Trends in 400G Optics for the Data Center

Christian Urricariet

NANOG 75

San Francisco, February 2019

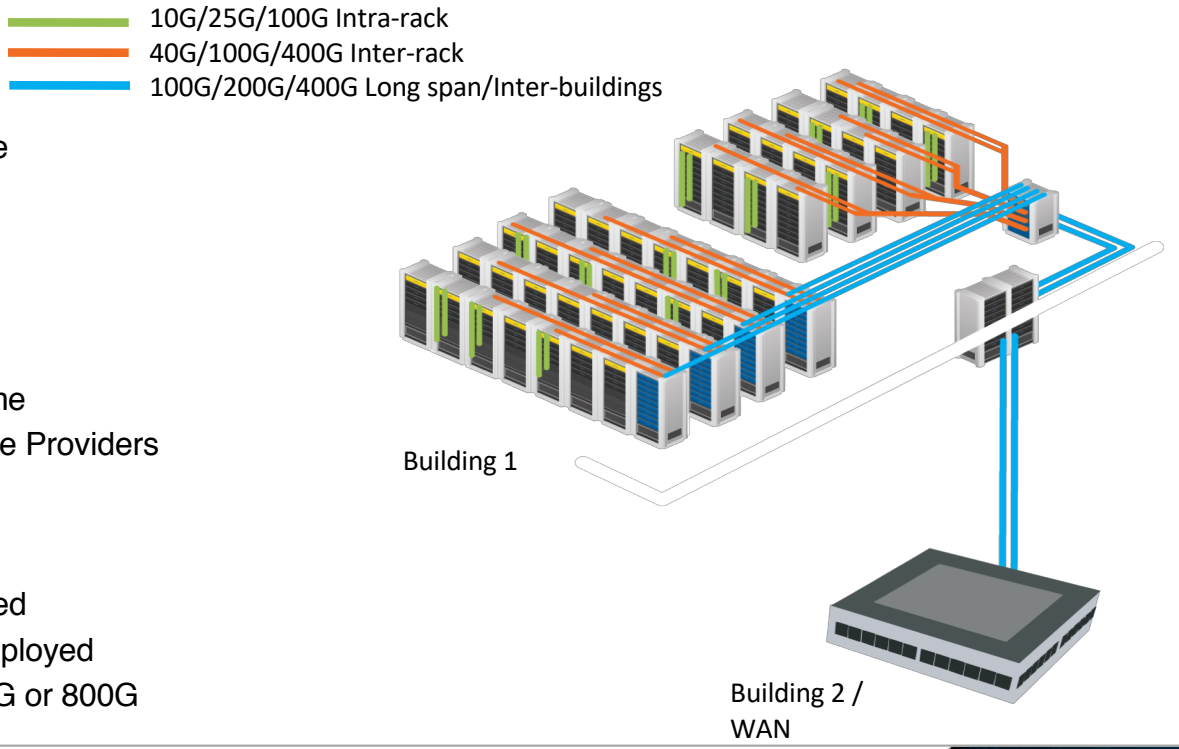
Data Center Connections are Driving Optics Volume

- Due to the ongoing large increase in bandwidth demand, Data Center connections are expected to move from 25G/100G to 100G/400G

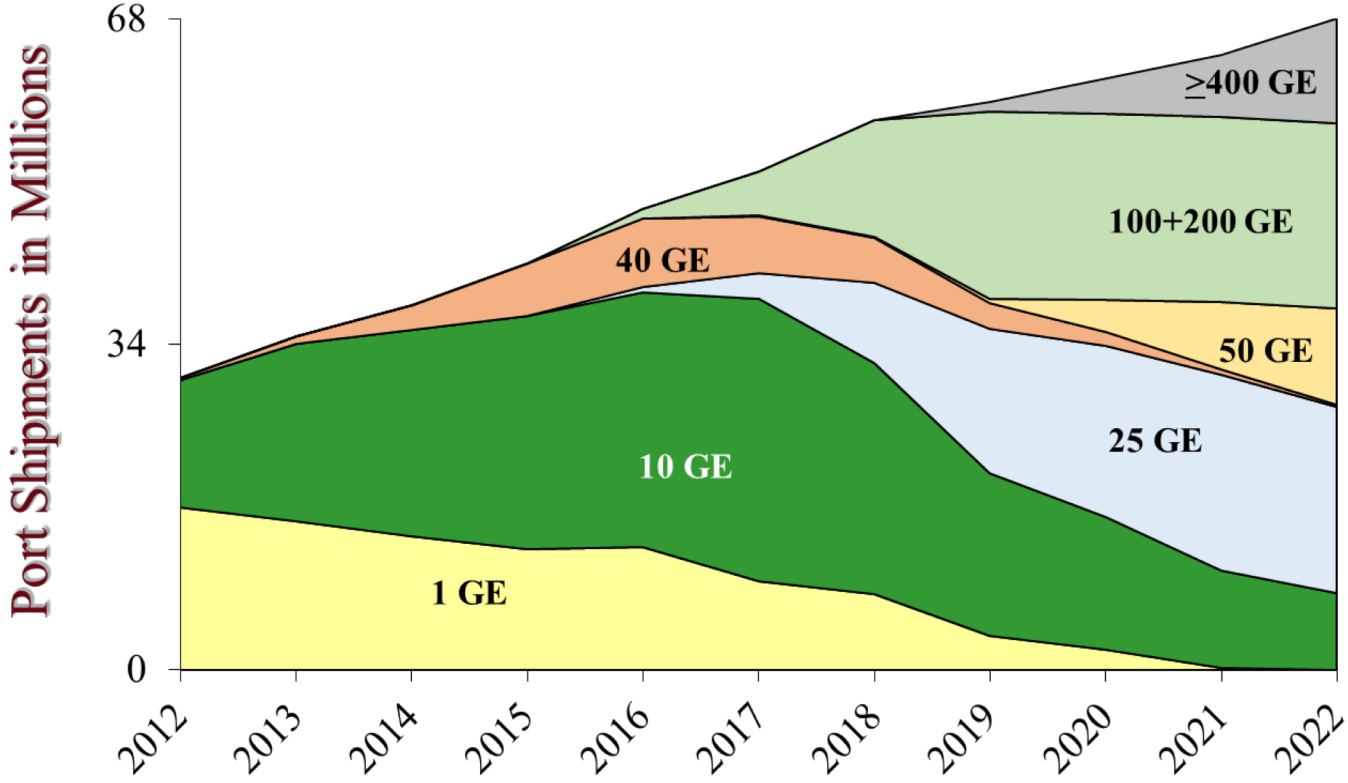
- Within the Data Center Racks
  - 10GE still being deployed
  - **25GE** starting to be deployed in volume
  - **100GE** (or 50G) will follow

- Between Data Center Racks
  - 40GE still being deployed
  - **100GE** starting to be deployed in volume
  - **400GE** will follow at large Cloud Service Providers

- Long Spans/DCI & WAN
  - 10G DWDM/Tunable still being deployed
  - **100G/200G Coherent** starting to be deployed
  - **400G Coherent** will follow – Then 600G or 800G

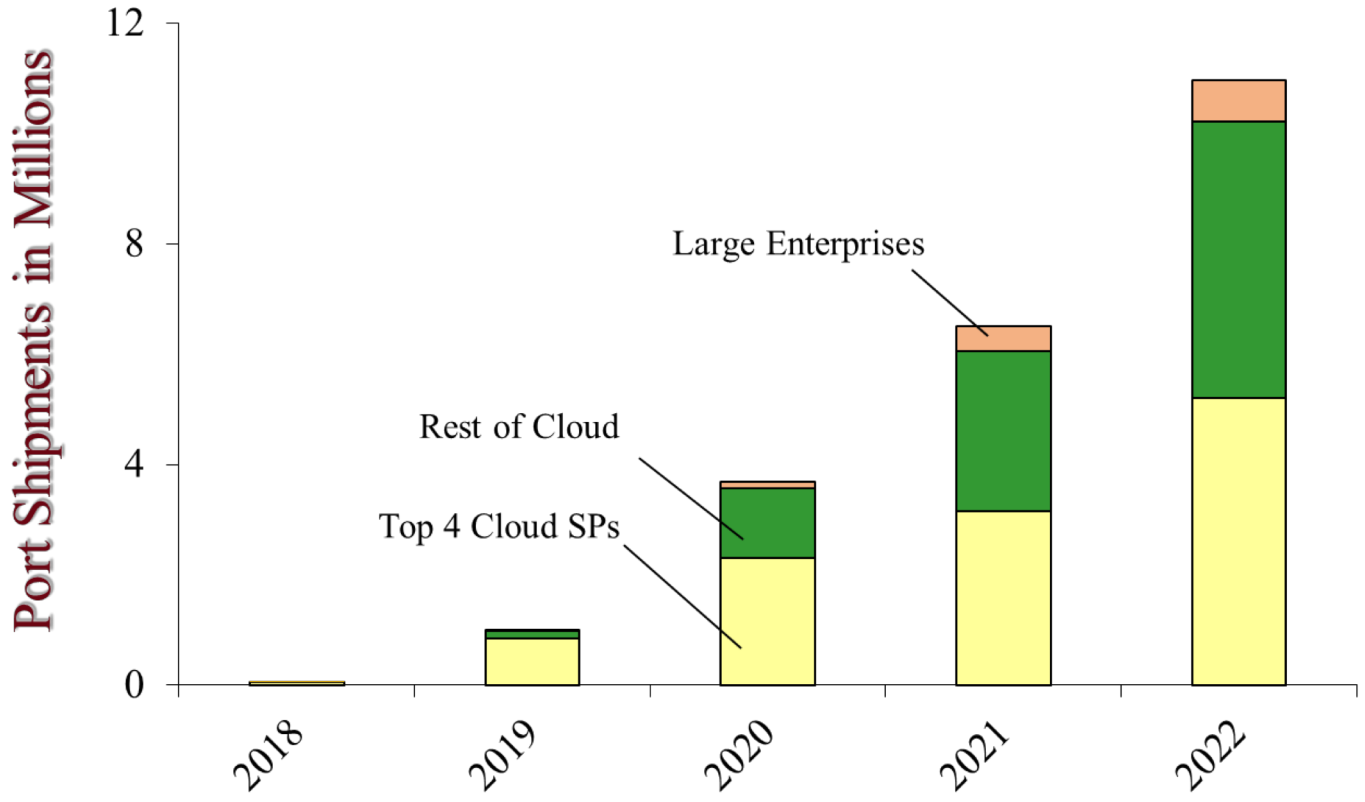


Forecasted Data Center Ethernet Port Shipments



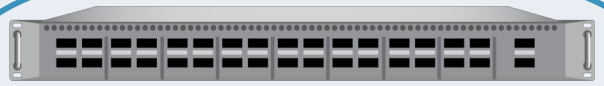
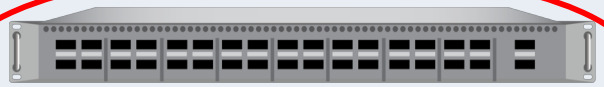
Source: Dell'Oro, 2018

Forecasted 400GE Shipments by Market Segment



Source: Dell'Oro, 2018

Mainstream 1RU Ethernet Switch Roadmap

First Deployed	Electrical I/O [Gb/lane]	Switching Bandwidth	TOR/Leaf Data Center Switch Configuration	
~2010	10G	1.28T	 32xQSFP+ (40G)	
~2015	25G	3.2T	 32xQSFP28 (100G)	3.2Tb/s switches based on 100G QSFP28 modules being deployed in cloud data centers today.
~2019	50G	6.4T	 32 ports of 200G	Given the multiple switching ICs expected to be available, the market is likely to be fragmented in the future.
~2020	50G	12.8T	 32 ports of 400G	

Large growth in bandwidth demand is pushing the industry to work on technologies and standards to support future 12.8T switches.

400G and Next-Gen 100G Ethernet Optical Standardization

Interface	Link Distance	Media type	Optical Technology
400GBASE-SR16	100 m (OM4)	32f Parallel MMF	16x25G NRZ Parallel VCSEL
400GBASE-DR4	500 m	8f Parallel SMF	4x100G PAM4 Parallel (SiP)
400GBASE-FR8	2 km	2f Duplex SMF	8x50G PAM4 LAN-WDM (DML)
400GBASE-LR8	10 km	2f Duplex SMF	8x50G PAM4 LAN-WDM (DML)

SR16 not expected to be deployed

400GE interfaces standardized in IEEE 802.3bs

Interface	Link Distance	Media type	Optical Technology
100GBASE-SR2	100 m (OM4)	4f Parallel MMF	2x50G PAM4 850nm (VCSEL)
100GBASE-DR	500 m	2f Duplex SMF	100G PAM4 1310nm (EML)

Next-Gen 100GE standardized in IEEE 802.3cd

Interface	Link Distance	Media type	Optical Technology
400GBASE-SR8	100 m (OM4)	16f Parallel MMF	8x50G PAM4 850nm (VCSEL)
400GBASE-SR4.2	100 m (OM4)	8f Parallel MMF	8x50G PAM4 BiDi 850 / 910nm (VCSEL)

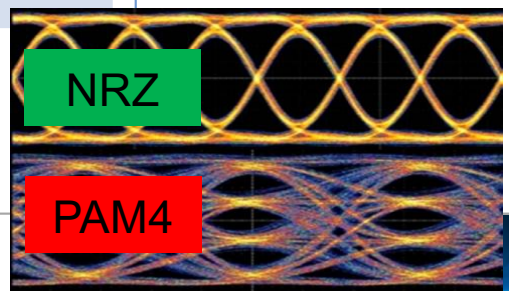
Multimode 400GE objectives in IEEE P802.3cm

Interface	Link Distance	Media type	Optical Technology
400G-FR4	2 km	2f Duplex SMF	4x100G PAM4 CWDM (EML)
100G-FR	2 km	2f Duplex SMF	100G PAM4 1310nm (EML)
100G LR	10 km	2f Duplex SMF	100G PAM4 1310nm (EML)

100G Lambda MULTI-SOURCE AGREEMENT

- VCSEL technology to be used <100m
- Silicon Photonics to be used <1km
- DML/EML technology to be used <40km

SWDM to enable 400GE over Duplex MMF in the future

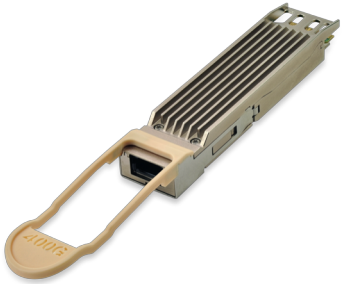


400GE Optical Transceiver Form Factor MSAs

CFP8 is the *1st-generation 400GE* module form factor, to be used in core routers and DWDM transport client interfaces.

Module dimensions are **slightly smaller than CFP2**

Supports either **CDAUI-16** (16x25G NRZ) or **CDAUI-8** (8x50G PAM4) electrical I/O



QSFP-DD and OSFP modules being developed as *2nd-generation 400GE,* for **high port-density data center switches**.

Enable **12.8Tb/s** in 1RU via 32 x 400GE ports

Support **CDAUI-8** (8x50G PAM4) electrical I/O only

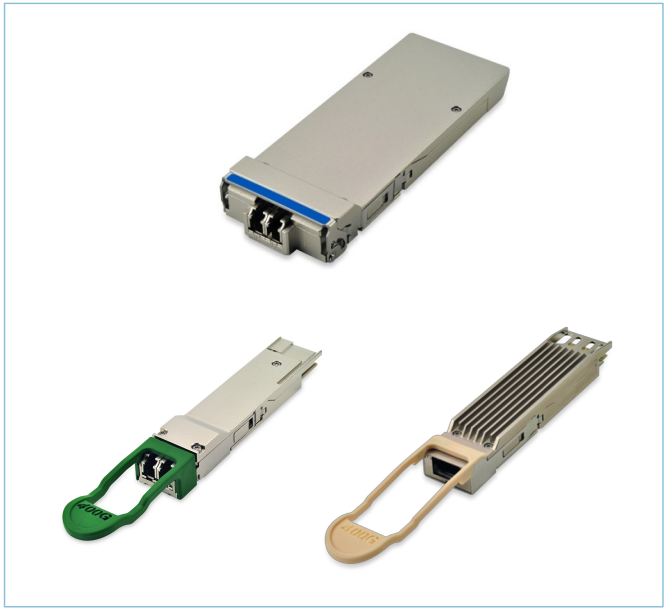
QSFP-DD host is backwards compatible with QSFP28

Potential Types of 400G Transceivers in the Market

	PARALLEL (MPO)	DUPLEX (LC)
MULTIMODE	**SR8 & 8x50G-SR (breakouts)** 70/100/100m **SR4.2** 70/100/150m	**TBD / TBD** ***Future support for existing enterprise duplex fiber infrastructure***
SINGLE MODE	**DR4** 500m	**FR8 / FR4 / CWDM8** 2km **LR8 / LR4** 10km **ER8** 40km **ZR** 80km

BLACK = Standard IEEE interfaces

RED = MSA and Proprietary interfaces



Multimode distances refer to OM3/OM4/OM5 MMF; Single mode distances refer to SMF.

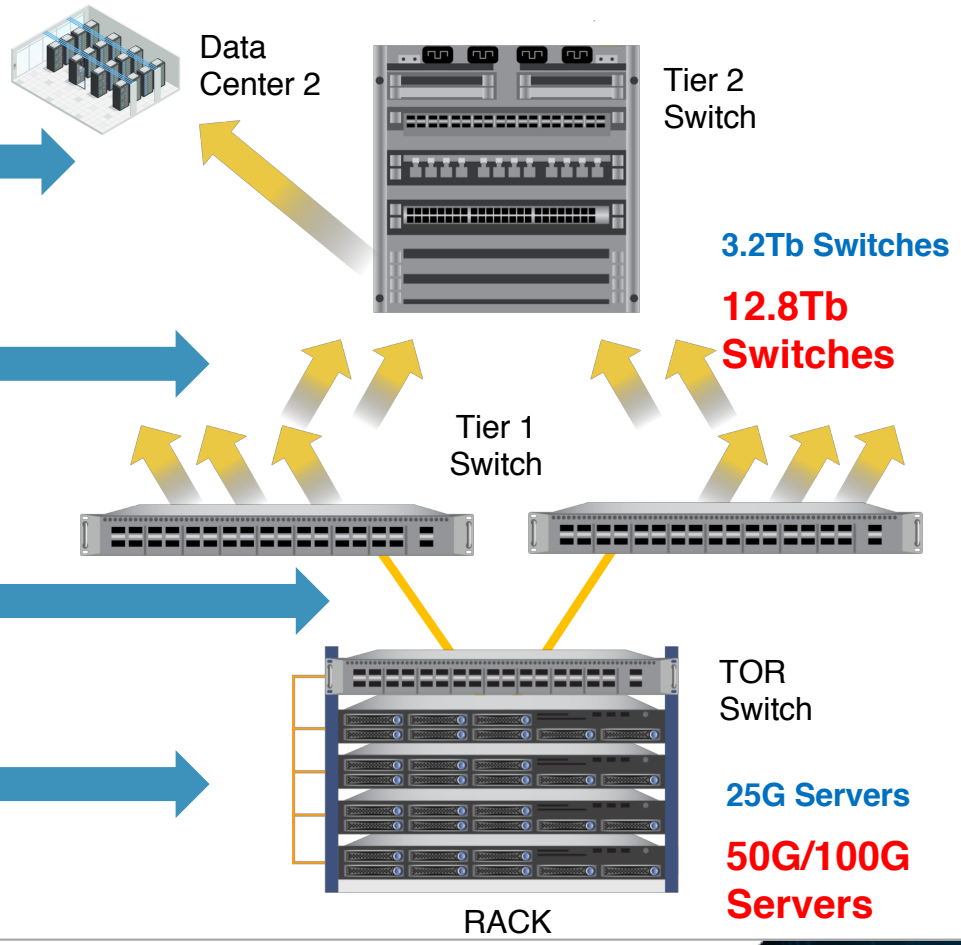
400G Ethernet Is Taking Shape in the Cloud Data Center

Metro DCI (< 80km)
Deploying 100G/200G Coherent
Roadmap is 400GE LR8/ER8, ZR Coherent

Tier 2 Switch to Tier 1 Switch links
Deploying 100GE CWDM4/PSM4
Roadmap is 400GE FR8/FR4/DR4

Tier 1 Switch to TOR Switch links
Deploying 100GE SR4/AOC
Roadmap is 400GE DR4/SR4.2/SR8/AOC

TOR Switch to Server links
Deploying 25GE SR/AOC/DAC(3m)
Roadmap is 50GE/100GE SR/AOC

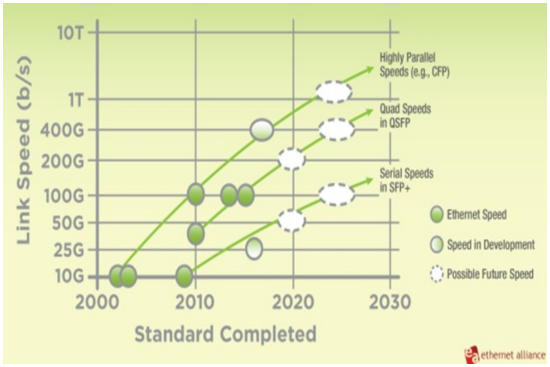


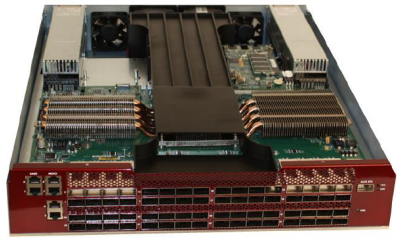
General Trends in Data Center Optical Interconnects

- Continuous increase in bandwidth density
  - On-board optics vs. pluggable optics discussion
- Increasing adoption of optics in Server-to-TOR Switch links
- Low-latency optics for certain niche cognitive-computing applications
- Maturity of key technologies
  - High-speed VCSELs
  - Silicon photonics
- Arrival of coherent optics for data center interconnects

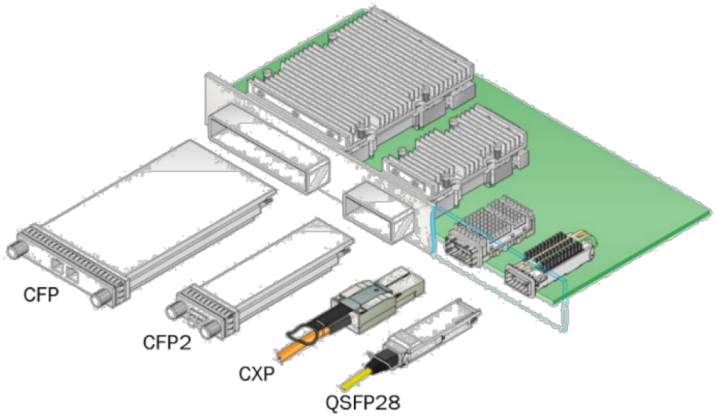
The Market Demands Continuous Improvement in Bandwidth Density

Module Type	# of I/O lanes	Electrical I/O	I/O Baud Rate	Module BW	Width (mm)
SFP+	1	10Gb/s-NRZ	10G	**10Gb/s**	13
QSFP+	4	10Gb/s-NRZ	10G	**40Gb/s**	18
QSFP28	4	25Gb/s-NRZ	25G	**100Gb/s**	18
QSFP56	4	50Gb/s-PAM4	25G	**200Gb/s**	18
QSFP-DD / OSFP	8	50Gb/s-PAM4	25G	**400Gb/s**	18
Form factor?	8	100Gb/s-PAM4	50G	**800Gb/s**	?

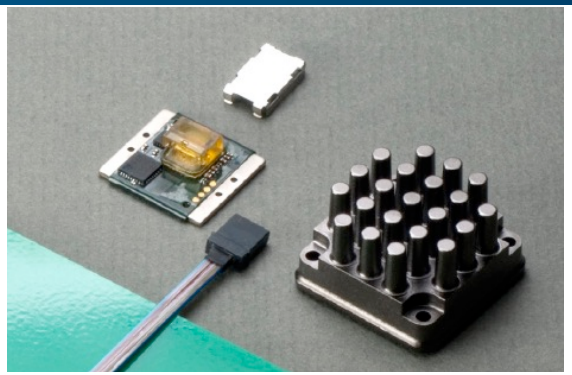


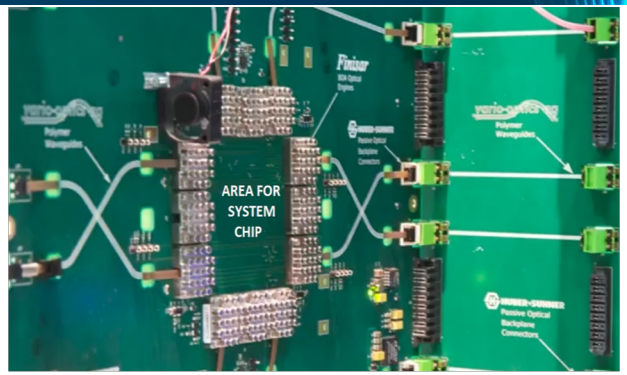


256 x 25G Switch System – 2 RU
(64 x QSFP28 interfaces)



Is Pluggability Still a Requirement for Optics?





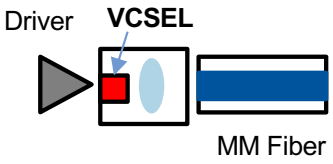
- Some optics are not pluggable; they are mounted directly on the system host PCB.
  - BOAs have been used for several years on core routers (inter-chassis) and supercomputers.
  - Very short host PCB traces enable low power dissipation and high port density.
- Higher bandwidth density can be achieved by:
  - More channels, e.g., up to 16 Tx and 16 Rx channels in a module.
  - Higher data rate per channel: 10G/ch and 25G/ch variants deployed today, 50G/ch in the future.
- The industry view however is that **pluggable optics** will be preferred for 400G Ethernet.
  - Facilitates maintenance and pay-as-you-grow model.

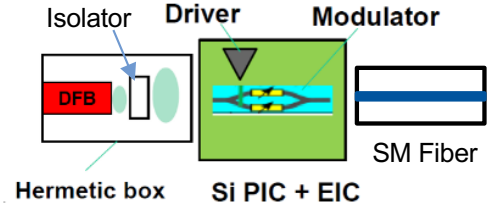
Optical Technologies for Next-Generation Data Centers

- Short Reach (0 to 100 meters)
  - Higher bandwidth VCSELs
  - VCSELs with sparing capability
  - VCSELs with low RIN
- Intermediate Reach (500 meters to 2 km)
  - Silicon photonics
- Long Reach (10 km and beyond)
  - DML/EMLs
  - Low-power coherent optics

Multimode Optics: Best Suited for Short Reach Links

For short reach links (<100 m), multimode optics have better cost, power, density than silicon photonics





VCSEL Tx components:

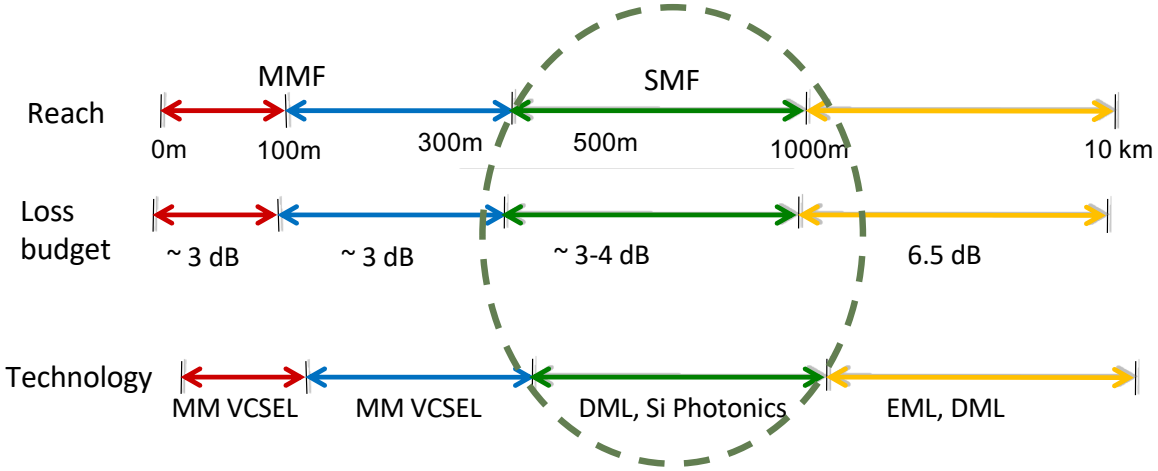
- High speed 850nm VCSEL, 150μm x 150μm die, ~8 μm diameter
- Coupling optics with 10 μm's tolerance
- VCSEL driver, <10 mA
- Non-hermetic packaging
- MM optical alignments: 1

Si Photonics Tx components:

- High power CW 1310nm DFB, >500 µm long
- Optical isolator
- External modulator
- Coupling optics with ~1 µm tolerance
- Si Photonic PIC + IC driver, several tens of mA
- Hermetic box or enclosure
- SM optical alignments: 2

Next-Gen Intermediate Reach Optical Interconnects: 500 m to 2 km

Well-suited for DML or silicon photonics (SiP) implementation: DR4, FR4



Today's application space is divided by fiber type and reach:

- 0-100m MM → MM VCSEL today (OM3/OM4 fiber types)
- 0-300m MM → MM VCSEL (OM4/OM5 fiber type for 4x25G)

850 nm

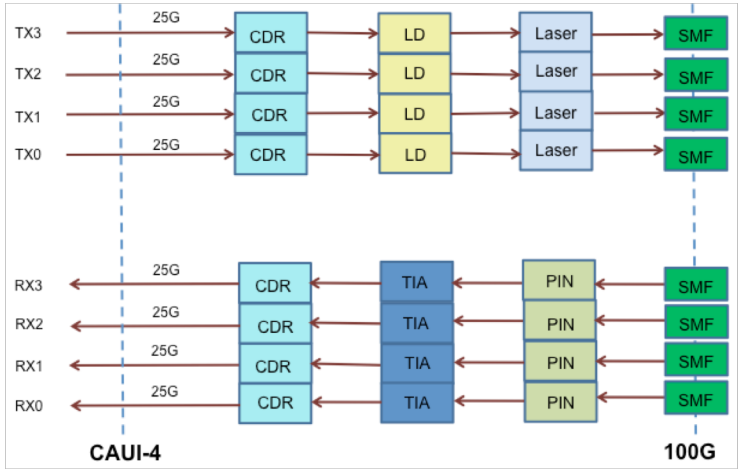
- 0-2 km SM → DML, Si Photonics *New application space since 100G*
- 0-10 km SM →EML, DML

1310nm

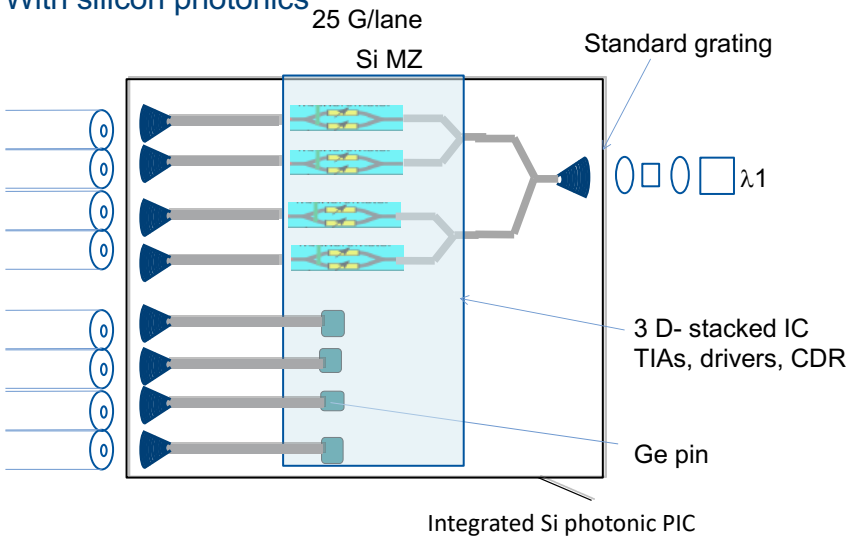
Implementation of 100G-PSM4 With Silicon Photonics

- Over 84% of the functions are integrated with high yield at wafer level in a Si fab
  - SiP integrates 4 modulators, 4 pin diodes, 4 x drivers, 4 x TIAs, CDR in a 3-D stacked self-hermetic, uncooled chip.
  - A similar architecture is being implemented for 400G-DR4 (“400G PSM4”), driven by the need for longer reach, future-proof singlemode cabling, and integration in optics.

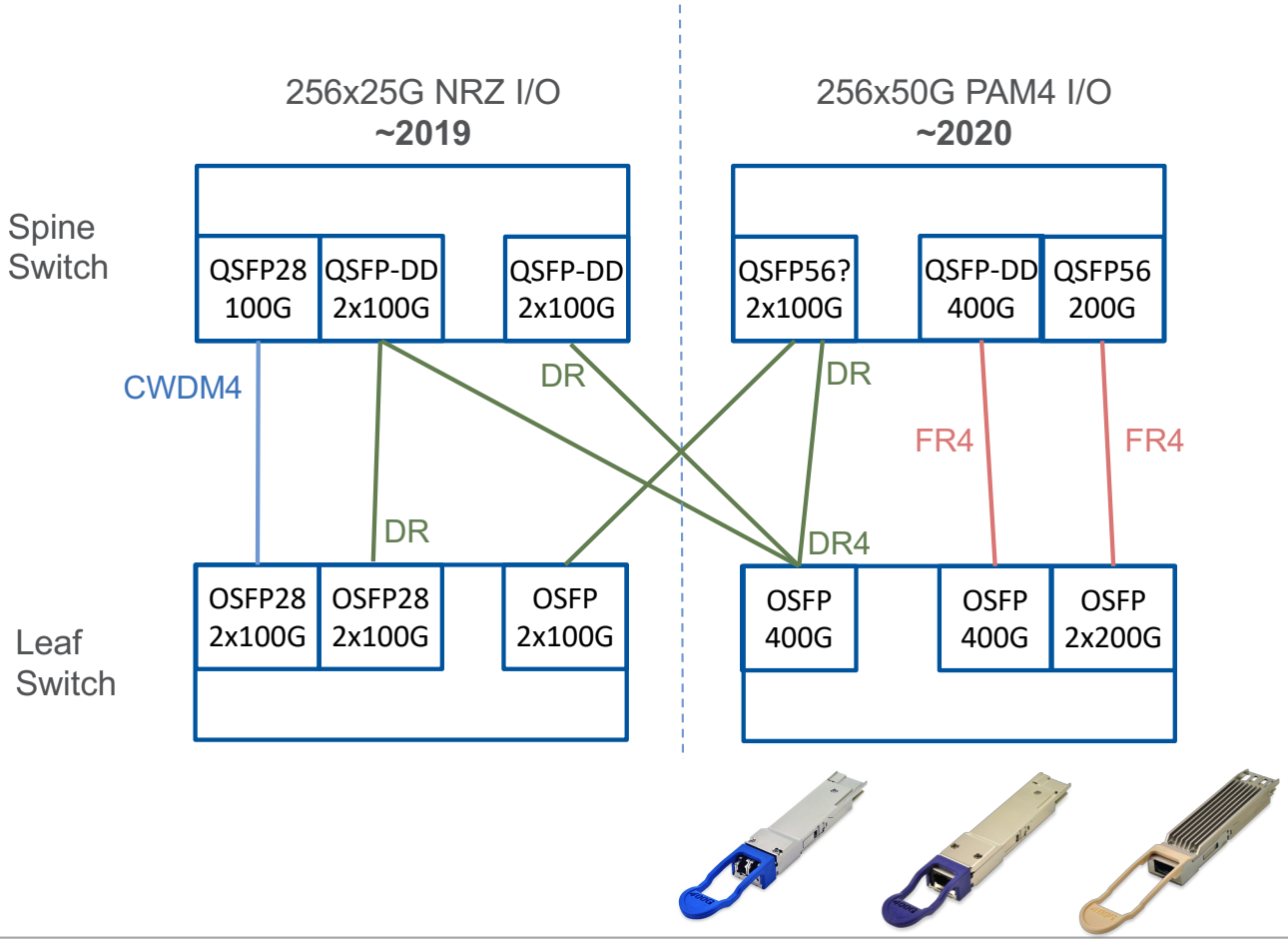
Traditional



With silicon photonics



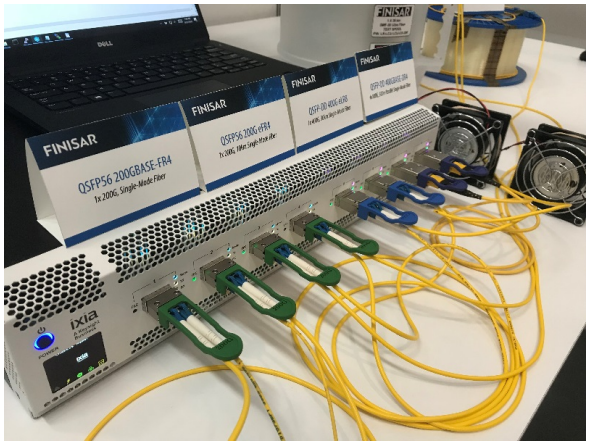
Several New Interface Types and Form Factors to be Deployed



- Enabled by high I/O count and by 400G-DR4 to 100G-DR breakout interoperability, high-density 100G implementations will thrive in Leaf-Spine topologies.
- Large I/O Line Cards will have QSFP-DD or OSFP sockets. OSFP slots may use QSFP adapters.

400G, 200G & 100G PAM4 Transceiver Demos at OFC/ECOC 2018

400G QSFP-DD LR8/FR8 (10km)



400G QSFP-DD AOC (70m)

100G QSFP28 DR/FR (2km)

200G SFP56 FR4 (2km)
200G QSFP56 eFR4 (10km)
400G QSFP-DD eLR8 (30km)
400G QSFP-DD DR4 (500m)

Additionally, several interoperability demos were done by the MSAs

80km DCI Space: Coherent vs. Direct Detection

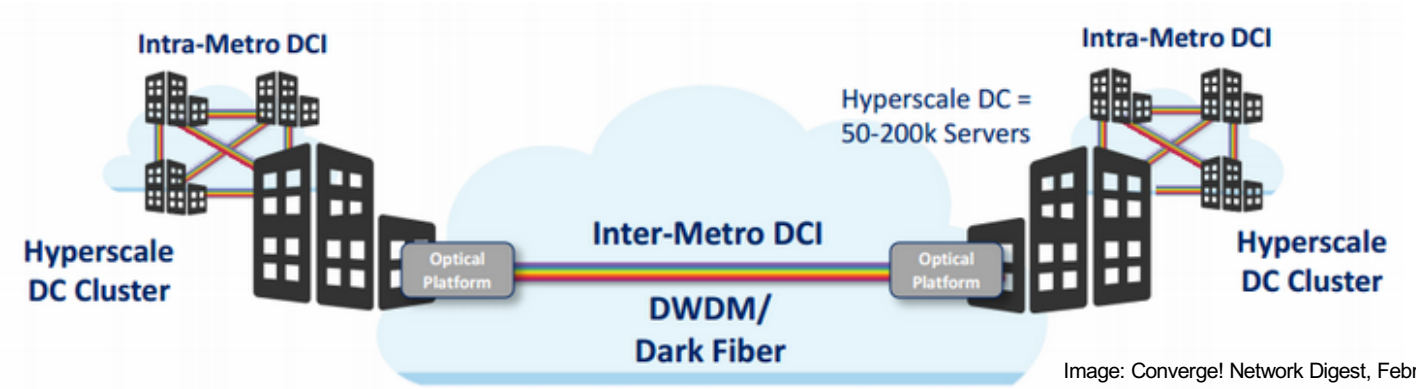


Image: Converge! Network Digest, February 2015

- Coherent interfaces are likely to capture the 80km market at 400Gb/s and higher rates.
- For 40km and shorter reaches, direct detection may be lower power and cost than coherent for the next few years. Example: 8x50Gb/s (PAM4) ER8 and eLR8 modules.
- Currently coherent technology is about 2x higher power and cost relative to 100Gb/lane direct detection.
- Standardization work by OIF 400ZR IA and IEEE P802.3cn/ct Task Forces (400G ER8 and ZR).
- Aggressive innovation will be required to maintain long-term trends to support 1.6 Tb/s ~2024.

Coherent Transmission for DCI Applications

- 100G/200G links require a transponder box to convert to coherent optical transmission in order to support 80~100km and beyond.
- Several system OEMs provide a 1RU transponder box for DCI applications, most of which use pluggable Coherent CFP2-ACO optical transceivers.





CFP2-ACO

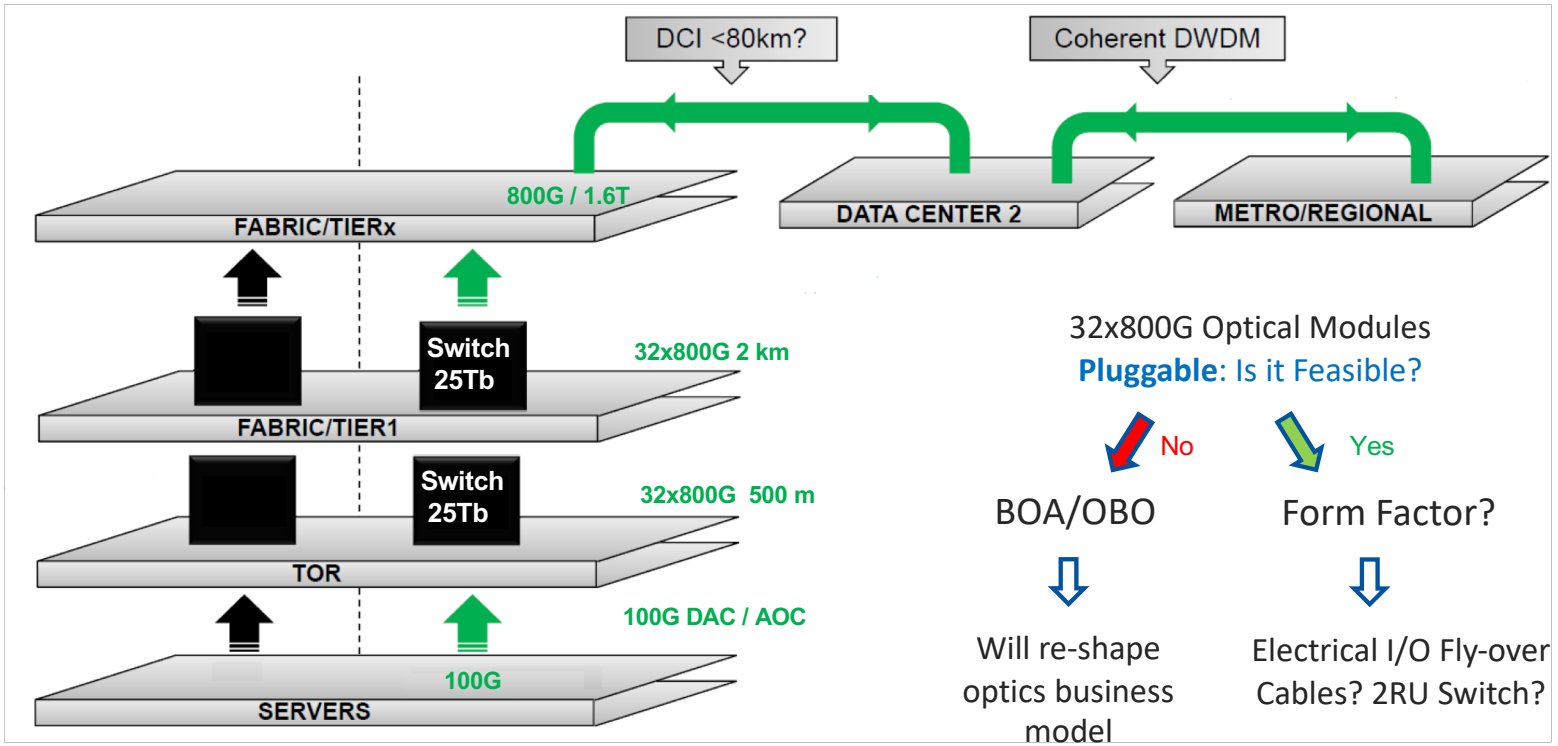
ACO = Analog Coherent Optics
DCO = Digital Coherent Optics

- Expected coherent transceiver evolution is driven by improvements in optical packaging and DSP power dissipation:

200G CFP2-ACO → **400G CFP2-DCO → 400G QSFP-DD DCO**

400G DCO transceivers are expected to be plugged directly into switches and routers

Coming Next: What Shape Will 800G Ethernet Take?



100G PAM4 electrical I/O being standardized in IEEE P802.3ck

Thank You

Christian Urricariet

christian.urricariet@finisar.com