Grant Agreement No. 619572

COSIGN

Combining Optics and SDN In next Generation data centre Networks

Programme: Information and Communication Technologies

Funding scheme: Collaborative Project – Large-Scale Integrating Project

Deliverable D1.2

Comparative Analysis of Optical Technologies for Intra-Data Centre Networks

Due date of deliverable: December 31, 2014
Actual submission date: January 16, 2015
Resubmission date: 3 September, 2015

Start date of project: January 1, 2014
Duration: 36 months

Lead contractor for this deliverable: UNIVBRIS

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Executive Summary

The purpose of this deliverable is to provide a review of the state-of-the-art optical data plane technologies employed in current Data Centre Networks (DCNs) or having potential to be employed in future DCNs, as well as some performance comparison analysis for these technologies. This deliverable aims to provide input to the architecture design in WP1 and data plane device/element development/selection in WP2.

In Section 2, this deliverable discusses figures of merit relevant to DCN performance, which are used in Sections 4, 6, and 7 to compare the performance of each technology. Then, different optical elements in the DCN data plane, including fibres, interfacing components, and optical transceivers, are investigated to explore their capabilities to meet the DCN requirements and their potential advantages for use in future DCNs. Finally an initial DCN architecture with optical technologies is presented as a starting point for the DCN architecture design (which will be addressed in detail in deliverable *D1.4 "Architecture Design")*, and some system level performance evaluations are also presented.

Document Information

Status and Version:	Final version 1.0	
Date of Issue:	03/09/2015	
Dissemination level:	Public	
Author(s):	**Name**	**Partner**
	Bingli Guo	UNIVBRIS
	George Saridis	UNIVBRIS
	Shuping Peng	UNIVBRIS
	Georgios Zervas	UNIVBRIS
	Reza Nejabati	UNIVBRIS
	Dimitra Simeonidou	UNIVBRIS
	Valerija Kamchevska	DTU
	Sarah Ruepp	DTU
	Michael Berger	DTU
	Adam Hughes	POLATIS
	Nick Parsons	POLATIS
	Tim Durrant	VENTURE
	Siyuan Yu	VENTURE
M12 Submission:		
Edited by:	Bingli Guo	UNIVBRIS
Checked by :	Sarah Ruepp	DTU
M20 Submission:		
Edited by:	Bingli Guo	UNIVBRIS
Reviewed by:	Katherine Barabash	IBM
	Oren Marmur	PHOTONX
Checked by:	Helene Udsen	DTU

Table of Contents

1 Introduction

COSIGN introduces novel optical networking solutions which facilitate the shift from a hardware centric DCN to a software centric one. In that respect, the technologies developed in the data plane will support the envisioned flexible, virtualized and ultra-high capacity DC networks. The shift to a software defined DCN together with the ever increasing demand for bandwidth and connectivity requires major changes to the networking infrastructure. The following breakthrough optical technological solutions will be developed within WP2 to support the vision of an integrated, automatic, and optimized DCN solution:

- Optical Switches: A flattened software defined mesh network topology based on large port count and low loss free space 3D beam steering switches for interconnecting TOR switches will be developed. The project will also explore semiconductor InP switches with very fast reconfiguration times (ns scale) as a long term solution.
- Fibres: Develop and fabricate the specialist interconnection fibres as required to realize the project subsystems and final system trials – likely requirements include Multicore fibres (MCFs), polarization maintaining MCFs and low-latency hollow-core photonic band gap fibres (HC-PBGF). MCF interfaces will be explored to enable flexible reconfiguration of the logical network connectivity on top of the installed physical network based on space division multiplexing. To tackle the subject of latency in large DCNs, HC-PBGFs will be developed to reduce propagation delays (30% reduction over conventional fibres).

This deliverable complements deliverable *D1.3 "Comparative Analysis of Control Plane Alternatives"*, providing input to COSIGN data plane and control plane technology selection/development respectively. More specifically, this deliverable provides a review of the state-of-the-art optical data plane technologies employed or having potential to be employed in DCNs, which could provide input to the architecture design in WP1 and data plane optical device/element development in WP2. Also, the technologies which will be leveraged by COSIGN will also influence choices to be made in the control plane developed by WP3.

Starting from the requirements defined in *D1.1 "Requirements for Next Generation Intra-Data Centre Networks Design"*, a definition of figures of merit relevant to DCN performance is presented, which are used in Sections 4, 6, and 7 to compare the performance of each technology. Then, different optical network elements in the DCN data plane, including optical fibres, transceivers, and some interfacing components, are investigated to show their component-level features/capabilities and gaps to meet the DCN requirements. Also, some system level performance of DCN architecture with optical technologies is analysed as a starting point for the DCN architecture design, addressed in detail in deliverable *D 1.4 "Architecture Design"*.

1.1 Reference Material

1.1.1 Reference Documents

[1]	COSIGN – Deliverable D1.1: Requirements for Next Generation intra-Data Centre Networks Design
[2]	COSIGN – Deliverable D1.3: Comparative Analysis of Control Plane Alternatives
[3]	COSIGN – Deliverable D1.4: Architecture design

1.1.2 Acronyms and Abbreviations

Most frequently used acronyms in the Deliverable are listed below. Additional acronyms can be specified and used throughout the text.

ADC	Analogue to Digital Converter
API	Application Programming Interface
AWG	Arrayed Waveguide Grating

BMR	Burst Mode Receivers
BVT	Bandwidth Variable Transceivers
CapEx	Capital expenditures
CDR	Clock and Data Recovery
CFP	C Form-factor Pluggable
CO-OFDM	Coherent Optical Orthogonal Frequency-Division Multiplexing
CO-WDM	Coherent Wavelength Division Multiplexing
CP	Control Plane
CWDM	Coarse Wavelength Division Multiplexing
DAC	Digital to Analogue Converter
DC	Data Centre
DCN	Data Centre Network
DEMUX	Demultiplexer, demultiplexing
DFB	Distributed Feedback laser
DMGD	Differential Mode Group Delay
DSP	Digital Signal Processing
EON	Elastic Optical Networking
EPS	Electrical Packet Switch
ER	Extinction Ratio
FCoE	Fibre Channel over Ethernet
FEC	Forward error correction
FMF	Few-Mode Fibre
FM-MCF	Few-Mode Multi-Core Fibre
FTL	Fast Tuneable Lasers
HA	High Availability
HC-PBGF	Hollow-Core Photonic BandGap Fibre
HPC	High Performance Computing
IC	Integrated Circuit
InP	Indium Phosphide
IFFT	Inverse Fast Fourier Transform
ISI	Intersymbol Interference
LP	Linearly Polarized
MCF	Multi-Core Fibre
MEF	Multi-Element Fibre
MEMS	Micro-Electro-Mechanical Systems
MIMO	Multiple Input – Multiple Output
MLF	Multi-Element Fibre
MMF	Multi-Mode Fibre
MPO/MTP	Multiple-Fibre Push-On/Pull-off
MTBF	Mean Time Between Failures
MTRJ	Mechanical Transfer Registered Jack
MTTR	Mean Time To Repair
MUX	Multiplexer, multiplexing
NAS	Network Attached Storage
NIC	Network Interface Card
OAM	Orbital Angular Momentum
OAWG	Optical Arbitrary Waveform Generation
OBS	Optical Burst Switch
OpEx	Operating Expenditure
OPS	Optical Packet Switch
OSNR	Optical Signal-To-Noise Ratio
OXS	Optical Cross Point Switch
PLC	Planar Waveguide Circuit
PLZT	Lead lanthanum zirconate titanate

QoS	Quality of Service
QSFP	Quad Small Form-Factor Pluggable
ROADM	Reconfigurable Optical Add/Drop Multiplexer
RRC	Root raised cosine
RTT	Round Trip Time
RWIN	Receive Window
SDE	Software Defined Environments
SDM	Spatial Division Multiplexing
SDN	Software Defined Networking
SE	Spectral Efficiency
SMF	Single-Mode Fibre
SOA	Semiconductor Optical Amplifier
SSE	Spatial Spectrum Efficiency
TCP	Transmission Control Protocol
TDM	Time-Division Multiplexing
TFB	Tapered fibre bundle
TFP	Time Frequency Packing
TIA	Transimpedance Amplifier
TMC	Tapered multicore connector
ToR	Top of the Rack
VCSEL	Vertical-cavity Surface-emitting Laser
VM	Virtual Machine
WDM	Wavelength-Division Multiplexing
WSS	Wavelength Selective Switch

1.2 Document History

Version	Date	Authors	Comment
00	01/09/2014		TOC first draft
01	01/11/2014		TOC ready and section assignment finished
02	01/12/2014		First round contribution
03	23/12/2014		Second round contribution
04	09/01/2015		Integrated version for review
05	15/08/2015		Restructured, added new Section 6
06	25/08/2015		Version ready for internal review
1.0	01/09/2015		Proofed, edited, reviewer comments addressed
1.1	11/09/2015		Minor correction + added reference

2 Figures of Merit Discussion

This section will briefly introduce the definition of figures of merit for DCNs. Also, we analyse/indicate what kind of data plane elements/devices and their features (especially the optical devices which COSIGN will investigate, i.e., fibre, fast switch, and high-radix switch) would have impact on these figures of merit.

2.1 Power Effectiveness

With the data centre's growth due to constant increase in traffic, the power consumption of intra-data centre networks becomes extremely important. Data centres are responsible for 18% of the ICT carbon footprint, or around 0.5% of the total world carbon footprint [*greenICT*]]. Therefore, it is of great importance to be able to determine how green a given data centre is, both in general and with respect to the different components that contribute to the overall power consumption. It has been demonstrated in [*energyDCN-10*] that in a typical Google data centre depending on the utilization of the servers, the fraction of power consumed by the network can vary from around 20% at full server utilization up to 50% when the servers are utilized at 15%. Thus, the network architecture and the equipment used significantly affect the overall energy efficiency of data centres. In order to be able to do comparative analysis of different network architectures, properly defined metrics are needed. For any intra-data centre network, the overall power consumption of the can be described as follows:

$$P = N_{transceivers} * P_{transceiver} + N_{switch\ ports} * P_{switch\ port}$$

where P is the total power consumed by the DCN, $N_{transceivers}$ is the total number of transceivers used, each with power consumption of $P_{transceiver}$, and $N_{switch\ ports}$ is the total number of switch ports, each contributing with power consumption of $P_{switch\ port}$. It is important to note that the power per switch port includes the control overhead, i.e., the power of control modules or fans installed in the switch chassis, shared among all the switch ports in that chassis. In networks that use electronic switches, the number of transceivers will be relatively high, since electrical switching requires optoelectronic conversion at each step. On the other hand, for all-optical architectures, this component completely disappears. With respect to the switch power consumption, optical switches consume relatively low power compared to electrical switches, but might have some limited functionalities. For different network architectures, this relation will have a different form. For a fat-tree network architecture based on Ethernet switches, the formulation reduces to:

$$P_{FatTree} = 4 * N * P_{transceiver} + 5 * N * P_{Ethernet\ switch\ port}$$

where N is the number of servers in the data centre. It is clear that every time a server is added that results with additional 4 transceivers and 5 switch ports. Typical power consumption of a 10G transceiver is 1W, while the average 10G switch port power consumption, including overhead can reach 20-30W. Going from 10G to 40G and 100G, results in almost linear increase of the power consumption of the Ethernet switch, since switching in the electrical domain is performed on each channel carrying lower data rate. A hybrid network architecture based on Ethernet switches at access level and optical switches in the remaining part of the DCN and a fully optical DCN would consume:

$$P_{hybrid} = N * P_{transceiver} + N * P_{Ethernet\ switch\ port} + N_{optical\ switch\ ports} * P_{optical\ switch\ port}$$

$$P_{optical} = N_{optical\ switch\ ports} * P_{optical\ switch\ port}$$

where N is the number of servers in the data centre, $P_{Ethernet\ switch\ port}$ is the power consumption of a single Ethernet switch port and $N_{optical\ switch\ ports}$ is the number of optical switch ports, each with power consumption of $P_{optical\ switch\ port}$. It can be seen that the hybrid architecture will be more energy efficient than the fat-tree architecture, if the power contribution of the optical switches including any control overhead is lower than $3 * N * P_{transceiver} + 4 * N * P_{Ethernet\ switch\ port}$. For comparison, optical switches usually consume around 1-10W per port including overhead, and

switching 40G or 100G WDM signals would still require the same amount of power, as in the case of single channels, unlike Ethernet switching [*opticalDCN-2013*] .

Furthermore, in fat-tree architecture where each switch has the same radix, the number of switches required for full interconnection increases radically when increasing the number of servers supported. If we assume a DCN composed of 100,000 servers, connecting them in a tier 3 fat-tree would require 6,845 switches with a radix of 72. If we ignore practical realizations and focus on the power per input port of a standard Ethernet switch [*power-eth-switch*], then assuming 12W per switch port and 1W for transceiver power consumption, this adds up to a total power consumption of 13W per port. As defined in Section 2.1, it can be calculated that this leads to a total power consumption of around 6.7 million W. For comparison, the Polatis 192×192 switch consumes only 75W, Finisar's WSS consumes 50W for different port ratios, and a standard 2×2 TDM switch consumes around 1W. If we make a general comparison, in order to consume the same amount of power a DCN would have to have deployed 53,435 optical switches of each kind. Since the Polatis switch itself is a high radix switch, by just using this type of switch in an optical DCN, only around 140 switches would be required. This demonstrates that optics can greatly contribute to reducing the power consumption in a data centre, and furthermore, by exploiting the different optical switching technologies it can enable energy efficient operation by still providing flows with different granularity and flexible bandwidth allocation for optimized resource allocation.

2.2 Network Latency

As the primary locus of data moves from disk to flash or even DRAM, the network is becoming the primary source of latency in remote data access. Network latency is an expression of how much time it takes for a packet of data to get from one point to another. Several factors contribute to network latency, including not only the time it takes for a packet to travel in the cable, but also the time the equipment/switch uses to transmit, receive, buffer, and forward the packet.

Total packet latency is the sum of all of the path latencies and of all the switch latencies encountered along the route (usually reported as RTT, Round Trip Time). A packet that travels over N links will pass through $N-1$ switches. The value of N for any given packet will vary depending on the amount of locality that can be exploited in an application's communication pattern, the topology of the network, the routing algorithm, and the size of the network. However, when it comes to typical case latency in a large-scale data centre network, path latency is a very small part of total latency. Total latency is dominated by the switch latency which includes delays due to buffering, routing algorithm complexity, arbitration, flow control, switch traversal, and the load congestion for a particular switch egress port. Note that these delays are incurred at every switch in the network, and hence these delays are multiplied by the hop count.

One of the possible ways to reduce hop count is to increase the radix of the switches. Increased switch radix also means fewer switches for a network of a given size and therefore a reduced CapEx cost. Reduced hop count and fewer switches also lead to reduced power consumption. For electrical switches, there is a fundamental trade-off due to the poor scaling of both signal pins and per pin bandwidth. For example, one could choose to utilize more pins per port which results in a lower radix, but a higher bandwidth per port. Another option is to use fewer pins per port which would increase the switch radix, but the bandwidth of each port would suffer. Photonics may lead to a better solution, namely the bandwidth advantage due to spatial/spectrum division multiplexing and the tighter signal packaging density of optics, i.e., high-radix switches are feasible without a corresponding degradation of port bandwidth.

2.3 Capacity and Throughput

In networking, throughput is defined as the average amount of traffic that can be transmitted between two given nodes. Throughput is different from capacity because capacity defines the theoretical capacity of the communication link/path, while the throughput defines the actual traffic that can be transmitted over the link/path.

Many factors impact the DCN capacity and throughput, for example server or switch interface, connectivity/topology and flow control/optimization strategy (e.g., routing algorithm, congestion control method). Some of them are listed as follows:

- Capacity of interfaces and links (e.g., NICs, cables)
- Average number of links for the path
- Switch dimension
- EPS forwarding capability
- Oversubscription ratio for different layers and topology design
- Traffic patterns

Regarding the interfaces and link capacity, the trend towards more powerful multi-core servers, server virtualization with high density of VMs per server, VM mobility and big data are accelerating the need to increase connectivity from GbE/10 GbE, to 40 GbE and even 100 GbE. Also, more than 10GbE connectivity will provide support for unified storage networking based on NAS, iSCSI, and Object Storage Systems. [*DCtraffic-13*] depicts the forecast for the server data rates inside the data centres by Intel and Broadcom, and it is estimated that by 2017, the majority of Ethernet transceivers will be based on 40G modules.

Generally, bisection bandwidth under random permutation traffic could be used as a metric to estimate topology capacity. Specifically, bisection bandwidth measures the worst-case bandwidth between two equal-size partitions of the network. This can be normalized to a value between 0 and 1 by dividing it by the total line-rate bandwidth of the servers in one partition.

To measure the throughput of a given network, average rate of successful message delivery over a communication channel (physical or logical link) or network are always employed, which means the amount of traffic that a network can accept per time unit. The throughput is usually measured in bits per second (bit/s or bps), and sometimes in data packets per second or data packets per time slot. For the latter, the throughput can be calculated as follows:

$$\text{Throughput} \leq \frac{\text{RWIN}}{\text{RTT}}$$

where RWIN is the TCP Receive Window and RTT is the round trip time for the path. The Max TCP Window size in the absence of TCP window scale option is 65,535 bytes.

We also need to consider the spectrum and cost efficiency of an optical network under physical layer impairments. Physical layer effects are incorporated in the definition of the feasible transmission configurations of the transponders, described by (*rate-reach-grid-spectrum-cost*) tuples [*transponder-ofc-2013*] . For example, we consider a bandwidth variable transponder (BVT) with two flexibility degrees, enabling: (a) the selection of the modulation format and (b) the choice of the spectrum they use (in contiguous spectrum slots). By adapting these, the flexible transponder can be tuned to transmit at a specific rate over a specific reach, using a specific amount of spectrum (in slots) and requiring a specific guard band (in the spectrum) from the spectrum-adjacent connections to exhibit acceptable transmission quality. The cost parameter is used for different types of transponders with different capabilities. Note that the rate/spectrum parameters of a tuple incorporate the choice of the t modulation format of the transmission.

2.4 Scalability

The DCN data plane should support large-scale data centres and allow the existing data centre to grow in a structured manner with simple and repeatable designs, both in number of servers, deployed workloads, supported amount of traffic, etc. Also, DCN scaling needs to be linear for performance, and with scale benefit of cost and power consumption.

Achieving this challenging goal in DCN includes the following aspects:

- Easy scaling up of architecture design with modular subsystem (e.g., rack/cluster/pod), as this allows more price/performance linearity.

- Extending or stretching the addressing of a network using open standards to embrace workload mobility across increased geographic distances and larger data centre facilities.
- Scaling the manageability and operational capabilities of the network infrastructure in the data centre network so that administrators can effectively provision and manage a larger number of systems. Additionally provide increased visibility to the network, and all of the devices that are required to deliver IT services. This includes integration with automation systems via open and extensible APIs.
- Topologies must continue to self-organize and self-heal while converging quickly in the event of link or node failure. So, technologies enabling full utilization of the available bandwidth under network failure and adapting topology change with running services are needed.

2.5 Flexibility

A desirable feature of any DCN is flexibility in resource allocation as well as the possibility to dynamically reconfigure the network equipment based on traffic demands. Establishing connections with different granularity allows for an effective use of the resources, matched to the actual requirements and increased bandwidth utilization that directly translate to lower network cost. With respect to the overall flexibility of a DCN it is necessary to consider the performance parameters of the switches used (such as insertion loss, switching speed, bit rate dependency, operational bandwidth, etc.) as well as the network requirements for the switching devices (such as port count, scalability, blocking, etc.). An Ethernet switch has packet granularity and enables good bandwidth utilization for sub-wavelength granularity demands; however it is ineffective when large chunks of data have to be transmitted, because packet parsing and processing at each switch causes increased latency. Deploying optical switches that can operate at different granularities and switch in different dimensions (space, wavelength, time, etc.) can enable better resource utilization and accommodate connections with different bandwidth requests. In order to improve the DCN flexibility, it is necessary to look into the specifications of these switches.

Connections that require only a fraction of a wavelength could share the resource in time. Therefore, a minimum requirement of an optical switch operating at sub-wavelength granularity is that the rise and fall time are in the order of a few nanoseconds, leaving a small fraction of the bandwidth unused in the switch reconfiguration period. Electro-optic LiNbO-based switches, SOA-based switches and PLZT switches have been demonstrated to operate with switching speed ranging from hundreds of picoseconds to only a few nanoseconds and could easily be used for sub-wavelength switching. Furthermore, they can operate with insertion losses of 0 dB for SOA-based switches up to 4-5 dB loss for LiNbO-based switches and PLZT switches and have size in the order of millimetres to centimetres. Connections that require a full wavelength capacity could be established by allocating the whole wavelength for a more long-term communication link, therefore the switching speed could be slower, since reconfiguration would be required less often. By proper aggregation of the traffic and grouping the wavelengths carrying traffic destined for the same end node, it is possible to further reduce power consumption and latency by switching a spectrum range, whole cores or even switching whole multi-core fibres. MEMS-based switches, piezoelectric beam-steering switches and thermal optical switches with switching speed of a few milliseconds, insertion loss of a few dB, high port count and size ranging from chip scale to one rack unit could fit these requirements.

2.6 Availability

In the networking world, a well-known figure of merit is the so-called x-nines availability metric (e.g., 5-nines), relating to 99 - 99.999% of availability. For data centres, this metric has been translated to various tiers of data centre availability. Thus, the downtime for the data centre is an obvious choice for a figure of merit. As shown in Table 1, 99.999% (5-nines) availability relates to a downtime of approx. 5 minutes per year. Tier 4 data centres are considered the most robust and less prone to fail. Tier 4 is designed to host mission critical servers and computer systems, with fully redundant subsystems (cooling, power, network links, storage, etc.) and compartmentalized security zones controlled by biometric access control methods. Naturally, the simplest is a Tier 1 data centre used by small businesses or shops.

- Tier 1 = Non-redundant capacity components (single uplink and servers).
- Tier 2 = Tier 1 + Redundant capacity components.
- Tier 3 = Tier 1 + Tier 2 + Dual-powered equipment and multiple uplinks.
- Tier 4 = Tier 1 + Tier 2 + Tier 3 + all components are fully fault-tolerant including uplinks, storage, chillers, HVAC systems, servers, etc. Everything is dual-powered.

Table 1: Data Centre Availability and Tiers

Availability	Downtime per year	Data Centre Tier
99% ("two nines")	3.65 days	Tier 1
99.9% ("three nines")	8.76 hours	Tier 2
99.99% ("four nines")	52.56 minutes	Tier 3
99.999% ("five nines")	5.26 minutes	Tier 4

From an availability point of view, a rating of data centres based on their availability metric could seem straight forward. However, from an operations point of view, there is a big difference whether a data centre experiences one failure of 5 minutes duration, or 10 failures of 30 seconds duration. As each failure requires repair times, reboots, data recovery and reassurance procedures. The actual number of failure events should be included as a figure of merit as well.

As a summary, the following are some parameters directly related with DCN availability:

- Mean Time Between Failures (MTBF)
- Mean Time To Repair (MTTR)
- Availability (calculated based on MTBF and MTTR)
- Number of Failure events
- Number of affected data centre components

2.7 Bandwidth Density

Bandwidth density is measured by the ratio of network transmission bit rate (excluding the error correction bits) per unit of transmission media (e.g., spectrum, core), which is sometimes called bandwidth efficiency or spectral efficiency. In DCNs, bandwidth density should not only take into consideration bandwidth density at port/link level, but also the port density on the switch side.

In order to identify Spectral-Spatial Efficiency (*SSE*) at optical port/link level as a metric, we propose a formula (as the following equation) expressing the aggregate Spectral Efficiency (*SE*) of the entire fibre divided by the area of its cross-section.

$$SSE = \frac{SE \cdot SM}{A_{cross}}$$

where *SE* is the Spectral Efficiency (*b*/*s*/*Hz*) of each spatial mode, *SM* the number of discrete Spatial Modes, and A_{cross} (mm^2) the area of the cross-section of the fibre. *SM* could be the number of cores in an MCF, the number of LP modes (single or dual polarization) in an FMF, the amount of elements in an MLF, the number of multiplexed modes carrying OAM in a Vortex Fibre or the number of cores multiplied by the number of LP modes in an FM-MCF.

With respect to the port density at switch level, future 40/100 Gb switches are projected to use more than 4,000 fibres per chassis where parallel optics are used. These high fibre count requirements demand high-density cable and hardware solutions that will reduce the overall footprint and simplify cable management and connections. High port densities allow for better use of rack space, power and square footage consumed by network hardware when both are in limited supply.

2.8 Network Cost Efficiency

The total number of transceivers will contribute significantly to the overall cost, and the number of transceivers could be several times higher than the number of servers, depending on the DCN architecture. A fat-tree DCN built with 24-port off-the shelf switches supporting 3,456 ports will need 17,280 transceivers. The cost of transceivers per Gbit/s in USD is shown below in Tables 2-4. The

numbers are based on price quotes taken from the public internet [*transcost-2014*] and should only be seen as very rough estimates.

Table 2: Cost of 10G transceivers for Long and Short Range

10G transceiver SFP+	LR	SR
Cost(USD)/Gb/s	$10	$2

Table 3: Cost of 40G transceivers for Long and Short Range

40G transceiver QSFP+	LR	SR
Cost(USD)/Gb/s	$50	$7

Table 4: Cost of 100G transceivers for Long and Short Range

100G transceiver QSFP+	LR	SR
Cost(USD)/Gb/s CFP	$100	$10

Cabling will contribute significantly to the cost of a DCN. A 144-fibre Trunk cable will cost in the range of USD 600. In the above data centre example (Fat-tree with 24 pods and 24*144=3456 servers) the number of required trunk cables is 24*2, so the total cost of trunk cables is ~30,000 USD.

To provide a rough estimate of transceiver cost, we assume 3456 LR transceivers and the remaining 4*3456 transceivers are SR. Thus $10*10*3456 + $2*10*4*3456 = ~600,000 USD.

Besides the cost of cables and transceivers, the cabling complexity will also have significant impact on the cost. In this regard, it is beneficial to make use of trunk cables with, e.g., 144 fibres, as the installation complexity will be almost the same as for a single fibre cable. A useful figure of merit for the cable installation cost is the amount of long cables needed to interconnect the DCN. If we return to the previous example, a standard fat-tree network will require 3,456 “long” cables between pod switches and core switches, but with a clever arrangement of pod and core switches, all the fibres from each pod could be managed by a single 144-fibre Trunk cable, thus reducing the cabling by a factor of 144. This way the employment of MCF and MEF in COSIGN will reduce the cabling complexity of DCN significantly.

3 Data Plane Requirements of Intra-DC Network

Based on the investigation in COSIGN D1.1[1], this section first reviews these requirements, then presents a deep analysis of these requirements by indicating their KPI, impact to the DCN performance, novelty, as well as the related functionality. Actually, this section intends to map these high level and generic requirements with the COSIGN data plane approach.

R-DP-01 Capacity

R-DP-01 Description	**Capacity** This requirement specifies both the aggregated and the link level DCN capacity. At link level DCN must support 10G to server today and in the near future. In COSIGN horizon, 40G to server links will be considered as well. On an aggregated level, we have to consider the amount of server ports that have to be supported in typical DCs. Few hundred thousands of servers are typical within the world's largest DCs, bringing the aggregated capacity requirement to be considered in COSIGN to millions of Gb/s.
KPIs	• data rate supported by server NIC card, switch interface and fibre/cable • switch dimension and port density • bisection bandwidth
Impact	High. The DCN capacity planning would have direct impact on amount of services that the DCN could accommodate (which reflects the cost efficiency) and their performance (e.g., implementation/response time, end-to-end delay).
Novelty	High. In previous DCN design, EPSs with different dimension and capability are structured in a hierarchical way, which causes the fundamental capacity constraints of DCN. By employing fibre with high bandwidth/spectrum efficiency and optical switch with high data forwarding/transmission capability in COSIGN, these constraints would be tackled in a cost-efficient and power-efficient way.
Functionalities	• Network resource utilization monitoring functionality
Layer	Infrastructure layer – Physical.
Related DP requirements	The data plane component and DCN topology design must be able to provide enough bandwidth to interconnect servers, as well as the connectivity provisioning scalability according to the dynamic data transmission requirements.

R-DP-02 Latency

R-DP-02 Description	**Latency** Depending on the application, very low (microsecond) latencies can be required to some types of traffic, while some other types can thrive with longer (tens of milliseconds) response times. It is therefore of the outmost importance to be able: 1) to provide the lowest possible latencies for the chosen flows and 2) to be able to distinguish the flows requiring the low-latency paths. "Always on" low latency connectivity is crucial. The reasoning for that is that from a data centre perspective, it would be valuable to always have "a little bit" of bandwidth available between the nodes, but the bandwidth should also be highly flexible so that it can be adjusted (cranked up or down) with very low latency.
KPIs	• Round Trip Time (RTT) and Jitter • Average hop count of server-to-server path • Switching/Forwarding delay
Impact	High. Latency is a critical QoS guarantee for cloud service/application
Novelty	Medium. Packets suffer from unpredictable delay caused by queuing and congestion in current EPS based DCN design, and reconfiguration time of device and network

	(e.g., path). In COSIGN, high bandwidth optical connectivity will be employed for elephant flow which would reduce the congestion to mice flows and further reduce the end-to-end packet delay for both types of flow.
Functionalities	• Per flow based service monitoring functionality
Layer	Infrastructure layer – Physical.
Related DP requirements	It is required that the DCN architecture could provide end-to-end high bandwidth connectivity with few hops (especially few electrical packet processing) for some specific service/traffic flows. Also, designing the DCN topology in a more flatten way would help to reduce the average hops of server-to-server path.

R-DP-03 Reconfigurability/Flexibility

R-DP-03 Description	**Reconfigurability/Flexibility** Traffic flow characteristics will have a significant impact on the network performance. Most flows are small <10 kB and last only a few 100 of milliseconds, requiring the network to be re-provisioned at a very high rate. Resource usage optimization is required for profitability. Resource optimization from the infrastructure owner perspective can come in conflict with the optimization goals of the deployed services. For example, workload optimizers tend to increase the amount of instances when the service experiences a peak in demand; for that it might be required to power on standby servers. Taking into account the wear and tear of frequent power on and power down operations is typically not part of the consideration of the workload manager, although it can be of outmost importance to the infrastructure operator.
KPIs	• Path (e.g., bandwidth) provision flexibility and time efficiency • Switch reconfiguration time • Service/traffic blocking rate
Impact	Medium. The reconfigurability and flexibility of DCN would significantly influence its service/traffic accommodation capability, since the traffic in data centre (between ToRs and servers) varies quickly in time and space domain.
Novelty	Medium. Through employing large scale optical switch, fast optical switch and spatial division multiplexing (SDM) based fibres (e.g., MEF, MCF), the dimension of network resource and end-to-end path provision will be increased significantly (e.g., hybrid SDM, Time-division multiplexing (TDM connection).
Functionalities	• Underlay (physical layer) resource utilization/performance monitoring • Flexible end-to-end path provision functionality
Layer	Infrastructure layer – Physical. Infrastructure level – CP layer.
Related DP requirements	The infrastructure should provide support for flexible resource allocation to satisfy the demands of different traffic flows and optimize resource utilization.

R-DP-04 Resiliency and HA

R-DP-04 Description	**Resiliency and HA** DCN data plane need provide high service availability and minimize the amount of systemic downtime events, i.e., it is required to have *n* fully redundant network paths between each pair of endpoints.
KPIs	• Amount of disjoint path between end points (ToRs or servers) • Network devices failure (e.g., link or switch port) awareness time • Network devices switch-over/reconfiguration time

Impact	High. An interruption of just seconds to normal data access can result in enormous cost to the business and if lengthy, may impact it to such a degree that it cannot recover. By building resilience into the DCN infrastructure, the risk of service interruption could be reduced.
Novelty	Medium. In DCN structure design, the infrastructure resiliency should be considered properly to recover from hardware layer failure. And a lot of previous research on survivable optical network and DCN design could be refereed.
Functionalities	• Failure recovery functionality
Layer	Infrastructure layer – Physical.
Related DP requirements	It is required to have fully redundant network paths between each pair of endpoints (e.g., servers, ToRs) and this is more related with DCN topology design.

R-DP-05 Traffic isolation

R-DP-05 Description	**Traffic isolation** DCN data plane should be capable of isolating the traffic on the prescribed granularity – workload owner, application, application transaction, application tier, etc. In addition to the physical isolation, the management and the performance isolation must be provided. Isolation here read more like a logical request (virtual) then a physical request. But there is also a requirement for physical isolation of parts of the data plane in the DC. For instance, physical isolation can be required in some cases (e.g., where we do not use overlays). This can be ensured by fully isolated paths, by using multiple cores in fibres, different wavelengths, etc.
KPIs	• Dimension of network resources (spatial, time, spectrum) • Granularity of resource allocation
Impact	High. Physical network isolation (e.g., path isolation) could provide a more trustable traffic isolation to support multi-tenant cloud service.
Novelty	High. SDM technology and optical switches with different features could provide different scale physical network isolation. Comparing with overlay based traffic isolation, it could provide more privacy, security, and robustness to isolation failure to client.
Functionalities	• Physical network isolation functionality • Combined overlay-based and underlay(physical)-based network virtualization
Layer	Infrastructure level – CP layer Infrastructure level – Physical layer
Related DP requirements	DCN supported switching dimension would enhance the physical isolation capability.

R-DP-06 Scalability and Extensibility

R-DP-06 Description	**Scalability and Extensibility** DCN data plane should support large-scale data centres and allow the existing data centre to grow organically, both in number of servers, in number of the deployed workloads, supported amount of traffic, etc.
KPIs	• Performance/cost linearity to involve more servers/switches • Topology/structure stability for involve more servers/switches
Impact	High. The DCN scalability implicates the DCN performance stability (e.g., latency) when extension happens, as well as the cost efficiency.
Novelty	High. The DCN design with optical switches should be investigated to better utilize

	the scalability of optical switch dimension (e.g., spatial, spectrum) and the capability of large scale optical switch.
Functionalities	• DCN structure extension functionality
Layer	Infrastructure layer – Physical.
Related DP requirements	The scalable structure design should allow an easy way to extend with modular subsystem, as well guarantee its flexibility.

4 Optical Fibres

Optical fibre is becoming a dominant transmission media for modern data centres. The vast number of interconnections for scale-out networks drives the need for compact cabling solution. At 10G, rack-to-rack communication in the data centre and high-performance computing environments have traditionally been the realm of VCSEL-based transmitters, and multi-mode fibre (MMF) primarily due to their low transceiver cost.

However, with the rising cost, bandwidth and reach limitation (approximately 10 Gb/s, several hundreds of metres) of these MMF-based interconnections, moving to single-mode fibre (SMF)-based interconnections for even the shorter, rack-to-rack distances provides significant benefits. Due to its simple structure and its prevalence for decades in the telecommunications industry, SMF is a low-cost, commodity technology. A single strand of fibre can support tens (to hundreds) of terabits per second of bandwidth. These high bandwidths per SMF are obtained not by a single transmitter–receiver pair, but by a number of pairs, each operating on a separate wavelength of light contained in the same fibre through WDM or SDM.

As a result of these characteristics, SMF-based interconnects provide a number of advantages over MMF-based interconnects within the data centre, contrary to the conventional viewpoint. There is a large saving in cable cost and volume through multiple generations of networking fabric when the bandwidth scales from 10GE, 40GE/100GE to 400GE. Thus, there is both a CapEx and OpEx advantage. The fibre is installed once for a particular interconnect speed. Subsequent increases in speed only require adding wavelength channels, with the same fibre infrastructure remaining in place. Fibre thus becomes a static part of the facility and requires only a one-time installation, similar to the electrical power distribution network. Considering the large number of fibres and time and cost to install them, this represents a huge cost saving. In addition, scalability in interconnect bandwidth is greatly enhanced as wavelengths in the same fibre are increased for higher speeds, and not the number of parallel fibres, as would be required in an MMF interconnection. The maximum reach of the interconnection is also significantly increased, along with reduction of count and patch panel space.

4.1 Single-Mode Fibre

Compared to multi-mode fibre, Standard single-mode fibre (SSMF) has a core diameter of typically 8-10 μm causing tighter requirements on mechanical alignment to optical sources and other components. However, SSMF has longer transmission reach than multi-mode fibre. This is particularly pronounced at higher data symbol rates. Single-mode fibre is mainly limited by chromatic dispersion whereas multi-mode fibre is mainly limited in transmission reach by mode dispersion. In the limit of maintaining 95% of the pulse energy within the timeslot given as 1/*B*, where *B* is the bit rate (or symbol rate for higher order modulation formats) the transmission reach for Gaussian pulses as function of *B* is given by: $L = \left(\frac{1}{4B}\right)^2 \frac{1}{|\beta_2|}$, where $\beta_2 \approx -20\ ps^2/km$ for standard single-mode fibre [*fmf-2012*] [*FOCS-10*] .

SSMF has a propagation loss of ~0.2 dB/km at 1.5-1.6 μm and is the technology of choice for most telecom systems which means that a great host of developed and matured components for telecom exist which is compatible with SSMF. This includes Distributed Feedback laser (DFB) lasers, Arrayed Waveguide Gratings (AWGs), high speed modulators, etc.

4.2 Multi-Mode/Few-Mode Fibre

Multi-mode fibres are optical fibres which support multiple transverse guided modes for a given optical frequency and polarization. Such fibres support tens of transverse guided modes for a given optical frequency and polarization (LP modes as in Figure 1), and the number of guided modes is determined by the wavelength and the refractive index profile. Particularly for fibres with a relatively large core, the number of supported modes can be very high. Launching light into a multi-mode fibre

is comparatively easy, because there are larger tolerances concerning the location and propagation angle of incident light, compared with a single-mode fibre. So, such fibres can guide light with poor beam quality (e.g., generated with a high-power diode bar), but for preserving the beam quality of a light source with higher brightness it can be better to use a fibre with smaller core and moderate numerical aperture, even though efficient launching can then be more difficult.

Also, the possible data rates and/or transmission distances achievable with such fibres are limited by the phenomenon of intermodal dispersion: the group velocity depends on the propagation mode (especially when having many modes co-propagating), the modal dispersion, modal interference and high DMGD, making long-haul transmission simply impossible. The only way to deal with that is to compensate those impairments through heavy MIMO DSP on the receiver side. In order to improve this situation, FMF has been proposed *[fmf-2012]* . FMFs are in principle the same as MMFs, but are made to carry fewer LP modes, thus lightening DSP load at the receiver end and making long-distance communication achievable [*[*fm-mdm-2012*]* [*mdm-16qam-2012]*]. However, unavoidable imperfections still set limits. There are ISO standards like OM1, OM2 and OM3, which quantify the residual level of intermodal dispersion, limiting the transmission bandwidth (or the bandwidth–distance product). The highest performance is achieved with OM3 50/125 μm laser-optimized fibres, having a very precisely controlled refractive index profile.

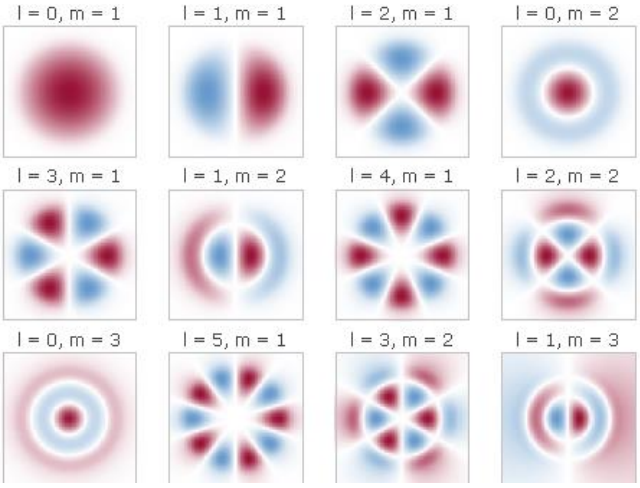


Figure 1: Some of the fundamental LP modes used in Mode-Division Multiplexing in MMF, FMF and FM-MCF

4.3 Multi-Core Fibre

As their name implies, multi-core fibres promise to significantly increase the bandwidth capacity of fibres by providing more light-carrying cores than the single core typical of conventional fibre. Multi-core fibres can be used to dramatically reduce the amount of space required and increase the bandwidth of the fibre-optic cable used in DCNs. Although MCFs are gaining increasing popularity lately, the notion of having multiple single-mode cores placed in a sole fibre structure is not that new. The first MCF was manufactured back in 1979 [*mlt-core79*]. However, the demand back then was not tremendous and the optical community not so mature to adopt it. Currently, MCF seems to be one of the most popular, accessible, in terms of designing and manufacturing *[crosst-mcf-2011*], and efficient ways to realize SDM. Many core arrangement styles for the inside of the cross-section of the fibre have been proposed (Figure 2); the One-ring [*ring-mcf-2012*] and the Dual-ring structure, the Linear Array [*line-mcf-2012*], the Two-pitch structure and finally the Hexagonal close-packed structure which is also the most prominent. Examples of this style, with 7 cores and 19 cores [*19-mcf-2013*] have already been proposed and used in experiments. Diameters of different MCFs vary between 150 and 400 μm, depending mostly in core pitch values. Multi-core fibres can deliver exceptionally high bandwidth, capacity up to Pb/s, supporting at the same time spatial super-channels (i.e., groups of same-wavelength subchannels transmitted on separate spatial modes but routed together) and the ability for switching also in the space dimension, other than time and frequency. For example, 3 cores of a MCF could be switched together at first creating a super-channel and then in the next network node, one data-stream propagating in one of those cores could be dropped or switched to another core, etc. In a real network environment, such a strategy could provide sufficient granularity for efficient

routing and facilitate ROADM integration, and could help simplify network design since the modes are routed as one entity, foster transceiver integration (e.g., share a single source laser in the transmitter and a single local oscillator in the receiver), and lighten the DSP load by exploiting information about common-mode impairments such as dispersion and phase fluctuations. In addition, MCFs have more or less the same attenuation values as common SMFs, so no extra amplification would be needed when replacing the old infrastructure, which is quite crucial from a network design point of view.

The main and most important constraint to be tackled in MCF design and furthermore in MCF networks is the inter-core crosstalk, i.e., the amount of optical signal power "leaking" from adjacent cores to a specific one, causing interference with the signal already propagating there. There are a lot of studies ongoing on how to minimize crosstalk in a MCF structure [*lea-mcf-2011*], [*mcf-design-2011*], showing that crosstalk can be successfully counteracted by using trench-assisted cores (Step-Index), by utilizing the fibre bend and by keeping the fibre cores well-spaced. Other solutions proposed are using cores with different refractive indexes, resulting in a heterogeneous MCF [*h-mcf-2013*], or even assigning bi-directional optical signals in adjacent cores to avoid long co-propagation in the same direction in order to reduce interference.

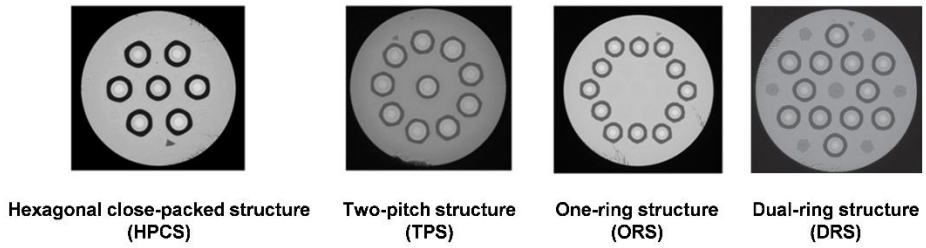


Figure 2: Different MCF core arrangement designs

4.4 Few-Mode Multi-Core Fibre

The combination of a multi-core fibre and a few-mode fibre, known as Few-Mode Multi-Core Fibre (FM-MCF) is becoming a very attractive SDM approach lately, since it incorporates benefits from both MCFs and MMFs without adopting all of their drawbacks. FM-MCF has increased capacity by a factor of = (number of cores) × (number of modes). According to [*fm-mcf-2014*], more than 300 FM-MCF channels can theoretically be supported. However, the prime advantage of FM-MCF compared to MMF-FMF is the significantly lighter MIMO DSP requirement in the receiver side. As shown in Figure 3, a MMF carrying 6 LP modes requires quite heavy DSP for the MIMO matrix, while in the case of having 3 cores each carrying only 2 modes, the matrix is much simpler and the DSP needed less. As in MCFs, FM-MCFs' most critical aspect to deal with, is inter-core crosstalk between the fundamental LP01 mode and higher order modes, such as LP11, LP21, etc. In a nutshell, FM-MCF is a very promising fibre technology for future SDM networks to deliver high capacity and scalability, provided that efficient outbound (Tx)/inbound (Rx) data transmission rates and mux/demux would be available in the next few years.

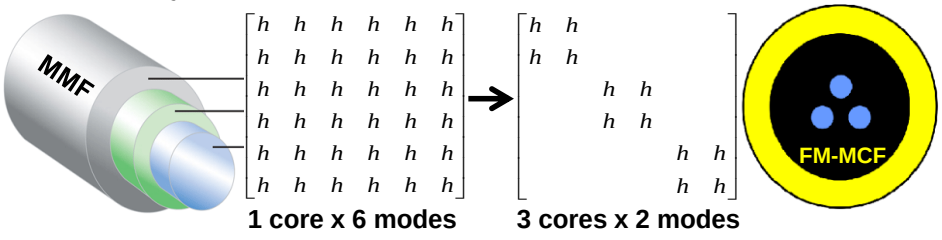


Figure 3: DSP complexity tables for a 6-mode MMF and a 3-core 2-mode FM-MCF..

4.5 Multi-Element Fibre

Another fresh approach for SDM implementation is using a Multi-Element Fibre, in which multiple fibre elements have been drawn and coated close together in a single coating [*mef-sdm-oe-2014*][*mef-*

ecoc-2013] as shown in Figure 4. Almost zero crosstalk has been detected between the spatial channels. However, the greatest advantage of MEF over MCF and MMF is that the fibre elements can easily be separated from the main structure, and coupled using conventional SMF connectors to any device of the existing infrastructure, avoiding the use of SDM mux/demux. That would keep the overall cost and power budget of the network low. Compared to MCF-MMF, the drawback of MEF is that it delivers fewer spatial channels than other fibre approaches, relative to its dimensions.

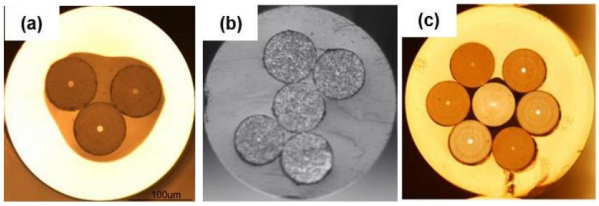


Figure 4: *(a) Cross-section of a 3-element MEF (b) Cross-section of a 5-element MEF (c) Cross-section of a 7-element Er-doped MEF used for core-pumped amplification [mef-ecoc-2013]*

4.6 Vortex Fibre Carrying Orbital Angular Momentum (OAM)

An upcoming technology that could contribute in the new era of SDM is the so called Vortex Fibre for OAM multiplexing [*oam-ofc-2014*]. Optical vortices are light beams made of photons that carry orbital angular momentum (OAM). In quantum theory, individual photons may have the following values of OAM:

$$Lz = l \ h \quad (1)$$

In Eq. (1), *l* is the topological charge and *h* is Planck's constant. The theoretically unlimited values of l (±16, ±14, ±12, ±10, ±8, ±4 …), in principle, provide an infinite range of possibly achievable and multiplex-able OAM states (see Figure 5). These OAM modes can be multiplexed in single wavelengths and then be multiplexed in the frequency domain (WDM) as well. Vortex fibres for OAM multiplexing are one of the most promising SDM solutions for the future networks, as it can potentially scale with reduced crosstalk compared to FMF between discrete modes.

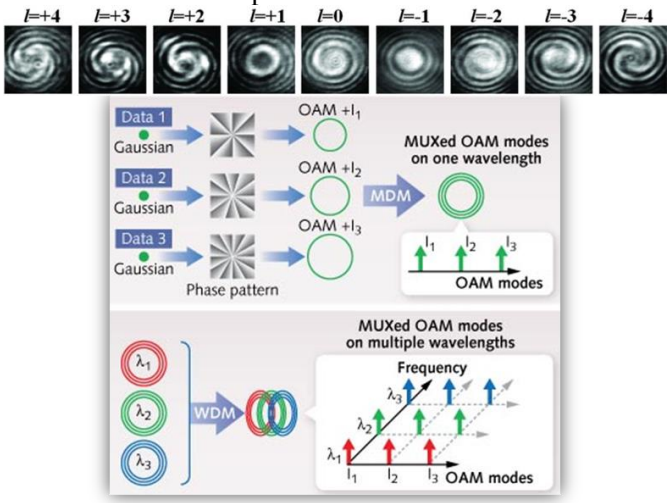


Figure 5: Multiplexing of OAM modes (SDM) in single wavelengths and then multiplexing in frequency domain (WDM) [oam-np-2012],[oam-mdm-ecoc-2012]].[oam-np-2012][oam-mdm-ecoc-2012].

4.7 Hollow-Core Photonic Band Gap Fibre (HC-PBGF)

Instead of a solid core, HCFs are hollow inside, as seen in Figure 6, containing air, and wave-guiding is achieved via a photonic bandgap mechanism. Initially, HCFs were not intended to be used for SDM, but as a substitute for SMF. As nearly 90% of the light propagates through air, HCF offers ultra-low latency (almost 30% reduction compared to SMF), enormous decrease in non-linear effects, and potential for extra-low loss, while at the same time supporting several LP modes, the number of which depends on the fibres dimensions and design [*mdm-hcpgf-jlt-2014*]. Finally, HCFs are theoretically found to have less loss around the 2 μm area, opening a new frequency band for transmission. All in

all, HCF seems to be the perfect candidate to combine SDM Mode-Multiplexing and low latency for future high-capacity latency-sensitive networks, i.e., HPC networks and high frequency trading applications.

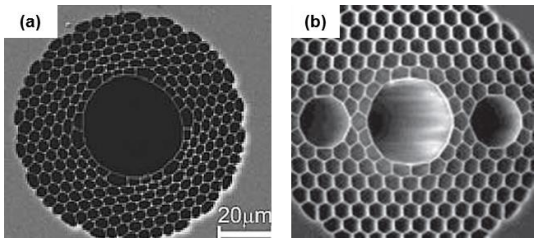


Figure 6: (a) Cross-section of a single-core 37-cell HC-PBGF able to carry three LP modes (b) Cross-section of a Tri-core HCF for low latency single-mode transmission [mdm-hcpgf-jlt-2014]

4.8 Comparison Analysis

Based on the descriptions above, the features of different fibres are compared in Table 5, as well as their preference in different DCN positions. However, it should be noted that it is hard to say which technology would have major advantage over others when considering all these features and various requirements in different scenarios.

First, it is clear that FMF/MMF, Vortex Fibre, HC-PBGF and MCF/MEF could provide better spectrum efficiency than others with negligible performance degradation. But fibres that support mode division would need more complex signal processing technologies (e.g., MIMO), which would increase the implementation difficulty and cost. On the other hand, MCF and MEF are more suitable for interconnection between aggregate/core switches due to their higher spectrum efficiency, while SMF and FMF/MMF are more suitable for interconnection closing on the server side (server-to-server and server-to-ToR) considering the cost efficiency (e.g., interfacing cost). Of course, with the technology advancement and increasing communication requirements, the cost of MCF/MEF would go down and an optical switch may directly interface with MCF, which would make the SDM-based fibres more popular. Also, as discussed above, with HC-PBGF, lower end-to-end latency could be achieved.

Table 5: Performance Comparison of different fibre technologies

	Technology						
	SMF	FMF/MMF	MCF	MEF	FM-MCF	Vortex Fibre	HC-PBGF
Figures of Merit							
Spectrum efficiency	low	medium	high	high	high	high	high
Fibre Loss	standard	Can be low as SMF	Can be low as SMF	Can be low as SMF	Higher than SMF	Higher than SMF	Higher than SMF
Intra-Mode Nonlinearity	no	low	standard or high	standard	high	high	standa rd
Inter-Mode Nonlinearity	no	low to medium	low	no	medium	medium	low to mediu m
Mode Coupling/	no	low to high, can be	medium	no	high	high	mediu m

Crosstalk		optimized					
Cost	low	as low as 1×SMF	medium	N×SMF	medium	high	mediu m
DSP complexity	low	medium to high	low to medium	low	high	high	high
Preference of DCN position							
Server-to-Server (Intra-Rack)	high	high	low	low	low	low	mediu m
Server-to-ToR	high	high	medium	low	low	low	mediu m
ToR-to-Aggregate Switch	low	low	high	medium	medium	medium	mediu m
Inter-cluster	low	low	high	high	high	high	high
Inter-DC	low	low	high	high	medium	high	high

The discussion below will further explore the benefits in terms of spectrum efficiency that may be achieved by leveraging SDM based fibre technology.

So far, SDM theoretical and experimental research was based on total bandwidth, capacity and aggregate spectral efficiency, not actually considering the space domain. Needless to say, it is essential to have a reference point in order to analyse and identify which SDM technology suits one's purpose better. In this section we aim to measure SE per cross-sectional area of the fibres (i.e., Spectral-Spatial Efficiency) [*SDM-survey-2015*]

As defined in Section 2, we calculated the Spectral-Spatial Efficiency for 10 fibre structures, each of which is implementing SDM in various ways. These fibres were reviewed qualitatively in this section. Figure 7 illustrates the SSE of those SDM fibre technologies for three discrete SE values, 1, 4 and 8 b/s/Hz. This figure depicts the expected trend that fibres with more spatial channels and smaller cross-section use the spectrum much more efficiently. Remarkable distinction is found between the SMF ribbon and the MC-FMF (5 and 5928 b/s/Hz/mm^2. That is due to the large difference between the sizes of the cross-section of the fibre of the two technologies, although SMF-bundle outnumbers FM-MCF in Spatial Modes. In the cases of the fibre bundle and MEF, coating diameters have been used, since the fibres and the elements do not have the same cladding as the rest of the instances, so taking their cladding diameter as a reference would be inaccurate. Adding coating diameters to the rest of the fibres would have made an insignificant difference to the final result, so we decided not to. The outcome of the above evaluation of SSE shows that there is enough room for future improvement in SDM networks. Especially in reducing cable complexity and conventional fibre mesh by having fewer SDM fibres still offering the same and higher spectral efficiency and capacity services.

As a summary, the SDM technologies are mature enough to successfully cope with most of those requirements. First of all, there is a wide variety of SDM fibres to serve the network capacity demands, and depending on those demands, the SSE figure of merit could be used to determine which fibre is more suitable. Not to mention, to pass from an SMF-based infrastructure to the SDM era, spatial multiplexers and de-multiplexers would be needed as well. In addition, a lot of SDM amplifiers have been developed and proposed for MCF, FMF and other uses, to meet the amplification challenges in a SDM metro/core network. Of course, there is progress still to be made in order to end up with a reliable solution. When it comes to reconfigurability, SDM ROADM experimental prototypes have been already presented that offer scalability and successfully deal with a high number of WDM and SDM channels. ROADMs are often the most crucial elements of metro/core networks, thus it is an absolute necessity to have some solid SDM implementations to rely on in the future. Finally, resilience and failure recovery functions can be supported by an SDM network, since there are a lot of spatial channels in parallel. If for some reason one channel fails, the adjacent channel can replace it instantly.

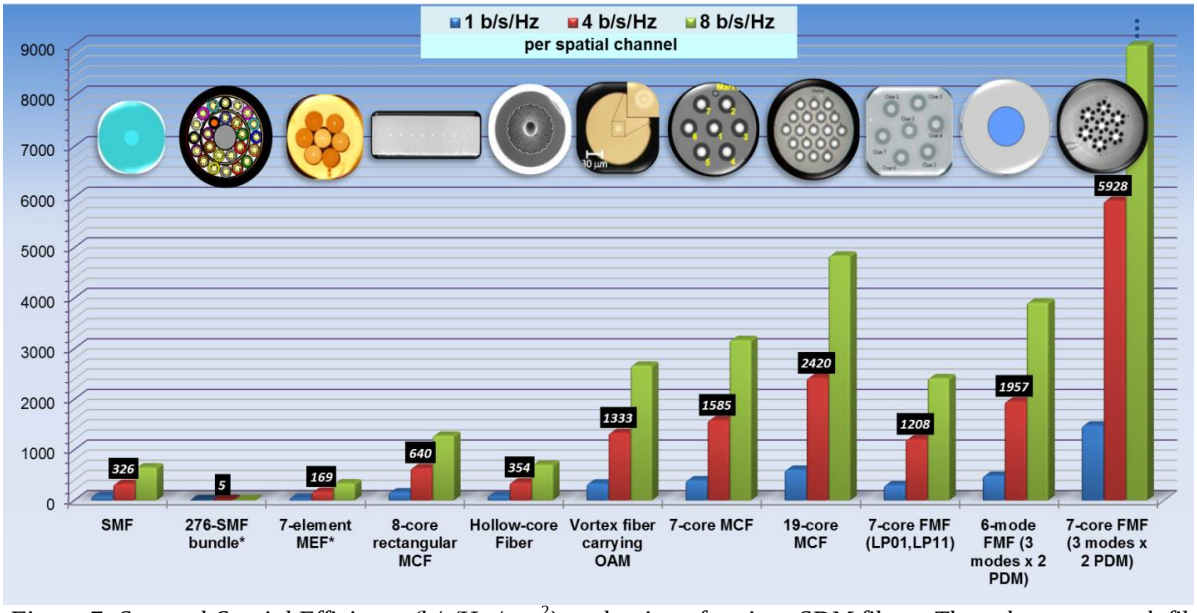


Figure 7: Spectral-Spatial Efficiency (b/s/Hz /mm^2) evaluation of various SDM fibres. The columns on each fibre represent 1, 4, 8 b/s/Hz Spectral Efficiency per spatial channel respectively. Indication labels (middle column) show the value of SSE for 4 b/s/Hz SE

5 Interfaces/Connectors

Interfaces supported by switches and connectors used to adapt the switch interface to the specific fibres (e.g., multi-core fibre) are directly related with the DCN design and its performance (e.g., cost efficiency, port density of switches). In this section, first the potential interfacing technologies utilized by COSIGN optical switch providers, e.g., Venture and Polatis, will be investigated. Also, one type of spatial multiplexer to interconnect MCF with SMF is reviewed, which may need to be employed to adapt the interface of COSIGN optical switches and MCF. At last, optical interfaces which have potential to be used in COSIGN optical switches are summarized, as well as their key features, e.g., loss, cost efficiency.

5.1 OXS Connector/Interface

With the decision to use ribbon single-mode fibre, deciding which optical connector to use is straightforward. The MTRJ only supports 2 or 4 fibres and has been made effectively obsolete by the MTP connector family. The MTP connector is currently available in 4, 8, 12, 24, and 72 fibre densities for multi-mode fibre (50 μm and 62.5 μm cores) and 4, 8, 12, and 24 fibre densities for single-mode fibre. The MTP Elite (low-loss) single-mode connector is available in both 8 and 12 fibre densities. Figure 8 presents the major features of various MTP connectors. For high performance characterization the angled variant has been selected..

	MM MT Elite® Multimode MT Ferrule	Standard Multimode MT Ferrule	SM MT Elite® Single mode MT Ferrule	Standard Single mode MT Ferrule
Insertion Loss	0.1dB Typical 0.35dB Maximum [2,3,5]	0.20dB Typical 0.60dB Maximum [2,3,5]	0.10dB Typical 0.35dB Maximum [1,4,5]	0.25dB Typical (All Fibers) 0.75dB Maximum (Single Fiber)[1,5]
Optical Return Loss	> 20dB [5]	> 20dB [5]	> 60dB (8° Angle Polish) [5]	> 60dB (8° Angle Polish) [5]

[1] As tested per ANSI/EIA-455-171 Method D3
[2] As tested per ANSI/EIA-455-171 Method D1
[3] As tested with proposed encircled flux launch condition on 50um fiber and 850nm per IEC 61280-4-1
[4] Compliant with proposed IEC 61755-3-31/GRADE B
[5] For 48-fiber MM MTs, 72-fiber MM MTs, or 24-fiber SM MTs, please see our website at www.usconec.com/resources/faq.htm#ques6.

Figure 8: Feature of MTP Optical Connector

These MTPs are fully compliant with IEC Standard 61754-7 and TIA 604-5 – Type MPO. Therefore the package concept is easily upgraded as high switch count devices are developed. As shown in Figure 9, a 12 fibre MTP connector takes up less space on the front panel than the six duplex SC connectors required for the same number of fibres.

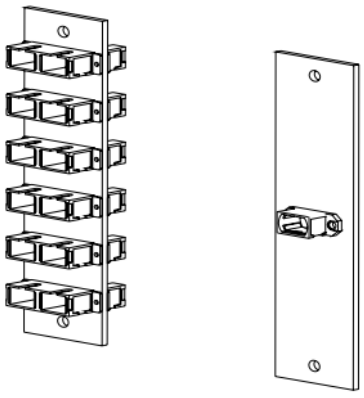


Figure 9: 12 Fibre SC Panel Density Compared to 12 Fibre MTP Panel Density

The connector and jumper fibre assemblies (making connection from the chip to the bulk head easy) are available from USConec in partnership with NTT [*AEN-1909 Rev 2.0*] .

Electrical

The OXS is a new product in a standard CFP2 case [*CFP*] . A new pin-out will be used, but the intention is to keep the ground leads the same as the standard using a 104 pin connector.

The optical connection from the chip to the outside world is by fibre. Industry requires the device to be demountable by connector. An optical jumper lead will connect the chip interface with a bulk head in the package wall. Fibres can be ribbon or multi-core with linear or hexagonal array core topography:

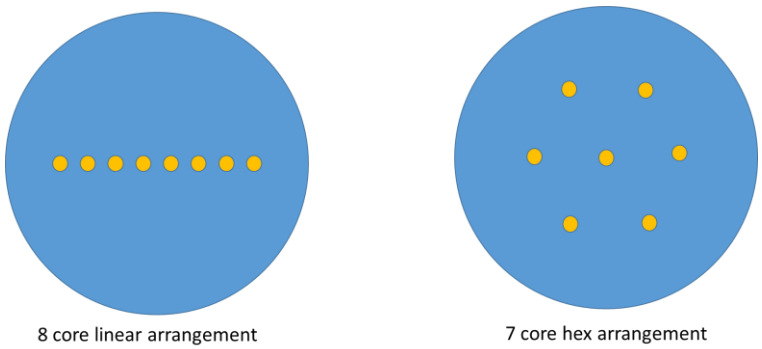


Figure 10: Multi-core Connector Arrangement

At this stage in device development the fundamental chip/fibre interface can be designed to suit either ribbon or linear multi-core fibre technology by suitable design of the light guides on the chip in terms of both dimension and spacing.

One option is leveraging a hexagonal core array layout which can be envisioned as a two-dimensional interface. This would however add cost, complexity and risk to the development, pushing the activity outside the bounds of COSIGN. The focus is on having a 4×4 switch technology demonstrated within the COSIGN project. Therefore light paths would ideally be multiples of 4, the resulting design being a linear 8-core structure. Single fibres with multiple single-mode cores are not yet available, and discussions related to rack topography suggest that there is little advantage in developing these in the short term.

Thus the design will focus on a slightly larger footprint of single-mode ribbon fibre. The ribbon is standard on a pitch of 250 µm centres. Whilst 4- and 8-fibre ribbons are available, the general standard is now 12 fibres as shown in Figure 11. This has been adopted, thus having 4 fibres 'spare'.

Figure 11: Single-Mode Ribbon Fibre Design

5.2 Beam Steering Switch Connector/Interface

The Polatis DirectLight optical circuit switch uses standard (SMF28e compatible) single-mode fibre internally and therefore can present the optical ports to the user in a wide variety of standard connector formats.

For the data centre user, the most popular are LC/UPC duplex connectors for low cost, loss and repeatability, together with compatibility with SFP+ transceivers. A recent addition is the LC/HD (high density) variant which uses pull tabs on the connector to allow minimal spacing between bulkhead adaptors. However, the connector format still requires significant rack space compared with multi-fibre formats.

Developments in multi-fibre connectors (MPO/MTP) now offer potential for lower loss mass connection – down to 0.2 dB for the MTP Super Elite - but at a premium.

For the nominally 400×400 port optical circuit switch being developed by Polatis in the COSIGN project, the front panel height in a standard 19" rack-mount chassis would be at least 12 RU with LC/UPC connectors, 8RU with LC-HD connectors and potentially less than 4RU with MTP-12 connectors.

5.3 Spatial Multiplexer

Connecting between SMF and MCF is made simple with the development of the MCF fanout products.

In [*seven-core-oe-10*] , fibre-based TMC was utilized to couple signals into and out of the MCF. Because several cores are densely packed in a small region, the connectivity of each individual core between MCF becomes very difficult. For example, splicing of the MCF requires careful precision alignment. Practical use of MCF requires coupling of signals into and out of each core independently. To overcome this problem a new TMC is designed and fabricated, the structure of which is illustrated in Figure 11. The TMC preserves individual cores at both ends of the connection, and seven single-core fibres are tapered together to match the MCF spacing. One end of the resulting taper can then be connected to the 7-core MCF via fusion splicing, while the other end consists of seven individual single-core fibres. It should be noted that the TMC is different from a tapered fibre bundle (TFB) which is often used for coupling pump light sources into cladding-pumped fibre lasers and amplifiers. As shown in Figure 12, the TMC keeps the individual cores at both ends of the connection, and so prevents crosstalk or optical power coupling between cores. In contrast, a TFB merges light from multiple cores into a single core.

To achieve low crosstalk and low insertion loss in TMC, a special single-mode fibre is used. Experimental results for insertion loss and crosstalk of two seven-core TMCs are shown in Table 6 [*seven-core-oe-10*] . The insertion loss for each connector of the TMCs ranges from 0.38 dB to 1.8 dB, and the crosstalk between cores is less than −38 dB. The slightly high coupling loss of some cores can be readily reduced by improving core matching.

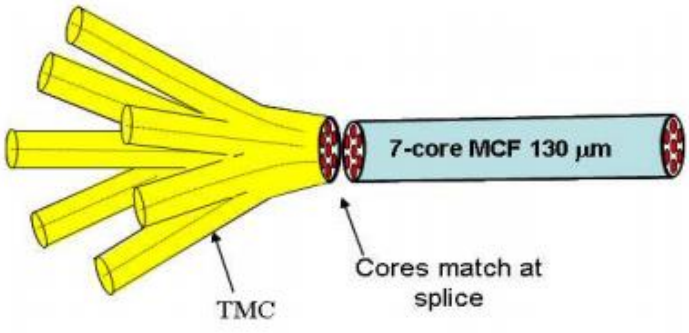


Figure 12: Schematic diagram of tapered multi-core fibre connectors [seven-core-oe-10]

Table 6: Insertion Loss and Crosstalk of TMCs [seven-core-oe-10]

	TMC #1		TMC #2	
Core #	Loss (dB)	Crosstalk (dB)	Loss (dB)	Crosstalk (dB)
Center core	1.6		0.38	
Outer core 1	1.8	–37.9	1.6	–40.8
Outer core 2	1.1	–39.0	0.9	–39.3
Outer core 3	1.4	–41.1	1.2	–43.8
Outer core 4	1.8	–39.6	1.0	–41.8
Outer core 5	1.3	–42.1	1.3	–41.8
Outer core 6	1.4	–38.7	0.9	–43.8
Average	1.48	–39.8	1.17	–41.8

5.4 Summary

Optical interfaces and connectors play a quite critical and essential role in interconnecting fibre and switch interfaces. To determine the interfacing technologies to be used for optical switches in DCNs, we need to consider not just the current fabricating technology constraints, but also the cost efficiency and their scalability requirements. Also, switches with different features may fit in different network positions where interconnection requirements vary. In this case, one specific switch must facilitate more than one interfacing technology to satisfy different networking requirements. Table 7

summarizes the potential interfaces which will be implemented in the large-scale core switch (by Polatis) and the optical cross-connect-based fast optical switch (by Venture) in the COSIGN project.

Regarding the figures of merit, insertion loss will limit the signal reach, while scalability mainly concerns the switch port density and capacity. As shown in Table 7, the beam steering switch from Polatis could support LC/UPC and MTP which are suitable for inter-rack and inter-cluster respectively considering their capacity and scalability features. For the optical cross-point switch (OXS), its *ns* level switching capability and dimensions make it perfect for intra-rack and inter-rack communication.

Table 7: Interfaces will be supported by Polatis and Venture switch

	Beam Steering Switch		**OXS**
Supported Connector/Interface	LC/UPC	MTP Elite	SMF
Figures of Merit			
Loss (db)	0.1	0.35	Low /zero
Scalability	Low	High	8×8
Cost efficiency	High	Medium	Not known yet

6 Optical Transceiver

An optical transceiver is a computer chip that uses fibre optic technology to communicate with other devices. This is opposed to a chip that transfers information electrically through metal wires and circuits or by the process of using various wave forms to communicate data. An optical transceiver chip is an Integrated Circuit (IC) that transmits and receives data using optical fibre rather than electrical wire. Optical transceivers are typically used to create high bandwidth links between network switches.

Many different kinds of transceivers are available for use in DCN and telecommunication applications. The different specs and designs are widely used to meet the changing requirements of services. Some widely deployed transceivers are available for these particular purposes, and in this section the existing optical transceiver technologies are reviewed.

Existing optical transceivers in data centres are usually offered in pluggable or board mounted packaging. According to the IEEE Standard for Ethernet, different optical interfaces for Short Reach (SR), Long Reach (LR) and Extended Reach (ER) have been defined for different data rates. Detailed specifications of these optical interfaces can be found in [*Ethernet-ieee-2012*] . In general, two types of optical technologies have been deployed, namely VCSEL based multi-mode optical transceivers and DFB based single-mode optical transceivers.

The common data rates at which these optical modules operate vary from 1 Gb/s up to 100 Gb/s. Current demand has not been enough to justify the cost and complexity of higher order advanced modulation formats, therefore until now only simple amplitude modulation has been used. Capacity scaling has been achieved by increasing the lane rate from 1 Gb/s to 10 Gb/s and 25 Gb/s, as well as by using parallel or WDM (Wavelength Division Multiplexing) channels for obtaining aggregate rates of 40 Gb/s and 100 Gb/s. All the links in data centres are point-to-point, thus bidirectional communication is attained over duplex MMF or SMF links.

Commodity VCSELs are cheap, easily integrated, consume low power and can operate uncooled, hence they are a proper choice for data centres. Due to the ease of direct coupling with the relatively large core of MMF (50 μm), they use MMF as a transmission medium. However, as a result of dispersion their reach is limited to relatively short range application of up to around 300 m on OM3 and around 550 m on OM4 for 10 Gb/s. Furthermore, relaxed coupling requirements as in 40 Gb/s and 100 Gb/s optical transceivers cause their reach to be limited to 100 m on OM3 and 150 m on OM4. The record bit rate demonstrated with VCSELs is 64 Gb/s [64g-vcsel-2014], but commercial directly modulated VCSELs exist only at 10 Gb/s. Since they are intrinsically incompatible with WDM, bandwidth scaling results with physical equipment scaling and higher aggregate rates are realized by deploying parallel transceivers and multi-mode ribbon fibres. The future challenges for this technology are the reach limitations due to dispersion, which will become even more important with further increase of the data rate. Shifting from 850 nm towards 1310 nm and single-mode operation holds the potential for both extended reach and WDM enabled operation.

To surpass the reach limitation of VCSEL based transceivers, DFB lasers combined with SMFs are used for LR and ER in data centres. They operate in two transmission windows, 1310 nm and 1550 nm. Although the loss at 1310 nm is twice the loss at 1550 nm (0.4 dB/km versus 0.2 dB/km), the dispersion penalty of operating at 1550 nm is much higher than the penalty at 1310 nm and considering that this penalty has a quadratic increase with the symbol rate, 1310 nm has been chosen as an operating wavelength for 40 Gb/s and 100 Gb/s modules. Bandwidth scaling for 40 Gb/s is achieved by using CWDM (4×10 Gb/s channels) and LAN WDM for 100 Gb/s (4×25 Gb/s channels). Transceivers for ER in general use external modulation, while recent advances in low threshold directly modulated DFB lasers has led to shifting from externally modulated to directly modulated DFB lasers for LR application. Although these transceivers may have some advantages over VCSEL based transceivers, they are in general more expensive and may require cooling. In addition, when comparing their footprint and power consumption, both of crucial importance for today’s and future

data centres, DFB based transceivers have the drawback that they consume more power and have larger footprint than VCSEL based interconnects.

In general, an optical module consists of a laser driver circuit, a VCSEL or DFB laser and an optional external modulator for the transmission part and a PIN photodiode with TIA for receiving data. Additionally, retimed electrical interfaces may have Clock and Data Recovery (CDR) blocks included in the module. Usually, 10 Gb/s electrical interfaces can be un-retimed, however 25 Gb/s electrical interfaces require a CDR for each electrical lane. Some of the early 100 Gb/s modules that operated at 4×25Gb/s but had to face electrical interface of 10×10 Gb/s also had a gearbox included to perform 10:4 and 4:10 operation. Another type of optical transceivers that has been mainly proposed for metro/access networks, but could also be used in data centres are transceivers based on Fast Tuneable Lasers (FTL) and Burst Mode Receivers (BMR). Having this type of transceivers enables sub-wavelength access to network resources and efficient bandwidth utilization.

Also, this section provides a brief review of the current available bandwidth variable transceiver which is widely discussed in future flexi-grid telecommunication network scenarios.

6.1 Existing 1 Gb/s and 10 Gb/s Optical Transceivers

The 1 Gb/s and 10 Gb/s optical transceivers come in a few pluggable packages, like XFP (10 Gigabit Small Form-Factor Pluggable), SFP (Small Form-Factor Pluggable) or SFP+ (Enhanced SFP). The standard SFP+ dimensions are 56.5×14.8×11.9 mm and power consumption is usually below 1 W [*sfp-10g-2014*] . More specification of commercial available 1 Gb/s and 10 Gb/s optical transceivers are summarized as follows:

- The SR module for 1 Gb/s and 10 Gb/s optical transceivers consists of directly modulated VCSEL laser at 850 nm and a PIN photodiode. The reach of SR transceivers is limited to 300 m on OM3 MMF and 550 m on OM4 MMF.
- The LR module uses a directly modulated DFB laser, operating at 1310 nm and a PIN photodiode. This module uses SMF and can support up to 10 km distance.

Regarding the scaling capability, the parallel architecture is not very efficient when network scaling is observed and results in new fibre deployments that are spatially inefficient and can be quite expensive. In addition, transiting from electrical to optical switching in the data centre might not be straightforward with this solution, due to the linear increase of interfaces needed for each new connection.

6.2 Existing 40 Gb/s Optical Transceivers

To provide 40 Gb/s data rates, either spatial or wavelength multiplexing is required. These transceivers usually come in a QSFP module, with dimensions of 72.4×18.4×13.5 mm [*qsfp-2014*] . In particular,

- The SR module for 40 Gb/s optical transceivers consists of four directly modulated VCSEL lasers at 850 nm operating at 10 Gb/s and four PIN photodiodes. Eight MMF fibres are needed for duplex communication, so usually a ribbon cable consisting of 12 MMFs is used. The reach due to relaxed coupling, is limited to 100 m on OM3 MMF and 150 m on OM4 MMF. The typical power consumption for the latest packaging QSFP2 of this type of optical module with un-retimed electrical interface is below 1.5 W.
- The LR module uses four directly modulated DFB lasers, typically operating uncooled at 1310 nm with 20 nm spacing (CWDM grid) and four PIN photodiodes. Two SMFs are needed to support bidirectional communication up to 10 km distance. The typical power consumption for the latest QSFP generation is below 3.5 W.

6.3 Existing 100 Gb/s Optical Transceivers

The 100 Gb/s optical transceivers come usually in CXP or CFP packaging. The standard CFP dimensions are 144.7×82×13.6 mm, however latest versions such as CFP4 can be as small as 92×21.5×9.5 mm [*cfp-2014*] .

The Ethernet specifications for the 100 Gb/s transceivers define the 100G-SR10 standard where special multiplexing is used in the same way as for 40 Gb/s. In addition, the current development of VCSELs operating reliably at 25 Gb/s has led to developing the 100G-SR4 standard, where wavelength multiplexing is exploited to achieve the desired data rate and reuse the same fibre link that has been used for 10 Gb/s. The most common long reach solution is as specified in 100G-LR4, where wavelength multiplexing of four wavelengths in the 1310 nm band with increased data rate of 25 Gb/s is used. In detail:

- The SR module optical transceivers consists of ten or twelve directly modulated VCSEL lasers at 850 nm operating at 10 Gb/s. 20 to 24 MMF fibres are needed for duplex communication, usually bundled in a ribbon cable. The reach is limited to 100 m on OM3 MMF and 150 m on OM4 MMF. The typical power consumption of this module is below 3.5 W.
- The LR module uses four externally or directly modulated DFB lasers, typically operating cooled at 25 Gb/s. The nominal wavelength is 1310 nm, and the four channels are spaced with 800 GHz spacing (LAN WDM grid). Two SMFs are needed to support bidirectional communication up to 10 km distance. The typical power consumption of the first generation CFP was around 20 W, however latest generation products can consume 4-5 W only. Though not standardized by IEEE, a 10×10 MSA module in CFP packaging with 10 electrical and optical lanes, each 10 Gb/s, also exists. The nominal operating wavelength is 1550 nm, and 10 DFB lasers are used with two SMF fibres for distances up to 2 km (10 km). This module consumes slightly below 20 W.
- The ER transceivers are also based on four DFB lasers, usually externally modulated at 25 Gb/s and operating cooled at 1310 nm with 800 GHz spacing (LAN WDM grid). SMF is used as a transmission medium with up to 40 km transmission distance.

6.4 Recent Approach on Bandwidth Variable Transceivers

BVT is a key enabling technology for future Elastic Optical Networking (EON) [eon-cm-2012]. These new networking paradigms require transceivers operating on a flexible wavelength grid (e.g., 12.5 GHz wide frequency slots with 6.25 GHz granularity for centre frequencies of the slots) and are able to accommodate traffic needs by flexibly adapting spectral efficiency, bit rate and reach.

The choice of architecture (e.g., trade-off between complexity in the electrical and optical domain) will also determine the most suitable technology to implement BVTs. Network operators would like to leverage the benefits of flexible transceivers, however without accepting a premium on capital expenditure. Hence, one of the main challenges for research and engineering is the cost-efficient realization of flexible functionalities required in future networks. The analogue electronic and optoelectronic parts of a transceiver are usually designed for a specific target symbol rate. It is therefore sensible to assume that a cost-efficient realization of a BVT operates at a fixed symbol rate. Under this assumption, several approaches exist to realize flexible variation of the three key parameters of a transceiver, which are: SE, bit rate and reach [*bvt-cm-2015*].

Multi-carrier solutions such as coherent WDM, CO-OFDM, Nyquist-WDM, as well as dynamic OAWG have been proposed as possible transponder implementations for EONs. These solutions rely on the generation of many low-speed subcarriers to form broadband data waveforms using lower-speed modulators so that terabit-per-second data can be generated using lower speed electronics at < 40 Gbaud. Despite the similarity of multicarrier signal generation between CO-OFDM, CO-WDM, Nyquist-WDM, and OFDM transmitters, there are fundamental differences with respect to their operation principles and capabilities [*bvt-cm-2015*].

- The main idea of Nyquist-WDM is to minimize the spectral utilization of each channel and reduce the spectral guard bands required between WDM channels generated from independent lasers. Using aggressive optical prefiltering with spectral shape approaching that of a Nyquist filter with a square spectrum minimizes the channel bandwidth to a value equal to the channel baud rate. Root raised cosine (RRC) is a popular choice for the pulse-shaping filter. Matched RRC filters at transmitter and receiver reduce the Intersymbol Interference (ISI) due to the

narrow filtering. The channels are then packed closely together such that the subcarrier spacing is equal to or slightly larger than the baud rate. And the bit rate carried by each NWDM subcarrier can be programmed individually (e.g., by varying the modulation format). However, significant power penalties arise from setting the channel frequency spacing equal to the baud rate, which requires Forward error correction (FEC) to achieve error-free performance. Hence, a trade-off exists between spectral efficiency and inter-carrier interference [*isi-2011*].

- In CO-WDM systems [*cowdm-leos-2009*][*cowdm-oe-2010*], to maintain orthogonality between subcarriers, the CO-WDM subcarrier symbol rate is set equal to the subcarrier frequency spacing. This restricts each modulator to generating only an integer number of subcarriers. Additionally, for both OFDM and CO-WDM, the chromatic dispersion tolerance scales with the subcarrier data rate instead of the total bit rate [*coofdm-oe-2008*]. However, the use of lower-bandwidth subcarriers (~100 MHz) in OFDM causes high peak-to-average power ratios, which increases susceptibility to nonlinear impairments. The larger subcarrier bandwidths (~10–40 GHz) typically used in CO-WDM have a lower peak-to-average-power ratio with comparable performance to isolated single-carrier systems [*cowdm-oe-2010*] .
- Optical OFDM is based on the transmission of multiple orthogonal subcarriers, which can be independently modulated by different formats. The orthogonal subcarriers are overlapped in the frequency domain, providing a high SE. In all-optical OFDM, the frequency multiplexing/demultiplexing is performed in the optical domain, and the number of subcarriers is limited to minimize cost and complexity. Electronic OFDM systems are based on DSP, and the OFDM subcarriers are generated in the electrical domain before optical modulation, exploiting an additional module with respect to Figure 13 [*bvt-cm-2015*]. Thus, also compared to all optical OFDM, a finer granularity (on the order of hundreds or even tens of megahertz) and narrower subcarrier spacing can be achieved, yielding unique bit rate/bandwidth scalability and spectral domain manipulation capability, with sub- and super-wavelength granularity, and thus suitable for EONs. A key element for enabling the transponder programmability and reconfigurability is the DSP, which allows adaptive modulation format selection at each subcarrier, and variable code rate and reach, according to the traffic requirements and channel profile. Both non-coherent and coherent optical front-ends are considered for the programmable flex subcarrier module. Either intensity modulation or linear field modulation can be implemented to use simple direct detection, for a cost-effective BVT design suitable for metro/regional networks. In this case the coherent receiver stage of Figure 13 is replaced with a single photodiode. On the other hand, coherent O-OFDM makes full use of the optoelectronic front-ends described in Figure 13. By combining coherent detection and DSP, the tolerance to transmission impairments can be significantly enhanced, achieving ultimate performance [*bvt-cm-2015*].
- With OAWG, the coherent combination of many spectral slices generated in parallel enables the creation of a continuous output spectrum [*SDM-survey-2015*. In this case, arbitrary-bandwidth single- and multicarrier channels can be generated, in which each channel can be in a different modulation format. The versatility of this signal generation technique enables customization of generated waveforms over the total operational bandwidth of the transmitter. This enables avoiding large peak-to-average-power ratios and incorporating pre-compensation for impairments such as chromatic dispersion. OAWG also removes the restriction that the modulator bandwidth must be a multiple or sub-multiple of any generated channel or subcarrier bandwidth.

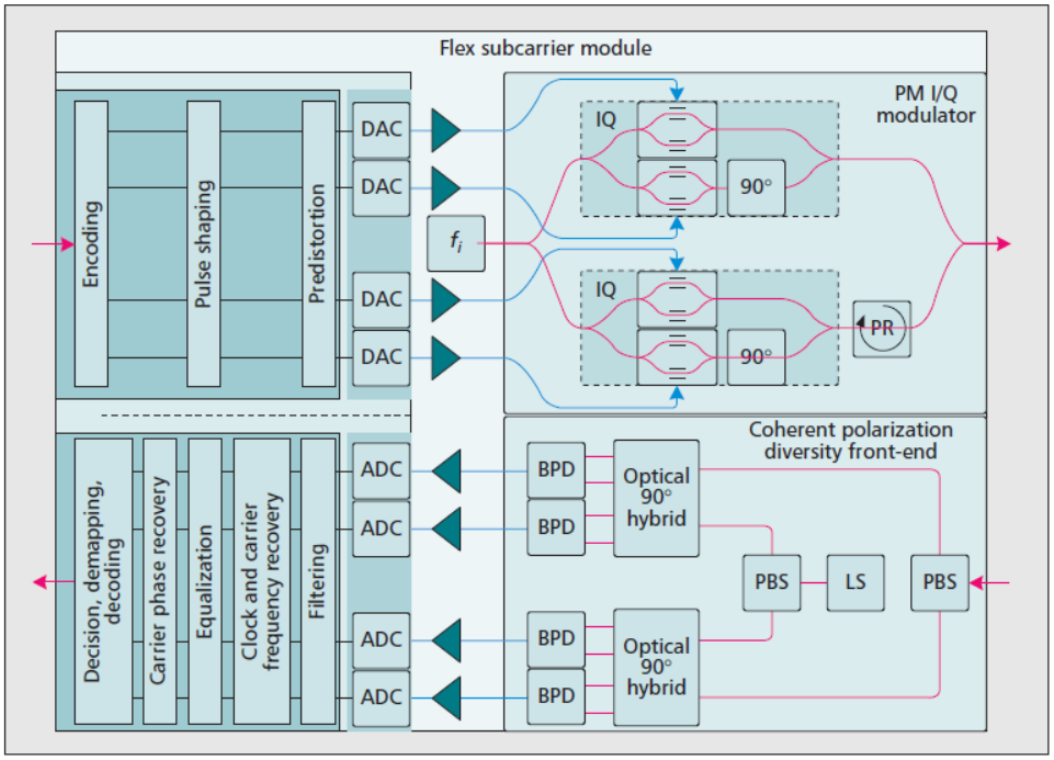


Figure 13: Flex subcarrier module building blocks [bvt-cm-2015]

6.5 Summary and Comparison Analysis

The overview given above refers to the specifications in the Ethernet standard related to the different means of providing various data rates and their practical realization in terms of transceivers and fibres used. Table 8 summarizes some of the most important characteristics of each alternative for 10 Gb/s, 40 Gb/s and 100 Gb/s. It can be seen that VCSELs provide the cheapest and most energy efficient solution, however they have limited reach and in general, scaling to more WDM channels is more difficult at 850 nm, as it requires special fibre engineering due to dispersion. On the other hand, DFB based transceivers can have higher power consumption, but provide intrinsic support for WDM and extended reach.

Table 8: 10 Gb/s, 40 Gb/s and 100 Gb/s Ethernet specifications

Standard	Transmitter	Fibre links	Channel data rate [Gb/s]	Aggregate data rate [Gb/s]	Power consumption [W]
10GBASE-SR	VCSEL	MMF	10	10	1
10GBASE-LR	DFB	SMF	10	10	1
40GBASE-SR4	4 × VCSEL	4 x MMF	10	40	1.5
40GBASE-LR4	4 × DFB	SMF	10	40	3.5
100GBASE-SR10	10 × VCSEL	10 x MMF	10	100	4
100GBASE-SR4	4 × VCSEL	MMF	25	100	4
100GBASE-LR4	4 × DFB	SMF	25	100	5

In addition, Table 9 summarizes the characteristics of existing BVT solution. These techniques present different levels of spectral efficiency and complexity. The choice of transmission technique may depend on the particular scenario, and can be influenced by the links and affordable or required costs.

Table 9: BVT Transmission techniques and characteristics

	Waveform Generation Technique	Transmitter/Receiver Implementation complexity
Nyquist-WDM	Nyquist-WDM combines many independently generated channels together with a minimum guard band	Mainly driven by DAC/ADC and DSP is required in receiver side (e.g., electronic bandwidth ≥ 0.5 × symbol rate) Sampling rate ≥ symbol rate
Optical OFDM	Generates many low-speed subcarriers using an IFFT to ensure orthogonality, mainly driven by DAC and DSP	Mainly driven by DAC/ADC and DSP (e.g., IFFT processing, sampling rate > symbol rate, in receiver side electronic bandwidth 0.5 × symbol rate). Sample rate: one sample per symbol
OAWG	OAWG operates over a continuous gapless spectrum, enabling the generation of any combination of single-carrier or multicarrier waveforms	Mainly driven by DAC/ADC and DSP Could completely control amplitude and phase over its operating bandwidth
CO-WDM	CO-WDM operates by combining many orthogonal subcarriers together to form a seemingly random waveform	Mainly driven by DAC/ADC and DSP

7 DCN Scenarios with Optical Technologies

Following the review and comparison analysis of optical technologies, this section provides input to the DCN structure design and system level performance comparison. First, a DCN scenario is presented utilizing a fast optical switch, large scale fibre switch and SDM-based fibres. This structure could provide server-to-server (within the same rack) optical TDM interconnection through intra-rack fast switch, and OCS or optical TDM connection for inter-rack communication. Also, it is worth noting that there could be several different designs depending on the technologies available and more scenarios will be discussed in D1.4 [3], including short-term, mid-term, and long-term scenarios. Finally, the benefit of leveraging optical technology (e.g., optical switches with different capabilities and SDM based fibres) in future DCN is investigated, especially regarding DCN topology scalability, capacity and power consumption. Since some of the network elements/devices and control plane solutions in COSIGN are still under developing/manufacturing, other figures of merit discussed in Section 2 will be addressed in future deliverables through data plane and control plane demonstration in WP5.

7.1 Possible DCN Architecture 1

7.1.1 DCN topology

Considering the requirements of future DCs and the advantage of above mentioned optical technologies, utilizing optical switches and SDM-based fibre for intra- and inter-rack connections could lead into a scalable architecture with server-to-server capacity. On the Tx side, which means inside each NIC, cheap low-power consumption VCSEL arrays could be used. A proposed architecture is shown in Figure 14, and it is worth noting that all-optical ToR may consist of three switches. One OCS switch for long duration flows, serving both intra-rack and inter-rack communication demands. There are several scenarios of what this OCS switch could be, e.g., AWG, Wavelength Selective Switch (WSS). In addition, two ultra-fast switches are used for optical TDM connection to avoid the overhead of electronic packet-header processing, one of which will be dedicated to intra-rack communication and the second for inter-rack. This selection is made because the large volume of server traffic (~80%) usually will stay within the rack. However, this large traffic volume may be distributed between different servers and divided into short data flows, making optical TDM switching the most suitable solution. Furthermore, these NICs will be fully-programmable and hybrid, supporting both Optical TDM and OCS transmission. They will have electronic buffers integrated that will aggregate traffic from each server to any possible destinations and either send it through optical TDM or OCS connection. To interconnect the server with ToR switch or ToR switch with large-scale core switch, SDM based fibre and SDM based multiplexer could be leveraged which could ease the cabling complexity with increased bandwidth density. In addition, to interconnect ToRs with large scale optical switches, both star and mesh topology could be employed, but with different performance. Their capacity and scalability will be analysed in the following subsection.

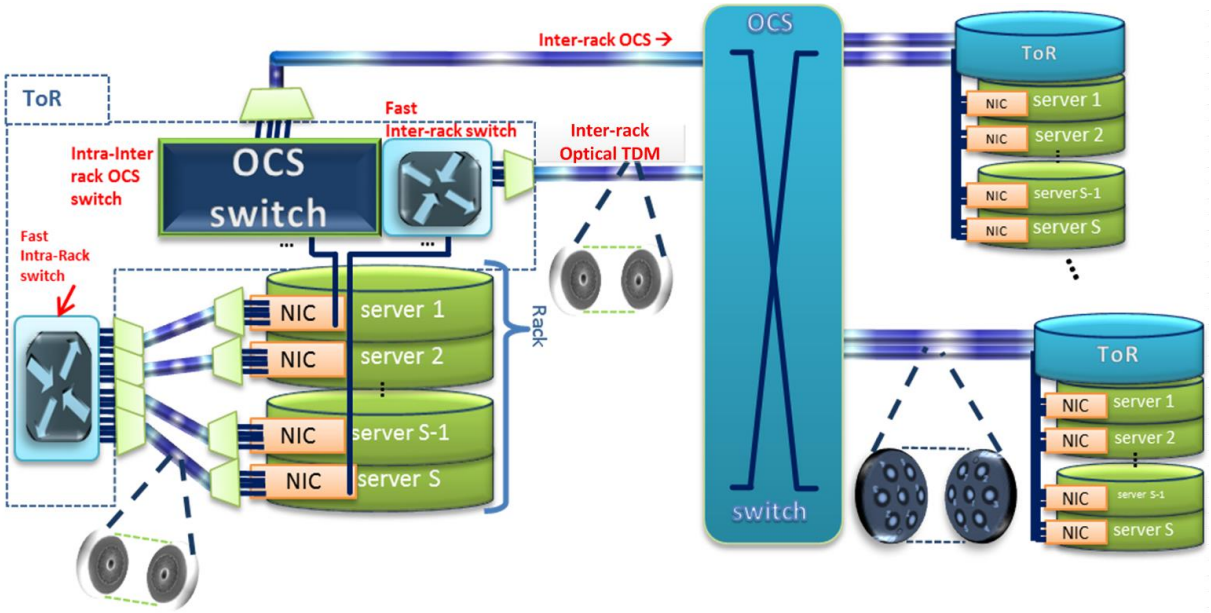


Figure 14: Proposed architecture utilizing SDM technology and Optical Switches

7.1.2 DCN Scalability and Capacity Analysis

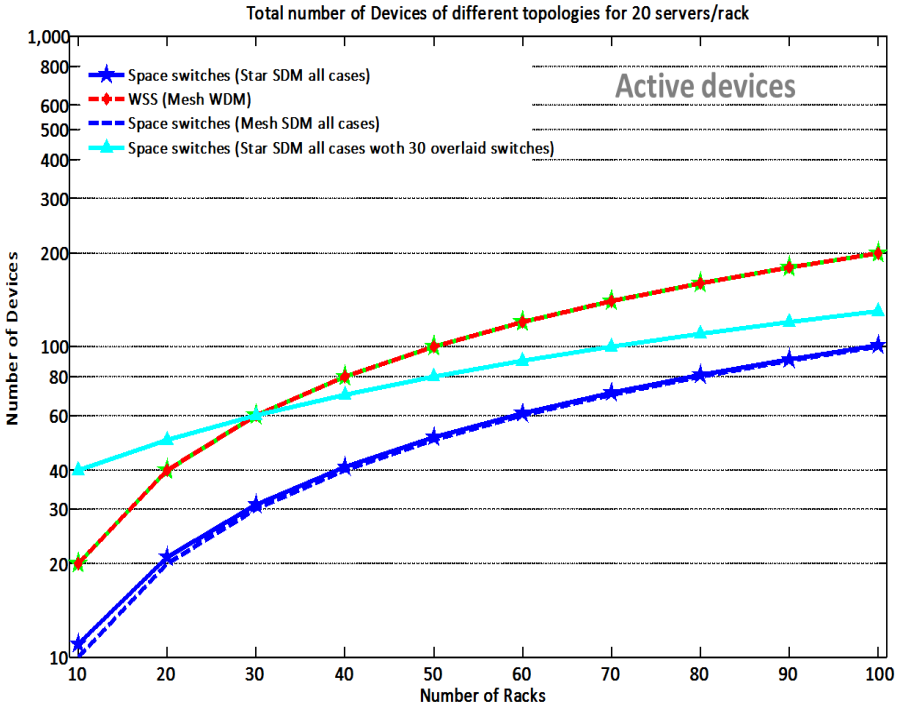
In this section a scalability analysis considering several instances of the above architectural design is presented. The scenarios investigated in terms of scalability vary from star and full mesh to spine-leaf topologies with diverse technologies, not only on the Tx/Rx side but also regarding fibre types. We consider pure SDM based architecture: utilizing cheap energy-efficient VCSEL-based transmitters, large port-count space switches as ToR and central switches and MCFs for the links. Also considered are WDM or WDM with SDM architectures where WDM Tx/Rxs are needed along with WSS as ToRs. Those architectures usually offer more flexibility in switching while requiring more expensive equipment.

The results of the analysis are presented in the figures below.

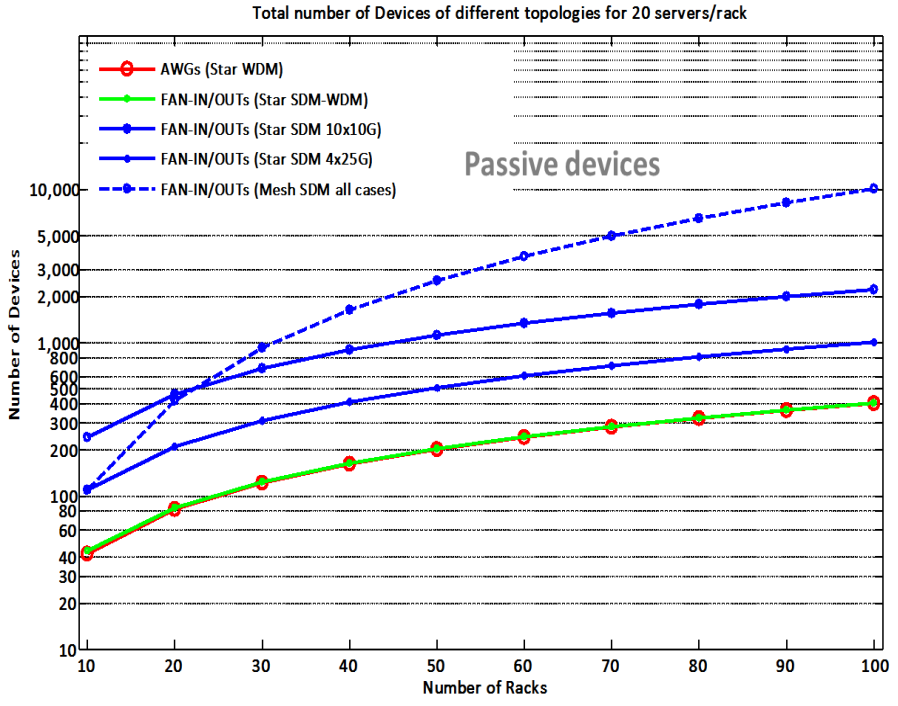
Figure 15 shows the number of active and passive devices required. The Mesh SDM 4×25G and the Star SDM 4×25G with overlaid switches (stacked switches) demonstrate very similar scalability, and they are both feasible as far as port count is concerned. Best performance is shown by star SDM-WDM and star SDM with multi-core switching, both with 30 overlaid switches in the centre, which is actually a spine-leaf topology. The mesh WDM solution is not directly comparable with the others, as it uses WSSs instead of a space switch, even though it is quite an impressive design for the future.

Figure 16 shows the scalability of the switches and the number of ports required to interconnect a certain number of servers and racks in a cluster. It is apparent that the mesh-based DCN designs require considerably larger number of devices, active such as WSS and passive such as SDM mux/demux (Spatial Multiplexer). Apparently, the considerable advantage of having a full-mesh topology is always-available all-to-all connectivity with lower blocking probability. That would definitely fit and serve well HPC and parallel computing systems where time boundaries are strict and there is no span for errors, re-transmissions or long buffering of data exchanges.

Figure 17 illustrates the number of fibres needed for each architecture. It is worth mentioning that different kinds of MCFs are used in the separate cases, depending on the Tx interfaces of each scenario. For example, for intra-rack connections, 4-core and 10-core MCF can be used to fit exactly the Tx modules in the NICs. Then different MCFs can be used for inter-rack communication, for instance 37-core or even more. The strong point here is the fact that the number of fibres, SMF or MCF, in all full mesh topologies scales aggressively with the number of racks per cluster, soaring for example above 9000 for 100 racks.



(a)



(b)

Figure 15: Comparison of different DCN topologies regarding the total number of devices needed [a] Active devices [b] Passive devices

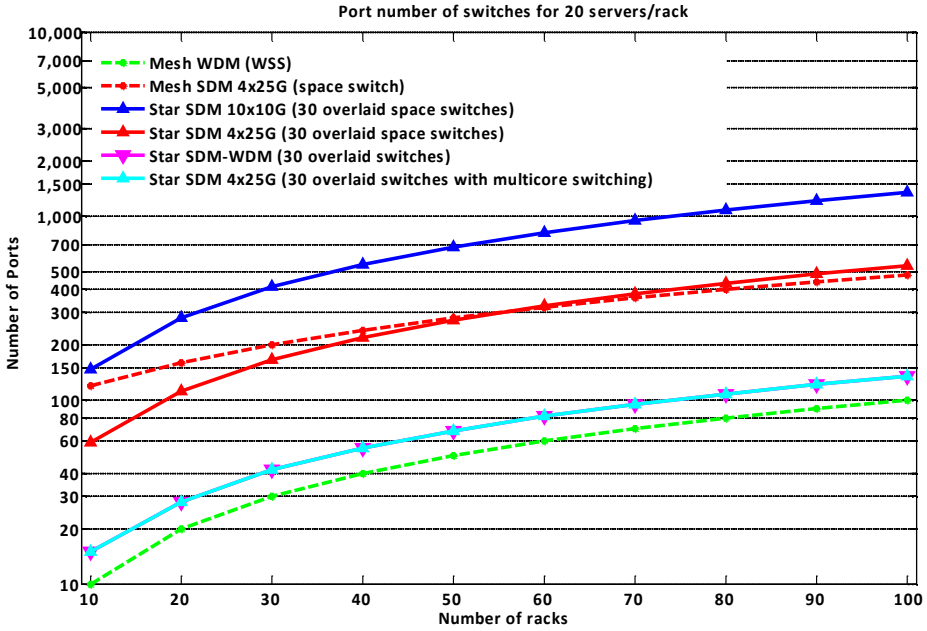


Figure 16: Comparison of switches port number for various Mesh (WDM, SDM) and Star (SDM,SDM-WDM) architectures and for 20 servers/rack vs. the number of racks per cluster

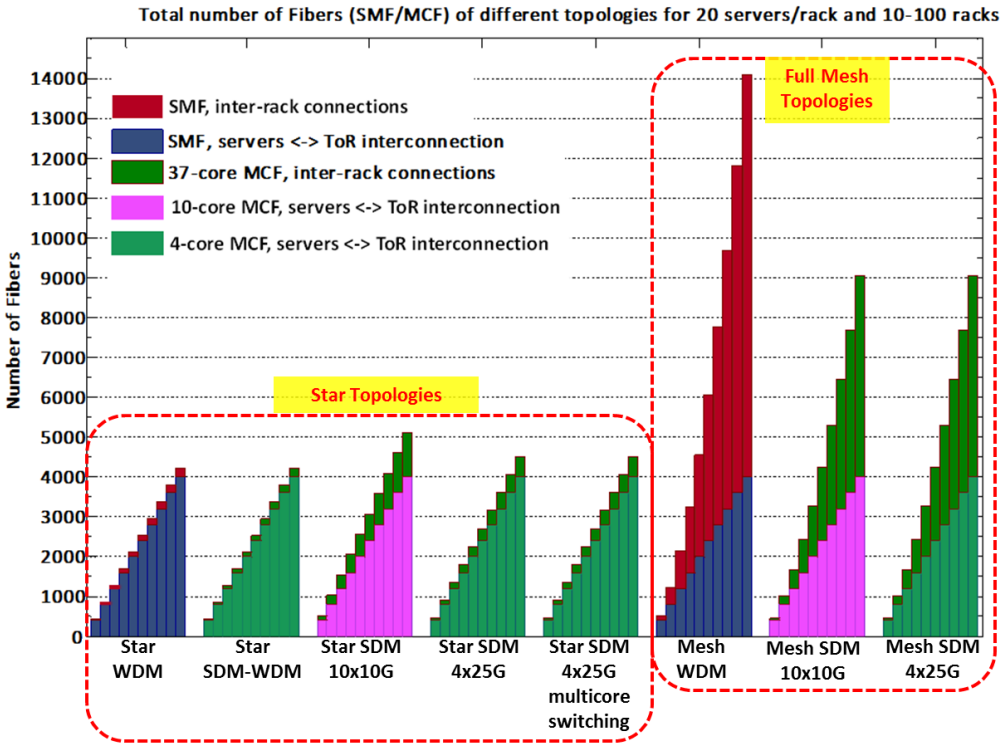


Figure 17: Comparison of various architectures presented before regarding their required number of fibres (SMF/MCF) for 20 servers/rack and 10-100 racks per cluster

Considering all the results from the above figures, the trade-offs for the star and mesh DCN topologies are crystal clear. For the Full-Mesh topologies the trade-off is between the number of ports of each ToR switch and the number of inter-rack fibres, whereas for all the Star topologies it is between the number of ports of central space switch (or switches when there are more than one overlaid, i.e., spine-leaf) and the number of inter-rack fibres once again.

7.2 Possible DCN Architecture 2

7.2.1 DCN topology

To achieve robust connectivity and high scalability, a ring network, as shown in Figure 18 (top), is a good candidate structure. Each network node serves a number of racks, where each rack contains a number of servers and a ToR switch. The ring network interconnects network nodes through (multicore/multiple) fibres, thus forming a cluster. Clusters can be interconnected through a cluster node (often duplicated for reliability) by, e.g., ring, tree, star or mesh structures. An example of a ring cluster interconnect is illustrated in Figure 18 (bottom), where a ring-of-rings (RoR) composition is proposed. The ring based architecture allows for fast and easy expansion of the data centre by either adding a new network node within a cluster, or a new cluster within the master ring. Concentrating the interconnections between nodes and clusters within a ring structure reduces the cabling complexity of the DCN dramatically, increases its scalability, and simplifies network deployment and management. Furthermore, the number of switches and cables/fibres needed within the DCN is significantly reduced compared to a fat-tree architecture, which improves the energy efficiency and the manageability of the infrastructure.

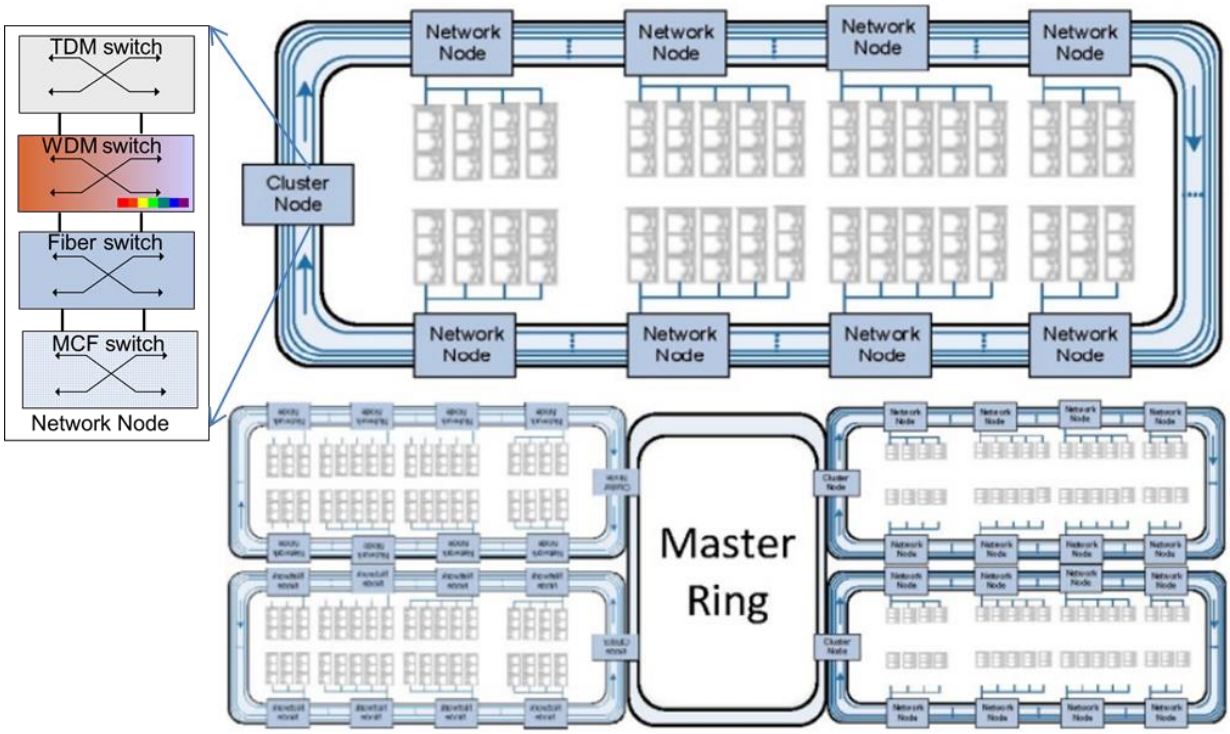


Figure 18: Generic ring-based cluster (top) and a ring-of-rings DCN configuration (bottom)

Four different levels exist in the node structure as shown in Figure 18. At the lowest level, switching of whole MCFs is performed; enabling traffic from full MCFs to be dropped at the Network Nodes. At the next level, individual core switching is performed, which allows for switching between cores in the MCFs, as well as add/drop of fibre cores. The top two levels perform wavelength and sub-wavelength TDM-based switching, thus allowing for efficient bandwidth utilization. The multi-level node structure allows establishing connections with capacity ranging from sub-wavelength to full MCF capacity, enabling any node to drop or add connections at different levels of granularity. Combined with SDN, this increases flexibility in the resource allocation, ultimately leading to reduced cost as well as adaptability to real-time demands, a feature highly desired for server-server and server-storage communication. The fundamental principle of the proposed node structure is to pass connections to the higher level only if they require switching at a finer granularity, allowing for bypass traffic to be largely unaffected by the presence of the node. This property of the Network Nodes translates directly to an inherently large amount of bypass traffic at any node, and therefore a ring topology is a suitable solution.

7.2.2 DCN Scalability and Capacity Analysis

Two main performance measures are considered: connection request blocking and network resource utilization. Furthermore, two scenarios are investigated. *Scenario 1* looks at the performance of both topologies under different capacity availability conditions in the DCN. The amount of available wavelengths per link is varied. The ratio of within-cluster/between-cluster traffic is set to 0.125 (i.e., 1/8 of the traffic is distributed within the cluster/pod and the rest is uniformly distributed between the other clusters/pods). *Scenario 2* investigates the performance of the topologies under varying traffic conditions, given fixed capacity availability in the network. In this scenario the ratio of within-cluster/between-cluster traffic is varied. The goal is to evaluate under what traffic patterns the corresponding topologies perform better. This will indicate the suitability of the topologies in handling different types of applications.

First, a numerical comparison between the proposed ring topology and the fat-tree network is performed, focusing on inventory and available capacity. From Table 10, it can be seen that for a network comprised of 8-port WDM switches, given the same number of supported ToRs, the RoR topology has 45.8% less links and 50% less nodes, compared to the fat-tree network. Switching from fat-tree to a ring-based topology reduces the amount of needed equipment significantly, which is a clear benefit with respect to inventory management, failure localization, cabling, occupied space, energy efficiency, etc. Furthermore, the bisection bandwidth (BW) of the RoR is more than 14 times lower.

Table 10: Inventory and capacity comparison

	# servers	# WDM nodes	# links	Bisection BW
FT	128	80	384	X
RoR	128	40	176	14.5*X

Next, we look at the performance of both networks under Scenario 1. Figure 19 shows the blocking ratio of the connection requests and the network resource utilization with varying resource availability in the DCN. The RoR topology outperforms the fat-tree network by 49% to 96% in terms of blocking and has 15% to 17% better utilization of its resources. Even though the fat-tree network has more links, and thus more paths between nodes, in a channel-switched scenario with millisecond long connections the available resources are not adequately utilized due to the concurrent nature of the resource reservation process. The RoR network on the other hand has less links, but the same amount of available capacity, which is better utilized to serve the requested connections. It is worth noting that the RoR network achieves lower blocking at more than 14 times lower bisection capacity.

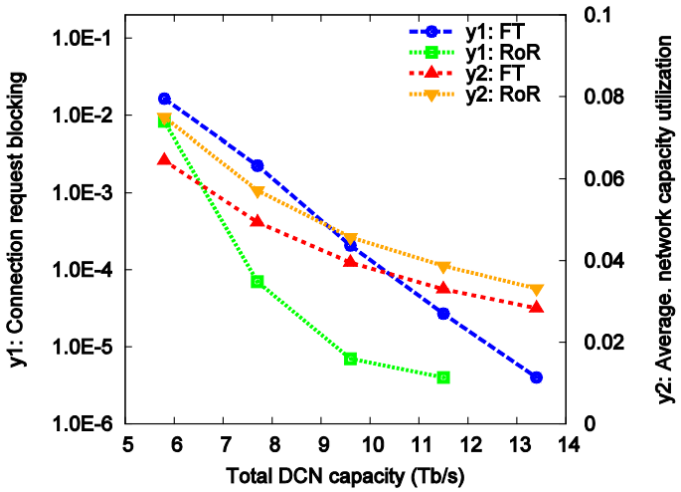


Figure 19: Performance evaluation under Scenario 1.

Finally, we look at the performance of the networks under Scenario 2, where different traffic conditions are simulated. Figure 20 shows the connection blocking ratio and Figure 21 the average

network capacity utilization as a function of local traffic ratio, i.e. when the ratio of within cluster/between-cluster traffic is changed, both for lightly loaded and for highly loaded networks. As before, the RoR DCN configuration outperforms the fat-tree architecture with respect to both performance measures (40% to 99% improvement in connection blocking and 2.6% to 17.6% improvement in resource utilization). It can be seen that at around 50% within-cluster/between cluster traffic distribution, the blocking is the lowest. At lower values, the core links of both DCN configurations get overloaded, whereas at higher values the links within a pod/cluster get overloaded, which results in higher blocking.

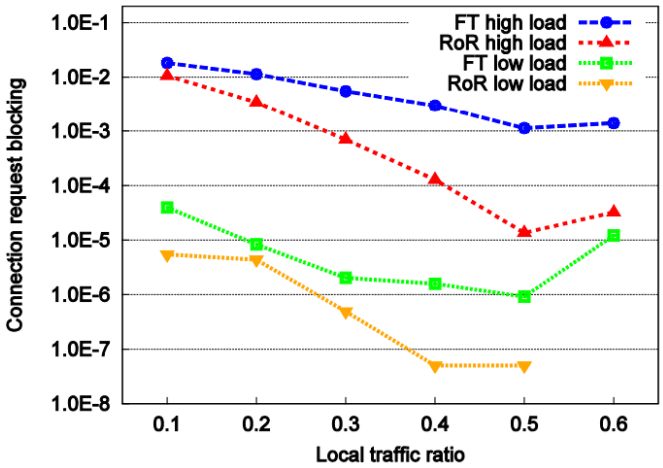


Figure 20: Connection request blocking vs. within cluster/between-cluster traffic ratio

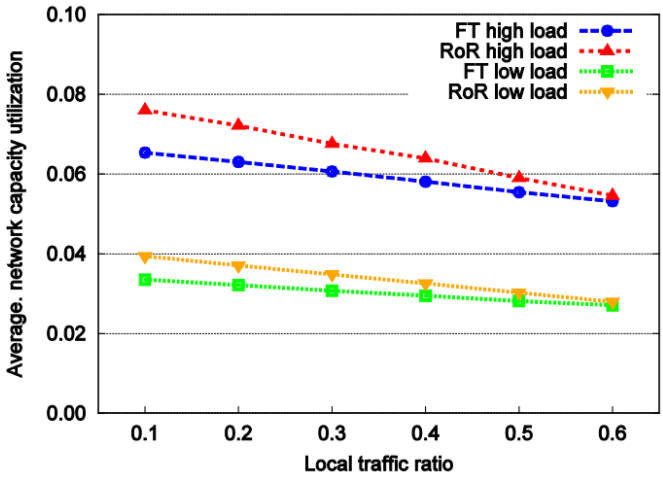


Figure 21: Network capacity utilization vs. within cluster/between-cluster traffic ratio

8 Summary

This deliverable has been focused on the analysis of figures of merit relevant for the COSIGN data plane and data plane requirements. Existing DCN related optical technologies, including transceivers, interfaces and fibres, have been evaluated. This study has been instrumental in defining the major functions for the COSIGN DCN data plane and in identifying the potential optical network devices/elements that will compose the overall COSIGN DCN data plane. Comparative results have been discussed with respect to various DCN topologies and employed optical technologies, showing that DCNs with optical devices/elements can achieve high scalability and provide better capacity to meet the future DCN requirements. The discussion on DCN architecture constitutes additional input for WP1, especially *D1.4 "Architecture Design",* which further analyses the potential DCN data plane architecture to support the final decision about the reference architecture for COSIGN developments.

.

REFERENCES

[*greenICT*] T. E. Klein, "Sustainable ICT Networks: The GreenTouch Vision," (available at http://www.greentouch.org/uploads/documents/3%20Thierry%20Klein_EU%20SEW%20-%20The%20GT%20Vision%20-%20v2.pdf)

[*power-eth-switch*] http://www.a-trac.com/documents/Cisco/Enterprise/brochures/Cisco_BROCHURE_Ethernet%20 Power%20Consumption%20for%20Switches_2.14.12.pdf

[*energyDCN-10*] D. Abts, M. R. Marty, P. M. Wells, P. Klausler, and H. Liu, "Energy proportional datacentre networks," *Proceedings of the International Symposium on Computer Architecture, pp. 338–347,* June 2010

[*opticalDCN-2013*] Kachris, Christoforos, Bergman, Keren, Tomkos, Ioannis, Optical Interconnects for Future Data Center Networks, 2013, Springer.

[*DCtraffic-13*] Fainman Y & Porter, G. "Directing Data Centre Traffic", *SCIENCE*, October 2013, Volume 342, Issue 6155, pp. 202-203.

[*FOCS-10*] Fiber-Optic Communication Systems, 4th Edition, Wiley, 2010.

[*transcost-2014*] http://www.digikey.com/product-search/en/optoelectronics/fiber-optics-transceivers/525358

[*fmf-2012*] Bigot-Astruc, Marianne; Boivin, D.; Sillard, Pierre, "Design and fabrication of weakly-coupled few-modes fibers," in *IEEE Photonics Society Summer Topical Meeting Series*, pp.189,190, 2012

[*fm-mdm-2012*] R. Ryf, S. Randel, A.H. Gnauck, C. Bolle, A. Sierra, S. Mumtaz, M. Esmaeelpour, E.C. Burrows, R. Essiambre, P.J. Winzer, D.W. Peckham, A.H. McCurdy, R. Lingle, "Mode-Division Multiplexing Over 96 km of Few-Mode Fiber Using Coherent 6 x 6 MIMO Processing," *J. Lightwave Technology*, vol.30, no.4, pp.521,531, 2012.

[*mdm-16qam-2012*] V.A.J.M. Sleiffer, Y. Jung, V. Veljanovski, R.G.H. van Uden, M. Kuschnerov, H. Chen, B. Inan, L. Grüner Nielsen, Y. Sun, D.J. Richardson, S.U. Alam, F. Poletti, J.K. Sahu, A. Dhar, A.M.J. Koonen, B. Corbett, R. Winfield, A.D. Ellis, and H. de Waardt, "73.7 Tb/s (96 x 3 x 256-Gb/s) mode-division multiplexed DP-16QAM transmission with inline MM-EDFA", *Optics Express*, Vol. 20, No. 26 , 2012

[*fm-mcf-2014*] T. Watanabe, Y. Kokubun, "Over 300 channels uncoupled few-mode multi-core fibre for space division multiplexing", in *OFC* h2A.50, 2014.

[*crosst-mcf-2011*] T. Hayashi, T. Taru, O. Shimakawa, T.Sasaki, E. Sasaoka, "Design and fabrication of ultra-low crosstalk and low-loss multi-core fiber", *Optics Express*, Vol. 19, No. 17, 2011.

[*ring-mcf-2012*] H. Takara, A. Sano, T. Kobayashi, H. Kubota, H. Kawakami, A. Matsuura, Y. Miyamoto, Y. Abe, H. Ono, K. Shikama, Y. Goto, K. Tsujikawa, Y. Sasaki, I. Ishida, K. Takenaga, S. Matsuo, K. Saitoh, M. Koshiba, and T. Morioka, "1.01-Pb/s (12 SDM/222 WDM/456 Gb/s) Crosstalk-managed Transmission with 91.4-b/s/Hz Aggregate Spectral Efficiency," in *ECOC*, paper Th.3.C.1., 2012.

[*line-mcf-2012*] Li Ming-Jun, B. Hoover, V.N. Nazarov, D.L. Butler, "Multicore fiber for optical interconnect applications," in *OECC* 2012.

[*19-mcf-2013*] J. Sakaguchi, B.J. Puttnam, W. Klaus, Y. Awaji, N. Wada, A. Kanno, T. Kawanishi, K. Imamura, H. Inaba, K. Mukasa, R. Sugizaki, T. Kobayashi, M. Watanabe, "305 Tb/s Space Division Multiplexed Transmission Using Homogeneous 19-Core Fiber," *J. Lightwave Technology*, vol.31, no.4, pp.554,562, 2013.

[*lea-mcf-2011*] K. Takenaga, Y. Arakawa, Y. Sasaki, S. Tanigawa, S. Matsuo, K. Saitoh, and M. Koshiba, "A large effective area multi-core fiber with an optimized cladding thickness", *Optics Express*, Vol. 19, No. 26, 2011.

[*mcf-design-2011*] S. Koshiba, M. Takenaga, "Multi- core fiber design and analysis: coupled-mode theory and coupled-power theory", Optics Express, Vol. 19, No. 26, 2011.

[*h-mcf-2013*] Y. Sasaki, Y. Amma, K. Takenaga, S. Matsuo, K. Saitoh, M. Koshiba, "Investigation of crosstalk dependencies on bending radius of heterogeneous multicore fiber," OFC, pp.1,3, 2013.

[*mlt-core79*] S. Iano, T. Sato, S. Sentsui, T. Kuroha, and Y. Nishimura, "Multicore optical fiber," OFC, OSA Tech. Digest Series, 1979.

[*CFP*] CFP2 Baseline Concept www.cfp-msa.org

[*seven-core-oe-10*] Seven-core multicore fiber transmissions for passive optical network, *Optics Express*, 24 May 2010, Vol. 18, No. 11.

[*sfp-10g-2014*] SFF Committee, SFF-8431 Specification for SFP+ 10 Gb/s and Low Speed Electrical Interface, ftp://ftp.seagate.com/sff/SFF-8431.PDF, [September 2014]

[*qsfp-2014*] SFF Committee, SFF-8436 Specification for QSFP+ 10 Gbs 4x Plugable transceiver, ftp://ftp.seagate.com/sff/SFF-8436.PDF, [September 2014]

[*cfp-2014*] CFP Multi-Source Agreement (MSA), http://www.cfp-msa.org/index.html, [September 2014]

[*Ethernet-ieee-2012*] IEEE 802.3™-2012 – IEEE Standard for Ethernet, http://standards.ieee.org/about/get/802/802.3.html, [September 2014]

[*AEN-1909 Rev 2.0*] Ref Application Notes AEN-1909 Rev 2.0, www.usconec.com/resources/AEN/AEN-1909.pdf

[*mef-sdm-oe-2014*] S. Jain, V. J. F. Rancaño, T. C. May-Smith, P. Petropoulos, J. K. Sahu, and D. J. Richardson, "Multi-Element Fibre Technology for Space-Division Multiplexing Applications", *Optics Express*, Vol. 22, pp. 3787-3796, 2014.

[*mef-ecoc-2013*] V. J. F. Rancaño, S. Jain, T. C. May-Smith, J. K. Sahu, P. Petropoulos, and D. J. Richardson, "First demonstration of an amplfied transmission line based on multi-element fibre technology," in *ECOC* 2013.

[eon-cm-2012] Gerstel, O.; Jinno, M.; Lord, A.; Yoo, S.J.B., "Elastic optical networking: a new dawn for the optical layer?," *Communications Magazine, IEEE* , vol.50, no.2, pp.s12,s20, February 2012

[*transponder-ofc-2013*] K. Christodoulopoulos, P. Soumplis, and E. Varvarigos, "Trading off Transponders for Spectrum in Flexgrid Networks," in *OFC 2013*, paper OTu2A.3.

[*oam-ofc-2014*] J. Wang and A. Willner, "Using Orbital Angular Momentum Modes for Optical Transmission," in *Optical Fiber Communication Conference*, 2014, paper W4J–5.

[*oam-np-2012*] J. Wang, J.-Y. Yang, I. M. Fazal, N. Ahmed, Y. Yan, H. Huang, Y. Ren, Y. Yue, S. Dolinar, M. Tur, and others, "Terabit free-space data transmission employing orbital angular momentum multiplexing," *Nat. Photonics*, vol. 6, no. 7, pp. 488–496, 2012.

[*oam-mdm-ecoc-2012*] N. Bozinovic, Y. Yue, Y. Ren, M. Tur, P. Kristensen, A. Willner, and S. Ramachandran, "Orbital angular momentum (OAM) based mode division multiplexing (MDM) over a Km-length fiber," in *ECOC*, 2012.

[*mdm-hcpgf-jlt-2014*] V. A. J. M. Sleiffer, Y. Jung, N. K. Baddela, J. Surof, M. Kuschnerov, V. Veljanovski, J. R. Hayes, N. V. Wheeler, E. R. N. Fokoua, J. P. Wooler, D. R. Gray, N. H.-L. Wong, F. R. Parmigiani, S.-U. Alam, M. N. Petrovich, F. Poletti, D. J. Richardson, and H. de Waardt, "High Capacity Mode-Division Multiplexed Optical Transmission in a Novel 37-cell Hollow-Core Photonic Bandgap Fiber," *J. Light. Technol.*, vol. 32, no. 4, pp. 854–863, Feb. 2014.

[*isi-2011*] X. Liu and S. Chandrasekhar, "High Spectral-Efficiency Transmission Techniques for Systems Beyond 100 Gb/s," *Sig. Proc. Photonic Commun*., 2011, paper SPMA1.

[*cowdm-leos-2009*] S. K. Ibrahim, A.D. Ellis, F.C.G. Gunning, J. Zhao, P. Frascella and F. Peters, "Practical implementation of coherent WDM", in Proc. IEEE Photonics Soc. (LEOS) Annual Meeting, 4-8 October 2009, ThM1.

[*cowdm-oe-2010*] P. Frascella *et al.*, "Unrepeatered Field Transmission of 2 Tbit/s Multi-Banded Coherent WDM Over 124 km of Installed SMF," *Opt. Express*, vol. 18, 2010, pp. 24745–52.

[*coofdm-oe-2008*] W. Shieh *et al.*, "Coherent Optical OFDM: Theory and Design," *Optics Express*, vol. 16, Jan 2008, pp. 841–59

[*bvt-cm-2015*] Sambo, N.; Castoldi, P.; D'Errico, A.; Riccardi, E.; Pagano, A.; Moreolo, M.S.; Fabrega, J.M.; Rafique, D.; Napoli, A.; Frigerio, S.; Salas, E.H.; Zervas, G.; Nolle, M.; Fischer, J.K.; Lord, A.; Gimenez, J.P.F.-P., "Next generation sliceable bandwidth variable transponders," *IEEE Communications Magazine*, vol.53, no.2, pp.163-171, Feb. 2015

[*oawg-oe-2011*] D. J. Geisler et al., "Bandwidth Scalable, Coherent Transmitter Based on the Parallel Synthesis of Multiple Spectral Slices Using Optical Arbitrary Waveform Generation," *Optics Express*, vol. 19, Apr. 2011, pp. 8242–53.

[*SDM-survey-2015*] Saridis, G.M.; Alexandropoulos, D.; Zervas, G.; Simeonidou, D., "Survey and Evaluation of Space Division Multiplexing: From Technologies to Optical Networks," in *IEEE Communications Surveys & Tutorials*, 2015, pre-print.