Performance Assessment of Optical Packet Switching System with Burst-Mode Receivers for Intra-Data Center Networks

Wang Miao[(1)], Xin Yin[(2)], Johan Bauwelinck[(2)], Harm Dorren[(1)] and Nicola Calabretta[(1)]

[(1)] COBRA Research Institute, Eindhoven University of Technology, the Netherlands, w.miao@tue.nl
[(2)] Ghent University, INTEC/IMEC-iMinds, Belgium

Abstract *We investigate the performance of a burst-mode receiver in an optical packet switching system. Experimental results indicate that a preamble of 25.6ns allows error-free operation of 10Gb/s asynchronous switched packets with 8dB dynamic range and 25ns minimum guard-time.*

Introduction

The exponential growth of the Internet traffic is boosting the requirements of higher capacity data centers networks (DCN)[1]. A flat DCN is demanded due to the bandwidth bottleneck and huge latency suffered by east-west traffic (~75%) in current tree-like DCNs[2]. Optical packet switch (OPS) exploiting sub-microseconds switching and statistical multiplexing could be a preferred solution to efficiently realize the flat DCN. An OPS-based flat DCN has been recently demonstrated which exploits semiconductor optical amplifier (SOA) switches for nanoseconds switching operation[3].

Although DCN is a closed environment with more controlled optical power variation, the receiver should handle packets with different length ranging from sub-microseconds to tens of microseconds, moderate different optical power levels, phase synchronization, and clock. Moreover, signal impairments due to the SOA switches, such as pattern dependent amplification and OSNR degradation, may affect the performance of the receiver. Thus, for practical implementation, packet-based networks need burst mode receivers (BM-RXs). A survey on different BM-RX techniques is reported[4].

Typical BM-RX includes several functions such as fast automatic gain control (AGC) and decision threshold extraction, clock and phase synchronization. Each of those functions contributes, with a different overhead, to the overall BM-RX operational time. This time determines the minimum length of the preamble and the packet guard-time to properly detect the signals[4]. From a network point of view, minimizing the preamble and packet guard-time (packet overhead) would result in higher throughput and lower latency. This is especially important in an intra-DC scenario where many applications produce short sub-microseconds traffic flows.

In this paper, the individual time contributions of the AGC and phase synchronization of a 10 Gb/s BM-RX are experimentally investigated in an OPS-based DCN scenario. Optimization results show that a preamble length of 25.6 ns guarantees error free operation for asynchronous switched packets with 25 ns guard-time and 8dB optical power range.

System operation

The system utilized for investigating the performance of BM-RX in OPS system is shown in Fig.1. Input packets at different wavelength (λ_1-λ_N) coming from top-of-rack switch are switched by the OPS node. Each packet contains a sequence of "1010…" preamble. A certain guard-time is placed in between consecutive packets. The OPS node has a modular architecture with highly distributed control[3]. The switching operation in each module is performed based on broadcasting and selecting procedure employing SOA gates. The on/off state of the SOA gates determines the forwarding/blocking process for the packets. In-

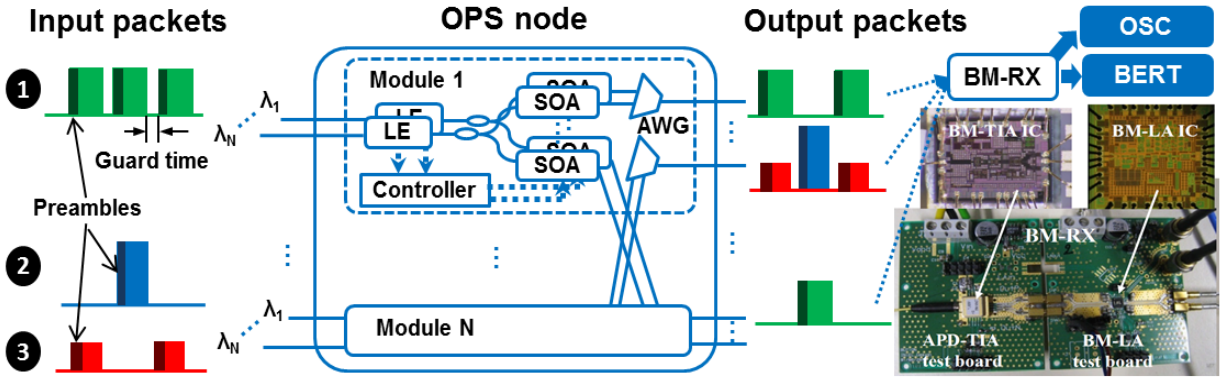

Fig. 1: Experimental set-up to evaluate the OPS system with burst-mode receiver

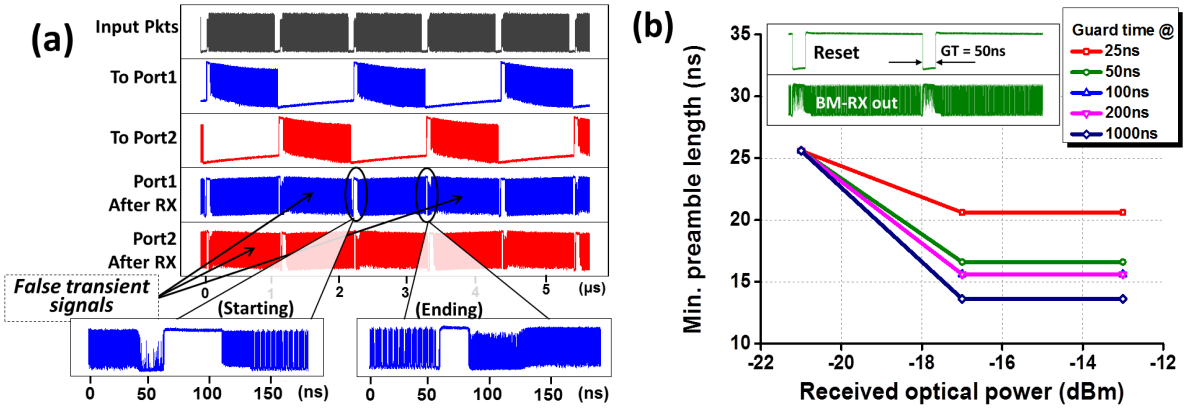

Fig. 2: (a) waveforms detected by two separate BM-RXs; (b) minimum preamble length vs. input power at different guard-time

band RF tone label is separated with payload at label extractor (LE) by using a narrow pass band filter. FPGA-based switch controller detects the destination information carried by the label and generates the controlling signal for SOA gates.

At the output of the OPS node, the switched packets are detected by the BM-RX. It consists of a BM trans-impedance amplifier (BM-TIA) featured with fast gain setting and a BM limiting-amplifier (BM-LA) recovering the amplitude[5]. A reset signal is applied externally in the experiments. Oscilloscope and bit-error rate (BER) tester are used for qualitative and quantitative evaluation of the detected switched packets. Three operational cases have been considered to experimentally evaluate the required preamble length including the effects caused by dynamic power, guard-time and SOA impairments.

Case I determines the preamble as a function of the dynamic power range and guard-time. In the experimental set-up shown in Fig. 1, packet flow 1 may be switched to all the possible output ports at a certain power level. There should be enough preamble length and guard-time for the proper operation of BM-RX and at the same time, without compromising the throughput and latency performance. The preamble length is investigated as a function of received optical power and guard-time and it would also provide the information on the minimum guard-time that could be placed in between the packets.

Case II investigates the capability of the BM-RX to detect asynchronous switched packets. Packet flows 2 and 3 in the set-up shown in Fig. 1 represent the situation that asynchronous packets with different wavelengths and power levels (representing packets that experience different link distances) are forwarded by the OPS to the same output port. As one of the key functions of the BM-RX, the gain and threshold should be fast settled to equalize the power fluctuation of incoming packets.

Case III investigates the preamble as a function of the clock data recovery (CDR) locking time. After the amplitude recovery conducted by the BM-RX, the clock phase alignment should also be realized by a BM-CDR. A fast-lock PLL-based CDR has been utilized in the experiment for this investigation.

Experimental results

We first investigate the BM-RX performance in a 4x4 OPS switching condition. 10Gb/s NRZ OOK packets at λ_1 = 1552.56nm and λ_2 = 1555.74nm consisting of 1200ns (1500Bytes) payload are generated and sent to OPS node. The attached packet labels are set so that the packets are alternatively forwarded by the OPS to Port 1 and Port 2. The labels are detected and processed by the FPGA-based switch controller, which consequently sets the SOA gates of the switch. At OPS output, the packets are detected by the BM-RX.

Figure 2 reports the results on the Case I investigation. The waveforms of the input packets, the switched packets at port 1 and port2, and the detected packets by the BM-RXs are shown in Fig. 2(a). The zoom-in at the starting and ending of a packet provides a better vision of the preamble and the clear payload bits indicates correct amplitude recovery. Original empty switched time slot are then filled with false transient signals due to high gain of the BM-LA. These results have been obtained for 50ns guard-time and -22dBm input power. BER curves for back-to-back (B2B) signal and switched packets after BM-RX are reported in Fig. 3. Error free operation has been obtained for both output ports with 1dB power penalty. In the next experiment, the guard-time and the optical power of the packets are varied to investigate the required preamble length that guarantees BER = 1E-9. The guard-times considered are 25ns, 50ns, 100ns, 200ns, and 1000ns. The optical power varies from -21dBm to -13dBm. The preamble length is optimized with 1ns step. Figure 2(b) indicates that input

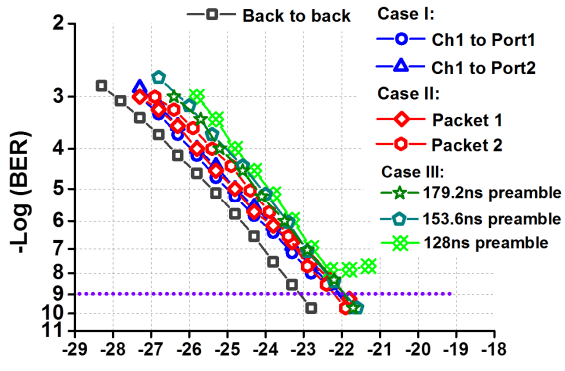

Fig. 3: BER curves for B2b and 3 cases

optical power ranging from -21dBm to -13dBm ensures a dynamic range larger than 8dB for the BM-RX with several nanoseconds decrease of the preamble. For a guard-time of 25ns, a preamble of 25.6 ns guarantees BER < 1E-9. Larger guard-time slightly reduces the required preamble because the charges at parasitic nodes in the circuits will be fully discharged. It also illustrates that the BM-RX would function properly after long empty packets sequence, as in the case of low traffic load.

The results of the Case II investigation are shown in Fig. 4(a). The waveforms of asynchronous packets at different wavelengths and optical power are detected by the BM-RX. Fast AGC is of great significance to guarantee the BM-RX could handle the power fluctuation that may occur in the DCN. The BM-RX output trace shows that the power of two packet flows are equalized. The zoom-in also indicates the successful recovery of both asynchronous flows coming from different sources with different power level. The false transient signal which is caused by the reset settling of the BM-TIA has also been observed in the guard-time. BER curves for the two asynchronous packets are plotted in Fig. 3. Power penalty of 1dB due to switching operation have been measured.

Case III evaluates the preamble length contributed by the BM-CDR time overhead. The employed BM-CDR is a fast-lock PLL-based CDR which is AC coupled to the BM-RX with a time constant of ~100ns[6]. The preamble length is therefore increased to ~150ns, mainly in line with the time constant of the CDR settling time. The output waveform of BM-CDR is reported in Fig. 4(b) which clearly shows the transient response after AC coupling. The BER curves as a function of different preamble lengths are also shown in Fig. 3. For preamble length > 153.6ns (including the 25.6ns due to the BM-RX) error free operation is achieved with 1dB penalty with respect to B2B signal.

Conclusions

We experimentally investigate the performance of a BM-RX in an OPS switching system. Results indicates that a preamble of 25.6ns could allow error free operation of 10Gb/s asynchronous switched packets with 8dB dynamic power range and 25ns minimum guard-time. Deployment of a fast-lock PLL-based BM-CDR introduces extra 128ns preamble. The Gated-VCO based CDR or Over-sampling based CDR would be a better solution to ultimately decrease the preamble contributed by BM-CDR.

Acknowledgements

The authors would like to thank the FP7 LIGHTNESS project (n° 318606) for supporting this work.

References

[1] S. Sakr et al., "A survey on large scale data management approaches in cloud environments," IEEE Commun.Surv. &Tutotials, Vol. **3**, no. 13, p. 311 (2011).

[2] T. Benson et al., "Network traffic characteristics of data centers in the wild," Proc. ACM, pp. 267-280 (2010).

[3] W. Miao et al., "Novel flat datacenter network architecture based on scalable and flow-controlled optical switch system," Optics Express, Vol. **22**, no. 3, p. 2465 (2014).

[4] X. Qiu et al., "Fast synchronization 3R burst-mode receivers for passive optical networks," J. Lightwave Technol., Vol. **32**, no. 4, p. 644 (2014).

[5] X. Yin et al., "Experiments on a 10Gb/s fast-settling high-sensitivity burst-mode receiver with on-chip auto-reset for 10G-GPONs," J. OPT. COMMUN. NETW., Vol. **4**, no. 11, p. B68 (2012).

[6] X. Yin et al., "A 10Gb/s APD-based linear burst-mode receiver with 31dB dynamic range for reach-extended PON systems," Optics Express, Vol. **20**, p. B462 (2012).

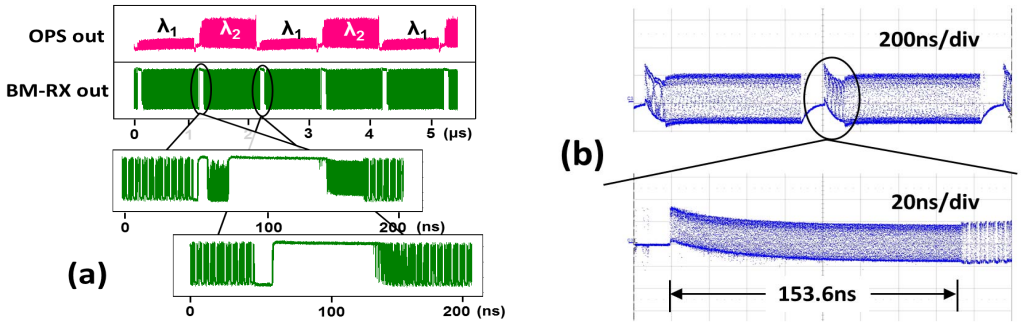

Fig. 4: (a) waveforms of asynchronous packets equalized by the BM-RX; (b) waveform of BM-CDR output