FINISAR®

Latest Trends in Data Center Optics

AusNOG 2016

Sydney, September 2016

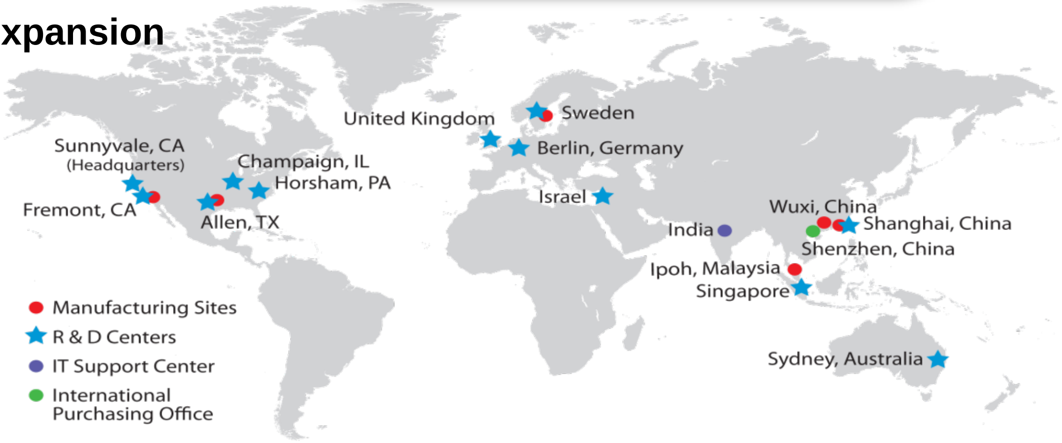
Christian Urricariet

Finisar Corporation

World's Largest Supplier of Fiber Optic Components and Subsystems

- **Optics industry leader with $1B+ in annual revenue**
- **Founded in 1988**
- **IPO in 1999 (NASDAQ: FNSR)**
- **14,000 employees**
- **Best-in-class broad product line**
- **Vertically integrated with low cost manufacturing**
- **Significant focus on R&D and capacity expansion**
- **Experienced management team**
- **1300+ Issued U.S. patents**

Corporate Headquarters: Sunnyvale, CA USA



Broad Product Portfolio and Customer Base

DATACOM

PRODUCTS

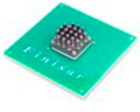
SFP

SFP+

QSFP/QSFP28

CFP2/ CFP4

CFP


Optical Engine (BOA)

CXP

Active Optical Cables

XFP

X2/XENPAK

CUSTOMERS

EMC² where information lives®
intel
extreme networks
cisco
BROCADE
JUNIPER NETWORKS
DELL
NetApp
IBM
EMULEX We network storage
H3C ITolP Solutions Expert
hp invent
QLOGIC
ORACLE
Mellanox TECHNOLOGIES

TELECOM

PRODUCTS

SFP

XFP

SFP+

CFP2-ACO

Coherent Transponder

ROADM line card

WSS

WDM Passives

Amplifiers

High speed components

Tunable laser

CATV

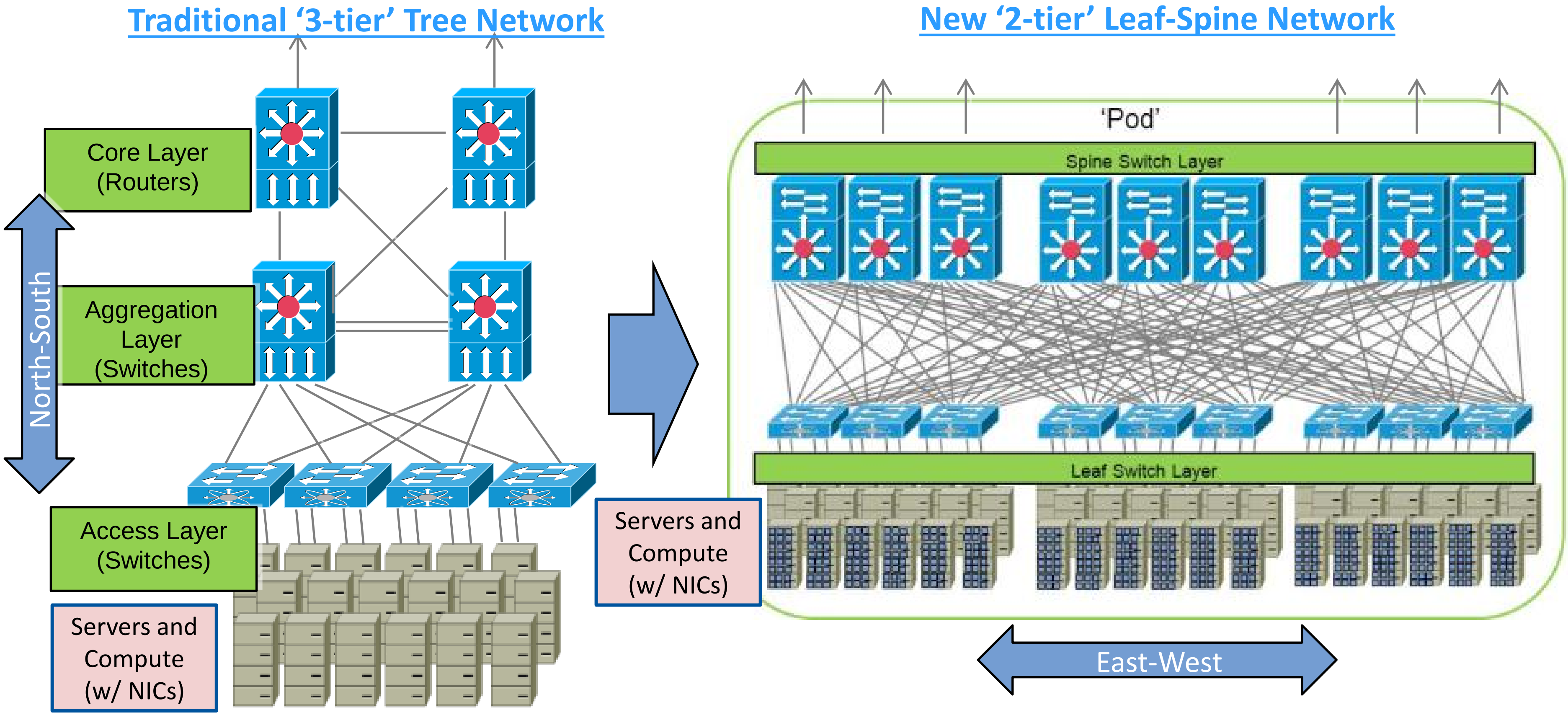
PON

CUSTOMERS

HUAWEI
ZTE中兴
Alcatel·Lucent
ERICSSON
HITACHI Inspire the Next
NOKIA
CIENA
NEC
Coriant
ADVA Optical Networking
transmode
eci
FUJITSU
cyan
infinera

New Architectures in Hyperscale Data Centers

- Most data center networks have been architected on a 3-tier topology
- Cloud data center networks are migrating from traditional 3-tier to flattened 2-tier topology
  - Hyperscale Data Centers becoming larger, more modular, more homogenous
  - Workloads spread across 10s, 100s, sometimes 1000s of VMs and hosts
  - Higher degree of east-west traffic across network (server to server)


Traditional '3-tier' Tree Network
New '2-tier' Leaf-Spine Network
Core Layer (Routers)
Aggregation Layer (Switches)
Access Layer (Switches)
Servers and Compute (w/ NICs)
North-South
Servers and Compute (w/ NICs)
'Pod'
Spine Switch Layer
Leaf Switch Layer
East-West

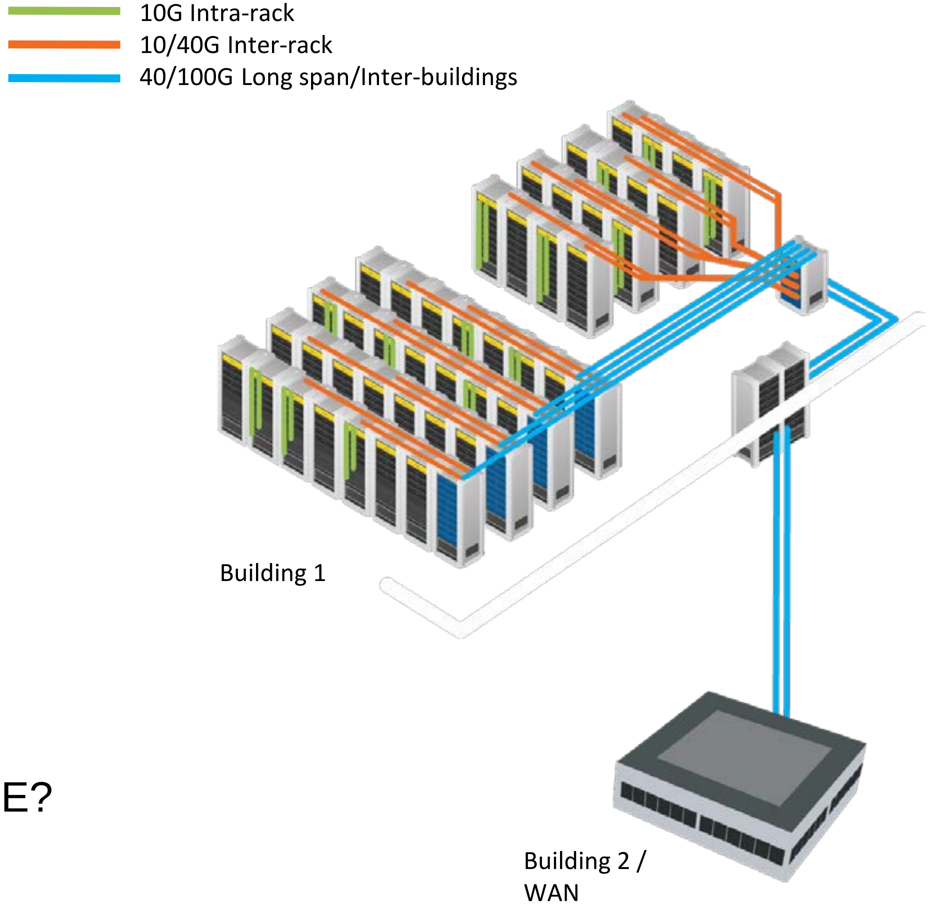
Data Center Connections are Evolving

- Due to the significant increase in bandwidth demand, Data Center connections are moving from 1G/10G, to 25G/40G/100G

- Within the Data Center Rack
  - **10GE** being deployed now
  - 25GE to be deployed soon
  - 50GE to the server will likely follow

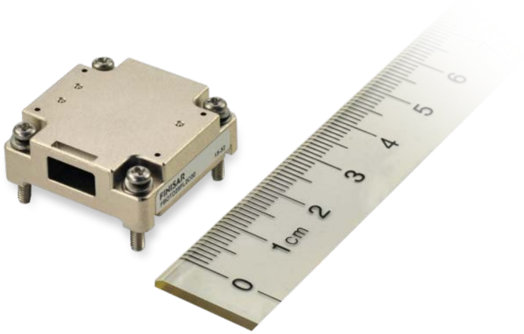
- Between Data Center Racks
  - **40GE** being deployed now
  - 100GE to be deployed soon
  - What follows? 200GE or 400GE?

- Long Spans/DCI & WAN
  - **100GE** being deployed now
  - 400GE being standardized now
  - What follows? 800GE, 1TE or 1.6TE?

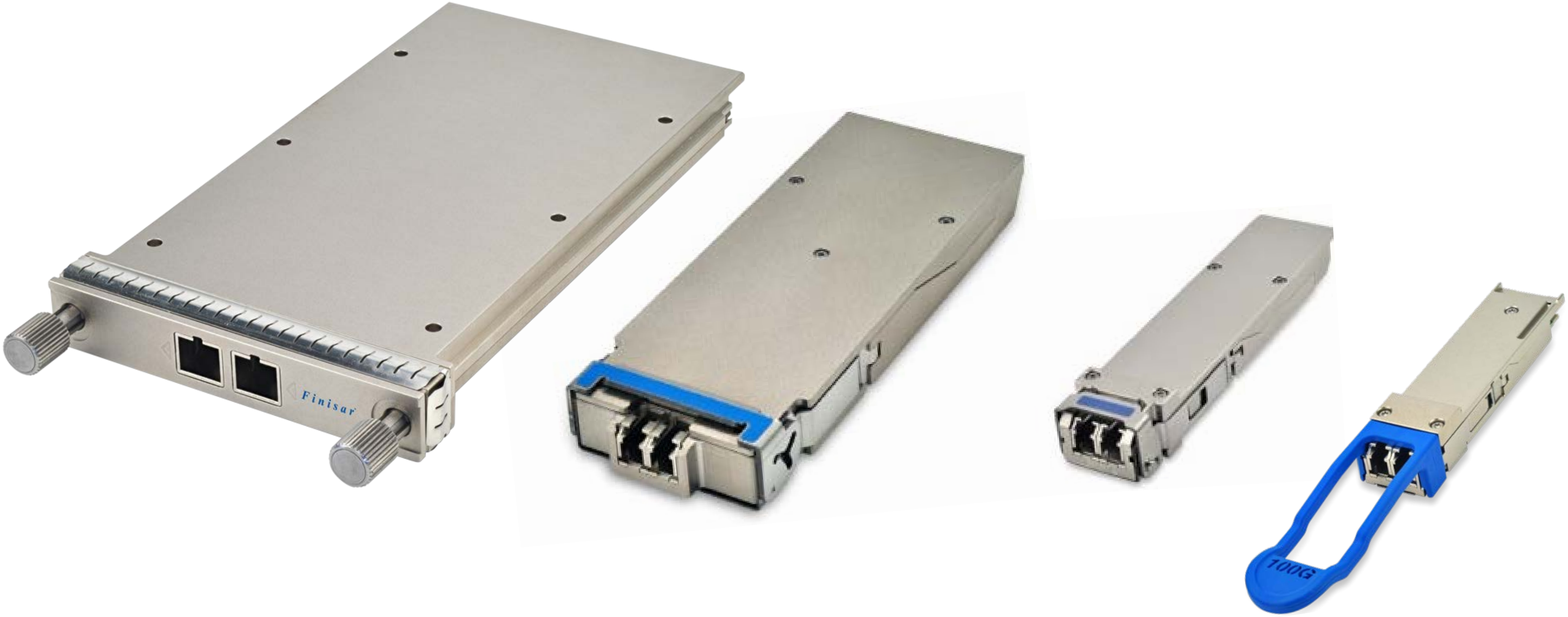


Optical Trends in the Data Center Market

- Significant increase in 100G and 25G port density
  - Smaller form factors, e.g., QSFP28 modules
  - 100G power dissipation <3.5W
  - Cost-effective Active Optical Cables
  - On-board optics for very high port density



100G Optical Module Form Factor Evolution



CFP
4 ports/row
16-24W

CFP2
8-10 ports/row
8W

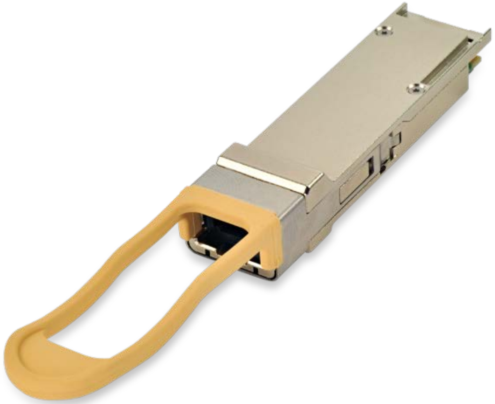
CFP4
16-18 ports/row
5W

QSFP28
18-20 ports/row
3.5W

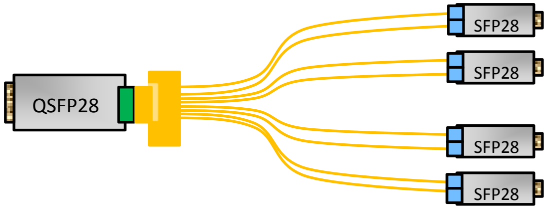
Deployments today

time

100G QSFP28 Optical Module



4x25G Breakout



- **100GE optical transceivers**
  - QSFP28 is standardized by SFF-8665 (SFF Committee)
  - It has a 4-lane, retimed 25G I/O electrical interface (CAUI-4)
  - Supports up to 3.5W power dissipation with standard cooling
  - Also used for 4x 25GE applications
- **100GE active optical cables (no optical connector)**

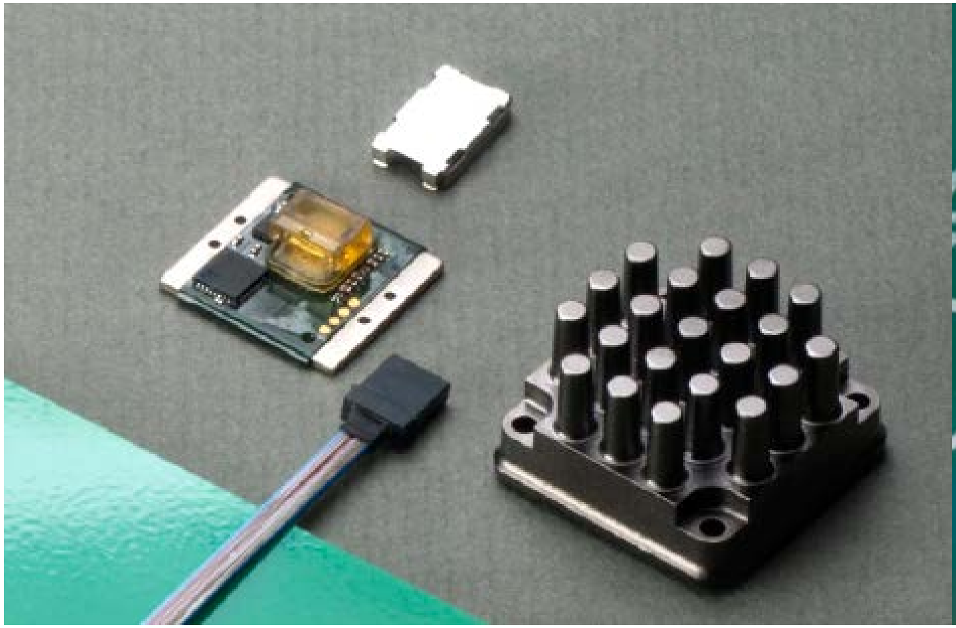
QSFP28 is the 100GE module form factor of choice for new data center switches

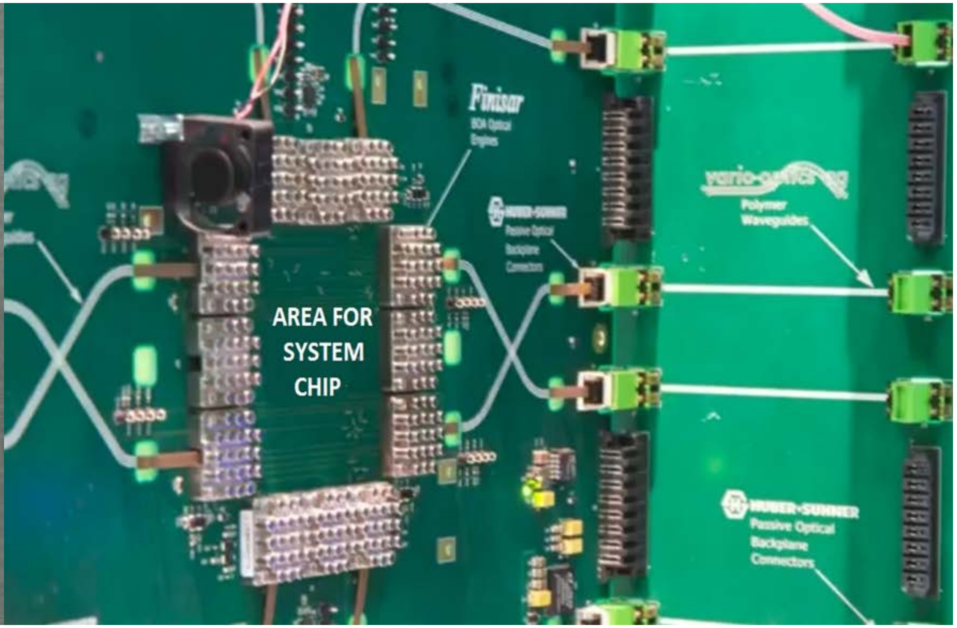
25G SFP28 Optical Module

- **25GE optical transceivers**
  - SFP28 is standardized by the SFF Committee
  - It has a 1-lane, retimed 25G I/O electrical interface
  - Supports up to 1W power dissipation with standard cooling
  - Used for 25GE ports in server and switches
- **25GE active optical cables**

SFP28 is the 25GE module form factor of choice for new Servers / NICs

Board-Mounted Optical Assembly (BOA)





- These optics are not pluggable; they are mounted on the host PCB
- Used today on supercomputers and some routers and switches
- Very short host PCB traces enable low power dissipation
- Higher bandwidth density can be achieved by:
  - More channels: Up to 12+12 Tx/Rx, or 24Tx and 24Rx
  - Higher data rate per channel: 10G/ch and 25G/ch variants today, 50G/ch in the future

Optical Trends in the Data Center Market

- Significant increase in 100G and 25G port density
- Extension of optical links beyond the Standards

40G Ethernet QSFP+ Modules

	Parallel (MPO)	Duplex (LC)
Multimode	SR4 • 100/150m eSR4 & 4xSR • 300/400m	A duplex multimode product is required to re-use the same fiber plant used for 10GE
Single Mode	4xLR • 10km 4xLR Lite • 2km	LR4 • 10km ER4 • 40km

Parallel links ***can*** be broken out to 4 separate 10G connections

Duplex WDM ***cannot*** be broken out to 4 separate 10G connections

Black = Standardized interfaces

Blue = MSA/Proprietary interfaces

Multimode distances refer to OM3/OM4
Single mode distances refer to SMF28

100G Ethernet QSFP28 Modules

	Parallel (MPO)	Duplex (LC)
Multimode	SR4 & 4x25G-SR • 70/100m SR4 without FEC • 30/40m	A duplex multimode product is required to re-use the same fiber plant used for 10GE
Single Mode	PSM4 • 500m	LR4 • 10km CWDM4/CLR4 • 2km

Parallel links ***can*** be broken out to 4 separate 25G connections

Duplex WDM ***cannot*** be broken out to 4 separate 25G connections

Black = Standardized interfaces

Blue = MSA/Proprietary interfaces

Multimode distances refer to OM3/OM4
Single mode distances refer to SMF28

Impact of Latency on 25G/100G Ethernet Optical Links

- Various recent 25G and 100G Ethernet standards and MSAs require the use of **RS-FEC** (aka, “KR4 FEC”) on the host to increase overall link length.

- RS-FEC does not increase the total bit rate, but it introduces an additional ***latency of ~100ns*** in the link.
  - Some applications like HFT have little tolerance for latency.

Standard	Link Length with RS-FEC
IEEE 802.3bm 100GBASE-SR4	100m on OM4 MMF
IEEE P802.3by 25GBASE-SR	100m on OM4 MMF
100G CWDM4 MSA	2km on SMF
100G PSM4 MSA	500m on SMF

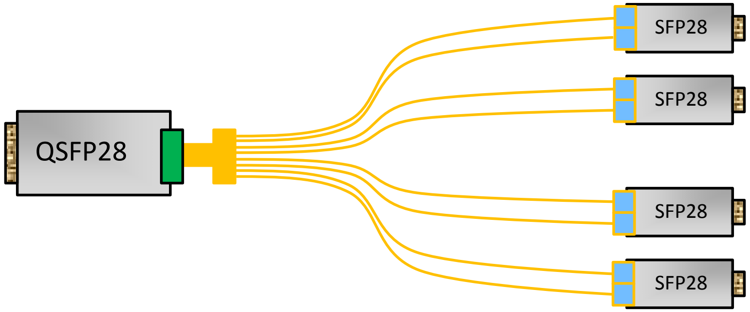
- The fiber propagation time of each bit over 100m of MMF is **~500ns**
  → The amount of additional latency introduced by RS-FEC may be significant for the overall performance of short links <100 meters (see next page).

- But the fiber propagation time of each bit over 500m of SMF is **~2500ns**
  → The amount of latency introduced by RS-FEC is **not** significant for the overall performance of links >500 meters.

Low-Latency QSFP28 SR4 and SFP28 SR without FEC

- Support of error-free 25G/100G Ethernet links ***without FEC***
  - Lower latency
  - Lower host power dissipation

- Standard QSFP28 and SFP28 form factors
- Supports 4:1 fan-out configuration
- Up to 30 meters on OM3 / 40 meters on OM4 MMF



Optical Trends in the Data Center Market

- Significant increase in 100G and 25G port density
- Extension of optical links beyond the Standards
- Reutilization of existing 10G fiber plant on 40G and 100G

Why Duplex Multimode Fiber Matters

- **For Brownfield Applications:**
  - Data centers today are architected around 10G Ethernet
  - Primarily focused on 10GBASE-SR using **duplex MMF (LC)**

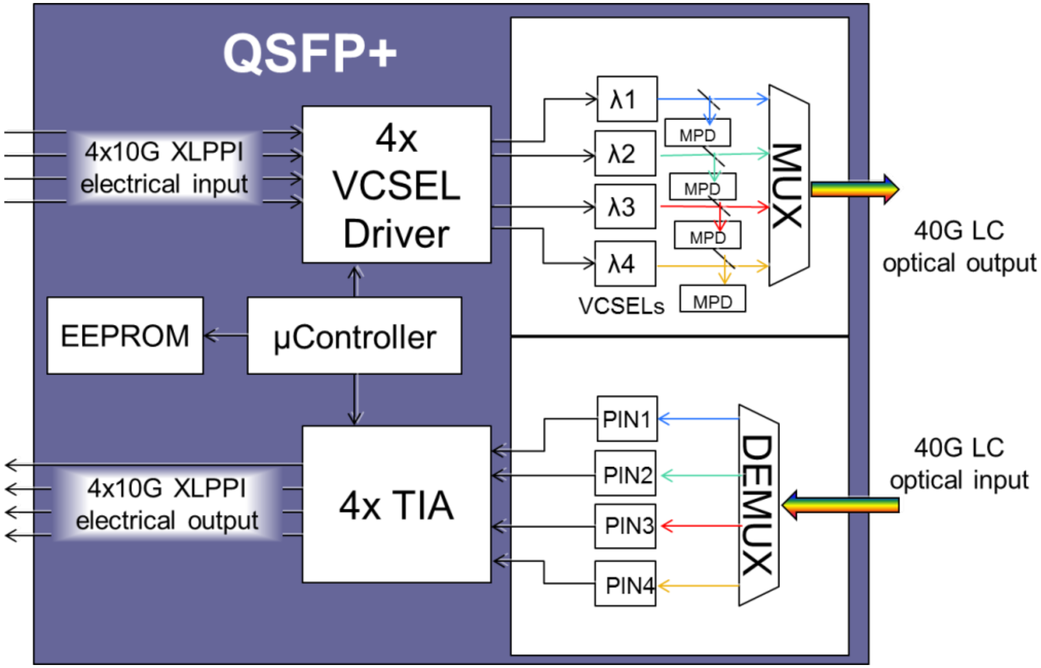
- Data center operators are migrating from 10G to 40G or 100G, but want to maintain their existing fiber infrastructure.
  - SR4 requires ribbon multimode fiber with an MPO connector.
    - ***Not provided by pre-installed fiber plant.***
  - LR4 requires single mode fiber.
    - ***Not provided by pre-installed fiber plant.***

Data centers want to upgrade from 10G to 40G and 100G
without touching the duplex MMF fiber infrastructure

Introducing Shortwave WDM (SWDM)

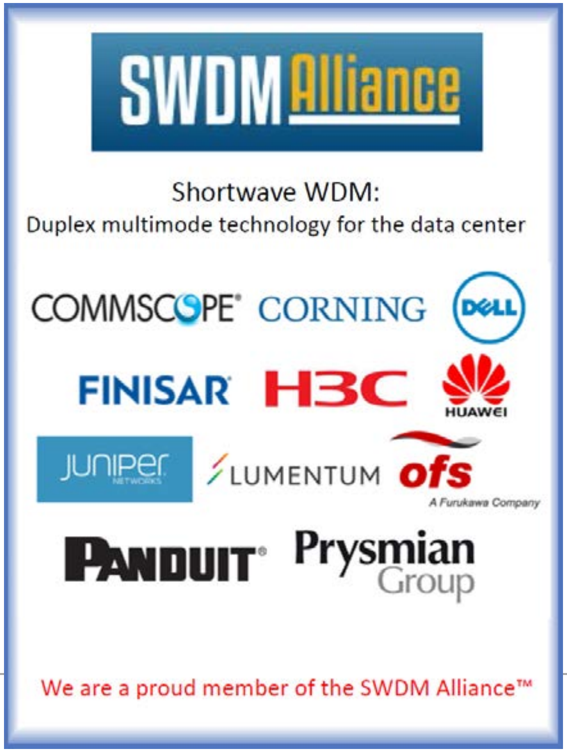
- SWDM uses 4 different wavelengths in the 850nm region, where MMF is optimized, which are optically multiplexed inside the transceiver.
- SWDM enables the transmission of 40G (4x10G) and 100G (4x25G) over existing duplex multimode fiber, using LC connectors.

Block diagram of a 40G SWDM QSFP+ Transceiver



SWDM Alliance

- Industry group to promote SWDM technology for duplex MMF in data centers.
- Finisar is a founding member of the SWDM Alliance.
- More information at www.swdm.org



SWDM Alliance

Shortwave WDM:
Duplex multimode technology for the data center

COMMSCOPE® CORNING DELL

FINISAR® H3C HUAWEI

JUNIPER NETWORKS LUMENTUM ofs A Furukawa Company

PANDUIT® Prysmian Group

We are a proud member of the SWDM Alliance™

swdm4™

40G Ethernet QSFP+ Modules

	Parallel (MPO)	Duplex (LC)
Multimode	SR4 • 100/150m eSR4 & 4xSR • 300/400m	Bi-directional • Limited use SWDM4 • Being tested LM4 • 140/160m/1km
Single Mode	4xLR • 10km 4xLR Lite • 2km	LR4 • 10km ER4 • 40km

Parallel links ***can*** be broken out to 4 separate 10G connections

Duplex WDM ***cannot*** be broken out to 4 separate 10G connections

Black = Standardized interfaces

Blue = MSA/Proprietary interfaces

Multimode distances refer to OM3/OM4
Single mode distances refer to SMF28

100G Ethernet QSFP28 Modules

	Parallel (MPO)	Duplex (LC)
Multimode	SR4 & 4x25G-SR • 70/100m SR4 without FEC • 30/40m	SWDM4 • Being tested
Single Mode	PSM4 • 500m	LR4 • 10km CWDM4/CLR4 • 2km
	Parallel links ***can*** be broken out to 4 separate 25G connections	Duplex WDM ***cannot*** be broken out to 4 separate 25G connections

Black = Standardized interfaces

Blue = MSA/Proprietary interfaces

Multimode distances refer to OM3/OM4
Single mode distances refer to SMF28

Optical Trends in the Data Center Market

- Significant increase in 100G and 25G port density
- Extension of optical links beyond the Standards
- Reutilization of existing 10G fiber plant on 40G and 100G
- Moving beyond 100G, to 200G and 400G
  - Service Provider applications
  - Data Center applications

400GE Standardization

- The 400GE Standard is already being defined by IEEE P802.3bs.

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

Electrical I/O:	CDAUI-8	8x50G PAM4
	CDAUI-16	16x25G NRZ

  - 400GE Standard is expected to be ratified in December 2017

- Optics suppliers are already working on components to support these new rates.
  - Based on VCSELs, InP DFB laser and Si Photonics technologies
  - ICs and test platforms that support PAM4 encoding

50G, 200G and Next-Gen 100G Ethernet Standardization

- 200GE PMD objectives being standardized by IEEE 802.3bs:

Interface	Link Distance	Media type	Technology
200GBASE-SR4	100 m	8f Parallel MMF	4x50G PAM4 850nm
200GBASE-DR4	500 m	8f Parallel SMF	4x50G PAM4 1300nm window
200GBASE-FR4	2 km	(2f) Duplex SMF	4x50G PAM4 CWDM
200GBASE-LR4	10 km	(2f) Duplex SMF	4x50G PAM4 LAN-WDM

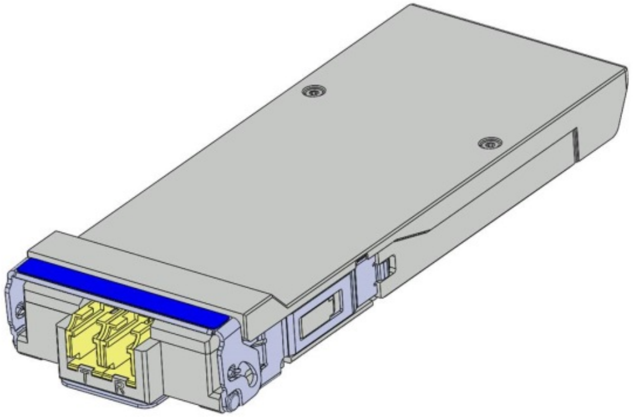
- 50GE PMD objectives being standardized by IEEE 802.3cd:

Interface	Link Distance	Media type	Technology
50GBASE-SR	100 m	(2f) Duplex MMF	50G PAM4 850nm
50GBASE-FR	2 km	(2f) Duplex SMF	50G PAM4 1300nm window
50GBASE-LR	10 km	(2f) Duplex SMF	50G PAM4 1300nm window

- Next-Gen 100GE PMD objectives being standardized by IEEE 802.3cd:

Interface	Link Distance	Media type	Technology
100GBASE-SR2	100 m	MMF	2x50G PAM4
100GBASE-FRx	2 km	(2f) Duplex SMF	TBD
100GBASE-LRx	10 km	(2f) Duplex SMF	TBD

400GE CFP8 Optical Transceiver Module

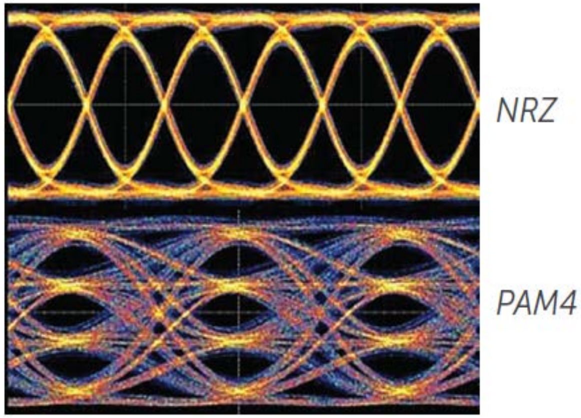


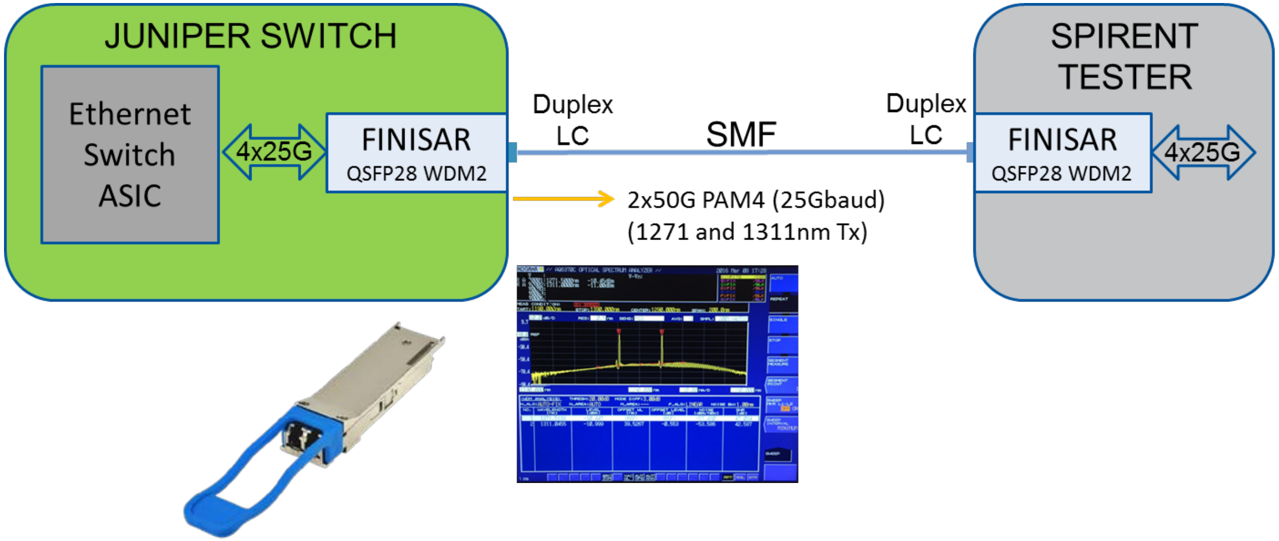
- **CFP8** is the *first-generation* 400GE form factor.
- Module dimensions are **slightly smaller than CFP2**.
- Supports standard IEEE 400G **multimode and single mode** interfaces.
- Supports either **CDAUI-16** (16x25G) or **CDAUI-8** (8x50G) electrical I/O.
- It is being standardized by the **CFP MSA.**

OFC 2016: 2x50G PAM4 100G Interoperability Demo

- Error-free **100G link** connecting Juniper Switch with Spirent Tester
- Using Finisar **QSFP28** prototype modules with **2x50G PAM4 technology**
- Demonstrates building blocks for future Nx50G PAM4 modules: 1x50G, 100G (2x50G), 200G (4x50G) and 400G (8x50G)

- DML technology transmitting **CWDM** wavelengths to enable duplex SMF
- **1271nm and 1311nm** for optimal performance
- Baseline configuration for **100G ‘WDM2’ (FR2/LR2)**

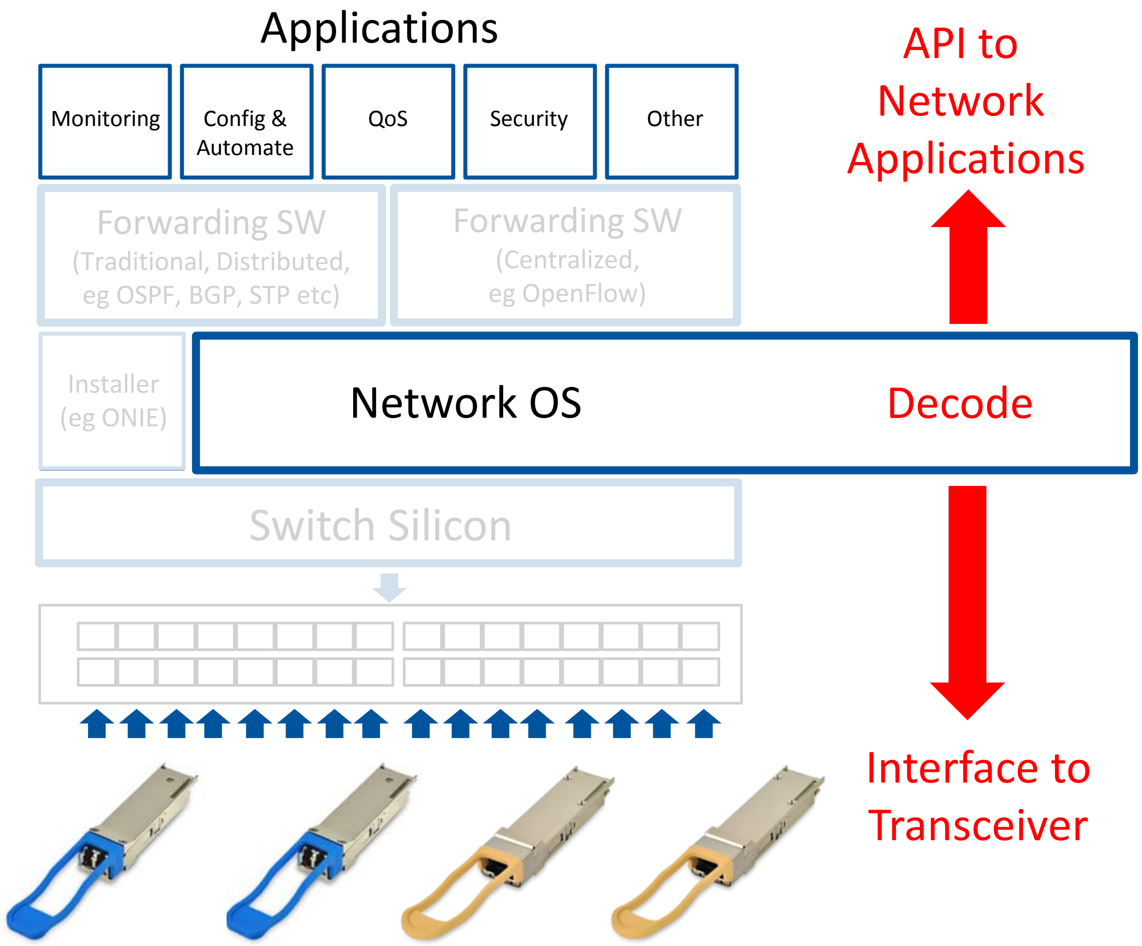




Optical Trends in the Data Center Market

- Significant increase in 100G and 25G port density
- Extension of optical links beyond the Standards
- Reutilization of existing 10G fiber plant on 40G and 100G
- Moving beyond 100G, to 200G and 400G
  - Service Provider applications
  - Data Center applications
- Open Optical Monitoring

Open Optical Monitoring and Control


Applications
Monitoring
Config & Automate
QoS
Security
Other
Forwarding SW (Traditional, Distributed, eg OSPF, BGP, STP etc)
Forwarding SW (Centralized, eg OpenFlow)
Installer (eg ONIE)
Network OS
Decode
Switch Silicon
API to Network Applications
Interface to Transceiver

Finisar is working on offering open APIs to enable broader use of digital diagnostics:

- Transceiver information
- Tx/Rx power
- Module temperature

As well as enable new features:

- Eye and BER monitoring
- Connectivity diagnostics
- And more

FINISAR

Accton
Making Partnership Work

BROADCOM.

cumulus networks

big switch networks

Open Optical Monitoring is now an OCP Project

Optical Layer Monitoring in Open Source

FINISAR sponsoring **TWO** initiatives to promote better access to optical layer diagnostic information in network SW stacks:

Open Optical Monitoring:

- Open Compute (OCP) Networking Project
- Provides access to monitors and controls inside optical modules and active cables
- Intuitive Python API for applications and agents
- Runs on any Linux-based NOS
- Access v0.5 spec and beta code at:

http://www.opencompute.org/wiki/Networking/SpecsAndDesigns

https://github.com/orgs/ocpnetworking-wip/oom

sFlow:

- sFlow.org project
- Extends sFlow to report optical module management information from SFP/QSFP optical modules
- A host sFlow agent (sflow.net) has been running without issue for over a month on three production Cumulus Linux switches in the SFMIX network
- Draft implementation:

http://sflow.org/draft_sflow_optics.txt

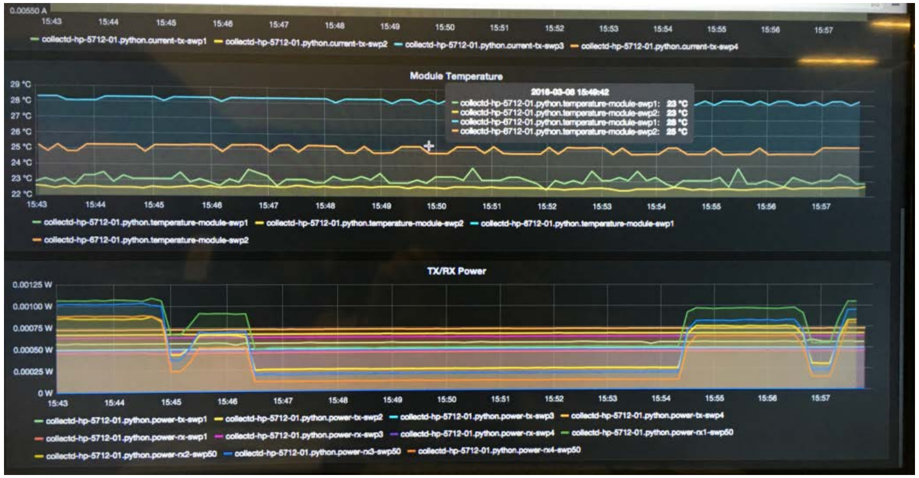
- Source code using the Linux ethtool API is available on github:

https://github.com/sflow/host-sflow/blob/master/src/Linux/readNioCounters.c#L291-L613

Intuitive APIs to Access Pluggable Modules

- Create an **inventory** of all ports SFP+ and QSFP+…
- Extract **Serial ID** information from each module…
- Access **Digital Diagnostic Monitoring** information from each module
- Access **new and value-added** functionality made available by module vendors… Example: Finisar Connectivity Diagnostics

- Connectivity Mapping
- Module Health Indication
- Link Troubleshooting
- Link Performance Indication



Example: Optical health metrics – in 4 lines of Python, 'out of the box'

```
from oom import *
for port in oom_get_portlist():
             # enumerate the ports on the switch
   status = oom_get_memory(port, 'DOM')
             # DOM = {TX, Rx}Power, temp, bias...
   display_module_status(port, status)
             # your display format here
```

Summary

- Large growth in web content and applications is driving:
  - Growth in bandwidth and changes in data center architectures
  - Subsequent growth in number of optical links
  - Large increase in bit rate and low-power requirements
- 25G, 40G and 100G optics support this growth today with:
  - Smaller module form factors for higher port density
  - Lower power consumption and cost per bit
  - Increased performance to leverage existing fiber infrastructure
- New Ethernet optics are being standardized and under development
  - 50G, 200G, 400G
- Open interfaces are coming to the optical layer.
- Questions?
- Contact Information
  - E-mail: christian.urricariet@finisar.com
  - www.finisar.com

FINISAR®

Thank You