VITESSE®

Making Next-Generation Networks a Reality.

Proposal for a Unifying MLD

Mike Scholten
Frank Chang
10-Mar-2008

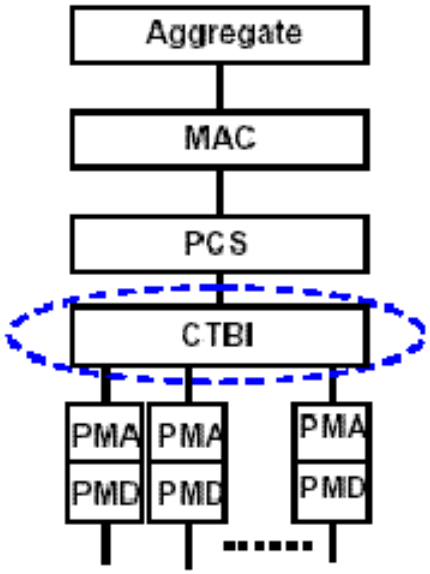
Presentation Outline

- 40GE / 100GE Applications using 10G Serial Lane Interface:
  - 40G/100G MAC-PCS-to-PMA/PMD Interface
  - 40G/100G across 10G backplane
  - 40G/100G extension across WAN using G.709 OTN
- Background:
  - CTBI/MLD, APL, PBL
  - Other 10G Serial Interfaces (XFI, SFI, OIF CEI, 10GBASE-KR)
- MLD Issues
  - 10GBASE-R based PCS
  - “Unfriendly” to 10GBASE-KR backplane and OTN support
- Proposal for a unified 10G Serial Interface for 40/100GE Applications
  - Goal: Simplify interfacing 40GE/100GE MACs to backplane and OTN devices

40 / 100GE Applications for 10G Serial Interface

- MLD and PBL propose 40GE & 100GE physical interfaces between MAC/PCS and PMA/PMD based on 64b/66b-encoded, 10.3125Gb/s serial streams.
- 10GBASE-KR is gaining popularity as a 10.3125Gb/s serial backplane interface for 10GE, and could potentially be applied to connect 40GE and 100GE across 10Gb/s serial backplanes.
- 10GBASE-KR transcodes 64b/66b into 64b/65b, reusing freed bits to provide a 32-bit FEC code to improve BER across 10G serial backplane.
- IEEE 802.3ba PAR indicates intention to support 40GE and 100GE transport over WAN using OTN; MLD could potentially be applied to connect 40GE/100GE MACs to 40G/100G OTN/FEC mappers.

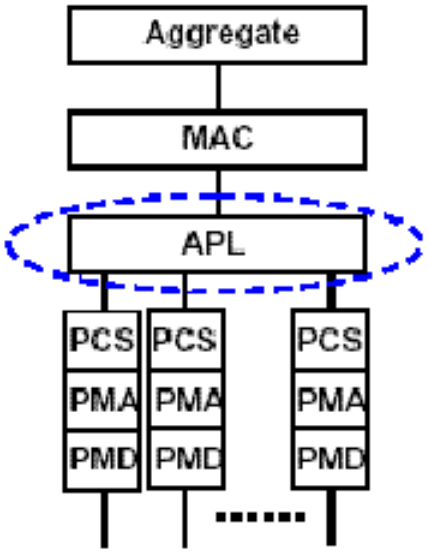
Background: Baseline 40/100GE Architecture



CTBI

- word (8 byte) striping below PCS
- one MAC and one PCS per interface
- can be extended using virtual lanes, to support simple optical modules at all rates and reaches

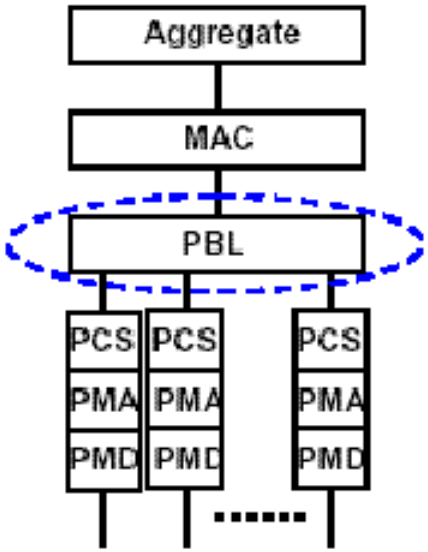
gustlin_01_0907.pdf



APL

- frame striping above PCS
- frames decimated into variable length fragments (SAR), tagged and distributed across lanes
- one MAC, but multiple PCSs (one per PMD lane)
- potential to reuse existing 10G PHY devices

frazier_01_1107.pdf

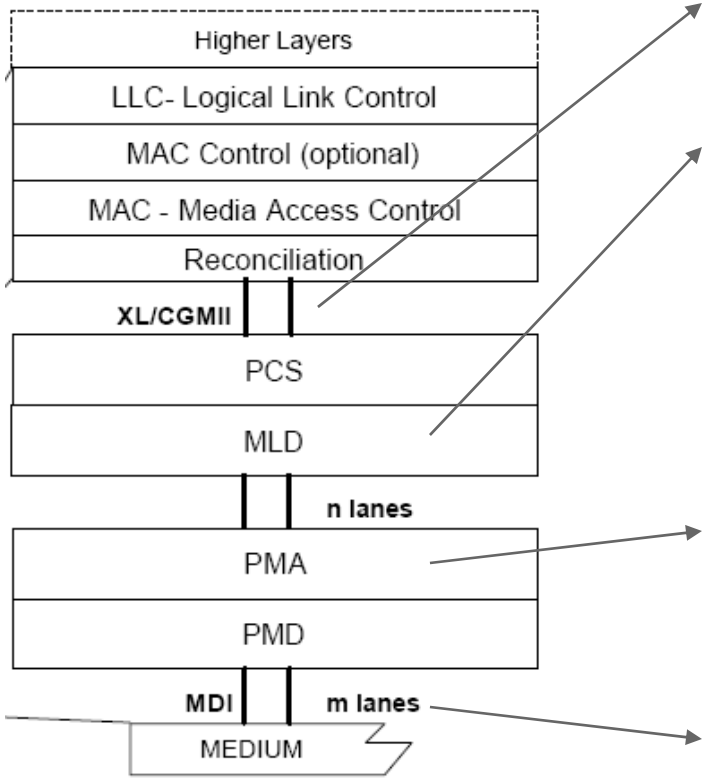


PBL

- word (8 byte) striping above PCS
- one MAC, but multiple PCSs (one per PMD lane)
- potential to reuse existing 10G PHY devices

Q: how to converge MLD and PBL?

Background: CTBI / MLD Approach



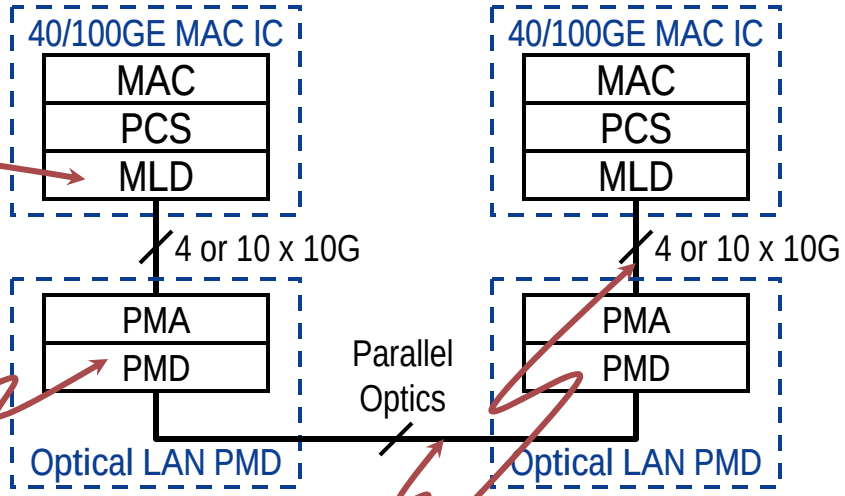
- XL/CGMII – 40G/100G Media Independent Interface
  - Logical interface to the MAC
- MLD – Multi-Lane Distribution (aka CTBI)
  - 4/10 Lane PCS to PMA/PMD Interface
  - Bit interleaves virtual lanes from PCS into MLD physical lanes
  - Deskews lanes in the receive direction
  - Each lane runs at 10.3125G
  - 40G/100G data is inverse muxed across the MLD lanes
- PMA – Physical Media Adaptor
  - the so-called “Gearbox”
  - Bit interleaves n MLD lanes into m MDI lanes
- MDI – Media Dependent Interface
  - XLGI: 4 x 10.3125G
  - CGI: 10 x 10.3125G, or 4 x 25G lanes

Background: CTBI / MLD Approach

- MLD uses virtual lanes with bit interleaving to simplify PMA/PMD
  - PCS & MLD layer not required in PMA/PMD layer
  - Lane alignment and deskew is only required & performed at MLD RX

MLD
- Remove / re-insert some IPG (make room for lane alignment codewords)
- Distribute / reassemble Virtual Lanes (VLs) (simple bit interleaving)
- Align & deskew VLs

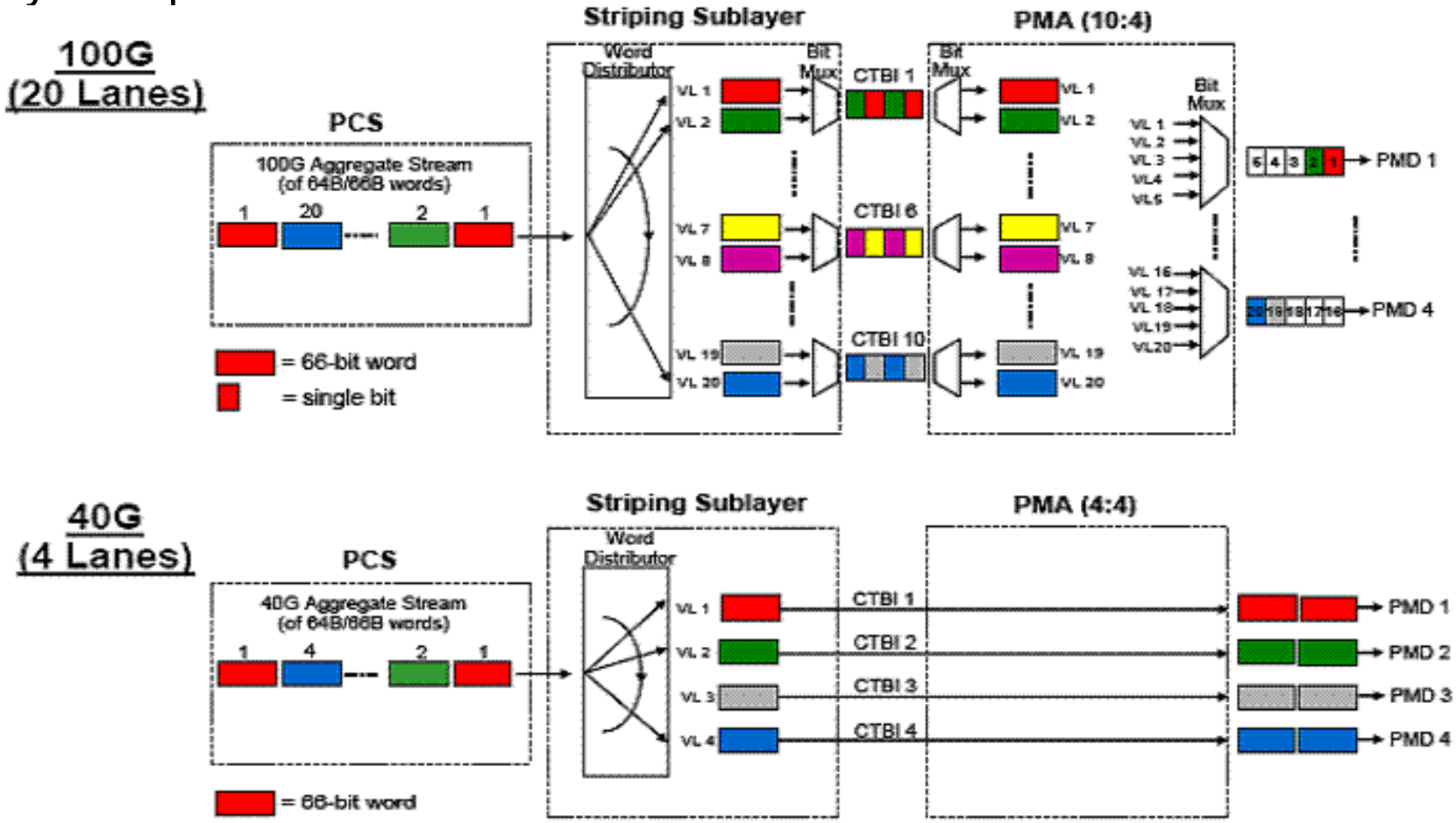
PMA/PMD
- 100GE LAN: 10:10 or 10:4 bit interleave / dis-interleave (10 x 10G or 4 x 25G parallel optics)
- 40GE LAN: 4:4 bit interleave / dis-interleave (4 x 10G parallel optics)
- 40G/100G WAN: Likely to use OTN plus serial optics



Skew
- Introduced by independent serial lanes (electrical and optical)
- Skew is only compensated at MLD RX

Background: MLD PCS Lane Bonding

Key concept - "Virtual lanes"

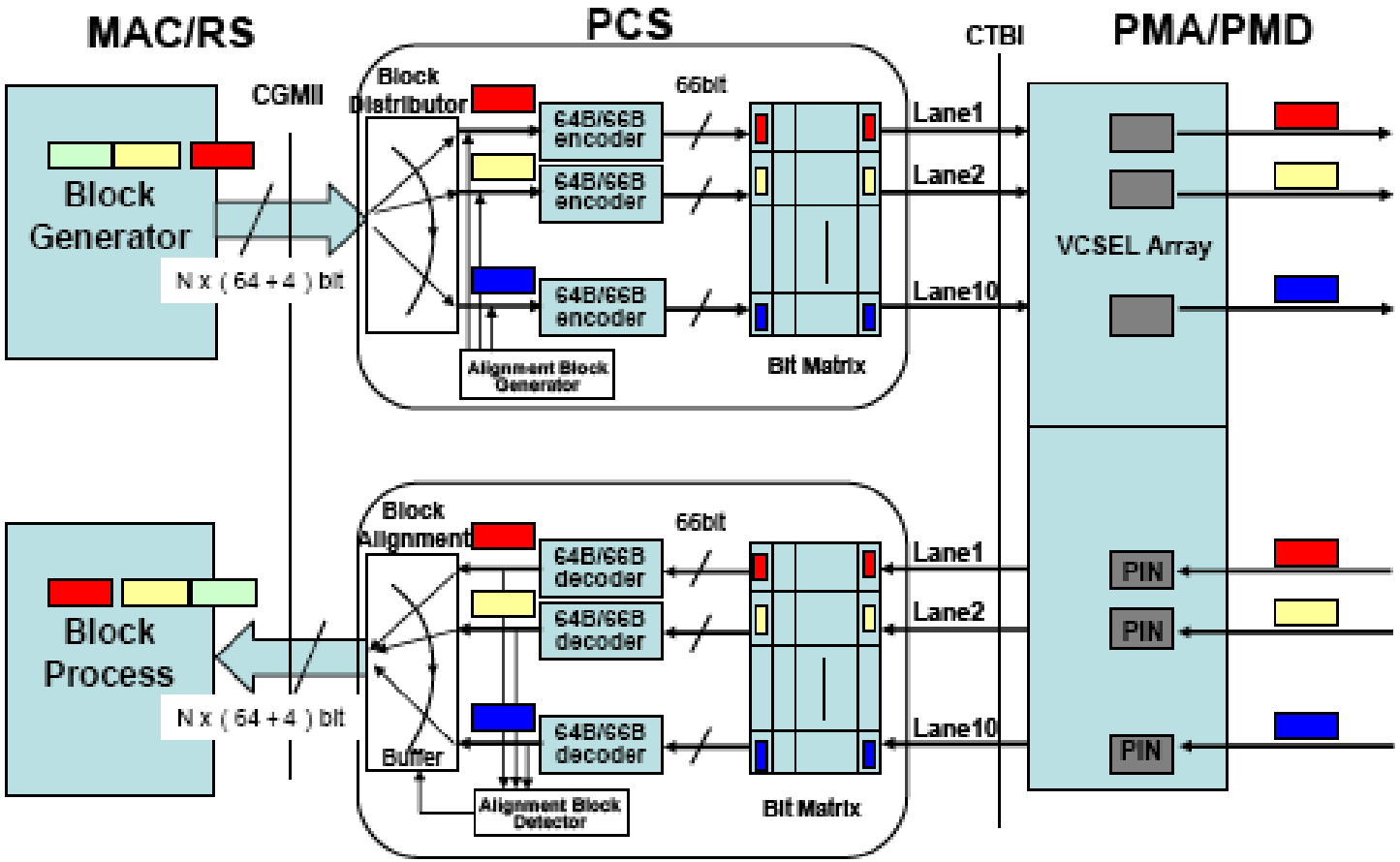


Key implication for implementation: [nicholl_01_1107.pdf]

MLD: Simplify host ASIC interface; PBL/APL: Re-use the existing PHY

Background: PBL PCS Lane Bonding

Key concept - “Bit Matrix” for Block integrity



Key implication for implementation: malpass_01_0108.pdf

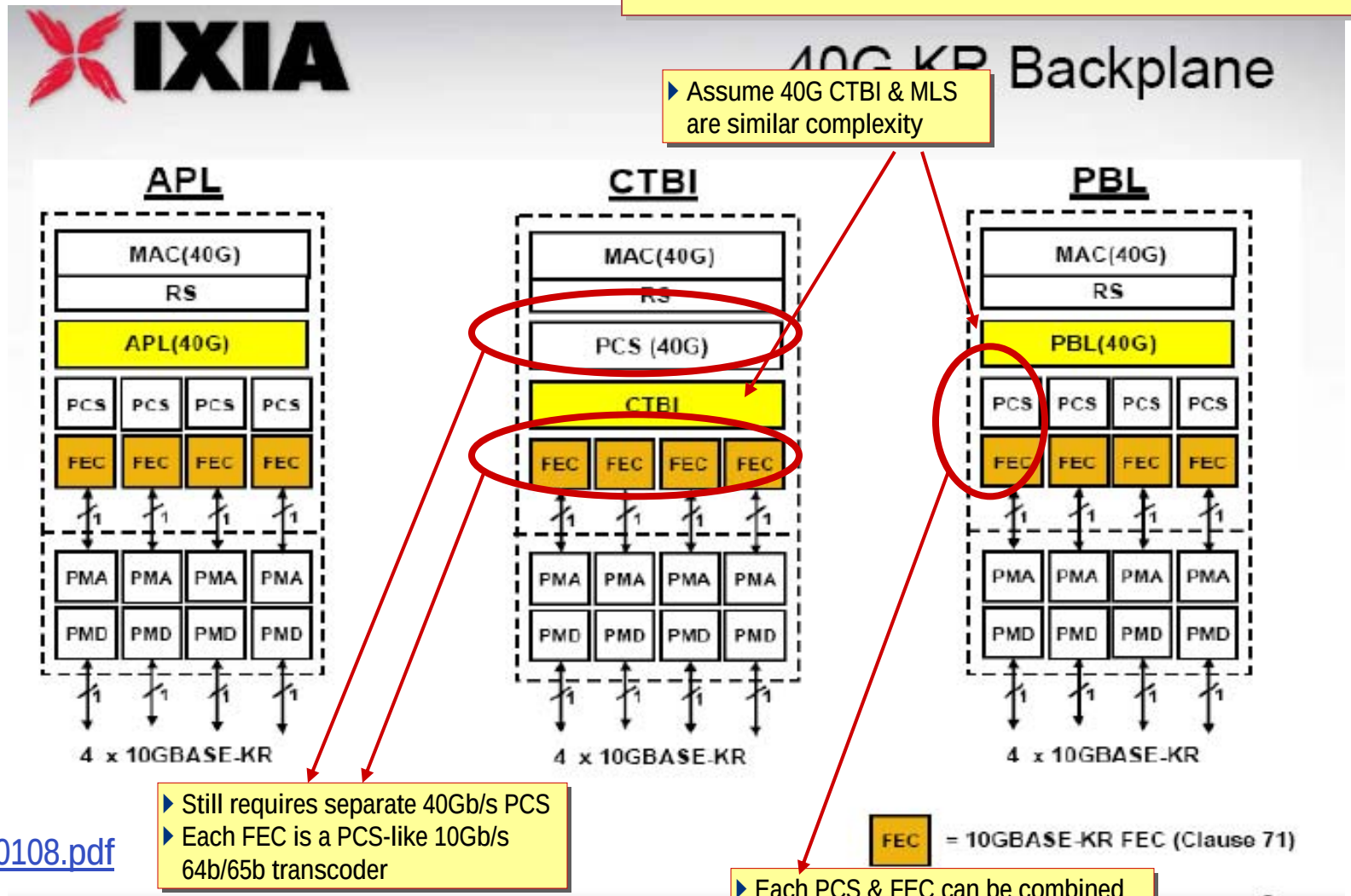
MLD: Simplify host ASIC interface; PBL/APL: Re-use the existing PHY

MLD Issues: 40/100GE Across 10G Serial Backplane (1 of 2)

- Issues for 40GE/100GE across 10G serial backplane
  - 10GBASE-KR recodes 64b/66b blocks into 64b/65b blocks, groups 32 blocks to create a "frame," then computes and appends 32-bit FEC parity check bits, resulting in same 10.3125Gb/s bit rate.
  - Requires visibility to 64b/66b blocks.
  - Could be applied to virtual lanes, but requires 66-bit frame detectors, 64b/65b transcoders, and FEC generators for each virtual lane (I.e. up to 20 of each in 100GE case with 20 VLs), and requires 10GBASE-KR receivers to recognize bit-interleaved KR blocks (since for 100GE, 2 virtual lanes per 10G serial lane).
  - Could be applied to aggregate 40/100GE signal, but requires implementing MLD deskew and reassembly function, then redistribute to 10 x 10GBASE-KR transcoder/FEC processors.
  - 40GE or 100GE transported across 4x10G or 10x10G serial lanes also requires lane alignment and deskew capability

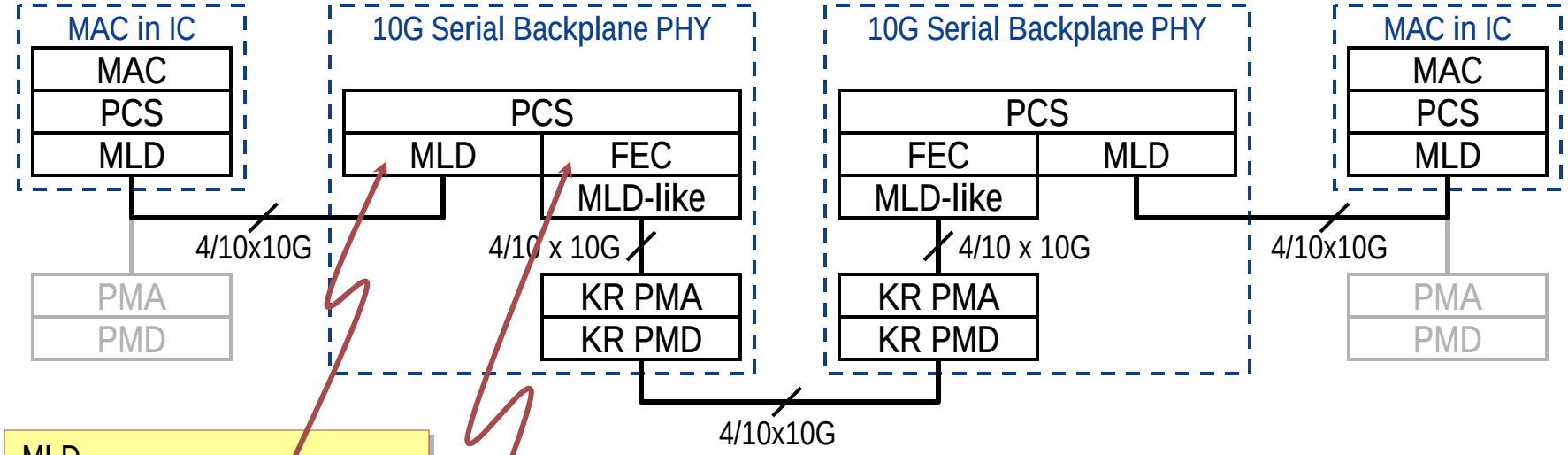
Ixia View: Include KR FEC in MAC

- Also consider 100G over 10G backplane (see next slide)



green_01_0108.pdf

MLD Issues: 40/100GE Across 10G Serial Backplane (2 of 2)



MLD
- Remove / re-insert some IPG
- VL Lane align / deskew
- Distribute / Reassemble VLs

FEC + MLD-like
- Transcode 64b/66b ↔ 64b/65b
- Compute / append FEC
- Lane align / deskew
- Distribute / Reassemble lanes

Issues for 40GE/100GE across 10G serial backplane

- KR transcoder requires visibility to 64b/66b blocks
- If FEC applied to 64b/66b-encoded virtual lanes:
  - Requires 66-bit aligner, 64b/65b transcoder, and FEC generator for each virtual lane (i.e. up to 20 of each in 100GE case with 20 VLs)
  - Requires 10GBASE-KR receivers to recognize bit-interleaved KR blocks (since for 100GE, 2 virtual lanes per 10G serial lane)
- If applied to aggregate 40/100GE signal:
  - Requires implementing MLD deskew and reassembly function, then redistribute to 10 x 10GBASE-KR transcoder/FEC processors.
- 40GE or 100GE transported across 4x10G or 10x10G serial lanes also requires alignment and deskew of 4 or 10 x 10G lanes sent across backplane
  - Skew budget may be higher than MLD supports

Electrical Interface Consideration

Conclusion

- 40G Optical Module Electrical Interface
  - Likely 1st generation based on 4 x 10.3125Gbps.
  - Expect possible 16 x 3.125Gbps or 8 x 6.25Gbps.
- 100G Optical Module Electrical Interface
  - Likely 1st generation based on 10 x 10.3125Gbps.
  - Likely 2st generation based on 4 x 25Gbps.
  - Likely someone will present a concept to negotiate between 4 x 10.3125Gbps and 4 x 25Gbps.
- Anticipate 25G proposals for >10m Cu Cabling
  - Anticipate 25G Backplane PHY in future.
- Consider looking at SR and LR channels separately for 25Gbps
  - LR channels may need more time to develop a solution.
  - LR channels may overburden SR applications so compatibility should not be a requirement.

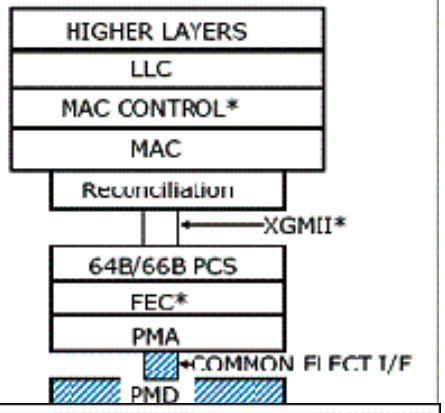
10

IEEE P802.3ba Task Force Jan 2008 Interim Meeting, Portland OR

LUXTERA
NANOPHOTONIC INTEGRATED CIRCUITS

Common Electrical I/F discussion

- CEI exists between PMA and PMD layers. (PMAI?)
- Needs to supports multiple PMD layers
  - Optical
  - Copper



Thoughts …

- 10G signaling today makes sense, but 25G signaling will be needed.
  - There is a future need for 4by25Gbps optics modules.
  - IEEE 802.3 needs to focus on 40Gbps and 100Gbps.
- Efforts such as OIF CEI-25 can be used to define SR and LR applications for some interfaces.
  - Allows IEEE 802.3 to refrain from some physical interface definitions.
  - Allows segments of the interface work to be done in parallel, saving time.
- IEEE P802.3ba should establish a closer rapport / liaison with OIF in support of the CEI-25 interface.

11

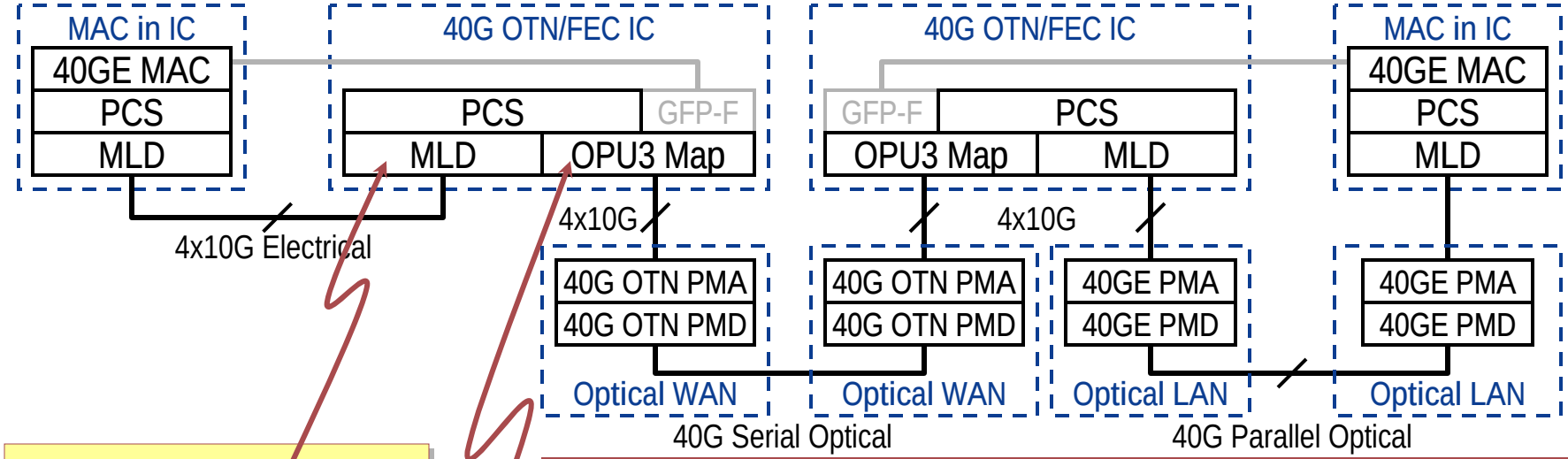
IEEE P802.3ba Task Force Jan 2008 Interim Meeting Portland OR

Virtual lanes also add value for future electrical backplane and cable applications

MLD Issues: 40GE WAN Transport (over OTN) (1 of 2)

- Issues for 40GE/100GE over OTN
  - ITU-T desires to fit 40GE into 40.150Gb/s OPU3 payload capacity of OTU3
  - 66/64 x 40Gb/s = 41.25Gb/s which exceeds OPU3 payload capacity
  - ALU Proposal: Transcode 64b/66b blocks into 512b/513b reducing BW to 40.08Gb/s, which fits, but requires visibility to 64b/66b blocks
    - Requires MLD encoder/decoder with each OTN/FEC transponder
  - Huawei Proposal: Define higher speed "stretched" OTU3s which can accommodate 64b/66b-encoded bandwidth in lanes sized for MLD lanes (or 10GE LAN)
    - Proposes to transport MLD lanes without MLD decode, then re-encode, requiring far-end MLD to support 2 x 40GE LAN skew
    - OTN prevents or compensates any skew introduced in optical transport across WAN

MLD Issues: 40GE WAN Transport (over OTN) (2 of 2)



MLD
- Remove / re-insert some IPG
- VL Lane align / deskew
- Distribute / Reassemble VLs

OPU3 Map + MLD-like
- Transcode 64b/66b ↔ 512b/513b
- Map into OPU3 payload of OTU3
- Compute / append OTN/FEC
- 4x10G OTN lane align / deskew
- Distrib / Reassem 4x10G lanes (assumes 4-lane PHY interface)

Issues for 40GE/100GE WAN Transport:

- OTN/FEC transcoder requires visibility to 64b/66b blocks if 512b/513b transcoding used to fit into OPU3 payload
  - Deskew / reassembly required if recoding applied to 40G serial (shown above);
  - 4 transcoders required if recoding applied per lane without MLD deskew / reassembly
- 4 x 10.3G lane-mapping into 4 x 10.3G timeslots within stretched "OPU3s" doesn't require OTN/FEC to apply MLD and transcode 64b/66b to 512b/513b, but does require far-end MLD to support deskew for 2 x 40GE LAN links
- MAC Frames can still be GFP-F-mapped into OPU3/4 payload
- *NOTE*: 512b/513b transcoding won't fit 10GE LAN into OTU2; Similarly, won't fit 4 x 10GE LAN into 4 x ODU2 into OPU3. These could be fit into "OPU3s".

Goals of a Unified 10G Serial Solution

- GOAL:
  - Unify 10G serial interface to support MAC/PCS ↔ PMA/PMD, 10G serial backplane, and MAC/PCS ↔ OTN/FEC interconnect using a common 10G serial lane format.
  - Independence between convenient 10G serial electrical lane MAC/PCS interface and future higher-speed backplane and optical module PMD interfaces
    - Justification for MLD virtual lane approach and PBL bit matrix approach.
  - Single-hop and multi-hop skew compensation, such that skew need only be compensated when connecting to a MAC
    - Backplane and cable applications introduce their own skew, which can potentially be compensated utilizing the same mechanism adopted for MLD/PBL, but which may require larger deskew buffers. If this can be expected to affect cost, "single-hop" and "multi-hop" flavors of MLD/PBL can be defined (in much the same way that short-reach and long-reach optics have been agreed).
  - Optional FEC
    - Backplane & 10G electrical cable applications benefit from FEC
    - Applications not requiring FEC insert known, non-FEC values in this field to assist in frame alignment without requiring FEC encoding/decoding to be implemented

Proposal for a Unified 10G Serial Solution

- POSSIBLE APPROACH:
  - Adapt 10GBASE-KR concept (64b/66b ↔ 64b/65b transcoding) to free up bits which can be used for lane alignment and optional backplane FEC
    - Option 1: Transcode 64b/66b to 64b/65b, then group 64 blocks, freeing up 64 bits which can be allocated for Frame/Lane Alignment and optional FEC
      - Reuse 64b/65b transcoding, but with weaker FEC (32 bit FEC over 64 blocks of 66 bits, instead of 32 blocks)
      - FEC decode latency 2-4 times current 10GBASE-KR (64 blocks instead of 32) per lane
    - Option 2: Transcode 2x64b/66b to 128b/129b, then group 32 blocks, freeing up 48 bits which can be allocated for Frame/Lane Alignment and optional FEC
      - New transcoding, but with existing 32-bit FEC over 32 blocks
      - FEC decode latency 1-2 times current 10GBASE-KR (2x only when VLs have been interleaved)
  - If present, FEC also protects Alignment bits and may be useful in short-reach optical apps
  - Requires PCS (or transcoding) with optional FEC generation per virtual lane
  - Faster lane alignment than current MLD due to small block size (<< 16384 66-bit blocks)
  - Encourage ITU to accept a "stretched" OTU3s format capable of supporting 10.3Gb/s timeslots which can accommodate either 10GE LAN signals or 10G serial lanes supporting 40GE & 100GE at 10.3125Gb/s rates without transcoding

Background: 10GBASE-KR Format

FEC block format

T_0	64 Bit Payload Word 0	T_1	64 Bit Payload Word 1	T_2	64 Bit Payload Word 2	T_3	64 Bit Payload Word 3
T_4	64 Bit Payload Word 4	T_5	64 Bit Payload Word 5	T_6	64 Bit Payload Word 6	T_7	64 Bit Payload Word 7
T_8	64 Bit Payload Word 8	T_9	64 Bit Payload Word 9	T_{10}	64 Bit Payload Word 10	T_{11}	64 Bit Payload Word 11
T_{12}	64 Bit Payload Word 12	T_{13}	64 Bit Payload Word 13	T_{14}	64 Bit Payload Word 14	T_{15}	64 Bit Payload Word 15
T_{16}	64 Bit Payload Word 16	T_{17}	64 Bit Payload Word 17	T_{18}	64 Bit Payload Word 18	T_{19}	64 Bit Payload Word 19
T_{20}	64 Bit Payload Word 20	T_{21}	64 Bit Payload Word 21	T_{22}	64 Bit Payload Word 22	T_{23}	64 Bit Payload Word 23
T_{24}	64 Bit Payload Word 24	T_{25}	64 Bit Payload Word 25	T_{26}	64 Bit Payload Word 26	T_{27}	64 Bit Payload Word 27
T_{28}	64 Bit Payload Word 28	T_{29}	64 Bit Payload Word 29	T_{30}	64 Bit Payload Word 30	T_{31}	64 Bit Payload Word 31
32 parity bits							

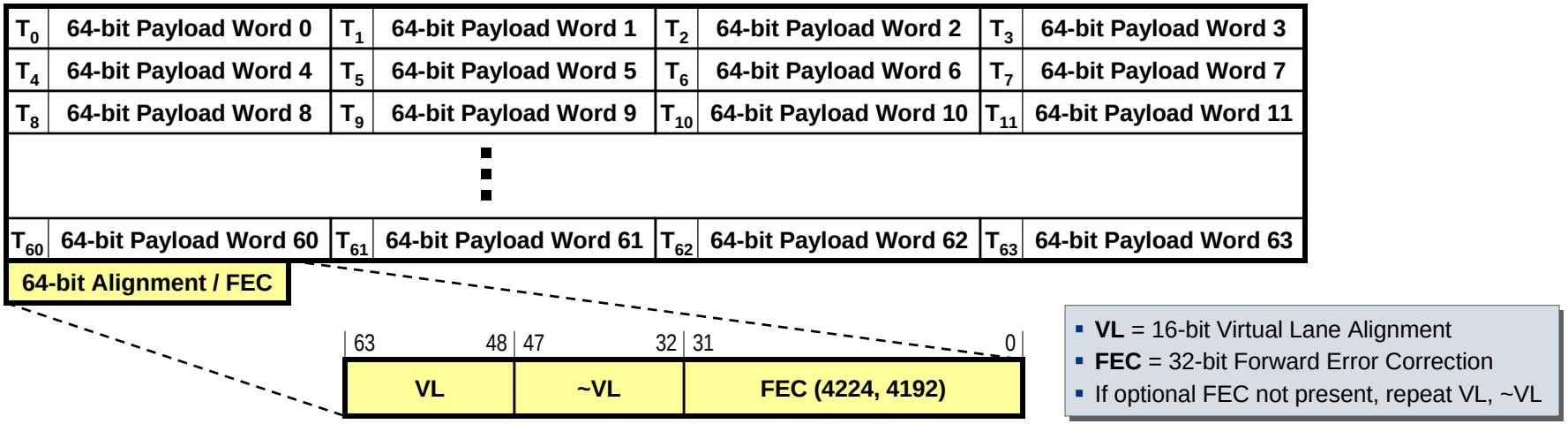
Total Block length = (32 x 65) + 32 = 2112 bits

- Payload words carry the 10GBASE-R scrambled payload words
- Tn = Transcode bit carries the state of the 10GBASE-R sync bits for the associated payload word
  - Sync bits are compressed to a single bit then scrambled to ensure DC balance
  - 64b/66b sync bits are either 10 or 01 hence can be reconstructed from the T bit
  - Synchronization is achieved at FEC block level
- Block has the same overhead as 64B/66B encoding

IEEE802 Plenary July 2006 10GBASE-KR FEC tutorial 13

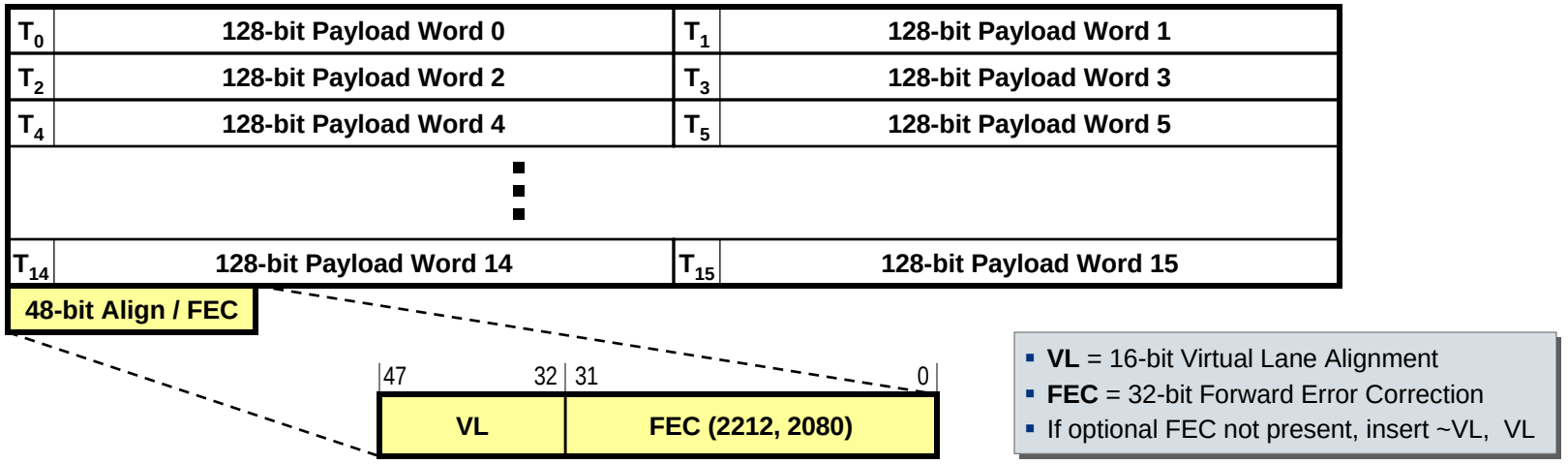
Proposed Unified 10G Serial Format (Option 1)

- Option 1: 64 x 64b/65b + 64-bits Alignment / FEC
  - Transcode 64b/66b to 64b/65b, then group 64 blocks per virtual lane, freeing up 64 bits which can be allocated for Frame/Lane Alignment and optional FEC
    - Applies same virtual lane and bit interleave approach suggested for MLD
    - Reuse 64b/65b transcoding, but with weaker FEC (32 bit FEC over 64 blocks of 66 bits, instead of 32 blocks)
    - Similar frame alignment as in 10GBASE-KR
    - FEC decode latency 2-4 x current 10GBASE-KR per lane (64 blocks instead of 32; 2 virtual lanes per physical lane for 100GE)



Proposed Unified 10G Serial Format (Option 2)

- Option 2: 16 x 128b/129b + 48-bits Alignment / FEC
  - Transcode 2 x 64b/66b into 128b/129b block, group 16 x 128b/129b blocks, freeing up 48 bits in 32 x 64b/66b blocks allocated as 16 bits for alignment + 32 bits for FEC
    - Applies same virtual lane and bit interleave approach suggested for MLD
    - New transcoding, but can reuse same 32-bit KR FEC over similar 32 blocks x 66 bits
    - Similar frame alignment as in 10GBASE-KR
    - FEC decode latency 1-2 x current 10GBASE-KR per lane (same 32 blocks, but 2 virtual lanes per physical lane for 100GE)

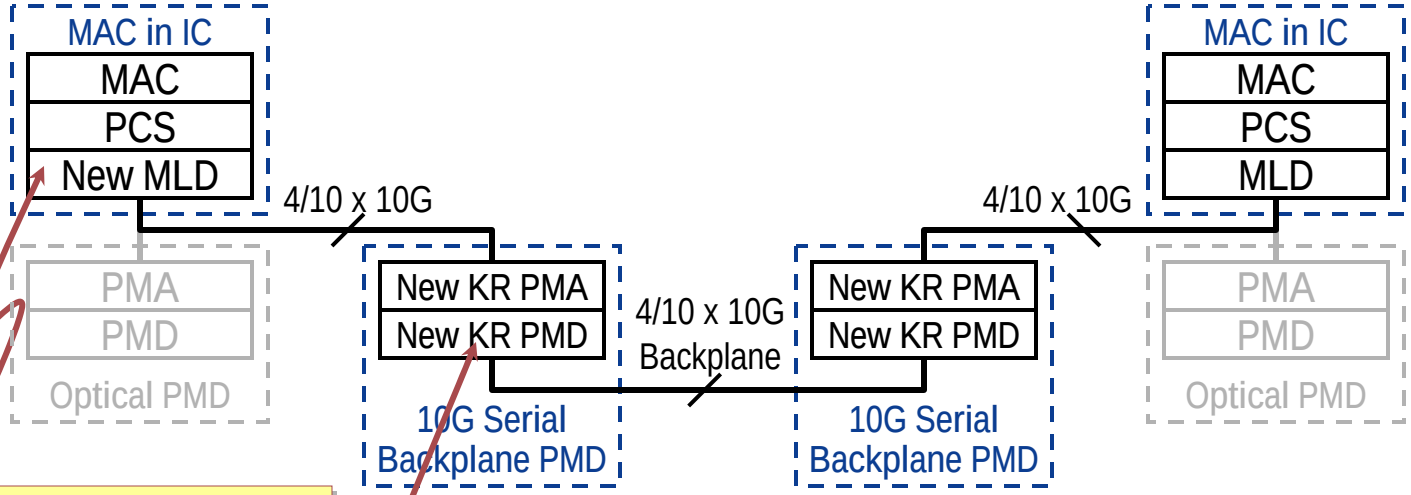


- **VL** = 16-bit Virtual Lane Alignment
- **FEC** = 32-bit Forward Error Correction
- If optional FEC not present, insert ~VL, VL

Benefits: Retain 10.3125Gb/s serial rates

- No Adjustments to IPG with no increase in lane rate
  - Both MLD and PBL delete IPG to accommodate inserted lane markers
  - IPG restoration not guaranteed to occur in the same locations
- Negative impact by the alignment block rate
  - Clocking rate complexity; possibly loss of SyncE synchronization.
  - Capability for adaptation to OTN
  - Affect the maximum skew that can be compensated
  - Less “bit-transparency” because of IPG removal and re-insertion.
  - Tougher link jitter budget

40/100GE Across Backplane with New, Improved MLD



MLD

- No adjustments to IPG
- Distribute / Reassemble VLs
- Virtual Lane align / deskew
- Transcode with optional FEC per lane (could eliminate Clause 49 PCS)

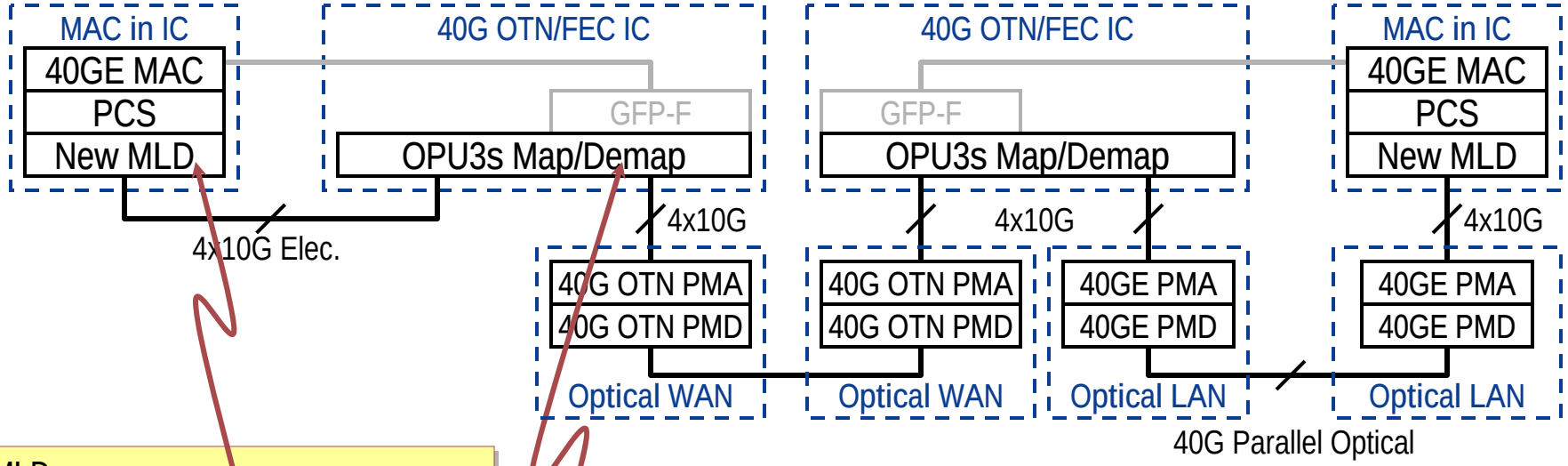
New KR PMA/PMD

- For 10G serial, 1:1 TX/RX
- For 25G serial, 10:4 bit interleave
- Transcoding, FEC in New MLD
- Apply TX pre-emphasis
- Apply RX equalization

Impact of New MLD

- No transcoder required in 10G Serial Backplane PHY
- FEC in New MLD can be applied to any PMA/PMD (optics, backplane, etc.)
- MLD lane alignment extended to 16-bit field and applied to deskew backplane, as well as LAN optics and O/E/O skew
- If FEC utilized:
  - FEC encoder/decoder required for each virtual lane
  - FEC decoder latency = FEC block size per virtual lane

40/100GE WAN Transport with New, Improved MLD



MLD

- No adjustments to IPG
- Distribute / Reassemble VLs
- Virtual Lane align / deskew
- Transcode with optional FEC per lane (or redefine PCS to include New MLD)

OPU3s Map + MLD-like

- Map into OPU3s payload of OTU3s
- Compute / append OTN/FEC
- 4x10G OTN lane align / deskew
- Distrib / Reassem 4x10G lanes (assumes 4-lane PHY interface)

Impact of New MLD & OPU3s Map/Demap:

- No transcoder required in 10G Serial Backplane PHY
- 4 x 10.3G lane-mapping into 4 x 10.3G timeslots within stretched "OPU3s" doesn't require New MLD in OTN/FEC
- 4 x 10.3G lanes can also be bit-interleaved into single 41.25Gb/s payload which is GMP-mapped into OPU3s
- New MLD required to support deskew for 2 x 40GE LAN links.
- 40GE GFP-F mapping into OPU3 or OPU3s still supported.

Summary

- Proposed MLD interface reuses and extends 10G serial technology
  - Reuses 10GBASE-R PCS 64b/66b-encoding
  - Adds valuable virtual lane concept for simple PCS-to-PMA adaptation
  - Requires deskew only at LAN termination
- Current MLD proposals present some difficulties
  - "Unfriendly" to 10GBASE-KR backplane applications due to bit-interleaved virtual lanes, limitation precluding application of MLD deskew to backplane transport
  - "Unfriendly" to OTN applications, again due to bit-interleaved virtual lanes and requirement to apply deskew at both ends of OTN transport
- Propose unifying 10GBASE-KR and MLD into common 10G serial format
  - Use transcoding to free up bits for lane deskew and optional FEC
  - Retain well-established 10.3125Gb/s serial rates
  - Define single-hop (short-reach) and multi-hop (long-reach) MLD
    - Concept similar to short-reach and long-reach flavors of optics
    - Common MLD format applies to all flavors
    - Multi-hop MLD requires larger deskew buffers