flows?

Switch Design

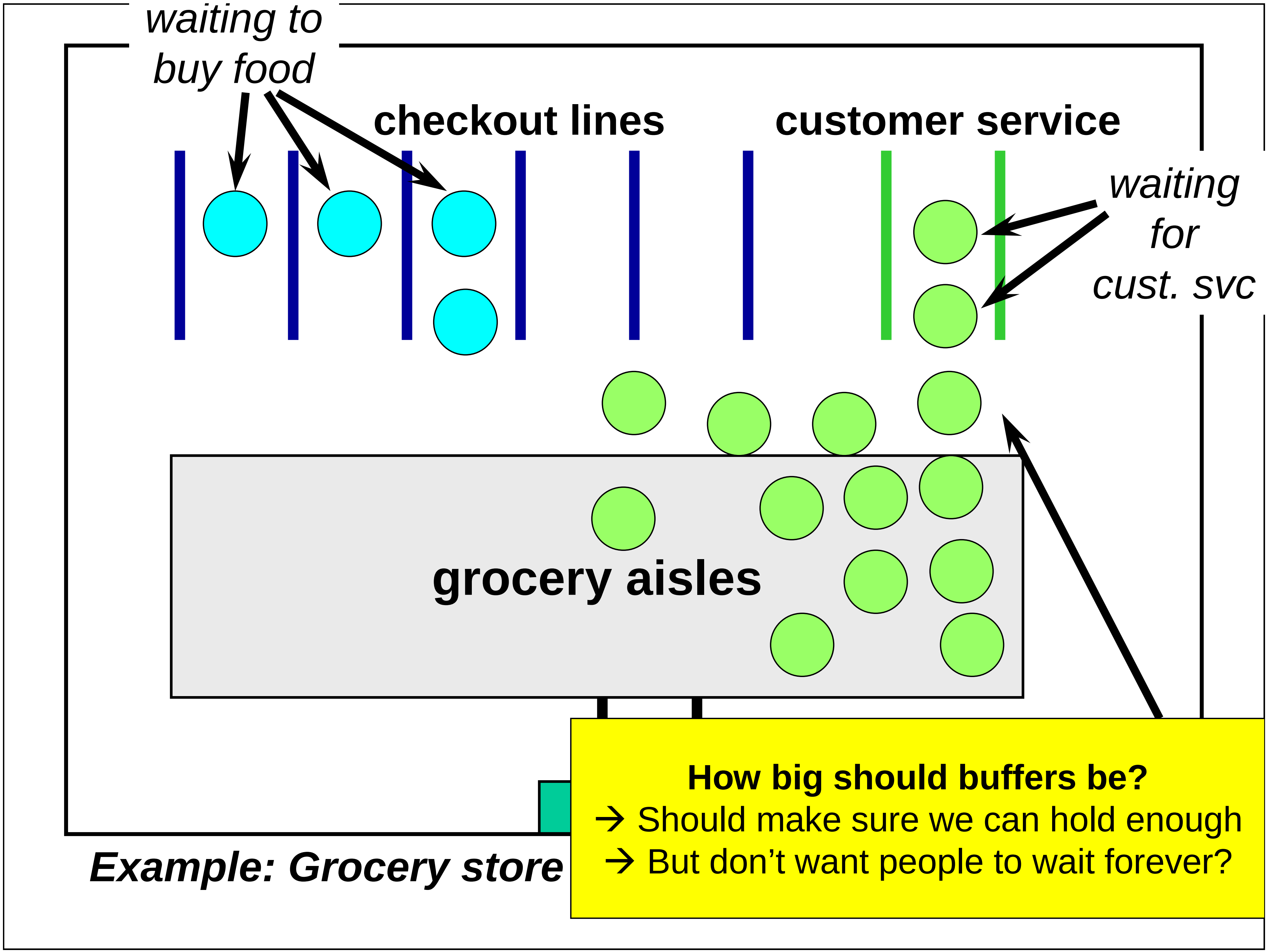


Contention – Output Port Buffering

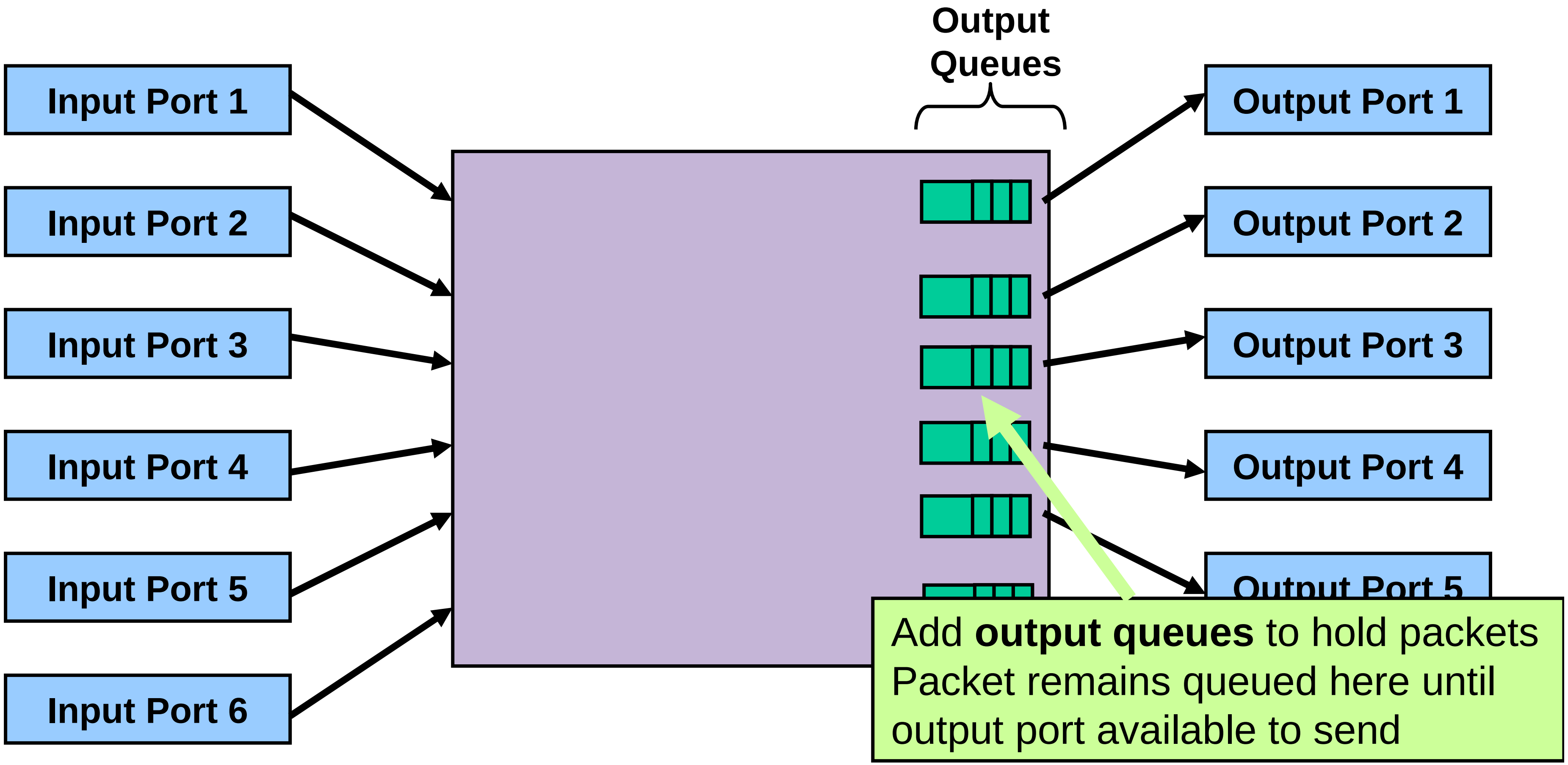
- Problem
  - Some packets may be destined for the same output port
- Solutions
  - One packet gets sent first
  - Other packets get delayed or dropped
- Delaying packets requires buffering
  - Buffers are finite, so we may still have to drop
  - Buffering at input ports
    - Increases, adds false contention
    - Sometimes necessary
  - Buffering at output ports
  - Buffering inside switch

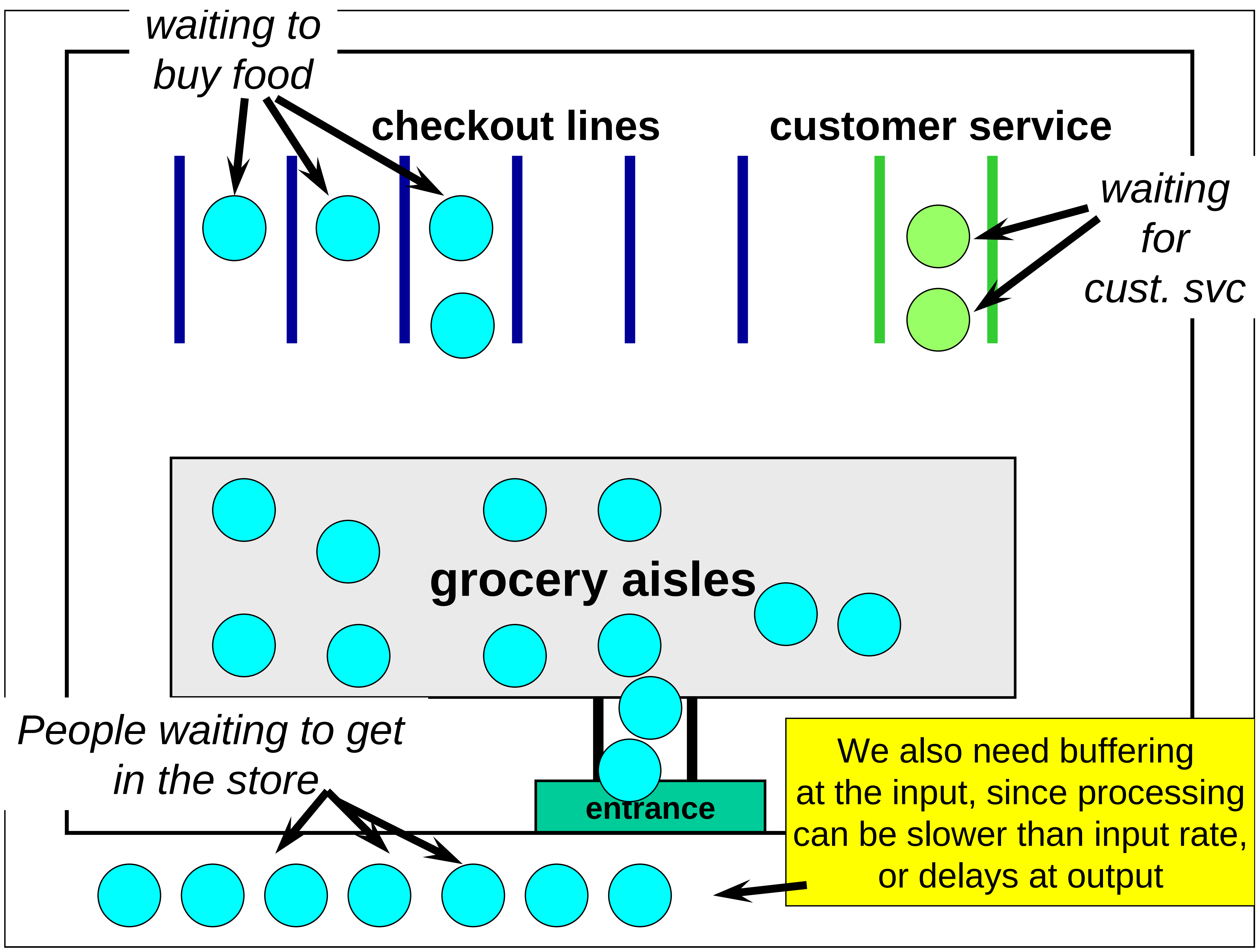


Example: Grocery store

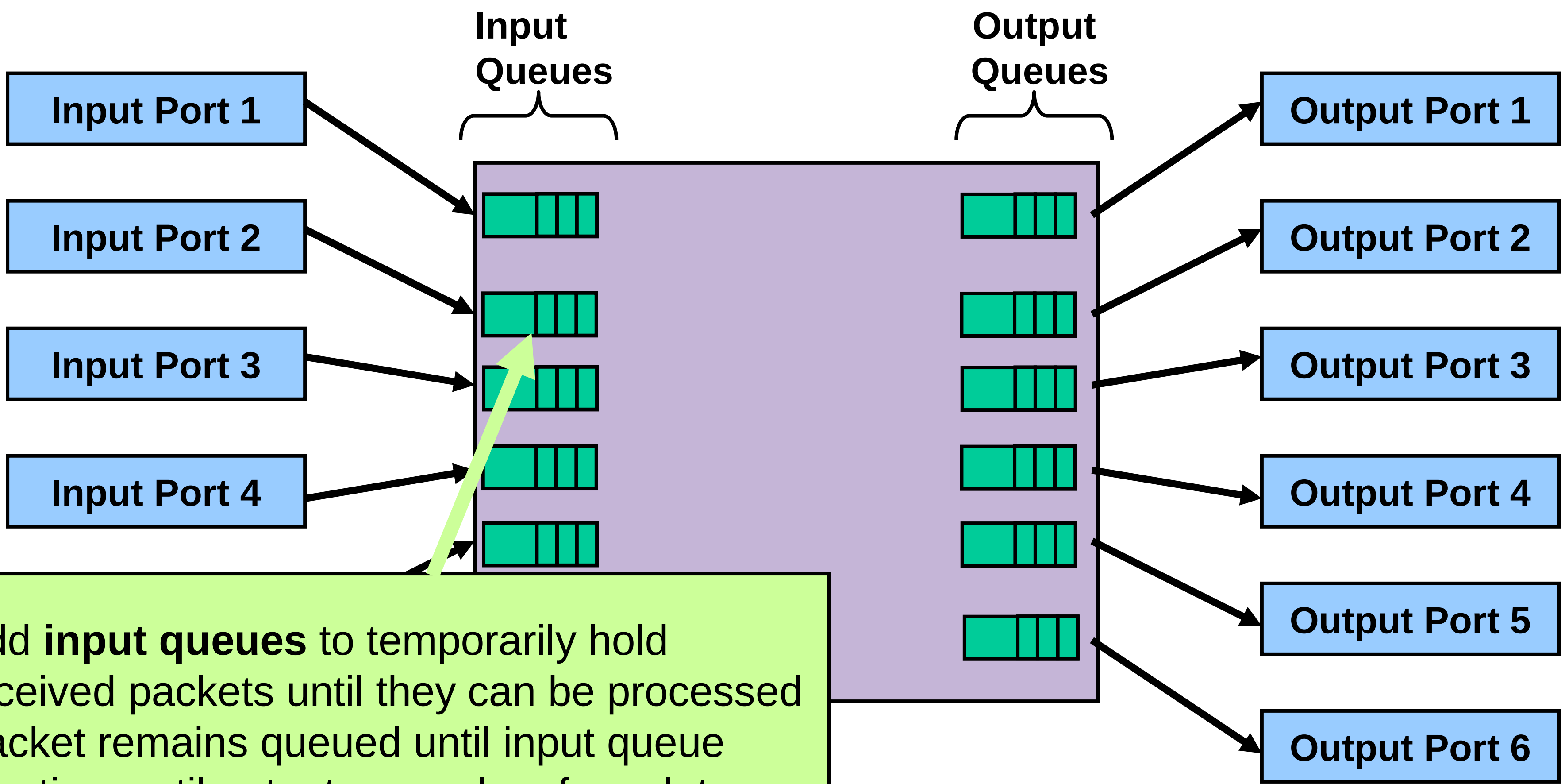

waiting to buy food
checkout lines
customer service
waiting for cust. svc
grocery aisles
Example: Grocery store
How big should buffers be?
→ Should make sure we can hold enough
→ But don't want people to wait forever?

Switch Design




waiting to buy food
checkout lines
customer service
waiting for cust. svc
grocery aisles
People waiting to get in the store
entrance
We also need buffering at the input, since processing can be slower than input rate, or delays at output

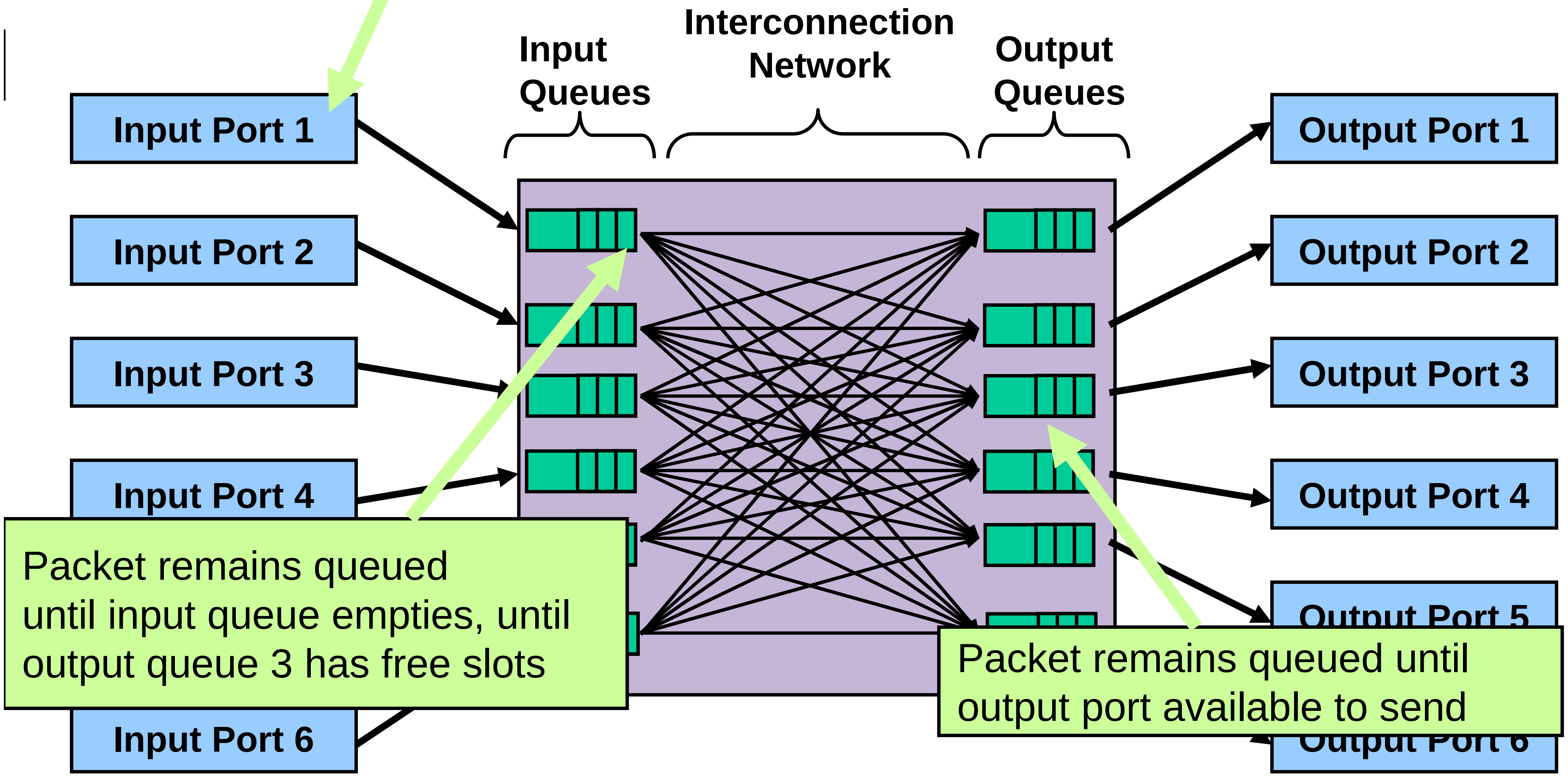
Switch Design



Add **input queues** to temporarily hold received packets until they can be processed
Packet remains queued until input queue empties, until output queue has free slots

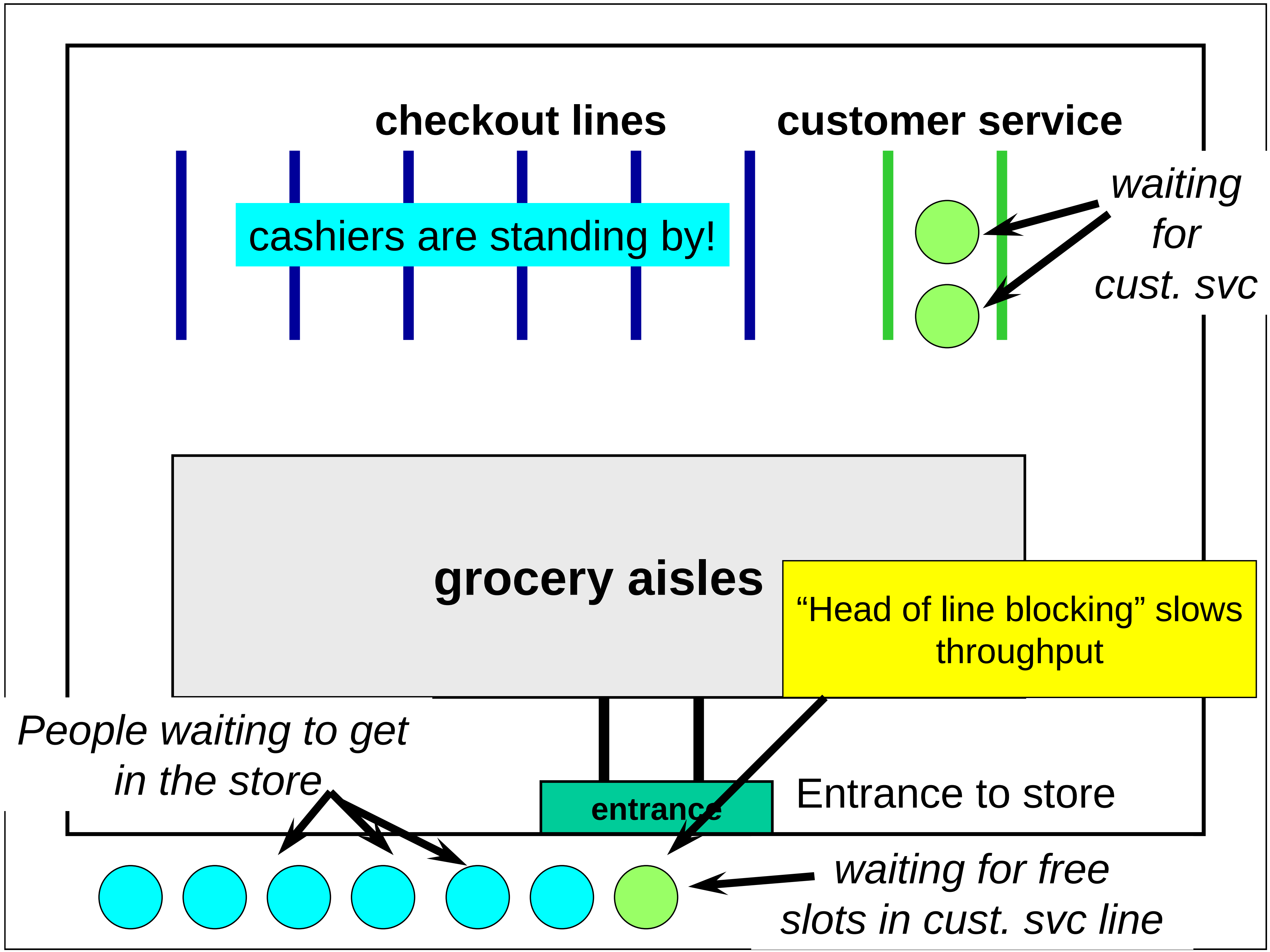
Switch design:
pu together

Looks in forwarding table,
finds output port 3 is associated
with destination

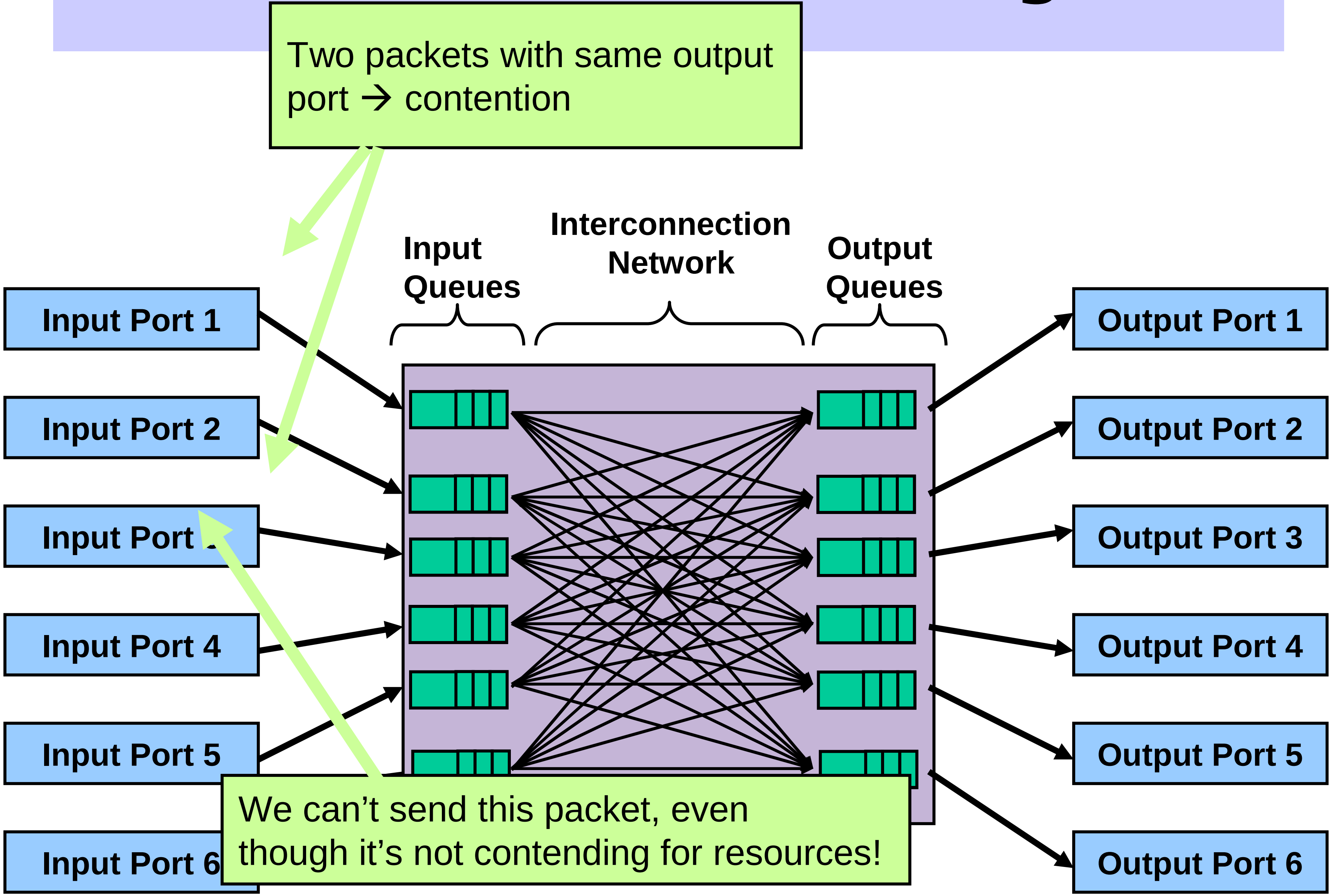


Packet remains queued
until input queue empties, until
output queue 3 has free slots

Packet remains queued until
output port available to send

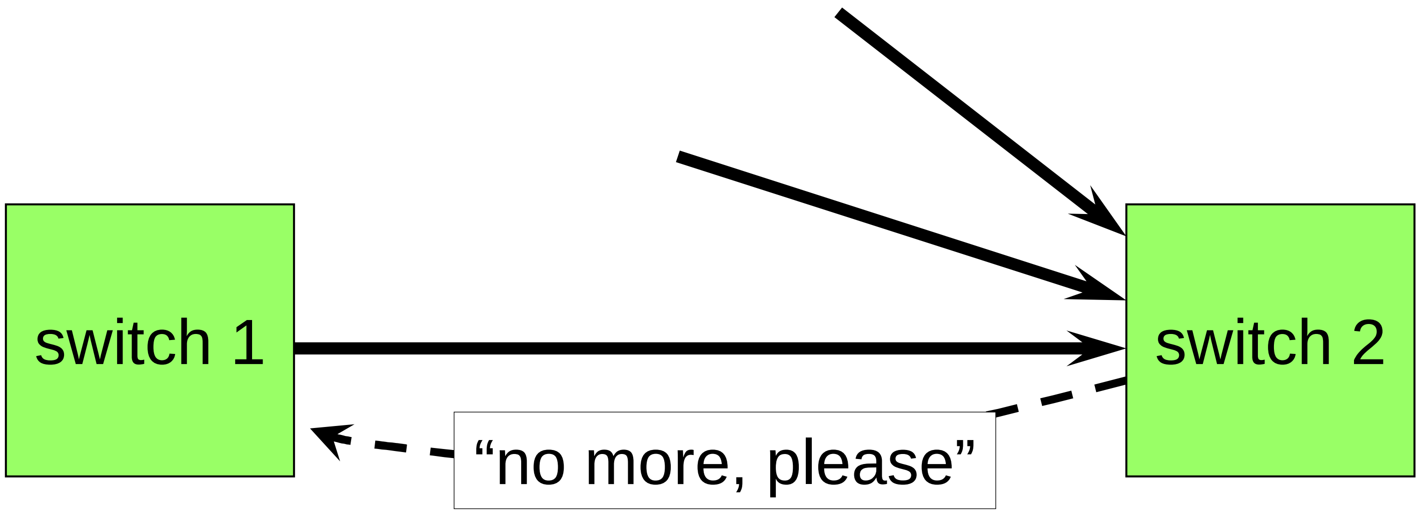

checkout lines
customer service
cashiers are standing by!
waiting
for
cust. svc
grocery aisles
“Head of line blocking” slows
throughput
People waiting to get
in the store
entrance
Entrance to store
waiting for free
slots in cust. svc line

Head of Line Blocking



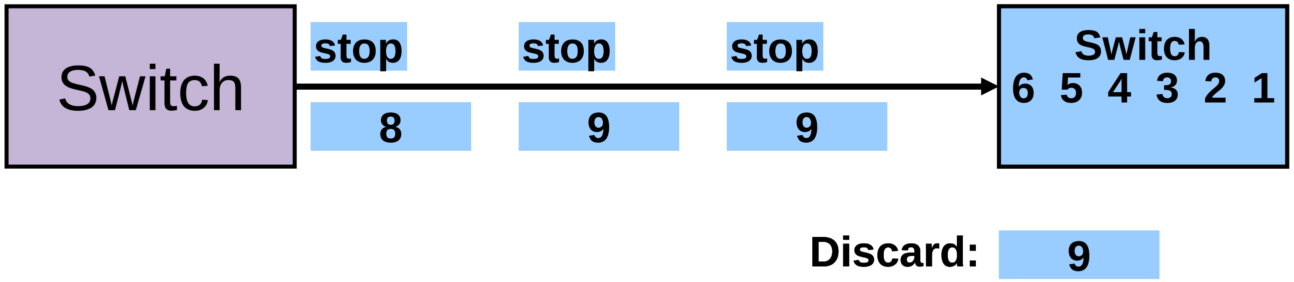
Contention – Back Pressure

- Let the receiver tell the sender to slow down
  - Propagation delay requires that the receiver react before the buffer is full
  - Typically used in networks with small propagation delay



Contention – Back Pressure

- NOTE
  - Propagation delay requires that switch 2 exert backpressure at high-water mark rather when buffer completely full. Backpressure is thus typically only used in networks with small propagation delays (e.g., switch fabrics).

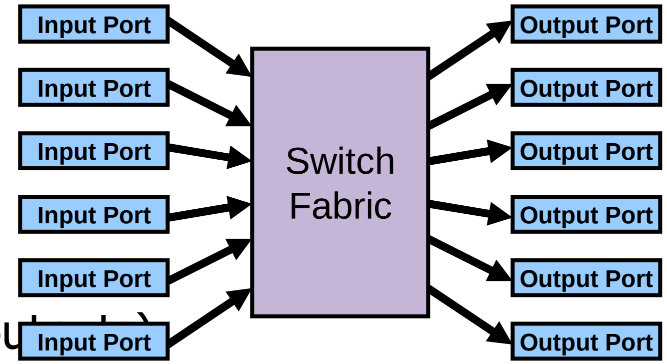


Switch Design Goals

- High throughput
  - Number of packets a switch can forward per second
- High scalability
  - How many input/output ports can it connect
- Low cost
  - Per port monetary costs

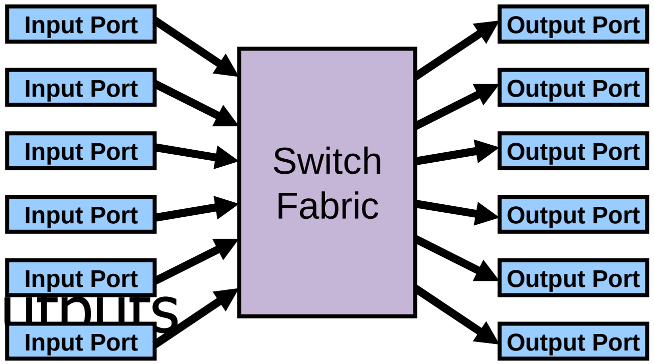
Special Purpose Switches

- Problem
  - Connect N inputs to M outputs
    - NxM ("N by M") switch
    - Often N = M
- Goals
  - High throughput
    - Best is MIN(sum of inputs, sum of outputs)
  - Avoid contention
  - Good scalability
    - Linear size/cost growth



Switch Design

- Ports handle complexity
  - Forwarding decisions
  - Buffering
- Simple fabric
  - Move packets from inputs to outputs
  - May have a small amount of internal buffering



Switch Design Goals

- Throughput
  - Main problem is contention
  - Need a good traffic model
    - Arrival time
    - Destination port
    - Packet length
  - Telephony modeling is well understood
    - Until faxes and modems
  - Data traffic has different properties
    - E.g., phone call arrivals are “Poisson”, but packet arrivals are “heavy-tailed”

Switch Design Goals

- Contention
  - Avoid contention through intelligent buffering
  - Use output buffering when possible
  - Apply back pressure through switch fabric
  - Improve input buffering through non-FIFO buffers
    - Reduces head-of-line blocking
  - Drop packets if input buffers overflow

Switch Design Goals

- Scalability
  - O(N) ports
  - Port design complexity O(N) gives $O(N^2)$ for entire switch
  - Port design complexity of O(1) gives O(N) for entire switch

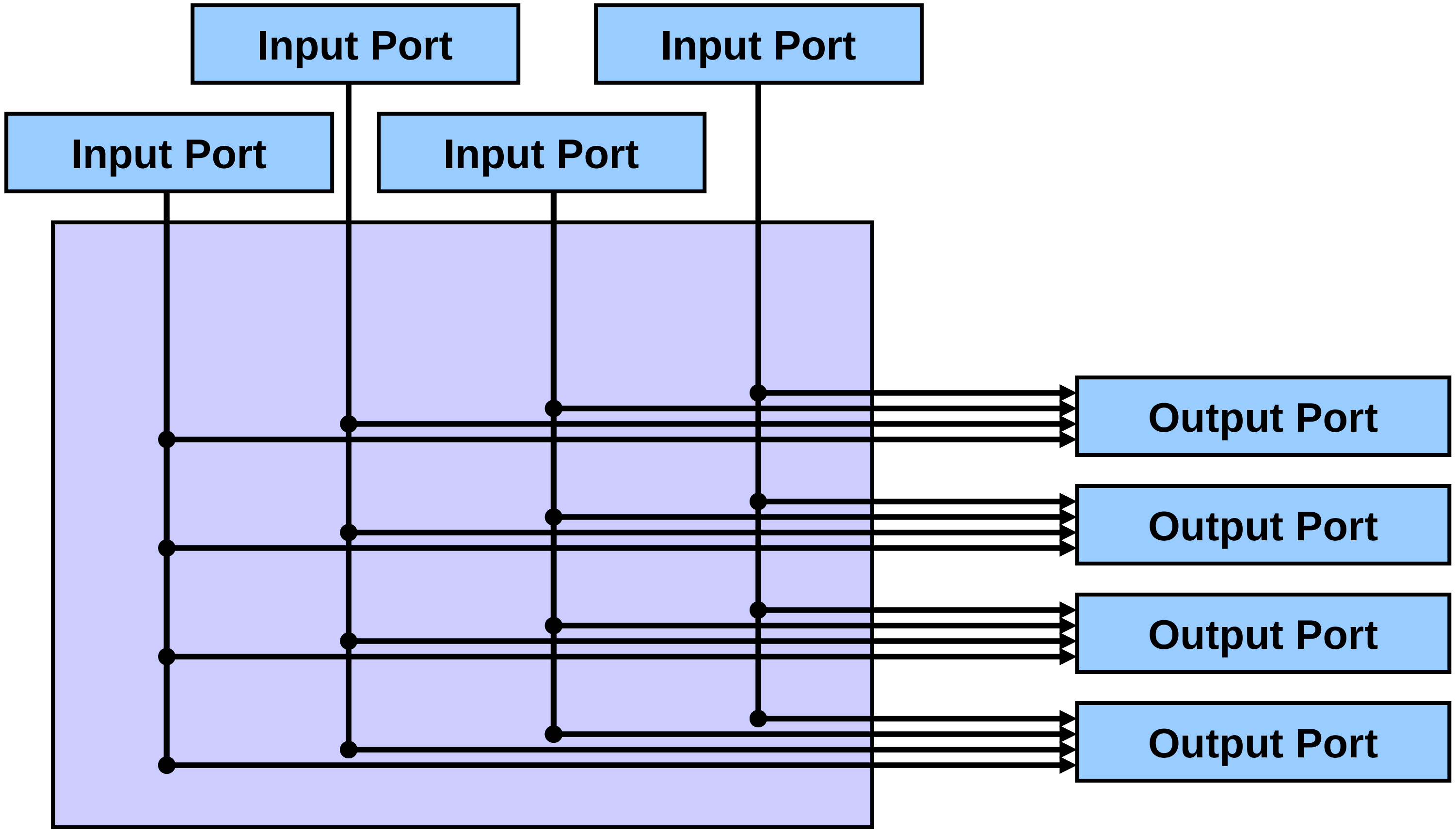
Switch Design

- Crossbar Switches
- Banyan Networks
- Batcher Networks
- Sunshine Switch

Crossbar Switch

- Every input port is connected to every output port
  - NxN
- Output ports
  - Complexity scales as $O(N^2)$

Crossbar Switch



Knockout Switch

- Full crossbar requires each output port to handle up to N input packets
- N simultaneous inputs for the same output is unlikely, especially in a large switch
- Instead, let's implement each port to handle L<N packets at the same time
- Hard issue: what value of L to use?

Knockout switch

- Components:
  - Packet filters (recognize packets destined for this output port)
  - Concentrator (selects subset of L packets, “knocks out” others)
  - A queue with capacity L packets

Knockout switch

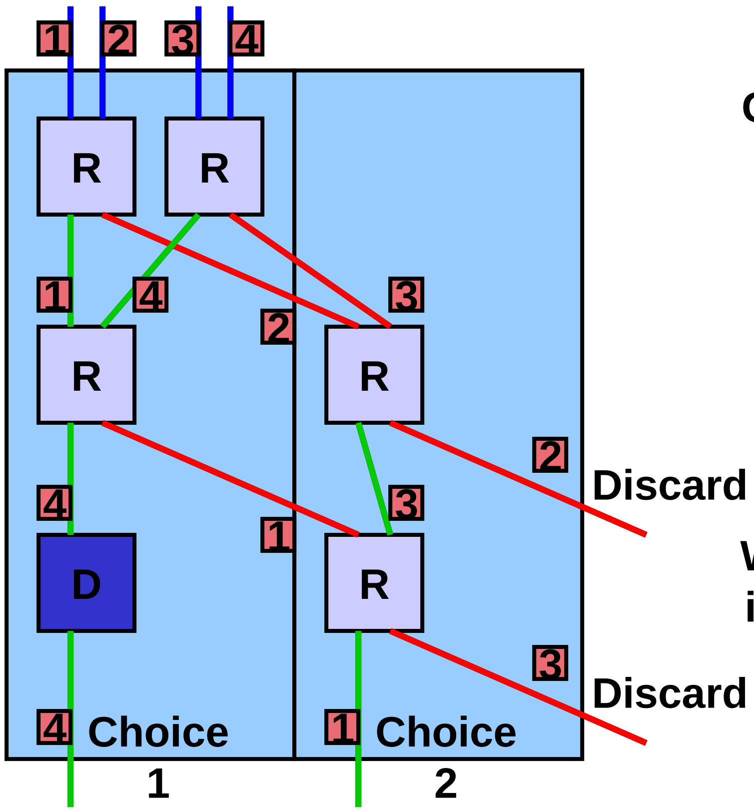
- Want some fairness: no single input should have its packets always "knocked out"
- Essentially a "knock out" tennis tournament with each game of 2 players (packets) chosen randomly
- Overall winner is selected by playing log N rounds, and keeping the winner

Knockout switch

- Pick L from N packets at a port
  - Output port maintains L cyclic buffers
  - Shifter places up to L packets in one cycle
  - Each buffer gets only one packet
  - Output port uses round-robin between buffers
  - Arrival order is maintained

- Output ports scale as O(N)

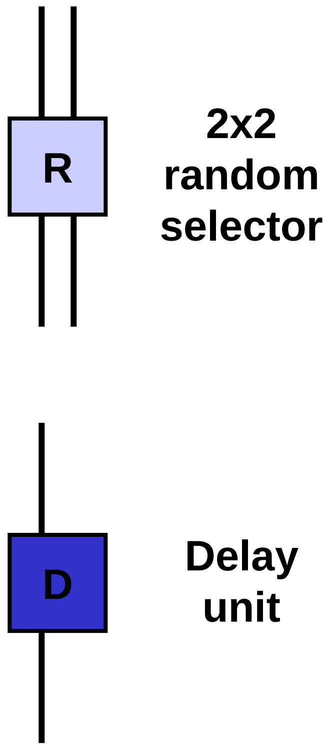
Knockout Switch



Choose L of N

Ex: 2 of 4

What happens if more than L arrive?

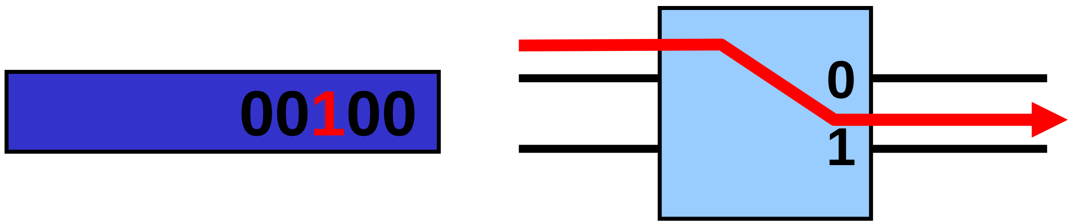


Self-Routing Fabrics

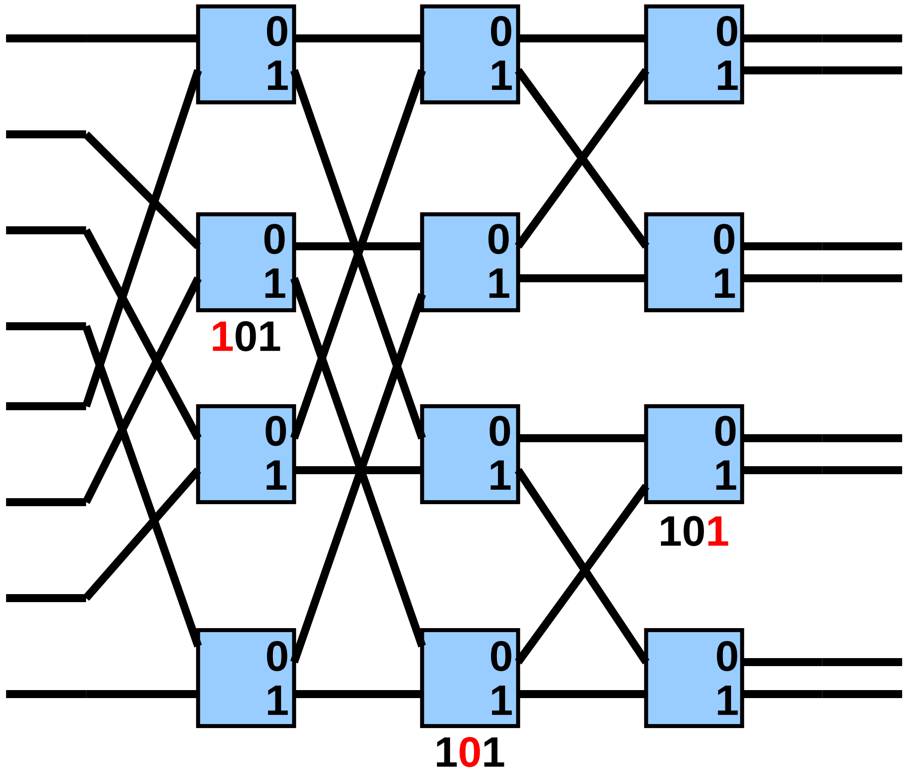
- Idea
  - Use source routing on “network” in switch
  - Input port attaches output port number as header
  - Fabric routes packet based on output port
- Types
  - Banyan Network
  - Batcher-Banyan Network
  - Sunshine Switch

Banyan Network

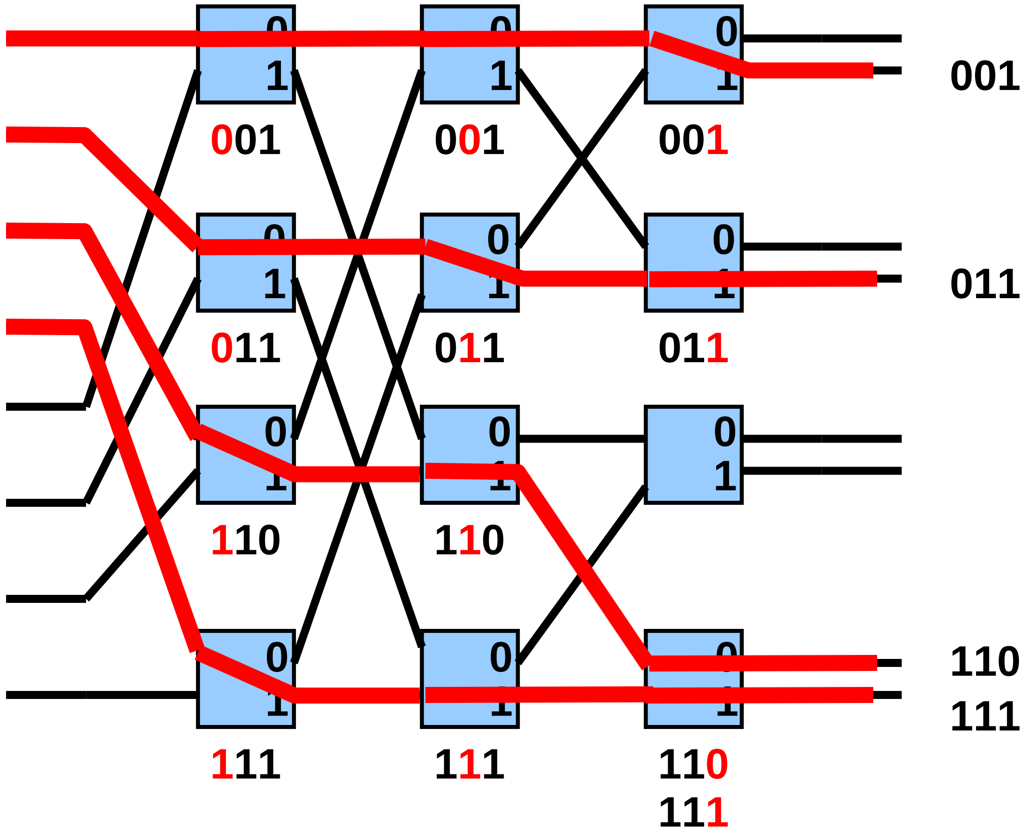
- A network of 2x2 switches
  - Each element routes to output 0 or 1 based on packet header
  - A switching element at stage i looks at bit i in the header



Banyan Network



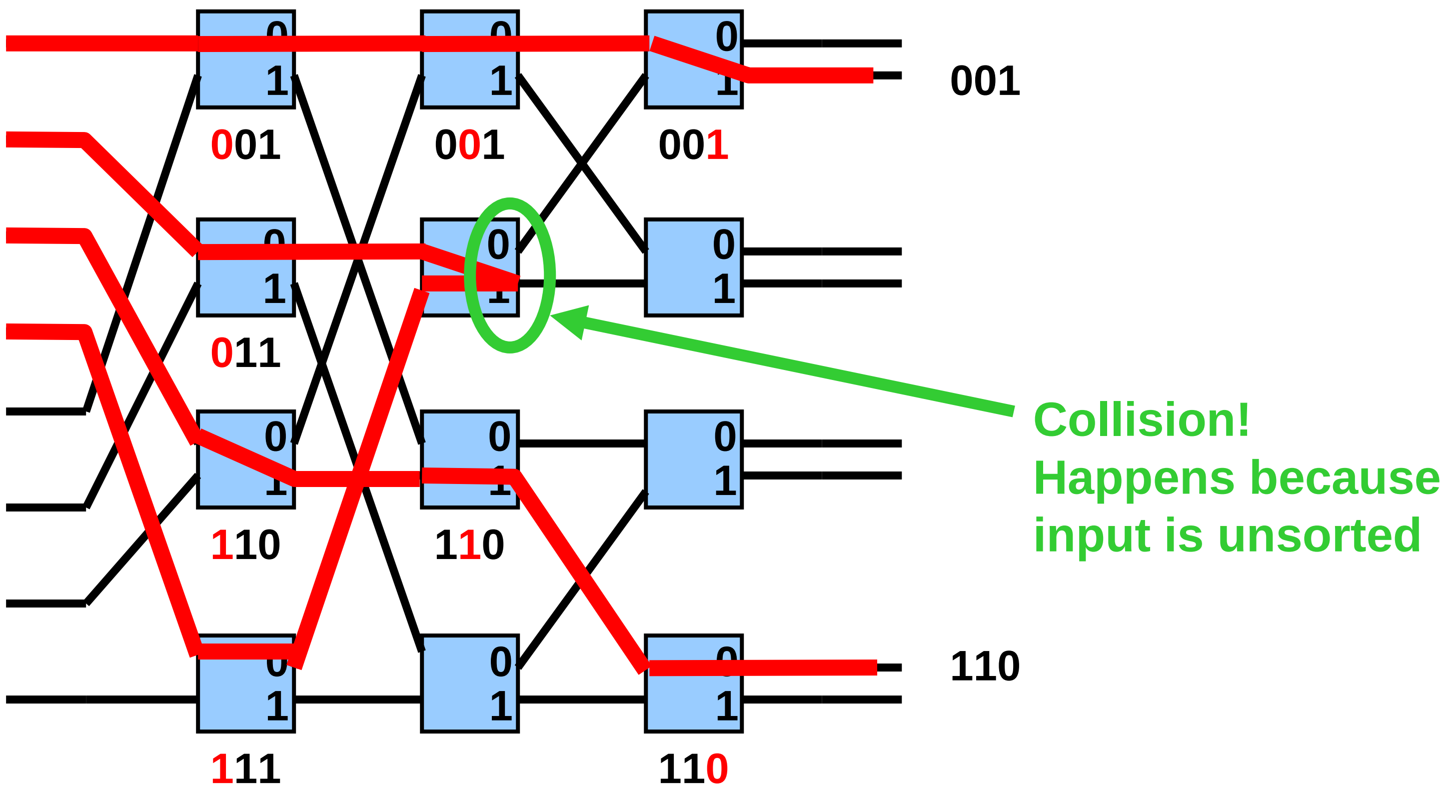
Banyan Network



Banyan Network

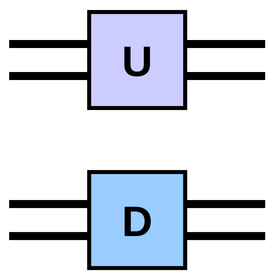
- Perfect Shuffle
  - N inputs requires $\log_2 N$ stages of N/2 switching elements
  - Complexity on order of $N \log_2 N$
- Collisions
  - If two packets arrive at the same switch destined for the same output port, a collision will occur
  - If all packets are sorted in ascending order upon arrival to a banyan network, no collisions will occur!

Collision in a Banyan Network

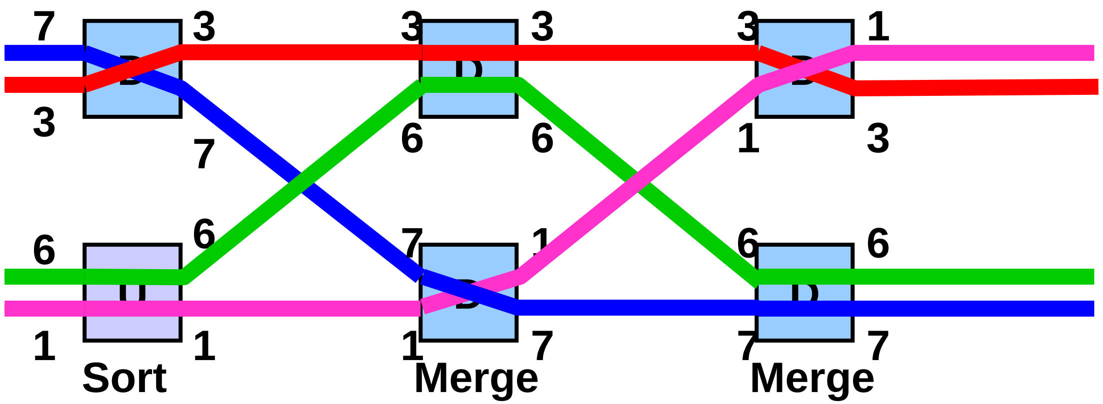


Batcher Network

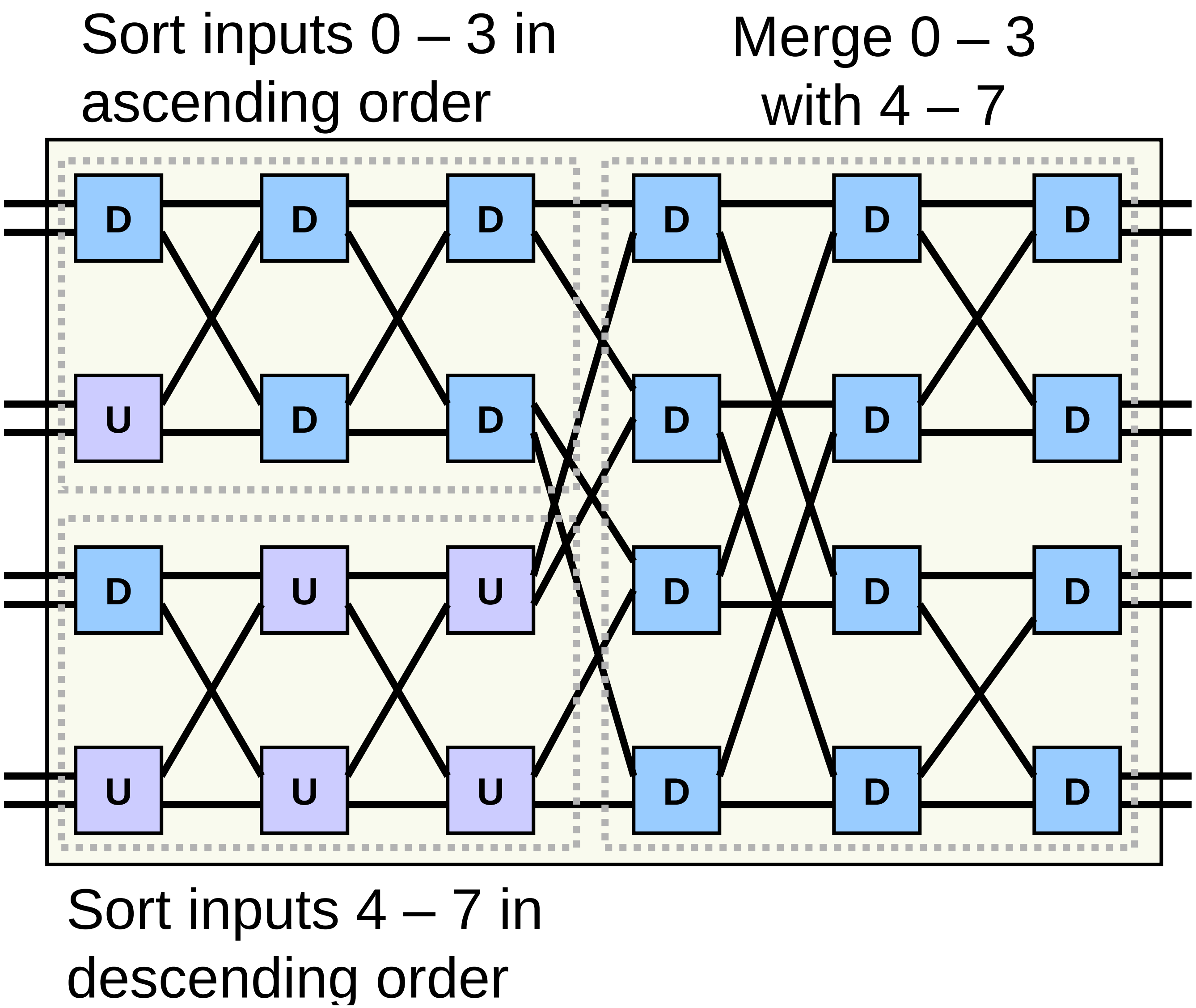
- Performs merge sort
- A network of 2x2 switches
  - Each element routes to output 0 or 1 based on packet header
  - A switch at stage i looks at the whole header
  - Two types of switches
    - Up switch
      - Sends higher number to top output (0)
    - Down switch
      - Sends higher number to bottom output (1)



Batcher Network



Batcher Network



8x8 Switch

Batcher Network

- How it really works
  - Merger is presented with a pair of sorted lists, one in ascending order, one in descending order
  - First stage of merger sends packets to the correct half of the network
  - Second stage sends them to the correct quarter
- Size
  - N/2 switches per stage
  - $\log_2 N \times (1 + \log_2 N)/2$ stages
  - Complexity = $N \log_2^2 N$

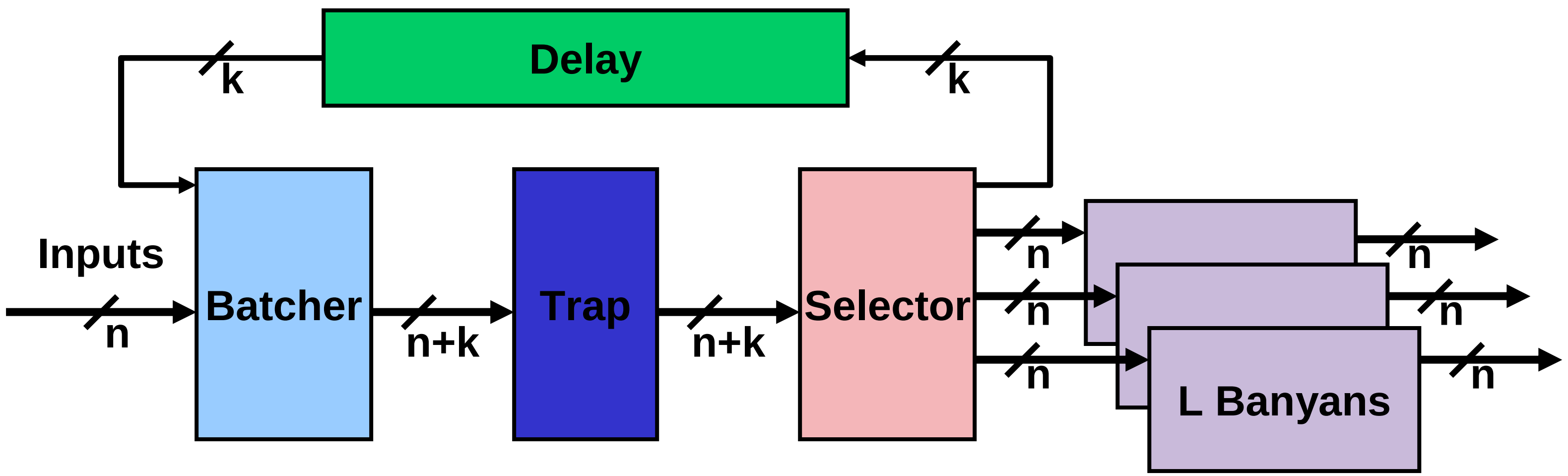
Batcher-Banyan Network

- Idea
  - Attach a batcher network back-to-back with a banyan network
  - Arbitrary unique permutations can be routed without contention
- Sunshine Switch
  - Like a knockout switch
    - Can handle up to L packets per output port
  - Recirculates overflow packets
    - If more than L packets arrive for any output port in one cycle

Sunshine Switch

- Elements
  - Multiple Banyan networks
    - Enables multiple packets per output port
  - Delay Box
    - Excess (K) packets are recirculated and resubmitted to the switch
  - Batcher network
    - N new packets
    - K delayed packets
  - Trap
    - Identifies packets destined for banyan
    - Identifies excess packets
  - Selector
    - Routes multiple packets for same output on separate banyans

Sunshine Switch



Sunshine Switch

- Can packets circulate for ever?
  - Priority bit is used to favor older packets
  - Priority bit also ensure packet order is preserved through the switch