New Data Center and Transport Interconnect Technology

Tom McDermott
Fujitsu Network Communications, Inc.
October 30, 2014
Changing Requirements

- Datacenters are expanding in size.
  - Continued need for 100m / MMF intra-DC connection.
  - Growing need for 500m~2km / SMF intra-DC connections.
  - Interest in breakout / fanout capabilities

- Ethernet WAN / Transport Interconnect.
  - 2km, 10km, 40km – continued need. SMF

- 100GE.
  - 2nd generation of devices becoming available.
  - Smaller size, lower cost, lower power.

- 400GE project started.
  - Intra-DC, Inter-DC, and Metro reach objectives.

- New class of coherent tunable pluggable T/Rs.
  - Metro reach & compatible with ROADM
  - Optical reach over ROADM still TBD.
Some Ethernet PMD Nomenclature

- **100GBASE-SR4**
  - 100GE
  - Short Reach (100 m) over MMF
  - 4 optical lanes

- **100GBASE-LR4 = 10 km over duplex SMF**
- **ER4 = 30 km (40 km with engineering).**
- **100GBASE-CR4 = 4 Copper twin-ax cables (each direction) ~ 5m**
- **100GBASE-KR4 & KP4 = 4 Copper backplane diff traces (each direction) ~ 1m**
Breakout / Fanout

- Growing interest in breakout / fanout functionality

- Increase density in top of rack (TOR) switches
  - Example: a 40GE 40GBASE-SR4 module could be operated in two modes:
    - 4 x 10GE – QSFP+ physically smaller than 4 separate 1 x 10GE plugs
    - 1 x 40GE – data is striped across the 4 lanes.
  - Provides high-density 10GE fanout from TOR switch
  - Fiber-ribbon to duplex fiber conversion is external.

- Recent discussion: 100GE 4-lane module (4 x 25G VCSEL):
  - 4 x 25GE - An IEEE 802.3 25GE Study Group is underway.
  - 1 x 100GE - 802.3bm currently balloting 100GBASE-SR4
Density Evolution 10GE → 100GE
2002 → 2014

Front-Panel Bandwidth Density (GB per 10 mm)

<table>
<thead>
<tr>
<th>Port Bandwidth</th>
<th>10G</th>
<th>40G</th>
<th>100G</th>
</tr>
</thead>
<tbody>
<tr>
<td>10G</td>
<td>SFP+</td>
<td>4 x SFP+</td>
<td>CPAK</td>
</tr>
<tr>
<td>40G</td>
<td>XFP</td>
<td>(x 4) XFP</td>
<td>CFP2</td>
</tr>
<tr>
<td>100G</td>
<td>QSFP (MMF)</td>
<td>QSFP+ (MMF)</td>
<td>CFP4</td>
</tr>
</tbody>
</table>

QSFP28 (2014)
CFP (2008-2010)

All Rights Reserved, ©2014 Fujitsu Network Communications
100GE Package Evolution

Today: CFP:
- 10x10MSA
- 100GBASE-LR4
- 100GBASE-ER4
- 100GBASE-SR10

Future: Digital DWDM

Today: CFP2:
- 100GBASE-LR4
- 100GBASE-ER4
- 10x10MSA? (v2)

Future: Analog Coherent DWDM

10 x 10G Digital Electrical I/O +
4 x 25G Digital Electrical I/O +
4 x 25G Analog I/O

QSFP28: 100GBASE-SR4, 100GBASE-CR4 (copper)

4 x 25G Digital Electrical I/O
CFP / CFP2 / CFP4 Client Devices

Gen 1
10:4 Gearbox + Retime
- 10 x 10G Digital
- 4 x Driver
- 4 x Receiver
- 4 x TOSA + WDM
- WDM + 4 x ROSA
- Optical I/O
- ~24w CFP Client

Gen 2
4 x 25G Digital
- 10x10MSA version is 10 parallel channels at 10G
- 4 x Retime + Driver
- 4 x Receiver + Retime
- Integrated Multidevice TOSA+ROSA + WDMs
- Optical I/O
- 8~12w CFP2 Client

Future
4 x 25G Analog
- 4 x Retime + Driver
- 4 x Receiver + Retime
- 4 x VCSEL 4 x PIN
- Optical I/O 100m
- ~4.5w CFP4 or QSFP28 Client
100GBASE-SR10 vs. 100GBASE-SR4
Generation 1, CFP form factor : Generation 2, QSFP28 form factor

Generation 1
100GBASE-SR10
10G / lane

Generation 2
100GBASE-SR4
25G / lane
**QSFP28 Client Device**

- IEEE 802.3bm considered retimer-less interface for 20m (resulting in ~1.5w QSFP28), but rejected.
  - 100m MMF requires retimer.
  - 20m market was too small to develop independent interface type from 100m.
  - Might be used in future very short reach applications.
- Proprietary Active Optical Cable interfaces adopted QSFP28 / 1.5w form factor.
Package Trends – 100GE Client Optics

- 1\textsuperscript{st} generation: CFP: for MMF and SMF (actual)
- 2\textsuperscript{nd} generation: CFP2, CPAK (actual)
- 3\textsuperscript{rd} generation: CFP4, QSFP28 (forecast, actual)
Some 400GE Adopted Objectives

Some Initial Ideas

- Support a BER of better than or equal to $10^{-13}$ at the MAC/PLS service interface (or the frame loss ratio equivalent)
- Provide appropriate support for OTN
- Specify optional Energy Efficient Ethernet (EEE) capability for 400 Gb/s PHYs
- Provide physical layer specifications which support link distances of:
  - At least 100 m over MMF
  - At least 500 m over SMF
  - At least 2 km over SMF
  - At least 10 km over SMF

- 25G x 16 VCSEL looks promising for MMF…
  - Small package (CDFP?), new 16-position MMF MT/MP connector.
  - Potential for very high density and lower cost.
- Several formats including $4\lambda$-DMT and $8\lambda$-PAM4 being studied for duplex SMF.
Form Factor Evolution 400GE Client

Ref: oif2012.346.02

Gen 1: CDP Client: 400GBASE-???
16 x 25G Digital Electrical I/O

Gen 2: CDP2 Client: 400GBASE-???
8 x 50G Digital Electrical I/O
16x25G electrical lanes 8x50G electrical lanes

CDP4
4x100G electrical lanes ??

Far Future: CDP4 Client: 400GBASE-???

Are electrical lanes practical at 100G?

Note: originally called CDFP in the OIF presentation. Renamed to CDP.
Provides 16 x 25G electrical interface.

Optimized for 16-wide VCSEL based MMF. TIA concurrently working on standard for 2 x 16 MMF fiber ribbon and MTO/MTP MMF connector.

Could be called 400GBASE-SR16 if IEEE standardized in 2017.

Desire by some in IEEE to have a 4 x 100GBASE-SR4 breakout mode of operation.

Card cage is small, but cage + plug in roughly the same size as 4 x QSFP28.
IEEE elected to work on 400GE as opposed to 1TE due to significant technical issues.
- Very small size, low power, low cost and high production volume are key requirements for intra-datacenter optics (TOR Ethernet switch interconnect).

Extrapolating current technology: 40 channels x 25G VCSEL.
- Would require 40-wide MMF fiber ribbon x 2 (80 total fibers).
- Electrical interconnect to the package would be massive.

Faster electrical signaling speeds and multi-level modulation may be possibilities.
- Not ready for the IEEE application.

Proposals for DMT and PAM-4 technology for 400GE.
- 500m – 10km duplex SMF.
- Future: Might form basis for 1TE Duplex SMF Ethernet client interconnect.
Coherent DWDM Packaging
Network Optics

- DWDM Coherent 100G optics moving towards CFP2 form factor.

- Two approaches:
  - Digital Interface to Optics
    - Places DSP inside the optics package.
    - Pros: well-quantified optical performance.
    - Cons: DSP power consumption makes thermal issues extremely challenging.
  - Analog Interface to Optics.
    - DSP is outside the optics package.
    - Pros: multivendor optical sourcing, low cost, thermal management fits CFP2 package well. CFP4 may become practical later.
    - Cons: PC board layout and DSP placement may impact optical span performance in strong or weak ways (unknown at this time).
CFP2 Analog DWDM Coherent Module

- Analog interface and copper traces between DSP and CFP2 optics will impact the span reach and optical performance.
  - Skew, reflection, attenuation, crosstalk.
- Quantification of the impact requires actual DSP, PC Board, and optical CFP2 modules, and characterization.
- Standard CAUI-4 interface and copper traces between Host and Module.
- Optical span performance not dependent on host.
- CFP package currently required due to high power dissipation.
Analog CFP2 Issues
Digital CFP Issues

- Interface to Analog CFP2 DWDM Coherent transceiver.
  - DSP is external to the optics package.
  - Copper traces connect analog I/O from DSP to Optics.
  - Signaling rate is about 32 Gbps. Digitizing rate of DSP ADC is about 64 Gsps.
  - Future generations (class 2, 3) planned to additionally support 16QAM.

- Analog interface, DSP, copper traces, optical module all have an impact on the optical span performance.
  - Too early to tell how tightly constrained the optical performance will be.
    - Performance may be weakly or strongly impacted by PCB motherboard, DSP, copper traces, and/or optics module changes.
    - Might be addressable with conservative design rules.
    - Might be addressable by complete chain characterization (but this can get complex).
      - Optics + DSP + PCB motherboard to determine performance.

- Digital module is large due to high power DSP inside optical envelope.

- Limited supply base for digital version.