How Photonic Networking Can Help Data Centres

Ivan Glesk, Senior *Member, IEEE*, Adrianus Buis[*], and Alan Davidson
Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XU, UK
[*] Department of Biomedical Engineering, University of Strathclyde, Glasgow G1 1XU, UK
Tel: (+44) 141-548-2529, Fax: (+44) 141-552-4968, e-mail: ivan.glesk@strath.ac.uk

ABSTRACT
In light of rapidly increasing demand for ultra-high speed data transmission, data centres are under pressure to provide ever increasing data transmission through their networks and at the same time improve the quality of data handling in terms of reduced latency, increased scalability and improved channel speed for users. However as data rates increase, present electronic switching technology using current data centre architecture is becoming increasingly difficult to scale despite improved data management. In this paper electronic scalability issues will be discussed and alternative optical solutions will be reviewed including a novel and highly scalable optical interconnect.
Keywords: electronic bottleneck, optical interconnects, datacentre networking, optical switching, OCDMA.

1. INTRODUCTION

The amount of data traffic passing through communication networks is rising at a phenomenal rate. Global data centre IP traffic is forecast to reach 7.7 Zetabytes/year by 2017 - 69% of which will be cloud traffic, from its present level of 2.6 Zetabytes/year. This is a compound annual growth rate of 25% [1]. To maintain this rapidly increasing demand, data centres must scale to provide higher bandwidths while maintaining low latency and improved scalability. This expansion is attributable to the rise of smart phones/tablets, e-services, e-health, and the emergence of the cloud. Smart phones alone generate between 10 and 20 times more data traffic than conventional mobile phones [2]. These current trends have resulted in new bottlenecks in both data centres and transmission networks. New developments in cloud computing will compound this problem even more.

The growing demand for high speed, low latency data transmission has generated a need for substantially increased capacity and improved connectivity within data centres. However current data centres performing all data processing based on electronic switching are incapable of fulfilling these demands [3]. Therefore new technological solutions to meet demand are required. It has been suggested that the fundamental limits of data centre switching which relies on bandwidth limited CMOS electronics is now perhaps being reached [4]. It is believed that all-optical systems using photonic integrated circuits and highly scalable optical interconnects may provide an answer with the promise of data rates exceeding Terabits per second.

2. LIMITS TO ELECTRONIC SCALABILITY

Back in 1974, Robert Dennard described the MOSFET scaling rules crucial to reduce transistor size and at the same time increase switching speed and reduce power consumption [5]. These scaling principles were adopted by the semiconductor industry to develop and improve silicon devices and the terms 'Dennard Scaling' and 'Dennard's Law' were coined. Dennard scaling states that as transistor size reduces, power consumption will reduce while maximum switching speed will be increased. An important observation of Dennard scaling is that power density (Watts/unit chip area) has remained constant as transistor density has increased. In addition, the density of transistors has been increasing by a factor of about two every 18 months for over 40 years. 'Moore's law' first suggested such an exponential improvement in transistor density back in 1965 with an estimated doubling of transistor density every two years [6]. What has actually happened to date has validated 'Moore's law' despite the increasing technical challenges of miniaturisation. This trend is forecast by Intel to continue until around 2020 [7]. However, despite the continuation of Moore's law, Dennard scaling came to an end in 2005 with the development of 90nm lithography. At this level, transistor gates become too thin to prevent current from leaking into the substrate, resulting in a rise in power density. As a result, power consumption has been increasing dramatically and the performance per watt, while still rising, is doing so at a slower rate. However the resultant increased power density runs the risk of causing thermal runaway and destruction of the chip in addition to creating cooling challenges and power consumption cost issues when scaled to the size of a data centre. In the absence of a mature alternative technology to overcome the leakage current problem and avoid the above consequences, it is therefore necessary to reduce power consumption by reducing the clock frequency and therefore the processing speed. Consequently, since 2005 chip manufacturers Intel and AMD have concentrated on introducing parallel processing CPU's using multicore processors to increase processing power. While parallel processing can, to a degree compensate for limited clock frequencies, it clearly results in at least a linear increase in consumption since an effective doubling in performance requires two processors.
However, Amdahl's parallelism law states that 'If a computation has a serial component *S*% and a parallel

component *P*%, then the maximum speed-up is *(S+P)/S'*. Clearly, the greater the parallel portion *P*, the higher the speed-up. Amdahl's law says that there is a fundamental maximum improvement in computational speed that is dependent upon the proportion of serial computation, beyond which further additional parallel processors will contribute a rapidly diminishing improvement in processing speed. Consequently the performance per watt which is initially constant will eventually decrease rapidly as the number of processors is increased beyond the optimum number. This principle is illustrated in Fig. 1 for four Parallel processing portions *P* [8]. As the number of parallel processors increases, a maximum speed-up is achieved, beyond which the addition of further processors provides no additional computational advantage and only serves to increase power consumption. As

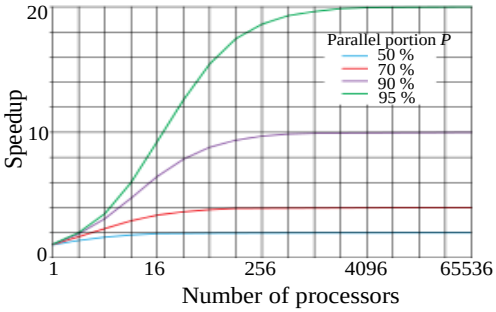

Figure 1. The effect of Amdahl's law.

demand grows, a data bottleneck will result with increases in contention and an associated increase in latency. It is therefore expected that data centres will become unable to comply with the minimum quality of service they are contracted to provide. In the longer term, the answer to satisfying the rapid increase in demand for processing power in data centres will therefore not be found using present MOSFET electronic technology. The fundamental limits of data switching in this bandwidth limited technology may be reached sooner rather than later. In the short term, while Moore's law holds and effective cooling can be achieved, processing speeds can continue to rise. However power density will no longer be constant as previously predicted by Dennard's law, but will increase with processor performance. However, the energy consumed per bit will continue to reduce, albeit at a slower rate than before. In the short term, increasing clock speeds to satisfy demand is therefore not a realistic solution. However measures to improve the power consumption of data centres have been suggested. For example, virtualisation can improve efficiency since fewer physical servers are required. In addition, research into reducing CMOS leakage currents has demonstrated improvements in data handling capacity and power efficiency [9]. Graphene [10] and nanowire [11] technologies are being investigated as a replacement for current CMOS based devices. However, research in these areas is in its infancy and operational devices are many years from market. If demand is to be met in the longer term, disruptive technologies will have to be developed that greatly reduce the energy consumption per bit. As a mature technology, photonics may hold the key to enable data centres to meet growing demand in the long term without incurring exponentially rising energy consumption and associated costs. At present, optical interconnects in the form of Intel's Light Peak technology are already possible for use in interconnecting servers to the TOR (top of the rack) switch with data rates of 10Gb/s over many tens of metres [12]. While this technology can overcome the attenuation and bandwidth limits of copper cables, it still requires E/O and O/E transceivers. However, since light cannot at present be stored, realisation of the all-optical data centre is still many years off. In the meantime, development of hybrid systems such as HELIOS has been suggested [13]. HELIOS is a 2-layer WDM based architecture using modular POD (performance optimised data centre) units. At the POD layer, packet switching is achieved using commodity switches; and at the core layer, optical circuit switching is used in parallel with commodity packet switches. Optical circuit switching is achieved using MEMS switches, facilitating the use of optical interconnects and reducing the requirement for power hungry transceivers and commodity switches.

It has been suggested that the fundamental limits of data centre switching which relies on bandwidth limited electronics is now perhaps being reached. It is becoming evident that all-optical systems using photonic integrated circuits and highly scalable optical interconnects will provide a long term answer with the promise of data rates exceeding Terabits per second. Recently, we demonstrated a novel method for increasing the scalability of an incoherent OCDMA system by employing the technique of OCDMA code reuse that enables different groups of OCDMA users to transmit in separate OTDMA channels [14]. We have shown that the scalability of such a hybrid system is increased by a factor of *M*, up to $M \times N$ where *N* is the number of OTDM channels and *M* is the number of simultaneous OCDMA users per OTDM time slot. In this paper we propose a new OCDMA architecture which together with the developed all-optical picosecond time gating based on ultra-fast silicon photonic switch will deliver an improved highly scalable interconnect.

3. IMPROVING SCALABILITY OF OPTICAL INTERCONNECTS

We will show that the scalability of the OCDMA system can be substantially enhanced by adding another

dimension to an OCDMA transmission by allowing multiple OCDMA user groups i where $i = (1,2,...,p)$ to transmit simultaneously within adjacent course C-WDM channels (see Fig. 2).

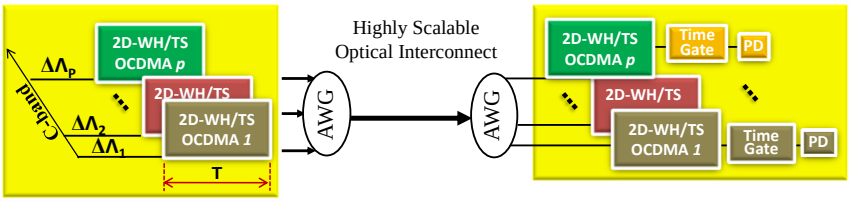

Figure 2. Conceptual diagram of a highly scalable optical interconnect. p 2D-WH/TS OCDMA subsystems broadcast over course C-WDM. Timing gates are used to suppress MAI: ($\Delta\Lambda_i$) spectral band occupied by individual OCDMA subsystems, (PD) photodetector.

Each C-WDM channel has been assigned a wavelength spacing of $\Delta\Lambda_i$. We will carry out scalability analysis where incoherent OCDMA subsystems are deployed and based on two dimensional wavelength-hopping time-spreading (2D-WH/TS) codes with multi-wavelength picosecond pulses as a code carrier. The code length is T where 1/T is the OCDMA subsystem data rate. Figure 2 shows a total of p such subsystems each having M users transmitting simultaneously over an optical interconnect. We now assume the OCDMA subsystem is uses a set of w wavelength pulses as the code carrier. Please note w (the number of wavelength used per code) is called the code weight. Each given code set with w wavelength pulses will therefore occupy one of the spectral bands $\Delta\Lambda_i$, where $i = (1,2,...,p)$. Here $\Delta\Lambda_i = \Lambda/w$, where Λ is the total spectral bandwidth available (in Fig. 2 shown as C-band). With this arrangement, the interconnect's overall capacity will be scaled up by a factor of P providing a total number of user connections $P \times M$. By way of example, the C-band consists of wavelengths ranging from 1535nm to 1565nm. By dividing the C-band using a 100 GHz (0.8nm) ITU standard will supply a total of 85 wavelengths available for use by all OCDMA subsystems. This allows a build-up to $P = 10$ (~ 85/8) of (8,47) 2D-WH/TS OCDMA or $P = 11$ *(~85/4)* of (4,47) 2D-WH/TS OCDMA subsystems, respectively.

Based on our calculations [14] it can be shown that a "standalone" 17-user (8,47) 2D-WH/TS OCDMA will achieve a raw BER = 10^{-9} at OC-48 data rates. Thanks to OCDMA soft blocking properties, if this system is designed for a raw BER = 10^{-4} which in turn is corrected back to 10^{-9} by a forward error correction technique (FEC), the same (8,47) 2D-WH/TS OCDMA will support up to $M = 37$ simultaneous user connections. Similarly it can be shown that a (4,47) 2D-WH/TS OCDMA subsystem using the same approach will support up to $M = 18$ simultaneous user connections. By taking into account all of the above, the novel OCDMA over a C-WDM interconnect will support a total of $P \times M = 10 \times 37 = 370$ user connections if based on ten (4,47) 2D-WH/TS OCDMA subsystems or $21 \times 18 = 378$ if based on twenty one (4,47) 2D-WH/TS OCDMA subsystems.

In summary, OCDMA over a C-WDM interconnect based on the proposed approach will be able to simultaneously transmit and receive data among up to $P \times M$ connections which is an increase of P if compared to a traditional "standalone" OCDMA approach. Depending on the target application the interconnect will be used for, the codes configuration of each OCDMA subsystem can either be the same (thereby achieving code reuse) or different code configurations can be arranged for each code set occupying its spectral band $\Delta\Lambda_i$. This will provide convenient flexibility allowing optimization of the OCDMA over a C-WDM optical interconnect to better fit the targeted application. It is also possible to implement coherent and/or incoherent OCDMA approaches including their combination as long as each OCDMA subsystem will occupy its own assigned

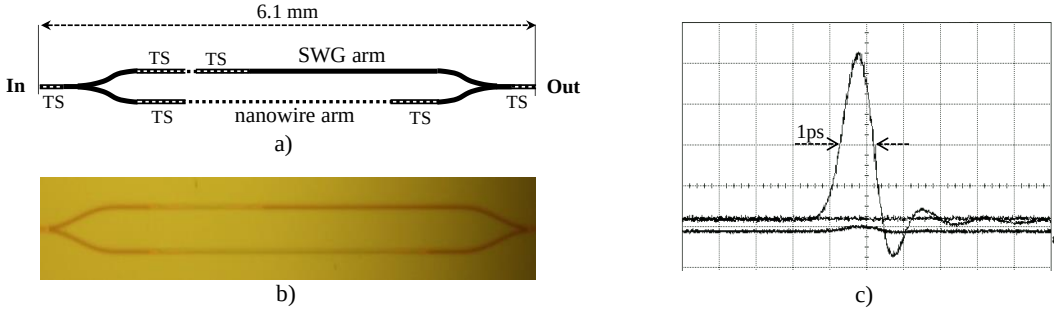

Figure 3. (a) Schematic diagram of a Timing Gate - picosecond all optical ultrafast switching device. (b) SMI image of manufactured device in (a). (c) experimental demonstration of picosecond switching capabilities of a device: (TS) tapered section, (SWG) subwavelength waveguide gating.

wavelength channel $\Delta\Lambda_i$. As with any OCDMA-based communication interconnect the overall performance will be impaired by the multi access interference noise (MAI) [15]. To suppress or even eliminate the MAI various techniques have been proposed. One very effective technique uses an ultrafast all-optical gating [16] based on a

time gate with a semiconductor optical amplifier (SOA). However the SOA suffers from gain recovery limitations [17]. To overcome this bottleneck, which ultimately limits the size of the gating windows to several picoseconds, we have developed a timing gate (TG) - which is an ultrafast all-optical photonic switch [18] which does not suffer the limitations imposed by the SOA and is capable of picosecond switching operations. The schematic diagram of the gate is shown in Fig. 3(a). The device has been fabricated and its SME image is shown in Fig. 3(b). Its performance has been tested and we have confirmed its picosecond switching capabilities (Fig. 3(c)).

The device has a Mach-Zehnder interferometric structure having one arm composed of a nanowire while the second arm is a subwavelength waveguide grating (SWG) structure. Tapered sections marked TS have been added to properly balance device properties and the loss to help achieve complete interferometric switching.

5. CONCLUSIONS

We have discussed the challenges data centres must face today. We have shown that an increasing amount of traffic and the electronic bottleneck are two major challenges which need to be overcome. In this challenging environment the role of advanced optical interconnects will play a major part. We have shown how optical interconnect scalability can be significantly improved by introducing Optical CDMA over CWDM. To further improve the performance we have developed and demonstrated a picosecond all-optical switching device capable of the elimination of MAI with a picosecond resolution.

REFERENCES

[1] "Cisco Global Cloud Index: Forecast and Methodology, 2012-2017," Internet: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index gci/Cloud_Index_White_Paper.html, [Mar. 30, 2014].

[2] "Explosive Growth in Data Traffic and the future of Global Communications Infrastructure," Internet: http://www.ntt.com/resource-centre/article/data/global-watch02.html, [Mar. 30, 2014].

[3] D. Barney, "The great cloud bottleneck: How capacity issues can kill your cloud project." Internet: http://redmondmag.com/articles/2011/12/01/cloud-bottleneck-issues.aspx, [Mar. 30, 2014].

[4] J. Hruska: The death of CPU scaling: From one core to many – and why we're still stuck, Internet: http//www.extremetech.com/computing/116561-the death-of-cpu-scaling-from-one-core-to-many-and-why-were-still-stuck Feb. 1, 2012 [Mar. 30, 2014].

[5] R.H. Dennard, et al.: "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9. pp. 256-268, 1974.

[6] G. Moore: Cramming More Components onto Integrated Circuits, *Electron. Mag*, vol. 38, April 1965.

[7] J. Hruska: "Intel's former chief architect: Moore's law will be dead within a decade," Internet: http//www.extremetech.com/computing/165331-intels-chief-architect-moores-law-will-be-dead-within-a-decade Aug, 30, 2013 [Mar. 30, 2014].

[8] Internet: http://en.wikipedia.org/wiki/Amdahl's_Law, [Mar. 30, 2014].

[9] S.K. Singh, B.K. Kaushik, D.S. Chauhan, S. Kum, Reduction of Subthreshold Leakage Current in MOS Transistors, *World Appl. Sciences J.*, vol. 25. pp. 446-450, 2013.

[10] "Better Graphene Transistors," Internet: http://www.technologyreview.com/news/409775/better-graphine-transistors/, [Mar. 30, 2014].

[11] "New 4-D Transistor is preview of future computers," Internet: http://www.purdue.edu/newsroom/releases/2012/Q4/new-4-d-transistor-is-preview-of-future-computers.html, [Mar. 30, 2014].

[12] Internet: http://www.intel.com/content/www/us/en/research/intel-labs-light-peak-tech-update-video.html, Mar. 30, 2014.

[13] N. Farrington, Helios: a hybrid electrical/optical switch architecture for modular data centres, *ACM SIGCOMM Computer Communication Rev.*, vol. 41. pp. 39-350, 2011.

[14] T. B. Osadola, S. K. Idris, I. Glesk, and W. C. Kwong, Network Scaling Using OCDMA over OTDM, *IEEE Photon. Tech. Lett.*, vol. 24. pp. 395-397 2012.

[15] G.-C. Yang and W. C. Kwong, *Prime Codes with Applications to CDMA Optical and Wireless Networks*, Artech House, Norwood, 2002.

[16] I. Glesk, et al., Incoherent Ultrafast OCDMA Receiver Design with 2 ps All-optical Time Gate to Suppress Multiple-Access Interference, *IEEE J. Sel. Top. Quantum Electron.* vol. 14. pp. 861-867, 2008.

[17] M. G. Kane, I. Glesk, J. P. Sokoloff, and P. R. Prucnal, Asymmetric Optical Loop Mirror: Analysis of an All-Optical Switch, *Applied Optics*, vol. 33. pp. 6833, 1994.

[18] I. Glesk, P. J. Bock, P. Cheben, J. H. Schmid, J. Lapointe, and S. Janz, "All-Optical Switching using Nonlinear Subwavelength Mach-Zehnder on Silicon," *Optics Express* **19** (15), 14031-14039 (2011).