Proposed responses to comments against Clause 91

Adam Healey
LSI Corporation

IEEE P802.3bj Task Force
September 2012

Topics

- Skew and Skew Variation (26, 222, 235, 409)
- Energy Efficient Ethernet (95, 99, 100, 242, 243, 208, 210)
- Exception handling (53, 162, 239)
- Management (196, 197, 244)

Skew and Skew Variation

Skew point	Maximum Skew[a], ns	Max. Skew Variation, ns
SP0	29	0.2
SP1	29	0.2
SP2	43	0.4
SP3	54	0.6
SP4	134	3.4
SP5	145	3.6
SP6	160	3.8
SP7	29	0.2
At RS-FEC transmit	49	0.4
At RS-FEC receive	180	4.0
At PCS receive	49	0.4

[a] The Skew limit includes 1 ns allowance for PCB traces that are associated with the Skew points.

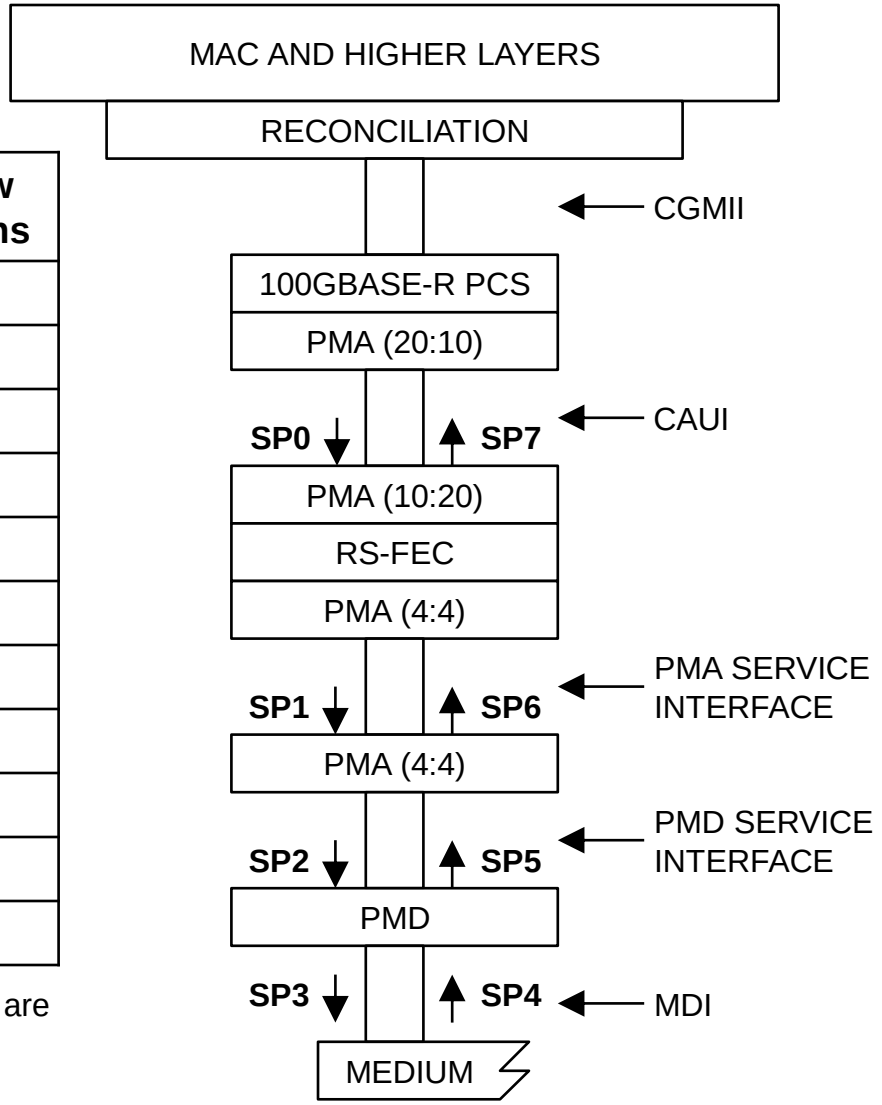


Figure 80–5a—100GBASE-R Skew points with RS-FEC and CAUI

Variable medium allocation

Skew point	Maximum Skew[a], ns	Max. Skew Variation, ns
SP0	29	0.2
SP1	29	0.2
SP2	43	0.4
SP3	54	0.6
SP4	54+medium	0.6+medium
SP5	65+medium	0.8+medium
SP6	80+medium	1.0+medium
SP7	29	0.2
At RS-FEC transmit	49	0.4
At RS-FEC receive	100+medium	1.2+medium
At PCS receive	49	0.4

[a] The Skew limit includes 1 ns allowance for PCB traces that are associated with the Skew points.

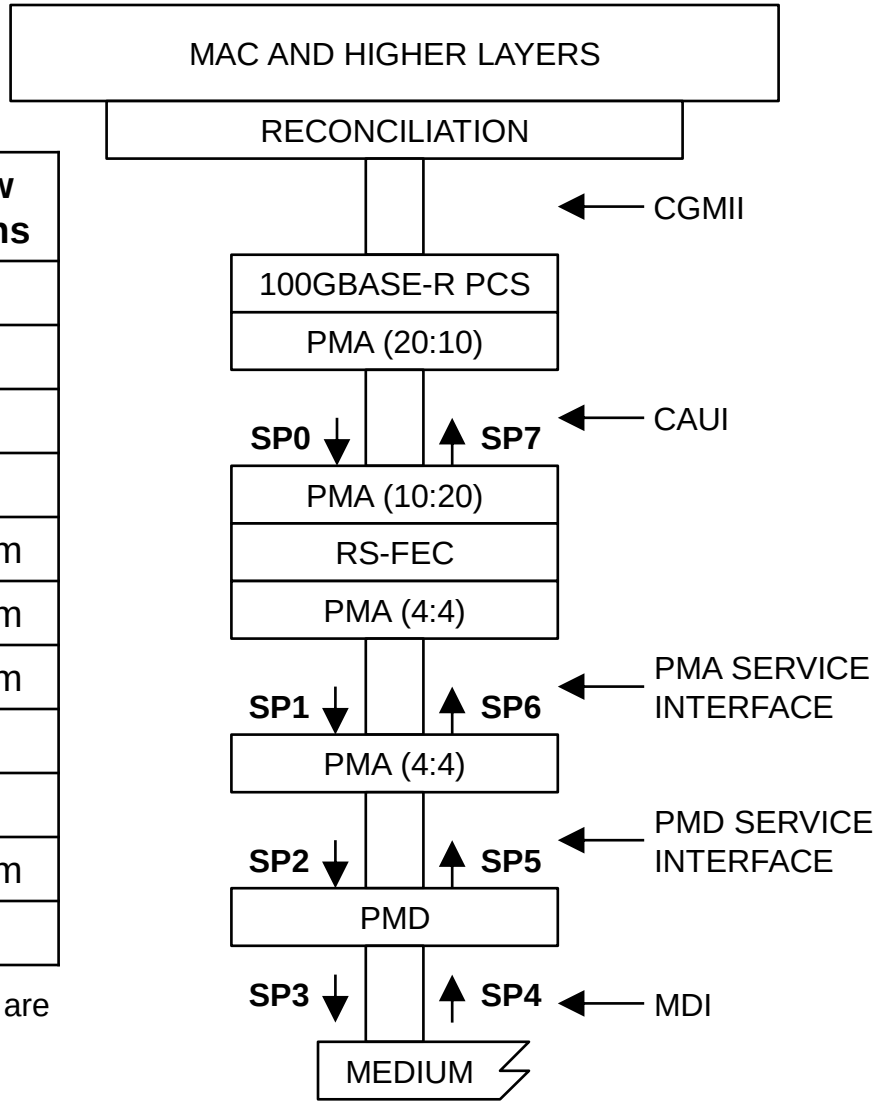


Figure 80–5a—100GBASE-R Skew points with RS-FEC and CAUI

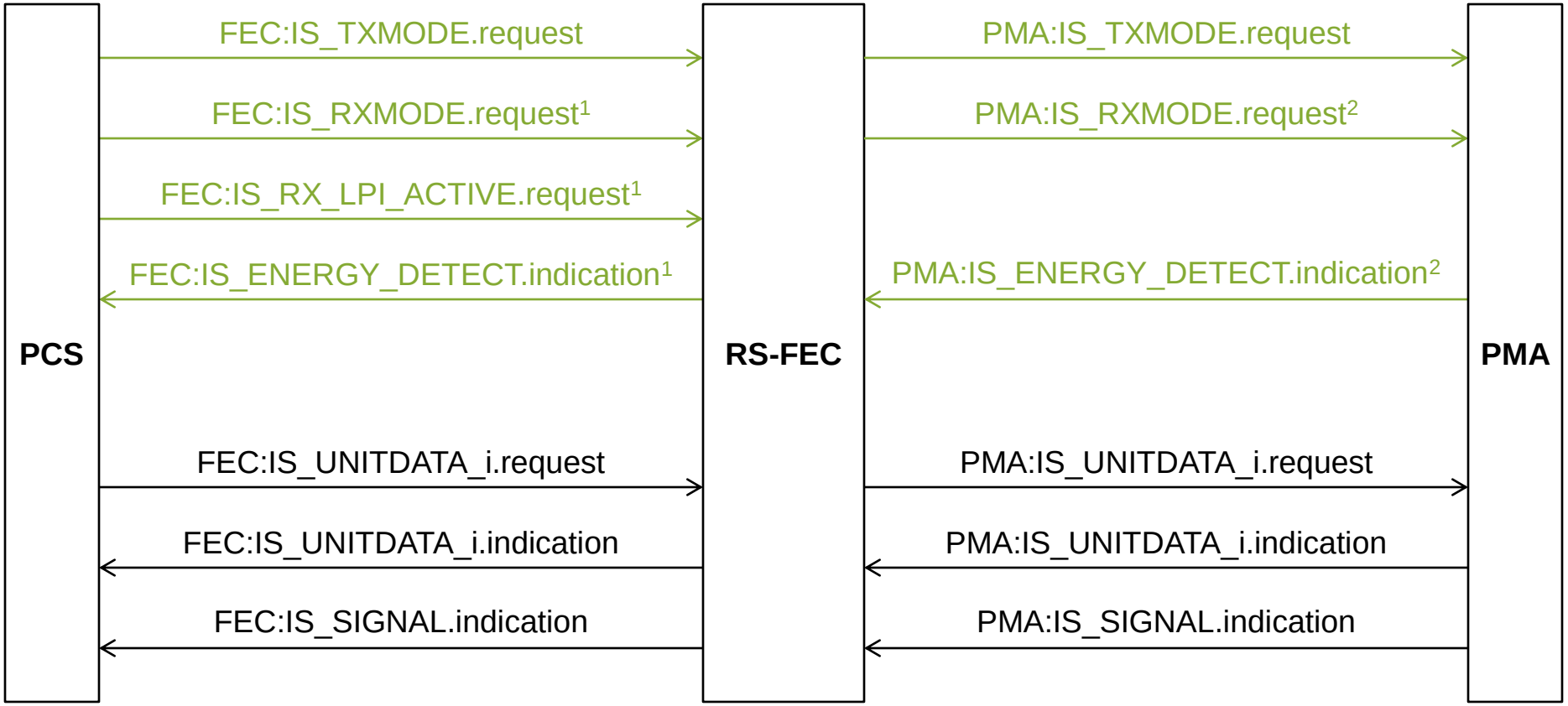
Alternative proposal

- 80 ns Skew and 2.8 ns Skew Variation allocated for medium
- Very generous for some media, e.g. electrical backplane
- Differentiate between “front-panel” and backplane interfaces?
  - Medium Skew 80 ns, Skew Variation 2.8 ns…
  - …other otherwise noted (backplane PMDs)

Topics

- Skew and Skew Variation (26, 222, 235, 409)
- Energy Efficient Ethernet (95, 99, 100, 242, 243, 208, 210)
- Exception handling (53, 162, 239)
- Management (196, 197, 244)

FEC/PMA service interface



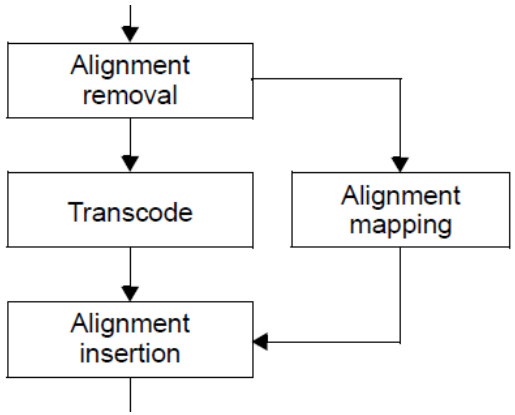
NOTE 1 – Direction reversed by or added by comment #99
NOTE 2 – Direction reversed by or added by comment #100

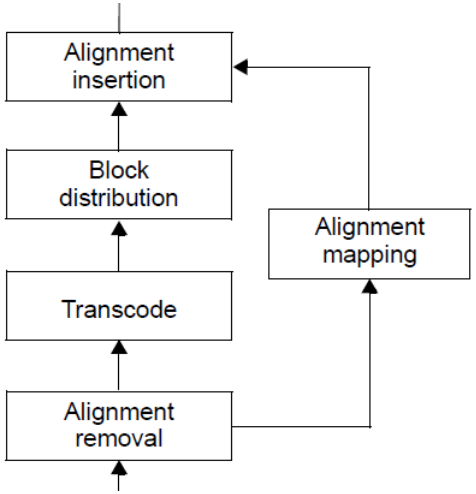
Problem statement

- RS-FEC defines a special mapping for alignment markers…
- …and therefore needs to know alignment marker positions

- Positions are learned via the FEC synchronization function...
- …but those positions are not continuously verified [1]

- Changes in alignment marker positions will not be tracked…
- …resulting in encoding and decoding errors during LPI



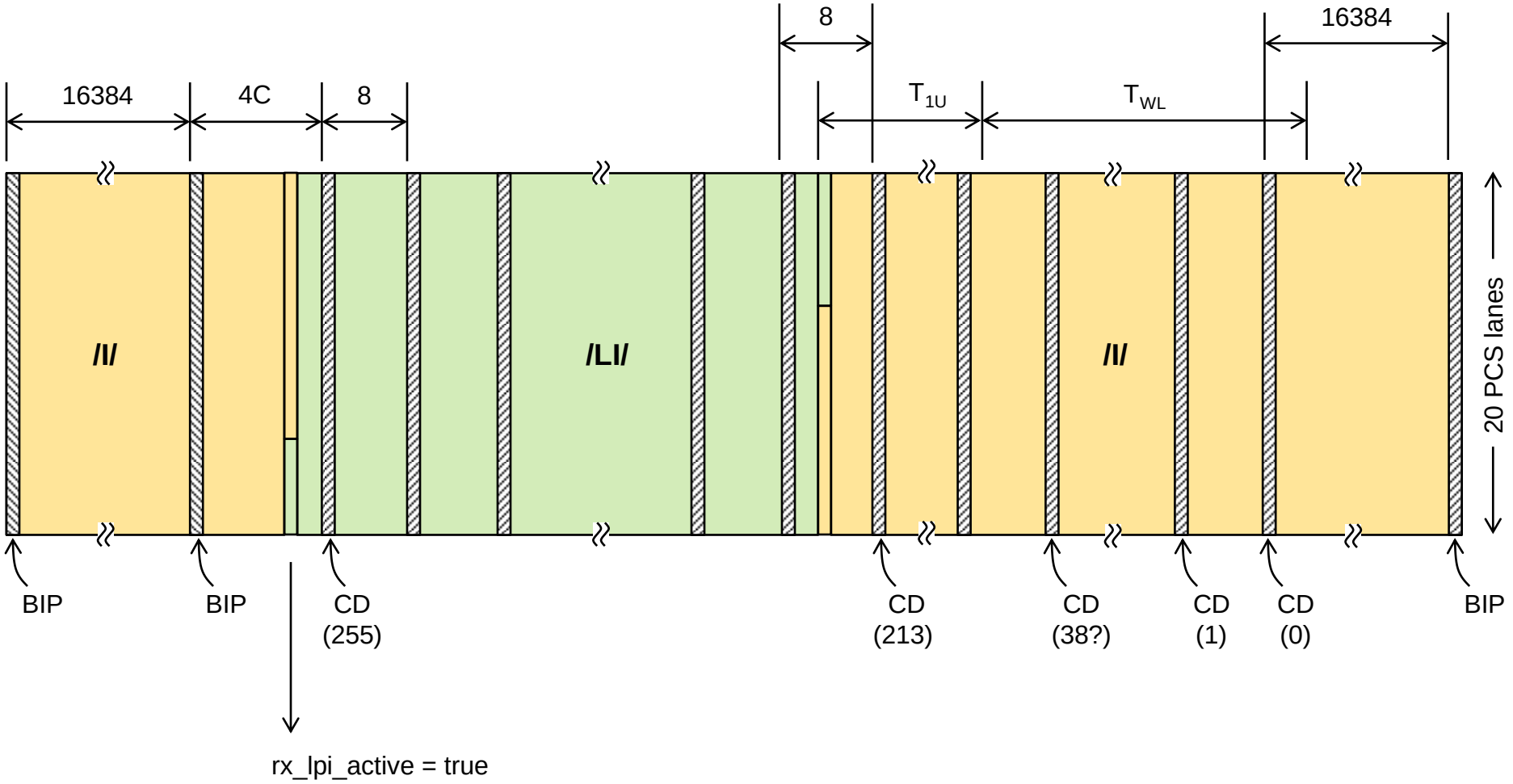


[1] As long as the codewords are correctable, the FEC lanes are deemed to be aligned

Normal to rapid alignment marker transition

- Transition from rapid to normal coordinated using down_count
- Must coordinate transition from normal to rapid to avoid errors
- Comment #195 suggests that the PCS sends the first RAM…
  - …after one block of /LI/ appears on PCS lane 0 and…
  - …when distance from last normal AM is a multiple of 4
- This ensures that…
  - …/LI/ precedes the first RAM and…
  - …RAMs coincide with the start of an FEC codeword

Graphical view



Methods for coordinated transition

- Service interface primitives
  - rx_lpi_active asserted by PCS upon receipt of /LI/
  - What about PCS transmit (define tx_lpi_active)?
  - However, no conveyance across CAUI

- Look for /LI/ as advance warning of transition but…
  - …this requires descrambling

- To use down_count…
  - …must already know alignment marker positions

- Continuously monitor potential RAM positions

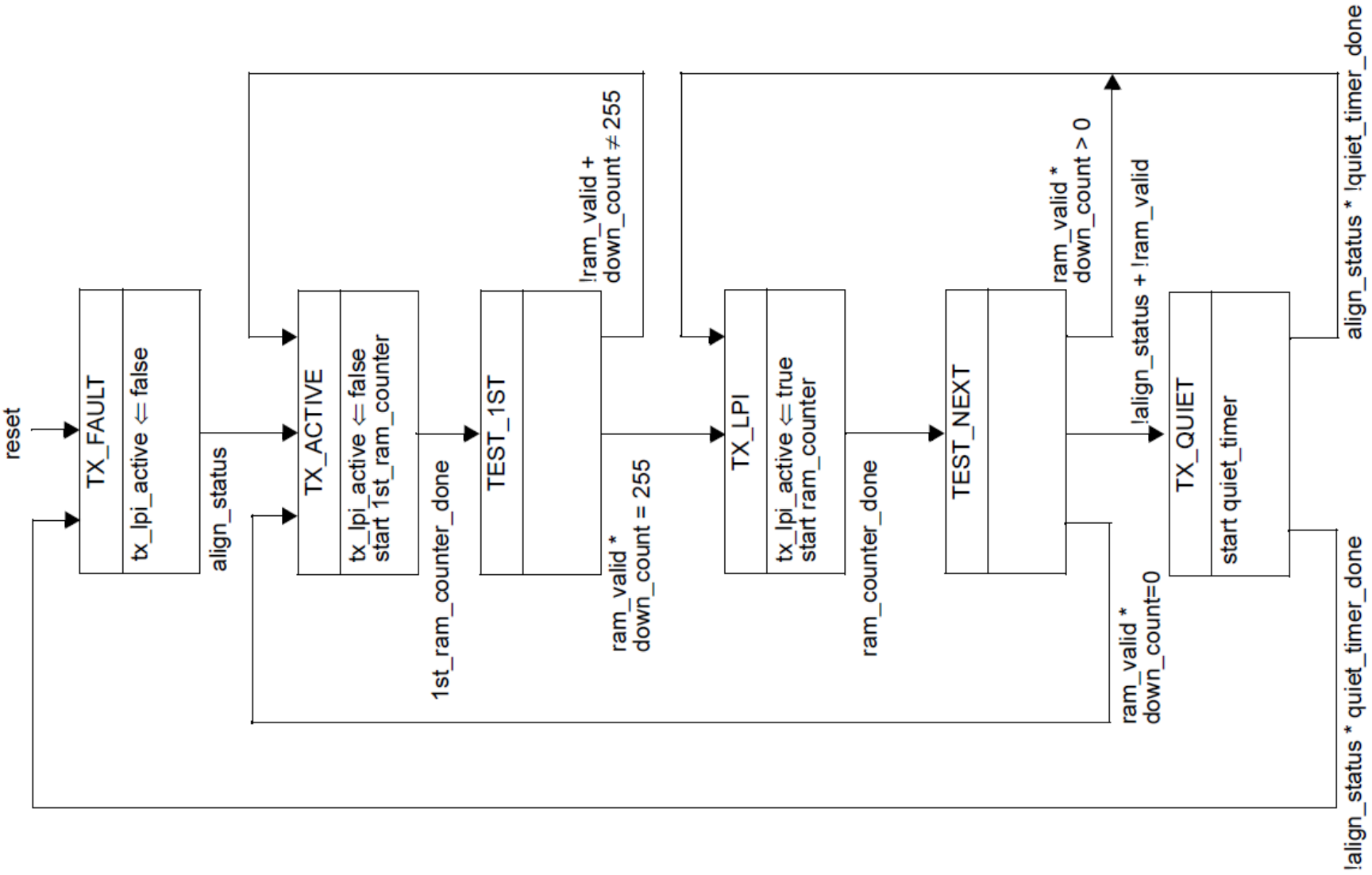
Principle of operation

- RS-FEC infers tx_lpi_active and rx_lpi_active
  - Asserted when RAM (down_count=255) detected
  - Only valid positions (per #195) checked
  - Verify across multiple lanes for robustness
  - De-asserted when down_count=0 detected

- When tx_lpi_active (or rx_lpi_active) is asserted...
  - Map RAMs at detected position/intervals
  - Enable rapid synchronization algorithm

- If quiescent mode supported, when tx(rx)_lpi_active asserted
  - Start timer when fec_align_status is deasserted
  - If fec_align_status reasserted before timer fails, no change
  - Otherwise, tx(rx)_lpi_active deasserted (link disconnected)

Proposed Transmit LPI state diagram



For the best viewing experience, please rotate your screen by 90º

Variables

1st_ram_counter_done

Boolean variable that indicates that 1st_ram_counter has reached its terminal count.

down_count

The value that results from the bit-wise exclusive-OR of the Count Down (CD_3) byte and M_0 byte of the current rapid alignment marker (see 82.2.8a).

tx_lpi_active

A Boolean variable that is set to true when the RS-FEC sublayer infers that the PCS is transmitting Low Power Idle and is set to false otherwise.

quiet_timer_done

Boolean variable that indicates that quiet_timer has reached its terminal count.

More variables

ram_valid

Boolean variable that is set to true when the 66-bit blocks concurrently received on at least TBD PCS lanes are valid alignment markers and is set to false otherwise.

ram_counter_done

Boolean variable that indicates that ram_counter has reached its terminal count.

Counters and timers

1st_ram_counter

This counter counts 4 66-bit blocks form the end of one candidate RAM position to the end of the next candidate RAM position. The first instance of the counter counts from the end of the last normal alignment marker received.

ram_counter

This counter counts 8 66-bit blocks form the end of the current RAM to the end of the next expected RAM position.

quiet_timer

The timer limits the maximum time align_status may be deasserted before the Transmit LPI state diagram concluded the link has failed [must be significantly longer than tx_tq_timer].

Completing the changes

- Add Receive LPI state diagram for rx_lpi_active

- Similar to Transmit LPI state diagram except…
  - …based on FEC lanes and not PCS lanes and…
  - …considers alignment maker payload sequences

- Draft refers to alternate behavior “for the optional EEE capability”
  - amp_valid, AMP_COMPARE, am_counter
  - Alignment marker insertion/removal (see 91.5.2.6 and 91.5.3.4)
  - Change to “when tx(rx)_lpi_active asserted” as appropriate

- Primitive FEC:IS_RX_LPI_ACTIVE.request is not required

Topics

- Skew and Skew Variation (26, 222, 235, 409)
- Energy Efficient Ethernet (95, 99, 100, 242, 243, 208, 210)
- Exception handling (53, 162, 239)
- Management (196, 197, 244)

Exception handling: RS-FEC transmit function

- tx_coded_*j*<1> = tx_coded_*j*<0> not handled
  - Invalid synchronization header

- Some suggestions…
  a) Substitute EBLOCK_T for tx_coded_*j*, but requires scrambling
  b) tx_xcoded<*j*+1> = 1, apply error mask to tx_coded_*j*<65:2>
  c) tx_xcoded<*j*+1> = 0, apply error mask to tx_coded_*j*<65:2>

- Recommendation: b) or c)
  - Error mask to be defined

Exception handling: RS-FEC receive function

- Failure to map g<3:0> to h<3:0> not handled
  - Invalid block type field

- Recommendation
  - Set h<3:0> to 0x0
  - Set rx_coded_*c*<1> = rx_coded_*c*<0> = 0

Loss of signal

- Consider FEC:IS_UNITDATA.indication(rx_bit)

- How to set rx_bit when SIGNAL_OK is FAIL?
  – This is a consideration for CAUI

- Some suggestions…
  a) EBLOCK_R, but requires scrambling
  b) LOCAL FAULT, but requires scrambling
  c) Demultiplex (4:20) rx_bit from PMA [1]

- Recommendation: c), similar to Clause 74

[1] Must discard every 34th bit from each 100GBASE-KP4 PMA lane to reconcile PMA and FEC rates.

Topics

- Skew and Skew Variation (26, 222, 235, 409)
- Energy Efficient Ethernet (95, 99, 100, 242, 243, 208, 210)
- Exception handling (53, 162, 239)
- Management (196, 197, 244)

Use Clause 74 as a template

Table 74–1—MDIO/FEC variable mapping

MDIO variable	PMA/PMD register name	Register/ bit number	FEC variable
BASE-R FEC ability	BASE-R FEC ability register	1.170.0	FEC_ability
BASE-R FEC Error Indication ability	BASE-R FEC ability register	1.170.1	FEC_Error_Indication_ability
FEC Enable	BASE-R FEC control register	1.171.0	FEC_Enable
FEC Enable Error Indication	BASE-R FEC control register	1.171.1	FEC_Enable_Error_to_PCS
FEC corrected blocks	10GBASE-R FEC corrected blocks counter register	1.172, 1.173	FEC_corrected_blocks_counter
FEC uncorrected blocks	10GBASE-R FEC uncorrected blocks counter register	1.174, 1.175	FEC_uncorrected_blocks_counter
FEC corrected blocks, lanes 0 through 19	BASE-R FEC corrected blocks counter register, lanes 0 through 19	1.300 through 1.339	FEC_corrected_blocks_counter_*i*
FEC uncorrected blocks, lanes 0 through 19	BASE-R FEC uncorrected blocks counter register, lanes 0 through 19	1.700 through 1.739	FEC_uncorrected_blocks_counter_*i*

RS-FEC is not optional

RS-FEC encoding is always transmitted

Replace with a symbol error count for each FEC lane

Derived from comment #196

- **91.6.1 FEC_corrected_blocks_counter**

  FEC_corrected_blocks_counter is a 32-bit counter that counts once for each corrected FEC block processed when fec_align_status is true. This variable is mapped to the registers defined in 45.2.1.TBD (1.TBD, 1.TBD).

- **91.6.2 FEC_uncorrected_blocks_counter**

  FEC_uncorrected_blocks_counter is a 32-bit counter that counts once for each uncorrected FEC block processed when fec_align_status is true. This variable is mapped to the registers defined in 45.2.1.TBD (1.TBD, 1.TBD).

- **91.6.3 FEC_symbol_error_counter_*i***

  FEC_symbol_error_counter_*i*, where i=0 to 3, is a 32-bit counter that counts once for each 10-bit symbol corrected on FEC lane *i* when fec_align_status is true. This variable is mapped to the registers defined in 45.2.1.TBD (1.TBD to 1.TBD).

Related to comment #197

- **91.6.4 BIP_error_counter_*i***

  BIP_error_counter_*i*, where i=0 to 19, is a 16-bit counter that holds the BIP error count for PCS lane *i* as calculated by the RS-FEC transmit function (see 91.5.2.4). This variable is mapped to the registers defined in 45.2.1.TBD (1.TBD, 1.TBD).

- **91.5.2.4 Alignment marker removal**

  Replace TBD with the register numbers defined above.

Pending the resolution of #241

- **91.6.5 FEC error indication ability**

  The RS-FEC sublayer may have the option to indicate decoding errors to the PCS sublayer by intentionally corrupting 66-bit block synchronization headers as defined in 91.5.3.3. This variable is set to one to indicate that the RS-FEC sublayer has the ability to indicate decoding errors to the PCS sublayer. The variable is set to zero if this ability is not supported. This variable is mapped to the bit defined in 45.2.1.TBD (1.TBD).

- **91.6.6 FEC error indication enable**

  This variable is set to one to enable indication of decoding errors to the PCS sublayer (see 91.5.3.3) when this feature is supported. When set to zero, the error indication function is disabled. This variable is mapped to the bit defined in 45.2.1.TBD (1.TBD).

Pending the resolution of #186

- **91.6.7 FEC bypass correction ability**

  The Reed-Solomon decoder may have the option to perform error detection without error correction (see 91.5.3.3) to reduce the delay contributed by the RS-FEC sublayer. This variable is set to one to indicate that the decoder has the ability to bypass error correction. The variable is set to zero if this ability is not supported. This variable is mapped to the bit defined in 45.2.1.TBD (1.TBD).

- **91.6.8 FEC bypass correction enable**

  When this variable is set to one the Reed-Solomon decoder performs error detection without error correction (see 91.5.3.3). When this variable is set to zero, the decoder also performs error correction. This variable is mapped to the bit defined in 45.2.1.TBD (1.TBD).

Clause 45

- RS-FEC to use MMD 1 register space

- Potential overlap with Clause 74 registers
  - ability/status (45.2.1.89), add FEC correction bypass ability?
  - control (45.2.1.90), add FEC correction bypass enable?
  - FEC_corrected_blocks_counter (45.2.1.91)
  - FEC_uncorrected_blocks_counter (45.2.1.92)

- New registers
  - BIP_error_counter_*i* (1 register per PCS lane = 1 x 20)
  - FEC_symbol_error_counter (2 registers per FEC lane = 2 x 4)
  - No contiguous address space (1.107+)

- Recommend contiguous address space starting at 1.200