Optical Networks

Manya Ghobadi
ghobadi@csail.mit.edu

Some slides are borrowed from:

Richard A. Steenbergen [NANOG'17]

Danyang Zhuo [SIGCOMM'17]

Mark Filer [OFC'17]

Why should we care about optics?

The Internet is largely based around optics

- 100s millions of dollars
- 100,000s miles of fiber
- 100s of Tbps capacity





Two million miles of optical fiber

Why should we care about optics?



Data centers



The Internet

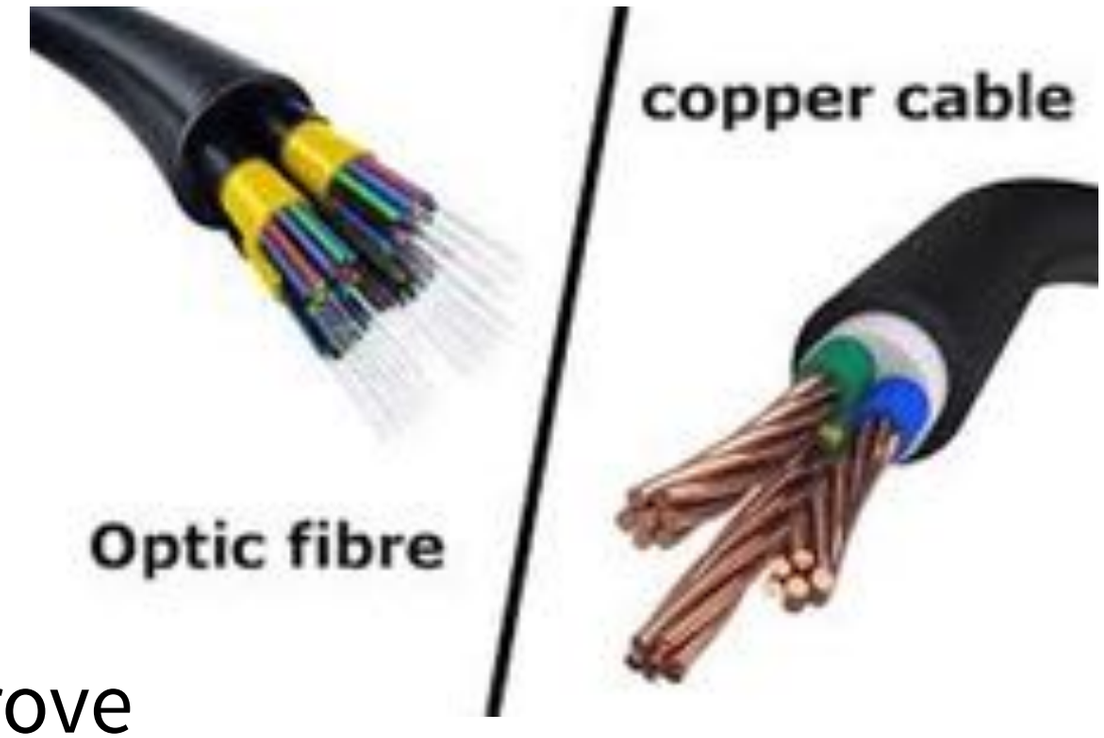


The basics of fiber optic transmission



What is fiber and why do we use it?

- Fiber is ultimately just a “waveguide for light”
- Benefits compared to copper:
  - Low-cost
  - Light
  - High bandwidth
  - Multiple wavelengths
  - Technology continues to improve

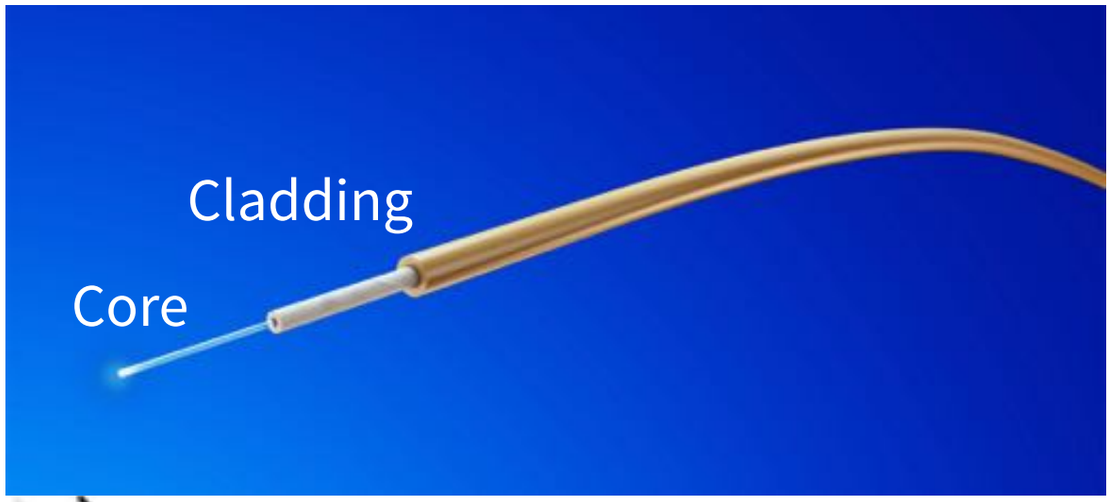


A quick flash back to high school physics

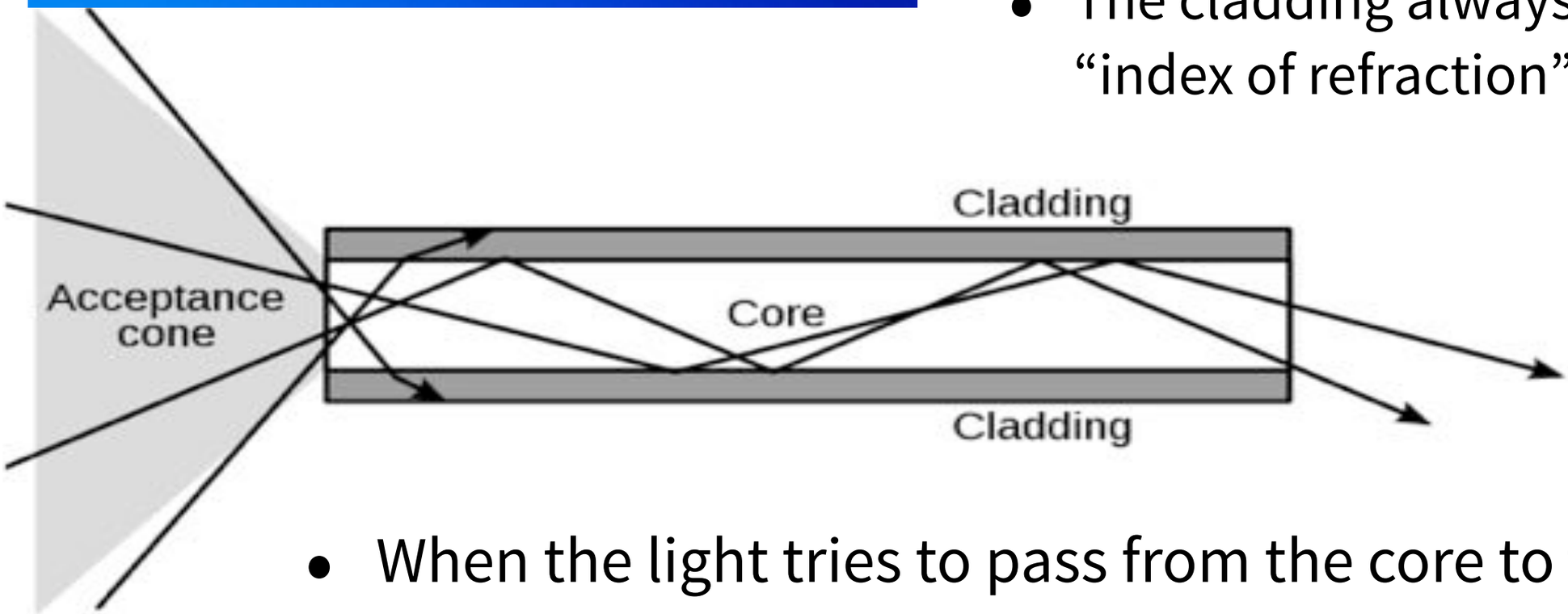
- Speed of light, “*c*”, in vacuum?
  - 300,000 km/sec
- What happens when light passes through materials that aren’t a perfect vacuum?
  - It propagates slower than *c*
- Refractive index: the speed of light in other material
  - Water has a refractive index of “1.33”, or 1.33x slower than *c*
- When light tries to pass from one medium to another with a different index of refraction, a reflection can occur instead

Slide credit: Richard A Steenbergen

Fiber works by “total internal reflection”

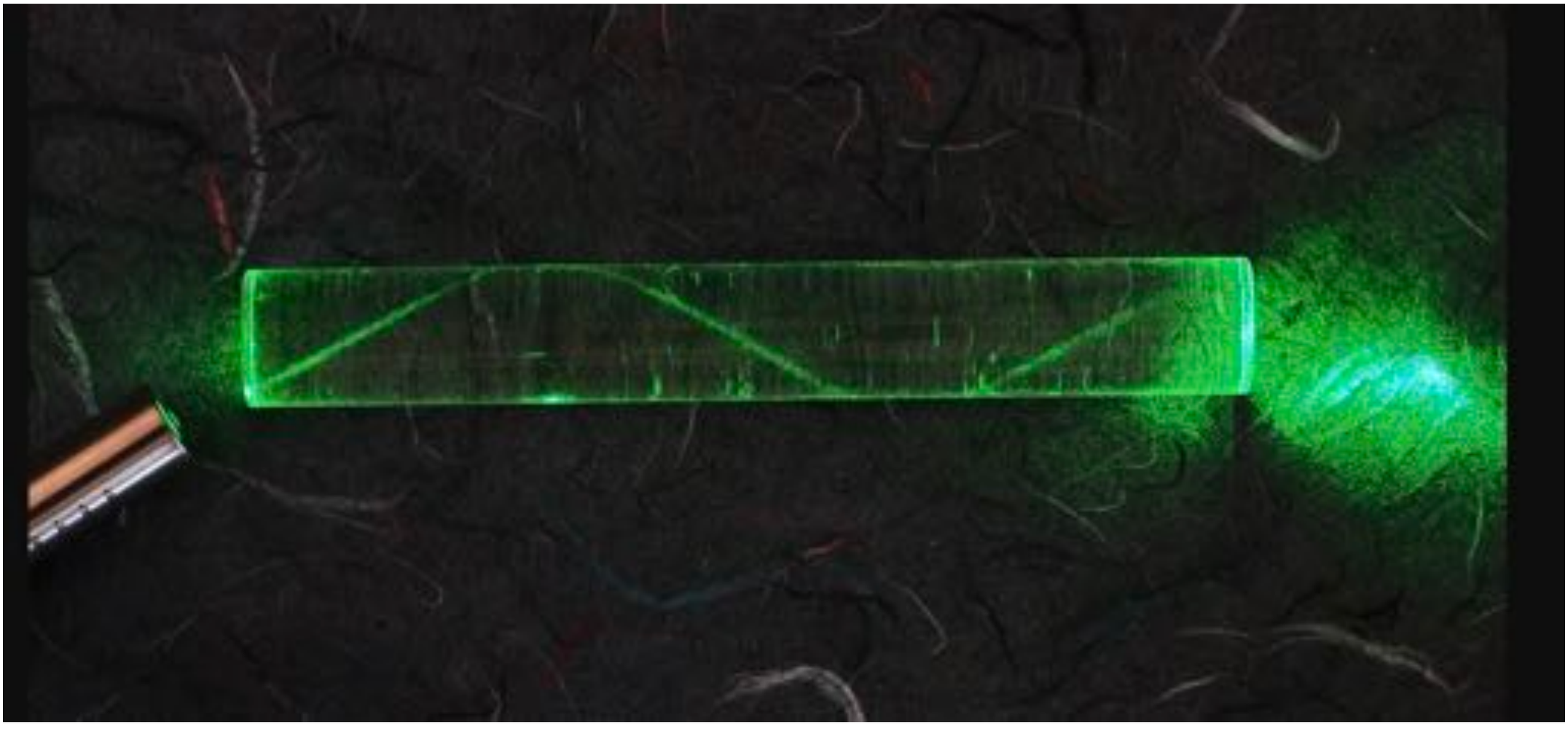


- Fiber optic cables are internally composed of two layers
  - A “core” surrounded by a different material known as the “cladding”
  - The cladding always has a higher “index of refraction” than the core



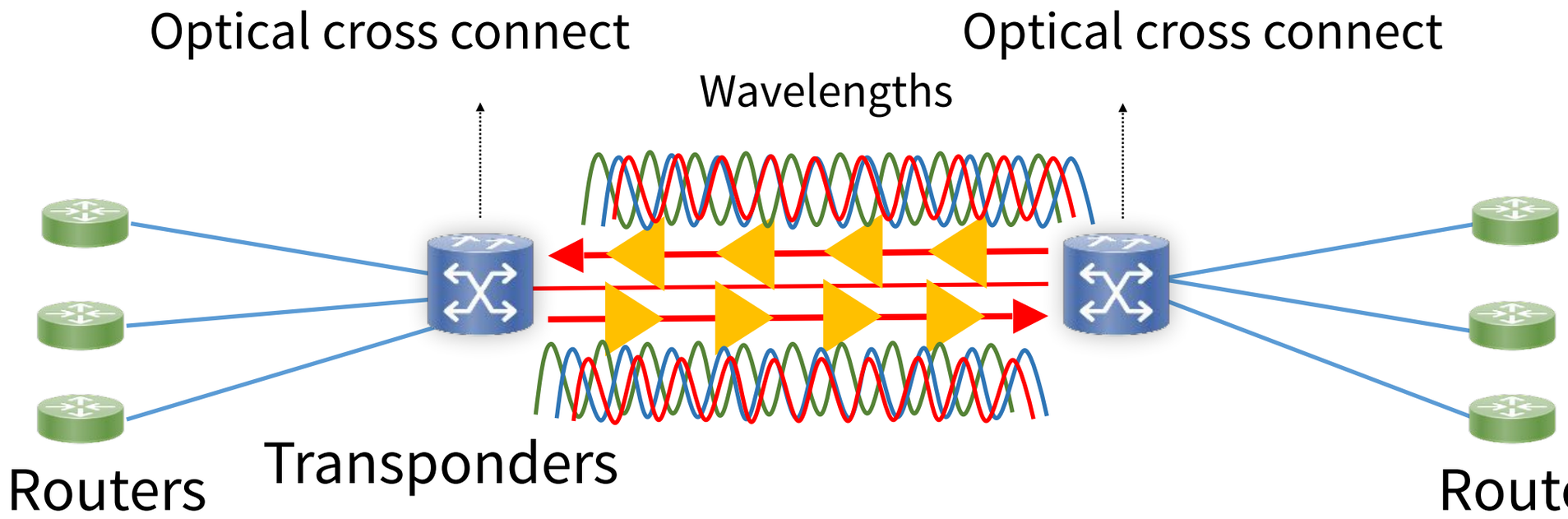
- When the light tries to pass from the core to the cladding, it is reflected back into the core.

Slide credit: Richard A Steenbergen



Source: https://en.wikipedia.org/wiki/Optical_fiber

How do we actually use the fiber?



- One strand of fiber is used to transmit signal, the other to receive one
- Incoming IP traffic is multiplexed into one or more optical wavelengths
- This results in simplest and cheapest components
- But fiber is perfectly capable of carrying many signals, in both directions, over a single strand

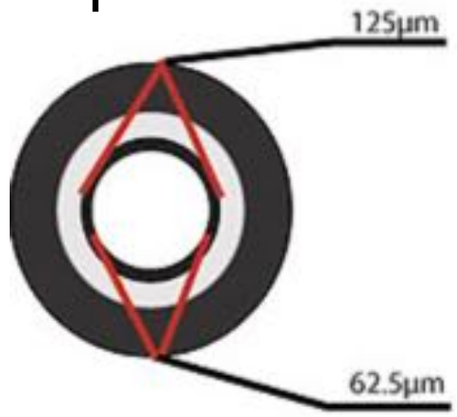
Distinction in Fiber:

Multi-Mode vs Single Mode

Multi-Mode Fiber

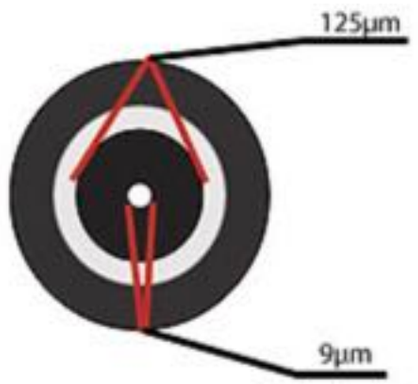


- Wide core allows the use of less precisely focused and cheaper light sources
- Short distance: 10-100s meters
- Types of Multi-Mode Fiber
  - OM1/OM2
  - OM3/OM4
    - Specifically designed for modern 850nm short reach laser sources.



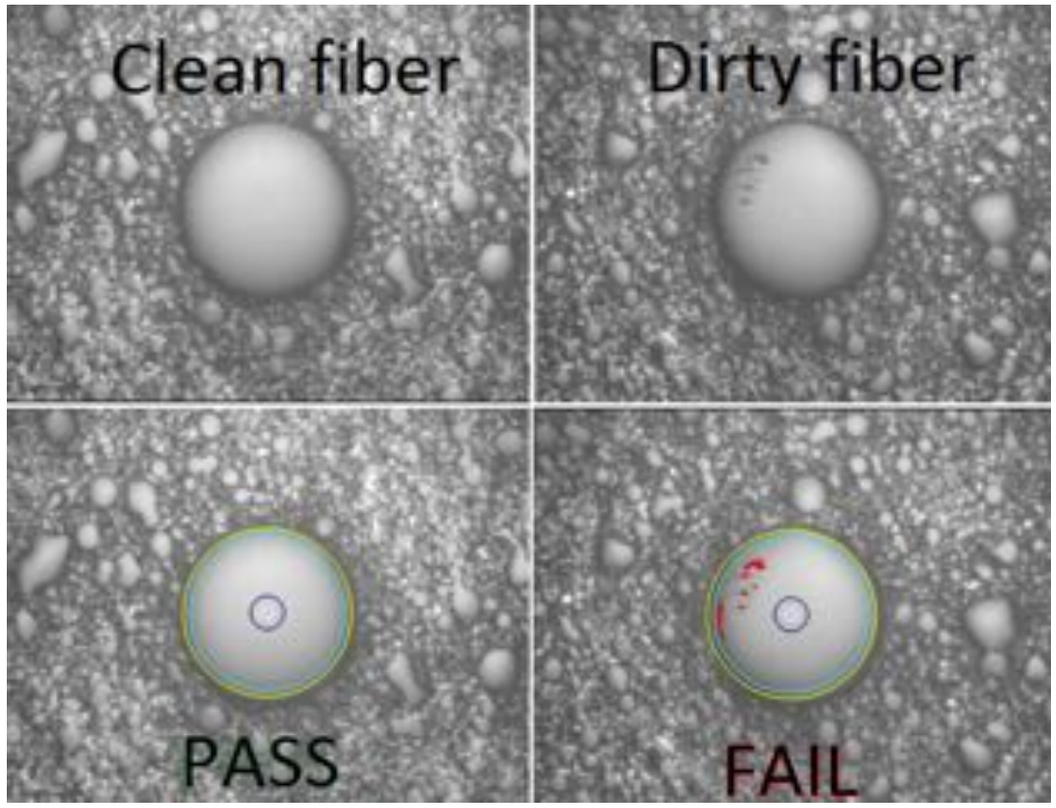
Single Mode Fiber



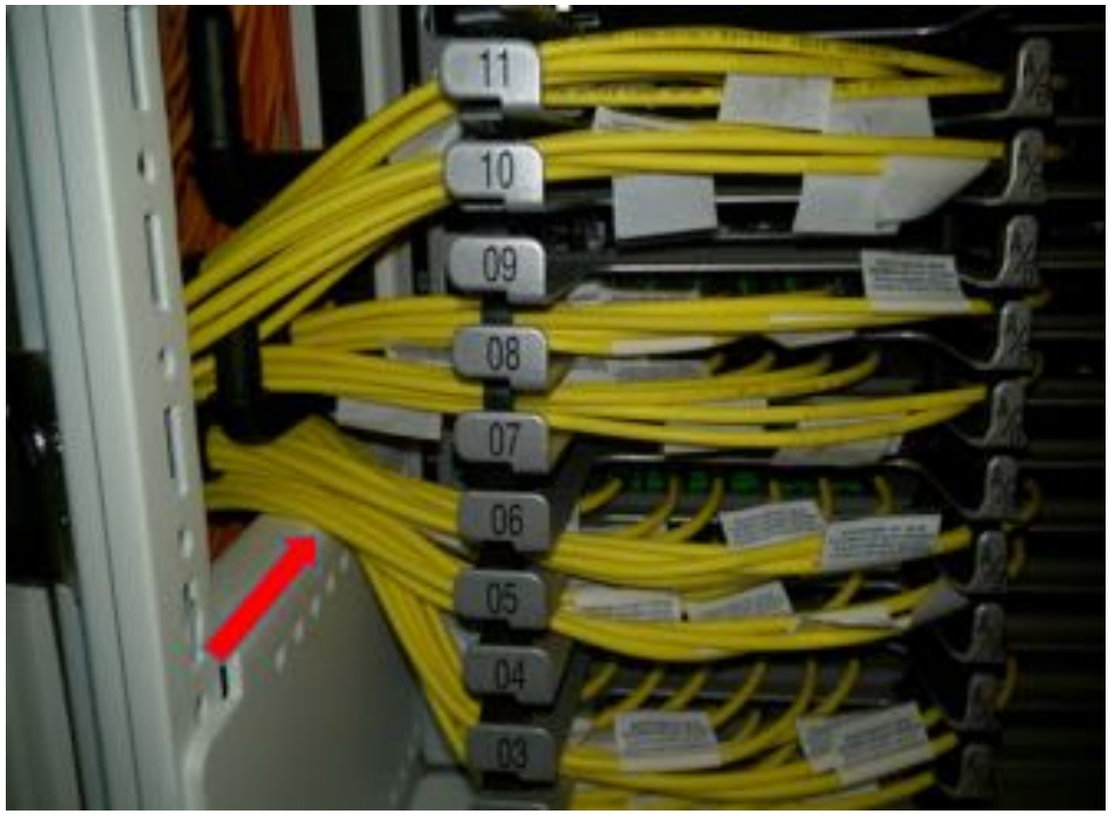


- The fiber used for high bandwidths, and long distances
  - Has a much smaller core size, between 8-10 µm
  - Typically supports distances of 80km (50 miles) without amplification
  - With amplification, can transmit a signal several thousand km
- “Classic” SMF can be called “SMF-28” (a Corning product name)

Slide credit: Richard A Steenbergen

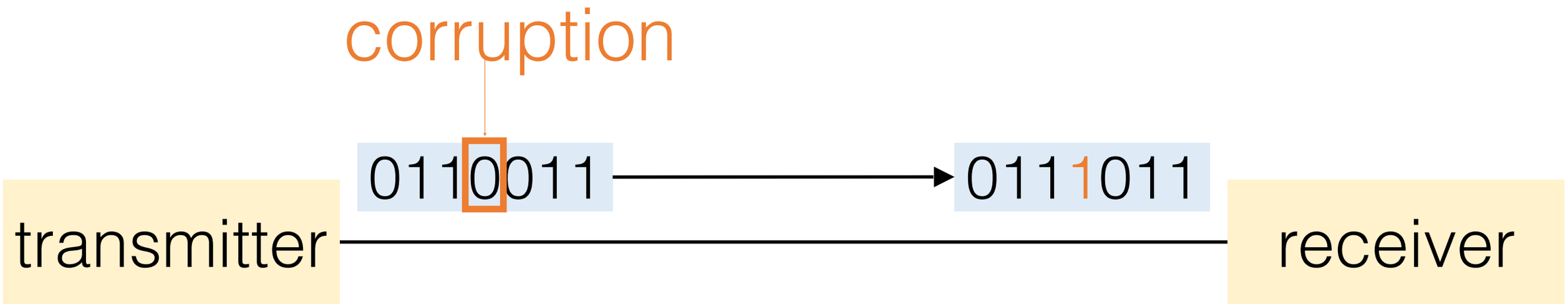


dirty optical connector



bent fiber

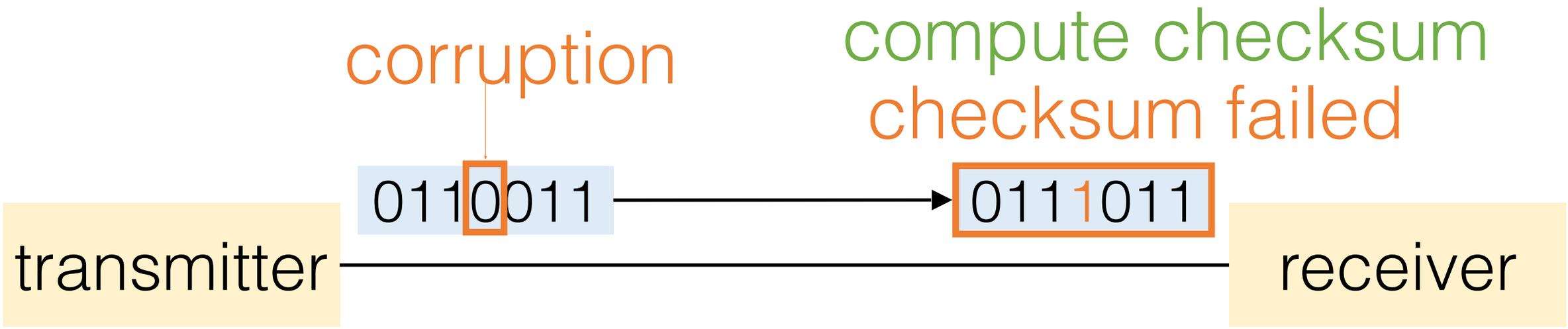
Packet Corruption



Understanding and Mitigating Packet Corruption in Data Center Networks
Zhuo et al. [SIGCOMM'17]

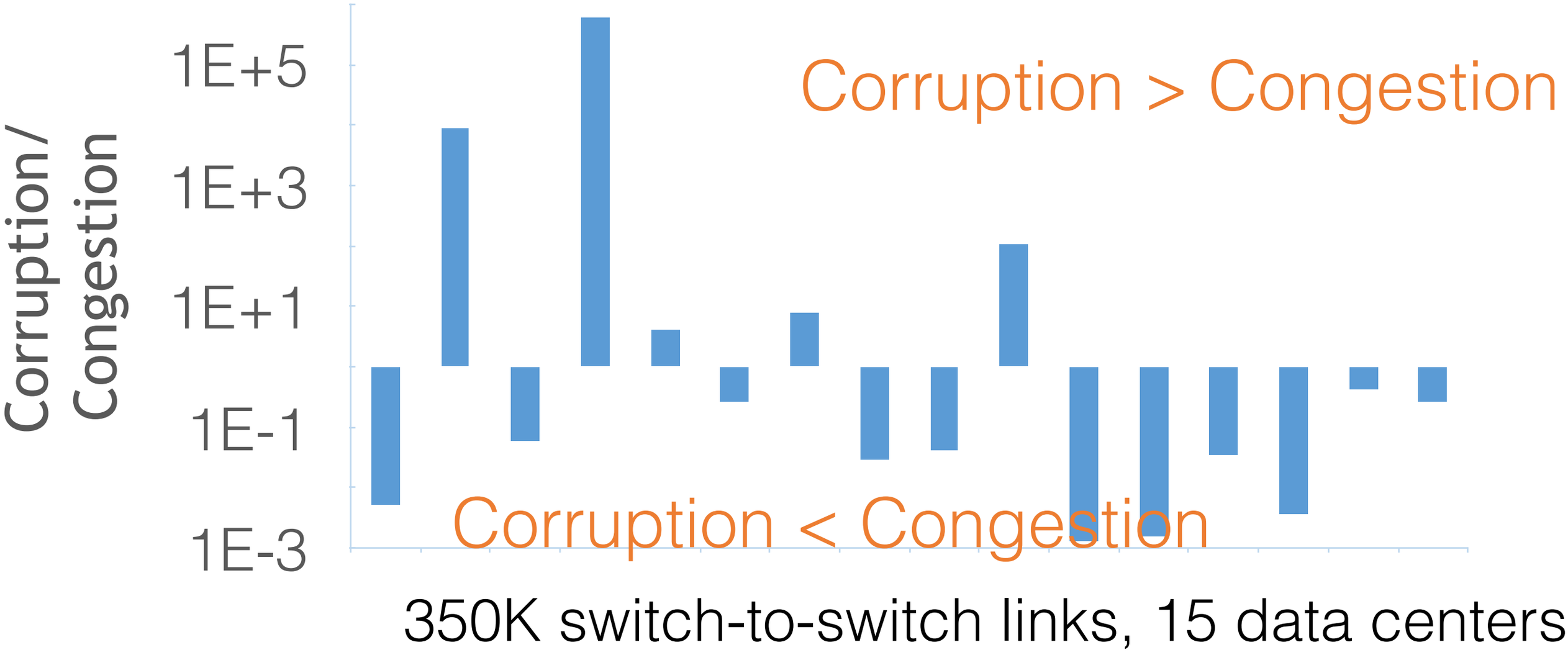
Slide credit: Danyang Zhuo

Packet Corruption



Slide credit: Danyang Zhuo

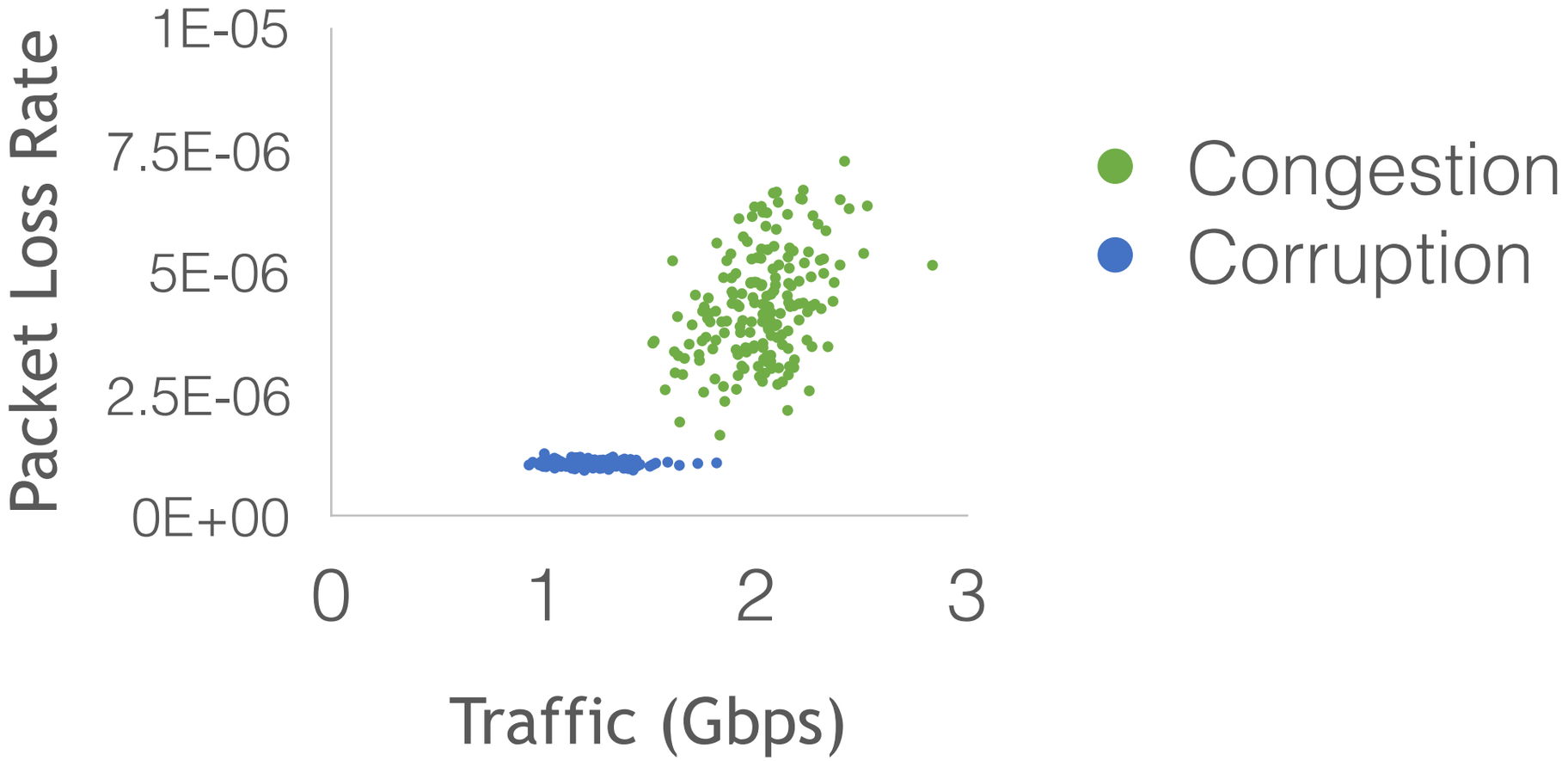
Packet Corruption is Significant



Slide credit: Danyang Zhuo

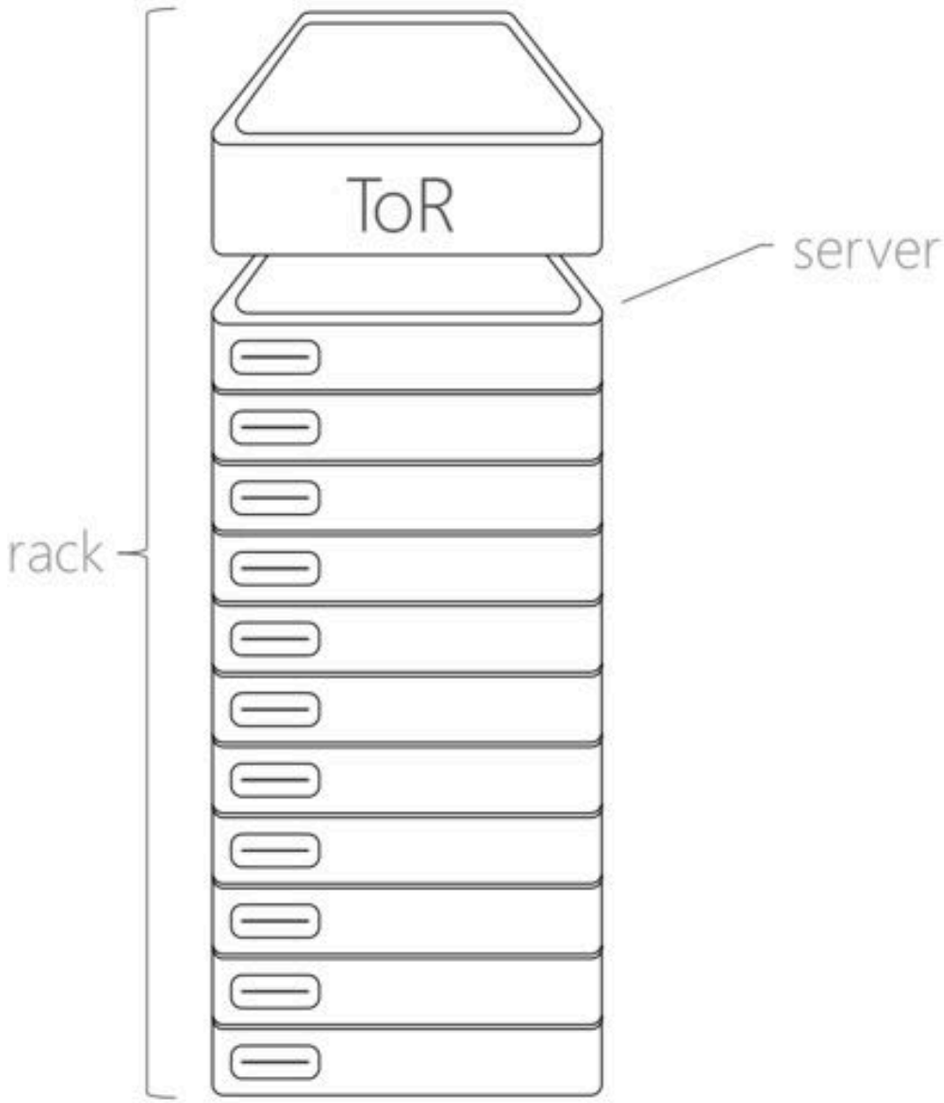
Quantifying corruption rate is easy

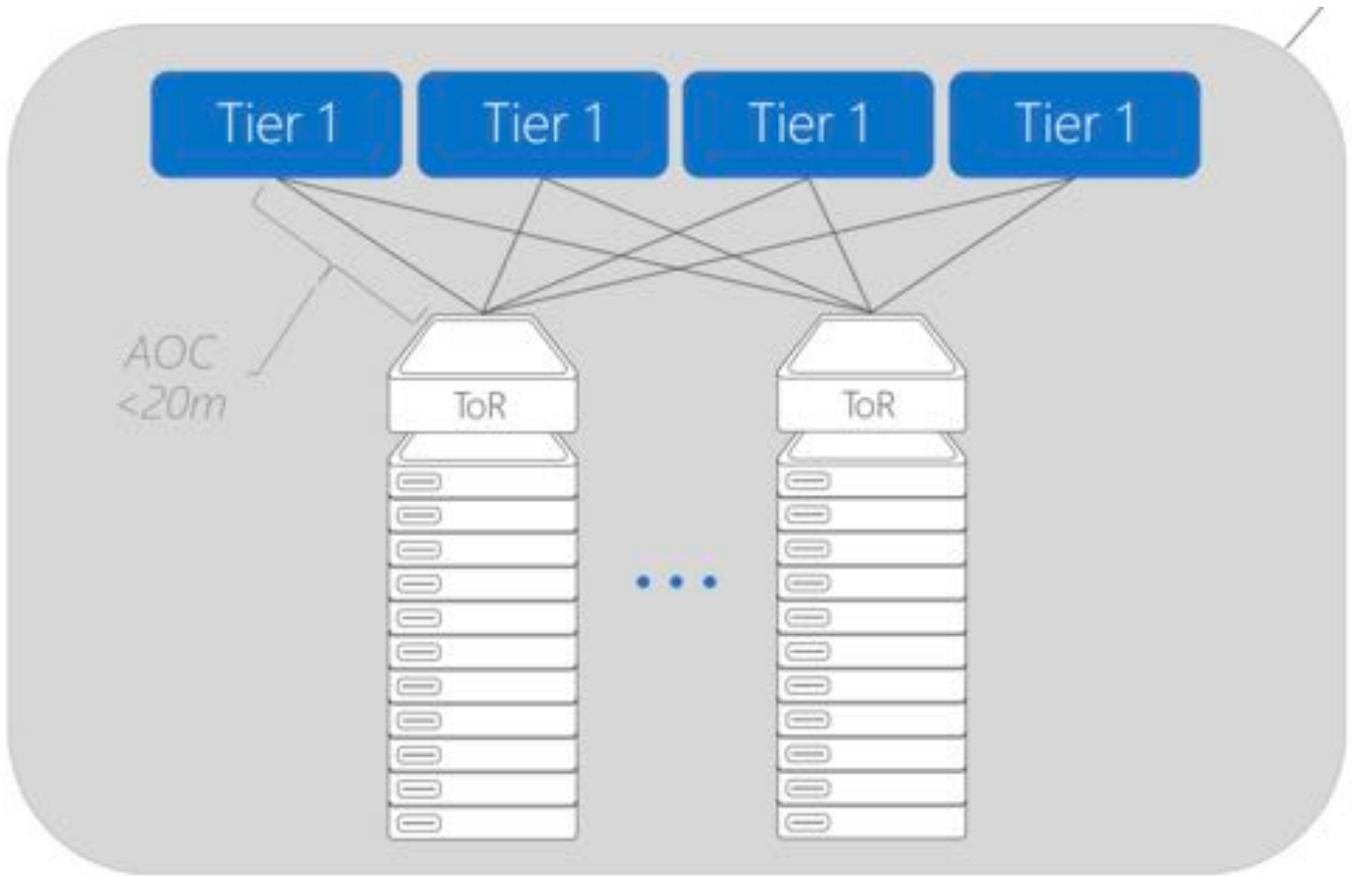
Corruption vs. congestion



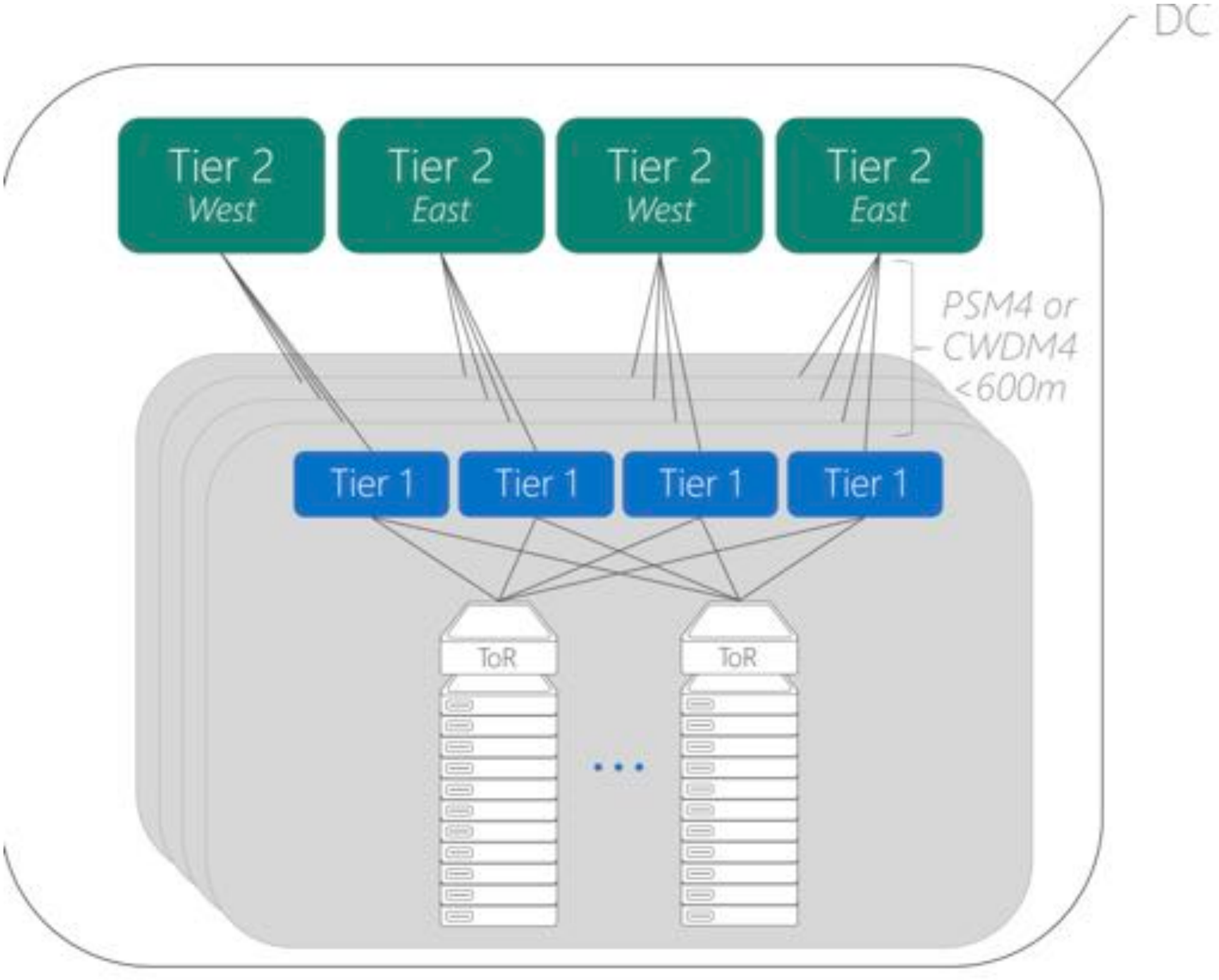
Slide credit: Danyang Zhuo

The pyramid of cabling



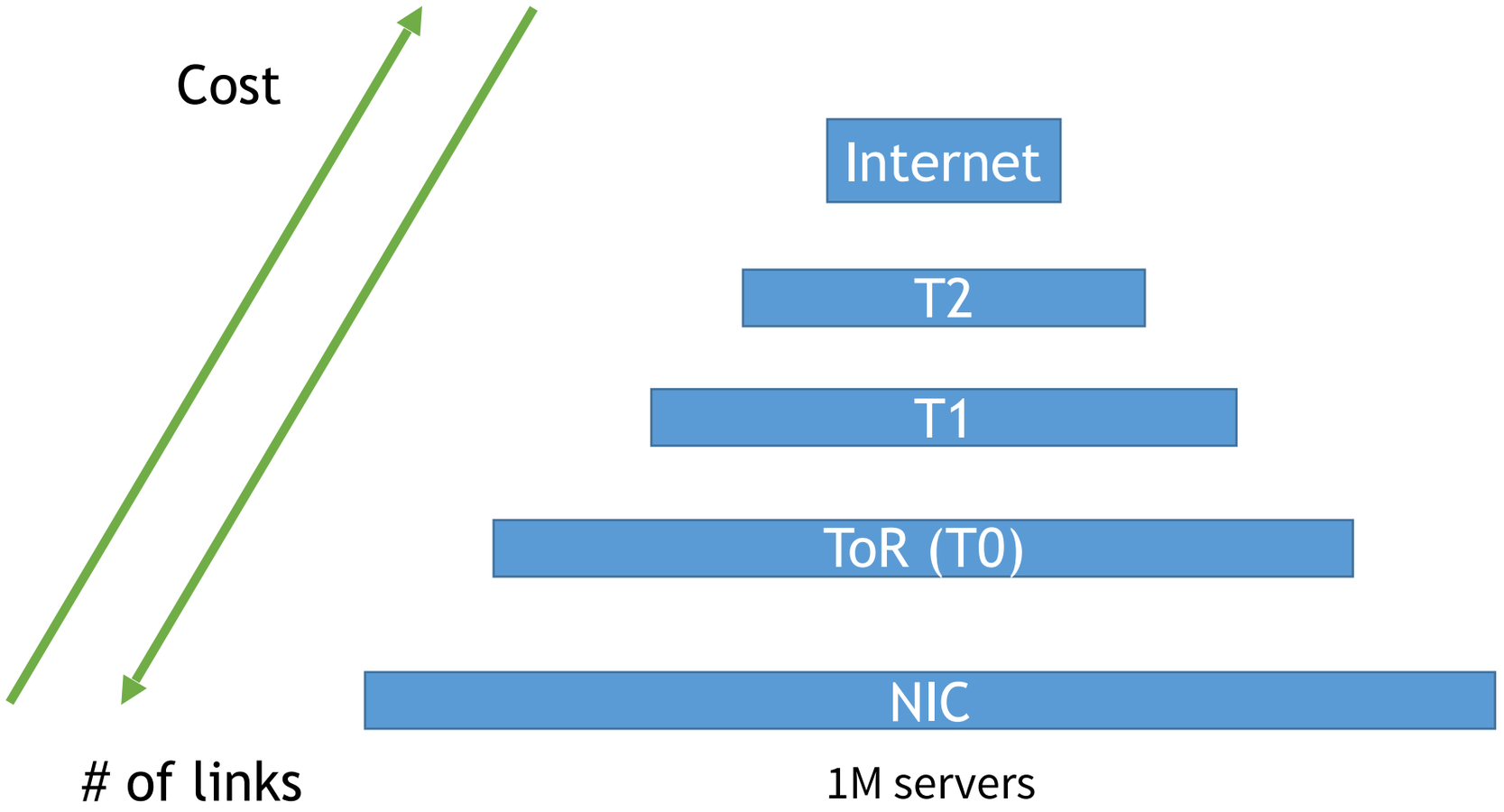


Slide credit: Mark Filer



Slide credit: Mark Filer

The pyramid of cabling



How fast does light travel in fiber?

- What is the “theoretical” RTT from Boston to LA?
- The speed of light is 299,792,458 m/sec
- SMF28 core has a refractive index of 1.4679
- Speed of light / 1.4679 = 204,232,207 m/sec
    - 204.2 km/ms
    - Cut that in half to account for round-trip times.
    - Approximately 1ms per 100km (or 62.5 miles) of RTT
    - BOS -> LA: 4800 km (2,982 miles) -> 48 ms RTT (not 4.8)
- Why do we see a much higher value in real life?
    - Fiber is rarely laid in a straight line.

Credit: Richard A Steenbergen

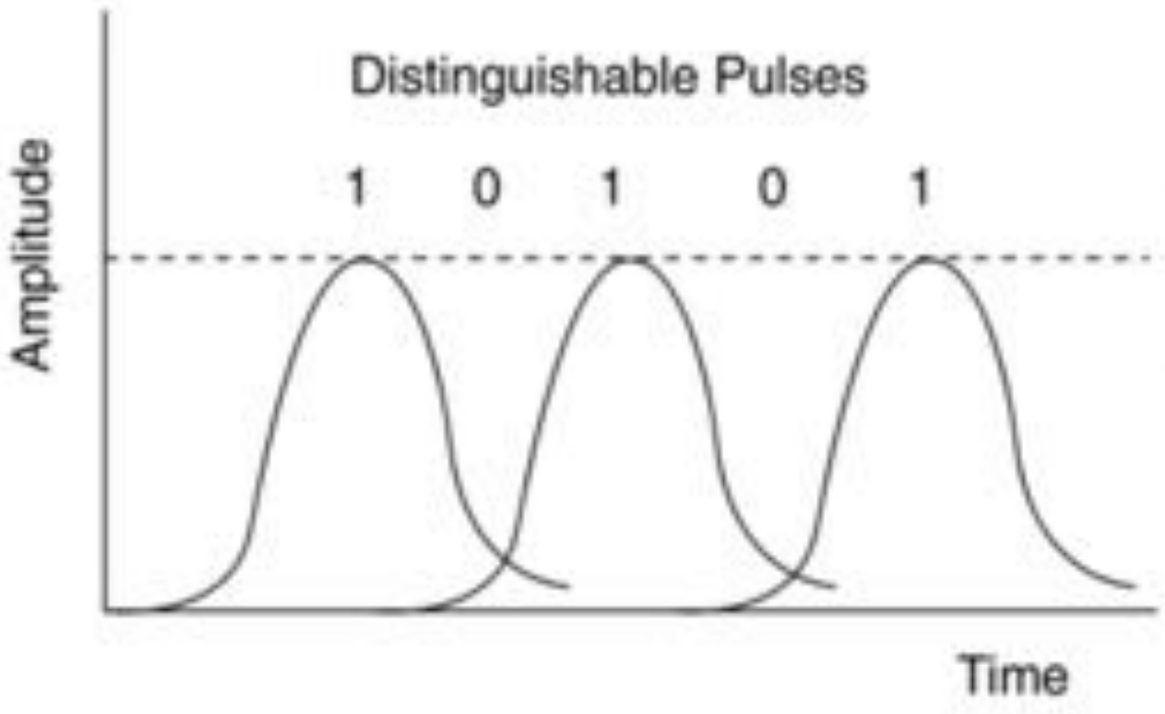


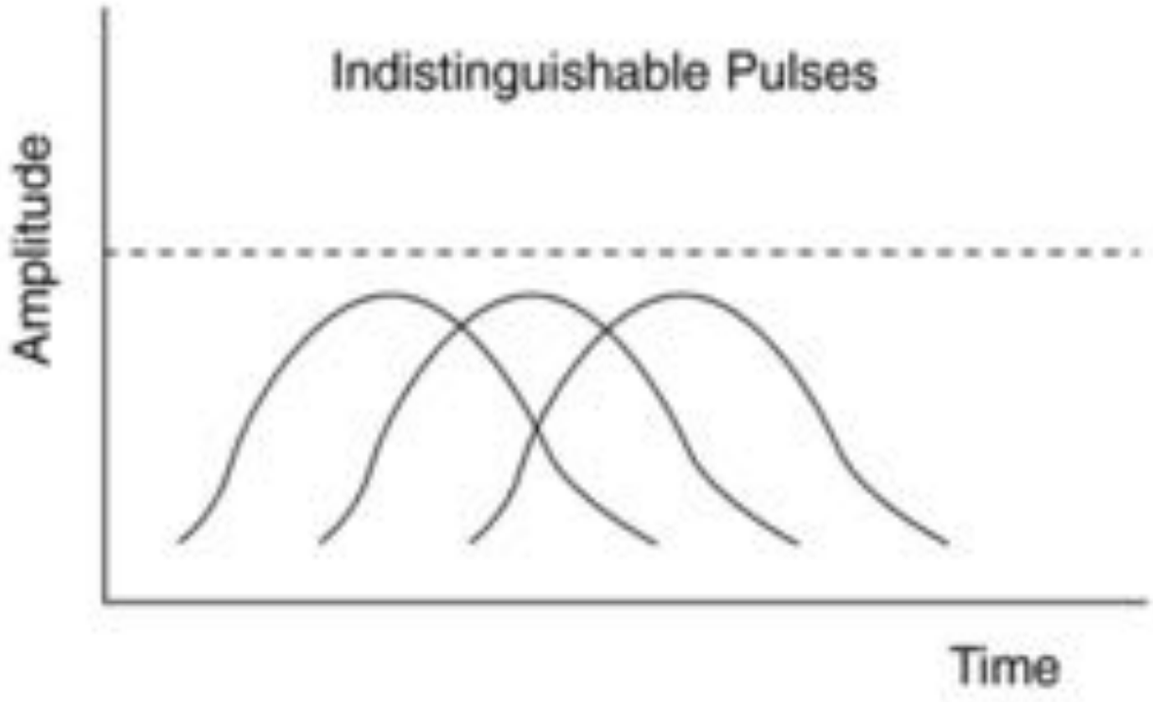
Credit: Level3 website

Basic optical networking terms and concepts

Dispersion

- Dispersion simply means “to spread out”
  - In optical networking, this results in signal degradation
  - As the signal is dispersed, it is no longer distinguishable as individual pulses at the receiver

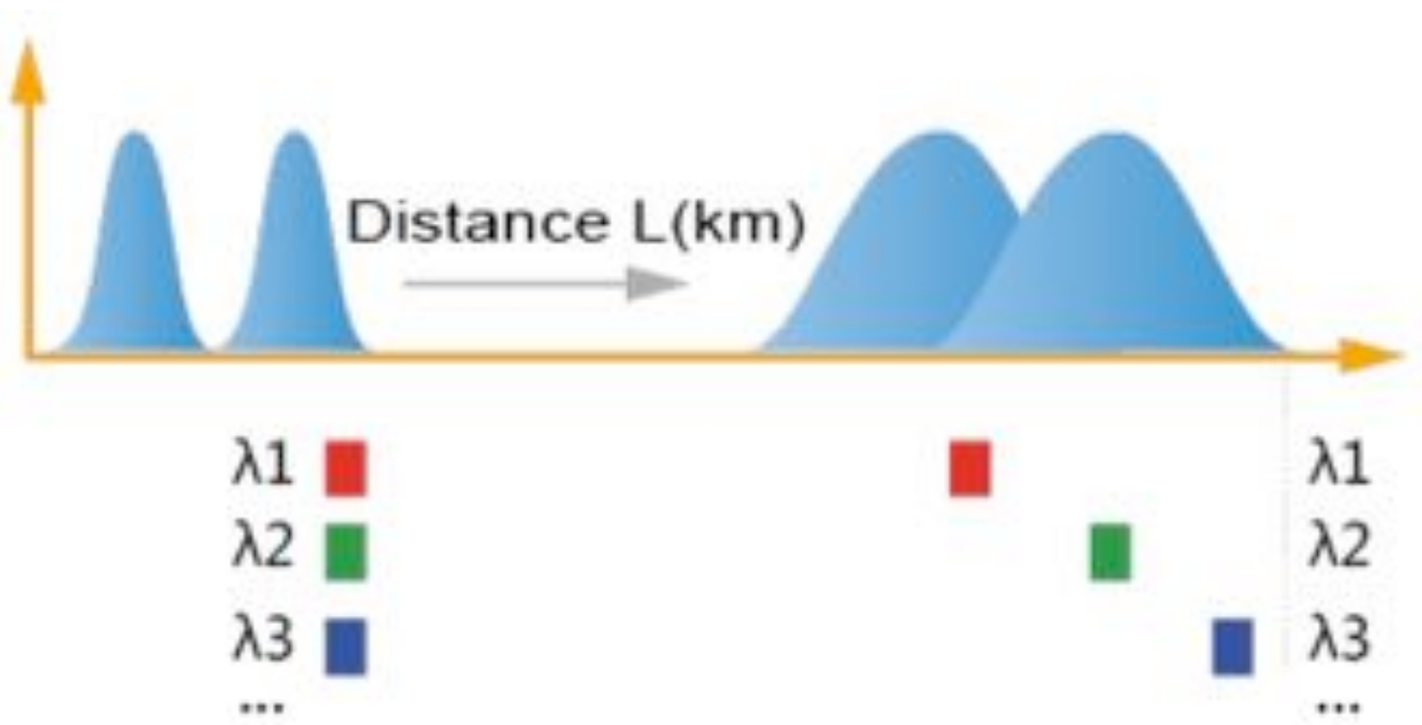




Slide credit: Richard A Steenbergen

Chromatic Dispersion (CMD)

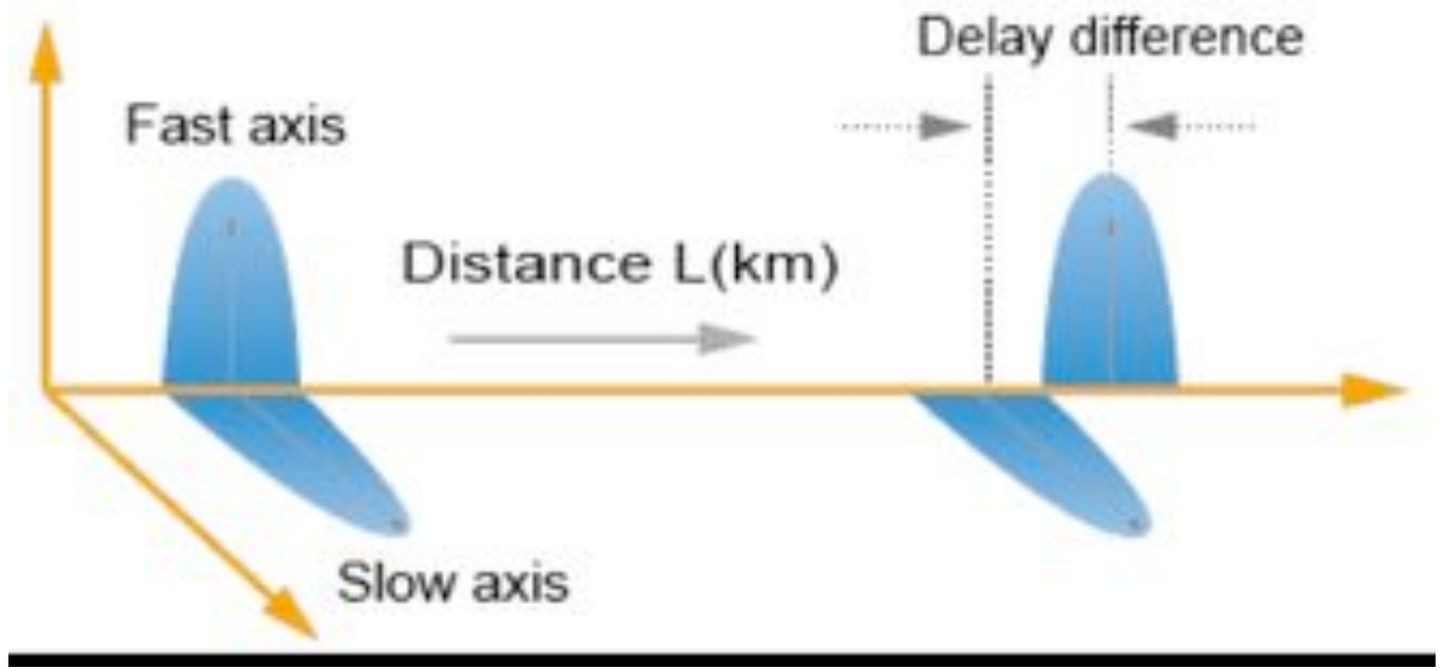
- Different frequencies propagate through a non-vacuum at different speeds. This is how optical prisms work
  - The wider your signal, the more CMD affects it
- Historically, a fundamental limiting factor in optical systems' speed



Slide credit: Richard A Steenbergen

Polarization Mode Dispersion (PMD)

- Not perfectly cylindrical fiber causes one polarization of light to propagate faster than the other
- The difference in arrival time between the polarizations is called “Differential Group Delay” (DGD)
- Makes it hard to recover the signal



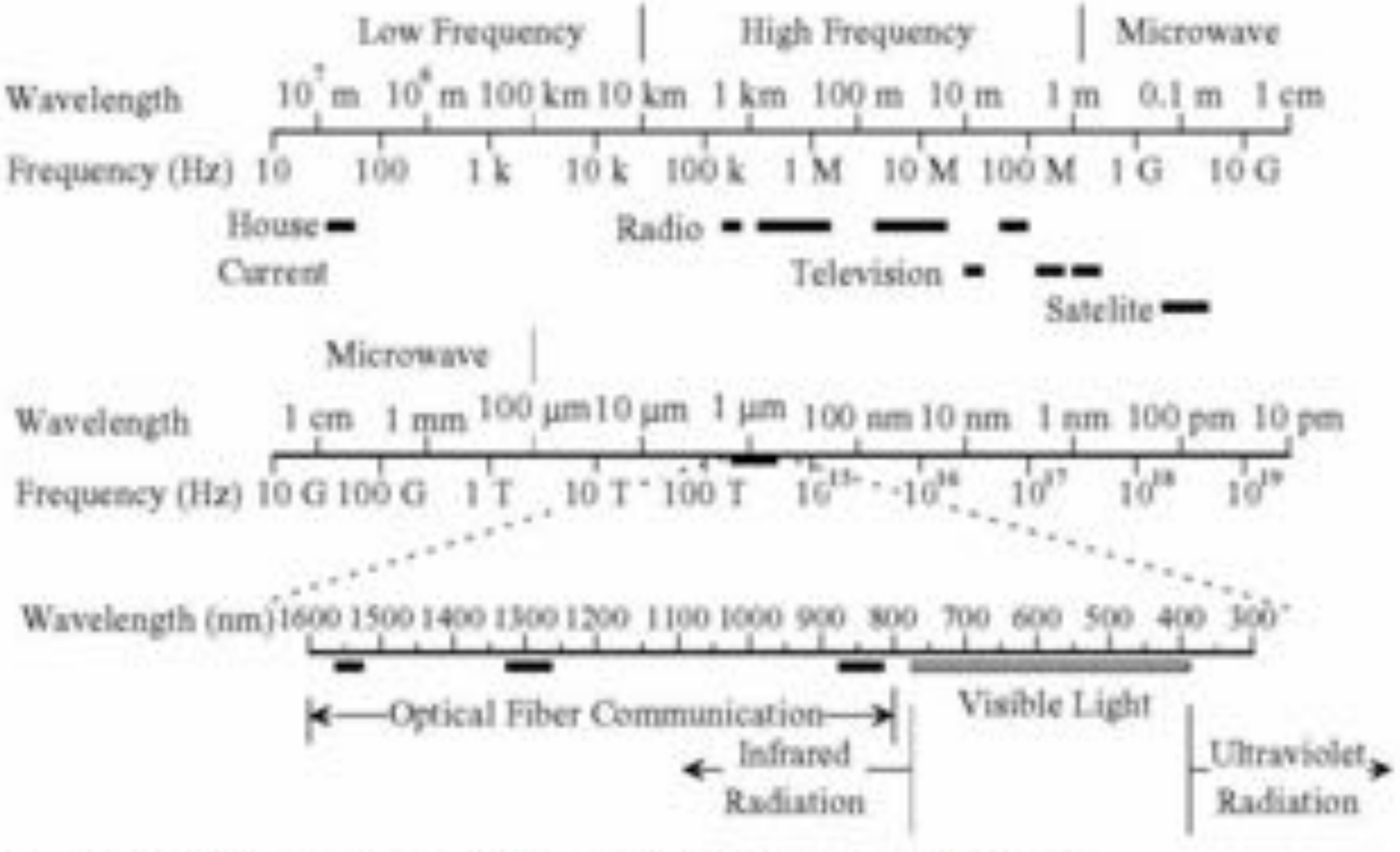
Slide credit: Richard A Steenbergen

Fiber Optic Transmission Bands

- There are several frequency “windows” available
  - 850nm – The First Window
  - 1310nm – The Second Window (O-band)
  - 1550nm – Third Window (C-band)
  - Fourth 1570-1610 nm (L-band)

Slide credit: Richard A Steenbergen

Electromagnetic Spectrum

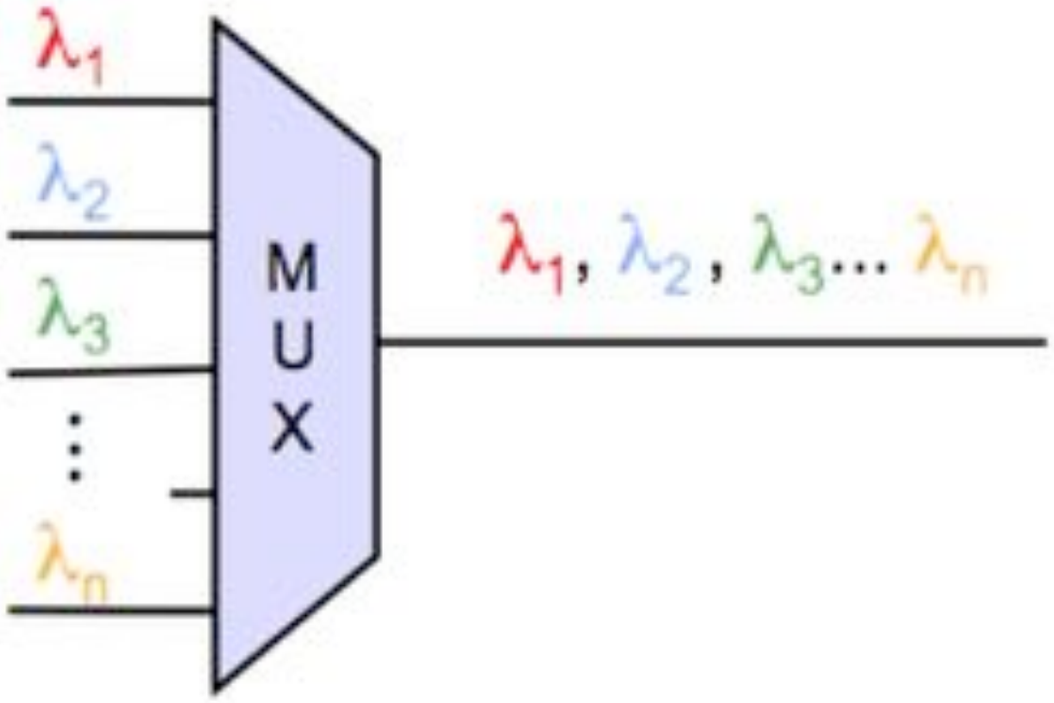


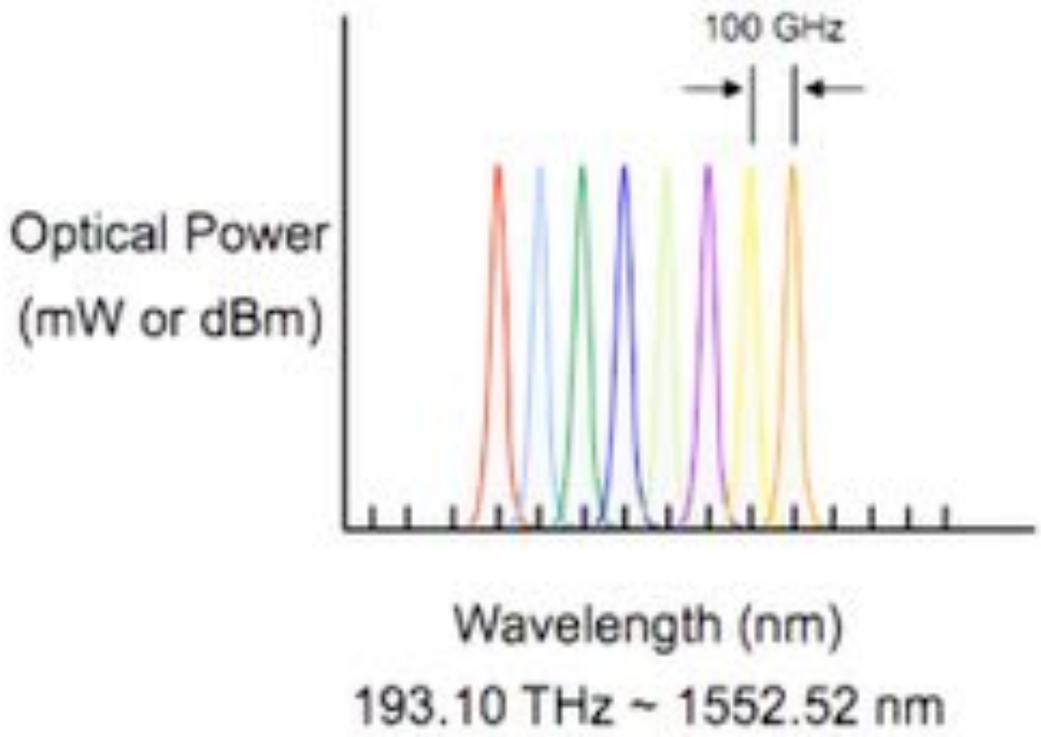
- Infrared light is used for optical communication

Wavelength Division Multiplexing

Wavelength Division Multiplexing (WDM)

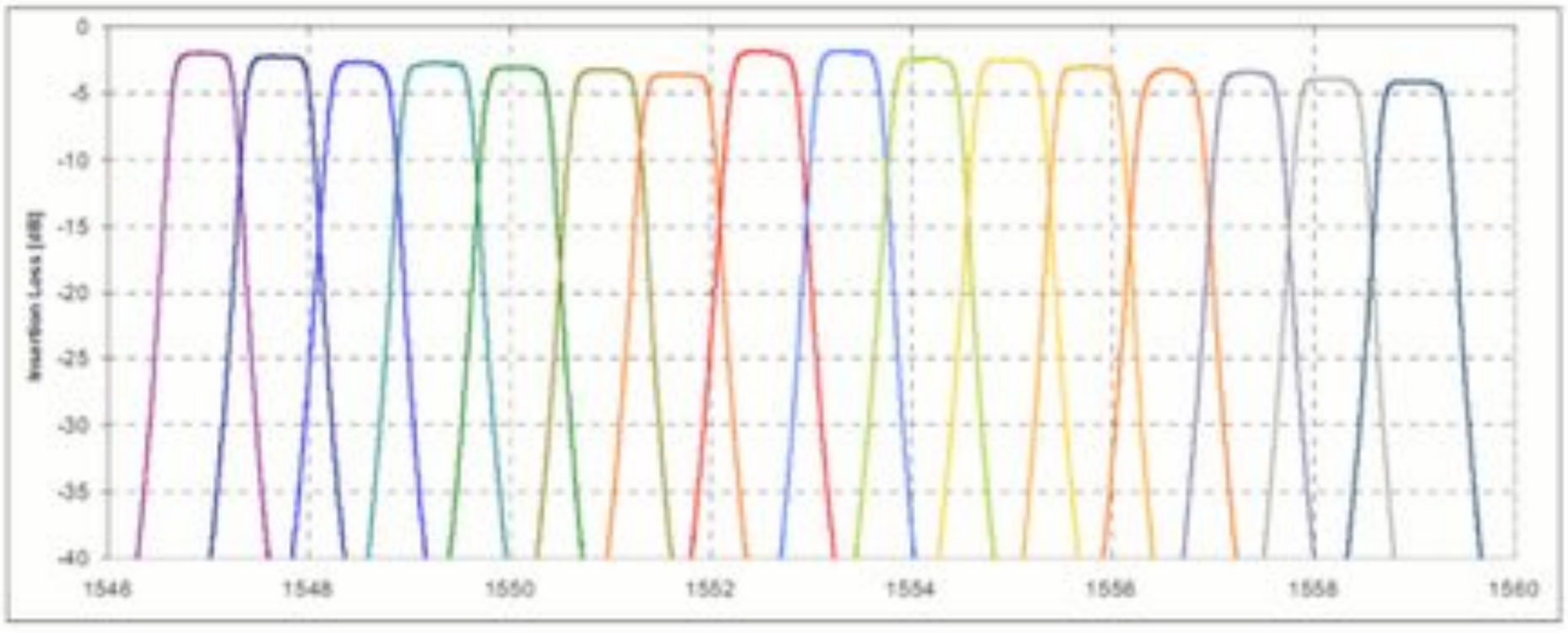
- Different colors can be combined on the same fiber.
- The goal is to put multiple signals on the same fiber





Coarse Wave Division Multiplexing (CWDM)

- CWDM is loosely used to mean “anything not DWDM”
  - One “ popular” meaning is 8 channels with 20nm spacing.
    - Centered on 1470 / 1490 / 1510 / 1530 / 1550 / 1570 / 1590 / 1610



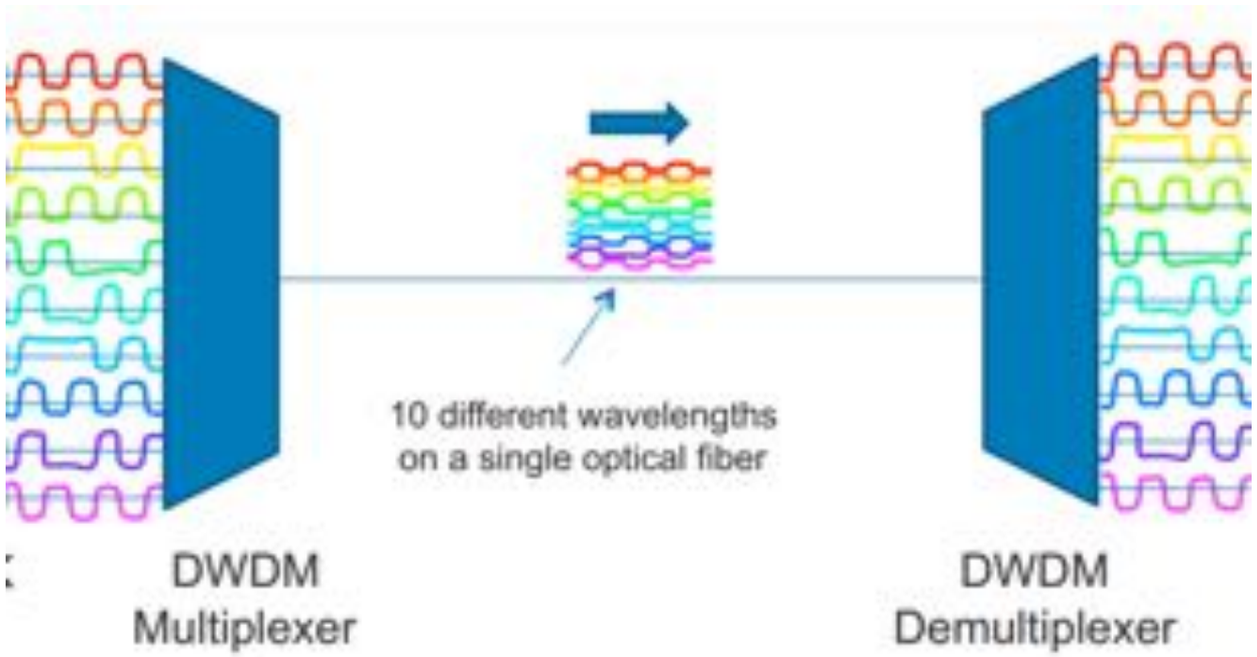
Dense Wave Division Multiplexing (DWDM)

- Defined by the ITU Telecommunication Standardization as a "grid" of specific channels.
- Within C-band, the follow channel sizes are common:
  - 200GHz – 1.6nm spacing, 20-24 channels (old 2000-era tech, rarely seen any more
  - 100GHz – 0.8nm spacing, 40-48 channels (still quite common)
  - 50GHz – 0.4nm spacing, 80-96 channels (common for long-haul 100G systems)
  - 25GHz – 0.2nm spacing, 160-192 channels (used briefly)

- Modern systems are moving towards flexible grids
  - 12.5GHz increments or smaller

Slide credit: Richard A Steenbergen

WDM in One Slide

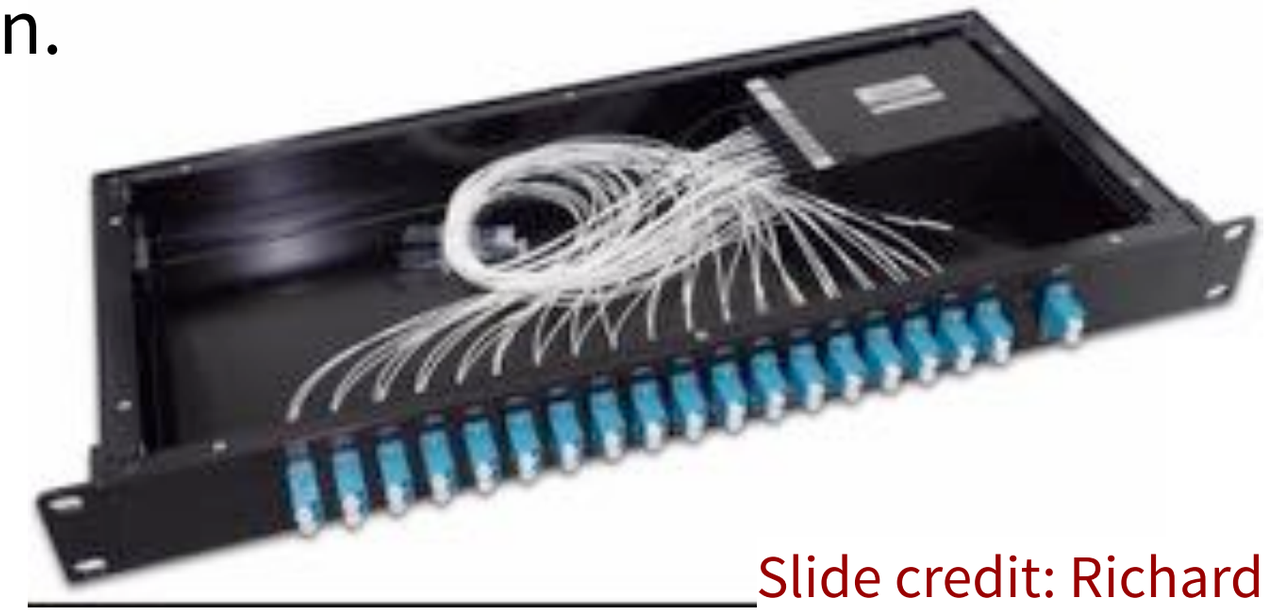


- Protocol and bitrate independent
- Dense WDM systems transmit 160 wavelengths
- Coarse WDM systems transmit 8 channels

WDM Networking Components

WDM Mux/Demux

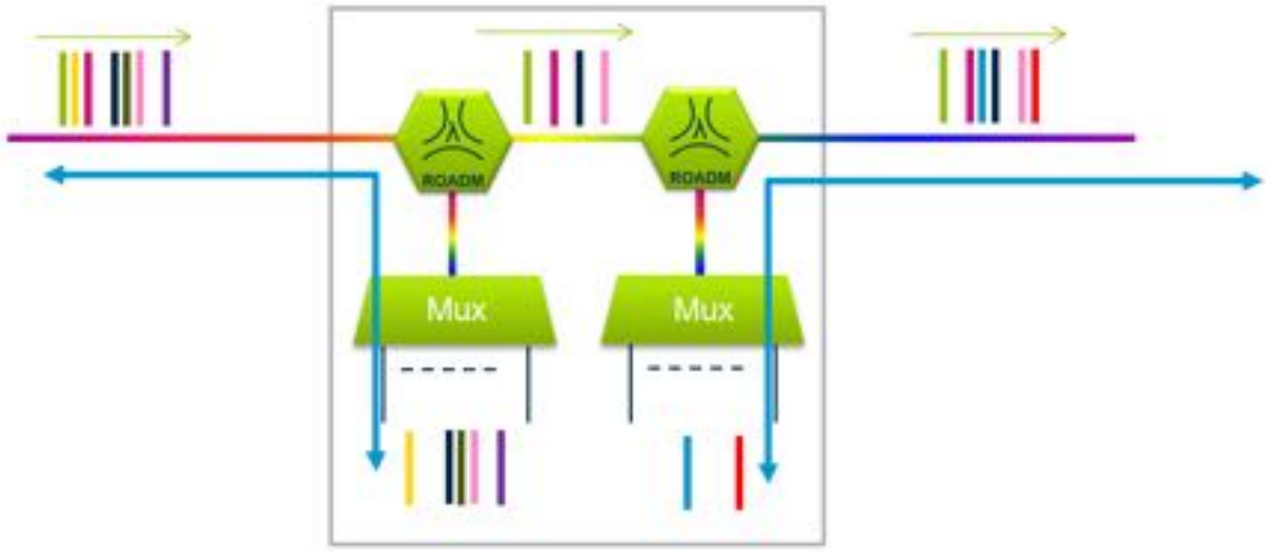
- First device you need to do any kind of WDM
- A passive (unpowered) device which combines/splits multiple colors of light to/from a single “common” fiber
- Short for “ multiplexer”, sometimes called a “filter”, or “prism”
- A “filter ” is how it actually works, by filtering specific colors
- But people conceptually understand that a prism splits light into its various component frequencies.
- A complete system requires both a mux and a demux, for the TX and RX operation.

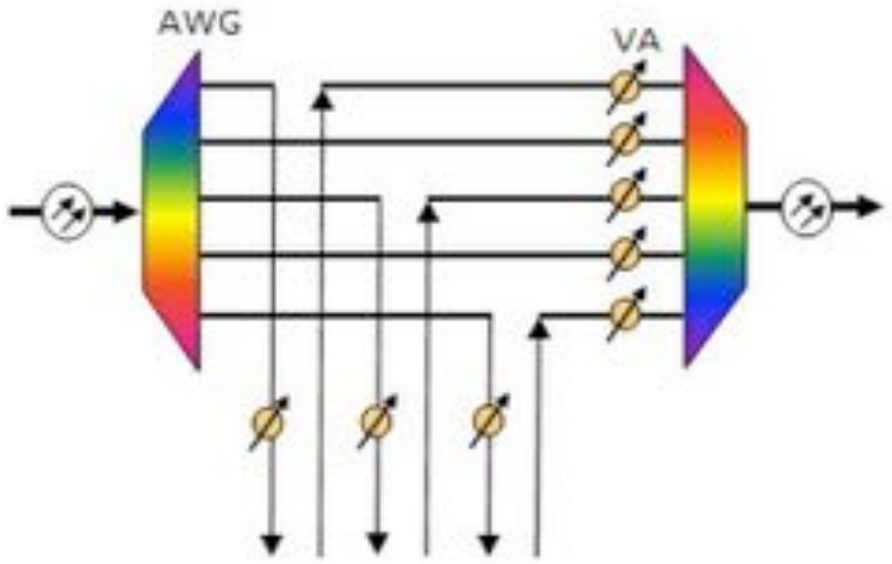


Slide credit: Richard A Steenbergen

The Optical Add/Drop Multiplexer (OADM)

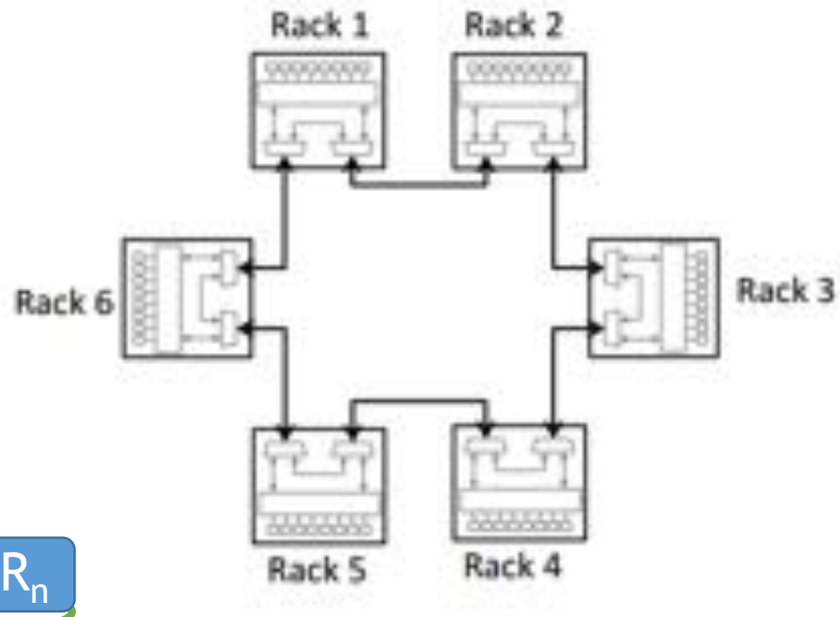
- Selectively Adds and Drops certain WDM channels, while passing other channels through without disruption.
- While muxes often used at major end-points to break out all channels, OADMs are often used at mid-points within rings



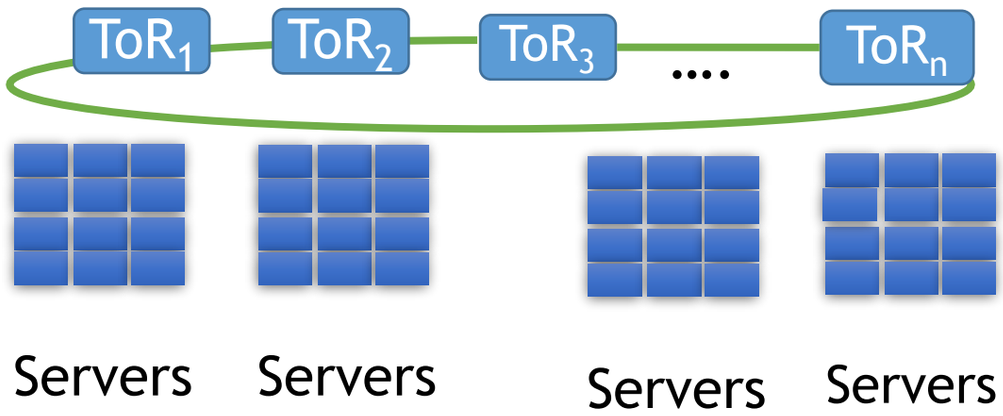


Slide credit: Richard A Steenbergen

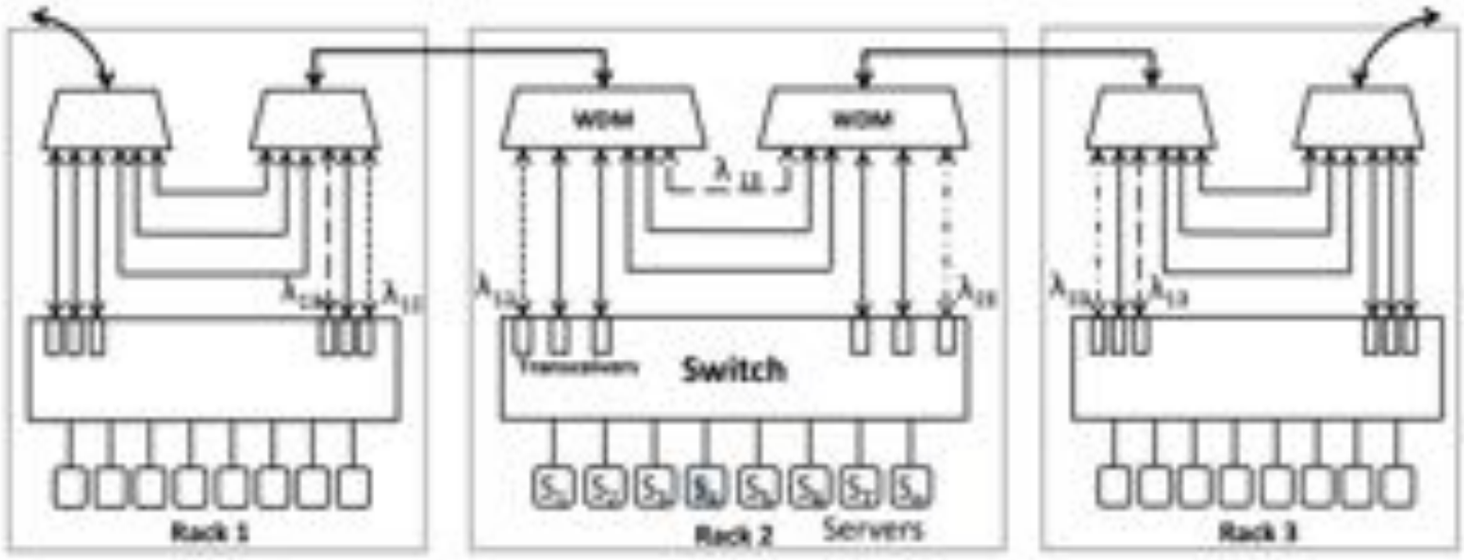
Let’s design a SIGCOMM paper together

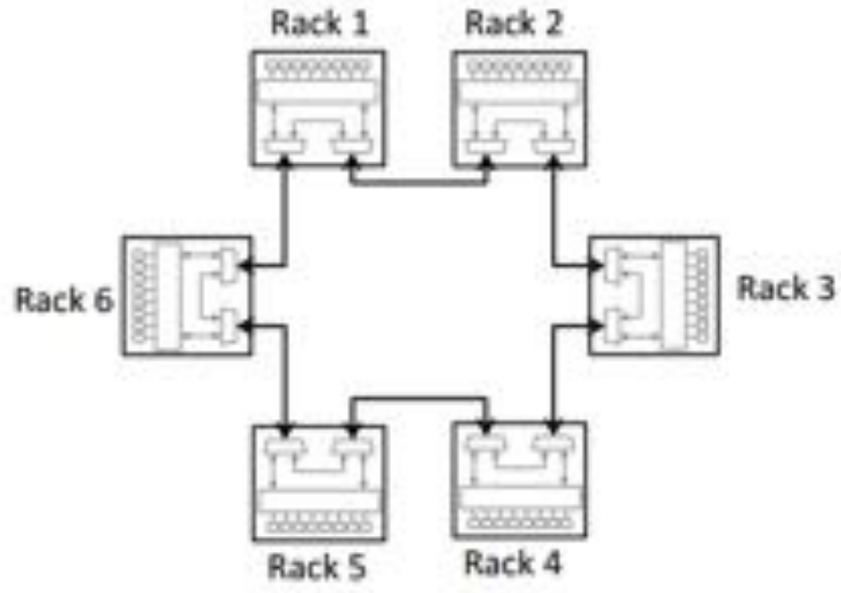


Ring topology



Quartz: A new design element for low-latency data center network [SIGCOMM’14]



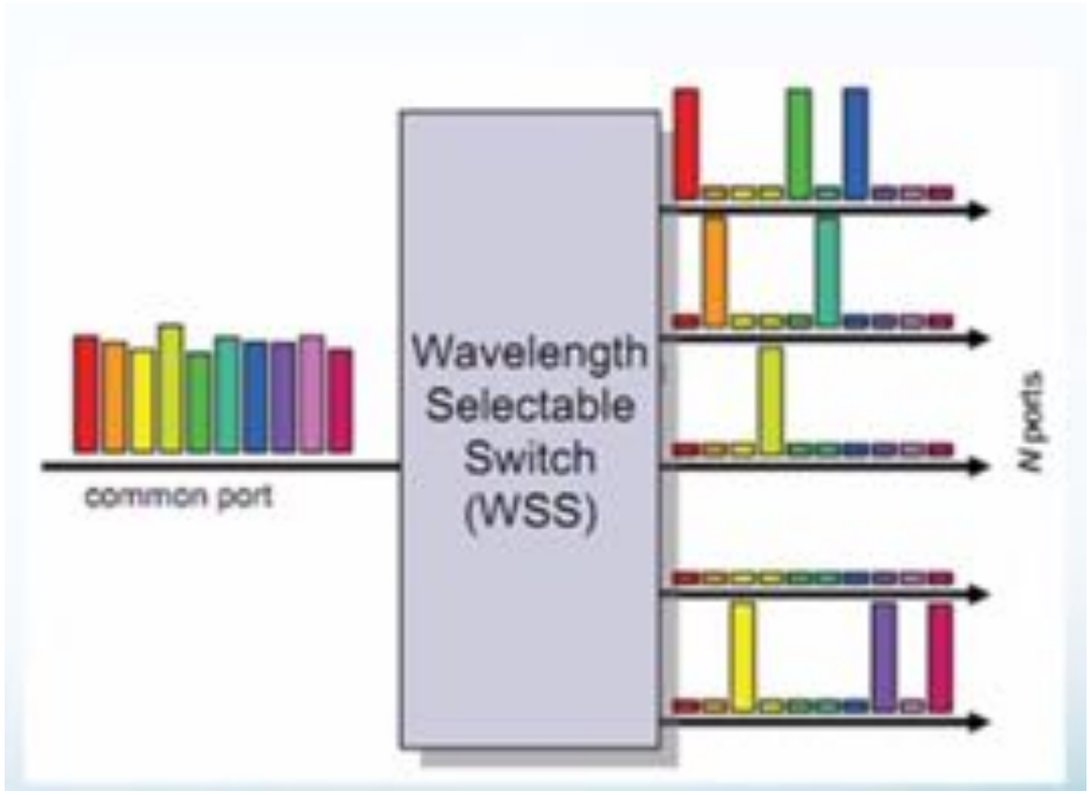


- Each switch gets dedicated wavelengths equal to the total number of servers
- Currently we can only multiplex 160 channels in an optical fiber : Maximum ring size is 35
- Wavelength planning is one time event that is done at design time

Quartz: A new design element for low-latency data center network [SIGCOMM’14]

Wavelength Selective Switch (WSS)

- 1 input port, K output ports
- Different channels from the input fiber can be independently switched to different output ports

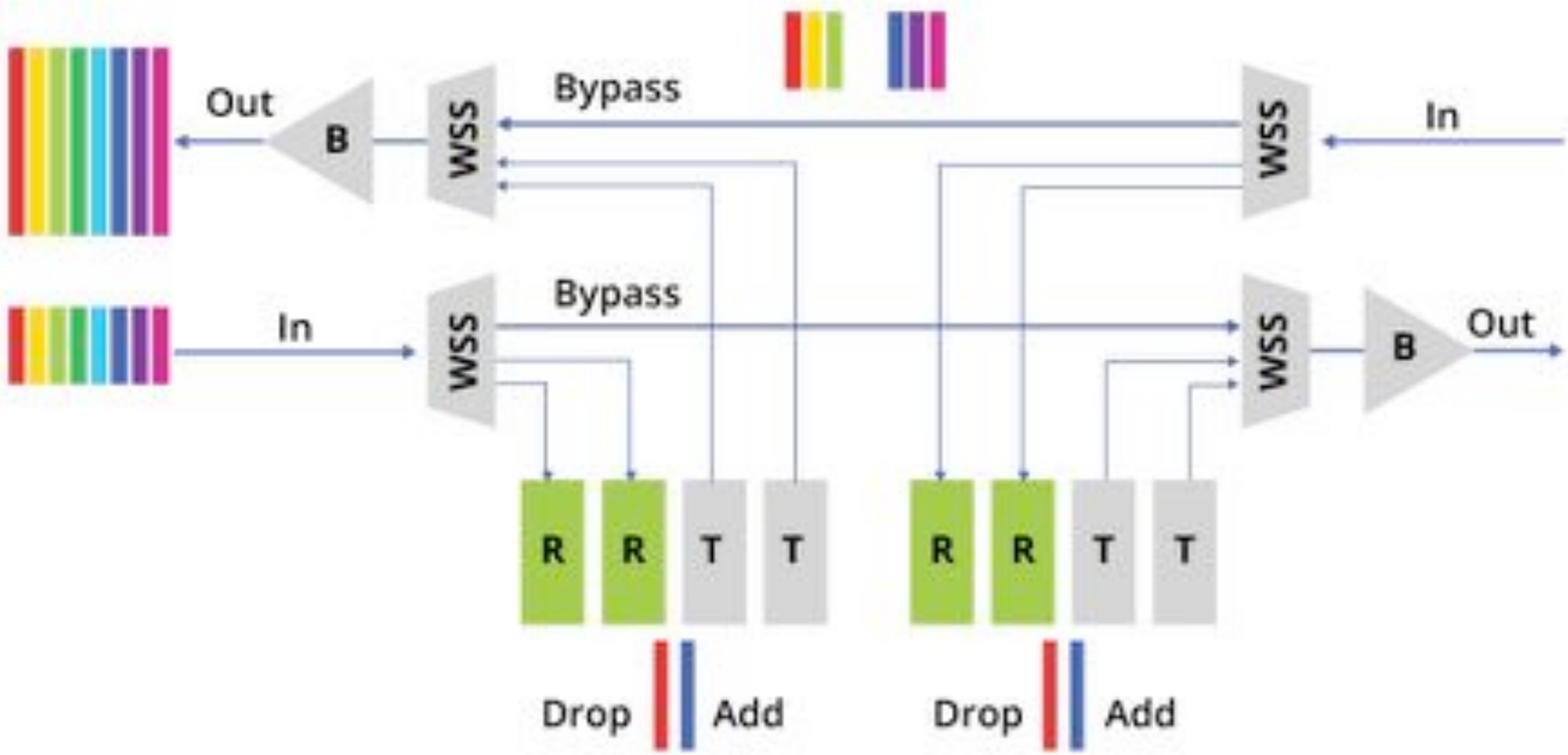


Finisar’s Wavelength Selective Switch (WSS) 4-20 ports, 10-400+ Gbps

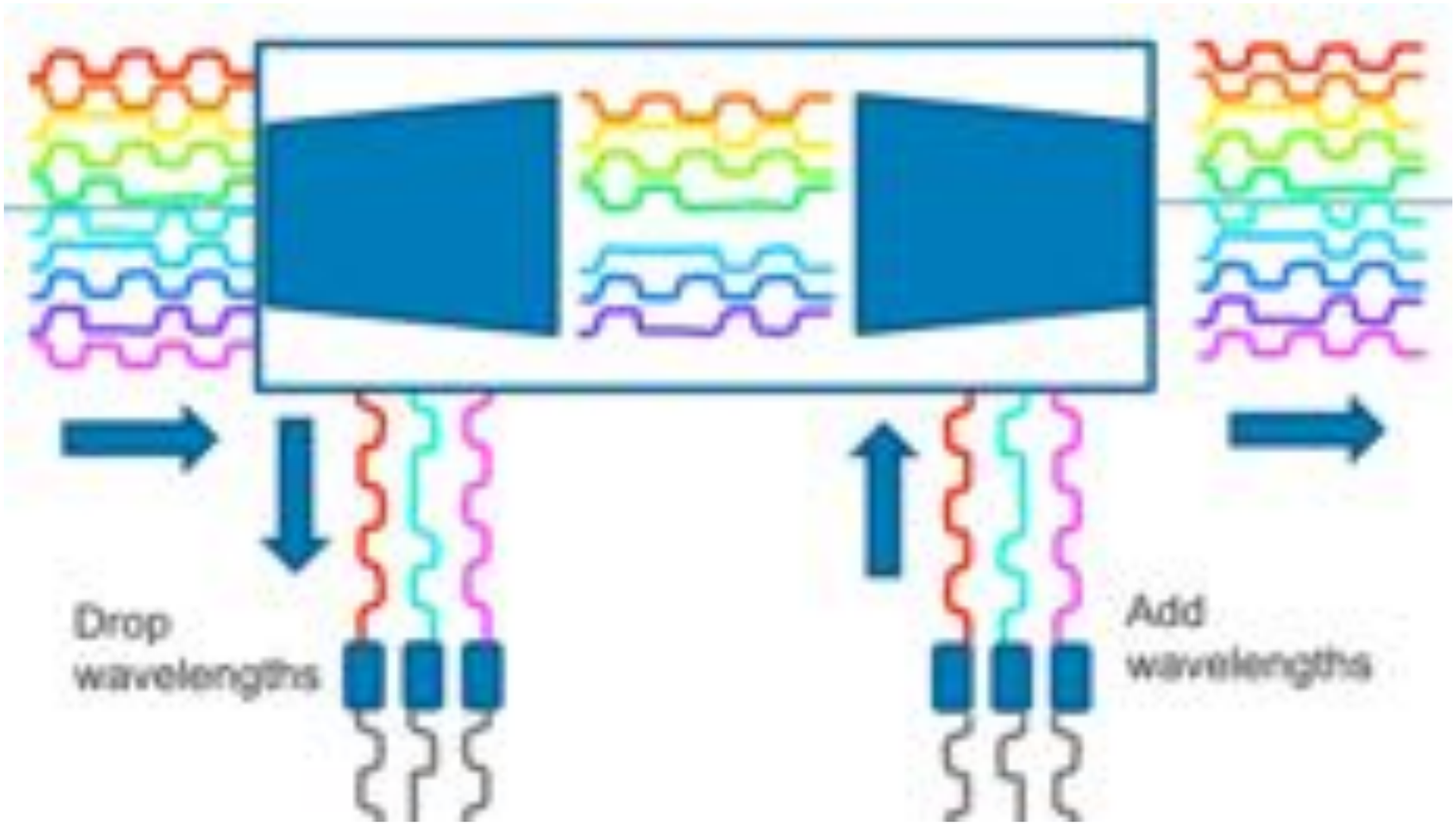


Reconfigurable OADM (ROADM)

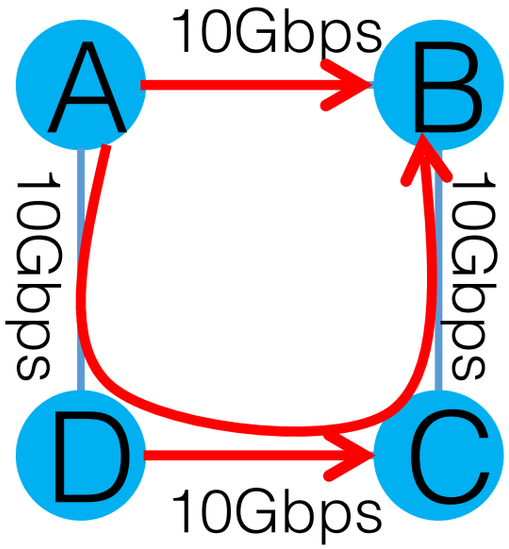
A ROADM is a software reconfigurable OADM



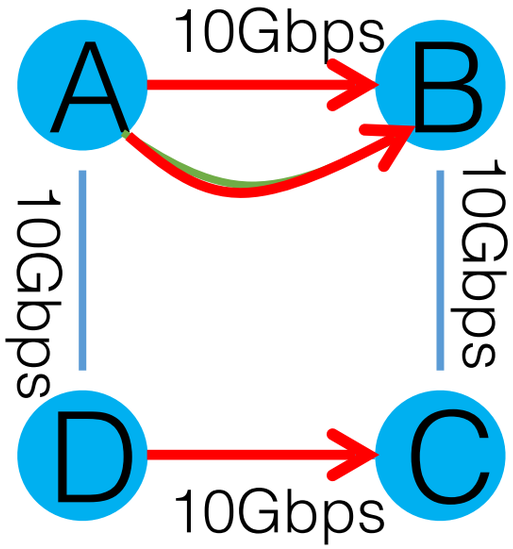
Reconfigurable OADM (ROADM)



The world we are headed



Throughput: 20 Gbps



Throughput: 30 Gbps

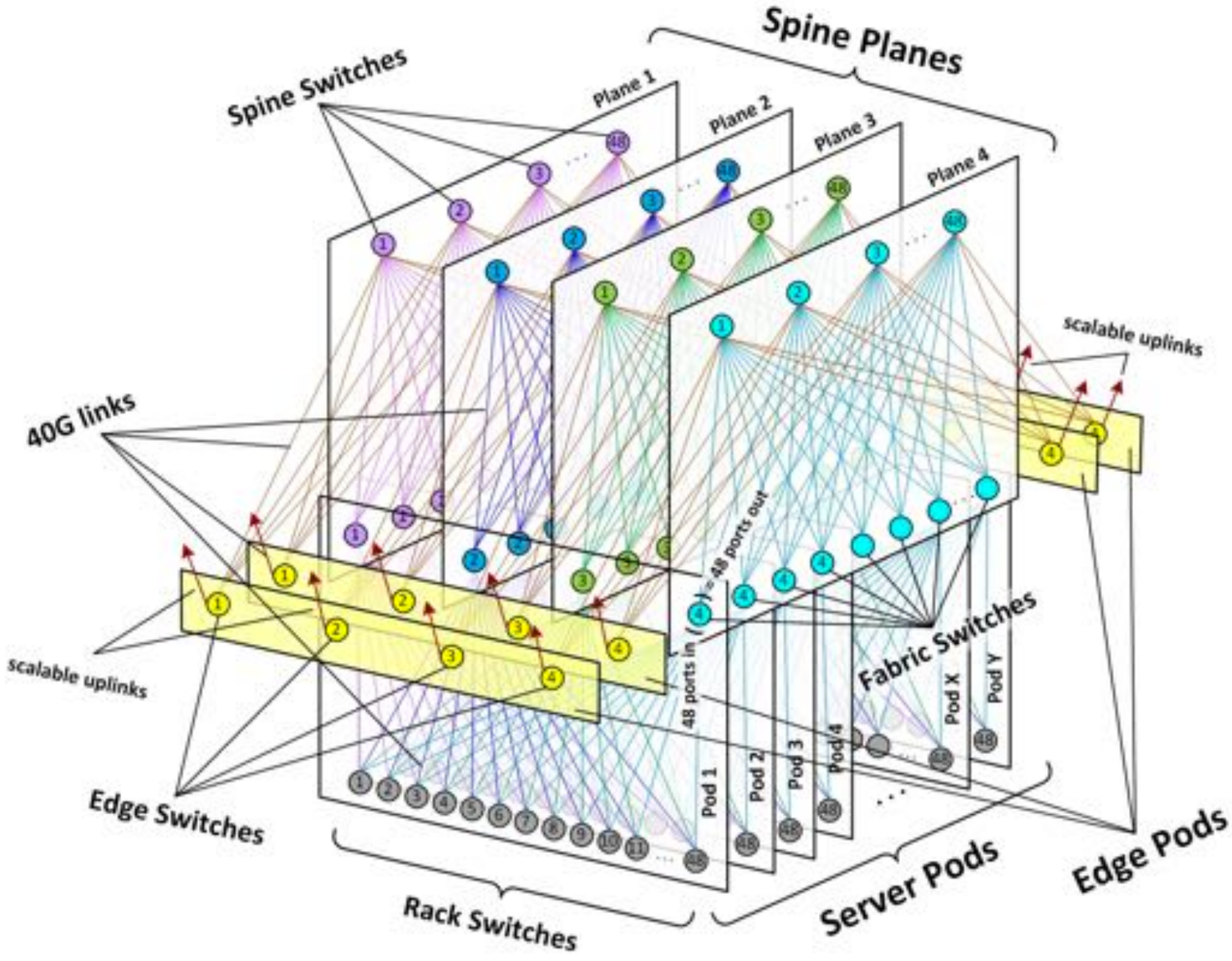
Source ->Destination	Demand
A->B	20 Gbps
D->C	10 Gbps

More on SIGCOMM papers

Data centers run the world

Google data center

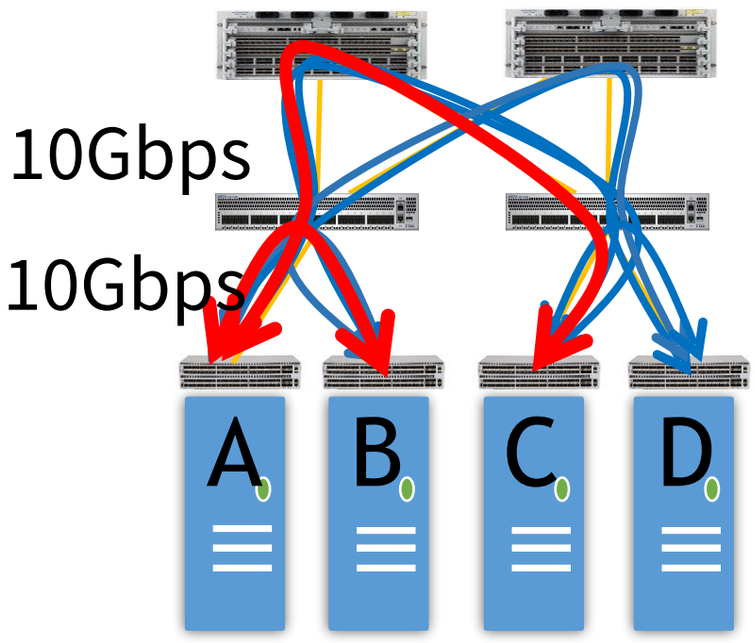
What is an ideal data center topology?



- https://code.facebook.com/posts/360346274145943/introducing-data-center-fabric-the-next-generation-facebook-data-center-network/

Facebook data center

Key observation



	A	B	C	D
A	0	3	3	3
B	3	0	3	3
C	3	3	0	3
D	3	3	3	0

demand matrix:
uniform and static

	A	B	C	D
A	0	6	6	0
B	0	0	0	0
C	0	0	0	0
D	0	0	0	0

demand matrix:
skewed and dynamic

Better topologies?



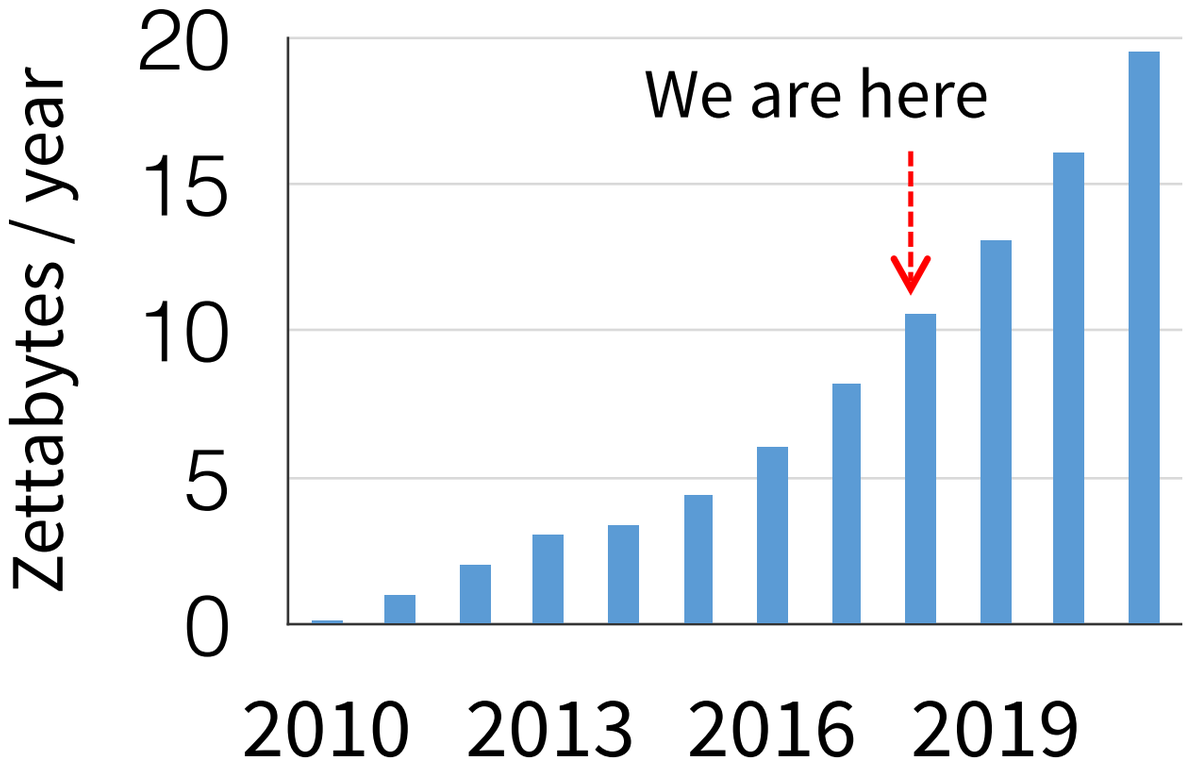
Calient S Series OCS (320 ports)

- Helios: A Hybrid Electrical/Optical Switch Architecture for Modular Data Centers [SIGCOMM’10]
- Integrating Microsecond Circuit Switching into the Data Center [SIGCOMM’13]
- Circuit Switching Under the Radar with REACToR [NSDI’14]
- RotorNet: A Scalable, Low-complexity, Optical Datacenter Network [SIGCOMM’17]

Another key observation for future research: Computing is shifting to the cloud



New workloads (ML, AI, IoT)



Cloud data center traffic growth

Source: Cicso Global Cloud Index

Zettabyte = 10^21 bytes

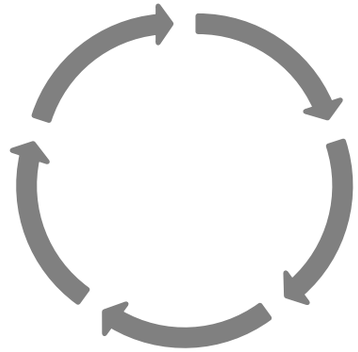
My research

High-performance cloud infrastructure
for emerging workloads (AI, ML, IoT, …)
using new algorithms and hardware.

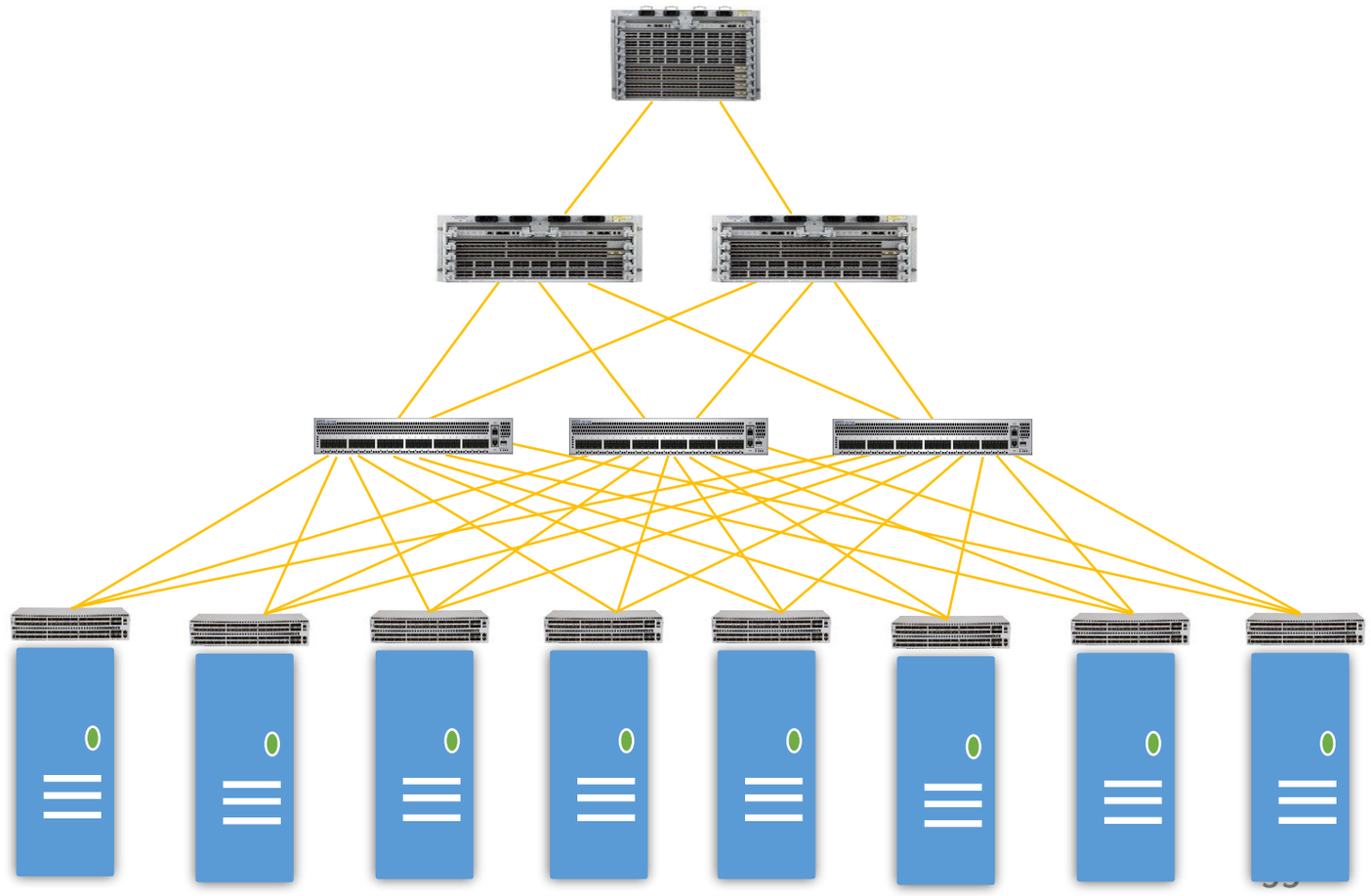
Internet

Data centers

Servers



Measurement/Theory
Prototype/Simulation
Real-world deployment

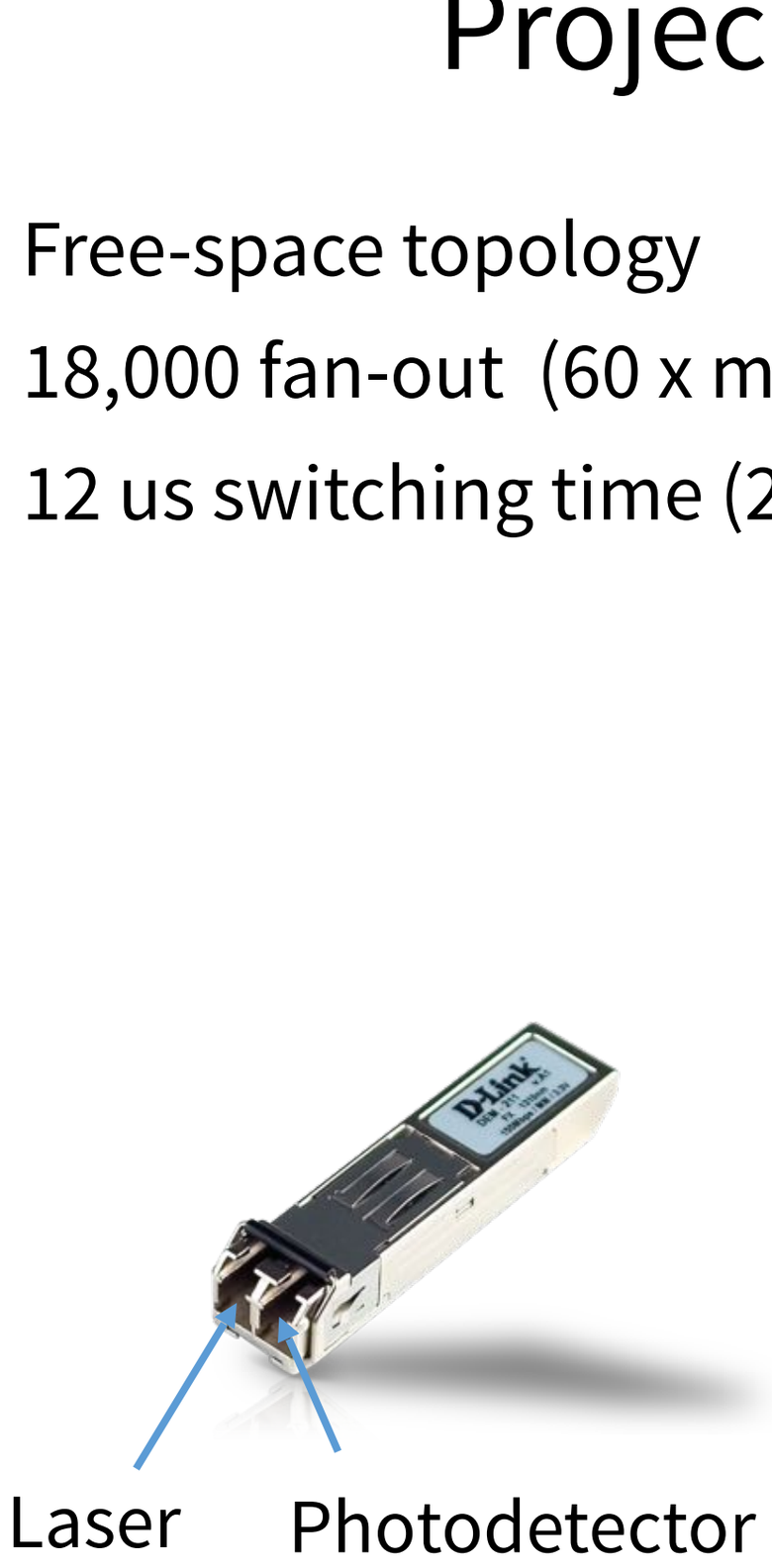


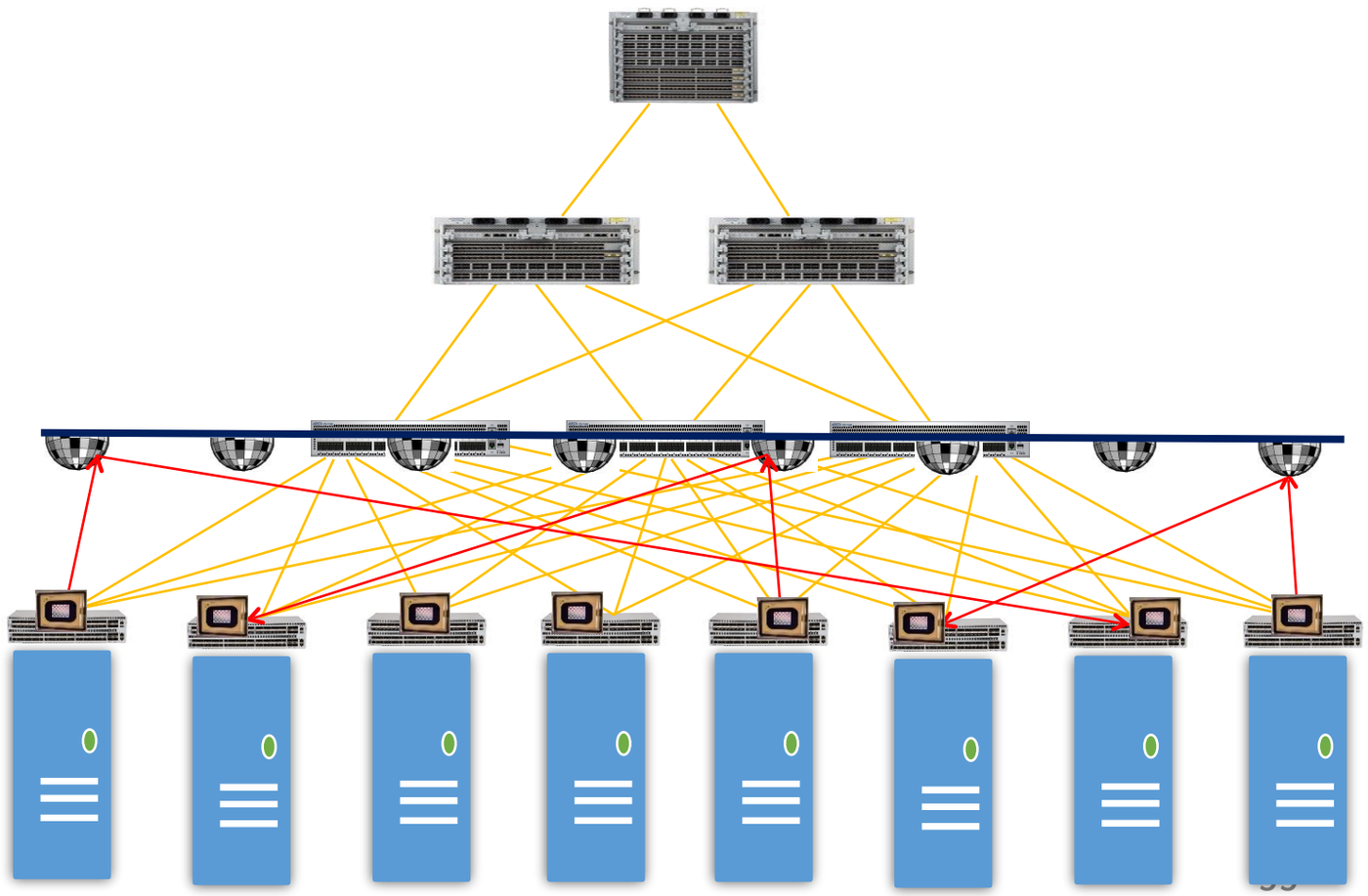
How are servers interconnected?



ProjecToR data center

- Free-space topology
- 18,000 fan-out (60 x more than optical circuit switches)
- 12 us switching time (2500 x faster than optical circuit switches)

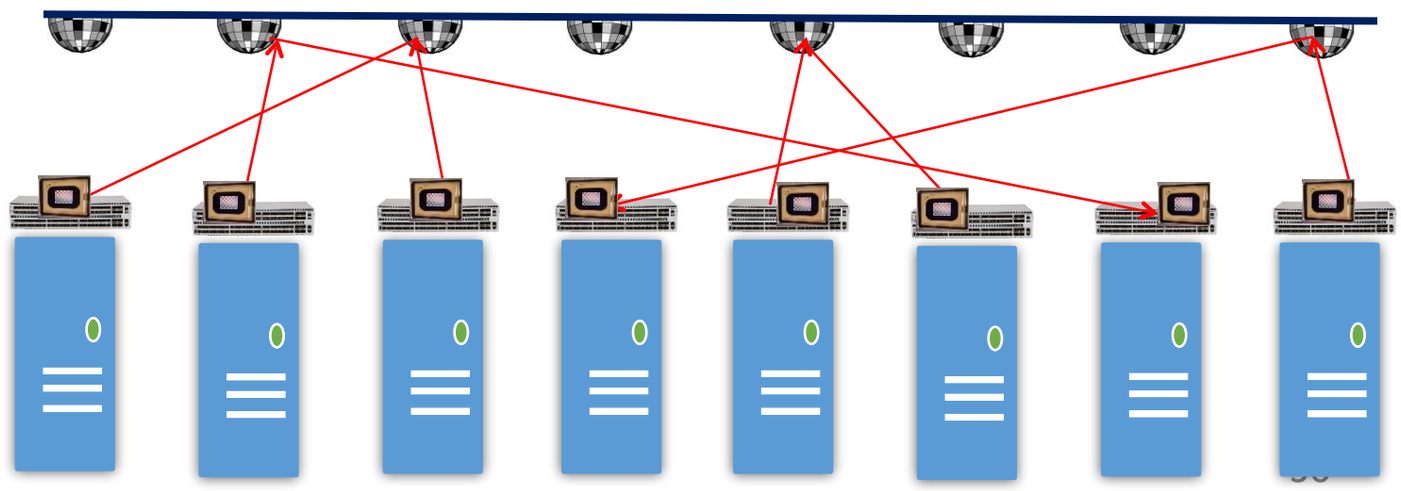




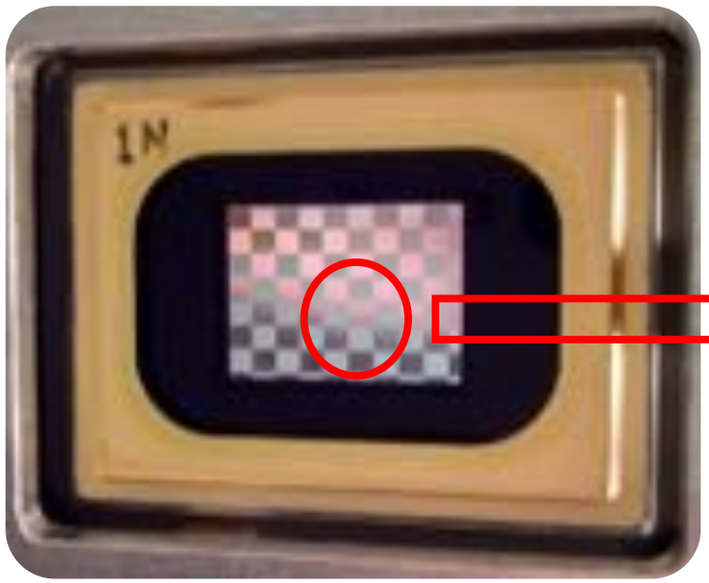
ProjecToR: Agile Reconfigurable Data Center Interconnect, Ghobadi et al. [SIGCOMM'16]

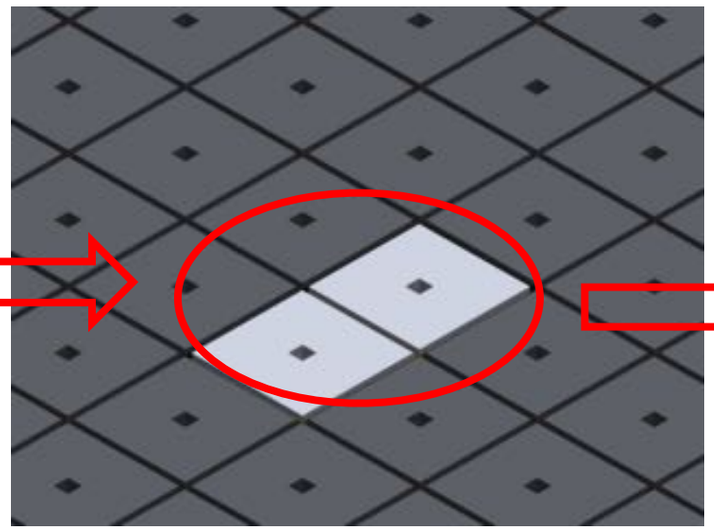
Reconfiguration in a ProjecToR interconnect

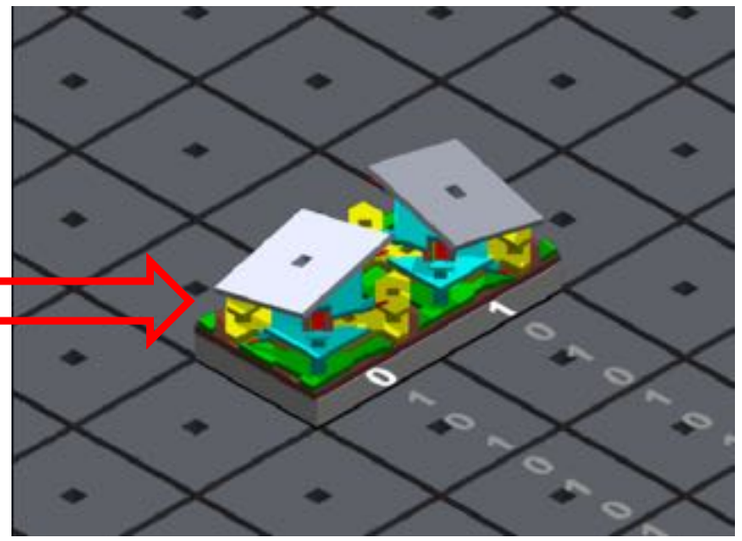
- Digital micromirror device to redirect light
- Disco-ball mirror assembly to magnify reach



Digital Micromirror Device (DMD)



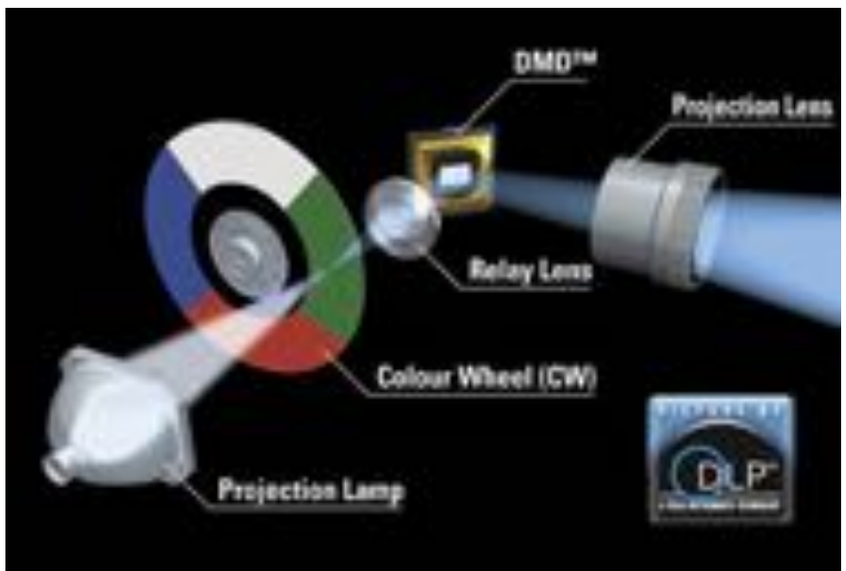




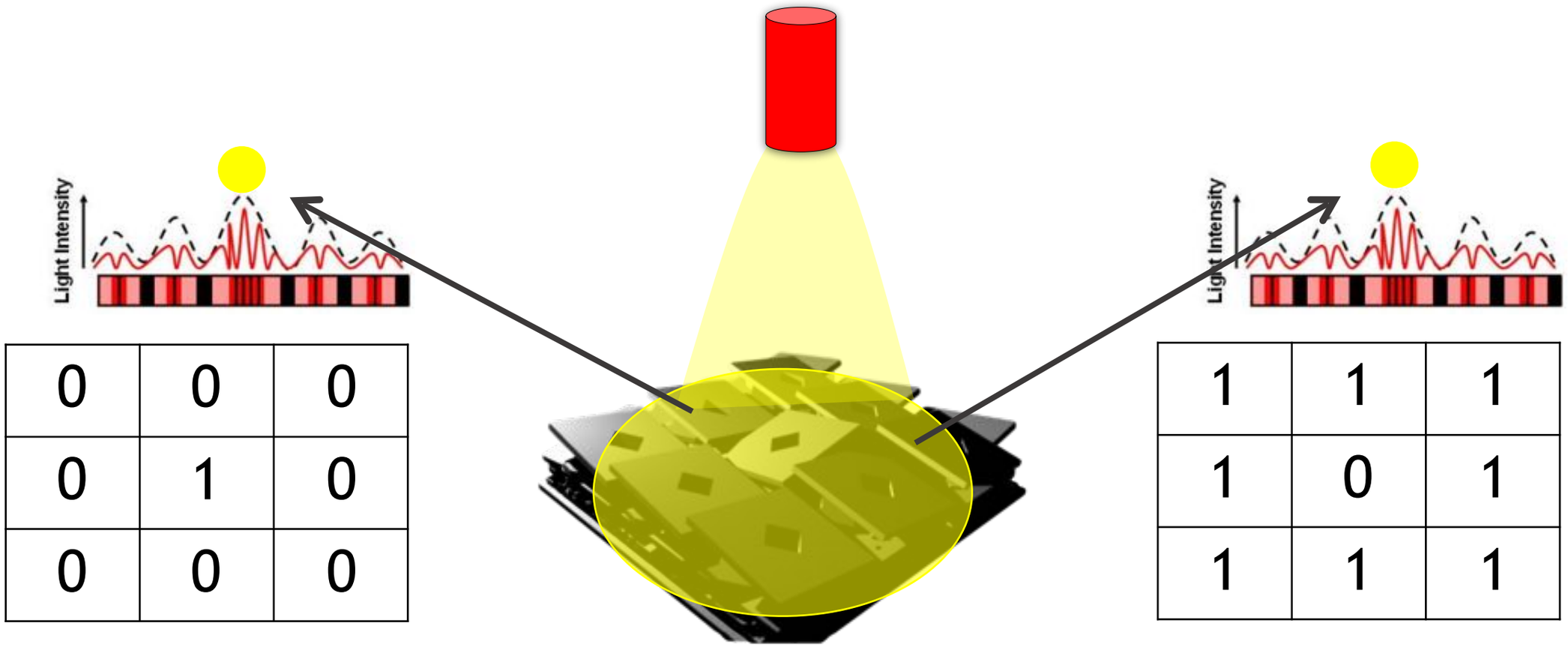
TEXAS INSTRUMENTS

Array of micromirrors (10 um)

Memory cell



Using DMDs to redirect light



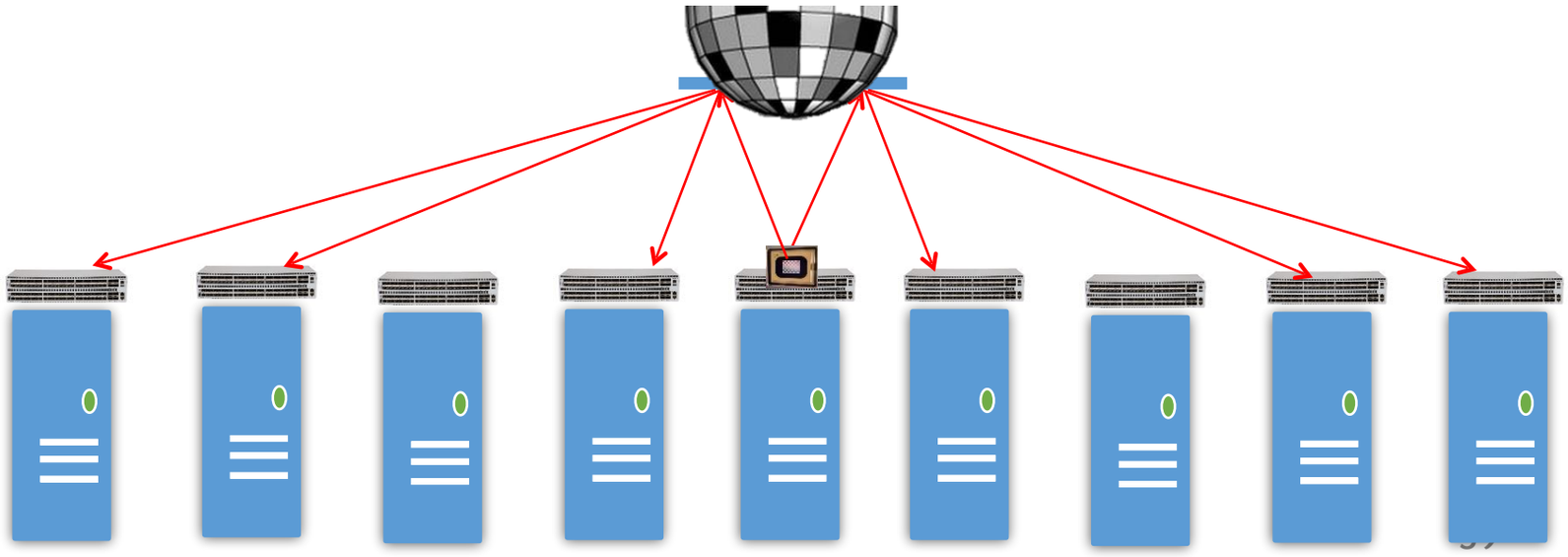
0	0	0
0	1	0
0	0	0

1	1	1
1	0	1
1	1	1

- Theoretical number of accessible locations: total number of micromirrors
  - 768x768 = 589824
- Cross-talk between adjacent locations
- Achievable number of accessible locations
  - 768x768 / 32 = 18,432

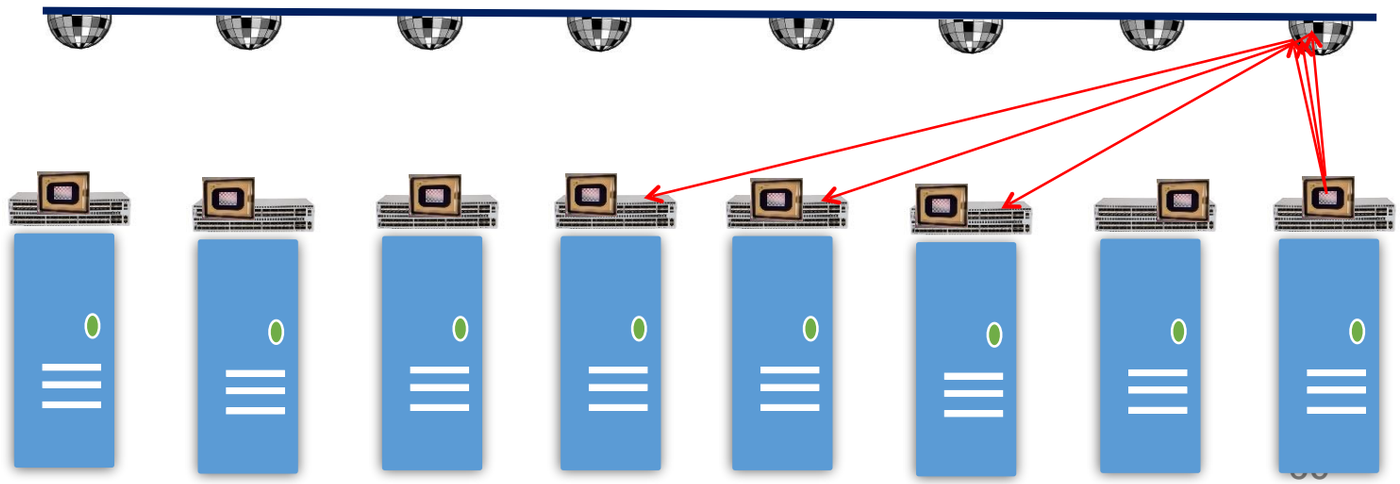
Using mirror assemblies to magnify reach

- Challenge: DMDs have a narrow angular reach
- Solution: Coupling DMDs with angled mirrors

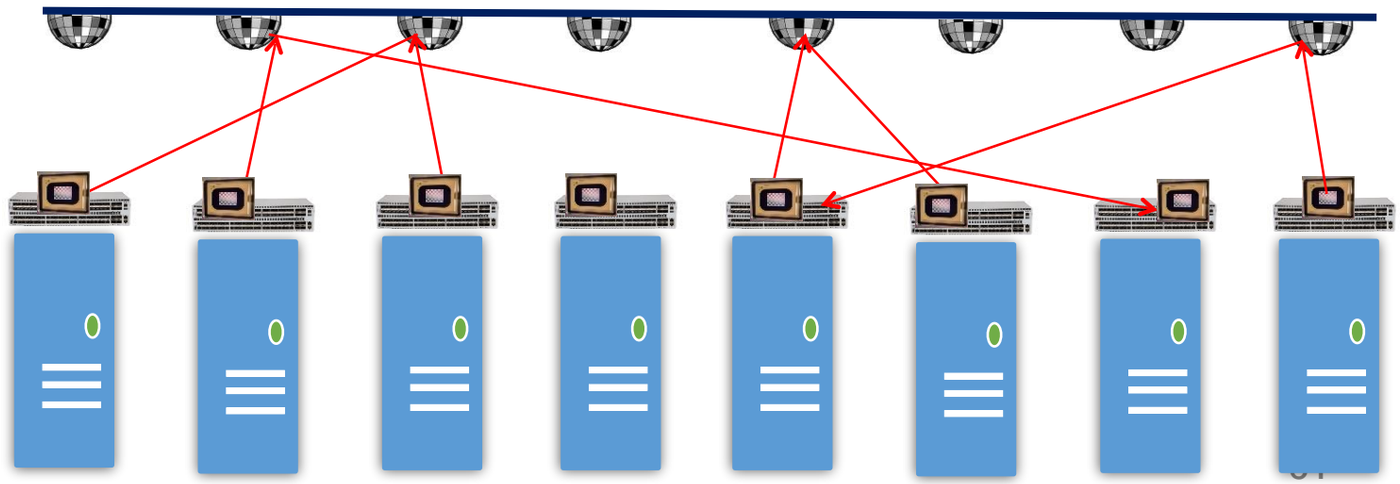


- To see the disco-ball come to my office G32-940!

ProjecToR interconnect



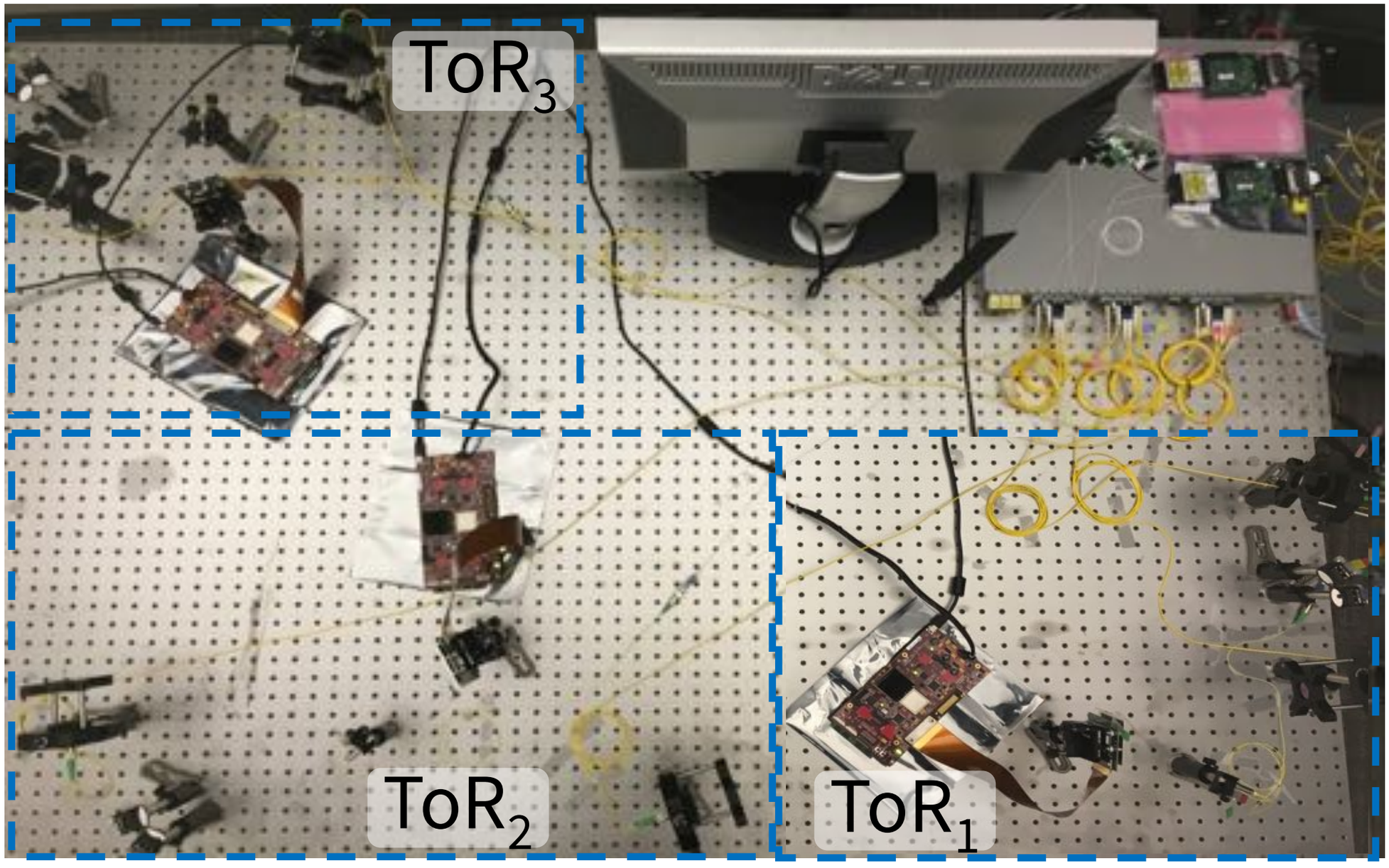
ProjecToR interconnect



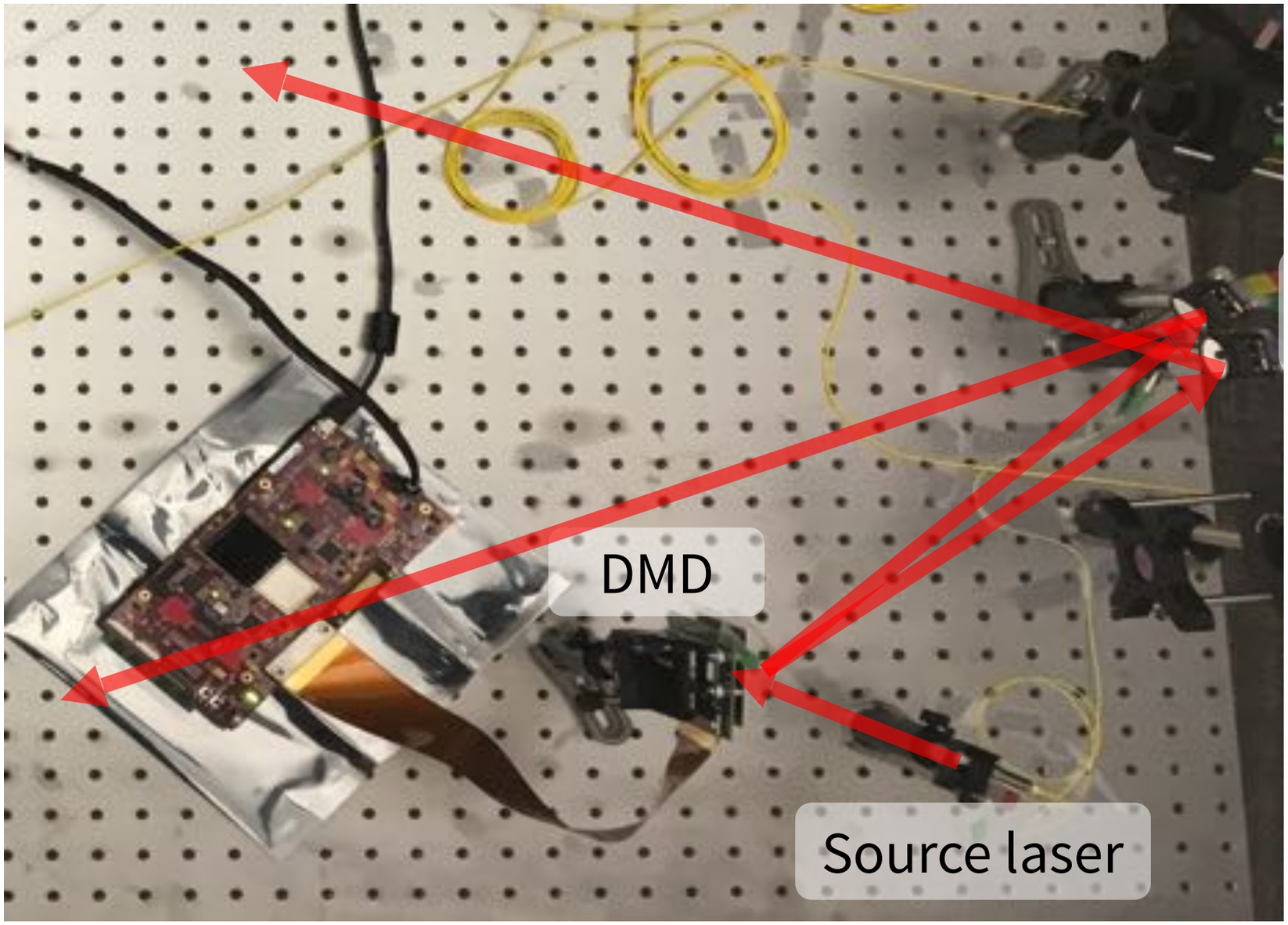
Questions to answer

- How feasible is a ProjecToR interconnect?
- How should packets be routed in a ProjecToR interconnect?
- How much does a ProjecToR interconnect cost?

Prototype: A 3-ToR ProjecToR interconnect



Prototype: A 3-ToR ProjecToR interconnect



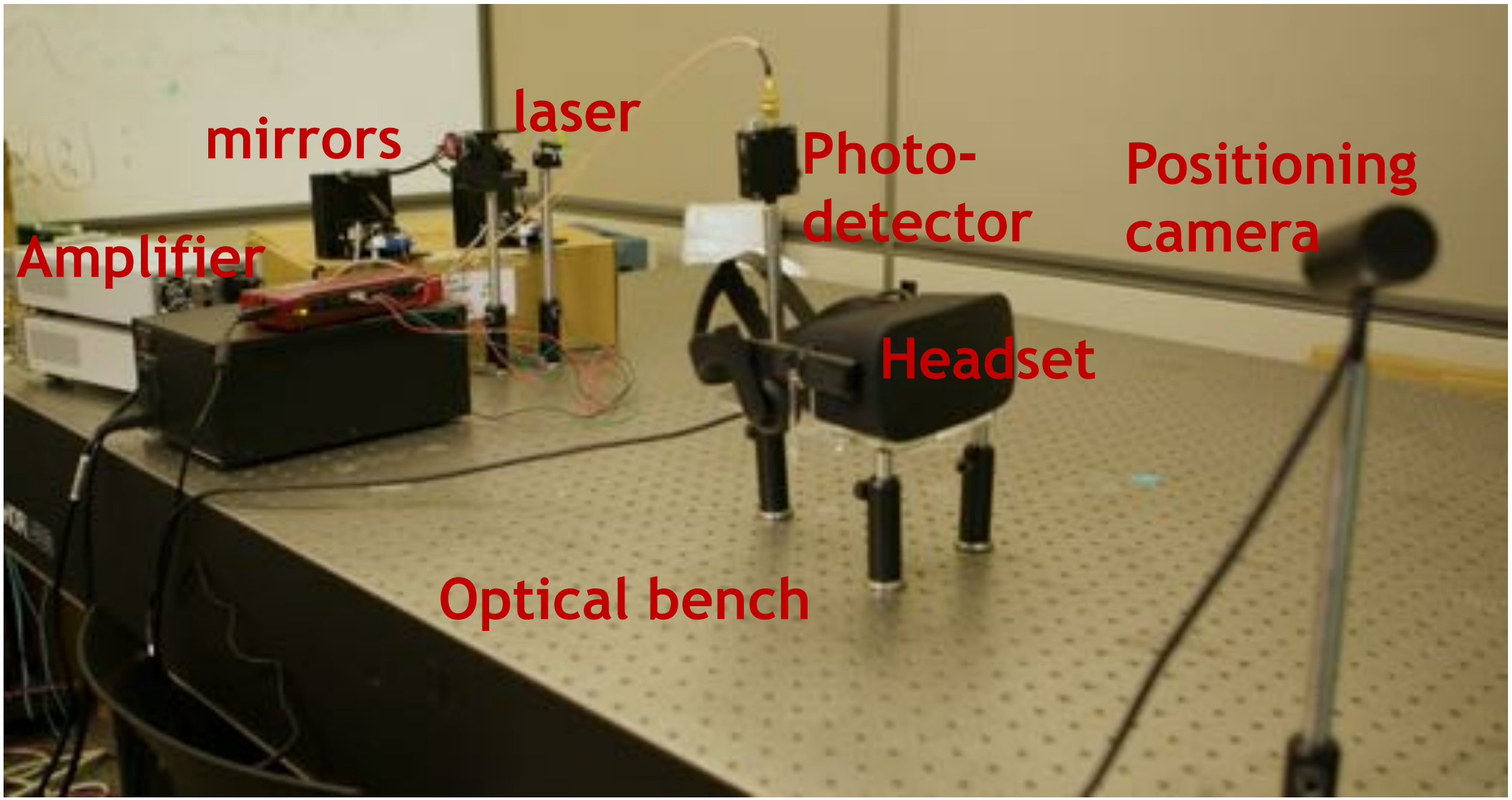
Mirrors reflecting to ToR_2 and ToR_3



Futuristic stuff:

Free-space optics for indoor IoT devices

Free-space lasers for virtual reality headsets



Slide credit: Manikanta Kotaru