Bit Error Ratio determination by failing TCP/IP transfers.
or
Stress your optical link and count your errors.

Erwin Kok
Henk Z. Peek
Ruud van Wijk

NIKHEF, Amsterdam, The Netherlands

February 2003

Abstract

The performance of Gigabit Ethernet physical layer components in the 40 km offshore data and communication connection of Antares largely depends on its signal to noise ratio at the receiver input of the used DWDM boards. This report describes a method to derive a figure of merit for these homemade produced DWDM boards. This figure of merit is expressed in terms of an estimated Bit Error Ratio at a distinct optical receiver input level. The presented method is also sensitive for the difference in nature of the detected communication errors that is it discriminates thermal noise errors from a possible presence of system errors.

Contents

1. Preamble

Antares physics data, accumulated in the deep sea, is transferred to shore over a distance of 40 km with help of Giga bit Ethernet data links. Multiple optical GbE links are combined into a single DWDM system, using different wavelengths for each of the links.

The inaccessibility of a sub-marine system for recovery operations requires a ten years lifetime expectation of the instrumentation. It is recommendable that during the production phase of the DWDM physical layer components a Quality of Signal number is determined for future system integrity statements. Unfortunately, a straightforward QoS number determination in a high quality system requires very long test times and is therefore in practice difficult to obtain.

A reasonable estimate of the QoS figure can be achieved by testing the produced components in a so-called reference test system. Herein the Bit Error Ratio of the GbE data transfer is measured by stressing the optical link, the reported Retransmit Request rate of the GbE link is then observed in relation to the optical input power of the receiver under test. This method is based on the worst-case assumption that a Retransmit Request of the TCP/IP protocol is caused by the detection of a checksum error in the physical layer, thus a bit error. This is not necessarily always the case as other sources of errors; for example, operating system occupancy also may be the source of a Retransmit Request. To eliminate this kind of errors in the BER calculation, it is essential that the typical Retransmit Request contribution of the test system itself is ignorable during the test period.

In a design environment however, the diagnostic power of a test like this is of limited value. Since the cause of an observed Retransmit Request cannot be deduced to one of its many possible bit error sources. Besides amplitude noise, suchlike bit errors can be generated by timing jitter, data dependency, firm/software imperfection etc.. For this kind of design-error diagnostics, a general purpose BER-tester is more appropriate.

2. Definition of terms

- data pattern; to increase the probability of catching physical layer timing and amplitude errors a fast repeating alternating bit pattern of zeros and ones is used as test pattern in the Gigabit link, like 10101010... . For verification purposes, such as gain-bandwidth check or timing versus amplitude diagnostics, the test pattern can be changed in ½ or a ¼ of the maximum bit change frequency, as in patterns containing a major component of 11001100…. or 1111000011110000… . like in hex 78787…. or hex fdfdf…. .
- bit transfer rate; the PCs operating systems are capable of transferring ~80 MB/s over the Gigabit link. The data is packaged in large blocks of 9 kB each. The bit-rate used in calculations is 600 Mb/s.
- the critical lower input test level of the receiver is achieved at that point where the status of the GbE link detection circuitry just remains stable Up during the test-period. This is checked, preferably as single user, by monitoring the Gbit Ethernet link status.

- error rate; all Retransmit Requests are considered to be caused by a checksum error, i.e. a bit error produced in the physical layer. This ignores the fact that Retransmits can also be caused by non-physical layer anomalies, such as operating system occupancies or firmware errors in link modules etc. These sources of error are presumed to be independent of the receiver input signal to noise ratio. This is most likely true as long as the link detection status remains stable Up of both the stressed and the non-stressed receiver input. The, by TCP/IP transfers generated, requests for retransmissions are accumulated in the Retransmit Counter. They are every 10 s reported as a standard error message and at the same time copied to a log file. At the limit of the lower system test level and having a bit transfer rate of 80 MB/s over the link, every 10 s a few tens of Retransmits can be expected. This is equivalent to one error bit per 100 Mb or more of transferred data bits. The error rate is then defined as the average per unit of time of the observed total of Retransmits gained in the test period.
- test time; for a highly accurate QoS number the required test time will rapidly increase with the optical power level at the input of the receiver. A couple of 100 errors are easily obtained within a few minutes time at the lowest optical level. Due to the limitation in practical available test time, and thus of the transferred bits, the observed Retransmits decreases strongly with increased optical receiver input levels. Nevertheless, reports of zero or one Retransmit are significant for the detection of a non-thermal noise presence. See Results text below.

3. Instrumental set-up

See the larger diagram in figure 1. Two Linux machines, Tricot and Triade version RH 6.2 kernel 2.4.14, are used respectively as test server and test client in a GbE test link. Both machines contain a GbE-PCI adapter type Acenic that provides signals conform the IEEE 802.3 SX standard, i.e. a short-range 850 nm optical GbE data connection.

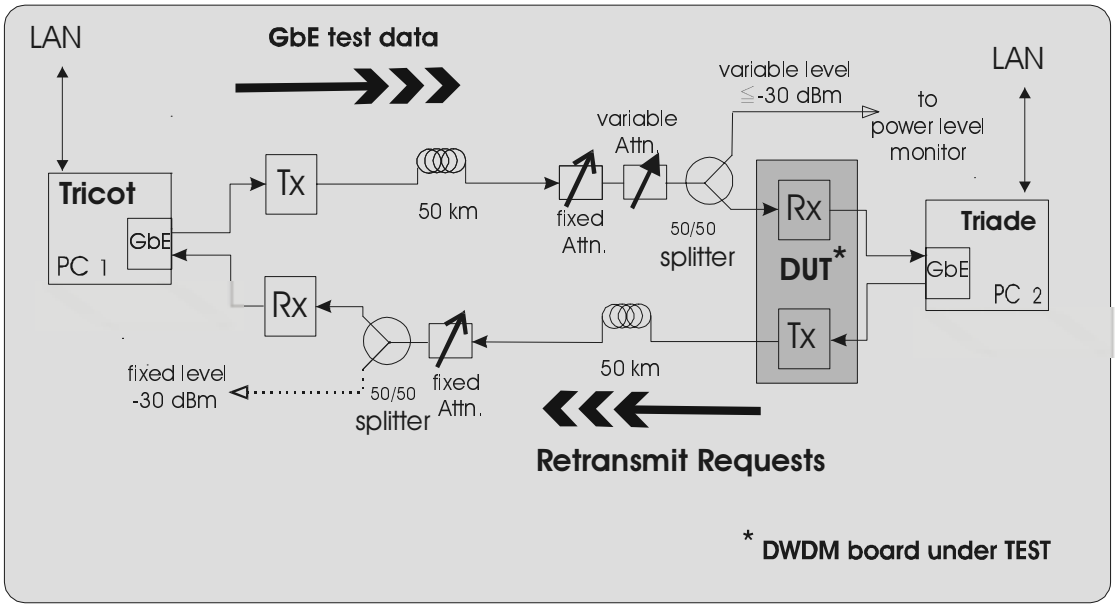

Fig. 1: Set-up diagram of the Giga bit Ethernet test link.

By means of homemade optical to electrical converter, SX2SMB, the fast electrical GbE signals are coupled to the laser and receiver plugs of a DWDM board. 1 GbE data packages are sent in one direction, reply packages are sent back in the reverse direction each on its own link. The length of the optical link is 50 km in both directions. In order to generate the highest bit error probability on the link the sent data stream contain a constant alternating pattern of zeros and ones, hex 4a.. Both links have in front of their receiver a fixed attenuator and an optical splitter. Moreover, a variable attenuator is added to the link of the receiver under test. The fixed attenuator adjusts the optical loss of the test link to a normalized value of -30 dBm at the receiver input. An undisturbed link operation at this level points towards a proper output level of the laser of the board under test. The variable attenuator is used to stress the link in an increasing degree to its limits. The actual optical input level of the receiver is monitored by an optical power level monitor, which in turn is connected to an optical splitter. Optionally, the GbE data path can be equipped with a pair of DWDM mux and demux filters.

The software consists of a traffic producing server tcpsrv and a traffic absorbing client tstclt. Both in the CVS tree: antares-daq/software/util/tstnet/choo. The server can run permanently (rc.local) while the client's can dynamically make/release connections from arbitrary machines to the server. The server as well as the clients will print every ten seconds info about the connection like:

- Thu Aug 1 16:58:45 triade tstclt [7554]: info 11332.04 Msg/sec 93104.06 Kb/sec

The server prints when possible also info about the eventually TCP/IP retransmission count like:

- Thu Aug 1 16:25:32 tricot tcpsrv [6675]: info Retransmit count: 46853 (+2988)

The increment of the retransmit count here (+2988) is accumulated over all connections of all Ethernet interfaces over the previous 10 seconds.

In /sbin/route the by the operating system used alias for the GbE-board can be found. Afterwards, the link status for that alias can be checked with:

> netstat –rn the status of the alias flag should be U(p).

With the following command a GbE (gigcot) data-stream of continues alternating bits of zero's and ones (4a) is send from test server (tricot) to a single (1) test client (triade):

> testnet4a 1 tricot triade gigcot >& log4a

Standard out and error out messages produced by testnet4a will both be filed in log4a. Instead of testnet4a, also testnet78 or testnetfd can be used. In this way the corner frequency, -3 dB down, of the analog system can be derived from respectively 600, 300 and 150 MHz signal measurements.
The PATH environmental should include: /project/antares/matra/export/linux/bin.

4. Measurements

For the given installed laser power on the DWDM boards and the calculated optical loss over the 40 km link to shore an optical power level of about –22 dBm is expected to be available at the receiver input under normal load conditions. Throughout the test a safe optical level of –30 dBm is maintained at the "non-stressed" receiver input. To determine the test-systems typical error rate the "stressed" receiver input is as well set at –30 dBm. An error-free period of over 100 hours is observed in our set-up preparation. System acceptance requires a non-error report over a period, which is at least a few times the actual foreseen test duration. Lowering the optical level of the "stressed" link will ultimately ignite the Loss off Signal LED of the optical receiver. A few dB before that level is reached link communication protocols already start to complain with link UP and link DOWN messages (Retransmit Counter increments with ~tens of Retransmits each time). The stable link up condition turns out to fade away at about –38 dBm or less. The lower test limit of the optical link is achieved slightly above this level, on condition that the GbE link remains in a stable UP position. A good starting point is the report of one or few Retransmits per 10 seconds. (In this stage, an eye pattern of the analogue input level on the scope does not provide any significant noise information). From this point on the Retransmit rate is measured as function of an increasing optical power at the receiver input, in steps of ~ 0.5 dB increments. Initially a fair amount of

[1] For tests on a fast-Ethernet DWDM board, the Giga-bit stream of the test server can be reduced to a 100 MHz stream with help of an extra-appended data-switch.

Retransmits can be accumulated in a few minutes time [2]. A few dB increase (typically 4 steps) of optical power will easily cause an error free transmit period of over 24 hrs. In that case, the latter is interpreted in the analysis as having one error in 24 hrs.

5. Results

In figurer 2 below; for each of the four, by variable attenuator tuned input levels, the reported Retransmit rates are converted into an average number of error-free transferred bits, according the formula:

transferred bits = S [sec/Retransmit] * bit rate [bit/sec] , where

S = average number of seconds, which equals to: total test time over accumulated Retransmits

bit rate = the systems maximum data transfer rate of ~80 MB/s.

This implies that the BitErrorRatio equals to one over the number of error free transferred bits.

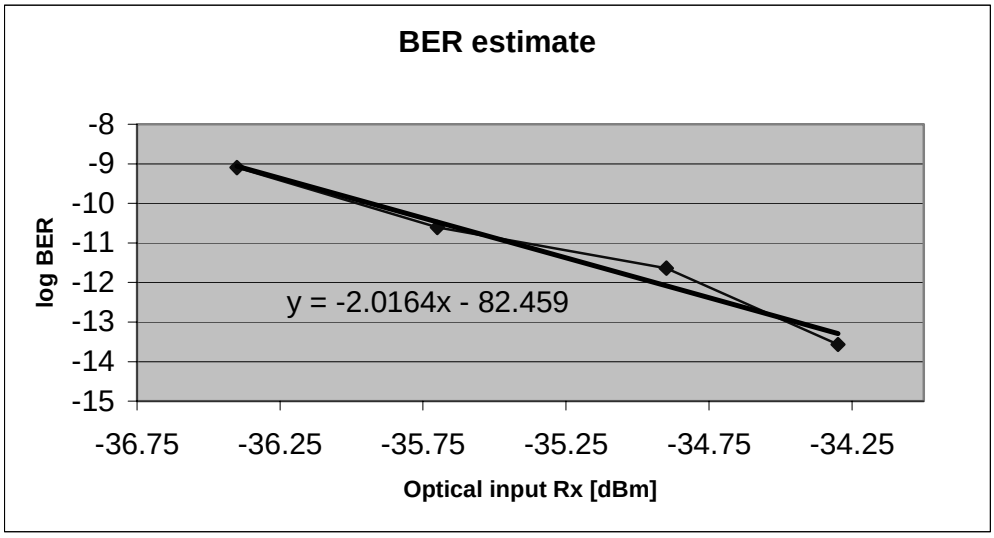

Fig 2: The measured BER value is plotted versus its optical power level at the receiver input. For graph reading convenience the log of BER is plotted.

A fit through the measured points should provide the equation of a straight line, which indicates that only receiver-noise errors are encountered. Otherwise, a broken line and or a floor in the optical receiver input points to system-noise errors. This kind of errors is independent of the S/N ratio at the receiver input and can for that reason not be cured by increasing the optical power level at the receiver input.

The linear equation can be used to extrapolate the BER estimate into a QoS number at a defined receiver input level of –30 dBm. From the data above, the logBER results in a QoS number of –22. In other words, at full data speed on average one by thermal noise caused bit error occurs every 500 millennia.

[2] Real physical layer bit errors induced by amplitude noise results in increments of one Retransmit per message report. The much lower probability that a checksum error turns up due to more then one bit error within the transmission time of a 9 kB TCP/IP package, ~ 150 μs, is ignored.

6. Summary

The given procedure is a relatively simple method to determine an upper limit of the BER estimate of Gigabit Ethernet physical layer components. The figure of merit of such a component, the QoS number, is described in terms of an estimated bit error at an agreed optical reference level. A batch of produced components can be tested by exchanging them one after the other in a so-called reference test system. At the end, the band gap of QoS numbers provides information of the estimated optical system margin. In this way, a production go/no-go test based on the outlook of having sufficient system margin can be realized within a couple of hours per DWDM board.