Consideration of a baseline spec for 400GBASE-ZR SD and SF Signaling

Bo Zhang

2020-10-08

P802.3cw/P802.3ct teleconference

Outline

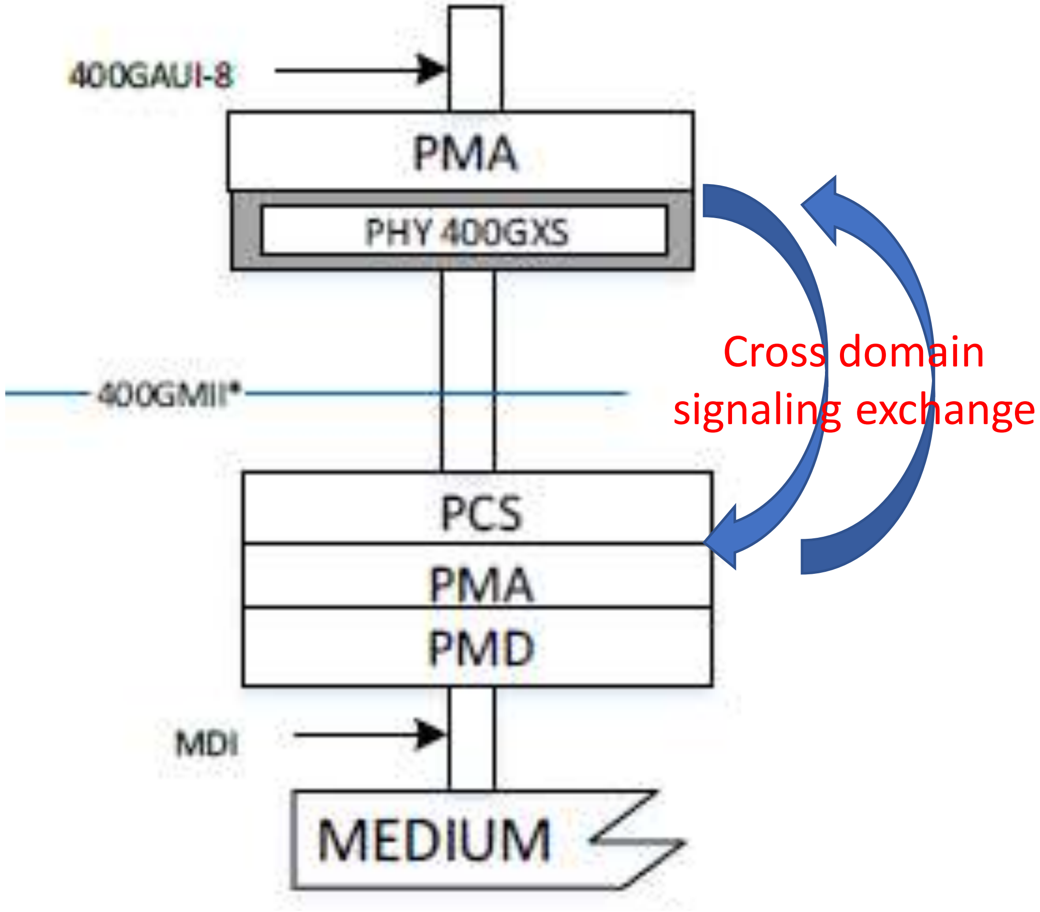
- From P802.3cw chief editor’s report
- IEEE 802.3 signaling baseline
- OIF 400ZR baseline consideration
- Proposal

From P802.3cw Chief Editor

- From chief editor’s report issenhuth_3cw_01_200813
  - New clause 155 - PCS (including FEC) and PMA for 400GBASE-ZR
  - PCS/PMA baseline adopted from lyubomirsky_3cn_01b_0119

- Recap adopted PCS/PMA baseline
  - Framing format for 400GBASE-ZR PCS and PMA were reviewed and proposed
  - Focus has been on the ZR specific PCS/PMA layer from 400GMII interface towards the ZR PMD and ZR medium
  - No proposal was covered yet for signal degrade/signal fault handling for 400GBASE-ZR

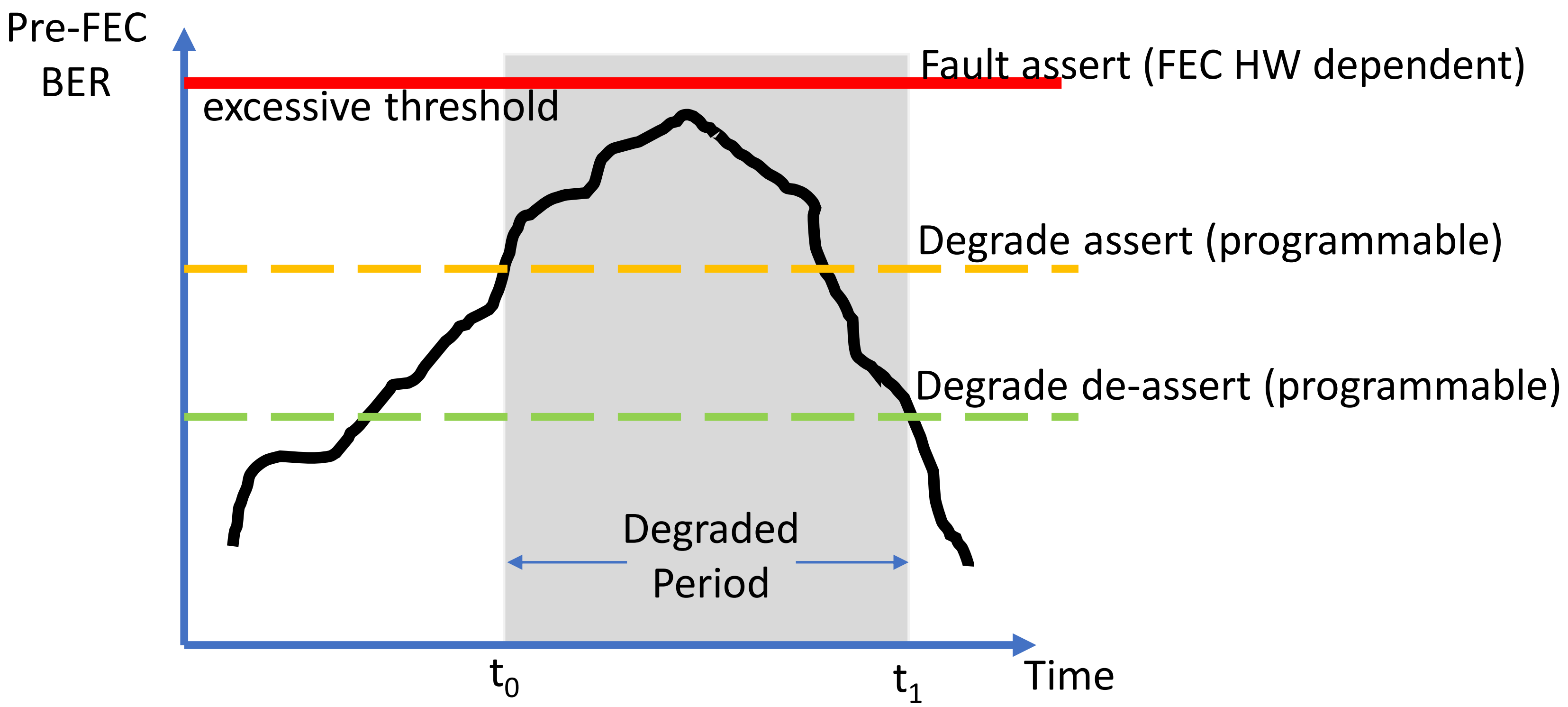
High Level View



- Signaling information exchange happens between PCS FEC layer with PHY 400G extended sublayer (400GXS), on both transmit and receive directions.

- This signaling exchange crosses the 400GMII interface and therefore will need to touch cross domain clauses such as 400GXS layer and PCS layer.

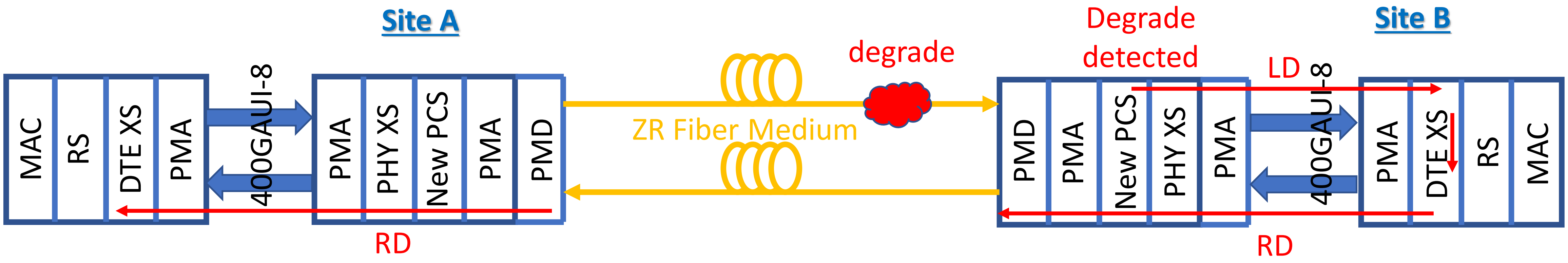
Signal Degrade (SD) and Signal Fault (SF)



- FEC excessive threshold crossing will result in the assertion of signal fault (SF).
- FEC degrade threshold crossing, defined by degrade assert and de-assert levels in observed interval periods, will result in the generation of signal degrade (SD) signaling.
- Both SF and SD can trigger Local Fault (LF)/Local Degrade (LD) as well as Remote Fault (RF)/Remote Degrade(RD).
- Consequent actions may include re-routing traffic away from deteriorated link.

https://www.ieee802.org/3/bs/public/15_11/maki_3bs_01a_1115.pdf

Examples of SF and SD (1)

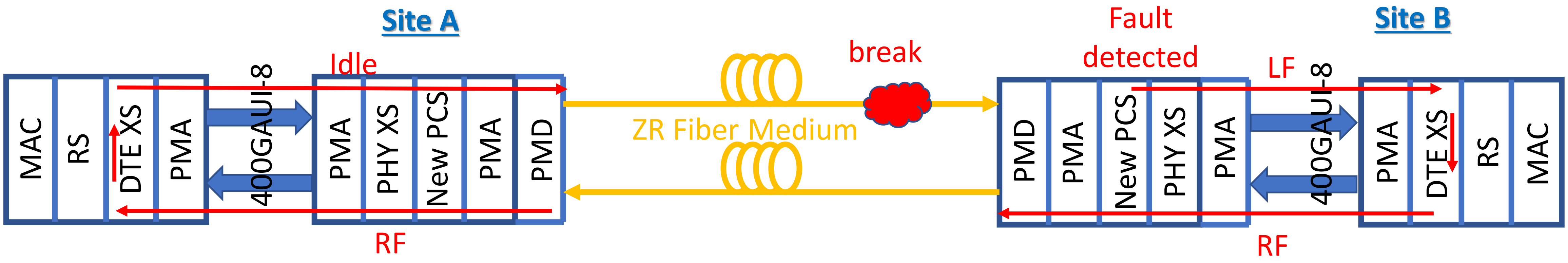


- ***DTE XS : Host side of the Extender Sublayer***
- ***PHY XS : External device side of the Extender Sublayer***

- New PCS at Site B exceeds pre-FEC BER *degrade assert threshold* and therefore sends *local degrade (LD)* to DTE XS at Site B.
- DTE XS at Site B sends *remote degrade (RD)* to DTE XS at Site A.
- Traffic is unaffected as long as the pre-FEC BER *excessive threshold* is not reached.
- Once BER lowered below *degrade de-assert threshold*, LD and RD signaling are cleared.
- In principle, degrades could also happen in the AUI-n interface and get detected in either in DTE XS or PHY XS.

https://www.ieee802.org/3/bs/public/16_05/ofelt_3bs_01a_0516.pdf

Examples of SF and SD (2)

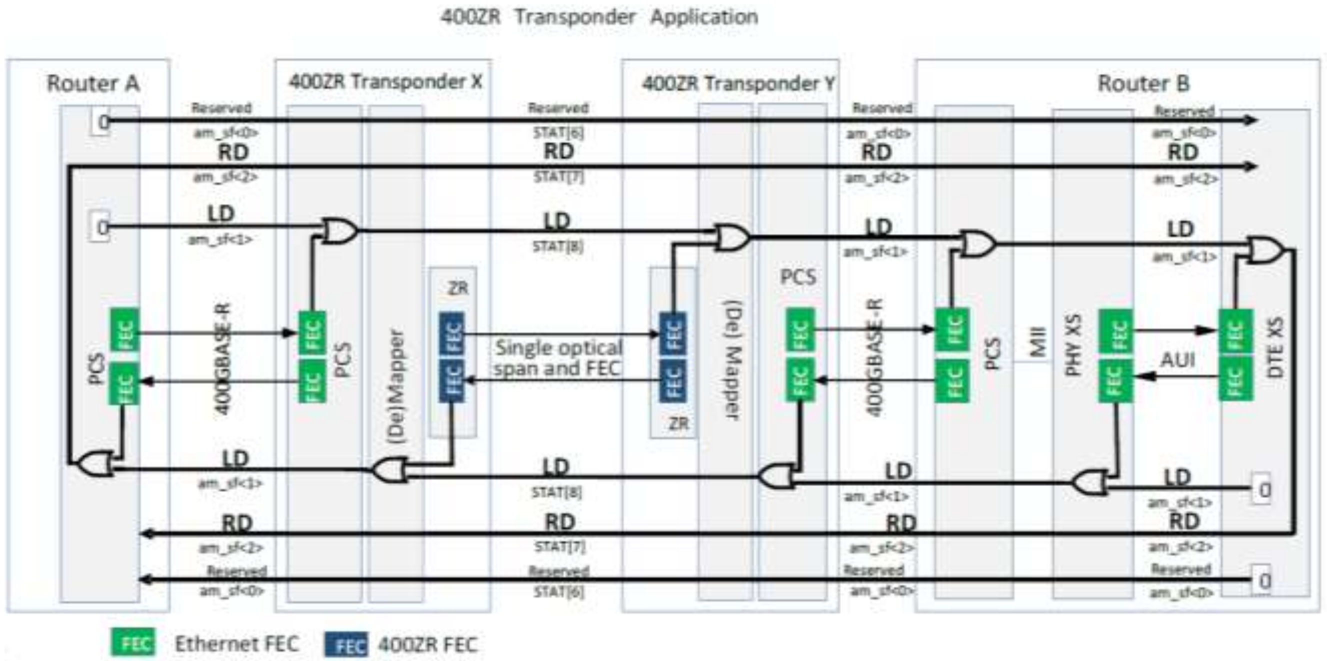


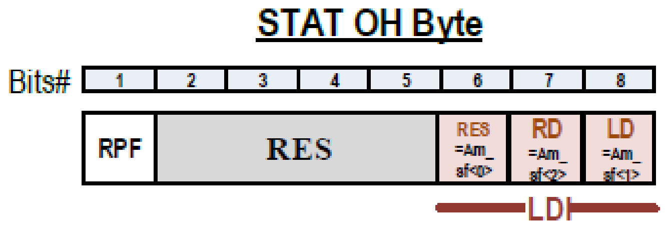
- ***DTE XS == Host side of the Extender Sublayer***
- ***PHY XS == External device side of the Extender Sublayer***

- New PCS at Site B exceeds pre-FEC BER *fault assert threshold* and therefore sends *local fault (LF)* to DTE XS at Site B.
- DTE XS at Site B sends *remote fault (RF)* to DTE XS at Site A.
- Traffic is interrupted, and Site A egress sends idle packets to Site B. Re-routing user traffic could be performed as a consequent action.

https://www.ieee802.org/3/cd/public/July16/ofelt_3cd_01_0716.pdf

Signaling interworking b/w host and ZR transceiver





STAT OH Byte

Bits#	1	2	3	4	5	6	7	8
	RPF	RES				RES =Am_sf<0>	RD =Am_sf<2>	LD =Am_sf<1>
						LDI		

- Pre-FEC BER monitors are used to detect and insert link degrade at both the 400ZR optical link media interface as well as the 400GBASE-R interface.
- IEEE 802.3 has specified three bits in the alignment marker (am) field to carry link degrade indicator (LDI). Near and far end ZR transceivers shall exchange signaling in RD (am_sf<2>) bits as shown in above diagram.
- The status information in LD (am_sf<1>) bit shall be carried after additional processing per below.
  - In host to media datapath, the processing consists of OR'ing the ingress LD status in am_sf<1> bit of the 400GBASE-R signal with the local host interface *RS(544,514) FEC degrade* status and signaling LD in STAT[8]
  - In media to host datapath, the STAT[8] bit from the media interface is OR'ed with the *ZR FEC degrade* status am_sf<1> bit to the local host.

https://www.oiforum.com/wp-content/uploads/OIF-400ZR-01.0_reduced2.pdf

ZR Link Degrade Warning and Alarming

- Link Degrade (LD) signaling shall be based on the FEC decoder statistics. Fault detection calculation and threshold settings could be implementation dependent.
- The Performance Monitoring (PM) parameters are defined for determining a link degrade (LD) condition over a PM interval.

FEC decoder block, bit counters:

- *pFECblkcount* = FEC blocks counted over PM interval
- *pFECbitcount* = total number of bits counted over PM interval = (*pFECblkcount* × bits per FEC block), 64-bit value
- *pFECcorrbitblk* = FEC corrected bits per block (min., avg., max.) over PM interval
- *pFECcorrbit* = total number of FEC corrected bits over PM interval = ∑*pFECcorrbitblk* over PM interval. (64-bit value).

Pre-FEC BER block, bit counters:

- *pFECblkBER = FEC block BER (min., avg., max.) over PM interval = (pFECcorrbitblk/pFECblkcount)*
- *pFECBER = FEC BER over PM interval = (pFECcorrbit / pFECbitcount)*

Pre-FEC threshold settings:

- *FEC_excessive_BER_activate_threshold (programmable)*
- *FEC_excessive_BER_deactivate_threshold (programmable)*
- *FEC_degraded_BER_activate_threshold (programmable)*
- *FEC_degraded_BER_deactivate threshold (programmable)*

FEC degrade settings:

- *FECdetectdegraded = FEC degraded status condition over PM interval.*
- *FECexcessdegraded = FEC excessively degraded status condition over PM interval.*

PM interval:

PM_Interval = (programmable); default = 1 second

Further considerations and proposal

- Like 400GBASE-R, 400GBASE-ZR interface should specify FEC fault and degrade signaling schemes. Link Degrade signaling based on pre-FEC BER monitoring is an established feature defined in 802.3 clauses 118, 119, etc. LD/RD interworking mechanism specified in OIF 400ZR could be served as a good baseline.

- PCS-layer FEC error statistics monitoring are specified in IEEE 802.3 Clause 119 for 400GBASE-R. Similarly, ITU-T G.798 requires the collection of pre-FEC statistics for OTN interfaces with RS FEC or higher coding gain FECs. The definition of 400GBASE-ZR FEC error monitoring counters could also be useful to provide a minimum set of unified statistics as defined in OIF 400ZR.

- Propose to adopt OIF 400ZR signaling mechanism (shown in slide 8 and 9) as 400GBASE-ZR SF and SD signaling baseline.