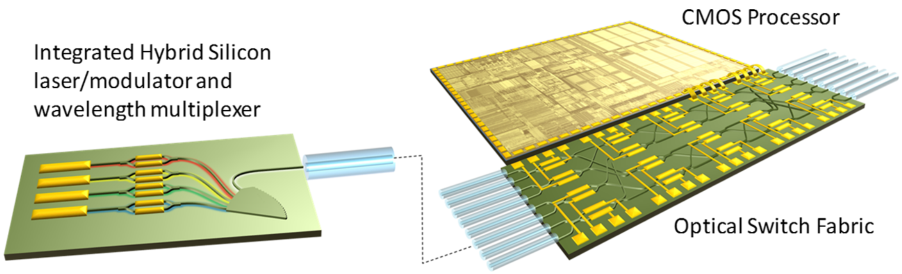
IEEE Photonics Journal
An IEEE Photonics Society Publication

Invited Paper

Photonic Switching for Data Center Applications

Volume 3, Number 5, October 2011

L. Chen
E. Hall, Member, IEEE
L. Theogarajan, Member, IEEE
J. Bowers, Fellow, IEEE


Integrated Hybrid Silicon laser/modulator and wavelength multiplexer
CMOS Processor
Optical Switch Fabric

DOI: 10.1109/JPHOT.2011.2166994
1943-0655/$26.00 ©2011 IEEE

Photonic Switching for Data Center Applications

L. Chen,[1] E. Hall,[2] *Member, IEEE*, L. Theogarajan,[1] *Member, IEEE*, and J. Bowers,[1] *Fellow, IEEE*

(Invited Paper)

[1]Department of Electrical and Computer Engineering, University of California at Santa Barbara,
Santa Barbara, CA 93106 USA
[2]Aurrion, Inc., Santa Barbara, CA 93117 USA

DOI: 10.1109/JPHOT.2011.2166994
1943-0655/$26.00 ©2011 IEEE

Manuscript received July 26, 2011; revised August 18, 2011; accepted August 24, 2011. Date of publication September 1, 2011; date of current version September 27, 2011. This work was supported in part by Intelligence Advanced Research Project Activity. Corresponding author: L. Chen (e-mail: lchen@ece.ucsb.edu).

Abstract: Switching fabrics in data centers that rely on traditional electrical switches face scaling issues in terms of power consumption. Fast optical switches based on a silicon photonics platform can enable the high port speed and high interconnection density needed while still maintaining a small footprint and low power consumption.

Index Terms: Data center switching, data warehouses, optoelectronic devices, silicon photonics.

1. Introduction

This paper proposes to radically increase the capacity of switch fabrics used to interconnect processor chips and memory units within a data center and to achieve this with dramatically lower power and cost requirements. The approach is scalable to the much larger capacities and much higher bisection bandwidths that will be required in the future.

The interconnect fabric is taking an ever-more dominant portion of the power budget and optical interconnects are already used today within data centers for rack to rack interconnection to overcome the limits of electrical signaling. However, the switches interconnecting these racks have inadequate capacity to continue to scale with complementary-metal-oxide semiconductor (CMOS)-based technology due to fundamental limitation of on-chip interconnects [1] and power consumption and pin count of scheduler application-specific integrated circuits [2]. Further, not only does the switch fabric require significant power, but the conversion from the optical signal to electrical and back again (OEO) adds a significant amount of power and space. We present here a unique switching technology that allows for high radix and high bandwidth, which is not achievable in conventional electrical interconnects, while dissipating very low power. We achieve these diametrically opposing goals by utilizing the large-bandwidth enabled by optics fabricated in a cost-efficient CMOS platform. Furthermore, we integrate conventional CMOS electronics within the optical platform using a chips-last integration process [3]. The grand vision of this paper is shown in Fig. 1.

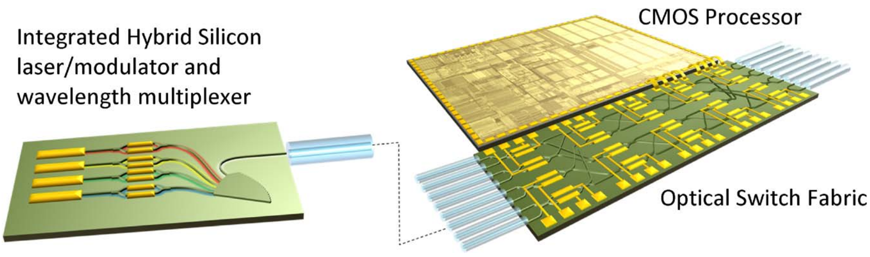

Integrated Hybrid Silicon laser/modulator and wavelength multiplexer
CMOS Processor
Optical Switch Fabric

Fig. 1. Grand vision. In the chip on the left, electrical data are modulated onto an optical link and are ported to the switching network on the right via optical fibers. The optical switching network is based on an array of hybrid-silicon Mach–Zehnder Interferometers. A processor that is integrated with the switch configures the switch utilizing packet switching protocols. Figure courtesy of M. Heck.

2. Electrical Switching Issues

The interconnection architecture is important in determining the cost-performance tradeoffs in data centers. All of this bandwidth is optical and must next pass through a switch chip and routed to another processing element in the system. Current state-of-the-art electrical switching chips for this function can be represented by the Vitesse VSC3144 Crosspoint Switch. This chip has 144 ports at 10.709 Gb/s, corresponding to a total switching capacity of 1.54 Tb/s. To achieve this switching throughput, the chip consumes 21 W of power or 13.6 pJ/bit. Additionally, the chip has 1072 electrical pins [4]. It is in both power consumption and pin count that future switching chips will face serious challenges as switching capacity scales to 10 Tb/s and to 100 Tb/s. Even at an aggressive target of 5 pJ/bit for electrical, a 100 Tb/s switch will require 500 W of power. Optical-to-electrical conversion plus deserialization that must occur at the chip edge must then also be added. The Vitesse's chip's inputs are electrical signals; therefore, OEO conversion would need to be added to the power calculation, which would likely double the power consumption to about 40 W if OEO are included, or about 25 pJ/bit, based on OEO power consumption estimates presented by Szymanski and Gourgy [5]. Similarly, pin counts will also be an issue. The number of balls just for signal input/output in the ball grid array of future switch chips at four different line rates per port and different port counts grows beyond practical limits in just a few years. For instance, to enable terabits bandwidth, the number of connections required exceeds 10 000 [6]. At very low port count, pin requirements are small, but at current switching bandwidth the pin count is already greater than 1000. This rapidly increases to greater than 10 000 at the bandwidths needed by future data centers. The International Technology Roadmap for Semiconductors (ITRS) has no roadmap for fabricating chips with pin counts greater than 6500 [7]. While building larger fabrics using smaller chips can overcome the pin count limitation, this would require high speed signals to be driven off-chip, increasing IO parasitics and power consumption. Another key figure-of-merit in interconnect topologies is the radix (number of input/output ports of a switch) of the network. It defines the minimum number of hops needed to route a packet of information from point-to-point. For a router with a radix k connecting N ports, the minimum number of hops is $2\log_k N$. The increased performance in newer technologies enables a high-radix network to have lower-latency and, hence, lower cost. Furthermore, a low-hop count enabled by a high-radix interconnects results in lower power dissipation of the network due to the lower number of nodes traversed. Current state-of-the-art electrical routers, like the Mellanox Infiniswitch IV, have a radix of 36 @ 40 Gb/s (i.e., 36 ports @ 40 Gb/s/port) [8]. Consider a network with 32 000 servers for instance, if the Mellanox switch is utilized to implement this network, it would yield a minimum of six hops. Given the electrical router delays of many clock cycles, this leads to large latencies which may be unacceptable.

3. Optical Switches

Optical switches offer several significant advantages over their electrical counterparts, which should allow systems to overcome these roadblocks:

TABLE 1

2 × 2 MZI switch device characteristics

Port	ER (dB)	Crosstalk (dB)	Rise Time (ps)
1 -> 3	25	-19	27
1 ->4	19	-24	35
2 -> 3	18	-29	30
2 -> 4	26	-15*	29

* Due to process error

- *Lower Power Consumption:* The switching element itself can be capacitive and, therefore, can have much lower power consumption ($<$ 1 μW/Gb/s compared with $\sim$5 mW/Gb/s), as will be discussed in more detail below.
- *Extremely High Port Bandwidth:* The input ports are data rate and modulation independent and even be compatible with Wavelength-Division Multiplexing (WDM) architectures, allowing the switch fabric to scale to extremely high data rates ($>$ 100 Gb/s per wavelength), as well as extremely high overall bandwidth per port (4 Tb/s when using 40 WDM wavelengths at 100 Gb/s or Polarization Multiplexed Quadrature Phase Shift Keying (QPSK), for example).
- *No OEO Conversion:* Since all of the data are kept optical, there is no OEO conversion required, eliminating the large power consumption and large circuit board area associated with these components as discussed above.

Although there are several approaches to optical switching which provide these basic advantages, not all are appropriate for packet switching application. Microelectromechanical system (MEMS) optical switches, for example, can be scaled for high throughput and high radix at low power consumption but at the expense of speed. Switching times are $>$ 1 ms, which is too slow for burst switching. The most appropriate choice would be an integrated semiconductor switch, which enables both high speed and low power consumption. Multiple demonstrations of such switching fabrics on InP, in fact, have been shown [9]. Unfortunately this approach does not scale well to larger port counts since the optical loss in the passive InP waveguides is very high and the chips themselves must be fabricated on limited wafer sizes. In order to avoid both of these issues, switch fabrics have been demonstrated on a silicon platform which takes advantage of the low loss in silicon waveguides, as well as the ability to fabricate these chips on large substrates within a foundry. The poor electrooptic effect in silicon, however, necessitates a tradeoff between narrower optical bandwidth, increased power consumption, or large device size. Power consumption is another key parameter when considering optical switching. High speed switching elements such as tunable wavelength converter (TWC) arrayed-waveguide grating (AWG) and semiconductor optical amplifier (SOA) gate arrays consume at least 100 mW of power [10], [11], making them marginally better than scaled CMOS counterpart at best [10]. Although, as noted by Yoo [12], today's CMOS electronics are limited by static power dissipation rather than dynamic power, and therefore, scaling an electronic switch to higher capacity by using a multichip approach will be limited by their idle rather than their active power.

4. Hybrid Silicon Switch

In order to still leverage the silicon platform while maintaining low power consumption in a small footprint, we propose using the hybrid integration of III–V (InP) material to a silicon platform [13]. This technology uses the III–V material to enhance the electrooptic effect while keeping the majority of the switch in the low-loss silicon waveguides. Since this technology is CMOS-compatible, it scales easily in terms of both wafer size and integration and can be fabricated in a standard foundry. Tables 1 and 2 summarize the measured results of a Mach–Zehnder Interferometer (MZI) switch built on the hybrid silicon platform [14] and compared with other silicon-based modulators [11]. Since the hybrid silicon switch uses a capacitive approach, the power consumption for each element is extremely small ($\ll$ 1 mW), allowing very large fabrics to be built with minimal power

TABLE 2

Comparison of silicon modulators

	Ring	Electroabsorption		Mach-Zehnder		
Platform	Si	Hybrid Si	SiGe	Hybrid Si	Si	Si
Reference	[6]	[7]	[8]	[9]	[10]	[11]
Vpp (V)	1.8	2	7	4	6	1
Length (μm)	10	100	50	500	1000	100
Vπ (Vmm)	—	—	—	2	40	0.06
ER (dB)	17	11	10	20	≥ 20	18
Optical BW (nm)	≤ 1	30	15	100	75	110

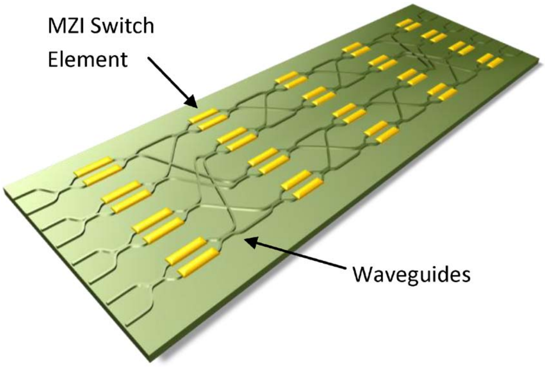


Fig. 2. Representative chip architecture of an 8 × 8 Benes switch.

consumption. In order to compensate for the optical loss of each element, however, optical amplifiers must be included using the same hybrid integration process. The addition of these amplifiers allows the overall optical insertion loss of the chip to be zero. Additionally, the amplifiers at the output of the chip can be coupled with simple on-chip photodetectors to both level the optical power across channels (eliminating the need for high dynamic range receivers in the system) and to self test the chip. A drawing of an 8 × 8 switch based on a Benes architecture and using the hybrid silicon switch is shown in Fig. 2. Using some basic assumptions about the individual components, the power consumption of the entire switch fabric for several switch sizes is shown in a later section.

5. Crosstalk

Waveguide crossings and MZI switches both introduce undesirable crosstalk that limits the signal to noise ratio, and the port count of the optical switch. Using a wavefront matching method and crossing at an angle of about 20°, waveguide crossing can be designed to have crosstalk of -38 dB and excess loss of 0.1 dB [15]. MZI switches, on the other hand, suffer more crosstalk. The switch presented in [16] has crosstalk of -19 dB, with process improvement, -20 dB is readily achievable. A Quantum-well electrorefraction-based MZI switch presented in [17] has already achieved -25 dB of crosstalk. We believe that with further research in process refinement and optimization, a projected crosstalk of -30 dB is achievable. Fig. 3 shows the Signal to Crosstalk ratio (SXR) performance as function of switch size. We assume a Benes network, and consider first order crosstalk only with 0 dB loss. The total amount of crosstalk is proportional to path length—the number of switches the input has to traverse to reach the output port. For Benes network, the path length $P = 2\log_2(N) - 1$, where N is the number of ports. We define $\mathrm{SXR(dB)} = 10\log(L/PC)$, where C and L are the switch element crosstalk and loss, respectively. With existing crosstalk level of about -20 dB, the switch size is limited to 16 × 16, resulting in a BER of about 10^{-8}. The largest packet

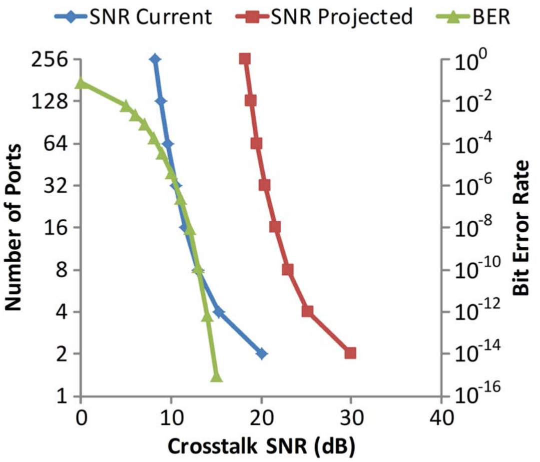


Fig. 3. SXR of current MZI switch (−20 dB isolation) and projected SXR performance (−30-dB isolation). BER is plotted as function of SXR assuming On/Off Keying data.

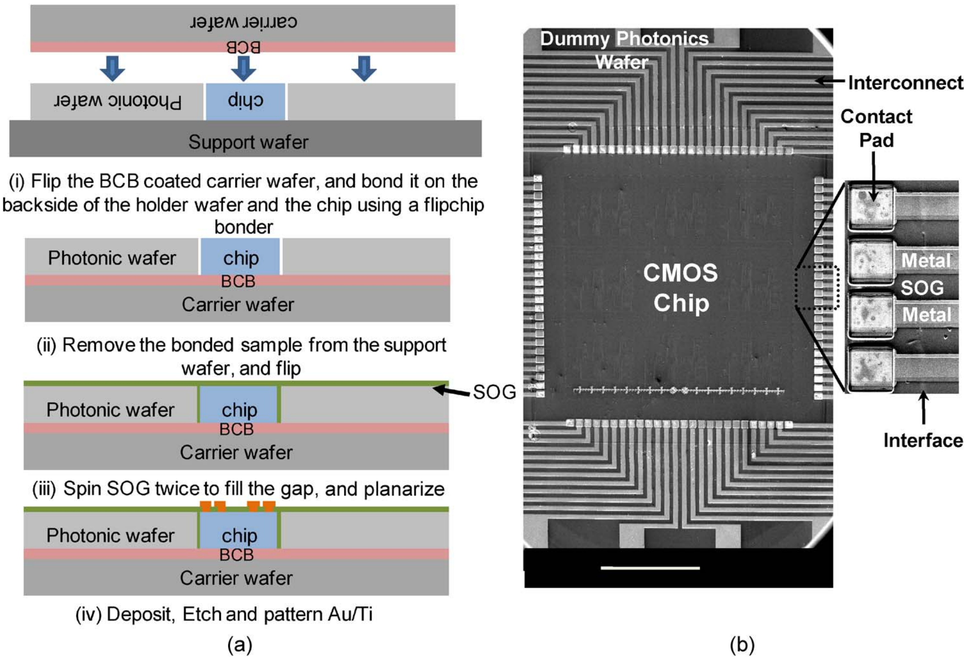
size is about 1500 bytes in a data center [18]. BER of 10^{-8} translates into a packet loss rate of about 0.01%. As SXR improves, we expect to be able to implement large port count switch fabric with little penalty due to crosstalk. The low loss property of the hybrid silicon platform enables the minimal use of SOAs, reducing the amount of amplified spontaneous emission (ASE) accumulated in the path. For instance, for 256 ports Benes network, 15 stages are needed, and the loss per stage is about 1 dB. If we compensate every six stages, which equates to three stages of SOA, and given that the typical noise produced by SOA is about −30 dBm, the output referred noise due to ASE is about −25 dBm. In contrast, noise due to crosstalk (assuming 30 dB isolation) is −18 dBm; therefore, the performance is dominated by the crosstalk.

6. Wafer Scale Integration

One of the key advantages of the proposed approach is our ability to intimately integrate the photonic and electronic integrated circuits on a single substrate, enabling lower driver power and higher bandwidth. We achieve this by modifying a process flow that we have developed for wafer scale integration of mixed foundry die [3]. The outline of our process is shown in Fig. 4. We start with the fully processed photonics wafer containing the switch fabric. An area is preallotted on the photonics wafer, for the electronic die. In this preallotted area, a cavity is etched into the photonics wafer where the electronic CMOS die is placed. The wafer and the chip are then flipped to ensure that alignment is nominally flat at the chip/wafer interface. Using a carrier wafer for support the photonics wafer and the chip are bonded onto the carrier wafer using benzocyclobutene. After completion of this process, we pattern and etch connectors from the chip to the photonic network. The entire process can also start with a blank wafer and the photonic network can be fabricated post-CMOS integration. The choice between a chips-last or chips-first approach depends on which process results in a higher yield and is the topic of current research in our group.

7. CMOS Electronics

To realize the full potential of a fast optical switch, low-power electronics can be integrated in a single CMOS chip to decode header packets, configure the switch fabric for transmission, and perform calibration to compensate for process, temperature and aging variations in the photonic IC. Fig. 5 shows a block diagram of our proposed network processor IC. For an *N*-port switch, the processor consists of *N* channel-transceivers, switch configuration logic, switch drivers, optical amplifier drivers, an analog-to-digital converter, and low-frequency transimpedance amplifiers (TIAs) to determine optical channel losses. The front end is responsible for performing optical-to-electrical conversion of the header packets. The header packets can be encoded in relatively low



(i) Flip the BCB coated carrier wafer, and bond it on the backside of the holder wafer and the chip using a flipchip bonder

(ii) Remove the bonded sample from the support wafer, and flip

(iii) Spin SOG twice to fill the gap, and planarize

(iv) Deposit, Etch and pattern Au/Ti

(a)

(b)

Fig. 4. (a) Process flow for the integration of the CMOS chip intimately with the photonics substrate. (b) SEM of initial CMOS integration into a dummy photonics wafer.

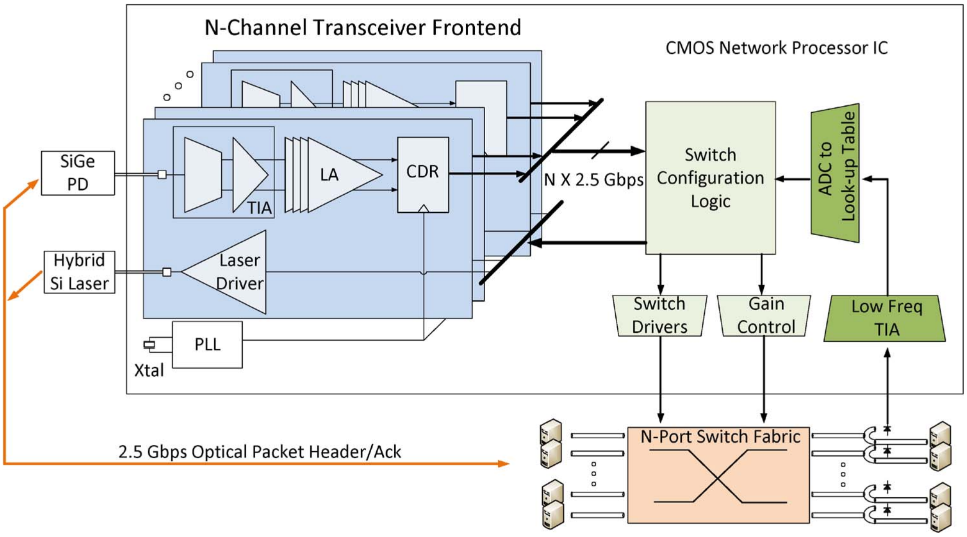


Fig. 5. Envisioned CMOS network processor IC integrating photonic header packet transceiver, switch configuration logic, switch drivers, and gain control electronics. The IC is capable of bidirectional communication with the host so that Acknowledge commands can be sent to the host once the switch fabric is ready for transmission.

data rate compared with the main payload [19]; hence, it is amenable to CMOS technology nodes that are relatively mature which reduces cost. Furthermore, the full integration of transceiver front end realizes substantial power savings, since it avoids driving high speed signals off-chip that would require 50-Ω impedance terminations. The switch configuration logic implements control algorithm such as the one implemented in [20] to enable packet routing and contention resolution. In addition, it determines the amount of gain compensation required to achieve minimal loss between input and

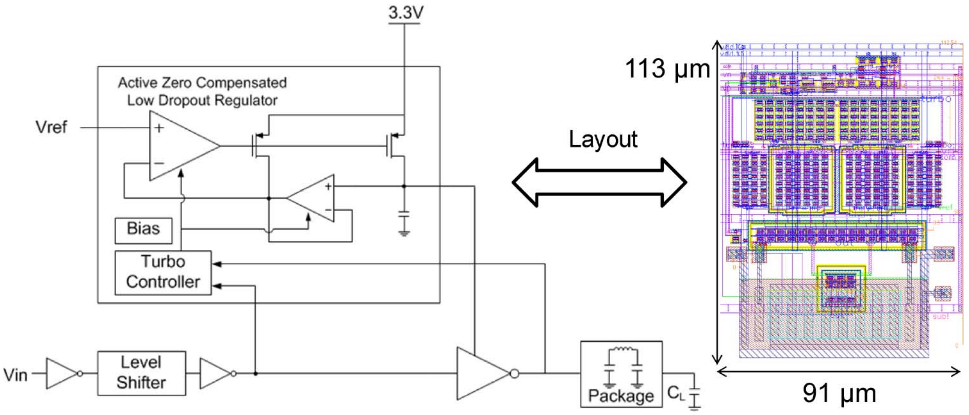


Fig. 6. Low-power, compact area capacitive switch driver.

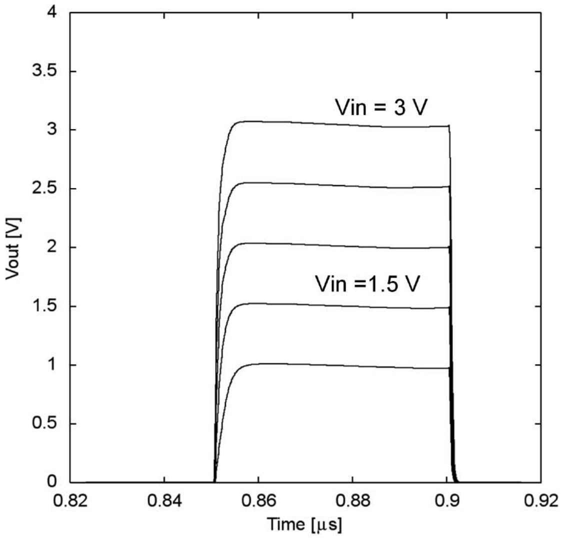
output ports. The channel loss is measured using a TIA operating at relatively low frequencies. Since the TIA senses unipolar optical input from the photodiode, the input signal's average is proportional to its amplitude, and therefore, the TIA can extract the DC component to determine the loss incurred in the switch fabric. This DC signal is then digitized using an ultralow-power Analog to Digital Converter that consumes 4.8 μW [21]. Once the amount of gain needed is determined, the gain control drivers adjust the integrated optical amplifiers to mitigate the channel losses. Finally, an array of switch drivers configures the switch fabric for packet transmission.

Fig. 6 shows the schematic diagram of the switch driver and layout implemented in IBM 0.13 μm CMOS process. The hybrid MZI switch presents a capacitive load of approximately 1 pF. The driver is designed to have a rise time of less than 5 ns. Multiple switches in an array are susceptible to interference from neighboring activity. To mitigate interference, the driver's power supply should be regulated using a low-dropout regulator (LDO). Furthermore, the LDO allows the output voltage to be set to a reference voltage, which is determined in the calibration stage. The calibration helps mitigate process variation in the fabrication of the optical switches by optimizing individual drive voltages to get maximum extinction ratio. To save power, the LDO incorporates a turbo mode in which the switch is given a power boost for approximately 10 ns during the rising edge of the input. This effectively increase the bandwidth of the amplifier, which in turn makes it respond much faster to the input step, thus providing a faster output rise time. This switching of bandwidth presents a challenge in the design of the LDO compensation network. Typically, a two-stage amplifier such as the one present in the LDO is compensated by lead-lag compensator, which introduces a zero in the transfer function near the unity gain frequency of the amplifier. Turbo mode moves the unity gain frequency of the amplifier, thus the zero location also needs to be moved accordingly. Otherwise, ringing could occur due to insufficient phase margin. To overcome this problem, we employ active zero compensation [22] to make the zero location proportional to the bias current, thus making sure that the regulator is stable under all switching conditions. Another added benefit of the AZC scheme is that the large compensation capacitor normally used for the lead-lag compensator can now be moved to the output of the regulator, reducing the high-frequency dropout of the regulator output. The entire switch driver, including the LDO and internal capacitor, is laid out in a compact space of 113 μm $\times$ 91 μm.

The simulated output of the switch driver is shown in Fig. 7. Vin is swept from 1 V to 3 V in 0.5-V increments. Maximum rise time is 3.4 ns on load capacitance of 1 pF. The dynamic power consumption is about 500 μW at 3 V output when driven with a 10-MHz waveform (100-ns burst).

8. Power Consumption

Since the power scales with the number of ports it is critical to minimize the power consumption per port. One of the key elements that conventionally dominates the power consumption is the switch itself, since optical switches normally require carrier injection for port switching. This is further



(a)

Output [V]	$t_{10\text{-}90}$ [ns]	Power [µW]
1	3.4	391
1.5	2.4	412
2	2.1	442
2.5	2.0	475
3	1.8	515

(b)

Fig. 7. Simulated result of the switch driver. (a) Input versus output step response. (b) Rise time and dynamic power consumption assuming 10-MHz switch rate and 3-V output.

TABLE 3

Power consumption of photonic switch in Benes architecture

Ports	No. of Stages	No. of SE	No. of SOA	Throughput (Tbps)		Total Power (W)	Energy/bit (pJ/bit)	
				4 x 50 Gbps	8 x 50 Gbps		4 x 50 Gbps	8 x 50 Gbps
16	7	56	16	3.2	6.4	1.61	0.50	0.25
32	9	144	64	6.4	12.8	6.41	1.00	0.50
64	11	352	128	12.8	25.6	12.84	1.00	0.50
128	13	832	128	25.6	51.2	12.88	0.50	0.25
256	15	1920	512	51.2	102.4	51.39	1.00	0.50

exacerbated by the fact that the drivers need to be terminated to prevent unwanted ringing due to impedance mismatch. In our architecture this is alleviated by the fact our switches are capacitive and use a field effect and, hence, do not require constant carrier injection, and the second is that the intimate integration of the electronics and the switch on the same substrate effectively makes the wire an *RC* rather than a transmission line. Table 3 summarizes the power efficiency of our proposed photonic switch. Commercially available electronic switching chips such as the Vitesse's VSC3144 consumes 21 W for a throughput of 1.54 Tb/s, resulting in 13.6 pJ/bit of energy efficiency. Assuming that the photonic switch is configured as a Benes network, for $N = 2^i$ ports, it requires $(2i - 1)2^{i-1}$ switches, spanning $2i - 1$ stages. Assume each switch element consumes 100 μW, and loss per stage is 1 dB, requiring 1 SOA per 6 dB of loss, or every six stages, and each SOA consumes 100 mW. Two data rate scenarios are presented. First one assumes a wavelength multiplexed data using four wavelengths at 50 Gb/s each, corresponding to an aggregate data rate of 200 Gb/s. The second scenario assumes eight wavelengths at 50 Gb/s each, or 400 Gb/s. This kind of data rate is feasible today with recently demonstrated technology such as 50 Gb/s hybrid silicon modulator [23] and integrated photonic link [24]. Based on the assumptions listed above, our proposed switch, compared with the state-of-the-art commercial part, can achieve an order-of-magnitude improvement in energy efficiency (see Table 3). However, as can be seen from Fig. 8 the power consumption of the amplifier can be considerably reduced by using modified SOA

TABLE 4

System power consumption

Ports	2	4	8	32
TIA/LA/CDR (mW)	10	20	40	160
Laser Driver (mW)	100	200	400	1600
Control Logic (mW)	10	20	40	160
Switch Driver (mW)	0.2	0.4	0.8	3.2
Optical Amplifier (mW)	80	160	320	1280
Calibration (mW)	5	10	20	80
Total (W)	0.21	0.41	0.82	3.28

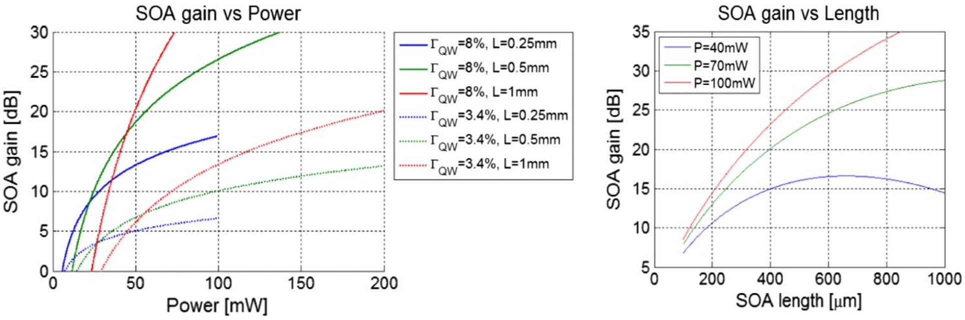


Fig. 8. SOA power consumption calculations.

designs. Table 4 lists the power consumption per component in our design with different port configurations. With the 4× improvement in SOA power, we can reduce the power even further.

9. Control Considerations

Our control protocol was designed to maintain high data throughput while allowing for short packet transfers. The main consideration was to hide the latency of the switch configuration. We assume that the delay of the packet processor's optoelectronic front end is about 1 ns, and we use 8 bits to encode the destination address. We further assume that the processor will take 8 cycles to load the destination bits, and about 50 cycles (logic and RAM access) to process the request. The switch will take about 5 ns to be fully configured and 2 ns to send acknowledge back to host. Assuming processor is running at 2.5 GHz, the processing time is 24 ns. The total switch overhead is 26 ns under contention-free conditions. With contention, round robin arbitration logic can be designed to resolve contention in about 2 ns [25]. Conservatively we assumed a rate of 50 ns as the minimum time between switch configurations, which allows for ~250 byte transfers for a 40-Gb/s channel. To ensure high degree of network utilization we propose to use an optical burst switching (OBS) algorithm [26]. In the OBS system, a burst header cell is first sent to the switch fabric on a separate wavelength. The header cell can also contain optional information about the size of packet transmission to allow for variable packet length transfers and allows the switch fabric to relinquish the route once the predetermined packet has been routed through the switch. The switch is configured by decoding the header cell using a custom clock and data recovery unit with eye tracking and passing the data to the control logic. The control logic computes the route and sends an acknowledge on the same optical wavelength back to the sender. When the sender receives the acknowledge transmission of data begins until an end of packet is detected. At this point, the route is retired and can be used by another sender. Fig. 9 shows a pictorial representation of the setup and tear down protocol. Alternatively a Tell-and-go [26] architecture can be designed where the sender waits a predetermined amount of time required for the switch to be configured and starts

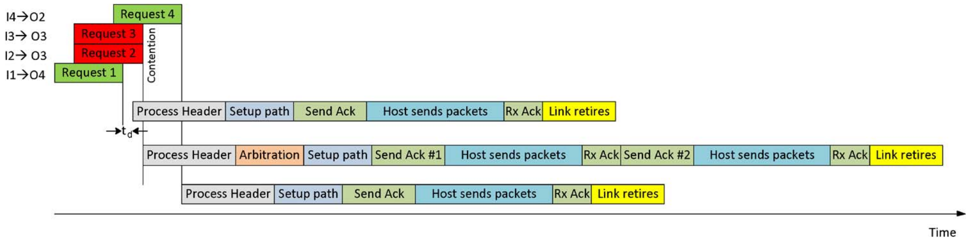


Fig. 9. Link setup/tear down protocol.

transmission. An optical header can also be inserted prior to egress of the packet if a multihop network protocol is desired.

10. Conclusion

The fundamental capabilities and energy efficiency of large-scale data centers rest heavily on their communication infrastructure. Electrical switching, however, will face serious obstacles in terms of power consumption and even pin count in its effort to scale and meet future interconnection demand. This paper proposed a design for the next generation of data center architectures enabled by a unique optical switching fabric which has very low overall power consumption. This design can be fully fabricated in CMOS compatible silicon technology, which allows the cointegration of CMOS electronics with the optical integrated circuit and can mitigate the inevitable process variations.

Acknowledgment

The authors thank E. Kintzer for useful suggestions.

References

[1] R. G. Beausoleil, P. J. Kuekes, G. S. Snider, S. Y. Wang, and R. S. Williams, “Nanoelectronic and nanophotonic interconnect,” *Proc. IEEE*, vol. 96, no. 2, pp. 230–247, Feb. 2008.
[2] I. Keslassy, S. T. Chuang, K. Yu, D. Miller, M. Horowitz, O. Solgaard, and N. McKeown, “Scaling Internet routers using optics,” in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun., ACM*, 2003, pp. 189–200.
[3] A. Uddin, K. Milaninia, C.-H. Chen, and L. Theogarajan, “Wafer scale integration of CMOS chips for biomedical applications via self-aligned masking,” *IEEE Trans. Components, Packaging, Manuf. Technol.*, vol. 2, 2011, DOI:10.1109/TCPMT.2011.2166395, (in press).
[4] Vitesse, *10.709 Gbps 144 × 144 Crosspoint Switch Datasheet*, 2011. [Online]. Available: http://www.vitesse.com/products/product.php?number=VSC3144
[5] T. H. Szymanski and A. Gourgy, “Power complexity of multiplexer-based optoelectronic crossbar switches,” *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 13, no. 5, pp. 604–617, May 2005.
[6] L. A. Polka, “Package technology to address the memory bandwidth challenge for terascale computing,” *Intel Technol. J.*, vol. 11, no. 3, pp. 197–204, Aug. 2007.
[7] ITRS, *2010 Overall Roadmap Technology Characteristics Tables*. [Online]. Available: http://www.itrs.net/Links/2010ITRS/Home2010.htm
[8] Mellanox Technologies, *Semiconductor Product Guide*. [Online]. Available: www.mellanox.com/pdf/products/AdapterSilicon_PSG.pdf
[9] G. A. Fish, B. Mason, L. A. Coldren, and S. P DenBaars, “Optical crossbar switches on InP,” in *Proc. 12th IEEE Annu. Meeting LEOS*, 1999, vol. 2, pp. 405–406.
[10] R. Tucker, “Optical packet switching: A reality check,” *Opt. Switching Netw.*, vol. 5, pp. 2–9, Mar. 2008.
[11] H.-W. Chen, “High-speed hybrid silicon Mach–Zehnder modulator and tunable microwave filter,” Ph.D. dissertation, Univ. California, Santa Barbara, CA, Mar. 2011.
[12] S. J. B. Yoo, “Energy efficiency in the future Internet: The role of optical packet switching and optical-label switching,” *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 2, pp. 406–418, Mar./Apr. 2011.
[13] H. W. Chen, Y. H. Kuo, and J. E. Bowers, “25Gb/s hybrid silicon switch using a capacitively loaded traveling wave electrode,” *Opt. Exp.*, vol. 18, no. 2, pp. 1070–1075, Jan. 2010.
[14] H.-W. Chen, “Low-power, fast hybrid silicon switches for high-capacity optical networks—OSA technical digest,” in *Proc. Photon. Switching, Opt. Soc. Amer.*, 2010, p. PMC3.
[15] Y. Sakamaki, T. Saida, M. Tamura, T. Hashimoto, and H. Takahashi, “Low loss and low crosstalk waveguide crossings designed by wavefront matching method,” *IEEE Photon. Technol. Lett.*, vol. 18, no. 19, pp. 2005–2007, Oct. 2006.
[16] H.-W. Chen, J. D. Peters, and J. E. Bowers, “Forty Gb/s hybrid silicon Mach–Zehnder modulator with low chirp,” *Opt. Exp.*, vol. 19, no. 2, pp. 1455–1460, Jan. 2011.

[17] B. H. P. Dorren, J. E. M. Haverkort, R. Prasanth, F. H. Groen, and J. H. Wolter, “Low-crosstalk penalty MZI space switch with a 0.64-mm phase shifter using quantum-well electrorefraction,” *IEEE Photon. Technol. Lett.*, vol. 13, no. 1, pp. 37–39, Jan. 2001.

[18] T. Benson, A. Anand, A. Akella, and M. Zhang, “Understanding data center traffic characteristics,” *SIGCOMM Comput. Commun. Rev., ACM*, vol. 40, no. 1, pp. 92–99, Jan. 2010.

[19] T. Lin, W. Tang, K. A. Williams, G. F. Roberts, L. B. James, R. V. Penty, M. Glick, and Mcauley, “80Gb/s optical packet routing for data networking using FPGA based 100Mb/s control scheme,” in *Proc. Eur. Conf. Opt. Commun.*, 2004, pp. 4–6.

[20] D. Rana, “A control algorithm for 3-stage non-blocking networks,” in *Proc. Global Telecommun. Conf.*, 1992, pp. 1477–1481.

[21] L. Wang and L. Theogarajan, “A micropower delta-sigma modulator based on a self-biased super inverter for neural recording systems,” in *Proc. IEEE Custom Integr. Circuits Conf.*, 2010, pp. 1–4.

[22] L. S. Theogarajan, “A low power fully implantable 15-channel retinal stimulator chip,” *IEEE J. Solid-State Circuits*, vol. 43, no. 10, pp. 2322–2337, Oct. 2008.

[23] Y. Tang, H.-W. Chen, S. Jain, J. D. Peters, U. Westergren, and J. E. Bowers, “50 Gb/s hybrid silicon traveling-wave electroabsorption modulator,” *Opt. Exp.*, vol. 19, no. 7, pp. 5811–5816, Mar. 2011.

[24] *The 50G Silicon Photonics Link*, Jul. 2010. [Online]. Available: http://download.intel.com/pressroom/pdf/photonics/Intel_SiliconPhotonics50gLink_WhitePaper.pdf?iid=pr_smrelease_vPro_materials2, white paper, Intel

[25] P. Gupta and N. McKeown, “Designing and implementing a fast crossbar scheduler,” *IEEE Micro*, vol. 19, no. 1, pp. 20–28, Jan./Feb. 1999.

[26] S. J. Ben Yoo, “Optical packet and burst switching technologies for the future photonic interconnect,” *J. Lightwave Technol.*, vol. 24, no. 12, pp. 4468–4492, Dec. 2006.