12th Optoelectronics and Communications Conference/16th International Conference on Integrated Optics and Optical Fiber Communication (OECC/IOOC2007)
Technical Digest, July 2007, Pacifico Yokohama

13A1-5

Optical Packet Interconnect System Using High-Speed SOA Switch for Peta-scale Computing System

Kyosuke Sone, Yasuhiko Aoki, Goji Nakagawa, Yutaka Kai, Setsuo Yoshida, Yutaka Takita,
Susumu Kinoshita and Hiroshi Onaka
Fujitsu Limited
4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki 211-8588, Japan (Company mail No. L50)
Tel: +81-44-754-2643, Fax: +81-44-754-2640, E-mail:sone.kyousuke@jp.fujitsu.com

Abstract

We confirmed the optical packet switching operation of our developed 2×2 optical packet interconnect system for peta-scale computing. The interconnect system consists of an arbiter for system control, leaf switches with WDM optical ports for optical packet generation, and semiconductor optical amplifier (SOA) switches for high bandwidth and fast switching.

1 Introduction

Peta-scale computing system consisting of more than 10,000 CPUs is indispensable for various kinds of simulations of nano-scale devices, multi-physics problems, and so on. In the Peta-scale computing system, a high throughput (bandwidth) interconnect system which allows high speed CPUs to freely exchange data is requested [1], [2]. The interconnect system has to support the data granularity ranging from 64-byte short length to several mega-byte huge data with microsecond-order latency.

Previously, we have proposed the WDM optical packet interconnect system and reported the feasibility of a 100-Tbps-class throughput by multi-gate broadcast-and-select SOA switch architecture [3]. In this paper, we describe the switching characteristics in our proposed system with a developed 2×2 optical packet interconnect system.

2 Optical packet interconnect system for peta-scale computing system

Our proposed optical packet interconnect system is shown in Fig. 1. A leaf switch at each computing-node group has two functions, namely the switching of the intra-rack signals and the aggregation of the inter-rack signals with packet queues for high bandwidth data interconnections. When the queue for a certain destination is filled, a queue manager at the node group sends a connection request to an arbiter attached to the opitcal packet switch fabric for scheduling, and the arbiter replies with a grant signal to the manager for the packet forwarding. Subsequently, the aggregated signals are converted to optical packet signals and interconnected between two computing node groups by the optical switch under the control of the arbiter. A WDM packet switching scheme is used to increase the bandwidth per switching port and reduce the number of OE/EO modules and cables. This system has a timing adjustment function which enables to send the optical packet signal to the destination node during the opening of the optical switch based on the system master clock of the arbiter.

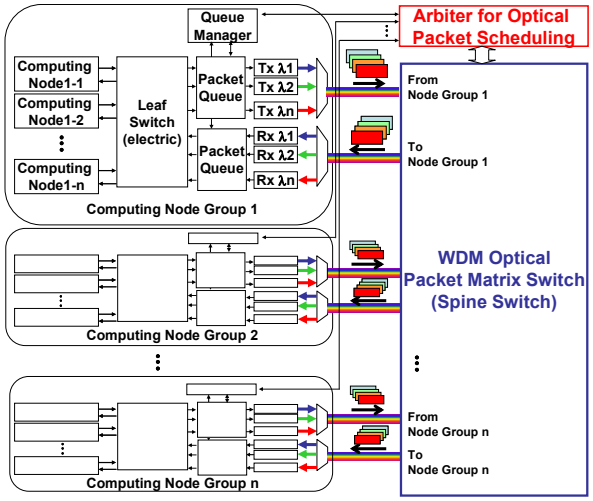

Fig.1 Optical packet interconnect system for peta-scale computing system

3 Demonstration of 2×2 optical packet interconnect system

To confirm the feasibility of our proposed interconnect system, we demonstrated its operation with

a 2×2 optical packet interconnect system. The system consists of two leaf switches, an arbiter, and a 2×2 optical switch subsystem as shown in Fig. 2. The optical switch is a broadcast−and−select type SOA gate switch. SOA has the fast switching characteristic (ns order) with a high extinction ratio (> 60 dB). The wavelengths of the data signal and the control signal are 1.55 μm and 1.3 μm, respectively, and they are coupled and separated by WDM couplers. The switching characteristic of a 2×2 optical switch subsystem is shown in Fig. 3. We obtained a high-speed switching (< 10 ns) at both rising and falling for all SOA switches. The detail system operation of the 2×2 optical packet interconnect system is shown Fig. 4. Upper 4 lines indicate control signals for 4 SOA switches. Bottom 2 lines show switching operation at out ports. In this experiment, we intentionally change the SOA gain, so that we can see all SOA operations. For example, at Out #1 there are 3 levels corresponding to SOA 2-1 on (highest level), SOA 1-1 on (middle level), and the both-off (bottom level) states. Signals are modulated at 10 Gb/s, however, those are averaged due to the small bandwidth of measurement instruments. We confirmed that 1.2 μs optical packet data was successfully switched. The amount of skew, which is the difference in arrival time from each leaf switch to the optical switch, can be suppressed to less than 6ns by the timing adjustment in the interconnect system. We can further shorten the current guard time (~45 ns) to 20 ns by improving the optical switching speed and the timing adjustment granularity.

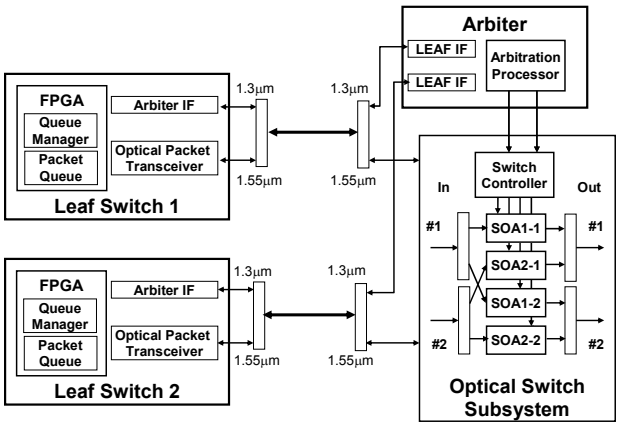

Fig.2 2×2 optical packet interconnect system

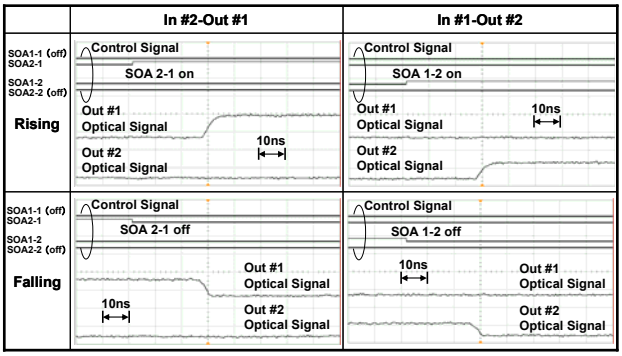

Fig.3 Switching characteristic of 2×2 optical switch

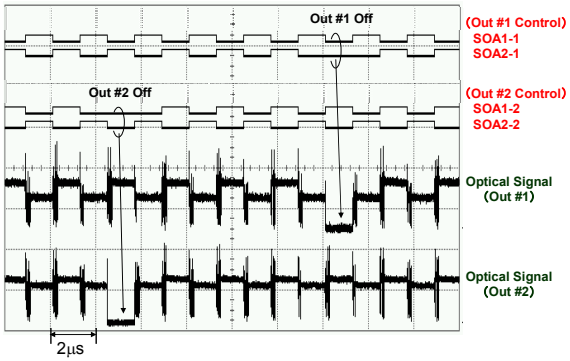

Fig.4 2×2 optical packet interconnect system

4 Conclusions

We have developed a 2×2 optical packet interconnect system using a high-speed SOA switch. We successfully demonstrated the operation of optical packet switching and confirmed the feasibility for peta-scale computing system. In future, we will enhance the number of switch ports up to 256.

Acknowledgements

This research was partly supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan for the design of the optical packet switch architecture and the National Institute of Information and Communications Technology of Japan for the development of the high speed optical switch fabric.

References

1 R. Hemenway et al., Journal of optical networking, Vol. 3, No. 12, pp. 900-913 (2004)

2 T. Lin et al., OFC2006, OWP4 (2006)

3 H. Onaka et al., ECOC2006, Tu4.6.6 (2006)