**Optical Networks**
*Series Editor:* Biswanath Mukherjee

Víctor López
Luis Velasco *Editors*

# Elastic Optical Networks

## Architectures, Technologies, and Control

Springer

Optical Networks

More information about this series at http://www.springer.com/series/6976

Víctor López • Luis Velasco
Editors

# Elastic Optical Networks

## Architectures, Technologies, and Control

Springer

*Editors*
Víctor López
Telefónica Global CTO
Madrid, Spain

Luis Velasco
Universitat Politècnica de Catalunya
Barcelona, Spain

ISSN 1935-3839 ISSN 1935-3847 (electronic)
Optical Networks
ISBN 978-3-319-30173-0 ISBN 978-3-319-30174-7 (eBook)
DOI 10.1007/978-3-319-30174-7

Library of Congress Control Number: 2016935860



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

# Contents

# Chapter 1
# Motivation

**Víctor López and Luis Velasco**

It is widely acknowledged by operators and market analysts from all parts of the telecommunication industry, bandwidth demand is increasing dramatically, year on year, with typical growth figures of up to 30 % for Internet-based traffic.

This current growth is predicted to continue for two main reasons:

- Proliferation of disruptive, high bandwidth applications. The Internet becoming potentially the most popular way to consume video and in the near there will be an increased demand for online High Definition. But there are many other applications, covered collectively by the 'cloud' umbrella that could fuel bandwidth growth to a similar degree.
- Fibre and other high bandwidth access provision. There are substantial plans to roll out Fiber to the Premises (FTTP) and high bandwidth copper access to residential customers. In fact, Spain had 749 thousand FTTH subscribers in 2014 and had reached 1895 million FTTH users in 2015, doubling the users and becoming the country with most FTTH users in Europe.

These two trends mean that the consumers will have both a wide range of high bandwidth services and the capacity to access them. Of course, the knock-on effect of this ensures that the current rate of growth of IP network traffic will continue and possibly even speed up. However, we believe that the traffic will not only increase but will also become much more dynamic, in both time and direction. We can anticipate large changes in traffic magnitude during a 24-h period, as daytime business users have very different demands to evening-time residential customers. Additionally, as multiple content and cloud service providers offer competing services, the traffic direction will also become dynamic.

V. López (✉)
Telefónica Investigación y Desarrollo, Madrid, Spain
e-mail: victor.lopezalvarez@telefonica.com

L. Velasco
Universitat Politècnica de Catalunya, Barcelona, Spain

 
V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_1

Recent data centre trends justify the dynamic behaviour of the traffic. Cloud Computing allows a user to install, in a data centre, virtual machines with any kind of software. In this scenario, any virtual machine can work as a video-streaming or back-up server, thus increasing the unpredictability in the traffic. The bandwidth utilization pattern is quite different in each data centre. Moreover, the data centers do not have to be out of the operators' network, but they would be part of the network not only to provide cloud services but also to support Network Functions Virtualization (NFV).

In summary, a future transport network will be required to provide connectivity between IP routers with a wide range of bandwidths, some of which are much larger than today, and will also be dynamic over both time and direction, with timescales of minutes to hours.

What if we were to continue with the current transport network development programs as operators are currently doing? To answer this requires an understanding of existing Dense Wavelength Division Multiplexing (DWDM) transmission systems and their current evolution path. The majority of worldwide DWDM systems currently use 10 Gb/s transponder technology. This has become a commodity with a resultant very low price. However, an overall link limit of 10 Gb/s × 80 wavelengths gives only 800 Gb/s per link, which is already insufficient for many operators on their key, high bandwidth links.

The immediate solution is to upgrade the transponder technology to either 40 or 100 Gb/s. This is beginning to take place despite the relative cheapness of 10 Gb/s due to the need for high bandwidth on some links. Therefore, in some cases we are seeing mixed line rate networks in which a range of bit rates is carried on the same optical fibres. Much work is taking place to ensure compatibility and to develop any special design rules that may be needed to guarantee operation of different bit rates simultaneously on the same network.

It is anticipated that 100 Gb/s will meet operator needs in the short to medium term, but that the forecast traffic growth will lead to a strong requirement for 400 Gb/s and 1 Tb/s in the long term. As and when this occurs, it is expected that the use of mixed line rates will become more popular, as will be reflected in the following discussion around Figs. 1.1 and 1.2.

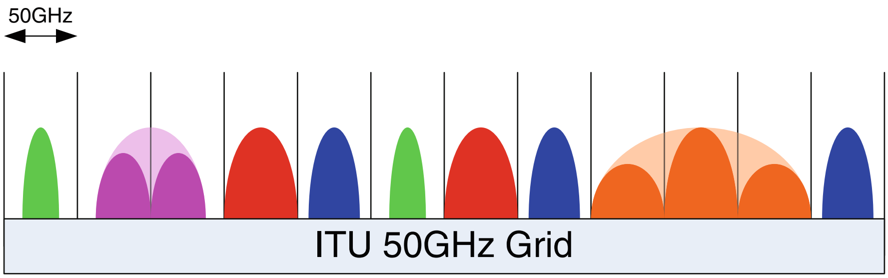



**Fig. 1.1** Impact of fixed grid filtering on large bandwidth demands

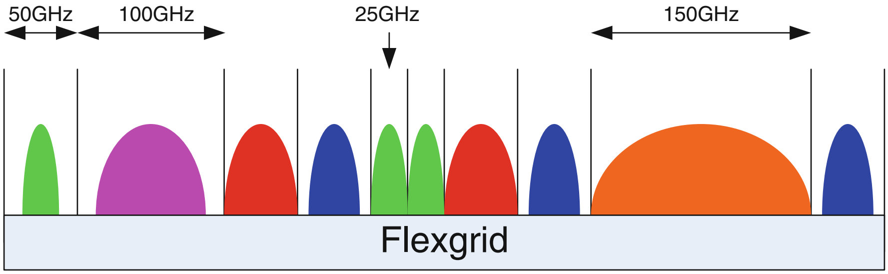



**Fig. 1.2** Flexgrid spectrum allocation makes sufficient space for large and small bandwidth demands

Existing DWDM systems divide the C-band optical spectrum into discrete bands, spaced usually by 50 or 100 GHz, and standardized by the ITU. A transponder provides an individual wavelength carrying a client demand (which might be Ethernet or Optical Transport Network (OTN), and might have a payload of anything up to 100 Gb/s), which can be accommodated in just one of these bands. This approach is inflexible in two distinct ways:

- The transponder has a fixed bit rate—for example of 10 or 40 Gb/s. In fact, it is possible to change some aspects of the transponder, such as different Forward Error Correction (FEC), but this does not affect the overall payload being carried.
- The spectral width of each wavelength signal cannot extend beyond the fixed ITU grid width used in the system (e.g. 50 GHz). The system will typically have wavelength-sensitive devices such as optical demultiplexers and Reconfigurable Add Drop Multiplexers (ROADMs) that require adherence to the fixed grid and would filter out any optical spectrum that crossed the grid boundaries.

Figure 1.1 shows a range of demands with different bandwidths, being fitted into a fixed grid network. Some of the demands fit comfortably between the 50 GHz grid boundaries, whereas others (pink and orange demands) are too broad. Optical filters, specified to the ITU fixed grid, will impose large filtering penalties as shown in Fig. 1.1.

This means that the resulting DWDM network is partially inflexible to changes in bandwidth demand. It is not completely inflexible however—for example, further transponders can be installed to cope with additional demands. However, this is a very slow process and typically can take many weeks to place the order for a new transponder and then conduct the necessary installation and provisioning. In addition, any very high bit rate transponders, such as 400 Gb/s and above, have a very broad optical spectrum, and are too large to fit into this grid.

Therefore, large bandwidth demands will have to be divided up so that they can be carried over the fixed grid. This results in a highly inefficient use of the network capacity, and consequently, in multi-layer networks will require a large amount of highly expensive, power-consuming IP/MPLS equipment to be installed for grooming/aggregation purposes.

If there were a simpler to manage, more cost effective, more flexible and dynamic solution than mixed line rate, fixed transponders, this would be a very attractive future alternative. Elastic optical networks (EON) are capable of fulfilling the requirements in terms of capacity and dynamicity of future core networks by combining two technologies: transponders and flexgrid transmission and switching technologies.

## 1.1 Bandwidth Variable Transponders

These transponders can adjust their bandwidth quickly, under software control by changing the modulation format. This is often referred to as software-defined optics. There are a wide range of options, each giving different bandwidth tunabilities. A major advantage of Bandwidth Variable Transponders (BVTs) is that they avoid the need to purchase many different bandwidth transponders in the future—there could simply be one type of transponder that delivers the wide range of bit rates required. This would drive development and production costs down.

BVTs potentially give another advantage: they could trade off reach and spectrum usage. For example if we only need to provide 100 Gbps over 400 km, we may use a highly spectrum-efficient modulation format such as DP-16QAM, which occupies only 25 GHz. However, if instead we need to reach 1500 km, then we can use double the spectrum with a modulation format DP-QPSK—thus avoiding regenerators and making highly efficient use of the available spectrum, all with the same transponder.

## 1.2 Flexgrid

The optical spectrum can be used in a more flexible way, where chunks of spectrum can be defined more arbitrarily than currently. Currently there is much discussion in the community about how the spectrum should be partitioned. Whilst this could be completely arbitrary, it is more likely that flexgrid will be implemented using two principles: (1) very fine bandwidth slices (6.25 or 12.5 GHz) and (2) individual slices will be able to be concatenated to make much larger contiguous portions of optical spectrum.

Figure 1.2 shows the capability of flexgrid to place grid boundaries in the most appropriate place so as to: (1) pass wide bandwidth channels without filtering, and (2) compactly pack channels to make maximum use of the spectrum. The industry has begun to use the term 'superchannel' to describe a general block of optical spectrum: this superchannel could contain multiple subchannels, grouped together for efficient transport across the network.

### 1.2.1 Structure of the book

In this book, we cover the above topics and present:

- The evolution from WDM to flexible or elastic optical networks as well as the new applications that EONs can support. Planning for EONs is presented, including its mathematical modeling and algorithms.
- From the data plane perspective, the transmission advances in EON are detailed, covering the next generation of bandwidth variable transponders, the novel modulation formats, the analysis and mitigation of physical impairments, and the node architectures.
- Regarding the control plane, the book introduces not only the extensions to GMPLS and OpenFlow for EONs, but also how Software-Defined Networking (SDN) can take advantage of the capabilities of EON, enabling even the in-operation planning.

# Chapter 2
# Evolution from Wavelength-Switched to Flex-Grid Optical Networks

**Andrew Lord, Yu Rong Zhou, Rich Jensen, Annalisa Morea, and Marc Ruiz**

## Acronyms

| | |
|---|---|
| APSK | Amplitude and phase-shift keying |
| AWG | Arrayed-Waveguide Gratings |
| BVT | Bit Rate Variable Transponders |
| C/D/C | Colourless Directionless Contentionless |
| CAPEX | Capital expenditure |
| DCM | Dispersion Compensation Module |
| DFT | Discrete Fourier Transform |
| DSP | Digital Signal Processing |
| DWDM | Dense Wavelength Division Multiplexing |
| FCAPS | Fault, configuration, administration, performance, and security |
| FPGA | Field Programmable Gate Array |
| FT | Fourier Transform |
| ITU-T | International Telecommunication Union |
| LCoS | Liquid-Crystal-On-Silicon |
| MEMS | Micro-Electromechanical System |
| MLR | Mixed Line Rate |

A. Lord (✉) • Y.R. Zhou
British Telecom, London, UK
e-mail: andrew.lord@bt.com

R. Jensen
Polatis, Bedford, MA, USA

A. Morea
Alcatel-Lucent, Boulogne-Billancourt, France

M. Ruiz
Universitat Politècnica de Catalunya, Barcelona, Spain



V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_2

| | |
|---|---|
| NMS | Network Management System |
| OA | Optical Amplifiers |
| OADM | Optical Add and Drop Multiplexer |
| OE | Opto-electronic |
| OXC | All-Optical Switch |
| PDM | Polarization-Division-Multiplexing |
| QAM | Quadrature Amplitude Modulation |
| QPSK | Quaternary Phase Shift Keying |
| ROADM | Reconfigurable OADM |
| SBVT | Sliceable Bit Rate Variable Transponders |
| SNR | Signal-to-noise ratio |
| WB | Wave Blockers |
| WSON | Wavelength switched optical networks |
| WSS | Wavelength Selective Switch |

This chapter reviews the basics of Dense Wavelength Division Multiplexing (DWDM) and shows how this will be superseded by a more flexible use of the optical fibre spectrum, together with more flexible transponders offering multiple bit rates from the same device. The chapter explains the benefits of this new approach and examines the optical filter technology that enables it. Finally, the chapter looks at how networks will migrate towards this new network operating paradigm.

## 2.1 Introduction

Until recently, the large available spectrum provided by optical fibre has been significantly more than required. Adding more data to a fibre was a simple matter of adding additional wavelengths, making use of the fact that at low enough power levels, multiple waves can be supported on the same fibre without interaction. Recently, continued internet-based exponential traffic growth has resulted in this spectrum to start to fill. This has focused attention on two related areas—how to more effectively manage the spectrum and how to fill the spectrum up as much as possible with light signals. These two revolutions in core transport form the theme of this chapter which forms an introduction to the more detailed discussions in later chapters.

## 2.2 The History of ITU Grids and Development from Fixed to Flex-Grid

Optical fibres are largely deployed in fibre-optic communications, as they enable transmission over longer distances and at higher bandwidths (data rates) than wire cables. These properties are due to lower losses and higher number of channels that can be simultaneously transported over their large spectrum window.

In optical fibre, the region between 1.3 and 1.6 μm is exploited for transmission. Within this region, the C band presents the lowest losses of the whole fibre spectrum and is exploited for transmitting over very long distances (from tens up to thousands kilometers). The C band refers to the wavelengths around 1550 nm and includes wavelengths between approximately 1525 nm (or a frequency of 195.9 THz) and 1565 nm (191.5 THz). (Dense) Wavelength Division Multiplexing ((D)WDM) denotes the technology enabling the transmission of a number of optical signal carriers onto a single optical fibre by using different wavelengths. To allow a universal use of the whole C band, the International Telecommunication Union Telecommunication Standardisation Sector (ITU-T) has defined a table of all the wavelengths (and their corresponding central frequencies); this list is detailed in the recommendation ITU-T G.694.1 [1]. In the beginning, DWDM wavelengths have been positioned in a grid having exactly 100 GHz (about 0.8 nm) spacing in optical frequency, with a reference frequency fixed at 193.10 THz (1552.52 nm).

To face the ceaseless traffic growth, in the last decade a large number of significant innovations have increased capacity by a factor of about 20 (compared to legacy DWDM systems at 10 Gb/s on a 100-GHz spacing). Firstly, it has been possible to squeeze channels by spacing them 50 GHz apart (about 0.4 nm) in core networks allowing the transport of around 80 (and most recently 96) channels. Furthermore, the channel capacity has been increased by installing optoelectronic devices at higher bit rates, always operating within the 50 GHz grid, passing from 2.5 to 10 Gb/s, then to 40 Gb/s, and since 2010 up to 100 Gb/s per wavelength; such channel spacing and bit rate allow 2 bit/s/Hz spectral efficiency.

Unlike the early days of DWDM systems, where the bandwidth of an optical fibre was considered to be infinite, in the near future, the optical spectrum will be a scarce resource and nowadays industry is inquiring how to improve the overall spectrum efficiency. Improvements in signal transmission techniques have permitted the reduction of the spectrum occupied by the optical signals: the combination of coherent detection techniques with Nyquist pulse shaping allows the transmission of 100 Gb/s in as little as 33 GHz channels.

Increasing the single carrier bit rate beyond 100 Gb/s requires the use of higher order Quadrature Amplitude Modulation (QAM)—for example 16-QAM doubles the bit rate compared to Quadrature Phase Shift Keying (QPSK) and hence provides 200 Gb/s capability. However, these higher QAM formats only work over shorter transmission distances. One way to implement higher rate channels such as 400 Gb/s and 1 Tb/s is by the adoption of multi-carrier signals with at least the same symbol rate than 100 Gb/s signals. As an example, a 400 Gb/s channel can be obtained with two 16-QAM-modulated sub-carriers (200 Gb/s each), each in 37.5 GHz, for a total of 75 GHz bandwidth; and a 1 Tb/s channel can be obtained with four 32-QAM-modulated sub-carriers, each within 43.75 GHz, for a total bandwidth of 175 GHz. The first consequence of this implementation is the need of larger channel spacing, breaking the standard 50 GHz grid per channel policy.

For the terrestrial networks, it became necessary to have a new standardized grid allowing the best spectral efficiency for the increased diversity of the spectrum requirements (e.g. a 100 Gb/s channel fitting in 37.5 GHz and 400 Gb/s in 75 GHz).

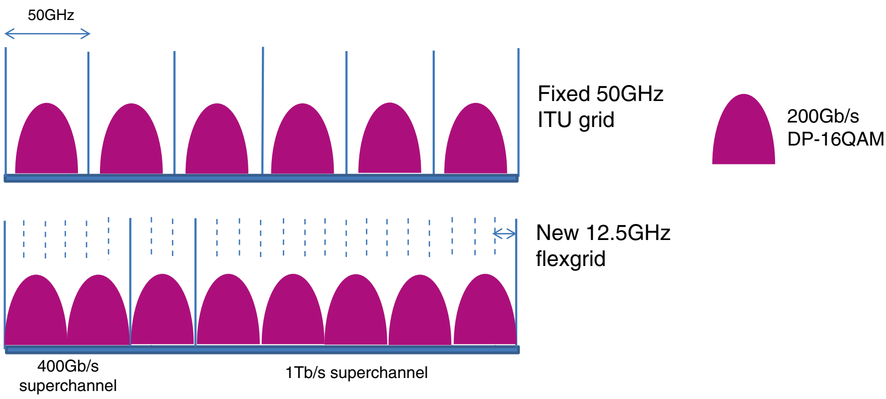



**Fig. 2.1** 12.5 GHz resolution for flex-grid allows closer packing of channels. Concatenation allows creation of a 400 Gb/s and 1 Tb/s super-channel

Addressing this issue, the ITU-T) proposed a finer grid associating a variable frequency slot to an optical connection, called flexible frequency grid, or more commonly flex-grid. Flex-grid allows the allocation of a variable number ($n$) of fixed-sized slots to an optical channel as a function of its requirements. A slot measures 12.5 GHz, allowing the transmission of 100 Gb/s channels in 37.5 GHz ($n = 3$), rather than 50 GHz in the fixed grid case. Figure 2.1 shows the impact of flex-grid on transmission. In this example, flex-grid is able to support 8 rather than 6 200 Gb/s channels, but additionally it is able to group these, if needed, into super-channels, which can be carried through an optical network as a single entity.

## 2.3 Point-to-Point Fixed-Grid DWDM Architectures

Before considering networks, we begin by describing point-to-point DWDM systems.

Figure 2.2 shows the schematic of a point-to-point DWDM system, where multiple wavelength channels are generated in optical transmitters, each being modulated by a data signal and then combined by a WDM multiplexer. The composite DWDM signal is then transmitted over an optical fibre link with optical amplifiers to boost signal before transmission, to compensate fibre loss at each span as well as to improve receiver sensitivity. The in-line amplifiers are usually two stage amplifiers as shown in the inset in Fig. 2.2, where the Dispersion Compensation Module (DCM) is used in each span to compensate fibre chromatic dispersion. When using coherent technology, the fibre dispersion can be compensated in the coherent receivers instead. Hence the new DWDM networks are designed to be DCM-less to enable optimum performance for higher speed using coherent technology. At the receiving end, a DWDM de-multiplexer is used to separate the DWDM signal into individual channels and data signals are recovered in the optical receivers.

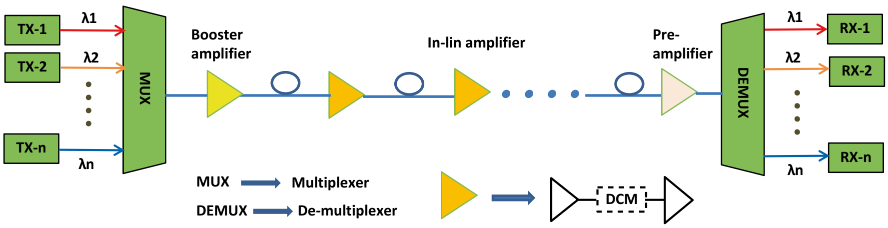



**Fig. 2.2** Schematic of a point-to-point DWDM system

**Table 2.1** Total fibre capacity using DWDM in C band with 50 GHz grid

| Data rate (Gb/s) | Total capacity (Tb/s) | | | Spectral efficiency (bits/s/Hz) |
|---|---|---|---|---|
| | 80 λs | 88 λs | 96 λs | |
| 2.5 | 0.2 | 0.22 | 0.24 | 0.05 |
| 10 | 0.8 | 0.88 | 0.96 | 0.2 |
| 40 | 3.2 | 3.52 | 3.84 | 0.8 |
| 100 | 8 | 8.8 | 9.6 | 2 |

DWDM technology provides an efficient way of increasing network capacity. Whilst the capacity is increased by the number of DWDM channels, the common network infrastructure, including optical fibre, optical amplifiers, DCMs, DWDM multiplexer, and de-multiplexer, is shared by all the DWDM channels. Thus the network cost is also shared, resulting in lower cost per channel. As traffic demand increases, network capacity can grow by adding extra transponders (i.e. transmitters and receivers) at the terminating points of the traffic. In addition, DWDM technology is transparent to the data signals carried by the wavelength channels. Therefore, new data rates and modulation formats can be introduced to further increase the network capacity. For instance, when DWDM technology deployment started around mid 1990, the predominant data rate was 2.5 Gb/s. As technology developed, the data rate increased to 10 Gb/s and then to 40 Gb/s using various new modulation formats. More recently, with advances in coherent technology and digital signal processing, 100 Gb/s has become mature and widely available. It is now being rapidly deployed in the networks. Table 2.1 shows the total fibre capacity using DWDM in a 50 GHz grid. For commercial DWDM systems, the total fibre capacity has increased over the years from up to 0.24 Tb/s at a data rate of 2.5 Gb/s to a total of 9.6 Tb/s at the higher data rate of 100 Gb/s. It is worth mentioning that utilization of the optical fibre has also increased significantly as demonstrated by the improvement of the spectral efficiency from 0.05 to 2 bit/s/Hz when the data rate increases to 100 Gb/s. The spectral efficiency continues to improve as the data speed increases beyond 100 Gb/s, as well as when the network technology moves from fixed-grid DWDM to flexible grid networking, which will be covered in more detail in the following sections.

In addition to increasing network capacity as described above, DWDM technology has also evolved from point-to-point DWDM systems to DWDM

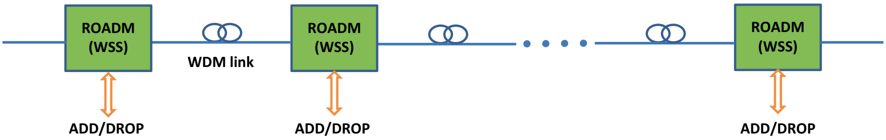



**Fig. 2.3** Schematic of a ROADM-based WDM network

networking with wavelength switching, initially with fixed wavelength Optical Add and Drop Multiplexers (OADM), and more recently with Reconfigurable OADM (ROADM), which can be configured remotely. Figure 2.3 illustrates a ROADM-based WDM network, where degree 2 ROADMs in WDM nodes route wavelength channels to different directions or to local ADD/DROP ports. Fixed-grid ROADMs using Wavelength Selective Switch (WSS) technology have now been widely deployed in WDM networks. ROADM nodes can have varying number of degrees, usually up to 8 degrees in the current WDM networks to enable meshes to be built. As bandwidth continues to increase and becomes more dynamic, ROADM-based WDM networks provide network operators more flexibility in adding new wavelengths or re-directing wavelengths, as well as restoration of traffic when a failure occurs. It also enables the monitoring, control, and management of power balancing among the WDM channels. All this can be carried out remotely via software or management configuration, resulting in significant operational cost reduction.

## 2.4 WSS Technology: Fixed and Flex

A key technology underpinning the emergence of flex-grid architectures is the Wavelength Selective Switch (WSS), with its wavelength ($\lambda$) switching and routing functionalities. This component typically has a single fibre input that contains many wavelength signals across the entire C band. Its job is to direct these wavelengths to any of a number of fibre outputs without any restrictions or blocking; any combination of the input wavelengths can be redirected to any of the fibre outputs. Such a component is already of critical importance for dynamic fixed-grid WDM networking, e.g. as found in a ROADM device. However, flex-grid requires the additional flexibility of variable passband, much finer spectral resolution, bandwidth control, and filtering. The huge majority of WDM technologies, e.g. Arrayed-Waveguide Gratings (AWGs) and Micro-Electromechanical System (MEMS) devices, while allowing important routing and switching functionalities, do not allow the possibility for variable passband widths. Rather, AWGs and MEMS devices are inherently fixed-grid WDM devices, with their fixed channel granularity (e.g. 50 GHz, 25 GHz, etc.) set at the time of their manufacture. In contrast, liquid crystal-based technologies have matured in recent years, and using the holographic principle can offer flexible passband filtering in conjunction with the wavelength switching and routing functionalities required for a flex-grid WSS. In this case, the key component within the WSS is a Liquid-Crystal-On-Silicon (LCoS) device, to enable active switching and beam steering of light.

The original concept for such holographic light switching was already first reported in 1995 [2], with the major functionalities and compact architectural design possibilities outlined by 1998 [3, 4], well before the advent of flex-grid as a WDM photonic networking design paradigm.

The specific functionalities that a flex-grid WSS based on the holographic LCoS principle include:

- Multiple wavelength routing and switching
- Power equalization of different wavelength channels
- Dispersion compensation/mitigation
- Variable channel bandwidths
- Polarization-independent operation
- Compactness
- Millisecond reconfiguration times.

Such a comprehensive set of features makes the LCoS-based WSS a highly attractive sub-system component, with considerable interest in it amongst the networking industry and academic communities. Indeed, the comprehensive set of light-conditioning functionalities offered by the holographic LCoS technology has led to it being referred to as a programmable optical processor [5], or even an optical Field Programmable Gate Array (FPGA). These are not unreasonable claims, since the theory underpinning the holographic principle is deeply embedded within Fourier Transform (FT) theory, and hence has strong analogies with Digital Signal Processing (DSP) theory, which itself relies on Discrete FTs (DFTs), and z-transforms. Hence, it is no surprise that the LCoS-based WSS can offer such a flexible and critical optical processing capability. It is interesting to note that the AWG also shares similar FT processing properties, since it can be considered to be the integrated, planar version of an LCoS WSS device, but of a reduced dimensionality and with a fixed wavelength spaced mode of operation [6].

Figure 2.4 indicates how the LCoS WSS device operates. Flex-grid WDM-multiplexed light from an input fibre is collimated by a lens and shone onto a fixed diffraction grating, which spatially spreads out the different colours (frequency components) of the composite signals so that they are imaged onto different locations of the LCoS device. This displays a hologram at each location of the separated light frequencies (colours). Each hologram is configured to act as a programmable diffraction grating, to selectively steer the light to the required output port. In so doing, for each particular colour, the hologram can adjust the power in the optical signal and (via its holographic phase characteristic) can also compensate or mitigate any dispersion degradation suffered by the signal.

The LCoS is a compact, pixelated device (utilising many of the technologies developed within the flat screen and displays sector, e.g. particularly micro-displays for virtual-reality helmets), where a liquid crystal matrix array is laid into an optically flat reflective surface, which in turn is located directly above a CMOS-based processing substrate, to allow individual electronic addressing of each pixel.

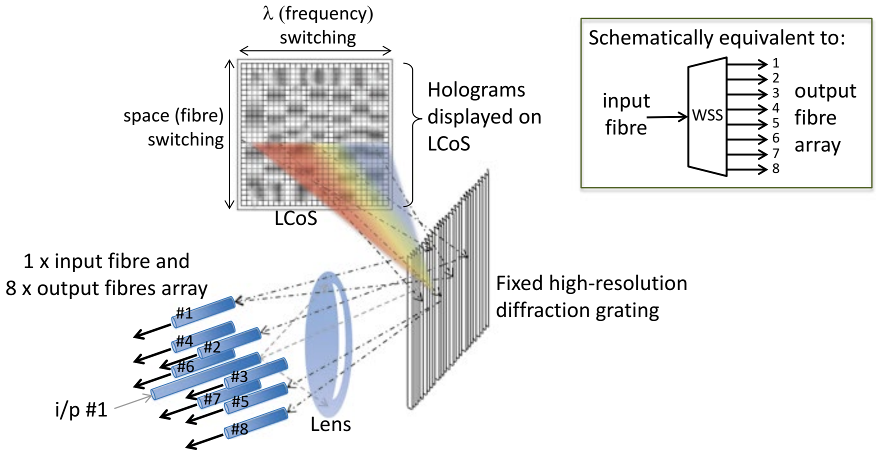



**Fig. 2.4** Schematic of Wavelength Selective Switch

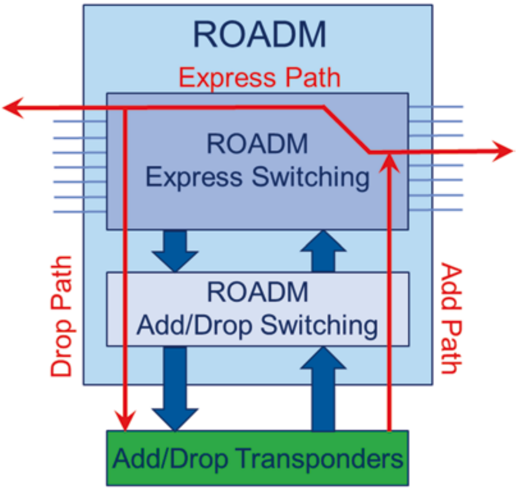
Since a flex-grid signal (i.e. a super-channel) contains a wide width or number of frequencies, when the super-channel is shone onto the LCoS device, it requires a wide width (i.e. a greater number of pixels) to be appropriately processed. This is where the flexibility of the holographic WSS switch becomes important, since for a varying flex-grid super-channel bandwidth, it is straightforward to dynamically assign more rows of the LCoS matrix channel as required, so as to be able to correctly steer the whole flex-grid super-channel to a desired output port, without any bandwidth limitations.

Typically, the number of output ports can range from 2 up to 20 or more. The holographic LCoS technology is not intrinsically limited in the number of output ports that can be addressed and routed to; i.e. the number of output ports can be expected to increase as the LCoS WSS technology continues to mature. Network requirements are determined by nodal degrees, connectivity (how meshed the photonic network is), number of fibres connecting node pairs, number of super-channels spanning the optical spectrum and their data bandwidths (e.g. 1 Tb/s, 400 Gb/s, 200 Gb/s, 100 Gb/s, etc.), as well as how the various super-channels are groomed as they transit through the network.

## 2.5 ROADM Architectures

A ROADM is a network element that allows for) dynamically adding or dropping of wavelengths at a network node. ROADM architectures are also able to switch DWDM wavelengths between the different express fibres. In the past, DWDM wavelengths were transmitted on a fixed 50 or 100 GHz bandwidth ITU grid. Flex-grid ROADMs have the additional advantage of being able to add and drop wavelengths with both fixed and variable channel optical bandwidths.

**Fig. 2.5** Express and add/drop functions of a ROADM




The two paths for traffic through a ROADM are shown in Fig. 2.5. The first is the express path for traffic that traverses the node and the other is the add/drop path for traffic that terminates or originates at the node.

ROADMs can be built using a variety of component technologies. Since ROADM architectures reflect advances in optical components, it is natural that flex-grid ROADM architecture development follows the development of flex-grid filtering components.

Some of the widely used components used today in both fixed and flex-grid ROADMs are:

- Wavelength Selective Switches (WSS)
- 1 × N and M × N All-Optical Switches (OXC)
- N × M Multicast Switches
- Optical Amplifiers (OA)
- Fixed and Tunable Filters
- Wave Blockers (WB)
- AWG Multiplexers
- Optical Splitters

Some of these components are intrinsically compatible with both fixed and flex-grid ROADMs, while others need to be adapted to work in flex-grid systems. Components, such as all-optical switches, splitters, circulators and optical amplifiers, are inherently flex-grid compatible since they are typically broadband devices that do not filter individual wavelengths. WSS devices do filter individual wavelengths but both fixed-grid and flex-grid versions are available.

M × N multicast switches combine multi-degree switching and filtering functions. They are used in the add/drop path to separate individual wavelengths from DWDM traffic on M fibres (M fibre degrees) and the individual wavelengths routed to N transponders. AWG filters are very popular mux/demux devices used to separate out the individual wavelengths on DWDM fibres. AWGs are inherently fixed-grid devices that are not compatible with flex-grid systems. While flex-grid all-optical wavelength multiplexers could be built, architects tend to prefer the flex-grid WSS that combines both filtering and switching in a single compact package.

Whilst WSS-based ROADM architectures have largely solved the problem of how to interchange wavelengths between different ROADM express fibres traversing a node, they have not resolved the issues with increasing add/drop flexibility. This flexibility has been addressed by Colourless Directionless Contentionless (C/D/C) ROADM designs:

- Colourless architectures allow any wavelength on an express fibre to be connected to any add/drop transponder associated with that fibre. In a colourless architecture the add/drop wavelengths on a single fibre share a group of transponders associated with that fibre.
- Colourless and Directionless ROADM architectures extend this concept to sharing a single group of transponders among all wavelengths from all express fibre directions.
- The word "Contentionless" was added to the definition because many proposed C/D architectures have some colour blocking and a way was needed to distinguish truly nonblocking from blocking architectures.

Network operators clearly want Colourless add/drop architectures to take full advantage of tunable transponders and improve network availability. They would also like Directionless so that spare transponders can be shared among all the waves in the node, and allowing advanced protection features. Currently, most of the proposed C/D/C networks involve using some combination of fixed or flexgrid WSSs and OXC (fibre switching) technologies. Optical switching technology has matured over the last few years and costs have dropped considerably—switches with optical losses of 1–2 dB are now commercially available.

## 2.6 Performance of Fixed and Flex-Grid Networks

Network operators have always investigated efficient ways to upgrade their deployed networks (e.g. by introducing ROADM), based on the fixed 50 GHz grid, and be able to transport the ever-increasing traffic. Until 2010 most of the operator networks multiplex their traffic over different wavelengths modulated at a fixed data rate. With each wavelength able to support higher and higher data rates (now 100 Gb/s) and with service connections requiring only a fraction of an optical channel capacity, the aggregation of several low-rate services into the high data rate optical circuit (traffic grooming) becomes important when designing the network to avoid over-dimensioning the channel capacity [7]. Meanwhile, grooming operations are costly and energy-greedy because they are performed by opto-electronic devices and routers; indeed router prices quadratically increase with their size [8], and the finer their grooming granularity is, the more energy they consume and the higher their price becomes [9, 10].

Conversely, recent improvements in transmission techniques and devices (such as coherent detection) and WSS at affordable prices make it possible to avoid systematic opto-electronic (OE) conversions at each intermediary node. An optical network enabling optical by-pass of intermediary cross-connects is called *transparent*, contrarily to *opaque* networks where systematic OE conversions are performed. The best compromise in terms of cost and energy consumption is obtained with the optical by-pass [10].

Although high data rate interfaces are becoming very attractive for network operators because of their huge capacity, one also has to consider that their transparent reach reduces with the increase of data rate, and this may result in a significant increase of required OE conversions [11]. Furthermore, the efficiency of channel filling decreases with the utilisation of huge pipes, especially if transparent networks are considered [12]. To improve the filling of optical channels and effectively upgrade network capacity without dramatically increasing the network cost per transported bit, several studies, both academic and industrial, have appeared in the literature, demonstrating the inefficiency of carrying such traffic with optoelectronic equipment operating at a single data rate and concerning the migration of optical networks towards infrastructures supporting different data rates [11–14], which optimizes the cost of ownership for a network.

In [12], the use of different data rates to cope with the increase of traffic in a nation-wide network is demonstrated; the choice of a specific data rate associated with a new connection considers both its capacity and the distance it has to cover, proving that the mix of data rates yields a cost-efficient solution and efficient resource utilization. In [11], the design of an optical network using three different modulation formats, each adapted to a given data rate (10, 40 and 100 Gb/s) is compared to a design of a network having a single data rate or multiple data rates but with only one modulation format. The mix of data rates and modulation formats enables an average reduction of the network cost up to 15 %. Moreover, in [15], it is shown that the mix of data rates also reduces the energy consumption of an optical transport network from 10 to 77 %, depending on the considered fixed data rate.

The migration of an optical network from single to multiple data rates is also studied in [14], where a routing and grooming strategy minimizing overall capital expenditure (CAPEX) of network migration is proposed and tested both for a national- and continental-wide network.

But, if using different types of transceivers for different data rate maximises the performance (typically the transparent reach) associated to each data rate, as mentioned before, this solution leaves little range for reconfiguration once the equipment has been deployed, thus requiring accurate forecast of the traffic evolution to efficiently handle network evolutions [14].

Indeed, networks mixing 10 and 100 Gb/s channels are up-grades of historical direct-detection networks supporting 2.5 and 10 Gb/s. In such networks, chromatic dispersion compensating fibres are deployed in each amplifier site so as to compensate a specific amount of chromatic dispersion following a certain chromatic management map maximizing the transmission distances (usually of 10 Gb/s channels). In these systems, 100 Gb/s signals are highly penalized in terms of transmission because of the deleterious effect of 10 Gb/s intensity-modulated channels on phase-modulated 100 Gb/s channels [16]. Moreover, dispersion-managed networks are not compatible with the deployment of channels carrying capacity higher than 100 Gb/s because of phase modulation, whose performance improves in dispersion unmanaged networks; indeed in dispersion unmanaged systems it is possible to slightly lower the noise figure of optical amplifiers and improve the resistance to nonlinearities [17].

To face this inconvenience which often is translated into inefficient configurations for future network evolution, an alternative option has been proposed. This relies on the creation of "elastic" optical networks in which a single type of transceiver can operate at various data rates according to the characteristics (capacity and signal degradation) of each connection providing similar benefits as mixed-rate solutions but with increased flexibility. In general, an elastic optical network can be based on Orthogonal Frequency-Division Multiplexing (OFDM) technologies [18, 19] or on 100 Gb/s Polarization-Division-Multiplexed Quaternary Phase Shift Keying (PDM-QPSK) transceivers adapted for modulation-format versatility, lowering its data rates [20].

This first concept of elastic networks is compliant with legacy systems (particularly ROADM) based on 50 GHz grid [19, 20]. Such elastic optical network architectures with flexible data rate and spectrum allocation ensure high resource efficiency, low cost, and low power consumption. The main advantage of these systems relies on the intrinsic rate tunability of an adaptive data rate interface, which increases the sharing of spare resources in restorable networks. This reduces the whole number of interfaces compared to fixed rate interfaces [21, 22] allowing one to reduce their number by up to 70 % [21]. The cost benefits of this interface sharing increase rapidly as the traffic load goes up, reaching 37 % [23]. The use of elastic interfaces not only guarantees interfaces sharing but also spectrum sharing. Indeed to mitigate the deleterious effect of intensity-modulated channels on phase-modulated 100 Gb/s channels [16], the optical comb can be partitioned into different bands, although this can have a negative impact on spectrum availability.

Elastic optical networks are up to 17.5 % more cost-efficient than single- and Mixed-Line Rate (MLR) architectures when the need for network upgrade is taken into account [24]. The built-in flexibility of bandwidth-variable transponders (BVT) allows a better adaptation than MLR to uncertainties in the growth of the traffic matrix, which more than offsets the initial higher CAPEX requirements.

Another advantage of BVTs is the possibility of adapting the power consumption of the network to the traffic daily fluctuation. Data rate adaptation can be performed by adapting the format modulation and symbol rate, separately or jointly; thanks to these rate adaptations, up to 30 % of energy savings are estimated [25, 26] when unprotected and protected networks are taken into account.

Until now, we have described the advantages of elastic networks compliant with the fixed 50-GHz grid; let us now consider the impact of flex-grid. The first spectrum adaptive architecture was proposed in [27, 28], referred to as SLICE, and based on an OFDM-based elastic devices. An optical channel is composed of a variable number of multiplexed sub-carriers; each sub-carrier has a symbol rate and modulation format that depends on the distance to be covered and transmitted rate. This concept allows the use of just enough spectrum (sub-carriers) resulting in high spectrum efficiency [29]. It relies on BVTs at the network edge and flex-grid ROADMs in the network core.

In the OFDM-based elastic optical network) architecture, multiple data rate sub-wavelength or super-wavelength paths are realized through flexible granular grooming and switching in the spectrum domain, using BVTs and bandwidth-variable wavelength cross-connects (WXCs). The BVT allocates just enough spectrum (sub-carriers) to accommodate sub-wavelength traffic, known as spectrum slicing [27]. The fundamental principle of this type of elastic network is the possibility to fit a wide range of spectrum efficiencies. Several OFDM channels can be merged together into a super-channel, transporting a multiple of the capacity of an individual OFDM channel without spectrum guard bands in between. This OFDM-based architecture supports multiple-rate accommodation (from Gb/s to Tb/s), provides higher spectrum efficiency through flexible spectrum allocation according to the transmitted data rate (up to 95 % spectrum improvement compared to a fixed-grid WDM network [30]), and enables optical network virtualization with the virtual link supported by OFDM sub-carriers.

In order to achieve high spectral resource utilization, the BVT needs to generate an optical signal using just enough spectral resources, according to the client data rate and the channel conditions. To create an OFDM channel, the usual scheme is to create a channel which adjusts the number of its sub-carriers and their rate so as to cope with the service requirements. The control of the sub-carrier number can be performed in the optical or digital domain, depending on what kind of signal synthesis method is used.

Unfortunately, filtering functions are not ideal and the tighter the filtering granularities of a ROADM, the higher the filtering penalties [31] suggesting the use of a guard-band separating OFDM channels in order to minimize the penalties due to both filtering functions and the channel cross-talk. The higher the guard-band value, the lower the advantages of flex-grid networks with respect to fixed-grid [32]. To increase the OFDM channel capacity, several OFDM channels can be merged together into a super-channel, where they are routed as a single entity. A super-channel provides a more efficient optical layer for aggregated layer 2 demands. In [33] seven optical paths are optically aggregated into a single spectrally-continuous super-channel with a bandwidth of 1 Tb/s; while in [34], a 420 Gb/s distance-adaptive super-channel with DP-QPSK and 8-APSK modulation was also demonstrated.

The super-channel concept where various carriers are steered together in a single unit was introduced in [34]. Here a super-channel is not based on OFDM channels, but on Nyquist channels (also called sub carriers). In that kind of super-channel, the super-channel capacity is adjusted by varying the number of subcarriers and their data rate. The spectral efficiency of the super-channel will depend on: the rate associated to each sub-carrier the spacing between the sub-carriers [35, 36] and the distance that has to be covered by the super-channel; the smaller it is, the narrower spacing that can be adopted [37, 38].

The concept of super-channels not only relates on the improvement of the spectrum efficiency of a network, but is also very promising for the realization of optical transmission beyond 100 Gb/s. Indeed, the usual trend of increasing the rate within the traditional 50 GHz grid for long-haul transmission (currently possible by adopting higher modulation formats) is no longer viable due to Signal-to-Noise Ratio (SNR) requirements for higher-modulation formats. Such channels can be realized by adopting multi-carrier channels (a particular case of the super-channel concept), with or without polarization division multiplexing, for guaranteeing a high spectral efficiency and optical reaches compatible with core networks [34, 39].

## 2.7 Migration to Flex-Grid

Whilst all the concepts of flex-grid technology appear to add some benefit, it is very unlikely that they all will be required across the entire network on day one. In fact, recent studies show that expected traffic volumes for short and medium term (few Tb/s of total traffic) are not enough to justify immediate flex-grid deployment. For example, authors in [40] estimate that capacity exhaustion in the current fixed-grid network will not occur before 2019. Notwithstanding, as also concluded in [40], the current spectrum capacity operated with flex-grid ROADMs could extend the network lifetime.

The most likely scenario is that some links in the network grow more quickly than others and become congested, acting as bottle necks to future network growth. It is in those links where flex-grid can be first deployed to extend network lifetime. Hence, a migration strategy is required so that new technology can be introduced where it has most benefit, without having to start again with a new network build. Unfortunately, such a migration has been unfeasible in the past, mainly due to the lack of interoperability between old and new network technologies. However, this is different between fixed-grid and flex-grid technologies partly due to the potential of coherent transmission and to flex-grid WSSs.

As traffic grows, some of the core network links will become congested and critical questions arise: is it worth simply upgrading the bottleneck links to be flex-grid enabled? Would a parallel fibre be a better option for these links? If the bottleneck links are not connected, does this reduce the benefits due to the inability to have a flex-grid path between them? And therefore, does it make more sense to introduce islands of flex-grid capability, and if so, where would those islands be, how large should they be to be of benefit, and at what stage should they be introduced?

Many operators structure their core networks in single-vendor sub-domains and in some cases optical transparency is not a requirement between such sub-domains. In such scenarios, each sub-domain can have its own migration strategy. However, when transparency is needed for some long-haul connections passing through domains with different technologies, the spectrum adaptability in flex-grid islands poses no additional constraints to the more restrictive spectrum allocation in fixed-grid domains.

The latest installed fixed-grid OXCs are already equipped with LCoS-based WSSs, which allow switching media channels of arbitrary size facilitating thus the migration towards flex-grid. Migrating fixed-grid OXCs equipped with LCoS-based

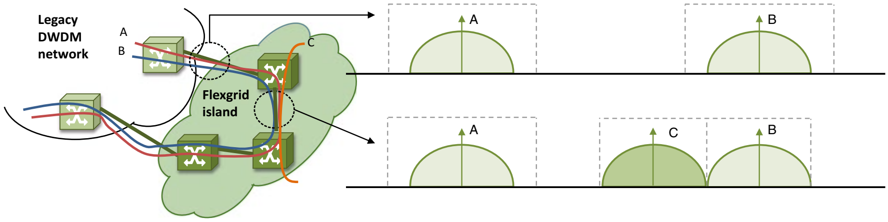



**Fig. 2.6** Fixed and flexgrid inter-operability

WSSs can be done, in general, by upgrading its software, resulting in a low cost migration. Older OXCs however, are equipped with MEMS-based WSSs. Many of those OXCs can be easily migrated to flex-grid by replacing WSSs or even having coexistence in the network with flex-grid-ready OXCs.

Therefore, looking at the optical spectrum, a mix of optical connections exists as a consequence of the creation of flex-grid islands. Figure 2.6 shows an example where 3 DP-QPSK 100 Gb/s optical connections are routed. While connection A is transported transparently through the flex-grid island using a 50 GHz media channel, the filter of connection B is shrunk to 37.5 GHz thus adjusting to the signal bandwidth. This allows connection C to use a portion of the fixed media channel allocated for connection B, which is possible since C will be dropped inside the flex-grid island.

Notwithstanding, capacity exhaustion is not the sole motivation behind migration. There are other drivers that justify the gradual deployment of flex-grid technology:

- In short term (2014–2016), the use of higher bit rates and advanced modulation formats for specific connections will allow cost-effective 400 Gb/s (and beyond) signals.
- In medium term (2017–2019), the advent of commercial Sliceable Bit Rate Variable Transponders (SBVT) will be a crucial event to increase the reach of flex-grid areas to those parts where, although spectrum will not be exhausted yet, the capability of splitting multiple flows will be beneficial. SBVTs are BVTs that have a level of integration which includes multiple modulators: these are therefore capable of creating either large super-channels or smaller individual channels as required.
- In the long term (>2020), capacity exhaustion as a result of dealing with expected traffic volumes of hundreds Tb/s or even some Pb/s will definitively require deploying flex-grid in the network possibly with different network architectures. Legacy fixed-grid equipment would then be completely upgraded to flexgrid in the core.

For all the aforementioned, migration to flex-grid technology as a set of gradual consecutive upgrading steps cannot be planned at the starting time because of several aspects, such as traffic uncertainty, which make it impossible to compute precise solutions for the future. As a result, solving each migration step taking as input precise data for the next period seems to be the most practical way to deal with the migration problem.

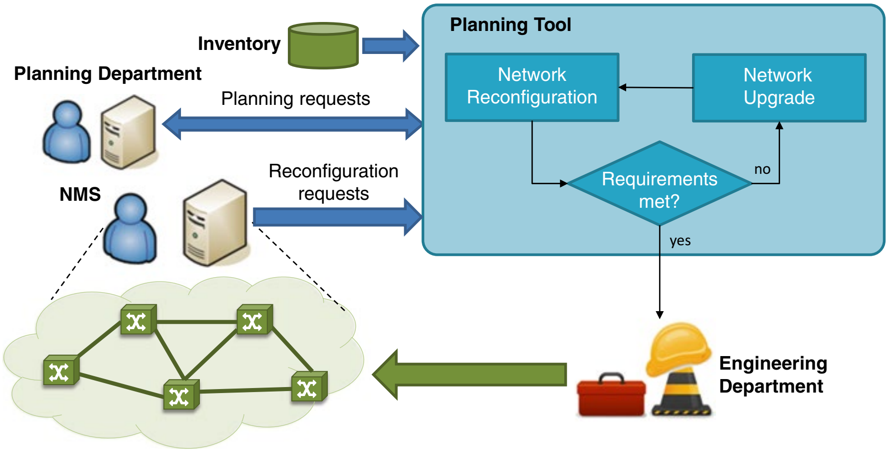



**Fig. 2.7** Migration flow chart

Figure 2.7 shows the migration flow chart considered for performing gradual network planning, where the following list of actors involved in the process is assumed: (1) the Network Management System (NMS) managing the core network, implementing fault, configuration, administration, performance, and security (FCAPS) functions; (2) a Planning Department administrating the planning process, i.e. analyzing the network performance and finding bottle necks, receiving potential clients' needs, evaluating network extensions and new architecture, etc.; (3) an inventory database containing all equipment already installed in the network, regardless they are in operation or not; (4) an Engineering Department, performing actions related to equipment installation and set-up; (5) a network planning tool in charge of computing solutions for each migration step. Since several sub-problems related to network reconfiguration, planning, and dimensioning, among others, need to be solved, the use of an integrated, adaptable, and service-oriented planning tool allows obtaining globally optimized solutions in a practical way [41].

As a general approach, we consider that a migration step begins when the planning tool receives a request that can be originated in different systems responding to different reasons:

- Operators analyzing data gathered by the NMS detect )that a migration step can be attempted to improve the performance of the current network. For example, bottlenecks have been detected in some parts of the network and its current configuration will not be able to allocate expected traffic, so reconfiguration can be attempted. Note that these triggers arise asynchronously (i.e. without a predefined schedule).
- Planners request network replanning to serve new clients or cover new areas. Contrary to reconfigurations coming from NMS, planning requests can be better synchronized with other network departments, such as the engineering department.

The planning tool solves the migration problem in two phases. Firstly, the *network reconfiguration* aims at reconfiguring the existing network resources and served traffic to meet target requirements. The solution of this problem consists in a set of actions that can be done in the network without purchasing and installing new flex-grid equipment. Therefore, the aim of this process is to exploit the possibilities of the current available resources as much as possible before purchasing and installing new equipment.

Among others, some of the possible actions that could form a solution of this reconfiguration phase are: (1) modifying the physical intra-connectivity at central stations; (2) moving physical devices, such as BVTs, from one location to another in a different part of the network; and (3) setting up and tearing down optical connections. Additionally, already purchased and installed network resources not yet activated can be put in operation in this phase. Thus, the reconfiguration process should process inventory data to decide whether some of these resources must be activated or not. It is worth highlighting that to implement such reconfigurations, the engineering department needs to perform manual actions that need to be scheduled and, therefore, not immediately processed. In fact, these manual interventions are usually performed during low activity periods since they might require temporally cutting some services, and therefore the whole reconfiguration process might last several weeks.

In the case when network reconfiguration is not sufficient to fulfill all the requirements, the *network upgrading* process, the second phase of the migration problem, is started. Network upgrading involves several network planning and dimensioning sub-problems, such as migrating selected regions to flex-grid, enlarging the network to cover new areas, extending the core towards the borders, etc. Obviously, the overall objective is to find solutions minimizing the total cost of the migration, including purchasing, installing and configuring new equipment.

When a solution is found, the reconfiguration phase is invoked to guarantee that all the requirements can be met. If the solution is acceptable, it usually requires to be accepted by operators at the planning department, who then send it to the engineering department, who, in turn, organize and schedule the set of processes that will physically implement the solution in the network. Although this whole process may take several weeks or even months to be completed, a new migration request can be started as soon as a subset of the new equipment is installed to partially reconfigure the network.

Therefore, the gradual migration from fixed grid to flex-grid is not only a feasible process from the technological point of view but also represents the optimal path to upgrade core optical networks.

## 2.8 Metro/Core Network Architecture

National IP/MPLS networks typically receive client flows from access networks and perform flow aggregation and routing. The problem of designing IP/MPLS networks consists of finding the configurations of the whole set of routers and links so as to transport a given traffic matrix whilst minimizing CAPEX. To minimize the number of ports, a router hierarchy consisting of *metro* routers performing client flow aggregation and *transit* routers providing routing flexibility is typically created.

As a consequence of link lengths, national IP/MPLS networks have been designed on top of fixed-grid DWDM optical networks and thus the design problem has been typically addressed through a multi-layer IP/MPLS-over-optical approach where transit routers are placed alongside optical cross-connects [42]. Besides, multi-layer IP/MPLS-over-DWDM networks take advantage of grooming to achieve high spectrum efficiency, filling the gap between users' flows and wavelength channels' capacity.

With the advent of flex-grid technology, this classical multi-layer network architecture has been commonly assumed by operators and researchers as part of the inherited fixed-grid background. For example, authors in [43] analysed the CAPEX needed to deploy a multi-layer IP/MPLS-over-Flex-grid architecture as a function of the frequency slice width used in the flex-grid network. Using different traffic profiles, they showed that investments in optical equipment capable of operating under narrow slice widths (12.5 GHz or lower) are more appropriate when lightpaths' capacity is low, whereas 25 GHz work better when lightpaths' capacity increases.

However, the flex-grid technology featuring a finer granularity allows performing grooming also at the optical layer and hence, the aggregation level of the incoming flows can be reduced. This fundamental difference with respect to fixed-grid networks opens the possibility to consider different network designs that could be better suited to flex-grid. Since most of the network operators have a relatively small core network (typically 20–30 nodes) serving a much larger metro and/or aggregation network, the key question to be answered nowadays is: should operators keep to this design or should they look at much larger core networks to better exploit flex-grid technology capabilities?

In contrast to such hierarchical multi-layer network, alternative network architecture can be devised by flattening previous multi-layer approaches and advancing towards single layer networks consisting of a number of IP/MPLS metro areas connected through a flex-grid-based core network (as illustrated in Fig. 2.8). The CAPEX associated with this architecture strongly depends on the number and size of metro areas, since they define, among others, the number and size of metro routers needed per area, the number and size of core routers to interconnect metro areas with the core, and the amount and characteristics of required BVTs and/or SBVTs. Therefore, a network planning procedure aiming at finding a global optimal solution is required.

Given a national network with hundreds or thousands of central offices, a network optimization procedure consisting in a single step is clearly impossible. Authors in [44] proposed a two-step procedure consisting of partitioning the network into a set of IP/MPLS areas interconnected through a flex-grid-based core network. Each IP/MPLS area and the core network were afterwards designed independently. The partitioning problem aims at maximizing the total aggregated traffic to be conveyed by the optical core network. In this way, the internal traffic of the resulting areas is indirectly minimized. That is important since the cost per bit of

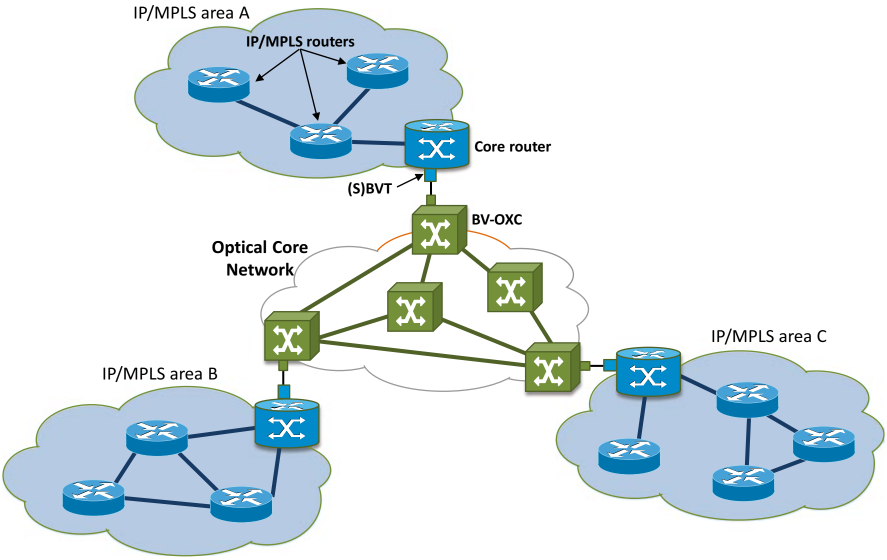



**Fig. 2.8** Flat metro/core network architecture

the optical switching technology is generally cheaper than that of the IP/MPLS and the aggregated flows in the optical core network occupy many spectral resources, so are most cost effectively groomed and routed optically.

Aiming at providing some figures regarding the proposed architecture, a numerical evaluation is presented from a close-to-real problem instance based on the BT network [44] consisting of 1113 locations. Those locations with a connectivity degree of 4 or above were selected as potential core locations (total amount of 323 locations). A traffic matrix was obtained by considering the number of residential and business premises in the proximity of each location. Locations could only be parented to a potential area if they were within a 100 km radius. Finally, three slice widths (50, 25 and 12.5 GHz) were considered.

The performance of several slice widths so as to obtain the highest spectral efficiency is investigated. Spectral efficiency is defined in Eq. (1), where $b_{aa'}$ represents the bit rate between locations $a$ and $a'$ belonging to two different areas in the set of areas $A$, $\Delta f$ is the considered slice width, and $B_{\text{mod}}$ is the spectral efficiency (b/s/Hz) of the chosen modulation format. Note that the ceiling operation computes the amount of slices to convey the requested data flow under the chosen slice width.

$$\sum_{a\in A}\sum_{\substack{a'\in A\\ a'\neq a}}\frac{b_{aa'}}{\Delta f\cdot B_{\text{mod}}}\Big/\sum_{a\in A}\sum_{\substack{a'\in A\\ a'\neq a}}\frac{b_{aa'}}{\Delta f\cdot B_{\text{mod}}} \tag{2.1}$$

In order to understand the behaviour of the core network spectral efficiency, Fig. 2.9 shows the number of core routers as a function of the spectral efficiency and the slice width. As illustrated, network spectral efficiency decreases sharply when the

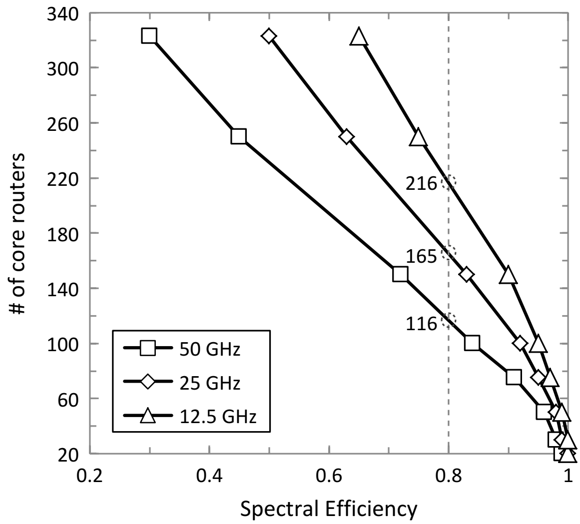



**Fig. 2.9** Spectral efficiency analysis

number of core routers (i.e. number of metro areas) is increased since more flows with lower amount of traffic are needed to be transported over the core network. Note that the traffic matrix to be transported by the core network has $|A| * (|A| - 1)$ unidirectional flows. Let us consider a threshold for spectral efficiency of 80 % (vertical dashed line in Fig. 2.9). Then, the largest number of areas is 116, 165, and 216 when the 50, 25, and 12.5 GHz slice width, respectively, is selected. Obviously, the coarser the grid granularity chosen for the optical network, the larger the areas have to be for the spectral efficiency threshold selected and thus the lower the number of core routers needed.

Table 2.2 focuses on the characteristics of those solutions in the defined spectral efficiency threshold for each slice width. Percentages in 25 and 12.5 GHz cells represent the relative difference with respect to 50 GHz values. As can be observed, the average number of locations in each area is lower than 10, being lower than 20 for the largest area. This limited area size simplifies the design of those IP/MPLS networks. In fact, using the 12.5 GHz grid allows reducing by 45 % the number of locations per area with respect to 50 GHz slice width.

Besides, the switching capacity of core routers is proportional to the size of the areas. They require capacities of up to 58 Tb/s (28.5 Tb/s on average) when the 50 GHz slice width is used, decreasing to 35.3 Tb/s (15.7 Tb/s on average) for 12.5 GHz. Hence, finer slice widths might keep routers to single chassis size—far more efficient and cost-effective.

The size of the area internal data flows is also detailed (flows where at least one end router is in a given area). In the reference interval, the size of the internal flows is up to 72 Gb/s, 32 Gb/s on average when the 50 GHz slice width is used, decreasing to 50 Gb/s, 21 Gb/s on average when the 12.5 GHz slice width is used. The average size of the areas traffic matrix is small as a result of the size of the areas: lower than 220 unidirectional flows.

**Table 2.2** Detailed results at spectral efficiency = 0.8

| | Slice width (GHz) | | |
|---|---|---|---|
| | 50 | 25 | 12.5 |
| # Core routers | 116 | 165 (42 %) | 216 (86 %) |
| Max metro area size | 19.66 | 15.04 (–23 %) | 15.04 (–23 %) |
| Avg metro area size | 9.94 | 6.97 (–30 %) | 5.46 (–45 %) |
| Max core router capacity (Tb/s) | 58.0 | 45.1 (–22 %) | 35.3 (–39 %) |
| Avg core router size (Tb/s) | 28.5 | 20.1 (–29 %) | 15.7 (–45 %) |
| Max area flows size (Gb/s) | 72.33 | 56.06 (–22 %) | 50.52 (–30 %) |
| Avg area flows size (Gb/s) | 32.68 | 23.84 (–27 %) | 21.12 (–35 %) |
| Max core flows size (Gb/s) | 989.06 | 602.16 (–39 %) | 404.51 (–59 %) |
| Avg core flows size (Gb/s) | 264.95 | 129.47 (–51 %) | 82.73 (–69 %) |

Eventually, the size of the aggregated data flows in the optical core network is examined. In the reference interval, the size of the aggregated flows is up to 989 Gb/s, 264 Gb/s on average when the 50 GHz slice width is used, decreasing to 404 Gb/s, 82 Gb/s on average when the 12.5 GHz slice width is used. Note that the size of the optical core network traffic matrix increases from 14,000 to 48,000 unidirectional flows (easily computed from the number of core routers) when the 50 and 12.5 GHz slice widths are selected. However, the mean size of the flows decreases until almost 70 % when moving to the finest spectrum granularity.

To summarize, results showed that simpler and smaller areas containing are enough to obtain good spectral efficiency in the flex-grid-base core network when 12.5 GHz slice width is used. The resulting grooming done in the flex-grid core network allows considerably reducing that done at the IP/MPLS one. In such scenarios, both the capacity and the number of IP/MPLS routers and ports can be reduced, being this the main reason to suggest network operators to reconsider their network design to gain the most from the change in technology from a network capacity and CAPEX basis.

## 2.9 Concluding Remarks

Transmission networks are changing rapidly with the onset of new technologies. Flex-grid allows the optical spectrum to be partitioned with much finer granularity and also arbitrarily wide spectral slots can be set-up and provisioned across a network. Flex rate allows the transceivers to function with different modulation formats to get close to the optimum spectral efficiency for given optical paths. Flex-grid requires novel LCoS technology to enable the fine optical filter control, whereas flexrate uses recent developments in coherent transmission to give the modulation format variations. Bringing these two concepts together allows large bandwidth pipes to be configured, in which multiple sub-channels are used together to form a super-channel, offering bandwidths in the Tb/s range. This is a radical new direction for core transmission networks, promising highly flexible future optical transport which makes full use of the available optical fibre spectrum resources.

This flexibility of optical connectivity will allow metro-core networks to become more flexible and potentially flatter in the sense that new direct connections between nodes will be possible as the network evolves and grows. Evolution from today's fixed-grid networks to more flexible networks of the future will require a migration strategy—perhaps starting with islands of flex-grid capability.

Core networks will ultimately evolve from static pipes of fixed optical spectrum, to dynamically flexible resources, capable of establishing arbitrary amounts of spectrum resources between nodes on demand, whilst carrying multiple Tb/s of data. These changes will be as radical as the original concept of DWDM and will enable our core networks to continue to meet exponentially growing bandwidth demands for the foreseeable future.

## References


1. ITU-T Rec. G.694.1, Spectral grids for WDM applications: DWDM frequency grid
2. S.T. Warr, M.C. Parker, R.J. Mears, Optically transparent digitally tunable wavelength filter. Electron. Lett. **31**(2), 129–130 (1995)
3. M.C. Parker, A.D. Cohen, R.J. Mears, Dynamic holographic spectral equalization for WDM. IEEE Photon. Technol. Lett. **9**(4), 529–531 (1997)
4. M.C. Parker, A.D. Cohen, R.J. Mears, Dynamic digital holographic wavelength filtering. IEEE J. Lightwave Technol. **16**(7), 1259–1270 (1998)
5. M.C. Parker, S.D. Walker, A. Yiptong, R.J. Mears, Applications of active AWGs in dynamic WDM networking and routing. IEEE J. Lightwave Technol. Spec. Issue Opt. Netw. **18**(12), 1749–1756 (2000)
6. K. Zhu, B. Mukherjee, Traffic grooming in an optical WDM mesh network. IEEE J. Sel. Areas Commun. **20**(1), 122–133 (2002)
7. M. Bertolini, O. Rocher, A. Bisson, P. Pecci, G. Bellotti, Benefits of OTN Switching Introduction in 100Gb/s Optical Transport Networks, in *Proceedings of OFC/NFOEC 2012*, Paper NM2F.2, March 2012
8. G. Shen, R.S. Tucker, Energy-minimized design for IP over WDM networks. J. Opt. Commun. Netw. **1**(1), 176–186 (2009)
9. A. Nag, M. Tornatore, B. Mukherejee, Optical network design with mixed line rates and multiple modulation formats. J. Lightwave Technol. **28**(4), 466–475 (2010)
10. G. Shen, R.S. Tucker, Sparse traffic grooming in translucent optical networks. J. Lightwave Technol. **27**(20), 4471–4479 (2009)
11. M. Batayneh, D.A. Schupke, M. Hoffmann, A. Kirstaedter, B. Mukherjee, Optical network design for a multiline-rate carrier-grade Ethernet under transmission-range constraints. IEEE/OSA J. Lightwave Technol. **26**(1), 121–130 (2008)
12. C. Meusburger, D.A. Schupke, A. Lord, Optimizing the migration of channels with higher bitrates. J. Lightwave Technol. **28**, 608–615 (2010)
13. P. Chowdhury, M. Tornatore, B. Mukherjee, On the energy efficiency of mixed-line-rate networks, in *Proceedings of OFC*, Paper OWY3, San Diego, March 2010
14. O. Bertran-Pardo J. Renaudier, G. Charlet, P. Tran, H. Mardoyan, M. Salsi, M. Bertolini, S. Bigo, Insertion of 100Gb/s coherent PDM-QPSK channels over legacy optical networks

relying on low chromatic dispersion fibres, in *Proceedings of Globecom'09*, Paper ONS.04.1, December 2009
15. G. Charlet, Coherent detection associated with digital signal processing for fiber optics communication. C. R. Phys. **9**, 1012–1030 (2008)
16. A. Bocoi et al., Reach-dependent capacity in optical networks enabled by OFDM, in *OFC'09*, San Diego, OMQ4, March 2009
17. A. Klekamp, O. Rival, A. Morea, R. Dischler, F. Buchali, Transparent WDM network with bitrate tunable optical OFDM transponders, in *Proceedings of OFC*, Paper NTuB5, San Diego, March 2010
18. O. Rival, A. Morea, Elastic optical networks with 25-100G format-versatile WDM transmission systems, in *Proceedings of OECC*, Paper, Sapporo, July 2010
19. A. Morea, O. Rival, Advantages of elasticity versus fixed data rate schemes for restorable optical networks, in *Proceedings of ECOC 2010*, Paper Th.10.F.5, September 2010
20. A. Klekamp, F. Buchali, R. Dischler, F. Ilchmann, Comparison of DWDM network topologies with bit-rate adaptive optical OFDM regarding restoration, in *Proceedings of ECOC 2010*, Paper P.5.05, September 2010
21. A. Morea, O. Rival, Efficiency gain from elastic optical networks, in *Proceedings of ACP*, November 2011
22. O. Rival, A. Morea, N. Brochier, H. Drid, E. Le Rouzic, Upgrading optical networks with elastic transponders, in *Proceedings of ECOC 2012*, Paper P5. 12, September 2012
23. A. Morea, O. Rival, N. Brochier, E. Le Rouzic, Datarate adaptation for night-time energy savings in core networks. J. Lightwave Technol. **31**(5), 779–785 (2013)
24. A. Morea, G. Rizzelli, M. Tornatore, On the energy and cost trade-off of different energy-aware network design strategies, in *Proceedings of OFC/NFOEC 2013*, Paper OM3A.4, March 2013
25. M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, S. Matsuoka, Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. IEEE Commun. Mag. **47**(11), 66–73 (2009)
26. M. Jinno, H. Takara, B. Kozicki, Concept and enabling technologies of spectrum-sliced elastic optical path network (SLICE), in *Proceedings, Communications and Photonics Conference and Exhibition (ACP),* Paper FO2, November 2009
27. K. Christodoulopoulos, I. Tomkos, E. Varvarigos, Spectrally/bitrate flexible optical network planning, in *Proceedings of ECOC 2010,* Paper We.8.D.3
28. B. Kozicki et al., Opt. Express **18**(21), 22105–22118 (2010)
29. A. Morea, O. Rival, A. Fen Chong, Impact of transparent network constraints on capacity gain of elastic channel spacing, in *Proceedings of OFC 2011*, Paper JWA.062, March 2011
30. B. Kozicki, H. Takara, Y. Tsukishima, T. Yoshimatsu, T. Kobayashi, K. Yonenaga, M. Jinno, Optical path aggregation for 1-Tb/s transmission in spectrum-sliced elastic optical path network. IEEE Photon. Technol. Lett. **22**(17), 1315–1317 (2010)
31. H. Takara, B. Kozicki, Y. Sone, T. Tanaka, A. Watanabe, A. Hirano, K. Yonenaga, M. Jinno, Distance-adaptive super-wavelength routing in elastic optical path network (SLICE) with optical OFDM, in *Proceedings, ECOC 2010*, Paper We.8.D.2
32. J.K. Fischer, S. Alreesh, R. Elschner, F. Frey, M. Nölle, C. Schubert, Bandwidth-variable transceivers based on 4D modulation formats for future flexible networks, in *Proceedings of ECOC'2013*, Paper Tu.3.C.1, London, September 2013
33. Q. Zhuge, X. Xu, M. Morsy-Osman, M. Chagnon, M. Qiu, D.V. Plant, Time domain hybrid QAM based rate-adaptive optical transmissions using high speed DACs, in *Proceedings of OFC/NFOEC'2013*, Paper OTh4E.6, Anaheim, Los Angeles, March 2013
34. G. Bosco, V. Curri, A. Carena, P. Poggiolini, F. Forghieri, On the performance of Nyquist-WDM Terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers. J. Lightwave Technol. **29**(1), 53–61 (2011)
35. A. Carena, V. Curri, G. Bosco, P. Poggiolini, F. Forghieri, Nyquist superchannels with elastic SD-FEC (32 Gbaud) subcarriers spectral spacing & constellation simulations on SMF. J. Lightwave Technol. **30**(10), 100–101 (2012)

36. R. Dischler, A. Klekamp, F. Buchali, W. Idler, E. Lach, A. Schippel, M. Schneiders, S. Vorbeck, R.-P. Braun, Transmission of 3x253-Gb/s OFDM-superchannels over 764 km field deployed single mode fibers, in *Proceedings of OFC 2010*, Paper PDPD2, San Diego March 2010
37. A. Mayoral, O. Gonzalez de Dios, V. López, J.P. Fernández-Palacios, Migration steps towards flexi-grid networks. J. Opt. Commun. Netw. **6**(11), 988–996 (2014)
38. M. Ruiz, A. Lord, D. Fonseca, M. Pióro, R. Wessäly, L. Velasco, J.P. Fernández-Palacios, Planning fixed to flexgrid gradual migration: drivers and open issues. IEEE Commun. Mag. **52**, 70–76 (2014)
39. M. Ruiz, O. Pedrola, L. Velasco, D. Careglio, J. Fernández-Palacios, G. Junyent, Survivable IP/MPLS-Over-WSON multilayer network optimization. IEEE/OSA J. Opt. Commun. Netw. **3**, 629–640 (2011)
40. O. Pedrola, A. Castro, L. Velasco, M. Ruiz, J.P. Fernández-Palacios, D. Careglio, CAPEX study for multilayer IP/MPLS over flexgrid optical network. IEEE/OSA J. Opt. Commun. Netw. **4**, 639–650 (2012)
41. L. Velasco, P. Wright, A. Lord, G. Junyent, Saving CAPEX by extending flexgrid-based core optical networks towards the edges (Invited Paper). IEEE/OSA J. Opt. Commun. Netw. **5**, A171–A183 (2013)
42. S. Frisken, H. Zhou, D. Abakoumov, G. Baxter, S. Poole, H. Ereifej, P. Hallemeier, High performance 'drop and continue' functionality in a wavelength selective switch, Paper PDP14, OFC'06, Anaheim March 2006
43. S.K. Korotky, Semi-empirical description and projection of internet traffic trends using a hyperbolic compound annual growth rate. Bell Labs Tech. J. **18**(3), 5–22 (2013)
44. J. Perelló, A. Morea, S. Spadaro, A. Pagès, S. Ricciardi, M. Gunkel, G. Junyent, Power consumption reduction through elastic data rate adaptation in survivable multi-layer optical networks. Photon. Netw. Commun. J. **28**(3), 276–286 (2014)

# Chapter 3
# Taking Advantage of Elastic Optical Networks

**Alexandros Stavdas, Chris Matrakidis, Matthias Gunkel, Adrian Asensio, Luis Velasco, Emmanouel Varvarigos, and Kostas Christodoulopoulos**

## Acronyms

| | |
|---|---|
| BV-OXCs | Bandwidth Variable Optical Cross Connect |
| BVT | Bandwidth Variable Transponder |
| CO | Coherent detection |
| DB | Database |
| DC | Data centre |
| EON | Elastic Optical Network |
| FRR | Fast Reroute |
| FTP | File transfer protocol |
| MLR | Multi-layer resilience |
| NRZ | Non-return-to-zero |
| OOK | On-off-keying |
| RML | Routing and Modulation Level |
| RSA | Routing and Spectrum Assignment |
| RWA | Routing and Wavelength Assignment |
| S-BVT | Sliceable BVT |
| SDN | Software Defined Network |
| SLA | Service Level Agreement |
| VM | Virtual Machine |

A. Stavdas (✉) • C. Matrakidis
University of Peloponnese, Tripoli, Greece
e-mail: astavdas@uop.gr

M. Gunkel
Deutsche Telekom, Bonn, Germany

A. Asensio • L. Velasco
Universitat Politècnica de Catalunya, Barcelona, Spain

E. Varvarigos • K. Christodoulopoulos
University of Patras, Rio Patras, Greece

 
V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_3

This chapter shows how the new paradigm of Elastic Optical Networks (EON) is changing the way we perceive networking. The mode of operation of EON is described and the potential gains are outlined, followed by Use-Cases outlining how flex-grid technology can be advantageous for multi-layer resiliency, for serving metro networks and finally for data centre interconnection.

## 3.1 Introduction

In this section, we summarize the features of EON that are opening up new directions in the way we perceive networking. Specifically, our goal is twofold: first to demystify EON's mode of operation and then to substantiate the usefulness of this platform.

### *3.1.1 Flex-Grid vs. EON*

Often in the literature the terms "flex-grid", "gridless" or EON are used interchangeably to designate systems (and the networks employing such systems) that are not complying with ITU-T's fixed-grid channel spacing. EON systems differentiate from the 50 GHz/100 GHz channel spacing, which was the outcome of two important technological constraints in legacy WDM systems: (a) the use of non-return-to-zero (NRZ) on-off-keying (OOK) as the primary modulation format, and (b) the difficulty to separate fine-spectral bandwidths over the entire C-band in the third attenuation window of the silica fibre. The deployment of NRZ OOK systems at line rates of 10 Gb/s and above is accompanied with considerable channel limitations (mainly dispersion and fibre nonlinearities) whilst it necessitates a channel spacing that is at least 2–3 times the channel's optical bandwidth, to ensure sufficient optical isolation for add/drop.

As elaborated in Chap. 2, EONs may overcome these limitations. However, as it can be observed from Fig. 3.1, it is very difficult to conceive a truly "gridless" system; one can observe that in EON the channel bandwidth may become tighter occupying less bandwidth and/or central frequency of a channel might be shifted (in a process similar to what was understood as "wavelength conversion in legacy WDM systems"), but there is always a fundamental pitch, associated to the limitations of the optical demultiplexing technology used, in the way we exploit the optical bandwidth. This "spectral pitch" is usually an integer fraction of the 100 GHz standard grid spacing e.g. 6.25 or 12.5 GHz. Moreover, the term "flex-grid" is less accurate in describing the essence of those systems since, as we show in Sect. 3.1.2, it is the optical bandwidth of the channel that varies, and this optical bandwidth is more efficiently exploited in EONs because of the narrow grids used (Fig. 3.1).

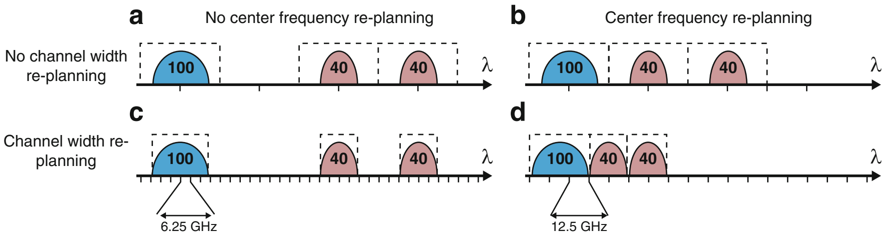



**Fig. 3.1** WDM under ITU-T grid and EON modes of operation

ITU-T REC G.694.1 [1] defines the following terms that will be used throughout this book.

- *Frequency grid*: A frequency grid is a reference set of frequencies used to denote allowed nominal central frequencies that may be used for defining applications.
- *Frequency slot*: The frequency range allocated to a slot and unavailable to other slots within a flexible grid. A frequency slot is defined by its nominal central frequency and its slot width.
- *Slot width*: The full width of a frequency slot in a flexible grid.

For the flexible DWDM grid, the allowed frequency slots have a nominal central frequency on a 6.25 GHz grid around the central frequency of 193.1 THz and a slot width in multiples of 12.5 GHz as seen in Fig. 3.1.

Any combination of frequency slots is allowed as long as no two frequency slots overlap.

### *3.1.2 Demystifying EONs*

To elaborate the properties of EONs, one needs to introduce the following quantities:

- *The baud rate*: it is the rate symbols which are generated at the source and, to a first approximation, equals to the electronic bandwidth of the transmission system. The baud rate is an important technology-dependent system performance parameter. This parameter defines the optical bandwidth of the transceiver and, following the discussion in Sect. 3.1.1, it specifies the minimum slot width required for the corresponding flow(s).
- *The modulation format*: For a given baud rate the modulation format defines the equivalent number of bits each symbol is transporting.
- *The line rate*: Effectively, it is the information-rate used for flow transportation between a source and a destination node (adjacent or remote). The dependence of the line rate on a number of sub-system parameters, mainly the nominal central frequency, the baud rate and the modulation format but also on whether or not there is polarization multiplexing and the type of the FEC used is schematically shown in Fig. 3.2a.

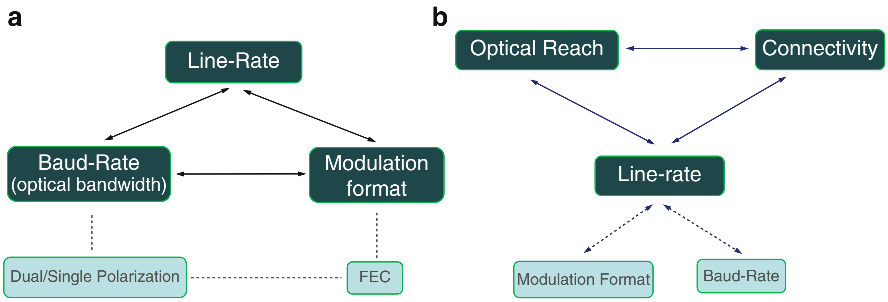



**Fig. 3.2** The interrelation of parameters in EON's (**a**) sub-system level and (**b**) system level

**Example:** Consider a system with a 50 GSymbols/s baud rate where the QPSK modulation format is used which corresponds to 2 bits/symbol. The equivalent information rate is 100 Gb/s and the electrical bandwidth of the system is 50 GHz. The two polarizations should be simultaneously modulated, the line rate per carrier would be 200 Gb/s although the actual information rate might be somewhat less, pending on the FEC used. Given the ITU-T REC G.694.1 [1] 12.5 GHz slot width granularity, the frequency slot can be as small as 50 GHz. In practice, the frequency slot may need to be larger, with additional bandwidth allocated at each side of the signal, to ensure sufficient optical isolation to facilitate add/drop and wavelength switching in intermediate nodes. The actual guard band needed depends on the impulse response of the Bandwidth Variable Optical Cross Connect (BV-OXCs) but for LCoS-based systems, 12.5 GHz per side is sufficient.

An important difference of EONs to legacy WDM/WSON networks is the following: Assume there is a necessity to transport a 1 Tb/s flow. Considering the previous example, a five-carrier system at 200 Gb/s line rate is needed. In this case, when multiple flows are generated and transported as a single entity (this is usually called *Superchannel*) from the source to the destination node, with no add/drop in intermediate nodes, the five-carrier system need not have intermediate spectral guard-bands to separate the corresponding sub-channels.

This is part of a general trend in EONs where connectivity is traded for (aggregate) line rate: with the aid of Fig. 3.3, one can observe that there are five carriers which in the previous example would transport 200 Gb/s each. In Fig. 3.3a these carriers are used to interconnect the source node A to four other nodes: from A to D by means of two independent carriers producing a total of 400 Gb/s, A to G with one carrier supporting 200 Gb/s, A to E with one carrier supporting 200 Gb/s, and A to F with one carrier supporting 200 Gb/s. On the other hand, as shown in Fig. 3.3b, the five carriers are used to interconnect node A to node C only, but this time with a superchannel at an aggregate rate of 1 Tb/s. In this latter case, no guard band is needed to separate the five carriers, while in the former case (Fig. 3.3a), the carrier directed to nodes F and the carriers directed to node D need to have lateral guard bands due to the possibility of being optically cross connected in the transit nodes E and B, C, respectively.

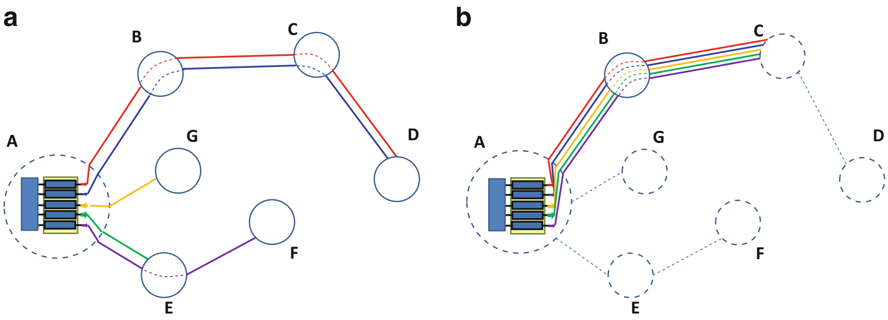



**Fig. 3.3** The trade-off between node connectivity and link capacity; (**a**) the connectivity degree is 4 and (**b**) the connectivity degree is 1

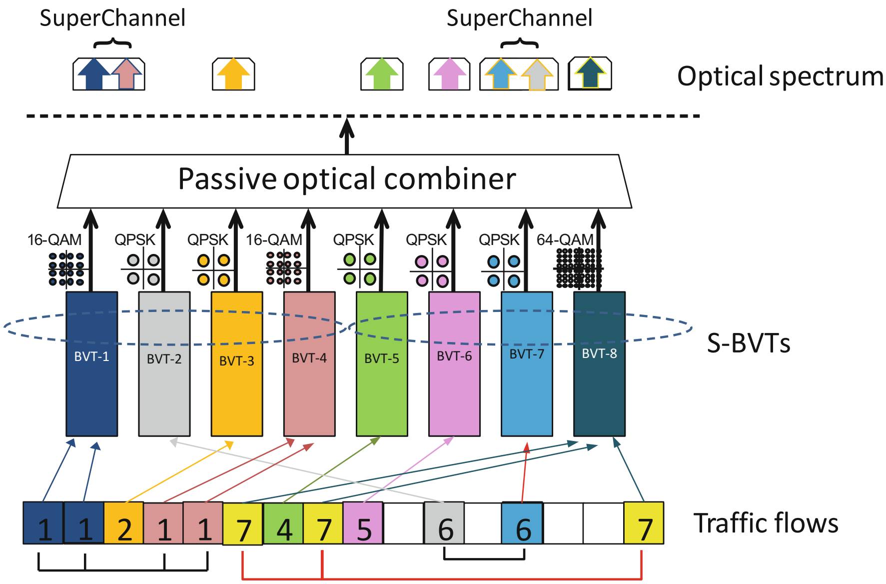



**Fig. 3.4** Example of node with 8 BVTs partitioned in two S-BVTs (Thanks to Emilio Riccardi)

Finally, as schematically illustrated in Fig. 3.2b, an EON is not only trading line rate for connectivity but also for optical reach: given a baud rate, a higher constellation modulation format increases the line rate but decreases the optical reach. The outcome is the same if the modulation format is fixed but the baud rate is varying, although the physical performance degradation is due to different causes in each case. Nevertheless, a variable baud rate system—also known as variable optical bandwidth—is more technologically challenging to engineer.

To illustrate the implications of these new features introduced with EONs, consider an EON node. As shown in Fig. 3.4, the client traffic of the node is forwarded to the BVTs, typically by means of OTN/MPLS framer/switch, but the number of flows assigned to them changes dynamically.

In our example, the node is equipped with two sliceable bandwidth variable transponders (S-BVTs) incorporating four bandwidth variable transponders (BVTs) each. There are flows of different capacity so these flows are decomposed, by means of the electronic sub-systems of the S-BVT (not shown in the figure), to a number of flows all at the same nominal rate. For example, there is a four-flow group designated as “1”, a three-flow group designated as “7”, a two-flow group designated as “6” and individual flows designated as “2”, “4” and “5”. The BVTs may support different modulation formats which typically are QPSK, 16-QAM and 64-QAM, as illustrated in Fig. 3.4. To this end, the following selections are made: the flows “2”, “4” and “5” are forwarded to the BVT-3, BVT-5 and BVT-6, respectively, using QPSK. The three “7”s are re-multiplexed constructing a larger capacity flow which is, then, forwarded to the BVT-8 and for this reason this BVT is now modulated with 64-QAM. Finally, four “1”s are multiplexed in two larger flows and they are directed to BVT-1 and BVT-4, respectively, which are modulated with 16-QAM using adjacent frequency slots and in this case a Superchannel is formed. The two-“6” flows are not multiplexed; instead they are directly forwarded to BVT-2 and BVT-7 employing QPSK which are forming up, again, a superchannel. Finally, there are four more potential input traffic flows (shown in white in Fig. 3.4) that are unused since there are no further BVTs available.

In this example, one identifies the following trade-offs: we have selected the flows “7” to be grouped together to be directed to a single BVT, freeing BVTs that can be used by other flows. At network level this is equivalent to adding resources that would be used to attain higher node connectivity. However, this is made feasible by employing a 64-QAM that significantly reduces the optical reach. If on the other hand one wishes to increase the optical reach, the three “7” flows should employ 16-QAM, but this is done at the expense of an additional BVT, decreasing the “connectivity resources” of the node by one degree (the BVT). To further increase the optical reach, flows “2”, “4” and “5” are employing QPSK but that decreases the connectivity degree of the node while using wider frequency slots.

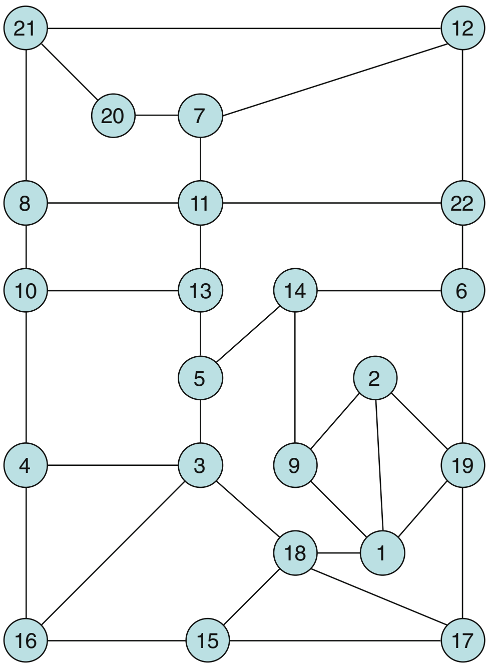
These examples illustrate the interrelations of the different parameters described in Fig. 3.2 as well as the emerging trade-offs between reach, connectivity and capacity in an EON.

Finally, to give some quantitative results, we compare two scenarios over the 22-node network shown in Fig. 3.5, with transparent end-to-end paths, with the fixed-grid and the EON case, respectively. The first scenario assumes a uniform traffic matrix, while the second one employs a traffic matrix generated using uniformly distributed random numbers.

The fixed-grid case has a channel spacing of 50 GHz and uses 100 Gb/s transceivers, while the EON case is the one described above, using only QPSK and 16-QAM modulations and polarization multiplexing (2000 km and 500 km reach, respectively) with 25 GHz guard bands between superchannels. Each link is assumed to comprise a single fibre pair where 4 THz of bandwidth can be allocated.

In this instance, the maximum capacity that can be transported over the network in each case was calculated, using shortest path routing and first fit wavelength/spectrum assignment [2]. The results are shown in Fig. 3.6. More sophisticated

**Fig. 3.5** 22-Node BT core network



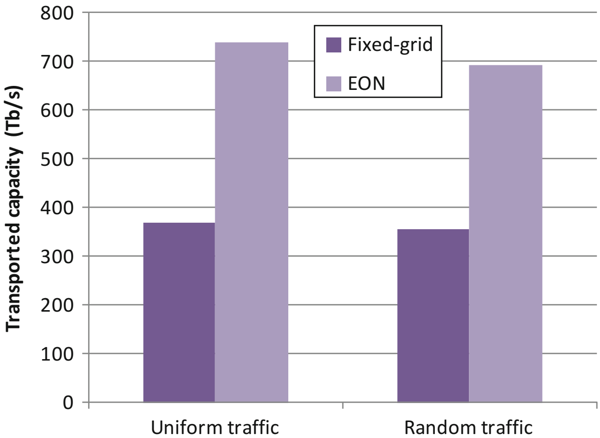




**Fig. 3.6** Maximum transported capacity obtained for each scenario for the fixed-grid and EON case over the 22-node BT core network

design methodologies will be presented in the next chapter while this gives a rough estimate of the potential gains.

Figure 3.6 shows that even with this simple approach, the EON effectively doubles the transported capacity for both traffic matrix scenarios, compared to the fixed-grid network case.

## 3.2 Advanced Networking Features Exploiting Flex-Grid

There are multiple application scenarios possible for leveraging EONs' capabilities in modern optical communication networks. More specifically, EONs' elasticity can be brought together with the idea of multilayer cooperation in reaction to a network failure.

### *3.2.1 Multi-layer Resilience in General*

Internet traffic in national backbones of European operators continues to grow with an annual rate of between 30 and 35 %. In this situation, operators are about to add new transport capacity connecting backbone routers over an agile DWDM layer. Today, the DWDM layer consists of fixed grid technology. In the future, say in 2–3 years, the technology of choice may be flex-grid based.

In traditional packet networks, it is the IP/MPLS layer which reacts to a failure, such as a fiber cut. A network failure on the optical layer may cause sudden, unavoidable and dynamic change over the transported traffic. Indeed, it is the most observed traffic dynamic in aggregation and backbone networks nowadays besides the normal constant and predictable traffic increase. At the high degree of traffic aggregation in these network hierarchies, unpredictable dynamics induced by failures happens much more frequently than dynamics induced by customers' behaviour (e.g. scheduled services).

Following the traditional approach of having reactive resilience exclusively over the packet layer might cause comparably bad utilization of both router interfaces and transponders that are equivalent to lambdas on the optical layer. The reason for this is the following: IP interfaces are loaded with traffic by adding up the packets which arrive randomly at the node. Forwarding the packets is subjected to a random process, too. Thus, multiplexing several flows together harvests a statistical efficiency gain. Though packet traffic is statistically multiplexed onto lambdas, those lambdas may be filled only up to 50 % in IP/MPLS networks relying on L3 recovery mechanisms only. The remaining 50 % capacity of a lambda is reserved for backup if a failure occurs. Without any dynamic countermeasure on the DWDM layer, optical robustness is conventionally assured by the creation of a second backup path $1+1$ disjoint from the first primary path. This comes along with a second transponder interface leading to more wavelengths and more links in the overall network. Directly related to these intrinsic inefficiencies are high capital expenditures inhibiting overall network profitability. Therefore, we can find that resilience mechanisms relying on the IP/MPLS layer exclusively are simple and widely adopted, but they are not cost-efficient.

The recently proposed converged transport network architecture (in the sense of an integration of packet layer devices, e.g. routers, and optical transport equipment, e.g. OXC or ROADMs), operators aim at building a network that has automated multi-layer resilience enabling considerably higher interface utilization. A key enabler for this new approach is the split of traffic into different service classes. For example, television, VoIP services, or financial transactions are critical applications that need to be delivered on time and with low packet loss ratio. Together with further services, they are pooled and assigned to high-priority traffic class. In contrast, a typical representative of best-effort traffic class is email or file transfer protocol (FTP) traffic which may experience some delay or packet loss.

A key idea behind multi-layer resilience (MLR) is to accept longer periods of packet losses for best-effort traffic than today. However, high-priority traffic maintains priority and is served in the same way as today. By treating different service classes unequally, capacity for best-effort class traffic is optically recovered after a considerable delay. Realistically, this recovery time is in the range between some seconds and up to a few minutes. This is huge when compared to IP resilience mechanisms like Fast Reroute (FRR) on one hand. On the other hand, it is very short when compared to the persistence of an optical cable cut or physical interface failure.

At the transmission layer, the MLR concept relies on agile photonic technology realized by multi-degree ROADMs with colourless and directionless ports (Fig. 3.7). That is where protection switching is related with the least cost when compared to packet layer protection switching. Several studies proved the feasibility and techno-economic superiority of this MLR concept under practical network scenarios (e.g. [3]). Currently, several operators are following this approach and roll-out MLR-based core networks. The general concept works fine already with conventional fixed-grid technology but EON can bring further improvements.

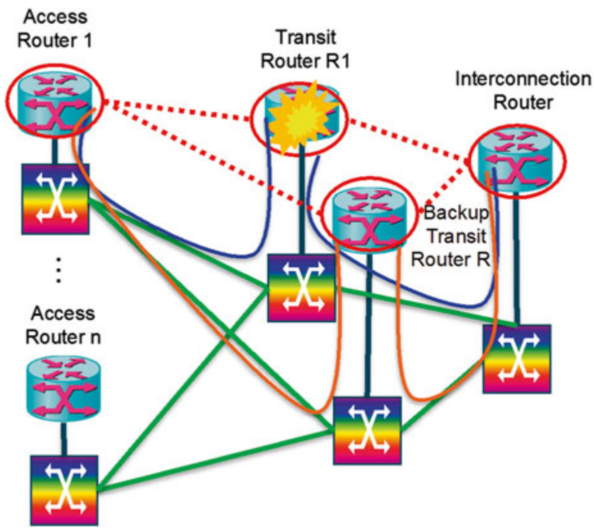



**Fig. 3.7** Multi-layer restoration using a remote backup transit router

### Case 1: Multi-layer Resilience Based on Elastic Flex-Rate Transceivers

As reaction to the failure, the conventional optical network tries to recover by finding a backup path that detours the non-functioning link. When routing the backup path through an optically rigid network, this new restoration path might exceed the reach specification of the previous light path. With a fixed and given modulation format the new light path might turn out to be infeasible. The consequence of that is the loss of the entire wavelength capacity (e.g. 100 Gbit/s). This is a quite big handicap for an operator and might induce further bottlenecks elsewhere in the network.

One promising application case of future software-configurable transceivers is to benefit from the reach-capacity trade-off (Fig. 3.2), to adjust the modulation format dependent on the reach requirement. Practically, this means a reduction of the original interface capacity.[1] Assuming the original capacity to be 400 Gbit/s realized by 2 × 200 Gbit/s with 16-QAM each, then the switch-over down to QPSK allows for a capacity of 2 × 100 Gbit/s instead of losing the entire interface capacity.

Thus, automatically lowering the existing line rate provides at least some part of the original capacity to be further exploited by IP routers accordingly. This type of photonic elasticity may save the operator a considerable amount of CapEx when compared to the case of a fixed-reach transceiver with its occasional necessity for intermediate signal regeneration. That is why flex-rate interfaces are considered as being of high importance to lower the network deployment cost.

Depending on the traffic volume and service classes, the operator may sometimes need the full capacity on the restoration path again. This can be accomplished by using an additional flex-rate BVT with again 2 × 100 Gb/s QPSK. But this additional interface will not be always needed, especially when the packet loss of best effort traffic is already beneath the acceptable threshold in accordance to the Service-Level Agreement (SLA). Therefore, after the failure repair, the additional restoring interface might be switched off or be used for other purposes.

Recently, a new approach was published [4] which leverages on the symbol-wise modulation adaptation in the time domain. With this approach, the considerable reach discrepancy between pure QPSK and pure 16-QAM (e.g. 2000 km and 500 km reach, respectively) can be closed by intermediate reach values. This is especially important for incumbent operators in smaller and medium countries where typical core distances fall just into the aforementioned reach gap. Techno-economic studies to quantify the benefit of this kind of flex-rate transceivers are on the way.

### Case 2: Transceiver Sliceability for Multi-layer Resilience

Due to the unpredictability of traffic arising from, for example, cloud services on one hand and the general burstiness of packet traffic on the other hand, there is always a significant amount of stranded capacity in today's networks which provide fixed-capacity

[1] Without further limitations (e.g. spectral scarcity), it is supposed here that in real network deployments optical interfaces run at their maximum capacity.

circuits between nodes. This leads to an under-utilization of interfaces on the optical layer. Especially in the early days of an optical transport network, the utilization of interfaces is inherently low.

In this situation, a high-capacity transponder (e.g. at 400 Gbit/s or 1 Tbit/s) being able to be logically and physically sliced into several virtual transponders targeting at different destinations with electronically adaptive bit rates may be highly beneficial for improving network economics. At day one, even a single S-BVT may provide enough total capacity for serving all destinations at a comparably low bit rate, say 100 Gbit/s [5]. Later when the traffic increases, it may serve only a few destinations each at a higher bit rate. Finally, it may support only a single huge flow to a single destination. All this is expected to become electronically controlled and adjustable.

The spectral slicing process is to be monitored and driven by the associated IP routers. They control the different service classes and assign them to spectral bandwidth slices enabling significant provisioning flexibility. In case of a failure in the optical layer, an adapted treatment of service classes is expected to be beneficial. An integrated control plane optimizes the assignment of high-priority and best-effort traffic unequally to optical subcarriers. In addition to parameters like path cost, latency, and shared risk groups, the optimization may also be accomplished dependent on various EON-specific parameters like availability of spectral slices or fibre fragmentation status.

Consequently, the next evolution step might be the investigation of the appropriateness of S-BVTs within a realistic MLR concept based on a service class differentiation.

For both application cases introduced above, the network architecture requires three main ingredients at least:

1. Firstly, an *agile DWDM layer* with colourless, directionless and flex-grid ROADMs providing faster service provisioning. Those photonic devices are also used for resilience switching at the physical layer (L0).
2. Secondly, an *integrated packet-optical control plane* offering a software-based flexibility including routing and spectrum assignment (RSA) instead of routing and wavelength assignment (RWA), signalling functionalities as well as elastic path selection in case of a network failure. An efficient integrated control plane solution also comprises just the right amount of information exchange between the packet and elastic optical layer.
3. Thirdly, a standard interface allowing configuring all IP routers using the same protocol. The advent of multi-layer control plane may ease the configuration of the MPLS and the GMPLS equipment through an UNI. However, IP layer services require configurations that go beyond control plane functionalities. After a failure, adding new routes or changing the metrics in the IP layer can help to optimize the IP topology. There are standardization efforts made, such as in the Internet Engineering Task Force (IETF), Network Configuration Protocol (NETCONF) and Interface to the Routing System (I2RS) working groups, but this is not yet supported by all router vendors.

## 3.3 Flex-Grid in Metro-Regional Networks: Serving Traffic to BRAS Servers

Metro architectures are currently composed of two main levels of aggregation: (1) The first level (named multitenant user or MTU level) collects traffic from the end users (optical line terminals, OLTs), while (2) the second level (named Access level) aggregates the traffic from the MTUs mainly through direct fiber connections (i.e. dark fiber). The IP functionality, namely traffic classification, routing, authentication, etc., is implemented in the Broadband Remote Access Servers (BRASs) that are usually located after the second level of aggregation.

In the last years, the main European network operators have been expanding their optical infrastructure to regional networks. As a consequence, it has been proposed to create a pool of BRASs located at, for example, two transit sites per region (the second site is for a redundancy purpose) as opposed to multiple regional BRAS locations. In this way, the needs for IP equipment and the associated CAPEX and OPEX costs would be reduced. These centralized servers can be *physical* or *virtual*, following the Network Function Virtualization (NFV) concept and instantiated at private or public data centres in the region. In this centralized scenario (Fig. 3.8a), the regional photonic mesh provides the transport capacity from the second aggregation level towards the remote BRASs. A representative scenario is Region-A of the Telefonica Spanish Network [6], which has 200 MTU switches, access level with 62 aggregation switches, and an optical transport network comprising 30 ROADMs that are connected to 2 BRAS sites over that network.

As flex-grid technology is considered an alternative to WDM [6, 7] for the core/backbone network, BRAS regional centralization appears to be an appropriate and interesting use case for examining the applicability of flex-grid technology in regional and metropolitan area networks (MAN). The examined scenario assumes the use of flex-grid transponders at the MTU switches and Sliceable Bandwidth Variable Transponders (S-BVT) at the BRAS servers. The aggregation level is thus removed, exploiting the finer spectrum granularity of flex-grid technology. This scenario is presented in Fig. 3.8b.

Since the metro network is quite different from the backbone network, the coherent (CO) transceiver technology that is used in backbone network applications might not be appropriate for the metro. It is worth examining alternatives to the CO technology in the metro, such as direct detection (DD) transponders.

We now investigate the detail of the BRAS centralization scenario. We assume a regional optical network consisting of ROADM nodes (WDM or flex-grid) and fiber links. The network supports the traffic generated by a set of MTU switches that are lower in the hierarchy and are connected to the regional optical network. Each MTU switch is connected to two different ROADMs for fault tolerance. The way an MTU switch is connected to the ROADMs changes according to the network scenarios (WDM or flex-grid), but since we focus now on the optical regional network, this does not play a role. In turn, the two ROADMs, the source and the chosen backup

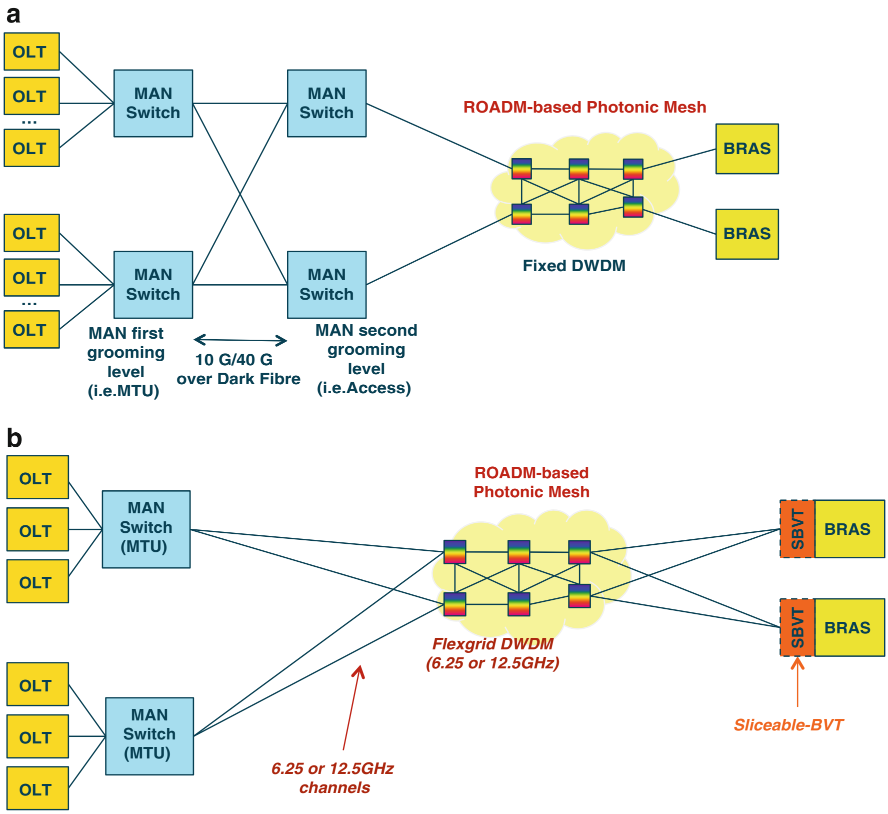



**Fig. 3.8** (**a**) WDM approach and (**b**) flex-grid approach to enable BRAS regional centralization

source, are connected via two light paths to two central BRAS nodes, and the chosen two paths are node-disjoint for protection purposes. There are two problem variations: (1) the location of the BRAS nodes is given and the goal is to minimize the maximum path length of all primary and backup paths; and (2) the maximum path reach is given (constrained by the transponders used) and the goal is to minimize the number of BRAS locations (which can be higher than 2). To solve the related problems, we developed optimal (ILP) and heuristic algorithms and we applied them to find the transmission reach requirements for the Region-A of Telefonica network [6].

We used our results as transmission requirements to develop an appropriate DD transponder model. The first attempt was an OFDM-based S-BVT presented in [8]. An improved and robust physical impairment approach with multiband (MB) OFDM was then proposed in [9], where the performance was numerically assessed according to the transmission requirements and experimentally validated within the 4-node ADRENALINE photonic mesh network. The proposed S-BVT is able to serve $N \times M$

MTUs: the array of modulators generates $N$ slices that can be directed towards different destinations and each BVT building block (each of the $N$ modulators) serves $M$ MTUs with a single optoelectronic front-end using only one laser source combined with simple and cost-effective DD. The number $M$ of MTUs served per MB-OFDM slice (corresponding to a single optical carrier) is limited by the bandwidth of DAC and optoelectronic components (10 GHz in [9]), and $N=8$ and $M=5$ are considered feasible with current technology (assuming components with, e.g. 20 GHz).

To compare the two network scenarios presented in Fig. 3.8, we performed a techno-economic analysis. In particular, the scenarios studied are: (1) based on WDM technology, and (2) based on flex-grid, with two variations: (2a) with coherent (CO), and (2b) with direct detection (DD) transponders. We calculate the cost of the first scenario (WDM) and use that as a reference to calculate the target costs of components used in the flex-grid scenarios to achieve 30 % cost savings. The cost calculations are based on the cost model presented in [8] (an extension of [6]).

*Scenario i—WDM solution*: The optical network consists of WDM ROADMs and the traffic from the MTU switches is aggregated at a second level of aggregation switches before it enters the WDM network. Taking the reference TID Region-A domain, we have 200 MTU switches and 62 aggregation switches. Each MTU switch forwards $C=10$ Gbps and is connected to two different aggregation switches for protection. Assuming equal distribution of load to the aggregation switches, each switch needs in total $14\times10$ Gbps ports (for down- and up-link). We also assume that the communication between the MTUs and aggregation switches is done with 10 Gbps grey short-reach transceivers and that the communication between the aggregation switches and the BRAS servers is done with 10 Gbps coloured transceivers (routed over the WDM regional network). In the cost calculation, we add the costs of: (1) the MTU switches' transceivers (grey short-reach); (2) the aggregation switches (140 Gbps capacity, grey transceivers facing the MTUs and colored transceivers facing the WDM network); and (3) transceivers (coloured) at the BRASs. According to our cost model, the total cost of the WDM case [8] is 219.82 ICU (IDEALIST Cost Unit [6, 10]).

*Scenario ii—Flex-grid solution*: We now assume that we replace the WDM network with a flex-grid network, and thus we include in the cost calculation the cost of the flex-grid wavelength selective switch (WSS). We examine two variations for the transceivers: (a) CO and (b) DD.

For case (a), we assume the use of 10 Gbps grey short-reach transceivers at the MTUs and coherent 100 Gbps flex-grid BVTs ($10\times10$ Gbps muxponders) at the add/drop ports of the flex-grid regional network, and coherent 400 Gbps S-BVT at the BRAS servers (functioning as $4\times100$ Gbps). We assume that each ROADM is equally loaded, so it serves 14 MTU switches. Thus, each ROADM utilizes two flex-grid BVTs. On the BRAS nodes, one 400 Gbps S-BVT serves two different ROADMs, which means that in total we need 15 S-BVTs. To obtain 30 % savings over the WDM scenario, the cost of 400G-SBVTs must be $\leq$0.01 ICU, which is impossible to achieve.

For case (b), we assume direct detected 10 Gbps MB-OFDM transceivers at the MTUs and DD MB-OFDM S-BVT at the BRAS servers (supporting $N=8$ flows, each flow serving $M=5$ MTUs at 10 Gbps). As above, we will assume that each ROADM is equally loaded, so it serves 14 MTUs. Thus, each ROADM needs three flows of the DD S-BVT, and in total we need 10 DD S-BVTs. To achieve cost savings of 30 % over the WDM network, the cost of 400G-S-BVT-DD must be ≤2.9 ICU, which is quite promising, since a same rate coherent S-BVT is expected to cost around 3 ICU in 2015 [6], and the cost of the DD S-BVT would be much lower.

Concluding, our comparison between a WDM and a flex-grid MAN network, showed that CO flex-grid transponders envisioned for core networks are expensive, while direct DD transponders seem a viable solution for the application of flex-grid solution in the MAN.

## 3.4 Multi-layer Network Planning for Cost and Energy Minimization

The increased flexibility of EONs can be exploited not only to save on spectrum but also on the cost and the energy consumption of the network. Currently, the optical network and the IP routers at its edges are not typically dimensioned with consideration of both technology domains, but this practice seems very inefficient for EONs. EONs envision the use of tunable (bandwidth variable) transponders (BVTs) that have a number of configuration options, resulting in different transmission parameters such as the rate, the spectrum, and the feasible reach of the light paths. Thus, decisions taken at the optical network often affect the dimensioning of the IP routers at its edges. This comes in contrast to the traditional fixed-grid WDM systems where the interdependence between the two layers was weaker and less dynamic. In EONs, we could still dimension both layers independently, for example sequentially, but there is vast space for optimization and gains if both layers are planned jointly. These gains come in addition to those obtained in multi-layer network operation such as restoration.

To give a closer look to the multi-layer planning problem, we model the network as follows. We assume that the optical network consists of ROADMs employing the flexgrid technology, connected with single or multiple fibres. At each optical switch, zero, one or more IP/MPLS routers are connected, and these routers comprise the edges of the optical domain. Short-reach transceivers are plugged to the IP/MPLS routers leading to flexible (tunable) transponders at the ROADMs. A transponder is used to transform the electrical packets transmitted from the IP router to the optical domain. We assume that a number of transmission parameters of the flexible transponders can be controlled, affecting their rate, spectrum and optical reach at which they can transmit. At the destination of a light path, the packets are converted back to electrical signal and are forwarded and handled by the corresponding IP/MPLS router. This can be: (1) the final destination, in which case traffic is further forwarded to lower hierarchy networks towards its final destination; or (2) an intermediate IP hop, in which case traffic will re-enter the optical network to be eventually forwarded to its domain destination. Note that connections are bidirectional and thus in the above description opposite directed light paths are also installed, and the transponders used act simultaneously as transmitters and receivers.

Following the cost model in [6] and related datasheets (Cisco CRS-3 [11]), we model an IP/MPLS router as a modular device, built-in (single or multi) chassis. A chassis provides a specified number of slots (e.g. 16) with a nominal transmission speed (e.g. 400 Gbps). Into each slot, a line card of the corresponding speed can be installed, and each line card provides a specified number of ports at a specified speed (e.g. 10 ports of 40 Gbps). A reference router model presented in [6] has 16 slots per line card chassis, and several types of chassis (fabric card, fabric card chassis) that can be used to interconnect up to 72 line card chassis in total.

Following the above model, the multi-layer network planning problem consists of problems in two layers: (1) the IP routing (IPR), (2) the Routing and Modulation Level (RML), and (3) the Spectrum Allocation (SA). In the IPR problem, we decide on the modules to install at the IP/MPLS routers, how to map traffic onto the light paths (optical connections), and which intermediate IP/MPLS routers will be used to reach the domain destination. In the RML problem, we decide how to route the light paths and also select the transmission configurations of the flexible transponders to be used. In the SA, we allocate spectrum slots to the light paths, avoiding slot overlapping (assigning the same slot to more than one light path) and ensuring that each light path utilizes the same spectrum segment (spectrum slots) throughout its path (spectrum continuity constraint). A variation of the IPR problem is also referred to as traffic grooming, while the RMLSA problem is also referred to as Distance-adaptive RSA. As discussed previously, distance adaptivity creates interdependencies between the routing at the optical (RML) and the IP layers (IPR) and the spectrum allocation (SA), making it hard to decouple these sub-problems, unless we can afford to sacrifice the efficiency and pay a higher CAPEX and OPEX.

Several algorithms have been developed to solve variations of the multi-layer planning problem, described above, that include both optimal and heuristic algorithms [12–14]. The objective is to minimize the CAPEX cost, while some algorithms also consider the spectrum used and/or the energy consumption [15].

Findings in [12] show that we can obtain saving of about 10 % for reference year 2020 for a national (Deutsche Telecom, DT) and a pan-European (GEANT) network, when comparing an EON to a WDM network. Note that these savings were calculated for conservative scenarios that assume a 30 % higher cost for the flexible (BVT) transponders over the cost of the fixed transponders of equal maximum transmission rate (400 Gbps). Planning the multi-layer network as a whole, as opposed to planning the IP and the optical network sequentially, can yield cost savings of 15 % for topologies that exhibit wide variation in path lengths, such as in GEANT, while the savings are less for national networks, where the majority of connections are established with similar transponders configuration parameters and the increased transmission options of BVTs are not utilized. The percentage of energy savings that was observed in [15] was 25 %, higher than the related savings in cost, since in this case we assumed similar reference energy consumption levels for BVTs and the fixed transponders.

## 3.5 Interconnecting Data Centres

### *3.5.1 Motivation*

As introduced above, transport networks are currently configured with big static fat pipes based on capacity overprovisioning. The rationality behind that is guaranteeing traffic demand and the committed quality of service. Considering a cloud-based scenario with federated data centres (DC) [16], the capacity of each inter-DC optical connection is dimensioned in advance according to some volume of foreseen data to transfer. Once in operation, scheduling algorithms inside the DC's local resource manager run periodically trying to optimize some cost function and organize data transfers among DC (e.g. Virtual Machine (VM) migration and database (DB) synchronization), as a function of the bit rate available. Since DC traffic varies greatly over time, static connectivity configuration adds high costs as a result of large connectivity capacity that remains unused in periods where the amount of data to transfer is low.

Figure 3.9 illustrates the normalized required bit rate along time of the day for VM migration and DB synchronization between DCs. Note that DB synchronization is mainly related to users' activities, whereas VM migration is related to particular policies implemented in scheduling algorithms running inside local resource managers.

Figure 3.10 represents the bit rate utilization along day time for DB synchronization, when the bit rate of the static optical connection between two DC is set to 200 Gb/s. As observed, during idle periods such connection is clearly under-utilized,

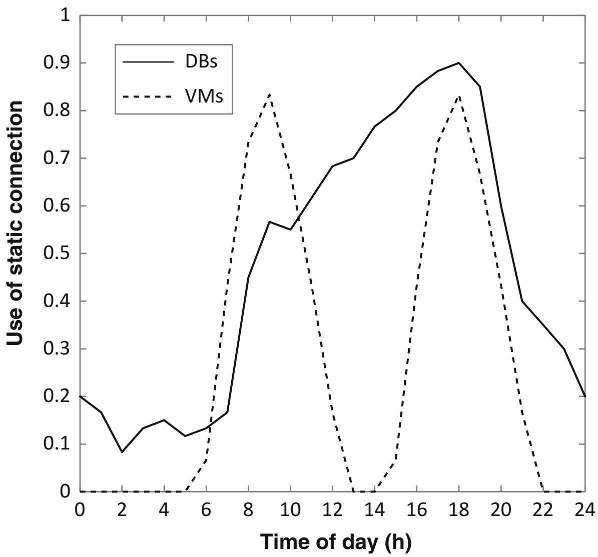



**Fig. 3.9** Use of static connections for DB synchronization and VM migration during time of day

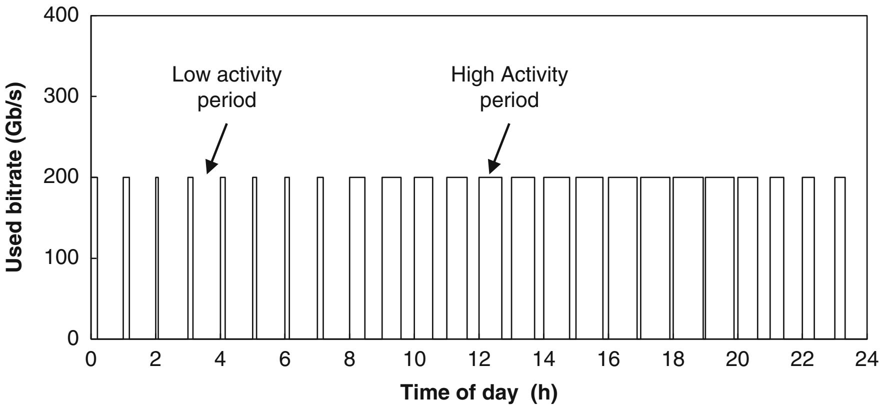



**Fig. 3.10** Used bit rate over a static 200 Gb/s optical connection during time of day

whereas during high activity periods the connection is continuously used. Note that network operators cannot re-allocate non-utilized resources to other customers during idle periods, whereas more resources would be used during high activity periods to reduce time-to-transfer.

In light of the above, it is clear that connectivity among DC requires new mechanisms to provide reconfiguration and adaptability of the transport network to reduce the amount of overprovisioned bandwidth. To improve resource utilization and save costs, dynamic inter-DC connectivity is needed.

### *3.5.2 Dynamic Connection Requests*

To handle dynamic cloud and network interaction, allowing on demand connectivity provisioning, the cloud-ready transport network was introduced in [17]. A cloud-ready transport network to interconnect DC placed in geographically dispersed locations can take advantage of EON providing bit rate on demand; optical connections can be created using the required spectral bandwidth based on users' requirements. Furthermore, by deploying EON in the core, network providers can improve spectrum utilization, thus achieving a cost-effective solution to support their services.

Figure 3.11 illustrates the reference network and cloud control architectures to support cloud-ready transport networks. Local resource managers in the DCs request connection operations to the EON control plane specifying the required bit rate.

In contrast to current static connectivity, dynamic connectivity allows each DC resource manager to request optical connections to remote DCs as data transfers need to be performed, and also to request releasing the connections when all data transfers completed. Furthermore, the fine spectral granularity and wide range of bit rates in EON make the actual bit rate of the optical connection match with the connectivity

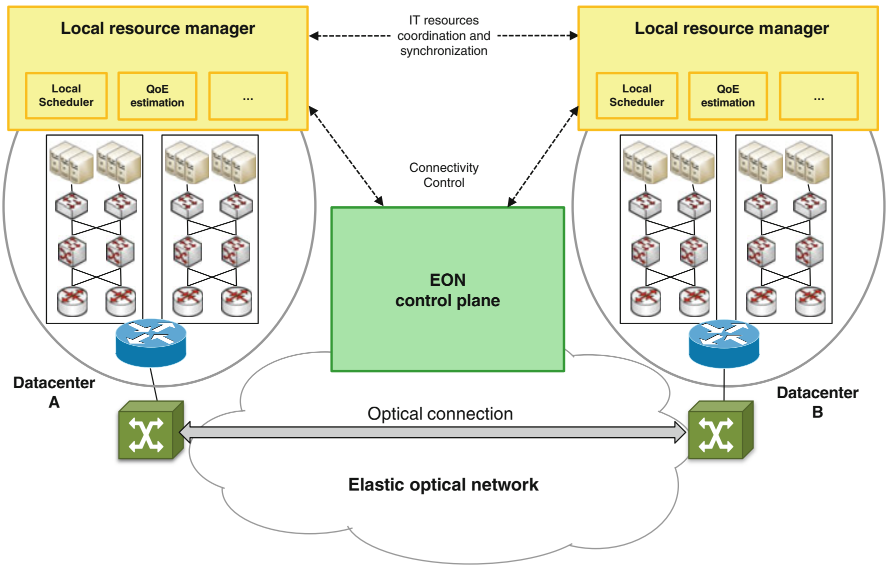



**Fig. 3.11** Architecture to support inter-DC connections

needs. After requesting a connection and negotiating its capacity as a function of the current network resources availability, the resulting bit rate can be used by scheduling algorithms to organize transfers.

Using the architecture shown in Fig. 3.11, local resource managers are able to request optical connections dynamically to the control plane controlling EON and negotiating its bit rate. Figure 3.12a illustrates the message sequence between the local resource manager and the EON control plane to set up and tear down an optical connection. Once the corresponding module in the local resource manager finished computing a data transfer to be performed, a request for an optical connection to a remote DC is sent. In the example, the local resource manager sends a request for an 80 Gb/s connection. Internal operations in the EON control plane are performed to find a route and spectrum resources for the request. Assuming that not enough resources are available for the bit rate requested, an algorithm in the EON control plane finds the maximum bit rate and, at $t_1$, it sends a response to the originating resource manager with that information. Upon the reception of the maximum available bit rate, 40 Gb/s in this example, the resource manager recomputes the transfer and requests a new connection with the corresponding available bit rate. When the transfer completes, the local resource manager sends a message to the EON control plane to tear down the connection, and the used resources are released ($t_2$) so that they can be assigned to any other connection.

Nonetheless, the availability of resources at the time of request is not guaranteed in the dynamic connectivity approach, and the lack of network resources may result in long transference time since transfers cannot be completed within the desired time. Note that connection's bit rate cannot be renegotiated and remains

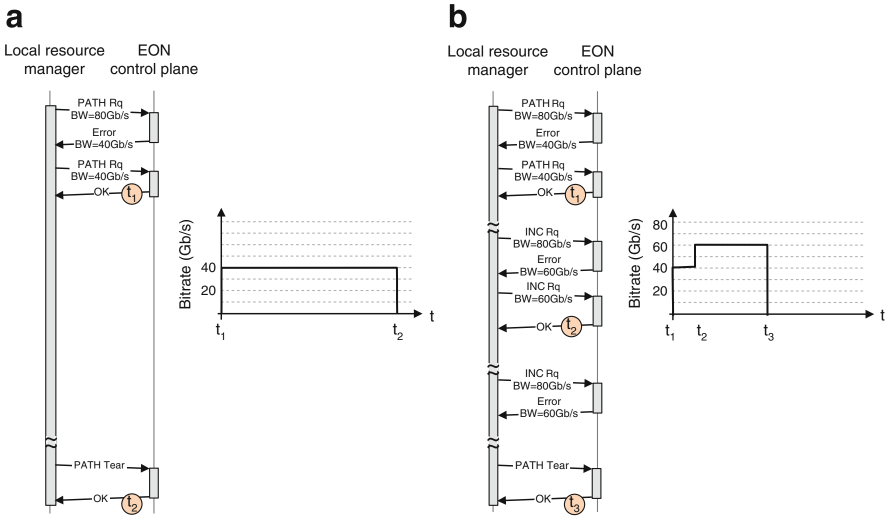



**Fig. 3.12** (**a**) Dynamic and (**b**) elastic connectivity

constant during the connection's holding time. To reduce the impact of the unavailability of required connectivity resources, authors in [18] proposed to use dynamic elastic connectivity.

In the dynamic elastic connectivity model, each local resource manager manages connectivity to remote DC so as to perform data transfers in the shortest total time. The resource manager requests not only connection set up and tear down operations but also elastic operations; the local resource manager is able to request increments in the bit rate of already established connections. Figure 3.12b illustrates messages exchanged between the local resource manager and the EON control plane to set up and tear down an optical connection as well as to increase its bit rate. In the example in Fig. 3.12b, after the connection has been established at $t_1$ with less of the initially requested bit rate (i.e. 40 Gb/s instead of 80 Gb/s initially requested), the local resource manager sends periodical retrials to increment connection's bit rate. In this example, some resources are released after the connection has been established. A request for increasing connection's bit rate is received and additional resources can be assigned to the connection ($t_2$), which reduces total transfer time; in this example, connection is increased to 60 Gb/s. Note that this is beneficial for both DC operators, since better performance could be achieved, and the network operator, since unused resources are immediately utilized.

It is worth noting that, although applications have full control over the connectivity process, physical network resources are shared among a number of clients; therefore, connection set up and elastic operations could be blocked as a result of lack of resources in the network. Hence, applications need to implement some sort of periodical retries to increase the allocated bit rate. These retries could impact negatively on the performance of the EON control plane and do not ensure achieving higher bit rate.

### 3.5.3 Transfer Mode Requests

With the aim to improve both network and cloud resource management, authors in [19] proposed to use Software-Defined Network (SDN), adding a new element, named as Application Service Orchestrator (ASO), in between local resource managers and the EON control plane. Resource managers can request transfer operations using their native semantic (e.g. volume of data and maximum time to transfer) liberating application developers from understanding and dealing with network specifics and complexity. The ASO stratum deployed on top of the EON control plane provides an abstraction layer to the underlying network and implements a northbound interface to request those transfer operations, which are transformed into network connection requests. It is worth noting that in the case where connection requests are sent to the ASO, the ASO would act as a proxy between the local resource manager and the network. Figure 3.13 illustrates both control architectures supporting dynamic and transfer mode requests.

The ASO translates transfer requests into connection requests and forwards them to the EON control plane which is in charge of the network. If not enough resources are available at the time of request, notifications (similar to interruptions in computers) are sent from the EON control plane to the ASO each time specific resources are released. Upon receiving a notification, the SDN takes decisions on whether to increase the bit rate associated to a transfer or not and notifies changes in the bit rate to the corresponding local resource manager. This connectivity model for inter-DC effectively moves from polling in the dynamic elastic model to a network-driven transfer model.

Figure 3.14 shows an example of the messages requesting transfer operations exchanged between a local resource manager, the ASO and the EON control plane.

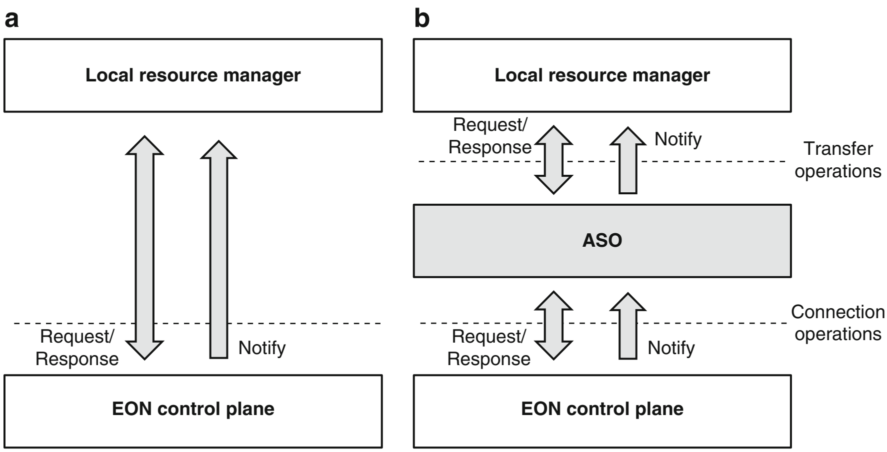



**Fig. 3.13** Control architectures supporting; (**a**) dynamic connections, and (**b**) transfer mode requests

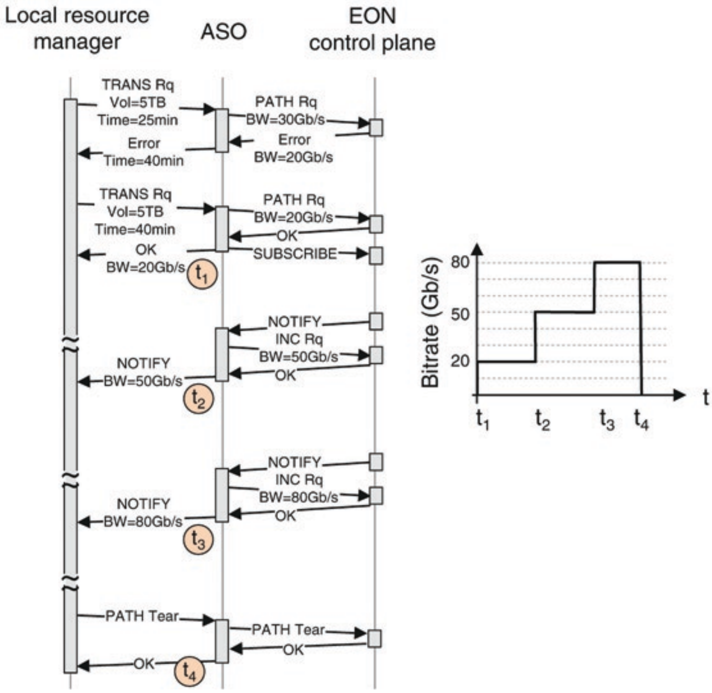



**Fig. 3.14** Transfer mode request

A local resource manager sends a transfer mode request to the ASO for transferring 5 TB of data in 25 min. Upon receiving the request, the) ASO translates it to network connection's semantic, sends a request to the EON control plane to find the biggest spectrum width available, taking into account local policies and current SLA, then sends a response back to the resource manager with the best completion time for transferring the required amount of data, which is 40 min in this example. The resource manager organizes data transfers and sends a new transfer request with the suggested completion time. A new connection is established and its capacity is sent in the response message at $t_1$; in addition, the SDN requests the EON control plane to keep informed upon more resources left available in the route of that connection. Algorithms deployed in the control plane monitor spectrum availability in the corresponding physical links. When resource availability allows increasing the allocated bit rate of some connection, the SDN controller autonomously performs elastic spectrum operations so as to ensure committed transfer completion times. Each time the SDN controller modifies bit rate of any connection by performing elastic spectrum operations, a notification is sent to the source resource manager containing the new throughput information. In this example, some network resources are released and the control plane notifies the SDN. Provided that those resources can be assigned to the connection, its bit rate is increased up to 50 Gb/s and the EON notifies the resource manager of the increased bit rate at $t_2$. The DC resource manager can then optimize transferences as a function of the actual throughput while delegating and ensuring completion transfer time to the SDN. Similarly, at $t_3$ more network resources are released and become available; the connection capacity is then increased to 80 Gb/s. Finally, at $t_4$, the connection is torn down and the corresponding network resources are released.

Finally, the proposed network-driven model opens the opportunity to network operators to implement policies so as to dynamically manage connections' bit rate of a set of customers and fulfil simultaneously their SLAs [20].

To conclude, EON can provide the infrastructure for interconnect DCs. The requirements for that relate to the capacity to support dynamic and elastic connectivity, both in the control and the data plane, since inter-DC traffic highly varies along the day.

## 3.6 Concluding Remarks

This chapter summarized the features of EON that are opening up new directions in the way we perceive networking. EON's mode of operation was demystified and the potential gains were outlined. The usefulness of the EON paradigm was substantiated by means of Use-Cases, outlining how flex-grid technology can be advantageous for multilayer resiliency, for serving metro networks and finally for data centre interconnection.

## References


1. ITU-T G.694.1, Spectral grids for WDM applications: DWDM frequency grid, February 2012
2. R. Ramaswami, K.N. Sivarajan, Routing and wavelength assignment in all-optical networks. IEEE/ACM Trans. Netw. **3**(5), 489–500 (1995)
3. M. Gunkel, A. Autenrieth, M. Neugirg, J. Elbers, Advanced multilayer resilience scheme with optical restoration for IP-over-DWDM core networks—how multilayer survivability might improve network economics in the future, in *International Workshop on Reliable Networks Design and Modeling (RNDM)*, 03.10.2012, St. Petersburg, Russia, 2012
4. Curri et al., Time-division hybrid modulation formats: Tx operation strategies and countermeasures to nonlinear propagation, in *Proceedings OFC, 2014*
5. A. Autenrieth, J. Elbers, M. Eiselt, K. Grobe, B. Teipen, H. Grießer, Evaluation of technology options for software-defined transceivers in fixed WDM grid versus flexible WDM grid optical transport networks, in *Proceedings ITG Photonic Networks*, 2013
6. F. Rambach, B. Konrad, L. Dembeck, U. Gebhard, M. Gunkel, M. Quagliotti, L. Serra, V. López, A multilayer cost model for metro/core networks. J. Opt. Commun. Netw. **5**, 210–225 (2013)
7. O. Gerstel, M. Jinno, A. Lord, S.J. Ben Yoo, Elastic optical networking: a new dawn for the optical layer? IEEE Commun. Mag. **50**(2), S12–S20 (2012)
8. M. Svaluto et al., Experimental validation of an elastic low-complex OFDM-based BVT for flexi-grid metro networks, in *ECOC*, 2013
9. M. Svaluto et al., Assessment of flexgrid technologies in the MAN for centralized BRAS architecture using S-BVT, in *ECOC*, 2014
10. Idealist deliverable D1.1, Elastic optical network architecture: reference scenario, cost and planning

11. Cisco CRS-3, http://www.cisco.com/c/en/us/products/collateral/routers/carrier-routing-system/data_sheet_c78-408226.html
12. V. Gkamas, K. Christodoulopoulos, E. Varvarigos, A joint multi-layer planning algorithm for IP over flexible optical networks. IEEE/OSA J. Lightwave Technol. **33**(14), 2965–2977 (2015)
13. O. Gerstel, C. Filsfils, T. Telkamp, M. Gunkel, M. Horneffer, V. Lopez, A. Mayoral, Multi-layer capacity planning for IP-optical networks. IEEE Commun. Mag. **52**(1), 44–51 (2014)
14. C. Matrakidis, T. Orphanoudakis, A. Stavdas, A. Lord, Network optimization exploiting grooming techniques under fixed and elastic spectrum allocation, in *ECOC*, 2014
15. V. Gkamas, K. Christodoulopoulos, E. Varvarigos, Energy-minimized design of IP over flexible optical networks. Wiley Int. J. Commun. Syst. (2015). doi:10.1002/dac.3032
16. I. Goiri, J. Guitart, J. Torres, Characterizing cloud federation for enhancing providers' profit, in *Proceedings IEEE International Conference on Cloud Computing*, 2010, pp. 123–130
17. L. Contreras, V. López, O. González De Dios, A. Tovar, F. Muñoz, A. Azañón, J.P. Fernandez-Palacios, J. Folgueira, Toward cloud-ready transport networks. IEEE Commun. Mag. **50**, 48–55 (2012)
18. L. Velasco, A. Asensio, J.L. Berral, V. López, D. Carrera, A. Castro, J.P. Fernández-Palacios, Cross-stratum orchestration and flexgrid optical networks for datacenter federations. IEEE Netw. Mag. **27**, 23–30 (2013)
19. L. Velasco, A. Asensio, J.L. Berral, A. Castro, V. López, Towards a carrier SDN: an example for elastic inter-datacenter connectivity. OSA Opt. Express **22**, 55–61 (2014)
20. A. Asensio, L. Velasco, Managing transfer-based datacenter connections. IEEE/OSA J. Opt. Commun. Netw. **6**, 660–669 (2014)

# Chapter 4
# Routing and Spectrum Allocation

**Luis Velasco, Marc Ruiz, Kostas Christodoulopoulos, Manos Varvarigos, Mateusz Żotkiewicz, and Michal Pióro**

## Acronyms

| | |
|---|---|
| B&B | Branch & Bound |
| BRKGA | Biased Random-Key Genetic Algorithm |
| CAPEX | Capital expenditures |
| CF | Central frequency |
| CG | Column generation |
| DAD | Dynamic Alternate Direction |
| DHL | Dynamic High expansion–Low contraction |
| DSP | Digital signal processing |
| GA | Genetic algorithm |
| GRANDE | Gradual Network Design |
| GRASP | Greedy Randomized Adaptive Search Procedure |
| ILP | Integer Linear Programming |
| LSO | Large-Scale Optimization |
| MILP | Mixed Integer Linear Programming |
| PLI | Physical layer impairments |
| PR | Path Relinking |
| QoT | Quality of the Transmission |
| RCL | Restricted Candidate List |
| RSA | Routing and spectrum allocation |

L. Velasco (✉) • M. Ruiz
Universitat Politècnica de Catalunya, Barcelona, Spain
e-mail: lvelasco@ac.upc.edu

K. Christodoulopoulos • M. Varvarigos
University of Patras, Rio Achaia, Greece

M. Żotkiewicz • M. Pióro
Warsaw University of Technology, Warszawa, Poland

 
V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_4

| | |
|---|---|
| RWA | Routing and Wavelength Assignment |
| SEC | Spectrum Expansion/Contraction policy |
| WSON | Wavelength switched optical networks |

To properly analyze, design, plan, and operate flex-grid networks, the routing and spectrum allocation (RSA) problem must be considered. The RSA problem involves two basic constraints: the continuity constraint to ensure that the allocated spectral resources are the same along the links in the route, and the contiguity constraint to guarantee that those resources are contiguous in the spectrum. Moreover, since advanced tunable transponders are envisioned for flex-grid networks, their configuration along with physical layer considerations need to be included in the optimization process. As a consequence of the RSA complexity, it is crucial that efficient methods are available for solving realistic problem instances in reasonable time. In this chapter, we review different variants of the RSA optimization problem; different methods to solve those variants are reviewed along with different requirements related to where those variants are applicable. Starting from a general RSA formulation, we analyze the network life-cycle and discuss different solving methods for the problems that arise at each particular network cycle: from offline to in-operation network planning. We tackle three representative use cases: (1) a use case for offline planning where a demand matrix need to be served taking into account physical layer impairments; (2) a use case for offline planning where a flex-grid network is designed and periodically upgraded; and (3) elastic bandwidth provisioning once the network is being operated.

## 4.1 Introduction

Before deploying a flex-grid network, certain activities need to be undertaken. Starting with inputs from the service layer and from the state of the already deployed resources, a planning phase needs to be carried out to produce recommendations that can be used to design the network for a given period of time. Next, the network design is verified and manually implemented. While the network is in operation, its capacity can be continuously monitored and the resulting data is used as input for the next planning cycle. In case of unexpected increase in demand or network changes, nonetheless, the planning process may be restarted.

Planning (i.e., network designing and dimensioning) generally consists in specifying the nodes and links to be installed and determining the equipment to be purchased to serve the foreseen traffic while minimizing network capital expenditures (CAPEX). To do so, traffic demands need to be routed over the network and a portion of the optical spectrum allocated to each of them, thus creating *lightpaths*. Note that the routing and spectrum allocation (RSA) problem is a part of the optimization problem that needs to be solved.

Once the network is being operated, provisioning algorithms need to solve the RSA in real time. In addition, Network reconfiguration requires solving optimization problems, where the obtained solutions are immediately implemented in the network [1]. Note that this is in contrast to traditional network planning, where results and recommendations require manual intervention and hence substantial time is needed until they are implemented in the network. *In-operation* planning involves the RSA problem; for example, recovery is a typical use case where the RSA problem encompasses a set of traffic demands.

Therefore, network planning problems can be solved *offline* since the network is not yet in operation and hence generally no strict time limit is applied to solve those problems. In contrast, when the network is *in operation*, stringent solving times are usually necessary.

Focusing on both, offline and in-operation planning, the contribution of this chapter is threefold: (1) we present two alternative Integer Linear Programming (ILP) formulations that can be used to model network planning variants of the RSA problem, named *link-path* and *node-link*. Several solving techniques to provide good trade-off between optimality and complexity are presented. In addition, a RSA algorithm for single demands is also reviewed; (2) two use cases for offline planning are presented: firstly the routing and spectrum allocation problem is extended with physical layer considerations, and secondly, the Gradual Network Design (GRANDE) problem, where the network is periodically updated to cope with yearly traffic increments, is presented; and (3) to illustrate online provisioning, the elastic bandwidth provisioning problem is presented.

## 4.2 Basic Offline Planning Problems

In this section we first introduce some basic concepts related to RSA and offline planning problems where the RSA need to be solved.

### *4.2.1 Basic Concepts*

The RSA problem consists in finding a feasible route and spectrum allocation for a set of demands. Similarly to the Routing and Wavelength Assignment (RWA) problem in fixed-grid networks, the spectrum continuity constraint must be enforced. In the case of flex-grid, the spectrum allocation is represented by a *slot* and thus, in the absence of spectrum converters, the same slot must be used along the links of a given routing path. Besides, the allocated *slices* must be contiguous in the spectrum; this is called as spectrum contiguity constraint. The RSA problem was proved to be *NP*-complete in [2] and [3]. As a consequence, it is crucial that efficient methods are available to allow solving realistic problem instances in practical times.

**Table 4.1** $C(d)$ precomputation

| **INPUT** $S$, $d$ | |
|---|---|
| **OUTPUT** $C(d)$ | |
| 1: | Initialize: $C(d) \leftarrow \mathbf{0}_{[|S|-nd+1 \text{ x } |S|]}$ |
| 2: | **for each** $i$ in $[0, |S|-n_d]$ **do** |
| 3: | **for each** $s$ in $[i, i+n_d-1]$ **do** |
| 4: | $C(d)[s]=1$ |
| 5: | **return** $C(d)$ |

Due to the spectrum contiguity constraint, RWA problem formulations developed for WDM networks are not applicable for RSA in flex-grid optical networks and they need to be adapted to include that constraint. Although several works can be found in the literature presenting ILP formulations for RSA, we rely on those in [4] since their approach, based on the assignment of slots, allows efficiently solving the RSA problem.

The definition of slots can be mathematically formulated as follows. Let us assume that a set of slots $C(d)$ is predefined for each demand $d$, which requests $n_d$ slices. Let $q_{cs}$ be a coincidence coefficient which is equal to 1 whenever slot $c \in C$ uses slice $s \in S$, and 0 otherwise. Hence, the spectrum contiguity constraint $\forall c \in C(d)$ is implicitly imposed by the proper definition of $q_{cs}$ such that $\forall i, j \in S$: $q_{ci}=q_{cj}=1$, $i<j \Rightarrow q_{ck}=1, \forall k \in \{i, \ldots, j\}$, $\sum_{s \in S} q_{cs}=n_d$. We consider that each set $C(d)$ consists of all possible slots of the size requested by $d$ that can be defined in $S$. Since $|C(n)|=|S|-(n-1)$, the size of the complete set of slots $C$ that needs to be defined is $|C|=\sum_{n \in N} [|S|-n+1] < |N| \cdot |S|$.

The algorithm in Table 4.1 computes $C(d)$. Note that slot computation is trivial and thus no additional complexity is added to the precomputation phase.

Therefore, we can define the RSA problem as the problem that finds a proper lightpath, i.e., a route and a slot, for each demand from a given set so that the number of active slices in the assigned slot guarantees that the bitrate requested by each demand can be transported. Note that by precomputing the set of slots that can be assigned to each demand, the complexity added by the contiguity constraint is removed.

Finally, without loss of generality, we can consider that guard bands are included as a part of the requested spectrum, i.e., in $n_d$.

## *4.2.2 Basic RSA Problem*

A basic RSA problem is to find a lightpath for every demand in a given traffic matrix with the objective of minimizing or maximizing some objective function. Several alternatives for this problem may exist, for instance, we can assume that the entire traffic specified by the traffic matrix needs to be served, or, alternatively, some demands can be blocked, in other words, not served. Additional characteristics, such as selecting the modulation format and/or accounting for the lightpath reach will be considered in the next subsection. The problem can be formally stated as follows.

*Given*:

- Connected graph $G(N, E)$, where $N$ is the set of locations and $E$ is the set of optical fibers connecting pairs of locations
- Characteristics of the optical spectrum (i.e., spectrum width and frequency slice width) and the set of modulation formats
- Traffic matrix $D$ with the amount of bitrate exchanged between each pair of locations in $N$.

*Output*: route and spectrum allocation for each demand in $D$.

*Objective*: one or more among:

- Minimize the amount of bitrate blocked
- Minimize the total amount of used slices
- etc.

In the following, we present an ILP model for the above problem, based on the formulations in [4]. Note that since the topology is given, we can precompute a set of $k$ distinct paths for each of the demands in the traffic matrix and hence, the formulation is usually known as *link-path* [5]. Moreover, because of the use of precomputed slots for each demand, we call this formulation Link-Path-Slot-Assignment (LP-SA).

The following sets and parameters are defined.Topology

$N$ Set of locations, index $n$

$E$ Set of fiber links, index $e$

Demands and Paths

$D$ Set of demands, index $d$. For each demand $d$, the tuple $\{o_d, t_d, b_d\}$ is given, where $o_d$ and $t_d$ are the origin and target nodes, and $b_d$ is the bitrate in Gb/s

$P$ Set of precomputed paths, index $p$

$P(d)$ Subset of precomputed paths for demand $d$, $|P(d)| = k\ \forall d \in D$

$r_{pe}$ Equal to 1 if path $p$ uses link $e$

Spectrum

$S$ Set of spectrum slices, index $s$

$C(d)$ Set of precomputed slots for demand $d$

$q_{cs}$ Equal to 1 if slot c uses slice s

The decision variables are:

$w_d$ Binary, equal to 1 if demand $d$ cannot be served

$x_{dpc}$ Binary, equal to 1 if demand $d$ is routed through path $p$ and slot $c$

The LP-SA formulation is as follows:

$$(\text{LP}-\text{SA}) \min \sum_{d \in D} b_d \cdot w_d \tag{4.1}$$

subject to:

$$\sum_{p \in P(d)} \sum_{c \in C(d)} x_{dpc} + w_d = 1, \quad \forall d \in D \tag{4.2}$$

$$\sum_{d \in D} \sum_{p \in P(d)} \sum_{c \in C(d)} r_{pe} \cdot q_{cs} \cdot x_{dpc} \leq 1, \forall e \in E, s \in S \tag{4.3}$$

Objective function (4.1) minimizes the amount of bitrate that is not served (rejected). Constraint (4.2) ensures that a lightpath is selected for each demand provided that the demand is served; otherwise the demand cannot be served and therefore is rejected. Constraint (4.3) guarantees that every slice in every link is assigned to at most one demand.

The size of the LP-SA formulation is O($|D| \cdot k \cdot |C|$) variables and O($|E| \cdot |S| + |D|$) constraints. As an example, the size of the above formulation for the 22-node, 35-link BT network, considering $|S| = 80$, $|D| = 100$, and $k = 10$, is 80,000 variables and 2900 constraints. Therefore, the size of this problem is noteworthy.

### *4.2.3 Topology Design as a RSA Problem*

The previous RSA problems assumed that all links in *E* are installed but it is not guaranteed that they are sufficient to carry the full traffic demand, and if they are not we are minimizing the traffic that must be rejected. Another version of RSA is considered below; we assume that the links in *E* are sufficient to carry the demand specified by the traffic matrix and in fact not all of them are needed for that. Consequently, our objective is to minimize the number of links that are necessary to carry the entire demand. Since each installed link increases network CAPEX as a result of optical interfaces, including amplifiers, to be installed in the end nodes and some intermediate locations, minimizing the number of links in the resulting network topology would reduce CAPEX cost.

The problem can be formally stated as follows:

*Given*:

- A connected graph *G*(*N*, *E*)
- The characteristics of the optical spectrum and the set of modulation formats
- A traffic matrix *D*

*Output*:

- The route and spectrum allocation for each demand in *D*
- The links that need to be equipped

*Objective*: Minimize the number of links to be equipped to transport the given traffic matrix.

Note that we could precompute $k$ distinct routes for each demand in the traffic matrix, as we did in the previous problem. However, since only part of the links will be eventually installed, the number of routes $k$ to be precomputed for each demand would need to be highly increased to counteract the fact that some of the routes would become useless. For that very reason, we present an ILP formulation known as *node-link* [5] that includes routing computations within the optimization process. Similarly as before, because of the use of precomputed slots for each demand, we call this formulation as node-link slot-assignment (NL-SA).
A new parameter has been defined:

$g_{ne}$ Equal to 1 if link $e$ is incident to node $n$

The decision variables are:

$x_{dec}$ Binary, equal to 1 if demand $d$ uses slot $c$ in link $e$
$z_e$ Binary, equal to 1 if link $e$ is installed

The NL-SA formulation is as follows:

$$(\text{NL}-\text{SA}) \ \min \ \sum_{e \in E} z_e \tag{4.4}$$

subject to:

$$\sum_{e \in E} \sum_{c \in C(d)} g_{ne} \cdot x_{dec} = 1, \quad \forall d \in D, n \in \{o_d, t_d\} \tag{4.5}$$

$$\sum_{e \in E} \sum_{c \in C(d)} g_{ne} \cdot x_{dec} \le 2, \ \ \forall d \in D, \ n \in N \setminus \{o_d, t_d\} \tag{4.6}$$

$$\sum_{\substack{e' \in E \\ e' \neq e}} g_{ne'} \cdot x_{de'c} \ge x_{dec}, \ \forall d \in D, c \in C(d), \ n \in N \setminus \{o_d, t_d\}, \ e \in E(n) \tag{4.7}$$

$$\sum_{d \in D} \sum_{c \in C(d)} q_{cs} \cdot x_{dec} \le z_e, \ \forall e \in E, \ \ s \in S \tag{4.8}$$

Objective function (4.4) minimizes the amount of links to be installed. Constraints (4.5), (4.6), (4.7) specify the lightpath for every demand. The lighpath for demand $d \in D$ is specified by the links $e \in E$ and the channel $c \in C(d)$ for which $x_{dec} = 1$. Specifically, constraint (4.5) ensures that one lightpath for each demand is created with end nodes equal to the source and destination of demand. Constraint (4.6) guarantees that each lightpath is a connected set of links using the same slot along the route, whilst constraint (4.7) assures that the route does not contain any loop. Finally, constraint (4.8) prevents that any slice in any link is used by more than one demand, while installing the link when any slice is used.

The size of the NL-SA formulation is $O(|D| \cdot |E| \cdot |C|)$ variables and $O(|D| \cdot |C| \cdot |N| \cdot |E|)$ constraints. The size of this formulation for the BT network is 280,000 variables and 6,200,000 constraints, remarkably higher than in the LP-SA formulation.

### *4.2.4 Network Dimensioning as a RSA Problem*

It is obvious that minimizing the number of links to be installed can be different than minimizing CAPEX, since some other costs need to be considered. For this very reason, the previous problem needs to be extended to take into account all the costs and to dimension all network resources.

The network dimensioning problem can be formally stated as follows.

*Given*:

- A connected graph $G(N, E)$
- The characteristics of the optical spectrum and the set of modulation formats
- A traffic matrix $D$
- The cost of every component, such as optical cross-connects (OXC), transponder (TP) types and regenerators specifying its capacity and reach. The cost of installing each link is also specified

*Output*:

- The route and spectrum allocation for each demand in $D$
- The links that need to be equipped
- Network dimensioning including the type of OXC, TPs, and regenerators in each location

*Objective*: Minimize the total CAPEX to transport the given traffic matrix.

Although we do not present any specific ILP model for this problem, it could be derived from the NL-SA formulation using an appropriate CAPEX model (see for example [6]). Notwithstanding, it is easy to realize that the size of the resulting formulation would be noticeable higher and might become intractable even using state-of-the-art computer hardware and the latest commercially available solvers, e.g., CPLEX [7]. In the next section we review some alternative solving techniques.

## 4.3 Solving Techniques

As we showed in the previous section, ILP or Mixed ILP (MILP) models for network planning might entail problems with literally thousands of millions of (integer or binary) variables. To deal with this complexity, in this section we present alternative methods to find near-optimal solutions.

**Table 4.2** Path generation algorithm

| | |
|---|---|
| **INPUT**: *G*, *D* | |
| **OUTPUT**: *Solution* | |
| 1: | $P^* \leftarrow$ initialize set of paths |
| 2: | $L \leftarrow$ initialize master problem (MILP with real variables) |
| 3: | **while** $P^* \neq \varnothing$ **do** |
| 4: | Add new columns to $L$ from all paths in $P^*$ |
| 5: | *Solution* $\leftarrow$ Solve $L$ |
| 6: | $[\lambda, \pi] \leftarrow$ dual variables in *Solution* |
| 7: | $P^* \leftarrow$ *PricingProblem*$(G, D, [\lambda, \pi])$ |
| 8: | *MILP* $\leftarrow L$ with binary variables |
| 9: | *Solution* $\leftarrow$ Solve *MILP* |

### 4.3.1 Large-Scale Optimization

The objective of the Large-Scale Optimization (LSO) methods is to improve the exact methodology based on the classical Branch & Bound (B&B) algorithm for solving MILP formulations [8]. Among different methods, decomposition methods such as column generation (CG) and Benders decomposition have been successfully used for solving communications network design problems.

When solving large instances of problems based on precomputed variables (for example, LP-SA where paths are precomputed), it is necessary to include enough of those variables to ensure, at least, a near-optimal solution. Instead, CG provides a way to find a reduced set of variables producing high-quality solutions. Basically CG consists in two subproblems that are iteratively solved (see Table 4.2): the *restricted master* (or just master) *problem*, which is the linear relaxation of the original MILP that grows at each iteration with new variables, and the *pricing problem* that finds new variables to feed the restricted master. At each iteration, the pricing problem is solved taking as input data the dual variables of the restricted master. The iterative algorithm ends when no more variables are found, i.e., the current solution of the restricted master problem cannot be further improved. Since CG does not ensure integer optimal solutions, the B&B algorithm needs to be ultimately applied. In this regard, when CG is applied at each node of the B&B tree, the resulting algorithm is known as Branch & Price. Finally, note that in the context of networking problems, where variables are mainly paths, this technique is also known as path generation. This method has been recently proved to be an efficient way to generate precomputed lightpaths for link-path RSA formulations [9].

Benders decomposition method is an iterative procedure based on fixing a subset of *difficult* variables and solving the *master problem*, which includes the remaining variables. In order to reach the optimal solution for the original problem, a subproblem is solved for finding new constraints to be added to the master problem and improve the overall solution. Note that, in contrast to column generation, Benders decomposition adds inequalities to the linear formulation being solved, thus strengthening lower bounds and speeding up the convergence to integer optimal solutions. The combination of B&B with this and other methods to generate inequalities or cuts, such as cutting plane, derives into the Branch & Cut algorithm.

Nonetheless, when the time to find a solution is critical, which happens when the network is in operation, a better trade-off between solutions' quality and time-to-compute can be obtained by relaxing optimality condition to find near-optimal solutions much more quickly.

### *4.3.2 Metaheuristics*

Heuristic algorithms are a practical way to produce suboptimal feasible solutions. In particular, metaheuristics (high-level strategies) guide a problem specific heuristic algorithm, to increase their performance avoiding the disadvantages of iterative improvement allowing escaping from local optima. Although a large variety of metaheuristic methods have appeared in the literature, we focus on describing two of them: the Greedy Randomized Adaptive Search Procedure (GRASP) [10] and the Biased Random-Key Genetic Algorithm (BRKGA) [11]. Some other *popular* metaheuristics are ant colony optimization, simulated annealing, and tabu search [12]. In addition, the Path Relinking (PR) intensification strategy is presented as a method to enhance heuristic solutions.

The GRASP procedure is an iterative two phase metaheuristic method based on a multi-start randomized search technique. In the first phase, a *constructive* algorithm is run to obtain a greedy randomized feasible solution of the problem. Roughly speaking, a solution is built by iteratively adding elements randomly chosen from a Restricted Candidate List (RCL) containing those elements with the best quality. The size of the RCL is determined by the parameter $\alpha \in [0, 1]$, being the extreme $\alpha$ values the pure greedy and the pure random configurations, respectively. Then, in the second phase, a local search technique to explore an appropriately defined neighborhood is applied in an attempt to improve the current solution. These two phases are repeated until a stopping criterion (e.g., number of iterations) is met, and once the procedure finishes, the best solution found over all GRASP iterations is returned. Table 4.3 presents an adaptation of the GRASP metaheuristic.

The BRKGA metaheuristic, a class of genetic algorithm (GA), has been recently proposed to effectively solve RSA-related optimization problems [13]. Compared to other metaheuristics, BRKGA has provided better solutions in shorter running times. As in GAs, each individual solution is represented by an array of *n genes* named *chromosome*, where each gene can take any value in the real interval $[0, 1]$. Each chromosome encodes a solution of the problem and a fitness value, i.e., the value of the objective function. A set of individuals, called a *population*, evolves over a number of *generations*. At each generation, individuals of the current genera-

**Table 4.3** GRASP algorithm

| |
|---|
| **INPUT** $G$, $D$, $\alpha$, *maxIter* |
| **OUTPUT** *BestSol* |
| *BestSol* ← Ø |
| **for** 1..*maxIter* **do** |
|   *Sol* ← Ø; $Q \leftarrow D$ |
|   **while** $Q \neq Ø$ **do** |
|     **for each** $d \in Q$ **do** |
|       evaluate the quality $q(d)$ |
|     $q^{min} \leftarrow \min\{q(d) : d \in Q\}$; $q^{max} \leftarrow \max\{q(d) : d \in Q\}$ |
|     $RCL \leftarrow \{d \in Q : q(d) \geq q^{max} - \alpha(q^{max} - q^{min})\}$ |
|     Select an element $d$ from *RCL* at random |
|     $Q \leftarrow Q \setminus \{d\}$; *Sol* ← *Sol* U $\{d\}$ |
|   *Sol* ← *doLocalSearch* (*Sol*) |
|   *Sol*.fitness ← computeFitness(*Sol*) |
|   **if** *BestSol* = Ø OR *Sol*.fitness > *BestSol*.fitness **then** |
|     *BestSol* ← *Sol* |
| **return** *BestSol* |

tion are selected to mate and produce offspring, making up the next generation. In BRKGA, individuals of the population are classified into two sets: the *elite* set with those individuals with the best fitness values and *non-elite* set. Elite individuals are copied unchanged from one generation to the next, thus keeping track of good solutions. The majority of new individuals are generated by combining two elements, one elite and another non-elite, selected at random (crossover). An *inheritance probability* is defined as the probability that an offspring inherits the gene of its elite parent. Finally, to escape from local optima a small number of *mutant* individuals (randomly generated) are introduced at each generation to complete a population. A deterministic algorithm, named *decoder*, transforms any input chromosome into a feasible solution of the optimization problem and computes its fitness value. In the BRKGA framework, the only problem-dependent parts are the chromosome internal structure and the decoder, and thus, one only needs to define them to completely specify a BRKGA heuristic.

Metaheuristic methods can be extended to create hybrid methods that improve the performance of the original metaheuristic algorithm. One of the most extended hybrid methods consist in adding PR as an intensification strategy that explores trajectories connecting heuristic solutions. It starts at a so-called *initiating* solution and moves towards a so-called *guiding* solution. To ensure that PR is only applied among high-quality solutions, an elite set ES must be both maintained and cleverly managed during all iterations. With the attribute high-quality we are not only referring to their cost function value but also to the diversity they add to the set ES. GRASP+PR have been successfully used in many applications including flexgrid network defragmentation [14].

### 4.3.3 RSA Algorithm for Single Demands

Finally, let us analyze the special case where the RSA problem needs to be solved for a single demand. In this case, shortest paths algorithms (e.g., $k$-shortest paths [15]), can be adapted to include spectrum availability; in a second step, spectrum allocation can be realized using any heuristic algorithms (e.g., first fit, random selection, etc.).

In the $k$-shortest paths, each node $i$ is labeled with the aggregated metric $m(i)$ from the source node $o$ and with its predecessor $pre(i)$. Consequently, the route $o$-$i$, defined by the subset of links $E(o,i) \subseteq E$, can be computed repetitively visiting the predecessor node starting from $i$, until source node $o$ is reached.

At this point, let $\eta_{es}$ be equal to 1 if slice $s$ in link $e$ is free, 0 otherwise, and be $S(c)$ the set of contiguous slices in slot $c$. Then, labels can be augmented with $\eta_s(i)$, the aggregated state of slice $s$ ($\eta_s(i) = \prod_{e \in E(o,i)} \eta_{es}$) for each $s \in S$. The downstream node $j$ of node $i$ updates the label only if at least one slot is available as described in Eq. (4.9), i.e., only if $\sigma(i,j) = 1$.

$$\sigma\left(e=\left(i,j\right)\right)=\begin{cases}1 & \exists c\in C\left(d\right):\eta_s\left(i\right)\cdot\eta_{es}=1 \quad \forall s\in S\left(c\right)\\0 & \text{otherwise}\end{cases} \tag{4.9}$$

Note that the complexity of the proposed spectrum availability extension is negligible. Besides, spectrum allocation is performed after the shortest route is found, adding flexibility to use any heuristic algorithms.

In the next sections we apply the presented techniques to different use cases for offline and in-operation planning.

## 4.4 Use Case I: Tunable Transponders and Physical Layer Considerations

In this use case we show how the ILP formulations in Sect. 4.2 can be extended to add extra considerations, such as physical layer impairments (PLIs).

Various implementations of flex-grid) transponders (referred to as bandwidth variable transponders—BVT) have been proposed [16], based on transmission techniques ranging from multi-carrier schemes, such as electrically or optically generated OFDM and Nyquist WDM, to single carrier schemes, usually employing some sort of digital signal processing (DSP) at the receiver but also at the transmitter side. These BVTs can adapt several transmission parameters, such as the modulation format, and/or the baudrate and/or the spectrum used. The term *software-defined optics* is also often used to describe these techniques.

The Quality of the Transmission (QoT) of a lightpath (measured by, for example, its BER) depends on its transmission parameters, the guardband left and the transmission parameters of the spectrum-adjacent lightpaths that share links with it. PLIs such as noise, dispersion, and interference effects accumulate and deteriorate the )QoT of the lightpath. To keep the QoT acceptable, a lengthy end-to-end connection may have to be regenerated and established in a multi-segment manner, with regenerators serving as "refueling stations" that restore signal quality. The multiple degrees of freedom present in flex-grid networks make connection establishment in such networks a more complicated problem than in fixed-grid WDM networks. The adaptability of the BVT and the interdependence between the transmission parameters, the PLIs and the transmission reach, add substantially to the complexity of the problem.

We propose a way to formulate the physical layer effects and the tunability of the transponders in a flex-grid network and outline an ILP algorithm that takes into account such input. We assume tunable transponders: each type of transponder has certain configurations, and each configuration has a specific optical reach, defined as the length at which the transponder can transmit with acceptable )QoT using that configuration. So, as also discussed above, the optical reach depends not only on the specific lightpath transmission configuration, but also on the presence of spectrum-adjacent interfering lightpaths, their transmission configurations and guardbands used. The number of combinations of possible configurations can be huge; also, PLI analytical models may not capture all effects or experimental measurements may be limited for some of the options. So, it seems that the only viable way to account for physical impairments is to resort to some sort of simplification that captures PLIs in a coarser but safe manner, reducing the parameters and the solution space without eliminating good solutions. In what follows we define the transmission reach assuming that a lightpath suffers worst case interference (four-wave-mixing, cross-phase modulation, cross-talk) by the adjacent lightpaths for a given transmission configuration and guardband distance.

Consider a BVT transponder of cost $c$ that can be tuned to transmit $r$ Gb/s using bandwidth of $u$ spectrum slices and a guardband of $g$ spectrum slots from its adjacent spectrum lightpaths, to reach $l$ km distance with acceptable QoT. This defines a *physical feasibility function* $l{=}f_c(r,u)$ that captures PLIs and can be obtained experimentally or using analytical models [17, 18].

Using the functions $f_c$ of the available transponders we define (*reach-rate-spectrum-cost*) transmission *tuples*, corresponding to feasible configurations of the transponders. The term "feasible" is used to signify that the tuple definition incorporates PLI limitations, while the cost parameter is used when there are transponders of different capabilities and costs. The above definition is very general and can be used to describe any type of flexible or even fixed-grid optical network. For example, the mixed-line-rate (MLR) fixed-grid network of [19] can be represented with the following transmission tuples: (3000 km, 10 Gb/s, 50 GHz, 1), (1600 km, 40 Gb/s, 50 GHz, 3), (800 km, 100 Gb/s, 50 GHz, 6). Using the above methodology the feasible transmission options can be enumerated so as to incorporate the physical layer effects.

The problem can be formally stated as follows:

*Given*:

- A connected graph $G(N, E)$
- The transmission configuration of the transponders, the physical layer and the availability or absence of regenerators, through the transmission tuples,

$t = (l_t, r_t, u_t, c_t)$ indicating a transmission of reach up to $l_t$, at rate $r_t$ (Gb/s), using $u_t$ spectrum slices, for a transponder of type (cost) $c_t$.

- A traffic matrix $D$

*Output*:

- The route and spectrum allocation for each demand in $D$

*Objective*: minimize the amount of bit-rate blocked

The formulation we present next is an extension of the one presented in [20] considering the definition of channel slots given in Sect. 4.2.1, and so it is called as transponder configuration-link-path-slot-assignment (TC-LP-SA). As in the previous LP-SA formulation, $k$ paths are precalculated for each source-destination pair. For each path $p$ and for each transmission tuple $t = (l_t, r_t, u_t, c_t)$, if the length of the path is shorter than $l_t$ we define a feasible path-transmission tuple pair $(p, t)$. We denote by $T$ the set of all available tuples and by $T(p)$ the feasible tuples over path $p$.

Two key differences from the formulation presented in the previous section are: (1) in the new formulation different options for the number of spectrum slots that can be used to serve a demand are allowed and (2) demands can be broken into more than one lightpath, depending on the selected path and transmission tuple. The solution, in addition to assigning routes and slots to the demands, also selects the transmission configuration of the transponders.

The following sets and parameters have been defined:Topology

| | |
|---|---|
| $N$ | Set of locations, index $n$ |
| $E$ | Set of fiber links, index $e$ |

Demands and Paths

| | |
|---|---|
| $D$ | Set of demands, index $d$. For each demand $d$, the tuple $\{o_d, t_d, b_d\}$ is given, where $o_d$ and $t_d$ are the origin and target nodes, and $b_d$ is the bitrate in Gb/s. |
| $P$ | Set of precomputed paths, index $p$ |
| $P(i, j)$ | Subset of precomputed paths between nodes $i$ and $j$, for all $(i, j) \in N^2$, with cardinality $\lvert P(i, j) \rvert = k$ |
| $t = (l_t, r_t, u_t, g_t, c_t)$ | A transmission tuple: $l_t$ maximum reach, $r_t$ rate (Gb/s), $u_t$ spectrum slices, $g_t$ guardband, transponder of type (cost) $c_t$. The total number of slices required is $n_t = u_t + g_t$ |
| $T$, $T(p)$ | Feasible transmission tuples for path $p$ |
| $(p, t)$ | A feasible path-transmission tuple. |
| $r_{pe}$ | Equal to 1 if path $p$ uses link $e$ |

Spectrum

| | |
|---|---|
| $S$ | Set of spectrum slices, index $s$ |
| $C(t)$ | Set of precomputed slots for transmission tuple $t$. Note that $n_t$ is the size of the slot |
| $q_{cs}$ | Equal to 1 if slot $c$ uses slice $s$ |

The decision variables are:

$w_d$ Binary, equal to 1 if demand $d$ cannot be served
$x_{dptc}$ Binary, equal to 1 if demand $d$ uses path $p$ and transmission tuple $t$ and slot $c$

The TC-LP-SA formulation is as follows:

$$(\text{TC LP SA}) \quad \min \sum_{d \in D} b_d \cdot w_d \tag{4.10}$$

subject to:

$$\sum_{n \in N} \sum_{p \in P(o_d, n)} \sum_{t \in T(p)} \sum_{c \in C(t)} r_t \cdot x_{dptc} + b_d \cdot w_d \geq b_d, \quad \forall d \in D \tag{4.11}$$

$$\sum_{n \in N} \sum_{p \in P(n, t_d)} \sum_{t \in T(p)} \sum_{c \in C(t)} r_t \cdot x_{dptc} + b_d \cdot w_d \geq b_d, \quad \forall d \in D \tag{4.12}$$

$$\sum_{i \in N} \sum_{p \in P(i,n)} \sum_{t \in T(p)} \sum_{c \in C(t)} x_{dptc} = \sum_{j \in N} \sum_{p \in P(n,j)} \sum_{t \in T(p)} \sum_{c \in C(t)} x_{dptc}, \quad \forall d \in D,\ n \in N \setminus \{o_d, t_d\},\ t \in T \tag{4.13}$$

$$\sum_{d \in D} \sum_{p \in P(d)} \sum_{t \in T(p)} \sum_{c \in C(t)} r_{pe} \cdot q_{cs} \cdot x_{dptc} \leq 1, \quad \forall e \in E,\ s \in S \tag{4.14}$$

The objective function (4.10), as in before, minimizes the amount of demands that cannot be served (rejected). Constraint (4.11) and (4.12) ensures that (one or more) lightpath(s) of enough rate are established to serve a demand provided that the demand is served; otherwise the demand cannot be served and therefore, is rejected. Constraint (4.13) conserves the lightpath flows at intermediate nodes where regenerators are placed. As written, it assumes that the regenerators would use the same transmission tuple as the initial transponder, and relaxes spectrum continuity, meaning that a different spectrum slots can be used after regeneration. Constraint (4.14) guarantees that every slice in every link is assigned to one demand at most.

The size of the TC-LP-SA formulation is $O(k \cdot |D|^2 \cdot |T| \cdot |C|)$ variables and $O(|E| \cdot |S| + |D| \cdot |N| \cdot |T|)$ constraints.

## 4.5 Use Case II: Gradual Network Design Problem

In this use case we focus on the planning phase )and study the problem of upgrading operators' core networks in order to extend its capacity as the traffic to be transported increases. We call this the Gradual Network Design (GRANDE) problem. In contrast to the models presented in section III, where the network planning was performed considering a green-field scenario, in the GRANDE problem the already deployed equipment can be reused to reduce upgrading CAPEX cost.

In the following subsections we first state the GRANDE problem and present its ILP formulation. As a result of its size, a method based on path generation is proposed. As an alternative, a BRKGA heuristic method is also presented. Illustrative numerical results are shown for two real network examples.

### 4.5.1 Problem Statement

The GRANDE problem can be formally )stated as follows.

*Given*:

- A connected graph $G(N, E)$, where $N$ represents the set of potential OXC locations, and $E$ is the set of links connecting pairs of locations
- Subsets $N_{in} \subseteq N$ and $E_{in} \subseteq E$ containing already installed OXCs and links, respectively
- The characteristics of the optical spectrum
- A traffic matrix $D$
- The cost of installing a new OXC and a new fiber link

*Output*:

- The network topology with the extra equipment needed to transport the set of demands $D$
- The route and spectrum allocation for each demand in $D$

*Objective*: Minimize the CAPEX cost resulting from upgrading the network to transport the given traffic matrix.

It is worth noting that to cope with new traffic requirements, manual operations causing service disruption are in general needed to implement the GRANDE solutions (see for example [21]). In this regard, the GRANDE problem focuses on optimizing traffic routing and network dimensioning admitting service disruption.

### 4.5.2 Mathematical Model

A node-link formulation was proposed in Sect. 4.2.3 for a network planning problem simpler than GRANDE. Although the use of node-link formulation may seem more effective than using link-path formulation when topology design is involved, it is hard to solve even for moderate-size network instances as a result its large size in terms of variables and constraints. Although link-path formulations need precomputed sets of paths for each demand, the path generation technique described in the next section can be applied to generate appropriate sets of paths that are required to achieve an optimal solution. For this very reason, we propose a link-path-based formulation for the GRANDE problem. The ILP model extends the LP-SA formulation presented in Sect. 4.2.2 adding constraints to tackle the already deployed equipment.

The approach we follow to minimize the upgrading CAPEX consists in setting the cost of the already installed equipment to zero and minimize total network CAPEX. CAPEX is divided into different components, such as the cost of installing a new node or a new link.

We note that the constraint to ensure that the whole traffic matrix is transported could result in the problem infeasibility when not enough resources are available. This constraint makes it difficult to apply path generation, so in the proposed formulation we allow demand rejection but with a large cost penalty. Note that when too few precomputed routes are used, the demands might be rejected.

In addition to the notation described so far, the following additional parameters are defined:

| | |
|---|---|
| $fn_n$ | Equal to 1 if an OXC is already installed in location $n$ |
| $fe_e$ | Equal to 1 if link $e$ is already installed |
| $cd_d$ | Penalty cost associated to demand $d$ rejection |
| $cn_n$ | Cost of installing a new OXC in location $n$ |
| $ct$ | Cost of adding a new link to an existing OXC |
| $ce_e$ | Cost of installing link $e$ |

New decision variables are also defined:

$y_n$ Binary variable equal to 1 if an OXC is installed in location $n$

The ILP formulation for the GRANDE problem is as follows:

$$(GRANDE)\ \min \quad \sum_{d\in D} cd_d \cdot w_d + \sum_{e\in E}(1-fe_e)\cdot(ce_e + 2\cdot ct)\cdot z_e + \sum_{n\in N}(1-fn_n)\cdot cn_n \cdot y_n \tag{4.15}$$

subject to:

$$\sum_{p\in P(d)}\sum_{c\in C(d)} x_{dpc} + w_d = 1, \quad \forall d \in D \tag{4.16}$$

$$\sum_{d\in D}\sum_{p\in P(d)}\sum_{c\in C(d)} r_{pe}\cdot q_{cs}\cdot x_{dpc} \le z_e, \quad \forall e\in E,\ s\in S \tag{4.17}$$

$$\sum_{d\in D}\sum_{p\in P(d)}\sum_{c\in C(d)} g_{ne}\cdot r_{pe}\cdot x_{dpc} \le |D|\cdot y_n, \quad \forall n\in N,\ e\in E \tag{4.18}$$

The objective function (4.15) minimizes the weighted sum of the cost of rejecting traffic (this term is assigned a coefficient $c$ which is large with respect to the equipment costs) and the CAPEX cost of the additional equipment. It is worth highlighting that the value of the objective function is 0 when all demands can be routed using the already deployed nodes and links.

**Table 4.4** Pricing problem algorithm

| |
|---|
| **INPUT**: $G$, $D$, $S$, dual variables [$\lambda$, $\pi$, $\mu$] |
| **OUTPUT**: $P^*$ |
| **for** each $d$ in $D$ **do** |
| $incCost \leftarrow 0$; $P^* \leftarrow \emptyset$ |
| **for** each slot $c$ in $C(d)$ **do** |
| Compute link metrics $h_e$ from Eq. (4.20) |
| $p^* \leftarrow$ Shortest path ($o_d$, $t_d$) |
| Compute reduced cost $u_{dp^*c}$ from Eq. (4.19) |
| **if** $u_{dp^*c} > incCost$ **then** |
| $incCost = u_{dp^*c}$ |
| $incPath = p^*$ |
| **if** $incPath \neq \emptyset$ **then** |
| $P^* \leftarrow P^* \cup incPath$ |

Constraint (4.16) ensures that either a lightpath is assigned to each demand or the demand is rejected. Constraint (4.17) guarantees that each slice in each link is used to convey one lightpath at most and, additionally, installs those links conveying any demand. Constraint (4.18) assures that an OXC is installed in those locations that are used by at least one lightpath. Note that the number of lightpaths using a node cannot exceed the number of demands.

The size of the GRANDE model is $O(|D| \cdot k \cdot |C| + |N| + |E|)$ variables and $O(|E| \cdot |S| + |E| \cdot |N| + |D|)$ constraints, in line to that of the LP-SA formulation.

As discussed above, the presence of not installed nodes and links in the original topology makes it difficult to precompute path sets. On the one hand, paths crossing the installed resources do not contribute to the CAPEX increment but increase probability of rejecting demands. On the other hand, paths containing non-installed resources lead to solutions with lower or zero demand rejection but increase the CAPEX cost. A way to solve this issue is presented in the next section, where a path generation algorithm designed to find the best paths to minimize the objective function by simultaneous minimization of demand rejection penalties and by reducing network CAPEX.

### *4.5.3 Path Generation Algorithm*

Aiming at finding the paths leading to good-quality solutions, we propose a path generation algorithm based on the one described in Table *4.4*. We have modified that algorithm to stop when either the pricing algorithm finds no more new paths or a maximum number of iterations (*maxIter*) is reached. This allows controlling the size of the problem, making this size proportional to the computational effort needed to obtain the optimal integer solution. Thus, it could be reasonable to stop generating paths and devote more time to solve the final primal problem.

The specific pricing problem behind path generation for GRANDE is as follows. First we derive the dual of the primal (master) problem (4.16)–(4.18). We define the following dual variables for the constraints of the relaxed GRANDE problem: $\lambda_d$ unconstrained in sign for constraint (4.16), $\pi_{es} \geq 0$ for constraint (4.17), and $\mu_{ne} \geq 0$ for constraint (4.18). Finally, all variables in GRANDE are relaxed to be continuous $0 \leq x_{dpc}, w_d, y_n, z_e \leq 1$ in the master problem $L$, i.e., $L$ contains the same set of variables and constraints as the original GRANDE formulation but with variables defined in the continuous domain.

From $L$, the dual problem can be easily derived from its Lagrangian function, which is obtained by moving the constraints to the objective function, multiplied by its associated dual variable. After grouping and re-ordering components, the Lagrangian function depending on primal and dual variables is as follows:

$$\begin{aligned} &L(x, y, z, \lambda, \pi, \mu, \gamma, \rho) \\ &= + \sum_{d \in D} w_d \cdot (cd_d - \lambda_d) + \sum_{n \in N} y_n \cdot \left( (1 - fn_n) \cdot cn_n - |D| \cdot \sum_{e \in E} g_{ne} \cdot \mu_{ne} \right) \\ &\quad + \sum_{d \in D} w_d \cdot (cd_d - \lambda_d) + \sum_{n \in N} y_n \cdot \left( (1 - fn_n) \cdot cn_n - |D| \cdot \sum_{e \in E} g_{ne} \cdot \mu_{ne} \right) \\ &\quad + \sum_{e \in E} z_e \cdot \left( (1 - fe_e) \cdot (ce_e + 2 \cdot ct) - \sum_{s \in S} \pi_{es} \right) + \sum_{d \in D} \lambda_d \end{aligned} \tag{4.19}$$

Now the dual problem ($D$) is defined as a maximization problem in the dual space, where the variables in brackets are the primal variables related to each constraint in the dual.

$$(D) \quad \max \sum_{d \in D} \lambda_d \tag{4.20}$$

subject to:

$$\begin{aligned} [x_{dpc} \geq 0] \quad \lambda_d \leq \sum_{e \in E} r_{pe} \cdot \left( \sum_{s \in S} q_{cs} \cdot \pi_{es} + \sum_{n \in N} g_{ne} \cdot \mu_{ne} \right), \\ \forall d \in D,\ p \in P(d),\ c \in C(d) \end{aligned} \tag{4.21}$$

$$[y_n \geq 0] \quad |D| \cdot \sum_{e \in E} g_{ne} \cdot \mu_{ne} \leq (1 - fn_n) \cdot cn_n, \quad \forall n \in N \tag{4.22}$$

$$[z_e \geq 0] \quad \sum_{s \in S} \pi_{es} \leq (1 - fe_e) \cdot (ce_e + 2 \cdot ct) \quad \forall e \in E \tag{4.23}$$

Since the objective of the pricing problem is to add paths not considered so far, this turns into adding new dual constraints so as to the current optimal dual solution infeasible. Looking at the formulation of the dual problem, constraint (4.21) is the only one defined for the path variables; therefore, the sole condition that a candidate path $p^*$ to be added to the problem must violate this constraint. Let $u_{dp^*c}$ be the reduced cost of demand $d$ using new path $p^*$ and slot $c$; such $p^*$ is a candidate path only if the reduced cost is strictly positive, i.e.,

$$u_{dp^*c} = \lambda_d - \sum_{e \in E} r_{pe} \cdot \left( \sum_{s \in S} q_{cs} \cdot \pi_{es} + \sum_{n \in N} g_{ne} \cdot \mu_{ne} \right) > 0 \tag{4.24}$$

Therefore, in light of Eq. (4.24), the pricing problem can be stated as follows: for each demand in $D$, find the path $p^*$ and the slot $c \in C(d)$ hat maximize the reduced cost $u_{dp^*c}$, provided that $u_{dp^*c} > 0$. Note that choosing the path with the highest positive reduced cost leads to the highest decrease rate in the objective function of the master problem.

An algorithm for solving the pricing problem is presented in Table *4.4*. For each demand $d$, a shortest path is computed for each of the slots in $C(d)$. For a given slot $c$, link metrics ($h$) used to compute shortest paths are setup as follows (line 4 in Table *4.4*):

$$h_e(c) = \sum_{s \in S} q_{cs} \cdot \pi_{es} + \sum_{n \in N} g_{ne} \cdot \mu_{ne} \tag{4.25}$$

Once the path is found (line 5), its reduced cost is computed using Eq. (4.24) and, if it is higher than the incumbent cost, the path is stored as incumbent path. After exploring all possible slots for the demand, the incumbent path is included in the set $P^*$ containing the generated paths for all demands (lines 6–9).

We have presented a path generation method as an alternative to precomputing shortest paths following natural link metrics. However, when problem instances are excessively large preventing its application, alternative heuristic methods are needed. Next subsection presents the details of a BRKGA heuristic to solve the GRANDE problem.

### 4.5.4 BRKGA Heuristic

As pointed out in Sect. 4.3.2, the only problem-dependent parts of the BRKGA heuristic method are the chromosome structure and the decoder algorithm. Table *4.5* presents the pseudo-code of the decoder algorithm. Such a decoder uses chromosomes to sort the set of demands and hence each chromosome contains one gene for each demand $d$ in $D$.

Starting from the topology with all nodes and links available, the metric of each node and link, which will be afterwards used by the RSA algorithm, are properly initialized to promote the use of the installed nodes and links (lines 1–3 in Table *4.5*). Next, the demands are sorted using the values of the genes in the input chromosome (line 4).

**Table 4.5** Decoder algorithm

| | |
|---|---|
| **INPUT**: *G*, *D*, Chromosome *chr*, *costInstall* | |
| **OUTPUT**: Solution, *fitness* | |
| 1: | Initialize *Solution* with installed nodes and links |
| 2: | Initialize metrics of nodes and links: |
| 3: | if installed set to 0 otherwise to their cost |
| 4: | Sort *D* according to genes in *chr* |
| 5: | **for** each *d* in *D* **do** |
| 6: | *d*.lightpath ← RSA(*G*, *d*) |
| 7: | **if** *d*.lightpath = Ø **then** |
| 8: | **return** *INFEASIBLE* |
| 9: | allocate(*G*, *d*) |
| 10: | *Solution.D* ← *Solution.D* ∪ *d* |
| 11: | **if** *new nodes and/or link have been installed* **then** |
| 12: | Set metrics of new installed equipment to 0 |
| 13: | *Solution.Equip* ← *Solution.Equip* ∪ {installed equip} |
| 14: | *fitness* ← ComputeCAPEX (*Solution*) |

Then, the decoder finds a lightpath for each of the demands following the given order (lines 5–6). Since the RSA algorithm finds a route with the minimum cost, the installed equipment will be reused before installing new nodes and/or links. After allocating the resources (line 9), metrics of the installed nodes and links are updated (lines 11–13). Finally, the fitness value is obtained by computing the CAPEX cost.

The next section evaluates performance of the methods proposed to solve GRANDE using network and traffic scenarios obtained from real network instances.

## 4.6 Use Case III: Elastic Bandwidth Provisioning

If we go back and consider the network life-cycle a planning-offline RSA algorithm (Sect. 4.2) is used to design the network. Then the network is gradually redesigned to adapt to traffic changes (Sect. 4.5). We envision a flexible network where dynamic traffic changes are accommodated at two different levels. The first level is the establishment of new connections that might need the deployment of new equipment (the GRANDE problem). Given the high rate that new generation transponders are expected to transmit (designs of 400 Gb/s or 1 Tb/s have appeared in the literature [16, 17]), relatively long periods of time will pass until a new connection requiring a new transponder is encountered. The second level is to absorb rate changes by adapting the tunable transponders, e.g., tuning the modulation format and/or the number of spectrum slots.

Therefore, in this use case we assume that lightpaths can expand/contract dynamically the spectrum they use to follow the small–medium scale dynamic traffic changes (e.g., daily traffic cycle). The slices that are freed by a lightpath can be

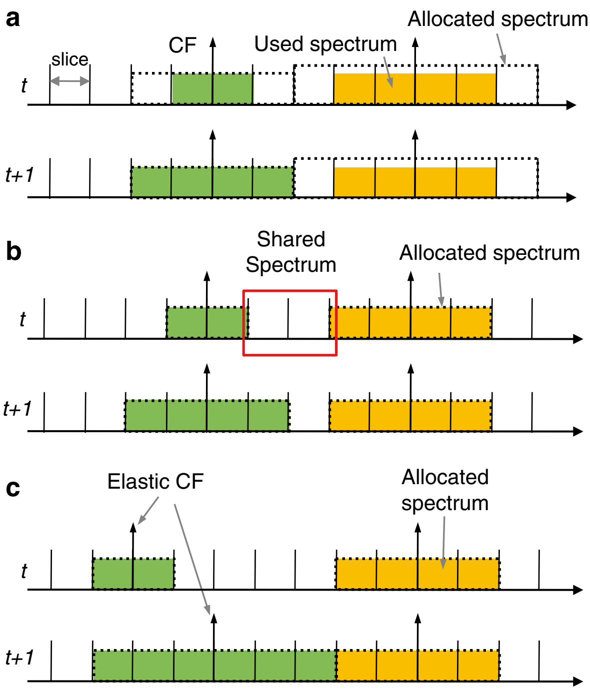



**Fig. 4.1** Three spectrum allocation policies for time-varying traffic in a flex-grid network. Two time intervals are observed: *t* before and *t*+1 after spectrum adaptation has been performed. (**a**) Fixed. (**b**) Semi-elastic. (**c**) Elastic

assigned to a different lightpath at different time instants, obtaining thus statistical multiplexing gains. Statistical multiplexing results in reducing the spectrum and the total number of slices required to serve a given traffic, or in the dual form of the problem, in reducing the blocking of the traffic changes for a given number of slices.

### *4.6.1 Spectrum Allocation Policies*

Authors in [22] discerned three alternative spectrum allocation policies for time-varying traffic demands (reproduced in Fig. 4.1). The spectrum allocation policies put the following restrictions on the assigned central frequency (CF) and the allocated spectrum width, in particular:

- *Fixed* (Fig. 4.1a): both the assigned CF and spectrum width do not change in time. At each time period, demands may utilize either whole or only a fraction of the allocated spectrum to convey the bit-rate requested for that period.
- *Semielastic* (Fig. 4.1b): the assigned CF is fixed but the allocated spectrum may vary. Here, spectrum increments/decrements are achieved by allocating/releasing frequency slices symmetrically, i.e., at each end of the already allocated spectrum while keeping invariant the CF. The frequency slices can be

**Table 4.6** Semielastic spectrum allocation algorithm

| |
|---|
| **INPUT**: $G(N,E)$, $S$, $L$, $p$, $(s_p)^{req}$ |
| **OUTPUT**: $(s_p)'$ |
| **if** $(s_p)^{req} \leq s_p$ **then** |
| $(s_p)' \leftarrow (s_p)^{req}$ |
| **else** |
| $L^+ \leftarrow \emptyset$, $L^- \leftarrow \emptyset$ |
| **for each** $e \in R_p$ **do** |
| $L^- \leftarrow L^- \cup \{l \in L: e \in R_l, \text{adjacents}(l, p), f_l < f_p\}$ |
| $L^+ \leftarrow L^+ \cup \{l \in L: e \in R_l, \text{adjacents}(l, p), f_l > f_p\}$ |
| $s_{\max} \leftarrow 2*\min\{\min\{f_p - f_l - s_l, l \in L^-\}, \min\{f_l - f_p - s_l, l \in L^+\}\}$ |
| $(s_p)' \leftarrow \min\{s_{\max}, (s_p)^{req}\}$ |
| **return** $(s_p)'$ |

shared between neighboring demands, but used by, at most, one demand at a time interval.

- *Elastic* (Fig. 4.1c): asymmetric spectrum expansion/ reduction (with respect to the already allocated spectrum) is allowed and it can lead to short shifting of the central frequency. Still, the relative position of the lightpaths in the spectrum remains invariable, i.e., no reallocation in the spectrum is performed.

The problem of dynamic lightpath adaptation was addressed in [23] and it can be formally stated as:

*Given*:

- A core network topology represented by a graph $G(N, E)$, being $N$ the set of optical nodes and $E$ the set of bidirectional fiber links connecting two optical nodes; each link consists of two unidirectional optical fibers
- A set $S$ of available slices of a given spectral width for every link in $E$
- A set $L$ of lightpaths already established on the network; each lightpath $l$ is defined by the tuple $\{R_l, f_l, s_l\}$, where the ordered set $R_l \subseteq E$ represents its physical route, $f_l$ its central frequency and $s_l$ the amount of frequency slices.
- A lightpath $p \in L$ for which spectrum adaptation request arrives and the required number of frequency slices, $(s_p)^{req}$.

*Output:* the new values for the spectrum allocation of the given lightpath $p$: $\{R_p, f_p, (s_p)'\}$ and $\{R_p, (f_p)', (s_p)'\}$, respectively, if the *Semielastic* and *Elastic* policy is used.

*Objective*: maximize the amount of bit-rate served.

For the *Fixed* spectrum allocation policy, the allocated spectrum does not change in time, therefore, any fraction of traffic that exceeds the capacity of the established lightpath is lost. Regarding the *Semielastic* and *Elastic* policies, the corresponding lightpath adaptation algorithms are presented in Tables 4.6 and 4.7, respectively. In the following, we discuss the details of these algorithms.

**Table 4.7** Elastic spectrum allocation algorithm

| | |
|---|---|
| **INPUT**: $G(N, E), S, L, p, (s_p)^{req}$ | |
| **OUTPUT**: $(f_p)', (s_p)'$ | |
| 1: | **if** $(s_p)^{req} \leq s_p$ **then** |
| 2: | $(s_p)' \leftarrow (s_p)^{req}$ |
| 3: | $(f_p)' \leftarrow f_p$ |
| 4: | **else** |
| 5: | $L^+ \leftarrow \emptyset, L^- \leftarrow \emptyset$ |
| 6: | **for each** $e \in R_p$ **do** |
| 7: | $L^- \leftarrow L^- \cup \{l \in L: e \in R_l$, adjacents$(l, p), f_l < f_p\}$ |
| 8: | $L^+ \leftarrow L^+ \cup \{l \in L: e \in R_l$, adjacents$(l, p), f_l > f_p\}$ |
| 9: | $s_{max} \leftarrow \min\{f_p - f_l - s_l, l \in L^-\} + \min\{f_l - f_p - s_l, l \in L^+\}$ |
| 10: | $(s_p)' \leftarrow \min\{s_{max}, (s_p)^{req}\}$ |
| 11: | $(f_p)' \leftarrow$ findSA_MinCFShifting $(p, (s_p)', L^+, L^-)$ |
| 12: | **return** $(f_p)', (s_p)'$ |

- *Semielastic* algorithm: the elastic operation is requested for a lightpath $p$ and the required amount of frequency slices to be allocated, maintaining $f_p$ invariant, is given. Since flex-grid is implemented, $(s_p)^{\text{req}}$ must be an even number. If elastic spectrum reduction is requested, the tuple for lightpath $p$ is changed to $\{R_p, f_p, (s_p)^{\text{req}}\}$ (lines 1–2). In the opposite, when an elastic expansion is requested, the set of spectrum-adjacent lightpaths at each of the spectrum sides is found by iterating on each of the links of the route of $p$ (lines 4–7). The greatest value of available spectrum without CF shifting, $s_{\text{max}}$, is subsequently computed and the value of spectrum slices actually assigned to $p$, $(s_p)'$, is computed as the minimum between $s_{\text{max}}$ and the requested one (lines 8–9). The tuple representing lightpath $p$ is now $\{R_p, f_p, (s_p)'\}$.
- *Elastic* algorithm: here, the CF of $p$ can be changed so the difference with the *Semielastic* algorithm explained above is related to that issue. Now, the value of $s_{\text{max}}$ is only constrained by the amount of slices available between the closest spectrum-adjacent paths. Then, $s_{\text{max}}$ is the sum of the minimum available slices along the links in the left side and the minimum available slices in the right side of the allocated spectrum (line 9). Finally, the returned value $(f_p)'$ is obtained by computing the new CF value so as to minimize CF shifting (line 11).

### *4.6.2 Spectrum Expansion/Contraction Policies*

To enable the dynamic sharing of spectrum, we need to define Spectrum Expansion/Contraction (SEC) policies to regulate the way this is performed [24]. Under the proposed spectrum sharing framework, a connection shares the spectrum slots with its spectrum-adjacent connections.

- A first policy is the fixed policy, in which each connection is given a specific amount of spectrum and does not share it with other connections. This policy forms the baseline for the other policies that enable dynamic spectrum sharing among connections.
- A second policy is the Dynamic High expansion–Low contraction (DHL) policy, according to which a connection over path *p* wishing to increase its transmission rate first uses its higher spectrum slots until it reaches a slot already occupied by an upper spectrum-adjacent connection on some link of *p*. Then, if additional bandwidth is needed, it expands its lower slots until it reaches a slot that is occupied by some bottom spectrum-adjacent connection on some link of *p*. If the connection needs to increase further its rate and there is no higher or lower free slot space, blocking occurs (for the excess rate). Note that the DHL policy performs indirectly slot defragmentation, since it fills the free higher spectrum slots in every chance it gets. When a connection decreases its spectrum slots due to a reduction in its rate, we first release lower spectrum slots and, if these have been reduced to zero, we release higher slots.
- Another SEC policy is the Dynamic Alternate Direction (DAD) policy which aims at the symmetrical use of spectrum around the reference frequencies. With DAD, a connection wishing to increase its transmission rate alternates between using its higher and lower spectrum slots starting from its higher slots, until it reaches a slot already occupied by an upper or bottom, respectively, spectrum-adjacent connection. Then, if additional slots are needed, it expands towards the other direction, in which case the symmetry is lost. After that point, it always examines if it can expand towards the direction that uses fewer slots, using slots that were freed in the meantime by other connections. Blocking occurs if the connection needs more slots and there is no higher or lower free slot space. When a connection decreases its slots due to a reduction in its rate, we first release spectrum slots from the direction that has used more slots, and once we have an equal number of higher and lower slots, we decrease the lower spectrum slots. Thus, both expansion and contraction processes are designed to yield symmetrical spectrum utilization so that the CF of the connections does not frequently change and remains close to the reference frequency used when congestion is low.

More advanced SEC policies that consider the utilization of the spectrum-adjacent connections are proposed in [25].

## Concluding Remarks

This chapter has reviewed RSA-related optimization problems and available solving techniques that can be used to solve those problems.

The notation used along the chapter was firstly introduced, where the spectrum contiguity constraint is modelled using slots, i.e., sets of contiguous frequency slices of fixed spectral width. Slots of the needed spectral width are assigned to optical connections. This greatly simplifies the spectrum allocation problem.

After the notation, the link-path, the transponder configuration-link path, and the node-link formulations were presented associated to simple offline planning problems. The sizes of these formulations were compared and we showed that those of the link-path formulations were much lower than those of the node-link. Nonetheless, the size of those really simple problems is so high that solving methods, apart from using commercial solvers for solving the mathematical models, need to be evaluated.

In view of the above, large-scale optimization techniques and metaheuristics were presented. Column generation and Bender's decomposition were presented to deal with large-scale problem instances, whereas GRASP and BRKGA metaheuristic frameworks to provide near-optimal solutions were detailed. In addition, an algorithm to compute the RSA problem for single-connection requests was introduced.

Three illustrative use cases were afterwards presented and solved: the provisioning problem considering PLIs, the GRANDE problem to design and periodically upgrade a flex-grid network, and the elastic bandwidth provisioning, where the spectrum allocated to a connection can be adapted to follow the time-varying required transmission rate.

## References

1. L. Velasco, D. King, O. Gerstel, R. Casellas, A. Castro, V. López, In-operation network planning. IEEE Commun. Mag. **52**, 52–60 (2014)
2. K. Christodoulopoulos, I. Tomkos, E. Varvarigos, Elastic bandwidth allocation in flexible OFDM based optical networks. IEEE J. Lightwave Technol. **29**, 1354–1366 (2011)
3. Y. Wang, X. Cao, Y. Pan, A study of the routing and spectrum allocation in spectrum-sliced elastic optical path networks, in *Proceedings IEEE INFOCOM*, 2011
4. L. Velasco, M. Klinkowski, M. Ruiz, J. Comellas, Modeling the routing and spectrum allocation problem for flexgrid optical networks. Photon. Netw. Commun. **24**, 177–186 (2012)
5. M. Pióro, D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks* (Morgan Kaufmann, San Francisco, 2004)
6. O. Pedrola, A. Castro, L. Velasco, M. Ruiz, J.P. Fernández-Palacios, D. Careglio, CAPEX study for multilayer IP/MPLS over flexgrid optical network. IEEE/OSA J. Opt. Commun. Netw. **4**, 639–650 (2012)
7. CPLEX optimizer, http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html
8. A. Land, A. Doig, An automatic method of solving discrete programming problems. Econometrica **28**, 497–520 (1960)
9. M. Ruiz, M. Pióro, M. Zotkiewicz, M. Klinkowski, L. Velasco, Column generation algorithm for RSA problems in flexgrid optical networks. Photon. Netw. Commun. **26**, 53–64 (2013)
10. T.A. Feo, M. Resende, Greedy randomized adaptive search procedures. J. Global Optim. **6**, 109–133 (1995)
11. J. Gonçalves, M. Resende, Biased random-key genetic algorithms for combinatorial optimization. J. Heuristics **17**, 487–525 (2011)
12. M. Gendreau, J. Potvin, *Handbook of Metaheuristics*, 2nd edn. (Springer, Berlin, 2010)
13. L. Velasco, P. Wright, A. Lord, G. Junyent, Saving CAPEX by extending flexgrid-based core optical networks towards the edges (Invited Paper). IEEE/OSA J. Opt. Commun. Netw. **5**, A171–A183 (2013)
14. A. Castro, L. Velasco, M. Ruiz, M. Klinkowski, J.P. Fernández-Palacios, D. Careglio, Dynamic routing and spectrum (re)allocation in future flexgrid optical networks. Comput. Netw. **56**, 2869–2883 (2012)

15. J. Yen, Finding the k shortest loopless paths in a network. Manag. Sci. **17**, 712–716 (1971)
16. A. Autenrieth, J.-P. Elbers, M. Eiselt, K. Grobe, B. Teipen, H. Griesser, Evaluation of technology options for software-defined transceivers in fixed WDM grid versus flexible WDM grid optical transport networks, in *ITG Symposium on Photonic Networks*, 2013
17. M. Svaluto Moreolo, J.J. Fabrega, L. Nadal, F.J. Vilchez, G. Junyent, Bandwidth variable transponders based on OFDM technology for elastic optical networks, in *International Conference on Transparent Optical Networks (ICTON)*, 2013
18. A. Klekamp et al., Limits of spectral efficiency and transmission reach of optical-OFDM superchannels for adaptive networks. IEEE Photon. Technol. Lett **23**(20), 1526–1528 (2011)
19. O. Rival, A. Morea, Cost-efficiency of mixed 10-40-100 Gb/s networks and elastic optical networks, in *OFC*, 2011
20. K. Christodoulopoulos, P. Soumplis, E. Varvarigos, Planning flexgrid optical networks under physical layer constraints. IEEE/OSA J. Opt. Commun. Netw. **5**, 1296–1312 (2013)
21. M. Ruiz, A. Lord, D. Fonseca, M. Pióro, R. Wessäly, L. Velasco, J.P. Fernández-Palacios, Planning fixed to flexgrid gradual migration: drivers and open issues. IEEE Commun. Mag. **52**, 70–76 (2014)
22. M. Klinkowski, M. Ruiz, L. Velasco, D. Careglio, V. Lopez, J. Comellas, Elastic spectrum allocation for time-varying traffic in flexgrid optical networks. IEEE J. Sel. Areas Commun. **31**, 26–38 (2013)
23. A. Asensio, M. Klinkowski, M. Ruiz, V. López, A. Castro, L. Velasco, J. Comellas, Impact of aggregation level on the performance of dynamic lightpath adaptation under time-varying traffic, in *Proc. IEEE International Conference on Optical Network Design and Modeling (ONDM)*, 2013
24. K. Christodoulopoulos, I. Tomkos, E. Varvarigos, Time-varying spectrum allocation policies in flexible optical networks. IEEE J. Sel. Areas Commun. **31**, 13–25 (2013)
25. E. Palkopoulou, I. Stakogiannakis, D. Klonidis, K. Christodoulopoulos, E. Varvarigos, O. Gerstel, I. Tomkos, Dynamic cooperative spectrum sharing in elastic networks, in *Optical Fiber Communications Conference (OFC)*, 2013

# Chapter 5
# Transmission in Elastic Optical Networks

**Antonio Napoli, Danish Rafique, Marc Bohn, Markus Nölle, Johannes Karl Fischer, and Colja Schubert**

## Acronyms

| | |
|---|---|
| ADC | Analog-to-Digital Converter |
| ASIC | Application-Specific Integrated Circuit |
| BPSK | Binary PSK |
| BVT | Bandwidth-Variable Transponder |
| CAPEX | Capital expenditure |
| CAZAC | Constant Amplitude Zero Autocorrelation |
| CMOS | Complementary Metal-Oxide-Semiconductor |
| CS | Comb Source |
| DAC | Digital-to-Analog Converter |
| DBP | Digital Back-Propagation |
| DD | Direct-Detection |
| DFT | Discrete Fourier Transform |
| DP | Dual-Polarization |
| DSP | Digital Signal Processing |
| DWDM | Dense wavelength division multiplexing |
| ENOB | Effective Number of Bits |
| EON | Elastic Optical Networks |
| ER | Extinction Ratio |
| FEC | Forward Error Correction |
| FEC-OH | FEC-overhead |

A. Napoli (✉) • D. Rafique • M. Bohn
Coriant R&D GmbH, Munich, Germany
e-mail: antonio.napoli@coriant.com

M. Nölle • J.K. Fischer • C. Schubert
Fraunhofer Institute for Telecommunications Heinrich Hertz Institute,
Department Photonics Networks and Systems, Berlin, Germany


V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_5

| | |
|---|---|
| GN | Gaussian Noise |
| GVD | Group Velocity Dispersion |
| IFFT | Inverse Fast Fourier Transform |
| ITU | International Telecommunication Union |
| LDPC | Low-density Parity Check |
| LS | Laser source |
| MIMO | Multiple-input multiple-output |
| NWDM | Nyquist WDM |
| OFDM | Orthogonal Frequency Division multiplexing |
| OIF | Optical Internetworking Forum |
| OSNR | Optical SNR |
| PBS | Polarization beam splitter |
| PM | Polarization Multiplexing |
| PMD | Polarization Mode Dispersion |
| PSK | Phase Shift Keying |
| QAM | Quadrature Amplitude Modulation |
| QPSK | Quadrature PSK |
| RFS | Recirculating Fiber Shifter |
| ROADM | Reconfigurable Optical Add-Drop Multiplexers |
| ROSNR | Required OSNR |
| S-BVT | Sliceable BVT |
| SE | Spectral efficiency |
| SI | Spectral inversion |
| SNR | Signal-to-noise ratio |
| SPM | Self-Phase Modulation |
| SSMF | Standard Single-Mode Fiber |
| TIA | Transimpedance Amplifiers |
| WDM | Wavelength Division Multiplexing |
| WSS | Wavelength Selective Switch |
| XPM | Cross-phase modulation |

## 5.1 Introduction

Optical systems experienced an impressive capacity increase since the dotcom bubble in 2001 [1]. Within this period, channel rates grew from 10G over 40G to nowadays commercial 100G [2–4, 104]. At the moment of writing, 200G solutions have been entering the market and the first prototypes of 400G and 1T transponders are at an advanced stage of development [5–7].

Technologically speaking, this represented a stepwise transition, moving from intensity-modulated Direct-Detection (DD) systems to multilevel multiphase Quadrature Amplitude Modulation (QAM) employing coherent detection and Digital Signal Processing (DSP). The realization of current commercial optical systems has been made possible by the simultaneous presence of these three elements: DSP, advanced modulation formats, and coherent detection.

Looking back into history, the first commercial Wavelength Division Multiplexing (WDM) optical systems were installed at the beginning of the 1990s. At the end of the last millennium, public and private sectors invested massive amounts of capital in telecommunication infrastructures. The former deployed a considerable amount of optical cables. The latter invested in research and development for advanced photonics technologies. When the bubble crashed in 2001, a considerable number of companies were seriously affected; forced to either bankrupt, got acquired by others, or considerably reduced in size. Some national operators were privatized.

The aftermath of the bubble was marked by the beginning of the Internet boom. New services such as Google and YouTube entered the online market and revolutionized the way we use the internet. This greatly increased the network capacity demand, and that motivated the carriers to deploy more optical channels with higher transmission data rate over their optical network (serving as the backbone network), for increased bandwidth, by meeting such capacity demand and minimizing capital expenditure (CAPEX) at the same time. At that time (~2003) high-speed electronics could not compete against optical equalization yet and consequently long-haul optical systems were all dispersion managed (e.g., the Group Velocity Despersion (GVD) was inline compensated through dispersion compensation fiber). Research was carried out to evaluate several options to provide high bandwidth at low cost, taking account of the clear constrain that new large investments, such as deployment of novel fibers, were not possible, therefore the most of those options took a digital approach to compensate the optical impairments. These investigations initially focused on DD systems; for example CoreOptics (now Cisco) developed a commercial chip with an equalizer for 10 G/s employing the maximum likelihood sequence estimation (MLSE) algorithm [8], but the equalizer was capable of compensating only for a limited amount of distortions due to lack of information on the phase (which is typical in DD). Contemporaneously, M. Taylor's experiment [9] highlighted the powerfulness of DSP when applied to coherently detected signals (where the preserved information on the phase and amplitude enables the full exploitation of advanced compensation techniques), and the capability of full compensation of linear impairments such as GVD and Polarization Mode Dispersion (PMD) [3, 4, 10]. These research ran in parallel up to ~2007, when the first Application-Specific Integrated Circuit (ASIC) prototypes for optical coherent receivers appeared on the market; since then, DD has been relegated to applications in low-cost metro area networks.

The recent evolution of high-speed electronic devices (e.g., Complementary Metal-Oxide-Semiconductor (CMOS) technology) had made it possible to fabricate ASICs for DSP, and the significance and impact of those were quickly recognized, leading complex and specialized ASICs for optical network devices to be commercialized. The similar movement happened to high-speed Digital-to-Analog Converter (DAC) and Analog-to-Digital Converter (ADC) devices, and excellent performance in sampling rate and resolution has been achieved [11]. Today's DSP ASICs comprise up to 150 million gates, size of each gate is 20 nm or below, with maximum signal processing capacity of 400 Gb/s achieved by multicore architecture, processing 200 Gb/s per core of a dual-core processor or 100 Gb/s signals per core of a quad-core processor. Commercial DAC/ADC boards can already enable transmission symbol rates ≥32 GBd with a resolution of ~5.5

Effective Number of Bit (ENOB) and a sampling rate of 64 GSamples/s [12]. Next-generation DSP ASICs target capacities of ~1 Tb/s; for example with a quad-core ASIC and symbol rates ≥40 GBd [5].

These innovations lead to numerous advantages: multilevel multiphase modulation formats with high spectral efficiency (SE) can be now generated; component imperfections can be mitigated by the DSP ASIC; fiber linear propagation effects can be fully compensated in the digital domain, thus simplifying the link design (i.e., the complexity is moved from the link to the transponder); system performance can be significantly enhanced by employing advanced DSP algorithms. Nevertheless, not all kinds of impairments can be digitally compensated. In particular, nonlinear fiber propagation effects still represent a major challenge to increase the system capacity. Effects such as Self-Phase Modulation (SPM) and Cross-Phase Modulation (XPM), as proposed in [13–16], can be only partially mitigated.

All the above is required to cope with the actual scenario of ever-growing bandwidth needs [17] and several strategies to enlarge the SE × (distance) product [b/Hz/s km] have been proposed. Among them, it is worth mentioning the following three approaches: (a) the production of high-performance fibers (e.g., with low-attenuation and low-nonlinearities) and hybrid amplification schemes (e.g., lumped and distributed amplification) [18–20], (b) multimode or multicore propagation [21], and finally (c) the Elastic Optical Network (EON) paradigm [17, 22].

It is commonly agreed that the EON represents one of the most promising candidates for the imminent upgrade of existing fixed-grid-based optical systems, because: (1) it supports gradual hence smooth deployment; (2) it optimizes the spectrum allocation; and (3) the overall cost for network upgrade would be affordable since only two network elements, the transponder and the optical node, require an upgrade.

The EON architecture originated from mobile communications and includes principles from Software-Defined Network (SDN) technology. The main idea behind EONs is the efficient utilization of the optical spectrum throughout a network, by implementing flexible channel allocation using frequency slots with reduced sizes. For example, the frequency grid recommended by the international telecommunication union telecommunication standardization sector (ITU-T) moved from current fixed 50/100 GHz Dense WDM (DWDM) grid to a flexible one with a granularity of ≤12.5 GHz.

In addition, recent advances in the Wavelength Selective Switch (WSS) technology (e.g., Liquid Crystal on Silicon) are paving the way to the commercial production of these filters. The increased flexibility and granularity enable the design of customized spectral grids whenever new light paths are requested, as well as the multiplexing of several subcarriers to form super-channel configurations and to minimize inter-channel crosstalk [23].

With this in mind, the ITU-T has adapted the recommendations G.694.1 and G.872 to include flexibility within the ITU grid [24]. A novel DWDM concept was defined by the ITU-T study group 15, starting from recommendation G.694.1, with the formalization of the nominal central frequencies (6.25 GHz granularity), the channel width (multiples of 12.5 GHz) and the concept of frequency slots. In such a scheme, a data plane connection is switched, based on allocated and variable size frequency ranges, within the optical spectrum (media channels), providing a first brick for EONs. This technology can be exploited, for example, by multicarrier DWDM transmission through the super-channel approach.

Once the ITU grid has been standardized for flexible architecture, current transponders and Reconfigurable Optical Add-Drop Multiplexers (ROADMs) will follow the same path. In this context, the optical transponder will move toward a Bandwidth-Variable Transponder (BVT) [25, 26], which is capable to dynamically adjust the spectrum by expanding or contracting the bandwidth, on demand. This is possible by varying the number of subcarriers, the symbol rate, and the employed modulation formats based on a dynamic trade-off between reach and capacity [26]. However, when BVTs transmit at low bit-rates, a part of its capability remains unused. Hence, the concept of sliceable-BVT (S-BVT) has been introduced [25], which further increases the level of elasticity and efficiency inside the network. S-BVTs can also transmit from one point-to-multiple destinations, changing the traffic rate to each destination and the number of destinations (e.g., multiflow transmission).

These functionalities can be achieved by utilizing high-quality hardware components combined with transparent DSP algorithms. A BVT must be able to: (1) generate a large set of modulation formats spanned from Low-SE Long-Reach formats such as Binary Phase Shift Keying (BPSK) up to High-SE QAM formats (2) adapt data-, symbol-, and code-rate according to actual traffic conditions, and finally (3) be capable to propagate over different networks and distances with sufficient system margins. For instance, high-order modulation formats such a e.g. 16-QAM to 256-QAM proved to be optimal for a point-to-point connection providing ultrahigh data rate over short distances (thus maximizing the SE). On the other hand, a high-capacity transmission over transoceanic links requires robust (thus lower order) modulation formats such as Quadrature Phase Shift Keying (QPSK) or BPSK, so that the reach is maximized at the expense of SE.

This chapter discusses the transmission in the context of EONs when employing (S)-BVT architectures, and it is structured as follows. Section 5.2 describes the main system impairments and challenges for EONs by paying attention to the current technology and components. The DSP architecture is described herein together with advanced solutions for the mitigation of components and propagation impairments. Section 5.3 discusses the BVT by focusing on the generation of advanced modulation formats with variable symbol rate and their optimization. Finally, Sect. 5.4 provides an outlook on future optical systems and Sect. 5.5 draws the conclusions of this chapter.

## 5.2 System Impairments and Their Mitigation

Section 5.2.1 differentiates between component and transmission medium impairments typical of next-generation elastic optical networks. The technological limitations are highlighted together with certain characteristics of the components, the transmission medium and network elements. Next Sect. 5.2.2 introduces some advanced mitigation techniques for modern optical communication systems both in terms of component and fiber propagation compensation.

### 5.2.1 System Impairments

An optical communication system consists of three main building blocks: a transmitter, a transmission channel, and a receiver. In the context of the recent exponential growth of bandwidth demand, the requirements for these three elements will be extreme. It is foreseen that next-generation EONs will transport, over a single fiber, capacities in excess of 40 Tb/s (a fourfold increase with respect to current deployed system) to satisfy current traffic forecast [5]. Consequently, component manufacturers, system vendors, and network operators worldwide are jointly working on improving the technology for components in lowering the system costs and investigating efficient network architectures.

#### Transmitter Impairments

Next-generation BVTs (see also Sect. 5.3) will be able to generate high-order modulation formats transmitted at variable symbol rates. Consequently, the building components must meet strict bandwidth requirements so that the degradation, due to bandwidth limitation, is minimized. The residual degradations will be compensated by a DSP unit placed within the transmitter (see Sect. 5.2.2).

One of the first important devices at the transmitter is the laser. Its characteristics, required for EONs, are: tunability over the entire C-band; minimum output optical power of 16 dBm; line-width ≤100 kHz; and a frequency stability ≤1.5 GHz. Next, we find the DSP unit that includes several functions such as:

- Digital pulse shaper (e.g., in case of Nyquist DWDM transmission it digitally pre-filters the individual subcarriers to form the super-channel)
- Digital GVD pre-compensation and, in some cases, nonlinear pre-distortion of the channel; and, finally, digital pre-compensation (see Sect. 5.2.2).

After the DSP, the main building blocks we encounter are the DAC, the driver amplifier, and the IQ-Mach-Zehnder modulator.

The DAC enables the generation of multilevel high-order modulation formats. Commercial DACs use a 20 nm and below CMOS technology, with electrical bandwidth of ~16 GHz and 64 GSamples/s. A further limit of the DACs is represented by the amplitude resolution. Current DACs can deliver up to eight quantization bits, which lowers down to ~5.5 ENOB, because of the degradations caused by frequency dependence, clock jitter, etc. This last characteristic is crucial for the generation of high-order modulation formats (e.g., 128-QAM).

The DAC is followed by the driver amplifier, which sets the amplitude level for the following modulator. Moreover, it introduces several degradations such as memory effects, electrical bandwidth limitation, and nonlinear behavior. These imperfections can be mitigated by using digital pre-distortion.

The last component is the IQ-Mach-Zehnder modulator, a nonlinear device characterized by three main parameters: the bandwidth, the extinction ratio, and the nonlinear characteristic. In case, no digital pre-distortion of the modulator is applied, the driver amplifier would stabilize the input to the modulator so that it operates in the linear regime. This produces an undistorted signal, but on the other hand it significantly lowers the optical power at the modulator output, so that an additional optical amplifier before the signal enters the fiber might be required. Apart from the sinusoidal characteristic, the modulator is affected also by finite extinction ratio (caused by imperfections of modulator arms originating from the production process) that further degrades its output. Nowadays, commercial modulators achieve ER on the order of 20–25 dB, values which do not introduce penalty for low-order modulation formats (e.g., QPSK), but that severely impact schemes such as 64-QAM.

Overall, the modules above present electrical bandwidth limitations. For example, by considering commercial devices the equivalent bandwidth of this component chain is below 10 GHz. This leads to severe degradation, since one of the main features of a BVT is to transmit variable symbol rates so that either the data rate can be changed or the FEC threshold can be adapted. In both cases, the bandwidth limitation implies the usage of digital pre-compensation techniques. Finally, we would like to conclude that the generation and reception of high-order modulation formats requires an increase of the optical power and consequently of the transmitted optical signal-to-noise ratio (OSNR). This could be achieved by positioning an additional optical amplifier before the optical field enters into the fiber. In the next section, we will explain why this approach cannot be always realized.

### Channel Impairments

The optical channel is affected by two types of impairments. The first is directly related to the transmission medium, where nonideal industrial processes and materials lead to the production of imperfect fibers that degrade the transmission. The second depends on components that for example are used to compensate for propagation impairments or which are utilized to build up the network. The first are commonly denominated fiber propagation impairments. Among them, we recall attenuation, linear effects (such as GVD, PMD), and nonlinear effects (SPM, XPM, and four-wave mixing (FWM)). In addition, there exist impairments like fiber splices, mechanical stresses, physical degradations, connectors, etc. that are typical for system deployment.

During the direct-detection era, propagation impairments were compensated through customized link design and inline optical compensators. With the coherent detection and DSP, linear effects are entirely compensated in the digital domain. Nonetheless, attenuation and nonlinearities cannot be compensated in DSP. The attenuation imposes to install Erbium Doped Fiber Amplifiers (EDFA) to keep the optical power constant. Nevertheless, this solution degrades the OSNR, i.e., the quality of the transmission, since each amplifier adds up noise thus lowering the OSNR. Concerning the compensation of nonlinearities, a multitude of methods has been proposed but without clear proposal for a realistic implementation. This topic is discussed in Sect. 5.2.2.

Both impairments (accumulation of noise due to EDFAs and nonlinear effects) could be reduced by deploying new engineered fibers. For example, optical cables such as ultralow-loss fiber have recently appeared on the market. This type of fiber significantly reduces the amount of required optical amplifiers, thus increasing the OSNR and enabling error-free transmission also for high-order modulation formats. Similarly, fibers with ultralow-nonlinearities have been realized as well, so that higher launch powers can be injected and high-order modulation formats can be received with sufficient OSNR levels. Both solutions are available on the market, but their deployment is still costly. Another possibility to recover the attenuation is the utilization of Raman amplifiers, which because of the distributed amplification presents the great advantage of lowering the amount of nonlinear effects in the fiber and they introduce a low noise figure. The optimum network design combines Raman and EDFA. For example, in [27] we demonstrate the great benefit provided by these devices jointly to novel optical fibers.

A further degradation in next-generation EONs is introduced by the cascade of filters. EONs will be characterized by novel WSS-based ROADMs, which employ WSS with granularity ≤12.5 GHz and with the great advantage of optimizing the bandwidth allocation within the network. Consequently, the guard band, which is in current 50 GHz fixed systems relatively large, can be significantly reduced. On the other hand, when optical channels pass through a cascade of narrow optical filters with such configurations, they will suffer degradations that could become dominant with respect to the aforementioned fiber propagation impairments. Finally, current WSS are far from being ideal and this considerably limits the number of filters that a channel can pass through (see Sect. 5.2.3).

### Receiver Impairments

The receiver presents a less problematic hardware compared to the transmitter. Apart the ADC, needed to convert the signal from analogue-to-digital, the remaining part is dedicated to the DSP ASIC unit, which is significantly larger and more complex than the one at the transmitter.

Furthermore, the local oscillator laser requires a low line-width to enable reception of high complexity modulation formats, as pointed out for the laser at the transmitter. The photoreceiver, in particular, the incorporated transimpedance amplifier, needs sufficient bandwidth to avoid distortions in case high symbol rates are received. Finally, the ADC is affected by the same characteristics of the DAC (ENOB, electrical bandwidth, and adequate sampling rate), which becomes relevant if we increase the symbol rate and/or the order of the modulation format.

The DSP ASIC compensates for system impairments through mathematical algorithms. Current ASICs are limited by the 20 nm CMOS technology, which sets up an upper bound to the maximum number of gates for a target power consumption and physical size. This condition reduces the possibility to practically implement nonlinear mitigation techniques, such as Digital Back-Propagation (DBP), which would require an enormous number of gates [14]. A possible way of downsizing the ASIC complexity is to partially compensate for propagation impairments in the optical domain or by using better components/fibers. A further strategy would be to customize the ASIC for specific needs, e.g., lower FEC-OH for metro optical networks, and higher for ultralong-haul. The same concept could be applied to the GVD compensation. Overall, DSP ASICs for EONs must be designed in order to build up a modular system that in principle can transmit over different distances, by employing adaptive symbol rates and variable modulation formats.

### *5.2.2 Digital Signal Processing for EONs*

In EONs, the role of DSP within the BVT is twofold: (1) to compensate for transmission impairments and (2) to reduce the penalty of nonideal hardware components so that advanced modulation formats can be generated. In the first section we focus on the DSP architecture, by comparing data-aided and blind algorithms; while in the second part, we explain novel DSP algorithms utilized at the transmitter to reduce the impact of component imperfections and at the receiver to compensate for nonlinear effects.

#### Digital Signal Processing Architectures

The common steps, to a typical signal processing architecture and within a coherent receiver are:

- Resampling and correction of the imperfections of the optical front-end (de-skew and orthonormalization)
- Correction of static channel impairments (GVD compensation)
- Equalization of dynamic channel impairments (polarization rotations, PMD, timing recovery, matched filtering)
- Compensation of the frequency offset between local oscillator and signal laser, carrier phase recovery, and finally
- Symbol estimation, decoding, and FEC [10]

Using the information derived within the equalizer, several channel and signal parameters can be estimated and thus exploited for optical monitoring. Within S-BVT, the DSP algorithms must be modulation format transparent, thus enabling the flexible choice of different bit-rates and spectral efficiencies.

### Blind DSP Architectures

Conventional blind DSP receivers utilize time-domain equalization and they present only two DSP functionalities which require information about the received modulation format: the dynamic channel equalizer [10] and the carrier phase estimation [28]. The basic building blocks of a blind DSP receiver are depicted in Fig. 5.1a [29]. Since the update criterion of the dynamic equalizer depends on the modulation format [28], this approach is less attractive for modulation format transparent BVTs.

### Data-Aided DSP Architectures

The main reason behind the interest for data-aided equalization approaches is related to the possibility of periodically inserting pilot or training symbols [30–32]. The inserted training sequences, which can be equal for every kind of payload

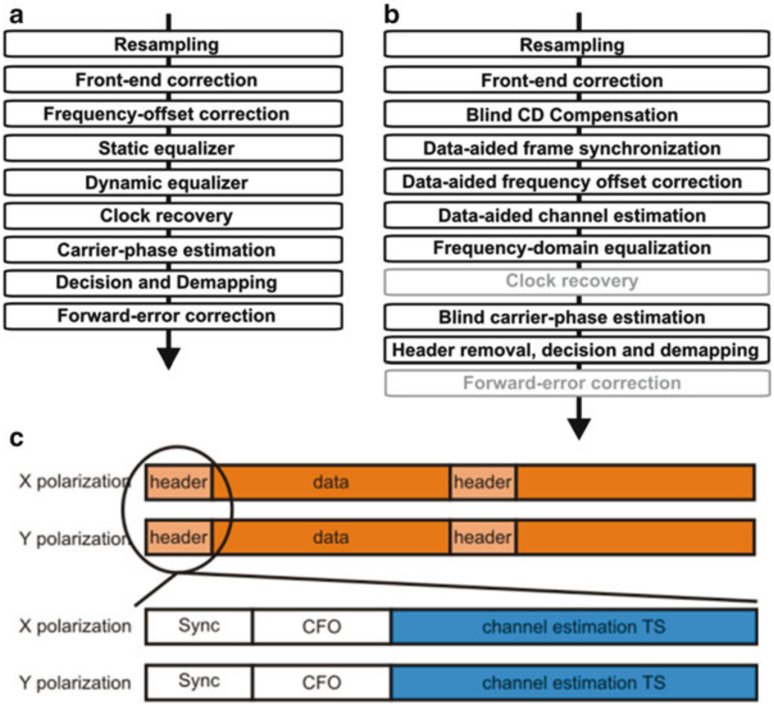



**Fig. 5.1** (**a**) DSP structure of a blind coherent receiver, (**b**) DSP structure of one exemplary data-aided DSP implementation [29], (**c**) structure and position of periodically inserted training sequences (TS) for frame synchronization (Sync), Carrier Frequency Offset Estimation (CFOE), and channel equalization

modulation format, require a small amount of overhead (typically ≤3 %) and enable format transparent channel equalization. Commonly, different types of training symbols are inserted and used for frame synchronization, Carrier Frequency Offset Estimation (CFOE), and channel estimation, respectively (see Fig. 5.1c). The DSP blocks for data-aided based coherent receivers are similar to the blind one and are reported in Fig. 5.1b. Different types of training sequences, already investigated in wireless communication systems, have recently been adopted into optics. Typically, used training sequences for channel estimation are either Constant Amplitude Zero Autocorrelation (CAZAC) [30] or pairs of Golay sequences. Comparisons between different training sequences and their performance differences for optical systems have been studied in [33]. Overall, the selection of the training sequences is still an open research topic.

## Digital Signal Processing at the Transmitter

In this section we first report the penalty induced by components when comparing system performance with commercial devices against ideal ones. In a second moment, we introduce some advanced DSP techniques to digitally pre-distort for component limitations.

### Impact of Component Quality on the System Performance

To assess this degradation at a system level, we numerically evaluated in [34] the performance for a set of advanced modulation formats when realistic values of some devices are considered. For example, the DACs are bandwidth-limited at ~18 GHz and the amplitude resolution is ~6 ENOB. Moreover, the IQ-modulator has a finite extinction ratio (ER) (i.e., ≤25 dB). These characteristics, if not compensated, will significantly reduce the maximum reachable distance. Overall, the introduced degradation grows as function of the order of the modulation format and symbol rate.

In [34] we targeted long-haul links and we selected as most suitable modulation formats, namely QPSK, 8-QAM, and 16-QAM. Afterwards, their transmission performance was evaluated for three given scenarios: *ideal* case (i.e., only fiber propagation effects are considered), *current* available (component limitations based on literature), and *within 3–4 years* (component limitations based on estimation). The numerical results were obtained by employing the software suite VPI Transmission Maker and the DSP was modeled as described in [10, 26]. Neither digital pre-compensation nor nonlinear mitigation was applied.

The numerical results were confirmed by comparison with the Gaussian Noise (GN) model. We propagated a continuous Ultradense WDM (UDWDM) spectrum with channel spacing = $1.15 \times R_s$ (where $R_s$ is the symbol rate) and with a Nyquist pulse shape and roll-off = 0.2. The same DWDM bandwidth of 1.23 THz was simulated for all cases, thus the number of super-channels varied according to the different modulation formats. An example of results is reported in Fig. 5.2a for the three scenarios with a $40 \times 200$ Gb/s 16-QAM Ultradense Wave Division Multiplexing (UDWDM) configuration. The bandwidth limitation of the individual transmitters in case of realistic components is visible when moving from the upper to the lower spectra.

Figure 5.2b shows the maximum transmittable distance vs. launch power for $\text{BER} = 1 \times 10^{-3}$. The solid black curve is obtained by using the GN approximation. From these results we observe that the ideal scenario and the one with estimated values of the components within 3–4 years achieve comparable performance, providing the clear recommendation that these values, for these components, are required.

### Advanced Digital Pre-compensation Techniques

EONs envision transmission with data rate ≥400 Gb/s at symbol rates ≥40 GBd. We know that state-of-the-art DACs present <18 GHz electrical bandwidth, values that cannot enable the generation of 40 GBd without significant back-to-back penalty [27]. Therefore, we must apply compensation at the transmitter to mitigate the electrical bandwidth limitations.

In [27], we reported the improved reach performance enabled by the joint digital pre-emphasis of the DAC and RF-driver as proposed in [12, 115]. This algorithm takes also into account, the quantization noise of the DAC and the employed symbol rate by a mean-square-error approach. We investigate several techniques, in [12, 20, 35], and all lead to OSNR gains, in back-to-back, of ~1 dB in OSNR, which substantially increases the reach [27]. In [36] we presented an alternative approach, where the constellation was designed to minimize the effect of the quantization noise, approach that might be useful for high-order modulation formats and low-amplitude-resolution DACs.

At the transmitter, besides the limitations in electrical bandwidth, other components introduce significant distortions, for instance, the driver amplifier and the IQ-modulator. In [37] we presented a novel method based on the Volterra series to digitally pre-distort the nonlinear characteristic of the transmitter (i.e., mainly from the RF-driver and the modulator). Through this method, a significant improvement of the signal quality was observed in simulations as well as in experiments for modulation formats up to 128-QAM. Moreover, we also showed that a linear pre-compensation is sufficient for low-order modulation formats, while the nonlinear one is needed for modulation formats beyond 16-QAM. Finally, in [38], we numerically evaluated a novel method based on mean-square-error for compensating the limitation introduced by a IQ-MZM with finite ER. Our simulations report a substantial improvement, either in terms of lower required OSNR or significantly decreased modulation losses.

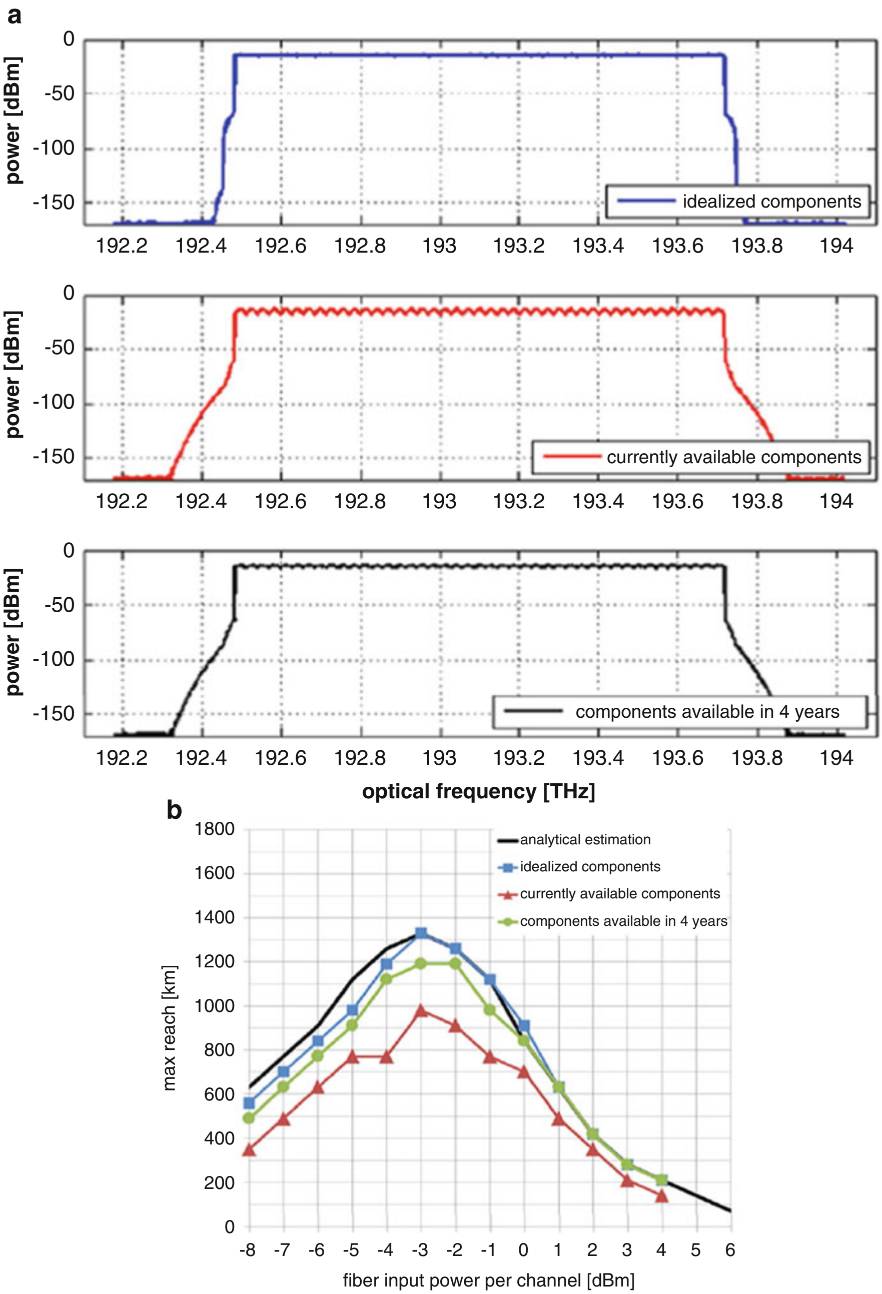



**Fig. 5.2** (**a**) 40×200 Gb/s PM-16-QAM spectra for the different considered scenarios and (**b**) reach versus power for different considered scenarios. The *black line* is obtained by utilizing the GN model

### *5.2.3 Digital Signal Processing at the Receiver*

This section presents some techniques to compensate for nonlinear effects [39], which represents one of the main limitations to further extend the system reach and to approach the Shannon limit [40]. Overall, EONs are greatly affected by nonlinear effects, because of the mix of densely packed channels, operating at variable modulation formats and data rates. The envisioned increased system capacity through EONs will be achieved at the cost of reduced transmission reach and system margins [41].

The main issues related to nonlinear fiber impairments is that distortions are induced by dynamic traffic and by the interplay between signal and noise power (e.g., from EDFAs). Both of them can be represented as stochastic processes, thus making their mitigation rather complex [42]. Overall, because of the aforementioned conditions, only a partial mitigation is possible by employing optical or electrical nonlinear compensation [13, 14, 43–45]. In particular, wide-band digital nonlinear compensation has been shown to enable significant performance margins, though it comes at the cost of prohibitive complexity. At the moment there are no cost-effective solutions to compensate for fiber nonlinearities. Nevertheless, this matter had been addressed by several groups and different methods have been proposed. Among them: Spectral Inversion (SI) [23, 46, 47], DBP [13, 14, 43], and Pilot-Based Nonlinearity Compensation (PB-NLC) [16].

**Spectral Inversion**

The concept of spectral inversion (SI) fits in nicely with the EON architecture, where SI may be employed to the super-channel, effectively cancelling any linear and nonlinear fiber impairments. The SI may be seamlessly utilized in the network using optical devices based on $\chi^2$ processes (e.g., PPLN) [48] or optoelectronic devices based on a regenerator-like solution as described in [49]. However, the advantage of a fully optical solution lies in high bandwidth of these devices, allowing up to 50 nm of spectral inverter bandwidth, compared to less than 100 GHz with currently studied optoelectronic solutions. Ideally, SI needs to be placed at the center of the transmission link and require symmetric signal profile in both halves of the link. However, recently it has been shown that power symmetry condition may be relaxed by pre-dispersing the spectral inverter [50], and that the central spectral inverter position may be relaxed using additional techniques as reported in [51]. Assuming various super-channel transmission configurations, the core concept of SI is illustrated in Fig. 5.3a. The figure shows super-channel nonlinearity compensation via SI (e.g., Intra-channel nonlinear compensation), at the center of the link, considering a ROADM site, where the WDM transmitted channels are demultiplexed, and SI can be applied to the super-channel before re-multiplexing the signals. This process effectively mitigates all linear and nonlinear fiber impairments that affected the

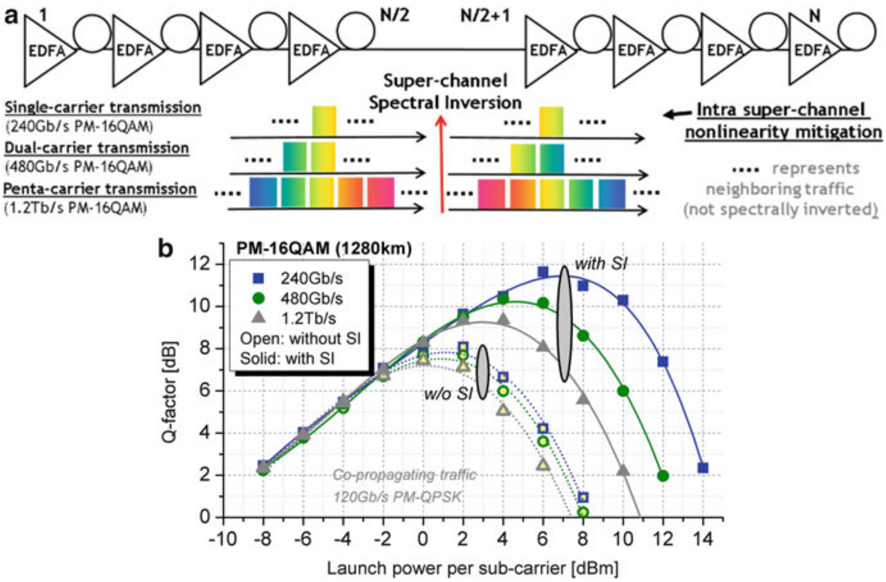



**Fig. 5.3** (**a**) Concept of a BVT super-channel spectral inversion and (**b**) Q-factor [dB] as a function of launch power per subcarrier for PM-16-QAM super-channels. *SI* Spectral Inversion, *N* number of spans

super-channel structure, essentially minimizing the impact of dynamic traffic allocation limited by channel nonlinearities. Figure 5.3b shows representative performance improvements for PM-16-QAM-based 200 Gb/s, 400 Gb/s, and 1 Tb/s super-channels, without (dashed lines) and with SI (solid lines).

### Digital Back-Propagation

DBP has been one of the most investigated nonlinear mitigation methods, being capable of compensating for both linear and nonlinear effects [13, 14]. The idea behind DBP is to invert the optical channel within the coherent DSP receiver. This method has been proposed for different scenarios ranging from single channel to DWDM as well as for uncompensated or compensated links. The numerous analyses carried out in the past years, highlighted that from one hand, DBP can extend the reach by 10–25 %, depending on the considered scenario [52, 53] and increase the optimal launch power by 1–2 dBm in case of WDM transmission. On the other hand, it presents the drawback of being rather complex, because it requires the implementation of multiple FFTs, and it is bandwidth limited, i.e., at most we can compensate only for SPM. Moreover, it might require information about the link, which is not commonly available. The remainder of this section focuses on how to limit the above drawbacks.

In [14] we presented two different methods, for uncompensated and compensated links, to decrease the complexity of the DBP method.

In case of long-haul transmission over uncompensated link, the usage of a full DBP (one DBP per span) could become unrealistic because of large number of spans. In [14] we investigated the effects of a considerable reduction of back-propagated spans and assessed the system performance. For example, we offline post-processed 11 × 112 Gb/s polarization multiplexing (PM)-QPSK channels over large effective area pure-silica core fiber (LA-PSCF) as reported in Fig. 5.4a. Clearly, the performance for the case with only 50 % digital back-propagated spans (DBP-28s) is comparable, at least for the area close to the optimal region, to the one with full DPB (DBP-56s). Moreover, the improvement against the frequency domain equalizer (FDE) is considerable.

Another interesting case of application of reduced DBP methods is represented by dispersion compensated links. In [14] we showed that this is the most likely

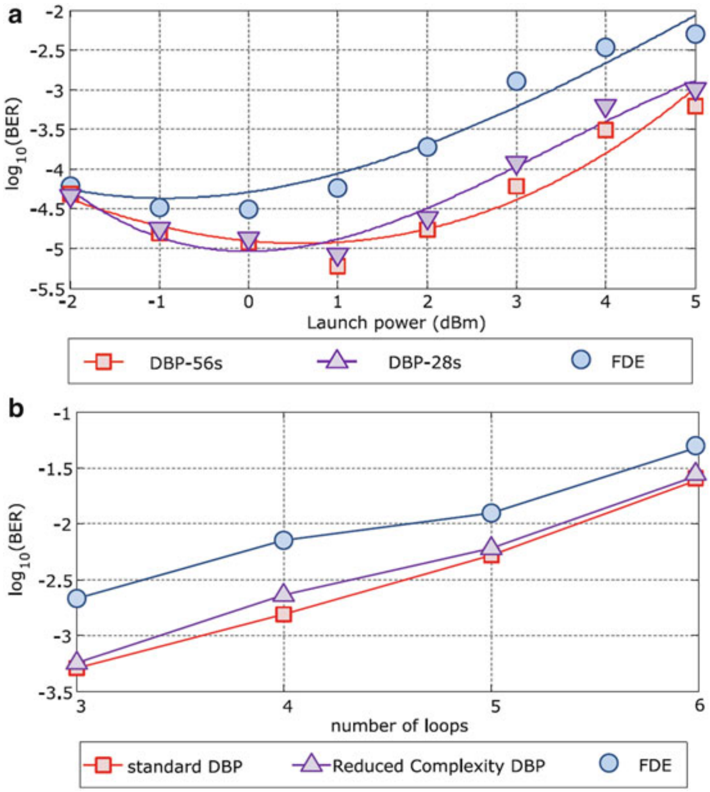



**Fig. 5.4** (**a**) $Log_{10}$(BER) versus launch power for the experimental transmission of 11 × 112 Gb/s PM-QPSK. FDE: frequency domain equalizer. (**b**) $Log_{10}$(BER) versus number of loops for the experimental transmission of 10 × 111 Gb/s PM-DQPSK. 1 loop = 475 km. Standard DBP = 1 DBP/span, reduced complexity DBP see [14]. $P_{TX}$ = 0 dBm [4]

scenario for a realistic implementation of DBP in a near future. Herein we proposed to exploit the characteristic of the dispersion map to develop an approach to lower the complexity of the DBP algorithm. The performance is reported in Fig. 5.4b and they clearly highlight that the method proposed in [14] approaches the full DBP. The results are showed for the experimental transmission of 10 × 111 Gb/s PM-DQPSK. In [14] we also compared the complexity of the standard DBP and the reduced complexity DBP finding out that for the analyzed link, the required complexity by DBP with respect to FDE would be just double. Finally, in [52] we investigated an adaptive DBP to self-derive, through signals already available within the DSP, the information required by DBP to run the algorithm.

**Radio-Frequency—Pilot Tone**

Pilot-tone (PT)-aided phase noise compensation was proposed to compensate for intra-symbol laser phase noise in coherent-OFDM systems [54]. In [16] we extended it to the mitigation of fiber nonlinearities.

The idea is to transmit an unmodulated PT whose position depends on the transmitted signals. In OFDM transmission, the PT would be placed between two subcarriers, while for single-carrier transmission it could be placed out of band as proposed in [16]. In both cases, it reduces the spectral efficiency.

The PT is filtered out at the receiver and used to estimate the phase noise that the signal experienced along the link. This method can theoretically compensate for any form of phase noise, since laser phase noise is treated identically to fiber nonlinearity. On the other hand, it presents intrinsic limitations, because it requires a guard-band so that the PT can be filtered out and only phase distortions within the bandwidth of the filter can be compensated. Furthermore, the in-band ASE noise will be also filtered together with the PT thus causing distortions.

In ASE-limited systems, the resulting SNR after pilot-based nonlinearity compensation (PB-NLC) is determined by the SNR of the signal and the one of the pilots. Therefore, the amount of power allocated to the pilot must be optimized to maximize the SNR after compensation. The optimum power and the obtainable SNR are also affected by the bandwidth of the filter employed to extract the PT, since a wider bandwidth will increase the amount of noise in the filtered pilot. The bandwidth must be optimized case-by-case [13, 16]. PB-NLC is therefore only suitable for the compensation of narrow-bandwidth phase noise. Because SPM effects are typically wider band, PB-NLC is enhanced if SPM is first compensated using the inverse model technique [16]. In [16] the PB-NLS was used in conjunction with DBP for an uncompensated link. The optimum power of the pilot was 20 dB below the signal with an optimized frequency gap of 24 GHz. The optimal filter bandwidth to extract information from the pilot was 100 MHz. PB-NLC alone extended the reach for both systems slightly and the benefit is enhanced if used after DBP. The broadband phase noise generated by SPM was compensated using DBP. The "*cleaned*" pilot gives a better estimation of narrowband phase noise generated by laser phase noise and XPM. In general, all phase compensation methods work for all sources of phase noise, such as from fiber nonlinearity or from the laser. Therefore, XPM can also be compensated using blind phase estimation techniques. However, the properties of the phase noises are different; for example, phase noise generated from XPM is mostly between 10 and 100 MHz. Therefore, the parameters will most likely need to be adjusted to compensate for XPM effectively.

### 5.2.4 Compensation of ROADMs Cascade Through Optical Pre-emphasis

Besides fiber propagation impairments, optical channels are transmitted over networks presenting different topologies and thus different degradations. For example, next-generation EONs will make large usage of filtering. When a light-path is established, a channel could traverse several WSSs before being detected, thus being degraded by tight optical filtering. In fact, a channel might transverse several ROADMs, which consists of two optical amplifiers (pre- and post-amplification) and two WSSs (at input and output) for the switching and routing of the channels.

In commercially deployed 50 GHz fixed-grid systems, optical signals do not suffer significant ISI because of add/drop and moreover, the filter penalty would most likely be negligible compared to the one of fiber nonlinearities. For example, if we considered a net symbol rate of 25 GBd, after FEC-overhead (OH), the gross rate might grow up to 34 GBd with a 16-QAM modulation and the traversing of a cascade of WSS with nominal bandwidth around 45 GHz would not degrade the signal performance. On the other hand, a way to increase the SE is to reduce the frequency slot (for instance going from 50 to 37.5 GHz, increasing the SE from 2 to 2.7 b/s/Hz). In this case, the same 34 GBd signal transmitted over a 37.5 GHz grid will suffer enormously from the traverse of filter, thus leading to a severe OSNR penalty already after a few WSS. For example the −3 dB bandwidth after 1 WSS will be ~34 GHz, and after 6 and 20 WSS, the equivalent bandwidth will drop to 25 GHz and 21 GHz respectively [55].

In [55], strategies to reduce this induced penalty, for the case of regional and national optical networks, have been investigated, by applying an adaptive optical filter in several locations of the networks to mitigate tight filtering penalties and the optical pulse shaping was obtained through an optical wave-shaper (WS). The analysis reported that the best location to significantly increase the number of hops that can be crossed is within each ROADM. This shows that the upgrade to long-haul as well as regional networks, employing flex-grid systems, would be possible. The findings were experimentally verified in [56], although the overall benefit was somewhat limited by the additional losses of the optical shaping.

## 5.3 Next-Generation Bandwidth-Variable Transponders

In order to realize EONs [22] and to exploit the capabilities of SDN [57], next-generation optical transponders must be able to allocate channels over a flexible wavelength grid [24]. This feature implies the possibility to flexibly adapt modulation formats, bit-rate and reach for a given traffic scenario. These three degrees of freedom characterize one of the key technologies enabling EON: the Bandwidth-Variable Transponder.

Figure 5.5 displays different ways to generate a super-channel with a certain data rate (the red and blue circles), by plotting the data rate as a function of the symbol rate, the modulation format and the number of subcarriers that constitute the super-channel. By varying these parameters, the data rate of a BVT and its performance over a link is varied and optimized, such that the BVT can fulfill the upcoming requests in terms of reach and capacity. The modulation format determines the maximum SE at which the BVT can operate, while the number of subcarriers and the symbol rate per subcarrier determines the total spectral occupancy and data rate. For example, a 1 Tb/s super-channel might consist of five subcarriers, modulated with 32 GBd Polarization Multiplexing (POM)-16QAM, transmits the data further than a 1 Tb/s super-channel consisting of just two subcarriers modulated with 32 GBd POM-1024-QAM. However, the latter super-channel configuration has 2.5 times higher SE occupies less bandwidth. In the remainder of this section, several options for realizing BVTs are discussed. In particular, the benefits and limitations arising from different technology options are highlighted. Section 5.3.1 describes the realization of flexible data rates by adaptive switching of the employed modulation format. Another degree of freedom, related to the employable symbol rate, is the usage of rate-adaptive Forward Error Correction (FEC) codes which is reviewed in

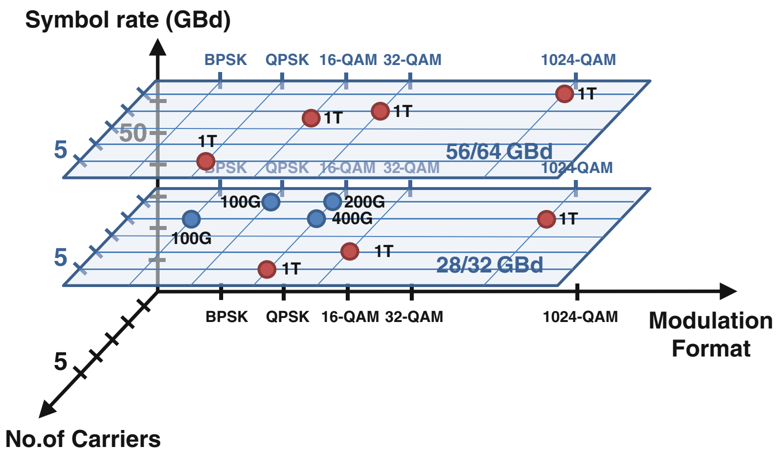



Fig. 5.5 Degrees of freedom to generate 100G, 200G, 400G, and 1T by varying the modulation format (i.e., the spectral efficiency), the number of subcarriers and the symbol rate within a BVT

Sect. 5.3.2. Finally, Sect. 5.3.3 compares several multiplexing techniques and their suitability for the realization of BVTs and the deployment of EONs.

In general, a BVT may consist of several optical modulation stages in parallel, where each stage modulates a single optical Continuous Wave (CW) carrier. Two possible options are sketched in Fig. 5.6. The optical carrier waves can be either generated by an optical Comb Source (CS) as shown in Fig. 5.6a or by individual Laser Sources (LS) depicted in Fig. 5.6b. If optical CSs are used, a wavelength selective element needs to be employed in order to separate the CW carriers, which are subsequently individually modulated. Both schemes have certain benefits and drawbacks. Optical CSs potentially offer lower costs due to a reduced number of components. Yet, there have been no reports so far of a commercial solution, which delivers on that promise. Furthermore, optical CSs produce carrier waves, which are locked in frequency. Such frequency locking may be exploited to reduce the frequency spacing between subcarriers for instance by applying advanced schemes such as offset QAM [58].

A typical setup of the transmitter stage (TX) is shown in Fig. 5.6c. It consists of a DSP block with a four-channel DAC implemented on a mixed-signal ASIC followed by modulator driver amplifiers and an electro-optic Dual-Polarization (DP) I/Q modulator. On the ASIC, a bit-rate flexible input data interface feeds the information bits to a channel encoder stage and subsequently to a flexible modulator, which supports various modulation formats. After modulation, the signal is pulse shaped. Optionally, advanced digital functions for compensation of GVD and/or Nonlinear Impairments (NL) as well as for alleviation of component limitations are performed.

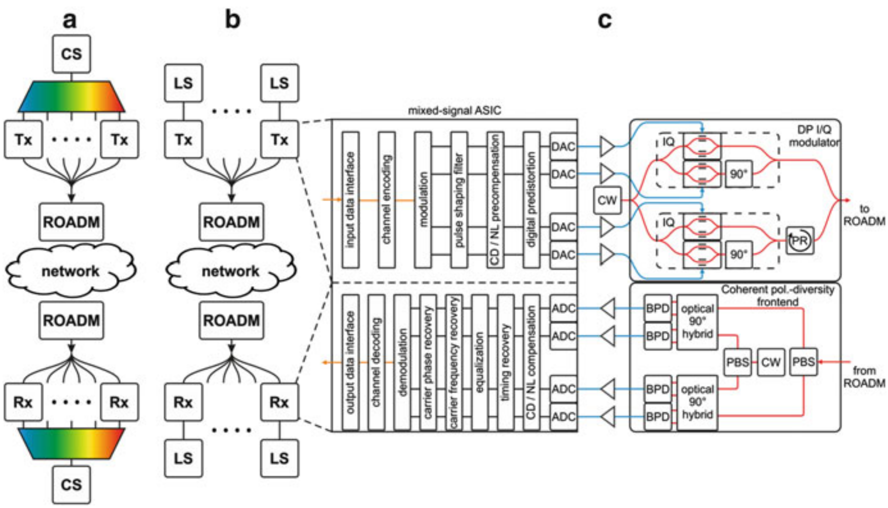



**Fig. 5.6** Schematic setup of BVTs based on (**a**) optical Comb Sources (CS) or (**b**) individual LS. (**c**) Components and DSP of the transmitter (TX) and receiver (RX) parts of a BVT

After the TX, the subcarriers are coupled and fed to the network via a ROADM. At the target destination, the super-channel is dropped by a ROADM and the different subcarriers are fed to several parallel optical receivers (RX). Transmitter and receiver optical components are already commercially available in highly integrated small form factor packages [59] and standardized, e.g., by the Optical Internetworking Forum (OIF) [60]. Usually, the receivers consist of an optical coherent polarization-diversity front-end, followed by linear Transimpedance Amplifiers (TIA) and a mixed-signal ASIC, which integrates a four channel ADC and the receiver DSP. The DSP performs functions for GVD compensation, timing, and carrier recovery as well as polarization demultiplexing, PMD compensation, and matched filtering. Mitigation of nonlinear impairments is a field of active research and it is being slowly adopted in products due to its computational complexity. After demodulation and channel decoding, a bit-rate flexible output interface forwards the received information bits.

### *5.3.1 Adaptive Choice of Modulation Format*

The choice of a particular modulation format (and thus the SE) is based on the trade-off between required transmission distance and capacity. For dispersion uncompensated links, the performance of different modulation formats can be quite precisely approximated by analytical approaches such as the GN model [61, 62]. This model considers the nonlinear distortions as additive Gaussian noise and it can be used to choose the appropriate modulation format for a particular transmission distance during the planning phase of a newly deployed link [63]. It has been proven that GN is highly accurate also when comparing it to experimental data [64].

In order to maximize the SE for any targeted link, BVTs are required to support fine granularity in terms of symbol rates, SE, and reach considering current standardization activities towards 25G ethernet, a granularity of 25 Gb/s in bit rate appears to be desirable. One option to enhance the granularity of the BVT in terms of SE is the time-domain interleaving of symbols belonging to QAM sets with different cardinality [65–67]. The resulting modulation formats are referred to as time-domain hybrid QAM and have been successfully demonstrated in several laboratory experiments [67–69]. A second option to enhance the SE granularity is to consider four-dimensional (4D) or even higher-dimensional modulation formats [26, 70]. In particular, 4D Set-Partitioning (SP) QAM formats are a suitable option due to their simple generation from POM-QAM formats [71] and their high performance [72, 73]. While keeping a constant SE, 4D SP-QAM formats have been shown to slightly outperform time-domain hybrid QAM formats [74, 75]. Fig. 5.7a shows the estimated transmission reach of typical POM modulation formats and of different 4D modulation formats at a symbol rate of 32 GBd as a function of SE. Optimal launch power and a FEC threshold of $2\times10^{-2}$ are assumed for all formats. For this analysis, a span length of 80 km of Standard Single-Mode Fiber (SSMF) without optical inline dispersion compensation was considered. The fiber parameters were: fiber loss = 0.2 dB/km, dispersion parameter = 16 ps/(nm km) and nonlinear coefficient = 1.3 (W km)$^{-1}$. Signal amplification is performed by lumped EDFAs, with a noise figure of 5 dB. The C-band (35 nm bandwidth) is filled with Nyquist-WDM (NWDM) channels. The channel spacing equals the symbol rate. The results obtained in Fig. 5.7a indicate that SEs ranging from below 2 bit/s/Hz up to almost 10 bit/s/Hz and transmission reaches between 800 km (POM-64-QAM) and 25,000 km (POM-BPSK) can be realized by switching between the different modulation formats.

Assuming a BVT that supports these various modulation formats, this allows adapting the supported net data rate to the desired reach. Figure 5.7b shows the maximum achievable net data rate as a function of reach for two different FEC scenarios ($2 \times 10^{-2}$ in blue and $3.8 \times 10^{-3}$ in red). Such a BVT is able to adapt the net bit rate from 50 Gb/s to 300 Gb/s in steps of 25 Gb/s.

### *5.3.2 Rate-Adaptive Coded Modulation*

Another option to flexibly adjust SE and net bit-rate to a specific demand is the use of rate-adaptive FEC codes or in a broader sense rate-adaptive coded modulation [76–80]. The FEC code-rate and the size of the modulation alphabet are adjusted to optimally support the desired net bit-rate for a target reach. The design parameters are: performance, implementation complexity and provided granularity (i.e., flexibility with respect to SE, net bit-rate, and reach). Reported rate-adaptive coded modulation schemes are based on hard-decision decoding using concatenated Reed-Solomon codes [77, 78] as well as soft-decision decoding using nonbinary

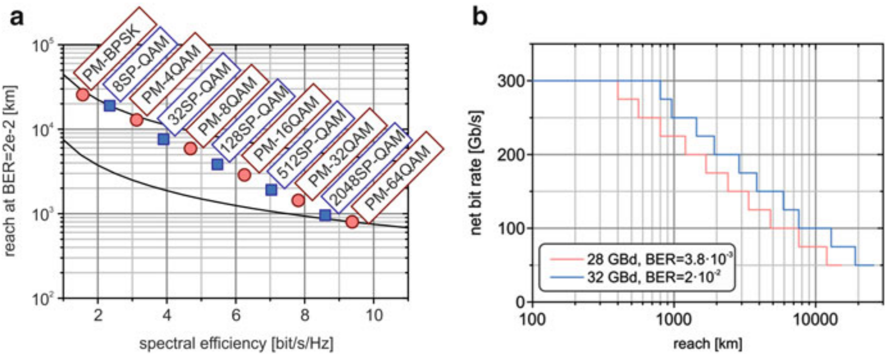



**Fig. 5.7** (**a**) Reach calculated by using the GN model versus SE for NWDM transmission of selected POM and 4D modulation formats over SSMF at a symbol rate of 32 GBd assuming SD-FEC with 23 % overhead and BER = $2 \times 10^{-2}$ threshold. Solid black lines denote the contours of constant SE × reach products belonging to the best and worst modulation format [26]. (**b**) Example of achievable net bit-rate as a function of reach for a BVT supporting the POM-m-QAM and mSP-QAM formats

low-density parity check (LDPC) codes [76, 80, 81], staircase LDPC codes [82] or concatenation of inner LDPC and outer RS code [79]. However, significant additional hardware effort is required when rate-adaptive operation can be only achieved by implementation of several FEC encoders and decoders. This may lead to an undesired increase of cost per transceiver [65]. Therefore, recent works focus on rate-adaptive coded modulation with reduced complexity. Instead of using a set of codes with different code-rates, the code-rate can also be adapted by using puncturing or shortening on a single mother code. Although this reduces the required hardware effort and allows for a fine granularity, it usually comes at the cost of a reduced performance. Other approaches employ flexible 4D bit-mapping together with concepts borrowed from polar-coded modulation and a single LDPC code [83], staircase LDPC codes [81] or Trellis-coded modulation [84]. Hardware-efficient rate-adaptive-coded modulation with suitable performance remains a highly competitive area of research.

### *5.3.3 Generation and Multiplexing of Super-Channels*

Super-channels have been recently proposed to overcome the opto/electronic bottleneck at the transmitter thus enabling the possibility to generate data rates per channel ≥400 Gb/s. In this context, different options to create super-channels through the usage of multiple optical carriers have been proposed and investigated. All these techniques are based on the multiplexing of optical carriers so that a high SE and high-capacity channels can be realized, while keeping the individual tributaries at a low bit-rate. However, the main difference of this method compared to regular DWDM is that the individual subcarriers are placed as close as possible, thus ensuring a higher SE with respect to currently deployed systems. Such a configuration implies that the individual subcarriers cannot be separated by existing ROADMs and that more transponders will be needed to fully detect data rates ≥1 Tb/s. Figure 5.8 illustrates three possible approaches to build a super-channel.

The simplest way to generate multiple optical carriers is by using individual laser sources similarly to regular WDM systems [85] (see Fig. 5.6). When using tunable laser sources, they enable a dynamic variation of the absolute frequency position and the frequency spacing between the optical lines. However, the main drawback of this method is that the individual carriers are not stable in frequency thus presenting a frequency drift with respect to the other individual lasers. A stable frequency and/or phase relation however is required for some of the below-discussed multiplexing techniques. A frequency locking of individual lasers is in principle possible, but requires, depending on the number of lasers, a complex hardware effort. This issue can be mitigated, to some extent, by DSP.

Therefore, several different methods to generate multiple optical carriers out of one single laser are being proposed.

- The first option is to use a pulsed laser source, followed by a nonlinear medium and a programmable optical filter [86]. The frequency comb of the pulsed laser

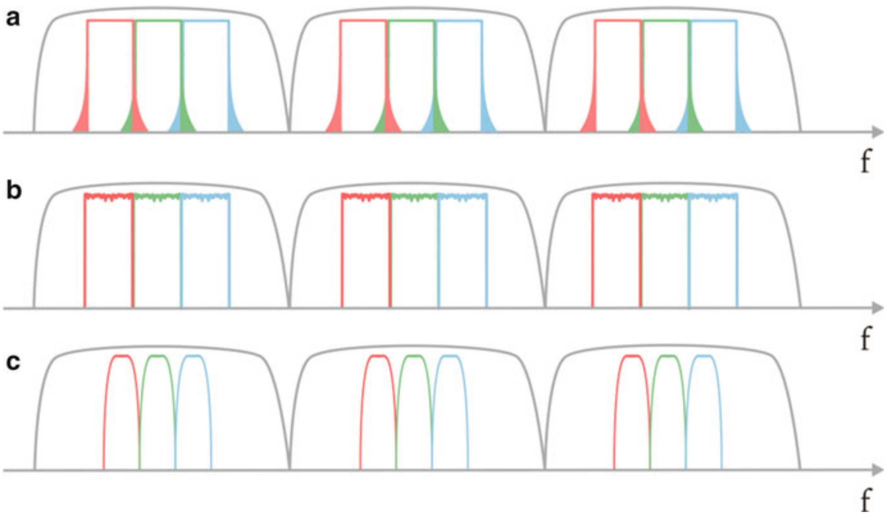



**Fig. 5.8** Illustrative examples of the spectrum of three super-channels each composed out of three subcarriers show (**a**) the spectrum for OFDM with 256 subcarriers, (**b**) the spectrum using the NWDM and (**c**) the Time-Frequency Packing approach

is broadened due to four wave mixing within the nonlinear medium, before the individual lines are equalized and the required number of lines is extracted with the optical filter. This method is able to produce several synchronized optical carriers with constant frequency spacing. The main downside is that the repetition rate of the laser (and consequently the spacing between the optical carriers) is fixed and therefore not simply tunable.

- A second option to generate a significant amount of frequency-locked carriers is to set up a Recirculating Fiber Shifter (RFS) [87]. This is a sinusoidal driven single-sideband modulator in a fiber loop configuration. A new carrier is added at each loop round trip and an optical filter within the loop is used to select the required number of carriers i.e., the overall bandwidth of the frequency comb. Similar to the first method, different carriers can be generated, but the frequency spacing can easily be varied by changing the frequency of the sinusoidally driving signal. However, the frequency spacing is limited to the electrical bandwidth of the single-sideband modulator. Another disadvantage of this approach is that it increases the OSNR of the carriers, due to the different number of round trips within the loop.
- A third commonly used method for frequency comb generation is a sinusoidally driven optical modulator optionally followed by a programmable filter in order to equalize the carriers [88, 89]. In order to create multiple optical lines, the driving signal requires a large amplitude, which can be challenging for extremely high frequencies. This method however, is only suitable to generate a moderate number of subcarriers. Similar to the RFS approach, the carrier spacing can easily be adjusted, in the range of the bandwidth of the optical modulator.

Recently, methods to produce different frequency spacing for the individual carriers have been proposed as well [90]. This, however, requires a more complex hardware setup and multiple radio-frequency generators. Clearly, all these methods present pros and cons in generating multiple optical carriers. Eventually the potential of integration and consequently the individual cost per Comb Source will decide in favor of a particular solution. Once the issue concerning the carrier frequency generation will be solved, the individual lines are separated and individually modulated by the BVT.

In order to spectrally efficient multiplex these individually modulated channels into a super-channel carrying the desired bit-rate, we can select among three different approaches: Nyquist WDM [91, 92], Orthogonal Frequency Division Multiplexing (OFDM) [85, 87, 88] and Time-Frequency Packing [93–95]. An illustrative example of the resulting spectra after multiplexing is reported in Fig. 5.8 for the three techniques.

Figure 5.8b shows the NWDM approach that employs a tight filter to reduce the required bandwidth of the signals. It is well known, that a band-limited signal can be sufficiently described by all frequencies between zero and the Nyquist frequency ($f_N$), which is given by half the symbol rate [96]. Therefore, filtering the upsampled modulation symbols with an ideal rectangular filter ranging from $-f_N$ up to $f_N$ leads to an inter-symbol-interference free signal with minimum spectral width. However, since ideal rectangular filters are not causal, a family of raised-cosine filters is used in practice. The bandwidth of this filter type can be adjusted with a single parameter called roll-off factor, which can be used to trade-off the spectral width against implementation complexity. To ensure matched filtering at the receiver, the raised-cosine filtering is often split between transmitter and receiver, leading to root-raised cosine filters at both sides. This filtering leads to well-defined narrow spectra of the individual subchannels and ideal subcarrier spacings equal to the symbol rate that are obtainable. In optical systems, this filtering can either be performed optically [92] or electrically [97, 98]. Since digital filters can be designed with higher steepness and accuracy than optical filters, the digital pulse shaping is the preferred method for generating NWDM signals. The Nyquist spectra of the individual subcarriers do not overlap in the frequency domain as illustrated in Fig. 5.8b. These channels can be separated by filtering them out at the receiver, regardless of their temporal alignment or phase relation to each other, an already substantial difference with respect to OFDM.

In contrast to NWDM, OFDM ideally uses rectangular data pulses which lead to a broad sinc-shaped modulation spectrum. However, by placing the neighboring channels exactly at the zero crossings of the sinc-shaped modulation spectrum (at a channel spacing equal to the symbol rate) and assuming proper demultiplexing (Discrete Fourier Transform (DFT) operation), crosstalk between them can be avoided. Basically, OFDM super-channels can be divided into optically multiplexed and electrically multiplexed super-channels. In this section, we will focus on electrically multiplexed OFDM super-channels only, since this option has recently gained more interest. Further information on optically multiplexed OFDM super-channels and recent demonstrations can be found in [86, 99]. In electrically multiplexed OFDM systems, the individual subcarriers are digitally generated by utilizing the Inverse Fast Fourier Transform (IFFT) operation and consist of multiple, commonly a few hundred, narrowband subcarriers (see Fig. 5.8a). The subcarriers are individually modulated onto an optical carrier with an IQ-modulator and then multiplexed with an optical coupler. Experimental demonstrations of such systems have been recently shown in [87, 89, 100]. Although the spectra of the individual subcarriers appear almost rectangular and well defined, the sinc-shaped modulation spectra of the subcarriers lead to side lobes, which can cause crosstalk between the subcarriers. To reduce the crosstalk, we can either increase the subcarriers spacing or the number of subcarriers. While the first approach decreases the SE, the latter only decreases the generated side lobes and therefore reduces the crosstalk while maintaining the SE. Another method to reduce the crosstalk between the subcarriers is to maintain the orthogonality among them [85, 87]. This however requires exact spectral and temporal alignment of the individual subcarriers and is therefore not compatible with the use of multiple independent laser sources.

A third multiplexing approach, namely Time-Frequency Packing, uses even tighter filtering than NWDM, which results in a spectral bandwidth smaller than the signals symbol rate (see Fig. 5.8c). This enables decreased channel spacing and offers a SE comparable to NWDM [93, 94], but with low-order modulation formats. However, the introduced tight filtering results in an increased inter-symbol interference (ISI), which has to be accounted for in the DSP. Finally, this method requires a sequence-by-sequence detector and special decoding schemes at the receiver instead of a simple symbol-by-symbol decision like in NWDM systems. This special treatment of the signal at the receiver drastically increases the DSP complexity.

Several efforts have been made to compare the performance of these different multiplexing techniques [64, 95, 101, 102] however a clear tendency is not yet visible. In [64], the comparison was carried out on the same identical test-bed, showing that TFP can achieve performance comparable to NWDM with:

- The advantage to generate QPSK which do not need a DAC
- The need for more complex transponder in terms of components, since it necessitates more subcarriers, and in terms of DSP, since TFP is based on sequence-by-sequence decision instead of a symbol-by-symbol like NWDM

## 5.4 Outlook and Roadmap

It is a basic demand in today's digital high-tech and mobility-oriented society that flexible and reliable communication and data connections are available anywhere at any time. The requirements on data rate, transmission distance, availability, latency, and security are rising continuously. An increasing number of communicating devices, either fixed (data centers, white goods for "Smart Living") or mobile (mobile phones, cars, trains, bicycles, clothing …) fuels this demand. In addition, industrial processes become more and more digitized to enable a flexible production on demand, which generates even more communicating entities.

To cope with the challenges, which result from these increasing requirements, the core networks will evolve towards an architecture with finer, more flexible spectral grids and flexible rate transponder arrays. Between the network nodes, parallel fibers/optical paths will be required together with new optical switching architectures to support flexible grid transmission systems. The DSP in the transponders will be the key to allow for a dynamic utilization of the available network resources, for instance by varying the data rate and bandwidth utilization, in response to dynamical network demands.

In the long-haul part of the core network, point-to-point technologies will be required, which support high spectral efficiencies and high data rates. The DSP in the transponders for long-haul systems will be a key building block. It needs to be designed to maximize the point-to-point capacity combined with FEC, utilizing coded modulation and nonlinear mitigation schemes. The FEC will remain a critical technology, specifically designed to cope with optical fiber nonlinearities in combination with the other DSP functions (coded modulation, nonlinearity mitigation). The symbol rate in the transponders is likely to increase as soon as mature and low power consumption converters between electronics and optics (modulator, photodetector) and electrical signal processing components (amplifier, ADC/DAC) become available.

For short-reach applications, such as communication intra- and inter-data centers, transponders with low complexity, cost, and power consumption will be needed. Optics will penetrate into the data centers, first for high-capacity multi-Gbps optical interfaces, utilizing VCSELs, multimode fiber or ribbon cables. Digital signal processing is likely to be used also in short-reach systems, as the complexity, power consumption, and footprint of the required ASICs is gradually decreasing. This will enable a co-design of DSP and photonics to trade performance against complexity, cost, and power consumption.

Overall, a key factor in the design of future transmission systems for all parts of the network will be the ability to multiplex and transmit multiple optical paths efficiently, by array integration of key components. The route towards integrated parallel systems is evident, but the realization at reduced cost per bit remains the big challenge. Technology improvements in the areas of optical fibers, optical amplifiers or optical signal processing can also play an important role in future elastic optical networks. In the context of elastic optical networking the improvements need to support the envisioned flexible and dynamic architectures.

## 5.5 Conclusions

The joint usage of DSP and advanced modulation formats enabled long-haul transmission at high data rates. Electronics played a fundamental role because it enables the realization ASIC engines that can recover through mathematical algorithms, several types of system impairments. DSP algorithms had been already applied to direct-detection receivers, but only with coherent systems (where the entire electrical field is available) they can significantly enhance the system performance and lower the overall costs. For instance, part of the system design can be now moved from the link to the transponder. Said this, new challenges are ahead of the optical communication world.

As previously highlighted, data traffic has literally exploded in the recent years when devices such as smartphones and tablets PC are changing the way and the position where the internet traffic is needed and generated. In addition, the amount of information and its quality is set for a sharp increase in the upcoming years. The first for example is derived from the dramatic increase of media services over the Internet, with particular attention to cloud computing; whereas the second can be referred to the advent of new video standards such as 4k/8k HD television. All these needs and requests push for a significant change within current deployed optical communication systems.

This chapter discussed the main technology options, impairments and design issues concerning transmission in the context of elastic optical networks. The EON paradigm was proposed as a straightforward ways to improve the current state of the art in a cost-wise and realizable way. In fact, EONs, in contrast to other solutions, require only the installation of new transponders and network nodes, so that the network efficiency can be maximized. In particular this chapter discusses new concepts for bandwidth-variable transponders, and challenges arising from the usage of next-generation ROADMs, where suitable solutions are proposed.

The main characteristic of a BVT is to be capable of generating several modulation formats (e.g. from BPSK to high order-QAM) at variable symbol rates (e.g. from ~30 to ~44 GBd). The first characteristic is useful to define the spectral efficiency, while the second decides the data rate and the type of correcting code (FEC) that can be used.

Within this chapter, several digital pre-distortion techniques were discussed that enable either to compensate for distortions at the transmitter and/or to increase the symbol rate. Furthermore, next-generation EONs will present spectrum, populated by channels with different transmission schemes and symbol rates, a condition where nonlinear impairments may severely limit the performance. We showed that such degrading effects cannot be fully recovered yet in the digital domain, and this represent one of the main limitations to increase the reach of modern optical communication systems. Nevertheless, BVTs will support dynamic adaptation of coded modulation formats, such that an optimum trade-off between desired distance and data rate can be obtained.

The massive use of electronics in optical communication enables flexible transmission systems and mitigation of various distortions (at the transmitter, receiver and distortions imposed by the fiber transmission). However it comes at the price of complex ASICs with high development costs and high power consumption. It remains a challenge to keep ASIC complexity and power consumption at a reasonable level by designing novel energy efficient algorithms and innovations in CMOS technology. An optimized functional split between processing in the electronic domain and in the optical domain may also prove to be a viable path for the future. This might imply either the usage of optical compensation or better fibers. A cost analysis is required to define the trade-off between the two possible solutions.

A further increase of the transmission capacity is still possible by developing and implementing lower loss, lower nonlinearity single-mode fiber, higher coding gain FEC, system architectures with flexible grids for better spectrum utilization and new amplification schemes to utilize the full bandwidth of the single-mode fiber. However, the estimated capacity increase enabled by those techniques is already rather limited and will not support a growth of the internet traffic, as we have seen it over the last decade, at a reasonable cost per bit in the future. Therefore, the development of integrated parallel systems is a possible valid way to opt for, but the realization at reduced cost per bit represents the next big challenge. The utilization of more sophisticated techniques and fibers such as multimode or hollow core fibers could become necessary. However, the challenges associated with these technologies are considerable, for example in terms of the required DSP to enable long-haul transmission in multimode fibers or the installation costs associated with the transition to new fiber types and amplification schemes.

## References


1. J. Cassidy, *Dot. con: how America lost its mind and money in the internet era* (HarperCollins Publishers, New York, 2003)
2. M.S. Alfiad et al., 111-Gb/s transmission over 1040-km field-deployed fiber with 10G/40G neighbors. IEEE Photon. Technol. Lett. **21**(10), 615–617 (2009)
3. C.R.S. Fludger et al., Coherent equalization and POLMUX-RZ-DQPSK for robust 100-GE transmission. J. Lightwave Technol. **26**(1), 64–72 (2008)
4. K. Roberts et al., 100 G and beyond with digital coherent signal processing. IEEE Commun. Mag. **48**(7), 62–69 (2010)
5. T. Rahman et al., Record field demonstration of C-band multi-terabit 16QAM, 32QAM and 64QAM over 762 km of SSMF, in *Optoelectronics and Communications Conference,* 2015
6. Y.R. Zhou et al., 1.4 Tb real-time alien superchannel transport demonstration over 410 km installed fiber link using software reconfigurable DP-16QAM/QPSK, in *OFC* (IEEE, 2014)
7. A. Pagano et al., 400 Gb/s real-time trial using rate-adaptive transponders for next-generation flexible-grid networks [Invited]. J. Opt. Commun. Netw. **7**(1), A52–A58 (2015)
8. J.-P. Elbers et al., Measurement of the dispersion tolerance of optical duobinary with an MLSE-receiver at 10.7 Gb/s, in *Optical Fiber Communication Conference*, 2005, Technical Digest, OFC/NFOEC, vol. 4 (IEEE, 2005)
9. M.G. Taylor, Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments. IEEE Photon. Technol. Lett. **16**(2), 674–676 (2004)
10. M. Kuschnerov et al., DSP for coherent single-carrier receivers. J. Lightwave Technol. **27**(16), 3614–3622 (2009)
11. W. Yan et al., 100 Gb/s optical IM-DD transmission with 10G-class devices enabled by 65 GSamples/s CMOS DAC core, in *Optical Fiber Communication Conference* (Optical Society of America, 2013)
12. A. Napoli et al., Novel DAC digital pre-emphasis algorithm for next-generation flexible optical transponders, in *Optical Fiber Communication Conference* (Optical Society of America, 2015)
13. E. Ip, J.M. Kahn, Compensation of dispersion and nonlinear impairments using digital backpropagation. J. Lightwave Technol. **26**(20), 3416–3425 (2008)

14. A. Napoli et al., Reduced complexity digital back-propagation methods for optical communication systems. J. Lightwave Technol. **32**(7), 1351–1362 (2014)
15. F.P. Guiomar et al., Fully blind linear and nonlinear equalization for 100G PM-64QAM optical systems. J. Lightwave Technol. **33**(7), 1265–1274 (2015)
16. L.B. Du et al., Digital fiber nonlinearity compensation: toward 1-Tb/s transport. IEEE Signal Process. Mag. **31**(2), 46–56 (2014)
17. A. Napoli et al., Next generation elastic optical networks: the vision of the European research project IDEALIST. IEEE Commun. Mag. **53**(2), 152–162 (2015)
18. V.A.J.M. Sleiffer et al., A comparison between SSMF and large-A eff Pure-Silica core fiber for Ultra Long-Haul 100G transmission. Opt. Express **19**(26), B710–B715 (2011)
19. J. Yu, X. Zhou, Ultra-high-capacity DWDM transmission system for 100G and beyond. IEEE Commun. Mag. **48**(3), S56–S64 (2010)
20. T. Rahman et al., Ultralong Haul 1.28-Tb/s PM-16QAM WDM transmission employing hybrid amplification. J. Lightwave Technol. **33**(9), 1794–1804 (2015)
21. V.A.J.M. Sleiffer et al., 73.7 Tb/s (96 × 3 × 256-Gb/s) mode-division-multiplexed DP-16QAM transmission with inline MM-EDFA. Opt. Express **20**(26), B428–B438 (2012)
22. O. Gerstel et al., Elastic optical networking: a new dawn for the optical layer? IEEE Commun. Mag. **50**(2), S12–S20 (2012)
23. D. Rafique et al., Intra super-channel fiber nonlinearity compensation in flex-grid optical networks. Opt. Express **21**(26), 32063–32070 (2013)
24. International Telecommunication Union, Telecommunication Standardization Sector (ITU-T), recommendation G.694
25. N. Sambo et al., Next generation sliceable bandwidth variable transponders. IEEE Commun. Mag. **53**(2), 163–171 (2015)
26. J.K. Fischer et al., Bandwidth-variable transceivers based on four-dimensional modulation formats. J. Lightwave Technol. **32**(16), 2886–2895 (2014)
27. T. Rahman et al., Long-haul terabit transmission (2272 km) employing digitally pre-distorted quad-carrier PM-16QAM super-channel, in *European Conference on Optical Communication,* 2014
28. A.J. Viterbi et al., Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission. IEEE Trans. Inf. Theory **29**, 543–551 (1983)
29. R. Elschner et al., Software-defined transponders for future flexible grid networks, in *Photonic Networks and Devices (NETWORKS 2013)*, Rio Grande, USA, 14–17 July 2013, Paper NT2C.4
30. M. Kuschnerov et al., Data-aided versus blind single-carrier coherent receivers. IEEE Photon. J. **2**, 387–403 (2010)
31. F. Pittalà et al., Data-aided frequency-domain 2×2 MIMO equalizer for 112 Gbit/s PDM-QPSK coherent transmission systems, in *Proceedings of Optical Fiber Communication Conference*, 2012, Paper OM2H.4
32. B. Spinnler, Equalizer design and complexity for digital coherent receivers. IEEE J. Sel. Top. Quantum Electron. **16**, 1180–1192 (2010)
33. M. Nölle et al., Investigation of CAZAC sequences for data-aided channel estimation considering nonlinear optical transmission, in *OFC* 2015, Paper Th3G.2
34. A. Napoli et al., On the next generation bandwidth variable transponders for future flexible optical systems, in *2014 European Conference on Networks and Communications (EuCNC)* (IEEE, 2014)
35. A. Napoli et al., Low-complexity digital pre-emphasis technique for next generation optical transceiver, in *OECC*, 2015
36. N. Markus et al., Performance comparison of different 8QAM constellations for the use in flexible optical networks, in *Optical Fiber Communication Conference* (Optical Society of America, 2014)
37. W.B. Pablo, T. Rahman, A. Napoli, M. Nölle, C. Schubert, J. Karl Fischer, Nonlinear digital pre-distortion of transmitter components, in *ECOC*, 2015

38. A. Napoli et al., Novel digital pre-distortion techniques for low-extinction ratio Mach-Zehnder modulators, in *Optical Fiber Communication Conference* (Optical Society of America, 2015)
39. G.P. Agrawal, *Nonlinear Fiber Optics* (Academic, New York, 2007)
40. A.D. Ellis, J. Zhao, D. Cotter, Approaching the non-linear Shannon limit. J. Lightwave Technol. **28**(4), 423–433 (2010)
41. J. Auge, Can we use flexible transponders to reduce margins? in *Optical Fiber Communication Conference/National Fiber Optic Engineers Conference,* 2013
42. N.V. Irukulapati et al., Stochastic digital backpropagation. IEEE Trans. Commun. **62**(11), 3956–3968 (2014)
43. D. Rafique, Fiber nonlinearity compensation: commercial applications and complexity analysis. J. Lightwave Technol. (2015)
44. D. Rafique et al., Performance improvement by fibre nonlinearity compensation in 112 Gb/s PM M-ary QAM, in *Optical Fiber Communication Conference* (Optical Society of America, 2011)
45. J. Zhao, Impact of dispersion map management on the performance of back-propagation for nonlinear WDM transmissions, in *OECC 2010 Technical Digest,* 2010, pp. 760–761
46. L.B. Du et al., Fiber nonlinearity compensation for OFDM super-channels using optical phase conjugation. Opt. Express **20**(18), 19921–19927 (2012)
47. E.F. Mateo et al., Electronic phase conjugation for nonlinearity compensation in fiber communication systems, in *Optical Fiber Communication Conference* (Optical Society of America, 2011)
48. S.L. Jansen et al., Optical phase conjugation for ultra long-haul phase-shift-keyed transmission. J. Lightwave Technol. **24**(1), 54 (2006)
49. B.-E. Olsson et al., Experimental demonstration of electro-optical mid-span spectrum inversion for mitigation of non-linear fiber effects, in *ECOC*, 2012
50. D. Rafique, A.D. Ellis, Various nonlinearity mitigation techniques employing optical and electronic approaches. IEEE Photon. Technol. Lett. **23**(23), 1838–1840 (2011)
51. H.C. Lim et al., Polarization-independent, wavelength-shift-free optical phase conjugator using a nonlinear fiber Sagnac interferometer. IEEE Photon. Technol. Lett. **11**(5), 578–580 (1999)
52. C.-Y. Lin et al., Adaptive digital back-propagation for optical communication systems, in *Tech. Digest of Optical Fiber Communications*, 2014
53. N. Antonio et al., Performance dependence of single-carrier digital back-propagation on fiber types and data rates, in *Optical Fiber Communications Conference and Exhibition (OFC)* (IEEE, 2014)
54. S.L. Jansen et al., 20-Gb/s OFDM transmission over 4,160-km SSMF enabled by RF-pilot tone phase noise compensation, in *Optical Fiber Communication Conference* (Optical Society of America, 2007)
55. T. Rahman et al., On the mitigation of optical filtering penalties originating from ROADM cascade. IEEE Photon. Technol. Lett. **26**(2), 154–157 (2014)
56. T. Rahman et al., Mitigation of filtering cascade penalties using spectral shaping in optical nodes, in *ECOC'14,* 2014, pp. 4–19
57. S. Gringeri, N. Bitar, T.J. Xia, Extending software defined network principles to include optical transport. IEEE Commun. Mag. **51**(3), 32–40 (2013)
58. J. Zhao, A.D. Ellis, Offset-QAM based coherent WDM for spectral efficiency enhancement. Opt. Express **19**(15), 14617–14631 (2011)
59. A. Beling et al., Fully-integrated polarization-diversity coherent receiver module for 100G DP-QPSK, in *Proceedings of Optical Fiber Communication Conference*, March 2011, Paper OML5
60. K. Roberts et al., Technologies for optical systems beyond 100G. Opt. Fiber Technol. **17**(5), 387–394 (2011)
61. A. Splett et al., Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities, in *ECOC*, 1993, Paper MoC2.4

62. P. Poggiolini et al., Analytical modeling of nonlinear propagation in uncompensated optical transmission links. IEEE Photon. Technol. Lett. **23**(11), 742–744 (2011)
63. P. Poggiolini, The GN model of non-linear propagation in uncompensated coherent optical systems. J. Lightwave Technol. **30**(24), 3875–3879 (2012)
64. T. Rahman et al., Experimental comparison of 1.28 Tb/s Nyquist WDM vs. time-frequency packing, in *Photonics in Switching*, 2015
65. X. Zhou et al., Rate-adaptable optics for next generation long-haul transport networks. IEEE Commun. Mag. **51**(3), 41–49 (2013)
66. D. van den Borne, S.L. Jansen, Dynamic capacity optimization using flexi-rate transceiver technology, in *Proceedings Opto-electronics Communication Conference (OECC)*, Busan, Korea, July 2012, Paper 6B4-1
67. X. Zhou et al., High spectral efficiency 400 Gb/s transmission using PDM time-domain hybrid 32–64 QAM and training-assisted carrier recovery. J. Lightwave Technol. **31**(7), 999–1005 (2013)
68. X. Zhou et al., 4000 km transmission of 50GHz spaced, 10 × 494.85-Gb/s hybrid 32-64QAM using cascaded equalization and training-assisted phase recovery, in *OFC*, 2012, Post-deadline paper PDP5C.6
69. Q. Zhuge et al., Time domain hybrid QAM based rate-adaptive optical transmissions using high speed DACs, in *Proceedings of Optical Fiber Communication Conference (OFC)*, Anaheim, USA, March 2013, Paper OTh4E.6
70. E. Agrell, M. Karlsson, Power-efficient modulation formats in coherent transmission systems. J. Lightwave Technol. **27**(22), 5115–5126 (2009)
71. L.D. Coelho, N. Hanik, Global optimization of fiber-optic communication systems using four-dimensional modulation formats, in *European Conference on Optical Communication*, September 2011, Paper Mo.2.B.4.
72. J. Renaudier et al., Experimental transmission of Nyquist pulse shaped 4-D coded modulation using dual polarization 16QAM set-partitioning schemes at 28 GBd, in *OFC*, 2013, Paper OTu3B.1
73. J.K. Fischer et al., Generation, transmission and detection of 4D set-partitioning QAM signals. J. Lightwave Technol. **33**(5), 1445–1451 (2015)
74. R. Rios-Müller et al., Experimental comparison between hybrid-QPSK/8QAM and 4D-32SP-16QAM formats at 31.2 GBd using Nyquist pulse shaping, in *ECOC*, September 2013, Paper Th.2.D.2
75. H. Sun et al., Comparison of two modulation formats at spectral efficiency of 5 bits/dual-pol symbol, in *Proceedings of 39th European Conference on Optical Communication (ECOC)*, London, United Kingdom, September 2013, Paper Th.2.D.3
76. M. Arabaci et al., Polarization-multiplexed rate-adaptive nonbinary-quasi-cyclic-LDPC-coded multilevel modulation with coherent detection for optical transport networks. Opt. Express **18**(3), 1820–1832 (2010)
77. G.-H. Gho et al., Rate-adaptive coding for optical fiber transmission systems. J. Lightwave Technol. **29**(2), 222–233 (2011)
78. G.-H. Gho, J.M. Kahn, Rate-adaptive modulation and coding for optical fiber transmission systems. J. Lightwave Technol. **30**(12), 1818–1828 (2012)
79. G.-H. Gho, M. Kahn, Rate-adaptive modulation and low-density parity-check coding for optical fiber transmission systems. IEEE/OSA J. Opt. Commun. Netw. **4**(10), 760–768 (2012)
80. M. Arabaci et al., Nonbinary LDPC-coded modulation for rate-adaptive optical fiber communication without bandwidth expansion. IEEE Photon. Technol. Lett. **24**(16), 1402–1404 (2012)
81. Y. Zhang et al., Rate-adaptive four-dimensional nonbinary LDPC-coded modulation for long-haul optical transport networks, in *Proceedings of Optical Fiber Communication Conference (OFC)*, Los Angeles, USA, March 2012, Paper JW2A.46
82. Y. Zhang, I.B. Djordjevic, Staircase rate-adaptive LDPC-coded modulation for high-speed intelligent optical transmission, in *Optical Fiber Communication Conference (OFC)*, San Francisco, USA, March 2014, Paper M3A.6

83. L. Beygi et al., Rate-adaptive coded modulation for fiber-optic communications. J. Lightwave Technol. **32**(2), 333–343 (2014)
84. E. Le Taillandier de Gabory et al., Experimental demonstration of the improvement of system sensitivity using multiple state Trellis coded optical modulation with QPSK and 16QAM constellations, in *OFC*, 2015, Paper W3K.3
85. M. Nölle et al., Techniques to realize flexible optical terabit per second transmission systems, in *Proc. SPIE 8646, Optical Metro Networks and Short-Haul Systems V,* 864602, 2013
86. D. Hillerkuss et al., 26 Tbit/s line-rate super-channel transmission utilizing all-optical fast Fourier transform processing. Nat. Photon. **5**, 364–371 (2011)
87. Y. Ma et al., 1-Tb/s single-channel coherent optical OFDM transmission with orthogonal-band multiplexing and subwavelength bandwidth access. J. Lightwave Technol. **28**, 308–315 (2010)
88. W. Shieh et al., 107 Gb/s coherent optical OFDM transmission over 1000-km SSMF fiber using orthogonal band multiplexing. Opt. Express **16**, 6378–6386 (2008)
89. X. Liu et al., 448-Gb/s reduced-guard-interval CO-OFDM transmission over 2000 km of ultra-large-area fiber and five 80-GHz-grid ROADMs. J. Lightwave Technol. **29**, 483–490 (2011)
90. N. Sambo et al., First demonstration of SDN-controlled SBVT based on multi-wavelength source with programmable and asymmetric channel spacing, in *ECOC*, 2014, Paper We.3.2
91. J.-X. Cai et al., Transmission of 96×100G pre-filtered PDM-RZ-QPSK channels with 300% spectral efficiency over 10,608 km and 400% spectral efficiency over 4,368 km, in *OFC*, 2010, Paper PDP B10
92. G. Gavioli et al., Investigation of the impact of ultra-narrow carrier spacing on the transmission of a 10-carrier 1Tb/s superchannel, in *Proceedings of Optical Fiber Communication Conference*, USA, March 2010, Paper OThD3
93. A. Barbieri et al., Time-frequency packing for linear modulations: spectral efficiency and practical detection schemes. IEEE Trans. Commun. **57**(10), 2951–2959 (2009)
94. M. Secondini et al., Optical time-frequency packing: principles, design, implementation, and experimental demonstration, ArXiv e-prints (2014)
95. A. Barbieri et al., OFDM versus single-carrier transmission for 100 Gbps optical communication. J. Lightwave Technol. **28**, 2537–2551 (2010)
96. C. Shannon, Communication in the presence of noise. Proc. Inst. Radio Engrs. **37**, 10–21 (1949)
97. R. Cigliutti et al., Ultra-long-haul transmission of 16×112 Gb/s spectrally-engineered DAC-generated Nyquist-WDM PM-16QAM channels with 1.05×(symbol-rate) frequency spacing, in *OFC*, 2012, Paper OTh3A.3
98. R. Schmogrow et al., 150 Gbit/s real-time Nyquist pulse transmission over 150 km SSMF enhanced by DSP with dynamic precision, in *Proceedings of Optical Fiber Communication Conference*, USA, March 2012, Paper OM2A.6
99. S. Chandrasekhar et al., Transmission of a 1.2-Tb/s 24-carrier no-guard-interval coherent OFDM superchannel over 7200-km of ultra-large-area fiber, in *ECOC*, 2009, Paper PD 2.6
100. R. Dischler, F. Buchali, Transmission of 1.2 Tb/s continuous waveband PDM-OFDM-FDM signal with spectral efficiency of 3.3 bit/s/Hz over 400 km of SSMF, in *OFC*, 2009, Paper PDPC2
101. S. Jansen et al., Optical OFDM, a hype or is it for real?, in *ECOC*, 2008
102. R. Freund et al., Single- and multi-carrier techniques to build up Tb/s per channel transmission systems, in *International Conference on Transparent Optical Networks (ICTON)*, 2010
103. E. Ip et al., Coherent detection in optical fiber systems. Opt. Express **16**(2), 753–791 (2008)
104. M.S. Alfiad et al., A comparison of electrical and optical dispersion compensation for 111-Gb/s POLMUX–RZ–DQPSK. J. Lightwave Technol. **27**(16), 3590–3598 (2009)
105. C. Schubert et al., New trends and challenges in optical digital transmission systems, in *Proc. 38th European Conference on Optical Communication (ECOC)*, September 16–20, 2012, Amsterdam (Netherlands), Paper We.1.C.1
106. A.V. Tran et al., 8×40-Gb/s optical coherent pol-mux single carrier system with frequency domain equalization and training sequences. IEEE Photon. Technol. Lett. **24**(11), 885–887 (2012)

107. X. Zhou et al., 12,000 km transmission of 100GHz spaced, 8 × 495-Gb/s PDM time-domain hybrid QPSK-8QAM signals, in *Proceedings of Optical Fiber Communication Conference*, Anaheim, USA, March 2013, Paper OTu2B.4
108. M. Karlsson, E. Agrell, Which is the most power-efficient modulation format in optical links? Opt. Express **17**(13), 10814–10819 (2009)
109. D.S. Millar, S.J. Savory, Blind adaptive equalization of polarization-switched QPSK modulation. Opt. Express **19**(9), 8533–8538 (2011)
110. S. Alreesh et al., Blind adaptive equalization for 6PolSK-QPSK signals, in *ECOC*, 2013, Paper Mo.4.D.3
111. M. Jinno et al., Multiflow optical transponder for efficient multilayer optical networking. IEEE Commun. Mag. **50**(5), 56–65 (2012)
112. N. Amaya et al., Introducing node architecture flexibility for elastic optical networks. J. Opt. Commun. Netw. **5**(6), 593–608 (2013)
113. A. Stavdas et al., A novel architecture for highly virtualised software-defined optical clouds, in *ECOC*, Mo.3.E.4, 22–26, London, UK, September 2013
114. M. Nölle et al., Transmission of 5×62 Gbit/s DWDM coherent OFDM with a spectral efficiency of 7.2 Bit/s/Hz using joint 64-QAM and 16-QAM modulation, in *OFC*, 2010
115. A. Napoli et al. Digital Compensation of Bandwidth Limitations for High-Speed DACs and ADCs, IEEE/OSA journal of lightwave technology, 2016

# Chapter 6
# Node Architectures for Elastic and Flexible Optical Networks

**Georgios Zervas, Emilio Hugues-Salas, Tanya Polity, Silvano Frigerio, and Ken-Ichi Sato**

## Acronyms

| | |
|---|---|
| AoD | Architecture on Demand |
| BPSK | Binary Phase Shift Keying |
| BV-WSS | Bandwidth Variable Wavelength Selective Switch |
| EDFA | Erbium-Doped Fiber Amplifier |
| FPGA | Field Programmable Gate Array |
| LAN | Local Area Network |
| LCoS | Liquid Crystal on Silicon |
| MEMS | Micro Electro Mechanical Systems |
| NFP | Network Function Programmability |
| NFV | Network Function Virtualization |
| OAM | Operation, Administration, and Management |
| OAM | Orbital Angular Momentum |
| ODU | Optical channel Data Unit |
| ODUflex | Flexible Optical channel Data Unit |

G. Zervas (✉) • E. Hugues-Salas
University of Bristol, Bristol, UK

T. Polity
University of Peloponnese, Peloponnese, Greece

S. Frigerio
Alcatel-Lucent, Vimercate, MB, Italy

K.-I. Sato
Nagoya University, Nagoya, Japan



V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_6

| | |
|---|---|
| OOS | Optical Overhead Signal |
| OTN | Optical Transport Network |
| OUT | Optical Transport Unit |
| OXC | Optical Cross-Connect |
| PDM | Polarization Division Multiplexing |
| QAM | Quadrature Amplitude Modulation |
| QPSK | Quadrature Phase Shift Keying |
| ROADM | Reconfigurable Optical Add Drop Multiplexer |
| RWA | Routing and Wavelength Assignment |
| SDH | Synchronous Digital Hierarchy |
| SDM | Space Division Multiplexing |
| SDN | Software-Defined Network |
| SONET | Synchronous Optical Networking |
| SSS | Spectrum Selective Switch |
| VLAN | Virtual LAN |
| WSON | Wavelength switched optical networks |

This chapter covers all critical aspects required to design node architectures able to support elastic and flexible optical networks. It first details the key requirements, design rules, and criteria for node architectures including key hardware components under Sect. 6.2. The next two Sects. 6.3 and 6.4, report on the two key subsystems designs, the bypass/express path subsystem, and the multi-layer add/drop transport technologies involving Optical Transport Network (OTN) with fixed and flexible optical layer. Last but not least, Sect. 6.5 reports on multidimensional and function programmable optical node architectures that propose more disruptive and long-term architectures where optical layer programmability plays a key role.

## 6.1 Introduction

As traffic demands become more uncertain and newer services continuously arise, novel network elements are needed to provide more flexibility, scalability, resilience, and adaptability to today's optical networks. Considering these requirements, next generation optical node architectures to support and address elasticity and flexibility in optical networks are required. As an evolution of existent Reconfigurable Optical Add Drop Multiplexers (ROADM) and Optical Cross-Connects (OXC), these optical nodes will establish a new paradigm in which the network requirements will be efficiently addressed considering various emerging dimensions. In this chapter, all required subsystems that are needed to construct a multi-layer flexible optical node. A high-level architecture is illustrated below (see Fig. 6.1).

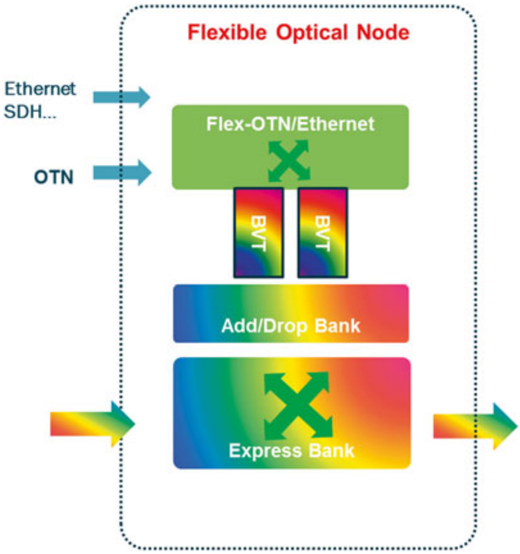



**Fig. 6.1** High-level multi-layer flexible node architecture

## 6.2 Requirements, Design Rules, and Criteria for Next Generation Flexible Optical Nodes

### *6.2.1 Literature Review on Legacy Architecture*

The notion of optical networking is based on automatic setup and tear-down of lightpaths that connect ingress and egress nodes in WDM optical transport systems. In order to achieve this functionality at the optical layer, whether ITU transmission grid is utilized or not, an optical node needs to efficiently and appropriately switch the corresponding optical channel. The first generation of optical transport systems were single wavelength transmission configurations where all switching and higher network layer functions were performed in the electronic domain, and examples of such networks are SDH networks deployed in Europe and SONET networks deployed in North America. The second generation of optical systems allowed multi-wavelength transmission via WDM technology. However, these high-capacity WDM systems only offer capacity increase and point-to-point connectivity in SDH/SONET ring networks. This means that an SDH/SONET ring comprised numerous WDM spans and corresponding digital nodes, where all traffic is terminated and passes through an electronic switch fabric, and that the only functions that were performed at the optical layer were *wavelength multiplexing/demultiplexing*. The third generation came with the advent of WDM technology deploying fixed filters that enable *optical bypass* and *wavelength add/drop*, known as Optical Add Drop Multiplexers (OADMs). The introduction of remotely reconfigurable OADM (ROADM) technology represents the fourth generation of optical systems, where optical wavelengths can be dynamically switched within an optical network, the so-called lightpaths. Now optical channels can be flexibly added and dropped by network operators through *wavelength switching* functions built into the transport node, in order to meet the everchanging bandwidth demands, scalability challenges, and unpredictable traffic patterns. The optical network topology becomes mesh at the wavelength level, as long as the multi-degree ROADMs are incorporated [1], designating a new era for optical networking.

As WDM transmission is future proofed through flex-grid technology, a traditional ROADM-based network will deploy flex-grid switching technology in order to support elastic networking, for the aim to capitalize on the trade-off between spectral efficiency and transmission reach. In elastic optical networks, superchannels that require shorter reaches deploy spectrally efficient modulation formats so that spectral resources are released for transporting high-capacity channels over longer distances. This superchannel connectivity is referred to as *flexible/variable spectrum switching*, and a new generation of multi-degree *flexible optical nodes* should be able to cope with such variable spectrum assignment and switching. Implementing this next generation flexible optical nodes requires new approaches of how optical devices implement specific optical functions on one hand while efficiently utilizing coherent technology to alleviate filter requirements on the other. Evidently a new generation of variable bandwidth components and devices is needed (e.g., WSS, switches, etc.)

A **flexible optical node architecture** comprises an express and an add/drop part (see Fig. 6.1) departing from the current generation of ROADMs. The legacy express part is based on a Broadcast-and-Select architecture comprising one splitter and a single WSS per ROADM degree. The splitter is used to provide a copy of the optical spectrum to all outputs while the WSS selects which parts of the spectrum (superchannels) are to express through that designated output, and optically blocks the rest [2]. Route-and-Select architectures that include two WSSs in tandem per ROADM degree have been proposed in order to reduce splitting losses and the required wavelength isolation of deployed WSS [3]. Although the core switching functions are colorless, the add/drop part comprises a "colored" (wavelength-specific) component either to distribute the non-express wavelengths onto corresponding drop points or to add traffic locally. Specifically in a fixed-grid ROADM, a single multiplexing device per degree combines all locally added channels aiming at the specific degree/output. Channel demultiplexing is generally accomplished using the same elements in the reverse direction imposing an extra WSS degree for this functionality.

Colorless and directionless designs put some more stringent criteria at the add/drop part. One cost-effective implementation of colorless and directionless demultiplexing for the flexible optical node relies on a combination of filterless power splitters and coherent receiver technology. Splitters provide copies of the dropped superchannels to several receivers, and the local oscillator at the receiver selects its corresponding channel. Another promising approach replaces the splitter with an extra WSS, limiting the part of spectrum that is dropped locally. However, both configurations suffer from wavelength contention at the add/drop end. One way to avoid this is to use a multicasting switch, with various configurations that are discussed in this chapter.

### *6.2.2 Key Enabling Technologies*

Developing flexible bandwidth ROADMs means implementing a new generation of optical devices that can support variable bandwidths and flex-grid spectrum allocation: WSSs, multicast switches and transponders together with application-specific optical amplifiers and monitors.

#### Wavelength Selective Switches

First-generation WSSs (based on Micro Electro Mechanical Systems (MEMS) [4] and/or liquid crystal [5] technologies) allocate a single switching element (mirror/pixel) to each channel which means that the channel bandwidth and center frequency are fixed and cannot be altered on demand. Additionally, many designs of first-generation WSS exhibit spectral dips in their transmission spectrum among spectrally adjacent channels due to the limited spectral “fill factor” [2]. This prevents concatenation of spectrally adjacent channels to create a superchannel that can be switched as an entity. In order to fully exploit the potential of elastic optical networks, superchannel bandwidth and center frequency should be dynamically configured. Second generation WSSs, based on Liquid Crystal on Silicon (LCoS) [6] or 2D MEMS arrays [7] permit dynamic control of channel center frequency and bandwidth through configuration of internal pixel panels [2]. These parameters can be finely controlled and dynamically set. The ability to independently control the center frequency of a channel is better than 1 GHz resolution in an LCOS-based WSS [2]. Another important feature of WSS is the support for larger number of degrees (ports). For current networks, this is, typically a $9\times1$ WSS, although $4\times1$ WSS are used in some smaller nodes.

#### (Sliceable) Bandwidth Variable Transponders

A component of immense importance is the transponder that lies at the add/drop end of the optical node [3]. The transponder technology and design vary. In order to capitalize on the spectral efficiency and reach trade-off to enhance elasticity in the networks, transponders need to support various modulation formats [8]. To give an example, currently coherent technology-based transmission systems have been proposed with modulation formats spanning from Dual Polarization-Binary Phase Shift Keying (DP-BPSK), DP-QPSK, and DP-16QAM formats to optical orthogonal frequency-division multiplexing (OFDM). The former three formats can generally be generated and received using the same transmitter and coherent receiver hardware designs [9]. A flexible (variable bandwidth/variable modulation format) transponder requires a finely tunable laser that can be tuned to the flexi-grid, at the transmitter side, and a coherent receiver with an appropriate-linewidth local oscillator. In order to support upgradability of modulation formats, it also requires an appropriately flexible and coherent pre- and post-processor with digital-to-analog converters so as to create the appropriate electrical modulator drive signal and receive the signal at the receiver end. Minimal alterations may be necessary within the coherent receiver resulting in a single transponder that can be provisioned to any of the three formats [10].

#### Multicasting Switch

A multicast switch resolves contention at the wavelength level at the drop point and can be comprised of a set of splitters combined with optical switches. Assuming that we need $M$ degrees to be able to reach the $N$ receiver ports, the traffic from each fiber is transmitted via a $1 \times N$ splitter to one port of an $M \times 1$ switch, which allows only the channels originating from a specific degree to reach the respective coherent receiver. In that sense, the multicasting switch (MCS) resolves wavelength contention [10].

### *6.2.3 Design Rules*

This section describes the design characteristics that a node should exhibit in order to be an appropriate candidate for a flexible node architecture.

#### Flexibility

The key property of a flexible ROADM is to independently, dynamically, and flexibly switch the variable bandwidth superchannels among the input and output ports and this translates into controlling the center frequency and spectral bandwidth of each port in the flexible optical node. With respect to the legacy ROADM architecture this means that both the WSS, which is the main colored element and resides at the heart of the ROADM, and the transponders that interface network clients should be able to operate in a flexible mode, in other words, to control the channel bandwidth and center frequency with GHz resolution. Advanced modulation formats allow for fine control of the optical spectrum, hence combining those with the flexible ROADM allows for new network paradigms like elastic optical networks. The exact degree of flexibility required for the resolution of channel position and bandwidth will be imposed by the complexity of the optical network solution that will comprise flexible optical nodes. Currently a resolution of 6.25 GHz for the channel position and 12.5 GHz for the bandwidth have been proposed.

Based on the aforementioned flexibility, the term *flexible node* has arisen. With this respect, an optical node within a network is considered flexible if it is able to cope with change, adapting dynamically to the modifications in the network caused by variations in traffic demands and network requirements. Thus, a measurement of flexibility has been proposed based on entropy maximization [11] and several types of flexibility in a node are recognized, such as:

- *Channel flexibility*: the node may support various bit rates and modulation formats
- *Expansion flexibility*: the node may be easily scaled to support higher traffic
- *Functional flexibility*: the node may be able to provide many different functions including spectrum defragmentation, format conversion, regeneration, etc.
- *Switching flexibility*: the node is able to map inputs to outputs in different ways and across various dimensions
- *Routing flexibility*: the node is capable of carrying signals from input to output along different routes
- *Architectural flexibility*: the node is capable of rearranging its building blocks in order to construct different architectures.

**Reconfiguration**

Supporting flexible-spectrum, and colorless and directionless multiplexing and demultiplexing requires new implementation approaches. Reconfiguration time of an optical node, however, is currently imposed by the optical network services and limited by the technology in hand. For flexible optical networks reconfiguration time is dictated by the reconfiguration time of the optical node and the flexible transponders. In order to offer upgradability without service disruption, reconfiguration time of milliseconds level should be available for these technologies.

**Colorless, Directionless, and Contentionless**

To enable flexibility and reconfigurability of the flexible optical nodes, flexible Colorless, Directionless, and Contentionless (CDC) architectures with remote spectrum assignment have been suggested.

- **Colorless architectures** comprise architectures that allow any wavelength or (super) channel to be dropped/added locally, permitting the connectivity of any wavelength channel to an add/drop point. Specific colorless architectures are discussed later in this chapter. Regardless of the configuration, colorless node architectures allow any wavelength to reach the add/drop point by automatically reconfiguring the ROADMs.
- **Directionless architectures** permit wavelength to be routed to any direction served by the node, as directed remotely by the control plane. This facilitates real flexibility and automation with respect to second generation ROADMs, where a single, fixed direction over which an outbound wavelength could be routed was already decided in advance. Contention can occur when two colors request the same output/direction.

- **Contentionless architectures** ensure that no wavelength blocking can occur in the case that two same colors coming from different fibers request the same output fiber, a scenario that may occur in colorless and directionless ROADMs as add/drop traffic is rerouted.
- **Colorless, Directionless, and Contentionless** ROADMs offer the same advantages as a colorless and directionless ROADM, and also relies on colorless ports to allow several superchannels of the same color to be added or dropped rather than only one.

#### Scalability: Modularity

The main reason behind introducing flexible ROADMs is to future proof the fixed-grid WDM networks. In order to guarantee that optical network scalability can be achieved, it must be ensured that node scalability is inherent in the node design. An optical network's capability to scale is typically defined by several of the optical network parameters, for example, the number of WDM channels/superchannel per fiber, capacity per fiber, number of fibers and the resulting number of optical nodes that can be reached, in order to support future traffic growth. In that sense flexible optical node architectures should be scalable in a similar manner. This means that flexible optical node architecture should have the ability to support growth with respect to the capacity per fiber and the number of ports/degrees which is related to both the number of fibers and the number of nodes that can be connected to a specific architecture. A modular flexible node architecture should be able to grow both in capacity and degrees without disrupting the initial configuration.

#### Resilience

Optical node resilience which relies on the ability to return to the original operation after a failure or a fault should be incorporated in the node architecture so that reconfiguration of the node and hence the network can be performed seamlessly in less than 50 ms.

### 6.2.4 Performance Criteria

#### Capacity/Throughput

Flexible optical nodes would be able to switch wavelength channels that occupy varying amounts of bandwidth with different data rates or modulation formats. The amount of maximum traffic that can be switched in a specific node architecture is dictated by the *ROADM degree* and *spectral efficiency* of the optical flex-grid transmission system and is denoted as the node capacity. Therefore, considering a colorless, directionless, and contentionless architecture, the capacity of the nodes is proportional to the number of input/output ports and the maximum amount of traffic that can be switched through the ROADM.

However, in realistic optical networks, the actual amounts of data that are switched through a node in a second (which indicates the node throughput) can be limited due to the node capacity. Routing of heterogeneous connections through the network adds a bandwidth continuity constraint to the existing requirement for wavelength continuity, potentially causing inefficiency over the network and the actual switch capacity. Optical node throughput must be estimated by taking into account the spectrum assignment policy that takes place at the control plane, along with any spectrum defragmentation that may occur in the node.

### Switching Granularity

Flex-grid optical transmission systems have been proposed for higher-spectral efficiency and elasticity of the optical layer. However, to take full advantage of this new possibility, full access, and manipulation of the finer spectral entity (i.e., spectral slot) should be available not only at the receiver or transmitter end but at the intermediate switching nodes. The main obstacle up to now for this spectral packing was the filtering required to switch adjacent WDM channels, which has been alleviated with the use of matrix-based switching (LCoS, etc.).

In flex-grid optical networks, two or more channels or subcarriers can be managed as a single entity, named as a *superchannel* or a multi-carrier, with a one-to-one mapping between a superchannel and a single-client interface. For instance, a 400 GbE client may be transported using $2 \times 200$ Gbit/s DP-16QAM carriers, which are considered a superchannel.

The subchannels comprising a superchannel do not necessarily need to be contiguous in wavelength, but need to be switched as one entity by the flexible optical node. The finer switching entity that the optical nodes can support is generally considered the same with the finer spectral slot of the optical transmission system; hence, the channel center frequency should be defined with 6.25 GHz resolution and the channel width with 12.5 GHz resolution.

### Physical Performance

The endeavor to continue increasing spectral efficiency faces some fundamental challenges, as previously mentioned. One of the primary limitations on the reach of a given signal is the degradation of its optical signal-to-noise ratio (OSNR) as it propagates through the transport system. Therefore, OTNs that reduce the rate of this degradation could transport signals further while maintaining a similar OSNR, enabling the transmission of higher-spectral efficiency formats. Although this is still an issue for the transmission system, OSNR requirements become stringent when the signal passes through a number of optical nodes, which will be the case when transmitted into the flex-grid optical network. Therefore, a good concatenation performance translates into especially higher OSNR requirements which, in turn, means less interference/crosstalk at the node leading to higher WSS isolation and less splitting requirements.

At the transmission level exploiting Raman amplification to compensate some of the transmission fiber loss in combination with Erbium-doped fiber amplifiers (EDFAs) can increase the OSNR of a signal by roughly 3–5 dB within a link depending on the fiber type, quality, and Raman pumping scheme deployed [3]. At the node however, stringent OSNR requirements translates into low crosstalk WSS. Typical isolation values in a 9-port WSS lies in the range of 35–45 dB, which means that crosstalk still accumulates as the channel goes through successive nodes [6]. Assuming coherent detection, this crosstalk together with the optical noise generated by optical amplification, is a primary source of signal degradation that inherently limits the reach of the transmission system with the actual crosstalk margin depending on the modulation format. Constructing high-port-count flexible-spectrum WSS with higher isolation is an open issue and for that only new architectural solutions must be sought, for example the Route-and-Select configuration with respect to the broadcast-&-select.

In any case, the architectural limitations of the flexible optical node should be resolved at the network level where concatenation and impairment accumulation are taken into account. This is enhanced when variable modulation formats with variable OSNR versus reach trade-offs are used.

## 6.3 Bypass/Express Node Architectures

### *6.3.1 Express Switch Architectures*

#### Brief History of OXC Development

Figure 6.2 shows that the generic OXC node architecture consists of the express switch part and add/drop switch parts. Various OXC switch architectures were investigated in the early 1990s, some of which are summarized in [11, 12]. Early studies mostly focused on the express switch part, since the first motivation for introducing OXCs was to eliminate expensive Optical/Electrical (OE) and Electrical/optical (EO)

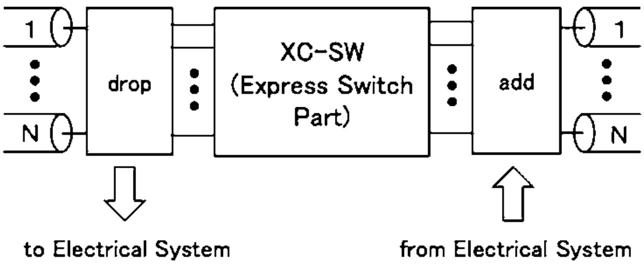



**Fig. 6.2** Generic OXC architecture

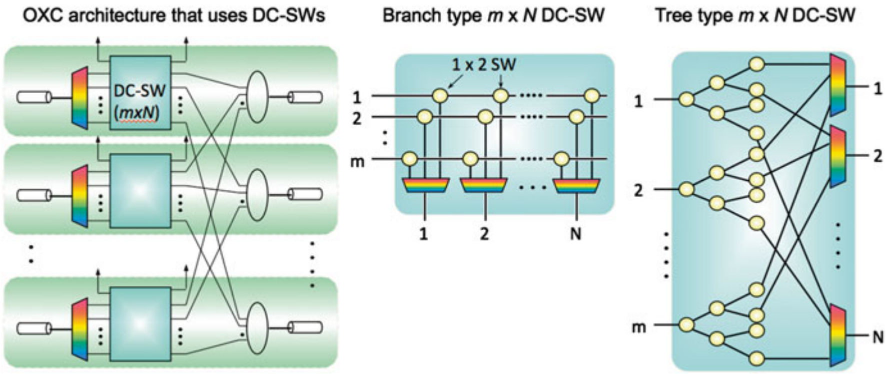



**Fig. 6.3** OXC architecture using Delivery and Coupling (DC) switches

conversion by realizing optical bypath of nodes. The wide deployment of OXCs in the 2010s strengthened demands for eliminating human intervention by automating the provisioning of new optical paths and for optical layer protection/restoration; CAPEX reduction remains a matter of great concern to the carriers. Automation encompasses the provisioning of future dynamic optical services, and this will become clearer with the evolution of SDN (Software-Defined Networking) and NFV (Network Functions Virtualization) technologies; agile and flexible OXCs are indispensable. To achieve the flexibility needed, add/drop parts of the OXCs should offer Colorless/Directionless/Contentionless (C/D/C) functions, which are elaborated in Sect. 6.3.1.2.

There are different ways to implement the express and add/drop switch part, and this section focuses on express switch architectures. Among the architectures developed in the early 1990s, only relatively simple ones were commercially deployed. Figure 6.3 shows one of the architectures proposed in 1992 [13–15]; it utilizes Delivery and Coupling (DC) type switches. The DC switch consists of multiple $1 \times 2$ (Mach–Zehnder interferometer) switches that form a branch or a tree configuration as shown in Fig. 6.2 or some variant thereof.

The first prototype OXC system, developed in 1996 [16], accommodates eight input and eight output fibers; each fiber carries 16 wavelengths and each channel is modulated at 2.5 Gb/s. A photograph of the system is presented in Fig. 6.4 [16]. The optical switches are fabricated with PLC (Planar Lightwave Circuit) technologies and the eight $1 \times 8$ thermo-optic switches are monolithically integrated on a PLC chip. The number of WDM wavelengths per fiber available then (early 2000s) was relatively small, 8–16, and hence the PLC approach was useful. The architecture, however, requires as many DC-switches as there are wavelengths multiplexed in a fiber. PLC-based DC-switches are rather expensive for massive implementation (i.e., more than 80 DC-switches), and the functional device was later replaced with WSSs (Wavelength Selective Switches) that utilize three-dimensional spatial optics.

WSSs can more easily process multiple wavelengths since most of the optical components can be shared among different wavelengths as seen in Fig. 6.4, which illustrates the schematics of (a) MEMS and (b) LCoS-based WSSs. By utilizing

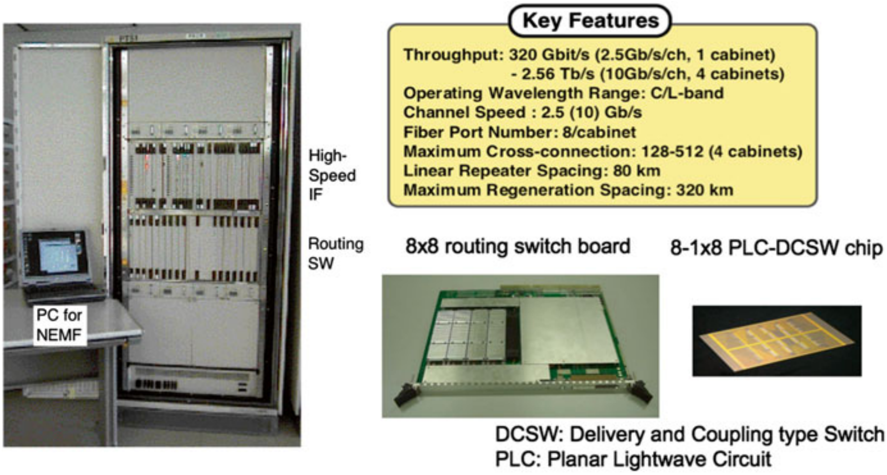



**Fig. 6.4** OXC prototype developed in 1996 (Courtesy of NTT)

WSSs, the Broadcast-and-Select OXC architecture that places optical couplers (OC) at the input fiber side and WSSs at the output fiber side has become widely deployed, see Fig. 6.5. The inverse configuration of WSS and OCs (Route-and-Couple) is equivalent to the one that was originally realized with DC-switches as depicted in Fig. 6.2. Please note that the optical systems that are designed to support single-mode fiber are symmetrical in terms of optical input and output directions. Thus both architectures work the same, but the Broadcast-and-Select architecture can implement input signal multicast. WSSs can cope with greater numbers of wavelengths, but the challenge lies in port-count expansion since WSSs utilize spatial optics. Indeed, maximum WSS port-counts are limited to 20+ and no cost-effective approach to substantially expanding this number is available.

## OXC Evolution

By the middle of the 2010s, the maximum OXC port count implemented in networks was eight, which corresponds to the maximum number of adjacent nodes. However, as traffic continues to increase, the necessary number of fibers per link has risen and, as a result, the OXC port-count needs to be increased. Starting with the current port count of eight, a yearly traffic increase of 40 % yields 84 ports in 8 years. This will render the present OXC architectures shown in Fig. 6.6a, b unworkable, since the OC loss will be excessive; for 84 fibers the intrinsic SC loss is 19 dB. To avoid this loss increase, the Route-and-Select architecture shown in Fig. 6.6c is preferable, however it needs twice as many WSSs as seen in Fig. 6.6c. Even with this architecture, the maximum available port count is limited to 20+ (the maximum WSS port counts). Cascading WSSs (Fig. 6.6c insertion) is a simple way to expand the port count. If we use $1 \times 9$ WSSs (one of the most commonly available

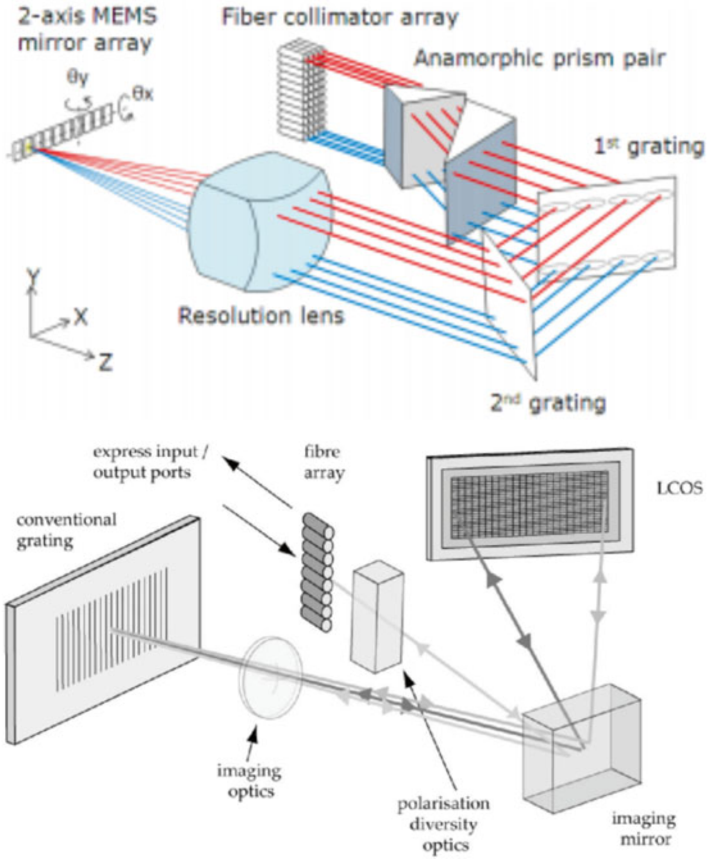



Fig. 6.5 Schematics of (**a**) MEMS [17] and (**b**) LCoS-based WSSs [6]

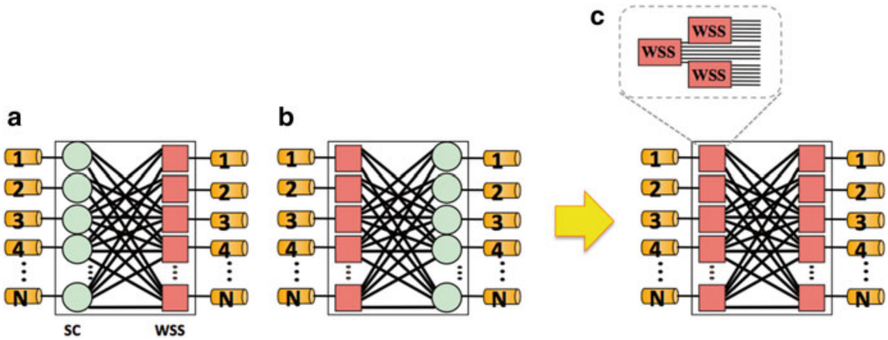



**Fig. 6.6** WSS-based OXC architectures. (**a**) Broadcast-and-Select. (**b**) Route-and-Couple. (**c**) Route-and-Select

sizes), a 1×84 WSS can be created with the cascaded architecture using 11 1×9 WSSs per incoming fiber. This approach demands a huge number of WSSs. If we use 1×9 (1×20) WSSs, a total of 1848 (840) WSSs will be required to create an 84×84 OXC; this does not include the additional WSSs that will be needed in the add/drop part. Furthermore, cascading WSSs doubles the loss. It is, therefore, necessary to develop new architectures that allow substantially higher OXC port counts, which demands non-traditional approaches to the non-blocking express switch.

### Impact of Intra-node Blocking on the Design of Large Port-Count OXCs

TDM networks avoid the problems, such as wavelength collision at transmission links and nodes that plague WDM networks that do not utilize wavelength conversion. Wavelength collision degrades network resource (fiber and node throughput) utilization (see Fig. 6.7). In order to minimize the loss, we apply new network design algorithms that achieve efficient RWA (Routing and Wavelength Assignment). The algorithms were originally designed to minimize link level collision (internode blocking), but can be extended to include node level collision (intra-node blocking), the intra-node blocking aware RWA (see Fig. 6.7). From the network performance point of view, the fiber utilization is of prime importance regardless of the origin of blocking. Intra-node blocking was originally discussed in terms of the add/drop part of OXCs [17] since the express switch part makes it easier to attain perfect non-blocking when input and output port numbers are relatively small. This is because the hardware requirements needed to realize non-blocking at the add/drop part or full C/D/C add/drop are rather intense, even when the number of input/output fibers is relatively small, say less than 8. As a result, various technology options were first

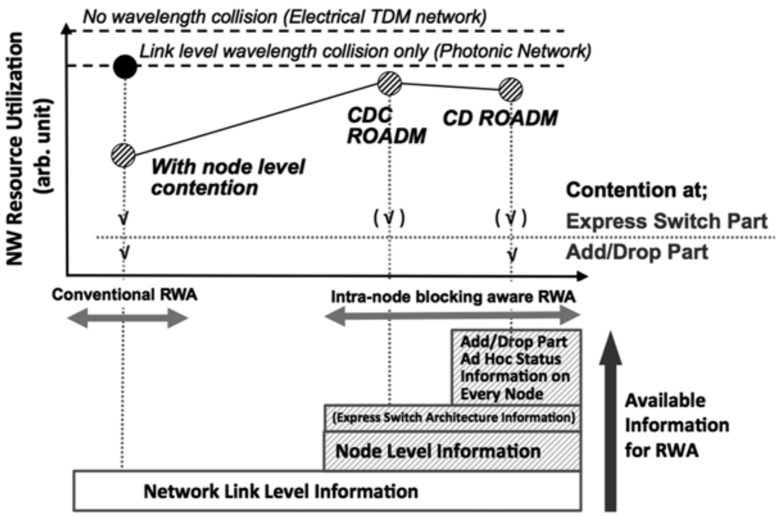



**Fig. 6.7** Network resources utilization and link/node level blocking

discussed for C/D architectures that abandoned the goal of contentionless capability or permitted blocking in the add/drop part [17–21]. The resulting C/D architectures combined with add/drop part contention aware RWA have been shown to attain almost the same performance as full C/D/C. This situation changes as the number of input/output fibers increases, which is discussed below.

**Architectures for Creating Large Port-Count OXCs [22, 23]**

In order to create large port-count OXCs, various express switch architectures have been proposed. The architectures exploit the fact that if we wisely limit the routing capability of, or permit blocking at the express switch, then the express switch hardware scale can be greatly reduced and the routing performance penalty can be marginal. In other words, degradation of the blocking ratios can be kept small, or when the target blocking ratio is set, the traffic volume supported by the node can be almost the same, if we employ the appropriate intra-node blocking aware RWA as illustrated in Fig. 6.7. Figure 6.8 explains the background of this approach. Figure 6.8a shows an example of a physical network topology, and (b) shows that the node degree of node A is 6, but the fiber degree is 14. Figure 6.8c shows an example of a node connection where node degree equals fiber degree, 14. The node degree is determined by geographical node configuration, which will rarely change, and the number is relatively small, say less than 8, for most practical networks, while fiber degree is determined by the total number of fibers on the links and can grow to a large number as traffic increases. Present WSS-based OXCs provide full routing capability in the express switch, or they are designed considering only the fiber degree. If we make use of the difference between node degree and fiber degree, or exploit practical node configurations (see Fig. 6.8b), we can introduce a slight restriction on the routing capability at the node. For example, if we want to route an optical path from node A to node B in Fig. 6.8b, any one of the three output fiber ports of OXC at node A can be utilized, while in Fig. 6.8c, only the specific output fiber port of the OXC is chosen to establish connection to the adjacent node. Therefore, even if we slightly limit the node routing capability, but adopt an intra-node blocking aware RWA, the change in network routing performance is minimal.

Various approaches have been tested to materialize efficient switch architectures for developing large-scale OXCs. The approaches fall into two classes; one introduces coarse granular routing (Fig. 6.9a–c) and the other utilizes the subsystem modular architecture (Fig. 6.9d). Both are briefly explained below.

Hierarchical Multi-granular Routing [24]

An optical switch can switch multiple optical paths simultaneously. Switching groups of optical paths or wavebands can reduce the total switch size (necessary number of cross-connect switch ports) substantially in matrix-type switches. The

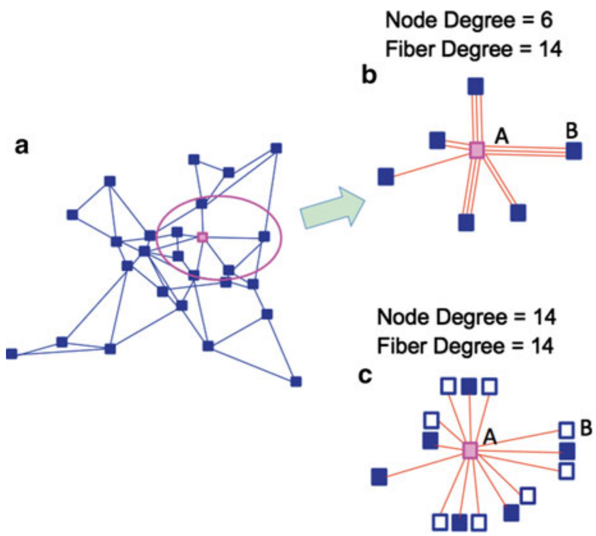



**Fig. 6.8** Node degree or fiber degree impacts OXC performance needed. (**a**) Physical network topology. (**b**) Adjacent nodes and multiple link fibers (fiber number allocation is just for explanation). (**c**) Node configuration with the same node and fiber degree

hardware scale reduction yielded by the introduction of wavebands can be substantial, more than 70 % for a matrix-switch-based cross-connect system [25], and more than 48 % for a 3-D MEMS-based WSS/WBSS (Wavelength/WaveBand Selective Switch) OXC [26]. The grooming ratio is an important parameter that determines routing performance of the node and the total switch hardware scale, and there is a trade-off relationship. Coarse granular routing (waveband routing) with supplemental intermediate grooming (Fig. 6.9a) can create cost-effective networks, and hence effective network design algorithms that incorporate a grooming ratio restriction are essential to make the best use of the architecture. Please note that carriers do not need to recognize the grooming ratio. The network design tool automatically determines the necessary switch hardware scale to accommodate the demands with consideration of the grooming ratio restriction. This architecture is viable when compact waveband cross-connects can be developed cost-effectively [27–29]; utilizing cyclic AWGs as waveband mux and demux is one example.

Grouped Routing and λ Selective Add/Drop [30]

The network architecture in [30] relies on coarse granularity routing (multiple wavelength path routing) and wavelength granularity level add/drop (Fig. 6.9b). Grouped routing is comparable to the situation where nodes are connected by virtual pipes whose bandwidth equals a bundle of multiple wavelength paths. The pipe is called "grouped routing entity pipe (GRE pipe)." A GRE pipe has or does not have end points; it may form a closed loop. No termination functions are defined, unlike the conventional waveband "path." A GRE pipe works as a "highway" where wavelength paths can be added/dropped at any node provided the

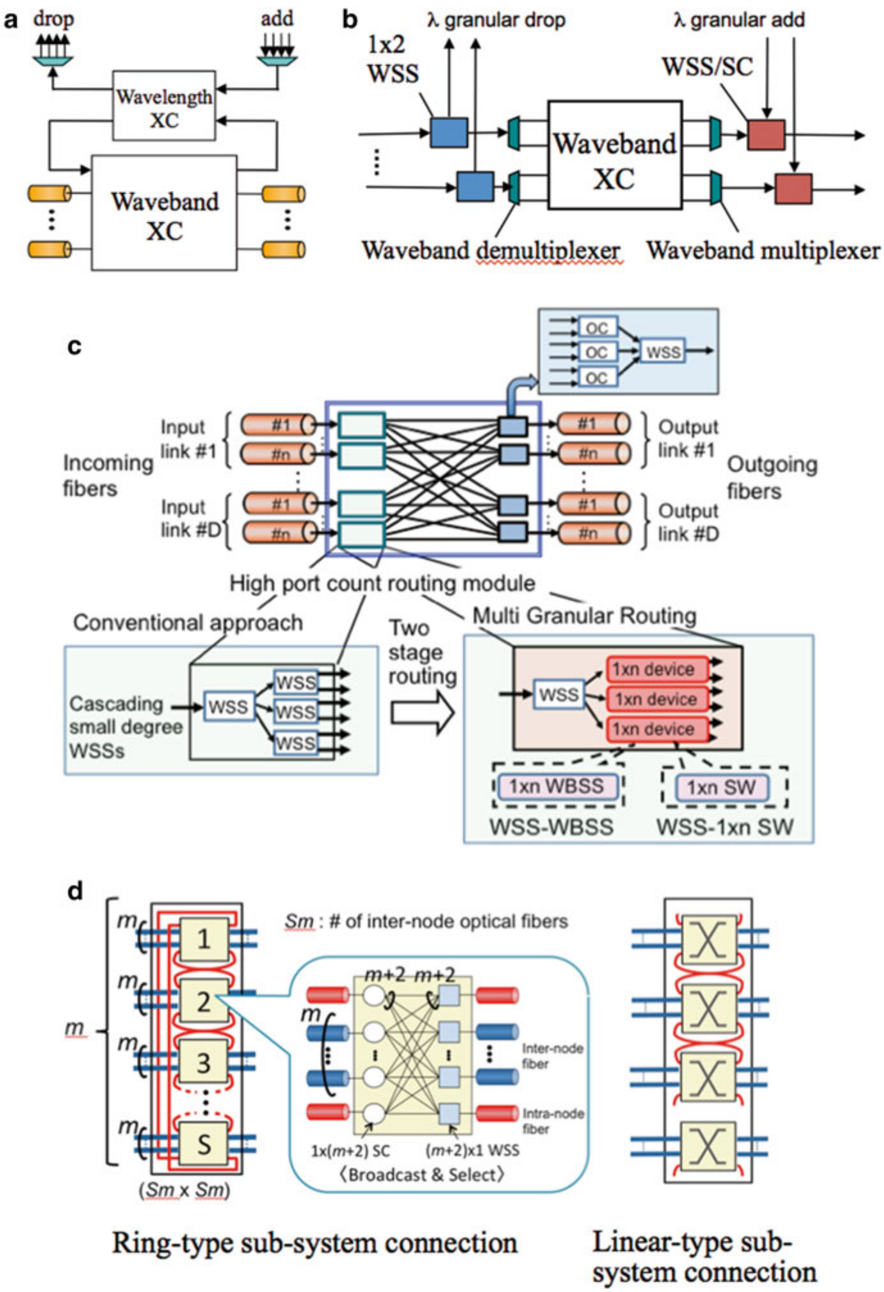



**Fig. 6.9** How to create large port-count OXCs. (**a**) Hierarchical multi-granular routing. (**b**) Grouped routing and λ selective add/drop. (**c**) Multi-granular two-stage switching. (**d**) Interconnected subsystem architecture

wavelength resource is vacant when adding. The express path routing part routes optical paths in terms of GRE, and can be constructed by using matrix switches or WBSS in a very compact manner as discussed in i). To make the best of the architecture, the efficient network design algorithm of [30] is suggested. Numerical experiments demonstrated that if the average traffic demand between each node pair exceeds six optical paths for a $7 \times 7$ regular mesh network with GRE pipe capacity of 10λs, the decrease in fiber utilization achieved by slightly restricting the routing capability can be held to less than 10 %, while the port count (in the case of matrix switch) is reduced by more than 88 %, compared to the comparable single layer optical path network. One particular point to be emphasized is that this scheme allows the use of WSSs instead of WBSSs, in which case the node architecture is the same as conventional WSS-based OXCs. With this, obviously no hardware simplification is possible, however the grouped routing operation minimizes the filtering (spectral narrowing) effect and this approach can minimize channel gaps in flexible grid networks; as a result, the total traffic accommodated in the network (per fiber) can be increased even if routing is done only on the GRE level [31]. These characteristics are particularly important in metro network applications where the number of nodes traversed end-to-end can be much larger than is true in core networks.

#### Multi-granular Two-Stage Switching [32, 33]

As explained before, wavelength paths arriving at a node from one of the input fibers that should be delivered to one of the adjacent nodes will not always need to be distributed among all parallel fibers on the link to the adjacent node. Setting slight restrictions on fiber selection on a link so that those wavelength paths can be routed on one or a few limited fibers on the link without allowing each wavelength path to freely choose a fiber, can yield an efficient solution. New OXC architectures have been presented [32, 33] that utilize a two-stage routing mechanism (Fig. 6.9c); WSSs for selective/dynamic wavelength grouping at the wavelength path level followed by $1 \times n$ optical switches (SWs) or $1 \times n$ WBSSs for selecting fibers on links between adjacent nodes. This architecture significantly reduces the necessary number or port counts of WSSs; up to *1/n* (*n*: number of parallel fibers on a link to the adjacent node), as detailed in [32, 33]. An optical path network design algorithm that can make the best use of the proposed architecture also plays an important role. The number of the required WSSs can be reduced to 40–60 %, while the fiber utilization decrease can be kept within 2–3 % [33].

#### Interconnected Subsystem Architecture [34–36]

Architectures that utilize the interconnected multiple OXC-subsystem, have been presented [34, 35]. The OXC-subsystems are created using small cost-effective WSSs as seen in Fig. 6.9d; the subsystems are interconnected with a limited

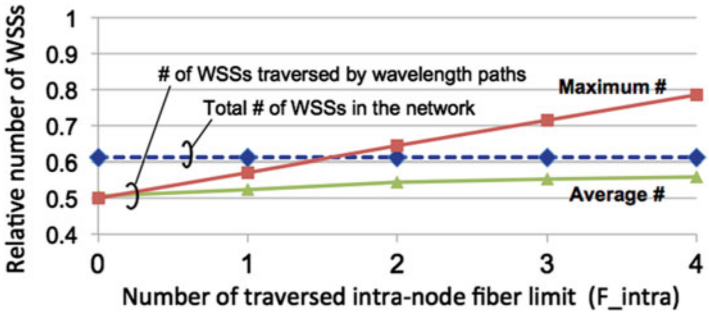



**Fig. 6.10** Relative # of WSSs traversed by wavelength paths, and the total number (rel.) of WSSs in the network when the blocking ratio is $10^{-3}$. Subsystem architecture assumes $1\times9$ WSSs, while the conventional one assumes large port-count WSSs, up to $1\times35$

number of intra-node fibers. The use of subsystems allows the use of both small port-count WSSs and the Broadcast-and-Select (SC + WSS) configuration, and as a result, the total number of WSSs needed in the network can be dramatically reduced. Figure 6.10 shows the relative number of WSSs traversed by wavelength paths for the 26-node pan European network (see Fig. 6.8) and the relative total number of WSSs in the network when the blocking ratio is $10^{-3}$. Here *F_intra* indicates the maximum number of intra-node fibers that each wavelength path can traverse end-to-end. Please note that subsystem architecture assumes $11\times n9$ WSSs, while the conventional one assumes large port-count WSSs, up to $11\times n35$. When we utilize $11\times n9$ WSSs for the conventional architecture and a larger port-count WSS is assumed to be constructed by cascading $11\times n9$ WSSs, then the superiority of the subsystem architecture is further enhanced. These reductions substantially mitigate the spectral narrowing generated by the WSS filtering effect, reduce total node loss, and substantially reduce node cost. With the subsystem architecture, both average and maximum loss values are reduced. All of these benefits are realized with a penalty of less than 1 % in terms of accepted traffic volume when the blocking rate is 0.1 % [34–36]. This performance can be attained by using the newly developed intra-node blocking RWA. A detailed analysis on where in the network, blocking occurs (link level and node-level blocking) is given in [35]; it was shown that F_intra is a key parameter that determines blocking performance. The architecture in [34, 35] assumed a symmetric ring-type interconnection between subsystems. The pay-as-you-grow and the dynamic expansion capabilities of the system, which should be supported with no service disruption (or hitless expansion), are critical in allowing this large-scale OXC to be introduced cost-effectively from the outset. One variant [36], adopts a linear-type subsystem interconnection Fig. 6.9d; it leaves one set of ports of each end subsystem open for later connection to additional subsystems without affecting existing optical path connections. By adapting a newly developed network design algorithm, it is shown that almost the same level of blocking performance can be achieved as the original ring-type interconnection [36].

## 6.4 Multi-layer Elastic and Flexible Add/Drop Node Architecture

### *6.4.1 Node Architectures and Interfaces That Address Different Multi-layer Use Cases*

An efficient network bandwidth utilization aiming at the minimum cost per transported bit calls for the aggregation of low rate services, such as data flows from IP edge routers (e.g., BRAS, IPTV, VoIP, etc.) or CBR signals coming from leased lines, storage (SAN) and legacy SONET/SDH/PDH networks, into the high rate, fat DWDM data pipes.

All optical sub-wavelength switching techniques, e.g., Optical Packet Switching (OPS) or Optical Burst Switching, though extensively investigated seem not yet enough mature in terms of manufacturing costs and reliability for a short-/medium-term widespread deployment. Thus the electrical grooming is still the only practical solution in which operators can extensively count on to optimize the bandwidth usage in their optical networks.

Keeping in mind that the transparency with respect to the transported services is a relevant requisite in many applications ranging from regional up to continental network sizes, OTH (Optical Transport Hierarchy) is the natural client-independent digital stack of choice for grooming or sub-λ multiplexing in the electrical domain. Though other options might be possible, it appears unavoidable that a transport entity functionally equivalent to an Optical channel Transport Unit (OTU) should anyway be introduced to adapt the service information content to the photonic layer (Fig. 6.11).

OTH is an integral part of the OTNs, the transport technology for the optical networks developed by ITU-T comprising all the functionalities needed to sup-

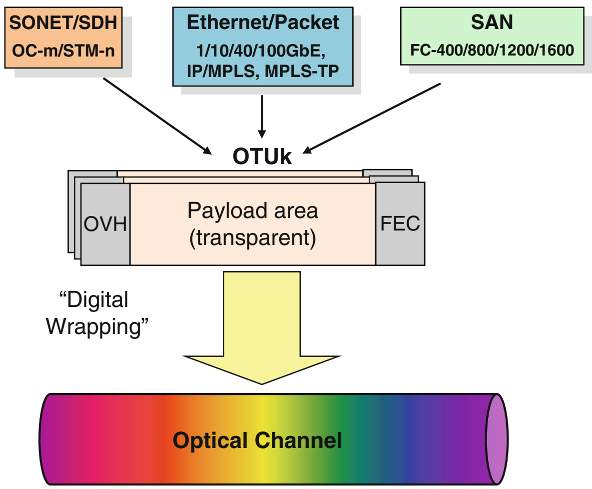



**Fig. 6.11** Protocol-agnostic switching and transport

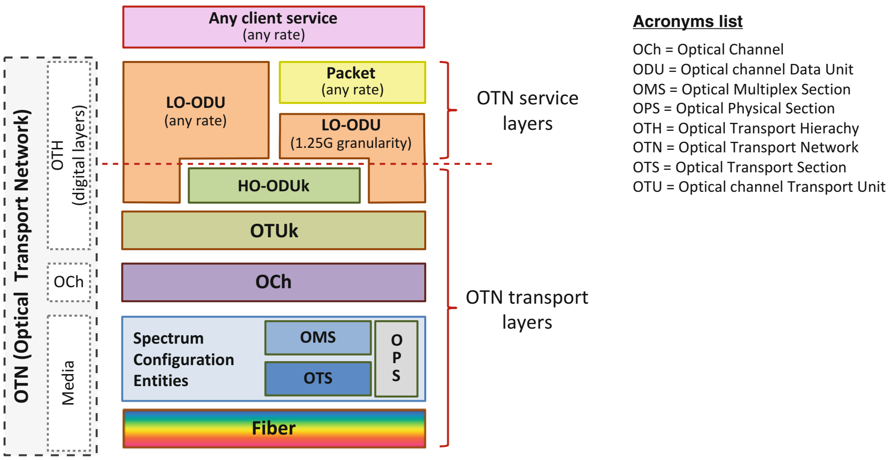



**Fig. 6.12** Overview of the OTN layering

port the transport, aggregation, routing, supervision, and survivability of client signals that are processed in both the photonic and digital domains. The OTN network architecture is defined in ITU-T G.872 "Architecture for the Optical Transport Network (OTN)" [37] where the network is decomposed into independent transport layers partitioned in a way that reflects their internal structure (Fig. 6.12).

ITU-T developed the OTN standard framework, with the overall objective to extend (or even enhance) the SDH/SONET-like features (performance monitoring, fault detection, communication channels, etc.) operators were familiar with, to WDM wavelengths. As result among the most important benefits OTN is bringing to the optical networking it is worth mentioning:

- Universal containers supporting any service type for an end-to-end optical transparent transport of the customer traffic.
- A standard multiplexing hierarchy accommodating client signals of various natures into a line interface served by a single optical channel.
- A frame structure supporting strong FEC capabilities for long-haul photonic applications and an overhead for robust multi-level OAM in the digital domain.
- Sub-50 ms protection schemes
- Wavelengths OAM through OOS (Optical Overhead Signal, only functionally specified) providing enhanced OAM&P features.

The basic characteristics of an OTN multiplexer are outlined in Fig. 6.13. The "Interfaces for the optical transport network" are specified in ITU-T G.709 standard, which is in turn based on G.872 [37].

Flexibility in foundation OTN (early 2000s) was supported by an Inverse-Multiplexing procedure, specified in the same way as the SONET/SDH Virtual

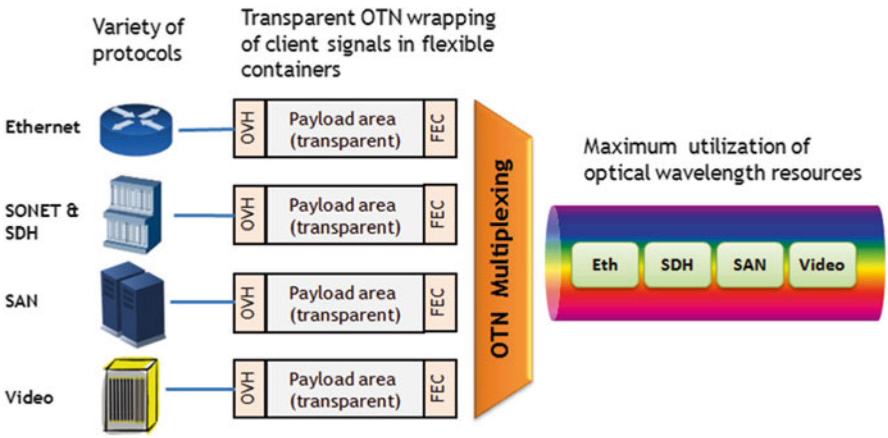



**Fig. 6.13** OTN multiplexing principle

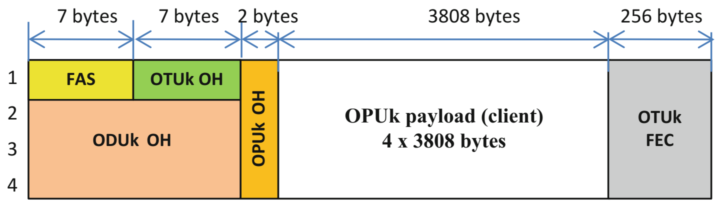
Concatenation (VCAT) [38]; the OPUk VCAT, specified for $k=1, 2, 3$, allowed the assignment of the transmission bandwidth in steps of 2.5 Gb/s ($k=1$) or 10 Gb/s ($k=2$) or 40 Gb/s ($k=40$).

In spite of its flexibility and resilience, VCAT is considered rather complex to build and manage (need of large buffers to compensate for the differential delay among VCAT members and many OAM&P networking instances to be administered for a single client signal), therefore the G.709 update occurred in 2009 introduced a new type of a flexible container named optical ODUflex (flexible Optical channel Data Unit) overcoming the VCAT hassles. An accompanying protocol for the ODUflex hitless resize was specified as well [39] for the case of those packet services mapped via GFP-F [40] having a variable bandwidth demand over time.

An elastic OTUflex line signal is however not specified, i.e., the ODUflex has always to be considered as "lower order" client signal carried within a "higher order" ODUk/OTUk ($k=1, 2, 3, 4$) line. The aim was indeed to keep the number of different types of line-interfaces as few as possible in order to minimize the capital expenditures (until recently to each optical line rate was corresponding one different Transponder module).

The new generation of optical transceivers for the Elastic Optical Network is undermining the rigidity of that approach. The number of carriers, the modulation format and baud rate are becoming configuration parameters that will enable network operators to have the right balance between the digital capacity or line rate and reach which is the best for their transport infrastructure (Fig. 6.14).

Thus the future OTN hierarchy advancement "beyond 100G" (B100G OTN) is expected to introduce a rate-flexible line interface format, finally complementing in the OTU layer the elasticity already provided by ODUflex in the LO-ODU service layer. The current predominant thinking in the OTN evolution debate is in favor to define an "$n \times 100$ Gb/s" (with $n \geq 2$) iterative structure, termed OTUCn shown in Fig. 6.15. The index "$n$" immediately suggests the possibility of variable bit rates in step of 100 Gb/s; while single vendor intra-domain OTUCn elastic interfaces (SV-IaDI) may take any $n \geq 2$ and be resizable, in similarity to




| OTU type | OTUk bit rate | OPUk payload rate | Example of typical clients |
|---|---|---|---|
| **OTU1** | ~ 2.666057 Gb/s | ~ 2.488320 Gb/s | STM-16, GbE, FC-100/200 |
| **OTU2** | ~ 10.709225 Gb/s | ~ 9.995277 Gb/s | STM-64, 10GbE WAN, 100GbE LAN FC-400/800 |
| **OTU3** | ~ 43.018414 Gb/s | ~ 40.150519 Gb/s | 40GbE LAN, STM256 |
| **OTU4** | ~ 111.809974 Gb/s | ~ 104.335975 Gb/s | 100GbE LAN |

**Fig. 6.14** OTN frame and OTUk ($k=1, 2, 3, 4$) rates as currently standardized

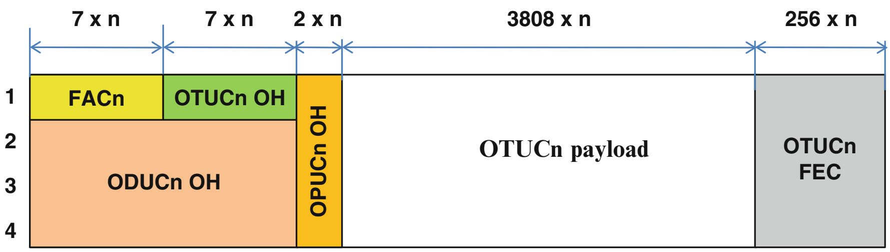



**Fig. 6.15** Example of flexible Optical Transport Unit for the OTN evolution "beyond" 100G

SDH only few values would possibly be object of standardization for multi-vendor intra-domain (MV-IaDI) interworking and IrDI interfaces (the initial standardized rate will very likely be for $n=4$ to carry future 400 GbE clients).

Although many of the basic modelling elements for representing an OTN network managing the elastic optical line can already be found in the in-force ITU-T G.872, the new flexible line paradigm breaks the simple 1:1 association between wavelength and OTU/ODU in the sense that an OTUCn/ODUCn can be distributed through several optical carriers occupying one or multiple frequency slots (FS). Fat signals (e.g., 400 Gb/s or above) are split into multiple thinner pipes that can be transported over multiple wavelengths (their number and modulation format will be the best compromise between spectral efficiency and reach), possibly fitting into a single media channel to minimize the spectrum occupancy (adjacent wavelengths can be closely packed with no or reduced guard-band) or alternatively into multiple, non-adjacent, media channels to deal with the possible spectral fragmentation on existing networks.

A new generation of Sliceable-Bandwidth Variable Transponder (S-BVT) will have the capability to source several elastic interfaces OTUCn with "$n$" varying over time to follow the typical network cyclic load fluctuations (daily, weekly, etc.). Therefore when the full available transponder capacity is not exploited by a single traffic demand, the unused transmission resources are assigned to other (low rate)

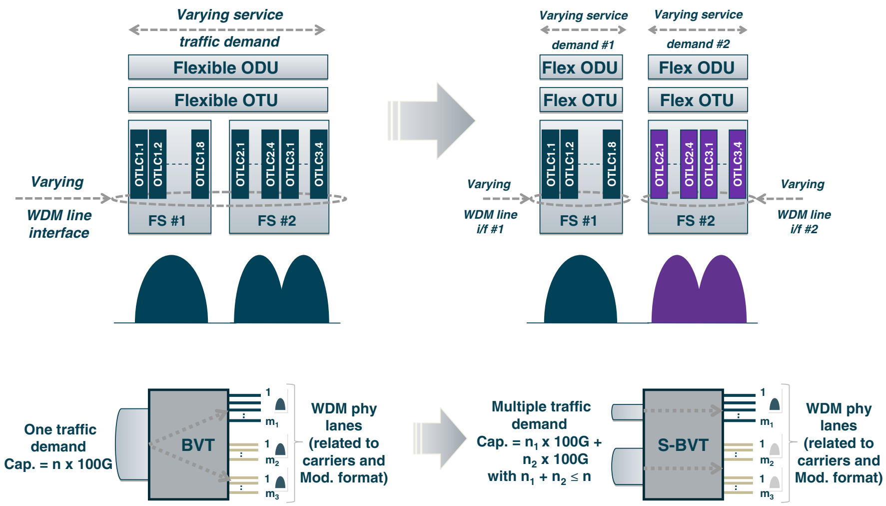



**Fig. 6.16** BVT and S-BVT description (from one traffic demand at a maximum capacity to two traffic demands at lower rate)

**Table 6.1** Comparison between S-BVT and BVT

| BVT basic characteristics | From BVT to S-BVT enhancements |
|---|---|
| • Follow the typical periodic load fluctuations (daily, weekly…) occurring in the network | • All the capacity and hardware resources can always be used up to the maximum even when at a certain time there is no a single traffic demand consuming the entire amount of the available bandwidth |
| • Dynamically reconcile the spectrum usage to actual used capacity and to reach the (e.g., higher efficient modulation format on the direct path, typically shorter, and switch to less efficient modulation format on the protection path, typically longer) | |
| • Avoid the need to purchase many different bandwidth transponders: the aim is to have one type of transponder that delivers the wide range of needed bit rates | • The utilized broad spectrum can then be "sliced" in the optical layer by flex-grid compatible ROADMs and the various (contiguous or not) portions of the spectrum sent to the different destinations |
| • Some capacity and precious hardware resources may remain unutilized when the required bandwidth is lower than the maximum supportable | |

demands. Figure 6.16 and Table 6.1 show a schematic comparison between S-BVT and BVT where the former is supporting multiple traffic demands simultaneously, while the latter is limited to only one.

This type of flexibility does involve not only the photonic layer (flexible grid, variable number of sub-carrier, variable modulation formats, flex-ROADM, etc.) but the electronic layer as well that has to be able to distribute the available bandwidth under software control, from one-to-several service demands.

The modern transport platforms are usually built as combination of L0 (photonic or lambda-based OXC), L1 (ODUk-based DXC and occasionally even legacy SDH

DXC) and some L2 packet switching capabilities for a convergent network model uniting packet, TDM and wavelength switching with robust OAM and FEC for reliable transport.

The example of Fig. 6.17 shows an OTN-based transport network supporting client networks including both packet and TDM. ODUs go straight onto wavelengths when no further grooming or aggregation is needed for the wavelength. The LO-ODUs allow instead further grooming or aggregation to improve the fill rate within the wavelength and multiple LO-ODU containers can be multiplexed into higher rate HO-ODUs. The ODU is also providing a clear demarcation boundary and a robust set of performance monitoring and fault detection.

From the just mentioned example becomes clear that photonic and electronic switching play complementary roles for maximum network efficiency:

- Optical signals transiting a network node are switched at the wavelength level. This is best suited for cases in which the granularity of the transported service is close to the wavelength capacity. Photonic switching is used primarily to provision and restore wavelength services.
- When the optical signal is terminated, the entire signal, or its service-specific traffic contributions, can be individually switched. Electronic switching is used primarily to provision and restore sub-wavelength services that consume less than a wavelength of bandwidth.

Nevertheless modular network element architectures should provide the ability to match the desired flexibility according to the underlying business problem by meeting the provider's specific service mix and traffic distribution. The possible configurations and grooming options are summarized in Table 6.2.

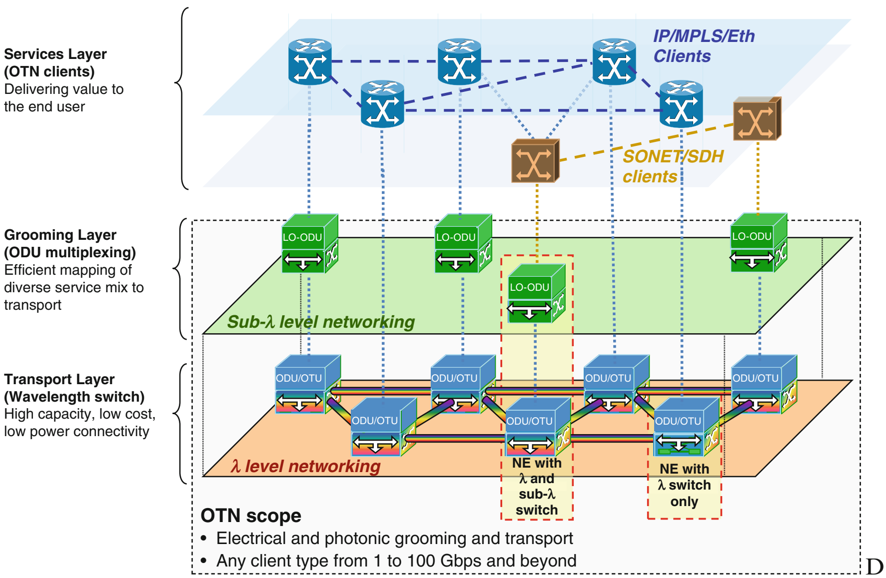
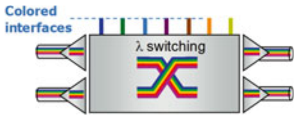
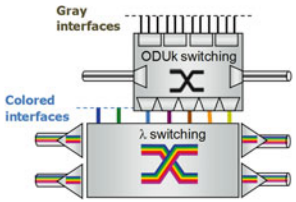
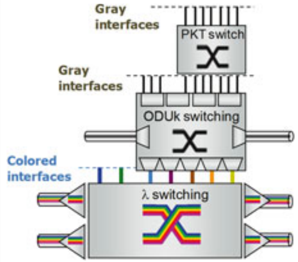



**Fig. 6.17** OTN network vision: switching at lambda or sub-lambda level

**Table 6.2** Possible configurations and grooming options

| Grooming model | λ Level grooming | Port-level grooming | Sub-port level grooming |
|---|---|---|---|
| Logical interfaces | None | Port | Sub-port, e.g., VLANs within a router port |
| Physical interfaces | Each port uses a wavelength | Ports that may be smaller than a wavelength capacity | Ports that may be smaller than a wavelength capacity |
| Granularity | Limited to the wavelength level, transport purely at photonic level | Limited to the port-level, electrical, and photonic processing | Down to virtual channels (e.g., VLANs) or port-level, electrical, and photonic processing |
| Transport switching layers |  |  |  |
| | Photonic networking | Integrated photonic and circuit networking | Integrated photonic and circuit networking including packet classification |
| | • Wavelength (OCh) | • Wavelength (OCh) | • Wavelength (OCh) |
| | | | • Fixed-rate circuit (ODU) |
| | | • Fixed-rate circuit (ODU) | • Adjustable (variable) rate circuit (GFP/ODUflex) |

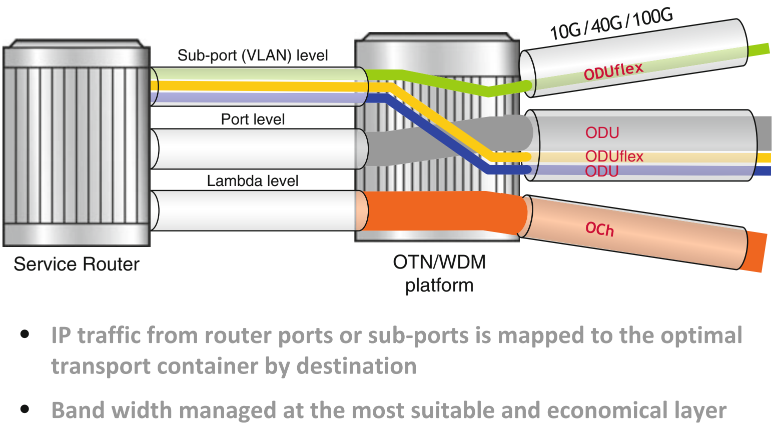



**Fig. 6.18** Example of IP traffic grooming from a router

The three possible grooming models are outlined in Fig. 6.18 taking as an example, the case of IP traffic grooming from a router.

### Sub-port Level Grooming

The optical node arrangement with an electrical switch, also providing packet aggregation and/or classification capabilities, on top of the photonic switch is the most flexible alternative as it allows the grooming of client traffics with finer granularity (packet services at variable rate included) in the electrical domain.

- It makes possible the handoff of channelized Ethernet interfaces from routers using a single high-speed interface, e.g., 40G, 100G or beyond, rather than requiring a different physical interface to every edge router. Different virtual channels, e.g., VLANs, from the same router physical port are mapped to different OTN virtual containers and through the transport network bounded to different destinations. VLAN shaping and resizable ODUflex are key enabler features.
- It deals with sub-port interfaces and sub-port level grooming such that low bandwidth services don’t have to consume the entire capacity of a wavelength. Many incumbent providers and large operators offer both leased lines and private data services, typically at CBR rates, including any-to-any business VPN services. Where the traffic load of these services represents a significant portion of the overall traffic matrix, the sub-wavelength grooming becomes a need for a better utilization of the network resources at the lowest cost per bit transported.

Finally a rate-flexible or elastic OTU will allow adapting the optical spectral spectrum usage of each wavelength to the actual aggregated bandwidth demand and reach according to cyclic traffic load variations (day and night, working days and week-end).

#### Port-Level Grooming

In network areas where the traffic distribution reveals mostly a direct hub and spoke type of connectivity between a few centrally located points of presence may become advantageous to move this traffic as directly as possible to its destination while driving fewer requirements for grooming flexibility. This is frequently characterized through leveraging to a lesser extent electrical grooming capabilities within a wavelength division multiplexing system. Functionally, port-level grooming represents a subset of the sub-port-level grooming architectural model, while achieving efficiencies in grooming at the physical port level.

#### Wavelength Level Grooming

As with port level grooming, wavelength level grooming implies a physical port-level interconnect from the client equipment with each physical port being the equivalent of a single wavelength. IP over point-to-point WDM is an example of implementation of this type of network configuration that may be simpler but be very expensive because of the large amount of transit traffic that has to pass through intermediate nodes. When the transit traffic approaches or exceeds the bandwidth that can be carried by a single wavelength, it becomes more efficient and cost-effective to use an optical through-path.

## 6.5 Multidimensional and Function Programmable Optical Node Architectures

### *6.5.1 Dimensions for Future Optical Nodes*

As mentioned in the first section of this chapter, early optical technologies were used mainly for point-to-point data transmission while the processing and networking operations were undertaken in the electronic domain. In these point-to-point transmissions, WDM was primarily used to increase the capacity of the optical network by using the *wavelength dimension*, where wavelength-based channels provided the scalability required over the optical transmission links. However, the potential of optical networking to reduce costs and increase performance was quickly recognized and different node architectures were proposed which consider the advantages of WDM with respect to increased capacity and flexibility to optical communication systems. For this new network architecture paradigm, new technologies were required such as colored laser sources and multiplexing/demultiplexing devices. In addition, optical amplifiers were important for power compensation due to losses generated by the optical fiber and switching devices at intermediate points. Based on these technologies, one-degree nodes (with one-input/one-output fiber components i.e., FBGs) were designed to cope with the requirement of adding/dropping individual wavelengths at intermediate nodes in the network. These nodes allowed the implementation of the frequency (*wavelength) dimension* for networking functionalities, being initially used in ring topologies but with the advantage of connecting one node to various other nodes using the same fiber by selecting different transmission wavelengths for different connections. The frequency dimension supported also the evolution towards mesh networks based on the development of multi-degree nodes with the intention of performing the switching function directly in the optical domain, thereby avoiding the use of expensive electronic switching at intermediate nodes. These multi-degree nodes use WSS to enable the effective operation of optical networks with fixed circuits or connections corresponding to different frequency channels.

Meanwhile, with the popularity of Internet, the optical circuit switching (OCS) model previously used for voice traffic was found to be non-ideal to handle the continuously growth of IP-centric services. Therefore, a new dimension was required to complement OCS (which relied on the *frequency dimension*) and could cope with the packet-oriented nature of the Internet traffic. This dimension is *time*, which has been used extensively for electronic packet switching, but now it is considered in the optical domain. This time dimension involves the duration of each data packet, and how the different optical packets can be routed to different node destinations within the optical network. Although all-optical packet switching is still challenging since most of the processing of the data packets needs to be converted into electronic format, this type of concept is very representative of the *time dimension* which can be exploited in future elastic optical networks.

Other particular dimension is employed to deal with the continuous increasing demand of higher transmission capacities. In traditional optical systems, high-capacity transmission systems were designed and implemented with transmitters, which will be easily modulated by conventional on–off keying (OOK) modulation formats. OOK leverages the amplitude modulation of the continuous wave (generated by a laser source) with the incoming electronic signal to the modulator, creating an optical signal instantaneously. However, these types of modulation format were highly affected by the noise, requiring additional subsystems to regenerate or detect the signals after transmission. To ameliorate these optical transmission systems, modulation formats considering the *phase domain* have been used to boost the capacity and spectral efficiency of optical fiber systems. Within this *phase dimension*, modulation formats based on phase shift keying (PSK) and Quadrature Amplitude Modulation (QAM) are commonly used to increase the number of bits per symbol.

*Polarization* is other type of dimension used to increase the transmission capacity and is considered to improve the flexibility of the optical network. The *polarization dimension* contemplates the modulation of one of the states of polarization by a single data signal. This type of dimension has been used to increase the total capacity of optical fiber systems. The use of this dimension has demonstrated to be very practical, as simple passive polarization beam splitters and combiners are used. The polarization technique that is used for high-speed optical communications is called Polarization Division Multiplexing (PDM).

Other type of dimension recently adopted is the *space dimension*, as future optical networks seem to utilize new types of fibers and technologies since the current transmission capacity based on single-mode optical fibers are exhausting its theoretical Shannon limit. This *space dimension* is supported by Space Division Multiplexing (SDM), which has been proposed as the solution to deal with the large network capacity growth [41]. This type of multiplexing relies on multi-core and multi-mode fibers to enable many channels to be transmitted through different cores or modes of the optical fiber. The *space dimension* is then considered since many channels carrying information are transmitted simultaneously. Furthermore, another category within SDM has more recently been proposed, which makes use of the Orbital Angular Momentum (OAM) [36]. Laser beams carrying an OAM, or vortex OAM, are used as data carriers in an optical communication link. Thus, multiple independent data streams can be carried by a different vortex charge, multiplexing these vortexes at the transmitter and demultiplexing them at the receiver.

Bearing in mind this evolution of optical networks and the possibility of inclusion of several dimensions (see Fig. 6.19) on their own or in combination; it is envisaged that the network performance will be enhanced by adding dimensions to cope with new requirements in the networks infrastructures. However, it is important to understand the needs of multidimensional nodes. These requirements are addressed in the following subsection.

### 6.5.2 Need for Multidimensionality in Future Optical Nodes

Multidimensionality in optical networks refers to the different number of dimensions that can be supported by the network elements or nodes in a network for transmission, control and management of optical signals carrying information. Mainly, these dimensions in combination offer salient advantages to the optical networks since not only the network capacity is increased considerably but also

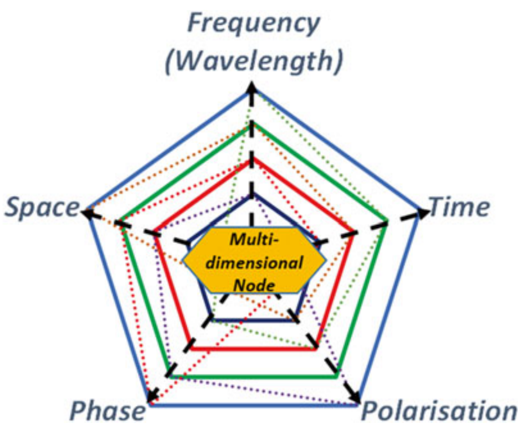



**Fig. 6.19** Multiple dimensions for optical nodes

optical networking system features are enhanced, allowing different network limitations to be overcome. However, advanced future optical nodes require certain system attributes to be able to cope with the dynamic network traffic and the expected new services demanded by the network users. In addition, it is important to provide a node architecture that is able to support a simple single-dimension single-function node evolvable and scalable to multiple ones where and when needed. To address this, the functional elements of the node should be decoupled from the input and output fibers. Thus, future network elements are expected to operate in combination with multiple dimensions, allowing a number of new benefits to the network.

Based on these node requirements, a node design paradigm has been presented in [37] and its architecture is shown in Fig. 6.20. This Architecture on Demand (AoD) node consists of an optical backplane, connected to several signal processing modules, such as SSS, fast switch, Erbium-Doped Fiber Amplifier (EDFA), spectrum defragmenter, etc., and the node's inputs and outputs. With this architecture, different arrangements of inputs, modules and outputs can be constructed by setting up appropriate cross-connections in the optical backplane. This AoD node observes a higher level of flexibility compared to Broadcast-and-Select (BS) and the Spectrum Routing (SR) architectures as the components used for optical processing are not hardwired like in a static architecture but can be interconnected together in an arbitrary manner.

#### AoD Node Functionality and Operation

During operation of the AoD node, the switching requirements (i.e., switching in spectrum, time, and space) of the optical signals are used to compute a suitable arrangement and interconnection of the modules that delivers the required functionality. Such arrangement constitutes the required architecture of the AoD node, which is implemented by means of internal cross-connections in the optical backplane.

Figure 6.20 shows the cross-connections used in the optical backplane to construct the example architecture of Fig. 6.21. As shown in Fig. 6.20, individual arrangements of components, with different levels of complexity, can be

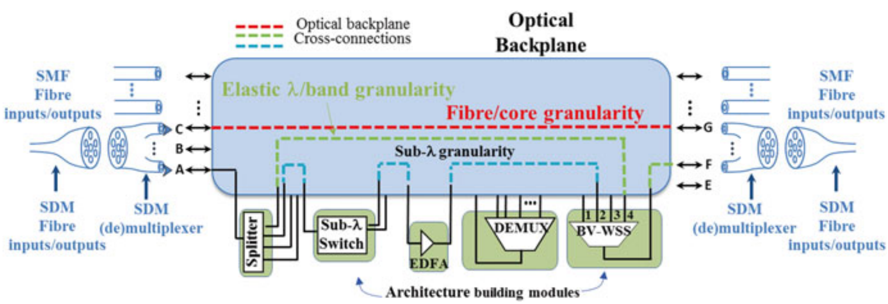



**Fig. 6.20** Architecture on Demand (AoD) node

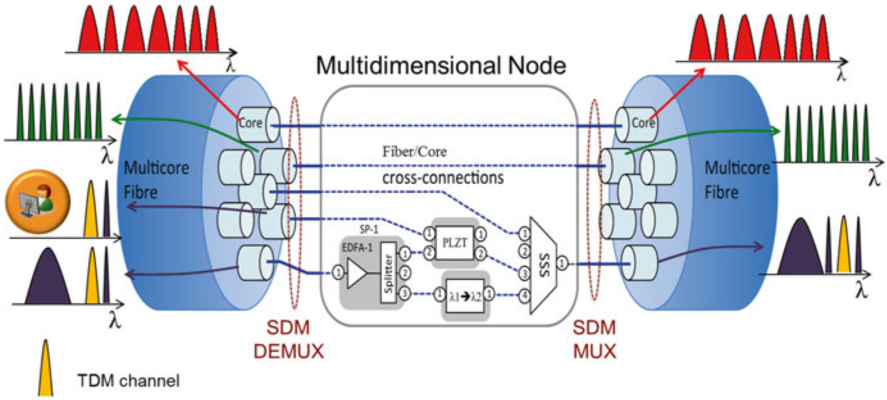



**Fig. 6.21** Multidimensional node (space, frequency, and time dimensions)

constructed on a per-port basis reflecting the switching requirements of the signals in each port. Such arrangements may be dynamically changed, by reconfiguring the backplane cross-connections, to fulfill the switching requirements of new traffic (Fig. 6.21).

Based on the AoD node architecture of Fig. 6.20, multidimensional nodes can be designed to provide increased flexibility in the network. For instance, in Figure B fiber/core switching granularity from input C to output G is presented. Elastic wavelength and waveband granularity from input A to output F are also implemented with a BV-WSS (SSS). Also, sub-wavelength switching granularity is implemented with a PLZT fast switch (e.g., 10 ns switching time), which is set to ON during time-slots occupied by sub-wavelength channels that need to be passed through to the output, and OFF elsewhere. Although at the PLZT input there is a copy of all signals from input A (Fig. 6.20), and they all undergo the same alternating ON/OFF process, the undesired signals are filtered out by the SSS. Thus, only the spectral components of the required sub-wavelength channels are passed through to output E.

Based on these AoD-based node attributes, improvements in node and network aspects are observed with respect to scalability, bandwidth granularity, and adaptability, in addition to the flexibility aspect previously mentioned.

## Scalability

A node must be able to scale when multiple dimensions are considered in the design. This node scalability refers to the ability of the node to be enlarged to accommodate a required growth with respect to different elements of the node such as the number of input or output ports, the number of optical components, number of cross-connections, etc. A simple way to evaluate the scalability of the node system is considering the number of cross-connections in the optical backplane or optical interconnect. Practically, the optical backplane has a specific number of ports, which will determine the scalability of the entire node architecture.

Taking into consideration the optical backplane of the AoD node architecture, in [42, 43] the scalability of the node has been assessed as a function of the number of cross-connections in the optical backplane. Furthermore, since the AoD design excludes hardwired cross-connections, fiber switching is possible. Fiber switching occurs when several or all the wavelength and waveband channels active in each input port can require the same destination or output port, allowing a simple fiber cross-connection and avoiding the inclusion of optical components such as SSS and power splitters. Thus, the use of a flexible node architecture, such as the one of the AoD, together with additional dimensions, will increase further the performance and scalability of the designed node. For instance, when an AoD node is used within an EON and in combination with SDM over multi-core fibers (MCF), the flexibility of the network is increased since on one hand the traffic demands varying from sub-wavelength to super-wavelength can be dealt with by flexibly allocating spectrum and, on the other hand, SDM will facilitate the network to encompass the capacity limitation and attain higher transmission throughput and spectral efficiency [44]. In addition, by introducing MCF in the networks, the spectrum continuity and contiguity constraints of the routing and spectrum assignment (RSA) will be mitigated considering that cores can be switched freely on different links during routing of the network traffic. With respect to the scalability, the cost of the network can be reduced by properly routing and assigning spectrum on different cores to traffic demands in such a way as to maximize fiber switching. In turn, this will lower the number of switching modules and minimize the number of cross-connections in the optical backplane (increasing the node scalability) for provisioning a given set of traffic demands. It has been demonstrated that even at high nodal degree, the number of cross-connections scales down by more than 60 % through enhancing fiber switching.

### Bandwidth Granularity and Adaptability

The network capacity highly depends on the node design being able to not only accommodate high-capacity channels but also to perform different operations with the ingressing and egressing of optical data flows, such as switching and routing, at different granularities, maximizing the throughput and handling multiple dimensions simultaneously. It has been demonstrated that with an AoD-based node in combination with SDM, spatial superchannels can support multiple bit rates by varying the number of aggregated spatial subcarriers and their modulation format [45]. This means that the spatial superchannel is adaptable to the user requirements, since some subcarriers can be programmed with a specific modulation format for a determined path within the network meanwhile other subcarriers can be programmed with different transmission capacities for different users' needs. Thus, the entire network can be adapted into different superchannel slices by combining the wavelength, phase and space domains (Fig. 6.22).

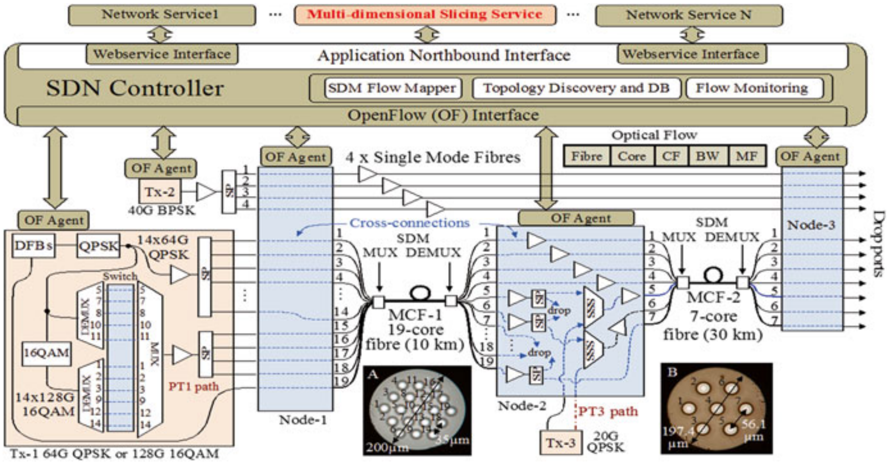



**Fig. 6.22** Multidimensional network based on AoD nodes

## *6.5.3 Optical Network Function Programmability*

Generally speaking, programmability is the ability of a particular system to change based on a set of provided new instructions that will modify the system performance. This set of instructions will follow a program logic for a particular application of the system being used. Thus, programmability in an optical network refers to the different and logical ways in which the network can be configured considering optical network elements that can be operated by a set of commands. This set of directions are used to synthesize the optical node and network, and the AoD node enables the dynamic synthesis of network architectures suited to switching and processing requirements of traffic.

### Node Level Synthesis and Programmability

The abovementioned AoD node dynamically provides customizable architectures. The working principle for the AoD for synthesis functionality is as follows: given a set of switching requirements of arbitrary input signals, a specific AoD architecture is devised and constructed interconnecting suitable building modules by means of backplane cross-connections. This AoD-based node requires fast and efficient synthesis and selection algorithms that take into account the availability of building modules and traffic requests to build up the architecture.

Figure 6.23 shows a synthesis example of an AoD node with heterogeneous traffic. The required output port for each channel is shown in red color. Channels at input port 1 are fed to a SSS to perform flexible-spectrum switching functionality. On the other hand, since all the channels fed to input port 2 require to be switched to output port 4, a simple cross-connection in the optical backplane is set. The same

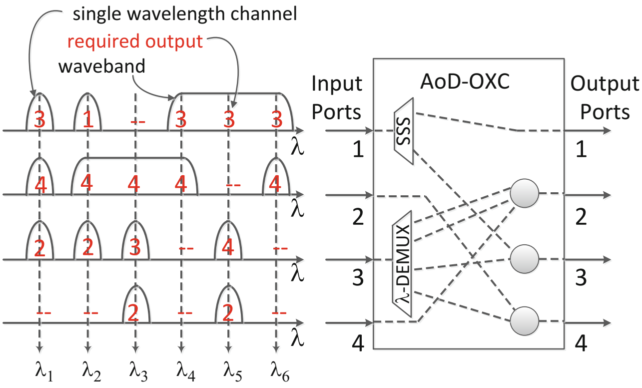



**Fig. 6.23** Example of AoD–OXC synthesis with respect to a particular traffic scenario

consideration is made at input port 4. A μ-demux is placed at input port 3 since signals fed require being wavelength-demultiplexed. Outputs 2, 3, and 4 need to couple signals from multiple sources. As a result of this synthesis mechanism, in this example the number of hardware modules (components) required is 5 and the number of cross-connections is 13 (considering all the ports from the components are connected to the optical backplane). Based on similar synthesis algorithms represented in Fig. 6.23, different scenarios have been analyzed in [46] with respect to the power consumption and the number of components synthesized. It is confirmed that more than 70 % reduction on the power consumption is achieved due to traffic aggregation into fiber switching of the AoD node compared to other architectures.

In addition to the synthesis enhanced by the programmability available in the AoD node, since flexible optical nodes are considered in the optical network and a series of on-demand optical layer functions are required by the end user, who is unaware of the optical hardware resources available, a new paradigm known as Network Function Programmability (NFP) emerges [47–49]. NFP paves the way for a new type of programmability complementary to SDN and NFV, which are enablers of network programmability at the software level focusing on control, management, services and assuming a hardware without intelligence and function-rigid. Thus, to understand NFP it is important to recognize its building blocks, which are the NFP nodes. This AoD-based NFP optical node establishes the foundations of programmable OXCs with abstraction of optical layer functions. NFP nodes are described in the following subsection.

#### Function Programmable OXCs

As it was previously mentioned, based on the programmable capabilities of the NFP nodes, the realization of programmable and synthetic optical white box nodes can be undertaken. In these nodes, a set of abstracted functions can be programmed utilizing the available resources. The functions in the NFP node can be at higher layers (e.g., routing, switching) and/or the physical layer (e.g., amplification, regeneration) under request of the network user. These NFP nodes consist of:

- *Function module pools*. A collection of resources (hardware systems) is programmed to satisfy various optical and electronic networking functions at different layers.
- *Synthesizing interconnection elements*. Interconnecting components to support the synthesis of a requested functionality (e.g., programmable optical backplane).
- *Add/drop interfaces*. To be able to insert or drop traffic locally, a sliceable transceiver with programmability characteristics can be adopted.

These NFP nodes are the fundamental building blocks for functional synthesizable networks. Based on the NFP nodes, the liberation of hardware including components and devices from fixed and static architectures is achieved, allowing software-defined hardware functions for network-wide operations. In addition, considering the available set of functions of the NFP node, the entire multidimensional network can be synthesized and sliced to provide a specific requirement. For instance, considering a SDM multi-core network, several fiber cores can be used in a functional network A and other fiber cores will be used for a functional network B. The bank of functions for the NFP nodes can be centrally or locally programmed to synthesize the appropriate network functionality supported by the node functionality adopted. Therefore, a network appears even at the node level, considering several functional modules available.

An example of an NFP node was demonstrated in [50]. This FPGA-based optical NFP node is implemented with functions, which can operate in aggregation as well as in isolated groups, enabling the node to fully support hardware virtualization and creating slices of the node associating them with arbitrary traffic types. The functions are categorized as intra and inter functions, and the node is capable to switch in between any of them. Hitless network function switchover was verified with guaranteed transport of parallel virtual slices, capable of 9.18 Gb/s out of 10 Gb/s throughput, flexible link granularity (6.85 Mb/s to 9.18 Gb/s) and single time-slice granularity (6.85 Mb/s to 2.4 Gb/s), variable latency (1.747–118.233 μs) and jitter (10–30 μs).

## Optical Layer NFP

As it was previously described, the main building blocks of a programmable node and network are components in each node which are interconnected to create various node functions. This programmability in the data plane is achieved by building the nodes with a modular and flexible architecture, such as the NFP node architecture, where functionalities such as fixed/flex-grid switching with or without time multiplexing are achieved on demand. For instance, in [51] it is demonstrated the use of NFP for the formation of different combination of components by cross-connecting the mounted modules in the optical backplane, resulting in different modes of operation and performance of the node. Based on this flexibility and empowered by programmability in the data and control planes, flex-grid and fixed-grid networking components are plugged into the optical backplane to support hybrid operations (see Fig. 6.24).

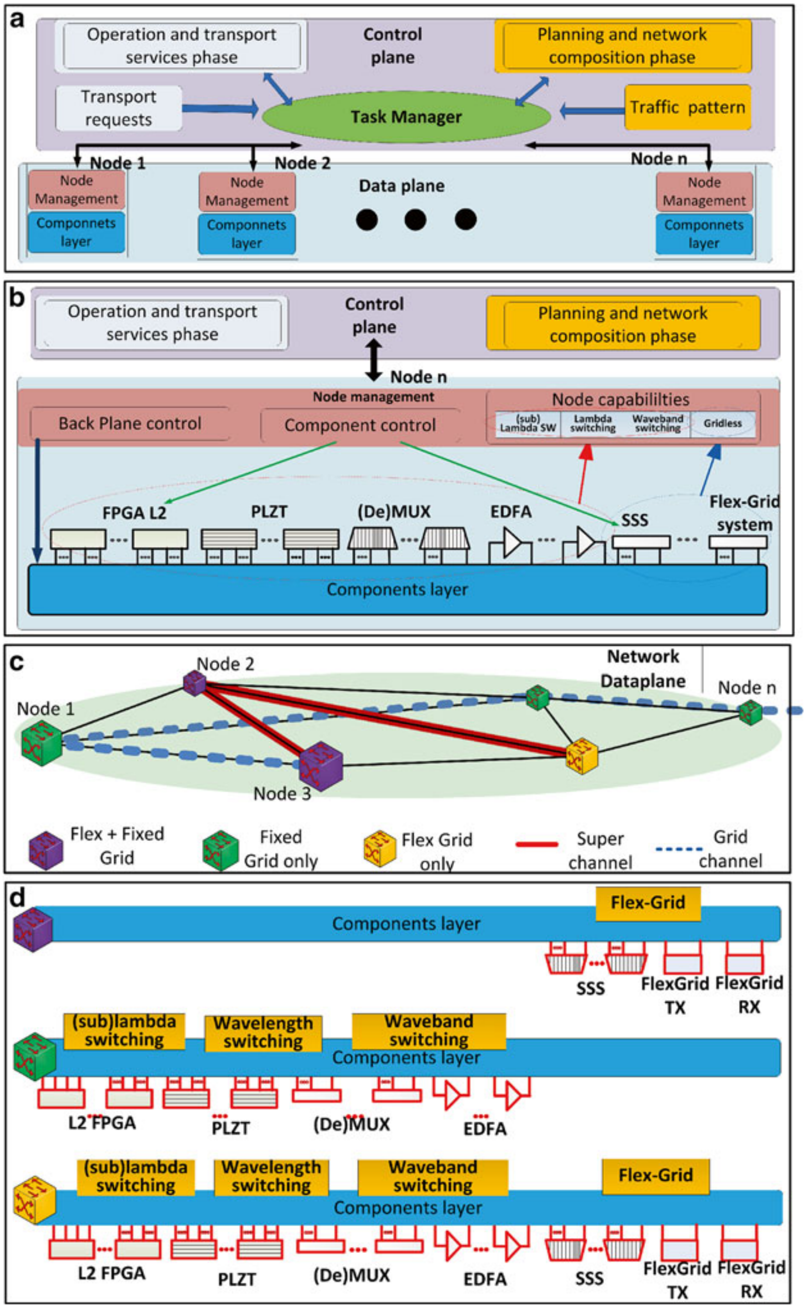



**Fig. 6.24** (**a**) High-level view of the proposed network architecture built using a multi-technology data plane and a modular central controller. (**b**) Flexible node architectures in the proposed network composed of components and management layers, communicating with the control plane. (**c**) Sample scenario of having nodes with different fixed-/flex-grid networking characteristics. (**d**) Composition of node components to enable different modes of operation in the data plane of the node

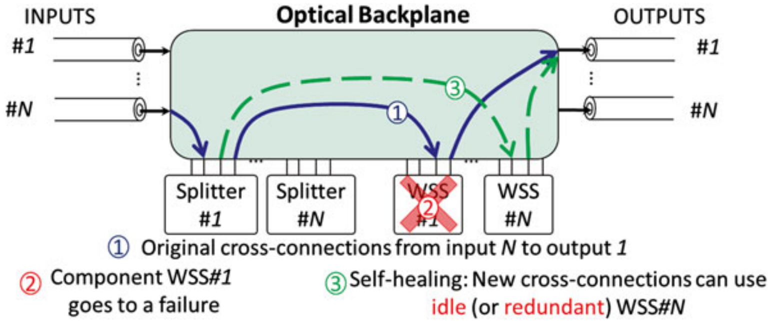



**Fig. 6.25** Self-healing concept based on an AoD ROADM configuration with 1:N protection

This programmability abovementioned together with the multidimensionality provided by the NFP node allows fully programmable multidimensional networks. This feature is observed in [41], where an elastic, SDM multi-granular network is designed and implemented, switching in three dimensions (space, time and frequency) and with over 6000-fold bandwidth granularity. This SDM multidimensional switching network supports high bandwidth flexibility, mixed traffic, conventional single core and new multi-core fibers, leveraging programmable NFP nodes, incorporating varying degrees of networking functionality. In this switching network, the end-to-end transport of 5.7-Tb/s traffic, using a combination of fixed/flex-grid, elastic band and sub-wavelength switching is demonstrated and channels include $4\times555$ Gb/s, $60\times42.7$ Gb/s, and $38\times10$ Gb/s wavelength channels, plus $12\times42.7$ Gb/s and $6\times10$ Gb/s time-multiplexed sub-wavelength.

Furthermore, enabled by this programmable-based synthetic feature of the NFP-OXC, self-healing or protection mechanisms can be undertaken in future optical nodes. It has been demonstrated in [52] that self-healing ROADM configurations with the NFP-OXC implementation are possible (Fig. 6.25).

In Fig. 6.25, the self-healing protection mechanism is illustrated. Based on the NFP-OXC concept, an optical backplane will interconnect input to output ports. Simultaneously, some ports of this optical backplane will allow to plug-in redundant units to cope with component failures. Based on this mechanism, additional devices such as bandwidth variable WSSs (BV-WSSs) are used to provide new redundant paths, thereby protecting from single BV-WSS failure. The synthetic programmable node deploys one redundant component of each type and provides on-the-fly protection from diverse failures by reconfiguring the optical backplane.

## 6.6 Summary

This chapter provides basic introduction to an important network entity, the optical node. An extensive list of requirements and design rules to deliver optical nodes for elastic optical networks are detailed. This provides the baseline and the link between subsystem designers and fabricators to node integrators and architects as well as network operators. In addition, and following the key structural elements of such node, a section was dedicated to numerous architectures able to construct the bypass/express system. A historical perspective is provided, reporting architectures that combine different type of components, i.e., LCOS, couplers, etc. To address future challenges of increased nodal degree and/or fiber number per node more advances approaches are reported. To interface such bypass structure with local access points, an add/drop multi-layer system is required. As such, numerous aspects of OTN over WDM and elastic optical networks are discussed, following numerous standardization activities and recent progress on ODUflex and beyond. Finally, and considering the extensive interest of the research community and commercial world on more programmable networks by means of SDN, NFV, and white boxes, a section was dedicated to optical nodes and systems able to offer NFP down at the optical layer. The fundamental principle based on AoD was described as the first optical white box system. A number of applications are reported as well as tools, control plane, and algorithms described to provide a dynamic synthetic and programmable optical infrastructure. There are considerable challenges identified and some initial node architectures and approaches when trying to scale elastic optical networks by means of SDM.

## References


1. S. Gringeri, B. Basch, V. Shukla, R. Egorov, T.J. Xia, Flexible architectures for optical transport nodes and networks. IEEE Commun. Mag. **1**, 40–50 (2010)
2. S. Poole, S. Frisken, M. Roelens, C. Cameron, Bandwidth-flexible ROADMs as network elements, in *OTuE1, OSA/OFC/NFOEC*, 2011
3. B. Collings, New devices enabling software-defined optical networks. IEEE Commun. Mag. **1**, 66–73 (2013)
4. D.M. Marom et al., Wavelength-selective 1_4 switch for 128 WDM channels at 50 GHzspacing, in *Proceedings of Optical Fibre Communications (OFC 2002)*, Anaheim, CA, Postdeadline Paper FB7
5. J. Kondis et al., Liquid crystals in bulk optics-based DWDM optical switches and spectral equalizers, in *Proceedings LEOS 2001*, Piscataway, NJ, 2001, pp. 292–293
6. G. Baxter et al., Highly programmable wavelength selective switch based on liquid crystal on silicon switching elements, in *Proceedings OFC/NFOEC 2006*, OTuF2, Anaheim, California, USA, 2006
7. T. Strasser, ROADM technology evolution. Presented at the LEOS annual meeting, Long Beach, CA, 2008, Paper TuH1
8. R.-J. Essiambre et al., Capacity limits of optical fibre networks. J. Lightwave Technol. **28**(4), 662–701 (2010)
9. S. Gringeri et al., Technical considerations for supporting data rates beyond 100 Gb/s. IEEE Commun. Mag. 2012
10. K. Roberts et al., Flexible transceivers, in *2012 ECOC*, Paper We3A3
11. A. Peters, E. Hugues-Salas, G. Zervas, D. Simeonidou, Design of elastic optical nodes based on subsystem flexibility measurement and other figures of merit, in *ECOC* 2015
12. K. Sato, *Advances in Transport Network Technologies* (Artech House, Norwood, 1996)
13. K. Sato, S. Okamoto, H. Hadama, Optical path layer technologies to enhance B-ISDN performance, in *Proceedings ICC'93*, Geneva, vol. 3, 1993, pp. 1300–1307

14. A. Watanabe, S. Okamoto, K. Sato, Optical switch using WDM, Patent No. 3416895
15. A. Watanabe, S. Okamoto, K. Sato, M. Okuno, Optical switch, Patent No. 3444548
16. M. Koga et al., 8×16 delivery and coupling type optical switches for a 320 Giga-bit/s throughput optical path cross-connect system, in *OFC '96*, ThN3, San Jose, February 25-March 1, 1996, pp. 259–261
17. M.D. Feuer, S.L. Woodward, Advanced ROADM networks, in *OFC/NFOEC 2012*, NW3F.3, March 2012
18. I. Kim, P. Palacharla, X. Wang, D. Bihon, M.D. Feuer, S.L. Woodward, Performance of colorless, non-directional ROADMs with modular client-side fibre cross-connects, in *OFC/NFOEC 2012*, NM3F.7, Los Angels, March 2012
19. T. Zami, Contention simulation within dynamic, colorless and unidirectional/multidirectional optical cross-connects, in *ECOC 2011*, We.8.K.4, Geneva, September 2011
20. H. Ishida, H. Hasegawa, K. Sato, An efficient add/drop architecture for large-scale subsystem-modular OXC, in *15th International Conference on Transparent Optical Networks, ICTON 2013*, We.A1.5, Cartagena, Spain, June 23–27, 2013
21. H. Ishida, H. Hasegawa, K. Sato, Hardware scale and performance evaluation of compact OXC add/drop architecture, in *OFC/NFOEC 2014*, W1C.7, San Francisco, March 9–14, 2014
22. K. Sato, How to create large scale OXC/ROADM for the future networks, in *16th International Conference on Transparent Optical Networks (ICTON 2014)*, Graz, Austria, July 6–10, 2014
23. K. Sato, Implication of inter-node and intra-node contention in creating large throughput photonic networks, in *IEEE Optical Network Design and Modeling Conference, ONDM 2014*, Stockholm, May 19–22, 2014
24. K. Sato, H. Hasegawa, Prospects and challenges of multi-layer optical networks. IEICE Trans. Commun. **E90-B**(8), 1890–1902 (2007)
25. S. Mitsui, H. Hasegawa, K. Sato, Hierarchical optical path cross-connect node architecture using WSS/WBSS, in *Photonics in Switching 2008*, S-04-1, Hokkaido, Japan, August 4–7, 2008
26. K. Ishii, H. Hasegawa, K. Sato, M. Okuno, S. Kamei, H. Takahashi, An ultra-compact waveband cross-connect switch module to create cost-effective multi-degree reconfigurable optical node, in *ECOC 2009*, Vienna, Austria, September 20–24, 2009, 4.2.2
27. T. Ban et al., Development of large capacity ultra-compact waveband cross-connect, in *16th Opto-Electronics and Communications Conference, OECC 2011*, 6A1-2, Kaohsiung, Taiwan, July 4–8 2011
28. K. Ishii et al., Monolithically integrated waveband selective switch using cyclic AWGs, in *ECOC 2008*, Mo.4.C.5, Brussels, September 22–25, 2008
29. Y. Taniguti, Y. Yamada, H. Hasegawa, K. Sato, Coarse granular optical routing networks utilizing fine granular add/drop. IEEE/OSA J. Opt. Commun. Netw. **5**(7), 774–783 (2013)
30. Y. Terada, Y. Mori, H. Hasegawa, K. Sato, Enhancement of fibre frequency utilization by employing grouped optical path routing, in *OFC/NFOEC 2014*, W1C.6, San Francisco, March 9–14, 2014
31. T. Ban, H. Hasegawa, K. Sato, T. Watanabe, H. Takahashi, A novel large-scale OXC architecture that employs wavelength path switching and fibre selection, in *ECOC 2012*, We.3.D.1, Amsterdam, September 16–20, 2012
32. L.H. Chau, H. Hasegawa, K. Sato, Performance evaluation of large-scale OXC architectures that utilize intra-node routing restriction, in *OECC/PS 2013*, MQ2-2, Kyoto, June 30–July 4, 2013
33. Y. Iwai, H. Hasegawa, K. Sato, Large-scale photonic node architecture that utilizes interconnected small scale optical cross-connect sub-systems, in *ECOC 2012*, We.3.D.3, Amsterdam, September 16–20, 2012
34. Y. Iwai, H. Hasegawa, K. Sato, A large-scale photonic node architecture that utilizes interconnected OXC subsystems. OSA Opt. Express **21**(1), 478–487 (2013)
35. Y. Tanaka, Y. Iwai, H. Hasegawa, K. Sato, Subsystem modular OXC architecture that achieves disruption free port count expansion, in *ECOC 2013*, Th.2.E.4, London, September 2013
36. H. Huang et al., 100 Tbit/s free-space data link enabled by three-dimensional multiplexing of orbital angular momentum, polarization, and wavelength. Opt. Lett. **39**(2), 197–200 (2014)
37. ITU-T Recommendation, Architecture of optical transport networks, Series G: Transmission Systems and Media, Digital Systems and Networks, Digital networks—Optical Transport Networks, October 2012

38. ITU-T G.707/Y.1322, Implementers' Guide, Series G: Transmission Systems and Media, Digital Systems and Networks, June 2010
39. ITU-T G.7044/Y.1347, Hitless adjustment of ODUflex (GFP), Series G: Transmission Systems and Media, Digital Systems and Networks, October 2010
40. ITU-T G.7041/Y1303, Generic framing procedure, Series G: Transmission Systems and Media, Digital Systems and Networks, April 2011
41. N. Amaya et al., Fully-elastic multi-granular network with space/frequency/time switching using multi-core fibres and programmable optical nodes. Opt. Express **21**(7), 8865–8872 (2013)
42. N. Amaya et al., Introducing node architecture flexibility for elastic optical networks. J. Opt. Commun. Netw. **5**(6), 593–608 (2013)
43. M. Garrich, N. Amaya, G. Zervas, J.R.F. Oliveira, P. Giaccone, A. Bianco, D. Simeonidou, J.C.R.F. Oliveira, Architecture on Demand Design for High-Capacity Optical SDM/TDM/FDM Switching. IEEE/OSA J. Opt. Commun. Netw. **7**(1), 21–35 (2015)
44. A. Muhammad et al., Flexible and synthetic SDM networks with multi-core-fibres implemented by programmable ROADMs, in *Proceedings of ECOC 2014*, Cannes, France, Paper P.6.6
45. N. Amaya et al., Software defined networking (SDN) over space division multiplexing (SDM) optical networks: features, benefits and experimental demonstration. Opt. Express **22**(3), 3638–3647 (2014)
46. M. Garrich et al., Power consumption analysis of architecture on demand, in *Proceedings ECOC 2012*, Amsterdam, Netherlands, Paper P5.06
47. A. Muhammad, G. Zervas, N. Amaya, D. Simeonidou, R. Forchheimer, Introducing flexible and synthetic optical networking: planning and operation based on network function programmable ROADMs. IEEE J. Opt. Commun. Netw. **6**(7), 635–648 (2014)
48. G. Zervas et al., Network function programmability and software-defined synthetic optical networks for data centres and future Internet, in *Proceedings Photonics in Switching (PS) 2014* San Diego, USA, Paper PM4C.3
49. A. Muhammad et al., Introducing flexible and synthetic optical networking: planning and operation based on network function programmable ROADMs. J. Opt. Commun. Netw. **6**(7), 660–669 (2014)
50. Y. Yan et al., FPGA-based optical network function programmable node, in *Proceedings OFC 2014*, San Francisco, USA, Paper W1C.1
51. B. Rahimzadeh Rofoee et al., All programmable and synthetic optical network: architecture and implementation. J. Opt. Commun. Netw. **5**(9), 1096–1110 (2013)
52. M. Dzanko, M. Furdek, G. Zervas, D. Simeonidou, Evaluating availability of optical networks based on self-healing network function programmable ROADMs. IEEE/OSA J. Opt. Commun. Netw. **6**(11), 974–987 (2014)
53. Transmode App Note, Transmode's Flexible Optical Networks, http://www.transmode.com/en/technologies/flexible-optical-networks
54. S. Okamoto, A. Watanabe, K. Sato, Optical path cross-connect architecture for photonic transport network. Special Joint Issue IEEE J. Lightwave Technol. IEEE J. Sel. Areas Commun. **14**(6), 1410–1422 (1996)
55. Y. Ishii, K. Hadama, J. Yamaguchi, Y. Kawajiri, E. Hashimoto, T. Matsuura, F. Shimokawa, MEMS-based 1×43 wavelength-selective switch with flat passband, in *ECOC* 2009
56. S. Kakehashi et al., IEICE Trans. Commun. **E91-B**(10), 3174–3184 (2008)

# Chapter 7
# Sliceable Bandwidth Variable Transponders

**Juan Pedro Fernández-Palacios, Víctor López, Beatriz de la Cruz, Ori Gerstel, Nicola Sambo, and Emilio Riccardi**

## Acronyms

| | |
|---|---|
| BCJR | Bahl-Cocke-Jelinek-Raviv |
| BVT | Bandwidth Variable Transponders |
| EON | Elastic Optical Network |
| MILP | Mixed Integer Linear Programming |
| MZM | Mach Zehnder Modulator |
| OTLC | Optical channels Transport Lanes |
| OTN | Optical Transport Network |
| OTU | Optical Transport Unit |
| OXC | Optical Cross-Connect |
| ROADM | Reconfigurable Optical Add–Drop Multiplexer |
| S-BVT | Sliceable Bandwidth Variable Transponder |
| TG | Tributary Group |
| WSON | Wavelength Switched Optical Networks |

J.P. Fernández-Palacios (✉) • V. López • B. de la Cruz
Telefónica I+D, Madrid, Spain
e-mail: juanpedro.fernandez-palaciosgimenez@telefonica.com

O. Gerstel
Sedona Systems, Ra'anana, Israel

N. Sambo
Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Parma, Italy

E. Riccardi
Telecom Italia, Milan, Italy



V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_7

## 7.1 Introduction

The Elastic Optical Networking (EON) approach advocates the use of new building blocks for an extended flexibility on resource assignment (be it capacity or spectrum) and an optimized use of the network capacity. The main blocks of EON are the *flexi-grid Reconfigurable Optical Add–Drop Multiplexer (ROADM)* and the *Bandwidth Variable Transponder (BVT)*, presented in Chaps 5 and 6, respectively. Let us emphasize that *BVTs* can adjust their transmission rate to the actual traffic demand by expanding or contracting the bandwidth of an optical path (i.e. varying the number of sub-carriers) and by modifying the modulation format, as depicted in Fig. 7.1a. There have been several demonstrations of bit-rate-variable transmitters where the number of sub-carriers or the modulation format is adapted to achieve the desired bit-rate and spectral efficiency (e.g. [1, 2]).

However, when a high-speed BVT is operated at lower than its maximum rate, for instance due to required reach or impairments in the optical path, part of the BVT capacity is wasted. In order to address this issue, the S-BVT has been proposed [3]. An S-BVT is able to allocate its capacity into one or several optical flows that are then transmitted to one or several destinations, as illustrated in Fig. 7.1b. Thus, when an S-BVT is used to generate a low bit-rate channel, its remaining capacity can also be exploited for transmitting other independent data flows. From the point of view of higher layers, an S-BVT may be viewed either as a high-capacity BVT or as a collection of multiple logically/virtually independent lower-capacity BVTs, depending on the mode of operation.

The S-BVT has gained a great attention in the scientific community as a new way to perform optical grooming, achieving an efficient use of the transponders in the network. There have been studies to find its target cost to justify its commercialization [4, 5], the potential savings in terms of operational savings [6], and its adoption in metro networks with economical architectures [7].

This chapter presents the architectures to build S-BVTs, its interface with IP routers, and finally a planning analysis of the economic justification to deploy these transponders in real scenarios.

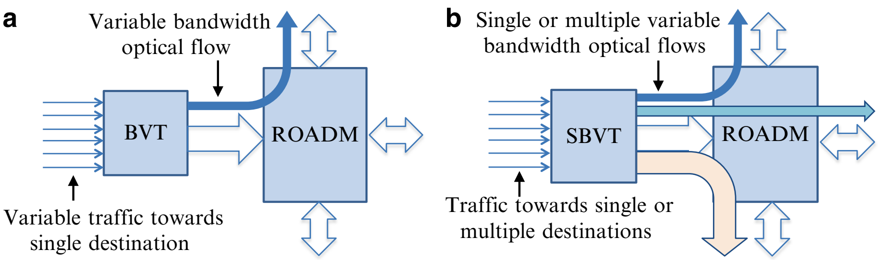



**Fig. 7.1** Schematics of (**a**) bandwidth variable transceiver (BVT) and (**b**) sliceable bandwidth variable transceiver (S-BVT)

## 7.2 Sliceable Bandwidth Variable Transponder Architectures

### 7.2.1 S-BVT Requirements

Several S-BVT implementations will be presented in this chapter to satisfy the following requirements:

(a) *Support of slice-ability*: the transponder should be able to generate sub-carriers that can be *sliced*, so that they can either form a super-channel (connection composed of several co-routed and adjacent sub-carriers) or be separated and directed towards different output ports (i.e. paths or destinations).
(b) *Ability to generate sub-carriers with configurable channel spacing*: the transponder should be able to generate sub-carriers with configurable channel spacing, which may vary depending on their bit-rate and modulation format. Moreover, this property may be also a requirement for slice-ability. As an example, by referring to the super-channel demonstrated in [8], sub-carriers are spaced by 28 GHz. Such spacing is allowed only when those sub-carriers are considered as a super-channel that, as a single entity, occupies overall bandwidth of 200 GHz (28 GHz times seven sub-carriers); when sub-carriers are sliced, sub-carrier spacing may become 37.5 GHz, which is the smallest ITU-T flex-grid spacing [9] larger than 28 GHz.
(c) *Multi bit-rate support*: the transponder should be able to support several bit-rate values, by varying the number of used sub-carriers or by changing the bit-rate of each sub-carrier.
(d) *Reach adaptation*: a trade-off between all-optical reach and spectral efficiency can be found through the support of multiple modulation formats and/or code adaptation. The transponder should support different modulation formats; for instance, each sub-carrier may be transmitted with PM-16QAM or PM-QPSK, depending on the optical reach requirements. The amount of redundancy transmitted within each sub-carrier (e.g. with time-frequency packing (TFP)) can be adapted to the physical characteristics of the path.

### 7.2.2 S-BVT Architecture

A general S-BVT architecture is presented in Fig. 7.2. Multiple traffic demands are associated with Optical Transport Network (OTN) transport units (i.e. optical transport unit—OTUCn as discussed below) through a configurable/sliceable flexible OTN framer module capable of dynamically partitioning its total throughput into dynamically created multiple OTUCn virtual interfaces. The same module should also be able to fragment each OTUCn among several Tributary Groups (OTUCnTG), if needed, to allow the distribution of load over different media channels. Data from

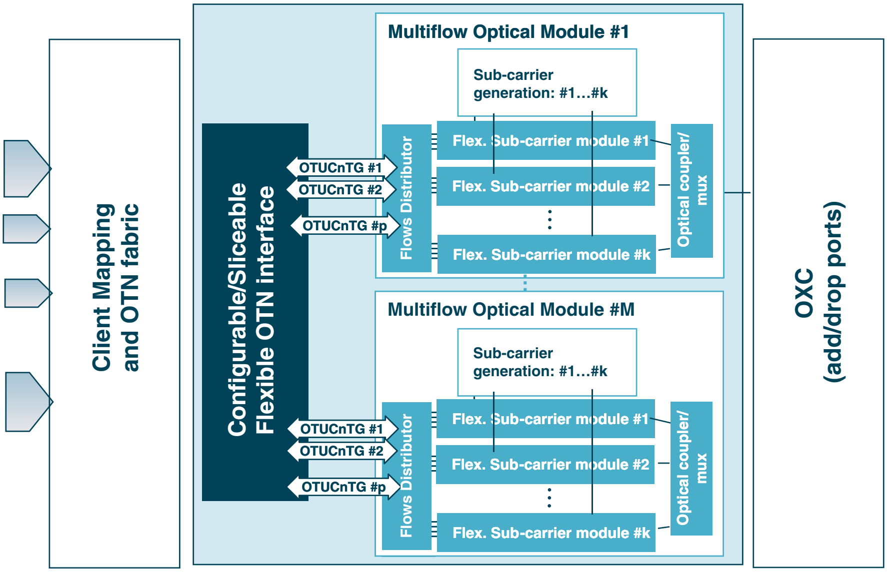



**Fig. 7.2** S-BVT architecture

all the generated OTUCnTGs feeds a variable pool of Multi-flow Optical modules (the S-BVT's optical front end), thanks to inverse multiplexing in a number of Optical channels Transport Lanes (OTLC), via a multi-lane interface.

The $p$ independent flows (in the form of multi-lane aggregates) are then used to modulate the optical sub-carriers, in the Multi-flow Optical module, which is composed of a Flow Distributor module, a Sub-carrier generation module generating up to $k$ ($k \geq p$) non-modulated sub-carriers, $k$ of Flex Sub-carrier modules (where each sub-carrier is modulated by a specific traffic portion), and a multiplexer module. First, the Flow Distributor (e.g. an electronic switching matrix) directs a suitable number of lanes of the multi-lane aggregate to a specific Flex Sub-carrier module. The Flow Distributor enables the traffic to modulate a single sub-carrier for transmission in the optical layer. The Flex Sub-carrier module comprises a pool of coherent front-ends (light sources, I/Q modulators, drivers, and receivers) associated with the processing electronics (digital to analogue converter—DAC, analogue to digital converter—ADC, transmitter, receiver, digital signal processing—DSP—at the receiver, FEC coding and decoding). The bit-rate and bandwidth can be tuned by changing the modulation format, the baud rate, the spectral shaping, and the coding. All generated sub-carriers are first logically grouped into super-channels distributed on different media channels and optically coupled together into a single add/drop port of the source Optical Cross-Connect (OXC) node. Then, with proper node configuration, media channels are routed independently and transparently across the photonic network up to specific (and different) destination nodes.

Several solutions can be considered for implementing the modules shown in Fig. 7.2, as detailed in the next sections.

### Flex Sub-Carrier Module

The Flex Sub-carrier module can be implemented focusing on specific transmission techniques, which allows the S-BVT suitable for different scenarios (e.g. long haul or metro/regional network), presenting different levels of spectral efficiency (measured with b/s/Hz) and complexity. An S-BVT, in particular the Flex Sub-carrier module, can be based on Nyquist WDM, Orthogonal frequency-division multiplexing (OFDM), or TFP transmission techniques. Table 7.1 summarizes the main characteristics (e.g. spectral efficiency) depending on the transmission technique implemented in the Flex Sub-carrier module.

### Sub-Carrier Generator Module

The generation of $N$ sub-carriers can be achieved through an array of $N$ lasers or the use of a multi-wavelength (MW) source. The two different solutions present advantages and drawbacks.

The array of tuneable lasers (fully in the C band) avoids any constraint in the Routing and Spectrum Allocation (RSA) algorithm, especially when slice-ability is exploited. The sub-carriers can be generated at any *central frequency*, so that they can be used in any portion of the spectrum, depending on *frequency slot* [9] availability along the path. As a drawback, the frequency drift of each laser composing the array may drive to the adoption of guard band (thus, reducing spectral efficiency) in order to avoid detrimental sub-carrier overlapping. In fact, normally a laser instability of 1 GHz can be assumed, and that can result in a guard band of 2 GHz between two adjacent sub-carriers. Thus, considering a super-channel of eight carriers, about 14 GHz would be wasted because of laser drift.

Conversely, an MW source is able to produce $N$ sub-carriers (e.g. $N=4$) from a single laser source. Such generated sub-carriers are locked together. This implies that a drift of the parent laser (i.e. the laser constituting the MW source) does not imply an undesired sub-carrier overlapping since the sub-carriers shift together in accordance with the drift. In this case, the guard band (of 14 GHz) used in the previous case would be avoided resulting in a higher spectral efficiency. Moreover, the reduction on the number of lasers ($N-1$ lasers can be saved with respect to an array of lasers) may reduce the cost of S-BVT. As a drawback, the sub-carriers generated by an MW source presents constraint in the spectrum assignment. Indeed, they are typically symmetrically spaced each other and their spacing can be limited to a maximum value (e.g. 50 GHz). As an example, a comparison between network blocking probability performance of an MW source and an array of lasers has been shown in [8]: in an EON, by fixing the same number of S-BVTs, the technology based on the array achieves lower connection blocking probability than the MW source at the expense of the adoption of a larger number of lasers.

Various techniques have been demonstrated for the MW source, based on mode-locked semiconductor and fibre lasers, and electro-optic modulators [10–13]. The former suffers from the bottleneck of fixed free spectral range (FSR) hence not feasible

**Table 7.1** Main S-BVT characteristics depending on the type of transmission techniques implemented in Flex Sub-carrier module

| Flex Sub-carrier module transmission technique | Max spectral efficiency[a] | All-optical reach | Cost & complexity of the Flex Sub-carrier module | Use cases |
|---|---|---|---|---|
| Nyquist WDM | Dependent on the modulation format; channel spacing is at least equal to the symbol rate: e.g. 4 b/s/Hz with PM-QPSK and 8 b/s/Hz withPM-16QAM | Dependent on the modulation format (several thousand km for PM-QPSK and PM-16QAM, less for higher formats) | Mainly driven by DAC, ADC, DSP | Long haul transmission with high spectral efficiency |
| | | | (e.g. electronic bandwidth ≈ 0.5 × baud rate) | |
| Time-frequency packing (with PM-QPSK) | Channel spacing can be smaller than the symbol rate: e.g. 8 b/s/Hz [22] | Thousands of km with variable spectrum efficiency: e.g. 3000 km (with 5.16 b/s/Hz), 5500 km (with 4.2 bit/s/Hz) [23] | Mainly driven by ADC and DSP (DAC is not required), e.g. electronic bandwidth <0.5 × baud rate | Long haul transmission with high spectral efficiency |
| | | | A sequence detector is required instead of a symbol-by-symbol detector | |
| O-OFDM | Dependent on the modulation format; channel spacing is at least equal to the symbol rate: e.g. 4 b/s/Hz with PM-4QAM and 8 b/s/Hz withPM-16QAM | Dependent on the modulation format and detection scheme (several thousand km for PM-4QAM and PM-16QAM and CO-OFDM, less for higher formats or cost-effective metro solutions) | Mainly driven by DAC, ADC, DSP | Core/long-haul |
| | | | (e.g. electronic bandwidth ≈ 0.5 × baud rate) | Metro/regional |

[a]Spectral efficiency (SE) is defined as the information rate over a given bandwidth. SE can vary depending on BL, FEC, and front-ends adopted

for S-BVT because carrier spacing cannot be tuned (see requirement b in Sect. 7.2.1), while the latter type provides tuneable carrier spacing [10] (equally-spaced sub-carriers are assumed) but suffers from bias drift in modulators. However, this drift limitation can be overcome y using bias controllers, a standard practice in commercial applications. Moreover, MW for S-BVTs should also offer an integrated design comprising an integrated tuneable laser and integrated multi-carrier generator. Hybrid III-V and Si tuneable integrated lasers were also recently demonstrated [14].

Regarding multi-carrier generator part, modulators based designs can provide the flexibility and performance required for S-BVTs, since designs involving fibres are not integrable. A block diagram of the schematic for an MW source based on design in [13] is shown in Fig. 7.3. A tuneable integrated laser provides the parent carrier source which is coupled with dual drive Mach Zehnder Modulator (DD-MZM). A sinusoidal RF signal is fed to one arm and its double frequency on the second arm by use of frequency doubler. This design has the capability to generate arbitrary number of lines in the range of 3, 4, 5, and 9 by simply adjusting RF drive signals amplitudes, thus meeting *requirement c* of supporting multiple rates. The sub-carrier spacing (symmetric) can be adjusted by changing the RF frequency (*requirement b*S-BVT) and the whole set of sub-carriers can also be tuned by simply tuning the parent laser. Once sub-carriers are generated, the proper sub-carrier will be selected and directed to the corresponding Flex Sub-carrier module. Currently optical filters (e.g. with tuneable micro ring resonators) are used to navigate the selected sub-carriers to through each spectral component to the proper Flex Sub-carrier module [15], until a new design for MW sources which does not require optical filters becomes available in future.

### 7.2.3 *Example of an S-BVT Supporting 400 Gb/s*

Figure 7.4 shows the S-BVT architecture proposed in [15]) designed specifically for a multi-rate, multi-format, code-adaptive S-BVT able to serve clients from four 100 GbE interfaces. At the transmitter side, by referring to the general S-BVT architecture shown in Fig. 7.2, the Flow Distributor is here implemented with an electronic switching matrix, while the Flex Sub-carrier module, supporting TFP, is decoupled in two parts: the electronic processing module and the Photonic Integrated Circuit (PIC), used as sub-carrier transmitter. The figure shows the architecture when the *N* sub-carriers are generated (by the sub-carrier generation module) by a single MW source; however, such a source may be replaced with an array of *N* lasers, one per sub-carrier. The architecture enables an information rate of 400 Gb/s.

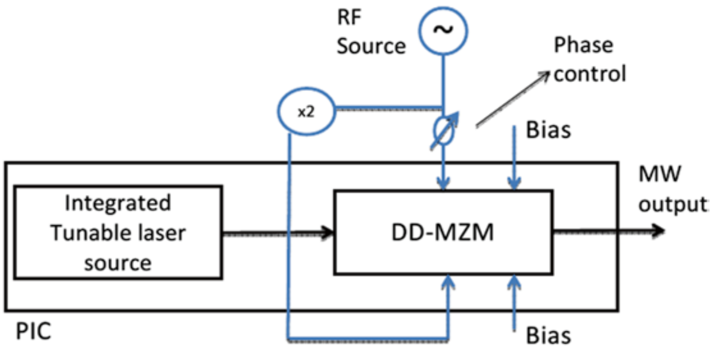



**Fig. 7.3** MW source design based on electro-optic modulators

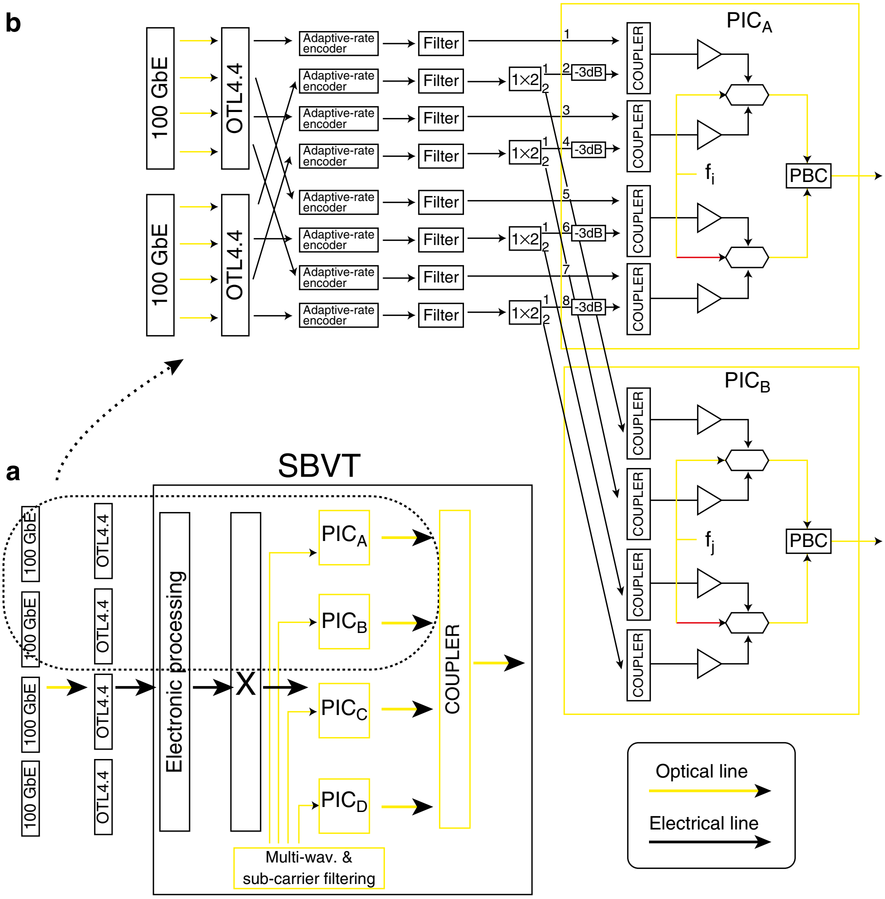



**Fig. 7.4** Example of S-BVT architecture supporting 4 × 100 GbE interfaces (**a**) and details (**b**)

As shown in Fig. 7.4a, client traffic coming from the 100 GbE ports can be encapsulated in OTN frames (e.g. for monitoring purposes), in particular OTL4.4 (i.e. four OTU4 physical lanes). Then, those frames are processed in the electronic domain of the S-BVT. This stage may involve electronic filtering or adding redundancy (through the adaptive-rate encoder) for the TFP technique (adding low-density parity-check—LDPC—code). The adaptive-rate encoder changes the amount of redundancy in the transmission, satisfying *Requirement d* by code adaptation. Then, the encoded clients enter, through the switching matrix, to the corresponding PICs to modulate sub-carriers; the 400 Gb/s S-BVT architecture contains four PICs. Modulated sub-carriers are coupled together forming a super-channel; however, sub-carriers can be sliced by filters in the add-node.

The part of the S-BVT architecture in the dotted line square is expanded in Fig. 7.4b. Each 100 GbE interface provides four on-off-keying optical lines at 25 Gb/s [16], each encapsulated in one of the four OTL4.4 lanes. Note that the four

25 Gb/s lines coming from a 100 GbE interface are not independent; thus they must modulate the same sub-carrier. Then, OTN clients are encoded and filtered in the electronic domain and directed to a PIC. Each PIC can generate PM-16QAM or PM-QPSK, dependent on the number of client traffic and required optical reach, making the S-BVT supporting *multi-format* modulations. The attenuators (−3 dB) permit to obtain four-level electrical signals (the minimum number of signals required to apply 16QAM) that modulate the sub-carrier through the IQ modulator. PM-16QAM can be applied if eight data streams are provided by two 100 GbE interfaces to the PIC. For this purpose, the 1 × 2 electronic switches are set to their output port 1 (Fig. 7.4b). This way, two 100 GbE interfaces can be served by a 200 Gb/s (plus overhead) PM-16QAM generated from $PIC_A$ (while $PIC_B$ is excluded). Four 100 GbE interfaces can be served with 400 Gb/s (plus overhead) super-channel achieved through two 200 Gb/s (plus overhead) PM-16QAM sub-carriers, using two PICs ($PIC_A$ and $PIC_C$, while $PIC_B$ and $PIC_D$ are not used). After the electrical signal modulation and grooming process, Polarization Beam Combiner (PBC) provides polarization multiplexing as the last stage within PIC.

If PM-16QAM does not satisfy optical reach and a more robust modulation format is required, PIC can be used as PM-QPSK transmitter, which is achieved by proving four data streams, instead of eight. The four data streams to a single PIC used as PM-QPSK are provided by a single 100 GbE interface. For this purpose, the 1 × 2 electronic switches are set to their output port 2. This way, as shown in Fig. 7.4b, client traffic coming from a 100 GbE is directed to $PIC_A$, while clients coming from the other 100 GbE to $PIC_B$. Thus, thanks to 1 × 2 switches, the proposed S-BVT architecture preserves the bit-rate while supporting *multi-format* adaptation (PM-16QAM/PM-QPSK). In the case of PM-QPSK transmission, the encoder can use LDPC code for TFP to increase spectrum efficiency (see Table 7.1).

Some considerations on the electronics after the encoder are here reported if TFP is adopted. Assuming a scenario with 100 GbE composed of four lines at 25 Gb/s and OTN frames in the form of OTL4.4, with the overhead the speed of the four OTL lines is around 28 Gb/s, thus resulting in 112 Gb/s for sub-carrier, if PM-QPSK is adopted. To evaluate the impact of LDPC overhead in the modulator, spectral efficiency has to be considered. Code rate (which affects spectral efficiency) varies with the optical signal-to-noise ratio (SNR). By referring to experiments in [8], for a 3000 km path the spectral efficiency achieved by TFP is 5.16 b/s/Hz. This means that a 112 Gb/s sub-carrier is transmitted over about 22 GHz (including information and LDPC coding). Without TFP, it would be transmitted at 28 GHz, for instance with Nyquist based transmission. Considering a path of 5000 km, spectral efficiency is around 4.25 b/s/Hz, thus implying an electronic of around 26 GHz, still below 28 GHz. Finally, the possibility to include LDPC directly in the OTN framing could be also adopted.

At the receiver side, a coherent receiver must be used for each sub-carrier. The same coherent receiver can be used for sub-carriers with both PM-16QAM and PM-QPSK. A tuneable laser used as local oscillator (LO) must be adjusted in order to match the incoming signal wavelength. The two optical signals beat into the opto/electronic (O/E) conversion module providing four analogue electrical signals where information is completely mixed in terms of phase, amplitude, and polarization. Analogue-to-digital (A/D) conversion at high sampling rate, such as 50 Gs/s, is performed and real-time DSP is used in order to recover for the data. DSP, with a two-dimensional adaptive feed forward equalizer, fully compensates for linear and partially for nonlinear fibre transmission impairments and provides a feedback to the LO in order to lock its central frequency to the incoming signal. In the case of TFP, the receiver iteratively exchanges information with a LDPC decoder. If TFP is adopted, inter-symbol interference (ISI) requires a receiver based on sequence detection, as the Bahl–Cocke–Jelinek–Raviv (BCJR) detector. However, that brings complexity in the detector design itself as demonstrated in [8]. If NWDM is assumed, a symbol-by-symbol detector can be used (Fig. 7.5).

Four output lines are simultaneously active if PM-QPSK is received, while eight output lines are simultaneously active in case PM-16QAM is received.

### 7.2.4 Component Technologies, Complexity, and Integration

To cope with future networks' flexibility requirements (e.g. high traffic rate and long reach), the S-BVT must transmit advanced modulation formats at high symbol rates meeting both high spectral efficiency and long reach requirements, by adaptively looking for a trade-off between them. The key component parameters, constituting the Multi-flow Optical module, are: analogue bandwidth, effective resolution of the DAC, extinction ratio of the I/Q modulator, and time-frequency laser stability. In order to cope with hardware limitations of off-the-shelf components, advanced DSP algorithms are required, such as pre-distortion methods for the most relevant components and for partially compensating fibre (linear and nonlinear effects) propagation impairments. Moreover, authors in [17] showed that it may be beneficial from a performance point-of-view to increase the number of sub-carriers to cope with electronics speed limitations. However, this comes at the cost of roughly doubling the number of hardware components/costs.

A related component technology challenge is the integration that provides cost reduction as well as reducing physical dimensions. The S-BVT design can be enhanced if all the above functions could be integrated on a single platform. High integration levels also permit better monitoring, management, and control of system performance. The enabling integration technologies could be based on Silicon Photonics which allows for a well-adapted matching between the electronic part (e.g. DAC, driver amplifier, ADC, DSP, and FEC) and the optical section (e.g.

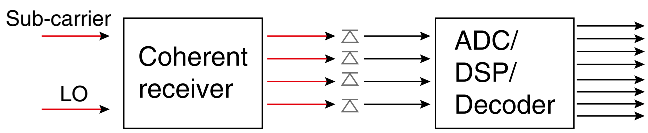



**Fig. 7.5** Coherent receiver

modulators, photodiode, laser source). Energy efficiency of an integrated system is a further benefit. Silicon photonics integration can help, for example, sharing thermal control and power dissipation functions among a subset of sub-channels, but with current hybrid approach the contribution to total power saving would be limited to about 10 %. To achieve significant power reduction, the more promising technology involves CMOS and photonics integration which will evolve in the next 5 years with a level of optoelectronic integration lower than 130 nm Silicon-On-Insulator (SOI) CMOS process (e.g. a single lithographic process can integrate hundreds of photonic components with millions of transistors, achieving significant power reduction benefits) [18]. For instance, based on CMOS photonics, an S-BVT sub-channel transceiver can be designed on a single chipset, equipped with a C-band tuneable laser (e.g. based on a Distributed Bragg Reflector laser), an optical MZM (e.g. based on Indium Phosphide), driven by an integrated electrical driver amplifier, a PIN photodiode followed by a trans-impedance amplifier and a clock and data recovery unit. All parts can be integrated into a thermally and power efficient package (i.e. the transceiver is intended as a pluggable module) and electrical interface to connect the transceiver on the S-BVT. Such integration will allow an S-BVT to transmit 1 Tb/s super-channel, while reducing system power consumption.

### *7.2.5 S-BVT Programmability Perspectives*

Given the requirements of flexibility, the S-BVT should also enable (remote) control and programmability. Indeed, an S-BVT finds application whenever the transmission rate is adapted to the actual traffic demand by expanding or contracting the bandwidth of an optical path (e.g. varying the number of sub-carriers), by adapting the optical reach, and by directing the generated super-channels towards specific destinations. To achieve these functionalities, several sub-carriers can be connected or disconnected and the modulation format or code rate can be modified based on the required optical reach. Therefore, several S-BVT transmission characteristics can be programmed by remote controllers, such as:

- Line rate.
- Number of sub-carriers.
- Association of physical lanes (i.e. OTN clients) to a specific Flex Sub-carrier module.
- Modulation format.
- Code rate.
- Required bandwidth.
- Optical carrier frequency.
- Specific detection strategies at the DSP (e.g. equalization).

A required traffic demand can be accommodated with a certain number of sub-carriers that can be flexibly adapted by activating a set of Flex Sub-carrier modules. The association of OTN flows with a set of optical sub-carriers is achieved by the automatic control of an electronic switch implementing the Flow Distributor shown in Figs. 7.2 and 7.4b. Optical reach adaptation can be obtained by using a proper modulation format (e.g. the more robust PM-QPSK instead of PM-16QAM) through the Flex Sub-carrier module. Adaptation of the code rate is done electronically via software by an encoder in the Flex Sub-carrier module. The optical carrier can be set by configuring the array of lasers or the MW source in the Sub-carrier generation module. Equalization in DSP is configured via software.

The sub-wavelength granularity of an OFDM-based S-BVT allows the programmability via DSP to be extended to the electrical sub-carrier level, enabling unique bandwidth manipulation, including arbitrary sub-carrier suppression, adaptive bit loading (BL), and power loading (PL). The BL consists of independently loading the sub-carriers with symbols mapped with different modulation formats. Thus, fine bit-rate selection and efficient use of spectral characteristics of transmission links can be achieved. Additionally, in order to optimize the system performance according to the channel profile, sub-carriers with lower SNR can be loaded with data mapped with the most robust modulation format. PL can also be implemented at the S-BVT DSP, in which each sub-carrier or a set of sub-carriers is multiplied by a gain coefficient to adaptively vary the sub-carrier amplitude. This introduces additional flexibility to the system, enhancing overall performance. Furthermore, due to the overhead from pilot tones, training sequences, and cyclic prefix, the programmable OFDM-based S-BVT intrinsically provides self-performance monitoring. The resulting acquisition of system (monitoring) parameters in the electrical domain (particularly at the network edge) enables channel estimation, equalization, and adaptive reconfigurability.

Based on these new paradigms, the programmable S-BVT (and node) plays an important role in the operation of next-generation optical networks, by supporting the on-demand configuration of programmable network functions, such as transmission rate, switching (slicing), bandwidth, etc. These functions will be decoupled from the hardware components (e.g. light sources, modulators, and filters) to provision the requested information rate.

## 7.3 S-BVT Architectures Involving Multiple Layers

The best way to judge the value of a technology is to consider how it will be used in the network and how it transforms the architecture of the network. This is the goal of this section. We will mainly focus on the main client layer for the DWDM layer—namely the IP layer. We look at how the IP layer is being designed today and how this design is constrained by the discrete nature of non-sliceable transponders. Then we consider how the additional flexibility offered by S-BVT will change the way the IP layer is designed and why this new architecture is more efficient—while still using the optical layer in a static manner to provide fixed connectivity between routers. We then consider how the optical network can be used in a more dynamic manner through multi-layer restoration and network re-optimization. While such dynamicity adds value without S-BVTs, it is further improved by use of S-BVTs. Once the network value is understood we focus on node level challenges, mostly around the interconnection of the S-BVT and the router, which must be flexible enough to enable the flexibility that S-BVTs provide. Finally we consider how S-BVTs impact the control plane between the layers.

### *7.3.1 IP Layer Architectures Without S-BVT*

The fundamental architecture of the IP layer over transport has not changed much over the last 20 years: routers connect to each other over a rigid wavelength plan using discrete transceivers with a fixed bit-rate per link. There are several variants of this approach, depending on the granularity of wavelengths used.

One approach is to use low granularity wavelengths (say, 10 G), in which case a large number of them may be used in parallel between routers using a load-balancing mechanism such as "link bundles." In this case, it makes sense to increase the connectivity of the IP topology using router bypass via ROADMs—as in Fig. 7.6a.

Another variant is to use high-speed wavelengths (such as 100 G today). In this case, the number of parallel links between routers is significantly reduced assuming the traffic cannot fill many such links. In fact, sometimes the traffic level on some links is so low that it does not make sense to keep them in the network, and the connectivity of the IP layer is reduced to ensure wavelengths are properly utilized as in Fig. 7.6b. A third variant is a hybrid of the above two approaches—in which some links use low bandwidth wavelengths while some use high-speed wavelengths—as in Fig. 7.6c.

Each of these options has its disadvantages: using low speed wavelengths for the network achieves low spectral efficiency, reduces the utilization of each connection due to inefficiencies of L2/L3 link bundling, and reduces the efficiency of IP protection due to the higher IP layer connectivity, as shown in [4]; migrating the network to higher speed wavelengths across the entire core footprint may still imply lower utilization of the IP links since there is a limit to allow reduction in the connectivity: the network topology must stay sufficiently connected to provide the right level of resiliency (note for example that Fig. 7.6b does not provide the redundancy needed

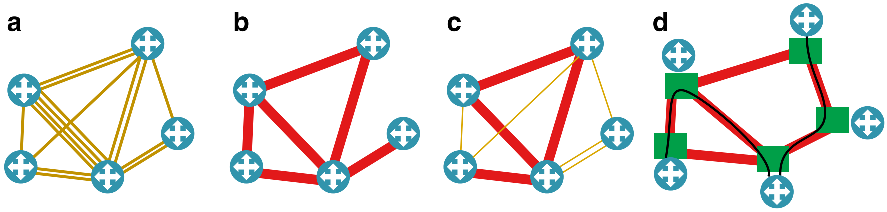



**Fig. 7.6** Fixed wavelength IP architectures. (**a**) High-density DWDM, (**b**) high-speed DWDM, (**c**) hybrid speed DWDM, (**d**) electrical time-based mux

in case the single link to the rightmost node fails); and the hybrid approach—while most cost effective—may be hard to operate due to the multiple generations of technology used (e.g. requiring guard bands for 10 G wavelengths).

A different approach is to use electrical transport layer multiplexing to provide sub-wavelength grooming over high-speed wavelengths, as shown in Fig. 7.6d. This architecture will tend to use a smaller number of high-speed wavelengths in order to utilize them well (same considerations as for Fig. 7.6b), and will rely on OTN or MPLS-TP to create sub-wavelength connections between routers that are not directly connected via a wavelength. These sub-wavelength connections are represented by the black lines in Fig. 7.6d. The cost of this solution will not necessarily be lower than the solutions provided by the optical wavelength multiplexing architecture due to the added electrical layer, the grey optics connecting the router to the switch, and—most importantly—the DWDM transceivers used for passing an OTN/MPLS-TP connection through each switch en-route to the destination. In essence, this approach substitutes router ports for MPLS-TP switch or OTN cross-connect ports, but does not save on DWDM transceivers. While this was economical in previous generation networks, when the cost of a router port was much higher than the cost of the DWDM transceiver, it will not make sense in most next-generation devices, in which the cost per bit of 100 G transceivers is projected to be similar to the cost of a 100 G router port.

In the above picture, transponders were assumed to have a fixed bit-rate. BVTs slightly change the picture, since they allow exploiting the trade-off between transmission reach and capacity. With BVTs, some of the links may have different capacities, but this difference does not stem from the need of the IP layer, but from the reach required in the optical layer. If the reach for certain links is very high, then the capacity of the wavelength could be low (say 50 G). Conversely, for short distances the capacity could be higher (say, 200 G). Either way, the transceiver is dedicated to the particular link, so there was no value by running a link at 200 G if the IP layer only requires 100 G for this link. One way to view such a network is by thinking of all transceivers as being fixed 200 G transceivers, but some running at lower than their maximum capacity to avoid regeneration in the middle. This view reflects the cost of the solution, since one has to pay for the maximum bit-rate even on links that do not require it.

Just like in the fixed transceiver case, the connectivity of the IP layer over BVTs is determined by the ratio between the capacity of the wavelength and the traffic needs over the wavelength: if the wavelengths have high capacity, it makes economic sense to have a sparser IP topology and if the wavelengths have lower capacity, it makes sense to have a denser IP network. Thus the IP topology is coupled to the capacity of wavelengths. This is an undesirable design constraint—the network can be more optimal if these two factors were to be independent of each other.

Note that the electrical transport multiplexing approach does allow the decoupling of the IP topology from the wavelength capacity: one can allocate as much capacity as needed to an IP link, irrespective of optical constraints, but this is done in the electrical domain and thus offers limited economic benefit as explained above. Can the same decoupling be achieved in the optical layer at a low cost? This will be discussed in the next section.

### *7.3.2 How Does S-BVT Enable a Better IP Layer Architecture?*

One of the main initial motivations for sliceable transponders is that they enable the desired decoupling of the IP layer density considerations from the capacity supported by the transceiver. This is because the S-BVT can be sliced based on capacity needs to create multiple independent links between routers. If routers need less than the maximum capacity offered by an S-BVT, only a portion of the S-BVT is allocated to this link, and the rest can be used by other links to other destinations.

Consider the example in the figure below. It shows three IP layer designs, each using two S-BVTs per node. These S-BVTs are shown as blue circles in Fig. 7.7a. All three designs are feasible for the same network and have the same cost, but they have different properties from an IP layer perspective. With S-BVTs, the IP designer can be freed from cost considerations when picking the optimal IP network for their needs.

However, the full potential of S-BVTs manifests itself in a dynamic network. In such a network, the IP topology can move from one topology to another topology (say, from Fig. 7.7a–c) in response to a change in traffic pattern or a failure. This is powerful since it allows for adaptation of the network to the needs of the traffic using a minimum amount of resources, instead of overprovisioning each of the links to support changes in the traffic. Clearly, such a change requires a multi-layer control plane—which will be discussed later.

It should be noted that such dynamic behaviour is possible in a network with a large number of low capacity connections—as in Fig. 7.6a, but with low spectral efficiency and high cost. As operators transition to the architecture of Fig. 7.6b—for example when transitioning to 100 G—they may lose the ability for dynamic changes in the network since the number of links in the topology is too low to allow for topology changes. S-BVTs enable this functionality without the drawbacks of a large number of low speed wavelengths, while still allowing the transition to high-capacity wavelengths.

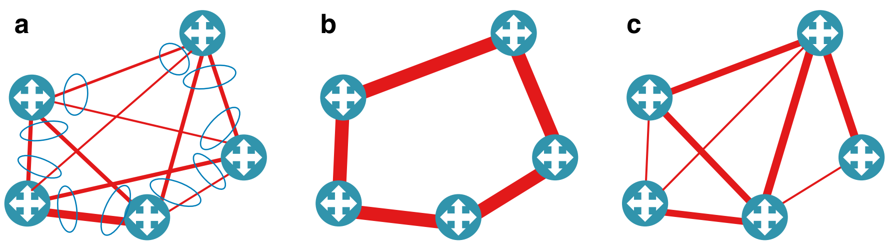



**Fig. 7.7** Different IP designs with the same two S-BVTs per node. (**a**) Dense IP design, (**b**) sparse IP design, (**c**) something in between

### *7.3.3 How to Interconnect Routers and S-BVTs*

So far we focused on the savings in the optical layer—in particular in the number of S-BVTs. We did not consider how the clients are connected to the optical layer. If we follow the same approach that is used today with fixed transponders, the client should be connected to the transponder using a single fixed capacity point-to-point link. As a result, flexible slicing of a transponder to different destinations will not be useful, since the client interface into the optical layer lacks this flexibility. Therefore, the interconnection between the layers should be done in a different fashion.

A few interconnect options are shown in Fig. 7.8. Options (a) and (b) are based on a number of fixed rate interfaces—either at a fine granularity (say 10 GE) or coarser granularity (say 100 GE). The S-BVT can map a varying number of these interfaces into a single lightpath to achieve the desired flexibility. This is done through TDM multiplexing (typically OTN). From a client perspective, these options look like a set of fixed links connecting client nodes (same as options (a) and (b) in Fig. 7.6). While the finer granularity option (a) provides for more flexibility in the IP layer topology, it suffers from a large number of patch chords connecting the boxes. This introduces operational complexity and—at some point—affects the density of both S-BVTs and the client platform due to faceplate limitations. Option (b) does not suffer from this operational complexity but may not be flexible enough to support the desired IP connectivity—especially if the overall traffic in the network is relatively low.

A better solution is shown in Fig. 7.8c and is based on a channelized interfaces between the systems. This allows for a single physical interface that is logically divided into multiple channels. This also enables more flexibility in granularity—assuming that the technology used for channelizing the interface is flexible enough. We will consider a few channelization options that exist or are being suggested.

The most common approach to channelizing transport links in TDM multiplexing, and the current standard for this technology is OTN, as defined in the ITU G.709 specification. A typical example for OTN multiplexing would be the mapping of a 10 G Ethernet signal into the payload of an ODU2. Ten such ODU2s can then be mapped into an ODU4, which can be carried in a 100 G wavelength (in an

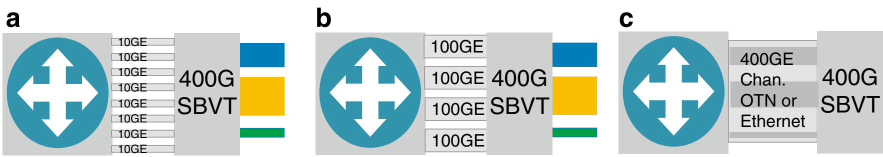



**Fig. 7.8** Client-S-BVT direct connectivity options. (**a**) Discrete low-speed connectivity, (**b**) discrete high-speed connectivity, (**c**) channelized connectivity

OTU4 container). ITU will also be standardizing 400 G wavelengths which will carry up to 40 such ODU2 payloads. If the lightpath requires some other bit-rate (neither 100 G nor 400 G), then there is a need for a flexible physical mapping, and a proposal of this nature—called OTUflex—is being discussed. This hierarchy is used as follows: suppose router A needs to establish a 70 G link to another router B and a 80 Gb/s link to router C. Router A will aggregate logically seven 10 GE interfaces. Those interfaces will be multiplexed into an ODU4 container and onto a 100 G wavelength (assuming no OTUflex) from A to B. Another eight 10 GE interfaces will be multiplexed into an ODU4 to connect A to C.

Another approach is based on packet technologies: a special tag is inserted to the header of the packet from the router towards the S-BVT, which indicates the channel. Typically this is done using an extra VLAN field in the header of the packet as per IEEE standard 802.1ad. The S-BVT takes all packets with the same VLAN tag and maps them into an ODU container and then the ODU into an OTU which is in turn mapped onto a lightpath. While this approach seems simpler that TDM multiplexing, it still requires a mechanism on the router to ensure that packets with a particular VLAN do not exceed the capacity of the particular transport container into which they are mapped. This is typically achieved via a special function in the hardware of router line card, which drops packets when exceeding the capacity of the channel, and is considered an expensive option compared to simple router interfaces that do not support this. An added complexity is the need to perform some minimal packet processing in the S-BVT—which is not needed in the first option.

An emerging variant of this solution is based on channelizing the traffic by exploiting the overheads in the Ethernet frame instead of adding a VLAN tag. The MAC and PCS levels are lower in terms of packet processing and therefore require less resources. This solution is referred to as “flex MAC.”

To complete this section, we note that all the above solutions still suffer from a fundamental problem: when the reach of lightpaths is long and the modulation format used is a lower order format such as BPSK to enable more robust transmission over a longer un-regenerated reach, the capacity of the S-BVT drops. If the solution must be flexible to automatically support shorter reaches as well as longer reaches, the client must be connected with the maximum supported capacity to the S-BVT. As a result, the client capacity connected to the S-BVT may be under-utilized if the S-BVT does not operate at maximum capacity. This is shown in Fig. 7.9a vs. Fig. 7.9b. A possible solution to this problem is depicted in Fig. 7.9c, in which a photonic switch is used to flexibly assign client capacity to S-BVTs without wasting their capacity. Such a photonic switch is likely to be a low cost and low power solution; however, it implies that the interface from the client is broken up into several separate optical links (since the switch can only switch an entire such link). This implies a non-channelized solution (as in Fig. 7.8a or b). This also implies use of single mode optics (optical switches do not support multi-mode optics today)—which today implies higher cost. Until the cost of single mode optics drops significantly, such an optical switch solution may be hard to justify.

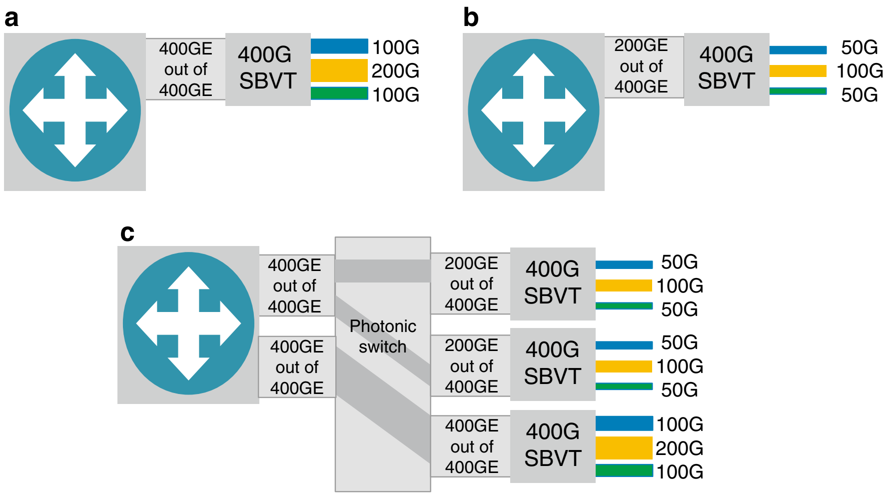



**Fig. 7.9** Client-S-BVT indirect connectivity. (**a**) Well utilized client-S-BVT interface, (**b**) poorly utilized client-S-BVT interface, (**c**) switched connectivity to improve utilization

## 7.4 Network Planning Process Using S-BVTs

Network problems appear in high variety of situations becoming in more and more frequent issue. These networks are widely used in diverse areas but which take importance here are communication networks (nodes, links, paths, traffics, etc.). This section presents a network planning process in an EON using S-BVTs. A Mixed Integer Linear Programming (MILP) is proposed to obtain the minimum number of transponders in a network.

During this section, we will talk about sliceable and non-sliceable transponders. The non-sliceable term is used to mention all transponders that are not S-BVTs, thus is fix rate transponders or BVTs.

### *7.4.1 Network Planning Example*

As stated before, S-BVTs enable transmitting from one point to multiple destinations, changing the traffic rate to each destination and the number of destinations on demand. The number of destination depends on the granularity, for instance a 400 Gb/s S-BVT with 40 Gb/s carriers could transmit up to ten different destinations, while for a granularity of 100 Gb/s this limit is four destinations. On the other hand, the number of non-sliceable transponders depends on the bit-rate and the strategy for dimension. With the intention of understanding better the planning process using S-BVTs, Fig. 7.10 shows the number of transponders required in two different situations: using non-sliceable transponders and using S-BVTs.

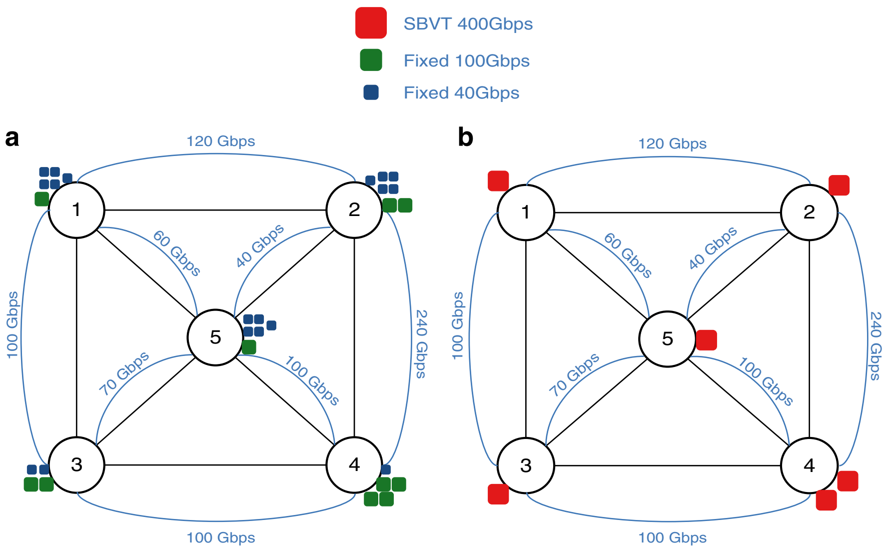



**Fig. 7.10** Example of transponders required in a network to supply the demanded traffic. (**a**) With non-sliceable transponders, (**b**) with S-BVTs

Non-sliceable transponders are fixed or BVT, which does not have sliceable capabilities of S-BVTs. Let us assume that the objective of the planning process is to minimize the capacity deployed in the network.

For example, in node 1 in case (a) six transponders are needed to supply all demands, concretely 3 × 40 Gb/s to provide 1–2 traffic, 1 × 100 Gb/s to provide 1–3 traffic, and 2 × 40 Gb/s to supply 1–5 traffic. On the other hand in case (b), it would be enough just with 1 × 400 Gb/s S-BVT with 40 Gpbs of granularity in node 1 to supply the same traffic, because the S-BVT advantage of changing the traffic (one carrier for each destination). Node 4 requires two S-BVTs because its demands are 440 Gb/s. In case of all demands from/for a node are 400 Gb/s if more than ten carriers were required two or more S-BVTs would be needed since the number of carriers in each S-BVT is ten. Thanks to this example, it is clear that the dimensioning process using S-BVTs is different than with non-sliceable transponders.

### *7.4.2 Mixed Integer Linear Programming Definition*

Once an example is presented, this section presents an optimization model in order to minimize the number of transponders required to comply with the requirements of the user's traffic. Optimal configuration is a classical problem in engineering field which tries to minimize or maximize a certain cost function at the same time that satisfies the constraints of the problem.

The definition of the MILP problem is presented below following a procedure based on node-link formulation. The differences between the next scenarios lie in the kind of transponders, as it was presented in the previous example.

Given a network denoted by $N(G, D)$ where $G = G(V, E)$ is the network graph and $D$ is a set of demands between the pairs of node of the graph. Graph $G$ is composed of the set of nodes $V$ and the set of links $E$. For ease of exposition, we usually assume that the graph does not contain loops nor parallel links, that is $E \subseteq V^{|2|} \setminus \{(v, v) : v \in V\}$ where $V^{|2|}$ is the set of all two-element subsets of the set $V$ in case of directed graph. The endpoints of a link $e \in E$ are denoted by $a(e)$ and $b(e)$. In the directed case, if $e = (v, w)$ $a(e) = v$ and $b(e) = w$.

In a directed graph the demands in set $D \subseteq V^{|2|}$ have to be also directed. It is important to point out that our demands are undirected but we have considered them directed and we have counted them two times due to both directions of a demands being routed throughout the same path (the shortest one). Therefore the endpoints of a demand are denoted by $s(d)$ and $t(d)$. In the directed case, if $d = (x, y)$ $s(d) = x$ and $t(d) = y$. The demand value is given by $h_d$.

In this problem, we also have a set of paths $P$ containing the shortest path precomputed for each demand. The elements of this set are defined as a subset of edges, $p \in P \mid p \subset E$.

For the S-BVT problem, we have just one kind of transponder with capacity $C$ and fixed granularity $G$. For the traditional non-sliceable transponder problem, we have a set $K$ that contains $k$ kinds of transponder with different capacity. In the case of BVTs, the dimensioning is done with the maximum capacity of the transponder.

(1) S-BVT problem statement.

**Set and Parameters**

$$e \in E e = (v, w) \begin{cases} a(e) = v \\ b(e) = w \end{cases} \quad \text{Link end nodes}$$

$$d \in D d = (x, y) \begin{cases} s(d) = x \\ t(d) = y \end{cases} \quad \text{Demand end nodes}$$

$p \in P \mid p \subset E \quad \forall d \exists! p \mid p = sp(d)$ shortest path of demand $d$

$h_d$ Demand value

$C$ S-BVTs capacity

$G$ S-BVTs granularity

**Decision Variables**

$T_u \quad u \in V$ Number of S BVTs per node (Integer)

$$Z_{de} = \begin{cases} 1 \; if \;\; e \in p \text{ being } p = sp(d) d \in D \\ \quad 0 \text{ otherwise} \end{cases} \quad \text{Demand link}$$

$S_e \quad e \in E$ Total used slices per link (Integer)

$S_u \quad u \in V$ Total used slices per node (Integer)

**Objective Function**

$$\text{Minimize}\left(\sum_{u \in V} T_u\right)$$

**Constraints**

$$\forall\ d \in D \quad \sum_{\substack{e \in E \\ a(e)=x}} Z_{de} - \sum_{\substack{e \in E \\ b(e)=x}} Z_{de} = 1$$

$$\forall\ d \in D \quad \sum_{\substack{e \in E \\ b(e)=y}} Z_{de} - \sum_{\substack{e \in E \\ a(e)=y}} Z_{de} = 1$$

$$\forall\ d \in D,\ \forall\ u \in V \mid u \neq x,\ u \neq y \quad \sum_{\substack{e \in E \\ a(e)=u}} Z_{de} - \sum_{\substack{e \in E \\ b(e)=u}} Z_{de} = 0$$

$$\forall\ e \in E \quad \sum_{\substack{d\ e\ D \\ a(e)=v \\ b(e)=w}} \left(h_d * Z_{de}\right) + \sum_{\substack{d\ e\ D \\ b(e)=v \\ a(e)=w}} \left(h_d * Z_{de}\right) \leq G * S_e$$

$$\forall\ u \in V \quad \sum_{\substack{e \in E \\ a(e)=u}} S_e \leq S_u$$

$$\forall\ u \in V \quad G * S_u \leq C * T_u$$

(2) Non-sliceable transponder problem statement

**Set and Parameters**

$$e \in E e = (v,\ w) \begin{cases} a(e) = v \\ b(e) = w \end{cases} \quad \text{Link end nodes}$$

$$d \in D\ \ d = (x,\ y) \begin{cases} s(d) = x \\ t(d) = y \end{cases} \quad \text{Demand end nodes}$$

$$p \in P \left| p \subset E \quad \forall d \exists ! p \right| p = sp(d) \text{ shortest path of demand } d$$

$h_d$ $d \in D$ Demand value

$C_k$ $k \in K$ Capacity of each kind of fixed transponder

**Decision Variables**

$T_{ke}$ $e \in E$ $k \in K$ Number of each kind of fixed transponder per link (Integer)

$T_{ku}$ $u \in V$ $k \in K$ Number of each kind of fixed transponder per node (Integer)

$$Z_{de} = \begin{cases} 1 \text{ if } e \in p \text{ being } p = sp(d) d \in D \\ 0 \text{ otherwise} \end{cases} \quad \text{Demand link}$$

**Objective Function**

$$\text{minimize}\left(\sum_{u \in V}\sum_{k \in K} C_k * T_{ku}\right)$$

**Constraints**

$$\forall\ d \in D \quad \sum_{\substack{e \in E \\ a(e)=x}} Z_{de} - \sum_{\substack{e \in E \\ b(e)=x}} Z_{de} = 1$$

$$\forall\ d \in D \quad \sum_{\substack{e \in E \\ b(e)=y}} Z_{de} - \sum_{\substack{e \in E \\ a(e)=y}} Z_{de} = 1$$

$$\forall\ d \in D,\ \forall\ u \in V \mid u \neq x,\ u \neq y \quad \sum_{\substack{e \in E \\ a(e)=u}} Z_{de} - \sum_{\substack{e \in E \\ b(e)=u}} Z_{de} = 0$$

$$\forall\ e \in E \quad \sum_{\substack{d\ e\ D \\ a(e)=v \\ b(e)=w}} \left(h_d * Z_{de}\right) + \sum_{\substack{d\ e\ D \\ b(e)=v \\ a(e)=w}} \left(h_d * Z_{de}\right) \leq \sum_{k \in K} C_k * T_{ke}$$

$$\forall\ u \in V,\ \forall\ k \in K \quad \sum_{\substack{e \in E \\ a(e)=u}} T_{ke} \leq T_{ku}$$

There are other targets for optimization (minimize the number of wavelengths or the number of hops), but based on previous definition the problem can be re-defined to cope with the requirements of any other optimization problem.

## 7.5 Expected Savings from S-BVTs

Previous MILP is defined with the objective of minimizing the number of transponders in the network. This section contains a use case to present the expected savings of using S-BVTs instead of non-sliceable transponders in the network.

### *7.5.1 Scenario Definition*

In order to quantify the benefits of optimizing the topology, two scenarios are compared: (1) current network topology and (2) optimized topology. Current network topology is non-optimal scenario, which is based on the IP demands of a realistic Spanish IP core network. The traffic is routed throughout the network using the classical shortest path in terms of number of hops at the IP layer. The optimal scenario considers an optimized topology, in which the optimizer can also change the route followed by the traffic. Besides of comparing the number of transponders for these scenarios, this study is also focused in granularity effects. Therefore, the comparison will be done for different S-BVTs granularities.

The problem defined above has been applied to the Spanish backbone represented in Fig. 7.11. This network is made up of 20 edge nodes that aggregate traffic from transit and access routers and forwards it over wavelength channels requested to the optical network, which is mesh.

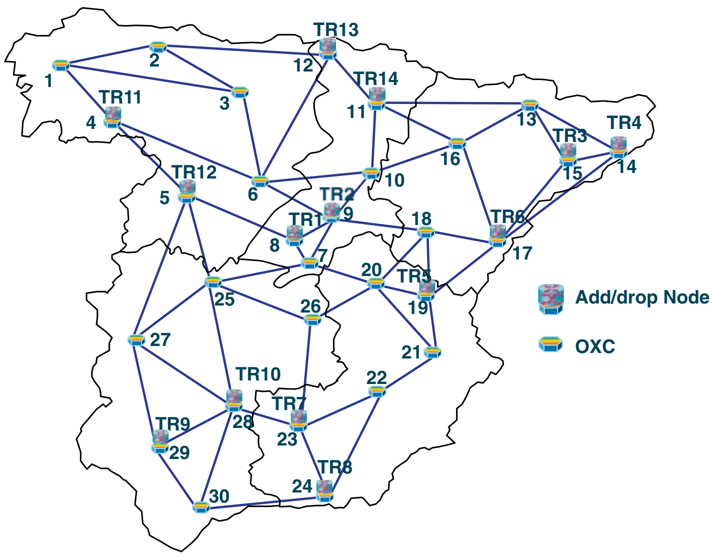



**Fig. 7.11** Reference network based on Spanish national backbone (reproduced from [21])

The initial traffic matrix is created based on information for the Telefonica backbone network in 2013. The link dimensioning is done using an over-dimensioning factor of 30 %. Traffic is incremented yearly by a 50 % factor in order to compare the results for the different kind of transponders proposed over the next 5 years (from 2013 to 2018) in a brownfield scenario.

### 7.5.2 *Impact on the Number of Transponders*

The results have been obtained for the two previously mentioned topologies (IP and optimized) for different values of the granularity, 10, 40, and 100 Gb/s. The S-BVT selected capacity is set to 400 Gb/s. In the case of granularity is equals to 40 Gb/s each transponder has ten carriers ($10 \times 40$ Gb/s). This allows to face up more low demands than with 100 Gb/s granularity. The reason is that there will be just four carriers per transponder as the total capacity is 400 Gb/s. A priori, it is natural to think that the lower the granularity is the less transponders are required. Below, we will see how the topology optimization affects to these initial beliefs. Table 7.2 represents the number of transponders required in each case 10, 40, and 100 Gb/s of granularity and for each topology.

Focusing on the results for the IP topology, it is possible to observe that much more transponders are required when granularity is higher. But the really important thing is the number of transponders is practically the same in case of the optimized topology. There is just a low increment is observed when the granularity goes from 10 to 100 Gb/s.

**Table 7.2** Number of S-BVTs per year for different topologies and granularities

| | Granularity = 10 Gb/s | | Granularity = 40 Gb/s | | Granularity = 100 Gb/s | |
|---|---|---|---|---|---|---|
| Year | IP | Opt | IP | Opt | IP | Opt |
| 2013 | 24 | 22 | 24 | 22 | 38 | 24 |
| 2014 | 32 | 32 | 34 | 32 | 44 | 33 |
| 2015 | 40 | 40 | 44 | 40 | 56 | 44 |
| 2016 | 62 | 60 | 62 | 60 | 68 | 62 |
| 2017 | 84 | 80 | 88 | 80 | 90 | 80 |
| 2018 | 120 | 114 | 120 | 115 | 130 | 118 |

Besides of advantages that topology optimization represents, the optimized topology also has another advantage over the IP topology. Looking again at their results, it is possible to observe that in any case, the optimized topology always needs less number of transponders.

### *7.5.3 Impact on IP Layer Networks*

Previous section in this chapter has analyzed different S-BVT architectures involving multiple layers. Due to the existence of different S-BVT layers, there are some requirements to interface with the upper layers and provide flexibility. For instance, the interface between the client equipment and S-BVT must also be flexible to allow S-BVT to take different aggregate traffic flows and send them in different directions, without requiring manual reconfiguration of the connectivity between the systems. There are several approaches to achieving such flexibility.

This section considers two architectures in order to evaluate the impact on IP layer. The first architecture uses an Ethernet switch or an OTN cross-connect between the router and the S-BVT. The switch can map VLANs in the S-BVT carriers and similarly OTN cross-connect can demultiplex a single router link into one or multiple carriers. Note that this approach adds a separate switching layer to the solution, which should not be needed given the router can perform the traffic switching function. This architecture is shown in Fig. 7.12a. This switching layer is considered later in the cost evaluation.

The second option is to use multiple 100 GE links between the router and S-BVT. Assuming the S-BVT granularity is in multiples of 100 GE, this approach can use existing router gear and allow S-BVT to map *N* 100 GE links into a single payload for transmission, e.g. a higher order ODU container. For example, three 100 GE links can be multiplexed together into a 300 G ODUflex frame and shipped in a 300 G OTUflex container (Fig. 7.12b). This approach also allows the router to exploit more flexible Ethernet frames based on MACflex concepts (an Ethernet frame that allows for any line rate instead of the fixed 40, 100, 400 GE etc.), instead of fixed 100 GE as in case b.

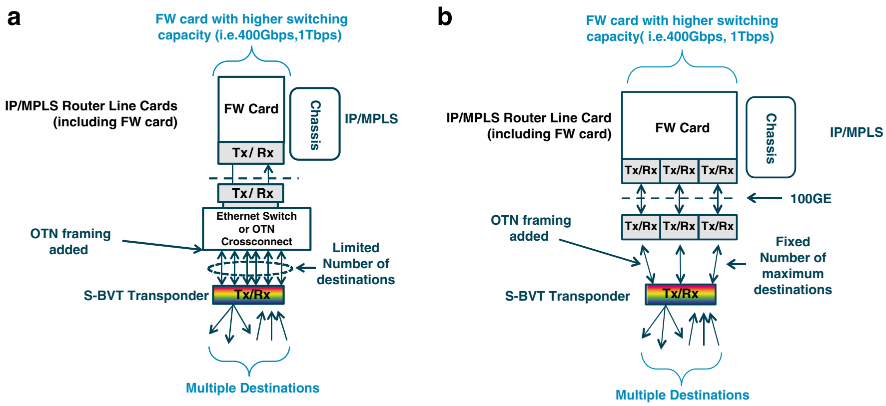



**Fig. 7.12** Architectural options for interconnecting routers and S-BVTs. (**a**) Using an Ethernet switch or an OTN cross-connect between the router and S-BVT, (**b**) using multiple 100 GE links between the router and S-BVT

**Table 7.3** IP/MPLS router line cards cost

| Line cards (Gb/s) | Cost | Year | Cost per Gb/s |
|---|---|---|---|
| 10 × 40 | 31.98 | 2014 | 0.07995 |
| 25 × 40 | 31.98 | 2018 | 0.03198 |
| 4 × 100 | 36 | 2014 | 0.09 |
| 10 × 100 | 36 | 2018 | 0.036 |
| 1 × 400 | 34,28 | 2014 | 0.0857 |
| 2 × 400 | 34,28 | 2018 | 0.04285 |
| 1 × 1000 | 36,42 | 2018 | 0.03642 |

According to the S-BVT architectures previously presented, the interface with the IP router is different. When there are non-sliceable transponders, the routers are connected to each transponder using directly to the transponders chassis (similar to Fig. 7.12b). When there is a switch (Fig. 7.12a), the router can use a card with the same capacity than the S-BVT, because the architecture allows the mapping between the IP/MPLS traffic and the carriers. However, it is clear that the router is limited to use fixed interfaces with the second configuration (Fig. 7.12b) or with non-sliceable transponders. Consequently, we assume in this section the architecture in Fig. 7.12a with an Ethernet switch, because it allows more flexibility to the node. The reason to choose Ethernet instead of OTN is because the demultiplexing of OTN channels is just proposed by some vendors, but there is not a real implementation to the best of the authors' knowledge.

Once the interface with the IP/MPLS layer is defined, we analyze the impact in the IP layer. For this section, the number of cards is computed for each scenario with non-sliceable transponders and S-BVT. Table 7.3 presents the values for the IP/MPLS line cards, based on the cost model presented on [19].

Figures 7.13 and 7.14 show the cost savings of IP/MPLS line cards by using S-BVT versus non-sliceable transponders for the two previously topologies: IP topology and optimized topology. Moreover, a topology connecting all routers is added to the results. This topology is the full mesh topology. Figure 7.13 shows the results for 400 Gb/s and a granularity of 40 Gb/s, while Fig. 7.14 does for 1 Tb/s with 40 Gb/s granularity. Figure 7.14 only represents the results just for the last 4 years, because the 1 Tb/s technology will not available until 2018 [19]. In light of the results, it can be seen that savings can reach up to 40 % in case of 400 Gb/s and slightly more in case of 1 Tb/s on IP/MPLS router line cards (43 %). For a granularity of 100 Gb/s, the maximum savings are 50 and 48 % for 400 Gb/s and 1 Tb/s. This potential savings using this technology are very important, thus motivating further research on this direction.

### 7.5.4 *Total CAPEX Savings of S-BVTs*

Once we have computed the investment reduction in IP/MPLS cards, the total savings are calculated considering the investment in the IP and the optical layer. Figures 7.15 and 7.16 present percentage of total savings in the IP and optical layers. To allow the flexibility in the S-BVT, an Ethernet switch is added in the architecture. To obtain the cost of the Ethernet ports, we have obtained the relation of an Ethernet port and an IP card in [20] and we have calculated its value in the relative cost units.

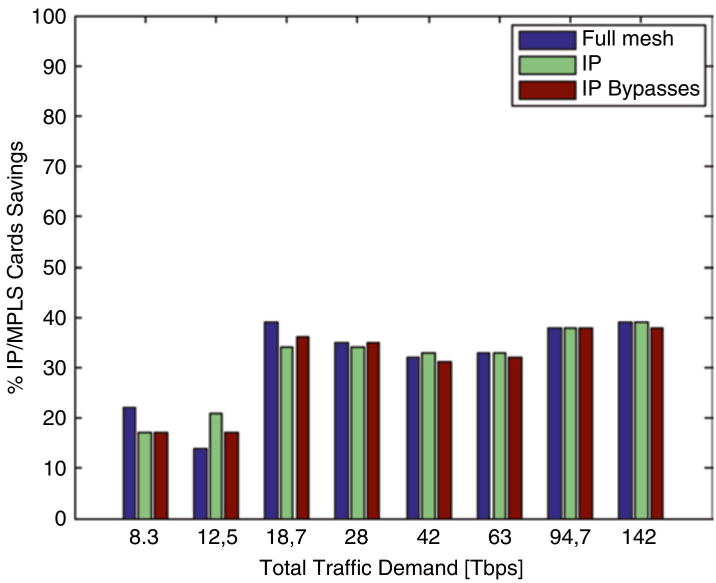



**Fig. 7.13** Percentage of line card cost saved by using S-BVTs 400 Gb/s versus non-sliceable

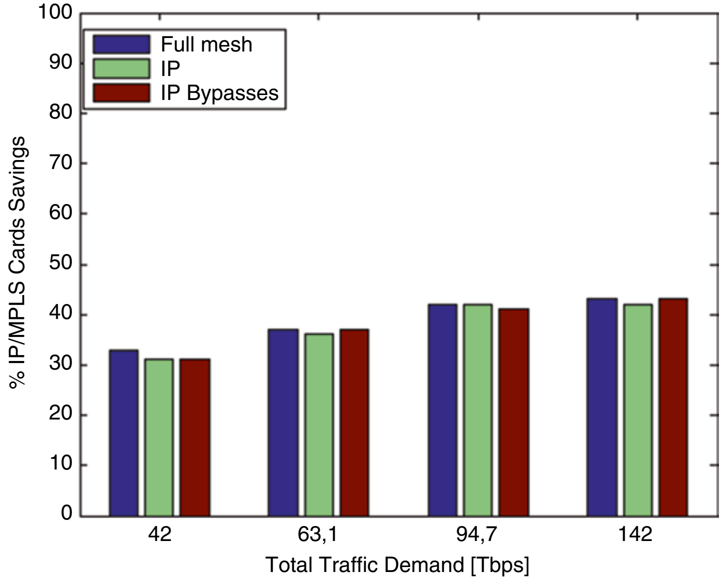



**Fig. 7.14** Percentage of line cards cost saved by using S-BVTs 1 Tb/s versus non-sliceable transponders

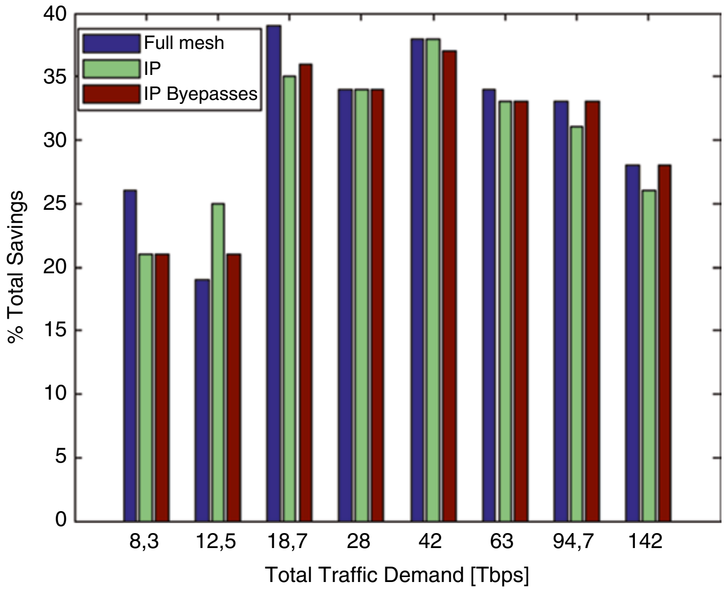



**Fig. 7.15** Percentage of total savings by using S-BVTs 400 Gb/s versus non-sliceable transponders

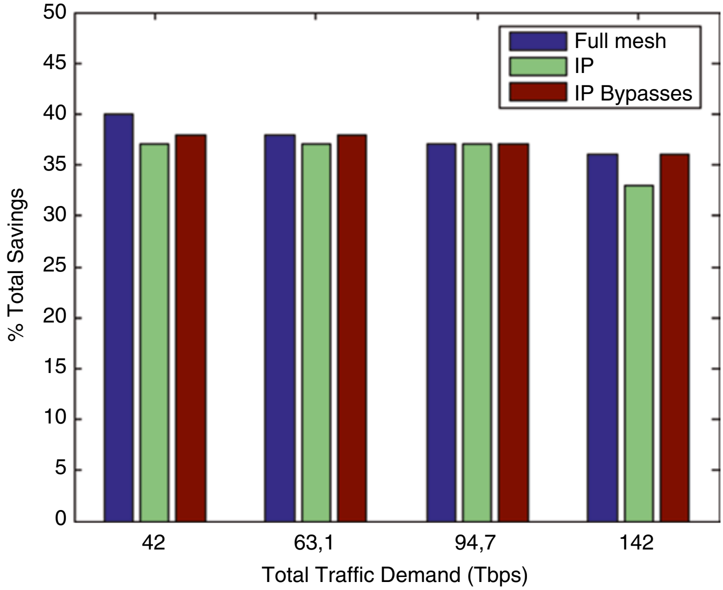



**Fig. 7.16** Percentage of total savings by using S-BVTs 1 Tb/s versus non-sliceable transponders

In light of the results, we can claim that the addition of S-BVTs can achieve around 30 % savings for the investment in the IP/MPLS cards and optical transponders, when using 400 Gb/s S-BVTs. When the capacity of the S-BVTs is 1 Tb/s, this savings are even higher reaching 35 %.

## 7.6 Conclusions

The Sliceable Bandwidth Variable Transponder (S-BVT) is proposed in real implementation and it is a new way to do optical grooming in EON. S-BVTs enable transmitting from one point to multiple destinations, changing the traffic rate to each destination and the number of destinations on demand. It can provide even higher levels of elasticity and efficiency to the network. This chapter presents the architecture of the transponder itself, as well as how it can be interconnected to the routers.

Once realistic architectures are presented, the chapter complements with planning examples and a detailed analysis of the potential savings in a network operator scenario. The analysis demonstrates that there are not only savings in the number of transponders, but also in the number of IP/MPLS cards.

## References


1. M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, T. Yoshimatsu, T. Kobayashi, Y. Miyamoto, K. Yonenaga, A. Takada, O. Ishida, and S. Matsuoka, *Demonstration of novel spectrum-efficient elastic optical path network with per channel variable capacity of 40 Gb/s to over 400 Gb/s*, in Proceedings of European Conference on Optical Communication, 2008
2. O. Gerstel, M. Jinno, A. Lord, S.J. Ben Yoo, Elastic optical networking: a new dawn for the optical layer? IEEE Commun. Mag. **50**, s12–s20 (2012)
3. M. Jinno, H. Takara, Y. Sone, K. Yonenaga, A. Hirano, Multiflow optical transponder for efficient multilayer optical networking. IEEE Commun. Mag. **50**(5), 56–65 (2012)
4. V. López, B. de la Cruz, O. González de Dios, O. Gerstel, N. Amaya, G. Zervas, D. Simeonidou, J.P. Fernandez-Palacios, Finding the target cost for sliceable bandwidth variable transponders. IEEE/OSA J. Opt. Commun. Networking **6**, 476–485 (2014)
5. L. Velasco, Ó. González de Dios, V. López, J.P. Fernández-Palacios, G. Junyent, *Finding an objective cost for sliceable flexgrid transponders*, in Proceedings of Optical Fiber Conference (OFC), San Francisco, 2014
6. B. de la Cruz Miranda, Ó. González de Dios, V. López, J.P. Fernández-Palacios, *OpEx savings by reduction of stock of spare parts with sliceable bandwidth variable transponders*, in Proceedings of Optical Fiber Conference (OFC), San Francisco, 2014
7. M. Svaluto, J.M. Fàbrega, L. Nadal, F.J. Vílchez, V. López, J.P. Fernández-Palacios, *Cost-effective data plane solutions based on OFDM technology for flexi-grid metro networks using sliceable bandwidth variable transponders*, in Proceedings of Optical Networking Design and Modeling (ONDM), Stockholm, May 2014
8. N. Sambo et al., Programmable transponder, code and differentiated filter configuration in elastic optical networks. J. Lightwave Technol. **32**, 2079–2086 (2014)
9. *Draft revised G.694.1 version 1.3*, Unpublished ITU-T Study Group 15, Question 6
10. P. Anandarajah, R. Zhou, R. Maher, M. Gutierrez Pascual, F. Smyth, V. Vujicic, L. Barry, *Flexible optical comb source for super channel systems*, in Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013, Anaheim, 2013
11. V. Ataie, B.-P. Kuo, E. Myslivets, S. Radic, *Generation of 1500-tone, 120 nm-wide ultraflat frequency comb by single CW source*, in Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013, Anaheim, 2013
12. A. Metcalf, V. Torres-Company, D. Leaird, A. Weiner, *Stand-alone high-power broadly tunable optoelectronic frequency comb generator*, in Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013, Anaheim, 2013
13. A. Mishra, R. Schmogrow, I. Tomkos, D. Hillerkuss, C. Koos, W. Freude, J. Leuthold, Flexible RF-based comb generator. IEEE Photon. Technol. Lett. **25**(7), 701–704 (2013)
14. E. Marchena, T. Creazzo, S. Krasulick, P. Yu, D. Van Orden, J. Spann, C. Blivin, J. Dallesasse, P. Varangis, R. Stone, A. Mizrahi, *Integrated tunable CMOS laser for Si photonics*, in Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013, Anaheim, 2013
15. N. Sambo et al., Sliceable transponder architecture including multi-wavelength source. IEEE/OSA J. Opt. Commun. Networking **6**, 590–600 (2014)
16. IEEE. *IEEE P802.3ba 40Gb/s and 100Gb/s Ethernet task force*, official web site, June 2010
17. D. Rafique, T. Rahman, A. Napoli, S. Calabro, B. Spinnler, Technology options for 400Gb/s PM-16QAM flex-grid network upgrades. IEEE Photon. Technol. Lett. **26**(8), 773–776 (2014)
18. D. Guckenberger, et al., *Advantages of CMOS photonics for future transceiver applications*, European Conference and Exhibition on Optical Communication (ECOC) 2010, Torino, Paper TU.4.C.2., 2010

19. F. Rambach, B. Konrad, L. Dembeck, U. Gebhard, M. Gunkel, M. Quagliotti, V. Lopez, A multilayer cost model for metro/core networks. IEEE/OSA J. Opt. Commun. Networking **5**, 210–225 (2013)
20. R. Huelsermann, M. Gunkel, C. Meusburger, D. Schupke, Cost modeling and evaluation of capital expenditures in optical multilayer networks. J. Opt. Commun. Networking **7**, 814–883 (2008)
21. FP7-STRONGEST project, *Efficient and optimized network architecture: requirements and reference scenarios,* http://www.ict-strongest.eu
22. G. Colavolpe, T. Foggi, *High spectral efficiency for long-haul optical links: time-frequency packing vs. high-order constellations*, in Proceedings of European Conference on Optical Communication (ECOC), 2013
23. L. Potì, G. Meloni, G. Berrettini, F. Fresi, M. Secondini, T. Foggi, G. Colavolpe, E. Forestieri, A. D'Errico, F. Cavaliere, R. Sabella, G. Prati, *Casting 1 Tb/s DP-QPSK communication into 200 GHz bandwidth*, in Proceedings of European Conference on Optical Communications (ECOC), Amsterdam, 2012

# Chapter 8
# GMPLS Control Plane

**Oscar González de Dios, Ramon Casellas, and Francesco Paolucci**

## Acronyms

| | |
|---|---|
| AFI | Address Family Identifier |
| API | Application Programming Interface |
| BGP | Border Gateway Protocol |
| BGP-LS | BGP-Link State |
| BPSK | Binary Phase Shift Keying |
| BVT | Bandwidth Variable Transponder |
| CCAMP | Common Control and Measurement Plane |
| cPCE | child Path Computation Element |
| CRUD | Create, Read, Update and Delete |
| DWDM | Dense Wavelength Division Multiplexing |
| E2E | End to End |
| ELC | Explicit Label Control |
| EON | Elastic Optical Network |
| ERO | Explicit Route Object |
| FA | Forwarding Adjacency |
| FA-LSP | Forwarding Adjacency LSP |
| FEC | Forward Error Correction |
| GMPLS | Generalized Multi-Protocol Label Switching |
| H-LSP | Hierarchical Label Switched Path |

O.G. de Dios (✉)
Telefónica, Madrid, Spain
e-mail: oscar.gonzalezdedios@telefonica.com

R. Casellas
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain

F. Paolucci
Scuola Superiore Sant'Anna, TeCIP, Pisa, Italy



V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_8

| | |
|---|---|
| H-PCE | Hierarchical Path Computation Element |
| IGP | Interior Gateway Protocol |
| IP | Internet Protocol |
| IPCC | IP Control Channel |
| IRO | Include Route Object |
| ISCD | Interface Switching Capability Descriptor |
| IS-IS | Intermediate System to Intermediate System |
| IS-IS-TE | IS-IS with Traffic Engineering |
| IV | Impairment Validation |
| KA | KeepAlive |
| LMP | Link Management Protocol |
| LS | Label Set |
| LSA | Link State Advertisement |
| LSC | Lambda-Switch-Capable |
| LSP | Label Switched Path |
| LSPDB | LSP-DataBase |
| LSR | Label Switched Router |
| MIB | Management Information Base |
| MPLS | Multiprotocol Label Switching |
| MPLS-TE | Multiprotocol Label Switching—Traffic Engineering |
| NCF | Nominal Central Frequency |
| NLRI | Network Layer Reachability Information |
| NMS | Network Management System |
| OFDM | Orthogonal Frequency Division Multiplexing |
| OSPF | Open Shortest Path First |
| OSPF-TE | Open Shortest Path First—Traffic Engineering |
| PCC | Path Computation Client |
| PCE | Path Computation Element |
| PCEP | Path Computation Element Protocol |
| PCInitiate | LSP Instantiation Request |
| PCReq | Path Computation Request |
| PCRpt | Path Computation State Report |
| PCUpd | Path Computation Update Request |
| pPCE | parent PCE |
| QAM | Quadrature Amplitude Modulation |
| QoS | Quality of Service |
| QoT | Quality of Transmission |
| QPSK | Quadrature Phase Shift Keying |
| RC | Routing Controller |
| REST | Representational State Transfer |
| RMSA | Routing, Modulation and Spectrum Assignment |
| RRO | Record Route Object |
| RSA | Routing and Spectrum Allocation |
| RSVP-TE | Resource Reservation protocol—Traffic Engineering |
| RWA | Routing and Wavelength Assignment |

| | |
|---|---|
| SAFI | Subsequent Address Family Numbers |
| SCC | Spectrum Continuity Constraint |
| SCSI | Switching Capability Specific Information |
| SDN | Software Defined Networking |
| SL | Suggested Label |
| SNMP | Simple Network Management Protocol |
| SRLG | Shared Risk Link Group |
| SSON | Spectrum Switched Optical Network |
| SVEC | Synchronization VECtor |
| TCP | Transport Control Protocol |
| TDM | Time-division multiplexing |
| TE | Traffic Engineering |
| TED | Traffic Engineering Database |
| TLV | Type, Length, Value |
| UL | Upstream Label |
| WSON | Wavelength Switched Optical Network |
| XRO | Exclude Route Object |

The control plane is the key element to utilize the full potential of Elastic Optical Networks (EONs). In this chapter, it is explained what are the functionalities of the control plane of a network, and the most popular control plane architecture, Generalized Multiprotocol Label Switching (GMPLS), is revisited. After the basics of routing, path computation, and signalling protocols are reviewed, the chapter explains how to architect a control plane to work in EONs. Moreover, the necessary extensions to the main GMPLS protocols to support EONs, including handling Bandwidth Variable Transponders (BVTs), are detailed. Finally, how to control a network with multiple domains is explained, focusing on the architecture and protocol improvements.

## 8.1 Introduction to Control Plane Functionalities

Throughout the book different kinds of data plane technologies, which are the base of EONs, have been described in detail; to name a few examples we can cite switching devices, modulators, or BVTs. The so-called *control plane* is the software that controls devices in network, such as switching devices, modulators, or BVTs, in real time and maintains the view of a "network." The control plane is able to react to changes in the network, and make it self-sustainable, without external human intervention. Also, upon management commands, control plane procedures are triggered to perform functions.

The control plane maintains a view of the state of the particular devices and makes a set of useful abstractions. For example, in a device, the control plane has to manage the configuration of a Wavelength Selective Switch, program the shape of a filter, or tune the wavelength of the transponder. When the optical components are configured, data can be sent from one transponder to another through the EON. The control plane abstracts a particular configuration of the network elements and has the notion of "connections."

The main functionalities that the control plane provides are:

- *Provisioning* (set up and tear down) of connections: The control plane automatically configures all necessary devices to create a connection between two (or more) points in the network. The process by which the control plane configures different elements to set up a connection is known as *signalling*.
- *Restoration*: Upon a failure in some element of the EON, a connection may no longer be able to meet the necessary quality required for the transmitted service. In this case, through restoration process the configuration of the network is changed so that the connection satisfies the desired quality again. The restoration process usually implies a change over the "physical" path of a connection.
- *Automatic network element discovery*: The control plane automatically discovers which elements are present in the network.
- *Routing*: The control plane automatically builds a topological view of the network; it discovers the connections among network elements and keeps the information up to date. Based on this discovery, a topological graph comprised of nodes and edges is built as an abstracted view of the topology. In addition, traffic engineering (TE) information (e.g. available spectrum and the shared risk link group information of a link representing a fibre) is also added to the graph.
- *Path computation*: Based on the network graph, traffic engineering capabilities of both edges (e.g. availability of spectrum) and vertexes (i.e. connectivity matrix between incoming/outgoing edges), the path of a service is computed. Constraints (e.g. Shared Risk Group (SRLG)) and optimization objectives, such as cost, can be applied to the computation.

Nowadays, the most extended control plane architecture is GMPLS/PCE, and we will look into its details in this chapter. This architecture relies on distributed communication between control elements and occasionally, between control elements and a central element such as in the network with the Path Computation Element (PCE) [1] and Software Defined Networks (SDN) [2], both of which require communications between all configurable elements and a centralized controller.

### *8.1.1 Protocols*

The control plane is software that relies on the communication between control elements to achieve the functionalities mentioned above, and the communication is determined and defined by protocols. Protocols are well-defined formats for exchanging messages. The control plane protocols have been defined to focus on specific functions. For example, in the Generalized Multi-Protocol Label Switching (GMPLS) architecture [3], signalling is performed using the Resource Reservation Protocol—Traffic Engineering (RSVP-TE) protocol [4] with some enhancements [5].

The control plane protocols have the syntax incorporated in the protocol, and define the messages and data structure to be exchanged. Thus, the protocol itself defines both the specific functions and the data model. The trend is to build control plane protocols directly over IP, as in Open Shortest Path First (OSPF) [6], RSVP-TE [4], or over TCP, like the Path Computation Element Protocol (PCEP) [7] or OpenFlow [8]. The control plane protocols are designed to be efficient on the wire and to work in real time, even with network elements that have little computational capability. Thus, efficiency on the protocol is preferred over simplicity to develop.

On the other hand, management protocols, like Simple Network Management Protocol (SNMP) [9], Netconf [10], or the popular REST APIs, are not designed for real time responses on interactions. They aim at Create, Read, Update and Delete (CRUD) operations, and other functionalities like rollback are incorporated. Speed in the response is not a must. However, with the advent of SDN and REST APIs [11], the difference between control and management protocols is beginning to blur.

## 8.2 GMPLS Control Plane Architecture

### *8.2.1 Introduction to the GMPLS/PCE Architecture*

As discussed in the previous section, the deployment of a control plane addresses the automation of optical network functionalities such as connection provisioning and recovery (i.e. protection and restoration), traffic engineering, or Quality of Service (QoS). Until the arrival of SDN-based control plane solutions, most control plane implementations for core and transport networks were based on the protocol suite defined by the GMPLS [3] architecture and so are, to date, the majority of commercial deployments. GMPLS was standardized in 2004 by the Common Control And Measurement Plane (CCAMP) working group of the Internet Engineering Task Force (IETF). One of the initial goals in the working group charter was to extend the MPLS control plane protocols, which had been defined specifically for packet switched networks. The original goal of GMPLS was to control multiple circuit-oriented switching technologies, such as time-division multiplexing (TDM), lambda switching (LSC), and fibre switching (FSC). Specifically, the basis of GMPLS involves enhancements to the MPLS-TE protocols: the resource reservation protocol (RSVP-TE) [5] signalling for setting up end-to-end quality-enabled connections, the open shortest path first (OSPF-TE) [12] routing for automatic topology and network resource dissemination, and the link management protocol (LMP) [13] for control channel connectivity maintenance, link property correlation, and fault localization.

The GMPLS control plane allows establishing explicitly routed TE Label Switched Paths (LSPs) whose routes follow a set of TE constraints. The constraints are used to achieve major TE objectives such as resource usage optimization, QoS guarantee, and fast failure recovery. Mainly, control plane consists of three main components: the *routing and topology dissemination* component, responsible for

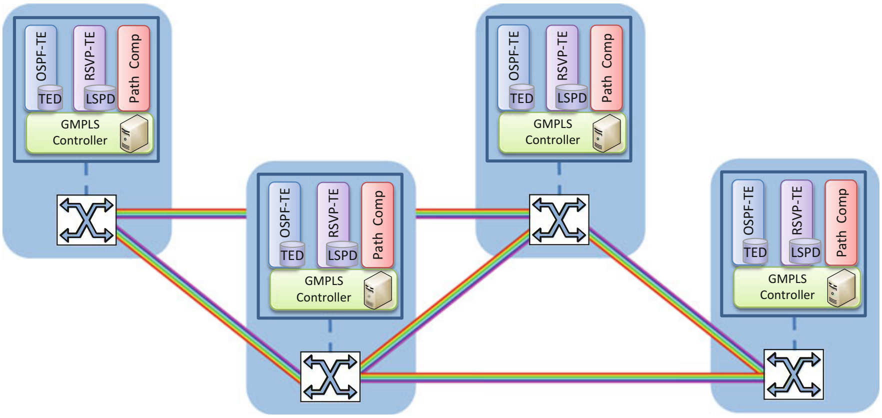



**Fig. 8.1** Example of GMPLS-controlled optical network

the automated discovery of the network topology and attributes in a given domain, the *path computation component*, responsible for computing the network path (route) and required resources, and the *signalling component*, responsible for the actual establishment of the LSP (resource reservation) driving the switching and forwarding behaviour of the network elements. Such components enable the following basic functions: path set up that requires route calculation and resource reservation; LSP modification, with the option to modify selected attributes without tearing down the LSP (depending on the constraints of the underlying technology); LSP rerouting that allows to change the routes of already set up LSPs and LSP preemption in order to handle different classes of service or priorities.

GMPLS included extensions that address specific aspects that were considered important in the context of transport networks, such as the minimization of the setup delay in optical networks: GMPLS permits an upstream node to suggest (propose) a label (wavelength) to a downstream node in the Path message in order to start reserving and configuring its hardware with the proposed wavelength from the source to the destination node, by means of the Suggested Label object. Second, GMPLS also supports bidirectional connection requests, rather than setting up and associating two unidirectional LSPs, and it is also possible to use a single set of `Path` and `Resv` messages from source to destination node and using the Upstream Label object that specifies the value of the label to use to the next node along the path for the upstream direction.

A GMPLS control plane is a distributed entity composed of controllers (commonly one per node) which execute several collaborative processes. The main involved components are the Connection Controller (CC) and the Routing Controller (RC), and optionally a path computation component (see Fig. 8.1). A data communication network based on IP control channels (IPCC) to allow the exchange of control messages between GMPLS controllers is also required, which can be deployed in-band or out-of-band (including, for example, a dedicated and separated physical network). A GMPLS-enabled node (both control and hardware) is named as Label Switched Router (LSR). Each GMPLS controller manages the state of all the connections (i.e. LSPs) originated, terminated, or passing-through a node, stored in the LSP-Database (LSPDB), and maintains its own network state information (topology and resources), collected in a local Traffic Engineering Database (TED) repository.

The *routing controller* (RC) is responsible for disseminating changes occurring in the network state through the GMPLS OSPF-TE routing protocol, allowing the nodes' routing controllers to update their local TEDs and maintain a global picture of the current network topology and resource availability. After a network state change, the node originating the change generates and floods a Traffic Engineering—Link State Advertisement (TE LSA) to all its neighbouring nodes. The neighbouring nodes receive the TE LSAs, forward them, and update their TED repositories. This mechanism allows synchronizing all the nodes' TED repositories so that all node's routing controllers have the same and unique view of the network state within a given time, referred to as the routing convergence time.

The *connection controller* (*CC*) is responsible for provisioning, modifying, rerouting, or release the optical connections through the GMPLS RSVP-TE signalling protocol. The GMPLS RSVP-TE signalling mechanism for provisioning consists of a label request within a RSVP PATH message from the source to the destination node (i.e. forward direction), and a generalized label assignment (reservation) sent in a RSVP RESV message which travels backwards to the source node (i.e. backward direction). This exchange of signalling messages allows each involved connection controller updating its LSPDB repository.

Finally, each GMPLS controller can execute efficient/optimum path computation algorithms (i.e. distributed source routing/path computation) to find a feasible end-to-end path based on the network state information. Thus, upon reception of a connection request at the source node (e.g. through a Network Management System—NMS), such a node executes a path computation algorithm to find a feasible end-to-end path. The computed path is passed to the connection controller as Explicit Route Objects (EROs). Alternatively, the path computation function can be delegated to dedicated entities, named Path Computation Elements (PCEs). A PCE is an architectural component and it can be collocated in the GMPLS nodes or in a separated physical node.

In the particular case of DWDM fixed-grid optical networks, also known as Wavelength Switched Optical Networks (WSON), the problem of computing a physical route (Routing) and allocating a wavelength is known as Routing and Wavelength Assignment (RWA), and it has been widely studied, considering both the wavelength continuity constraint and optical physical impairments. The equivalent assignment in DWDM Flexible grid optical networks, the so-called Spectrum Switched Optical Network (SSON), involves the computation of a physical route but also of a contiguous optical spectrum (i.e. a frequency slot) and a set of modulation parameters, such as the modulation format (e.g. BPSK, QAM), bits per symbol, the number of OFDM sub-carriers, or the Forward Error Correction (FEC), required for configuring the elastic optical transmitters and receivers. By analogy, this is

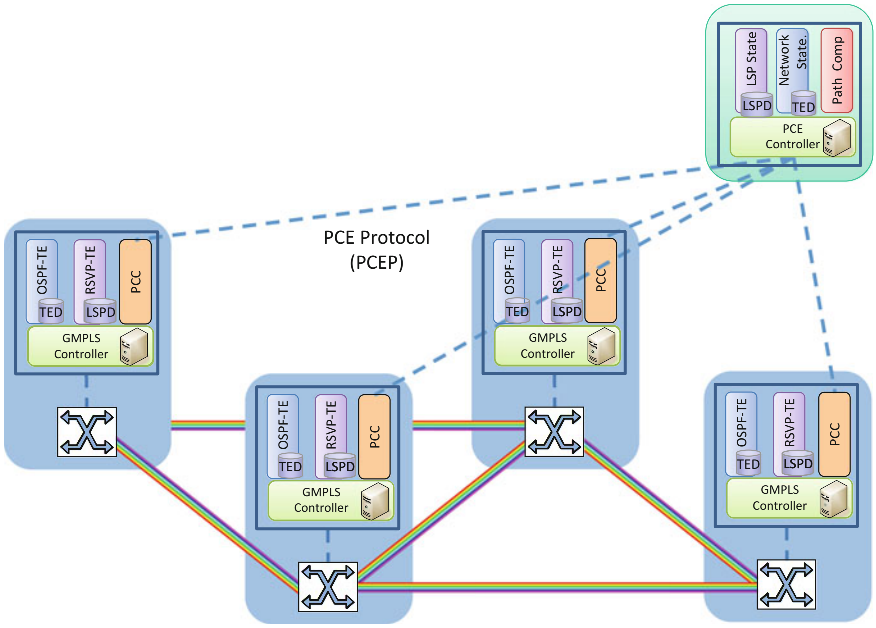



**Fig. 8.2** GMPLS-controlled optical network with PCE based path computation (single PCE deployment model)

commonly referred to as Routing, Modulation and Spectrum Assignment (RMSA), or simply RSA (see Chap. 4). It is worth noting that the allocated basic slots of a frequency slot must be consecutive (in frequency), and the same frequency slot must be available across all fibre links along the route, allowing an end-to-end continuous spectrum (we refer to this term as Spectrum Continuity Constraint, or SCC).

A stateless PCE is an entity (component, application, or network node) that is capable of computing a network path or route based on a network graph (TED) and applying computational constraints. The main idea is to decouple the path computation function from the GMPLS controllers into a centralized and dedicated PCE with an open and well-defined interface and protocol. The IETF defined a standard and functional formalization of the PCE global architecture [1] and the communication interface and protocol (PCEP) [7] in 2006 for MPLS path computation, and it has been recently extended and adapted for GMPLS. By doing so, Path Computation Clients (PCCs) such as GMPLS Controllers or the NMS may request the computation of an explicit route. This is shown in Fig. 8.2.

One of the key aspects of the PCE is the synchronization mechanism by which the PCE obtains a copy of the TED in order to perform path computation. Information on network topology and resource status to build the TED may be provided by participating in interior gateway protocol (IGP) distribution of TE information, or through an out-of-band synchronization mechanism. In the former, a PCE may collect TE information by maintaining a routing adjacency with a GMPLS controller in the domain (act as an IGP passive listener). In the latter, some mechanism (e.g. a topology server) is used by the PCE to retrieve the TED. Such a mechanism could be either incremental (like IGP) or involving a bulk transfer of the complete TED. In general, out-of-band mechanisms may lead to TED synchronization problems. It is worth noting that the TED may also be enhanced to include additional information obtained from other means, such as physical impairment information which may be gathered by dedicated monitors.

When an operator requires to provision/reroute a new/existing LSP, it uses the PCE protocol (PCEP) to send a Path Computation Request (`PCReq`) message to the PCE. The PCReq message specifies the endpoints (source and destination node addresses) and objective functions (requested algorithm/optimization criteria), and the associated constraints such as traffic parameters (e.g. requested bandwidth), the switching capability, and the encoding type. It is also possible to include or exclude network nodes, links, or whole domains (Include Route Object—IRO—and Exclude Route Object—XRO—, respectively), or re-optimize existing paths avoiding resource double booking using the Reported Route Object (RRO). A PCReq message can request the computation of a path or a set of paths. The use of the Synchronization VECtor (SVEC) list in the `PCReq` message allows requesting synchronized computation, that is, the PCE avoids assigning the same resources to the previous computed paths of the same set. In both cases, each path computation request included in the PCReq requests a specific objective function and defines some constraints. If the path computation succeeds, the PCE replies (`PCRep` message) with the computed path(s) specified by means of EROs. Then, the NMS requests the provisioning of new LSPs, or rerouting of existing LSPs to the source GMPLS controllers (head-end of the connections).

A passive stateful PCE allows for efficient path computation and increased path computation success, considering both the network state (TED) and the LSP state (LSPDB) (i.e. set of computed paths and reserved resources in use in the network). Consequently, it is required a strict synchronization mechanism to allow the stateful PCE to build the global LSPDB, based on the local LSPDBs stored in the GMPLS controllers. In [14], it is proposed to extend PCEP with a new message named Path Computation State Report (`PCRpt`) to allow a stateful PCE to learn the LSP state (i.e. the PCCs report about the state of the LSP whenever there is a change in their LSPDB, or under request). Since a passive stateful PCE maintains a global LSPDB that is only used as input for new path computations, it does not have control (modification, rerouting) of path reservations stored in the LSPDB. However, the lack of control of these LSPs may result in sub-optimal algorithms, as we will detail next.

To introduce the concept of active stateful PCE, let us recap that, for instance, flexi-grid DWDM networks show two well-known problems that penalize the network performance, namely, the SCC and spectrum fragmentation. To satisfy the SCC, the PCE tends to compute longer paths and, since low spectral efficiency modulation formats are required to compensate the optical impairments, such paths end up requiring more bandwidth. On the other hand, the spectrum fragmentation is due to the dynamic establishment and release of elastic connections. A fragmented spectrum leads PCE to block the new path computation requests, since no continuous spectrum can be allocated. These problems could be solved if the PCE could optimize (i.e. reroute and reassign the frequency slot) some of the existing elastic connections during the path computation process. To this end, an active stateful PCE allows for optimal path computation considering the LSPDB, but not only as an input of the path computation process, but also for the control of the state (e.g. increase of bandwidth, rerouting) of the stored LSPs. Since under distributed control LSPs are only managed by the GMPLS controllers, this approach requires that the GMPLS controllers temporally delegate the control of a set of active LSPs to an Active Stateful PCE. A delegation mechanism based on the `PCRpt` is proposed in [14] for PCEP. Moreover, an active stateful PCE may also request the modification/rerouting of the existing LSPs to the source GMPLS controllers (i.e. head-end PCC of the connection). To this end [14] also proposes a new PCEP message named Path Computation Update Request (`PCUpd`) sent from the PCE to the PCC. Upon reception of a `PCUpd` message, the PCC triggers the modification of the LSP in the GMPLS controller. Additionally, an active stateful PCE could not only modify existing LSPs but also setup/release new LSPs, referred to as an active stateful PCE with instantiation capabilities. PCEP has been extended in [15] with a new message named LSP Instantiation Request (`PCInitiate`), allowing a PCE to request the creation and deletion of new LSPs to the GMPLS controllers.

Finally, the CCAMP working group is also in charge of producing SNMP Management Information Base (MIB) for the management of GMPLS networks. Several MIBs have been or are being defined, such as the definitions of Textual Conventions for GMPLS Management, GMPLS LSR MIB, GMPLS TE MIB, LMP MIB, or extensions to the OSPF MIB in support of GMPLS.

### 8.2.2 Signalling (RSVP-TE)

The Resource Reservation Protocol (RSVP) is a bandwidth reservation protocol that enables resource reservation to provide different QoS for data flows. The RSVP protocol, with TE extensions, is used as the GMPLS signalling protocol, allowing for explicit path selection or constrain based routing support. GMPLS employs a hop-by-hop signalling scheme through extensions to RSVP-TE signalling protocol. In GMPLS RSVP-TE protocol [5], the signalling phase consists of a generalized label request, sent in a RSVP Path message, traversing hop-by-hop from the source to destination node, followed by a generalized label assignment (reservation), sent in a RSVP `Resv` message, traversing in the opposite direction back to the source node.

To satisfy the continuity constraint in optical networks, GMPLS defines the Label Set object (`LABEL_SET`) [5] which is included in the Path message at the source node. The source node includes a label set specifying the available wavelengths or frequency slots on its outgoing downstream fibre link, and this label set allows an upstream node to restrict the set of wavelengths (labels) that a downstream node can

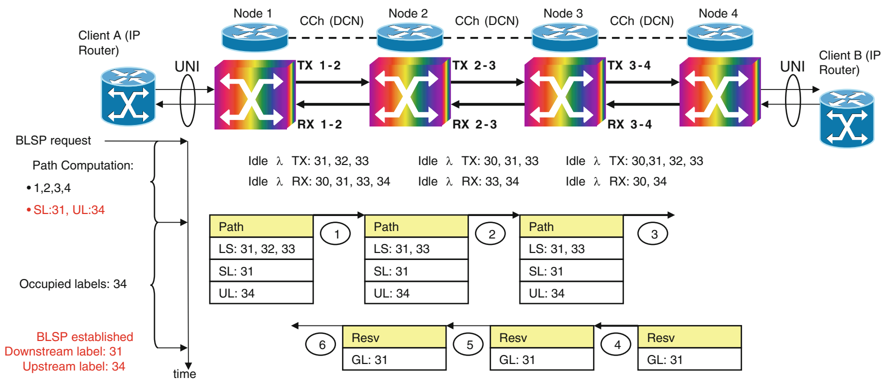



**Fig. 8.3** Signalling of a bidirectional LSP (LSP) using LABEL_SET (LS), SUGGESTED_LABEL (SL) and UPSTREAM_LABEL (UL) objects

choose. Therefore, it is based on a backward reservation protocol scheme, i.e. it collects the available wavelengths in the entire path, but no wavelength reservation is performed. Each intermediate node updates the received `LABEL_SET` object removing currently occupied wavelengths from the label set according to its local wavelength resource information, and then forwards the packet to the next hop. When the Path message reaches the destination node, a wavelength or frequency slot is selected from the resulting label set according to a wavelength assignment algorithm (e.g. First-Fit, Random). This selected wavelength is then notified to the upstream nodes by means of the Generalized Label object (`GENERALIZED_LABEL`). Indeed, before forwarding the `Resv` message towards the source node with the chosen label, each node cross-connects properly the switching fabrics to establish internal data paths. Once the source node receives the `Resv` message, the channel can be immediately used to transport data traffic. If the label is computed by a PCE or the source node, the control plane may use the Explicit Label Control (ELC) feature: the ERO (`EXPLICIT_ROUTE`) contains a list of sub-objects containing the labels to apply at each hop. Each node enforces the allocation by, for example, restricting the outgoing `LABEL_SET` object to the label present in the ERO sub-object.

Figure 8.3 shows an example of the signalling process: a client A requests the establishment of a bidirectional LSP. Node 1 sends a Path message containing the idle (free) labels, including in the label set, 31, 32, and 33. It also includes its preferred label as 31. Finally, since it is a bidirectional label, it adds 34 as upstream label. The Path message is processed by intermediate nodes and the label set trimmed along the way (e.g. value 32 is removed from the set since it is not available). Intermediate nodes must make sure that the suggested label is still contained in the label set, otherwise the suggested label object is removed. In the example, when the path object reached node 4, it allocates generalized label 31 and sends it upstream. The example shows the difficulty of meeting the continuity constraint if detailed information about label (e.g. wavelength or frequency slot) availability is not known when there are continuity constraints. If node 1 selects an upstream wavelength that is not available along the path, it will also cause the set up to fail.

### *8.2.3 Resource Discovery and Topology Dissemination*

**Introduction to Routing, OSPF-TE and IS-IS-TE**

As of today, the term routing is somehow overloaded and can mean different concepts: the routing function can involve the topology management and dissemination function, or even refer to the path computation function. The former deals with the management of network topology and resource information, commonly in terms of graph nodes and links and their attributes; the latter deals with the effective use of that information in order to compute paths across a network that need a set of constraints. Part of this confusion comes from the original routing protocols in which both functions were tightly integrated and coupled.

Nowadays, a routing protocol in the scope of GMPLS is mainly used to disseminate information regarding the network topology and TED, in view of the control and data plane separation mandated by the GMPLS architecture. The information flooded by the routing protocol must be controlled, in terms of volume and periodicity, as it affects the scalability of the solution. A trade-off may involve some kind of bundling or aggregation. As we will detail, there are mainly two routing protocols for GPMPLS: OSPF-TE and IS-IS-TE.

OSPF is an Interior Gateway Protocol (IGP) within a single routing domain, such as an autonomous system (which can, in turn, be segmented into areas). It is based on the concept of "link state routing" in which each router announces to neighbouring nodes the status of its links and those, in turn, disseminate this information to their neighbours thus "flooding" the domain with topological information, allowing each participating node to construct a topology of the network. OSPF was conceived for IPv4 networks in RFC 2328 (1998) [6] and later extended also for IPv6 OSPF Version 3 in RFC 5340 [16]. Once synchronized, the topology information or network graph is further processed—using Dijkstra shortest path algorithm—to obtain a routing table that mapped destination addresses to outgoing interfaces and next hops for IP datagram forwarding, featuring variable-length subnet masking and Classless Inter-Domain Routing addressing models.

Similarly, IS-IS is an alternative link-state routing protocol. Like the OSPF protocol, IS-IS uses Dijkstra's algorithm for computing the best path through the network. The IS-IS protocol was developed by Digital Equipment Corporation as part of DECnet Phase V. It was standardized by the ISO in 1992 as ISO 10589 for communication between Intermediate Systems, allowing the forwarding of datagrams using the ISO-developed OSI protocol stack. IS-IS was later extended to support routing of datagrams in the Internet Protocol (IP) as detailed in RFC 1195 [17]. Although both link-state protocols are conceptually similar (both are dynamic link-state IGPs, and are able to detect changes in the topology such as link failures and converge on a new loop-free routing structure within seconds), there are some differences between both, details of which are out of the scope for this book. The similarity is what caused both protocols to be simultaneously extended for Traffic Engineering purposes and later on for GMPLS. In the following, we will focus on OSPF and its extensions.

In OSPF [6], the segmentation into areas is the main means to scale with network size. Area 0 represents the core or backbone area of an OSPF network. Each additional area must have a direct or virtual connection to the OSPF backbone area, maintained by an area border router (ABR) that maintains separate link-state databases for each area it serves and summarized routes for all areas in the network.

OSPF-TE is an extension to OSPF traffic engineering extensions and use on non-IP networks detailed in RFC 3630 [18]. Traffic Engineering attributes for link and nodes are conveyed into OSPF-TE opaque LSAs carrying type-length-value elements. These extensions allow OSPF-TE to run completely out of band of the data plane network and to describe non-IP networks. OSPF-TE extensions were originally meant for MPLS, and were further refined for GMPLS.

In short, OSPF-TE extensions for GMPLS enable the dissemination of TE link attributes, on a per-link basis, inside a TE LSA. Attributes include link type, link identifier, local and remote interface IP addresses (if numbered interfaces), local and remote unnumbered interfaces, traffic engineering metric, maximum bandwidth, maximum reservable bandwidth, maximum LSP bandwidth, shared risk link groups (SRLGs), and administrative groups. Given the generic nature of GMPLS, the concept of Interface Switching Capability Descriptor (ISCD) was also introduced to describe what kind of switching a node performs in a given link (e.g. packet switching, TDM).

## BGP-LS

The routing protocols, such as OSPF-TE or IS-IS-TE, are able to build and disseminate a TED within a domain. There are certain occasions when it may be useful to export the TED to another control element, such as a PCE, which can perform CPU exhaustive tasks, like complex optimizations, or multi-domain path selection. The BGP-4 protocol has been extended by the IETF to support the exchange of link-state information between two entities [19]. This particular extension of BGP-4 is known as BGP-LS. To summarize, BGP-LS fulfils the function of exporting a TED.

### BGP-LS Session Establishment

In order to proceed with the exchange of TE information, a BGP-LS session between the exporter and the consumer is established. The BGP-LS session between two peers starts as a regular BGP-4 session with an exchange of `OPEN` Messages. The indication that the session is a BGP-LS session (and not a regular BGP-4 session) is performed by adding a Capabilities Optional Parameter (Parameter

Type 2) with a Multiprotocol Extensions capability (Capability Code 1) with Address Family Number (AFI) of 16388 and a Subsequent Address Family Numbers (SAFI) of 71.

Each BGP speaker (peer) acts both as a client and as a server. Thus, the speakers initiate connections and attend others trying to connect to it. After the exchange of `OPEN` messages, `KEEPALIVE` (KA) messages are sent from both sides if peers accept the capabilities.

### Topology Exchange

In order to distribute the TED, a peer sends node and link information in BGP-4 `UPDATE` messages. A peer sends these `UPDATE` messages without waiting for any reply from the remote peer. After the BGP session has been established, the peer that wants to export the topology starts sending the `UPDATE` messages (right after sending the KA messages). The BGP-LS does not define the frequency of the updates, which is left as an implementation choice.

Each `UPDATE` message is encoded as a BGP-4 `UPDATE` message [20]. In BGP-LS, one update message is sent per node or link to be exported. Which particular node or link information has to be exported is based on a policy decision. For example, a peer may decide to export all the details of the topology while another may decide to export only inter-AS links or even send abstracted information.

### Describing Nodes and Links in BGP-LS

The information of an Optical Node or Link is sent within an `UPDATE` message. The node and link information is carried in the Path Attributes of the message. Specifically, when informing of a new node/link (or updating information a new node), the `MP_REACH` attribute has to be used for part of the information, while the rest of the information, related to traffic engineering, is carried in the BGP-LS Attribute. If the node or link is no longer available, the `MP_UNREACH` attribute is used instead.

The part of information in the `MP_REACH` attribute is related to the addressing of the node/link and the domain it belongs to. The `MP_REACH` and `MP_UNREACH` attributes are BGP's containers for carrying opaque information. The `MP_REACH` is characterized by an Address Family Identifier (AFI), subsequent Address family Identifier (SAFI), and a variable Network Layer Reachability Information (NLRI). The NLRI contains the information of the addresses of nodes and links. A link needs to be described by its source and destination (node, interface, and domain it belongs to), and the TE information.

The traffic engineering information is carried in the BGP-LS Attribute. The information to send is independent on the source of the information (OSPF-TE, IS-IS-TE, direct configuration). The goal is to have a generic representation. The information is for example the router IDs of both local and remote nodes, maximum link bandwidth, unreserved bandwidth, metric, default metric, SRLGs, etc.

**Table 8.1** Proposed flexgrid 64-bit label

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid | C.S.  |    Identifier   |              n                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              m                |             Reserved          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

## 8.3 Architecting the Control Plane for EON

### *8.3.1 Extending Signalling for EON*

The reference protocol considered in GMPLS for resource reservation is RSVP-TE. With respect to standard GMPLS (e.g. as in the context of WSON), the general distributed resource reservation mechanism in EON, employing `Path` and `Resv` messages (along with `PathErr`, `ResvErr` messages in the case of signalling block, contention, and errors) is practically unchanged. RSVP-TE extensions have been proposed to fully support the provisioning and the modification of an EON-enabled LSP [21]. The key concept that extends WSON lambda reservation is the reservation of a variable and contiguous spectrum portion, called *frequency slot*, multiple of a reference basic frequency slice. Three aspects have been addressed for the full EON support:

- An extended label supporting flexible grid channel allocation
- An extended requested resource reservation
- An extended description of reservable spectrum set (`LABEL_SET`)

The flexible grid extended label, proposed in [22] and depicted in Table 8.1, is encoded in 64 bit and identifies the reserved channel in terms of the following parameters:

- Grid: the type of grid: fixed (e.g. CWDM, DWDM), flexible (e.g. flexi-grid).
- CS: the channel spacing indication (e.g. 50/100, 6.25 GHz in case of current flexi-grid).
- Identifier: local attribute of different lasers operating at the same frequencies.
- *n*: The central frequency of the slot, expressed as an integer offset from the 193.1 THz reference frequency. The value is computed as *central frequency* = 193.1 + *CS*n* THz.
- *m*: The slot width, expressed as integer multiple of the basic 2*CS frequency slot. The value is expressed as *slot width* = *m**2*CS GHz.

This label is employed in both `Path` and `Resv` messages. In Path messages, it can be carried whenever spectrum assignment is pre-computed or suggested before the signalling phase (e.g. when spectrum is computed at the source node or by a PCE) and in the case of explicit label control (ELC) option (i.e. when label is enclosed in the ERO object). In `Resv` messages, the flexgrid label is used in the

**Table 8.2** Proposed reservation SENDER_TSPEC and FLOWSPEC

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               m               |            Reserved           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

**Table 8.3** Bitmap set representation of label set

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   4   |      Num Labels       |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Base Label                           |
|                             . . .                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Bit Map Work #1 (Lowest numerical labels)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Bit Map Work #N (Lowest numerical labels)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

reservation label to actually perform filter configuration on the selected outgoing interface. Note that the extended label can be also applied to standard fixed-grid and legacy DWDM networks.

In standard GMPLS, the requested type of resource reservation (e.g. type of signal, type of hierarchy, nominal bit-rate) is carried out within the `SENDER_TSPEC` object, enclosed in Path messages. In EON, the requested reservation resource corresponds to the frequency slot width, uniquely identified by the *m* parameter. Similarly, the extended assigned reservation resource is carried out in the `FLOWSPEC` object enclosed in `Resv` messages. The extended objects format is depicted in Table 8.2.

Finally, the `LabelSet` object has to be extended to describe the available and reservable spectrum portions. The possible information structures used to describe the label set can be obtained in many ways, as proposed in [23]:

- Inclusive/exclusive list
- Inclusive/exclusive range
- Bitmap set

Since, with flexgrid, the number of available central frequencies could be high (due to reduced channel spacing), the bitmap set is a strong candidate to represent the label set in a compact, effective, and fixed-size way. In particular, in this case, arrays of Booleans (i.e. bitmap words) describe nominal central frequencies (bit 0: occupied, bit 1: free), starting from a base nominal central frequency identified by a base label (see Table 8.3).

In multi-layer and multi-domain contexts, the reference signalling mechanisms adopted for WSON continue to be valid for EON. Specifically, in multi-domain, nesting, stitching, and contiguous LSPs mechanisms can be employed for end-to-end resource reservation among different areas/domains/vendors.

### *8.3.2 Path Computation in Flexi-Grid Networks*

In EON, path computation represents a key aspect; it enables the benefits of new transponders capabilities and the novel flexible grid granularity. With respect to standard WSON, the number of input/output parameters of an EON-enabled LSP is potentially higher. In addition, novel advanced network operation, available thanks to next-generation BVTs, requires refined path computation resorting to additional network information with respect to standard TED. Finally, as in fixed-grid optical networks, effective path computation is strictly related with impairment validation and may require additional information on optical impairment models, typically available at the management plane. For all the aforementioned reason, particular effort has been dedicated to propose scalable solutions compatible with centralized path computation, performed by the PCE.

Thanks to the flexgrid extensions proposed for routing protocols (e.g. OSPF-TE extensions for flexgrid), the PCE can resort to the TED with link-state information at the flexgrid granularity. For example, it is aware of the occupied central frequencies with the granularity of the considered channel spacing (e.g. 6.25 GHz). Such extended TED enables constrained path computation accounting for typical optical networks requirements (e.g. spectrum continuity constraint for transparent LSPs) [36]. Different PCE architectural schemes are supported:

- Joint routing and spectrum suggestion/assignment (RSA) within a single PCE with a combined algorithm.
- Separated routing and spectrum suggestion/assignment (R + SA) performed either with independent algorithms or at different dedicated PCEs.
- Routing performed at the PCE and distributed spectrum assignment performed by signalling protocol (R + dSA).

A number of path computation routing algorithms (e.g. shortest path, least congested path) available in the literature for WSON can be extended to support novel flexgrid granularity (see Chap. 4).

PCE-based path computation, as in WSON for wavelength suggestion, may also include spectrum suggestion (e.g. in R&SA and R + SA schemes). In EON, the flexgrid label includes not only the suggested central frequency (*n*), but also the frequency slot width (*m*). In this case, two PCE-based spectrum suggestion options can be considered:

- Central frequency suggestion, given an input slot width
- Combined central frequency and slot width suggestion

This represents one of the main extensions of EON-based path computation with respect to WSON. In fact, the path computation request message (i.e. `PCReq`) carries the `BANDWIDTH` object specifying the amount of bandwidth required by the connection. If the bandwidth object carries an input spectrum slot width (*m*), the first option is considered. Conversely, if it carries a requested bit-rate (expressed in Gb/s), the second option is considered. For the first case, the PCE computes the central frequency suggestion based on a path computation request including a rigid constraint on the channel width. Thus, in this case, *m* is an input of path computation. This option can occur whenever a pre-determined amount of spectrum is required for the connection (e.g. connection compatible with fixed-grid—enabled nodes). For the second case, the PCE computes both the central frequency and the frequency slot width. Based on the requested bit-rate, the features/capabilities of the transponders equipped within source and destination nodes should be considered in selecting *m*. Different strategies can be adopted during the computation process: *m* can be selected by referring to a rigid look-up table (for any requested bit-rate a certain slot width is selected) or to a flexible look-up table (for a given requested bit-rate, different values of *m* are possible). The flexible look-up table is employed if network nodes support BVTs, i.e. including the capability to adapt transmission parameters (e.g. modulation format and coding), according to the necessary optical reach. In fact, given a bit-rate, the more efficient the modulation format is for the occupied spectrum, the less robust the signal becomes in terms of quality of transmission (QoT) (which in turn determines the overall optical reach). As an example, 16-quadrature amplitude modulation (16-QAM) halves the required spectrum occupancy with respect to quadrature phase shift keying (QPSK), but 16-QAM presents a limited optical reach with respect to QPSK. As a consequence, for a given requested bit-rate (e.g. 200 Gb/s), different *m* values can be selected, based on selectable modulation format, codes, the modulation format, and other transmission parameters, which may be an output of the RSA [24]. Thus, the PCEP protocol should include modulation format information, in the PCReq message as in traditional optical networks to communicate transponder constraints and moreover it should indicate the selected solution. Works [25] and [24] report experimental results on the aforementioned proposed approach, evaluated in optical network testbeds employing coherent detection applied on single carrier and orthogonal frequency division multiplexing (OFDM) signals, respectively.

In EON, as in WSON, path computation is further complicated by the need to account for physical impairments. A number of end-to-end physical parameters have to be considered to assess the expected QoT. Several PCE-based architectures for impairment-aware RSA (IA-RSA) have been proposed as a solution and successfully validated:

- *Combined impairment validation* (*IV) and RSA Process (IV&RSA)*: the processes of impairment validation and RSA are aggregated into a single PCE.
- *IV-Candidates + RSA Process (IV-candidate + RSA)*: the impairment validation and RSA processes are separated and performed by two different PCE entities. In this case, the IV PCE provides the RSA PCE with a set of validated candidate paths with guaranteed QoT. Thus, the IV PCE exploits physical parameter information besides routing information. Then, the RSA PCE performs RSA on the validated set without considering parameters and QoT.

- *Routing + Distributed Spectrum Assignment and IV*: a PCE, unaware of spectrum resource availability information and physical parameters, is assumed. Spectrum assignment and impairment validation are performed in a distributed way by exploiting either signalling or routing protocol extensions.

While with distributed IV all network nodes must store network physical parameters, with centralized IV it is only the PCE that is required to store (in some special database) and maintain physical parameters. Thus, IV&RSA and IV-candidate + RSA relax the amount of information stored in the nodes. Moreover, in these centralized cases, network resources are typically better utilized. On the other hand, the network may not be able to operate autonomously (e.g. in the case of a PCE failure).

The adoption of an active stateful PCE (storing the state of the LSPs currently installed in the network and directly suggesting modifications of their attributes) increases the control of EON-based lightpaths whereas simplifies re-optimization procedures, as oppose to a standard stateless PCE. In particular, with reference to specific features supported by EON, advanced path computation/reoptimizations can be employed through active stateful PCE:

- Elastic operations, dynamically enlarging or reducing the connection bit-rate (realized through single carrier spectrum expansion/contraction or subcarrier switched on/off in a superchannel).
- Defragmentation (online hitless or global re-optimization), enabling procedures to compact reserved spectrum and improve network utilization performance.
- LSP adaptation, dynamically changing the physical parameter attributes of a LSP (e.g. modulation format, code) upon network degradation events.
- Sliceability functions, enabling configurable multipath transmission.
- Efficient shared protection.
- Advanced QoT-enabled computation.

Different performance can also be experienced according to the stateless or stateful condition. With reference to QoT, the knowledge of the attributes of established LSPs (including physical parameters) enables advanced impairment-aware path computation, with respect to QoT evaluation in the stateless condition, for which worst case scenario and guard bands (also implicitly included) have to be considered leading to waste of spectral resources. Moreover, thanks to the active functionality, the PCE can perform re-optimization in such a way that the waste of spectrum resources dedicated to guard bands is minimized (e.g. by creating pools of contiguous light paths of the same type).

### 8.3.3 Resource Discovery

As explained above, routing and topology dissemination (for example, by means of the OSPF-TE routing protocol) deals with the synchronization between network nodes of a TED that encompasses TE attributes characterizing network nodes and links. One of the main issues to address is that GMPLS/OSPF-TE Type Length Values (TLVs) that convey such information are based on link characterizations in terms of bandwidth (bytes per second), and in flexible/elastic networks, link resources are considered in terms of optical spectrum. The relationship between a link optical spectrum (in Hz) and GMPLS announced bandwidth values (in bytes or bits per second) is not straightforward, and depends on the chosen modulation formats, FEC overheads, and similar considerations.

In consequence, to define protocol extensions for resource discovery (we will focus on the OSFP-TE protocol), the control plane needs to have a model of all the switching elements and their restrictions (e.g. devices may have a different minimum slot size or cannot support all slot sizes). It should be noted that the computation entity needs to get the detailed network information, i.e. connectivity topology, node capabilities, and available frequency ranges of the links. Route computation is performed based on the connectivity topology and node capabilities; spectrum assignment is performed based on the available frequency ranges of the links. Since the computation entity may get the detailed network information using the GMPLS routing protocol, it is proposed to extend and adapt the protocol [26], with information regarding: (1) available Frequency Ranges/slots of DWDM Links. In the case of flexible grids, slot central frequencies span from 193.1 THz towards both ends of the C band spectrum with 6.25 GHz granularity. Different LSPs could make use of different slot widths on the same link. Hence, the available frequency ranges should be advertised, and (2) available slot width ranges of DWDM links. The available slot width ranges need to be advertised, in combination with the Available frequency ranges, in order to verify whether an LSP with a given slot width can be set up or not; this is constrained by the available slot width ranges of the media matrix. Depending on the availability of the slot width ranges, it is possible to allocate more spectrum than strictly needed by the LSP.

For EONs based on flexi-grid, a set of non-overlapping available frequency ranges should be disseminated in order to allow efficient resource management of flexi-grid DWDM links and RSA procedures, i.e. in the flexi-grid case, the available frequency ranges are advertised for the link instead of the specific "wavelengths." The proposed extensions mainly disseminate the status of the Nominal Central Frequencies. Such extensions are carried into the Interface Switching Capability Descriptor (ISCD), and more specifically in the Switching Capability Specific Information (SCSI), shown in Table 2.1 (Table 8.4).

The Switching Capability Specific Information (SCSI) is used to carry the technology specific part of the flexi-grid DWDM, such as Available Labels Set sub-TLV [27], which can be used for specifying the available central frequencies of flexi-grid DWDM links. The bitmap format of Available Labels Set sub-TLV can be used, given its compact form. In this case, the Base Label field specifies the lowest supported central frequency. Each bit in the bitmap represents a particular central frequency with a value of 1/0 indicating whether the central frequency is in the set or not. Bit position zero represents the lowest central frequency and corresponds to the base label, while each succeeding bit position represents the next central frequency logically above the previous (Table 8.5).

With the sub-TLV defined as (Table 8.6):

**Table 8.4** Switching capability specific information

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Switching Cap |   Encoding    |             Reserved          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Max LSP Bandwidth at priority 0                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Max LSP Bandwidth at priority 1                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Max LSP Bandwidth at priority 2                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Max LSP Bandwidth at priority 3                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Max LSP Bandwidth at priority 4                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Max LSP Bandwidth at priority 5                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Max LSP Bandwidth at priority 6                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Max LSP Bandwidth at priority 7                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Switching Capability-specific information              |
|                  (variable)                                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

**Table 8.5** Contents of switching capability specific information

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Type = 1 (Available)   |           Length              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                 Available Label Sub-TLV                       |
|                                                               |
```

**Table 8.6** Available labels set sub-TLV

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     PRI       |              Reserved                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Label Set Field                         |
:                                                               :
```

**Table 8.7** Label set field

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Action|     Num Labels          |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       ncf-number 1                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       ncf-number 2                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       ncf-number ..                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       ncf-number M                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

**Table 8.8** Encoding for bitmap

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    4   |    Num Labels          |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         start NCF                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Bit Map Word #1 (Lowest numerical labels)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:                                                               :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Bit Map Word #N (Highest numerical labels)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

**Table 8.9** Start label nominal central frequency (NCF)

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Grid | C.S.  |    Identifier     |              n                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The PRI field specifies a bitmap used to indicate which priorities are being advertised. The bitmap is in ascending order, with the leftmost bit representing priority level 0 (i.e. the highest) and the rightmost bit representing priority level 7 (i.e. the lowest). The Label Set field is as shown in Table 8.7, where Action can be inclusive list (value 0) or bitmap, with value 4; the Num Labels field specifies the total number of wavelengths supported (e.g. 128 nominal central frequencies); length is the total size in bytes of the field. For the specific Action case set to inclusive list, each "entry" has the format: grid channel spacing–identifier-*n* value.

Finally, Table 8.8 shows the encoding for the specific action case set to the Bitmap.

With start label nominal central frequency (NCF) as shown in Table 8.9, where *n* is set to the lowest supported central frequency. This is in line with the wavelength definition encoding proposed in [28] and with the GMPLS label encoding part that determines the NCF [22].

At the time of writing, there are several proposals being discussed on how to extend OSPF-TE. Of notable importance let us mention extensions for port label and port control restrictions, physical layer impairments, and control of BVTs.

## 8.4 Multi-Domain EON Control Plane GMPLS & H-PCE Architecture

Even when under the control of a single administrative entity, transport networks may be segmented into domains for technical or scalability reasons (e.g. in the form of vendor islands). Such multi-domain networks are characterized by the fact that no single entity has full TE topology visibility, affecting optimality and efficient resource usage. Moreover, each network domain runs its own control plane. Hence, there is a collaboration of control planes in which details of each domain can be hidden, ensuring confidentiality either of the network administrator or the vendor island.

The control of multi-domain flexi-grid networks needs to perform the functions of multi-domain signalling (which allows the provisioning and restoration of multi-domain LSPs), path computation, and routing.

The architecture considered for the multi-domain EON Control plane is built around the hierarchical PCE (H-PCE) framework for an optimal Path Computation, BGP-LS for topology exchange, and RSVP-TE for signalling.

### *8.4.1 Hierarchical PCE*

The Hierarchical Path Computation Element (H-PCE) architecture provides a mechanism to allow the optimum sequence of domains to be selected, and the optimum end-to-end path to be derived through the use of a hierarchical relationship between domains. A parent PCE (pPCE) coordinates several children PCEs (cPCE), one per network domain. The pPCE is in charge of the domain selection and inter-domain path computation. Children PCEs are responsible for segment expansion, i.e. for path computation in their respective domains.

The calculation of optimal paths based on hierarchical PCE is based on three basic aspects: the existence of a parent PCE having access to the topology and inter-domain traffic engineering (and in an optional way, summarized information about the domains or even the whole domain information), the presence of child PCEs in the domains, and the route request exchange between cPCE and pPCE based on the PCEP standard protocol.

The PCEP extensions for the purpose of implementing Hierarchical PCE procedures are described in [29]. The PCE needs to indicate the qualification of the PCE Request, that is if an end-to-end path is wanted, or just the domain sequence. Also, the Multi-domain Objective Functions (OF) and Multi-domain Metrics need to be included.

### 8.4.2 Hierarchical PCE Topology Construction

Each one of the domains must provide information about the nodes which have been designated as border nodes, and as such it may progress connections to other domains, and of the inter-domain links starting from those border nodes. In this way, a multi-domain traffic engineering database may be maintained, accessible from the H-PCE, and which must be permanently updated. This database must have global visibility of the inter-domain focused when the optimal routes for multi-domain optical connections are calculated. A variety of methods are available to provide the parent PCE with the necessary topological and Traffic Engineering information.

The first option is to statically configure all inter-domain link and topology information. This option is not adequate to track changes in the network. The second option is the membership of the parent PCE in an IGP instance; by joining the IGP instance of each child PCE domain, the pPCE could get the necessary topological information. However, by doing so, it would break the domain confidentiality principles and is subject to scalability issues. The third option is to send the information from cPCE to pPCE in PCEP Notification Messages. One approach is to embed in PCEP Notifications the Inter-AS OSPF-TE LSA to send the Inter-Domain Link information from child PCEs to the parent PCE complemented with reachability information (list of endpoints in each domain). However, it is argued that the utilization of PCEP to disseminate topology is beyond the scope of the protocol.

The fourth option is to have a separate IGP instance. RFC 6805 [30] points out that in models such as ASON it is possible to consider a separate instance of an IGP running within the parent domain where the participating protocol speakers are the nodes in the domain and the PCEs (parent and child PCEs). The OSPFv2 protocol [6] with the TE extensions [18], GMPLS extensions [12], and inter-AS extensions [31] can be used to achieve this. Since it is a topological update and synchronization mechanism, it complies with the architecture needs. In this way, an OSPF adjacency is established between the border nodes and the cPCE, and an additional adjacency between each D-PCE and the H-PCE.

The last option is the use of the northbound distribution of TE information (BGP-LS), explained in this chapter. In each domain, a BGP speaker has access to such domain TED and acts as BGP-LS route reflector to provide network topology to the pPCE. Next to the pPCE, there is a BGP speaker that maintains a BGP session with each of the BGP speakers in the domains to receive the topology and build the parent TED.

### 8.4.3 Inter-Domain Signalling

The signalling mechanisms are responsible for the establishment, maintenance, restoration, and release of the LSPs configured to support end-to-end services. GMPLS proposes several alternatives for the establishment of end-to-end LSPs.

The IETF proposes the contiguous RSVP-TE signalling methods, described in RFC 3209 [4] and RFC 3473 [5], hierarchical RSVP-TE, presented in RFC 4206 [32], and concatenated RSVP-TE, detailed in RFC 5150 [33].

On the other hand, the Optical Internetworking Forum (OIF) proposes in [34] a series of abstract messages defined from a mapping with the messages suggested by the IETF in the RFCs related to RSVP-TE, although it introduces some extensions. Until now, the OIF does not cover aspects related to standardization in terms of E2E restoration, which makes it more feasible to develop a solution based on IETF standards.

For contiguous RSVP-TE signalling, a single TE-LSP is established through the multiple domains it crosses using the signalling procedures described in [4] and [5]. No additional TE LSPs are required to create and maintain an E2E LSP. The model based on concatenating segments (stitching) results in a single E2E LSP in the data plane. Nevertheless, in the Control plane, each segment is signalled as a different TE-LSP (different RSVP-TE sessions) and an additional session for the E2E LSP.

RFC 4206 [32] details the signalling mechanism based on the E2E LSP hierarchy. This mechanism allows for nesting one or more E2E LSPs in an intra-domain hierarchical LSP called H-LSP. As described in RFC 5151 [35], H-LSP is the term used to refer to an LSP which allows other LSPs to nest within it. In addition, this H-LSP may be announced in the same instance as the routing protocol used in the intra-domain, in which case it may be called Forwarding Adjacency LSP (FA-LSP).

An FA-LSP is a type of LSP in which its ingress and egress nodes are adjacent. The LSP created as FA-LSPs present additional features against the LSPs related to the calculation, restoration, and re-optimization of optical paths. This way, for example, an FA-LSP enables a better compliance of the control policies. Other example is the ability to implement an admission control for the E2E LSPs crossing through the FA-LSP based on its available resources, instead of looking at every link crossed by the E2E LSP.

Inter-domain TE LSPs can be supported by any of three options, contiguous LSPs, stitched LSPs, and nested LSPs. A single end-to-end RSVP-TE signalling session allows simplified setup and teardown procedures, especially in case of exceptions handling, and in the case of separated signalling sessions (one per domain). As an alternative, taking advantage of the H-PCE structure, the pPCE can orchestrate the cPCEs, acting as the responsible within its own domain, for the establishment (and release) of connections to an underlying GMPLS control plane. By this approach, all PCEs are stateful and have instantiation capabilities. That is, every domain has its own “local” RSVP-TE session and the connectivity at the data plane level is insured by the concatenation of media channels at each domain, while the coordination among the domains (i.e. ingress/egress ports, labels, etc.) is the responsibility of the pPCE. In this case, interoperability requirements are scoped to PCEP extensions for stateful PCE with instantiation capabilities and no protocols are required at the inter-domain boundaries. Thus, the pPCE may coordinate the establishment of segments in a vertical top-down manner, ensuring continuity of the assigned frequency slot as required.

## 8.5 Conclusions

This chapter has described the main concepts of a control plane for optical transport networks and presented in depth how to architect a control plane for EON. In particular, the details of how to extend the GMPLS architecture to support EON are deeply analysed in the chapter. Three aspects have been addressed for the full EON support:

- An extended label supporting flexible grid channel allocation.
- An extended requested resource reservation.
- An extended description of reservable spectrum set (LABEL_SET).

Finally, a desirable feature for the commercial deployment of an EON is to support multi-domain environments. The architecture considered for the multi-domain EON Control plane is built around the hierarchical PCE (H-PCE) framework for an optimal Path Computation, BGP-LS for topology exchange, and RSVP-TE for signalling.

## References

1. A. Farrel, J.P. Vasseur, J. Ash, *A Path Computation Element (PCE)-Based Architecture*, IETF RFC 4655 (2006, August)
2. ONF Foundation, *Software-Defined Networking: The New Norm for Networks*, Open Networking Foundation White paper (2012, April 13)
3. E. Mannie (ed.), *Generalized Multi-Protocol Label Switching (GMPLS) Architecture*, IETF RFC 3945 (2004, October)
4. D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow, *RSVP-TE: Extensions to RSVP for LSP Tunnels*, IETF RFC 3209 (2001, December)
5. L. Berger (ed.), *Generalized Multi-Protocol Label Switching (GMPLS) Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions*, IETF RFC 3473 (2003, January)
6. J. Moy, *OSPF Version 2*, IETF RFC 2328 (1998, April)
7. J.P. Vasseur, J.L. Le Roux (ed.), *Path Computation Element (PCE) Communication Protocol (PCEP)*, IETF RFC 5440 (2009, March)
8. N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, J. Turner, OpenFlow: enabling innovation in campus networks. ACM SIGCOMM Comput. Commun. Rev. **38**(2), 69–74 (2008)
9. J. Case, M. Fedor, M. Schoffstall, J. Davin, *A Simple Network Management Protocol (SNMP)*, IETF RFC 1157 (1990, May)
10. R. Enns, M. Bjorklund, J. Schoenwaelder, A. Bierman (eds.), *Network Configuration Protocol (NETCONF)*, IETF RFC 6241 (2011, June)
11. R.T. Fielding, R.N. Taylor, Principled design of the modern Web architecture, in *Proceedings of the 22nd International Conference on Software Engineering (ICSE '00)*, pp. 407–416 (2000)
12. K. Kompella, Y. Rekhter, *OSPF Extensions in Support of Generalized Multi-Protocol Label Switching GMPLS*, IETF RFC 4203 (2005, October)
13. J. Lang (ed.), *Link Management Protocol (LMP)*, IETF RFC 4204 (2005, October)
14. E. Crabbe, I. Minei, J. Medved, R. Varga, *PCEP Extensions for Stateful PCE*, IETF draft-ietf-pce-stateful-pce (work in progress) December 2015
15. E. Crabbe, I. Minei, S. Sivabalan, R. Varga, *PCEP Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model*, IETF draft-ietf-pce-pce-initiated-lsp (work in progress) October 2015

16. R. Coltun, D. Ferguson, J. Moy, A. Lindem, *OSPF for IPv6*, IETF RFC 5340 (2008, July)
17. R. Callon, *Use of OSI IS-IS for Routing in TCP/IP and Dual Environments*, IETF RFC 1195 (1990, December)
18. D. Katz, K. Kompella, D. Yeung, *Traffic Engineering (TE) Extensions to OSPF Version 2*, IETF RFC 3630 (2003, September)
19. H. Gredler et al., *North-Bound Distribution of Link-State and TE Information using BGP*, IETF draft-ietf-idr-ls-distribution (work in progress) October 2015
20. Y. Rekhter, T. Li, S. Hares, *A Border Gateway Protocol 4 (BGP-4)*, IETF RFC 4271 (2006)
21. F. Zhang, X. Zhang, A. Farrel, O. Gonzalez de Dios, D. Ceccarelli, *RSVP-TE Signaling Extensions in support of Flexi-grid DWDM networks*, IETF draft-ietf-ccamp-flexible-grid-rsvp-te-ext (work in progress) November 2015
22. A. Farrel, D. King, Y. Li, F. Zhang, *Generalized Labels for the Flexi-Grid in Lambda Switch Capable (LSC) Label Switching Routers*, RFC 7699, November 2015
23. G. Bernstein, Y. Lee, D. Li, W. Imajuku, *General Network Element Constraints Encoding for GMPLS Controlled Networks*, IETF draft-ietf-ccamp-general-constraint-encode (work in progress)
24. F. Cugini, G. Meloni, F. Paolucci, N. Sambo, M. Secondini, L. Gerardi, P. Castoldi, Demonstration of flexible optical network based on path computation element. J. Lightwave Technol. **30**(5), 727–733 (2012)
25. R. Casellas, R. Muñoz, J.M. Fabrega, M.S. Moreolo, R. Martinez, L. Liu, T. Tsuritani, I. Morita, Design and experimental validation of a GMPLS/PCE control plane for elastic CO-OFDM optical networks. IEEE J. Sel. Areas Commun. **31**(1), 49–61 (2013)
26. Y. Lee (ed.), *Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSONs)*, IETF, RFC 6163 (2011, April)
27. X. Zhang, H. Zheng, R. Casellas, O. Gonzalez de Dios, D. Ceccarelli, *GMPLS OSPF-TE Extensions in Support of Flexible Grid*, IETF draft-ietf-ccamp-flexible-grid-ospf-ext (work in progress)
28. T. Otani (ed.), *Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers*, RFC 6205 (2011, March)
29. F. Zhang et al., *Extensions to Path Computation Element Communication Protocol (PCEP) for Hierarchical Path Computation Elements (PCE)*, IETF draft-ietf-pce-hierarchy-extensions (work in progress) January, 2015
30. D. King, A. Farrel, *The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS*, IETF RFC 6805 (2012)
31. M. Chen, R. Zhang, X. Duan, *OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering*, IETF RFC 5392 (2009)
32. K. Kompella, Y. Rekhter, *Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)*, RFC 4206 (2005, October)
33. A. Ayyangar, K. Kompella, J.P. Vasseur, A. Farrel, *Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)*, RFC 5150 (2008, February)
34. O.I.F. Forum, *User Network Interface (UNI) 2.0 Signaling Specification* (2008, February)
35. A. Farrel, A. Ayyangar, J.P. Vasseur (eds.), *Inter-Domain MPLS and GMPLS Traffic Engineering—Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions,* RFC 5151 (2008, February)
36. O. Gonzalez de Dios, R. Casellas (eds.), *Framework and Requirements for GMPLS-Based Control of Flexi-Grid Dense Wavelength Division Multiplexing (DWDM) Networks*, RFC 7698, November 2015

# Chapter 9
# Software Defined Networking (SDN) in Optical Networks

**Filippo Cugini, Piero Castoldi, Mayur Channegowda, Ramon Casellas, Francesco Paolucci, and Alberto Castro**

## Acronyms

| | |
|---|---|
| ABNO | Application-Based Network Operations |
| API | Application Programming Interface |
| EMS | Element Management System |
| EON | Elastic optical network |
| GMPLS | Generalized Multi-Protocol Label Switching |
| MAC | Medium Access Control |
| NMS | Network Management System |
| OBS | Optical Burst Switching |
| OCS | Optical Circuit Switching |
| OF | OpenFlow |
| OFDM | Orthogonal Frequency Division Multiplexing |
| OFP | OpenFlow Protocol |
| ONF | Open Networking Foundation |
| ONOS | Open Source Network Operating System |
| OPS | Optical Packet Switching |

F. Cugini (✉)
CNIT, Pisa, Italy
e-mail: filippo.cugini@cnit.it

P. Castoldi • F. Paolucci
Scuola Superiore Sant'Anna, Pisa, Italy

M. Channegowda
University of Bristol, Bristol, UK

R. Casellas
Optical Networks and Systems Department, CTTC, Castelldefels, Spain

A. Castro
Universitat Politecnica de Catalunya, Barcelona, Spain



V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_9

| | |
|---|---|
| OS | Operating System |
| PCE | Path Computation Element |
| PCEP | Path Computation Element Communication Protocol |
| ROADM | Reconfigurable Optical Add Drop Multiplexer |
| RSA | Routing and Spectrum Assignment |
| RWA | Routing and Wavelength Assignment |
| SDN | Software Defined Networking |
| WAN | Wide Area Network |

Software Defined Networking (SDN) has recently emerged as a relevant network architecture enabling the direct programmability of forwarding functions and the effective abstraction of the underlying infrastructure. SDN has been successfully introduced in the context of packet switching (e.g., within data centers) and relevant effort has been recently devoted to extend the SDN architecture to efficiently operate on optical networks.

In this chapter, the general concepts of the SDN architecture are first introduced. Then, the main features of OpenFlow, the most relevant SDN communications protocol that gives access to the forwarding plane of network nodes, are presented. Both the SDN architecture and the OpenFlow protocol are then described and discussed in the specific context of optical networks. Additional solutions, e.g., based on the NETCONF protocol, are also highlighted. Finally, use cases and preliminary implementation solutions are presented and discussed.

## 9.1 SDN Architecture

### *9.1.1 General Concepts*

The concept of software defined networking (SDN) has its bases derived from the computer engineering area, where an abstraction-based architecture has )simplified programming problems of writing and maintaining software. Abstraction is the process by which data and programs are defined with a representation similar in form to its meaning, while hiding away the implementation details. Different level of abstractions can be defined providing different amount of details (high and low levels) and, in turn, creating different interfaces (instances of abstractions).

As stated in [1, 2], the objective of the SDN paradigm is realized by redesigning the architecture of networks in order that they operate in a similar way to that of computing architectures. In fact, as shown in Fig. 9.1 left, computing architectures have a hardware that is generic enough to host different operating system (OS), i.e., the hardware is unbundled from the OS (e.g., Windows or Linux). Similarly, the operating system is generic enough to support a large variety of applications on top of it, i.e., the operating system is unbundled from the applications. In fact, since a long ago, programmers are able to implement complex systems without having to handle the technicalities of individual devices involved or interact in machine language, thanks to the introduction of the most appropriate among the aforementioned levels of abstraction in the architecture.

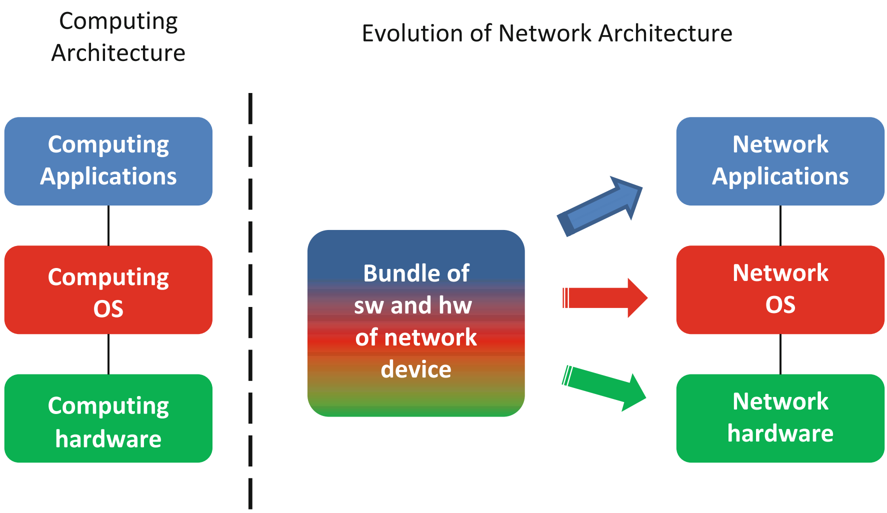



**Fig. 9.1** SDN partitioned architecture (*left*). From integrated to partitioned architecture (*right*)

The same type of logical separation can be applied to the world of networks. As shown in Fig. 9.1 right, most current network devices have software that is highly "integrated" with the network hardware and any hardware modification may require modification to the relevant software of the device. The evolution that is depicted is toward a separation of the basic element of a network element; the hardware and the software that are partitioned in a general network operating system can host different advanced applications. This approach allows for an easy programmability of network elements and an easy replacement of any of the logical blocks in the partitioning.

### *9.1.2 SDN Logical Partitioning*

Figure 9.2 shows a logical view of the) SDN architecture [3]. Three layers can be identified in this architecture: an application layer, a control layer, and an infrastructure layer. Hence, the SDN paradigm envisions a network architecture where the network devices (routers, switches, optical nodes, etc.) become programmable. This objective is realized introducing appropriate levels of abstraction that are accessible by Applications through the use of control interfaces, specifically Application Programming Interfaces (APIs). The interface between the Application and Control Layer goes under the name of northbound interface and is specified by the API, while the interface between Control and Data Layer is the southbound interface defined by various protocol specifications, being OpenFlow [4] the most popular.

Network intelligence is (logically) centralized in SDN controllers, which maintain global view of the network. As a result, the network appears to the applications and policy engines as a single, logical switch. With SDN, enterprises and carriers

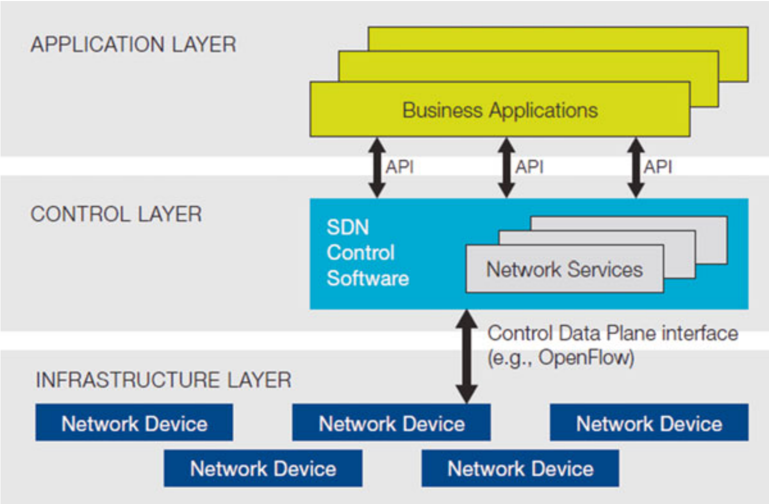



**Fig. 9.2** SDN reference architecture [3]

gain vendor-independent control over the entire network from a single logical point, which greatly simplifies the network design and operation. SDN also greatly simplifies the network devices themselves, since they no longer need to understand and process thousands of protocol standards but merely accept instructions from the SDN controllers [3].

The centralized approach for the control layer has several advantages: single view of the network, no need of synchronization. It also presents some disadvantages especially regarding the aspect of fault localization and failure recovery. A few approaches have been proposed to have multiple controllers from an implementation point of view still maintaining a (logically) centralized architecture.

The SDN approach has become very natural for store and forward packet network especially in the core-metro segments [5] and more recently also in the access segment [6]. The extension of the SDN architecture for optical networks poses several challenges that are considered in the next sections.

## 9.2 OpenFlow Protocol

The SDN architecture encompasses different possible southbound technologies between Control and Data Layer, including the OpenFlow protocol (OFP) [4]. OpenFlow, originally )designed for packet switching, enables the controller to directly configure the forwarding tables of network elements. One OpenFlow controller typically manages multiple OpenFlow switches and an OpenFlow switch can be controlled by one or more controllers. Through OpenFlow, a controller can add, update, and delete flow entries in flow tables both reactively (in response to packets)

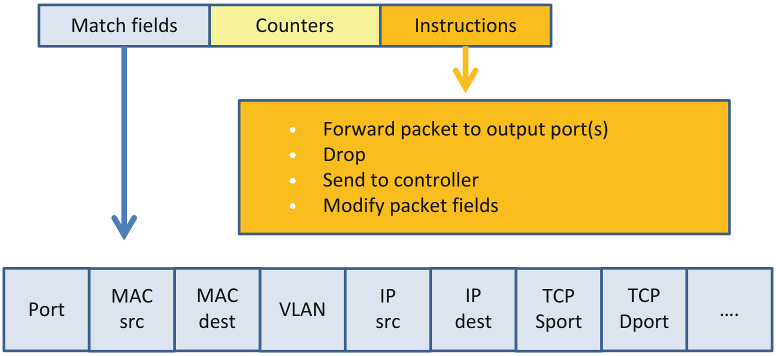



**Fig. 9.3** Main components of a flow entry in a flow table

and proactively. Multiple flow tables per switch may be supported. Each flow table contains a set flow entries. Each flow entry consists of three fields: match fields, counters, and instructions (Fig. 9.3).

When a packet arrives, table lookup is performed. If matching is found, the associated instructions are performed (e.g., forwarding toward a specific output port) and the related counters updated. Otherwise, according to the configuration of the table-miss flow entry, the packet may be forwarded to the controller through the OpenFlow protocol.

Among others, one technical aspect of OpenFlow has triggered relevant interest on this technology. OpenFlow has the capability to define in a standard unique way a flow match and related instructions according to any possible packet field. This potentially enables a single OpenFlow network element to perform any networking action, without requiring, as in today's networks, dedicated devices for layer 2 switches, routers, layer 7 switches, firewalls, etc.

For example, the behavior of a layer 2 switch can be achieved by configuring the flow match according to MAC addresses, a router according to IP addresses, and a layer 7 switch according to other packet headers. Moreover, the behavior of a firewall can be achieved by including within the matching action the packet drop rather than its forwarding, and multicast and broadcast are natively supported by specifying multiple outgoing ports. Furthermore, instructions may also update the content of some packet fields, thus easily enabling operations like address natting or time-to-live decrement.

To efficiently support the aforementioned behavior, the OpenFlow protocol relies on the following solutions and functionalities:

- Secure channel between controller and network element. Through this encrypted channel, the controller configures and manages the switch and its flow entries, it receives events and statistics from the switch, and it sends packets out the switch (e.g., for topology discovery).
- Multiple flow tables, groups, action buckets, and pipeline processing are introduced to support instruction sets and improve scalability performance.

- Both physical and logical ports are defined.
- Flow removal: a flow expiry mechanism enables the switch to remove unused entries, upon timeout expirations.
- A per-flow meter mechanism is supported to enable OpenFlow to implement various simple QoS operations, such as rate-limiting. It can be combined with per-port queues to implement more complex QoS frameworks, such as DiffServ.

### 9.2.1 Main OpenFlow Messages

The following most relevant OpenFlow messages )have been defined:

- *Hello*: exchanged between the switch and controller upon connection startup.
- *Echo request/reply*: mainly used to verify the liveness of a controller-switch connection. They may also be used to measure communication latency or bandwidth.
- *Features*: the controller requests the identity and the basic capabilities of a switch. Reply is expected.
- *Configuration*: the controller sets and queries configuration parameters in the switch.
- *Modify-state*: the controller adds, deletes, and modifies flow/group entries in the OpenFlow tables.
- *Read-state*: the controller collects information from the switch (current configuration, statistics, and capabilities).
- *Packet-out*: used by the controller to send packets out of a specified port on the switch.
- *Packet-in*: the switch transfers the control of a packet to the controller.
- *Port-status*: inform the controller of a change on a port (up/down).
- *Error*: used by the switch or the controller to notify the other side of the connection of problems.

## 9.3 SDN for Optical Networks: Reference Architecture

The potential of SDN, owing to its separation )of control and data layer, is already visible in the packet world with major vendor and service providers lunging forward to support products and services based on SDN technologies [7]. The advantages of SDN for optical transport over existing control plane solution namely GMPLS have shown to be substantial on several specific aspects [8–10]. SDN-based architectures can simplify the complexities of handling traffic among various networking technologies. It allows the underlying infrastructure to be abstracted and used by applications and network services as a virtual entity. This allows network operators to define and manipulate logical maps of the network, creating multiple co-existing network slices (virtual networks) independent of underlying transport

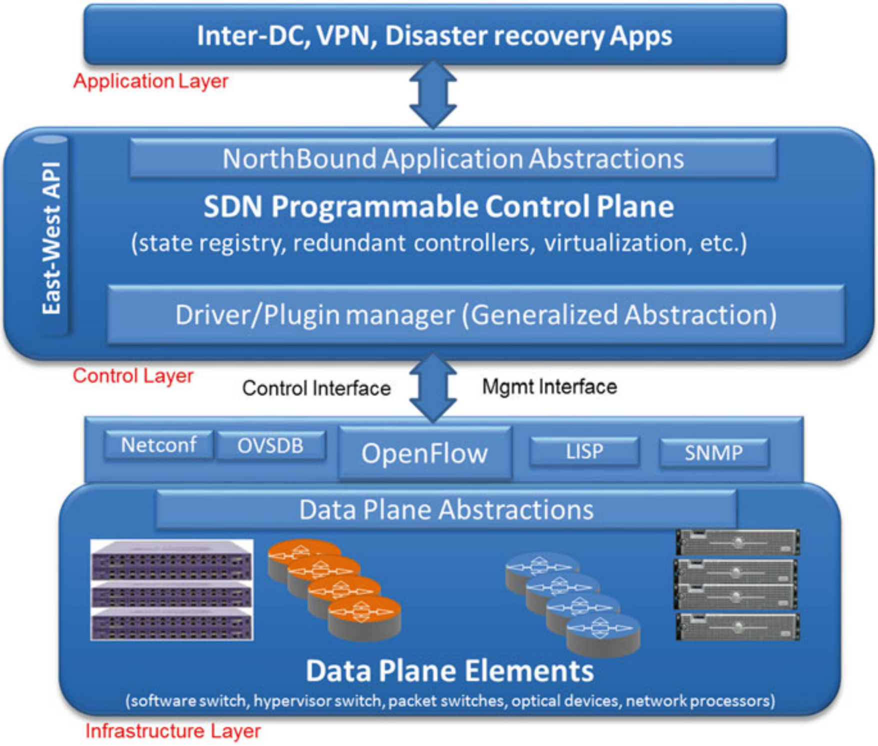



**Fig. 9.4** SDN reference architecture

technology and network protocols [11]. Furthermore, the separation of the control plane and data plane makes SDN a suitable candidate for an integrated control plane supporting multiple network domains and multiple transport technologies. The architecture is not only well suited to address the present optical networking problems related to supporting different administrative and technology segments, bridging gap between packet and optical layers but also support new operations such as virtualization, cross layer orchestration, bandwidth on demand (BoD), load balancing, and many more. Figure 9.4 extends this concept and provides an overview of a reference optical SDN architecture based on principle of network abstraction, its associated drivers, network virtualization, and an application interface to dictate dynamic and flexible network usage.

## *9.3.1 Challenges for Transport Network Abstraction*

Due to the unique analogue features of the optical layer, )various static and dynamic attributes and properties (e.g., modulation format, capacity, power, impairments)) need to be considered in the process of optical resource abstraction. Therefore, the abstraction layer should also be adaptive (real-time) and cognitive (predictive) to reflect the dynamic changes of the physical layer, and continuously monitor physical layer information for learning, analyzing, and characterizing the physical layer parameters. Depending on the type of the involved optical transmission & networking elements and the requirements of the upper layer operations, different levels of abstraction (i.e., the amount of information to be exposed to the upper layer) need to be defined and the corresponding abstraction algorithms need to be designed and evaluated.

One of the key challenges of optimally using the optical infrastructure is to create an abstracted optical resource model and describe it in a way that is simple enough for the higher control layer to utilize it while capturing the true properties of the distributed nonlinear, noisy, and dispersive analogue optical channel. Existing models are often conservative or over-simplified, and targeted to direct-detection systems, now superseded by digital coherent transmission systems. A key challenge here is to develop simplified models for the optical physical layer, starting from the understanding of the underlying physics of the optical channel and the optical layer analogue characteristics.

### *9.3.2 Flow Abstraction*

SDN control plane abstractions heavily )rely on the key concepts of re-usability, modularity, and extensibility. It supports a) common-flow abstraction that fits well with different networks types (packet, optical) and provides a common paradigm for control. It is based on a data abstraction of switch flow tables or cross-connection tables, manipulated by a common switch-API, such as OpenFlow [12], NETCONF [13], etc.

This mechanism (OpenFlow, NETCONF-YANG, etc.) enables generalization of flow switching concept for the underlying heterogeneous optical transport technologies as well as its integration with packet-switched domains. The Control plane using the controller is able to abstract the switching entity and transport format of each technological domain in the form of generic flows and to configure network elements using technology specific flow tables. For example, in packet domains a flow can be any combination of the L2–L4 headers and in optical (fixed or flexible grid) network a flow can be identified by port, frequency slot, modulation format, signal type fields, etc. [14, 15].

### *9.3.3 Drivers*

A programmable control plane needs to know how to configure and operate a variety of hardware appliances and network elements )using mentioned )network/node abstractions. Optical networks vendors bring their own vendor configuration APIs and Element/Network Management System (EMS/NMS) that deter innovation.

Generic techniques such as ontology definition, ontology mapping mechanisms among network elements of different vendors, language modeling, and semantic adaptations and processing, are just a few examples of the many attempts realized by the research community to overcome vendor agnostic barriers. The IETF proposed a more pragmatic approach, and came up with NETCONF [13]. NETCONF, as described in the following sections, is a protocol that allows manipulating the configuration of network elements. YANG is positioned as a good data model candidate, but it remains to be seen if NETCONF and YANG together will become widely used not only for configuration and management purposes, but also for dynamic networking. The most popular of the SDN protocol driver is OpenFlow (current version OpenFlow 1.4) which includes the support of optical devices, as detailed in the next section. There are many other noteworthy drivers such as PCEP [16], OVSDB [17] etc., but their adoption has been slow or limited to specific applications.

### 9.3.4 Virtualization

One of the key drivers of SDN has been virtualization and it brings forward one of the biggest gain for optical transport networks. )Virtualization )enables the creation of logically isolated network partitions over abstracted physical networks and shares them in a flexible and dynamic way. But traditional optical networks are tightly integrated with the underlying physical base (e.g., wavelength), making it difficult to enjoy the benefits of the virtualization concept at the same level as traditional TDM or packet networks [18]. With advent of Elastic Optical networks (EONs), virtualization over emerging flexible optical technologies will play a major role in defining future services. In [18, 19], the role of optical network virtualization and its implication toward future optical technologies focused toward data center services is introduced. Google SDN WAN [20] and NTT SDN intra data center [21] have already showcased the potential of virtualized WAN resources, hence proving the necessity of virtualization in optical networks.

### 9.3.5 Northbound Application Layer Abstraction

The northbound abstraction is crucial )not only for defining )application-based services and policies but also for integrating different technology and administrative domains which are typical of optical-based core networks. The two main aspects to consider for the northbound interface are the set of functionalities that are exposed to external applications, together with the different interfaces that determine the different mechanisms used to interact with them.

The application abstraction is a common-map abstraction, based on a data abstraction of a network wide common map manipulated by a network API. The common map has full visibility into both packet and circuit switches networks, allowing creation of network applications that work across multiple layers and multiple optical technologies. Full visibility allows applications to jointly and globally optimize network-functions and services across multiple layers and technologies. Implementing network-functions as centralized-applications is simple and extensible, as the common-map abstraction hides the details of state-distribution (including east-west API for multiple controllers) from the applications [9], allowing unified operation over multiple layers.

One way to represent information to applications is as graphs. Few implementations of transport network controllers such as ONOS [22] and PAC [9] use network graph as the method to define northbound abstraction. These graph abstraction help optical network control and management function to utilize the wealth of graph computation algorithms that are widely used in the optical world. Another added benefit is to use existing and emerging PCE orchestration architectures (ABNO [23], etc.) to allow applications to seamlessly define network requirements over multiple administrative and technology domains.

### *9.3.6 Performance*

In addition to network resource virtualization) and )flow abstraction, SDN has the potential to guarantee effective traffic engineering solutions due to its centralized view of network resources. Moreover, the capability to directly configure each specific network element (e.g., through OpenFlow or NETCONF) may improve the control of optical connections, avoiding some complexity and limitations of the GMPLS signaling protocol, forced to encompass all networking attributes in a single message along the head-end optical circuit.

On the other hand, as for any centralized architecture, SDN may suffer from scalability and reliability issues. Scalability issues could be overcome by organizing the whole network in several domains of limited visibility and using a hierarchical architecture as adopted by the PCE architecture. Controller reliability could be addressed as well by providing hot-standby mechanisms. However, fast recovery upon network failure seems to represent the most critical challenge to address. Indeed, differently from the distributed GMPLS architecture which allows each node to autonomously react upon failure detection, the SDN technology have to rely on pre/planned restorations and on the involvement of the centralized controller, with relevant risks of scalability issues and long recovery time.

## 9.4 OpenFlow for Optical Networks: Protocol Extensions

As detailed in the previous sections, despite its origins as a protocol) to configure the OpenFlow (OFP)forwarding in the scope of packet-switched networks, there has been a recurring goal of applying the OpenFlow protocol and procedures to transport networks, and to optical networks in particular.

Let us note that this section assumes the choice of the OFP as the protocol to be used as a southbound interface of a SDN controller toward an optical device. Such a choice seems straightforward and well justified, given its availability, industry support, and existence of open source implementations. However, as reported in the next section, this may change in the future, with, for example, different approaches based on NETCONF (or RESTCONF).

Within this technological context, the target of a unified, centralized control plane involves, on the one hand, extending the OpenFlow protocol to support circuit switching (that is, where a dedicated communications channel or circuit is established between endpoints) and, on the other hand, ensuring that an operator use cases and workflows are fully covered. This is of importance since the operation of an optical transport network is significantly different of the operation of a packet-switched network. For example, the former usually involves the operator intervention before provisioning a new optical connection, and the lifetimes of services are in different timescales.

From a standardization perspective, extending OpenFlow for optical networks is perceived to be somehow a difficult task, given the intrinsic complexity of the optical technology. One of the reasons is that there is no common, comprehensive abstract model of optical devices, such as reconfigurable optical add/drop multiplexers (ROADMs), nor the associated information and data models. That said, initial efforts are being made at the ONF Optical Transport working group and other standard developing organizations as well as in related research initiatives.

### *9.4.1 Basic OpenFlow Protocol Requirements*

OpenFlow extensions for optical networks need to cover), as a minimum, a set of basic functional requirements. These requirements cover main control plane functions such as )resource and topology discovery; programming cross-connections; performing the required signal mappings and adaptation and reporting errors, failures, and dynamic status. In other words, a logical OpenFlow switch needs to be able to report to its controller its features, capabilities, and dynamic status of its ports; configure, upon request of the controller, the optical cross-connections and the optical parameters. The main model of operation is assumed proactive. In such mode, service is provisioned as part of an operator controlled process in line with the operator business and operation support policies, as opposed to a reactive model, in which new flows may be data triggered upon detection of a new flow upon reception of a packet that triggers an incoming packet event. Additional requirements for the protocol extensions are related to enabling post-processing, power leveling, equalization, etc. of the path, as well as to consider hardware constraints such as tunability limitations, asymmetric switching capabilities, signal compatibility, or physical impairments.

Despite such complexity, both from a research and prototyping perspective, there have been different proposals that we describe next. It is difficult to evaluate to what extent the proposed extensions meet the requirements, due to the lack of a formal complete standard specification, with the notable exception of the basic support in the recently published OpenFlow v.1.4 specification.[1]

### 9.4.2 OpenFlow for Optical Circuit Switching

The required extensions )of OpenFlow for Optical Circuit Switching (OCS) were partially addressed by the same group that conceived the original) OpenFlow specification. The (now archived and of historical interest) document covering the extensions to the OpenFlow Protocol in support of circuit switching [24], also known as PAC.C, described the requirements of an OpenFlow circuit switch and extended OpenFlow v1.0 messages and data structures to support basic circuit switching operations and signal adaptations. The specification introduced the concept of circuit flow and circuit switch flow table (or cross-connect table), and included several extensions. Such extensions are new information objects describing logical switch features; the notion of circuit switched ports as a new type of physical port; the means to configure cross-connections; new action types for adapting packet flows to circuit flows and new status and error reporting.

However, this circuit switching addendum does not support features that are considered key functions of an optical control plane like accounting for switching constraints and optical impairments, nor supports advanced and emerging optical transport technologies such as flexible DWDM grid. Some encodings do not consider all technology choices and they are rigid (for example, per port bandwidth encoding only supports a limited number of wavelengths assuming a channel spacing of 100 GHz, and a spacing of 50 GHz is not supported).

Despite its shortcomings and limited actual deployments, the circuit switching addendum has been used, mainly in the context of research activities, as the basis to extend OpenFlow to cover optical networks including those based on ITU-T DWDM flexi-grid [10, 14, 15, 25]. The following list tries to capture, from a high-level perspective, what types of extensions are proposed, how they are implemented and encoded. Most of the implementations share the main underlying approaches and either extend the aforementioned circuit switched addendum or proceed by using the "experimental" tagged reserved fields and information ranges.

- The initial handshake between an agent and its controller(s) is adapted, to convey, in the payload of the `OFP_FEATURES_REPLY` message (the ofp_switch_features structure) the number of circuit ports, what capabilities are supported, and a variable list of port descriptors for each of the circuit ports including, mainly, the port name, identifier, and basic switching type.
- This initial handshake, along with subsequent status messages that are sent by the OpenFlow agent to controller (described later) are commonly used as a mechanism for automated topology discovery and dynamic updates. This

[1] OpenFlow extensions covering optical networks were provided by Vello, a member the ONF. Such extensions are included in version 1.4, the latest published version at the time of writing.

mechanism allows the controller to build a network connectivity directed graph that can be used for path computation and Routing and Wavelength/Spectrum Assignment (RWA/RSA). Of course, this does not preclude the case in which the controller authoritatively manages the network topology and link/node status.

- The basic circuit flow table is extended to support a diversity of technologies and related parameters. For example, in a flexi-grid network, the entries contain the characterization of the switched frequency slots, in terms of nominal central frequency and frequency slot width (the so-called *n* and *m* parameters, respectively) and, optionally, signal types, modulation formats, etc.
- If neither a supervisory channel nor a discovery protocol between neighbors can be assumed (thus precluding the ability to inject test messages for port identifier correlation), local and remote port identifiers are typically pre-configured, either at the nodes or at the controller.
- A range of port number identifiers is reserved to specify mapper ports that map, e.g., Ethernet packets to TDM time-slots, used mainly in signal adaptation.
- Common procedures are clarified, to name a few: in order to guarantee the liveliness of a connection between a network element and the controller, echo request and echo reply messages are used. The result of cross-connection request operations (described next) is specifically acknowledged rather than assumed working, etc.
- A new message is defined and used for the configuration of cross-connections, that is, for the addition and removal of circuit flows. Such message (`OFPT_CFLOW_MOD`, although sometimes a new message e.g., `SLICE_MOD` is used for such purpose [15]) includes an ofp_connect structure that conveys associations of the type (incoming port, incoming wavelength channel, outgoing port, and outgoing lambda channel). The message also contains a set of actions, of limited use in the scope of circuit switching, except in the particular case when inserting or removing packet flows to/from circuit flows. For optical networks, the newest extensions encode better the parameters that characterize a service, such as the central wavelength in a fixed DWDM grid or the frequency slot on a flexible grid. Finally, specific extensions address additional capabilities such as central frequency, spectrum range, number of ports and wavelength channels of switches, peering connectivity inside and across multiple domains, signal types, optical constraints, etc.
- A new asynchronous message (`OFP_CPORT_STATUS`) is used to report the dynamic status of the optical ports, adding specific new reasons to report bandwidth (i.e., wavelengths).

### *9.4.3 OpenFlow for Optical Transceiver Configuration and Monitoring*

An inherent aspect of optical networks )is the fact )that the transmission technology uses optical transponders that are incorporated either directly into the IP routers cards or into the optical transport nodes themselves, interfacing with client signals.

Such devices have the ability to configure some parameters dynamically, by software, such as the used modulation format or FEC. Furthermore, new modulation formats are being used in specific scenarios. For example, Optical Orthogonal Frequency Division Multiplexing (O-OFDM) transmits high speed optical signals using multiple spectrally overlapped lower-speed sub-carriers, either as Coherent Optical OFDM (CO-OFDM) or with direct-detection (DDO-OFDM), potentially offering a low-cost solution for EONs, especially in metro networks, and short or medium distance core networks. Finally, different transmission techniques (e.g., modulation formats and bitrates) may co-exist on the same network infrastructure, increasing the need for effective and automatic control and monitoring.

In consequence, the OpenFlow protocol is also being adapted to convey such parameters in OPFT_CFLOW_MOD messages, which are sent to dedicated agents controlling such transceivers, although the encodings and parameters are quite (vendor) specific [26]. Likewise, the monitoring functionalities are re-designed for effective optical networks. In OpenFlow, the controller sends request messages (`OFPT_STATS_REQUEST`) to the switches, which reply with `OFPT_STATS_REPLY`. In [27], lightpaths are monitored thanks to a specific field within a new OpenFlow structure, here called OFPST_STATS_PORT_LP, retrieved by the controller by activating periodic requests of statistics, at configurable time intervals. Such statistics are effectively utilized for management and correlation purposes, for example by enabling the successful localization of possible sources of QoT degradation.

### *9.4.4 OpenFlow for Optical Burst and Packet Switching*

Finally, and mainly in the context of research projects, OpenFlow) is also being extended to support Optical Burst Switching (OBS) and Optical Packet Switching (OPS), either as) standalone technologies or integrated with OCS [4, 28–31]. Such extensions have several aspects in common, such as the use and extension of the `OFPT_PACKET_IN` message to act as at trigger to set up optical resources, when a new flow is detected by edge nodes, and the adaptation of the existing `OFPT_FLOW_MOD` message. The `OFPT_FLOW_MOD` message is used to convey, for example, incoming optical label information and allowing the controller to install a flow in the optical packet switch. In particular, in such reported experimental testbeds, OF enabled OPS nodes that are at the edge of the OPS network have both electrical and optical interfaces and, upon reception of an unmatched electrical packets, they will send an `OFPT_PACKET_IN` to the controller, which will modify the flow tables accordingly within the OPS domain and configure the transceivers accordingly. The details of such protocol extensions remain mainly undisclosed and only reported at a descriptive level, or are conceived for specific use cases and subject to change.

From a standardization perspective, the OpenFlow specification version 1.4 [4] officially contemplates optical networks, where a new set of port properties add support for Optical ports, including fields to configure and monitor the transmit and receive frequency of a laser, as well as its power and optical port descriptions to describe optical port capabilities. These extensions use encodings not available in older (i.e., v1.0) protocol definitions such as property types, and can be summarized as:

- The controller uses the OFPT_PORT_MOD message to modify the behavior of the port, using a variable list of properties. This allows the controller to configure the central frequency, offset, grid span, and transmission power. The grid span is the amount of bandwidth consumed by this port, useful for flexi-grid as well as other tuning information.
- New properties are used to report the optical port statistics, such as current frequency, power, or laser temperature.

Finally, let us note that the ONF Optical Transport working group has internally being working in preliminary OpenFlow extensions for multi-technology switches with GMPLS label encodings, complementing [24] and adapted to OpenFlow 1.1. That said, it is not specified whether such extensions will be published and the only published version at the time of writing remains version 1.4. It is expected that newer versions will improve the support of optical networks, in order to address, in a more comprehensive way, identified requirements both for fixed- and flexi- grid, notably filter configuration and configuration of cross-connections of media channels in media matrices.

## 9.5 NETCONF Protocol

### *9.5.1 Configuration of Packet Switches*

In traditional OpenFlow architectures designed for packet) flows, a minimum set of parameters is assumed to be pre-configured at the network switches. Typical examples of these parameters include the )assignment of one or more OpenFlow controllers to a switch, the enabling/disabling of ports, the configuration of certificates for secure communication between switches and controllers.

In currently deployed OpenFlow switches, these parameters are typically configured through manual intervention. To enable remote configurations through network protocols and avoid inefficiencies due to manual operations, the Open Networking Foundation (ONF) is standardizing automatic mechanisms to provide remote settings. However, ONF is not considering the OpenFlow protocol as particularly suitable for configuration and management purposes. Indeed, changes in configuration parameters are not expected to be frequently performed or to be handled at the fast timescale of a packet flow. Moreover, and most important, differently with respect to packet flow encodings that are based on bit-level protocol specifications, configuration parameters generally rely on more complex data

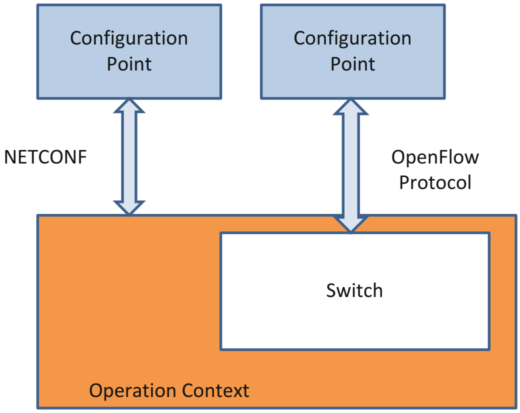



**Fig. 9.5** Example of a traditional OpenFlow switch supporting, besides the OpenFlow protocol, the OpenFlow Configuration and Management Protocol (OF-CONFIG) [32]

models, typically described through eXtensible Markup Language (XML) specifications.

To define the exchange of configuration parameters between network switches and management platforms, ONF has recently introduced the OF-CONFIG specification [32]. In OF-CONFIG, rather than extending the OpenFlow protocol, the NETCONF protocol [13] is proposed as the configuration transport mechanism enabling the configuration of OpenFlow switches by a remote configuration entity, called Configuration Point in the context of OF-CONFIG. Figure 9.5 shows the reference OF-CONFIG architecture, including both the OpenFlow protocol for packet flow handling and the NETCONF protocol for management and configuration purposes.

NETCONF, first introduced by RFC 4741 in 2006 [13], uses a data encoding based on XML to provide configuration information through protocol messages. NETCONF enables functionalities to install, manipulate, and delete configurations. Moreover, it supports both synchronous and asynchronous event notifications and the exchange of operational parameters. Security and reliability in message exchange is guaranteed by the use of the SSH transport protocol. NETCONF relies on YANG as a human-friendly data modeling language [33].

NETCONF has been widely deployed in most of networking devices, mainly routers and optical nodes. However, so far, many implementations adopted proprietary solutions for the data modeling, making difficult the implementation of a multi-vendor interoperable networking solutions and controllers.

To address this issue, standardization bodies are nowadays posing relevant efforts on the definition of common YANG models. To this extent, OF-CONFIG also defines the XML-based data model specifying classes, attributes, and descriptions of individual elements in the context of OpenFlow architectures. For example, a switch is defined through several types of data resources including ports, queues, certificates, and flow tables (Fig. 9.6). The data model is then based on XML strings, adopting a namespace which uniquely identifies the individual elements.

```
<!-- Example for a physical port -->
<port>
      <resource-id>Port214748364</resource-id>
      <number>214748364</number>
      <name>name0</name>
      <current-rate>10000</current-rate>
      <max-rate>10000</max-rate>
      <configuration>
            <admin-state>up</admin-state>
            <no-receive>false</no-receive>
            <no-forward>false</no-forward>
            <no-packet-in>false</no-packet-in>
      </configuration>
      <state>
            <oper-state>up</oper-state>
            <blocked>false</blocked>
            <live>false</live>
            </state>
            …
```

**Fig. 9.6** Example of XML description for a physical port of a packet switch [32]

### *9.5.2 Configuration of Optical Nodes*

In optical networks, management and configuration aspects are significantly more relevant and complex than in packet flow networks. Indeed, the modularity and complexity of optical node architectures and technologies, as well as the presence of advanced functionalities and transmission techniques typically make the startup and maintenance phases of optical nodes extremely critical.

XML-based configurations are already supported within the large majority of optical nodes (Fig. 9.7). Moreover, remote configurations through NETCONF protocol are typically already enabled from a centralized network management system. However, also in this case, different implementations adopt completely different proprietary solutions for the data modeling, affecting the implementation of multi-vendor interoperable networking controllers.

Specific effort is in place to define a common YANG model defying classes and attributes for the main optical node components and capabilities. The related namespace and XML strings will enable the standard configuration of the most relevant and common parameters. The possibility to support some vendor-specific solutions will be also guaranteed, given the aforementioned intrinsic complexity of optical nodes and devices.

The YANG model is expected to encompass also standard attributes and parameters to be used for the setup of optical connections, such as frequency slot, signal type, modulation format, etc. However, these parameters may be controlled also through the extended version of the OpenFlow protocol.

```
<!-- Example for a physical port of an optical node-->
<port>
       <resource-id>Port003018323</resource-id>
       <number>003018323</number>
       <name>name0</name>
       <fabricId>1</fabricId>
       <OCHCtp_cdb_array>
              <OCHCtp_cdb_entry>
                     <class>OCHCtp</class>
                     <lineId>1</lineId>
                     <ochId>1</ochId>
                      <data>
                            <mappingMode>2</mappingMode>
                            <nIndex>-272</nIndex>
                     </data>
              </OCHCtp_cdb_entry>
              ...
```

**Fig. 9.7** Example of a portion of XML description for a physical port of an optical node

At the time of writing, some vendors are arguing about the actual need to support the extended version of OpenFlow for optical networks and not just limit remote configurations to take place through NETCONF. The main motivations in support of this approach are:

- NETCONF implementations are typically already available while the support to OpenFlow needs to be introduced on optical nodes.
- NETCONF will be used anyway for configuration purposes.
- NETCONF will evolve to adopt standard YANG models also encompassing attributes for the setup of optical connections.
- NETCONF implementations are expected to become more efficient and reactive. For example, NETCONF is a candidate technology also in the context of the Interface to the Routing System (I2RS), expected to provide fast access to node parameters.
- Dynamic operations on optical circuits are not expected to take place at the timescale of packet flows.
- As described in previous sections, with respect to the original SDN architecture designed for packet switching, OpenFlow requires relevant changes and enhancements, not only in terms of protocol objects, but also in the full concept of flows, tables, node features (e.g., stateful controller) and constraints, error reporting and monitoring.
- The presence of both NETCONF and OpenFlow protocols will introduce additional complexity and the need to implement an intermediate abstraction layer to account for possible attributes handled by both protocols.

## 9.6 Examples and Use Cases

In this section, some preliminary use cases and implementation experiments are presented to show the SDN capability to provide efficient centralized configuration while simplifying the control operations performed by networks nodes. Indeed, in contrast to control planes based on GMPLS, flooding of traffic engineering information among network nodes is avoided as well as distributed signaling to perform node configurations. In these use cases, the SDN centralized controller triggers node configurations by directly providing the flow switching information through a communication protocol. In particular, an extended version of OpenFlow for optical networks is here considered.[2]

### *9.6.1 Use Case I: Restoration*

The main feature of SDN in support of restoration is its inherent capability of configure several elements (e.g., ROADMs and transponders) in parallel, thus reducing total configuration times. Note that reducing configuration time is of paramount importance to reduce total restoration times.

In this example, the use of bitrate squeezing and multipath restoration (BATIDO) in the context of dynamic SDN-controlled EONs is presented [34]. In that regard, first, the BATIDO problem is formally stated; the BATIDO problem aims at maximizing the amount of restored bitrate by exploiting the available spectrum resources also along multiple routes. Second, SDN architecture is introduced to support the configuration of sliceable bandwidth-variable transponders in EONs. OpenFlow extensions are presented to control the specific EON transmission parameters.

#### Bitrate Squeezing and Multipath Restoration

To illustrate the restoration schemes that can be applied to the set of connections affected by a failure in the context of EONs, Fig. 9.8 shows a simple network topology where a lightpath is set-up between nodes $s$ and $t$. In normal conditions (Fig. 9.8a), let us assume that the lightpath uses a slot consisting of 16 frequency slices to convey the requested bitrate, in our example 400 Gb/s.

Let us imagine that a link failure occurs and the lightpath is affected. Upon failure detection and localization (e.g., exploiting a distributed monitoring system and the Link Management Protocol—LMP), restoration is performed. If a restoration lightpath, including route and frequency slot, can be found in the network for the

[2] The introduced protocol messages, fields, and objects reported in this section, as part of advanced experimental implementations, have to be considered open for changes and interoperability agreements.

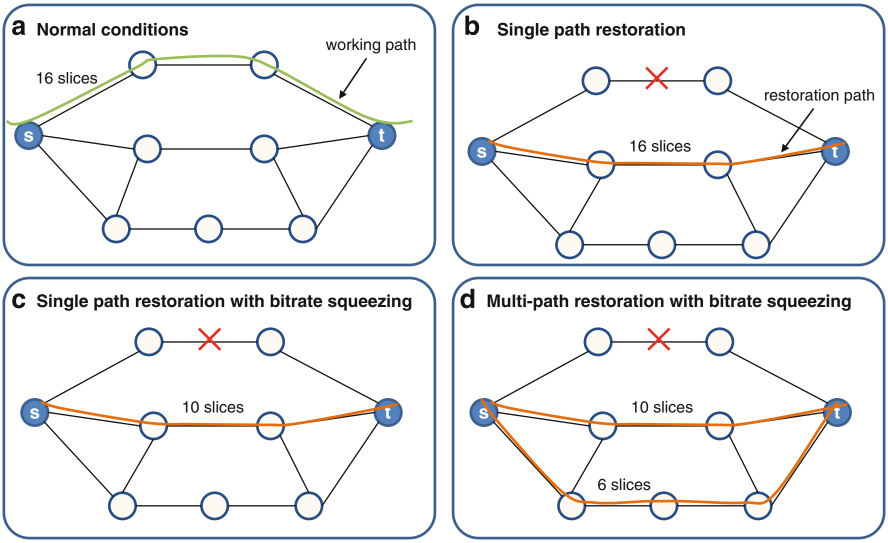



**Fig. 9.8** Bitrate squeezing and multipath restoration

required 16 slices, the disrupted lightpath is obviously restored using the restoration path. This is the normal restoration scheme that has been traditionally used in optical networking; we call this scheme as single path restoration (Fig. 9.8b).

However, in contrast to protection schemes, the availability of 16 contiguous frequency slices at failure time is not generally guaranteed in restoration schemes. In such case, the disrupted lightpath can be restored in part. This case, named single path restoration with bitrate squeezing in this paper, is illustrated in Fig. 9.8c. Note that the restoration lightpath uses just ten frequency slices and thus, the conveyed bitrate has been squeezed to 200 Gb/s.

Another possibility is to use several parallel lightpaths, each conveying part of the total bitrate, so restore the original bitrate of the disrupted lightpath (Fig. 9.8d). Note that in this case, although restoration lightpaths use 16 frequency slices, the total bitrate cannot be recovered, since 200 Gb/s can be conveyed within 10 slices and only 100 Gb/s within 6 slices. This illustrates the fact that spectral efficiency decreases when multiple lightpaths are used. Although for that very reason network operators prefer not using multipath for provisioning, it can be exploited to improve restorability, provided that the number of parallel lightpaths is kept limited.

### The Bitrate Squeezing and Multipath Restoration Problem Statement

The problem can be formally stated as follows:

*Given*:

- A network topology represented by a graph G($N$, $E$), where $N$ is the set of optical nodes and $E$ is the set of fiber links connecting two optical nodes,

- A set $S$ of available frequency slices of a given spectral width in each fiber link in $E$,
- A set $D$ of failed demands to be restored, each requesting a fixed bitrate $b^d$.

*Output*: the routing and spectrum assignment for each restored demand $d \in D$.
*Objective*: maximize the total restored bitrate.

As previously discussed, the BATIDO problem can be faced using three different approaches:

- Single path restoration, where the total requested bitrate is either restored using a single lightpath or blocked,
- Bitrate squeezing restoration, where part of the originally requested bitrate is restored whilst the rest is blocked,
- Multipath restoration with bitrate squeezing, where several restoration lightpaths can be established to recover partially or totally one single demand.

### Example of SDN Implementation Supporting BATIDO

The support of multipath restoration and bitrate ),squeezing is here provided through a SDN architecture enhanced to operate over an EON. In this implementation, the OpenFlow protocol has been extended to enable the configuration of flexi-grid lightpaths by configuring flexible transmitters, receivers, and intermediate nodes. Three novel messages are hereafter considered: the `LIGHTPATH_IN` message (extended from standard `PACKET_IN` message), the `LIGHTPATH_OUT` message, and the `FLOW_ACK` message. Moreover, the existing `FLOW_MOD` message is extended to support the configuration of the spectrum-flow entry.

The spectrum-flow entry stores the currently active cross-connections of the switch (input port, output port), along with the related reserved spectrum, expressed as the tuple {central frequency, channel width in terms of number frequency slices}, i.e., $n$ and $m$ values of the ITU-T flexible grid label. The structure of the extended OpenFlow messages is depicted in Fig. 9.9.

The `LIGHTPATH_IN` message is sent by the source node to the controller requesting the provisioning or restoration of a lightpath. It is inherited by the `PACKET_IN` message and includes the most utilized parameters of a flexible lightpath, such as the endpoints and the requested bitrate. The optional duration time is included in the message. The extended `FLOW_MOD` message is sent by the controller to the switches in order to set up a new flow entry. The novel *com* field specifies the kind of operation (i.e., new flow, modification of existing flow, deletion) and includes the cross-connections indications (i.e., output port, input port), the flexible grid frequency slots (i.e., grid/CS/identifier, $n$ and $m$), the modulation format (MF), and forward error correction (FEC) indication. The latter parameters are utilized in the case of source or destination node, indicated by the source/destination/transit (SDT) flag. Moreover, the information bitrate and optional optical channel specification (e.g., in case of multiple sub-carriers the number of sub-carriers and their displacement in terms of central frequency and width) are carried out [35]. A specifically introduced

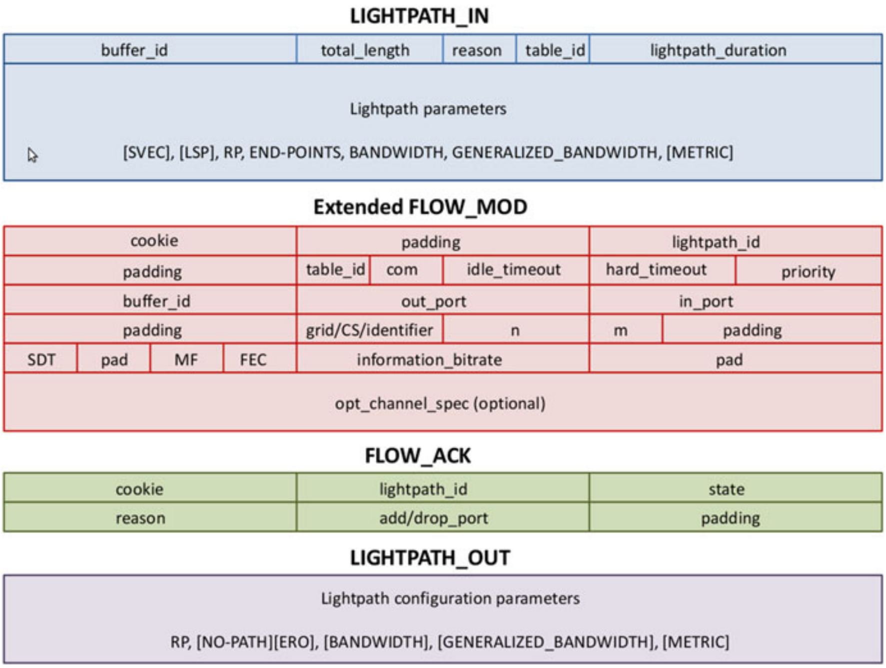



**Fig. 9.9** OpenFlow message extensions (experimental)

`FLOW_ACK` message is sent by the switch upon the reception of the related `FLOW_MOD` message, when the data-plane cross-connection configuration is performed. The result of such configuration is reported in this message, specifying the state of the related flow entry (i.e., enabled/disabled) and the possible reason of a failed configuration (e.g., software error, hardware error). The `LIGHTPATH_OUT` message reports the outcome of the lightpath setup and the actual configuration parameters to the source node. In the case of successful setup, its reception triggers the data plane to start data transmission onto the active lightpath.

### Evaluation on Recovery Time

Figure 9.10 depicts the recovery time as a function of the node configuration time (Pan-European network is considered). The figure shows that with both GMPLS&PCE and SDN the average recovery time linearly increases with the node configuration time. However, the obtained slopes are different. In particular, the GMPLS/PCE case is less scalable when node configuration time typical of EONs is considered (i.e., higher than 100 ms). Conversely, the SDN control plane significantly reduces the recovery time thanks to the parallel signaling procedure adopted for performing the configuration of nodes.

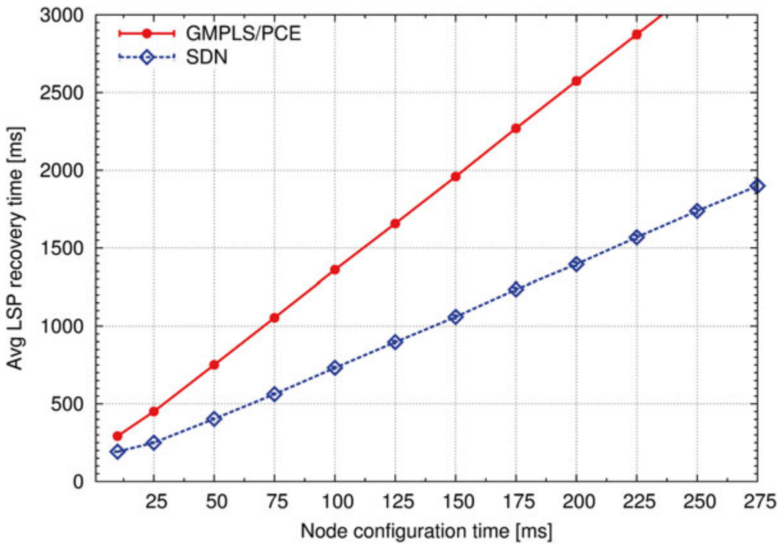



**Fig. 9.10** Average recovery time as a function of the node (ROADM) configuration time

## *9.6.2 Use Case II: Filter Configuration Optimization*

In this example, a recently proposed technique, called super-filter [36], is presented. The technique aims at reducing detrimental filtering effects and improving the overall network spectrum efficiency. A super-filter consists in the aggregation of multiple independent filter configurations related to different lightpath connections that flow through common filters (i.e., output ports).

Super-filters are suitable for SDN architecture thanks to its capability to directly perform independent node configurations. On the other hand, super-filters could be more difficultly implemented in the context of PCE and GMPLS architectures where node resources (including traversed filters) are configured according to head-end lightpath parameters.

### Proposed Super-Filter Technique

Figure 9.11 shows two lightpaths *A–B* and *E–F* routed through transit nodes *G–L*. A broadcast-and-select node architecture is here considered, where one filter per node is traversed by any incoming optical signal. In this example, each lightpath occupies 37.5 GHz of spectrum resources. An amount of spectrum of 25 GHz is available between the two lightpaths along transit nodes *G–L*.

A new lightpath request arrives from *C* to *D*. With traditional strategies, the path computation only accounts for the available 25 GHz along *G–L*. Thus,

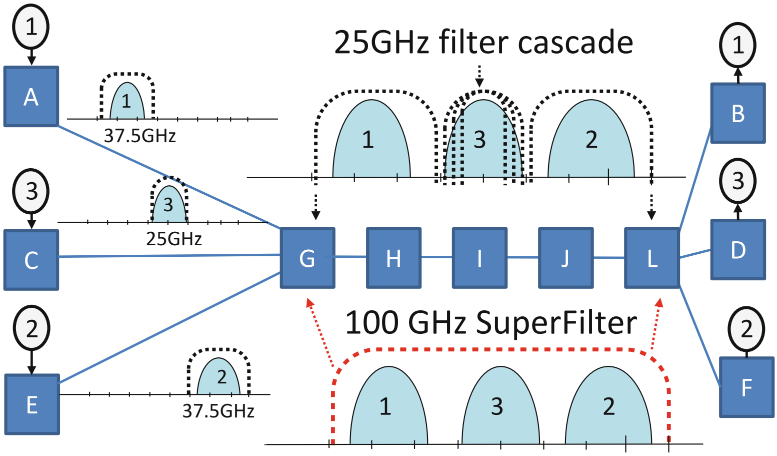



**Fig. 9.11** Example of super-filter application

impairments are evaluated accounting for 25 GHz of narrow filtering per traversed node. Given the excessive degradation introduced by the filter cascade, the *C*–*D* path computation fails along the considered resources and the request is rejected.

The proposed super-filter technique consists in a path computation strategy which specifically accounts for other existing lightpaths and in particular, for the actual configuration of filters in the network nodes. That is, filter configuration is computed such that different lightpaths, also with *different* source–destination pairs, can co-exist within the same flat region of a single filter configuration. With reference to the example above, when the new lightpath request arrives from *C* to *D*, the technique enables the path computation to also account for the whole filter configuration in all candidate nodes. In particular, the path computation can also consider a unique filter configuration in nodes *G*–*L* among all three lightpaths. That is, the path computation can be successfully achieved with only 25 GHz of available spectrum resources. Indeed, a filter of value 100 GHz is computed and applied to all three lightpaths in transit *D*–*G* nodes, as shown in Fig. 9.11 (bottom). Lightpath *C*–*D* then traverses a cascade of filters in their flat region, without experiencing significant transmission degradation along the *G*–*L* route.

Differently with respect to traditional networking solutions, also in case of tear down, the behavior is modified. If the *A*–*B* lightpath is torn-down, the SDN controller has to account for the presence of the considered *C*–*D* connection. Thus, all the *A*–*B* spectrum resources will not be completely released. In particular, a slice of 12.5 GHz contiguous to *C*–*D* resources will be maintained reserved along *G*–*L*.

### Super-Filter Implementation, OpenFlow Extensions, and Experimental Demonstration

Once the request from source *C* to destination *D* arrives, the node controller at *C* sends the `LIGHTPATH_IN` message to the OF-controller to request both path computation and lightpath set up. Upon message reception, the OF-controller computes the path *C* and *D*. The OF-controller exploits a physical parameter database (PPD) to have knowledge of the QoT experienced by the lightpath along the network for the considered modulation format (PM-QPSK in this example). The OF-controller, considered under stateful conditions, also exploits a label switch path database (LSP-DB) storing the attributes of existing active lightpaths, including their routes and related filter configurations. This way, the OF-controller is able to account for filtering cascade effects as well as QoT parameters (e.g., OSNR) during path computation.

As described above, when super-filters are not considered, the OF-controller relies only on available network resources (i.e., 25 GHz along *G*–*L*). Thus, the *C*–*D* request cannot be computed along *G*–*L* due to unacceptable QoT.

On the other hand, by accounting for super-filters, the OF-controller successfully computes the path through that spectrum resources along *G*–*L*. In particular, the OF-controller accounts for the presence of contiguous spectrum resources reserved along the path and it computes their expansions to incorporate the new request. The OF-controller then communicates to the node controllers at *C*, *D* and to all the intermediate nodes along *G*–*L* to perform the lightpath setup. In particular, the OF-Controller sends `FLOW_MOD` messages extended for flexi-grid networks to configure the optical switches, the transmitter, and the receiver. The `FLOW_MOD` messages directed to the transit nodes change the filter configuration identified by the nominal central frequency (*n*) and the channel width in terms of frequency slices (*m*).

As shown in the Wireshark captures of Figs. 9.12 and 9.13, a spectrum reservation of 25 GHz is sent to node *C* and *D*, while 100 GHz is configured in nodes *G* to *L*. In the latter case, a specifically introduced `SUPER_FILTER` flag is enclosed within the `FLOW_MOD` to bypass local filter contention and admission control and allow the filter overlap on the existing filters of *A*–*B* and *E*–*F* lightpaths. This way, the path is successfully activated.

If the *A*–*B* lightpath is torn-down, the OF-controller will not completely release the *A*–*B* resources. In particular, a slice of 12.5 GHz contiguous to C–D resources will be maintained along *G*–*L*. To this extent, the `FLOW_MOD` of type "delete" will include the `SUPER_FILTER` flag and will enclose the (*n*) and (*m*) values (different from those installed within the flow entry) corresponding to the portion of the filter to be released.

As shown by the experiment above, node configurations decoupled from head-end lightpath parameters are natively and easily supported by the SDN architecture, without requiring complex signaling functionalities as in the case of the GMPLS protocol suite.

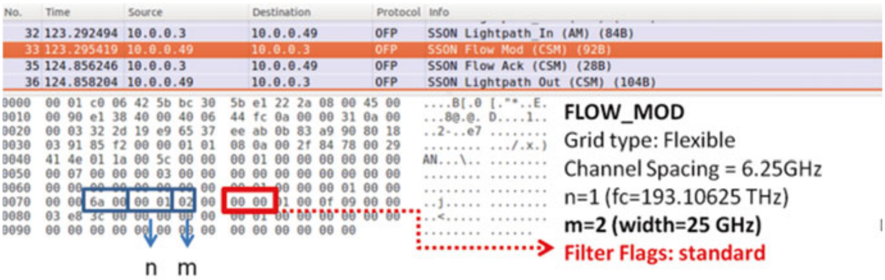



**Fig. 9.12** Message capture at node controller C

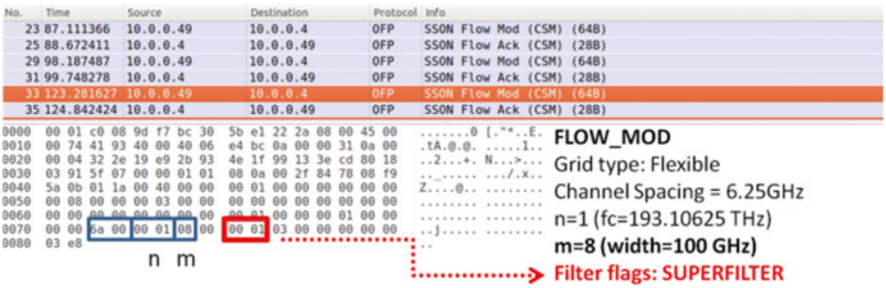



**Fig. 9.13** Message capture at node controller H

## 9.7 Conclusions

In this chapter, the Software Defined Networking (SDN) architecture has been presented and discussed. SDN enables the direct programmability of forwarding functions, guaranteeing effective abstraction of the underlying infrastructure. SDN has been successfully experimented and applied in packet-switched networks (e.g., data centers) and significant work is on-going to extend the SDN architecture to operate on optical networks.

SDN relies on southbound communication protocols to provide direct programmability of the forwarding plane of network nodes. The two most relevant southbound protocols are presented: OpenFlow and NETCONF. The former is widely adopted and mainly designed to operate in packet-switched networks. The latter presents interesting flexibility to be applied in the specific context of optical networks.

Finally, use cases and first implementation solutions are presented and elaborated.

## References

1. S. Shenker, *The Future of Networking, and the Past of Protocols*. Open Networking Summit 2011, keynote speech (2011)

2. N. McKeown, *Software Defined Networks and the Maturing of the Internet* (IET Appleton Lecture, London, 2014)
3. *Software-Defined Networking: The New Norm for Networks*, ONF White Paper (2012, April 13). https://www.opennetworking.org/sdn-resources/sdn-library/whitepapers. Accessed 14 Jun 2014
4. Open Networking Foundation, *OpenFlow Switch Specification, version 1.4 (Wire protocol 0x5)* (2013, October 14). https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.4.0.pdf
5. D. Mcdysan, Software defined networking opportunities for transport. IEEE Commun. Mag. **51**(3), 28–31 (2013)
6. K. Kerpez, J. Cioffi, G. Ginis, M. Goldburg, S. Galli, P. Silverman, Software-defined access networks. IEEE Commun. Mag. **52**(9), 152–159 (2014)
7. *SDN Based Companies*. http://www.sdncentral.com/sdn-directory/
8. S. Das, G. Parulka, N. Mckeown, Why OpenFlow/SDN can succeed where GMPLS failed, in *38th European Conference and Exhibition on Optical Communications (ECOC 2012)*, Technical Digest (CD) (Optical Society of America, Washington, DC, 2012), paper Tu.1.D. 1
9. S. Das, G. Parulkar, N. McKeown, P. Singh, D. Getachew, L. Ong, Packet and circuit network convergence with OpenFlow, in *Conference on 2010 Optical Fiber Communication (OFC), Collocated National Fiber Optic Engineers Conference* (IEEE, San Diego, 2010)
10. M. Channegowda, R. Nejabati, D. Simeonidou, Software-defined optical networks technology and infrastructure: enabling software-defined optical network operations [invited]. IEEE/OSA J. Opt. Commun. Netw. **5**, A274–A282 (2013)
11. D.E. Simeonidou, R. Nejabati, M. Channegowda, Software defined optical networks technology and infrastructure: enabling software-defined optical network operations, in *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC)* (IEEE, Anaheim, 2013)
12. N. McKeown et al., OpenFlow: enabling innovation in campus networks. ACM Commun. Rev. **38**(2), 69–74 (2008). Accessed 02 Nov 2009
13. R. Enns, *NETCONF (Network Configuration Protocol)*, IETF RFC 4741 (2006)
14. M. Channegowda et al., Experimental demonstration of an OpenFlow based software-defined optical network employing packet, fixed and flexible DWDM grid technologies on an international multi-domain testbed. Opt. Express. **21**, 5487–5498 (2013)
15. L. Liu, R. Muñoz, R. Casellas, T. Tsuritani, R. Martínez, I. Morita, OpenSlice: an OpenFlow-based control plane for spectrum sliced elastic optical path networks. Opt. Express. **21**, 4194–4204 (2013)
16. J.P. Vasseur, J.L. Le Roux, *Path Computation Element (PCE) Communication Protocol (PCEP)*, IETF RFC 5440, (2009)
17. B. Pfaff, B. Davie, *Open vSwitch Database Management Protocol*, IETF RFC 7047 (2013)
18. M. Jinno, H. Takara, K. Yonenaga, A. Hirano, Virtualization in optical networks from network level to hardware level [invited]. IEEE/OSA J. Opt. Commun. Netw. **5**(10), A46–A56 (2013). doi:10.1364/JOCN.5.000A46
19. R. Nejabati, E. Escalona, P. Shuping, D. Simeonidou, Optical network virtualization, in *15th International Conference on Optical Network Design and Modeling (ONDM)*, (IEEE, Bologna, 2011), pp. 1–5
20. S. Jain, et al., B4: experience with a globally-deployed software defined WAN, in *Proceedings of the ACM SIGCOMM 2013, Hong Kong*, (2013)
21. *O3 Project SDN WAN*. http://www.sdnjapan.org/
22. *Network Operating System*. http://onlab.us/tools.html#os
23. A. Farrel, D. King, *PCE-Based Architecture for Application-Based Network Operations*, IETF RFC 7491 (2015)
24. S. Das, *Extensions to the OpenFlow Protocol in Support of Circuit Switching*, addendum to OpenFlow Protocol Specification (v1.0)—Circuit Switch Addendum v0.3, (2010, June). http://archive.openflow.org/wk/images/8/81/OpenFlow_Circuit_Switch_Specification_v0.3.pdf

25. R. Casellas, R. Martínez, R. Munoz, L. Liu, T. Tsuritani, I. Morita, An integrated stateful PCE/OpenFlow controller for the control and management of flexi-grid optical networks, in *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC)* (IEEE, Anaheim, 2013)
26. L. Liu, W.R. Peng, R. Casellas, T. Tsuritani, I. Morita, R. Martínez, R. Muñoz, S.J. Yoo, Design and performance evaluation of an OpenFlow-based control plane for software-defined elastic optical networks with direct-detection optical OFDM (DDO-OFDM) transmission. Opt. Express. **22**, 30–40 (2014)
27. F. Paolucci, F. Cugini, N. Hussain, F. Fresi, L. Poti, OpenFlow-based flexible optical networks with enhanced monitoring functionalities, in *38th European Conference and Exhibition on Optical Communications (ECOC 2012)*, paper Tu.1.D.5, (Optical Society of America, Washington, DC, 2012)
28. L. Liu, D. Zhang, T. Tsuritani, R. Vilalta, R. Casellas, L. Hong, I. Morita, H. Guo, J. Wu, R. Martinez, R. Muñoz, First field trial of an OpenFlow-based unified control plane for multilayer multi-granularity optical networks, in *Optical Fiber Communication Conference and Exposition and National Fiber Optic Engineers Conference (OFC/NFOEC 2012)*, paper PDP5D.2, (Optical Society of America, Washington, DC, 2012)
29. T. Miyazawa, *Architecture for Interworking Between an Optical Packet & Circuit Integrated Network and OpenFlow-based Networks,* iPOP 2013, Tokyo, (2013, May)
30. H. Harai, *Optical Packet and Circuit Integrated Networks and SDN Extensions*, ECOC 2013, Mo.4.E.1, (2013, September)
31. X. Cao, N. Yoshikane, T. Tsuritani, I. Morita, T. Miyazawa, N. Wada, Openflow-Controlled Optical Packet Switching Network with Advanced Handling of Network Dynamics, in *2014 European Conference on Optical Communication (ECOC)*, Cannes (2014, September)
32. Open Networking Foundation, *OF-CONFIG 1.2—OpenFlow Management and Configuration Protocol*, version 1.2 (2014)
33. RFC 6020, *YANG—A Data Modeling Language for the Network Configuration Protocol (NETCONF)*, IETF, (2010, October)
34. F. Paolucci, A. Castro, F. Cugini, L. Velasco, P. Castoldi, Multipath restoration and bitrate squeezing in SDN-based elastic optical networks, (invited paper). Photon. Netw. Commun. **28**(1), 45–57 (2014)
35. N. Sambo, F. Paolucci, F. Cugini, M. Secondini, L. Potì, G. Berrettini, G. Meloni, F. Fresi, G. Bottari, P. Castoldi, Software Defined Code-Rate-Adaptive Terabit/s Based on Time-Frequency Packing, in *Optical Fiber Communication Conference*, (Optical Society of America, Washington, DC, 2013) OTh1H. 5
36. F. Paolucci, F. Cugini, F. Fresi, G. Meloni, A. Giorgetti, N. Sambo, L. Poti, A. Castro, L. Velasco, P. Castoldi, Superfilter technique in SDN-controlled elastic optical networks [Invited]. IEEE/OSA J. Opt. Commun. Netw. **7**(2), A285–A292 (2015)

# Chapter 10
# Application-Based Network Operations (ABNO)

**Daniel King, Víctor Lopez, Oscar Gonzalez de Dios, Ramon Casellas, Nektarios Georgalas, and Adrian Farrel**

## Acronyms

| | |
|---|---|
| ABNO | Application-based Network Operations |
| ASON | Automatically Switched Optical Network |
| BGP-LS | Border Gateway Protocol Link State |
| GMPLS | Generalized Multi-protocol Label Switching |
| H-PCE | Hierarchical Path Computation Element |
| IP/MPLS | Multi-protocol Label Switching over Internet Protocol |
| LSP | Label-switched Path |
| LSP-DB | LSP Database |
| NFV | Network Function Virtualization |
| NMS | Network Management System |
| OF | Open Flow |
| OXC | Optical cross-connect |
| PCE | Path Computation Element |
| QoS | Quality of Service |

D. King (✉)
Lancaster University, Lancaster, UK
e-mail: daniel@olddog.co.uk

V. Lopez • O.G. de Dios
Telefonica, Madrid, Spain

R. Casellas
CTTC, Castelldefels, Spain

N. Georgalas
British Telecom, London, UK

A. Farrel
Old Dog Consulting, Llangollen, UK



V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_10

| | |
|---|---|
| SDN | Software Defined Networking |
| TE | Traffic Engineering |
| TED | TE Database |

Networks today integrate multiple technologies, allowing network infrastructure to deliver a variety of services to support the different characteristics and dynamic demands of applications. There is an increasing goal to make the network responsive to service requests issued directly from the application layer and high-layer client interfaces. This differs from the established model where services in the network are instantiated in response to management commands driven by a human user using a wide variety of Operational Support Systems (OSS), and where networks are typically over-provisioned to ensure minimal traffic loss, even at peak traffic periods.

## 10.1 General Concepts

An idealized network resource controller would be based on an architecture that combines a number of technology components, mechanisms, and procedures. These include:

- Policy control of entities and applications for managing requests for network resource information and connections
- Gathering information about the resources available in a network
- Consideration of multilayer resources and how topologies map to underlying network resources
- Handling of path computation requests and responses
- Provisioning and reserving network resources
- Verification of connection and resource setup

## 10.2 Network Abstraction

A major purpose of Software Defined Networks (SDN) is to bury complexity and make service deployment and overall network operation simpler without invoking the management and provisioning software of the many manufacturers deployed in the network. Consequently, allowing higher-layer applications to automate requests and creation of services simpler and more direct.

### *10.2.1 Logically Centralized Control*

We use the term "logical centralized" to signify that network control may appear focused in a single entity, independent of its possible implementation in distributed form. The centralized control principle states that resources can be used more efficiently when viewed from a global perspective.

A centralized SDN controller would be able to orchestrate resources that span a number of subordinate domains or in cooperation with other entities, and thereby offer resource efficiency when setting up services and overall operation of network resources. Other reasons for logically centralized control include scale, optimization of information exchange and minimization of propagation delay.

Given constraints of not being able to always deploy green field networks, it is necessary that a controller co-exist with both native SDN forwarding technologies (OpenFlow) non-native SDN traffic engineered technology (MPLS, GMPLS, etc.).

### 10.2.2 Application-Driven Use-Cases

Dynamic application-driven requests and the services they establish place a set of new requirements on the operation of networks. They need on-demand and application-specific reservation of network connectivity, reliability, and resources (such as bandwidth) in a variety of network applications (such as point-to-point connectivity, network virtualization, or mobile back-haul) and in a range of network technologies from packet (IP/MPLS) and optical transport networks, to Software Defined Networks (SDN) forwarding technologies, application-driven use cases include:

- *Virtual Private Network (VPN) Planning*—Support and deployment of new VPN customers and resizing of existing customer connections across packet and optical networks
- *Optimization of Traffic Flows*—Applications with the capability to request and create overlay networks for communication connectivity between file sharing servers, data caching or mirroring, media streaming, or real-time communications
- *Interconnection of Content Delivery Networks (CDN) and Data Centers (DC)*—Establishment and resizing of connections across core networks and distribution networks
- *Automated Network Coordination*—Automate resource provisioning, facilitate grooming and regrooming, bandwidth scheduling, and concurrent resource optimization
- *Centralized Control*—Remote network components allowing coordinated programming of network resources through such techniques as Forwarding and Control Element Separation (ForCES) OpenFlow (OF)

An SDN Controller framework for network operator environments must combine a number of technology components, mechanisms and procedures including:

- Policy control of entities and applications for managing requests for network resource information and connections.
- Gathering information about the resources available in a network.

- Consideration of multilayer resources, and how these topologies map to underlying network resources.
- Handling of path computation requests and responses.
- Provisioning and reserving network resources.
- Verification of connection and resource setup.

The overall objective is to develop a control and management architecture of transport networks to allow network operators to manage their networks using the core principles of Software Defined Networks and to allow high-layer applications and clients to request, reconfigure and re-optimize the network resources in near real time, and in response to fluid traffic changes and network failures.

This chapter outlines the core network control principles required for application-based network operations of transport networks and discusses key control plane principles and architectures. It introduces the Application-Based Network Operations (ABNO) Framework [1], and how this framework and functional components are combined for Adaptive Network Manager (ANM) [2], used to address the requirements for operating Elastic Optical Networks (EONs) [3]. Finally, the chapter provides a view of the research challenges and areas for investigation to continue development of Transport SDN and control of EONs.

## 10.3 Network Control

A central principle of SDN is the separation of network forwarding and control planes (Fig. 10.1). By separating these functions, a set of specific advantages in terms of centralized or distributed programmatic control might arise. Firstly, there is a potential economic advantage by using commodity hardware rather than proprietary specific hardware. Secondly, remove the need for a fully distributed control

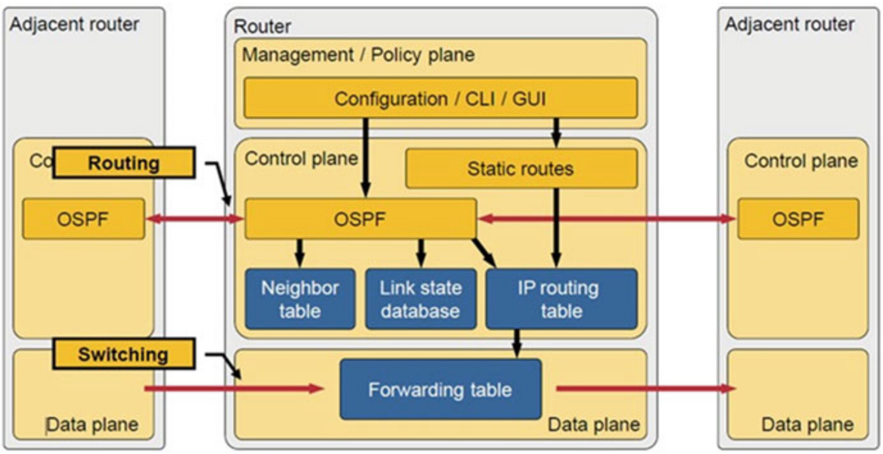



**Fig. 10.1** Management, control, and forwarding example

plane with capability often requiring senior engineering experience to deploy and operate, with a wide range of features, which are very often underutilized. Thirdly, the ability to consolidate in one or a few places what is often a considerably complex piece of OSS software to configure and control network resources.

Typically, the network operator has followed a prescribed path for hardware upgrade to circumnavigate the networking scaling issues. This requires the operator to consider the node forwarding performance versus price-to-performance numbers to pick just the right time to participate in an upgrade. Conversely, as network topologies increase, the complexity of the control plane and scalability also need consideration.

The Internet represents an example of a significant scaling problem. Vast numbers of administrative regions loosely tied with the interconnections changing constantly as traffic patterns fluctuate and failures occur. Therefore, to address the control paradigm, the Internet was designed accordingly. Its structure was federated, where individual nodes participate together to distribute reachability information in order to develop a localized view of a consistent, loop-free network using IP forwarding. The Internet forwarding paradigm, where routes and reachability information are exchanged, later results in data plane paths being programmed to realize those paths; however, paths are often suboptimal and prone to traffic congestion, so clearly this approach has weaknesses which might be addressed using a centralized approach.

As network technology evolved and the concepts of SDN were invented (centralized control, superstation of control and forwarding, and network programmability), the cycle of growth and scaling management and upgrade in the control plane to accommodate scale, was a clear objective. It is much easier to pursue solutions for a centralized management environment controlling distributed, but simple, forwarding elements.

### *10.3.1 Control Plane*

The control plane is the part of the node architecture that is concerned with establishing the network map. Control plane functions, such as participating in routing protocols, are control elements. This establishes the local rule set used to create the forwarding table entries, interpreted by the data plane, to forward traffic between incoming and outgoing ports on a node (Fig. 10.2). The foundation of the current IP control plane model is to use an Interior Gateway Protocol (IGP). This normally is in the form of a link-state protocol such as Open Shortest Path First (OSPF) or Intermediate-System-to-Intermediate-System (ISIS). The IGP will establish layer 3 reachability between the IP forwarding elements.

Layer 3 network reachability information primarily concerns itself with the reachability of a destination IP prefix. In all modern uses, layer 3 is used to segment or stitch together layer 2 domains in order to overcome layer 2 scaling problems. In most cases, the routing table contains a list of destination layer 3 addresses and the

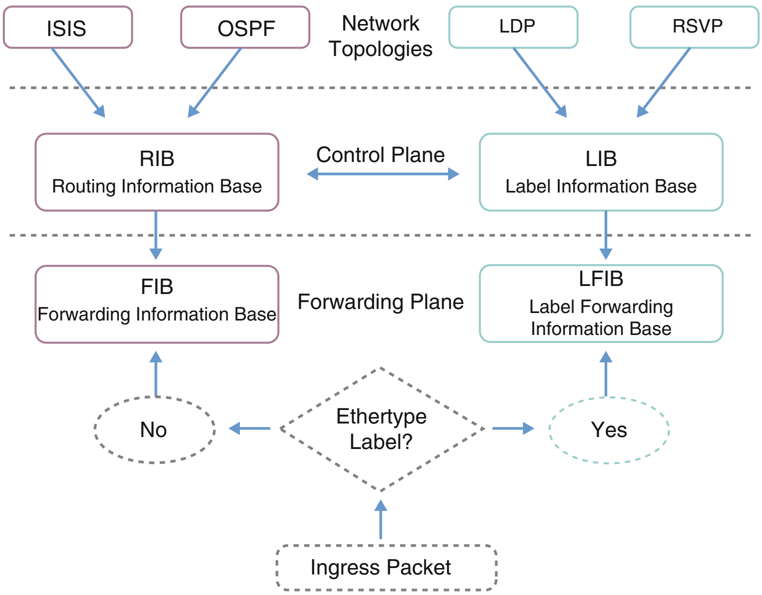



**Fig. 10.2** Relationship of control and forwarding plane

outgoing interface(s) associated with them. Control plane logic can define certain traffic rules, for priority treatment of specific traffic for which a high quality of service is defined and known as differentiated services. Forwarding focuses on the reachability of network addresses.

The role of the control plane includes:

- Network topology discovery (resource discovery)
- Signaling, routing, address assignment
- Connection setup/teardown
- Connection protection/restoration
- Path Computation and Traffic engineering

## *10.3.2 Management Plane*

The Management Plane is responsible for managing the control plane. It performs a number of responsibilities, including configuration management and applying policy. It also provides Fault Management, Performance Management, and Accounting and Security Management functions.

In their early deployments, optical transport networks were inherently managed, deployed in a single administrative domain, and locked to a single vendor hardware solution (i.e., arranged into *vendor islands*). Such small- and mid-sized networks, in terms of number of nodes, were relatively homogeneous, thus reducing interoperability issues. A single, vendor-specific Network Management System (NMS) was deployed, being responsible for the management of the optical network, tailored to the underlying hardware, and using proprietary interfaces and extensions.

Those systems were perceived as closed, bundled together as a whole, and with a limited set of functionalities that were dependent on a given release. The provisioning of a network connectivity service involved manual processes, where a service activation or modification could involve human intervention, with a user requesting the service provider, which was then manually planning and configuring the route and resources in the network to support the service.

Several challenges motivated the evolution towards the control plane. First, network operators continuously have specific requirements to reduce operational costs, while ensuring that the network still meets the requirements of the supported services. Second, the manual, long-lasting processes associated to NMS-based networks did not seem adapted for the dynamic provisioning of services with recovery and Quality of Service (QoS). In short, the introduction of a dynamic control plane was justified, from an operational perspective, for the automation of certain tasks, freeing the operator from the burden of manually managing and configuring individual nodes, leading to significant cost reductions.

In this context, the introduction of a control plane aims at fulfilling the requirements of fast and automatic end-to-end provisioning and rerouting of flexi-grid connections, while supporting different levels of quality of service. Regardless, of the actual technology, a control plane needs to address common functions like addressing, automatic topology discovery, network abstraction, path computation, and connection provisioning, as stated earlier in this chapter. From a high level perspective, and as any software system that automates tasks and processes, the functions of a control plane can, from a simplistic point of view, be distributed or centralized, although we will later see that this separation is becoming blurry. This dichotomy applies not only from a functional perspective but also from a resource allocation perspective. Both models are viable; both have their own strengths and weaknesses, and both are being extended to address the new requirements associated to the aforementioned emerging optical technologies, such as flexible spectrum allocation, efficient co-routed connection setup, and configuration of related optical parameters. Thus, the selection of a centralized or distributed control plane is conditioned by diverse aspects, such as the desired functions, flexibility and extensibility, availability, etc., as well as by more concrete aspects such as the inherent constraints of the optical technology (e.g., the need to account for physical impairments which are collected from monitoring systems and not standardized), already installed deployments, and actual network size and scalability.

The network elements participating in distributed control plane environment exchange the accumulated advertisements from other nodes in a state database (e.g., OSPF database) and run a Dijkstra (shortest path) algorithm to establish a reachability graph of best paths to destinations. This process uses a distributed flooding algorithm within the IGP protocol procedure to propagate attachment information, thus, all nodes speaking a particular IGP protocol in the domain remain connected to each other (directly or indirectly) and participate with timely reachability information and establish a network topology that reports change in connectivity in the event of failure. A key aspect is thus convergence, which is the time it takes when a network element introduces a change in reachability of a destination due to a network. A variety of methods exist in various IGP mechanisms and procedures to address scaling of the control plane state (memory and CPU) in the network, both for physical and logical design. These methods include summarization, filtering, recursion, and segregation.

### *10.3.3 Control Elements for Operating Optical Networks*

#### Path Computation

Path computation manages aspects related to finding a physical route between two network nodes, commonly referred to as endpoints. Path computation is a functional component of a control plane, invoked for the purposes of (dynamic) provisioning, rerouting, restoration, as well as advanced use-cases such as overall optimization, adaptive network planning or, in the particular case of DWDM flexi-grid networks, spectrum de-fragmentation.

#### Service Provisioning

This would include the node and interface configuration, specifically known as service provisioning, the setup and teardown of connections. The control element would automatically configure the required hops between the source and destination nodes required to create a connection between two (or point to multipoint) points in the network. The procedure and protocols used via the controller to configure different elements to set up a connection is known as either distribute via the signaling mechanisms available (such as RSVP-TE) or direct using a flow provision process (such as OpenFlow).

#### OAM and Performance Monitoring

Operations, Administration, and Maintenance (OAM) is often used as a general term to describe a collection of tools for fault detection and isolation, and for performance measurement. Many OAM tools and capabilities have been defined for various technology layers [4].

OAM tools may, and quite often do, work in conjunction with a control plane and management plane. OAM provides instrumentation tools for measuring and monitoring the data plane. OAM tools often use control-plane functions, e.g., to initialize OAM sessions and to exchange various parameters. The OAM tools communicate with the management plane to raise alarms, and often OAM tools may be activated by the management plane (as well as by the control plane), e.g., to locate and localize problems, and initiate performance measurement of an optical segment, or end-to-end service.

## 10.4 Distributed and Centralized Control Planes

### *10.4.1 Control Plane Architecture Evolution*

In their early deployments, optical transport networks were inherently managed, deployed in a single administrative domain, and locked to a single vendor hardware solution (i.e., arranged into *vendor islands*). Such small- and mid-sized networks, in terms of number of nodes, were relatively homogeneous, thus reducing interoperability issues. A single, vendor-specific Network Management System (NMS) was deployed, being responsible for the management of the optical network, tailored to the underlying hardware, and using proprietary interfaces and extensions.

For example, the Internet represents an example of a significant scaling problem. Vast numbers of administrative regions are loosely tied with the interconnections changing constantly as traffic patterns fluctuate and failures occur. To address this, the Internet control paradigm was designed to be distributed. On the other hand, SDH/Optical core transport networks, while geographically spanning national or continental regions, are still relatively small in size/number of elements when compared to IP networks, and are commonly under the control of a single entity or operator. Services offered were relatively stable, characterized by long holding times, coupled to slow traffic dynamics, and service provisioning delays of the order of days/ weeks was acceptable. Such deployments models were, arguably, best addressed with a centralized control paradigm.

While the need of a control plane does not seem to present significant opposition, the choice of the technology is still debatable. From a historical perspective, the evolution of the control plane for optical networks started augmenting NMS-based networks with a distributed control plane, based on the ASON (Automatically Switched Optical Networks) [5–7] architecture with Generalized Multi-Protocol Label Switching GMPLS [8] suite of protocols, as detailed next. Recently, the application of Software Defined Networking (SDN) principles to the control of optical networks is presented as a means to enable the programmability of the underlying network (in any case, the formal separation of the data and control planes is a key concept in optical network control). To some extent, there is an analogy between a Transport SDN architecture and a centralized NMS, although the former insists on using modern system architectures, open and standard interfaces, and flexible and modular software development.

### Distributed Control

In this setting, the control plane is implemented by a set of cooperating entities (control plane controllers) that execute processes that communicate. Control plane functions such as topology management, path computation, or signaling are distributed (for the first one, each node disseminates the topological elements that are directly under its control, and the IGP routing protocol enables the construction of a unified view of the network topology. Path computation is carried out by the ingress node of the connection and signaling is distributed along the nodes involved in the path). The protocols ensure the coordination and synchronization functions, autonomously (although commonly, the provisioning of a new service is done upon request from a NMS).

The reference architecture is defined by the ITU-T, named ASON enabling dynamic control of an optical network, automating the resource and connection management. ASON relies on the GMPLS set of protocols defined by the IETF (with minor variations). In short, the ASON/GMPLS architecture defines the transport, control, and management planes. In particular, the control plane is responsible for the actual resource and connection control, and consists of Optical Connection Controllers (OCC), interconnected via Network to Network Interfaces (NNIs) for network topology and resource discovery, routing, signaling, and connection setup and release (with recovery). The Management Plane is responsible for managing and configuring the control plane and fault management, performance management, accounting, and security.

As seen in Fig. 10.3, the main involved processes are the Connection Controller (CC) and the Routing Controller (RC), and optionally a path computation component. A data communication network, based on IP control channels (IPCC) to allow the exchange of control messages between GMPLS controllers, is also required, which can be deployed in-band or out-of-band (including, e.g., a dedicated and separated physical network). A GMPLS-enabled node (both control and hardware) is named Label Switched Router (LSR). Each GMPLS controller manages the state of all the connections (i.e., Label Switched Path—LSPs) originated, terminated, or passing through a node, stored in the LSP Database (LSPDB), and maintains its own network state information (topology and resources), collected in a local Traffic Engineering Database (TED) repository.

The network elements participating in distributed control plane environment exchange the accumulated advertisements from other nodes in a state database (e.g., OSPF database) and run a Dijkstra (shortest path) algorithm to establish a reachability graph of best paths to destinations. This process uses a distributed flooding algorithm within the IGP protocol procedure to propagate attachment information; thus, all nodes speaking a particular IGP protocol in the domain remain connected to each other (directly or indirectly) and participate with timely reachability information and establish a network topology that reports change in connectivity in the event of failure. A key aspect is thus convergence, which is the time it takes from when a network element introduces a change in reachability of a destination due to a network change, such as a failure. A variety of methods exist in various IGP

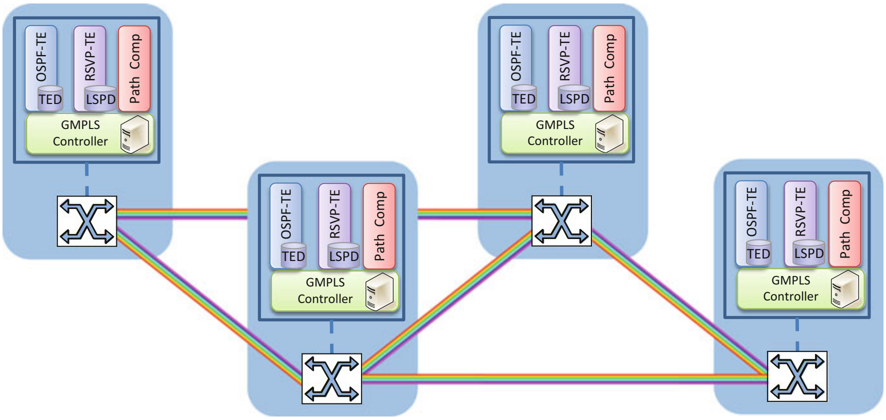



**Fig. 10.3** Example of GMPLS-controlled optical network

mechanisms and procedures to address scaling of the control plane state (memory and CPU) in the network, both for physical and logical design. These tools include summarization, filtering, recursion, and segregation.

### Centralized Control

In a centralized control, a single entity, usually called controller, is responsible for the control plane functions, commonly using open and standard protocols, such as those defined by the SDN architectures and protocols, e.g., OpenFlow protocol (OF/OFP) [9]. The controller performs path computation and service provisioning, and proceeds to configure the forwarding and switching behavior of the nodes. A centralized control plane provides a method for programmatic control of network resources and simplification of control plane process. Deployment and operation of connections requires an interaction with control points to establish the forwarding rules for specific traffic. These are not recent innovations; separation of the control and data planes occurred with the development of ForCES [10] and Generalized Switch Management Protocol (GSMP) [11] many years ago.

By deploying the control plane intelligence in the controller, resources allocated in hardware nodes for CP functions are reduced significantly. Moreover, such solutions involve deploying hardware (computational and storage) in a centralized location which is orders of magnitude more powerful than individual controllers are. Although a centralized controller does not seem significantly different from an NMS, it is worth noting the aspects such as the automation of processes, and programmability, as well as the use of open interfaces and standard architectures, terminology, models and protocols. Note that a logically centralized controller may, itself, be implemented as a distributed system, while appearing, programmatically, as a single entity. Finally, SDN principles bring new opportunities such as joint allocation of IT and network resources, or the orchestration of heterogeneous control technologies, or the unified control of access and core network segments.

### Comparison of Distributed Versus Centralized

In a distributed control approach, individual nodes participate together to distribute reachability information in order to develop a localized view of a consistent, loop-free network. Routes and reachability information are exchanged that later result in data plane paths being programmed to realize those paths; however, paths are often suboptimal and prone to traffic congestion, so clearly this approach has weaknesses which might be addressed using a centralized approach. Mainly, a distributed control plane is affected by the latencies in the propagation and synchronization of data. Changes occurring at a given network element need to be propagated and the transitory may affect network performance.

On the other hand, in a distributed model, each node element is mainly self-sustained. There is no bottleneck or single point of failure, such as SDN controller, and this model seems most appropriate when there is no central authority and functional elements need to cooperate. Each node can survive failures at other nodes as long as the network remains connected.

The benefits of a centralized model are lower capital and operational cost, involving, in the case of a control plane, minimal control plane hardware and software at each node, while enabling computational scaling at the controller location. A centralized controller may be easier to implement, given the tight coupling of components and the less stringent requirements of internal interfaces not subject to interoperability issues. It simplifies automation and management, enables network programmability, and is less subject to latencies and out-of-date information due to the need for synchronizing entities. It provides more flexibility, a single point of extension for operators' policies and customizations, and improved security. There is less control plane overhead, and arguably, network security is increased, with less complexity and greater control over potential risk areas. The downside is that centralized elements are always points of failure.

### Hybrid Control Plane Models

In view of the current trends and evolutions of control plane architectures, it seems too simplistic to tag a control plane as distributed or centralized. Control plane architectures are evolving towards hybrid control-plane models, in which some elements may be centralized and some elements may be distributed, sometimes following the mantra "distribute when you can, centralize when you must." Even if a given control plane entity is centralized, it can be logically centralized, where a system is implemented in terms of the composition of functional components that appear as one. A given function can be centralized in a given domain (e.g., the path computation function can be centralized in a Path Computation Element (PCE) assuming a single PCE per domain deployment model), but the same function can

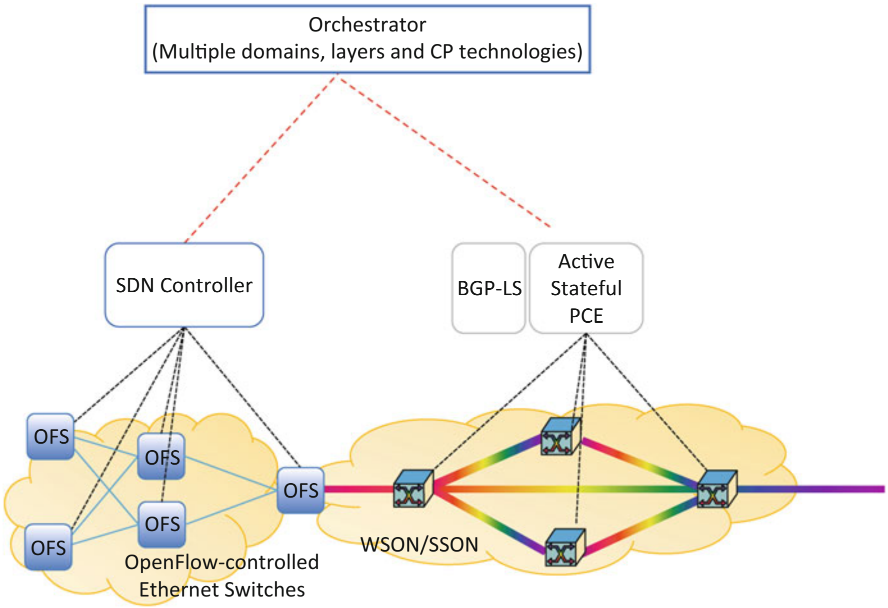



**Fig. 10.4** The use of an orchestrator for the over-arching control of heterogeneous control technologies

be distributed among several children PCE in Hierarchical PCE (H-PCE) architecture [12] within a multi-domain scenario.

New use-cases, such as remote data center interconnection, highlight the need for multi-domain service provisioning and heterogeneous CP interworking, potentially requiring an overarching control (see Fig. 10.4). Additionally, network operators aim at addressing the joint control and allocation of network and IT resources (e.g., networking, computing, and storage resources), or the joint optimization of different network segments, such as access, aggregation, and core. Different alternatives, with varying degrees of integration and flexibility, are available: straightforward approaches characterized by the adaptation of one control model to the other or more advanced interworking requiring the definition of common models (e.g., a subset of attributes for network elements) and of coordination and orchestration functions. Such orchestrator may in turn, be (logically or physically) centralized while delegating specific functions, to subsystems that may be distributed (such as the provisioning of connectivity delegated to a GMPLS control plane) [8].

Finally, let us mention that the adoption of new computing and interworking models, and concepts, such as those of server consolidation, host virtualization or Network Function Virtualization (NFV), are challenging common approaches and existing practice: for example, a GMPLS control plane could be run as a Virtual Network Function running in a datacenter, for legacy purposes, in which a distributed system could run on a centralized physical infrastructure.

## 10.5 Framework for Application-Based Network Operations

The three tenants of SDN are programmability, the separation of the control and data planes, and the management of ephemeral network state in a centralized control model [1], regardless of the degree of centralization. In an ideal world, it should be possible to utilize a distributed control plane as well, providing the best practices of centralized control and distributed control plane for ephemeral state management.

Application-Based Network Operations (ABNO) was designed using the following architectural principles:

1. Loose Coupling: For ease of implementation and fast development, we do not attempt to tightly integrate the functional components of the network controller. Instead, we use well-defined APIs and protocol mechanisms.
2. Low Overhead: The goal is to ensure that each management and control function is not duplicated, which reduces the overall platform overhead.
3. Modular: A modular design enables easier composition of existing features into new capabilities.
4. Intelligent: Designing the framework around the Path Computation Element and Traffic Engineered principles provides significant benefits for controlling a range of network technologies and maximizing resource utilization.
5. Resource Management: The framework allows for various network and node state to be discovered and stored. This state information is collected using the protocol mechanisms provided by traditional and already existing network and service management tools.
6. Dynamic Management: A key goal of an SDN controller is actual dynamic control based on application demands and other network events.
7. Policy Control: It is important to implement policy management to provide the mechanisms for specifying connection requirements (e.g., QoS, security) for various applications. It also allows network operators to associate different service levels.
8. Technology Agnostic: The ABNO framework communicates with the network nodes using a variety of Southbound APIs and protocols, allowing for a wide variety of forwarding mechanisms to be managed using ABNO.

Figure 10.5 presents an example of network architecture using ABNO.

### *10.5.1 Functional Components*

#### NMS and OSS

A Network Management System (NMS) or an Operations Support System (OSS) can be used to control, operate, and manage a network. Within the ABNO framework, an NMS or OSS may issue high-level service requests to the ABNO Controller.

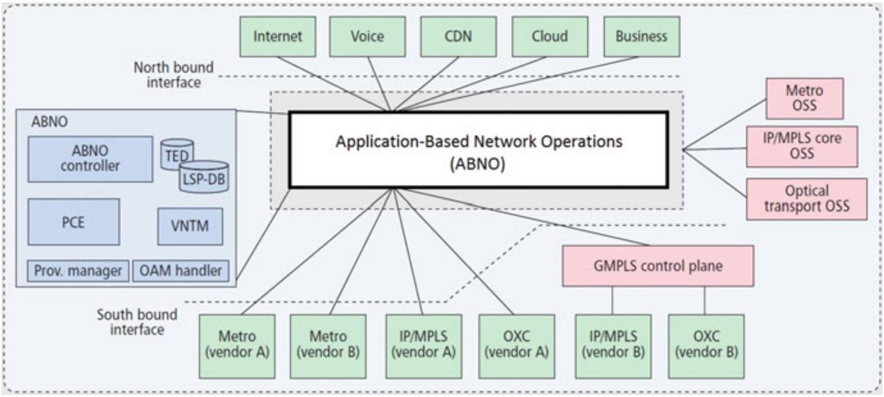



**Fig. 10.5** ABNO architecture example

It may also establish policies for the activities of the components within the architecture.

The NMS and OSS can be consumers of network events reported through the OAM Handler and can act on these reports as well as displaying them to users and raising alarms. The NMS and OSS can also access the Traffic Engineering Database (TED) [13] and Label Switched Path Database (LSP-DB) to show the users the current state of the network.

Lastly, the NMS and OSS may utilize a direct programmatic or configuration interface to interact with the network nodes within the network.

## Application Service Coordinator

The Application Service Coordinator communicates with the ABNO Controller to request operations on the network. Requests may be initiated from entities such as the NMS and OSS, services in the ABNO architecture may be requested by or on behalf of applications. In this context, the term "application" is very broad. An application may be a program that runs on a host or server and that provides services to a user, such as a video conferencing application. Alternatively, an application may be a software tool that a user uses to make requests to the network to set up specific services such as end-to-end connections or scheduled bandwidth reservations. Finally, an application may be a sophisticated control system that is responsible for arranging the provision of a more complex network service such as a VPN. For the sake of ABNO architecture discussion, all of these concepts of an application are grouped together and are shown as the Application Service Coordinator, since they are all in some way responsible for coordinating the activity of the network to provide services for use by applications. In practice, the function of the Application Service Coordinator may be distributed across multiple applications or servers.

### ABNO Controller

The ABNO Controlleris the main gateway to the network for the NMS, OSS, and Application Service Coordinator for the provision of advanced network coordination and functions. The ABNO Controller governs the behavior of the network in response to changing network conditions and in accordance with application network requirements and policies. It is the point of attachment and invokes the right components in the right order.

### Policy Agent

Policy plays a very important role in the control and management of the network. It is, therefore, significant in influencing how the key components of the ABNO architecture operate. The Policy Agent is responsible for propagating those policies into the other components of the system. Simplicity in this discussion necessitates leaving out many of the policy interactions that will take place. In our example, the Policy Agent is only discussed interacting with the ABNO Controller; in reality, it will also interact with a number of other components and the network elements themselves. For example, the Path Computation Element (PCE) will be a Policy Enforcement Point (PEP) [14], and the Interface to the Routing System (I2RS) Client will also be a PEP as noted in [15].

### OAM Handler

Operations, Administration, and Maintenance (OAM) plays a critical role in understanding how a network is operating, detecting faults, sand taking the necessary action to react to problems in the network. Within the ABNO architecture, the OAM Handler is responsible for receiving notifications (often-called alerts) from the network about potential problems, for correlating them, and for triggering other components of the system to take action to preserve or recover the services that were established by the ABNO Controller. The OAM Handler also reports network problems and, in particular, service-affecting problems to the NMS, OSS, and Application Service Coordinator. Additionally, the OAM Handler interacts with the devices in the network to initiate OAM actions within the data plane [4], such as monitoring and testing.

### Path Computation Element

The Path Computation Element (PCE) is a functional component that services request to compute paths across a network graph. In particular, it can generate traffic-engineered routes for MPLS-TE and GMPLS Label Switched Paths (LSPs). The PCE may receive these requests from the ABNO Controller, from the Virtual Network Topology Manager (VNTM), or from network elements themselves.

The PCE operates on a view of the network topology stored in the Traffic Engineering Database (TED). A more sophisticated computation may be provided by a Stateful PCE that enhances the TED with a database (the LSP) containing information about the LSPs that are provisioned and operational within the network.

Additional functionality in an Active PCE allows a functional component that includes a Stateful PCE to make provisioning requests to set up new services or to modify in-place services as described in [16]. This function may directly access the network elements or channelled through the Provisioning Manager. Coordination between multiple PCEs operating on different TEDs can prove useful for performing path computation in multi-domain or multilayer networks. A domain in this case might be an Autonomous System (AS), thus enabling inter-AS path computation.

In the latter case, the ABNO controller will need to request an optimal path for the service. If the domains (ASes) require path setup to preserve confidentiality about their internal topologies and capabilities, they will not share a TED and subsequently each domain (AS) will operate its own PCE. In such a situation, the Hierarchical PCE (H-PCE) architecture, described in [12], is necessary.

### Network Database

The ABNO architecture includes a number of databases that contain information stored for use by the system. The two main databases are the TED and the LSP Database (LSP-DB), but there may be a number of other databases used to contain information about topology (ALTO Server), policy (Policy Agent), services (ABNO Controller), etc.

Typically, the IGP (like OSPF-TE or IS-IS-TE) is responsible for generating and disseminating the TED within a domain. In multi-domain environments, it may be necessary to export the TED to another control element, such as a PCE, which can perform more complex path computation and optimization tasks.

### Virtual Network Topology Manager

A Virtual Network Topology (VNT) is defined as a set of one or more LSPs in one or more lower-layer networks that provide information for efficient path handling in an upper-layer network. For instance, a set of LSPs in a wavelength division multiplexed (WDM) network can provide connectivity as virtual links in a higher-layer packet switched network.

The creation of virtual topology for inclusion in a network is not a simple task. Decisions must be made about which nodes in the upper layer it is best to connect, in which lower-layer network to provision LSPs to provide the connectivity, and how to route the LSPs.

**Provisioning Manager**

The Provisioning Manager is responsible for making or channeling requests for the establishment of LSPs. This may be instructions to the control plane running in the networks, or may involve the programming of individual network nodes.

### *10.5.2 South Bound Interfaces*

The network devices maybe configured or programmed directly from the NMS/OSS. Many protocols already exist to perform these functions, including the following:

- SNMP [17]
- The Network Configuration Protocol (NETCONF) [18, 19]
- RESTCONF [20]
- ForCES [10]
- OpenFlow [9]
- PCEP [21]

The role of the protocols described is to assign state to the forwarding element, either by programming each node individually or via a distributed signaling mechanism. Indeed the previous list is not an exhaustive representation of protocol methods and procedures available, and over time, new forwarding mechanisms will be developed. Therefore, the ABNO framework has been designed to be forwarding mechanism agnostic.

## 10.6 Adaptive Network Manager

The European Commission-funded project "IDEALIST" identified the need for a control architecture to combine the best of distributed routing and signaling protocols, to provide real-time adaption and to survive against failures, and a centralized intelligence that, on the one hand, provides a point for optimization (e.g., interfacing with the planning tool), and also capable of interfacing with the higher applications, including cloud platforms and data center (WAN) inter-connections.

The distributed functions are based on the well-known GMPLS architecture, while the centralized intelligence and interface with applications follows a SDN approach. Thus, the ANM is the IDEALIST network controller (based on the ABNO framework) [22] that considers not only the Flexi-grid Network (the main focus of IDEALIST) but also a wider scope, a multilayer IP/MPLS over optical Network.

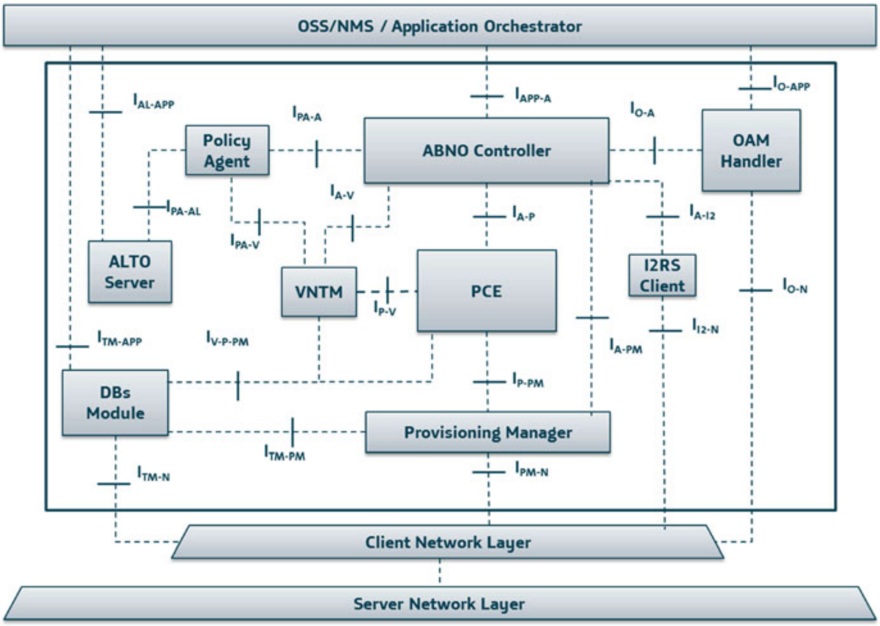



**Fig. 10.6** Adaptive network manager functional components and interfaces

## 10.6.1 Interfaces

As the ABNO architecture was generic in its intent, most of the interfaces are defined as concepts. In ANM architecture, HTTP/JSON interfaces will be used in these interfaces not already defined (Fig. 10.6). There are two reasons: easy development and flexibility for the workflows definition. These interfaces will help to have a modular design, which can be adapted to the future requirements that may come during the project. If during the project, there are some other solutions in the standardization fora, this have been assessed and where applicable, included in the ANM architecture.

- *IN-APP*—This is the interface between the application layer/NMS/OSS and the ABNO controller. Application layer makes requests to set up connections or to trigger any other workflow using HTTP/JSON. This interface is currently under development in the Internet Engineering Task Force (IETF). The parameters of the request change depending on the workflow, but the operation type is always mandatory.
- *IAL-APP*—This is the interface between the ALTO Server and Application layer/NMS/OSS, where the Application layer acts as an ALTO Client. They communicate using the ALTO Protocol [23]. They communicate over HTTP/JSON. An

information model has to be defined for this interface to support TED, LSPs, and inventory requests.

- *IA-I2*, *II2-N*—The Interface to the Routing System (I2RS).
- *IPA-A*, *IPA-V*, *IPA-AL*—All the interfaces between the Policy Agent and the modules that request it for permission using a HTTP/JSON request.
- *IA-P*—This is the interface between the ABNO controller and the PCE. The ABNO controller queries the PCE using PCE; Stateless and Stateful PCEs may be used' this interface will support requests for both PCEs.
- *IA-V*—This interface connects the ABNO controller and the VNTM. They communicate through PCEP.

## 10.7 Adaptive Network Manager Use-Cases

### *10.7.1 Catastrophic Network Failure*

While most networks are designed to survive single failures without affecting customer service level agreements (SLAs), they are not designed to survive large-scale disasters, such as earthquakes, floods, wars, or terrorist acts, simply because of their low failure probability and the high cost of overprovisioning to address such events in today's network.

Since many systems might be affected, large network reconfigurations are necessary during large-scale disaster recovery. The disaster recovery process is similar to that of the virtual topology reconfiguration after a failure. However, multiple optical systems, IP links, and possible routers and OXCs (assuming central offices are affected) may be taken offline during the disaster. Several additional planning and operation requirements in response to large-scale disasters are highlighted below:

- Consideration of potential IP layer traffic distribution changes, either using MPLS-TE tunnels or by modification of IP routing metrics, and evaluating benefits based on the candidate topology.
- It may be impossible to reach the desired network end state with one-step optimization. Therefore, two or more step optimizations may be necessary, for example, to reroute some other optical connections to make room for some new connections.
- The system must verify that the intermediate configuration after each such step is robust and can support the current traffic and possibly withstand additional outages.
- Based on preemption and traffic priorities, it might be desirable to disconnect some virtual links so as to reuse the resources for post-disaster priority connections and traffic.

We have described the creation of one disaster recovery plan, but in a real network, there may be several possible plans, each with its pros and cons. The tool must present all these plans to the operator so that the operator can select the best plan, and possibly modify it and understand how it will be behave.

To summarize, the above process consists of several steps:

1. Immediate action by the network to recover some of the traffic
2. Dissemination of the new network state
3. Root cause analysis to understand what failed and why
4. An operator-assisted planning process to come up with a disaster recovery plan
5. Execution of the plan, possibly in multiple steps
6. Reconvergence of the network after each step and in its final state

This scenario for recovering from catastrophic network failures may also be known as "In-Operation Network Planning" [24]. The ANM platform and use-cases are also discussed in-depth in the next chapter.

## 10.8 Next Steps for ABNO-Based Control and Orchestration

We can assume that SDN is well-defined as a logically centralized control framework and architecture. It supports the programmability of network functions and protocols by decoupling the data plane from the control plane through a well-defined control South Bound Interface (SBI) protocol. These SBIs exist in many forms, and assist in the hiding of technology or vendor-specific forwarding mechanisms. As network evolution continues, a new technology area known as "Network Functions Virtualization" (NFV) [25] is developing in parallel to SDN.

The development of NFV is to leverage Information Technology (IT) virtualization techniques to migrate entire classes of network functions typically hosted on proprietary hardware onto virtual platforms based on general compute and storage servers. Each virtual function node is known as a Virtualized Network Function (VNF), which may run on a single or set of Virtual Machines (VMs), instead of having custom hardware appliances for the proposed network function.

Furthermore, this virtualization allows multiple isolated VNFs or unused resources to be allocated to other VNF-based applications during weekdays and business hours, facilitating overall IT capacity to be shared by all content delivery components, or even other network function appliances. Industry, via the European Telecommunications Standards Institute (ETSI), has defined a suitable architectural framework [25], and has also documented a number of resiliency requirements and specific objectives for virtualized media infrastructures.

Utilizing the benefits of enabling technologies (i.e., ABNO-based control principles and NFV-based infrastructure), we have the potential to fundamentally change the way we build, deploy, and control broadcast services built on top of flexible optical networks allowing dynamic and elastic delivery and high-bandwidth broadcast and media resources.

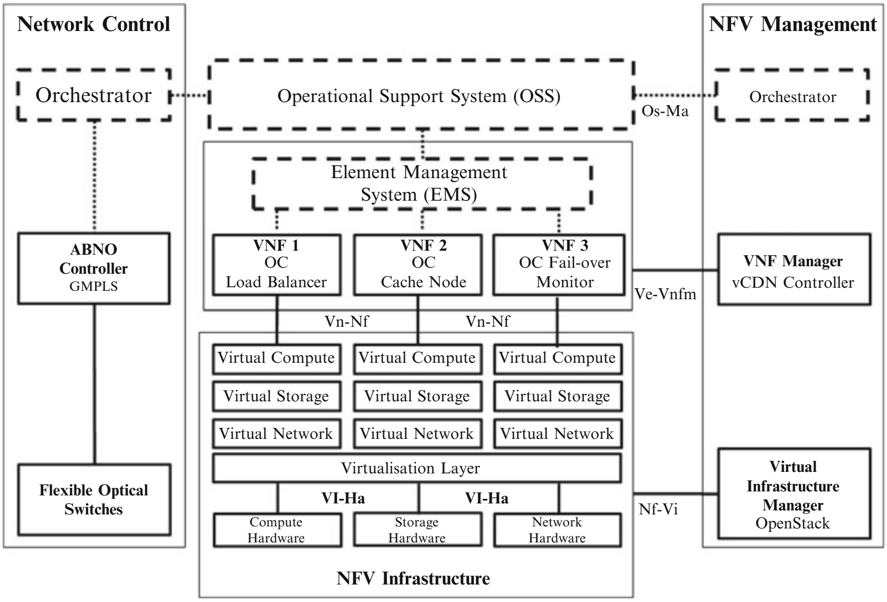



**Fig. 10.7** Candidate SDN & NFV framework based on ETSI NFV ISG model

### *10.8.1 Control and Orchestration of Virtual Content Distribution Network*

Virtualization of Content Distribution Networks (CDNs) components is a core design principle necessary to create a content network that can be deployed rapidly and in a scalable way. The first element to be virtualized is the cache node itself, and then required services such as content monitors and load balancers [26]. A key requirement of the Virtual Content Distribution Network (vCDN) is reconfigurable bandwidth as content moved from HD content at 1080p to 4k streams demands change based on time of day and week [27]. Deploying the various infrastructure elements of a CDN as a collection of virtual appliances (VNFs) and connecting content and access (user networks) with a flexible optical network infrastructure offers significant benefits.

Figure 10.7 describes how an ABO-enabled network controller would integrate with an NFV-based CDN.

Using the ABNO-based controller in conjunction with the NFV Management and Infrastructure itself would provide the VNFs connectivity over a high-bitrate optical infrastructure, and similar flexibility that exists in the IP and Ethernet layer, which until recently and the advent of EONs, simply not previously available in optical transport domain.

## References


1. D. King, A. Farrel, *A PCE-Based Architecture for Application-Based Network Operations*, IETF Internet RFC 7491, (2015, March)
2. R. Muñoz, et al., IDEALIST control and service management solutions for dynamic and adaptive flexi-grid DWDM networks, in *Proceedings of Future Network and Mobile Summit*, Lisbon, 3–5 July 2013
3. Ó. González de Dios, R. Casellas, *Framework and Requirements for GMPLS Based Control of Flexi-grid DWDM Networks*, RFC 7698, (2015, December)
4. N. Sprecher et al., *An Overview of Operations, Administration, and Maintenance (OAM) Tools*, RFC 7276, (2014, June)
5. ITU-T Recommendation G.8080/Y.1304, *Architecture for the Automatically Switched Optical Network (ASON) 02/2012*
6. ITU-T Recommendation G.872, *Architecture of Optical Transport Networks 10/2012*
7. ITU-T Recommendation G.709/Y.1331, *Interface for the Optical Transport Network* (OTN) 02/2012
8. E. Mannie (ed.), *Generalized Multi-Protocol Label Switching (GMPLS) Architecture*, IETF RFC 3945, (2004, October)
9. Open Networking Foundation, *OpenFlow Switch Specification Version 1.4.0 (Wire Protocol 0x05)*, (2013, October)
10. J. Halpern, J. Hadi Salim, *Forwarding and Control Element Separation (ForCES) Forwarding Element Model*, RFC 5812, (2010, March)
11. A. Doria, K. Sundell, F. Hellstrand, T. Worster, *General Switch Management Protocol (GSMP) V3*, RFC 3292, (2002, June)
12. D. King, A. Farrel (eds.), *The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS*, RFC 6805, (2012, November)
13. O. Dugeon, et al., *Path Computation Element (PCE) Database Requirements*, IETF Internet Draft draft-dugeon-pce-ted-reqs-03, (2014, February)
14. I. Bryskin, et al., *Policy-Enabled Path Computation Framework*, RFC 5394, (2008, December)
15. A. Atlas, T. Nadeau, D. Ward (eds.), *Interface to the Routing System Problem Statement*, draft-ietf-i2rs-problem-statement (2015, March)
16. E. Crabbe, I. Minei, S. Sivabalan, R. Varga, *PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model*, draft-ietf-pce-pce-initiated-lsp, (2015, October)
17. J. Case, D. Harrington, R. Presuhn, B. Wijnen, *Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)*, STD 62, RFC 3412, (2002, December)
18. R. Enns, et al., *Network Configuration Protocol (NETCONF)*, RFC 6241, (2011, June)
19. M. Bjorklund, (ed.), *YANG—A Data Modeling Language for the Network Configuration Protocol (NETCONF)*, IETF Request or Comments 6020, (2010, October)
20. A. Bierman, M. Bjorklund, K. Watsen, *RESTCONF Protocol*, draft-ietf-netconf- restconf, (2015, July)
21. J.P. Vasseur, J.L. Le Roux (eds.), *Path Computation Element (PCE) Communication Protocol (PCEP)*, RFC 5440 (2009, March)
22. A. Aguado, et al., ABNO: a Feasible SDN Approach for Multi-Vendor IP and Optical Networks, in *OFC Conference*, Th3I.5 (2014, March)
23. J. Seedorf, E. Burger, *Application-Layer Traffic Optimization (ALTO) Problem Statement*, RFC 5693, (2009, October)
24. L. Velasco, D. King, O. Gerstel, R. Casellas, A. Castro, V. López, In-operation network planning. IEEE Commun. Mag. **52**(1), 52–60 (2014)
25. ETSI GS NFV 002. *Network Functions Virtualization (NFV); Architectural Framework* (2014)
26. ETSI GS NFV 001. *Network Functions Virtualization (NFV); Use Cases* (2013)
27. M. Broadbent, D. King, S. Baildon, N. Georgalas, N. Race, OpenCache: a software-defined content caching platform, in *1st IEEE Conference on Network Softwarization (NetSoft)*, London (2015, April)

# Chapter 11
# In-Operation Network Planning

**Ramon Casellas, Alessio Giorgetti, Lluís Gifre, Luis Velasco, Víctor López, Oscar González, and Daniel King**

## Acronyms

| | |
|---|---|
| ABNO | Application-based Network Operations |
| AFRO | After Failure repair Re-optimization Problem |
| BER | Bit Error Rate |
| BGP | Border Gateway Protocol |
| BGP-LS | BGP Link-State |
| b-PCE | Back-end PCE |
| BV-WSS | Bandwidth Variable Wavelength Selective Switching |
| ERO | Explicit Route Object |
| FEC | Forward Error Correction |
| f-PCE | Front-end PCE |
| FRR | Fast Re-Route |
| GCO | Global Concurrent Optimization |
| GMPLS | Generalized Multi-protocol Label Switching |
| IP/MPLS | Multi-protocol Label Switching over Internet Protocol |
| IRO | Include Route Object |

R. Casellas (✉)
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain
e-mail: ramon.casellas@cttc.es

A. Giorgetti
Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), Pisa, Italy

L. Gifre • L. Velasco
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

V. López • O. González
Telefónica, Madrid, Spain

D. King
Old Dog Consulting, Lancaster, United Kingdom



V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7_11

| | |
|---|---|
| LSP | Label-switched Path |
| LSP-DB | LSP Database |
| MP-AFRO | Multipath AFRO Problem |
| NMS | Network Management System |
| OF | Objective Function Object |
| OOK | On–Off Keying |
| OSS | Operations Support System |
| OXC | Optical cross-connect |
| PCE | Path Computation Element |
| PCEP | PCE Protocol |
| PCInit | Path Computation Initiate |
| PCRep | Path Computation Reply |
| PCReq | Path Computation Request |
| PCRpt | Path Computation Report |
| PCUpd | Path Computation Update |
| PM-QPSK | Polarization Multiplexed Quadrature Phase Shift Keying |
| QoS | Quality of Service |
| RP | Request Parameters Object |
| RRO | Reported Route Object |
| SPRESSO | Spectrum Re-allocation Problem |
| SVEC | Synchronization Vector Object |
| TE | Traffic Engineering |
| TED | TE Database |
| TLV | Type-Length-Value |
| TX | Transmitter |
| VNT | Virtual Network Topology |
| VNTM | VNT Manager |
| WSON | Wavelength-Switched Optical Networks |

Current transport networks are statically configured and managed, because they experience a rather limited traffic dynamicity. As a result, long planning cycles are used to upgrade the network and prepare it for the next planning period. Aimed at guaranteeing that the network can support the forecast traffic and deal with failure scenarios, spare capacity is usually installed, thus increasing network expenditures.

Moreover, results from network capacity planning are manually deployed in the network, which limits the network agility. In this chapter, we propose a control and management architecture to allow the network to be dynamically operated. Employing those dynamicity capabilities, the network can be reconfigured and re-optimized in response to traffic changes in an automatic fashion; hence, the resource overprovisioning can be minimized and overall network costs reduced.

# 11.1 Towards In-operation Network Planning

## *11.1.1 Drivers and Motivations for In-operation Network Planning*

The classical network planning life-cycle typically consists of several steps that are performed sequentially (Fig. 11.1). The initial step receives inputs from the service layer and from the state of the resources in the already deployed network and configures the network to be capable of dealing with the forecast traffic, for a period of time. That period is not fixed and actual time length usually depends on many factors, which are operator- and traffic-type-specific. Once the planning phase produces recommendations, the next step is to design, verify, and manually implement the network changes. While in operation, the network capacity is continuously monitored and that data is used as input for the next planning cycle. In case of unexpected increases in demand or network changes, nonetheless, the planning process may be restarted.

However, operation of carriers' transport networks is very complex; multiple manual configuration actions are needed for provisioning purposes (e.g., hundreds of thousands of node configurations per year in a mid-size network). In fact, transport networks are currently configured with big static *fat pipes* based on capacity overprovisioning, since they are needed for guaranteeing traffic demand and QoS. Furthermore, network solutions from different vendors typically include a centralized service provisioning platform, using vendor-specific NMS implementations along with an operator-tailored umbrella provisioning system, which may include a technology-specific Operations Support System (OSS). Such complicated

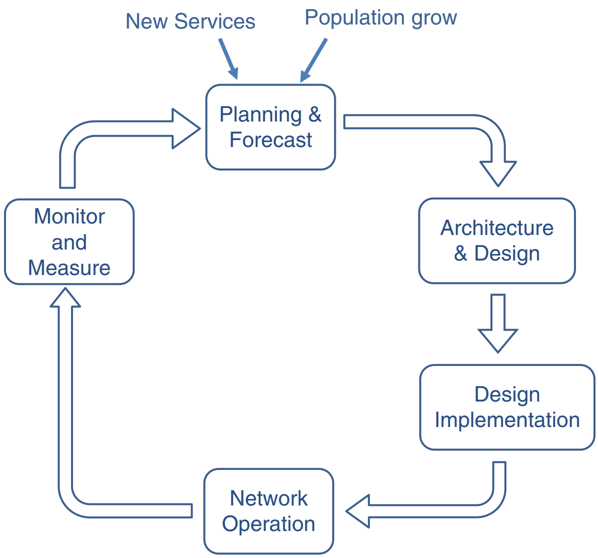



**Fig. 11.1** Classical networks life-cycle

architectures (Fig. 11.2) generate complex and long workflows for network provisioning: up to 2 weeks for customer service provisioning and more than 6 weeks for core routers connectivity services over the optical core.

Figure 11.3 illustrates the fact that such static networks are designed to cope with the requirements of several failure scenarios, and predicted short-term increases in bandwidth usage, thus requiring capacity overprovisioning and significantly increasing CAPEX. It shows a simple network consisting in three routers connected to a central one through a set of lightpaths established on an optical network. Two different scenarios are considered, although the same amount of IP traffic is conveyed in each of them. In the scenario A, router R3 needs three lightpaths to be established to transport its IP traffic towards R4, whereas R1 and R2 need only one lightpath each. In contrast in the scenario B, R1 and R2 need two lightpaths whilst R3 needs only one lightpath. In static networks, where lightpaths in the optical network are statically established, each pair of routers has to be equipped with the number of interfaces for the worst case, resulting in R4 being equipped with seven interfaces. However, if the optical network can be dynamically reconfigured setting up and tearing down lightpaths on demand, each router can be dimensioned separately for the worst case, regardless of the peering routers. As a result, R4 would need to be equipped with only five interfaces, thus saving 28.5 % of interfaces.

### *11.1.2 Migration Towards In-operation Network Planning*

As technologies are developed to allow the network to become more agile, it may be possible to provide response to traffic changes by reconfiguring the network near real-time. In fact, some operators have deployed GMPLS control planes, mainly for

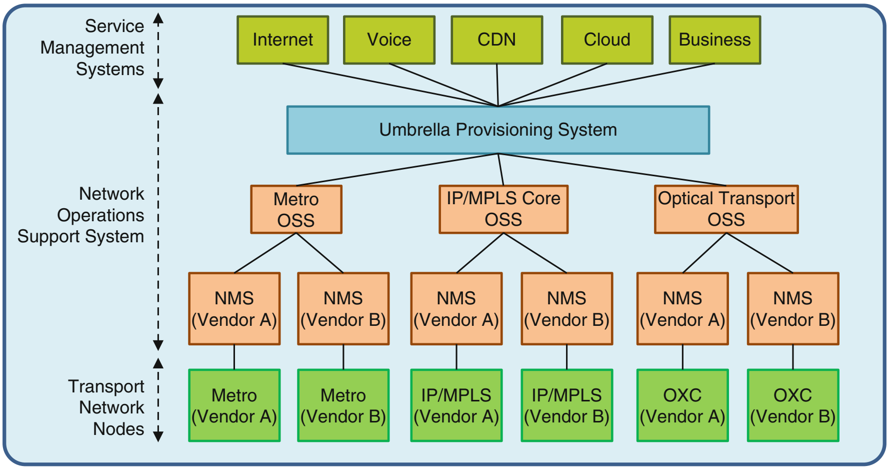



**Fig. 11.2** Current static architecture

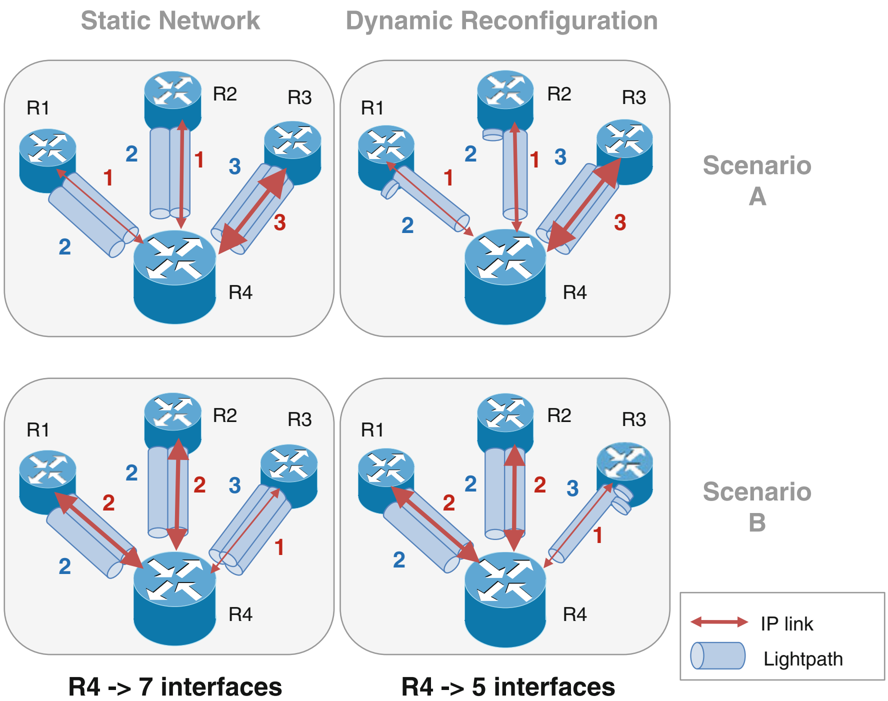



**Fig. 11.3** Example of dynamic planning and reconfiguration

service setup automation and recovery purposes. However, those control only parts of the network and do not support holistic network reconfiguration. This functionality will require an in-operation planning tool that interacts directly with the data and control planes and operator polices via OSS platforms, including the NMS.

Assuming the benefits of operating the network in a dynamic way are proven, the classical network life-cycle has to be augmented to include a new step focused on reconfiguring and re-optimizing the network, as represented in Fig. 11.4. We call that step *in-operation planning* and, in contrast to the traditional network planning, the results and recommendations can be immediately implemented on the network.

To support dynamicity, however, the current network architecture depicted in Fig. 11.2 will need to evolve to include a functional block between the service layer and the network elements to support multi-service provisioning in multi-vendor and multi-technology scenarios; two standard interfaces are required. Firstly, the *north-bound* interface that, among other tasks, gives an abstracted view of the network, enabling a common entry point to provision multiple services and to provision the planned configuration for the network. Moreover, this interface allows coordinating network and service layer according to service requirements. Secondly, the *south-bound interface* covering provisioning, monitoring, and information retrieval.

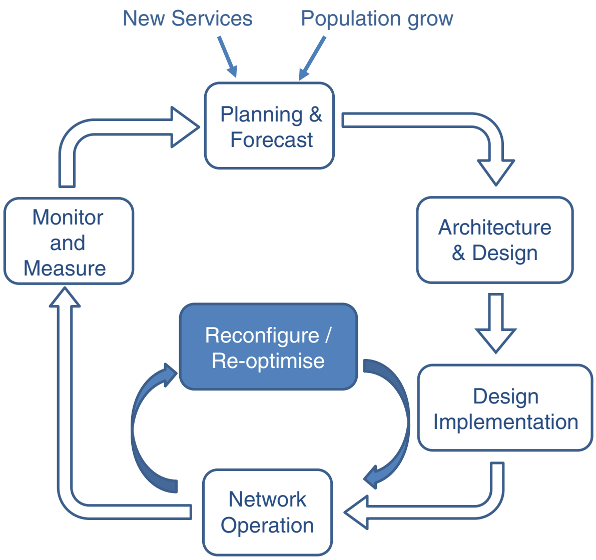



**Fig. 11.4** Augmented networks life-cycle

Finally, operators will typically require human–machine interaction, this is to ensure new configurations and network impact are reviewed and acknowledged, before being implemented in the network.

## 11.2 Architectures to Support In-operation Planning

### *11.2.1 Requirements to Support In-operation Planning*

Network planning processes aim to optimize the deployment of resources for the operator. This activity is done periodically in the operators to have enough capacity to support the traffic engineering requirements from the customers. "Internet traffic engineering is defined as that aspect of Internet network engineering dealing with the issue of performance evaluation and performance optimization of operational IP networks" [1]. According to this definition, traffic engineering are techniques that help the network operator to obtain a better network utilization with subsequent performance improvement. To do so, there are five inputs to the planning process:

- *Traffic demands*. Traffic demands are the information that customers need the network transport. There are many features that determine a given traffic demand. A fixed matrix with end-to-end peak rates can be the input to the problem, but more complicated processes may be defined, for instance, with random traffic changes or with TE parameters.
- *Network status*. Network status provides the information about the status of the connections and occupation of the network resources that are already in the net-

work. This information can be the bandwidth in MPLS networks, or the spectrum/wavelength occupation in the optical networks.
- *Network equipment*. This is all kind of physical devices used in the process of transporting the information, including routers, ROADMs and the cables or fibers that connect them and their software. Operators invest on infrastructure and planification helps to this network capacity task, but once the equipment is deployed on the network, planning algorithms must deal with the resources available only.
- *Objective*. This is a crucial aspect of every engineering problem. It is very important to properly define what aspects the planning problem must either to solve or improve. CAPEX minimization is usually the target of many real planning problems in operators, but there can be other objectives like energy efficiency or delay minimization.
- *Network configuration*. After the planning process a solution to the previous objective is provided. This solution can have two actions: (1) installation of new hardware in the network or (2) modification/activation of the configuration in the network.

The main difference between classical planning process and in-operation planning is that the loop between the network and the planning tool exists. This means that there must be some mechanisms to provide to the planning tool this information. Based on the previous list with the information required for a planning process, we can highlight that the only parameter that is not related with the network is the objective function. Therefore, to support the online process the following requirements are defined:

- *Connection status information*. The in-operation planning process requires knowing the connections that are setup in the network. This is not the whole traffic demands, because there will be new connections, but this is the information from the connections in the network.
- *Traffic Engineering Database*. The planning tool requires knowing the resources that are free in the network. Based on the connections, the traffic engineering can be obtained, but as the information is already in the network, this will reduce the complexity of planning tool.
- *Inventory information*. There are some resources that are not available in the network, because they are not activated, but they are deployed. Moreover, there are physical limitations like the number of slots in the chassis. This information is not in the network state, but it is important for the planning tool so it can optimize where to place the new cards in the nodes or to use some resources that are already deployed.
- *Commands to deploy*. Once the optimization process is run in the planning tool, the planning tool requires a mechanism to deploy such changes in the network. A human in the loop will be required by the operators, but at least through a GUI there must be a protocol to download the new configuration in the network.

The in-operation planning process requires having the same information than traditional planning tool. However, it is required also to have a set of protocols to obtain all information from the network and to download the desired configuration in the network. This information must contain; (1) the connection status information, (2) the Traffic Engineering database, (3) the inventory information, and (4) commands to deploy.

### *11.2.2 Existing Approaches for In-operation Network Planning*

The activity of in-operation planning is performed either as a consequence of an event, that is the so-called reactive in-operation planning approach, at regular intervals, in the periodic planning approach, or before know events, in the preventive planning approach.

#### Reactive In-operation Planning

In the reactive approach, the planning process is triggered by a relevant event that has happened in the network. The event notifies that either there is an unsolved problem in the network, such as a fiber cut or node failure, or informs about the degradation of relevant quality parameters, such as packet loss ratio or video quality, which require attention and demand corrective actions in the network.

It is important to highlight that failure events are first handled in real time by the control plane that let the network survive and are performed automatically. However, after the control plane has taken the necessary actions to recover as much as possible the provided services, the network might not be in an optimal situation or can be in a potential high risk. Some services can be even not be recovered by the control plane. The in-operation planning process, after the control plane has finished its automatic recovery tasks, looks at the traffic engineering and service databases and performs a global optimization that can recommend movements of services to allocate room for the non-recovered demands. The operator can decide to apply the changes depending on the estimated reparation of the source of the problem (e.g., repairing the fiber, replacing a damaged card).

The event that triggers the in-operation planning process is generated by a correlation engine that receives all the alarams and real-time quality indication paramenters. Note that after a failure, multiple alarms will be generated in the network elements and monitoring devices. The correlation engine has to find out that all alarms and warnings are related to the same problem, and should trigger a single in-operation event. In order to decide when to send the re-planning event, a set of policies have to be defined by the operator.

Summarizing, the sources that trigger the re-planning event can be:

- Alarms and warnings from network elements (e.g., Loss of Signal alarms).
- Alarms from monitoring devices (e.g., OTDRs or monitoring devices).
- Threshold of KPI surpassed. Key Performance Indicators (KPIs) can be measured in real time.

The ability of the network planner to obtain real-time information of the network and access fast optimization tools is essential to ensure a high service availability and improve the network quality.

### Periodic Planning

On the other hand, the operator can decide to perform periodic re-plannings of the network, as a preventive check to early-detect risks and capacity exhaustion. Every time the re-planning process if performed, the real situation of the network is benchmarked against the potential best situation of the network. The operator, in his decision on whether to apply the necessary changes or leave the network in the same state, has to consider the degree of non-optimality of the network, the complexity and risk of performing the proposed changes and the impact on running services. The classical rule of thumb of a network operation is "*if it isn't broken, don't try to fix it*." However, even though avoiding unnecessary risks is the main rule, a policy needs to be set to allow the planning tool to recommend perform the actions.

The periodicity of the periodic planning determines the complexity of the algorithm that can be executed by the computation engine. While the yearly process can be performed by powerful CPU-hungry algorithms that take several days to give the best scenarios, a nightly or weekly planning must be focused on some key parameters, such as current network availability, find what happens in case of the most likely failure events, evolution of quality KPIs, available capacity for new demands. If the check finds a risk, for example capacity for new services to be exhausted in less than 1 month, the operator is notified and a deeper re-planning is conducted.

The periodic planning will try re-optimize the network to create space for new demands. However, should that not be possible anymore, the planning tool will launch an alarm and recommend the acquisition of new resources.

For the latter, the in-operation planning has to deal with not only control-plane information, but with the network inventory.

### Preventive Planning

An interesting approach to the in-operation planning concept is triggering the planning process before some know event, such as summer holidays or a big sports event. Bank holidays and seasonal holidays cause relevant movements of people, who take their smartphones and laptops to the destination. It is common that a village with a few permanent residents, multiplies by several times, the number of inhabitants in the summer months. Coastal towns are especially crowded in summer. As the infrastructure is deployed and dimensioned for the regular network state, a big change in the traffic pattern can cause a high level of congestion, as well as thousands of complaints in the customer care center.

Before a know event happens, the network can be replanned taken into account the estimated new traffic demand. For example, a set of lighpaths or microwave links can be re-oriented towards vacation places. After the first set of actions, a periodic in-operation planning has to be performed to ensure that the network runs within the desired quality KPIs.

As opposed to the classic planning, the preventive planning is fully automated and counts with real-time information. The network planner, however, has to indicate the relevant events and the best-known information, even not accurate. With the classic approach, to avoid any risks, due to the lack for continuous feedback and quality information, a high overprovisioning has to be made.

### *11.2.3 Relationship with the Control Plane*

#### The Front-End/Back-End Architecture

The path computation element (PCE) architecture allows for a lot of flexibility when considering PCE deployments. Although it is commonly implied that a PCE is deployed in a centralized way (i.e., one PCE per domain), nothing precludes deploying several PCEs, or even a PCE per node. A network may have more than one PCE, which may provide redundancy for load-sharing, resilience, or partitioning of computation features.

Even in the scope of a logically centralized deployment, more complex architectures involve the design of a flexible, policy-enabled architecture in which PCEs may exchange detailed information regarding their specific capabilities in path computation, providing network operators with further control in what concerns the path computation function. This architecture must allow advanced deployments ranging from load balancing mechanisms to partition the path computation function into specialized PCEs. For example, we can mention the deployment of one or more PCE specialized in computations used in network planning by means of global concurrent optimization; a PCE specialized in real-time restoration, etc.

The so-called Front-end/Back-end architecture (FBA) [2] is an architecture based on the concept of a front-end and one or more dedicated back-ends, where the end client of path computation client (PCC) only sees the front-end, and operators may deploy different capabilities at back-ends. Likewise, complex computations can be delegated to dedicate PCEs. A common use case is when one or more PCEs are deployed in the same TE domain, so the back-end PCEs may use the same TED, although it is not mandatory.

The main motivations behind this work are related to scalability and load-sharing policies while enabling some degree of specialization. The proposed FBA (see Fig. 11.5) also provides a higher level of robustness and redundancy: even though the front-end PCE (f-PCE) is still a single point of failure, its implementation is significantly simpler than a back-end-PCE (b-PCE). Additionally, several b-PCEs can present the same set of capabilities, and several f-PCEs could be deployed managing several back-ends.

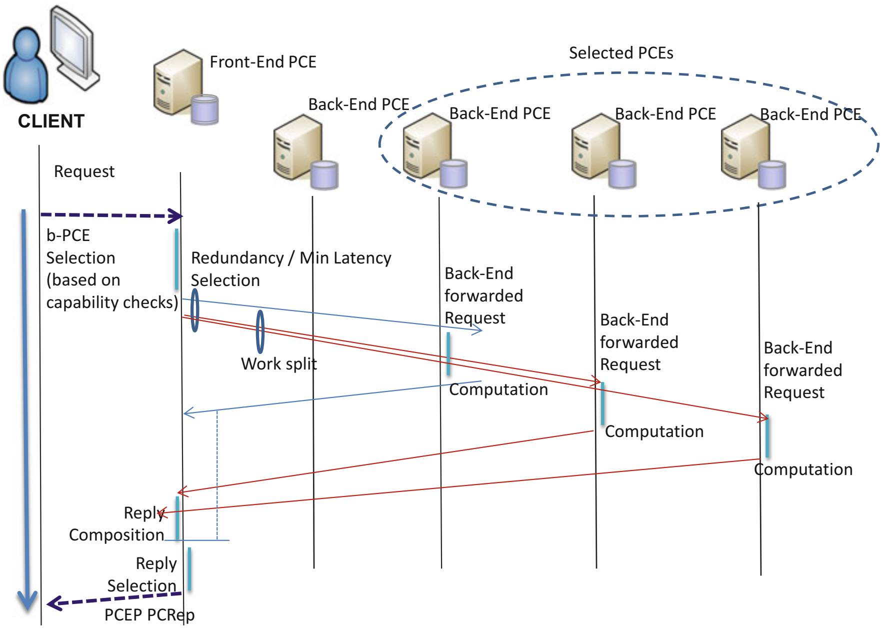



**Fig. 11.5** An f-PCE/b-PCE architecture as used by ABNO to offload complex computations to one (or more) dedicated PCE. The end client interfaces to the front-end PCE (which can be the main PCE in ABNO) and this one is responsible for selecting candidate back-end PCEs and forwarding the request to them. The architecture is flexible in terms of the algorithms and policies used to select back-end PCEs

In this setting, the role of the f-PCE is to interface with end-clients that perform path computation requests and the selection of candidate and viable b-PCEs: after an initial deployment (e.g., configured) of candidate b-PCEs, the dynamic selection (e.g., on request) of viable b-PCEs is based on the end-client request and applicable policies, e.g., to select the min-latency response or to select a "posteriori" best-path. It is also possible to apply Round Robin or more advanced approaches such as using the Least-loaded PCE and its response time. Note that, when several b-PCEs are selected, and a given request is forwarded to them, the f-PCE must ensure that a single reply is forwarded to the end client. For this reason, the first response is forwarded, the rest is discarded, or *N* responses are stored, and the best one is then selected (using a timer).

### Stateless PCE with GCO and ABNO

In view of the previous, there are different ways in which the concept of "in-operation planning tool" can be deployed, along its relationship with the control plane. The straightforward approach is to deploy it as a dedicated back-end PCE for performance improvements and optimizations. The back-end PCE is accessible via the PCEP interface, so the application-based network operations (ABNO) [3] components can forward requests to the planning tool.

Depending on the capabilities of the PCEs (either as standalone elements or integrated within an ABNO control), different interfaces and options should be used. For example, without stateful capabilities, the f-PCE or the ABNO controller may rely on global Concurrent Optimization (GCO). With GCO extensions to PCEP, the b-PCE is able to simultaneously consider the entire topology of the network and the complete set of existing TE LSPs, and their respective constraints, and look to optimize or re-optimize the entire network to satisfy all constraints for all TE LSPs. A GCO may also be applied to some subset of the TE LSPs in a network. For this, the set of active connections is provided in the request message. Alternatively, if stateful extensions are enabled, both f-PCE /ABNO and b-PCE may synchronize the LSPDB. Finally, the use of BGP-LS may allow the b-PCE to retrieve the TED from a BGP-LS speaker, located at the f-PCE, ABNO topology server or directly from a node mapping the IGP.

## 11.3 Main Use Cases

Let us now to analyze three in-operation planning use cases: (1) virtual network topology reconfiguration, (2) spectrum defragmentation, and (3) after failure repair optimization.

### *11.3.1 Use Case I: Virtual Network Topology Reconfiguration*

In this use case, we consider a multilayer network where a set of routers are connected through a set of lightpaths over a flexi-grid optical network to create a virtual network topology (VNT). We analyze the reconfiguration of the VNT after a failure in an optical link. Fig. 11.6a illustrates a scenario consisting of four optical cross-connects (OXCs) in the optical layer and three routers in the IP/MPLS layer; physical topologies are depicted on the top, where IP/MPLS routers are connected through 10 Gb/s lightpaths to create the VNT depicted at the bottom. Three bidirectional packet label-switched paths (LSPs) are established on the virtual topology.

After an optical link failure occurs, Fast Re-Route (FRR) [4] can be used to recover part of the affected packet LSPs immediately after the failure. In addition, the state of the network after the failure can be updated in the control plane also within seconds. However, the capacity of some packet LSPs might be reduced or even remain disconnected as a consequence of high congestion in some links in the VNT and a reconfiguration needs to be performed to groom and distribute traffic away from choke points and heavily utilized resources.

An example is presented in Fig. 11.6b, where link X1–X2 has failed entailing a re-routing of packet LSP R1–R2 through R3 and getting its capacity reduced from

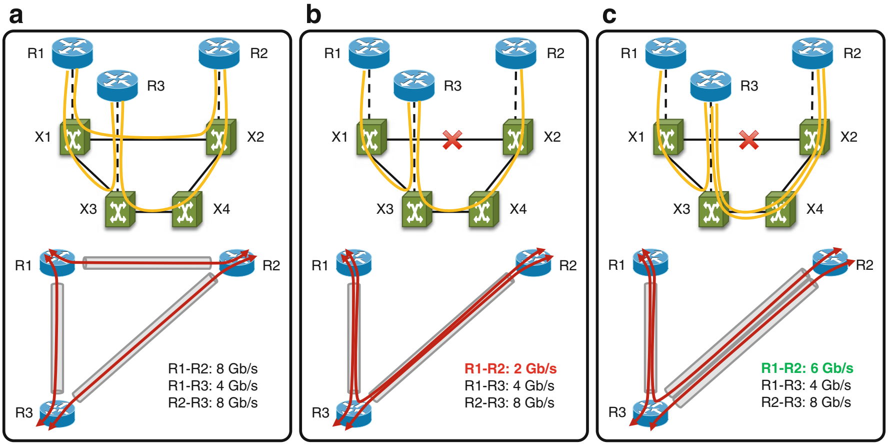



**Fig. 11.6** Example of VNT reconfiguration. (**a**) State before failure. (**b**) A failure affects an optical link and part of the IP/MPLS traffic has been recovered. (**c**) VNT is reconfigured

8 Gb/s to only 2 Gb/s. The reason for this capacity squeezing is that lightpaths R2–R3 and R1–R3 are conveying respectively the 8 Gb/s packet LSP R2–R3, and the 4 Gb/s packet LSP R1–R3. The established lightpaths have 10 Gb/s, so the remaining capacity of lightpaths R2–R3 and R1–R3 are only 2 Gb/s and 6 Gb/s, respectively. The available capacity for the re-routed packet LSP R1–R2 is the minimum between available capacities on used lightpaths, in this case, 2 Gb/s. Note that if one of those working lightpaths used to re-route packet LSP R1–R2 had no available capacity, new lightpaths should need to be established.

In Fig. 11.6c, a new lightpath is created between R2 and R3 and as a result the packet LSP R1–R2 can be re-routed and its capacity expanded to 6 Gb/s. packet LSP R1–R3 is consuming 4 Gb/s from lightpath R1–R3, so the available capacity for packet LSP R1–R2 is 6 Gb/s. To restore its original capacity of 8 Gb/s a new LSP should be created between R1 and R3.

To cope with VNT reconfigurations our approach relies on the Application-based Network Operations (ABNO) architecture presented in Fig. 11.7, where the several steps to perform a VNT are also represented.

When the network operator needs to perform a network-wide VNT reconfiguration, a request is sent from the Network Management System (NMS) or the Operations Support System (OSS) to the ABNO Controller (1), who then forwards it to the Path Computation Element (PCE) (2). To alleviate PCE workload, a back-end PCE (b-PCE) which contains an active solver is responsible for computing new virtual topology layout taking into account the current state of the network. Therefore, the front-end PCE (f-PCE), formerly the PCE, sends a Path Computation Request (PCReq) message towards the in-operation planning tool running in that b-PCE to compute the new virtual topology (3).

The tool considers all the surviving resources, which may include router interfaces and transponders connected to failed optical connections, spare interfaces that

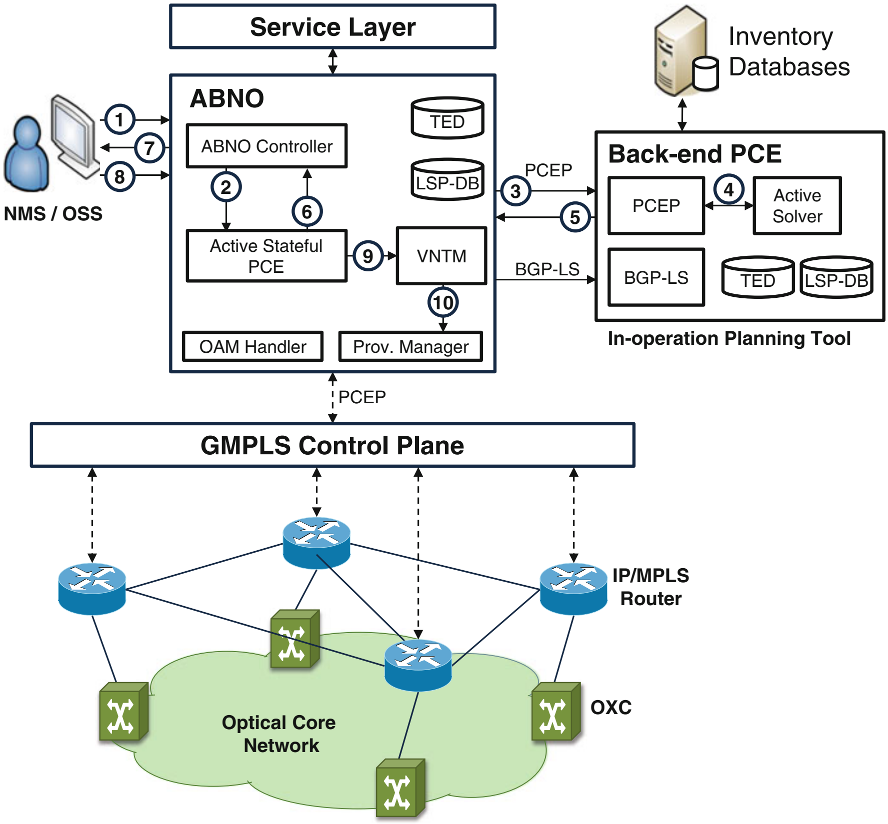



**Fig. 11.7** The VNT re-optimization process

typically exist in the network for normal growth, and possibly some spare routers that have been installed ahead of time; all those resources can be stored in an inventory database. In addition, it must consider how to implement the desired IP/MPLS connectivity over the optical layer. To this end, it needs to know which optical links and nodes are up and which connections are optically feasible, considering optical impairments.

When a result is obtained (4), the set of lightpaths is replied in a Path Computation Reply (PCRep) message (5) towards the originating front-end PCE (f-PCE). In case that an operator needs to approve the new (virtual) layout, the f-PCE forwards it to the ABNO Controller (6). The computed layout is then presented to the operator for final approval (7). When the operator acknowledges the new optimized layout (8), it is passed to the VNT Manager (VNTM) (9) which computes the sequence of operations to carry out in terms of creating new lightpaths and/or updating existing ones, and re-routing existing packet LSPs minimizing disruption.

The sorted sequence of actions on existing packet LSPs is passed to the provisioning manager (10), which is able to interact with each head-end node. The provisioning interface, by which the provisioning manager is able to suggest re-routing of packet LSPs, is based on the PCE Protocol (PCEP) interface using Path Computation Update (PCUpd) messages [2]. The new allocated resources are reported back to the provisioning manager and ultimately the VNTM, using Path Computation Report (PCRpt) messages. Note that after a successful re-optimization, the LSP database (LSP-DB) in the ABNO is updated accordingly.

### *11.3.2 Use Case II: Spectrum Defragmentation*

In this use case we study the spectrum defragmentation problem, a specific problem that arises in flexi-grid networks [5] and where re-optimization could bring clear benefits. In such networks, lightpaths can be allocated using variable-sized frequency slots, whose width (usually a multiple of a basic width such as 12.5 GHz) is a function of the requested bit rate, Forward Error Correction (FEC), and modulation format. Such frequency slots must be contiguous in the spectrum and the same along the links in its route. As a consequence of the unavailability of spectrum converters, spectrum fragmentation appears increasing the blocking probability of connection requests, making worse the network grade of service.

An example is shown in Fig. 11.8 where the optical spectrum of a link is represented in three sequential situations. In Fig. 11.8a three already established lightpaths share that link, each one using a different frequency slot width. If a new lightpath needing 50 GHz is requested, it would be blocked as a consequence of lack of spectrum contiguity. In such scenario, re-optimization could be applied to the network before a connection request is blocked, by re-allocating already established lightpaths in the spectrum (Fig. 11.8b) to make enough room for the triggering connection requested (Fig. 11.8c).

Authors in [6] describe the spectrum re-allocation (SPRESSO) algorithm to efficiently compute the set of connections to be re-allocated. In [7] the SPRESSO algorithm was integrated into an active stateful PCE and re-allocations were performed in a hitless manner by using the *push–pull* technique.

The push–pull technique has been introduced in [8] to perform in-operation defragmentation. It operates at the physical layer by gradually re-tuning the transmitting laser source of a selected lightpath. Push–pull technique has been experimentally demonstrated considering control plane and data plane aspects for both on–off keying (OOK) modulation with direct detection, and polarization multiplexed quadrature phase shift keying (PM-QPSK) modulation with coherent detection.

The testbed in Fig. 11.9 has been used in the experimental demonstration of the push–pull technique for PM-QPSK 112 Gbps (100 Gbps plus overhead) transmission with coherent detection. This testbed includes two 80 Km links and three GMPLS-controlled nodes. The first node acts as transmitter, the second node is implemented with a bandwidth variable wavelength selective switching (BV-WSS), and the third node acts as receiver. The utilized transmitter and receiver are detailed in [9].

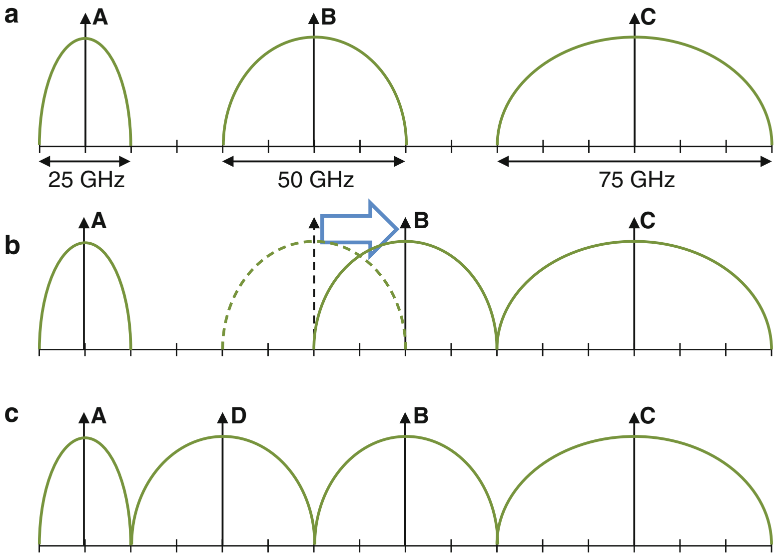



**Fig. 11.8** An example of spectrum defragmentation optimization. (**a**) Initial spectrum of the link. (**b**) LSP B is re-allocated to make room for LSP D. (**c**) New LSP D is setup

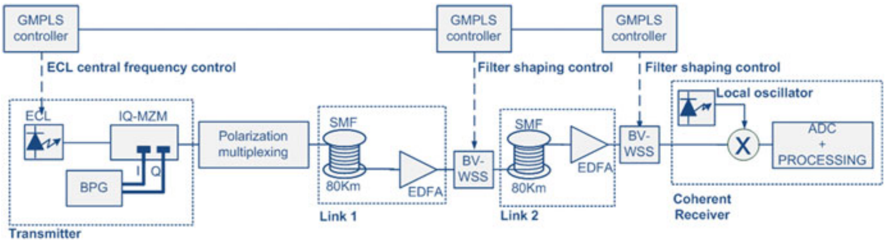



**Fig. 11.9** Push–pull experimental testbed including control and data planes with PM-QPSK with coherent detection

In the performed experiment, the transmitter (TX) laser is gradually tuned to shift the PM-QPSK channel. An overall shift of 50 GHz is applied in 40 iterations each imposing a frequency shift of 1.25 GHz at the TX laser. For each iteration, the receiver first performs a data acquisition of 40 s. Then, it elaborates offline the received data obtaining an estimation of frequency shift. Finally, it triggers the tuning of the local oscillator to nullify the estimated frequency offset.

Figure 11.10 shows the estimated frequency offset values at the receiver, for each acquisition. At the beginning, the frequency offset is zero. When the signal frequency begins to move, a different value of frequency offset is estimated and subsequently compensated. This process is repeated for each iteration. The figure shows that an almost constant frequency shift of about −1.25 GHz is estimated in each iteration during the push–pull operation at the receiver. When the push–pull is

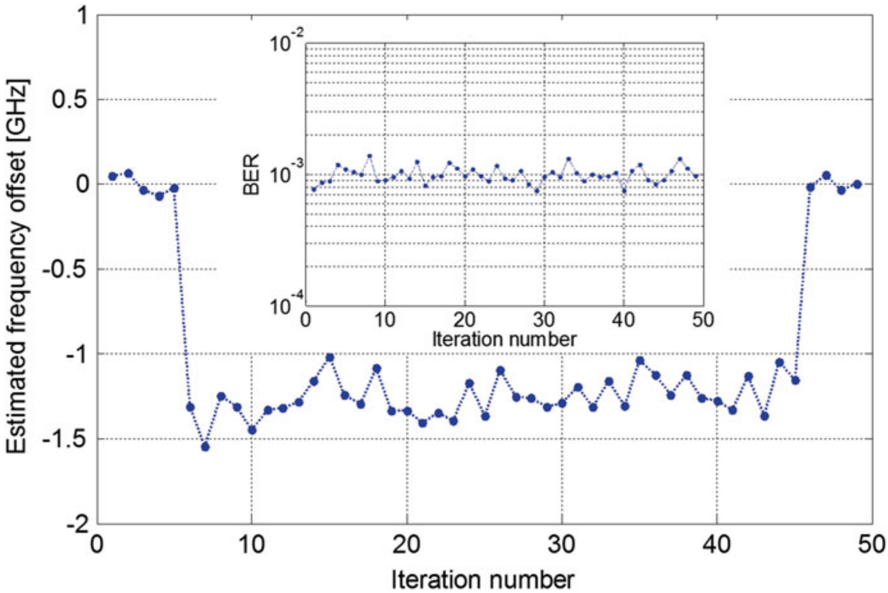



**Fig. 11.10** Estimated frequency offset between the signal and the local oscillator and achieved BER, as a function of the iteration number

terminated and the overall 50 GHz shift is performed, the loop back control successfully estimates a null frequency offset. The bit error rate (BER) value for each pushing step is also reported in the inset of Fig. 11.10 demonstrating no performance degradation during the push–pull operation. Figure 11.11 reports the recovered PM-QPSK signal constellation including both polarization components.

Figure 11.12 illustrates the overall control-plane architecture proposed in [7] to support flexi-grid network re-optimization (including push–pull operation), while Fig. 11.13 depicts the meaningful PCEP messages exchanged between architecture components in this experiment. Moreover, message numbers are correlated to steps in Fig. 11.12.

When the NMS needs to setup a new connection from Router A to Router B, it sends a request to the ABNO Controller (IP = 172.16.4.2) (1). After checking admission policies, a PCReq message is sent to the f-PCE (IP = 172.16.1.3) (2), which invokes its local provisioning algorithm (3).

In the event of insufficient resources being available due to spectrum fragmentation, the active f-PCE reports a PCRep to the ABNO Controller notifying the unability to establish the path (4). The ABNO Controller then requests a path computation with network defragmentation of relevant nodes and connections (5), utilizing the right algorithm to provide such re-optimizations. Similarly, as in the previous use case, we assume that the b-PCE (IP = 172.16.50.2) providing such algorithm will perform the computation (6) upon receipt of a PCReq message containing a Global Concurrent Optimization (GCO) [10] request. When a result is obtained (7), it is sent back to the f-PCE (8) by means of a PCRep message.

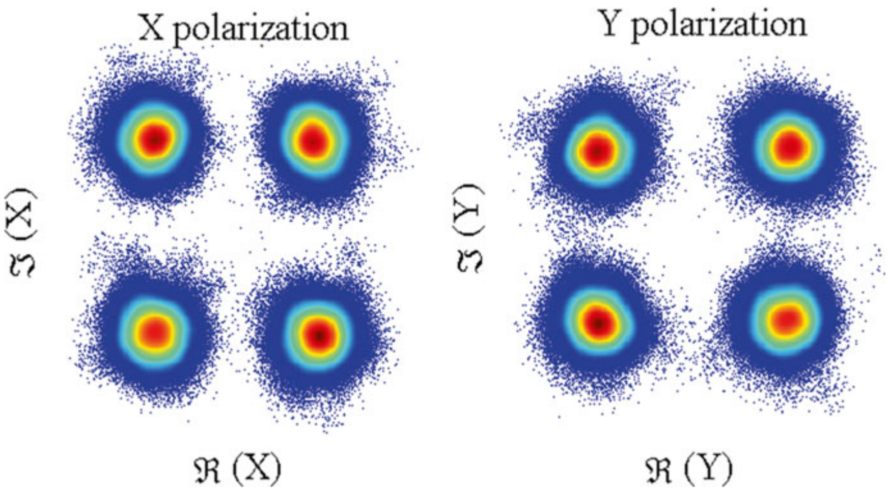



**Fig. 11.11** Recovered signal constellation with the two polarization components

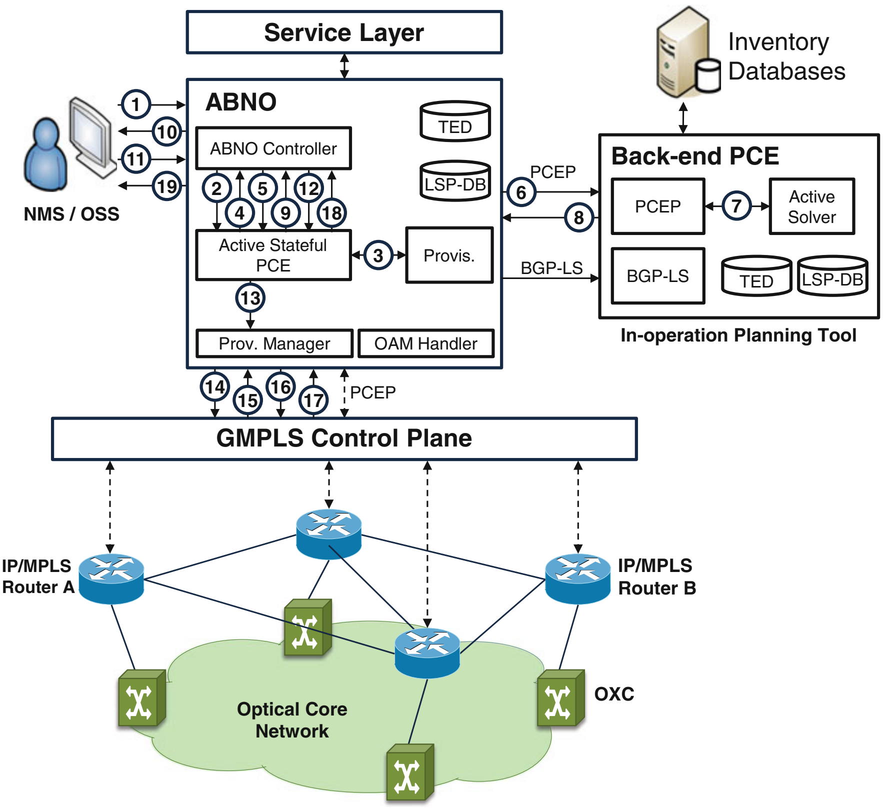



**Fig. 11.12** The spectrum defragmentation re-optimization process

| No. | Time | Source | Destination | Protocol | Length | Info |
|---|---|---|---|---|---|---|
| 2 | 181.073832 | 172.16.4.2 | 172.16.1.3 | PCEP | 110 | PATH COMPUTATION REQUEST MESSAGE |
| 4 | 181.073980 | 172.16.1.3 | 172.16.4.2 | PCEP | 98 | PATH COMPUTATION REPLY MESSAGE |
| 5 | 181.146152 | 172.16.4.2 | 172.16.1.3 | PCEP | 110 | PATH COMPUTATION REQUEST MESSAGE |
| 6 | 181.147138 | 172.16.1.3 | 172.16.5.2 | PCEP | 730 | PATH COMPUTATION REQUEST MESSAGE |
| 8 | 181.282308 | 172.16.5.2 | 172.16.1.3 | PCEP | 574 | PATH COMPUTATION REPLY MESSAGE |
| 14 | 186.505937 | 10.0.0.49 | 10.0.0.5 | PCEP | 78 | PATH COMPUTATION UPDATE MESSAGE |
| 15 | 186.508890 | 10.0.0.5 | 10.0.0.49 | PCEP | 78 | PATH COMPUTATION REPORT MESSAGE |
| 14 | 188.735930 | 10.0.0.49 | 10.0.0.6 | PCEP | 78 | PATH COMPUTATION UPDATE MESSAGE |
| 15 | 188.738888 | 10.0.0.6 | 10.0.0.49 | PCEP | 78 | PATH COMPUTATION REPORT MESSAGE |
| 16 | 192.452368 | 10.0.0.49 | 10.0.0.4 | PCEP | 194 | PATH COMPUTATION INITIATE MESSAGE |
| 17 | 192.455502 | 10.0.0.4 | 10.0.0.49 | PCEP | 78 | PATH COMPUTATION REPORT MESSAGE |
| 18 | 195.426888 | 172.16.1.3 | 172.16.4.2 | PCEP | 190 | PATH COMPUTATION REPLY MESSAGE |

**Fig. 11.13** Spectrum defragmentation experimental captures

When the solution has been received, in case that an operator needs to approve the new (virtual) layout, the f-PCE forwards it to the ABNO Controller (9). The computed re-optimization is then presented to the operator for final approval (10). When the operator acknowledges the new optimized connections (11), it is passed to the f-PCE (12) that forwards it to the provisioning manager (13) which computes the sequence of operations to carry out in terms of creating new lightpaths and/or updating existing ones, and re-routing existing packet LSPs minimizing disruption.

Existing connection re-allocations are requested to its corresponding head-end routers (IP sub-network = 10.0.0.X) sending them PCUpd messages (14) and waiting for the corresponding PCRpt messages (15).

Once the dependent connections have been updated, a Path Computation Initiate (PCInit) message is sent to Router A to establish the new connection (16). The connection setup is confirmed when Router A sends the corresponding PCRpt message to the provisioning manager (17), who in turn forwards it to the f-PCE. Then the f-PCE confirms by means of a PCRep message to the ABNO Controller, the connection establishment (18). The process concludes when the ABNO Controller confirms to the NMS, the establishment of the connection (19). Further details on this experiment can be found in [11].

### *11.3.3 Use Case III: After Failure Repair Re-optimization*

In this use case, we face the after failure repair re-optimization (AFRO) problem, defined in [12]. When a fiber cut occurs, connections affected are restored; when the link is repaired not only those connections that were restored, but also any other connection that might use the repaired link, can be re-routed.

In addition, to increase restorability, multipath restoration can be applied [13] dividing original connections into several parallel disaggregated connections (*sub-connections*). Although multipath restoration has benefits, it provides poor resource utilization and it is spectrally inefficient. Hence, re-optimizing the network just after a link is repaired, becomes essential. To that end, the AFRO problem needs to be enhanced with the capability of merging sub-connections; we call it as MP-AFRO [14].

For illustrative purposes, Fig. 11.14 reproduces a small flexi-grid network topology where several connections are currently established; in particular, the route of connections *P1* and *P2* is shown. The spectrum usage is also provided in Fig. 11.14, where the spectrum allocation for all five established connections is specified. Three snapshots with the state of the network are shown: the link 6-7 has failed in Fig. 11.14a; multipath restoration has been applied in Fig. 11.14b and connection *P2* has been split into two parallel sub-connections *P2a* and *P2b*; failed link 6-7 has been repaired in Fig. 11.14c, and re-optimization has been performed by solving MP-AFRO.

It is clear, in view of the example, that multipath restoration allows increasing restorability, in particular when no enough contiguous spectrum can be found along a single path, as happened when restoring *P2*. Nonetheless, this benefit is at the cost of an increased resource usage, not only as a result of using (not shortest) parallel routes, but also because the spectral efficiency is degraded when connections are split. For instance, a 400 Gb/s aggregated flow can be conveyed on one single 100 GHz connection or on four parallel 37.5 GHz sub-connections (150 GHz in total).

For this very reason, resource utilization can be improved by applying MF-AFRO, i.e., by re-routing established connections on shorter routes and by merging parallel sub-connections to achieve better spectrum efficiency. Figure 11.14c illustrates an example of such re-optimization, where connection *P1* has been re-routed using a shorter route that includes the repaired link, whilst sub-connections *P2a* and *P2b* have been merged on a single connection conveying the originally requested bandwidth.

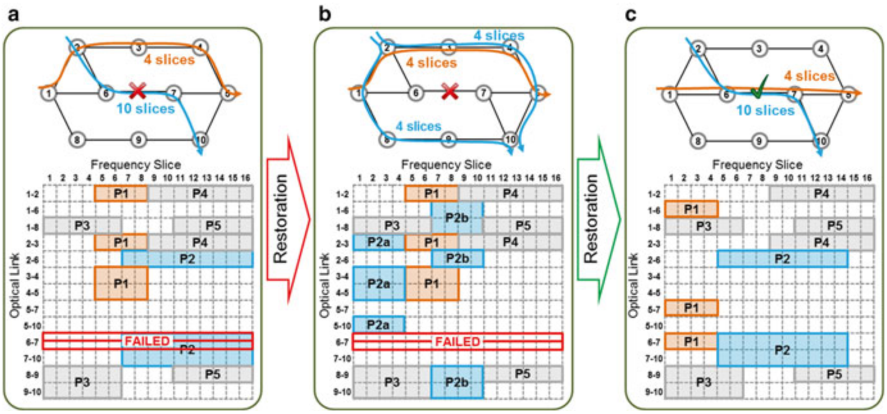



**Fig. 11.14** An example of after-failure repair optimization. (**a**) Initial network state before link failure. (**b**) Network state after restoring failed links. (**c**) Network state after repairing failed link and re-optimization

To deal with this network re-optimization use case, like in previous ones, we consider a control plane based on the ABNO architecture, including the f-PCE responsible of computing provisioning requests and dealing with network data plane and the b-PCE capable of performing complex computations to solve optimization problems. The diagram in Fig. 11.15 reproduces the re-optimization sequence involving ABNO components.

In light of the MP-AFRO problem statement, the planning tool in the b-PCE needs to know what the current network topology is, with information of nodes and links, and state of the resources, i.e., current demands and its corresponding sub-connections. This information is stored in f-PCE's TED and LSP-DB databases, respectively. Different solutions can be considered to synchronize this information. In our case, we're considering Border Gateway Protocol—Link-State (BGP-LS) [15], to synchronize the TED, and the PCEP Extensions for Stateful PCE [2] to synchronize the LSP-DB.

Figure 11.15 also illustrates the proposed control plane architecture to support MP-AFRO, while Fig. 11.16 depicts the meaningful PCEP messages exchanged between architecture components in this experiment. Moreover, message numbers are correlated to steps in Fig. 11.15.

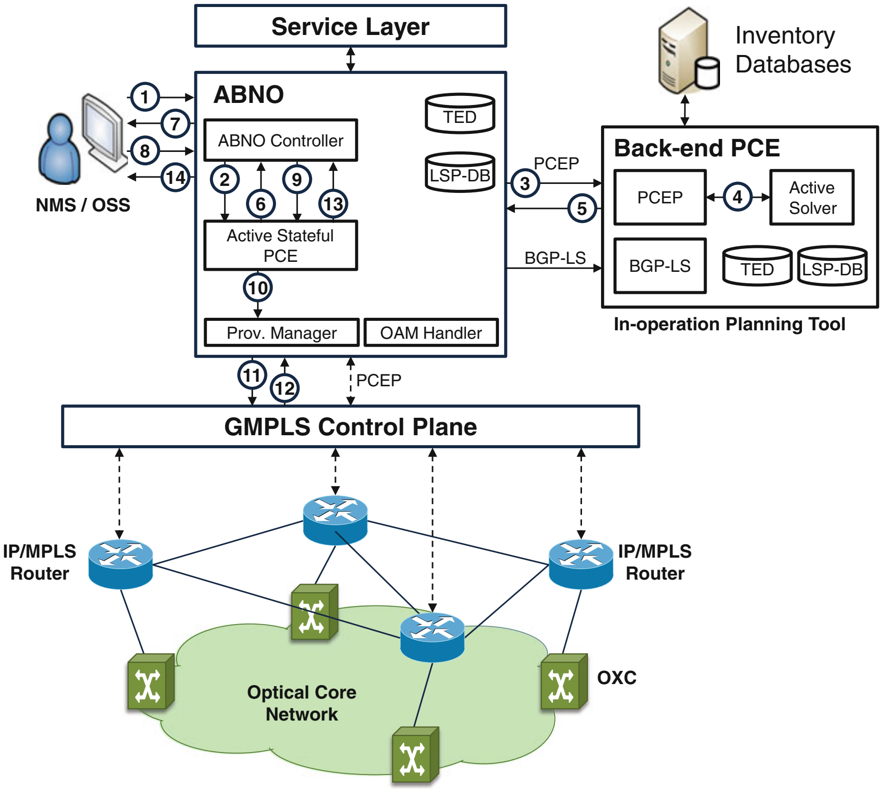



Fig. 11.15 The after-failure repair optimization process

| No. | Time | Source | Destination | Info |
|---|---|---|---|---|
| 2 | 8.097681 | 172.16.4.2 | 172.16.1.3 | PATH COMPUTATION REQUEST |
| 3 | 8.097845 | 172.16.1.3 | 172.16.5.2 | PATH COMPUTATION REQUEST |
| 5 | 8.199351 | 172.16.5.2 | 172.16.1.3 | PATH COMPUTATION REPLY |
| 11 | 8.220174 | 10.0.0.49 | 10.0.0.1 | PATH COMPUTATION UPDATE |
| 12 | 8.223918 | 10.0.0.1 | 10.0.0.49 | PATH COMPUTATION REPORT |
| 11 | 11.281405 | 10.0.0.49 | 10.0.0.8 | PATH COMPUTATION UPDATE |
| 12 | 11.285044 | 10.0.0.8 | 10.0.0.49 | PATH COMPUTATION REPORT |
| 13 | 16.639926 | 172.16.1.3 | 172.16.4.2 | PATH COMPUTATION REPLY |

**Fig. 11.16** After-failure repair optimization experimental captures

We assume that an operator in the NMS triggers the MP-AFRO workflow after a link has been repaired. To that end, the NMS issues a service request (labeled as 1 in Fig. 11.15) towards the ABNO Controller (IP = 172.16.4.2). When the request from the NMS arrives at the ABNO Controller, re-optimization is requested by sending a PCReq message (2) to the f-PCE (IP = 172.16.1.3). Upon receiving the request, the f-PCE collects relevant data to be sent to the b-PCE (IP = 172.16.5.2) in the form of a PCReq message containing a GCO [10] request (3).

The PCReq message to be sent to the b-PCE has to have the list of identifiers for demands candidate to be re-optimized including the original bandwidth, remember that current connection’s bandwidth could be less than the requested bandwidth due to the restauration process, and the identifier of the fixed link on which re-optimization should be focused. Therefore, an algorithm to find the candidate demands in the LSP-DB is needed. Our algorithm selects the demands that have been previously split or whose shortest path traverses the repaired link.

Upon receiving the PCReq message, the b-PCE runs the MP-AFRO algorithm (4). To solve the MP-AFRO problem, a heuristic algorithm can be implemented, computing a number of iterations each one randomly sorting the list of candidate demands and eventually returning the best solution found, i.e. maximizing the bitrate that is served and minimizing the amount of used resources.

On each iteration the set of candidate demands are deallocated from a copy of the local Traffic Engineering Database (TED), and a new route and spectrum allocation (RSA) [16] is computed for each of them in the specified order allocating the new route found so subsequent computations of the same iteration consider those resources in use.

The solution of the MP-AFRO problem is encoded in a PCRep (5); each individual request is replied specifying the bitrate that could be served and the RSA of the connections related to each request. Note that the solution might entail merging several existing sub-connections to create one or more new connections. For each re-optimized demand, the order in which the solution needs to be implemented in the data plane should be provided.

Upon receiving the PCRep message with the solution from the b-PCE, the f-PCE compares the re-optimized routes received with the connections currently in its

LSP-DB. In case updates are received and an operator needs to approve the re-optimization solution, the f-PCE forwards it to the ABNO Controller (6). The computed re-optimization is then presented to the operator for final approval (7). When the operator acknowledges the new optimized layout (8), it is passed to the f-PCE (9) which forwards it to the provisioning manager (10).

The provisioning manager computes the sequence of operations to carry out in terms of creating/updating new/existing lightpaths minimizing disruption, and sends the corresponding PCUpd/PCInit messages towards source GMPLS controllers to the head-end routers (IP sub-network = 10.0.0.X) through the provisioning manager (11) in the required order. When the corresponding PCRpt messages are received (12), the provisioning manager forwards to the f-PCE the confirmation. Finally, the f-PCE replies the completion of the requested operation to the ABNO Controller using a PCRep message (13), which eventually informs the NMS (14). Further details on this experiment can be found in [17]

## 11.4 Conclusions

A control and management architecture for transport networks has been proposed to support in-operation planning. The architecture is based on ABNO, and allows carriers to operate the network in a dynamic way, and to reconfigure and re-optimize the network in near real-time in response to changes like traffic or failures.

Network life-cycle is extended, achieving a better resource utilization and thus reducing network CAPEX. Moreover, process automation reduces manual interventions and, consequently, OPEX. Use cases have been used to illustrate how the proposed in-operation planning and control and management architecture work together. In a multilayer network scenario we analyzed VNT reconfiguration after a failure and for disaster recovery. In a flexi-grid network scenario, we studied LSP re-allocation to reduce connection blocking.

## References


1. E. Mannie (ed.), *Generalized Multi-Protocol Label Switching (GMPLS) Architecture*, IETF RFC 3945 (2004, October)
2. R. Casellas, R. Muñoz, R. Martínez, R. Vilalta, Applications and status of path computation elements [invited]. IEEE/OSA J. Opt. Commun. Netw. **5**(10), A192–A203 (2013)
3. D. King, A. Farrel, *A PCE-Based Architecture for Application-Based Network Operations*, IETF RFC 7491 (2015, March)
4. P. Pan, G. Swallow, A. Atlas, *Fast Reroute Extensions to RSVP-TE for LSP Tunnels*, IETF RFC 4090 (2005)
5. M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, S. Matsuoka, Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. IEEE Commun. Mag. **47**, 66–73 (2009)

6. A. Castro, L. Velasco, M. Ruiz, M. Klinkowski, J.P. Fernández-Palacios, D. Careglio, Dynamic routing and spectrum (re)allocation in future flexgrid optical networks. Comput. Netw. **56**, 2869–2883 (2012)
7. A. Castro, F. Paolucci, F. Fresi, M. Imran, B. Bhowmik, G. Berrettini, G. Meloni, A. Giorgetti, F. Cugini, L. Velasco, L. Poti, P. Castoldi, Experimental demonstration of an active stateful PCE performing elastic operations and hitless defragmentation, in *Proceedings of European Conference on Optical Communication (ECOC)*, London (2013)
8. F. Cugini, F. Paolucci, G. Meloni, G. Berrettini, M. Secondini, F. Fresi, N. Sambo, L. Poti, P. Castoldi, Push-pull defragmentation without traffic disruption in flexible grid optical networks. J. Lightwave Technol. **31**(1), 125–133 (2013)
9. F. Cugini, G. Meloni, F. Paolucci, N. Sambo, M. Secondini, L. Gerardi, L. Poti, P. Castoldi, Demonstration of flexible optical network based on path computation element. J. Lightwave Technol. **30**(5), 727–733 (2012)
10. Y. Lee, J.L. Le Roux, D. King, E. Oki, *Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization*, IETF RFC 5557 (2009)
11. L.L. Gifre, F. Paolucci, A. Aguado, R. Casellas, A. Castro, F. Cugini, P. Castoldi, L. Velasco, V. López, Experimental assessment of in-operation spectrum defragmentation. Photon. Netw. Commun. **27**, 128–140 (2014)
12. M. Ruiz, M. Zotkiewicz, A. Castro, M. Klinkowski, L. Velasco, M. Pioro, After failure repair optimization in dynamic flexgrid optical networks, in *Proceedings of IEEE/OSA Optical Fiber Communication Conference (OFC)*, San Francisco (2014)
13. A. Castro, L. Velasco, M. Ruiz, J. Comellas, Single-path provisioning with multi-path recovery in flexgrid optical networks, in *International Workshop on Reliable Networks Design and Modeling (RNDM)* (2012)
14. L. Velasco, F. Paolucci, L.L. Gifre, A. Aguado, F. Cugini, P. Castoldi, V. López, First experimental demonstration of ABNO-driven in-operation flexgrid network re-optimization, post-deadline paper in IEEE/OSA Optical Fiber Communication Conference (OFC) (2014)
15. H. Gredler, J. Medved, S. Previdi, A. Farrel, S. Ray, *North-Bound Distribution of Link-State and TE Information Using BGP*, IETF draft, work in progress (2014) November 2015
16. L. Velasco, A. Castro, M. Ruiz, G. Junyent, Solving routing and spectrum allocation related optimization problems: from off-line to in-operation flexgrid network planning, (Invited Paper). J. Lightwave Technol. **32**(16), 2780–2795 (2014)
17. L.L. Gifre, F. Paolucci, L. Velasco, A. Aguado, F. Cugini, P. Castoldi, V. López, First experimental demonstration of ABNO-driven in-operation flexgrid network re-optimization. J. Lightwave Technol. **33**(3), 618–624 (2015)

# Index





V. López, L. Velasco (eds.), *Elastic Optical Networks*, Optical Networks,
DOI 10.1007/978-3-319-30174-7