PHY error rate monitoring for 40GbE and 100GbE

Gary Nicholl

Cisco Systems

IEEE 802.3ba, May 2008 Interim, Munich

The setup …….

PHY (PCS) level bit error rate monitoring:

- **ability to continuously monitor (estimate) the BER on an Ethernet link.**
- **independent of the traffic on the link**
- **can almost do this today at 10G by using the 802.3ae PCS error counters**
- **BUT .. need a slight tweak**

Background and Requirements

- Customers would like to have the ability to continuously monitor (estimate) the bit error rate on an Ethernet link, and set a user defined threshold with which to drive proactive notification and/or consequent action (e.g. take the link down, trigger restoration).
- Many customers are used to having error rate thresholds that are user settable in the range from 10-12 to 10-4.
- This will likely become more important to customers as we move to higher speeds and longer distances (40km,etc).
- The error monitoring needs to be completely independent of the traffic being carried on the link (i.e. cannot use CRC errors or any frame based statistics)
- PHY error counters need to be suitable sized to allow host software to poll them at a reasonable rate (say once per second)

Proposed solution for 40GbE/100GbE

- The error rate on the PHY can be easily monitored by counting invalid 64B66B sync headers (i.e. 00 or 11).
- Counting the invalid sync headers provides a sampled view (2/66) of the errors on the link
- Assuming the errors are random and Gaussian in nature, the total error count on the link (and thus the BER) can be estimated by simply multiplying the invalid sync header count by 33.
- A 24 bit counter for invalid sync headers, would allow host software to poll the PHY at a 1 second rate and without overflowing for an error rate of up to 10-4 on a 100G link.
- Invalid sync headers can also be monitored on individual virtual lanes as well (virtual lanes are 66 bit structured). This might prove to be useful information for debugging and troubleshooting purposes.

What about 10GbE today ?

- But isn't something like this already provided at 10GbE with the 802.3ae PCS error counters (Reg 3.33 .. see backup) ? Yes and no. There are issues with both of the PCS counters in 802.3ae:
  - the 'errored block' counter, counts 'error'ed code words (blocks) and not just invalid sync headers. This makes it impossible to estimate the actual BER (for a given error rate the error count depends on the actual traffic pattern)
  - the 'BER counter' appears to only count invalid sync headers (?), but the counter is too small at 6 bits.
- But isn't something like this already provided with Ethernet Link OAM in 802.3ah ?
  - same problem as above, as 802.ah leverages the same PCS error coutners

Recommendation

- **PCS should include a single counter to count the total number of invalid sync headers at the aggregate level, and sized to be at least 24 bits**
- **PCS should also include invalid sync header monitoring and reporting at the virtual lane level (to be used primarily for debugging and troubleshooting purposes)**

Backup

10G BASE-R PCS Error Counters

Table 45–39—10GBASE-R PCS status 2 register bit definitions

Bit(s)	Name	Description	R/W[a]
3.33.15	Latched block lock	1 = 10GBASE-R PCS has block lock 0 = 10GBASE-R PCS does not have block lock	RO/LL
3.33.14	Latched high BER	1 = 10GBASE-R PCS has reported a high BER 0 = 10GBASE-R PCS has not reported a high BER	RO/LH
3.33.13:8	BER	BER counter	RO/NR
3.33.7:0	Errored blocks	Errored blocks counter	RO/NR

[a]RO = Read Only, LL = Latching Low, LH = Latching High, NR = Non Roll-over

45.2.3.12.3 BER(3.33.13:8)

The BER counter is a six bit count as defined by the ber_count variable in 49.2.14.2. These bits shall be reset to all zeros when the BER count is read by the management function or upon execution of the PCS reset. These bits shall be held at all ones in the case of overflow.

45.2.3.12.4 Errored blocks (3.33.7:0)

The errored blocks counter is an eight bit count defined by the errored_block_count counter specified in 49.2.14.2. These bits shall be reset to all zeros when the errored blocks count is read by the management function or upon execution of the PCS reset. These bits shall be held at all ones in the case of overflow.

10G BASE-R PCS Error Counters

49.2.14.2 Counters

The following counters are reset to zero upon read and upon reset of the PCS. When they reach all ones, they stop counting. Their purpose is to help monitor the quality of the link.

ber_count:

6-bit counter that counts each time BER_BAD_SH state is entered. This counter is reflected in MDIO register bits 3.33.13:8. Note that this counter counts a maximum of 16 counts per 125 μs since the BER_BAD_SH can be entered a maximum of 16 times per 125 μs window.

errored_block_count:

8-bit counter. When the receiver is in normal mode, errored_block_count counts once for each time RX_E state is entered. This counter is reflected in MDIO register bits 3.33.7:0.

test_pattern_error_count:

16-bit counter. When the receiver is in test-pattern mode, the test_pattern_error_count counts errors as described in 49.2.12. This counter is reflected in MDIO register bits 3.43.14:0.

10G BASE-R PCS Error Counters

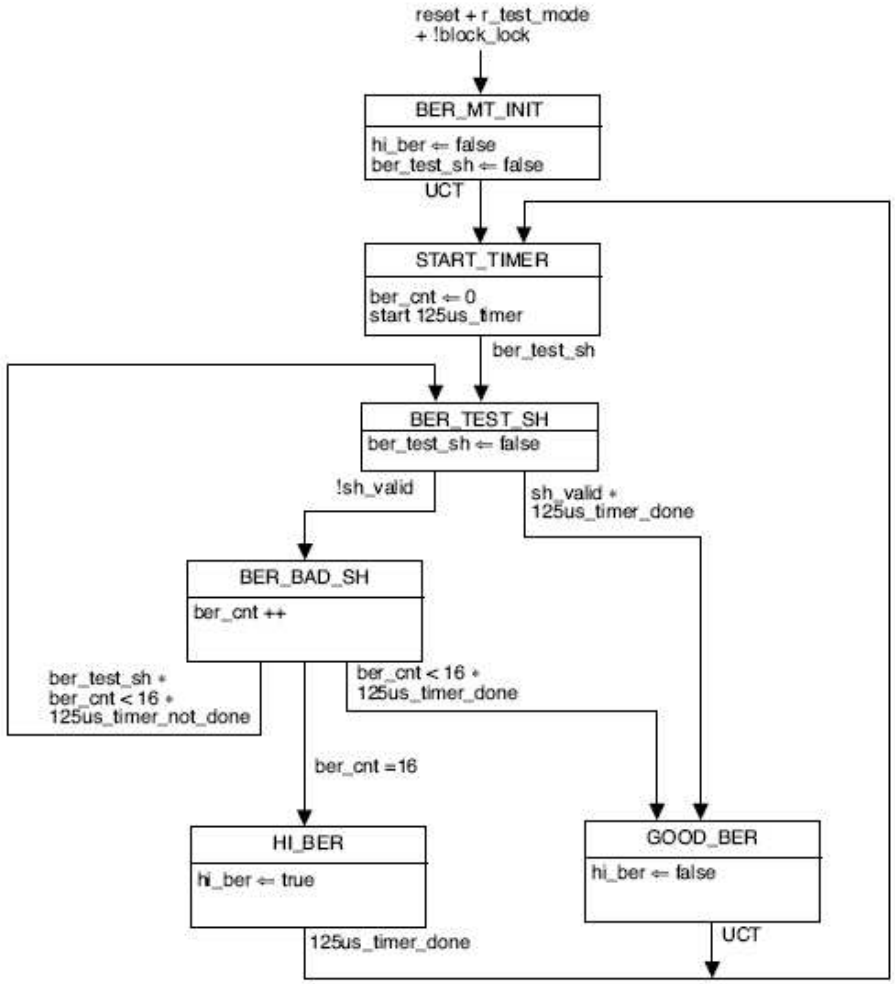

Figure 49–13—BER monitor state machine