Chapter 4

Hybrid Optoelectronic Router for Future Optical Packet-Switched Networks

Salah Ibrahim and Ryo Takahashi

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/67623

Abstract

With the growing demand for bandwidth and the need to support new services, several challenges are awaiting future photonic networks. In particular, the performance of current network nodes dominated by electrical routers/switches is seen as a bottleneck that is accentuated by the pressing demand for reducing the network power consumption. With the concept of performing more node functions with optics/optoelectronics, optical packet switching (OPS) provides a promising solution. We have developed a hybrid optoelectronic router (HOPR) prototype that exhibits low power consumption and low latency together with high functionality. The router is enabled by key optical/optoelectronic devices and subsystem technologies that are combined with CMOS electronics in a novel architecture to leverage the strengths of both optics/optoelectronics and electronics. In this chapter, we review our recent HOPR prototype developed for realizing a new photonic intra data center (DC) network. After briefly explaining about the HOPR-based DC network, we highlight the underlying technologies of the new prototype that enables label processing, switching, and buffering of asynchronous arbitrary-length 100-Gbps (25-Gbps × 4λs) burst-mode optical packets with enhanced power efficiency and reduced latency.

Keywords: optical packet switching, optical signal processing, optoelectronic devices

1. Introduction

Optical networks have been playing a pivotal role in achieving the current unprecedented capability of communications that has transformed human experience. In addition to their conventional role in the infrastructure core, metro, and access networks, optical networks are also indispensable in enabling other vital applications such as large-scale data centers (DCs) and supercomputers. With such diversity of application domains, the traffic handled by the

network varies between stable and bursty traffic. The network should thus utilize the optical transmission scheme that matches its traffic nature.

Among different schemes of optical transmission, optical circuit switching (OCS) and optical packet switching (OPS) are two basic schemes that possess complementary features. The OCS scheme allows uninterrupted data transmission where a link is established between two network nodes before starting to transmit data in between. The optical link is realized in a way similar to reserving a closed circuit and thus the scheme is entitled circuit switching. The OCS scheme is suitable for transmission of stable traffic and hence it is widely utilized in core networks. Differently, no link establishment is required with the OPS scheme as data is transmitted as individual packets in a connectionless manner. Each packet is equipped with a given label and based on that label, the packet is forwarded along network nodes until arriving at its destination. More importantly, the packets are forwarded without going through optical-electrical-optical (OEO) conversion in a real photonic network. This feature together with the elimination of the link establishment time makes the OPS scheme very suitable for handling bursty traffic.

On the other hand, the capacity of optical links has been significantly boosted [1] by using higher data rates and complex data formats, which increases the burden on current network nodes that are mostly relying on electrical packet switching (EPS). The resulting extensive dependence on electronic processing is the reason for several shortcomings that are difficult to overlook such as the high power consumption and end-to-end latency. The EPS approach can be also identified as a limiting factor for the network scalability which is the problem faced by current large-scale DC networks. This condition can be overcome by realizing a photonic network instead, and in this regard, we have proposed a new photonic intra DC network based on the hybrid optoelectronic router (HOPR).

In this chapter, we review our recent HOPR prototype developed for realizing the DC network. We briefly explain about the HOPR-based photonic DC network, and highlight the underlying technologies of the new prototype that enable label processing, switching, and buffering of asynchronous arbitrary-length 100-Gbps (25-Gbps × 4λs) burst-mode optical packets with enhanced power efficiency and reduced latency.

2. New photonic data center network

DCs have considerably evolved to become main players in the big-data era, providing a diversity of services with unprecedented volumes of data traffic [2–4]. However, as mentioned earlier, current intra DC networks that are mainly based on EPS have been facing increasing difficulties to cope with the growing demand. The advancement of the EPS-based DC networks has continuously relied on the progress of the CMOS and transceiver technologies. But these technologies have already reached an advanced level after which it becomes difficult to achieve further improvement at the quick pace required for fulfilling the ongoing demands [5]. Realizing a photonic DC network is a radical solution that can let DCs surpass their current difficulties. To serve this end, we have proposed the photonic intra DC network

http://dx.doi.org/10.5772/67623

[6] illustrated in **Figure 1**. The network has a torus topology and it depends on the deployment of HOPRs and a centralized network controller. More about the key network aspects can be found as follows.

- **Basic operation:** A HOPR unit is located at each node of the torus DC network and connected to neighboring HOPRs via optical links. High-speed burst-mode optical packets are transmitted over these links. Each HOPR unit is also connected to a group of Top-of-Rack (ToR) switches that handle Ethernet packets from the servers connected to them. An Ethernet packet is transformed into a burst-mode optical packet at the corresponding HOPR unit and from there on it is transmitted in the optical domain throughout the torus network until it reaches the HOPR unit attached to the destination server. Thus, each HOPR unit has a twofold role as an optical-packet switch that forwards the optical packets of the torus network and an aggregation switch that handles the servers' packets.
- **Topology:** A highly dimensional torus topology is considered for realizing the network. This topology has been widely adopted in supercomputers such as the CRAY (3D), Blue Gene (3–5D), and K Computer (6D). It enables a highly scalable network that can strongly support fault tolerance with redundancy of links. The torus topology also exhibits several features that are advantageous for the OPS scheme; as (1) it is a direct topology in which a low-radix switch is sufficient for deployment at each network node, (2) the uniform arrangement of the network nodes allows the utilization of a simple deterministic algorithm for forwarding incoming packets, and (3) the presence of multiple equidistant routes between the source and destination nodes with the same latency can efficiently help in resolving packet contention when it occurs.
- **Transmission schemes:** Three data transmission schemes are supported by the photonic DC network; namely the OPS, OCS, and virtual V-OCS schemes. Unlike current DC networks, the OPS scheme is used to allow latency-sensitive applications over a wide network scale without being limited to nodes in close proximity.

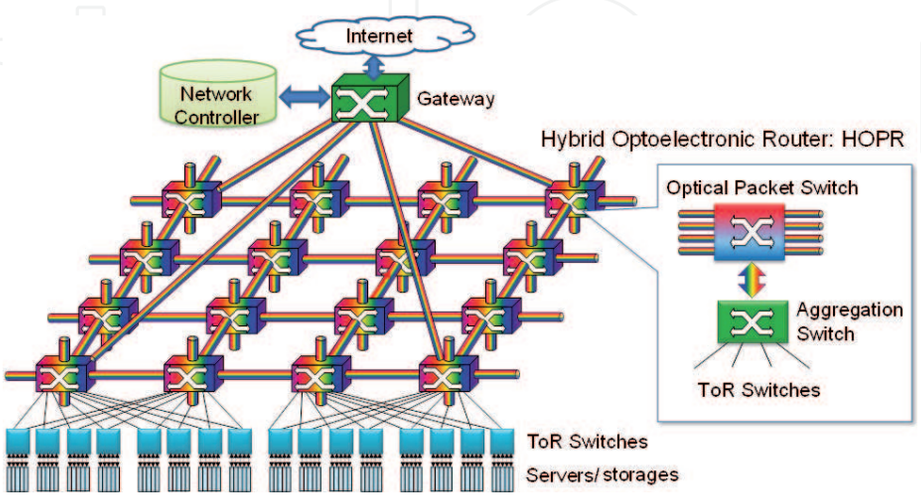

Figure 1. Illustration for the torus DC network based on HOPR and NW controller.

However, an OPS packet can reach its destination via different routes and might encounter collision with other packets and thus the probability of packet loss cannot be completely avoided. For the OCS case, an optical path is reserved before transmission is started to avoid packet loss and the sent packets arrive with the same order and same latency. However, as the number of OCS paths increases, the normal OPS packets are strongly counteracted due to the increasing difficulty of finding a vacant route to reach their destination. Hence, the virtual OCS scheme has been devised [7, 8] to enable the coexistence of the OPS and OCS packets almost without counter effects. To realize such DC network, we have developed a new HOPR prototype that can meet the demand of handling burst-mode optical packets without OEO conversion unless necessary.

3. Hybrid optoelectronic router

In this section, we highlight HOPR's operation, architecture, and implementing approach. **Figure 2a** shows HOPR unit for a 3-D torus network, and **Figure 2b** shows a 6-D network where a 16 × 16 optical switch is used and the shared buffer interface is upgraded to 200 Gbps allowing the aggregation of more servers' traffic. Using the WDM configuration as shown in **Figure 2c**, the link capacity can be increased as the product of the number of wavelength layers and data rate of generated packets to allow achieving 0.4–1 Tbps.

When an optical packet arrives at HOPR, the destination address is recognized by the label processor (LP) while keeping the packet in the optical domain. The optical switch is then configured to forward the packet via the desired output switch-port. If the port is occupied, a contention resolution plan is followed to resolve the condition. At the torus DC network, the packet can be forwarded to another output switch-port and still goes through one of the shortest routes toward its destination, that is, deflection routing. HOPR is also equipped with optical and optoelectronic buffers to help with resolving contention, where (1) the packet can have a fixed time delay by going through a fiber delay line (FDL) without OEO conversion, or (2) it can be electronically stored in the optoelectronic buffer for an arbitrary storage time.

Unlike most conventional electrical routers that adopt the store-and-forward mechanism for packet forwarding, HOPR relies on the cut-through mechanism where the optical switch is configured once the packet label is recognized without demanding the whole packet to be received first. This matches well with forwarding the packets without OEO conversion. The packet path via HOPR only includes the LP and optical switch, whereas the shared buffer is located in parallel without obstructing the normally forwarded packets. The role of these three basic functional units is summarized in the following.

- **Label processor:** The LP's role is divided into: (1) extraction and recognition of the incoming packet label, (2) deciding the output switch-port accordingly, (3) performing arbitration for colliding packets, and then (4) configuring the switch with the corresponding control signals. On the other hand, the LP should be realized with (1) low power consumption to enable a widely scalable network, (2) low latency as the packet remains in the optical domain and the label processing time is compensated for just by delaying it, (3) a sufficiently high dynamic range to exhibit tolerance for power variation among the incoming packets.

http://dx.doi.org/10.5772/67623

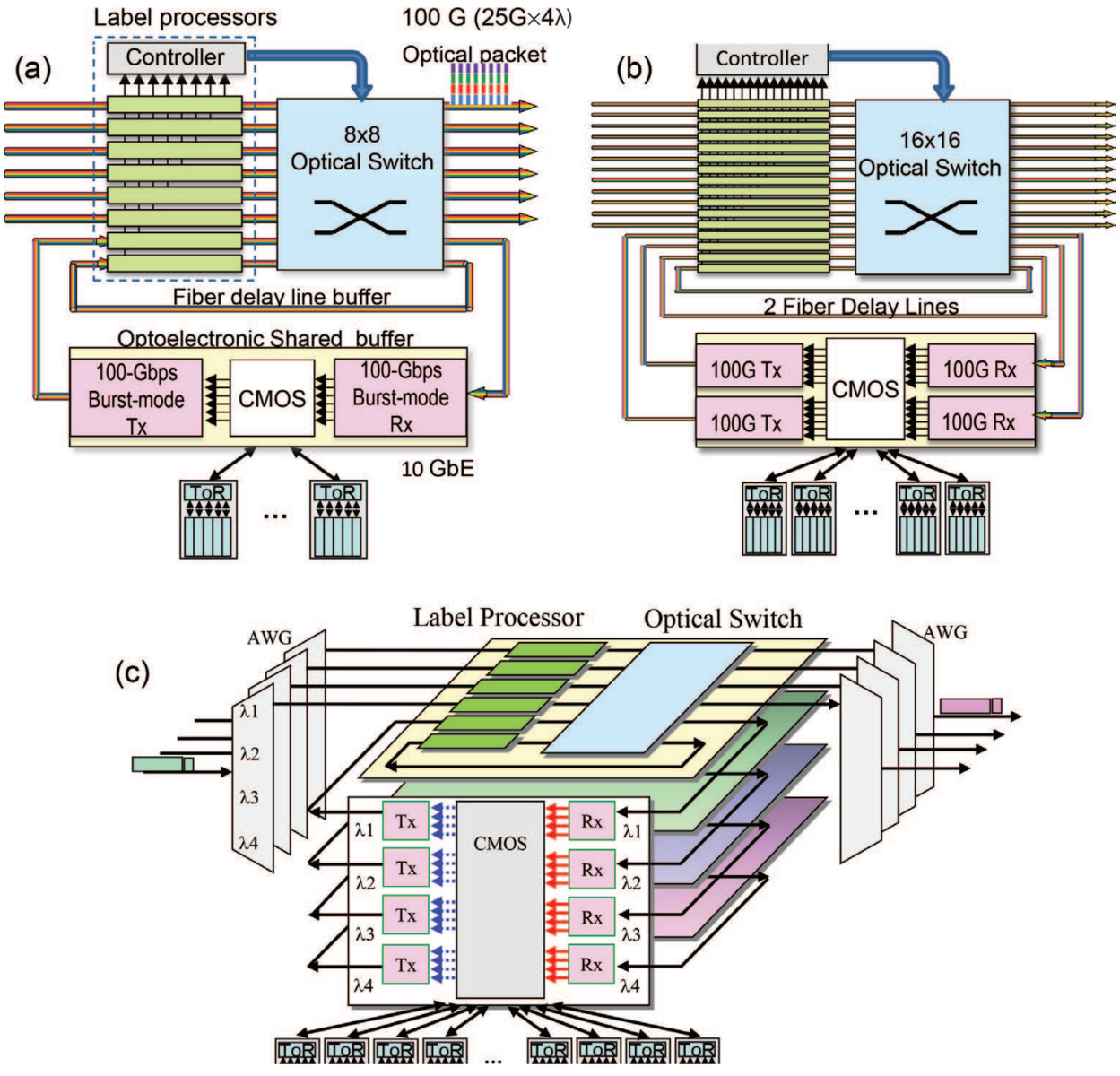

Figure 2. HOPR's architecture and basic subsystems in different configurations.

- **Optical switch**: There is a general trade-off between the switch speed and its maximum port count. For example, mechanical switches such as MEMS-based switches [9] can exhibit large port-counts, that is, over 1000 ports, but their reconfiguration time lies in the millisecond order. Differently, the OPS switch should operate in the nanosecond order; and with current technologies, only low port-count switches, for example, 16 × 16 ports, are available. A fast switch with low port-count can still enable realizing a highly scalable network with good performance if a suitable network topology is selected. It is beneficial to realize a switch with independent characteristics in terms of bit-rate, packet format, wavelength, and polarization. The switch is also demanded to possess low power consumption, high extinction ratio, low crosstalk, and ease of controllability. To meet these demands, several switch types have been considered by researchers [10–12]. Examples include the matrix switch that consists of cascaded smaller switches,

phased-array switch, wavelength routing switch, and broadcast-and-select (B&S) switch. Among them, the B&S switch is selected as it meets most of the requirements to a good extent. We have also considered wavelength-routing switches to achieve a switch with higher port-counts.

- **Optoelectronic shared buffer:** Realizing an optical memory has attracted a lot of research interest [13, 14] but a reliable solution is still missing. Available solutions are either passive all-optical buffering via an FDL [15] for a fixed duration or electrical buffering where the optical packets are interfaced to an electronic memory. The role of shared buffer exceeds the demand for packet storage as it is essential for resolving contention and for enabling higher network functionalities such as packet regeneration, Quality of Service (QoS) control, and format conversion. Another important role for the optoelectronic buffer is interfacing the photonic network to other transmission domains with different data formats and data rates, where for the DC network case, it should act as an aggregation switch for the servers Ethernet packets.

Unlike conventional electrical routers handling continuous data streams to prevent clock loss at the receiver side, HOPR should operate in a burst-mode fashion where a packet suddenly arrives after a period of no received signal. Knowing the bit timing of the incoming packet is essential, but there is no prior synchronization between the incoming packets and HOPR. To enable clock recovery, burst-mode routers rely on preamble bits [16] that precede the packet, where a conventional way for clock recovery is used such as the phase locked loop (PLL) method that demands a long locking time [17], or the phase picking method [18] that is widely selected for handling lower data rates (~10 Gbps). The over-sampling method [19] is another alternative, where a clock recovery time of 31 ns has been recently reported [20] by combining this method and an approximation algorithm. However, the continuous increase of data rates reduces the packet duration, and keeping the dependence on preamble bits will degrade the efficiency of utilizing the optical link. Thus to allow HOPR to handle preamble-free optical packets, special burst-mode optoelectronic devices are developed to interface optical packets to electronic circuits with a novel optical clocking (triggering) method.

HOPR prototype (**Figure 2a**) is developed with (1) total throughput of 1.28 Tbps, where six input/output ports handle 100-Gbps optical packets and four input/output ports handle the 10-GbE connections, (2) total power consumption of 110 W, with ~40 W for the optical packet switch part that includes 8 LPs and an 8 × 8 optical switch, and ~70 W for the aggregation switch (optoelectronic shared buffer), including the contribution of the control plane, cooling, GUI, 10-GbE transceiver modules, and so on, and (3) latency of 140 ns, with ~60 ns for the transmission delay via the optical switch and EDFA, and ~80 ns for the LP dominated by the arbitration time for resolving contention. Electrical switches have also been significantly enhanced, but still they require a Network Interface Card (NIC) and optical transceiver module at each port. The CFP 100-GbE optical transceiver, for example, demands 6–20 W for different transmission ranges. HOPR's optical packet switch part consumes 5 W/100-G-port which is lower than a single CFP module. The future increase in packets data-rates, the power of optical packet switch would remain almost unchanged. The power consumed by the

http://dx.doi.org/10.5772/67623

shared buffer can also be further reduced if instead of combining a Field Programmable Gate Array (FPGA) and discrete electronic components, the state-of-the-art Application Specific Integrated Circuit (ASIC) technologies are used in a way similar to the electrical switches.

As the name implies, the hybrid optoelectronic router is based on a hybrid implementation approach, where both electrical and optical technologies are used, each where it fits most. This approach has been vital for fulfilling the basic demand of reducing HOPR's power consumption and latency without sacrificing its performance. The hybrid approach can be seen at HOPR's underlying device level, where, for example, the operation principal of some essential devices is based on using optical timing pulses, that is, clock pulses, for triggering electronic circuits to enable handling the packets' ultrafast bits in an efficient way. Moreover, the hybrid approach can also be seen at the subsystem level as for instance in the presence of different packet buffering options, represented by the all-optical FDL-based buffer and the optoelectronic buffer. These alternatives provide the flexibility demanded to efficiently cope with diversified conditions of operation.

4. Optical label processing technologies

In this section, we explain about our latest LP [21] that handles 25-Gbps burst-mode optical labels without preamble bits. **Figure 3** shows the structure of the LP that consists of a set of label extractors (LE) connected to a shared controller.

At each input port of HOPR, a split from the incoming packet is directed to the attached LE, whereas the main part goes toward the optical switch but first passes through an FDL to compensate for the label processing time. Unlike traditional Serializer/Deserializer (SerDes)

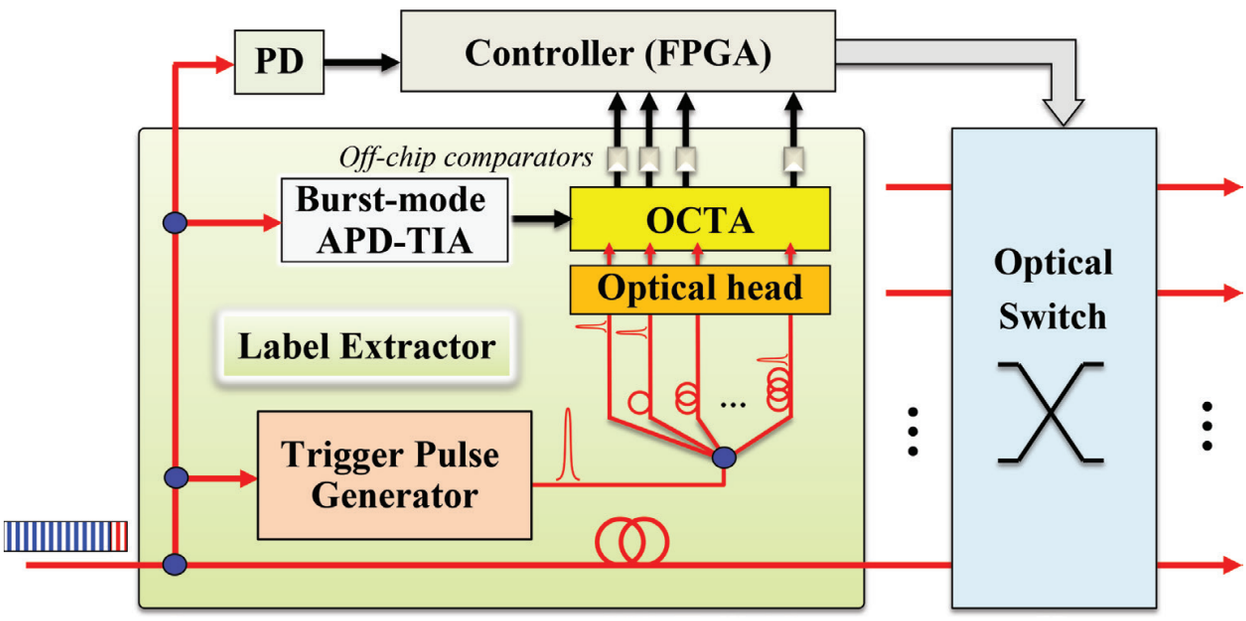

Figure 3. The overall structure of the burst-mode label processor.

that converts the bits of the whole packet, the LE handles only the packet label. The label's ultrafast bits are separated in parallel and their speed is sufficiently slowed down by undergoing serial-to-parallel conversion to allow their direct latching into the CMOS based LP's shared controller. At the LP controller, the forwarding table is looked up with the extracted label and when a match is found and the corresponding output port of the switch comes out as the result. The occupancy of the resulting output port is examined; and if the port is free, the control signals necessary for configuring the optical switch are generated and applied, whereas if the port is occupied, the second priority output is selected and so on until finding an available port. To enable label processing of preamble-free packets, the LE adopts an operating mechanism that relies on two basic elements: (1) a burst-mode serial-to-parallel converter (SPC) that operates once supplied by optical triggers and it is referred to as the optically clocked transistor array (OCTA), and (2) a burst-mode optical trigger pulse generator (TPG) that selectively utilizes the first bit of incoming packet to produce a synchronized optical trigger pulse for OCTA.

4.1. Optical trigger pulse generator

The TPG produces an optical trigger pulse by selectively amplifying the first bit of incoming packet with a semiconductor optical amplifier (SOA). Being originally a part of the incoming cket, the resulting optical pulse is a synchronized trigger that allows jitter-free serial-to-parallel conversion when applied to OCTA. **Figure 4a** shows the SOA's driver integrated circuit developed to drive a narrow current pulse (~1 ns) of high peak current (>600 mA).

A split from the incoming packet as shown in **Figure 4b** is applied to the metal-semiconductor-metal (MSM)-PD at the driver's discharge-based (DB) circuit and produces a single electrical pulse that undergoes reshaping before turning on a set of integrated high electron mobility transistors (HEMTs) shortly to enable the flow of high electric current through the SOA. The red curve in **Figure 4c** shows the normalized amplified spontaneous emission (ASE) of the

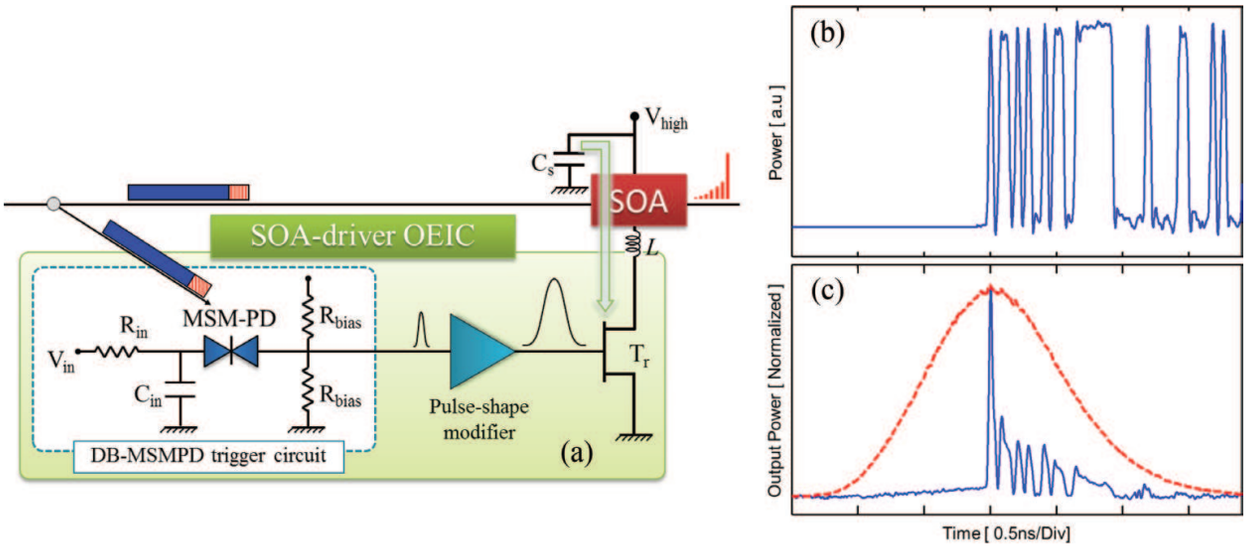

Figure 4. (a) The TPG circuit, (b) incoming packet, and (c) SOA output.

SOA that corresponds to its gain profile when the current pulse of the driver is applied to the SOA without having it supplied with any optical input. The packet arrival at the SOA is roughly adjusted to be around the gain peak. The packet's first optical pulse experiences strong amplification and quenches the accumulated carriers at the SOA, and the following pulses fade out due to lack of gain as shown in **Figure 4c**. As the SOA's power dissipation is limited only to a short instant after packet arrival, the power consumption of the TPG has been significantly reduced to 120 mW compared to an older TPG [21] that consumed 3 W.

The DB circuit [22] is a fundamental element in our burst-mode devices, used for converting an optical trigger signal into a narrow electrical pulse that can control the gate of a HEMT transistor. It utilizes an MSM-PD that has a fabrication advantage due to its surface structure. Unlike PIN photodetectors whose response is limited by an RC time constant, the MSM-PD has a very quick rise-up time only limited by the electron transit time due to its ultralow capacitance. However, the MSM-PD response suffers from a long tail due to the low mobility of holes. The discharge-based configuration is thus employed to allow the MSM-PD to generate a sufficiently narrow electrical pulse. The input capacitor C_{in} of the DB circuit is initially charged maintaining a high bias voltage. When an optical pulse is then applied to the MSM-PD, it causes the flow of photocurrent. A corresponding electric current cannot be injected by V_{in} due to the high resistance R_{in}, and thus C_{in} is discharged resulting in the reduction of the bias voltage. One important feature of the MSM-PD is that in the absence of bias voltage, the current cannot flow even when the carriers are still present. The first pulse of incoming packet almost depletes the carriers of C_{in} leaving the following packet bits ineffective. Thus even with the MSM-PD direct irradiation with the whole packet, the DB circuit can produce a single short electrical pulse.

4.2. Optically clocked transistor array

Figure 5 illustrates the structure of OCTA which is a monolithically integrated circuit that consists of 16 serial-to-parallel conversion channels attached to a common transmission line (TL) each via a separate HEMT (T_m).

An avalanche photodetector and trans-impedance amplifier (APD-TIA) burst-mode module is used to convert the packet split at the LE unit (**Figure 3**) into an electrical signal that is then coupled to OCTA and propagated along its TL. The channels are used in turn, each for converting a different label bit. The timing of a separate optical trigger pulse is adjusted to match the presence of a given bit at the TL, and the bit is converted by applying that trigger pulse to the respective channel. OCTA's correct operation demands T_m to be turned on shortly to convert only a single bit, and hence the DB-MSMPD trigger circuit is used to produce a single narrow electrical pulse to control T_m. OCTA-based serial-to-parallel conversion was initially done by the sample-and-hold (S&H) scheme as shown in **Figure 5a**, where an electric charge corresponding to the considered bit level is sampled into the capacitor C_{hold} through T_m. The voltage-change induced at C_{hold} is then amplified to produce the channel's final output. This SPC scheme suffers the limited difference between the sampled charge that corresponds to the "1" and "0" bits, respectively. The reason is that charging C_{hold}, that is, in case of a "1" bit, increases the voltage at T_m source terminal and hence forces it to get turned-off. Moreover,

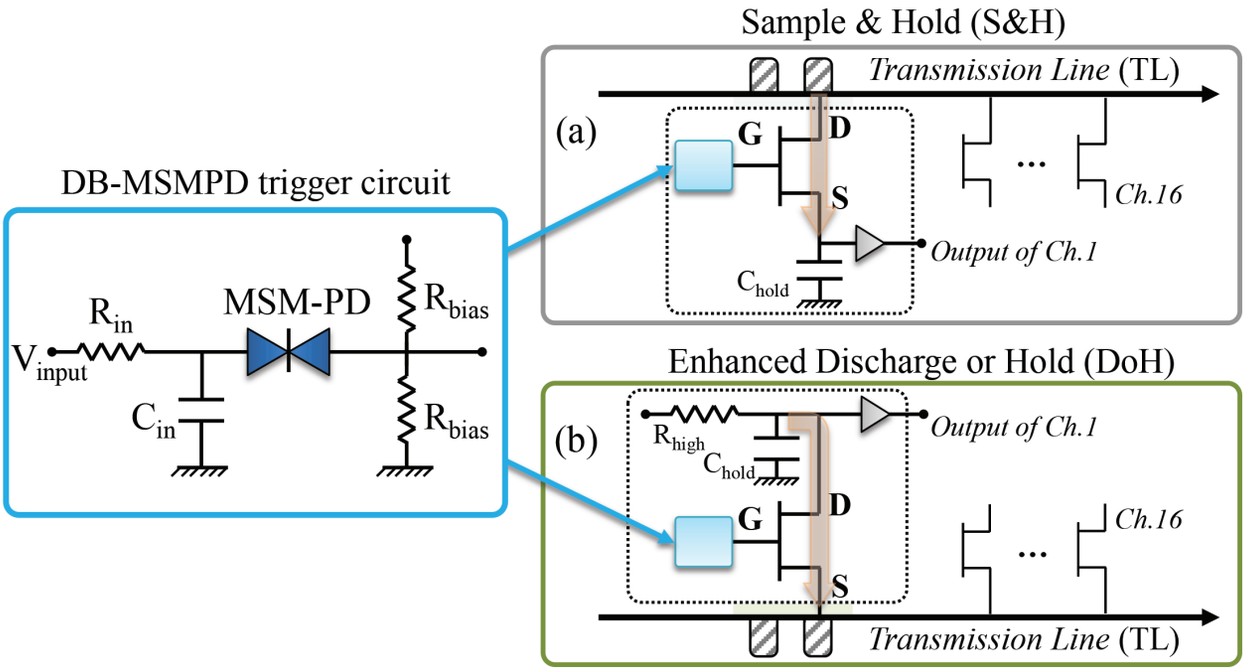

Figure 5. The structure of OCTA and illustration of its operation schemes (a), and (b).

charging C_{hold} cannot be done efficiently as a part of the bit voltage is unavoidably dissipated in the TL's characteristic impedance that is present in parallel to the turned-on conversion channel. The S&H scheme was then replaced by the discharge-or-hold (DoH) scheme [21] illustrated in **Figure 5b**, where the charge initially present at C_{hold} is either discharged into the TL or kept unchanged. Discharging C_{hold} is done more efficiently than charging it, and a much higher difference of charge is produced at C_{hold}. The benefit of the DoH scheme has been further elevated by using a label signal with negative voltage span, that is, negative voltage for a "1" bit and zero voltage for a "0" bit.

Figure 6 shows a comparison for using the DoH scheme with the TL signal having either a positive or a negative polarity. The gate voltage signal generated by the optical trigger pulse is shown in solid line. When the transistor T_m is turned on as V_{GS} exceeds the threshold voltage V_{th}, the bias voltage between its drain and source terminals, that is, ΔV_{DS}, is obviously higher in case of negative polarity. This allows a more efficient discharge for C_{hold} with the higher electrical current enabled by the higher ΔV_{DS}. Then if the energy of optical trigger pulse is reduced, a corresponding reduction in the gate pulse amplitude takes place as highlighted by the dotted line. Even with this reduction in ΔV_{GS}, the initially higher value of ΔV_{DS} enables conversion as efficient as in the case of a positive polarity signal with unreduced optical trigger energy.

4.3. Enhanced packaging

Optoelectronic integrated circuits (OEICs) are attractive for their high-speed operation and low power consumption. To make the best use of these key features, the optical trigger pulses

http://dx.doi.org/10.5772/67623

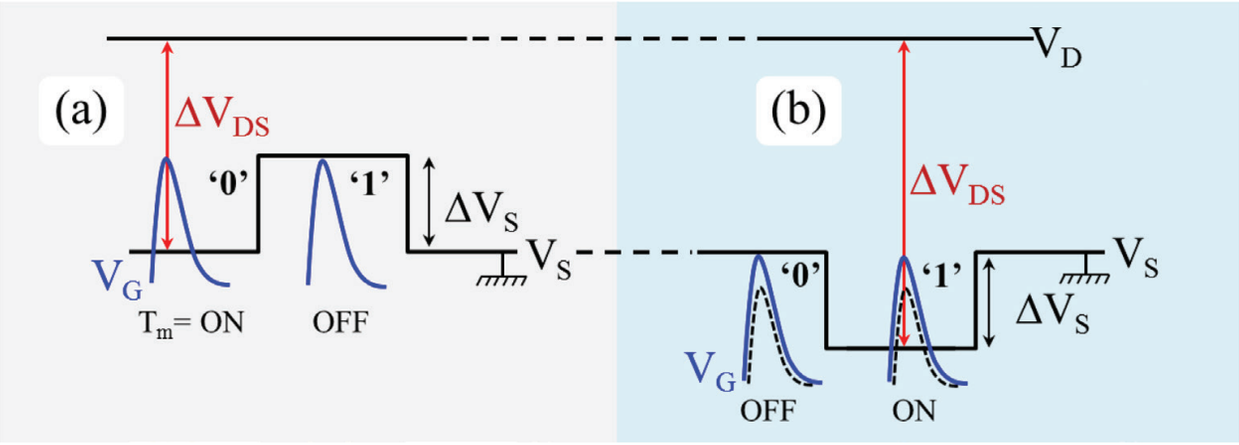

Figure 6. The DoH scheme with TL signals of (a) positive and (b) negative polarities.

necessary for OEIC operation should be provided with suitable device packaging. The older way of device packaging was complicated and costly, and thus it had to be revised.

Figure 7a shows a photo for the old packaged OCTA, where a set of lens array is used to focus the optical triggers on the MSM-PDs located at the front side of the chip. The old packaging method was wasting the trigger energy due to the shadowing effect that occurs for front-illuminated light by the MSM-PD's interdigitated metal electrodes [23]. This method also had other shortcomings such as the need for active alignment and an expensive lens array. To overcome these issues, a new method was developed as shown in **Figure 7b**. The optical head employed is a commercially available fiber-array block, where SMF fibers are placed in grooves with the same pitch as the chip's MSM-PDs. The chip is directly attached to the optical head after reducing its thickness to ~130 μm. The high refractive index of the chip's InP substrate limits the divergence of the optical beams launched from the SMF fibers. The $1/e^2$

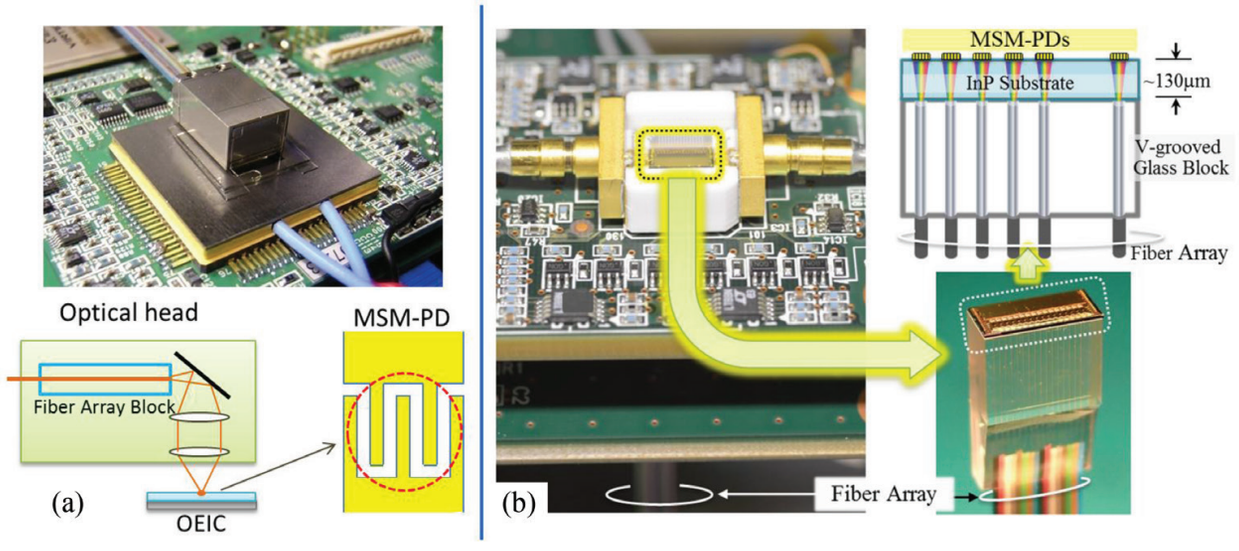

Figure 7. (a) Old packaging of OCTA versus (b) new packaging.

diameter of the optical spot resulting at the MSM-PD is ~13 μm, which is narrow enough to fit into the 20-μm active area of the MSM-PD without using lenses and while still exhibiting a large misalignment tolerance. Another key feature of this method is that optical triggers are supplied via the backside of the chip. With back illumination, the responsivity of the MSM-PDs is almost doubled and the trigger pulse energy necessary for performing serial-to-parallel conversion for 25-Gbps label bits has been reduced to 0.35 pJ/bit.

When an optical signal is applied to the MSM-PD via its backside, a part of the signal goes though the MSM-PD and can still be seen from the chip's front side. A simple alignment process has been developed by making use of this feature, where the optical beams coming out of the MSM-PDs of all the conversion channels are simultaneously observed by using an infrared camera. The best alignment position is achieved simply when a full set of clear optical spots is visually observed.

5. Optical switching technologies

In this section, we present two types of fast optical switches that we have been considering, namely the wavelength-routing switch and the broadcast-and-select switch.

5.1. Wavelength routing switch

Figure 8a illustrates the basic architecture of an N × N optical switch that operates with the wavelength routing mechanism. The switch consists of a cyclic arrayed waveguide grating (AWG) equipped with a tunable wavelength converter (TWC) at each input port and a fixed wavelength converter (FWC) at each output port. By changing the wavelength of the optical signal input to the AWG with the TWC, the signal can be directed to a different output port, whereas the FWC is used to return the signal back into its original wavelength. Based on the well-established AWG technology, it is feasible to realize such switch with a medium port-count, for example, 64 × 64 ports, provided that the used TWC can cover the whole C-band. The full exploitation of this switching method demands a TWC with independent characteristics in terms of modulation format and data-rate, but it is still hard to realize such TWC.

Currently, the most reliable TWCs are based on signal regeneration with OEO conversion where the wavelength of a TL is changed to match the desired output port, and the data of incoming packet is converted into an electrical signal that is then used to modulate the new wavelength. The OEO-based TWC does not allow the switch usage for handling WDM packets or coherent packets.

The components required for realizing the OEO-based TWC are (1) a fast tunable laser for generating an optical carrier with desired wavelength, (2) a burst-mode photodetector and an electrical amplifier for converting the incoming optical packet into electrical signal with sufficient amplitude, and (3) a modulator for modulating the desired wavelength with the electrical packet signal. We have developed a parallel-ring resonator-based tunable laser (PRR-TL) that is integrated with an electro-absorption modulator (EAM) on the same InP chip. The

http://dx.doi.org/10.5772/67623

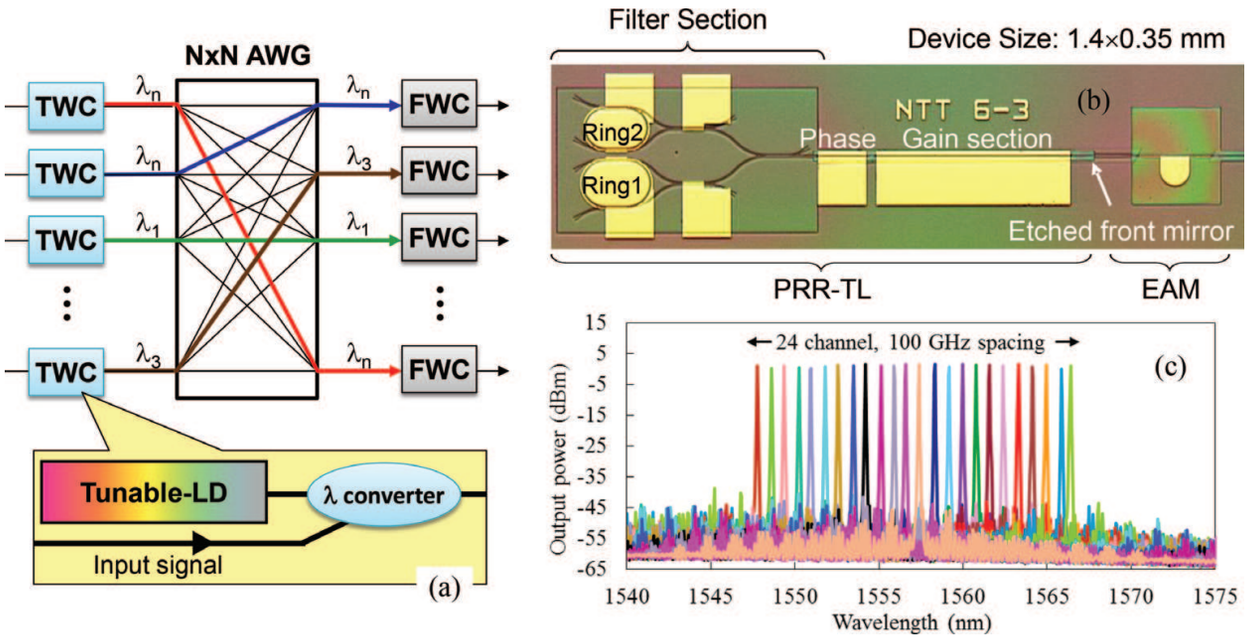

Figure 8. (a) Illustration for wavelength-routing switch, (b) photo of tunable transmitter chip, and (c) superposed spectrum of the PRR-TL.

PRR-TL relies on the Vernier effect for providing a wide tuning range [24]. The chip's photo is shown in **Figure 8b** where the TL consists of a phase section, gain section, and parallel-ring resonator (PRR). The TL has a partially reflective output port, whereas the PRR section is the other end of its laser cavity that selectively reflects the optical signal with desired wavelength. The wavelength is tuned with low current injection, that is, mA order. This enables the PRR-TL to exhibit a very low wavelength drift of less than 5 GHz. The low current injection and parallel ring design results in a laser cavity with low optical loss. This allows the PRR-TL to exhibit an output power variation of less than 1 dB over a wide tuning range of 35 nm as shown in **Figure 6c**. Moreover, it also allows the TL to generate a high output power while exhibiting a high-speed wavelength tuning of less than 6 ns.

5.2. Broadcast-and-select switch

As the wavelength-routing switch is not suitable for handling WDM packets (25-Gbps × 4λs), we have considered the B&S switch instead. **Figure 9a** illustrates the basic structure of the B&S switch, where an optical 1 × N passive splitter is located at each input port, with an optical gating unit placed at each of the splitter outputs. On the other hand, an optical N × 1 coupler is placed at each output port, with each coupler input connected to one output of the splitter that serves for a different input switch port. The packet undergoes switching by turning on the optical gate that can lead it to the desired output port.

- **Optical gate:** SOAs are widely selected as optical gating units in the B&S switches as they can provide optical gain to compensate for the optical loss while unavoidably adding some ASE noise. SOAs exhibit pattern-dependent effects for high bit-rate packets and inter-channel crosstalk for DWDM-based packets due to four-wave mixing and cross-gain modulation,

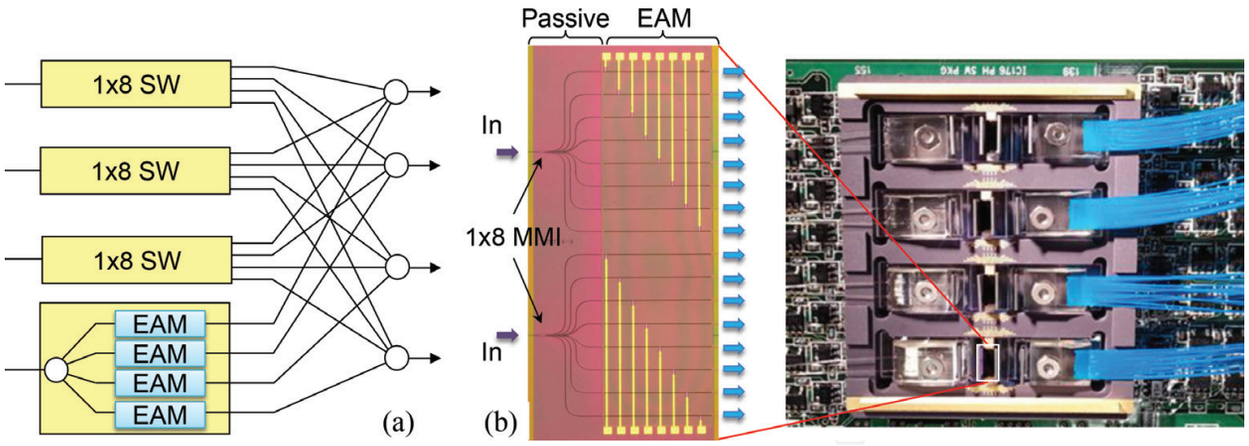

Figure 9. (a) Illustration for the B&S switch and (b) realized switch with inset of fabricated chip.

and they also require high-speed drivers with large current modulation which leads to high power consumption. Instead of SOAs, we have considered EAMs as optical gates, and the full switch system requires an EDFA for loss compensation at each output port. But yet still the total power consumption is lower than that of the SOA-based switch as the power consumption of an EDFA can be reduced (~1 W) by optimizing it for a given gain range (~25 dB). In addition, the power consumption of the reverse-biased EAM and its electrical driver is very low compared to the SOA.

- **Device and module**: **Figure 9b** shows the device photo for two monolithically integrated 1 × 8 EAM gate switches. The EAM section has a shallow-ridge waveguide structure and the passive waveguide section including the 1 × 8 multimode interference splitter is formed as a deep-ridge to reduce the required bending radius. The bulk InGaAsP (λ_{gap} = 1.4 μm) used for the absorption layer of the EAM section operates based on the Franz-Keldysh (FK) effect. To simplify the fabrication process, the EAM and passive sections share the same core/cladding layers, in which EAMs are not electrically isolated from each other. Thus when voltage is applied to a given EAM, an undesirable change is caused in the output power of other switch ports due to the electro absorption by FK effect induced in the passive waveguide section. One way to eliminate such electrical crosstalk is to replace the p-InP cladding layer by a high-impedance material such as Fe-doped InP. However, this would demand an additional step of regrowth process. To avoid that, a surface-ground electrode is used to cover the MMIs of the passive waveguide section. The simple addition of the surface ground electrode has been successful in suppressing the electrical crosstalk [25].

A set of micro-lenses is used for the optical coupling between an optical fiber array and the switch waveguide array. High coupling loss and loss variation among different waveguides were observed due to the large and asymmetric numerical aperture of ridge waveguide (NA ~ 0.8). To improve coupling loss and alignment tolerance, spot size converters (SSCs) were then added at both switch sides to decrease the waveguide's NA by expanding its optical mode field. The core size is set to 0.4 μm × 0.4 μm to achieve a symmetric NA of 0.5. The fabricated module exhibits low WDL and PDL of typically ±0.5 dB in the wavelength range of

http://dx.doi.org/10.5772/67623

1540–1560 nm. The module is attached to an electrical control board including an FPGA and EAM drivers that enable a switching rise/fall time of less than 10 ns. With a reverse bias of ~7 V, all the output ports exhibited extremely high extinction ratios of more than 50 dB, which is sufficient to avoid any signal degradation caused by inter-symbol interference.

6. Optical buffering technologies

HOPR comprises an FDL-based all-optical and optoelectronic shared buffer. This section is devoted to the optoelectronic shared buffer where packet processing is performed to enable higher network functions such as packet regeneration, QoS control, and format conversion.

6.1. Operation

Figure 10 illustrates the optoelectronic buffer used to handle 100-Gbps WDM optical packets composed of four wavelengths each modulated at 25 Gbps. A CMOS processor is located at the buffer core, and it can independently handle the servers' Ethernet packets delivered through the ToR switches, and the network OPS packets. The OPS packets are fed into and out of the processor by using a set of burst-mode serial-to-parallel converters (SPCs) and parallel-to-serial converters (PSCs), respectively. The incoming packet four wavelengths are de-multiplexed with an AWG at the buffer input. Each packet part belonging to a different wavelength is converted into fast electrical signal by using an APD-TIA module. Each two

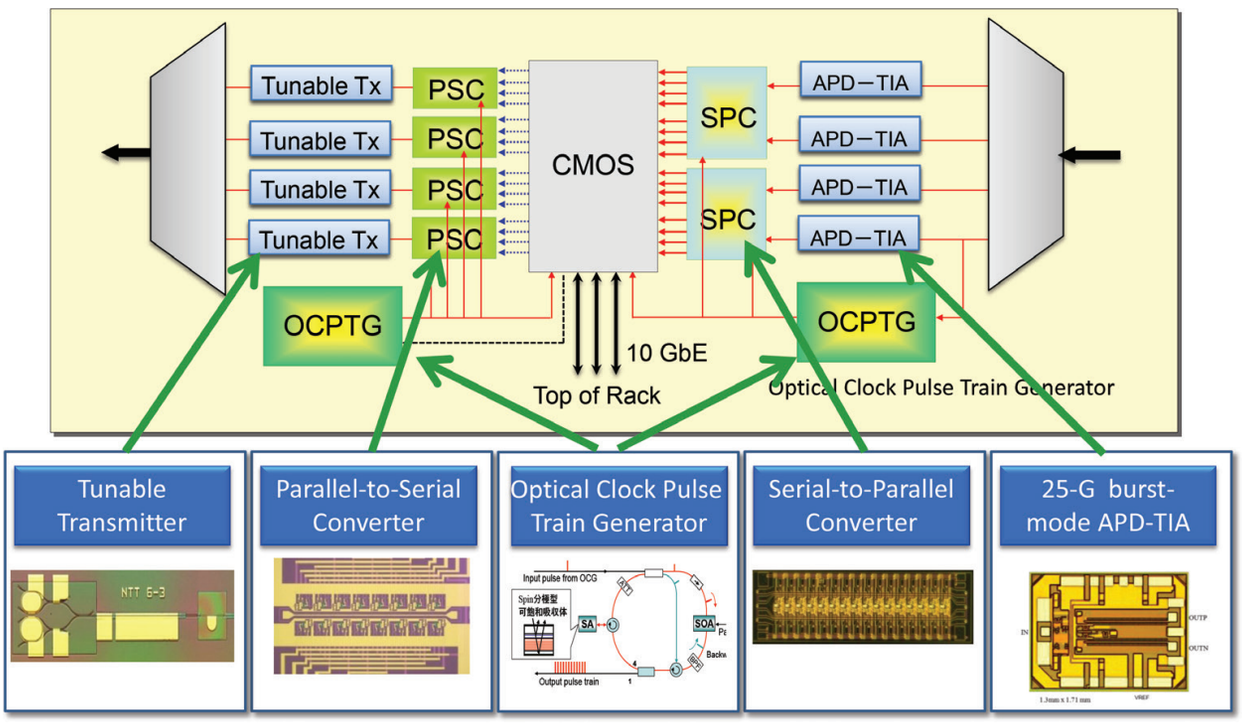

Figure 10. The optoelectronic shared buffer that can handle burst-mode WDM packets.

packet parts are then simultaneously fed into the same SPC device, where the speed of their ultrafast bits is reduced by the ratio of 1:16.

The two SPC devices are supplied with optical clock pulses to keep them operating until the serial-to-parallel conversion is done for the whole packet. The optical clock pulse train generator (OCPTG) module present at the buffer input side provides the optical clock pulses necessary for operating the SPCs and also the optical clock pulses required to latch the SPCs output to the CMOS processor.

To generate a 100-Gbps optical packet, four different PSCs separately receive the electrical data for each wavelength from the processor, and operated in parallel to reduce the speed of the data bits by the ratio of 16:1. Each PSC output is used at a separate tunable transmitter that generates an optical carrier adjusted to a desired wavelength. After modulation, the four wavelengths are multiplexed by an AWG to form the output 100-Gbps optical packet. The OCPTG module at the buffer output side provides the optical clock pulses required for reading out data from the processor and for operating the PSCs.

6.2. Enabling devices

The devices necessary for realizing the shared optoelectronic buffer are briefly highlighted here, and more details can be found in the corresponding references.

- **OCPTG:Figure 11** illustrates the OCPTG structure that has an optical clock generator (OCG) at the input followed by the pulse train generator (PTG)'s optical loop that includes an SOA and spin-polarized saturable absorber (SA) with high-extinction ratio. The OCG generates a single optical pulse in response to the first bit of incoming packet. This pulse is then fed into the PTG module to produce a train of optical pulses with fixed separating intervals between pulses throughput the whole length of the packet. The OCPTG [26] can thus handle preamble-free asynchronous optical packets with variable lengths.

The optical pulse train is generated by tapping out a portion of the circulating seed pulse initially provided by the OCG. The loop is made to have a round-trip gain function with

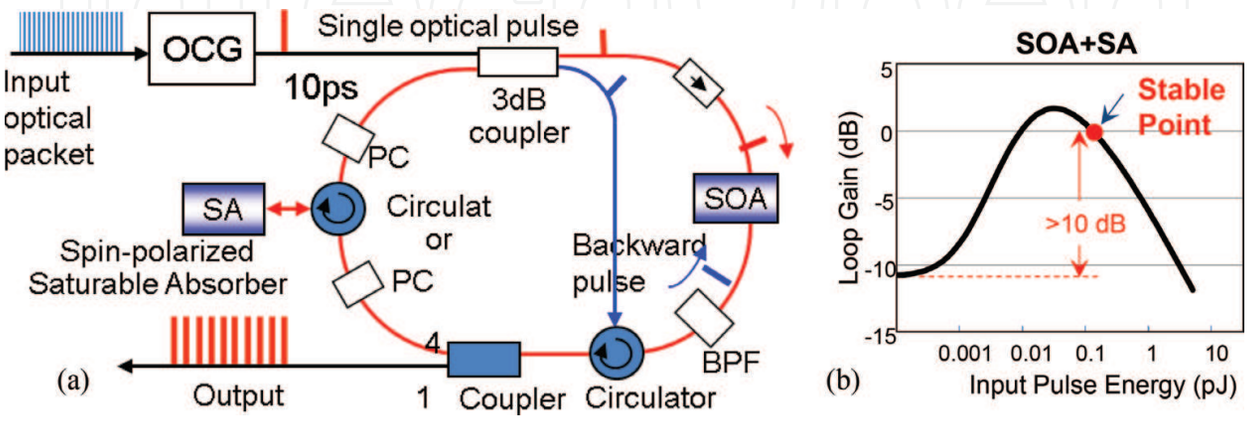

Figure 11. (a) The structure of OCPTG and (b) loop transfer function of PTG.

http://dx.doi.org/10.5772/67623

negative slope based on the SOA gain saturation (**Figure 11b**). This allows the circulating pulse energy to converge even if some fluctuation occurs, where the gain and absorption of the SOA and SA, respectively, finally balance out. The ASE of the SOA is suppressed by the SA. The SA is characterized by fast absorption recovery due to spin relaxation, but still after the pulse comes out of the SA, a backward optical pulse is launched into the SOA to prevent the increase of the ASE inside the loop with the recovery of SOA gain. Without any external control for the loop gain, a stable optical pulse train with low jitter is generated from a single 10-ps seed optical pulse even with a 10-dB variation in the seed pulse intensity.

- **APD-TIA module:** The high-speed avalanche photodetector (APD) and burst-mode transimpedance amplifier (TIA) module have been newly developed [27]. The APD has a new p-down inverted structure [28] to prevent undesirable edge breakdown and to suppress surface leakage current. A back-illuminated InAlAs/InGaAs APD based on this structure exhibited a 3-dB bandwidth of 27 GHz. On the other hand, the TIA was designed and fabricated by using 0.13-μm SiGe BiCMOS technology with a cut-off frequency (f_t) and maximum frequency (f_{max}) of 200 and 270 GHz, respectively. Both chips are DC coupled to enable burst-mode operation with a record of high sensitivity.

- **SPC:** The basic structure of the SPC used at the optoelectronic buffer resembles that of the SPC used at the LP, and operates based on the same discharge-or-hold scheme. But unlike the LP's SPC that is used in a single shot fashion every several 10's of nanoseconds when a new label comes in, each conversion channel of the buffer SPC is operated repeatedly every 640 ps, that is, 16 bits at 25 Gbps, until the whole packet is converted, and thus the implementation circuits are different. More details about the buffer's SPC can be found in [29].

- **PSC:** The parallel-to-serial converter (PSC) is an optoelectronic integrated circuit (OEIC) that resembles the SPC in consisting of several conversion channels attached to a common TL. The input to each channel of the PSC is a low speed electrical signal provided by the CMOS processor, whereas the TL carries the final device output which is the high-speed electrical pulses that form a packet part used in a following step to modulate a given wavelength. Each channel comprises a customized MSM-PD discharge-based circuit operated with an optical trigger pulse to produce a positive electrical pulse and a negative electrical output from the opposite sides of the MSM-PD. Both pulses are used to generate a non-return-to-zero (NRZ)-like output electrical pulse. More details about the buffer's PSC can be found in [30].

- **Tunable transmitter**: A fast tunable transmitter is realized by monolithically integrating an EAM section composed of InGaAlAs MQWs [31] and a PRR-based tunable laser (**Figure 8b**). The InGaAlAs MQWs allow for a steep extinction curve and large E/O frequency bandwidth [32], and hence can support operation at 25-Gbit/s with a sufficient extinction ratio and over a wide wavelength range. **Figure 12a** shows the NRZ eye diagram for a 25-Gbps pseudorandom bit stream (PRBS) of 2^{31}-1. A dynamic ER larger than 10 dB was achieved for wavelengths up to 1570 nm with DC bias levels ranging from -0.8 to -1.5 V, and with a constant voltage swing of 2.0 V maintained at all wavelengths (**Figure 12b**).

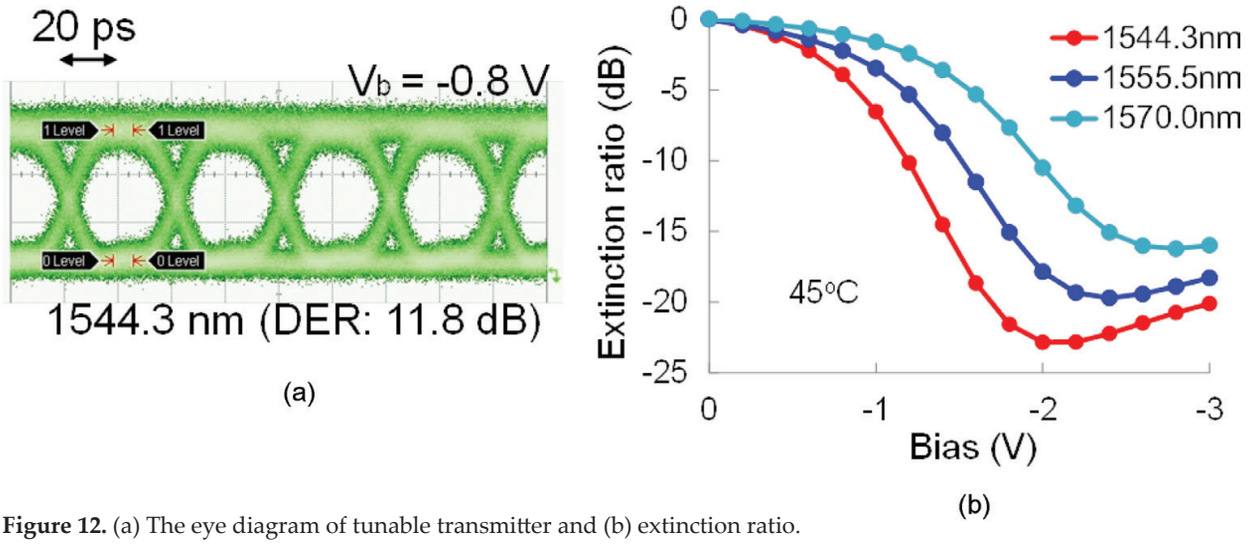

Figure 12. (a) The eye diagram of tunable transmitter and (b) extinction ratio.

7. Conclusion and outlook

In this chapter, we have reviewed our recent HOPR prototype developed to realize a new photonic intra DC network. The basic aspects of the network have been briefly introduced, followed by highlighting the enabling technologies of the new prototype. HOPR's role at the DC network has two folds as an optical-packet switch for forwarding high-speed burst-mode optical packets in a torus-topology network, and as an aggregation switch for handling the traffic of the DC servers. HOPR's operation relies on three functional units namely the label processor, optical switch, and shared buffer. HOPR is realized based on a hybrid implementation approach with an attempt to make optimal use of optics and electronic for enabling low power consumption and latency without sacrificing performance. The use of both all-optical and optoelectronic shared buffers shows the hybrid approach adopted at the subsystem level. The hybrid approach can also be seen at the device level where, for example, special optoelectronic devices are developed to efficiently interface preamble-free optical packets to electronic circuits with a novel optical clocking mechanism.

To meet the pressing demand for power reduction, HOPR is developed to have the power dissipation limited to incoming packet duration. The dissipation time has been even further reduced as, for example, in the label processor case where it occurs only during the short label duration. HOPR's latency is dominated by performing arbitration for resolving packet contention where a fast arbitration algorithm is demanded. On the other hand, to maintain a high-quality signal after multi-hop transmission, EDFAs and not SOAs are used for packet amplification, whereas other than amplification, the SOA attractive features of high-speed operation and compactness are employed for signal processing. New device packaging has been also considered as for instance in the label processor to enhance the responsivity of used photodetectors and to enable an easy lens-free and low-cost alignment.

http://dx.doi.org/10.5772/67623

OPS-based photonic DC networks provide a radical solution that can take data centers into new frontiers, and let them surpass their current difficulties caused by excessive electronic processing. Performing more node functions by combining optics/electronics is thus pivotal, and to enable that the demand for new innovative concepts and implementation strongly exists. HOPR's new prototype is a step toward fulfilling this objective which together with other community achievements reveals the high potential of this networking approach.

Acknowledgements

This work has been funded by the National Institute of Information and Communications Technology (NICT) R&D program Basic Technologies for High-Performance Hybrid Optoelectronic Router (2011–2016).

Author details

Salah Ibrahim* and Ryo Takahashi

*Address all correspondence to: ibrahim.salah@lab.ntt.co.jp

NTT Device Technology Laboratories, NTT Corporation, Kanagawa, Japan

References

[1] D.J. Richardson, J.M. Fini, and L.E. Nelson. Space division multiplexing in optical fibers. Nature Photonics 7: 353-362, 2013.

[2] Cisco visual networking index: forecast and methodology, 2013-2018. White paper c11-481360.pdf, 10 June 2014.

[3] M. AI-Fares, A. Loukissas, and A. Vahdat. A scalable, commodity data center network architecture. In Proceedings of the ACM SIGCOMM, USA, pp. 1-12, 2008.

[4] C. Kachris, K. Kanonakis, and T. Tomkos. Optical interconnection networks in data centers: recent trends and future challenges. IEEE Communications Magazine 51(9), pp. 39-45, 2013.

[5] K. Bernstein, R.K. Cavin, W. Porod, A. Seabaugh, and J. Welser. Device and architecture outlook for beyond CMOS switches. Proceedings of the IEEE 98(12), pp. 2169-2184, 2010.

[6] R. Takahashi, T. Segawa, S. Ibrahim, T. Nakahara, H. Ishikawa, A. Hiramatsu, Y-C. Huang, and K-I. Kitayama. Torus data center network with smart flow control enabled by hybrid optoelectronic routers. Journal of Optical Communications and Networking 7: B141–B152, 2015.

[7] Y-C. Huang, Y. Yoshida, S. Ibrahim, R. Takahashi, A. Hiramatsu, and K-I. Kitayama. Novel virtual OCS in OPS data center networks. Photonic Network Communications 31(3): 448-456, 2016.

[8] Y-C. Huang, Y. Yoshida, S. Ibrahim, R. Takahashi, A. Hiramatsu, and K-I. Kitayama. Bypassing route strategy for optical circuits in OPS-based data center networks. IEEE Photonics Journal 8(2): 1-10, 2016.

[9] W. Mellette, G. Schuster, G. Porter, G. Papen, and J. Ford, "A scalable, partially configurable optical switch for data center networks," J. Lightwave Tech., 35(2), pp. 136-144, Jan.15, 2017.

[10] S. J. Ben Yoo. Optical packet and burst switching technologies for the future photonic internet. Journal of Lightwave Technology 24, pp. 4468-4492, 2006.

[11] G. I. Papadimitriou, C. Papazoglou, and A. S. Pomportsis. Optical switching. Hoboken, NJ: Wiley-Interscience, 2006.

[12] M.-J. Kwack, T. Tanemura, A. Higo, and Y. Nakano. Monolithic InP strictly non-blocking 8×8 switch for high-speed WDM optical interconnection. Paper presented at the European Conference and Exhibition on Optical Communication, Amsterdam, The Netherlands, 2012, Paper Th.3.B.3.

[13] R. S. Tucker, P.-C. Ku, and C. J. Chang-Hasnian. "Slow-light optical buffers: capabilities and fundamental limitations," J. Lightwave Tech., 23(12), 2005.

[14] K. Nosaki, A. Shinya, S. Matsuo, Y. Suzaki, T. Segawa, T. Sato, Y. Kawaguchi, R. Takahashi, and M. Notomi. Ultralow power all-optical RAM based on nanocavities. Nature Photonics 6: 248-252, 2012.

[15] T. Zhang, K. Lu, and J. R. Jue. Shared fiber delay line buffers in asynchronous optical packet switches. IEEE Journal on Selected Areas in Communications 24(4): 118-127, 2006.

[16] S. Porto, C. Antony, A. Jain, D. Kelly, D. Carey, G. Talli, P. Ossieur, and P. D. Townsend, "Demonstration of 10 Gbit/s burst-mode transmission using a linear burst-mode receiver and burst-mode electronic equalization," J. Opt. Comm. and Net. 7(1), A118–A125 (2015).

[17] Roland E. Best. Phase locked loop: design, simulation and applications, 4th ed. New York: McGraw-Hill, 1999.

[18] R. Yu, R. Proietti, S. Yin, J. Kurumida, and S. J. Ben Yoo. "IO-Gb/s BM-CDR circuit with synchronous data output for optical networks," IEEE Photon. Technol. Lett., 25, pp. 508-511, 2013.

[19] N. Suzuki, K. Nakura, T. Suehiro, M. Nogami, S. Kosaki, and J. Nakagawa. Over-sampling based burst-mode CDR technology for high-speed TDM-PON systems. In Optical Fiber Communication Conference and Exposition (OFCINFOEC), 2011 and the National Fiber Optic Engineers Conference, Los Angeles, CA, 2011, pp. 1-3.

[20] A. Rylyakov, J. Proesel, S. Rylov, B. G. Lee, J. Bulzacchelli, A. Ardey, C. Schow, and M. Meghelli. A 25 Gb/s burst-mode receiver for low latency photonic switch network. IEEE Journal of Solid-State Circuits 50(12), pp. 3120-3132, 2015.

[21] S. Ibrahim, T. Nakahara, H. Ishikawa, and R. Takahashi. Burst-mode optical label processor with ultralow power consumption. Optics Express 24: 6985-6995, 2016.

[22] K. Takahata, R. Takahashi, T. Nakahara, H. Takenouchi, and H. Suzuki. 3.3 ps electrical pulse generation from discharge-based metal-semiconductor-metal photodetector. Electronics Letters 41(1): 38-40, 2005.

[23] H. Kim Jae, H. T. Griem, R. A. Friedman, E. Y. Chan, and S. Ray. High-performance back-illuminated InGaAs/InAlAs MSM photodetector with a record responsivity of 0.96 A/W. IEEE Photonics Technology Letters 4(11): 1241-1244, 1992.

[24] T. Segawa, S. Matsuo, T. Kakitsuka, T. Sato, Y. Kondo, and R. Takahashi. Semiconductor double-ring-resonator-coupled tunable laser for wavelength routing. IEEE Journal of Quantum Electronics 45(7): 892-899, 2009.

[25] Y. Muranaka, T. Segawa, Y. Ogiso, T. Fujii, and R. Takahashi. Performance improvement of an EAM-based broadcast-and-select optical switch module. In The Proceedings of the International Conference on Photonics in Switching (PS), Florence, 2015.

[26] T. Nakahara, and R. Takahashi. Self-stabilizing optical clock pulse-train generator using SOA and saturable absorber for asynchronous optical packet processing. Optics Express 21, pp. 10712-10719, 2013.

[27] M. Nada, M. Nakamura, and H. Matsuzaki. 25-Gbit/s burst-mode optical receiver using high-speed avalanche photodiode for 100-Gbit/s optical packet switching. Optics Express 22, pp. 443-449, 2014.

[28] M. Nada, Y. Muramoto, H. Yokoyama, N. Shigekawa, T. Ishibashi, and S. Kodama. Inverted InAlAs/InGaAs avalanche photodiode with low–high–low electric field profile. Japanese Journal of Applied Physics 51, p. 02BG03-1, 2012.

[29] S. Ibrahim, H. Ishikawa, T. Nakahara, and R. Takahashi. A novel optoelectronic serial-to-parallel converter for 25-Gbps burst-mode optical packets. Optics Express 22, pp. 157-165, 2014.

[30] H. Ishikawa, T. Nakahara, H. Sugiyama, and R. Takahashi. A parallel-to-serial converter based on a differentially-operated optically clocked transistor array. IEICE Electronics Express 10(20); pp. 20130709, 2013.

[31] T. Segawa, W. Kobayashi, T. Sato, S. Matsuo, R. Iga, and R. Takahashi. A flat-output widely tunable laser based on parallel-ring resonator integrated with electroabsorption modulator. Optics Express 20: B485-B492, 2012.

[32] W. Kobayashi, M. Arai, N. Fujiwara, T. Fujisawa, T. Tadokoro, K. Tsuzuki, Y. Kondo, and F. Kano. Design and fabrication of 10-/40-Gb/s, uncooled electroabsorption modulator integrated DFB laser with butt-joint structure. Journal of Lightwave Technology 28(1): 164-171, 2010.

IntechOpen