DEVELOPMENT OF PAM-4 SIGNALING FOR HIGH PERFORMANCE COMPUTING, SUPERCOMPUTERS AND DATA CENTER SYSTEMS

Duy-Tien Le[*], Ngoc-Minh Nguyen[+] and Trung-Thanh Le[*]

[+]Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam
[*]International School, Vietnam National University (VNU-IS), Hanoi, Vietnam

Abstract: We propose a new scheme for multilevel pulse amplitude modulation (PAM-4) signaling for optical interconnects and data center networks. Our approach is to use only one 4x4 multimode interference (MMI) structure with two phase shifters in push-pull configuration. An extreme high bandwidth and compact footprint can be achieved. The whole device is designed using the existing VLSI technology.

Keywords: data center, high performance computing, optical interconnect, supercomputer.

I. INTRODUCTION

Over the last few years, the explosive increase of internet service driven from applications, such as streaming video, social networking and cloud computing, the demand for high bandwidth, throughput interconnection networks is required. As conventional electronic interconnection has reached its capacity limit, it is rather challenging to improve the performance of throughput and latency while maintaining low power consumption. In recent years, many significant advances and approaches have been undertaken to overcome the limitation.

Optical interconnection network is a promising means of high bandwidth and low latency routing for future high performance computing platforms. Data centers are large-scale computing systems with high-port-count networks interconnecting many servers, typically realized by commodity hardware, which are designed to support diverse computation and communication loads while minimizing hardware and maintenance costs. Contemporary data centers consist of tens of thousands of servers, or nodes, and new mega data centers are emerging with over 100,000 nodes [1].

A data center consists of computer systems and associated components used for high performance computing as shown in Fig. 1 [2, 3]. The majority of optical interconnection architectures for data center are based on devices used in optical communication networks. Optical technologies will be required across the entire computer system, including processors, memory, storage, interconnects, and system software. For interconnects, the power required to communicate a bit across many distance scales (rooms, racks, boards, and chips) must be lowered dramatically as requirements for bandwidths per link increase. Photonics will play a key role in meeting power goals at all levels of granularity in future high-performance computing (HPC) and data centers [4].

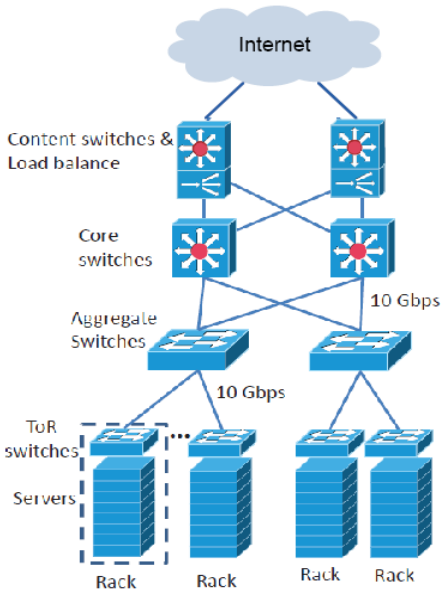

Fig.1. Architecture of data centers

The most promising approach to improve performance across the entire installation is to provide higher bandwidths through the installed infrastructure. Using photonic or optical co-packaged with processors, switches, and future systems-on-chip (SOCs), we can increase the bandwidth to all nodes and endpoints in the datacenter without any changes to the racks or boards—and without requiring more fiber connections to chips.

HPC systems and data centers have quite similar architectures: a large number of many-core processing nodes

Corespondence: Duy-Tien Le
email: ldtien82@gmail.com
Received: 31/08/2017, corrected: 07/09/2017, accepted: 08/09/2017

are connected by scalable interconnect networks [2, 5]. Recent trends in data center consolidation as well as the growth of cloud-based computation and storage, have resulted in datacenters with node counts that exceed that of most supercomputer systems. But HPC systems are usually dedicated to a single application at a time, while data centers typically run a large number of concurrent applications. As a result, a key difference between HPC systems and commercial data centers is the utilization of the interconnect networks: as data centers make less use of fine-grain distributed processing, they can require less network bandwidth to support a given amount of processing power.

By 2020, deployment of exascale systems with as many as 100,000–1,000,000 nodes is expected to be underway [6]. By that time, single-chip processors with sustained performance exceeding 10 Teraflops will be available, exploiting both high levels of thread parallelism and SIMD parallelism (similar to today's GPUs) within the floating-point units. With memory bandwidths as high as 4 Terabytes/s (TB/s), one of the most critical aspects of the node design shown in Fig. 2 will be providing sufficient memory bandwidth to sustain the processor within an acceptable power budget (e.g., 200 W).

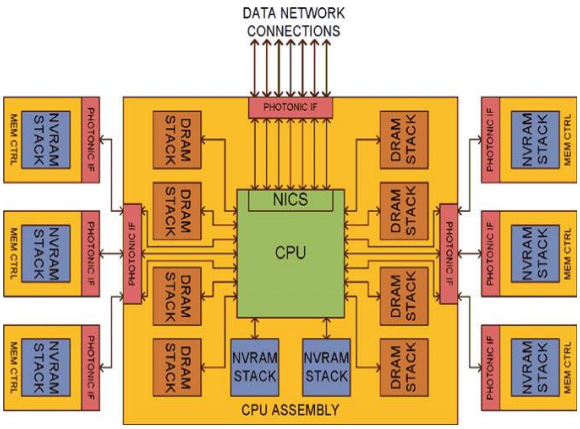

Fig. 2. Extrascale computing node

This will be achieved be either stacking "near" memory directly on the processor, or locating it within the processor package itself. As the amount of memory that can be connected in this way is limited, additional "far" memory (provided by nonvolatile RAM) will be provided by memory modules connected to the processor through high-speed links. Distributed memory programming techniques, such as MPI message passing, are used across a network spanning 100,000 nodes with required bandwidths of at least 1 Terabyte/s per connection.

One of the current top supercomputer IBM Sequoia uses over 1.5 million cores. With a total power consumption of 7.9 MW, Sequoia is not only 1.5 times faster than the second-ranked supercomputer, the K computer, but also 150% more energy ecient. The K computer, which utilizes over 80,000 SPARC64 VIIIfx processors, results in the highest total power consumption of any Top500 system (9.89 MW). IBM Sequoia achieves its superior performance and energy eciency through the use of custom compute chips and optical links between compute nodes. Each compute chip shown in Fig. 3 contains 18 cores: 16 user cores, 1 service, and 1 spare. The chips contain two memory controllers, which enable a peak memory bandwidth of 42.7 GB/s, and logic to communicate over a 5D torus that utilizes point-to-point optical links.

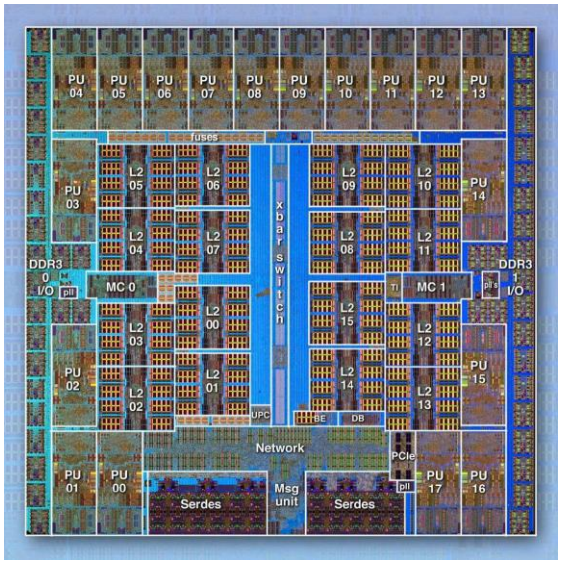

Fig.3. Blue Gene Q Compute Chip - IBM's Blue Gene Q compute chip contains 18 cores and dual DDR3 memory controllers for 42.7 GB/s peak memory bandwidth.

One of the most important approach used for optical interconnects used in data center and high performance computing systems is to use multilevel modulation systems such as PAM or QAM [7, 8].

4-PAM modulation is one of the most modulation schemes used in the data center. In recent years, there are two approaches to implement optical PAM-4 modulation schemes. For example, microring resonator [9-14] or MZI with multiple electrodes [15-18] can be used for 4-PAM modulation. However, these structures based on MZI structure, so they have a large footprint, low fabrication tolerance and they are very sensitive to the fabrication error.

Therefore, in this study, we propose a new architecture to implement a 4-PAM signaling system by using only one 4x4 MMI coupler to solve the above limitation. Here we show that the comsumption power of our structure is very small compared to the conventional structure. In addition, we use two phase shifters and two data bit b0b1 will control the phase shifters with a length of the ring resonator waveguide is exemely short, therefore a very compact device can be achieved.

II. THEORY AND SIMULATION RESULTS

Our proposed device schematic for PAM-4 signaling using a 4x4 MMI coupler is shown in Fig. 4(a). Here we use two PN junction phase shifter segments, which use the plasma dispersion effect in silicon waveguides. The structure of the optical silicon waveguide and PN phase shifters are shown in Fig. 4(b).

The change in index of refraction is phenomenologically described by Soref and Bennett model [19]. Here we focus on the central operating wavelength of around 1550nm.

The change in refractive index is described by:

$$\Delta n \text{ (at 1550nm)} = -8.8\text{x}10^{-22}\Delta N - 8.5\text{x}10^{-18}\Delta P^{0.8} \quad (1.1)$$

The change in absorption is described by:

$$\Delta\alpha \text{ (at 1550nm)} = 8.5\text{x}10^{-18}\Delta N + 6\text{x}10^{-18}\Delta P[\text{cm}^{-1}] \quad (1.2)$$

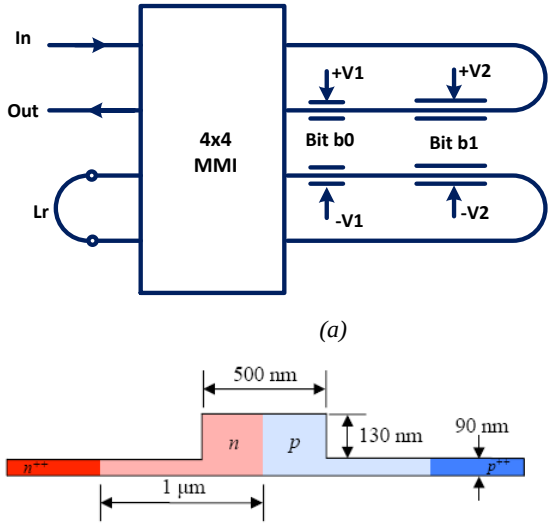

Fig 4. (a) Scheme of a PAM-4 signaling based on a 4x4 MMI coupler and (b) PN junction phase shifter with reserve bias and the structural parameters of the waveguide

The mode profile of the optical waveguide at 1550nm is shown in Fig.5, where the effective refractive index is $n_{eff} = 2.612016$ by using the EME method.

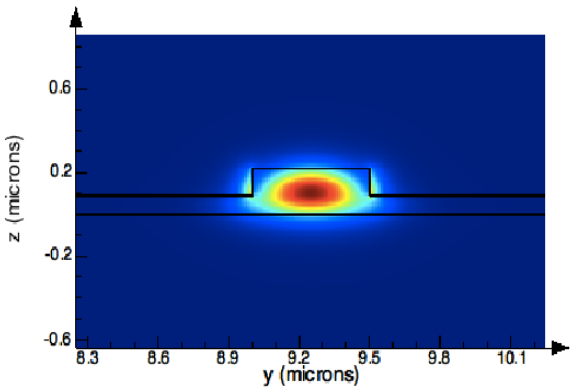

Fig 5. Mode profile calculated by the EME method

Optical power transmission of the proposed device can be modulated from theoretical 0 to unity by varying the phase difference in right two arms of Fig.4(a), $\Delta\phi$, between $2\sin^{-1}(\alpha)$ and π for direct connection Lr. Here Lr is particulary small, so the loss factor α is high and neary unity.

By segmenting the length of the phase shifter into L1 and L2, where $L_2 = 2L_1$ with applied voltage V_1 and V_2 respectively in Fig. 4, multilevel optical modulation can be achieved. It is assumed that the phase shifter with the length L_1 is for LSB bit and L_2 is for MSB bit of input data bits b_1b_0.

By using the mode propagation method, the length of 4x4 MMI coupler with the width of W_{MMI} is to be $L_{MMI} = \frac{3L_\pi}{2}$ [20]. Then by using the BPM simulation, we showed that the width of the MMI is optimized to be W_{MMI} =6µm for compact and high performance device. The calculated length of a 4x4 MMI coupler is found to be $L_{MMI} = 141.7$ µm as shown in Fig. 6 when input signal is at port 1.

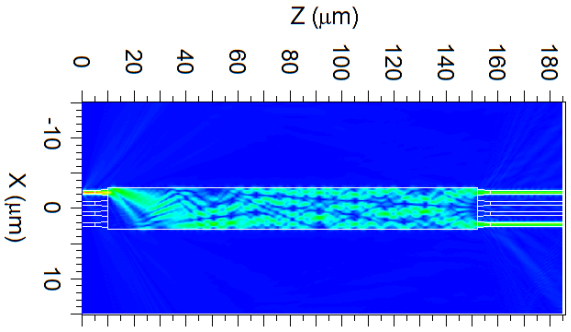

Fig. 6 Power transmissions through the 4x4 MMI at the optimized length 141.7 µm *, input signal is at port 1*

The FDTD simulation of the whole device is shown in Fig. 7(a). We take into account the wavelength dispersion of the silicon waveguide. A Gaussian light pulse of 15fs pulse width is launched from the input to investigate the transmission characteristics of the device. The grid size $\Delta x = \Delta y = 0.02\text{nm}$ and $\Delta z = 0.02\text{nm}$ are chosen in our simulations. The VLSI mask design of the device is shown in Fig. 7(b). Our design showed that a very compact device can be achieved.

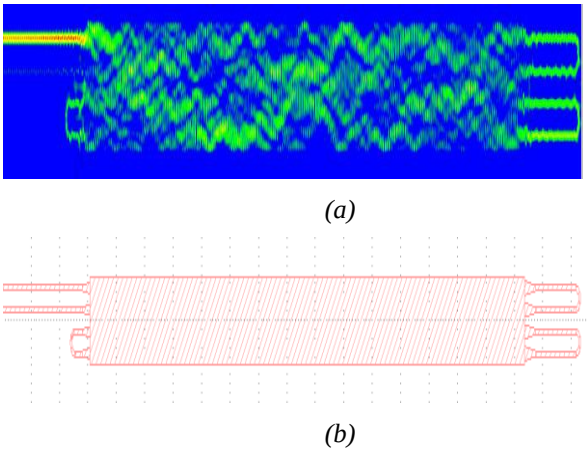

Fig. 7 FDTD simulation of the whole device when input signal is at port 1

By using transfer matrix method, the normalized transmission of the device can be expressed by

$$T = \frac{P_{out}}{P_{in}} = \frac{\alpha^2 + \cos^2(\frac{\Delta\phi}{2}) - 2\alpha\left|\cos(\frac{\Delta\phi}{2})\right|\cos(\theta)}{1+\alpha^2\cos^2(\frac{\Delta\phi}{2}) - 2\alpha\left|\cos(\frac{\Delta\phi}{2})\right|\cos(\theta)} \quad (1.3)$$

Where the transmission loss factor α is $\alpha = \exp(-\alpha_0 L_r)$, where $L_r = \pi R$ is the length of the microring waveguide in Fig.4, R is the radius of the microring resonator and $\alpha_0(dB/cm)$ is the transmission loss coefficient. $\theta = \beta_0 L_r$ is the phase accumulated over the microring waveguide, where $\beta_0 = 2\pi n_{eff}/\lambda$, λ is the optical wavelength and n_{eff} is the effective refractive index.

At resonance, $\theta = 2m\pi$, $\cos(\theta) = 1$, m is an integer, the transmission can be expressed by [21]

$$T = \frac{P_{out}}{P_{in}} = \frac{\left|\alpha - \left|\cos(\frac{\Delta\phi}{2})\right|\right|^2}{\left|1-\alpha\left|\cos(\frac{\Delta\phi}{2})\right|\right|^2} \quad (1.4)$$

The normalized transmission of the device at resonance when the loss factor $\alpha = 0.995$ is shown in Fig. 8. This result shows that the power consumption to achieve multilevel PAM-4 is much lower than the conventional structure based on Mach Zehnder modulator in the literature.

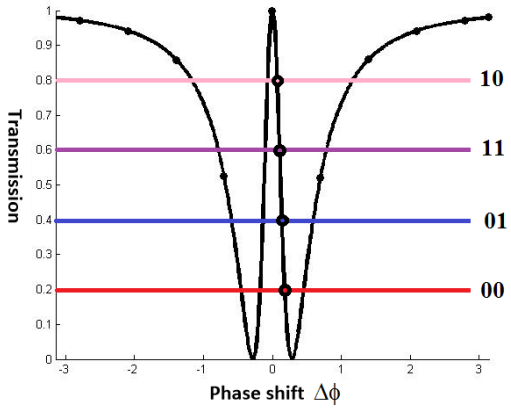

Fig. 8 Transmission at resonance with different phase shifters

The simulation results in Fig.8 show that for data bits 00, 01, 11, 10, the total phase difference between two arms of Fig.1 must be 0.0558π, 0.0428π, 0.0323π and 0.0215π, respectively.

The effective index change was achieved by the plasma dispersion effect in silicon waveguide due to the applied voltage. For example, we use a phase shift total length of 10um, the required phase shift for PAM-4 can be easily achieved as shown in Fig. 9.

Fig.10 shows the normalized transmissions at for input data streams of 00, 01, 11, 10. The normalized outputs at resonant wavelength is 0.2, 0.4, 0.6 and 0.8, respectively. It assumed that the mirror can be used at the corner of the waveguide at the left hand side of Fig.4, the ring radius of 3um can be used. As a result, a very high free spectral range of 72nm can be achieved with our proposed structures. This means that our approach can offer a very high bandwidth and it allows us to use multiple channels in the same waveguide. Therefore, it is very useful for multicore micrprocessors, high performance computing and data center systems in the future.

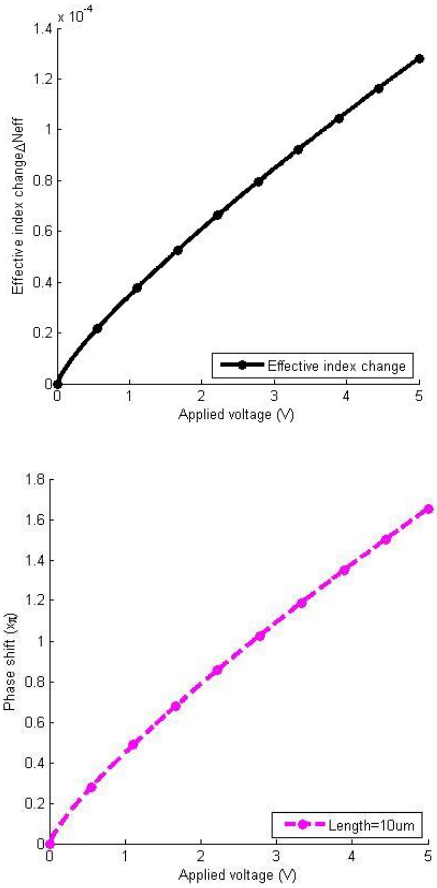

Fig. 9 Effective index change and phase shift with the electrode length of 10um

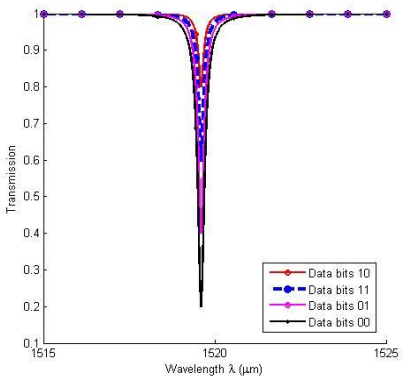

Fig. 10 Transmission of the proposed structure for input data bits 00, 01, 10, 11

III. CONCLUSION

We have presented a new approach for PAM-4 signaling implementation using only one 4x4 MMI coupler based on CMOS technology. The design is suitable for VLSI design. Our proposed approach requires a low power comsumption and compactness. The proposed approach is suitable and useful for high performance computing, multicore and high speed data center systems.

REFERENCES

[1] Tolga Tekin, Richard Pitwon, Andreas Håkansson *et al.*, *Optical Interconnects for Data Centers*: Woodhead Publishing, 2016.

[2] M. A. Taubenblatt, "Optical Interconnects for High-Performance Computing," *Journal of Lightwave Technology*, vol. 30, pp. 448-457, 2012.

[3] Laurent Vivien and Lorenzo Pavesi, *Handbook of Silicon Photonics*: CRC Press, 2013.

[4] R. Lytel, H. L. Davidson, N. Nettleton *et al.*, "Optical interconnections within modern high-performance computing systems," *Proceedings of the IEEE*, vol. 88, pp. 758-763, 2000.

[5] Agam Shah. IBM Chip Breakthrough May Lead to Exascale Supercomputers [Online].

[6] Sébastien Rumley, Meisam Bahadori, Robert Polster *et al.*, "Optical interconnects for extreme scale computing systems," *Parallel Computing*, vol. 64, pp. 65-80, 2017.

[7] Alan Benner, "Optical Interconnect Opportunities in Supercomputers and High End Computing," in *Optical Fiber Communication Conference*, Los Angeles, California, 2012, p. OTu2B.4.

[8] Jürgen Jahns, Sing H. Lee, and Sing H. Lee, *Optical Computing Hardware: Optical Computing*: Academic Press, 1994.

[9] Sajjad Moazeni and Vladimir Stojanovic, *A 40Gb/s PAM4 Transmitter based on a Ring-resonator Optical DAC*: Technical Report of University of California at Berkeley, 2017.

[10] S. Palermo, P. Chiang, C. Li *et al.*, "Silicon Photonic Microring Resonator-Based Transceivers for Compact WDM Optical Interconnects," in *2015 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, 2015, pp. 1-4.

[11] A. H. K. Park, A. S. Ramani, L. Chrostowski *et al.*, "Comparison of DAC-less PAM4 modulation in segmented ring resonator and dual cascaded ring resonator," in *2017 IEEE Optical Interconnects Conference (OI)*, 2017, pp. 7-8.

[12] Raphaël Dubé-Demers, Sophie LaRochelle, and Wei Shi, "Low-power DAC-less PAM-4 transmitter using a cascaded microring modulator," *Optics Letters*, vol. 41, pp. 5369-5372, 2016.

[13] Rui Li, David Patel, Eslam El-Fiky *et al.*, "High-speed low-chirp PAM-4 transmission based on push-pull silicon photonic microring modulators," *Optics Express*, vol. 25, pp. 13222-13229, 2017.

[14] M. A. Seyedi, C. H. J. Chen, M. Fiorentino *et al.*, "Data rate enhancement of dual silicon ring resonator carrier-injection modulators by PAM-4 encoding," in *2015 International Conference on Photonics in Switching (PS)*, 2015, pp. 363-365.

[15] Jianfeng Xu, Jiangbing Du, Rongrong Ren *et al.*, "Optical interferometric synthesis of PAM4 signals based on dual-drive Mach–Zehnder modulation," *Optics Communications*, vol. 402, pp. 73-79, 2017.

[16] Alireza Samani, David Patel, Mathieu Chagnon *et al.*, "Experimental parametric study of 128 Gb/s PAM-4 transmission system using a multi-electrode silicon photonic Mach Zehnder modulator," *Optics Express*, vol. 25, pp. 13252-13262, 2017.

[17] M. A. Seyedi, Yu Kunzhi, Li Cheng *et al.*, "Silicon Mach-Zehnder Interferometer modulator with PAM-4 data modulation at 64 Gb/s," in *2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2015, pp. 1-3.

[18] A. Samani, V. Veerasubramanian, E. El-Fiky *et al.*, "A Silicon Photonic PAM-4 Modulator Based on Dual-Parallel Mach–Zehnder Interferometers," *IEEE Photonics Journal*, vol. 8, pp. 1-10, 2016.

[19] S.J. Emelett and R. Soref, "Design and Simulation of Silicon Microring Optical Routing Switches," *IEEE Journal of Lightwave Technology*, vol. 23, pp. 1800-1808, 2005.

[20] Trung-Thanh Le, *Multimode Interference Structures for Photonic Signal Processing: Modeling and Design*: Lambert Academic Publishing, Germany, ISBN 3838361199, 2010.

[21] Duy-Tien Le and Trung-Thanh Le, "Coupled Resonator Induced Transparency (CRIT) Based on Interference Effect in 4x4 MMI Coupler," *International Journal of Computer Systems (IJCS)*, vol. 4, pp. 95-98, May 2017.

PHÁT TRIỂN PHƯƠNG PHÁP ĐIỀU CHẾ PAM-4 ỨNG DỤNG CHO HỆ THỐNG KẾT NỐI, TÍNH TOÁN HIỆU NĂNG CAO VÀ HỆ THỐNG TRUNG TÂM MẠNG DỮ LIỆU

Tóm tắt: Bài báo đề xuất một phương pháp mới thực hiện điều chế 4 mức biên độ xung (PAM-4) ứng dụng cho các hệ thống kết nối quang và các mạng trung tâm dữ liệu lớn. Cấu trúc điều chế sử dụng chỉ một bộ ghép giao thoa đa mode 4 cổng vào, ra kết hợp với hai bộ dịch pha cho 2 bits thông tin. Bộ điều chế mới có ưu điểm kích thước nhỏ, băng thông cao. Toàn bộ cấu trúc của bộ điều chế có thể thiết kế, chế tạo bằng công nghệ vi mạch VLSI.

Từ khóa: *Trung tâm dữ liệu, tính toán hiệu năng cao, kết nối quang, siêu máy tính.*

Duy-Tien Le received MSc degrees of Information Systems in 2014 from Hanoi VNU University of Engineering and Technology. He is a currently PhD student of Computer Engineering, Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam. His research interests include DSPs and Photonic Integrated Circuits.

Ngoc-Minh Nguyen received PhD degree of Electronic Engineering in 2007 from La Trobe University, Australia. His research interests include DSP, FPGA, embeded systems.
He is working at Faculty of Electronic, Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam.

Trung-Thanh Le received PhD degree of Electronics and Telecommunications in 2009 from La Trobe University, Australia. His research interests include Computer Science, Laser and Optical Fiber Systems, Photonic Integrated Circuits, and Sensors. He is now with International School, Vietnam National University (VNU-IS), Hanoi.