802.3dj D3.2 Comment Resolution Common Topics

Matt Brown (Qualcomm), 802.3dj Chief Editor
Tom Huber (Nokia)
Leon Bruckman (Nvidia)

Introduction

- This slide package was assembled by the 802.3dj editorial team to provide background and detailed resolutions to aid in comment resolution.
- Specifically, these slides are for the various common-topic comments.

local_rx_ready assignments
Comment R2-49

Cl **178B** *SC* **178B.8.3.1** *P* **898** *L* **6** # R2-49

Hajduczenia, Marek RG Nets

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

The D3.2 split of the D3.1 variable local_rx_ready into the new rx_ok (page 899, which took over the substantive "receiver is ready / no further adjustments required" definition and all the state-diagram conditions) plus a retained-but-redefined local_rx_ready leaves local_rx_ready inconsistent with its remaining consumers. Its definition is now reduced to "Boolean variable that indicates the state of the receiver. Its value is set by the training control state diagram (Figure 178B-10a).", which does not say which state of the receiver is indicated, nor the conditions for true and false. In Figure 178B-10a, local_rx_ready is set true only on entry to TRAIN_REMOTE and set false on entry to SEND_TRAINING, RX_READY (nearest numbered line, page 901 line 42) and TRAIN_FAIL - Figure 178B-10b never assigns it. It is therefore false throughout RX_READY, PATH_READY, WAIT_DATAPATH and PATH_UP, i.e. during normal data transmission. Three consumers were not updated and still cite local_rx_ready: isl_ready (178B.8.2.1, page 893, "set to true when local_rx_ready and remote_rx_ready are true for all lanes"), PMD_signal_detect_i (179.8.5, page 426 line 35, "the logical AND of the variables local_rx_ready and remote_rx_ready"), and register bit 1.1476.8 "isl ready" (45.2.1.169c, page 102 lines 24 and 26). Since isl_ready is also a condition in the RTS update state diagram (Figure 178B-9, page 896 lines 10 and 22) and on the ISL_READY exit (page 901 line 38), as drafted isl_ready and PMD_signal_detect_i can never remain true in the operational states. Separately, the SEND_LOCAL branch of Figure 178B-10a (entered from QUIET on "!mr_training_enable * quiet_timer_done") never assigns local_rx_ready at all, leaving it undefined whenever mr_training_enable is false, although PMD_signal_detect_i and management bit 1.1420.15 read it with no mr_training_enable guard.

SuggestedRemedy

Resolve the split consistently. Either (a) do not clear local_rx_ready on entry to RX_READY (and TRAIN_FAIL and the PATH states as applicable) so that it stays true from TRAIN_REMOTE through PATH_UP, and give it a substantive definition on page 898 stating the conditions for true and false - or (b) if local_rx_ready is intended only as the transmitted "receiver ready" status bit of the training frame, restate the consumers in terms of rx_ok: change isl_ready in 178B.8.2.1 (page 893) to "set to true when rx_ok and remote_rx_ready are true for all lanes of the interface", change PMD_signal_detect_i in 179.8.5 (page 426 line 35) to "the logical AND of the variables rx_ok and remote_rx_ready", and update the 1.1476.8 descriptions in 45.2.1.169c (page 102 lines 24 and 26). In either case add "local_rx_ready <= false" to the action block of the SEND_LOCAL state in Figure 178B-10a (page 901 line 13) so the variable is defined on the non-training path.

Proposed response:

ACCEPT IN PRINCIPLE.
Make the following changes to Figure 178B-10a/10b.

- In the SEND_TRAINING and TRAIN_FAIL states, delete "local_rx_ready <= false".
- In the INITIALIZE and ISL_FAIL states, add "local_rx_ready <= false".
- In the RX_READY state, add "local_rx_ready <= true".

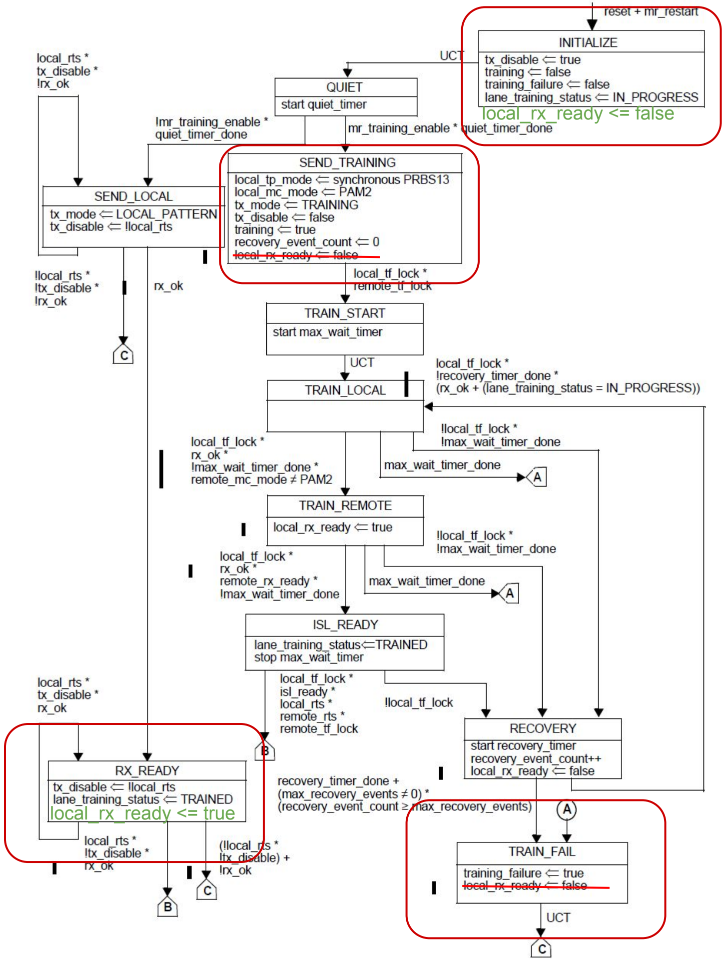


Figure 178B–10a—Training control state diagram, part 1

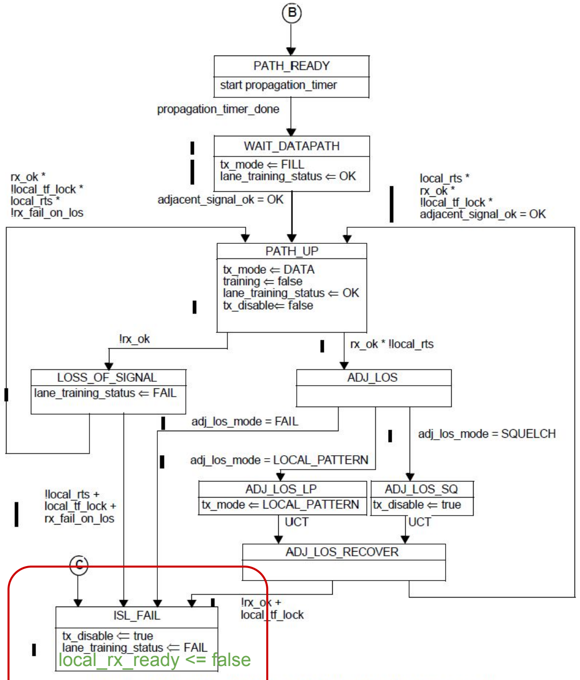


Figure 178B–10b—Training control state diagram, part 2

Comment R2-110

Cl **178B** *SC* **178B.8.2.1** *P* **894** *L* **42** # R2-110

Ran, Adee Cisco Systems, Inc.

Comment Type **T** *Comment Status* **D** *Restart (CA)*

rx_fail_on_los is defined as an ability (per the register bit, as noted in the editor's note), but in fact it is a "disability" - the receiver can't do anything but fail upon a LOS event.
Even if this is treated as an "inverted abilty bit", it is possible that the receiver can, by design, recover without failing (so rx_fail_on_los is false), but management prefers to force a restart on every failure. This is not supported by the current definition (read-only).

The common way to specify management variables is ability + enable as separate variables. The suggested remedy adds these management variables while keeping the semantic of rx_fail_on_los, so the state diagram is not changed.

SuggestedRemedy

Add the following variable definitions:
rx_los_recovery_ability: Boolean variable that indicates that when the training control state diagram (Figure 178B–10b) is in the PATH_UP state, the receiver can recover from loss of signal (rx_ok = false) without a restart of the ILT function. The value of this variable is implementation dependent.
mr_rx_los_recovery_enable: Boolean variable that is set by management to enable recovery from loss of signal. The default value is false.

Change the definition of rx_fail_on_los to:
rx_fail_on_los: Boolean variable that is true when either rx_los_recovery_ability or mr_rx_los_recovery_enable is false, and false otherwise.
Delete the editor's note after the definition.

In Table 178B-3, add a row for mr_rx_los_recovery_enable | 178B.8.2.1 | 1.1479.11 | 45.2.1.169f.
In Table 178B-4, replace rx_fail_on_los with rx_los_recovery_ability (keep mapping the MDIO register 1.1479.15)
In 45.2.1.169f, in the definition of bit 15, swap the 0 and 1 definitions (such that 1 indicates ability), and add a row for bit 11: Receiver recovery | 0=mr_rx_los_recovery_enable is false; 1=mr_rx_los_recovery_enable is true | R/W. Keep only bit 12 as reserved.

Proposed Response *Response Status* **W**

PROPOSED ACCEPT IN PRINCIPLE.

Implement suggested remedy except define mr_rx_los_recovery_enable as follows.

"mr_rx_los_recovery_enable: Boolean variable that is set by management to enable recovery from loss of signal without a restart of the ILT function. The default value is false."

Implement all with editorial license.

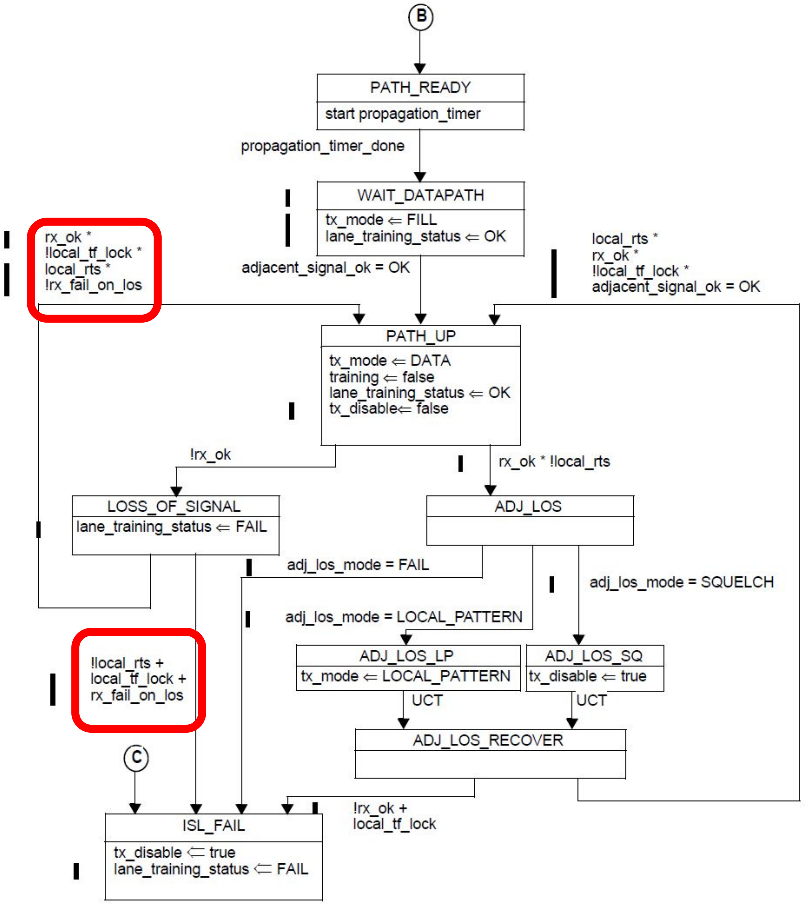


Figure 178B–10b—Training control state diagram, part 2

Comment R2-110 - Variable definition change and additions

rx_fail_on_los

Boolean variable that indicates the receiver requires a restart of the ILT function in order to recover from a loss of signal event (rx_ok becomes false) when the training control state diagram (Figure 178B–10b) is in the PATH_UP state. The value of this variable is implementation dependent. Default for this variable is true.

~~*Editor's note: The rx_fail_on_los variable was added per Draft 3.1 comment R1-248. The register bit for rx_fail_on_los is defined as a read-only ability bit. Therefore defining a default is not meaningful. Proposals to address this are encouraged.*~~

rx_fail_on_los:

Boolean variable that is true when either rx_los_recovery_ability is false or mr_rx_los_recovery_enable is false, and false otherwise.

rx_los_recovery_ability

Boolean variable that indicates that when the training control state diagram (Figure 178B–10b) is in the PATH_UP state, the receiver can recover from loss of signal (rx_ok = false) without a restart of the ILT function. The value of this variable is implementation dependent.

mr_rx_los_recovery_enable

Boolean variable that is set by management to enable recovery from loss of signal without a restart of the ILT function. The default value is false.

Comment R2-110 - MDIO

Table 45–133f—PMA/PMD interface training control bit definitions

Bit(s)	Name	Description	R/W[a]
1.1479.15	Receiver recovery ability	0 = supports recovery from loss of signal without training restart 1 = requires restart of training for recovery from loss of signal	RO
1.1479.14	Adjacent loss of signal squelch ability	0 = does not support mr_adj_los_mode = SQUELCH 1 = supports mr_adj_los_mode = SQUELCH	RO
1.1479.13	Adjacent loss of signal local pattern ability	0 = does not support mr_adj_los_mode = LOCAL_PATTERN 1 = supports mr_adj_los_mode = LOCAL_PATTERN	RO
1.1479.12:11	Reserved	Value always 0	RO
1.1479.10:9	Adjacent loss of signal mode	10 9 1 1 = Reserved 1 0 = mr_adj_los_mode is SQUELCH 0 1 = mr_adj_los_mode is LOCAL_PATTERN 0 0 = mr_adj_los_mode is FAIL	R/W
1.1479.8	Uses recovered clock	1 = Use recovered clock for transmission 0 = Use local clock for transmission	R/W
1.1479.7:0	Maximum recovery events	Maximum allowed number of transitions into the RECOVERY state in the training control state diagram (Figure 178B–10a and Figure 178B–10b)	R/W

[a] R/W = Read/Write, RO = Read only.

Swap

1.1479.11	Receiver recovery	0 = mr_rx_los_recovery_enable is false 1 = mr_rx_los_recovery_enable is true	R/W

Comment R2-93 - Breaking the comment into pieces

Cl **186** *SC* **186.2.2** *P* **650** *L* **22** # R2-93

Ran, Adee Cisco Systems, Inc.

Comment Type **TR** *Comment Status* **D** *ER1 payload (CA)*

In Table 186-1 the last column "Payload" contains values that are inconsistent with other specifications of ER1.

("Payload" is not defined in the text preceding the table, but based on the text in 186.2.3.3, the payload is what is mapped into the tributary frame)

In the first row, the value is "FEC:IS_UNITDATA.request" but this is a name of a service interface primitive, which communicates the input from the client (PCS or PMA). It does not describe the payload, because the payload is the output of the "Pre-mapping distribution" block (based on Figure 186-3) which has gone through the "inverse RS-FEC transmit" processing.

The 2nd through 4th rows have "Scrambled Idle", but there is no definition of scrambled idle in this clause. If the input is scrambled idle as generated by the PCS, then scrambling is removed by the inverse RS-FEC transmit processing, and the input to the mapping will be unscrambled idle characters. But there is no specified functionality to generate idle characters in clause 186.

The 5th row has "LF", which is not defined anywhere in the draft. I assume it means Local Fault, an ordered set that is defined in the Reconciliartion Sublayer. It is mentioned in 186.2.4.7.3 (without definition or reference), part of the receive function. But the _transmit_ function is not supposed to send local faults to the link partner... that would contradict the established use of this ordered set.

Note that "local fault" in 186.2.4.7.3 should be the 257-bit blocks created by transcoding the Local Fault ordered sets. Also, the replacement by local fault is repeated twice, in 186.2.4.7.3 and in 186.2.4.8, but occurs only once - the second instance is redundant.

Furthermore, 186.2.6 states that for some values of c_tx_mode, "the payload of the ER1 FEC frame is filled with the scrambled idle test pattern defined in 186.2.3.12". But 186.2.3.12 does not define a scrambled idle test pattern - it defines an encoded PRBS31 (there is no scrambled idle generator function in this sublayer).

Furthermore, when SIGNAL_OK of FEC:IS_SIGNAL.request is FAIL (meaning the adjacent interface has no data and is not training) it will be reflected into the RTS function by c_adjacent_signal_ok=FAIL, and thus c_adjacent_intf_rx_ready=false and c_local_rts=false. These values prevent the RTS state diagram (Figure 178B-15) from progressing past TRAINING_COMPLETE, but it is either there or in TRAINING, so c_tx_mode will be either TRAINING or NOT-RTS. These values are both specified in 186.2.6 to generate the test pattern defined in 186.2.3.12 (encoded PRBS31), not local fault.

In summary, the payload depending on FEC:IS_SIGNAL.request(SIGNAL_OK) is either the output of the "Pre-mapping distribution" block, or the encoded PRBS31 test pattern generator; this would match Figure 186-3. There is no "scrambled idle" or "LF" generator in the transmit direction.

Table 186–1—Mapping FEC:IS_SIGNAL.request to MNT bit values

FEC:IS_SIGNAL. request(SIGNAL_OK)	IFEC_align_status	Transmitted MNT field	Payload
OK	true	000 (OK)	FEC: IS_UNITDATA.request
OK	false	001 (READY)	Scrambled Idle
READY	don't care	001 (READY)	Scrambled Idle
IN_PROGRESS	don't care	010 (IN_PROGRESS)	Scrambled Idle
FAIL	don't care	011 (FAIL)	LF

The information in the last column of the table is incorrect

Concerns whether it is clear that 'local fault ordered sets' means 257b blocks comprising 4 LF ordered sets

Relates to the issue with the table, concerns consistency of text wrt the pattern that is inserted when the transmit function is not yet in data mode (i.e., there is no data available from the AUI, and the function has to insert a pattern so the GMP mapper has something to map)

Comment R2-93 - intended behavior

The GMP mapper must always have data to map. This is not adequately described in the current text.

- In normal operation that is the signal coming from the AUI, as transformed by the processes prior to the GMP mapper (which is not FEC:IS_UNITDATA.request).
- During APSU, the signal from the AUI is not forwarded, so a replacement is needed.
    - In typical implementations of this type of interface, this is a test pattern or Idle signal
    - Exactly what it is doesn't really matter since it won't be forwarded and there is a scrambler before FEC encoding
- If the AUI is failed, LF should be inserted in the transmit function

If the ER1 PHY is failed, LF should be inserted by the receive function. This is covered by text in 186.2.4.7.3 and 186.2.4.8.

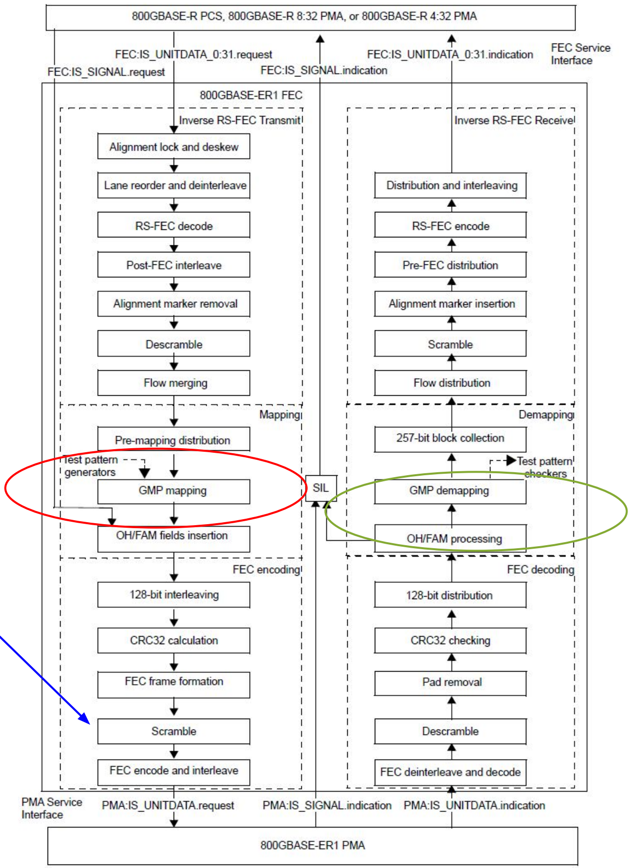


Figure 186–3—800GBASE-ER1 FEC functional block diagram

Comment R2-93 - Proposed changes (1)

Table 186-1 is about mapping service interface to overhead bits; this is not the optimal place to describe what goes into the payload area of the frame

Delete the last column of the table

Table 186–1—Mapping FEC:IS_SIGNAL.request to MNT bit values

FEC:IS_SIGNAL.request(SIGNAL_OK)	IFEC_align_status	Transmitted MNT field	~~Payload~~
OK	true	000 (OK)	~~FEC: IS_UNITDATA.request~~
OK	false	001 (READY)	~~Scrambled Idle~~
READY	don't care	001 (READY)	~~Scrambled Idle~~
IN_PROGRESS	don't care	010 (IN_PROGRESS)	~~Scrambled Idle~~
FAIL	don't care	011 (FAIL)	~~LF~~

186.2.6 Autonomous path startup (APSU) functions

The ER1 FEC shall provide the ready-to-send (RTS) function using the RTS without squelch state diagram (see 178B.11) as specified in Annex 178B.

The RTS status is signaled via the MNT bits using the mapping defined in 186.2.2. The variable c_rts_status is set as defined for FEC:IS_SIGNAL.indication(SIGNAL_OK) in Table 186–2.

The transmit function transmits ER1 FEC frames regardless of the value of c_tx_mode. When c_tx_mode = TRAINING, NOT-RTS, RTS, or FILL, the payload of the ER1 FEC frame is filled with the scrambled idle test pattern defined in 186.2.3.12.

The variable c_isl_ready is set to true when the FEC sublayer has aligned to the ER1 FEC frame and the ER1 tributary frame, and c_local_rx_ready is true. Note that this requires that the PMA be aligned to the DSP frame. The variable c_isl_ready is set to false if the FEC sublayer has not aligned to the ER1 FEC frame and ER1 tributary frame, or if c_local_rx_ready is false.

The variable c_remote_rts is set to true if the received SIGNAL_OK status in the MNT bits from the peer interface indicate a status of READY or OK (see Table 186–2), and is set to false otherwise.

Corresponding text in 186.2.6 needs to be updated; change the second sentence of the 3rd paragraph as shown:

When c_tx_mode = TRAINING, NOT-RTS, RTS, or FILL, the payload area of the ER1 ~~FEC~~ tributary frame is filled with ~~the scrambled idle~~ a locally generated signal such as the test pattern defined in 186.2.3.12.

Comment R2-93 - Proposed changes (2)

186.2.3.1.7 Flow merge

Following descrambling, the two flows of 257-bit blocks shall be merged in a round-robin fashion into a single stream of 257-bit blocks, starting with a block from flow 0.

186.2.3.2 Pre-mapping distribution

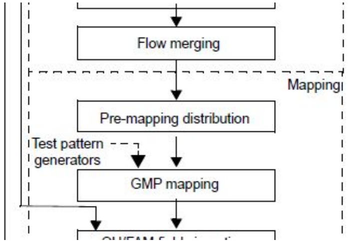
The stream of 257-bit blocks shall be distributed to 8 FEC flows (FEC flows 0 to 7) in a round-robin manner. The distribution shall be such that FEC flows 0, 2, 4, and 6 receive 257-bit blocks that originated from PCS flow 0, and FEC flows 1, 3, 5, and 7 receive 257-bit blocks that originated from PCS flow 1. The GMP mapper (see 186.2.3.3), FAM and pad insertion (see 186.2.3.4), and OH insertion (see 186.2.3.5) functions are performed individually on each FEC flow. The 8 FEC flows are then interleaved as specified in 186.2.3.6.

186.2.3.3 GMP mapper

Based on the incoming SIGNAL_OK status (see Table 186–1), either the 8 FEC flows or a locally generated signal is mapped into the payload area of the 800GBASE-ER1 tributary frame.

GMP is a generic mechanism that uses a sigma/delta distribution algorithm to accommodate an arbitrary bit rate difference between a payload and the space in which it is carried. The principles of the GMP mapper used in the 800GBASE-ER1 FEC sublayer are described in ITU-T G.709 Annex D and ITU-T G.709.1 Annex B. Details of the overhead encoding for 800GBASE-ER1 are described in 186.2.3.5.

The GMP mapper shall insert the serialized stream of 257-bit blocks corresponding to each FEC flow into the payload area of a 800GBASE-ER1 tributary frame. The 800GBASE-ER1 tributary frame is a structure



Reference to Table 186-1 no longer valid. Text between first and third paragraphs is not self-consistent in terms of what is mapped into the payload area. No mention anywhere of LF as a replacement signal

Add to the end of 186.2.3.1.7:

"When IFEC_align_status is true, this stream is provided to the pre-mapping distribution function. Otherwise, a stream of 257-bit blocks containing local fault ordered sets is provided to the pre-mapping distribution function."

Replace first paragraph of 186.2.3.3:

“The input to the GMP mapper is either the stream of 257b blocks output from the pre-mapping distribution function or the output of the test pattern generator (see 186.2.3.12).”

Replace first sentence of third paragraph of 186.2.3.3:

“The GMP mapper shall insert the stream of 257-bit blocks at its input into the payload area of an 800GBASE-ER1 tributary frame.”

Comment R2-93 - LF definition

LF is mentioned 3 times in the text:

- Proposed new text in 186.2.3.1.7 (slide 11)
- Existing text in 186.2.4.7.3
- Existing text in 186.2.4.8

> When any of the GID, IID, MAP, PT, or MSI values do not match the expected value, the payload area of the frame is replaced with local fault ordered sets.

> If the payload type or multiplex structure identification overhead (see 186.2.4.7.3) did not match the expected value, the 257-bit data blocks extracted from the GMP frame are replaced with LF ordered sets.

In all cases, “local fault ordered sets” or “LF ordered sets” is intended to mean a stream of 257b blocks, each of which is composed of four local fault ordered sets (i.e., the replacement signal is a stream of LF ordered sets that has been 257b-transcoded).

The intended meaning is clear, so there is no need to make any clarification to this text.