UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

PhD program in Signal Theory and Communications

Study and Application of Spectral Monitoring Techniques for Optical Network Optimization

Author Fabiano LOCATELLI

Supervisor Dr. Josep M. FÀBREGA
Co-supervisor Dr. Michela SVALUTO MOREOLO
Co-supervisor Dr. Konstantinos CHRISTODOULOPOULOS

Tutor Prof. Salvatore SPADARO

A thesis submitted in fulfillment of the requirements for the International doctoral degree

Teoria del Senyal i Comunicacions

Barcelona, September 2021

Abstract

One of the possible ways to address the constantly increasing amount of heterogeneous and variable internet traffic is the evolution of the current optical networks towards a more flexible, open, and disaggregated paradigm. In such scenarios, the role played by Optical Performance Monitoring (OPM) is fundamental. In fact, OPM balances performance and specification mismatches resulting from the disaggregation adoption and provides the control plane with the necessary feedback to grant the optical networks an adequate automation level. Therefore, new flexible and cost-effective OPM solutions are needed, as well as novel techniques to extract the desired information from the monitored data and process and apply them.

In this dissertation, we focus on three aspects related to OPM. We first study a monitoring data plane scheme to acquire the high-resolution signal optical spectra in a nonintrusive way. In particular, we propose a coherent detection-based Optical Spectrum Analyzer (OSA) enhanced with specific Digital Signal Processing (DSP) to detect spectral slices of the considered optical signals.

Then, we identify two main placement strategies for such monitoring solutions, enhancing them using two spectral processing techniques to estimate signal- and optical filter-related parameters. Specifically, we propose a way to estimate the Amplified Spontaneous Emission (ASE) noise or its related Optical Signal-to-Noise (OSNR) using optical spectra acquired at the egress ports of the network nodes and the filter central frequency and 3/6 dB bandwidth, using spectra captured at the ingress ports of the network nodes. To do so, we leverage Machine Learning (ML) algorithms and the function fitting principle, according to the considered scenario. We validate both the monitoring strategies and their related processing techniques through simulations and experiments. The obtained results confirm the validity of the two proposed estimation approaches. In particular, we are able to estimate in-band the OSNR/ASE noise within an egress monitor placement scenario, with a Maximum Absolute Error (MAE) lower than 0.4 dB. Moreover, we are able to estimate the filter central frequency and 3/6 dB bandwidth, within an ingress optical monitor placement scenario, with an MAE lower than 0.5 GHz and 0.98 GHz, respectively. Based on such evaluations, we also compare the two placement scenarios and provide guidelines on their implementation. According to the analysis of specific figures of merit, such as the estimation of the Signal-to-Noise Ratio (SNR) penalty introduced by an optical filter, we identify the ingress monitoring strategy as the most promising. In fact, when compared to scenarios where no monitoring strategy is adopted, the ingress one reduced the SNR penalty estimation by 92%.

Finally, we identify a potential application for the monitored information. Specifically, we propose a solution for the optimization of the subchannel spectral spacing in a superchannel. Leveraging convex optimization methods, we implement a closed control loop process for the dynamical reconfiguration of the subchannel central frequencies to optimize specific Quality of Transmission (QoT)-related metrics. Such a solution is based on the information monitored at the superchannel receiver side. In particular, to make all the subchannels feasible, we consider the maximization of the total superchannel capacity and the maximization of the minimum superchannel subchannel SNR value. We validate the proposed approach using simulations, assuming scenarios with different subchannel numbers, signal characteristics, and starting frequency values. The obtained results confirm the effectiveness of our solution. Specifically, compared with the equally spaced subchannel scenario, we are able to improve the total and the subchannel minimum SNR values of a four subchannel superchannel, of 1.45 dB and 1.19 dB, respectively.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Dr. Josep Maria Fàbrega and Dr. Michela Svaluto Moreolo from CTTC, and Dr. Konstantinos Christodoulopoulos from Nokia Bell Labs, for their continuous support during my Ph.D. studies. Without your help and guidance, this work would not have been possible.

I would also like to extend my gratitude to all the people involved in the ONFIRE project. Special thanks to Dr. Laia Nadal for her valuable advice and contributions and Prof. Salvatore Spadaro for his help and academic mentorship. I also wish to thank Dr. Raul Muñoz for his support throughout the project.

I must also thank Javier Vílchez from CTTC, Karsten Schuh, Dr. Fred Buchali, and Dr. Qian Hu from Nokia Bell Labs, for their assistance during lab experiments.

Many thanks to Ankush Mahajan, who accompanied me along this journey, for his friendship and all the constructive advices he gave me in these years.

I also wish to thank Prof. Pierpaolo Boffi, who, back in 2017, gave me the opportunity to work with Dr. Michela Svaluto Moreolo for the first time. That collaboration started the chain of events that brought me to this point today.

I would also like to thank my family and lifelong friends for their unfailing support.

I would also like to acknowledge the European Commission for supporting my work with a Marie Skłodowska-Curie Fellowship.

Finally, I would like to thank all the people I met over the past three years, there are too many of you to list, but you are definitely the ones who made this experience unforgettable!

Contents

List of Figures

List of Tables

List of Abbreviations

AAH	Asynchronous Amplitude Histogram
ADC	Analog-to-Digital Converter
ADTS	Asynchronous Delay-Tap Sampling
AGC	Automatic Gain Control
AH	Amplitude Histogram
AI	Artificial Intelligence
ANN	Artificial Neural Network
APC	Automatic Power Control
API	Application Programming Interface
ASE	Amplified Spontaneous Emission
BER	Bit Error Ratio
BOSA	Brillouin Optical Spectrum Analyzer
BV-WSS	Bandwidth Variable-Wavelength Selective Switch
B2B	Back-to-Back
CAPEX	Capital Expenditure
CD	Chromatic Dispersion
CDF	Cumulative Distribution Function
COSA	Coherent Optical Spectrum Analyzer
DAC	Digital-to-Analog Converter
DEMUX	Demultiplexer
DSP	Digital Signal Processing
DWDM	Dense Wavelength Division Multiplexing
EDFA	Erbium-Doped Fiber Amplifier
EID	European Industrial Doctorate
FCE	Filter Cascade Effect
FEC	Forward Error Correction
FFT	Fast Fourier Transform
GMM	Gaussian Mixture Model
GN	Gaussian Noise
GP	Gaussian Process
GPR	Gaussian Process Regression
HPF	High-Pass Filter
HT	Hilbert Transform
ICP	Internet Content Provider
ICT	Information Communication Technology
IEC	International Electrotechnical Commission

ITN	Innovative Training Network
K-MC	K-Mean Clustering
K-NN	K-Nearest Neighbor
LPF	Low-Pass Filter
MAE	Maximum Absolute Error
MAX	Maximum Error
MIN	Minimum Error
ML	Machine Learning
MP	Muxponder
MSE	Mean Squared Error
MUX	Multiplexer
MZDI	Mach-Zehnder Delay Interferometer
M2M	Machine-to-Machine
NBI	North Bound Interface
NEB	Noise Equivalent Bandwidth
NF	Noise Figure
NFV	Network Function Virtualization
NLI	Nonlinear Interference
OCM	Optical Channel Monitor
ODM	Original Design Manufacturer
OE	Optical-Electrical
OEO	Optical-Electrical-Optical
OLS	Optical Line System
ONFIRE	Optical Network for Innovation Research and Experimentation
OPEX	Operation Expenditure
OPM	Optical Performance Monitoring
OSA	Optical Spectrum Analyzer
OSNR	Optical Signal-to-Noise Ratio
OTC	Optical Testing Channel
PBS	Polarization Beam Splitter
PCA	Principal Component Analysis
PDF	Probability Density Function
PDL	Polarization-Dependent Loss
PDM	Polarization-Division Multiplexing
PLM	Physical Layer Model
PM	Polarization Multiplexed
PMD	Polarization Mode Dispersion
PRBS	Pseudorandom Binary Sequence
PSD	Power Spectral Density
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service

QoT	Quality of Transmission
QPSK	Quadrature Phase Shift Keying
Q-L	Q-Learning
RBF	Radial Basis Function
RF	Radio Frequency
RL	Reinforcement Learning
ROADM	Reconfigurable Optical Add/Drop Multiplexer
RRC	Root Raised Cosine
RX	Receiver
SBI	South Bound Interface
SBS	Stimulated Brillouin Scattering
SCI	Self-Channel Nonlinear Interference
SDM	Space Division Multiplexing
SDN	Software-Defined Networking
SL	Supervised Learning
SMO	Sequential Minimal Optimization
SNR	Signal-to-Noise Ration
SOP	State of Polarization
SP	Switchponder
SSMF	Standard Single-Mode Fiber
SUT	Signal Under Test
SVM	Support Vector Machine
SVR	Support Vector Regression
S-BVT	Sliceable-Bandwidth Variable Transceiver
S-OM	Self-Organizing Maps
TF	Transfer Function
TIP	Telecom Infra Project
TP	Transponder
TX	Transmitter
UL	Unsupervised Learning
VOA	Variable Optical Attenuator
WDM	Wavelength Division Multiplexing
WSS	Wavelength Selective Switch
XCI	Cross-Channel Nonlinear Interference
YANG	Yet Another Next Generation

Chapter 1

Introduction

The work presented in this thesis was carried out in the context of the "Future Optical Network for Innovation Research and Experimentation (ONFIRE)" project. The ONFIRE project is a Marie Skłodowska-Curie Action as an Innovative Training Network (ITN) European Industrial Doctorate (EID) supported by the European Union Horizon 2020 Research and Innovation Programme. ONFIRE focuses on the study, design and experimental evaluation of cost-effective and flexible solutions for future automated and disaggregated optical networks.

1.1 Motivation

During the past few decades, optical networks continuously evolved to handle an ever-increasing traffic growth demand. Nowadays, such a trend is far from over. According to [1], nearly two-thirds of the global population will have Internet access by 2023, when 5.3 billion total Internet users are expected. In 2018, this amount was more than 20% lower. In western Europe alone, 370 million Internet users are expected by 2023, up from 345 million in 2018. In particular, Machine-to-Machine (M2M) will constitute half of the global connections by 2023, with connected home applications such as home automation and video surveillance expected to be the most popular M2M connection. Finding ways to address such an enormous amount of heterogeneous and variable data requests in a cost-effective but reliable manner is fundamental to sustain the expected traffic growth. The increase in transmitter data rates and the use of coherent-based receivers are certainly two viable solutions to address this evolution.

Moreover, Space Division Multiplexing (SDM) represents a promising technology to raise the network capacity limits by employing multi-fiber links or multicore/multimode fibers [2]. In parallel, keeping the costs low is definitely the most difficult aspect to deal with. One of the possible ways to

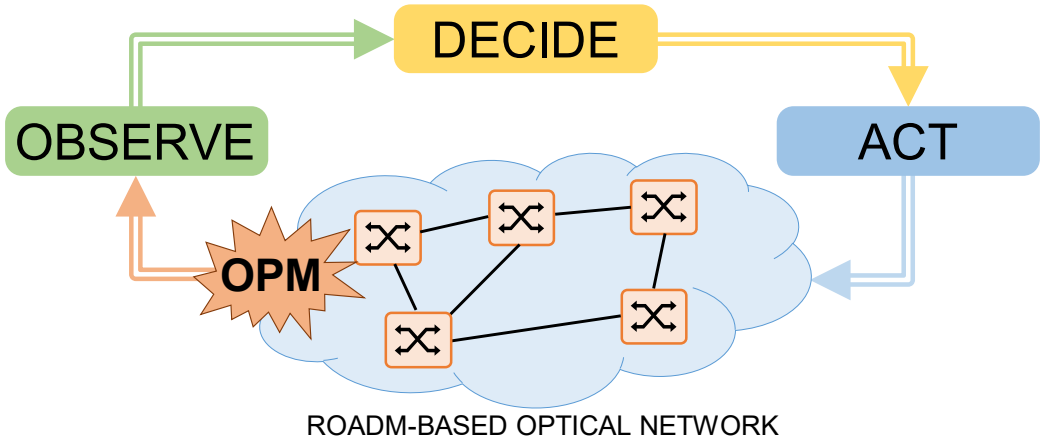

FIGURE 1.1. The role of OPM within the Observe-Decide-Act loop. OPM allows to close the control loop in which a central controller, based on the network monitored information, decides what actions to perform in order to optimize the considered ROADM-based optical network.

tackle such a challenge is represented by the disaggregation approach, which in the past few years gained a foothold in data center environments and recently has become appealing also for optical network scenarios. In fact, telecommunication operators also started to consider disaggregating the different elements of the optical transport networks into off-the-shelf components [3]. According to the disaggregation paradigm, several elements of the network, including the software to manage them, are provided by different vendors. This enables a relatively free market, where the network operators can choose the best equipment from different providers overcoming the vendor lock-in and reducing their cost [4]. In order to cope with the network specification and performance mismatches introduced by elements provided by different vendors, appropriate advanced Optical Performance Monitoring (OPM) techniques are needed.

Moreover, in disaggregated but also in traditional/aggregated optical networks, ensuring a suitable degree of automation is crucial [5]. This is necessary to guarantee an adequate answer to the always more demanding end-to-end Quality of Transmission (QoT) and Quality of Service (QoS) requests and mainly to improve the efficiency of the optical connections. Typically, following a *set and forget* strategy, optical connections are provisioned with high margins, which allow them to remain almost untouched until their end-of-life and rarely reconfigured. Instead, automation would enable dynamic adjustment and control of the margins of existing and future connections [6], allowing their reduction. Indeed, a reduction of the margins required to operate the networks translates to an increment of network efficiency and an erosion of the cost [7]. To cope with all these aspects, the

control plane constantly requires a certain amount of updated network-related information, based on which new decisions are taken. OPM again fulfills this role, delivering the feedback needed to close the so-called "*observe-decide-act*" loop and enabling the optical networks automation, as shown in FIGURE 1.1 [6]. In particular, the role of OPM within such a loop is to assist the observation phase, constantly providing the required optical network information. Then, a central controller, which acts as a coordinating entity for the network, decides how to react to such monitored data. Finally, if one or more actions are required, the central controller interacts with the physical network, providing adequate adjustments and eventually operating it in a real-time manner in a closed control loop mode.

Furthermore, during the past years, the classic Dense Wavelength Division Multiplexing (DWDM) paradigm evolved towards a new flexible approach, named flex-grid DWDM or simply flex-grid, which was defined in [8]. Thanks to such a new paradigm, frequency slots with different widths became available. Key elements of the deployed flex-grid channels are the Reconfigurable Optical Add/Drop Multiplexers (ROADMs), which allows individual wavelengths to pass through the various nodes of the network or to be added or dropped. To do so, the ROADM-based nodes, according to their architecture, include/adopt a certain number of optical filters. Therefore, along with all the benefits deriving from such flexibility, new challenges connected with the optical impairments introduced by the filters need to be faced. Accordingly, advanced OPM techniques able to constantly monitor filter and signal-related parameters need to be developed. Finally, since the flex-grid channels also allow the development of superchannels, in which several single channels are jointly transmitted and received, new ways to control and optimize the frequency allocation of the superchannel subchannels are necessary. OPM can embody the critical factor of such optimization, providing the feedback required by the algorithms to calculate the most suitable frequency configuration.

Due to these reasons, OPM represents one of the most important elements of future optical networks [9]. Many physical parameters related to the transmitted signals and the components of an optical network can be estimated and acquired through OPM [9], [10]. Among the several available OPM techniques, this thesis mainly focuses on those that leverage optical spectral data to retrieve different network-related parameters. In addition, since one of the key requirements that future OPM techniques must comply with is cost-effectiveness, this work deals with developing new cost-efficient solutions to monitor the optical network parameters, identifying a suitable placement strategy and possible ways to process such information. In particular, an ideal candidate to support the data processing phase is Machine

Learning (ML), a flexible and highly adaptable instrument, which recently has been widely adopted in several scientific fields, also including optical communications [11], [12]. Finally, to demonstrate the advances that OPM and adequate processing can achieve, we also present a closed control loop process that uses OPM and monitoring to optimize the performance of optical superchannels.

1.2 Objectives and Methodology

The objective of this thesis is threefold and revolves around the use of OPM for the optimization of optical networks. Specifically, on one side, we aim to investigate a novel technique to acquire high-resolution optical spectral data. On another side, we seek to develop new solutions to process the spectral data collected with cheap monitoring devices and optimize their placement within the optical network. Finally, on the last side, we provide an example of how the parameters retrieved by using OPM can be applied to optimize some of the transmission elements.

More in detail, in this thesis, we present the theoretical study of an agile optoelectronic coherent-based scheme for the acquisition of high-resolution optical spectral data. A detailed analysis of the requirements needed to eventually implement similar front-end solutions is also proposed. Then, we study how cheap and flexible optical monitors can be adopted within the optical network. In particular, we identify two possible scenarios for their placement and for each of them, we present a technique to process the related acquired spectral data. We propose an ML-based spectral processing technique to retrieve signal-related parameters from optical spectra acquired with monitoring devices placed at the egress ports of the network nodes. Moreover, we propose a spectral processing technique based on the function fitting principle to retrieve filter-related parameters from optical spectra acquired with monitoring devices placed at the ingress ports of the network nodes. A comparison between the two placement strategies and their related processing solutions is also presented. Finally, to optimize the transmission in superchannel configurations, we propose an approach based on a closed control loop process, to optimize the superchannel subchannel spectral spacing. Specifically, we implement an algorithm based on convex optimization methods that probes the physical layer and monitors the related outcomes to maximize specific figures of merit.

In this dissertation, to support all the aforementioned solutions, a thorough review of the theoretical background and the state of the art of the considered topics is also presented. In fact, an initial in-depth analysis of the investigated

subjects is essential to detect potential gaps for new technological developments. More in detail, in the initial part of this work, we review the background concepts related to the two mathematical tools employed within this thesis, namely ML and convex optimization. Moreover, the state of the art of disaggregation paradigm applied to optical networks, signal- and filter-related OPM, and spectral processing techniques are also reported.

The methodology adopted to draft the works presented in this thesis always included an initial identification phase, where a new promising approach and its eventual feasibility were assessed. After such initial analysis, the conceived solution was evaluated using simulations. To do so, two main simulation environments were considered: MATLAB [13] and VPIphotonics [14]. The first one is a well-known numerical analysis tool useful for developing numerical models and the simulation of the system under investigation. Instead, VPI is a simulation software that allows the user to design and test almost all kinds of optical elements in the transmission layer in a fast and easy way. In most cases, after verifying the conceived solution through simulations, a final experimental validation was also carried out. To this extent, the facilities made available by the ONFIRE partners were exploited. In particular, the experimental data used in this thesis were collected in experiments carried out at CTTC Optical Networks and Systems lab and in Nokia Bell Labs Stuttgart optical lab. In particular, at CTTC optical lab, the ADRENALINE network testbed was used. ADRENALINE is a photonic mesh network composed of four (white box) nodes connected by five amplified bidirectional links for a total length equal to 600 km [15].

Regarding the adopted notation, in this thesis we will denote scalars with italic lower- and upper-cases (e.g., x, Y), vectors with bold lowercases (e.g., $\mathbf{x}$, $\mathbf{y}$), and matrices with bold uppercases (e.g., $\mathbf{X}$, $\mathbf{Y}$).

1.3 Outline of the Thesis

This section provides the list of chapters composing the thesis, along with a brief description of their content. In addition, for each chapter, the papers presenting the related obtained results are also referenced. In particular, Chapter 2 and Chapter 3 are introductory. They cover the theoretical background of the main topics addressed in this dissertation and the state of the art. Instead, Chapter 4, Chapter 5, and Chapter 6 include our contributions about the identified monitoring solution, the optical monitor placement within the network and the related spectral processing techniques, and the superchannel optimization, respectively. Finally, Chapter 7 concludes the thesis.

Chapter 2 – Mathematical Framework

Chapter 2 provides a generic overview of the two main mathematical tools employed within this work. Specifically, we review the different existing ML solutions, with a particular focus on Support Vector Machine (SVM) and Gaussian Process Regression (GPR) models, which we employ in Chapter 5 to estimate from optical spectral data signal- and filter-related parameters. In addition, we introduce the convex optimization methods, particularly the gradient descent and the stochastic subgradient algorithms, which we adopt in Chapter 6 to optimize the transmission of a superchannel.

Chapter 3 – State of the Art

In Chapter 3, we present a thorough review of the state of the art of the topics addressed in this dissertation to support the technical solutions proposed in Chapters 4, 5, and 6. Specifically, this chapter introduces the concepts related to disaggregated optical networks, OPM, and superchannels. Regarding the disaggregated optical networks, we review the available operational models and the related network architectures. Concerning OPM, a particular focus is reserved to those techniques that allow monitoring the Optical Signal-to-Noise Ratio (OSNR) and the filter-related parameters. In addition, a specific analysis of the state of the art of those methods that leverage optical spectral data to estimate network-related parameters is also presented. Furthermore, concerning superchannels, we provide the motivations for their implementation and the challenges that still need to be addressed when such technology is considered. Finally, starting from the theoretical concepts introduced in this chapter, we identify the gaps that the methods we propose aim to fill, and we introduce the reference network scenario considered for the development of our solutions.

Chapter 4 – Data Plane Architectures for High-Resolution Optical Spectrum Estimation

In Chapter 4, a theoretical study of a nonintrusive, cost-effective, and coherent detection-based Optical Spectrum Analyzer (OSA) system to be deployed at the network nodes is presented. The proposed monitoring probe is agnostic to the modulation format, and it can provide an arbitrary spectral resolution.

The work reported in this chapter was presented in the following paper:

- J. M. Fàbrega, F. Locatelli, L. Nadal, K. Christodoulopoulos, M. Svaluto Moreolo, and S. Spadaro, **Data Plane Elements for Optical Performance Monitoring Agnostic to the Modulation Format**

for Disaggregated Optical Networks, in *Proc. 2020 22nd International Conference on Transparent Optical Networks (ICTON),* July 2020, Online Event, doi: 10.1109/ICTON51198.2020.9203369.

Chapter 5 – Optical Monitor Placement Strategies and Assessment

In Chapter 5, a detailed study about the placement of the optical monitors within the network is provided. In particular, the two main placement scenarios that have been identified are here described and analysed. Moreover, the solutions to estimate signal- and filter-related parameters associated with the considered scenario are presented in this chapter. Their assessment, performed with spectral data collected in simulations and experiments, is also reported. Finally, a comparison between the two proposed approaches is presented.

The methods and the results reported in this chapter were presented in the following papers:

- F. Locatelli, K. Christodoulopoulos, J. M. Fàbrega, M. Svaluto Moreolo, and S. Spadaro, **Machine learning-based in-band OSNR estimation from optical spectra**, *IEEE Photonics Technology Letters (PTL)*, Volume 31, Issue 24, Pages 1929-1932, December 2019, doi: 10.1109/LPT.2019.2950058.

- F. Locatelli, K. Christodoulopoulos, J. M. Fàbrega, M. Svaluto Moreolo, L. Nadal, and S. Spadaro, **Experimental demonstration of a machine learning-based in-band ONSR estimator from optical spectra**, in *Proc. 2020 International Conference on Optical Network Design Modelling (ONDM)*, May 2020, Online Event, doi: 10.23919/ONDM48393.2020.9133001.

- F. Locatelli, K. Christodoulopoulos, J. M. Fàbrega, M. Svaluto Moreolo, L. Nadal, and S. Spadaro, **Filter features extraction from optical spectra**, in *Proc. 2020 European Conference on Optical Communication (ECOC)*, December 2020, Online Event, doi: 10.1109/ECOC48923.2020.9333342.

- F. Locatelli, K. Christodoulopoulos, M. Svaluto Moreolo, J. M. Fàbrega, L. Nadal, and S. Spadaro, **Spectral processing techniques for efficient monitoring in optical networks,** *IEEE/OSA Journal of Optical Communications and Networking (JOCN)*, Volume 13, Issue 07, Pages 158-168, July 2021, doi: 10.1364/JOCN.418800.

Chapter 6 – Frequency Optimization in Superchannels

Chapter 6 describes a possible application for the data acquired through OPM. In particular, it focuses on the frequency optimization of the superchannel subchannels. We propose a solution based on a probe and monitor approach, in which specific figures of merit are maximized, leveraging monitoring information provided by the superchannel receivers.

The work reported in this chapter was presented in the following paper:

- F. Locatelli, K. Christodoulopoulos, M. Svaluto Moreolo, J. M. Fàbrega, L. Nadal, A. Mahajan, and S. Spadaro, **Feedback-Based Channel Frequency Optimization in Superchannels,** submitted to *IEEE/OSA Journal of Optical Communications and Networking (JOCN)*, September 2021.

Chapter 7 – Conclusions and Future Works

This final chapter concludes the dissertation, summarizing the obtained results and outlining the perspective for future works.

Chapter 2

Mathematical Framework

In this chapter, we provide an overview of the mathematical framework adopted in this thesis. In particular, we describe two efficient and widely employed mathematical tools that ease the resolution of complex problems in many scientific areas, namely ML and mathematical convex optimization methods.

This chapter is organized as follows. In Section 2.1, ML is introduced and its application to optical networks is motivated. In particular, special emphasis is put on SVM and on Gaussian Process (GP)-based algorithms. Section 2.2 discusses the convex optimization problems, and the gradient descent and subgradient algorithms to solve them, with a special focus on the stochastic subgradient method.

2.1 Machine Learning

In this section, we provide a brief introduction on the available ML solutions and on their potential application in optical networks. In addition, we present an overview of the different ML-based algorithms, that were used in this thesis to estimate signal- and filter-related parameters. Specifically, in Section 5.2, ML is leveraged for the prediction of the OSNR/Amplified Spontaneous Emission (ASE) noise, a crucial parameter for the optimization of the physical layer performance.

2.1.1 Introduction

As per many technological scientific fields, also in the context of optical networks plenty of potential applications for ML have been proposed in the past few years. On the one hand, aspects such as the nonlinear signal propagation through the optical channel and the plethora of complex systems

and subsystems used to build the network (Transponders (TPs), fiber spans, amplifiers, and ROADM nodes) make the optical network modelling and its optimization a really challenging task. On the other hand, the high degree of flexibility enabled by the employment of technologies, such as the flex-grid channels and tuneable TPs, in wavelength switched optical networks, have increased the complexity of the resource allocation problem. When properly combined with features such as OPM, ML could represent a key element to fill these gaps. In fact, ML-based algorithms allow inferring features that otherwise could not be easily retrieved from the large amount of data monitored within the network [11], since they can solve problems in which the system is analytically intractable, or the solution space is too vast to be evaluated [12]. Tasks like QoT estimation, link and network capacity optimization, low-margin network design, optical network automation, and pattern recognition for failure identification and prediction from raw monitored optical data are just few examples of the many application fields that ML has within optical networks. Of course, to enable the adoption of ML in deployed optical networks, some aspects need to be further improved [11], [12], and [16]. These include the lack of available and optimized datasets for model training and testing and the necessity for ML-driven solutions to follow standardization processes. In addition, special care must be taken to trade off the complexity and computational effort with performance. Moreover, further research is required to explore the feasibility and performance of an ML-based unified control framework approach, in which the estimated parameters are taken into account when making new network-related decisions (e.g., where to route a new lightpath, or when to modify the parameters of the transmission). Finally, the development of visualization tools that allow the end-users to access and understand easily the ML algorithm outputs is also needed.

2.1.2 Overview of Machine Learning Algorithms

ML is a branch of Artificial Intelligence (AI), and according to [17], it can be defined as a computer program that learns from experience E with respect to some class of tasks T and performance measure P, whose performance at tasks T, as measured by P, improves with experience E. ML techniques can be particularly useful in scenarios where an explicit mathematical or physical problem description is too hard to be modelled with other classical approaches. Usually, as schematically represented in FIGURE 2.1, depending on the considered learning process, ML algorithms can be divided into three main categories: Supervised Learning (SL), Unsupervised Learning (UL), and Reinforcement Learning (RL) [18]. In addition, a second classification of the ML methods can be done based on the objectives of the learning task. Such objectives can be to assign a class to a set of new input data, or to estimate any

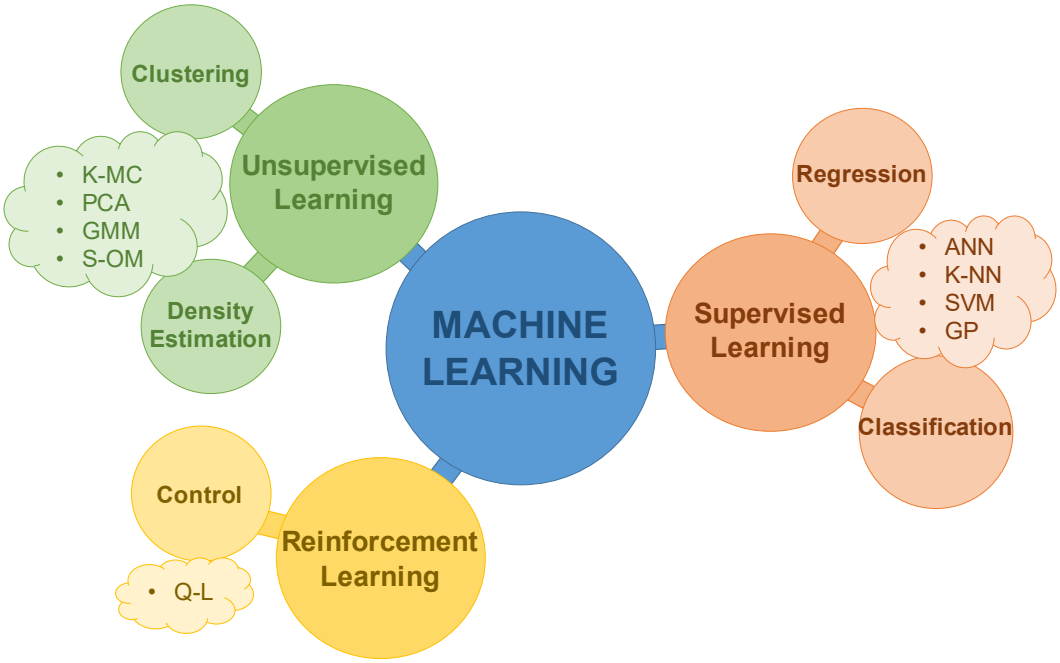

FIGURE 2.1. Machine Learning categories, their main tasks, and some of their most representative algorithms. ANN: Artificial Neural Network, K-NN: K-Nearest Neighbors, SVM: Support Vector Machine, GP: Gaussian Process, K-MC: K-Mean Clustering, PCA: Principal Component Analysis, GMM: Gaussian Mixture Model, S-OM: Self-Organizing Maps, Q-L: Q-Learning.

eventual relationship among given input data. The ML methods belonging to the first group are known as classification methods, while those related to the latter are known as regression methods. In the following paragraphs, we will briefly introduce each one of the three main ML categories, providing examples of their related algorithms and their most common applications within optical networks.

Supervised Learning

SL is a ML category in which the input data are labelled with their known corresponding output values. Therefore, relying on such labelled data, the ML models are trained to find proper generalized mapping functions that predict the output values, which correspond to new and unseen input data. Depending on the output data, the learning problem can be categorized as a regression or a classification problem. The former considers output variables, which are continuous values, whereas the latter deals with discrete output values. For example, the task of estimating the OSNR value of an optical channel from its optical spectrum falls under the regression problem category [19]. On the contrary, recognizing the modulation format of a connection from its eye-diagram represents an example of a classification problem, since the values

assumed by the output variables (the possible modulation formats in this case) are finite [20].

Depending upon the number of parameters to be used by the mapping function, two main models can be identified within the SL algorithm family: the parametric and the nonparametric ones. In parametric models, a fixed number of parameters, which is independent from the training data size, is considered. On the contrary, in nonparametric models, the number of parameters depends on the training data, resulting in most cases in more powerful and accurate predictions. However, this accuracy comes with certain drawbacks, such as the necessity of bigger data sets for the training phase or the time required to train the model. On the other hand, the parametric models are typically faster but less accurate. The most common SL parametric models are Artificial Neural Network (ANN) and Convolutional Neural Network. Regarding the nonparametric models, some of the most common are K-Nearest Neighbors (K-NN), SVM, Gaussian Process Regression (GPR), and decision trees techniques, as also shown in FIGURE 2.1. Specifically, SVM and GPR algorithms will be discussed in detail in Sections 2.1.3 and 2.1.4 of this thesis, respectively.

In addition to the already mentioned OSNR estimation and modulation format recognition [21], typical optical networks applications for supervised ML algorithms can include fault identification [22], traffic analysis [23], [24], QoT estimation [25] and the prediction of several other classical network performance parameters, such as symbol rate [26], Chromatic Dispersion (CD) [27], and Polarization Mode Dispersion (PMD) [28].

Unsupervised Learning

In many cases, due to the complexity of the labelling operation, having a large labelled dataset is not a viable option. In such situations UL represents a better choice, since the algorithms directly search for patterns and structures in unlabelled input data. In other words, the UL models have to learn by themselves, without any external guidance.

The most common task that UL algorithms can address is clustering, in which the data are grouped based on their intrinsic characteristics. Samples belonging to the same cluster will exhibit high similarities between them. Conversely, samples of different clusters will exhibit low degrees of similarity. In addition, UL algorithms can also be employed to determine the distribution of data within the input space, also known as density estimation. As also shown in FIGURE 2.1, representative UL algorithms are K-Mean Clustering (K-MC), Principal Component Analysis (PCA), Gaussian Mixture Model (GMM), and Self-Organizing Maps (S-OM). Representative applications of such models in optical networks include spectral slot identification [29], traffic prediction

[30], and nonlinearity mitigation [31]. Moreover, UL algorithms are also often employed as preprocessing tools before applying SL algorithms [12].

Reinforcement Learning

In RL, a decision maker named agent directly interacts with the environment, receiving feedbacks that activate the learning process. The environment rewards or penalizes the agent according to the actions it has undertaken. Hence, the goal of the agent is to maximize the received rewards over time. To do so, the agent has to discover which actions yield rewards and which others not, following a trial-and-error process. Such process could take a long time to be completed and this is why RL should be avoided in those scenarios where error-free decisions are a crucial requirement. On the contrary, employing RL might be beneficial when searching for optimal solutions in constantly changing environments.

Although RL is not a novel paradigm within ML, its application in the physical layer of optical communication systems has not yet been fully investigated. Mainly, RL algorithms are used in network self-configuration, more specifically for path computation [32] and service reconfiguration tasks [33]. The most common RL technique is Q-Learning (Q-L).

2.1.3 Support Vector Machine

SVM concepts were first identified by Vladimir Vapnik and his colleagues in 1992 [34]. As mentioned in the previous section, SVM is a nonparametric SL approach that can be used to address both regression and classification problems, although it is best known for its application in the second category [18]. When applied to regression problems, SVM is also referred to as Support Vector Regression (SVR). The main advantage of employing SVM, both for classification or regression, is that only few of the training data are used to build the decision surfaces on which the predictions are based. Thus, once trained, the rest of the training data become irrelevant, making SVM and SVR very memory efficient and attractive approaches. In addition, as described in the next paragraphs, SVM and SVR represent a flexible solution to handle nonlinear problems, thanks to the possibilities guaranteed by the kernel functions. On the other hand, SVM and SVR are not suitable for handling very large or noisy datasets [35].

In this section, we will first introduce the SVM algorithm for classification and then we will generalize its description to cover also the regression aspect. The derivations proposed in this section are adapted from [18], [35], and [36].

Support Vector Machine for Classification

As per many other classification algorithms, the goal of SVM is also to define decision boundaries that separate the points belonging to different data classes. If we consider a space with more than two dimensions, such decision boundaries become surfaces, and they are generally referred to as hyperplanes. As one can imagine, many possible hyperplanes exist and can be identified when classifying the data points, as depicted in FIGURE 2.2(a). SVM achieves such classification by searching for those hyperplanes that maximize their distance to the borderline points of each class, keeping at the same time an equal margin between them, as shown in FIGURE 2.2(b). The borderline points of each class are known as support vectors, and in FIGURE 2.2(b) we represented them as solid dots. In particular, in the example of FIGURE 2.2(b), the blue data points class has two support vectors, while the orange one has only one.

In SVM, the role assumed by the support vectors within the training set is crucial, since the hyperplane choice is only influenced by their position. Moreover, in order to ease the hyperplane identification, often the input data are first transformed into a higher-dimensional space, also known as feature space. This operation allows to classify data which eventually are nonlinearly separable. However, in case of large dataset, such transformation can take a very long time to be completed. Thus, to avoid this issue, the so-called kernel trick can be employed. Indeed, leveraging a kernel function, the algorithm can avoid to explicitly transform the data, which would be extremely computationally expensive, but can still exploit the benefits that would derive

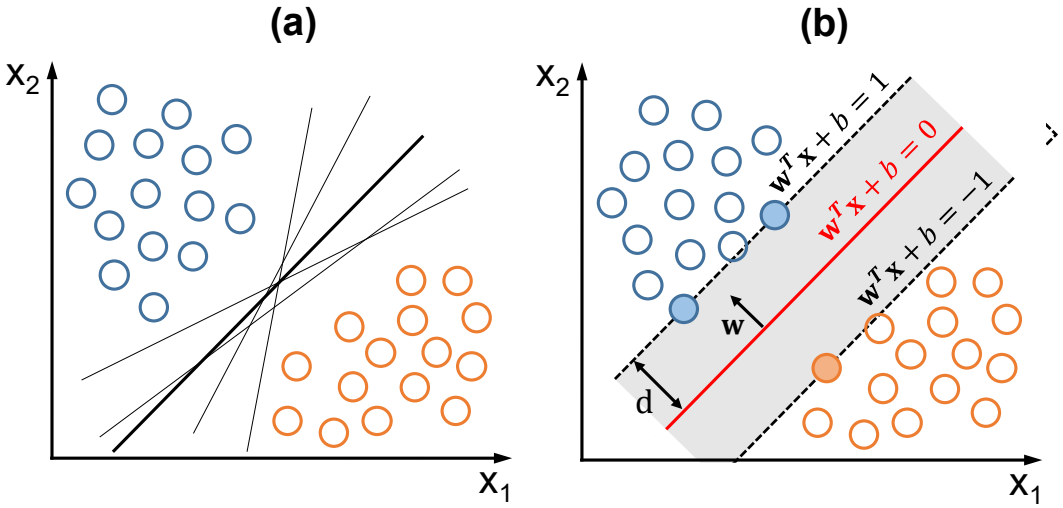

FIGURE 2.2. (a) Hyperplane identification: many options are available for dividing the two classes of points. (b) Optimal hyperplane (red line) separating two data classes (blue and red circles). The solid circles represent the borderline points of each class, also known as support vectors.

from the transformation. The most common kernel functions adopted in SVM are the polynomial, the hyperbolic and the Gaussian one, which is also known as Radial Basis Function (RBF).

In the following paragraphs, we will describe how SVM is trained and how it works, translating the basic concepts introduced above into mathematical formulas.

Following the notation of FIGURE 2.2(b), a generic hyperplane of the feature space can be defined as

$$\mathbf{w}^{\mathrm{T}}\mathbf{x} + b = 0\,, \tag{2.1}$$

where $\mathbf{x}$ is the vector representing the input samples of the SVM, such that $\mathbf{x} \in \mathbb{R}^n$ with n representing the dimension of the $\mathbf{x}$ vector (e.g., in the example of FIGURE 2.2(b) $n = 2$), $\mathbf{w}$ is the weight vector representing the polynomial coefficients, $\mathbf{w}^{\mathrm{T}}$ is the transpose of $\mathbf{w}$, and b represents the so-called bias parameter. $\mathbf{w}$ and b are the SVM parameters to be learnt during the training phase. SVM goal is to find a hyperplane that maximizes the margin d between the hyperplane itself and the support vectors. It can be demonstrated that such margin can be expressed as

$$d = \frac{1}{\|\mathbf{w}\|}\,, \tag{2.2}$$

where $\|\mathbf{w}\| = \mathbf{w}^{\mathrm{T}}\mathbf{w}$ represents the Euclidean norm of the weight vector $\mathbf{w}$.

Therefore, the research of the optimal hyperplane reduces to the research of its corresponding values $\mathbf{w}$ and b for which the distance d is maximized, under the constraint that all the input data points are classified correctly. Such conditions can be mathematically derived as follows. Assuming to classify the samples belonging to the blue class with 1 and those belonging to the orange class with -1, one can write

$$y_l = \begin{cases} -1, & \text{if } \mathbf{w}^{\mathrm{T}}\mathbf{x}_l + b \le -1 \\ 1, & \text{if } \mathbf{w}^{\mathrm{T}}\mathbf{x}_l + b \ge 1 \end{cases}\,, \tag{2.3}$$

where y_l represents the class assigned by the SVM algorithm (i.e., the SVM output) to the l-th input $\mathbf{x}_l$. Then, the condition of making only correct decisions for all the input data points can be written as

$$y_l(\mathbf{w}^{\mathrm{T}}\mathbf{x}_l + b) \ge 1, \qquad l = 1,2,\dots,L\,, \tag{2.4}$$

where L represents the total length of the input data.

Thus, the optimization problem related to the optimal hyperplane research becomes

$$min_{\mathbf{w},b}\ \|\mathbf{w}\| \tag{2.5}$$

$$\text{subject to} \quad y_l(\mathbf{w}^{\mathrm{T}}\mathbf{x}_l + b) \geq 1, \; l = 1,2,\dots,L \; .$$

Equation 2.5 represents a quadratic optimization problem also known as hard-margin SVM. As previously mentioned, in case of nonlinearly separable data a transformation to a higher dimensional feature space must be considered. For such purpose, a mapping function $\varphi(\cdot)$ has to be used and the optimization problem described in Equation 2.5 becomes

$$\begin{aligned} & min_{\mathbf{w},b} \; \|\mathbf{w}\| \\ & \text{subject to} \quad y_l(\mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_l) + b) \geq 1, \; l = 1,2,\dots,L \; . \end{aligned} \tag{2.6}$$

In addition, since there is the possibility that no such hyperplane exists, the constraints can be relaxed employing the so-called slack variables ξ. Considering that each l-th input sample needs its own slack variables, the optimization problem of Equation 2.6 becomes

$$\begin{aligned} & min_{\mathbf{w},b} \quad \|\mathbf{w}\| + C\sum_{l=1}^{L}\xi_l \\ & \text{subject to} \quad y_l(\mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_l) + b) \geq 1 - \xi_l, \qquad l = 1,2,\dots,L \\ & \qquad\qquad \xi_l \geq 0, \qquad\qquad\qquad\qquad l = 1,2,\dots,L \; , \end{aligned} \tag{2.7}$$

where C, also known as penalty of the error term, is a hyperparameter representing the cost assigned to any misclassification (i.e., a higher C will imply a more strict data separation). The quadratic optimization problem described by Equation 2.7 is also known as soft-margin SVM.

As previously mentioned, often the data to be classified are nonlinear separable. Therefore, a shift towards a higher dimensional features space is required. In order to avoid computing the mapping function $\varphi(\cdot)$, the so-called kernel trick can be leveraged. To do so, the primal problem described by Equation 2.7 has to be transformed in a dual problem by means of the Lagrange dual formulation. This will allow the chosen nonlinear kernel function to be employed.

The most common solutions available to solve the optimization problem described by Equation 2.7 are the CVXOPT Python package [37] for convex optimization and the Sequential Minimal Optimization (SMO) approach. In Section 2.2, we will also provide a brief overview of convex optimization problems. For a given input x_l, the output of the optimization problem will be a set of weights **w**, whose linear combination will correspond to the predicted class y_l. The analysis presented above is valid for binary classification problems only. To address multiclass classification problems, they simply need to be reduced in a series of binary ones.

Support Vector Machine for Regression

As for the SVM for classification, SVR also reduces to an optimization problem. However, unlike the previous case, the minimization of $\mathbf{w}$ is subject to the condition that the label y_l (which now assumes continuous values) assigned to the l-th input x_l deviates from the hyperplane within the accuracy boundary ϵ, for all the samples l, as shown in FIGURE 2.3(a). In other words, as long as the prediction errors fall within the interval defined by ϵ, the prediction will be considered accurate. Therefore, by tuning ϵ, the exact accuracy level of the model can be chosen. Consequently, while in SVM for classification the hyperplane represented the optimal decision boundary, in SVR it represents the objective fitting function $f(x)$ to be identified. In addition, in order to minimize any possible deviation larger than ϵ, slack variables can be also considered for those points lying outside the accuracy boundaries, as shown in FIGURE 2.3(b). Bearing in mind such assumptions, the optimization problem described by Equation 2.7 can be restated as

$$\begin{aligned} & min_{\mathbf{w},b} \quad \|\mathbf{w}\| + C\sum_{l=1}^{L}\xi_l \\ \text{subject to} \quad & |y_l - f(x)| \leq \epsilon + \xi_l, \qquad l = 1,2,\dots,L \\ & \xi_l \geq 0, \qquad\qquad\qquad l = 1,2,\dots,L\,, \end{aligned} \tag{2.8}$$

where ξ_l represents the slack variable associated to the l-th input point which falls outside the error margin defined by the hyperparameter ϵ, C is the

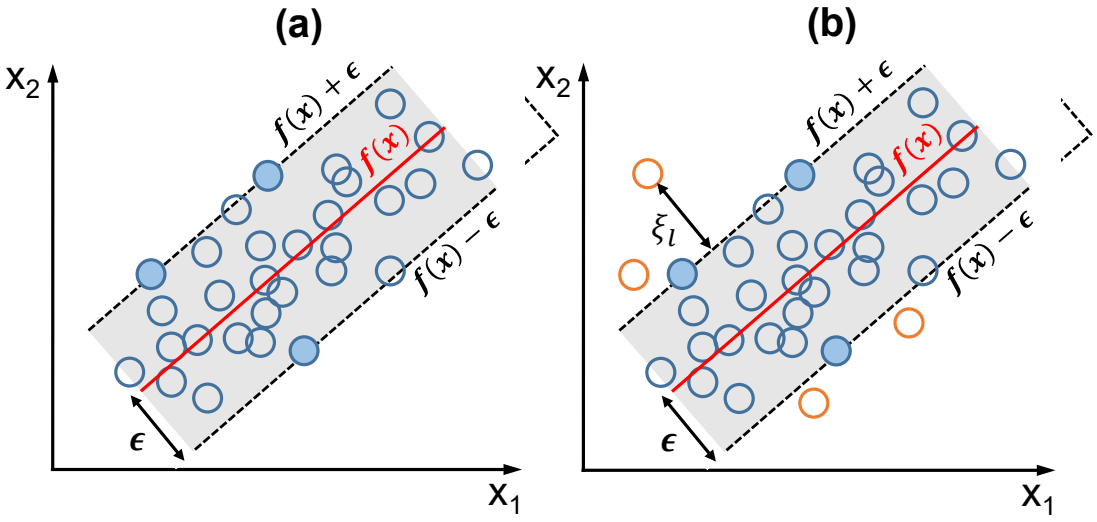

FIGURE 2.3. (a) SVMs for regression. Each circle represents an input point, the solid circles represent the support vectors, the red line represents the fitting function $f(x)$, and the dashed black lines represent the error margins defined by ϵ. (b) The orange circles represent the input points falling outside the error margins ϵ, which are taken into account by slack variables ξ.

hyperparameter representing the tolerance for such points (i.e., a higher C will imply a higher tolerance for points lying outside the error margin), and $f(x) = (\mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_l) + b)$ represents the line (or the hyperplane) fitting the input data. Again, as per the classification SVM, when dealing with nonlinear data a transformation into a higher dimensional space should be considered. To do so, the same kernel functions presented in the previous section can be employed.

In this thesis, we leverage SVR to estimate the OSNR values of optical connections starting from their optical spectra. Such implementation will be detailed within an egress monitoring scenario in Section 5.2.

2.1.4 Gaussian Process

GPs can be leveraged to address both regression and classification problems [18]. In this section we mainly focus on the first application, which is also known as GPR, providing a description of its operation principles and listing the aspects that make it an interesting and attractive approach for regression purposes. In addition, a brief introduction about GP for classification is also presented. The derivations proposed in this section are adapted from [38].

A GP is a stochastic process, that is a collection of random variables such that every finite linear combination of them have a joint Gaussian distribution. A GP $f(\mathbf{x})$ can be defined as

$$\{f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n\}, \tag{2.9}$$

where $\mathbf{x}$ represents the feature vector, the symbol $\mathbb{R}$ represents the set of all the real numbers, and n is $\mathbb{R}$ dimensional space. Given L observations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L$, the joint distribution of the random variables $f(\mathbf{x}_1)$, $f(\mathbf{x}_2), \dots, f(\mathbf{x}_L)$ is Gaussian. Moreover, being a GP completely specified by its mean $m(\mathbf{x})$ and its covariance function $k(\mathbf{x}, \mathbf{x}')$, it can be written as

$$f(\mathbf{x}) \sim GP\big(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\big). \tag{2.10}$$

The covariance function $k(\mathbf{x}, \mathbf{x}')$, which is also known as kernel function, specifies the statistical relationship between two input points of the input space.

Gaussian Process for Regression

Considering the GP described by Equation 2.10, a GPR model can be represented as

$$h(\mathbf{x})^{\mathrm{T}}\boldsymbol{\beta} + f(\mathbf{x}), \tag{2.11}$$

where $h(\mathbf{x})$ represents a set of basis functions whose role is to transform the original feature vector $\mathbf{x} \in \mathbb{R}^n$ into a new feature vector $\in \mathbb{R}^p$ (with p being the dimension of the new feature space), such as $h : \mathbb{R}^n \to \mathbb{R}^p$, $\boldsymbol{\beta}$ represents the $p \times 1$ vector of the basis function coefficients, and $f(\mathbf{x})$ represents a set of random variables from $GP(0, k(\mathbf{x}, \mathbf{x}'))$, which is a GP with zero mean and $k(\mathbf{x}, \mathbf{x}')$ covariance function.

Therefore, any instance l of the GPR label set y, can be modelled as the following probability with Gaussian distribution

$$P(y_l | f(\mathbf{x}_l), \mathbf{x}_l) \sim N(y_l | h(\mathbf{x}_l)^{\mathrm{T}} \boldsymbol{\beta} + f(\mathbf{x}_l), \sigma^2) \,, \tag{2.12}$$

where $f(\mathbf{x}_l)$ represents a latent variable and σ^2 represents the variance of the Gaussian distribution. Since a latent variable $f(\mathbf{x}_l)$ is introduced for each observation $\mathbf{x}_l$, with $l = 1,2, \dots, L$, the GPR is considered as a probabilistic nonparametric model. The fact of being a probabilistic model implies that, once the model is trained, from the standard deviations σ of the predicted responses a prediction interval can be computed. The prediction interval can be useful when dealing with noisy observations. In addition, as mentioned in Section 2.1.2, a model is defined as nonparametric when the number of its parameters is not fixed, but depends on the training dataset (like in this case, where the number of latent variables is directly correlated to the number of observations).

The joint distribution of the latent variables $f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_L)$ in the GPR model can be written in vector form as

$$P(f|\mathbf{X}) \sim N\big(f\big|0, \mathbf{K}(\mathbf{X}, \mathbf{X})\big) \,, \tag{2.13}$$

where $\mathbf{X}$ is the matrix composed by the transpose of the L observation vectors $\mathbf{x}_l$ and $\mathbf{K}(\mathbf{X}, \mathbf{X})$ is the kernel matrix (i.e., the covariance matrix) composed by all the kernel functions (i.e., the covariance functions) $k(\mathbf{x}, \mathbf{x}')$, such as

$$\mathbf{K}(\mathbf{X}, \mathbf{X}) = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_L) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_L, \mathbf{x}_1) & \cdots & k(\mathbf{x}_L, \mathbf{x}_L) \end{pmatrix}. \tag{2.14}$$

Usually, the covariance function $k(\mathbf{x}, \mathbf{x}')$ is parametrized using a set of hyperparameters $\boldsymbol{\theta}$, also known as kernel parameters, such as $k(\mathbf{x}, \mathbf{x}'|\boldsymbol{\theta})$.

Therefore, fitting a GPR model requires, during the training phase, the estimation of three parameters from the data, which are the coefficients vector $\boldsymbol{\beta}$ of the basis functions, the variance σ^2, and the hyperparameters vector $\boldsymbol{\theta}$ of the covariance functions $k(\mathbf{x}, \mathbf{x}'|\boldsymbol{\theta})$.

Several aspects of GPR make it an interesting solution for solving regression problems, among which we can mention the kernel parameters learning, which

happens directly from the data, and allows to shape the fitted function in many different ways, and the fact that, for each predicted value, it returns a distribution rather than just the single value. This last feature is particularly important, since it allows to have confidence intervals for the predictions. On the other hand, being a nonparametric model, GPR needs to take into account the whole training data every time it makes a prediction, therefore requiring an important computational cost which scales cubically with the number of training samples.

As per the SVM, in this thesis we employed GPR to estimate the signal OSNR value starting from its optical spectrum. We will explain in detail GPR implementation within our work in Chapter 5.

Gaussian Process for Classification

Usually, in a probabilistic approach to classification, the posterior probabilities of the target variable for a new input vector assume values in the interval (0, 1). However, the predictions made by a GP model can assume any value on the real axis. Therefore, to address classification problems with GPs, their outputs have first to be adapted using appropriate nonlinear activation functions, such as the logistic sigmoid one [18]. Once this transformation is performed, if a two-class classification problem is considered, the two probabilities to be predicted will be

$$p(y_{L+1} = 1|\mathbf{y}_L)\,, \tag{2.15}$$

$$p(y_{L+1} = 0|\mathbf{y}_L) = 1 - p(y_{L+1} = 1|\mathbf{y}_L)\,, \tag{2.16}$$

where $\mathbf{y}_L = (y, y_2, \dots, y_L)^{\mathrm{T}}$ is the vector representing the observed target variables related to the training set inputs $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L$ and y_{L+1} represents the target variable related to a single test point $\mathbf{x}_{L+1}$.

As shown by Equation 2.16, $p(y_{L+1} = 1|\mathbf{y}_L)$ can be simply derived by Equation 2.15, which thus becomes the only probability to be predicted. For a detailed explanation on how the prediction of such probability is assessed, the reader is referred to [18]. To extend such results to multi-class classifications, the simplest option is to use several binary one-versus-rest classifiers [39].

2.2 Convex Optimization

In this section we provide a brief overview of convex optimization, a special class of mathematical optimization problems. Then, we define and formulate the main characteristics of gradient descent and subgradient methods, two of the most common approaches to address optimization problems. In Chapter

6, we adopt such kind of algorithms to optimize the spectral spacing within an optical superchannel.

2.2.1 Introduction

In general, an optimization problem assumes the following form [40]:

$$\begin{aligned} &\text{minimize} && f_0(x) \\ &\text{subject to} && f_i(x) \leq b_i, \qquad i = 1, \dots, m\,, \end{aligned} \tag{2.17}$$

where $x = (x_1, \dots, x_n)$ is the vector representing the optimization variable of the problem, $f_0 : \mathbb{R}^n \to \mathbb{R}$ represents the objective, or cost, function, $f_i : \mathbb{R}^n \to \mathbb{R}$, with $i = 1, \dots, m$, represent the (inequality) constraint functions, and $b_1, \dots, b_m$ are the limits, or bounds, for the constraints. If there are no constraints, the problem is defined as an unconstrained optimization problem. A vector x^* is called optimal, or a solution of the optimization problem, if it has the smallest objective value among all vectors that satisfy the constraints: for any z with $f_1(z) \leq b_1, \dots, f_m(z) \leq b_m$, we have $f_0(z) \geq f_0(x^*)$. By convention, optimization problems are always represented as minimization problems. We can translate any maximization problem into a minimization one, simply by minimizing the negative of the cost function.

In particular, if the objective and constraint functions $f_0, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ satisfy the following condition

$$f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y)\,, \tag{2.18}$$

for all x, with $y \in \mathbb{R}^n$, and all the coefficients α and $\beta \in \mathbb{R}$, they are considered as linear functions, and the optimization problem described by Equation 2.17 is called a linear program. On the contrary, if the problem is not linear, it is called a nonlinear program.

In addition, if the objective and constraint functions $f_0, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ satisfy the following inequality

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)\,, \tag{2.19}$$

for all x, with $y \in \mathbb{R}^n$, and all the coefficients α and $\beta \in \mathbb{R}$, with $\alpha + \beta = 1$, $\alpha \geq 0$, and $\beta \geq 0$, they are considered as convex functions, and the optimization problem described by Equation 2.17 is called a convex optimization problem. Geometrically, the inequality described by Equation 2.19 means that a function is convex if the line segment between $(x, f(x))$ and $(y, f(y))$, which represents the chord from x to y, lies above the graph of f, as we depict in FIGURE 2.4. By contrast, a function f is defined as concave if $-f$ is convex.

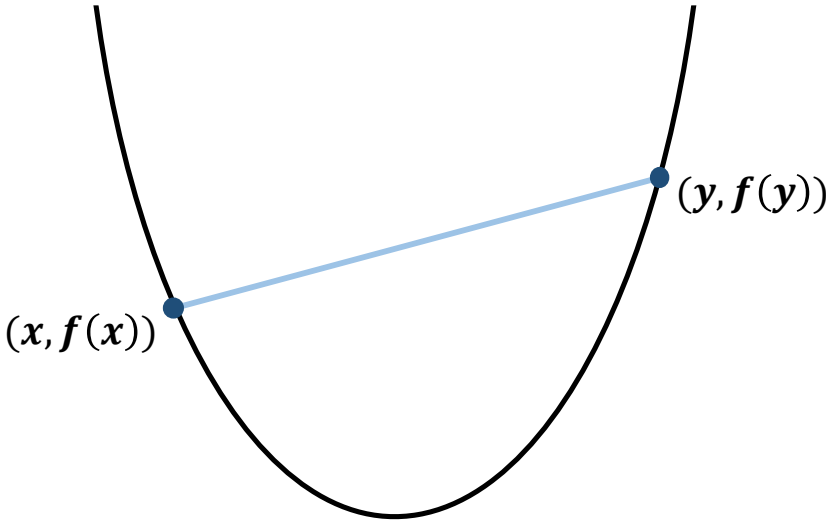

FIGURE 2.4. Example of a convex function. The chord between any two points of the graph lies above the graph itself.

Since convexity is more general than linearity, convex optimization is a generalization of linear programming, and therefore any linear program can be considered as a convex optimization problem.

According to [40], we can interpret the optimization problem described by Equation 2.17 as an abstraction of the problem of choosing the best possible vector in $\mathbb{R}^n$, from a set of candidate choices. In turn, the variable x represents the choice made, whereas the constraints $f_i(x) \leq b_i$ and the objective value $f_0(x)$ represent the specifications that limit the possible choices and the cost of choosing x, respectively. On the contrary, we can also think of $-f_0(x)$ as representing the value, or utility, of choosing x. Finally, we can think a solution x^* of the optimization problem, as the choice that has the minimum cost (or the maximum utility), among all the choices that meet the required specifications.

An example of optimization problem is data fitting, in which the task is to find a model, from a family of potential models, that best fits some observed data and prior information [40]. In particular, here the variables are the parameters of the model, and the constraints can represent the prior information. We leverage data fitting in Section 5.3, in order to retrieve the optical filter Transfer Functions (TF) spectra.

Unluckily, no analytical formulas are available to solve convex optimization problems. Nevertheless, several effective methods can help address such problems [40]. In general, if a problem can be formulated as a convex optimization one, then an efficient way to solve it exists. However, the

challenging part is to recognize and formulate convex optimization problems, or those that can be transformed into such ones.

2.2.2 Gradient Descent Method

In case of unconstrained optimization problems, one of the most common method to solve them is known as gradient descent method. In the following paragraphs, we derive the gradient descent method formulation starting from the definition of an unconstrained optimization problem, and defining a general descent method. However, with small extensions, similar solutions can also be employed to address constrained problems. The following derivation is reported from [40].

Consider the following unconstrained optimization problem:

$$\text{minimize} \quad f(x)\,, \tag{2.20}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ represents a convex and twice continuously differentiable cost function, which means that the domain of f, **dom** f (i.e., the subset of $\mathbb{R}^n$ of points x for which $f(x)$ is defined) is open. It can be demonstrated [40], that a necessary and sufficient condition for a point x^* to be the optimal solution of the problem described by Equation 2.20 is

$$\nabla f(x^*) = 0\,, \tag{2.21}$$

where ∇ represents the gradient of f. Therefore, solving the problem described by Equation 2.20 is equivalent to find a solution to the one described by Equation 2.21, which is a set of n equations in the n variables $x_1, \ldots, x_n$. Usually, such problems are solved using iterative algorithms, that are algorithms that compute a sequence of points $x^{(0)}, x^{(1)}, \ldots \in$ **dom** f with $f(x^{(k)}) \to p^*$ as $k \to \infty$, where $p^* = f(x^*)$ is the optimal value of f. This sequence of points is known as minimizing sequence. The iterative algorithm stops when $f\left(x^{(k)}\right) - p^* \leq \epsilon$, where $\epsilon > 0$ represents some specified tolerance.

In addition, such iterative algorithms requires a suitable starting point $x^{(0)}$. In particular, the starting point must lie in **dom** f, and in addition the sublevel set S, defined as

$$S = \left\{x \in \textbf{dom}\, f \middle| f(x) \leq f\left(x^{(0)}\right)\right\}, \tag{2.22}$$

must be closed. Such condition is satisfied for all $x^{(0)} \in$ **dom** f if f is closed, which means that all its sublevel sets are closed. For instance, continuous functions with **dom** $f = \mathbb{R}^n$ are closed, so if **dom** $f = \mathbb{R}^n$, the initial sublevel set condition is satisfied by any $x^{(0)}$.

The iterative algorithms we consider produce a minimizing sequence of the form $x^{(k)}$, $k = 1, \dots$, where

$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)}\,, \tag{2.23}$$

and $t^{(k)} > 0$, except when $x^{(k)}$ is optimal. In Equation 2.23, Δx represents a vector in $\mathbb{R}^n$, which is known as the step or search direction, $k = 0,1, \dots$ denotes the iteration number, and the scalar $t^{(k)}$ is called the step size or step length at iteration k.

An iteration method is referred to as descent, if

$$f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right), \tag{2.24}$$

except when $x^{(k)}$ is optimal. This implies that for all k we have $x^{(k)} \in S$, where S is the initial sublevel set defined in Equation 2.22, and in particular we have $x^{(k)} \in \mathbf{dom}\, f$. It can be demonstrated [40], that the search direction in a descent method must satisfy the condition

$$\nabla f\left(x^{(k)}\right)^T \Delta x^{(k)} < 0\,, \tag{2.25}$$

which means that the search direction must make an acute angle with the negative gradient. Such direction is defined as the descent direction.

From [40], we report Algorithm 2.1, the outline of a general descent method, which alternates between two steps: determining a descent direction Δx, and selecting a proper step size t.

Algorithm 2.1: General descent method.

given a starting point $x \in \mathbf{dom}\, f$.
repeat
1. Determine a descent direction Δx.
2. *Line search*. Choose a step size $t > 0$
3. *Update*. $x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)}$.

until stopping criterion is satisfied.

Usually, the stopping criterion is of the form $\|\nabla f(x)\| \leq \eta$, where η represents a small and positive value, and often it is checked while, or immediately after, the descent direction Δx is computed.

If as search direction we select the negative gradient $\Delta x = -\nabla f(x)$, the resulting descent algorithm is known as the gradient descent method, which we report from [40] as Algorithm 2.2.

Algorithm 2.2: Gradient descent method.

given a starting point $x \in$ **dom** f.
repeat
1. $\Delta x^{(k+1)} = -\nabla f(x^{(k)})$
2. *Line search*. Choose step size t via exact or backtracking line search.
3. *Update*. $x^{(k+1)} = x^{(k)} + t^{(k)}\Delta x^{(k)}$.

until stopping criterion is satisfied.

In Algorithm 2.2, exact and backtracking line search represent two common line search methods [40]. As per the general descent method, also the stopping criterion for the gradient descent method is often checked after step 1.

2.2.3 Subgradient Method

In the minimization problem described by Equation 2.20, we assume the function f to be convex and differentiable. To minimize a convex nondifferentiable function, we can employ the subgradient method, a simple algorithm originally developed by Shor in the 1970s [41]. Unlike the gradient method, the subgradient one is not a descent method, in fact the objective function can also increase, as we will show in Chapter 6 for an optimization problem applied to superchannels. Apart from this aspect, the main difference between gradient and subgradient methods is that the latter uses step lengths that are fixed in advance, instead of exact or approximate line searches as in Algorithm 2.2. In this section, we first recall the definition of subgradient and then we leverage it to describe the subgradient method, as done in [42].

The subgradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ (not necessarily convex) at x can be defined as any vector $g \in \mathbb{R}^n$ that satisfies, for all y, the inequality

$$f(y) \geq f(x) + g^T(y - x)\,. \tag{2.26}$$

Considering the same minimization problem described by Equation 2.20, but with the function f being only convex, the subgradient method minimize f using the iteration

$$x^{(k+1)} = x^{(k)} - t^{(k)} g^{(k)}\,, \tag{2.27}$$

where, similarly to the minimizing sequence described by Equation 2.23, $x^{(k)}$ represents the k-th iteration, $t^{(k)} > 0$ is the k-th step size, and $g^{(k)}$ represents any subgradient of f at $x^{(k)}$. What Equation 2.26 expresses is, that at each iteration of the subgradient method, we take a step in the direction of a negative subgradient. If f is differentiable, the only possible choice for $g^{(k)}$ would be $\nabla f\left(x^{(k)}\right)$, and the subgradient method would then reduce to the gradient method presented in Section 2.2, apart for the step size choice.

As we previously mentioned, the subgradient method is not a descent method. Hence, it is common to keep track of the best achieved points, which are those with the smallest function value. At each step, we set [42]

$$f_{\text{best}}^{(k)} = \min\left\{f_{\text{best}}^{(k-1)}, f(x^{(k)})\right\}. \tag{2.28}$$

And finally, we have

$$f_{\text{best}}^{(k)} = \min\left\{f\left(x^{(1)}\right), \dots, f(x^{(k)})\right\}, \tag{2.29}$$

where $f_{\text{best}}^{(k)}$ represents the best objective value found in k iterations.

Within the subgradient method, several types of step size can be chosen, among which two of the most common are the constant step size and the constant step length. In the former, $t^{(k)} = h$, where h is a constant independent from k. Instead, in the latter, $t^{(k)} = h/\left\|g^{(k)}\right\|$, which translates into $h = \left\|x^{(k+1)} - x^{(k)}\right\|$.

For these two step size rules, it can be demonstrated [42], that the subgradient algorithm is guaranteed to converge within some range of the optimal value, which translates into:

$$\lim_{k\to\infty} f_{\text{best}}^{(k)} - f^* < \epsilon\,, \tag{2.30}$$

where the value ϵ is a function of the step size parameter h, and decreases with it.

2.2.4 Stochastic Subgradient Method

The stochastic subgradient method is similar to the subgradient method, but it uses noisy subgradient and a more limited set of step size rules [43]. The noise can represent error in computing a true subgradient, or errors related to the measurement process. In this section, we first introduce a noisy subgradient, and then, following the steps proposed in [44], we use it to define the stochastic subgradient method. In Chapter 6, we apply such method to optimize the subchannels spectral spacing in a superchannel.

Considering the convex function $f : \mathbb{R}^n \to \mathbb{R}$, a random vector $\tilde{g} \in \mathbb{R}^n$ is defined as a noisy subgradient of f at x, if its expected value satisfy the following condition:

$$g = \mathbf{E}\tilde{g} \in \partial f(x)\,, \tag{2.31}$$

where $\partial f(x)$ represent the subdifferential of f at x, which is the set of all subgradients of f at x.

Rewriting Equation 2.26 considering Equation 2.31, the following condition about the noisy subgradient, which has to be valid for all y, can be stated

$$f(y) \geq f(x) + (\mathbf{E}\tilde{g})^T(y - x)\,. \tag{2.32}$$

Therefore, $\tilde{g}$ is defined as a noisy unbiased subgradient of f at x, if it can be written as $\tilde{g} = g + v$, where $g \in \partial f(x)$ and v, represents a zero mean noise.

If x is a random variable, then we say that $\tilde{g}$ is a noisy subgradient of f at x, if for all y, the following condition holds almost surely (i.e., the set of possible exceptions may be nonempty, but has probability 0):

$$f(y) \geq f(x) + \mathbf{E}(\tilde{g}|x)^T(y - x)\,. \tag{2.33}$$

Similar to Equation 2.31, the condition expressed by Equation 2.33 can be written in a more compact way, as

$$g = \mathbf{E}(\tilde{g}|x) \in \partial f(x)\,. \tag{2.34}$$

Therefore, the stochastic subgradient method is defined as the subgradient method, but using noisy subgradients. The slow convergence of subgradient methods, which means high amounts of considered steps, averages out the statistical errors in the subgradients evaluations [44]. Considering the usual unconstrained minimization problem of a convex function $f : \mathbb{R}^n \to \mathbb{R}$, the stochastic subgradient method minimize f using the iteration

$$x^{(k+1)} = x^{(k)} - t^{(k)}\tilde{g}^{(k)}\,, \tag{2.35}$$

where $t^{(k)} > 0$ is the k-th step size and $\tilde{g}^{(k)}$ is any noisy subgradient of f at $x^{(k)}$, such as

$$\mathbf{E}\left(\tilde{g}^{(k)}\middle|x^{(k)}\right) = g^{(k)} \in \partial f\left(x^{(k)}\right). \tag{2.36}$$

As per the ordinary subgradient method, also in the stochastic subgradient one, the objective function can increase during the research for the minimum, so we again keep track of the best found points as

$$f_{\text{best}}^{(k)} = \min\left\{f\left(x^{(1)}\right), \ldots, f(x^{(k)})\right\}, \tag{2.37}$$

where $x^{(k)}$ and $f_{\text{best}}^{(k)}$ represent stochastic processes.

It can be demonstrated [44], that for $k \to \infty$, the expected $f_{\text{best}}^{(k)}$ value converges to the optimum as

$$\mathbf{E} f_{\text{best}}^{(k)} \to f^* \,. \tag{2.38}$$

In addition, convergence in probability can also be demonstrated [44], in fact, for any $\epsilon > 0$,

$$\lim_{k \to \infty} \mathbf{Prob}(f_{\text{best}}^{(k)} \geq f^* + \epsilon) = 0 \,. \tag{2.39}$$

Chapter 3

State of the Art

In this chapter, we review the background concepts and the state of the art of those topics around which our technical proposals revolve. This chapter lays the foundation for Chapter 4, Chapter 5, and Chapter 6, in which the solutions we identified are presented. In particular, in this chapter, we review the literature on disaggregated optical networks, OPM techniques, with a special focus on OSNR, optical filter and spectral-based monitoring solutions, and superchannels. Building on those, we identify the open challenges and define the network scenario considered in the remainder of this dissertation.

More in detail, this chapter is organized as follows. In Section 3.1, we present the disaggregation paradigm and its related operational models and network architectures. In Section 3.2, we review the main aspects of OPM, mainly focusing on those techniques that allow monitoring the OSNR and the filter-related parameters, and those monitoring approaches that leverage optical spectral data. In Section 3.3, we review the superchannel technology, the advantages resulting from its adoption, and the challenges that it presents. Finally, relying on the literature reviewed in the previous sections, in Section 3.4, we present the open tasks that we identify as worthy of addressing, along with the considered network reference scenario.

3.1 Disaggregated Optical Networks

In this section, we present how the trend among telecommunication operators is evolving from aggregation to a disaggregation paradigm. We list the key enablers of such transformation and the benefits that it could bring and the challenges that still need to be faced. Moreover, we present a detailed description of the envisioned network disaggregation architecture, focusing on the white box concept.

3.1.1 Paradigm Evolution: From Aggregation to Disaggregation

Aggregation, that is, taking functions that once resided in separate devices and combining them in a single system, has always been a trend for telecommunication operators [45]. Such approach allows saving both Capital Expenditure (CAPEX) and Operation Expenditure (OPEX). The former term describes the funds used by a company to buy or upgrade its fixed assets, while the latter is related to the money spent by the company to maintain its business. A cost reduction related to the optical aggregated network approach translates for instance into fewer boxes to be acquired (CAPEX reduction) or in a more simplified network management (OPEX reduction).

Of course, as pointed out in [45], to guarantee the success of the aggregation approach (i.e., a significant cost reduction) an adequate unified element/network management system had to be developed. Historically, such unified control system has been provided by the equipment vendor itself and it has been used to manage all its network elements. However, this aspect represents the downside of the aggregation strategy, given the proprietary and therefore closed nature of the element/network management systems, which hinders the innovation process by reducing the competition.

Nowadays, the arrival at an increased rate of new generations of coherent DWDM TPs, and the presence of new large-scale network operators, such as the Internet Content Providers (ICPs), is pushing the telecom industry to consider a complete opposite approach with respect to the past: the disaggregation paradigm [46]. In fact, the growing need for ICPs to optimize their network components requires a more dynamic and flexible scenario than the old aggregated one. In addition, the rise of Software-Defined Networking (SDN) and Network Functions Virtualization (NFV) has enabled a potential disaggregation of the control plane from the forwarding plane and of the network hardware from software functions. Similarly, when considering the WDM transport segment, a traditional integrated/aggregated WDM system can be separated into different functional blocks, like for instance the so-called Open Line System (OLS) and the TP/Muxponder (MP) platform [47]. In particular, the OLS refers to the fiber, the amplifiers, and the optical network nodes (ROADMs). Moreover, new generations of TPs are usually released faster than new generations of line systems, not to mention that their replacement cost and lifespan is also way lower [48]. Therefore, network operators would like to have the possibility to deploy TPs from multiple suppliers and belonging to multiple generations, over a single line system generation [46]. An example of the benefits coming from the OLS employment is reported in [49], where the authors compared the routing performance of OLS with respect to proprietary line systems. Their results demonstrated that

disaggregated systems can be very competitive, especially in multi-domain deployment scenarios.

In general, the benefits of adopting the disaggregation approach within optical networks are several [45], [46], [47]. Among them, the most significant is certainly the possibility for the telecom operators to efficiently scale and size their infrastructure according to their needs. In fact, they can incrementally expand the network dimensions as traffic increases and more capacity is required, better matching the targeted requirements using appropriate components, which also allow to avoid any vendor lock-in issue. Furthermore, disaggregation allows to share functions across different hardware resources (e.g., using commodity servers). Therefore, a more efficient use of the network capacity is envisioned, along with a reduction of the power consumption and footprint. All these aspects will translate in an OPEX and CAPEX reduction [50]. Finally, disaggregation guarantees a high level of flexibility, a fundamental requirement for the future optical transport networks. On the other hand, hardware suppliers and service providers still need to face some challenges before being able to fully embrace the disaggregation paradigm [45], [46]. The main one certainly is to maintain the network and services required performance/reliability, while migrating toward a disaggregated environment. In this regard, in [48], when asked about the primary market target for disaggregated systems, the service providers identified metro-access (i.e., <100 km) as the most likely application. In fact, according to the surveyed service providers, for long haul (i.e., >600 km) applications performance remains the key requirement to address, while in short-medium range scenarios interoperability and cost savings have relatively higher priority.

In the past few years, several initiatives for the development of common specifications for the disaggregated optical equipment have been launched, among them we can cite OpenROADM [51], OpenConfig [52], and Telecom Infra Project (TIP) [53].

An experimental demonstration of a disaggregated network scenario adopting SDN-enabled Sliceable Bandwidth/Bitrate Variable Transceivers (S-BVTs) can be found in [54]. There, the authors implemented OpenConfig SDN agents, which are developed to reconfigure and program a disaggregated transceiver according to the network condition.

According to the 2019 survey presented in [55], in which the authors interviewed several service providers over the evolution of their strategies to address the 5G arrival, the 66% of them were likely to consider disaggregated architectures for their future optical transport networks. The main advantages for justifying such choice were the possibility to reduce both CAPEX and OPEX and the deployment flexibility guaranteed by the disaggregation approach.

These results prove once more how the optical network disaggregation paradigm represents a possible key enabler for the upcoming Information and Communication Technologies (ICTs).

3.1.2 The White Box Operational Model

The disaggregation paradigm can be defined at different levels within the optical networks. At the optical layer, such approach involves an evolution from chassis-based (proprietary) network elements to commodity (off-the-shelf) components, the so-called white boxes, where generic off-the-shelf hardware can be purchased from any vendor and customized with software from different sources [4].

More in detail, in [4], the authors identified 4 different operational models for the metro/regional WDM transport system. On one side, there is the bare metal operational model, in which the operators source their hardware directly from the Original Design Manufacturers (ODMs) adapting free and open source software to it. Instead, on the other side, we have the traditional aggregated model, also known as black box approach, where the single vendor closely aggregates both the hardware and software and assumes oversight of the hardware components. In between these two extremes, two other operational models are identified: an evolution of the bare metal model, known as the white box approach and a slightly different paradigm known as brite box or branded white box. In the former, the operating system and application software are disaggregated from the hardware. Nevertheless, the ODMs can provide devices to the operators with some preinstalled operating system on them along with a certain level of support. In the latter, as for white boxes, software and hardware are disaggregated, but some preinstalled third-party operating system is present and a tailored level of lifecycle support from the vendor is envisioned. All the above mentioned approaches imply a different level of involvement for the telecommunication operators and the equipment vendors. In particular, the bare metal and white box ones require the biggest commitment from the operators.

Therefore, white boxes represent a key element for providing the required disaggregation between software and hardware, as well as to support autonomic optical networking architectures [56]. An example of white box is reported in [57], where the authors experimentally demonstrated the configuration and the dynamic reconfiguration of a fully disaggregated device in which the hardware was separated from the control/management plane. In that case, the white box was composed by two types of TPs, performance monitors, switches, add-drop multiplexers (MUXs), and an agent for

interfacing with the control/management plane. All the modules listed above were provided by different vendors.

As pointed out in [56] and [57], performance monitors play a fundamental role in white box-based disaggregated optical networks. In fact, in order to take accurate decisions, updated monitoring data must be periodically conveyed from different points of the network to the controller. Therefore, proactive and reactive automation of optical disaggregated networks with white box switches represents another big challenge that needs to be addressed [58]. Such automation is based on processing real-time network monitoring parameters and learning from the effects of the decisions previously taken. Thus, a fundamental task to accomplish is the development of new optical monitoring techniques able to deliver the feedback needed for guaranteeing a particular end-to-end QoT and QoS. In this context, the monitoring probe scheme and the techniques described within this thesis represent an important starting point.

Finally, to support white boxes control and management, standardized data models are required. One of the most common is Yet Another Next Generation (YANG) [59], a data modelling language that enables interfacing the hardware with the control and the management planes. YANG is supported by NETCONF, an emerging protocol standardized by the IETF [60].

3.1.3 Disaggregated Optical Networks Architectures

The main elements of a metro/regional WDM transport system can be classified as: client to WDM adapters, which include TPs, MPs, and Switchponders (SPs), ROADMs, amplifiers, line terminals, i.e., MUXs and Demultiplexers (DEMUXs), a suitable interconnection for all these elements, and finally, a WDM transport controller/management software [4].

Telecommunication operators can be involved in several degrees of depth in the design, assembly, integration and testing of the whole WDM transport system. According to [4], based on such implication level, three main network visions can be defined. The first one refers to the classical aggregated scenario, where a fully aggregated optical network is considered, as shown in FIGURE 3.1(a), where a horizontal aggregated data plane structure is depicted. In this case, the optical system lifecycle management is responsibility of the system vendor that provides both the proprietary WDM transport controller and all the network elements, as foreseen by a pure black box approach.

A second approach, defined as the partially disaggregated one, considers a hybrid scenario in which the terminal equipment (i.e., the TPs) is

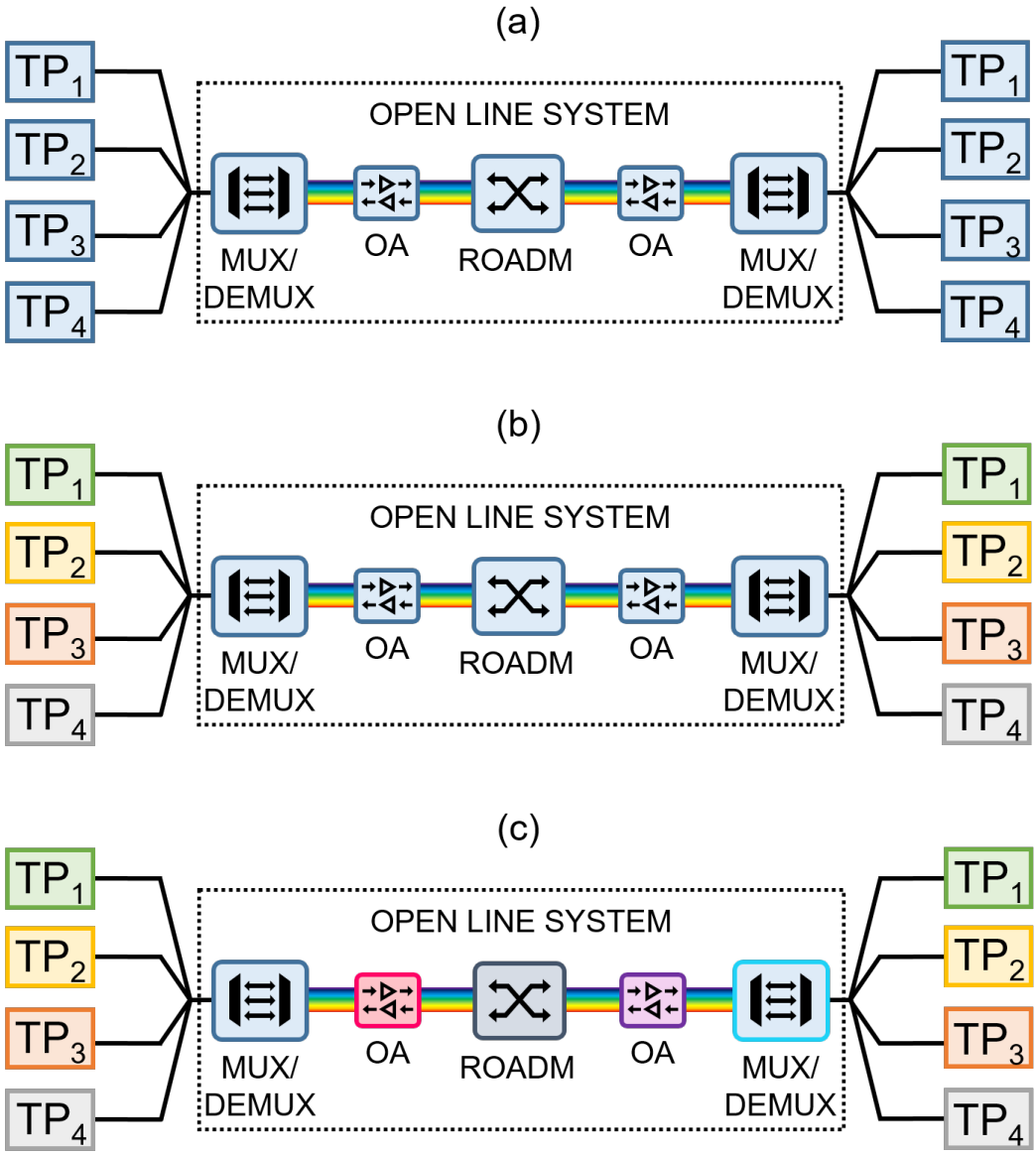

FIGURE 3.1. (a) The classical fully aggregated data plane: all the TPs and the OLS elements are provided by the same vendor (i.e., they all have the same colour). (b) Partially horizontally disaggregated data plane: while each TP is provided by a different vendor, all the OLS elements are provided by the same vendor. (c) Fully horizontally disaggregated data plane: all the TPs and all the OLS elements are provided by a different vendor. TP: transponder, MUX/DEMUX: multiplexer/demultiplexers, OA: optical amplifier, ROADM: reconfigurable optical add/drop multiplexer.

disaggregated from the line system. In fact, as introduced in Section 3.1.1, within the WDM transport network the concept of disaggregation implies a separation between the TPs/MPs platform, which adapt the digital client signals to the analogical media channels, and the rest of the transport layer, also known as OLS. Thanks to such strategy, the TPs, whose technological

lifecycle is shorter than that of the line system components, can be independently upgraded [61]. In this hybrid approach, the guidelines defined by the black box paradigm presented in Section 3.1.2 are mainly adopted within the line system. A horizontal partially disaggregated data plane is depicted in FIGURE 3.1(b).

Finally, the third approach envisions a fully disaggregated optical network. In such scenarios, each element of the network, including the TPs, is purchased from a different vendor and the control intelligence is moved to a vendor-agnostic WDM controller, as foreseen by the white box operational model. A strong presence of the telecommunication operators in the WDM transport system lifecycle is also required. FIGURE 3.1(c) depicts an example of a fully disaggregated horizontal data plane. Potentially, further disaggregation levels can be reached decomposing each ROADM into set of ROADM degrees or even more elementary components, also according to the bare metal operational model presented in Section 3.1.2.

All the elements composing the disaggregated optical network are seen by the control and management plane as a single management entity. Such task is accomplished through a suitable Open Application Programming Interface (OpenAPI), also named South Bound Interface (SBI). On the other hand, by means of a North Bound Interface (NBI), the management and control planes communicate with higher-level entities (e.g. the SDN controller) enabling enhanced network programmability [4].

3.2 Optical Performance Monitoring

In this section, we introduce the concept of OPM along with the reasons that make it a potential game changer in the context of optical networks. In particular, we focus on OPM techniques that allow to monitor the OSNR and the filter-related parameters, such as the filter central frequency or the 3/6 dB filter bandwidth. Moreover, we describe those OPM techniques that allow to retrieve several network-related information, leveraging optical spectral data.

3.2.1 An Introduction to Optical Performance Monitoring

OPM plays a key role in the current optical network management [9]. In fact, it allows the network operators to quickly identify faults occurring in their networks, their locations, and the reasons that caused them. Indeed, some of the faults affecting an optical network are the so-called soft failures, which do not bring the whole system down, but rather result in noise-like conditions that deteriorate the communications [22]. Soft failures are caused by optical

impairments, such as for instance the frequency drift of the transmitter laser. On the other hand, the term hard failure refers to scenarios where "catastrophic" signal degradations occur. Usually, such situations can be easily identified simply by monitoring any abrupt loss of power occurring in the network. The ability to identify failures provided by OPM, allows the operators to develop more robust and stable optical networks [9].

Since an optical communication can be affected by several optical impairments, many parameters must be monitored within an optical network. Some of them can be retrieved directly from the monitored optical network information (e.g., the optical power of a signal), some others need raw data to undergo a certain degree of processing before they can be recovered (e.g., PMD). Optical physical layer impairments include the ASE noise (see Section 3.2.2), CD, PMD, Polarization-Dependent Loss (PDL), fiber nonlinearities, frequency chirp, and the distortions related to the optical filters (see Section 3.2.3) [9].

Focusing on OPM provided by the receivers of the connections, and depending on the kind of receiver available at the end of the considered communication link, the OPM techniques can be divided between noncoherent and coherent [10], [62]. The former set of techniques is applied in connections where the optical signals are directly detected (i.e., no local lasers are present at the OPM site). On the contrary, the latter refers to those connections in which digital coherent receivers are employed. In fact, in such scenarios, several OPM techniques became obsolete due to the tight channel spacing, the Nyquist pulse shaping, and the polarization multiplexing enabled by the coherent detection (e.g., the polarization-based OSNR monitoring techniques, see Section 3.2.2) [10]. Of course, this second solution guarantees more options in terms of monitored parameters and techniques. A vast amount of relevant techniques for monitoring many different parameters in both the above presented scenarios have been proposed [10]. Furthermore, monitoring can also be performed at the intermediate nodes of the network or even at the EDFA locations, by means of OSAs, coherent receiver-based solutions or leveraging flexible solutions, such as the front-end scheme presented in Chapter 4. The spectral data acquired with this kind of devices, can be enhanced using adequate spectral processing techniques, such as those presented in Chapter 5.

In Chapter 5 of this thesis, we will focus on monitoring the OSNR and the filter-related parameters, including the filter central frequency and the 3/6 dB filter bandwidth. Several OPM techniques were proposed in the past to retrieve such parameters. In the following subsections, we will review the main ones along with their benefits and limitations. Moreover, in Chapter 6, we will exploit the

monitoring features of coherent receivers to optimize the channel frequencies of superchannels.

3.2.2 Optical Signal-to-Noise Ratio Monitoring

In 2016, the International Electrotechnical Commission (IEC), through the technical report IEC 61282-12, defined the spectrally integrated in-band OSNR, $OSNR_{int}$, as the spectrally integrated ratio of time-averaged power spectral density of a signal to the power spectral density of the ASE noise, normalized to a chosen reference bandwidth [67]:

$$OSNR_{int} = 10\,log\left(\frac{1}{B_r}\int_{\lambda_1}^{\lambda_2}\frac{s(\lambda)}{p(\lambda)}d\lambda\right), \tag{3.1}$$

where B_r represents the reference bandwidth expressed in nm (usually 0.1 nm), λ_1 and λ_2 represent the extremes of the integration range, in nm, which usually is chosen to include the total signal spectrum (e.g., 0.4 nm in the case of a 50 GHz grid), $s(\lambda)$ represents the time-averaged signal spectral power density not including ASE, expressed in W/nm, $p(\lambda)$ represents the ASE spectral power density, independent of polarization, also expressed in W/nm, and $OSNR_{int}$ is expressed in dB.

Generally, the OSNR is considered one of the most critical parameters to be monitored within an optical network since it provides information about the quality of the network channels. In addition, the OSNR is transparent to both the bit rate and the modulation format of the optical signal, and it also contributes to the Bit Error Ratio (BER) calculation [10], [63]. Thus, the capability to monitor each WDM channel OSNR is a strong requirement in

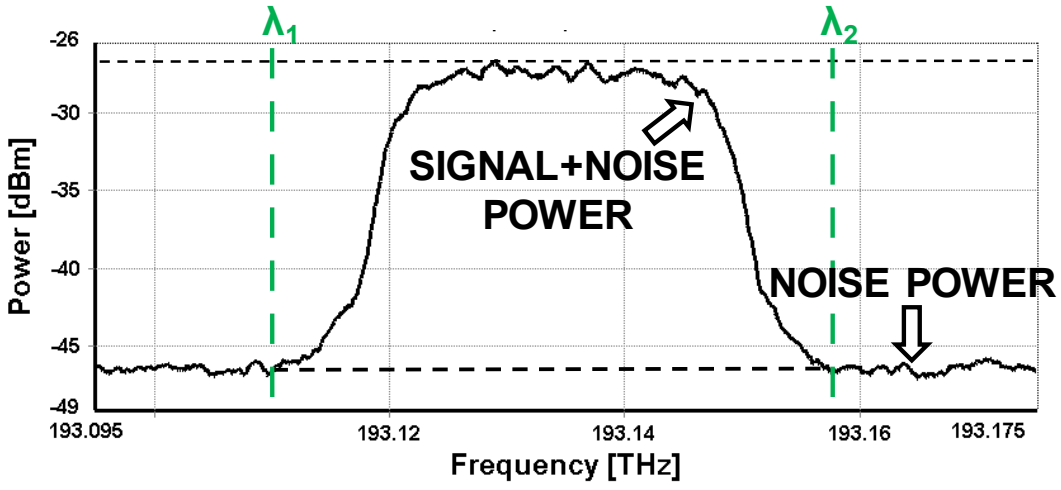

FIGURE 3.2. Example of a signal with the noise ground and the integration range extremes.

current optical networks. A schematic representation of a signal OSNR is shown in FIGURE 3.2.

In practical terms, the OSNR quantifies the ASE noise added to the signal by the optical amplifiers during the amplification process. In optical networks, optical amplifiers have the crucial role of compensating for the losses introduced by the optical fiber. Therefore, an accurate analysis of their contribution in term of noise is needed.

The OSA is a device that measures the power of each spectral component simply sweeping the central frequency of a narrowband optical filter. Then, relying on the measured power values, the OSA reconstructs the spectral shape of the considered spectral component [65].

OSAs are the key enabler of one of the most common and used techniques to measure the OSNR, the so-called interpolation method [66]. In such technique, the ASE noise is estimated interpolating the noise levels measured at the sides of the considered signal. Therefore, the interpolation technique is by definition an out-of-band method, since it relies on values measured outside the signal bandwidth, as shown in FIGURE 3.3. Estimating the ASE noise from spectral points outside the channel bandwidth becomes challenging in wavelength switched optical networks employing DWDM or flex-grid channels [67]. Indeed, in DWDM configurations, the signals are spaced closer to increase the overall transmission capacity, making the ASE noise identification in between channels very challenging. In addition, according to the number of ROADM nodes it has traversed, each channel will exhibit a different level of noise. Therefore, unless very high-resolution OSA are employed, the real out-of-band noise level of each channel could be hard to

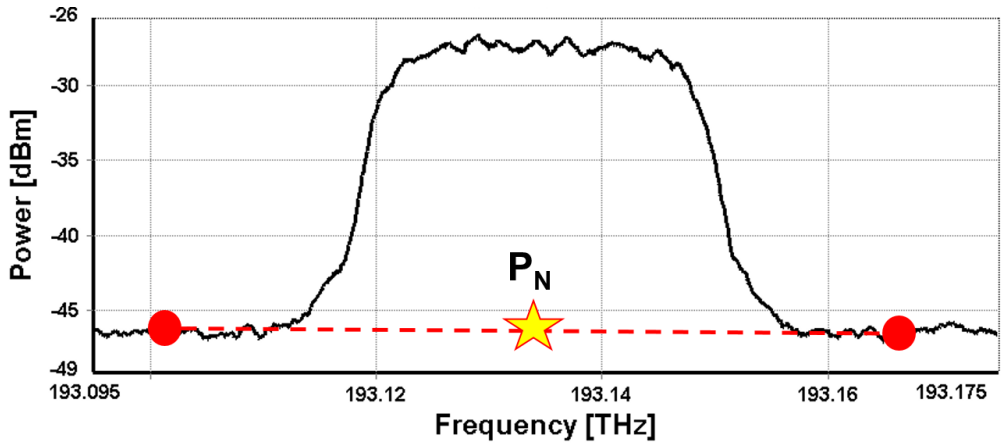

FIGURE 3.3. Graphical representation of the interpolation technique. The in-band noise level is found interpolating the left and right out-of-band noise levels.

recover. Finally, every time a signal crosses a ROADM node, according to the considered ROADM architecture (see Section 3.2.3), the signal traverses one or more optical filters [68]. The filters introduce sharp power drops in between the channels, especially when several nodes have been crossed and the Filter Cascade Effect (FCE) must also be considered [69]. All these factors make the OSNR interpolation technique unfeasible in a DWDM or flex-grid operating network. Thus, the estimation of the OSNR must be done in-band [67]. For this purpose, in Chapter 5, we will present an OSNR monitoring solution that leverages optical spectral data and ML techniques, to allow an in-band estimation of the OSNR in DWDM scenarios.

Other common approaches to monitor the OSNR include the techniques based on polarization, interferometers, beat noise analysis, asynchronous amplitude histograms, and those relying on the DSP of digital coherent detectors. In the following paragraphs, we will briefly review their main characteristics along with the challenging aspects, which hinder their implementation.

The polarization-based OSNR monitoring techniques represented a popular choice in the past, being transparent to data rates and modulation formats. They exploit the fact that the ASE noise is completely depolarized (i.e., it has an even power distribution for every polarization direction), whereas usually an optical signal contains the entire power in a precise state of polarization [9]. An example of a typical polarization-based method is the polarization-nulling technique, where the different polarization-related properties of signal and ASE noise are exploited to precisely differentiate the noise value from the signal one [70]. However, with the advent of coherent detection receivers, the demultiplexing process of Polarization Multiplexed (PM) signals became simpler with respect to direct detection, allowing Polarization-Division Multiplexing (PDM) to be widely employed in almost all the DSP-based coherent transmission systems [10]. Thus, this aspect severely limits the application range of polarization-based OSNR monitoring techniques, which cannot be employed in PM-based transmission systems.

The interferometer-based technique uses Mach-Zehnder Delay Interferometers (MZDIs). This method exploits the high coherence of the optical signals and, on the contrary, the incoherence of the in-band noise, which therefore is insensitive to constructive and destructive interference [71]. However, considering WDM signals with high spectral density makes such approach unfeasible, due to the destructive MZDI bandwidth, which would reject not only the optical signal, but also the ASE noise [9].

The beat noise analysis techniques leverage the receiver electrical noise, which is mainly caused by the beating between the optical signal and the ASE noise [72]. However, when low-frequency noise analysis is considered, the signal is

supposed to be made by periodic bit patterns and therefore its spectrum is assumed to be composed by discrete spectral components. In reality, user traffic has random patterns rather than repetitve ones, and therefore an evolution of this scheme, which relied on high-frequency noise analysis, was also proposed [73].

Another group of OSNR monitoring techniques leverages the Asynchronous Amplitude Histograms (AAHs) [9]. Amplitude Histograms (AH) are plots created sampling the electrical signal generated by the photodetector at an arbitrary sampling rate. In particular, an AH is defined as asynchronous when the adopted sampling rate is much lower than the symbol rate. Such feature allows to avoid any clock recovery system. The key point of these techniques is that the shape of an AAH reflects the properties of the optical signal that generated it. For instance, the ASE noise maps on the AAHs reducing the amplitude of its peaks. Therefore, several factors that degrade an optical signal can be detected analyzing its related AAHs, among which the OSNR is certainly one of the most important [74]. The most straightforward approach is to compare the measured AAH with a reference one, obtained from the same signal during a calibration stage. Thus, the OSNR penalty between the two AAH can be derived simply determining the ASE noise needed for the two AAH to match [75]. The main limitations of this kind of technique are related to the sampling process, which introduces noise. In particular, because of the sampler finite aperture time, during the time in which the sampling operation captures a sample of the signal under analysis, the signal itself could vary, leading to sampling errors. Therefore, the sampling aperture time conditions the performed measurements by restricting the noise or distorting the signal under analysis [9].

Finally, following the trend of applying ML algorithms to solve optical network-related tasks, several ML-based solutions have also been proposed for OPM (see Chapter 2), and in particular for OSNR monitoring. In general, the procedure to achieve ML-based OSNR monitoring requires three steps. First, the optical signal has to be converted into an electrical one and be sampled to build a dataset. Then, specific signal features have to be extracted from it, such as the AH, the Cumulative Distribution Function (CDF), the Asynchronous Delay-Tap Sampling (ADTS) plots, or the eye diagram statistics. Finally, the chosen ML model has to be trained offline. Once the offline training process is done, the ML model can be used for real-time monitoring in deployed optical networks [12]. In case the chosen ML algorithm is a SL approach, the training phase will consist of labelling the acquired features with their corresponding OSNR values. Otherwise, in case of UL approach, the training phase simply consists of providing the unlabelled data to the algorithm.

In both [76] and [77], the authors leveraged ANNs for OSNR monitoring. In those two works, the dataset included features retrieved from the signal eye-diagrams. Instead, the authors of [28], applied an ANN algorithm to asynchronous sample signal amplitudes. A similar approach was employed in [78], where as a feature to train the ANN, the authors used AAH. An evolution of this technique was proposed in [79], where two dimensional histograms were provided as features to the ANN. The 2D histograms were obtained employing ADTS, a technique in which the signal amplitude is sampled using two clocks delayed by a delay time, which is also known as delay tap. Similarly, in [80], the 2D histograms retrieved with ADTS were used as features to train a CNN. Finally, an example of OSNR monitoring with UL techniques was proposed in [81], where PCA and statistical distance measurement-based pattern recognition were applied to ADTS plots.

One of the goals of this dissertation is to deal with cost-effective and flexible OSNR monitoring schemes that can be implemented across the whole network, including intermediate nodes, such as that presented in Chapter 5. Thus, despite the availability of several OSNR monitoring techniques that leverage the DSP blocks within the digital coherent receivers located at the connection ends [10], we are not considering this kind of approach. For the same reason, we also do not consider any ML-based OSNR monitoring techniques that require features resulting from the coherent detection process. For more complete surveys on such methods, the reader can refer to [12] and [82]. It is worth noting that in Chapter 5, we will focus on OSA-based solutions for OSNR monitoring, whereas the solution that we will propose in Chapter 6 exploits the coherent receiver monitoring features to retrieve the subchannel SNR values within a superchannel.

3.2.3 Optical Filter-Related Parameters Monitoring

As mentioned in Section 1.1, ROADMs are key enablers for the current DWDM-based wavelength switched core and metro optical networks [83]. A ROADM allows individual channels to pass-through the optical nodes, or to be added/dropped. Such functionalities enable termination and entry of services, but also to transparently bypass the optical nodes, avoiding costly OEO conversions [84], [85]. In current-generation ROADMs, these features are implemented by means of Wavelength Selective Switches (WSSs). A WSS is a 1 x N optical device that works in both directions. In the 1 x N direction, it allows any entering wavelength on its common input port to be switched to any of its N available output ports. When operated in the opposite direction, the WSS selects out of its N input ports the wavelengths to be forwarded on its common output port. These functionalities are enabled by the optical filters

included in the WSS. The number of WSSs (i.e., of optical filters) within a ROADM depends on the ROADM architecture. In fact, in the past few years, their architectures evolved from a "Broadcast and Select" (B&S) approach to a more flexible and better performing paradigm called "Route and Select" (R&S) [86]. In the B&S approach, a broadcasting power splitter and only a single WSS per degree (i.e., per direction) at the egress fiber are employed. Such configuration allows to employ a reduced number of filters, thus reducing the amount of filtering penalties and the final network cost [50]. Conversely, in the R&S paradigm, two independent WSSs per degree are considered, both at the ingress and egress fibers, yielding better connections isolation and lower insertion losses. Although R&S introduces some benefits from the performance point of view, it comes with the drawback of higher cost and a more critical passband narrowing effect. Therefore, when crossing a ROADM node, the channels/signals pass through one or more optical filters, which in addition to filtering them, slightly distorts their spectral shape and therefore their characteristics [69]. Thus, it is of primary importance to have a way for constantly monitor the features that characterize an optical filter, such as its central frequency and its 3 or 6 dB bandwidth. Of course, as also mentioned in Section 3.2.2, when a channel/signal crosses many optical nodes along its path, the distortions introduced by the node filters have a huge impact on the channel/signal characteristics. In fact, on the one hand, the chances of frequency misalignments between the signal and the filter cascade (but also between the cascaded filters) increase. On the other hand, the bandwidth tightening caused by the FCE has also to be taken into account [69], [87]. For these reasons, filter-related impairments are considered as one of the causes of soft failures (see also Section 3.2.1) [22]. Therefore, in such scenarios, monitoring the filter-related parameters assumes even more importance [88]. For this purpose, in Chapter 5, we will present a technique that allows to retrieve this kind of parameters, relying on optical spectral data and on the curve fitting principle. Currently, in order to handle a larger number of optical channels, N x M Add/Drop WSS solutions are also available. However, in [89], the filter-induced penalties associated with the use of such devices are studied, showing how far this technology still is from representing a reliable solution, due to the too smooth filtering functions that it implements.

Although ROADM and filtering effects represent a real issue in today's optical networks, in literature, not many works have been dedicated to retrieve filter-related parameters at intermediate network nodes. In fact, most of the proposed solutions rely on the DSP modules of digital coherent receivers. For instance, in [90], leveraging the Q-factor metric retrieved at the coherent receiver, the authors studied the effects of the filters central frequencies shifts. In particular, the results of [90] indicated that the optical filtering penalties do not significantly change when variations of the filter Gaussian order between

3 and 4 are considered. Instead, such penalties mainly depend on the filter bandwidth. Similarly, in [91], the authors studied the performance degradation caused by frequency detuning between transmitter laser and in-line filtering, analyzing the BER retrieved at the coherent receiver. The results showed that, to limit the impact of such misalignments, frequency control with granularity finer than 1.25 GHz is needed, at both the transmitter and the filters. Instead, in [92], the authors proposed an extended Signal-to-Noise Ratio (SNR)-OSNR relation to better study the coherent receiver behavior, in presence of strong filtering conditions. In fact, accurate performance predictions in such scenarios are needed to design and optimize optical networks with improved spectral efficiency. Often, the filter-related impairments are associated to soft failures, and therefore, some new approaches to monitor the filter parameters were proposed in order to identify and prevent these unwanted events. For instance, solutions leveraging ML were proposed in [93] and [94]. In particular, in [93], the authors monitored pre-Forward Error Correction (pre-FEC) BER, to study impairments such as signal overlaps, filter tightening, and gradual/cyclical drifts. Their solution used a finite state machine to detect suspicious pre-FEC BER fluctuations and reported them to a central controller, that in turns, leveraging ML techniques, identified the cause behind the degradation. Similarly, in [94], the authors experimentally demonstrated a ML-based solution for the detection and identification of soft-failures caused by filter misalignments and/or undesired amplifier-gain reductions. The method proposed in [94] envisioned a continuous pre-FEC BER monitoring performed at the coherent receiver, over long durations time (>20 minutes). More in detail, in their experimental setup, the authors adopted a Bandwidth Variable-WSS (BV-WSS) to introduce filter bandwidth tightening effects. On the contrary, a soft failure identification method, which leveraged an UL technique, was proposed in [95]. There, the authors used one-class SVM to classify optical network-related abnormalities, such as filter impairments, which they emulated reducing the bandwidth and detuning the central frequency of a WSS.

In addition, some solutions to mitigate the filtering effects were also presented in the literature. Specifically, in [96], the authors proposed the employment of an optical WSS within every ROADM node, to spectrally shape the outgoing signals. Such configuration allowed them to transmit a 32 GBd signal through a cascade of 14 WSSs with average bandwidth equal to 33 GHz, ensuring the target performance. Similarly, in [97], the adoption of a duobinary shaping was proposed. Employing an alternative carrier modulation, more tolerant to narrow filtering, the authors demonstrated a 27% increment in the transmission distance reach. A similar approach to [96] was employed in [98] to reduce the filter-related penalties. In particular, the authors of [98] presented a method that uses only a single digital pre-equalization filter to

compensate filtering effects caused by multiple concatenated WSSs. In fact, in a real transmission system, signals may experience different degree of filtering-related penalties according not only to the cascade length, but also to the different filter shapes variations of the traversed WSSs. Finally, in [99], a different approach to mitigate the filter-related penalties was presented. There, the authors proposed and experimentally demonstrated a solution to address the detuning between signal and filter cascade, based on information monitored inside the coherent receiver. In particular, the signal-filter detuning was retrieved starting from the signal Power Spectral Density (PSD) estimated at the output of the Analog-to-Digital Converter (ADC) block. Then, such information was sent to a controller, where an algorithm calculated the corrections to be applied to the transmitter laser, in order to maximize the SNR and to close the control loop. In Chapter 6, we will present a similar approach, but applied to superchannels, in which we will leverage information acquired from the optical receivers to spectrally adjust the superchannel subchannels.

3.2.4 Spectral-Based Optical Performance Monitoring

A possible way to monitor signal and filter-related parameters also at the intermediate nodes and not only at the receiver side, relies on the processing of optical spectral data. From the optical spectra, one can retrieve information not only about the signal itself, but also indirectly, about the network conditions or the elements that compose it, like the optical amplifiers or the optical filters (i.e., monitoring any eventual soft failure). For instance, the interpolation method presented in Section 3.2.2, through which the OSNR of a channel is estimated starting from its optical spectrum, represents one of such OPM spectral-based solution. Apart from that, some spectral-based techniques have been proposed in the literature for monitoring purposes. In [100], an SDN-based QoT monitoring solution leveraging optical spectral data was experimentally demonstrated. In particular, the authors presented an experimental setup where the source and destination ROADM nodes were equipped with OSAs, which tapped at all their input and output ports. Once the controller detected a BER degradation, it requested spectral information from both transmitter and receiver. After the extraction of features like signal central frequency and 3 dB bandwidth from such spectra, the controller compared the transmitted and received values, deciding for eventual lasers reconfigurations. In [22], the authors proposed a set of ML-based solutions that exploited features retrieved from optical spectra to detect and identify soft failures. In particular, those soft failures resulting from transmitter laser drift, filter central frequency shift, and filter tightening. They considered a network-wide infrastructure, which relied on an Optical Testing Channel (OTC) for active monitoring during the commissioning testing phase, and on OSA for

passive monitoring during an operative phase. More in detail, they considered OSAs placed at every network node outgoing link, resulting in a number of OSAs equal to the nodal degree. Similarly, in [101] and [102], the authors proposed a ML and optical spectra-based method to detect and identify filter-related soft failures, such as filter shift/laser drift and filter tightening. However, unlike the approach proposed in [22], [101] and [102] presented a way to prevent misclassification of valid signals, whose bandwidth has narrowed due to FCE, from those whose bandwidth has narrowed due to failures. Such approaches relied on a series of frequency-power pairs, retrieved directly from the optical spectra captured with OSAs placed at the egress port of every node of the network. In [103], the authors compared the performance of four different ML algorithms (SVM, ANN, K-NN, and decision tree) for the estimation of parameters, such as central wavelength, OSNR, and signal bandwidth, from optical spectral data. However, they considered wide optical spectra, not available in deployed filtered networks. In fact, in realistic scenarios, we would have the spectrum of the considered signal/channel limited at its sides by the spectra of the adjacent signals/channels, as per the spectral grid standard. In [104], the authors applied optical spectral data analysis to filterless optical networks. Specifically, they proposed a method to monitor power fluctuations and TPs laser drifts, exploiting optical spectra collected by a single OSA per filterless segment. A solution based on monitoring information of existing connections to estimate filtering uncertainties and therefore improve the QoT estimation of future connections was proposed in [105]. In that paper, the authors leveraged a ML regression model and processed spectral data acquired with cheap and flexible monitors, to show in simulations an 80% reduction of the margins for new connections. In [106], an optical spectrum-based approach for joint linear and nonlinear noise monitoring was proposed and experimentally demonstrated. The authors leveraged the correlation between two slices of the signal spectrum. In particular, they considered the lower and the upper signal spectrum sidebands, which they collected using narrowband filters. Finally, optical spectra processing has also found an application for network security purposes, as shown in [107]. There, the authors implemented a ML-based approach able to detect unauthorized signals traveling in the network. Instead, a different kind of spectra were considered in [108], where the authors presented an ANN-based OSNR monitoring method, which relied on Radio Frequency (RF) spectra. They proposed to directly detect a Polarization Multiplexed-Quadrature Phase Shift Keying (PM-QPSK) signal, and to use the low-frequency part of its RF spectrum to train an ANN model.

Similarly to the approaches presented in this section, in Chapter 5, we will propose two solutions to retrieve signal- and filter-related parameters from optical spectra captured by optical monitors placed near the network nodes.

More in detail, to obtain the signal-related parameters, we will apply ML-based algorithms to spectral data collected at the node output ports, where, due to filtering effects and the proximity of adjacent channels, classic solutions such as the interpolation method cannot be employed. Regarding the filter-related parameters, the solutions reported in Chapter 5 always employed optical monitors at the node output ports, where the ASE noise effects are hidden by the filter. Instead, we will process the optical spectra acquired at the node input ports, and to overcome the presence of the ASE noise, we will leverage the curve fitting principle.

3.3 Superchannels

The flex-grid DWDM paradigm defined in [8] allows the implementation of frequency slots with different spectral widths [109]. Such standard envisioned for the spectral slots a central frequency and a width granularity of 6.25 and 12.5 GHz, respectively. In addition, the slots can be combined to satisfy the different requirements that each connection can have, enabling the transmission of heterogeneous traffic (e.g., channels with mixed widths) over the same network.

In Section 3.2.3, we analyzed the key role played by the ROADM nodes within the flex-grid optical network and the filtering issues introduced by the WSSs composing them. These aspects exacerbate even more when the aforementioned spectral densification is considered. A possible way to cope with such filtering-related penalties is represented by superchannels. A superchannel includes multiple channels (or, subchannels, which is the term used in the rest of this thesis) allocated into a continuous set of spectral slots with a limited guard band between them [110]. In FIGURE 3.4, we show a

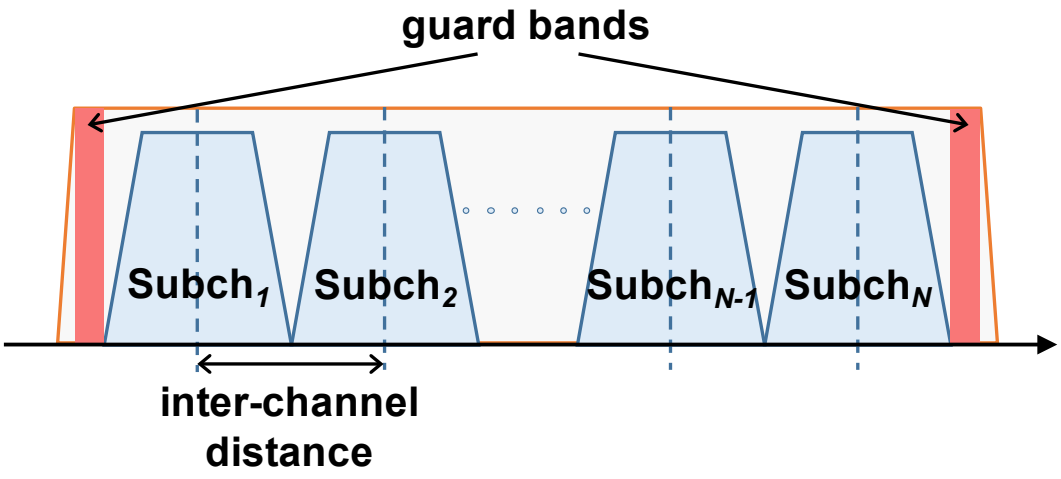

FIGURE 3.4. Schematic representation of a N-subchannel superchannel.

schematic representation of a superchannel composed by a set of *N* subchannels. When the superchannel traverses the optical network nodes (i.e., when it crosses the ROADM-based nodes and the WSSs that constitute them), the set of subchannels that composes it behaves as a single entity [111]. So, the WSS filters within the ROADM nodes are configured to switch all the slots occupied by the superchannel. Such configuration assumed within the WSS is sometimes referred to as superfilter. It allows to reduce the inter-channel distances and to minimize those between the two outer channels and the filter, which are also referred to as guard bands, therefore improving the total spectral efficiency.

Additionally, nowadays, SDM technology represents one of the most attracting solution for the telecommunication operators to meet the ever increasing capacity requests [2]. In fact, in SDM-based networks, either multi-fiber link or multicore/multimode fibers are considered. Hence, the single channel granularity of the ROADM-based networks will most likely evolve towards a multichannel approach. In this context, since superchannels are the next level of switching granularity, we can envision their wide employment in the near future [112].

A common metric to describe the superchannel spectral efficiency is to refer to their corresponding WDM class. In turn, the WDM classes are categorized according to the value assumed by the ratio between the distance of the subchannels and the symbol rate of the signal travelling within them [113]. In particular, if such ratio is between 1 and 1.2 (e.g., a 64 GBd PM-QPSK signal with a subchannel spacing of 75 GHz), then the WDM class is known as quasi-Nyquist-WDM. On the contrary, if the distance between subchannels equals the symbol rate (i.e., a ratio equal to 1), the WDM class is known as Nyquist-WDM. Finally, if the spacing is lower than the symbol rate (i.e., ratio less than 1), the WDM class is known as super-Nyquist-WDM. It is important to note that this classification applies to equidistant and uniform superchannels. The former feature refers to superchannel with equally spaced subchannels, whereas the latter refers to superchannels, where all the subchannels composing them have the same symbol rate and roll-off factor.

Although superchannels seem to represent an agile and performing solution, they also present some drawbacks. In particular, their main limitations are the interference occurring between their subchannels, due to the considered reduced inter-channel distances, and the penalties introduced by the filter, which affect the two external subchannels. The former impairment comes in two forms: i) Cross-Channel Nonlinear Interference (XCI), which although occurs among all channels mainly affects the close adjacent ones, and ii) linear crosstalk, which only occurs between adjacent ones and depends on the considered inter-channel distances and on the signal roll-off factors. Linear

crosstalk can be caused by uncontrolled drifts of the transmitter lasers, for instance as consequence of the network ageing, and translates into degradations of the received channels SNR values [114]. On the other hand, the filtering penalties mainly affect the two external superchannel subchannels (i.e., subchannels *1* and *N* in FIGURE 3.4). In fact, especially in presence of laser/filter cascade misalignment, the two outer subchannels SNR values could be strongly degraded by the filter effect. To compensate for such impairment, the superchannel can be operated at lower modulation formats and/or an additional spectral guardband between the subchannels can be adopted, degrading the spectral efficiency. Of course, the best solution would be to monitor and eventually correct any misalignment occurring between lasers and filters.

Usually, the most common way to optimize the transmission of a superchannel is to employ and target to maintain an equidistant configuration [110]. However, such frequency spacing can be limited by the eventual imperfections of the subchannel transmitters and receivers and/or by their laser central frequency drifts, and by any wavelength dependent loss/performance variation. All these factors result in variations of the subchannel SNR and BER values, and make the equidistant configuration not the optimal solution. To address such issues, an approach that leverages the network feedback need to be considered, therefore taking into account its current conditions and act accordingly. To this aim, in Chapter 6, we will present a solution that uses a closed control loop to optimize the spectral spacing of the subchannels in a superchannel.

3.4 Reference Optical Network Architecture

Relying on the concepts introduced in the previous sections of this chapter, here we identify an optical network architecture to be used as a reference for the rest of this thesis. In particular, we consider such reference architecture to be constituted by a ROADM-based WDM optical network, widely equipped with spectral monitors and supervised by a central controller. We also assume that the links of the considered optical network are constituted by optical fiber spans and Erbium-Doped Fiber Amplifiers (EDFAs). Furthermore, we do not assume any specific aggregation/disaggregation scenario, but we define it individually in each considered case. Finally, we assume that within the central controller, the information is processed and leveraged for any eventual applications, for example, those related to optimization. In FIGURE 3.5, we depict a schematic view of the considered reference scenario. There, we also define each task addressed within this thesis.

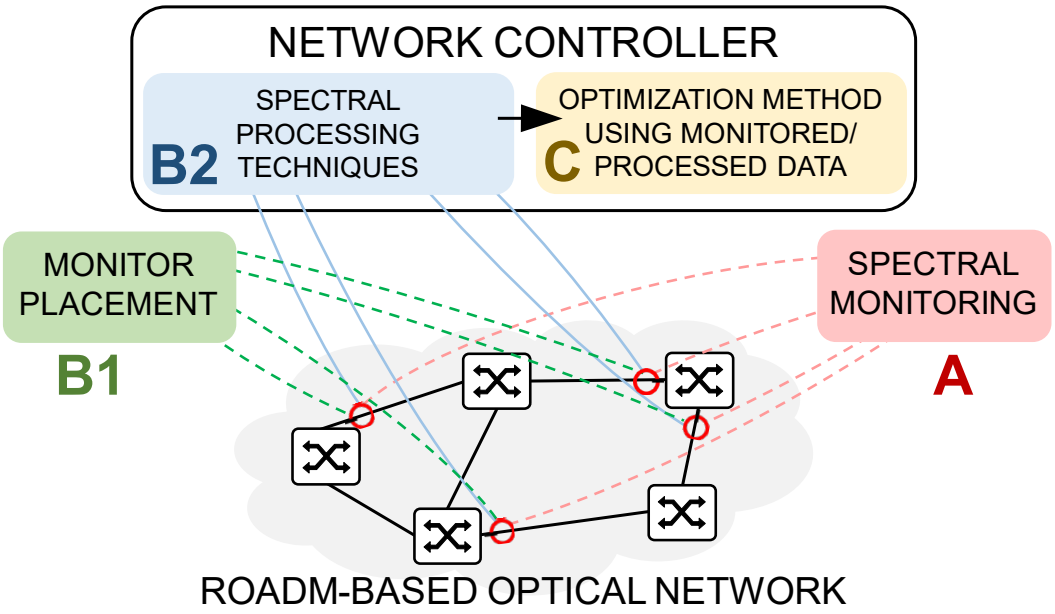

FIGURE 3.5. Reference optical network architecture scenario along with the task addressed within the thesis. Task A: spectral monitoring; Task B1: optical monitors placement; Task B2: spectral processing technique; Task C: optimization done leveraging the monitored/processed data.

In particular, task A corresponds to the definition and the study of a monitoring scheme for spectral data collection. We report about such work in Chapter 4. Then, task B is divided into two main subtasks: identifying adequate optical monitors placement scenarios and investigating processing techniques suitable for such considered placement scenarios. In particular, the spectral data collected through the optical monitors spread throughout the network are first transferred to the controller and then processed in that same framework. An alternative solution would be to process the acquired spectral data directly at the monitor locations and then transfer the extracted features to the network controller. The latter approach would reduce the transferred data / the management overhead but would require some processing power at the monitor locations and would lose some accuracy in solutions requiring the combination/correlation of information from various monitors. Comparing these two control/management scenarios could also represent a possible direction to extend further the work presented in this thesis. Tasks B1 and B2 are addressed in Chapter 5. Finally, optimization algorithms can be implemented within the network controller framework to leverage the monitored data or the extracted features, corresponding to task C. In Chapter 6, we will show an example of such optimization applied to superchannels.

Chapter 4

Data Plane Architectures for High-Resolution Optical Spectrum Estimation

Based on the reference architecture proposed in Section 3.4, in this chapter we present a solution to address the spectral monitoring task, i.e., task A of FIGURE 3.5. In particular, in this chapter, we identify the data plane elements for modulation format-agnostic OPM [115].

More in detail, this chapter is organized as follows. In Section 4.1 we provide an overview of the main concepts and components of the proposed scheme. In Section 4.2 we present a theoretical study of such solution, corroborated by its mathematical description. In Section 4.3, we review the requirements and the potential benefits deriving from the implementation of the proposed solution, and we provide a list of commercially available devices that could be eventually employed for its experimental validation. Finally, in Section 4.4, we draw the conclusions.

4.1 Introduction

In [115], we proposed the theoretical study of an agile optoelectronic front-end scheme for nonintrusive monitoring, transparent to the modulation format and able to deliver arbitrary spectral resolution. The solution we proposed, which we schematically depicted in FIGURE 4.1, was composed by two main blocks:

1. An OSA, whose main parts are an Optical-Electrical (OE) front-end, an ADC, and adequate DSP;

2. A suitable monitoring agent, whose roles are:

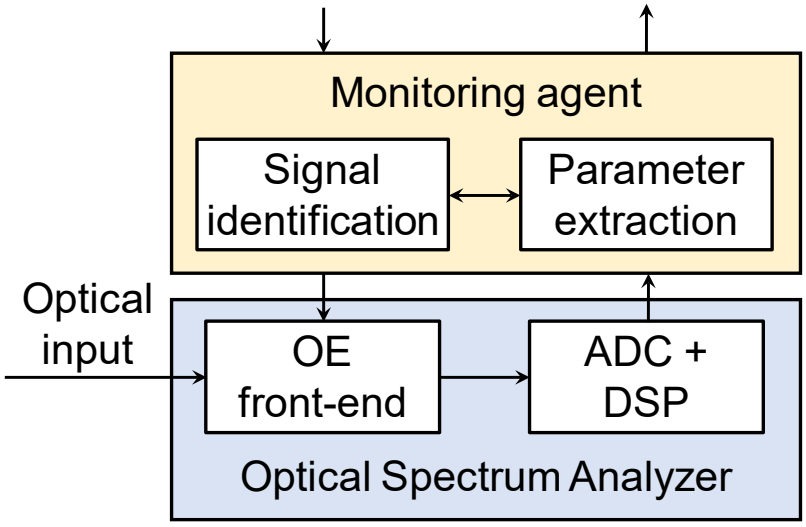

FIGURE 4.1. Scheme of the proposed monitoring solution. O-E: Optical-Electrical; ADC: Analog-to-Digital Converter; DSP: Digital Signal Processing.

a. Control of the OSA by means of an agreed API or, in case of a full custom design, by low-level commands interfacing the different devices composing the OSA;

b. Extraction of the different parameters and figures of merit from the optical spectra provided by the OSA (e.g., signal power level, OSNR, etc.);

c. Provision of a communication interface for the interaction of the physical device with the control plane.

In particular, our solution leverages a custom OSA design which is based on coherent detection for the acquisition of different spectral slices of a Signal Under Test (SUT). Once detected, such slices are processed and merged together, enabling the reconstruction of the full original optical signal spectrum with a high spectral resolution that depends on the width of the spectral slices and on the phase noise.

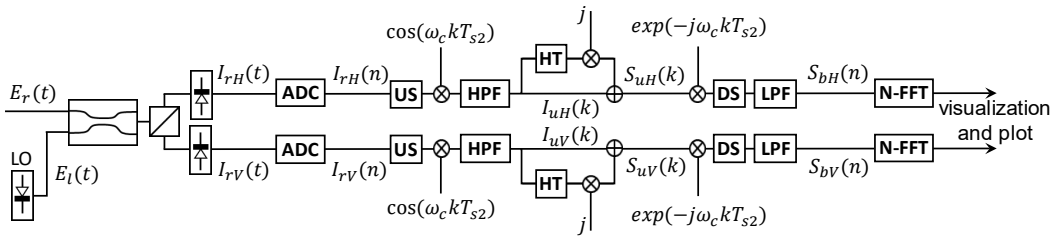

FIGURE 4.2. Block-diagram of the proposed coherent detection-based front-end. LO: Local Oscillator; ADC: Analog-to-Digital Converter; US: Up-Sampler; HPF: High-Pass Filter; HT: Hilbert Transform; DS: Down-Sampler; LPF: Low-Pass Filter; N-FFT: N-Fast Fourier Transform.

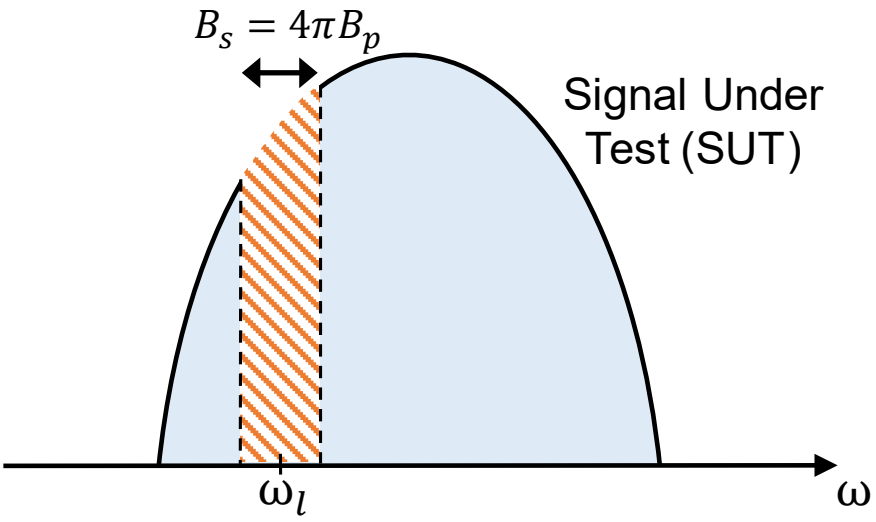

FIGURE 4.3. Principle of the SUT slicing along with the relation between slice width and photodiode bandwidth. SUT: Signal Under Test; B_s: slice width; B_p: photodiode bandwidth.

A block diagram of the proposed scheme is shown in FIGURE 4.2. Briefly, the first part of such solution consists mainly of a coherent detector featuring polarization diversity [116]. Here, the optical SUT is coupled with the signal of a local oscillator and their output is fed to a Polarization Beam Splitter (PBS). In particular, the role of the PBS is to separate the two polarization components. Then, in order to detect such two signal polarization components, we place two photodetectors in cascade to the PBS. After the digitalization stage, which we implement using two ADCs, the DSP part takes place. In particular, we store the two digitalized output currents in two N samples arrays, upsampling, upconverting and then filtering them, by means of a High-Pass Filter (HPF). Hence, in order to reconstruct the two complex signals, we apply the Hilbert Transform (HT) and again downconvert, downsample and low-pass filter the signals. Finally, applying the Fast Fourier Transform (FFT) we obtain the desired optical spectral slice of the SUT. In such proposed scheme, the SUT spectral slice width B_s only depends on the frequency of the coherent detection laser ω_l and on the bandwidth B_p of the two photodiodes, as shown in FIGURE 4.3. In the upcoming section we will formally describe all these stages.

4.2 Optical front-end concept

As mentioned in Section 4.1, the aim of the proposed solution is to detect different spectral slices of a SUT and to calculate their optical spectra. To do so, we first have to couple the SUT with the signal of a local laser, which oscillates at a frequency ω_l, as shown in FIGURE 4.2.

Leveraging the corresponding Jones vectors, we represent the sliced SUT and the local oscillator signals with $E_r(t)$ and $E_l(t)$, respectively, such as

$$E_r(t) = \begin{bmatrix} \sqrt{P_r(t)}\cos\psi_r \exp\left(j\omega_r t + j\phi_r(t)\right) \\ \sqrt{P_r(t)}\sin\psi_r \exp(j\omega_r t + j\phi_r(t) + \theta_r) \end{bmatrix}, \tag{4.1}$$

$$E_l(t) = \begin{bmatrix} \sqrt{P_l(t)}\cos\psi_l \exp\left(j\omega_l t + j\phi_l(t)\right) \\ \sqrt{P_l(t)}\sin\psi_l \exp(j\omega_l t + j\phi_l(t) + \theta_l) \end{bmatrix}, \tag{4.2}$$

where $P_r(t)$ and $P_l(t)$ represent the SUT slice power and the local oscillator signal power, ω_r and ω_l are the angular frequencies, $\phi_r(t)$ and $\phi_l(t)$ represent two generic phases accounting for phase noise and others, ψ_r and ψ_l are the azimuth of the states of polarization (SOPs), θ_r and θ_l are the polarization ellipticities, and t represents the time.

Considering the PBS as the SOP reference, we assume the SOP and the ellipticity of $E_l(t)$, to be

$$\psi_l = 45°\,, \tag{4.3}$$

$$\theta_l \simeq 0\,. \tag{4.4}$$

Hence, the Jones vector components of $E_l(t)$ become equal and Equation 4.2 becomes

$$E_l(t) \cong \begin{bmatrix} \sqrt{\frac{P_l}{2}}\exp\left(j\omega_l t + j\phi_l(t)\right) \\ \sqrt{\frac{P_l}{2}}\exp\left(j\omega_l t + j\phi_l(t)\right) \end{bmatrix}. \tag{4.5}$$

Then, the two polarization components of the coupled signal are detected by two photodiodes, whose output currents can be expressed as

$$I_{rH}(t) = \frac{R}{2}\left((1-\alpha)P_r(t)cos^2\psi_r + \alpha\, P_l/2\right) + R\sqrt{\frac{\alpha(1-\alpha)P_r(t)P_l}{2}}\cos\psi_r\cos\left(\Delta\omega t + \phi_e(t)\right), \tag{4.6}$$

$$I_{rV}(t) = \frac{R}{2}\left((1-\alpha)P_r(t)sin^2\psi_r + \alpha\, P_l/2\right) + R\sqrt{\frac{\alpha(1-\alpha)P_r(t)P_l}{2}}\sin\psi_r\cos(\Delta\omega t + \phi_e(t) + \theta_r), \tag{4.7}$$

where R represents the photodiode responsivity, α represents the coupling ratio, $\Delta\omega$ represents the difference between the frequency of the considered SUT slice and the frequency of the local oscillator, such as

$$\Delta\omega = \omega_r - \omega_l\,, \tag{4.8}$$

and $\phi_e(t)$ represents the difference between the phase of the considered SUT slice and the phase of the local oscillator, such as

$$\phi_e(t) = \phi_r(t) - \phi_l(t)\,. \tag{4.9}$$

Employing a 50:50 coupler (i.e. considering a coupling ratio $\alpha = 1/2$) and assuming that in Equation 4.6 and 4.7 the terms related to the coherent detection (i.e., the second addends) are bigger than those representing the direct detection (i.e., the first addends), we can simplify them as

$$I_{rH}(t) = R\sqrt{\frac{P_r(t)P_l}{8}}\cos\psi_r\cos\big(\Delta\omega t + \phi_e(t)\big), \tag{4.10}$$

$$I_{rV}(t) = R\sqrt{\frac{P_r(t)P_l}{8}}\sin\psi_r\cos(\Delta\omega t + \phi_e(t) + \theta_r). \tag{4.11}$$

Observing Equations 4.10 and 4.11, it can be noticed that for both the currents, the only component left is the real one, along with the SOP information. In addition, their bandwidths are also limited by the photodiode bandwidth B_p.

Thus, the two signals are digitized using the ADC and stored in arrays of N samples, becoming

$$I_{rH}(n) = R\sqrt{\frac{P_r(nT_s)P_l}{8}}\cos\psi_r\cos\big(\Delta\omega nT_s + \phi_e(nT_s)\big), \tag{4.12}$$

$$I_{rV}(n) = R\sqrt{\frac{P_r(nT_s)P_l}{8}}\sin\psi_r\cos(\Delta\omega nT_s + \phi_e(nT_s) + \theta_r), \tag{4.13}$$

where $1 \le n \le N$ and T_s, which represents the sampling period, is defined as

$$T_s = \frac{1}{f_s}\,, \tag{4.14}$$

with f_s representing the sampling frequency.

Then, we upsample I_{rH} and I_{rV} to a frequency $f_{s2} \gg f_s$, upconvert them to an angular frequency $\omega_c \gg 2\pi B_p$, and finally high-pass filter them. The results of such operations lead to the following:

$$I_{uH}(k) = R\sqrt{\frac{P_r(kT_{s2})P_l}{32}}\cos\psi_r\cos\big((\Delta\omega+\omega_c)\mathrm{k}T_{s2}+\phi_e(kT_{s2})\big), \quad (4.15)$$

$$I_{uV}(k) = R\sqrt{\frac{P_r(kT_{s2})P_l}{32}}\sin\psi_r\cos\big((\Delta\omega+\omega_c)\mathrm{k}T_{s2}+\phi_e(kT_{s2})+\theta_r\big), \quad (4.16)$$

where $1 \le k \le M$ with M representing the number of samples of the array, M/N is the upsampling ratio, and T_{s2} represents the upsampling period, such as

$$T_{s2} = \frac{1}{f_{s2}}. \quad (4.17)$$

To reconstruct the corresponding complex signals, we first need to retrieve the imaginary parts of each detected signal. To do so, we can use the Bedrosian theorem and apply the HT to I_{uH} and I_{uV} [117], obtaining

$$S_{uH}(k) = R\sqrt{\frac{P_r(kT_{s2})P_l}{32}}\cos\psi_r\exp\big(j(\Delta\omega+\omega_c)kT_{s2}+j\phi_e(kT_{s2})\big), \quad (4.18)$$

$$S_{uV}(k) = R\sqrt{\frac{P_r(kT_{s2})P_l}{32}}\sin\psi_r\exp\big(j(\Delta\omega+\omega_c)kT_{s2}+j\phi_e(kT_{s2})+j\theta_r\big). \quad (4.19)$$

Then, we downconvert, downsample and low-pass filter $S_{uH}(k)$ and $S_{uV}(k)$, obtaining

$$S_{bH}(n) = R\sqrt{\frac{P_r(nT_s)P_l}{32}}\cos\psi_r\exp\big(j\Delta\omega nT_s+j\phi_e(nT_s)\big), \quad (4.20)$$

$$S_{bV}(n) = R\sqrt{\frac{P_r(nT_s)P_l}{32}}\sin\psi_r\exp\big(j\Delta\omega nT_s+j\phi_e(nT_s)+j\theta_r\big), \quad (4.21)$$

where S_{bH} and S_{bV} are the baseband representation of the detected SUT slice. Such signals depend on the frequency and phase differences between local oscillator and SUT slice, respectively $\Delta\omega$ and $\phi_e(nT_s)$.

Finally, by applying the N-point FFT, the desired slice of the SUT optical spectrum can be obtained.

4.3 Requirements and Performance

The proposed scheme comes with some requirements, a few of which already mentioned in the previous sections. For example, the slice frequency width B_s depends on the photodetectors bandwidth B_p. The relation between these two values is

$$B_s = \omega_l \pm 2\pi B_p \,. \tag{4.22}$$

The main consequence of such relation is that B_p limits the step width taken by the local oscillator to a maximum of $4\pi B_p$, as shown in FIGURE 4.3. Therefore, in the case of very coarse central frequency tuning sensibility local lasers, photodiodes with very large bandwidth are needed. In addition, B_p is also related to the ADC sampling frequency f_s. Hence, to avoid any aliasing effect, the following condition must be satisfied:

$$f_s \geq 2B_p \,. \tag{4.23}$$

In turn, f_s limits the frequency resolution that can be obtained after the FFT to

$$f_{RES} \leq \frac{f_s}{N} \,. \tag{4.24}$$

Another critical relation to consider is the one that binds the reconstructed signal slices S_{bH} and S_{bV} to the phase and frequency differences between the local oscillator and the original signal slice. In order to mitigate this dependency, phase and frequency estimation techniques should be applied.

Finally, since the frequency range covered by the proposed technique is mainly limited by the local laser, a broadly tuneable laser is preferable.

The main advantage of our proposed scheme is that there is no need for a continuous frequency sweep, as in the so-called Coherent Optical Spectrum Analyzer (COSA)-based techniques [118]. Furthermore, being the proposed front-end scheme transparent to the optical signal waveform (i.e., independent from the used modulation format and multiplexing scheme), it represents a valuable solution to provide high-resolution optical spectra, from which the most important figures of merit can be accurately estimated. Finally, since it is a modular approach, the presented DSP modules could also be enhanced by leveraging one of the ML algorithms described in Chapter 2.

The solution we proposed constitutes a trade-off between cost and performance. Although a precise analysis might need detailed simulations and eventual experimentation, here we give hints of an actual implementation using commercially available devices.

The employed optical sources should be stable, broadly tuneable, and provide low linewidth. Ideally this is achieved with external cavity lasers, which can provide narrow linewidths down to 150 KHz. Also, they are slow-tuning and high-performant devices, and it would take around 1 s to be tuned at a precise wavelength with a ~1 pm resolution [119]. It is worth noting that these devices can cover different bands (e.g., S, C, and L). Also, low linewidth is needed to maintain high spectral resolution after the FFT. In case we can tolerate linewidths in the MHz range, we can approach a (fast) tuneable laser, able to be finely tuned in less than 3 ns [120]. Such devices can be stabilized using an appropriate wavelength locking circuit, with a resolution accuracy of ±64 MHz [121]. Therefore, with such resolution, the whole C-band (i.e., from 1530 to 1565 nm) can be scanned within 100 μs approximately.

Regarding the ADCs, we envision using common high-speed ADCs up to 10 GSa/s [122]. However, such devices require low bandwidth photodetectors and finely tuneable lasers. In case high-performance ADCs are considered, a proposal could be to have 103 GSa/s ADCs [123], which represent a more expensive option, but less stringent in terms of laser tunability/stability. Considering the former solution, we could easily achieve high resolutions with small FFT sizes. For instance, a 128-point FFT would result in ~78 MHz spectral resolution, whereas 64-point would lead to ~156 MHz. On the contrary, adopting the high performance ADCs, in order to achieve ~100 MHz resolution, we would need a 1024-point FFT.

Finally, a broad range of balanced photodetectors is commercially available, featuring bandwidths up to 100 GHz [124]. So, these components are found to be the less limiting of all.

In the following, we propose two different implementation examples. Implementation #1 includes a fast tuneable laser with a 10G photodetector and a 10 GSa/s ADC. Instead, implementation #2 involves using an external cavity laser with a high-speed photodetector and a 103 GSa/s ADC. We report a summary of the performance of such proposed schemes in TABLE 4.1.

TABLE 4.1. Performance summary of the proposed implementation examples.

	Implementation #1	Implementation #2
ADC sampling rate	10 GSa/s	103 GSa/s
FFT size	64 points	1024 points
Resolution bandwidth	156 MHz	100 MHz
Maximum power	20 dBm	20 dBm
Sweeping speed	37.7 ns/nm	1.2 s/nm
Wavelength Range	1530 nm – 1565 nm	1460 nm – 1565 nm

More in detail, in the first implementation, we found the maximum slice bandwidth to be B_s=10 GHz due to the ADC sample rate,. Hence, the step width adopted by the laser to cover the monitored spectral region is 10 GHz. In addition, in this implementation, the laser can cover only the C-band [120], whereas each point can be set within around 3 ns. Thus, it can sweep the entire spectrum within 1.32 μs (at a speed of around 37.7 ns/nm). Regarding the power range, we retrieved P_r by comparing the direct and coherent detection terms of Equation 4.6. In particular, we assumed a local oscillator signal power P_l=10 dBm, a photodiode responsivity R=0.9 [125], and a coupling ratio α=0.5. Such an equation resulted in a maximum power of 20 dBm.

Concerning the second considered implementation scheme (i.e., implementation #2), we proposed to use a 103 GSa/s ADC, entailing a maximum slice bandwidth B_s=103 GHz. So, the external cavity laser is expected to cover the S, C, and L bands with steps of 103 GHz. Since setting to each point requires 1 s, it can take about two minutes (126 s) for the laser to cover the whole spectrum (S, C, and L bands) at a speed of 1.2 s/nm. As per implementation #1, we calculated the power range by comparing the direct and coherent detection terms of Equation 4.6. However, this time we considered a photodiode responsivity R=0.5 [124], which again resulted in a maximum power of 20 dBm.

4.4 Conclusions

In this chapter, we presented a method to estimate the signal optical spectra at high resolutions. Since the technique we presented relies on detecting spectral slices of the considered signal, it does not require a continuous frequency sweep. Along with the theoretical study of the proposed solution, we also presented the technical requirements and two possible examples of its implementation.

Chapter 5

Optical Monitor Placement Strategies and Assessment

In this chapter, we focus on the second major task addressed in this thesis, task B, which according to the definition introduced in Section 3.4 is composed of two subtasks, as shown in the scheme of FIGURE 3.5. The first one, subtask B1, refers to identifying adequate strategies for the placement of optical spectral monitors within the optical network. To this aim, in [126] and in [127], we defined an ingress and an egress placement scenario. The second subtask, B2, identifies two spectral-based solutions for signal and filter-related parameters estimation within the two previously identified placement scenarios. In particular, in [19], we proposed and experimentally validated an ML-based method for estimating the ASE noise relying on optical spectral data collected in the egress monitor placement scenario. We further experimentally verified such a solution under different circumstances in [128] and [127]. In [126], we proposed and experimentally validated a method to extract filter-related parameters from optical spectral data acquired within the ingress placement scenario. We further validate such a solution in [127]. Finally, in [127], we compared the two identified placement scenarios and the accuracy of the related methods for the parameter estimation, providing some guidelines on their employment. In this chapter, when we mention the monitored spectral data, we refer to the monitored PSD; therefore, we will use these two terms interchangeably.

More in detail, this chapter is organized as follows. In Section 5.1, we introduce the two identified optical monitor placement scenarios. Section 5.2 describes the proposed solution to estimate the ASE noise value from the optical spectral data acquired within the egress placement scenario. Along with the method principles, we also describe the implemented simulation and experimental setups for validating it. Section 5.3 reports about the filter feature extraction method, which relies on optical spectral data collected within the ingress placement scenario. In addition, we also describe the related simulation and

experimental setup that we implemented to validate the proposed solution. Then, in Section 5.4, we describe the simulation and experimental setups we implemented to jointly validate the two proposed solutions, reporting the obtained results. Furthermore, we compare the two methods, providing guidelines on their implementation and their impact on the SNR penalty reduction. Finally, in Section 5.5, conclusions are drawn.

5.1 Introduction

In an ideal scenario, an optical network would be equipped with high-spectral resolution and omnipresent optical spectral monitors. In general, to better monitor the relevant network parameters, the most appropriate position to place the spectral monitoring devices is close to the network ROADM nodes. Hence, such an ideal scenario would require optical spectral monitors placed before and after every node of the network, as shown in FIGURE 5.1. We assume this kind of solution to be expensive and impractical to be applied in real optical networks. Thus, more realistic configurations with a reduced number of spectral monitors are desirable. A possible way to do so is to limit or suitably select the placement of the monitoring devices within the optical network, as we proposed in [126] and [127].

In particular, we identified two main placement strategies, as we also show in FIGURE 5.2. The first one, which we defined as the egress placement scenario, envisioned the presence of spectral monitors after the egress ports of every network ROADM node, that is, after their egress WSSs. On the contrary, the second identified monitoring strategy, which we named as the ingress placement scenario, envisioned spectral monitors placed before the ingress

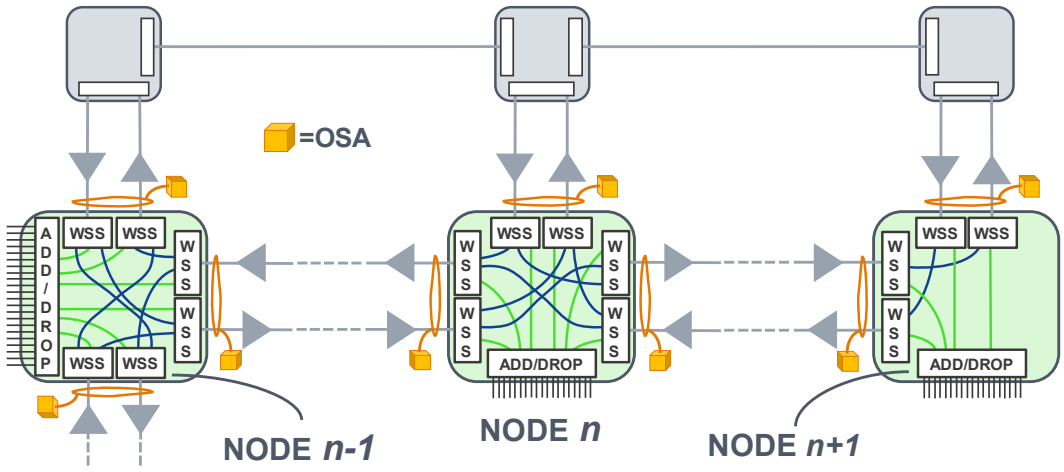

FIGURE 5.1. Ideal optical spectral monitors placement scenario.

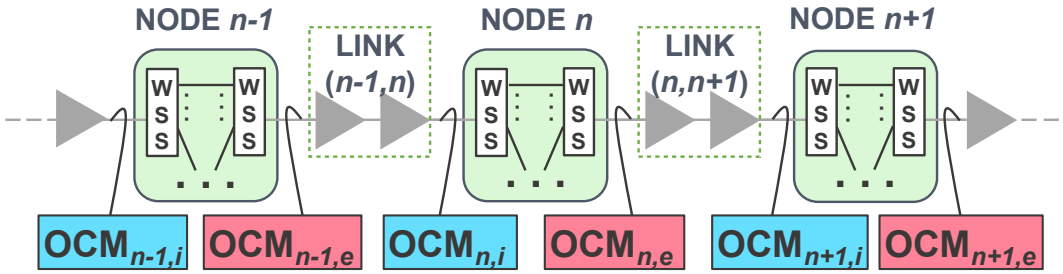

FIGURE 5.2. Considered network architecture with the two proposed optical monitors placement strategies: ingress (blue boxes) and egress (red boxes). WSS: Wavelength Selective Switch; OCM: Optical Channel Monitor.

ports of every network ROADM node, that is, before their ingress WSSs. In both scenarios, the spectral monitoring devices we considered were the Optical Channel Monitors (OCMs), cost-effective monitoring solutions available nowadays on the market, providing resolutions up to the sub-GHz order [129].

Depending on the chosen monitoring strategy, some parameters can be easily estimated from the optical spectral data, whereas others cannot. Thus, in order to retrieve the missing information of each monitoring scenario, new processing solutions are needed. To this aim, in the following sections, we will present two approaches to enhance the collected spectral data with proper data analytics and ML methods. It is worth mentioning that both the estimation methods we will report focused on different aspects of the transmission without considering the consequences of the Nonlinear Interference (NLI) caused by the Kerr effect. In fact, all the spectra we collected and processed referred to a single-channel configuration. An accurate study of the NLI impact on our estimation solutions is left for future investigations.

5.2 OSNR Estimation in an Egress Monitoring Placement Scenario

As introduced in Section 5.1, the egress monitoring scenario envisions the placement of the OCMs after the output ports of each network ROADM node, as represented in FIGURE 5.2 through the red boxes. Hence, in such scenarios, we monitor the signals just after being filtered by the WSSs placed within the ROADM nodes. Consequently, filter-related parameters such as the filter 3/6 dB bandwidth and central frequency can be easily estimated from the collected spectral data. In fact, the filter directly affects the signal spectra traversing the node, as shown in FIGURE 5.3, where we depicted a signal spectrum before and after being filtered. On the other hand, signal-related information such as its 3/6 dB bandwidth and the associated ASE noise

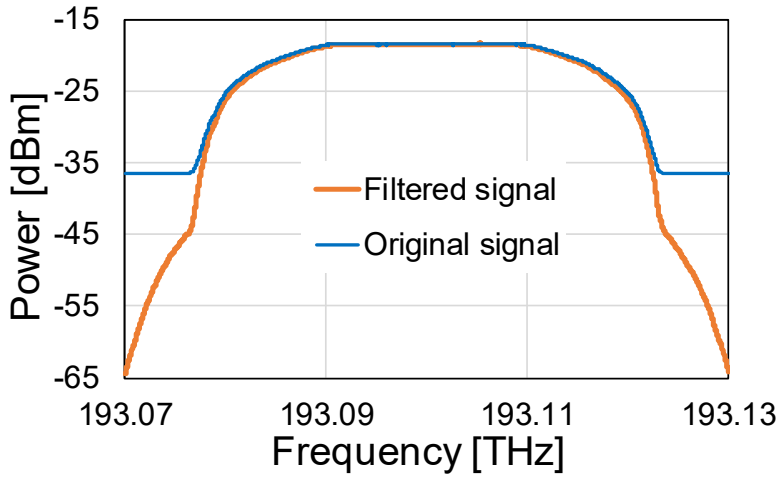

FIGURE 5.3. Signal optical spectra acquired before (orange plot) and after (blue plot) the filtering process. The original signal bandwidth and its related ASE noise cannot be directly measured from the blue plot, whereas the orange plot clearly shows the filter characteristics.

(directly mapped into the OSNR metric) cannot be directly measured from the acquired spectral data, as shown in FIGURE 5.3. A possible solution to overcome this limitation is represented by ML-based algorithms, which can infer such information with a certain accuracy even from filtered spectral data. To this aim, in [19], we proposed and validated using simulations and an experimental setup, a supervised ML-based method to estimate in-band the ASE noise/OSNR from optical spectral data collected in such egress optical monitor placement scenario. In addition, in [127] and in [128], we carried out further experimental demonstrations of such a solution under different circumstances.

As a first step, our method requires the collection of different sets of optical spectra to be classified according to some particular signal parameters (e.g., roll-off factor, symbol rate, etc.) and labeled with their corresponding ASE noise/OSNR values. Then, leveraging the labeled data, we train separate ML regression models for the different considered classes of parameters and predict the OSNR of unlabeled spectral data. The following paragraph will formally present such an OSNR estimation solution, referring to the formulation we initially presented in [19] and [128].

We represent an optical spectrum instance acquired right after an optical filter, with a vector $\mathbf{s}$ of length l, and its corresponding OSNR value with the scalar y. Our goal is to find the mapping f, between the spectrum $\mathbf{s}$ and its OSNR value y, that is $y = f(\mathbf{s})$. In addition, we denote with $\mathbf{S}_c$, the matrix of dimensions $l \times m$ representing the set of m collected optical spectra with the same group of parameters c, for which the model is valid. In particular, $c = (r, b, q)$, with r being the spectral resolution of the optical monitor that captures the spectra,

b being the bandwidth of the optical filter that filtered the considered signals, and q being the connection symbol rate. The length l of the optical spectrum vector $\mathbf{s}$ depends on r and on b, so that $l = b/r$. We also denote with $\mathbf{y}_c$ the vector of length m containing the OSNR values corresponding to the $\mathbf{S}_c$ spectra. In order to approximate the estimation function f, we implement a ML model Q_c specific for the group of parameters c. We then train Q_c with the sets of monitored spectra and labeling OSNR values $(\mathbf{S}_c, \mathbf{y}_c)$. Note that the parameter c mentioned above corresponds to nominal TP and filter values, so we can create a training set either during a lab calibration phase or directly in the field, right before commissioning a connection. Concerning the first option, we could implement different scenarios simply by tuning the available parameters (e.g., channel roll-off factor, modulation format, symbol rate, monitor spectral resolution, filter 3/6 dB bandwidth) and collecting the corresponding optical spectra (see the following sections). Instead, several limitations should be considered when operating in the field to avoid interfering with existing connections. In fact, parameters like those related to the node filters cannot be varied in operating networks, where such changes could influence the adjacent connections. Furthermore, if we denote the estimated OSNR values with the vector $\hat{\mathbf{y}}_c = Q_c(\mathbf{S}_c)$, then we can represent the estimation errors with the vector $\boldsymbol{\varepsilon}_c = \hat{\mathbf{y}}_c - \mathbf{y}_c$. Hence, the goal of the training process will be to identify the model Q_c, which minimizes some functions of $\boldsymbol{\varepsilon}_c$, such for example, the Mean Squared Error (MSE) function. Finally, once the ML algorithm is trained with the set of spectra $\mathbf{S}_c$ and their reference OSNR values $\mathbf{y}_c$, it will be able to estimate from the optical spectrum $\mathbf{s}_c^h$, of an operating channel h with parameters c, its OSNR value $\hat{y}_c^h$.

5.2.1 OSNR Estimation Method Assessment

First Experimental Setup

We carried out the first experimental validation of the proposed solution in [19], where we leveraged high-resolution spectral data captured using a Brillouin Optical Spectrum Analyzer (BOSA) [130]. This monitoring device exploits the Stimulated Brillouin Scattering (SBS), a nonlinear optical effect that causes very narrow filtering, allowing the BOSA to achieve spectral resolutions up to 12.5 MHz (0.1 pm) and a dynamic range greater than 80 dB [131]. We used the BOSA-collected spectral data to train two ML regression algorithms: a GPR and an SVR model. In addition, to evaluate the impact of the monitor spectral resolution on the OSNR estimation accuracy, we also employed lower-resolution spectral inputs obtained post-processing the collected optical spectra. The experimental setup we implemented is shown in FIGURE 5.4. We generated a 28 GBd PM-QPSK modulated signal with a

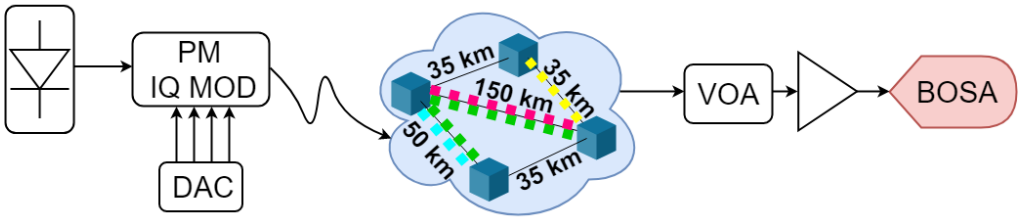

FIGURE 5.4. Block diagram of the experimental setup we proposed in [19]. PM-IQ-MOD: Polarization Multiplexed-IQ-Modulator; DAC: Digital-to-Analog Converter; VOA: Variable Optical Attenuator; BOSA: Brillouin Optical Spectrum Analyzer.

tuneable laser centered at 1550.918 nm (i.e., 193.3008 THz) and no pulse-shaping. We considered a Back-to-Back (B2B) scenario and, leveraging CTTC's ADRENALINE testbed, transmitted the signal over four different distance paths: 35, 50, 150, and 200 km. In order to emulate a further cascade of fiber spans, at the testbed output, we placed a Variable Optical Attenuator (VOA) and then an EDFA operating in Automatic Power Control (APC) mode. Using the VOA, we obtained sixteen different OSNR levels. Finally, we acquired the optical spectra with the BOSA working at its highest spectral resolution, that was, 12.5 MHz. We retrieved the OSNR values to be used as references during the labeling phase from such high-resolution spectral data, employing the integral method. The integral method calculates the signal and noise integrals within two different bandwidths (usually 0.4 nm for the signal and 0.1 nm for the noise). After that, the noise integral value is subtracted from the signal one in the linear domain, and their result is again divided by the noise value. For this operation, we considered signal and noise bandwidths of 50 and 12.5 GHz (i.e., 0.4 and 0.1 nm), respectively. In operating networks, we could have measured such values during the signal provisioning phase, that is, before the channel operates, adopting the On/Off method [67].

For each of the five distance scenarios, we collected a total of 160 high-resolution optical spectra, specifically ten for each VOA level (five for each polarization state). We also excluded from the considered data set those optical spectra whose reference OSNR values were lower than 8 dB. We did such a separation to emulate a real network scenario, where low OSNR signals would not be kept in operation. It is worth noting, that every time the optical signals entered or left the ADRENALINE testbed, they also crossed an optical filter. In particular, the 35 and 50 km distance scenarios had an entry/exit optical filter with 100 GHz 3 dB bandwidth, whereas the 150 and 200 km distance scenarios had an entry filter with 100 GHz 3 dB bandwidth and an exit one with 50 GHz 3 dB bandwidth.

Hence, we post-processed the collected high-resolution optical spectra. First, we filtered the spectral data by multiplying their PSDs with the TF of a 50 GHz

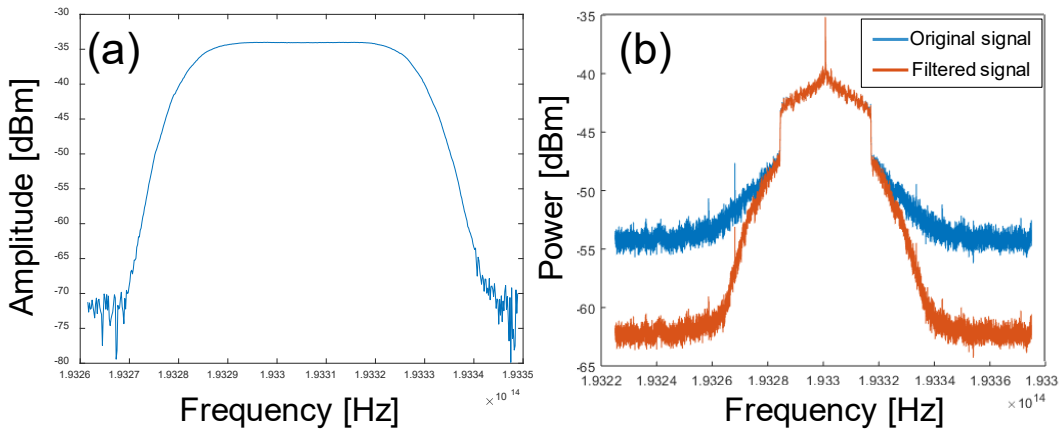

FIGURE 5.5. (a) Transfer function of the considered optical filter with 50 GHz 3 dB bandwidth. (b) Comparison of pre-filtered (blue plot) and filtered (orange plot) high-resolution optical spectra.

3 dB bandwidth optical filter, which we previously characterized. Such TF is shown in FIGURE 5.5(a). This operation was done to recreate a variety of possible realistic egress placement scenario conditions, such as filter central frequency shift/transmitter laser drift and filter bandwidth tightening. We emulated the first impairment by introducing a ±1 GHz misalignment between the filter TF and the signal central frequencies. Moreover, to emulate the bandwidth tightening, we reduced the filter 3 dB bandwidth to 46 GHz. We considered three bandwidth and central frequency shift variations for each spectrum. This operation caused a narrowing of the area in which the OSNR can be detected. In FIGURE 5.5(b), we show an example of experimentally collected optical spectra, in which the filtering effects are clearly visible. In particular, the ASE noise is easily measurable from the original pre-filtered signal optical spectrum, the blue plot in FIGURE 5.5(b). On the other hand, the ASE noise level cannot be identified from the filtered signal spectrum, which is the orange plot of FIGURE 5.5(b). This example again shows how the filters remove information at the channel sides, making the actual noise level identification often challenging. Once the filtering stage was completed, we cut the edges of the collected optical spectra to replicate a real DWDM spectral grid, where each channel is bounded by its adjacent (in this case, we considered 50 GHz slots). To do so, we only considered the spectra within the ±25 GHz interval centered around the channel central frequency identified through the carrier peaks.

Furthermore, to evaluate the proposed solution with spectral data at a lower spectral resolution, we processed the high-resolution collected optical spectra by averaging their samples in the linear domain. In this way, we obtained a reduced spectral resolution r_1, proportional to the new optical spectra length l_1, as shown in the example of FIGURE 5.6. In particular, the original BOSA-

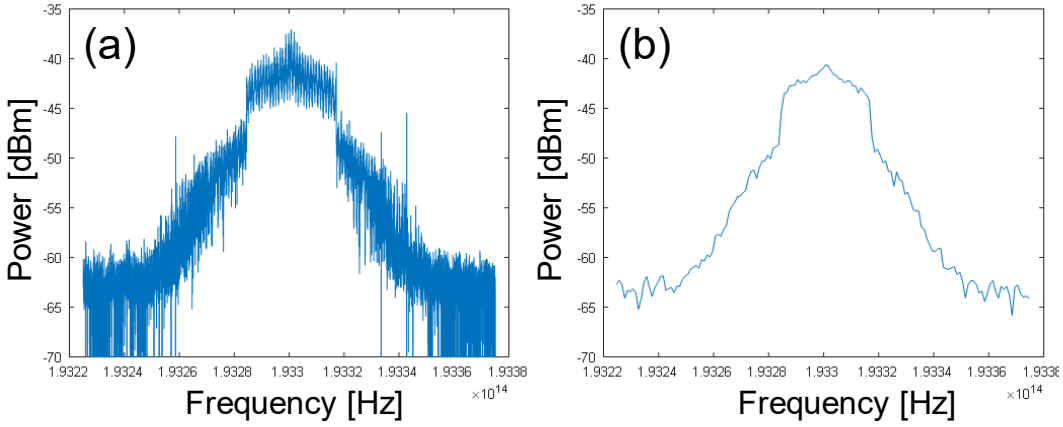

FIGURE 5.6. (a) Original BOSA-collected high-resolution (i.e., 12.5 MHz) optical spectrum. (b) Post-processed low-resolution (i.e., 1.25 GHz) optical spectrum.

collected data had a spectral resolution of 12.5 MHz (0.1 pm) and a length equal to 4000 samples, whereas the processed spectral data had a spectral resolution of 1.25 GHz (0.01 nm) and a length of 40 samples. Finally, we time-averaged the five different optical spectra we collected for each of the different case states of polarization. Time averaging is a typical method used to reduce the random noise affecting the measurements and is often applied directly within the OSA at the time of the spectral acquisition.

The sets of parameters for which we trained our ML model were those related to the high and low-resolution optical spectra, c_1 = (12.5 MHz, 50 GHz, 28 GBd) and c_2 = (1.25 GHz, 50 GHz, 28 GBd), respectively. We used 85% of our data set for the training phase and the remaining 15% for the testing. The total amount of spectra we fed to the algorithm was 198: 169 were used to train it, whereas the remaining 29 were for testing it. In addition, to better evaluate the estimation accuracy of the models, we also randomly shuffled training and

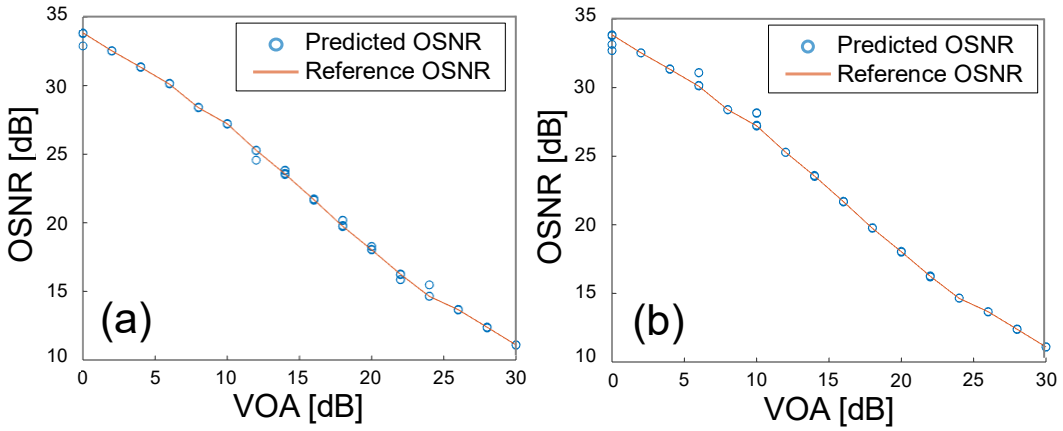

FIGURE 5.7. Reference and predicted OSNR values as function of the VOA levels for the 50 km distance scenario, with (a) high-resolution and (b) low-resolution optical spectral data.

testing data sets 200 times, trained a different Q_c ML model each time, and tested it with the corresponding testing set. The estimation accuracy we obtained with the GPR method was higher than the one we obtained with SVR. Therefore, for this experimental validation, we report and discuss only the GPR-related results. In particular, in FIGURE 5.7, we show the comparison between the reference and the predicted OSNR values as a function of the VOA levels in the 50 km distance scenario for the original BOSA-collected high-resolution optical spectra (a) and the processed low-resolution ones (b). We obtained FIGURE 5.7 results, training the GPR model with a spectral data set comprising the optical spectra of all the five considered distance scenario and then plotting only the 50 km-related spectra of the testing set. It can be noted from FIGURE 5.7 that sometimes, the predicted OSNR values of a specific VOA configuration (e.g., 12 dB for the high-resolution spectrum results on the left) are different. The different spectral sets used in the training phase of the ML-based algorithm caused these prediction variations. We obtained an MSE and a Maximum (MAX) error equal to 0.007 and 1.142 dB, respectively, considering the high-resolution spectral data. Instead, we obtained an MSE and MAX error equal to 0.012 and 0.954 dB concerning the low-resolution optical spectra, respectively. In FIGURE 5.8, we plot the obtained OSNR estimation error Probability Density Functions (PDFs) for the two considered spectral resolutions for all the five distance scenarios. We expected higher ASE noise estimation accuracies employing high- resolution spectral data, but the results we obtained disproved such an assumption. In fact, the considered ML-based algorithm maintained the same performance even when we reduced the amount of input data through the averaging operation. Moreover, we did not observe any dependence of the estimation error with respect to the reference OSNR value in the range between 10 and 30 dB of the experimentally acquired data.

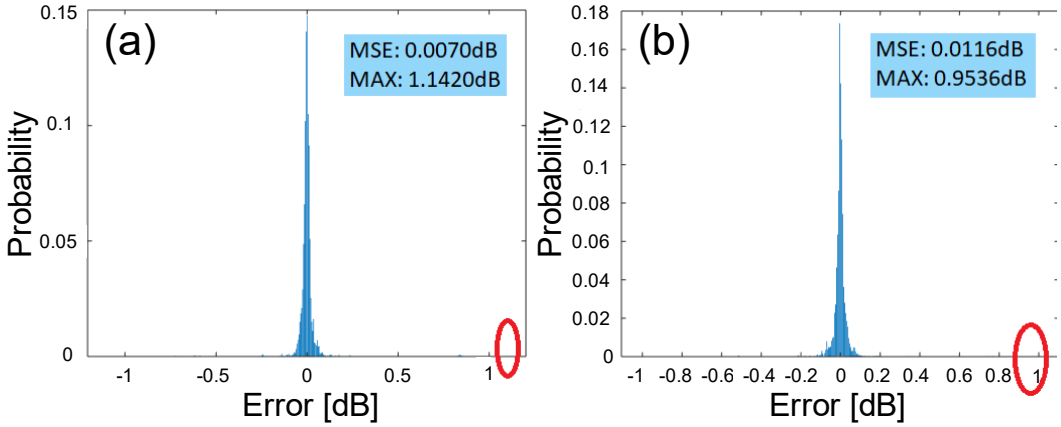

FIGURE 5.8. PDFs of the OSNR estimation errors obtained training the GPR model with all the five distance scenarios, for (a) high and (b) low-resolution spectral data. The red circles highlight the maximum errors (MAXs).

Simulation Setup

In addition to the experiment described in the previous paragraphs, in [19] we also implemented the VPI simulation setup depicted in FIGURE 5.9. The purpose of these further simulations was to verify if the post-processing we applied to the original high-resolution BOSA-collected optical spectra influenced in some ways the obtained results. We generated a 28 GBd PM-QPSK signal with roll-off factor equal to 1, yielding a 112 Gb/s connection. By placing a set OSNR module right after the transmitter, we adjusted the signal OSNR value increasing the related noise level. Then, in cascade, we placed a VOA and an EDFA, again considering sixteen different attenuation levels. In particular, we swept the VOA with 2 dB steps, within the interval between 0 and 30 dB, exactly as we did for the experimental case. Following the amplification stage, using an OSNR meter that employed the integral method described in the previous subsection, we measured the OSNR values used as references during the labeling phase. In addition, we also placed a tuneable optical filter with a second-order super-Gaussian TF and 37.5 GHz 3 dB bandwidth. In order to emulate the two most common optical impairments, that are, filter central frequency shift and the filter bandwidth tightening, we shifted the filter central frequency of ±1 GHz and reduced its bandwidth up to 25 GHz. Finally, we collected the optical spectra employing an OSA module with two spectral resolutions, 12.5 MHz (0.1 pm) and 1.25 GHz (0.01 nm). We acquired a total of 96 optical spectra for each considered spectral resolution. Using the same VPI simulation setup, we also generated a 224 Gb/s PM-16 Quadrature Amplitude Modulation (QAM) signal. We kept all the setup parameters as in the QPSK case and collected 128 optical spectra for each considered spectral resolution. Such acquisition process resulted in a total of four parameters sets c, one for each considered modulation format and spectral resolution. For all the four sets we only considered a B2B distance scenario. Again, we split the data set between training and testing sets maintaining the same ratio as before, that was, 85% of the data to train the algorithms and the remaining 15% to test it. In addition, we applied the same shuffling procedure of the experimental case. For the labeling phase, we

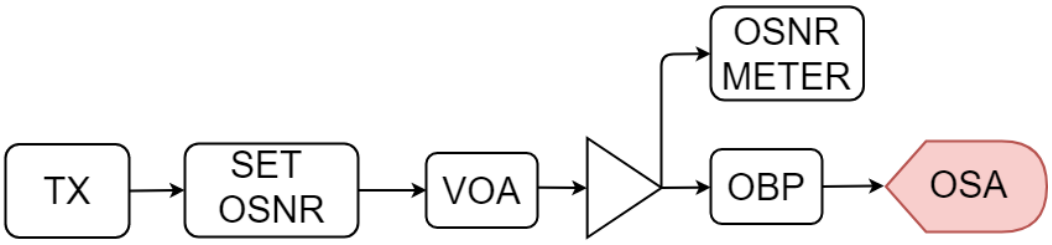

FIGURE 5.9. Diagram of the VPI implanted setup. TX: Transmitter; VOA: Variable Optical Attenuator; OBP: Optical Bandpass Filter; OSA: Optical Spectrum Analyzer.

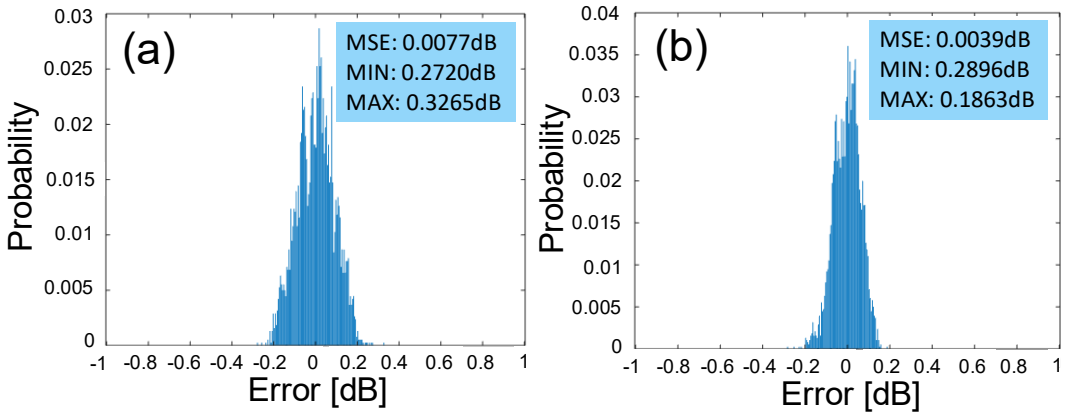

FIGURE 5.10. PDFs of the OSNR estimation errors obtained training the GPR model with the PM-16 QAM spectral data at (a) high and (b) low spectral resolution.

associated each collected optical spectrum with its related OSNR value measured by the OSNR meter. We obtained excellent estimation accuracies for both the considered ML algorithms (i.e., GPR and SVR), spectral resolutions (i.e., 12.5 MHz and 1.25 GHZ), and modulation formats (i.e., PM-QPSK and PM-16 QAM). In particular, concerning the OSNR estimation with the GPR model, in all the four considered cases we obtained a MSE and a MAX error lower than 0.009 and 0.33 dB, respectively. In FIGURE 5.10, we report the PDFs of the OSNR estimation errors for the PM-16 QAM format related to the two considered spectral resolutions. The results we obtained proved that our solution is able to identify the OSNR values in different scenarios, independently of the considered spectral resolution. In particular, if we compare the results obtained with the simulation-acquired spectral data to those obtained using experimental optical spectra, we notice lower error margins. In our opinion, such better estimation accuracy is mainly due to the more accurate and less realistic spectral shapes of the VPI-related spectra. In fact, more symmetric and smoother optical spectra surely help the ML algorithm to identify specific matches between the spectral shapes and the reference OSNR values.

Second Experimental Setup

The experimental setup we previously presented in this section did not include any optical filtering stage right before the BOSA. In fact, we collected the optical spectra after the amplification stage, and only during the post-processing we filtered them. To cope with such lack, in [128] we proposed a new experimental setup, also shown in FIGURE 5.11, which included a tuneable optical filter placed right before the OSA. In this way, we collected a new set of optical spectra, reproducing an egress placement scenario, and further tested our proposed ML-based solution under such conditions. In

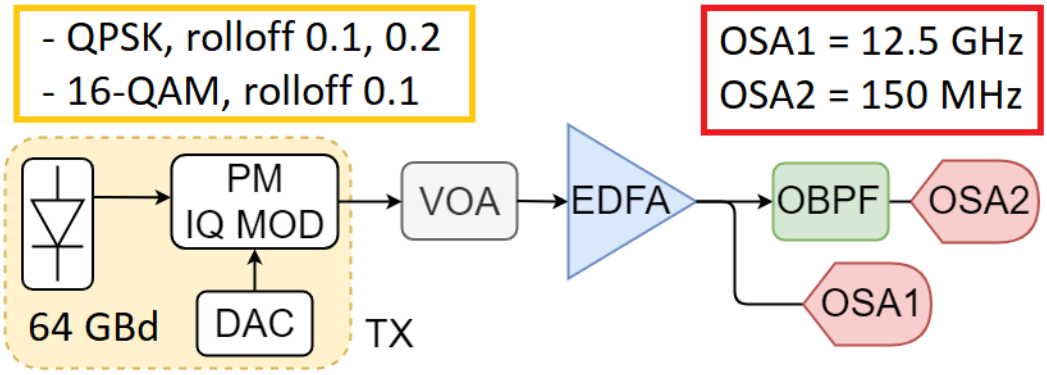

FIGURE 5.11. Schematic diagram of the second implemented experimental setup. TX: Transmitter; PM-IQ-MOD: Polarization Multiplexed-IQ-Modulator; DAC: Digital-to-Analog Converter; VOA: Variable Optical Attenuator; EDFA: Erbium-Doped Fiber Amplifier; OBPF: Optical Bandpass Filter; OSA: Optical Spectrum Analyzer.

particular, in [128] we only considered the GPR model, which in [19], between the ML methods we considered, was the one returning the best estimation accuracy. In this new experimental setup, leveraging a tuneable laser working at 1550.116 nm, we generated a 64 GBd PM-QPSK modulated signal with a roll-off factor equal to 0.1. After the transmitter, we cascaded a VOA and an EDFA operating in APC mode. With this configuration, we obtained nine different OSNR levels, which ranged between 13 dB and 28 dB. Right after the EDFA, we placed a first OSA with a spectral resolution of 12.5 GHz (0.1 nm). With such low-resolution spectral monitor, we acquired the signal ASE noise levels to be used as reference during the labeling phase. Then, we placed a 75 GHz 3 dB bandwidth optical bandpass filter, aligned with the transmitter laser frequency. Note that, in a real network scenario, for the considered 64 GBd signal the control plane would configure the filter bandwidth at 75 GHz. By shifting the filter central frequency and tuning its bandwidth, we reproduced the two classic optical impairments that we also considered in the previous experiment, filter central frequency shift and filter bandwidth tightening, respectively. In particular, we narrowed the filter 3 dB bandwidth up to 63 GHz, with steps of 3 GHz, and for each of the five obtained bandwidth values we also shifted the filter central frequency of ±1 and ±2 GHz. Finally, by means of a second OSA configured at a higher spectral resolution, that was 150 MHz (~1 pm), we acquired the optical spectra to be used as data set for the OSNR estimation. Since in this experimental setup we did not consider any fiber span, all the measurements we took referred to a B2B scenario. We acquired a total of 1125 optical spectra. In particular, we repeated each acquisition five times and finally averaged them during the post-processing phase. In addition, maintaining the whole setup unchanged, we also generated a 64 GBd PM-QPSK signal with 0.2 roll-off factor. For this second scenario we collected a total of 180 optical spectra, with filter 3 dB bandwidth set at 75 and

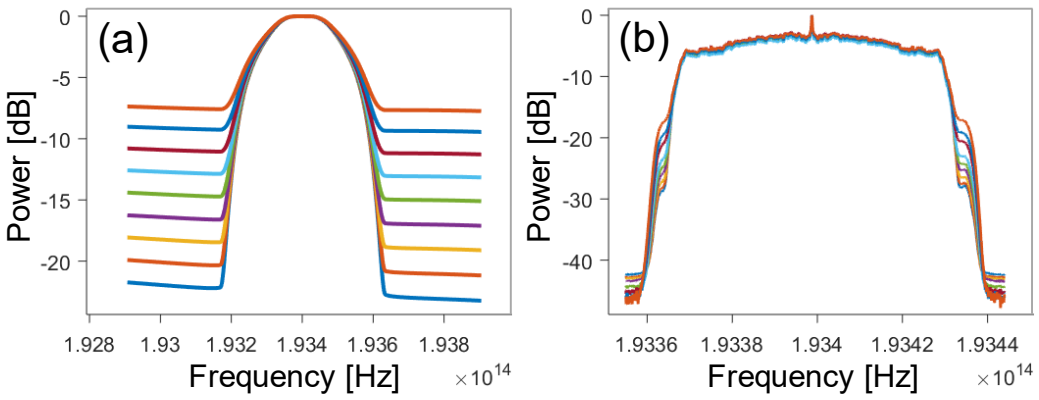

FIGURE 5.12. (a) The nine normalized low-resolution optical spectra collected with the OSA placed at the filter ingress port. (b) The nine normalized filtered optical spectra collected with OSA placed at the filter egress port and a 3 dB filter bandwidth equal to 75 GHz.

69 GHz and filter shift of ±2 GHz. Finally, we collected an additional 180 optical spectra, which referred to a PM-16 QAM format with 0.1 roll-off factor and the same filter configurations as per the 0.2 roll-off factor scenario.

As for the first experimental setup, we post-processed the acquired optical spectra to replicate some of the real network conditions that are difficult to reproduce in the lab. In particular, to emulate a WDM network scenario, we normalized all the collected optical spectra to align their peaks at 0 dBm. We show the results of such operation in FIGURE 5.12(a) and FIGURE 5.12(b). The former figure represents an example of nine normalized low-resolution optical spectra related to the nine different considered VOA levels, which we collected with the low-resolution OSA place at the ingress port of the filter. Instead, in FIGURE 5.12(b), we show the same nine normalized spectra, this time acquired at the egress filter port, with the high-resolution OSA. From FIGURE 5.12(b), we can observe a slight asymmetry in the optical spectra, which probably originated from the imperfect alignment between transmitter laser and filter central frequency. Such misalignment happened although we set both devices at the same frequency. These behaviors are quite typical in real networks, and further exacerbate when long paths including several cascaded filters are considered. Moreover, when collected with the egress OSA, the length l of the spectral data was equal to 50000 samples. Therefore, in order to reduce their size and speed up the ML processing phase, we cut the optical spectra maintaining only the portions at their sides, that were, the rising and falling parts. Such operation resulted in a reduction of the optical spectra length up to 800 samples.

Before feeding the spectral data to the GPR algorithm, we labeled them with their related OSNR values, which we measured using the spectra captured at the filter ingress port. As for the previous assessments, we used 85% of the

TABLE 5.1. Summary of the OSNR estimation results for the different trained ML models.

Mod. format (k)	Roll-off factor (a)	Filter 3 dB bandwidth (b)	MSE	MIN [dB]	MAX [dB]
PM-QPSK	0.1	63 GHz	0.427	-1.880	1.693
		66 GHz	0.187	-1.573	1.430
		69 GHz	0.115	-1.194	0.813
		72 GHz	0.055	-0.966	0.549
		75 GHz	0.029	-0.627	0.497
PM-QPSK	0.2	69 GHz	0.228	-1.167	1.207
		75 GHz	0.172	-0.992	0.868
PM-16QAM	0.1	69 GHz	0.139	-1.280	1.172
		75 GHz	0.030	-0.923	0.895

spectral data set to train the algorithm and the remaining 15% to test it. Furthermore, we randomly shuffled the training and testing set 200 times, each time training a different model and testing it with its corresponding test set. However, unlike the previous assessments, this time the parameters set that we considered also included the roll-off factor a, since the spectral data we acquired referred to two different roll-off factor. In addition, having spectral data related to two different modulation formats k, with the same symbol rate q, the set of parameters c for which we trained the GPR model was $c = (r, q, k, a)$. In particular, we trained three different models, according to the three different data set we collected, that were, c_1 = (150 MHz, 64 GBd, PM-QPSK, 0.1), c_2 = (150 MHz, 64 GBd, PM-QPSK, 0.2), and c_3 = (150 MHz, 64 GBd, PM-16 QAM, 0.1). The results we obtained for these three models showed a MAX estimation error around 2.5 dB. We considered such performance degradation to be the result of a lack of a priori knowledge of the filter bandwidth values, which we believed to be a key parameter for the ASE noise estimation. Therefore, we retrained our models also taking into account the related 3 dB bandwidth values. Hence, our set of parameters became $c = (r, b, q, k, a)$, with b representing the considered filter 3 dB bandwidth. We summarize the results obtained with this new configuration, for the nine considered models, in TABLE 5.1. In particular, there, we report the MSE, the Minimum (MIN), and the MAX errors for each of the nine models. From such results, we observe how the Maximum Absolute Error (MAE) in all the considered cases, never exceeded 1.9 dB. As expected, the estimation errors for wider filter bandwidth values were lower, since in those cases the ASE noise

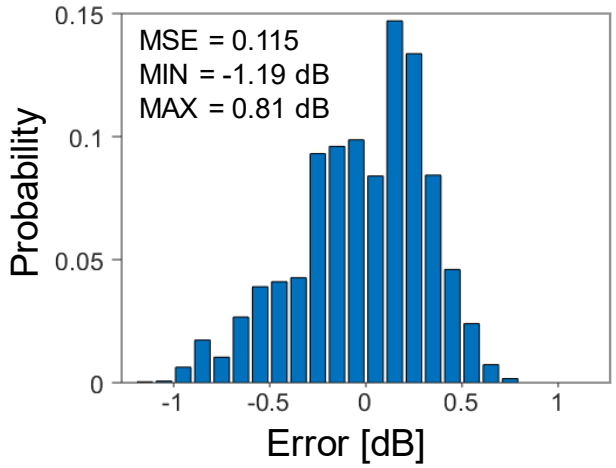

FIGURE 5.13. PDF of the OSNR estimation error for the 69 GHz 3 dB filter bandwidth case with PM-QPSK modulation format and roll-off factor equal to 0.1.

levels are less affected by the filtering effects, and traditional methods can also be employed. On the other hand, our ML-based technique could still retrieve the ASE noise with acceptable accuracy when narrower bandwidth values were considered, whereas traditional methods, such as the interpolation one, would struggle [66]. More in detail, considering only filter 3 dB bandwidth values equal or larger than 69 GHz, the obtained MAE reduced to 1.3 dB. In FIGURE 5.13, we report the PDF of the OSNR estimation error related to the 69 GHz 3 dB bandwidth case of the PM-QPSK modulation format with 0.1 roll-off factor. In the figure inset we also report MSE, MIN, and MAX errors values.

5.3 Filter-Related Parameters Estimation in an Ingress Monitoring Placement Scenario

In Section 5.1, we introduced the ingress optical monitor placement scenario, in which the OCMs are placed before the ingress WSS of each network ROADM node, as we showed in FIGURE 5.2 through the blue boxes. From spectral data collected within such a placement scenario, we can easily retrieve information related to the signal, such as the ASE noise accumulated in the adjacent link. On the other hand, when estimating the filter-related parameters, we cannot directly measure any such information from the optical spectra. This occurs because we are monitoring the signals before they undergo any filtering process within the ROADM node. Therefore, advanced processing techniques are required. For this purpose, in [126], we presented a solution to extract the filter-related features, such as the filter central frequency and the filter bandwidth, from optical spectral data acquired from OSMs placed before the ingress ROADM ports. There, we also validated this method leveraging

spectral data collected through VPI simulations and an experimental setup. In addition, in [127], we further enhanced this solution proposing an advanced way to identify the noise and again validating it with simulations and an experiment. In the following paragraphs, we will formally present our filter features extraction technique, relying on the formulation we initially presented in [126].

In FIGURE 5.2, we depicted an optical network with OCMs related to both the placement scenarios. As we saw in Section 5.1, a configuration where both the ingress and egress optical spectral monitors are present would be unfeasible because of its cost. We also know that to estimate any features of a ROADM node filter, as a first step, we should retrieve the spectrum of its TF. Thus, hypothetically, within such ideal scenario, to retrieve the filter TF of a node we would simply divide in the linear domain the PSD captured by the OCM at the node egress port, by that captured at the ingress one. Relying on the formulation presented in FIGURE 5.2 and assuming that no ASE noise is added when crossing the ROADM nodes, we can express the TF of a generic node n in the ideal placement scenario, as

$$\mathrm{TF}_{n,ideal} = \frac{\mathrm{PSD}_{OCM_{n,e}}}{\mathrm{PSD}_{OCM_{n,i}}}, \tag{5.1}$$

where $\mathrm{PSD}_{OCM_{n,e}}$ and $\mathrm{PSD}_{OCM_{n,i}}$ represent the PSDs acquired at the egress and at the ingress ports of node n, respectively. However, since the ingress monitoring placement scenario does not envision OCMs at the egress ports, we have to find another way to obtain the egress-related PSD. In order to cope with this lack, in [126], we proposed to replace it with the PSD monitored at the ingress port of the following node. The result of such operation represents not only node n filter TF, but also the noise introduced by the link connecting n and n+1, which in FIGURE 5.2, we named as Link $(n, n+1)$. Therefore, according to FIGURE 5.2 formulation, we can restate Equation 5.1 expressing the TF of a generic node n in the ingress placement scenario, as

$$\mathrm{TF}_{n,ingress} = \frac{\mathrm{PSD}_{OCM_{n+1,i}} - \mathrm{ASE}_{n,n+1}}{\mathrm{PSD}_{OCM_{n,i}}}, \tag{5.2}$$

where $\mathrm{PSD}_{OCM_{n+1,i}}$ represents the PSD acquired at the ingress port of the node $n+1$, which is the one following node n, and $\mathrm{ASE}_{n,n+1}$ represents the ASE noise accumulated within Link $(n, n+1)$. The operation described by Equation 5.2 removes the effects accumulated by the connection over its path up to node n OCM, leaving as the only noise source Link $(n, n+1)$. Therefore, to reduce such unwanted contribution, we estimated from the optical spectra acquired at node $n+1$ ingress port, the ASE noise accumulated within Link $(n, n+1)$,

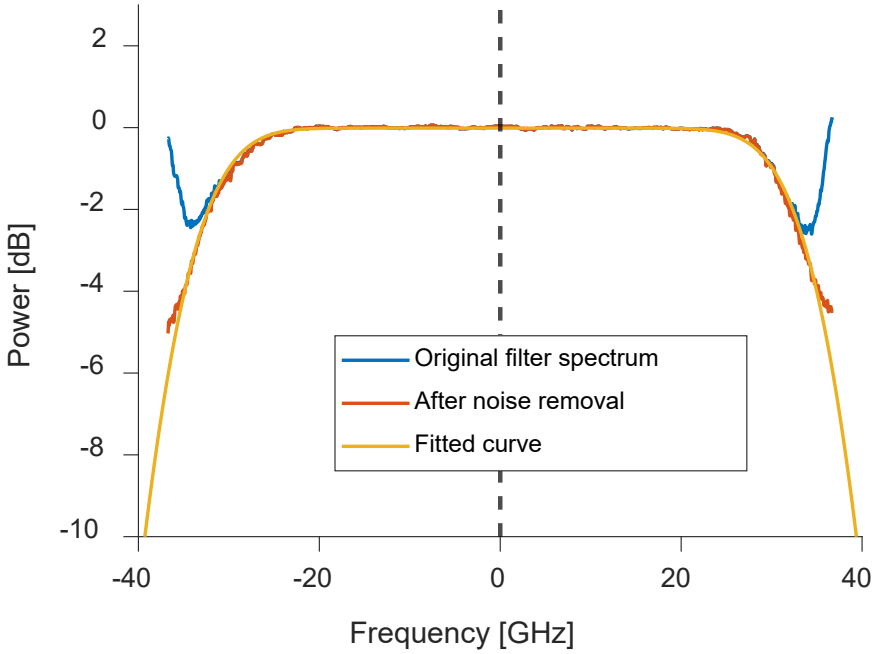

FIGURE 5.14. Example of an optical filter TF spectrum reconstruction. Original noisy portion of the filter TF spectrum (blue) and after the noise removal (orange). The yellow curve represents the filter TF spectrum after the fitting process. The fitting parameter values for this example were: α=5.4, β=73.51, γ=-18.55, and δ=-0.04.

and removed it, in the linear domain, from those same spectra acquired by $\mathrm{OCM}_{n+1,i}$.

It is worth to note that, as we discussed in Section 3.2.3, depending on the considered ROADM architecture, the number of WSSs contained in the ROADM node can vary between one (B&S) and two (R&S) per degree. We treated both cases the same: if the node contained two filters, we modeled an equivalent filter for the specific ingress-egress direction. By doing so, Equation 5.2 would express the TF of the single filter, in case of a B&S configuration, or the TF equivalent to both the filter involved in the specific ingress-egress direction, for the R&S case.

We show the spectra resulting from the operations described by Equation 5.2 in FIGURE 5.14. There, we plot in blue the noisy filter TF spectrum and in orange the TF spectrum after the noise removal process, as described by Equation 5.2. Observing FIGURE 5.14, it is clear how, despite the above described operation, the filter TF spectrum sides still do not reach deep values, making the 6 dB bandwidth identification impossible. We observed this behavior in most of the cases we considered. It occurred, because at the spectra edges we processed the noise instead of the signal contained in the channel. Hence, in order to overcome this obstacle and be able to evaluate the filter bandwidth, we have to reconstruct the full filter TF. To do so, we proposed to

fit the available TF spectral portion with a function that corresponds to the ideal filter shape. Typically, optical filters are considered to have a high-order super-Gaussian shape. In [132], the authors, leveraging the error function $\mathrm{erf}(x)$, proposed a model to characterize the optical field spectrum $S(f)$ of the bandpass filter created by a WSS. The function modelling an ideal symmetric filter centered at 0 Hz frequency, is

$$S(f) = \frac{1}{2}\alpha\sqrt{2\pi}\left[\mathrm{erf}\left(\frac{\beta/2 - f}{\sqrt{2}\alpha}\right) - \mathrm{erf}\left(\frac{-\beta/2 - f}{\sqrt{2}\alpha}\right)\right], \tag{5.3}$$

where α is the parameter related to the steepness of the filter edges, β represents the 6 dB bandwidth of the filter, and f represents the frequency.

Because of filter/signal misalignments or power leveling issues, often the portions of filter TF spectra that we wanted to fit were not centered. Thus, to also take into account the shifts in both axis directions, we extended the model described by Equation 5.3 including two new parameters, such as

$$S(f) = \frac{1}{2}\alpha\sqrt{2\pi}\left[\mathrm{erf}\left(\frac{\beta/2 - f - \delta}{\sqrt{2}\alpha}\right) - \mathrm{erf}\left(\frac{-\beta/2 - f - \delta}{\sqrt{2}\alpha}\right)\right] + \gamma\,, \tag{5.4}$$

where γ is a normalization factor for the y-axis shifts and δ is the parameter related to the filter central frequency shift.

Finally, by tuning the parameters α, β, γ, and δ, we were able to fit the filter TF spectral portions with the logarithmic squared version of Equation 5.4, as also shown in FIGURE 5.14. In addition, in order to better fit the filter TF, we can also narrow down the range in which the parameters are tuned, simply relying on the filter specifications. Once the fitting process completed, the values assumed by β and δ represented the estimated filter 6 dB bandwidth and the estimated filter central frequency shift, respectively. We focused our analysis on the two filter parameters that we considered as the most representative, but we retrieved the full filter TF shape, therefore making any other filter feature of interest available.

5.3.1 Filter-Related Parameters Estimation Method Assessment

Simulation Setup

In [126], we carried out several simulations and implemented an experimental setup to validate the proposed solution. In particular, concerning the simulations, we implemented in VPI the setup depicted in FIGURE 5.15. We generated a 32 GBd PM-QPSK modulated signal, with a 0.1 roll-off factor, centered at 193.4 THz (1550.116 nm). In order to simulate the optical links, after the transmitter we cascaded a number of spans which included 80 km

length Standard Single-Mode Fibers (SSMFs) and EDFAs with 5 dB Noise Figure (NF). We set the output power of the transmitter to 0 dBm as well as the output power of all the EDFAs. The optimal launch power varies depending on several factors, such as the amount of ASE noise, the fiber attenuation, the amplifier NF, and the NLI. In particular, since we only considered a single-channel configuration and the only marginal NLI contribution was that related to Self-Channel Interference (SCI), we set the launch power to 0 dBm. With no other channels that could contribute to the XCI, the reason behind the fiber span presence in the setup was to verify the role of fiber dispersion and SCI on the performance of the proposed solution. By varying the number of considered fiber spans and EDFAs, we simulated different ASE noise levels, and according, different OSNR values. Each link was followed by an optical node, which we implemented as a cascade of two optical filter with second-order super-Gaussian TF. We considered three different 6 dB filter bandwidth values: 36.5, 37.5, and 38.5 GHz. Moreover, to emulate the laser drift/filter central frequency shift impairment, we shifted the central frequency of each filter of a value ranging in between -1 and +2 GHz. To simplify the considered cases, we assumed both the filters of each node to have the same characteristics. Finally, at each node input and output ports, we placed optical spectral monitors configured with 1 GHz spectral resolution. Although we focused on the ingress monitor placement scenario, in order to retrieve the ideal filter TF to be used as a reference within the estimation accuracy assessment phase, we also assumed egress placed optical monitors. In total, our setup included three nodes, four optical links, and seven OCMs.

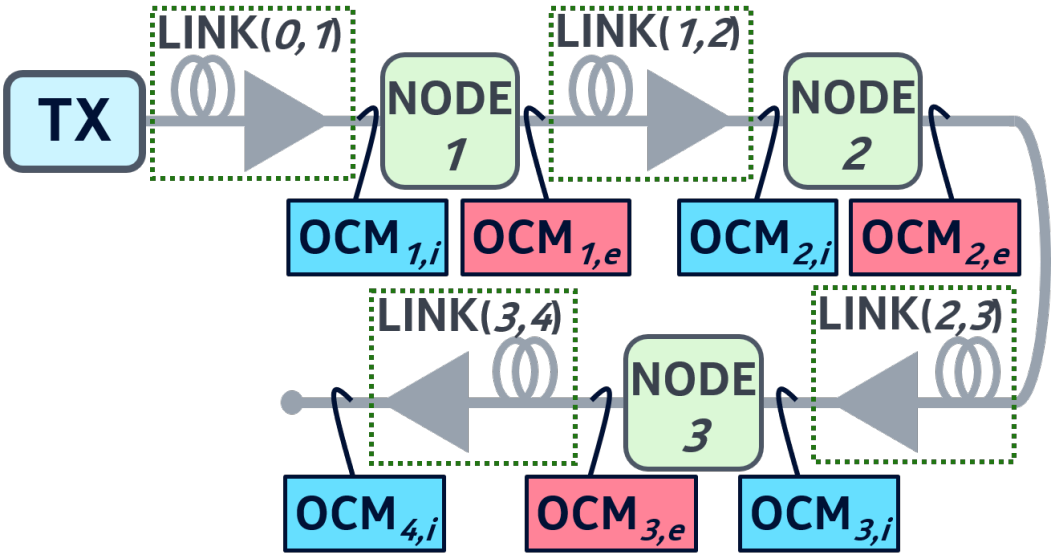

FIGURE 5.15. Schematic diagram of the VPI-implemented simulation setup. TX: Transmitter; OCM: Optical Channel Monitor.

TABLE 5.2. The sixteen considered configurations for the simulation setup.

Links (# of spans)				Filters (6 dB bandwidth [GHz] + central freq. shift [GHz])		
0-1	1-2	2-3	3-4	1	2	3
5	2	2	3	37.5 (+0, +1, +2, +2)	37.5 (+2, +2, +1, -1)	37.5 (+1, +2, +0, +2)
2	5	2	3	37.5 (+0, +1, +2, +2)	37.5 (+2, +2, +1, -1)	37.5 (+1, +2, +0, +2)
2	5	2	3	36.5 (+0, +1, +2, +2)	36.5 (+2, +2, +1, -1)	36.5 (+1, +2, +0, +2)
2	5	2	3	38.5 (+0, +1, +2, +2)	38.5 (+2, +2, +1, -1)	38.5 (+1, +2, +0, +2)

In TABLE 5.2, we summarized the sixteen configurations that we considered, listing the number of fiber spans and EDFAs per link, and each filter 6 dB bandwidth and central frequency shift. Through such 16 configurations and their related random filter central frequency shifts, we tried to cover as many combinations as possible.

To assess the proposed filter parameter extraction solution with the spectral data related to the sixteen considered cases, we retrieved the noisy filter TFs corresponding to each setup node. Then, we identified the ASE noise amounts of each case by averaging the residual noise values at the signal sides. In fact, we collected the considered optical spectra before the signals entered the optical nodes, and therefore before the signals and their related ASE noise amounts were filtered. According to the steps presented in Section 5.3, after identifying the noise values, we removed them from the spectra and then fitted the resulting spectral portions with the logarithmic squared version of Equation 5.4. We estimated the filter 6 dB bandwidth and central frequency shift values by observing those assumed by Equation 5.4 parameters β and δ, respectively. Finally, we compared such values with those retrieved employing the OCM placed at the egress port of the nodes and calculated the estimation errors. In TABLE 5.3, we list the MSE, σ, and the MAE for the filter 6 dB bandwidth and central frequency shift estimations for the three nodes we considered. Theoretically, dividing in the linear domain the PSDs captured at two adjacent OCMs (i.e., at two adjacent nodes) should nullify the effect of the path up to the first of the two considered OCMs. For instance, in FIGURE 5.15 setup, dividing the PSD acquired at $\mathrm{OCM}_{3,i}$ by that acquired at $\mathrm{OCM}_{2,i}$ should nullify the effect of the path up to $\mathrm{OCM}_{2,i}$ location. Nevertheless, the estimation errors we summarize in TABLE 5.3 indicate a small residual effect of the cascade. In fact, the 6 dB bandwidth estimation error related to Node *3* is slightly higher than those corresponding to Node *1* and Node *2*. In general, at

TABLE 5.3. Estimation errors for the filter-related features, considering the simulation-acquired optical spectra.

Node	Estimated Feature	MSE	σ [GHz]	MAE [GHZ]
1	Centr. freq. shift	0.0019	0.0334	0.0655
	6 dB BW	0.0183	0.0807	0.1937
2	Centr. freq. shift	0.0008	0.0178	0.0454
	6 dB BW	0.0024	0.0479	0.1057
3	Centr. freq. shift	0.0026	0.0482	0.0997
	6 dB BW	0.0163	0.1247	0.2962

the first two nodes, the spectra are almost not affected by the cascade, and the estimation errors are also very low. Hence, keeping such accuracy as the cascade increases, for instance, at Node *3*, is quite challenging.

Experimental Setup

To further validate the proposed solution, we also implemented the experimental setup shown in FIGURE 5.16. By means of a tuneable laser working at 193.4 THz (1550.116 nm), we generated a 64 GBd PM-QPSK modulated signal with a roll-off factor equal to 0.1. We set the transmitter output power to -11 dBm. In addition, in order to simulate the optical impairments that could occur in a real network scenario, right after the transmitter we placed an optical filter and varied the values of its bandwidth and central frequency. Following this first filter, with the aim of emulating different fiber span lengths and therefore different ASE noise contributions, we cascaded a VOA and an EDFA operating in APC mode with an output power of 0 dBm and a NF of 5 dB. In FIGURE 5.16, we referred to these two components as Link (1, 2). As per the simulation case, we chose such output power considering the single channel configuration that we planned. After

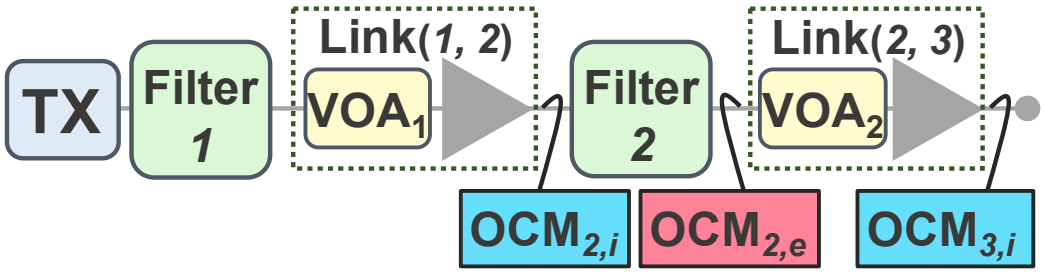

FIGURE 5.16. Schematic diagram of the considered experimental setup. TX: Transmitter; VOA: Variable Optical Attenuator; OCM: Optical Channel Monitor.

Link (1, 2), we placed a second optical filter, namely Filter 2, that was the device on which we actually tested our proposed technique. We implemented Filter 2 using a Finisar WaveShaper 4000S, a device able to emulate the flex-grid capability of a WSS [133]. We set Filter 2 to have a high-order super-Gaussian TF (i.e., "rectangular" in the Finisar software) and varied its 6 dB bandwidth and central frequency to obtain different optical impairment combinations. After Filter 2, we placed a second link, which we called Link (2, 3), cascading a VOA and an EDFA operating in APC mode, and varying the attenuation value. Finally, we tapped the ingress and the egress ports of Filter 2, and the output of Link (2, 3), with an OCM. In particular, we referred to the OCM placed at Filter 2 ingress and egress ports, with $\mathrm{OCM}_{2,i}$ and $\mathrm{OCM}_{2,e}$, respectively, whereas we named the last OCM placed after the second link, $\mathrm{OCM}_{3,i}$, as shown in FIGURE 5.16. The optical monitor we used was the Finisar WaveAnalyzer 1550S, a high-resolution coherent OSA able to reach spectral resolutions down to 150 MHz [134]. However, to emulate the performance of an OCM, which generally operates at 1 GHz spectral resolution, we collected all the spectral data with such resolution.

In total our experimental setup included two links, two optical nodes represented by two optical filters, and three monitoring locations. By varying the attenuation values introduced by the two VOAs and the characteristics of the two filters, that were, their central frequencies and their 6 dB bandwidths, we clustered the acquired spectral data in three cases. Each case corresponded to a different data set and consisted of nine subcases. We resume such three considered combinations in TABLE 5.4.

As per the simulation scenarios, the first step to implement our filter extraction method was to calculate the filter TFs. In particular, leveraging $\mathrm{OCM}_{2,e}$, we retrieved the ideal Filter *2* TFs to be used as references for the estimation accuracy evaluations. On the other hand, we retrieved the noisy Filter *2* TFs

TABLE 5.4. The three considered cases for the experimental setup.

Case	Filter *1* 6 dB BW (shift) [GHz]	Link (*1, 2*) VOA_1 [dB]	Filter *2* 6 dB BW (shift) [GHz]	Link (*2, 3*) VOA_2 [dB]
1	74 (-2)	0	73, 75, 77 (-1, 0, +1)	10
2	69 (-2)	10	73, 75, 77 (-1, 0, +1)	10
3	74 (-2)	0	73, 75, 77 (-1, 0, +1)	20

TABLE 5.5. Estimation errors for the Filter 2-related features, considering the experimental-acquired optical spectra.

Case	Estimated Feature	MSE	σ [GHz]	MAE [GHz]
1	Centr. freq. shift	0.0228	0.1341	0.2433
	6 dB BW	0.0204	0.1433	0.3933
2	Centr. freq. shift	0.0597	0.2256	0.5054
	6 dB BW	0.1978	0.4508	0.6811
3	Centr. freq. shift	0.0133	0.1018	0.1863
	6 dB BW	0.1383	0.3470	0.7627

considering the optical spectra acquired within the two ingress scenario-related monitors. We estimated and removed the noise amounts from the identified TFs through the same procedure we described for the simulation setup. Then, we fitted the obtained spectral portions with the logarithmic squared version of the Equation 5.4 function. Finally, to obtain the new estimation accuracies, we compared the values assumed by the parameters β and δ with those assumed by the same two parameters within the corresponding reference scenarios. We summarize the results of such comparison in TABLE 5.5, where we report the MSE, σ, and the MAE, for the filter 6 dB bandwidth and central frequency shift estimations related to Filter *2*, in the three considered cases.

From the obtained results, we can notice how the accuracy related to the 6 dB bandwidth estimation tended to worsen as the filter bandwidth narrowed or the attenuation increased. In particular, case 2 emulated bigger FCE and higher attenuation (i.e., longer spans) in Filter *1* and Link $(1, 2)$, respectively, whereas case 3 emulated higher attenuation in Link $(2, 3)$.

5.4 Estimation Methods Joint Assessment and Comparison

Experimental Setup

In order to compare the accuracies of the ASE noise/OSNR and the filter-related parameters estimation solutions reported in Sections 5.2 and 5.3, respectively, in [127] we further assessed them under a unified experimental setup. In particular, we considered the same experimental setup described in Section 5.3.1 and depicted in FIGURE 5.16, with just the addition of an OCM

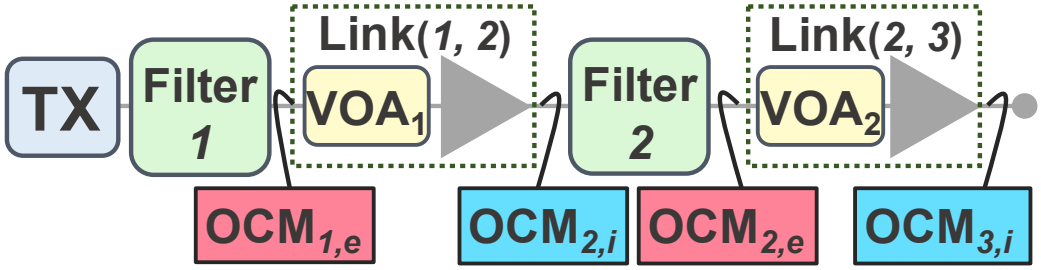

FIGURE 5.17. Schematic diagram of the unified experimental setup. TX: Transmitter; VOA; Variable Optical Attenuator; OCM: Optical Channel Monitor.

placed at the egress port of Filter *1*, as shown in FIGURE 5.17. Moreover, we also proposed two ways to enhance the previously introduced estimation solutions, improving their accuracies. Thus, in this section, we will first describe such enhancements, report the joint assessment-related results, and finally compare the two approaches.

Concerning the ASE noise/OSNR estimation method within an egress optical monitor placement scenario, we proposed to exploit the spectral knowledge at previous nodes, in order to neglect the effect of the network portion up to the considered OCM. We did this, in a similar manner to what we already implemented within the ingress placement scenario. Referring to the notation introduced in FIGURE 5.2, for the OSNR estimation, instead of the optical spectra captured at node n egress port, i.e., using $\mathrm{OCM}_{n,e}$, we used the PSDs resulting from the division, in the linear domain, of the $\mathrm{OCM}_{n,e}$ captured spectra by those acquired with $\mathrm{OCM}_{n-1,e}$, such as

$$\mathbf{s} = \frac{\mathrm{PSD}_{OCM_{n,e}}}{\mathrm{PSD}_{OCM_{n-1,e}}}. \tag{5.5}$$

Leveraging also node $n-1$ egress OCM allowed us to neglect the effect of the link/node cascade up to the location of node $n-1$, and to focus only on what happened between the nodes $n-1$ and n. Assuming that no frequency drift affects the employed OCMs, the only requirement for such operation is the collection of both spectra at a single point, which would typically be the central controller. In this way, spectra resolution variations from different OCMs could be easily accounted with some simple spectral processing, e.g., upsampling the lowest resolution. Performing the above described divisions along an optical connection path, would allow to see the effect of each link/node on that path, nullifying those of the previous network elements along that same lightpath.

On the other hand, to enhance the filter-related parameters estimation solution to be applied in ingress monitoring placement scenarios, we proposed

an improved way to identify the noise to be removed before fitting the TF spectra [127]. In fact, the approach we presented in Section 5.3, envisioned the measurement of such values from the sides of the spectra collected by the ingress OCMs. However, because of filtering and/or the OCM response, sometimes such estimations might not be precise. Hence, we implemented a function that searched within a set of noise values that we previously defined, based on the monitored spectra or from a basic knowledge of the link/span configurations and an accuracy correction factor. The function selected the noise amount that resulted in the lowest fitting error. In fact, removing a lower or higher amount of noise would return a wrong filter shape, yielding to a worse match with respect to the case where a correct noise amount is subtracted. Referring to FIGURE 5.2, the assumption behind such operation was that the filter at node n suppressed the noise introduced by the amplifiers/links cascade up to that point, and therefore the noise at the sides of the $\mathrm{OCM}_{n+1,i}$ collected PSDs is mainly caused by Link $(n, n+1)$ contributions. This noise identification process allowed us to better estimate the ASE noise with respect to the previous approach we presented (see Section 5.3), and thus to better estimate the final filter TF and its related parameters.

As we previously mentioned, the setup we considered for the joint assessment of the two estimation methods was the same to that one we presented in Section 5.3.1, with the addition of a further OCM at Filter *1* egress port, namely $\mathrm{OCM}_{1,e}$. We represent such experimental setup in FIGURE 5.17. In addition, we generated the 64 GBd PM-QPSK modulated signal with two different roll-off factor values, that were, 0.1 and 0.2. Furthermore, in order to verify any dependence of our proposed estimation methods from the OSA spectral resolution, we acquired the spectra data at the four different monitoring locations, with 600 MHz and 1 GHz spectral resolutions. For each of the two roll-off values and resolutions, we collected a total of 252 optical spectra, 63 for each of the four optical monitoring locations. TABLE 5.6 shows how we clustered the collected spectra into seven different cases, according to the attenuation values assumed by the VOAs and the parameters of the filters. In turn, each of the seven cases consisted of nine subcases. In particular, the 6 dB bandwidth and central frequency shift of Filter *2* always assumed the same values in every cluster, whereas the attenuation of the two links and the parameters related to Filter *1* varied in each case. It is worth noting that cases 1, 2, and 4 of TABLE 5.6 correspond to the three cases which we already described in Section 5.3.1 and reported in TABLE 5.4. This time, we assumed case 1 to be the "default" situation, in which we set the 6 dB bandwidth of Filter *1* at 74 GHz, did not consider any additional attenuation for Link $(1, 2)$, and considered 10 dB of attenuation for Link $(2, 3)$. All the other remaining six cases represented a worsening of case 1, where at least one of the three varying

TABLE 5.6. The seven considered experimental cases for the joint assessment of the two proposed methods.

Case	Filter *1* 6 dB BW (shift) [GHz]	Link (*1*, *2*) VOA_1 [dB]	Filter *2* 6 dB BW (shift) [GHz]	Link (*2*, *3*) VOA_2 [dB]
1	74 (-2)	0	73, 75, 77 (-1, 0, +1)	10
2	74 (-2)	0	73, 75, 77 (-1, 0, +1)	20
3	74 (-2)	10	73, 75, 77 (-1, 0, +1)	10
4	69 (-2)	10	73, 75, 77 (-1, 0, +1)	10
5	74 (-2)	5	73, 75, 77 (-1, 0, +1)	10
6	74 (-2)	2.5	73, 75, 77 (-1, 0, +1)	15
7	69 (-2)	7.5	73, 75, 77 (-1, 0, +1)	20

parameters (Filter *1* bandwidth and the two VOAs) assumed a value worse than in the default case.

5.4.1 OSNR Estimation Method Joint Experiment Results

To validate the improved OSNR estimation method within the unified experimental setup shown in FIGURE 5.17, we leveraged the optical spectra collected through the egress placed OCMs, namely, $OCM_{1,e}$ and $OCM_{2,e}$, with a spectral resolution of 600 MHz. In addition, we only considered the spectra related to the 0.1 roll-off factor scenario, and to the cases where Filter *1* 6 dB bandwidth was equal to 74 GHz, that were, cases 1, 2, 3, 5, and 6 of TABLE 5.6. In order to implement the enhancement presented in Section 5.4, we divided the PSDs captured at $OCM_{2,e}$ by those captured at $OCM_{1,e}$. Then, we labeled the resulting 45 optical spectra, with their corresponding reference ASE noise values, that we retrieved from the spectra captured with the Filter *2* ingress node OCM, namely $OCM_{2,i}$. Formulating the estimation as a regression problem enabled us to employ SVR as ML algorithm for the ASE noise prediction. In particular, we observed better estimation accuracies by training

the SVR model with a linear kernel function than with a Gaussian one. We used 80% of the 45 total spectra to train the model, 10% to cross-validate it, and the remaining 10% to test it. Moreover, we used the cross-validation to tune the parameter ϵ of the SVR model. As we saw in Section 2.1.3, in SVM and SVR, ϵ represents half the width of the insensitive band, which is that tolerance area where no penalty is assigned to the errors. In addition, to precisely assess the model estimation accuracy, we also randomly shuffled the training and the testing data sets 4000 times, trained a different model each time, and tested it with its corresponding testing set. Employing such configuration, we achieved a MSE equal to 0.0136 and a MAE lower than 0.4 dB. It is important to stress, that in the experimental setup depicted in FIGURE 5.17, we considered the filter 6 dB bandwidth, whereas in Section 5.2.1, we considered the 3 dB bandwidth of the filters. This resulted in filtering effects which affected way more severely the ASE noise with respect to the past, because of the narrower 3 dB bandwidths we assumed. Nevertheless, comparing the current results with a similar case we reported in TABLE 5.1 of Section 5.2.1 (i.e., 64 GBd PM-QPSK signal with 0.1 roll-off and 72 GHz 3 dB bandwidth), we can notice how the accuracy improved. In fact, the MAE corresponding to the past case was almost equal to 1 dB, whereas the current one is lower than 0.4 dB. Another aspect worth to be observed, is that the results of Section 5.2.1 referred to optical spectra collected with a spectral resolution of 150 MHz, while the current ones were collected with a spectral resolution of 600 MHz. This further reinforces the observation we previously made about the independence of our proposed OSNR/ASE noise estimation method from the spectral resolution of the employed optical spectra.

5.4.2 Filter Features Estimation Method Joint Experiment Results

As we previously showed in Section 5.3.1, the first step to identify bandwidth and central frequency of any optical filter is the identification of its TF. To do so, we considered the optical spectra acquired with the ingress-placed OCMs, which in the unified experimental setup depicted in FIGURE 5.17, we named as $\mathrm{OCM}_{2,i}$ and $\mathrm{OCM}_{3,i}$. In addition, we also employed $\mathrm{OCM}_{2,e}$, in order to retrieve the ideal Filter *2* TFs to be used as reference for the accuracy evaluation of our method. Once we obtained the noisy TFs, from the optical spectra acquired with $\mathrm{OCM}_{3,i}$ we estimated the noise amounts to be removed, according to the enhancement steps reported in Section 5.4. Finally, by fitting the resulting TF spectral portions, we estimated the 6 dB bandwidth and central frequency of each considered filter and compared them with the previously identified reference values. The estimation errors we obtained through such process, for the filter central frequency and 6 dB bandwidth, are graphically shown in FIGURE 5.18 and FIGURE 5.19 for roll-off factor of 0.1 and in FIGURE 5.20

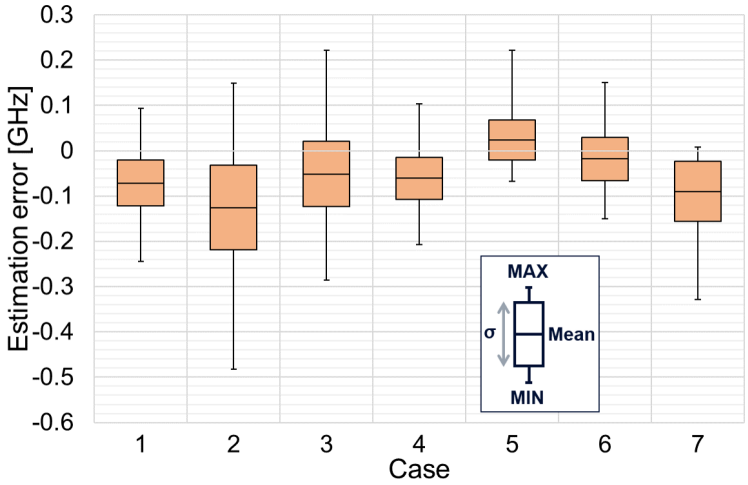

FIGURE 5.18. Filter central frequency shift estimation errors for the seven experimental cases with roll-off factor = 0.1.

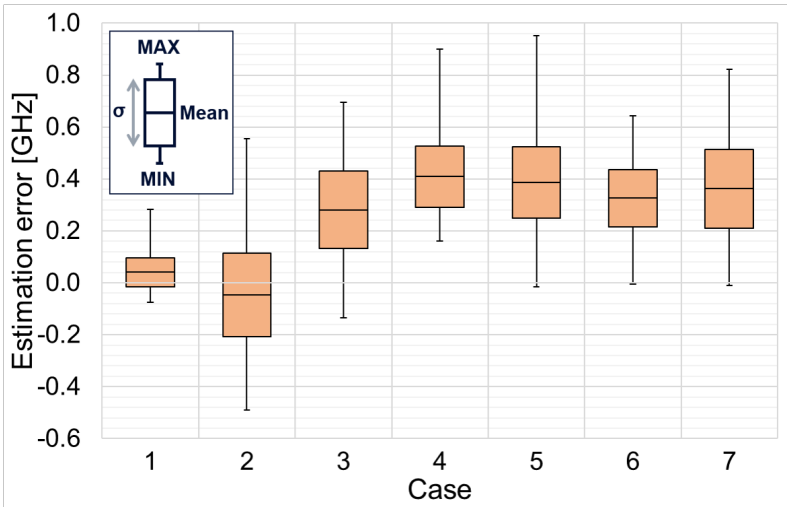

FIGURE 5.19. Filter 6 dB bandwidth estimation errors for the seven experimental cases with roll-off factor = 0.1.

and FIGURE 5.21 for roll-off factor of 0.2. To represent them we employed the box and whiskers plots, which graphically depict σ, the mean, the MIN, and the MAX errors for each of the seven considered cases. From such results, it is clear how the default case (i.e., case 1) showed the best performance, especially concerning the filter bandwidth estimation. On the other hand, when we considered the cases with narrower Filter *1* bandwidth (i.e., cases 4 and 7) or higher link attenuation values (i.e., cases 2 and 7), the estimation accuracy tended to degrade. In general, we obtained a lower accuracy for the filter 6 dB bandwidth estimation compared to the accuracy of the filter central frequency shift estimation for both the considered roll-off values. The results we report in this section referred to optical spectra acquired with a spectral resolution of 600 MHz. We repeated the fitting/estimation process employing optical

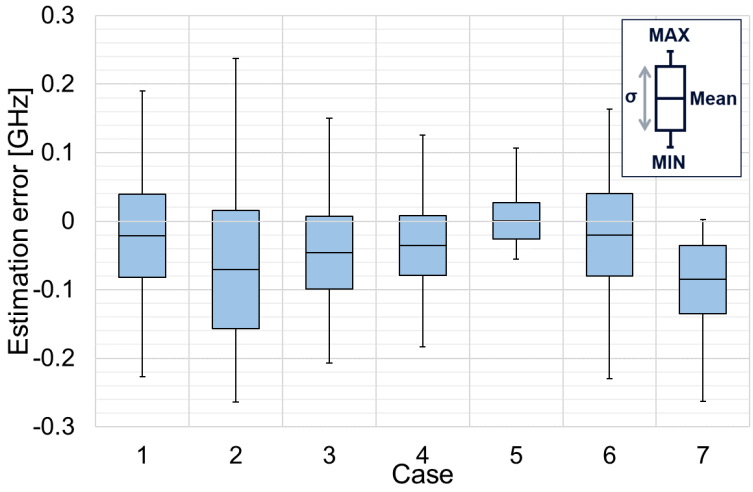

FIGURE 5.20. Filter central frequency shift estimation errors for the seven experimental cases with roll-off factor = 0.2.

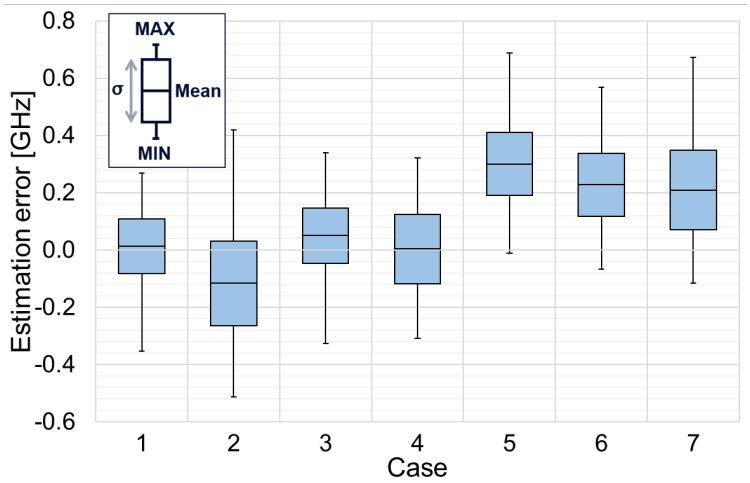

FIGURE 5.21. Filter 6 dB bandwidth estimation errors for the seven experimental cases with roll-off factor = 0.2.

spectra at 1 GHz resolution and did not observe any substantial difference in the estimation accuracy.

5.4.3 Monitoring Placement Scenarios and Related Estimation Methods Comparison

In the previous sections, we proposed two optical monitor placement scenarios enhanced by appropriate estimation methods to complement the monitoring parameters achieved in each scenario. This section aims to compare such two solutions and to select the one that brings more benefits. However, the parameters we estimated within the two placement scenarios have different

units of measurement. On the one hand, the OSNR/ASE noise estimation method for the egress placement scenario presented in Section 5.2 provides values in decibel. On the other hand, the filter-related parameters extraction method for the ingress placement scenario presented in Section 5.3 estimates quantities in Hertz. Hence, in order to compare the two solutions and their estimation accuracies, we converted the estimation errors obtained with the unified experimental setup presented in Section 5.4, into relative errors. To do so, we first transformed the estimation errors into percentages with respect to their nominal values, and then we translated the MAEs into relative errors. Of course, as a first step, we identified the values to be used as reference. Concerning the egress scenario, we obtained a reference for the OSNR/ASE noise estimation, evaluating the amplifier NF as follows. We considered the cascade of a number of fiber spans with EDFAs at the end of each of them. We assumed the EDFAs to have a NF equal to 5 dB, with fluctuations of ±0.5 dB [135], [136]. This assumption yielded a noise reference error with a 1 dB range. Instead, for the ingress scenario, we considered the datasheet of the Finisar filter that we employed in the unified experimental setup [133]. In particular, such reference reported a central frequency setting accuracy of ±2.5 GHz and a bandwidth setting accuracy of ±5 GHz. Nevertheless, we also considered that WSSs deployed in real networks could have better characteristics. Thus, we decided to adopt for both the filter-related parameters a reference accuracy equal to ±2 GHz [84], [99].

Relying on the filter- and signal-related parameter estimation errors reported in Sections 5.4.1 and 5.4.2, we identified the worst-case error ranges and calculated accordingly the relative errors with respect to the aforementioned reference values. We summarize the results of such calculations in TABLE 5.7. In particular, we found that the egress monitoring strategy improved the ASE noise estimation accuracy of 22%, with respect to the case in which no monitoring strategy was implemented. Likewise, for the ingress monitoring one, we observed an improvement in the filter central frequency estimation

TABLE 5.7. Error range and relative error comparisons for the two different considered scenarios.

Scenario	Parameter	Roll-off factor	Parameter error range	Reference error range	Relative error
Egress	ASE noise	0.1	0.78 dB	1 dB	78%
Ingress	Filter centr. freq. shift	0.1	0.63 GHz	4 GHz	15.8%
		0.2	0.5 GHz	4 GHz	12.5%
	Filter 6 dB BW	0.1	1 GHz	4 GHz	25%
		0.2	0.93 GHz	4 GHz	23.3%

greater than 84% for 0.1 roll-off factor, and greater than 87% for 0.2 roll-off factor. As per the egress case, these improvements were with respect to scenarios where no monitoring strategy was adopted, and the nominal parameters provided by the vendors/datasheet are considered. Moreover, considering the filter 6 dB bandwidth estimation, we were able to improve it by 75% for 0.1 roll-off factor and by a factor greater than 76% for 0.2 roll-off factor.

It is worth mentioning that the relative error we considered is a valuable metric, but not perfectly suitable for eventually evaluating the impact of the network elements on the QoT of the connections (e.g., on the OSNR/SNR). Hence, to evaluate the QoT estimation related benefits, we further translated the obtained estimation errors into SNR estimation errors. Of course, concerning the egress scenario, there was no need for such a translation, since we evaluated the improvements in OSNR/noise estimation. In fact, our proposed egress monitoring placement strategy and its related processing/estimation method resulted in 0.22 dB improvement in OSNR estimation per link, as we also reported in TABLE 5.7. Unlike this case, to obtain a similar metric in the ingress placement strategy scenario, we simulated in VPI the transmission of a 64 GBd QPSK modulated signal, with roll-off factor equal to 0.1 and 0.2, which crossed a single optical filter. In particular, we implemented the filter with a 3.5-order super-Gaussian TF and 75 GHz bandwidth and calculated the SNR penalty it introduced as a function of its central frequency shift and its bandwidth variation. To do so, we first measured the SNR values at the optical filter input and output when no central frequency shift and bandwidth tightening were present, thus retrieving the SNR penalty introduced by the filter in the ideal configuration. Then, sweeping the filter central frequency and bandwidth values with step sizes of 0.5 GHz and 1 GHz, respectively, we again measured the SNR at the filter output. Finally, by subtracting from the values obtained in these conditions those of the ideal configuration (i.e., no filter central frequency shift and no bandwidth tightening), we obtained the SNR penalties introduced by an optical filter as a function of its central frequency shift and its bandwidth variation. We plot the results of such simulations in FIGURE 5.22. Then, leveraging the obtained curves, we measured the improvement that our solutions brought on the SNR penalty estimation. Since the differences between the penalty variations of the two roll-off factor cases were negligible, we considered them as a unique case. In detail, the reference filter-related parameters error ranges, which were equal to ±2 GHz, translated into SNR penalty estimation of 0.12 dB for the filter central frequency shift and of 0.16 dB for the filter bandwidth. Instead, applying the proposed ingress monitoring strategy and the related processing solution, we reduced the error of the estimated filter-introduced SNR penalties down to 0.01 dB and 0.03 dB for the central frequency shift and the 6 dB

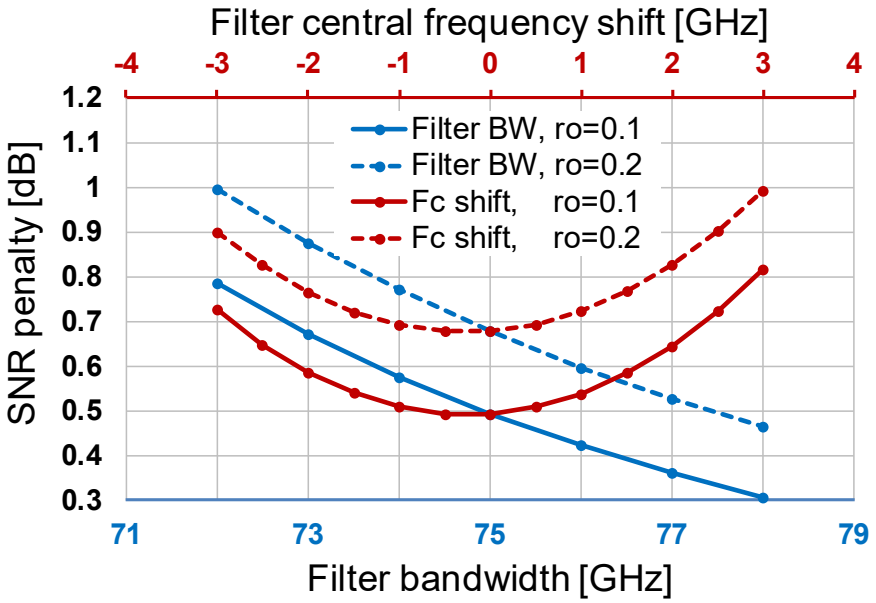

FIGURE 5.22. SNR penalty introduced by a 3.5-order super-Gaussian filter as a function of its central frequency shift and its bandwidth variation for an input 64 GBd QPSK signal with roll-off factor equal to 0.1 and 0.2.

bandwidth parameters, respectively. Therefore, the total estimated SNR penalty reduction yielded by our solution was equal to 0.24 dB for the two considered parameters.

5.5 Conclusions

In this chapter, we studied different scenarios for the placement of optical spectral monitors within DWDM and flex-grid optical networks. In particular, we defined an egress and an ingress placement scenario, in which the optical spectral monitors are positioned after and before the network nodes, respectively. Our goal was to optimize the placement of the employed spectral monitoring devices in order to minimize their number. In addition, we developed two spectral processing techniques related to the two identified placement scenarios to enhance their monitoring features. We validated the proposed solutions leveraging spectral data generated through several simulation and experimental setups. Finally, we compared the two monitoring strategies and provided guidelines in terms of SNR penalty reduction for their adoption. It is worth noting that in our evaluations, we assumed the NLI not to impact the accuracy of the proposed solutions. Therefore, further investigations, which include the transmission of multiple channels, are needed.

From this analysis, we can draw the following conclusions. The obtained results always confirmed the validity of both the proposed techniques. In particular, using the spectral data acquired within the egress placement scenario, we were able to estimate in-band the OSNR/ASE noise of a signal with an MAE lower than 0.4 dB. Similarly, leveraging optical spectra collected within the ingress placement scenario, we were able to estimate filter-related features, such as the filter central frequency and the filter 3/6 dB bandwidth, with MAEs lower than 0.5 GHz and 0.98 GHz, respectively. In addition, comparing the two monitoring strategies, we identified the ingress one as the most promising solution. In fact, compared against scenarios where no monitoring strategies are adopted, the ingress one reduced the estimation of the SNR penalty introduced by an optical filter up to 92%.

Chapter 6

Frequency Optimization in Superchannel

We devote this chapter to address the third major task of this thesis, which consists in optimizing the network, leveraging the information directly monitored from it, or exploiting those resulting from processing methods. In Section 3.4, in which we defined the reference scenario, we named such an optimization task as task C, as depicted in FIGURE 3.5. With respect to Chapter 5, where we focused on OSNR monitoring using OSAs, in this chapter, we assume a slightly different monitoring approach that exploits SNR monitoring at the coherent receivers. In particular, the approach presented in this chapter, which we initially proposed in [137], deals with the frequency optimization of the subchannels in superchannel scenarios. To address such a problem, we implement a closed control loop employing an iterative optimization algorithm based on the stochastic subgradient method introduced in Section 2.2.4.

More in detail, this chapter is organized as follows. In Section 6.1, we introduce the main aspects of our proposed solution, providing a context for the following sections. Then, in Section 6.2, we formulate the considered optimization problem, detailing its objectives and constraints. Section 6.3 describes how we addressed the identified problem, leveraging the stochastic subgradient method. Then, Section 6.4 shows the integration of the developed optimization algorithm within the VPI simulation tool and the validation of the proposed solution based on such implementation. We present the obtained results in Section 6.5. Finally, in Section 6.6, we draw the conclusions.

6.1 Introduction

In Section 3.2.3, we reported about [99], in which the authors presented a closed control loop approach to align a single optical channel to a cascade of optical filters to overcome filter-related impairments in a single channel scenario. Furthermore, in Section 1.1, we also reported about [5] and [6], in which a similar solution was leveraged in a field demonstration of an autonomic, self-reconfigurable network. In this chapter, we extend these works, presenting a method to optimize the subchannel central frequencies of a superchannel in order to minimize their interference. Note that similar strategies can also be adopted at a network level, where adjacent channels entering a ROADM-based node at the same ingress port are forwarded to the same egress port, therefore sharing a superfilter configuration, as we already explained in Section 3.3. In fact, any misalignment of the channels sharing such superfilter would result in an interference similar to that faced in superchannels. So, the method we propose can also be potentially extended to reduce this interference and align the whole network channels.

In particular, we formulated the problem of optimizing the superchannel frequencies as a nonlinear optimization one. Our objective was to identify the set of subchannel frequencies that optimized a function of the subchannels QoT. To be more specific, we considered maximizing the superchannel total SNR value and the minimum SNR value between all the subchannels. The nonlinear behavior came from the dependence on the interference (with two components: XCI and linear crosstalk) and the filtering penalties affecting the two edge subchannels, as we explained in Section 3.3. As a subchannel moves closer to its adjacent ones or the filter edges, its SNR value decreases nonlinearly. To this aim, in Section 6.2.1, we report the results of simulations showing the channel SNR function to be concave around the typical operation point (equal distance). According to the definitions presented in Section 2.2.1, a concave function can be easily transformed to convex by taking its negative. Hence, convex optimization techniques can be applied to solve the considered optimization problem. However, due to several reasons, such as random noise, imperfections of transmitters and receivers, and monitoring errors, the convex function might not be completely smooth. Thus, the proposed method leveraged a stochastic subgradient, an algorithm able to solve convex problems with zero mean noise [43] and several other nonconvex problems [138], as we described in detail in Section 2.2.4. We chose the proposed algorithm to be robust and able to optimize under the uncertainties that are expected to appear in a real network. Relying on the stochastic subgradient algorithm, we proposed an iterative closed control loop process to optimize the superchannel frequencies. Specifically, the automatized optimization process we assumed used subchannel-related information (i.e., their SNR values monitored at their

corresponding coherent receivers). Then, the algorithm probed the superchannel with new frequencies, monitored the outcomes again, and moved to new frequencies towards the optimum (even though the stochastic subgradient might not move at each step towards the optimum).

We evaluated the proposed solution integrating the considered optimization algorithm within the VPI simulation tool, which has the advantage of simulating the physical layer of an optical network in a more detailed and complex way with respect to analytical Physical Layer Models (PLMs), such as the well-known Gaussian Noise (GN) model [139]. To maximize the desired objective function, we fed the information monitored from VPI to our optimization algorithm. Then, relying on the algorithm outputs, we adjusted the subchannel frequencies accordingly and monitored the corresponding VPI outputs, passing them to the algorithm. In turn, a new algorithm iteration started, and the control loop was finally closed. We believe this approach to be universal and applicable in different contexts, regardless of the considered ROADM architecture (see Section 3.2.3), network disaggregation level (see Section 3.1.3) or its specific segment (e.g., metro, core, etc.).

6.2 Problem Formulation

To formulate the optimization problem, we assumed a superchannel with an allocated spectrum BW_{SC} centered around the frequency f_{SC} and composed by a set of N subchannels, as shown in FIGURE 6.1. In addition, we considered the spectral shaping of the subchannels to be defined by a square Root Raised Cosine (RRC) with a roll-off factor equal to a. If we assume a single channel with symbol rate R_S and roll-off factor a to pass through an optical filter with

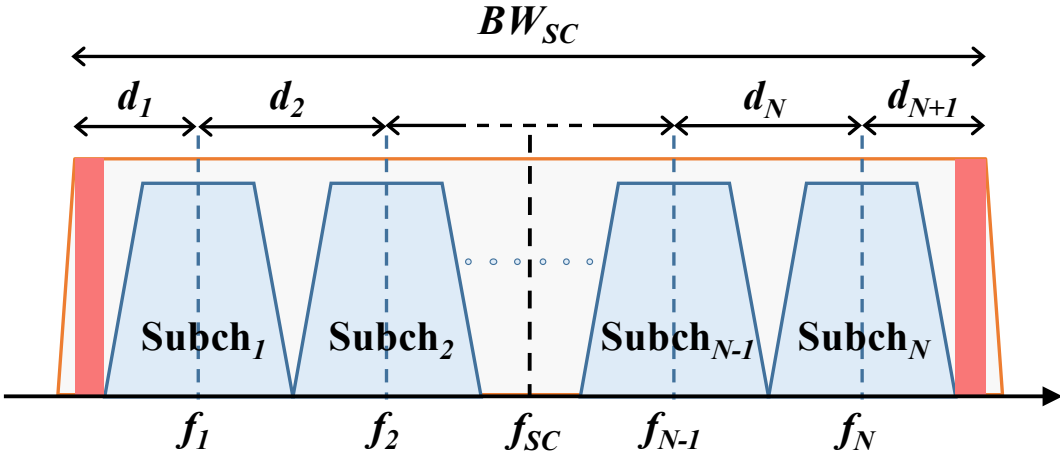

FIGURE 6.1. Schematic representation of the considered N subchannels superchannel.

bandwidth BW, the minimum required filter bandwidth BW_{min}, for the channel to not heavily suffer the filtering effects, would be

$$BW_{min} = R_S(1 + a). \tag{6.1}$$

For instance, considering a 32 GBd single channel with a 0.1 roll-off factor that crosses an optical filter, the minimum required filter bandwidth to avoid filtering-related penalties would be 35.2 GHz. Hence, according to the flex-grid standard that considers filter 3 dB bandwidths equal to 37.5 GHz, such transmission would have enough spectral guardband space to avoid considerable filtering effects.

Translating such a relation to the considered superchannel scenario, the subchannel minimum distance d_{min} to have crosstalk-free (or very low crosstalk) interactions with the adjacent subchannels is defined as

$$d_{min} = R_S(1 + a)\,. \tag{6.2}$$

For instance, in a superchannel with subchannels defined by a 32 GBd symbol rate and 0.1 roll-off factor, considering distances between subchannels larger than 35.2 GHz would result in a very low crosstalk effect. It is worth mentioning that such distance calculation assumes equidistant and uniform superchannels, but it can also be extended to cover nonequidistant and nonuniform configurations. Finally, we assumed the superchannel to be established in a DWDM/flex-grid optical network, crossing at least one filter with 3 dB bandwidth BW_{SC} and central frequency f_{SC} (also referred to as superfilter).

Extending the notation to cover nonequidistant superchannels, we denoted by $\mathbf{d} = [d_1, d_2, \ldots, d_{N+1}]$ the vector of length $N + 1$, where d_1 and d_{N+1} represent the distances between the superchannel filter sides (3 dB) and the first and last superchannel subchannel central frequencies, f_1 and f_{N+1}, respectively. Moreover, d_n, with $2 \leq n \leq N$, represents the generic distance between the $(n-1)$-th and the n-th subchannel central frequencies, f_{n-1} and f_n, respectively.

We assumed the SNR values of the superchannel subchannels to be functions of the distances between their central frequencies. We denote the SNR value of the n-th subchannel as $SNR_n(\mathbf{d})$. In particular, $SNR_n(\mathbf{d})$ mainly depends on the distances d_n and d_{n+1}, which are those separating subchannel n from its two adjacent ones, $n-1$ and $n+1$, respectively. In fact, when their spectra overlap, (linear) crosstalk is created, and the highest XCI also occurs. Moreover, since a small contribution is also given by the XCI caused by all subchannels, SNR_n also depends on all subchannel distances $\mathbf{d}$. Bringing two subchannels closer would increase their crosstalk and interference, resulting

in a reduction of their SNR values with respect to their distance. On the other hand, such a shift would reduce their crosstalk and interference with their adjacent outer ones. Specifically, in Section 6.2.1, we will show the SNR function to be concave around the point corresponding to the equal distance of a subchannel from its adjacent ones. Hence, the optimization problem aims to identify the proper distance set $\mathbf{d}^*$ that maximizes specific functions of the subchannel SNR values.

In particular, for the problem at hand, we considered two optimization functions: the maximization of the superchannel total SNR value, which we named Obj#1, and the maximization of the subchannel minimum SNR value, which we named Obj#2. The former objective is directly correlated to maximizing the superchannel total capacity and is equivalent to the summation of all the subchannel SNR values. Instead, Obj#2 aims to make all the superchannel subchannels feasible with respect to a specific SNR threshold related to the considered modulation format.

We formulated the optimization problem as follows:

$$\max\big(h(\mathbf{d})\big), \tag{6.3}$$

where

$$Obj\#1 \qquad h(\mathbf{d}) = \sum_{n=1}^{N} SNR_n(\mathbf{d}), \tag{6.4}$$

or

$$Obj\#2 \qquad h(\mathbf{d}) = min_n\big(SNR_n(\mathbf{d})\big)\,. \tag{6.5}$$

Equations 6.4 and 6.5 correspond to Obj#1 and Obj#2, respectively.

In addition, the following set of constraints applies:

$$\sum_{n=1}^{N+1} d_n = BW_{SC} \quad n = 1, \dots, N\,, \tag{6.6}$$

$$d_n = \begin{cases} \dfrac{R_S}{4}, & n = 1, N+1 \\ \dfrac{R_S}{2}, & 1 < n < N+1 \end{cases}. \tag{6.7}$$

Specifically, the constraint defined by Equation 6.6 keeps the subchannel distances within the filter bandwidth. Instead, the constraints defined by Equation 6.7 put a lower limit on the distances between adjacent subchannels and outer subchannels and filter sides, based on the subchannels symbol rates. Note that in the above relations, we assumed the SNR as the key metric, but

we can convert it to a corresponding SNR margin by considering the SNR threshold of the employed modulation format. It is also worth reminding that the above formulation applies to uniform superchannels, where all the subchannels have the same symbol rate, roll-off factor, and modulation format. However, this model and our proposed solution could also be extended to cover nonuniform superchannels.

6.2.1 Channels Distance and SNR Function Properties

To understand the relation between the subchannel distances and their SNR values, we implemented in VPI a simulation setup considering three channels with nonidentical Pseudo-Random Binary Sequences (PRBSs). We generated three 32 GBd QPSK modulated signals, pulse shaped with an RRC filter with 0.1 roll-off factor. In FIGURE 6.2, we plot the SNR values of the second channel as a function of its distance d_1 from the first channel. To do so, we kept the first and third channel central frequencies constant and shifted the second channel one. The obtained SNRs referred to a B2B scenario and a transmission over two and ten SSMF spans. In particular, we considered SSMF spans of 80 km length, with attenuation and dispersion coefficients equal to 0.2 dB/km and 16.7 ps/nm/km, respectively. Moreover, each SSMF span was followed by an EDFA with an NF of 5.5 dB and gain that matched the previous span loss (i.e., Automatic Gain Control, AGC).

From FIGURE 6.2, we observe that the SNR of the central channel (i.e., the second) is a concave function of d_1 around the equidistant point ($d_1 = d_2$ ±2 GHz). As so, the considered objective functions, which are combinations of such functions, are also concave. Varying the channel distance, the resulting change in the XCI is relatively small, so the dominant effect observed in FIGURE 6.2 is the linear crosstalk between adjacent channels. Therefore, the

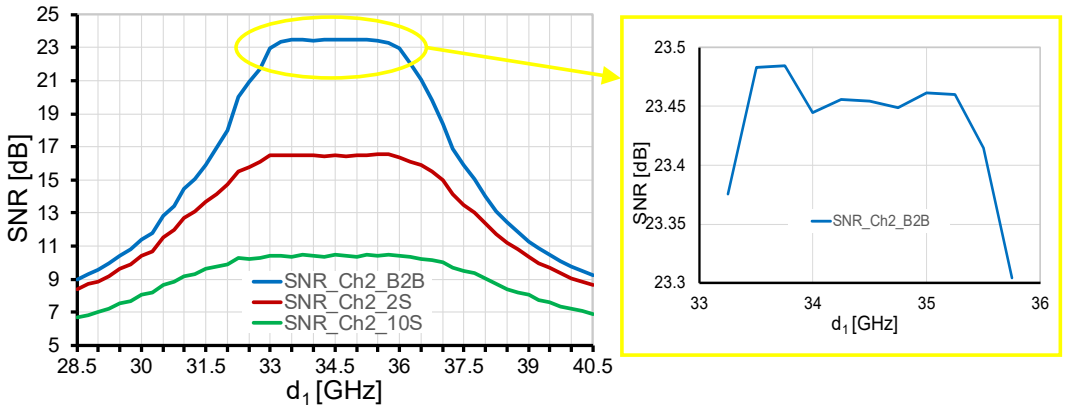

FIGURE 6.2. Relation between the SNR value of the second channel and its distance from the first one, for a B2B, two and ten SSMF spans scenarios. The yellow box presents a zoomed view of the central part of the B2B SNR graph.

shape of the SNR function depends on the crosstalk, which in turn depends on the channels shaping. In particular, the crosstalk contribution from the adjacent channel is convex: as we move the second channel towards the first, the amount of area in which the signals overlap increases. The crosstalk is proportional to the overlapping area integral, and the integral of an increasing function is convex, as shown in FIGURE 6.3. Similar observations would still hold if we moved to the other side. The sum of two convex functions is convex, so the total crosstalk contribution is convex. Moreover, the SNR depends inversely on the crosstalk and thus is concave.

So, based on the above, we expect a single or a continuous set of distances d_1 that exhibits the maximum SNR value (or almost reaches the maximum) for the second channel, making gradient-based methods suitable to solve the considered problem. Note that a similar shape appears when one of the channel sides faces the filter. Furthermore, we observed that transmitting the signal over an increasing number of SSMF spans reduced the reference SNR value and flattened the shape of the SNR curve. However, the concavity of the function is maintained, and the proposed optimization process can still be applied, as indicates by our results. The causes of the SNR function flattening are the noise generated by the EDFAs, which accumulates between the channels, and the widening of the channel spectra caused by the dispersion. Thus, the channel spectra are shallower and their sides less sharp, resulting in less crosstalk area when they overlap and in flatter SNR curves.

To be more specific, although from the main plots of FIGURE 6.2 it might seem that there is a set of distances d_1 for which the maximum SNR values flatten out, zooming in that area, we see minor SNR variations, as shown in FIGURE6.2 yellow box. Such fluctuations can be caused by random noise,

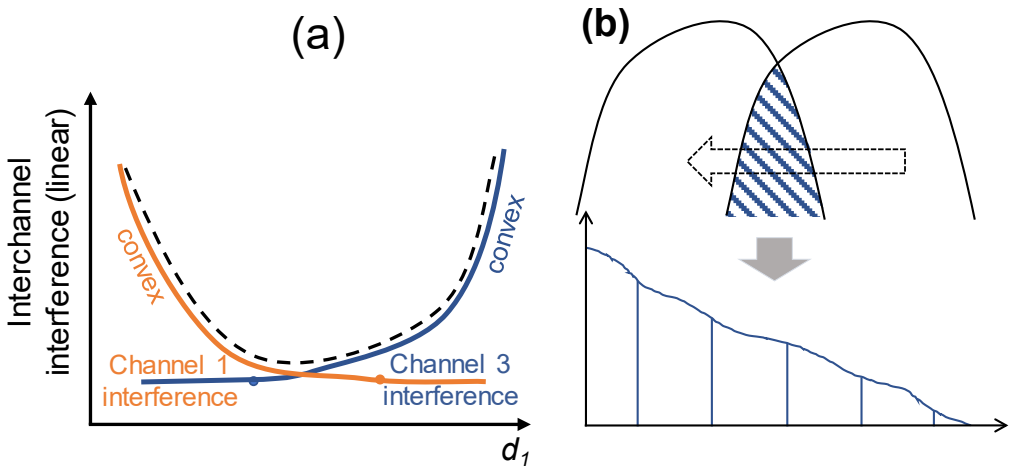

FIGURE 6.3. (a) Graphical representation of the interchannel interference convexity. (b) The integral of an increasing function is convex.

wavelength-dependent penalties, and other factors. In real systems, we would even have to consider additional effects coming from transmitter/receiver imperfections (as mentioned in Section 6.1, we assume to monitor the subchannel SNR values at the coherent receivers). Hence, in that area, there are several local maxima close to the global, which are hard to avoid when optimizing the subchannel distances, and that would even vary as a function of time. For these reasons, we decided to use the stochastic subgradient method with a fixed step length, which is robust, can optimize in the presence of noise and uncertainties, and works even with nonlinear and nonconcave (or nonconvex) problems.

In our optimization, we target to reach the flat area and any local maxima close to the global one. In particular, using the proposed optimization method, the results we obtained consistently ranged within 0.2 dB from the optimum.

In addition to the relation between the subchannel distances and their SNR values, we also need to consider that one between the two outer subchannels and the filter sides. To this aim, we implemented a VPI simulation setup similar to the previous one, with only two subchannels. In this setup, in addition, following each SSMF span and EDFA, we placed a 3.5th-order super-Gaussian TF optical filter, symmetrically centered around the two signals, with 92 GHz 3 dB bandwidth. We kept all the system parameters and the considered configurations (i.e., B2B, 2 and 10 fiber spans) as they were in the previous simulation. In FIGURE 6.4, we plot the SNR values of the left subchannel as a function of its distance d_2 from the right one (please note that in this case d_1 represents the distance between the left filter side and the left subchannel central frequency). Of course, filtering effect becomes stronger when longer filter cascades are considered, resulting in SNR degradations.

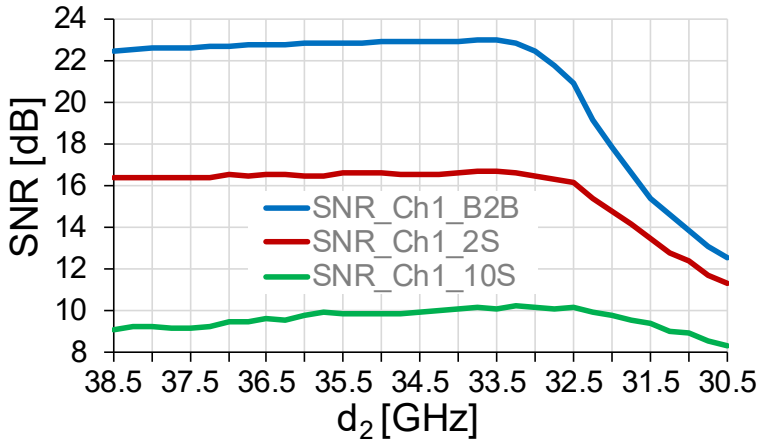

FIGURE 6.4. Relation between the SNR values of an outer subchannel and its distance from the adjacent one for a B2B, two and ten SSMF spans scenarios.

From FIGURE 6.4, we observe that the outer subchannel SNR is a concave function of the distance d_2 between the two subchannels. As per the previous analysis, similar observations would still hold if we moved to the other side. As a result, also in this case, gradient-based methods are suitable to solve the considered optimizations problem.

6.2.2 Dynamic Superchannel Optimization

Typically, a superchannel would be designed with appropriate transmitter and receiver configuration and equal distances between subchannels [111]. However, imperfections such as those related to the subchannel transmitters and receivers, dynamic impairments, wavelength-dependent losses / performance variations (e.g., amplifier gain ripple), and ageing of the various network elements must also be considered. All such factors result in the subchannel QoT variations (e.g., their SNR values) and make equidistant superchannels suboptimal and unattainable during the network operation. In fact, the control plane would configure the superchannel as equidistant, but due to the above imperfections and laser/filter drifts, it would not actually be. For this reason, some margins in QoT or additional spectra between the subchannels are typically used. Note that such effects are unknown before the superchannel is established and operates, and they might also vary with time. For example, assuming a ±2 GHz drift for the second channel of FIGURE 6.2 in the two spans case, we end up with a 0.7 dB lower SNR. To compensate for this, the same dB amount should be considered as a margin when selecting the feasible superchannel modulation format (actually, the margin should be higher because we should consider the worst case, in which the first and third channels also drift).

To remove the margin or increase the efficiency of the superchannel, we need a feedback-based approach that dynamically interacts with the superchannel and the network, understands the current conditions, and corrects and compensates these effects. In the following section, we propose such an approach that dynamically optimizes the superchannel as the network operates.

6.3 Closed Control Loop-Based Solution

As introduced in Section 6.1, our solution envisions implementing a closed control loop leveraging an iterative optimization algorithm based on the stochastic subgradient method with a fixed step length. In particular, we assumed the optimization algorithm to reside in the central SDN controller.

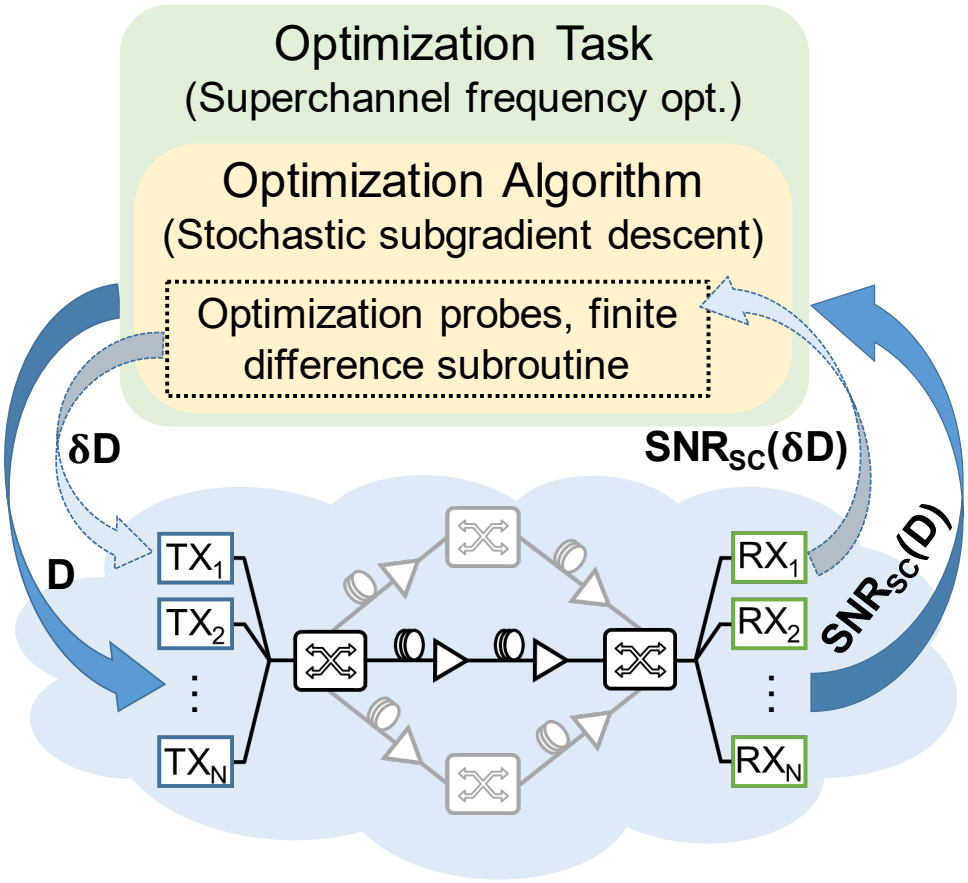

FIGURE 6.5. Schematic representation of the implemented optimization loop with monitoring probes.

Here, appropriate interfaces to configure the transmitter lasers and communicate with the coherent receivers are present to retrieve the subchannels SNR values. Our optimization algorithm employs a subroutine to specify the probes, which are those configurations applied to the network through the control plane. The outcomes of such applied probes, which are the monitored SNR values, are in turn delivered back to the optimization algorithm to provide it with the information it needs to perform the intermediate optimization steps. This process is repeated at each algorithm iteration. We also refer to this technique as optimization with monitoring probes. A representation of such a scheme is shown in FIGURE 6.5.

In Section 6.2, we showed how the problem of maximizing the two considered objectives, namely Obj#1 and Obj#2, is concave with respect to the distances between subchannels. Hence, we can transform it into the minimization of a convex function simply by taking the negative of the SNR function. In fact, as we saw in Section 2.2.1, several known optimization techniques exist in the literature to address such convex problems. In particular, we chose to use a method based on a stochastic subgradient that can optimize a convex problem even in the presence of noise. Optimization algorithms such as (sub)gradient

method, interior point, and trust-region-reflective are iterative, meaning that they need to calculate the first- or second-order partial derivatives at each iteration. For optimization problems involving the optical physical layer and the capability to probe the network (configure and monitor the outcome), a typical way to find the partial derivatives is by using a subroutine that implements the finite differences method [140].

To calculate the (sub)gradient of the objective function h, which we defined with Equations 6.4 and 6.5, we need to configure new frequencies and monitor the related changes in the subchannels SNR values. Specifically, following the notation we previously introduced and referring to FIGURE 6.1, we assume a superchannel consisting of a set of N subchannels separated by a set of distances $\mathbf{d}$. We denote by $\boldsymbol{\delta}\mathbf{d_n}$ the distance vector for which the central frequency of the n-th subchannel has increased by f_{step}, which we refer to as the frequency probe step, and in the optimization process is considered to be a fixed step length. This implies that f_{step} is subtracted from d_{n-1} and added to d_n.

We denote the SNR value of the shifted subchannel n with $SNR_n(\boldsymbol{\delta}\mathbf{d_n})$, whereas the SNR of all subchannels with $\mathbf{snr_{SC}}(\boldsymbol{\delta}\mathbf{d_n}) = \{SNR_1(\boldsymbol{\delta}\mathbf{d_n}), \ldots, SNR_N(\boldsymbol{\delta}\mathbf{d_n})\}$. Note that the change in the central frequency of the single subchannel n mainly affects the SNR values of the adjacent subchannels ($n-1$ and $n+1$) through (linear) crosstalk, but more lightly also of all the other subchannels through XCI. Therefore, the partial derivative g_n of the optimization function h for the n-th subchannel is given by

$$g_n = \frac{h(\mathbf{d}) - h(\boldsymbol{\delta}\mathbf{d_n})}{f_{step}}. \tag{6.8}$$

Depending upon the chosen objective function h, the partial derivative identification involves specific operations with the vectors $\mathbf{snr_{SC}}(\mathbf{d})$ and $\mathbf{snr_{SC}}(\boldsymbol{\delta}\mathbf{d})$, as indicated in the two considered objectives equations, namely Equations 6.4 and 6.5. Thus, to calculate the (sub)gradient with the finite difference method, we need to probe with $\boldsymbol{\delta}\mathbf{d_n}$ and monitor $\mathbf{snr_{SC}}(\boldsymbol{\delta}\mathbf{d_n})$. We repeat this for the subchannels that are chosen (stochastically) for each algorithm iteration.

The algorithm at each iteration calculates the subgradients of a set of randomly selected subchannels through the finite differences subroutine. Specifically, we use the minibatch option of the stochastic method, in which at each iteration, we randomly choose a set of M subchannels to probe and monitor. In order to converge or approach the optimum, the stochastic subgradient method with a fixed step length f_{step} and monitoring probes runs I iterations, stopping when the objective improvement rate falls below a specified threshold. In particular,

if $f_{step} = 2\varepsilon/G$, where G represents a bound on the gradient such as $|h(\mathbf{d_u}) - h(\mathbf{d_v})| < G\|\mathbf{d_u} - \mathbf{d_v}\|_2$, valid for any $\mathbf{d_u}$ and $\mathbf{d_v}$, then the subgradient method converges to the optimum within ε. Specifically

$$h_{best}^{i} - h^{*} \leq \varepsilon\,, \tag{6.9}$$

where, h_{best}^{i} and h^{*} represent the objective functions related to the i-th iteration and the optimal solution, respectively. Therefore, the required number of steps to converge towards the optimum within ε is given by $(RG/\varepsilon)^2$, where R is the distance of the starting point from the optimum, such as $R \geq \left\|\mathbf{d^1} - \mathbf{d}^{*}\right\|_2$. Note that the subgradient method discussed above is a batch method; it calculates the subgradients for all the functions summed in the objective (e.g., the SNR function of the different subchannels). The stochastic subgradient method that we adopt has a bound on the convergence rate given by $O(N^2/\varepsilon^2)$. When we use minibatches of size M, we have a bound on the convergence rate given by $O(N^2/M^2\varepsilon^2)$ [141]. So, the bound on the expected number of iterations of the stochastic subgradient method with minibatches is given by

$$I = \left(\frac{NRG}{M\varepsilon}\right)^2\,, \tag{6.10}$$

In turn, within each iteration, the (sub)gradient identification subroutine is called M times, corresponding to the times the algorithm probes the network and monitors the SNR values.

We denote by t_{mon} the monitoring time needed to monitor all the N subchannels simultaneously. Monitoring, is typically done every 15 minutes in today networks, but the new generation of telemetry-based solutions can speed up this process to subsecond timescales [142]. However, the monitoring time here is constrained by other factors. Once a reconfiguration is decided and applied, the time needed by the transponders to adapt to such change and by the network to reach a stable state (e.g., for transient effects to settle) must also be considered. Moreover, monitoring the SNR is not instantaneous but requires to average over a specific period. For random Gaussian errors the monitoring accuracy increases with the monitoring time. If monitoring is fast and the error is high, the algorithm will have to perform more iterations. Thus, although we will have a small t_{mon}, we will pay that with a higher iteration number I. Note that the stochastic subgradient method finds the optimum in a polynomial number of iterations I. Based on the above, we expect that t_{mon} may range from tens of seconds to minutes, depending upon the network complexity, the monitoring plane, the targeted monitoring error, and other factors [142]. However, once the monitoring information is forwarded to the algorithm, the time t_{calc} that it needs to calculate the (sub)gradients and the

following set of frequencies is relatively low (within seconds) compared to the monitoring time (i.e., $t_{mon} \gg t_{calc}$). So, with the proposed approach, under the assumption that the N subchannels are monitored in parallel, the total optimization time T_{opt} is given by

$$T_{opt} = I \cdot (M \cdot t_{mon} + \ t_{calc}) \approx I \cdot M \cdot t_{mon} \,. \tag{6.11}$$

Returning to the optimization problem, the set of distances **d** is defined by the central frequencies of the subchannel transmitter lasers. However, the lasers cannot be configured with infinite accuracy. Moreover, since we rely on feedback/physical layer monitoring information and because of physical layer variations (our optimization is mainly affected by short and medium terms) and monitoring errors, we are not able to observe fine differences in the monitored SNR values. For these reasons, as discussed above, we use for the gradient identification in both the optimization problems a frequency step equal to f_{step}. For example, such a step can be equal to 0.25, 0.5 or 1 GHz. This value should be chosen according to the laser configuration capabilities (i.e., the laser frequency tuning resolution) and the uncertainties coming from variability and monitoring errors. In particular, smaller values tend to make the stochastic subgradient method slower but more precise. Moreover, laser stability represents a crucial factor when considering small amounts of frequency shift like those we employed in our evaluations (e.g., 0.25 GHz). However, fast drifts will be averaged in the monitoring process, whereas slower drifts/biases will be corrected through the proposed process. Moreover, consider that we based our solution on the stochastic subgradient method which is robust to noise and uncertainties.

Compared to the optima found with a brute-force approach, our results indicated good convergence performance for frequency steps of 0.25 GHz and 0.5 GHz, whereas, for 1 GHz steps, the performance tended to degrade slightly.

Note that the monitoring probes used to identify the (sub)gradient are not a universal solution. A monitoring probe here refers to the configuration of new frequencies for one or more subchannel lasers and monitoring the subchannel SNR values at their receivers. Thus, each optimization problem requires a specific definition of monitoring probes. For some problems, monitoring probes might not be available, e.g. problems involving establishing or releasing connections. Moreover, the monitoring probes approach involves several interactions with the network, which might be time-consuming. Therefore, alternative methods can include the use of PLMs or QoT estimators [139]. It is also worth mentioning that our approach based on monitoring probes achieves the optimal performance as long as the problem has specific properties (convexity or unimodularity) and an appropriate algorithm is used. Nevertheless, such a solution might take a long time due to the delay

introduced by the monitoring process. On the other hand, PLM has the benefit of being much faster and without monitoring noise, but its inaccuracies might mislead the optimization. Such concepts are studied in [143] for another optimization problem and are left for future investigations concerning superchannel optimization.

Regarding the optimization algorithm, we adapted the stochastic subgradient algorithm with minibatches. Given the subgradients of M subchannels, specified with the monitoring probes / finite difference method mentioned above, the algorithm decides on the next move and identifies the new central frequencies of all the subchannels. Note that the algorithm might decide to make a more complex action (move several subchannels) than the simple movements performed in the gradient identification subroutine. The algorithm output is mapped to the network, and a new probing and monitoring phase starts. The stochastic subgradient algorithm converges to the optimum, with a specific accuracy, in a polynomial number of iterations [141]. However, although the algorithm selects/calculates new variable values to improve the objective at each step, when applied to the actual system they might not be effective because of the noise. So, the best result is memorized and reused as starting point in case the objective falls. Once the algorithm converges (no further improvement of the objective is calculated for a certain amount of iterations, with a defined accuracy set, for instance, to 0.2 dB), the process terminates. The configuration saved as the best one up to that point corresponds to the optimized frequency set for the superchannel subchannels.

6.4 Optimization Algorithm and VPI Setup Integration

To quantify the benefits of the proposed optimization solution, we carried out several simulations using both VPI and MATLAB. In particular, we co-simulated the superchannel and its transmission in VPI, whereas the finite differences subroutine and the stochastic subgradient algorithm were developed in MATLAB. To be more specific, in MATLAB, we implemented the frequency optimization algorithm to maximize the two objective functions presented in Section 6.2, namely the sum of the SNR values of the superchannel subchannels (i.e., Obj#1) and the lowest subchannel SNR value (i.e., Obj#2).

In the co-simulations, VPI takes as input the subchannel frequencies, performs the detailed transmission simulations and calculates the SNR values of the received subchannels. Instead, the MATLAB code implementing the finite differences subroutine first identifies which frequencies $\boldsymbol{\delta d}$ to probe to the superchannel, then passes them to VPI, as schematically represented in

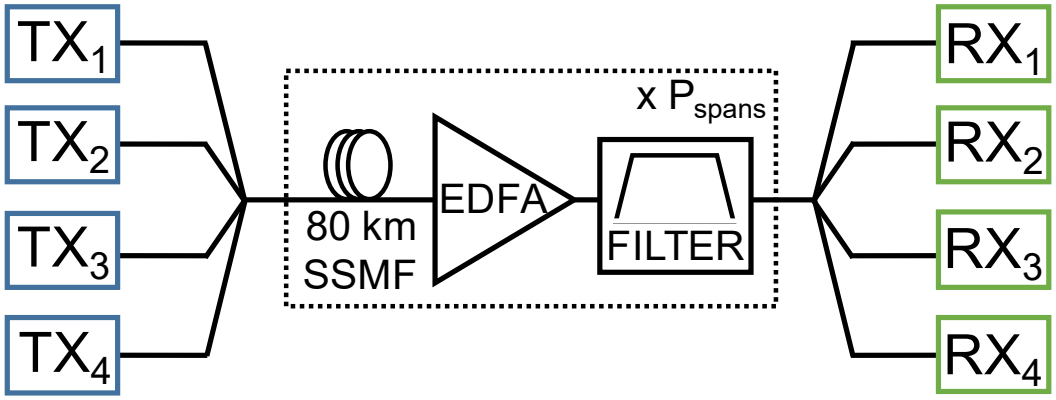

FIGURE 6.6. Schematic diagram of the VPI-implemented superchannel simulation setup. TX: Transmitter; SSMF: Standard Single-Mode Fiber; EDFA: Erbium-Doped Fiber Amplifier; RX: Receiver.

FIGURE 6.5, and finally receives from VPI the corresponding monitored SNR values $\mathbf{snr_{SC}}(\boldsymbol{\delta d_n})$. After a specific number of such probes (in our simulations, we had a minibatch of $M = 2$ subchannels), the optimization algorithm identifies their subgradients and uses them to calculate the following frequencies. This process is repeated until the optimization algorithm converges to the optimal solution, which without considering the monitoring stage, takes a time in the order of seconds. In addition, we kept track of the best-achieved solution and then stopped the algorithm once the estimated improvement was below a specific threshold. In particular, regarding the result presented below, we set an optimality tolerance of 0.2 dB.

To simulate a four-subchannel superchannel, we implemented in VPIphotonics the setup depicted in FIGURE 6.6. We generated four 32 GBd QPSK modulated signals that formed four quasi-Nyquist-WDM subchannels symmetrically centered around 193.1 THz. For their pulse shaping, we employed an RRC filter with roll-off factors equal to 0.1 and 0.15. We set the launch power of each transmitter laser at 0 dBm, which is the optimized value to guarantee a good compromise between linear and nonlinear effects in our simulations. As starting condition, we always considered the four signals equally spaced by 34.5 GHz, resulting in a quasi-Nyquist-WDM transmission, according to the definition we presented in Section 6.1 (i.e., a ratio between inter-channel distance and symbol rate between 1 and 1.2). Note that in addition to the equally distanced configuration, we also considered ten further random starting frequencies set, which were ±2 GHz shifted from the equally distanced setting. This was done in order to emulate soft failure scenarios. Then, we multiplexed the four subchannels and transmitted them in a loop composed by the cascade of an 80 km length SSMF, with 0.2 dB/km attenuation coefficient and 16.7 ps/nm/km dispersion coefficient, an EDFA with NF of 5.5 dB and gain tilt set to 0 dB/Hz, and a tuneable optical filter with

3.5th-order super-Gaussian TF. We set the 3 dB bandwidth of the optical filter with values ranging between 137.5 GHz and 200 GHz. To study the performance of the proposed solution on higher-order signal modulation formats, we also considered a 16QAM format. In addition, by varying the number of loops, we simulated a different number of optical spans. In particular, to simulate a B2B scenario, we removed the loop maintaining only a single optical filter. Moreover, before receiving the signal, we placed a dispersion-compensating optical fiber of length equal to the loop fiber length to fully compensate for the chromatic dispersion effect. Then, at the receiver side, we detected each of the four subchannels by tuning the frequency of four coherent receiver local oscillators. In [111], the optimum value for the receiver low-pass electric filter 3 dB bandwidth was found to be equal to half of the symbol rate. Accordingly, we considered 3 dB bandwidth values for the receiver filters equal to 16 GHz. Once detected, we monitored the SNR of each subchannel and sent all of them to the optimization algorithm. To obtain the measurements needed to derive the gradient, the algorithm probed several times the VPI setup. Then, based on the calculated gradient, it identified a new set of frequencies and fed it to VPI, triggering a new simulation setup / iterative optimization process iteration.

6.5 Results and Discussion

For the four-subchannels scenario discussed above, we considered nine different transmission cases corresponding to different parameter configurations. We summarized these nine cases in TABLE 6.1. In particular, we considered case 1 of TABLE 6.1 as the default case, with BW_{SC} equal to 137.5 GHz, roll-off factor a, for all the four transmitted signals, equal to 0.1,

TABLE 6.1. Simulation parameters of the nine considered cases.

# Case	Mod. Format	Roll-off	# Spans	# Filter	Bw [GHz]
1 (Def.)	QPSK	0.1	2	2	137.5
2	16-QAM	0.1	2	2	137.5
3	QPSK	0.1	2	2	150
4	QPSK	0.15	2	2	137.5
5	QPSK	0.1	10	5	137.5
6 (B2B)	QPSK	0.1	0	1	137.5
7 (B2B)	QPSK	0.1	0	1	200
8 (B2B)	QPSK	0.15	0	1	137.5
9 (B2B)	QPSK	0.15	0	1	200

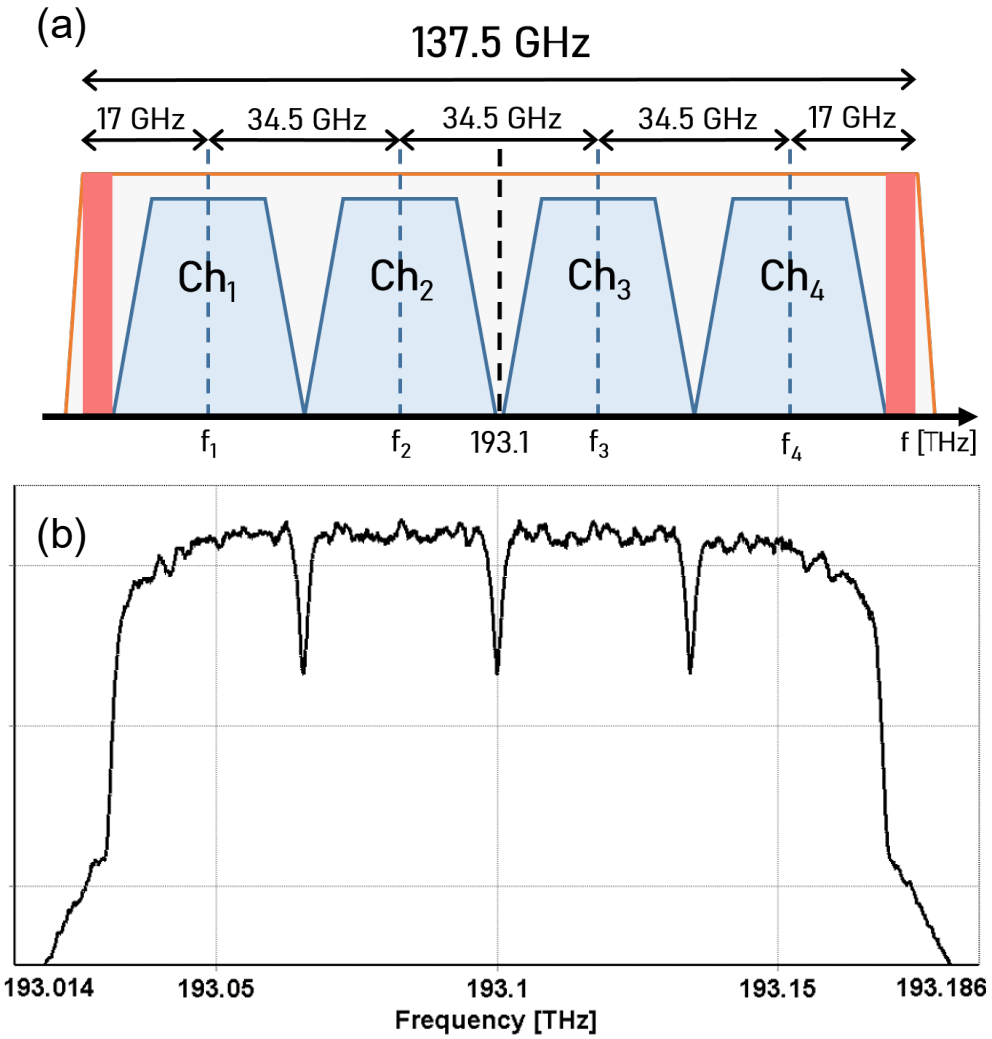

FIGURE 6.7. (a) Schematic representation of the considered default superchannel configuration with four equally spaced subchannels. (b) VPI-generated superchannel optical spectrum for the default case.

and two loops (i.e. twice the cascade of an 80 km length SSMF, an EDFA, and a superchannel filter). In addition, for this default case, we assumed as a starting condition the subchannels to be equally spaced, i.e. $\mathbf{d}_{\mathbf{eq\ dis}}$=[17, 34.5, 34.5, 34.5, 17] GHz. We depict a schematic representation of such default configuration in FIGURE 6.7(a), whereas, in FIGURE 6.7(b), we show the simulated superchannel spectrum acquired before the reception stage. Considering such default parameters, the received four subchannel SNR values were equal to $\mathbf{snr}_{\mathbf{SC}}(\mathbf{d}_{\mathbf{eq\ dis}})$=[13.55, 16.86, 16.91, 13.15] dB. This resulted in a superchannel total SNR value of 60.47 dB, with 13.15 dB being the minimum among the four subchannel SNR values. After the integration with the optimization algorithm presented in the previous sections, we found the optimized sets of distances $\mathbf{d}^*_{\mathbf{obj1}}$ and $\mathbf{d}^*_{\mathbf{obj2}}$ related to the two considered objective functions. In particular, the set $\mathbf{d}^*_{\mathbf{obj1}}$=[18.25, 33.25, 33.25,

33, 19.75] GHz corresponded to the solution that maximized the superchannel total SNR value, found with a step size of 0.25 GHz. The superchannel total SNR value associated with such configuration was 61.92 dB, yielding a 1.45 dB improvement with respect to the equally spaced initial configuration represented by $\mathbf{d}_{\mathbf{eq\ dis}}$. Regarding the maximization of the subchannel minimum SNR, again considering a step size of 0.25 GHz, we obtained the set of distances $\mathbf{d}^*_{\mathbf{obj2}}$=[19, 33, 31.75, 32.75, 21] GHz and its corresponding set of SNR values $\mathbf{snr}^*_{\mathbf{obj2}}(\mathbf{d}^*_{\mathbf{obj2}})$=[14.34, 14.58, 14.37, 14.38] dB. We see that the algorithm flattened the SNR values of all the subchannels. This is expected in such an optimization objective, where the algorithm iteratively tries to improve the minimum, eventually bringing all the subchannels to the same level. In particular, the optimal minimum SNR value was 14.34 dB. Such a result translated in a 1.19 dB improvement with respect to equally spaced initial distances $\mathbf{d}_{\mathbf{eq\ dis}}$. The total number of algorithm iterations needed to obtain the previous solutions were 23 for Obj#1 and 9 concerning Obj#2. Note that the number of iterations is crucial because it affects (linearly) the number of probes and monitoring calls, as we showed in Section 6.3.

The evolution of the two considered objectives as a function of the iteration number for the default case and equidistant subchannels as a starting point is shown in FIGURE 6.8. As introduced in Section 2.2.3, from FIGURE 6.8, we can observe that the stochastic subgradient method is not a descent method (like gradient descent), and therefore slight deteriorations of the objective function within the path towards the optimum are tolerated. Note that in this figure, we observe the negative of a convex/descent problem; we search for the

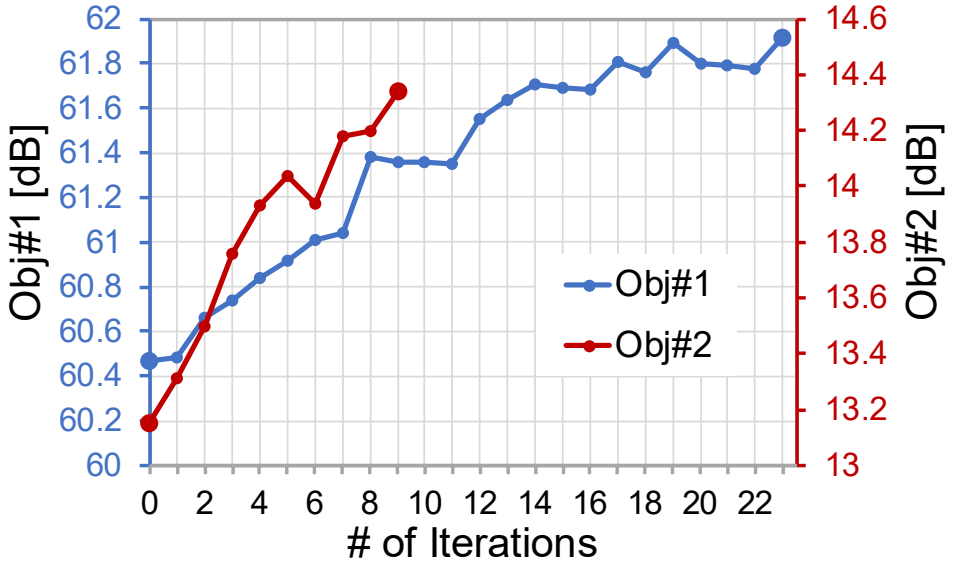

FIGURE 6.8. Evolution of the two considered objectives as a function of the iteration number.

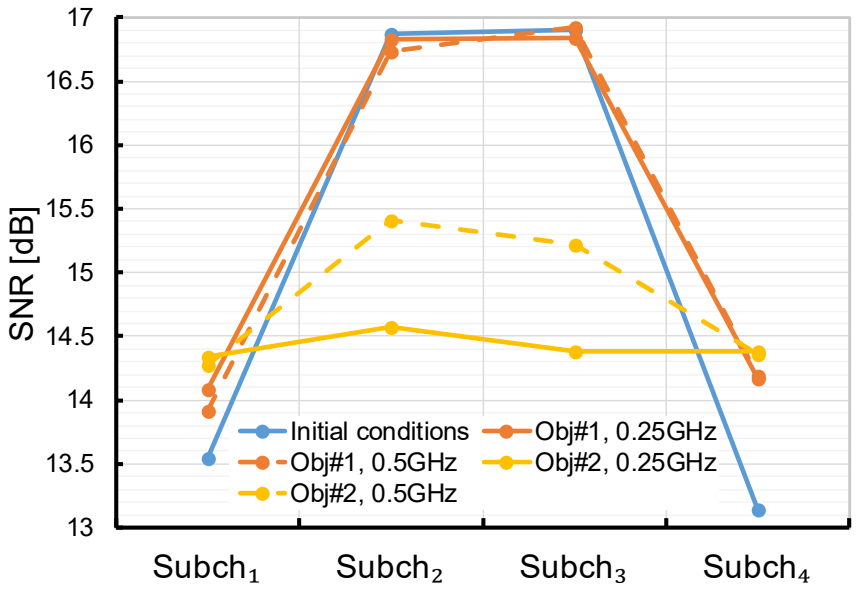

FIGURE 6.9. SNR values of the four subchannels considering the default case parameters. The blue line represents the SNR values obtained with the equally spaced subchannels, whereas the brown and yellow lines represent those obtained after the optimization of Obj#1 and Obj#2, respectively. For both the objectives, the solid and dashed lines represent a frequency step size of 0.25 and 0.5 GHz, respectively.

maximum, so the objective ascends. Also, note that the number of iterations depends on the considered starting point. In FIGURE 6.9, we report the complete results related to the default case, graphically representing the SNR values of the four subchannels before and after the optimization process for two different frequency step sizes. Concerning the results represented in FIGURE 6.9 for Obj#2, it is worth noting that even though the curve representing the 0.5 GHz step size seems to reach a higher total SNR value than the 0.25 GHz one, in Obj#2, we focused on the maximization of the subchannel minimum SNR value, for which 0.25 GHz gave better results.

The results related to the nine considered cases of TABLE 6.1 are summarized respectively in TABLE 6.2 for the maximization of Obj#1 and TABLE 6.3 for the maximization of Obj#2. With respect to the default case (case 1), in cases 2 to 5, we considered variations of the: modulation format (case 2), superchannel filter bandwidth (case 3), roll-off factor (case 4), and span number (case 5). Regarding this last configuration, increasing the number of spans also implied the presence of NLI and other wavelength-dependent impairments, such as the gain ripple. In this case, employing the proposed algorithm, we improved Obj#1 and Obj#2 of 1.99 dB and 1.91 dB, respectively, with respect to the scenario with equally distanced subchannels. More in general, the obtained results reported in TABLE 6.2 and TABLE 6.3 show how, independently from the variation of the aforementioned parameters, the

TABLE 6.2. Obj#1 optimization results for the four subchannels scenario.

Case	Total SNR start [dB]	Step size [GHz]	# Iter.	Total SNR Obj#1 [dB]	Brute-force		Save wrt worst [dB]	Save wrt start [dB]
					Best [dB]	Worst [dB]		
1	60.47	0.25	23	61.92	61.92	52.03	9.89	1.45
		0.50	7	61.77				
2	62.60	0.25	17	64.52	64.52	48.40	16.12	1.92
		0.50	10	64.24				
3	61.61	0.25	43	64.59	65.12	54.22	10.90	3.51
		0.50	24	65.12				
4	61.19	0.25	9	61.92	61.92	50.85	11.07	0.73
		0.50	12	61.83				
5	34.25	0.25	24	36.24	36.24	31.16	5.08	1.99
		0.50	22	36.23				
6	85.01	0.25	31	86.46	86.47	62.19	24.28	1.45
		0.50	16	85.31				
7	93.16	0.25	36	93.30	93.31	64.32	28.99	0.15
		0.50	19	93.31				
8	83.67	0.25	47	83.94	84.15	64.49	19.66	0.48
		0.50	21	84.15				
9	91.45	0.25	43	91.31	91.45	75.46	15.99	0.00
		0.50	21	91.25				

proposed optimization solution was always able to improve the considered objectives. In addition, as per the default case, we ran simulations considering ten different random starting frequencies for cases 2, 3, 4, and 5. In all such cases, the algorithm always converged on the same optimal solution.

TABLE 6.1 cases from 6 to 9 represent four different B2B scenarios. We considered them to understand the role played by the filtering and the crosstalk effects. As we showed in FIGURE 6.2, when the distances between subchannels ranged between 34.5 ±2 GHz, the crosstalk from the two adjacent subchannels is negligible. In particular, in cases 7 and 9, where we set a very broad filter 3 dB bandwidth (i.e., 200 GHz), almost no improvement was observed after the optimization with respect to the equidistant superchannel configuration. This happens because our starting conditions (i.e., 34.5 GHz distanced subchannels) already guaranteed an impairment-free transmission, and the optimization increases further all the distances, but with negligible improvement. As expected, higher roll-off factors (e.g., case 9) implied higher crosstalk effects between the signals. On the other hand, when considering B2B and tighter filter bandwidth (i.e., cases 6 and 8), crosstalk- and filter-related penalties appears, and therefore our solution can improve the considered

TABLE 6.3. Obj#2 optimization results for the four subchannels scenario.

Case	Min SNR start [dB]	Step size [GHz]	# Iter.	Min SNR Obj#2 [dB]	Brute-force		Save wrt worst [dB]	Save wrt start [dB]
					Best [dB]	Worst [dB]		
1	13.15	0.25	9	14.34	14.34	10.61	3.73	1.19
		0.50	4	14.28				
2	13.63	0.25	8	14.86	14.86	10.03	4.83	1.23
		0.50	4	14.82				
3	13.74	0.25	23	15.76	15.78	10.61	5.17	2.04
		0.50	10	15.78				
4	13.21	0.25	8	14.17	14.18	10.50	3.68	0.97
		0.50	3	14.18				
5	6.19	0.25	17	8.10	8.10	6.60	1.50	1.91
		0.50	6	8.00				
6	18.97	0.25	10	20.08	20.08	12.84	7.24	1.11
		0.50	5	20.03				
7	23.04	0.25	26	23.13	23.19	15.89	7.30	0.15
		0.50	16	23.19				
8	18.68	0.25	10	19.41	19.41	13.58	5.83	0.73
		0.50	5	19.30				
9	22.55	0.25	26	22.58	22.66	18.64	4.02	0.11
		0.50	24	22.66				

objectives. As expected, case 8, where we considered a higher roll-off factor, has a lower initial (average) SNR value than case 6. However, its improvement with respect to the equidistant configuration is relatively lower since both crosstalk and filtering effects are strong, and not enough optimization space is present.

As presented in Section 6.3, the proposed solution approaches the optimum within a specific accuracy interval. Hence, we might reach results closer to the global optimum using coarser frequency step sizes than finer ones, which, theoretically, should be more precise. In particular, this happened for Obj#2, in cases 3, 4, 7, and 9, as shown in TABLE 6.3.

In order to provide a benchmark for the obtained results, in TABLE 6.2 and TABLE 6.3, we also report the best and the worst SNR values obtained employing a brute-force approach. In particular, we considered a ±2 GHz interval around the optimal solutions found by the algorithm (i.e., $\mathbf{D}^*_{\mathbf{obj1}}$ and $\mathbf{D}^*_{\mathbf{obj2}}$) and collected the SNR values related to all the possible distance combinations in such interval, with a frequency step size of $f^{BF}_{step} = 1$ GHz. For each considered case listed in TABLE 6.2 and TABLE 6.3, the total number of

brute-force iterations was 5^N. A smaller step, e.g., $f_{step}^{BF} = 0.25$ GHz, would result in a huge number of scenarios and prohibitive simulation time. In a real network scenario, the worst SNR value found in this frequency range could represent the result of a soft failure occurring while the network is operating (e.g., ±2 GHz drift of each laser). Usually, these situations are addressed foreseeing some margins for the SNR when planning the connection and selecting an adequate modulation format. Instead, with the proposed solution, we remove such need. In particular, regarding the default case, with respect to such a soft failure scenario, we achieved improvements of 9.89 dB and 3.73 dB for the two considered objectives, respectively.

From TABLE 6.2 and TABLE 6.3, we can see that our proposed method reached almost always the optimal SNR values identified by the brute-force approach. However, in both cases where this did not happen (i.e., cases 6 and 9 of TABLE 6.2), the optimal found solutions almost equalized the best ones. In particular, the solution related to case 6 missed the best one by 0.01 dB, whereas case 9 missed it by 0.14 dB. Note that such a brute-force approach is unsuitable to be used in a real network, mainly for two reasons. On the one hand, because it has an execution time that is exponential to the subchannel number. In fact, considering an explored spectrum range $\Delta F = 4$ GHz (i.e., ±2 GHz) and a $f_{step}^{BF} = 1$ GHz, we would have to configure and monitor the network $[1 + (\Delta F / f_{step}^{BF})]^N$ times. Assuming a monitoring time of 5 minutes, this would take more than two days to complete, even for four subchannels. This increases even further for finer granularity. On the other hand, because it operates blindly, that is, without considering the subchannel SNR values and thus potentially making unfeasible some of the connections. Comparing that with a few tens of monitoring calls (i.e., IM, with $M = 2$) used by the proposed solution gives a clear benefit of employing our approach instead of the brute-force one.

To further assess the effectiveness of the proposed method, we also integrated the optimization algorithm in simulations with six, eight, and ten subchannels. Apart from the number of transmitters and receivers, we kept the rest of the setup precisely as shown in FIGURE 6.6. For such subchannel scenarios (i.e., six, eight and ten subchannels), we considered only one transmission case, similar to the default one of the four subchannels configuration (i.e., case 1 in TABLE 6.1). In particular, for the scenario with six subchannels (N=6), we transmitted six 32 GBd QPSK equally spaced signals, with roll-off factor equal to 0.1, in a loop of two spans, filter 3 dB bandwidth of 200 GHz, and a minibatch number M=3. Similarly, for the scenarios with eight and ten subchannels (i.e., N=8 and N=10, respectively), we assumed the same configuration as before, but a filter 3 dB bandwidth equal to 275 GHz and

340 GHz, respectively, and $M = N/2$. Note that we used a frequency step size of 0.5 GHz for all the considered configurations and objectives.

As per the scenario with four, also with six, eight, and ten subchannels, our achieved results showed an improvement of the considered figure of merit with respect to the starting conditions for both the objectives. In particular, with respect to the equally spaced scenario and concerning Obj#1, we improved the superchannel total SNR for the six, eight, and ten subchannels cases of 1.72 dB, 1.44 dB, and 0.55 dB, respectively. In addition, regarding Obj#2, such improvements were equal to 1.67 dB, 1.33 dB, and 1.19 dB, respectively. We summarize the obtained results related to the two objectives in TABLE 6.4 and TABLE 6.5, respectively. There, for the sake of comparison, we also report the corresponding case with four subchannels. As per the previous results, also in this case, it is worth reminding that the number of iterations required to reach the optimal solution depends on the considered subchannel starting positions. For instance, referring to TABLE 6.4 results, we can see that the iterations needed to reach Obj#2 optimum in the case of six subchannels were lower than those of the eight subchannels scenario. Furthermore, from TABLE 6.4 results, we can also notice a reduction of the SNR improvements with respect to the equidistant superchannel, which grows with the number of considered subchannels N. We obtain SNR improvement with respect to the equidistant

TABLE 6.4. Obj#1 optimization results for the considered subchannel scenarios.

Case	Total SNR start [dB]	Step size [GHz]	# Iter.	Total SNR Obj#2 [dB]	Brute-force		Save wrt worst [dB]	Save wrt start [dB]
					Best [dB]	Worst [dB]		
4 subchs., 137.5 GHz, 2 spans, 0.1 r.o.	60.47	0.50	7	61.77	61.92	52.03	9.74	1.30
6 subchs., 200 GHz, 2 spans, 0.1 r.o.	91.86	0.50	11	93.58	93.65	76.84	16.74	1.72
8 subchs., 275 GHz, 2 spans, 0.1 r.o.	126.93	0.50	14	128.37	128.54	107.54	20.83	1.44
10 subchs., 340 GHz, 2 spans, 0.1 r.o.	159.29	0.50	10	159.84	159.84	141.67	18.17	0.55

TABLE 6.5. Obj#2 optimization results for the considered subchannel scenarios.

Case	Min SNR start [dB]	Step size [GHz]	# Iter.	Min SNR Obj#2 [dB]	Brute-force		Save wrt worst [dB]	Save wrt start [dB]
					Best [dB]	Worst [dB]		
4 subchs., 137.5 GHz, 2 spans, 0.1 r.o.	13.15	0.50	4	14.28	14.34	10.61	3.67	1.13
6 subchs., 200 GHz, 2 spans, 0.1 r.o.	11.90	0.50	16	13.57	13.57	9.53	4.04	1.67
8 subchs., 275 GHz, 2 spans, 0.1 r.o.	13.07	0.50	11	14.40	14.40	9.84	4.56	1.33
10 subchs., 340 GHz, 2 spans, 0.1 r.o.	13.11	0.50	14	14.30	14.44	9.81	4.49	1.19

configuration, by redistributing the spectral space initially placed between the subchannels to those between the outer subchannels and the filter sides, and vice versa. Therefore, as the number of subchannels increases, the effect of this optimization will become less significant. Note, however, that maintaining equidistant might be a challenging task due to laser drifts and filter detuning (i.e. soft failures), while the proposed method finds the optimum in an operating network under any such issue. Moreover, as shown in TABLE 6.4, it is worth underlining that the improvements with respect to a soft failure scenario remain stable for all the considered multi-subchannel scenarios, making our solution effective even in those cases.

Given the high amount of considered subchannels N, we ran the brute-force approach around the optimal found solution only within an interval of ±1 GHz. This resulted in a total number of brute-force iterations equal to 3^N. Such reduction was deemed necessary to complete the simulations in a reasonable time (i.e., within hours). This is also a further indication of the infeasibility of the brute-force approach for real scenarios. As per the previous four subchannel scenario, even in these further cases, our optimal found solutions almost always corresponded to the brute-force found optima. In particular, when this did not happen, we had the following accuracies: in scenarios with six and eight subchannels, for Obj#1, our optimal solutions were 0.07 dB and

0.17 dB away from the brute-force found optima, respectively. Instead, in the ten subchannels scenario, for Obj#2, our solution was 0.14 dB far from the optimum found with the brute-force approach.

Concerning real network scenarios (i.e., considering other WDM channels), we expect the SNR improvement amounts brought by our solution to be affected by the increased channel complexity and the growing NLIs. However, we believe this reduction to be of a specific amount, and therefore it could be modelled as a constant reduction in each SNR value, not affecting our process. Further investigations are needed to clarify such aspects more exhaustively.

6.6 Conclusions

In this chapter, we presented a closed control loop process to optimize the spectral distances of the superchannel subchannels. We observed that the problems of maximizing the values of the superchannel total SNR and the subchannel minimum SNR are concave around the operation point. We developed a solution that probes and monitors the network and then uses the stochastic subgradient method to optimize the chosen objectives. We elaborate such an approach as to be robust to monitoring errors and noise. To validate it, we implemented several co-simulations in VPI and MATLAB.

The results we obtained showed excellent performance for all the considered configurations, independently from the number of subchannels, signal characteristics, superchannel filter bandwidth values, modulation format, or starting frequencies. In particular, we considered superchannels composed of four, six, eight, and ten subchannels, with uniform characteristics, such as symbol rate and roll-off values. We compared the optimal solutions found through our method with the optima retrieved using an exhaustive search approach. In all the assumed scenarios, we were able to approach such values, improving both the chosen objectives with respect to the case where equal distances between the superchannel subchannels were considered. In particular, for a uniform four-subchannel superchannel with a 32 GBd symbol rate, 0.1 roll-off factor, and a superfilter bandwidth of 137.5 GHz, we improved the superchannel total SNR and the subchannel minimum SNR values of 1.45 dB and 1.19 dB, respectively. Furthermore, for a soft failure scenario in which a ±2 GHz subchannel frequency drift around the optimum was considered, we improved the two objectives of 9.89 dB and 3.73 dB, respectively.

Chapter 7

Conclusions and Future Works

In this final chapter, which conclude the thesis, we summarize the main contributions presented within this dissertation. In addition, we also provide possible directions for the development of additional future works.

This chapter is organized as follows. In Section 7.1, we summarize the achieved results, whereas in Section 7.2, we identify possible future lines of investigation.

7.1 Summary of the Results

In this thesis, we addressed three main problems which revolve around the optimization of optical networks through the use of OPM. Firstly, we investigated a cost-effective method for the spectral acquisition of optical signals. Secondly, we proposed two strategies to optimize the placement of the monitoring solutions within the optical network and accordingly developed two spectral processing techniques to retrieve network-related parameters from the acquired optical spectra. Finally, we presented a potential application for the monitored network information, developing a solution for optimizing the superchannel transmission.

More in detail, this thesis is composed of a preliminary part and a technical part. Excluding Chapter 1, where we provided objectives, methodology and the outline of this dissertation, the preliminary part includes Chapter 2 and Chapter 3. There, we introduced the mathematical foundations as well as the context of this work and presented a study of the state of the art literature for the investigations carried out in this thesis. Regarding the technical part, in the following, we provide a summary of each chapter contributions.

In Chapter 4, we presented the theoretical study of a spectral monitoring method to estimate the optical signal high-resolution spectra in a simple and flexible manner. Such a proposed technique relies on detecting different

spectral slices of an optical signal spectrum by means of a coherent receiver enhanced with adequate DSP. Nevertheless, the proposed scheme comes with some requirements, among which is the need for a broadly tuneable laser, and it is mainly limited by the stability of the optical sources employed to detect the spectral slices.

In Chapter 5, we proposed two monitoring placement scenarios, according to the position of the considered optical monitors. In particular, we identified egress and ingress monitoring scenarios. The former considers optical monitors placed after the egress ports of the optical nodes, whereas the latter envisions the monitors placed before the ingress ports of such optical nodes. Both the presented monitoring strategies allow retrieving directly from the acquired optical spectra specific network information. In particular, the filter parameters can be easily measured from the egress-related spectra, whereas those related to the signal, such as the ASE noise/OSNR value, cannot because of the filtering effect. On the contrary, from the spectral data acquired within the ingress placement scenario, the ASE noise/OSNR can be easily estimated, but the filter central frequency and the filter 3/6 dB bandwidth cannot. Hence, along with the two placement scenarios, we propose two spectral processing techniques to enhance the monitoring strategies retrieving the missing parameters. The first one, which applies in egress monitoring placement scenarios, leverages two ML regression algorithms, namely SVM and GPR, to estimate in-band the ASE noise values from the optical spectra. The second technique, which applies to the ingress monitoring placement scenarios, leverages the curve fitting principle to retrieve the filter TFs and the related filter parameters. We experimentally validated both the proposed methods and finally compared them in terms of estimation reduction of the SNR penalty introduced by an optical filter. We identify the ingress placement strategy and its related approach as the most promising solutions.

Finally, in Chapter 6, we proposed a closed control loop-based process to optimize the frequency spacing of the superchannel subchannels. The solution we developed probes and monitors the network, and optimize specific chosen objectives leveraging the stochastic subgradient method. In particular, we aim to maximize the total superchannel SNR value and the minimum superchannel subchannel SNR value. The former goal is directly related to maximizing the total superchannel capacity, whereas the latter makes all the subchannels feasible within the considered superchannel. We validate the proposed solution by simulating the superchannel transmission in VPI and integrating it with the optimization algorithm developed in MATLAB. We assumed scenarios with different subchannel numbers, signal characteristics, and starting frequencies. Our optimization process shows an improvement of the two considered objectives for all the configurations mentioned above.

7.2 Future Works

OPM and its role in optimizing future optical networks will certainly represent hot research topics for the upcoming years. During the investigations we carried out to develop the methods and the solutions presented within this dissertation, we identified several open issues that deserve further study. In the following, we list such potential lines of investigation.

- Comparison of the centralized and distributed management architectures outlined in Section 3.4 (in terms of implementation and communication overhead and monitoring / feature extraction quality that can be achieved).

- Simulation and experimental validation of the front-end optical spectrum estimation scheme presented and studied in Chapter 4 and investigation of novel ways to overcome the limitation introduced by the optical source stability.

- The two monitoring placement strategies and the corresponding processing solutions presented in Chapter 5 could be further evaluated within the QoT estimation, considering NLI effects deriving from multiple channels transmission.

- In addition to the two monitoring placement scenarios that we identified in Chapter 5, a third potential strategy is the sparse one. In such a scenario, some of the network nodes are equipped with monitors at their egress ports, some others with monitors at their ingress ports, and some others again are not monitored at all. According to this monitoring configuration, adequate spectral processing solutions should be developed and validated.

- Extension of the solution proposed in Chapter 6 to optimize the superchannel subchannel frequencies to address nonuniform superchannels, i.e., superchannel transmitting signals with different characteristics, such as different modulation formats or roll-off factors.

- Extension of the optimization method proposed in Chapter 6 to a network level to reduce the interference and align the channels of the whole network.

- Development of a new superchannel optimization algorithm for scenarios where monitoring probes do not represent the best choice due

to the high overhead of configuring and monitoring the network that such approach requires and the issues related to the monitoring noise and the high probing time. A candidate solution to replace the monitoring probes strategy is represented by PLMs or QoT estimators.

Bibliography

[1] Cisco, “Cisco annual internet report (2018-2023),” 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf

[2] Y. Miyamoto and R. Kawamura, “Space division multiplexing optical transmission technology to support the evolution of high-capacity optical transport networks,” *NTT Technical Review*, vol. 5, no. 6, pp. 1-7, June 2017.

[3] S. Perrin, “Open and disaggregated packet and optical networks,” 2021. [Online]. Available: https://cdn.brandfolder.io/D8DI15S7/as/383s7z9m94rh8hm3ksqgtks/HR_Open_and_Disaggregated_Packet_and_Optical_Networks_February_2021.pdf

[4] E. Riccardi, P. Gunning, Ó. González de Dios, M. Quagliotti, V. López, and A. Lord, “Operator view on the introduction of white boxes into optical networks,” *IEEE/OSA Journal of Lightwave Technology*, vol. 36, no. 15, pp. 3062-3072, August 2018.

[5] N. Sambo, K. Christodoulopoulos, N. Argyris, P. Giardina, C. Delezoide, D. Roccato, A. Percelsi, R. Morro, A. Sgambelluri, A. Kretsis, G. Kanakis, G. Bernini, E. Varvarigos, and P. Castoldi, "Field trial: demonstrating automatic reconfiguration of optical networks based on finite state machine," *IEEE/OSA Journal of Lightwave Technology*, vol. 37, no. 16, pp. 4090-4097, August 2019.

[6] K. Christodoulopoulos, C. Delezoide, N. Sambo, A. Kretsis, I. Sartzetakis, A. Sgambelluri, N. Argyris, G. Kanakis, P. Giardina, G. Bernini, D. Roccato, A. Percelsi, R. Morro, H. Avramopoulos, P. Castoldi, P. Layec, and S. Bigo, "Toward efficient, reliable, and autonomous optical networks: the ORCHESTRA solution," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 11, no. 9, pp. C10-C24, September 2019.

[7] Y. Pointurier, "Design of low-margin optical networks," in *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A9-A17, January 2017.

[8] ITU-G.694.1, “Spectral grids for WDM application: DWDM frequency grid,” October 2020.

[9] C. C. K. Chan, *Optical Performance Monitoring*, Elsevier, USA, 2010.

[10] Z. Dong, F. N. Khan, Q. Sui, K. Zhong, C. Lu, and A. P. T. Lau, “Optical performance monitoring: a review of current and future technologies,” *IEEE/OSA Journal of Lightwave Technology*, vol. 34, no. 2, pp. 525–543, January 2016.

[11] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383-1408, Second quarter 2019.

[12] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," *IEEE/OSA Journal of Lightwave Technology*, vol. 37, no. 2, pp. 493-516, January 2019.

[13] Matlab, ver. R2018b. [Online]. Available: https://mathworks.com/products/new_products/release2018b

[14] VPIphotonics, 2021. [Online]. Available: https://vpiphotonics.com

[15] [Online]. Available: http://networks.cttc.es/ons/adrenaline

[16] D. Rafique and L. Velasco, "Machine learning for network automation: overview, architecture, and applications [Invited Tutorial]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D126-D143, October 2018.

[17] T. M. Mitchell, Machine Learning, McGraw-Hill, 1997.

[18] C. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag, USA, 2006.

[19] F. Locatelli, K. Christodoulopoulos, J. M. Fàbrega, M. Svaluto Moreolo, and S. Spadaro, “Machine learning-based in-band OSNR estimation from optical spectra,” *IEEE Photonics Technology Letters*, vol. 31, no. 24, pp. 1929-1932, December 2019.

[20] D. Wang, M. Zhang, Z. Li, J. Li, M. Fu, Y. Cui, and X. Chen, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photonics Technology Letters*, vol. 29, no. 19, pp. 1667-1670, October 2017.

[21] J. Thrane, J. Wass, M. Piels, J. C. M. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals," *IEEE/OSA Journal of Lightwave Technology*, vol. 35, no. 4, pp. 868-875, February 2017.

[22] A. P. Vela, B. Shariati, M. Ruiz, F. Cugini, A. Castro, H. Lu, R. Proietti, J. Comellas, P. Castoldi, S. J. B. Yoo, and L. Velasco, "Soft failure localization during commissioning testing and lightpath operation," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 1, pp. A27-A36, January 2018.

[23] S. Troia, R. Alvizu, Y. Zhou, G. Maier, and A. Pattavina, "Deep learning-based traffic prediction for network optimization," in *International Conference on Transparent Optical Networks (ICTON), 2018*.

[24] F. Morales, M. Ruiz, L. Gifre, L. M. Contreras, V. Lopez, and L. Velasco, "Virtual network topology adaptability based on data analytics for traffic prediction," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A35-A45, January 2017.

[25] A. Mahajan, K. Christodoulopoulos, R. Martínez, S. Spadaro, and R. Muñoz, "Modeling EDFA gain ripple and filter penalties with machine learning for accurate QoT estimation," *IEEE/OSA Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2616-2629, May 2020.

[26] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, “Data-analytics-based optical performance monitoring technique for optical transport networks,” in *Optical Fiber Communications Conference (OFC)*, 2018.

[27] T. B. Anderson, A. Kowalczyk, K. Clarke, S. D. Dods, D. Hewitt, and J. C. Li, "Multi impairment monitoring for optical networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3729-3736, August 2009.

[28] F. N. Khan, T. S. R. Shen, Y. Zhou, A. P. T. Lau, and C. Lu, "Optical performance monitoring using artificial neural networks trained with empirical moments of asynchronously sampled signal amplitudes," *IEEE Photonics Technology Letters*, vol. 24, no. 12, pp. 982-984, June 2012.

[29] S. G. Petridou, P. G. Sarigiannidis, G. I. Papadimitriou, and A. S. Pomportsis, "On the use of clustering algorithms for message scheduling in WDM star networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 26, no. 17, pp. 2999-3010, September 2008.

[30] S. Troia, Gao Sheng, R. Alvizu, G. A. Maier, and A. Pattavina, "Identification of tidal-traffic patterns in metro-area mobile networks via Matrix Factorization based model," in *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2017.

[31] J. Zhang, W. Chen, M. Gao, B. Chen, and G. Shen, "Novel low-complexity fully-blind density-centroid — tracking equalizer for 64-QAM coherent optical communication systems," in *Optical Fiber Communications Conference (OFC)*, 2018.

[32] Y. V. Kiran, T. Venkatesh, and C. S. Ram Murthy, "A reinforcement learning framework for path selection and wavelength selection in optical burst switched networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 9, pp. 18-26, December 2007.

[33] N. Fernández, R. J. Durán, I. de Miguel, N. Merayo, J. C. Aguado, P. Fernández, T. Jiménez, I. Rodríguez, D. Sánchez, R. M. Lorenzo, E. J. Abril, M. Angelou, and I. Tomkos, "Survivable and impairment-aware virtual topologies for reconfigurable optical networks: A cognitive approach," in *International Congress on Ultra-Modern Telecommunications and Control Systems*, 2012.

[34] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, USA, 1995.

[35] N. Christianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, UK, 2000.

[36] R. Berwick, *An Idiot Guide to Support Vector Machines (SVMs)*, notes for 6.034 Artificial Intelligence, Massachusetts Institute of Technology (MIT), 2011. [Online]. Available: http://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf

[37] M. S. Andersen, J. Dahl, and L. Vandenberghe, *CVXOPT: A Python Package for Convex Optimization*. [Online]. Available: http://cvxopt.org.

[38] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, USA, 2006. [Online]. Available: http://krasserm.github.io/2020/11/04/gaussian-processes-classification

[39] M. Krasser, *Gaussian Processes for Classification*, 2020. [Online]. Available: http://krasserm.github.io/2018/03/19/gaussian-processes/

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, UK, 2009. [Online]. Available: https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

[41] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer, 1985.

[42] S. Boyd, L. Xiao, and A. Mutapcic, *Subgradient Methods*, Notes for EE3920, Stanford University, 2003. [Online]. Available: https://web.stanford.edu/class/ee3920/subgrad_method.pdf

[43] N. Z. Shor, *Nondifferentiable Optimization and Polynomial Problems,* Springer, USA, 1998.

[44] S. Boyd, A. Mutapcic, and J. Duchi, *Stochastic Subgradient Methods*, Notes for EE364b, Stanford University, 2018. [Online]. Available: https://web.stanford.edu/class/ee364b/lectures/stoch_subgrad_notes.pdf

[45] Perrin, “Bringing disaggregation to transport networks,” Fujitsu, 2015. [Online]. Available: https://www.fujitsu.com/caribbean /Images/HR-Fujitsu%20Disaggregation-WP.pdf

[46] Infinera, “Future-proof open line systems for cloud scale networks,” 2017. [Online]. Available: https://www.infinera.com/wp-content/uploads/infinera-wp-Future-proof-Open-Line-Systems-Cloud-Scale-Networks.pdf

[47] Coriant, “The case for open line systems,” 2017. [Online]. Available: http://photonics-complete.eu/wp-content/uploads/2018/01/Coriant_WP _The_Case_for_Open_Line_Systems.pdf

[48] H. Adams, “Trends in metro optical networks,” IHS Markit, 2018. [Online]. Available: https://www.juniper.net/assets/us/en/local/pdf/industry-reports/2000696-en.pdf

[49] J. Santos, N. Costa, and J. Pedro, "On the impact of deploying optical transport networks using disaggregated line systems," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 1, pp. A60-A68, January 2018.

[50] M. Filer, J. Gaudette, M. Ghobadi, R. Mahajan, T. Issenhuth, B. Klinkers, and J. Cox, “Elastic optical networking in the Microsoft cloud,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 7, pp. A45-A54, July 2016.

[51] OpenROADM, “Open ROADM v2 general whitepaper,” 2018. [Online]. Available: https://0201.nccdn.net/4_2/000/000/05e/0e7/Open-ROADM-whitepaper-v2_2.pdf

[52] OpenConfig, “Data models and API,” 2016. [Online]. Available: https://www.openconfig.net/projects/models/

[53] Telecom Infra Project (TIP). [Online]. Available: https://www.telecominfraproject.com

[54] L. Nadal, J. M. Fàbrega, M. Svaluto Moreolo, R. Casellas, R. Muñoz, L. Rodríguez, R. Vilalta, F. J. Vílchez, and R. Martínez, "SDN-enabled sliceable transceivers in disaggregated optical networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 37, no. 24, pp. 6054-6062, December 2019.

[55] Infinera, “Operator strategies for disaggregation in 5G transport network,” 2019. [Online]. Available: https://www.infinera.com/wp-

content/uploads/Operator-Strategies-for-Disaggregation-in-5G-Transport-Networks.pdf

[56] L. Velasco, A. Sgambelluri, R. Casellas, L. Gifre, J. L. Izquierdo-Zaragoza, F. Fresi, F. Paolucci, R. Martínez, and E. Riccardi, "Building autonomic optical whitebox-based networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 36, no. 15, pp. 3097-3104, August 2018.

[57] N. Sambo, K. Christodoulopoulos, N. Argyris, P. Giardina, C. Delezoide, A. Sgambelluri, A. Kretsis, G. Kanakis, F. Fresi, G. Bernini, H. Avramopoulos, E. Varvarigos, and P. Castoldi, "Experimental demonstration of fully disaggregated white box including different types of transponders and monitors, controlled by NETCONF and YANG," in *Optical Fiber Communications Conference (OFC)*, 2018.

[58] J. M. Fàbrega, M. Svaluto Moreolo, and L. Nadal, "Optical performance monitoring systems in disaggregated optical networks," in *International Conference on Transparent Optical Networks (ICTON)*, 2018.

[59] M. Björklund, "YANG – A data modeling language for the network configuration protocol (NETCONF)," IETF RFC 6020, October 2010. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc6020

[60] R. Enns, M. Björklund, A. Bierman, and J. Schönwälder, "Network configuration protocol NETCONF," IETF RFC 6241, June 2011. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc6241

[61] A. Giorgetti, A. Sgambelluri, R. Casellas, R. Morro, A. Campanella, and P. Castoldi, "Control of open and disaggregated transport networks using the Open Network Operating System (ONOS)," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 12, no. 2, pp. A171-A181, February 2020.

[62] D. Wang, Q. Sui, and Z. Li, "Toward universal optical performance monitoring for intelligent optical fiber communication networks," *IEEE Communications Magazine*, vol. 58, no. 9, pp. 54-59, September 2020.

[63] G. P. Agrawal, *Fiber-Optic Communication Systems*, 5th edition, Wiley, USA, 2021.

[64] B. Chomicz, *Planning Fiber Optic Networks*, The McGraw-Hill Companies, USA, 2009.

[65] D. Derickson, *Fiber Optic Test and Measurement*, Pearson, USA, 1998.

[66] IEC, “Optical signal-to-noise ratio measurement for dense wavelength-division multiplexed systems,” IEC 61280-2-9, 2009.

[67] IEC, “In-band optical signal-to-noise ratio (OSNR),” IEC TR 61282-12, 2016.

[68] A. Morea, J. Renaudier, T. Zami, A. Ghazisaeidi, and O. Bertran-Pardo, "Throughput comparison between 50-GHz and 37.5-GHz grid transparent networks [Invited]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 2, pp. A293-A300, February 2015.

[69] J. M. Fàbrega, M. Svaluto Moreolo, L. Martín, A. Chiadò Piat, E. Riccardi, D. Roccato, N. Sambo, F. Cugini, L. Potì, S. Yan, E. Hugues-Salas, D. Simeonidou, M. Gunkel, R. Palmer, S. Fedderwitz, D. Rafique, T. Rahman, H. de Waardt, and A. Napoli, "On the filter narrowing issues in elastic optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 7, pp. A23-A33, July 2016.

[70] J. H. Lee, H. Y. Choi, S. K. Shin, and Y. C. Chung, "A review of the polarization-nulling technique for monitoring optical-signal-to-noise ratio in dynamic WDM networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 24, no. 11, pp. 4162-4171, November 2006.

[71] X. Liu, Y. H. Kao, S. Chandrasekhar, I. Kang, S. Cabot, and L. L. Buhl, "OSNR monitoring method for OOK and DPSK based on optical delay interferometer," *IEEE Photonics Technology Letters*, vol. 19, no. 15, pp. 1172-1174, August 2007.

[72] S. K. Shin, K. J. Park, and Y. C. Chung, "A novel optical signal-to-noise ratio monitoring technique for WDM networks," in *Optical Fiber Communications Conference (OFC)*, 2000.

[73] C. J. Youn, S. K. Shin, K. J. Park, and Y. C. Chung, "OSNR monitoring technique based on high-frequency receiver noise," in *Asia-Pacific Optical and Wireless Communications Conference*, 2001.

[74] N. Hanik, A. Gladisch, C. Caspar, and B. Strebel, “Application of amplitude histograms to monitor performance of optical channels,” *Electronic Letters*, vol. 35, no. 5, March 1999.

[75] R. S. Luis, A. Teixeira, and P. Monteiro, "Optical signal-to-noise ratio estimation using reference asynchronous histograms," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 6, pp. 731-743, March 2009.

[76] X. Wu, J. A. Jargon, R. A. Skoog, L. Paraschis, and A. E. Willner, "Applications of artificial neural networks in optical performance monitoring," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3580-3589, August 2009.

[77] Y. Huang, Y. Chen, and J. Yu, "Optical performance monitoring of 56Gbps optical PAM4 signal using artificial neural networks," in *Asia Communications and Photonics Conference (ACP)*, 2017.

[78] T. S. R. Shen, K. Meng, A. P. T. Lau, and Z. Y. Dong, "Optical performance monitoring using artificial neural network trained with asynchronous amplitude histograms," *IEEE Photonics Technology Letters*, vol. 22, no. 22, pp. 1665-1667, November 2010.

[79] J. A. Jargon, X. Wu, and A. E. Willner, "Optical performance monitoring by use of artificial neural networks trained with parameters derived from delay-tap asynchronous sampling," in *Optical Fiber Communications Conference (OFC)*, 2009.

[80] T. Mrozek, "Simultaneous monitoring of chromatic dispersion and optical signal to noise ratio in optical network using asynchronous delay tap sampling and convolutional neural network (deep learning)," in *International Conference on Transparent Optical Networks (ICTON)*, 2018.

[81] M. C. Tan, F. N. Khan, W. H. Al-Arashi, Y. Zhou, and A. P. T. Lau, "Simultaneous optical performance monitoring and modulation format/bit-rate identification using principal component analysis," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 6, no. 5, pp. 441-448, May 2014.

[82] W. S. Saif, M. A. Esmail, A. M. Ragheb, T. A. Alshawi, and S. A. Alshebeili, "Machine learning techniques for optical monitoring and modulation format identification: a survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2839-2882, Fourth quarter 2020.

[83] Viavi, "ROADM and wavelength selective switches," 2019. [Online]. Available: https://www.viavisolutions.com/en-us/literature/ roadm-and-wavelength-selective-switches-application-notes-en.pdf

[84] T. A. Strasser and J. L. Wagener, "Wavelength-selective switches for ROADM applications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1150–1157, September-October 2010.

[85] R. Shankar, M. Florjanczyk, T. Hall, A. Vukovic, and H. Hua, "Multi-degree ROADM based on wavelength selective switches: architectures and scalability," *Optics Communications*, vol. 279, no. 1, pp. 94–100, November 2007.

[86] B. Collings, "New devices enabling software-defined optical networks," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 66-71, March 2013.

[87] T. Zami, B. Lavigne, and B. Faure, "Extra penalty when fitting the filtering bandwidth of successive traversed WSS's to a lower channel symbol rate," in *Optical Fiber Communications Conference (OFC)*, 2017.

[88] T. Zami, I. F. de Jauregui Ruiz, A. Ghazisaeidi, and B. Lavigne, "Growing impact of optical filtering in future WDM networks," in *Optical Fiber Communications Conference (OFC)*, 2019.

[89] I. F. De Jauregui Ruiz, T. Zami, and B. Lavigne, "Implications of NxM WSS in terms of filtering," in *OptoElectronics and Communications Conference (OECC)* and *International Conference on Photonics in Switching and Computing (PSC)*, 2019.

[90] Y. Sakamaki, T. Kawai, T. Komukai, M. Fukutoku, and T. Kataoka, "Evaluation of optical filtering penalty in digital coherent detection system," *IEICE Communication Express*, vol. 1, no. 2, pp. 54-59, July 2012.

[91] A. Ghazisaeidi, P. Tran, P. Brindel, O. Bertran-Pardo, J. Renaudier, G. Charlet, and S. Bigo, "Impact of tight optical filtering on the

performance of 28 Gbaud Nyquist-WDM PDM-8QAM over 37.5 GHz grid," in *Optical Fiber Communication Conference (OFC)*, 2013.

[92] C. Delezoide, P. Ramantanis, and P. Layec, "On the performance prediction of optical transmission systems in presence of filtering," in *International Conference on Transparent Optical Networks (ICTON)*, 2017.

[93] A. P. Vela, M. Ruiz, F. Fresi, N. Sambo, F. Cugini, G. Meloni, L. Potì, L. Velasco, and P. Castoldi, "BER degradation detection and failure identification in elastic optical networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 35, no. 21, pp. 4595-4604, November 2017.

[94] S. Shahkarami, F. Musumeci, F. Cugini, and M. Tornatore, "Machine-learning-based soft-failure detection and identification in optical networks," in *Optical Fiber Communications Conference (OFC)*, 2018.

[95] S. Varughese, D. Lippiatt, T. Richter, S. Tibuleac, and S. E. Ralph, "Identification of soft failures in optical links using low complexity anomaly detection," in *Optical Fiber Communications Conference (OFC)*, 2019

[96] T. Rahman, A. Napoli, D. Rafique, B. Spinnler, M. Kuschnerov, I. Lobato, B. Clouet, M. Bohn, C. Okonkwo, and H. de Waardt, "On the mitigation of optical filtering penalties originating from ROADM cascade," *IEEE Photonics Technology Letters*, vol. 26, no. 2, pp. 154-157, January 2014.

[97] Q. Hu, F. Buchali, M. Chagnon, K. Schuh, and H. Bülow, "3.6-Tbps duobinary 16-QAM transmission with improved tolerance to cascaded ROADM filtering penalty," in *European Conference on Optical Communication (ECOC)*, 2018.

[98] J. Pan and S. Tibuleac, "Real-time ROADM filtering penalty characterization and generalized precompensation for flexible grid networks," *IEEE Photonics Journal*, vol. 9, no. 3, pp. 1-10, June 2017.

[99] C. Delezoide, P. Layec, and S. Bigo, "Automated alignment between channel and filter cascade," in *Optical Fiber Communications Conference (OFC)*, 2019.

[100] M. Dallaglio, Q. Pham Van, F. Boitier, C. Delezoide, D. Verchere, P. Layec, A. Dupas, N. Sambo, S. Bigo, and P. Castoldi, "Demonstration of a SDN-based spectrum monitoring of elastic optical networks," in *Optical Fiber Communications Conference (OFC)*, 2017.

[101] L. Velasco, B. Shariati, A. P. Vela, J. Comellas, and M. Ruiz, "Learning from the optical spectrum: soft-failure identification and localization [Invited]," in *Optical Fiber Communications Conference (OFC)*, 2018.

[102] B. Shariati, M. Ruiz, J. Comellas, and L. Velasco, "Learning from the optical spectrum: failure detection and identification," *IEEE/OSA Journal of Lightwave Technology*, vol. 37, no. 2, pp. 433-440, January 2019.

[103] D. Wang, M. Zhang, Z. Zhang, J. Li, H. Gao, F. Zhang, and X. Chen, "Machine learning-based multifunctional optical spectrum analysis technique," *IEEE Access*, vol. 7, pp. 19726-19737, January 2019.

[104] M. Ruiz, A. Sgambelluri, F. Cugini, and L. Velasco, "Smart filterless optical networks based on optical spectrum analysis," in *International Conference on Transparent Optical Networks (ICTON)*, 2019.

[105] A. Mahajan, K. Christodoulopoulos, R. Martinez, S. Spadaro, and R. Munoz, "Modeling filtering penalties in ROADM-based networks with machine learning for QoT estimation," in *Optical Fiber Communications Conference (OFC)*, 2020.

[106] J. Li, G. Zhou, J. Zhou, and C. Lu, “Joint linear and nonlinear noise monitoring techniques based on spectrum analysis,” *Optics Express*, vol. 28, no. 24, pp. 36953-36971, November 2020.

[107] Y. Li, N. Hua, J. Li, Z. Zhong, S. Li, C. Zhao, X. Xue, and X. Zheng, “Optical spectrum feature analysis and recognition for optical network security with machine learning,” *Optics Express*, vol. 27, no. 17, pp. 24808-24827, August 2019.

[108] T. S. R. Shen, Q. Sui, and A. P. T. Lau, "OSNR monitoring for PM-QPSK systems with large inline chromatic dispersion using artificial neural network technique," *IEEE Photonics Technology Letters*, vol. 24, no. 17, pp. 1564-1567, September 2012.

[109]G. Bennett, K. Wu, A. Malik, S. Roy, and A. Awadalla, "A review of high-speed coherent transmission technologies for long-haul DWDM transmission at 100g and beyond," *IEEE Communications Magazine*, vol. 52, no. 10, pp. 102-110, October 2014.

[110] T. Zami, "Co-optimizing allocation of Nyquist superchannels and physical impairments aware placement of regenerators in elastic WDM networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 16, pp. 2830-2840, August 2014.

[111] G. Bosco, V. Curri, A. Carena, P. Poggiolini, and F. Forghieri, "On the performance of Nyquist-WDM terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers," *IEEE/OSA Journal of Lightwave Technology*, vol. 29, no. 1, pp. 53-61, January 2011.

[112] A. C. Jatoba-Neto, D. A. A. Mello, C. E. Rothenberg, S. Ö. Arik, and J. M. Kahn, "Scaling SDM optical networks using full-spectrum spatial switching," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 12, pp. 991-1004, December 2018.

[113] X. Liu and S. Chandrasekhar, "Superchannel for next-generation optical networks," in *Optical Fiber Communications Conference (OFC)*, 2014.

[114] J. Pan, C. Liu, T. Detwiler, A. J. Stark, Y. Hsueh, and S. E. Ralph, "Inter-channel crosstalk cancellation for Nyquist-WDM superchannel applications," *IEEE/OSA Journal of Lightwave Technology*, vol. 30, no. 24, pp. 3993-3999, December 2012.

[115] J. M. Fàbrega, F. Locatelli, L. Nadal, K. Christodoulopoulos, M. Svaluto Moreolo, and S. Spadaro, "Data plane elements for optical performance monitoring agnostic to the modulation format for disaggregated optical networks," in *International Conference on Transparent Optical Networks (ICTON)*, 2020.

[116] J. A. Altabas, L. F. Suhr, G. Silva Valdecasa, J. A. Lázaro, I. Garces, J. B. Jensen, and A. T. Clausen, "25Gbps quasicoherent receiver for beyond NG-PON2 access networks," in *European Conference on Optical Communication (ECOC)*, 2018

[117] E. Bedrosian, "A product theorem for Hilbert transforms," *Proceedings of the IEEE*, vol. 51, no. 5, pp. 868-869, May 1963.

[118] D. M. Baney, B. Szafraniec, and A. Motamedi, "Coherent optical spectrum analyzer," *IEEE Photonics Technology Letters*, vol. 14, no. 3, pp. 355-357, March 2002.

[119] Anritsu, “Tunics Purity,” 2008. [Online]. Available: https://dl.cdn-anritsu.com/en-us/test-measurement/files/Brochures-Datasheets-Catalogs/datasheet/Tunics_Purity_V2.pdf.

[120] J. M. Fàbrega, B. Schrenk, F. Bonada Bo, J. A. Lázaro, M. Forzati, P. J. Rigole, and J. Prat, "Modulated grating Y-structure tunable laser for λ-routed networks and optical access," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 17, no. 6, pp. 1542-1551, November-December 2011.

[121] II-VI Photonics, “Application Note AN-2095 Controlling the S7500 CW Tunable Laser,” November 2011. [Online]. Available: https://ii-vi.com/product/cw-tunable-laser-butterfly-package/

[122] Texas Instruments, “RF-sampling 12-bit ADC with dual-channel 5.2 GSPS or single-channel 10.4 GSPS,” 2021. [Online]. Available: https://www.ti.com/product/ADC12DJ5200RF

[123] Socionext, “103GSa/s 8-bit ADC H-Family,” 2019. [Online]. Available: https://www.eu.socionext.com/assets/downloads/37/16nm%20ADC%20IP%20H-Family%20V2.2.pdf

[124] II-VI Photonics, “High-speed detectors and receivers,” 2021. [Online]. Available: https://ii-vi.com/product-category/products/optical-communications/communication-components/high-speed-detectors-and-receivers.

[125] Spectrolab, “10 Gb/s InGaAs P-I-N Photodetector Die,”. [Online]. Available: https://www.spectrolab.com/sensors/SR_PIN_10G_Die.pdf.

[126] F. Locatelli, K. Christodoulopoulos, J. M. Fàbrega, M. Svaluto Moreolo, L. Nadal, and S. Spadaro, "Filter features extraction from optical spectra," in *European Conference on Optical Communications (ECOC)*, 2020.

[127] F. Locatelli, K. Christodoulopoulos, M. Svaluto Moreolo, J. M. Fàbrega, L. Nadal, and S. Spadaro, "Spectral processing techniques for efficient

monitoring in optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 13, no. 7, pp. 158-168, July 2021.

[128] F. Locatelli, K. Christodoulopoulos, J. M. Fàbrega, M. Svaluto Moreolo, L. Nadal, and S. Spadaro, "Experimental demonstration of a machine learning-based in-band OSNR estimator from optical spectra," in *International Conference on Optical Network Design and Modeling (ONDM)*, 2020.

[129] II-VI Photonics, "High resolution optical channel monitor," 2015. [Online]. Available: https://ii-vi.com/product/high-resolution-optical-channel-monitor-focm-series/

[130] Aragon Photonics, "BOSA high-resolution optical spectrum analyzer," 2021. [Online]. Available: https://aragonphotonics.com/bosa -optical-spectrum-analyzer/

[131] J. M. S. Domingo, J. Pelayo, F. Villuendas, C. D. Heras, and E. Pellejer, "Very high resolution optical spectrometry by stimulated Brillouin scattering," *IEEE Photonics Technology Letters*, vol. 17, no. 4, pp. 855-857, April 2005.

[132] C. Pulikkaseril, L. A. Stewart, M. A. F. Roelens, G. W. Baxter, S. Poole, and S. Frisken, "Spectral modeling of channel band shapes in wavelength selective switches," *Optics Express*, vol. 19, no.9, pp. 8458-8470, April 2011.

[133] Finisar, "WaveShaper 4000A multiport optical processor," 2021. [Online]. Available: https://ii-vi.com/product/waveshaper-4000a-multi port-optical-processor/

[134] Finisar, "WaveAnalyzer 1500S high-resolution optical spectrum analyzer," 2021. [Online]. Available: https://ii-vi.com/product/waveanal yzer-1500s-high-resolution-optical-spectrum-analyzer

[135] P. Ramantanis, C. Delezoide, P. Layec, and S. Bigo, "Revisiting the calculation of performance margins in monitoring-enabled optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 11, no. 10, pp. C67-C75, October 2019.

[136] Arris, “CHP EDFA headend erbium doped fiber amplifiers,” 2021. [Online]. Available: https://www.commscope.com/globalassets/digizuite/61636-chp-edfa.pdf

[137] F. Locatelli, K. Christodoulopoulos, M. Svaluto Moreolo, J. M. Fàbrega, L. Nadal, A. Mahajan, and S. Spadaro, “Feedback-based channels frequencies optimization in superchannels,” submitted to *IEEE/OSA Journal of Optical Communications and Networking*, September 2021.

[138] V. Patel and S. Zhang, “Stochastic gradient descent on nonconvex functions with general noise models,” 2021. [Online]. Available: https://arxiv.org/abs/2104.00423.

[139] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *IEEE/OSA Journal of Lightwave Technology*, vol. 32, no. 4, pp. 694-721, February 2014.

[140] J. R. R. A. Martins and J. T. Hwang, “Review and unification of methods for computing derivatives of multidisciplinary computational models,” *AIAA Journal*, vol. 51, no. 11, pp. 2582-2599, November 2013.

[141] R. Tibshirani, R. Chen, D. Schwartz, and S. Yao, “Convex Optimization,” 2015. [Online]. Available: http://www.stat.cmu.edu/~ryantibs/convexopt-F15/lectures/07-sg-method.pdf.

[142] F. Paolucci, A. Sgambelluri, F. Cugini, and P. Castoldi, "Network telemetry streaming services in SDN-based disaggregated optical networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 36, no. 15, pp. 3142-3149, August 2018.

[143] A. Mahajan, K. Christodoulopoulos, R. Martínez, R. Muñoz, and S. Spadaro, "Quality of transmission estimator retraining for dynamic optimization in optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B45-B59, April 2021.