

# Hybrid Optical Switching for Data Center Networks

by

Matteo Fiorani<sup>a</sup>, Slavisa Aleksic<sup>b</sup>, Maurizio Casoni<sup>a</sup>

<sup>a</sup> *Department of Engineering “Enzo Ferrari”, University of Modena and Reggio Emilia, Via Vignolese 905  
41125 Modena, Italy*

<sup>b</sup> *Institute of Telecommunications, Vienna University of Technology, Favoritenstrasse 9-11/E389  
1040 Vienna, Austria*

# Hybrid Optical Switching for Data Center Networks

Matteo Fiorani<sup>a,\*</sup>, Slavisa Aleksic<sup>b</sup>, Maurizio Casoni<sup>a</sup>

<sup>a</sup>Department of Engineering "Enzo Ferrari", University of Modena and Reggio Emilia, Via Vignolese 905, 41125 Modena, Italy

<sup>b</sup>Institute of Telecommunications, Vienna University of Technology, Favoritenstrasse 9-11/E389, 1040 Vienna, Austria

---

## Abstract

Current data centers networks rely on electronic switching and point-to-point interconnects. When considering future data center requirements, these solutions will raise issues in terms of flexibility, scalability, performance and energy consumption. For this reason several optical switched interconnects, which make use of optical switches and wavelength division multiplexing (WDM), have been recently proposed. However, the solutions proposed so far suffer from low flexibility and are not able to provide service differentiation. Furthermore, very few studies evaluate scalability and energy consumption and make extensive comparison with current data center networks. In this paper we introduce a novel data center network based on hybrid optical switching (HOS). HOS combines optical circuit, burst and packet switching on the same network. In this way different data center applications can be mapped to the optical transport mechanism that best suits to their traffic characteristics. Furthermore, the proposed HOS network achieves high transmission efficiency and reduced energy consumption by using two parallel optical switches; a slow and low power consuming switch for the transmission of circuits and long bursts, and a fast switch for the transmission of packets and short bursts. We consider the architectures of both a traditional data center network and the proposed HOS network and present a combined analytical and simulation approach for their performance and energy consumption evaluation. We demonstrate that the proposed HOS data center network achieves high performance and flexibility while considerably reducing the energy consumption of current solutions.

**Keywords:** Data center network, Optical switched interconnect, Optical circuit switching, Optical burst switching, Optical packet switching

---

## 1. Introduction

A data center (DC) refers to any large, dedicated cluster of computers that is owned and operated by a single organization. Mainly driven by emerging cloud computing applications data center traffic is showing an exponential increase. It has been estimated that for every byte of data transmitted over the Internet, 1 GByte are transmitted within or between data centers [1]. Cisco [2] reports that while the amount of traffic crossing the Internet is projected to reach 1.3 zettabytes per year in 2016, the amount of data center traffic has already reached 1.8 zettabytes per year, and by 2016 will nearly quadruple to about 6.6 zettabytes per year. This corresponds to a compound annual growth rate (CAGR) of 31% from 2011 to 2016. The main driver to this growth is cloud computing traffic that is expected to increase six-fold by 2016, becoming nearly two-thirds of total data center traffic. To keep up with these trends, data centers are improving their processing power by adding more servers. Already now large cloud computing data centers owned by on-line service providers such as Google, Microsoft, and Amazon host tens of thousands of servers in a single facility. With the expected growth in data center traffic, the number of servers per

facility is destined to increase posing a significant challenge to the data center interconnection network.

Another issue rising with the increase in the data center traffic is energy consumption. The direct electricity used by data center has shown a rapid increase in the last years. Koomey estimated [3, 4] that the aggregate electricity use for data centers worldwide doubled from 2000 to 2005. The rates of growth slowed significantly from 2005 to 2010, when the electricity used by data centers worldwide showed an increase by about 56%. Still, it has been estimated that data centers accounted for 1.3% of worldwide electricity use in 2010, being one of the major contributor to the worldwide energy consumption of the ICT sector.

The overall energy consumption of a data center can be divided in energy consumption of the IT equipment, energy consumption of the cooling system and energy consumption of the power supply chain. The ratio between the energy consumption of the IT equipment and the overall energy consumption represents the power efficiency usage (PUE). The PUE is an important metric that shows how efficiently companies exploit the energy consumed in their data centers. The average PUE among the major data centers worldwide is estimated to be around 1.80 [5], meaning that for each Watt of IT energy 0.8 W are consumed for cooling and power distribution. However, modern data centers show higher efficiency. Google declares that its most efficient data center shows a PUE as low as 1.12 [5]. We can then conclude that the major energy savings in

---

\*Corresponding author.

Email addresses: [matteo.fiorani@unimore.it](mailto:matteo.fiorani@unimore.it) (Matteo Fiorani),  
[slavisa.aleksic@tuwien.ac.at](mailto:slavisa.aleksic@tuwien.ac.at) (Slavisa Aleksic),  
[maurizio.casoni@unimore.it](mailto:maurizio.casoni@unimore.it) (Maurizio Casoni)

modern data centers can be achieved by reducing the power consumption of the IT equipment. The energy consumption of IT equipment can be further divided in energy consumption of the servers, energy consumption of the storage devices and energy consumption of the interconnection network. According to [6] current data centers networks consume around 23% of the total IT power. When increasing the size of data centers to meet the high requirements of future cloud services and applications, the internal interconnection network will most likely become more complex and power consuming [7]. As a consequence, the design of more energy efficient data center networks is of utmost importance for the scope of building greener data centers.

Current data centers networks rely on electronic switching elements and point-to-point (ptp) interconnects. The electronic switching is realized by commodity switches that are interconnected using either electronic or optical ptp interconnects. Due to the high cross talk and distance dependent attenuation very high data-rates over electrical interconnects can be hardly achieved. As a consequence, a large number of copper cables are required to interconnect a high-capacity data center, thereby leading to low scalability and high power consumption. Optical transmission technologies are generally able to provide higher data rates over longer transmission distances than electrical transmission systems, leading to increased scalability and reduced power consumption. Hence, recent high-capacity data centers are increasingly relying on optical ptp interconnection links. According to an IBM study [8] only the replacement of copper-based links with VCSEL-based ptp optical interconnects can reduce the power consumption of a data center network by almost a factor of 6. However, the energy efficiency of ptp optical interconnects is limited by the power hungry electrical-to-optical (E/O) and optical-to-electrical (O/E) conversion required at each node along the network since the switching is performed using electronic packet switching.

When considering future data center requirements, optical switched interconnects that make use of optical switches and wavelength division multiplexing (WDM) technology, can be employed to provide high communication bandwidth while reducing significantly the power consumption with respect to ptp solutions. It has been demonstrated in several research papers that solutions based on optical switching can improve both scalability and energy efficiency with respect to ptp interconnect architectures for data centers have been recently presented [11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. Some of the proposed architectures [11, 12] are based on hybrid switching with packet switching in the electronic domain and circuit switching in the optical domain. The others are based on all-optical switching elements, and rely either on optical circuit switching [14, 17] or on optical packet/burst switching [13, 15, 16, 18, 19]. In [20] electronic ToR switches are employed for intra-rack communications, while a WDM PON is used for inter-rack communications. Only a few of these studies evaluate the energy efficiency of the optical interconnection network and make comparison with existing solutions based on electronic switching [12, 17, 20]. Furthermore, only a small fraction of these architectures are proven to be scalable enough to keep up with the

expected increase in the size of the data centers [15, 18, 19]. Finally, none of this study addresses the issue of flexibility, i.e. the capability of serving efficiently traffic generated by different data centers applications.

With the worldwide diffusion of cloud computing, new data center applications and services with different traffic requirements are continuously rising. As a consequence, future data center networks should be highly flexible in order to serve each application with the required service quality while achieving efficient resource utilization and low energy consumption. To achieve high flexibility, in telecommunication networks hybrid optical switching (HOS) approaches have been recently proposed [21, 22]. HOS combines optical circuit, burst and packet switching on the same network and maps each application to the optical transport mechanism that best suits to its traffic requirements, thus enabling service differentiation directly in the optical layer. Furthermore, HOS envisages the use of two parallel optical switches. A slow and low power consuming optical switch is used to transmit circuits and long bursts, and a fast optical switch is used to transmit packets and short bursts. Consequently, employing energy aware scheduling algorithms, it is possible to dynamically choose the best suited optical switching element while switching off or putting in low power mode the unused ones.

Extending the work presented in [23], in this paper we propose a novel data center network based on HOS. The HOS switching paradigm ensures a high network flexibility that we have not found in the solutions proposed so far in technical literature. We evaluate the proposed HOS architecture by analyzing its performance, energy consumption and scalability. We compare the energy consumption of the proposed HOS network with a traditional network based on optical ptp interconnects. We demonstrate that HOS has potential for satisfying the requirements of future data centers networks, while reducing significantly the energy consumption of current solutions. The rest of the paper is organized as follows. In Section 2 we describe the optical ptp and HOS network architectures. In Section 3 we present the model used for the evaluation of energy consumption. In Section 4 we describe the performed analysis and discuss data center traffic characteristics. In Section 5 we present and discuss the results and, finally, in Section 6 we draw conclusions.

## 2. Data Centers Networks

### 2.1. Optical ptp Architecture

Figure 1 shows the architecture of a current data center based on electronic switching and optical ptp interconnects. Here, multiple racks hosting the servers are interconnected using a fat-tree 3-Tier network architecture [24]. The 3 tiers of the data center network are the edge tier, the aggregation tier and the core tier. In the edge tier the Top-of-Rack (ToR) switches interconnect the servers in the same rack. We assume that each rack contains  $N_S$  servers and that each server is connected to a ToR switch through a 1 Gbps link. Although in future data centers, servers might be connected using higher capacity links, the majority of current data centers still use 1 Gbps links, as reported

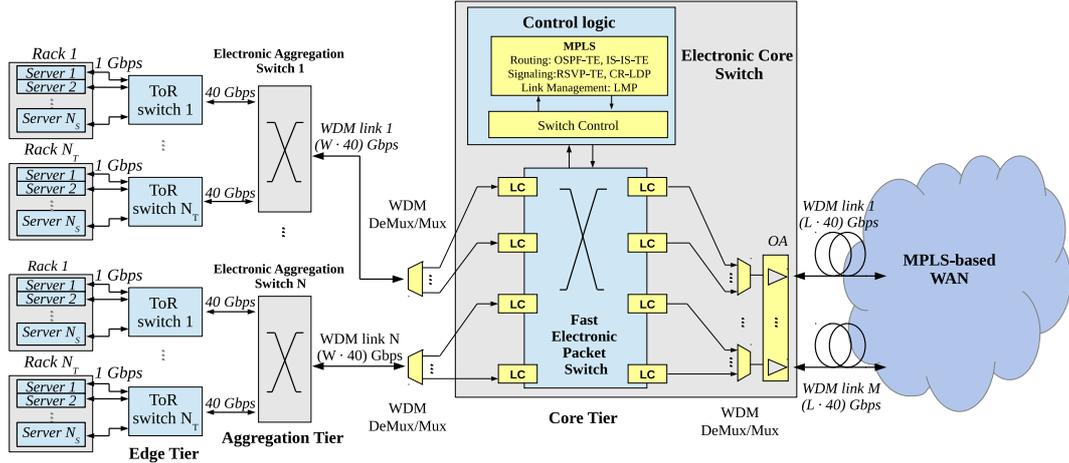


Figure 1: Architecture of a data center employing an optical ptp interconnection network. ToR: top of the rack, OA: optical amplifiers.

in [25]. In future works we plan to consider higher capacity per server port and evaluate the effect of increased server capacity. However, it is worth noting that the network performance (e.g., throughput and loss) do not depend on the line data rate, but on the link load, which we consider here as the percentage of the maximum link capacity.

As many as  $N_T$  ToR switches are connected to an aggregation switch using 40 Gbps links. The aggregation switches interconnect the ToR switches in the edge tier using a tree topology and are composed of a CMOS electronic switching fabric and electronic line cards (LC), that include power regulators, SRAM/DRAM memories, forwarding engine and LASER drivers. Each aggregation switch is connected to the electronic core switch through a WDM link composed of  $W$  wavelengths channels operated at 40 Gbps. The core switch is equipped with  $N \cdot W \cdot 40$  Gbps ports for interconnecting as many as  $N$  aggregation switches. Furthermore, the core switch employs  $M \cdot L \cdot 40$  Gbps ports for connecting the data center to a wide area network (WAN). We assume that the data center is connected to a WAN employing the MPLS control plane. It is worth noting that the considered optical ptp architecture employs packet switching in all the data center tiers.

The electronic core switch is a large electronic packet switch that comprises three building blocks, namely control logic, switching fabric and other optical components. The control logic comprises the MPLS module and the switch control unit. The MPLS module performs routing, signaling and link management as defined in the MPLS standard. The switch control unit performs scheduling and forwarding functionalities and drives the electronic switching elements. The switching fabric is a single electronic switch interconnecting a large number of electronic LCs. Finally, the other optical components include the WDM demultiplexers/multiplexers (WDM DeMux/Mux) and the optical amplifiers (OA) used as boosters to transmit toward the WAN.

In data centers with many thousands of servers, failures in the interconnection network may lead to losses of a high amount of important data. Therefore, resilience is becoming an increasingly critical requirement for future large-scale data center net-

works. However, the resilience is out of scope of this study and we do not address it in this paper, leaving it as an open issue for a future work.

## 2.2. HOS Architecture

The architecture of the proposed HOS optical switched network for data centers is shown in Figure 2. The HOS network is organized in a traditional fat-tree 3-Tier topology, where the aggregation switches and the core switches are replaced by the HOS edge and core node respectively. The HOS edge nodes are electronic switches used for traffic classification and aggregation. The HOS core node is composed by two parallel large optical switches. The HOS edge node can be realized by adding some minimal hardware modifications to current electronic aggregation switches. Only the electronic core switches should be completely replaced with our HOS core node. As a consequence, our HOS data center network can be easily and rapidly implemented in current data centers representing a good mid-term solution toward the deployment of a fully optical data center network. When higher capacities per server, e.g. 40 Gbps, will be required, operators can just connect the servers directly to the HOS edge switches without the need of passing through the electronic ToR switches. In this way it will be possible to avoid the electronic edge tier, meeting the requirements of future data centers and decreasing the total energy consumption. In the long term, it is possible also to think about substituting the electronic HOS edge switches with some optical devices for further increase the network capacity. This operation will not require any change in the architecture of the HOS core node, which can be easily scaled to support very high capacities. Furthermore, for increased overall performance and energy efficiency we assume that the HOS core node is connected to a HOS WAN [21, 22], but in general the core node could be connected to the Internet using any kind of network technology.

The architecture of a HOS edge node is shown in Figure 3. In the direction toward the core switch the edge node comprises three modules, namely classifier, traffic assembler and resource allocator. In the classifier, packets coming from the

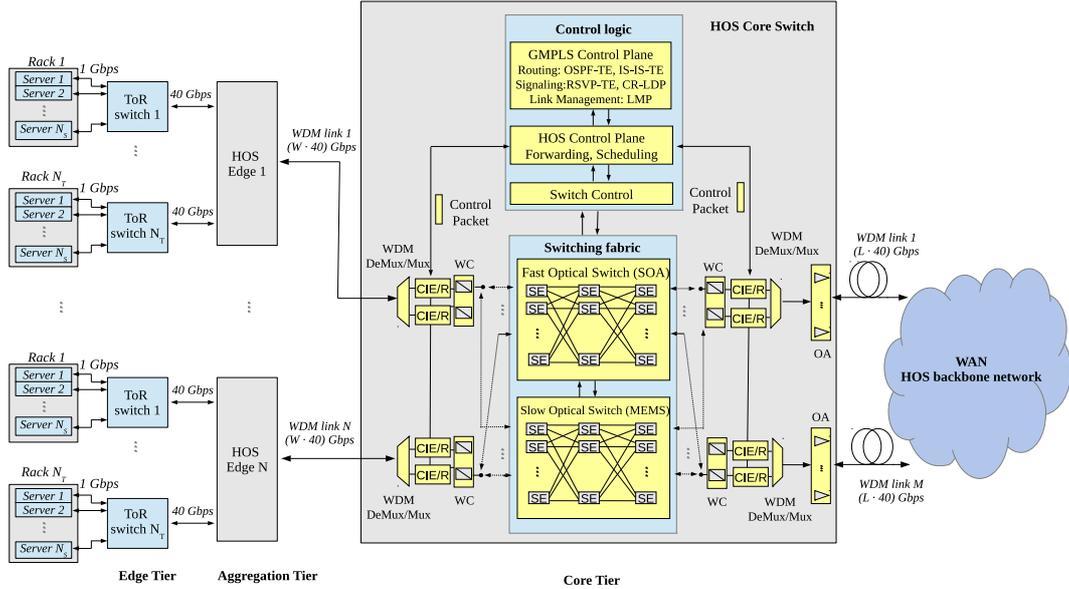


Figure 2: Architecture of a data center employing a HOS interconnection network. CIE/R: control information extraction/reinsertion, WC: wavelength converters.

ToR switches are classified based on their application layer requirements and are associated with the most suited optical transport mechanism. The traffic assembler is equipped with virtual queues for the formation of optical packets, short bursts, long bursts and circuits. Finally, the resource allocator schedules the optical data on the output wavelengths according to specific scheduling algorithms that aim at maximizing the bandwidth usage. In the direction toward the ToR switches a HOS edge node comprises packet extractors, for extracting packets from the optical data units, and an electronic switch for transmitting packets to the destination ToR switches.

As for the electronic core switch, we can divide the HOS core node in three building blocks, i.e. control logic, switching fabric and other optical components. The control logic comprises the GMPLS module, the HOS control plane and the switch control unit. The GMPLS module is used to ensure the interoperability with other core nodes connected to the WAN. The GMPLS module is needed only if the HOS core node is connected to a GMPLS-based WAN, such as the WAN proposed in [21, 22]. The HOS control plane manages the scheduling and transmission of optical circuits, bursts and packets. Three different scheduling algorithms are employed, one for each different data type, for optimizing the resource utilization and minimizing the energy consumption. A unique feature of the proposed HOS control plane is that packets can be inserted into unused TDM-slots of circuits with the same destination. This technique introduces several advantages, such as higher resource utilization, lower energy consumption and lower packet loss probability. For a detailed description of the HOS scheduling algorithms the reader is referred to [26]. Finally, the switch control unit creates the optical paths through the switching fabric. The switching fabric is composed of two optical switches, a slow switch for handling circuits and long bursts and a fast switch for the transmission of packets and short bursts. The

fast optical switch is based on semiconductor optical amplifiers (SOA) and its switching elements are organized in a non-blocking three-stage Clos network. In order to achieve high scalability, 3R regenerators are included after every 9th SOA stages to recover the optical signal, as described in [27]. The slow optical switch is realized using 3D micro electromechanical systems (MEMS). Finally, the other optical components include WDM DeMux/Mux, OAs, tunable wavelength converters (TWCs), and control information extraction/reinsertion (CIE/R) blocks. TWCs can convert the signal over the entire range of wavelengths, and are used to solve data contentions.

### 2.3. HOS Transport Mechanisms

In this Section we report a brief description of the HOS concept. For a more detailed explanation regarding the HOS data and control plane, we refer the reader to [22, 23, 26].

The proposed HOS network supports three different optical transport mechanisms, namely circuits, bursts and packets. The different transport mechanisms share dynamically the optical resources by making use of a common control packet that is sub-carrier multiplexed with the optical data. The use of a common control packet is a unique feature of the proposed HOS network that ensures high flexibility and high resource utilization. Each transport mechanism employs then a particular reservation mechanism, assembly algorithm and scheduling algorithm according to the information carried in the control packet. For a detailed description of the control plane the reader is referred to [26].

Circuits are long lived optical connections established between the source and destination servers. Circuits are established using a two-way reservation mechanism, with incoming data being queued at the HOS edge node until the reservation has been made through the HOS network. Once the connection has been established data are transmitted transparently toward

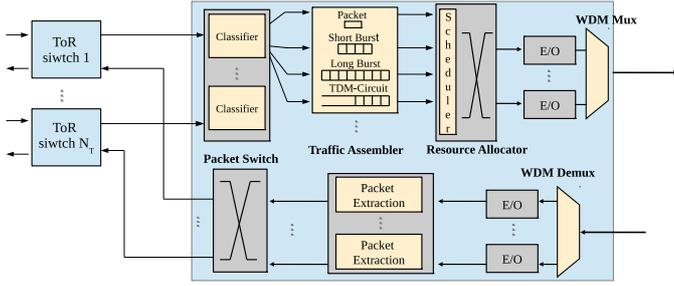


Figure 3: HOS edge node architecture.

the destination without any losses or delays other than the propagation delay. In the HOS network circuits are scheduled with the highest priority ensuring a very low circuit establishment failure probability. As a consequence, circuits are well suited for data centers applications with high service requirements and generating long-term point-to-point bulk data transfer, such as virtual machine migration and reliable storage. However, due to relatively long reconfiguration times, optical circuits provide low flexibility and are not suited for applications generating bursty traffic.

Optical burst switching has been widely investigated in telecommunication networks for its potential in providing high flexibility while keeping costs and power consumption bounded. In optical burst switching, before a burst is sent a control packet is generated and sent toward the destination to make an one-way resource reservation. The burst itself is sent after a fixed delay called offset-time. The offset-time ensures reduced loss probability and enables for the implementation of different service classes. In this paper we distinguish between two types of bursts, namely short and long bursts, which generate two different service levels. Long bursts are characterized by long offset-times and are transmitted using slow optical switching elements. To generate a long burst incoming data are queued at the HOS edge node until a minimum queue length  $L_{min}$  is reached. After  $L_{min}$  is reached, the burst is assembled using a mixed timer/length approach, i.e. the burst is generated as soon as the queue reaches  $L_{max} > L_{min}$  or a timer expires. The long offset-times ensure to long bursts a prioritized handling in comparison to packets and short bursts leading to lower loss probabilities. On the other side, the long offset-times and the long times required for burst assembly lead to large end-to-end delays. Short bursts are characterized by shorter offset-times and are transmitted using fast optical switching elements. To generate a short burst we use a mixed/timer length approach. The short burst is assembled as soon as the queue length reaches a fixed threshold or a timer expires. No minimum burst length is required, as was the case for the long bursts. The shorter offset-times and faster assembly algorithm lead to a higher loss probability and lower delays with respect to long bursts. In [23] we observed that bursts are suited only for delay-insensitive data center applications because of their high latency. Here, we were able to reduce the bursts latency by acting on the thresholds used in the short and long burst assemblers. Still, the bursts present remarkably higher delays than

packets and circuits and thus are suited for data-intensive applications that have no stringent requirement in terms of latency, such as MapReduce, Hadoop, and Dryad.

Optical packets are transmitted through the HOS network without any resource reservation in advance. Furthermore, packets are scheduled with the lowest priority. As a consequence they show a higher contention probability with respect to bursts, but on the other hand they also experience lower delays. However, the fact that packets are scheduled with the lowest priority leads to extra buffering delays in the HOS edge nodes, giving place to higher latency with respect to circuits. Optical packets are mapped to data centers applications requiring low latency and generating small and rapidly changing data flows. Examples of data center applications that can be mapped to packets are those based on parallel fast Fourier transform (MPI FFT) computation, such as weather prediction and earth simulation. MPI FFT requires data-intensive all-to-all communication and consequently requires frequent exchange of small data entities.

For a more detailed description of the HOS traffic characteristics we refer the reader to [21, 22].

### 3. Energy Consumption

We define the power consumption of a data center as the sum of the energy consumed by all of its active elements. In our analysis we consider only the power consumed by the network equipment and thus we exclude the power consumption of the cooling system, the power supply chain and the servers.

#### 3.1. Optical ptp Architecture

The power consumption of the optical ptp architecture is defined through the following formula:

$$P_{Net} = N_T \cdot N \cdot P_{ToR} + N \cdot P_{Aggr} + P_{Core} \quad (1)$$

where  $P_{ToR}$  is the power consumption of a ToR switch,  $P_{Aggr}$  the power consumption of an aggregation switch and  $P_{Core}$  the power consumption of the core switch. The ToR switches are conventional electronic Ethernet switches. Several large companies, such as HP, Cisco, IBM and Juniper, offer specialized Ethernet switches for use as ToR switch in data center networks. We estimated the power consumption of a ToR switch by averaging the values found in the data sheets released by these companies. With reference to Figures 1 and 2, without loss of generality we assume  $N_T = W$ . As a consequence we can assume that the aggregation switches are symmetric, i.e. they have the same number of input and output ports. From now on we will then use  $N_T$  to indicate also the number of wavelengths in the WDM links connecting the aggregation and core tiers. The power consumption of an aggregation switch  $P_{Aggr}$  is then given by the following formula:

$$P_{Aggr} = N_T \cdot (P_{CMOS} + P_{LC}) \quad (2)$$

Here,  $N_T$  is the number of input/output ports,  $P_{CMOS}$  is the power consumption per port of an electronic CMOS-based electronic switch and  $P_{LC}$  is the power consumption an electronic LC at 40 Gbps.

The power consumption of the electronic core switch is given by the sum of the power consumed by all its building blocks:

$$P_{Core} = P_{CL} + P_{SF} + P_{OC} \quad (3)$$

where  $P_{CL}$  is the power consumption of the control logic,  $P_{SF}$  is the power consumption of the switching fabric and  $P_{OC}$  is the power consumption of the other optical components.  $P_{CL}$  includes the power consumption of the MPLS module and the switch control unit. When computing  $P_{SF}$  we assume that the electronic ports are always active. This is due to the fact that current electronic switches do not yet support dynamic switching off or putting in low power mode of temporarily unused ports. The reason for that is because the time interval between two successive packets is usually too short to schedule the switching off of the electronic ports. As a consequence, we compute  $P_{SF}$  through the following formula:

$$P_{SF} = (N \cdot N_T + M \cdot L) \cdot (P_{CMOS} + P_{LC}) \quad (4)$$

where  $P_{LC}$  is the power consumption of an electronic LC and  $P_{CMOS}$  is again the power consumption per port of an electronic CMOS-based electronic switch. Finally,  $P_{OC}$  includes the power consumption of the OAs only, since the WDM De-Mux/Mux are passive components. In Table 1 the power consumption of all the elements introduced so far is reported. The values were obtained by collecting and averaging data from a number of commercially available components and modules of conventional switching and routing systems as well as from research papers. The Table shows that the main power drainers in a traditional data center network are the electronic LCs, which include the components for packet processing and forwarding [27]. A more detailed explanation on how to compute the power consumption of the electronic core switch is given in [26].

### 3.2. HOS Architecture

The power consumption of the HOS network architecture is obtained through the following formula:

$$P_{Net}^{HOS} = N_T \cdot N \cdot P_{ToR} + N \cdot P_{Edge}^{HOS} + P_{Core}^{HOS} \quad (5)$$

where  $P_{Edge}^{HOS}$  is the power consumption of the HOS edge node and  $P_{Core}^{HOS}$  is the power consumption of the HOS core node. The power consumption of the HOS edge node is obtained by summing the power consumption of all the blocks shown in Figure 3:

$$P_{Edge}^{HOS} = N_T \cdot (P_{Cs} + P_{As} + P_{PE} + P_{CMOS}) + P_{RA} \quad (6)$$

where  $P_{Cs}$  is the power consumption of the classifier,  $P_{As}$  is the power consumption of the traffic assembler, and  $P_{PE}$  is the power consumption of a packet extraction module. To compute the power consumption of the classifier and assembler we evaluated the average buffer size that is required for performing correct classification and assembly. We obtained an average required buffer size of 3.080 MByte. The assembler and classifier are realized with two large FPGAs equipped with external RAM blocks for providing the total required memory size of 3.080 MByte.  $P_{RA}$  represents the power consumption of the

Table 1: Values of power consumption of the components within the optical ptp and the HOS data center networks.

Components	Power [W]
Top of the Rack Switch ( $P_{ToR}$ )	650
<b>Aggregation Switch</b>	
Electronic Switch ( $P_{CMOS}$ )	8
Line card ( $P_{LC}$ )	300
<b>Electronic Core Switch</b>	
Control logic ( $P_{CL}$ )	27,096
Optical Amplifiers ( $1 \times$ port)	14
<b>HOS Edge Node</b>	
Classifier ( $P_{Cs}$ )	62
Assembler ( $P_{As}$ )	62
Resource Allocator ( $P_{RA}$ )	296
Packet Extractor ( $P_{PE}$ )	25
<b>HOS Core Switch</b>	
Control logic ( $P_{CL}^{HOS}$ )	49,638
SOA switch ( $P_{SOA}$ )	20
MEMS switch ( $P_{MEMS}$ )	0.1
Tunable Wavelength Converter ( $1 \times$ port)	1.69
Control Info Extraction/Re-insertion ( $1 \times$ port)	17

resource allocator. Again,  $P_{CMOS}$  is the power consumption per port of an electronic CMOS-based electronic switch. The power consumption of the HOS core node is obtained by summing the power consumption of the control logic, switching fabric and other optical components:

$$P_{Core}^{HOS} = P_{CL}^{HOS} + P_{SF}^{HOS} + P_{OC}^{HOS} \quad (7)$$

Here,  $P_{CL}^{HOS}$  is the sum of the power consumed by the GM-PLS module, the HOS control plane and the switch control unit. When computing  $P_{SF}^{HOS}$ , we assume that the optical ports of the fast and slow switches are switched off when they are inactive. This is possible because when two parallel switches are in use, only one must be active to serve traffic from a particular port at a specified time. In addition, because circuits and bursts are scheduled a priori, the traffic arriving at the HOS core node is more predictable than the traffic arriving at the electronic core switch. We then compute the power consumption of the HOS switching fabric through the following formula:

$$P_{SF}^{HOS} = N_{fast}^{AV} \cdot P_{SOA} + N_{slow}^{AV} \cdot P_{MEMS} \quad (8)$$

Here,  $N_{fast}^{AV}$  and  $N_{slow}^{AV}$  are respectively the average number of active ports of the slow and fast switches obtained through simulations.  $P_{SOA}$  and  $P_{MEMS}$  are respectively the power consumption per port of the SOA-based and MEMS-based switches. The average number of active ports for a specific configuration is obtained through simulations. Finally,  $P_{OC}^{HOS}$  includes the power consumption of OAs, TWCs and CIE/R blocks. The values used for the power consumption evaluation of the HOS data center network are included in Table 1. A more detailed expla-

nation on how to compute the power consumption of the HOS core node is given in [26, 27].

## 4. Modeling Approach

To evaluate the proposed HOS data center network we developed an event-driven C++ simulator. The simulator takes as inputs the parameters of the network and the data center traffic characteristics. The output produced by the simulator includes the network performance and energy consumption.

### 4.1. Data Center Traffic

In general traffic flowing through data centers can be broadly categorized into three main areas: traffic that remains within the data center, traffic that flows from data center to data center and traffic that flows from the data center to end users. Cisco [2] claims that the majority of the traffic is the one that resides within the data center accounting for 76% of all data center traffic. This parameter is important when designing the size of the data center and in particular the number of ports of the core node that connects the data center to the WAN. Basing on the information provided by Cisco, we designed our data center networks so that the number of ports connecting the core node to the WAN is 24% of the total number of ports of the core node.

In this paper we analyze the data center interconnection network, thus we simulate only the traffic that remains within the data center. To the best of our knowledge a reliable theoretical model for the data center network traffic has not been defined yet. However, there are several research papers that analyze data collected from real data centers [28, 29, 30]. Basing on the information collected in these papers, the inter-arrival rate distribution of the packets arriving at the data center network can be modeled with a positive skewed and heavy-tailed distribution. This highlights the difference between the data center environment and the wide area network, where a long-tailed *Poisson* distribution typically offers the best fit with real traffic data. The best fit [30] is obtained with the *lognormal* and *weibull* distributions that usually represent a good model for data center network traffic. We run simulation using both the *lognormal* and *weibull* distributions. In order to analyze the performance at different network loads, we considered different values for the mean and standard deviation of the *lognormal* distribution as well as for the shape and scale parameters of the *weibull* distribution.

In the considered data center networks, the flows between servers in the same rack are handled by the ToR switches and thus they do not cross the aggregation and core tiers. We define the *intra-rack traffic ratio (IR)* as the ratio between the traffic directed to the same rack and the total generated traffic. According to [28, 29, 30], the IR fluctuates between 20% and 80% depending on the data center category and the applications running in the data center. The IR impacts both performance and energy consumption of the HOS network and thus we run simulations with different values for the IR. The IR ratio has instead a negligible impact on the energy consumption of the optical ptp network. This is due to the fact that in the optical ptp network we do not consider switching off of the core switch ports

when inactive and thus the power consumption is constant with respect to the network traffic characteristics.

In our analysis we set the number of blade servers per rack to 48, i.e.  $N_S = 48$ , that is a typical value used in current high-performance data centers. Although a single rack can generate as much as 48 Gbps, the ToR switches are connected to the HOS edge nodes by 40 Gbps links leading to an over-subscription ratio of 1.2. Over-subscription relies on the fact that very rarely servers transmit at their maximum capacity because very few applications require continuous communication. It is often used in current data center networks to reduce the overall cost of the equipment and simplify data center network design. As a consequence, the aggregation and core tiers of a data center are designed to have a lower capacity with respect to the edge tier.

When simulating the HOS network, we model the traffic generated by the servers so that about 25% of the flows arriving at the edge nodes require the establishment of a circuit, 25% are served using long bursts, 25% are served with short bursts, and the remaining 25% are transmitted using packet switching. We do not consider in this paper the impact of different traffic patterns, i.e. the portions of traffic served by circuits, long bursts, short bursts and packets. In fact, we already evaluated this effect for core networks in [21], where we showed that an increase in traffic being served by circuits leads to slightly higher packet losses and a more evident increase of burst losses. Since in this paper we employ the same scheduling algorithms as in [21], we expect a similar dependence of the performance on the traffic pattern.

### 4.2. Performance Metrics

In our analysis we evaluate the performance, scalability and energy consumption of the proposed HOS data center network.

As regards the performance, we evaluate the average data loss rates and the average delays. When computing the average loss rates, we assume that the ToR switches and HOS edge nodes are equipped with electronic buffers with unlimited capacity and thus they do not introduce data losses. As a consequence, losses may happen only in the HOS core node. The HOS core node does not employ buffers to solve data contentions in the time domain, but is equipped with TWCs for solving data contentions in the wavelength domain. We consider one TWC per port with full conversion capacity, i.e. each TWC is able to convert the signal over the entire range of wavelengths. We define the packet (burst) loss rate as the ratio between the number of dropped packets (bursts) and the total number of packets (bursts) that arrive at the HOS core switch. Similarly, the circuit establishment failure probability is defined as the ratio between the number of negative-acknowledged and the total number of circuit establishment requests that arrive at the HOS core switch. The delay is defined as the time between a data packet is generated by the source server and when it is received by the destination server. We assume that the IR traffic is forwarded by the ToR switches with negligible delay, and thus we analyze only the delay of the traffic between different racks, i.e. the traffic that is handled by the HOS edge and core nodes. The delay is given by the sum of the propagation delay

Table 2: Reference data center configuration.

Parameter	Value
Number of servers per rack ( $N_S$ )	48
Number of ToR per HOS edge node ( $N_T$ )	64
Number of HOS edge nodes ( $N$ )	32
Number of wavelengths per fiber ( $W$ )	64
Number of servers in the data center ( $N_S^{tot}$ )	98,304
Intra-rack traffic ratio (IR)	40%
Traffic distribution	Lognormal
Network load	[65,80]%

and the queuing delay, i.e.  $D = D_p + D_q$ . The propagation delay  $D_p$  depends only on the physical distance between the servers. The physical distance between servers in a data center is usually limited to a few hundreds of meters, leading to negligible values for  $D_p$ . We then decided to exclude  $D_p$  from our analysis and consider  $D = D_q$ . The queuing delay includes the queuing time at the ToR switch and the delays introduced by the traffic assembler and resource allocator in the HOS edge switch ( $D_q = D_{ToR} + D_{as} + D_{ra}$ ). The HOS optical core switch does not employ buffers and thus does not introduce any queuing delay. We refer to the packet delay as to the average delay of data packets that are transmitted through the HOS core node using packet switching. Similarly, we define the short (long) burst delay as the average delay of data packets that are transmitted through the HOS core node using short (long) burst switching. Finally, the circuit delay is the average delay of data packets that are transmitted through the HOS core node using circuit switching.

As regards the scalability, we analyze our HOS network for different sizes of the data center. In general, data centers can be categorized in three classes: university campus data centers, private enterprise data centers, and cloud computing data centers. While university campus and private enterprise data centers have usually up to a few thousands of servers, cloud computing data centers, operated by large service providers, are equipped with up to tens or even hundreds thousand of servers. In this paper we concentrate on large cloud computing data centers. As a consequence, we vary the data center size from a minimum of 25K servers up to a maximum of 200K servers.

As regards the energy consumption, we compute the total power consumed by the HOS and the optical ptp networks using the analytical model described in Section 3. To highlight the improvements introduced by our HOS approach, we compare the two architectures in terms of energy efficiency and total greenhouse gas (GHG) emissions. The energy efficiency is expressed in Joule of energy consumed per bit of successfully transmitted data. The GHG emissions are expressed in metric kilotons (kt) of carbon dioxide equivalent ( $CO_{2e}$ ) generated by the data center networks per year. To compute the GHG emissions, we apply the conversion factor of 0.356  $KgCO_{2e}$  emitted per kWh, which was found in [31].

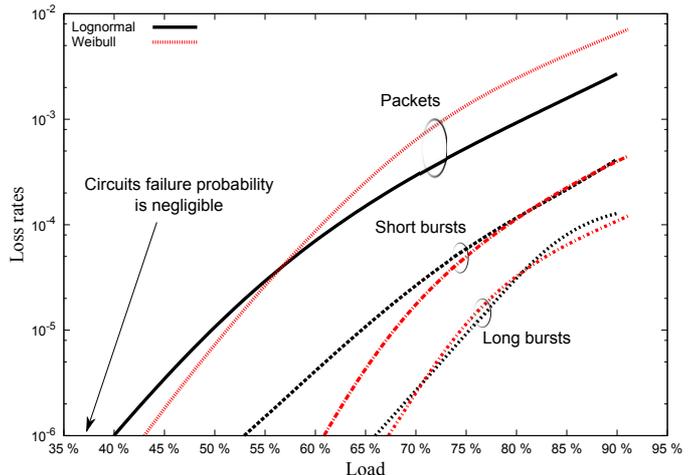


Figure 4: Average data loss rates in the HOS network as a function of the input load. Two different traffic distributions, i.e. lognormal and weibull, are considered.

## 5. Numerical Results

In this Section we show and discuss the results of the performed study. Firstly, we present the data loss rates, secondly, we report the network delays, and, finally, we analyze the energy consumption. We take into consideration several parameters, namely: network load, number of servers, traffic distribution, and IR ratio. We define the load as the ratio between the total amount of traffic offered to the network and the maximum amount of traffic that can be handled by the network. On the other side, the number of servers is given by  $N_S^{tot} = N_S \cdot N_T \cdot N$  and represents the sum of all the servers in the data center. Finally, for the traffic distribution and the IR ratio, we refer to the definitions provided in Section 4.1. The reference data center configuration is reported in Table 2. In the following, we evaluate the response of the network in terms of performance and energy consumption.

### 5.1. Loss Rates

In this Section we show and discuss the average data loss rates in the HOS network.

In Figure 4 we show the average data loss rates in the HOS network as a function of the input load. Two different distributions for the inter-arrival time of the traffic generated by the servers are considered, i.e. lognormal and weibull. Figure 4 shows that the data loss rates with the lognormal and weibull distributions present the same trend and very similar values. In the case of the weibull distribution the loss rates are slightly lower at low and medium loads, but they increase more rapidly with increasing the load. At high loads the loss rates obtained with the weibull distribution are similar or slightly higher than the loss rates obtained with the lognormal distribution. This effect is particularly evident for the packet loss probability, where the loss rates obtained with the two distributions are more different. Figure 4 also shows that the packet loss rates are always higher than the burst loss rates. This is due to the fact that for packets there is no resource reservation in advance. Due

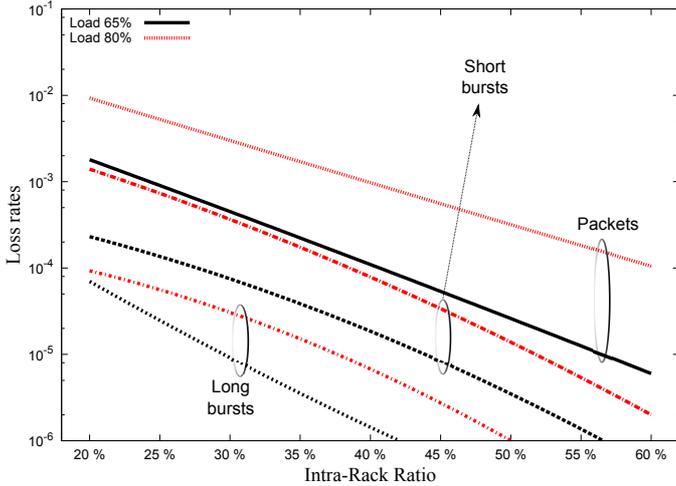


Figure 5: Average data loss rates in the HOS network as a function of the IR at 65% and 80% of offered load.

to shorter offset-times, the short bursts show higher loss rates with respect to long bursts, especially for low and moderate loads. Finally, we observe that the circuit establishment failure probability is always null. We conclude that data center applications having stringent requirements in terms of data losses can be mapped on TDM-circuits or long bursts, while applications that are less sensitive to losses can be mapped on optical packets or short bursts.

In Figure 5 the average data loss rates as a function of the IR are shown. The IR has been varied from 20% to 60%. The Figure shows that the higher is the IR and the lower are the data loss rates. This is due to the fact that a higher IR leads to a lower amount of traffic passing through the core switch, thus leading to a lower probability of data contentions. While increasing IR from 20% to 60% the packet and short burst loss rates decrease respectively by two and three orders of magnitude. It can also be observed that the difference between the loss rates at 65% and 80% of input load becomes more evident at higher IRs. The circuit established failure probability is always null.

Finally, in Figure 6 we show the data loss rates as a function of the number of servers in the data center. When changing the size of the data center, we changed both the number of ToR switches per HOS edge node ( $N_T$ ) and the number of HOS edge nodes ( $N$ ). We always consider  $N_T = W$ , in order to have symmetric HOS edge nodes. As a consequence, the higher is  $N_T$  and the higher is the number of wavelengths in the WDM links. The smallest configuration was obtained by setting  $N = 22$  and  $N_T = 24$ , achieving a total number of 25,344 servers in the data center. The largest configuration was obtained by setting  $N = 50$  and  $N_T = 84$  achieving a total number of 201,600 servers. Figure 6 shows that the higher is the size of the data center network and the lower are the loss rates introduced by the HOS core node. This is due to the fact that in our analysis a higher data center size corresponds to a higher number of wavelengths per WDM link. Since the HOS core node relies on TWCs to solve data contentions, the higher is the number of wavelengths per fiber and the higher is the probability to

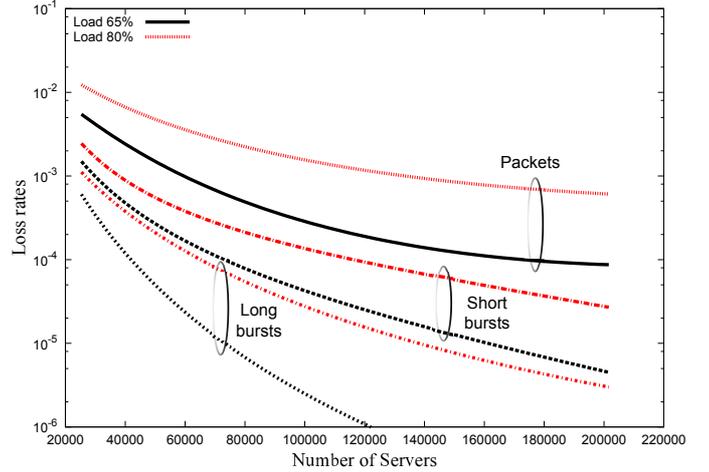


Figure 6: Average data loss rates in the HOS network as a function of the number of servers in the data center. Two different values for the input load, namely 65% and 80%, are considered.

find an available output resource for the incoming data. This is a unique and very important feature of our HOS data center network, that results in high scalability. In fact, increasing the number of wavelengths per fiber ( $N_T = W$ ) we can scale the size of the data center while achieving an improvement in the network performance. Figure 6 shows that the loss rates, especially the loss rates of the long bursts, decrease by more than one order of magnitude while increasing the number of servers from 25K to 200K.

## 5.2. Delays

In this Section we address the network latency. Since there are differences of several orders of magnitude between the delays of the various traffic types, we plotted the curves using a logarithmic scale.

In Figure 7 the average delays as a function of the input load are shown for two different distributions of the inter-arrival times of packets generated by the servers. The Figure shows that the delays obtained with the lognormal and weibull distributions show the same trends. The largest difference is observed for the delays of packets at high input loads, with the delays obtained with the weibull distribution being slightly higher. Figure 7 also shows that circuits introduce the lowest delay. To explain this result let us recall the definition of end-to-end delay  $D = D_{ToR} + D_{as} + D_{ra}$ . For circuits the assembly delay  $D_{as}$  is related to the circuit setup delay. Since in our network the circuit setup delay is several orders of magnitude lower than the circuit duration, its effect on the average end-to-end delay is negligible. Furthermore, circuits are scheduled with the highest priority by the resource allocator resulting in negligible  $D_{ra}$ . As a consequence, the circuit delay is determined mainly by the delay at the ToR switches  $D_{ToR}$ . As can be seen from Figure 7, circuits ensure an average delay below  $1.5 \mu s$  even for network loads as high as 90%. These values are in line with those presented in [15, 18, 19], where very low-latency optical data center networks are analyzed, and are suitable for applications with very strict delay requirements. Packets also do not suffer from

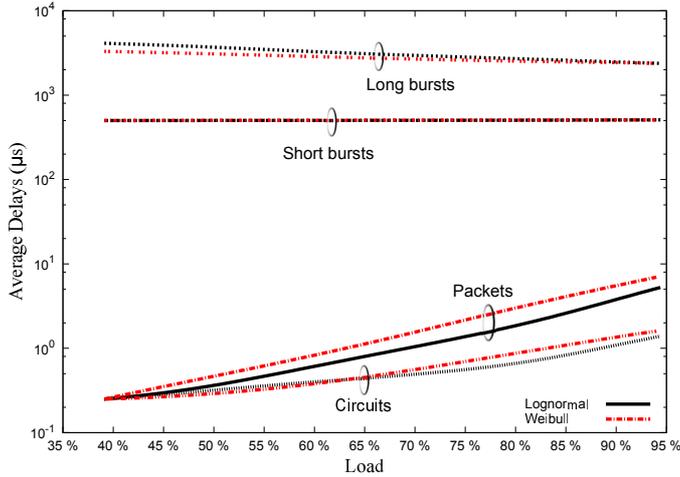


Figure 7: Average delays in the HOS network as a function of the input load and for two different input traffic distributions, namely lognormal and weibull.

any assembly delay, i.e.  $D_{as} = 0$ , but they are scheduled with low priority in the resource allocator resulting in non negligible values for  $D_{ra}$ . However, it can be observed that the packet delay remains below  $1 \mu s$  up to 65% of input load. For loads higher than 65% the packet delays grow exponentially, but they remain bounded to a few tens of  $\mu s$  even for loads as high as 90%. These values are similar to those presented for other optical packet switched architectures, e.g. [20], and are suitable for the majority of today's delay-sensitive data center applications.

Short and long bursts are characterized by very high traffic assembler delays  $D_{as}$ , which are given by the sum of the time required for the burst assembly and the offset-time. The traffic assembler delay is orders of magnitude higher than  $D_{ToR}$  and  $D_{ra}$  and thus the end-to-end delay can be approximated with  $D_{as}$ . In order to reduce the bursts delays obtained in [23] we acted on the timers and the length thresholds of the burst assemblers. We optimized the short and long burst assemblers and strongly reduced the bursts delays. Still, short and long bursts delays are respectively one and two order of magnitude higher than packets delays, making bursts suitable only for delay-insensitive data center applications. Figure 7 shows that short bursts present an almost constant delay attested around  $500 \mu s$ . Instead, the long burst delay decreases while increasing the input load. This is due to the fact that the higher is the rate of the traffic arriving at the HOS edge node and the shorter is the time required for reaching the long burst threshold  $L_{min}$  and starting the process for generating of the burst. The minimum long burst delay, which is obtained for very high input loads, is around 2 ms. This delay is quite high for the majority of current data center applications and raises the question if it is advisable or not to use long bursts in future data center interconnects. On the one hand long bursts have the advantage of introducing low loss rates, especially at low and moderate loads, and reducing the total power consumption, since they are forwarded using slow and low power consuming switching elements. On the other hand, it may happen that a data center provider does not have any suitable application to map on long bursts due to their high latency. If this is the case, the provider

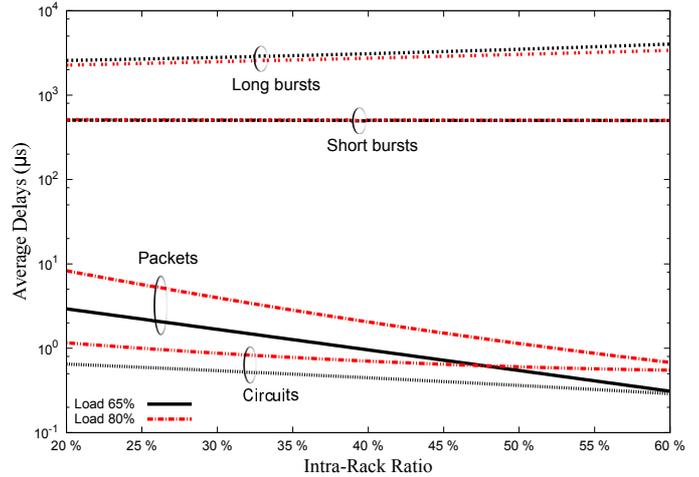


Figure 8: Average delays in the HOS network as a function of the IR at 65% and 80% of offered load.

could simply switch off the long burst mode and run the data center using only packets, short bursts and circuits. This highlights the flexibility of our HOS approach, i.e. the capability of the HOS network to adapt to the actual traffic characteristics.

In Figure 8 we show the average delays in the HOS network as a function of the IR. The Figure shows that the circuits and packets delay decrease while increasing the the IR traffic. This is due to the fact that the higher is IR and the lower is the traffic that crosses the ToR switches and the HOS edge nodes in the direction toward the HOS core node. This leads in turn to lower  $D_{ToR}$  and lower  $D_{ra}$ . In particular, when IR is as high as 60% the  $D_{ra}$  for packets becomes almost negligible and the packets delays become almost equal to the circuits delays. As for the long bursts, the higher is IR and the higher are the delays. In fact, a higher IR leads to a lower arrival rate at the HOS edge nodes and, consequently, to a longer assembly delay  $D_{as}$ . Finally, the short burst delay is almost constant with respect to IR.

In Figure 9 we show the average delays as a function of the number of servers in the data center. The Figure shows that increasing the size of the HOS data center leads to a slight decrease of the end-to-end delays. To explain this fact it is worth remembering that when increasing the number of servers we also increase the number of wavelengths per fiber  $W = N_T$  in the WDM links. The higher is  $N_T$  and the lower is the time required by the resource allocator to find an available output resource where to schedule the incoming data, i.e. the higher is  $N_T$  and the lower is  $D_{ra}$ . This fact again underlines the scalability of the proposed HOS solution.

### 5.3. Energy Consumption

In this Section we present and compare the energy efficiency and GHG emissions of the HOS and the optical ptp data center networks.

In Figure 10 the energy consumption per bit of successfully delivered data is shown as a function of the input load. In the case of the HOS network we consider three different values for IR, namely 20%, 40% and 60%. The energy consumption of

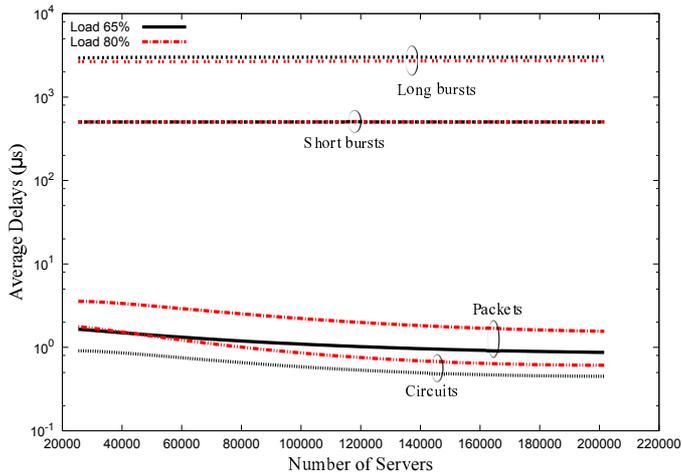


Figure 9: Average delays in the HOS network as a function of the number of servers in the data center. Two different values for the input load, namely 65% and 80%, are considered.

the optical ptp network is independent with respect to the IR. Firstly, we consider the overall energy consumption of the data center network and thus we include in our analysis the power consumption of the ToR switches. The electronic ToR switches are the major contributor to energy consumption especially for the HOS network where they consume more than 80% of the total. In the optical ptp network ToR switches are responsible for around 50% of the total energy consumption. Figure 10 shows that the proposed HOS network provides energy savings in the range between 31.5% and 32.5%. The energy savings are due to the optical switching fabric of the HOS core node that consumes considerably less energy with respect to the electronic switching fabric of the electronic core switch. Furthermore, the HOS optical core node is able to adapt its power consumption to the current network usage by switching off temporarily unused ports. This leads to additional energy savings especially at low and moderate loads when many ports of the switch are not used. However, the improvement in energy efficiency provided by HOS is limited by the high power consumption of the electronic ToR switches. In order to evaluate the relative improvement in energy efficiency provided by the use of HOS edge and core switches instead of traditional aggregate and core switches, we show in Figure 10 also the energy efficiency obtained without the energy consumption of the ToR switches. It can be seen that the relative gain offered by HOS is between 75% and 76%. The electronic ToR switches limit then by more than two times the potential of HOS in reducing the data center power consumption, raising the issue for a more energy efficient ToR switch design. Finally, Figure 10 shows that the energy efficiency of the HOS network depends only marginally on the IR traffic ratio. While increasing the IR ratio the energy consumption decreases because a higher IR ratio leads to a lower amount of traffic crossing the HOS core node. Due to the possibility of switching off unused ports, the lower is the amount of traffic crossing the HOS core node and the lower is its energy consumption.

Figure 11 shows the GHG emissions per year of the HOS

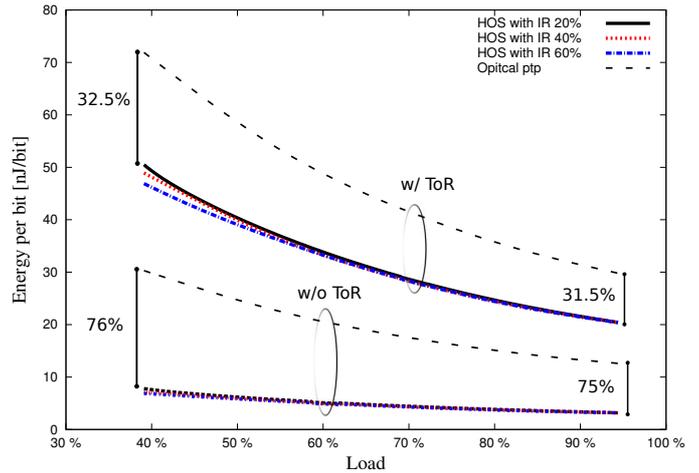


Figure 10: Energy consumption per bit of successfully transmitted data for the HOS and optical ptp networks. Both the cases with and without the ToR switches are shown.

and the optical ptp networks versus the number of servers in the data center. Again we show both the cases with and without the ToR switches. The Figure illustrates that the GHG emissions increase linearly with the number of servers in the data center. In both the cases with and without the ToR switches the GHG emissions of the HOS architecture are significantly lower than the GHG emissions of the optical ptp architecture. In addition, the slopes of the GHG emission curves of the HOS network are lower. In fact, while increasing the number of servers from 25K to 200K the reduction in GHG emissions offered by the HOS network increases from 30% to 32.5% when the power consumption of the ToR switches is included and from 71% to 77% when the power consumption of the ToR switches is not included. This is due to the fact that the power consumption of all the electronic equipment depends linearly on the size, while the power consumption of the optical slow switch does not increase significantly with the dimension. As a consequence, the power consumption of the HOS core node increases slower than the power consumption of the electronic core switch. This leads to a higher scalability of the HOS network with respect to the optical ptp network. Figure 11 also shows that when including the energy consumption of the ToR switches the gain offered by the HOS architecture is strongly reduced, highlighting again the need for a more efficient ToR switch design.

## 6. Conclusions

To address the limits of current ptp interconnects for data centers, in this paper we proposed a novel optical switched interconnect based on hybrid optical switching (HOS). HOS integrates optical circuit, burst and packet switching within the same network, so that different data center applications are mapped to the optical transport mechanism that best suits to their traffic characteristics. This ensures high flexibility and efficient resource utilization. The performance of the HOS interconnect, in terms of average data loss rates and average delays, has been evaluated using event-driven network sim-

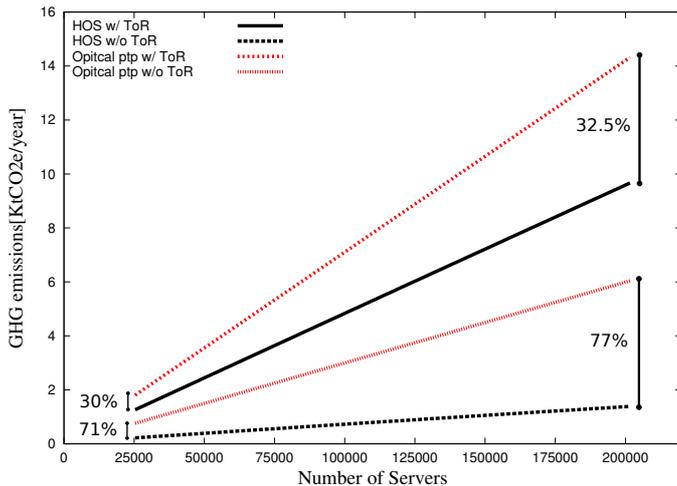


Figure 11: Greenhouse gas emissions per year of the HOS and the optical ptp networks as a function of the size of the data center.

ulations. The obtained results prove that the HOS network achieves relatively low loss rates and low delays, which are suitable for today's data center applications. In particular, we suggest the use of circuits for carrying premier traffic and packets for serving best-effort traffic. Bursts can be used to provide different QoS classes, but their characteristics should be carefully designed to avoid the risk of high network delays.

The proposed HOS architecture envisages the use of two parallel optical core switches for achieving both high transmission efficiency and reduced energy consumption. Our results show that the HOS interconnect reduces by a great extent the energy consumption and GHG emissions of data center interconnects with respect to current point-to-point solutions. Furthermore, the HOS interconnect requires limited hardware modifications to existing architectures and thus can be implemented in the short/mid term and with modest investments for the operators. An open question that we plan to address in detail in future works is how efficiently the HOS interconnect is able to scale with the increase in the servers capacity.

## References

- [1] Greg Astfalk, "Why optical data communications and why now?," *Applied Physics A* (2009) 95: 933-940. DOI 10.1007/s00339-009-5115-4.
- [2] Cisco white paper: "Cisco Global Cloud Index: Forecast and Methodology, 2011-2016," 2012.
- [3] J. Koomey, "Worldwide electricity used in data centers.," *Environmental Research Letters*, Vol. 3, No. 034008, September 2008.
- [4] J. Koomey, "Growth in Data Center Electricity use 2005 to 2010," *The New York Times* (2011), Volume: 49, Issue: 3, Pages: 24.
- [5] <http://www.google.com/about/datacenters/efficiency/internal/>
- [6] Where does power go? GreenDataProject, available online at: <http://www.greendataport.org>, 2008.
- [7] S. Aleksic, G. Schmid, N. Fehratovic, "Limitations and perspectives of optically switched interconnects for large-scale data processing and storage systems," (invited), *MRS Proceedings*, Cambridge University Press, Vol. 14382012, pp. 1-12.
- [8] A. Benner, "Optical Interconnect Opportunities in Supercomputers and High End Computing," *OFC 2012, OTu2B.4*.
- [9] N. Fehratovic, S. Aleksic, "Power Consumption and Scalability of Optically Switched Interconnects for High-Capacity Network Elements,"

- Proc. OFC 2010, pp. JWA84-1-JWA84-3, Los Angeles, USA, March, 2010.
- [10] S. Aleksic, N. Fehratovic, "Requirements and Limitations of Optical Interconnects for High-Capacity Network Elements," *Proc. ICTON 2010*, pp. We.A1.2-1-We.A1.2-4, Munich, Germany, June, 2010.
- [11] G. Wang, D. G. Andersen, M. Kaminsky, K. Papagiannaki, T. E. Ng, M. Kozuch, and M. Ryan, "c-Through: Part-time Optics in Data Centers," *Proc. ACM SIGCOMM 2010*, pp. 327-338.
- [12] N. Farrington, G. Porter, S. Radhakrishnan, H. H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, and A. Vahdat, "Helios: a hybrid electrical/optical switch architecture for modular data centers," *Proc. ACM SIGCOMM 2010*, 2010, pp. 339-350.
- [13] X. Ye, Y. Yin, S. J. B. Yoo, P. Mejia, R. Proietti, and V. Akella, "DOS: A scalable optical switch for datacenters," *Proc. ACM/IEEE Symposium on Architectures for Networking and Communications Systems*, 2010, pp. 24:1-24:12.
- [14] A. Singla, A. Singh, K. Ramachandran, L. Xu, and Y. Zhang, "Proteus: a topology malleable data center network," *Proc. ACM SIGCOMM*, 2010, pp. 8:1-8:6.
- [15] K. Xia, Y.H. Kaob, M. Yangb, and H. J. Chao, "Petabit Optical Switch for Data Center Networks," Technical report, Polytechnic Institute of NYU, 2010.
- [16] R. Luijten, W. E. Denzel, R. R. Grzybowski, and R. Hemenway, "Optical interconnection networks: The OSMOSIS project," 17th Annual Meeting of the IEEE Lasers and Electro-Optics Society, 2004.
- [17] O. Liboiron-Ladouceur, I. Cerutti, P. Raponi, N. Andrioli, and P. Castoldi, "Energy-efficient design of a scalable optical multiplane interconnection architecture," *IEEE J. Sel. Topics Quantum Electron.*, no. 99, pp. 1-7, 2010.
- [18] A. Shacham and K. Bergman, "An experimental validation of a wavelength-stripped, packet switched, optical interconnection network," *J. Lightwave Technol.*, vol. 27, no. 7, pp. 841-850, April 2009.
- [19] J. Luo, S. Di Lucente, J. Ramirez, H. Dorren and N. Calabretta, "Low Latency and Large Port Count Optical Packet Switch with Highly Distributed Control," *Proc. OFC/NFOEC*, pp. 1-3, 2012.
- [20] C. Kachris, I. Tomkos, "Power consumption evaluation of hybrid WDM PON networks for data centers," *Proc. NOC 2011*, pp.118-121, 20-22 July 2011.
- [21] S. Aleksic, M. Fiorani, M. Casoni, "Adaptive Hybrid Optical Switching: Performance and Energy Efficiency," (invited), *Journal of High Speed Networks*, Vol.19, No.1, pp. 85-98, 2013.
- [22] M. Fiorani, M. Casoni, S. Aleksic, "Hybrid Optical Switching for Energy-Efficiency and QoS Differentiation in Core Networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol.5, no.5, pp.484,497, May 2013.
- [23] M. Fiorani, M. Casoni, S. Aleksic, "Large Data Center Interconnects Employing Hybrid Optical Switching", *Proc. of the 18th IEEE European Conference on Network and Optical Communications (NOC)*, July 10-12, 2013, Graz, Austria.
- [24] C. Kachris, I. Tomkos, "A Survey on Optical Interconnects for Data Centers," *IEEE Communications Surveys & Tutorials*, Vol. 14, Issue 4, pp. 1021-1036, Fourth Quarter 2012.
- [25] Emulex white paper: "Connectivity Solutions for the Evolving Data Center," May, 2011.
- [26] M. Fiorani, M. Casoni, S. Aleksic, "Performance and Power Consumption Analysis of a Hybrid Optical Core Node", *OSA/IEEE Journal of Optical Communications and Networking*, Vol. 3, Issue 6, pp. 502-513, June 2011.
- [27] S. Aleksic, "Analysis of Power Consumption in Future High-Capacity Network Nodes," *IEEE/OSA Journal of Optical Communications and Networking (JOCN)*, Vol. 1, No. 3, pp. 245 - 258, 2009.
- [28] S. Kandula, S. Sengupta, A. Greenberg, P. Patel, and R. Chaiken, "The nature of data center traffic: measurements & analysis," in *Proc. 9th ACM SIGCOMM conference on Internet measurement conference*, *Proc. IMC 2009*, pp. 202-208.
- [29] T. Benson, A. Anand, A. Akella, and M. Zhang, "Understanding data center traffic characteristics," *Proc. ACM workshop on Research on enterprise networking*, 2009, pp. 65-72.
- [30] T. Benson, A. Akella, and D. A. Maltz, "Network traffic characteristics of data centers in the wild," *Proc. IMC 2010*, 2010, pp. 267-280.
- [31] Guidelines to Defra/DECCs GHG Conversion Factors for Company Reporting, *Am. Economics Assoc. (AEA)*, 2009.