#	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://grouper.ieee.org/groups/802/3/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://grouper.ieee.org/groups/802/3/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://grouper.ieee.org/groups/802/3/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://grouper.ieee.org/groups/802/3/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://grouper.ieee.org/groups/802/3/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://grouper.ieee.org/groups/802/3/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://grouper.ieee.org/groups/802/3/ba/public/may08/trowbridge_01_0508.pdf
8	100GE 40KM PMD	Move to adopt 4x25G LAN WDM (as per cole_02_0508) as the baseline proposal for the 100GE 40km SMF PMD.	#12 of May 2008	http://grouper.ieee.org/groups/802/3/ba/public/may08/cole_02_0508.pdf
9	100GE 10KM PMD	Move to adopt 4x25G LAN WDM (as per cole_01_0508) as the baseline proposal for 100GE 10km SMF PMD.	#15 of May 2008	http://grouper.ieee.org/groups/802/3/ba/public/may08/cole_01_0508.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://grouper.ieee.org/groups/802/3/ba/public/may08/mellitz_01_0508.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

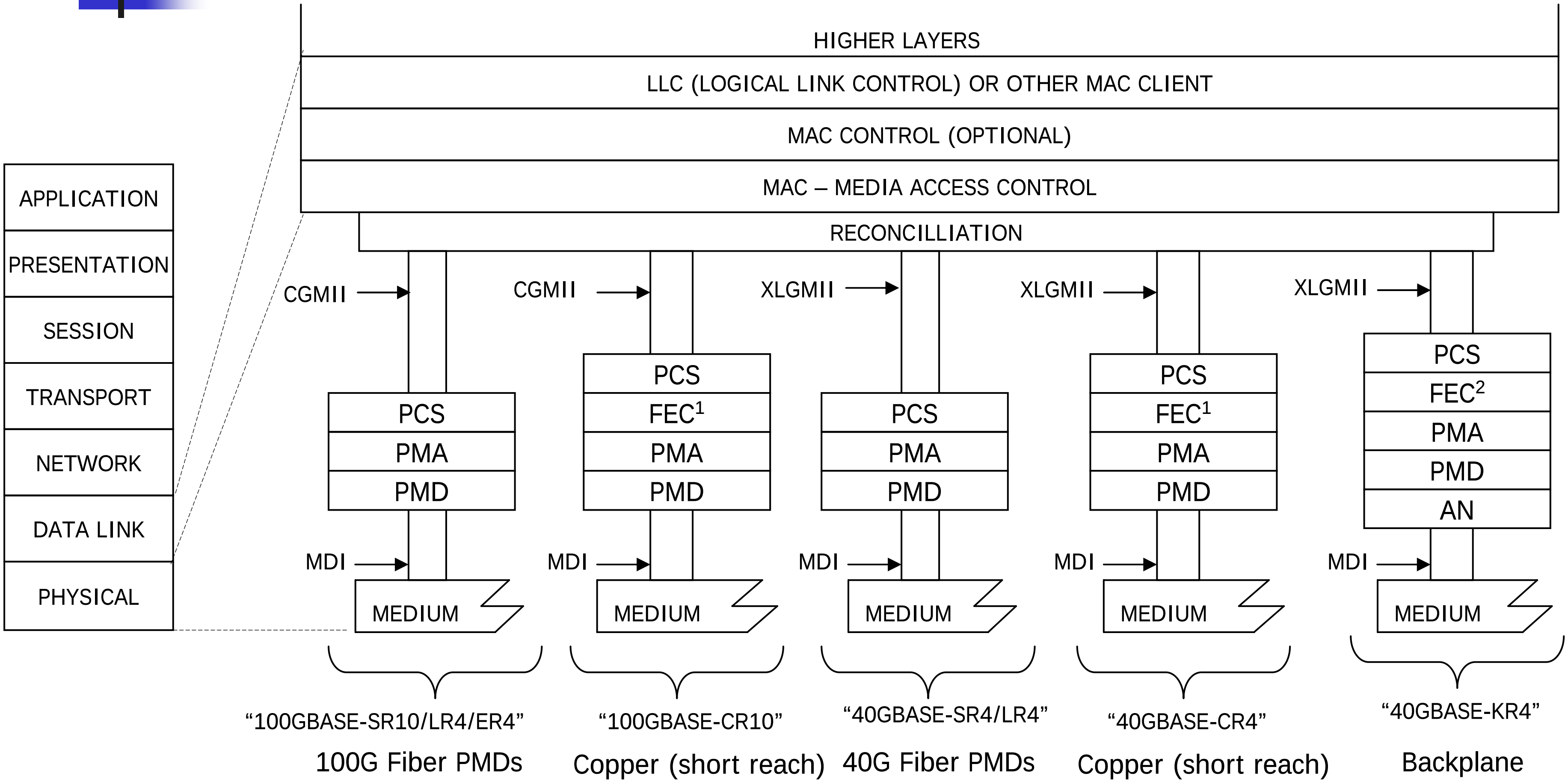
PHY description	**Port Type**
40G Backplane PHY	40GBASE-KR4
40G Cable Assembly PHY 100G Cable Assembly PHY	40GBASE-CR4 100GBASE-CR10
40G MMF 100m PHY (Ribbon) 100G MMF 100m PHY (Ribbon)	40GBASE-SR4 100GBASE-SR10
40G SMF 10Km PHY 100G SMF 10Km PHY	40GBASE-LR4 100GBASE-LR4
100G SMF 40Km PHY	100GBASE-ER4

40/100G Architecture and Interfaces proposal

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

May 13, 2008

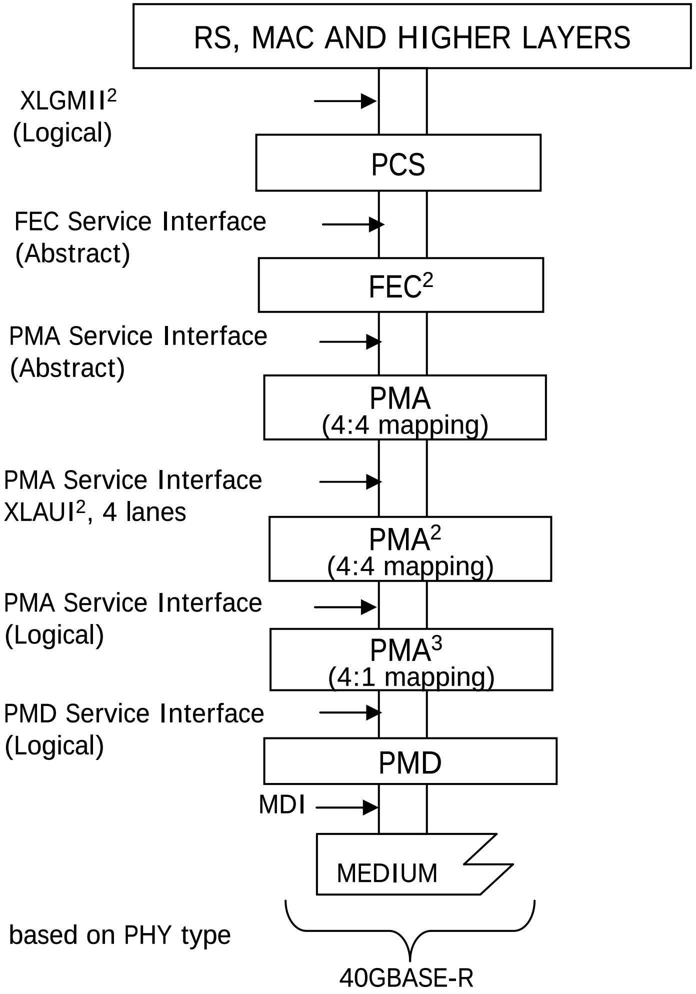
Proposed 40/100GbE layer model



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

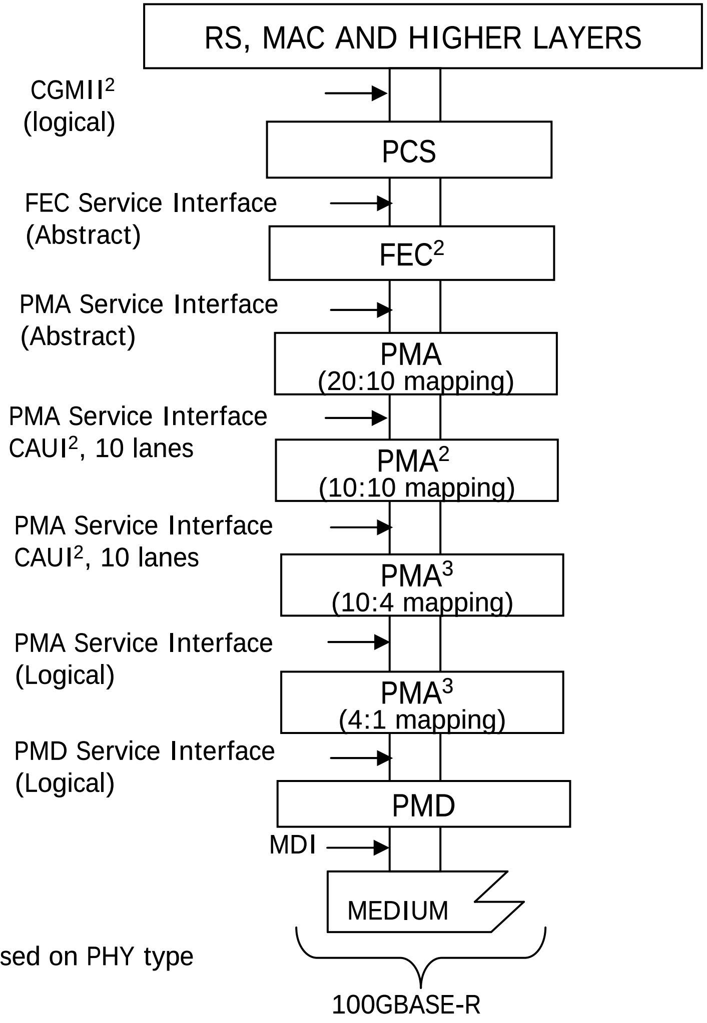


Note: 2. Optional

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



RS, MAC AND HIGHER LAYERS

CGMII[2] (logical)

PCS

FEC Service Interface (Abstract)

FEC[2]

PMA Service Interface (Abstract)

PMA (20:10 mapping)

PMA Service Interface CAUI[2], 10 lanes

PMA[2] (10:10 mapping)

PMA Service Interface CAUI[2], 10 lanes

PMA[3] (10:4 mapping)

PMA Service Interface (Logical)

PMA[3] (4:1 mapping)

PMD Service Interface (Logical)

PMD

MDI

MEDIUM

100GBASE-R

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

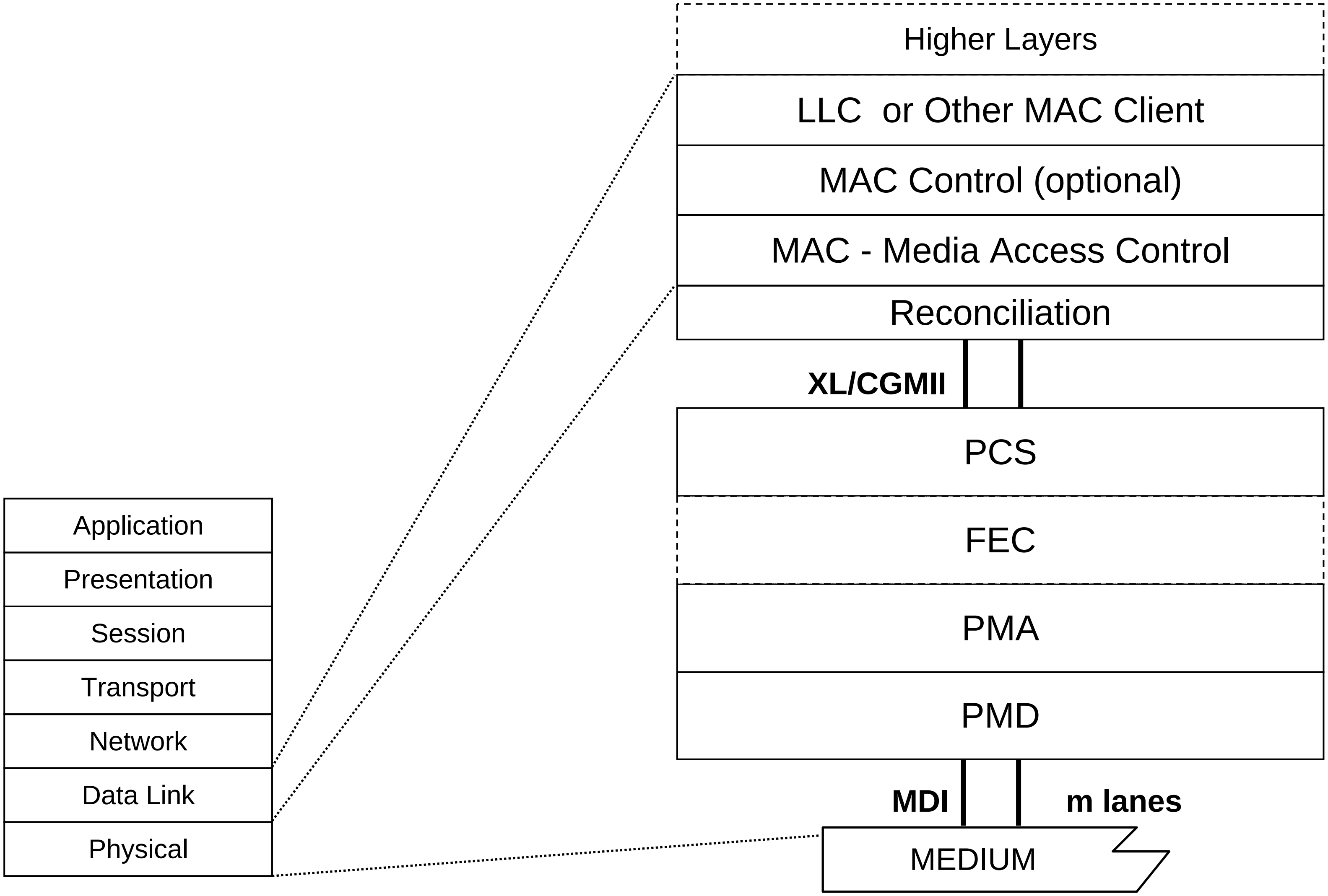
Mark Gustlin - Cisco

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

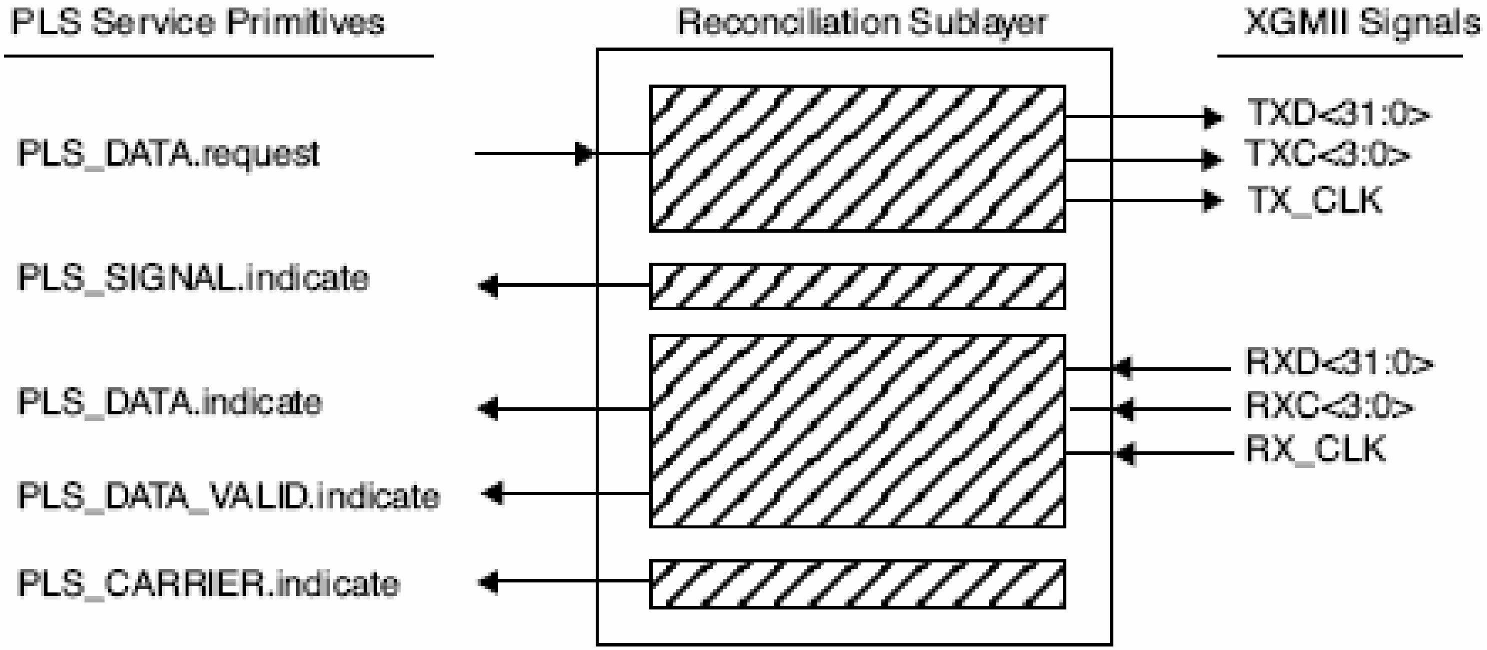
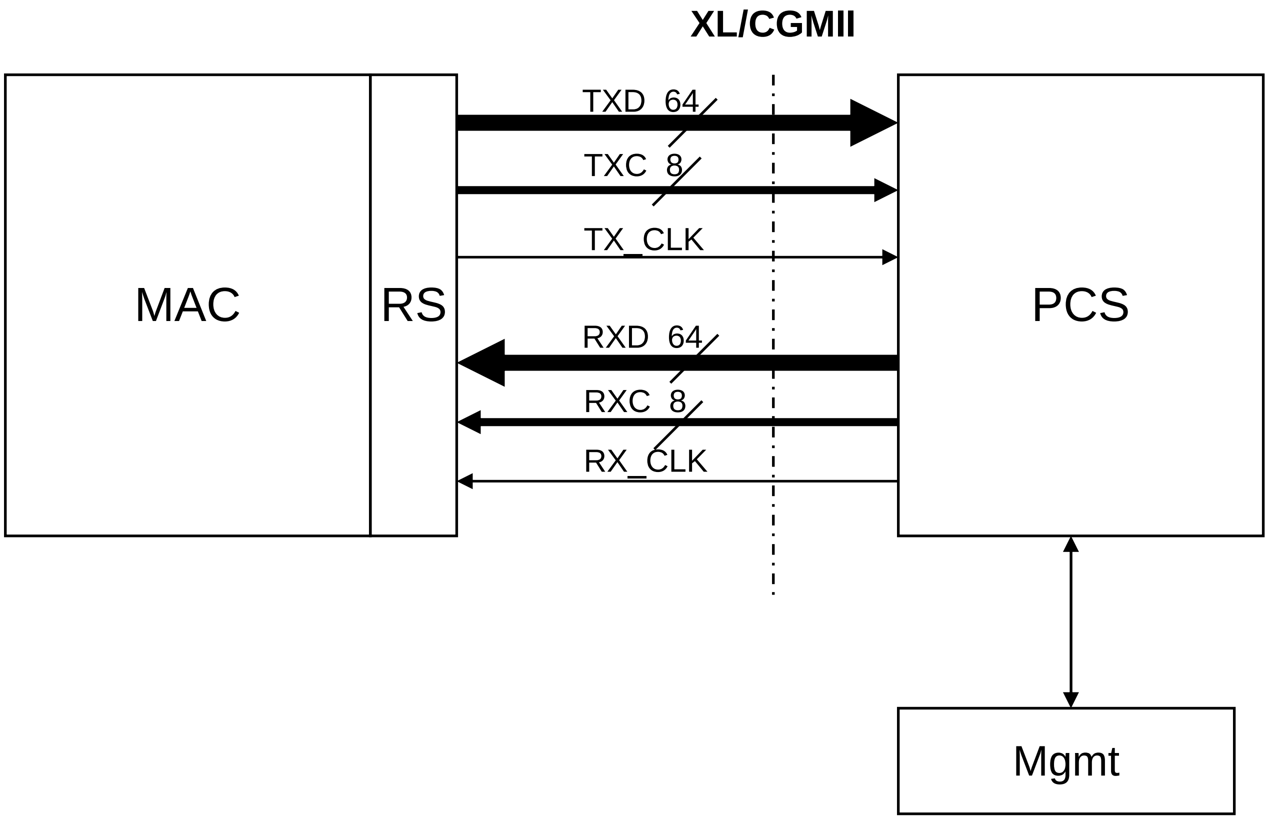


Figure 46–2—Reconciliation Sublayer (RS) inputs and outputs

From 802.3ae

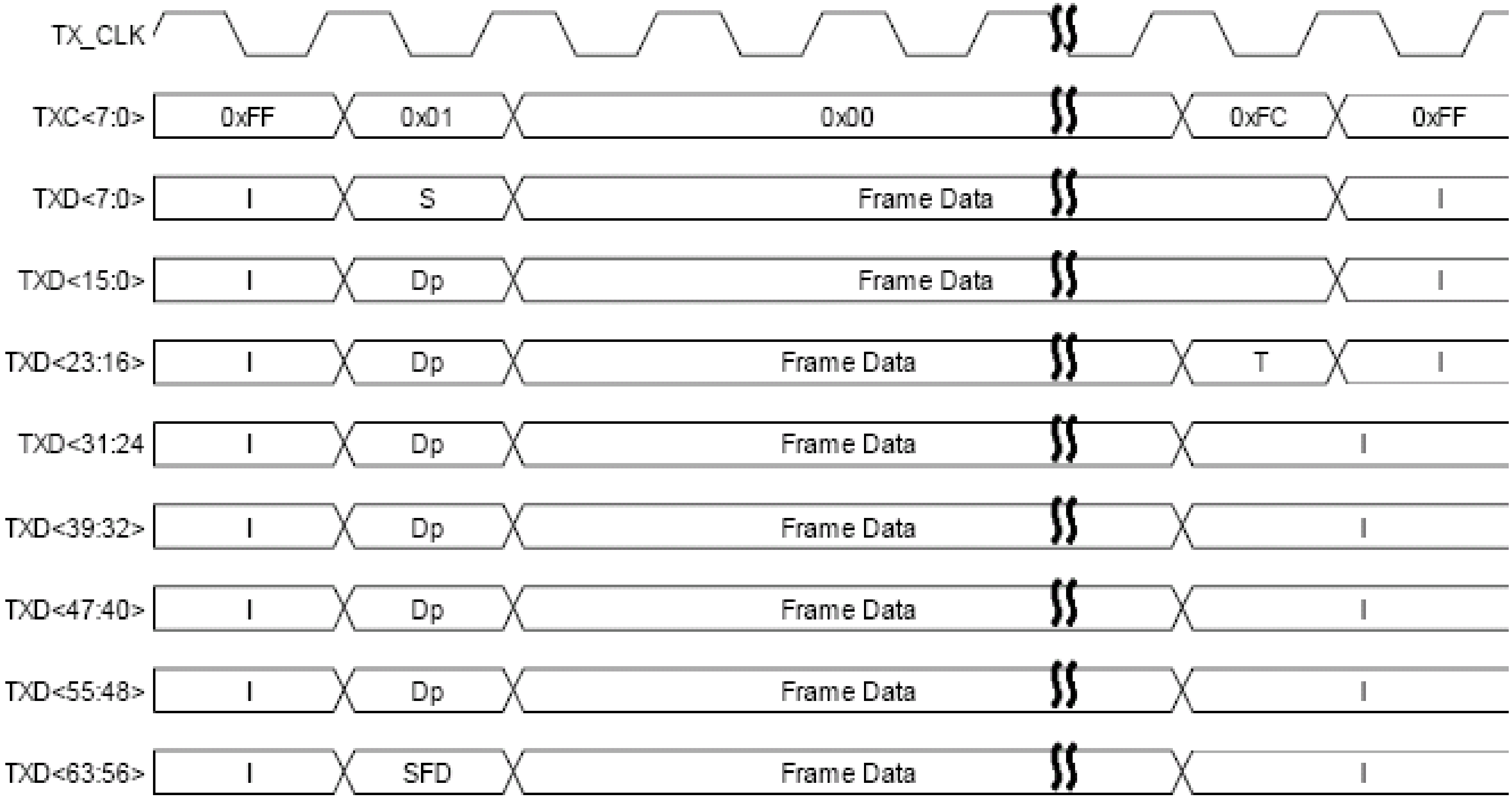
XL/CGMII Interface

- 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

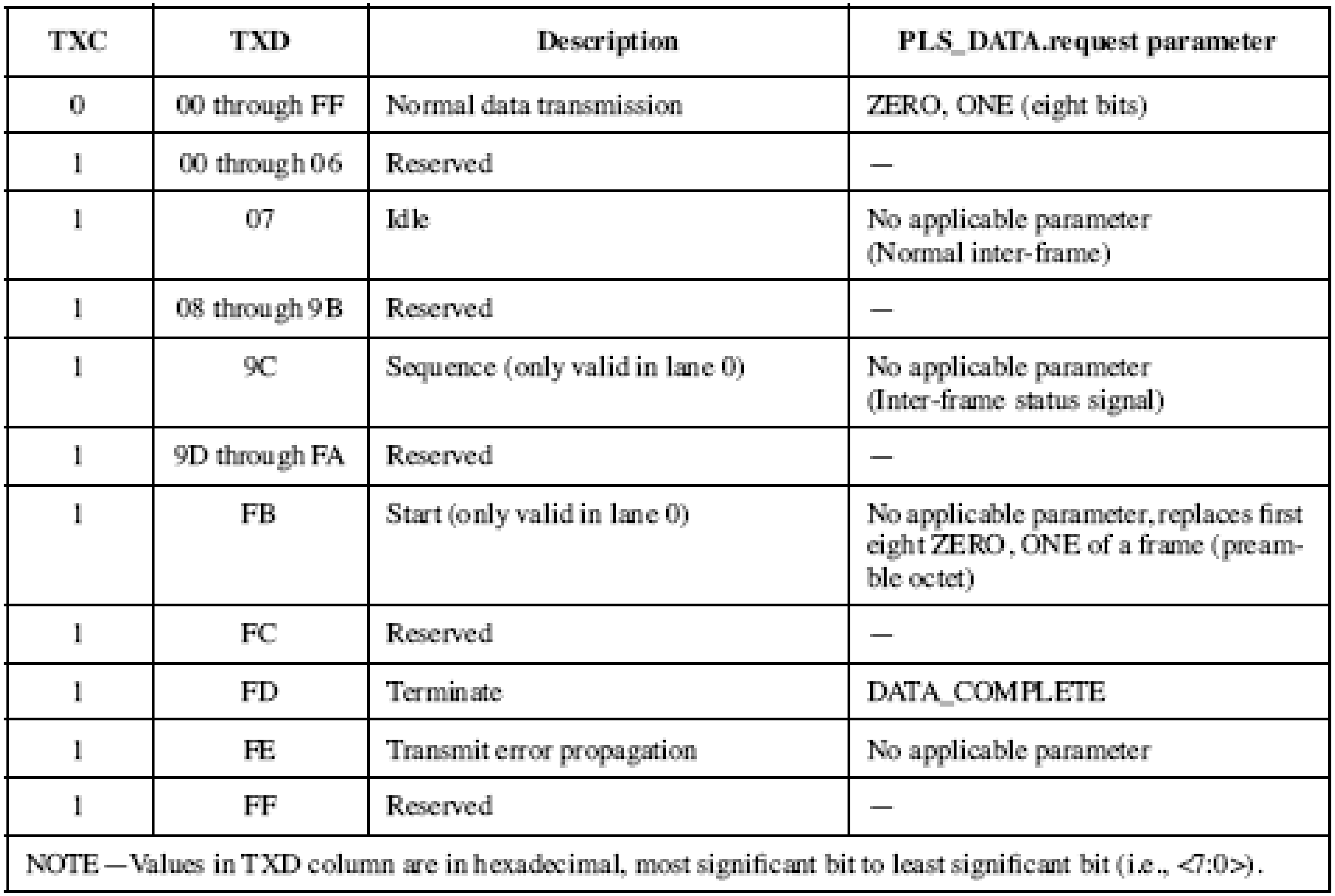


I: Idle control character, S: Start control character, Dp: preamble Data octet, T: Terminate control character, SFD: Start of Frame Delimiter

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

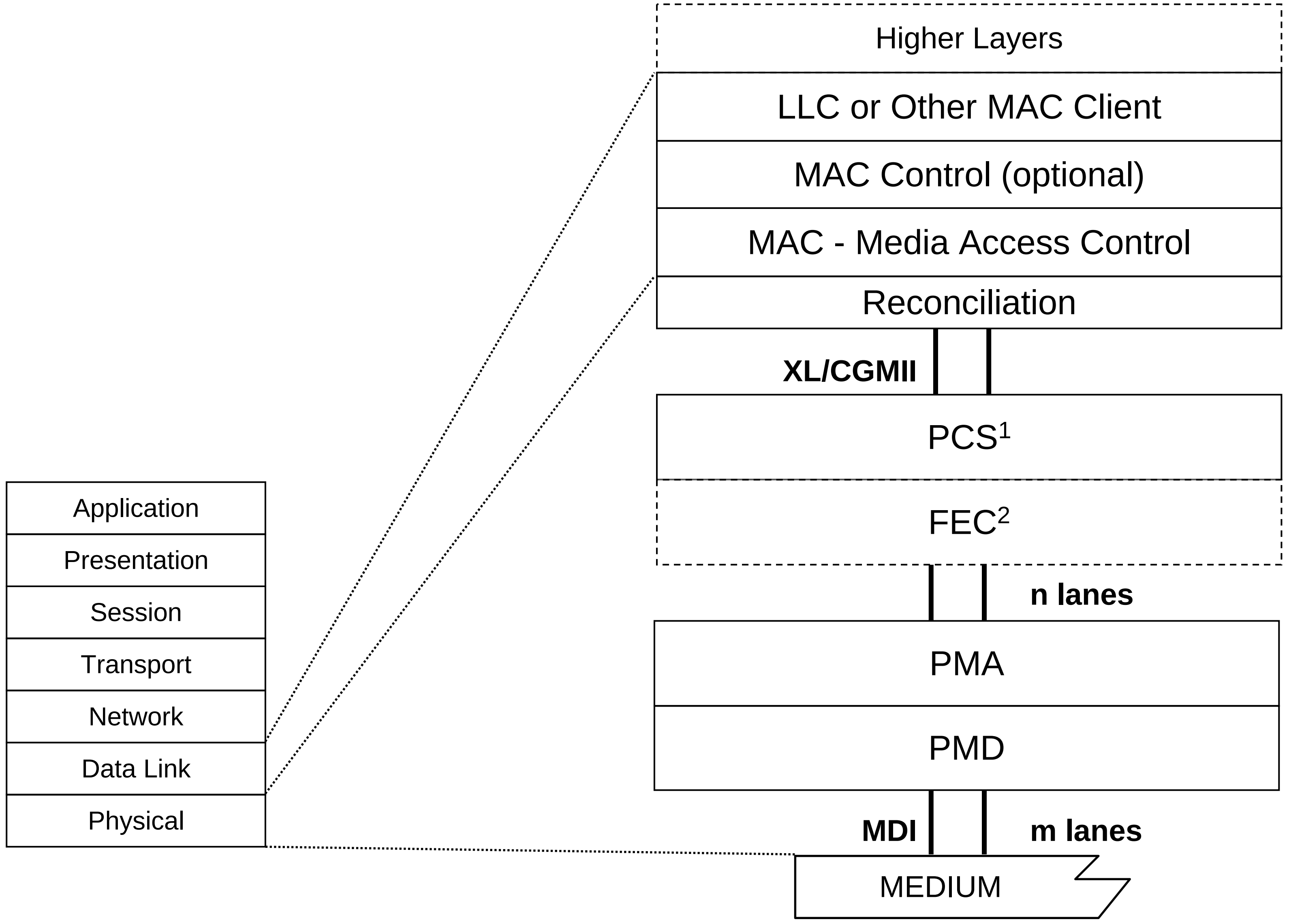
Martin Carroll - Verizon

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



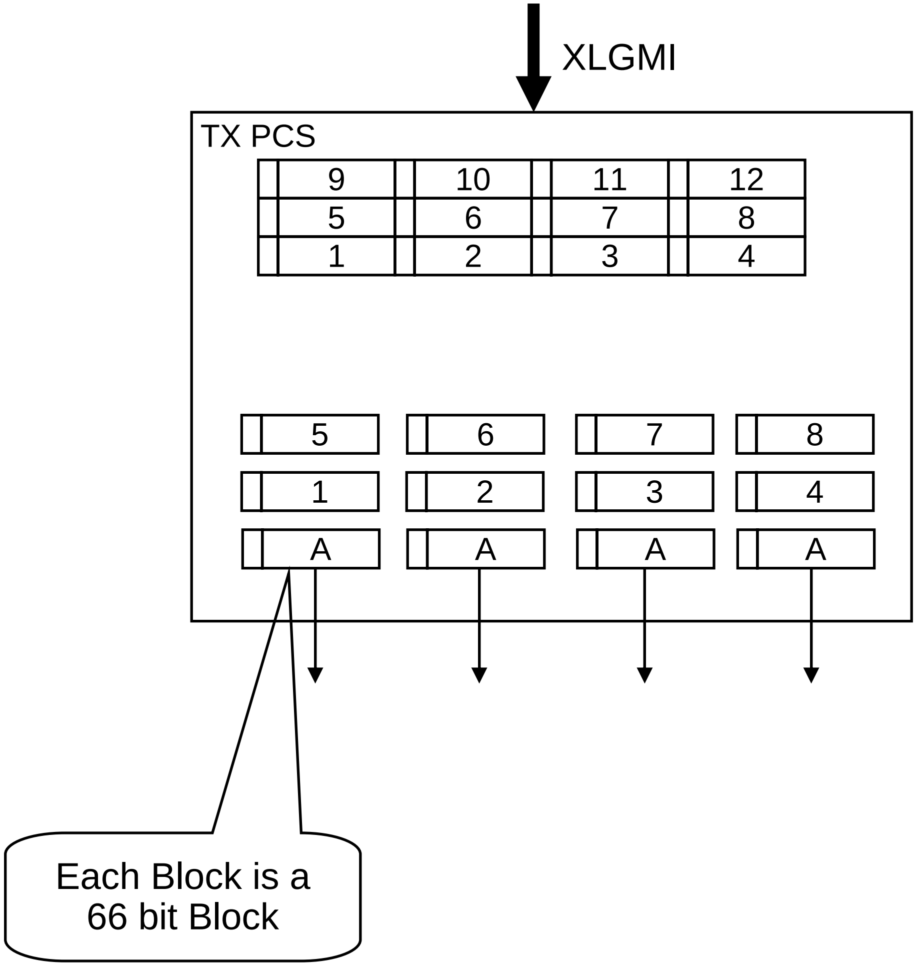
1: Includes MLD functionality 2: For 40GE Backplane

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

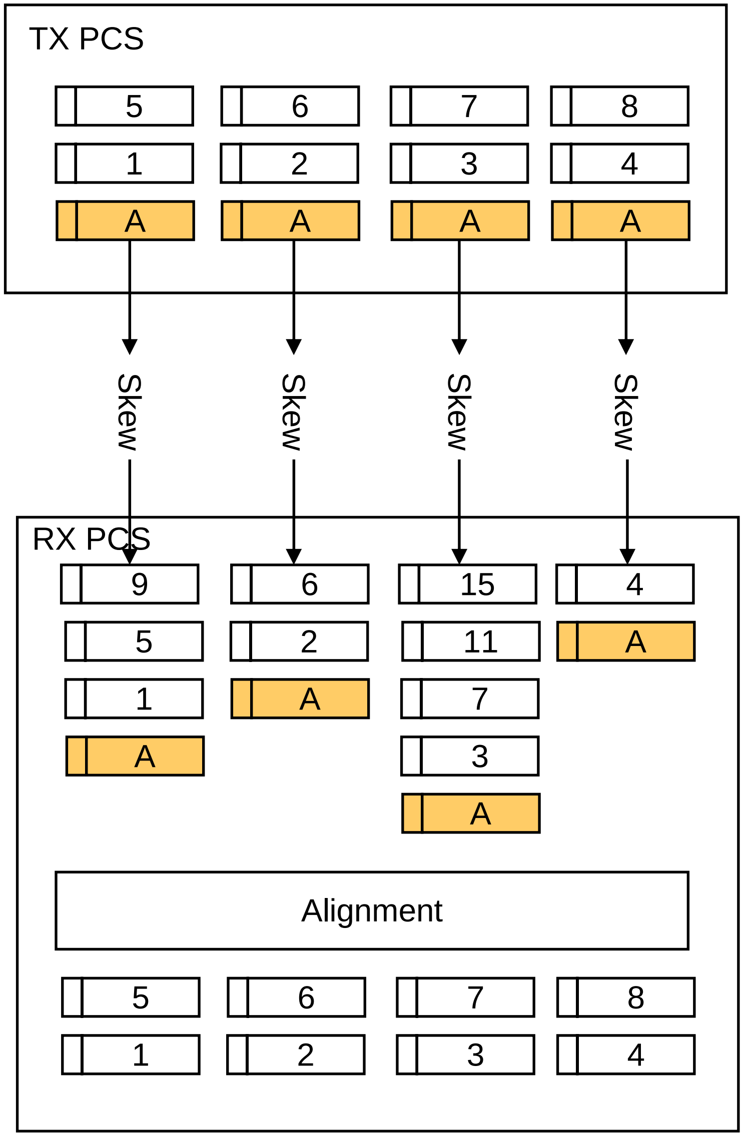
This example is 40GE with 4 electrical and 4 optical lanes



PCS Functions:

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

Alignment Mechanism – 40GE Example



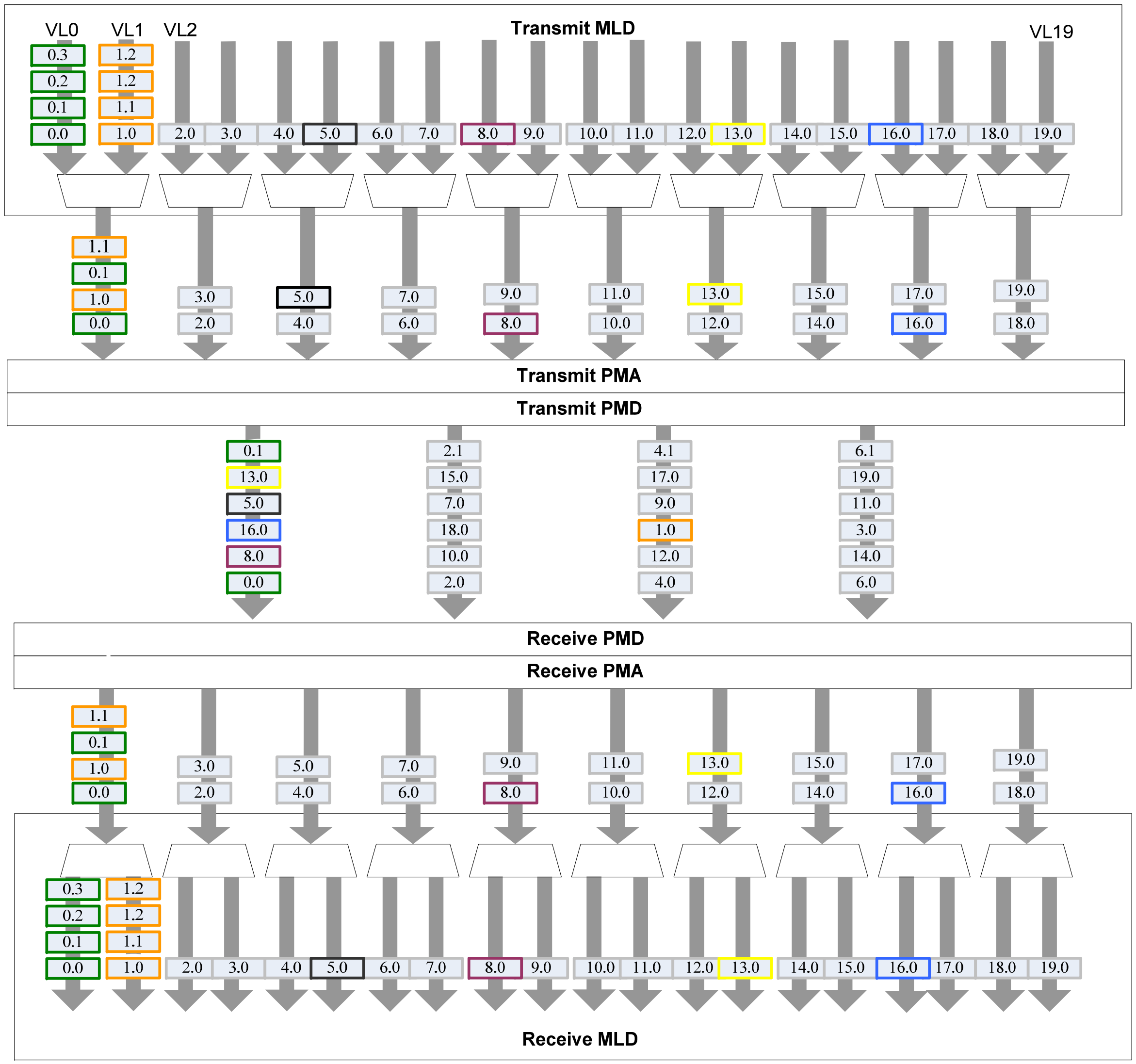
RX PCS Functions:

- 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:

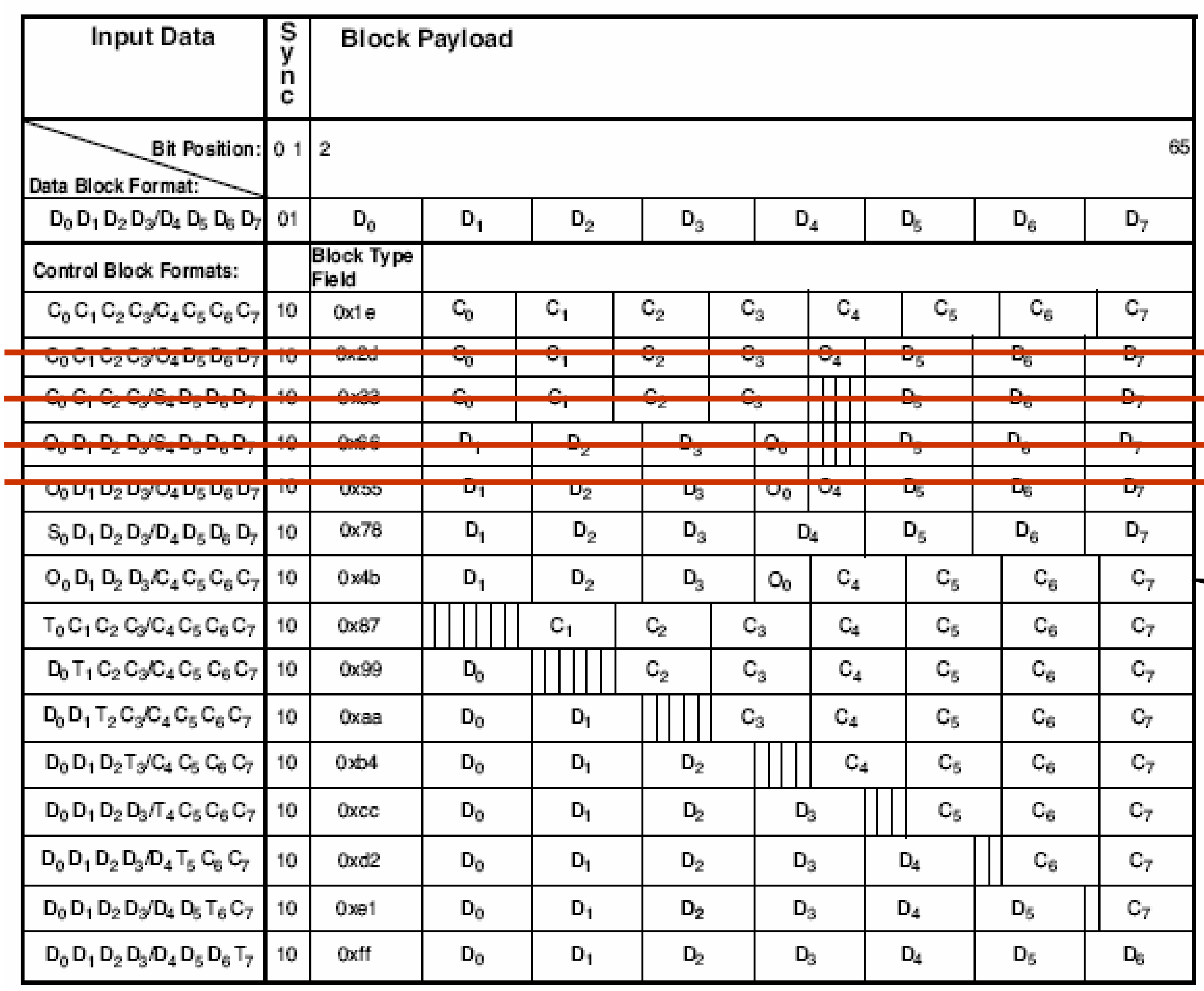
Electrical Lane Widths	PMD Lane Widths	Virtual Lanes Needed
4, 2, 1	4, 2, 1	4

- 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
~~$C_0 C_1 C_2 C_3/O_4 D_5 D_6 D_7$~~	~~10~~	~~0x2d~~	~~C_0~~	~~C_1~~	~~C_2~~	~~C_3~~	~~O_4~~	~~D_5~~	~~D_6~~	~~D_7~~
~~$C_0 C_1 C_2 C_3/S_4 D_5 D_6 D_7$~~	~~10~~	~~0x33~~	~~C_0~~	~~C_1~~	~~C_2~~	~~C_3~~		~~D_5~~	~~D_6~~	~~D_7~~
~~$O_0 D_1 D_2 D_3/S_4 D_5 D_6 D_7$~~	~~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	

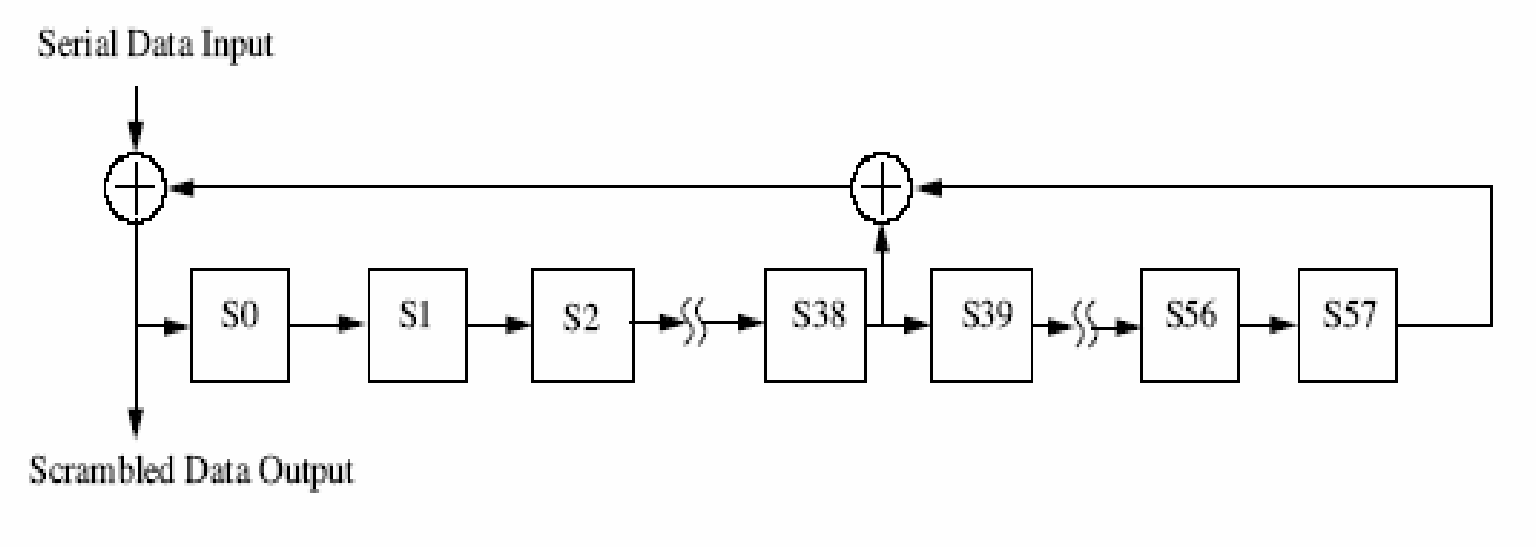
Not used since we have 8B alignment

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

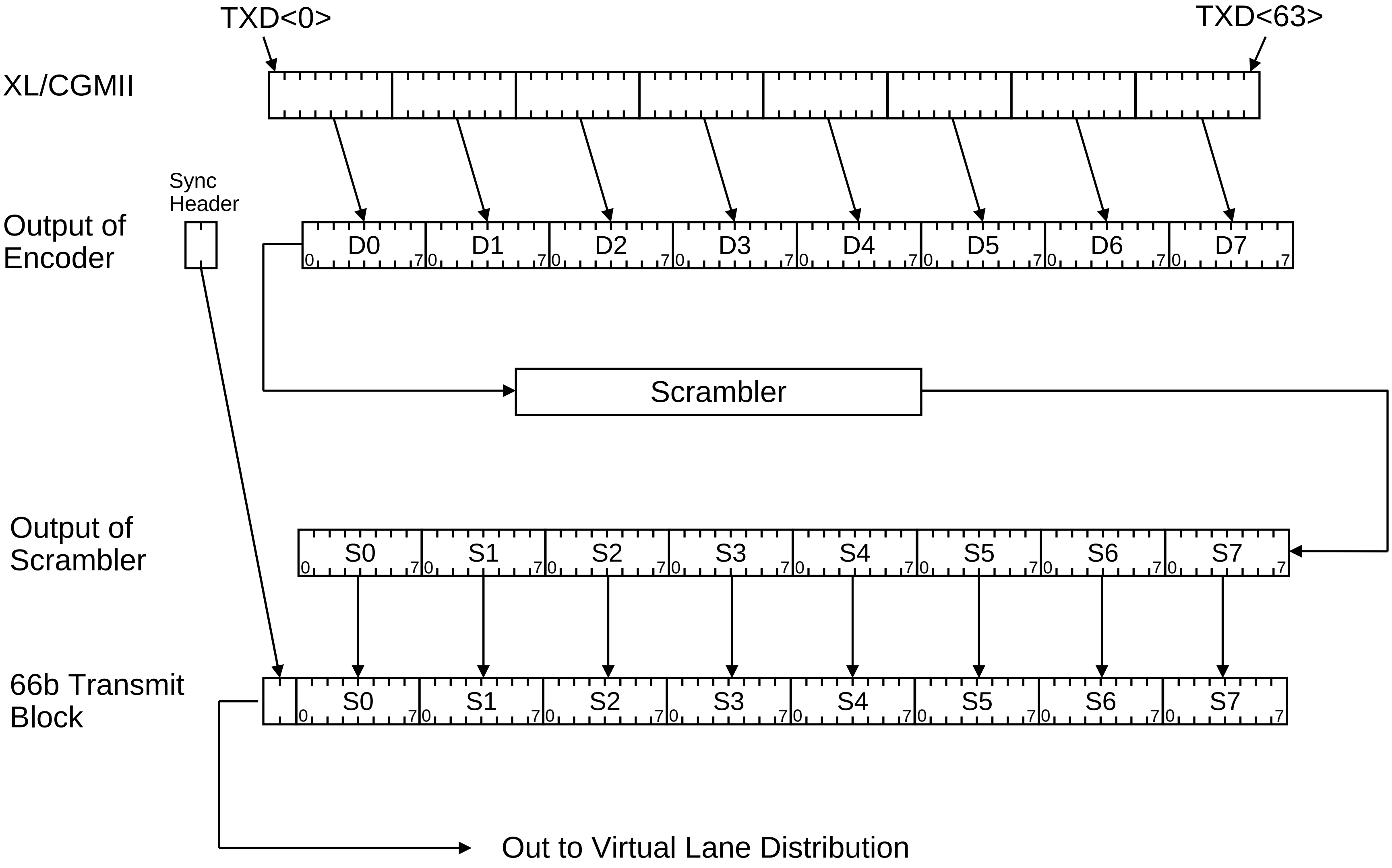
  When deleting idles, the minimum IPG of one character is maintained

  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

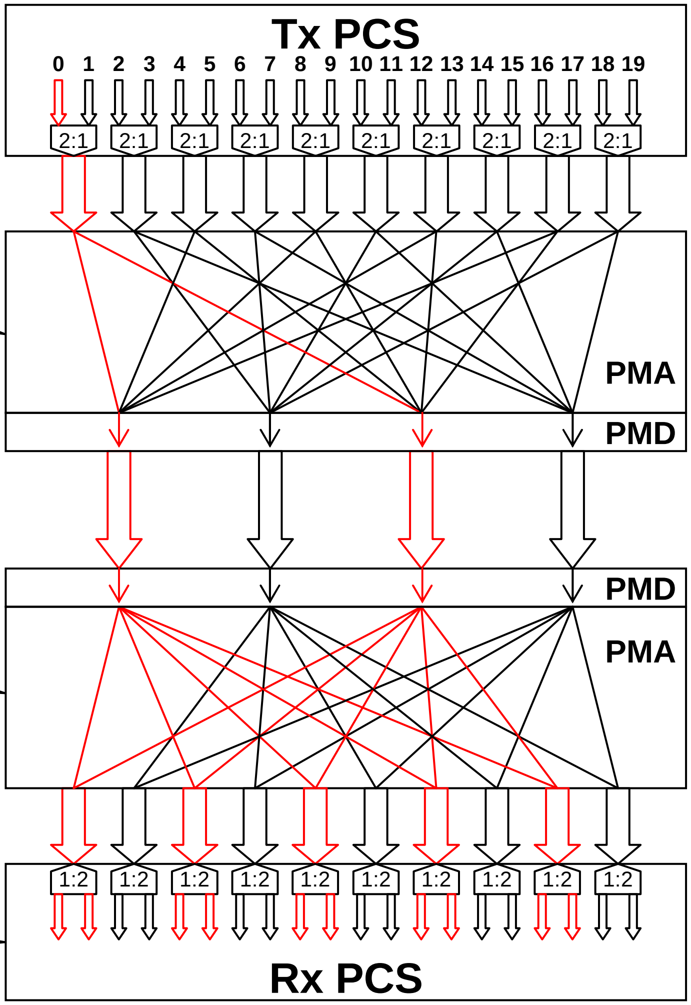
Note that data is played out in VL order, 0, 1, 2, 3, 0…

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

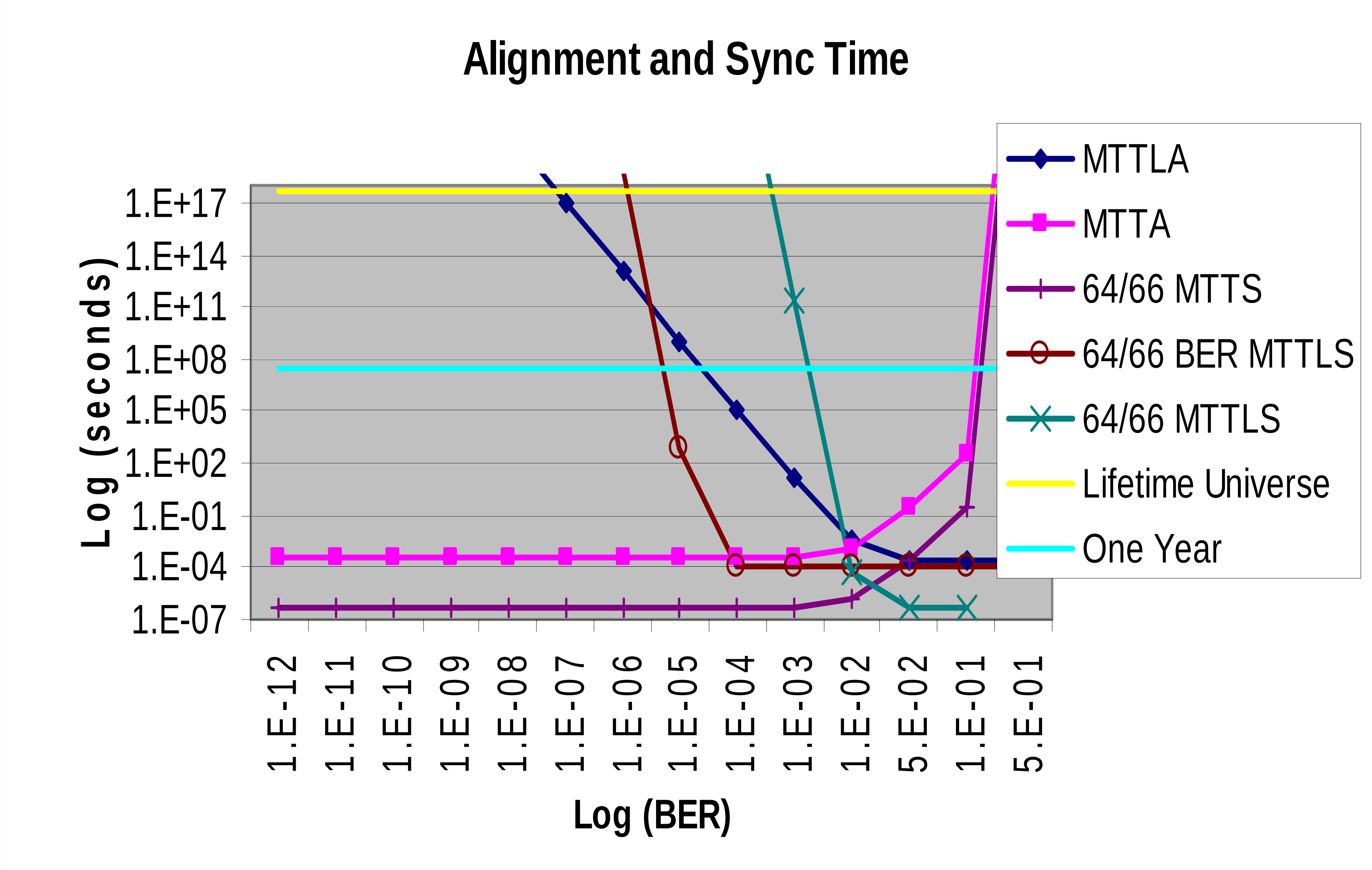
- Also have 64/66 sync stats on the graph for comparison

  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

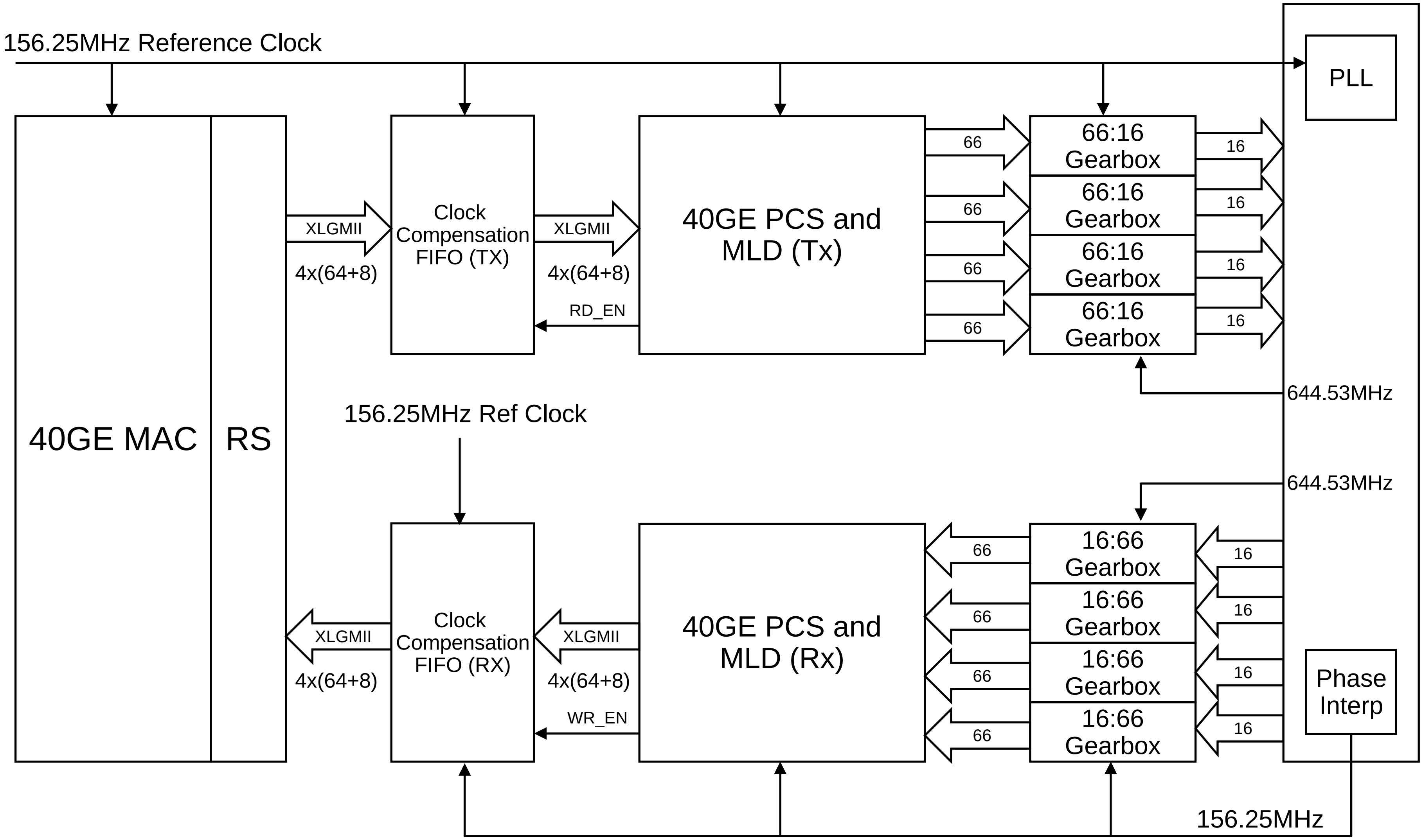
  MTTLS - Mean Time To Lose Sync

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

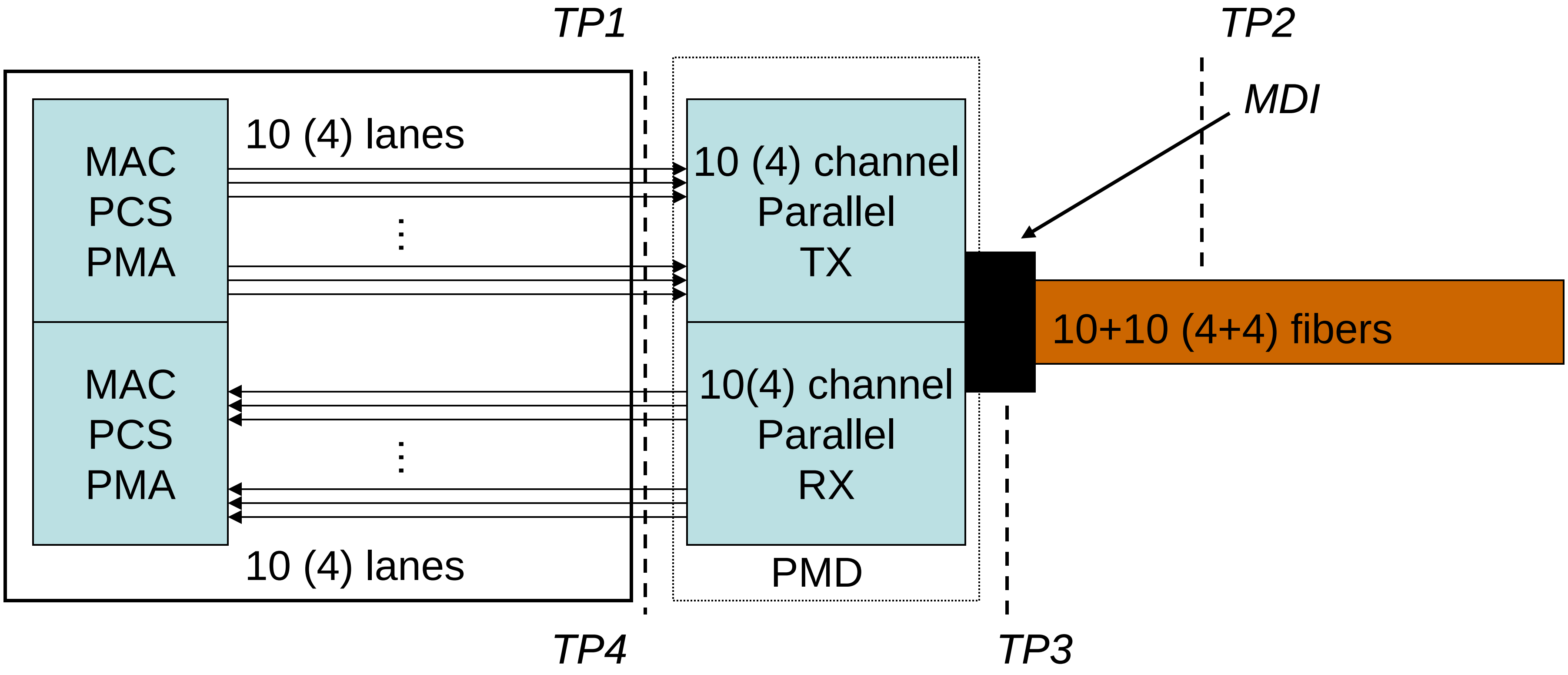
- Simple 10GBASE-R based PCS
- MLD layer to support multiple physical lanes/lambdas
- Complexity is low within the MLD layer
  - Simple block data striping
- Complexity in the optical module is low
  - Simple bit muxing even when m != n
- Based on proven 64B/66B framing and scrambling
- Electrical interface 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

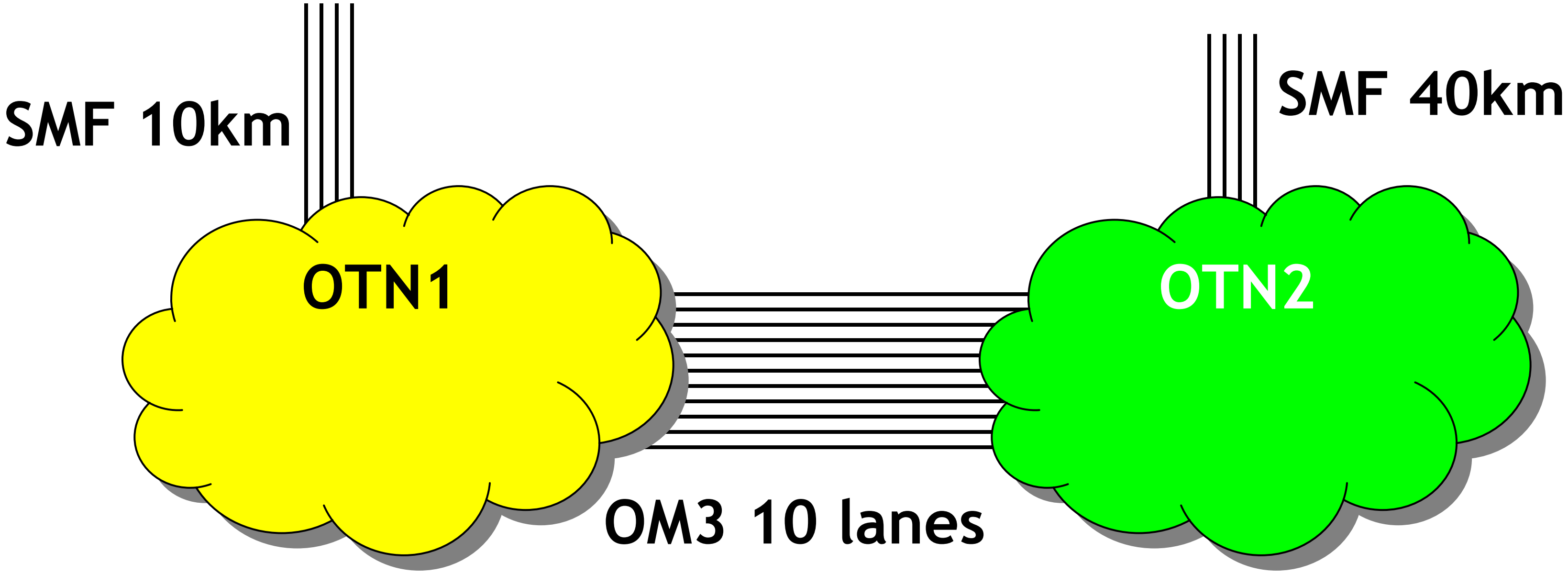
Supporters

- 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
  - o Key feature of MLD
- 40 GbE must fit into the OPU3 payload with a minimum of PCS codeword and timing transparency
  - o Limitation on control block types to permit transcoding
- Lane Marker transparency for 40 GbE
  - o 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

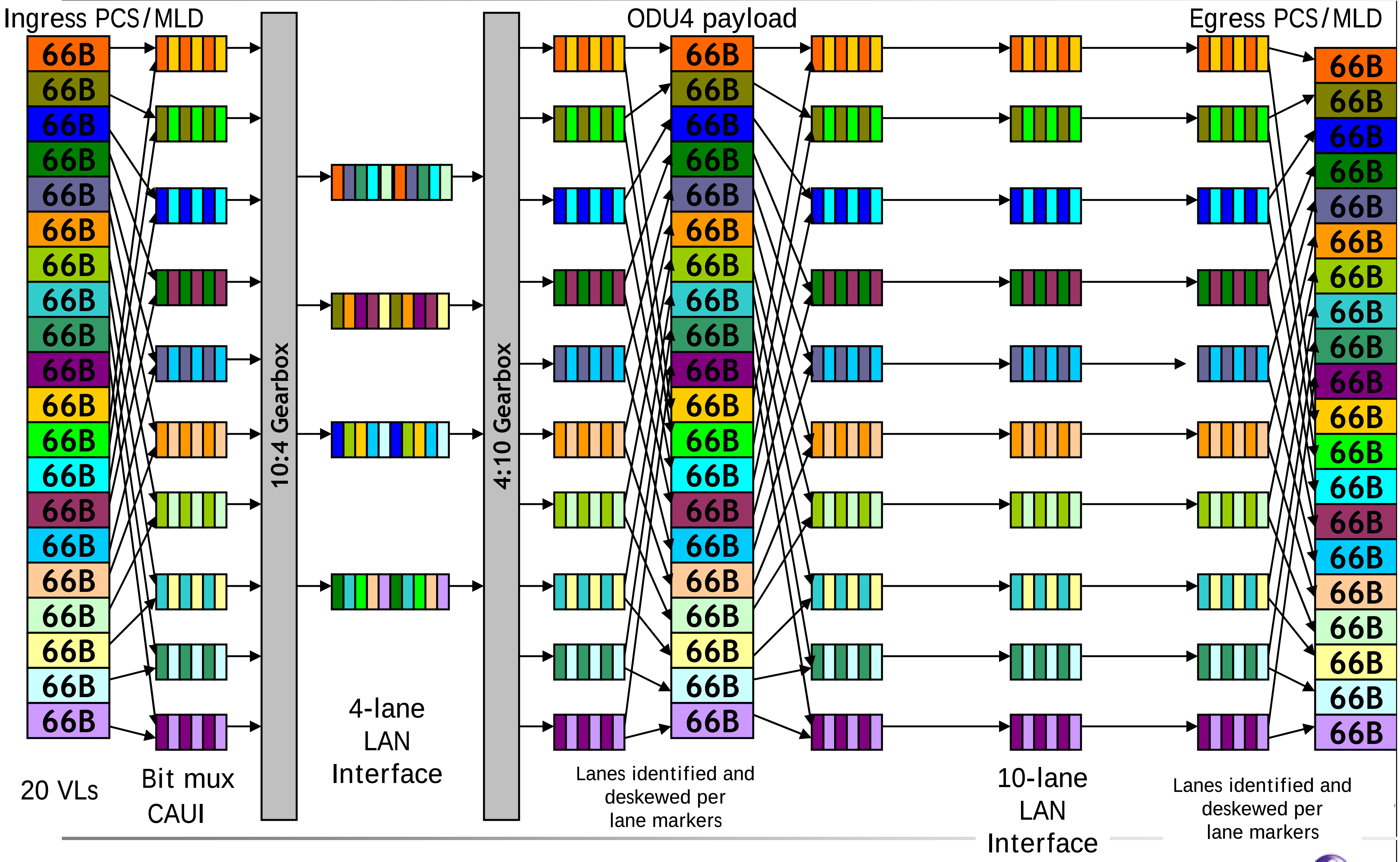


The sequence of bits transported across OTN should not depend on which physical interface is chosen for Ethernet at the ingress or egress

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.



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:
  - o A single canonical form to be used for mapping of any 802.3ba interface with at least codeword transparency over OTN
  - o Selection of different Ethernet PMDs at the OTN ingress and egress

OTN support for 40 GbE

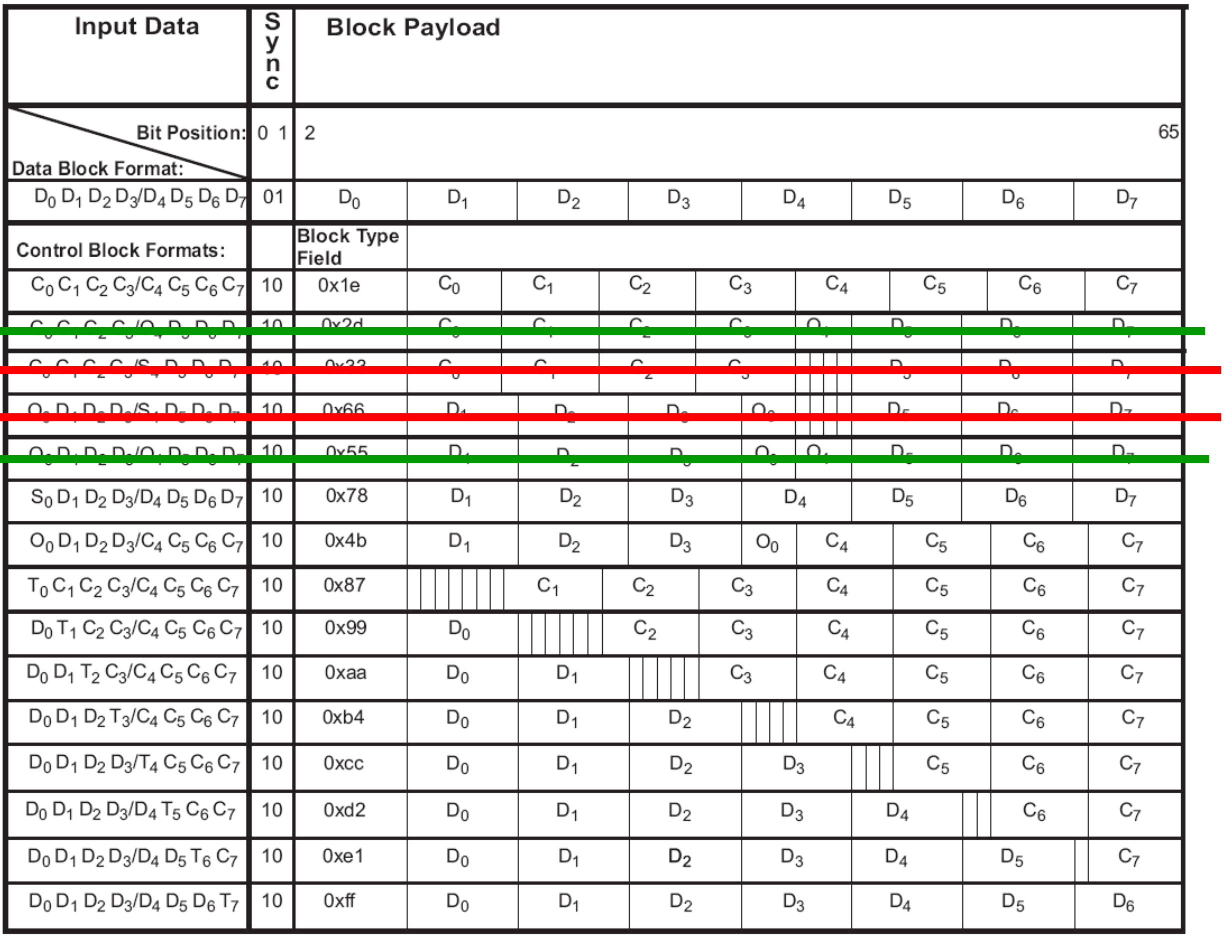
- **Two ways to provide 40 GbE transparent transport over OTN:**
  - Choose a MAC bit-rate (e.g., 38.9 Gbit/s) such that 64B/66B coding and lane marker insertion results in a bit-stream that fits the payload area of an OPU3 (not preferred)
  - Impose strict requirements on PCS codeword set that permits 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

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



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
$C_0 C_1 C_2 C_3/O_4 D_5 D_6 D_7$	10	0x2d	C_0	C_1	C_2	C_3	O_4	D_5	D_6	D_7
$C_0 C_1 C_2 C_3/S_4 D_5 D_6 D_7$	10	0x33	C_0	C_1	C_2	C_3		D_5	D_6	D_7
$O_0 D_1 D_2 D_3/S_4 D_5 D_6 D_7$	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	

Figure 49–7—64B/66B block formats

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

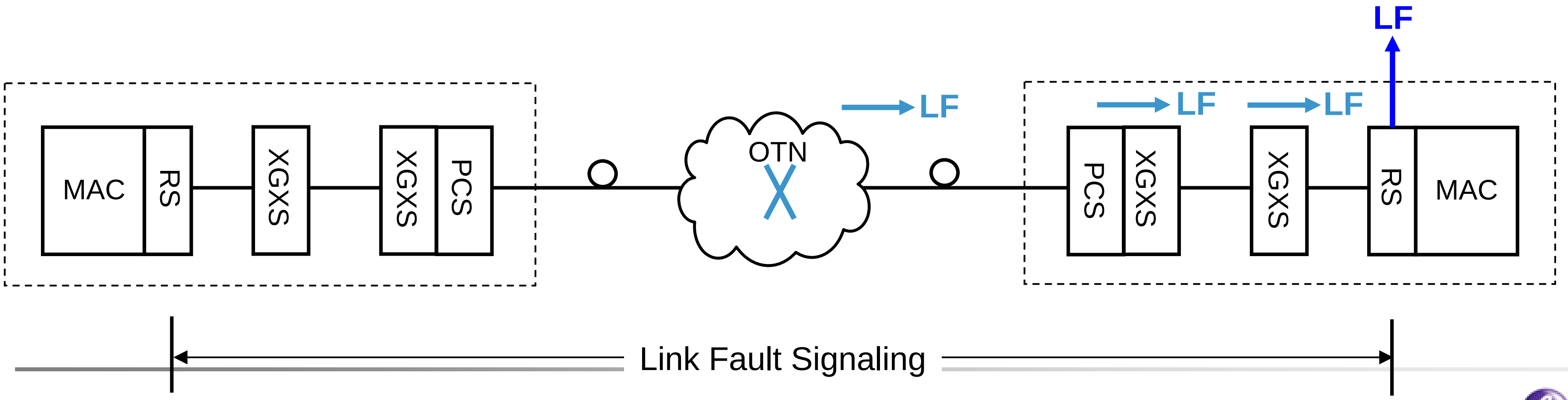
OTN support for 40 GbE proposal

- 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



100GE 40km SMF PMD

IEEE 802.3ba Task Force

13-15 May 2008

Chris Cole - Finisar

Pete Anslow - Nortel

Ramon Gutierrez - UNAM

Wenbin Jiang - Huawei

John Johnson - CyOptics

Radha Nagarajan - Infinera

Hirotaka Oomori - Sumitomo

Matt Traverso - Opnext

Supporters

- Ghani Abbas - Ericsson
- Arne Alping – Ericsson
- Ralf-Peter Braun – Deutsche Telekom
- Martin Carroll – Verizon
- Mike Dudek – JDSU
- Jörg-Peter Elbers – Adva
- Joel Goergen – Force10
- John Jaeger – Infinera
- Jack Jewell - JDSU
- Jeff Maki – Juniper Networks
- George Young - AT&T
- Mark Nowell – CISCO
- Gary Nicholl – CISCO
- Shoichi Ogita – Eudyna
- Thomas Paatzsch - Cube Optics
- Shashi Patel – Foundry Networks
- Bill Ryan – Foundry Networks
- Sam Sambasivan – AT&T
- Henk Steenman – AMS-IX
- Eddie Tsumura – ExceLight
- George Young - AT&T
- Ted Woodward – Telcordia

40km SMF Outline

- Status
- Architecture
- LAN WDM Baseline (-10nm) Grid
- 40km Baseline Grid Link and Power Budget

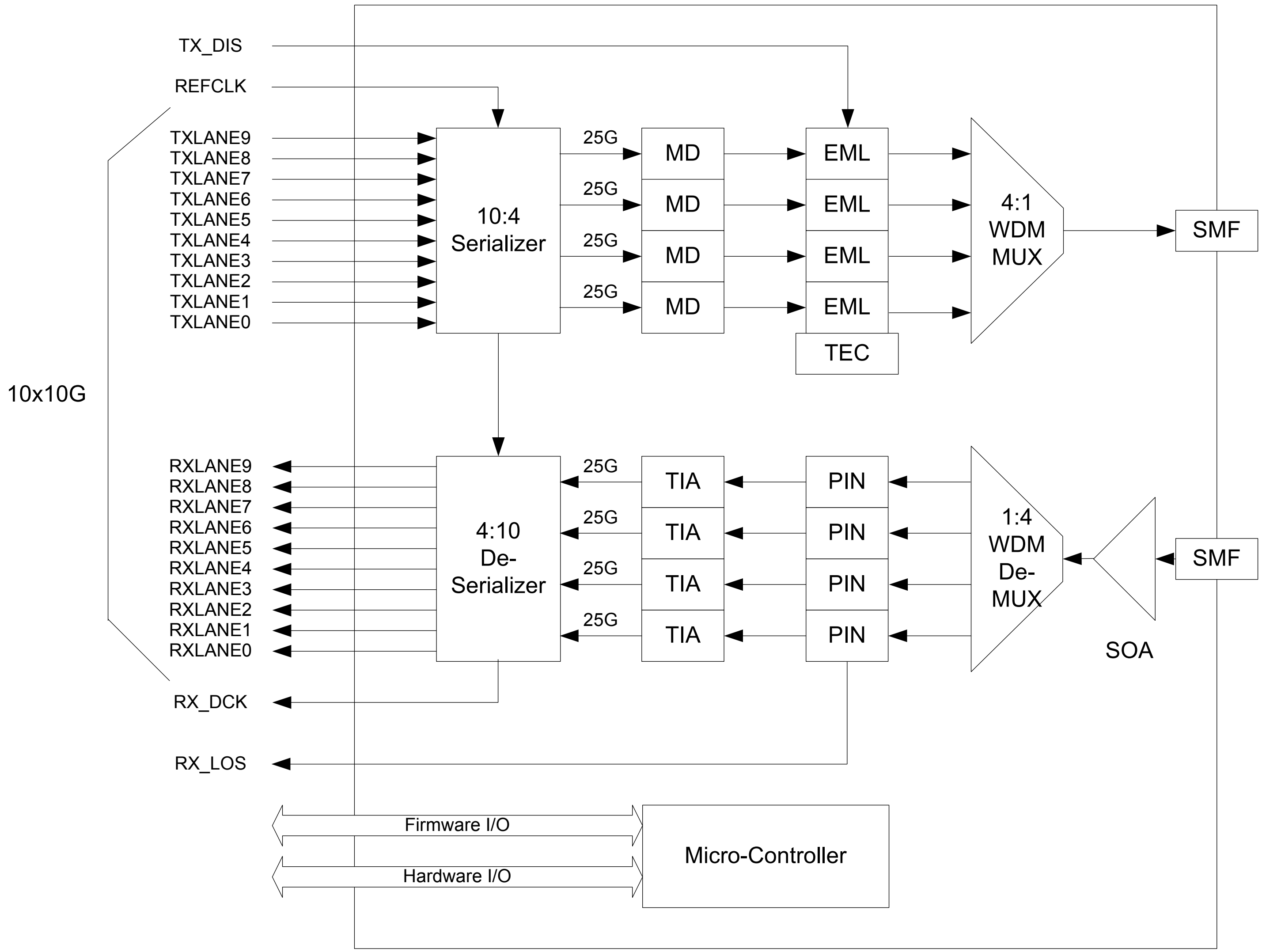
The following appendices have NOT been reviewed by the presentation co-authors (other then the lead author) and supporters, so their co-authorship and support does not necessarily apply to any of the appendices

- Appendix 1: LAN WDM Reference (0nm) Grid and Power Budget
- Appendix 2: LAN WDM -5nm Grid and Power Budget
- Appendix 3: LAN WDM -15nm Grid and Power Budget
- Appendix 4: 1310nm EML 40km SMF Dispersion Tolerance Measurements
- Appendix 5: SOA Overload Performance Simulation

40km SMF Status

- Baseline Approach to 40km SMF reach
  - TX: 4x25G MD-EML → LAN WDM Mux
  - RX: 4x25G PIN-TIA ← LAN WDM DeMux ← SOA
- Technical presentations discussing baseline approach
  - cole_01_1106, cole_02_0107, cole_01_0407, cole_01_0507, cole_01_0907, cole_02_0108
  - traverso_01_0407
  - jiang_01_0507, jiang_01_0907
  - gutierrez_01_0507, gutierrez_01/02/03/04_1107
  - matsumoto_01_1107
  - nagarajan_01_1107
  - johnson_01_0108
  - anslow_01_0308
- Key Issues analyzed
  - Min receiver sensitivity
  - Non-linear effects
  - Overload
  - PMD Penalty

Gen1 40km 4x25G 1310nm Transceiver Architecture



LAN WDM Baseline (-10nm) Grid

- ITU G.694.1 specification
- 800GHz spacing (193.1THz base)
- 4 wavelengths shifted by -10nm from Reference Grid
- Exact wavelength values: 1295.56 1300.05 1304.58 1309.14 nm
- Shorthand wavelength values: 1295, 1300, 1305, 1310 nm
- 2nm window (precise pass-band TBD)
- G.652 A&B 40km SMF worst dispersion and fiber loss
  - Max positive dispersion (1310nm) = 36ps/nm
  - Max negative dispersion (1295nm) = -114ps/nm
  - Max Loss (1310nm) = 16.8dB
  - Max Loss (1295nm) = 17.3dB

40km Baseline Grid Power Budget

25G Link Budget
40km SMF
TP2 → TP3
dB
	LAN WDM
	EML chirp α = -0.5
	λ = 1295nm
	ER = 8dB

Fiber Loss (G.652 A&B)	17.3
Connector loss	2.0
Dispersion Penalty	1.5
Other Penalties (TX, PMD)	1.7
Total budget	22.5 dB

25G Pwr. Budget
40km SMF
TP2 → TP3
OMA (Average) dBm
	LAN WDM
	EML chirp α = -0.5
	λ = 1295nm
	ER = 8dB

TX Min [Max]	2.6 (1.0) [5.6 (4.0)]
TP2 TX Min [Max] 2.5dB Mux loss	0.1 (-1.5) [4.1 (2.5)]
Link Budget (dB)	22.5 dB
TP3 RX Min	-22.4 (-24.0)
RX Min (with 1dB crosstalk penalty)	-10.2 (-11.8) dBm

- EML chirp range assumption: $-0.5 \leq \alpha \leq 1.0$
- 1.5 dB Dispersion Penalty and 1dB PMD in Other Penalties needs further quantification
- Min attenuation = 0dB assumption subject to verification of SOA WDM overload at low bias
- RX overload, max difference in power between wavelengths, other specs TBD

Appendix 1: LAN WDM Reference (0nm) Grid

- ITU G.694.1 specification
- 800GHz spacing (193.1THz base)
- 4 wavelengths selected for minimum dispersion in 1310nm window
- Exact wavelength values: 1305.72, 1310.28, 1314.88, 1319.51 nm
- Shorthand wavelength values: 1305, 1310, 1315, 1320 nm
- 2nm window
- G.652 A&B 40km SMF worst dispersion and fiber loss
  - Max positive dispersion (1320nm) = 75ps/nm
  - Max negative dispersion (1305nm) = -74ps/nm
  - Max Loss (1320nm) = 17dB
- Reference Grid is used as basis for comparison of alternate grid proposals

40km Reference Grid Power Budget

25G Link Budget
40km SMF
TP2 → TP3
	LAN WDM
	EML chirp α = 1.0
	λ = 1320nm
	ER = 8dB

Fiber Loss (G.652 A&B)	17 dB
Connector loss	2.0
Dispersion Penalty	2.0
Other Penalties (TX, PMD)	1.7
Total budget	22.7 dB

25G Pwr. Budget
40km SMF
TP2 → TP3
OMA (Average)
	LAN WDM
	EML chirp α = 1.0
	λ = 1320nm
	ER = 8dB

TX Min	2.6 (1.0) dBm
TP2 TX Min 2.5dB Mux loss	0.1 (-1.5)
Link Budget (dB)	22.7 dB
TP3 RX Min	-22.6 (-24.2)
RX Min (with 1dB crosstalk penalty)	-10.2 (-11.8) dBm

Appendix 2: LAN WDM -5nm Grid

- ITU G.694.1 specification
- 800GHz spacing (193.1THz base)
- 4 wavelengths shifted by -5nm from Reference Grid
- Exact wavelength values: 1300.62, 1305.15, 1309.71, 1314.3 nm
- Shorthand wavelength values: 1300, 1305, 1310, 1315 nm
- 2nm window
- G.652 A&B 40km SMF worst dispersion and fiber loss
  - Max positive dispersion (1315nm) = 56ps/nm
  - Max negative dispersion (1300nm) = -92ps/nm
  - Max Loss (1315nm) = 16.6dB
  - Max Loss (1300nm) = 17.1dB

40km -5nm Grid Power Budget

25G Link Budget
40km SMF
TP2 → TP3
dB
	LAN WDM
	EML chirp α = -0.5
	λ = 1300nm
	ER = 8dB

Fiber Loss (G.652 A&B)	17.1
Connector loss	2.0
Dispersion Penalty	1.2
Other Penalties (TX, PMD)	1.7
Total budget	22 dB

25G Pwr. Budget
40km SMF
TP2 → TP3
OMA (Average) dBm
	LAN WDM
	EML chirp α = -0.5
	λ = 1300nm
	ER = 8dB

TX Min [Max]	2.6 (1.0) [5.6 (4.0)]
TP2 TX Min [Max] 2.5dB Mux loss	0.1 (-1.5) [4.1 (2.5)]
Link Budget (dB)	22 dB
TP3 RX Min	-21.9 (-23.5)
RX Min (with 1dB crosstalk penalty)	-10.2 (-11.8) dBm

- EML chirp assumption: -0.5 ≤ α ≤ 1.0
- 1.2 dB Dispersion Penalty and 1dB PMD in Other Penalties needs further quantification
- EML λ = 1315nm, chirp = 1.0: Dispersion Penalty = 1.5dB, Fiber Loss = 16.6dB

Appendix 3: LAN WDM -15nm Grid

- ITU G.694.1 specification
- 800GHz spacing (193.1THz base)
- 4 wavelengths shifted by -15nm from Reference Grid
- Exact wavelength values: 1290.54, 1295.00, 1299.49, 1304.01 nm
- Shorthand wavelength values: 1290, 1295, 1300, 1305 nm
- 2nm window
- G.652 A&B 40km SMF worst dispersion and fiber loss
  - Max positive dispersion (1305nm) = 19.2ps/nm
  - Max negative dispersion (1290nm) = -134ps/nm
  - Max Loss (1305nm) = 16.9dB
  - Max Loss (1290nm) = 17.6dB

40km -15nm Grid Power Budget

25G Link Budget 40km SMF
TP2 → TP3
dB
	LAN WDM
	EML chirp α = -0.5
	λ = 1290nm
	ER = 8dB

Fiber Loss (G.652 A&B)	17.6
Connector loss	2.0
Dispersion Penalty	1.7
Other Penalties (TX, PMD)	1.7
Total budget	23.0 dB

25G Pwr. Budget
40km SMF
TP2 → TP3
OMA (Average) dBm
	LAN WDM
	EML chirp α = -0.5
	λ = 1290nm
	ER = 8dB

TX Min [Max]	2.6 (1.0) [5.6 (4.0)]
TP2 TX Min [Max] 2.5dB Mux loss	0.1 (-1.5) [4.1 (2.5)]
Link Budget (dB)	23.0 dB
TP3 RX Min	-22.9 (-24.5)
RX Min (with 1dB crosstalk penalty)	-10.2 (-11.8) dBm

- EML chirp assumption: -0.5 ≤ α ≤ 1.0
- 1.7dB Dispersion Penalty and 1dB PMD in Other Penalties needs further quantification
- EML λ = 1305nm Dispersion Penalty = 0.6dB, Fiber Loss = 16.9dB

Appendix 4: Dispersion Penalty Measurements

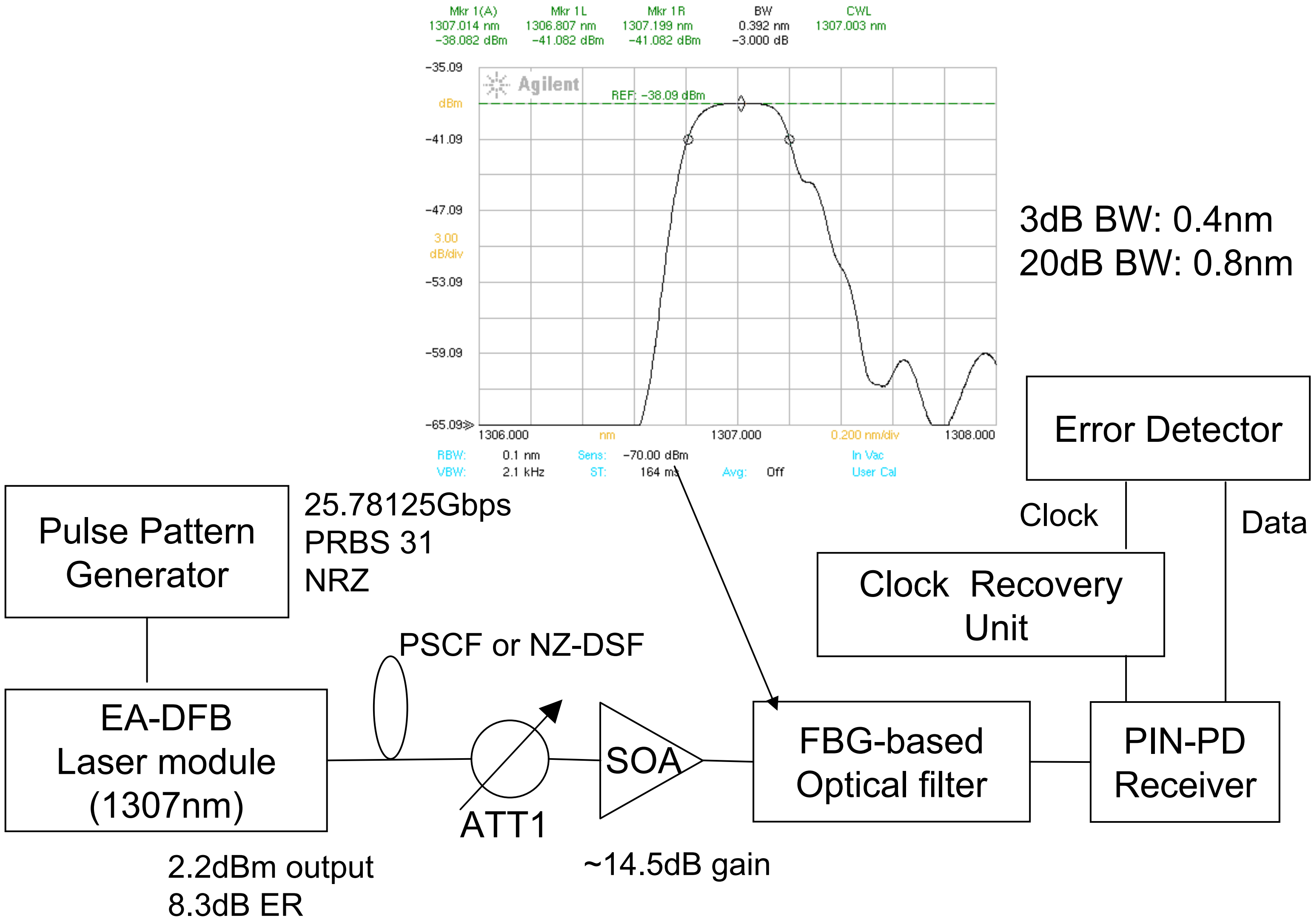
1310nm band EML Dispersion Tolerance Measurement Result over 40km SMF

100GE 40km PMD

Hirotaka Oomori (Sumitomo Electric Industries, Ltd.)

Eddie Tsumura (ExceLight Communications, Inc.)

Test Setup



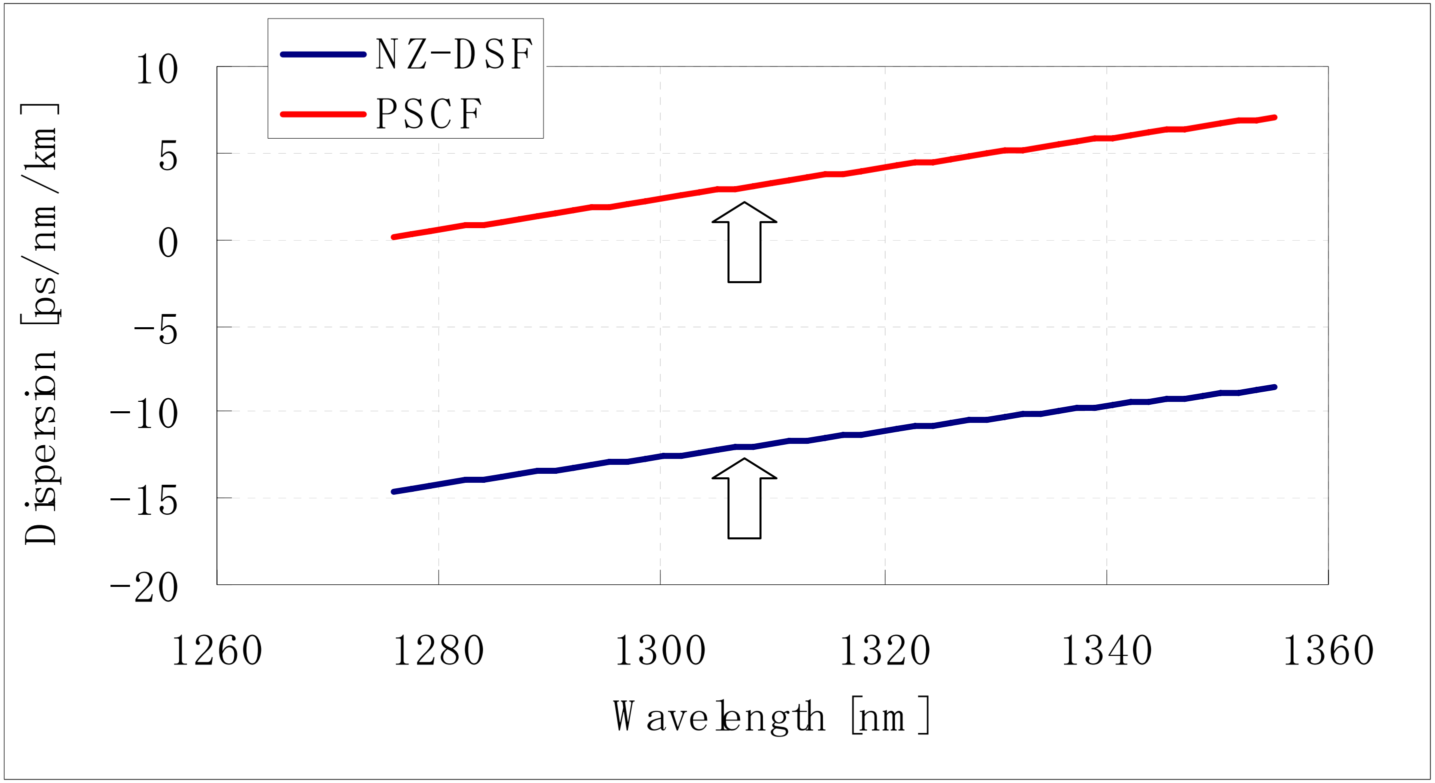
3dB BW: 0.4nm
20dB BW: 0.8nm

25.78125Gbps
PRBS 31
NRZ

2.2dBm output
8.3dB ER

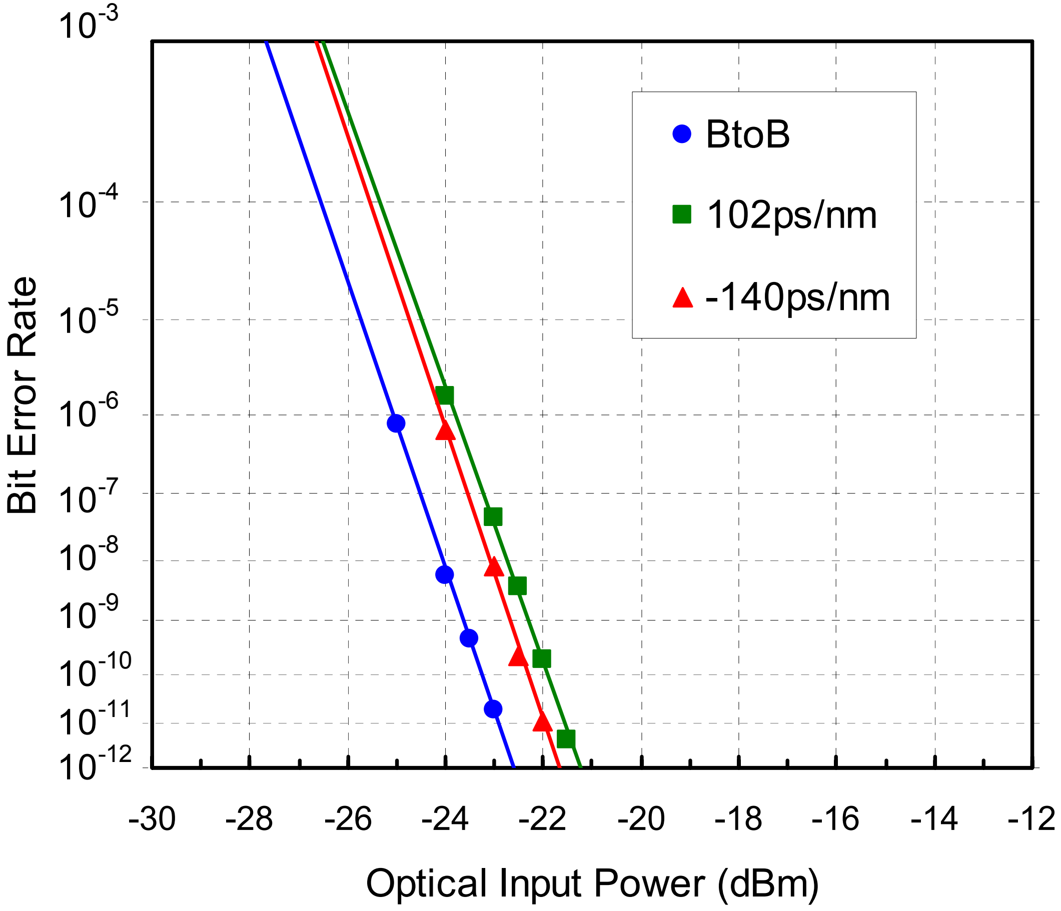
~14.5dB gain

Dispersion Value of NZ-DSF and PSCF



- NZ-DSF: Non Zero - Dispersion Shifted Fiber
- PSCF: Pure Silica Core Fiber

BER Measurement Result



- 1.4dB penalty @102ps/nm
- 0.9dB penalty @-140ps/nm

Appendix 5: Overload Performance Simulation

Gain-controlled SOA performance

High input power conditions

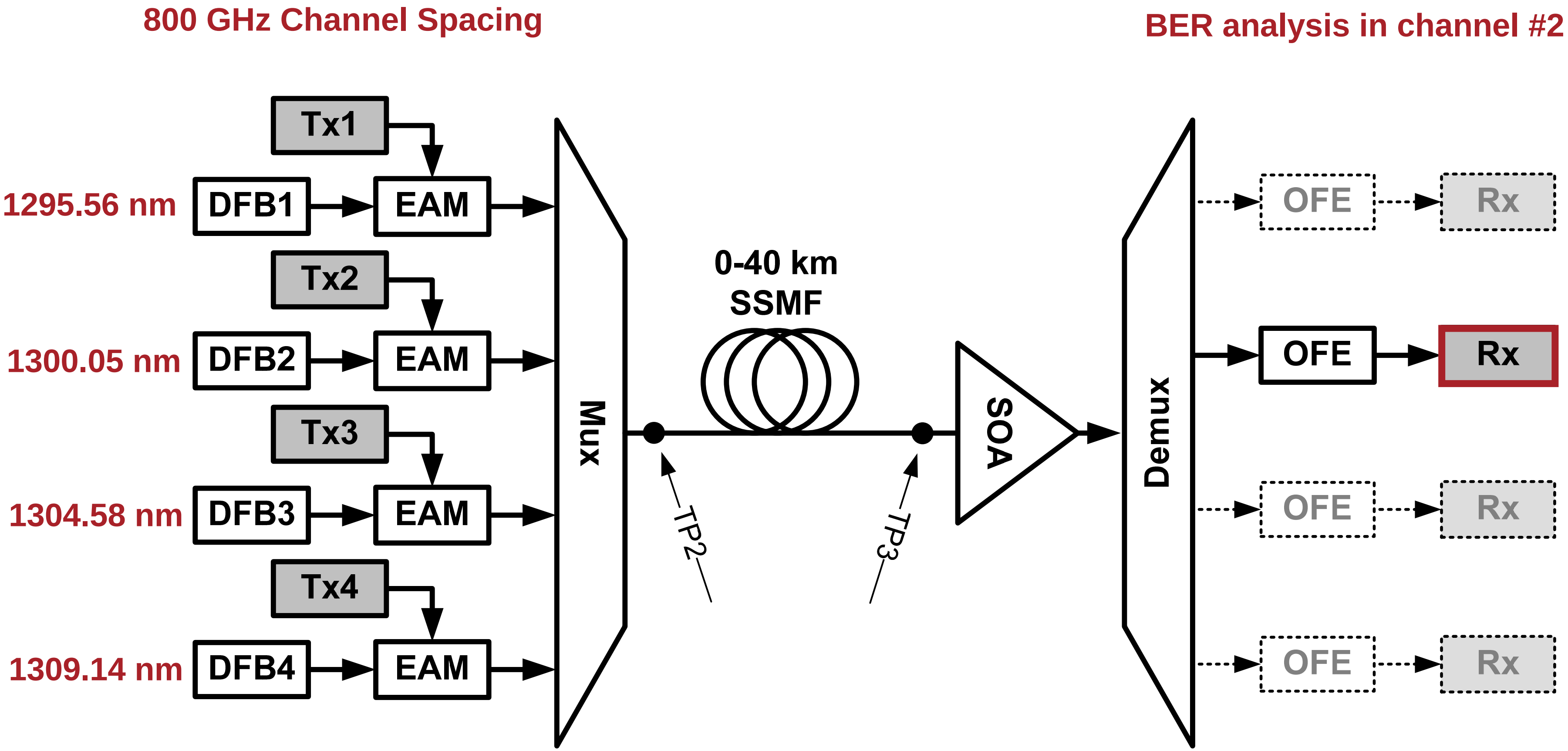
100 GbE 40-km PMD

Ramón Gutiérrez-Castrejón

Universidad Nacional Autonoma de Mexico-UNAM

email: RGutierrezC@ii.unam.mx

Optical Link Setup: 4x25-Gb/s EMLs & SOA Pre-Amp



EML Transmitters Characteristics

For the analysis we have considered:

- Extinction ratio = 8 dB
- Optical signal-to-noise ratio = 40 dB
- High and low EML output powers = +5.6 dBm, +2.6 dBm
- Insertion loss MUX = 2.5 dB
- Insertion loss DEMUX = 5.2 dB

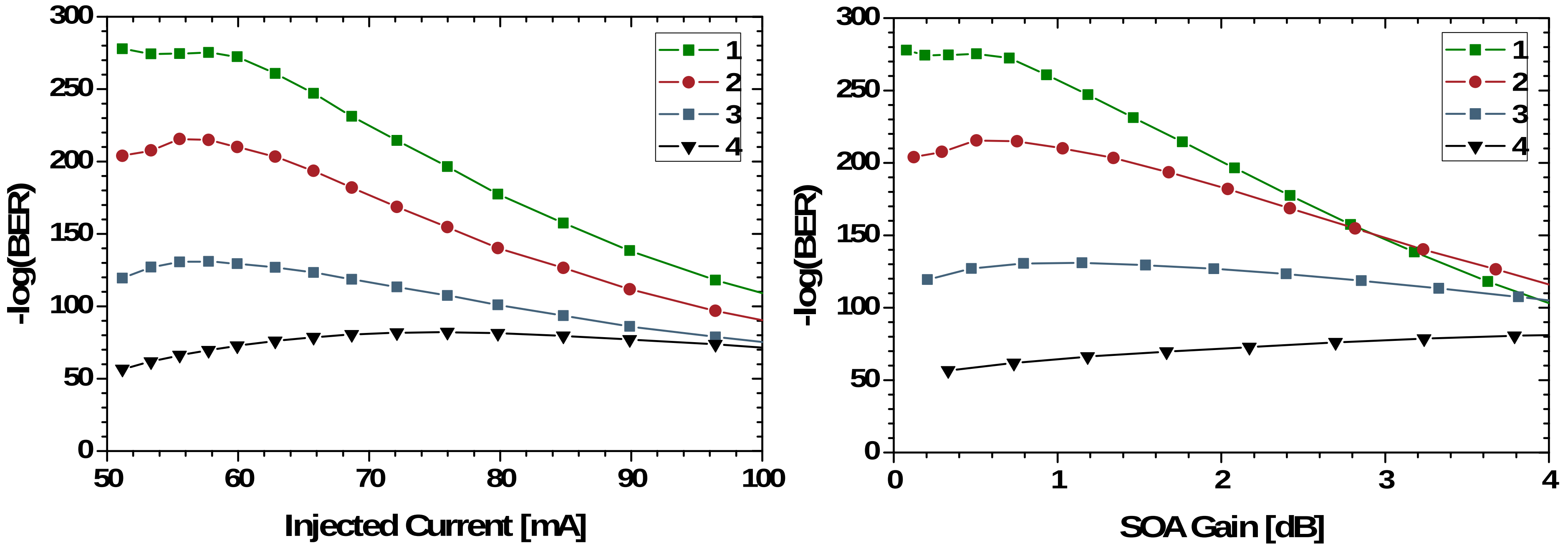
EML Output Power	+5.6 dBm	+2.6 dBm
Per Channel Power at TP2	**+3.1 dBm**	**+0.1 dBm**
Total Power at TP2	**9.1 dBm**	**6.1 dBm**

Simulation Characteristics

- BER vs. SOA injection current analysis
- Current varied in (50 mA ,…,100 mA), corresponding to small-signal gain in (4 dB,…,18 dB). Lower bound determined by SOA model.
- Four fiber lengths analyzed: 0, 0.001, 5 and 10 km
- Fiber Characteristics: losses: 0.45 dB/km (+ 2 dB connector), dispersion coefficient @ 1310 nm: D = -0.20 ps/nm/km, dispersion slope @ 1310 nm: S = 0.090 ps/nm^2/km
- Analysis for
  - High power transmitters: All channels at 5.6 dBm
  - Low power Transmitters: All channels at 2.6 dBm
  - Combined power: All channels at 5.6 dBm, but Tx2** at 2.6 dBm
- Special test bit pattern of 2^10-1 bits. See gutierrez_01_1107.

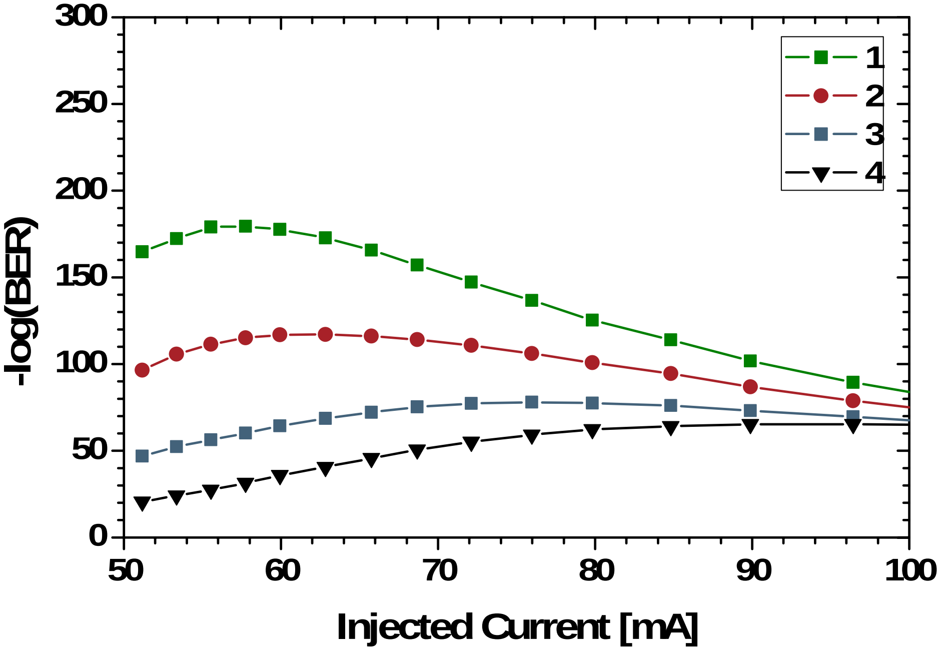
**** Note: BER Performance carried out in Channel 2 (Tx2)**

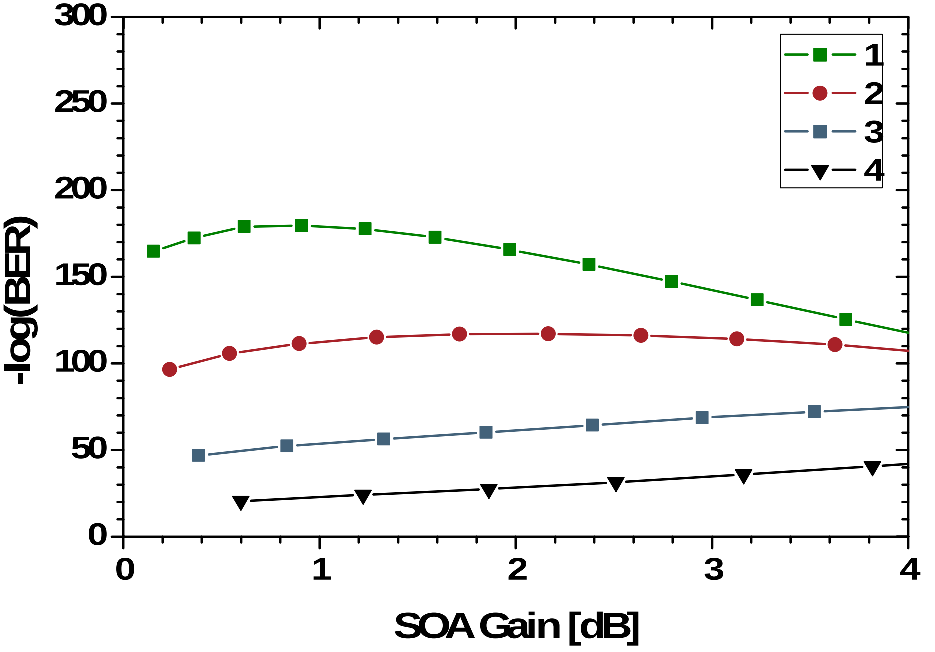
BER Performance: High Power Transmitter (5.6 dBm)



Curve	Fiber Length	Fiber Losses	SOA Input Power (Tot)
1	**0 km**	**0 dB**	**+9.10 dBm**
2	**0.001 km**	**2 dB**	**+7.10 dBm**
3	**5 km**	**4.25 dB**	**+4.85 dBm**
4	**10 km**	**6.50 dB**	**+2.60 dBm**

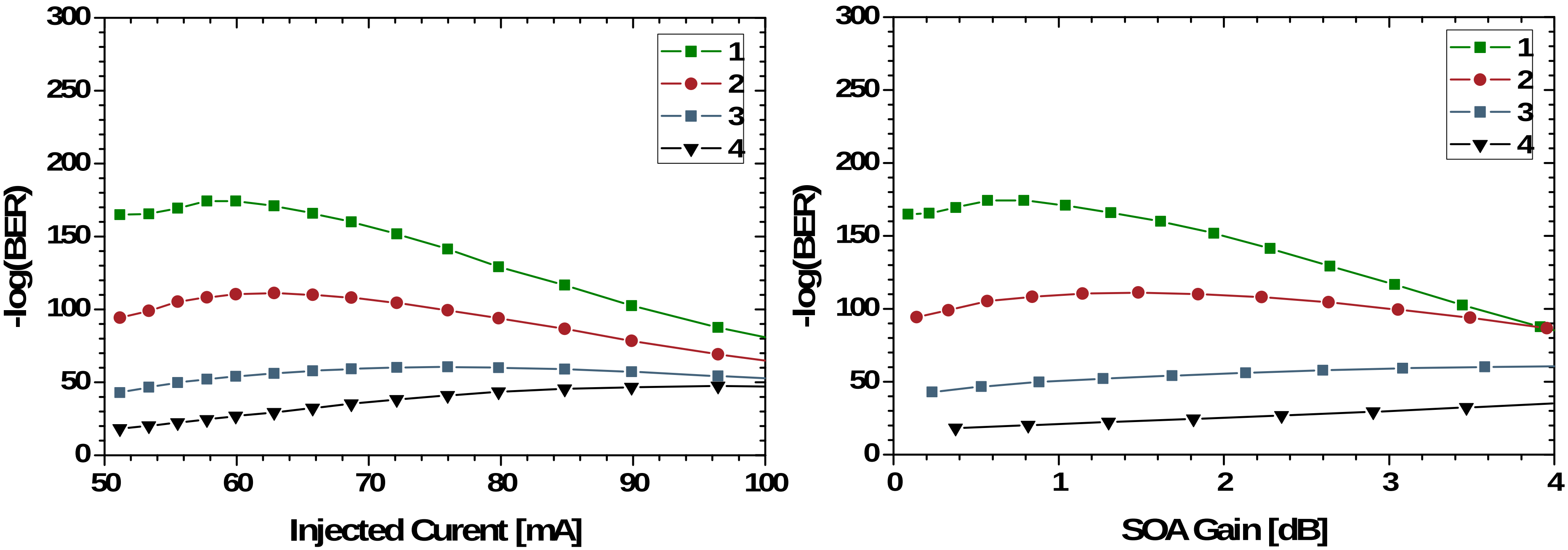
BER Performance: Low Power Transmitter (2.6 dBm)





Curve	Fiber Length	Fiber Losses	SOA Input Power (Tot)
1	**0 km**	**0 dB**	**+6.10 dBm**
2	**0.001 km**	**2 dB**	**+4.10 dBm**
3	**5 km**	**4.25 dB**	**+1.85 dBm**
4	**10 km**	**6.50 dB**	**-0.40 dBm**

BER Performance: High Power w/Low Power @ Tx8



Curve	Fiber Length	Fiber Losses	SOA Input Power (Tot)
1	**0 km**	**0 dB**	**+8.50 dBm**
2	**0.001 km**	**2 dB**	**+6.50 dBm**
3	**5 km**	**4.25 dB**	**+4.25 dBm**
4	**10 km**	**6.50 dB**	**+2.00 dBm**

Appendix 5 Conclusion

- The SOA gain-control scheme exhibits excellent performance for high optical powers
- Good system BER performance for a wide range of current values - no need for highly accurate control
- The SOA gain-control scheme operates correctly even above the transparency point (Gain > 0 dB)
- Single intermediate current value (e.g. 100 mA correspond to 18 dB of small-signal gain) is good enough for fiber lengths ranging from 0 to 10 km and even longer
- Results depends on SOA characteristics
- Measurements required to confirm findings

100GE 10km SMF PMD

IEEE 802.3ba Task Force

13-15 May 2008

Chris Cole - Finisar
Bernd Huebner - Finisar
Pete Anslow - Nortel
John Johnson - CyOptics
Radha Nagarajan - Infinera
Hirotaka Oomori - Sumitomo

Supporters

- Ghani Abbas - Ericsson
- Arne Alping – Ericsson
- Ralf-Peter Braun – Deutsche Telekom
- Mike Dudek – JDSU
- Jörg-Peter Elbers – Adva
- Joel Goergen – Force10
- John Jaeger – Infinera
- Jack Jewell - JDSU
- Jeff Maki – Juniper Networks
- Arlon Martin – Kotura
- Mark Nowell – CISCO
- Gary Nicholl – CISCO
- Thomas Paatzsch - Cube Optics
- Shashi Patel – Foundry Networks
- Bill Ryan – Foundry Networks
- Sam Sambasivan – AT&T
- Henk Steenman – AMS-IX
- Eddie Tsumura – ExceLight
- George Young - AT&T
- Ted Woodward – Telcordia

Outline

- Status
- Architecture
- LAN WDM Baseline (-10nm) Grid
- 10km Baseline Grid Link and Power Budget

The following appendices have NOT been reviewed by the presentation co-authors (other then the lead author) and supporters, so their co-authorship and support does not necessarily apply to any of the appendices

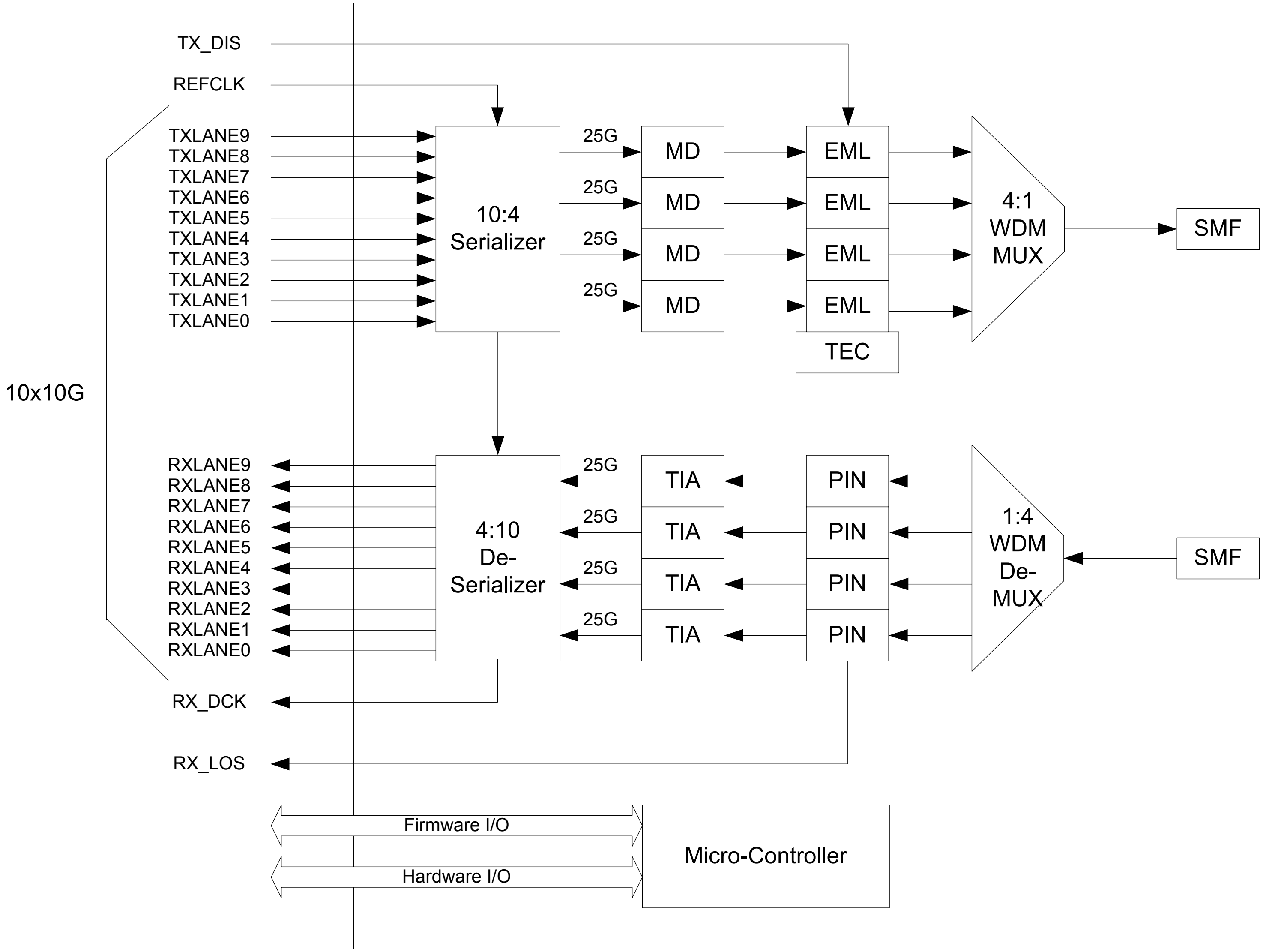
- Appendix 1: LAN WDM Reference (0nm) Grid
- 10km Reference Grid Link and Power Budget
- Appendix 2: LAN WDM -5nm Grid
- 10km -5nm Grid Link and Power Budget
- Appendix 3: LAN WDM -15nm Grid
- 10km -15nm Grid Link and Power Budget

10km SMF Status

- Baseline Approach to 10km SMF reach
  - TX: 4x25G MD-EML → LAN WDM Mux
  - RX: 4x25G PIN-TIA ← LAN WDM DeMux
- Technical presentations discussing baseline approach
  - cole_01_1106, cole_01_0307, cole_01_0907, cole_01/02_0108, cole_01/02_0308
  - jiang_01_0507, jiang_01_0907
  - nagarajan_01_1107
  - johnson_01_0108

  Many other presentations discussed 4x25G MD-EML approach, but on a CWDM Grid
- Key Issues analyzed
  - Min receiver sensitivity
  - Transmitter output power
  - DML feasibility
  - Dispersion Penalty
  - Integration approaches

Gen1 10km 1310nm EML PMD



LAN WDM Baseline (-10nm) Grid

- ITU G.694.1 specification
- 800GHz spacing (193.1THz base)
- 4 wavelengths shifted by -10nm from Reference Grid
- Exact wavelength values: 1295.56 1300.05 1304.58 1309.14 nm
- Shorthand wavelength values: 1295, 1300, 1305, 1310 nm
- 2nm window (precise pass-band TBD)
- G.652 A&B 10km SMF worst corner dispersion and fiber loss
  - Max positive dispersion (1310nm) = 9ps/nm
  - Max negative dispersion (1295nm) = -28ps/nm
  - Max Loss (1310nm) = 4.2dB
  - Max Loss (1295nm) = 4.3dB

10km Baseline Grid Power Budget

25G Link Budget
10km SMF
TP2 → TP3
	LAN WDM
	EML α = -1.0
	λ = 1295nm
	ER = 7dB
		LAN WDM
		DML α = 3.5
		λ = 1310nm
		ER = 4.5dB

Fiber Loss (G.652 A&B)	4.3 dB	4.2 dB
Connector loss	2.0	2.0
Dispersion Penalty	0.5	0.6
Other Penalties	0.7	0.7
Total budget	7.5 dB	7.5 dB

25G Pwr Budget
10km SMF
TP2 → TP3
OMA (Average)
	LAN WDM
	EML α = -1.0
	λ = 1295nm
	ER = 7dB
		LAN WDM
		DML α = 3.5
		λ = 1310nm
		ER = 4.5dB

TX Min	2.3 dBm (1.1)	2.3 dBm (2.5)
TP2 TX Min 2.5dB Mux loss	-0.2	-0.2
Link Budget (dB)	7.5	7.5
TP3 RX Min 2.5dB Demx loss	-7.7	-7.7
RX Min (w/ 1dB xtalk penalty)	-10.2 dBm (-11.4)	-10.2 dBm (-10.0)

- EML chirp range assumption: $-1.0 \leq \alpha \leq 1.0$
- EML λ = 1310nm, chirp α = 1.0: Dispersion Penalty = 0.2, Loss = 4.2dB (not limiting)
- DML λ = 1295nm, chirp α = 3.5: Dispersion Penalty = 0.5, Loss = 4.3dB (equivalent)
- RX overload, max difference in power between wavelengths, other specs TBD

Appendix 1: LAN WDM Reference (0nm) Grid

- ITU G.694.1 specification
- 800GHz spacing (193.1THz base)
- 4 wavelengths selected for minimum dispersion in 1310nm window
- Exact wavelength values: 1305.72, 1310.28, 1314.88, 1319.51 nm
- Shorthand wavelength values: 1305, 1310, 1315, 1320 nm
- 2nm window
- G.652 A&B 10km SMF worst corner dispersion and fiber loss
  - Max Positive Dispersion (1320nm) = 19ps/nm
  - Max Negative Dispersion (1305nm) = -18ps/nm
  - Max Loss (1320nm) = 4.2dB
  - Max Loss (1305nm) = 4.2dB
- Reference Grid is used as basis for comparison of alternate grid proposals

10km Reference Grid Power Budget

25G Link Budget
10km SMF
TP2 → TP3
	LAN WDM
	EML α = 1.0
	λ = 1320nm
	ER = 7dB
		LAN WDM
		DML α = 3.5
		λ = 1320nm
		ER = 4.5dB

Fiber Loss (G.652 A&B)	4.2 dB	4.2 dB
Connector loss	2.0	2.0
Dispersion Penalty	0.3	1.3
Other Penalties	0.7	0.7
Total budget	7.2 dB	8.2 dB

25G Pwr Budget
10km SMF
TP2 → TP3
OMA (Average)
	LAN WDM
	EML α = 1.0
	λ = 1320nm
	ER = 7dB
		LAN WDM
		DML α = 3.5
		λ = 1320nm
		ER = 4.5dB

TX Min	2.0 dBm (0.8)	3.0 dBm (3.2)
TP2 TX Min 2.5dB Mux loss	-0.5	0.5
Link Budget (dB)	7.2	8.2
TP3 RX Min 2.5dB Demx loss	-7.7	-7.7
RX Min (w/ 1dB xtalk penalty)	-10.2 dBm (-11.4)	-10.2 dBm (-10.0)

Appendix 2: LAN WDM -5nm Grid

- ITU G.694.1 specification
- 800GHz spacing (193.1THz base)
- 4 wavelengths shifted by -5nm from Reference Grid
- Exact wavelength values: 1300.62, 1305.15, 1309.71, 1314.3 nm
- Shorthand wavelength values: 1300, 1305, 1310, 1315 nm
- 2nm window
- G.652 A&B 10km SMF worst corner dispersion and fiber loss
  - Max positive dispersion (1315nm) = 14ps/nm
  - Max negative dispersion (1300nm) = -23ps/nm
  - Max Loss (1315nm) = 4.2dB
  - Max Loss (1300nm = 4.3dB

10km -5nm Grid Power Budget

25G Link Budget
10km SMF
TP2 → TP3
	LAN WDM
	EML α = -1.0
	λ = 1300nm
	ER = 7dB
		LAN WDM
		DML α = 3.5
		λ = 1315nm
		ER = 4.5dB

Fiber Loss (G.652 A&B)	4.3 dB	4.2 dB
Connector loss	2.0	2.0
Dispersion Penalty	0.4	0.8
Other Penalties	0.7	0.7
Total budget	7.4 dB	7.7 dB

25G Pwr Budget
10km SMF
TP2 → TP3
OMA (Average)
	LAN WDM
	EML α = -1.0
	λ = 1300nm
	ER = 7dB
		LAN WDM
		DML α = 3.5
		λ = 1315nm
		ER = 4.5dB

TX Min	2.2 dBm (1.0)	2.5 dBm (2.7)
TP2 TX Min 2.5dB Mux loss	-0.3	0.0
Link Budget (dB)	7.4	7.7
TP3 RX Min 2.5dB Demx loss	-7.7	-7.7
RX Min (w/ 1dB xtalk penalty)	-10.2 dBm (-11.4)	-10.2 dBm (-10.0)

- EML chirp range assumption: -1.0 ≤ α ≤ 1.0
- EML λ = 1315nm, chirp α = 1.0: Dispersion Penalty = 0.3, Loss = 4.2dB

Appendix 3: LAN WDM -15nm Grid

- ITU G.694.1 specification
- 800GHz spacing (193.1THz base)
- 4 wavelengths shifted by -15nm from Reference Grid
- Exact wavelength values: 1290.54, 1295.00, 1299.49, 1304.01 nm
- Shorthand wavelength values: 1290, 1295, 1300, 1305 nm
- 2nm window
- G.652 A&B 10km SMF worst corner dispersion and fiber loss
  - Max positive dispersion (1305nm) = 4.8ps/nm
  - Max negative dispersion (1290nm) = -33.5ps/nm
  - Max Loss (1305nm) = 4.2dB
  - Max Loss (1290nm) = 4.4dB

10km -15nm Grid Power Budget

25G Link Budget
10km SMF
TP2 → TP3
	LAN WDM
	EML α = -1.0
	λ = 1290nm
	ER = 7dB
		LAN WDM
		DML α = 3.5
		λ = 1290nm
		ER = 4.5dB

Fiber Loss (G.652 A&B)	4.4 dB	4.4 dB
Connector loss	2.0	2.0
Dispersion Penalty	0.6	0.5
Other Penalties	0.7	0.7
Total budget	7.7 dB	7.6 dB

25G Pwr Budget
10km SMF
TP2 → TP3
OMA (Average)
	LAN WDM
	EML α = -1.0
	λ = 1290nm
	ER = 7dB
		LAN WDM
		DML α = 3.5
		λ = 1290nm
		ER = 4.5dB

TX Min	2.5 dBm (1.3)	2.4 dBm (2.6)
TP2 TX Min 2.5dB Mux loss	-0.0	-0.1
Link Budget (dB)	7.7	7.6
TP3 RX Min 2.5dB Demx loss	-7.7	-7.7
RX Min (w/ 1dB xtalk penalty)	-10.2 dBm (-11.4)	-10.2 dBm (-10.0)

- EML chirp range: -1.0 ≤ α ≤ 1.0
- EML λ = 1305nm, α = 1.0: Dispersion Penalty = 0.3, Loss = 4.2dB
- DML λ = 1305nm, α = 3.5: Dispersion Penalty = 0.5, Loss = 4.2dB

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)

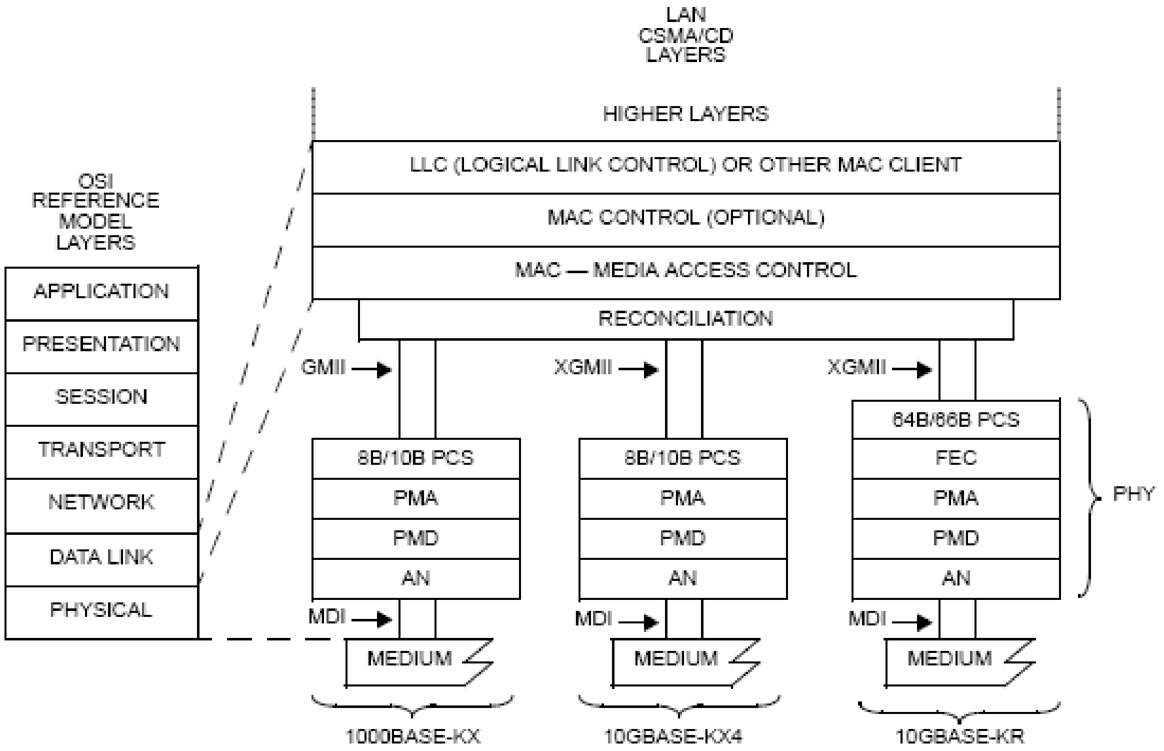
Considerations for 40G Backplane Ethernet PHY

- 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

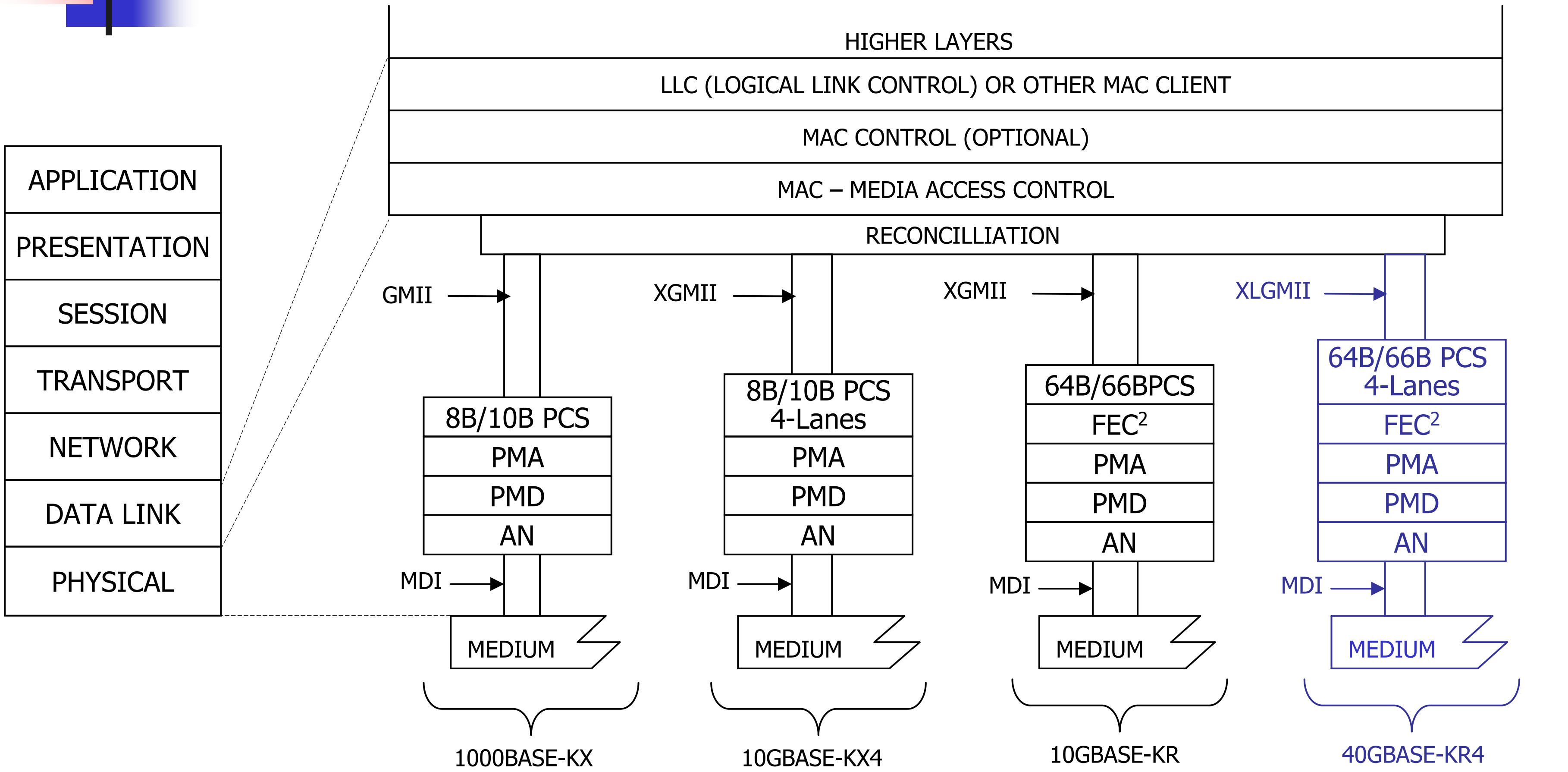


AN = AUTO-NEGOTIATION
GMII = GIGABIT MEDIA INDEPENDENT INTERFACE
MDI = MEDIUM DEPENDENT INTERFACE
PCS = PHYSICAL CODING SUBLAYER
FEC = FORWARD ERROR CORRECTION

PHY = PHYSICAL LAYER DEVICE
PMA = PHYSICAL MEDIUM ATTACHMENT
PMD = PHYSICAL MEDIUM DEPENDENT
XGMII = 10 GIGABIT MEDIA INDEPENDENT INTERFACE

Figure 69–1—Architectural positioning of Backplane Ethernet

Proposed backplane architecture with 40GbE



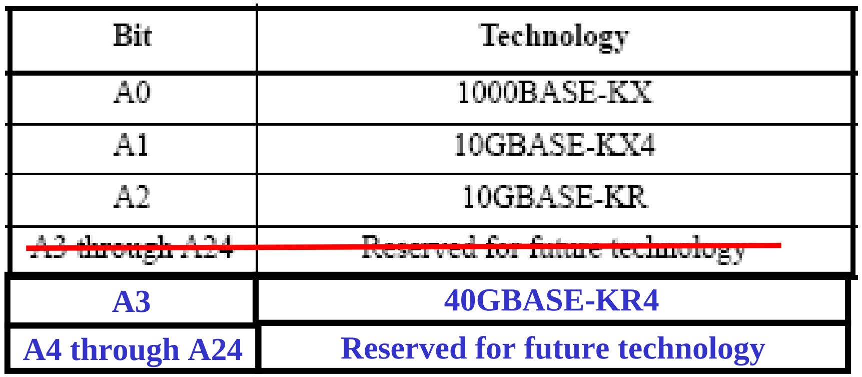
Note: 2. Optional

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

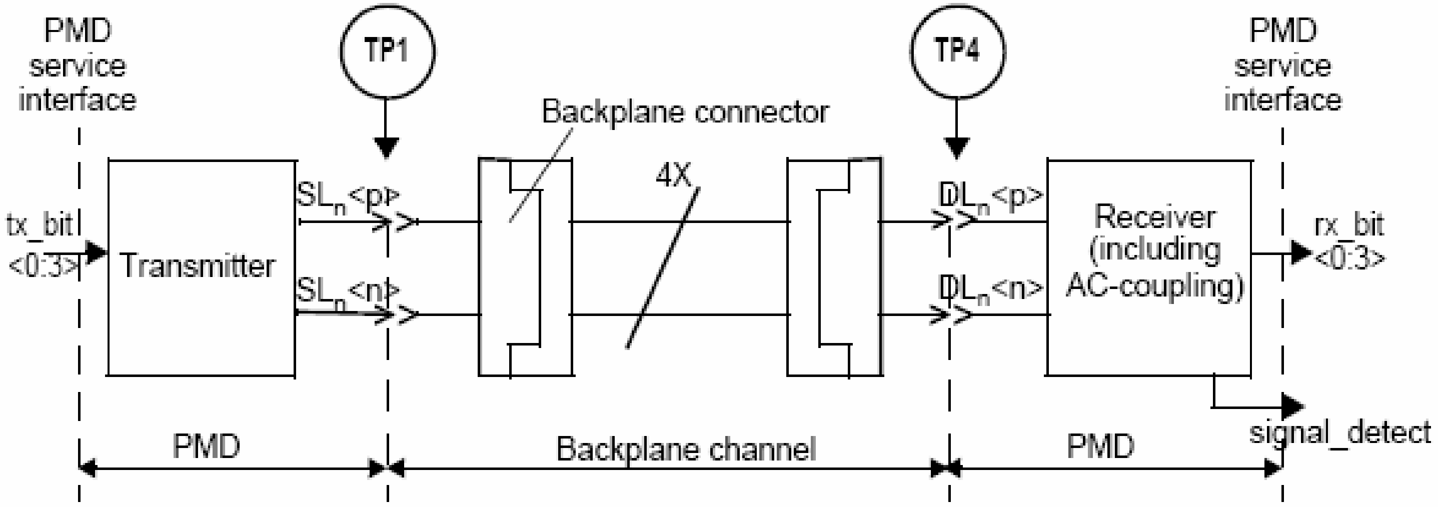


Figure ~~71–1~~—Link block diagram

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

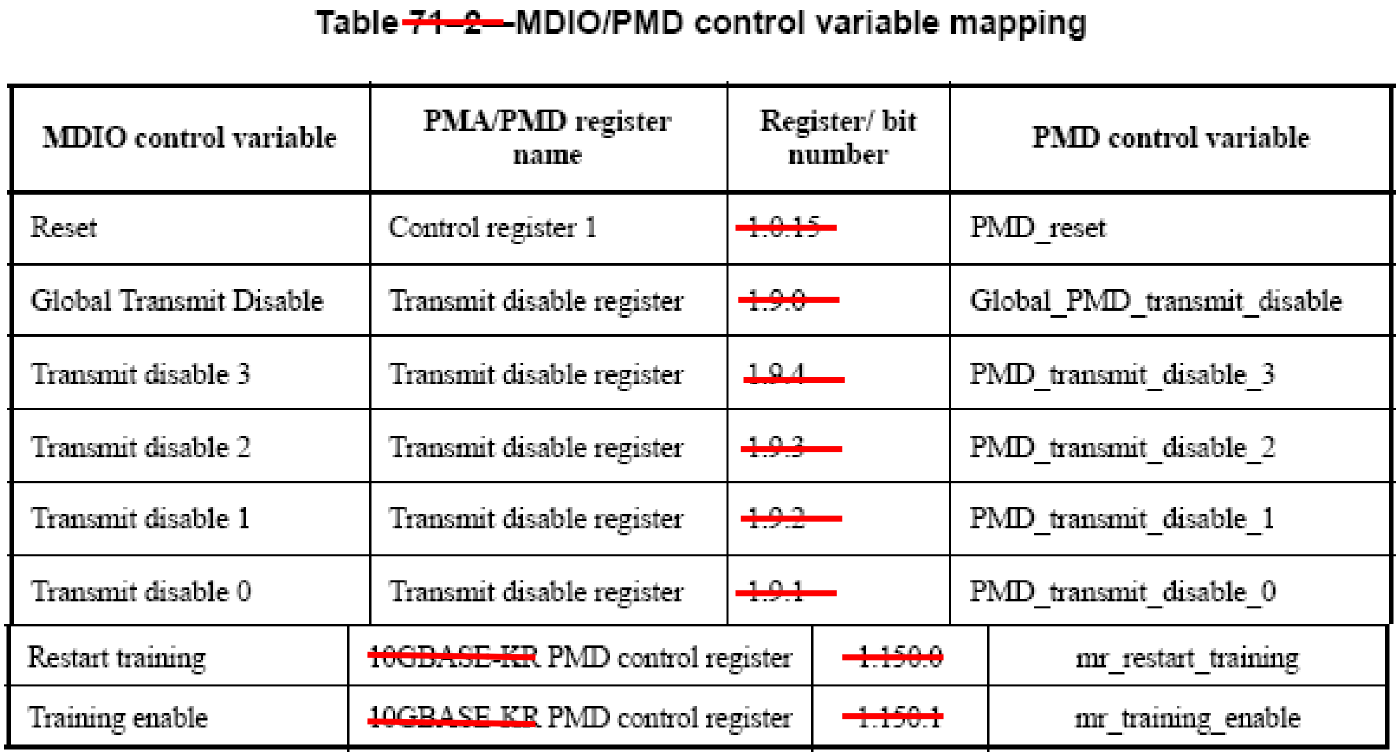


Table ~~71–2—~~MDIO/PMD control variable mapping

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

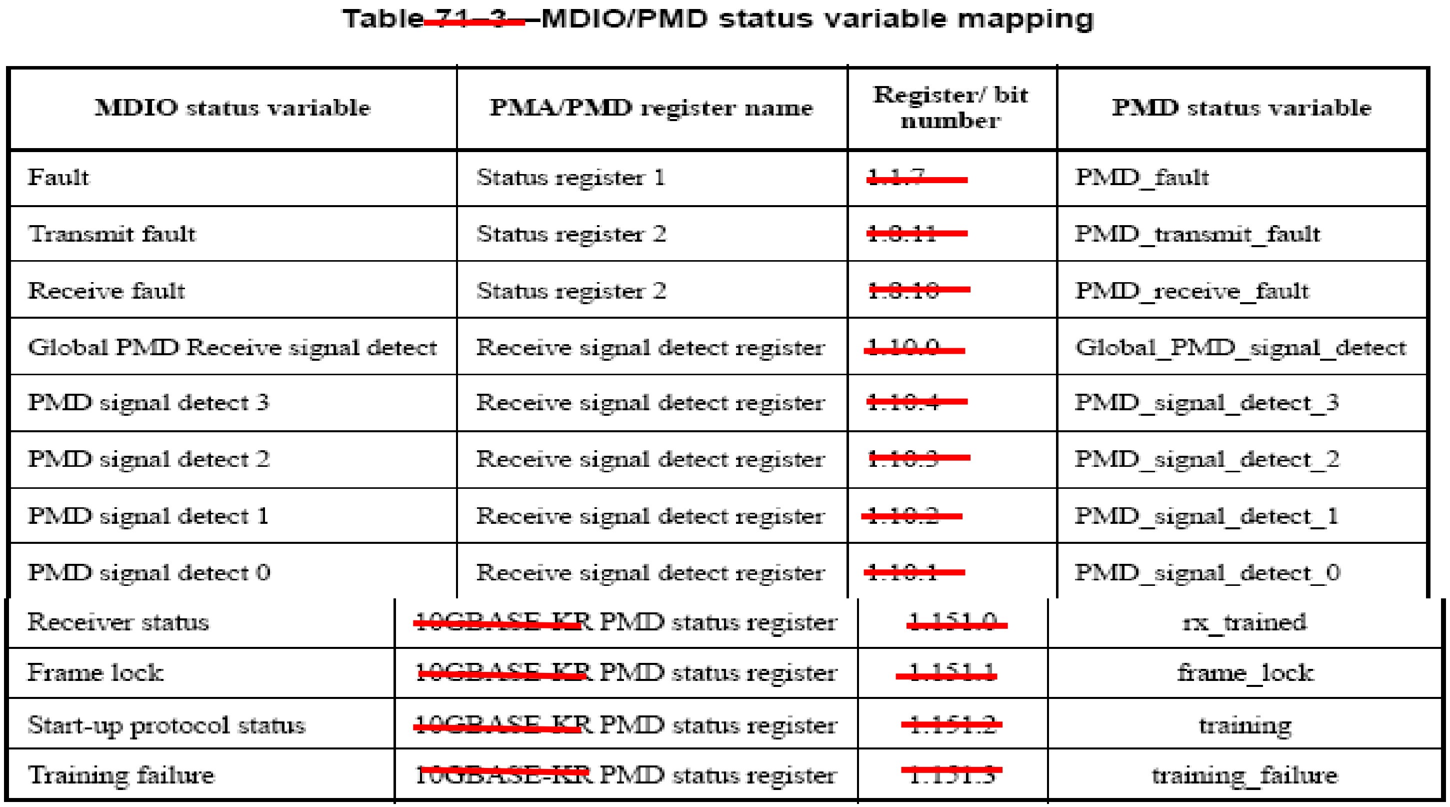


Table ~~71–3~~—MDIO/PMD status variable mapping

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

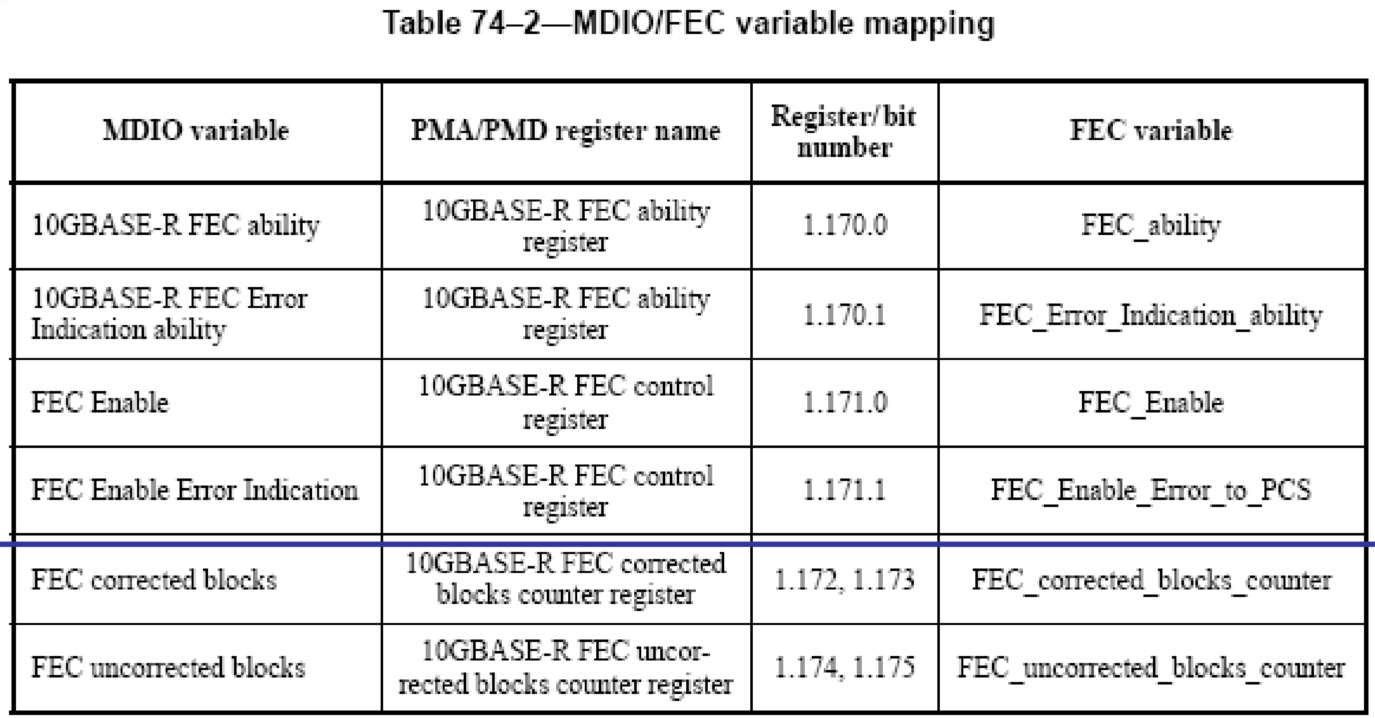


Table 74–2—MDIO/FEC variable mapping

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

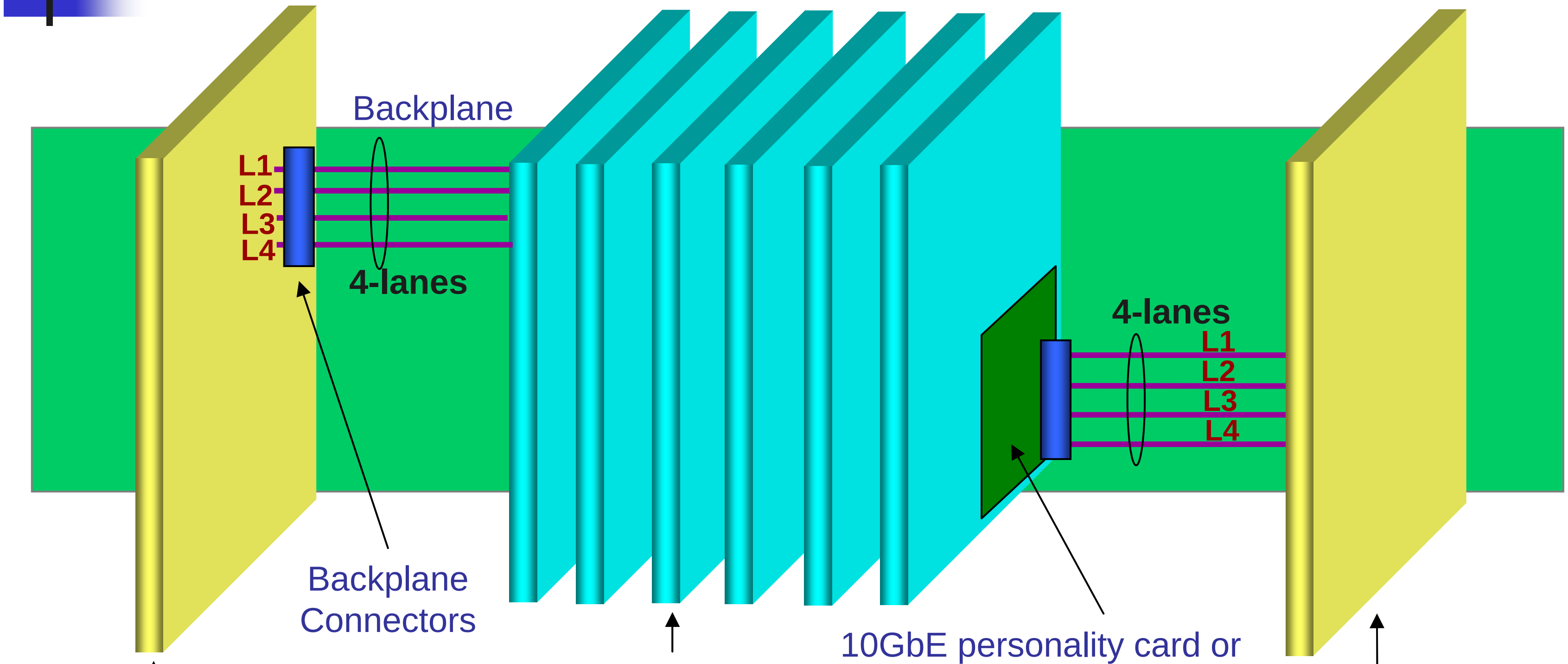
- Appropriate changes to add EEE feature, when adopted by 802.3az for KR
- PCS proposals and interface definitions to accommodate backplane Ethernet architecture (including FEC and AN)

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



Backplane Connectors

10GbE Switch card

Compute Blades or Line cards

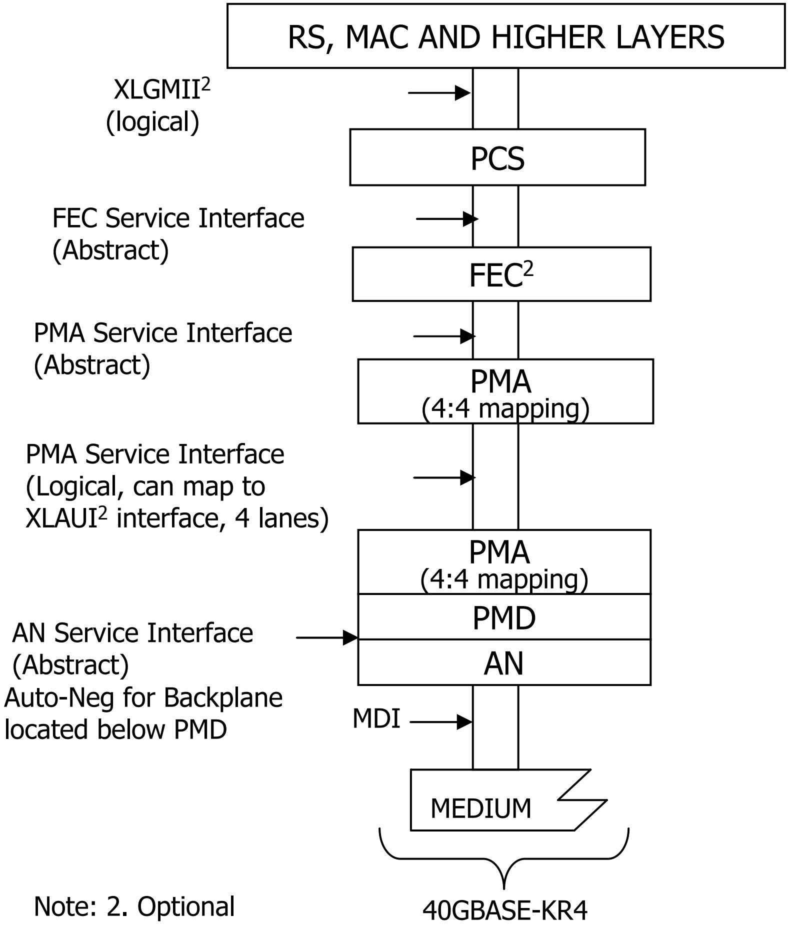
10GbE personality card or mezzanine (Some configurations may not include a mezzanine)

10GbE Switch card (Redundant)

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



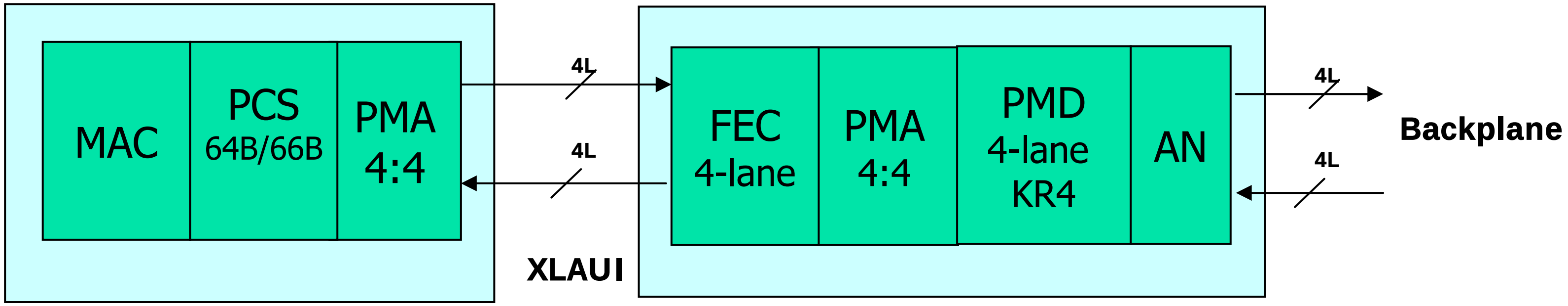
RS, MAC AND HIGHER LAYERS
XLGMII² (logical)
PCS
FEC Service Interface (Abstract)
FEC²
PMA Service Interface (Abstract)
PMA (4:4 mapping)
PMA Service Interface (Logical, can map to XLAUI² interface, 4 lanes)
PMA (4:4 mapping)
PMD
AN
AN Service Interface (Abstract)
Auto-Neg for Backplane located below PMD
MDI
MEDIUM
40GBASE-KR4

Note: 2. Optional

See ganga_01_0508 for 40/100G architecture and interfaces

Possible implementation examples

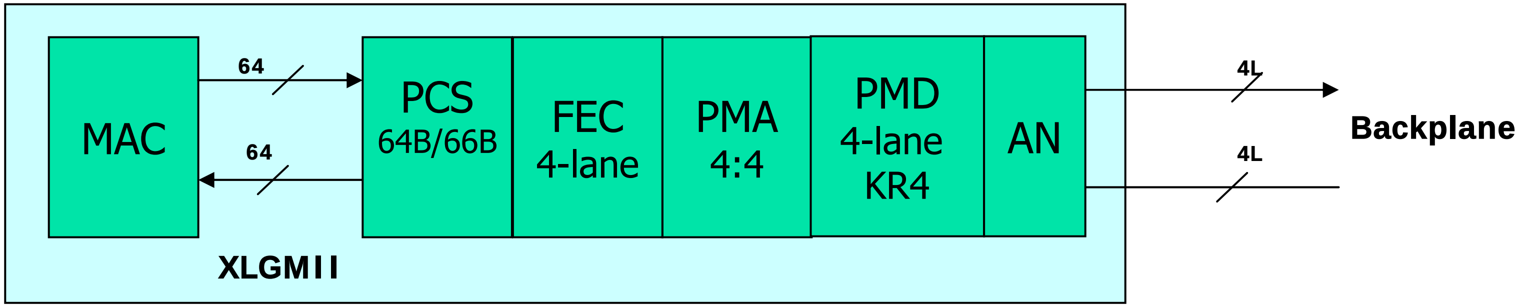
40GbE
separate
Backplane
PHY



MAC
PCS
64B/66B
PMA
4:4
4L
4L
XLAUI
FEC
4-lane
PMA
4:4
PMD
4-lane
KR4
AN
4L
4L
Backplane

MAC-PHY interface
(Chip to Chip)

40GbE
integrated
backplane
PHY



MAC
64
64
XLGMII
PCS
64B/66B
FEC
4-lane
PMA
4:4
PMD
4-lane
KR4
AN
4L
4L
Backplane

MAC-PHY interface
(intra-chip)