Scaling the Cloud Network

Andreas Bechtolsheim
Chairman, Arista Networks Inc
The World has Moved to the Cloud

Billions of Smartphones

Millions of Servers in the Cloud
Creating the Hyper-scale Datacenter Era
Hyper-scale Cloud Network Challenge

How do you interconnect 100,000s of servers such that cloud applications can easily scale?
Idealized Cloud Network

• Ideal cloud network is truly transparent to applications
  • Predictable bandwidth and low latency between all servers
  • 10+ Gbps Bandwidth/server, a few microseconds latency

• This avoids the need for data placement
  • Compute can be anywhere, data can be anywhere
  • Location does not matter since all servers are equal distant

• Old approach was to divide datacenter into clusters
  • Creates a significant burden on application developers
  • It was clear quickly that this was not practical
Consistent bandwidth and latency from any server to any server, allowing applications to scale across the entire data center
Facebook Multi-Level Leaf-Spine Fabric

Layer3 From ToR to Edge
ECMP Load Balancing

Flow based Hashing
Large number of flows

40G -> 100G -> 400G
10X speedup in 5 years

Consistent Performance
No more clusters
Growth in Cloud Network Bandwidth at Facebook
Cloud Network Bandwidth Demand Doubling/Year

Intra-datacenter Bandwidth Growth

Traffic generated by servers in our datacenters

Aggregate traffic

Time

Source: Urs Hoelzle, Google

Driven by Video, AI and ML
Ethernet Speed Transitions are the easiest way to scale the throughput of data center networks, in particular hyper-scale cloud networks.
40G - 100G - 400G Switch Port Transition

100G has passed 40G Ethernet in Ports end of 2017
400G Volume ramps in 2020, passes 100G bandwidth in 2022

Source: Dell’Oro Group Jan 2018 Ethernet Switching Forecast
400G Timeline

First 400G Switch silicon and 400G optics in lab now
Typically one year from first silicon to production release, allowing for one silicon spin on switch chip and optics

Ramping 400G optics is required for volume deployment
Nobody wants a replay of the 100G-CWDM4 experience
Volume availability of 400G optics expected in 2H2019

400G Ports Market Forecast (Dell’Oro Market Research)
2019: 500K
2020: 3M
2021: 5M
400G In the Next-generation Cloud Network

- 400G-ZR
- 400G-LR4
- 400G-FR4
- 400G-DR4
- 400G-AOC
- 400G-CR8
- 8x50G-CR

Source: Brad Booth and Tom Issenhuth Microsoft, IEEE 802.3bs 400G
No Single 400G optics technology addresses all market requirements
In a hyper scale cloud data center, need at least the following:

1. Copper cables for TOR-SERVER (3m max)
2. 400G-SR8 or AOC cables for TOR-LEAF (30m max)
3. 400G-DR4 or 400G-FR4 for LEAF-SPINE (500m - 2km)
4. 400G-LR8 or 400G-CWDM8 for Campus Reach (10km)
5. 400G-ZR for Metro Reach DCO (40km-100km)
Merchant Switch Silicon and Optics
# The Expanding Merchant Silicon Roadmap

<table>
<thead>
<tr>
<th></th>
<th>2008</th>
<th>2012</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optical</td>
<td>Transport</td>
<td>Transport</td>
<td>Transport</td>
</tr>
<tr>
<td>Routing</td>
<td>Core</td>
<td>Core</td>
<td>Core</td>
</tr>
<tr>
<td></td>
<td>Edge</td>
<td>Edge</td>
<td>Edge</td>
</tr>
<tr>
<td>Switching</td>
<td>Spine</td>
<td>Spine</td>
<td>Spine</td>
</tr>
<tr>
<td></td>
<td>Leaf</td>
<td>Leaf</td>
<td>Leaf</td>
</tr>
</tbody>
</table>

- **Proprietary Chips**
- **Merchant Silicon**
Merchant Silicon Driving Network Growth

Source: The 650 Group, Jan 2017
Merchant Silicon Leading Industry in Performance

2008: First ultra-low latency 24-port 10G single chip
2010: First Large Buffer 10G Chip with VOQ Fabric
2011: First 64-port 10G single chip switch
2012: First 32-port 40G single chip
2013: First Large Buffer 40G Chip with VOQ Fabric
2015: First 32-port 100G single chip
2016: First Router 100G Chip with VOQ Fabric
2017: First 64-port 100G single chip
2018: First 32-port 400G single chip
Switch Silicon Bandwidth Growth

[Gbps]

- 2.66X / 3Y
- 2X / 2Y
- 2.5X / 2Y
- 2X / 2Y
- 2X / 2Y
- 2X / 2Y

- 2008
- 2011
- 2013
- 2015
- 2017
- 2019
## Switch Silicon Speed Transitions

<table>
<thead>
<tr>
<th>Lanes</th>
<th>10Gbps</th>
<th>25Gbps</th>
<th>50Gbps</th>
<th>100Gbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>1X</td>
<td>10G</td>
<td>25G</td>
<td>50G</td>
<td>100G</td>
</tr>
<tr>
<td>2X</td>
<td>—</td>
<td>50G</td>
<td>100G</td>
<td>200G</td>
</tr>
<tr>
<td>4X</td>
<td>40G</td>
<td>100G</td>
<td>200G</td>
<td>400G</td>
</tr>
<tr>
<td>8X</td>
<td>—</td>
<td>—</td>
<td>400G</td>
<td>800G</td>
</tr>
</tbody>
</table>

**Server Interface**

**Leaf-Spine Interface**

First Product: 2012, 2016, 2019, 2021

- 4 Years
- 3 Years
- 2 Years
Why are SERDES transitions so Important?

1. They are the easiest way to scale switch performance
2. They drive Optics Standards and the Optics Ecosystem
3. Next Serdes Speed replaces previous one fairly quickly
SERDES Speed Transition Over the Years [% Mix]

- 10G
- 25G
- 50G
- 100G

Yearly Mix:
- 2017: 30% 25G, 40% 50G, 30% 100G
- 2018: 35% 25G, 35% 50G, 30% 100G
- 2019: 40% 25G, 40% 50G, 20% 100G
- 2020: 45% 25G, 35% 50G, 20% 100G
- 2021: 50% 25G, 30% 50G, 20% 100G
- 2022: 55% 25G, 25% 50G, 20% 100G
- 2023: 60% 25G, 20% 50G, 20% 100G
- 2024: 65% 25G, 15% 50G, 20% 100G
Four-Lambda SMF Optics Transitions

The relentless march of Merchant Silicon drives rapid Transitions
The Three Most Important 400G Optics Modules for SMF
400G-DR4

400G Over pSFM (8 Fibers)

- 500m Reach
- MTP Parallel Fiber Connector

Estimated Power: 8W in 2020

8x50G to 4x100G Gearbox

Works across same Fibre Plant as 100G-pSMF today
400G-DR4 can be split into four 100G-DR ports
400G-FR4

400G Over Duplex Fiber

2km Reach (10km with LR4)
Standard LC Fiber Connector

Estimated Power: 8W in 2020

Works across same fiber plant as 100G-CWDM4 today
400G-ZR: 100km Reach DCO

400G-16QAM DSP + Coherent Laser

- 20+ Terabits bandwidth per dark fiber

**Pluggable Form Factor, 15W Power**

- Plugs into standard Switch Router Port

**400G Coherent at the same port density as other Datacenter Optics**
Three Key Optics Transition for 400G SMF

<table>
<thead>
<tr>
<th>FIBER</th>
<th>100G</th>
<th>400G</th>
</tr>
</thead>
<tbody>
<tr>
<td>500m pSMF (8F)</td>
<td>100G-pSMF</td>
<td>400G-DR4</td>
</tr>
<tr>
<td>2km SMF Duplex</td>
<td>100G-CWDM4</td>
<td>400G-FR4</td>
</tr>
<tr>
<td>100km Reach</td>
<td>100G-ColorZ</td>
<td>400G-ZR</td>
</tr>
</tbody>
</table>

Three Key Benefits of making these Optics Transitions:
1. **4X Bandwidth** without Change to Fiber Infrastructure
2. **Forwards Compatible** with 100G Lane Switch Chips
3. **High Volume** drives best availability and economics
Co-packaged Optics
Placement of Optics

- Pluggable optics
- Move optics on-board (COBO)
- Optics Co-packaged with Switch Chip

Co-packaged optics enable much lower-power electrical I/O with a potential 30% power reduction at the system level
Co-Packaged Optics Switch

Packaging Study
(not an actual product)

51.2 Tbps in 1U
128 400G ports

Four Optical Tiles
128 lanes each

Four Laser Sources
driving 128 lanes each

Double Density
compared to pluggable

Image Courtesy of Luxtera
Co-Packaged Optics Benefits

Lower Power / Higher Density
Eliminate high-power SERDES I/O

Cost Advantages
Sub-linear scaling of cost/channel

Greater Reliability
Separating out the laser sources

Diagram:
- Substrate + BGA
- Switching chip die
- P-die
- E-die
- Silicon substrate
Co-Packaged Optics Challenges

Technical Challenges
  Picking the best low-power electrical Interface

Multi-vendor Standardization
  Need to enable multiple vendors to work together

Supply Chain (Switch Chip, Optics, CM)
  Who owns the yield at each manufacturing stage
Solution: Electrical Interposer Connector for Optics

- Optics Engine
- Switch Chip
- Optics Engine

BGA/LCA Array Connector, 0.25mm thick
Interposer Solves the Co-Packaging Problem

Makes Product Manufacturable
High yield merge of fully tested Optics and fully tested switch chip at the CM

Enables Repairability
Failed Optics can be replaced
In manufacturing or even in the field

Supports Configurability
Different Optics can be Configured
For example: 400G-DR4, FR4, LR4, etc
Co-Packaged Optics Summary

**Workable Solution Must Solve all Problems**
Manufacturability, Serviceability, Configurability

**Standardized Electrical Connector is Key**
Easiest solution to the above challenges

**Need Multi-Vendor Standardization**
Define electrical interface and physical form factor

This is a multi-year project, let’s start now
Optics and Standards
Standards Drive New Optics Schedules

Need Standards to drive Volumes
Without Volume, Economics don’t work

Silicon and Optics Developments take a long time
Typical 2-3 years from start of product development

Standards are gating the Speed of Progress
Can’t start product development without a standard

Time needed to develop new Optics Modules is 2-3 Years
## IEEE 802.3 LAN Standards Group

### Table 1

<table>
<thead>
<tr>
<th>Standard</th>
<th>Year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>802.3</td>
<td>1983</td>
<td>10BASE5 10 Mbit/s (1.25 MB/s) over thick coax. Same as Ethernet II (above) except Type field is replaced by Length, and an 802.2 LLC header follows the 802.3 header. Based on the CSMA/CD Process.</td>
</tr>
<tr>
<td>802.3a</td>
<td>1985</td>
<td>10BASE2 10 Mbit/s (1.25 MB/s) over thin coax (a.k.a. thinnet or cheapernet)</td>
</tr>
<tr>
<td>802.3b</td>
<td>1985</td>
<td>10BROAD36</td>
</tr>
<tr>
<td>802.3c</td>
<td>1985</td>
<td>10 Mbit/s (1.25 MB/s) repeater specs</td>
</tr>
<tr>
<td>802.3e</td>
<td>1987</td>
<td>10BASE or FastLAN</td>
</tr>
<tr>
<td>802.3d</td>
<td>1987</td>
<td>Fiber-optic inter-repeater link</td>
</tr>
<tr>
<td>802.3i</td>
<td>1990</td>
<td>10BASE-T 10 Mbit/s (1.25 MB/s) over twisted pair</td>
</tr>
<tr>
<td>802.3j</td>
<td>1993</td>
<td>10BASE-T Ethernet at 10 Mbit/s (1.25 MB/s) over Fiber-Optic</td>
</tr>
<tr>
<td>802.3k</td>
<td>1995</td>
<td>10GBASE-T 10 Gbit/s Ethernet over Fiber-Optic at 1 Gbit/s (125 MB/s) with autonegotiation</td>
</tr>
<tr>
<td>802.3l</td>
<td>1995</td>
<td>Full Duplex and flow control; also incorporates DIX framing, so there’s no longer a DIX/802.3 split</td>
</tr>
<tr>
<td>802.3m</td>
<td>1999</td>
<td>10GBase-SR, 10GBase-LR, 10GBase-ER, 10GBase-SW, 10GBase-LW, 10GBase-EW</td>
</tr>
<tr>
<td>802.3n</td>
<td>1999</td>
<td>A revision of base standard incorporating the above amendments and errata</td>
</tr>
<tr>
<td>802.3o</td>
<td>1998</td>
<td>Max frame size extended to 1502 bytes (to allow “Q-tag”) The Q-tag includes 802.1Q VLAN information and 802.1p priority information.</td>
</tr>
<tr>
<td>802.3p</td>
<td>1999</td>
<td>10GBase-T Ethernet over twisted pair at 1 Gbit/s (125 MB/s)</td>
</tr>
<tr>
<td>802.3q</td>
<td>2000</td>
<td>Link aggregation for parallel links, since moved to IEEE 802.1AX</td>
</tr>
<tr>
<td>802.3r</td>
<td>2002</td>
<td>10 Gigabit Ethernet over fiber; 10GBASE-SR, 10GBase-LR, 10GBase-ER, 10GBase-SW, 10GBase-LW, 10GBase-EW</td>
</tr>
<tr>
<td>802.3s</td>
<td>2002</td>
<td>A revision of base standard incorporating the three prior amendments and errata</td>
</tr>
<tr>
<td>802.3t</td>
<td>2003</td>
<td>Power over Ethernet (15.4 W)</td>
</tr>
<tr>
<td>802.3u</td>
<td>2004</td>
<td>10GBASE-CX4 10 Gbit/s (1,250 MB/s) Ethernet over twinaxial cables</td>
</tr>
<tr>
<td>802.3v</td>
<td>2004</td>
<td>Ethernet in the First Mile</td>
</tr>
<tr>
<td>802.3w</td>
<td>2005</td>
<td>A revision of base standard incorporating the four prior amendments and errata.</td>
</tr>
<tr>
<td>802.3x</td>
<td>2006</td>
<td>10GBASE-LRM 10 Gbit/s (1,250 MB/s) Ethernet over multimode fiber</td>
</tr>
<tr>
<td>802.3y</td>
<td>2006</td>
<td>10GBase-T 10 Gbit/s (1,250 MB/s) Ethernet over unshielded twisted pair (UTP)</td>
</tr>
<tr>
<td>802.3z</td>
<td>2006</td>
<td>Frame expansion</td>
</tr>
<tr>
<td>802.3aa</td>
<td>2006</td>
<td>Isolation requirements for Power over Ethernet (802.3-2005/Cor 1)</td>
</tr>
<tr>
<td>802.3ab</td>
<td>2007</td>
<td>Backplane Ethernet (1 and 10 Gbit/s (125 and 1,250 MB/s) over printed circuit boards)</td>
</tr>
<tr>
<td>802.3ac</td>
<td>2007</td>
<td>Fixed an equation in the publication of 10GBASE-T (released as 802.3-2005/Cor 2)</td>
</tr>
<tr>
<td>802.3ad</td>
<td>2008</td>
<td>A revision of base standard incorporating the 802.3an/ap/aq/as amendments, two corrigenda and errata. Link aggregation was moved to 802.1AX.</td>
</tr>
<tr>
<td>802.3ae</td>
<td>2008</td>
<td>10 Gbit/s EPON</td>
</tr>
<tr>
<td>802.3af</td>
<td>2008</td>
<td>Power over Ethernet (15.4 W)</td>
</tr>
<tr>
<td>802.3ag</td>
<td>2008</td>
<td>10GBASE-CX4 10 Gbit/s (1,250 MB/s) Ethernet over twinaxial cables</td>
</tr>
<tr>
<td>802.3ah</td>
<td>2008</td>
<td>Ethernet in the First Mile</td>
</tr>
<tr>
<td>802.3ai</td>
<td>2008</td>
<td>A revision of base standard incorporating the four prior amendments and errata.</td>
</tr>
<tr>
<td>802.3aj</td>
<td>2008</td>
<td>10GBASE-LRM 10 Gbit/s (1,250 MB/s) Ethernet over multimode fiber</td>
</tr>
<tr>
<td>802.3ak</td>
<td>2008</td>
<td>10GBase-T 10 Gbit/s (1,250 MB/s) Ethernet over unshielded twisted pair (UTP)</td>
</tr>
<tr>
<td>802.3al</td>
<td>2008</td>
<td>Frame expansion</td>
</tr>
<tr>
<td>802.3am</td>
<td>2008</td>
<td>Isolation requirements for Power over Ethernet (802.3-2005/Cor 1)</td>
</tr>
<tr>
<td>802.3an</td>
<td>2008</td>
<td>Backplane Ethernet (1 and 10 Gbit/s (125 and 1,250 MB/s) over printed circuit boards)</td>
</tr>
<tr>
<td>802.3ao</td>
<td>2008</td>
<td>Fixed an equation in the publication of 10GBASE-T (released as 802.3-2005/Cor 2)</td>
</tr>
<tr>
<td>802.3ap</td>
<td>2008</td>
<td>A revision of base standard incorporating the 802.3an/ap/aq/as amendments, two corrigenda and errata. Link aggregation was moved to 802.1AX.</td>
</tr>
<tr>
<td>802.3aq</td>
<td>2008</td>
<td>10 Gbit/s EPON</td>
</tr>
<tr>
<td>802.3ar</td>
<td>2008</td>
<td>Power over Ethernet (15.4 W)</td>
</tr>
<tr>
<td>Standard</td>
<td>Year</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>802.3ba</td>
<td>2010</td>
<td>40 Gbit/s and 100 Gbit/s Ethernet. 40 Gbit/s over 1 m backplane, 10 m Cu cable assembly (4×25 Gbit or 10×10 Gbit lanes) and 100 m of MMF or 40 km of SMF, respectively</td>
</tr>
<tr>
<td>802.3az</td>
<td>2010</td>
<td>Energy-efficient Ethernet</td>
</tr>
<tr>
<td>802.3bd</td>
<td>2010</td>
<td>Priority-based Flow Control. An amendment by the IEEE 802.1 Data Center Bridging Task Group (802.1Qbb) to develop an amendment to IEEE Std 802.3 to add a MAC Control Frame to support IEEE 802.1Qbb Priority-based Flow Control.</td>
</tr>
<tr>
<td>802.3.1</td>
<td>2011</td>
<td>MIB definitions for Ethernet. It consolidates the Ethernet related MIBs present in Annex 30A&amp;B, various IETF RFCs, and 802.1AB annex F into one master document with a machine readable extract. (workgroup name was P802.3be)</td>
</tr>
<tr>
<td>802.3bg</td>
<td>2011</td>
<td>Provide a 40 Gbit/s PMD which is optically compatible with existing carrier SMF, 40 Gbit/s client interfaces (OTU3/STM-256/OC-768/40G POS).</td>
</tr>
<tr>
<td>802.3bf</td>
<td>2011</td>
<td>Provide an accurate indication of the transmission and reception initiation times of certain packets as required to support IEEE P802.1AS.</td>
</tr>
<tr>
<td>802.3-2012</td>
<td>2012</td>
<td>A revision of base standard incorporating the 802.3at/av/az/ba/bc/bd/bf/bg amendments, a corrigenda and errata.</td>
</tr>
<tr>
<td>802.3bk</td>
<td>2013</td>
<td>This amendment to IEEE Std 802.3 defines the physical layer specifications and management parameters for EPON operation on point-to-multipoint passive optical networks supporting extended power budget classes of PX30, PX40, PRX40, and PR40 PMDs.</td>
</tr>
<tr>
<td>802.3bj</td>
<td>2014 (June)</td>
<td>Define a 4-lane 100 Gbit/s backplane PHY for operation over links consistent with copper traces on &quot;improved FR-4&quot; (as defined by IEEE P802.3ap or better materials to be defined by the Task Force) with lengths up to at least 1 m and a 4-lane 100 Gbit/s PHY for operation over links consistent with copper twinaxial cables with lengths up to at least 5 m.</td>
</tr>
<tr>
<td>802.3bw</td>
<td>2015</td>
<td>100BASE-T1 – 100 Mbit/s Ethernet over a single twisted pair for automotive applications</td>
</tr>
<tr>
<td>802.3bm</td>
<td>2015</td>
<td>100G/40G Ethernet for optical fiber</td>
</tr>
<tr>
<td>802.3-2015</td>
<td>2015</td>
<td>802.3bk – a new consolidated revision of the 802.3 standard including amendments 802.3at/ba/bc/bd/bf/bg</td>
</tr>
<tr>
<td>802.3bp</td>
<td>2016 (June)</td>
<td>1000BASE-T1 – Gigabit Ethernet over a single twisted pair, automotive &amp; industrial environments</td>
</tr>
<tr>
<td>802.3bn</td>
<td>2016</td>
<td>10G-EPON and 10G-PASS-XR, passive optical networks over coax</td>
</tr>
<tr>
<td>802.3bz</td>
<td>2016 (Sep.)</td>
<td>2.5GBASE-T and 5GBASE-T – 2.5 Gigabit and 5 Gigabit Ethernet over Cat 5e/Cat 6 twisted pair</td>
</tr>
<tr>
<td>802.3bq</td>
<td>2016 (June)</td>
<td>25G/40GBASE-T for 4-pair balanced twisted-pair cabling with 2 connectors over 30 m distances</td>
</tr>
<tr>
<td>802.3by</td>
<td>2016 (June)</td>
<td>Optical fiber, twinax and backplane 25 Gigabit Ethernet</td>
</tr>
<tr>
<td>802.3bu</td>
<td>2016</td>
<td>Power over Data Lines (PoDL) for single twisted-pair Ethernet (100BASE-T1)</td>
</tr>
<tr>
<td>802.3br</td>
<td>2016</td>
<td>Specification and Management Parameters for Interspersing Express Traffic</td>
</tr>
<tr>
<td>802.3bs</td>
<td>2017 (Dec.)</td>
<td>200GBASE-FC (200 Gbit/s) over single-mode fiber and 400GBASE-FC (400 Gbit/s) over optical physical media</td>
</tr>
<tr>
<td>802.3bc</td>
<td>2017 (Dec)</td>
<td>25 Gbit/s over Single Mode Fiber</td>
</tr>
<tr>
<td>802.3bv</td>
<td>2017</td>
<td>Gigabit Ethernet over plastic optical fiber (POF)</td>
</tr>
<tr>
<td>802.3ce</td>
<td>2017 (March)</td>
<td>Multilane Timestamping</td>
</tr>
<tr>
<td>802.3cb</td>
<td>2018 (TBD)</td>
<td>2.5 Gbit/s and 5 Gbit/s Operation over Backplane</td>
</tr>
<tr>
<td>802.3cd</td>
<td>2018 (TBD)</td>
<td>Media Access Control Parameters for 50 Gbit/s, 100 Gbit/s, and 200 Gbit/s Operation</td>
</tr>
<tr>
<td>802.3bt</td>
<td>2018 (TBD)</td>
<td>Power over Ethernet enhancements up to 100 W using all 4 pairs balanced twisted-pair cabling</td>
</tr>
<tr>
<td>802.3cf</td>
<td>2018 (TBD)</td>
<td>YANG Data Model Definitions</td>
</tr>
<tr>
<td>802.3cg</td>
<td>2019 (TBD)</td>
<td>10 Mb/s Single Twisted Pair Ethernet</td>
</tr>
<tr>
<td>802.3ca</td>
<td>2019 (TBD)</td>
<td>100G-EPON – 25 Gbit/s, 50 Gbit/s, and 100 Gbit/s over Ethernet Passive Optical Networks</td>
</tr>
</tbody>
</table>
IEEE P802.3bs 400GbE Task Force adopted the IEEE 802.3bs 400GbE standard at its September 2015 Interim Meeting.
# History of IEEE 802.3 Ethernet Standards

<table>
<thead>
<tr>
<th>Ethernet Speed</th>
<th>PAR</th>
<th>Standard Ratified</th>
<th>Time (Years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Mbps</td>
<td>1981</td>
<td>1983</td>
<td>2</td>
</tr>
<tr>
<td>100 Mbps</td>
<td>1992</td>
<td>1995</td>
<td>3</td>
</tr>
<tr>
<td>1 Gbps</td>
<td>1995</td>
<td>1998</td>
<td>3</td>
</tr>
<tr>
<td>10 Gbps</td>
<td>1999</td>
<td>2002</td>
<td>3</td>
</tr>
<tr>
<td>40/100 Gbps</td>
<td>2007</td>
<td>2010</td>
<td>3</td>
</tr>
<tr>
<td>400 Gbps</td>
<td>2014</td>
<td>2017</td>
<td>3</td>
</tr>
</tbody>
</table>

**Problem:** New Optics can’t wait for three years of standards process.
Problems with IEEE 100G Optics Standards

IEEE 802.3ba (100G Ethernet) standardized two 100G optics:
100G-LR4 (10km reach duplex fiber) and 100G-SR10 (100m reach 10x10)
Neither addressed the large cloud network market potential

IEEE 802.3bm (lower cost 100G optics standards) tried to correct this
Proposed 4x25G 500m reach duplex SMF (100G-CWDM4) and parallel SMF
After 2 years of meetings, neither proposal was accepted as an IEEE standard

IEEE Voting rules prevented standardization of the most common 100G Optics in use today
Similar Situation with 400G Optics

<table>
<thead>
<tr>
<th>802.3bs Standard</th>
<th>Description</th>
<th>Reach</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>400G-SR16</td>
<td>16x25G lambda, 32-MMF</td>
<td>100m</td>
<td>Nobody will use this</td>
</tr>
<tr>
<td>400G-FR8/LR8</td>
<td>8x50G lambda, duplex SMF</td>
<td>2/10km</td>
<td>Limited Market Potential</td>
</tr>
<tr>
<td>400G-DR4</td>
<td>4x100G lambda, 8-SMF</td>
<td>500m</td>
<td>High-volume for pSMF</td>
</tr>
</tbody>
</table>

IEEE 802.3bs did not standardize the highest volume 400G optics for cloud, including 400G-FR4 and 400G-LR4.
The 100G Lambda MSA defines 100G PAM-4 optical signaling and encoding, FEC and link characteristics for 100G and 400G applications using 100Gb/s per optical channel for 2km and 10km reaches. The MSA will leverage the IEEE 802.3 draft specifications and methodology for similar 100 Gb/s single channel specifications over 500m of single-mode fiber – specifically 100GBASE-DR and 400GBASE-DR4. The MSA specifications will be consistent with the IEEE specifications in PAM4 signaling and RS (528,544) FEC to maintain compatibility with existing system and PHY technology currently under development.

Currently, the most popular 100Gb/s optical standards such as 100GBASE-LR4, 100G-CWDM4, 100G-PSM4, rely on 25 Gb/s optical lanes that align with 25Gb/s SERDES commonly used on ASICs for the switching, routing and transport applications. As ASIC SERDES increase in speed, it is necessary to increase the optical channel speeds to avoid additional cost that comes from needing to translate to slower speeds. Cost savings can further be realized by reducing the number of optical lanes and increasing the speed from four times 25Gb/s per lane to a single lane of 100Gb/s. It has been recognized by the IEEE and the members of this MSA that a single optical lane of 100Gb/s can be at least 40% lower cost than four lanes of 25G. The MSA members expect 400 Gb/s specifications defined by the MSA using 100Gb/s per optical channel will be much more conducive to high density, and lower cost, implementations in module form factors and networking systems.

Source: www.100glambda.com
### 100G Lambda MSA SMF Optics Standards

<table>
<thead>
<tr>
<th>Speed/Fiber</th>
<th>500m</th>
<th>2km</th>
<th>10km</th>
</tr>
</thead>
<tbody>
<tr>
<td>100G Duplex Fiber</td>
<td>100G-DR</td>
<td>100G-FR</td>
<td>100G-LR</td>
</tr>
<tr>
<td>400G Parallel Fiber</td>
<td>400G-DR4</td>
<td>400G-DR4</td>
<td>TBD</td>
</tr>
<tr>
<td>400G Duplex Fiber</td>
<td>400G-FR4</td>
<td>400G-FR4</td>
<td>TBD</td>
</tr>
</tbody>
</table>

**IEEE 802.3 Specs**

- 100G Lambda MSA
- Future Work

**Timeline from announcement of 100G Lambda MSA to release of first set of specifications was four months (9/12/2017 to 1/9/2018)**
How do Optics MSAs work?

- The outcome of any standards group activity can be predicted by (1) the group constituency and (2) its voting rules
- With MSAs, members have a shared goal to get a spec done. There are typically weekly meetings with active participation
- As a result, time lines become compressed. Most MSAs complete their specification work in a couple of months, not years.
- MSAs are driven by members that have shared goals. There are no dissenting parties blocking progress
Optics MSA and Related Standard Efforts

400G Optics
- 4x100G-LAMBDA
- 400G-ZR
- 400G-CWDM8
- 400G-SR8
- 400G-SR4.2

100G Optics
- 100G-LAMBDA
- 100G-CWDM4
- 100G-PSM4

Form Factors
- OSFP
- QSFP-DD
- uQSFP
- D-SFP
- SFP-DD

Need Standards for everything not included in 802.3
One cannot build new products without a standard
Next-gen Optics Standards Summary

Standards for Next-gen 400G Optics are needed now
400G switch silicon is in the lab, products will ship in volume in 2019

MSAs are taking the initiative to create these standards
This is working well, specifically with the 100G Lambda MSA

Traditional Standards Bodies have not worked well for optics
Multi-year processes are simply too slow to make good choices

OCP can play a major role promoting and advocating optics standards that are good for cloud networks