#	Item	Details	Motion	Reference File
1	Document Outline	Move to adopt the draft outline, based on slides 3 thru 7 of "ganga_02_0508.pdf" as the basis for the first draft of P802.3ba.	#1 of May 2008	http://www.ieee802.org/3/ba/public/may08/ganga_02_0508.pdf
2	Nomenclature	Move to adopt the nomenclature, based on slide 8 of "ganga_02_0508.pdf" as the basis for the first draft of P802.3ba	#2 of May 2008	http://www.ieee802.org/3/ba/public/may08/ganga_02_0508.pdf
3	Architecture	Move that the 802.3ba Task Force adopt slides 4 thru 9 as the 40/100G architecture as proposed in "ganga_01_0508.pdf" with the inclusion of an optional n-lane x 10.3125GBd electrical interface for PMD service interface (to slides 5,6, and 9), as baseline.	#3 of May 2008	http://www.ieee802.org/3/ba/public/may08/ganga_01_0508.pdf
4	XLGMII / CGMII	Move to adopt "gustlin_02_0508.pdf" as the baseline for the XLGMII and CGMII logical interfaces.	#4 of May 2008	http://www.ieee802.org/3/ba/public/may08/gustlin_02_0508.pdf
5	PCS	Move to adopt "gustlin_01_0508.pdf" as the baseline for the 40GbE and 100GbE PCS.	#5 of May 2008	http://www.ieee802.org/3/ba/public/may08/gustlin_01_0508.pdf
6	MMF PMD	Move that the 802.3ba Task Force adopt the parallel PMD proposal and tables on pages 6, 8, 9 and 10 of (pepeljugoski_01_0508) as the baseline proposal of the work of the task force towards writing the first draft standard for 40GBASE-SR4 and 100GBASE-SR10.	#6 of May 2008	http://www.ieee802.org/3/ba/public/may08/pepeljugoski_01_0508.pdf
7	OTN Compatibility	Move to adopt "trowbridge_01_0508.pdf" as the baseline for the "Appropriate support for OTN" with the inclusion of "and pending concurrence of the 802.3 working group" prior to the last bullet of slide 11.	#8 of May 2008	http://www.ieee802.org/3/ba/public/may08/trowbridge_01_0508.pdf
8	100GE 40km PMD	Move to adopt the proposals in "cole_02_0708.pdf" as the baseline proposal for the 100GE 40km SMF PMD in place of "cole_02_0508.pdf".	#2 of July 2008	http://www.ieee802.org/3/ba/public/jul08/cole_02_0708.pdf
9	100GE 10km PMD	Move to adopt slides 9 to 11 of "anslow_01_0708.pdf" as the baseline proposal for the 100GE 10km SMF PMD in place of "cole_01_0508.pdf".	#1 of July 2008	http://www.ieee802.org/3/ba/public/jul08/anslow_01_0708.pdf
10	Backplane PMD	Move to adopt mellitz_01_0508.pdf as the baseline for the 40GbE backplane PHY (40GBASE-KR4).	#16 of May 2008	http://www.ieee802.org/3/ba/public/may08/mellitz_01_0508.pdf
11	PMA	Move to adopt the PMA baseline proposal described in slides 5-15 of "trowbridge_01_0708.pdf" for the first draft of 802.3ba with the following amendments: • Remove the word "logical" from the PMD service interface on slides 5 and 6. • Remove the word "abstract" from the PMD service interface on slide 10 and 13. • Replace "clock source" with "clock information" on slides 10, 12, 13, 14.	#3 of July 2008	http://www.ieee802.org/3/ba/public/jul08/trowbridge_01_0708.pdf
12	XLAUI / CAUI	Adopt ghiasi_01_0708.pdf pages 17, 18, 19, 20, and 29 as the baseline proposal for the XLAUI/CAUI Electrical Interface with the following addition: All links are AC Coupled with maximum single ended voltage difference from -0.3V to 4.0V.	#4 of July 2008	http://www.ieee802.org/3/ba/public/jul08/ghiasi_01_0708.pdf
13	PMD Service Interface	Move to adopt the PMD service interface proposal for 40GBASE-SR4 and 100GBASE-SR10 as defined in petrilla_01_0708 pages 24 & 25 with the following additions: 1) Use associated templates in pages 26, 27 & 28, and 2) use the template in page 22 with limit characteristics using 802.3ap KR channel equations (coefficients TBD) referencing nicholl_01_0708 for guidance.	#5 of July 2008	http://www.ieee802.org/3/ba/public/jul08/petrilla_01_0708.pdf
14	Cu PMD	Move to adopt "diminico_01_0708.pdf" as the baseline for 40GBASE-CR4 and 100GBASE-CR10 with the revision to bullet 2 slide 19 in "diminico_02_0708.pdf".	#3 of July 2008	http://www.ieee802.org/3/ba/public/jul08/diminico_02_0708.pdf
15	40GE 10km PMD	Move to adopt 4x10G CWDM as per "cole_01_0908.pdf" as the baseline proposal for the 40GE 10km SMF PMD objective.	Motion #3 Reconsidered of Sept 2008	http://www.ieee802.org/3/ba/public/sep08/cole_01_0908.pdf

Chief Editor's Report

Ilango Ganga, Intel
Editor-in-Chief, IEEE P802.3ba Task Force

May 13, 2008

Existing clauses

- Clause 1 – Introduction to 802.3
  - Add appropriate normative references, definitions, description of compatibility interfaces, and abbreviations
- Annex A –Bibliography
  - Add appropriate informative references
- Clause 4, Annex 4A –Media access control
  - Mostly speed independent, update Table 4-2 MAC parameters
- Clause 30, Annex 30A & 30B –Management
  - Need presentation - Add new objects, attributes, and enumerations for 40Gb/s and 100Gb/s functions

Existing clauses (cont'd)

- Annex 31B –MAC control PAUSE operation
  - Need presentation - Update timing considerations for PAUSE
- Clause 45 Management data input/output (MDIO) interface.
  - Add new registers for the control and management of 40Gb/s and 100Gb/s PHY types
  - Add new MMDs if any, control/status of PMA/PMD and PCS
  - Update Backplane Auto-Negotiation and FEC registers
  - Presentations to other clauses to include the required management variables

Existing clauses (cont'd)

- Annex 69A –Interference tolerance testing
  - Need presentation - 40GbE test methodology
- Annex 69B –Interconnect characteristics
  - Need presentation - 40GbE cross-talk limits if needed
- Clause 72 –10GBASE-KR PMD
  - Changes if any due to 40GbE
- Clause 73 –Auto-Negotiation for Backplane Ethernet
  - Add technology ability bit for new 40GbE PHY
- Clause 74 –Forward error correction for 10GBASE-KR
  - Changes for 4 lane KR operation
- Clause 74A –FEC block coding examples
  - Additional patterns for 4 lanes if needed
- Need to select a proposal for 40Gb/s Backplane Ethernet

New Clauses

- Introduction to 40Gb/s and 100Gb/s operation
  - Based on presentations for other new clauses
  - Global PICs - separate PICS tables for 40 and 100Gb/s Sub-layers
  - Need to select an architecture proposal for baseline
- Reconciliation Sublayer and Media Independent Interface(s)
  - Need presentation to reference for baseline
- Physical Coding Sublayer clause(s)
  - Need to select a proposal for baseline
- PMA Sublayer clause(s)
  - Need presentation to reference for baseline(s)
- nAUI Electrical interface if included in adopted baseline proposals
  - Need presentation to reference for baseline
- FEC sublayer for optical PMDs if included in adopted baseline proposals
  - Need presentation to reference for baseline

New Clauses

- 40G Backplane PMD Sublayer
  - Need to select a proposal for baseline
- 40G / 100G Cu Cable PMD(s) Sublayer
  - Need to select a proposal for baseline
- 40G / 100G MMF PMD(s) Sublayer
  - Need to select a proposal for baseline
- 40G 10Km MMF PMD(s) Sublayer
  - Need presentation to reference for baseline
- 100G 10km SMF PMD(s) Sublayer
  - Need presentation to reference for baseline
- 100G 40km SMF PMD(s) Sublayer
  - Need presentation to reference for baseline
- Additional annexes to describe test methods, channel characteristics, coding details, etc.,
  - Need presentations to reference for baseline(s)

Proposed Nomenclature

- Nomenclature for the 3 part suffix
  - Speed
    - 40 = 40Gb/s, 100 = 100Gb/s
  - Medium type
    - Copper
      - K = Backplane
      - C = Cable assembly
    - Optical
      - S = Short Reach (100m)
      - L = Long Reach (10Km)
      - E = Extended Long Reach (40Km)
  - Coding scheme
    - R = 64B/66B block coding
  - Number of lanes or wavelengths
    - Copper: n = 4 or 10
    - Optical: n = number of lanes or wavelengths
    - n=1 not required as serial is implied

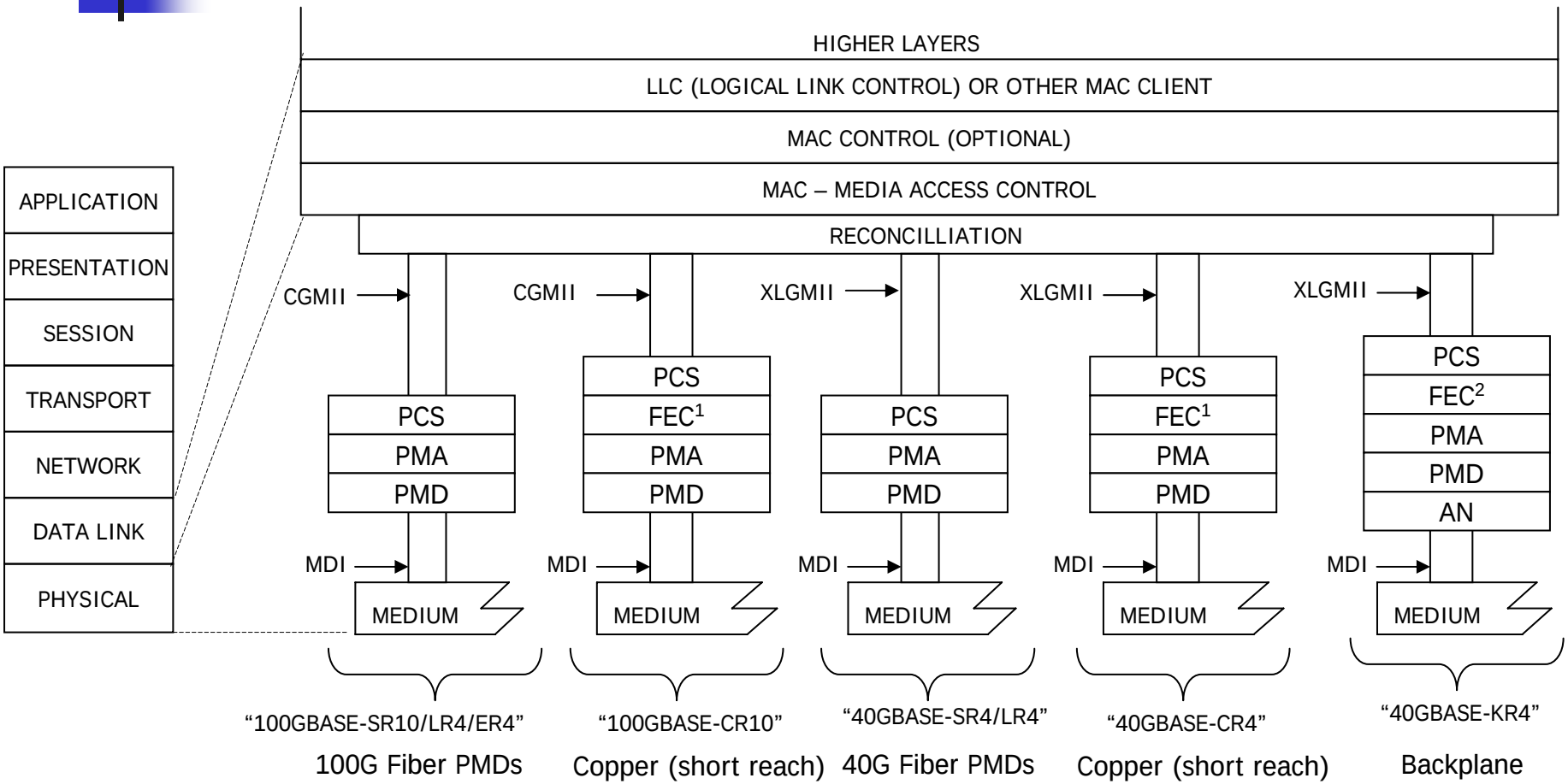
PHY description	Port Type
40G Backplane PHY	40GBASE-KR4
40G Cable Assembly PHY 100G Cable Assembly PHY	40GBASE-CR4 100GBASE-CR10
40G MMF 100m PHY (Ribbon) 100G MMF 100m PHY (Ribbon)	40GBASE-SR4 100GBASE-SR10
40G SMF 10Km PHY 100G SMF 10Km PHY	40GBASE-LR4 100GBASE-LR4
100G SMF 40Km PHY	100GBASE-ER4

40/100G Architecture and Interfaces proposal

Ilango Ganga, Intel
Brad Booth, AMCC
Howard Frazier, Broadcom
Shimon Muller, Sun
Gary Nicholl, Cisco

May 13, 2008

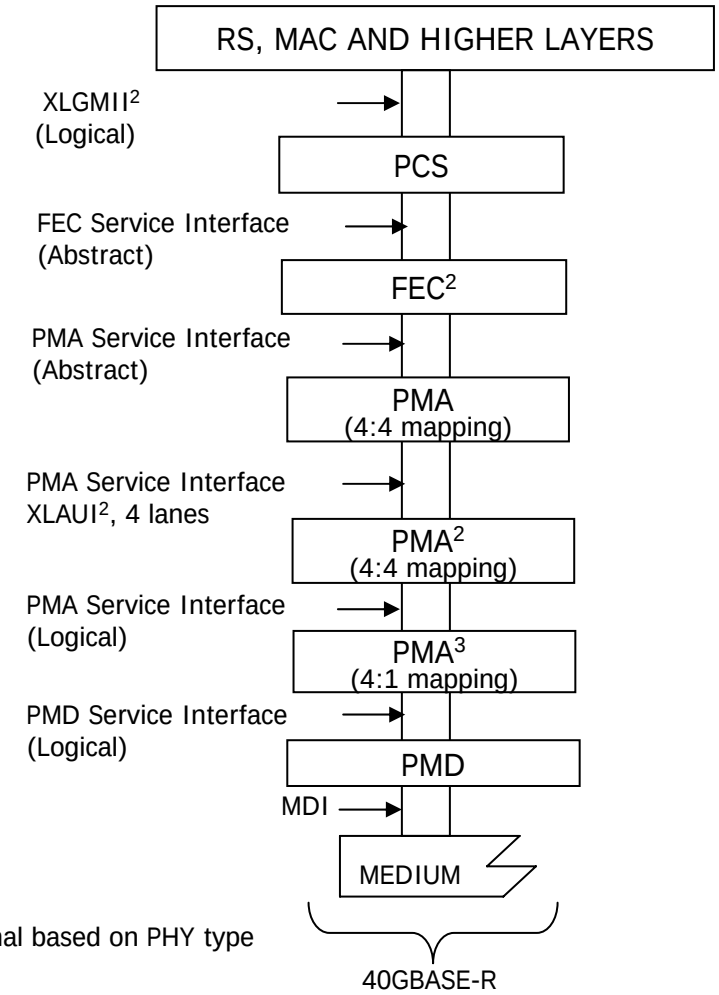
Proposed 40/100GbE layer model



Note: 1. CR4 & CR10 may use optional FEC
2. Optional

Proposed 40GbE architecture

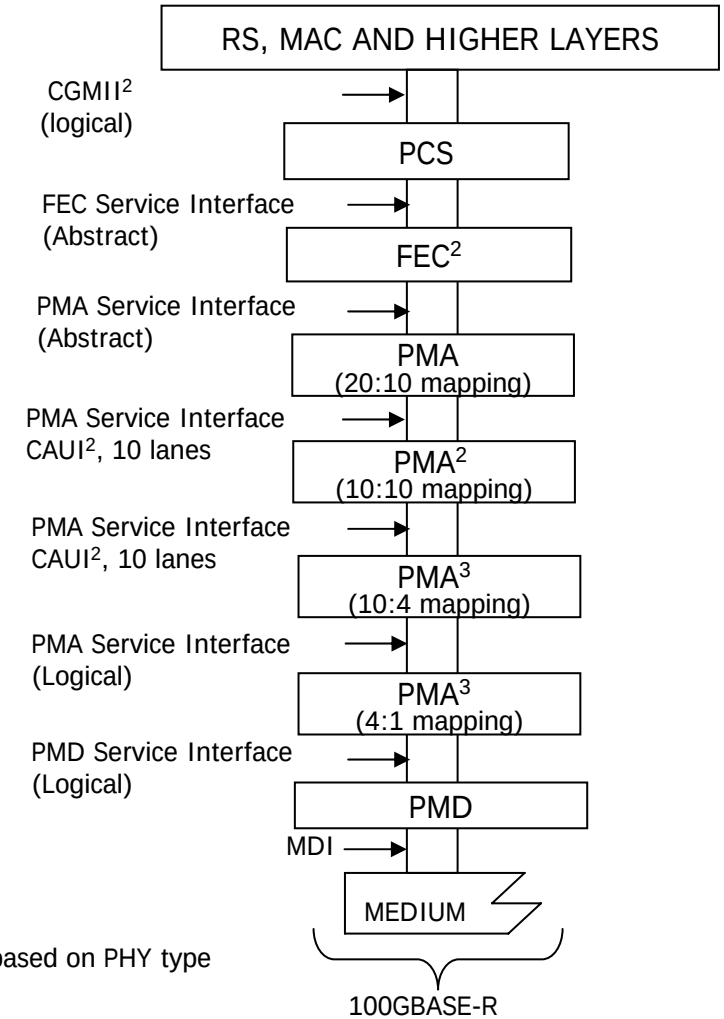
- XLGMII (intra-chip)
  - Logical, define data/control, clock, no electrical specification
- PCS
  - 64B/66B encoding
  - Lane distribution and alignment
- XLAUI (chip-to-chip)
  - 10.3125 GBaud electrical interface
  - 4 lanes, short reach
- FEC service interface
  - Abstract, can map to XLAUI electrical interface
- PMA Service interface
  - Logical n lanes, can map to XLAUI electrical interface
- PMD Service interface
  - Logical



Note: 2. Optional
3. Conditional based on PHY type

Proposed 100GbE architecture

- CGMII (intra-chip)
  - Logical, define data/control, clock, no electrical specification
- PCS
  - 64B/66B encoding
  - Lane distribution and alignment
- CAUI (chip-to-chip)
  - 10.3125 GBaud electrical interface
  - 10 lanes, short reach
- FEC service interface
  - Abstract, can map to CAUI electrical interface
- PMA Service interface
  - Logical n lanes, can map to CAUI electrical interface
- PMD Service interface
  - Logical



Note: 2. Optional
3. Conditional based on PHY type

Interface description (1)

- XLGMII (Forty Gigabit MII) or CGMII (100 Gigabit MII) – PCS interface
  - Interface between MAC and PHY layers needed for intra-chip connectivity
  - Need for Compatibility interface
    - Multiple vendors develop IP blocks for system on chip implementations
    - Provides a point of interoperability for multi vendor implementations
  - Logical definition, data width, control, clock frequency, no electrical specification
  - XLGMII and CGMII will have same logical behavior
  - Allows XLGMII/CGMII implementations with different data/control widths at either end of a link
  - See gustlin_02_0508 for further details on XL/CGMII

Interface description (2)

- XLAUI or CAUI interface (Chip-to-Chip)
  - 10.3125 GBaud electrical interface
    - Lane width: 4 lane for 40G, and 10 lane for 100G
  - Provides a point of interoperability for multi vendor implementations
    - Similar to XAUI, for 10GbE, which is widely used as MAC-PHY interface
  - Low pin count, low power interface, for example PHYs, Switches, LAN controllers
  - Common electrical definition for XLAUI/CAUI
    - 10.3125 GBaud differential signaling
    - Short reach channel: e.g. around 10 inches with 1 connector
  - Same electrical definition can be optionally used with multiple Service interfaces (e.g. PMA, FEC, etc.,)
  - This is not an MDI

Interface description (3)

- FEC Service interface
  - Interface between PCS and optional FEC sub-layer
    - Used for backplane PHYs, may be used with other PHY types (e.g. copper cable assy)
  - FEC Service interface is similar to PMA interface
  - Possible implementations: FEC integrated with MAC/PCS, or with PMA/PMD device
  - Abstract definition, with an option to map to XLAUI/CAUI electrical interface
- PMA Service interface
  - Interface between PMA and PCS
  - Logical definition with n Lanes, can map to XLAUI/CAUI electrical interface
- PMD Service interface
  - Interface between PMD and PMA
  - PMA and PMD may be implemented together in the same device
  - Logical definition

XL/CGMII and RS Proposal

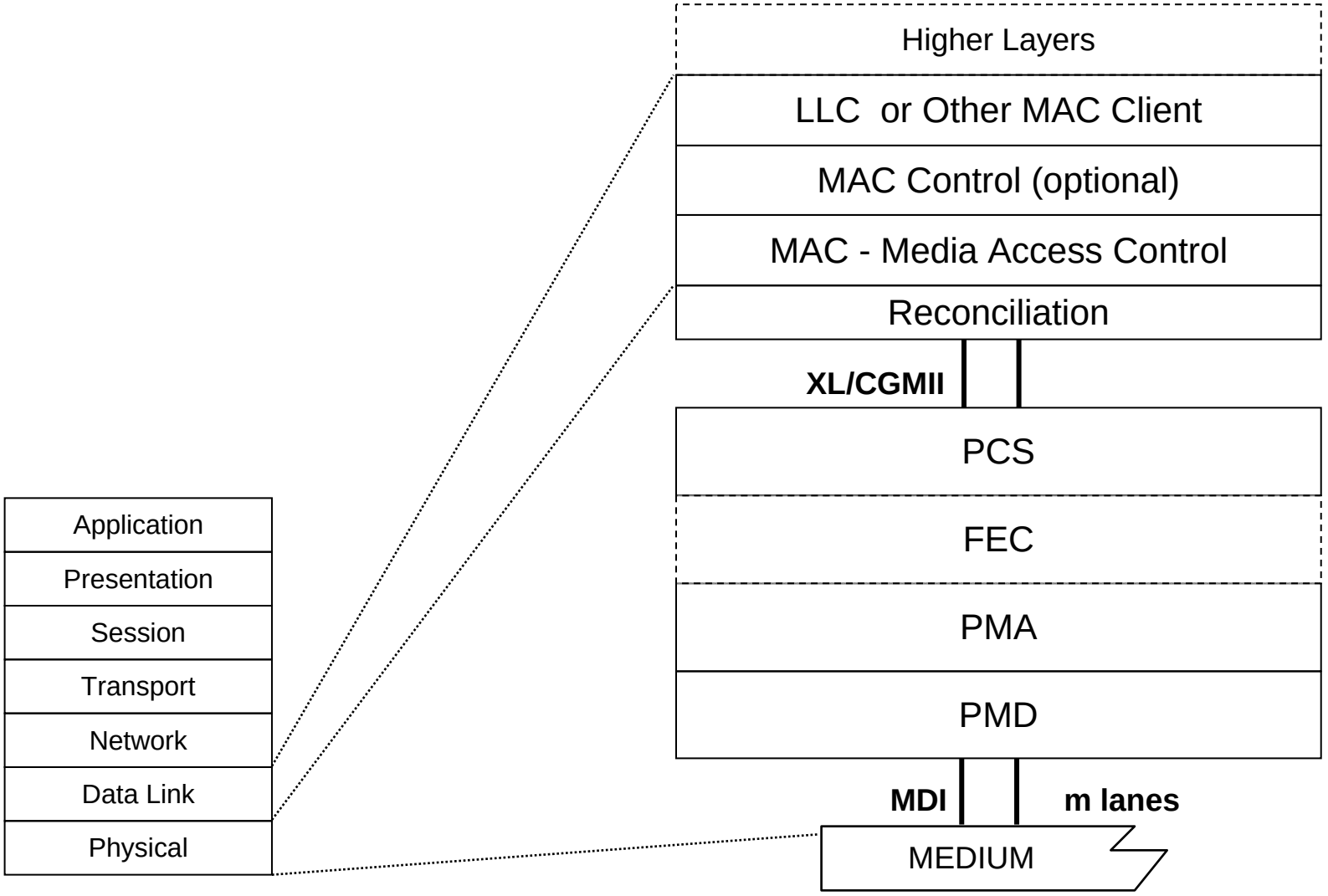
Mark Gustlin - Cisco

IEEE 802.3ba May 2008 Munich

Contributors and Supporters

- Steve Trowbridge - Alcatel-Lucent
- Brad Booth – AMCC
- Dimitrios Giannakopoulos - AMCC
- Piers Dawe – Avago
- Howard Frazier - Broadcom
- Arthur Marris - Cadence Design Systems
- Gary Nicholl – Cisco
- Med Belhadj – Cortina Systems
- Chris Cole - Finisar
- Subi Krishnamurthy - Force10 Networks
- Aris Wong – Foundry
- Shashi Patel - Foundry
- Ryan Latchman - Gennum
- Shinji Nishimura - Hitachi
- Hidehiro Toyoda – Hitachi
- John Jaeger - Infinera
- Andy Moorwood - Infinera
- Thananya Baldwin – Ixia
- Jerry Pepper – Ixia
- Faisal Dada - JDSU
- Norbert Folkens – JDSU
- Jack Jewell – JDSU
- Jeffery J. Maki - Juniper Networks
- David Ofelt – Juniper Networks
- Adam Healey - LSI
- Avigdor Segal – Marvell
- Martin White - Marvell
- Pete Anslow – Nortel
- Song Shang - SMI
- Shimon Muller – Sun
- Farhad Shafai – Sarance
- Andre Szczepanek – TI
- Frank Chang - Vitesse

40GE/100GE Architecture



XL/CGMII Interface

- Why define it?

  Electrically it won't see the light of day

  Some want it for RTL to RTL connections within devices

- The interface is naturally scaled based on speed targets of an implementation

  FPGAs run slower, ASICs faster, next generation ASICs even faster…

- Define it as a logical interface only
  - Service primitives (function calls, pseudo code) +
  - Signals, code-points, syntax, sequences, true/false

XL/CGMII Interface

- Leverage XGMII, but make it 8 lanes instead of 4
- Preserve use of encoded rather than discrete delimiters

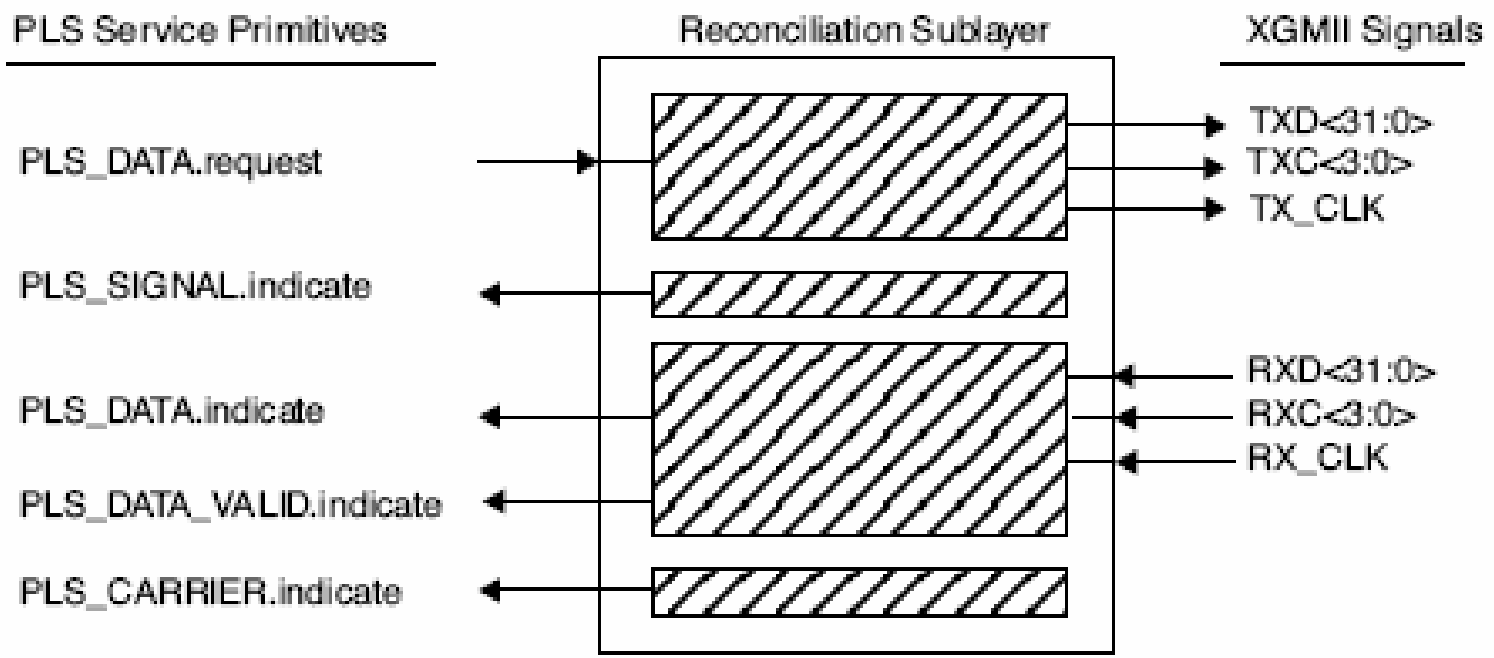
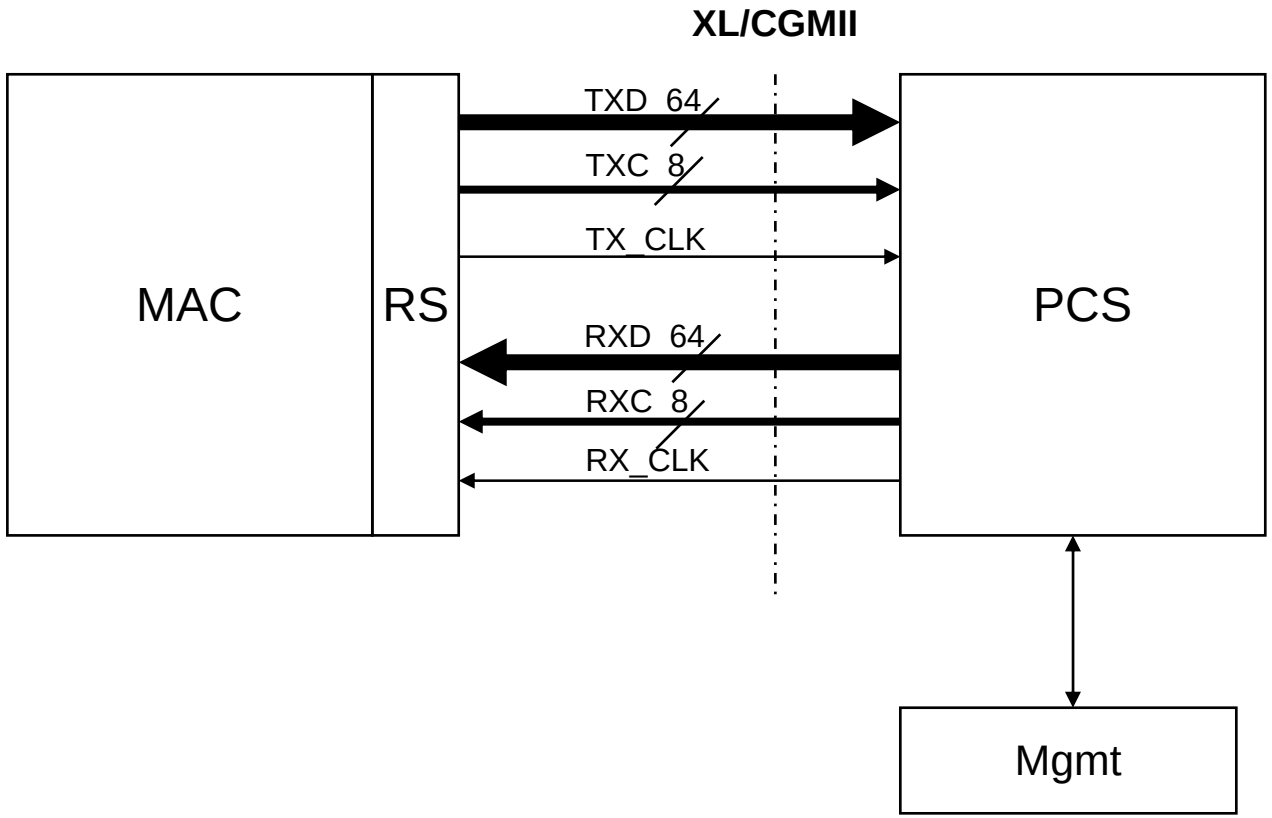


Figure 46–2—Reconciliation Sublayer (RS) inputs and outputs

From 802.3ae

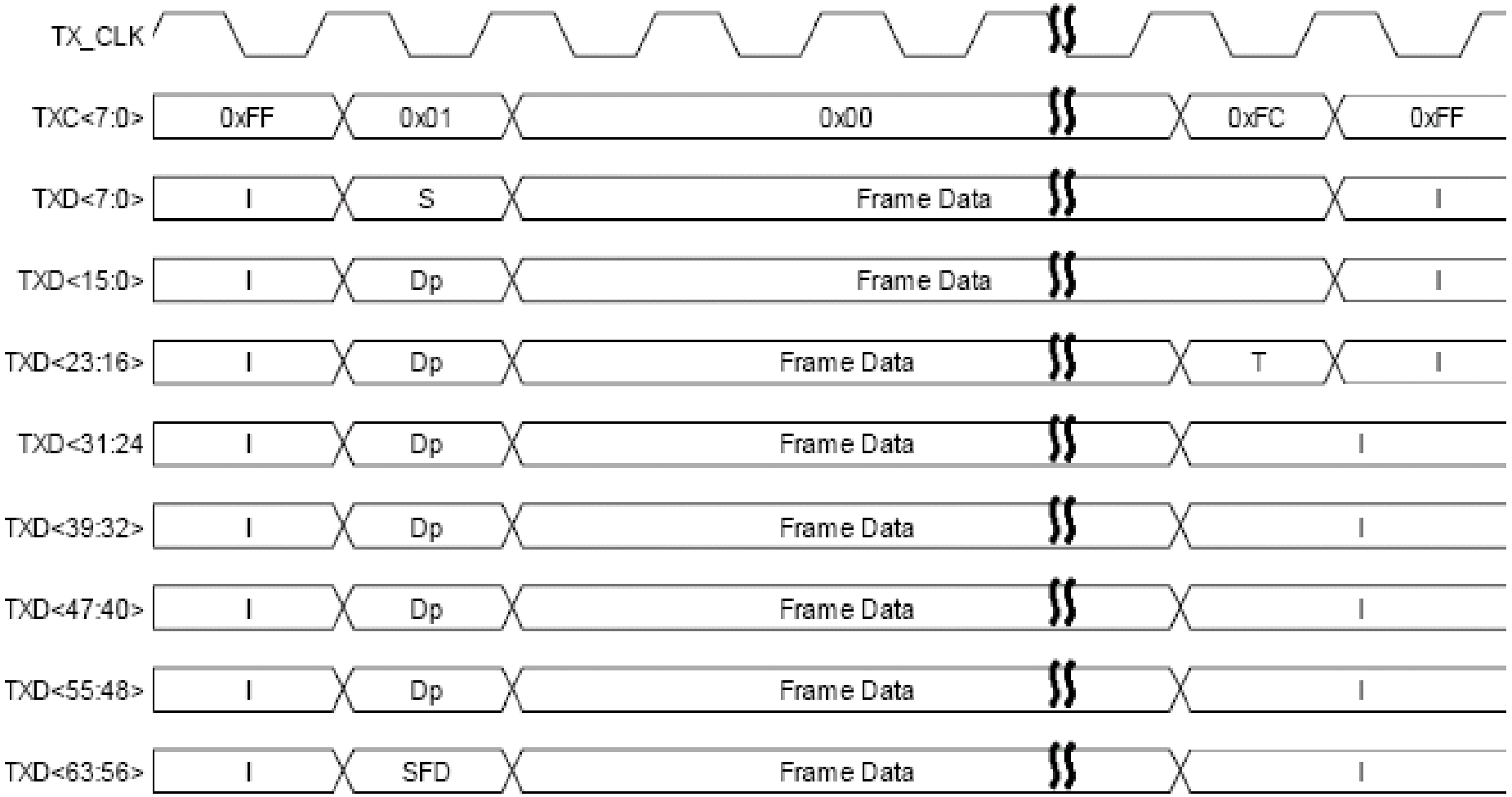
XL/CGMII Interface

- Leverage XGMII, but make it 8 lanes instead of 4
- CLK = 625MHz for 40GE, 1.5625GHz for 100GE
- Clock may be scaled down in frequency by increasing the width from 8 lanes to 16, 24, 32 etc.



XL/CGMII Interface

RX Diagram is Identical



I: Idle control character, S: Start control character, Dp: preamble Data octet, T: Terminate control character, SFD: Start of Frame Delimiter

XL/CGMII Interface

Same encoding as XGMII (for both tx and rx):

Table 46–3—Permissible encodings of TXC and TXD

TXC	TXD	Description	PLS_DATA.request parameter
0	00 through FF	Normal data transmission	ZERO, ONE (eight bits)
1	00 through 06	Reserved	—
1	07	Idle	No applicable parameter (Normal inter-frame)
1	08 through 9B	Reserved	—
1	9C	Sequence (only valid in lane 0)	No applicable parameter (Inter-frame status signal)
1	9D through FA	Reserved	—
1	FB	Start (only valid in lane 0)	No applicable parameter, replaces first eight ZERO, ONE of a frame (preamble octet)
1	FC	Reserved	—
1	FD	Terminate	DATA_COMPLETE
1	FE	Transmit error propagation	No applicable parameter
1	FF	Reserved	—

NOTE—Values in TXD column are in hexadecimal, most significant bit to least significant bit (i.e., <7:0>).

8B vs. 4B alignment

- We could keep the legacy 4B alignment even with the new 8B wide bus
- Or we could go to 8B alignment
  - Only start packets in lane 0
  - Significant gate savings for 100GE, especially in FPGAs
  - Deficit counter goes from 0-7 for 8B alignment (vs. 0-3 for 4)
  - Doubles the buffering required for clock compensation when compared to 4B alignment
- Recommended to go with 8B alignment
- If interface is to be scaled down in frequency (and up in width), packet starts are still on 8B boundaries (lane 0, 8, 16 etc).

IPG Rules for 8B Alignment

- A MAC implementation may be designed to always insert additional idle characters to align the start of preamble on an eight byte boundary.

  Note that this will reduce the effective data rate for certain packet sizes separated with minimum inter-frame spacing.

- Alternatively, the RS may maintain the effective data rate by sometimes inserting and sometimes deleting idle characters to align the Start control character.

  When using this method the RS must maintain a Deficit Idle Count that represents the cumulative count of idle characters deleted or inserted. The counter is incremented for each idle character deleted, decremented for each idle character inserted, and the decision of whether to insert or delete idle characters is constrained by bounding the counter to a minimum value of zero and maximum value of seven.

Summary

- Simple logical interface based on XGMII
- Extended to 8 Bytes
- Naturally scales up and down in width and frequency
- Packet Starts on 8 Byte boundaries

100GE and 40GE PCS (MLD) Proposal

IEEE 802.3ba May 2008 Munich

Contributors and Supporters

David Law – 3com
Steve Trowbridge - Alcatel-Lucent
Jesse Simsarian - Alcatel-Lucent
Brad Booth – AMCC
Dimitrios Giannakopoulos – AMCC
Francesco Caggioni – AMCC
Keith Conroy – AMCC
Piers Dawe – Avago
Rita Horner – Avago
Howard Frazier - Broadcom
Arthur Marris – Cadence
Mike Shahine - Ciena
Mark Nowell - Cisco
Gary Nicholl - Cisco
Hugh Barrass - Cisco
Steve Swanson - Corning
Med Belhadj – Cortina
Chris Cole - Finisar
Krishnamurthy Subramanian – Force10

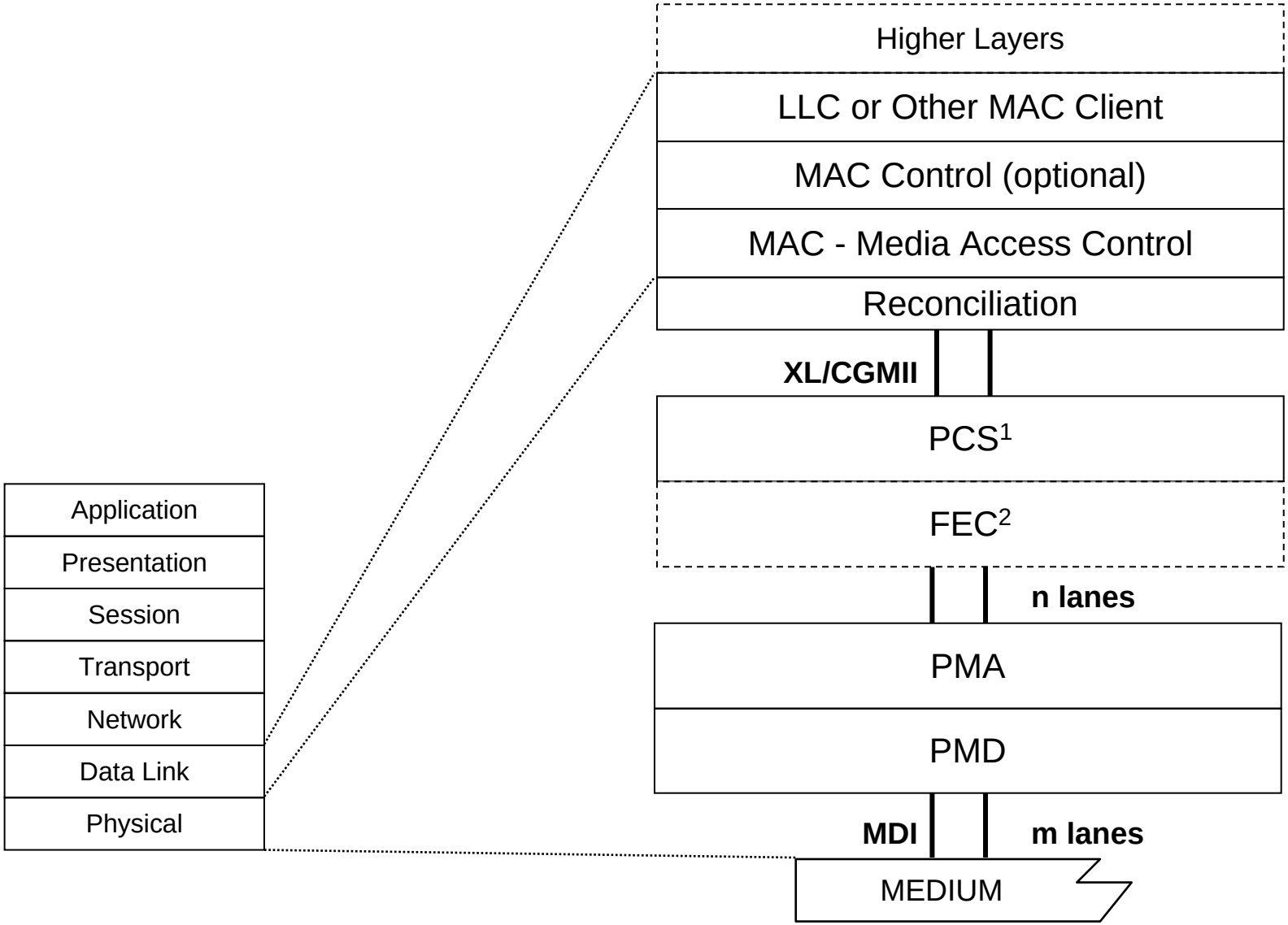
Aris Wong – Foundry Networks
Shashi Patel – Foundry Networks
Bill Ryan – Foundry Networks
Ryan Latchman - Gennum
Justin Abbott - Gennum
Hong Liu – Google
Ashby Armistead – Google
Shinji Nishimura - Hitachi Ltd
Hidehiro Toyoda - Hitachi Ltd
Dan Dove – HP
Petar Pepeljugoski – IBM
John Jaeger - Infinera
Andy Moorwood - Infinera
Drew Perkins - Infinera
Jerry Pepper - Ixia
Thananya Baldwin - Ixia
Faisal Dada - JDSU
Jack Jewell - JDSU
Mike Dudek - JDSU

Jeffery J. Maki - Juniper Networks
David Ofelt - Juniper Networks
Brad Turner - Juniper Networks
Adam Healey - LSI
Martin White – Marvell
Andy Weitzner – Marvell
Pete Anslow – Nortel
David W. Martin – Nortel
Osamu Ishida - NTT
Shoukei Kobayashi - NTT
Matt Traverso – Opnext
Farhad Shafai - Sarance Technologies
Farzin Firoozmand – SMI
Craig Hornbuckle – SMI
Song Shang - SMI
Ted Seely - Sprint
Kengo Matsumoto - Sumitomo Electric
Shimon Muller - Sun
Andre Szczepanek – TI
Martin Carroll - Verizon
Frank Chang - Vitesse

Agenda

- 40GE/100GE Architecture
- PCS and MLD layer details
- Possible XL/CGMII Interface
- Alignment details
- Alignment performance metrics
- Clocking example
- Skew
- Summary

40GE/100GE Generic Architecture



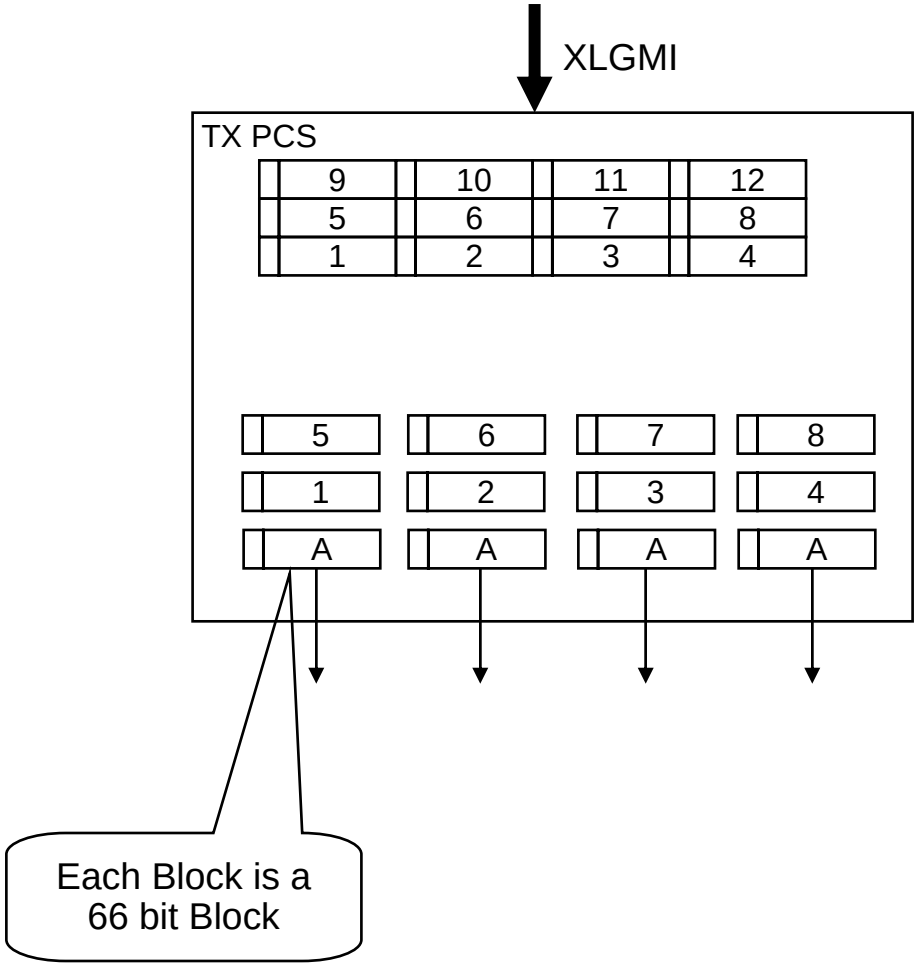
1: Includes MLD functionality 2: For 40GE Backplane

Proposed 100GE/40GE PCS

- 10GBASE-R 64B/66B based PCS
  - Run at 100Gbps or 40Gbps serial rate
  - Includes 66 bit block encoding and scrambling
- Multi-Lane Distribution
  - Data is distributed across n virtual lanes 66 bit blocks at a time
  - Round robin distribution
  - Periodic alignment blocks are added to each virtual lane to allow deskew in the rx PCS
- PMA maps n lanes to m lanes
  - PMA is simple bit level muxing
  - Does not know or care about PCS coding
- Alignment and static skew compensation is done in the Rx PCS only

Striping Mechanism

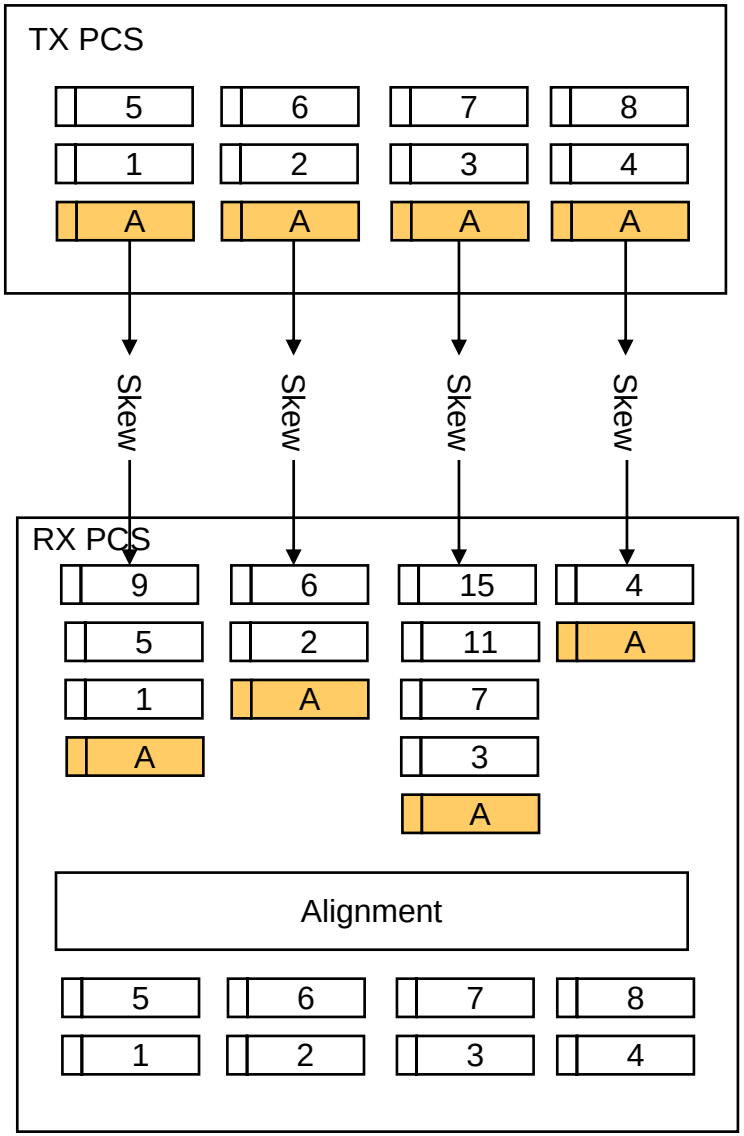
This example is 40GE with 4 electrical and 4 optical lanes



PCS Functions:

- 66 bit encoding
- Scrambling
- Periodic alignment block addition
- Round robin block distribution

Alignment Mechanism – 40GE Example



RX PCS Functions:

- Re-Align 66 bit blocks
- Remove the Alignment blocks
- Then descramble and decode

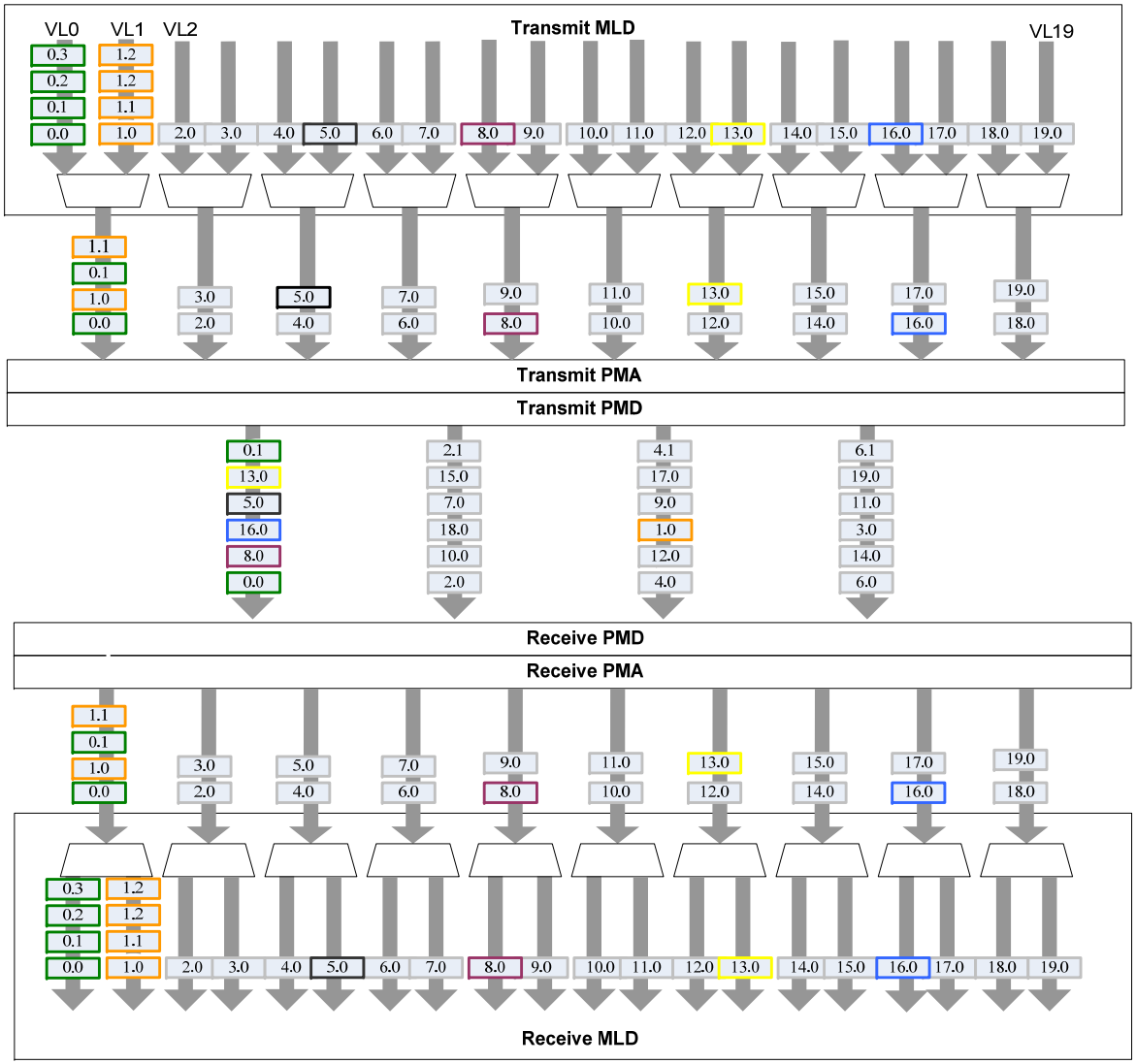
Key Concept – Virtual Lanes

- Virtual lanes may or may not correspond to physical lanes
- Virtual lanes are created by distributing PCS encoded data in a round robin fashion, on a 66 bit block basis
- The number of virtual lanes generated is scaled to the Least Common Multiple (LCM) of the n lane electrical interface and the m lane PMD

  This allows all data (bits) from one virtual lane to be transmitted over the same electrical and optical lane combination

  This ensures that the data from a virtual lane is always received with the correct bit order at the Rx MLD
- The alignment markers allow the Rx PCS to perform skew compensation, realign all the virtual lanes, and reassemble a single 100G or 40G aggregate stream (with all the 64B/66B blocks in the correct order)

Bit Flow Through – 100GE 4 lane PMD



- 20 VLs
- 10 Electrical lanes
- 4 Optical lanes
- With Skew, VLs move around
- RX MLD puts things back in order

How Many Virtual Lanes are Needed?

- **4 VLs For 40GE, this covers all of the possible combinations of lanes:**

Electrical Lane Widths	PMD Lane Widths	Virtual Lanes Needed
4, 2, 1	4, 2, 1	4

- **20 VLs For 100GE, this covers all of the possible combinations of lanes:**

Electrical Lane Widths	PMD Lane Widths	Virtual Lanes Needed
10, 5, 4, 2, 1	10, 5, 4, 2, 1	20

PCS Encoding

- Same 10GBASE-R PCS (Clause 49) encoding

Input Data	Sync	Block Payload							
Bit Position: / Data Block Format:	0 1	2							65
$D_0 D_1 D_2 D_3/D_4 D_5 D_6 D_7$	01	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7
Control Block Formats:		Block Type Field							
$C_0 C_1 C_2 C_3/C_4 C_5 C_6 C_7$	10	0x1e	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7
~~$C_0 C_1 C_2 C_3/C_4 D_5 D_6 D_7$~~	~~10~~	~~0x2d~~	~~C_0~~	~~C_1~~	~~C_2~~	~~C_3~~	~~C_4~~	~~D_5~~	~~D_6~~	~~D_7~~
~~$C_0 C_1 C_2 C_3/D_4 D_5 D_6 D_7$~~	~~10~~	~~0x33~~	~~C_0~~	~~C_1~~	~~C_2~~	~~C_3~~		~~D_5~~	~~D_6~~	~~D_7~~
~~$O_0 D_1 D_2 D_3/D_4 D_5 D_6 D_7$~~	~~10~~	~~0x66~~	~~D_1~~	~~D_2~~	~~D_3~~	~~O_0~~		~~D_5~~	~~D_6~~	~~D_7~~
~~$O_0 D_1 D_2 D_3/O_4 D_5 D_6 D_7$~~	~~10~~	~~0x55~~	~~D_1~~	~~D_2~~	~~D_3~~	~~O_0~~	~~O_4~~	~~D_5~~	~~D_6~~	~~D_7~~
$S_0 D_1 D_2 D_3/D_4 D_5 D_6 D_7$	10	0x78	D_1	D_2	D_3	D_4		D_5	D_6	D_7
$O_0 D_1 D_2 D_3/C_4 C_5 C_6 C_7$	10	0x4b	D_1	D_2	D_3	O_0	C_4	C_5	C_6	C_7
$T_0 C_1 C_2 C_3/C_4 C_5 C_6 C_7$	10	0x87		C_1	C_2	C_3	C_4	C_5	C_6	C_7
$D_0 T_1 C_2 C_3/C_4 C_5 C_6 C_7$	10	0x99	D_0		C_2	C_3	C_4	C_5	C_6	C_7
$D_0 D_1 T_2 C_3/C_4 C_5 C_6 C_7$	10	0xaa	D_0	D_1		C_3	C_4	C_5	C_6	C_7
$D_0 D_1 D_2 T_3/C_4 C_5 C_6 C_7$	10	0xb4	D_0	D_1	D_2		C_4	C_5	C_6	C_7
$D_0 D_1 D_2 D_3/T_4 C_5 C_6 C_7$	10	0xcc	D_0	D_1	D_2	D_3		C_5	C_6	C_7
$D_0 D_1 D_2 D_3/D_4 T_5 C_6 C_7$	10	0xd2	D_0	D_1	D_2	D_3	D_4		C_6	C_7
$D_0 D_1 D_2 D_3/D_4 D_5 T_6 C_7$	10	0xe1	D_0	D_1	D_2	D_3	D_4	D_5		C_7
$D_0 D_1 D_2 D_3/D_4 D_5 D_6 T_7$	10	0xff	D_0	D_1	D_2	D_3	D_4	D_5	D_6	

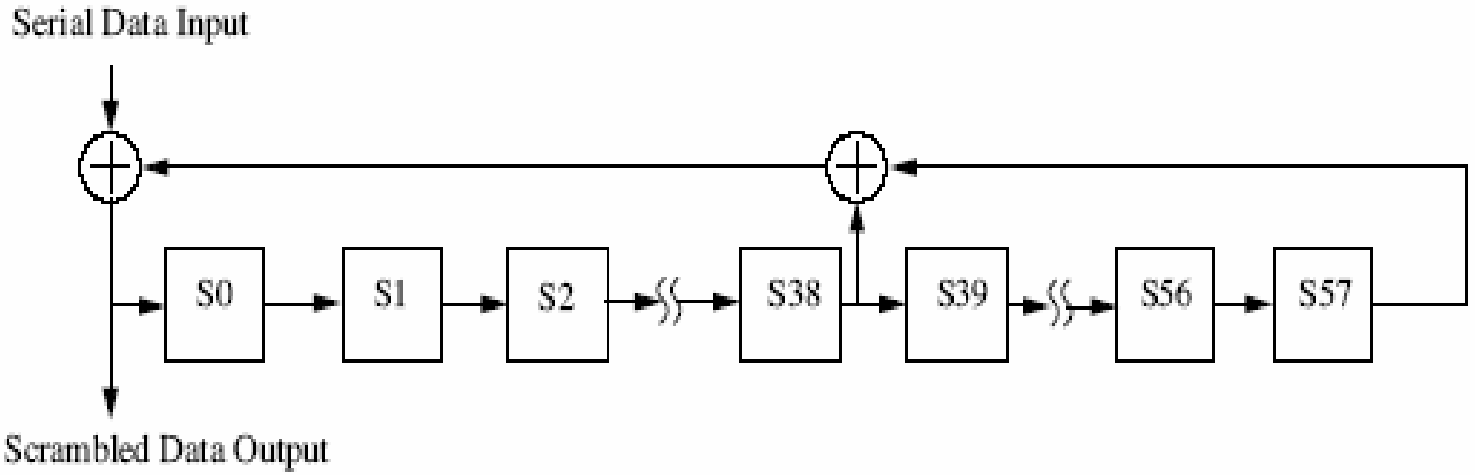
Not used since we have 8B alignment

Only block type used for ordered sets

PCS Scrambling

- Identical 10GBASE-R PCS (Clause 49) scrambler

  Runs at 40Gbps or 100Gbps now



PCS Idle Deletion/Insertion rules

- Straight from 802.3ae (except for highlighted text):

  Idle insertion or deletion occurs in groups of eight Idle characters

  Idle characters are added following idle or ordered_sets

  Idle characters are not added while data is being received

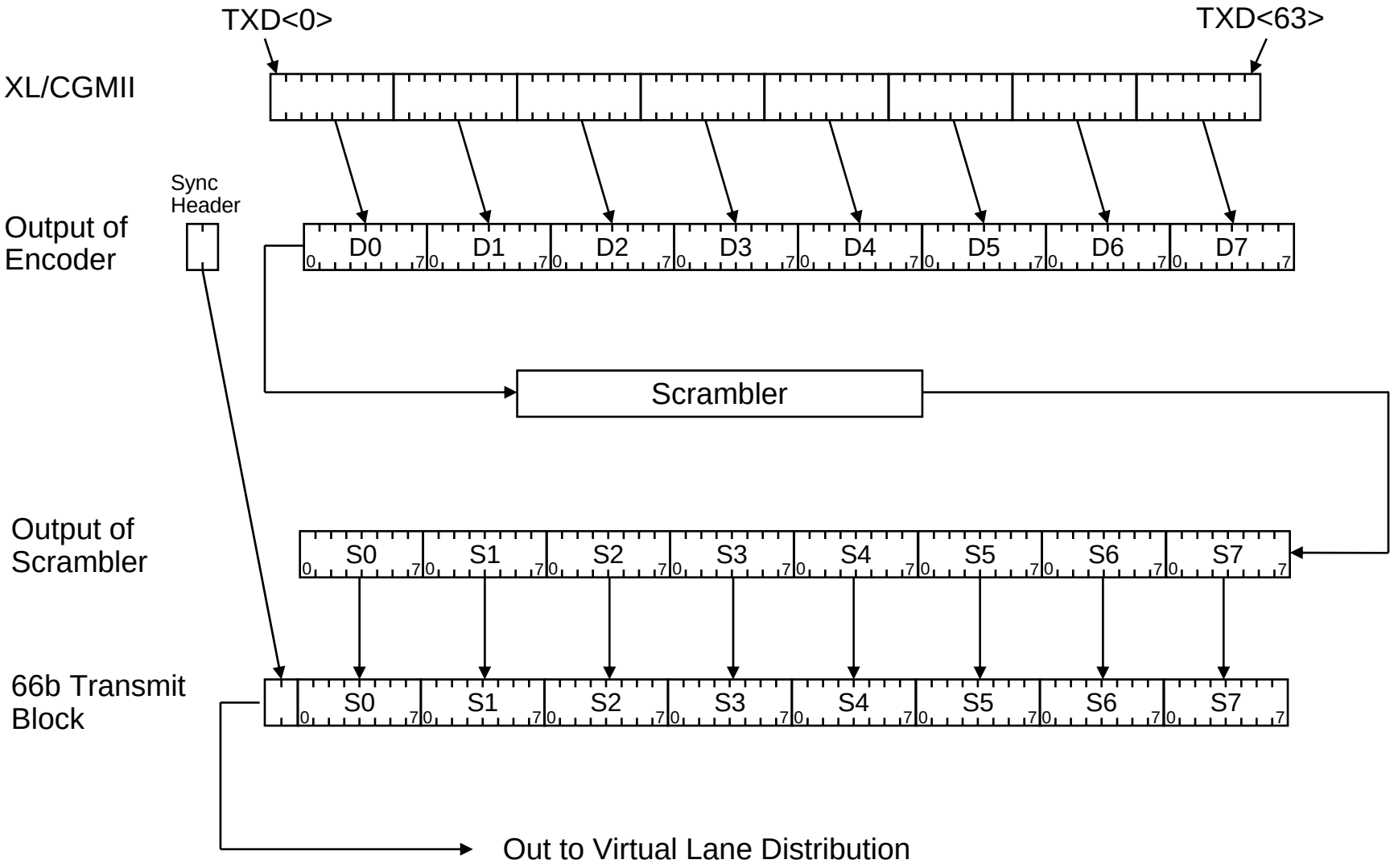
  When deleting idles, the minimum IPG of one character is maintained

  Sequence ordered_sets are deleted to adapt between clock rates

  Sequence ordered_set deletion occurs only when two consecutive sequence ordered_sets have been received and deletes only one of the two

  Only idles are inserted for clock compensation

PCS Bit Order



Alignment Proposal

- Send alignment on a fixed time basis
- Alignment word also identifies virtual lanes
- Sent every 16384 66bit blocks on each virtual lane at the same time
  - ~216usec for 20 VLs @ 100G
  - ~108usec for 4 VLs @ 40G
- It temporarily interrupts packets
- Takes only 0.006% (60PPM) of the Bandwidth
- Rate Adjust FIFO will delete enough IPG so that the MAC still runs at 100.000G or 40.000G with the interface running at 10.3125G

Alignment Word Proposal

Requirements:

- Significant transitions and DC balanced – word is not scrambled
- Keep in 66 bit form, but no relation to 10GBASE-R is needed
- But why not keep it close? – Because of the clock wander concerns
- Contains Virtual Lane Identifier

Proposed Alignment Word

10	VL	~VL

- This is DC balanced
- No relationship to the normal 10GBASE-R blocks
- Added after and removed before 64/66 processing
- Alignment block is periodic, no Hamming distance concerns with 64/66 block types

Alignment Word Proposal – 100GE

The encoding of the VL markers is as follows (based on $x^{58} + x^{39} + 1$ scrambler output):

VL Number	32 Bit encoding	VL Number	32 Bit encoding
0	C1,68,21,F4	10	FD, 6C, 99, DE
1	9D, 71, 8E, 17	11	B9, 91, 55, B8
2	59, 4B, E8, B0	12	5C, B9, B2, CD
3	4D, 95, 7B, 10	13	1A, F8, BD, AB
4	F5, 07, 09, 0B	14	83, C7, CA, B5
5	DD, 14, C2, 50	15	35, 36, CD, EB
6	9A, 4A, 26, 15	16	C4, 31, 4C, 30
7	7B, 45, 66, FA	17	AD, D6, B7, 35
8	A0, 24, 76, DF	18	5F, 66, 2A, 6F
9	68, C9, FB, 38	19	C0, F0, E5, E9

Note that data is played out in VL order, 0, 1, 2, …19, 0, 1…

Alignment Word Proposal – 40GE

The encoding of the VL markers is as follows (based on $x^{58} + x^{39} + 1$ scrambler output):

VL Number	32 Bit encoding
0	C1,68,21,F4
1	9D, 71, 8E, 17
2	59, 4B, E8, B0
3	4D, 95, 7B, 10

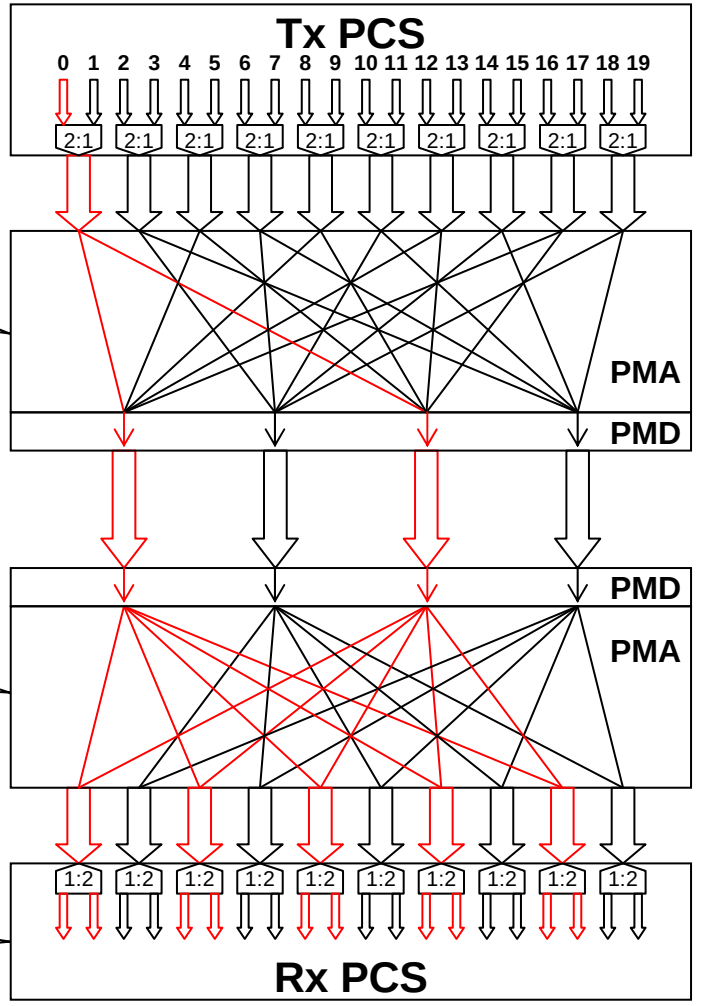
Note that data is played out in VL order, 0, 1, 2, 3, 0…

Possible Paths Through the Link

Skew on input electrical interface determines which optical lane VL 0 passes through

Skew on optical interface and previous electrical interface determines which output electrical lane VL 0 passes through

VL 0 can appear on any of the red outputs depending on the skew of the electrical and optical interfaces



Note: These possible paths are based on a 10:4 and 4:10 function based on round-robin distribution. Other arrangements which give different paths are possible.

Virtual Lane Location on the Receive Side

- Due to how virtual lanes are multiplexed, and due to skew, and in order to be future proof:
  - All receivers must support receiving a transmitted virtual lane on any received virtual lane
  - This is true for 100GE and 40GE

Finding VL Alignment

- After reception in the rx MLD, you have x VLs, each skewed and transposed
- First you find 66bit alignment on each VL

  Each VL is a stream of 66 bit blocks

  Same mechanism as 10GBASE-R (64 valid 2 bit frame codes in a row)
- Then you hunt for alignment on each VL

  Look for one of the 20 VL patterns repeated and inverted
- Alignment is declared on each VL after finding 2 consecutive non-errored alignment patterns in the expected locations (16k words apart)
- Out of alignment is declared on a VL after finding 4 consecutive errored frame patterns
- Once the alignment pattern is found on all VLs, then the VLs can be aligned

Alignment Performance Parameters – 100GE

- Mean Time To Alignment (MTTA)

  Mean time it takes to gain Alignment on a lane or virtual lane for a given BER

  Nominal time = 314usec

- Mean Time To Loss of Alignment (MTTLA)

  Mean time it takes to lose Alignment on a lane or virtual lane for a given BER

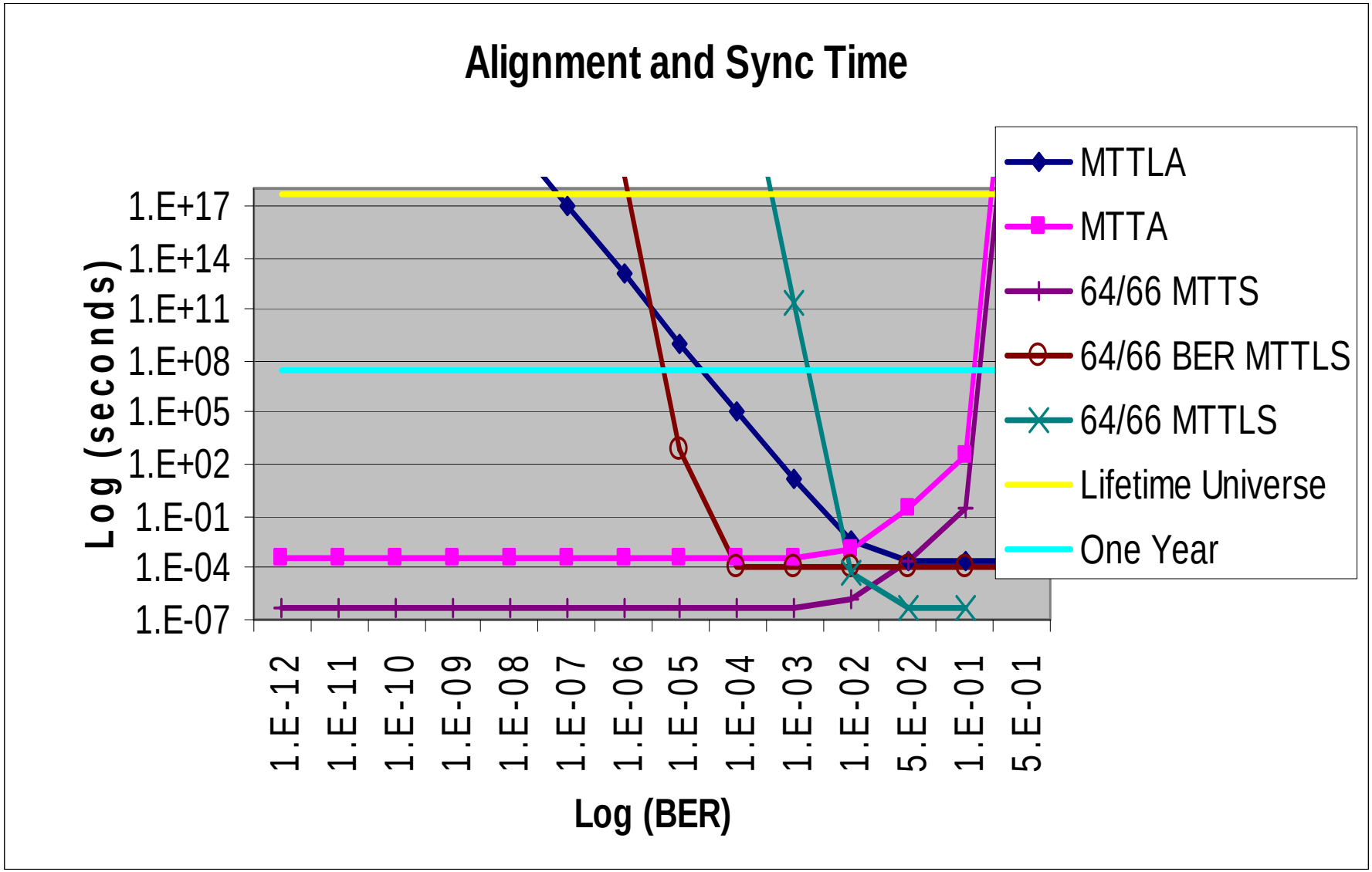
- Probability of False Alignment (PFA) = 3 E-40
- Probability of Rejecting False Alignment (PRFA) = ~1
- Also have 64/66 sync stats on the graph for comparison

  MTTS – Mean Time To Sync (64 non errored syncs in a row)

  BER MTTLS – With the 125usec BER window, what is the Mean Time To Lose Sync

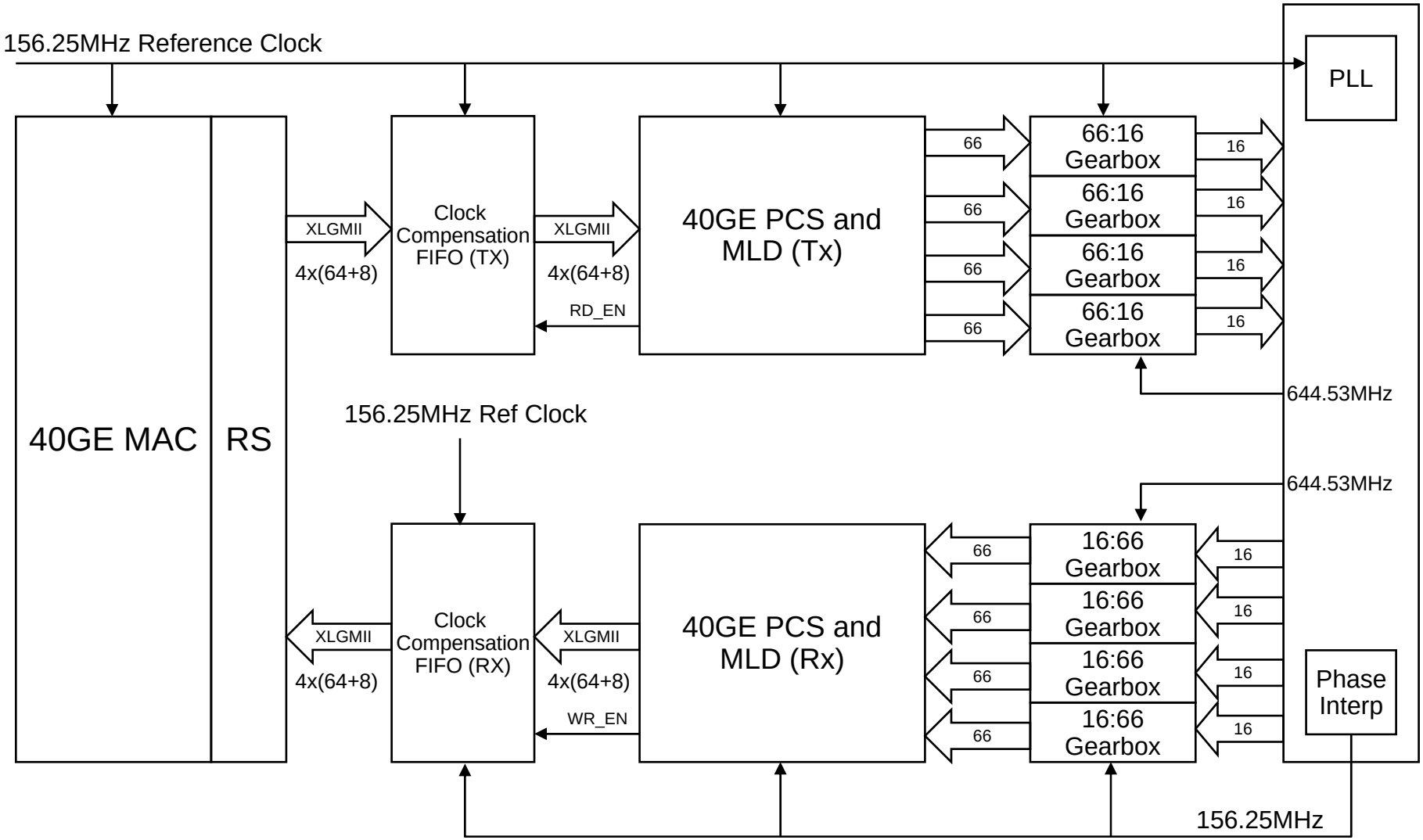
  MTTLS - Mean Time To Lose Sync

Alignment Performance Parameters – 100GE



40GE Alignment Performance will be similar

Clocking Example – 40GE



Skew Handling

- Both dynamic and static skew budgets need to be identified
- See other presentations for details

Summary

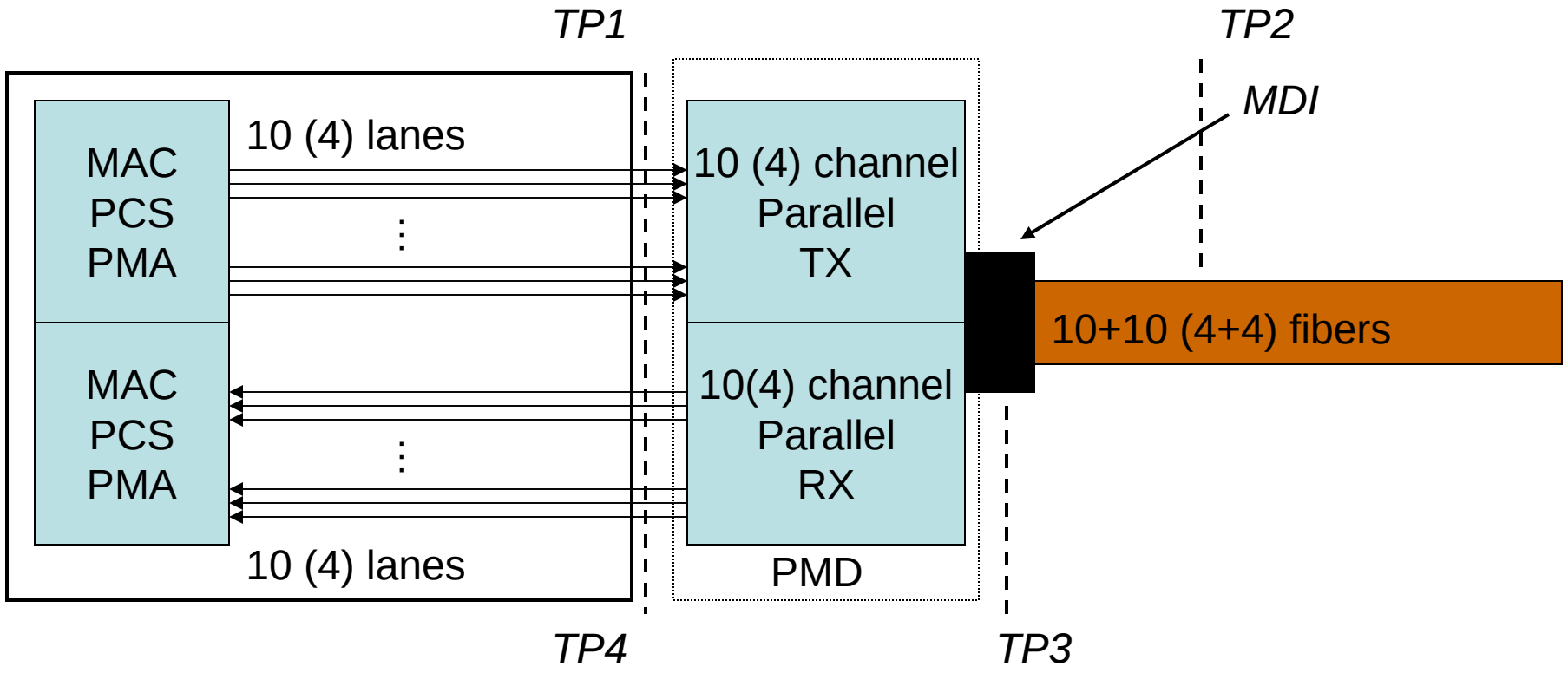
- Simple 10GBASE-R based PCS
- MLD layer to support multiple physical lanes/lambdas
- Complexity is low within the MLD layer
  - Simple block data striping
- Complexity in the optical module is low
  - Simple bit muxing even when m != n
- Based on proven 64B/66B framing and scrambling
- Electrical interface is feasible at 10x10G or 4x10G
- Allows for a MAC rate of 100.000G or 40.000G
  - Overhead very low and independent of packet size
- Supports an evolution of optics and electrical interfaces

Proposal for a PMD for 100GBASE-SR10 and 40GBASE-SR4 and Related Specifications

Petar Pepeljugoski - IBM
Piers Dawe, John Petrilla - Avago Technologies
John Dallesasse, Kenneth Jackson - Emcore
Lew Aronson, Jonathan King, Chris Cole - Finisar
Mike Dudek, Jack Jewell – JDSU
Phil McClay - Zarlink

Proposal

- 10 parallel lanes @ 10.3125 GBd for 100GBASE-SR10 over OM3 fiber
- 4 parallel lanes @ 10.3125 GBd for 40GBASE-SR4 over OM3 fiber
- No glue chip required
  - See also last slide



Transmitter specifications (each lane)

Description	Value	Unit
Signaling speed (nominal)	10.3125	GBd
Signaling speed variation from nominal (max)	±100	ppm
Center wavelength (range)	840-860	nm
RMS spectral width (max)	0.65	nm
Average Launch Power (max) *(2)*	1 *(1)*	dBm
Launch Power (min) in OMA	-3 *(1), (3)*	dBm
Average launch power of OFF transmitter (max)	-30	dBm
Extinction ratio (min)	3	dB
$RIN_{12}OMA$ (max)	-128 to -132 *(1),(3)*	dB/Hz
Optical Return Loss Tolerance (max)	12	dB
Encircled Flux	> 86% @ 19um, < 30% at 4.5um *(1)*	
Transmitter eye mask definition	TBD	
Aggregate TP2 signal metrics (4)	TBD	TBD
TP1 jitter allocation	0.3 (5)	U.I.

(1) subject to further study
(2) see presentation on eye safety by J. Petrilla at March 2008 meeting
(3) to be made informative if aggregate signal parameter includes the effect
(4) for further study, e.g. TDP, TWDP, etc.
(5) *for further study, intermediate between 10G SFP+ and 8GFC*

Receiver characteristic (each lane)

Description		
Signaling speed (nominal)	10.3125	GBd
Signaling speed variation from nominal (max)	± 100	ppm
Center wavelength (range)	840-860	nm
Average receiver power (max)	1[(1)]	dBm
Average power at receiver input (min)	-7.9[(1),(2)]	dBm
Receiver reflectance (max)	-12	dB
Stressed receiver sensitivity in OMA (max)	TBD	dBm
- Vertical eye closure penalty (target)	TBD	dB
- Stressed eye jitter (target)	TBD	UI pk-pk
TP4 jitter allocation	0.7	UI

(1) For further study

(2) Depends on connector loss

Link and Cable Characteristic

Parameter	Value	Unit
Supported fiber types	50μm OM3	
Effective Modal Bandwidth	2000*	MHz*km
Power Budget	>8.3**	dB
Operating Range	0.5-100	m
Channel insertion loss	1.9***	dB

** - depends on launch conditions*
*** - for further study*
**** - connector loss under study*

Appropriate Support for OTN Baseline Proposal

Stephen J. Trowbridge
Alcatel-Lucent

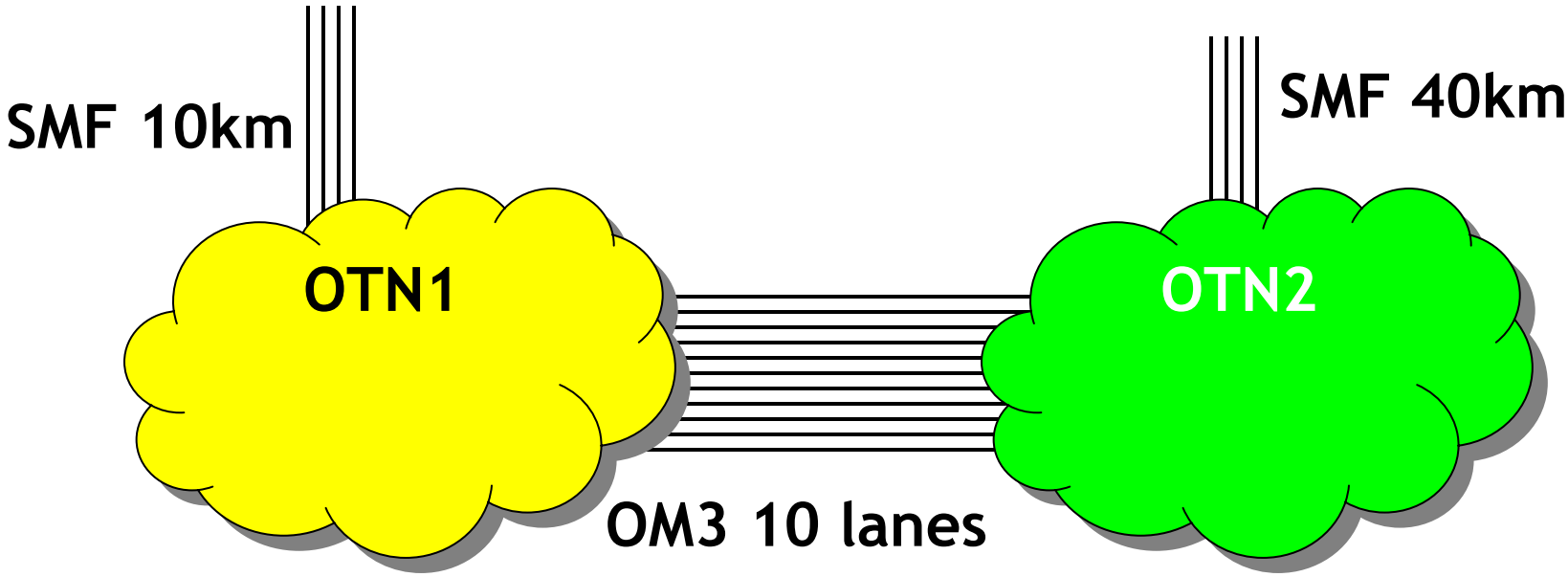
Supporters

- Thomas Fischer – Nokia-Siemens Networks
- Pete Anslow – Nortel Networks
- Ralf-Peter Braun – Deutsche Telekom
- Martin Carroll – Verizon
- Ghani Abbas – Ericsson
- Arne Alping – Ericsson
- Chris Cole – Finisar
- Mark Gustlin – Cisco
- Osamu Ishida – NTT
- George Young – AT&T
- Gary Nicholl - Cisco

Key elements of OTN support

- Use a Lane Independent PCS to enable different Ethernet PMDs to be used at the OTN ingress/egress
  - Key feature of MLD
- 40 GbE must fit into the OPU3 payload with a minimum of PCS codeword and timing transparency
  - Limitation on control block types to permit transcoding
- Lane Marker transparency for 40 GbE
  - ITU-T decision, but maintain spare value in 4-bit representation of control block types available for encoding lane markers if necessary
- Link fault signaling for 802.3ba Ethernet over OTN can use existing mechanisms from 802.3ae

Independence of Ethernet PMDs in OTN mapping

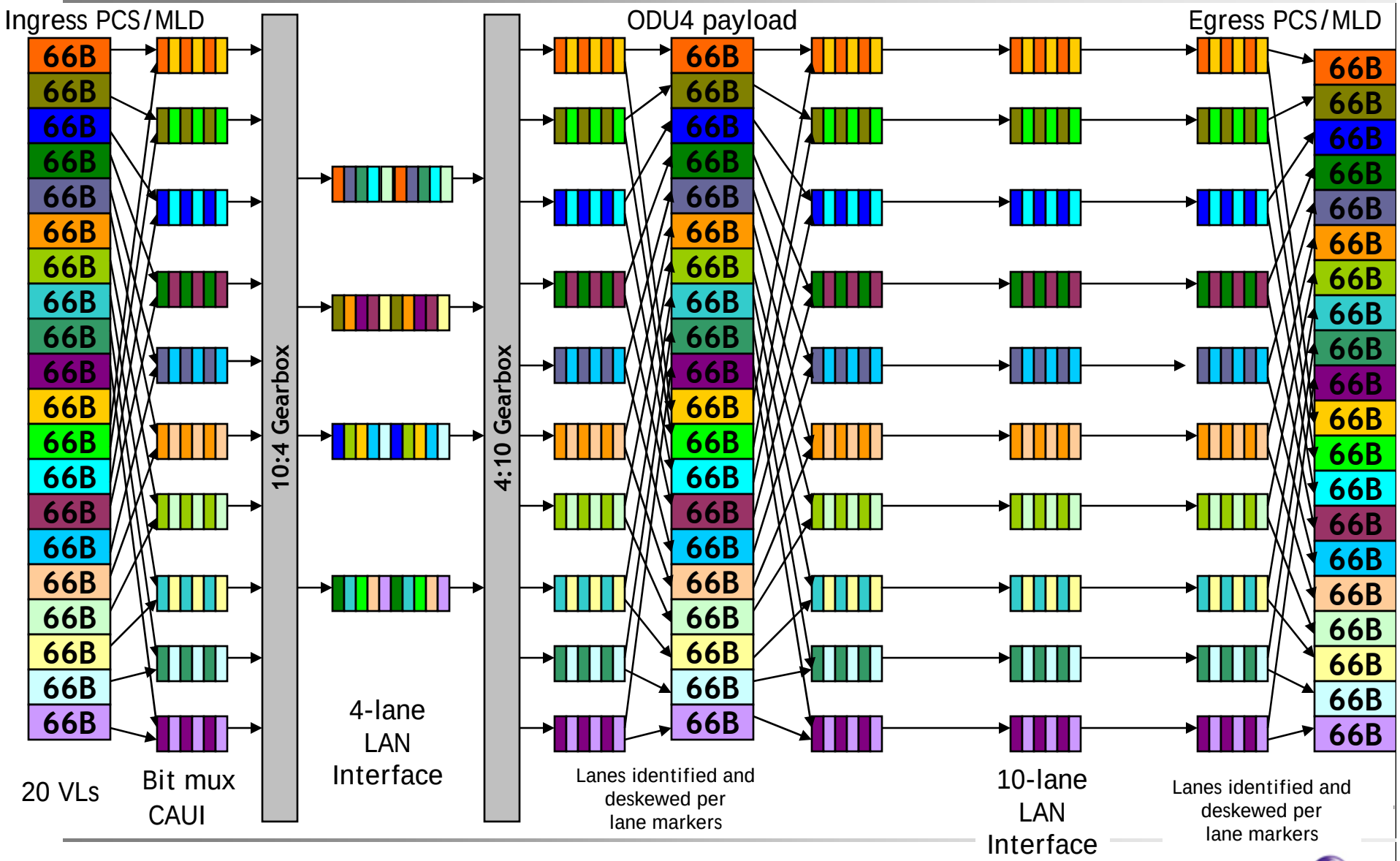


The sequence of bits transported across OTN should not depend on which physical interface is chosen for Ethernet at the ingress or egress

Common PCS – Good news

- The MLD proposal comprises a common PCS that is used across all Ethernet PMDs - see gustlin_01_0308
- As the complexity of using the MLD PCS is no more than that of managing skew to within 32UI (see shafai_01_0308), consensus is moving towards using the MLD PCS for all PHY types including 40 GbE backplane
- Skew in OTN must be managed so that Ethernet over OTN does not exceed LAN deskew budget (OTN must deskew)

Example: Four-Lane 100 GbE LAN interface at OTN ingress; 10-lane 100 GbE LAN interface at OTN egress.



Alcatel·Lucent

Common PCS Proposal (100 GbE and 40 GbE)

- Adopt MLD with 64B/66B coding as the common PCS for all 802.3ba interfaces. This enables:
  - A single canonical form to be used for mapping of any 802.3ba interface with at least codeword transparency over OTN
  - Selection of different Ethernet PMDs at the OTN ingress and egress

OTN support for 40 GbE

- **Two ways to provide 40 GbE transparent transport over OTN:**
  - Choose a MAC bit-rate (e.g., 38.9 Gbit/s) such that 64B/66B coding and lane marker insertion results in a bit-stream that fits the payload area of an OPU3 (not preferred)
  - Impose strict requirements on PCS codeword set that permits codeword transparent mapping of 40 GbE into payload of OPU3 (preferred)
- **Feasibility for codeword transparent mapping from 40.0 Gbit/s MAC rate into capacity of OPU3 payload demonstrated in trowbridge_01_0707, with possible improvements shown in trowbridge_01_0308 (actual standard to be specified as mapping of 40 GbE into OTN by ITU-T SG15)**
  - The proposed transcoding method requires 15 or fewer control block types to be used in underlying 64B/66B code
  - A single additional (among the 16 available) control block type can be used to encode a lane marker, with 56 bits available for a very sparse coding of the lane number
- **10G Base-R 64B/66B coding uses 15 control block types. 40GbE/100GbE may use fewer control block types if packet and ordered set start is restricted to an 8-byte boundary**
- **To rely on transcoding, a fixed, limited set of control block types understood by both IEEE and ITU-T is essential to specification of the mapping of 40 GbE into OPU3 and interoperable implementations**

Possible Changes to 64B/66B for 40 GbE and 100 GbE given 8-byte boundary for packet start and ordered sets

Ordered sets can't start in 5th lane

Packets can't start in 5th lane

Input Data	Sync	Block Payload							
Bit Position: / Data Block Format:	0 1	2							65
$D_0\ D_1\ D_2\ D_3/D_4\ D_5\ D_6\ D_7$	01	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7
Control Block Formats:		Block Type Field							
$C_0\ C_1\ C_2\ C_3/C_4\ C_5\ C_6\ C_7$	10	0x1e	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7
$O_0\ D_1\ D_2\ D_3/S_4\ D_5\ D_6\ D_7$	10	0x33	O_0	D_1	D_2	D_3		D_5	D_6	D_7
$O_0\ D_1\ D_2\ D_3/O_4\ D_5\ D_6\ D_7$	10	0x66	D_1	D_2	D_3	O_0		D_5	D_6	D_7
$S_0\ D_1\ D_2\ D_3/O_4\ D_5\ D_6\ D_7$	10	0x55	D_1	D_2	D_3	O_0	O_4	D_5	D_6	D_7
$S_0\ D_1\ D_2\ D_3/D_4\ D_5\ D_6\ D_7$	10	0x78	D_1	D_2	D_3	D_4		D_5	D_6	D_7
$O_0\ D_1\ D_2\ D_3/C_4\ C_5\ C_6\ C_7$	10	0x4b	D_1	D_2	D_3	O_0	C_4	C_5	C_6	C_7
$T_0\ C_1\ C_2\ C_3/C_4\ C_5\ C_6\ C_7$	10	0x87		C_1	C_2	C_3	C_4	C_5	C_6	C_7
$D_0\ T_1\ C_2\ C_3/C_4\ C_5\ C_6\ C_7$	10	0x99	D_0		C_2	C_3	C_4	C_5	C_6	C_7
$D_0\ D_1\ T_2\ C_3/C_4\ C_5\ C_6\ C_7$	10	0xaa	D_0	D_1		C_3	C_4	C_5	C_6	C_7
$D_0\ D_1\ D_2\ T_3/C_4\ C_5\ C_6\ C_7$	10	0xb4	D_0	D_1	D_2		C_4	C_5	C_6	C_7
$D_0\ D_1\ D_2\ D_3/T_4\ C_5\ C_6\ C_7$	10	0xcc	D_0	D_1	D_2	D_3		C_5	C_6	C_7
$D_0\ D_1\ D_2\ D_3/D_4\ T_5\ C_6\ C_7$	10	0xd2	D_0	D_1	D_2	D_3	D_4		C_6	C_7
$D_0\ D_1\ D_2\ D_3/D_4\ D_5\ T_6\ C_7$	10	0xe1	D_0	D_1	D_2	D_3	D_4	D_5		C_7
$D_0\ D_1\ D_2\ D_3/D_4\ D_5\ D_6\ T_7$	10	0xff	D_0	D_1	D_2	D_3	D_4	D_5	D_6	

Figure 49–7—64B/66B block formats

It is expected that the 64B/66B coding for 40GbE and 100 GbE will use between 11 and 15 control block types, leaving one 4-bit code free for encoding of lane markers if necessary

40 GbE into OPU3 – what can break the mapping?

- Someone could implement a proprietary extension that used non-standard control block types
  - o Extremely unlikely area for proprietary extension as proper packet delineation depends on control block types and misuse could lose packet framing and impair MTTFPA; however
  - o As a safeguard, the standard should contain extremely strong language to prevent proprietary extensions in this area
- Evolution of the standard could allocate new control block types that are not anticipated by the OTN mapper
  - o As a safeguard, the relationship between IEEE 802.3 and ITU-T Recommendation should be clearly noted in the standards

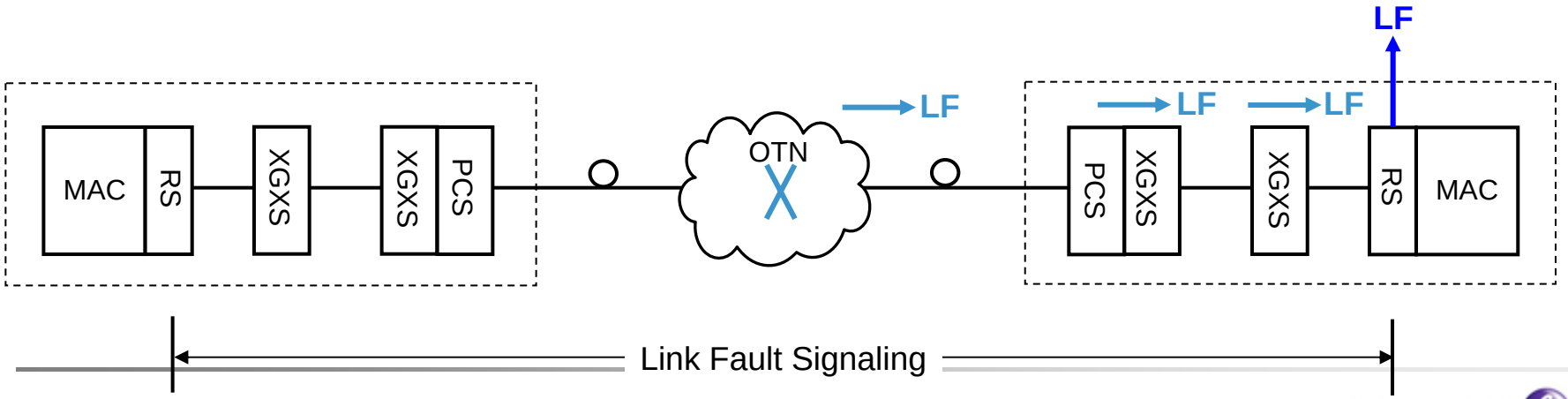
OTN support for 40 GbE proposal

- The aggregate PCS encoded bit-rate for 40 GbE including 64B/66B coding with inserted MLD lane markers shall be no more than 41.25 Gbit/s ±100ppm
- Aside from MLD lane markers, PCS codewords are 64B/66B encoded blocks similar to those used in 10G Base-R (IEEE Std 802.3 clause 49)
- The PCS coding for 40 GbE shall use no more than the 15 control block types specified for 10G Base-R (likely fewer, if 8-byte alignment for packet start and/or ordered sets)
- The equivalent of Figure 49-7 for the 40 GbE PCS shall include the following text:
  - **"Control block types not listed in Figure xx-yy shall not be transmitted and shall be considered an error if received"**
- and Pending concurrence of the 802.3 working group
  - **"The mapping of 40G Base-?? signals into OPU3 (to be) specified in ITU-T Recommendation G.709 depends on the set of control block types shown in Figure xx-yy. Any change to the coding specified in Figure xx-yy must be coordinated with ITU-T Study Group 15."**

Link Fault Signaling for Ethernet over OTN

LF will be transmitted on the Ethernet interface as the forward defect indication when failures are detected within the OTN network (using the same sequence ordered set as in 802.3ae)

- **Consistent with the definition of link fault signaling (LF/RF) in Clause 46**
- **An OTN failure is treated no differently than any other failure between remote and local RS (Clause 46)**
- **Nothing needs to be added or changed for 802.3ba**
- **The equipment functions specified by ITU-T SG15 supporting the OTN mappings for 40GE and 100GE should clarify that Local Fault (LF) should be inserted on the downstream (egress) ethernet interface in the event of OTN failures**



Update to Adopted 100GE 40km SMF PMD Baseline

IEEE 802.3ba Task Force

15-17 July 2008

Chris Cole - Finisar

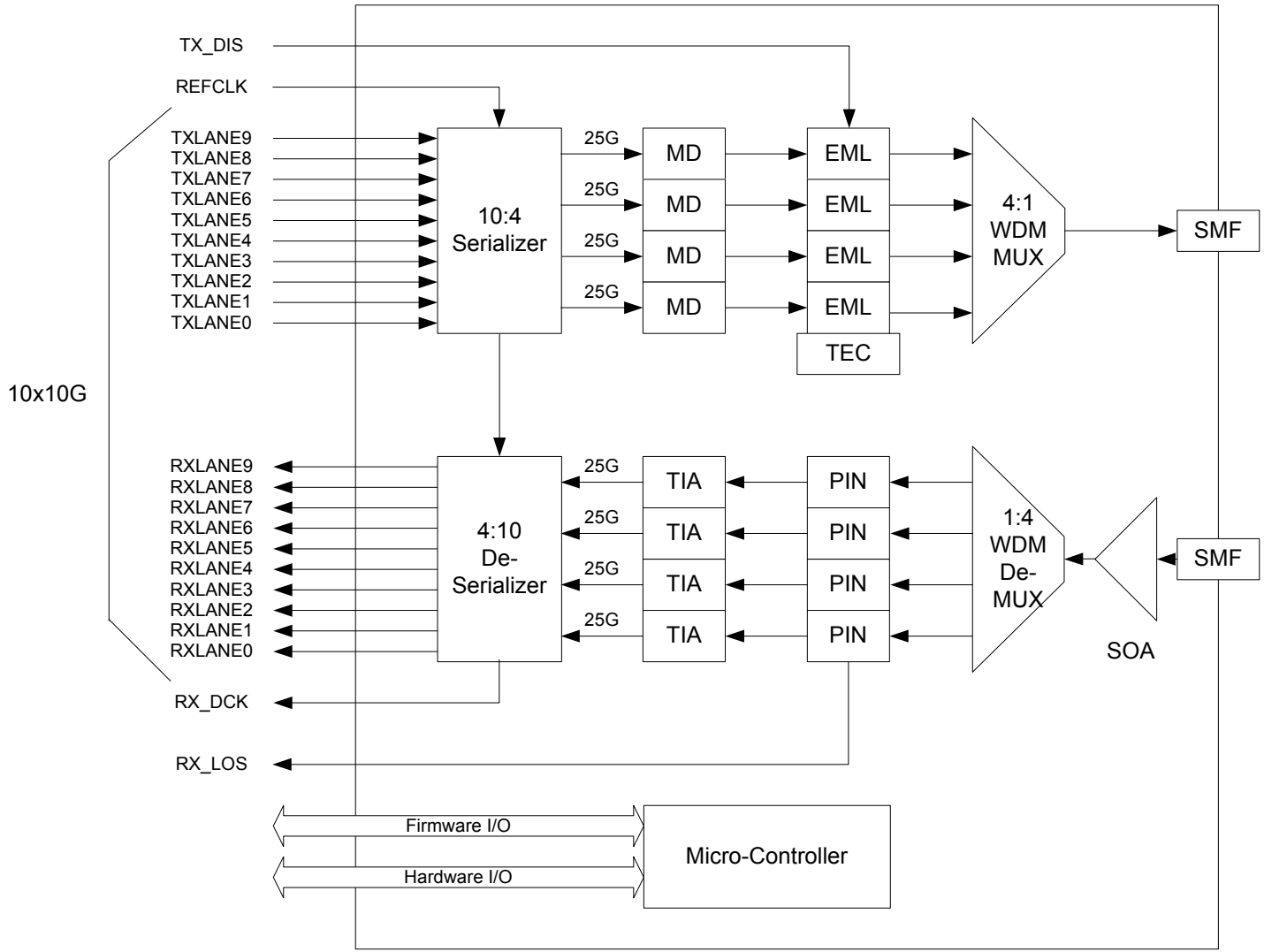
Pete Anslow – Nortel

Jonathan King - Finisar

Introduction

- Baseline Proposal for 100GE 40km SMF PMD objective was adopted at the May 802.3ba meeting (http://www.ieee802.org/3/ba/public/may08/cole_02_0508.pdf.)
- This presentation proposes updates to the Baseline Proposal.
- It lists refinements and additions required for a complete 802.3ba specification.
- All numbers should be viewed as subject to change as a result of continued discussion by 802.3ba participants, through Task Force review.
- Proposed updates to the Baseline Proposal
  - Power Budget reformatted to 802.3ae format specification tables
  - Exact wavelength range set to 2.1nm (changed from tentative 2nm in May)
  - Link Power Budget decreased by 1dB to support 30km operating distance and 40km engineered link operating distance, as per 802.3ae 10GBASE-ER (40km) methodology and format
  - Max sensitivity values increased by 1dB to match the 1dB Budget decrease
  - Maximum transmitter and minimum receiver power values added
  - Eye mask, SMSR, ER, RIN, ORLT, TR, RL, 3dB BW limits added.

40km 1310nm EML 4x25G PMD



LAN WDM Baseline (-10nm) Grid

- ITU G.694.1 specification
- 800GHz spacing (193.1THz base)
- 4 wavelengths shifted by -10nm from minimum dispersion Grid
- Exact wavelengths: 1295.56 1300.05 1304.58 1309.14 nm
- Shorthand wavelengths: 1295, 1300, 1305, 1310 nm
- TX and RX wavelength range: 2.1 nm (2 nm shorthand)
- G.652 A&B 40km SMF worst dispersion and fiber loss
  - Max positive dispersion (1310nm) = 38ps/nm
  - Max negative dispersion (1295nm) = -114ps/nm
  - Max Loss (1310nm) = 16.8dB
  - Max Loss (1295nm) = 17.3dB

100GBASE-ER4 lane assignments

Lane	Center frequencies	Center wavelengths	Wavelength ranges [a]
L_0	231.4 THz	1295.56 nm	1294.53 – 1296.59 nm
L_1	230.6 THz	1300.05 nm	1299.02 – 1301.09 nm
L_2	229.8 THz	1304.58 nm	1303.54 – 1305.63 nm
L_3	229.0 THz	1309.14 nm	1308.09 – 1310.19 nm

[a] Wavelength ranges calculated for center frequencies $\pm$ 23% of 800GHz spacing

100GBASE-ER4 transmit characteristics

Description	100GBASE-ER4	Unit
Signaling speed per lane	25.78125 ±100 ppm	GBd
Lane wavelengths (range)	1294.53 – 1296.59 1299.02 – 1301.09 1303.54 – 1305.63 1308.09 – 1310.19	nm
Transmitter eye mask definition {X1, X2, X3, Y1, Y2, Y3}[a]	TBD	
Side Mode Suppression Ratio (SMSR), (min)	30	dB
Total average launch power (max)	8.4	dBm
Difference in launch power between any two lanes (max)	3.0	dB
Average launch power per lane (max)[b]	2.4	dBm
Average launch power per lane (min)[b]	-2.9	dBm
Optical Modulation Amplitude (OMA), per lane (max)	4.0	dBm
Optical Modulation Amplitude (OMA), per lane (min)	0.1	dBm
Extinction Ratio (min)	8.0	dB
Average launch power of OFF transmitter, per lane (max)	-30	dBm
$RIN_{12}OMA$ (max)[c]	-132	dB/Hz
Optical Return Loss Tolerance (max)	12	dB
Transmitter Reflectance (max)[d]	-12	dB

[a] Tx eye mask spec to be specified as per eye mask methodology discussions

[b] Informative

[c] RIN is scaled by 10*log(10/4) to maintain SNR out of transmitter

[d] -12dB transmitter reflectance helps relax RX reflection spec

100GBASE-ER4 receive characteristics

Description	100GBASE-ER4	Unit
Signaling speed per lane	25.78125 ±100 ppm	GBd
Lane wavelengths (range)	1294.53 – 1296.59 1299.02 – 1301.09 1303.54 – 1305.63 1308.09 – 1310.19	nm
Difference in receive power between any two lanes (max)	4.0	dB
Receive power, per lane (OMA) (max)	4.0	dBm
Average receive power, per lane (max)[a]	4.0	dBm
Average receive power, per lane (min)[b]	-20.9	dBm
Return loss (min)[c]	-26	dB
Receive sensitivity (OMA), per lane (max)	-21.4	dBm
Stressed receive sensitivity (OMA), per lane	-17.9	dBm
Vertical eye closure penalty, per lane	3.5	dB
Receive electrical 3 dB upper cutoff frequency, per lane (max)	31	GHz

[a] The receiver shall tolerate, without damage, the Average Receive Power (max) plus 1 dB

[b] Informative, equals min Tx OMA with infinite ER and max channel insertion loss

[c] Prevents excess coherent interference due to Tx Rx reflectance

100GBASE-ER4 link power budget

Description	100GBASE-ER4		Unit
Power budget	21.5		dB
Operating distance	30	40[c]	km
Channel insertion loss [a]	15	18	dB
Maximum Discrete Reflectance (max)	-26	-26	dB
Allocation for penalties [b]	3.5[d]	3.5	dB
Additional insertion loss allowed	3.0	0.0	dB

[a] Channel insertion loss includes fiber and connector losses for worst case wavelength lane

[b] Dispersion and other penalties for worst case wavelength lane, DGD_{max} = T.B.D.

[c] Links longer then 30km are considered engineered links. Attenuation for such links needs to be less then that guaranteed by 802.ba reference SMF.

[d] Assumes 1.5dB CD Penalty, 1.0dB PMD Penalty, 1.0dB Other Penalties.

100GBASE-xR4 Discussion

Pete Anslow

IEEE P802.3ba, Denver, July 2008

Introduction

This presentation contains the result of informal discussions held during the Denver meeting of the IEEE 802.3ba Task Force on the parameter values for the applications 100GBASE-LR4 and 100GBASE-ER4 aimed at achieving consensus on a suitable parameter set to be used for the first draft version of clause 158

100GBASE-xR4 lane assignments

- **Baseline in cole_01/02_0508 had:**

  “2nm window (precise pass-band TBD)”

- **Consensus to use 2.1 nm values from cole_01/02_0708 for first draft:**

Lane	Center frequencies	Center wavelengths	Wavelength ranges
L_0	231.4 THz	1295.56 nm	1294.53 – 1296.59 nm
L_1	230.6 THz	1300.05 nm	1299.02 – 1301.09 nm
L_2	229.8 THz	1304.58 nm	1303.54 – 1305.63 nm
L_3	229.0 THz	1309.14 nm	1308.09 – 1310.19 nm

[a] **Wavelength ranges calculated for center frequencies ± 23% of 800GHz spacing**

100GBASE-xR4 transmit characteristics

- **Baseline in cole_01/02_0508 had no values for red parameters**

- **Consensus to use values from cole_01/02_0708 as below for first draft**

Description	100GBASE-xR4	Unit
Signaling speed per lane	25.78125 ±100 ppm	GBd
Lane wavelengths (range)	1294.53 – 1296.59 1299.02 – 1301.09 1303.54 – 1305.63 1308.09 – 1310.19	nm
Transmitter eye mask definition {X1, X2, X3, Y1, Y2, Y3}[a]	TBD	
Side Mode Suppression Ratio (SMSR), (min)	30	dB
Average launch power of OFF transmitter, per lane (max)	-30	dBm
RIN_{12} OMA (max)	-132	dB/Hz
Optical Return Loss Tolerance (max)	12	dB
Transmitter Reflectance (max)	-12	dB

100GBASE-xR4 receive characteristics

- **Baseline in cole_01/02_0508 had no values for red parameters**

- **Consensus to use values from cole_01/02_0708 as below for first draft**

Description	100GBASE-xR4	Unit
Signaling speed per lane	25.78125 ±100ppm	GBd
Lane wavelengths (range)	1294.53 – 1296.59 1299.02 – 1301.09 1303.54 – 1305.63 1308.09 – 1310.19	nm
Return loss (min)	-26	dB
Receive electrical 3 dB upper cutoff frequency, per lane (max)	31	GHz

[a] The receiver shall tolerate, without damage, the Average Receive Power (max) plus 1 dB

100GBASE-LR4 power budget

- **Values in cole_01_0708 and proposed changes in isono_01_0707 were discussed**

- **Consensus to use values from cole_01_0708 as modified by isono_01_0707 for first draft**

- **The resulting power budget is shown in the right hand column on the next slide. The red values are those that have changed with respect to cole_01_0708**

100GBASE-LR4 power budget with TDP

	cole_0508	cole_0708	isono_0708	Consensus	
Total ave power (max)		10.0		10.0	dBm
Ave power per lane (max)		4.0		4.0	dBm
Ave power per lane (min)		-3.0		-3.8	dBm
OMA per lane (max)		4.0		3.2	dBm
OMA per lane (min)	-0.2	0.0	-0.8	-0.8	dBm
TDP per lane (max)			2.0 - 2.5	2.2 *	dB
OMA-TDP per lane (min)			-1.8	-1.8	dBm
Extinction Ratio (min)	4.5	4.0	4.0	4.0	dB
Power budget	7.5	8.1	8.3 - 8.8	8.5	dB
Channel insertion loss	6.3	6.3	6.3	6.3	dB
Allocation for penalties	1.2	1.8	2.0 - 2.5	2.2 *	dB
Rx power, per lane OMA (max)		4.0		4.0	dBm
Rx power, per lane ave (max)		4.0		4.0	dBm
Rx power, per lane ave (min)		-9.3		-10.1	dBm
Sensitivity OMA per lane (max)	-7.7	-8.1	-8.1	-8.1	dBm
Stressed sens. OMA per lane		-6.3		-6.3	dBm
Vertical eye closure pen. per lane		1.8	2.0 - 2.5	1.8	dB

The value of 2.2 dB for TDP per lane (max) is a placeholder. The final value is expected to be between 1.8 and 2.5 dB

100GBASE-LR4 tables

- **The result of making these changes to the power budget for 100GBASE-LR4 is shown in the next three slides**

100GBASE-LR4 transmit characteristics

Description	100GBASE-LR4	Unit
Signaling speed per lane	25.78125 ±100 ppm	GBd
Lane wavelengths (range)	1294.53 – 1296.59 1299.02 – 1301.09 1303.54 – 1305.63 1308.09 – 1310.19	nm
Transmitter eye mask definition {X1, X2, X3, Y1, Y2, Y3}	TBD	
Side Mode Suppression Ratio (SMSR), (min)	30	dB
Total average launch power (max)	10	dBm
Average launch power per lane (max) [a]	4.0	dBm
Average launch power per lane (min) [a]	-3.8	dBm
Optical Modulation Amplitude (OMA), per lane (max)	3.2	dBm
Launch power per lane (min) in OMA minus TDP	-1.8	dBm
Optical Modulation Amplitude (OMA), per lane (min) [b]	-0.8	dBm
Transmitter and dispersion penalty per lane (max)	2.2 *	dB
Extinction Ratio (min)	4.0	dB
Average launch power of OFF transmitter, per lane (max)	-30	dBm
RIN_{12} OMA (max)	-132	dB/Hz
Optical Return Loss Tolerance (max)	12	dB
Transmitter Reflectance (max)	-12	dB

[a] Informative

[b] Even if the TDP < 1 dB, the OMA (min) must exceed this value.

* The value of 2.2 dB for TDP per lane (max) is a placeholder. The final value is expected to be between 1.8 and 2.5 dB

100GBASE-LR4 receive characteristics

Description	100GBASE-LR4	Unit
Signaling speed per lane	25.78125 ±100ppm	GBd
Lane wavelengths (range)	1294.53 – 1296.59 1299.02 – 1301.09 1303.54 – 1305.63 1308.09 – 1310.19	nm
Receive power, per lane (OMA) (max)	4.0	dBm
Average receive power, per lane (max)[a]	4.0	dBm
Average receive power, per lane (min)[b]	-10.1	dBm
Return loss (min)	-26	dB
Receive sensitivity (OMA), per lane (max)	-8.1	dBm
Stressed receive sensitivity (OMA), per lane	-6.3	dBm
Vertical eye closure penalty, per lane	1.8	dB
Receive electrical 3 dB upper cutoff frequency, per lane (max)	31	GHz

[a] The receiver shall tolerate, without damage, the Average Receive Power (max) plus 1 dB
[b] Informative, equals min Tx OMA with infinite ER and max channel insertion loss

100GBASE-LR4 link power budget

Description	100GBASE-LR4	Unit
Power budget	8.5	dB
Operating distance	10	km
Channel insertion loss [a]	6.3	dB
Maximum Discrete Reflectance (max)	-26	dB
Allocation for penalties [b]	2.2 *	dB
Additional insertion loss allowed	0.0	dB

[a] Channel insertion loss includes fiber and connector losses for worst case wavelength lane

[b] Dispersion and other penalties for worst case wavelength lane

* The value of 2.2 dB for allocation for penalties is a placeholder. The final value is expected to be between 1.8 and 2.5 dB

100GBASE-ER4 power budget

- **The values in cole_02_0708 were discussed**
- **Consensus to use values from cole_02_0708 for first draft**

40GBASE-KR4 backplane PHY proposal and Next Steps

Richard Mellitz & Ilango Ganga
Intel Corporation

May 13, 2008

Supporters

- Andre Szczepanek, Texas Instruments
- Arne Alping, Ericsson
- Arthur Marris, Cadence Design Systems
- Brad Booth, AMCC
- David Koenen, HP
- Frank Chang, Vitesse
- Gourgen Oganessyan, Quellan
- Jeff Lynch, IBM
- Scott Kipp, Brocade
- Tom Palkert, Luxtera

Supporters for mellitz_01_0308

- Jeff Cain, Cisco Systems
- Chris DiMinico, MC Communications
- Pravin Patel, IBM

Key messages

- Proposal to adopt 10GBASE-KR as a baseline for specifying 40GBASE-KR4 with the following changes
  - Backplane layer diagram (Clause 69)
  - Leverage 10GBASE-KR PMD for operation over 4 lanes (Clause 72)
  - Auto-Negotiation (Clause 73)
  - Forward Error correction (Clause 74)

Considerations for 40G Backplane Ethernet PHY

- To be architecturally consistent with the Backplane Ethernet layer stack illustrated in Clause 69
- To interface to a 4-lane backplane medium with interconnect characteristics recommended in IEEE Std 802.3ap (Annex 69B)
  - Most generation 2 blade systems are built with 4-lanes (10Gbaud KR ready)
- Leverage 10GBASE-KR technology/specifications (Clause 72 and Annex 69A) to define 40GBASE-KR4 PHY:
  - 64B/66B block coding
  - Startup protocol (per lane)
  - Signaling speed 10.3125Gbd (per lane)
  - Electrical characteristics
  - Test methodology and procedures
- Optional FEC sublayer
  - PCS to interface to optional FEC sublayer consistent with Clause 74 specification
- Compatible with Backplane Ethernet Auto-Neg (Clause 73)
  - Enhancement to indicate 40GbE ability

Backplane Ethernet overview

- IEEE Std 802.3ap-2007 Backplane Ethernet defines 3 PHY types
  - 1000BASE-KX : 1-lane 1 Gb/s PHY (Clause 70)
  - 10GBASE-KX4: 4-lane 10Gb/s PHY (Clause 71)
  - 10GBASE-KR : 1-lane 10Gb/s PHY (Clause 72)
- Forward Error Correction (FEC) for 10GBASE-R (Clause 74) – optional
  - Optional FEC to increase link budget and BER performance
- Auto-negotiation (Clause 73)
  - Auto-Neg between 3 PHY types (AN is mandatory to implement)
  - Parallel detection for legacy PHY support
    - Automatic speed detection of legacy 1G/10G backplane SERDES devices
  - Negotiate FEC capability
- Clause 45 MDIO interface for management
- Channel
  - Controlled impedance (100 Ohm) traces on a PCB with 2 connectors and total length up to at least 1m.
  - Channel model is informative (Annex 69B)
- Interference tolerance testing (Annex 69A)
- Support a BER of 10^{-12} or better

Existing backplane architecture

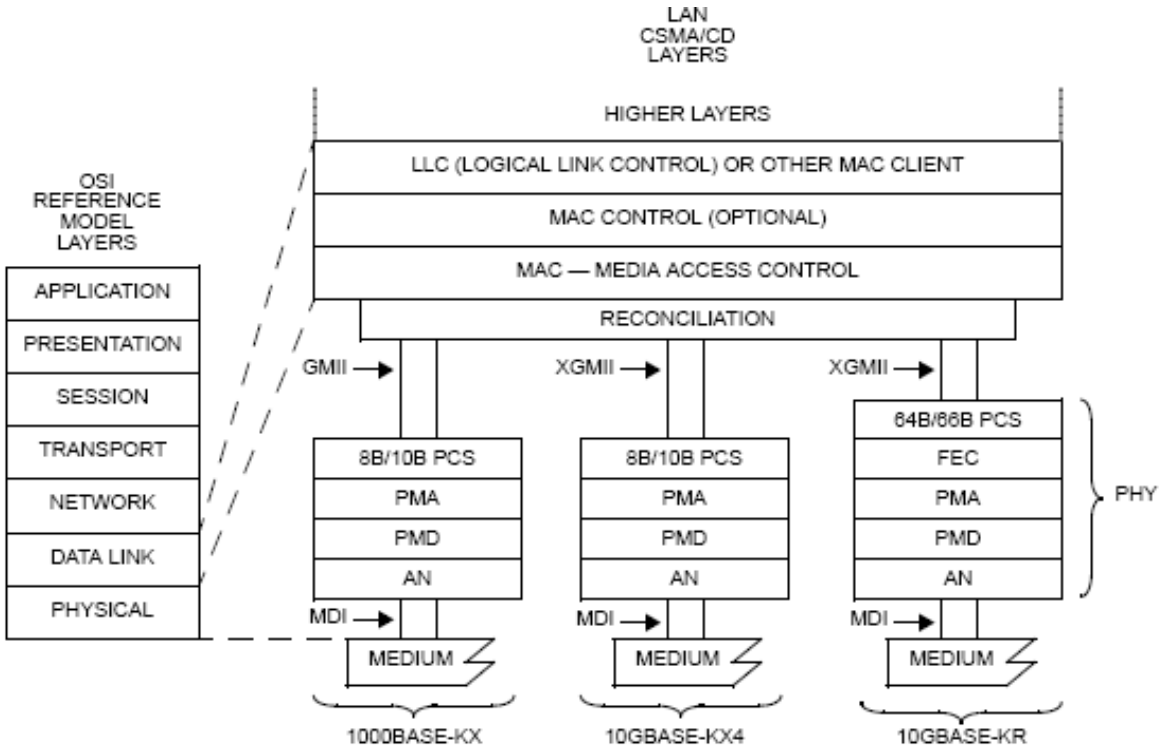
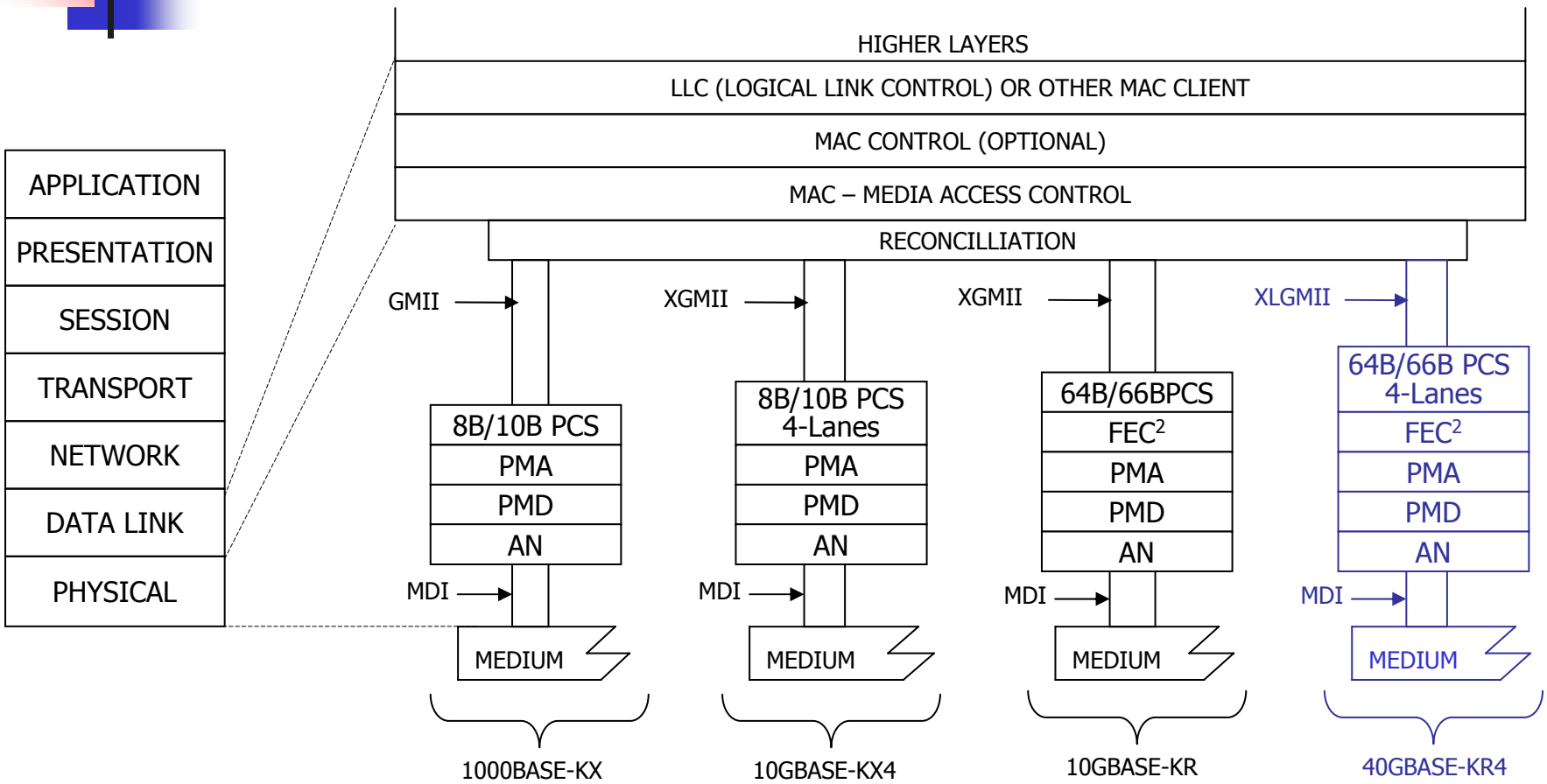


Figure 69–1—Architectural positioning of Backplane Ethernet

Proposed backplane architecture with 40GbE



Note: 2. Optional

Figure 69-1 Architectural positioning of Backplane Ethernet

Proposed Auto-Neg changes

- IEEE Std 802.3ap defines Auto-Negotiation for backplane Ethernet PHYs
  - AN uses DME signaling with 48-bit base pages to exchange link partner abilities
  - AN is mandatory for 10GBASE-KR backplane PHY, negotiates FEC ability
  - Lane 0 of the MDI is used for Auto-Negotiation, of single or multi-lane PHYs
- Proposal for 40GBASE-KR4 (Ability to negotiate with other 802.3ap PHYs)
  - Add a Technology Ability bit A3 to indicate 40GbE ability (A3 is currently reserved)
  - No changes to backplane AN protocol or management register format
  - No change to negotiate FEC ability, FEC when selected to be enabled on all 4 lanes
  - AN mandatory for 40GBASE-KR4, no parallel detect required for 40G

Table 73–4—Technology Ability field encoding

Bit	Technology
A0	1000BASE-KX
A1	10GBASE-KX4
A2	10GBASE-KR
~~A3 through A24~~	~~Reserved for future technology~~
A3	**40GBASE-KR4**
A4 through A24	**Reserved for future technology**

Proposed 40GBASE-KR4 PMD

- Leverage 10GBASE-KR (Clause 72) to specify 40GBASE-KR4 with following changes for 4 lane operation
  - Change KR Link diagram for 4 lanes (similar to KX4)
  - Change KR PMD service interface to support 4 logical streams (similar to KX4)
  - Change PMD control variable mapping table to include management variables for 4 lanes

40GBASE-KR4 Link block diagram

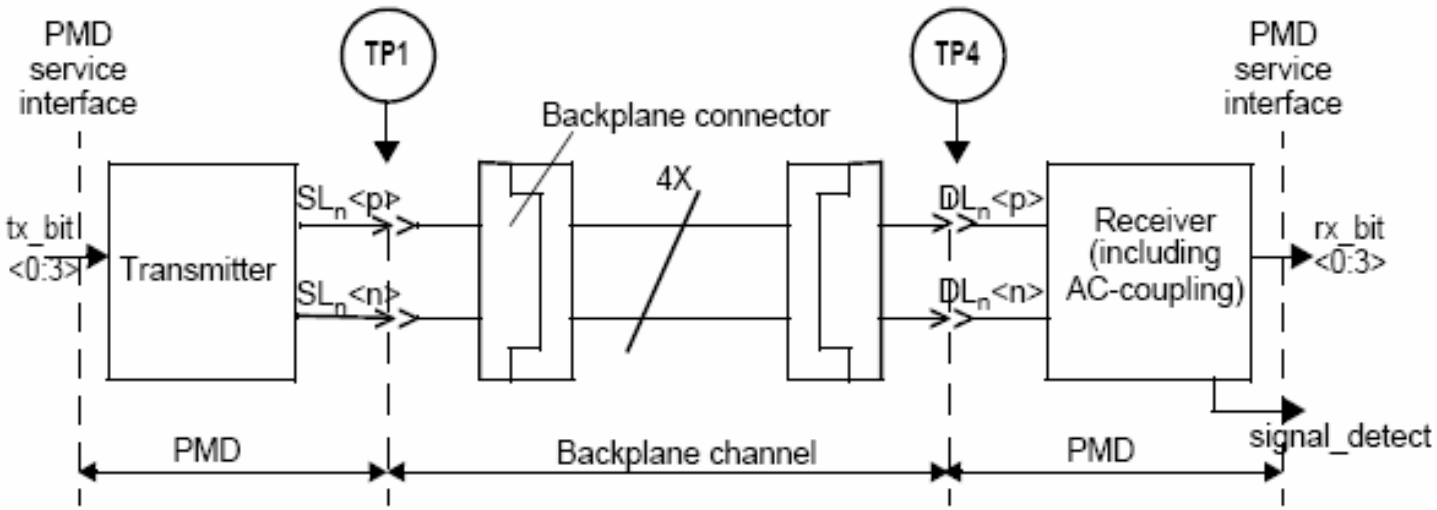


Figure ~~71–1~~—Link block diagram

Service Interfaces for KR4 PMD

- PMD Service Interface
  - Service interface definition as in Clause 72
  - Specify 4 logical streams of 64B/66B code groups from PMA
    - PMD_UNITDATA.request (txbit<0:3>)
    - PMD_UNITDATA.indication (rxbit<0:3>)
    - PMD_SIGNAL.indication (SIGNAL_DETECT<0:3>)
      - "While normally intended to be an indicator of signal presence, is used by 10GBASE-KR to indicate the successful completion of the start-up protocol". Enumerate for 4 lanes
- AN Service Interface (Same as Clause 73)
  - Support AN_LINK.indication primitive
  - Requires associated PCS to support this primitive

PMD MDIO function mapping (1)

- Support management variables for 4 lanes
- Include lane by lane Transmit disable

Table ~~71–2~~–MDIO/PMD control variable mapping

MDIO control variable	PMA/PMD register name	Register/ bit number	PMD control variable
Reset	Control register 1	~~1.0.15~~	PMD_reset
Global Transmit Disable	Transmit disable register	~~1.9.0~~	Global_PMD_transmit_disable
Transmit disable 3	Transmit disable register	~~1.9.4~~	PMD_transmit_disable_3
Transmit disable 2	Transmit disable register	~~1.9.3~~	PMD_transmit_disable_2
Transmit disable 1	Transmit disable register	~~1.9.2~~	PMD_transmit_disable_1
Transmit disable 0	Transmit disable register	~~1.9.1~~	PMD_transmit_disable_0
Restart training	~~10GBASE-KR~~ PMD control register	~~1.150.0~~	mr_restart_training
Training enable	~~10GBASE-KR~~ PMD control register	~~1.150.1~~	mr_training_enable

PMD MDIO function mapping (2)

- Support management variables for 4 lanes
  - Add lane by lane signal detect
  - Enumerate status indication per lane as appropriate

Table ~~71–3~~—MDIO/PMD status variable mapping

MDIO status variable	PMA/PMD register name	Register/ bit number	PMD status variable
Fault	Status register 1	~~1.1.7~~	PMD_fault
Transmit fault	Status register 2	~~1.8.11~~	PMD_transmit_fault
Receive fault	Status register 2	~~1.8.10~~	PMD_receive_fault
Global PMD Receive signal detect	Receive signal detect register	~~1.10.0~~	Global_PMD_signal_detect
PMD signal detect 3	Receive signal detect register	~~1.10.4~~	PMD_signal_detect_3
PMD signal detect 2	Receive signal detect register	~~1.10.3~~	PMD_signal_detect_2
PMD signal detect 1	Receive signal detect register	~~1.10.2~~	PMD_signal_detect_1
PMD signal detect 0	Receive signal detect register	~~1.10.1~~	PMD_signal_detect_0
Receiver status	~~10GBASE-KR~~ PMD status register	~~1.151.0~~	rx_trained
Frame lock	~~10GBASE-KR~~ PMD status register	~~1.151.1~~	frame_lock
Start-up protocol status	~~10GBASE-KR~~ PMD status register	~~1.151.2~~	training
Training failure	~~10GBASE-KR~~ PMD status register	~~1.151.3~~	training_failure

KR4 PMD transmit & receive functions

- PMD transmit function (enumerate for 4 lanes)
  - Converts 4 logical streams from PMD service interface into 4 separate electrical streams delivered to MDI
  - Separate lane by lane TX disable function in addition to Global TX disable function
- PMD receive function (enumerate for 4 lanes)
  - Converts 4 separate electrical streams from MDI into 4 logical streams to PMD service interface
  - Separate lane by lane signal detect function in addition to Global signal detect function
- Same electrical specifications as defined in Clause 72 for 10GBASE-KR PMD
  - Receiver Compliance defined in Annex 69A (Interference Tolerance Test) and referenced in Clause 72

PMD Control function
Startup & Training

- Reuse Clause 72 control function for KR4 PMD (Startup & Training)
  - Used for tuning equalizer settings for optimum backplane performance
  - Use Clause 72 training frame structure
  - Use same PRBS 11 pattern, with randomness between lanes
- Same Control channel spec as in Clause 72, enumerated per lane
  - All 4 lanes are independently trained
  - Report Global Training complete only when all 4 lanes are trained
  - Same Frame lock state diagram (Fig 72-4)
  - Same Training state diagram with enumeration of variables corresponding to 4 lanes (Fig 72-5)
  - Enumerate the management registers for coefficient update field and status report field for 4 lanes

Electrical characteristics

- 40GBASE-KR4 Transmit electrical characteristics
  - Same as 10GBASE-KR TX characteristics and waveforms as specified in Clause 72
  - Same test fixture setup as in Clause 72
- 40GBASE-KR4 Receiver electrical characteristics
  - Same as 10GBASE-KR RX characteristics specified in Clause 72 and Annex 69A

Receiver Interference tolerance test

- Test procedure specified in Annex 69A
- Receiver interference tolerance parameters for 40GBASE-KR4 PMD
  - Same as Receiver interference tolerance test parameters as in Clause 72
  - No change to broadband noise amplitude for KR4

Forward Error Correction

- Reuse FEC specification for 10GBASE-R (Clause 74)
  - The FEC sublayer transparently passes 64B/66B code blocks
  - Change to accommodate FEC sync for 4 lanes
    - Same state diagram for FEC block lock
    - Report Global Sync achieved only if all lanes are locked
    - Possibly add a FEC frame marker signal that could be used for lane alignment

FEC MDIO variable mapping

Table 74–2—MDIO/FEC variable mapping

MDIO variable	PMA/PMD register name	Register/bit number	FEC variable
10GBASE-R FEC ability	10GBASE-R FEC ability register	1.170.0	FEC_ability
10GBASE-R FEC Error Indication ability	10GBASE-R FEC ability register	1.170.1	FEC_Error_Indication_ability
FEC Enable	10GBASE-R FEC control register	1.171.0	FEC_Enable
FEC Enable Error Indication	10GBASE-R FEC control register	1.171.1	FEC_Enable_Error_to_PCS
FEC corrected blocks	10GBASE-R FEC corrected blocks counter register	1.172, 1.173	FEC_corrected_blocks_counter
FEC uncorrected blocks	10GBASE-R FEC uncorrected blocks counter register	1.174, 1.175	FEC_uncorrected_blocks_counter

- Enumerate the following counters for 4 lanes
  - FEC_corrected_blocks_counter
  - FEC_uncorrected blocks_counter
  - Possibly use indexed addressing to conserve MDIO address space

Interconnect Characteristics

- Interconnect characteristics (informative) for backplane is defined in Annex 69B
  - No proposed changes
- 40GBASE-KR4 PHY to interface to the 4 lane backplane medium to take advantage of 802.3ap KR ready blade systems in deployment

Summary

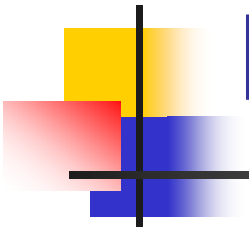
Summary

- 40GbE backplane PHY to be architecturally consistent with IEEE Std 802.3ap layer stack
- Adopt 10GBASE-KR as baseline to specify 40GBASE-KR4 PHY with appropriate changes proposed in this document
- Interface to 4 lane backplane medium to take advantage of 802.3ap KR ready blade systems in deployment

- Appropriate changes to add EEE feature, when adopted by 802.3az for KR
- PCS proposals and interface definitions to accommodate backplane Ethernet architecture (including FEC and AN)

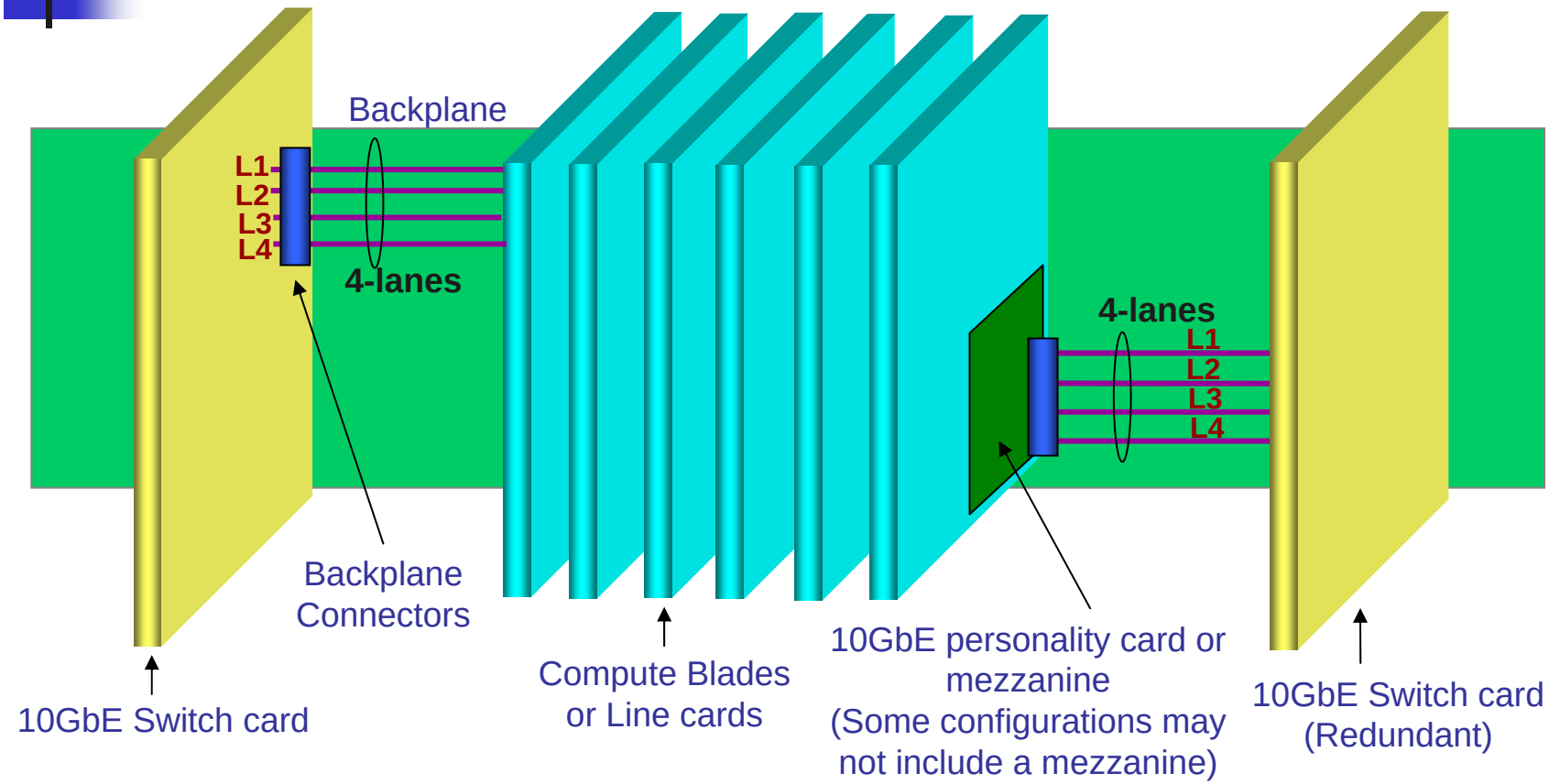
Next Steps

- Make a second generation blade channel model (IEEE Std 802.3ap KR compatible) available to the P802.3ba task force by July '08
- Simulations showing technical feasibility of 40GBASE-KR4 over 40G ready IEEE Std 802.3ap compatible 4 lane backplane system with compliant receivers



Backup

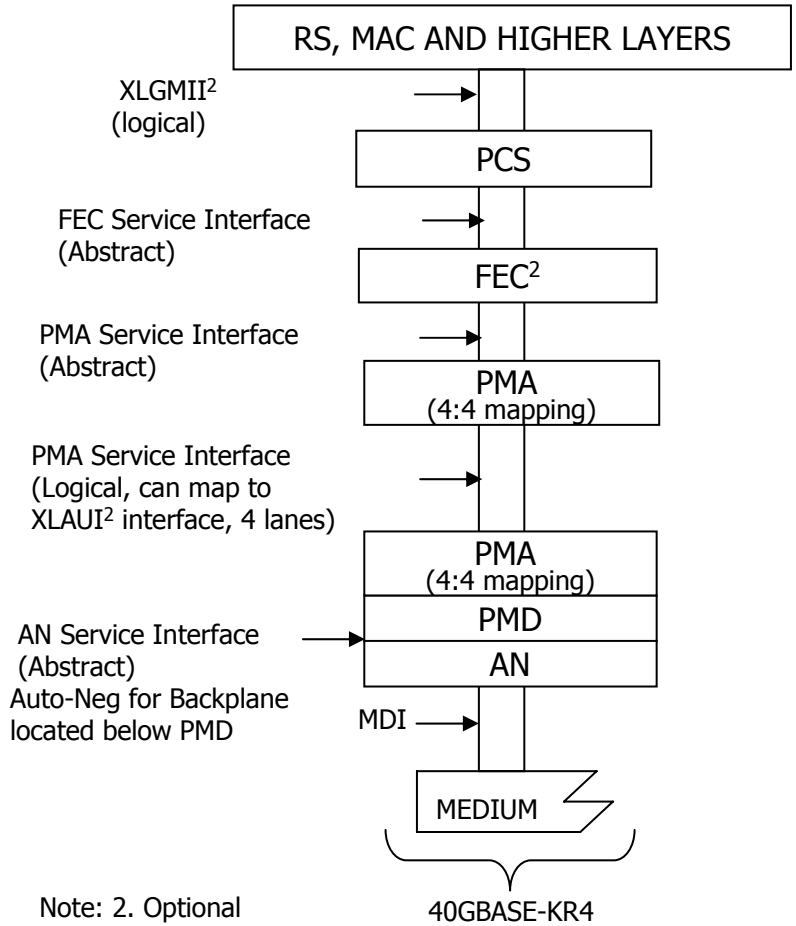
Typical backplane system illustration



Note: The switch cards are shown at the chassis edge for simplicity.
In real systems there could be multiple fabrics located at the center, edge, or rear of the chassis

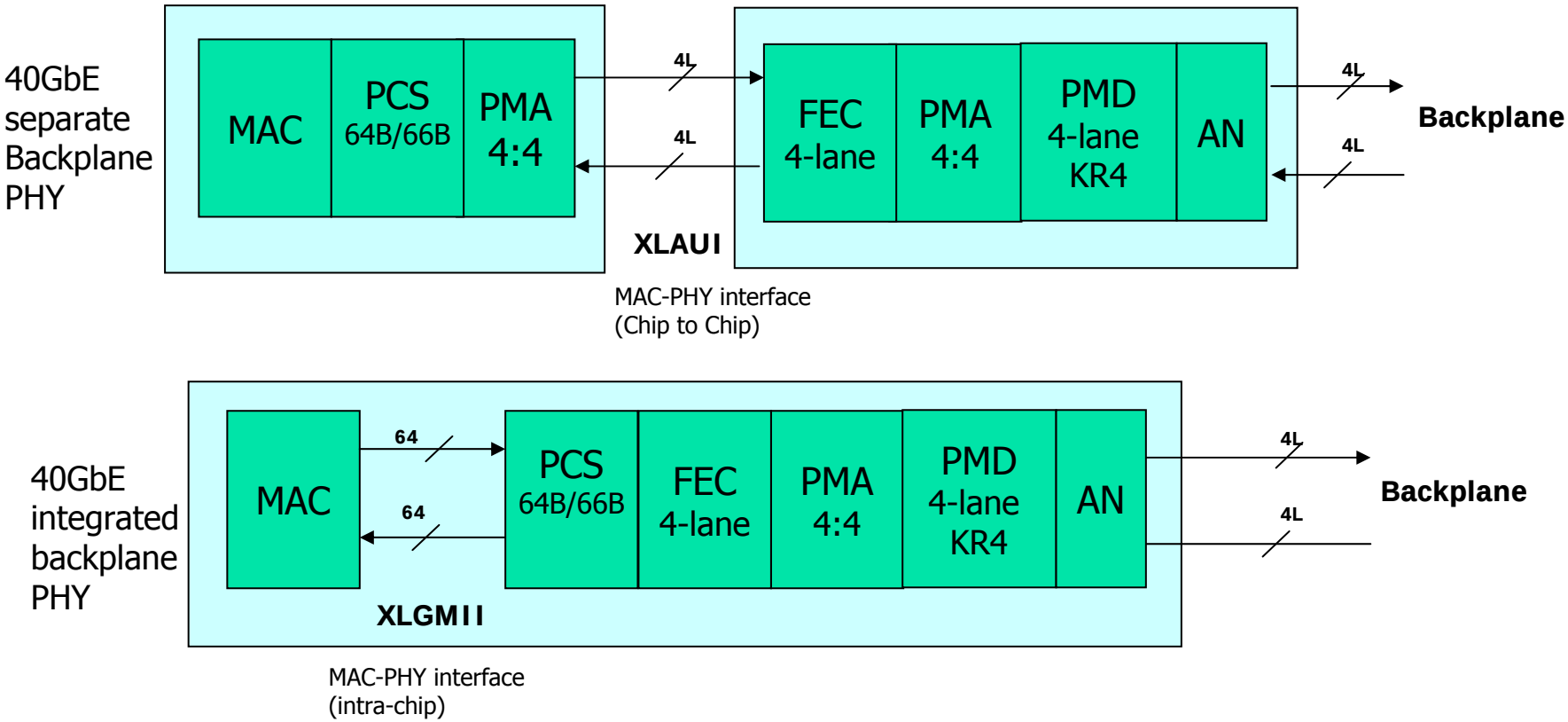
Proposed 40GbE architecture

- XLGMII (intra-chip)
  - Logical, define data/control, clock, no electrical specification
- PCS
  - 64B/66B encoding
  - Lane distribution and alignment
- XLAUI (chip-to-chip)
  - 10.3125 GBaud electrical interface
  - 4 lanes, short reach
- FEC service interface
  - Abstract, can map to XLAUI electrical interface
- PMA Service interface
  - Logical n lanes, can map to XLAUI electrical interface
- PMD Service interface
  - Logical



See ganga_01_0508 for 40/100G architecture and interfaces

Possible implementation examples



100/40 GbE PMA Proposal

Mark Gustlin – Cisco
Steve Trowbridge – Alcatel-Lucent

IEEE P802.3ba July 2008 Denver

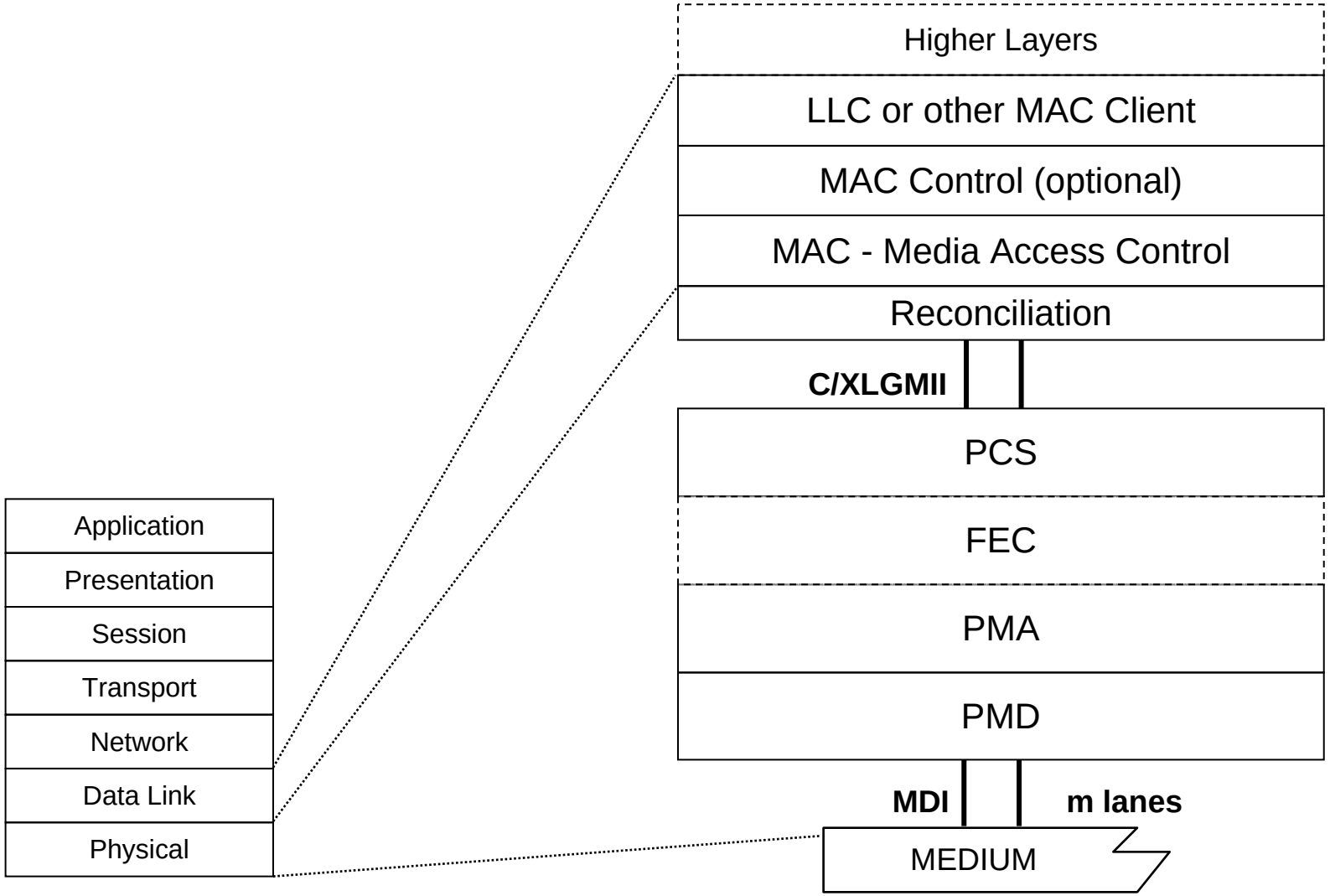
Contributors and Supporters

- Brad Booth - AMCC
- Gary Nicholl - Cisco
- Chris Cole – Finisar
- Subi Krishnamurthy - Force10 Networks
- Shashi Patel - Foundry
- Ryan Latchman – Gennum
- Shinji Nishimura - Hitachi
- Hidehiro Toyoda – Hitachi
- Pete Anslow - Nortel
- Farhad Shafai - Sarance

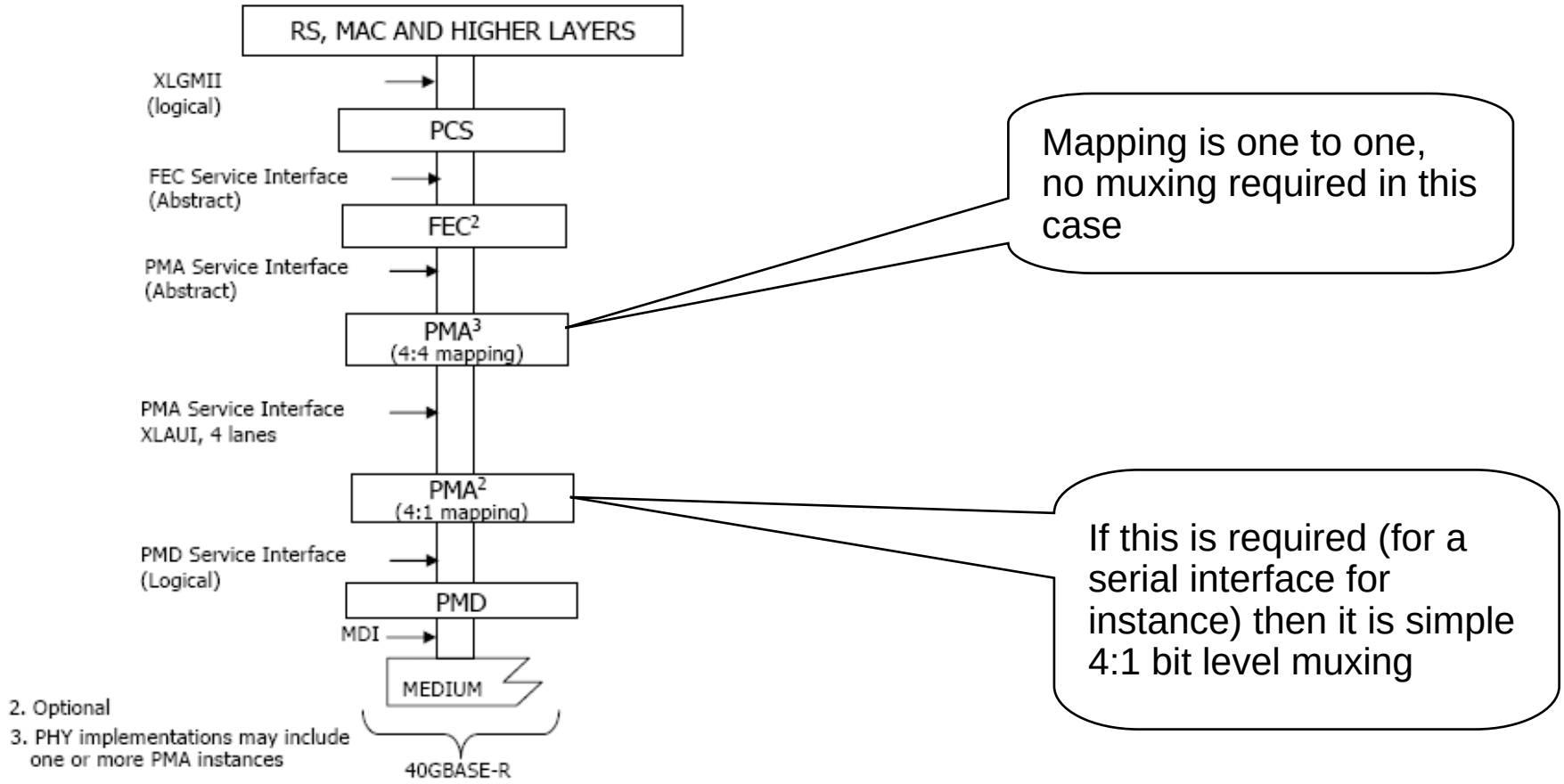
Overview

- This is a proposal for the logical portion on the PMA for 100GE and 40GE
- It does not cover the electrical interfaces (CAUI/XLAUI)

40GE/100GE Generic Architecture

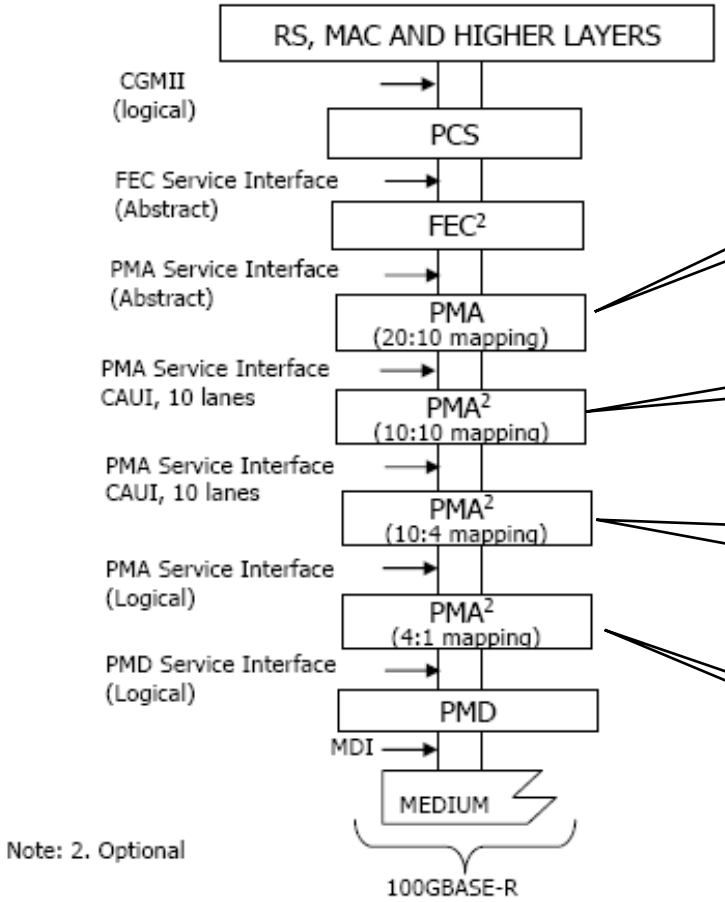


40GE PMA Variants



From ganga_01_0508

100GE PMA Variants



Mapping is two to one, Simply done at the bit level with 2:1 muxes.

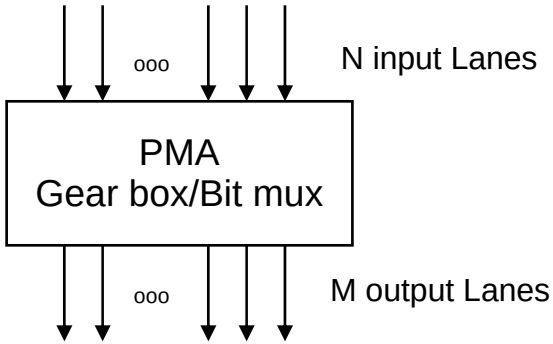
Mapping is one to one, no muxing required in this case

If this is required (for the single mode interface for example) then it is 10:4 gearbox. It could also be implemented as 2x 5:2 muxes

If this is required (for a future serial interface for example) then it is a 4:1 mux.

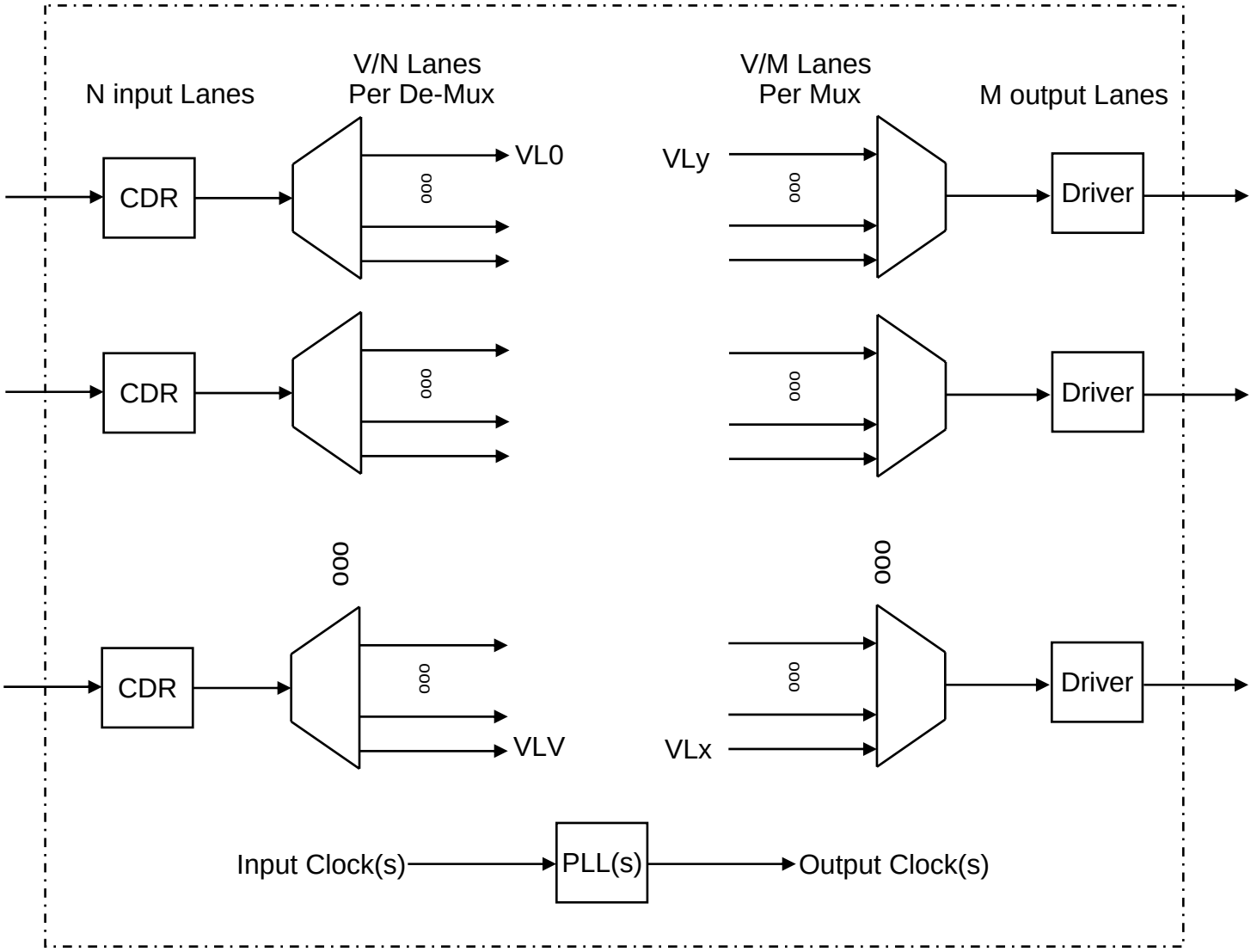
From ganga_01_0508

A Parameterized PMA



- V is the number of Virtual Lanes
- N input lanes, each with V/N virtual lanes
- M output lanes, each with V/M virtual lanes
- The muxing/gearboxing follows the rules as stated later in this presentation

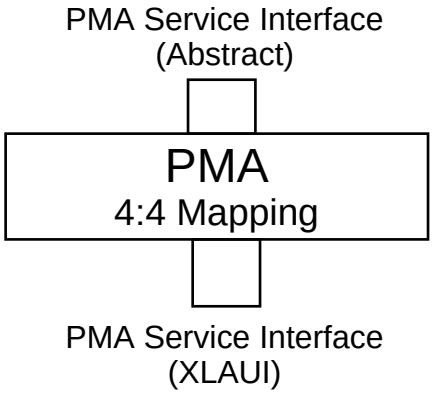
A Parameterized PMA - Details



All implementations that map every input VL to an output VL position are valid, even if they do not completely demux and remux the VLs

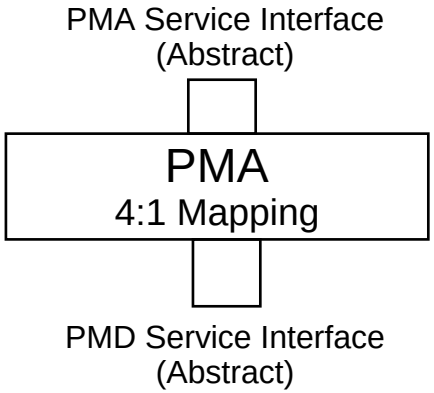
40GE PMA Variant #1

- In the transmit direction the following is provided:
  1. Direct 1:1 mapping of the interface.
  2. Transmission of parallel data to PMD.
- In the receive direction the following is provided:
  1. Direct 1:1 mapping of the interface
  2. Transmission of parallel data to PMA client.
  3. Provide link status information.



40GE PMA Variant #2

- In the transmit direction the following is provided:
  1. Bit level multiplexing of the 4 input lanes into a single output lane
  2. Provide a clock source to the PMA client.
  3. Transmission of serial data to the PMD.
- In the receive direction the following is provided:
  1. Reception of serial data from the PMD
  2. Provides receive clock to PMA client
  3. Bit level de-multiplexing of the serial data into four output lanes
  4. Transmission of parallel data to PMA client.
  5. Provide link status information.



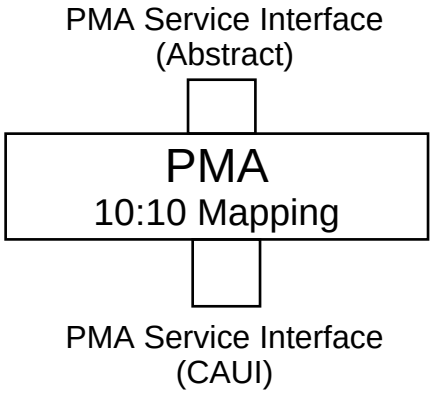
PSI 0	PSI 1	PSI 2	PSI 3
8	9	10	11
4	5	6	7
0	1	2	3

Serial I/F
5
4
3
2
1
0

One possible bit muxing order. PSI = PMA Service Interface

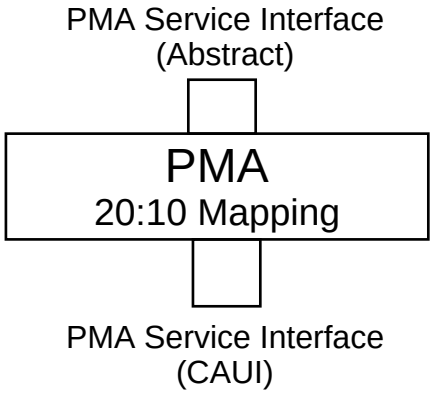
100GE PMA Variant #1

- In the transmit direction the following is provided:
  1. Direct 1:1 mapping of the interface.
  2. Transmission of parallel data to PMD.
- In the receive direction the following is provided:
  1. Direct 1:1 mapping of the interface
  2. Transmission of parallel data to PMA client.
  3. Provide link status information.



100GE PMA Variant #2

- In the transmit direction the following is provided:
  1. Bit level multiplexing of the 20 input lanes into a 10 output lanes
  2. Provide a clock source to the PMA client.
  3. Transmission of parallel data to the PMD.
- In the receive direction the following is provided:
  1. Reception of parallel data from the PMD
  2. Bit level de-multiplexing of the 10 input lanes into 20 output lanes
  3. Provides receive clock to PMA client
  4. Transmission of parallel data to PMA client.
  5. Provide link status information.



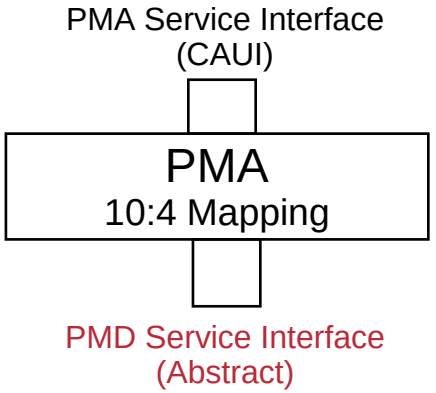
PSI 0	PSI 1	PSI 2	PSI 3	PSI 4	PSI 5	PSI 6	PSI 7	PSI 8	PSI 9	PSI 10	PSI 11	PSI 12	PSI 13	PSI 14	PSI 15	PSI 16	PSI 17	PSI 18	PSI 19
20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

CI 0	CI 1	CI 2	CI 3	CI 4	CI 5	CI 6	CI 7	CI 8	CI 9
30	31	32	33	34	35	36	37	38	39
20	21	22	23	24	25	26	27	28	29
10	11	12	13	14	15	16	17	18	19
0	1	2	3	4	5	6	7	8	9

Here is one bit muxing order. PSI = PMA Service I/F, CI = CAUI I/F. Others are ok also, rx must expect any lane to show up anywhere.

100GE PMA Variant #3

- In the transmit direction the following is provided:
  1. Bit level gearboxing of the 10 input lanes into a 4 output lanes
  2. Provide a clock source to the PMA client.
  3. Transmission of parallel data to the PMD.
- In the receive direction the following is provided:
  1. Reception of parallel data from the PMD
  2. Bit level gearboxing of the 4 input lanes into 10 output lanes
  3. Provides receive clock to PMA client
  4. Transmission of parallel data to PMA client.
  5. Provide link status information.



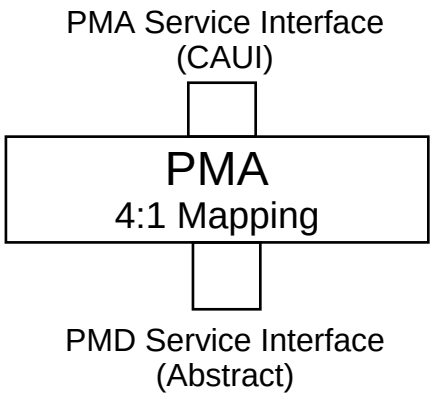
CI 0	CI 1	CI 2	CI 3	CI 4	CI 5	CI 6	CI 7	CI 8	CI 9
30	31	32	33	34	35	36	37	38	39
20	21	22	23	24	25	26	27	28	29
10	11	12	13	14	15	16	17	18	19
0	1	2	3	4	5	6	7	8	9

PI 0	PI 1	PI 2	PI 3
12	13	14	15
8	9	10	11
4	5	6	7
0	1	2	3

Here is one possible bit gearbox order. PI = PMD I/F, CI = CAUI I/F. Others are possible and supportable.

100GE PMA Variant #4

- Same as 40GE PMA variant #2, except for a faster speed
- In the transmit direction the following is provided:
  1. Bit level multiplexing of the 4 input lanes into a single output lane
  2. Provide a clock source to the PMA client.
  3. Transmission of serial data to the PMD.
- In the receive direction the following is provided:
  1. Reception of serial data from the PMD
  2. Bit level de-multiplexing of the serial data into four output lanes
  3. Provides receive clock to PMA client
  4. Transmission of parallel data to PMA client.
  5. Provide link status information.



PSI 0	PSI 1	PSI 2	PSI 3
8	9	10	11
4	5	6	7
0	1	2	3

Serial I/F
5
4
3
2
1
0

One of the possible bit muxing orders. PSI = PMA Service Interface.

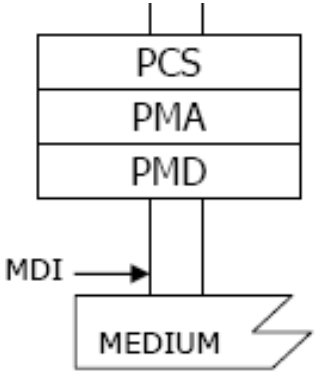
A Note on Bit Muxing Requirements

- All PCS receivers must support receiving a virtual lane on any physical lane
  - This allows flexibility now and in the future for how we bit mux and what widths of the interfaces we have today and tomorrow
- This means that there is more than one valid way to multiplex the virtual lanes at all stages
- All are supportable, with the requirement that:
  - When multiplexing from n to m lanes, any given virtual lane is always only sent on one physical lane, which particular lane does not matter
  - On each of M output lanes, every nth bit on a given physical lane must be a given Virtual Lane, where n = V/M (where V = number of total Virtual Lanes)
- With the above multiplexing rules, and with the requirement that the number of virtual lanes is the Least Common Multiple of all the to be supported lane widths, then everything works
  - For 100GE we can support any combination of lane widths of: 20, 10, 5, 4, 2, 1 with 20 VLs
  - For 40GE we can support any combination of lane widths of: 4, 2, 1 with 4 VLs

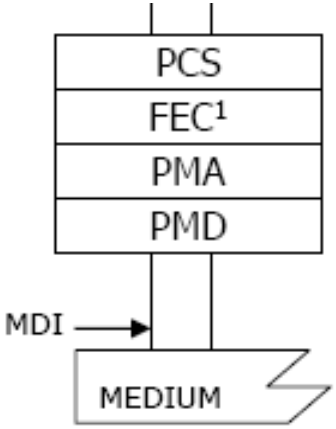
Key Differences from Clause 51 (802.3ae PMA)

- Parameterized Specification
  - Same specification covers both rates, input/output lane counts
- Layer Adjacency
  - In clause 51, PCS is above the PMA and PMD is below the PMA
  - For 802.3ba, above could be PCS, FEC, or another (stacked) PMA. Below could be another PMA or PMD. An appropriate naming scheme for primitives will be introduced to describe this
- Unidirectional Specification
  - Clause 51 documents bi-directional behavior (Tx direction from XSBI to PMD service interface, Rx direction from PMD service interface to XSBI)
  - 802.3ba will use a different instance of the single parameterized specification in each direction, e.g., 10:4 in the Tx direction and 4:10 in the Rx direction
- Independent bit arrival per lane
  - XSBI uses a vector of aligned lanes. For 802.3ba, Dynamic skew results in varying independent arrival of bits on each lane (even though all lanes originate at the Tx with the same clock)

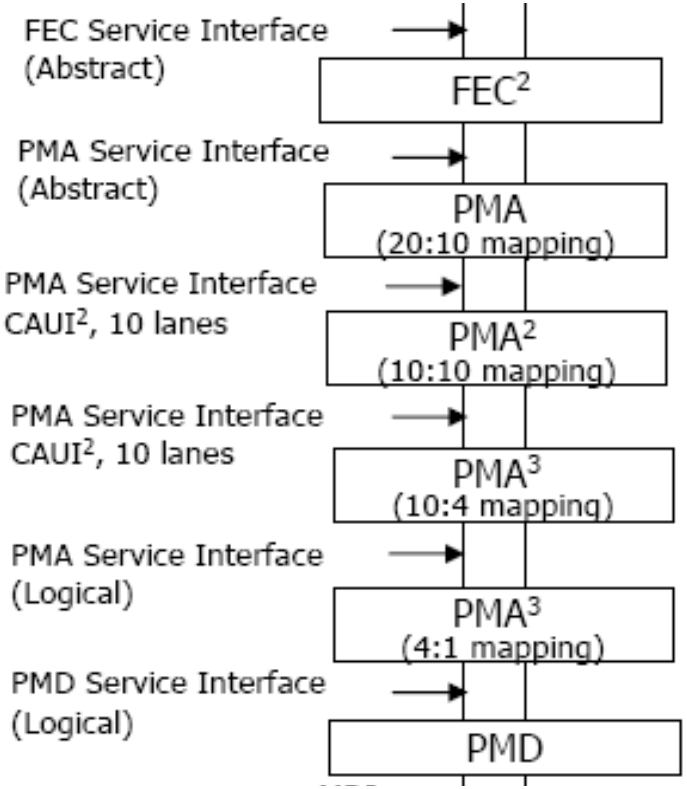
Layer Adjacency aspects



Classical:
-PCS above PMA
-PMD below PMA



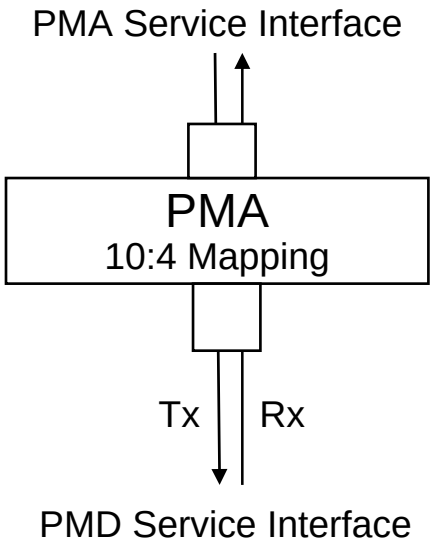
Variant –
- PMA client could be FEC rather than PCS



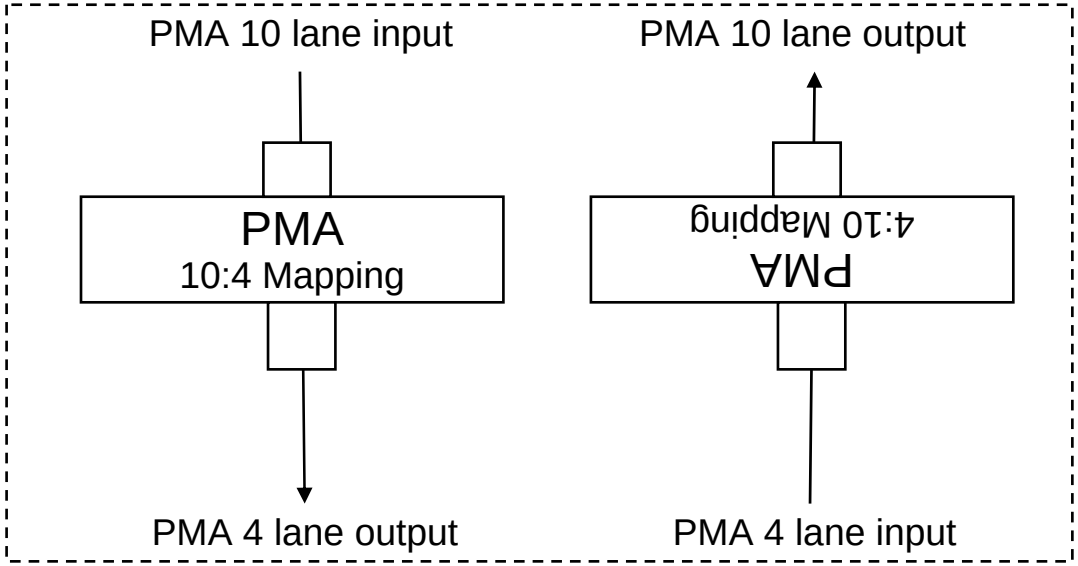
Variant –
- Can also have PMA as client or server for another PMA

Unidirectional Specification

Classical:
- Tx and Rx specified as different aspects of the same PMA



Instead we propose to describe in this way:



Possible Nomenclature

- ***r*PMA*d_n_m* is a PMA for rate *r*=C or XL with *n* input lanes ane *m* output lanes in direction *d* (T or R)**
- **CPMA*d_n_m* has the following characteristics:**
  - **20 VLs at 5.15625 Gbit/s ±100ppm**
  - ***n* and *m* are divisors of 20**
  - **Each of the *n* input lanes carries 20/*n* bit-muxed VLs at 103.125/*n* Gbit/s ±100ppm**
  - **Each of the *m* output lanes carries 20/*m* bit-muxed VLs at 103.125/*m* Gbit/s ±100ppm**
  - **in direction *d*=T, must tolerate 2ui of dynamic skew between VLs**
  - **in direction *d*=R, must tolerate 15ui of dynamic skew between VLs**
- **XLPMA*d_n_m* has the following characteristics:**
  - **4 VLs at 10.3125 Gbit/s ±100ppm**
  - ***n* and *m* are divisors of 4**
  - **Each of the *n* input lanes carries 4/*n* bit-muxed VLs at 41.25/*n* Gbit/s ±100ppm**
  - **Each of the *m* output lanes carries 4/*m* bit-muxed VLs at 41.25/*m* Gbit/s ±100ppm**
  - **in direction *d*=T, must tolerate 2ui of dynamic skew**
  - **in direction *d*=R, must tolerate 30ui of dynamic skew**

PMA Primitives

- ***r*PMA*d_n_m*_UNITDATA.request (*lane, bit*) is used to indicate the arrival of a bit with value *bit* on one of the *n* input lanes.**

  Note: in 802.3ae, UNITDATA.request would only come from the PMA client, but as we are using the same specification in both directions, it could be either a request from above in the transmit direction or, e.g., a PMD_UNITDATA_indication in the receive direction

- ***r*PMA*d_n_m*_UNITDATA.indication (*lane, bit*) is used to emit a bit with value *bit* on one of the *m* output lanes**

  Note: in 802.3ae, UNITDATA.indication would only be sent to the PMA client above, but as we are using the same specification in both directions, it could either be an indication to the PMA client, or a bit sent to the server below, e.g., a PMD_UNITDATA.request in the transmit direction

PMA primitives - continued

- ***r*PMA*R*_*n*_*m*_SIGNAL.request (SIGNAL_OK) is used to indicate whether the received signal from the server layer is OK**

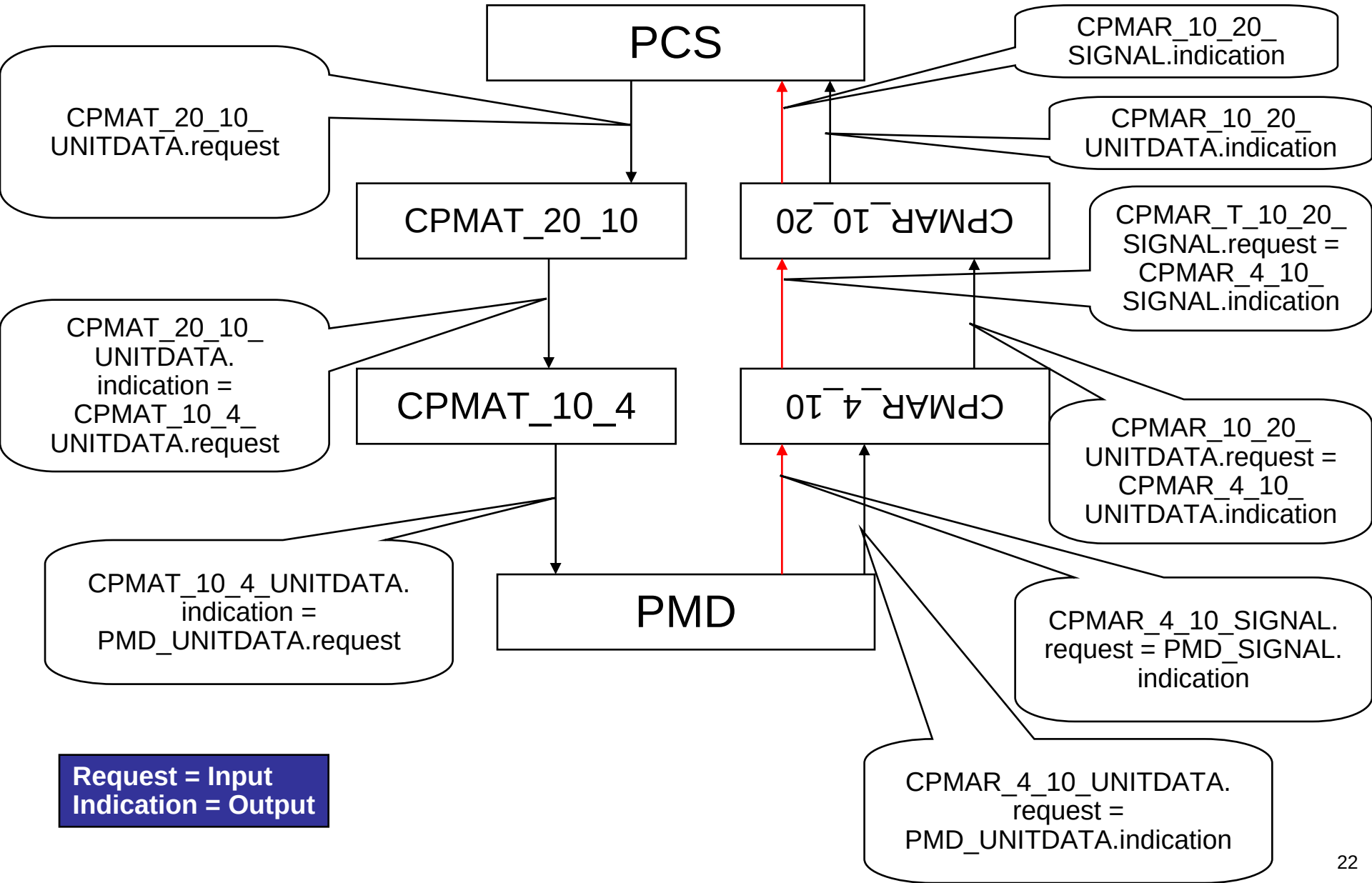
  Note: in 802.3ae, this would be a PMD_SIGNAL.indication, but for 802.3ba it could come from a PMD or another stacked PMA. Should we define this signal in both directions? 802.3ae only defines in receive direction. Initial assumption is that this only applies in the receive direction.

- ***r*PMA*R*_*n*_*m*_SIGNAL.indication (SIGNAL_OK) is used indicate to the PMA client whether the received signal is OK**

  Semantics are similar to 802.3ae, PMA_SIGNAL.indication if we only define this in the receive direction. The signal is indicated as OK if both SIGNAL_OK is being received from the server layer below AND data is being successfully recovered by the PMA on all lanes, PMA fifos are in limits, and therefore data is being transmitted on the lanes of the PMA output

Nomenclature/Primitives Example

Same interface may be known by different names by sender/receiver



Other PMA Aspects

Given that PMA is a logical interface, these next several items could be left as exercises for the designer. But they could be addressed at a high level in the standard.

- **PMA partitioning: If *n* and *m* share any common factors, an implementation may partition the PMA into several smaller PMAs. For example, a CPMAT_20_10 could be partitioned into ten CPMAT_2_1s, or CPMAT_10_4 could be partitioned into two CPMAT_5_2s. Then:**
  - **Virtual lane rearrangement input to output is only within a partition (although input and output lanes might be rearranged, e.g., for routing convenience knowing that the receive logic is general)**
  - **Clocking architecture is local to each partition**
  - **FIFO management for dynamic skew compensation is local to a partition**
  - **If m and n are equal, the partition size is one**

Other PMA Aspects - continued

- **Clocking Architecture – Several options based on context:**
  - **Single reference clock, e.g., for a transmit direction PMA implemented synchronously in the same device as the PCS**
  - **Per-input lane clock recovered in an upstream layer implemented in the same device**
  - **Per-input lane CDR in the PMA itself**
- **The PMA output clock is derived from the reference clock or the input clock on one of the input lanes (within a partition) using an m/n clock multiplier/divider circuit.**

Other PMA Aspects - Continued

- **Dynamic skew compensation and FIFO management**
  - **Amount of skew to be tolerated and/or compensated depends on context:**
    - **A transmit side PMA implemented synchronously in the same device as the PCS may not experience any dynamic skew**
    - **Skew compensation only required within a partition. If m and n are equal, no skew compensation is required since the partition size is one**
    - **2ui per VL dynamic skew tolerance in transmit direction**
    - **15ui per VL dynamic skew tolerance in receive direction (100 GbE)**
    - **30ui per VL dynamic skew tolerance in receive direction (40 GbE)**
  - **Requirement is that after startup, bit rotation order is maintained on each output lane as long as dynamic skew budget is not exceeded on input lanes**

Other PMA Aspects - Continued

- **Dynamic skew compensation implementation possibilities**
  - **FIFOs/buffers are needed between the PMA input and output to compensate dynamic skew (within a partition) where m and n are not equal. Equivalent implementations could use appropriately sized input lane FIFOs, output lane buffers, or per VL FIFOs. The FIFO/buffer depth (per VL) needs to be double the dynamic skew tolerance amount as the clock might be derived from a lane that is leading or lagging**
  - **The PMA output (within a partition) doesn't start until all input lanes have recovered clock and data and FIFOs/buffers (within the partition) are centered.**

Summary

- PMA Functions include:
  - Clock and data recovery
  - Bit level multiplexing/gearboxing
  - Clock generation
  - Signal drivers
- Key differences from Clause 51 (802.3ae) PMA description
  - Unified parameterized specification covers all rates and input/output lane counts
  - Matching interface descriptions may have different names at sending and receiving layers to allow layer stacking flexibility
  - Uni-directional specification uses same text to describe Tx and Rx directions
  - Independent bit arrival per lane (even though lanes originate from common Tx clock) rather than a bus (e.g., XSBI) operating on a common clock

XLAUI/CAUI Electrical Specifications

IEEE 802.3ba
Denver 2008

July 15 2008

Ali Ghiasi
Broadcom Corporation
aghiasi@broadcom.com

Ryan Latchman
Gennum Corporation
ryan.latchman@gennum.com

List Supporters

- Mike Peng - Altera
- Kieth Conroy – AMCC
- Francesco Caggioni – AMCC
- John Petrilla – Avago Tech
- Mark Gustlin – Cisco
- Eddie Tsumura - Excelight
- Chris Cole - Finisar
- Jim Tavacoli - Santur
- Frank Chang - Vitesse

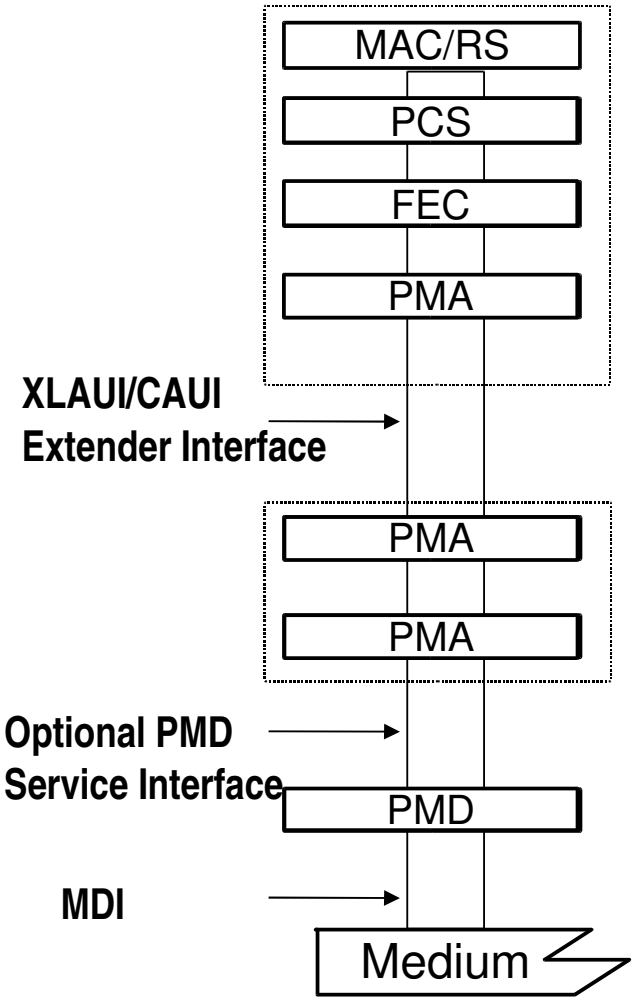
Overview

- XLAUI/CAUI interface
- Optimized electrical interface for XLAUI/CAUI
- Channel simulation
- Channel measurement
- Jitter transfer

The proposed XLAUI/CAUI are not final and are subject to the IEEE review process.

XLAUI/CAUI Interface

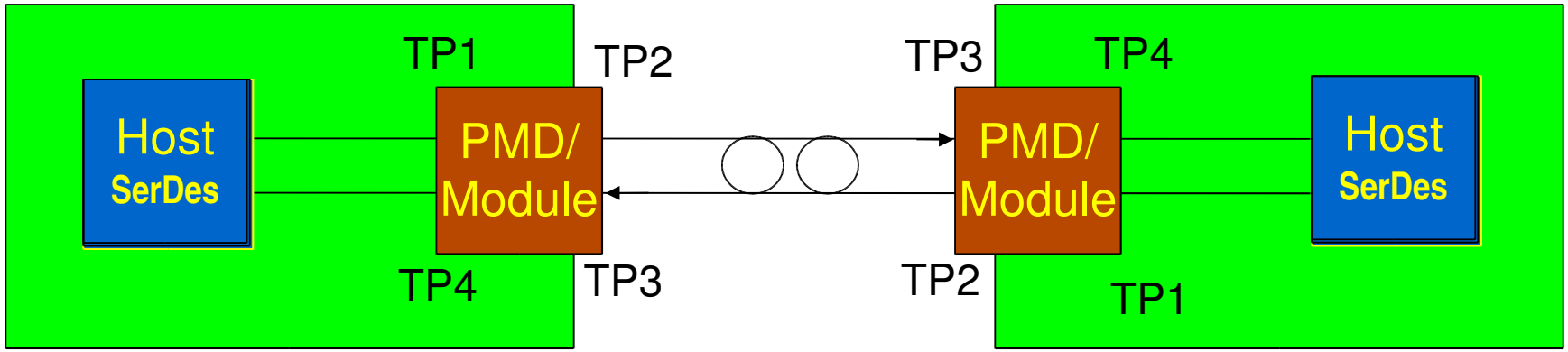
- Simple, low power chip to chip or chip to module interface.
  - Simplifies ASIC SerDes by not requiring to support TP1/TP4 requirements for all PMD's.
- Retimed interface with relax jitter budget and ASIC friendly.
- Take advantage of pre-emphasis to increase PCB loss with simple CDR receiver.
- Operate over ~250 mm FR4-8 stripline or ~375 mm FR4-13 stripline.
- XLAUI/CAUI will be the bolting point for future electrical interface based on 25 GBd/lane.



see ganga_01_0508.pdf for XLAUI and CAUI layer definition

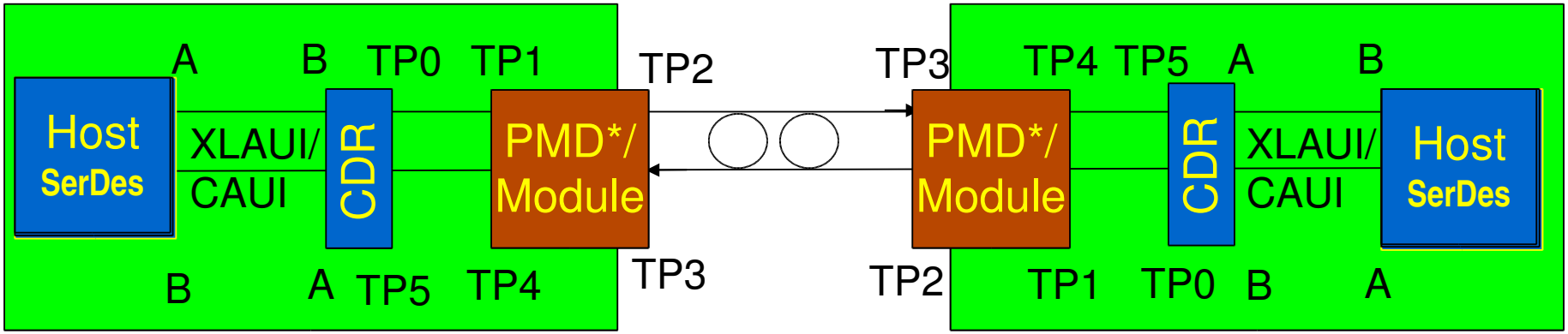
Application Not Using XLAUI/CAUI

- **ASIC capable of interfacing with the PMD are not required to use the XLAUI/CAUI retimer, see petrilla_01_0508.pdf.**
- **Likely scenarios are:**
  - **A simplified ASIC supporting single PMD type**
  - **ASIC SerDes support all PMDs when the technology is mature and there is little power penalty (i.e. SFP+ now).**

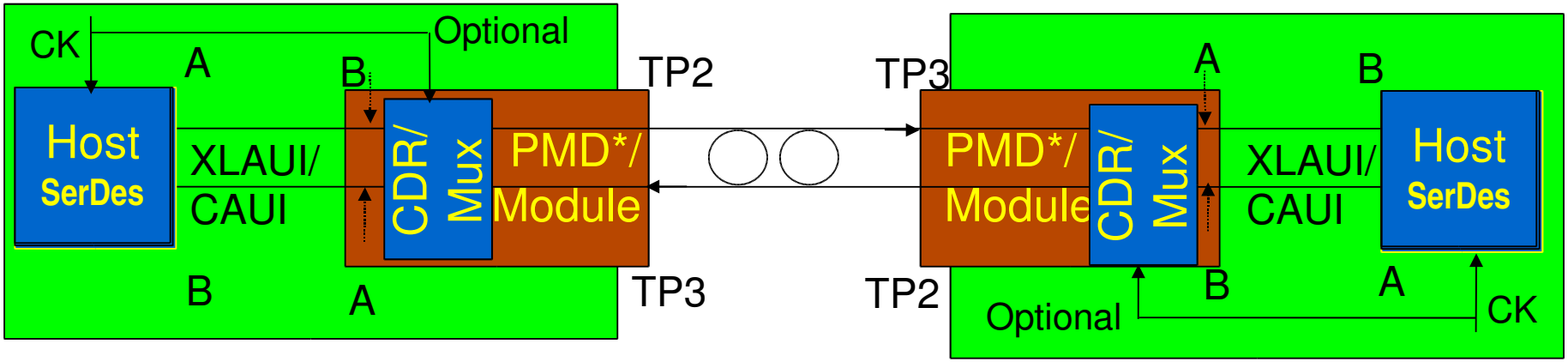


Application XLAUI/CAUI Extender for Front Ports (PMD nx10Gbaud)

- **Application with CDR on the host PCB (QSFP/CSFP)**

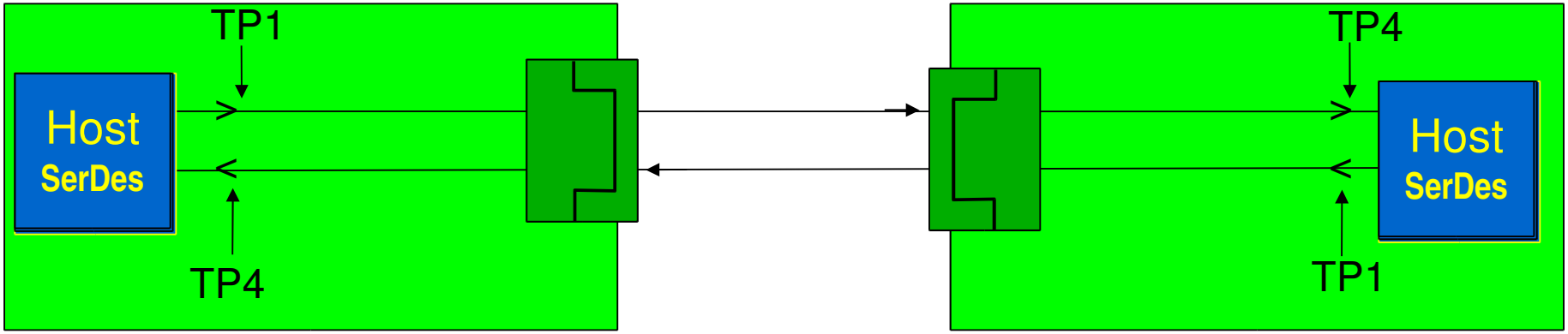


- **Application with CDR or Mux/De-mux in the module (QFP/CFP) with optional XLAUI/CAUI clock**

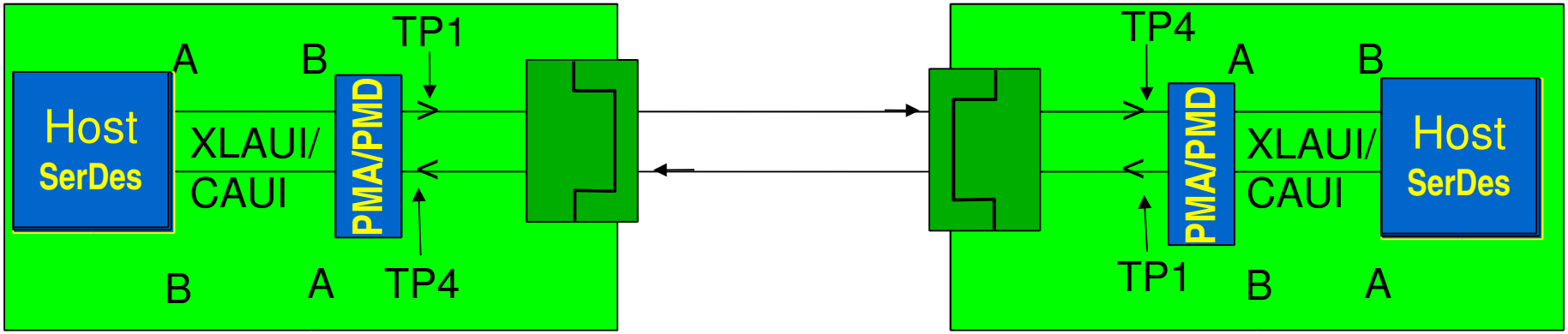


XLAUI/CAUI Extender for KR Application Ports

- **KR application without XLAUI/CAUI Retimer**

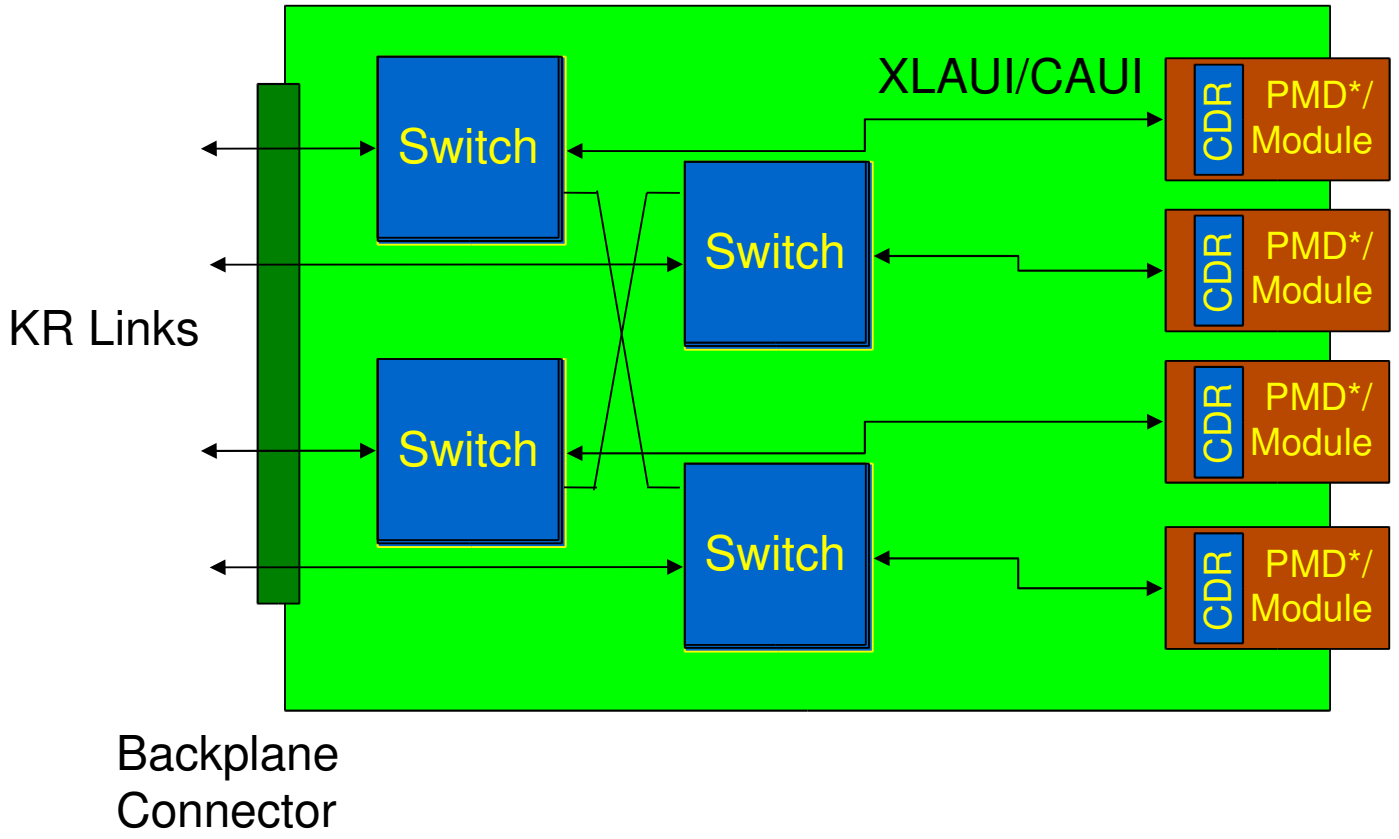


- **KR application with XLAUI/CAUI Retimer**



Linecard Application of XLAUI/CAUI

- **Typical CAUI implementation can be supported with 250 mm on FR4.**
- **In the implementation shown below 375 mm on improved FR4 may be required.**

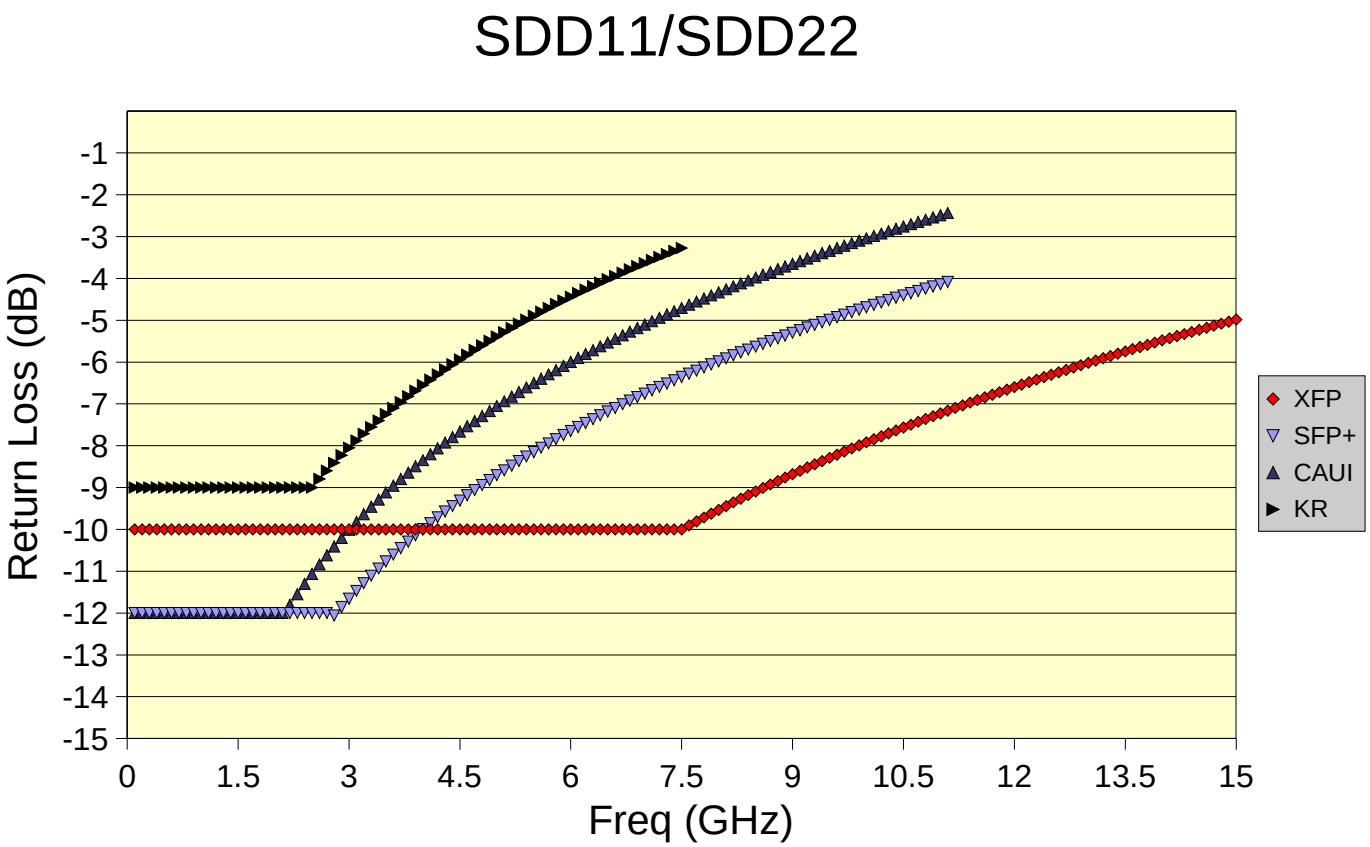


XLAUI/CAUI Electrical Interface

- **XFI has very stringent SDD return loss -10 dB up to 7.5 GHz.**
  - **Flat return loss up to 7.5 GHz is not realistic, too tight at high frequency, but too loose at the low frequency!**
  - **The proposed XLAUI/CAUI specifications is based on SFF-8431 with corner frequency pulled back to 2.125 GHz from 7.5 GHz but with improved return loss up to 2.125 GHz.**
- **XFI has very stringent common mode return loss -6 dB up to 15 GHz.**
  - **The proposal here is for SCC to follow SDDxx – 3 dB.**
  - **SCC for the receiver is not required as it limits the implementation.**
- **XFI and SFI allocate only 6 dB of channel loss at ½ the baudrate which has either limited the host PCB and/or require improved FR4**
  - **The proposed channel increases the loss at ½ the Baudrate to 10 dB for more flexible PCB design supporting ~250 mm on FR4 (Isola FR4-8) or ~375 mm on improved FR4 (Nelco N4000-13).**

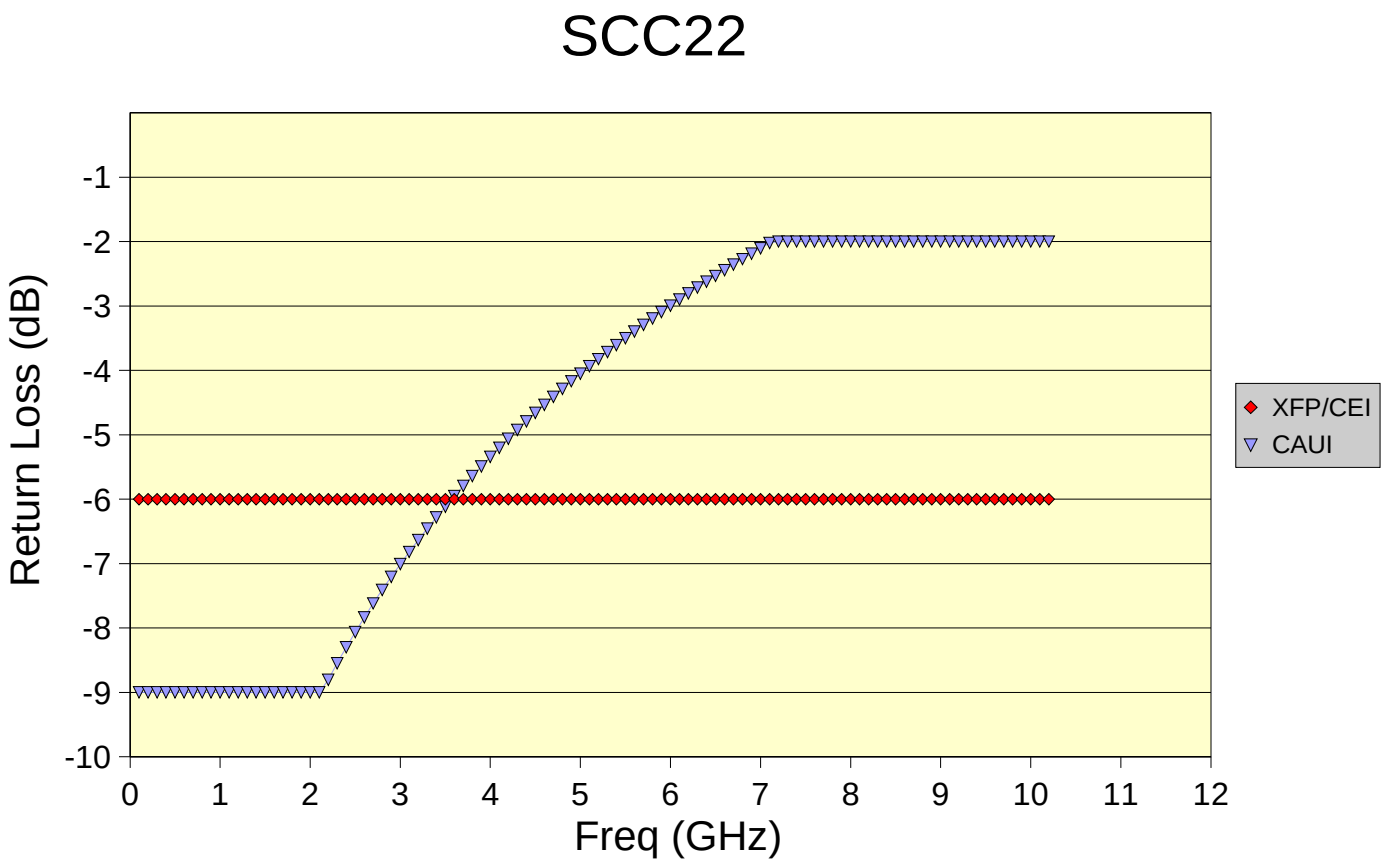
XFI, CEI, SFP+, CAUI Return Losses

- **Physical limitation of the IC parasitics makes it difficult to meet XFI return loss at high frequency but low frequency too relaxed.**
  - **This proposal uses 8.5 SFP+ Host return loss s4p available asT11-838v0.**
    - **-12 dB up to 2.125 GHz. -6.5 + 13.33LOG10(f/5.5) from 2.125 GHz to 11.1 GHz.**



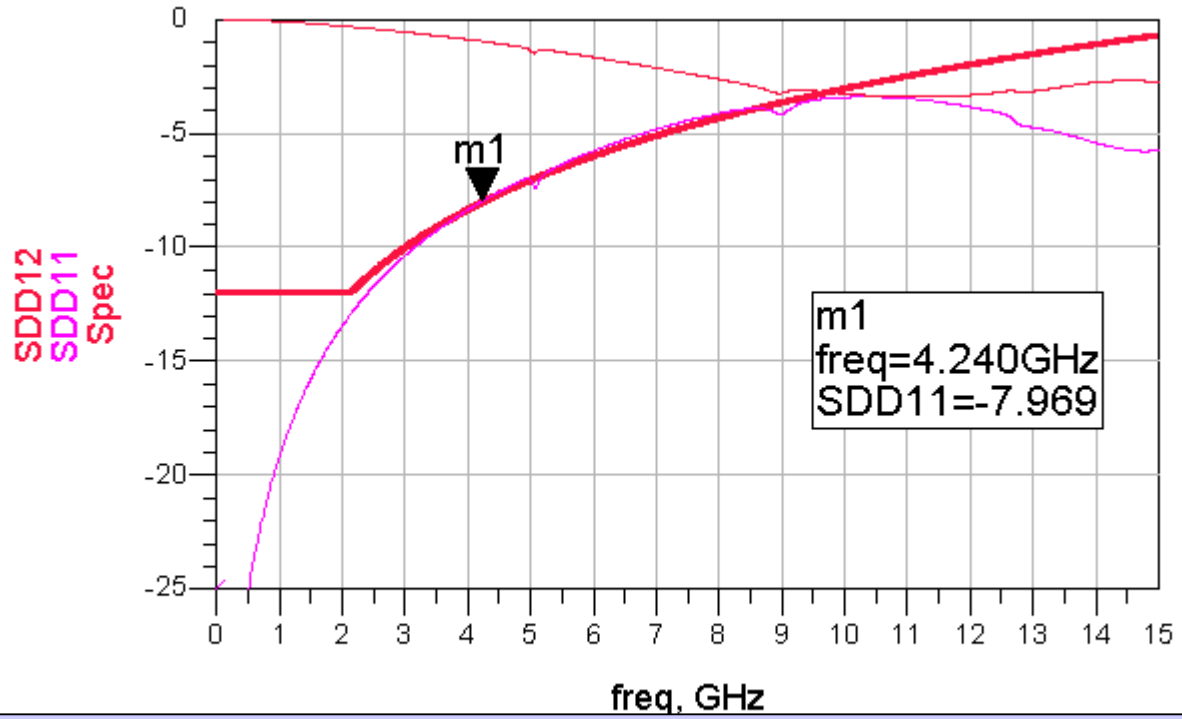
XFI, CEI, and CAUI/XLAUI RL

- **XFP/CEI common mode more difficult than SDD and not practical!**
- **SFP+ defines SCC to follow the SDD mask but 3 dB worse:**
  - **-9 dB from 0.1 to 2.125 Ghz, (-3.5 + 13.333LOG10(f/5.5)) from 2.125 to 7.1 Ghz, -2 from 7 to 11.1 GHz.**



RX/TX Chip Return Loss

- **CAUI SDDxx mask overlayed on top of the 8.5Gig SFP+ Host return loss, s4p file is available from T11 website as T11-838v0.**



SDD Mask=-12 if <2.125 GHz else -6.5 + 13.33*LOG10(f/5.5), where f is in GHz

XLAUI/CAUI Channel Loss Budget

- XFI, SFP+, and CEI 11G-SR operate only over ~150 mm of FR4-8 (Isola Fr4-8) stripline
- The proposed channel loss for XLAUI/CAUI is 10 dB at Nyquist, with following estimated PCB trace reach:
  - About 250 mm on FR4 (Isola FR4-8)
  - About 375 mm on improved FR4 (N4000-13)

Parameter	Channel Loss @ 5.15 GHz
Channel Loss (SDD21) Including one Connector	10 dB *
Reflection and other penalties	2.5 dB
Total Loss	12.5 dB
* SDD21 = -0.144 – 1.323*SQRT(f) – 1.333*(f/1e9), where f is given in GHz.	

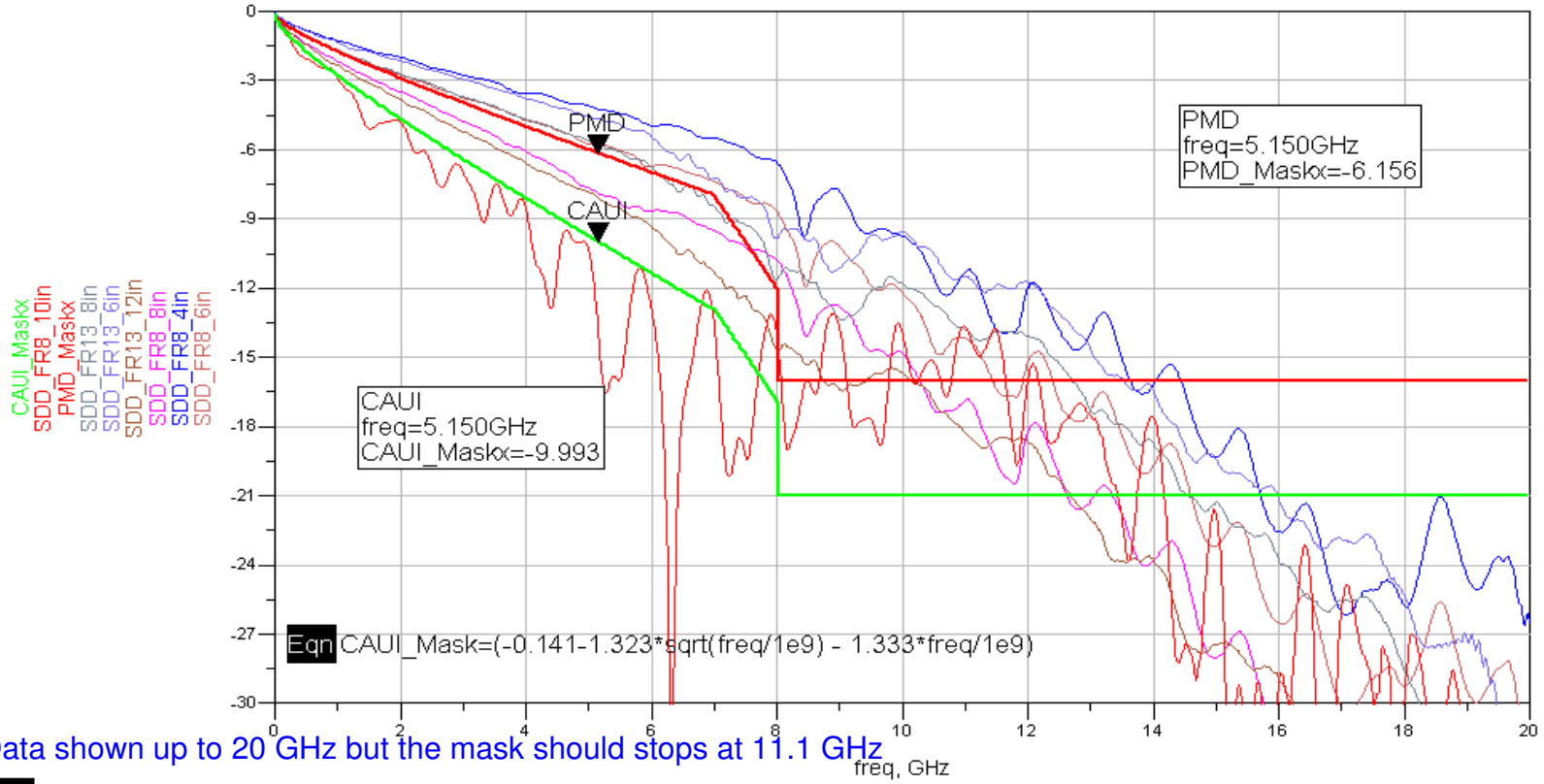
PCB Trace Reach

- **Current proposal allows for 250 mm of Isola FR4-8 or 375 mm of Nelco N4000-13 stripline traces with one connector.**
  - **Use of transmit pre-emphasis and relaxed far end jitter allows increasing the channel loss budget compare to SFI or petrilla_02_0508.**
  - **All channels are routed on lower stripline with two short stubs ~ 13 mils.**

Interface	Isola FR4-8 *	N4000-13 **	Loss at Nyquist	Host PCB Loss at Nyquist 2	Relative to XFI
XFP/SFI	150 mm	200 mm	6 dB	4.500	0.00%
petrilla_02_0508	150 mm	200 mm	6 dB	4.500	0.00%
XALUI/CAUI Proposal	250 mm	375 mm	10 dB	8.500	189.00%
*. Assumes 5 mils moderately coupled ~ 7% wide 0.5 oz striplines					
**. Connector loss and HCB loss subtracted					

XLAUI/CAUI Channel SDD21 (Informative)

- **XLAUI/CAUI supports about 250 mm of FR4-8 or about 375 mm of 5.5 mils FR4-13 striplines**
  - **The 10 dB channel was created by cascading 2nd PCB with 2 dB loss at Nyquist with the 8" Fr4-8 channel which is adding some ripple.**



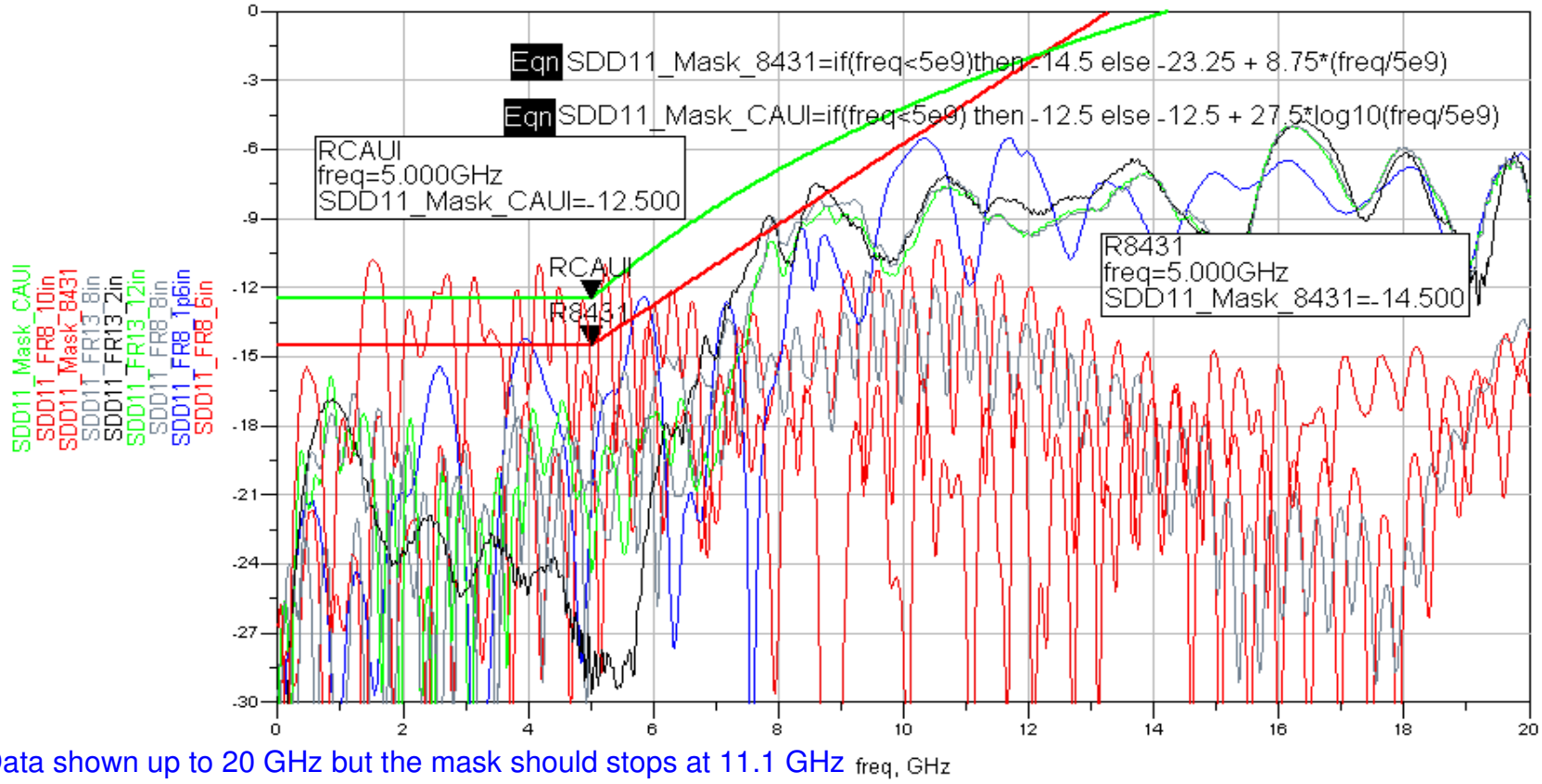
Data shown up to 20 GHz but the mask should stops at 11.1 GHz

Eqn CAUI_Maskx=if(freq<7e9) then (-0.141-1.323*sqrt(freq/1e9) - 1.33*freq/1e9) elseif (freq<=8e9) then 15.1-4.*freq/1e9 else -21 endif

Eqn PMD_Maskx=if(freq<7e9) then (-0.108-0.845*sqrt(freq/1e9) - 0.802*freq/1e9) elseif (freq<=8e9) then 20-4*freq/1e9 else -16 endif

XLAUI/CAUI Channel SDD11(Informative)

- **The CAUI informative channel SDD11/SDD22 is ~ 2 dB more relaxed than SFF-8431.**
  - **The cascaded channel with 10 dB loss at Nyquist its SDD11 is degrades about 3 dB.**



Data shown up to 20 GHz but the mask should stops at 11.1 GHz

XLAUI/CAUI Transmitter Electrical Specifications (Point A)

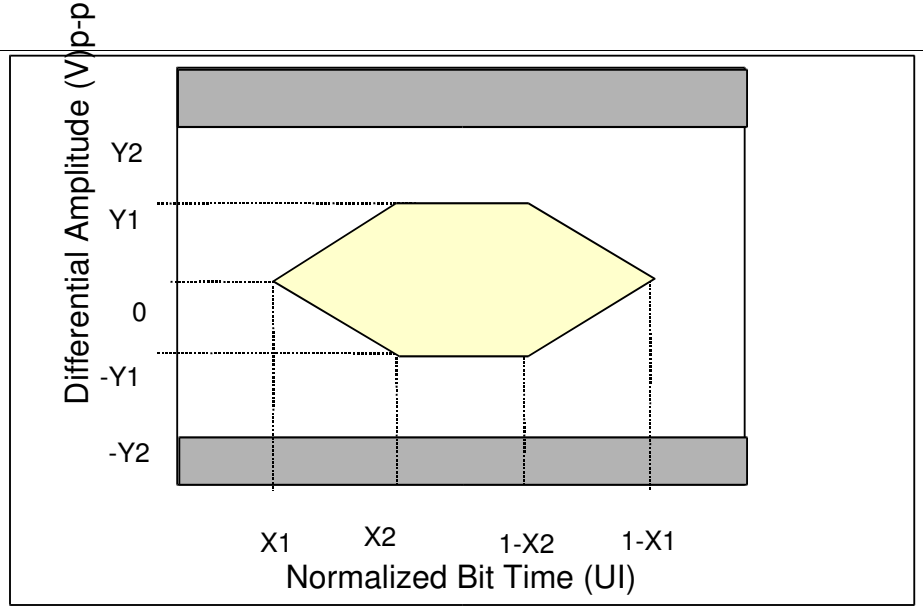
- **Starting with SFF-8431 specification at A**

Parameter	*Symbol*	*Conditions*	*Min*	*Typ*	*Max*	*Units*
Differential Output Voltage, p-p	Vdiff				see [1]	mV
Termination Mismatch		at 1 MHz			5	%
Output AC Common Mode Voltage (RMS)	Vcm				15	mV
Output Rise and Fall time (20% to 80%)	t_{RH}, t_{FH}		24			ps
Differential Output S-parameters	SDD22	0.01 to 2.125 GHz			-12	dB
		2.125-11.1 GHz			*[2]	dB
Common Mode Output S-parameters	SCC22	.01-2.125 GHz			-9	dB
		2.125-7.1 GHz			* [3]	dB
		7.1-11.1 GHz			-2	dB

1. **Must meet eye mask parameter Y1 and Y2.**
2. **SDD22(dB) = -6.5 + 13.33*LOG10(f/5.5), f is given in GHz**
3. **SCC22(dB)= -3.5 + 13.33*LOG10(f/5.5), f is given in GHz**

XLAUI/CAUI Transmit Eye Mask (Point A)

Parameter	*Symbol*	*Conditions*	*Min*	*Typ*	*Max*	*Units*
Determinstic Jitter					0.17	UI
Total Jitter	TJ				0.32	UI
Eye Mask	X1				0.16	UI
Eye Mask	X2				0.38	UI
Eye Mask	Y1				190	mV
Eye Mask	Y2				380	mV



XLAUI/CAUI Receiver Electrical Specifications (Point B)

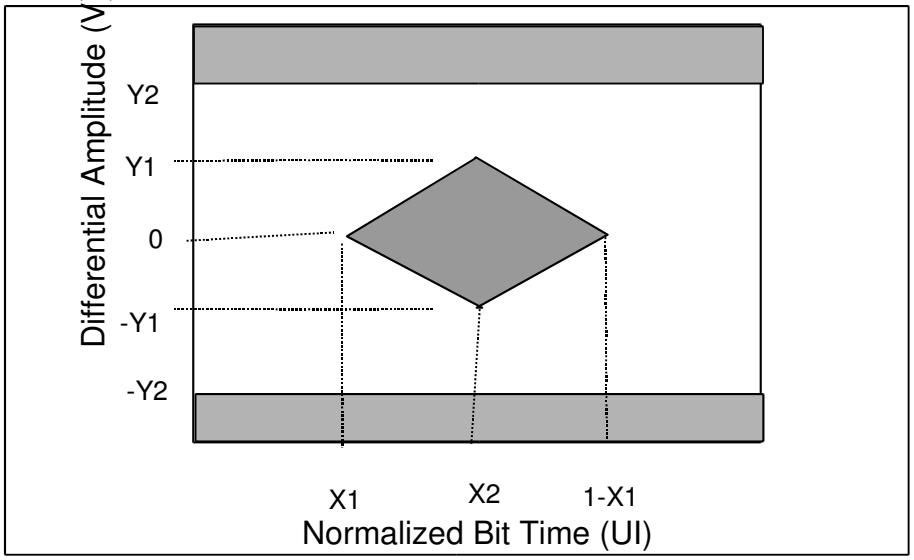
- **Starting with SFF-8431 Receiver specification**

Parameter	*Symbol*	*Conditions*	*Min*	*Typ*	*Max*	*Units*
Differential Input Voltage , Differential p-p	Vdiff	see 1	90		850	mV
Input AC Common Mode Voltage (RMS)	Vcm		20			mV
Input Rise and Fall time (20% to 80%)	t_{RH}, t_{FH}		24			ps
Differential Input S-parameters	SDD11	0.05 to 2.125 GHz			-12	dB
		2.125-11.1 GHz			*[2]	dB
Differential to Common Mdoe Input Conversion S-parameters	SCD11	0.01-11.1 GHz			-15	

1. **Max value is 850 mV for compatibility with TP4 see petrila_01_0508.pdf**
2. **SDD22(dB) = -6.5 + 13.33*LOG10(f/5.5), f is given in GHz.**

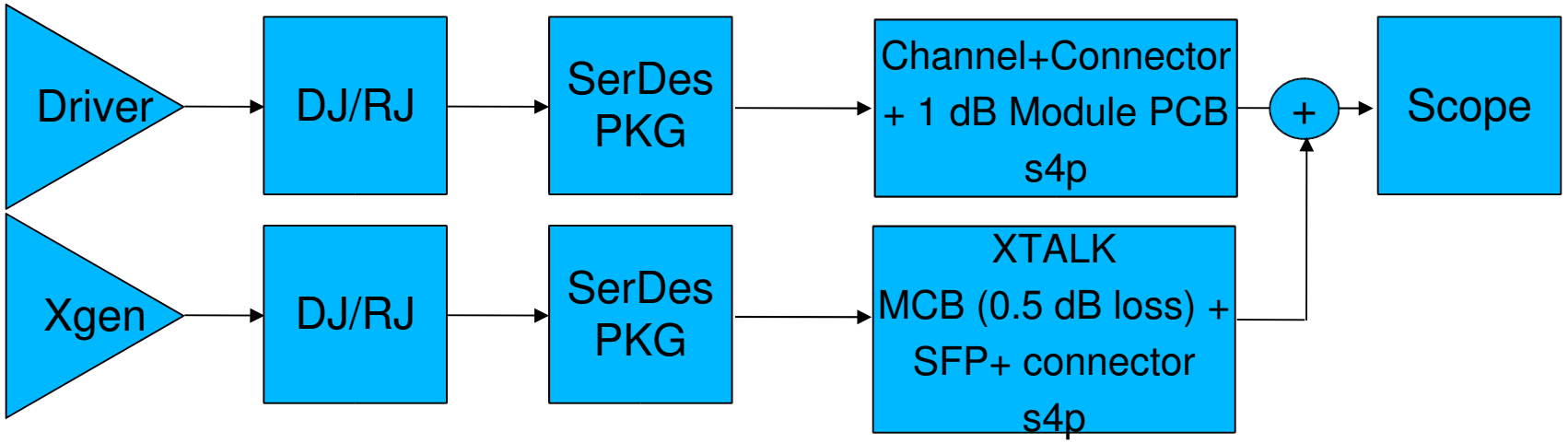
XLAUI/CAUI Receive Eye Mask Specifications

Parameter	*Symbol*	*Conditions*	*Min*	*Value*	*Max*	*Units*
non-EOI Jitter (TJ – ISI)	DJ	Corner Frequncy>4 MHz			0.42	UI
Total Jitter	TJ	Corner Frequncy>4 MHz			0.62	UI
Eye Mask	X1				0.31	UI
Eye Mask	X2			0.5		UI
Eye Mask	Y1		45			mV
Eye Mask	Y2				425*	mV
*same as SFP+						

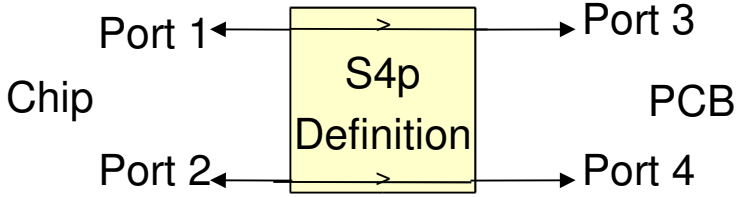


Link Simulation Set-up

- **Host SerDes package model was cascaded with the channel**
  - **The crosstalk source Xgen amplitude was set to 3x the driver to account for additional XTALK due to multi-aggressor and/or connectors possibly worse than SFP+.**

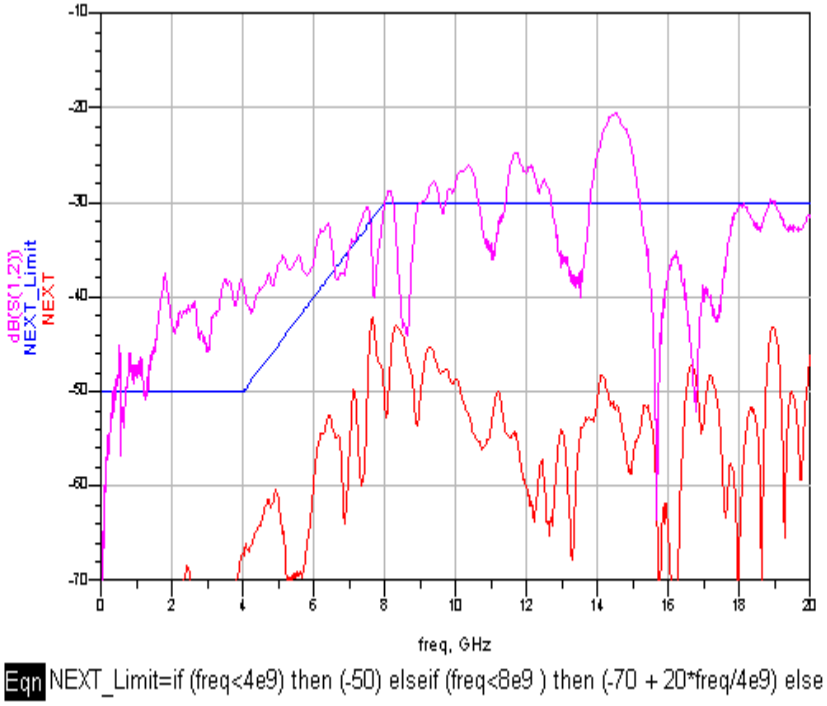


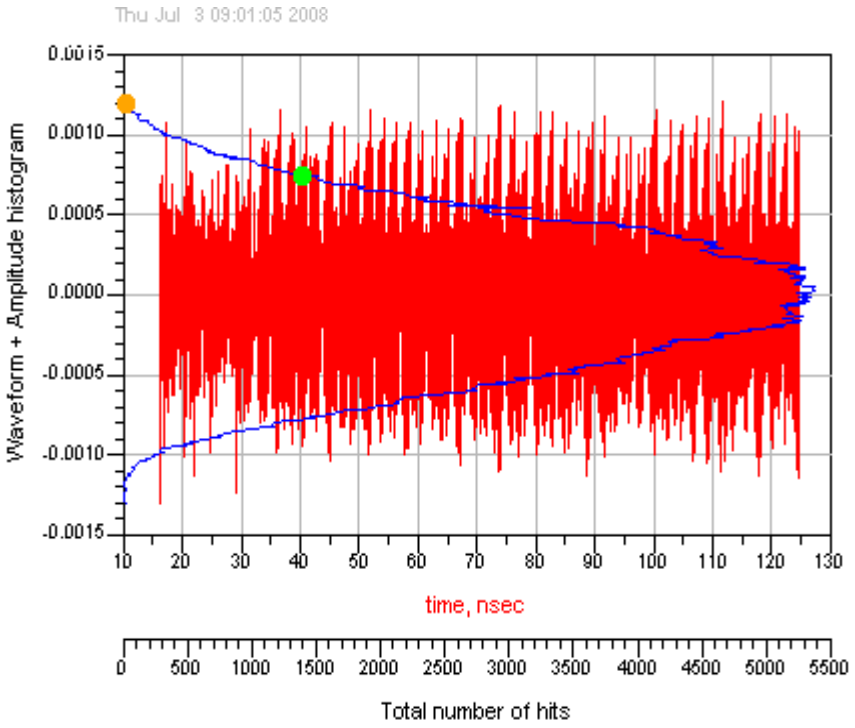
S4P File Definition for Through Channel



Connector XTALK

- **Simulation here uses 3x the amplitude with SFF-8083 (SFP+) connector till more suitable Xtalk data available.**

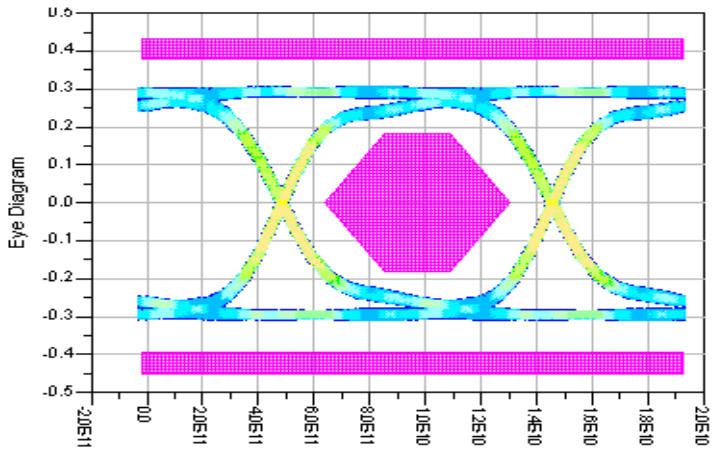




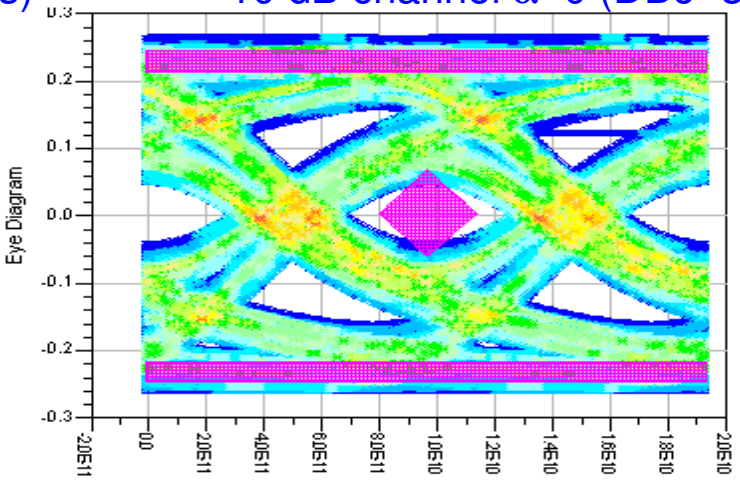
Meeting Far End CAUI/XLAUI Mask with Single Pre-emphasis Setting

- **TX launch was 600 mV pk-pk differential**

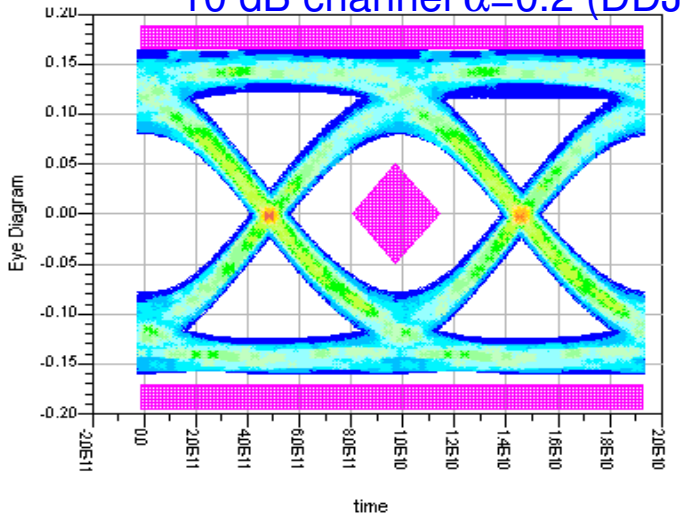
Near End Eye α=0 (DDJ=0.8 ps)



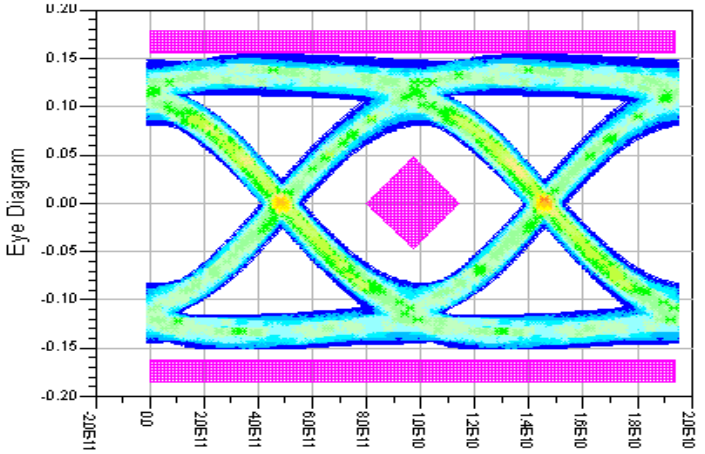
10 dB channel α=0 (DDJ=39.6 ps)



10 dB channel α=0.2 (DDJ=12.7 ps)



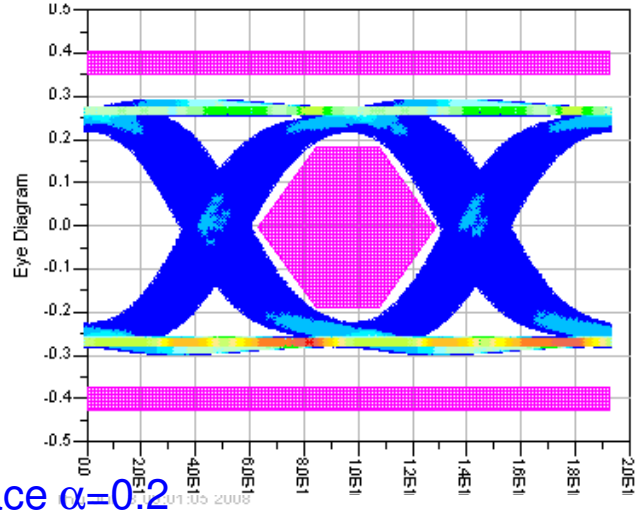
10 dB channel α=0.24 (DDJ=12.5 ps)



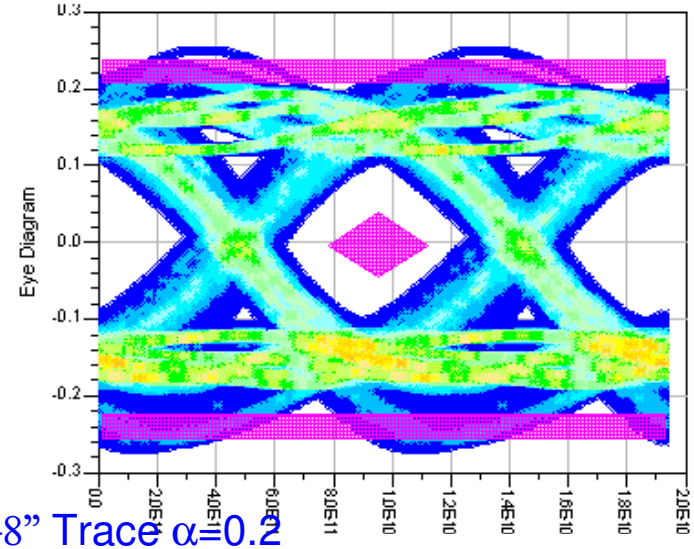
Meeting Eye Mask with Maximum Transmitter DJ and RJ

- **TX launch was 550 mV pk-pk differential**

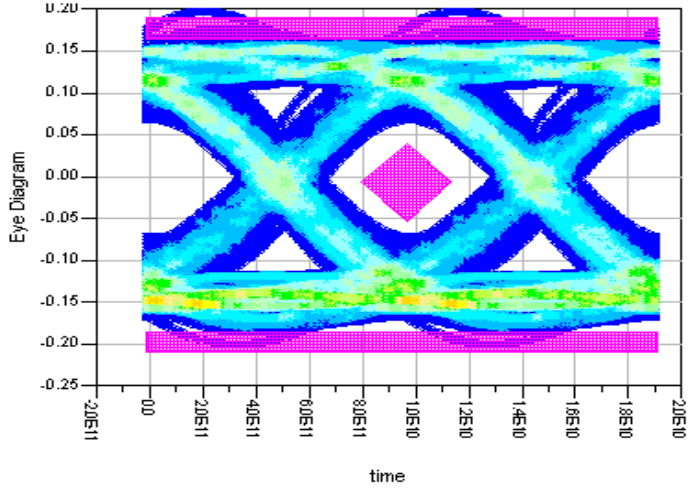
Near End Eye α=0



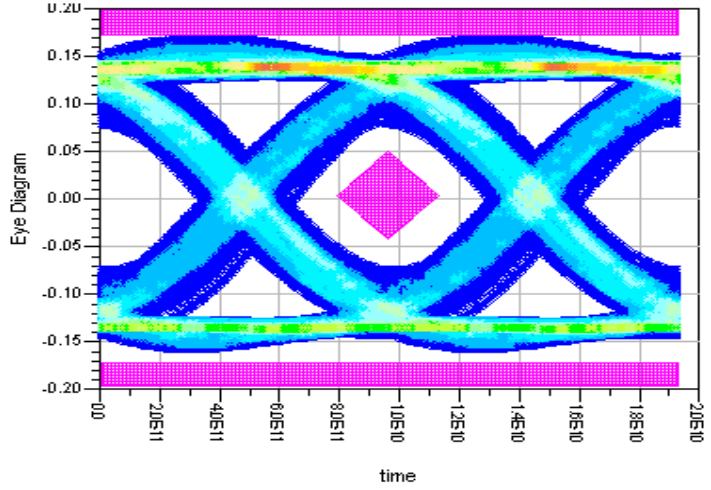
2” Trace α=0.2



6” Trace α=0.2



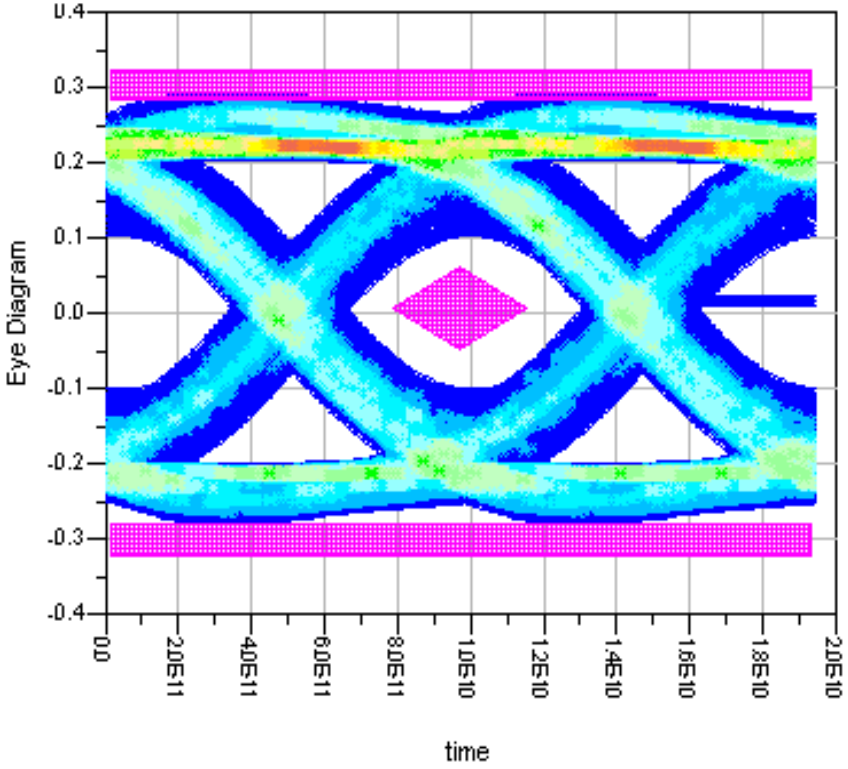
2”+8” Trace α=0.2



Meeting Eye Mask with Maximum Transmitter DJ and RJ

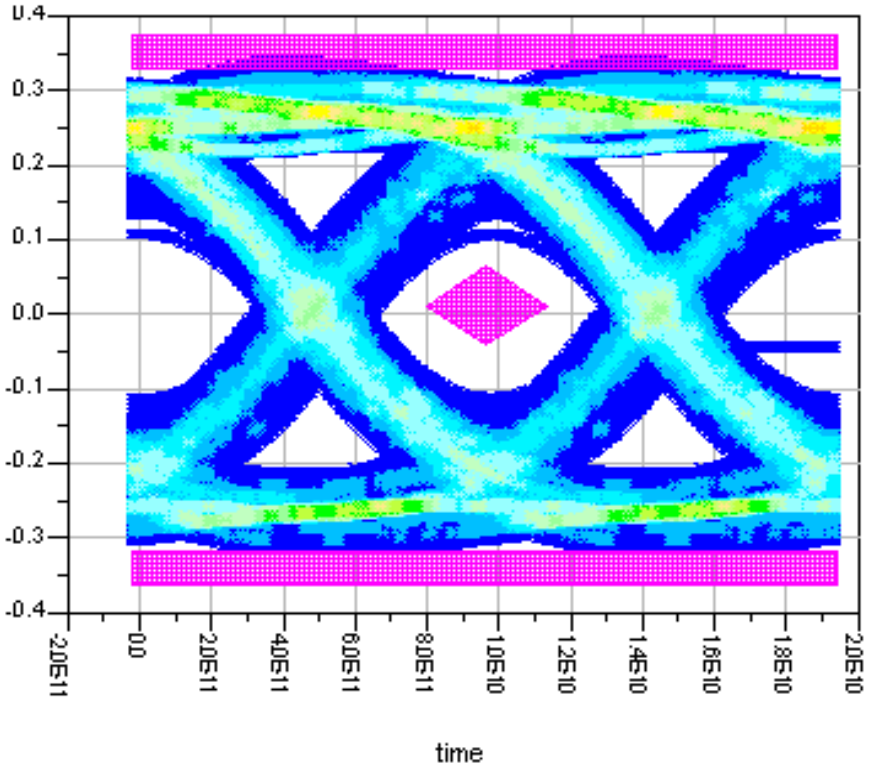
- **TX launch was 550 mV pk-pk differential**

2"+8" Channel XTALK off α=0.2



TJ ~ 0.45 UI

2"+8" Channel XTALK on α=0.2



TJ ~ 0.52 UI

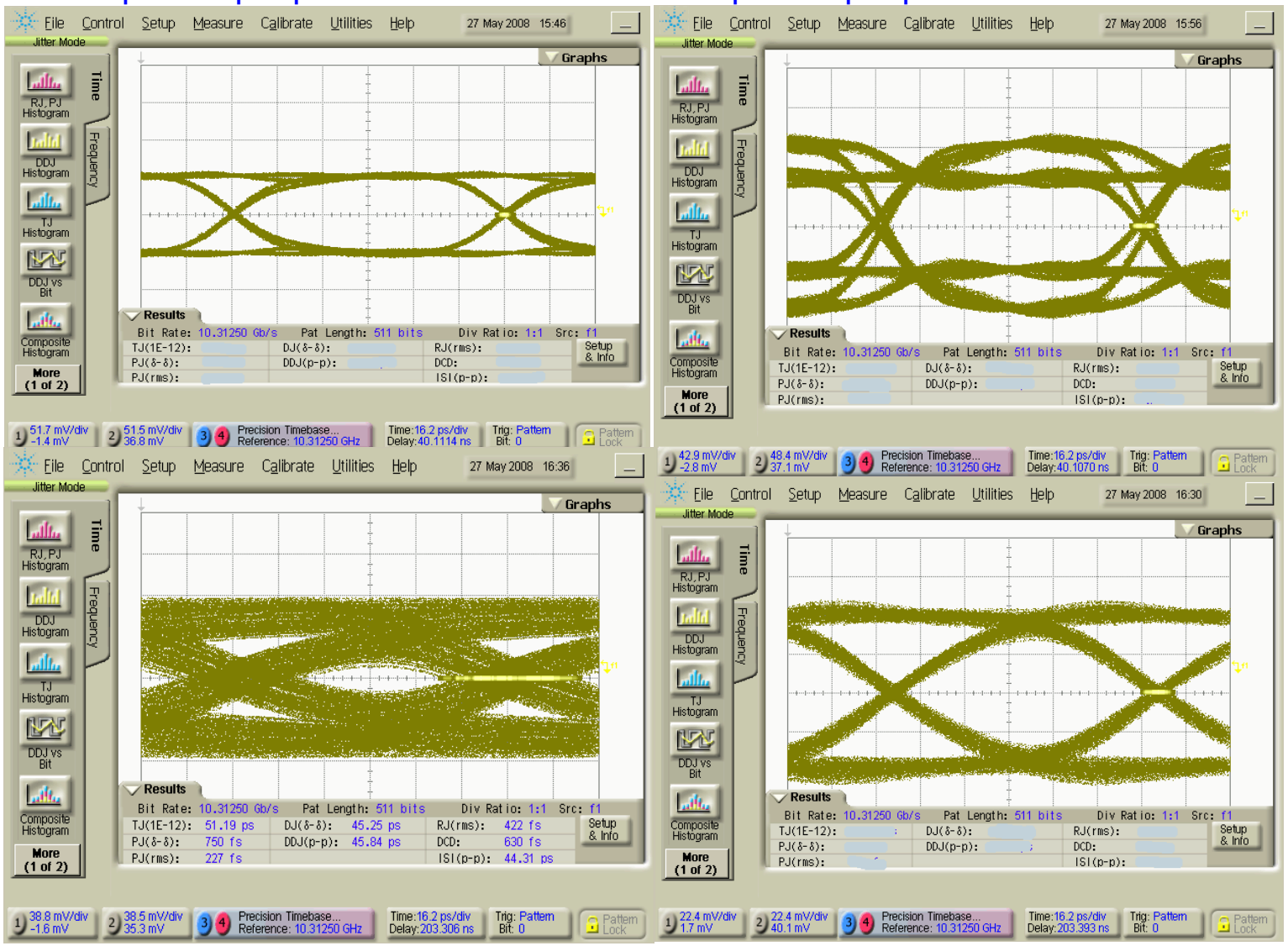
Output Eye Digram for Near and Far End

pre-emp Optimum for Near End

pre-emp Optimum for Far End

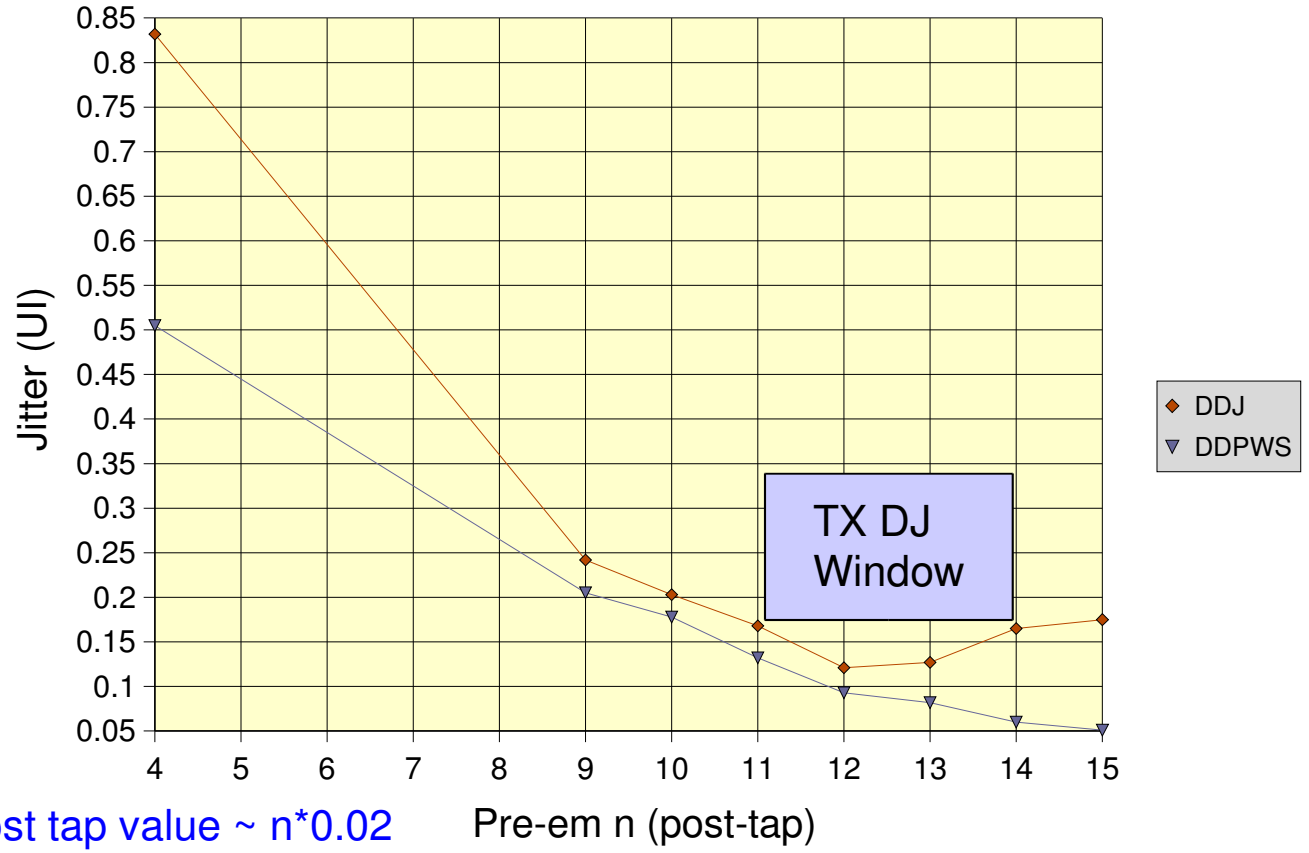
Chip Out

8” FR4-8



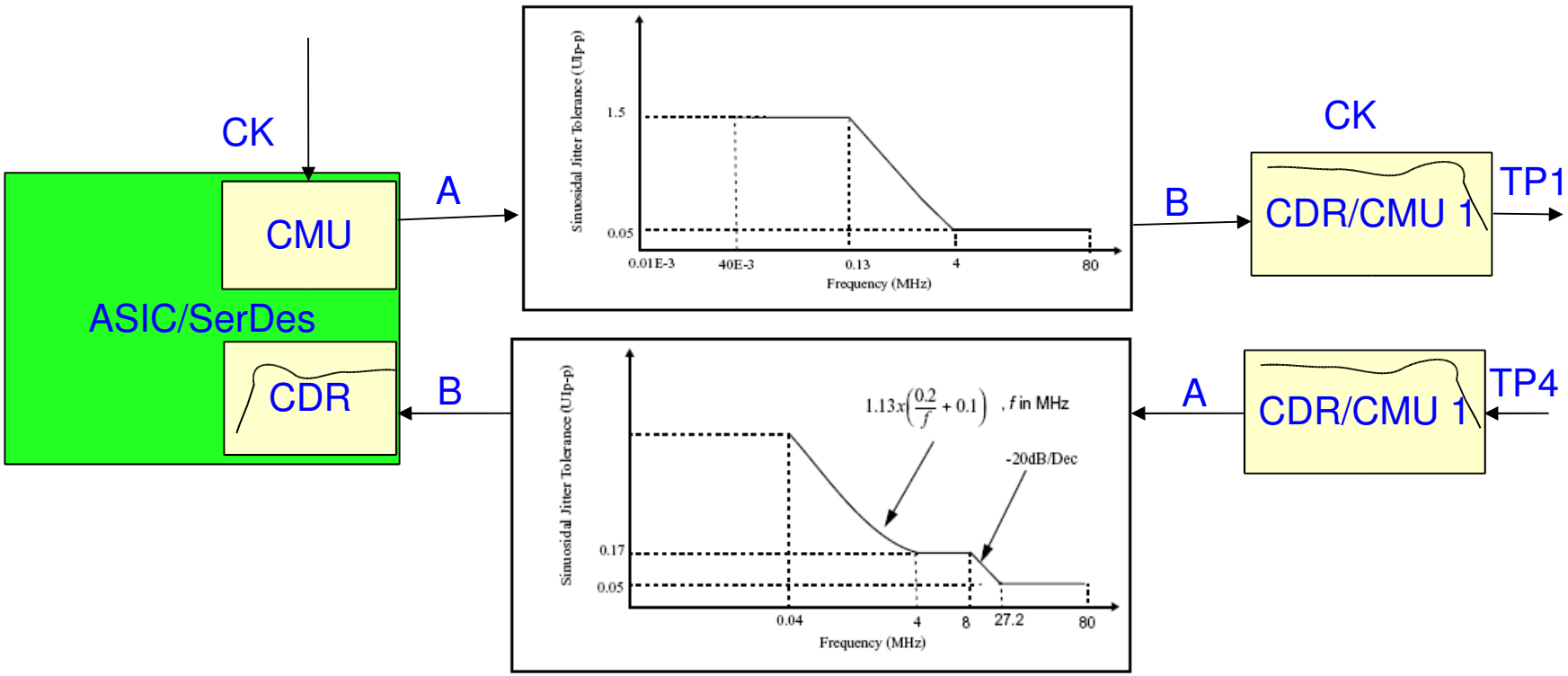
8” Fr4-8 Jitter

- **DDJ and DDPWS for PRBS9 as function of pre-emphasis**
  - **SerDes near end DDJ and DDPWS subtracted from the results.**



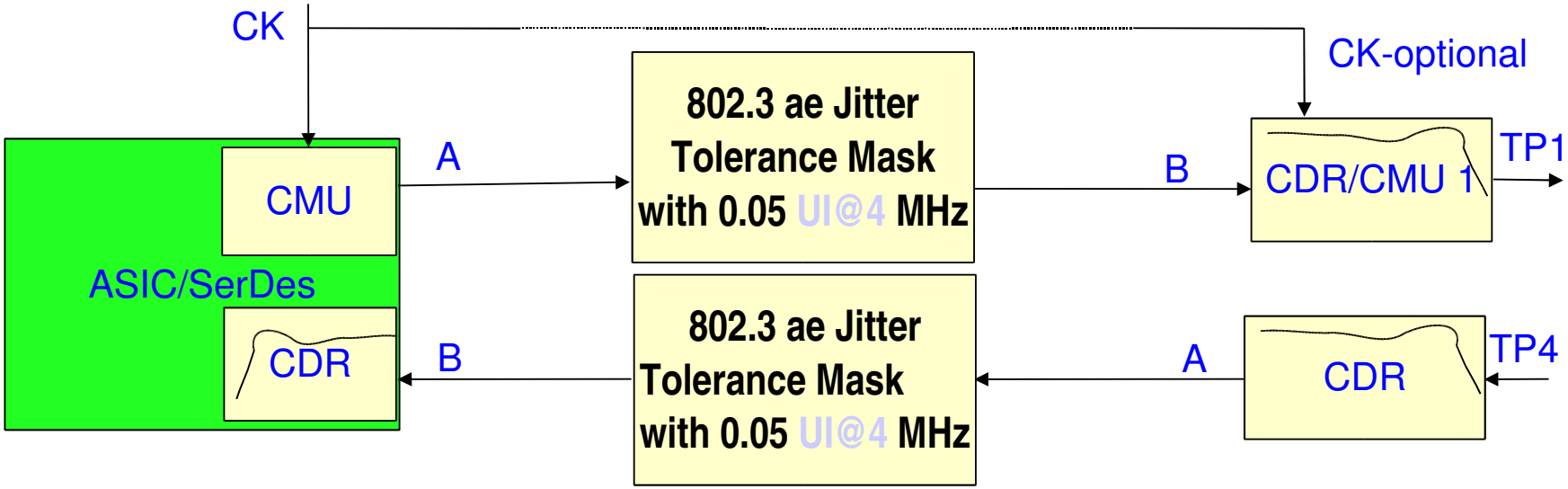
XFI Jitter Transfer Concept

- **XFP assumes CDR has BW of 4-8 Mhz with 1 dB peaking**
- **As the diagram below illustrates this creates additional jitter tolerance penalty on the receiving host.**



CAUI/XALUI Jitter Transfer Concept

- To not penalize the host ASIC the CAUI/XALUI retimer will be based on one of the following implementations:
  - CMU mode (can only be done on the TX path)
  - CDR several implementation are possible low BW with high peaking or high BW with low peaking. No need to define an implementation, just meeting A, B, TP1, and TP4 is sufficient
  - E-FIFO (latency and complexity)



Conclusion

- XALUI/CAUI electrical interface leverage XFI and SFI work, but with several key improvements:
  - Channel length increased to 250 mm of Isola FR4-8 or 375 mm of Nelco N4000-13 by levering pre-emphasis
  - More ASIC friendly return loss than SFI or XFI.
- Simulation shows with XTALK 3x the SFP+ and as transmitter with maximum DJ/TJ the far end eye mask is met with margin.
  - Measurement shown here as well as in latchman_01_0708 confirms the budget.
- Based on simulation and measurement shown here single pre-emphasis setting is adequate for channels with 10 dB loss at Nyquist.
  - Optionally pre-emphasis may be adjusted for each trace for more margin.
- Propose to use 802.3ae jitter tolerance mask with 0.05 UI amplitude at 4 Mhz for jitter tolerance.

40GBASE-SR4 & 100GBASE-SR10 PMD Service Interface Update

John Petrilla, Rita Horner,
Brian Misek, Piers Dawe
Avago Technologies
July 2008

Supporters

- Ali Ghiasi, Broadcom Corporation
- Petar Pepeljugoski, IBM
- Tom Palkert, Luxtera
- Jonathan King, Finisar
- Kenneth Jackson, Emcore Fiber Optics
- Mike Dudek, JDS Uniphase
- Gourgen Oganessyan, Quellan
- Ryan Latchman, Gennum
- Frank Chang, Vitesse

Outline

Focus will be on the Host IC to PMD channel

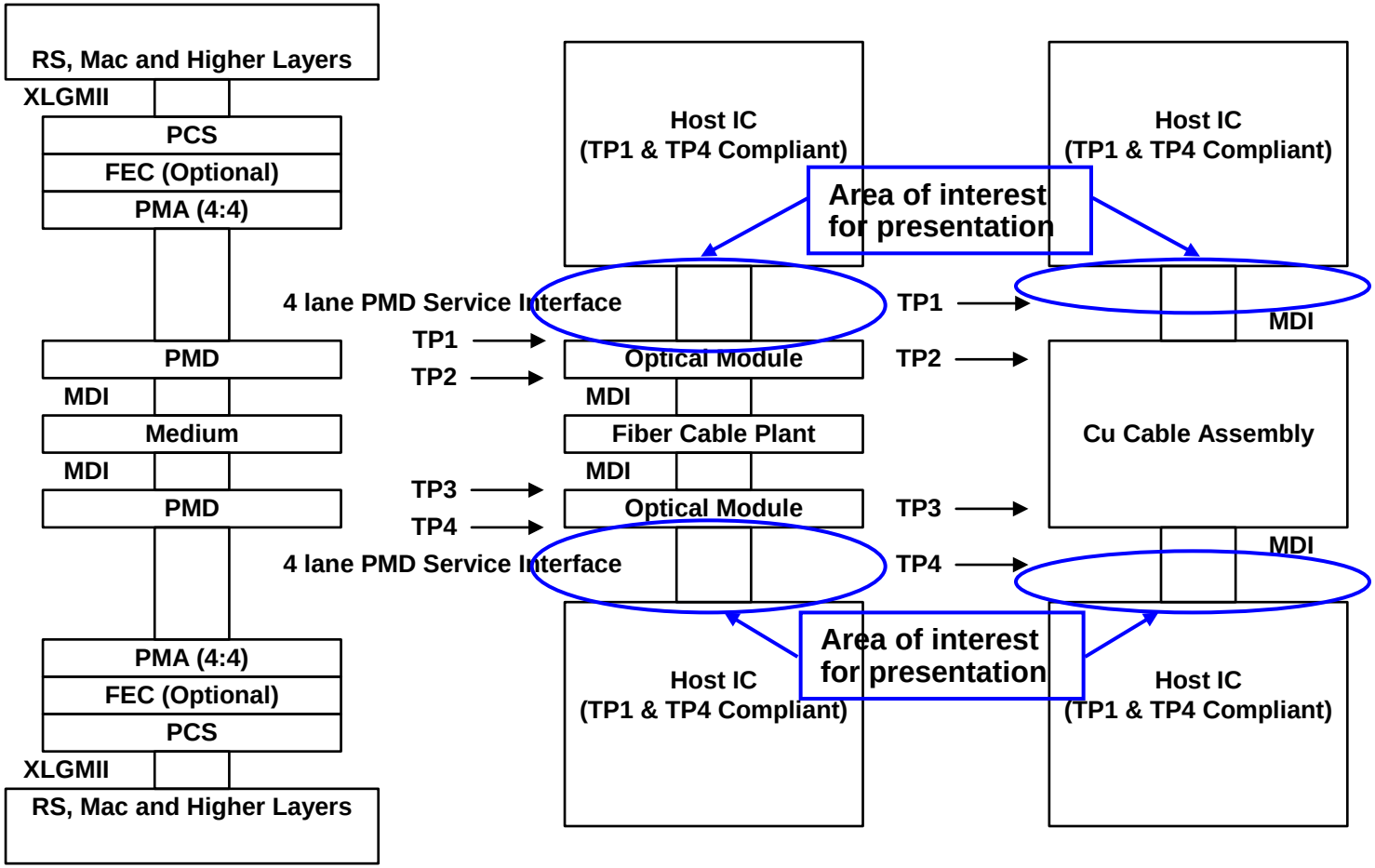
- 802.3ba Alignment
- Elements for success
  - Opportunities for and issues with common form factors
- 40GBASE-SR4 & 100GBASE-SR10 Proposal
  - PMD service interface jitter and electrical characteristics
- Conclusions, Recommendations & Next Steps
- References
- Backup

802.3ba Alignment
802.3ba Objectives Addressed in Presentation

- Support MAC data rates of 40 Gb/s & 100 Gb/s
- Support a BER better than or equal to 1E-12 at the MAC/PLS service Interface
- Provide … at least 100 m on OM3 MMF
- Provide … at least 10 m over a copper cable assembly

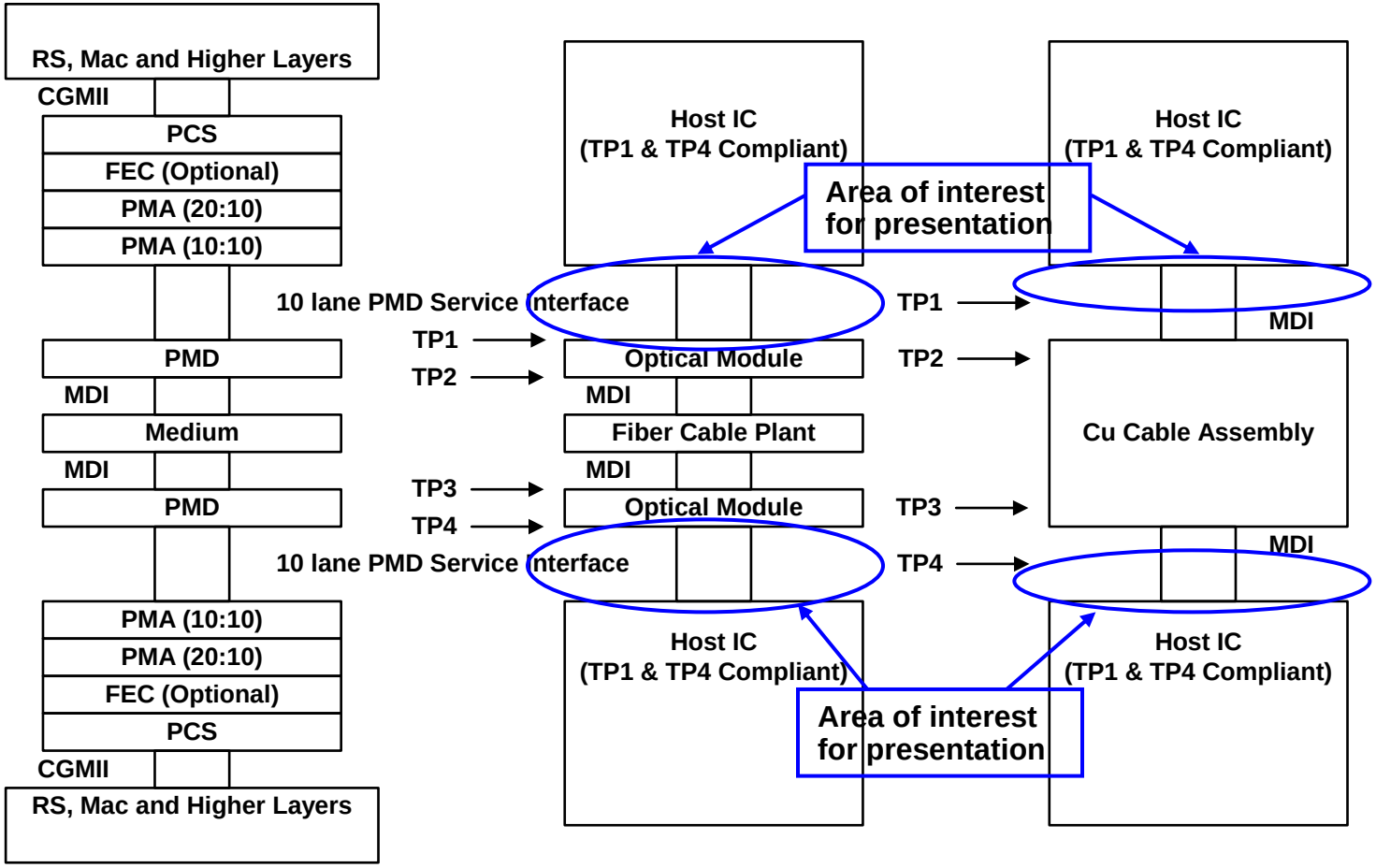
802.3ba Alignment

802.3ba Architectural Layers & Interfaces – 40G



802.3ba Alignment

802.3ba Architectural Layers & Interfaces – 100G

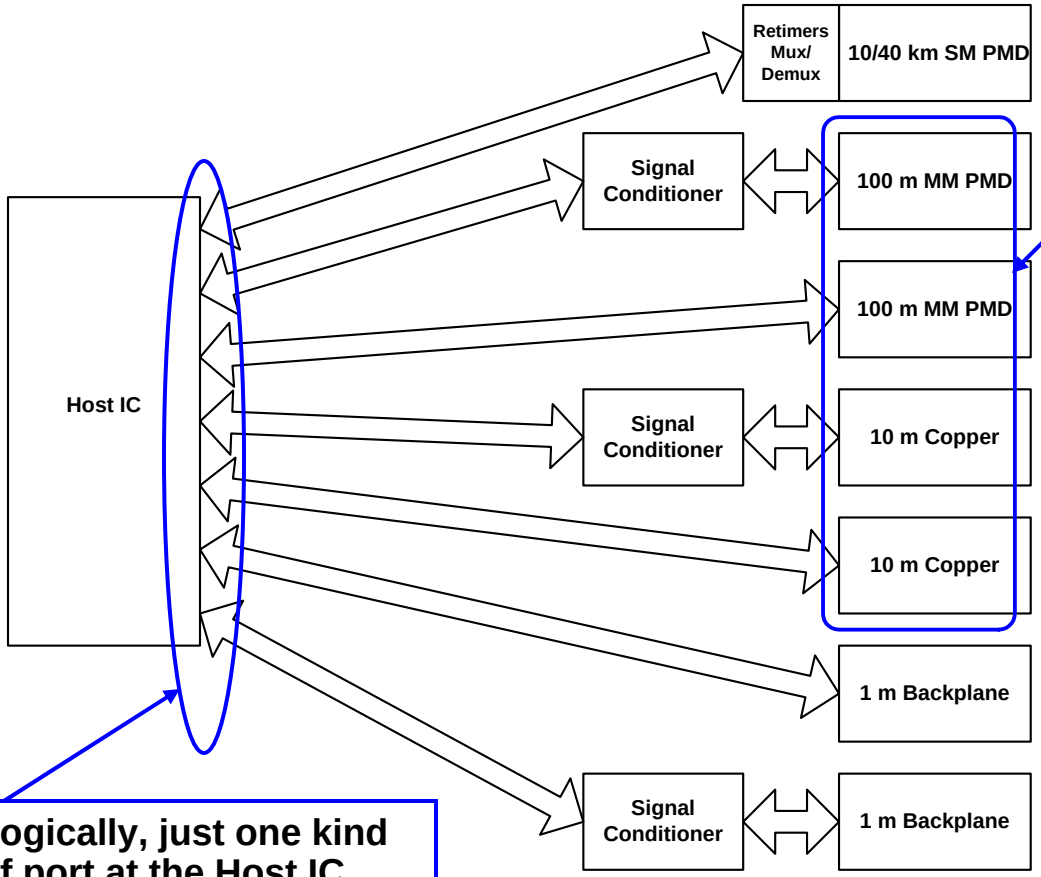


Elements for Success
Overview

- Total cost: less than four/ten 10 GbE solutions
  - A pluggable form factor that is common to multiple variants, e.g. copper cable and MM fiber, leads to lower costs by sharing piece parts and footprints and by accelerating market acceptance and increasing volumes. This enables a single build standard for DTE that can be connected by a choice of the two PMD/media types that will dominate the data center. These advantages are compelling and the opportunities should not be overlooked.
- Power consumption: less than four/ten 10 GbE solutions
- High module density: higher than 10 GbE solutions
- Cable plant: 100 m of OM3 & up to 4 intermediate connections
- Reliability: better than ten/four 10 GbE solutions
- Appropriate design points
  - Support ~6 dBe channel insertion loss (SDD21 at Nyquist frequency) between the host IC and pluggable module without an intermediate connector for 6" to 8" of PCB signal traces. This loss occurs twice, once on each end.
  - Use experience gained in SFP+ (SFF-8431) and 8GFC (FC-PI -4) as well as 802.3ap developments to guide choices for electrical attributes.

May be too aggressive. Making smart choices is essential.

Opportunities for a Common Form Factor



Common form factor

- A common pluggable form factor can support the multilane MM and copper cable variants due to similarities in size, contact count and power consumption. By comparison, the SM modules are expected to have ~4x the power consumption of the MM modules.
- For some applications the host environment will permit direct connection between a Host IC and the pluggable module or backplane connector. For other environments some form of signal conditioner will be required. Both cases are shown.
- Here the term signal conditioner is used to cover all forms of link extenders, CDRs, equalizers, etc.

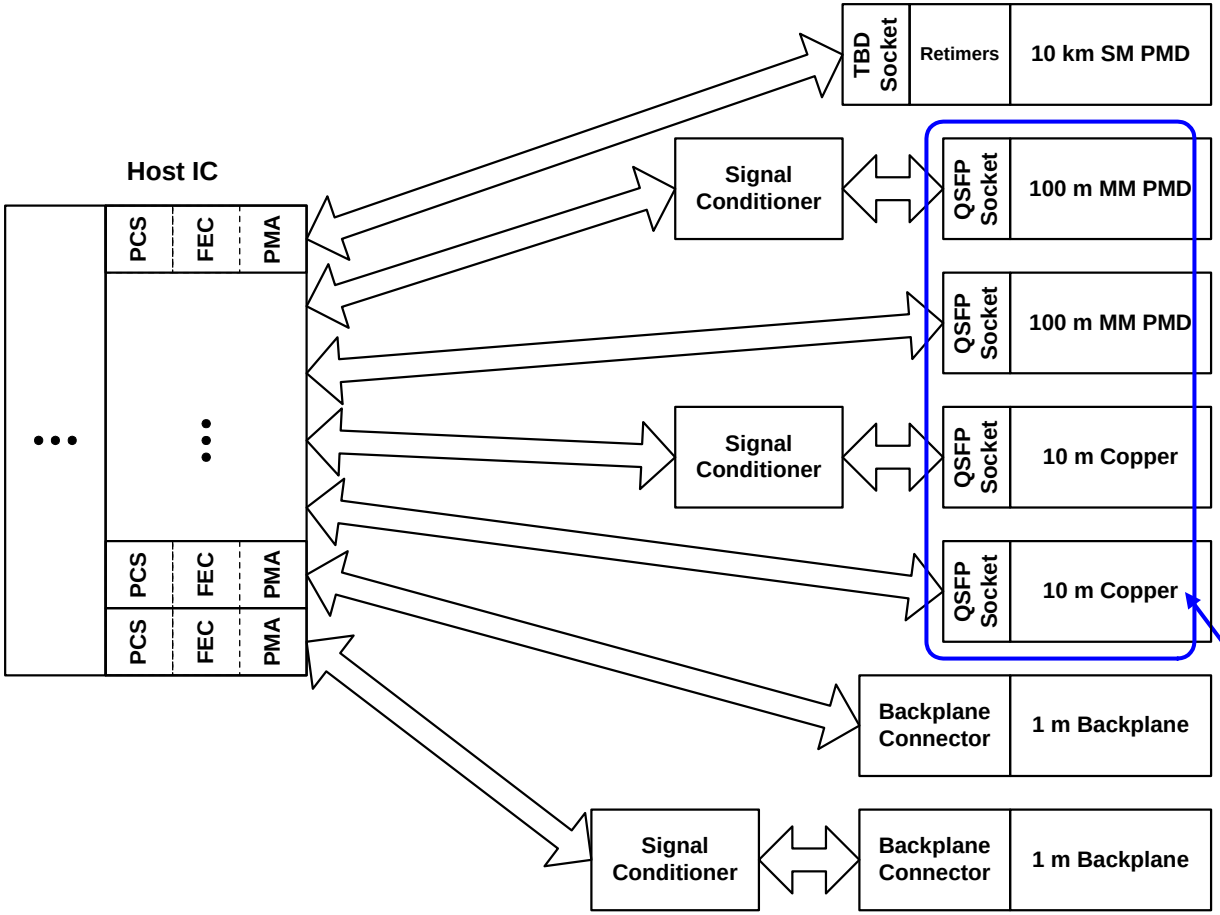
Logically, just one kind of port at the Host IC

Identical functional, logical, digital and management spec

Compatible analog and auto-negotiation

Opportunities for a Common Form Factor

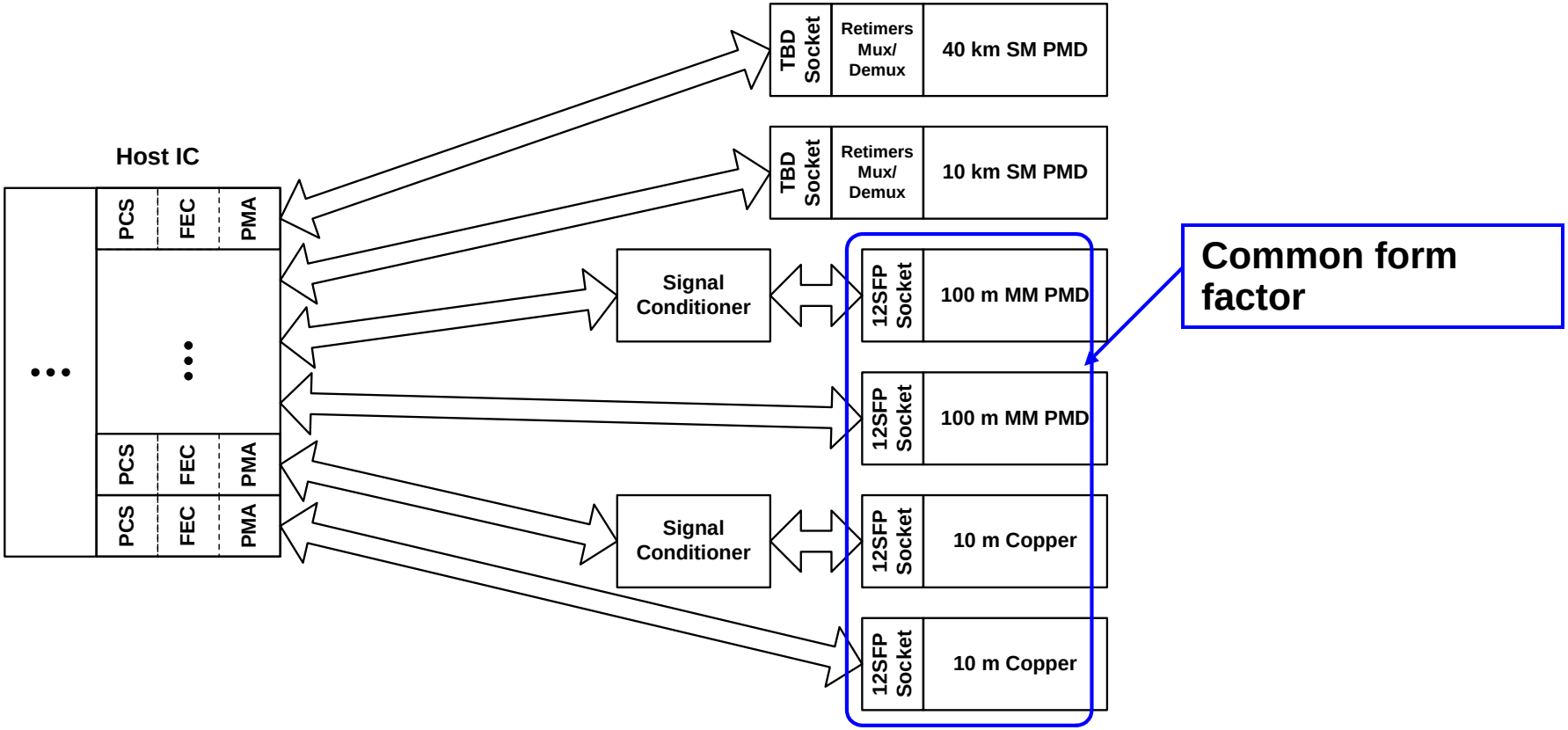
40G Variants & QSFP



- For the 40G variants, the QSFP form factor looks to be a good fit for pluggable 100m MM and 10m Copper variants.
- One issue to address is defining a channel between the Host IC and QSFP socket that supports both the 100m MM and 10m Copper variants. Here the existing terminology differences between copper and optics can be confusing.

Common form factor

Opportunities for a Common Form Factor
100G Variants & 12SFP



- For the 100 m MM PMD and 10 m copper variants, a common pluggable form factor seems reasonable. While there is currently no popular form factor, reasonable 10 lane or 12 lane form factors appear plausible. Here a 12 lane form factor, 12SFP for short, is presumed. As above, an issue to address is defining a channel between the Host IC and 12SFP socket that supports both the 100 m MM and 10 m Copper variants and not being deterred by existing terminology.

Service Interface
Definition Rationale

- For the 100 m OM3 variant, pluggable, multilane, non-retimed, limiting fiber optic modules are expected to provide the lowest power, highest density and lowest cost solution.
  - Direct connection between the module and host IC without intermediate signal conditioners is required to maximize the power, density and cost advantages.
- For the 10 m copper cable assembly variant, direct connection between the host IC and cable assembly offers similar cost, density and power savings as for the 100m OM3 variants.
- For the 1 m backplane variant, direct connection between the host IC and backplane connector offers similar cost, density and power savings as for the 100 m OM3 variants.
- To enable direct connection for the 100 m OM3 variant, the approach and characteristics on the following pages are proposed for consideration.
- Requirements for the 10 m cable assembly and 1m backplane variants are being included as they become apparent.

Beginning to look challenging

PMD Service Interface

Prologue – Jitter Allocation Targets

	GbE	8GFC	10G SFP+ SFF-8431	**40/100G Targets**
TP1 DJ, UI	0.100	0.170	0.100*	**0.150**
TP1 DDPWS, UI		0.110		
TP1 TJ, UI	0.240	0.310	0.280	**0.300**
TP4 DJ, UI	0.462	0.420	0.420	**0.400**
TP4 DDPWS, UI		0.360		
TP4 TJ, UI	0.749	0.710	0.700	**0.700**

* SFF-8431 specifies DDJ instead of DJ.

- Perhaps the most critical attributes of TP1 & TP4 are the jitter allocation. The above table provides information regarding previous choices.
- The targets were chosen to provide a better design point than those from either 8GFC or 10GBASE-SR as implemented in SFF-8431. With respect to 10G SFP+ in SFF-8431, significant relief is expected.

PMD Service Interface

Goal and Approach

Goal: Establish an approach and specifications that enable direct connection and inter-operability between host ICs and pluggable, multilane, non-retimed, limiting fiber optic modules for a reasonable range of equipment designs

- The proposed approach does not place explicit requirements on the host IC, but, rather, provides far-end characteristics over worst case channels.
- The far-end characteristics for the host IC, TP1 & TP4, are based on experience gained from SFP+ (SFF8431) and 8GFC (FC-PI-4) developments. Included are jitter, signal levels, and reflection coefficients.
- The proposed channel is intended to support 150 mm to 200 mm of PCB traces without an intermediate connector and is defined by an SDD21 template.
- TP1 requirements are defined in an input characteristics table and TP4 requirements are defined in an output characteristics table that follow.
- Compliance, as with SFF-8431, is based on use of compliance test boards to improve measurement accuracy and reproducibility.

May be too aggressive

Multilane channels are challenging

The intention is to maintain the same TP1 & TP4 requirements for 40GBASE-SR4 and 100GBASE-SR10.

PMD Service Interface
Host IC – PMD Channel

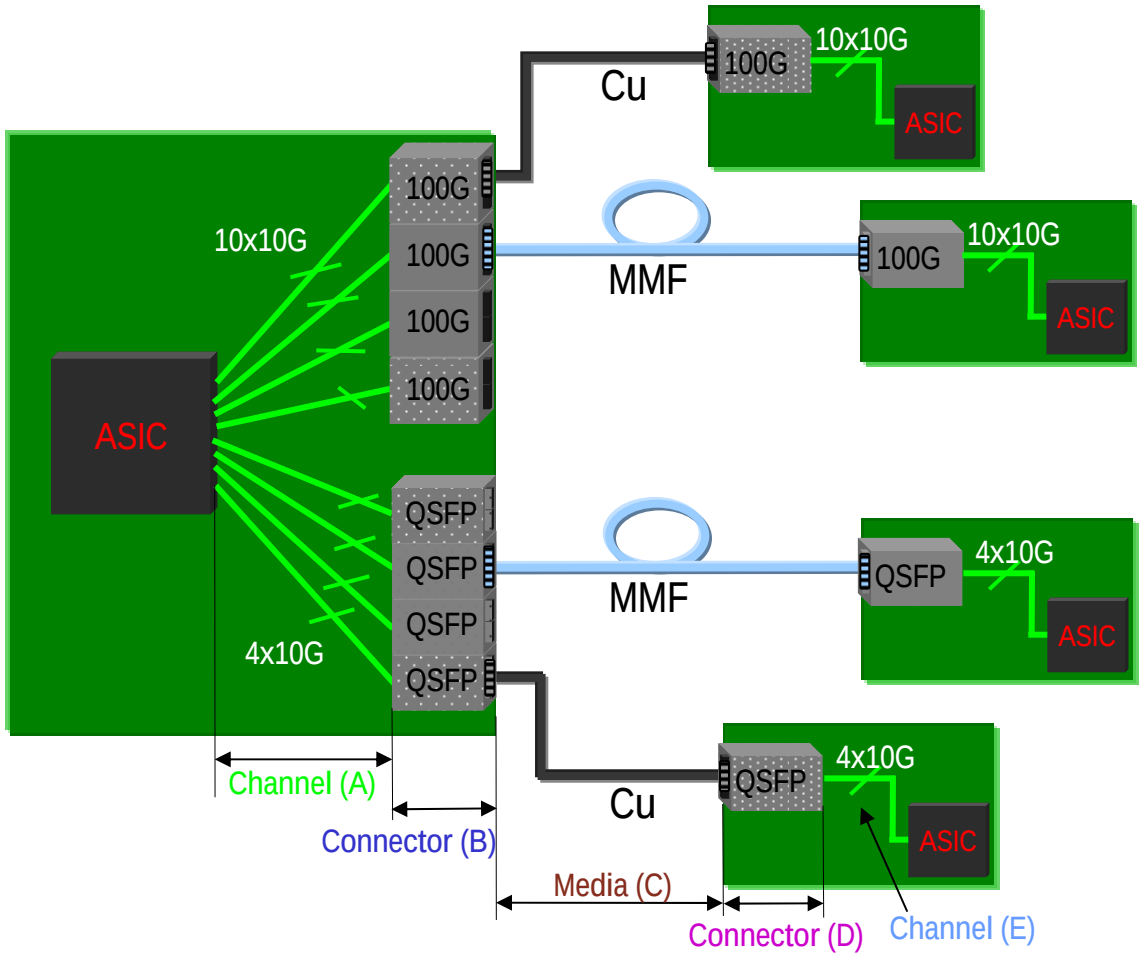
Goal: To define a common port and channel that accommodates both pluggable 100 m MM fiber and 10 m cable assemblies from a single ASIC

- For cable assemblies using 24 AWG, SDD21 loss at the Nyquist frequency is expected ~ 22 dBe. Accommodating another 6 dBe between the port socket and the ASIC on each side may be un-necessarily challenging for the host IC.
- Three issues will be presented.
  - PCB trace lengths that may be required in end use to support multi-port equipment
  - PCB trace lengths that may be required in evaluation boards and/or compliance test boards.
  - Appropriate compliance points for host IC, i.e. where to locate TP0 and TP5.

Need for Well Defined Port Side Interface Specification

Optics and Cu

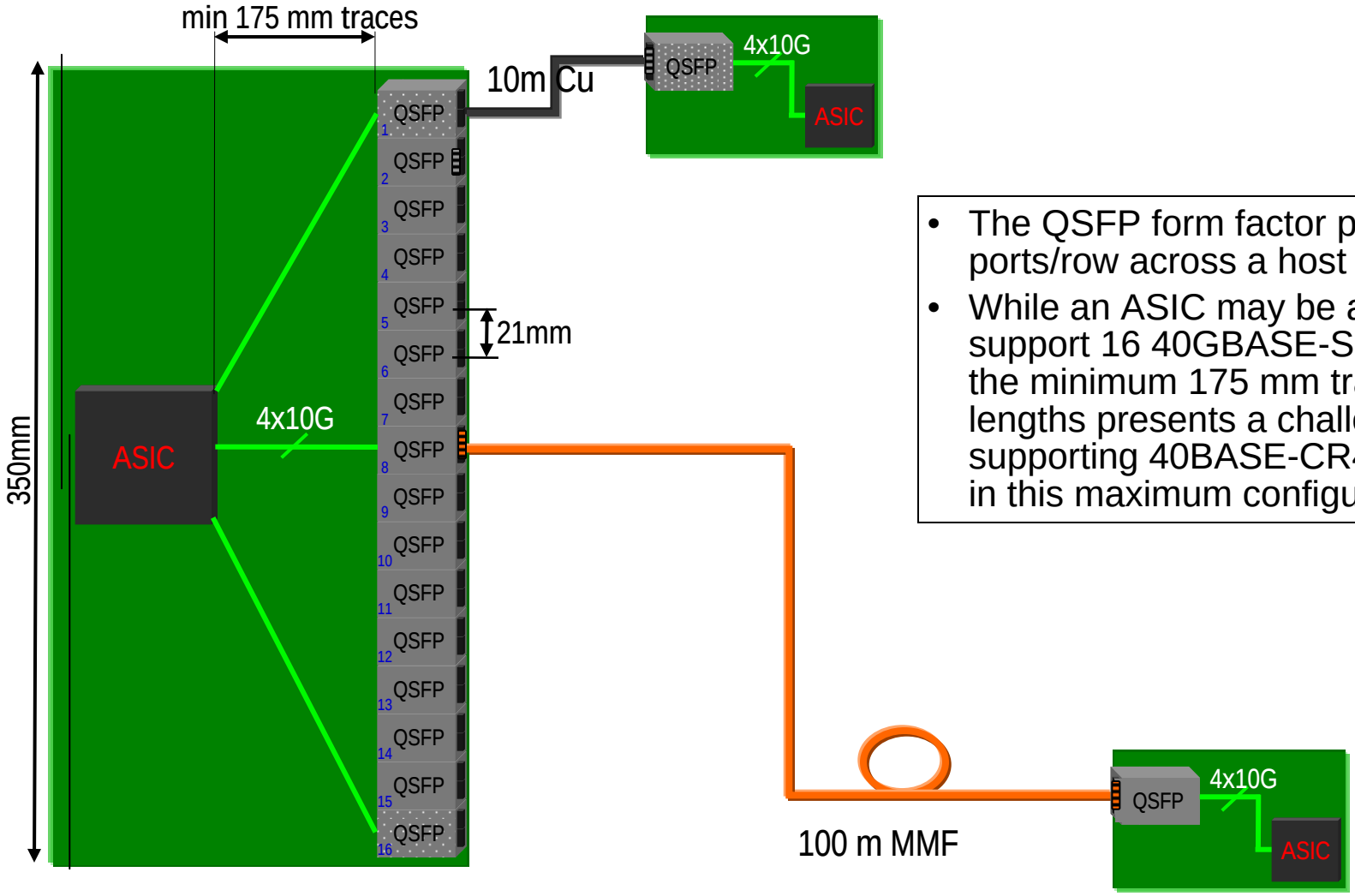
- 100 m MMF links appear jitter limited.
- 10 m Copper links appear insertion loss or ICR limited.
- Can a common test point (e.g. TP0 or TP1) with a common channel (e.g. Channel (A)) be defined that covers both variants?



- For 100 m MMF Total SerDes Link Jitter = ASIC Generation + Channel Contribution (A) + MMF Tx Contribution (B) + MMF Contribution (C) + MMF Rx Contribution (D) + Channel Contribution (E) + ASIC Tolerance
- For 10 m Cu Total SerDes Link Loss = Channel Loss (A) + Connector Loss (B) + Interconnect Media Loss (C) + Connector Loss (D) + Channel Loss (E)

40G SerDes Port Side Connections to QSFP

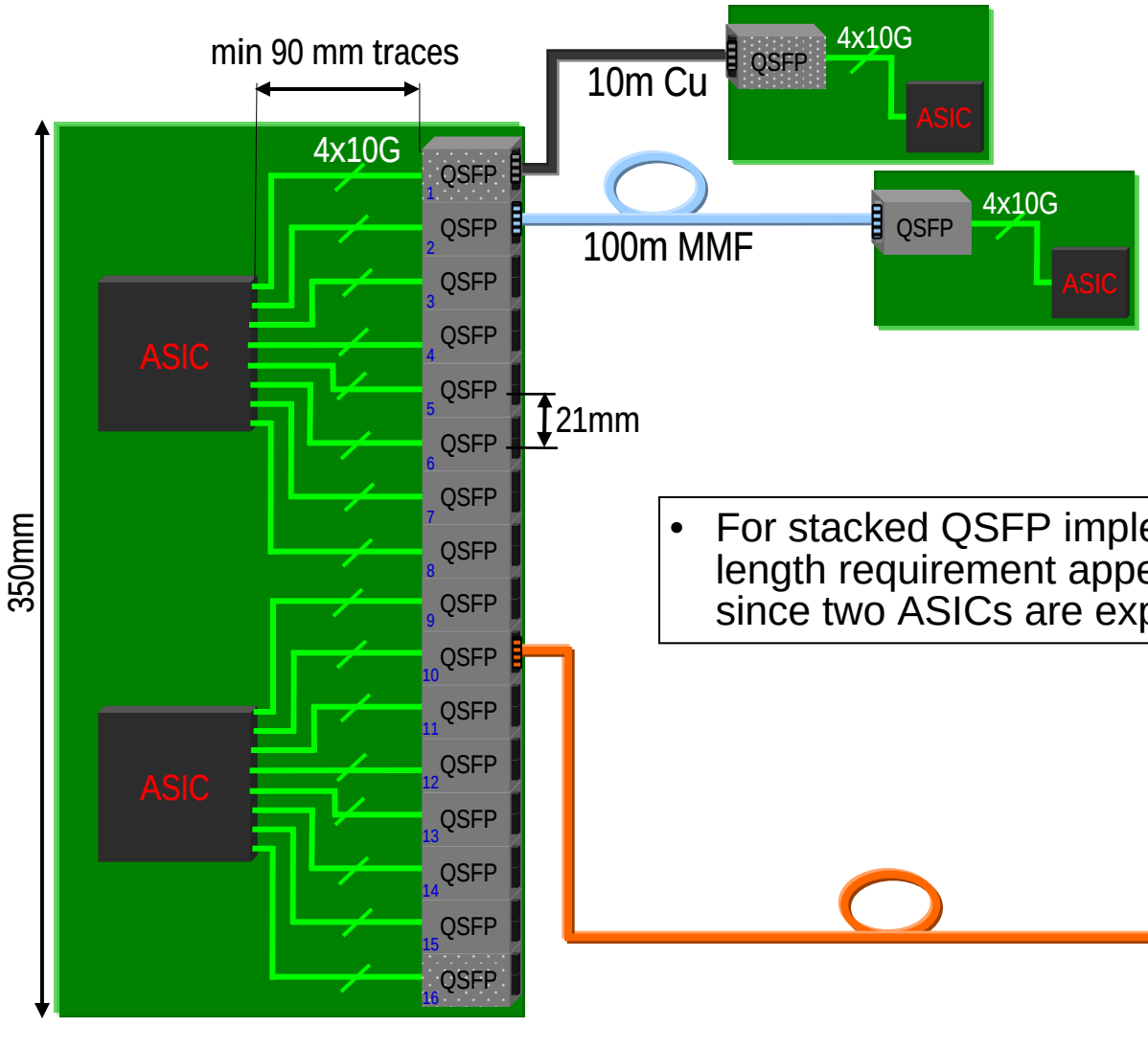
Single ASIC – Max Non-Stacked Port Case



- The QSFP form factor permits 16 ports/row across a host PCB.
- While an ASIC may be able to support 16 40GBASE-SR4 ports, the minimum 175 mm trace lengths presents a challenge when supporting 40BASE-CR4 variants in this maximum configuration.

40G SerDes Port Side Connections to QSFP

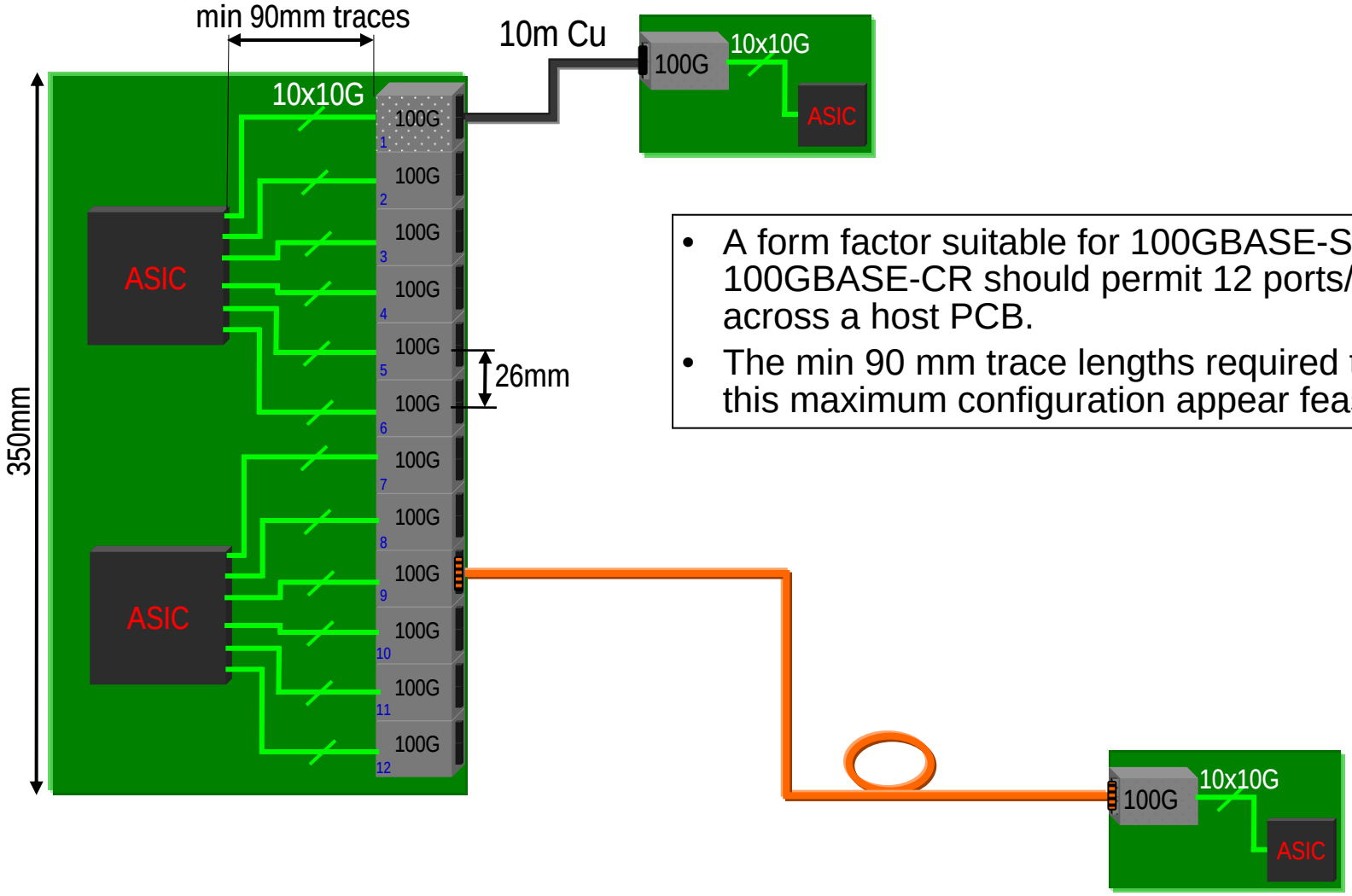
Maximum Stacked Port Case



- For stacked QSFP implementations the trace length requirement appears more supportable since two ASICs are expected.

100G SerDes Port Side Connections

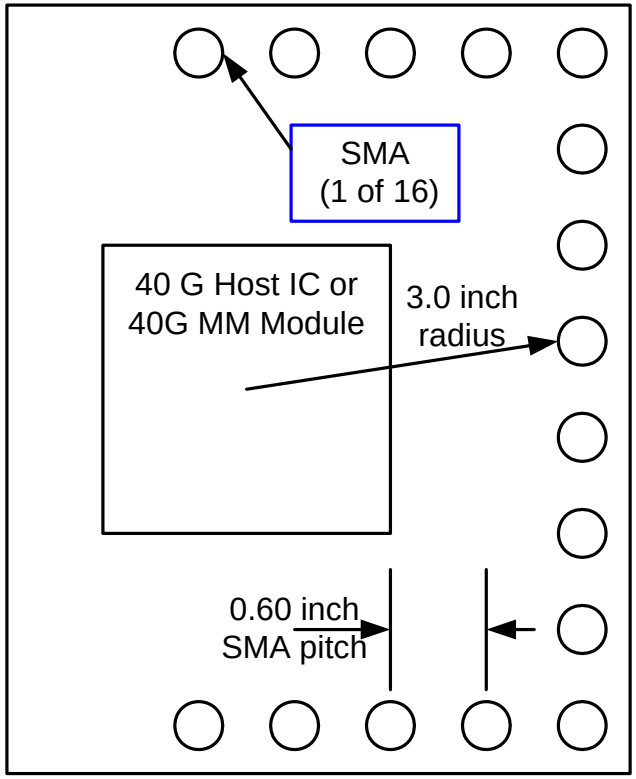
Two ASICs – Max Port Case



- A form factor suitable for 100GBASE-SR10 and 100GBASE-CR should permit 12 ports/row across a host PCB.
- The min 90 mm trace lengths required to support this maximum configuration appear feasible.

TP1 & TP4 PMD Service Interface

4 Lane Compliance and/or Evaluation Boards

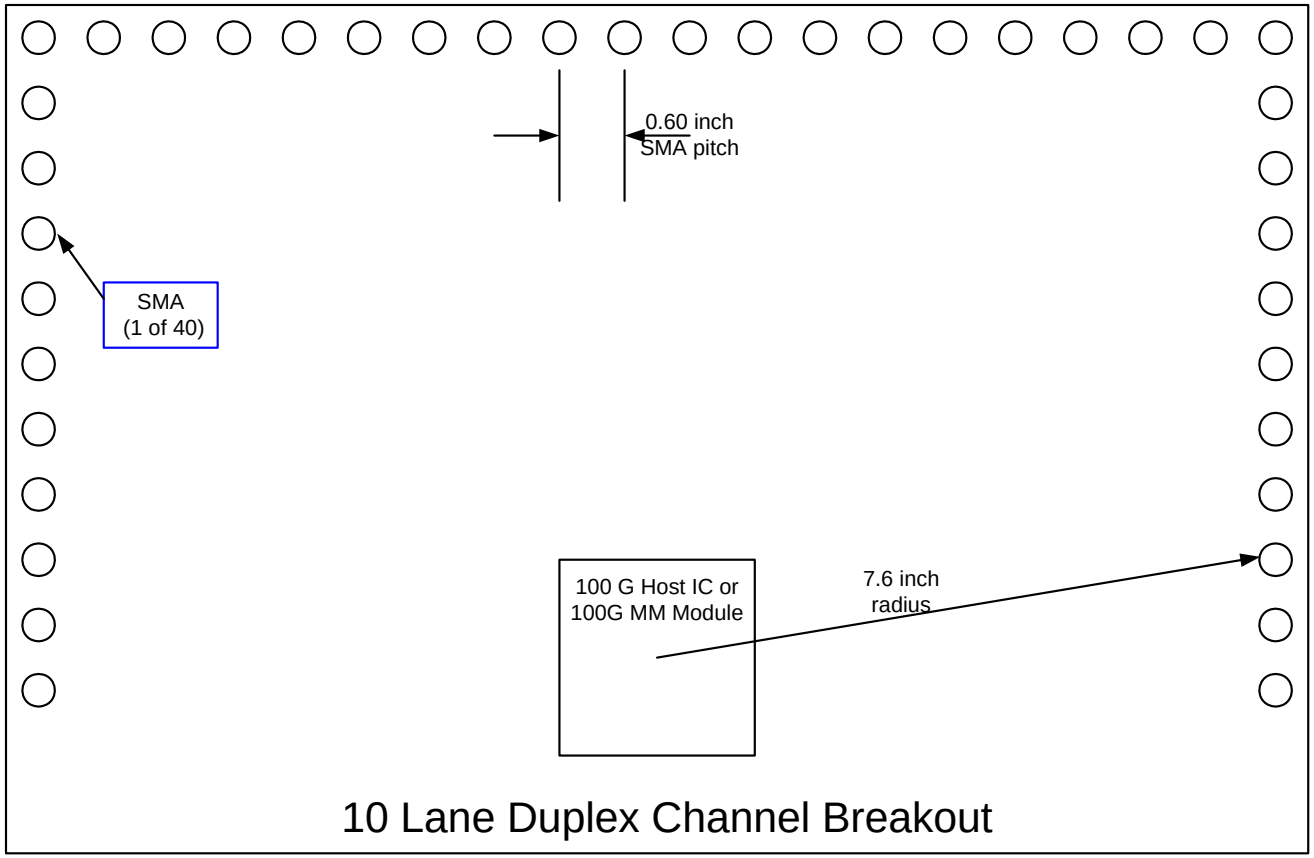


4 Lane Duplex Channel Breakout

- Evaluation and/or compliance test cards for multilane channels face the challenge of breaking out all channels. In the figure all eight differential pairs of a full duplex, four lane channel are made available with SMA connectors. For an SMA pitch of 0.60 inch and semicircular pattern, a minimum radius of 3.0 inches is required. This is larger than the trace length in SFF-8431 module compliance boards but may be tolerable.
- The pattern in the drawing is rectangular for simplicity.

TP1 & TP4 PMD Service Interface

10 Lane Compliance and/or Evaluation Boards

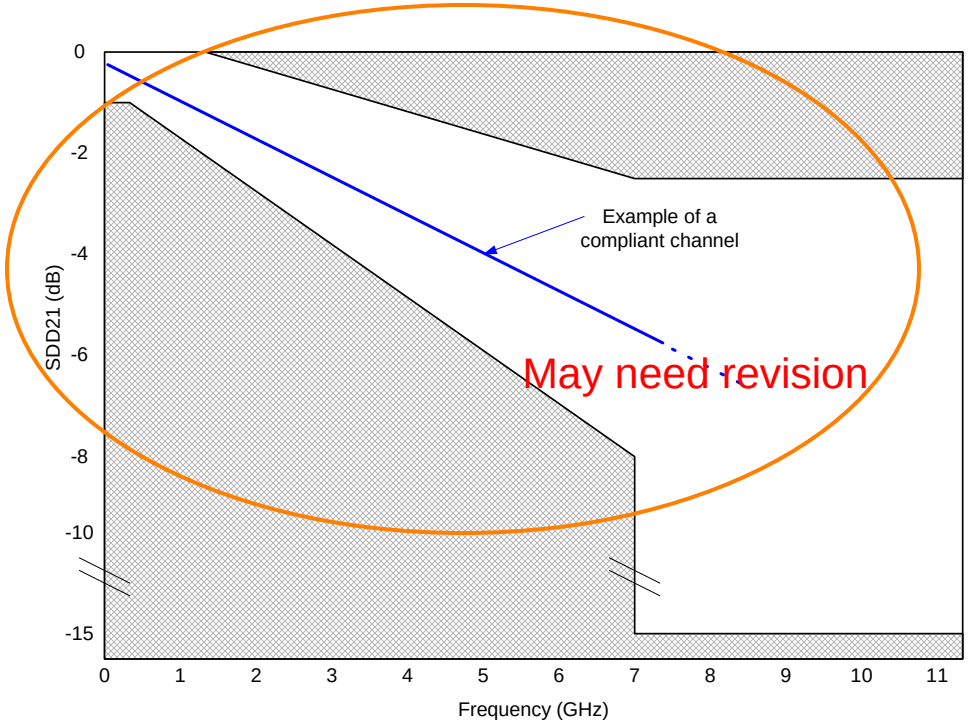


- In the figure all 20 differential pairs of a full duplex, ten lane channel are made available with SMA connectors. For an SMA pitch of 0.60 inch and semicircular pattern, a minimum radius of 7.6 inches is required. This trace length is significantly longer than trace lengths in SFF-8431 module compliance boards and presents a significant challenge.

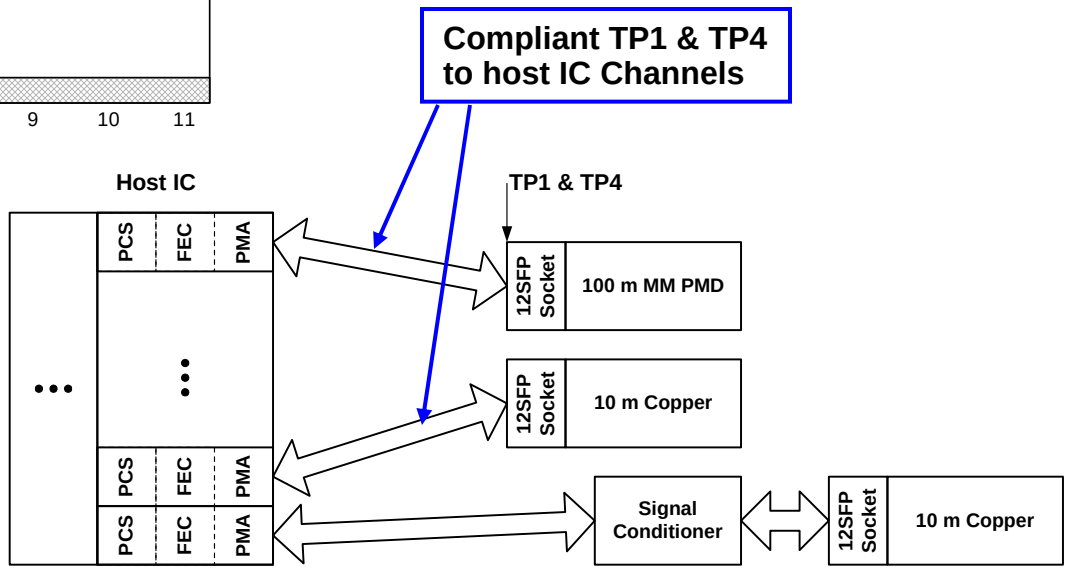
PMD Service Interface
Host IC – PMD Channel Summary

- While for 100 m MMF variants, jitter appears to the limiting factor, for 10 m Copper links, insertion loss (or insertion loss crosstalk ratio) appears to be the limiting factor.
  - Finding a way to reduce the combined insertion loss without compromising the needed reaches will be key.
- While supporting some possible maximum port population cases requires a minimum 175 mm of PCB traces, most can be supported with a minimum of 90 mm.
- The short trace lengths found in SFF-8431 compliance cards may not be feasible for 10-lane channels. Connectors offering a higher packing density and only breaking-out a subset of the lanes should be explored.
- Finding optimum compliance points for the PMD Service Interface that account for the needs of the ASIC, 100 m MMF pluggable modules and the 10 m Copper pluggable assemblies as well as pragmatic consideration of compliance test cards requires further attention. For the ASIC, a single set of test points that are common to copper and MM optics variants is important.

TP1 & TP4 to Host IC Channel SDD21 Compliance Template

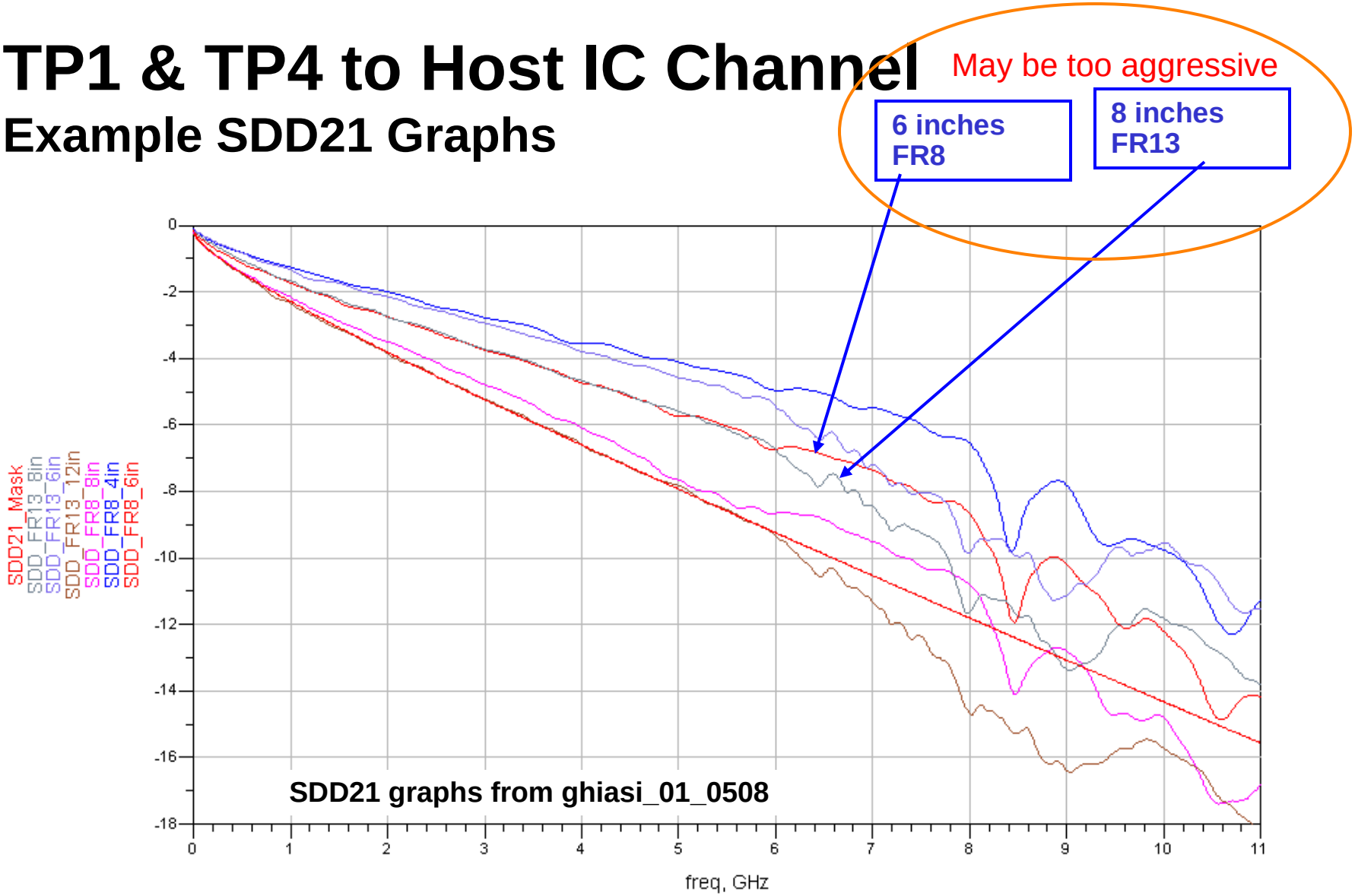


- Characteristics are for each lane individually and are normative except where noted.
- All values are provisional, shown for example, and will benefit from additional study.



TP1 & TP4 to Host IC Channel

Example SDD21 Graphs



- The above examples for FR13- 8in, FR13- 6in, FR8-6in and FR8-4in all meet the proposed template.

TP1

40GBASE-SR4 & 100GBASE-SR10 Input Characteristics

Parameter Description	Value	Units	Conditions
Single ended input voltage tolerance range	-0.3 to 4.0	V	Ref'd to module signal common
AC common mode input voltage tolerance (min)	15	mV (RMS)	
Differential input reflection coefficient, SDD11 (max)	See Template A	dB	0.01 to 11.1 GHz
Reflected differential to common mode conversion, SCD11 (max)	-12	dB	0.01 to 11.1 GHz
Total jitter tolerance	0.30	UI	At BER = 1E-12
Deterministic jitter tolerance	TBD	UI(p-p)	
Eye mask coordinates: X1, X2, Y1, Y2	0.15, TBD, 90, 350		TBD

Characteristics are for each lane individually and are normative except where noted.

All values are provisional, shown for example, and will benefit from additional study.

As requirements for 100 m MMF and 10 m Cu are harmonized, DJ jitter tolerance may be relaxed and the maximum input signal amplitude tolerance increased depending on the maximum supported PCB trace length. Alternatively, the ASIC could generate different signal amplitudes for copper and fiber

TP4

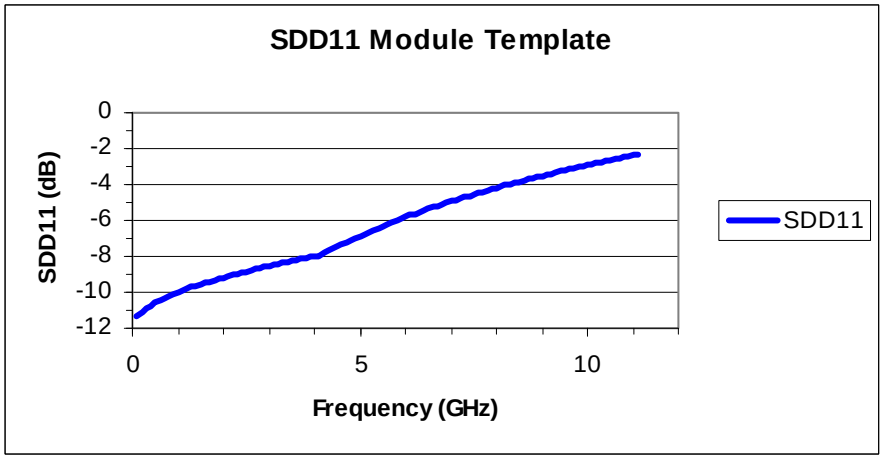
40GBASE-SR4 & 100GBASE-SR10 Output Characteristics

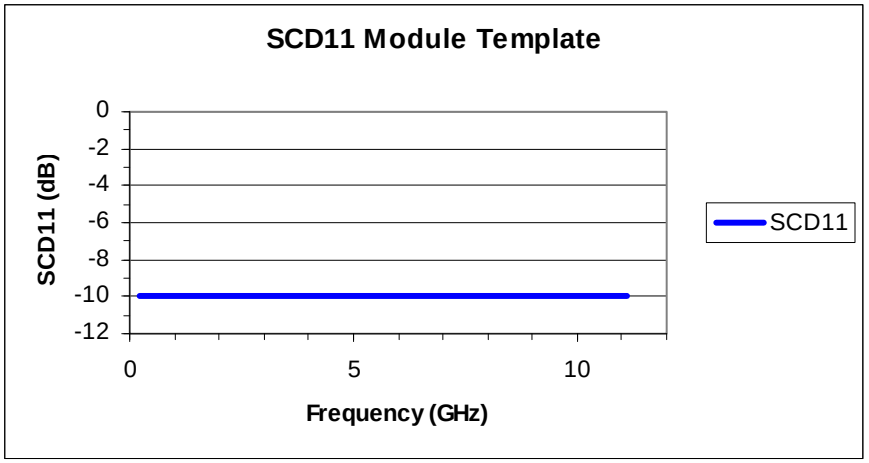
Parameter Description	Value	Units	Conditions
Single ended output voltage tolerance range	-0.3 to 4.0	V	Ref'd to module signal common
AC common mode output voltage (max)	7.5	mV (RMS)	
Termination mismatch at 1 MHz	5	%	
Differential output reflection coefficient, SDD22(max)	See Template B	dB	0.01 to 11.1 GHz
Common mode output reflection coefficient, SCC22 (max)	See Template C	dB	0.01 to 11.1 GHz
Output transition time, 20% to 80%, (min)	28	ps	
Total jitter	0.70	UI	At BER = 1E-12
Deterministic jitter	0.40	UI(p-p)	
Eye mask coordinates: X1, X2, Y1, Y2	0.35, TBD, 150, 425		TBD

Characteristics for are each lane individually and are normative except where noted.

All values are provisional, shown for example, and will benefit from additional study.

TP1
Reflection Coefficient Characteristics





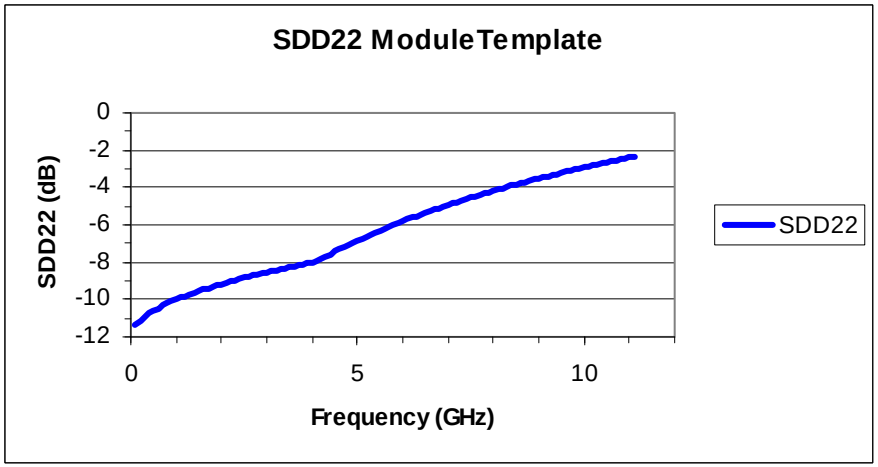
Template A
(updated for SFF-8431 r3.0)

Characteristics are for each lane individually and are normative except where noted.

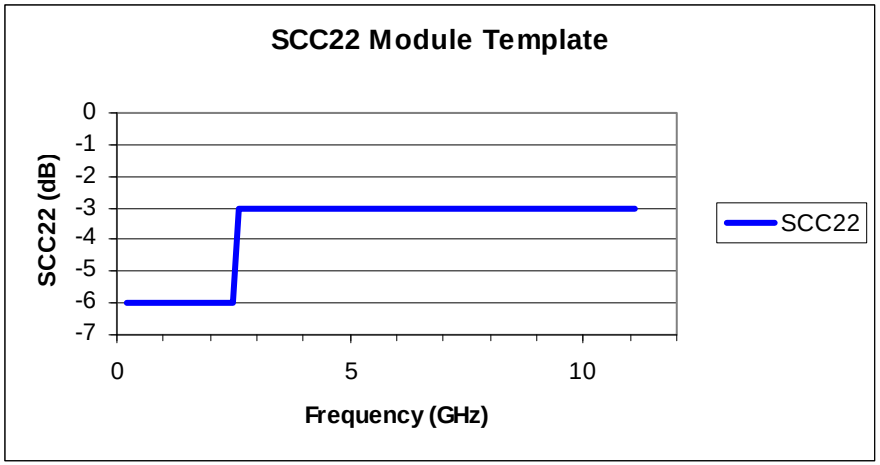
All values are provisional, shown for example, and will benefit from additional study.

TP4

Reflection Coefficient Characteristics



Template B
(updated for SFF-8431 r3.0)

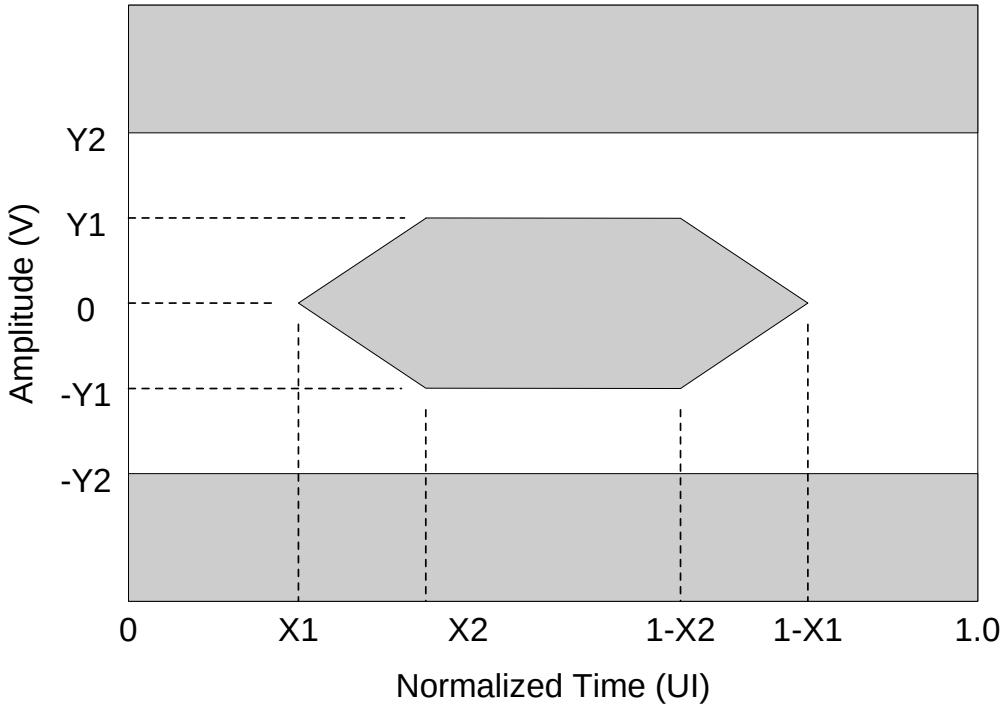


Template C

Characteristics are for each lane individually and are normative except where noted.

All values are provisional, shown for example, and will benefit from additional study.

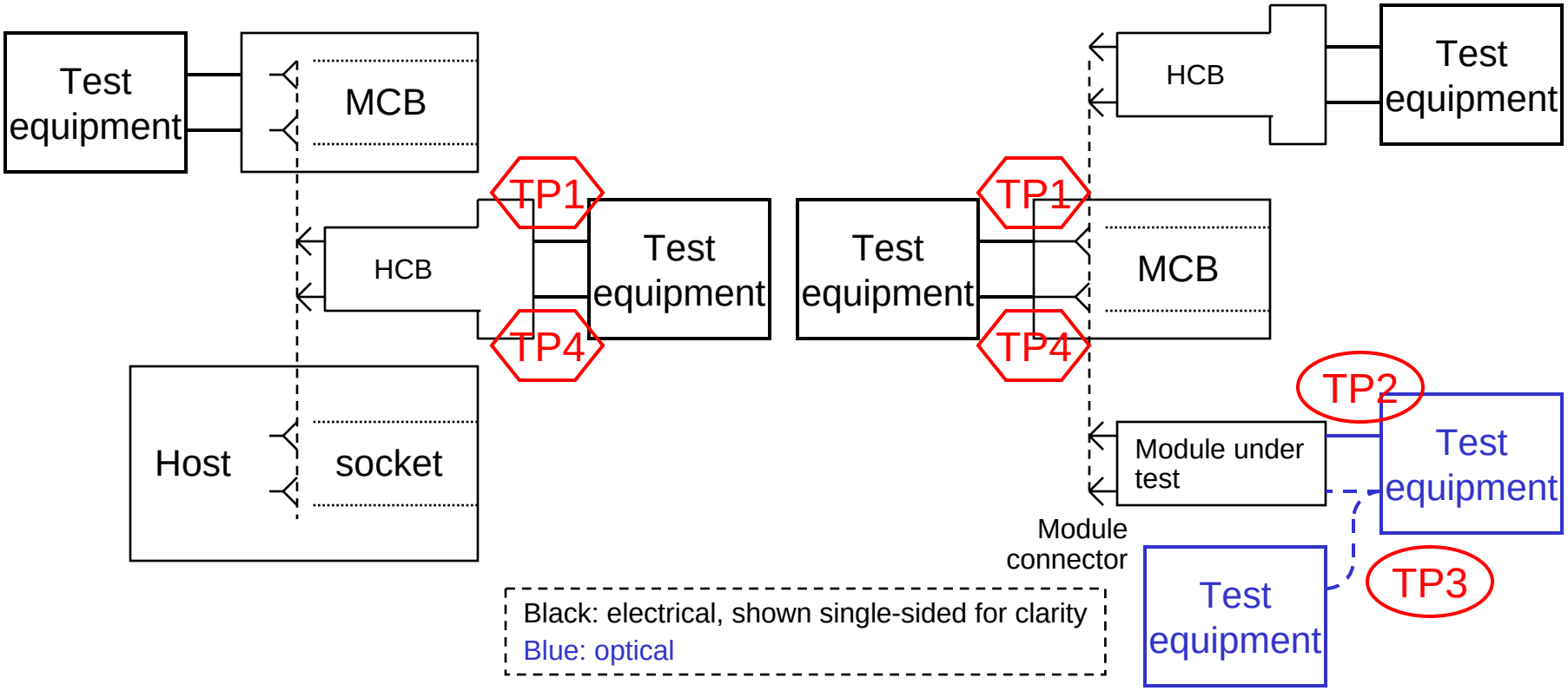
TP1 and TP4
Eye Mask



Characteristics are for each lane individually and are normative except where noted.

All items will benefit from additional study.

Use of compliance boards



- HCB (Host Compliance Board) Used to provide access at the far-end, TP1 & TP4, of Host IC signals and for calibration of module compliance test signals. Host system transmit and receive signal compliance are defined with the HCB inserted in the pluggable interface of the host.
- MCB: (Module Compliance Board) Used to provide access at TP1 & TP4 to module signals and for calibration of host compliance test signals. Module transmit and receive signal compliance are defined with the module inserted in the pluggable interface provided by the MCB.

Conclusions, Recommendations & Next Steps

Conclusions:

- Cost, power and density advantages of 100m OM3, 10m copper cable assembly and 1m backplane variants are maximized by direct connection with the host IC and a single build standard for DTE that can be connected by a choice of the two PMD/media types that will dominate the data center.
- Robust solutions, inter-operability of pluggable modules and cable assemblies and market acceptance are enabled by well chosen required interface characteristics.
- This presentation provides a set of characteristics sufficient to enable direct connection between the proposed 100m OM3 variant and a host IC.

Recommendations:

- The specification approach as outlined in the tables, pages 24 & 25, and associated templates, pages 22, 26, 27 & 28 should be considered for inclusion in 802.3ba.
- Coordinate TP1 and TP4 requirements with other PMD variants, looking for commonality in use for ASICs, 100 m MM and 10 m copper variants. Re-activation of the Test Point ad hoc should be considered.

Next Steps:

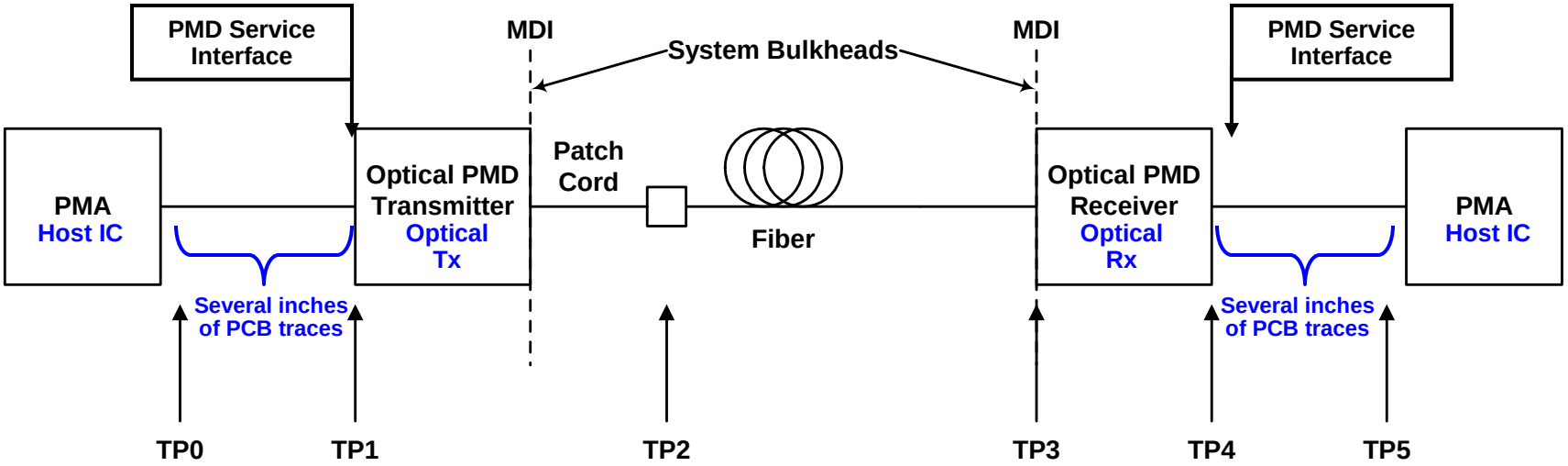
- Continue to gather and incorporate feedback regarding host IC capabilities
- Collect information regarding crosstalk and other impairments related to multilane channels, stacked connectors and modules and incorporate.
- Upgrade proposed specifications.

References

- SFF-8431 Specifications for Enhanced 8.5 and 10 Gigabit Small Form Factor Pluggable Module ”SFP+”: ftp://ftp.seagate.com/sff/SFF-8431.PDF
- FC-PI-4 Physical Interface-4 (8GFC): http://www.t11.org/index.htm
- ghiasi_01_0508
- pepeljugoski_01_0508
- petrilla_01_0508
- petrilla_01_0308

802.3ba Alignment

802.3ba 100 m MM PMD Block Diagram



- The above block diagram shows relevant elements and interfaces for an optical link between two PMAs. The patch cord is included for the definition of TP2. Otherwise intermediate fiber connectors are not shown.
- P802.3ba, in Munich, agreed to include “an optional n-lane x 10.3125GBd electrical interface for PMD service interface … as baseline”.
- Here the PMAs may be host ICs and the PMDs, fiber optic modules. TP1, TP2, TP3 and TP4 are traditional labels for interfaces of a fiber optic link. TP0 and TP5 may be used as labels of the host IC interfaces.

802.3ba copper cable assembly baseline proposal

Chris Di Minico
MC Communications
cdiminico@ieee.org

Contributors

- Galen Fromm, Jay Neer - Molex
- Jens Aumann, Leoni Special Cables
- Vivek Telang, Broadcom
- Howard Baumer, Mobius Semiconductor
- Amir Mezer, Intel

Supporters

- Dan Dove, ProCurve Networking by HP
- Shimon Muller, Sun Microsystems
- Tom Palkert, Luxtera
- Gourgen Oganessyan, Quellan
- Ed Cady, Meritec
- Herb Van Deusen, Gore
- Hugh Barrass, Cisco
- ilango Ganga, Intel
- Rich Mellitz, Intel
- Greg McSorley, Amphenol
- Bob Thornton, Fujitsu
- Bill MacKillop, Cinch Connectors
- Jim McGrath, Cinch Connectors

Summary

- Considerations for 802.3ba Cu cable assembly specifications for 802.3ba baseline proposal.

- Measurement models and simulation models developed to validate usage of 10GBASE-KR (Clause 72) for 10 Gb/s lane options for both 40GBASE-CR4 and 100GBASE-CR10 cable assemblies.

- CX4 twinaxial cable assembly differential parameters proposed as basis for 40GBASE-CR4 and 100GBASE-CR10 link specification (i.e., S-parameters).

- Considerations for configuring QSFP low speed electrical hardware pins for 40GBASE-CR4 operation.

- Two independent demonstrations of 10GBASE-KR operation over 10 meters of passive copper cable assemblies.

802.3ba objectives

- Support full-duplex operation only
- Preserve the 802.3 / Ethernet frame format utilizing the 802.3 MAC
- Preserve minimum and maximum FrameSize of current 802.3 standard
- Support a BER better than or equal to 10^{-12} at the MAC/PLS service interface
- Provide appropriate support for OTN

- Support a MAC data rate of 40 Gb/s
- Provide Physical Layer specifications which support 40 Gb/s operation over:
  - at least 10km on SMF
  - at least 100m on OM3 MMF
  - at least 10m over a copper cable assembly
  - at least 1m over a backplane
- Support a MAC data rate of 100 Gb/s
- Provide Physical Layer specifications which support 100 Gb/s operation over:
  - at least 40km on SMF
  - at least 10km on SMF
  - at least 100m on OM3 MMF
  - at least 10m over a copper cable assembly

Copper cable assembly: lane options considered

- Support a MAC data rate of 40 Gb/s
- Provide Physical Layer specifications which support 40 Gb/s operation over:
  - at least 10m over a copper cable assembly
  - 4 x 10 Gb/s lane

- Support a MAC data rate of 100 Gb/s
- Provide Physical Layer specifications which support 100 Gb/s operation over:
  - at least 10m over a copper cable assembly
  - 10 x 10 Gb/s lane

802.3ba Cu cable assembly proposal

- Utilize 10GBASE-KR (Clause 72) and 10GBASE-CX4 to specify 40GBASE-CR4 and 100GBASE-CR10.
  - 64B/66B PCS
  - Signaling speed 10.3125 Gbd (per lane)
  - 4x and 10x - KR transmit and receive functions
  - Commonality with 40 GbE backplane proposal
  - S-parameters - cable assembly differential parameter
  - x4 MDI considered: QSFP and IEC 61076-3-113 mechanical mating interface (10GBASE-CX4 mechanical)
    - QSFP- module and connector dimensions common for both fiber and copper
      - For 40GBASE-CR4, QSFP low speed control and sense signals set to non-operational QSFP state.
    - CX4 – connector mechanicals for copper (allows for backward compatibility)
  - SFF-8092 MDI considered for 100GBASE-CR10: (proposals evaluated in IBTA)

- Optional FEC sublayer - PCS to interface to optional FEC sublayer - consider Clause 74 specification – commonality with 40 GbE backplane proposal

- Auto-Negotiation – consider Clause 73 specification - negotiate FEC capability through Auto-Negotiation

40GBASE-CR4 and 100GBASE-CR10 layer diagrams

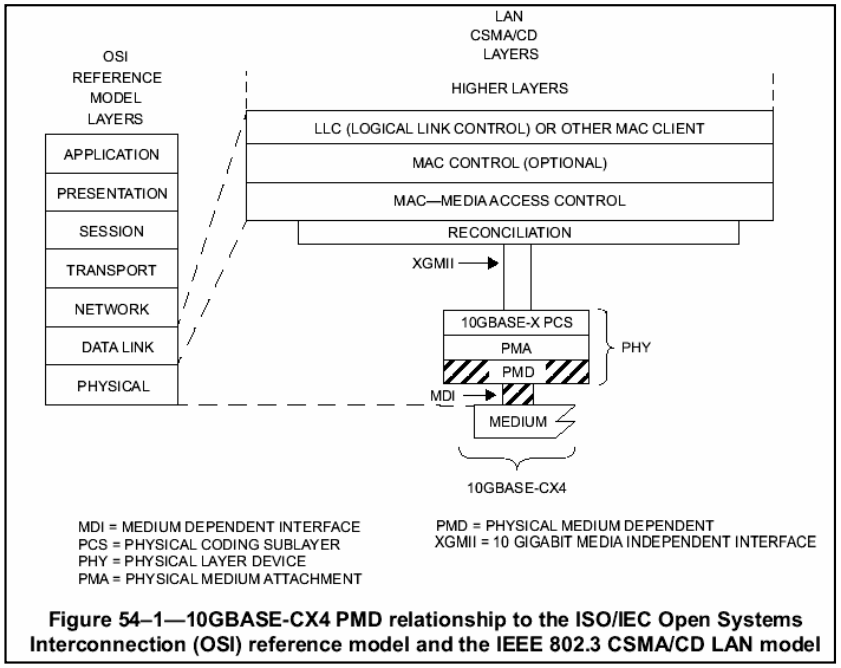


Figure 54–1—10GBASE-CX4 PMD relationship to the ISO/IEC Open Systems Interconnection (OSI) reference model and the IEEE 802.3 CSMA/CD LAN model

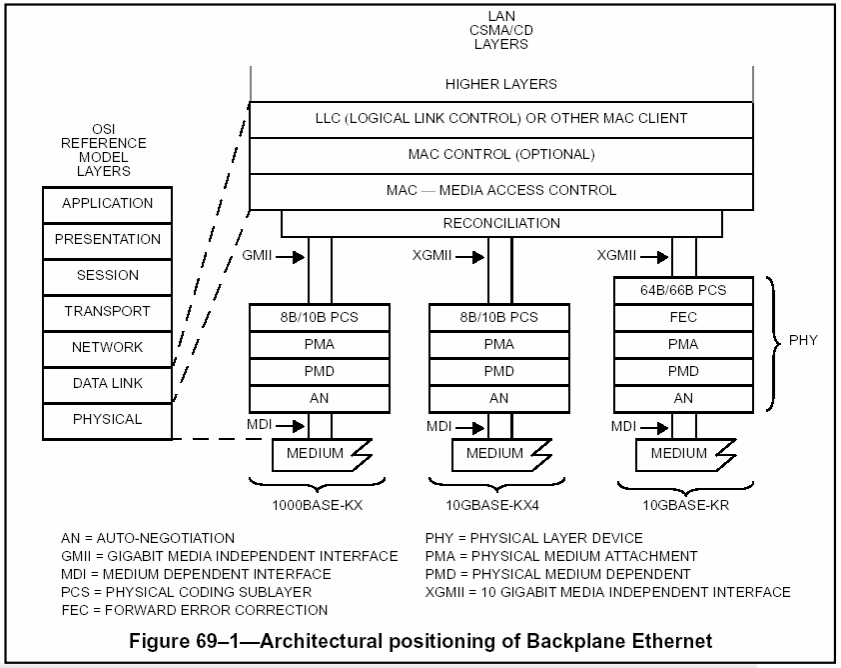
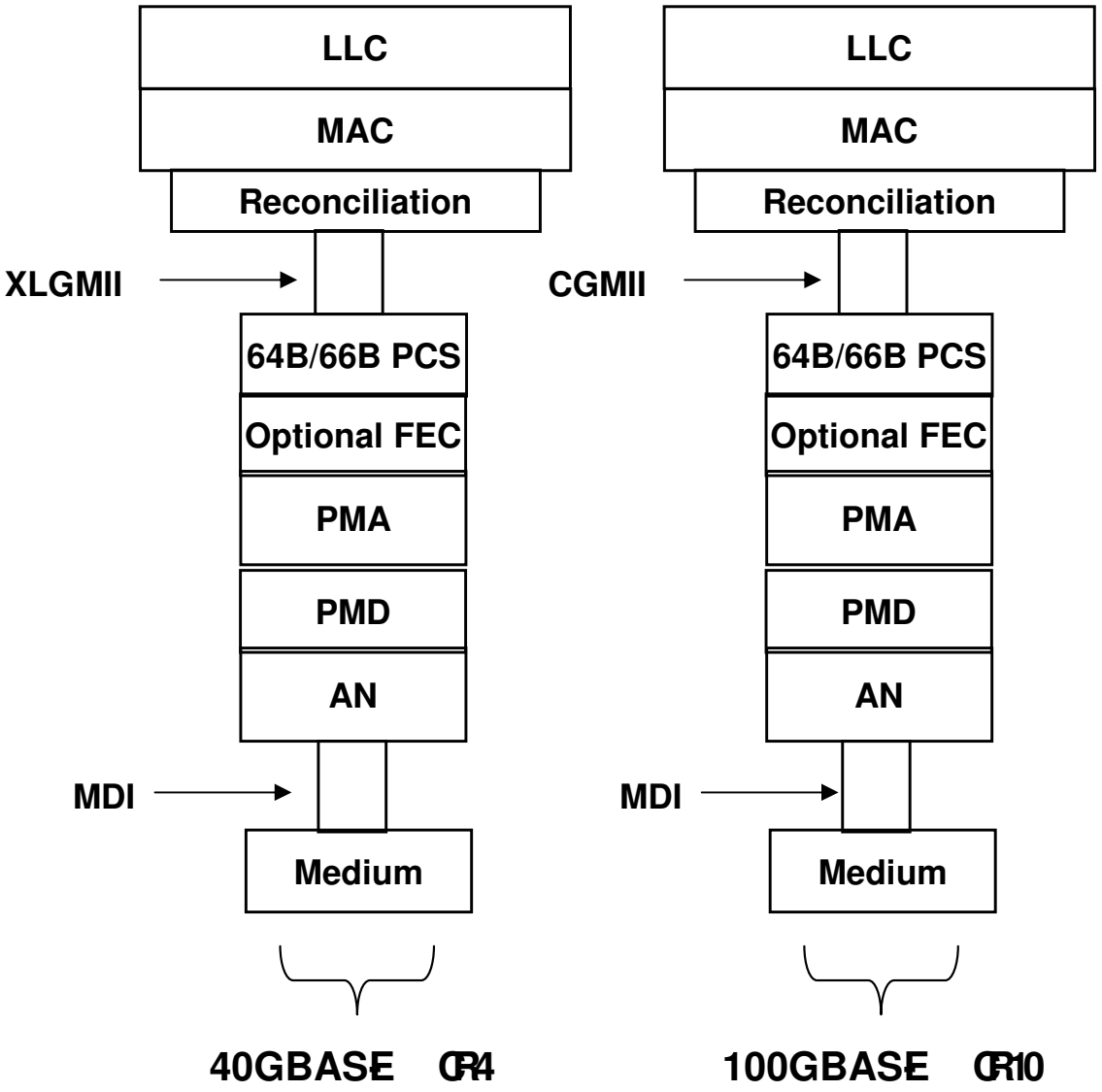


Figure 69–1—Architectural positioning of Backplane Ethernet



802.3ba copper cable assembly link diagram

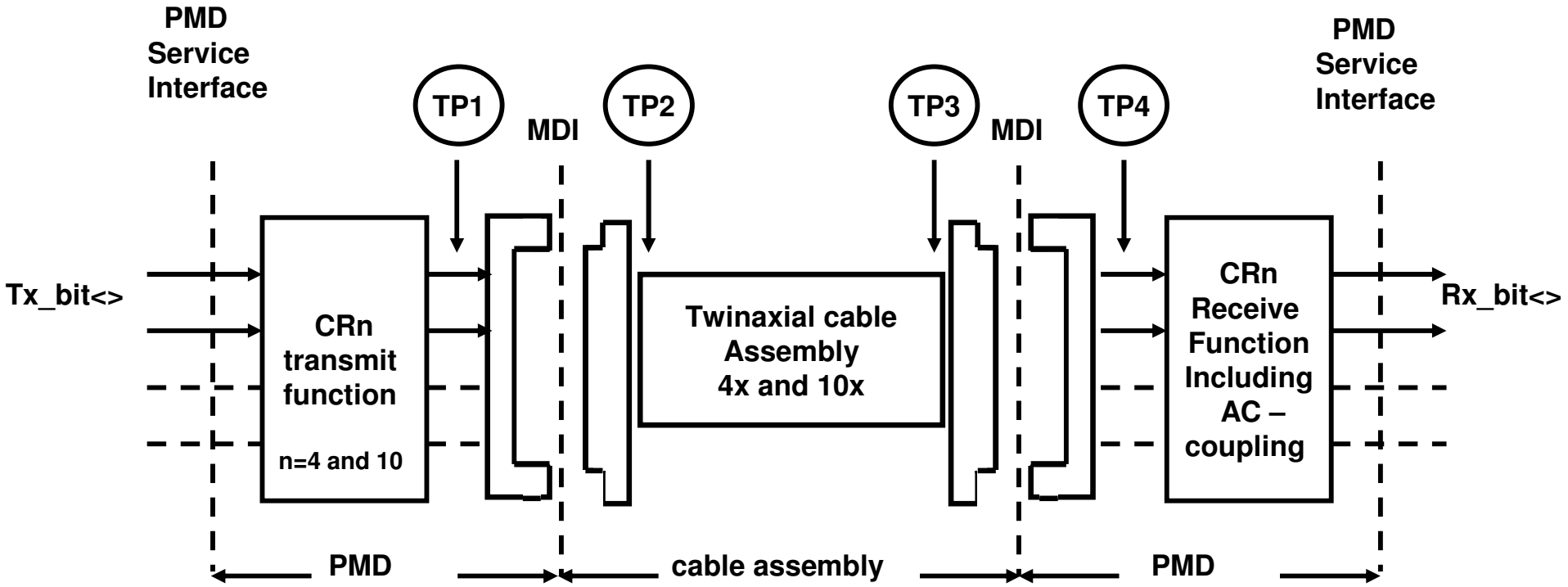


Figure XX–X—40GBASE-CR4 and 100GBASE-CR10 link

40GBASE-CR4 and 100GBASE-CR10 cable assembly

- **Cable assembly differential parameters**

Description	Value	Unit
$Insertion\ \ Loss\ (f) \le TBD\sqrt{f} + TBD \times f + \frac{TBD}{\sqrt{f}}$	TBD	dB
$NextLoss\ (f) \ge TBD - TBD \times \log\left(\frac{f}{TBD}\right)$	TBD	dB
$ReturnLoss(f) \ge TBD$	TBD	dB
$MDNextLoss(f) \ge TBD - TBD \times \log\left(\frac{f}{TBD}\right)$	TBD	dB
$ELFEXT\ \ (f) \ge TBD\ - TBD \times \log\left(\frac{f}{TBD}\right)$	TBD	dB
$MDELFEXT\ (f) \ge TBD - TBD \times \log\left(\frac{f}{TBD}\right)$	TBD	dB

- **TBD's > to be determined from measurement models.**

100GBASE-CR10 - MDI considered: SFF-8092

- SFF-8092 Specification for Mini Multilane Series: Shielded High Density Connector (mechanicals).
  - Scope: The specification defines the plug, guide/strain relief shell, mating interface, footprint, and latching requirements.
  - x12 proposals currently under consideration in IBTA EWG-QDR

- 40GBASE-CR4 and 100GBASE-CR10 for cable assembly differential parameters.

40GBASE-CR4 and 100GBASE-CR10 Auto-Neg

- Adopt Clause 73 (Auto-Negotiation) as a baseline for 40GBASE-CR4 and 100GBASE-CR10 with applicable changes for CR4 and CR10 operation.
  - Use Auto-Neg to Negotiate FEC capability
  - Auto-Neg allows backward compatibility with legacy 10 GbE CX4 PHYs
- Clause 73 provides parallel detection function for compatibility with legacy PHYs that do not support Auto-Negotiation
  - New 40 GbE PHY can use parallel detection for auto-detection of legacy CX4 devices
  - No impact to 10GBASE-CX4 devices

- See ganga_03_05_08.pdf ("FEC and Auto-Neg Proposal for 40/100G Copper Cable Assembly")

40GBASE-CR4 and 100GBASE-CR10 Auto-Neg

- Proposed changes for 40GBASE-CR4 and 100GBASE-CR10
  - Add Technology Ability bits from the reserved space to indicate
    - 40GBASE-CR4 ability
    - 100GBASE-CR10 ability
  - Reuse AN management registers
  - No change to negotiate FEC ability
    - FEC when selected to be enabled on all lanes
    - FEC is enabled when both sides advertise FEC ability and at least one side requests to enable FEC
  - No change to Pause ability and Remote Fault bits
  - Parallel detection function to detect legacy 10GBASE-CX4 PHYs

•See ganga_03_05_08.pdf ("FEC and Auto-Neg Proposal for 40/100G Copper Cable Assembly")

40GBASE-CR4 and 100GBASE-CR10 FEC

- Adopt Clause 74 FEC as baseline for an optional sublayer for 40GBASE-CR4 and 100GBASE-CR10 with appropriate changes for CR4 and CR10 operation.
  - Negotiate FEC capability through Auto-Negotiation
  - FEC is optional- allows lowering BER of 10^{-12} for integration into systems which require lower BER
  - Correction of burst errors up to 11 bits
  - 2-2.5 dB coding gain
  - No penalty in signaling rate
- Enumerate the FEC encode and decode functions for 4 lane and 10 lane operation
  - Each lane is encoded and decoded independently
  - The coding is performed on a virtual lane basis
  - 4 in case of 40 Gb/s
  - 20 in case of 100 Gb/s
- Commonality with 40 Gb/s backplane solution
- Reuse the management register format

- See ganga_03_0508.pdf ("FEC and Auto-Neg Proposal for 40/100G Copper Cable Assembly") for details
- See http://www.ieee802.org/802_tutorials/july06/10GBASE-KR_FEC_Tutorial_1407.pdf for FEC tutorial

QSFP low speed electrical hardware pins

- **For 40GBASE-CR4 copper QSFP low speed control and sense signals set to non-operational QSFP state**

3.1.1.4 ModPrsL

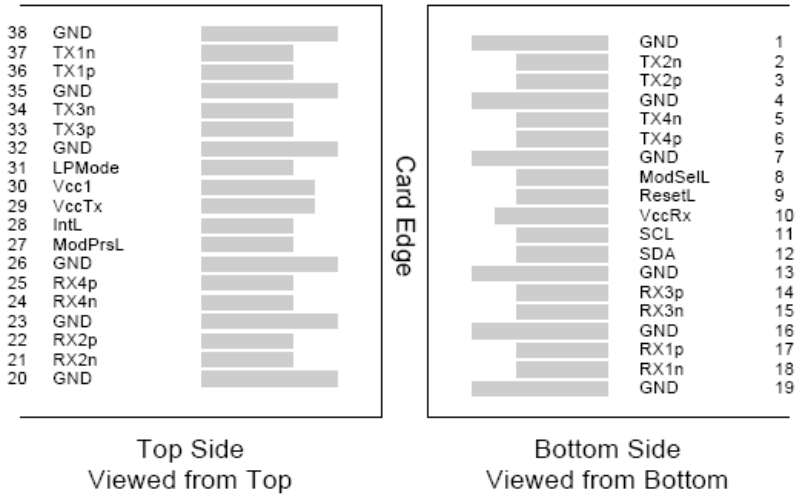
ModPrsL is pulled up to Vcc_Host on the host board and grounded in the module. The ModPrsL is asserted "Low" when inserted and deasserted "High" when the module is physically absent from the host connector.

3.1.1.5 IntL

IntL is an output pin. When "Low", it indicates a possible module operational fault or a status critical to the host system. The host identifies the source of the interrupt using the 2-wire serial interface. The IntL pin is an open collector output and must be pulled to host supply voltage on the host board.

ModPrsL	IntL	Condition	Signal state
1	0	copper module	ModPrsl open, IntL set low
1	1	no module	both signals open
0	x	module present	ModPrsL set low, IntL either state operational

Figure 2 — QSFP Transceiver Pad Layout



10GBASE-KR cable assembly demonstration - Intel

- **10GBASE-KR based device tested over 10 m passive copper assembly under the following setup and conditions**
  - **10 meter Leoni 26 AWG passive twinaxial cable with 2 x SFP+ connectors and 1 x 2" and 1 x 4" FR-4 traces on test boards (6" total); ~5 dB worse @ 5 GHz than QSFP 10 meter cable assembly**
  - **Single NEXT aggressor**
  - **Adaptive TXFFE with the 10GBASE-KR protocol**
  - **5-tap DFE at the receiver**
- **Test results**
  - **BER=0 with PRBS31 was measured for 1500 seconds**
- **Summary**
  - **Feasibility demonstrated at 10 Gb/s, very promising results with single NEXT aggressor**
  - **Margin should be sufficient for QSFP Xtalk environment**

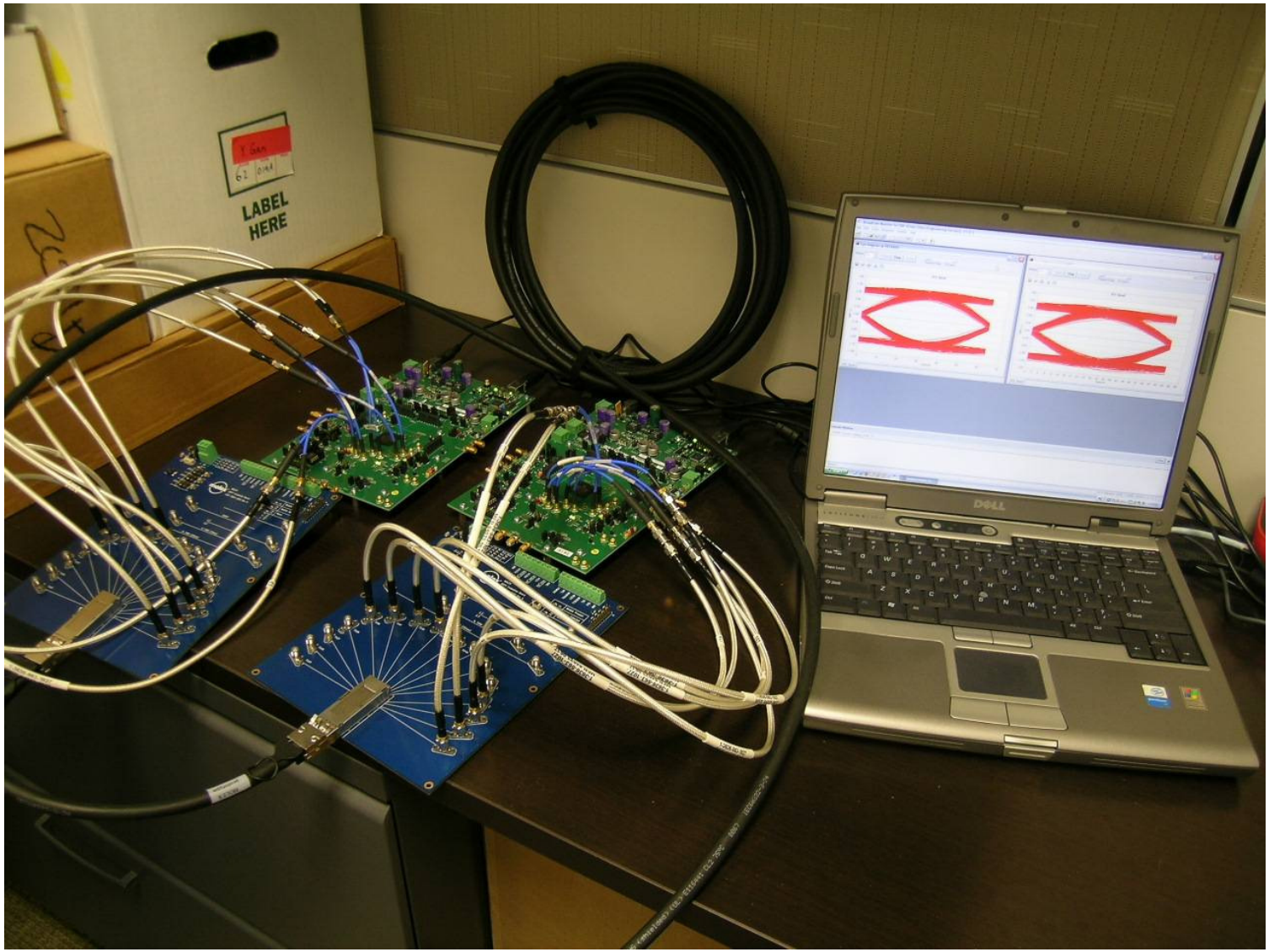
Source: Amir Mezer, Intel

10GBASE-KR cable assembly demonstration - Broadcom

- **10 meter QSFP passive cable assembly including test fixtures (Molex connectors and Leoni cable) tested with Broadcom PHYs designed for compliance to the 10GBASE-KR specification.**
- **Operation over two lanes; simultaneously transmitting and receiving.**
- **Lanes selected in closest proximity; pair-to-pair crosstalk but not multi-disturber.**
- **Additional insertion loss of demonstration:**
  - **2 x device verification board trace (2x (1-1.5 in))**
  - **2 x 2 ft SMA cables, 2 x .5" SMP cables**
- **Test ran for more than a day with 0 errors exceeded 10-12 BER objective.**
- **10 meter QSFP passive cable assembly including test fixtures:**
  - **utilized in the 802.3ap ICR analysis validating 802.3ap KR operation over 10 meters of twinaxial copper cable assembly**
  - **utilized to generate measurement models for Broadcom simulations**

Source: Vivek Telang, Broadcom

10GBASE-KR demonstration setup - Broadcom



Source: Vivek Telang, Broadcom

40GBASE-CR4 and 100GBASE-CR10 baseline

- Adopt 10GBASE-KR electrical specifications (Clause 72) for 40GBASE-CR4 and 100GBASE-CR10 baseline electrical specifications with applicable revisions to account for differences in channel parameters e.g., copper cable assembly versus backplane and the 4-lane and 10-lane operation versus serial operation.

- Adopt 10GBASE-CX4 (Clause 54) cable assembly characteristic transmission parameters for 40GBASE-CR4 and 100GBASE-CR10 with TBD's (slide 10) to be determined from measurement models utilized in the feasibility analysis with the additional consideration of specifying the PCB loss between the transmit function and TP1 and the PCB loss between the receiver function block and TP4 in Figure xx-x. TP2 will be used as test reference point for the transmit function which will include the additional specified PCB loss to be measured with the appropriate test fixture; nicholl_01_0708.pdf to be used as guidance on minimum PCB length. The channel parameters are expected to fall within the high confidence region as defined for 10GBASE-KR in 802.3ap Annex 69B.

- Adopt 40GBASE-CR4 (x4) MDI - QSFP and IEC 61076-3-113 mechanical mating interface (10GBASE-CX4 mechanical) and 100GBASE-CR10 MDI - SFF-8092 Specification for Mini Multilane Series: Shielded High Density Connector (mechanicals).

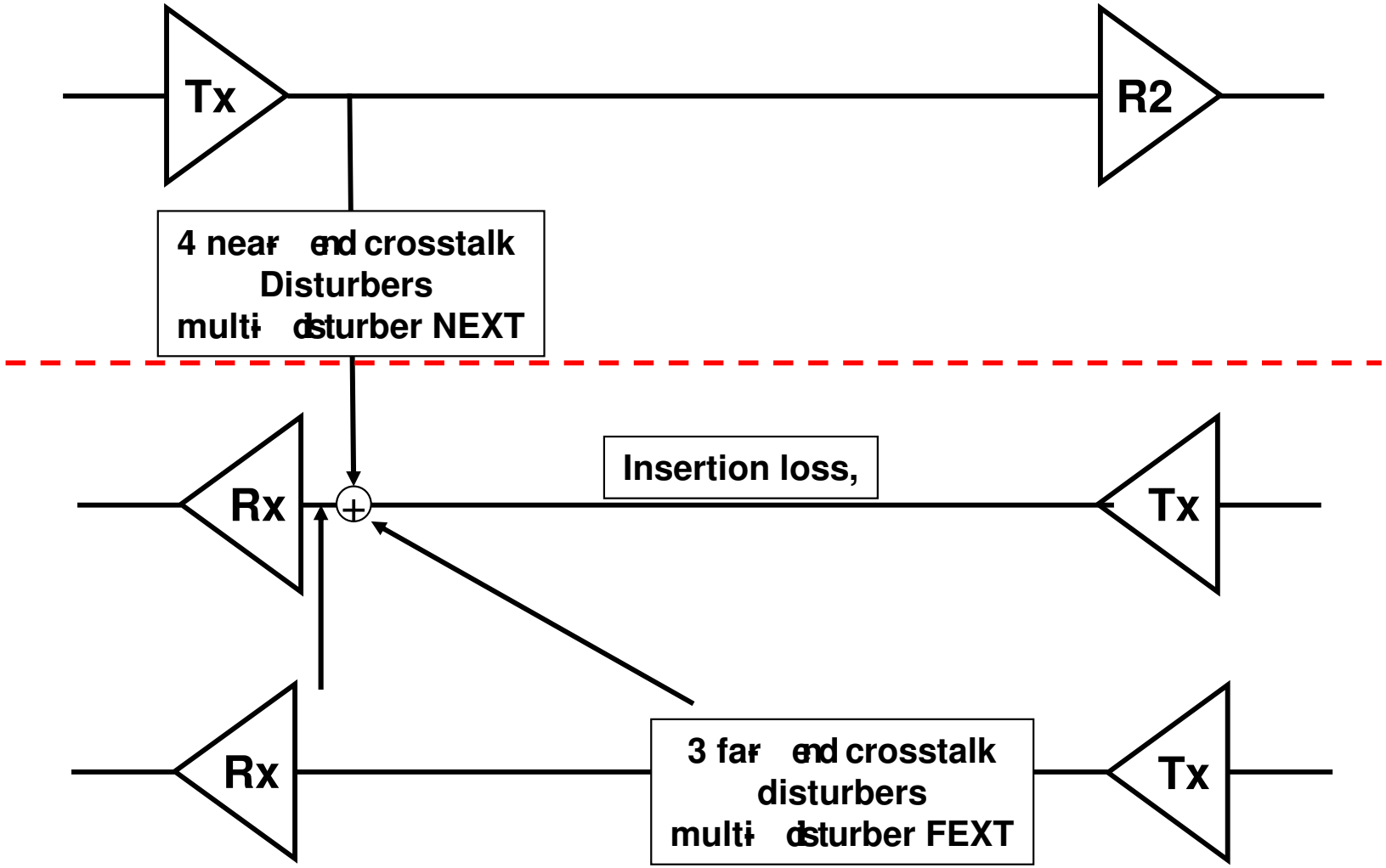
- Adopt Clause 73 (Auto-Negotiation) as a baseline for 40GBASE-CR4 and 100GBASE-CR10 with appropriate changes for CR4 and CR10 operation.

- Adopt Clause 74 FEC as baseline for an optional sublayer for 40GBASE-CR4 and 100GBASE-CR10 with appropriate changes for CR4 and CR10 operation.

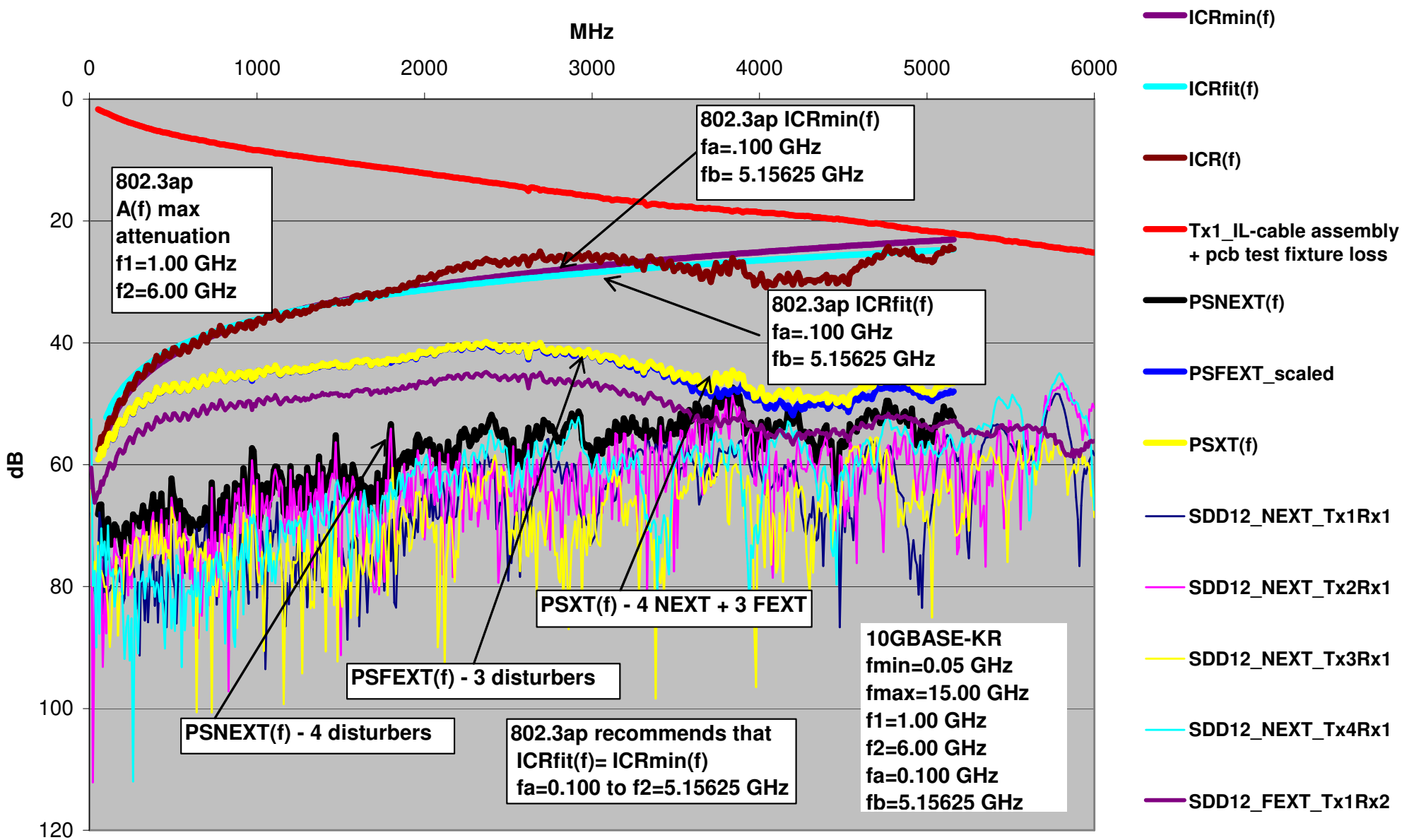
Backup

802.3ap – channel parameter comparisons

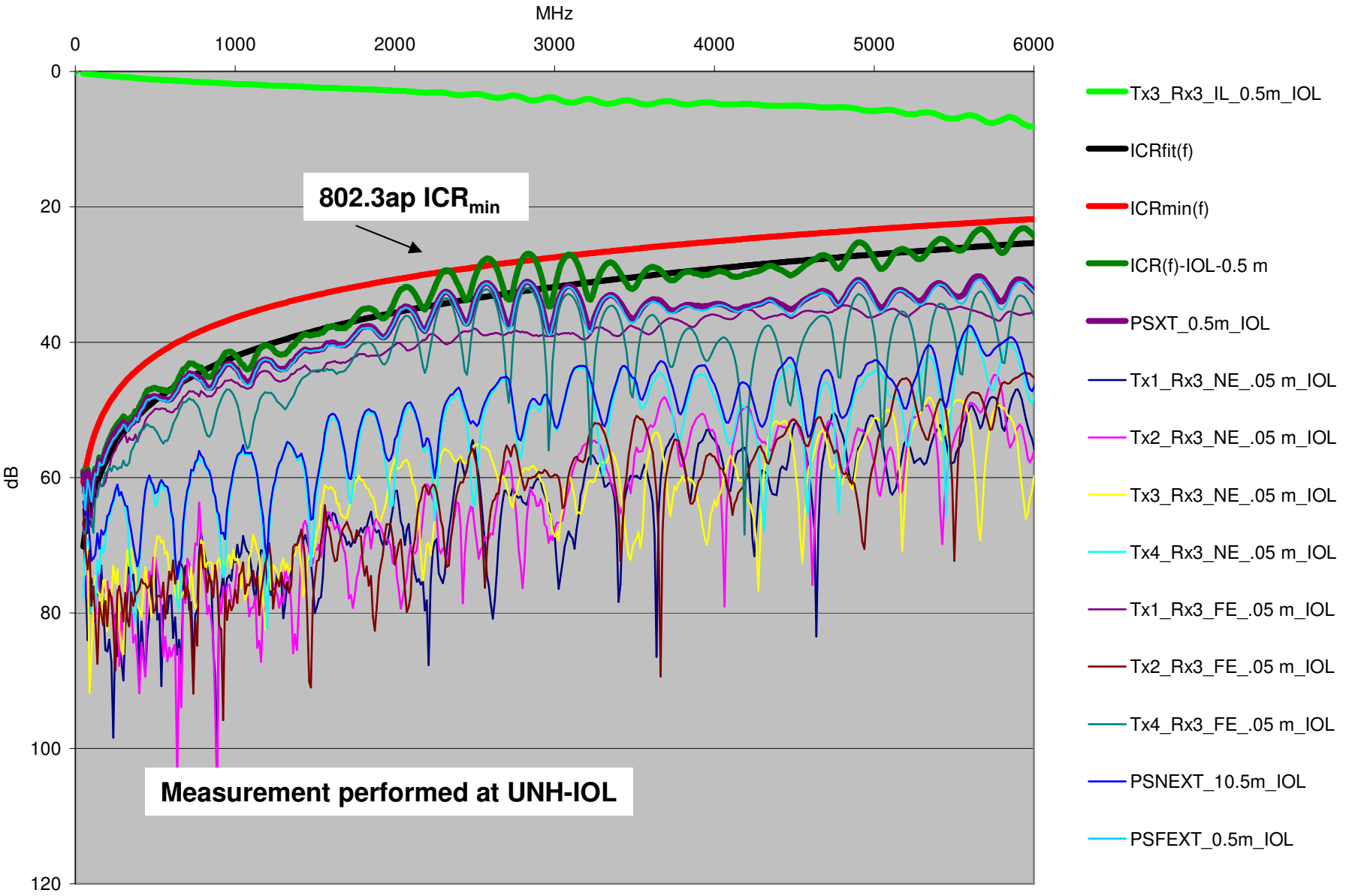
- Insertion loss to crosstalk ratio (ICR) computed from S-parameter measurements and models of QSFP 10 meter copper cable assembly (24 AWG).



802.3ap ICR limits vs 10 m QSFP cable assembly 24 AWG including test fixture



802.3ap ICR limits vs 0.5 m QSFP cable assembly 24 AWG including test fixture

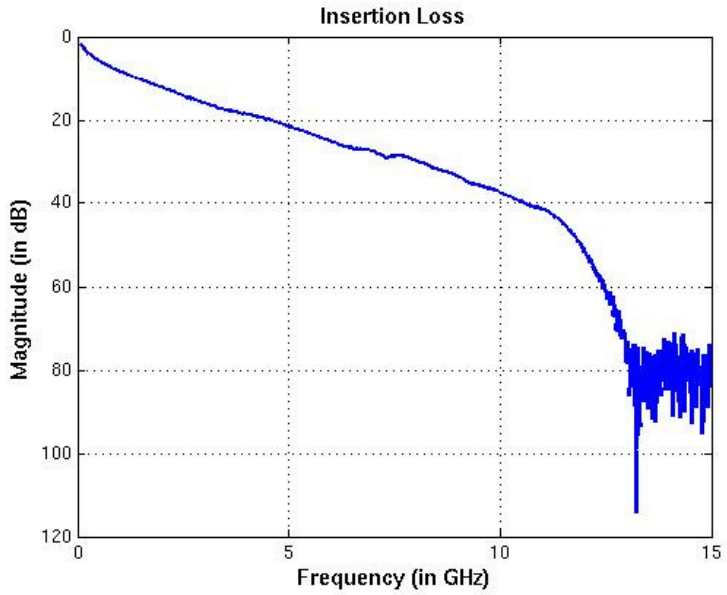


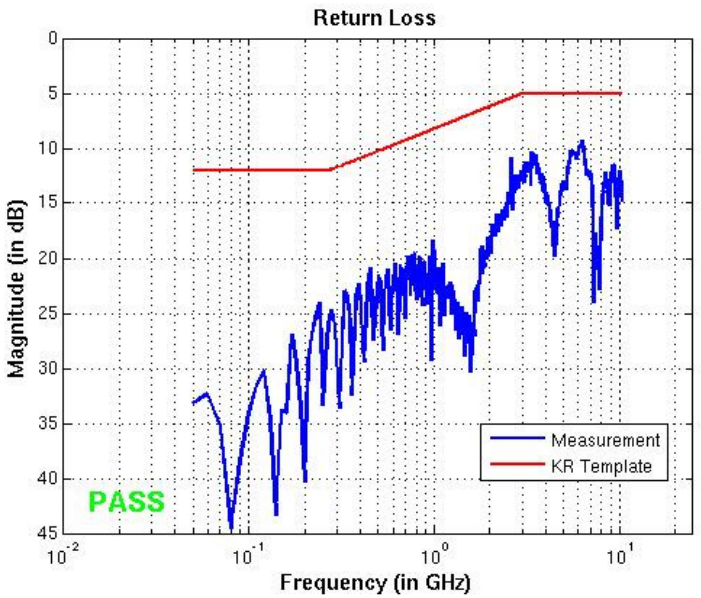
Simulation Setup

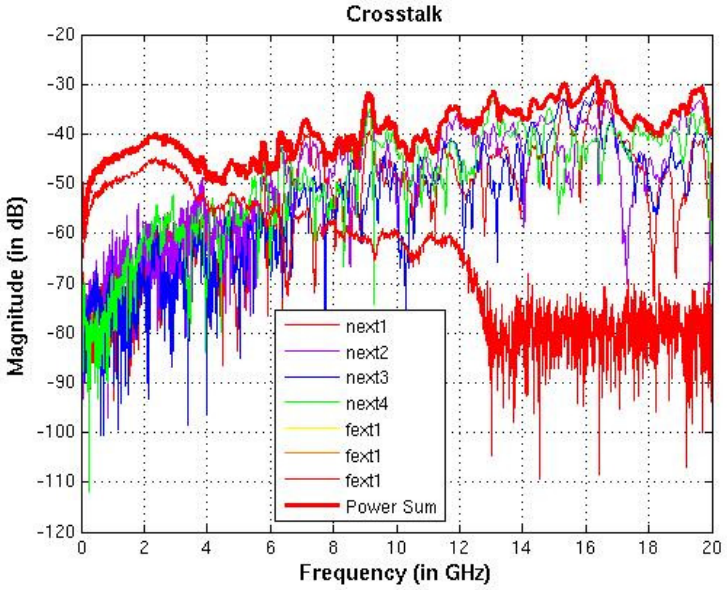
- **Insertion Loss, Return Loss, Crosstalk per data from Chris DiMinico**
- **Package models based on measured data**
- **Receiver architecture same as that used in KR group (802.3ap)**
- **MATLAB simulations**
  - **Pulse Response "Frequency-domain" Analysis, with MMSE optimization**
- **Performance evaluation based on detailed, worst-case error probabilities (not simple Gaussian assumption)**
- **On-chip impairments included**
  - **Clock jitter, Offsets, Front-end noise, Detailed analog circuit models, Detailed equalizer implementation penalties**
- **Worst-casing of ISI data patterns and crosstalk phase**

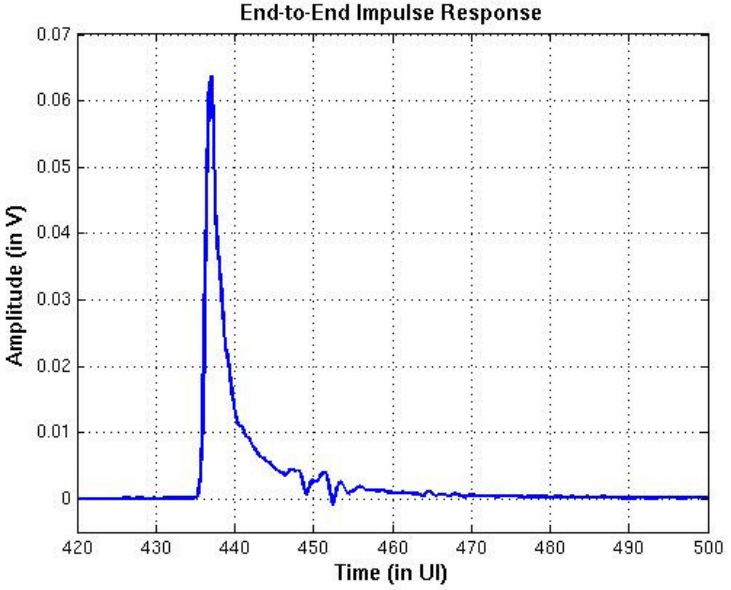
Source: Vivek Telang, Broadcom

Channel models



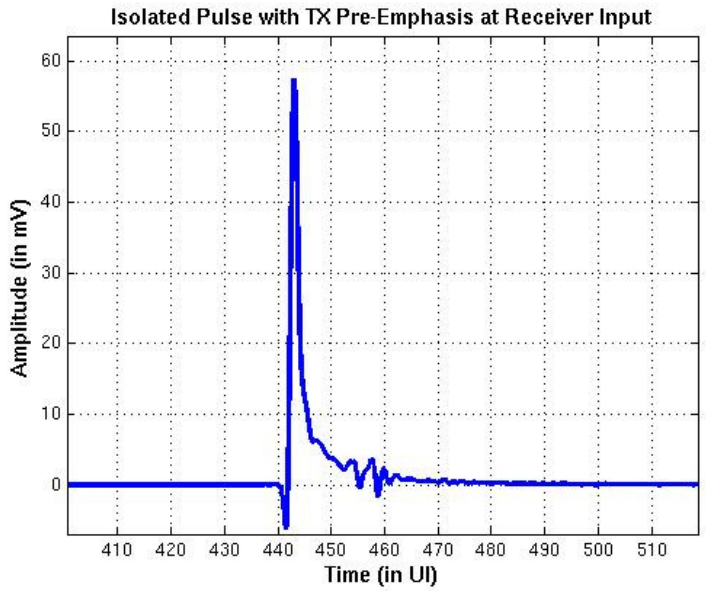




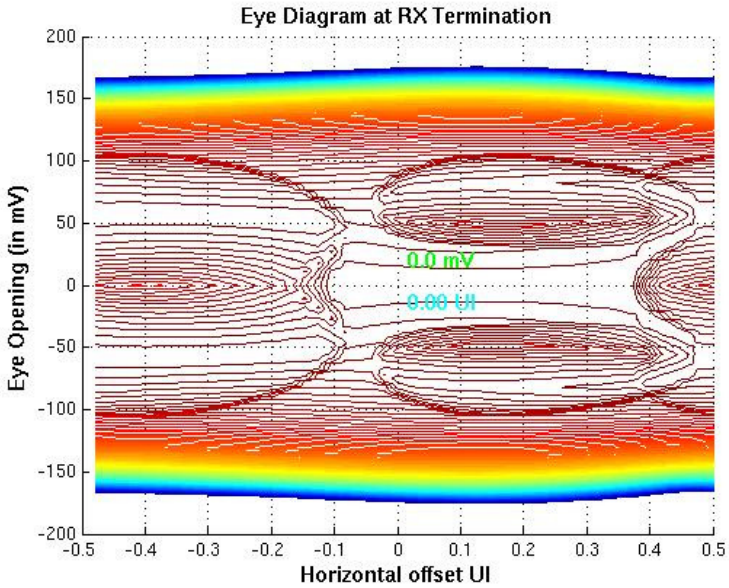


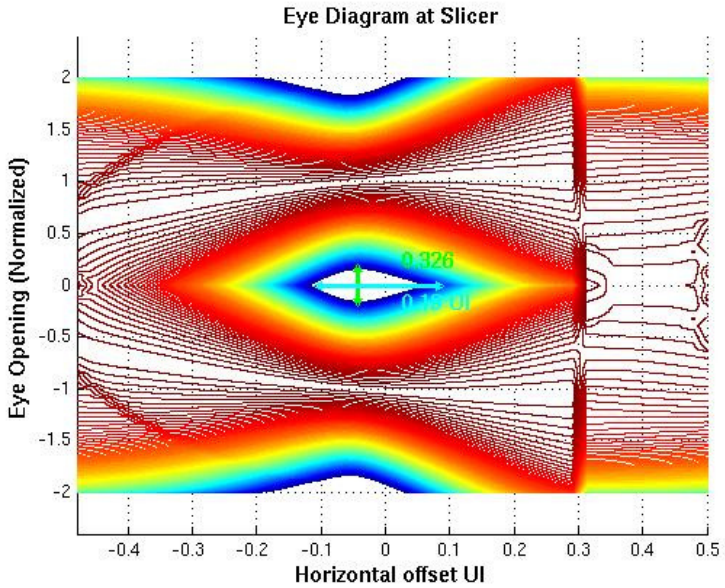
Source: Vivek Telang, Broadcom

Simulation results



Slicer SNR & BER	
SNR (dB)	BER
18.5	1.4×10^{-17}





Source: Vivek Telang, Broadcom

1000BASE-CX (short-haul copper) – MDI

39. Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-CX (short-haul copper)

Connectors meeting the requirements of 39.5.1.1 (Style-1) and 39.5.1.2 (Style-2) shall be used as the mechanical interface between the PMD of 39.3 and the jumper cable assembly of 39.4. The plug connector shall be used on the jumper cable assembly and the receptacle on the PHY. Style-1 or style-2 connectors may be used as the MDI interface. To limit possible cross-plugging with non-1000BASE-CX interfaces that make use of the Style-1 connector, it is recommended that the Style-2 connector be used as the MDI connector.

39.8.3 Major capabilities/options

39.8.4 PICS proforma tables for Physical Medium Dependent (PMD) sublayer and baseband medium, type 1000BASE-CX (short-haul copper)

39.8.4.1 PMD functional specifications

***STY1 Style-1 MDI 39.5 Either the style-1 or the style-2**

MDI must be provided O/1 Yes [ ] No [ ]
***STY2 Style-2 MDI 39.5 O/1 Yes [ ] No [ ]**

40GBASE-LR4 Specification Proposal

IEEE 802.3ba Task Force

16-19 September 2008

Chris Cole – Finisar

Hirotaka Oomori – Sumitomo

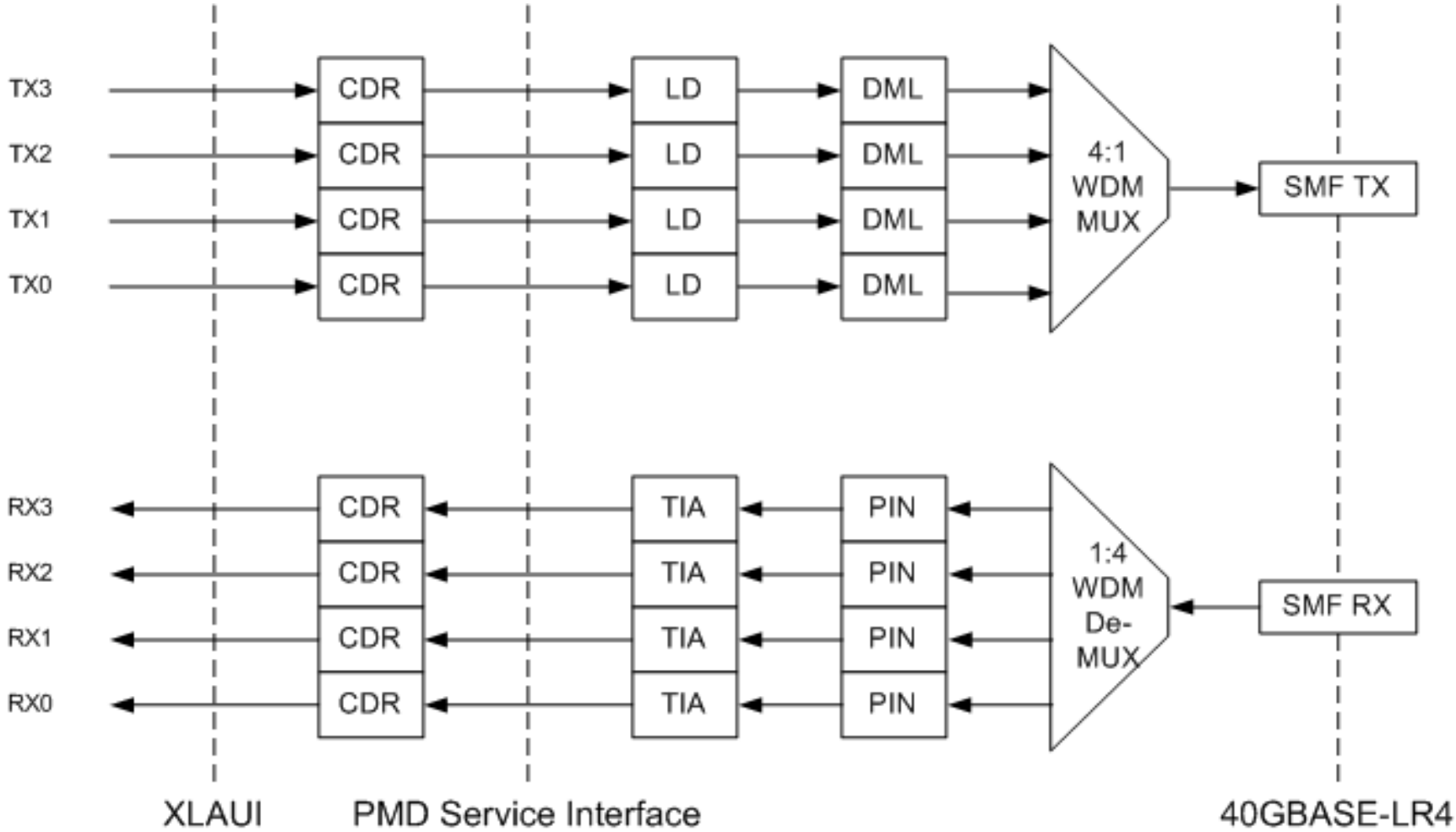
Supporters

- Ghani Abbas – Ericsson
- Justin Abbott – Gennum
- Arne Alping – Ericsson
- Pete Anslow - Nortel
- Alessandro Barbieri – CISCO
- Mike Bennett – LBL
- Francesco Caggioni – AMCC
- Martin Carroll - Verizon
- Adam Carter – CISCO
- Dave D'Andrea - Lightwire
- Dan Dove - HP
- Ali Ghiasi – Broadcom
- Joel Goergen – Force10
- Larry Green – Ixia
- Robert Hays – Intel
- John Jaeger – Infinera
- John Johnson - CyOptics
- Jonathan King – Finisar
- Ryan Latchman – Gennum
- Donn Lee - Facebook
- Jeff Maki - Juniper
- Arlon Martin – Kotura
- Shantanu Mitra – Netlogic
- Andy Moorwood – Infinera
- Jay Moral - AOL
- Ed Nakamoto - Spirent
- David Ofelt – Juniper
- Thomas Paatzsch - CubeOptics
- Shashi Patel – Foundry
- Petar Pepeljugoski - IBM
- Bill Ryan – Foundry
- Ted Seely - Sprint
- Jesse Simsarian - Alcatel-Lucent
- Henk Steenman - AMS-IX
- Steve Trowbridge - Alcatel-Lucent
- Bill Trubey – Time-Warner
- Eddie Tsumura – ExceLight
- Jason Weil - Cox

Introduction

- The 40GBASE-LR4 proposal closely follows 10GBASE-LR 802.3ae specification to provide maximum re-use of existing technology, development, and test infrastructure.
- The major changes from 10GBASE-LR are:
  - update to link power budget to account for improvement in DFB technology (speed,)
  - update to TP2 and TP3 power levels to account for CWDM Mux and DeMux losses.
- The change from cole_03_0708 is 0.5dB increase in TP2 power range.
- All numbers should be viewed as subject to change/refinement as a result of detailed review and discussion by 802.3ba participants.
- Support of this presentation is for a baseline for the 40GBASE 10km SMF PMD objective, but does not necessarily imply exact agreement with every specification number.

Example 10km 1310nm DML 4x10G Implementation



CWDM Baseline Grid

- ITU G.694.2 specification
- Exact wavelengths: 1271, 1291, 1311, 1331 nm
- Shorthand wavelengths: 1270, 1290, 1310, 1330 nm
- TX and RX wavelength range: 13 nm
- G.652 A&B 10km SMF worst dispersion and fiber loss
  - Max positive dispersion (1337.5nm) = 33ps/nm
  - Max negative dispersion (1264.5nm) = -59ps/nm
  - Max Loss (1337.5nm) = 4.3dB
  - Max Loss (1264.5nm) = 4.7dB

40GBASE-LR4 lane assignments

Lane	Center wavelengths	Wavelength ranges
L_0	1271 nm	1264.5 – 1277.5 nm
L_1	1291 nm	1284.5 – 1297.5 nm
L_2	1311 nm	1304.5 – 1317.5 nm
L_3	1331 nm	1324.5 – 1337.5 nm

40GBASE-LR4 transmit characteristics

Description	40GBASE-LR4	Unit
Signaling speed per lane	10.3125 ±100 ppm	GBd
Lane wavelengths (range)	1264.5 – 1277.5 1284.5 – 1297.5 1304.5 – 1317.5 1324.5 – 1337.5	nm
Transmitter eye mask definition {X1, X2, X3, Y1, Y2, Y3} [a]	TBD	
Side Mode Suppression Ratio (SMSR), (min)	30	dB
Total average launch power (max)	8.3	dBm
Average launch power per lane (max)	2.3	dBm
Average launch power per lane (min)[b]	-7.0	dBm
Optical Modulation Amplitude (OMA) - TDP, per lane (min)	-4.8	dBm
Optical Modulation Amplitude (OMA), per lane (min) [c]	-4.0	dBm
Transmitter and dispersion penalty (max)	2.3	dB
Average launch power of OFF transmitter, per lane (max)	-30	dBm
Extinction Ratio (min)	3.5	dB
RIN_{12} OMA (max)	-128	dB/Hz
Optical Return Loss Tolerance (max)	12	dB
Transmitter Reflectance (max) [d]	-12	dB

[a] Tx eye mask spec to be specified as per eye mask methodology discussions

[b] Informative, average launch power (min) and is not the principle indicator of signal strength.

[c] Even if the TDP < 0.8dB, the OMA (min) must exceed this value.

[d] Transmitter reflectance is defined looking into the transmitter.

40GBASE-LR4 receive characteristics

Description	40GBASE-LR4	Unit
Signaling speed per lane	10.3125 ±100ppm	GBd
Lane wavelengths (range)	1264.5 – 1277.5 1284.5 – 1297.5 1304.5 – 1317.5 1324.5 – 1337.5	nm
Average receive power, per lane (max) [a]	2.3	dBm
Average receive power, per lane (min) [b]	-13.7	dBm
Receive sensitivity (OMA), per lane (max)	-11.5	dBm
Return loss (min)[c]	-26	dB
Stressed receive sensitivity (OMA), per lane (max) [d]	-9.9	dBm
Vertical eye closure penalty, per lane [f]	1.6	dB
Receive electrical 3 dB upper cutoff frequency, per lane (max)	12.3	GHz

[a] The receiver shall tolerate, without damage, the Average Receive Power (max) plus 1 dB

[b] Informative, equals min Tx OMA with infinite ER and max channel insertion loss

[c] Prevents excess coherent interference due to Tx Rx reflectance

[d] Measured with conformance test signal at TP3 for BER = 10^{-12}

[f] Informative. Penalty for testing stressed receiver sensitivity

40GBASE-LR4 link power budget

Description	40GBASE-LR4	Unit
Power budget	9.0	dB
Operating distance	10	km
Channel insertion loss [a]	6.7	dB
Maximum Discrete Reflectance (max)	-26	dB
Allocation for penalties (TDP (max)) [b]	2.3 [c]	dB
Additional insertion loss allowed	0.0	dB

[a] Channel insertion loss includes fiber and connector losses for worst case wavelength lane

[b] Dispersion and other penalties for worst case wavelength lane

[c] Assumes Ts = 40ps, 1.6dB ISI Penalty, 0.7dB other penalties.