Regarding the Views Expressed

The views expressed on IEEE standards and related products should NOT be considered the position, explanation, or interpretation of the Ethernet Alliance.

Per IEEE-SA Standards Board Bylaws, Dec 2016

“At lectures, symposia, seminars, or educational courses, an individual presenting information on IEEE standards shall make it clear that his or her views should be considered the personal views of that individual rather than the formal position of IEEE.”
Our Mission and Priorities

We are a global community of end users, system vendors, component suppliers and academia

Our Mission

• Promote existing and emerging IEEE 802 Ethernet standards
• Accelerate industry adoption
• Demonstrate multi-vendor interoperability

2017 Strategic Priorities

• Support Existing Technology Deployment
• Support IEEE 802 Standards Development
• Marketing & Education

The Voice of Ethernet
### Agenda

<table>
<thead>
<tr>
<th>Topic</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>The State of Ethernet</td>
<td>John D’Ambrosia, Huawei</td>
</tr>
<tr>
<td>Enabling Future Ethernet Connectivity</td>
<td>Nathan Tracy, TE Connectivity</td>
</tr>
<tr>
<td>Maximizing Ethernet Performance for Most Demanding Workloads</td>
<td>Ran Almog, Mellanox</td>
</tr>
<tr>
<td>Test and Measurement Considerations for Ethernet Applications</td>
<td>David Rodgers, Teledyne LeCroy</td>
</tr>
<tr>
<td>Q &amp; A</td>
<td></td>
</tr>
</tbody>
</table>
THE STATE OF ETHERNET

John D’Ambrosia
Huawei

November 16, 2017
Top500 “Segments”

55%
Top500 “Industry” Segment

Ethernet 81%
<table>
<thead>
<tr>
<th>Electrical Interface</th>
<th>Backbone</th>
<th>Twin-ax</th>
<th>BASE-T (4 Pair)</th>
<th>MMF</th>
<th>500m SMF</th>
<th>2km SMF</th>
<th>10km SMF</th>
<th>40km SMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>10GBASE- XGB, XAU, XFI, SFI</td>
<td>KX4</td>
<td>CX4</td>
<td>T</td>
<td>SR</td>
<td>LR</td>
<td>ER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40GBASE- XLAUI</td>
<td>KR4</td>
<td>CR4</td>
<td>T</td>
<td>SR4/aSR4</td>
<td>PSM4</td>
<td>FR</td>
<td>LR4</td>
<td>ER4</td>
</tr>
<tr>
<td>50GBASE- LAUI-2/50GAU-2 50GAU-1</td>
<td>KR</td>
<td>CR</td>
<td>SR</td>
<td>FR</td>
<td>LR</td>
<td>ER</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Gray Text = IEEE Standard   Red Text = In Standardization   Green Text = Future Possible Standard
Blue Text = Non-IEEE standard but complies to IEEE electrical interfaces
I/O Escape Forcing Transition to Higher Lane Speeds

Single ASIC IO capacity doubling every ~ 2 years

Current BGA practical maximum ~ 70mm package (due to coplanarity / warpage)

The New Ethernet Paradigm: Follow the SerDes
# IEEE 802.3 Standards Activities

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Schedule</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE p802.3bs</td>
<td>200 Gb/s and 400Gb/s Ethernet (electrical interfaces / optical PHYs)</td>
<td>Dec 2017</td>
</tr>
<tr>
<td>IEEE p802.3bt</td>
<td>4 Pair Power-Over-Ethernet</td>
<td>Sept 2018</td>
</tr>
<tr>
<td>IEEE p802.3ca</td>
<td>25Gb/s, 50 Gb/s, 1000Gb/s EPON</td>
<td>Apr 2020</td>
</tr>
<tr>
<td>IEEE p802.3cb</td>
<td>2.5Gb/s and 5Gb/s Backplane</td>
<td>Jun 2018</td>
</tr>
<tr>
<td>IEEE p802.3cc</td>
<td>25 Gb/s Ethernet over SMF (10 / 40 km)</td>
<td>Dec 2017</td>
</tr>
<tr>
<td>IEEE p802.3cd</td>
<td>50Gb/s, 100 Gb/s, 200Gb/s Ethernet (electrical interfaces, Copper PHYs, Optical PHYs)</td>
<td>Sept 2018</td>
</tr>
<tr>
<td>IEEE p802.3.2</td>
<td>YANG Data Models</td>
<td>June 2018</td>
</tr>
<tr>
<td>IEEE p802.3cg</td>
<td>10 Mb/s Single Twisted Pair</td>
<td>June 2019</td>
</tr>
<tr>
<td>IEEE p802.3ch</td>
<td>Multi-gig Automotive Ethernet</td>
<td>Undefined</td>
</tr>
<tr>
<td>Study Group</td>
<td>Beyond 10km Optical PHYs (50Gb/s, 100Gb/s, 200Gb/s, and 400Gb/s Ethernet)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 Gb/s Electrical Interfaces and Electrical PHYs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 Mb/s Backplane Ethernet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Next-gen 200G &amp; 400G PHYs for MMF</td>
<td></td>
</tr>
</tbody>
</table>
The Importance of Multi-vendor Interoperability

**Industry investment**

• **On-going Work**
  – PoE (802.3af / 802.3at)
  – 2.5G / 5G / 10G BASE-T
  – 25GbE
  – 100GbE

• **Future**
  – 4 Pair PoE
  – 25 GbE (10 km / 40 km)
  – 50 GbE
  – 200 GbE / 400 GbE
  – New Signaling
  – New Optical Form Factors
Ethernet Alliance PoE Certification Program

- According to Dell’Oro
  - 750M PoE Enabled Switch Ports over Next 5 Years
  - Hundreds of Millions of PoE Devices over Next 5 Years
- Distinguishes products based on IEEE 802.3 standards in the market
- Open to general industry
- [https://ethernetalliance.org/poe-cert/](https://ethernetalliance.org/poe-cert/)

Example – Class 3 PSE

Example – Class 1 PD

THE ETHERNET ALLIANCE™, EA™, THE EA LOGO™, EA CERTIFIED & PD Logo™, and EA CERTIFIED & PSE Logo™ are trademarks, service marks, and certification marks of The Ethernet Alliance in the United States and other countries. Unauthorized use strictly prohibited.
Connectivity Challenges For The Next Generation

• With 100G Ethernet, we started with 10 x 10Gbps electrical interfaces, then transitioned to 4x25Gbps
  – Enabled a narrower interface (more density)
  – Enabled lower power
• What comes next?
• 50Gbps signaling further reduces the interface width
• But the per port rate needs to get to 400Gbps
Where Are The Data Rates Headed?

Ethernet Switch – Data Center: Total Shipments

Hyperscale performance is transitioning to 50Gbps

Data provided by 650 GROUP. Further distribution is prohibited.
• Current QSFP Input Output (I/O) port allows 32 to 36 ports per 1RU enclosure
• Each port has 4 electrical channels
• At 25Gbps, enables 3.6Tbps via 100Gbps ports
• At 50Gbps, enables 7.2Tbps via 200Gbps ports
• But we need 400Gbps ports........
Challenges to 400Gbps Connectivity

• 400Gbps requires 50Gbps signaling and 8 electrical channels
• With a doubling of electrical channels from 4 to 8, how do we fit at least 32 ports in a 1RU enclosure to yield 12Tbps?
• Need a new module/connector form factor solution
  • More lanes
  • Improved electrical performance
  • Improved thermal performance
• The Ethernet Community is up to the task
Three New Form Factors Being Developed

• QSFP-DD
  – Goes to 8 channels, maintains ability to accept legacy QSFP ports (backwards capability)

• OSFP
  – New form factor with 8 channels. Includes integrated heat sink for improved thermal performance, uses an adapter for backwards performance

• COBO
  – Defines embedded optics modules. Fits 32 400G modules on a 1RU linecard, allowing more airflow and improved thermal performance
Pluggable or Embedded?

• Pluggable modules are the traditional approach but put all the thermal dissipation of the modules at the face plate and block some airflow

• Embedded optics spread the thermal dissipation, allow larger heat sink and create more area for airflow at the faceplate
Critical Requirements of a Connector

- **Signal Integrity**
  - 4 row style connector
  - 2 row style connector
  - 0.6mm contact pitch for COBO and OSFP, 2 row
  - 0.8mm contact pitch for QSFP-DD, 4 row
  - More channel margin with the simpler 2 row solution
  - Copper cable reach is 0.5m longer for OSFP

All are designed to the same IEEE electrical requirements (industry consensus)

- **Thermal management (sliding, integrated, attached)**
  - COBO has more room for heat sink, 15+ W
  - QSFP-DD sliding heat sink, 12W?
  - OSFP integrated heat sink, 15W

- **Backwards compatible**
  - COBO is disruptive, no backward capability
  - QSFP-DD does accept QSFP, i.e. 40G, 100G
  - OSFP with adapter accepts QSFP
400G Summary

• There will be choices for 400Gbps implementation
• This discussion has focused on IO, but backplane solutions are available in the market as well
• Important decisions must be made in the connector selection regarding future rates, design margin, thermal performance and cable reach
• Ethernet’s silicon, optic, connector and cable suppliers keep on innovating .........
What’s Next in Connectivity?

• Remember this slide?

  Ethernet Switch – Data Center: Total Shipments

  Data provided by 650 GROUP. Further distribution is prohibited

• 100Gbps per differential pair developments have started!
• IEEE has a Study Group
• OIF has a project
• Component suppliers have development programs

Further innovations to address the challenges of signal integrity, reach, thermal, and density will continue!
MAXIMIZING ETHERNET PERFORMANCE FOR MOST DEMANDING WORKLOADS

Ran Almog
Mellanox Technologies

November 16, 2017
Delivering Highest DC Return on Investment

90% of the Top 10 Oil and Gas Companies
60% of Top 5 Pharmaceutical Companies
100% of Top 10 Automotive Manufacturers

Connects All of 40G Ethernet Systems
Enabling Most Efficient AI Platforms

- OCP Big Sur Artificial Intelligence Platform
- Real Time Fraud Detection
- Machine Learning System with 400Gb/s
- 18X Speedup For Image Recognition
- 4X Speedup For Image Recognition
- Data Analytics Image Recognition
- World Record For Data Sort, 3X Faster

The First 100G Ethernet System on The TOP500 List
Maximizing Ethernet Performance with RoCE

**High Performance Network**
- Highest throughput
- Lowest latency

**Advanced Congestion Control**
- Early detection and prevention
- Reliable and predictable

**Efficient Network Utilization**
- Higher server productivity, cost and power savings
- Higher availability of CPU resources to the application

**Extensive Visibility**
- Optimize network behavior
- Detect, prevent and troubleshoot
RoCE Simplified

- Soft RoCE
- RoCE for Lossy networks
- Out of the box experience
- ECN capable network
- Automatic PFC configurations
Protect Against Congestion Victims

- Nodes 1-4 send traffic in 50% BW each
- 1, 3 and 4 send to 5
- 2 sends to 6

- Port C is a root of congestion and starts marking the ECN bit
- Port B is also congested due to back pressure from C and sends Flow Control to A

- A identifies it’s a “victim”, hence doesn’t set the ECN bit
- 5 gets the ECN mark and echoes 1, 3 and 4

- 1, 3 and 4 receives the ECN echo and reduce their sending rate
- 2 sending rate was not affected
Open Composable Networks

- Network OS Choice
- Open APIs
- Automation
- End-to-End Interconnect
TEST AND MEASUREMENT CONSIDERATIONS FORETHERNET APPLICATIONS

David J. Rodgers
Teledyne LeCroy PSG

November 16, 2017
Basic Considerations!

• Testing and Validation Needs to Keep Up
• Integrating Higher Speeds in the SAN
  – 10/40GbE, now 25/100GbE and 50/200GbE right around the corner
  – Closing in on the “Holy Grail” of 100GbE
• Ethernet Fabrics Fueling Storage Explosion
  – Speed and Optimization meeting QOS Expectations
    • iSCSI, FCoE, NVMf, NFS, IBXoE, FCIP, iSER, iWARP, RoCE, Routable RoCE (v2)
• Conforming to Standards
• Keep on Budget and Keep Users Happy!
Conformance to Standards

• Ethernet Standards Evolving at Breakneck Pace
  • Automotive
  • 25GbE to 100GbE, now 50GbE to 200GbE
  • Soon, 100GbE to 400GbE

• Storage Solutions Leveraging Speed
  – FCoE, iSCSI
  – NVMf

• Standards beget Interoperability?
  – Interpretation and implementation differences abound
    • Increase in speed has added complexities
Interoperability in the Real World

• No two vendors implementations are identical
• There is a “protocol” to the phy
  – Auto-negotiation
  – Link Training
  – FEC
• New Speeds adding new complexities
  – NRZ vs PAM4 signaling
• Testing needs to be “standardized” and repeatable
  – Interop PlugFests, 3rd party testing services
Key Interoperability Challenges

• Identifying Participants
  – Characterizing Functionality of All Ecosystem Players

• Determining Root Cause

• Crafting the Solution

• Remediation Validation
  – Test the fix

• Timely Resolution!
Effective Observation

Fabric Management
Utility/Hypervisor

Traffic Tap and DPI (Wire Shark)

Line Rate Analysis
Adding Line Rate Analysis

• **Purpose Built Protocol Tools!**
  – Compliment to, not replacement for Traditional Tools

• **Invisible to the Fabric Under Test**
  – Unbiased traffic capture of all layers

• **Agnostic to Participants, Traffic Type, and Transport Media**
  – Real Time Triggering, Post Capture Data Analysis

• **Traffic Modification/Error Injection Functions**
  – *Determine Root Cause!*
  – Proof Remediation before deployment
Testing the Fix

• Once remediation is applied, does it work?
• Lab Recreation of the offending condition(s)
  – Proof of concept on the bench – Analyzer/Jammer
  – Observation on the link - Analyzer
• Test out additional corner case scenarios
  – Prescreen pending releases
  – Reuse profiles/test cases
Investigative Challenge

“Often times (the problem) requires the recreation of a given fault in a lab environment, which is problematic without an appropriate toolset. For this specific purpose an in-band protocol analyzer and error injection utility is now an integral part of my troubleshooting arsenal.”

“The error injection capability of these tools is of even higher importance to me however.

Reference excerpted from Teledyne LeCroy user case study.
Error Injectors

- Provide effective, programmatic recreation of faulty conditions/events
- Drop commands, responses
- Insert Errors at all levels
- Counters and Timers for true-to-traffic conditions
- Best when tightly integrated to the analysis tools
ReCap

- Ethernet is a Juggernaut
- Content delivery and Storage Demands are High
- Consistent and predictable interoperation is mandatory
- Speed adds exponential Influences on the EcoSystem
- Testing, Testing, Testing

*Tool Sets and Methodologies Must Evolve*
QUESTIONS?