25GSG Architecture Ad Hoc Technical Feasibility Summary

IEEE 802.3 25 Gb/s Ethernet Study Group

Matt Brown, APM
Joel Goergen, Cisco
Vineet Salunke, Cisco
Mark Gustlin, Xilinx
Gary Nicholl, Cisco
David Ofelt, Juniper
Chris Diminico, MC Communication/Panduit
Rich Mellitz, Intel
Adee Ran, Intel
Dave Chalupsky, Intel
Jonathan King, Finisar

Supporters

- Chris Collins, APM
- Rick Rabinovich, Alcatel-Lucent
- Erdem Matoglu, Amphenol
- Adam Healey, Avago
- John Petrilla, Avago
- Jeff Slavick, Avago
- Scott Kipp, Brocade
- Dan Dove, DNS
- Andrew Zambell, FCI
- Vipul Bhatt, Inphi
- Sudeep Bhoja, Inphi
- Arash Farhood, Inphi
- Brad Booth, Microsoft
- Scott Sommers, Molex
- Keith Conroy, MultiPhy
- Neal Neslusan, MultiPhy
- Megha Shanbhag, TE Connectivity
- Nathan Tracy, TE Connectivity

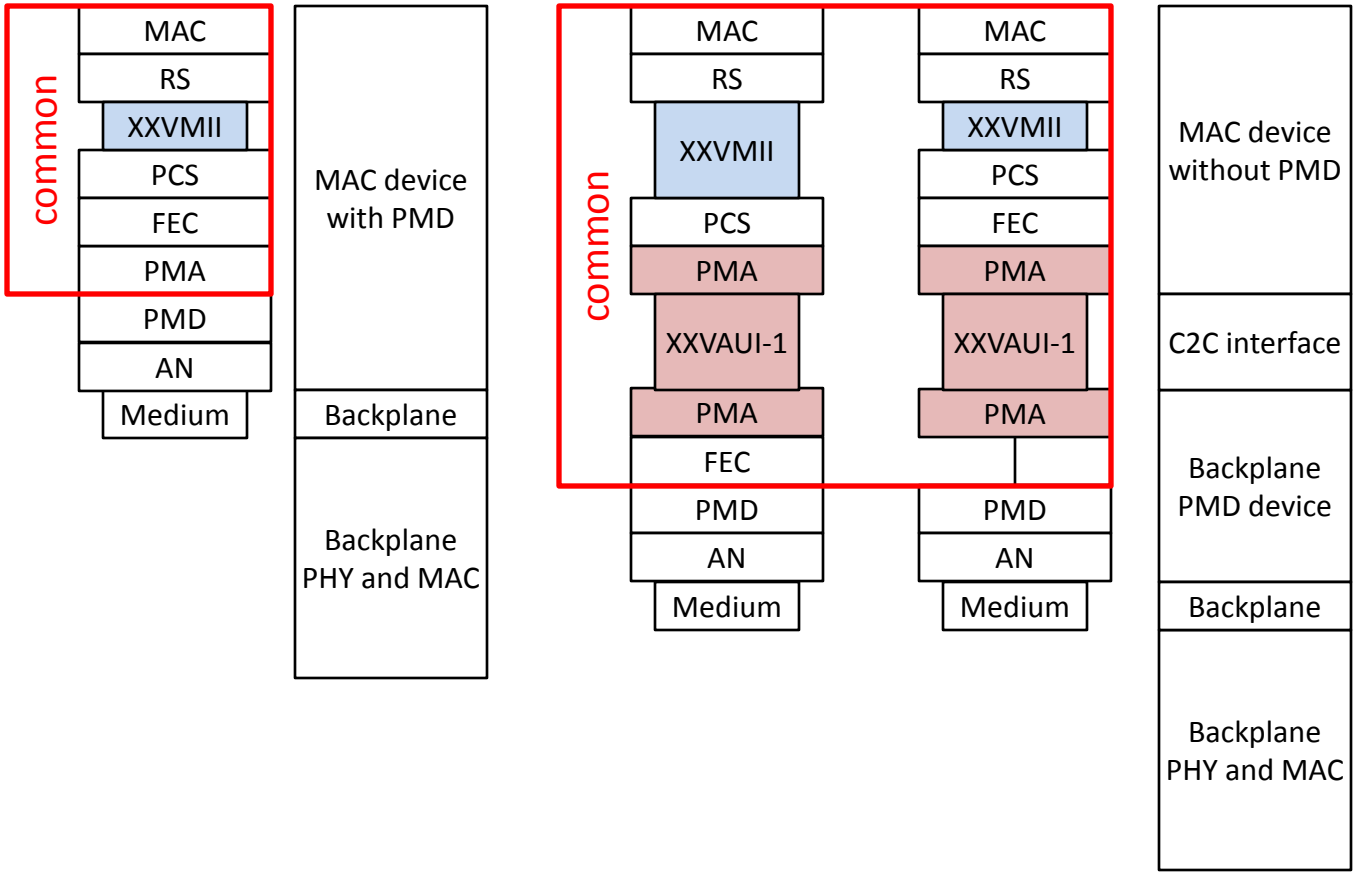
Introduction

- Demonstrate viable, example architectures and specifications based on current technologies that could be used as a basis to meet the objectives.
- Where they have been identified, multiple options are listed to be used as examples.
  - We can debate the merits of each in task force.
- Intent is to substantiate technical feasibility.

General Architecture

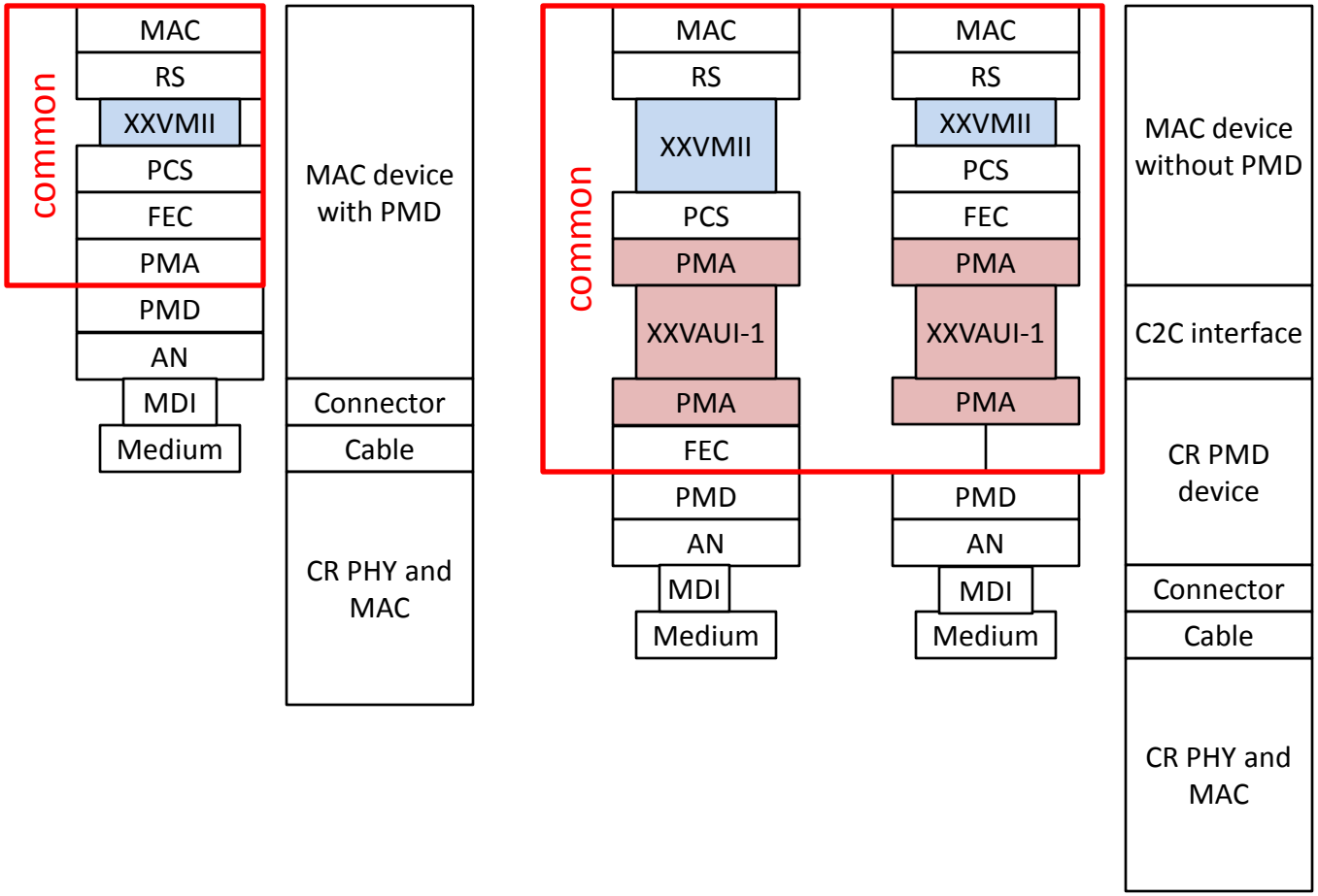
Examples

Backplane 25GBASE-KR



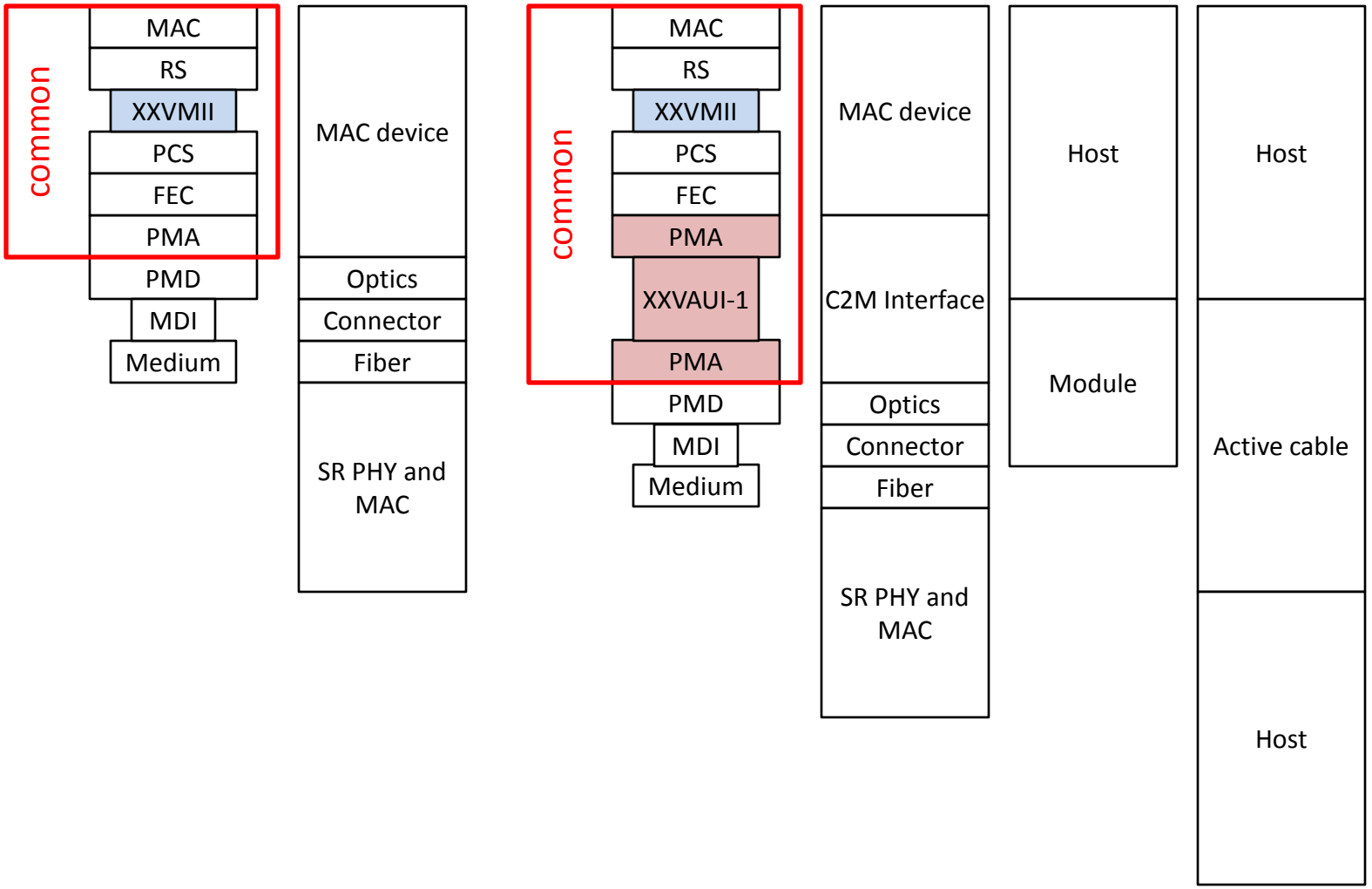
based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/brown_081214_25GE_adhoc.pdf

Cable 25GBASE-CR



based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/brown_081214_25GE_adhoc.pdf

Optical 25GBASE-SR



based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/brown_081214_25GE_adhoc.pdf

25 Gb/s specification sources

Sublayer/Function	Specification Sources
MAC	Clause 4
RS	Clause 46 (10G) or Clause 81 (40G/100G)
PCS	Clause 49 (10G) or Clause 82 (40G/100G)
FEC	Clause 74 (10G) and/or Clause 91 (100G)
PMA	Clause 52 (10G) or Clause 83 (40G/100G)
PMD	25GBASE-CR: Clause 92 Copper Cable 25GBASE-KR: Clause 93 Backplane PAM2 25GBASE-SR: Clause 95 Optical MMF and 32G Fibre Channel
XXVAUI Chip-to-Chip XXVAUI Chip-to-Module	Annex 83D Annex 83E with SFP28, QSFP28, or CFP4 connectors
AN	Clause 73
MDI	25GBASE-CR: SFP28, QSFP28, or CFP4 connectors 25GBASE-KR: No MDI required. 25GBASE-SR: LC and MPO connectors (same as for SFP+ and QSFP optical modules)
Medium	Same as PMD (alternatives for CR provided in later slides)
Management	Clause 45 MDIO Clause 30 Management Objects
EEE	Reuse EEE specifications in source clauses.

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/brown_081214_25GE_adhoc.pdf and http://www.ieee802.org/3/25GSG/public/adhoc/optical/king_090314_25GE_adhoc.pdf

RS/PCS/FEC Common Digital Specification examples

Assumptions

- Possible PMDs of interest are: CR, KR, SR
- Channel assumptions are similar if not identical to 100GBASE-CR4, KR4 and SR4
  - Loss budgets are the same as a single xR4 channel
  - Assume crosstalk is similar (multiple 25GbEs run next to each other)
- Assuming no KP channel needed?
  - But architecture should support it if needed
- Therefore a moderate strength FEC is required, assuming at this point that RS(528,514) is sufficient
  - If the assumptions change then this might change also
- For some applications a no-FEC configuration could be provided
- Goal is to maximize re-use from previous projects
  - Many devices will need to support 100GbE/40GbE/25GbE/10GbE on a given interface/port

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/gustlin_081214_25GE_adhoc.pdf

Option 1

1. 64b/66b only, leveraging 40/100GBASE-R but run at 25.78125G
   - But without Alignment Markers
   - 64b alignment for encoding (leveraging clause 82)
2. Use the 256B/257B transcoding as defined in 802.3bj
3. RS-FEC encoded data always
   - Just sync up FEC correctable match, with 256b/257b transcoding
   - Bit slips until n FEC correctable blocks are found, loses lock after m FEC blocks are uncorrectable
   - Similar to clause 74 KR FEC
   - Provide option for no-FEC configuration, if necessary.

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/gustlin_081214_25GE_adhoc.pdf

Option 2

1. 64b/66b only, leveraging 10GBASE-R but run at 25.78125G
   - No Alignment Markers
   - 32b alignment for encoding
2. Use the 256B/257B transcoding as defined in 802.3bj
3. RS-FEC encoded data always
   - Just sync up FEC correctable match, with 256b/257b transcoding
   - Bit slips until n FEC correctable blocks are found, loses lock after m FEC blocks are uncorrectable
   - Similar to clause 74 KR FEC
   - Provide option for no-FEC configuration, if necessary.

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/gustlin_081214_25GE_adhoc.pdf

Option 3

1. 64b/66b only, leveraging 40GBASE-R but run at 25.78125G
   - Single Alignment Marker (or single group of five)? Single PCS lane?
     - AMs provide means of detecting errors in no-FEC configurations
   - 64b alignment for encoding
2. Use the 256B/257B transcoding as defined in 802.3bj
   - No remapping of AMs needed though
3. RS-FEC encoded data always
   - With Alignment markers you can sync up the same as you do for 100G
   - Provide option for no-FEC configuration, if necessary.

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/gustlin_081214_25GE_adhoc.pdf

25GBASE-KR Backplane PHY

Example specifications

25GBASE-KR Summary

- Could use RS, PCS, and FEC specifications per common architecture slides.
- PMD and AN specified per 100GBASE-KR4 Clause 93 specifications adapted for a single lane.
- Specify EEE per the adapted specifications.

25GBASE-CR Copper Cable PHY

Example specifications

25G CR copper cables summary

- Three potential cable applications

  (1) Long cables
  - could use all electrical specifications for 100GBASE-CR4, could use 25G RS-FEC.
  - optimized for cable lengths up to 5m.

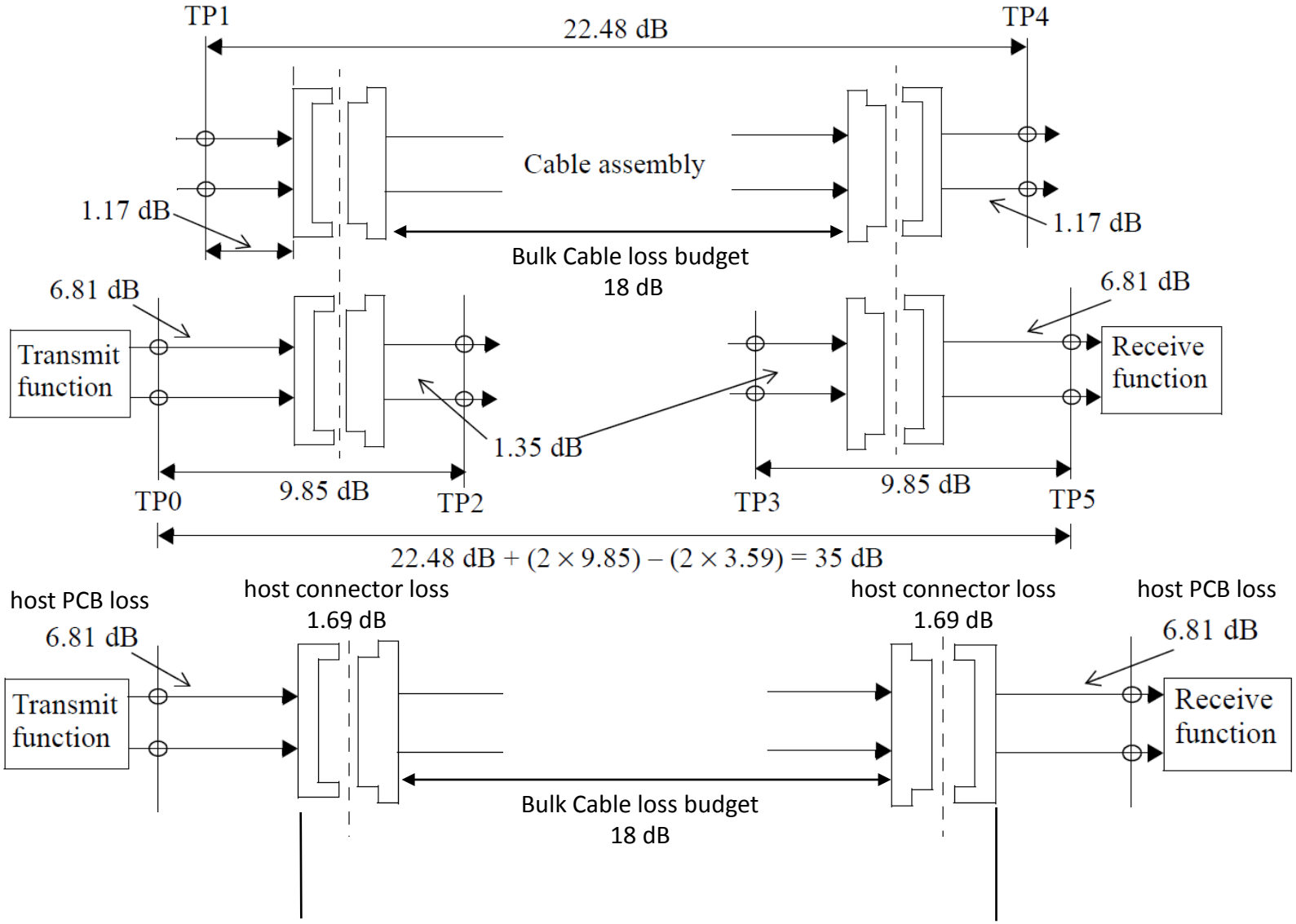
  (2) Short cables application #1
  - reduced cable loss, optimized for cable lengths up to 3m.
  - without FEC to reduce latency

  (3) Short cables application #2
  - reduced cable loss, optimized for cable lengths up to 3m.
  - with FEC, higher loss allocation for host boards

  - RS-FEC could be optional for copper cables (avoids latency of 250ns).
  - RS-FEC selection could be done by Auto Neg protocol.
- Could use RS, PCS, and FEC specifications per common architecture slides.
- Could specify EEE per adapted specifications.

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/salunke_082814_25GE_adhoc.pdf and
http://www.ieee802.org/3/25GSG/public/adhoc/architecture/mellitz_01_081914a_25GE_adhoc.pdf

Example 25GBASE-CR “long” channel loss budget
Same IL budget as 100GBASE-CR4 92A.5



based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/salunke_082814_25GE_adhoc.pdf

Example Loss Budget
"long" cables, 5 m with FEC

- Possible end to end loss Budget (per 92A.5)
  - end to end (TP0 to TP5) budget = 35 dB
  - host PCB loss = 6.81 dB per host
  - host connector loss = 1.69 dB per host
  - bulk cable loss budget = 35-6.81*2-1.69*2 = 18 dB
- To show that the cable assembly budget agrees with this
  - cable assembly loss budget (TP1 to TP4) = 22.48 dB
  - test fixture PCB loss = 1.17 dB per fixture
  - test fixture connector loss = 1.07 dB per fixture
  - bulk cable loss budget = 22.48 – 1.17*2-1.07*2 = 18 dB
- Cable length / gauge examples that can meet this
  - 3m cable @ 5.5 dB/m @ 30 AWG = 16.5 dB
  - 5m cable @ 3.5 dB/m @ 26 AWG = 17.5 dB ← optimum diameter
  - 6m cable @ 3.0 dB/m @ 24 AWG = 18 dB
  - 7m cable @ 2.5 dB/m @ 22 AWG = 17.5 dB ← maximum diameter.

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/salunke_082814_25GE_adhoc.pdf

Example Loss Budget
"short" cables, application #1, no FEC and low latency

- Possible end to end loss Budget
  - end to end (TP0 to TP5) budget = 30 dB
  - host PCB loss = 6.81 dB per host
  - host connector loss = 1.69 dB per host
  - bulk cable loss budget = 30-6.81*2-1.69*2 = 13 dB
- To show that the cable assembly budget agrees with this
  - cable assembly loss budget (TP1 to TP4) = 22.48-5 = 17.48 dB
  - test fixture PCB loss = 1.17 dB per fixture
  - test fixture connector loss = 1.07 dB per fixture
  - bulk cable loss budget = 17.48 – 1.17*2-1.07*2 = 13 dB
- Cable length / gauge examples that can meet this
  - 2m cable @ 5.5 dB/m @ 30 AWG = 10.5 dB
  - 3m cable @ 3.5 dB/m @ 26 AWG = 10.5 dB ← optimum diameter.
  - 4m cable @ 3.0 dB/m @ 24 AWG = 12 dB
  - 5m cable @ 2.5 dB/m @ 22 AWG = 12.5 dB ← maximum diameter.

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/salunke_082814_25GE_adhoc.pdf

“short” cables, application #1, no FEC and low latency considerations

- Since FEC is used MTTFPA needs to be analyzed.
  - Concern of MTTFPA from DFE burst errors, for operating without FEC
  - Option (1) – use COM procedure to verify if BER and MTTFPA is acceptable
  - Option (2) – use 10G KR FEC to correct DFE burst errors (latency 82 ns)
  - (coding gain ~2 dB, is less than one meter additional cable length)

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/salunke_082814_25GE_adhoc.pdf

Example loss Budget
"short" cables, application #2, with FEC and higher host loss

- Possible Loss Budget
  - bulk cable budget = (18 dB / 5 m) * 3 m = 10.8 dB
  - reduction in bulk cable loss = 18 – 10.8 = 7.2 dB
  - host PCB loss= 6.81 + 7.2/2 = 6.81 + 3.6 = 10.41 dB per host
  - host connector loss = 1.69 dB per host
  - end to end (TP0 to TP5) loss budget = 10.41*2 + 1.69*2 + 10.8 = 35 dB
- To show that the cable assembly budget agrees with this
  - bulk cable budget = (18 dB / 5 m) * 3 m = 10.8 dB
  - test fixture PCB loss = 1.17 dB per fixture
  - test fixture connector loss = 1.07 dB per fixture
  - cable assembly loss budget (TP1 to TP4) = 10.8+2*1.17+1.07*2 = 15.28 dB
- Cable length / gauge examples that can meet this
  - 2m cable @ 5.5 dB/m @ 30 AWG = 10.5 dB
  - 3m cable @ 3.5 dB/m @ 26 AWG = 10.5 dB

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/mellitz_01_081914a_25GE_adhoc.pdf

FEC selection

- RS-FEC selection could be done by Auto Neg protocol.
- Host port configured in “short cable” mode (SCM) or “long cable” mode (LCM), based on which PHY can advertise availability of RS-FEC (Y or N).

	Cable type	Server Port (FEC = Y / N)	Switch Port (FEC = Y / N)	RS-FEC selection by AN
(1)	25G CR short cable	SCM (N)	SCM (N)	N
(2)	25G CR short cable	SCM (N)	LCM (Y)	N
(3)	25G CR short cable	LCM (Y)	SCM (N)	N
(4)	25G CR short cable	LCM (Y)	LCM (Y)	Y
(5)	25G CR long cable	LCM (Y)	LCM (Y)	Y

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/salunke_082814_25GE_adhoc.pdf

Other related topics

- Host electrical connectors (MDI)
  - could consider 2 connectors in 25G CR channel model.
  - SFP28, and QSFP28 (for 4x25G breakout).
- 2 types of RX tolerance test needed for host –
  - (1) Long cables (with RS-FEC) @ BER ~ 1E-5.
  - (2) Short cables application #1 (no FEC) @ BER ~ 1E-12.
  - (3) Short cables application #2 (higher host loss) @ BER ~ 1E-5.

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/salunke_082814_25GE_adhoc.pdf and http://www.ieee802.org/3/25GSG/public/adhoc/architecture/mellitz_01_081914a_25GE_adhoc.pdf

What we can leverage from 100G CR4

	Spec or Function	100G CR4	25G CR
(1)	PCS encoding	Clause 82	See common digital slides
(2)	RS-FEC encoding	Clause 91	See common digital slides
(3)	Auto Negotiation	Clause 73	Updated to include 25G CR and optional RS-FEC.
(4)	Link Training	Clause 92	Single lane version of Clause 92.
(5)	Link block diagram and test points	Clause 92.7	Exactly same.
(6)	Host TX, RX compliance tests	Clause 92.8	Exactly same, add RX test for "No FEC" mode.
(7)	Cable assembly compliance tests	Clause 92.10	Exactly same for "long cable", add test for "short cable", including COM.
(8)	Test fixtures	Clause 92.11	Exactly same.
(9)	Host connector (MDI)	92.12	keep QSFP28, add SFP28.
(10)	Channel loss budget and allocations	Annex 92A	Exactly same for "long cable", add section for "short cable".
(11)	Host TX, RX chip electrical spec	Annex 92A	Exactly same.

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/salunke_082814_25GE_adhoc.pdf

25GBASE-SR Optical PHY

Example specifications

25GBASE-SR Summary

- RS, PCS, and FEC
  - Could use RS, PCS, and FEC specifications per common digital architecture slides.
- Chip-to-module interface
  - Interface could be specified per CAUI-4 chip-to-module Annex 83E specifications adapted for a single lane.
- Electrical connector
  - Could use copper twin-ax cables port interfaces & form factors: SFP28, QSFP28, CFP4
- Optical interface specs
  - Could use 32GFC and 100GBASE-SR4, both of which include applicable ~25 Gb/s optical lane specifications.
  - No new component developments.
  - <1 Watt SFP+ form factor has been demonstrated (32GFC samples)
- Optical MDI
  - Could use same MDI as SFP+ and QSFP optical modules: LC and MPO connectors

No technical risk + extensive industry experience + full suite of existing standards near completion to draw from = rapid standardization

based on http://www.ieee802.org/3/25GSG/public/adhoc/optical/king_090314_25GE_adhoc.pdf

25G Attachment Interfaces

Example specifications

XXVAUI C2C

- For any PHY to attach MAC device to SERDES device.
- Interface could be specified per CAUI-4 chip-to-chip Annex 83D specifications adapted for a single lane.
- Specify EEE per the adapted specifications.

XXVAUI C2M

- For 25GBASE-SR PHY and Active Cable attachment.
- Interface could be specified per CAUI-4 chip-to-module Annex 83E specifications adapted for a single lane.
- Specify EEE per the adapted specifications.

Management

Management Summary

- Registers per MDIO Clause 45
- Managed Objects per Clause 30.

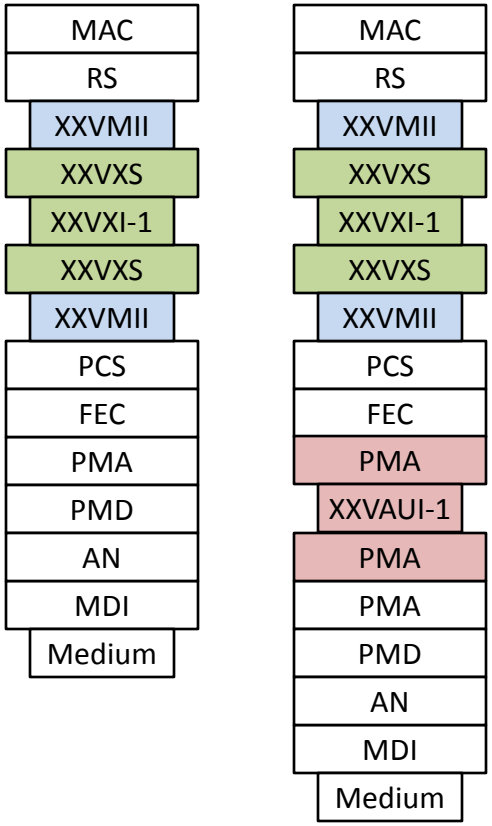
Conclusion

- Objectives are served by mature technology
- Options exist to further reduce cost
- We have demonstrated technically feasibility.

Thank You

Backup slides

Other possible 25GE PHY sublayer stacks



based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/brown_081214_25GE_adhoc.pdf

Example 25GBASE-CR Link block diagram
Same as 100GBASE-CR4

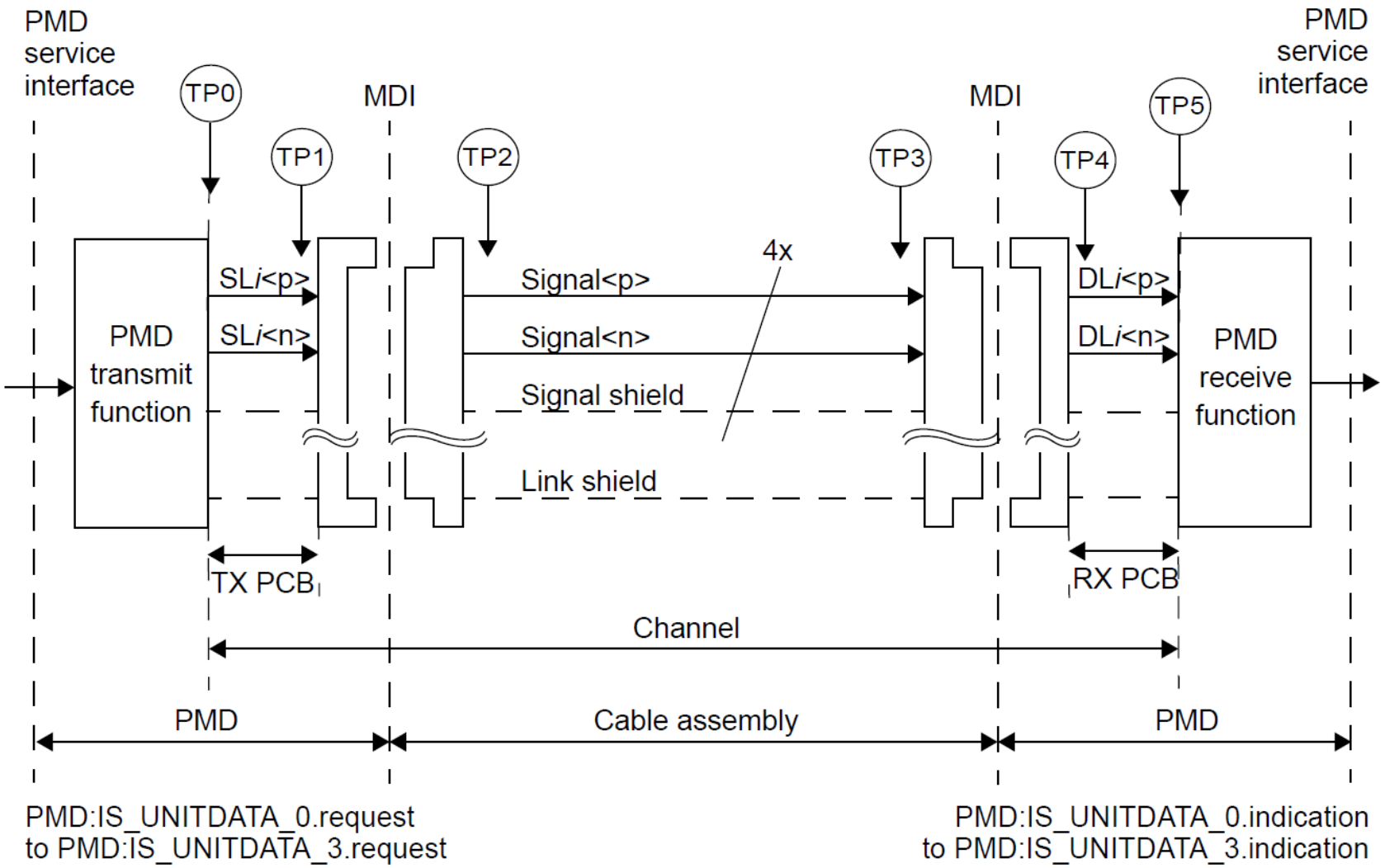


Figure 92–2—100GBASE-CR4 link (one direction is illustrated)

based on http://www.ieee802.org/3/25GSG/public/adhoc/architecture/salunke_082814_25GE_adhoc.pdf