CR Asymmetric Host Discussion

Beth Kochuparambil

Comments Against D2.0 - Recap

Cl **162** *SC* **162.9.3** *P* **154** *L* **21** # 166

Dawe, Piers Nvidia

Comment Type **TR** *Comment Status* **R** *CR port type*

The draft loss budget wastes over 3 dB in nearly every case.
The recommended maximum insertion loss allocation for the host traces plus BGA footprint and host connector footprint, of 6.875 dB, compares very poorly with C2M's host insertion loss up to 11.9 dB, making passive copper expensive and unattractive for a switch, while a full range of NICs can be made within only 3.75 dB. Server-switch links will get made with an asymmetric loss budget, so it would be better for the standard to regularise what will happen anyway. By the way, many server-switch links will be asymmetric anyway (different form factors at server and switch ends), and that's already allowed in this draft.
This change would also benefit CR switch-switch links because the shortest ports would get credit for their low loss.

SuggestedRemedy

As we have done for C2M, create two kinds of CR ports. Host loss allocations of 3.75 dB and 10 dB. Short can connect to short or long with same cable as today; long to long is not supported. Add entries in Clause 73 Auto-Negotiation to advertise short and long to the other end.
In Table 162-10, provide separate limits for Linear fit pulse peak (min).
In Table 162-14, provide separate rows for Test channel insertion loss: for testing the short host input the values for Test 2 are 10-6.875 = 3.125 dB higher (26.75 dB and 27.75 dB), while for the long host input the values for Test 2 are 6.875-3.75 = 3.125 dB lower (20.5 dB and 21.5 dB). No change needed for Test 1.
In 162A.4, provide two equations for each of IL_PCBmax and for ILHostMax and show them in Fig 162A-1 and 2. In 162A.5, provide two Value columns in Table 162A-1. Adjust figures 162A-3 and 4.

For discussion: should a "long" cable, 19.75+2*(6.875-3.75) = 19.75+6.25 = 26 dB max (maybe 3 m) be defined? A CR link could have no more than one of the three host, cable, and host being "long".

We could choose other names than "short" and "long" for the ports, possibly "short" and "medium" (as a C2M host can be "longer"), or A and B, somewhat like USB.

In 162.11.7.1.1, zp, representing the extra loss a host has above an MCB, could be made asymmetric but I believe that would not bring an improvement in accuracy.
There could be a third kind of CR port with 6.875 dB but this would not be useful for server-switch links, would be useful for only a subset of switch-switch links, for which passive copper is a subset anyway, so it doesn't seem worthwhile.

Response *Response Status* **U**

REJECT.

The following presentation was reviewed by the task force:
https://www.ieee802.org/3/ck/public/adhoc/apr28_21/dawe_3ck_adhoc_01_042821.pdf

The suggested remedy would require two or three different CR port types.

The assymetric-port approach was discussed early in this project.
Straw Poll #1 from the July 2018 Task Force meeting indicated strongest support for the current specification.
https://www.ieee802.org/3/ck/public/18_07/minutes_3ck_0718_approved.pdf

Based on discussion and straw poll 6 and 7, there is interest in exploring this proposal further. However, the proposal is not sufficiently complete at this time. A complete proposal and consensus is required.

Straw poll #6 (direction, chicago rule)
Straw poll #7 (direction, pick one)
I would support a new pair of CR port types with reduced host insertion loss limit on one end (e.g., NIC) and increased host loss limit on the other end (e.g., switch) similar to slide 7 of dawe_3ck_adhoc_01_042821.

Strawpoll #6
A: Yes 27
B: No 13
C: Need more information 29
D: Abstain 7

Straw poll #7
A: Yes 22
B: No 11
C: Need more information 11
D: Abstain 6

Comments Against D2.1

Cl **162** *SC* **162.9.3** *P* **163** *L* **18** # 92

Dawe, Piers — Nvidia

Comment Type **TR** *Comment Status* **X**

The draft CR loss budget wastes over 3 dB in nearly every case. The relative range of host losses, 6.875/2.3 = 3:1, is too small for switch layout yet not needed for NICs.
The recommendation for the host traces plus BGA footprint and host connector footprint, 6.875 dB, compares very poorly with C2M's host insertion loss up to 11.9 dB, making passive copper to this draft expensive and unattractive for a switch, yet a full range of NICs can be made with only 3.75 dB. Server-switch links are asymmetric in form factor (e.g. QSFP-DD to 2 x QSFP) and will get made with an asymmetric loss budget, so it would be better for the standard to regularise what will happen anyway. C2M already has short and long ports.
This change would also benefit CR switch-switch links because the shortest ports would get credit for their low loss.
The symmetric budget is used for some designs under way and may be useful in future for LOM, so it is kept here, and the better way added.

SuggestedRemedy

3 classes of CR ports, host loss allocations of A 10, B 6.875, C 3.75 dB. B is as D2.1.
A connects to C, B to B or C, C to A, B or C.
Use 2 bits in Clause 73 Auto-Negotiation Link codeword Base Page to advertise A, B or C to the other end. In the Priority Resolution function, an A port ignores a 100G/lane Technology Ability Field bit from an A or B port, a B port ignores a 100G/lane Technology Ability Field bit from an A port.
In Table 162-10, add limits A and C for linear fit pulse peak ratio (min). Change text in 162.9.3.1.2 to refer to the table.
In Table 162-14, add columns for Test 2 (high loss), A and C, with test channel insertion loss: A: 6.875-3.75 = 3.125 dB lower (20.5 dB to 21.5 dB), and C: 10-6.875 = 3.125 dB higher (26.75 dB to 27.75 dB). No change needed for Test 1.
In 162A.4, add equations for IL_PCBmax and ILHostMax A and B and show them in Fig 162A-1 and 2. In 162A.5, add Value columns A, C in Table 162A-1 (ILChmin and ILMaxHost differ). Adjust figures 162A-3 and 4.

Cl **162** *SC* **162.11** *P* **177** *L* **29** # 93

Dawe, Piers — Nvidia

Comment Type **T** *Comment Status* **X**

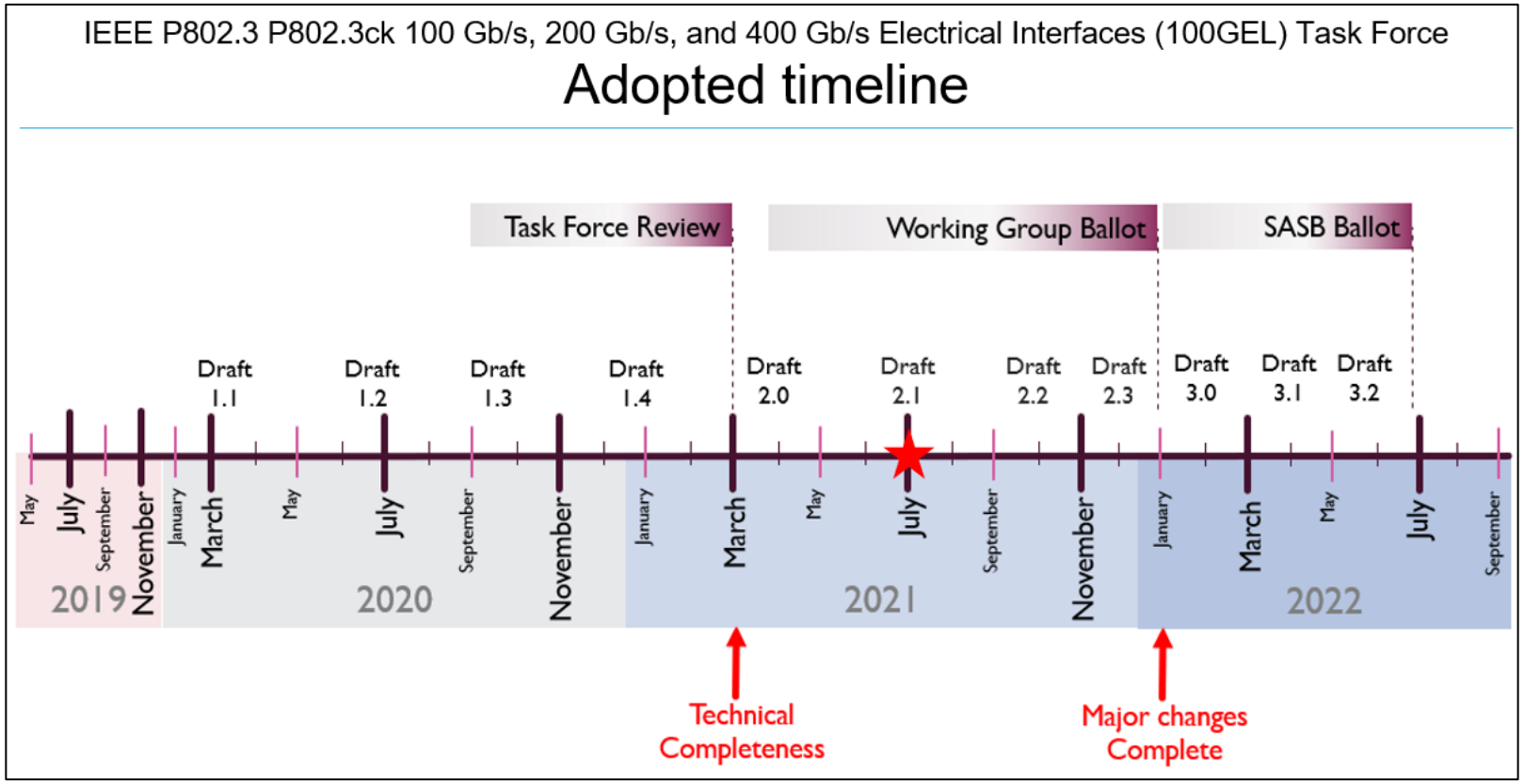
The poor max cable loss makes CR unattractive, while all NICs and some ports on any switch have host loss going to waste. Enabling longer cables on a minority of links is needed.
In the remedy, each host knows the other host's loss class through AN and the cable's loss class from its I2C compliance code, so the situation is just like any other CR scenario, no extra management features needed in the spec for the long cable class.

SuggestedRemedy

2 classes of cable, which could be called "short" (19.75 dB, as today) and "long", 19.75+2*(6.875-3.75) = 19.75+6.25 = 26 dB max (achievable cable length 3 m). Long cables connect port types C (see another comment) at both ends, short cables connect a valid combination of A, B, C.
In 162.11.2, cable assembly insertion loss, change text to refer to Table 162-17.
In 162.11.7.1.1, add zp = 30.7 mm for the "short" cable.
In Table 162A-1, add a column for the A-short-A scenario (ILCamax differs).

Schedule Impacts

- This would be a significant change to draft and industry understanding
- Would need several iterations for “clean up”
- If not made this round, we’re likely to impact draft cycles, thus ratification date.



Presentations

- May & July 2018
  - Piers, Rob, Jane, Ali, etc
- May 2021 by Piers
- ~~Ad Hoc 2021 by Kent~~
- Lisa in B400G Study Group

Straw Poll Language

- At this time, I support a new pair of CR port types with reduced host insertion loss limit on one end (e.g., NIC) and increased host loss limit on the other end (e.g., switch) similar to ~~slide 7 of dawe_3ck_adhoc_01_042821 for P802.3ck.~~ dawe_3ck_adhoc_01_071421.
  - Yes
  - No
  - Abstain

  **note this is just to capture current thoughts on the subject.