THE STATE OF ETHERNET OPTICS

Scott Kipp, Brocade, President of the Ethernet Alliance
Brad Smith, Mellanox
Chris Cole, Finisar
Mark Nowell, Cisco
OFC 2016, March 23, 2016

ethernet alliance

Disclaimer

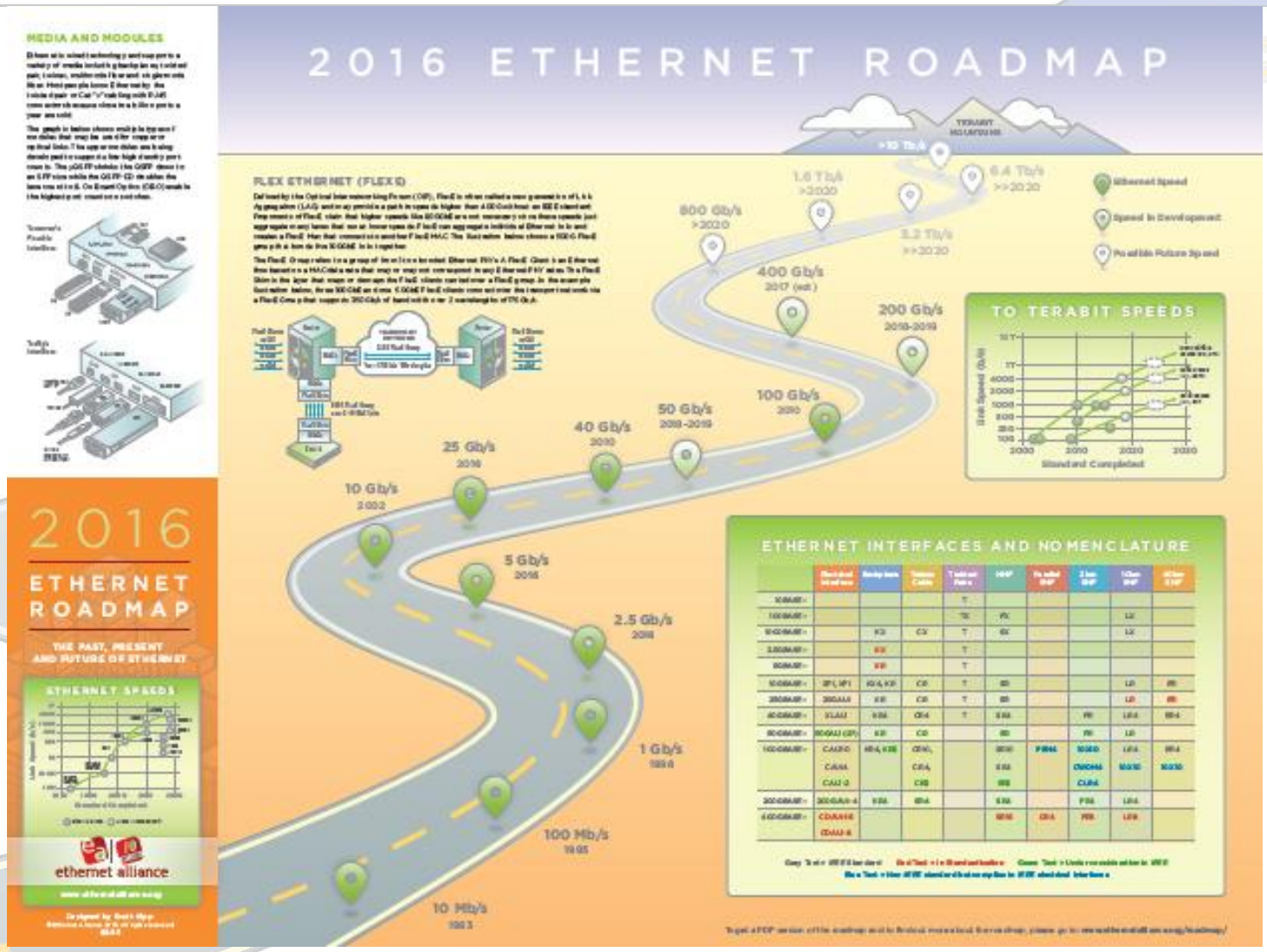
- Opinions expressed during this presentation are the views of the presenters, and should not be considered the views or positions of the Ethernet Alliance.

Agenda

- 2:30-2:40 – The 2016 Ethernet Roadmap – Scott Kipp, Brocade
- 2:40-2:52 – The Ethernet Landscape Today - Brad Smith, Mellanox
- 2:52-3:04 – The Ethernet Landscape Tomorrow – Chris Cole, Finisar
- 3:04-3:16 – Systems Use of Ethernet speeds – Mark Nowell, Cisco
- 3:16-3:30 – Q&A

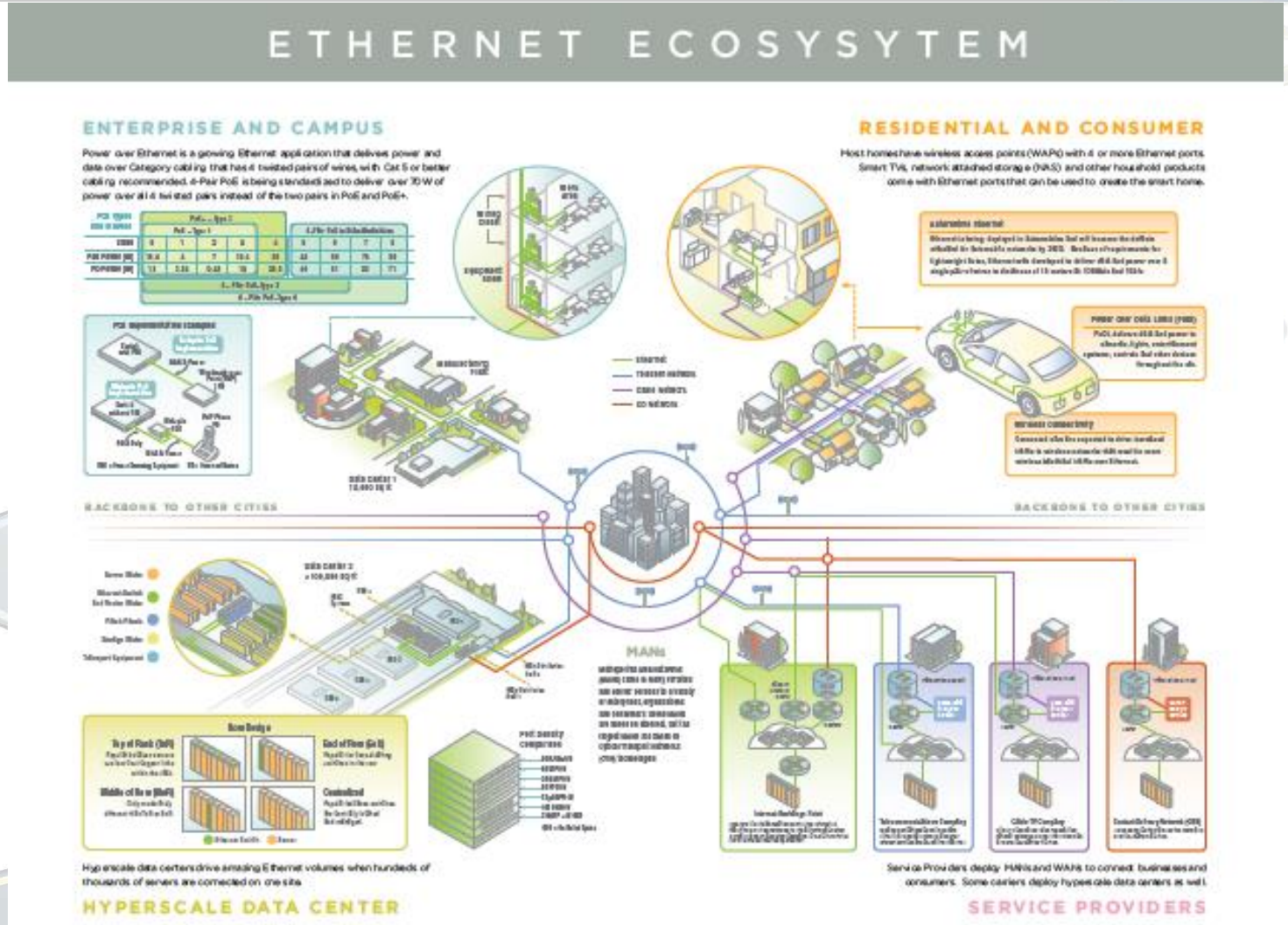
The 2016 Ethernet Roadmap

- The Front



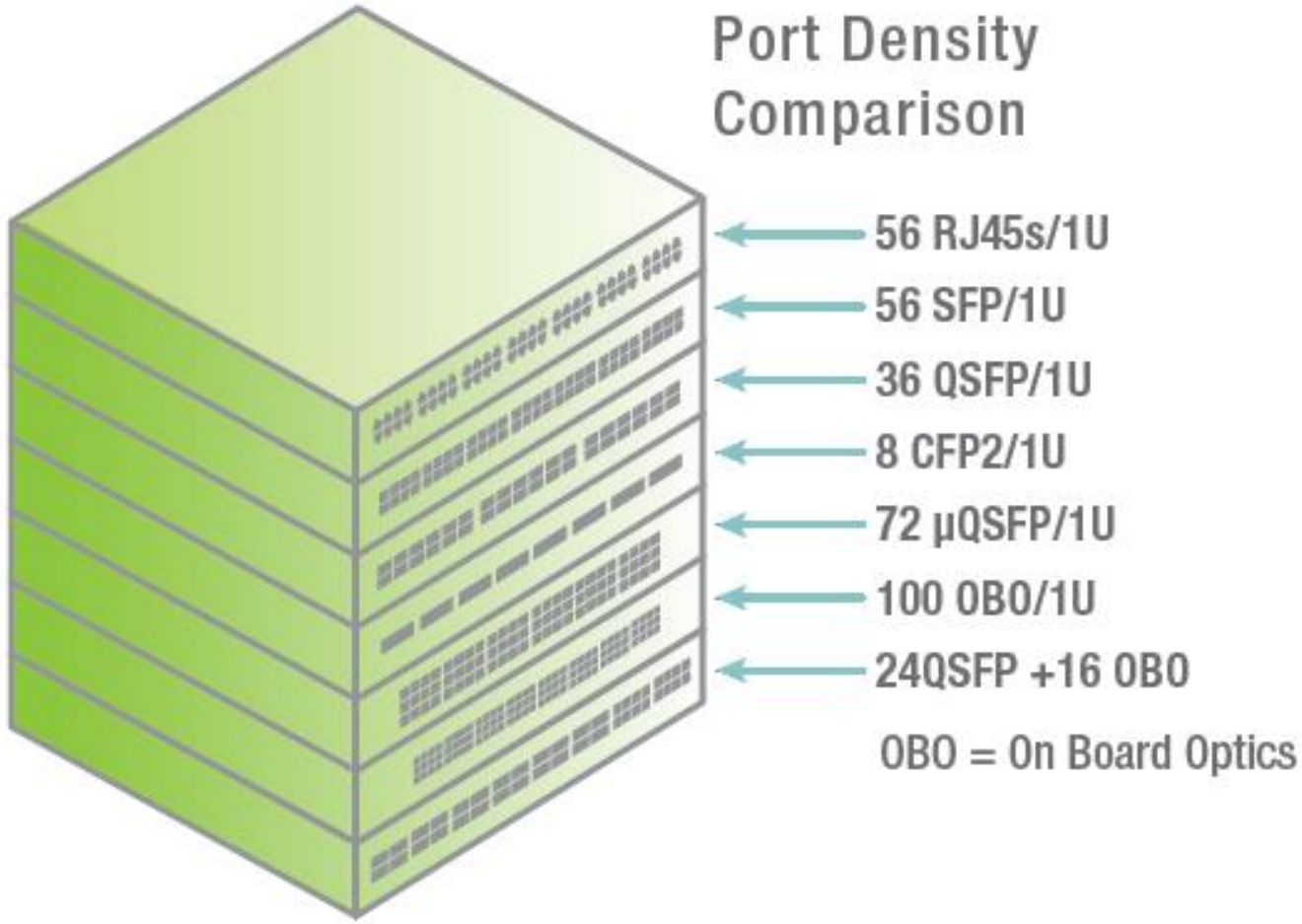
The 2016 Ethernet Roadmap

- The Back



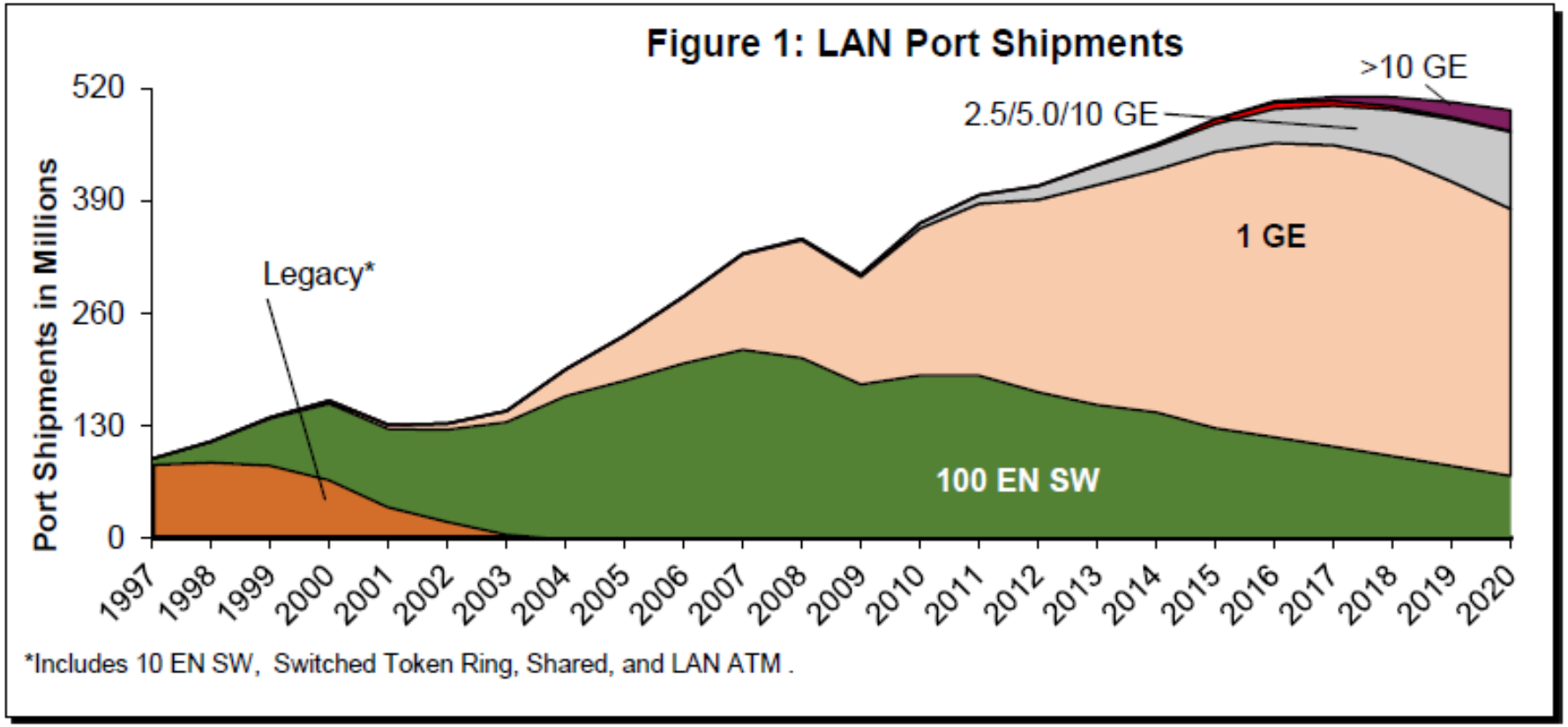
Port Density Comparison

- How many ports can you fit in 1U?



The Ethernet Landscape

- Ethernet shipping over 1B ports/year
- Over $2B in Ethernet modules sold every year



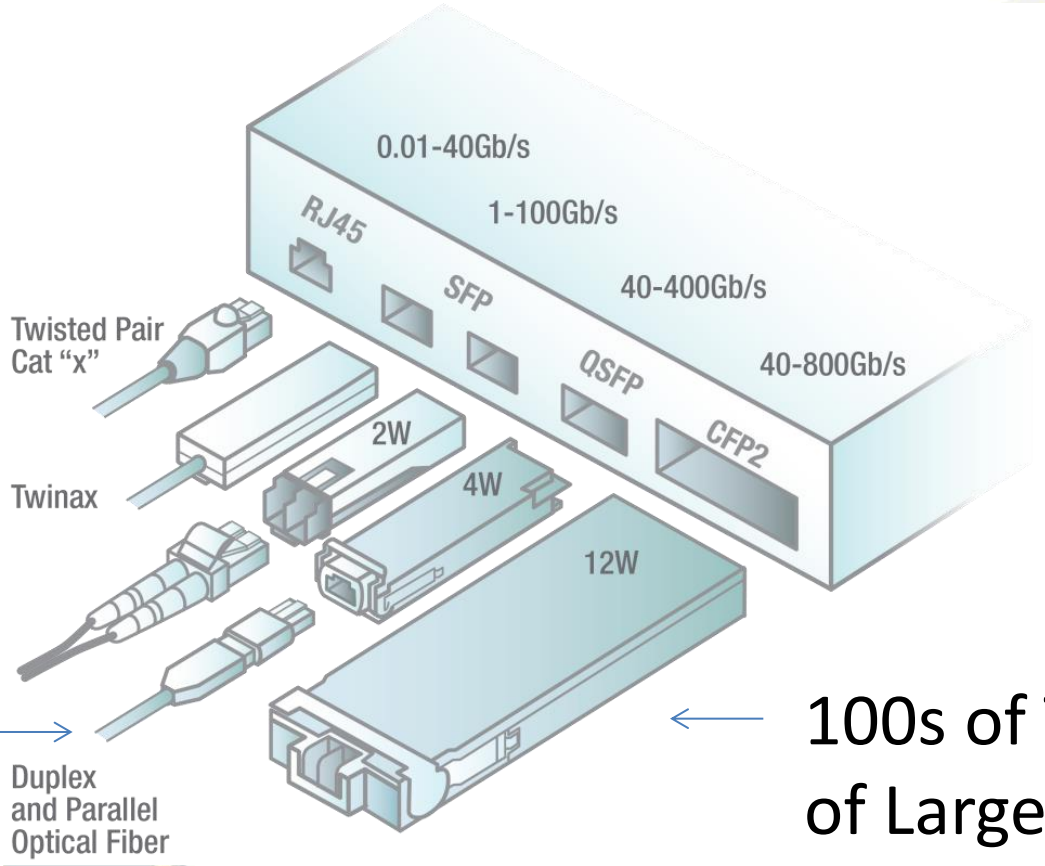
Source: Dell'Oro Ethernet Switch Forecast

Ethernet Port Volumes

100s of Millions of BASE-T ports/year

10s of Millions of SFP ports/year

Millions of QSFP ports/year



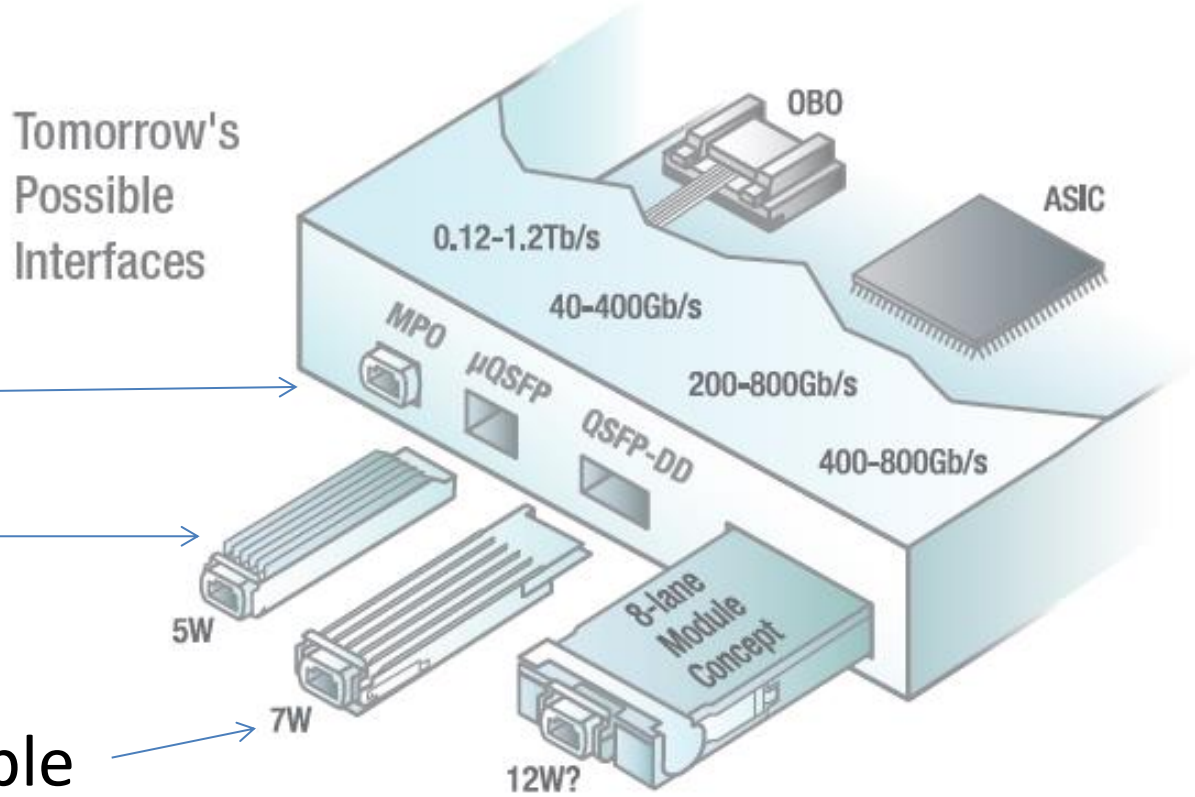
100s of Thousands of Larger than QSFP ports/year

Modules of the Future

Will On Board Optics (OBO) finally reach volume shipments

Will µQSFP replace many QSFP?

Will QSFP-DD enable 400GbE and surpass QSFP?



What will be the 400G module of choice?

ETHERNET OPTICS TODAY: 25G NRZ

THE STATE OF ETHERNET OPTICS PANEL

Brad Smith,

BradS@Mellanox.com

Director of Marketing, LinkX Interconnects, Mellanox

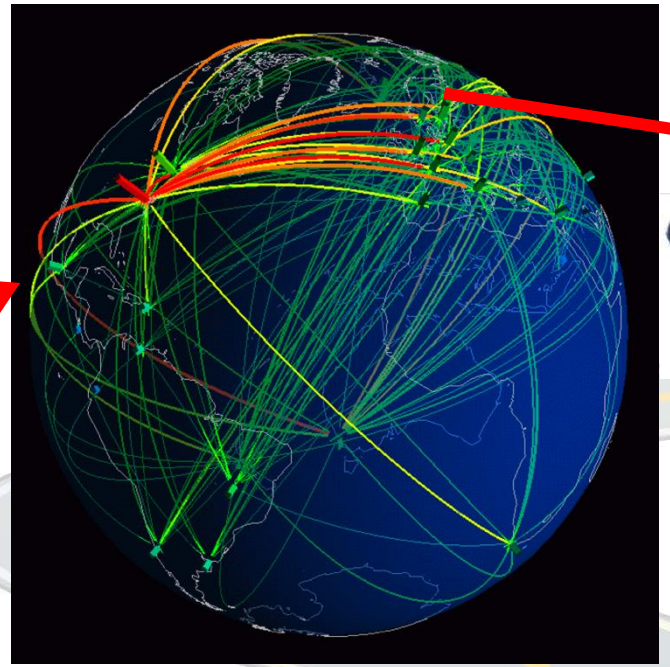
March 23, 2016 OFC 2016 Anaheim, CA

ethernet alliance

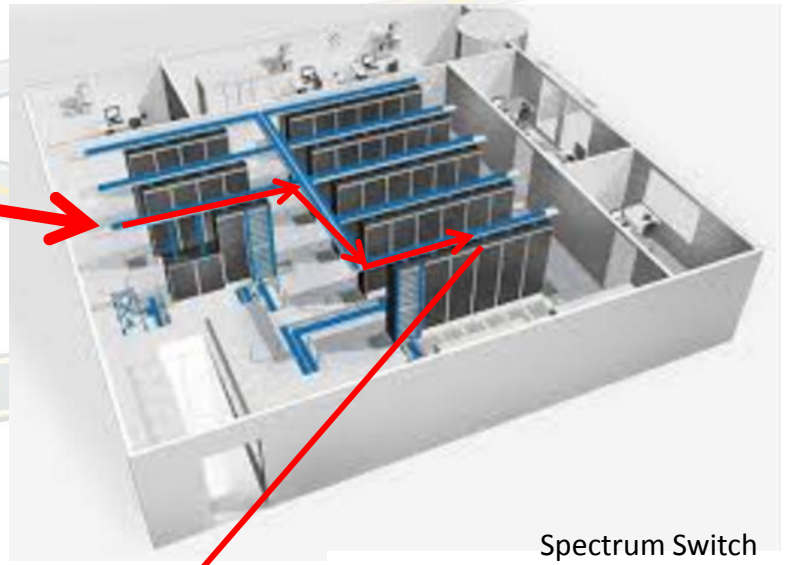
1GBASE-T CAT5 is s REALLY big in Asia

IT’S YUGE!

All Internet Traffic Flow Through Optical Transceivers

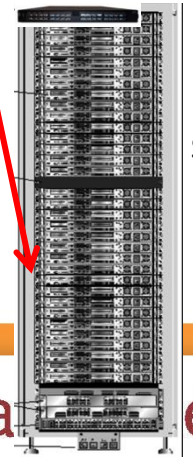


Data Center



Spectrum Switch

Optical Transceiver



Server Rack

New Industry Mantra

The money today may be in 10G/40G optics …but

"25G is the new 10G"

"100G is the new 40G"

WHY?

BOM Costs are Almost the Same;

40G (4x10G)
SR4, AOC, CWDM4, LR4

100G (4x25G)
SR4, AOC, CWDM4, LR4

4-lasers
Laser Driver

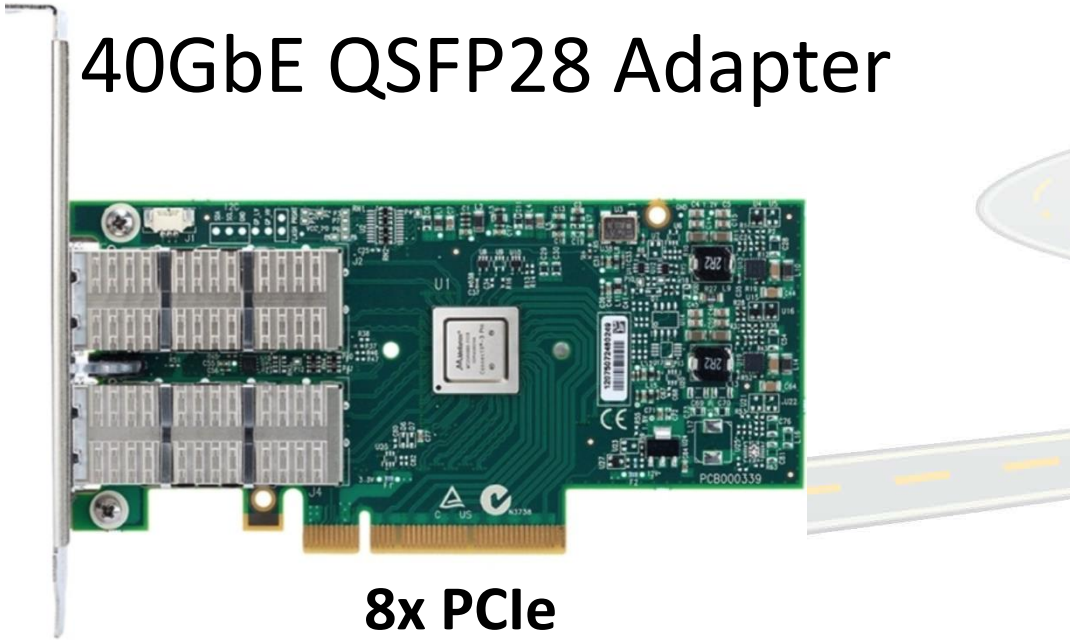
4 Detectors
Detector Amplifier

PCB
Power controller
Microcontroller
QFP shell

MPO optical connector
8-Multi-mode fibers

Switches & Network Cards –Almost the same BOM costs

40GbE QSFP28 Adapter



8x PCIe

100GbE QSFP28 Adapter

16x PCIe

32-port 40G Switch

32-port 100G Switch

Compelling 25G Economics vs 10G with Minimal Changes

- **25G Exploits the same hardware infrastructure as 10G**
- **Same 32-ports** in switch or **2-ports** in NIC configuration
- Same **QSFP/SFP/CXP** form factors

- Same **DAC** copper cable + a little more shielding
- Same **AOC** configurations
- Same **MPO** or Duplex **LC** optical connectors

- Same **fibers** – OM3/OM4 multi-mode and OS2 single-mode
- Same **Reaches:**
  - DAC drops from 7m to 3m at 25G (but most use <3m in the rack anyway)
  - Multi-mode (100m) & single-mode reaches stay the same (10Km+)

- **Soon, 25GBASE-T**

What's Driving the 10G-25G Transition? Compelling Economics

- **Costs:** 2.5X bandwidth at <2X increase in price
- **Tomorrow Future proofing:**
  - 25G line rates for today, 2x25G (50G), then 4x25G (100G)
  - 50G 2x25G = **4** fibers –vs- 4x10G = **8** fibers – lower costs
  - 25G/lane bandwagon for futures (PAM4)
- **Hardware infrastructure changes are minimal**
  - Electrical connectors improve
  - Shielding & PCB materials improve
  - Electronics and Lasers speeds increase

2.5X Speed/Bandwidth with Minimal Infrastructure Impact

10G/40G





25G/50G/100G





Most Common Interconnects Schemes In Modern Data Centers Today

Data center Interconnects 101: “Plugs”

SFP28

1-Channel
2 Fibers or wires
Small **F**ormFactor **P**luggable
1-1.5W
Duplex LC optical Connector

2-fiber Serial

“+” 10G;
“28” for 28G;
“56” for 56G

Both use MMF or SMF

QSFP28

4-Channels
8 Fibers or wires
Quad **S**mall **F**ormFactor **P**luggable
3.5W (5W future)
MPO 8-fiber parallel Optical connector

8-fiber Parallel

2-fiber Serial

Data Centers =SFP/QSFP “CXP”

Data center Interconnects 101: Wires and Fibers

Direct Attach Copper DAC "TwinAx"

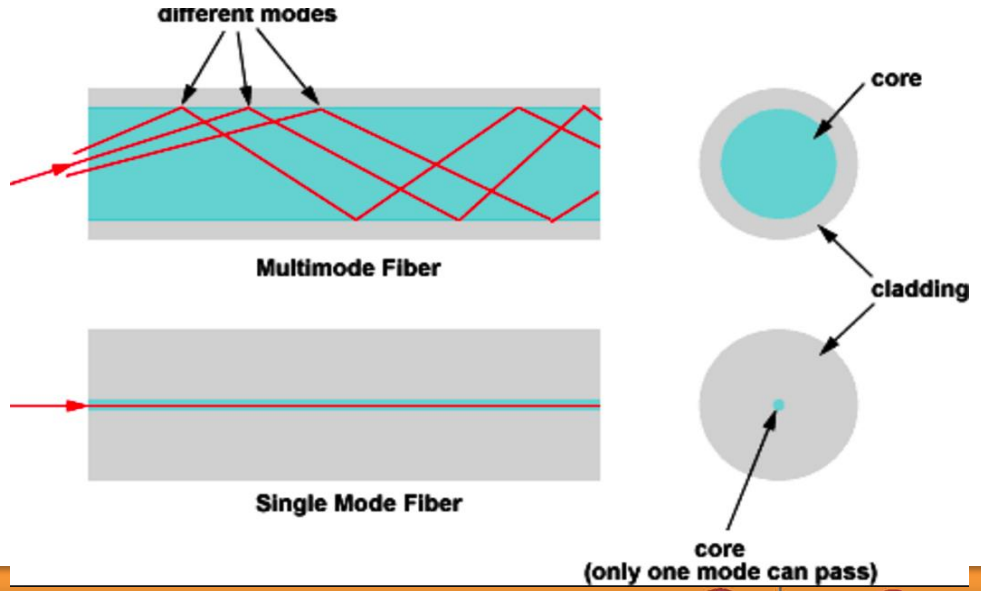
Copper wires & shielding
3m (9m) reach
2-wires/Channel

Multi-Mode Fiber

50-um Large core fiber
100m (300m) reach
Easy to attach components
Transceiver are low cost
Fiber 3x cost of SMF

Single-Mode fiber

9-um Tiny core fiber
2/10Kmreach
Hard to attach components
Transceivers are expensive
SMF cost less than dental floss!



Multi-Mode Fiber

Single-Mode fiber

Data center Interconnects 101:

Cables

DAC

3m

AOC

<30m

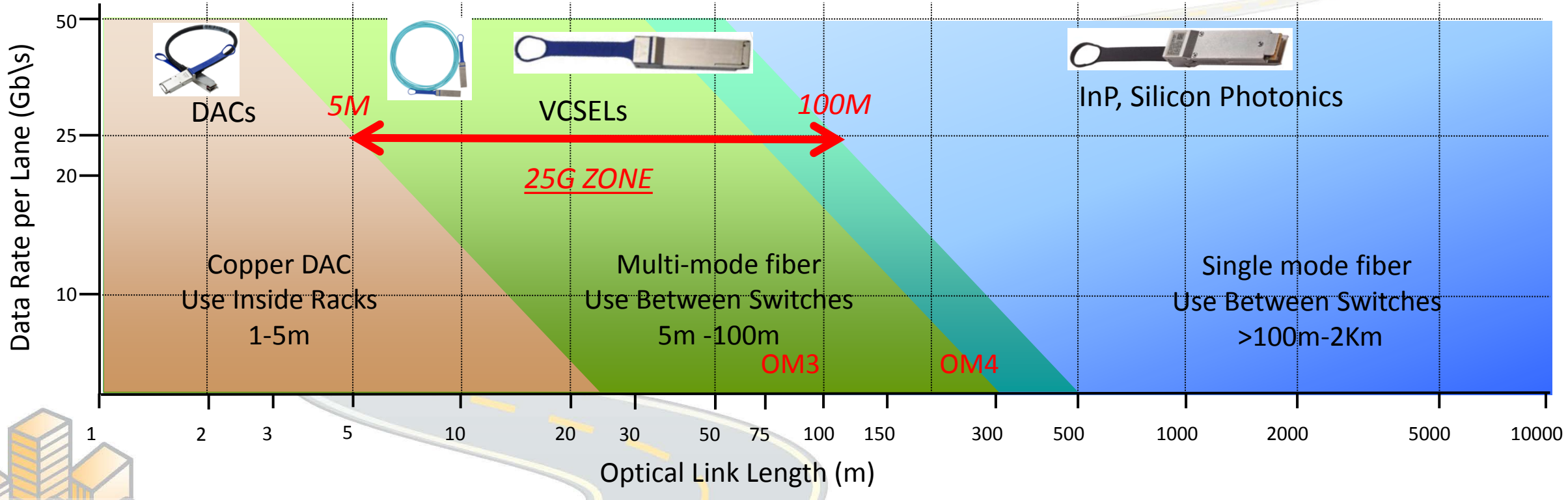
Transceivers

Multi-mode
VCSEL Laser
GaAs

Single-mode
FP, DFB or Ext Modulated Laser
InP, PLCs, Silicon Photonics

<100m

2Km, 10Km



Direct Attach Copper
- Zero power
- Best fit 3m

Active Optical Cables
- VCSEL 100m
- Best fit for 3-20m

SR4 VCSEL Transceivers
- Reaches to 100m
- Best fit for MMF
- Structured cabling

Silicon Photonics Transceivers
- Reaches to 2km
- Best fit for SMF
- Parallel PSM4 or WDM4

MMF= MULTI-MODE FIBER SMF = SINGLE-MODE FIBER

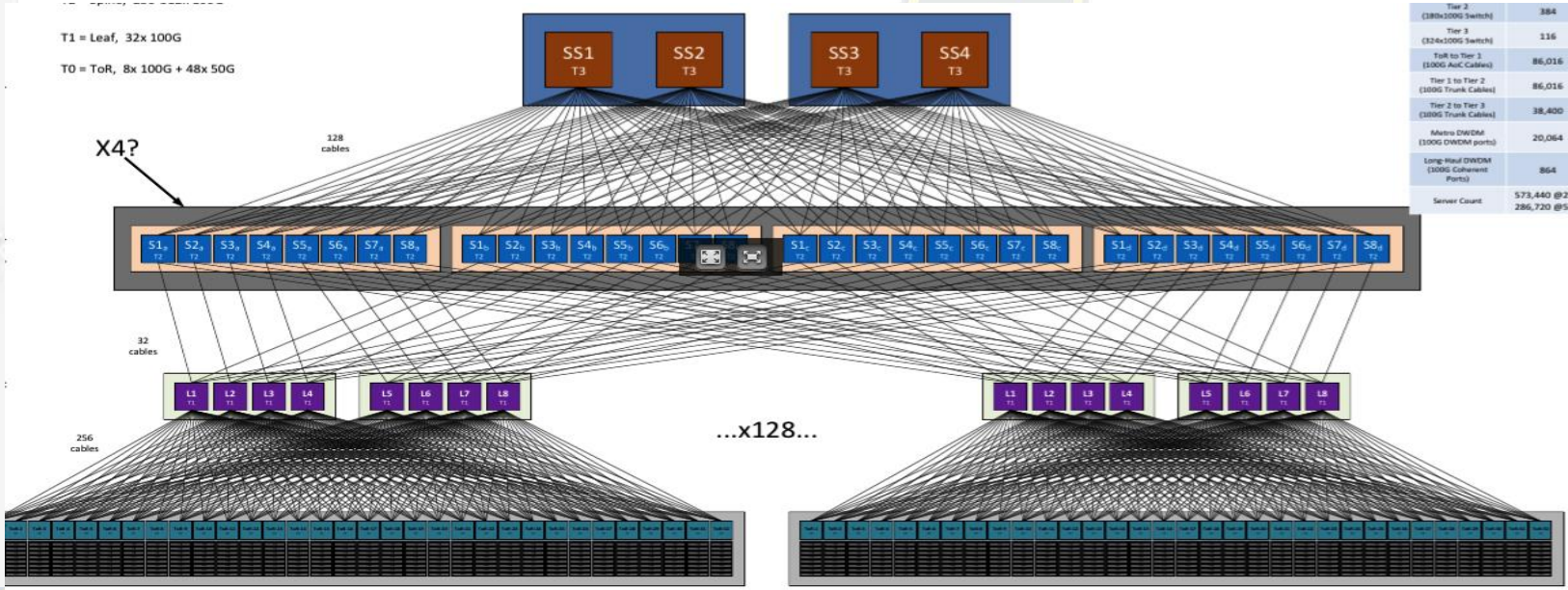
How 25G/50G/100G Interconnects are Deployed in Data Centers

Different Sized Data Centers

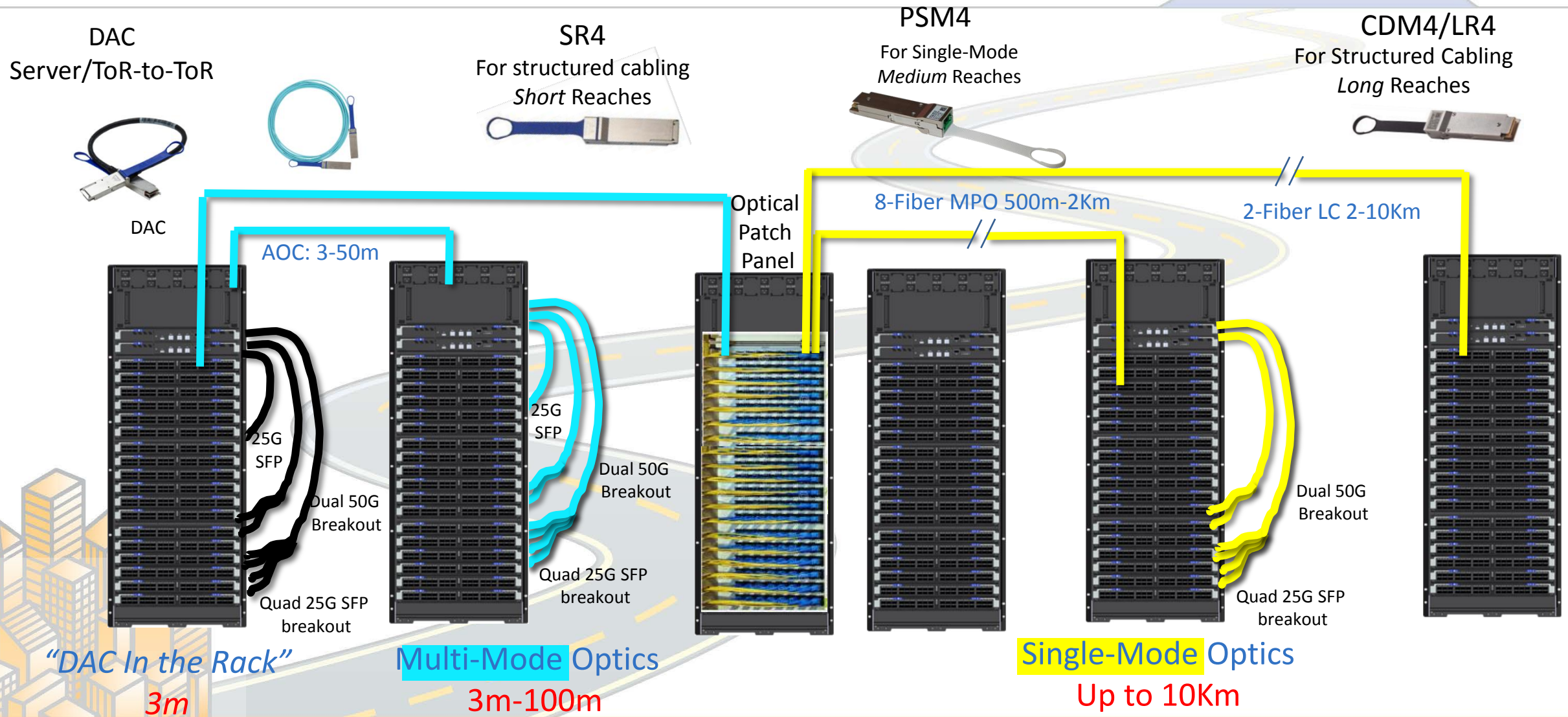
Multi-mode & DAC Territory
Reaches Typically < 100m

Hyperscale Data Centers

Single-Mode & DAC Territory
Reaches From 1-10Km

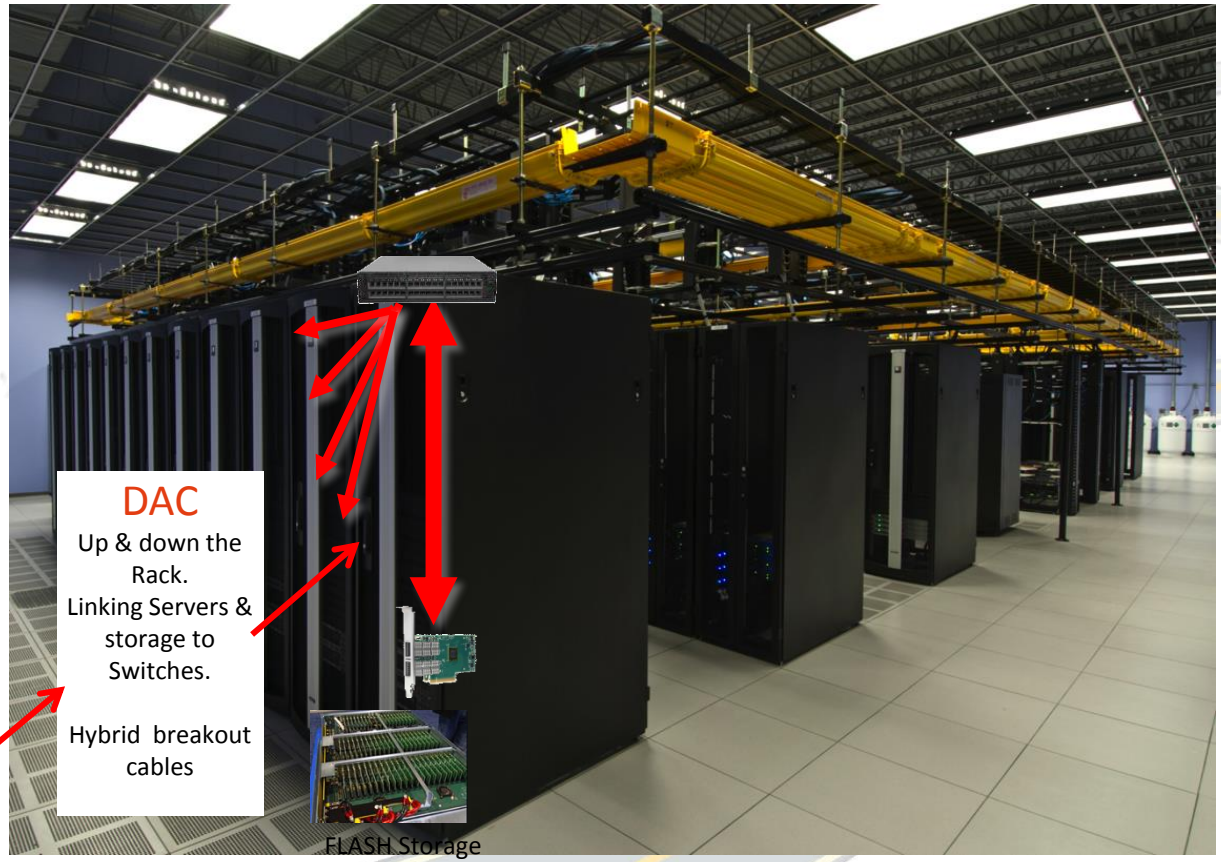


How Interconnects are Being Used in DC



Where Are DAC Links Used?

Up & Down the rack
Servers & Storage linked to
ToR Switches



HPCs may use AOCs everywhere
including up & down the rack

Where AOCs Are Used?

AOC

Between Switches over short reaches <20m where access is easy (cable trays)

HPCs may use AOCs everywhere including up & down the rack

Where SR4/MPO Links Used?

SR4

Transceivers to optical patch panels and in Structured Cabling Pipes & under raised floors where connectors are needed.
<100m

Under the Floors
Requires optical connectors & SR4s

SR4 - To Optical Patch Panels

SR4 - Into Structured Cabling

PSM4 too

Where PSM4, CWDM4 & LR4 Links Used?

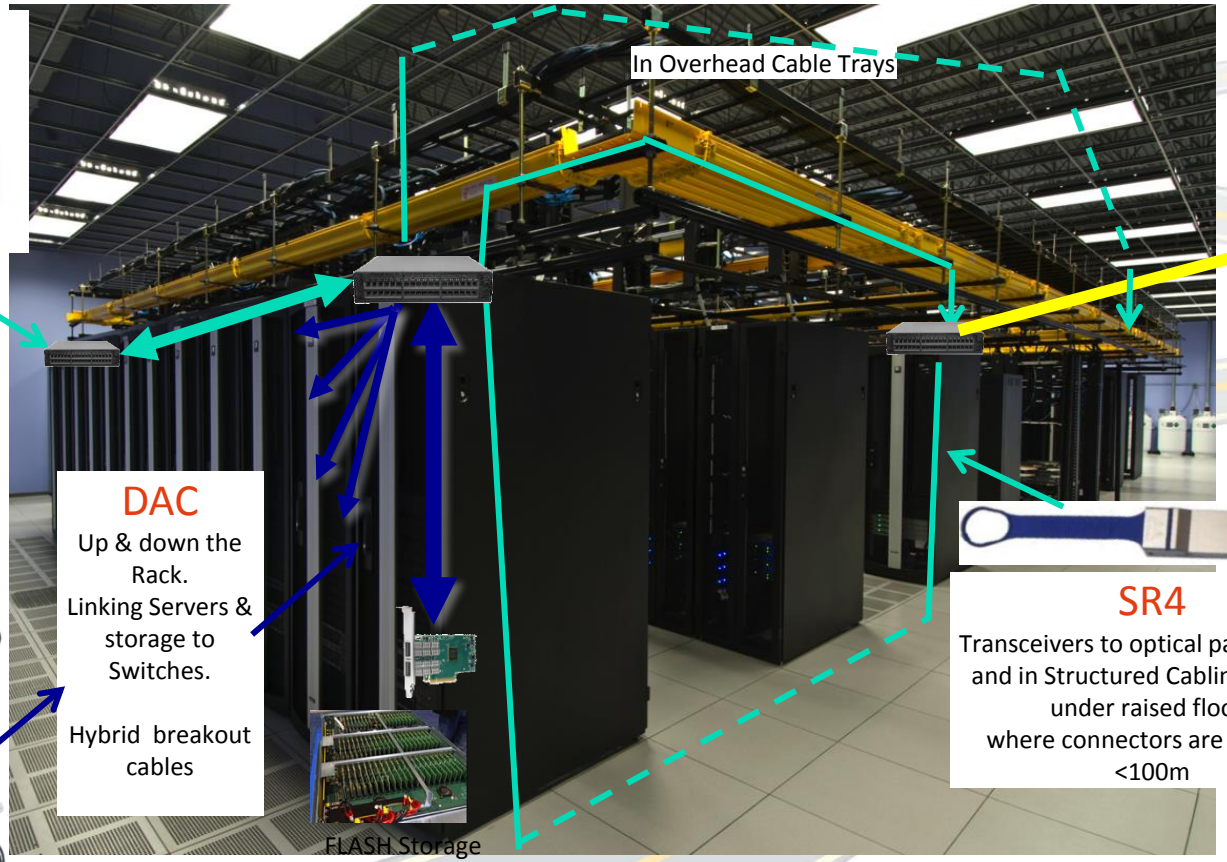
PSM4/CWDM4/ LR4

single-mode Transceiver linking to other buildings/floors up to 2Km/10Km.

All together 25G/50G/100G Links

AOC

Between Switches over short reaches <20m where access is easy (cable trays)

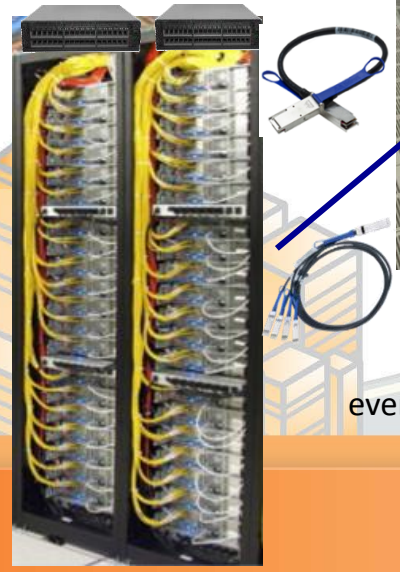


DAC

Up & down the Rack.
Linking Servers & storage to Switches.

Hybrid breakout cables

Up & Down the rack
Servers & Storage linked to ToR Switches



HPCs may use AOCs everywhere including up & down the rack

PSM4/CWDM4/LR4

single-mode Transceiver linking to other buildings/floors up to 2Km/10Km.

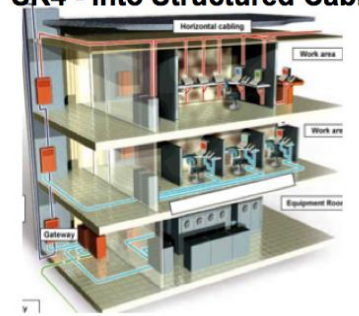
SR4

Transceivers to optical patch panels and in Structured Cabling Pipes & under raised floors where connectors are needed.
<100m

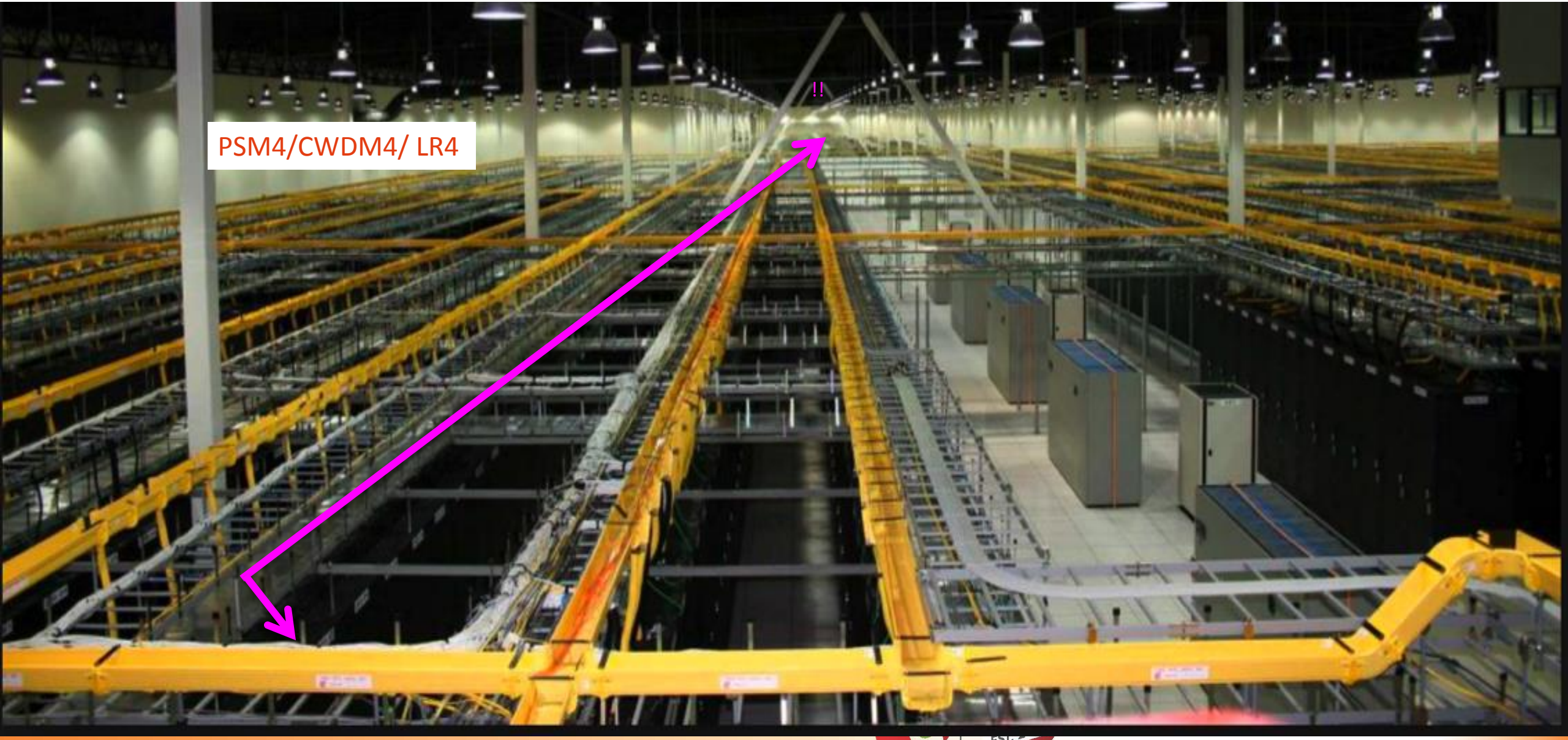
Under the Floors
Requires optical connectors & SR4s

SR4 - To Optical Patch Panels

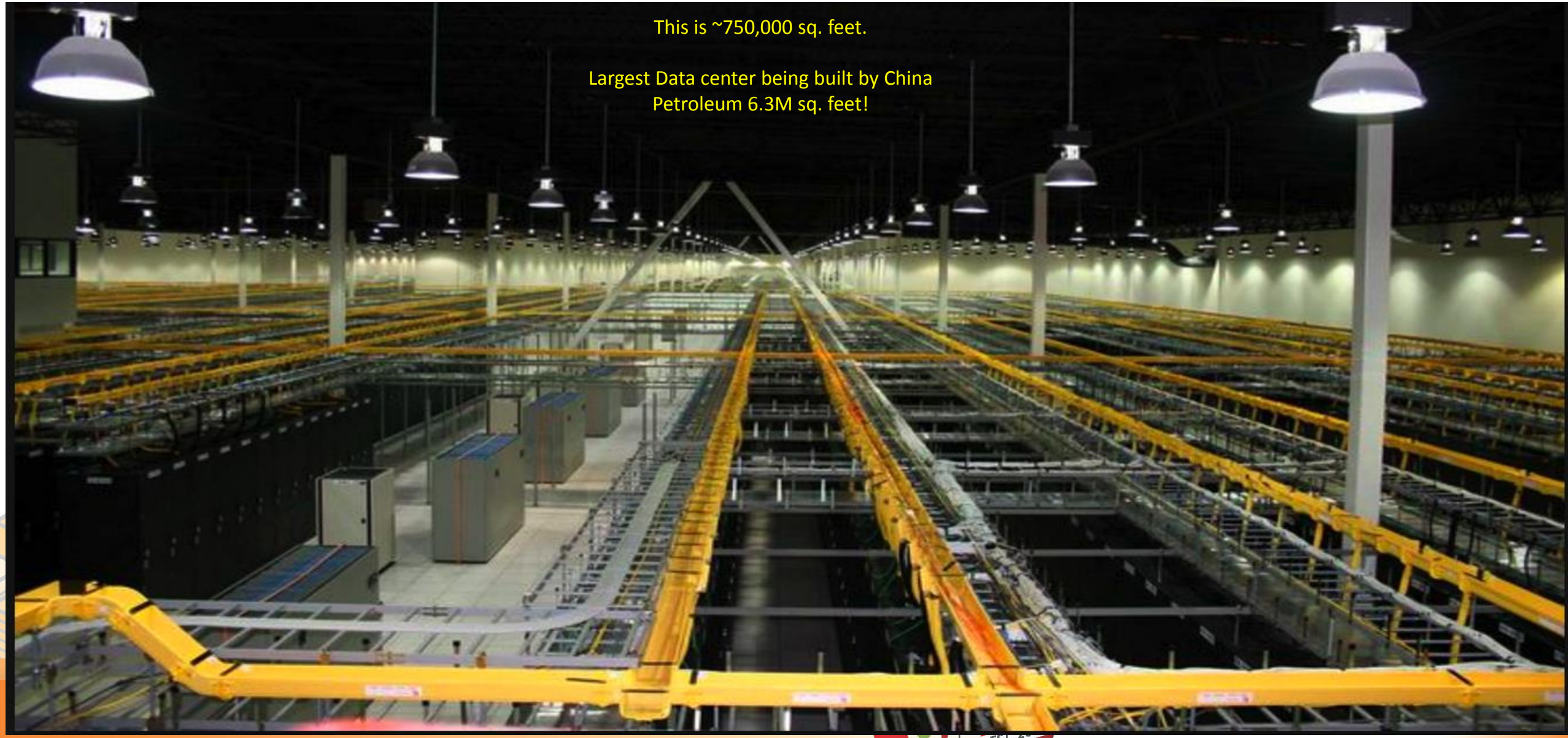
SR4 - Into Structured Cabling



Hyper Scale Data centers –Single-mode Territory



What it looks like IN THE OTHER DIRECTION!



ETHERNET OPTICS TOMORROW: 50G PAM4

The State of Ethernet Optics

Chris Cole, Finisar

March 23, 2016

OFC 2016

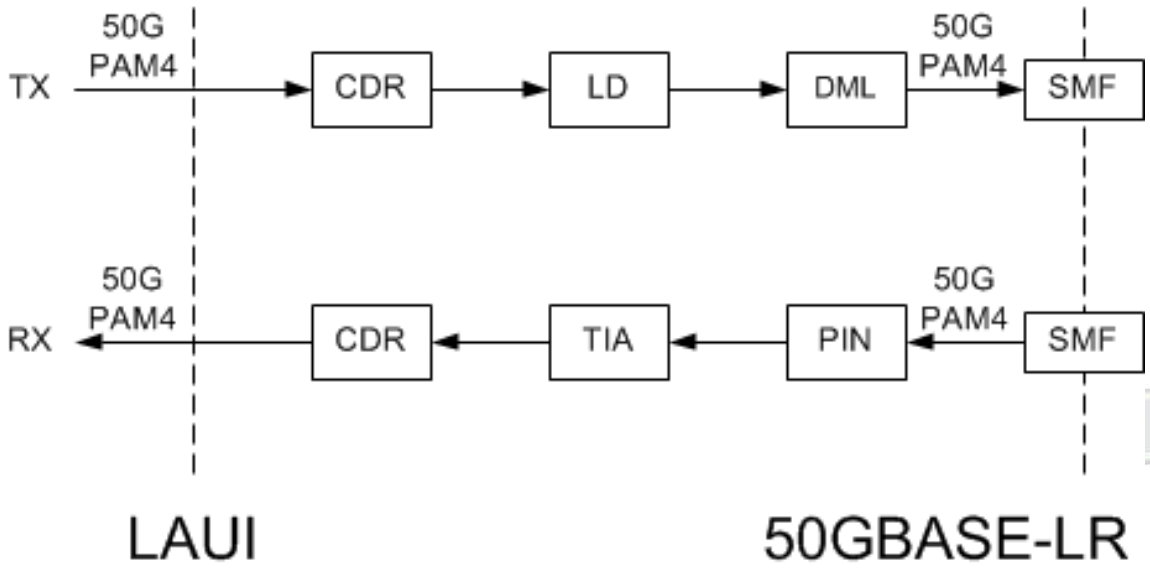
Anaheim, CA

ethernet alliance

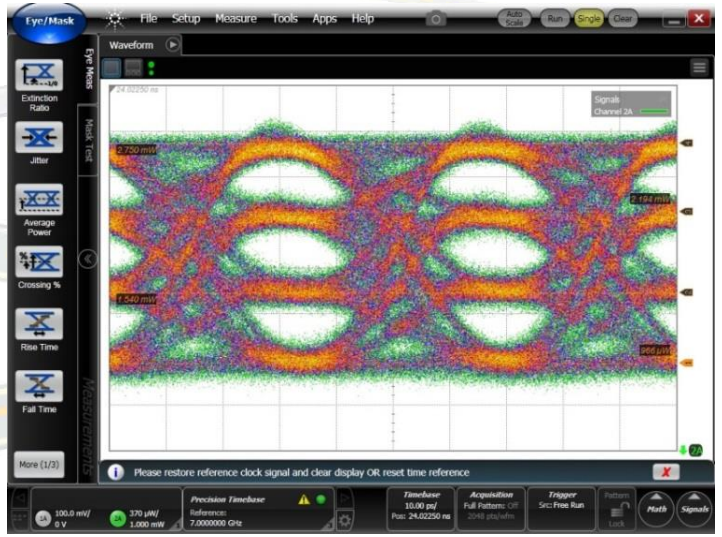
Disclaimer

- Opinions expressed during this presentation are the views of the presenters, and should not be considered the views or positions of the Ethernet Alliance.

50G 1310nm SMF Optics



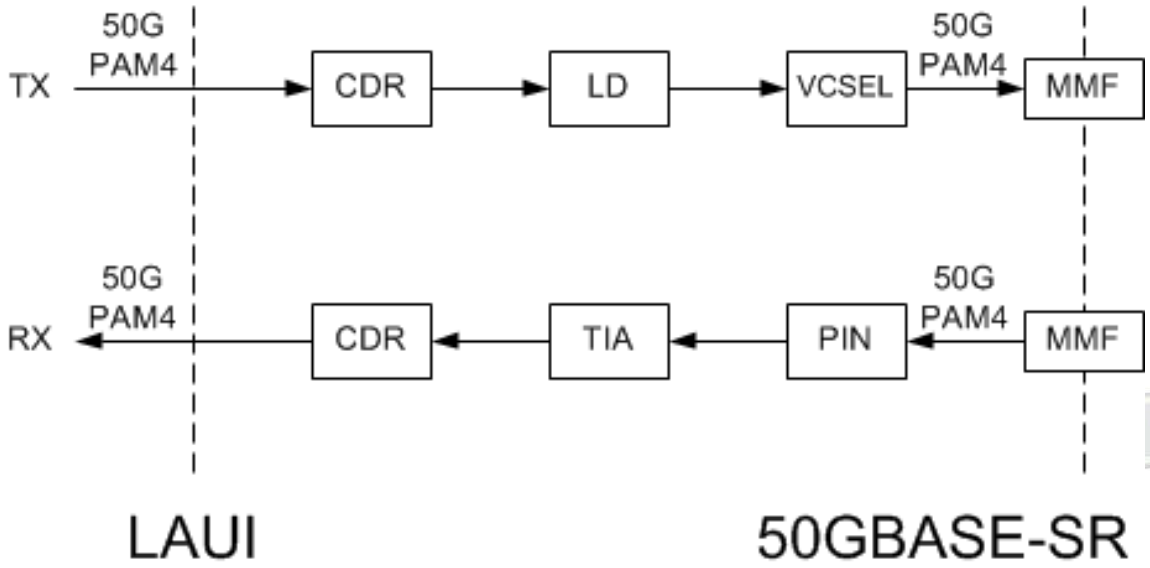
56Gb/s PAM4 optical eye



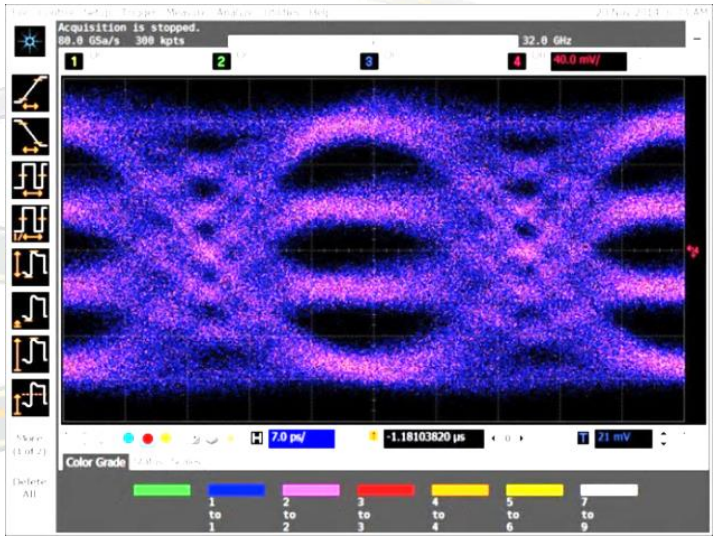
Form Factor: SFP56 w/ LC

Lane Rate	No. of Lanes		Data Rate
Gb/s	fiber pairs	λ	Gb/s
50	1	1	50

50G 850nm MMF Optics



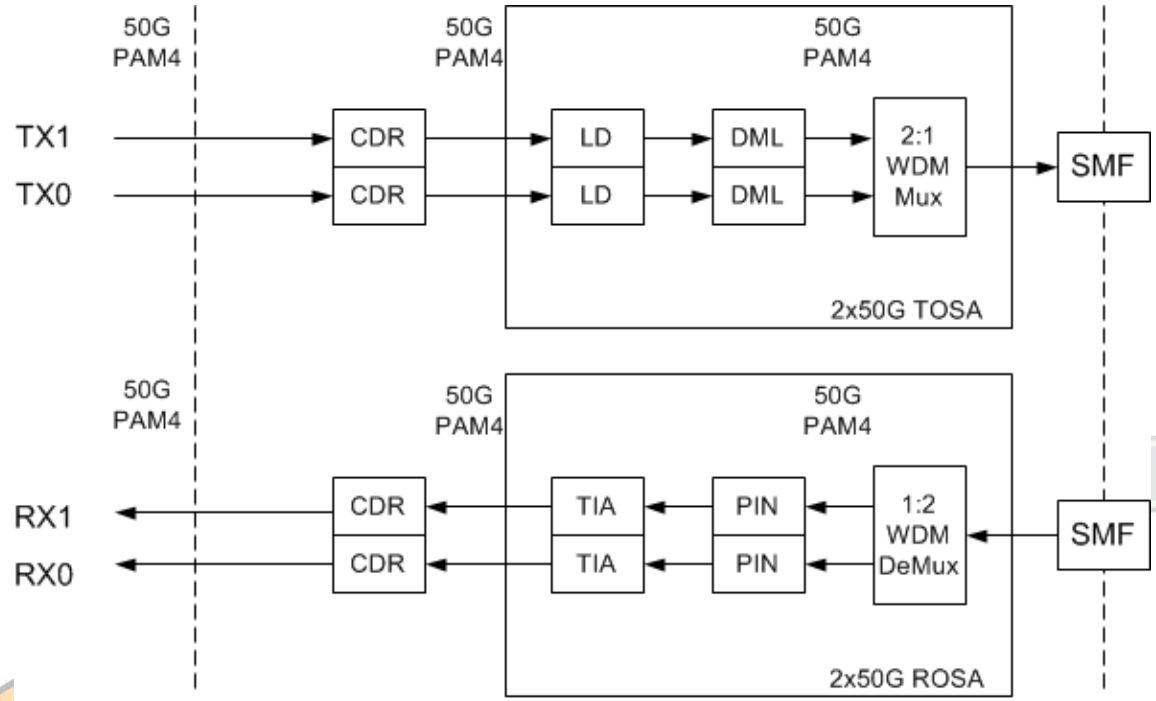
56Gb/s PAM4 optical eye



Form Factor: SFP56 w/ LC

Lane Rate	No. of Lanes		Data Rate
Gb/s	fiber pairs	λ	Gb/s
50	1	1	50

100G 1310nm SMF Optics

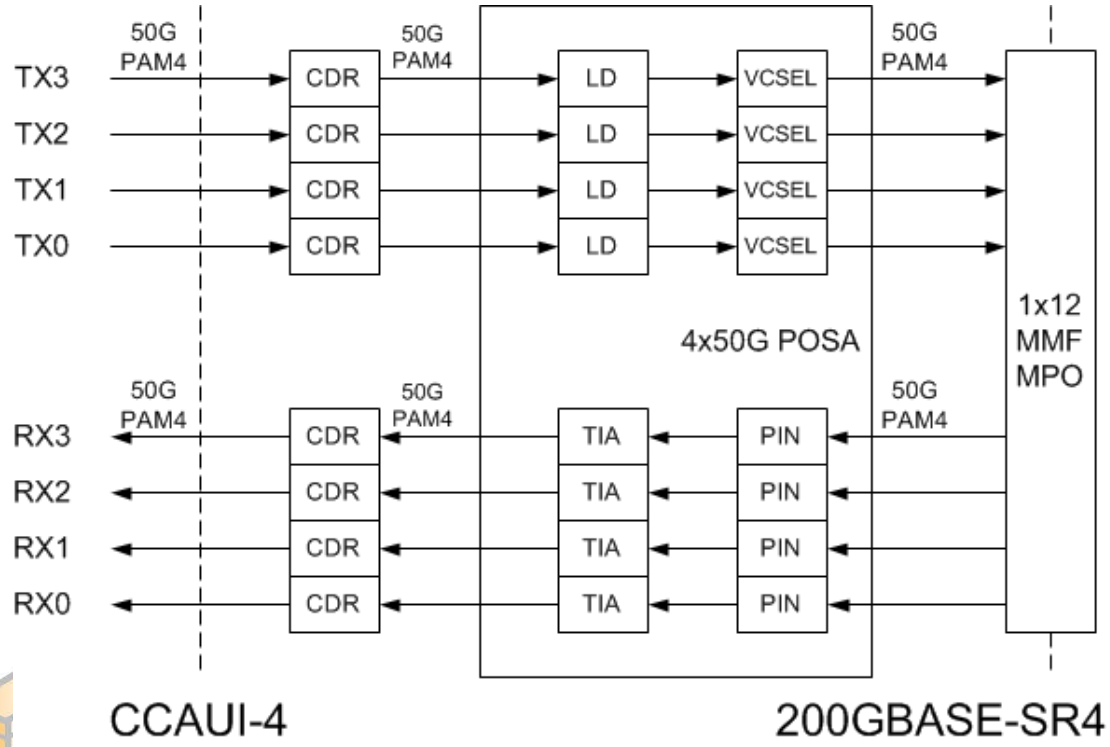


- Use with CAUI-4 (4x25G) I/O requires 4:2 Mux CDR
- QSFP56 can support two 100G WDM2 channels with MPO

Form Factor: QSFP56 w/ LC

Lane Rate	No. of Lanes		Data Rate
Gb/s	fiber pairs	λ	Gb/s
50	1	2	100

200G MPO 850nm MMF Optics



Applications:

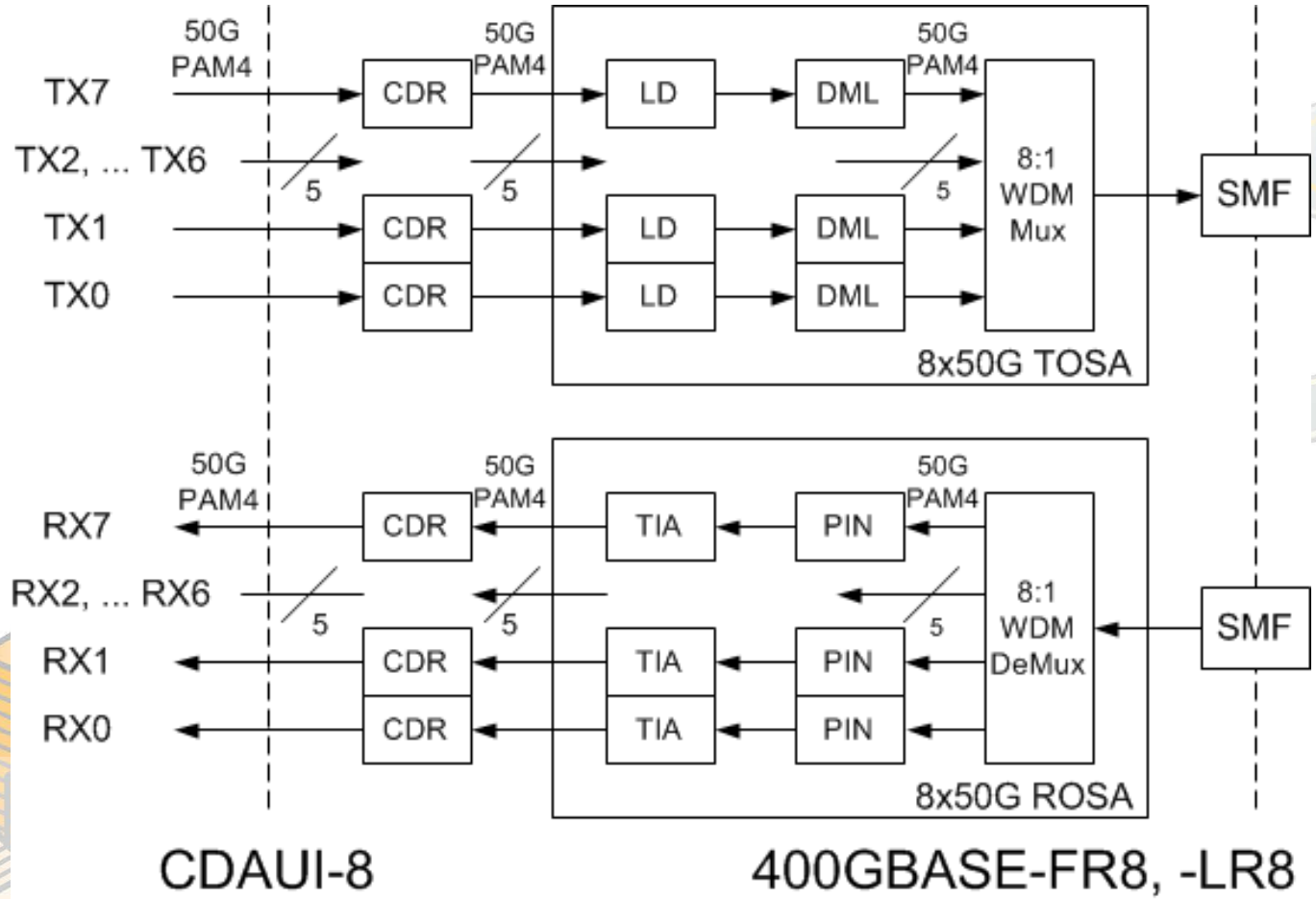
- 4x 50G SR
- 2x 100G SR2
- 1x 200G SR4

(PSM4 MPO similar use)

Lane Rate	No. of Lanes		Data Rate
Gb/s	fiber pairs	λ	Gb/s
50	4	1	200

Form Factor: QSFP56 w/ MPO

400G 1310nm SMF Optics



Potential Form Factors:

- CFP8
- CFP16
- QSFP-DD
- OSFP

Lane Rate	No. of Lanes		Data Rate
Gb/s	**fiber pairs**	**λ**	**Gb/s**
50	1	8	400

50G PAM4 Ethernet Optics

Lane Rate	No. of Lanes		Data Rate	SW code	LW code
Gb/s	fiber pairs	λ	Gb/s	(MMF)	(SMF)
50	1	1	50	SR	LR
50	2	1	100	SR2	PSM2
50	1	2	100	SWDM2	LR2
50	4	1	200	SR4	PSM4
50	1	4	200	SWDM4	FR4, LR4
50	1	8	400		FR8, LR8

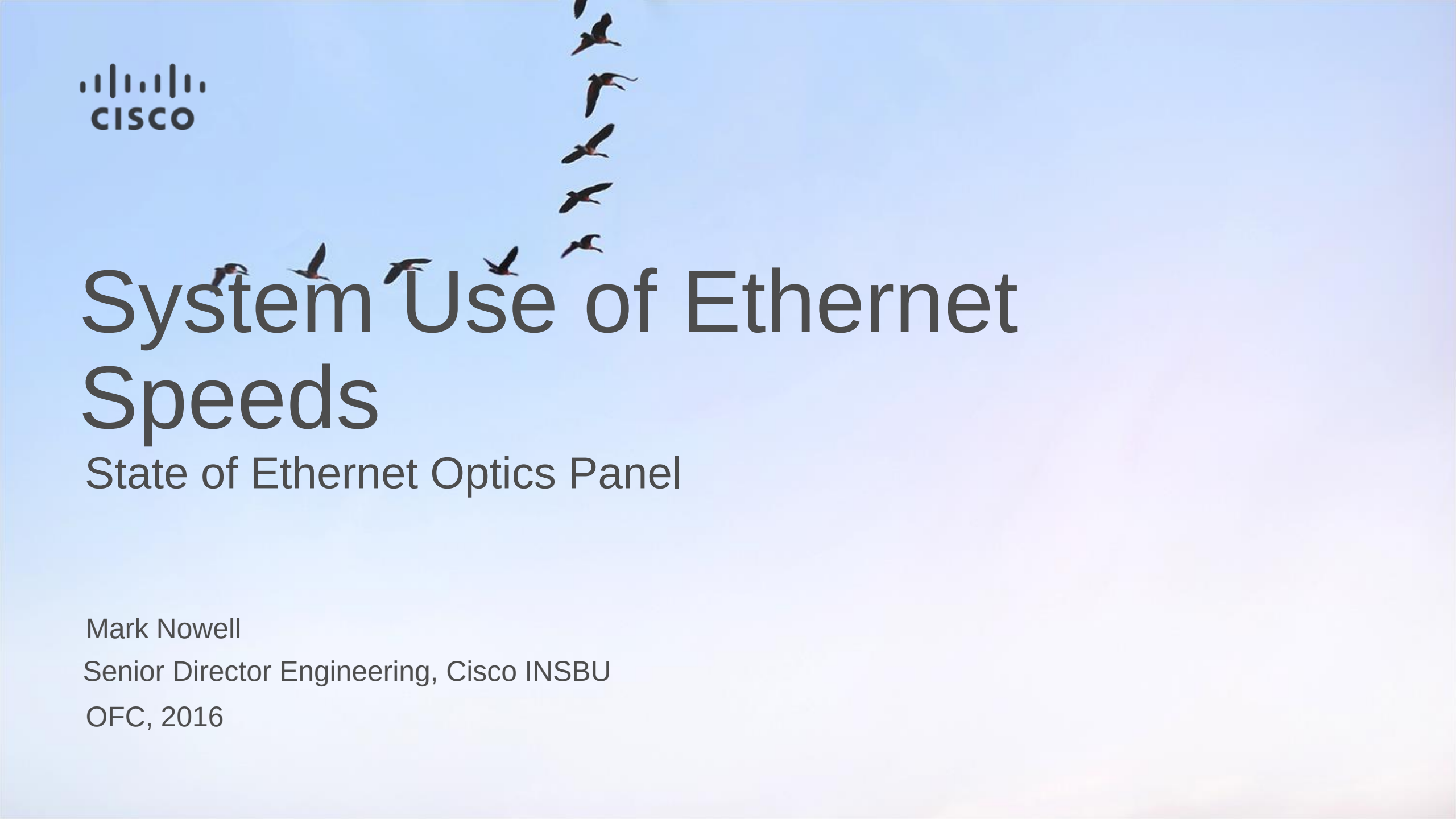
If you have any questions or comments, please email admin@ethernetalliance.org

Ethernet Alliance: visit www.ethernetalliance.org

Join the Ethernet Alliance LinkedIn group

Follow @EthernetAllianc on Twitter

Visit the Ethernet Alliance on Facebook



System Use of Ethernet Speeds

State of Ethernet Optics Panel

Mark Nowell

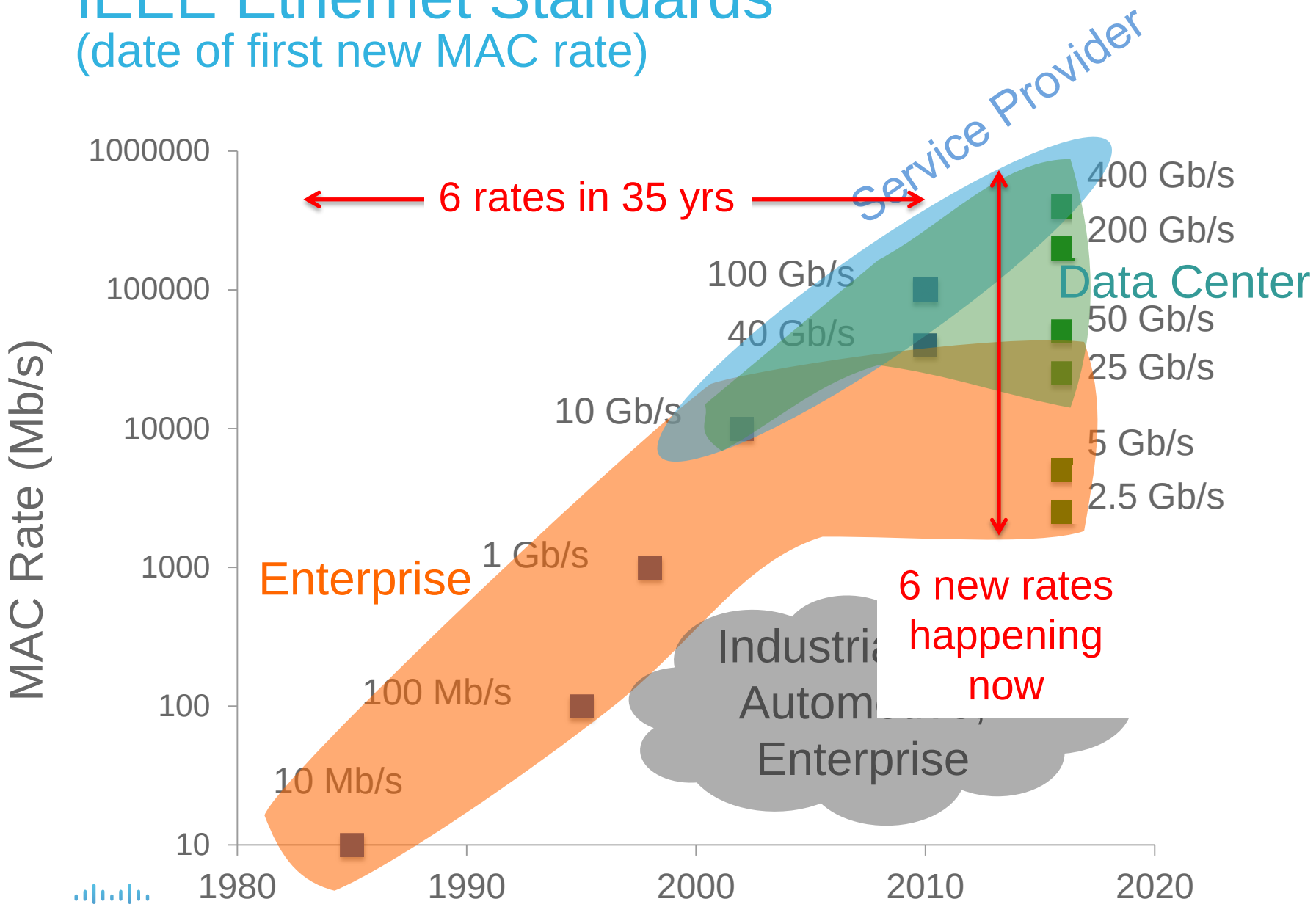
Senior Director Engineering, Cisco INSBU

OFC, 2016

Agenda

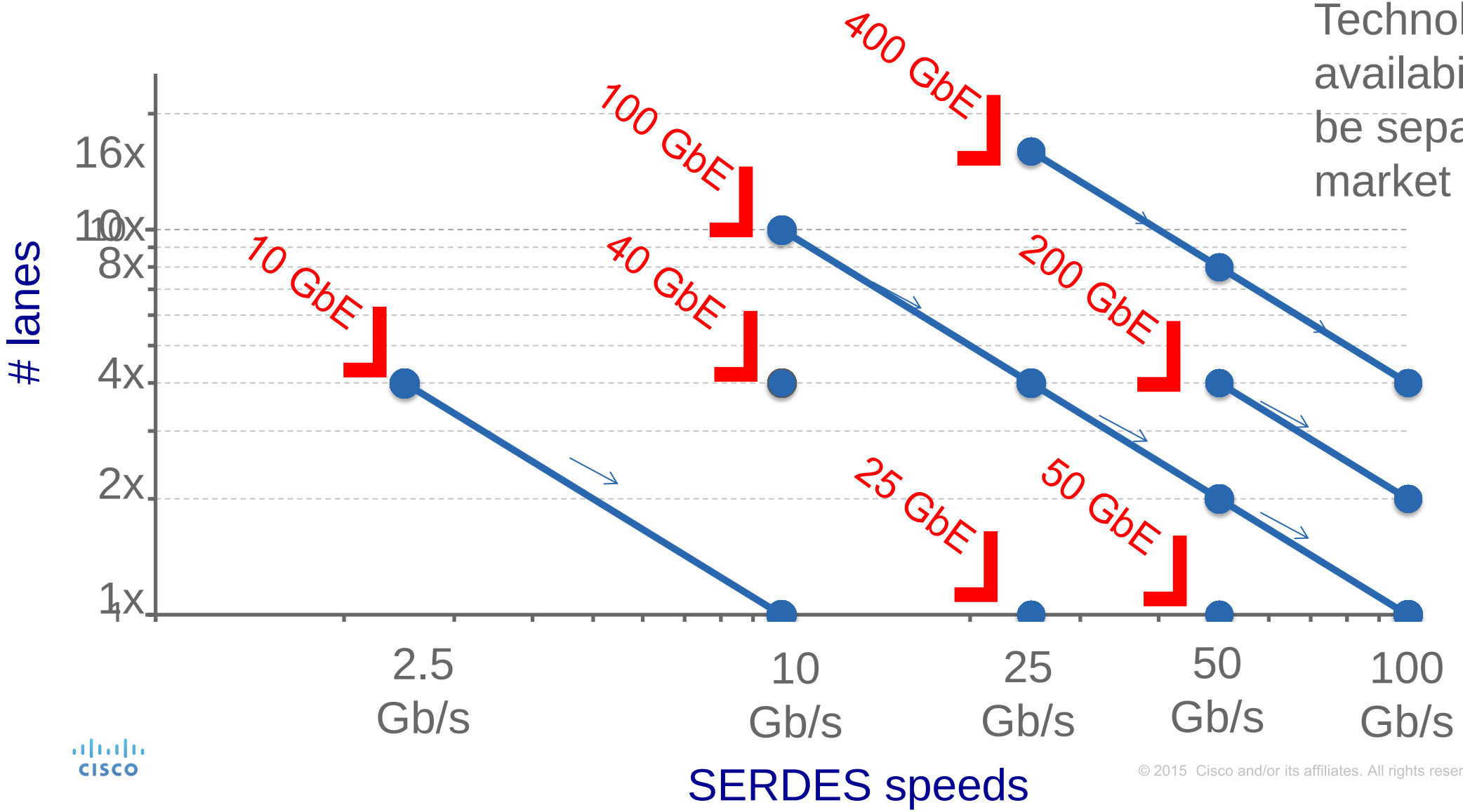
- Port density requirements
- Ethernet Rates
  - ASIC/PHY
- Form Factors
- Optics
  - Standard Optics vs. MSA Optics

IEEE Ethernet Standards
(date of first new MAC rate)



*Only shows the first time a new rate is standardized. Many subsequent variants are standardized over time.

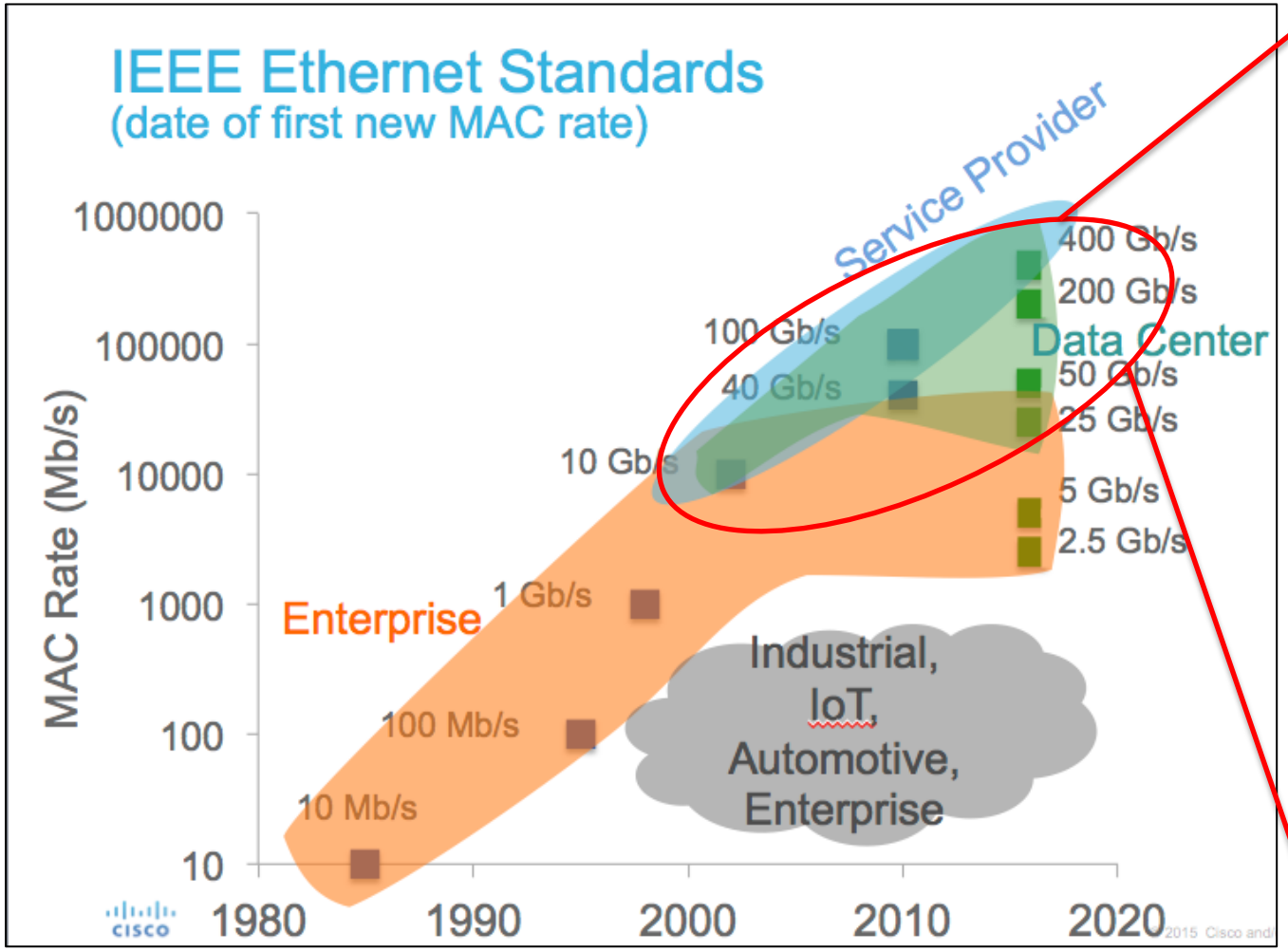
Why can so much be happening at once? **SERDES reuse**



Technology availability can not be separated from market direction

47

The impact of Cloud Data Center



Operational challenges

- Scale and upgrade requirements drove new architectures (aka Leaf-Spine)

Interconnect Challenges

- New architectures droves higher interconnect density
- Flatter hierarchy – more homogeneity

Ecosystem Challenges

- High volume
- Quicker cadence
- Higher emphasis on cost & density
- Switch ASIC re-architecture

Building High Density Systems for Cloud DC applications

High Density switching Silicon

- Gen1: 40 GbE ports w/ 10G serdes
- Gen2: 100 GbE ports w/ 25G serdes
- Gen3: higher ports counts 100GbE w/ 25G serdes
- Gen4: 200 GbE / 400 GbE w/ 50G serdes

High Density Pluggable Form Factors

- SFP & QSFP are the work horse form factors
  - Everything else is transitory

Scaling Switch Silicon to meet market needs

Application requires high port count silicon and high density interconnect

→ DC market initially adopted 40 GbE
  → It was the only high density switch silicon option.
  → Single lane 10GbE server IO & virtualization.

→ Current market need is dense 100 GbE
  → 25Gb/s serdes available → single lane 25 GbE servers

→ Next market need is dense 400 GbE
  → 50 Gb/s serdes coming. Single lane 50GbE servers will align

4x was a consequence of market need and technology availability

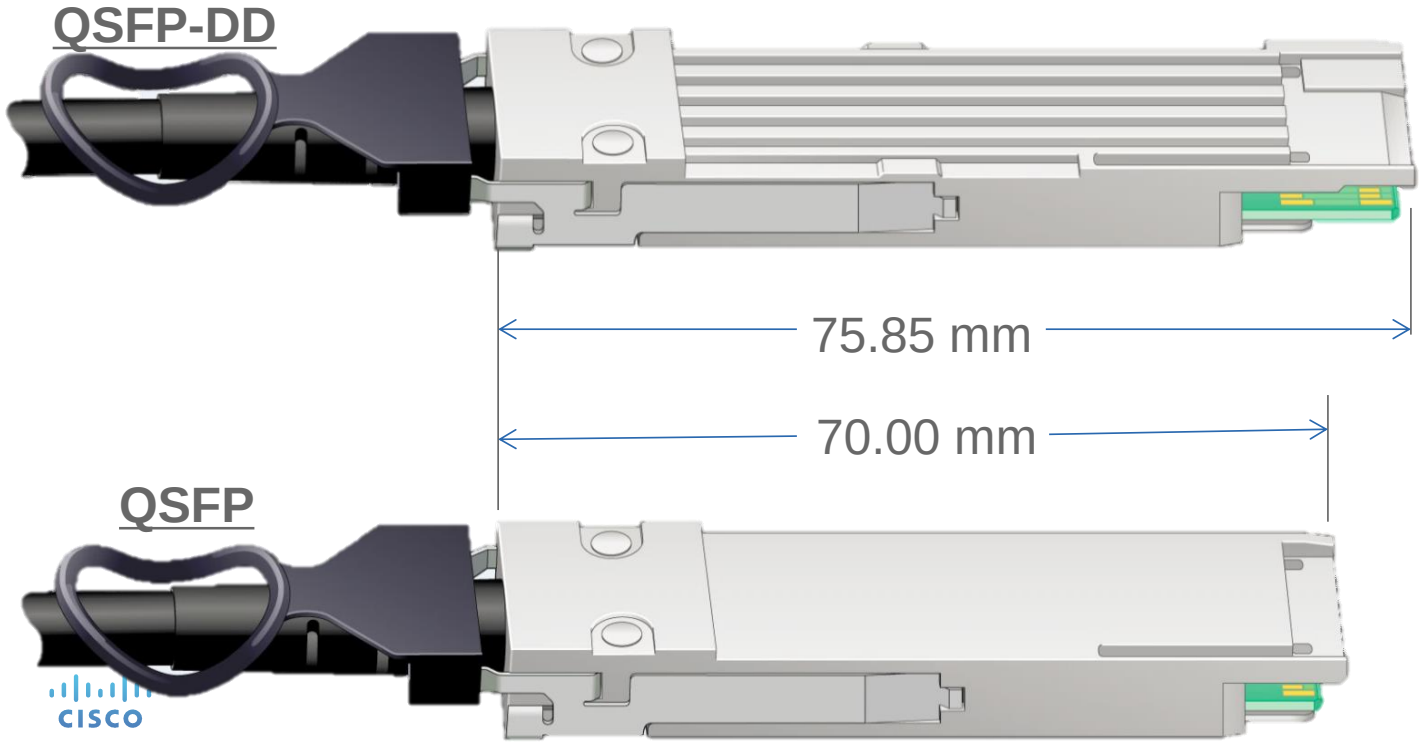
Pluggable Form Factors

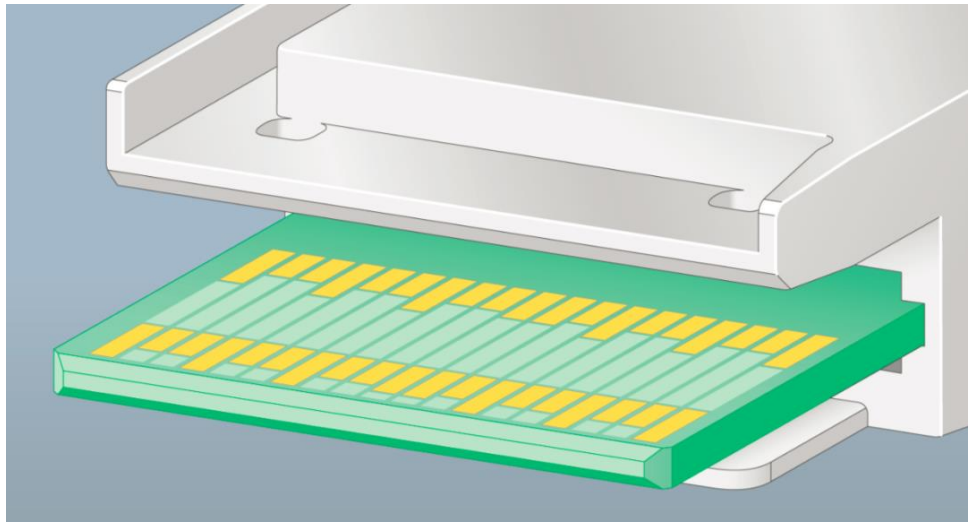
- Pluggable Form Factors continue to be the norm
- SFP & QSFP provide system densities consistent with Cloud DC architecture requirements
- Backwards compatibility offers huge value
  - Customer flexibility/refresh cycles
  - System design re-use
  - Economies of scale
- A key enabler for highly dense 100G and 400G is an upgrade for QSFP…

Introducing QSFP-DD (new 2x 100 GbE and 400 GbE capable pluggable module)

QSFP-DD

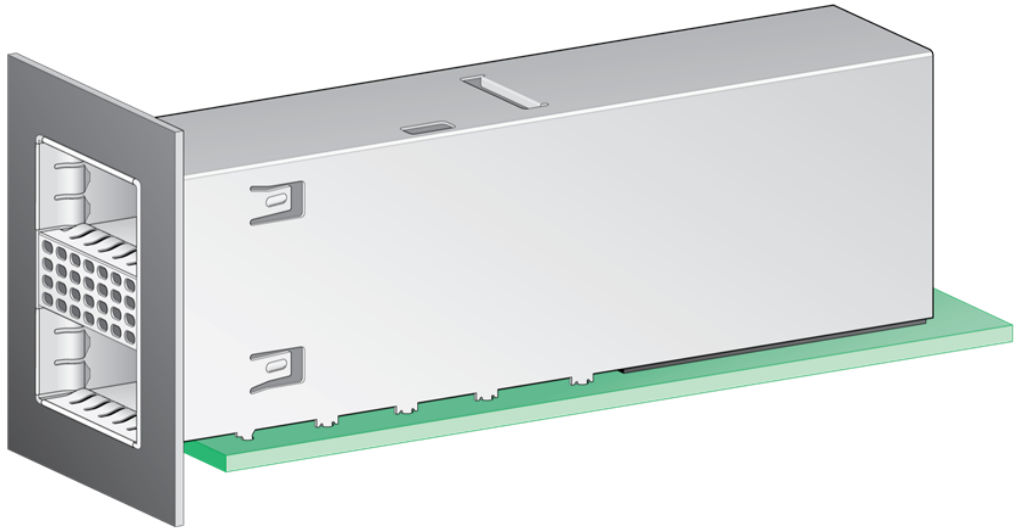
Improved thermals supports >2.5x QSFP power





Essentially the same as QSFP but with extra row of contacts. Allows boards to be backwards compatible to both.

QSFP-DD

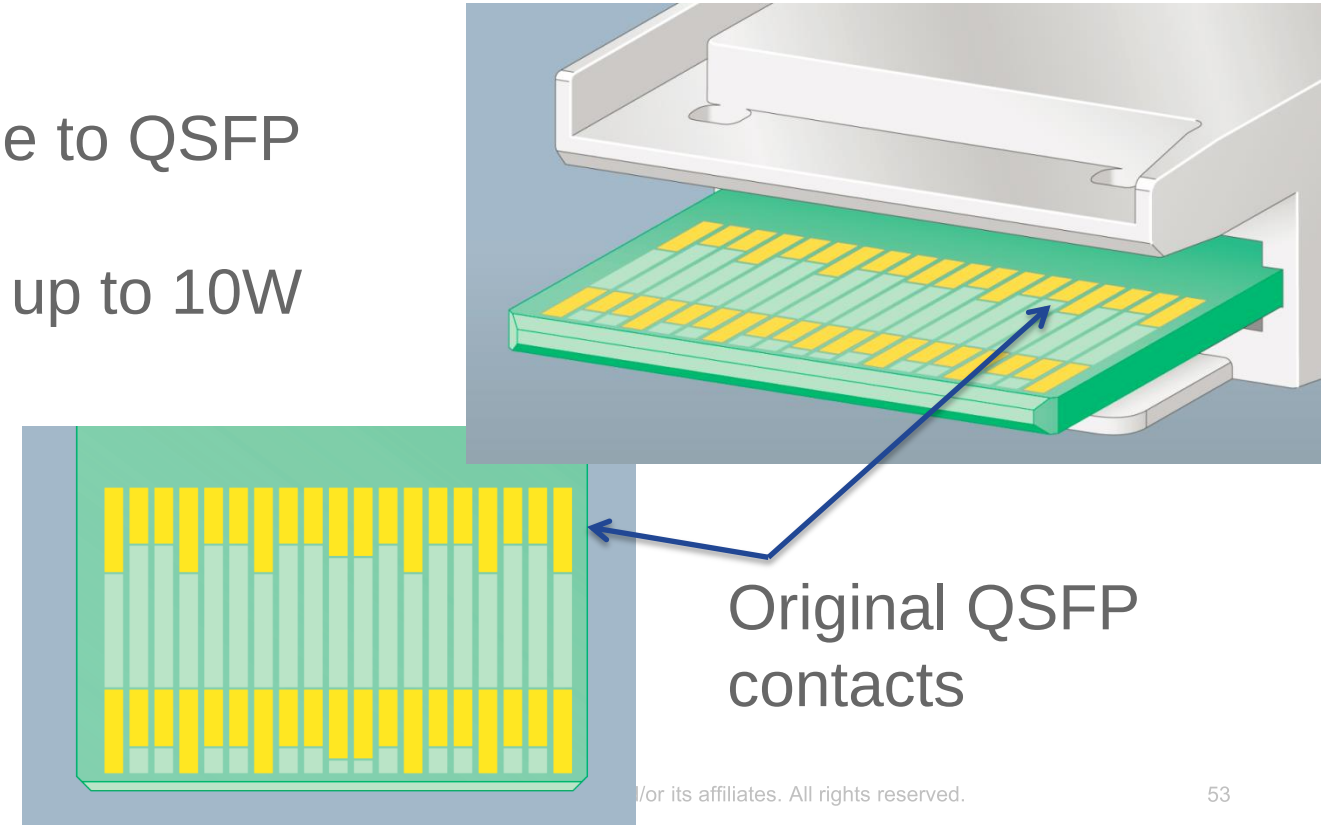


Supports 8 electrical IO
- 8x50G (CDAUI) → 400 GbE, 8x 50GbE
- Dual 4x25G (CAUI) → 2x 100 GbE

Host System fully backwards compatible to QSFP

Cooling/Thermal improvements enable up to 10W
- Advanced 2x1 cage design

MSA has been announced
- 13 founding companies
- Spec under development



Original QSFP contacts

Ethernet Optics: Standards vs. MSA

- 100 GbE optics has been unique in its breadth of options
  - IEEE has been unable to define additional specs
  - Transition to 3rd party optics to end users provides limited refinement
  - Multiple MSAs form to promote solutions
- Diluted supplier investments and volume
- System vendor view
  - Identify form factor requirements
  - Able to qualify multiple variants –takes time & energy – focus on customer needs
  - Greater concern is the dilution of resources that slow the cost reduction curve

Summary

- Market applications drive technology
- Technology does not drive a market but can enable a market
- Cloud DC Market is looking for 400 GbE and dense 100 GbE
  - Form factor identified
- Innovation required to address cost/integration challenges
- Finally… Cloud DC isn't the only market, it's just the newest. Do not overlook the high volume Enterprise markets.

If you have any questions or comments, please email
admin@ethernetalliance.org

Ethernet Alliance: visit www.ethernetalliance.org

Join the Ethernet Alliance LinkedIn group

Follow @EthernetAllianc on Twitter

Visit the Ethernet Alliance
on Facebook