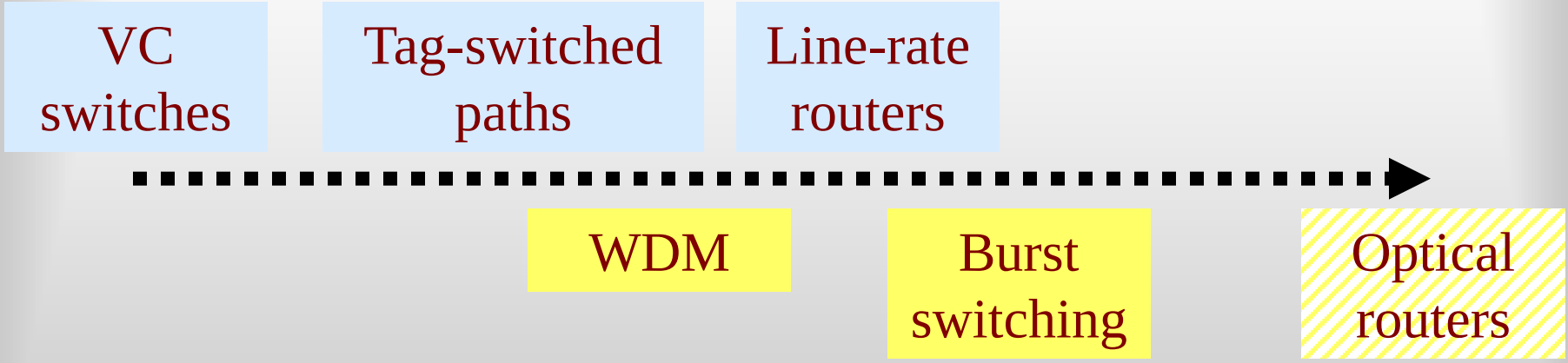
Building an All-Optical Internet Router

Joe Touch
Postel Center
Computer Networks Division
USC/ISI

Path to Optical Routers

- **Evolution of electronics**



- **Evolution of optics**

Current optical focus

- **WDM as a bonus**
  - Needed to overcome dispersion
  - Can be used to partition? or route?
- **Connection-based/-like traffic**
  - ATM/MPLS flow-based setups (MPλS, SWAP)
    - BUT: Setup doubles connection latency
  - Packet-train setup on-the-fly (OBS, TBS)
    - BUT: Setup requires large gap after first packet
  - BUT: Both expect long flows or aggregation

Goal – Optical Internet

(the rest of this talk)

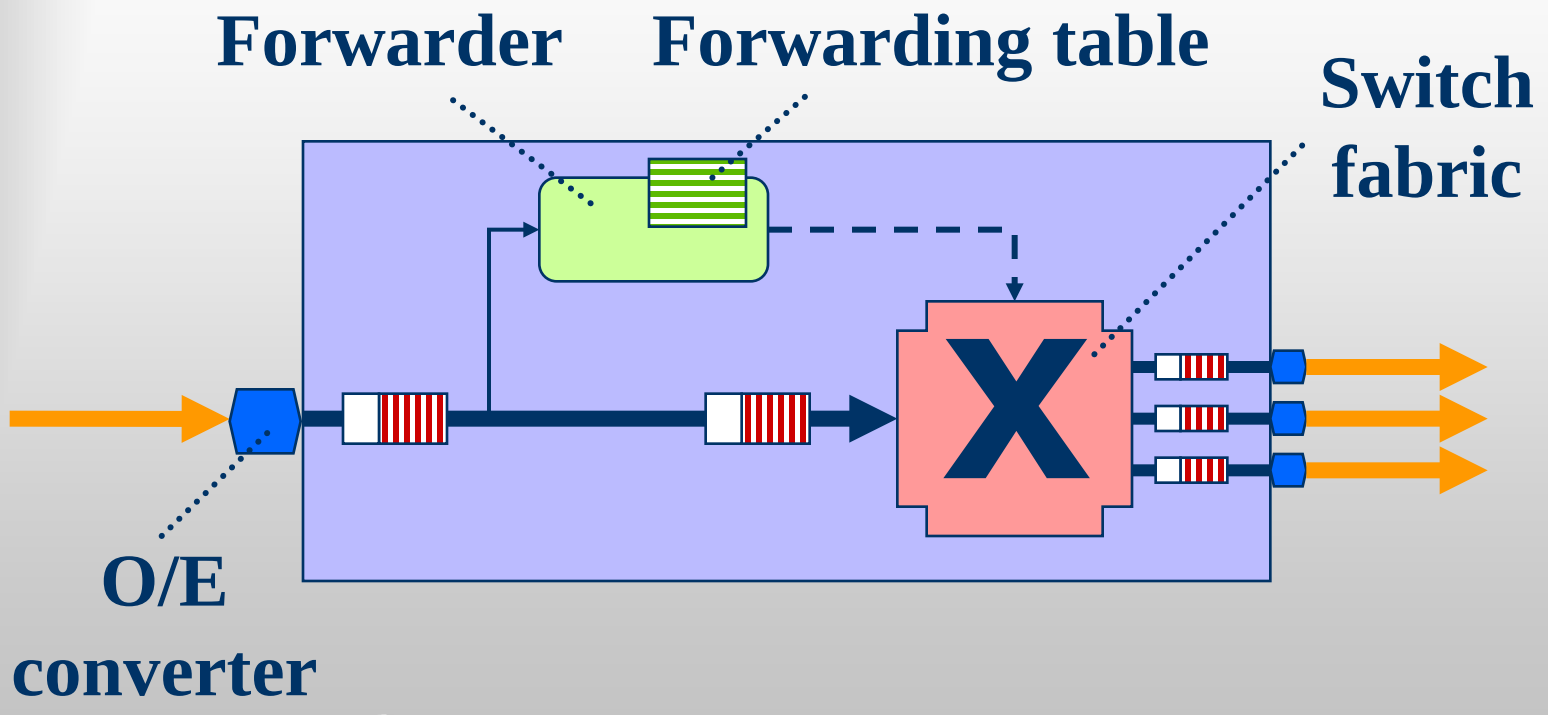
- **IP over light**
  - No setup
  - Single terabit channels (*no WDM*)
  - Works for short flows, or for single packets

Challenges

- **Router Design**
  - Queue-free architecture
  - Forwarding via partial filters
  - TTL decrement
  - IP checksum
- **LAN Protocols (another talk…)**
  - OCDMA MAC design

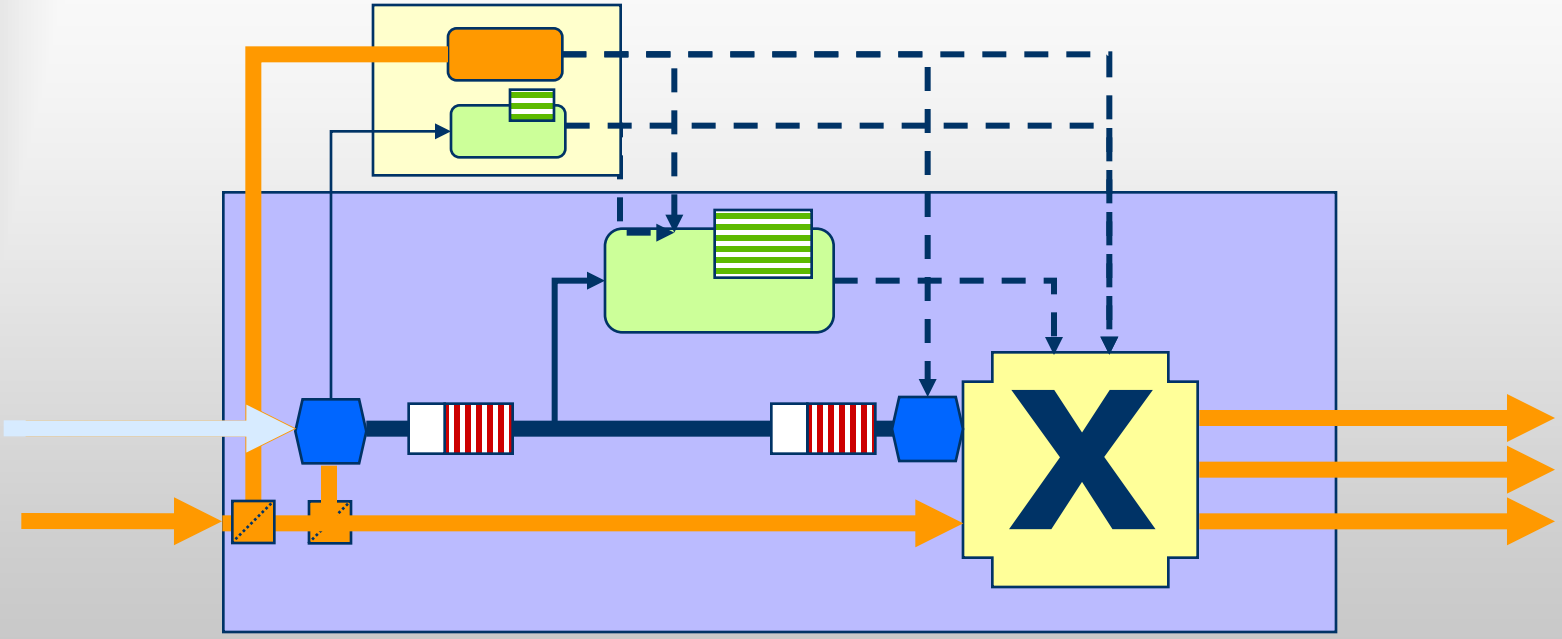
Inside current routers

- Forwarder + switch fabric

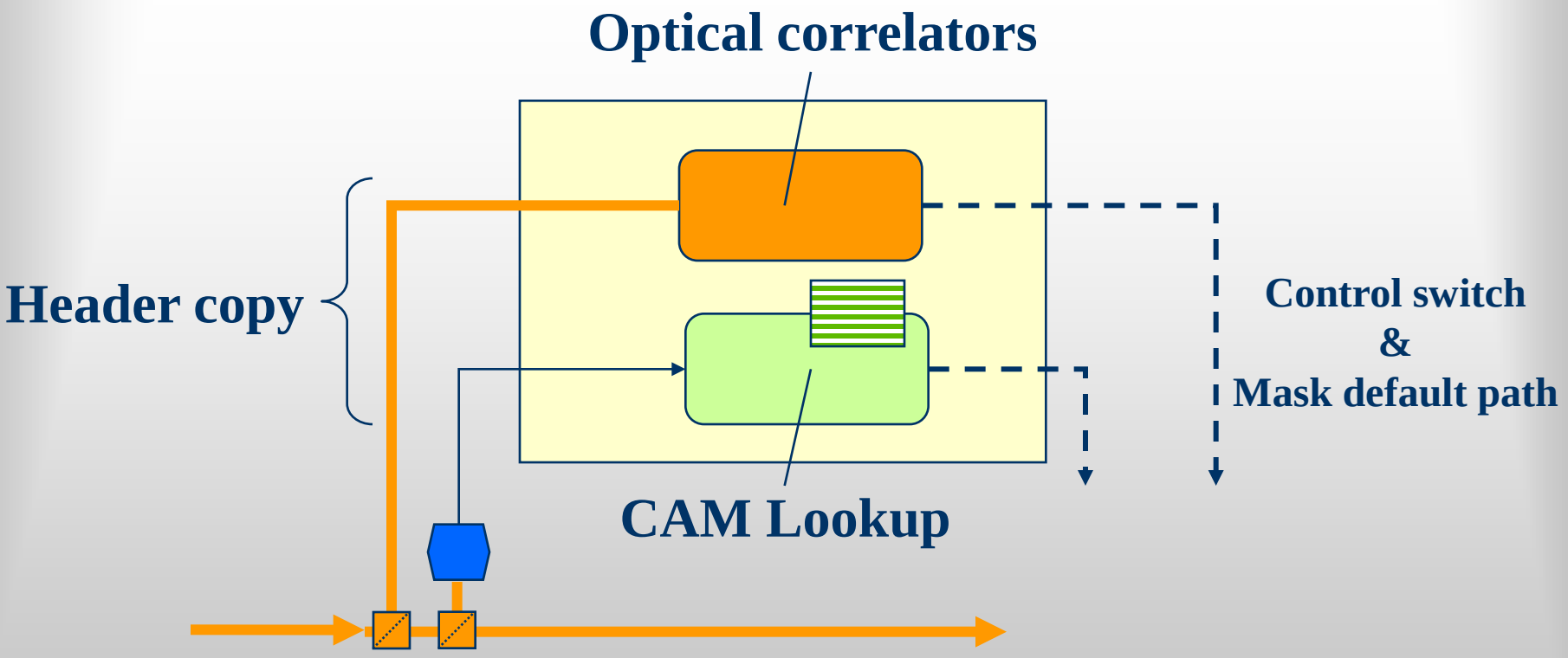


Optical Assist

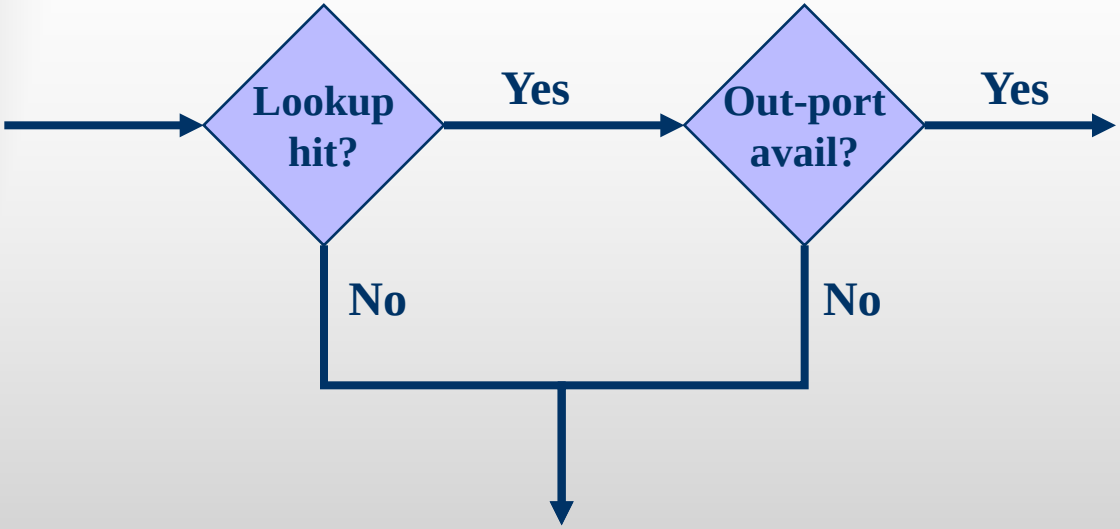
- Optics as ‘fast path’; electronics as backup



Bypass Design

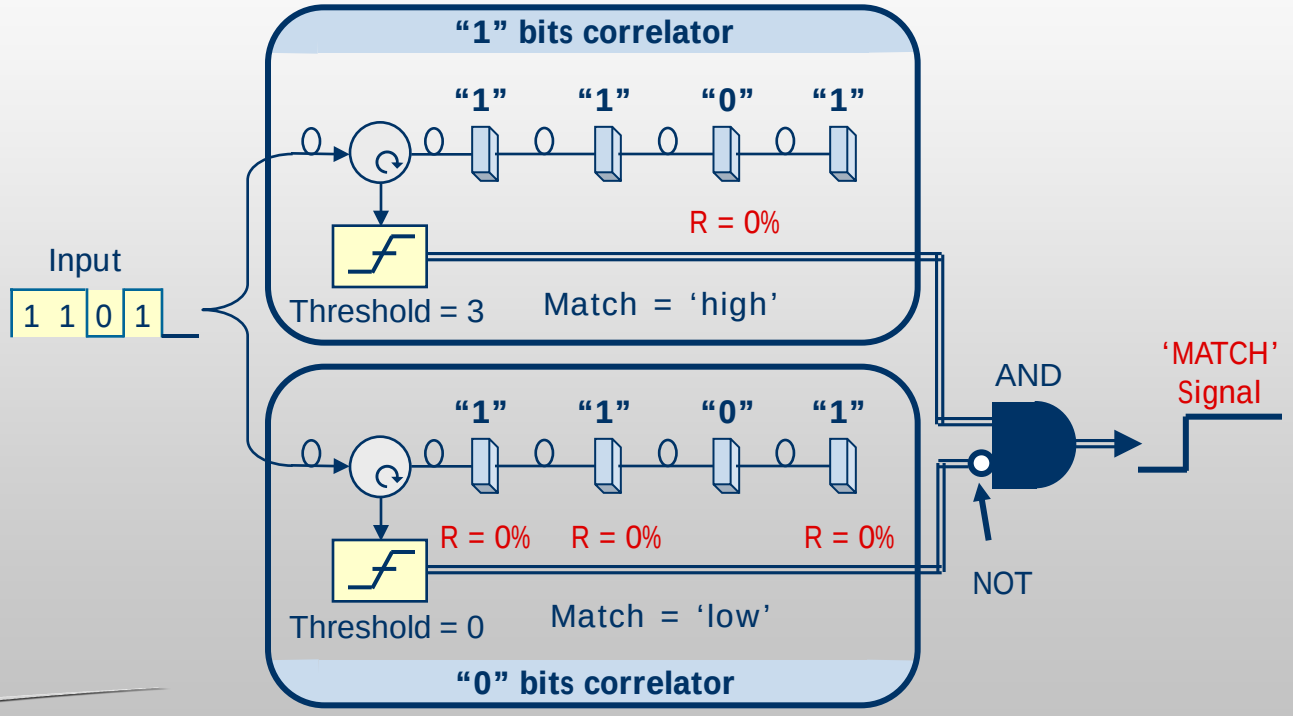


Bypass decision tree



Forward via Filters

- **Bit-subset groups share next-hops**
- **Remainder to helper router**

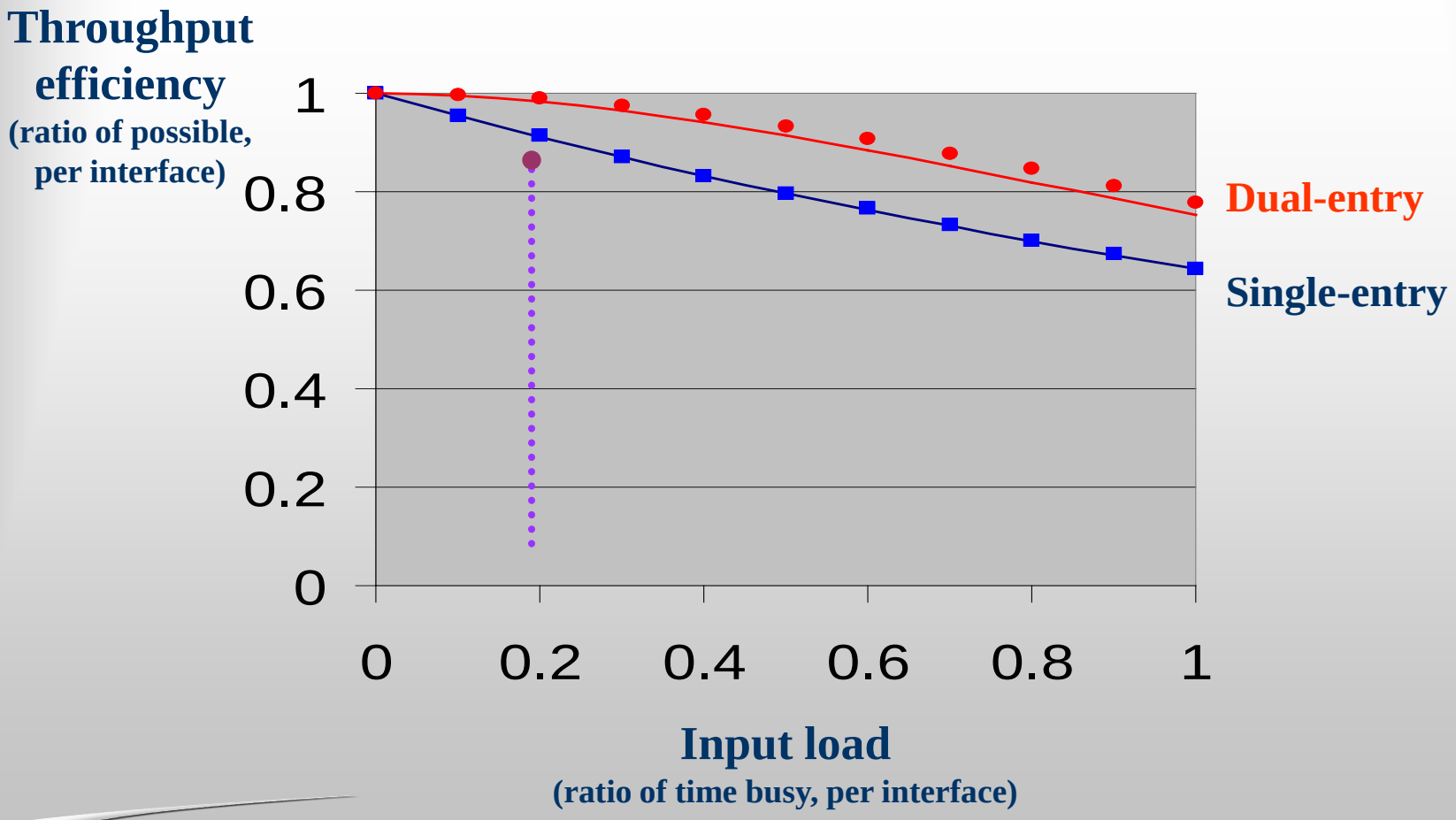


Some limits

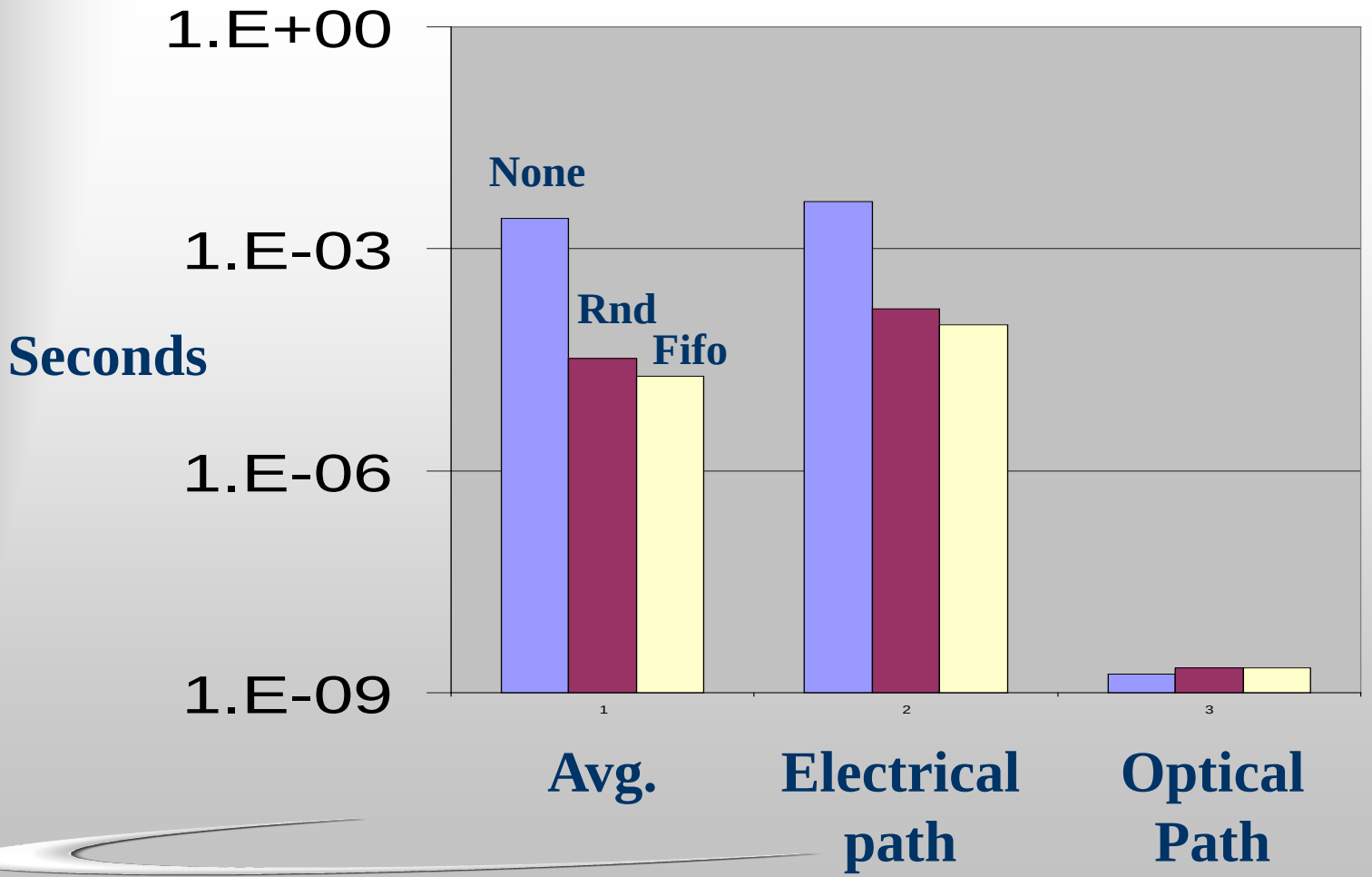
- Correlators
  - ~8 correlators
  - ~8 bits per correlator
- CAM
  - 1-8K entries, 1ns lookup (7x pipeline)
- Switch
  - $LiNbO_3$ at 1ns currently (parallelize?)

How useful?

Switching gain



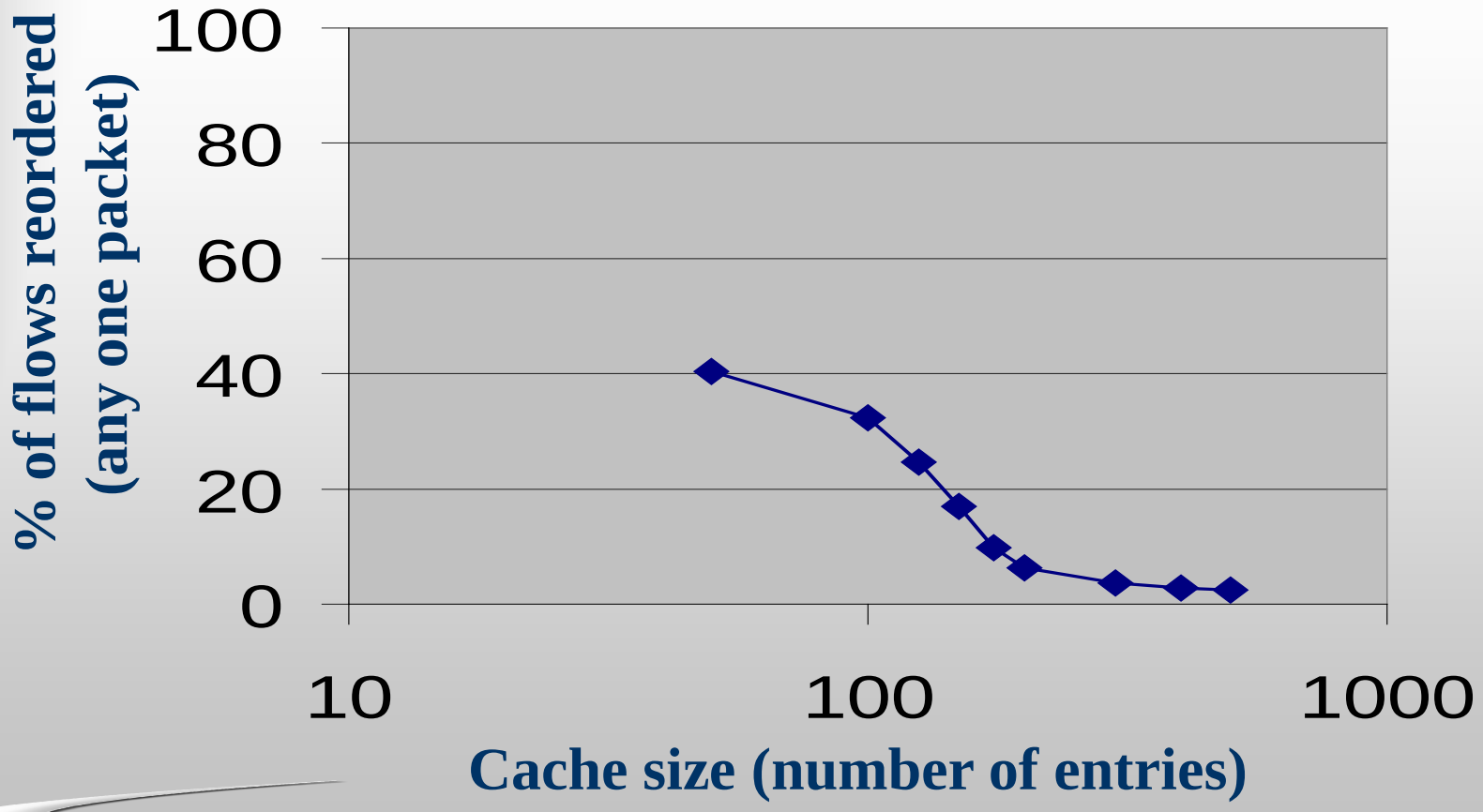
Latency


1.E+00
1.E-03
1.E-06
1.E-09
Seconds
None
Rnd
Fifo
1
2
3
Avg.
Electrical path
Optical Path

Cost

- Packet reordering
  - Two separate paths,
    each path non-reordering,
    but together can reorder
- HOL blocking
  - Optical has priority on outputs
  - Use second-best paths

Reordering


% of flows reordered (any one packet)
100
80
60
40
20
0
10
100
1000
Cache size (number of entries)

Partition function

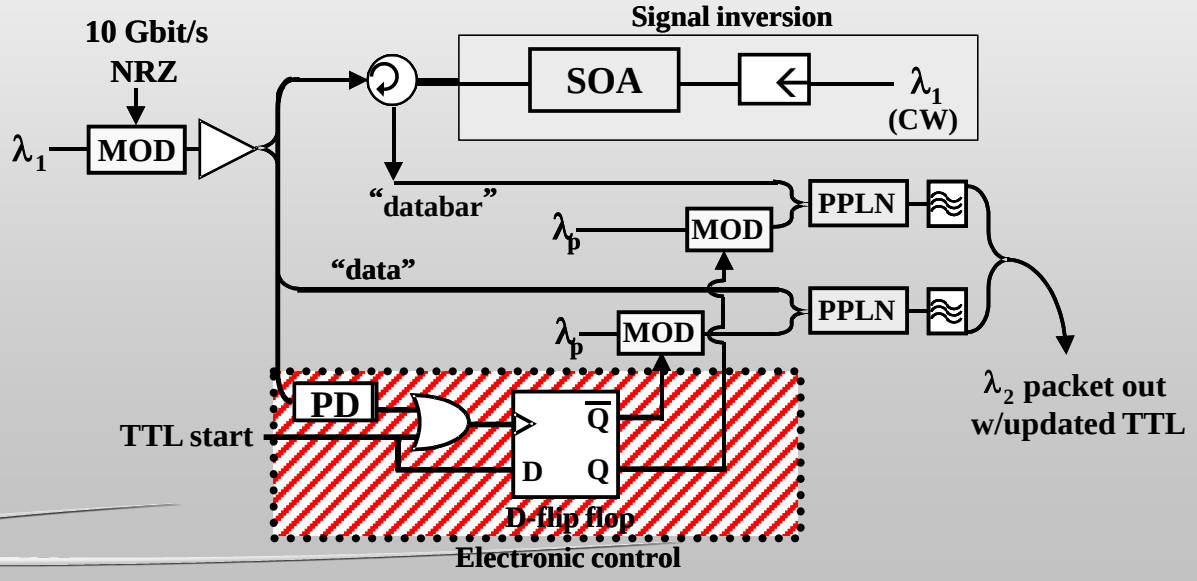
- **Take advantage of optics**
  - Easier to match 0's than 1's
  - Hard to match all 24 bits, easy to pick movable subset
- **Incomplete function is OK**
  - Avoid false positives
  - False negatives just reduce efficiency

TTL Decrement

- Unsigned, 8-bit field
  - Decrement by 1 each IP hop
  - Drop if zero before decrement
- Current design:
  - Arithmetic subtract-by-1

Optical Decrementer

- **LSB-first:**
  - Invert until 1
  - Stop @ 1st "1 (delete if no "1")

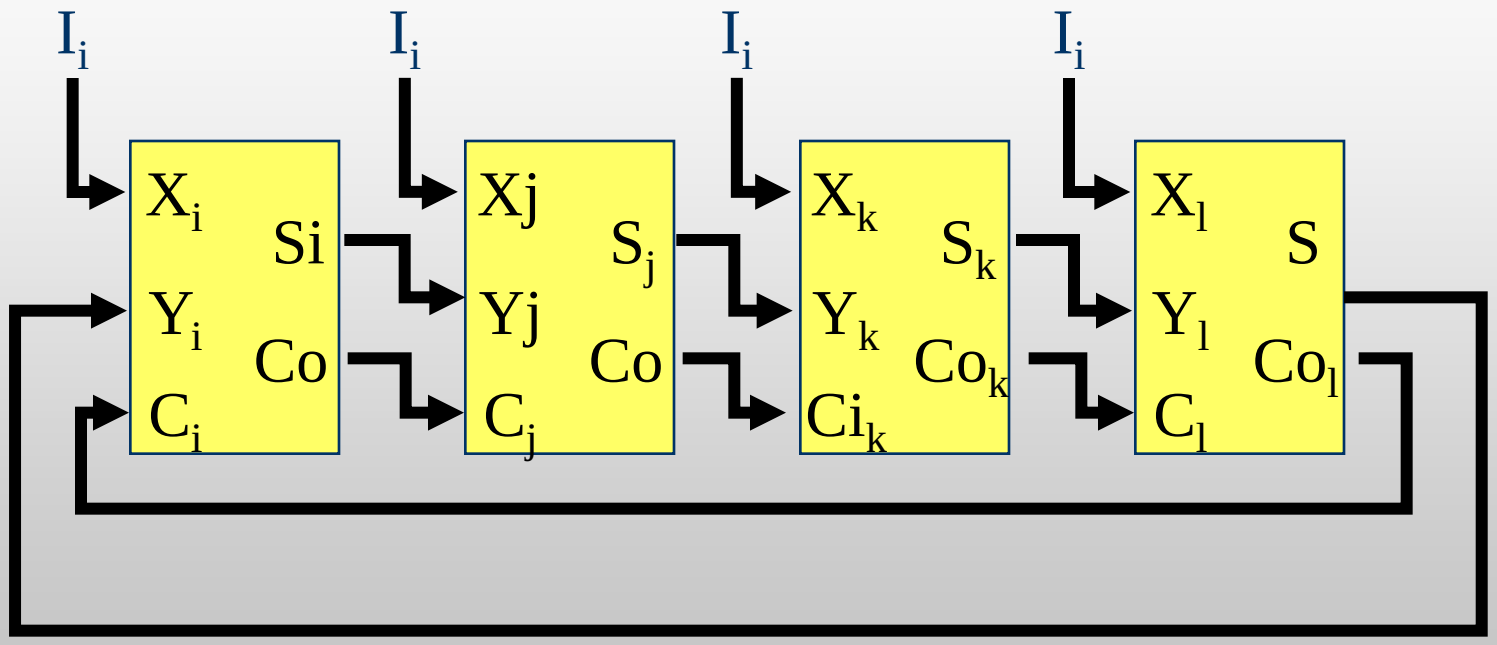


Internet Checksum

- 16-bit, 1's complement sum
  - In 2's complement sum
  - Add carry back in
  - Can be done in words, doubles, etc. with a folded result...
- Current electronic hardware:
  - 2's complement accumulators
  - Groups of full-adds; carries wired in a loop

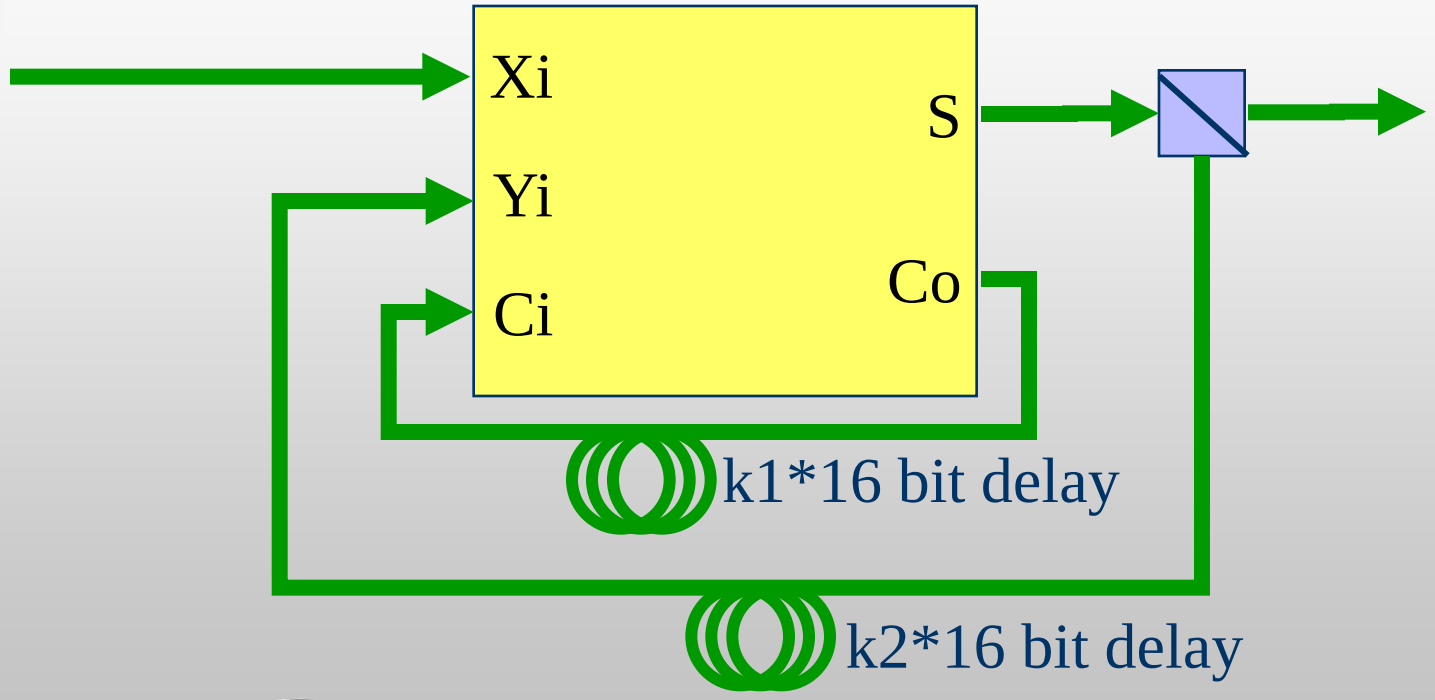
Parallelized Checksum

- Wire carries in a loop



Optical Checksum

- **Serial 1-bit full-adder**



www.isi.edu/pow
www.isi.edu/ocdma

- ISI
  - Joe Bannister, Joe Touch
  - GRAs: Purushotham Kamath,
  - GRA alumni: Stephen Suryaputra, Aatash Patel
- USC:
  - Alan Willner (advisor)
  - Michelle Hauer (LUT), Deniz Gurkin (CSum), John McGeehan (TTL)
- Papers:
  - OCDMA Collision Sense – ICC 2004 (to appear)
  - OCDMA MACs – Infocom 2004.
  - Correlator lookups – JLT 2003 / OFC 2002.
  - Optical TTL – JLT 2003 / OFC 2003.
  - Limits of WDM – Optical Nets 2000.
  - OCDMA via 2D codes – OFC 2003.
  - Booster routers – HPC 2001.
  - SWAP protocol – Photonics East 1999.