Details and possible resolutions for fault signaling comments

(In support of comments #13, #14 against D2.0)

Adee Ran, Cisco

Acknowledgements

Support

- Matt Traverso, Cisco

Contribution

- Piers Dawe
- Eric Maniloff

Comments addressed by this presentation (against D2.0)

Cl **171** *SC* **171.4** *P* **193** *L* **42** # 13

Ran, Adee Cisco

Comment Type **T** *Comment Status* **D** *fault signaling*

The standard should be explicit about what happens in a PHY connected to an 800GMII Extender when there is no valid input signal.

The precedence is set in 802.3cw: D2.1 states (in 155.2.6.7.2) that the 400GBASE-ZR PCS sends local fault ordered sets to the 400GMII when there is no signal; this means the PHY XS transmits these local fault over the 400GAUI-n toward the DTE XS. There is no provision for "shutting down" the PHY XS output, so the 400GAUI-n in an Extender is never silent.

The behavior of the 800GMII extender should be the same as that of the 400GMII extender as described above.

Note that this behavior is different from existing optical modules that are connected with any AUI-C2M to a PCS (as part of the PHY, not an extender), where it is common to squelch the module electrical output (aka disable the AUI's transmitter) when there is no optical input (PMD:IS_SIGNAL.indication is not_ok); that is indicated to by PCS via PMA:IS_SIGNAL.indication on its adjacent PMA. That would not be compliant behavior when the AUI is within a 800GMII Extender.

The different behavior required from Extender modules may not be obvious and should be mentioned.

Note: if the task force wants to allow squelching the Extender's AUI, it may require more significant changes; as an alternative, an editor's note can be added to capture that intent until a detailed proposal is presented
(such as "Editor's note: the behavior of the Extender when there is no input signal from the PHY is to be determined").

SuggestedRemedy

Add the following paragraph at the end of 171.4:

NOTE-link fault signaling generated by the PHY (see 170.3 and 81.3.4) is transmitted to the RS through the 800GMII Extender. Therefore, the electrical interface used within the 800GMII Extender sends valid PHY 800GXS data regardless of the link state of the PHY below the 800GMII.

Cl **173** *SC* **173.4.8.3** *P* **236** *L* **19** # 14

Ran, Adee Cisco

Comment Type **T** *Comment Status* **D** *fault signaling*

"Otherwise the SIL reports the signal status as FAIL"

In the case of 8:8 PMA, this FAIL status typically indicates that data is not being received on all 8 input lanes (inst:IS_UNITDATA_0:7.indication). When this happens, the data on the output lanes (PMA:IS_UNITDATA_0:7.indication) cannot be determined from the standard. Apparently it is unspecified, but it isn't stated explicitly.

In optical modules (a common implementation of PMAs similar to this one), the typical behavior is to turn off the electrical output of the AUI-C2M; but this functionality is not specified in the standard, and there is no specification of "output disabled" in 120G.3.2. It can be argued that this common behavior is non-compliant.

With no specification of behavior in this condition, the signal status is not conveyed to the PMA client (host ASIC) in a specified and consistent manner. Moreover, SerDes designers cannot assume what signal appears on the AUI when there is no input, and that is a repeating source of confusion, often leading to bad design or unnecessary over-design.

We need to specify the AUI behavior when signal status is FAIL such that the PMA client can detect this situation. Based on existing module behavior, it is suggested to state that a PMA with a physically instantiated interface disables the transmitters on all lanes of that interface when signal status is FAIL on the other interface, for some minimum time. The PMA client can infer the status by detecting that its input signal corresponds to a disabled transmitter. This requires adding the missing "output disabled" mode in the module output characteristics (120G.3.2).

A possible alternative is to allow the PMA to transmit the PRBS31Q test pattern (120.5.11.2.2), if implemented, instead of disabling the transmitter. The PMA client can then infer the link status by detecting that its input corresponds to a PRBS31Q test pattern. This would not require adding "output disabled" mode, but it is likely not the existing behavior, and would be more disruptive.

Note that this isn't just an 802.3df problem (ambiguity of the module output is a long-standing issue), but since we are defining a new PMA it is a good opportunity to close this gap.

SuggestedRemedy

Add the following paragraph at the end of 173.4.8.3:

"When the signal status is FAIL, an 8:8 PMA shall disable the output on all lanes of its
120G–3. When the module output is disabled, the Differential peak-to-peak output voltage
shall be less than 35 mV."<paragraph break>

Change the title of Table 120G-3 to "Module output characteristics in enabled state at TP4"
"The module output shall meet the specifications given in Table 120G–3"
to
"When the module output is enabled, it shall meet the specifications given in Table

Outline

- The issue with signal indication in modules
- New developments in 802.3cw
- Proposed remedy

Problem statement

- When there is no signal, the received data (bit stream) is undefined!
- Receiver logic should be informed of this situation
  - To prevent unexpected behavior (e.g., adaptive parameters diverging, timing violations)
  - To enable quick response when a signal does appear
  - To save power
- In 802.3, signal indication interface primitive is defined in all PMDs
  - But what do you do with it?
- 802.3 has largely been silent about this topic, leaving implementers to handle the situation any way they see fit
  - From a SerDes design point of view, this is unfortunate

Signal indication and SIL – history

- Primitive and description of SIL goes back as far as clause 51 (10G Serial PMA)
  - Primitive semantics are similar to what we have on all later PMAs
  - SIL had a dedicated (optional) signal on the XSBI PMA service interface – as shown Figure 51-3

51.2.3 PMA_SIGNAL.indication

This primitive is sent by the PMA to its client to indicate the status of the receive process. PMA_SIGNAL.indication is generated by the PMA receive process to propagate the detection of severe error conditions (e.g. no valid signal being received from the PMD sublayer) to the PMA client.

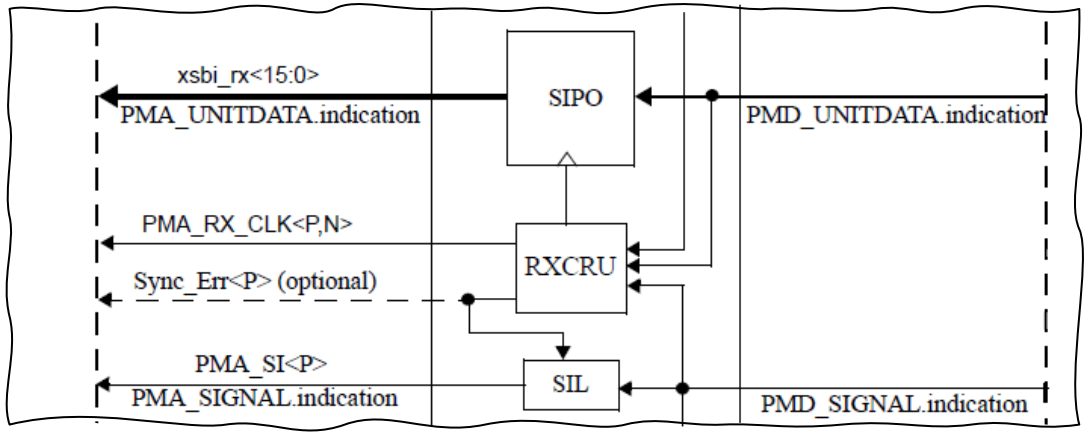
51.2.3.1 Semantics of the service primitive

PMA_SIGNAL.indication (SIGNAL_OK)

The SIGNAL_OK can take one of two values: OK or FAIL. A value of FAIL denotes that invalid data is being presented to the PMA client. A value of OK does not guarantee valid data is being presented to the PMA client.

51.2.3.2 When generated

The PMA generates a PMA_SIGNAL.indication primitive to the PMA client whenever there is change in the value of the SIGNAL_OK parameter.



Excerpt from Figure 51-3

Signal indication and SIL – present

- More recent PMA specifications, starting in Clause 83, still have the IS_SIGNAL.indication primitive, but do not specify a physical signal for it
  - Physical instantiations of the service interface (collectively known as AUIs) do not include it
- The signal indication logic (SIL) is specified very vaguely
- AUI specifications do not address the case where the PMA has no valid data to send

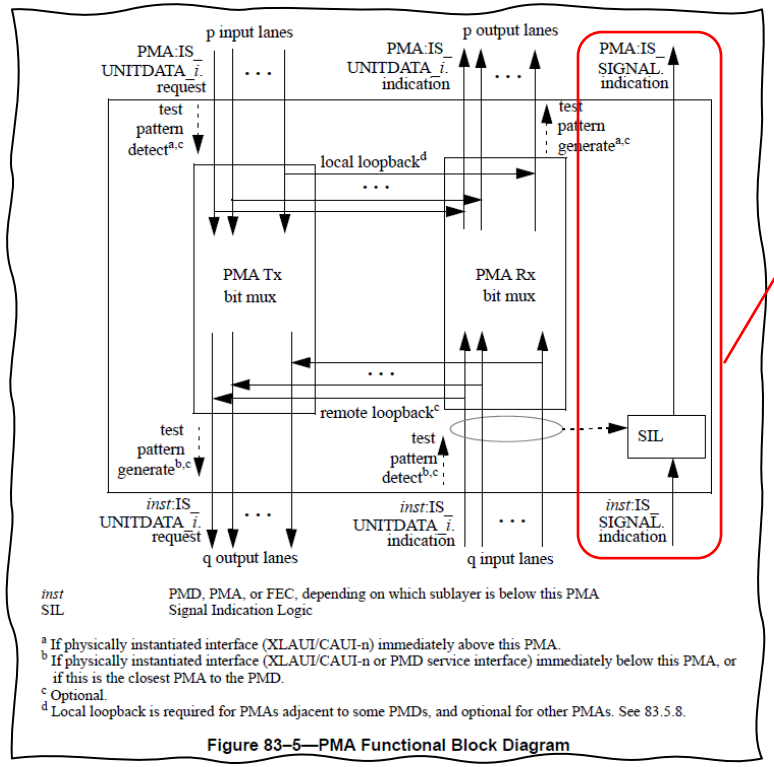


Figure 83–5—PMA Functional Block Diagram

Either the PMA service interface or the interface below is always physically instantiated (AUI)

IS_SIGNAL indication is not propagated!

83.5.7 Link status

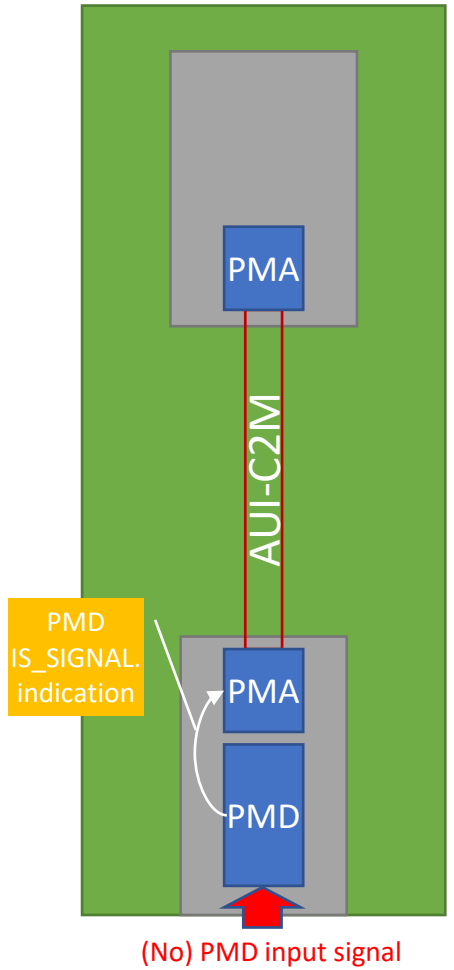
The PMA shall provide link status information to the PMA client using the PMA:IS_SIGNAL.indication primitive. The PMA continuously monitors the link status reported by the service interface below from the *inst*:IS_SIGNAL.indication primitive, and uses this as input to Signal Indication Logic (SIL) to determine the link status to report to the layer above. Other inputs to the SIL may include the status of clock and data recovery on the lanes from the service interface below the PMA and whether buffers/FIFOs have reached the required fill level to accommodate Skew Variation so that data is being sent on the output lanes.

What happens in practice when the host has no data to send

- PCS output is always defined, even in reset (LBLOCK_T=local fault)
  - But 82.2.19.3 says “Though the Transmit state diagram sends Local Fault ordered sets **when reset is asserted**, the scrambler may not be operational during reset. Thus, **the Local Fault ordered sets may not appear on the PMA/FEC service interface**”
- The PMA is supposedly always functional, but:
  - When the PCS is held at reset, there is no valid data that the PMA can transmit
  - In this situation, the AUI (interface below the PMA) is typically *squelched* (output disabled)
    - **AUI-C2M Annexes starting in 83E, including 120G, have a specification for “transmitter disabled” on host output**
    - This is useful as a way to “reset” the module’s receiver, although it is not specified

What happens in practice when the module has no incoming optical signal

- The module interface is AUI-C2M (starting with the original XLAUI and CAUI-10), connecting two PMAs
  - The bottom PMA is in the module and has a (mandatory) specified signal indication logic (SIL) to generate SIGNAL_OK (the PMA:IS_SIGNAL.indication primitive parameter, see 120.3)
  - The SIL can use the (mandatory) PMD signal detect function (architecturally, through the SIGNAL_OK received from the PMD:IS_SIGNAL.indication primitive)
  - But it can't communicate SIGNAL_OK to the upper PMA through AUI-C2M
- In practice, many modules squelch their AUI-C2M output to indicate that there is no valid signal (SIGNAL_OK=FAIL)
  - CMIS (7.6.1.3) calls this behavior "automatic squelching" and defines it as mandatory (with an option to disable it)
  - … but 802.3 doesn't specify that, and the AUI-C2M Annexes do not have a provision to silence the module output (unlike the host output)
  - Arguably, squelching is incompliant with AUI-C2M module output specifications!
- **This inconsistency between 802.3 specifications and industry practice should be fixed**



Host and module output characteristics

Table 120G–1—Host output characteristics at TP1a

Parameter	Reference	Value	Units
Signaling rate, each lane (range)		53.125 ± 50 ppm[a]	GBd
DC common-mode output voltage (max)	120G.5.1	2.8	V
DC common-mode output voltage (min)	120G.5.1	–0.3	V
Single-ended output voltage (max)	120G.5.1	3.3	V
Single-ended output voltage (min)	120G.5.1	–0.4	V
Peak-to-peak AC common-mode voltage (max) Low-frequency, VCM_{LF} Full-band, VCM_{FB}	120G.5.1	 32 80	mV
Differential peak-to-peak output voltage (max) Transmitter disabled Transmitter enabled	120G.5.1	 35 750	mV
Steady-state voltage, v_f (max)	120G.5.3	375	mV
Eye height (min)	120G.3.1.5	10	mV
Vertical eye closure, VEC (max)	120G.3.1.5	12	dB
Common-mode to differential-mode return loss, *RLdc* (min)	120G.3.1.1	Equation (120G–1)	dB
Effective return loss, ERL (min)	120G.3.1.2	7.3	dB
Differential termination mismatch (max)	120G.3.1.3	10	%
Transition time (min) Host is requesting short mode[b] Host is requesting long mode[b]	120G.3.1.4	 10 15	 ps ps

[a] For a PMA in the same package as the PCS sublayer. In other cases, the signaling rate is derived from the signaling rate presented to the PMA input lanes (see Figure 135–3 and Figure 120–3) by the adjacent PMA or FEC sublayers.
[b] Short and long modes are defined in 120G.3.2.1.

"Transmitter disabled" is defined

Table 120G–3—Module output characteristics at TP4

Parameter	Reference	Value	Units
Signaling rate, each lane (nominal)		53.125[a]	GBd
Peak-to-peak AC common-mode voltage (max) Low-frequency, VCM_{LF} Full-band, VCM_{FB}	120G.5.1	 32 80	 mV
Differential peak-to-peak output voltage (max) Short mode Long mode	120G.5.1	 600 845	 mV mV
Eye height (min)	120G.3.2.2	15	mV
Vertical eye closure, VEC (max)	120G.3.2.2	12	dB
Common-mode to differential-mode return loss, *RLdc* (min)	120G.3.1.1	Equation (120G–1)	dB
Effective return loss, ERL (min)	120G.3.2.3	8.5	dB
Differential termination mismatch (max)	120G.3.1.3	10	%
Transition time (min)	120G.3.1.4	8.5	ps
DC common-mode voltage tolerance (range) Upper limit Lower limit	120G.3.2.4	 2.85 -0.35	 V V

[a] The signaling rate range is derived from the PMD receiver input.

No specification for a disabled Transmitter

New development in P802.3cw

- In D2.1 of P802.3cw, 155.2.6.7.2 includes the following paragraph:

In the case of a DSP framing loss or 400GBASE-ZR frame or multi-frame loss, the PCS receive path inserts a stream of 257-bit blocks carrying local fault ordered sets.

- So, in a 400GBASE-ZR module:
  - The PHY XS (client of the PCS) always receives data; if there is no signal, the data is local fault (LF) ordered sets.
  - The PMA within the 400GMII Extender always has SIGNAL_OK=OK.
  - To enable the host (DTE XS) to receive the LF's, the AUI-C2M should not be squelched when there is no signal.
- This behavior is different from other modules (that do not include a PCS)
  - In some respects, it is better than squelching
  - A comment has been submitted to P802.3cw to add a note highlighting this difference
- Assuming this requirement is retained in P802.3cw, for consistency, the same behavior should be specified in P802.3df for modules that include a PCS and PHY XS.
  - Note that it requires the capability to generate local faults inside the module.

Cl **155** *SC* **155.2.6.7.2** *P* **53** *L* **42** # 71

Ran, Adee Cisco

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

The standard should be more explicit about what happens in a PHY connected to a 400GMII Extender when there is no input signal.

The text here suggests that the PCS sends local fault to the 400GMII; this means the PHY XS should be able to generate local fault signaling over the 400GAUI-n toward the DTE XS. Moreover, there is no IS_SIGNAL.indication across the 400GMII. Apparently it means that the 400GAUI-n in an Extender should never be silent.

In existing optical modules that are connected with any AUI-C2M to a PCS (as part of the PHY, not an extender), it is common to squelch the module electrical output (aka disable the AUI's transmitter) when there is no optical input (PMD:IS_SIGNAL.indication is not_ok); that is indicated to by PCS via PMA:IS_SIGNAL.indication on its adjacent PMA. That would not be compliant behavior when the AUI is within an XS.

Ignoring this detail may lead to "surprising" module implementations that squelch the module's output when there is no input, and may create interoperability issues with hosts that stick to the standard.

SuggestedRemedy

Assuming this is the intent, please add a NOTE emphasizing that the adjacent PHY 400GXS generates PHY_XS:IS_UNITDATA.indication and does not squelch the 400GAUI-n even when PMA_IS_SIGNAL.indication is FAIL.

How to fix it?

- In P802.3df, we should do the following:
  - For modules that are part of a type 1 or type 2 PHY (800GBASE-R PMD + PMA with 800GAUI-8 C2M interface):
    - **Allow, and possibly recommend, turning off the 800GAUI-8 C2M output to indicate SIGNAL_OK=FAIL; specify a minimum time of output off, to enable the host to detect it and reset its receiver**
    - This change relaxes the specification by allowing new behavior that is already implemented in deployed modules; therefore, it does not break compliance of existing modules with Annex 120G
  - For modules that are full PHYs (containing a PHY 800GXS with 800GAUI-8 C2M interface):
    - **Option A: Specify that the PHY indicates SIGNAL_OK=FAIL by sending local fault ordered sets, and never turns off the 800GAUI-8 C2M output**
    - **Option B: Allow "full PHY" modules too to disable their output to indicate SIGNAL_OK=FAIL (as an alternative to sending ordered sets, or after some time)**
- If such changes to 200GAUI-2, 400GAUI-4 C2M, and 200G/400GXS are considered in scope, apply them too; otherwise, it can be done in maintenance
- In the future we should have similar specifications into P802.3dj (and make them mandatory).

Proposal, part 1 (for “partial PHY” modules)

- Change 173.4.8.3 as follows:

173.4.8.3 8:8 PMA signal status

The 8:8 PMA contains a signal indication logic (SIL) function that continuously monitors the quality of the signal from the sublayer below the PMA (see 173.3 and Figure 173–5).

The SIL function reports the signal status as OK when all of the following conditions are met:

— data is being received on all 8 input lanes (*inst*:IS_UNITDATA_0:7.indication)

— data is being sent on all 8 output lanes (PMA:IS_UNITDATA_0:7.indication)

— the SIGNAL_OK parameter of the PMD:IS_SIGNAL.indication primitive is set to OK, if there is a PMD immediately below the PMA

Otherwise the SIL reports the signal status as FAIL.

If the 8:8 PMA service interface or the interface below it is physically instantiated (800GAUI-8 C2C or C2M), the PMA may indicate SIGNAL_OK=FAIL by disabling the output of some or all lanes of the physically instantiated interface. If the output is disabled, it shall remain disabled for a minimum of 50 ms.

- Add 120G.3.2 to the draft, and change its first paragraph as follows:

120G.3.2 Module output characteristics

~~The module output~~ When the module output is enabled, it shall meet the specifications given in Table 120G–3. When the module output is disabled, the differential peak-to-peak output voltage shall be less than 35 mV.

The low-frequency 3 dB cutoff of the output AC-coupling within the module should be less than 100 kHz. A test system with a fourth-order Bessel-Thomson low-pass response with 40 GHz 3 dB bandwidth is to be used for all output signal measurements, unless otherwise specified.

- Change the title of Table 120G-3 to "Module output characteristics at TP4 in enabled state“.

To be implemented with editorial license

Proposal, part 2 (for “full PHY” modules): option A

- Change 171.4 as follows:

171.4 800GAUI-n

An 800GMII Extender may use any of the following electrical interfaces for the connection between its PMA sublayers, as shown in Figure 171–1:

— 800GAUI-8 chip-to-chip (Annex 120F)

— 800GAUI-8 chip-to-module (Annex 120G)

NOTE—When a PHY connected to a PHY 800GXS has no valid received signal, local fault ordered sets (see 170.3 and 81.3.4) are transmitted to the RS through the 800GMII Extender. The electrical interface used within the 800GMII Extender carries valid PHY 800GXS output regardless of the link state of the PHY.

To be implemented with editorial license

Proposal, part 2 (for “full PHY” modules): option B

- Change 171.3.2 as follows:

171.3.2 PHY 800GXS service interface

The PHY 800GXS service interface primitives are summarized as follows:

PHY_XS:IS_UNITDATA_*i*.request(tx_symbol)
PHY_XS:IS_UNITDATA_*i*.indication(rx_symbol)
PHY_XS:IS_SIGNAL.request(SIGNAL_OK)
PHY_XS:IS_SIGNAL.indication(SIGNAL_OK) (optional)

and update Figure 171–2 to include the PHY_XS:IS_SIGNAL.indication primitive.

- Insert a new subclause after 171.4, as follows (renumbering subsequent subclauses as necessary):

171.5 Link fault signaling

Link fault signaling generated by the PHY (see 170.3 and 81.3.4) is transmitted to the RS through the 800GMII Extender. A link fault condition may optionally be communicated to the client sublayer through the service interface primitive PHY_XS:IS_SIGNAL.indication with SIGNAL_OK=FAIL.

To be implemented with editorial license

Proposal, part 3 (200GAUI-2, 400GAUI-4 C2M, and 200G/400GXS)

- Apply part 1 also in 120.3 by changing the final paragraph as follows:

The PMA:IS_SIGNAL.indication primitive is generated through a set of Signal Indication Logic (SIL) that reports signal health based on receipt of the *inst*:IS_SIGNAL.indication from the sublayer below, data being received on all of the input lanes from the sublayer below, buffers filled (if necessary) to accommodate Skew Variation, and symbols being sent to the PMA client on all of the output lanes. When these conditions are met, the SIGNAL_OK parameter sent to the PMA client via the PMA:IS_SIGNAL.indication primitive has the value OK. Otherwise, the SIGNAL_OK primitive has the value FAIL. In the case where the sublayer below the PMA is a PHY XS the PMA ~~does~~ might not receive a PHY_XS:IS_SIGNAL.indication as an input to the SIL. If the PMA service interface or the interface below it is physically instantiated (400GAUI-4 C2C or C2M or 200GAUI-2 C2C or C2M), the PMA may indicate SIGNAL_OK=FAIL by disabling the output of some or all lanes of the physically instantiated interface. If the output is disabled, it shall remain disabled for a minimum of 50 ms.

- Apply part 2 also in clause 118 (according to the selected option, A or B).

To be implemented with editorial license, if considered in scope