UNIVERSIDAD DE ANTIOQUIA

Estimation and Reduction of Inter−Channel Interference and Nonlinear Effects in Nyquist−WDM systems using Machine Learning

Alejandro Escobar Pérez

Cita: [1] Referencia Estilo IEEE (2020) [1] A. Escobar Pérez, "Estimation and Reduction of Inter-Channel Interference and Nonlinear Effects in Nyquist-WDM systems using Machine Learning" Tesis de maestría, Maestría en Ingeniería de telecomunicaciones, Universidad de Antioquia, Medellín, 2022.

Maestría en Ingeniería de Telecomunicaciones.
Grupo de Investigación en Telecomunicaciones Aplicadas.

Centro de Documentación Ingeniería.
Universidad de Antioquia - www.udea.edu.co

Rector: Jhon Jairo Arboleda Céspedes.

Decano/Director: Francisco Vargas Bonilla.

Jefe departamento: Augusto Enrique Salazar Jimenez.

Estimation and Reduction of Inter−Channel Interference and Nonlinear Effects in Nyquist−WDM systems using Machine Learning

Master's Thesis in Telecommunication Engineering

Submitted by

Alejandro Escobar Pérez

UNIVERSIDAD DE ANTIOQUIA

1 8 0 3

Advisor: Prof. Dr.−Ing. Jhon James Granada Torres

Medellín, Colombia

Faculty of Engineering

Department of Electronics and Telecommunications

University of Antioquia.

Abstract

The increment of data traffic demand by end users will require to increase the capacity of next generation optical networks, both, long−haul and optical access networks (OANs). The use of high-level modulation formats, wavelength division multiplexing (WDM) and digital coherent receivers, will enable the increase of the OANs capacity. On the other hand, for long-haul networks, the paradigm called Elastic Optical Networks (EONs) will be essential to increase the capacity, thanks to the dynamic use of networks resources. For instance, a dynamic use of the optical spectrum will enable the gridless networks which jointly with Nyquist pulses that present rectangular spectral shapes, will improve the Spectral Efficiency (SE) and thus, the capacity of the networks.

Nevertheless, nonlinear phase noise, Kerr effects and inter-channel interference (ICI), are some of the challenges that future OANs and gridless networks would face. The nonlinear effects are present when high launch power (needed for high order modulation formats) is induced into the optical fiber. These effects distort the transmitted symbols generating errors after demodulation and, thus, increasing bit error rate (BER). On the other hand, the ICI effects increase the BER due to the interaction of the optical channels when the spectral spacing is reduced up to the baud rate.

Machine Learning (ML)−based techniques have shown improvement of traditional monitoring and mitigation of different effects in optical communications. ML is regarded as one of the most promising methodological approaches to perform network-data analysis and has shown relevant early contributions in optical communications, allowing channel estimation, mitigation of nonlinear effects and foresees to be a powerful tool into the optical performance monitoring (OPM) field.

Thus, the goal of this work is to develop techniques using ML algorithms for i) estimation of spectral spacing in gridless Nyquist-WDM systems and ii) mitigation of nonlinear and ICI effects. For the former, two spectral overlapping estimation tools are proposed, both use information obtained from fuzzy clustering algorithms applied to 10k received symbols frames. The first one uses the membership degrees of the partition matrix resulted by an unsupervised learning algorithm to construct different features which are called "counting vectors". The second uses clustering validation indexes after the unsupervised learning algorithms are applied. It is tested two unsupervised learning algorithms that use fuzzy clustering: Fuzzy c-Means (FCM) and Gustafson-Kessel Means (GKM). On the other hand, in order to reduce nonlinear and ICI effects, three supervised learning algorithms: Support Vector Machine (SVM), Artificial Neural Networks (ANN) and k-Nearest Neighbors (KNN) are proposed to carry out nonsymmetrical demodulation and thus, minimize the BER.

Experimental results show that it is possible to determinate if any constellation diagram is affected by spectral overlapping using any of the proposed methods. Accuracies up to 91% are achieved in binary classification knowing a priori the OSNR value in the case of 32 GBd with fiber transmission. Moreover, when the OSNR is not known by the classifier, accuracies up to ~80% are obtained using the 16 GBd data. Besides, classification of different scenarios as high or low spectral overlapping jointly with detection of close channels or single channel is obtained also. Whereby, results show that it is possible to mitigate nonlinear effects as well as ICI effects achieving gains in terms of the OSNR. Additionally, the BER reduction is obtained at low training lengths for all the three proposed methods. Simulation results show the effectiveness of the methods by presenting gains up to ~0.5 dB in terms of the OSNR at a FEC limit of $log_{10}(BER) = -2.4$. Besides, at high launch power, traditional demodulation method reaches BER values of $log_{10}(BER) = -2$ while the proposed methods obtain ~$log_{10}(BER) = -2.6$. Moreover, the experimental results show mitigation of ICI effects achieving gains up to ~4 dB in terms of OSNR at spectral overlapping of 18%.

Consequently, it is demonstrated that the use of ML algorithms will be useful in future OANs and gridless networks by allowing mitigation of signal impairments and could be helpful to control lasers frequencies by estimating the spectral overlapping among channels.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Context

The increasing internet content given by the new generation end users devices and the need of high-quality data transmissions in telemedicine, internet of things and smart-cities, demands a parallel increment in the capacity of the optical fiber networks [1]. The next generation networks based on 5G standards, will require 100 times the capacity of current deployed networks, representing data rates in the order of several Gbps for access networks and Tbps for long haul links [1, 2].

For instance, nowadays optical access networks (OAN) reach up to 2.5 Gbps using passive optical networks (PON) standards with limitations in the intensity modulation [3]. On the other hand, long haul networks based on digital coherent detection and Wavelength Division Multiplexing (WDM), achieve data rates up to 100 Gbps per channel [4]. Thence, for OANs capacity increment, it has been proposed: i) the introduction of digital coherent detection into the PONs, using phase modulations as quadrature amplitude (QAM) formats [5, 6], ii) the use of WDM which increase capacity due to the possibility of transmitting several information channels into the same fiber (WDM-PON) [3, 6]. Moreover, for capacity increment in long haul links, the paradigm of Elastic Optical Networks (EONs) has been foreseen, where the networks resource management and use of the optical spectrum would be dynamic, improving the spectral efficiency (SE) with the use of gridless networks [7]. For instance, traditional WDM standard, recommended by the International Telecommunications Union (ITU), allows optical channels separated from each other by 50 GHz [8]. Also, in order to increase the SE, flexi-grid arrangements with spectral spacing of $n \times 6.25$ GHz have been proposed and standardized [9]. However, due to the dynamic use of resources that the EONs promise, the bit rate of each channel would change at any moment according to the network traffic needs. These traffic changes could lead to blank frequencies slots in the optical spectrum. Thence, the paradigm of gridless networks would fill the frequency slots reassigning carrier frequencies of the adjacent channels or adding new channels. Figure 1 shows spectral comparison among traditional WDM, flexi-grid and gridless.
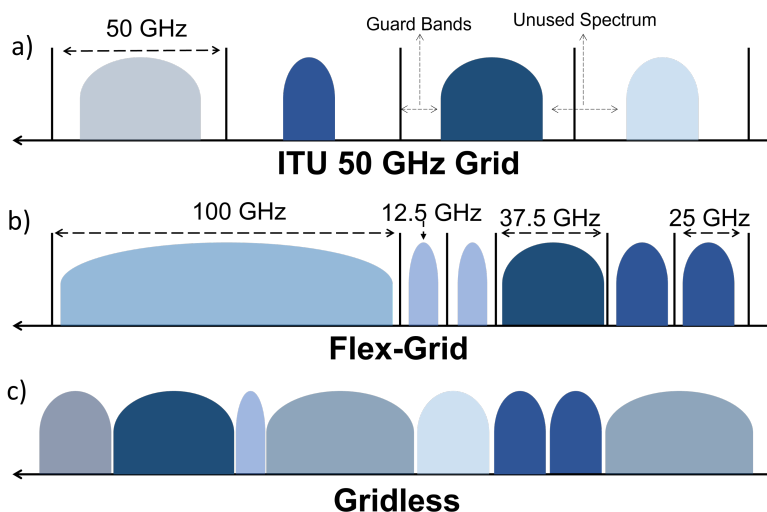
Figure 1.1: Comparison in the use of spectrum for a) ITU 50 GHz grid, b) flex-grid and c) gridless.

1.2 Challenges & Problem Statement

The new capacity increment strategies must face several challenges, mainly, nonlinear effects as i) phase noise due to the lack of monochromatic sources, ii) Kerr effects due to the high input power needed to maintain high optical signal-to-noise ratio (OSNR) in m-QAM formats and iii) Inter-channel Interference (ICI) due to the optical carriers closeness that results in spectral overlapping in gridless optical networks [10]. The nonlinear effects distort the transmitted symbols values generating demodulation errors and, therefore, increasing bit error rate (BER). For instance, Fig. 2a shows a 16-QAM constellation diagram where the symbols have been affected by Gaussian linear noise, meanwhile Fig. 2b shows the distortion given by nonlinear phase noise. Moreover, Fig. 2c shows traditional demodulation which uses hard-decision rule according to ideal symbol positions. Therefore, several approaches in receivers with Digital Signal Processing (DSP) modules have been proposed using nonlinear digital equalizers for channel estimation and nonlinear effects mitigation. Nonetheless, these efforts would not be enough for data rates of Terabits order because the required computational cost to perform digital back-propagation becomes unfeasible [11, 12].
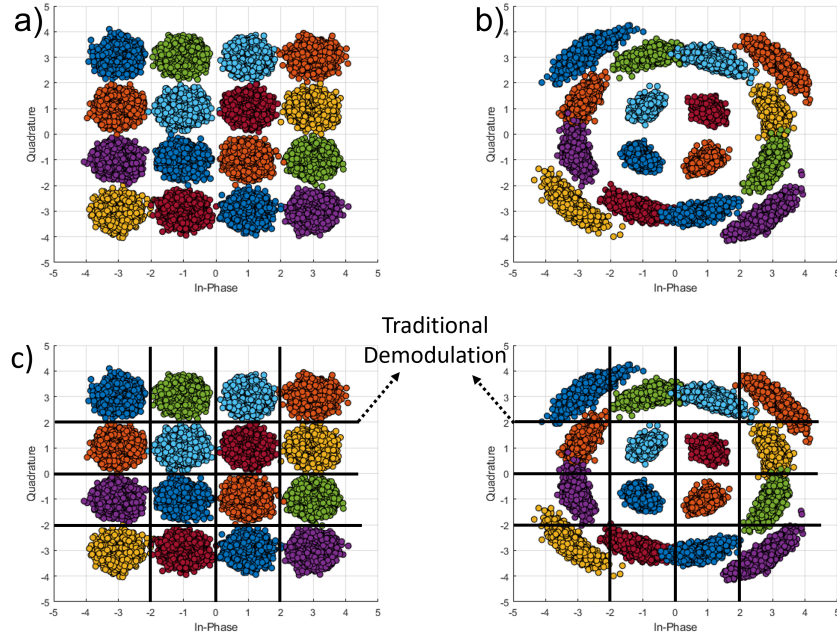
Figure 1.2: a)16-QAM constellation affected with Gaussian Noise. b) 16-QAM constellation affected with Nonlinear Noise. c) Traditional demodulation.

On the other hand, the ICI effects increment the BER due to the interaction of the adjacent channels which alters the symbols position causing changes in the in-phase and quadrature components as well. Additionally, mitigation and monitoring of ICI effects are challenging because its effects seen in a constellation diagram, are similar to the well-known Gaussian noise. For example, Fig. 1.3a shows a constellation diagram of a single-channel scenario affected by linear Gaussian noise given by OSNR of 19.3 dB, meanwhile, Fig. 1.3b shows a constellation diagram of a highly overlapped channels with OSNR of 36.3 dB. Both constellations

differ from 17 dB of OSNR, and it can be seen similar circular-like distortion. Consequently, ICI effects have been modeled as Gaussian-like distortions [13,14], nevertheless, it can produce slight temporal asymmetrical movements of the symbol's centroid seen in short time windows, thus, traditional demodulation techniques, does not effectively mitigate ICI effects [15].
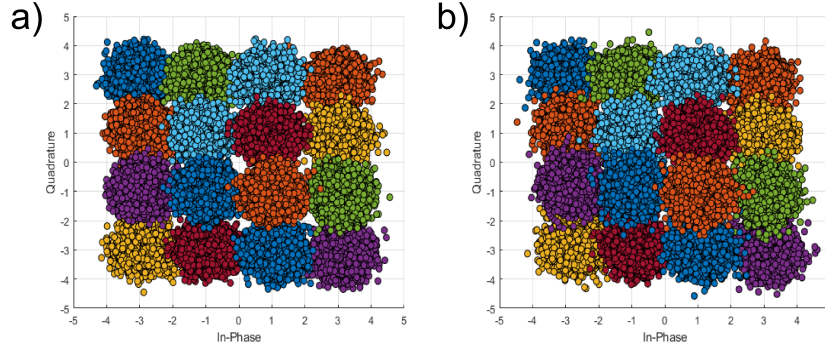
Figure 1.3: 16-QAM constellation diagram with a) OSNR of 19.3 dB and b) OSNR of 36.3 dB and spectral overlapping of 18%.

Multiple-Inputs Multiple-Outputs (MIMO) equalizers have been proposed to mitigate ICI effects using information of the adjacent channels [16–19]. However, the flexibility granted by the future gridless networks with reconfigurable optical add/drop multiplexers (ROADMs) at different nodes, will not guarantee that a channel affected by ICI could be equalized using information of its adjacent channels. Figure 1.4 show an schematic of a gridless network where in each node, a ROADM adds and/or removes optical carriers. The ROADMs are based on wavelength selective switches that assigns different lighpaths to incoming channels. Note that the gray information channel at the node A is firstly affected by the red and purple channels. Then, at the destination node (B), these channels are not available for ICI mitigation using MIMO equalizers. On the other hand, ICI effects have been mitigated using bit overheads for channel encoding [20] and by optical means using wave mixing approaches [21, 22]. Though, in order to increment channel capacity in terms of transmitted goodput, the robust encoding overheads must be avoided. Furthermore, the use of optical devices for optical mixing tends to be expensive.

The need for mitigation of complex signal impairments and monitoring the optical networks, has led to a new application of Machine Learning (ML) that has shown up into scene as a branch of Artificial Intelligence (AI) that have motivated the concept of machines that learn to solve specific problems, in some cases, at low computational cost. ML is regarded as one of the most promising methodological approaches to perform network-data analysis and has shown relevant early contributions in optical communications, reaching for some scenarios and conditions, channel estimation [23], mitigation of nonlinear effects [24,25] and foresees to be a powerful tool into the optical performance monitoring (OPM) field [26, 27].

Thus, the goal of this work is to propose two methods based on ML: i) to estimate the spectral overlapping of received signals in m-QAM gridless Nyquist-WDM systems and ii) to reduce nonlinear impairments of received signal in AONs scenarios using WDM systems and to minimize ICI effects in order to decrease the BER in gridless Nyquist-WDM systems. The proposed methods will be validated using simulated and experimental data.
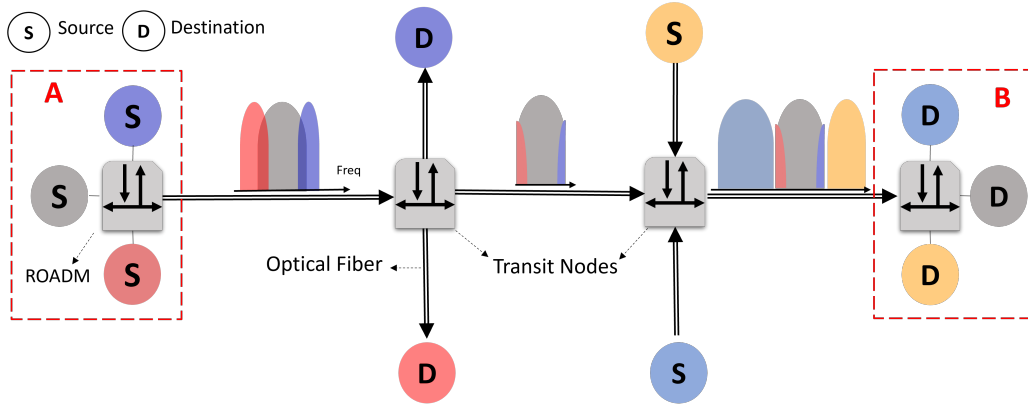
Figure 1.4: Inter-channel Interference in future gridless WDM scenarios due to the adding-dropping process when channel spacing is reduced.

1.3   State of the Art

In the found literature, both technological paradigms: optical communications and ML had crucial moments in 2006 when the first study of digital coherent detection was published [28] and the reinventing of the 1980's Artificial Neural Network (ANN) with the introduction of Deep Neural Networks (DNN) had place at the same year [29]. It was just a matter of time when both fields started to work together, first, as monitoring tools that were first published in 2007 with the application of ANN [30] and demodulation approaches that were published in 2010 with a k-Means adaptation in Radio-over-Fiber (RoF) link [31].

This state of the art is divided in two sections: the works focused on optical performance estimation and the ones in signal impairments mitigation. In this sense, Table 1.1 shows the recent works in which optical network or signal performance is estimated using ML algorithms. On the other hand, Table 1.2 shows works related to mitigation of signal impairments using ML approaches. In both tables it is specified different characteristics as type of ML algorithm, modulation format, fiber length or data rate used.

1.3.1   Optical Performance Estimation

The performance of a telecommunication system could be measured by the product of data rate and transmission distance, although higher distances, leads to strong attenuation issues which are faced by using several amplifiers in different fiber spans. However, the optical amplifiers add noise to the signal resulting in reduction of the OSNR values which increases the BER. Therefore, in order to retain the OSNR, the launch power is increased which can be unfavorable due to the stimulation of fiber nonlinearities. Consequently, an accurate OSNR measurement is required in order to present an optimal trade-off between power consumption and signal quality. On the other hand, high-level modulation formats increase the capacity of the networks without using a broad bandwidth due to the fact that the information is included into the phase of the signal carrier. Thus, in order to increase the capacity of the channel high level modulation formats will be mostly used and, according to EONs flexibility, an optimal identification of modulation formats will be necessary.

According to the above and in order to present a wider view of the tendencies of the ML role into the optical performance estimation, Table 1.1 shows the works related to the optical performance estimation us-

ing ML-based techniques in which it is specified what network or signal parameter is estimated, modulation format used, information of the ML algorithm, fiber length and data rate.

Table 1.1: Review of Optical Performance Estimation using Machine Learning.

Monitored	Type of Learning and used Algorithm	Simulated / Experimental	Modulation Format(s)	Fiber Length	Data Rate	Year	Reference
OSNR, Chromatic Dispersion, Polarization Mode Dispersion	Supervised / ANN	Simulated	QPSK	-	-	2010	[32]
OSNR, Chromatic Dispersion, Modulation Format	Supervised / PCA	Experimental	OOK NRZ-DPSK RZ-DPSK	-	5 GBd	2015	[33]
OSNR	Supervised / DNN	Experimental	QPSK	B2B	16 GBd	2016	[34]
OSNR, Modulation Format	Supervised / DNN	Experimental	QPSK, 16-QAM, 64-QAM	80 km	60 GBd, 40 GBd, 30 GBd	2017	[35]
OSNR Nonlinear Noise	Supervised / ANN	Simulated	16-QAM	3200 km	56 GBd	2017	[36]
OSNR, Modulation Format	Supervised / Convolutional ANN	Simulated and Experimental	QPSK, 8-PSK, 16-QAM, 32-QAM, 64-QAM	B2B	25 GBd	2017	[37]
OSNR, Nonlinear Noise	Supervised / ANN	Simulated	16-QAM	1500 km	-	2018	[38]
OSNR, Data rate, Modulation Format	Supervised / Convolutional ANN	Experimental	QPSK, 16-QAM, 64-QAM	-	14 GBd, 16 GBd	2018	[39]
OSNR, Nonlinear Noise	Supervised / ANN	Simulated	QPSK, 16-QAM, 64-QAM	1000 km	26 GBd	2018	[40]
OSNR	Supervised / Convolutional ANN	Simulated and Experimental	QPSK, 16-QAM, 64-QAM	B2B	14 GBd, 16 GBd	2019	[41]
Resource Assignment	Unsupervised / Fuzzy c-Means	Simulated	BPSK, QPSK	-	-	2019	[42]
Spectral Overlapping	Unsupervised / DB-SCAN	Experimental	16-QAM	B2B	16 GBd	2019	[43]
Nonlinear Noise	Supervised / Support Vector Regression	Simulated	16-QAM	3200 km	60 GBd	2019	[44]
OSNR, Chromatic Dispersion	Supervised / ANN	Experimental	16-QAM, 64-QAM	120 km	10 GBd	2019	[45]
OSNR	Supervised / DNN	Simulated and Experimental	QPSK	80 km	28 GBd	2019	[46]
OSNR, Center Wavelength, Data Rate	Supervised / Support Vector Machine	Simulated	QPSK	-	-	2019	[47]
OSNR, Nonlinear Noise	Supervised / ANN	Experimental	16-QAM	1000 km	-	2019	[48]
OSNR	Supervised / PCA and Support Vector Machine	Simulated	QPSK	1000 km	100 GBd	2020	[49]
OSNR, Modulation Format	Supervised / CNN	Simulated and Experimental	QPSK, 8-QAM, 16-QAM, 64-QAM	270 km	32 GBd	2020	[50]
Quality of Transmission	Supervised / Decision Tree and ANN	Experimental	QPSK, 8-QAM, 16-QAM, 64-QAM	-	200 GBd	2021	[51]
OSNR	Supervised / ANN	Simulated	64-QAM	1600 km	70 GBd	2021	[52]

Marked in red: works where OSNR is monitored.
Marked in gray: works using Artificial Neural Networks.
Marked in blue: works using 16-QAM modulation format.

It is clear that the concern of an accurate OSNR measurement have been attended by the authors (see marked in red in Table 1.1). Traditional methods of OSNR estimation are based on optical spectrum analyzers (OSAs), where linear interpolations between signal spectrum power and noise floor is performed, this measurement tends to be inaccurate when multiple signals with narrow spectral spacing are transmitted, for instance in gridless or dense-WDM networks [49]. Thus, the versatility of the ML-based techniques allowed deeper constellation diagram analysis in which the OSNR could be measured. These advantages are

exploited by several authors [32–34, 36, 39, 41, 45, 46, 50], in which improvements in OSNR estimation are reported with low bias values up to 0.2 dB [36] without using spectrum analysis but constellation diagrams. Moreover, improvement on spectrum analysis has been reached also [40,49]. On the other hand, only in [43], differentiation between OSNR and ICI effects is carried out in a gridless network. Here, the spectral spacing is varied obtaining up to 30% of spectral overlapping in a Nyquist-WDM system of three channels at 16 GBd data rate. Then, using analysis of diagram constellation affected by ICI, unsupervised algorithm which finds different densities in the symbols groups is applied. Thereby, the algorithm classifies different symbols as noise (do not belong to any of the found clusters) and considering those noise symbols, an overlapping index is calculated. Furthermore, detection of ICI effects at low OSNR values is achieved in scenarios with more than 12.5% spectral overlapping. However, transmission was not considered and spectral spacing estimation with fiber distortions are challenging. The state of the art review shows that OSNR estimation has been attended by the author, thereby, approaches of ICI estimation are still needed where assumption of an accurate OSNR could be forseen.

Additionally, marked in grey in Table 1.1, are the works where the main used algorithm relies on neural networks, this is because they are the most promising and convincing approaches not only for optical communications, but technological advances in the data analysis field. Moreover, in blue, are the works that used 16-QAM modulation formats, pointing the next generation networks will use quadrature-based modulations.

1.3.2   Nonlinear Effects Mitigation

The demodulation process is focused on a precise symbol detection according to minimum distances to ideal symbol positions. Thence, in linear noise scenarios in which distorted symbols surrounds ideal positions following Gaussian-like distortions, traditional demodulation techniques are suitable in order to obtain lower BER values. Even then, the presence of nonlinear noises produces distortions that do not necessarily surround ideal positions or circular shapes, thus, mitigation of nonlinear effects is a concern that have been attended, mostly, in long-haul links, this can be seen marked in red on Table 1.2. Nonetheless, for OANs, it is required more research considering the relative low distances and high OSNR that OANs would require for m-QAM constellations. Consequently, low-complexity strategies to reduce nonlinear effects in OANs will be needed in order to increase its capacity for new generation technologies, for example, RoF and internet of things (IoT). Furthermore, minimization of ICI effects using ML in gridless Nyquist-WDM systems has been carried out only in [15, 53] using data rates of 16 GBd and 32 GBd, respectively.

In [53], supervised algorithm is applied to minimize ICI effects where gains up to ~1.3 dB (in terms of OSNR) is achieved in the scenario of 24% spectral overlapping, however, no fiber transmission is analyzed. Moreover, in [15] an unsupervised algorithm is applied for ICI effects minimization using fiber transmission of 270 km and gains up to 0.6 dB are achieved for spectral spacing of less than 32 GHz in terms of OSNR.

The review presents a lack of research in ICI effects mitigation, the currently reported approaches [54] will still need a wider exploration and constellation diagram characterization must be studied for a better spectral overlapping estimation. Besides, more research at different data rates, transmission distance, modulation formats are needed due to the dynamic use of resources that EONs would require.

Table 1.2: Review of optical signal impairments mitigation using Machine Learning.

Mitigation	Type of Learning and used Algorithm	Simulated / Experimental	Modulation Format(s)	Fiber Length	Data Rate	Year	Reference
Nonlinear Effects	Supervised / ANN	Simulated	16-QAM	1000 km	5 GBd	2015	[55]
Nonlinear Phase Noise	Supervised / Support Vector Machine	Simulated	BPSK	480 km	40 GBd	2016	[56]
ICI Effects	Unsupervised / k-Means	Experimental	16-QAM	270 km	32 GBd	2017	[15]
Nonlinear Effects	Supervised / Support Vector Machine	Experimental	16-QAM	2000 km	40 GBd	2017	[57]
Nonlinear Effects	Supervised and Unsupervised / Hierarchical Clustering and ANN	Experimental	QPSK	3200 km	20 GBd	2017	[58]
Nonlinear Effects	Supervised / ANN	Experimental	4-QAM	3200 km	20 GBd	2017	[59]
Nonlinear Effects	Supervised / DNN	Simulated	16-QAM	3200 km	20 GBd	2018	[25]
Nonlinear Effects	Supervised and Unsupervised / k-Means, FCM, Hierarchical clustering, ANN, SVM	Simulated and Experimental	BPSK	2000 km	5 GBd	2018	[23]
Nonlinear Effects	Supervised / Convolutional ANN	Experimental	4PAM	40 km	56 GBd	2018	[60]
Nonlinear Effects	Supervised / ANN	Simulated	16-QAM	800 km	40 GBd	2018	[61]
Nonlinear Effects	Supervised / DNN	Simulated and Experimental	QPSK	B2B	10 GBd	2019	[62]
ICI Effects	Supervised / KNN	Experimental	16-QAM	B2B	16 GBd	2019	[53]
Nonlinear Phase Noise	Supervised and Unsupervised / KNN, k-Means and Fuzzy c-Means	Simulated	16-QAM	50 km	32 GBd	2019	[63]
Nonlinear Phase Noise	Unsupervised/ GK-means	Experimental	4+12 PSK	80 km	1 GBd	2019	[64]
Nonlinear Effects	Unsupervised / Affinity Propagation	Experimental	QPSK	3200 km	2.5 GBd	2019	[65]
Nonlinear Effects	Supervised / ANN	Simulated	4-PAM	-	100 GBd	2019	[66]
Nonlinear Effects	Supervised / Parzen Window	Simulated	16-QAM, 64-QAM	1600 km	56 GBd	2019	[67]
Nonlinear Effects	Supervised / CNN	Simulated	16-QAM	3200 km	100 GBd	2021	[68]
Nonlinear Effects	Supervised / ANN	Simulated and Experimental	16-QAM	480 km	32 GBd	2021	[69]

Marked in gray: works using Artificial Neural Networks.
Marked in blue: works using 16-QAM modulation format.
Marked in red: works using long-haul links.

1.4 Objectives

Due to the dynamism of gridless networks that would cause spectral overlapping, the computational cost that nonlinear equalizers would require to mitigate nonlinear effects and the unavailability of adjacent channels at the end nodes for ICI mitigation, the following objectives using ML approaches are proposed:

General: To propose and simulate two methods based on Machine Learning to reduce and estimate Inter-Channel Interference and nonlinear effects in Nyquist-WDM systems.

Specific:

- To simulate a m-QAM Nyquist-WDM system including nonlinear impairments and Inter-channel Interference effects.

- To propose a method based on machine learning to minimize ICI and nonlinear effects in terms of BER in Nyquist-WDM systems.

- To propose a method based on machine learning to estimate the percentage of spectral overlapping in Nyquist-WDM systems.

- To validate the proposed methods using the simulated data and experimental data (acquired by the advisor in previous works) in terms of BER and accuracy as functions of different system parameters such as: optical channel spacing, OSNR, launch power and transmission distance.

1.5   Contributions

In this research, there were proposed and tested i) two methods for estimation of spectral spacing and ii) three methods to reduce nonlinear and ICI effects. The former pursues the improvement of a gridless network performance in which the optical spectrum will be used dynamically. Besides, the latter aims to minimize the BER in order to increment transmissions capacity of the OANs and gridless networks. Moreover, the spectral overlapping estimation and ICI effects minimization will be applied into the gridless networks (Fig. 1.5a), on the other hand, the nonlinear effects mitigation, will be held into the OANs (Fig. 1.5b).
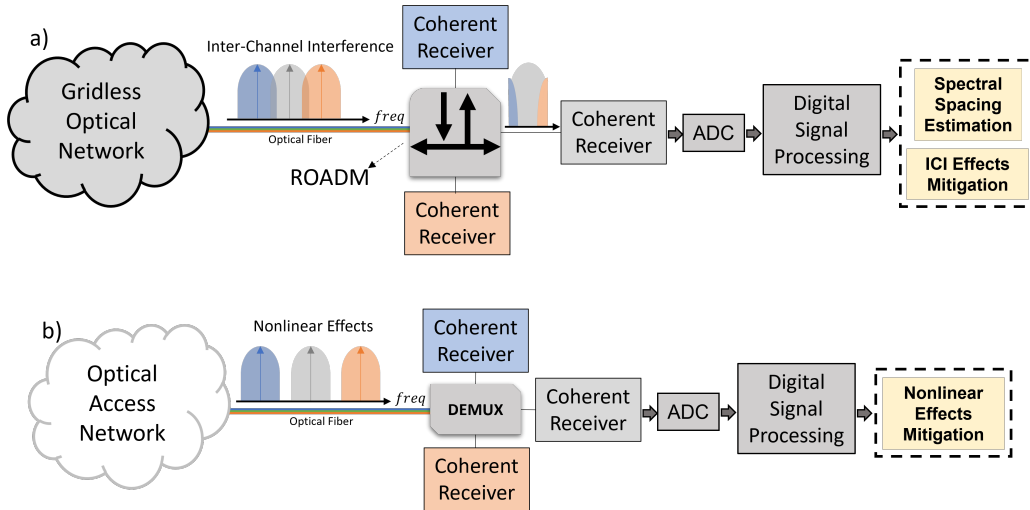
Figure 1.5: Contribution allocation of the proposed methods in the case of a) gridless optical networks and b) optical access networks.

(i) **Contribution of the methods for spectral overlapping estimation:**
   Two different methods are proposed to estimate the spectral overlapping using constellation diagram analysis. Both methods use information of unsupervised learning algorithms that are applied to received symbols affected or not by ICI effects. After the algorithms perform the classification, several features are extracted using the information of the classification and then, estimation of the spectral spacing is carried out using supervised learning. Specifically, the first method use the membership degrees resulting of the classification by unsupervised learning algorithms. On the other hand, the second

method uses validation indexes considering a clustering evaluation. Both methods are tested using experimental data from two different gridless setups a) 3×16 GBd Nyquist-WDM system (CPqD) and b) 3×32 GBd Nyquist-WDM (GeorgiaTech) which are described in Section 3, in both, spectral spacing among the three channels were varied obtaining different spectral overlapping distortions. The performance of the methods is evaluated by means of accuracy metric vs different configurations of spectral spacing scenarios, for example, binary case in which the classifier detects if a received frame is affected or not by spectral overlapping.

(ii) **Contribution of the methods for minimization of nonlinear and ICI effects:**
Three different supervised learning algorithms are proposed to perform nonsymmetric demodulation in order to reduce nonlinear and ICI effects. The methods are tested using simulated data generated in VPIDesignSuite v11.2 specialized software and using experimental gridless networks data. The methods performance is evaluated employing BER vs OSNR results and gains in terms of OSNR are analyzed. Additionally, a training strategy is detailed for future algorithms implementations.

Publications:
This research work resulted in the following publications:

Journals

- **Alejandro Escobar Perez**, D. Zabala-Blanco, C. A. Azurdia Meza, N. Guerrero González, and J. J. Granada Torres, "Training strategies to minimize interchannel interference effects using supervised learning in gridless Nyquist-WDM systems," Appl. Opt., vol. 60, no. 28, p. 8939, 2021, doi: 10.1364/ao.428856.

- **Alejandro Escobar Perez**, N. G. González, and J. J. G. Torres, "Spectral overlapping estimation based on machine learning for gridless Nyquist-wavelength division multiplexing systems," Opt. Eng., vol. 59, no. 07, p. 1, 2020, doi: 10.1117/1.oe.59.7.076116.

- **Alejandro Escobar Perez**, Karen A. Giraldo, Jhon. A. L. Cortés, and Jhon. J. Granada. T., "Mitigation of Nonlinear Effects using Machine Learning in Coherent Optical Access Networks," INGE CUC, vol. 17, no. 2, 2021.

Conferences

- **Alejandro Escobar Perez**, A. Estrada Moscoso, N. G. Gonzalez, and J. J. G. Torres, "Training Approaches in Supervised Learning for ICI Mitigation in Gridless Nyquist-WDM," 2020 IEEE Photonics Conf. IPC 2020 - Proc., no. 53, pp. 2020–2021, 2020, doi: 10.1109/IPC47351.2020.9252366.

- Marquez-Viloria, D., Castano-Londono, L., **Alejandro Escobar Perez**, Granada-Torres, J. J., & Guerrero-Gonzalez, "Scalable Multi-Core FPGA Design for Maximum Concurrency: The Case of KNN for ICI Mitigation." Signal Processing in Photonic Communications. Optical Society of America, 2020.

- **Alejandro Escobar Perez**, N. G. González, and J. J. Granada Torres, "Inter-Channel Interference Estimation based on IQHistograms including Machine Learning," no. 53, p. SpTh2I.6, 2020, doi: 10.1364/sppcom.2020.spth2i.6.

- **Alejandro Escobar Perez**, N. G. González, Stephen E. Ralph and J. J. Granada Torres, "Spectral Spacing Estimation using Fuzzy Logic and KNN algorithms in Gridless Nyquist-WDM Systems," no. , p. SpTh1D.7, 2021, (Presented on July 29, published on OSA proceedings).

- **Alejandro Escobar Perez**, Omar D. Vargas, Stephen E. Ralph, and J. J. Granada Torres, "Spectral Spacing Estimation in Gridless Nyquist-WDM Systems using Local Binary Patterns," 2021 IEEE Photonics Conf. IPC 2021 - Proc., no. , pp. 2020–2021, 2021, (Presented on October 20, published on IEEE proceedings).

1.6 Document Organization

The remainder of this work comprises six chapters. In the first, the motivation jointly with the problem statement and objectives are presented. Besides, a brief introduction to the gridless networks and ICI effects are mentioned. In Chapter 2, an introductory and theoretical description about main optical communications technologies and networks, signal impairments and Machine Learning (ML) is given. In Chapter 3, the three different setups in which the objectives are developed are detailed. Here, the first one follows a simulated WDM system of three channels where launch power, laser linewidth and noise levels are varied in order to stimulate nonlinear distortions. The second and third ones compose experimental gridless WDM-systems (16 GBd and 32 GBd, respectively) in which spectral overlapping is induced concerning ICI and fiber effects. In Chapter 4, the proposed methods are exposed including two techniques for estimation of spectral overlapping and three for mitigation of nonlinear and ICI effects. Additionally, different hyperparameter selection and training strategies for possible future implementations are presented. In Chapter 5, results and discussions are given. There, the performance of the methods for spectral overlapping estimation is evaluated in terms of accuracy, on the other hand, the mitigation of nonlinear and ICI effects is analyzed in terms of OSNR. Finally, conclusions, main outlines and future works are discussed in Chapter 6.

2   Theoretical Framework

In this section a brief theoretical basis of the main topics exposed in this work will be explained in order to contextualize the proposed methods.

2.1   Optical Communications

The optical communications are based on modulation of optical signals which are transmitted through an optical fiber. The use of optical fiber gives a lot of advantages, for instance, its prime matter, the silicon (Si), is the second most abundant element on the planet, behind the oxygen. Besides, the optical fiber presents electromagnetic immunity and astounding low attenuation (typically 0.2 dB/km). Operating at the optical regime (C band, 1530 to 1565 nm), allows the possibility to exploit exceptional data rates due to the availability of high bandwidth. Figure 2.1a shows a basic optical communication system which modulates the light intensity with direct detection (IM-DD). With the use of photodetectors, the received optical signal is transferred to the electrical field, then, the bit decision is made according to electrical signal amplitude using assigned thresholds. Since the information is carried on the signal amplitude, this type of transmission is very vulnerable to high dispersion owing to pulses' broadening and attenuation, making difficult an error-free transmissions. On the other hand, IM-DD systems have allowed bit rates over 10 Gbps at ~40 km in a single channel configuration [70]. However, higher data rates technologies use the quadrature modulations which
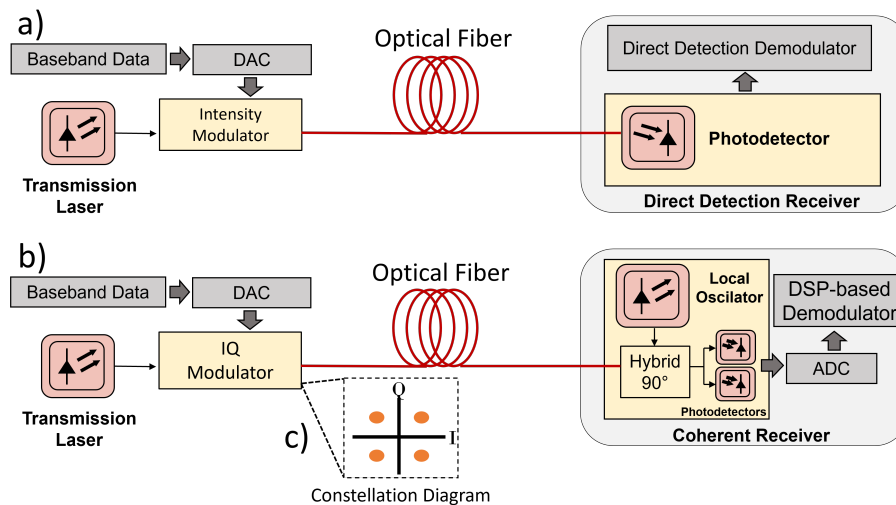
Figure 2.1: a) Intensity Modulation – Direct Detection scheme. b) Coherent Detection scheme c) Constellation Diagram of 4-QAM.

combines phase and amplitude modulations known as Quadrature and Amplitude Modulation (QAM). The use of QAM increased the capacity of the networks because it allows higher SE and greater noise tolerance compared to amplitude modulation. This modulation formats requires coherent detection techniques (see Fig. 2.1b) which need at the reception, an optical source with comparable operating frequencies of the transmitter source (local oscillator). Then, the detection of the phase changes is made correlating the received signal with the local oscillator where a recovering of the real and complex components are carried out.

The transmitted symbol information can be mapped in constellation diagram which is a two-dimensional plane where X and Y axes components are called In-phase (I) and Quadrature (Q), respectively (see Fig. 2.1c). Besides, after the symbols IQ information is recovered, the DSP modules are used for compensation of several channel distortions as well as for clock recovery and frequency offset correction with the use of a digital equalizers with which it is possible to estimate the optical channel and thus, minimize channel impairments [71].

2.1.1   Wavelength Division Multiplexing

WDM technology allows the transmission of multiple carrier signals in one single fiber (see Fig. 2.2). The main principle is to merge the multiple carriers spaced one from each other by several GHz. The merging is based on the prism effect using arrayed waveguide gratings, then, the reception is carried out making the inverse. Furthermore, traditional optical WDM systems follow ITU recommendations where each carrier frequency is spaced from another by 50 GHz, this concept is called a 50 GHz-grid where a total of 80 channels are allowed transmitting at ~10 Gbps [8]. However, an extension of this standard is to use smaller channel spacing (25 or 12.5 GHz) to increase the amount of transmitted carriers, these systems are called dense-WDM (DWDM).
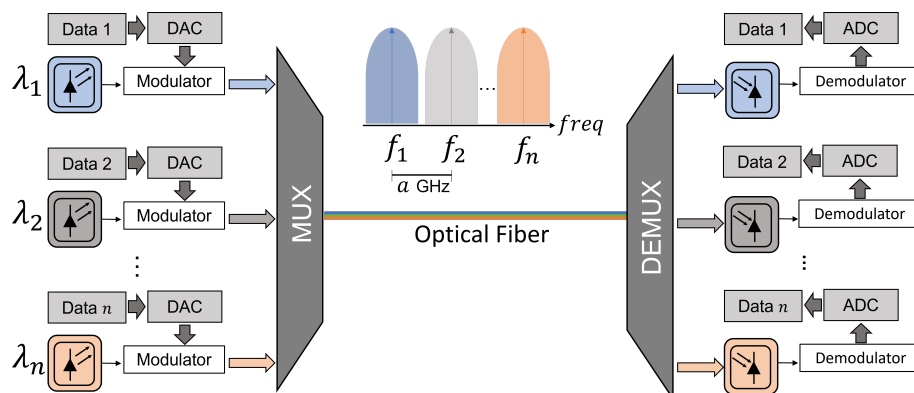
Figure 2.2: Wavelength Division Multiplexing system.

2.1.2   Nyquist-WDM Systems

Usually, traditional WDM systems use fixed modulation formats and data rates for every multiplexed channel, thus, the available spectrum could not be fully used as seen in Fig. 1.1a. Hence, in order to increase the SE, an advanced scenario of WDM systems, called "Flex-grid" has been proposed, where carriers will not be equally spaced, but considering multiples of 6.25 GHz, improving, even more, the SE as it is shown in Fig 1.1b [8]. The new techniques based on dense frequency allocations, enabled the concept of 'super-channel' schemes allowing the transmission of multiple identical channels (with same bandwidth) retaining a maximum of spectral orthogonality for a single user [72, 73]. Nonetheless, having channels at a low spectral spacing (up to the baud-rate spacing), leads to undesirable ICI, thus, the best candidate to obtain an ICI-free operation for the super-channel, is the use of temporal-Nyquist pulses which brings a rectangular spectral shape, as Fig. 2.3a shows [73, 74], plus this waveform allows reduction of the inter-symbol interference (ISI). The base-band pulses are filtered using Root Raised Cosine (RRC) transfer function. However, due

to the limitations of finite digital filter taps, the resultant spectrum does not fall instantly, instead, a roll-off factor induces a window broadening from the ideal data-rate bandwidth as Fig. 2.3b shows. The resultant bandwidth can be calculated using Eq. 1 where R_s is the base-band baud rate in Hz and γ is the roll-off which is a dimensionless value between 0 and 1. The roll-off factor could reduce SE in super-channels due to the spectral spacing being close to the baud-rate.

$$BW = R_s(\gamma + 1) \tag{1}$$
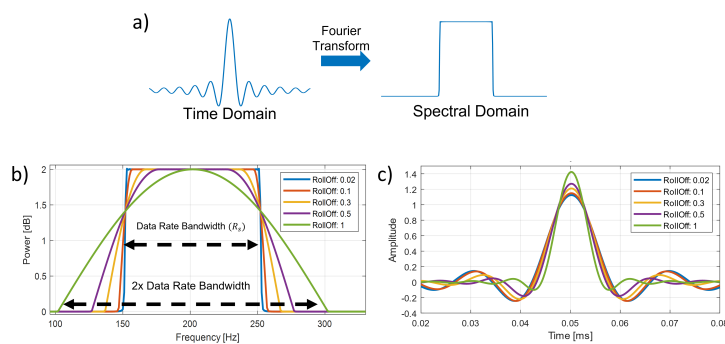
Figure 2.3: a) Nyquist pulse in time and spectral domain. b) Spectrum of Nyquist pulses at different roll-off values. c) Temporal Nyquist pulses at different roll-off values.

Moreover, according to the promising EONs paradigm, in which the assignment of systems resources would follow the dynamic traffic demands, for instance, the adjacent channels will not be identical as in the superchannels, therefore, the frequencies assignments and channel allocation would be carried out without frequency grids using ROADMs (see Fig. 1.4) which consist in arrangements of wavelength selective switches, constant optical spectrum monitoring and optical amplifiers [75].

2.2   Nonlinear and ICI effects in Optical Communications

Two main nonlinear effects can be seen in a constellation diagram: i) the phase noise due to the lack of monochromatic lasers and ii) the Kerr effect given by high launch power applied to avoid attenuation by the fiber. The former results since it is particularly difficult and expensive to generate narrow carrier frequencies, thus, the resultant laser spectrum, has a linewidth associated to the total emitted frequencies. Therefore, considering fiber dispersion effects, the different frequency components have different transmission speeds which generate signal phase mismatches at the receiver [76]. Hence, nonlinear phase noise gives signal distortion which unlike Gaussian noise (see Fig. 1.2a), that creates circular shapes around the ideal symbol position, distorts symbols positions following non-circular shapes (see Fig. 1.2b). On the other hand, the Kerr effect is caused by changes in the refractive index of the fiber core when high launch power is applied. The high launch power is required in order to keep high OSNR in transmissions with complex modulation formats such as m-ary QAM [77]. The Kerr effect leads to signal nonlinearities as i) self-phase modulation (SPM) which alters the phase of the temporal pulse due to its own intensity causing broadening due to

the signal chirp, thus SPM is strongly related to inter-symbol interference (ISI), ii) cross-phase modulation (XPM) that is significant when the pulses of two spectrally close channels get a temporal synchronization, and iii) four-wave mixing (FWM) which is caused when three information channels get spectrally close, generating a fourth wave which alters the phase of the middle signal [77, 78]. The nonlinear effects seen in a constellation diagram cause irregular symbols group shape and centroid jittering. Moreover, the ICI effects are seen when two or more information channels get spectrally overlapped as Fig. 2.4 shows where the gray and black dots are arbitrary received symbols affected by ICI, this, it is one of the most significant issues that future gridless optical networks must resolve. The randomness of the bit sequences and variation in the channel's power, induce distortions in the symbols IQ values that leads to errors in demodulation [79].
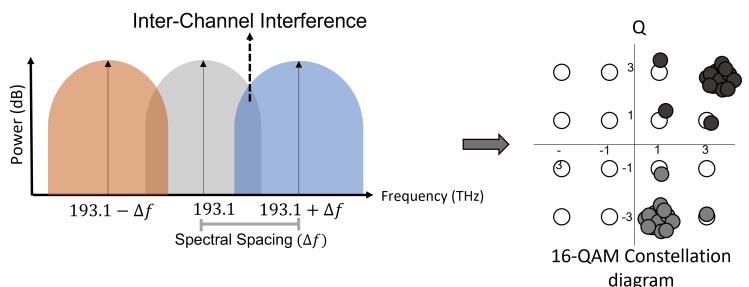
Figure 2.4: Inter-Channel Interference effects on received symbols.

2.3   Digital Signal Processing

Digital signal processing of the down-converted (optical to electrical) received signal presents several advantages in the optical communication systems [71]. It could include clock recovery, IQ imbalance correction, digital equalization and carrier recovery [71]. Specifically, the digital equalization refers to the channel estimation in order to compensate linear and nonlinear impairments. For instance, linear equalization comprehends symbol timing offset correction, ISI mitigation or compensation of chromatic and polarization mode dispersion (PMD) which are linear and time-invariant effects [80]. On the other hand, nonlinear equalization includes the signal impairments mentioned in 2.2. In the last decade, channel estimation using digital back-propagation (DBP) have been proposed for linear and nonlinear signal impairment mitigation. DBP is based on the principle of solving the inverse nonlinear Schrödinger equation (NLSE) which is strongly related to the signal propagation through the fiber [81, 82]. Nonetheless, main drawbacks of DBP relies on its excessive computational cost and the difficulty of applying it in presence of linear PMD [82]. Additionally, the digital equalization could be data-aided, this means that several pilot symbols known by the receiver, are used to adjust the digital filter coefficients based on error calculation between ideal symbols values and affected ones. Furthermore, the non-data-aided equalization uses gradient descent-based methods and is also known as blind-equalization.

2.4   Machine Learning in Optical Communications

ML algorithms allow to create a model of a specific system using historical data of a previous characterization. These models could give predictions of system states and find relation among the data that could be hardly found by traditional methods. Moreover, the increasing number of network monitoring tools has

improved the possibility to obtain historical data of the optical networks [27, 83]. Hence, ML will be a new segment of the DSP module of an optical communication system [83–85] (Fig. 2.5), where the measurements represent different network properties as traffic demand, noise levels, routing, modulation format, carrier frequencies, etc. These measurements feed the historical data block in which the possible network and signal states are labeled in order to supply the training of the ML algorithms. Furthermore, the livestream data will be classified presenting signal and network monitoring as prediction of traffic demands, recognition of modulation format, system fault detection or demodulation based on constellation diagrams analysis [86]. Two main categories are found in the ML-related literature: supervised and unsupervised algorithms.
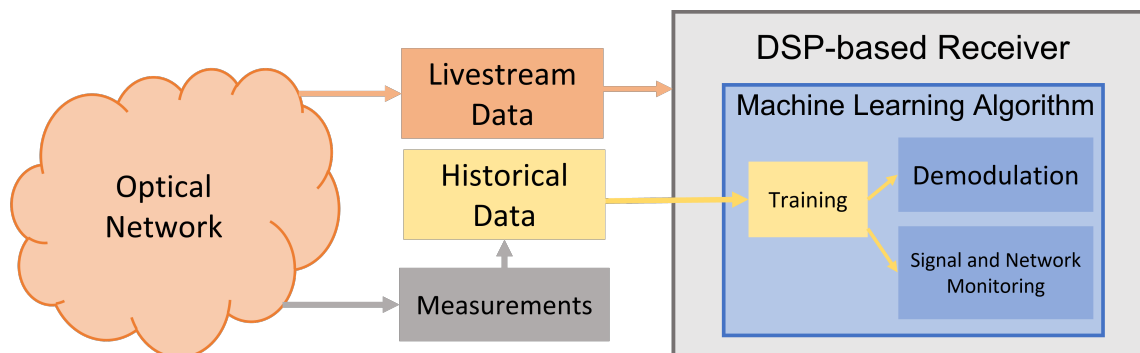
Figure 2.5: Machine Learning-based receiver module in an optical communication system.

2.4.1   Supervised Learning

The goal of these techniques is to classify a certain group of data labeling them from some finite number of possible classes. Its advantages are highlighted by the statistical and mathematical models that separate training data in such a way that prediction or labeling of input data is the most accurate as possible. According to different scenarios, these algorithms can improve performances in decision making of network states (resource assignment or network monitoring) or creating nonlinear boundaries in a constellation diagram for demodulation at the receiver (mitigation of signal impairments). Although, extra memory would be needed at receivers side for storing training data or decision boundaries values according to the training method used, the algorithm could be used as channel estimator similar to other digital equalization techniques. Besides, in most cases, when more data is supplied in the training stage, the accuracy is usually better, nonetheless, there are some scenarios where a minimum amount of data gives a constant accuracy, this behavior can be seen in the use of K-Nearest Neighbor (KNN) algorithm in demodulation [53]. Figure 2.6 shows a typical supervised algorithm functionality. The training data is given to the algorithm where 3 classes are labeled, after training process is finished, the calculated threshold will classify the input data.

2.4.2   Unsupervised Learning

Unlike supervised training, the labels or tags are not given in the unsupervised learning, the algorithms group the data in different clusters based on the similarity of its features and could give labels after a training process. Although the strength of these algorithms is seen in the capacity of finding new classes, some of them can be forced to work with a finite number of classes, for example, in demodulation, where the number of clusters is given by the modulation format, every received symbol is classified into one single

Training Data
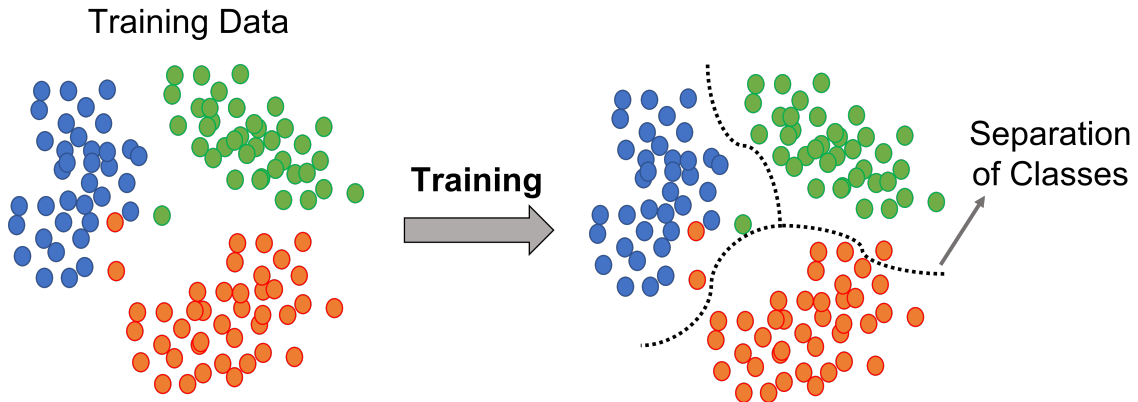
Figure 2.6: Typical supervised Machine Learning algorithm functionality.

class (modulation index) [15]. Figure 2.7 shows an unsupervised algorithm functionality where at first stance, unlabeled data is given, then, the algorithm identifies 3 different groups or classes. One of the most known unsupervised algorithms is the k-Means, which after incoming data enters for classification, measures distance to every existing cluster centroid and closest is labeled, then, the centroid position is updated by the mean of total data belonging to that cluster, besides, different algorithms using this type of classification logic differ in the distance calculation. Moreover, unsupervised learning counts with a different subcategory called fuzzy clustering algorithms , which gives to every analyzed data, a membership degree to every existing cluster, using the same distance-centroid logic.
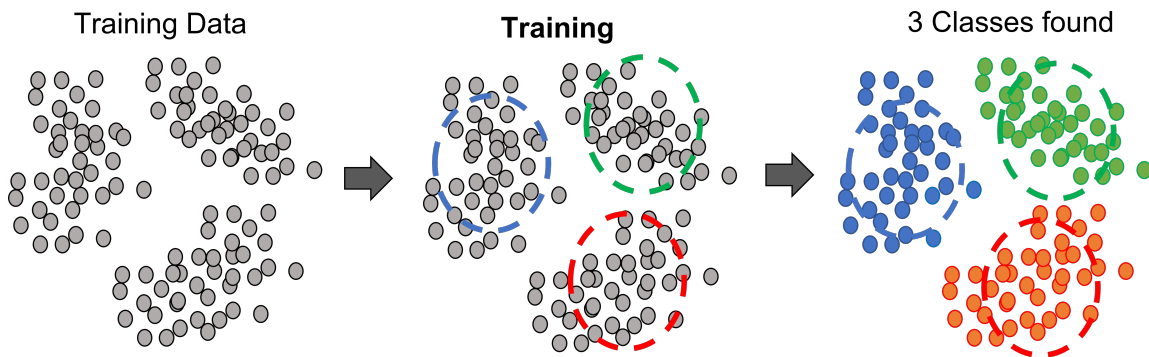
Figure 2.7: Typical unsupervised Machine Learning algorithm functionality.

3   Experimental and Simulation Setup

3.1   Simulation

Using the specialized software VPIDesignSuite v11.2, the numerical setup (Fig. 3.1) simulates a WDM OAN system composed of three channels modulated at 16 GBd. For every WDM carrier, the transmitted data bits are generated by encoding equiprobable random bit sequences of 2^{18} length. The generated bits are mapped in a Gray-coded 16-QAM symbols. Then, the signals are upsampled to 16 samples/symbol and pulse shaped with a root-raised cosine (RRC) filter with a roll-off factor of 0.1. The transmitting lasers count with linewidth of 100, 500 and 1000 kHz and the launch power is controlled by an ideal amplifier ensuring 0, 3, 6, 9 and 12 dBm that excites nonlinearities in a 40 km span of single-mode fiber which has an attenuation parameter of 0.2 dB/km [87]. Additionally, the fiber nonlinear index and dispersion are chosen as 2.6 W-1km-1, and 16 ps/nm/km, respectively. Additive white gaussian noise (AWGN) generator is yield to vary the OSNR with values from 5 to 31 dB. At the receiver, the center channel is chosen for evaluation using local oscillators with identical transmitter configuration where the analog-digital converter (ADC) uses bipolar function based on eight bits to digitalize the signal. The DSP modules based on MATLAB® executes IQ imbalance correction, chromatic dispersion compensation, clock timing recovery, digital multi-modulus algorithms (MMA) equalization of 25 taps, finishing with carrier phase and frequency recovering.
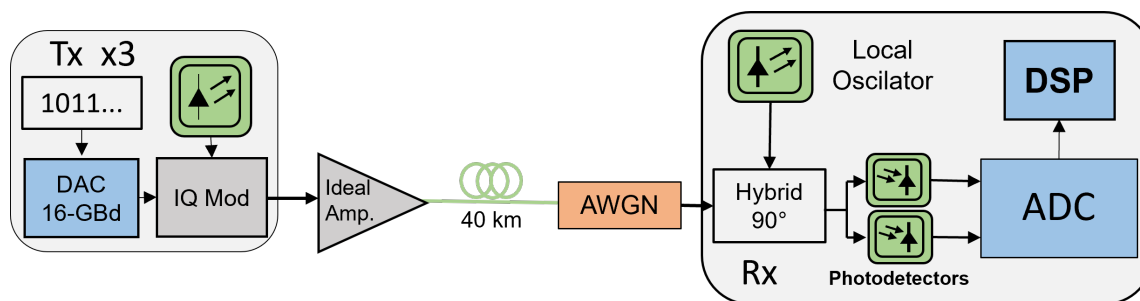
Figure 3.1: Simulation setup of an optical access link.

3.2   Experimental

3.2.1   16 GBd

Figure 3.2 shows the experimental setup from the CPqD labs that is based on three tunable lasers arrangement with 100 kHz linewidth that are used to generate 3×16 Gbaud 16-QAM Nyquist WDM system. The digital-analog converter (DAC) operates at 64 GSa/s with an electrical bandwidth of 14 GHz. The center and side frequency channels are generated using different DACs to produce uncorrelated signals. the four DAC outputs are amplified by four RF drivers with 32 GHz bandwidth before being sent to the IQ modulator. Two 133152 random symbols sequences, mapped in 16-QAM, were generated in Matlab® for X-Y polarizations. Root raised cosine filter with roll-off factor of 0.1 is used to implement Nyquist pulse shaping, resulting in spectral widths of 17.6 GHz for each channel. The channel spacing is varied shifting the laser frequency, from 15 GHz to 18 GHz to obtain different percentage of spectral overlapping: 18 GHz (close channels), 17.6 GHz (0% overlap), 17 GHz (6.8% overlap), 16.5 GHz (12.5% overlap), 16 GHz (18% overlap), 15.5 GHz (24% overlap) and 15 GHz (30% overlap), including the single channel case. In a back-to-back scenario,

ASE (amplified spontaneous emission) noise is added to yield the OSNR ranging from 14.3 to 36.3 dB. At the receiver side, a wideband coherent receiver of electrical bandwidth with 40 GHz is used to detect the optical carriers. The central carrier is the channel of interest. The received signal is digitized by an 80 GSa/s real-time oscilloscope, where the 4 inputs, namely the I and Q in X and Y polarizations, have a bandwidth of 35 GHz each. This is followed by offline DSP in MATLAB®.

Figure 3.2: Experimental setup of the 3×16 GBd gridless Nyquist-WDM system.

3.2.2   32 GBd

Figure 3.3 shows the experimental setup from the GeorgiaTech labs which is based on three tunable lasers arrangement with 25 kHz linewidth that are used to generate 3×16 Gbaud 16-QAM Nyquist WDM system. The digital-analog converter (DAC) operates at 88 Gsamp/s with an electrical bandwidth of 32 GHz. The center and side frequency channels are generated using different DACs to produce uncorrelated signals. The four DAC outputs are amplified by four RF drivers with 32 GHz bandwidth before being sent to the IQ modulator. Root raised cosine filter with roll-off factor of 0.1 is used to implement Nyquist pulse shaping, resulting in spectral widths of 35.2 GHz for each channel. The channel spacing is varied shifting the laser frequency, from 30 GHz to 37.5 GHz to obtain different percentage of spectral overlapping, including the single channel case (50 GHz). Transmission over 250 km of standard single-mode fiber [87] is carried on including B2B and ASE noise is added to yield the OSNR ranging from 15 to 32 dB. At the receiver side, a wideband coherent receiver of electrical bandwidth with 40 GHz is used to detect the optical carriers. The central carrier is the channel of interest. The received signal is digitized by an 80 GSa/s real-time oscilloscope, where the 4 inputs, namely the I and Q in X and Y polarizations, have a bandwidth of 35 GHz each. This is followed by offline digital signal processing in MATLAB®.
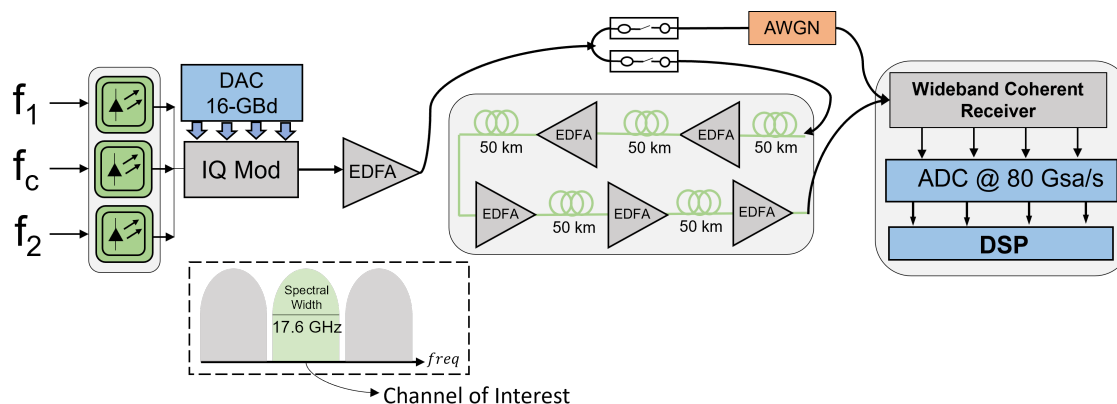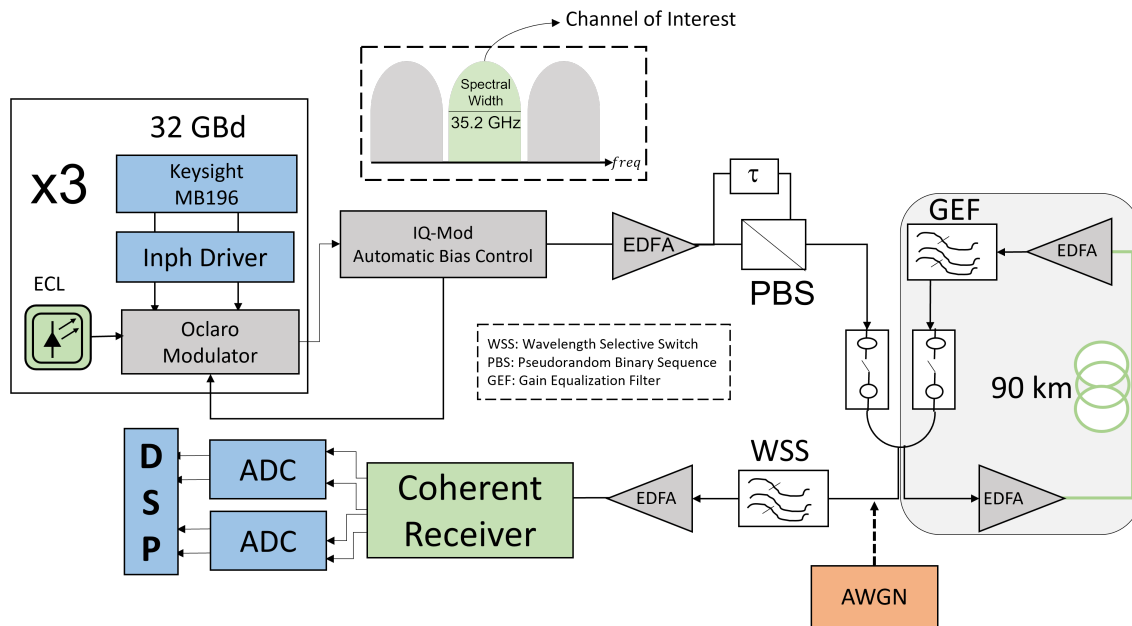
Figure 3.3: Experimental setup of the 3×32 GBd gridless Nyquist-WDM system.

4 Proposed Methods

The proposed methods aim to improve the gridless networks performance by i) estimation of the spectral overlapping by evaluating the received symbol frames and ii) minimizing the ICI effects by applying non-symmetrical demodulation techniques using ML algorithms. Figure 4.1 shows where the proposed methods would be operating in the digital signal processing (DSP) module at the scenario of a gridless optical network. The methods use the received IQ symbols affected by ICI and then, determinate the spectral spacing of the adjacent channels or to mitigate its ICI or nonlinear effects.
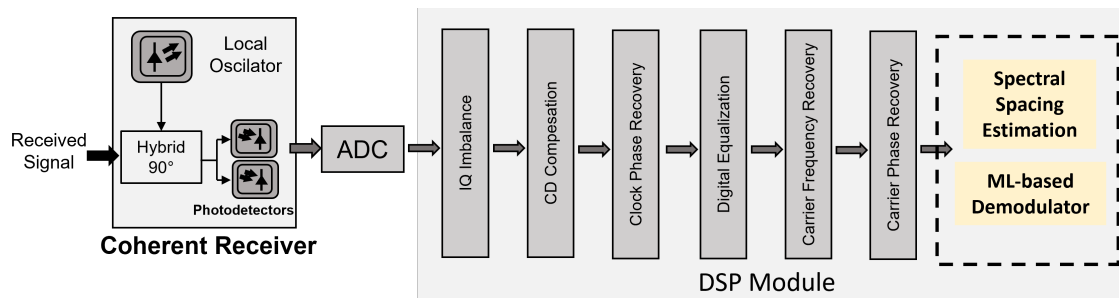
Figure 4.1: Allocation of the proposed methods in a digital coherent receiver with digital signal processing module.

4.1 Spectral Spacing Estimation

Two methods for spectral estimation are proposed. Both of them consist in two stages: i) feature extraction from constellation diagrams using an unsupervised (fuzzy clustering algorithm) learning algorithms and ii) spectral spacing estimation carried out by a supervised learning algorithms as Fig. 4.2 shows.

4.1.1 Counting Vectors

Unlike the boolean classification which assigns the data a "1" (belonging) or "0" (not belonging) to certain class, the fuzzy clustering algorithms give the data a membership degree to every existing class. For
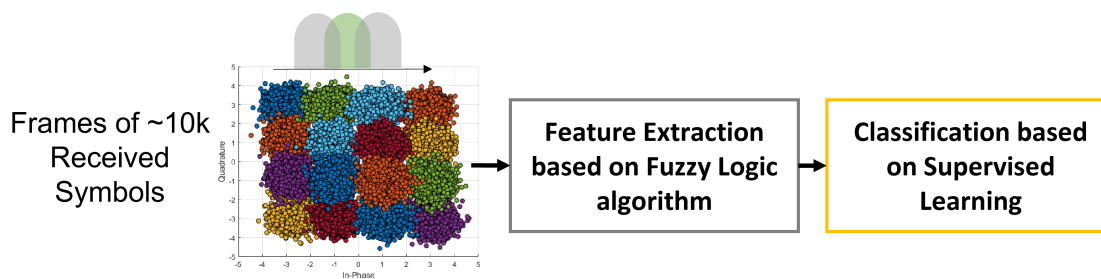
Figure 4.2: Spectral spacing estimation of a diagram constellation.

instance, Fig. 4.3a shows the assignation of membership degrees to a received symbol in a 16-QAM constellation diagram. Note that the closest modulation index (9) has the higher membership degree (0.6221), on the other hand, the farthest (2) has the lower membership degree (0.0047) according to the Euclidean distance. Whereby, the fuzzy clustering algorithms differ in the kind of distance used to classify the data. The Euclidean distance assumes that the data is isotropically distributed, i.e. it will treat each feature equally resulting on an analysis of circular-shaped clusters (see Equation (2)) where $x_j = x_I + x_Q$ is the *jth* analyzed data point respect to the cluster c_i and B is the identity matrix, moreover, in the Mahalanobis distance (see Equation (3)), B_i (Equation (4)) is the covariance matrix of the classified where μ_{ij} is the membership degree of data point j to the cluster i, in other words, the algorithm measures the correlation among the data and relaxes the assumption of a Gaussian distribution, this allows the identification of different clusters shapes. Figure 4.3b shows an illustrative comparison between the Euclidean and Mahalanobis distances. Thus, both algorithms are tested by its data partition to carry out the demodulation.

$$\left\| x_j - c_i \right\|_B^2 = \left(x_j - c_i \right)^T B \left(x_j - c_i \right) \tag{2}$$

$$\left\| x_j - c_i \right\|_{B_i}^2 = \left(x_j - c_i \right)^T B_i \left(x_j - c_i \right) \tag{3}$$

$$B_i = \frac{1}{N} \begin{bmatrix} \sum_{j=1}^{N} (\mu_i x_I^2)_j & \sum_{j=1}^{N} (\mu_j x_I x_Q)_j \\ \sum_{j=1}^{N} (\mu_i x_Q x_I)_j & \sum_{j=1}^{N} (\mu_i x_Q^2)_j \end{bmatrix} \tag{4}$$
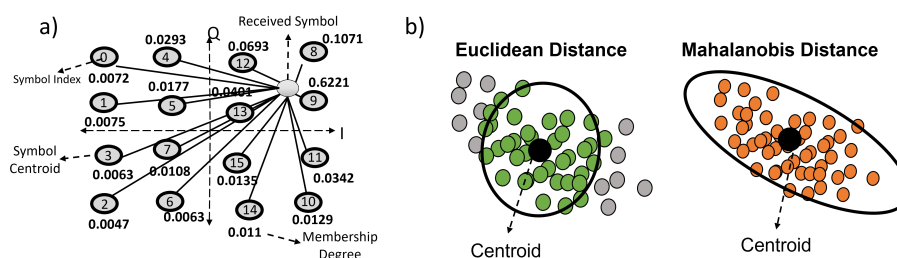
Figure 4.3: a) Fuzzy clustering algorithm applied to a received symbol in a constellation diagram. b) Illustrative differences between Euclidean and Mahalanobis distances.

Fig. 4.4 shows the feature extraction using an hypothetically constellation diagram of 4 symbols. Each black dot represents the ideal symbol position and the initial centroid to every cluster. When a received symbol is analyzed, the algorithm will calculate the respective distances to every centroid. Then, according to the membership degree, we can relate how far or close the received symbol is to its correct modulation index. We carried out a counting of the symbols which its correct modulation index is the closest, second closest, third closest and the rest resulting in *"counting vectors"* of 16 features (for 16-QAM) for frames of 10k symbols. The counting vectors can characterize how ICI effects distorts the symbols position. We tested two fuzzy clustering algorithms: Fuzzy c-Means (FCM) which uses Euclidean distance and Gustafson-Kessel Means (GKM) that uses Mahalanobis distance. Both algorithms' centroids were fixed to be 16 and its initial values are the ideal 16-QAM constellation. Additionally, the fuzzifier parameter which relates the

impact of the distances to the membership degree calculation, is set to standard value of 2. Finally, resultant counting vectors are taken as inputs to the k-Nearest Neighbors (KNN) or Support Vector Machine (SVM) algorithm to estimate the spectral spacing. We used these algorithms due to their simplicity and versatility in nonlinear classifications. The resultant data data was divided into 70% training and 30% testing.
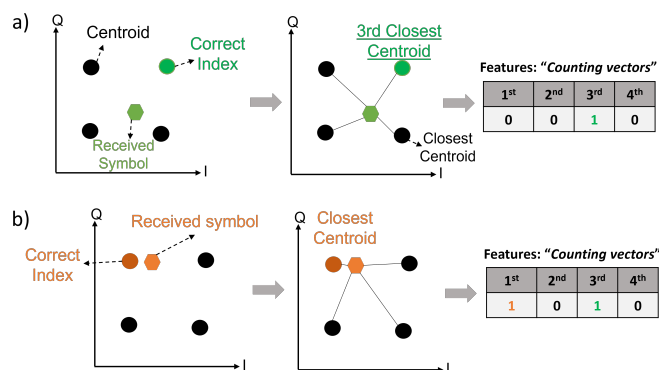
Figure 4.4: Feature extraction of a constellation diagram of 4 symbols when a) the correct modulation index is the 3rd closest to the received symbol and b) when the correct modulation index is the closest.

4.1.2   Cluster Validation

The cluster validation refers to the evaluation of each found clusters in certain data set. Thus, the cluster validation is usually used to find the best number of clusters. However, in the case of a constellation diagram, the number of clusters is set according to the number of symbols (16 for 16-QAM). In this sense, it checks the quality of the results of the fuzzy clustering algorithms on how well grouped are the data according to the membership degrees values, thus, information of the actual structure of the data could be given [88]. For instance, in a constellation diagram, it can be noted well-defined clusters at high OSNR values (low noise levels), this could be related to BER estimation or quality of the signal. On the contrary, diffuse clusters are seen at low OSNR values or when the signal is affected by high nonlinear effects. Here, the validation aims to evaluate different partitions of the data (FCM and GKM) and not the number of clusters. Thus, three validation indexes were considered: i) ratio between compactness and separation of clusters (Fig. 4.5a), ii) fuzzyness of each cluster (Fig. 4.5b) and iii) overlapping among the clusters (Fig. 4.5c), which are described by the following equations:

Being μ_{ij} the membership degree of data point j to the cluster i.

- Compactness and Separation: Partition Index

$$PI(c) = \sum_{i=1}^{C} \frac{\sum_{j=1}^{N} (\mu_{ij})^2 \left\| x_j - c_i \right\|^2}{N \sum_{k=1}^{c} \left\| c_k - c_i \right\|^2} \tag{5}$$

- Fuzzyness: Classification Entropy

$$CE(c) = -\frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{N} \mu_{ij} \log \mu_{ij} \tag{6}$$

- Overlapping: Partition Coefficient

$$PC(c) = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{N} \mu_{ij}^2 \tag{7}$$

A signal quality mapped in an diagram constellation could be characterized using the three indexes together. However, there is an evident correlation between the partition coefficient and the classification entropy, thus, in order to avoid information redundancy, principal component analysis (PCA) is applied to the three features 4.5d. The PCA involves a mathematical procedure that performs dimension reduction by replacing the variables for principal components which are orthogonal to each other and organized by descending variance [89]. Therefore, it could be achieved two features with all the information of the three mentioned indexes which are inputs to the classification algorithm. The explained covariance which show how much information the PCA components have, is shown in Fig. 4.5e.
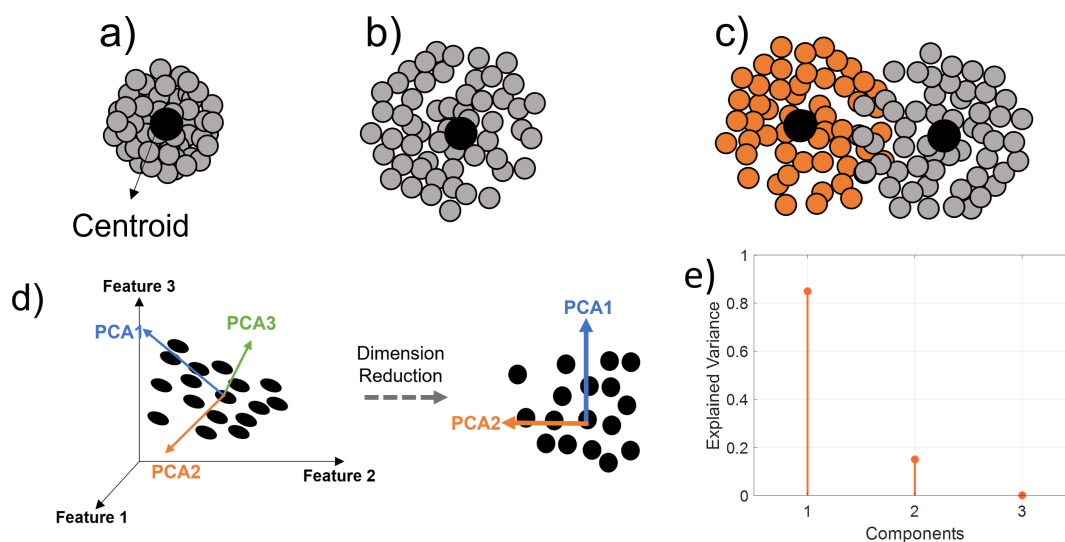
Figure 4.5: a) Compact cluster. b) Fuzzy cluster. c) Overlapped clusters. d) Dimension reduction using principal component analysis.

4.2   ICI and Nonlinear Effects Minimization

The minimization of ICI and nonlinear effects is carried out using supervised learning algorithms to demodulate 16-QAM constellations. The main interest of using these algorithms is due to i) its adaptability and availability of nonsymmetrical classification of distorted symbols in a constellation diagram and by ii) the late interest that the supervised learning have gained in the last years for nonlinear classification [27].

4.2.1 Algorithms Adaptation

(a) Support Vector Machine

SVM is a supervised learning algorithm that adopts hyperplanes as thresholds in order to separate the data classes [90]. These planes are constructed by using support vectors subject to a nonlinear function. These support vectors are training points usually located in the data-space limits among the classes (see Fig. 4.6a). On the other hand, the nonlinear function is based on kernels which for an m-QAM constellation, the best choice of kernel comes to be the Gaussian-radial-basis function due to it can bring semi circular-shaped thresholds. Nonetheless, the number of support vector considered to calculate the thresholds can be varied by the Gamma parameter with which a higher or lower curvature adjustment is controlled avoiding overfitting issues [90]. For example, Fig. 4.6b shows the thresholds calculation for a single symbol in a constellation diagram affected by nonlinear effects with a high Gamma value, whereas Fig. 4.6c shows the same classification with a low Gamma value.
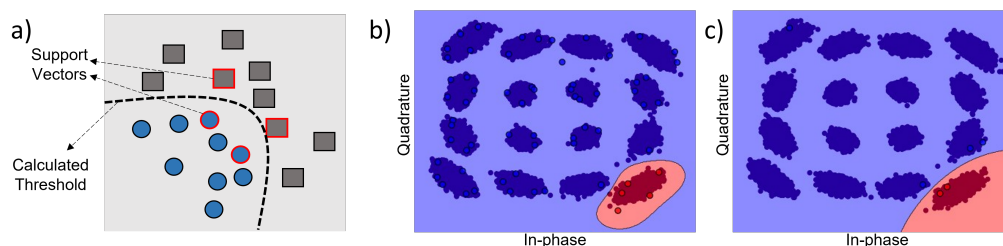
Figure 4.6: Thresholds calculation for a single symbol in a constellation diagram affected by nonlinear effects where the Gamma parameter is a) high and b) low.

Moreover, in a multiclass classification, the SVM proceeds as 'one vs rest', by doing a two-class thresholds calculation as can be seen in the example of the Fig. 4.7, where the 16-QAM format is susceptible to 6.8% spectral overlapping and 36.3 dB-OSNR. Hence, the demodulation is made according to the position of the received symbol in the partitioned constellation diagram.
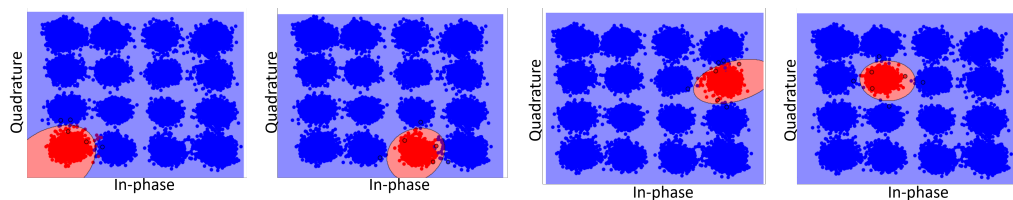
Figure 4.7: 'One vs rest' thresholds calculation for different 16-QAM symbols.

(b) Artificial Neural Networks

Figure 4.8 shows the adopted single-hidden layer ANN [91]. We explored a single-hidden layer network for aiming lower computational complexity. On the contrary, we avoided deep neural networks that need high-performance computing techniques [91]. The input layer X corresponds to the in-phase (x_I) and quadrature (x_Q) components of each received symbol, whilst the output layer $Y = (y_1, .., y_{16})$ brings the modulation index (1-16 for the 16-QAM format) by following a softmax classification. The

hidden layer contains n neurons subject to the rectified linear Unit (ReLU) activation function. Although ReLU presents equal performance than the benchmarking hyperbolic tangent function, ReLU functions has a lower computation complexity [92]; hence, the activation function relates the product between inputs and weights $\omega(k, j)$ (k indicates input or hidden layer and j goes from 1 to n) and then, sums the result with bias value $b(1, j)$. The training process is carried out by introducing received symbols frames X associated with its respective modulation index Y. Then, the weights and biases are adjusted based on the backpropagation (BP) algorithm [91]. The BP algorithm connects all the neurons carrying on a prediction with the initial weights and biases values in a forward way, namely by starting from the inputs X to the hidden layer and then, from the hidden layer to the outputs Y. The difference between outputs and predictions are computed by following a backward process (weights and biases values are updated). The adjusting of these parameters is iteratively made according to an
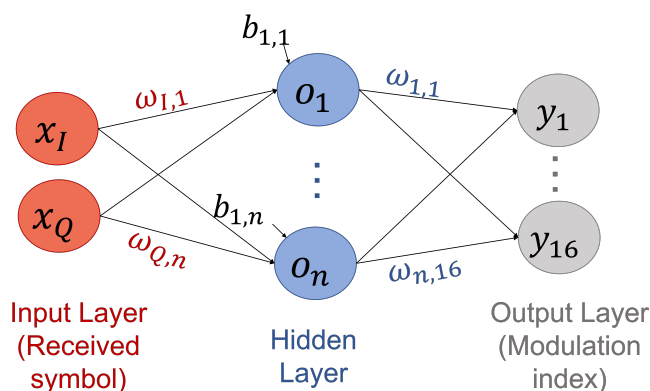
Figure 4.8: Artificial Neural Network used to demodulate 16-QAM symbols.

epoch parameter, which means how many rounds the BP algorithm is executed. Besides, the updating rate of weights and biases follows optimizer function adam, which in comparison to classical stochastic gradient descent, does not maintain a single learning rate for all weights and biases [93]. Therefore, at the receiver side of the optical communication system, the ANN model (calculated weights, biases, and trained topology of the network) will be stored to predict modulation indexes of the incoming received symbols.

(c) k-Nearest Neighbors:
   KNN is one of the simplest supervised algorithms [90]. Figure 4.9 shows the demodulation using KNN in a 16-QAM constellation, where the colored points are the training symbols affected by ICI effects. Considering a received symbol (Fig. 4.9a) and using the Euclidean distance, the algorithm identifies the k closest training symbols; then, it compares the modulation index of every considered training symbol to assign the most common one to the received symbol (Fig. 4.9b). Thereby, in the receiver side, all the training symbols will be stored in a memory in order to calculate the respective distances from the received symbol.
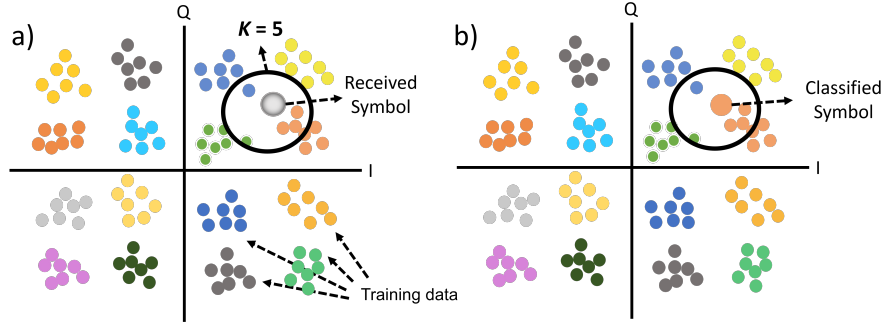
Figure 4.9: KNN classification using k=5 for a received 16-QAM symbol. a) Comparative of classes before classification. b) Classification.

4.2.2 Training Strategy

We propose a training strategy that will be useful for future gridless networks implementations in order to mitigate ICI effects. It uses training symbol sequences added to the transmitted signal when changes in spectral spacing are detected. Consequently, this method would require a diagnostic tool for detection and estimation of spectral spacing as mentioned previously in Section 4.1. We called this strategy the updating strategy because every time when a change in carrier frequencies is detected, the training symbols affected by the new distortions will be updated. Fig. 4.10 illustrates this strategy, when changes in spectral spacing occurs, the receiver should send an 'alert' to the transmitter and the training sequence will be again transmitted. The training symbols affected by the new spectral spacing will be used to update the support vectors that create the demodulation thresholds for SVM, to update the neurons weights and biases of the ANN algorithm, and for KNN, the symbols will be stored to perform the digital demodulation.
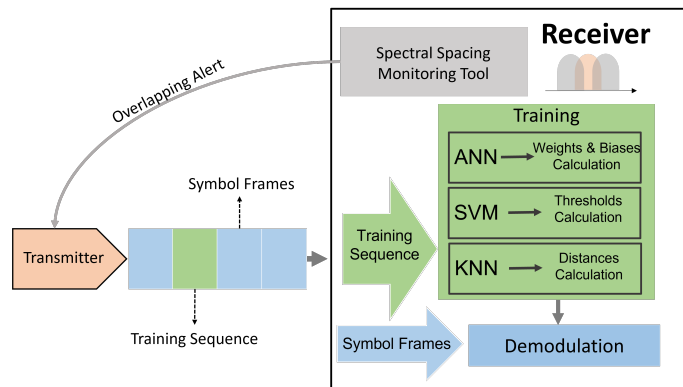
Figure 4.10: Training approach by employing updating training sequences.

In order to reduce computational costs of all strategies, we calculated the lowest training data without penalizing the BER metric. Thus, based on the average of ten iterations where the frames of 100k symbols are randomized, we estimated the BER for various scenarios. For the training strategy, Fig. 4.11 shows the BER vs. training length by considering three representative scenarios that have BER values close to a

common hard-decision forward error correction (FEC) limit of $log_{10}(BER) = -2.4$. The FEC techniques uses data redundancy for error correction based on codification with which it is possible to obtain BER values of $log_{10}(BER) = -9$ [94]. In order to take into account ICI and/or additive noise effects, the chosen scenarios correspond to a) a single channel with OSNR = 19.3 dB, b) 12% spectral overlapping with OSNR = 21.3 dB, and c) 18% spectral overlapping with OSNR = 36.3 dB. It can be seen that the BER is lower whilst training length increases. Besides, ANN and SVM are more dependent on training data than KNN to minimize the system performance. Moreover, the three ML algorithms allow lower BER than traditional demodulation, which uses symmetrical decision thresholds in the constellation diagram. Additionally, all above observations occur in the three scenarios., by allowing the generalization of the method. Finally, it is noticed that for the three ML algorithms, <7000 symbols are required in order to get the best BERs.
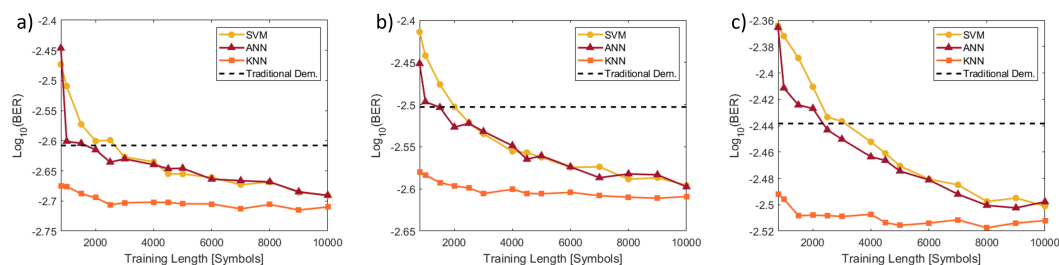
Figure 4.11: BER versus training length for characterization method in (a) single channel at 19.3 dB, (b) 12% spectral overlapping at 21.3 dB, and (c) 18% spectral overlapping at 36.3 dB

4.2.3   Hyperparameter Selection

In this section we briefly discuss and show how the main hyperparameters of SVM, ANN and KNN algorithms were chosen for a minimum BER value in order to reduce system complexity. Although all the hyperparameters were chosen using gridsearch, we present here the BER impact changing the main hyperparameters. Thus, for the ANN algorithm, Fig. 4.12a shows the BER as a function of the hidden layer neurons (n) for the mentioned scenarios i) single channel at high noise levels, ii) 12.5% spectral overlapping at high noise level and iii) 18% spectral overlapping at low noise. Once again, theses cases were selected in order to obtain a BER value around the FEC limit of $log_{10}(BER) = 2.4$. It can be seen that from 13 neurons, no differences in BER are evidenced. On the other hand, the analysis of the SVM's hyperparameters is carried out based on the type of kernel we used which is Gaussian−radial−basis suitable for diagrams constellations. The Gamma hyperparameter (Fig. 4.12b) relates how high or low the curvature of the Gaussian threshold will be therefore, the best Gamma value is 0.1. Finally, the best k (amount of neighbors) value for KNN is 9 (Fig. 4.12c). Moreover, the choosing of the hyperparameters is carried out looking for best BER performance and for minimum computational cost. For example a higher number of neurons leads to a higher number of weight and biases calculation and a higher k value, means more class comparison among training points.
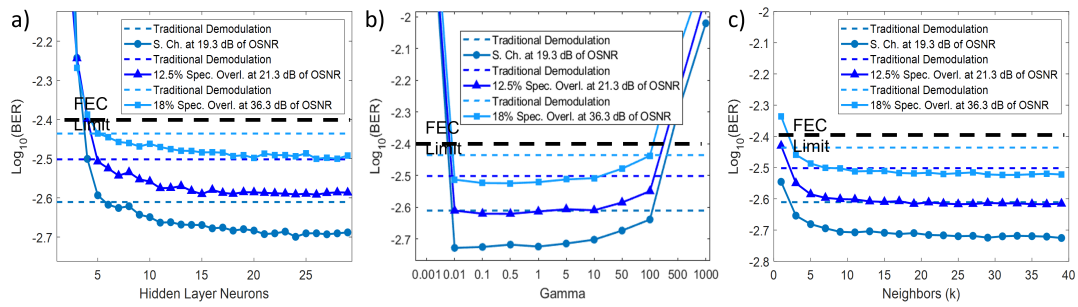
Figure 4.12: BER vs a) Neurons for ANN, b) Gamma for SVM and c) Neighbors (k) for KNN algorithms.

5 Results and Discussions

5.1 Spectral Spacing Estimation

Figure 5.1a shows the spectral overlapping between channels that are spectrally spaced less than the baud rate. The spectral overlapping percentage is calculated using Eq. (8) where R is the baud rate and C_{sp} is the channel spacing. These overlapping values could be negative which means the channels are not overlapped. Furthermore, Figs. 5.1b and 5.1c shows the resultant overlapping cases using (8) applied to the 16 GBd and 32 GBd data in which the spectral spacing were varied in steps of 0.5 GHz. Besides, the overlapping percentages were grouped in order to carry out different estimations, for instance, in the 32 GBd setup (Fig. 5.1c), it is made five different configurations: i) for 2 classes, the classifier detects if any constellation diagram was affected by ICI or not, ii) 3 classes means two overlapping levels are considered, iii) in the case of 4 classes, three overlapping levels are considered, iv) for 5 classes, the single channel scenario (50 GHz of spectral spacing) is contemplated as a separated case, finally, v) all separated overlapping cases are also classified. On the other hand, for the 16 GBd setup, it is also considered a binary classification, two overlapping levels and the case of all possible overlapping cases.

$$Ov = 2 \times \frac{R - C_{sp}}{R} \tag{8}$$
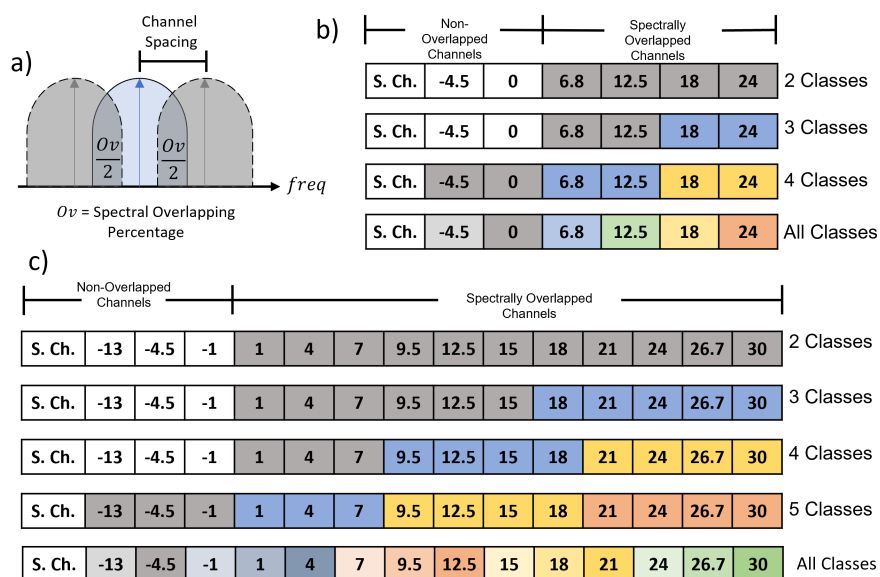
Figure 5.1: a) Spectral overlapping according to the channel spacing. Grouping of different spectral overlapping scenario for b) 32 GBd setup and c) 16 GBd setup.

After the counting vectors or validation indexes are calculated using any of the unsupervised algorithms (FCM or GKM) described in Section 4.1, we tested two classifiers: KNN and SVM. Thus, the results are presented comparing performances in terms of accuracy percentages. Figures 5.2a, 5.2b and 5.2c shows the accuracy performance using counting vectors features vs each grouping case for the scenarios 16 GBd B2B,

32 GBd B2B and 32 GBd with fiber transmission, respectively. Moreover, the OSNR is given as a feature or not.

Figure 5.2: Accuracy vs classes using counting vectors in a) 16 GBd setup, b) 32 GBd setup in B2B scenario, c) 32 GBd setup with fiber transmission scenario.

The experimental results show that it is possible to determinate if any constellation diagram is affected by ICI effects using classification after the counting vectors are extracted. For the 16 GBd setup (Figs. 5.2a), accuracies up to ~84% were achieved using any of the proposed method with SVM or KNN classification when only 2 classes are considered and knowing a priori the OSNR value. The importance of an OSNR knowledge is enlightened as ~80% of accuracy is achieved for 4 classes. Besides, the classification performance by KNN is slightly better than the SVMs even if the OSNR is not included as a feature. On the other hand, when the OSNR is not known by the classifier, accuracies up to ~77% were achieved. Furthermore, the classification performance using the 32 GBd B2B data (Fig. 5.2b) was similar to the 16 GBd (Fig. 5.2a) results, for instance, KNN preented better accuracy than SVM. Whereas, in a transmission over 250 km of fiber (Fig. 5.2c) the classification was better by obtaining ~92% of accuracy in a binary classification, knowing the OSNR. In this case, the KNN and SVM obtained same accuracy of ~85% in the case of 5 classes. Finally, the classification using the features extracted by FCM was better than the features extracted using GKM in the case of B2B scenarios, this is because the distortions tend to follow circular shapes than the distortions in fiber transmission which are nonsymmetric due to the fiber nonlinearities.

Figures 5.3a to 5.3c depict the classification accuracy using the validation indexes features vs the number of classes according to the grouping in Fig. 5.1. The results show that it is possible to determinate if any constellation diagram is affected or not by spectral overlapping effects. It can be seen that the use of SVM classifier obtained better performance than KNN in the 32 GBd scenario with fiber transmission or not (Fig. 5.3b and 5.3c, respectively), for example, in the case of five classes for the 32 GBd with fiber data, the SVM classifier obtained ~87% accuracy, whilst the KNN achieved ~81%. Though, in the case of 16 GBd data, SVM and KNN showed same performance, besides, accuracy up to 85% was obtained in the case of 4 classes.

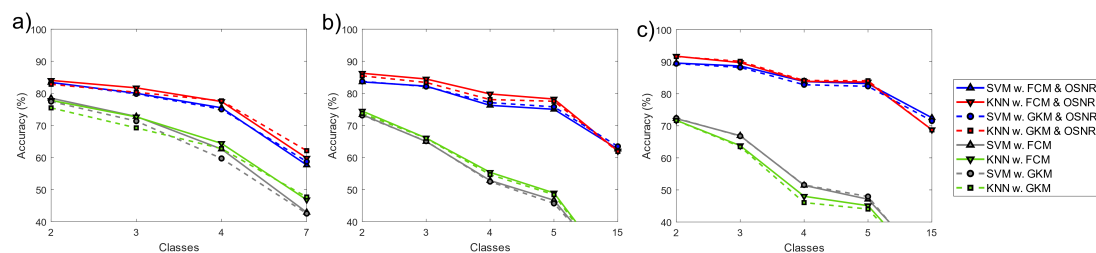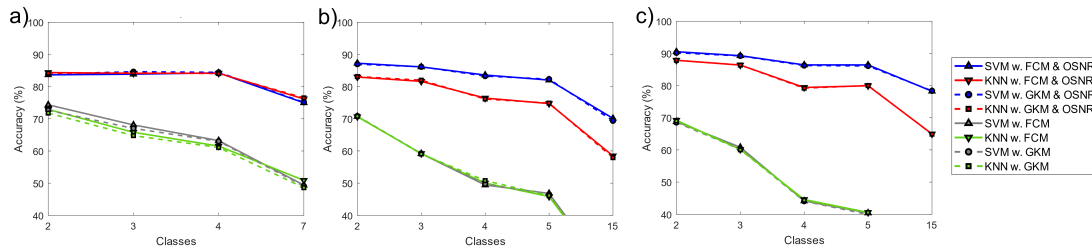Figure 5.3: Accuracy vs classes using validation indexes in a) 16 GBd setup, b) 32 GBd setup in B2B scenario, c) 32 GBd setup with fiber scenario.

Additionally, the use of FCM or GKM partitions for feature extraction did not show variation in terms of accuracy. Figures 5.4 and 5.5 show the confusion matrix when 3 and 4 classes are considered, respectively. Specifically, the confusion matrices shows the classification using KNN. Also, Fig. 5.4a and Fig. 5.4c show the cases when the features were extracted using FCM with and without OSNR, respectively. On the other hand, Fig. 5.4b and Fig. 5.4d show the cases when the features were extracted using GKM with and without OSNR, respectively. Moreover, the class 1 represents that there is not spectral overlapping, whilst the class 2 and class 3 represent low and high spectral overlapping, respectively. Specifically, the classification of the class 3 (high overlapping) is mistakenly classified as class 1 (no overlapping) using FCM features (see Fig. 5.4a). Whereas, using GKM features (Fig. 5.4b), this misclassification is equally seen as class 1 and class 2. This is due to the circular assumption of cluster that FCM gives. Besides, when the OSNR is known by the classifier (5.4c and 5.4d) it can be seen that the easiest class to classify is the one of high spectral overlapping cases, this is because the ICI effects increase the clusters fuzzyness more than low OSNR effects; and those distortions are most evidenced in the first and second class.
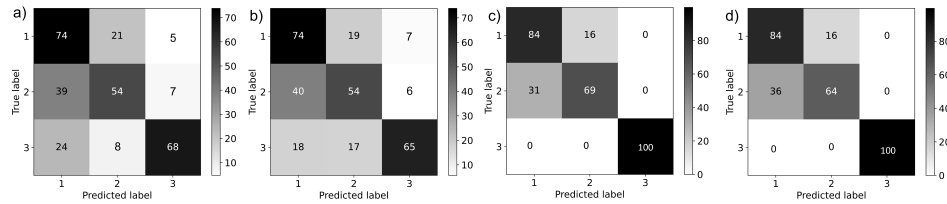
Figure 5.4: Confusion matrix by using validation indexes as features and classifying with KNN for the case of three classes using 16 GBd data, where: a) FCM without OSNR. b) GKM without OSNR. c) FCM with OSNR and d) GKM with OSNR.

Furthermore, Fig. 5.5 shows the confusion matrices when 4 classes are considered. It can be seen the contrast of an OSNR knowledge by the classifier comparing for instance Fig. 5.5a (OSNR not known by the classifier) and Fig. 5.5c (OSNR known by the classifier), it is shown that single channel scenario and highly spectral overlapping cases are almost perfectly classified, hence, the confusion is found between close channels scenarios and low level of spectral overlapping.

In this sense, Figs. 5.6a and 5.6b shows the mapping of the data points where PCA is applied when the OSNR is not known and when it is known by the classifier, respectively. It can be seen that the data of high
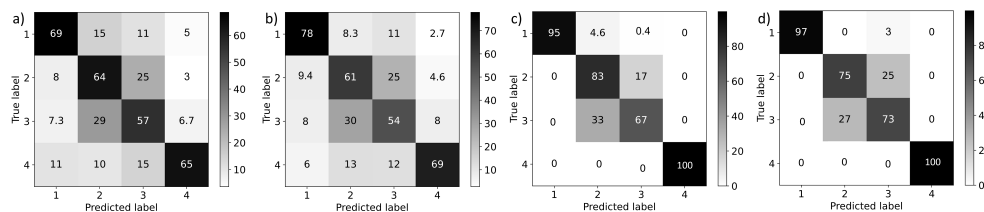
Figure 5.5: Confusion matrix by using validation indexes as features and classifying with KNN for the case of four classes using 16 GBd data, where: a) FCM without OSNR. b) GKM without OSNR. c) FCM with OSNR and d) GKM with OSNR.

spectral overlapping is more likely separable than the cases of low spectral overlapping and close channels cases, this when the OSNR is known (Fig. 5.6b), however, when it is not known, the data could achieve similar positions even if the channels are overlapped, for example, blue dots (single channel) are seen close to red dots (high spectral overlapping), this is due the similarity of gaussian noise and ICI effects.
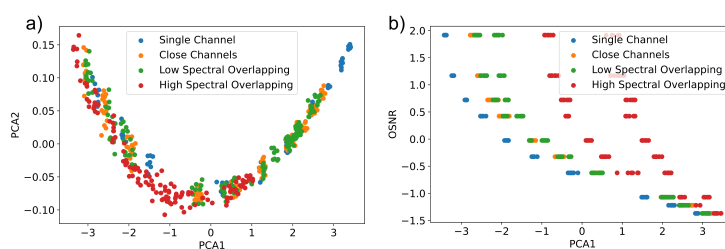
Figure 5.6: Mapping of data after PCA is applied to features extracted using the validation indexes when: a) OSNR is not known and b) OSNR is known by the classifier.

Table 5.1 shows accuracy comparison between the two proposed methods when the OSNR is known by the classifier. It can be seen that in B2B scenarios, the use of the validation indexes presented better performance than the counting vectors by about ~0.7%, moreover, this difference is higher with the increment of classes. On the other hand, when fiber transmission is considered, the counting vectors showed slightly a higher accuracy by about ~1.5%. Furthermore, in order to achieve these accuracies, the receiver will need a precise OSNR measurement, and, considering the ML approaches in OSNR monitoring seen in Section 1.3, this technique becomes feasible.

In contrast, Table 5.2 shows accuracy comparison when the OSNR is not known by the classifier. In this case, the use of the counting vectors preented better accuracy even when fiber transmission is considered, thus, it would be the best option when a precise OSNR measurement is not available, due to the remarkable accuracy of 78.5% only using the receiver symbols.

The use of any of the proposed methods would be helpful to control carrier frequencies in future gridless networks. Whilst the use of the counting vectors would require training sequences where the receiver needs to know the correct modulation index of each receiver symbol, the method using the validation index will only need a clustering analysis based on how compact, fuzzy and separated are the group of symbols in the

Table 5.1: Comparison of best accuracy results between Counting Vectors and Validation Indexes methods when the OSNR is known by the classifier.

Case	Number of Classes	Counting Vectors Accuracy	Validation Indexes Accuracy
16 GBd	2	84.0%	84.3%
	3	81.7%	84.6%
	4	77.7%	84.4%
32 GBd B2B	2	86.2%	87.3%
	3	84.4%	86.1%
	5	78.2%	82.3%
32 GBd Fiber	2	91.8%	90.5%
	3	90.1%	89.2%
	5	84.0%	86.4%

constellation diagram.

Table 5.2: Comparison of best accuracy results between Counting Vectors and Validation Indexes methods when the OSNR is not known by the classifier.

Case	Number of Classes	Counting Vectors Accuracy	Validation Indexes Accuracy
16 GBd	2	78.5%	74.4%
	3	72.7%	68.0%
	4	64.4%	63.2%
32 GBd B2B	2	74.4%	70.8%
	3	66.0%	59.2%
	5	49.0%	46.8%
32 GBd Fiber	2	72.4%	69.2%
	3	66.8%	60.8%
	5	48.0%	40.5%

5.2   ICI and Nonlinear Effects Mitigation

The ICI and nonlinear effects mitigation were carried out in the simulation and experimental data. We used as reference the FEC limit of $log_{10}(BER) = -2.4$ and evaluation of the methods was carried out as a function of OSNR.

5.2.1   Simulation Results

The proposed methods were applied in order to mitigate nonlinear Kerr effects in an simulated OAN link. Figures from 5.7a to 5.7d show BER vs OSNR results for four different launch power values: 0, 3, 6 and 9 dBm, respectively. As mentioned in Section 2.2, high launch power stimulates nonlinear effects which results in nonsymmetrical distortions. Thereby, when no nonlinear effects are present, for instance, 0 dBm of launch power (Fig. 5.7a), the distortions are symmetric and the methods do not present a relevant gain in terms of OSNR. The gains obtained by the methods are measured as it is shown in the Fig. 5.7c in terms of OSNR. Whereby, the use of the three proposed methods improves the BER when the nonlinear effects increase. For example, at a launch power of 3 dBm (Fig. 5.7b), a gain of ~0.1 dB is achieved by the proposed methods, on the other hand, when the launch power is 6 dBm (Fig. 5.7c), gains up to 0.5 dB are achieved respect to the traditional demodulation. Besides, in the case of 9 dBm of launch power (Fig. 5.7d), the FEC limit of $log_{10}(BER) = -2.4$ could not be obtained by traditional demodulation, however, it could be achieved using any of the three methods at an OSNR of ~22 dB.
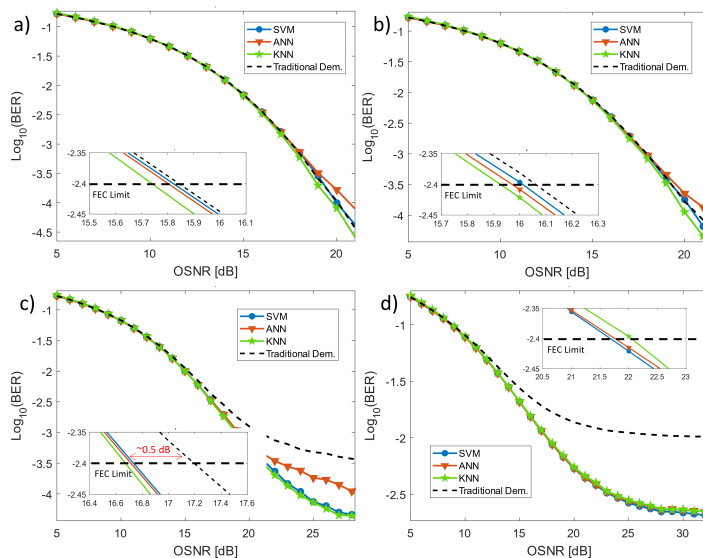
Figure 5.7: BER vs OSNR for launch power of: a) 0, b) 3, c) 6 and d) 9 dBm.

Furthermore, the nonlinear phase noise mitigation is also evaluated. Figures 5.8a to 5.8c shows BER vs OSNR for laser linewidth of 100, 500 and 1000 kHz, respectively. It can be observed that mitigation of nonlinear phase noise due to low-monochromatic sources can be achieved by any of the proposed methods. Besides, higher OSNR gains are achieved when the laser linewidth is greater. For instance, when the laser

linewidth is 100 kHz (Fig. 5.8a), gains of ~0.2 dB are achieved, on the other hand, for a linewidth value of 1000 kHz (Fig. 5.8c), gains up to ~0.5 dB were obtained. Moreover, it is noted that at high OSNR values, the method based on ANN algorithm, does not reach same BER values as using KNN or SVM. For example, Fig. 5.8d shows the constellation diagram at an OSNR of 23 dB in the case of 100 kHz linewidth. Marked in yellow are the transmitted symbols of an arbitrary index. Hence, Figs. 5.8e to 5.8g shows the demodulation of the received symbols using KNN, SVM and ANN, respectively. In Fig. 5.8g it can be seen that the ANN algorithm does not perform a properly nonsymmetric demodulation at high OSNR values. However, at low OSNR values, its performance is similar to KNN or SVM.

Thus, we demonstrated that it is possible to increase the network capacity only adding ML modules in the coherent receiver.
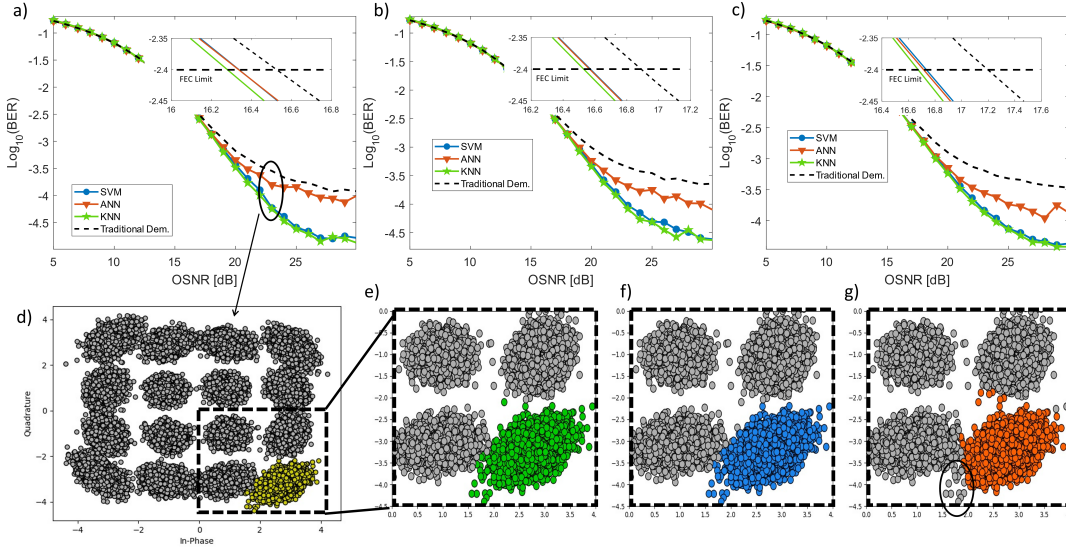
Figure 5.8: BER vs OSNR for laser linewidth of: a) 100, b) 500 and c) 1000 kHz. d) Constellation diagram at 23 dB of OSNR. Demodulation of a single symbol using e) KNN, f) SVM and g) ANN.

5.2.2   16 GBd Results

Figures from 5.9a to 5.9g depict the BER versus OSNR after the proposed methods are applied in demodulation. The BER impact in different spectral overlapping, including the single-channel scenario, is discussed. As a reference curve, the results with traditional demodulation (hard-decision) are also exposed. As mentioned in Section 3, these and previous BER curves correspond to B2B links, namely, the optical fiber effects are discarded in these tests. It is noticed that the use of the methods improves BER performance in comparison to the traditional demodulation. As expected, higher performance gains are achieved by the methods when the spectral spacing is reduced. For example, in the single-channel scenario (Fig. 5.9a), the use of ML methods allowed gains up to 0.3 dB in terms of OSNR at the FEC limit. Meanwhile, for a spectral overlapping of 0% (baud rate spacing in Hz), 6.8%, and 12.5%, the gains resulted from the adoption of the KNN algorithm are 0.4, 0.5, and 0.6 dB, respectively. Moreover, the use of KNN and ANN showed superior

performances to SVM, since SVM would need more training data per scenario in order to update the thresholds. In particular, for the scenario of 24% spectral overlapping (Fig. 5.9g), the FEC limit is not achieved due to the high spectral overlapping distortions regardless of high OSNR values. However, gains up to 2.5 dB were achieved by the three methods at a $log_{10}(BER)$ value of -1.7.

Transmission over 250 km of optical fiber (a typical WDM distance link) without OSNR variations was carried out in order to focus on the contribution of fiber impairments in the gridless scenario. Spectral spacing was varied from 12.6% to 0% of overlapping by including the single channel as a reference. Figure 5.9i shows the BER versus spectral overlapping and it can be seen a minimization of ICI effects by using the three methods. Moreover, it was possible to achieve similar BER values in a non-overlapped scenario without using any of the methods that in overlapped scenarios by using the proposed methods. For instance, with traditional demodulation at the case of 0% spectral overlapping a $log_{10}(BER)$ value of -2.8 is achieved. Instead, using any of the methods, this BER value can be obtained with spectral overlapping up to 12.6%. Besides, superior performances are achieved by the method based on KNN. This is because fiber nonlinearities can be mitigated by KNN due to the consideration of neighbors at the borders of symbols groups.
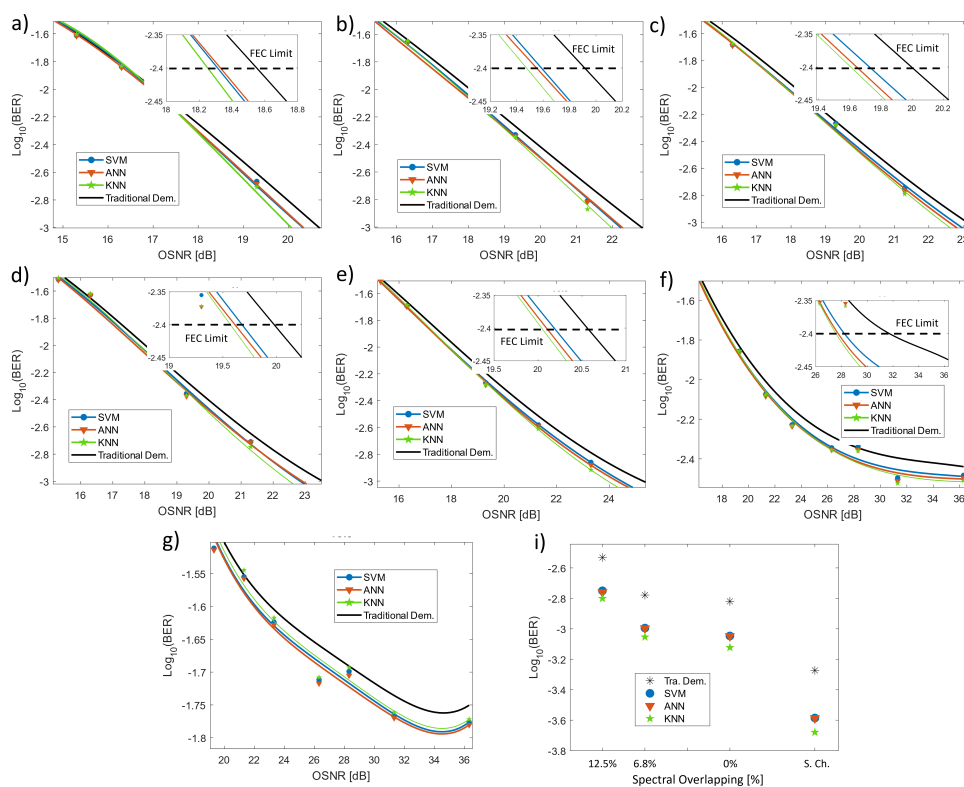
Figure 5.9: BER versus OSNR for different spectral overlapping (a) single channel, (b) close channels (18 GHz spacing), (c) 0%, (d) 6.8%, (e) 12.5%, (f) 18%, and (g) 24%. (i) BER versus spectral overlapping for the fiber transmission scenario.

5.2.3 32 GBd Results

(a) Back-to-back

Figures from 5.10a to 5.10j shows BER vs OSNR for different spectral overlapping scenarios including single channel. It can be evidenced that the use of the proposed methods allows mitigation of ICI effects. For instance, when the channels are spectrally overlapped by 1% (Fig. 5.10d) gains up to ~0.3 dB are achieved at the FEC limit of $log_{10}(BER) = -2.4$, on the other hand, at a spectral overlapping of 15% (Fig. 5.11i), gains up to ~0.8 dB in terms of OSNR. Moreover, the use of the proposed methods presented a minimum performance equal to the traditional demodulation. Besides, in the case of a high spectral overlapping of 18%, the FEC limit could not be reached for the OSNR applied. This is because the overlapped spectrum given the high baud rate is high and the distortions tend to be difficult to mitigate. Nevertheless, in this case, a gain up to ~1 dB is achieved at a $log_{10}(BER)$ value of -1.9.
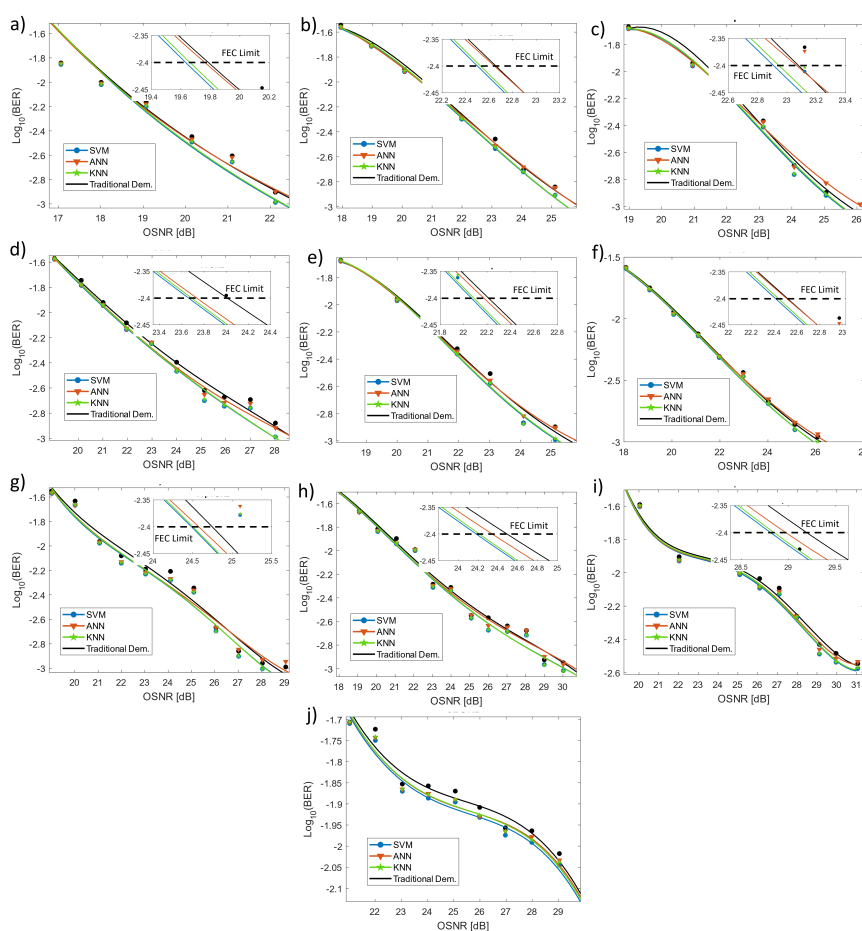
Figure 5.10: BER versus OSNR for different spectral overlapping for (a) single channel, (b) -4.5, (c) -1%, (d) 1%, (e) 4%, (f) 7%, (g) 9.5%, (h) 12.5%, (i) 15%, (j) 18%, applied to the 32 GBd setup at B2B.

(b) Fiber Transmission

Figures from 5.11a to 5.11i shows BER vs OSNR for different spectral overlapping scenarios including single channel. It is also seen that the use of the proposed methods allow mitigation of ICI effects. For example, when the channels are spectrally overlapped by 1% (Fig. 5.10d) gains up to ~0.2 dB are achieved at the FEC limit of $log_{10}(BER) = -2.4$, whilst at a spectral overlapping of 15% (Fig. 5.11i), gains up to ~0.5 dB in terms of OSNR. Moreover, the use of the methods allowed mitigation of fiber nonlinear effects which is seen when gains up to ~0.6 dB are achieved for example at Fig. 5.11b where there is not spectral overlapping.
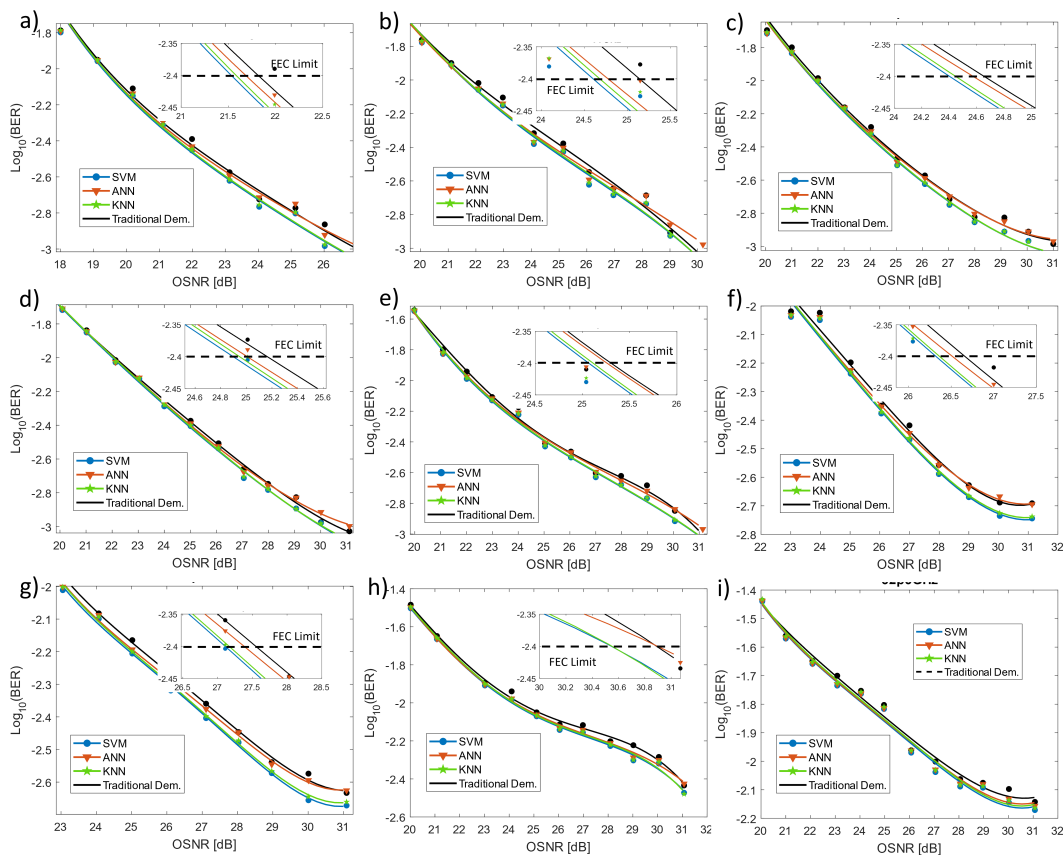
Figure 5.11: BER versus OSNR for different spectral overlapping for (a) single channel, (b) -4.5, (c) -1%, (d) 1%, (e) 4%, (f) 7%, (g) 9.5%, (h) 12.5%, (i) 15%, applied to the 32 GBd setup at 270 km of fiber transmission.

5.2.4 Proportional Execution Time

The computational time taken by the algorithms to train and demodulate is briefly evaluated in order to analyze future implementations. We calculated the overall processing time taken by the algorithms to train and

demodulation of 7k and 93k symbols, respectively. For comparison purposes, we normalized the highest
processing time (given by ANN), and the results presented by the KNN and SVM algorithms are propor-
tional. The results were obtained used the same computer. Table 5.2.4 shows the respective results by each
ML algorithm. It is seen a high elapsed time by ANN, due to the BP process in the training stage that
tends to be computationally complex. Besides, KNN requires 28% of the ANN time to classify the symbols
because it does not count with a training stage. However, it uses the training symbols to demodulate each
received symbol. Thus, the computational time is mostly expended in the sorting process of all the distances
calculated from the training symbols. Furthermore, SVM presented a low computational time carrying the
threshold calculations in the training stage, which is lower than sorting 7000 distances. Additionally, analy-
sis of computational complexity by means of computational operations, for instance, Big-O notation, is one
of the main future works.

Table 5.3: Proportional Time Taken by Each Supervised Algorithm to Train and Demodulate Symbols.

Algorithm	Proportional Time
ANN	100%
SVM	6.2%
KNN	28%

6   Conclusions and Future Work

6.1   Conclusions

In this research work, methods for estimation of spectral overlapping jointly with reduction of nonlinear and ICI effects were exposed. The methods use ML algorithms which have presented excellent results either in the performance estimation and in mitigation of signal impairments into the optical communications systems such as OANs and gridless networks.

Nonlinear effects will be relevant issues in high−capacity OANs because they increase the BER due to the nonsymmetrical distortions that are difficult to mitigate using traditional m−QAM demodulation, besides, nonlinear equalizers tends to be unfeasible due to its high computational cost. On the other hand, the ICI effects will be one of the most relevant issues in future gridless optical networks because in these networks, the channels will be spectrally close or even overlapped. Besides, ICI effects present similar distortions as the linear Gaussian noise resulting in BER increments and making difficult its detection.

The proposed methods were tested using simulated and experimental data. The simulated data performed an WDM OAN link at 16 GBd of 40 km fiber transmission with variation of laser linewidth and launch power. On the other hand, the experimental data was focused on 16-QAM Nyquist-WDM systems at 16 GBd and 32 GBd where the spectral spacing among three channels was varied reaching ICI effects.

Firstly, it was introduced two methods for estimation of spectral overlapping. The methods use received symbols frames affected by ICI effects and then, a feature extraction using information of unsupervised learning algorithms is carried out. Finally, these features are inputs to supervised learning algorithms which classifies if the received frame was affected by spectral overlapping. Two unsupervised learning algorithms which uses fuzzy clustering algorithms were used for feature extraction: FCM and GKM. On the other hand, two supervised algorithms for classification were evaluated also: SVM and KNN.

Experimental results showed that it is possible to determinate if any constellation diagram is affected by spectral overlapping using any of the proposed methods. Accuracies up to 91% were achieved in binary classification knowing a priori the OSNR value in the case of 32 GBd with fiber transmission. Moreover, when the OSNR is not known by the classifier, accuracies up to ~80% were obtained using the 16 GBd data. Besides, classification of different scenarios as high or low spectral overlapping jointly with detection of close channels or single channel was obtained also with accuracies up to ~90% or ~85% using the 32 GBd and 16 GBd data respectively. Besides, when the OSNR is not known by the classifier, ~78% of accuracy was obtained in a binary classification.

The use of FCM or GKM for feature extraction presented no relevant differences, this is due GKM could identify different clusters shapes and the ICI effects tend to form circular distortions. Additionally, in relation to the classification algorithms and using the counting vectors, the KNN algorithm presented better classification than SVM in the B2B scenarios. On the other hand, when fiber transmission was considered, the best classification was obtained by the SVM algorithm. In contrast, using the validation indexes as features, the SVM obtained better performance than KNN. Furthermore, the use of the counting vectors requires the use of training sequences whilst the use of the validation indexes needs only frames of received symbols. Besides, the feature extraction from the constellation diagrams will only require the initializing the clusters centroids in the ideal constellation format position using standard algorithms hyperparameters.

Furthermore, it was proposed three supervised ML algorithms to carry out nonsymmetrical demodulation in order to minimize: i) nonlinear distortions and ii) ICI effects and thus, reduce the BER. For the former, the methods were applied into the simulated OAN link, for the latter, ICI effects were mitigated using the experimental data.

Results showed that the ML algorithms could mitigate nonlinear effects as well as ICI effects allowing

gains in terms of the OSNR. Additionally, the BER reduction was obtained at low training lengths. Hence, a training strategy was also proposed in which a estimation tool will detect spectral overlapping level and the training sequences will be send carrying on updating of the algorithms. Simulation results preented the effectiveness of the methods by presenting gains up to ~0.5 dB in terms of OSNR at a FEC limit of $log_{10}(BER) = -2.4$. Besides, at high launch power of 9 dBm and OSNR of 25 dB, traditional demodulation method reached BER values of -2 while the proposed methods obtained ~-2.6. In those cases the three proposed methods showed same BER performance.

Besides, with the use of the KNN algorithm, it was possible to achieve the best BER improvement. For the B2B scenario, the ANN and KNN algorithms outperformed the SVM results in all spectral spacing cases. Furthermore, the KNN algorithms required less training symbols than ANN and SVM to achieve their best performance in terms of BER. Additionally, at a spectral overlapping of 18%, the SVM, KNN and ANN algorithms allowed gains up to ~4 dB in terms of OSNR. On the other hand, for the transmission scenario of 250 km of optical fiber, the use of KNN allowed the best BER improvement because the adaptability of KNN to classify nonsymmetrical symbols groups given by nonlinear fiber effects. Furthermore, it was aimed to reduce ICI effects at a minimal training length; thus, we presented the proportional time taken by the three algorithms to train and demodulate with a 100 k symbols frame, where the SVM algorithm showed the lowest computation time (equivalent to 6.2% of the ANN computational time).

Lastly, the methods aim to improve the capacity of future gridless optical networks and to enhance OANs. Additionally, the proposed methods for estimation and mitigation can be implemented for any constellation diagram-based modulation formats as *m*-QPSK and *m*-QAM. The use of these ML-based techniques will be useful to improve the performance of future OANs and gridless networks.

6.2   Future Works

In order to achieve higher estimation accuracy, different contributions using variants of characterization of diagram constellations could be explored. In [95], we explored the use of digital image processing where several features were extracted using heat maps and gray scaling including a texture analysis. We explored alternatives to the trend of neural networks with convolutional layers where feature extraction in carried out automatically. However, only preliminar results were obtained and the path for deep explorations was opened. Additionally, another ML algorithms related to deep learning will be explored using image processing. Besides, the inclusion of regression models for spectral spacing estimation will be investigated due to the fact that in a gridless network, the spectrum granularity will be close to be continuous.

On the other hand, future work may put forward the inclusion of higher data-rate scenarios with advanced modulation formats as well as the exploration of low-computational-complexity solutions in order to obtain advanced gains in terms of OSNR. Besides, hardware implementations of any ML approach into the optical communications would still need more research as we partially explored in [96]. Additionally, analysis of computational complexity by means of computational operations, for instance, Big-O notation, is one of the main future works.

7 References

References

[1] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5G Backhaul Challenges and Emerging Research Directions: A Survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.

[2] ITU-R, "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," *Itu-R M.2083-0*, vol. 0, 2015.

[3] S. Pachnicke, M. H. Eiselt, K. Grobe, and J. P. Elbers, "The frontiers of optical access networks," *Conference Proceedings - 2015 International Conference on Optical Network Design and Modeling, ONDM 2015*, pp. 12–15, 2015.

[4] G. Bosco, A. Carena, V. Curri, P. Poggiolini, and F. Forghieri, "Performance limits of nyquist-WDM and CO-OFDM in high-speed PM-QPSK systems," *IEEE Photonics Technology Letters*, vol. 22, no. 15, pp. 1129–1131, 2010.

[5] N. Suzuki and H. Miura, "Digital Coherent DSP based PON Technology for Ultimate Capacity Optical Access Systems," *2020 IEEE Photonics Society Summer Topical Meeting Series, SUM 2020 - Proceedings*, pp. 52–53, 2020.

[6] D. Lavery, R. Maher, D. S. Millar, B. C. Thomsen, P. Bayvel, and S. J. Savory, "Digital coherent receivers for long-reach optical access networks," *Journal of Lightwave Technology*, vol. 31, no. 4, pp. 609–620, 2013.

[7] D. Amar, E. Le Rouzic, N. Brochier, J.-L. Auge, C. Lepers, N. Perrot, and S. Fazel, "How problematic is Spectrum Fragmentation in operator's Gridless network?" *International Conference on Optical Network Design and Modeling (ONDM)*, no. July 2015, pp. 67–72, 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6855776/

[8] International Telecommunication Union - ITU-T, "G.694.1 (06/2002), Spectral grids for WDM applications: DWDM frequency grid," *Series G.694.1*, p. 14, 2002.

[9] P. Wright, A. Lord, and L. Velasco, "The network capacity benefits of flexgrid," *International Conference of Optical Networking Design and Modeling*, no. c, pp. 7–12, 2013.

[10] J. J. Granada Torres, J. P. López Martínez, E. Avendaño Fernández, A. M. Cárdenas Soto, and N. Guerrero González, "Impact of interchannel interference in gridless nyquist-wdm systems with and without nonlinear impairment compensation," *Ciencia e Ingeniería Neogranadina*, vol. 29, no. 2, pp. 9–23, 2019.

[11] L. Beygi, E. Agrell, P. Johannisson, M. Karlsson, and H. Wymeersch, "The Limits of Digital Backpropagation in Nonlinear Coherent Fiber-Optical Links," *2012 38th European Conference and Exhibition on Optical Communications*, no. 1, p. P4.14, 2013.

[12] L. Li, Z. Tao, L. Dou, W. Yan, S. Oda, T. Tanimura, T. Hoshida, and J. C. Rasmussen, "Implementation Efficient Nonlinear Equalizer Based on Correlated Digital Backpropagation," *2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference*, no. 2, p. OWW3, 2013.

[13] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: Modeling and mitigation," *Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1044–1053, 2015.

[14] P. Serena and A. Bononi, "An alternative approach to the gaussian noise model and its system implications," *Journal of Lightwave Technology*, vol. 31, no. 22, pp. 3489–3499, 2013.

[15] J. J. Granada Torres, S. Varughese, V. A. Thomas, A. Chiuchiarelli, S. E. Ralph, A. M. Cárdenas Soto, and N. Guerrero González, "Mitigation of time-varying distortions in Nyquist-WDM systems using machine learning," *Optical Fiber Technology*, vol. 38, no. March, pp. 130–135, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.yofte.2017.09.008

[16] V. Vgenopoulou, N. P. Diamantopoulos, I. Roudas, and S. Sygletos, "MIMO Nonlinear Equalizer based on Inverse Volterra Series Transfer Function for Coherent SDM Systems," *2019 Optical Fiber Communications Conference and Exhibition, OFC 2019 - Proceedings*, vol. 1, no. Cd, pp. 1–3, 2019.

[17] L. Wang, F. Jiang, M. Chen, H. Dou, G. Gui, and H. Sari, "Interference Mitigation Based on Optimal Modes Selection Strategy and CMA-MIMO Equalization for OAM-MIMO Communications," *IEEE Access*, vol. 6, pp. 69 850–69 859, 2018.

[18] W. Wang, J. Zhao, H. Yu, Z. Yang, Y. Zhang, Z. Zhang, C. Guo, and G. Li, "Demonstration of 6 × 10-Gb/s MIMO-Free Polarization- and Mode-Multiplexed Transmission," *IEEE Photonics Technology Letters*, vol. 30, no. 15, pp. 1372–1375, 2018.

[19] M. Sato, R. Maher, D. Lavery, K. Shi, B. C. Thomsen, and P. Bayvel, "Frequency diversity MIMO detection for DP-QAM transmission," *Journal of Lightwave Technology*, vol. 33, no. 7, pp. 1388–1394, 2015.

[20] T. Koike-Akino, K. Kojima, D. S. Millar, K. Parsons, S. Kametani, T. Sugihara, T. Yoshida, K. Ishida, Y. Miyata, W. Matsumoto *et al.*, "Han–kobayashi and dirty-paper coding for superchannel optical communications," *Journal of Lightwave Technology*, vol. 33, no. 7, pp. 1292–1299, 2015.

[21] F. Alishahi, A. Mohajerin-Ariaei, A. Fallahpour, Y. Cao, A. Almaiman, P. Liao, C. Bao, B. Shamee, K. Zou, H. Zhou, A. N. Willner, J. D. Touch, M. Tur, C. Langrock, M. M. Fejer, and A. E. Willner, "Optical mitigation of interchannel crosstalk for multiple spectrally overlapped 20-GBd QPSK/16-QAM WDM channels using nonlinear wave mixing," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 548–554, 2019.

[22] A. Mohajerin-Ariaei, M. Ziyadi, Y. Cao, A. Almaiman, F. Alishahi, A. Fallahpour, C. Bao, P. Liao, B. Shamee, J. Touch, M. Tur, C. Langrock, M. M. Fejer, and A. E. Willner, "Demonstration of Tunable Mitigation of Interchannel Interference of Spectrally Overlapped 16-QAM/QPSK Data Channels using Wave Mixing of Delayed Copies," 2017, p. Th3J.5.

[23] E. Giacoumidis, A. Matin, J. Wei, N. J. Doran, L. P. Barry, and X. Wang, "Blind Nonlinearity Equalization by Machine-Learning-Based Clustering for Single- and Multichannel Coherent Optical OFDM," *Journal of Lightwave Technology*, vol. 36, no. 3, pp. 721–727, 2018.

[24] D. Wang, M. Zhang, M. Fu, Z. Cai, Z. Li, H. Han, Y. Cui, and B. Luo, "Nonlinearity Mitigation Using a Machine Learning Detector Based on k -Nearest Neighbors," *IEEE Photonics Technology Letters*, vol. 28, no. 19, pp. 2102–2105, 2016.

[25] C. Häger and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," *Optics InfoBase Conference Papers*, vol. Part F84-O, no. 2, pp. 38–40, 2018.

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

[27] 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, 2018.

[28] D. S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi, "Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation," *Journal of Lightwave Technology*, vol. 24, no. 1, pp. 12–20, 2006.

[29] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," vol. 313, no. July, pp. 504–507, 2006.

[30] K. Lai, R. Y. Surampalli, R. D. Tyagi, S. Banerji, and S. Yan, "Applications of Artificial Neural Networks in Optical Performance Monitoring," *Remediation Technologies for Soils and Groundwater*, vol. 27, no. 16, pp. 395–437, 2009.

[31] N. Gonzalez, D. Zibar, A. Caballero, and I. Monroy, "Experimental 2.5-Gb/s QPSK WDM Phase-Modulated Radio-Over-Fiber Link With Digital Demodulation by a k-Means Algorithm," *IEEE Photonics Technology Letters*, vol. 22, no. 5, pp. 335–337, 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5378581

[32] J. A. Jargon, X. Wu, H. Y. Choi, Y. C. Chung, and A. E. Willner, "Optical performance monitoring of QPSK data channels by use of neural networks trained with parameters derived from asynchronous constellation diagrams," *Optics Express*, vol. 18, no. 5, pp. 4931–4938, 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-5-4931

[33] F. N. Khan, Y. Yu, M. C. Tan, W. H. Al-Arashi, C. Yu, A. P. T. Lau, and C. Lu, "Experimental demonstration of joint osnr monitoring and modulation format identification using asynchronous single channel sampling," *Optics express*, vol. 23, no. 23, pp. 30337–30346, 2015.

[34] T. Tanimura, T. Hoshida, J. C. Rasmussen, M. Suzuki, and H. Morikawa, "OSNR monitoring by deep neural networks trained with asynchronously sampled data," *2016 21st OptoElectronics and Communications Conference, OECC 2016 - Held Jointly with 2016 International Conference on Photonics in Switching, PS 2016*, pp. 39–41, 2016.

[35] F. N. Khan, K. Zhong, X. Zhou, W. H. Al-Arashi, C. Yu, C. Lu, and A. P. T. Lau, "Joint osnr monitoring and modulation format identification in digital coherent receivers using deep neural networks," *Optics express*, vol. 25, no. 15, pp. 17767–17776, 2017.

[36] A. S. Kashi, Q. Zhuge, J. C. Cartledge, A. Borowiec, D. Charlton, C. Laperle, and M. O'Sullivan, "Fiber nonlinear noise-to-signal ratio monitoring using artificial neural networks," pp. 1–3, 2017.

[37] D. Wang, M. Zhang, J. Li, Z. Li, J. Li, C. Song, and X. Chen, "Intelligent constellation diagram analyzer using convolutional neural network-based deep learning," *Optics Express*, vol. 25, no. 15, pp. 17150–17166, 2017. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-25-15-17150

[38] F. J. Caballero, D. J. Ives, C. Laperle, D. Charlton, Q. Zhuge, M. O'Sullivan, and S. J. Savory, "Machine learning based linear and nonlinear noise estimation," *Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D42–D51, 2018.

[39] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, "Data-analytics-based optical performance monitoring technique for optical transport networks," *Optics InfoBase Conference Papers*, vol. Part F84-O, pp. 1–3, 2018.

[40] Z. Wang, A. Yang, P. Guo, and P. He, "Osnr and nonlinear noise power estimation for optical fiber communication systems using lstm based deep learning technique," *Optics express*, vol. 26, no. 16, pp. 21 346–21 357, 2018.

[41] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, "Convolutional neural network-based optical performance monitoring for optical transport networks," *Journal of Optical Communications and Networking*, vol. 11, no. 1, pp. A52–A59, 2019.

[42] H. Yang, Q. Yao, A. Yu, Y. Lee, and J. Zhang, "Resource Assignment based on Dynamic Fuzzy Clustering in Elastic Optical Networks with Multi-core Fibers," *IEEE Transactions on Communications*, vol. 6778, no. c, pp. 1–1, 2019.

[43] J. J. Torres and N. G. González, "Overlapping estimation based on DBSCAN algorithm in Nyquist-WDM systems," in *Optics InfoBase Conference Papers*, vol. Part F137-, 2019, pp. 2–3.

[44] K. Zhang, Y. Fan, T. Ye, Z. Tao, S. Oda, T. Tanimura, Y. Akiyama, and T. Hoshida, "Fiber Nonlinear Noise-to-Signal Ratio Estimation by Machine Learning," *2019 Optical Fiber Communications Conference and Exhibition, OFC 2019 - Proceedings*, vol. 1, no. 1, pp. 1–3, 2019.

[45] C. Wang, S. Fu, H. Wu, M. Luo, X. Li, M. Tang, and D. Liu, "Joint osnr and cd monitoring in digital coherent receiver using long short-term memory neural network," *Optics express*, vol. 27, no. 5, pp. 6936–6945, 2019.

[46] L. Xia, J. Zhang, S. Hu, M. Zhu, Y. Song, and K. Qiu, "Transfer learning assisted deep neural network for osnr estimation," *Optics express*, vol. 27, no. 14, pp. 19 398–19 406, 2019.

[47] 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. 19 726–19 737, 2019.

[48] A. Salehiomran, G. Gao, and Z. Jiang, "Linear and nonlinear noise monitoring in coherent systems using fast BEr measurement and neural networks," *IET Conference Publications*, vol. 2019, no. CP765, pp. 32–34, 2019.

[49] Y. Zhong, S. Cui, H. Lu, C. Ke, H. Wang, P. Jiang, and D. Liu, "A Robust Reference Optical Spectrum Based in-Band OSNR Monitoring Method Suitable for Flexible Optical Networks," *IEEE Photonics Journal*, vol. 12, no. 3, pp. 1–10, 2020.

[50] H. J. Cho, S. Varughese, D. Lippiatt, R. Desalvo, S. Tibuleac, and S. E. Ralph, "Optical performance monitoring using digital coherent receivers and convolutional neural networks," *Optics Express*, vol. 28, no. 21, pp. 32 087–32 104, 2020.

[51] R. Pousa, P. Georgieva, J. Pina, P. Cruz, and P. André, "Machine learning approach for online monitoring of quality of transmission performance indicators in optical fiber networks," in *2021 European Conference on Optical Communication (ECOC)*. IEEE, 2021, pp. 1–4.

[52] A. S. Kashi, J. C. Cartledge, and W.-Y. Chan, "Neural network training framework for nonlinear signal-to-noise ratio estimation in heterogeneous optical networks," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*.   IEEE, 2021, pp. 1–3.

[53] A. E. Pérez, J. J. Torres, and N. G. González, "KNN-based demodulation in gridless Nyquist-WDM systems affected by interchannel interference," in *Optics InfoBase Conference Papers*, vol. Part F137-, 2019, pp. 1–2.

[54] J. J. G. Torres, A. M. C. Soto, and N. G. González, "Mitigation of linear inter-channel interference for sub-nyquist spacing in optical multicarrier systems," pp. 1–5, 2015.

[55] M. A. Jarajreh, E. Giacoumidis, I. Aldaya, S. T. Le, A. Tsokanos, Z. Ghassemlooy, and N. J. Doran, "Artificial neural network nonlinear equalizer for coherent optical OFDM," *IEEE Photonics Technology Letters*, vol. 27, no. 4, pp. 387–390, 2015.

[56] D. Wang, M. Zhang, Z. Cai, Y. Cui, Z. Li, H. Han, M. Fu, and B. Luo, "Combatting nonlinear phase noise in coherent optical systems with an optimized decision processor based on machine learning," *Optics Communications*, vol. 369, pp. 199–208, 2016.

[57] E. Giacoumidis, S. Mhatli, M. F. Stephens, A. Tsokanos, J. Wei, M. E. McCarthy, N. J. Doran, and A. D. Ellis, "Reduction of Nonlinear Intersubcarrier Intermixing in Coherent Optical OFDM by a Fast Newton-Based Support Vector Machine Nonlinear Equalizer," *Journal of Lightwave Technology*, vol. 35, no. 12, pp. 2391–2397, 2017.

[58] E. Giacoumidis, A. Matin, J. Wei, N. J. Doran, and X. Wang, "Unsupervised hierarchical clustering for blind nonlinear equalization in WDM coherent optical OFDM," *Optics InfoBase Conference Papers*, vol. Part F83-A, pp. 4–6, 2017.

[59] E. Giacoumidis, S. Mhatli, J. Wei, S. T. Le, I. Aldaya, M. F. C. Stephens, M. E. MacCarthy, A. D. Ellis, N. J. Doran, and B. J. Eggleton, "Intra and inter-channel nonlinearity compensation in WDM coherent optical OFDM using artificial neural network based nonlinear equalization," in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, 2017, pp. 1–3.

[60] C. Y. Chuang, L. C. Liu, C. C. Wei, J. J. Liu, L. Henrickson, W. J. Huang, C. L. Wang, Y. K. Chen, and J. Chen, "Convolutional neural network based nonlinear classifier for 112-Gbps high speed optical link," *Optics InfoBase Conference Papers*, vol. Part F84-O, pp. 1–3, 2018.

[61] L. Liu, M. Bi, S. Xiao, J. Fang, T. Huang, and W. Hu, "OLS-based RBF neural network for nonlinear and linear impairments compensation in the CO-OFDM system," *IEEE Photonics Journal*, vol. 10, no. 2, pp. 1–8, 2018.

[62] X. Dai, X. Li, M. Luo, and S. Yu, "Numerical simulation and experimental demonstration of accurate machine learning aided iq time-skew and power-imbalance identification for coherent transmitters," *Optics express*, vol. 27, no. 26, pp. 38 367–38 381, 2019.

[63] A. E. Pérez and J. J. G. Torres, "Knn, k-means and fuzzy c-means for 16-qam demodulation in coherent optical systems," in *2019 IEEE Colombian Conference on Communications and Computing (COLCOM)*.   IEEE, 2019, pp. 1–4.

[64] E. A. Fernández, J. J. G. Torres, A. M. C. Soto, and N. G. Gonzalez, "Radio-over-Fiber signal demodulation in the presence of non-Gaussian distortions based on subregion constellation processing," *Optical Fiber Technology*, vol. 53, no. October, p. 102062, 2019. [Online]. Available: https://doi.org/10.1016/j.yofte.2019.102062

[65] E. Giacoumidis, J. Wei, I. Aldaya, C. Sanchez, H. Mrabet, and L. P. Barry, "Fiber-induced nonlinearity compensation in coherent optical systems by affinity propagation soft-clustering," *arXiv preprint arXiv:1812.05600*, 2018.

[66] L. Yi, T. Liao, L. Huang, L. Xue, P. Li, and W. Hu, "Machine Learning for 100 Gb/s/λ Passive Optical Network," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1621–1630, 2019.

[67] A. Amari, X. Lin, O. A. Dobre, R. Venkatesan, and A. Alvarado, "A Machine Learning-Based Detection Technique for Optical Fiber Nonlinearity Mitigation," *IEEE Photonics Technology Letters*, vol. 31, no. 8, pp. 627–630, 2019.

[68] O. Sidelnikov, A. Redyuk, S. Sygletos, M. Fedoruk, and S. Turitsyn, "Advanced convolutional neural networks for nonlinearity mitigation in long-haul wdm transmission systems," *Journal of Lightwave Technology*, vol. 39, no. 8, pp. 2397–2406, 2021.

[69] P. J. Freire, V. Neskornuik, A. Napoli, B. Spinnler, N. Costa, G. Khanna, E. Riccardi, J. E. Prilepsky, and S. K. Turitsyn, "Complex-Valued Neural Network Design for Mitigation of Signal Distortions in Optical Links," *Journal of Lightwave Technology*, vol. 39, no. 6, pp. 1696–1705, 2021.

[70] A. Ellis, M. McCarthy, M. Al Khateeb, M. Sorokina, and N. Doran, "Performance limits in optical communications due to fiber nonlinearity," *Advances in Optics and Photonics*, vol. 9, no. 3, pp. 429–503, 2017.

[71] E. M. Ip and J. M. Kahn, "Fiber Impairment Compensation Using Coherent Detection and Digital Signal Processing," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 502–519, 2010.

[72] J. Rahn, S. Kumar, M. Mitchell, H. Sun, K. T. Wu, G. Goldfarb, M. Kato, D. Krause, R. Nagarajan, F. Kish, and D. Welch, "Super-channels: DWDM transmission beyond 100 Gb/s," *2012 IEEE Photonics Conference, IPC 2012*, vol. 1, pp. 854–855, 2012.

[73] G. Bosco, A. Carena, V. Curri, P. Poggiolini, E. Torrengo, and F. Forghieri, "Investigation on the robustness of a Nyquist-WDM Terabit superchannel to transmitter and receiver non-idealities," *European Conference on Optical Communication, ECOC*, vol. 1-2, no. 1, pp. 19–21, 2010.

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

[75] N. Praveen, S. A. Vasista, and T. S. Indumathi, "Spectrum analysis for performance evaluation of wavelength selective switches and gridless ROADMs for next generation networks," *2012 International Conference on Optical Engineering, ICOE 2012*, 2012.

[76] G. P. Agrawal, *Fiber-Optic Communications Systems, Third Edition.*, 2002, vol. 6.

[77] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity Limits of Optical Fiber Networks PART 1: DIGITIZATION, CAPACITY, AND CONSTELLATIONS," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, 2010.

[78] M. Wu and W. I. Way, "Fiber nonlinearity limitations in ultra-dense WDM systems," *Journal of Lightwave Technology*, vol. 22, no. 6, pp. 1483–1497, 2004.

[79] J. J. G. Torres, A. M. C. Soto, and N. G. González, "Evaluation and compensation of interchannel interference effects in a 16-QAM Nyquist-WDM system with LMS equalization," *2014 IEEE Latin-America Conference on Communications, IEEE LATINCOM 2014*, pp. 2–7, 2014.

[80] S. J. Savory, "Digital filters for coherent optical receivers," *Optics express*, vol. 16, no. 2, pp. 804–817, 2008.

[81] S. Oda, T. Tanimura, T. Hoshida, C. Ohshima, H. Nakashima, Z. Tao, and J. C. Rasmussen, "112 gb/s dp-qpsk transmission using a novel nonlinear compensator in digital coherent receiver," in *Optical Fiber Communication Conference*.   Optical Society of America, 2009, p. OThR6.

[82] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *Journal of Lightwave Technology*, vol. 26, no. 20, pp. 3416–3425, 2008.

[83] Q. Zhuge and W. Hu, "Application of machine learning in elastic optical networks," in *2018 European Conference on Optical Communication (ECOC)*.   IEEE, 2018, pp. 1–3.

[84] F. Shen, J. Zhou, Z. Huang, and L. Li, "Going deeper into osnr estimation with cnn," in *Photonics*, vol. 8, no. 9.   Multidisciplinary Digital Publishing Institute, 2021, p. 402.

[85] J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey," *Optical Switching and Networking*, vol. 28, pp. 43–57, 2018.

[86] L. Velasco and D. Rafique, "Fault Management Based on Machine Learning," *2019 Optical Fiber Communications Conference and Exhibition, OFC 2019 - Proceedings*, pp. 1–3, 2019.

[87] ITU-T, "Recommendation ITU-T G.652: Characteristics of a Single-Mode Optical Fibre and Cable," *Itu-T G652*, pp. 1–28, 2016. [Online]. Available: https://www.itu.int/rec/T-REC-G.652-201611-I/en

[88] A. Bensaid, L. Hall, J. Bezdek, L. Clarke, M. Silbiger, J. Arrington, and R. Murtagh, "Validity-guided (re)clustering with applications to image segmentation," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, pp. 112–123, 1996.

[89] G. James, D. Witten, T. Hastie, and R. Tibshirani.   Springer, 2013, vol. 112.

[90] R. James, G., Witten, D., Hastie, T., Tibshirani, *An Introduction to Statistical Learning - with Applications in R | Gareth James | Springer*, 2013. [Online]. Available: https://www.springer.com/gp/book/9781461471370%0Ahttp://www.springer.com/us/book/9781461471370

[91] C. C. Aggarwal, *Neural Networks and Deep Learning*.   Cham: Springer International Publishing, 2018. [Online]. Available: http://link.springer.com/10.1007/978-3-319-94463-0

[92] "Activation Functions: Sigmoid, Tanh, ReLU, Leaky ReLU, Softmax | by Mukesh Chaudhary | Medium." [Online]. Available: https://medium.com/@cmukesh8688/activation-functions-sigmoid-tanh-relu-leaky-relu-softmax-50d3778dcea5

[93] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, 2015.

[94] L. M. Zhang and F. R. Kschischang, "Staircase codes with 6% to 33% overhead," *Journal of Lightwave Technology*, vol. 32, no. 10, pp. 1999–2002, 2014.

[95] A. E. Pérez, O. D. Vargas, S. Ralph, and J. J. G. Torres, "Spectral spacing estimation in gridless nyquist-wdm systems using local binary patterns," in *2021 IEEE Photonics Conference (IPC)*. IEEE, 2021, pp. 1–2.

[96] D. Marquez-Viloria, L. Castaño Londoño, A. Escobar-Perez, J. J. Granada-Torres, and N. Guerrero-Gonzalez, "Scalable multi-core fpga design for maximum concurrency: The case of knn for ici mitigation," in *Signal Processing in Photonic Communications*. Optical Society of America, 2020, pp. SpTu1I–5.