THE STATE OF ETHERNET OPTICS

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

www.ethernetalliance.org
• The Front
The 2016 Ethernet Roadmap

• The Back
Port Density Comparison

• How many ports can you fit in 1U?

Port Density Comparison

- 56 RJ45s/1U
- 56 SFP/1U
- 36 QSFP/1U
- 8 CFP2/1U
- 72 µQSFP/1U
- 100 OBO/1U
- 24QSFP +16 OBO

OBO = On Board Optics
The Ethernet Landscape

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

Source: Dell’Oro Ethernet Switch Forecast
www.ethernetalliance.org
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 \( \mu \)QSFP 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, Director of Marketing, LinkX Interconnects, Mellanox
March 23, 2016  OFC 2016  Anaheim, CA
1GBASE-T CAT5 is **REALLY** big in Asia

IT’S YUGE!
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

- 4-lasers
- Laser Driver
- 4 Detectors
- Detector Amplifier
- PCB
- Power controller
- Microcontroller
- QFP shell
- MPO optical connector
- 8-Multi-mode fibers

100G (4x25G)
SR4, AOC, CWDM4, LR4
Switches & Network Cards – Almost the same BOM costs

40GbE QSFP28 Adapter

32-port 40G Switch

100GbE QSFP28 Adapter

32-port 100G Switch

www.ethernetalliance.org
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 FormFactor Pluggable**
- 1-1.5W
- Duplex LC optical Connector
- “+” 10G;
- “28” for 28G;
- “56” for 56G
- Both use MMF or SMF

**QSFP28**
- 4-Channels
- 8 Fibers or wires
- **Quad Small FormFactor Pluggable**
- 3.5W (5W future)
- MPO 8-fiber parallel Optical connector

---

www.ethernetalliance.org
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
2/10Km reach
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
- AOC

**Transceivers**
- Multi-mode
  - VCSEL Laser
  - GaAs

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

Distance Range:
- DAC: 3m
- AOC: <30m
- Transceivers:
  - Multi-mode: <100m
  - Single-mode: 2Km, 10Km
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

**DAC**
- Server/ToR-to-ToR
- For structured cabling
  - Short Reaches

**SR4**
- For structured cabling
  - Short Reaches

**PSM4**
- For Single-Mode
  - Medium Reaches

**CDM4/LR4**
- For Structured Cabling
  - Long Reaches

- 8-Fiber MPO 500m-2Km
- 2-Fiber LC 2-10Km

**“DAC In the Rack”**
- 3m

**Multi-Mode Optics**
- 3m-100m

**Quad 25G SFP**
- Breakout

**Single-Mode Optics**
- Up to 10Km

**Dual 50G Breakout**

**25G SFP**

**Optical Patch Panel**

**25G SFP Breakout**

**Quad 25G SFP Breakout**

**AOC: 3-50m**

- DAC In the Rack
- Multi-Mode Optics
- Single-Mode Optics

www.ethernetalliance.org
Where Are DAC Links Used?

- DAC Up & down the Rack.
- Linking Servers & storage to Switches.
- Hybrid breakout cables

HPCs may use AOCs everywhere including up & down the rack.
Where AOCs Are Used?

AOCs are used 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?

In Overhead Cable Trays

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

PSM4 too

SR4 - To Optical Patch Panels

SR4 - Into Structured Cabling
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
Servers & Storage linked to ToR Switches

SR4
Transceivers to optical patch panels and in Structured Cabling Pipes & under raised floors where connectors are needed.
<100m

PSM4/CWDM4/LR4
single-mode Transceiver linking to other buildings/floors up to 2Km/10Km.

Under the Floors
Requires optical connectors & SR4s

HPCs may use AOCs everywhere including up & down the rack
Hyper Scale Data centers – Single-mode Territory

PSM4/CWDM4/ LR4
What it looks like IN THE OTHER DIRECTION!

This is ~750,000 sq. feet.

Largest Data center being built by China Petroleum 6.3M sq. feet!
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

<table>
<thead>
<tr>
<th>Lane Rate</th>
<th>No. of Lanes</th>
<th>Data Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gb/s</td>
<td>fiber pairs</td>
<td>λ</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
50G 850nm MMF Optics

Form Factor: SFP56 w/ LC

56Gb/s PAM4 optical eye

<table>
<thead>
<tr>
<th>Lane Rate</th>
<th>No. of Lanes</th>
<th>Data Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gb/s</td>
<td>fiber pairs</td>
<td>λ</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Lane Rate</th>
<th>No. of Lanes</th>
<th>Data Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gb/s</td>
<td>fiber pairs</td>
<td>λ</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
200G MPO 850nm MMF Optics

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

(PSM4 MPO similar use)

Form Factor: QSFP56 w/ MPO

<table>
<thead>
<tr>
<th>Lane Rate</th>
<th>No. of Lanes</th>
<th>Data Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gb/s</td>
<td>fiber pairs</td>
<td>(\lambda)</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>
400G 1310nm SMF Optics

Potential Form Factors:
- CFP8
- CFP16
- QSFP-DD
- OSFP

### Data Rate Table

<table>
<thead>
<tr>
<th>Lane Rate</th>
<th>No. of Lanes</th>
<th>Data Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gb/s</td>
<td>fiber pairs</td>
<td>λ</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>
## 50G PAM4 Ethernet Optics

<table>
<thead>
<tr>
<th>Lane Rate</th>
<th>No. of Lanes</th>
<th>Data Rate</th>
<th>SW code</th>
<th>LW code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gb/s</td>
<td>fiber pairs</td>
<td>λ</td>
<td>Gb/s</td>
<td>(MMF)</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>1</td>
<td>50</td>
<td>SR</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>1</td>
<td>100</td>
<td>SR2</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>2</td>
<td>100</td>
<td>SWDM2</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>1</td>
<td>200</td>
<td>SR4</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>4</td>
<td>200</td>
<td>SWDM4</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>8</td>
<td>400</td>
<td>FR8, LR8</td>
</tr>
</tbody>
</table>
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 @EthernetAlliance 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)

6 rates in 35 yrs

6 new rates happening now

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

© 2019 Cisco and/or its affiliates. All rights reserved.
Why can so much be happening at once? SERDES reuse

Technology availability cannot be separated from market direction.

- 2.5 Gb/s
- 10 Gb/s
- 25 Gb/s
- 50 Gb/s
- 100 Gb/s

# lanes

SERDES speeds
The impact of Cloud Data Center

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

Interconnect Challenges
• New architectures drove 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)

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
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 @EthernetAlliance on Twitter

Visit the Ethernet Alliance on Facebook

www.ethernetalliance.org