

IEEE P802.3bs Baseline Summary

John D'Ambrosia, Dell
Chair, IEEE P802.3bs 400 GbE Task Force

February 6, 2015

Topic Matter	Motion	Reference Presentation
Architecture	Motion #3, Jan 15: Move to adopt slides 4 and 8 from dambrosia_3bs_02b_0115 as baseline architecture.	http://www.ieee802.org/3/bs/public/15_01/dambrosia_3bs_02b_0115.pdf
RS / CDMII	Motion #3, July 14: Move to adopt the baseline for the CDMII logical interface as shown in slide 5 of gustlin_3bs_03_0714.pdf.	http://www.ieee802.org/3/bs/public/14_07/gustlin_3bs_03_0714.pdf
Electrical Interfaces (C2C and C2M)	Motion #4, Sept 14: Move to adopt 16 x 25Gb/s and 8 x 50Gb/s as the basis for the lane rates for any optional C2C and C2M electrical interfaces	
C2C / C2M 25G Electrical	Motion #6, Sept 14: Move to adopt the P802.3bm C2C and C2M specifications with current values (except that the BER requirement is TBD) as a baseline draft for the 16 x 25Gb/s electrical interfaces	
C2C Informative Channel	Motion #6, Jan 15: Move to adopt the following equation as the informative insertion loss equation for CDAUI-8 chip-to-chip electrical I/O interface $IL \leq \{ 1.083 + 2.543\text{SQROOT}(f) + 0.761f \quad 0.01 \leq f \leq 28.05\text{GHz} \} \text{ dB}$	
C2M Informative Channel	Motion #8, Jan 15: Move to adopt the following equation as the informative insertion loss equation for CDAUI-8 chip-to-module electrical I/O interface $IL \leq \{ 1.076(0.075 + 0.537\text{SQROOT}(f) + 0.566f) \quad 0.01 \leq f \leq 28.05\text{GHz} \} \text{ dB}$	
EEE	Motion #4, Jan 15: Move to adopt the EEE baseline proposed in marris_3bs_01_0115.pdf slide 7.	http://www.ieee802.org/3/bs/public/15_01/marris_3bs_01_0115.pdf
OTN	Motion #5, Jan 15: Move to adopt slide 10 of trowbridge_3bs_01a_0115.pdf as the baseline for the OTN mapping reference point	http://www.ieee802.org/3/bs/public/15_01/trowbridge_3bs_01a_0115.pdf
100m MMF	Motion #3, Nov 14: Move to adopt the proposal in slides 6 to 16 in king_3bs_02a_1114.pdf as the baseline proposal for the P802.3bs objective to “provide physical layer specifications which support link distances of at least 100 m of MMF” (400GBASE-SR16)*	http://www.ieee802.org/3/bs/public/14_11/king_3bs_02a_1114.pdf
10km SMF	Motion #4, July 14: Move that 10km 400GbE SMF PMD will use a duplex fiber solution.	

400GbE Architecture Baseline Proposal (Update)

IEEE P802.3bs 400 Gb/s Ethernet Task Force

January 2015

Pete Anslow - Ciena
John D'Ambrosia – Dell
Mark Gustlin – Xilinx
Adam Healey – Avago
David Law – HP
Gary Nicholl - Cisco
Dave Ofelt – Juniper
Steve Trowbridge - ALU

What Needs to be Supported in the Architecture?

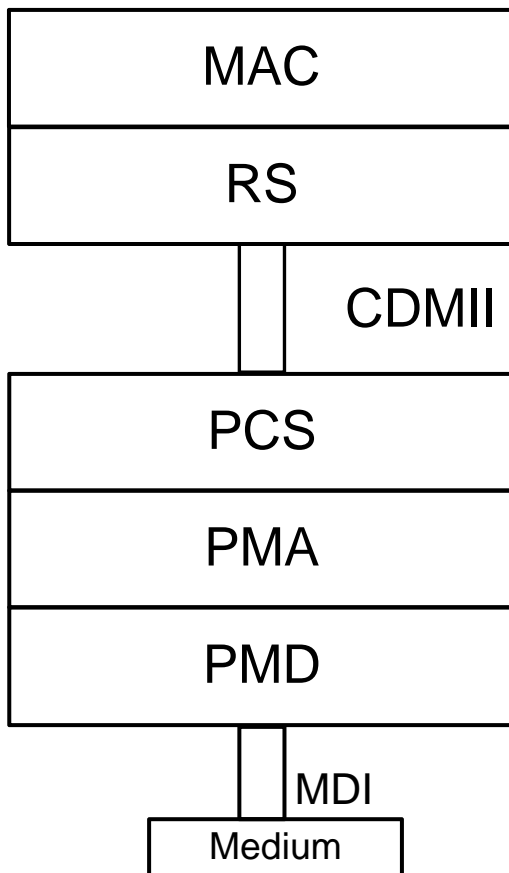
- The coding needs of the electrical interface may vary independently from the PMD interface
- The requirements for each interface can be different, both the FEC, modulation and number of lanes can change over time for each interface
- We need a single high level architecture which can support the evolving requirements of the interfaces over time
 - This does not mean it requires a complicated implementation
- A Media Independent interface needs to be specified to enable standardization of different PHYs today and future, “unknown”, PHYs tomorrow.
- We need an electrical interface between different devices, CDAUI (C2C & C2M)
- IEEE 802.3 supports two “levels” of implementers
 - The system implementer
 - The component implementer

Sublayer Functions (at a high level)

Sublayer	10GbE	100GbE	400GbE (proposed)
MAC	Framing, addressing, error detection	Framing, addressing, error detection	Framing, addressing, error detection
Extender	XGXS (PCS + PMA function)	N/A	CDXS (PCS + FEC function)
PCS	Coding (X: 8B/10B, R: 64B/66B), lane distribution, EEE	Coding (64B/66B), lane distribution, EEE	Coding, lane distribution, EEE, FEC
FEC	FEC, transcoding	FEC, transcoding, align and deskew	N/A
PMA	Serialization, clock and data recovery	Muxing, clock and data recovery, HOM	Muxing, clock and data recovery, HOM??
PMD	Physical interface driver	Physical interface driver	Physical interface driver

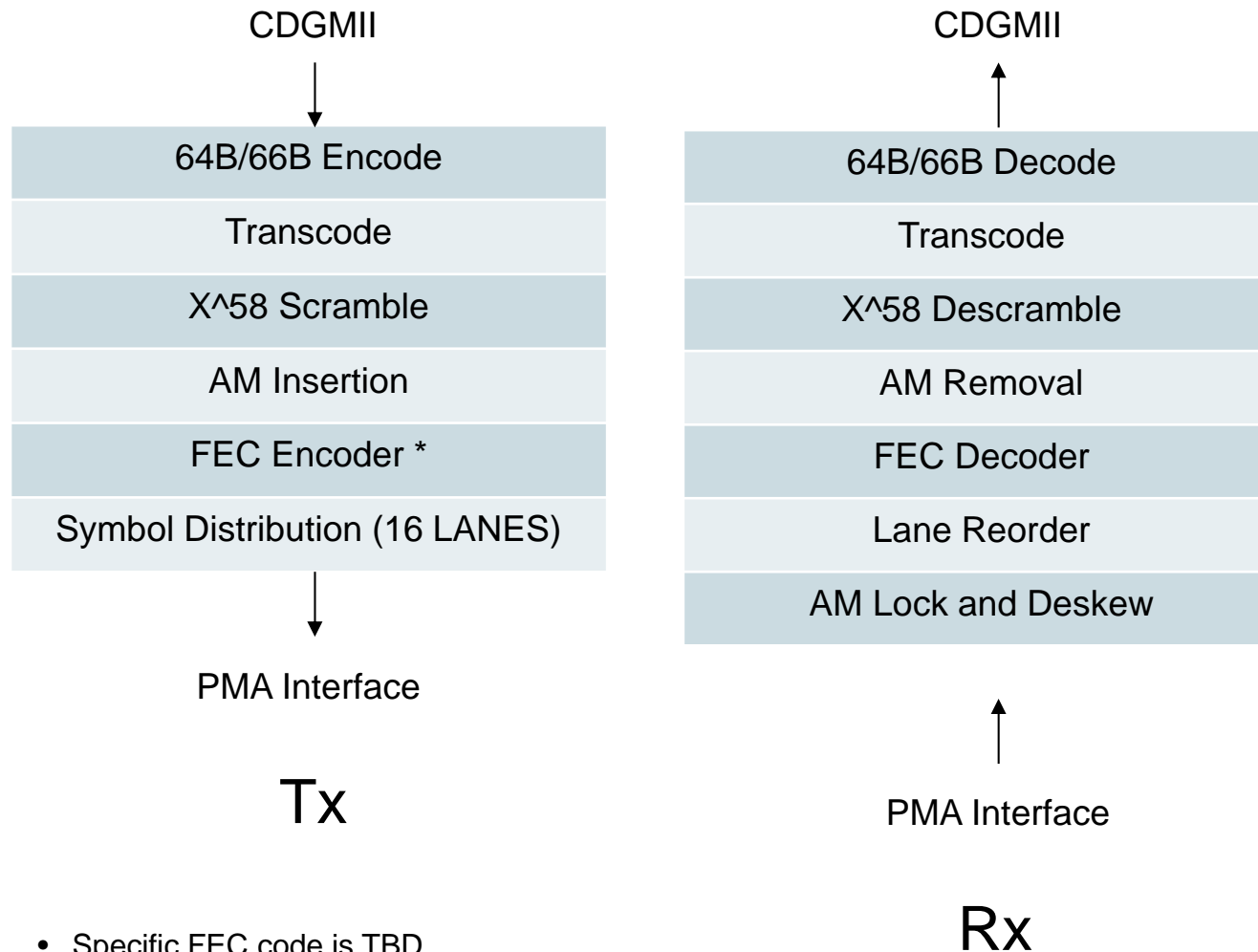
Note that there are variations with a single speed, not all are captured in this table

The 400GbE Basic Layer Diagram



- But...
 - To enable flexibility for future efforts, an extender sublayer for the CDMII is desirable, but there is no physical instantiation of the CDMII.
 - From a standardization perspective, it can leverage a CDAUI, which is an optional physical instantiation of the PMA service interface

PCS Block Diagrams



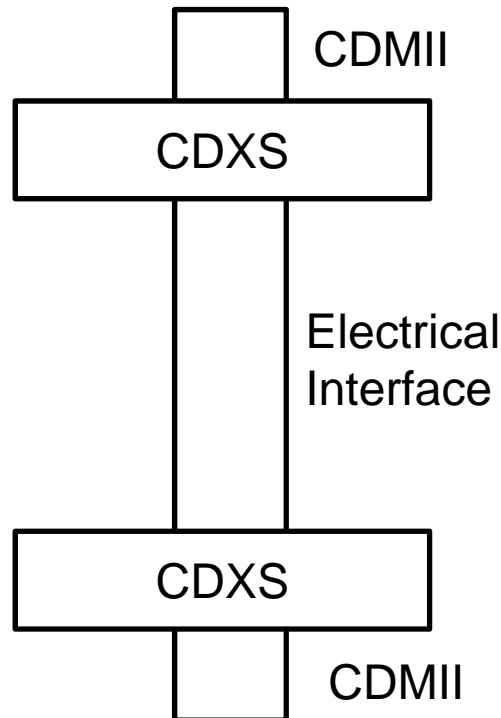
- Specific FEC code is TBD
- From gustlin_3bs_02_0115

PMA

The following are the functions performed by the PMA sublayer

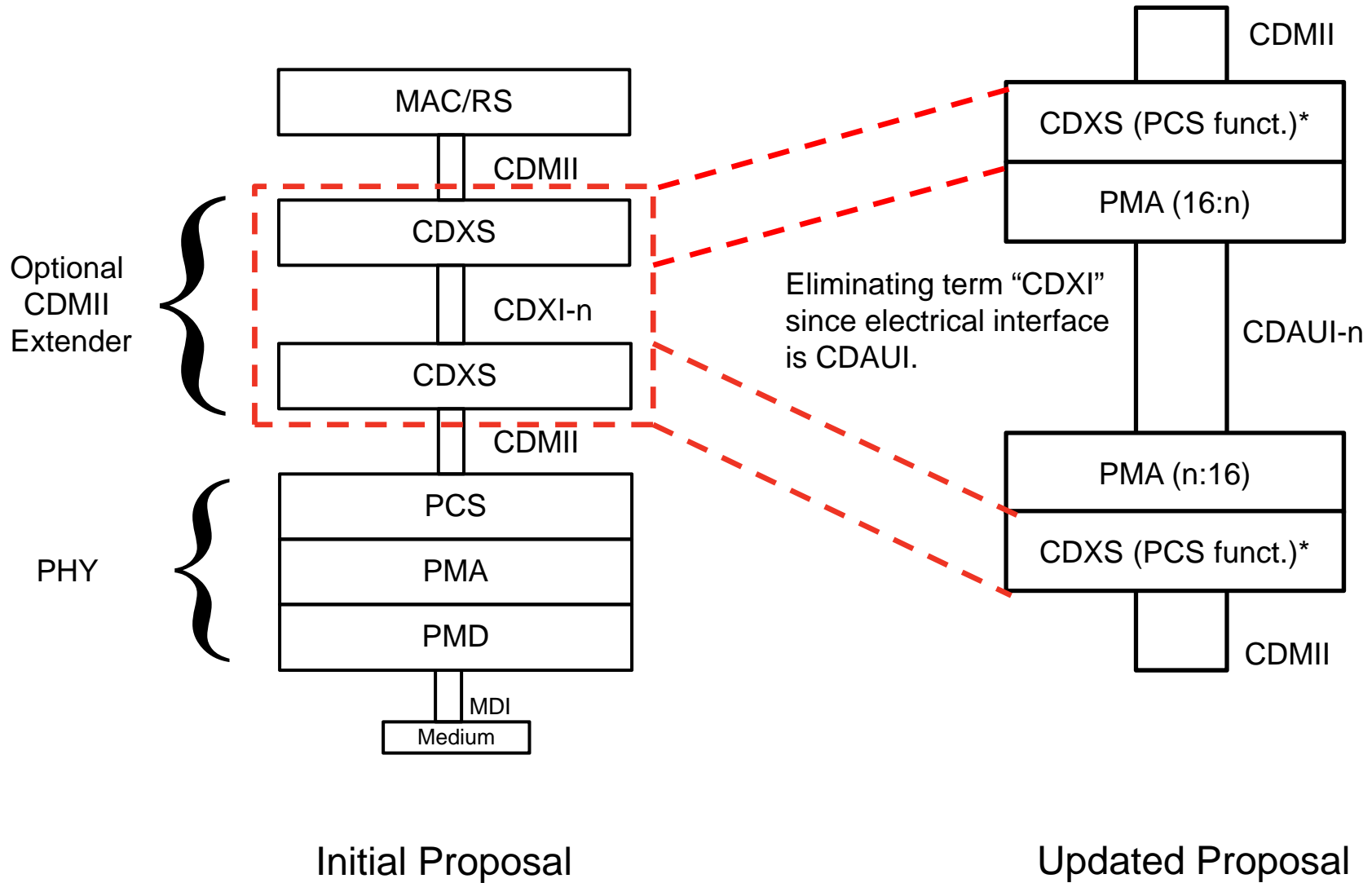
- Provide appropriate multiplexing
 - Provide appropriate modulation (PAM4 for instance if required)
 - Provide per input-lane clock and data recovery
 - Provide clock generation
 - Provide signal drivers
 - Optionally provide local loopback to/from the PMA service interface
 - Optionally provide remote loopback to/from the PMD service interface
 - Optionally provide test-pattern generation and detection
 - Tolerate Skew Variation
-
- From `gustlin_3bs_02_0115`

Comments on CDXS



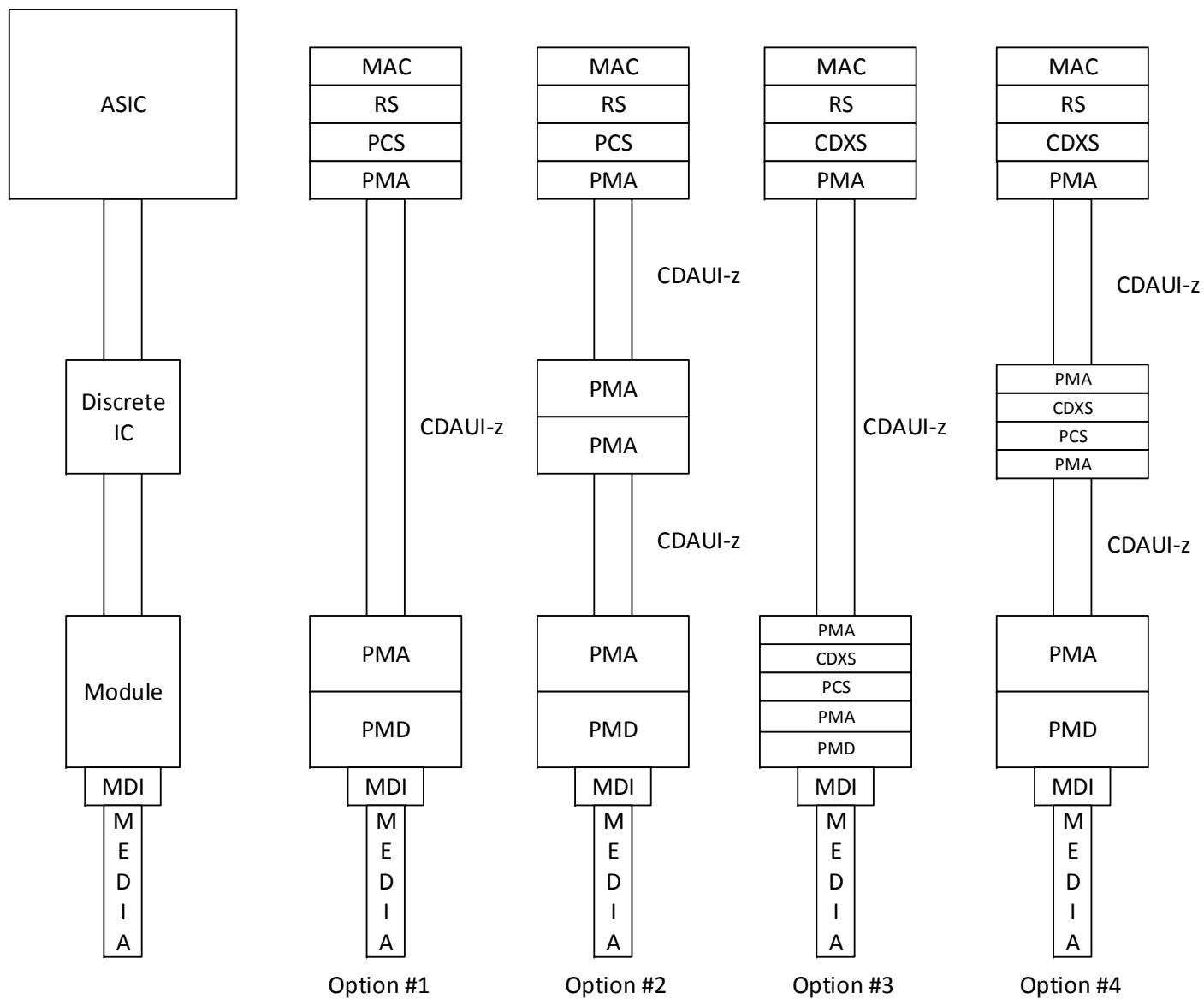
- **CDMII is the only media independent interface**
- **Different implementations or future PHYs may require changing FEC, which would require a return to CDMII (from a standardization perspective)**
- **The CDXS, as shown, is an extension of the CDMII.**
- **This allows support for new PCS / PMA functionality below the extended CDMII, if needed.**
- **The CDXS provides the coding / FEC of the electrical interface, not the coding / FEC of the PHY.**

CDMII Extender Functional Concept



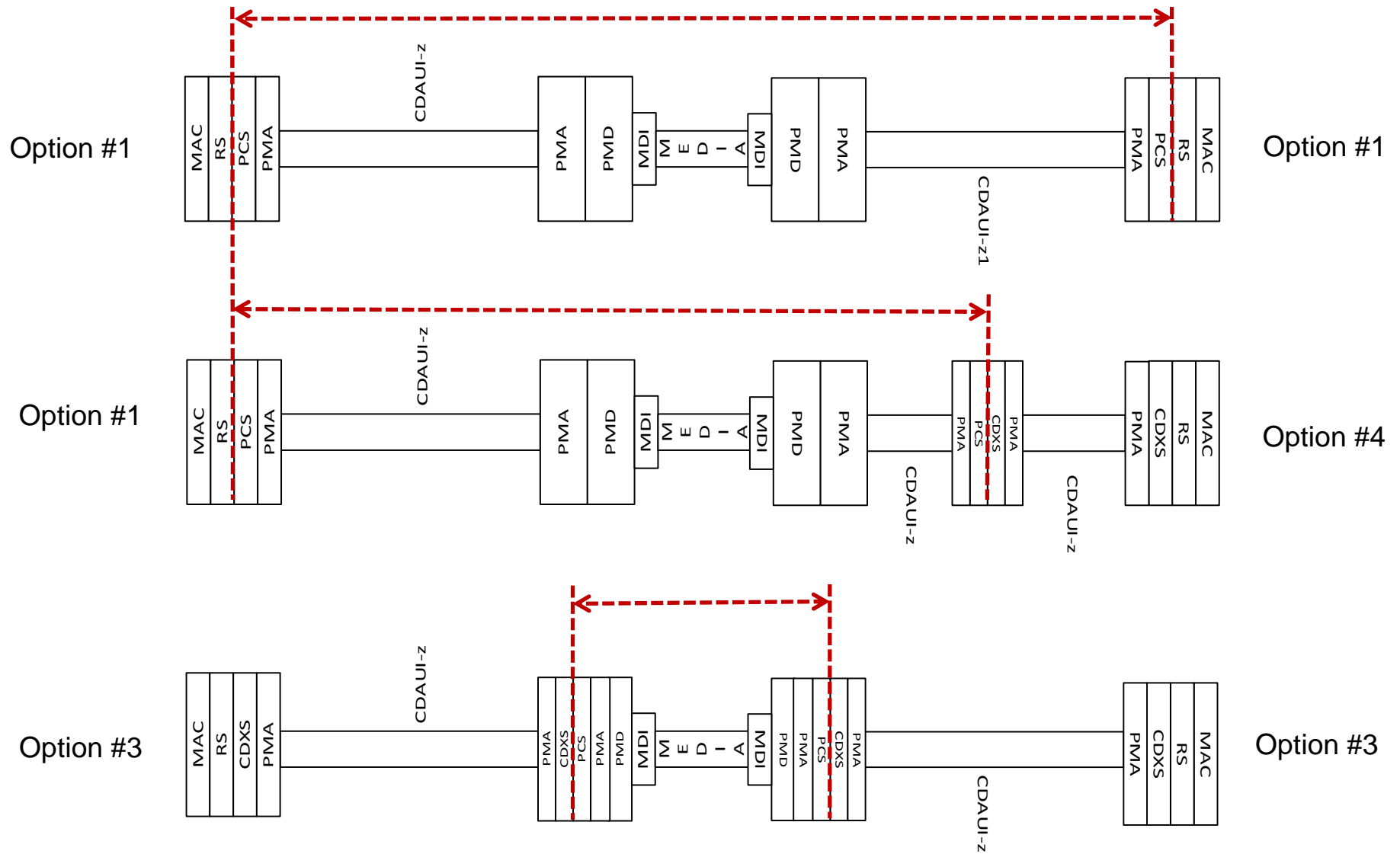
*** Note - Same as PCS (including FEC) to be defined.**

400GbE Example Implementations



“z” may be different for various interfaces cited for CDAUI

Leveraging the Proposed Architecture





Thanks!

400GbE MII Baseline Proposal

IEEE P802.3bs 400 Gb/s Ethernet Task Force

July 2014 San Diego

Mark Gustlin – Xilinx

Supporters

- John D'Ambrosia – Dell
- Arthur Marris – Cadence
- Dave Ofelt – Juniper
- Steve Trowbridge - ALU

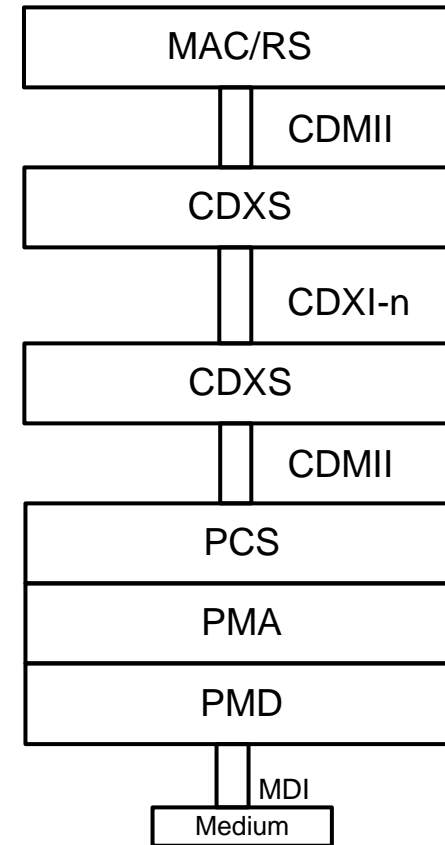
Proposed 400GbE Architecture

- The protocol stack diagram shows one possible implementation
- The CDMII connects the MAC/RS sublayer to the Extender sublayer or the PCS

CDMII is the 400 Gb/s Media Independent Interface

CDXS is the 400 Gb/s extender sublayer

CDXI is the interface between two extender sublayers



CDMII Interface

➤ Why define it?

- Electrically it won't be directly instantiated, but in the proposed 400GbE architecture it can be extended with an extender sublayer (CDXS) and interface (CDXI-n)
- Some will want it for RTL to RTL connections within devices

➤ Define it as a logical Interface only

- Unless it is extended, then there is a physical instantiation via an extender sublayer

What is it?

- Base it directly on clause 81
- Same signal structure as shown below, just run faster, or in parallel

81.1.6 XLGMII/CGMII structure

The XLGMII/CGMII is composed of independent transmit and receive paths. Each direction uses 64 data signals (TXD<63:0> and RXD<63:0>), 8 control signals (TXC<7:0> and RXC<7:0>), and a clock (TX_CLK and RX_CLK). Figure 81–2 depicts a schematic view of the RS inputs and outputs.

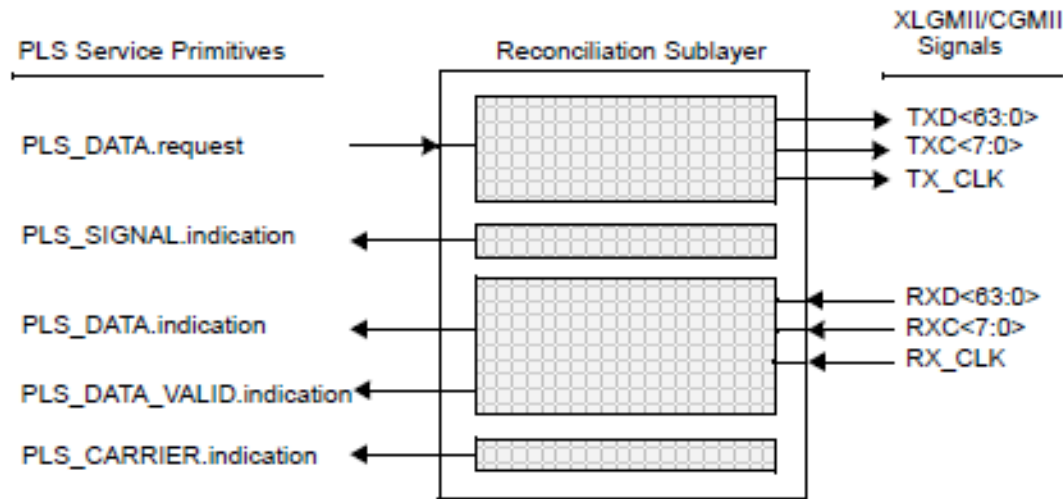


Figure 81–2—Reconciliation Sublayer (RS) inputs and outputs

Extender Sublayer (CDXS)

- The CDXS is the proposed extender sublayer to extend the CDMII
 - A typical instantiation is a high speed parallel SerDes interface
- It is optional, only used if the PCS does not cover both the electrical and optical interface needs
- The CDXS can contain PCS, FEC, and PMA functionality

Thanks!

Energy Efficient Ethernet (EEE) for 802.3bs

Arthur Marris - Cadence

Supporters

- Mark Gustlin – Xilinx
- Andre Szczepanek – Inphi
- Dave Ofelt – Juniper
- Gary Nicholl - Cisco

Key features of Energy Efficient Ethernet (EEE)

- If a system has nothing to transmit it can power down its transmit path after it has indicated its intention to do so.
- The partner device can also power down its receive path when it detects that the path is going to be powered down.
- The link stays up while in low power mode and no frames are dropped.
- EEE is asymmetric. One direction can be powered up while the other is powered down.

How does EEE work?

- The client (i.e. system) requests Low Power Idle (LPI) from the reconciliation sublayer (RS).
- The RS then signals LPI on the MII (media independent interface).
- The transmit PCS (physical coding sublayer) encodes the LPI signal using a special symbol.
- The receive PCS decodes the LPI symbol and signals LPI on the receive MII.
- When the client ceases requesting LPI the RS continues inhibiting transmission for a fixed period to allow time for the local transmit path and remote receive path to recover from their low power modes.

What is “Fast Wake” mode of operation?

- In Fast Wake mode of operation the transmit and receive path remain active and continuously transmit and receive LPI when it is requested by the client. (However Clause 82 BIP is not calculated)
- Fast Wake is compulsory for 40G and 100G PHYs that support EEE.
- Fast Wake is controlled by LLDP (Link Layer Discovery Protocol) rather than AN (auto-negotiation)
- Fast Wake is suitable for optical PHYs that unable to power up and down quickly and do not support AN.

What is “Deep Sleep” mode of operation?

- In Deep Sleep mode the transmit path stops transmitting but periodically sends refresh indication while LPI is requested.
- In Deep Sleep mode the receiver checks for the occurrence of refresh indication and will assert link failure if refresh does not appear.
- Deep Sleep mode requires the receive PMA and PCS to resume operation within a determined time period.
- In Deep Sleep mode the transmit and receive PCS generate tx_quiet and rx_quiet signals to allow the PHY to periodically power down its transmit and receive path.
- Clause 83 defines a mechanism for sending the tx_quiet signal over the CAUI/XLAUI interface and for synthesizing the rx_quiet signal.
- As only optical PHYs are included in the 802.3bs objectives, none of the PHYs specified in 802.3bs will support Deep Sleep operation, however the architecture should not preclude support for Deep Sleep in the future.

What should be done for 802.3bs?

- Adopt Fast Wake mode of operation for the 802.3bs PHY types.
- Add these PHY types to “[Table 78-1 Clauses associated with each PHY or interface type](#)” and indicate that they do not support deep sleep with the “b” suffix.
- The CDMII will need to be able to signal LPI and the RS will need to include a transmit LPI state machine to defer transmission for the wake time period after de-assertion of LPI.
- The PCS will need to be able to encode and decode LPI.

Further considerations for supporting EEE in 802.3bs

- Consider whether the receive PCS should have a “RX_FW” state that it goes into when receiving LPI from its link partner. In this state it could disable error checking.
- When specifying the CDAUI electrical interface consider defining a mechanism for transmitting alert and quiet signalling similar to how it is done in 83.5.11.1 to allow for future support of Deep Sleep, and also consider generation of an energy detect signal.
- To support Deep Sleep mode the PCS needs to achieve synchronization and alignment quickly. This will most likely be done through the use of “rapid alignment markers” so make sure the architecture does not preclude their use in the future.
- If Deep Sleep is supported in the future then the CDXI may need to be powered down. In this case it would be necessary to preserve the MII signalling of LPI to the PCS in the transmit direction and LPI to the MAC in the receive direction until the interface powers up. Also it will be necessary for the CDXI to power up quickly.

Summary

- 802.3bs will use EEE fast wake functionality which requires the CDMII to signal LPI and the PCS to encode and decode it
- To allow for future support of deep sleep functionality consideration needs to be given to how the PCS and electrical interfaces can resume operation quickly after power down

OTN Support Proposal

P802.3bs 400 Gb/s Ethernet Task Force

Steve Trowbridge
Alcatel-Lucent

Supporters

- Pete Anslow (Ciena)
- Gary Nicholl (Cisco)
- Tongtong Wang (Huawei)
- Xinyan Wang (Huawei)
- David Ofelt (Juniper)
- Andre Szczepanek (Inphi)

Key Elements of OTN Support

- See “[OTN Support: What is it and why is it important?](#)”, July 2013
 - A new rate of Ethernet (e.g., 400 Gb/s) fits into the corresponding rate OTN transport signal
 - All Ethernet PHYs of a given rate are mapped the same way and can be interconnected over the OTN (e.g., same PCS for all 100 Gb/s PHYs gives a single canonical format (“characteristic information” in ITU-T terminology) that can be mapped
 - Optical modules for Ethernet can be reused for OTN IrDI/client interfaces at the corresponding rate

A new rate of Ethernet (e.g., 400 Gb/s) fits into the corresponding rate OTN transport signal

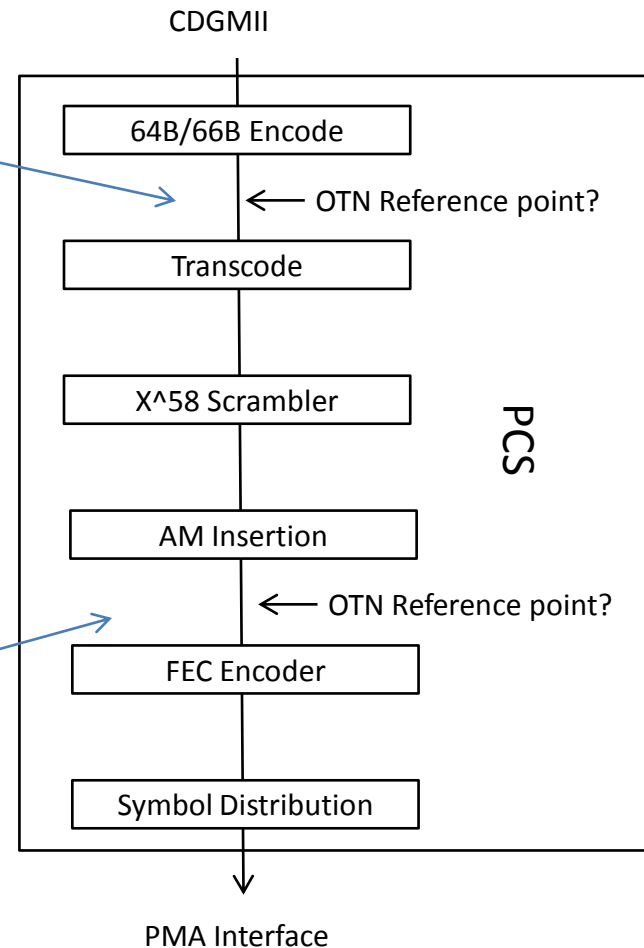
- Assumption – the OTN mapper/demapper will terminate and regenerate any Ethernet FEC code, correcting errors at the OTN ingress since the FEC is chosen to correct single-link errors but not double-link errors
- Assumption – the OTN mapper/demapper may trans-decode/trans-encode back to 64B/66B to avoid MTTFPA reduction for OTN transported signal
- Based on these assumptions, the encoded data rate of the OTN-mapped 400 Gb/s Ethernet would be no more than $400 \text{ Gb/s} \times 66 / 64 = 412.5 \text{ Gb/s} \pm 100\text{ppm}$. Since the 400 Gb/s OTN container would presumably be designed to also transport four “lower order” ODU4s, there should be no concern that it is large enough to carry 400 Gb/s Ethernet based on the assumption that the canonical form is near this rate.
- Any Ethernet bits in excess of this rate are likely to be part of a FEC that is not carried over OTN

All PMDs of a given rate are mapped in the same way into OTN:

Candidate Locations for the OTN Reference point

Option B

Option A



OTN Reference Point Option A

- Pro
 - Most similar to what ITU-T chose for mapping of 40GbE into OPU3 and 100GbE into OPU4 based on 802.3ba assumption that all PMDs of a given rate used exactly the same logical lane striping
 - Capable of carrying end-to-end information in the alignment markers if all PMDs are striped the same
- Con
 - Even if all P802.3bs PMDs use the same logical lane striping, if all future 400GbE PMDs are not striped in the same way, requires re-striping into the canonical format before mapping into OTN
 - Relies on OTN FEC for satisfactory MTTFPA

OTN Reference Point Option B

- Pro
 - High certainty to be a common format regardless of whether all 400GbE PMDs (P802.3bs and future) use the same logical lane striping
 - No need to convert between logical lane formats in OTN mapper/demapper
 - Robust MTTFPA regardless of what ITU-T does with FEC
- Con
 - Can't carry end-to-end information in alignment markers. But note that there doesn't seem to be a compelling reason to support BIP inside of FEC-enabled Ethernet PMDs
 - Higher bit-rate to map as 64B/66B rather than mapping the transcoded signal.

Straw Polls relevant to OTN support

- Straw Poll #1 - I support FEC for optical PMDs
 - **FEC mandatory – 69**
 - FEC optional – 7
 - Some PMDs may not need FEC – 0
 - Mandatory for some, optional for others – 10
 - Need more information – 10
- Straw Poll #9 - If all PMDs developed in P802.3bs include mandatory FEC and FEC error statistics are available, do we also require BIP?
 - Y: 4; **N: 24; A: 69**
- Straw Poll #10 – If BIP is required, should it be:
 - Segment by segment (optimized for fault isolation) – 2
 - End to end (optimized for service assurance) – 6
 - **Need more information – 35**
 - **Not required/don't care - 33**

Observations

- Strong majority believe that all optical PMDs developed by P802.3bs will have mandatory FEC
- Most think that if all optical PMDs have mandatory FEC, that BIP is not necessary, i.e., you get a better view of link quality from FEC corrected errors
 - Note that any BIP inside of a FEC exhibits a “cliff” behavior, going from zero errors to quite a lot the instant the error ratio exceeds the correction capability of the FEC. In addition, this information is available from the FEC uncorrected codewords counter
- For those who still think there would be BIP, the number who express an opinion on how it is used (fault isolation or service assurance) is statistically insignificant
- For those who still think they need more information, please study slides 5-12 of [trowbridge 3bs 01 0714.pdf](#) and ask questions!

OTN Reference Point Proposal

- Propose that the OTN reference point for 400GbE is a logically serial stream of 64B/66B blocks, unstriped, without lane alignment markers (Option B). An implication of this choice is that there is no BIP carried end-to-end (note that a segment BIP for a particular lane striping and FEC would still be possible).
- Any idle insertion/deletion to provide room for striping overhead must occur between the CDGMII and the OTN reference point. No idle insertion/deletion should occur between the OTN reference point and the PMD. Assuming 16K frequency of lane markers (regardless of the lane count or lane rate), the effective rate of this logical interface is the *nominal MAC rate* $\times 66/64 \times (1 - 1/16384)$ so that any physical instantiation has room to insert lane markers as needed without idle insert/delete elsewhere in the stack

Module Reuse

- Module reuse was facilitated by the fact that nothing below a CAUI chip-to-module interface cared about the or manipulated the bit values on the lanes – as long as OTN was striped into the same number of logical lanes as Ethernet, everything would work
- The following likely can be preserved: no idle insertion/deletion occurs below a CDAUI chip-to-module interface
- The following are possibly not be precluded by the 400GbE architecture:
 - Logical to physical lane multiplexing in a module may be on a block or FEC symbol basis rather than a bit basis
 - One (possibly Ethernet Frame Format dependent) FEC code may be replaced with another)

Options for Module Reuse

- Option I: Preserve the 802.3ba rule that no sublayers below a CDAUI care about bit values or manipulate the bit values on logical lanes (bit multiplexing only). Any FEC is done on the host board above a CDAUI. OTN may use a different FEC than Ethernet if it needs a stronger FEC to compensate for the higher bit-rate
- Option II (most general, described in Norfolk) encode the OTN frame as 66B blocks (all data) and use whatever striping and FEC encoding mechanisms are used for Ethernet. OTN and Ethernet use the same FEC
- Option III (potentially very complex) Allow OTN to use a different (stronger) FEC than Ethernet but do not require bit multiplexing of logical lanes. This would constrain that OTN and Ethernet choose FEC (or pairs of FECs, if not all 400GbE PMDs use the same FEC) with the same symbol size and that the marking to discover the FEC symbol alignment is common between OTN and Ethernet

Option II Amplification

- Use the fact that the OTN reference point, as proposed, is a logically serial stream of 64B/66B blocks.
- Note that before this reference stream can be physically instantiated, it must be striped over multiple physical or logical lanes
- Maintain the principle, as in 802.3ba, that idle insertion/deletion is not done below this reference point.

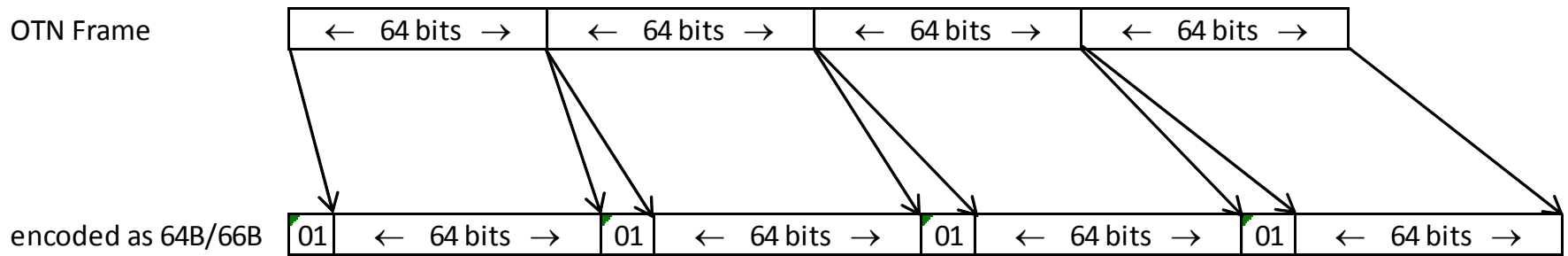
Option II Example

- Example physical instantiation could be something like [gustlin 400 02a 1113.pdf](#), produced by transcoding 64B/66B to 256B/257B, striping first into 100G groups, striping within each 100G group into 4 logical lanes on 10-bit symbol boundaries, inserting alignment markers on each lane, and applying an RS(528,514) code based on 10-bit symbols with alignment markers appearing in the first of each of 4096 Reed Solomon code blocks (essentially 4 instances of P802.3bj 100G FEC)

Option II Implications for OTN

- Likely only possible if the same FEC code can be used for OTN applications as for Ethernet applications at about 6% higher bit-rate
- Would need to make OTN look like 66B blocks. Easiest way to do this and not lose any information in transcoding is to insert a “01” sync header after every 64 bits (all data)
- Since this is just part of the logical frame format, this doesn't waste as many bits as it appears. 8 sync header bits are added to every 256 data bits in the “logical” frame format, but 7 of those bits are immediately recovered in 256B/257B transcoding and reused for the FEC code. So 0.39% net is added to the OTN frame to make it look like 66B blocks, then 2.724% overhead RS FEC added

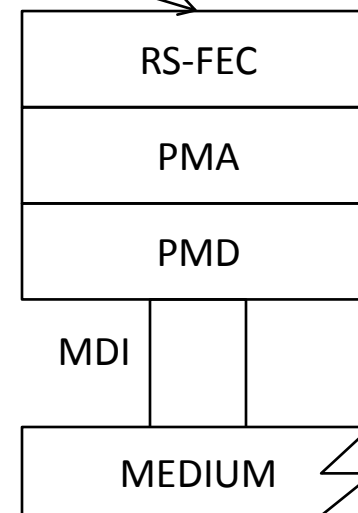
Option II - Illustration of turning OTN frame into 64B/66B blocks



Use the Ethernet Stack to stripe and FEC encode the OTN frame when carrying over an Ethernet Module for an OTN IrDI or client interface

Could be OTN frame aligned as an OTUC4 frame without FEC is exactly 7648×64 bits, but not essential with scrambling

Scramble



Option II: OTN Bit-rates using this scheme

	Working Assumption Bit-Rate
OTUC4 bit-rate without FEC	422.904 Gb/s
64B/66B encoded	436.120 Gb/s
256B/257B transcoded	424.556 Gb/s
Insert Lane Markers	424.582 Gb/s
Add RS(528,514) FEC	436.146 Gb/s
Logical Lane Rate (well within CEI-28G)	27.259 Gb/s
Ethernet Nominal Bit-rate	412.5 Gb/s
400G OTN Increase in bit-rate	5.73 %
100G OTN Increase in bit-rate	8.42 %

Smaller increase for 400G than for 100G, mainly due to RS(528,514) FEC rather than RS(255,239) FEC. Proportion remains the same even for Ethernet PMDs that use a higher overhead FEC

Option II – Recommended module reuse mechanism for OTN

- There is an Ethernet sublayer reference point such as the that is logically equivalent to a serial stream of 64B/66B blocks (the same as the recommended OTN reference point)
- No idle insertion/deletion occurs below the that reference point, and hence the rest of the stack can deal with a constant-bit-rate (CBR) bitstream that is effectively an infinite-length packet.
- Note that any logical to physical lane interleaving that works for Ethernet also works for OTN since they are encoded the same way
- The link parameters and FEC coding gain have sufficient margin to meet the error performance target when running at approximately 5.73% higher bit-rate than necessary for 400G Ethernet

THANKS!

400 Gb/s 100 m MMF reach objective draft baseline proposal

MMF ad hoc

IEEE P802.3bs, San Antonio, TX, Nov 2014

Outline

- Baseline proposal of a retimed PMD to address the 802.3bs objective to ‘Provide physical layer specifications which support link distances of at least 100 m over MMF’
 - 16 lane parallel, short wavelength based PMD for 400GBASE-SR16.
 - Leveraging 100GBASE-SR4 technology, compatible with 16 x 25 Gb/s electrical interface, and breakout applications.
 - Assumed use of 100GBASE-SR4 FEC, or similar strength FEC (to be defined in 802.3bs), to enable 100 m reach.
 - Architecture, parameters and specifications for optical interfaces, and the proposed MDI, follow.

Supporters and contributors

- John Abbott, Corning
- Piers Dawe, Mellanox
- Mike Dudek, QLogic
- Ali Ghiasi, Ghiasi Quantum LLC
- Mark Gustlin, Xilinx
- Jack Jewell, Commscope
- Jonathan King, Finisar
- Scott Kipp, Brocade
- Paul Kolesar, Commscope
- Brett Lane, Panduit
- Robert Lingle Jr., OFS
- Valerie Maguire, Siemon
- Slobodan Milijevic, Microsemi
- John Petrilla, Avago technologies
- Rick Pimpinella, Panduit
- Rick Rabinovich, Alcatel-Lucent Enterprise
- Steve Swanson, Corning
- Mike Zhang, Siemon

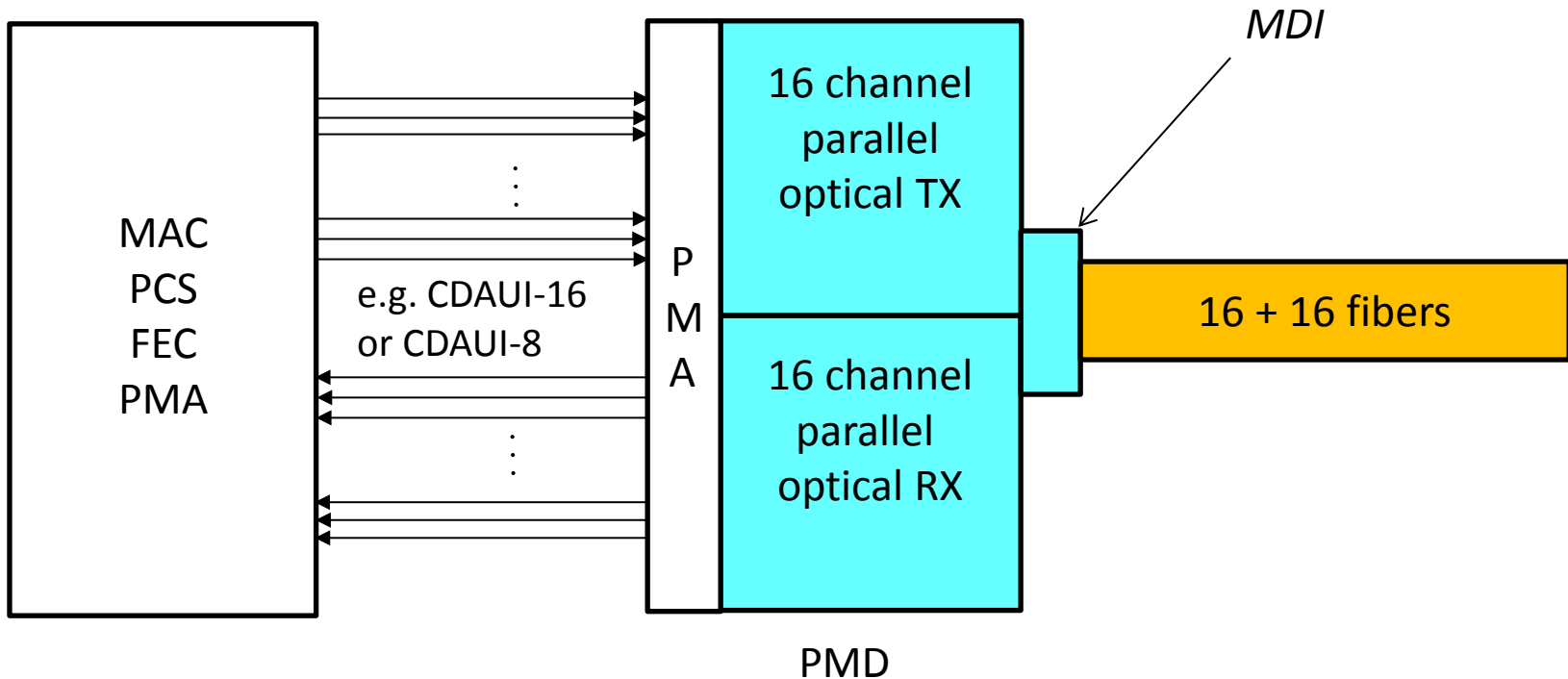
and a very nice letter of support
from the CDFP MSA

Motivation

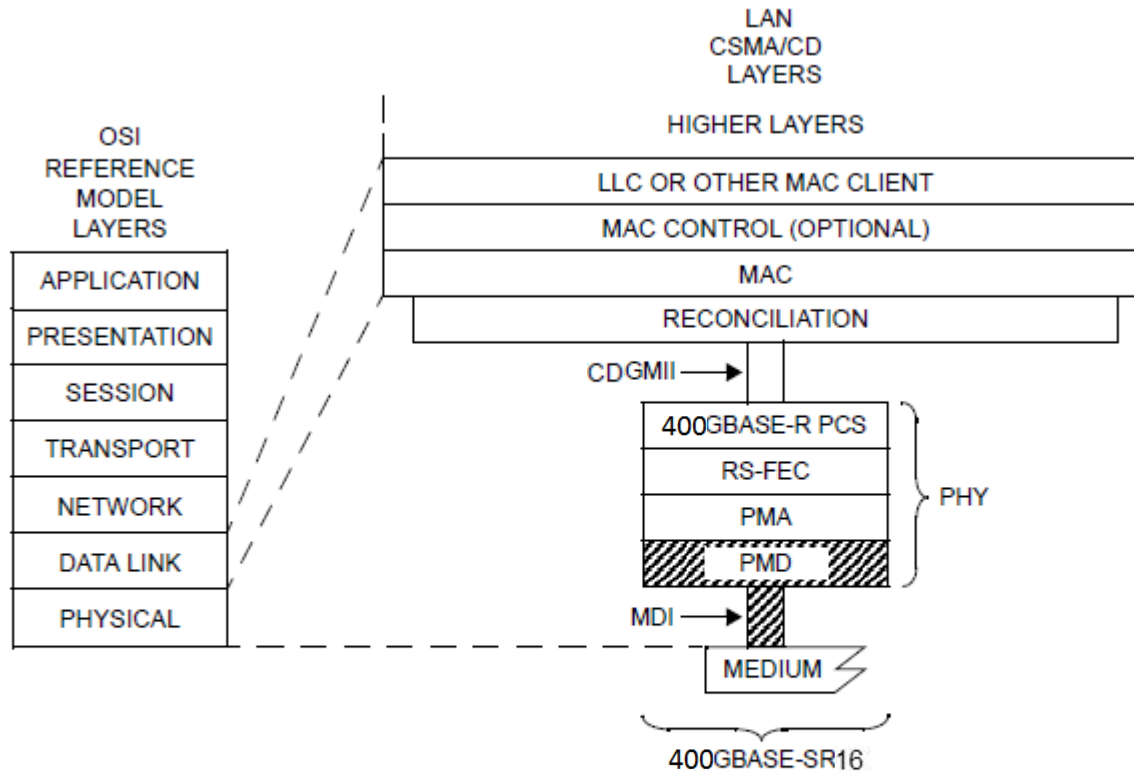
- 16 parallel links operating at 25.78125 GBd utilize low cost, high performing multimode fiber compatible optics and electronics
 - Leverages 100GBASE-SR4 technology
 - FEC supported retimed interface enables a lowest power, lowest cost, 100m solution today
 - Uses existing, viable semiconductor technologies and uncooled VCSELs
- The 16 optical lanes can directly map the 16 electrical lanes of CDAUI-16, without requiring multiplexing, translation, or de-skewing inside the module.
- Compatible with ‘break out’ application
- This proposal is supported by multiple vendors and users, and is economically feasible and competitive compared to other alternatives.

Proposal

- 16 parallel lanes @ 25.78125 GBd for 100GBASE-SR16 over 100 m OM4 fiber.
 - Exact signaling rate is determined by project's choice of FEC.
- 850 nm sources, re-use of 100GBASE-SR4 specifications.
 - Assumes PMD target BER (prior to error correction) around 5×10^{-5} , similar to 100GBASE-SR4.



Position in 802.3 architecture

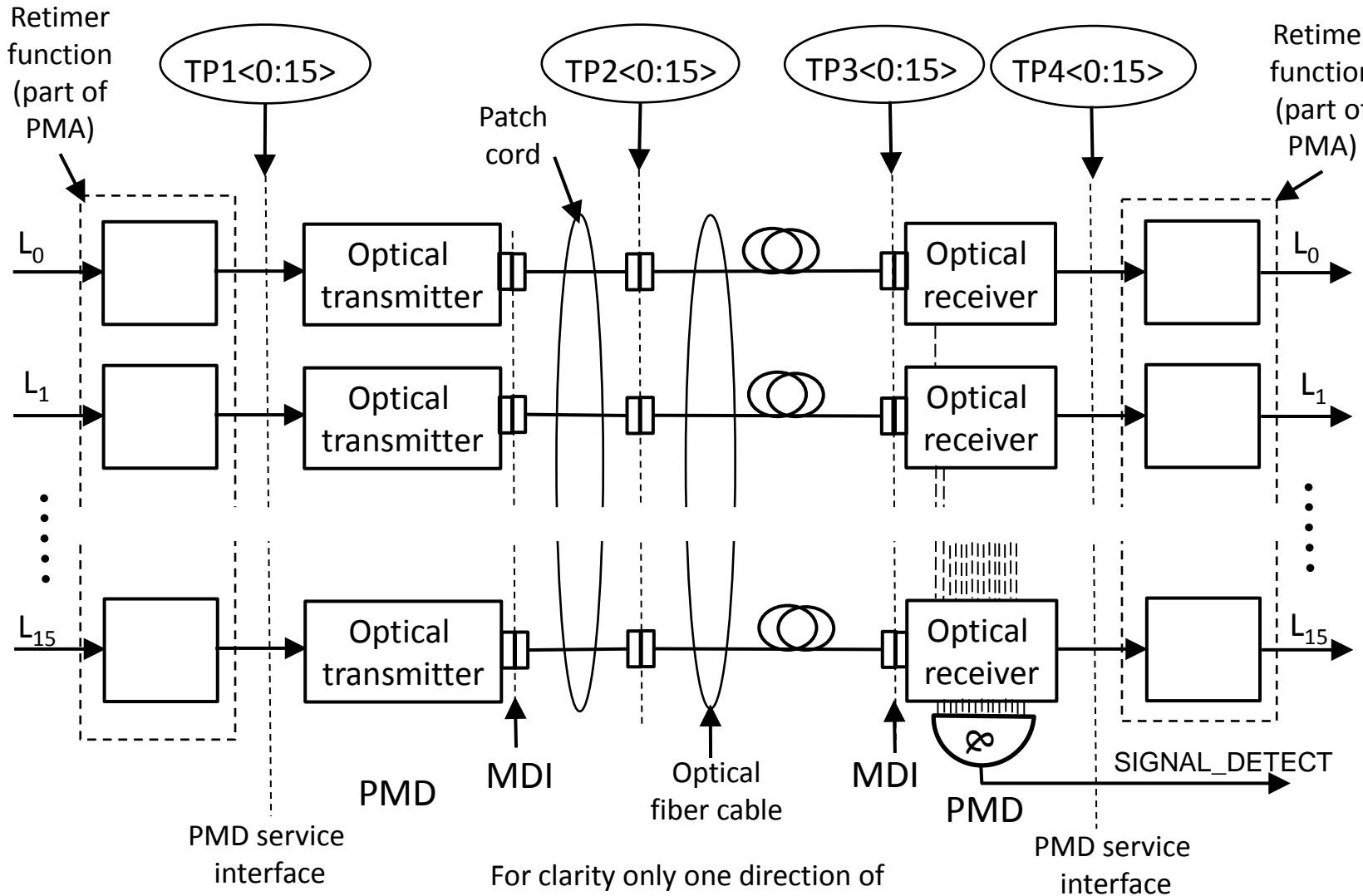


CDGMII = 400 Gb/s MEDIA INDEPENDENT INTERFACE
 LLC = LOGICAL LINK CONTROL
 MAC = MEDIA ACCESS CONTROL
 MDI = MEDIUM DEPENDENT INTERFACE
 PCS = PHYSICAL CODING SUBLAYER

PHY = PHYSICAL LAYER DEVICE
 PMA = PHYSICAL MEDIUM ATTACHMENT
 PMD = PHYSICAL MEDIUM DEPENDENT
 RS-FEC = REED-SOLOMON FORWARD ERROR CORRECTION
 SR = PMD FOR MULTIMODE FIBER

Editor's note: The RS-FEC layer may be merged into 400GBASE-R PCS layer, depending on the choice of architecture by the Task Force.

Block diagram for 400GBASE-SR16 transmit/receive path



PMD:IS_UNITDATA_0.request
to PMD:IS_UNITDATA_15.request

PMD:IS_UNITDATA_0.indication
to PMD:IS_UNITDATA_15.indication

PMD Optical specifications

- Transmitter characteristics (each lane) at TP2 follow 100GBASE-SR4, Clause 95, Table 95-6.
- Receiver characteristics (each lane) at TP3 follow 100GBASE-SR4, Clause 95, Table 95-7.
- Illustrative link power budget follows 100GBASE-SR4, Clause 95, Table 95-8.
 - Current status of these tables shown on next 3 slides

Transmitter characteristics (each lane) at TP2:
follow 100GBASE-SR4, Clause 95, Table 95-6 (D3.2 illustrated below)

Description	Value	Unit
Signaling rate, each lane (range)	25.78125 ± 100 ppm	GBd
Center wavelength (range)	840 to 860	nm
RMS spectral width ^a (max)	0.6	nm
Average launch power, each lane (max)	2.4	dBm
Average launch power, each lane (min)	-9	dBm
Optical Modulation Amplitude (OMA), each lane (max)	3	dBm
Optical Modulation Amplitude (OMA), each lane (min) ^b	-7	dBm
Launch power in OMA minus TDEC (min)	-7.9	dBm
Transmitter and dispersion eye closure (TDEC), each lane (max)	4.9	dB
Average launch power of OFF transmitter, each lane (max)	-30	dBm
Extinction ratio (min)	2	dB
Optical return loss tolerance (max)	12	dB
Encircled flux ^c	≥ 86% at 19 μm ≤ 30% at 4.5 μm	
Transmitter eye mask definition {X1, X2, X3, Y1, Y2, Y3} Hit ratio 1.5 × 10 ⁻³ hits per sample	{0.3, 0.38, 0.45, 0.35, 0.41, 0.5}	

^aRMS spectral width is the standard deviation of the spectrum.

^bEven if the TDEC < 0.9 dB, the OMA (min) must exceed this value.

^cIf measured into type A1a.2 or type A1a.3 50 μm fiber in accordance with IEC 61280-1-4.

Receiver characteristics (each lane) at TP3: follow 100GBASE-SR4, Clause 95, Table 95-7 (D3.2 illustrated below)

Description	Value	Unit
Signaling rate, each lane (range)	25.78125 ± 100 ppm	GBd
Center wavelength (range)	840 to 860	nm
Damage threshold ^a (min)	3.4	dBm
Average receive power, each lane (max)	2.4	dBm
Average receive power, each lane ^b (min)	-10.9	dBm
Receive power, each lane (OMA) (max)	3	dBm
Receiver reflectance (max)	-12	dB
Stressed receiver sensitivity (OMA), each lane ^c (max)	-5.6	dBm
Conditions of stressed receiver sensitivity test: ^d		
Stressed eye closure (SEC), lane under test	4.9	dB
Stressed eye J2 Jitter, lane under test	0.39	UI
Stressed eye J4 Jitter, lane under test	0.53	UI
OMA of each aggressor lane	3	dBm
Stressed receiver eye mask definition {X1, X2, X3, Y1, Y2, Y3} Hit ratio 5×10^{-5} hits per sample	{0.28, 0.5, 0.5, 0.33, 0.33, 0.4}	

^aThe receiver shall be able to tolerate, without damage, continuous exposure to an optical input signal having this average power level on one lane. The receiver does not have to operate correctly at this input power.

^bAverage receive power, each lane (min) is informative and not the principal indicator of signal strength. A received power below this value cannot be compliant; however, a value above this does not ensure compliance.

^cMeasured with conformance test signal at TP3 (see 95.8.8) for the BER specified in 95.1.1.

^dThese test conditions are for measuring stressed receiver sensitivity. They are not characteristics of the receiver.

Illustrative link power budget:
 follow 100GBASE-SR4, Clause 95, Table 8 (D3.2 illustrated below)

Parameter	OM3	OM4	Unit
Effective modal bandwidth at 850 nm ^a	2000	4700	MHz.km
Power budget (for max TDEC)	8.2		dB
Operating distance	0.5 to 70	0.5 to 100	m
Channel insertion loss ^b	1.8	1.9	dB
Allocation for penalties ^c (for max TDEC)	6.3		dB
Additional insertion loss allowed	0.1	0	dB

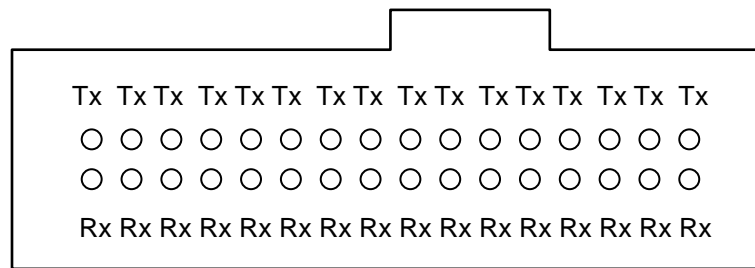
^aper IEC 60793-2-10.

^bThe channel insertion loss is calculated using the maximum distance specified in Table 95–5 and cabled optical fiber attenuation of 3.5 dB/km at 850 nm plus an allocation for connection and splice loss given in 95.11.2.1.

^cLink penalties are used for link budget calculations. They are not requirements and are not meant to be tested.

Medium Dependent Interface (MDI) for 400GBASE-SR16 and lane assignments

- Similar to the MDI defined for 100GBASE-SR10 (Clause 86.10.3.3, Recommended Option A), but using MPO-16, a 16-wide version of the 100G-SR10 MDI.
- Transmitters occupy the top row and receivers occupy the bottom row for better heat dissipation



400GBASE-SR16 optical lane assignments for the MDI receptacle when viewed looking into the receptacle with keyway feature on top.

- The following 4 slides show draft text and figures which describe the MDI and lane assignments, using clause 86 and 95 content as basis, with modifications for 400GBASE-SR16 and MPO-16

MDI (1 of 4)

xx.m.n Medium Dependent Interface (MDI)

The 400GBASE-SR16 PMD is coupled to the fiber optic cabling at the MDI. The MDI is the interface between the PMD and the “fiber optic cabling” (as shown in Figure xx-a). The 400GBASE-SR16 PMD is coupled to the fiber optic cabling through one connector plug into the MDI optical receptacle as shown in Figure xx-b. Example constructions of the MDI include the following:

- a) PMD with a connectorized fiber pigtail plugged into an adapter;
- b) PMD with receptacle.

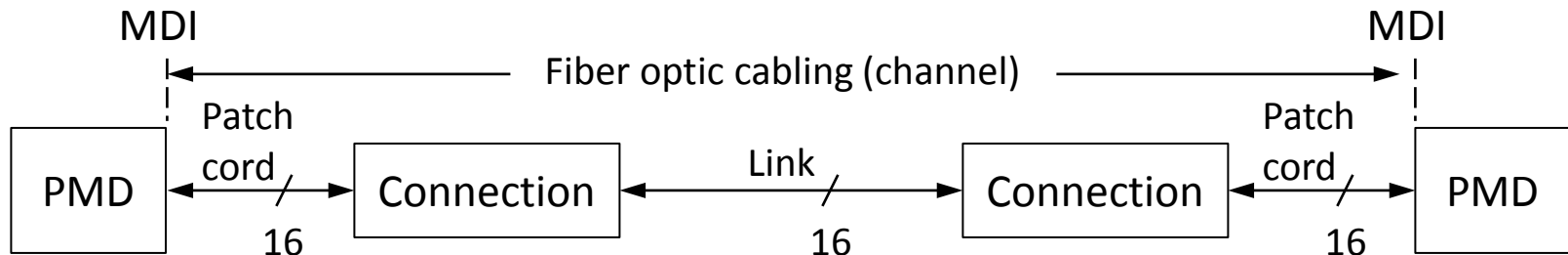


Figure xx-a – Fiber optic cabling model

Editor's note: Figure xx-a may be placed in a preceding subclause

MDI (2 of 4)

xx.m.n.1 Optical lane assignments

The sixteen transmit and sixteen receive optical lanes of 400GBASE-SR16 shall occupy the positions depicted in Figure xx-b viewed looking into the MDI receptacle with the connector keyway feature on top. The interface contains 32 active lanes within 32 total positions. The transmit optical lanes occupy the top row. The receive optical lanes occupy the bottom row. See clause xx.m.n.2 for MDI optical connector requirements.

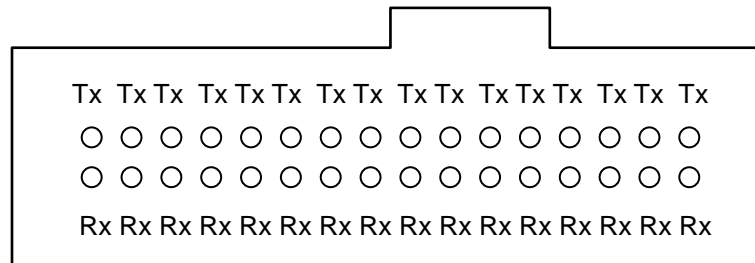


Figure xx-b -- 400GBASE-SR16 optical lane assignments viewed looking into the MDI receptacle with keyway feature on top.

MDI (3 of 4)

xx.m.n.2 Medium Dependent Interface (MDI) requirements

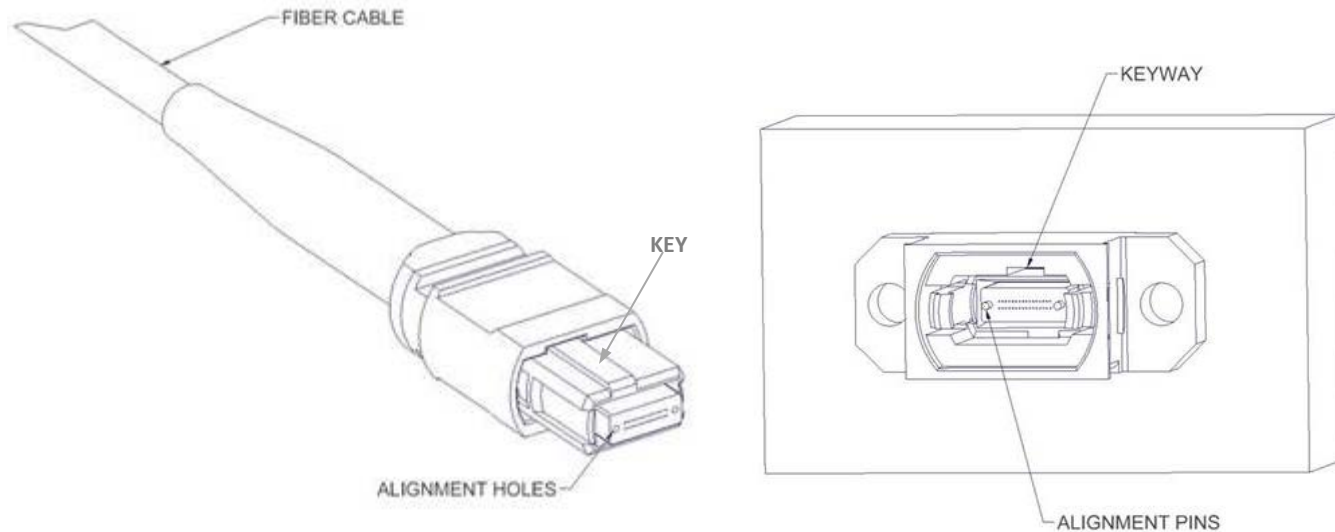
The MDI adapter or receptacle shall meet the dimensional specifications of ANSI/TIA-604-18 adapter designation FOCIS 18A-k-0. The plug terminating the optical fiber cabling shall meet the dimensional specifications of ANSI/TIA-604-18 female plug connector flat interface designation FOCIS 18P-2x16-1-0-2-2. The MDI shall optically mate with the plug on the optical fiber cabling. Figure xx-c shows an MPO-16 female plug connector with flat interface, and an MDI.

The MDI connection shall meet the interface performance specifications of IEC 61753-1 and IEC 61753-022-2.

NOTE— Transmitter compliance testing is performed at TP2 as defined in xx.k.j, not at the MDI.

*Editor's note: ANSI/TIA-604-18 presently entering third ballot.
IEC has not yet initiated ballot on the equivalent connector.*

MDI (4 of 4)



MPO-16 female plug connector flat interface

MDI

Figure xx-c – MPO-16 female plug connector flat interface and MDI

Editor's note: Figure is in public domain so may be used "as is". It is also acceptable redrawn in a form like Figure 86-8 with keying adjustment.

Further work

- The PMD target BER is likely to deviate from 5×10^{-5} , so some fine tuning of parameters may be required.
 - The project's choice of FEC will determine the pre-FEC BER target, and may also affect the exact signaling rate.
- Confirm skew budget