BER Implication for TCP/IP Network Throughput over a 10x10Gbps Wavelength-Striping Cascaded SOA-Switch

E. Rodriguez-Colina†, A. W. Moore‡, M. Glick*, A. Wonfor†, I. H. White†, R. V. Penty†

*† University of Cambridge, Photonics Comm., 9 JJ Thompson Av., Cambridge, CB3OFA U.K, ‡ Queen Mary, University of London, Dept of Computer Science, Mile End Road, E14NS U.K., *Intel Research Cambridge, 15 JJ Thompson Avenue, Cambridge, CB3 OFD U. K.*

Abstract: A model is reported allowing one to investigate the TCP/IP performance of a Semiconductor Optical Amplifier (SOA) base switch with 10x10Gbps wavelength striped traffic. Bit Error Rate measurements and Packet Error Rate equivalences were assessed to demonstrate TCP throughput analysis in a cascaded SOA for an optical link.
Keywords: Bit-Error-Rate (BER), packet-loss, TCP/IP in Optical Networks, Semiconductor Optical Amplifier (SOA) cascadability, wavelength striping.

1 Introduction

The requirement for future optical switches and optical interconnections pose a problem where future physical infrastructure works with current and widely used transport protocols such as TCP. Presently in electronic switches the signal regeneration is a straightforward process but for Semiconductor Optical Amplifiers (SOA) in a switch-fabric, the signal regeneration is not possible; this has a consequence for error-rate increase. We show that cascading SOAs in a switch provokes greater Packet Error Rate (PER), which consequently impacts upon the network throughput.

We study the repercussion of the Bit Error Rate (BER) in the throughput for several TCP/IP concurrent flows transmitted in a 10x10Gbps wavelength striped link with a reliability of 10^{-7} to 10^{-2} Packet Error Rate (PER) and using a 64B/66B line encoding.

Systems with SOA switches have been show-cased in a number of optical packet interconnection networks [2, 3, 4], and in recent years there has been much interest in the performance of wavelength striped optical networks [3, 4, 5] which offer latency reduction by allowing bit parallel transmission of the data.

For the reason that optical packet switching (OPS) and optical burst Switching (OBS) remain topics of research interest [1, 7] we have proposed the research of an OPS prototype based on a wavelength striped optical link [3, 5, 6], avoiding optical buffering as was done in the SWIFT demonstrator [5, 6, 12, 13, 14].

As previously stated in other publications, SWIFT demonstrates an architecture that is suitable for storage applications, desk area networks and optical-chip-interconnects based on a Semiconductor Optical Amplifier (SOA) switch fabric.

The SOA simulation using VPI [10] was validated with laboratory experiments [3]. Considering that the lab-experiments are constrained by current equipment and technological limitations a more complex system analysis is proposed with simulations for the study of 10 wavelengths where each wavelength carries 10Gbps data. A similar optical link that has been tested, has achieved 100Gbps [3] and the cascadability of an SOA switch fabric has previously been reported by Yang, Williams et. al [4]. In contrast to this past work we show the cascadability of the SOA with TCP traffic behaviour, comparing BER and PER, results that have not yet been reported.

2 Methodology

A connection is made between two results of two simulations which suggest a linear relationship of Packet Error Rate with Bit Error Rate. It is assumed that every bit with an error represents an octet of data that is damaged in a layer above. A BER of 10^{-9} thus implies the mean number of bits in error for each sample of 10^{9} bits is one bit in error. This 1 bit-error may be detected by the Cyclic-Redundancy-Check (CRC) that is a function used to produce a checksum. The distribution of the errors occurring within the physical layer may be irregular although standards as Ethernet at 10Gbps ("IEEE-802.3ae") use 64B/66B encoding which incorporates a scrambler that in-turn whitens user data with respect to line errors. This subsequently removes the non-uniformity of the data-errors [7]. The impact of whitening is to cause a normalization of the relationship between line-errors and errors that occur in the photonics layer.

After the link layer transit, the bit-error is encapsulated in the IP datagram. The presence of an error in the IP layer is not corrected thus is practically the same error received in the TCP layer, so will lead to a TCP segment retransmission.

To enable cascadability analysis in the switch fabric, SOA is used as the switching device so each SOA characterizes the number of switch-stages or end-points that may be interconnected. We aim to measure the performance of cascading SOAs while maintaining minimal degradation of the signal. Degradation is expressed in terms of the Bit Error Rate measurement. The system was analysed where the path between two end-terminals had been established hence we do not need to consider the control analysis for switching or the switching itself. In the figures below schematic representations of the path between point A and point B are presented where the path has been pre-defined, figure 1 and with the TCP traffic generated from 50 servers and 50 clients, figure 2.

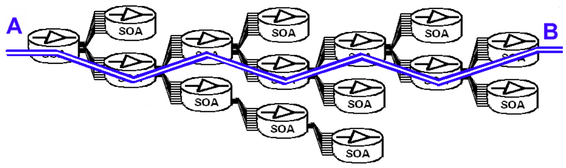

Fig. 1. Schematic of the SOA base switch established path from node A to node B

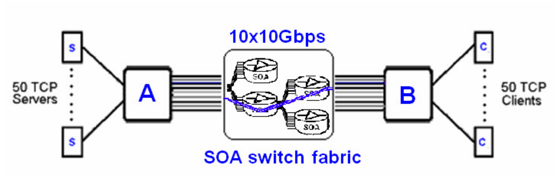

Fig. 2. TCP flows connecting 50 servers and 50 clients through A to B link-path

We evaluated for the worst scenario where each 1 bit-error provoke a packet-error, so for the conversion from Bit-Error-Rate to Packet-Error-Rate a simple calculus is done, assuming PER is equivalent to the multiplication of BER by minimum-transmission-unit (MTU) equal to 1500bytes times 8bits/bytes multiplied by the 66B/64B encoding. Several examples of this association between BER and PER are shown in table 1.

BER	MTU (bytes)	bit/bytes	66B/64B encoding	PER
1.E-08	1500	8	1.03125	1.24E-04
1.E-09	1500	8	1.03125	1.24E-05
1.E-10	1500	8	1.03125	1.24E-06
1.E-11	1500	8	1.03125	1.24E-07
1.E-12	1500	8	1.03125	1.24E-08

Table 1. Equivalence BER to PER for the link

2.1 Simulation

A VPI [10] simulation was performed with 10 wavelengths where each one is carrying a PRBS 2^{15}-1 pattern. This pattern is injected through a SOA model previously validated with experimental results obtained in a test-bench made in our laboratory. The BER and the Q factor were measured for the signal in each wavelength after through the 8 cascade semiconductor optical amplifiers (SOA).

The losses between SOAs are due to the passive components utilized as isolators and couplers which in total is around 15dB loss per stage.

For the networking architecture 2 scenarios were tested using SSFNet simulator [9]. The first one is a single connection between a server and a client using the TCP for several flows over the link-path between A and B in fig. 2; the simplest case when the 100Gbps is not at full capacity utilization. The second scenario uses 50 TCP concurrent flows travelling from 50 servers to 50 clients using the same link-path; this is chosen because the TCP servers and clients sending concurrent flows of data provide full-pipe utilization for the 100Gbps link, matching with a real sample for a high-traffic demand over a high-bandwidth link.

3 Results

Figure 4 shows the Q factor and the comparative Bit Error Rate (BER) for the 10 wavelengths up to 8 SOAs in a row. It can be observed that for cascaded 7 SOAs, the BER is lower than 10^{-7} (1E-07) for the 10 wavelengths (λ1 to λ10). The frequencies are distributed from λ1=1556.21nm to λ10=1562.61nm with 8nm gap between each other.

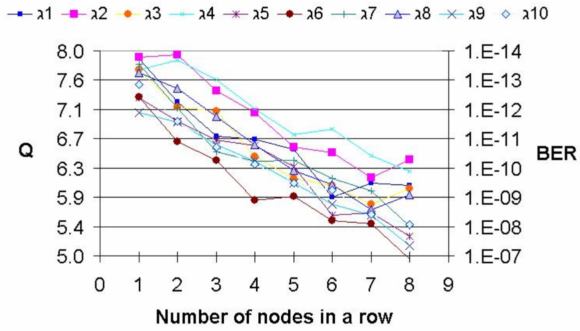

Fig. 4. Q factor and BER versus number of cascaded SOAs for 10 wavelengths (λ1 to λ10).

The system was modelled with one-way-transfer delay of 50 nano-seconds, which corresponds to a short link length (suitable for targeted applications) and which means that the model does not have to deal with fibre dispersion impairments.

The samples of traffic utilized for the 2 scenarios are divided. In the first scenario "single-flow transfer" is divided into 3 test groups:

- The transfer-file-size of 10Mbytes
- Big transfer-file-size of 100Mbytes
- Small transfer-file-size, with a mean of 10Kbytes.

In the second scenario "concurrent-flows from 50 servers and 50 clients" is divided in 2 groups:

- Big transfer-file-size of 100Mbytes
- Small transfer-file-size, with a mean of 100Kbytes

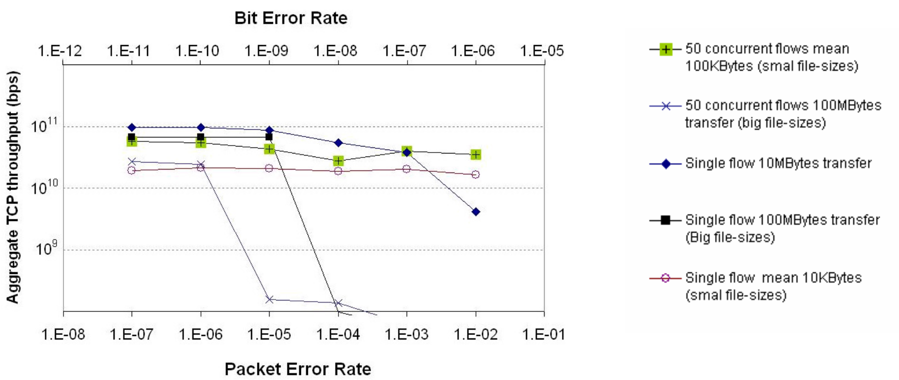

Fig. 5. Aggregated throughput vs. BER, PER, for a single and multiple TCP client-servers, different transfer-file-sizes.

Some implications can be deduced from the results shown in figure 5. These are that the aggregated network throughput does not decrease notably when there are concurrent flows and the file size of each transfer is close to a mean of 100kBytes even if the Packet Error Rate is high. We conjecture that this is because the continual renewal of flows means that even a small number of flows may be stalled with the mean throughput being maintained. Contrast this with the 100MByte transfer where stalled flows are operating for sufficient time that recovery from loss will significantly impact throughput. A further example where the renewal of flows allows high throughput to be maintained is in the case of a random transfers with a mean of 10kbytes the throughput remains stable despite Packet Error Rate being high.

If the number of flows is high and the files are small (smaller than 10kBytes) the aggregate throughput is more stable and does not decrease in a significant way. In contrast to the previous result sending few flows with file sizes bigger than 10Mbytes is that the efficiency level falls 100 fold when the PER is around 10^{-6} or higher.

The aggregated network throughput goes down when the PER is high but TCP is stable and can manage a PER around 10^{-6}, corresponding to a BER of 10^{-10} without a significant drop in the general performance.

4 Conclusions

We have shown cascadability of up to 7 SOAs in a switch matrix, evaluated in terms of the Bit Error Rate and Packet Error Rate, and how the performance can be related to the throughput of several clients and servers sending TCP flows. It is shown that TCP is robust in face of high packet loss PER<10^{-5} when the transfer-file-size is smaller than 10 Mbytes but is inefficient for big file sizes with a small number of flows.

It can be concluded that the Semiconductor Optical Amplifier (SOA) switch fabric carrying 100Gbps can be cascadable due to the fact that the aggregate throughput does not decrease significantly using TCP protocol.

The methodology introduced for Gigabit Ethernet using 64B/66B encoding is useful in resolving the relationship between the optical physical layer and the layers above due to the equalization of the data errors. More analysis is required in case of non-uniformity of errors.

Acknowledgments

We would like to thank J. Crowcroft, K.A. Williams, L.B. James, from University of Cambridge, Computer Lab. and Photonics Communications group for their comments and help and R.L. Allen for valuable proofreading.

References

[1] C. Cameron, H. Le Vu, J.Y. Choi, S. Bilgrami, M. Zukerman and M. Kang: TCP over OBS -fixed-point load and loss, *Optics Express, OSA, Vol. 13, No. 23 / 9172, 2005.*

[2] W. Lu, O. Liboiron-Ladouceur, B.A. Small and K. Bergman: Cascading switching nodes in data vortex optical packet interconnection network, *Electronics Letters, Vol.40 No.14, 2004*

[3] T. Lin, K.A. Williams, R.V. Penty, I.H. White, M. Glick, D. McAuley: Performance and Scalability of a Single Stage SOA Switch for 100Gb/s Wavelength Striped Packet Routing, *Photonics Technology Letters, 2005.*

[4] F. Yang, K.A. Williams, R.V. Penty, I.H. White.: Cascaded performance of semiconductor optical amplifiers for high port count multistage switch fabrics, *Semiconductor Conference. CAS 2004 Proceedings. 2004 International, vol.1, no. pp.- 216, 4-6, 2004.*

[5] M. Glick, M. Dales, D. McAuley, T. Lin, K. Williams, R. Penty, I. White: SWIFT: A testbed with optically switched data paths for computing applications, *Transparent Optical Networks, ICTON 2005. Proceedings of 2005 7th International Conference, vol. 2, no. pp. 29-32 Vol. 2. July 2005.*

[6] K.A. Williams, G.F. Roberts, T. Lin, R.V. Penty, I.H. White, M. Glick, D. McAuley: Integrated Optical 2x2 Switch for Wavelength Multiplexed Interconnects, *IEEE Journal of Selected Topics in Quantum Electronics, vol. 11, no. 1, January/Feb. 2005.*

[7] A.W. Moore, L. B. James, M. Glick, A. Wonfor, R. Plumb, I. H. White, D. McAuley, and R.V. Penty: Optical Network Packet Error-Rate due to Physical Layer Coding, *IEEE/OSA Journal of Lightwave Technology, Volume 23, Number 10, 2005.*

[8] T. Lin, K. A. Williams, P. V. Penty, I. H. White, M. Glick, and D. McAuley: Self-Configuring Intelligent Control for Short Reach 100GB/s Optical Packet Routing. *Optical Fiber Communications, 2005.*

[9] "SSFNet simulator", *http://www.ssfnet.org/SSFdocs/dmlReference.html*

[10] "VPI simulator", *http://www.vpisystems.com/*

[11] James, L.B.: Error Behaviour in Optical Networks, PhD thesis, *Department of Engineering, University of Cambridge, UK, 2005.*

[12] L.B. James, G. Roberts, M. Glick, D. McAuley, K.A.Williams, R. V. Penty, I.H. White: Wavelength Striped Semi-synchronous Optical Local Area Network, *Intel Research SOAPS project, IRC-TR-03-010, 2003.*

[13] M. Dales, M. Glick: SWIFT: A High Capacity Wavelength-striped Optically Switched Network with Electronic Control, *IEEE INFOCOM 2005 Poster/Demo session, 2005*

[14] K. A. Williams, G. F. Roberts, R. V. Penty, I. H. White, M. Glick, D. McAuley, D. J. Kang and M. G. Blamire: Monolithic 2x2 Amplifying Add/Drop Optical Switch for Data Networking, *in Proc. SIOE, 2003.*