

## Dynamic Load-Balancing Algorithm Incorporating Flow Distributions and Service Levels for an AOPS Node

Fuding Zhang, Xu Zhou, and Xiaohan Sun\*

*National Research Center for Optical Sensing/Communications Integrated Networking,  
Department of Electronics Engineering, Southeast University, Nanjing 210096, China*

(Received July 11, 2014 : revised September 1, 2014 : accepted September 2, 2014)

An asynchronous optical packet-switching (AOPS) node with load-balancing capability can achieve better performance in reducing the high packet-loss ratio (PLR) and time delay caused by unbalanced traffic. This paper proposes a novel dynamic load-balancing algorithm for an AOPS node with limited buffer and without wavelength converters, and considering the data flow distribution and service levels. By calculating the occupancy state of the output ports, load state of the input ports, and priorities for data flow, the traffic is balanced accordingly. Simulations demonstrate that asynchronous variant data packets and output traffic can be automatically balanced according to service levels and the data flow distribution. A PLR of less than 0.01% can be achieved, as well as an average time delay of less than 0.46 ns.

*Keywords* : Asynchronous optical packet switch, Service level, Load balancing  
*OCIS codes* : (060.4510) Optical communications; (060.1155) All-optical networks; (060.4259) Networks, packet-switched

### I. INTRODUCTION

Load balancing becomes more and more important problem in current high-speed optical networks [1]. The exponential growth of Internet traffic and the emergence of cloud computing could lead to link overload and imbalance at each node of a network, resulting in network performance deterioration, such as increased packet-loss ratio (PLR) and time delay [2-6]. Meanwhile, network traffic is shifting from the “North-South” type to “West-East”, and network resources need to be reassigned with stricter reliability [7-9]. Accordingly, balanced scheduling must be implemented for better performance.

Several studies have been made of routing algorithms for networks to achieve load balancing. A few routing schemes were analyzed in data centers, with improved and guaranteed throughput [10]. An integrated routing algorithm was proposed to achieve load balancing in a hybrid architecture [11]. A topological routing scheme with load balancing and superior quality of service (QoS) support and resource utilization was presented in [12]. Also, the Ameba network architecture with a simple routing algorithm was proposed to achieve load balancing [13]. However, the

assignment and distribution of optical flow inside a node with multiple ports could not be taken into consideration in the works above, so it is difficult to ensure the load balancing inside a node.

Considering the load-balancing implementation inside the node, some expensive components were utilized in several studies. A two-stage switching node fabric with load-balancing algorithm with tunable wavelength converters supporting asynchronous variant data was presented in [14]. The load balancing was performed by hardware configurations in a switch architecture [15]. The packet-loss performance was analyzed in a cloud data center with load-balancing switching using a wavelength converter [1]. The performance enhancement of switching fabric employing load balancing was achieved with wavelength converters in [16]. Balance switching can be implemented for synchronous fixed-length packets with wavelength converters [17]. However, it is ineffective to strike the load balance using just hardware with the average optical flow at each wavelength for each port in such works. Meanwhile, the key components such as tunable wavelength converters are still costly or immature for practical deployment [18-21].

In this paper a dynamic load-balancing algorithm is pro-

\*Corresponding author: [xhsun@seu.edu.cn](mailto:xhsun@seu.edu.cn)

Color versions of one or more of the figures in this paper are available online.

posed, based on data distribution and service levels in the AOPS node, without wavelength converters and with a limited buffer. It can flexibly schedule data with different priorities according to their corresponding requirements, and the simulated demonstration can be implemented with asynchronous and variant optical packets. Data can be dynamically balanced with this algorithm, the PLR and time delay can be reduced effectively, and QoS can be guaranteed.

## II. DYNAMIC LOAD-BALANCING ALGORITHM

Figure 1 illustrates the architecture of an AOPS node, which consists of a time-delay unit, control-management unit, load-balancing unit, and an optical switching array. It is composed of  $m$  in/out optical fiber links,  $n$  wavelength channels per fiber link, and  $k$  optical packets per wavelength channel in time domain. The control-management unit is composed of an information-extraction unit, decision unit, algorithm unit, and processing and configuration unit. The load-balancing unit consists of couplers, optical switches, and a degenerate buffer. The FDLs have level  $L$  and the granularity is  $T$ , with maximum buffer length  $(L-1) \times T$ . The switching units are composed of  $1 \times L$  optical switches in the load-balancing unit, and the optical switching array is

composed of optical switches. Arriving asynchronous optical flow goes through the splitter at any input port, and then is divided two. One enters the control-management unit that can generate the corresponding control signals according to its algorithms; another enters the delay unit with duration  $\tau$  until the control-management unit confirms the switching path. To reduce contention and balance the traffic, the control-management unit should decide whether or not the optical packets enter the buffer of the switching unit, and then the ultimate path to output.

The control-management unit gets the load state of the inflow, monitors the balancing state of output optical packets in real time, and calculates the occupancy state of the wavelength channels. The load-balancing algorithm is executed according to the load threshold, service levels, and other control messages after extraction and analysis of the information from inflow headers. Then the output load balance is optimized, and the corresponding balance-state information database is updated accordingly. Whether the load-balancing algorithm is started depends on the load state, and the flow can be balanced according to the service levels to satisfy the corresponding QoS for each level. Consequently the PLR and time delay are reduced, and the total throughput is increased.

As shown in Fig. 2, the control-management unit first judges the load state of the input flow and starts the load-balancing algorithm if the load is greater than the

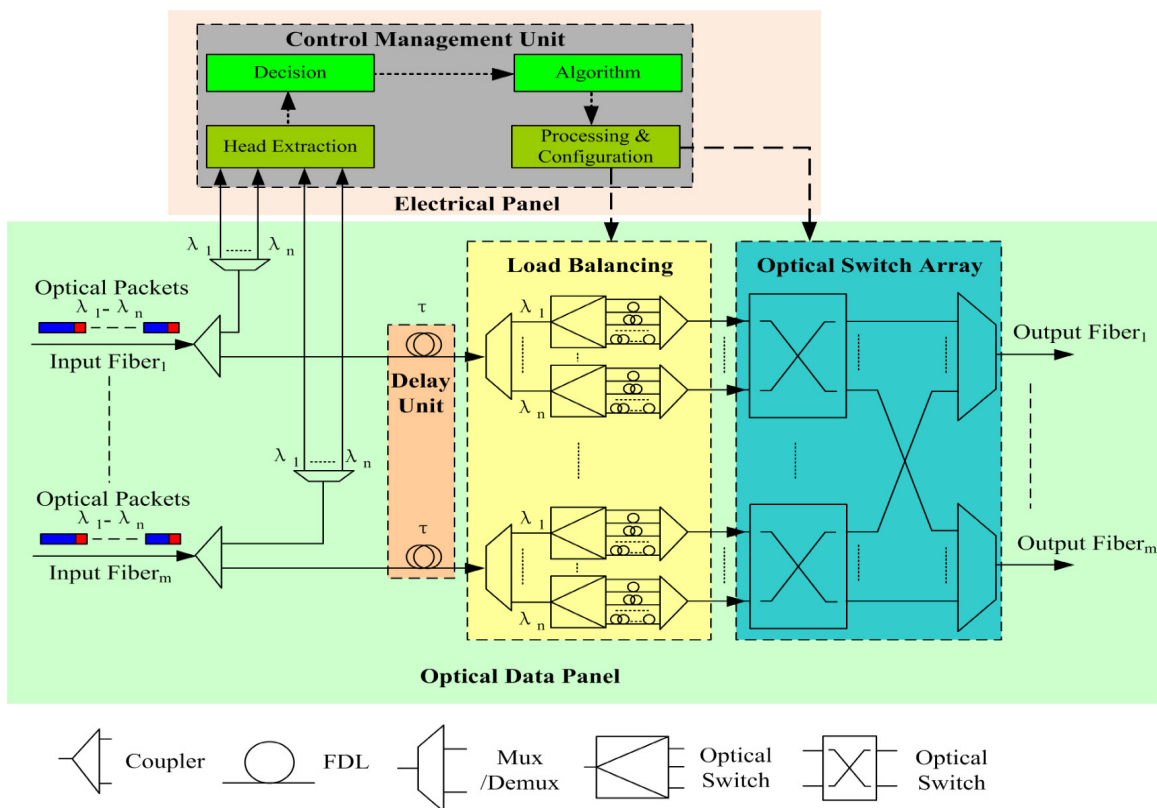


FIG. 1. The architecture of an AOPS node.

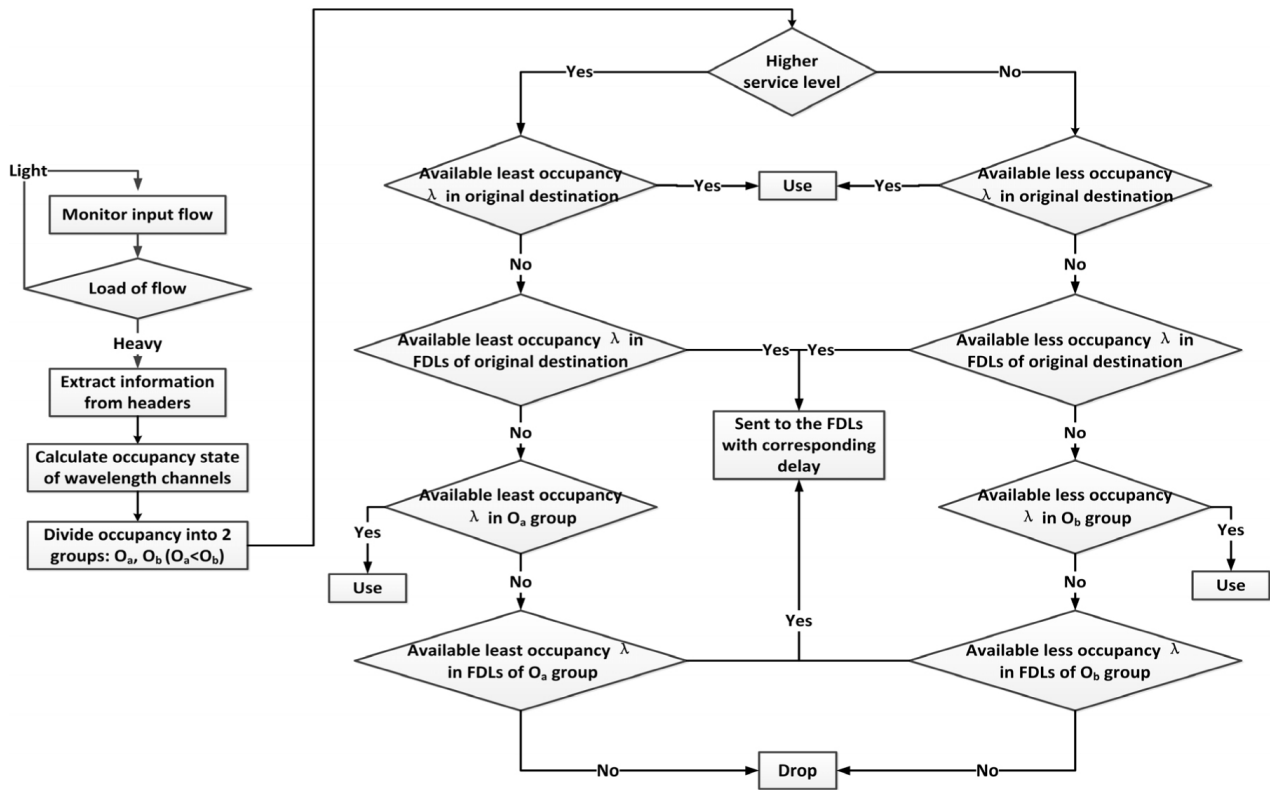


FIG. 2. Flow chart for the load balancing algorithm.

threshold. After extracting information from the packet headers, the current occupancy state of the output ports is calculated and divided into two groups in ascending order,  $O_a$  and  $O_b$ . Then the service level is judged, when the level is too high, the least occupied wavelength channel in the direct output wavelength or FDL array of the original destination port can be searched and confirmed according to the algorithm. If none is available, then the search turns to the group  $O_a$  to find another possible channel in the direct output wavelength channels or FDL array. If the service level is low, then the corresponding requirement is lowered. First the algorithm searches for the least occupied wavelength channel in the direct output wavelengths or FDL array of the original destination port. If none can be found, then the search is switched to the group  $O_b$  to find the suitable path in the direct output wavelength channels or FDL array. Finally the control-management unit updates timely the balancing status of the node in the information database, and refreshes the occupancy states.

Compared to the others, the proposed algorithm may be executed flexibly according to the traffic load, and it can divide the occupancy state into two groups according to the service levels, the data with higher priority entering the first occupancy group while the data with lower priority enter the other. In this way the QoS of higher-level data can be assured, and scheduling time can be saved without the need to traverse the entire occupancy state. The data

will only be sent to the corresponding group according to its priority, and then the data in the queue are divided again into two groups, one being sent to output directly and the other buffered in the FDLs. The packets are dealt with according to the balance state and priorities. In this way the occupancy states are analyzed precisely and the data can be scheduled more effectively.

### III. SIMULATION RESULTS

The simulation platform is built based on the switch fabric shown in Fig. 1, which has 50 input/output ports, 50 fiber channels, and 3 service levels of data flow. The packet length and interval time are asynchronous and variant, subject to the negative exponential distribution shown in Table 1. Figure 3 illustrates the PLR curves of the three service levels and their total, with and without the balancing algorithm in the simulation platform. Compared to the situation without balancing algorithm, the performance with load balancing is better and improved in terms of PLR. Figure 3(a) shows the total PLR curve. We can observe that with the balancing algorithm the PLR is reduced when the load increases. Especially when the load is bigger than 0.5 Erlang, the PLR with the algorithm is obviously lower than that without one. The PLR is reduced by 1.41%, keeping below 0.01%. Figure 3(b) shows the PLR versus

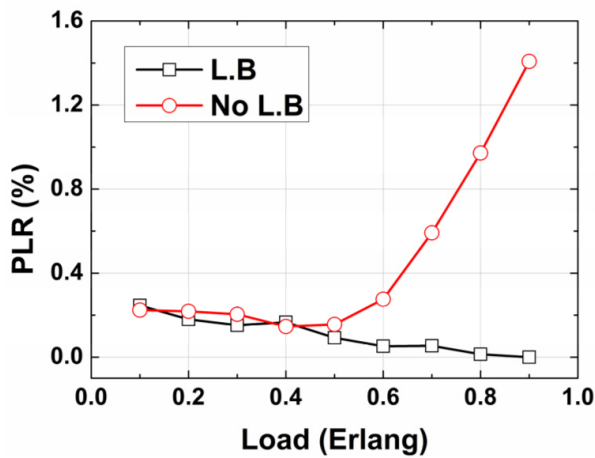
load for the three different service levels. We can observe that when the load is greater than the threshold, the PLR with the balancing algorithm is also far lower than that without one, and the PLR for all three levels decrease as

TABLE 1. Simulation parameters

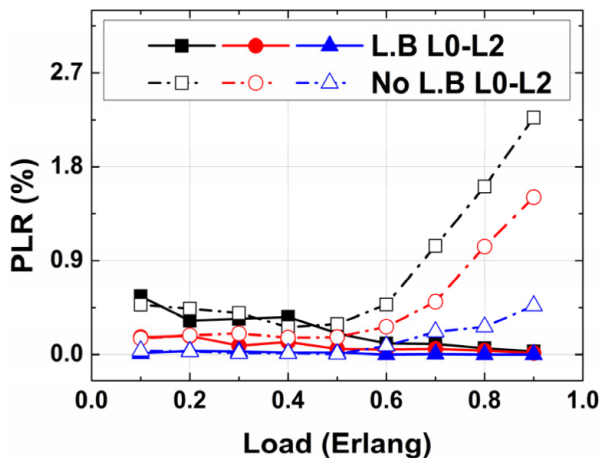
Parameter description	Value
Number of input/output ports	50
Number of link wavelengths	50
Number of service levels	3
Data speed	40 Gbit/s
Power of source	4 dB m
Distribution of packet length	Exponential, mean 200 bytes
Distribution of interval time	Exponential, mean 300 ns
FDL granularity	1 ns
FDL Level	50
Traffic load	0.1-0.9

the load increases. The L0 service with the lowest level has the maximum PLR reduction of 2.24%; the L2 service with the highest level has the least PLR reduction of 0.47%, while the L1 has a reduction of 1.5%.

Figure 4 illustrates the average time delay for the three service levels and their total, with and without the balancing algorithm in the simulation platform. Compared to the situation without balancing algorithm, the performance with load balancing is better and improved in terms of average time delay. Figure 4(a) shows the total average time delay curve. We can observe that the average time delay increases when the load increases. Especially when the load is bigger than 0.5 Erlang, the average time delay with the algorithm is obviously lower than that without one. The average time delay is reduced by 0.09 ns, keeping below 0.46 ns. Figure 4(b) shows the average time delay versus load for the three different service levels. We can observe that when the load is greater than the threshold, the average time delay with the balancing algorithm is also far lower than that without one, and the average time delay of the three different levels increases as the load increases.

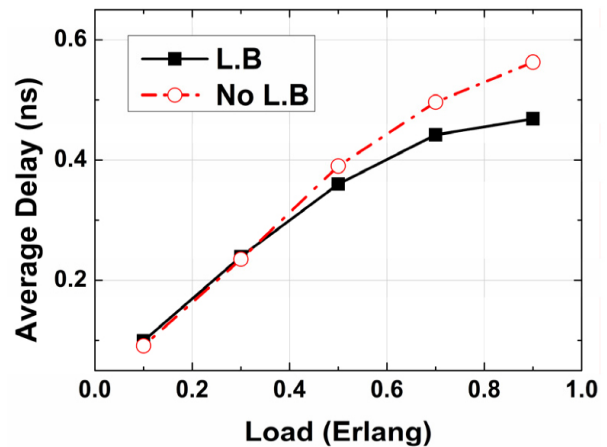


(a)

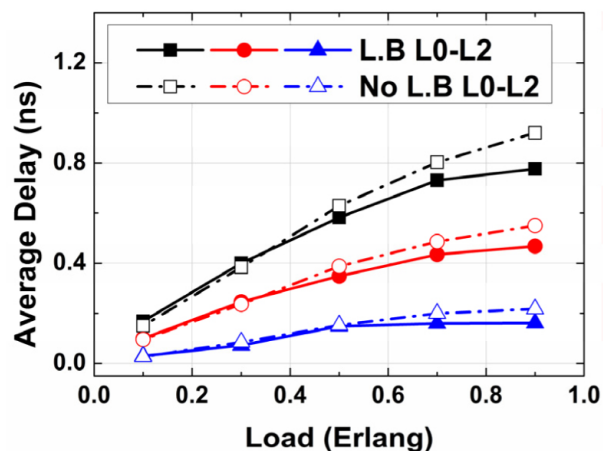


(b)

FIG. 3. PLR: (a) total, (b) for different service levels. “L.B.” stands for “Load Balancing”.



(a)



(b)

FIG. 4. Time delay: (a) total, (b) for different service levels. “L.B.” stands for “Load Balancing”.

The L0 service with the lowest level has the maximum average time delay reduction of 0.14 ns; the L2 service with the highest level has the least average time delay reduction of 0.05 ns, while L1 has a reduction of 0.08 ns.

The simulation results show that the balancing algorithm can reasonably balance the asynchronous variant data flow according to the service levels and traffic distribution, and optimize the PLR and time delay for the different service levels correspondingly. The PLR and time delay can be reduced effectively, and the throughput increases when the load becomes higher.

#### IV. CONCLUSION

This paper proposes a novel dynamic load-balancing algorithm to improve the performance, in terms of PLR and time delay, for an AOPS node with limited buffer and no wavelength converters, based on the data flow distribution and the service levels. The algorithm is demonstrated by simulation, the results showing that it can enhance the performance for the different service levels and overall, in terms of PLR and time delay in the node. The performance is improved with PLR reduction of 2.24% for different services maximally and 1.41% for the total, staying below 0.01%, and the time-delay reduction of 0.14 ns for different services maximally and 0.09 ns for the total, staying below 0.46 ns. QoS of high level is guaranteed, with optimized output balance and enlarged throughput.

#### ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 60972025 and Grant 61271206.

#### REFERENCES

1. H. Rastegarfar, L. A. Rusch, and A. Leon-Garcia, "Load balancing in wavelength-routing cloud data centers," in *Proc. HPSR* (Regent Taipei, Taipei, China, 8-11 July 2013), pp. 211-212.
2. N. Bitar, S. Gringeri, and T. J. Xia, "Technologies and protocols for data center and cloud networking," *IEEE Commun. Mag.* **51**, 24-31 (2013).
3. X. Jie, W. Bin, J. Xiaohong, A. Pattavina, W. Hong, and Z. Lei, "Scalable data center network with distributed placement of component sets in optical networks," in *Proc. ICC* (Xi'an International Conference Center, Xi'an, China, 12-14 Aug. 2013), pp. 204-208.
4. L. Yi, X. Gaoxi, and H. Ghafouri-Shiraz, "On traffic allocations in optical packet switches," *IEEE J. Sel. Areas Commun.* **25**, 108-117 (2007).
5. L. M. Peng, C. H. Youn, W. Tang, and C. M. Qiao, "A novel approach to optical switching for intradata-center networking," *J. Lightwave Technol.* **30**, 252 (2012).
6. S. Yoo, S. Mun, J. Kim, and C. Lee, "1.25 Gb/s broadcast signal transmission in WDM-PON based on mutually injected Fabry-Perot laser diodes," *J. Opt. Soc. Korea* **16**, 101-106 (2012).
7. W. Miao, J. Luo, S. Di Lucente, H. Dorren, and N. Calabretta, "Novel flat datacenter network architecture based on scalable and flow-controlled optical switch system," *Opt. Express* **22**, 2465-2472 (2014).
8. C. Min, J. Hai, W. Yonggang, and V. C. M. Leung, "Enabling technologies for future data center networking: A primer," *IEEE Netw.* **27**, 8-15 (2013).
9. N. Charbonneau and V. M. Vokkarane, "Performance modeling of HS-RR-TCP over load-balanced optical burst-switched (OBS) networks," *Opt. Switch. Netw.* **8**, 116-128 (2011).
10. N. Maksic and A. Smiljanic, "Improving utilization of data center networks," *IEEE Commun. Mag.* **51**, 32 (2013).
11. S. Wei-Tao, W. Shing-Wa, C. Ning, K. Balasubramanian, Z. Xiaoqing, M. Maier, and L. G. Kazovsky, "Hybrid architecture and integrated routing in a scalable optical-wireless access network," *J. Lightwave Technol.* **25**, 3443 (2007).
12. J. H. Broberg, K. Balachandran, S. Hede, J. Pedersen, M. T. Riaz, and J. M. Pedersen, "Load balancing in grid networks," in *Proc. ICACT* (Phoenix Park, PyeongChang, Korea, 7-10 Feb. 2010), pp. 1041-1046.
13. H. Tode, K. Tada, and S. Kohama, "Ameba network architecture based on advanced multi-layer network and its configuration algorithm," in *Proc. ICC* (Budapest Marriott Hotel, Budapest, Hungary, 9-13 June 2013), pp. 3481-3486.
14. H. J. Yang, J. Q. Hu, Z. Pan, F. Xue, and S. Yoo, "An asynchronous, variable length two-stage packet switch fabric architecture with look-ahead and load-balance scheduling schemes for optical-label switching networks," in *Proc. GLOBECOM '04* (Hyatt Regency Dallas, Dallas, USA, 29 Nov.-3 Dec. 2004), pp. 1828-1834.
15. S. Arekapudi, C. Shang-Tse, I. Keslassy, and N. McKeown, "Configuring a load-balanced switch in hardware," in *Proc. High Performance Interconnects* (Stanford Univ., CA, USA, 25-27 Aug. 2004), pp. 48-53.
16. H. Yang and S. J. B. Yoo, "New optical switching fabric architecture incorporating load balanced parallel rapidly switching all-optical variable delay buffer arrays," in *Proc. OFC/NFOEC* (Anaheim Convention Center, Anaheim, USA, 6-11 March 2005), p. 3.
17. J. Gripp, J. E. Simsarian, J. D. LeGrange, P. Bernasconi, and D. T. Neilson, "Photonic terabit routers: The IRIS project," in *Proc. OFC/NFOEC* (San Diego Convention Center, San Diego, USA, 21-25 March 2010), pp. 1-3.

18. H. Moon and S. Jung, "Multi-coded variable PPM for high data rate visible light communications," *J. Opt. Soc. Korea* **16**, 107-114 (2012).
19. H. Aghababaeian and M. Vadjed-Samiei, "Compact and temperature independent electro-optic switch based on slotted silicon photonic crystal directional coupler," *J. Opt. Soc. Korea* **16**, 282-287 (2012).
20. A. Dolatabady and N. Granpayeh, "All optical logic gates based on two dimensional plasmonic waveguides with nanodisk resonators," *J. Opt. Soc. Korea* **16**, 432-442 (2012).
21. R. Asif, M. K. Islam, and M. Zafrullah, "All-optical signal processing of fiber impairments in dual-polarization 112 Gbit/s m-ary QAM coherent transmission," *J. Opt. Soc. Korea* **17**, 57-62 (2013).