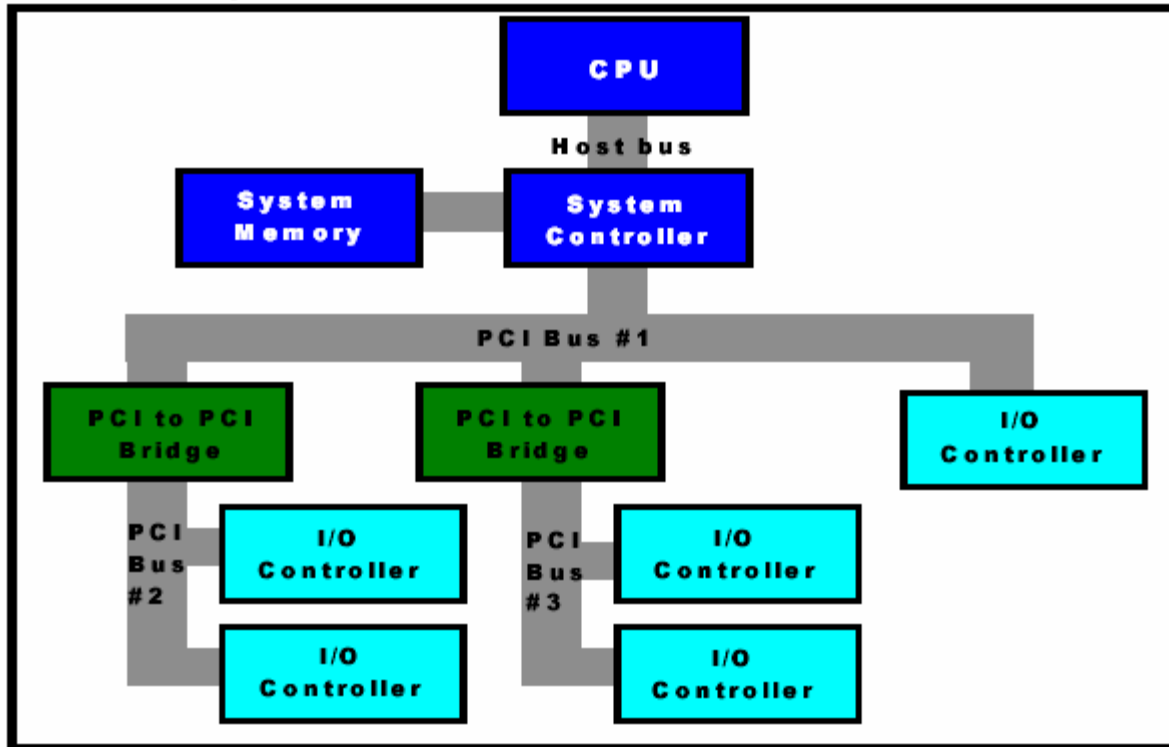


- System-level Interconnects
  - PCI, PCI-X, PCI-Express
  - HyperTransport, RapidIO
  - Advanced Switching

# System-level interconnect



PCI – dominant standard

standard includes physical interface, signalling, and read/write data semantics – address + data

# PCI

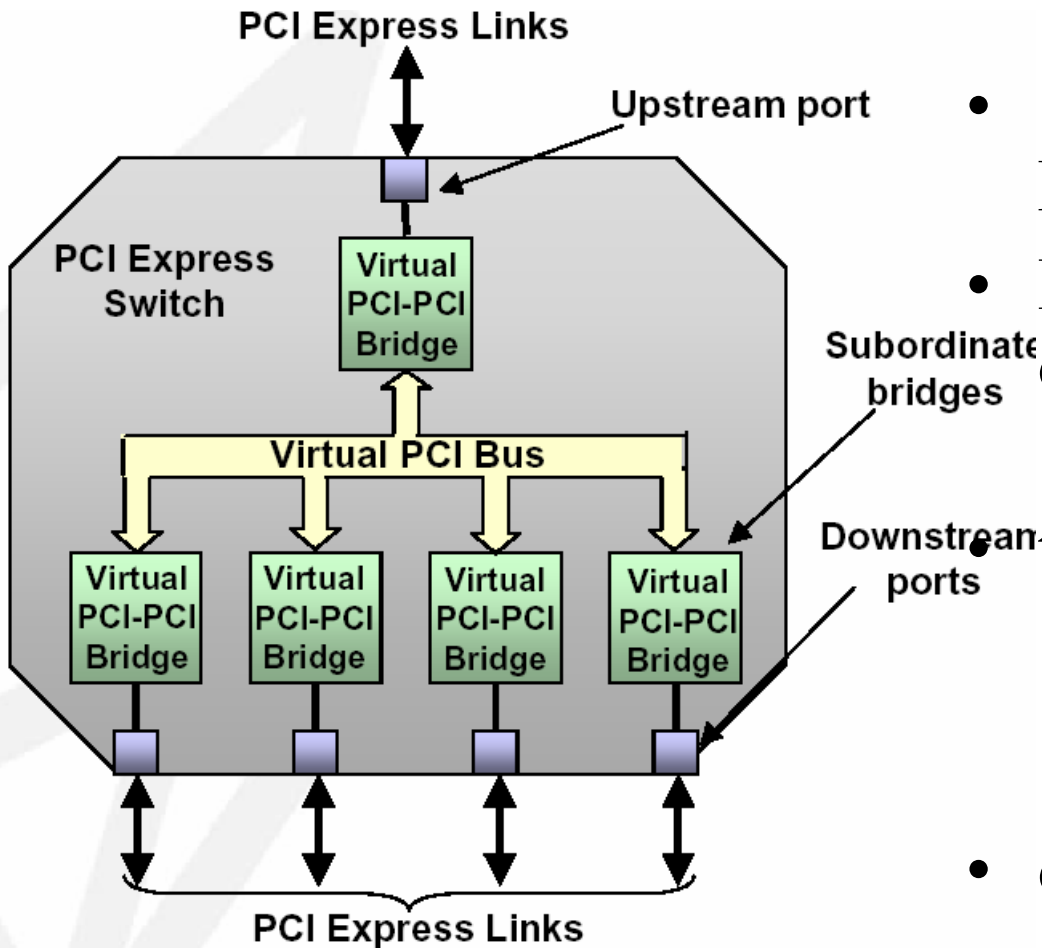
- around since early '90s, originally Intel, backed by PCI-SIG
- PCI – 32 or 64bit /33 or 66MHz
- parallel, shared bus
  - synchronized clock signal
  - scalability in terms of # of devices
  - limit on physical distance (clock skew)
  - electrical and speed limits due to bus loads
  - reliability – one misbehaving guy on shared bus...
- limitation drive domain targeted solutions (e.g., graphics ports, chip-to-chip busses... )
- load/store (read/write) transactions
- address + data phase, with address decode happening in first
  - 7ns for 33MHz, 3ns for 66MHz!

# PCI-X

- PCI-X – still a shared parallel bus,
  - 64bit @ 133,266,533(1066) MHz
  - fully compatible with PCI, physically, signals, systems, software...
    - weakest device on the bus will set performance limits
  - signal in registers to extend decoding time budget
    - but faster clocks, so latency increase - not necessarily
  - clock seed + throughput + additional QoS
    - error correction codes (link-layer), device ID messages, split transactions, reordering of transactions, attribute field with byte count -> better buffer management...

# PCI-Express (PCIe)

- HP serial I/O interconnect
  - packet-based, serial switched (not shared, but point-to-point)
- Software compatible with PCI, but not otherwise
- Logically – strict hierarchy of PCI bridges
  - just like before
- target domain – system level
  - 10s, 100s, few thousands nodes; not very dynamic topology
- one host & root switch
  - device-to-device comm



# PCIe - layers

- PCIe Standard specifies layers:
- Physical
  - lane – Rx/Tx pairs @2.5Gbps with 8b/10b encoding (with code words for delimiters and control)
  - 1, 2, 4, 8, 12, 16, 32 lanes (reaches up to 16Gbytes/s in each direction!)
    - Configurable widths
  - 32bit aligned data
  - fewer pins -> less power, space

- Data Link
  - managed by 64b DLLP
  - synchronize link, acks, credit info...
  - link-layer CRC and checking and retry if necessary
- Transport
  - TLP: sequence#, header, CRC, data, of maximum 4kB
  - header specifies operation options (split transaction, messages...)
  - virtual channels and traffic classes (mapping not fixed)
  - credit-based flow control on per VC
    - will not starve for data
- Originally – 3GIO, Arapahoe Group

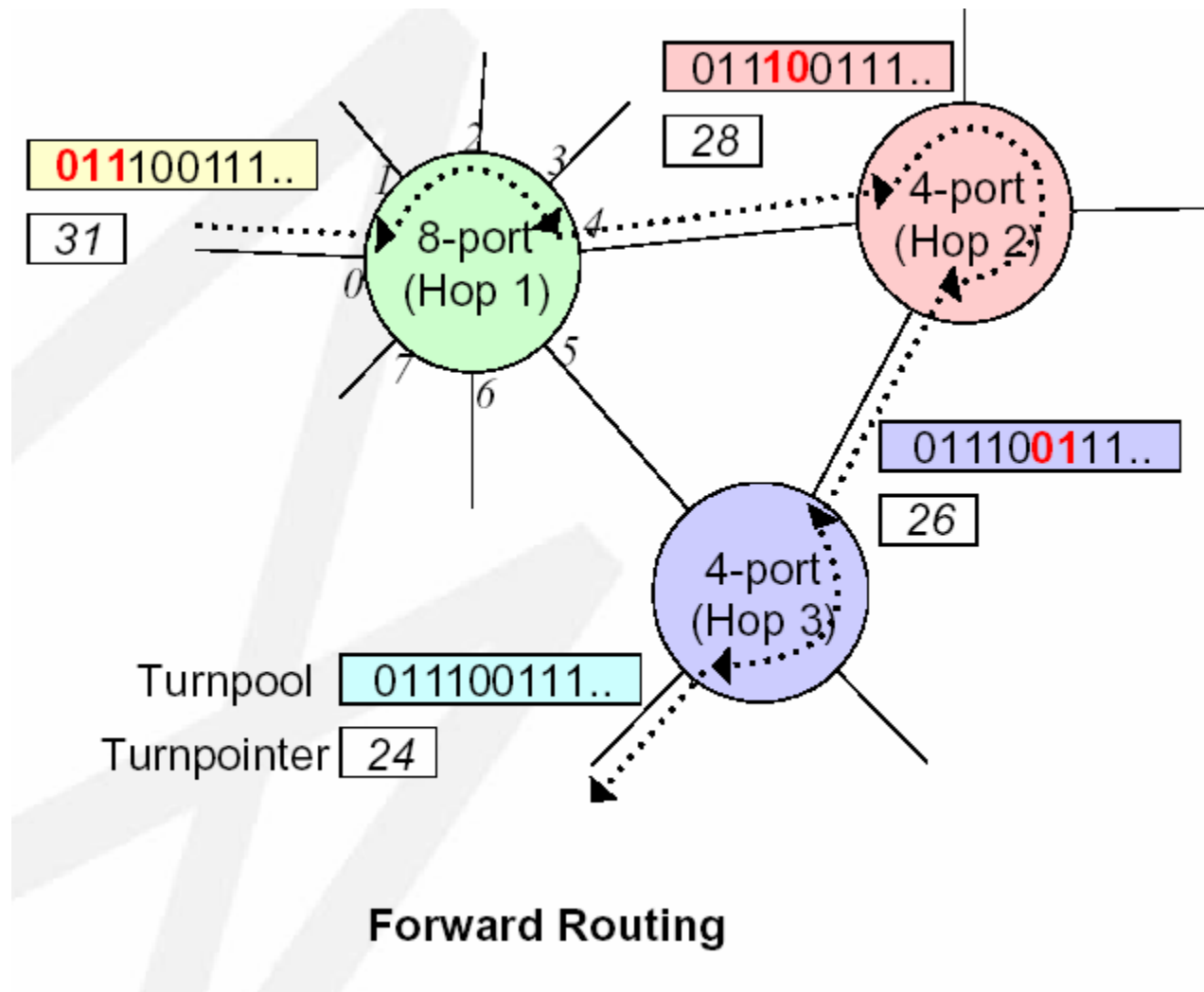
# Advanced Switching

- PCIe – single host, not for peer-to-peer and multiprocessing
- PCI-Express AS -> into networking domain (arbitrary topologies, increased reliability...)
- AS – routing model, protocol agnostic, shared physical and link layer with PCIe, but can encapsulate other traffic
  - header: route-specific-part + protocol-specific-part
  - AS – uses first part only
  - unicast – source-based routing (provide encoding of path, can backtrace route to source)
  - multicast – destination-based routing (support number of mcast groups)

turnpool  
encodes path

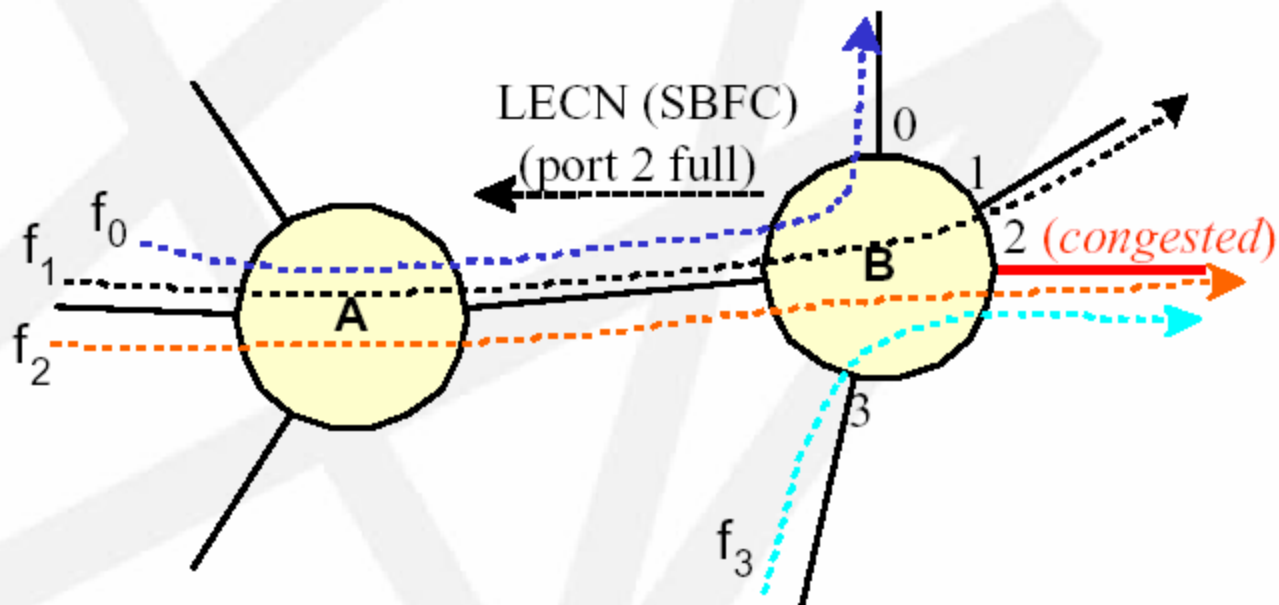
turnpointer  
keeps track of  
status

bits depend on #ports



# Congestion management

- switches detect (buffer thresholds...) and can act upon
- Forward/Backward/Local Explicit Congestion Notification



Standard	Bus Width	Clock	Transfer
PCI 2.3	32 Bit	33 MHz 66 MHz	133 MB/s 266 MB/s
PCI 64	64 Bit	33 MHz 66 MHz	266 MB/s 533 MB/s
PCI-X 1.0	64 Bit	66 MHz 100 MHz 133 MHz	533 MB/s 800 MB/s 1066 MB/s
PCI-X 2.0 (DDR)	64 Bit	133 MHz	2132 MB/s
PCI-X 2.0 (QDR)	64 Bit	133 MHz	4264 MB/s
PCI Express	1 Lines, 8 Bit	2.5 GHz	512 MB/s
PCI Express	2 Lines, 8 Bit	2.5 GHz	1 GB/s (Duplex)
PCI Express	4 Lines, 8 Bit	2.5 GHz	2 GB/s (Duplex)
PCI Express	8 Lines, 8 Bit	2.5 GHz	4 GB/s (Duplex)
PCI Express	16 Lines, 8 Bit	2.5 GHz	8 GB/s (Duplex)
PCI Express	32 Lanes, 8 Bit	2.5 GHz	16 GB/s (Duplex)

# RapidIO and HyperTransport

- RapidIO (Motorola), HyperTransport (AMD)
  - mainly embedded and communications market
    - (HyperTransport also chip-to-chip in AMD server systems, Xbox)
  - switched, packet-based, originally parallel, now serial
    - trade-off is #pins and range vs. serialization/packetization logic & latency
  - link error and flow control in hardware
  - shared distributed memory, coherency protocols
  - request/transaction queue for QoS management
  - RapidIO/HT-Bridge-to-PCI\*
- Hypertransport 3.0
  - Up to 2.6GHz clock rates (1GHz)
  - Up to 41.6GB/s aggregate rate (8GB/s)
- RapidIO
  - Parallel and serial (up to 10Gbps)
  - Design geared towards communications/streaming apps

From hypertransport TA:

