Machine Learning based Reliability Assurance in Optical Networks

Dissertation
zur Erlangung des akademischen Grades
Doktor der Ingenieurswissenschaften (Dr.-Ing.)
der Technischen Fakultät
der Christian-Albrechts-Universität zu Kiel

Lars Eike Kruse

Kiel 2024

Betreuer:	Prof. Dr.-Ing. Stephan Pachnicke
Zweitgutachterin:	Univ.-Prof. Dr. Carmen Mas Machuca
Datum der Disputation:	10. Juli 2024

Vorwort

Die vorliegende Arbeit ist im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Nachrichtenübertragungstechnik der Christian-Albrechts-Universität zu Kiel entstanden. Die gesamte Forschung wurde mit Hinblick auf das Forschungsprojekt AI-NET im Teilprojekt AI-NET PROTECT des EUREKA CELTIC-NEXT Clusters durchgeführt. Während der Zeit des Projekts hatte ich die Möglichkeit viele interessante und lehrreiche Projekttreffen sowie internationale Konferenzen zu besuchen und somit meine Erkenntnisse mit der Forschungsgemeinschaft zu teilen.

Mein Dank gilt als erstes meinem Betreuer Herrn Prof. Dr.-Ing. Stephan Pachnicke für die Möglichkeit in diesem Projekt mitzuwirken, für die Vielzahl an produktiven Diskussionen und das Vertrauen während der gesamten Zeit. Weiterhin möchte ich mich bei meinen Kollegen am Lehrstuhl Tom, Sebastian, Annika, Jonas, Rebekka, Olaf, Alexandr, Silas, Tim, Ali, Mohammad, Sandra und Petra für die angenehme Atmosphäre, die konstruktiven Diskussionen, die gute Zusammenarbeit und das Korrekturlesen bedanken. Außerdem danke ich den Projektpartnern von AI-NET PROTECT für die Anregungen und fachlichen Diskussionen.

Ich danke Frau Prof. Dr. Carmen Mas Machuca für die Übernahme des Zweitgutachtens und die damit einhergehende investierte Zeit. Des Weiteren möchte ich mich bei Herrn Prof. Dr. Michael Höft und Herrn Prof. Dr. Hermann Kohlstedt für die Bildung einer freundlichen und fairen Prüfungskommission bedanken.

Für die uneingeschränkte Unterstützung während meines gesamten Studiums und meiner Promotion danke ich vor allem meiner Familie. Ohne Euch wäre dies nicht möglich gewesen.

Zuletzt möchte ich mich bei meiner Frau Clara für die ständige Unterstützung und Beistand bedanken. Danke, dass ich mich immer auf Dich verlassen kann und du mir den Rücken freihältst!

Abstract

Effectively handling failures and ensuring the quality of transmission are essential elements in maintaining optimal performance for high-speed optical networks in the digital age. With the constant increase in data demands, any disruption of optical connections can result in significant financial consequences and data loss. In response to the complex and continuously evolving nature of optical networks, there is an urgent need for automated and dynamic methods to manage failures. While traditional approaches rely on conservative designs, machine learning algorithms offer proactive solutions for early detection of failures and aging dependent degradations in the quality of transmission. Accurately estimating the quality of transmission of a certain connection is crucial for maximizing capacity and possible self-management of future networks. Monitoring data for the training of machine learning algorithms is difficult to obtain from wavelength division multiplexed signals without demultiplexing the channels. However, optical spectrum analyzers can be used to obtain signal information in a transparent way, i.e., without demultiplexing.

This work presents novel approaches for quality of transmission estimation and soft-failure management based on the optical spectrum obtained through optical spectrum analyzers. Multiple machine learning algorithms were evaluated for quality of transmission estimation under uncertain parameters using optical spectral data. The algorithms were trained on simulations and tested on experimental data. Spectral data utilization increased the quality of transmission estimation accuracy. Recursive interpretation of optical spectra for each node in conjunction with recurrent neural networks further improved accuracy. Furthermore, this work investigates optical spectra driven soft-failure management using experimentally emulated component failures. Machine learning algorithms analyzing optical spectra effectively detected, identified, and localized emulated soft-failures, showcasing their potential when running on spectral data. For fault detection, a variational autoencoder was used, yielding high accuracy. Integrating a generative adversarial network into the framework, enabled by the variational autoencoder's generative capabilities, achieved exceptional performance in identifying soft-failures with very low amounts of available training data. In combination with advanced training methods for generative adversarial networks, unknown spectra were reliably identified enabling unknown failure detection and classification.

Zusammenfassung

Effektiver Umgang mit Ausfällen und die Sicherstellung der Übertragungsqualität sind wesentliche Elemente zur Aufrechterhaltung optimaler Leistung für Hochgeschwindigkeits-Glasfasernetzwerke im digitalen Zeitalter. Angesichts der stetig steigenden Datennachfrage kann jede Störung der optischen Verbindungen erhebliche finanzielle Folgen und Datenverluste zur Folge haben. Als Reaktion auf die komplexe und sich ständig weiterentwickelnde Natur optischer Netzwerke besteht ein dringender Bedarf an automatisierten und dynamischen Methoden zur Bewältigung von Ausfällen. Während traditionelle Ansätze auf konservativen Designs beruhen, bieten Machine-Learning-Algorithmen proaktive Lösungen für frühzeitige Erkennung von Ausfällen und altersabhängigen Verschlechterungen der Übertragungsqualität. Eine genaue Schätzung der Übertragungsqualität einer bestimmten Verbindung ist entscheidend, um die Kapazität zu maximieren und eine mögliche Selbstverwaltung zukünftiger Netzwerke zu ermöglichen. Die Überwachungsdaten für das Training des Machine-Learning-Algorithmus sind schwer zu erhalten, wenn sie aus Wellenlängenmultiplex-Signalen ohne Demultiplexing der Kanäle stammen. Allerdings können optische Spektrumanalysatoren verwendet werden, um Signalinformationen auf transparente Weise zu erhalten, d.h. ohne Demultiplexing.

Diese Arbeit präsentiert neuartige Verfahren für die Schätzung der Übertragungsqualität und für das Soft-Failure-Management, basierend auf dem optischen Spektrum, das durch optische Spektrumanalysatoren gewonnen wird. Mehrere Machine-Learning-Algorithmen wurden für die Schätzung der Übertragungsqualität unter unsicheren Parametern unter Verwendung optischer Spektraldaten evaluiert. Die Algorithmen wurden auf Simulationen trainiert und auf experimentellen Daten getestet. Die Nutzung von spektralen Daten erhöhte die Genauigkeit der Übertragungsqualitätsschätzung. Darüber hinaus untersucht diese Arbeit ein auf optischen Spektren basierendes Soft-Failure-Management unter Verwendung von experimentell emulierten Komponentenausfällen. Machine-Learning-Algorithmen, die optische Spektren analysieren, haben emulierte Soft-Failures effektiv erkannt, identifiziert und lokalisiert, was ihr Potenzial bei der Verwendung von Spektraldaten zeigt. Für die Fehlererkennung wurde ein Variational Autoencoder verwendet, der eine hohe Genauigkeit lieferte. Die Integration eines erzeugenden gegnerischen Netzwerkes (engl.: generative adversarial network) in das Framework, ermöglicht durch die generativen Fähigkeiten des Variational Autoencoders, erzielte eine außergewöhnliche Leistung bei der Identifizierung von Soft-Failures bei sehr geringen verfügbaren Trainingsdatenmengen. In Kombination mit fortschrittlichen Trainingsmethoden für erzeugende gegnerische Netzwerke wurden unbekannte Spektren zuverlässig identifiziert, was eine Erkennung und Klassifizierung unbekannter Ausfälle ermöglichte.

Contents

Chapter 1

Introduction

Managing failures and ensuring the quality of transmission (QoT) are critical aspects of maintaining high-speed optical networks in the digital era. With the data demand growing exponentially [1] (Fig. 1.1), challenges arise for future optical networks. In this regard, accurately estimating the QoT is vital for maximizing capacity by reducing system margins and enabling self-management of networks in the future. Furthermore, disruptions to optical links can result in significant data loss and service level agreements no longer being met. To address the increasing complexity and dynamic nature of today's and future optical networks, automated and dynamic techniques for QoT estimation and failure management are required [2].

With the evolution of optical networks to meet increasing bandwidth demands, flexible add-drop multiplexers and reconfigurable networks have become the norm. However, current networks include large operating margins to ensure service level agreements and promised capacity, resulting in wasted capacity [3]. This encourages the use of performant QoT estimation for optimizing the performance and capacity of optical networks.

In multi-vendor optical networks, where equipment parameters may be confidential or unknown, QoT estimation becomes even more challenging. These networks can be considered as scenarios where exact component parameters are agnostic, further complicating the estimation task due to parameter uncertainties and fiber nonlinearities [4, 5]. Various factors affect signal quality, such as amplified spontaneous emission (ASE) noise, signal power, individual channel power, or channel spacing, all of which contribute to the signal-to-noise ratio (SNR). Additionally, the inherent nonlinearity of the fiber causes these factors to introduce intersymbol and interchannel interference, making QoT estimation a complex undertaking.

Traditional fault management approaches relying on conservative designs and fault detection alarms based on QoT thresholds are no longer sufficient in such complex optical networks. However, machine learning algorithms offer a promising solution for proactive

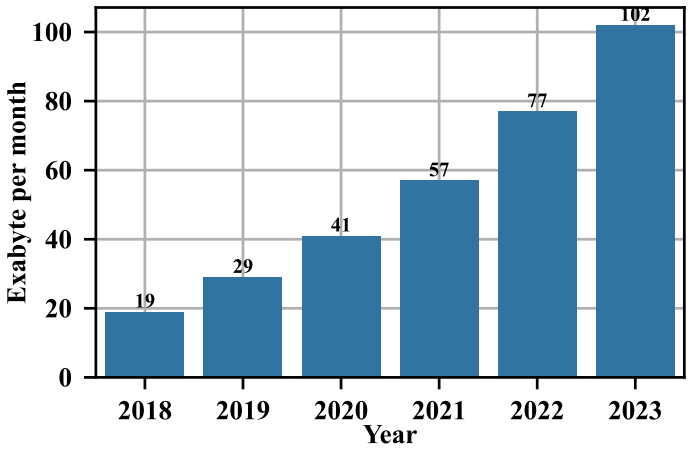


Figure 1.1: Total amount of global internet traffic over the years from 2018 to 2023 [1].

maintenance of future networks. These algorithms leverage a substantial amount of training data to operate reliably. Obtaining this training data and developing fault management machine learning algorithms require effective optical performance monitoring (OPM).

Since OPM of wavelength division multiplexing (WDM) systems is difficult, the usage of optical spectrum analyzers (OSAs) can play a crucial role in extracting optical spectrum information for machine learning frameworks. By incorporating transmission quality degrading failures, i.e., soft-failures, into the fault management processes, machine learning algorithms extend the capabilities beyond the limitations of QoT metrics like the bit error ratio (BER), optical signal-to-noise ratio (OSNR) or generalized optical signal to noise ratio (GOSNR). While fault detection can be achieved relying on thresholds of those metrics, the inability to adapt to changing network conditions inherently limits QoT-based fault identification and localization.

Today, fault localization is mostly relying on optical time-domain reflectometry (OTDR) measurements. While OTDR allows fault localization [6], it only provides span-by-span measurements and necessitates technical staff. This process is time-consuming and may lead to delays in addressing issues. By leveraging machine learning techniques, fault detection and identification capabilities are enhanced, enabling adaptive QoT thresholds and the recognition of changing patterns [7].

It can be seen that effective failure management and accurate QoT estimation are essential for ensuring the reliable operation and optimal performance of high-speed optical networks. By leveraging machine learning algorithms and innovative monitoring techniques, network

operators can proactively address failures, adapt to changing network dynamics, maximize capacity utilization, and enable the self-management of future networks [8, 9].

1.1 Goals and Applications

The goals of this work are based on the project description for the AI-NET-PROTECT subproject of the AI-NET project, which encompasses a set of defined objectives. In essence, the AI-NET-PROTECT subproject focuses on establishing automated resilient and secure networks on trusted devices for critical infrastructure and enterprises. The primary goal of the project is to safeguard critical data while ensuring optimal performance and high resilience of the network infrastructure. To achieve this, scalable network and node architectures are developed, utilizing both hardware and software, enabling the optimization and monitoring of essential performance parameters. These architectures leverage network telemetry and software defined networking (SDN) to attain complete automation of end-to-end services within the network using artificial intelligence techniques. Artificial intelligence (AI) is employed to optimize performance, proactively detect faults and anomalies, conduct vulnerability testing, and manage security. The security management aspect also encompasses the utilization of multi-layer cryptography and quantum-safe algorithms.

The main goals for this work are

- the development of novel AI-based techniques for enabling end-to-end performance predictions of different optical link paths in complex, meshed networks,
- developing AI-based methods for QoT estimation, and
- the development of AI-based methods for enhanced failure and anomaly detection, identification and localization.

These goals are achieved using in-depth literature research, simulations, failure modeling, laboratory experiments and measurements.

The algorithms and frameworks developed in this work, can be further used for automating performance optimization of optical networks as well as fully-automated soft-failure handling. The combined framework with QoT estimation and failure management could be deployed in an SDN controller with a monitoring data pipeline to unfold their full potential in disaggregated optical networks.

Table 1.1: Selected ML-based soft-failure management and QoT estimation literature overview.

Literature	Sim.	Exp.	OPM Data	(ML-)Algorithm	SFD Acc.	SFI Acc.	SFL Acc.	QoT MAE
Vela et al. [7]	√		Rx power, BER	Analytical model	99.06%	99.55%		
Furdek et al. [10]	√		BER, block errors, etc.	DBSCAN, SVM	96.2%			
Shariati et al. [11]	√		Optical spectrum (1 Ch.)	SVM	up to 100%	up to 100%		
Sun et al. [12]	√		Residual spectrum	AE+SVM		97.61%		
Lun et al. [13]		√	PSD	CNN		up to 100%		
Zhang et al. [14]		√	OTN data	BiLSTM		98.73%		
Mayer et al. [15]	√		Tx Power, OSNR	ANN			up to 100%	
Abdelli et al.[6]	√		OTDR	LSTM	92±1.06%		2.2±0.13 m	
Sun et al. [16]		√	Residual spectrum	AE+SVM		96.20%		
Aladin et al. [17]	√		Transmission parameters	GRU, LSTM				0.285 dB
Müller et al. [18]	√		Transmission parameters	XGB				0.1 dB
D'Amico et al. [19]		√	Transmission parameters	GNPy				0.6±0.9 dB
Kruse et al. [20]	√		Optical spectrum	LSTM				0.18 dB
Kruse et al. [21]	√	√	Optical spectrum	VAE+LSTM				0.2 dB
Kruse et al. [22]	√		Optical spectrum	VAE+SVM	98.9%	99.6%	90.08%	0.17 dB
Kruse et al. [23]		√	Optical spectrum	VAE+SVM	99.96%	99.05%	99.13%	0.29 dB

1.2 State of the Art

In the following, the state of the art regarding QoT estimation and soft-failure detection, identification and localization is presented briefly. Table 1.1 provides a concise overview of the literature discussing failure management and the estimation of transmission quality with the help of machine learning (ML).

1.2.1 Quality of Transmission Estimation

There have been several proposed approaches for estimating the QoT in optical links, with the goal of evaluating nonlinear impairments. However, these techniques always involve a trade-off between accuracy and speed. The most accurate method is the full-fiber propagation simulation using the split-step Fourier method (SSFM) [24], but it is computationally complex and unsuitable for real-time implementation.

Analytical tools like the Gaussian noise (GN)-model [25] offer acceptable accuracy with low computation time, although they are less accurate than the SSFM-based approach. Extensions of the GN-model, such as incoherent (IGN) and close-form methods, approximate the nonlinear interference to reduce computation time. However, both numerical and analytical approaches require precise knowledge of link parameters [19, 26], which can lead to less accurate QoT estimation [27]. Additionally, both models cannot adapt to the presence of soft-failures in the network. In contrast, ML-based estimation combines high accuracy and fast computation by training an estimator on input features that are correlated to the target. Although the training process can take minutes to days, once trained, the estimation is quick and suitable for real-time environments.

Recently, various approaches, including analytical [25], machine learning-based [17, 28], and hybrid methods [29, 30], have been explored to evaluate the performance of lightpaths

in optical networks using different metrics. The choice of a performance metric is crucial for proactive response to degradation or failures. The primary metric of interest for network designers is the BER of the lightpath, usually expressed as pre-forward error correction (FEC) BER, which determines performance acceptability [3].

What all of the mentioned approaches are missing is the consideration of uncertain parameters as mentioned in [4], meaning not exactly known component parameters in the field.

1.2.2 Failure Detection, Identification and Localization

In recent years, there has been a significant focus on exploring machine learning algorithms for the management of soft-failures, including their detection, identification, and localization. Soft-failures occur due to the degradation and aging of components in the networks, and if not addressed, they can evolve into hard failures that disrupt the service [7].

Vela et al. [7] considered signal overlap, tight filtering, gradual drift, and cyclic shift of filters as examples of soft-failures. They proposed an adaptive threshold mechanism for detecting changes in the BER and a pattern recognition algorithm for identifying these failures.

In [11], Shariati et al. focused on soft-failures caused by filters. They utilized features extracted from the optical spectrum and employed a support vector machine (SVM) for detection and identification purposes.

Lun et al. [13] adopted a one-dimensional convolutional neural network (CNN) applied to the power spectral density (PSD) obtained from a coherent receiver. Their approach aimed at identifying variations in the PSD that can indicate soft-failures, such as filter shift, filter tightening, increased ASE noise, and Kerr nonlinear effects resulting from an increase in launch power.

Mayer et al. [15] used an artificial neural network (ANN) to localize soft-failures based on limited telemetry data. They classified the failures by considering features such as OSNR and transmitter power. The identified failures included amplifier gain degradation, transponder power degradation, and additional fiber losses.

Fully automated failure management will need highly accurate OPM with metrics that mirror the actual state of the connection. This is why we present the concept of using OSAs as OPM and using the optical spectra for failure detection, identification and

localization. Furthermore, there are no frameworks that combine all previously mentioned stages into one. Our final approach combines not only failure management but also QoT estimation into one framework.

1.3 Structure of this Work

This work is divided into seven chapters. After the introduction, the fundamentals of coherent optical transmission systems are described in Chapter 2. First, the transmitter side is explained followed by the transmission channel, including the optical fiber with its linear and nonlinear effects, optical filters and optical amplifiers. Afterwards, the receiver side and in particular the optical-to-electrical conversion are described.

Chapter 3 introduces the basics of machine learning and basic machine learning algorithms of supervised and unsupervised learning are briefly presented. This is followed by the description of the advanced machine learning algorithms used in this work and an explanation of the metrics for performance evaluation. The chapter ends with a complexity analysis of machine learning algorithms with regards to big-O complexity.

In Chapter 4, the approach of using optical spectra as OPM data for QoT estimation is proposed, investigated and experimentally validated. Subsequently, this approach is used further for failure prediction, identification and localization with regards to soft-failures in optical networks in Chapter 5 based on experimental data. In Chapter 6, QoT estimation and soft-failure management are combined in a single framework. Finally, Chapter 7 ends the thesis with a conclusion and an outlook.

Chapter 2

Coherent Optical Transmission Systems

Optical transmission networks are the backbone of today's digital world. Their low latency and high bit rates enable global communication to be possible in the first place. The complex electric field of the optical signal E_S is defined as [31]

$$E_\mathrm{S}(t) = A_\mathrm{S}(t) \cdot e^{j2\pi f_\mathrm{S} t}, \tag{2.1}$$

where A_S is the amplitude and f_S is the frequency of the signal. Optical transmission systems can be divided into two groups, depending on which parts of Eq. (2.1) are used for modulation and detection as well as their underlying technologies. Simpler systems use the intensity of the optical signal for modulation (IM) together with direct-detection (DD) with a simple photodiode (PD). Such IM/DD systems are inexpensive, but they are severely limited. By modulating the intensity, in this case the amplitude, only one degree of freedom can be used. From Eq. (2.1) it can be seen that a change of further components of the optical signal brings a further modulation dimension with it, i.e., the phase of the signal. Using phase modulation ensures higher robustness against fiber propagation effects and extending transmission reach. By combining intensity and phase modulation, the number of phase states can be reduced for the same number of symbols leading to larger distances between the symbols which increases the spectral efficiency of the system as well as offering a higher robustness [32]. However, for phase modulated systems a simple detection with a PD is not sufficient since only the amplitude can be extracted with the absolute square operation employed which loses the phase information. For extracting the phase, coherent detection is needed.

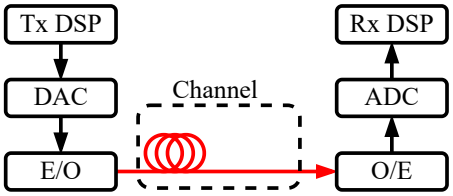


Figure 2.1: Basic setup of an optical transmission system.

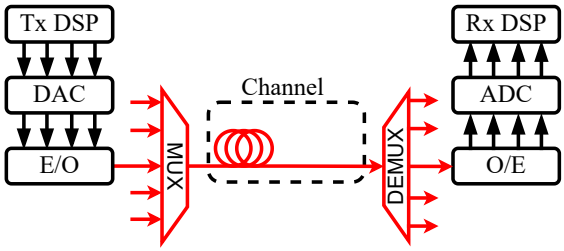


Figure 2.2: Basic setup of an WDM system with PolMux IQ modulation.

A simplified block diagram of an optical transmission system is depicted in Fig. 2.1. First, the transmitter digital signal processing (DSP) is applied, followed by the digital-to-analog converter (DAC) generating the analog electrical signal. This electrical signal is used to drive the external modulator which is used for the electrical-to-optical (E/O) conversion. The optical signal is transmitted over the channel in which optical amplification, filtering and the transmission over the fiber may be applied. At the end of the transmission, the signal is converted back to the electrical domain by a detector (optical-to-electrical (O/E) conversion). The analog-to-digital converter (ADC) generates the digital signal being processed by the receiver DSP.

In a transmission system where more than one channel is to be transmitted over the optical link, the technique of wavelength division multiplexing (WDM) is used. Different signals are assigned to different wavelengths with a specific channel spacing and are sent over the same fiber. The wavelength assignment is done in a multiplexer with either a fixed frequency grid or a flexible frequency grid configuration. Most of today's networks still use fixed grid channel assignments, since multiplexing and demultiplexing is way easier and more cost efficient than in the flexible grid case. A simplified WDM system with polarization multiplex (PolMux) inphase-quadrature (IQ) configuration for the example channel is depicted in Fig. 2.2.

In this chapter, the building blocks from Fig. 2.1 are discussed in their order of appearance. The relevant components as well as the underlying technologies are explained for deeper understanding of coherent optical transmission systems.

2.1 Transmitter Side

To transmit data over an optical link, the data has to be pre-processed in order to use the transmission medium as efficiently as possible. The pulse shaper is then used to convert the data signal into a continuous time electrical signal. Afterwards the electrical baseband signal is converted into an optical bandpass signal using a modulator which is mapping

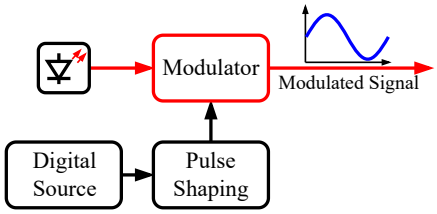


Figure 2.3: Basic setup of an optical transmitter.

the electrical signal onto an optical carrier from the continuous wave (CW) laser. A basic structure of an optical transmitter can be seen in Fig. 2.3. There are two main types of modulation used in optical systems: direct modulation and external modulation. In direct modulation, the laser itself performs the modulation, while external modulation relies on external devices like the Mach-Zehnder modulator (MZM) or the electro-absorption modulator (EAM). Direct modulation is a more cost-effective option but has limited bandwidth due to chirping, making it suitable for low data rates only. On the other hand, external modulation is well-suited for high data rates and long transmission distances. Due to the fact that only external modulation is used in this thesis, direct modulation is omitted. In the following, the transmitter DSP including modulation formats and digital pre-compensation, the digital-to-analog conversion including pulse shaping, and E/O conversion with information on MZM and polarization multiplexed IQ modulators will be described.

2.1.1 Transmitter DSP

For data that is to be transmitted over an optical link, DSP is applied in today's optical systems to optimize the spectral efficiency, the transmission reach, and the robustness of the optical signal. Transmitter DSP algorithms enable the use of advanced modulation formats and with this, higher data rates can be achieved. In this section, the main aspects of modulation formats, digital pre-compensation and pulse shaping are described.

Modulation Formats

As mentioned in the introduction of this chapter, higher order modulation formats offer the possibility of enlarging the overall throughput of an optical transmission. Assigning different amounts of data bits to one symbol is called modulation. The number of symbols M is usually a power of two. The amount of bits per symbol is called order and is calculated by $\log_2(M)$. Following this definition, higher order modulation means that a higher number of bits are contained in each symbol. The overall bit rate R_B is defined

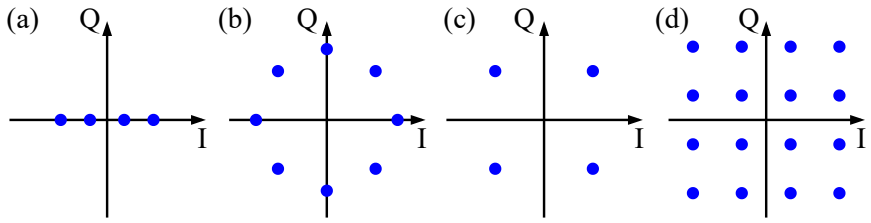


Figure 2.4: Constellation diagrams for (a) bipolar 4-PAM, (b) 8-PSK, (c) 4-QAM, and (d) 16-QAM.

as

$$R_{\mathrm{B}} = R_{\mathrm{S}} \log_2(M), \tag{2.2}$$

where R_{S} is the symbol rate. In order to improve R_{B}, R_{S} and/or M can be increased [32]. The mapping of the bits to the symbols is defined by the used modulation format. Modulation formats can be divided into real and complex formats. Those rely on the definition of the electrical field in Eq. (2.1): A modulation of the intensity will only change the amplitude of the signal, which means a shift on the real axis, of the inphase (I-) component. Such a modulation format is called pulse amplitude modulation (PAM). Complex modulation formats change the imaginary part of Eq. (2.1), i.e., the phase component. If the symbols are placed equidistant from the zero point, the modulation format is called phase shift keying (PSK). Combining amplitude and phase modulation reduces the number of phase states and the Euclidean distances of the symbols are increased. These modulation formats are called quadrature amplitude modulation (QAM) and can be easily extended for more constellation points and with this to very high data rates. The constellation points contain an I-component and a quadrature (Q-) component. Each symbol can thus be defined on a complex plane in a constellation diagram, also called IQ-diagram. Figure 2.4 contains examples of real and complex modulation formats.

Digital Pre-Compensation

For the compensation of impairments on the signal, the signal can be pre-distorted in the transmitter. This includes for example I-Q-imbalances or low pass characteristics of the transmitter caused by the arbitrary waveform generator (AWG), cables, electrical amplifiers or MZMs. The task of the pre-compensation is therefore to optimize the electrical signal before it is applied to the optical signal in the external modulator. Digital pre-compensation can be realized in practice by measuring the impulse response of the transmitter and then implementing a finite impulse response (FIR) filter.

Within the context of this work, the frequency response is evaluated by utilizing a multi-tone signal in an optical back-to-back (B2B) setup. In Section 4.4.1, where we present the coherent laboratory setup, the pre-distortion process takes into account both the frequency response and the skew of the AWG, exclusively [33, 34].

2.1.2 Digital-to-Analog Conversion (DAC)

A basic task is the conversion from a digital signal to an analog signal by a DAC, since only in this way a signal can be transmitted over a real channel. A DAC is characterized by the number of bits/sample N_S and its maximum sampling rate f_DAC as well as its output power [35]. An optical transmission system using a DAC is most likely limited towards higher data rates by f_DAC. Due to this, the DAC should be used with one sample per symbol, resulting in equal symbol and DAC rates [36]. In the likely case that the symbol rate is not equal to the DAC rate, resampling has to be applied to match both frequencies. In this case, the pulse shape of the transmitted signal is dependent on the frequency response of the DAC. However, it is possible to define the pulse shape before the DAC by pre-processing the signal using DSP. In this work, the only pulse shape used is the root-raised cosine (RRC) pulse [35]. Pulse shaping is needed to ensure that the detection of a symbol is not affected by neighboring pulses. This is achieved, if the spectrum of a pulse is as steep as possible. The RRC pulse, is parameterized by the roll-off factor $0 < \alpha_\mathrm{RRC} \leq 1$. α_RRC determines the additional bandwidth required compared to an ideal low-pass. For a roll-off factor of 0, the result is a Si-pulse, which corresponds to an ideal rectangular function in the frequency domain [35]. In contrast, $\alpha_\mathrm{RRC} = 1$ results in a bandwidth which is double of the ideal low-pass.

2.1.3 Electro-to-Optical (E/O) Conversion

For a transmission over an optical fiber, the electrical baseband signal has to be converted into an optical bandpass signal. This can either be done by a directly modulated laser (DML) or by an external modulator. DMLs suffer from frequency chirping above 2.5 Gb/s which makes them unsuitable for high data rate applications [37]. Moreover DMLs and single external MZMs can only be used to modulate one-dimensional information on the optical signal. External modulators change the incoming CW laser signal. If complex valued modulation formats are desired, IQ modulators which combine multiple MZMs, are used. These offer up to four degrees of freedom in the case of a PolMux IQ modulator. In the following, first the basic principle of an MZM is described followed by more information on PolMux IQ modulators, since these are the only type of modulators used in the remainder of this work.

Mach-Zehnder Modulator (MZM)

The MZM is typically based on Lithium-Niobate ($LiNbO_3$) and is built from one Mach-Zehnder interferometer (MZI). The function of the MZI is based on the Pockels effect, which describes the change of the refractive index of a waveguide by applying an external voltage [38]. By changing the refractive index, the phase of the optical signal is changed. The structure of an MZM is shown in Fig. 2.5(a). The incoming signal is split by a 3 dB-coupler and travels through two separate waveguides which are driven by electrical voltages individually. Afterwards, the two signals are recombined by a 3 dB-combiner and the output signal is formed.

The applied voltage to the MZM is given by

$$u(t) = u_{\mathrm{D}}(t) + u_{\mathrm{bias}}, \tag{2.3}$$

where $u_{\mathrm{D}}(t)$ is a DC-free drive signal and u_{bias} the bias voltage [38]. The MZMs used in the experiments in this work are driven in push-pull configuration. This means that they are used entirely for amplitude modulation. The voltages applied to the MZIs are defined in this case as $u_1(t) = -u_2(t)$. Although complex modulation is possible with an MZM, the more robust push-pull operation is preferred due to the sensitivity of the output signal depending on the bias voltage, ambient temperature and other influences. Furthermore, in push-pull configuration the MZM is completely chirp-free, theoretically [39]. The transfer function H_{MZM} can be derived as [39]:

$$H_{\mathrm{MZM}} = \cos\left(\frac{\pi}{2} \cdot \frac{u_{\mathrm{D}}(t) + u_{\mathrm{bias}}}{u_{\mathrm{D}}}\right). \tag{2.4}$$

Here, u_{D} represents the switching voltage, which is a characteristic for each MZM.

(PolMux) IQ Modulation

Realizing higher order modulation schemes that use amplitude and phase for symbol generation is challenging with a simple MZM. For modulating such complex signals on an optical carrier, IQ modulators are used. In Fig. 2.5(b), the setup of an optical IQ modulator is shown. In order to modulate the real and imaginary components of the signal individually, a nested MZM structure is employed. This modulator incorporates a 90-degree phase shift between its upper and lower branches. By operating the MZMs in push-pull configuration, the amplitude of both branches can be modulated [37].

If the X- and Y-polarization of the light should be utilized for two additional degrees of

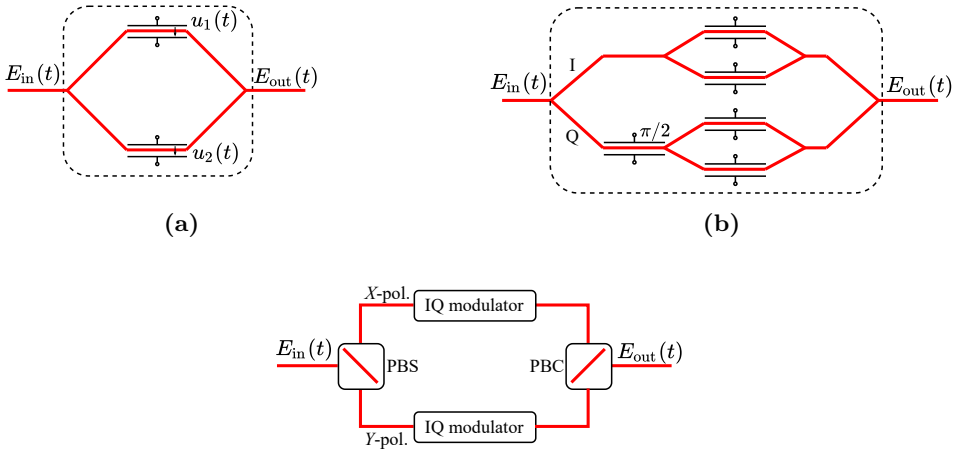


Figure 2.5: Schematic illustrations of (a) an MZM, (b) an IQ modulator composed of two MZMs and a 90° phase shift, and (c) a PolMux IQ modulator utilizing two IQ modulators, a polarization beam splitter (PBS) and a polarization beam combiner (PBC).

freedom for modulating the signal, two of these IQ-modulators are used in combination with a polarization beam splitter (PBS) as depicted in Fig. 2.5(c). First, the incoming signal undergoes polarization splitting by the before mentioned PBS to obtain two orthogonal polarizations, X and Y. Afterwards each polarization is modulated individually using an IQ modulator before the outputs are combined by a polarization beam combiner (PBC).

2.2 Transmission Channel

After looking at the transmitter setup, the transmission channel containing the optical fiber as the transmission medium, optical filters and optical amplifiers will be looked at in more detail.

2.2.1 Optical Fiber

As mentioned in the introduction of the chapter, the optical fiber is the preferred transmission medium when it comes to high data rates over long distances. Its low loss and wide usable bandwidth make it highly suitable for meeting the ever-increasing demand for data. Therefore, the optical fiber is the transmission medium used in this work. As shown in Eq. (2.1), the optical signal in an optical fiber propagates as an electromagnetic wave. For mathematically describing the propagation of light as an electromagnetic wave in a fiber, the nonlinear Schrödinger equation (NLSE) is used. Originated from the Maxwell equations the NLSE describes the behavior of the optical envelope $A(z,t)$ given

by [38]

$$\frac{\partial A}{\partial t} + \underbrace{\frac{\beta_1}{2}\frac{\partial A}{\partial t} + j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - \frac{\beta_3}{6}\frac{\partial^3 A}{\partial t^3}}_{1^{\text{st}},\, 2^{\text{nd}},\, \text{and } 3^{\text{rd}} \text{ order dispersion}} = \underbrace{j\gamma|A|^2 A}_{\text{nonlinear Kerr effect}} - \underbrace{\frac{\alpha}{2}A}_{\text{attenuation}} . \tag{2.5}$$

The NLSE includes the modeling of the most important linear effects, i.e., attenuation and dispersion, and nonlinear impairments which limit the transmission distances for fiber communications.

The following sections will further discuss the linear and nonlinear effects that affect the light propagating through the fiber. For a more in-depth view on effects in the optical fiber and their representation in the coupled NLSE, the reader is referred to [38, 40].

Linear Effects

There are two major linear effects on a fiber, namely attenuation and dispersion. The fiber losses were the main limiting factor in fiber optic communications since they reduce the signal power which is reaching the receiver before low-loss fibers and low-cost optical amplifiers were introduced.

The attenuation in a fiber can be described by the attenuation coefficient α which does not only include material absorption but also other sources of power attenuation [38]. The output power P_{out} is given by

$$P_{\text{out}} = P_{\text{in}} \cdot e^{-\alpha L}, \tag{2.6}$$

where P_{in} denotes the input power into the fiber and L is the length of the fiber. In general, α_{dB} is expressed in units of dB/km and is referred to as the fiber-loss parameter defined as [38]:

$$\alpha_{\text{dB}}\,[\text{dB/km}] = -\frac{10}{L}\log_{10}\left(\frac{P_{\text{out}}}{P_{\text{in}}}\right), \tag{2.7}$$

$$\alpha_{\text{Np}}\,[\text{Neper/km}] = \frac{\ln(10)}{10} \cdot \alpha_{\text{dB}}. \tag{2.8}$$

The fiber losses depend on the wavelength of the transmitted light. They emerge mostly from Rayleigh scattering, infrared and ultra-violet absorption, OH absorption, and bad splices or connectors. Typical values for the standard single mode fiber (SSMF) are $\alpha = 0.4$ dB/km at $\lambda = 1300$ nm in the O-band and $\alpha = 0.2$ dB/km at $\lambda = 1550$ nm in the C-band . In this work, the C-band is used for the transmission through the fiber.

For the simplified case of a linear fiber, meaning all nonlinear effects are omitted, the propagation of the light along the fiber can be modeled as the frequency response of the SSMF as follows [37]:

$$H_{\mathrm{SSMF}}(\omega, L) = e^{-\frac{1}{2}\alpha_{\mathrm{Np}} \cdot L} + e^{-j\beta(\omega L)}. \tag{2.9}$$

Here $\omega = 2\pi f$ denotes the angular frequency, L the fiber length, α_{Np} the fiber loss in Neper/km (see Eq. (2.8)), and $\beta(\omega)$ the propagation constant which includes the dispersion influences.

Dispersion leads to pulse broadening in time domain. This pulse broadening results from the frequency dependence of β. The propagation constant β can be expressed as a Taylor series with n components [37, 38]:

$$\beta(\omega) = \beta_0 + \beta_1\omega + \frac{1}{2}\beta_2\omega^2 + \frac{1}{6}\beta_3\omega^3 + ... + \frac{1}{n!}\beta_n\omega^n. \tag{2.10}$$

The first element of the Taylor series, β_0, is the phase constant at the carrier frequency and β_1 is the reciprocal value of the group velocity [37]. The higher-order components describe the dispersive fiber effects. The group velocity dispersion is described by β_2 which is related to the dispersion parameter:

$$D = -\frac{2\pi c}{\lambda}\beta_2. \tag{2.11}$$

Another typical value for describing the dispersion is the dispersion slope

$$S = \frac{dD}{d\lambda} = \frac{(2\pi c)^2}{\lambda^3}\left(\frac{1}{\lambda}\beta_3 + \frac{1}{\pi c}\beta_2\right), \tag{2.12}$$

which is related to the third-order dispersion parameter β_3 and the derivative of the dispersion parameter D to the wavelength λ. The dispersion slope is of increasing importance in today's WDM systems with large spectral bandwidth since the different channels suffer from different amounts of dispersion [37].

The lightwave travelling through a fiber is composed of two orthogonal polarization modes called X- and Y- polarization. The electrical field of the wave can then be described as

$$\mathbf{E}(\omega, L) = \begin{bmatrix} E_X(\omega, L) \\ E_Y(\omega, L) \end{bmatrix}. \tag{2.13}$$

The so-called state of polarization (SOP) of the electrical field is not constant during

the transmission along a fiber. Random rotations of the SOP occur along the fiber and polarization mode dispersion (PMD) is present. The effect of the rotating polarizations can be described using the rotation matrix $\mathbf{R}$, which includes the rotation angle ϕ_r [33]:

$$\mathbf{R} = \begin{bmatrix} \cos\phi_\mathrm{r} & \sin\phi_\mathrm{r} \cdot e^{-j\phi_\mathrm{p}} \\ -\sin\phi_\mathrm{r} \cdot e^{-j\phi_\mathrm{p}} & \cos\phi_\mathrm{r} \end{bmatrix}. \tag{2.14}$$

Here, ϕ_p is the phase shift along the X-axis and $-\phi_\mathrm{p}$ along the Y-axis.

PMD is related to fiber birefringence, which is caused by small deviations from perfect cylindrical symmetry because of different mode indices associated with orthogonally polarized components of the fundamental fiber mode [38]. Thus, the propagation constants differ for X- and Y-polarization due to non-perfectly cylindrical shaped fiber cores. An input pulse becomes broader as the two polarization parts of the pulse disperse differently along the fiber because of the underlying different group velocities. Since the fiber birefringence is randomly distributed along the fiber, PMD can be described by the standard deviation of the time delay between the polarization components [38]:

$$\sigma_T \approx D_\mathrm{PMD}\sqrt{L}, \tag{2.15}$$

where D_PMD is the PMD parameter which is measured to be in the range of $D_\mathrm{PMD} =$ 0.01 to 10 ps/$\sqrt{\mathrm{km}}$ and L is the length of the fiber.

For fiber lengths around 100 km the time delay σ_T is around 1 ps and can be ignored for pulse widths $>$ 10 ps [38]. In systems with high bit rates and long transmission distances, PMD can be a limiting factor [37].

Nonlinear Effects

When transmitting light over long distances, it is necessary to have a high input power to the fiber to compensate for attenuation and insertion losses. In any dielectric medium the refractive index gets power dependent in presence of an electric field. At high intensities these materials behave nonlinearly. This physical effect is called the Kerr effect [37, 38, 40] and can be described as

$$n(t) = n_\mathrm{linear} + n_2(P/A_\mathrm{eff}), \tag{2.16}$$

where n is the refractive index, n_linear is the linear refractive index, n_2 is the nonlinear refractive index, A_eff is the effective mode area and P is the power of the light. The

numerical value of n_2 varies for different core materials, however, for silicia fibers n_2 is in the range of $2.6 \cdot 10^{-20}$ m^2/W. Effects emerging from the Kerr effect are also called elastic effects, since the Kerr effect is energy conserving regarding the overall optical power [37].

When a single channel is transmitted, the Kerr effect induces self-phase modulation (SPM), resulting in a phase shift caused by the time-varying power of the optical signal. The phase shift is then converted into an amplitude shift through dispersion, i.e., group-velocity dispersion. If more than one channel is transmitted over a fiber, for example in a WDM system, cross-phase modulation (XPM) occurs. In WDM systems, the nonlinear phase shift for a specific channel depends not only on the power of that channel but also on the aggregate power of other channels. This induces a phase shift of other WDM channels based on the changing refractive index through the power exchange. The third effect caused by the Kerr effect is four-wave mixing (FWM). The name arises from its property of generating a fourth wave if three waves are propagating over a fiber. Especially for WDM systems, FWM is crucial, because the generated wave can fall into WDM channels and leads to undesired crosstalk. However, in systems with a moderate amount of dispersion, i.e., an SSMF in the C-band, FWM is less impactful. A typical WDM system is mostly distorted by XPM which is shown to be noise-like in the constellation diagrams [37]. The dependency of the SPM, XPM, and FWM is defined by the fiber-specific nonlinear coefficient γ given by

$$\gamma = \frac{n_2 \omega_0}{c_0 A_{\text{eff}}}, \tag{2.17}$$

where ω_0 is the angular carrier frequency and c_0 is the speed of light in vacuum.

Besides elastic nonlinear effects, optical transmission over a fiber is prone to inelastic effects. Those include scattering processes like stimulated Brillouin scattering (SBS) and stimulated Raman scattering (SRS). Both phenomena involve the scattering of a photon, resulting in the generation of a lower energy photon accompanied by a phonon that accounts for the energy difference. Besides SBS and SRS being quite similar from their origin, there are some major differences between them:

1. SRS is caused by optical photons while SBS involves acoustic phonons [40],
2. SBS only occurs in the backward direction of the incoming light, whereas SRS can happen in both directions [37], and
3. for SBS the frequency of the light is shifted by about 10 GHz but for SRS the frequency shift is around 13 THz (Stokes shift) [38].

The fact that SRS can lead to a power shift of the incoming light to a lower frequency

leads to its use in amplifier technologies like the Raman amplifier.

Since in this work long-haul WDM transmissions are assumed for most investigations, the nonlinear effects of XPM and FWM dominate. For more information on nonlinear effects and their understanding, the reader is referred to [38] and [40].

2.2.2 Optical Filters

Optical filters are crucial to WDM systems and are an important part of the optical communications infrastructure. They are essential for selecting specific wavelengths out of complex multiplexed signals at the receiver side of the system. The proper functioning of optical filters in WDM systems depends on critical considerations and design principles.

One of the main goals when designing tunable optical filters in WDM systems is to achieve a fine balance between transmitting the desired channel and effectively blocking unwanted neighboring channels. This task necessitates precise selection of the filter bandwidth, as a too narrow bandwidth may result in insufficient signal transmission, while a too wide bandwidth may allow unwanted signals to pass through.

Based on the physical mechanisms that determine their operation, optical filters can be divided into different types. These mechanisms primarily fall into two categories: optical interference and diffraction [38].

When considering tunable optical filters for use in WDM systems, it is important to identify and prioritize certain desirable characteristics. These characteristics include a wide tuning range to allow selection of a greater number of channels within the available spectrum [41]. Additionally, it is crucial to have minimal crosstalk between channels to avoid interference from adjacent channels, ensuring the transmission of data remains unaffected. It is also important to have a fast tuning speed for quick access to specific channels to improve system responsiveness.

Furthermore, reducing insertion loss is critical to minimize signal attenuation during the filtering process. Another important attribute is polarization insensitivity, which allows optical filters to accommodate different polarizations and assures compatibility with diverse signal sources. Stability against environmental changes, such as humidity, temperature, and vibrations, is essential for reliable and consistent filter performance. Cost-effectiveness is also a significant consideration, particularly in large-scale optical communication networks where efficiency and affordability are key.

There are many different types of optical filters. In the following, four types, each utilizing

different interferometric and diffractive devices to achieve their filtering capability are presented. These four kinds are

1. Fabry-Perot (FP) filters,
2. Mach-Zehnder filters,
3. Bragg grating filters, and
4. Acousto-optic filters.

Each of these filter types has its own advantages and limitations, making them suitable for particular applications and scenarios within the broader domain of WDM systems. The selection of a filter type depends on the specific requirements of the system, considering factors such as channel count, bandwidth, and performance criteria.

Fabry-Perot (FP) filters are based on a cavity formed by using two mirrors [42]. The tunability is achieved through electronic control, typically utilizing a piezoelectric transducer. By electronically controlling the length of the cavity, the filter can be precisely tuned to select specific channels within the WDM spectrum. As an example, for a 10-channel WDM signal with 0.8 nm channel spacing, the cavity length should be smaller than 100 μm [38]. Such a short length together with the requirement of high mirror reflectivities underscores the complexity of the design of FP filters for WDM applications. Other implementations of an FP filter include an all-fiber design with an air gap between two optical fibers [43], liquid crystals as mirrors [41, 44], semiconductor waveguides [45], and dielectric films for narrow-band interference filters [46].

Mach-Zehnder filters are optical filters that can be adjusted using a one or a series MZIs, where only one MZI is enough for intensity modulated transmission scenarios. To construct an MZI, two output ports of a 3-dB coupler are connected to two input ports of another 3-dB coupler. The input signal is then divided into two paths with different phase shifts by the first coupler. The phase shift is created by intentional variations in the length of the arms within the interferometer. The introduced phase shift is dependent on wavelength, resulting in the filter transmittance being inherently dependent on wavelength and enabling accurate wavelength discrimination. The cascaded chain of MZIs with adjusted relative delays acts then as a tunable optical filter. It can select closely spaced channels and can be built using fiber couplers or silica waveguides on a silicon substrate. Tuning in Mach-Zehnder filters can be achieved through a chromium heater deposited on one arm of each MZI, resulting in a slow response with a switching time of about 1 ms [38]. More information on Mach-Zehnder filters can be found in [38]

and examples of implementations can be found in [47–50].

A fiber Bragg grating acts as an optical filter because of the existence of a stop band [38]. This stop band is the frequency region in which most of the incident light is reflected and is centered at the so-called "Bragg wavelength". So basically, a fiber grating acts as a reflection filter. To make a fiber Bragg grating tunable, the Bragg wavelength has to be controlled by changing the grating period and the bandwidth has to be adjusted by changing the grating strength or by chirping the grating period slightly [38]. Wavelength selectivity can be achieved by a build-in grating whose Bragg wavelength is tuned electrically through electrorefraction [51]. In [52], a phase shift is applied in the middle of the grating to convert the fiber grating into a narrow band transmission filter. Many other schemes can be used to make transmission filters based on fiber gratings, like MZI-based filters with a grating in each arm [53], FP filters with gratings as mirrors [54], and Saganc and Michelson interferometers [55].

Tunable filters are for example used if a certain channel should be added or dropped from a WDM signal. Such a filter is called an optical add and dropp multiplexer (OADM). Furthermore, the overall multiplexing and demultiplexing of WDM channels is achieved using filters. Multiplexing and demultiplexing can be realized with all techniques described above, however, fiber Bragg gratings offer the advantage of all-fiber demultiplexers offering low losses and fast conversion times [38]. Multiplexers can be simply implemented using multiple directional couplers [38].

Other applications of filters are, for example, gain flattening, noise reduction and single-sideband (SSB) creation. New emerging reconfigurable optical add-drop multiplexer (ROADM) provide the feature of adjusting and changing the channel assignment of a WDM signal according to the traffic pattern. Today, fixed-grid ROADMs, meaning every wavelength has to fit inside a particular frequency slot, are replaced with wavelength selective switch (WSS) technology enhanced ROADMs. This allows the usage of a flexible grid approach enabling different channel bandwidths and spacings to be used in order to increase the spectral efficiency [3].

2.2.3 Optical Amplifiers

As discussed in Section 2.2.1, the reach of an optical transmission is limited by fiber losses. Historically, the first long-haul systems used optoelectronic repeaters for regenerating the signal [38]. However, this technology got way too slow and expensive with the emergence of WDM and high data rate systems. Today, long distance transmission is enabled by

the usage of optical amplifiers relying on scattering processes, for example Erbium-doped fiber amplifiers (EDFAs) [56]. Those devices amplify the optical signal without the conversion to the electrical domain. Amplifiers are used either before (pre-amplifier), inside (inline) or after the transmission link (post-amplifier). Since EDFAs are the only optical amplifiers used in this work, their underlying technology will be explained in the following. For other amplifier types and their technologies, the reader is referred to [38] and [57].

EDFAs consist of a rare-earth doped fiber, as the name suggests. This rare-earth dopand is chosen according to the desired output frequency range of the amplifier. For the case of a conventional EDFA, this frequency range is most of the C- and L-band. There are different structures of EDFAs, however, all are built up on the same technology: A pump laser with a significantly different wavelength than the incoming light (mostly 980 nm or 1480 nm) excites the erbium ions in the amplifier's fiber to their high-energy state. Upon encountering photons belonging to the signal at a different wavelength from the pump light, the excited erbium ions relinquish some of their energy to the signal and revert back to their lower-energy state [56]. In the process, it is noteworthy that the erbium releases its energy as additional photons, precisely matching the phase and direction of the signal being amplified. Consequently, the amplification of the signal occurs exclusively along its direction of travel [56]. As mentioned, the amplification effect of an EDFA is based on stimulated emission. However, spontaneous emission, i.e., a energy state change without any influence from outside, also occurs. This spontaneous emission is also amplified in the EDFA, resulting in undesired ASE noise. In a link with EDFAs, the ASE noise is the major noise source overtaking thermal noise or shot noise sources by far.

An amplifier is characterized by its ability to boost the power of the incoming signal. This measure is called the gain G and is defined as

$$G = \frac{P_{\text{out}}}{P_{\text{in}}}. \tag{2.18}$$

For an EDFA, the gain is wavelength dependent resulting in a non-flat gain characteristic. However, modern EDFAs employ gain flattening filters after amplification to account for possible imbalances in the amplification of high bandwidth WDM signals in the C-band. The influence of the ASE noise of an amplifier is characterized by its impact on the SNR. The relation between the SNR at the input and the output of the amplifier is called the noise figure and is defined as [38]

$$F_{\text{N}} = \frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}}. \tag{2.19}$$

Commercially available EDFAs offer a noise figure as low as $F_{\mathrm{N}} = 5$ dB. The noise power spectral density N_{ASE} of an EDFA is calculated as [37]

$$N_{\mathrm{ASE}} = n_{\mathrm{sp}} \cdot \mathrm{h} \cdot f_{\mathrm{c}} \cdot (G - 1), \tag{2.20}$$

where n_{sp} is the spontaneous emission factor, f_{c} is the carrier frequency, h is the Planck constant and G is the amplifier gain.
The optical power of the noise is measured over a reference bandwidth B_{ref} (usually 0.1 nm) and is given by [37]

$$P_{\mathrm{ASE}} = 2 \cdot N_{\mathrm{ASE}} \cdot B_{\mathrm{ref}}. \tag{2.21}$$

For a more in-depth look on EDFAs the reader is referred to [56].

2.3 Receiver Side

When looking back at Fig. 2.1, the next part of a transmission system is the receiver side containing the O/E conversion, the ADC and the receiver side DSP. In the following, the different parts of the receiver side will be explained in more detail.

2.3.1 Optical-to-Electrical (O/E) Conversion

In a simple intensity modulation/direct detection (IM/DD) system, the detection is realized with a single PD. If a (PolMux) IQ modulator is used to employ higher order modulation formats, a receiver which is able to extract amplitude and phase is required. Such a receiver is called a coherent receiver. In this work, coherent detection is used in every investigation. Due to this, only the coherent receiver will be introduced as a device for O/E conversion. With coherent detection the phase is preserved which allows for an

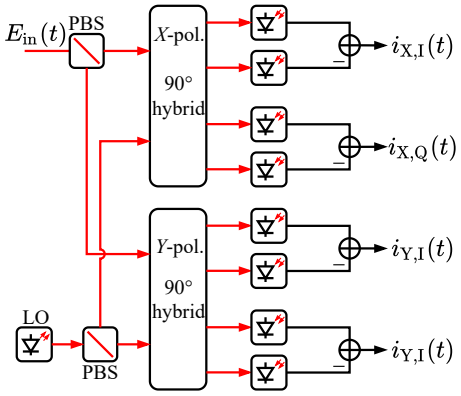


Figure 2.6: Block diagram of a coherent PolMux receiver.

electronic dispersion compensation (EDC) as well as allowing for carrier synchronization by the use of a local oscillator (LO) for down-conversion.

A coherent receiver (Fig. 2.6) is built out of four pairs of balanced PDs to extract the four quadrature components of the X- and Y-polarization. The incoming light is first separated for the two polarizations before being mixed in two 90° hybrids with the LO signal. The outputs are then detected by the mentioned pairs of balanced PDs.

The photocurrents at the output of the PDs are calculated as follows [31, 38, 58]:

$$i_{\mathrm{I}}(t) = R_{\mathrm{PD}}\sqrt{P_{\mathrm{LO}}P_{\mathrm{sig}}} \cdot \cos(\Phi_{\mathrm{sig}}(t) - \Phi_{\mathrm{LO}}(t)), \tag{2.22}$$

$$i_{\mathrm{Q}}(t) = R_{\mathrm{PD}}\sqrt{P_{\mathrm{LO}}P_{\mathrm{sig}}} \cdot \sin(\Phi_{\mathrm{sig}}(t) - \Phi_{\mathrm{LO}}(t)), \tag{2.23}$$

where R_{PD} denotes the responsivity of the photodiode, P_{LO} the local oscillator power, P_{sig} the signal power, Φ_{sig} the signal's phase, and Φ_{LO} the phase of the local oscillator. The difference in phase between the LO and the signal must be compensated in the DSP by carrier frequency offset (CFO) recovery algorithms [59].

2.3.2 Analog-to-Digital Conversion (ADC)

In order to extract the digital information contained in the analog signal, the signal has to be converted back to a digital sequence. As the counterpart to the DAC described in Section 2.1.2, the ADC samples and digitizes the analog input signal. It can be modeled as a low-pass filter, a sampler and a quantizer. Like the DAC, the ADC is characterized by its maximum sampling frequency f_{ADC}, the resolution in bits and the bandwidth. A signal processed by an ADC suffers from the same impairments as its counterpart at the transmitter side, i.e., the DAC. These impairments are the quantization noise originating from the limited number of amplitude levels being available by the converter and inter-symbol interference (ISI) which occurs due to bandwidth limitations of the converters.

2.3.3 Receiver DSP

The digital sequence after the ADC can be further processed to extract the information. The simplest way would be the use of a simple decider for symbol extraction. In a system with no noise or other impairments, these symbols would not be distorted in any way meaning they would be located exactly on their transmitted constellation points. In a real-world transmission system, however, the signal is distorted by the fiber effects described in Section 2.2.1 and noise originating from the amplifiers, DAC and ADC, the PDs and many more electrical components. This leads to a deviation from the ideal

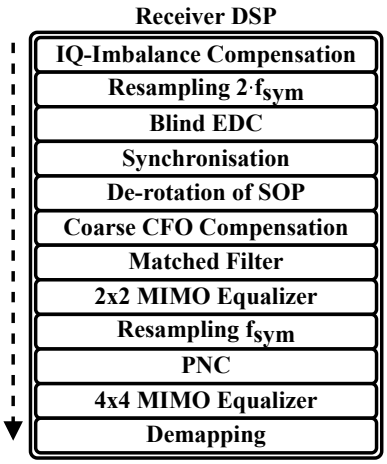


Figure 2.7: DSP chain for coherent PolMux transmission used in this work; EDC: electrical dispersion compensation; SOP: state of polarization; CFO: carrier frequency offset; PNC: phase noise compensation; MIMO: multiple input multiple output.

constellation points.

Some of these impairments can be compensated for in the DSP. In Fig. 2.7, the DSP chain used in this work is shown. The DSP includes the dispersion compensation, IQ imbalance compensation, SOP de-rotation, phase noise compensation, and equalization. However, all this is only possible, if the signal is synchronized. If this is not the case, no information can be extracted, since it is not recognizable when the data block starts. Synchronization is achieved using pilot sequences within the data blocks.

Although the receiver DSP described in the following sections can compensate for most impairments, there are more accurate DSP chains that exist. These include joint detection and decoding or iterative techniques, where the output of one DSP block is fed back to a preceding block for another iteration of already executed algorithms [60].

Within the receiver side DSP, all algorithms operate at two samples per symbol (SpS), except for carrier recovery using blind phase search (BPS) and the 4x4 multiple input multiple output (MIMO) equalizer. These are executed at symbol intervals (1 SpS). The arrangement of the DSP chain is designed to counteract impairments in the reverse order in which they occur. Thus, compensation starts with receiver-based impairments and finishes with transmitter-based impairments. This approach recognizes that some effects in the system are influenced by later effects. For example, when evaluating transmitter skews, compensating for IQ-mixing from CFO and phase noise is crucial since they directly affect the estimation of the skews.

IQ-Imbalance Compensation

As a first step in the receiver side DSP, the IQ-imbalances originating from the receiver front-end are compensated. The skews of the receiver setup can be measured using cross-correlation [33] and are set directly in the software of the real-time oscilloscope.

In an ideal scenario, the IQ components of each polarization should be mutually orthogonal, resulting in a received constellation that exhibits circular rotation. This rotation is mainly caused by phase variations introduced by factors such as CFO and laser phase noise. However, when there are IQ imbalances at the receiver, the received constellation deviates from circularity and takes on an elliptical shape. To address these IQ imbalances, a corrective approach involves adjusting the scaling of the quadrature components and introducing a phase shift to the imaginary part of the signal, which restores the desired orthogonality between the quadratures. Like this, the receiver side IQ-distortions are compensated for using the Gram-Schmidt orthogonalization procedure (GSOP) [61, 62].

To correct the received signal $s_{\mathrm{Rx,X/Y}}$, the imaginary and real parts of the signal have to be corrected for both polarizations X and Y. The corrected real part is given by

$$\hat{s}_{\mathrm{I,X/Y}}(k) = \frac{\mathrm{Re}\{s_{\mathrm{Rx,X/Y}}(k)\}}{\sqrt{E_{\mathrm{I}}}}, \tag{2.24}$$

$$\text{with } E_{\mathrm{I}} = \frac{1}{N}\sum_{k=0}^{N-1} \mathrm{Re}\{s_{\mathrm{Rx,X/Y}}(k)\}^2, \tag{2.25}$$

where N is the length of the data sequence. The corrected imaginary part can be derived as

$$\hat{s}_{\mathrm{Q,X/Y}}(k) = \left(\mathrm{Im}\{s_{\mathrm{Rx,X/Y}}(k)\} - \left(\frac{E_{\mathrm{IQ}}}{E_{\mathrm{Q}}} \cdot \mathrm{Re}\{s_{\mathrm{Rx,X/Y}}(k)\}\right)\right) \cdot \frac{1}{\sqrt{E_{\mathrm{I}}}} \tag{2.26}$$

$$\text{with } E_{\mathrm{Q}} = \frac{1}{N}\sum_{k=0}^{N-1} \mathrm{Im}\{s_{\mathrm{Rx,X/Y}}(k)\}^2, \text{ and} \tag{2.27}$$

$$E_{\mathrm{IQ}} = \frac{1}{N}\sum_{k=0}^{N-1} \mathrm{Re}\{s_{\mathrm{Rx,X/Y}}(k)\} \cdot \mathrm{Im}\{s_{\mathrm{Rx,X/Y}}(k)\}. \tag{2.28}$$

Combining imaginary and real parts, the corrected signal $\hat{s}_{\mathrm{Rx,X/Y}}$ can be computed as

$$\hat{s}_{\mathrm{Rx,X/Y}} = \hat{s}_{\mathrm{I,X/Y}}(k) + j \cdot \hat{s}_{\mathrm{Q,X/Y}}(k). \tag{2.29}$$

The IQ-imbalances of the receiver can be considered as almost static. This leads to

an adjustment of the correction terms being only necessary for every new measurement.

Blind Electrical Dispersion Compensation

Coherent detection allows for electronic dispersion compensation (EDC) since phase information can be retrieved. For EDC, an algorithm for blind estimation of accumulated dispersion can be used. Blind EDC is based on the delay-tap sampling method that can be used for signals with an oversampling factor of 2 SpS [33]. In this algorithm, the standard deviations σ of two consecutive samples are compared [63].

Coarse EDC can be achieved using an all-pass filter which has the inverse phase response of the fiber. However, this is only possible when the accumulated dispersion is known. The chromatic dispersion (CD) can be computed by multiplying the dispersion coefficient D and the transmission length L. Nevertheless, for blind EDC several test values for the accumulated dispersion are computed following a pre-defined step size. With those values for the CD, the following cost function can be computed for a sequence with length N_{FFT} [63]:

$$F_D(DL) = \frac{\sigma(\delta_{\mathrm{X}}(DL) + \sigma(\delta_{\mathrm{Y}}(DL))}{2}, \tag{2.30}$$

where $\delta_{\mathrm{X/Y}}$ is defined as

$$\delta_{\mathrm{X/Y}} = \frac{|s_{\mathrm{Rx,X/Y}}(2k+1, DL)|^2 - |s_{\mathrm{Rx,X/Y}}(2k, DL)|^2}{\sqrt{2}},\ k \in \{0, ..., \frac{N_{\mathrm{FFT}}}{2} - 1\}. \tag{2.31}$$

If the power of sample k is identical to the power of sample $k+1$, $\delta_{\mathrm{X/Y}}$ is 0. The variable $\delta_{\mathrm{X/Y}}$ can thus be interpreted as the power difference between two consecutive samples. The cost function averages this difference over both polarizations for higher stability of the estimation. The cost function F_D has a minimum at the correct test value of the accumulated dispersion [33].

There are multiple approaches to determine this minimum value [33]. The first approach involves an iterative process where potential values for the cumulative dispersion are tested until the cost function F_D falls below a specified threshold. The choice of this threshold depends on factors such as the modulation format, pulse shape, and OSNR. Alternatively, when there is approximate knowledge of the link parameters, another method involves testing predefined distance ranges and searching for the minimum of the cost function within this range. This approach supports parallel implementation and may offer greater efficiency. A third method employs a two-stage estimation process, initially

using a coarse step size in the first estimation, followed by a more refined search [33]. In all these methods, the accuracy and complexity of the estimation are influenced by the step size used for the cumulative dispersion.

Frame Synchronization

A modified Schmidl and Cox approach is used to address the difficulty in performing cross-correlation based synchronization when a CFO or dispersion is present in the signal [64]. For this purpose, a training sequence of length N_{TS} is used which consists of two identical halves:

$$d_{\mathrm{CAZAC}}(k) = d_{\mathrm{CAZAC}}(k + N_{\mathrm{TS}}/2), k = 0, 1, ..., N_{\mathrm{TS}}/2 - 1. \tag{2.32}$$

The method for detecting the beginning of a frame involves sliding auto-correlation of the received signal. This repetitive structure enables the identification of the frame's start. First, a window of size $N_{\mathrm{TS}}/2$ is extracted from the received signal, and its correlation with a subsequent sequence of equal length is calculated. The correlation with the matching (identical) half is computed when the sliding window aligns with the initial half of the received training sequence. Since these two halves are only correlated within the signal, this leads to a significantly greater correlation compared to the rest of the signal, resulting in a distinctive peak in the timing metric.

A constant amplitude zero auto-correlation (CAZAC) sequence is used because of its special properties regarding the auto-correlation. The CAZAC sequence is given as [65]

$$d_{\mathrm{CAZAC}}(k) = \begin{cases} e^{jM\pi k^2/N_{\mathrm{TS}}}, & N_{\mathrm{TS}} \text{ even} \\ e^{jM\pi(k+1)^2/N_{\mathrm{TS}}}, & N_{\mathrm{TS}} \text{ odd} \end{cases} \text{with } k = 0, 1, ..., N_{\mathrm{TS}}/2 - 1. \tag{2.33}$$

Here, M is an integer that is relatively prime to $N_{\mathrm{TS}}/2$. In Schmidl and Cox synchronization schemes, so-called timing plateaus can occur. To overcome this, a pseudo noise sequence of length N_{TS} is used to weight the CAZAC-sequence [33, 66]. This pseudo noise sequence d_{PN} has the values ± 1 and is also known at the receiver for descrambling.

The timing metric at symbol d using a sliding correlation window of length $N_{\mathrm{TS}}/2$ is computed based on the auto-correlation of the received signal [33, 66]:

$$J(d) = \frac{|A(d)|^2}{R(d)^2}, \tag{2.34}$$

where $A(d)$ is the auto-correlation of the descrambled received signal given by

$$A(d) = \sum_{k=0}^{N_{\mathrm{TS}}/2} d_{\mathrm{PN}}(k) s_{\mathrm{Rx}}^{*}(d+k) \cdot d_{\mathrm{PN}}\left(k + \frac{N_{\mathrm{TS}}}{2}\right) \cdot s_{\mathrm{Rx}}\left(d + k + \frac{N_{\mathrm{TS}}}{2}\right). \tag{2.35}$$

The normalization factor $R(d)$ is defined as

$$R(d) = \frac{1}{2} \sum_{k=0}^{N_{\mathrm{TS}}-1} |s_{\mathrm{Rx}}(d+k)|^2. \tag{2.36}$$

The maximum of $J(d)$ indicates the starting point of the frame and thus synchronization is achieved.

De-rotation of the State of Polarization

It is necessary to de-rotate the state of polarization (SOP), to ensure steady compensation of the CFO and to achieve swift convergence of the channel estimation during equalization [33, 67]. Since the SOP is given by the rotation matrix $\mathbf{R}$ from (2.14), the training symbols $N_{\mathrm{TS,SOP}}$ with special properties are used. Those are 4-QAM symbols which are multiplied with a 2x2 Walsh-Hadamard matrix [33]. A Walsh-Hadamard matrix is a square matrix whose entries are either +1 or -1 and whose rows are mutually orthogonal. As a result, two sequential symbols are identical in the X-polarization, while in the Y-polarization, the same symbol is used with the second symbol having a phase shift of π. The different symbols in the training symbol alphabet $\mathbf{d}_{\mathrm{SOP}}$ are given by [67]:

$$\mathbf{d}_{\mathrm{SOP}} = \begin{bmatrix} d_X(k) & d_X(k+1) \\ d_Y(k) & d_Y(k+1) \end{bmatrix} = e^{j\theta} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \tag{2.37}$$

with $\theta \in \{\pm\pi/4, \pm 3\pi/4\}$. Because the matrix is orthogonal between the polarizations, blind estimation of the SOP can be achieved.

The compensation of the SOP is done using the received training symbols from $s_{\mathrm{Rx}}(k)$ and multiplying them with the compensation matrix $\mathbf{R}_{\mathrm{c}}$. The compensated signal $\hat{s}_{\mathrm{Rx}}(k)$ can be derived as [67]:

$$\hat{s}_{\mathrm{Rx}}(k) = \mathbf{R}_{\mathrm{c}} \cdot s_{\mathrm{Rx}}(k), \tag{2.38}$$

where

$$\mathbf{R}_{\mathrm{c}} = \begin{bmatrix} |a| & |b| e^{j(\arg[a]+\arg[b])} \\ |b| e^{j(\arg[a]+\arg[b])} & |a| \end{bmatrix}. \tag{2.39}$$

The correction terms a, which can be interpreted as a correlation between the two received symbols of each pair of training symbols, and b are represented as [33, 67]:

$$|a| \approx \sqrt{\frac{1}{2} + \frac{1}{2N_{\mathrm{TS,SOP}}} \mathrm{Re}\left\{ \sum_{n=0}^{\frac{N_{\mathrm{TS,SOP}}^2}{2}-1} s_{\mathrm{Rx},S}(2n) s_{\mathrm{Rx},X}^*(2n+1) - s_{\mathrm{Rx},Y}(2n) s_{\mathrm{Rx},Y}^*(2n+1) \right\}}, \tag{2.40}$$

$$|b| = \sqrt{1-|a|^2}, \tag{2.41}$$

and

$$\arg[a] + \arg[b] = \arg\left[- \sum_{n=0}^{\frac{N_{\mathrm{TS,SOP}}}{2}-1} s_{\mathrm{Rx},X}(2n) s_{\mathrm{Rx},Y}^*(2n+1) - s_{\mathrm{Rx},Y}^*(2n) s_{\mathrm{Rx},Y}(2n+1) \right]. \tag{2.42}$$

In all cases, the product of $\mathbf{R} \cdot \mathbf{R}_{\mathrm{c}}$ results in an identity matrix with positive or negative sign so that the SOP is rotated back [33].

Compensation of Carrier Frequency Offset

As it can be seen in Fig. 2.7, the compensation of the carrier frequency offset (CFO) is done in two stages. At first, a coarse compensation of the CFO is done before the matched filter and the second stage is located after the first equalizer to compensate for the residual CFO and phase noise.

The coarse CFO compensation is done using a periodogram-based estimation [68], since this technique does not require training symbols. In this technique, the property of QAM symbols of the constant fourth power of the symbols, $\mathrm{E}[d(k)^4] = \text{const.}$, is used. With this, the phase modulation of the data signal is removed so that the expectation value of the fourth power of a QAM-signal is constant [33].

In case of a frequency offset, a distinct spectral component appears in the expectation value of the fourth power of the received signal, which exactly matches the carrier and LO frequency discrepancy multiplied by four. Therefore, determining the CFO can be effectively accomplished using the following method:

A sequence of N_{FFT} samples of the signal is taken and the fourth power is applied. Using a fast Fourier transform (FFT), the signal is converted into the frequency domain. As the last step, a periodogram is generated by squaring the absolute values of the

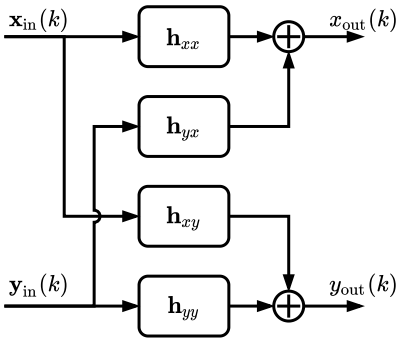


Figure 2.8: Block diagram of an 2x2 MIMO equalizer.

frequency spectrum. This periodogram shows a prominent peak at precisely four times the frequency difference between the carrier and the LO. With this peak, the CFO can be easily computed by identifying the maximum value in the periodogram.

However, it is essential to recognize that the accuracy of this estimate depends on the frequency resolution of the periodogram, which is determined by the dimensions of the FFT used in the process.

There are also other methods based, for example, on training symbols for compensating the CFO. For more information, the reader is referred to [33, 64].

2x2 MIMO Equalization

For compensation of residual ISI, PMD and residual rotation of the SOP, a 2x2 MIMO equalizer is used. Its butterfly structure can be seen in Fig. 2.8. The equalizer is operated at 2 SpS. The complex valued output sequences for the X- and Y-polarizations are given by [69]:

$$\begin{bmatrix} \mathbf{x}_{\text{out}}(k) \\ \mathbf{y}_{\text{out}}(k) \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{xx}{}^{\text{T}} & \mathbf{h}_{yx}^{\text{T}} \\ \mathbf{h}_{xy}^{\text{T}} & \mathbf{h}_{yy}^{\text{T}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\text{in}}(k) \\ \mathbf{y}_{\text{in}}(k) \end{bmatrix}, \tag{2.43}$$

where $\mathbf{x}_{\text{in}}$ and $\mathbf{y}_{\text{in}}$ are the input signals for the X- and Y-polarization, respectively. The filtering in the four complex-valued filters can be performed in the frequency domain using the overlap-save technique [33]. Since the equalizer has the purpose of erasing residual distortions, the number equalizer coefficients can be chosen to be low.

The initial channel estimation is performed with the training symbols using the minimum mean square error (MMSE)-criterion [33]. Changes of the channel over time are tracked with the decision-directed least mean squares (LMS)-algorithm.

The channel's impulse response is typically not known in advance, necessitating a channel estimation process at the receiver. The algorithms presented in the following are designed to be applicable to complex-valued signals as well.

Channel estimation can be achieved using the MMSE-criterion which aims to minimize the mean square error (MSE) at the output of the equalizer, resulting in the cost function F_{MSE}:

$$F_{\mathrm{MSE}} = \mathrm{E}\{|x_{\mathrm{out}}(k) - x_{\mathrm{in}}(k)|^2\} \Rightarrow \min_{\mathbf{h}}. \tag{2.44}$$

With the help of training symbols d_{TS} known at the receiver, the channel estimation can be computed in the time domain as follows [33]:

$$\mathbf{h}_{\mathrm{MMSE}} = {\mathbf{R}_{xx}}^{-1} \cdot \mathbf{r}_{xd}, \tag{2.45}$$

where $\mathbf{R}_{xx}$ is the auto-correlation matrix of the received training symbols and $\mathbf{r}_{xd}$ is the cross-correlation vector of the received training symbols and the undisturbed training symbols.

When the channel is changing over time, the LMS-algorithm can be used to apply an adaptive filter to update the filter coefficients during the transmission based on the equalized signal points [64]. The update of the filter coefficients can be computed from the current coefficients $\mathbf{h}_{\mathrm{LMS}}(k)$ and the derivation of the MSE [33]:

$$\mathbf{h}_{\mathrm{LMS}}(k+1) = \mathbf{h}_{\mathrm{LMS}}(k) - \mu \frac{\delta F_{\mathrm{MSE}}}{\delta \mathbf{h}^{\mathrm{H}}_{\mathrm{LMS}}(k)}. \tag{2.46}$$

Here, the superscript H is the Hermitian transposition and μ is the step-size of the LMS-algorithm.

Usually, the instantaneous error, i.e., the deviation from the decided symbol to the output of the equalizer, $\varepsilon(k) = \hat{d}(k) - x_{\mathrm{out}}(k)$ is used instead of F_{MSE} because the expectation value of the MSE cannot be computed at the receiver. The so-called decision-directed LMS can be described as follows [33, 70]:

$$\mathbf{h}_{\mathrm{LMS}}(k+1) = \mathbf{h}_{\mathrm{LMS}}(k) - \mu \varepsilon^*(k) \cdot \mathbf{x}_{\mathrm{in}}(k). \tag{2.47}$$

It has to be noted, that the LMS-algorithm increases the convergence time significantly, if the step size is chosen too small. A large step size, however, leads to a less accurate estimation and thus to a worse equalization performance [33].

Phase Noise Compensation

For the compensation of residual CFO and laser frequency deviations, the second stage of the carrier recovery from the phase noise compensation (PNC) uses blind phase search (BPS) [33, 71]. The following algorithm is designed to operate at 1 SpS and for modulation formats with a rotational symmetry of $\pi/2$. When the received signal is only distorted by phase noise and white noise, the received signal contains the transmitted symbols with a certain phase offset $\Delta\varphi(k)$ at the time instance k.

Because of the rotational symmetry of the modulation formats, the BPS can be performed using a set of N test-angles to rotate the received symbols. The test-angles are defined as $\varphi_{\text{test}} = \frac{n\pi}{2N} - \frac{\pi}{4}$ with $n \in \{0, 1, ..., N-1\}$. After the rotation by the test-angle, the difference between the rotated symbols and the rotated symbols after the decider is computed. Using a window on the received signal, the influence of the additive white noise can be removed. Also, the distance to the nearest constellation point can be computed. With minimizing this distance, the estimated phase difference of the received symbol with respect to the transmitted symbol is derived. To avoid the uncertainty of $\pi/2$ at the initialization of the algorithm, the decision operation can be replaced using the known training symbols [33]. A more in-depth look on PNC can be found in [33, 71].

4x4 MIMO Equalization

In high-speed communication systems, especially those operating at high symbol rates, various challenges arise, including signal degradation caused by timing mismatches and IQ-skews. These issues, often in the range of tens of picoseconds, are significant limitations to achieving reliable data transmission.

The IQ imbalances, resulting from propagation delays in both X- and Y-polarizations, can be attributed to a range of factors. These include differences in radio frequency (RF) cable lengths, variations in signal paths, varying RF driver phase responses, and timing mismatches within the DAC channels. As modulation format orders increase, the system's tolerance against these IQ imbalances decreases, making the need for mitigation strategies more pronounced. Furthermore, IQ imbalances are not only the result of propagation delays and timing issues, but are also influenced by amplitude responses in DAC channels, variations in RF cable characteristics, differences in RF amplifier performance, and biases in amplitude modulators. It is crucial to address these imbalances since they can significantly degrade the system's overall performance. In general, skews remain constant over time, IQ imbalances, however, exhibit more rapid variations, making their block wise compensation necessary.

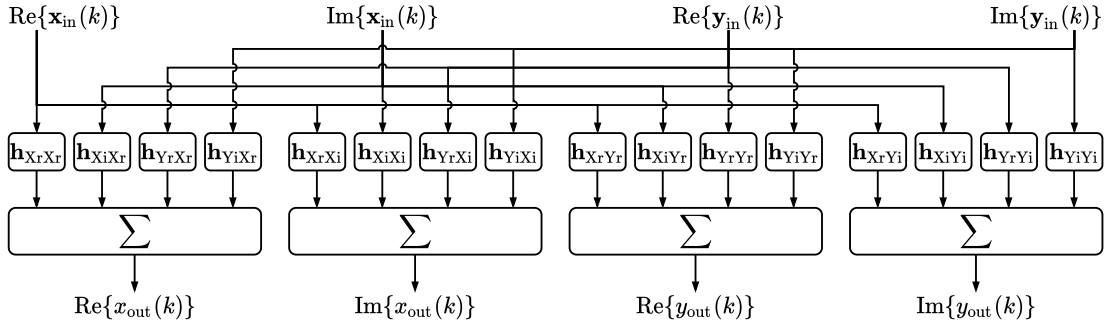


Figure 2.9: Block diagram of an 4x4 MIMO equalizer.

To overcome these challenges, various strategies are used. DSP algorithms are used in the skew compensation stage to minimize distortions by introducing delays that align signals correctly. To address IQ imbalances, it is essential to track and estimate them in real-time within the DSP at the receiver. Employing the previously described techniques such as GSOP or blind moment estimation can help to accurately assess and correct these imbalances.

To compensate for residual IQ-imbalances, skews, ISI, polarization rotations, PMD, and phase noise (PN), a 4x4 MIMO equalizer can be used. It also uses adaptive channel estimation techniques to guarantee an optimal performance in dynamic and challenging transmission scenarios.

While the 2x2 MIMO equalizer jointly filters the inphase and quadrature components, the 4x4 equalizer processes the four quadratures of the incoming signal independently [33, 72]. By this the imbalances and mixing between the quadrature components can be removed.

An overall structure of the equalizer can be seen in Fig. 2.9. The four quadratures of the received signal after the PNC are used as the input signals for the 16 real-valued filters. It follows for the output signals of the MIMO equalizer:

$$\begin{aligned}
\operatorname{Re}\{x_{\text{out}}\} &= \mathbf{h}_{\text{XrXr}}{}^{\text{T}} \operatorname{Re}\{\mathbf{x}_{\text{in}}\} + \mathbf{h}_{\text{XiXr}}{}^{\text{T}} \operatorname{Im}\{\mathbf{x}_{\text{in}}\} + \mathbf{h}_{\text{YrXr}}{}^{\text{T}} \operatorname{Re}\{\mathbf{y}_{\text{in}}\} + \mathbf{h}_{\text{YiXr}}{}^{\text{T}} \operatorname{Im}\{\mathbf{y}_{\text{in}}\}, \\
\operatorname{Im}\{x_{\text{out}}\} &= \mathbf{h}_{\text{XrXi}}{}^{\text{T}} \operatorname{Re}\{\mathbf{x}_{\text{in}}\} + \mathbf{h}_{\text{XiXi}}{}^{\text{T}} \operatorname{Im}\{\mathbf{x}_{\text{in}}\} + \mathbf{h}_{\text{YrXi}}{}^{\text{T}} \operatorname{Re}\{\mathbf{y}_{\text{in}}\} + \mathbf{h}_{\text{YiXi}}{}^{\text{T}} \operatorname{Im}\{\mathbf{y}_{\text{in}}\}, \\
\operatorname{Re}\{y_{\text{out}}\} &= \mathbf{h}_{\text{XrYr}}{}^{\text{T}} \operatorname{Re}\{\mathbf{x}_{\text{in}}\} + \mathbf{h}_{\text{XiYr}}{}^{\text{T}} \operatorname{Im}\{\mathbf{x}_{\text{in}}\} + \mathbf{h}_{\text{YrYr}}{}^{\text{T}} \operatorname{Re}\{\mathbf{y}_{\text{in}}\} + \mathbf{h}_{\text{YiYr}}{}^{\text{T}} \operatorname{Im}\{\mathbf{y}_{\text{in}}\}, \\
\operatorname{Im}\{y_{\text{out}}\} &= \mathbf{h}_{\text{XrYi}}{}^{\text{T}} \operatorname{Re}\{\mathbf{x}_{\text{in}}\} + \mathbf{h}_{\text{XiYi}}{}^{\text{T}} \operatorname{Im}\{\mathbf{x}_{\text{in}}\} + \mathbf{h}_{\text{YrYi}}{}^{\text{T}} \operatorname{Re}\{\mathbf{y}_{\text{in}}\} + \mathbf{h}_{\text{YiYi}}{}^{\text{T}} \operatorname{Im}\{\mathbf{y}_{\text{in}}\}.
\end{aligned} \tag{2.48}$$

The filtering in the 4x4 equalizer is done at 1 SpS in this work [33]. As for the 2x2

equalizer, the MMSE-criterion with training symbols is used for initial channel estimation and adaption is done using the LMS channel estimation. Again, the filtering can also be done in frequency domain using the overlap-save technique [33]. In [33], the performance of the 4x4 equalizer is evaluated. The results show that the usage of a second MIMO-equalizer is necessary in our lab environment to ensure proper BER during the measurements.

2.4 Performance Metrics

In order to evaluate the performance of a transmission, measurable quantities must be available. The optimization of different performance metrics for optical links is an essential part of this work. For this reason, this section derives the GOSNR, which is based on the SNR, the OSNR, the BER and the Q-factor.

Signal-to-Noise Ratio

The ratio between the signal power P_{S} and the total noise power P_{N} in the system is called the signal-to-noise ratio (SNR) and is defined as [73]

$$\mathrm{SNR} = \frac{P_{\mathrm{S}}}{P_{\mathrm{N}}}. \tag{2.49}$$

In general, the SNR represents the overall signal quality after the receiver side sampling and can be improved, for example, by equalization.

Optical Signal-to-Noise Ratio

If now only optical power and optical noise is assumed to be present in both polarizations, the OSNR can be defined. However, calculating optical noise requires a certain reference bandwidth B_{ref} which is set to be 0.1 nm or 12.5 GHz for a carrier wavelength of 1550 nm [37]. It follows for the OSNR in a system with an optical power P_{opt} over the optical bandwidth B_{opt} and the noise power P_{ASE}:

$$\mathrm{OSNR} = \frac{P_{\mathrm{opt}}}{P_{\mathrm{ASE}}} \cdot \frac{B_{\mathrm{opt}}}{B_{\mathrm{ref}}}. \tag{2.50}$$

In practice, the OSNR is measured using an OSA. The OSA obtains the optical PSD with a fine resolution bandwidth. The optical PSD should contain the signal as well as the out-of-band noise floor. From the noise and the signal power levels, i.e., integrating over the different parts of the signal, and normalizing with B_{ref}, the OSNR can be obtained. It has to be noted, that for the case of a WDM transmission, the measurement of the OSNR is difficult. Since the OSNR of one channel has to be obtained the other channels should

not be considered when measuring the signal power. Another problem is the measurement of the out-of-band noise floor, which might be out of the bandwidth of the OSA in a dense WDM system. A solution for this problem would be switching off different channels in order to measure noise power more accurately. Another technique would be increasing the channel spacing of the WDM in such a way, that the measurement bandwidth of the OSA only includes one channel. By this method, the signal power in the window of the OSA corresponds to one channel while the overall optical power is not reduced. The latter is used in this work for OSNR measurements of WDM systems. It has to be noted that in a real-world deployment both techniques are not applicable. This is one reason for the problem of optimizing the OSNR in order to reduce margins.

Bit Error Ratio

Since the information in a transmission system are contained in bits, errors in bits are undesired, because information can be lost. A metric to evaluate the signal quality after the transmission is the so-called bit error ratio (BER). It is defined as the ratio between the bit errors to the total number of sent bits. To calculate the number of errors, the detected bit sequence is compared to the originally sent bit sequence. For a reliable estimation of the overall BER at least 100 bit errors have to occur [37]. This approach is also called the "Monte Carlo" method. In today's networks, however, schemes are used to correct bit errors like FEC which lowers the BER towards zero. This is why for the evaluation of the quality of a transmission, the BER before the FEC is used as a metric; the so-called pre-FEC BER.

The requirement for the BER in modern optical communication systems lies in the range of BER $\leq 10^{-12}$ or preferably BER $\leq 10^{-15}$ [74]. If such a BER is reached, the transmission is considered error free. However, to reach such regions, FEC needs to be applied. To lower the effort in simulations and experiments, FEC limits are assumed. Those limits are set by the correction capabilities of the underlying FEC code. This means, if a pre-FEC BER below this limit is reached, it can be assumed, that the FEC code with a specific overhead can correct the errors to reach the desired BER.

In general, two types of FEC codes can be distinguished, based on the decoding at the receiver: hard decision (HD)-FEC and soft decision (SD)-FEC. For the case of HD-FEC an overhead of 7 % is assumed which enables a compensation of an BER of $3.8 \cdot 10^{-3}$. For reaching lower BER boundaries, the more complex SD-FEC codes are used which add more overhead to the payload while lowering the pre-FEC treshold. For the long-haul transmission scenarios in this work, the SD-FEC limit defined in the OpenROADM

standard is assumed [75]. The so-called openFEC code adds an overhead of 15.3 % lowering the pre-FEC limit to $2 \cdot 10^{-2}$.

Q-Factor

Another metric is the quality factor Q which is directly derived from the BER by [76]

$$Q = \sqrt{2}\mathrm{erfc}^{-1}(2 \cdot \mathrm{BER}). \tag{2.51}$$

The Q-factor is typically expressed in decibel (dB), which provides a logarithmic scale that allows for convenient representation of a wide range of values. This logarithmic scale helps to highlight small differences in performance, particularly in systems with very low BER values.

Generalized Optical Signal-to-Noise Ratio

Recently, the performance of an optical transmission is evaluated using the GOSNR. The GOSNR is calculated with the received power P_{Rx}, the ASE noise power P_{ASE} and the power of the nonlinear impairments P_{NLI}:

$$\mathrm{GOSNR} = \frac{P_{\mathrm{Rx}}}{P_{\mathrm{ASE}} + P_{\mathrm{NLI}}}. \tag{2.52}$$

The power of the nonlinear impairments, however, is not easy to measure. Besides the nonlinearities caused by the optical link, the OSNR incorporates impairments originating from the transceiver, as it is unable to differentiate between the various components of the noise term within the OSNR. This constraint necessitates the introduction of the GOSNR, which denotes the OSNR value at which an equivalent BER is attained in the back-to-back transmission using the back-trace method [21, 77]. Consequently, the GOSNR exclusively accounts for the optical impairments resulting from the optical link, encompassing both noise and nonlinear interference. Furthermore, due to the challenges associated with obtaining the OSNR within a dense WDM signal, the estimation of GOSNR is limited to the destination node exclusively.

Optical Spectrum Analysis

As described in the previous section, the performance of an optical link can be evaluated using different metrics. However, the signal can also be monitored using the optical spectrum. In practice, the optical spectrum can be extracted from the signal using an OSA. An exemplary spectrum is depicted in Fig. 2.10, showing five channels of 32 Gbaud transmission in a WDM signal with 37.5 GHz spacing. The spectrum shows the frequency components of the signal with their magnitude. This allows different information to be

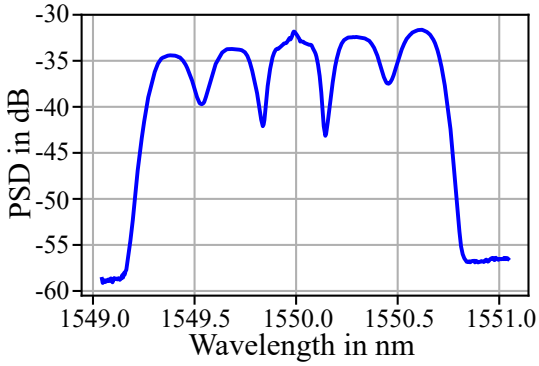


Figure 2.10: Exemplary spectrum of an experimentally obtained spectrum for a five channel WDM bandwidth-loaded signal after 88.4 km transmission; signal channel at center, neighboring channels generated by shaping of ASE noise.

extracted and derived. First, the noise power can be extracted by extrapolating and integrating the left and right noise floor. Second, the overall signal power can be extracted by integrating over the entire spectrum. From this, the individual channel powers can be derived. Furthermore, frequency information is visible like the individual channel bandwidths, the channel spacings, the center frequencies, and the overall signal bandwidth as well as the number of channels in a WDM system by simply counting the peaks in the spectrum. Overall, much information regarding the state of the signal is contained in the spectrum. The advantage of using an OSA in a real-world environment is that the status of the signal and its channels can be monitored without a total demodulation of the signal. This gives the opportunity to connect the optical spectrum with the other performance metrics to construct a comparable monitoring approach.

2.5 Simulation Techniques

Solving the NLSE is not a trivial problem due to its nonlinear partial differential behavior. There are analytical methods established on assumptions regarding physical behavior (e.g. the GN-model) and numerical solutions like the SSFM. In this work, the SSFM has been employed to simulate the light propagation over a fiber for single-channel and multi-channel as well as PolMux multi-channel transmission. For multi-channel system modeling, the GN-model has emerged as a fast and easy way to calculate system margins. However, it comes with major downsides, especially for the scope of this work. To give an overview of the simulation techniques available, this section describes the functionality of the SSFM and elaborates on the advantages and disadvantages of both the GN-model and the SSFM.

2.5.1 Split-step Fourier Method (SSFM)

For simulating X- and Y-polarizations using the SSFM, the general NLSE from Eq. (2.5) has to be extended [37, 38]. Since the SSFM solves the linear and nonlinear parts of the NLSE independently, the linear part $\hat{L}$ and a nonlinear part $\hat{N}$ are added to the equation:

$$\frac{\partial A(z,t)}{\partial z} = (\hat{L} + \hat{N})A(z,t) \tag{2.53}$$

$$\Rightarrow \frac{\partial}{\partial z}\begin{pmatrix} A_X \\ A_Y \end{pmatrix} = (\hat{L} + \hat{N})\begin{pmatrix} A_X \\ A_Y \end{pmatrix}. \tag{2.54}$$

Here, the linear part $\hat{L}$ is defined as

$$\hat{L} = -\frac{\alpha}{2} + \frac{1}{2}\begin{pmatrix} \Delta\beta_1 & 0 \\ 0 & -\Delta\beta_1 \end{pmatrix}\frac{\partial}{\partial t} - j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2} - \frac{\beta_3}{6}\frac{\partial^3}{\partial t^3}, \tag{2.55}$$

and the nonlinear part $\hat{N}$ as

$$\hat{N} = j\gamma\begin{pmatrix} |A_X|^2 + \frac{2}{3}|A_Y|^2 & 0 \\ 0 & |A_Y|^2 + \frac{2}{3}|A_X|^2 \end{pmatrix}. \tag{2.56}$$

The linear part $\hat{L}$ is solved in the frequency domain, while the nonlinear part $\hat{N}$ is solved in the time domain [37]. The transformation between time and frequency domains, and vice versa, is obtained by fast Fourier transforms (FFTs) or inverse fast Fourier transforms (IFFTs), respectively.

To understand the influence of the step size on the result of the SSFM, the analytical solution of (2.53) resulting from the local integration over the step width l is given by:

$$A(z+l,t) = \exp\left(\int_z^{z+l}[\hat{L} + \hat{N}(z',t)]dz'\right). \tag{2.57}$$

If the step size is chosen to be small, the linear and nonlinear operators can be calculated according to [34]

$$A(z+l,t) \approx \exp(l\hat{L})\exp(l\hat{N})A(z,t). \tag{2.58}$$

It can be seen, that if the step size would be too large, the results will be inaccurate because the interactions of the linear and nonlinear parts cannot be considered adequately.

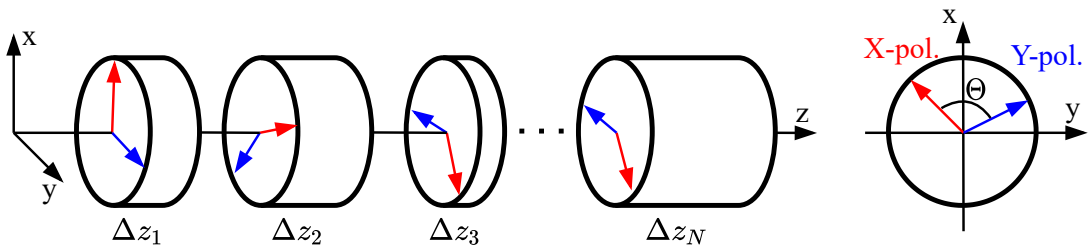


Figure 2.11: Visualization of the wave plate model; adapted from [37].

However, if the step size is chosen too small, the computational efficiency is compromised. Typically, multiple criteria are employed to restrict the maximum step-size, and the strictest criterion (resulting in the smallest step-size) is always used. These criteria encompass limiting the maximum nonlinear phase shift $\varphi_{\mathrm{NL,max}}$ between two split-steps, the maximum amount of artificial FWM, and the maximum walk-off between two WDM channels [37]. This nonlinear phase shift is defined as:

$$\varphi_{\mathrm{NL,max}} = \gamma |A_{\mathrm{max}}|^2 \Delta z. \tag{2.59}$$

For a high accuracy, $\varphi_{\mathrm{NL,max}}$ is typically set to be less than 0.1 mrad. Following from Eq. (2.59), the maximum nonlinear phase shift is amplitude dependent. The amplitude of the signal is degrading over the transmission distance due to the attenuation. Due to this, the accuracy can be maintained by enlarging the step width at higher transmission distance leading to a decreased computational complexity.

The SSFM also suffers from overestimating the influence of FWM on the signal if the step size is constant over the entire simulation. To conquer this, alternating random step sizes within the allowed maximum nonlinear phase shift are used together with an overall small step size. By providing an overall maximum FWM threshold, an overestimation is also prevented [33, 34, 37].

As described in Subsection 2.2.1, small differences from a perfect cylindrical shape of the fiber lead to fiber birefringence. This is also expressed by different propagation constants β_X and β_Y for X- and Y-polarizations, respectively. Due to the origin of fiber imperfections, birefringence is varying randomly along the fiber which leads to the definition of PMD through its standard deviation in Eq. (2.15). Simulating a randomly distributed PMD influence in the SSFM is achieved using the so-called wave plate model. Here the fiber is modeled as a series of birefringence elements resulting in random rotation of the angle Θ [37]. Additionally, the length of each element Δz as well as the phase of

the mode coupling are also random variables. By applying these randomly distributed elements, the behavior of PMD can be modeled. The wave plate model assumes that the fiber consists of a limited number of wave plates which contribute to the given overall PMD value. Thus, the higher the number of wave plates, the more accurate the estimation. A high level visualization of the wave plate model is depicted in Fig. 2.11. However, the computational complexity also increases with including PMD in the SSFM. More information on the wave plate model can be found in [37].

2.5.2 Gaussian Noise (GN) Model

As shortly described in the introduction of this section, the GN-model is an analytical approach for modeling optical transmission systems. Since its introduction in 2012 [25], the GN-model has become one of the most popular methods for QoT estimation, especially with the open source implementation in GNPy [78].

The GN-model finds an approximate solution for the NLSE with three assumptions being made:

1. The transmitted signal behaves statistically like stationary Gaussian noise [25],
2. the nonlinear interference (NLI) behaves like additive Gaussian noise [25, 79], and
3. the overall contribution of the nonlinearity is relatively small compared to the overall signal power [25].

Furthermore, the simple GN-model is restricted to equal span lengths within a link and equal channel spacing in a WDM signal.

The Gaussian noise model reference formula (GNRF) for a WDM system is given by [79]

$$\begin{aligned} F_{\mathrm{NLI}}(f) = \frac{16}{27}\gamma {L_{\mathrm{eff}}}^2 \\ \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{\mathrm{WDM}}(f_1)F_{\mathrm{WDM}}(f_2)F_{\mathrm{WDM}}(f_1+f_2-f) \\ \cdot \rho(f_1,f_2,f)\cdot\chi(f_1,f_2,f)df_2df_1, \end{aligned} \tag{2.60}$$

with $F_{\mathrm{WDM}}(f)$ describing the PSD of the transmitted signal at the frequency f and L_{eff} being the effective fiber length, i.e., the 'strength' of the frequency components. The normalized FWM efficiency ρ assuming EDFA amplification is defined as

$$\rho(f_1,f_2,f) = \left|\frac{1-e^{-2\alpha L_{\mathrm{s}}}e^{j4\pi^2\beta_2 L_{\mathrm{s}}(f_1-f)(f_2-f)}}{2\alpha - j4\pi^2\beta_2(f_1-f)(f_2-f)}\right|^2 \cdot {L_{\mathrm{eff}}}^{-2}, \tag{2.61}$$

where L_{eff} is used to normalize the maximum of ρ to 1. The factor χ takes into account the coherent interference at the receiver location of the NLI produced in each span. It is derived as

$$\chi(f_1, f_2, f) = \frac{\sin^2(2N_s\pi^2(f_1 - f)(f_2 - f)\beta_2 L_s)}{\sin^2(2\pi^2(f_1 - f)(f_2 - f)\beta_2 L_s)}. \tag{2.62}$$

By calculating the NLI contribution, together with the launch power and the power from the ASE noise, the GOSNR is calculated following Eq. (2.52). For more information on the behavior of the GN-model, the reader is referred to [25, 79–81].

There are several extensions to the simple GN-model that either increase its accuracy at the cost of computational power (closed form GN-model), removing limitations such as identical span lengths [79] or removing the Gaussian approximation from all NLI to form a much more complex model (extended GN-model or EGN) [81]. Overall the GN-model is based on assumptions which makes it less accurate than a detailed SSFM simulation. However, the computation time of a whole WDM signal takes between 1 second and 2 hours, depending on the GN-model extension used. In general, the higher accuracy of the model should be, the longer it takes for the computation. Another limitation of the GN-model is the fact that it can only output a certain performance metric instead of the whole signal envelope like the SSFM does. This makes the GN-model unsuitable for the generation of optical spectra in simulations.

2.6 Experimental Techniques

Hardware cost is a significant constraint for experiments of optical transmission systems in laboratories. With the need for cost-effective experimental solutions, different approaches are being considered. One challenge is the replication of dense WDM systems, where each channel would require a laser, modulator and receiver, so the need for transceivers drastically increases as the number of channels increases, placing a significant demand on laboratory setups. Another challenge arises for experiments with long transmission distances where additional fibers and amplifiers are required. As a result, more affordable and efficient techniques have emerged to enable the advancement of optical transmission research and development. The experimental techniques used in this work are described in the following including the concept of bandwidth loading and the description of a recirculating loop.

2.6.1 Bandwidth Loading

As presented in Subsection 2.2.1, the optical fiber channel is subject to Kerr nonlinearity, causing nonlinear phase shifts, and the signal is distorted from the interaction of the propagating WDM channels, the nonlinearity and ASE noise [82]. In order to make predictions for the system performance of WDM optical transmission systems, these predictions are commonly derived from experimental investigations wherein the transmission link is rarely provided with a fully loaded WDM spectrum or even independent channels [83]. To reduce the required hardware resources and the complexity in the experiments, simplifications have to be applied to the system. Individual optical channels would require large laser banks and high amounts of optical modulators. A common approach to reduce the number of modulators is the odd-and-even-channel approach [84]. However, a higher or lower degree of correlation between the channels will be present depending of the implementation [83]. To further reduce the hardware demand in the experiments, the usage of ASE noise for bandwidth loading to emulate WDM channels has emerged [82]. The ASE shaping can be implemented such that the spectral shape of actual data channels is replicated or as flat continuous ASE bands without gaps between neighboring emulated WDM channels [83]. The bandwidth loading scheme in this work is based on shaping ASE noise using a Waveshaper to emulate WDM channels.

2.6.2 Recirculating Loop

A recirculating loop is a combination of optical components that can be used to emulate long optical fiber transmissions in laboratory environments. A recirculating loop instead of a traditional transmission structure offers several advantages for the investigation of optical long-haul systems, e.g. long transmission distances can be achieved with only small amounts of available fiber, which takes up less space than actual long-haul fiber, and multiples of the loop length can be evaluated without much effort regarding rebuilding the experimental setup. A block diagram of the used recirculating loop in this work is depicted in Fig. 2.12. Recirculating loops can be built in different ways as it is described in [85–87], however, the basic functionalities and the underlying components will be described in the following.

To couple an optical signal into the recirculating loop, the first acousto-optic modulator (AOM) is opened, and then closed after a duration longer than the signal duration to ensure that a complete signal sequence is present in the loop. The signal is subsequently directed to both outgoing branches at the 2x2 coupler. It goes to the Boston Applied Technologies, Inc. (BATi) switch on one hand, and passes through the loop on the other hand. After the loop's completion, the signal is permitted through the coupler using

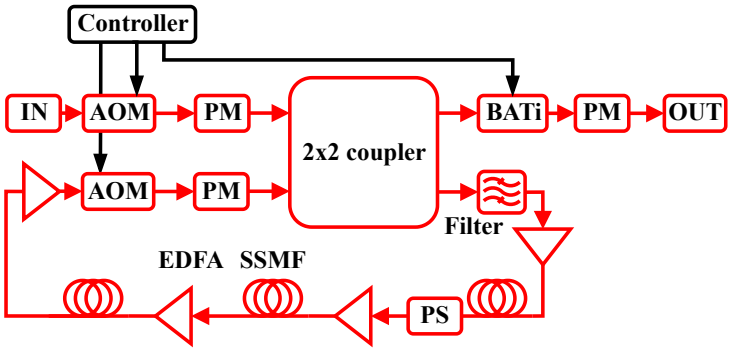


Figure 2.12: Block diagram of a recirculating loop; AOM: acousto-optic modulator; EDFA: Erbium-doped fiber amplifier; PM: power meter; BATi: switch from Boston Applied Technologies, Inc.; SSMF: standard single mode fiber; PS: polarization scrambler.

the second AOM. It is subsequently transmitted to the BATi switch on one end, and is redirected back into the loop on the opposite end. This course of action is repeated until the second AOM is shut down, which occurs once the desired maximum number of loop passes is successfully reached.

In the receiver branch at the BATi switch input, the signal from the transmitter arrives with a time delay during each loop round trip. However, the BATi switch forwards only the specific loop pass designated for consideration at the receiver. To ensure the loop runs properly, it is necessary to know the round trip time (RTT), i.e., the time required for a loop round trip.

All of the timing of switching the AOMs and the BATi switch is done by a delay generator,

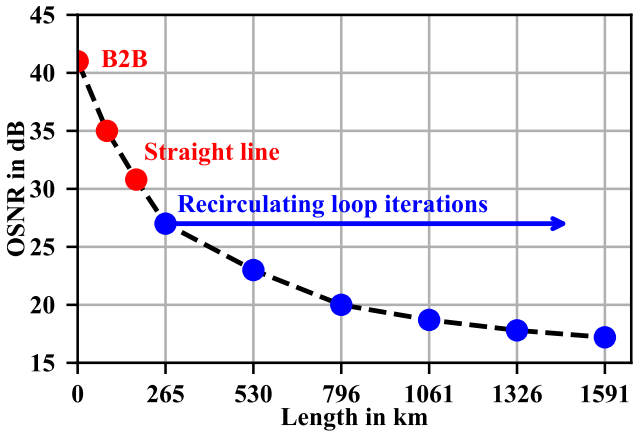


Figure 2.13: Experimentally measured OSNR of a 32 Gbaud dual-polarization (DP)-quadrature phase shift keying (QPSK) signal for different link lengths with launch power of 0 dBm and a single channel.

also called the control unit. Other tasks involve providing the electrical signals for the polarization scrambler and for synchronization of the receiver side real-time scope or the OSA.

In addition to the effects that occur in a fiber optic link, such as ISI, CD, or nonlinear effects, there are additional challenges when operating a recirculating loop. The installation of additional components within the link under test causes additional losses in the system, which limits the OSNR budget as it can be seen in Fig. 2.13. Additionally, the components cause unintended effects on the signal. Thus, using AOMs leads to a frequency offset of 27.12 MHz, which can be either compensated by another inverse AOM or during the receiver side DSP. The recirculating loop serves as a useful tool for studying long-range optical transmission, although it has certain limitations. For example, the delay generator restricts the minimum RTT, since the switching has to be at least as fast as the RTT of the loop. This limits the length of the fibers and the number of spans in the loop.

Chapter 3

Machine Learning Algorithms

AI is a field that utilizes computers and machines to perform cognitive tasks, including knowledge processing, perception, learning, reasoning, and understanding. An AI system is designed to store and utilize knowledge, apply it to solve problems, and acquire new knowledge through experience. The key components of an AI system include knowledge representation, ML, and automated reasoning [88].

ML, which is a branch of AI, employs algorithms to automatically learn patterns and trends from datasets [89]. Although ML algorithms have been around since the 1960s [90], they have experienced significant advancements due to improved computational power, developments in theory and algorithms, and the availability of vast amounts of data. One advanced field within ML is deep learning, which has demonstrated superior performance in various domains [88, 90].

The impact of ML on society is anticipated to be even more substantial as it is today with it's usage in medicine, weather forecasting, translation and much more. The emergence of ChatGPT and other large-language models (LLMs) provides a good glimpse into the future. Thus, the possible applications of ML are numerous, with applications in areas such as web searches, computer translation, content filtering on social media, healthcare, finance, and law. It can be seen, that ML is an interdisciplinary field, intersecting with statistics, optimization, information theory, and game theory [91].

ML algorithms typically focus on two types of pattern recognition tasks: regression and classification. Regression predicts values for new inputs, while classification assigns data points into categories. ML excels in scenarios where explicit descriptions of the underlying physics and mathematics are challenging. This is the case for the exact solving of the NLSE which is crucial for optimal tuning of the optical communication networks.

In the field of optical communication systems the applications of ML are plenty. In the past ten years, coherent detection and DSP techniques have played crucial roles in optical transceivers for fiber-optic communication. Advanced modulation formats like

16-QAM and above, along with DSP-based estimation and compensation of transmission impairments, drive innovation in optical communication. Parameter estimation and symbol detection involve regression and classification tasks, where probability theory and an understanding of the underlying physics are utilized to derive estimation and decision rules. High-capacity optical transmission links are often limited by transmission impairments, such as fiber nonlinearity. ML techniques are being researched and applied for fiber nonlinearity compensation. ML is also flourishing in short-reach direct detection systems affected by chromatic dispersion, laser chirp, and imperfections in transceiver components. Currently, ML-based QoT estimation, symbol detection or equalization are a research area [88].

In general, ML algorithms require a large amount of data to be trained properly. This data can be obtained using monitoring techniques. In OPM, ML is employed to acquire real-time information about channel impairments across the network. OPM often faces cost limitations and relies on simple hardware components and partial signal features. ML becomes necessary in OPM when the mapping between input and output parameters is complex and cannot be easily derived from underlying physics and mathematics. SDN frameworks are driving major shifts in optical network architectures and operations. This enables big data analysis to estimate network states and facilitate adaptive resource provisioning and fault discovery. The data obtained in SDN frameworks encompasses parameters from the physical layer to the network layer, enabling the use of data-driven algorithms like ML for pattern extraction and analysis. However, OPM of the physical layer aspect of WDM systems is especially challenging because a demultiplexing of all channels is not desirable. In this work, the usage of OSAs for the extraction of optical spectra is proposed as an OPM technique. The optical spectrum is then analyzed with the help of ML algorithms and further processed with the same.

In this chapter, first, the basics of ML are described with respect to important terms, nomenclature, and training algorithms. Then, the ML algorithms are roughly divided into two classes, i.e., supervised and unsupervised learning. For each of those classes, the algorithms used in this thesis are briefly introduced and their basics are explained. Furthermore, advanced ML algorithms are explained, which are also used in this thesis. At the end, the performance metrics, which are used to evaluate the ML algorithms, are explained.

3.1 Machine Learning Basics

As pointed out in the introduction to this chapter, an ML algorithm is able to learn from data. Mitchell provides a definition of "learning" which describes the basics of ML [92]:

> "A [ML algorithm] is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Basically, an ML algorithm is set up to solve the task T. For this, it is trained using the features contained in the experience E. Its performance is then evaluated during training, improved with experience E and afterwards in the testing phase with the performance measure P. A typical scenario contains a set of measured values, called **features**, and the observations of the outcome of the task T, called **target**. A training set of data, in which the outcome for the ML algorithm is observed given the **features**, is applied to predict quantitative or categorical outcomes.

ML algorithms are used in this work to solve the following tasks:

- **Classification**, i.e., a task, where the algorithm is asked to specify which of the given categories the input belongs to,
- **regression**, i.e., a prediction of a numerical value given some input,
- **anomaly detection**, i.e., the labeling of given inputs as unusual or atypical, and
- **synthesis and sampling**, i.e., the generation of new examples that are similar to those in the input data set.

Machine learning algorithms can be classified into two main categories: unsupervised learning and supervised learning, which are differentiated based on the type of experience E they acquire during the learning process [90].

Unsupervised learning algorithms operate by analyzing datasets and extracting valuable properties from the underlying structure of the data. These algorithms do not rely on labeled examples but rather focus on uncovering patterns, relationships, and intrinsic characteristics of the dataset. They aim to understand the data's inherent organization and discover underlying concepts or clusters. By examining the dataset's distribution, unsupervised learning algorithms can perform tasks such as density estimation, synthesis, denoising, and clustering. This type of learning is particularly useful when dealing with large and unlabeled datasets, as it can reveal valuable insights and enable data-driven

decision-making [89, 90].

On the other hand, supervised learning algorithms work with datasets that provide both input features and their corresponding labels or targets. This labeled information serves as a guide for the algorithm during the learning phase. By training on labeled examples, supervised learning algorithms can learn to recognize patterns, establish relationships between input features and output labels, and generalize from the provided information to make predictions or classify new, unseen instances. The primary objective of supervised learning is to develop models capable of accurately mapping inputs to desired outputs based on the available labeled data [91].

3.1.1 Capacity, Overfitting and Underfitting

A central and critical challenge in ML is achieving generalization. Generalization is the ability of an algorithm to perform well on novel, previously unseen inputs [90]. To achieve this, the process of training a machine learning model involves accessing a large, comprehensive training set and iteratively refining the model's performance by minimizing training errors through optimization techniques.

However, ML efforts go beyond optimization. The goal is a low test error, which is a measure of the true predictive power of a model on new data. The generalization error is essentially the expected error value on a novel input, given various possible inputs drawn from the expected distribution. In order to estimate the generalization error, a separate test set is used to measure the performance of the model independent of the training set [90]. This partitioning of the data into training and test sets allows for a fair assessment of the algorithm's ability to generalize.

Under generalization, two central challenges in ML are present: underfitting and overfitting:

- Underfitting occurs when the model struggles to achieve a sufficiently low error on the training set, indicating its inability to effectively capture the underlying patterns in the data [90].
- On the other hand, overfitting poses the opposite problem, where the model's performance becomes overly tuned to the training set, resulting in a significant gap between the training error and the test error. This discrepancy means that the model is unable to generalize well to new inputs. [91].

The performance of a machine learning model depends on its ability to reach a balance between minimizing the training error and reducing the gap between the training and

test errors. The key factor influencing this balance is the capacity of the model, which defines its potential to fit a wide range of functions [90]. Models with low capacity may struggle to fit the training set, leading to underfitting, while models with high capacity may be prone to overfitting by memorising feature combinations of the training set. The capacity of a model can be changed by increasing or reducing the number of input features as well as optimizing its hyperparameters. There are many more methods of changing the capacity, however, this is not the scope of this work and the reader is referred to [89, 90].

3.1.2 Hyperparameters and Optimization Algorithms

ML algorithms have typically different settings to be tuned outside of the training to achieve the best performance possible, the so-called hyperparameters. Those parameters do not modify the model's capacity and are not adapted in the learning process [88, 89]. They encompass parameters such as for example the number of coefficients, and, for the case of neural networks, the learning rate, the activation function, the batch size, and the number of hidden layers. These choices wield substantial influence over a model's performance, guarding against overfitting or underfitting. Hyperparameter tuning can be done via manual exploration or automated techniques like grid search, random search, Bayesian optimization, or evolutionary algorithms, i.e., one algorithms learns the best hyperparameters for the other [90]. The objective of this tuning is to maximize the model's performance metrics on a validation set or through cross-validation.

Optimization algorithms play a crucial role in machine learning by enabling the training and fine-tuning of models. These algorithms are used to find the optimal values of parameters or variables that minimize or maximize an objective function. Their target is for example to optimize loss functions, tuning hyperparameters, optimizing neural networks, and incorporating regularization techniques [88]. They enable the iterative process of finding the best set of parameters or variables that lead to improved model performance.

Especially the updating of weights and biases in neural networks is a field of application for optimization algorithms. In this field the stochastic gradient descent (SGD) optimization algorithm or other algorithms based on the SGD are used. The SGD updates model parameters iteratively by adjusting them in proportion to the negative gradient of the objective function with respect to the parameters. The gradient indicates the direction of the steepest ascent of the function, allowing the algorithm to update parameters or coefficients for convergence towards the optimal solution through iterative steps. A comprehensive overview of gradient descent optimization algorithms is given in [93]. In

this work, however, only the so-called adaptive moment (Adam) optimizer is used. Adam estimation [94] is a method that computes adaptive learning rates for each parameter. Furthermore, it keeps an exponentially decaying average of the past gradients. More information on the Adam optimizer can be found in [93, 94].

3.2 Supervised Learning

As described in Section 3.1, supervised learning algorithms learn from previously labeled datasets. Because the targets of the input data is known during training, this error can be measured and the algorithm can be adjusted accordingly. In the following, the supervised learning algorithms used in this work will be explained briefly.

K-Nearest Neighbors

These kinds of classifiers are firstly mentioned in 1967 [95]. k-nearest-neighbor (k-NN) classifiers are memory-based and require no model to be fit [91]. Given a query point x_0, it finds the k training points $x_{(r)}$ with $r = 1, ..., k$ closest in distance to x_0. Then, the classification is done among the k neighbors regarding the majority of the present classes. If the Euclidean distance is taken as a distance measure, the distance $d_{(i)}$ of the ith sample $x_{(i)}$ to x_0 is defined as

$$d_{(i)} = ||x_{(i)} - x_0||. \tag{3.1}$$

In this case, the features have to be standardized to have a mean value of zero and a variance of 1, since it is possible that they are measured in different units [91]. Besides the simplicity of the k-NN, it has been used for various numbers of classification tasks like handwritten digit classification or electrocardiogram analysis even before high performance hardware for ML was available.

Support Vector Machine

An SVM [91, 96–99] is a powerful supervised machine learning algorithm used for classification and regression tasks. It aims to find an optimal hyperplane that separates data points into different classes, maximizing the margin between the classes [96] as depicted in Fig. 3.1. An SVM is a binary classifier but can be extended to handle multi-class classification problems using techniques like one-vs-one or one-vs-rest [91].

The SVM utilizes a subset of training data points called support vectors. These vectors are the closest to the decision boundary and have the most influence on the classification. The choice of the hyperplane depends on the support vectors. During training, the SVM

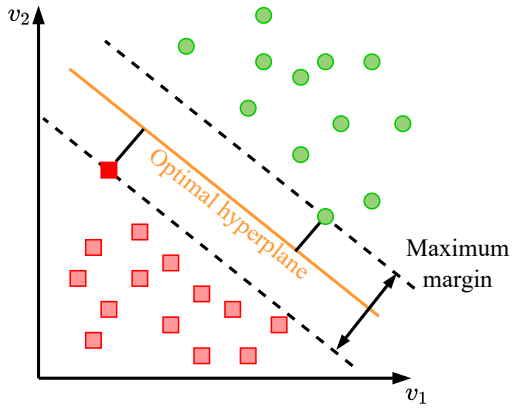


Figure 3.1: Separation of two data classes in the feature space through an optimal hyperplane; solid dot and square are referred to as support vectors; adapted from [88].

tries to find the hyperplane that maximizes the margin between the support vectors of different classes [96]. In addition, SVMs have the ability to handle both linearly separable and nonlinearly separable data [34]. This is achieved by using different kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid. These kernel functions transform the input data into a higher-dimensional space where the data becomes separable. Due to the application of kernels, SVMs are particularly effective in high-dimensional spaces and can handle datasets with a small number of samples but a large number of features [88].

The regularization parameter C in an SVM plays a crucial role in finding the balance between achieving a larger margin and allowing misclassified points and is a hyperparameter of the SVM which has to be optimized. A smaller C value allows more misclassifications but results in a larger margin, while a larger C value leads to a smaller margin but fewer misclassifications [91].

Decision Tree

Decision trees have become increasingly popular in machine learning as a versatile algorithm for classification and regression tasks. They have a distinct structure resembling a flowchart, with internal nodes representing features, branches representing decisions based on those features, and leaf nodes representing outcomes or predicted values. The key advantage of decision trees is that they are easy to interpret and to visualize, making it easier to understand the underlying decision process [100]. Their interpretability and simplicity make decision trees a simple solution to complex problems, especially in problems where the importance of different features is of interest.

Each decision tree begins with a root node, which determines the optimal feature to split the data. The process of constructing a decision tree involves recursively splitting the data using different features until a stopping criterion is met. This criterion may include reaching a maximum depth, maintaining a minimum number of samples in a leaf, or observing no further improvement in impurity or information gain [91, 100–102].

Decision trees and their evolutions offer flexibility in handling both categorical and numerical features. When it comes to numerical features, the data is typically divided into distinct intervals based on threshold values. However, decision trees are prone to overfitting especially if the depth of the tree is chosen to be large [102]. To address this issue, techniques such as pruning, enforcing a minimum number of samples per leaf, or employing ensemble methods like Random Forests can be employed. Despite their advantages, decision trees do have limitations. They may exhibit a bias towards features with a large number of levels and encounter challenges with class imbalance issues [100, 102]. To enhance their performance and address these limitations, ensemble methods such as Random Forests and gradient boosting are commonly employed [103].

Ensemble Learning

Ensemble learning combines multiple ML models, known as base learners, to enhance prediction accuracy and generalization [104]. By leveraging diverse algorithms or variations of the same algorithm trained on different data subsets, ensembles capture different aspects of the problem, reducing individual biases [105]. The predictions of the base learners are combined using voting or averaging methods. Popular ensemble methods, like Random Forest, use decision trees to create robust and accurate models. Ensemble learning improves predictive performance, increases resistance to overfitting and noise, but can be computationally expensive. It is a valuable approach when diverse base learners and sufficient training data are available [106].

Artificial Neural Networks

ANNs are, as the name suggests, computational models inspired by the structure and functioning of the biological neural networks in the human brain [107]. Biological brain structures are way more advanced and superior to conventional computers, since they overcome the limitations of traditional computers in the way of processing information. ANNs consist of interconnected nodes, called neurons, which are organized in layers. Those layers include an input, one or more hidden layers, and an output layer. Those neurons, the basic building blocks of neural networks, can be seen as processing units.

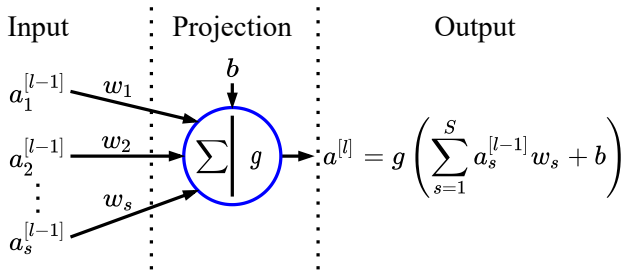


Figure 3.2: Single artificial neuron; ${a_s}^{[l-1]}$: input from the connected layer; $a^{[l]}$: weighted sum of the input signals through activation g with weights w_s and bias b; adapted from [109].

A single neuron is depicted in Fig. 3.2. Each neuron receives weighted information $w_s \cdot {a_s}^{[l-1]}$ from the connected neurons as the input and computes the output $a^{[l]}$ by passing the weighted sum of the input signals through an activation function g [108]. The activation function is used to introduce nonlinearity into the output of the neuron as well as to decide, whether a neuron should be activated or not. This nonlinear transformation enables the neuron to learn and perform more complex tasks, leading to ANNs being able to perform any task [107].

Different types of ANNs exist, each of which are used for specific problem domains. For example, feed-forward neural networks (FF-NNs) process information in a unidirectional manner, while recurrent neural networks (RNNs) incorporate memory to handle sequential or temporal data. Convolutional neural networks (CNNs) as a more complex structure, are used for analyzing visual data.

An FF-NN, also known as multilayer perceptron (MLP), is a popular type of ANNs which is widely used in ML applications. As the name suggests, the data flows unidirectionally through the network, i.e., from the input to the output without a backward connection. A simple structure of an FF-NN is depicted in Fig. 3.3. During training, feed-forward neural networks employ backpropagation, a technique that adjusts the weights associated with the connections between neurons. This adjustment is based on the error between the predicted output and the desired output. By iteratively updating the weights, the network learns to minimize the prediction errors and improve its performance. This optimization is done with gradient descent optimization algorithms such as SGD or Adam [94] as explained in 3.1.2.

In contrast to FF-NNs, RNNs allow backward connections between neurons. The connections of an RNN are thus a directed graph along a sequence of inputs, which allows for temporal dynamic behavior [110, 111]. The backward connections are used as an input for each neuron together with the forward paths, enabling the RNN to work as a memory,

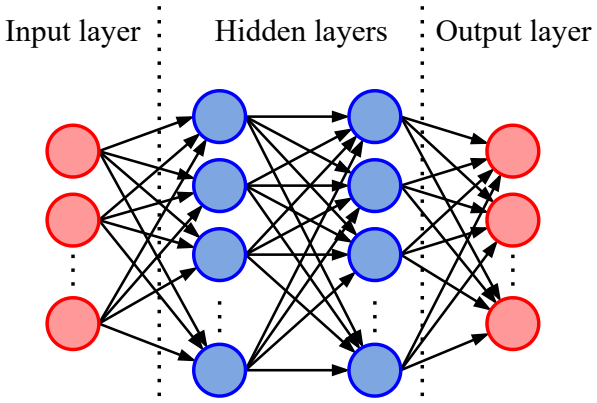


Figure 3.3: Example of an FF-NN with two fully connected hidden layers.

since previous data has an influence on the activation of the neuron. This structure allows the RNNs to store, remember, and process past complex signals for long time periods [111]. RNNs can map an input sequence to the output sequence at the current step and predict the sequence in the next step [111]. More information on the basics of RNNs can be found in [110, 111]. In this work, the two most used types of RNNs, i.e., the long-short term memory (LSTM) and the gated recurrent unit (GRU), are used. Thus they are explained in more detail.

Long-Short Term Memory

LSTM networks were initially proposed by Hochreiter and Schmidhuber [112] in 1997. Since their introduction, LSTMs have become one of the most popular and effective methods for modeling sequential dependencies in input features. The basic structure of an LSTM is depicted in Figure 3.4(a). In comparison to traditional RNNs, each layer of an LSTM is expanded to include memory cells, which are controlled by gates. These gates manage the flow of information and retain information from previous time steps [112].

An LSTM cell consists of input, forget, and output gates, as well as a cell activation component [111]. These gates regulate the information flow between memory cells based on past inputs to the network. The different gates and their weights at time step t are

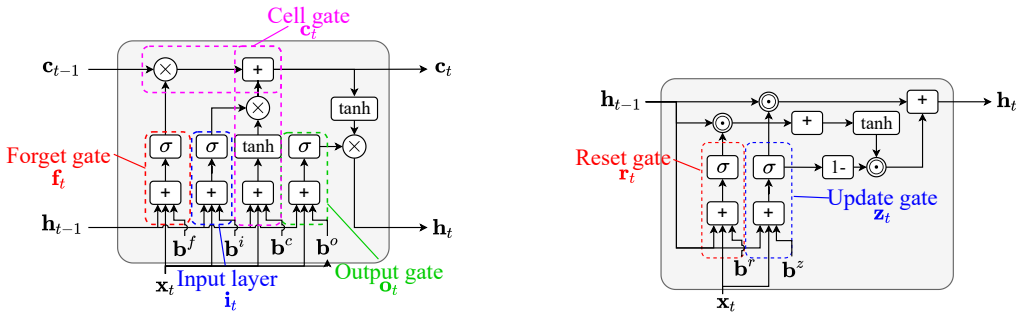


Figure 3.4: (a) Basic structure of an LSTM cell; (b) basic structure of a GRU cell; $\mathbf{x}_t$: input at time t, $\mathbf{h}_{t-1}$: output of the unit (last hidden state) at time $t-1$, $\mathbf{h}_t$: output (new hidden state) at time t. Each gate has a certain bias value $\mathbf{b}$.

defined as follows [112]:

$$\mathbf{i}_t = \sigma(\mathbf{W}_I^i \mathbf{x}_t + \mathbf{W}_H^i \mathbf{h}_{t-1} + \mathbf{W}_A^i \mathbf{c}_{t-1} + \mathbf{b}^i), \tag{3.2}$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_I^f \mathbf{x}_t + \mathbf{W}_H^f \mathbf{h}_{t-1} + \mathbf{W}_A^f \mathbf{c}_{t-1} + \mathbf{b}^f), \tag{3.3}$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_I^o \mathbf{x}_t + \mathbf{W}_H^o \mathbf{h}_{t-1} + \mathbf{W}_A^o \mathbf{c}_{t-1} + \mathbf{b}^o), \tag{3.4}$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_I^c \mathbf{x}_t + \mathbf{W}_H^c \mathbf{h}_{t-1} + \mathbf{b}^c), \tag{3.5}$$

$$\mathbf{c}_t = \mathbf{i}_t \tilde{\mathbf{c}}_t + \mathbf{f}_t \mathbf{c}_{t-1}, \tag{3.6}$$

$$\mathbf{h}_t = \mathbf{o}_t \tanh(\mathbf{c}_t). \tag{3.7}$$

The weight matrix from the input layer to the corresponding gates (i: input layer, f: forget gate, c: cell gate, o: output gate) is denoted as $\mathbf{W}_I$. The weight matrix from the hidden state to the corresponding gates is represented by $\mathbf{W}_H$. $\mathbf{W}_A$ signifies the weight matrix from the cell activation to the corresponding gates. $\mathbf{x}$ represents the input vector, $\mathbf{h}$ denotes the output vector, $\tilde{\mathbf{c}}$ represents the candidate hidden state, and $\mathbf{b}$ is the bias associated with the corresponding gates. In the equations above, $\sigma(\cdot)$ denotes the activation function of the gates, and $\tanh(\cdot)$ represents the output activation function.

Gated Recurrent Unit

While LSTMs takle the problem of vanishing or exploding gradients, they require a high amount of memory due to multiple memory cells in the architecture. Similar to the LSTM unit, the GRU, which was first introduced by Cho et al. in 2014 [113], has gating units that modulate the flow of information within the unit, but without having separate memory cells.

The basic structure of a GRU cell is depicted in Fig. 3.4(b). Unlike the LSTM, the GRU exposes the entire state at each time step by forming a linear sum between the existing state and the newly calculated state [111, 114, 115]. In a similar manner to the LSTM gates' equations, the updated GRU cells at each time step t are given as:

$$\mathbf{z}_t = \sigma(\mathbf{W}^z \mathbf{x}_t + \mathbf{W}^z \mathbf{h}_{t-1} + \mathbf{b}^z), \tag{3.8}$$

$$\mathbf{r}_t = \sigma(\mathbf{W}^r \mathbf{x}_t + \mathbf{W}^r \mathbf{h}_{t-1} + \mathbf{b}^r), \tag{3.9}$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}^h \mathbf{x}_t + \mathbf{W}^h (\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}^h), \tag{3.10}$$

$$\mathbf{h}_t = \mathbf{z}_t \odot \mathbf{h}_{t-1} + (1 - \mathbf{z}_t) \odot \tilde{\mathbf{h}}_t, \tag{3.11}$$

where $\mathbf{z}$ is the the update gate, $\mathbf{r}$ denotes the reset gate, $\mathbf{x}$ represents the input vector, $\mathbf{h}$ is the output vector and $\mathbf{W}$ and $\mathbf{b}$ represent the weight matrix of the corresponding gate and the bias vector of the corresponding gate, respectively [114]. As for the LSTM, $\sigma(\cdot)$ represents the activation function of the gate and $\tanh(\cdot)$ is the output activation function. In addition, the '$\odot$' denotes an element-wise product operation.

In [116], the performance of LSTMs and GRUs are compared. Several similarities and differences are presented, concluding that none of the models is inherently better than the other. Both models perform better than the other only on certain tasks. However, the GRU requires less memory than the LSTM in general.

3.3 Unsupervised Learning

As mentioned in Sec. 3.1, unsupervised learning algorithms operate by analyzing datasets and by extracting valuable properties from the underlying structure of the data. They aim to discover underlying concepts or clusters of the data which makes them suitable for density estimation, denoising, and clustering of data points. In the following the most used unsupervised learning algorithm called k-means will be explained as well as a more advanced algorithm aiming for a more accurate clustering of the data, i.e., density-based spatial clustering of applications with noise (DBSCAN).

k-Means

The k-means clustering algorithm is used for grouping data points into distinct clusters by minimizing the within-cluster sum of squares, i.e., the squared Euclidean distance (see Eq. (3.1)) [91]. By setting the number k of clusters to find in the dataset, the algorithm iteratively assigns data points to a centroid based on the assigned distance metric. The process of assigning data points to clusters and updating the centroids is repeated until

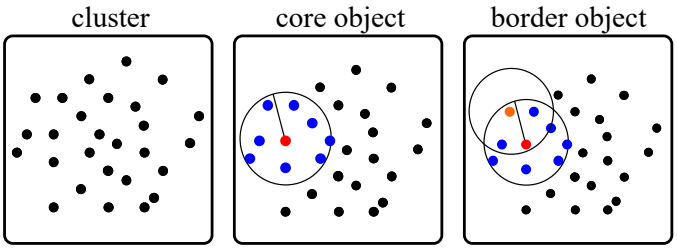


Figure 3.5: Visual example of DBSCAN objects.

the predefined number of clusters is reached [91]. One downside of using k-means is that prior knowledge of the domain is needed so that k can be chosen properly. However, k-means is widely used, especially for customer segmentation and image compression. More information on the algorithm can be found in [91].

DBSCAN

Unlike k-means, DBSCAN does not require specifying the number of clusters in advance of the training for the clustering of data. The algorithm defines clusters as dense regions of data points separated by regions of lower density [117]. By this, DBSCAN is able to find clusters of similar points from the input data. It assigns different labels to the data which include [118]

- core objects, i.e., data points with a sufficient number of neighboring points with a specified distance,
- border objects, i.e., points with less neighboring points than core objects but within the specified distance, and
- noise points, which are points that have not enough neighboring points and are not within the neighborhood radius of any core point.

A visual example is depicted in Fig. 3.5. Like k-means, the algorithm is of iterative manner. It starts by randomly selecting a data point and expanding the neighborhood radius in order to find all, so-called, reachable points within the specified distance. At the end of the iterations, the algorithm is able to find clusters with different densities and arbitrary shape, while for example k-means is restricted to spherical shapes [91, 118]. By setting the distance for neighboring points and the minimum number of neighboring points of a core object, DBSCAN is tuned for specific tasks. However, DBSCAN is complex on a large number of input data because of its iterative behavior. Furthermore, it is hard to be tuned properly, because, like k-means, a rudimentary knowledge of the input data is required for proper setting of the distance and the minimum number of points.

3.4 Advanced Machine Learning Algorithms

The supervised and unsupervised learning algorithms from the previous sections are limited when it comes to certain tasks, like dimensionality reduction, anomaly detection, and generation of new samples. In recent years, popular advanced ML algorithms have emerged to conquer the mentioned tasks. The presented algorithms in the following are basically architectures containing ANNs, which are then trained and connected in special ways.

3.4.1 Variational Autoencoder

Autoencoders (AEs) are a type of ANN architecture comprising an encoder $E : X \to Z$ and a decoder network $D : Z \to X$, where $Z \in \mathbb{R}^n$, $n \in \mathbb{N}^+$. These are jointly trained to reconstruct unlabeled data $X \in \mathbb{R}^m$. By selecting a lower dimension $n < m$, represented by the multivariate latent vector $\mathbf{z} = E(\mathbf{x})$ with $\mathbf{x} \in X$, the encoder E learns to encode the input data X in a way that enables reconstruction with the decoder $\hat{\mathbf{x}} = D(z)$, where $z \in Z$. Trained AEs enable applications such as dimensionality reduction by using z instead of x in the following algorithms, denoising by utilizing $\hat{\mathbf{x}}$, and anomaly detection by measuring the discrepancy between $\mathbf{x}$ and $\hat{\mathbf{x}}$.

Kingma and Welling [119] introduced variational autoencoders (VAEs) as an extension of AEs. VAEs share a similar architecture with AEs, but have a key difference in their objective. Instead of directly reconstructing the data, VAEs aim to learn the distribution of the data using a prior distribution p_θ parameterized by θ. The latent vector $\mathbf{z}$ is typically assumed to follow a multivariate Gaussian distribution. This Gaussian assumption enables additional capabilities beyond conventional AEs, such as data generation by decoding samples drawn from a Gaussian distribution using the probabilistic decoder. Typically, VAEs demonstrate better generalization due to the fact, that encoded samples are not reconstructed directly, but parameterize the distribution from which the input of the decoder is drawn. As the true posterior $p_\theta(\mathbf{z}|\mathbf{x})$ is often intractable, it is approximated by a function $q_\phi(\mathbf{z}|\mathbf{x}) \approx p_\theta(\mathbf{z}|\mathbf{x})$ parameterized by the probabilistic encoder $E_\phi(\mathbf{x})$. The multivariate latent vector is calculated as follows:

$$\mathbf{z} = \mu + \sigma \odot \epsilon \tag{3.12}$$

where μ represents the mean value, σ is the standard deviation, and ϵ is a sample drawn from a normal distribution with a mean value of 0 and a standard deviation of 1. The basic structure of a VAE is illustrated in Fig. 3.6. During training, the goal is to find optimal parameters θ and ϕ that minimize the reconstruction error at the decoder output

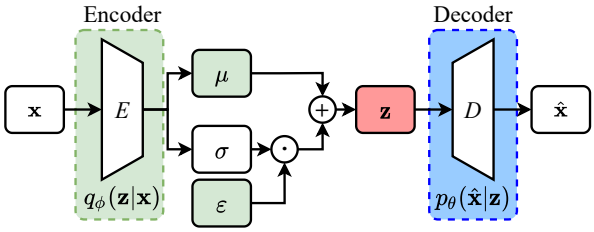


Figure 3.6: Basic structure of a VAE; $\mathbf{x}$: input vector, $\mathbf{z}$: multivariate latent vector, $\hat{\mathbf{x}}$: reconstructed input vector; μ: mean value, σ: standard deviation, and ε: sample drawn from multivariate normal distribution.

while preserving the Gaussian probability distribution in the latent space. In general, by utilizing a well-trained AE's encoder, the input dimension m can be effectively reduced to n ($n < m$) with minimal information loss. Thus, the latent space represents a set of meaningful features for describing the input data, which reduces the need for manual feature selection in other machine learning algorithms. AEs are called semi-supervised learning algorithms, especially for anomaly detection, because they can be trained on normal data and can reliably react to anomaly data due to the reconstructive training.

3.4.2 Generative Adversarial Network

Generative adversarial networks (GANs) have emerged as a powerful tool for generating realistic and high-quality synthetic data in various domains, including computer vision and natural language processing. GANs, introduced by Goodfellow et al. in 2014 [120], consist of two neural networks: a generator and a discriminator, which compete against each other in a "zero sum game" [120]. A "zero-sum game" is a situation in game theory where one participant's gain or loss is exactly balanced by the losses or gains of other participants. In other words, the total gains and losses within the system sum to zero [90].

The basic structure of a GAN is depicted in Fig. 3.7. The generator network aims to produce synthetic data samples that resemble the real data distribution, while the discriminator network strives to differentiate between the real and fake samples. During training, the generator produces synthetic samples, and the discriminator provides feedback by labeling each sample. The two networks are trained simultaneously in an adversarial manner, with the goal of improving the generator in its ability to deceive the discriminator, and the discriminator being trained to accurately discriminate between real and fake samples.

By leveraging this adversarial training process, GANs have been successful in producing

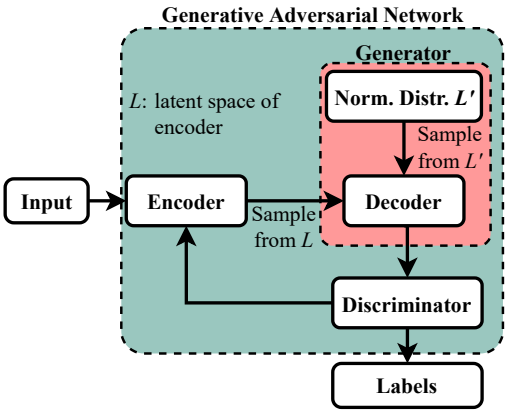


Figure 3.7: Basic structure of a GAN; L: latent space.

images indistinguishable from real photos by humans [121], synthesizing natural language and generating music. However, GANs are prone to training instability. The hyperparmeters of the networks have to be chosen carefully since GANs suffer from convergence oscillations and vanishing gradients [122]. Achieving a balance between the generator and discriminator networks during training can be challenging, leading to suboptimal results or failed convergence.

3.5 Performance Metrics

Evaluating the performance of a trained ML algorithm requires metrics that are suitable for the task, meaning either regression or classification metrics. There are many different performance metrics. The following section focuses on the metrics, which are used in this work for evaluating the performance of the proposed ML frameworks.

3.5.1 Regression Metrics

Regression models aim to approximate a mapping function from input variables to a continuous output variable. Thus, the output variable is a scalar, which can be, for example, amounts or sizes. The results of a regression task can be evaluated regarding their deviation to the expected or real values.

A common metric for evaluating the absolute deviation of a regression task is the mean absolute error (MAE). The MAE is calculated as

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}, \tag{3.13}$$

where n is the number of samples, y_i is the prediction, x_i is the true value, and $|y_i - x_i|$,

thus, is the absolute error of the i^{th} sample.

Another common metric for the quality of an estimator is the MSE. It is derived from the square of the Euclidean distance. If y_i gives the predictions of the estimator, then the MSE is given by

$$\text{MSE} = \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}. \tag{3.14}$$

It can be seen that this error measure decreases to 0 when $y_i = x_i$. Also, the error increases when the Euclidean distance $||.||$ between the predictions and the target increases as shown by [90]

$$\text{MSE} = \frac{1}{n}||y_i - x_i||^2. \tag{3.15}$$

While the MAE is a metric in the same scale as the input data, the R^2-score, also known as the coefficient of determination, is a more intuitive performance metric with its output being between 0 and 1, where a score of 1 is indicating a perfect prediction [123]. The R^2-score is defined as

$$R^2(y, x) = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \overline{y})^2}, \tag{3.16}$$

where $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the mean over all predictions. It provides an indication of how well unseen samples are likely to be predicted accurately by the model.

3.5.2 Classification Metrics

For a classification task, the classifier distinguishes between two classes, i.e., a binary classification. The result can be either "true" or "false".
Decisions from the classifier lead to

- true positives (TPs), i.e., the number of samples classified as "true" while being labeled "true",
- true negatives (TNs), i.e., the number of samples classified as "false" while being labeled "false",
- false positives (FPs), i.e., the number of samples classified as "true" while being labeled "false", and
- false negatives (FNs), i.e., the number of samples classified as "false" while being labeled "true".

The overall effectiveness of a classifier can be evaluated using different metrics arising from the numbers of TP, TN, FP and FN [124]:

- Accuracy is the percentage of correct predictions:

$$\mathrm{A} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}}. \tag{3.17}$$

- Precision is the percentage of positive predictions that are actually positive:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{3.18}$$

- Recall is the percentage of actual positives that are predicted positive:

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{3.19}$$

- F1-score is the harmonic mean of the precision and recall defined as:

$$\mathrm{F1} = \frac{2\mathrm{TP} + \mathrm{TN}}{2\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}. \tag{3.20}$$

Especially the F1-score is of high interest, because it provides one concise metric that summarizes the model's performance, even for multi-class scenarios. Furthermore, in situations where the classes are imbalanced (i.e., one class significantly outnumbers the other), which might be the case for soft-failure management tasks, accuracy alone might not be a suitable metric. The F1-score is robust in such cases because it accounts for both false positives and false negatives through incorporating recall and precision, making it suitable for evaluating model performance in imbalanced classification tasks. Class imbalance is mostly present in failure datasets, since more normal data can be retrieved than faulty data.

3.6 Complexity Analysis

The complexity of ML algorithms is crucial for their implementation in the field. The higher the complexity, the more operations of the processing unit have to be performed in order to use the algorithm. This means a higher power draw or a high performance processing unit has to be used.

The computational efficiency of an algorithm is typically measured by considering the number of basic operations it executes in relation to the length of its input [125]. This measurement is captured by a function, denoted as T, which maps natural numbers $\mathbb{N}$ to

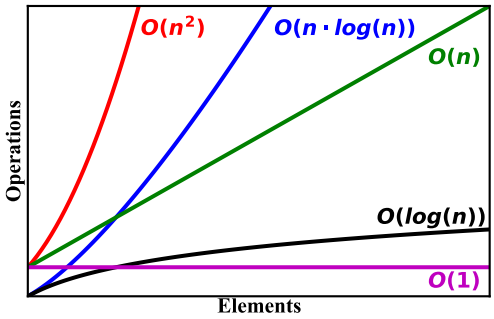


Figure 3.8: Schematic view of number of operations for different big-O complexity orders.

themselves. Specifically, $T(n)$ represents the maximum number of basic operations that the algorithm performs when given inputs of length n [126]. However, it is important to note that the definition of a basic operation can significantly influence the efficiency function T. To overcome this dependency, the so-called "big-O" notation is used in the literature. In [126], the big-O notation is defined as follows: If f and g are two functions from $\mathbb{N}$ to $\mathbb{N}$, then $f = O(g)$ if there exists a constant c such that $f(n) \leq \mathrm{c} \cdot g(n)$ for every sufficiently large n. Often $f(n) = O(g(n))$ is used as the nomenclature for this relationship. Typical big-O functions and their complexity are depicted in Fig. 3.8. It has to be noted, that for ML, the time complexity regarding training and prediction is relevant as well as the auxiliary space, i.e., the temporary space used by the algorithm during the runtime [127]. With this in mind, Table 3.1 is derived for n being the number of training samples or rows in a dataset, m is the number of features or columns in a dataset, i denotes the number of iterations, n_{sv} is the number of support vectors, d the depth and p the number of nodes in a tree. Regarding clustering algorithms k represents the number of clusters. In FF-NNs the complexity is hard to determine, since it is depending on the overall structure, however, there are some complexities regarding the number of epochs e and the average number of weights per neuron w which are included in the table.

Table 3.1: Big-O computational and space complexities of ML algorithms (adapted from [127]).

Algorithm	Training	Prediction	Auxiliary space
Decision Tree	$n\log(n)d$	d	p
Random Forest	$n\log(n)dn_{\mathrm{tree}}$	dn_{tree}	pn_{tree}
k-NN	1	nm	nm
SVM	$n^2m + n^3$	mn_{sv}	n_{sv}
FF-NN	$enmw$	m	w
k-means	$nmki$	mk	$nm + km$
DBSCAN	$nlog(n)$	mn	n

Chapter 4

Quality of Transmission Estimation

Utilizing ML techniques for QoT estimation is highly valued within the scientific community. The applications for QoT estimation are diverse, such as optimizing links and networks by reducing margins and allocating resources based on knowledge of link behavior in a network.

ML techniques for QoT estimation have been shown to be effective in so-called "brownfield" networks, where data can be collected directly from the field to train models. Typical scenarios for "brownfield" networks include capacity upgrades by adapting modulation formats or symbol rates [4]. In such networks the retrieval of monitoring data, which enables accurate ML models, is difficult, especially for dense WDM systems. Hence, the usage of OSAs as OPM devices for the use-case of QoT estimation by using the optical spectrum is presented in this chapter.

In contrast to "brownfield" networks, field data is not available in so-called "greenfield" networks. Hence, ML models must be trained on simulated or experimental data. The adaptation from simulated trained data to real-world scenarios is difficult, which means these ML models need to show a high generalization capability. Using experimental data for ML model training can lead to better performance in the field, however, those models are also needed to generalize well. The presented ML model in this chapter can also be applied in a "greenfield" scenario, since simulation data is easily available and ML shows computational speed advantages over other approaches like the GN-model.

First, the benefit of using spectral input features for QoT estimation in "brownfield" networks is shown using simulations. Afterwards, the ML-based QoT estimation framework is experimentally validated and compared to other ML algorithms in order to investigate their generalization capabilities by training on simulation data and testing on experimental data. The robustness of the framework is then been looked at by stochastic variation of the input features. The chapter is finished with an investigation of impact of the OSA resolution on the QoT estimation accuracy.

Table 4.1: Varied uncertain simulation parameters.

Parameter	Symbol	Mean Value	Standard Deviation
Span lengths	L_S	80 km	5 km
EDFA output power	P_E	P_L	0.5 dBm
EDFA noise figure	NF	5 dB	0.5 dB
Attenuation	α	0.2 dB/km	0.02 dB/km
Dispersion	D	17 ps/(nm·km)	0.2 ps/(nm·km)
Nonlinear coefficient	γ	1.295 $(\text{W·km})^{-1}$	0.05 $(\text{W·km})^{-1}$

4.1 Exact Component Parameter Agnostic Scenarios

Multi-vendor optical networks make accurate QoT estimation a non-trivial task, as exact equipment parameters are often considered confidential or are not exactly known. As a result, such a multi-vendor network can be considered a so-called "exact component parameter agnostic" network scenario. In addition to the unknown component parameters, parameter uncertainties and fiber nonlinearities further increase the complexity of the QoT estimation task [5]. For such scenarios, analytical solutions for the QoT estimation are more difficult [26]. The nonlinearities distort the channels and lead to inter-symbol and inter-channel interference. The signal quality, as represented by different QoT metrics like SNR, BER or GOSNR, depends not only on the linear ASE noise from the EDFAs in the network, but also on the signal power, the power of the individual channels, and the channel spacing. Uncertain parameters in a deployed network ("brownfield") include fiber specifications, EDFA gain and noise figure, and transponder penalties [4, 20]. To simulate those uncertainties, the parameters are calculated using a heuristic approach with a certain mean and standard deviation based on realistic assumptions and margins as shown in Tab. 4.1.

4.2 Framework Structure

For precise QoT estimation, choosing an appropriate ML algorithm is crucial, particularly as optical networks become increasingly intricate. The primary challenge arises from the varied nature of data obtained from links. Every link can be depicted as a sequence of intermediate nodes, with each node signifying a spectrum captured through OSAs. The analysis will consider the length of spans between the current and previous nodes, total power within the spectrum, and channel powers. It is important to note that the number of intermediate nodes present in each link has an impact on the size of input features.

Consequently, conventional ML algorithms designed for a fixed number of features may not be practical.

Because each link is interpreted as a series of values, this series can be processed by a RNN, such as an LSTM or GRU, which solves the problem of varying input sizes of the input features. By leveraging the inherent sequential dependencies within the data, RNNs can effectively model the varying sizes of input features.

The QoT estimation framework is depicted in Fig. 4.1. It can be seen that it consists of LSTM and FF-NN layers. Selecting the optimal size for the LSTM layers is crucial for achieving both accuracy and computational efficiency. Our previous investigations have highlighted the need to find the right balance between model size and accuracy, finding that increasing the size of the layers does not guarantee a significant improvement in performance.

During this work, an LSTM-based solution was compared to a GRU-based approach. The results showed that the LSTM exhibited superior estimation accuracy when compared to the GRU approach. After a grid search over the layer sizes, the best configuration for the QoT estimation framework comprised LSTM recurrent layers with sizes of 24 and 12.

Additionally, to facilitate the learning process, a VAE is employed to extract a compact and informative representation of the data. The VAE is composed out of an encoder with an input layer with a size equal to the 10,000 points of the spectrum, followed by a fully-connected layer with 100 neurons, a batch normalization layer and another fully-connected layer with an output size of 12, i.e., the latent space size. The decoder is

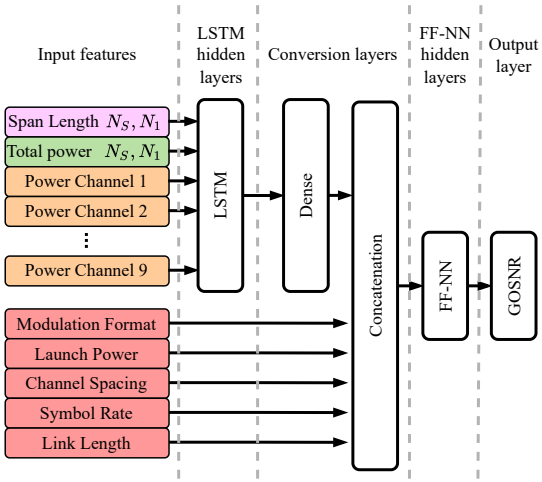


Figure 4.1: QoT estimation framework with the used layer types of LSTM and FF-NN layers [21].

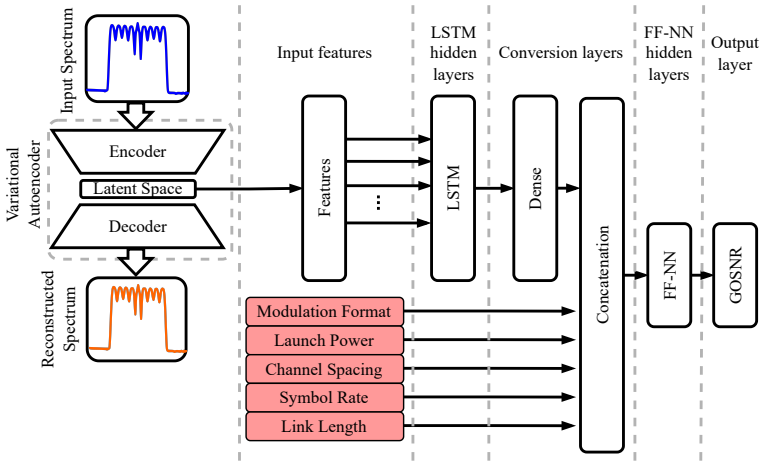


Figure 4.2: QoT estimation framework extension with the spectrum pre-processed by a VAE [21].

built to reverse the functions of the encoder.

The extended framework is depicted in Fig. 4.2. It can be seen that the latent space of the VAE replaces the spectral input features from the framework before. This means that each spectrum is now represented by the latent space.

4.3 Simulative Investigation

The following section covers the underlying simulation setup and the approach of varying parameters per run to simulate agnostic network scenarios according to Tab. 4.1. Afterwards, the developed QoT estimation framework will be compared to other ML approaches to show the benefits of using spectral information as input features. Different ML algorithms for QoT estimation are evaluated for their performance on the COST266 European network topology [128] which is shown in Fig. 4.3 in order to evaluate their application in a real network topology. The network is simulated using the varying component parameters from Tab. 4.1, for the different configurations from Tab. 4.2. The presented content is based on our research published in [20–22].

4.3.1 Simulation Setup

The simulations are done using the MATLAB-based simulation tool "Move-It" [129]. A high-level model of the simulation setup is depicted in Fig. 4.4. The setup is used for generating data for ML algorithm training in order to estimate the QoT and is built to mirror the experimental setup from Section 4.4.1.

Up to 9 channels are transmitted over a coherent DP-WDM link with fixed channel

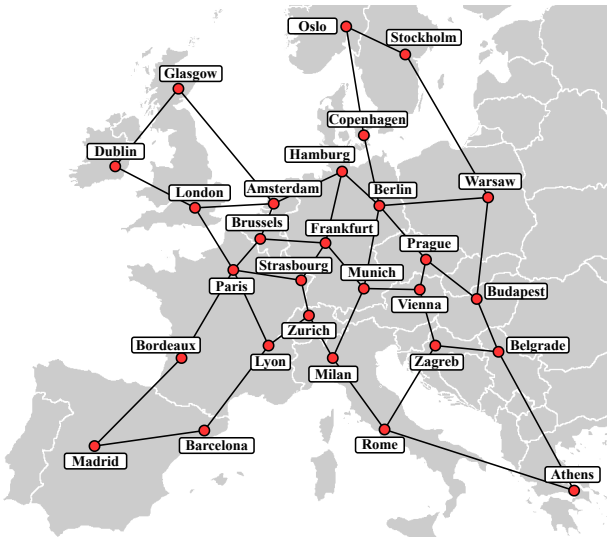


Figure 4.3: COST266 European network topology; adapted from [128].

spacing and equal launch powers per channel. The different links with up to 15 spans are analyzed for the configurations of parameters summarized in Table 4.2. As mentioned before, uncertainties of transmission parameters are for example considered in the span lengths by randomly choosing a length with a mean of 80 km and a standard deviation σ of 5 km. The varied, uncertain parameters are summarized in Table 4.1. Therefore, the parameters are different for each span according to the random distribution.

The nonlinearities for the propagation of the signal through the fiber are calculated using the SSFM with a randomly chosen number of wave plates for the PMD calculation ranging between 50 and 200 per span and a PMD coefficient of 0.03 ps/km$^{1/2}$. The maximum nonlinear rotation angle in the nonlinear step is $\varphi_{\mathrm{rot,max}} = 0.05°$ and the step-size of the SSFM is chosen accordingly. The usage of the SSFM ensures exact modeling of the transmission and gives the opportunity to extract the spectrum which would not be

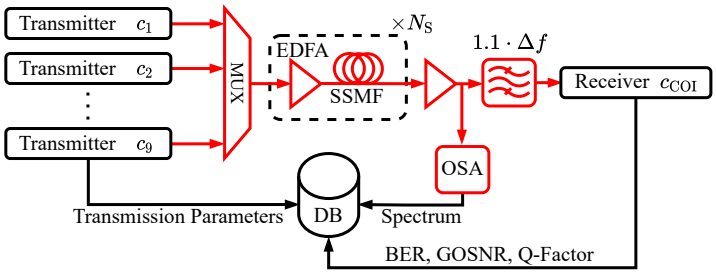


Figure 4.4: Simulation setup with a central database (DB) containing the obtained feature vectors; transmission parameters include modulation format, launch power, channel spacing, symbol rate, total link length, lengths between start, intermediate, and end node [20, 21].

Table 4.2: Simulation parameters.

Parameter	Symbol	Value
Modulation format	MF	DP-QPSK, DP-8-QAM, DP-16-QAM
Symbol rate	b	32, 64, 69 Gbaud
Channel spacing	Δf	37.5, 100 GHz
Launch power	P_{L}	-3 to 3 dBm
Center wavelength	λ_{c}	1550 nm
Number of spans	N_{S}	1 to 15
OSA resolution	R_{OSA}	13 pm
Channel assignment		Neighboring pairs to COI
SSFM waveplates		50 to 200 (random)
SSFM max. nonlinear rotation angle	$\varphi_{\mathrm{rot,max}}$	0.05°

possible with other (faster) simulation methods.

To simulate more complex transmission scenarios, the number of spans is increased with every iteration step of the simulation, i.e., one span is added for every step to a total of 15 spans. Every link can be interpreted in a different way: For example, if the link contains 4 spans, we represent a set of links with 0, 1, 2, and 3 intermediate nodes and their various possible distance variations for the distances to and from the intermediate node. This means, that there are $2^{N_{\mathrm{i}}}$ different combinations for each link with N_{i} intermediate nodes considered [20].

At the receiver side, the BER, Q-factor, and GOSNR are calculated and stored in a database. The number of channels transmitted over the fiber varies from 1 to 9, with only neighboring pairs to the center channel are added or dropped. Additionally, the simulation considers different scenarios for adding and dropping channels at intermediate nodes. Up to four neighboring pairs are removed from all intermediate nodes in the centered channel configuration. However, in the add scenario, these channels are added to configurations that have free slots for the channels. Including only neighboring pairs ensures worst-case scenarios while minimizing simulation effort. Additionally, a maximum of $N_{\mathrm{i}} = 5$ intermediate nodes are assumed to be present in a link.

4.3.2 Simulation Results

The simulation setup from the previous section is used to obtain a dataset based on the interpreted links. It is utilized to train the QoT estimation frameworks depicted in

Table 4.3: QoT estimation results for different ML algorithms.

ML Algorithm	R^2-score	MAE
Regression Tree	0.814	1.93 dB
Support Vector Regressor	0.881	0.76 dB
FF-NN	0.879	0.81 dB
LSTM-FF-NN-hybrid	0.959	0.18 dB
VAE-LSTM-FF-NN-hybrid	0.996	0.16 dB

Fig. 4.1 and Fig. 4.2. To ensure the effectiveness of the training, the dataset is split into 60% for training, 20% for validation, and 20% for testing purposes. The models undergo training for 800 epochs utilizing the Adam optimizer [94], which is a widely-used stochastic gradient descent algorithm for optimization.

In the following, various ML algorithms are compared for their QoT estimation performance on the COST266 European network topology with the specified 378 transmission links to show their performance in an optical network with agnostic component parameters. Since conventional ML algorithms cannot handle series of values, those are only using the transmission-based features (red inputs in Fig. 4.1). The compared algorithms include a regression tree, a support vector regressor, which is an SVM designed for regression tasks, the QoT estimation framework without the LSTM branch, the QoT estimation framework itself, and the extension of the framework with the pre-processed spectrum from the VAE. The overall performance is summarized in Tab. 4.3.

Overall the estimation performance is high with R^2-scores above 0.8 across all investigated ML algorithms. The regression tree shows the lowest performance followed by the FF-NN using only the transmission-based features. The support vector regressor achieves slightly better performance. Including the spectral features to the inputs for the developed framework, the estimation gets more accurate shown by R^2-scores above 0.95 and low MAEs under 0.2 dB.

For regression tasks it is common to use so called "actual versus predicted" plots as visualization. For this purpose, the actual values of the QoT metric are plotted against the values predicted by the ML algorithm. A baseline, which represents the actual values, serves as a reference. To show the impact of including spectral features in QoT estimation tasks, for each neural network-based estimator, i.e., the FF-NN, LSTM-FF-NN-hybrid, and the VAE-LSTM-FF-NN-hybrid, such a plot is generated and depicted in Fig. 4.5. The predictions are represented by the blue crosses, whereas the actual values are depicted by the red baseline. Thus, the nearer the prediction is to the baseline, the more accurate

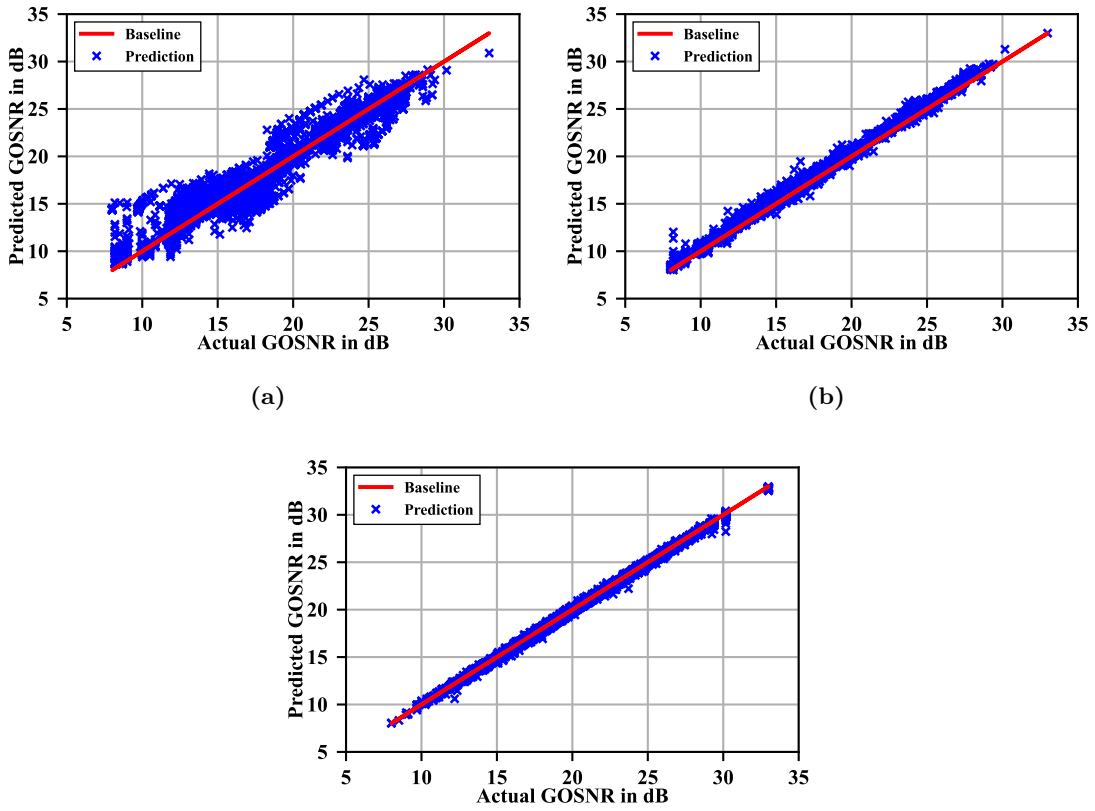


Figure 4.5: Actual versus predicted GOSNR for the proposed framework with (a) no spectral features [20], (b) with spectral features [20], (c) and with spectral features extracted by a VAE [22].

the estimation.

It can be seen that the usage of no spectral features (Fig. 4.5a) leads to a wide spread of the predictions. This mirrors the R^2-score of 0.879 as well as the MAE of 0.81 dB. If the spectral features are included, the number of outliers is reduced and the predictions move near the baseline. The LSTM-FF-NN-hybrid shows an R^2-score of 0.959 and a low MAE of 0.18 dB (Fig. 4.5b). When the VAE is used to extract valuable features from the spectrum, the estimation errors can be even more reduced to an MAE of 0.16 dB and an R^2-score of 0.996, which is also represented by the low deviation of the predictions from the baseline (Fig. 4.5c).

The usage of spectral features increases the QoT estimation performance by up to 0.65 dB to a mean deviation of only 0.16 dB. Such a low deviation is well suited for margin optimization in "brownfield" networks.

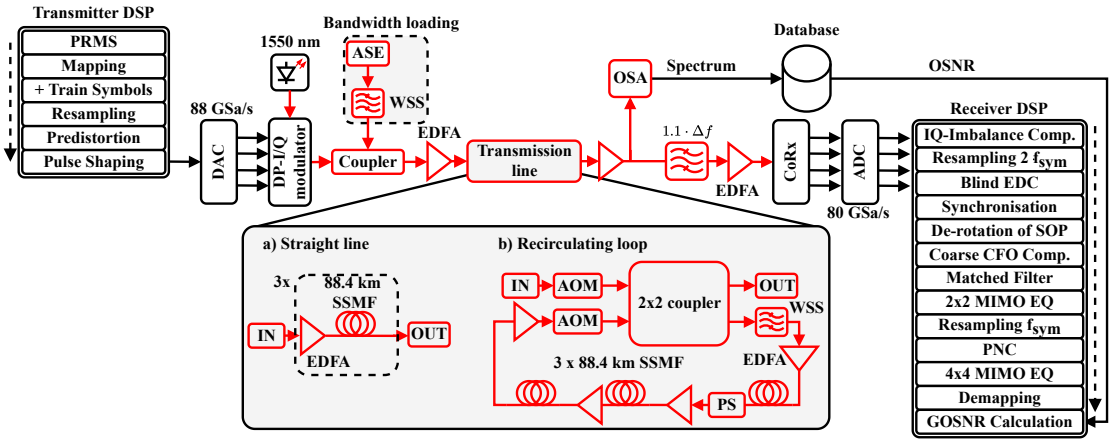


Figure 4.6: Experimental transmission setup with a) a straight line and b) a recirculating loop. PRMS: pseudo-random multilevel sequence, DAC: digital-to-analogue converter, ASE: amplified-spontaneous emission, WSS: wavelength selective switch, EDFA: Erbium-doped fiber amplifier, PS: polarization scrambler, EDC: electrical dispersion compensation, SOP: state of polarization, CFO: carrier frequency offset, PNC: phase-noise compensation [23].

4.4 Experimental Validation

Simulations can often explore a wider variety of complex network structures compared to experiments and can also assist in selecting the most appropriate ML algorithms and data sets. Nevertheless, experimental data is critical for authenticating ML algorithms and their usage in practical situations to guarantee seamless function within deployed networks. Experimental data can be used to approximate real-world conditions in optical transmission and to potentially train ML algorithms.

This section aims to validate the proposed LSTM-based QoT estimator experimentally and to show its high generalization capability by training it on simulation data and testing it on experimental data. This enables the algorithm to be trained before the deployment in a real network. By exploiting the high generalization capability of the ML algorithm, the QoT estimation gets more accurate and faster in "brownfield" networks and component parameter agnostic networks compared to other QoT estimation approaches like the GN-model. Also, the high generalization capability enables the usage of the framework in a "greenfield" network where not monitoring data from the field is available.

To generate experimental data, the experimental setup with the recirculating loop explained in Section 4.4.1 is used. The experimental investigations presented in this section have been published in [21, 23, 77, 130].

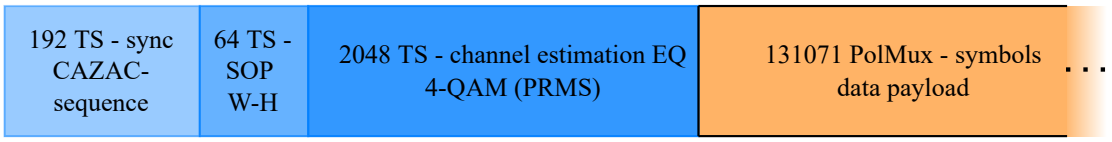


Figure 4.7: Dataframe structure for the DSP; TS: training symbols; CAZAC: constant amplitude zero auto-correlation; SOP: state of polarization; W-H: Walsh-Hadamard matrix; sync: frame synchronization; PRMS: pseudo-random multilevel sequence; EQ: equalizer; adapted from [33].

4.4.1 Experimental Setup

The experimental setup of the used coherent DP-WDM system is depicted as a high-level black-box model in Fig. 4.6. The primary purpose of this setup is to generate experimental data for this work and to evaluate the possibilities of using ML algorithms in optical communication systems.

To create the channel of interest (COI), a pseudo-random multilevel sequence (PRMS) with a length of $2^{17} - 1$ is generated. This sequence is then mapped to DP-QPSK, DP-8-QAM, or DP-16-QAM symbols, and training symbols are added. The dataframe structure can be seen in Fig. 4.7. In total, 2304 training symbols are sent before the data payload. The preamble consists of a CAZAC sequence with 192 symbols for signal synchronization purposes, 64 symbols for estimating the Walsh-Hadamard matrix for a SOP de-rotation [33] and 2048 4-QAM symbols for channel estimation for equalization.

Before transmission, pre-distortion techniques are applied to account for the characteristics of electrical amplifiers and the AWG. The signal is further shaped using a root-raised cosine filter with a roll-off factor α_{RRC} of 0.2. The digital signal is converted to analog using an AWG operating at 88 GSa/s, with an effective number of bits (ENOB) of 5.5. The COI is generated by an external laser operating at 1550.004 nm, coupled with a DP-IQ modulator which is driven by the DAC via four driver amplifiers. In addition to the COI, other WDM channels (loaders) are generated using a programmable wavelength-shaping filter (II-VI WS4000A) with an ASE noise source as input. The wavelength-shaping filter ensures alignment of the channels' spacing and equalization of all channels at the output. This technique is called bandwidth loading as described in Section 2.6.1. This offers some advantages over traditional channel generation. It reduces complexity on the transmitter side, requiring only one modulator, one laser, and one DAC. Moreover, the bandwidth loaded signal characteristics closely resemble those of a conventional WDM signal [83].

The COI and loaders are combined using a 3 dB-coupler before undergoing amplification through an EDFA. Depending on the experimental configuration, the amplified signal is either transmitted over three spans of 88.4 km SSMF with an EDFA per span or through

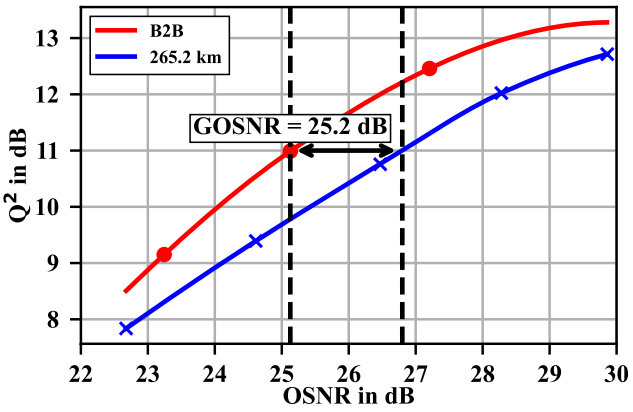


Figure 4.8: Exemplary illustration of the GOSNR back-trace procedure for an experimental measurement of a WDM signal with five, 37.5 GHz spaced channels with 1 dBm launch power per channel and a QPSK modulated COI [77].

a recirculating loop. The overall functionality of a recirculating loop is described in more detail in Section 2.6.2. The loop consists of a waveshaper as a gain-flattening filter, followed by three spans and a polarization scrambler that randomizes the polarization shift effects from the fibers to emulate random fibers. An additional EDFA is used for further signal amplification, ensuring sufficient signal power for downstream processing. After amplification, the COI is filtered and detected using a coherent receiver.

The received signal may suffer from impairments, primarily due to noise and nonlinearities. To compensate for these impairments, various techniques are employed, including IQ-skew and IQ-imbalance compensation from the transmitter and receiver, as well as laser phase noise compensation from both ends. Additionally, chromatic dispersion, PMD, rotation of the SOP, and carrier frequency offset are also compensated by offline receiver DSP. Here, standard algorithms for coherent dual-polarization WDM systems are used [21] which are explained in more detail in Section 2.3.3.

The computation of the GOSNR involves using pre-measured lookup tables that establish the relationship between OSNR and Q-factor for specific configurations. Those lookup tables are obtained using the back-trace method. An example for this technique is depicted in Fig. 4.8. Through the definition of the GOSNR in Eq. (2.52) follows, that the GOSNR is defined as the OSNR in the B2B case which achieves the same BER as over the transmission. Thus, the GOSNR is always smaller than the measured OSNR of the distorted signal.

To obtain the optical spectrum an OSA with an optical resolution of 10 pm is used before

the extraction filter. The spectrum, other performance metrics and the transmission configuration are stored in a database for further use, for example, in the ML algorithms. The explained experimental setup enables to investigate and evaluate the performance of a coherent dual-polarization WDM system under realistic conditions. Through the recirculating loop, arbitrary transmission distances can be emulated in the lab.

The performance of the experimental setup is summarized in Fig. 4.9 as box plots over transmission distances. The minimum and maximum values of the metrics for each modulation format are shown by the whiskers, while the values in the range between the first and third quartiles are shown by the boxes. The median is shown by the horizontal line within the boxes.

Figure 4.9a shows the Q-factor distribution over length in the obtained dataset. The dataset includes different modulation formats, number of channels, and launch powers per channel. As expected, the lower-order modulation formats have higher Q-factors than the higher-order modulation formats. Additionally, the Q-factor decreases for higher lengths. It is important to note that the hard decision FEC limit is always surpassed for QPSK for any length. For 8-QAM, only the median for 265.2 km and 530.4 km are above the HD-FEC limit, while for 16-QAM it is not reached for any length. This is mainly due to non-optimal precompensation of the electrical amplifiers and amplifier noise after the ADC, as well as shot noise and thermal noise of the coherent receiver. However, in an optical long-haul transmission system, strong FEC algorithms are used anyways. Therefore, the soft decision FEC limit (15% overhead) is added to the graph [75]. It can be seen that for 8-QAM, the maximum reach is increased to over 1326.0 km, while for 16-QAM, over 795.6 km transmission reach can be achieved.

The experimentally obtained Q-factors are then used to back-trace to a Q-factor graph of a B2B configuration to obtain the GOSNR as described before. The GOSNR distribution over length is depicted in Fig. 4.9b. In general, it can be observed that the curves characterized by the medians decrease with higher transmission length getting more flat at high distances, as expected. This is because the curves of the Q-factor over the OSNR become steep below a certain low OSNR value. A similar behavior could also be observed at high OSNR values, since the Q-factor-OSNR curve is almost flat there. Furthermore, it is noticeable that the modulation formats QPSK and 16-QAM have smaller boxes than 8-QAM, and generally the medians of the 8-QAM values are below those of 16-QAM. This is due to the fact that 8-QAM is more prone to nonlinearities from multi-channel transmissions, reasoned by the non-equal symbol distances.

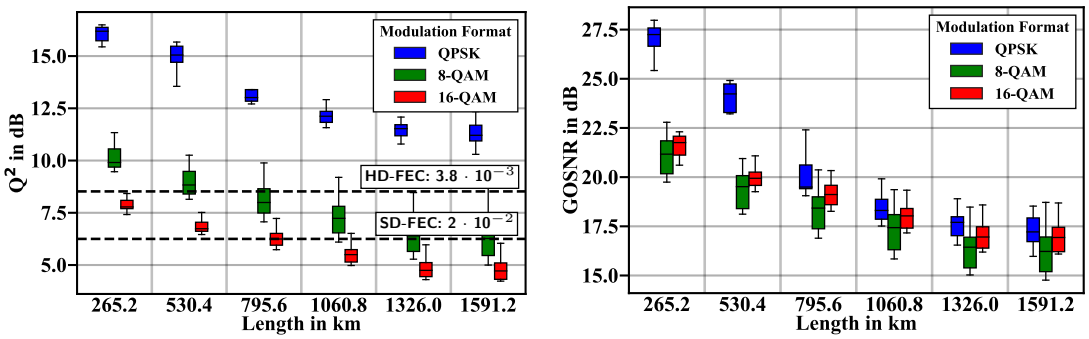


Figure 4.9: Measurements of (a) Q-factor distribution over length; SD-FEC limit for 15% overhead according to OpenROADM specifications [75]; (b) GOSNR distribution over length for QPSK, 8-QAM and 16-QAM; HD-FEC: hard decision forward-error-correction; SD-FEC: soft decision forward-error-correction [21].

For generating an experimental dataset, the configurations from Tab. 4.4 are swept. The obtained dataset is then used for the experimental validation of the proposed QoT estimation approach.

4.4.2 Machine Learning Algorithm Comparison

Precise QoT estimation enables margin reduction in deployed networks which increases the spectral efficiency and the overall throughput of the network. ML algorithms offer the advantage of being fast once trained correctly. However, ML algorithms need a large amount of training data which is not available in the field due to insufficient physical layer monitoring. Due to this, the application of an ML algorithm is desired which achieves a

Table 4.4: Experimental system parameters for QoT estimation.

Parameter	Symbol	Value
Modulation format	MF	DP-QPSK, DP-8-QAM, DP-16-QAM
Symbol rate	b	32 Gbaud
Channel spacing	Δf	37.5 GHz
Number of channels	N_{Ch}	1, 3, 5
Launch power	P_{L}	-3, -2, -1, 0 dBm
Center wavelength	λ_{c}	1550.004 nm
Loop length	L_{Loop}	265.2 km
Loop iterations	N_{Loop}	1 to 6
OSA resolution	R_{OSA}	10, 20, 50, 100, 200, 500 pm
OSA points	N_{OSA}	501

high estimation accuracy on experimental data while only trained on simulation data. Simulation data can be obtained before the deployment of the algorithm even with varying component parameters to ensure proper functionality in such a component parameter agnostic networking scenario.

To investigate the application of different ML algorithms for such tasks, the performance of several non-recurrent ML algorithms is compared based on QoT estimation in the measure of GOSNR for experimental data, while training the algorithms with simulated data. Additionally, a comparison of these algorithms to the developed recurrent structures that include manually and automatically selected features from the spectrum is done.

The comparison of non-spectral algorithms includes the following estimators:

- An FF-NN with 2 hidden layers and 40 neurons per layer,
- a support vector regressor (SVR) with an RBF kernel,
- a decision tree regressor (DTR),
- an XGradientBoost (XGB) regressor, and
- an LSTM-based framework without spectral features.

Additionally, algorithms using spectral features were compared including

- a one-dimensional CNN,
- an LSTM-based framework, and
- an LSTM-based framework with feature extraction by a VAE.

All algorithms are trained on simulation data and evaluated on the experimental dataset. The hyperparameters of non-spectral estimators are optimized using a grid search with 250 configurations.

The choice of the one-dimensional CNN was due to varying spectral data input sizes based on the number of intermediate nodes in the network topology. The LSTM structures are built following the description in Section 4.2.

The results of the comparison are depicted in Fig. 4.10. Among the non-spectral algorithms, the R^2-scores are generally low, showing scores below 0.5. However, ML solutions based on neural networks perform relatively better, with R^2-scores above 0.5. The tree structures (DTR and XGB) perform the worst, followed by the SVR. The simple FF-NN achieved the highest R^2-score among non-recursive ML algorithms, reaching 0.51137. This indicates that the algorithms cannot compensate for the discrepancy

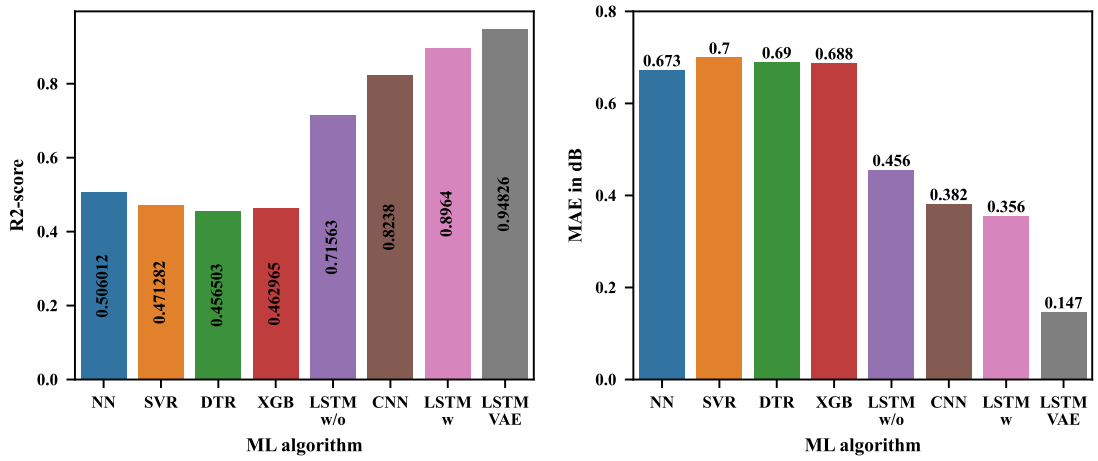


Figure 4.10: Estimation performance of the QoT estimators regarding (a) R^2-score (higher is better) and (b) MAE (lower is better).

between simulation and experimental data, which may result in poor generalization ability.

The framework without spectral features achieved a higher R^2-score of 0.71563. This improvement was attributed to the recursive structures and the inclusion of individual lengths between nodes.

However, algorithms that incorporated spectral features outperformed the algorithms using non-spectral features significantly. The CNN achieved an R^2-score of 0.8238, while the LSTM framework achieved R^2-scores of 0.8964 (with manually selected features) and an even better score of 0.94826 with features extracted by the VAE.

The MAE followed a similar performance order, reinforcing the findings that spectral information usage is beneficial for QoT estimation.

Overall, the LSTM framework without spectral feature inputs showed proficiency in learning the dependency between lengths and nonlinearities, while the inclusion of spectral features in the LSTM framework improved the GOSNR estimation further. The experimental results show the advantages of using spectral features for QoT estimation in optical networks. Furthermore, it shows the generalization capability of the LSTM framework because the framework shows very good performance on experimental data while being trained on simulation data only.

If it is assumed that monitoring data is available in the field, the accuracy of the regression is even more increased. In such a scenario the QoT estimation framework is trained on ex-

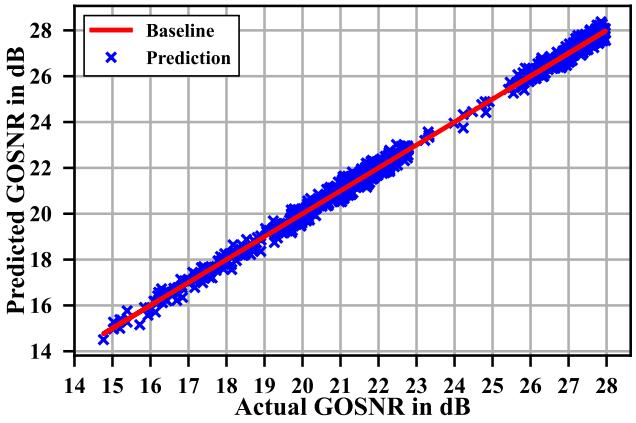


Figure 4.11: Actual versus predicted GOSNR for the VAE-based LSTM framework trained and tested on experimental data.

perimental data only and the prediction performance is significantly improved. The dataset is split into 60% training data, 20% validation data, and 20% test data. After training, the QoT estimator is tested on the test data. The results of the prediction are shown in Fig. 4.11. It shows that the predicted values are very close to the actual values, with a small deviation of only 0.145 dB. The R^2-score of 0.9961 also indicates a high degree of accuracy.

The high density of GOSNR values in the range from 16 to 23 dB and between 25.5 and 28 dB is due to the different modulation formats used. The higher values arise from the QPSK modulated samples, while the lower values are populated by the samples modulated as 8-QAM and 16-QAM symbols.

In total, the experimental investigation of the QoT estimation framework showed, that including spectral data in the information for the ML algorithm improves the accuracy of the prediction. Also, the recurrent structure of the framework increases the generalization capability of the QoT estimator.

4.4.3 Investigation of Generalization Capabilities

The generalization of an ML algorithm is not only shown by its performance on experimental data while being trained on simulation data only. It can also be shown by varying the input features of the algorithm and observing the variation of the output. By this a more in-depth view of the dependency of the input features and the output of the algorithm can be achieved. The results presented in this section have been published in [130].

In this section, it is shown how the LSTM-FF-NN-hybrid with manually selected features responds to variations in its input features. This allows conclusions to be drawn about

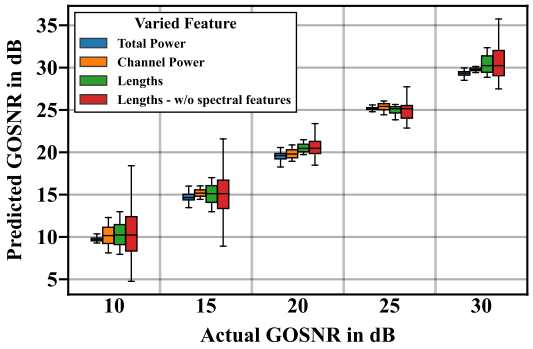


Figure 4.12: Impact of varying the input features with their standard deviations on the estimator trained on spectral data and on the estimator trained without spectral data for DP-QPSK [130].

the generalizability and robustness of the framework. Such investigations are important to provide an understanding of the importance of the component parameters and the input parameters to enable deployment in a real-world scenario.

If an ML-based estimator possesses the ability to respond reliably to variations in input features without significantly compromising accuracy on unseen datasets, it is considered both generalizable and robust. To assess the effects of the QoT estimator, it becomes crucial to identify the parameters that could potentially change when dealing with a different dataset, particularly in scenarios where exact fiber lengths are unknown, as in an agnostic network setup. In the following, the generalization capabilities of the LSTM-FF-NN-hybrid are investigated, since the spectral features can be varied individually in contrast to the VAE-based approach.

The parameters focused on for potential changes in the dataset are the total link length, the length of each span, the total power, and the channel powers. These features may vary, and to explore their impact, deliberate variations are introduced using Gaussian distributed random processes with standard deviations equivalent to 10% of the assumed accurate values. Whenever the total link length is altered, corresponding adjustments are made to the span lengths. This procedure is repeated 1000 times for each parameter to thoroughly investigate its specific influence on the GOSNR prediction.

Additionally, training of another estimator without incorporating the spectral features, namely the total power and channel powers, is done for the purpose of comparison.

Figure 4.12 displays the results of the generalization investigation. The size of the error bars indicates the impact of different factors on estimating the GOSNR. Specifically, the span lengths and the total link length have the largest impact, as indicated by their larger

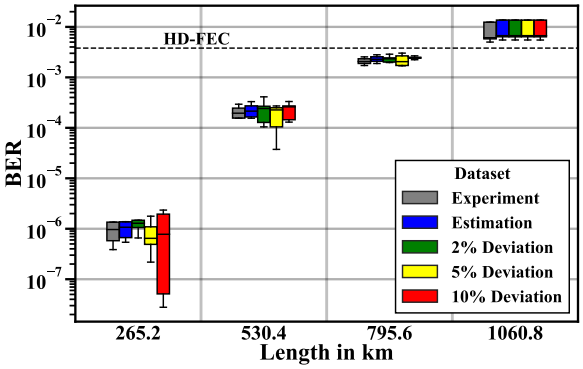


Figure 4.13: BER over length for 5 different scenarios; grey: distribution of BER in the experimental dataset; blue: estimation of the QoT estimation on the dataset; green: 2% change in the standard deviation of the dataset; yellow: 5% change in the standard deviation of the dataset; red: 10% change in the standard deviation of the dataset [130].

error bars, followed by the channel powers derived from the spectrum. On the other hand, small fluctuations in the total power, when adjusted for its standard deviation, lead to only minor deviations in the estimator. This is because most of the information about the total power is already included in the channel powers.

In the lower GOSNR regime (at 10 dB), the estimator's robustness is slightly diminished due to the dataset containing fewer high distance transmissions. However, despite this, no significant outliers are visible in the graphs, showcasing the overall reliability of the estimator.

Furthermore, when comparing the estimator trained without spectral features to the spectral data-driven estimator, it becomes evident that the former experiences much higher deviations when varying the span lengths. This observation underscores the significantly enhanced robustness of our spectral data-driven estimator in comparison to an estimator trained without spectral features.

Figure 4.13 shows an overview of the estimation process, displaying the reference data alongside four scenarios. These scenarios consist of the experiment's measured values, the corresponding estimated values, and estimations derived from input features with deviations of 2 %, 5 %, and 10 % of their standard deviation. The accuracy of the estimation is represented by the size of the error bars, with smaller error bars indicating more accurate estimates.

Changing the standard deviation for different transmission distances reveals another observation: Changes in the standard deviation have a greater effect at shorter distances

than at longer distances. This distinction is due to the inherent relationship between BER and GOSNR, where lower GOSNR values cause a significant increase in BER. As a result, modifications to the standard deviation have a more significant impact on the estimation accuracy for shorter distances due to the heightened sensitivity of the system.

However, as the distance surpasses the 1000 km mark, the situation changes. Deviations in both the experimental data and estimations become more substantial. This shift can be attributed to the presence of lower GOSNR values in the experimental dataset at these extended distances. These lower GOSNR values contribute to larger deviations in both the experimental data and the estimations, making accurate estimation a greater challenge under such conditions.

In summary, the investigation into generalization provides valuable insights regarding how the estimation process behaves in different scenarios. It showcases the impact of changes in input features at varying distances. Additionally, it is demonstrated that the estimation performance is generally good, even with changes in the input features. This is largely due to the fact that certain spectral features also incorporate the transmission distance.

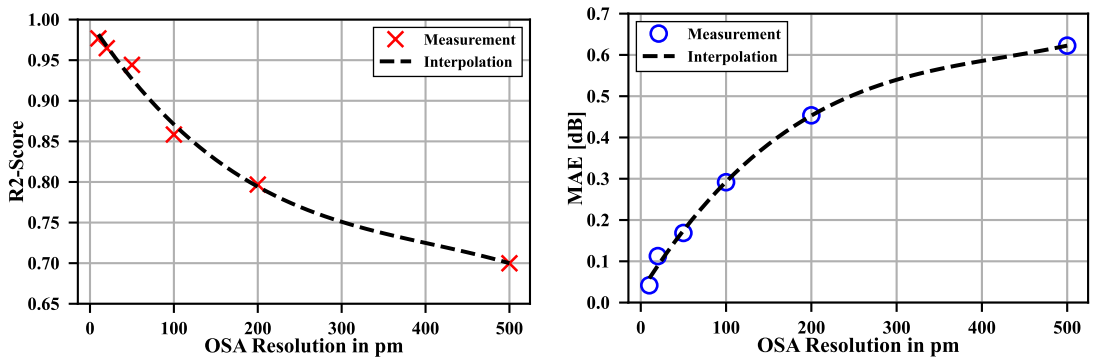
4.4.4 Performance Impact of OSA Resolution

In the previous sections, the benefit of using spectral features for QoT estimation has been shown. However, a spectral-data driven QoT estimation framework is only enabled by the availability of OSAs in the link. To investigate how the OSAs has to be chosen in terms of resolution, the performance of the VAE-based QoT estimator is examined with respect to OSA resolution. With this investigation potentially low-cost OSAs can be chosen for deployment depending on the desired QoT estimation performance. The results have been published in our work in [21].

The QoT estimator, utilizing spectral features extracted by a VAE, undergoes extensive training with simulation data and is subsequently put to the test using experimental datasets acquired at various OSA resolutions. In Fig. 4.14, a detailed evaluation of the estimator's performance, assessed by its R^2-score and MAE, is presented.

The results demonstrate that the estimator can accurately estimate GOSNR values using simulation data, and can generalize this ability to experimental datasets. This showcases the effectiveness and robustness of the estimator's training process.

Another finding emerges when examining the impact of the OSA resolution on the es-



(a) (b)

Figure 4.14: Impact of the OSA resolution on the estimation performance represented by (a) R^2-score and (b) MAE [21].

timator's performance. Lower resolutions yield superior estimation performance. This observation can be attributed to the fact that the estimator is trained using simulation spectra that were generated with an OSA resolution of 13 pm. This allows the estimator to better handle data at comparable or lower resolutions.

Figure 4.14 provides evidence of the estimator's accuracy, demonstrating low error rates even up to a resolution of 50 pm. Such a level of precision further supports the estimator's reliability and highlights its potential for practical application in real-world scenarios.

The developed QoT estimator can accurately estimate GOSNR using simulation data for training and experimental datasets for testing. Additionally, the estimator's performance is enhanced at lower OSA resolutions, demonstrating its ability to provide accurate estimates even at resolutions as low as 50 pm. If smaller deviations are acceptable, then using lower-cost OSAs with lower resolution can lead to cost savings, particularly when integrating spectrum analyzers into key nodes across the network.

4.5 Summary

In this chapter, the performance of various ML algorithms for QoT estimation was analyzed. The investigation includes scenarios where these algorithms are trained on simulation data and then tested on experimental data. In addition, the investigation extends to evaluating the impact of spectral data and recursive ML architectures on the overall effectiveness of these estimators.

The group of machine learning algorithms being evaluated includes various techniques, such as feed-forward neural networks, support vector regressors, multi-layer perceptrons, and tree structures like XGradientBoost. Additionally, LSTM networks are also part of this list. Two different training modes were examined for the use of LSTM: one trained with spectral data and the other without. The spectral data includes either selected features or features extracted automatically through a VAE.

The evaluation was based on experimental data obtained through a recirculating loop setup that included various configurations such as 32 GBaud DP-QPSK, DP-8-QAM, and DP-16-QAM, with up to 5 channels spaced at 37.5 GHz. The findings of this in-depth investigation were conclusive showing that algorithms based on spectral features show excellent performance on experimental data, even when trained on simulation data. This is highlighted by achieving R^2-scores exceeding 0.9.

In addition, by utilizing VAE-based spectral feature selection it is possible to achieve estimations with an MAE of less than 0.2 dB with only 50 pm of OSA resolution. This novel approach highlights the potential for accurate QoT estimation, even when faced with limited spectral information. It was observed that combining input features distributed heuristically to encapsulate component parameters of an uncertain nature with spectral features derived from OSAs serves to enhance the accuracy of QoT estimation.

This chapter presents an important aspect in estimating QoT, which has the potential to provide reliable estimates in complex settings, like multi-vendor networks. Furthermore, this advancement could accelerate the transition to fully disaggregated networks without any need to share confidential component information.

Chapter 5

Soft-Failure Detection, Identification and Localization

In optical networks, the need for uninterrupted service and quick failure recovery has led to a focus on failure management, which includes failure detection, identification, and localization. This has significant impact on network resilience. By detecting the source of failures, the duration of repair efforts can be significantly reduced, which helps to maintain optimal network performance [11]. A high-level view on potential component failures is shown in Fig. 5.1.

This applies not only to immediate hard failures, such as service disruptions on the optical layer, but also extends to subtle and gradual degradations that may eventually manifest as hard failures, known as soft-failures. Predicting and addressing these issues before they become hard failures offers significant benefits and allows for quick actions such as traffic rerouting or equipment changes.

Aging of deployed equipment decreases the overall QoT. Commissioning tests are usually performed on site to confirm the proper functioning of resources that support lightpaths, like optical switching and amplification [132]. These tests often require manual interven-

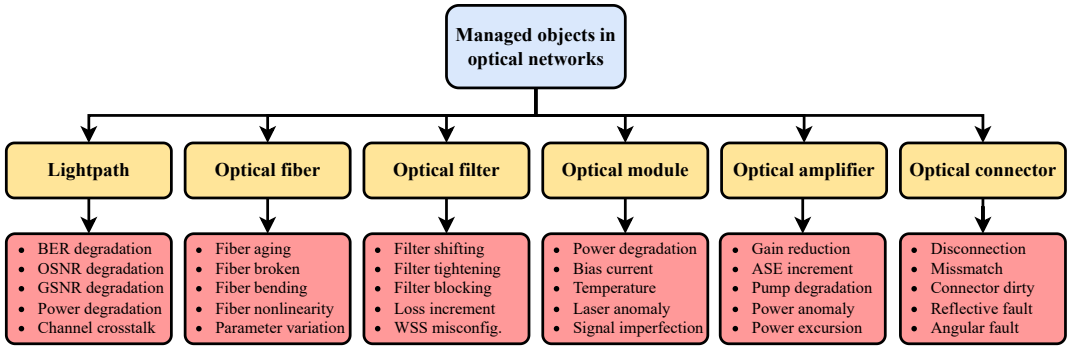


Figure 5.1: Failure sources in optical networks with the components and their corresponding potential failure sources; adapted from [131].

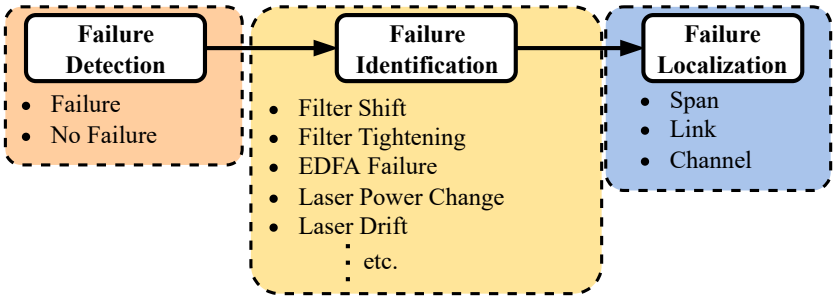


Figure 5.2: Overview of a general soft-failure management framework including failure detection, identification and localization stages.

tion from technicians and specialized equipment. In cases of high BER during testing, additional measurements may be necessary at intermediate nodes, which can increase the complexity of the process [11].

5.1 Failure Sources in Optical Networks

For soft-failures, not only power-dependent but also frequency-dependent effects are relevant. For analyzing the optical spectrum, OSAs are essential. The initial cost constraints have been reduced with the availability of cost-effective sub-gigahertz models. Real-time monitoring of optical spectra and SNRs has become crucial, with features characterizing the optical spectrum itself. ML algorithms, trained on these features, are capable of detecting and identifying faults.

Typically, failure management consists of failure detection, identification, localization and reaction. In this work, however, the latter is not examined, since the reactions are undertaken individually by the network operator. The overall structure of an ML-based soft-failure management framework is depicted in Fig. 5.2. Such a framework consists of three stages, each of which is responsible for one of the three different tasks of detection, identification and localization. Each stage activates the next stages, so that if a failure is detected, it can be identified and localized. Localization can be either on a per-span level, on a per-link level or even on a per-channel level. To identify the failing component in the network, a per-span localization is of high interest.

Typical soft-failures to investigate are

- the shifting of the center frequency of a filter, called filter shift,
- the narrowing of a filter characteristic, called filter tightening,
- the increase of ASE noise from an EDFA through a power degradation in the pump

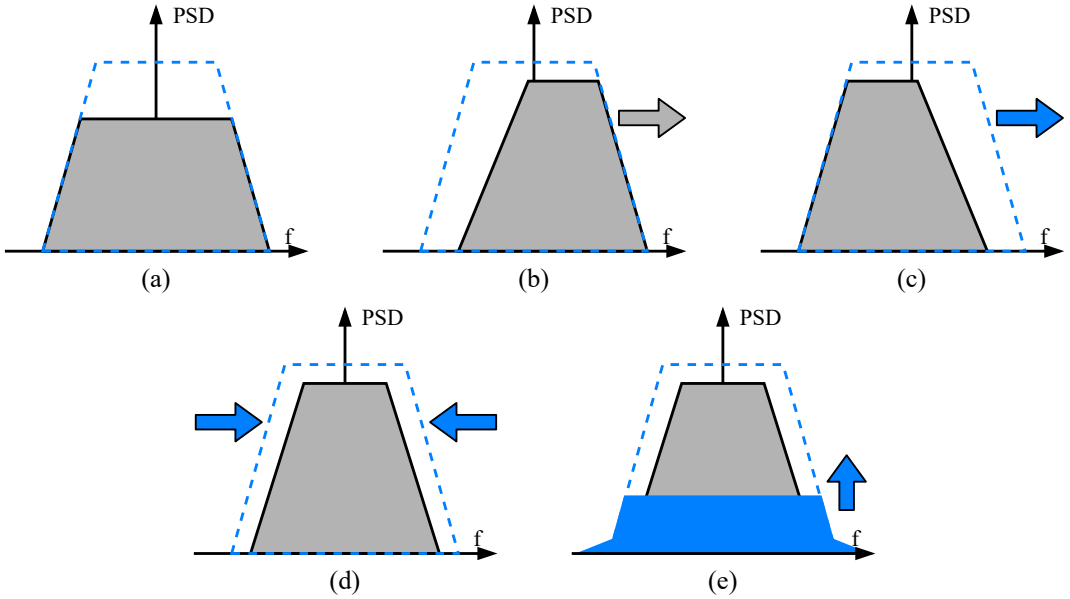


Figure 5.3: Spectrum shape varation behaviors of the five considered soft-failures: (a) laser power change; (b) laser drift; (c) filter drift; (d) filter tightening; (e) ASE noise increase; adapted from [133].

laser,

- the change of the output power of a transmission laser, called laser power change, and
- the permanent shift of the center frequency of a laser, called laser drift.

Figure 5.3 shows schematic examples for the different failure types and their influences on the spectrum.

In general, simulations provide a wider parameter sweep and a more comprehensive perspective than experiments. However, simulating failures of specific components is not straightforward, because some behaviors are hard to replicate. This includes, for example, the behavior of an EDFA, as properties such as spectral hole burning, unpredictable gain tilts, and rapid on/off responses are difficult to model accurately. Additionally, simulations do not consider all factors such as component bandwidth constraints, nonlinear tendencies of lasers and modulators at low power levels, and limitations of ADCs and DACs. Although this work does not address the simulation of these components, it is worth noting that some partial implementations have been done for more exact simulations. The implementations include bandwidth limitations, transmitter and receiver skews, IQ-imbalances, PN and some more easily simulated distortions on the signal. Our experimental investigations showed that machine learning algorithms performed better than expected compared

Table 5.1: Experimentally emulated soft-failures.

Soft-failure	Range	Steps
EDFA noise figure increase	0.2 to 2 dB	0.2 dB
Transmit laser drift	-2.5 to 2.5 GHz	0.5 GHz
Transmit laser power change	-2.5 to 2.5 dB	0.5 dB
Filter tightening	1 to 5 GHz	1 GHz
Filter shift	-2 to 2 GHz	1 GHz

to the same algorithms running on simulation data. This difference can be attributed to the aforementioned limitations of simulations, since identical algorithms were used for comparison. Therefore, the following investigations will only include experimental investigations based on datasets collected in the laboratory.

5.1.1 Experimental Soft-Failure Emulation

Since it is not feasible to physically damage laboratory equipment to simulate faults, the faults are generated by manipulating component controls and utilizing additional components.

Five different types of soft-failures are considered, in particular: an increase in the noise figure of the EDFA, a shift in the transmit laser frequency, a decrease in the transmit laser power, a tightening of the optical filter, and a shift in the filter characteristics. A straight-line experiment is conducted to simulate these fault scenarios, as shown in Fig. 4.6(a), by using three 88.4 km long SSMFs to replicate various failure conditions. Table 5.1 summarizes these soft-failures with their magnitudes.

To emulate an increase in EDFA noise caused by degradation of the pump laser, a variable optical attenuator (VOA) is introduced at the midstage access of the inline EDFAs. The VOA's attenuation ranges from 0.2 to 2 dB in increments of 0.2 dB. For the purpose of simulating a laser drift, the transmit laser's frequency for the center channel is adjusted within a range of -2.5 to 2.5 GHz in increments of 0.5 GHz.

To simulate a change in laser power, it is varied by -2.5 to 2.5 dB, in increments of 0.5 dB. This procedure is also applied to several randomly selected channels within the waveshaper, which is responsible for shaping ASE noise for generating the signal loaders.

For the case of filter tightening, the waveshaper controlling the loaders is utilized to narrow the channels by 1 to 5 GHz, in increments of 1 GHz. Shifting the center frequency of the waveshaper in increments of 1 GHz from -2 to 2 GHz allows to achieve the effect

Table 5.2: Experimental system parameters for soft-failure emulation.

Parameter	Symbol	Value
Modulation format	MF	DP-QPSK, DP-8-QAM, DP-16-QAM
Symbol rate	b	32 Gbaud
Channel spacing	Δf	37.5 GHz
Number of channels	N_{Ch}	1, 3, 5
Launch power	P_{L}	-3, -2, -1, 0 dBm
Center wavelength	λ_{c}	1550.004 nm
SSMF length	L_{Loop}	$3 \cdot 88.4$ km
OSA resolution	R_{OSA}	10 pm

of filter shift.

With a sweep over the experimental parameters in Table 5.2 and the sweep over the different failure scenarios, a dataset is obtained containing almost 800 spectra per failure type. For a comparison to the non-faulty case, the dataset obtained with the recirculating loop from Fig. 4.6b is used. To accommodate for class imbalances, the F1-score is used as an accuracy metric and the classes are weighted within the algorithms based on their occurrences.

5.2 Autoencoder-based Spectrum Assessment

As introduced in Section 3.4.1, AEs are trained to map high dimensional data to a lower dimensional space, called the latent space, and to reconstruct the input data as precise as possible from the latent space. Through the training procedure, the latent space is a low dimensional representation of the high dimensional input data. Basically, the encoder of the AE can thus be used for input feature reduction while preserving the most important information. When using input feature reduction, the size of the following algorithms is reduced, thereby decreasing training time and overall complexity. Another advantage of autoencoders is their semi-supervised learning, which is particularly useful for anomaly detection. The learning of autoencoders for anomaly detection purposes is considered semi-supervised as training is performed in a supervised manner using labeled data. However, the AE can respond to abnormal inputs because the reconstruction fails when the input data differs significantly from the training data.

AEs have already been used for semi-supervised anomaly detection based on its reconstruction error in previous work (e.g. [12]). In this work, an extension of conventional AEs

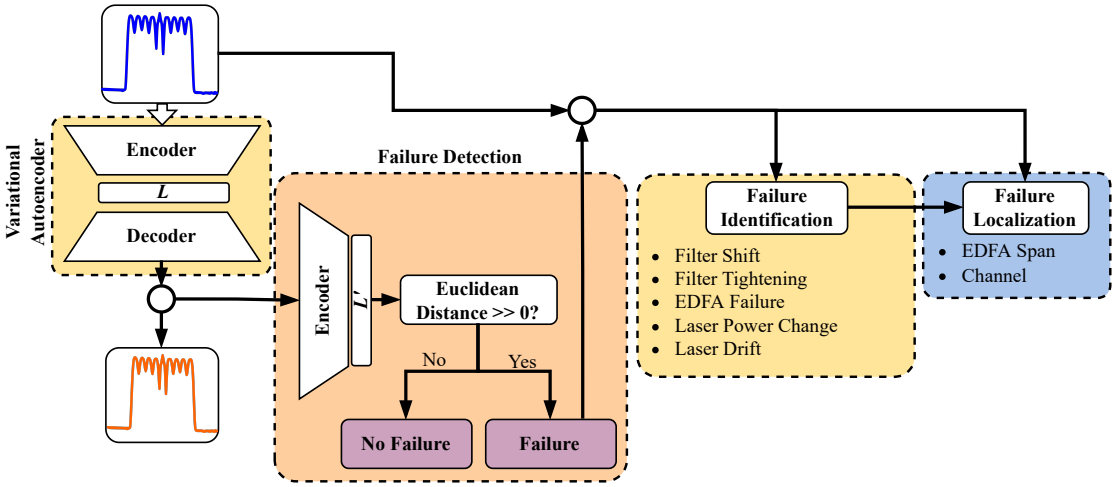


Figure 5.4: Soft-failure management framework based on optical spectra with failure detection, identification and localization stages; detection realized with Euclidean distance; L: latent space.

is used, called a VAE. A VAE utilizes stochastic variables as latent variables due to its probabilistic encoder. This property enhances the VAE's anomaly detection capabilities since normal and anomalous data may exhibit similar mean values but differ in variance meaning coming from another distribution [134]. The stochastic nature of the latent space allows for generating outputs from the decoder by sampling latent space variables from its known normal distribution. However, the latter will be used to generate a GAN in the extension of the framework in Section 5.3.

The proposed framework was extensively optimized using a grid search with 80,000 configurations. The VAE encoder was configured with an input layer of size 501, one hidden layer of size 25, batch normalization, and an output layer of size 12. Consequently, the latent space size is also 12, which matches the input layer size of the decoder part in the VAE. The decoder structure mirrors that of the encoder. The ReLU function serves as the activation function for all layers.

In the following the overall framework structure will be explained on a higher level first before going into more detail on the three different stages which also include results from the experimental investigations. The results in this chapter have been partly published in [23, 135].

5.2.1 Framework Structure

The proposed soft-failure management method consists of three stages: detection, identification and localization. It can be seen in Fig. 5.4. These stages work together to effectively manage the soft-failures. The detection stage uses a VAE to detect soft-failures.

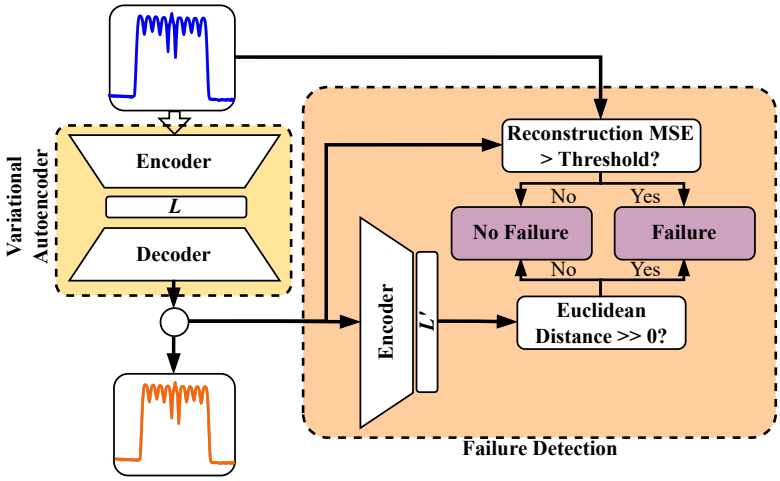


Figure 5.5: Soft-failure detection stage with MSE-based and Euclidean distance-based detection.

This approach aims to identify deviations from the expected optical spectrum and labels possible failures for further analysis.

Central to this approach is the VAE's ability to detect soft-failures in the optical spectrum. As mentioned above, the VAE is able to identify anomalies that are an indication of underlying issues in a semi-supervised manor. The VAE enhances this detection capability with the probabilistic distribution in the latent space, because normal and anomalous data may exhibit similar mean values but differ in variance. These advantages make it particularly applicable in the situation in optical networks. To examine each stage in more detail, the following sections will delve into the particular facets of the entire process.

Once a potential fault is detected, the framework moves to the next stage. This stage involves an ML-based classifier that determines the cause of the soft-failure event. If a particular anomaly, such as an increase in EDFA noise figure or a decrease in channel power, is successfully identified, the stage for failure localization is triggered. To locate the fault precisely, another ML-based classifier is trained to accurately identify the location of the underlying problem. This localization process applies to individual spans and channels, providing high spatial accuracy in identifying the source of the issue. The output of this stage will indicate either the span number of the fault location or the specific implicated channel.

5.2.2 Soft-Failure Detection

The detection of soft-failures depends on the semi-supervised ability of the VAE, as explained in the introduction to this section. During training, the VAE is optimized to

reduce the reconstruction error for non-faulty data. When a fault occurs, the resulting reconstruction MSE exceeds that of non-faulty data. This distinction permits the use of a fault detection mechanism utilizing a threshold-centered approach, where the threshold for the reconstruction MSE is fine-tuned for optimal performance. Another technique involves comparing the Euclidean distance between the latent space of the encoded input spectrum (L) and the latent space of the encoded reconstructed spectrum (L'). An anomaly is detected if the Euclidean distance is significantly greater than zero. This method offers an advantage over MSE threshold-based methods for comparing reconstruction errors by eliminating the need for threshold optimisation, since the faulty spectra might be slightly different from the non-faulty spectra. The MSE of different error spectra may deviate only slightly from the error-free spectra, while the Euclidean distance differs significantly in such cases. Figure 5.5 illustrates the two approaches used for failure detection. It is clear that using the Euclidean distance method requires accessing the VAE encoder, which takes more time to run during the detection process and adds more complexity.

The detection capabilities of the framework are assessed by partitioning the dataset into 60% training data, 20% validation data, and 20% test data. The threshold optimization for the MSE-based detection is depicted in Fig. 5.6. The MSE-based detection achieves an F1-score of 0.9881, while the Euclidean distance-based mechanism attains a perfect F1-score of 1. This discrepancy can be explained as follows: During failure occurrences, the Euclidean distance becomes notably large, often in the order of 10^4, whereas a reconstructed non-faulty spectrum from the VAE typically exhibits an Euclidean distance within the range of 10^2. In consequence, the utilization of a well-trained VAE under this approach results in notably high accuracy.

Nevertheless, the threshold is set to be at 10^2 to cover a broader range of failures and

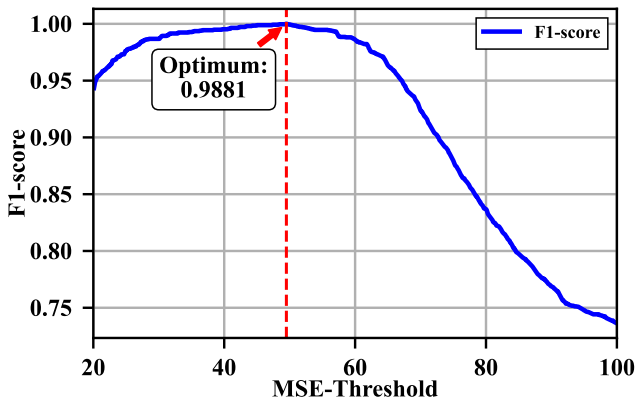


Figure 5.6: MSE-threshold over F1-score with indicated optimal threshold [23].

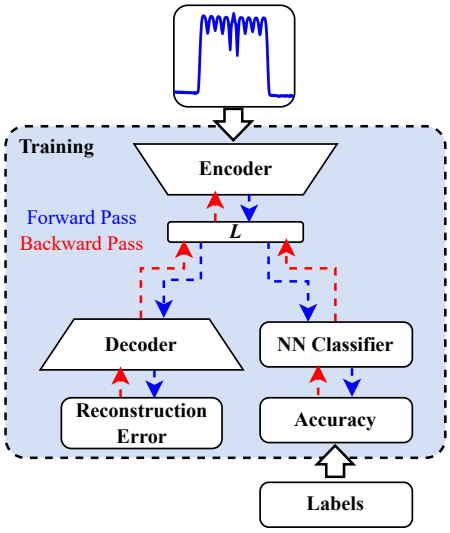


Figure 5.7: Variational autoencoder and neural network classifier in a two-step training approach [135].

identify minor deviations in the optical spectrum, leading to an F1-score of 0.9941. The slight inconsistencies in the dataset are responsible for the slight deviation from perfect detection, as it intends to encompass a broader spectrum of failure scenarios and variations.

5.2.3 Soft-Failure Identification

To evaluate the performance of the proposed framework, a comparative analysis is conducted against less advanced ML algorithms and a hybrid model that integrates VAE and NN. The ML algorithms include a linear classifier, a k-nearest neighbor classifier, a support vector machine classifier using an RBF kernel, a decision tree, and a random forest classifier. The architecture of the VAE-NN hybrid is illustrated in Fig. 5.7. This design allows for a two-step training approach. First, the VAE is trained to minimize the reconstruction error between the input and output spectra. Next, the encoder of the VAE and the neural network-based classifier for soft-failure identification are trained. The two-step training method creates a beneficial synergy between the encoder and the NN classifier. This happens because it separates failure types within the latent space while clearly distinguishing non-faulty data from faulty data. Additionally, this methodology can lead to a reduction in the scale of the neural network since the latent space serves as input. Consequently, this results in a less complex NN-based classifier, which speeds up training and improves generalization capabilities thanks to the unified training approach.

To perform a comprehensive evaluation using all available training data, the dataset is partitioned again, with 60% for training, 20% for validation, and another 20% for

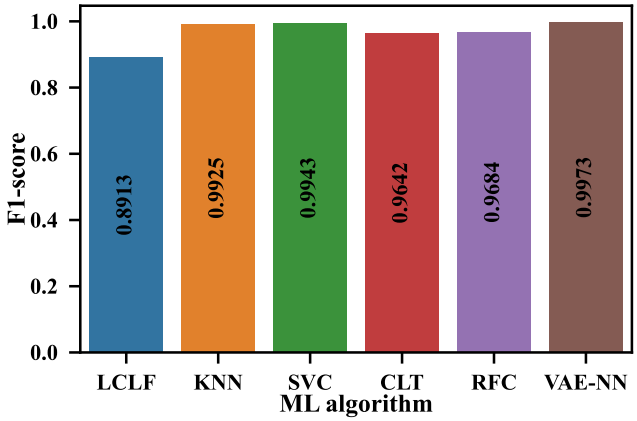


Figure 5.8: F1-scores for the identification task of different ML-algorithms on total training data; LCLF: linear classifier, k-NN: k-nearest neighbors, SVC: support vector machine classifier, CLT: classification tree, RFC: random forest classifier.

testing. To optimize the machine learning algorithms, an exhaustive grid search with over 400 configurations per algorithm is conducted to get the best F1-score performance as possible. The results of the optimization process are depicted in Fig. 5.8 and show distinct F1-scores for different classifiers.

Apparently, the linear classifier has the worst F1-score of 0.8913. In contrast, the tree-based models, specifically the decision tree, show a fairly improved F1-score of 0.9642, while the random forest classifier achieves an even higher F1-score of 0.9684. The k-

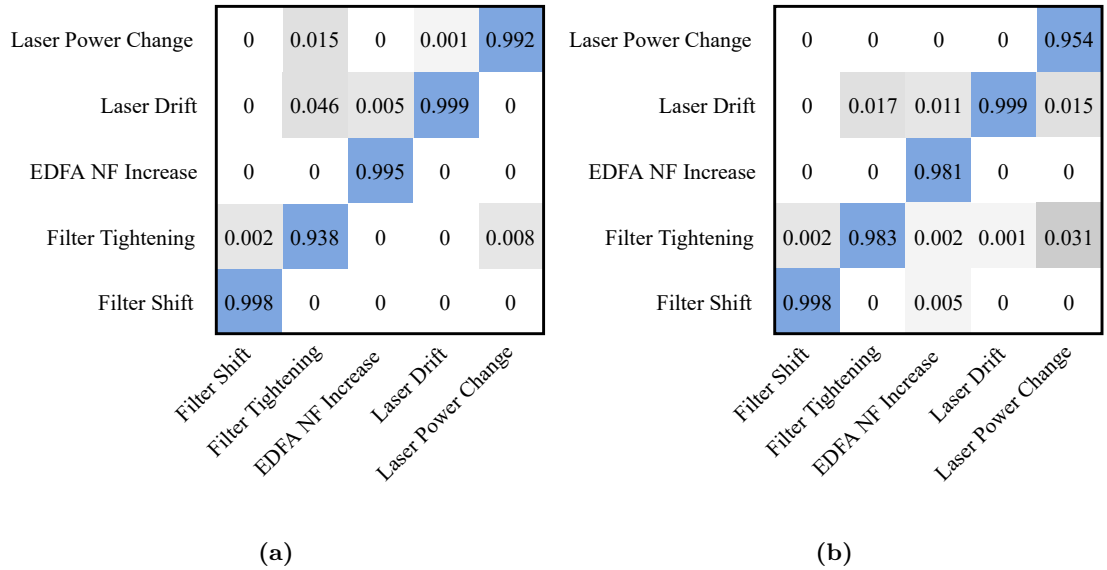


Figure 5.9: Confusion matrices for failure identification by (a) the SVM-based classifier [23] and (b) the k-NN classifier.

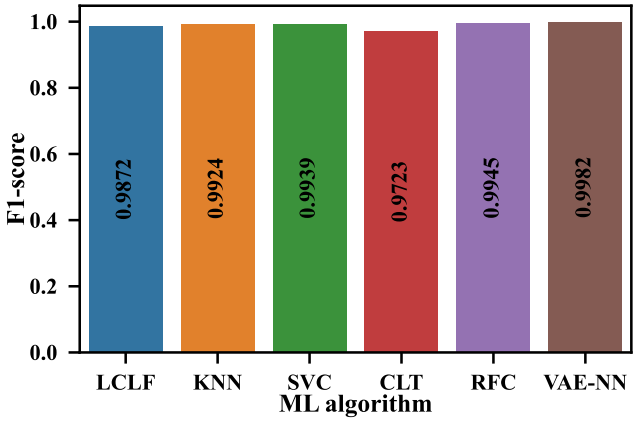


Figure 5.10: F1-scores for the EDFA failure localization task of different ML-algorithms on total training data; LCLF: linear classifier, k-NN: k-nearest neighbors, SVC: support vector machine classifier, CLT: classification tree, RFC: random forest classifier.

nearest neighbor approach reaches an F1-score of 0.9925, which makes it one of the best performing approaches. However, the support vector machine with an RBF kernel surpasses it, achieving an even higher F1-score of 0.9943. The VAE-NN hybrid achieves an F1-score of 0.9973, making it the best performing approach in this evaluation.

As an example of the classification performance of the ML algorithms, the identification confusion matrices for the SVM and k-NN classifiers are shown in Fig. 5.9. The overall classification accuracy is high, with all diagonal values exceeding 0.93. The SVM-based classifier, for example, misclassifies a filter shift as a filter tightening in 0.2% of the cases. False classifications are highly unlikely, according to the confusion matrices. However, the SVM exhibits false positives for the laser drift of 4.6%, which could be caused by sub-optimal SVM parameters. This outlier is also visible for the k-NN classifier, which incorrectly classifies laser power change as filter tightening in 3.1% of cases. This may be due to the fact that if the laser center frequency is close to the filter slope, tightening the filter leads to a decrease in the overall signal power.

Overall, the identification of soft-failure causes in the experimental data using the optical spectrum yields high F1-scores with few misclassifications.

5.2.4 Soft-Failure Localization

The same machine learning algorithms mentioned earlier are used for the analysis and comparison of the localization capabilities. The data partitioning strategy remains consistent with the one used in the identification phase when examining the complete training dataset. The details concerning optimal F1-scores can be seen in Fig. 5.10. Notably, the

combined F1-scores are remarkably high, easily exceeding the 0.97 threshold.

Among the algorithms evaluated, the decision tree has the lowest F1-score of 0.9723. It is closely followed by the linear classifier. Conversely, the k-nearest neighbor, SVM-based classifier, random forest classifier, and the VAE-NN hybrid all register F1-scores above the 0.99 mark. The VAE-NN hybrid shows good localization abilities, resulting in a F1-score of 0.9982.

These outstanding F1-scores can be attributed to the nature of the task itself. Detecting a change in laser power within a given channel through an in-depth analysis of the optical spectrum is relatively straightforward. However, a challenge arises when attempting to accurately determine an EDFA's noise figure increase on a span-by-span basis. This difference in difficulty can be seen in the analysis of the confusion matrix shown in Fig. 5.11, which illustrates the performance of the SVM classifier.

Within this matrix, there is a clear difference between accurately localizing the first EDFA versus the final EDFA in the link. This variation is due to the limitations of the SVM that has an RBF kernel, which has difficulty distinguishing between an increase in the noise figure of an EDFA and a simple fluctuation in the noise figure. Adding to this complexity is the fact that even a slight reduction of 0.2 dB in the mid-stage access of an optical amplifier triggers a small increase in ASE noise. This complex interplay of factors emphasizes the multifaceted challenges of the localization effort. The localization of the EDFA, which causes the noise figure increase, can be physically explained as follows: An increase in noise figure and change in gain spectrum due to pump laser degradation leads to different non-linearities in the following span affecting the signal. This leads to an individual distortion on the signal which is also visible in the spectrum obtained by

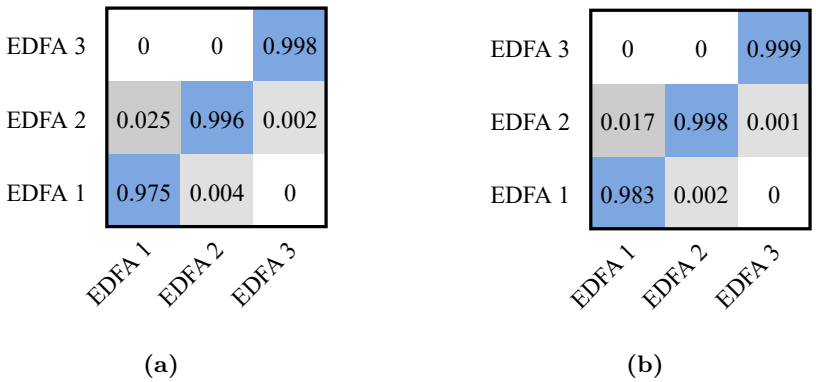


Figure 5.11: Confusion matrices for EDFA failure localization by (a) the SVM-based classifier [23, 135] and (b) the k-NN classifier.

the OSA and used by the ML algorithms. The next EDFA in the link would need to pump more which leads to an additional increase in the ASE noise. The k-NN shows the same behavior, however, the overall accuracy is higher. It has to be noted, that k-NN tends to be inaccurate in generalization tasks, i.e., using the ML algorithm on a more diverse and unseen dataset, while an SVM-classifier is more accurate in those scenarios. Overall, the localization accuracy is high showing that the small increase in noise can be recognized by ML algorithms while not being visible to the human eye. It has to be noted, that for a higher number of EDFAs in the link, the localization accuracy will be lower for the first EDFAs in the link. This may lead to a mislocalization of the underlying issued span. This problem can be addressed by dividing the link into segments, where each segment can contain multiple EDFAs. This approach allows reducing the accuracy error in long links and to increase the precision of localization.

5.3 Extension with Generative Adversarial Network

In the field of machine learning, the effectiveness of algorithms depends on the availability of extensive training data. Typically, a diverse and substantial dataset serves as the basis for these algorithms to operate successfully and yield great results when used in real-world applications.

However, when it comes to optical networks, unique challenges arise. Unlike typical scenarios, faults are rare due to conservative network designs. Additionally, fault measurement and quantification are inherently difficult. The typical approach of relying on extensive training data may not be practical given the infrequency and complexity of problems in optical networks. The goal is to develop a solution capable of achieving optimal performance in detecting failures without requiring large data sets.

In this section, the soft-failure management framework is extended to work with low amounts of training data without sacrificing on identification accuracy. Furthermore, the new approach enables the identification of unknown failures. The results in this section are based on our findings published in [135–137].

5.3.1 Training Data Augmentation

To achieve a good performance for soft-failure management without high amounts of available soft-failure data, data augmentation is a promising approach. Data augmentation has two primary objectives for which it can be used: Enhanced generalization and

improved model performance. By introducing divers data samples into a training dataset, machine learning models become more robust and achieve higher generalization to unseen or real-world scenarios. Furthermore the overall performance and accuracy of machine learning models can be improved as the model is exposed with a wider range of scenarios and variations of the input data during training. Both objectives are desired in the field of soft-failure management in optical networks.

For the case of using the optical spectrum as the input for the machine learning framework, i.e., for soft-failure detection, identification and localization, different spectra have to be generated to cover a wider range of possible spectrum variations. As mentioned in the introduction to 5.2, the stochastic nature of the latent space of a VAE allows for generating outputs from the decoder by sampling the latent space variables from its known normal distribution and transform it to a spectrum through the VAE's decoder. However, generating spectra like this will not always result in real-world optical spectra since the sampling from the latent space is done randomly. To ensure the best possible generated spectrum being used for the processing in further identification algorithms, the Euclidean distance between each input and each output spectrum is calculated and the k generated spectra with the smallest Euclidean distance to the inputs are selected, where k is the desired augmentation size. The used VAE is of the same size as the one used in Section 5.2.

An alternative strategy involves leveraging a GAN to enhance the robustness of the model to missing training data. The inherent generative capabilities of the VAE make it a valuable component in the context of a GAN for producing more realistic output spectra. As discussed in Section 3.4.2 and elaborated in the work of Goodfellow et al. on generative models [120], GANs are rooted in game theory. The primary objective of a GAN is to train a generator network capable of generating samples that mimic a specified data distribution by manipulating noise vectors. In parallel, a discriminator network is also trained to distinguish between authentic and generated inputs. This adversarial setup allows both models to improve collectively by learning from each other.

During the training process, gradients flow through the VAE, the generator (utilizing the VAE's decoder), and finally, the discriminator. This approach ensures that the VAE optimizes for both reconstruction performance and the disentanglement of the latent space, enabling the discriminator to effectively discriminate between genuine and unknown samples. It's worth noting that the GAN's discriminator comprises an input layer with 501 units, two hidden layers with 85 and 42 units, respectively, and an output layer matching the number of failure classes, which in this case, is five.

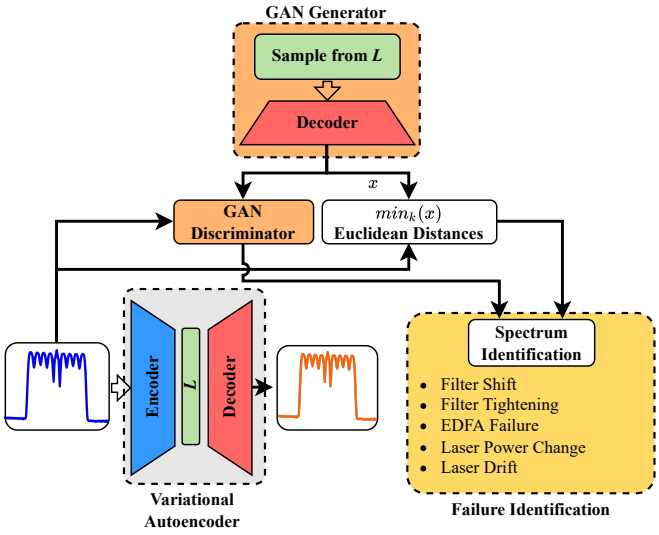


Figure 5.12: Data augmentation chain with a GAN and a VAE for soft-failure identification, L: latent space [137].

To assess the effectiveness of the two data augmentation methods, a comparative evaluation of their impact on soft-failure identification performance across various soft-failure dataset sizes is conducted. In the context of the VAE-based approach, soft-failure identification is performed using three distinct classifiers: an SVM-based classifier, a classification tree, and an NN. This evaluation encompasses both non-augmented scenarios and cases where data augmentation with a specified augmentation size, denoted as k, is applied. The data augmentation chain including both methods is depicted in Fig. 5.12.

It is important to emphasize that in the augmentation process, the number of failure samples is utilized to train both the VAE and the GAN. Subsequently, testing is carried out on the remaining portion of the experimentally generated dataset.

The results of the investigation are presented in Fig. 5.13. To underscore the advantages of data augmentation, both the VAE and GAN are trained using a specific number of failure samples. Then the F1-score is calculated for soft-failure identification, comparing scenarios with and without the utilization of augmented data.

Initially, a search for the optimal number of augmentation samples is done for each algorithm through a systematic exploration of various augmentation sizes, as visualized in Fig. 5.13a. This process helps in determining the minimum number of failure samples required to achieve an F1-score of 0.9 across all three classifiers. It becomes evident that an augmentation size of 100 necessitates a minimum of 98 failure samples to attain an F1-score of 0.9. Furthermore, all classifiers exhibit improved performance with an

augmentation size exceeding 1000. Among them, the SVM-based classifier stands out as the top performer, closely followed by the classification tree, with the neural network performs the least effectively.

To gain a more comprehensive understanding of the relationship between the number of failure samples and achieved accuracies, the algorithms are trained on varying amounts of available soft-failure training data and tested on the remaining experimental dataset. Figure 5.13b shows the achieved F1-score as a function of the number of available failure samples for both the non-augmented scenario and data augmentation involving 1000 samples.

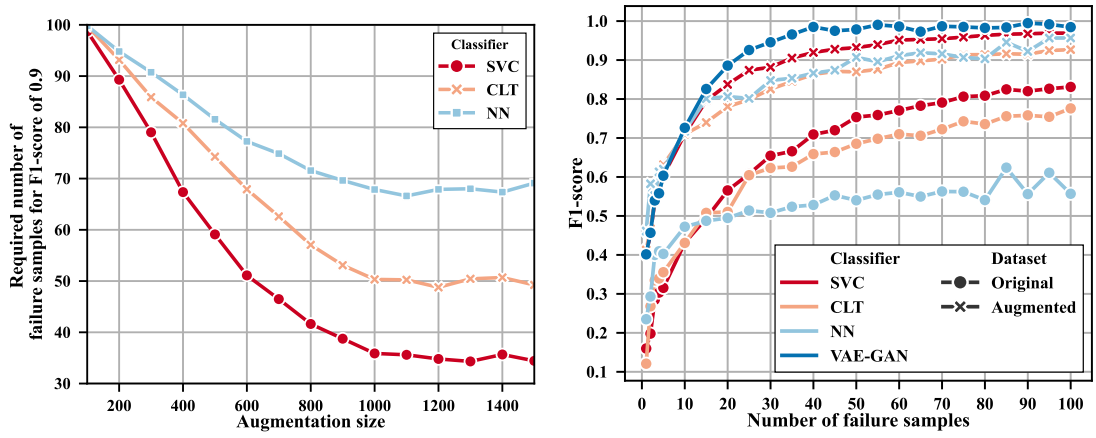
The comparison indicates that classifiers operating without augmented data consistently exhibit suboptimal performance. The SVM-based classifier achieves an F1-score slightly above 0.8 with 80 failure samples, while the classification tree reaches an F1-score of 0.7521 with the same number of samples. The neural network appears to demand a substantial amount of training data, resulting in low F1-scores across all training sample sizes, peaking at 0.6213 with 85 training samples.

However, the scenario changes when employing VAE-based data augmentation. In this context, all three classifiers reach F1-scores surpassing 0.8 when the available failure samples range from 15 to 25, eventually exceeding 0.9 in all instances. The SVM-based classifier reaches an impressive F1-score of 0.9726, while the classification tree and neural network both achieve a maximum around 0.94. Outperforming all these algorithms is the VAE-based GAN, which attains F1-scores above 0.9 with just 25 available failure samples for training, corresponding to 5 failure samples per failure class. The GAN achieves a remarkable maximum F1-score of 0.9873 with only 40 failure samples.

5.3.2 Extended Framework Structure

As previously discussed, the benefits of using a VAE make it a suitable component to be integrated into a GAN, in order to enhance the generation of more authentic output spectra. To achieve this, the dual branch strategy proposed in [138] is adopted. This strategy consists of an unsupervised branch to detect unknown faults, and a supervised branch to identify faults within the discriminator. This concept is shown in Fig. 5.14, which illustrates the two different branches of the discriminator: the lambda layer, a specialized activation function, and the spectrum identification component.

First, a supervised model is built using a softmax activation function for five unique failure classes. Then, an unsupervised model with a lambda layer is used, allowing the



(a) (b)

Figure 5.13: (a) Required number of real failure samples to reach an F1-score of 0.9 over the augmentation size from those samples; (b) F1-score over number of failure samples used to train the VAE and the GAN for data augmentation with an augmentation size of 1000; SVC: Support vector machine-based classifier, CLT: Classification tree, NN: Neural network [137].

usage of a custom activation function. The lambda layer processes the softmax output from the supervised model and calculates a normalized sum of exponential inputs, as described by [138]:

$$D(x) = \frac{Z(x)}{Z(x)+1}, \text{ with } Z(x) = \sum_{k=1}^{K} \exp[l_k(x)], \tag{5.1}$$

where D represents the discriminator, K is the number of classes, and l are the logits of the classes. Consequently, the output of the lambda layer spans the range of 0 to 1, enabling a clear differentiation between known and unknown samples. Both branches share weight connections within the hidden layers, establishing a mutually beneficial relationship where their classification performance becomes interdependent.

5.3.3 Performance Evaluation

As mentioned earlier, the proposed framework has the ability to identify previously unknown failures through its interaction with the discriminator of the GAN. Just like in Section 5.2.2, a failure is detected, if the Euclidean distance between the encoded output spectrum and the encoded input spectrum is above the predefined threshold. To identify unknown faults, this condition has to be met as well as the additional condition that the output of Eq. (5.1), i.e., the softmax output of the lambda layer, is ≈ 1. This combination of conditions effectively identifies spectra as potential unknown failures. To gain insight

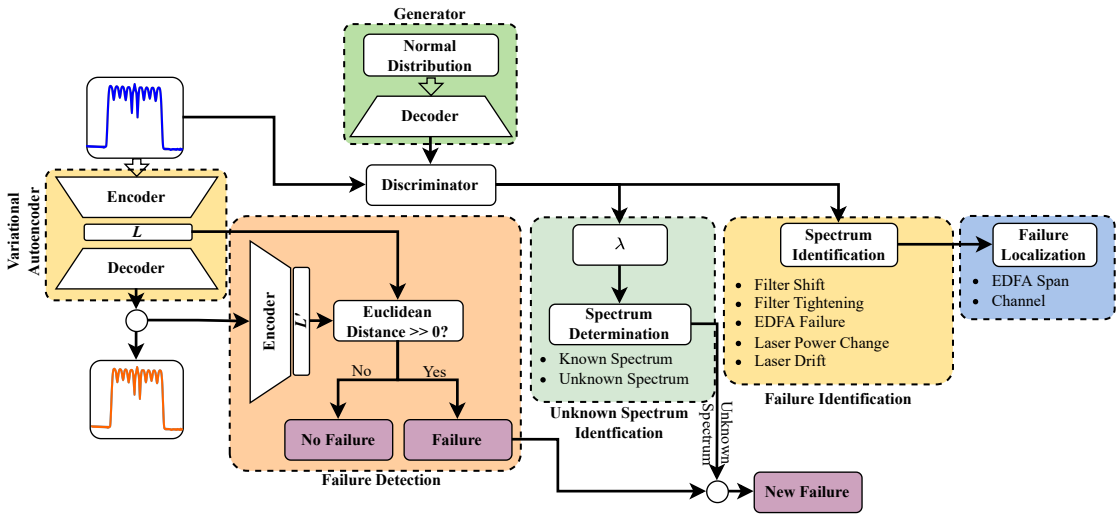


Figure 5.14: Soft-failure management framework with failure detection, identification, and localization stages in combination with a GAN for unknown spectrum identification; λ: layer with a custom activation function, L: latent space [135, 136].

into such scenarios, a random frequency misalignment of up to 0.5 GHz in all WDM loaders is done in experiments following the swept experimental parameters from Table 5.2.

For comparison purposes, unknown failures are also detected using the k-means and DBSCAN algorithm as outlined in detail in [10]. The algorithm's parameters are tuned through an exhaustive grid search to identify five different failure classes. This adjustment ensures that any outlier data points are identified as potential unknown failures. On our dataset, the k-means algorithm achieves only an F1-score of 0.5812, while, when calibrated correctly, the DBSCAN algorithm achieves an F1-score of 0.8352 without supervision.

However, the full potential for unknown failure detection of the proposed framework can be seen when it is tested against the previously mentioned algorithms. The GAN shows an F1-score of 0.9912 for the unknown failure detection. The performance of the framework relies on the dynamic interplay between the generator and discriminator of the GAN. The generator produces synthetic samples, also called "fake" samples, that are then selected to be used to train the discriminator. This cooperative approach enhances the ability of the framework to accurately identify unknown spectra. The GAN serves as an optimized filter, enabling the framework to distinguish between known and unknown failure spectra with high accuracy. However, this interplay of generator and discriminator makes the GAN in-stable during training.

The above insights are more evident when assessing the framework's applicability in scenarios where only a limited portion of the entire training data is available. For compar-

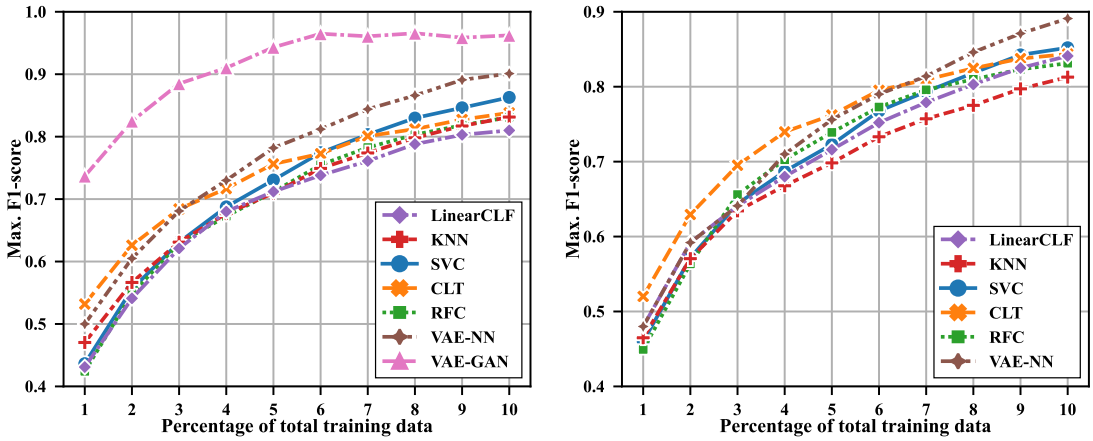


Figure 5.15: Maximum F1-score for the different machine learning algorithms in the soft-failure (a) identification and (b) localization stage over the percentage of used training data from the total number of training data; LinearCLF: linear classifier, k-NN: k-nearest neighbors, SVC: support vector classifier, CLT: decision tree, RFC: random forest classifier [135, 136].

ison, the algorithms are trained on subsets, ranging from 1% to 10% of the full training data. Through repetitive training and testing, the impact of outliers is reduced; in this case 100 repetitions were chosen. For this, 100 sub-datasets are created by drawing random samples from the training data. Each of the (sub-) datasets incorporates the corresponding percentage of the total training data amount and is evaluated individually. The results, illustrated in Fig. 5.15, reveal insights to the performance of the different algorithms when limited training data is available.

In various areas, traditional machine learning algorithms have been unable to deal with lower percentages of available training data. Although the k-nearest neighbors, random forest classifier, and linear classifier vary in their techniques, they perform similarly in the end. This complexity of the classification task highlights the need for a deeper analysis, beyond linear classifiers and simple clustering, to achieve high accuracy.

Also, it can be seen that the decision tree is able to handle limited training data quiet well. It remains competitive with its counterparts until the SVM-based classifier surpasses it at 7% training data. Similarly, the VAE-NN hybrid consistently outperforms other machine learning algorithms once the training data exceeds 4%. This performance is reflected in an F1-score of 0.902 using a 10% threshold, demonstrating the adaptability and learning efficiency of the hybrid model.

However, the proposed structure of the VAE-GAN exhibits superior performance across

all levels of training data. This is due to the inherent generative capabilities of the GAN architecture. The interplay between the generator and discriminator improves classification accuracy, as the discriminator is exposed to the generated spectra. The method peaks at about 6% of training data, beyond which additional data availability does not lead to improved results.

When dealing with limited training data in the localization task, algorithms display similar performance differences as in the identification case. The k-NN, due to its simplicity, performs the worst overall, followed by the linear classifier and random forest classifier. However, the decision tree algorithm proves to be effective in the low percentage range and performs the best, second only to the VAE-NN hybrid after the 7% training data threshold. The hybrid architecture achieves a very good F1-score of 0.891 using 10% of the total training data. It is worth noting that the SVM-based classifier outperforms the decision tree after the 9% training data threshold.

In summary, all algorithms show a consistent trend that shows a direct relation between data abundance and F1-scores. This trend confirms previous findings and highlights the critical role of training data in improving algorithm performance. The proposed VAE-GAN approach exhibits exceptional adaptivity and generative power. It shows to be a unique solution that is capable of handling identification tasks regardless of the available training data. Furthermore, the VAE-NN hybrid shows the advantages of advanced training methods for tolerance to limited training data in the localization tasks.

5.4 Summary

In this chapter, different machine learning algorithms were compared for identifying and locating soft-failures, using the optical spectrum as input data. Also it was analyzed how the limited availability of training data affects classification accuracy.

Various machine learning algorithms were considered in the comparison. These algorithms were a linear classifier, a k-nearest neighbors classifier, a support vector machine-based classifier utilizing a radial bias function kernel, a decision tree classifier, a random forest classifier, and the presented hybrid approaches of VAE-NN and VAE-GAN. The hybrid structures are contained in the proposed framework.

To evaluate these methods, actual experimental data of soft-failures were used collected from 32 GBaud DP-QPSK, DP-8-QAM, and DP-16-QAM transmission scenarios. The dataset encompassed up to 5 channels, each with a 37.5 GHz spacing, spanning three segments of SSMF with a total length of 265.4 km. The emulated soft-failures encom-

passed a variety of failure scenarios, such as EDFA noise figure increase, transmit laser frequency drift, transmit laser power change, filter tightening, and filter shift.

The results demonstrate the ability to detect, identify, and localize soft defects using the optical spectrum as input. Furthermore, a GAN helped in identifying unknown spectra. The proposed VAE-GAN architecture achieved outstanding results, surpassing those of conventional machine learning algorithms, especially when dealing with limited training data. It achieved an F1-score of 0.9821 in identifying soft-failures at only 6% of training data which corresponds in total to 40 failure samples.

For localization purposes, the proposed two-step training approach for the VAE-NN hybrid showed the best results. These results can be attributed to the mutually beneficial interplay between the autoencoder and the neural network during the training.

In conclusion, the potential of leveraging the generative capabilities of GANs combined with VAEs was shown to establish robust soft-failure management based on the optical spectrum, even in situations with limited training data.

Chapter 6

Combination of QoT Estimation and Soft-Failure Management

To date, the estimation of quality of transmission (QoT) and the management of soft-failures have been examined separately. The goal is to identify optimal algorithms suited to specific applications and verify the usefulness of optical spectrum data for those tasks. However, it is important to note that these two facets, i.e., QoT estimation and soft-failure management, are inherently interconnected. This is because soft-failures inevitably cause a decrease in QoT.

The following chapter therefore combines the previously separate frameworks into a unified approach. The elaborated framework is presented and experimentally validated. The objective is to show how QoT estimation and soft-failure management are connected and to demonstrate the effectiveness of an integrated approach. Nevertheless, it should be pointed out that while such integration offers advantages, it also has inherent drawbacks like instability during training and trade-offs in accuracy due to the combined approach.

This chapter contributes to the ongoing research on optical network management by explaining the relationship between QoT estimation and soft-failure management. It will increase the comprehension by emphasizing the synergy between these previously separate areas. Through the use of experiental data, this chapter aims to expand the current state of knowledge and provide a comprehensive perspective that recognizes the benefits and limitations associated with the interplay of these two basic concepts.

6.1 Framework Extension

The proposed framework shown in Fig. 6.1 combines the two frameworks explained in Chapters 4 and 5. To enhance clarity, the classifiers used for soft-failure detection, identification, and localization have been replaced with SVMs. The utilization of an SVMs as a classifier has been studied in the previous chapters, and it has been shown that

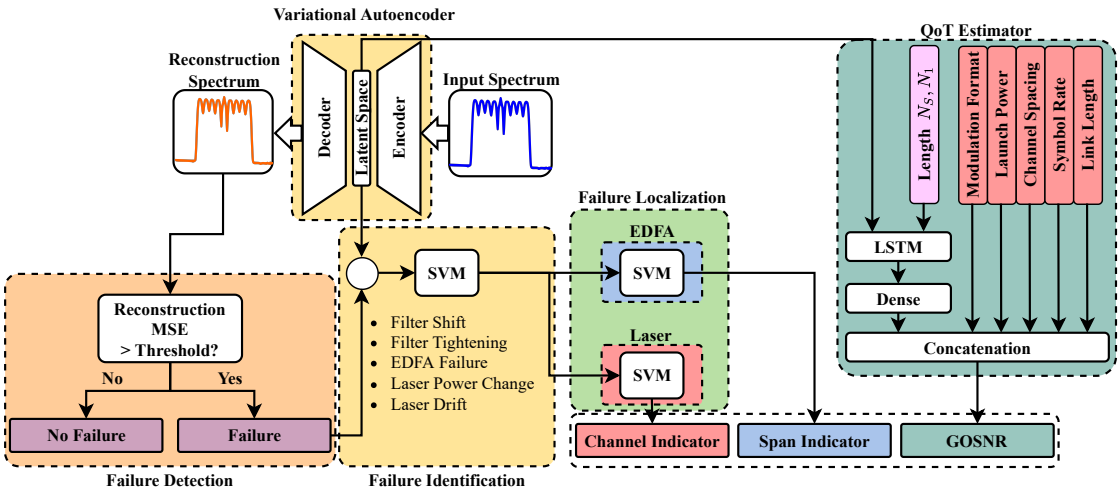


Figure 6.1: Fault management and QoT estimation framework based on optical spectrum interpreted by a variational autoencoder; N_S: current node number, N_1: starting node [23].

it can provide highly accurate results, especially when enough training data is available.

This section differs from the previous ones as it focuses on the recalibration of the QoT estimator. The new configuration gives the QoT estimator the ability to adapt to faulty data. This enables the framework and the QoT estimator not only to detect failures but also to identify the effects of soft-failures on the QoT.

This integration demonstrates changes in optical network management. The framework shows the significance of utilizing ML, specifically SVMs, to address the complexities of soft-failures. Additionally, broadening the scope of the QoT estimator offers a detailed perspective on identifying failures and their impact on the transmission quality. By combining QoT estimation and soft-failure management, the framework demonstrates a step towards automatically optimizing optical networks.

6.2 Simulative Investigation

To determine if it is possible to combine ML framework for QoT estimation and soft-failure management, simulations are done first. It should be noted that the limits of simulating soft-failures, as discussed in the previous chapter, are still applicable. The simulative investigation is based on our results published in [22, 23].

6.2.1 Simulation Setup

The simulation setup used for this investigation is the same as in Section 4.3.1 in Fig. 4.4. However, the simulation parameters are changed in a way to obtain a balanced dataset

with regards to the overall transmission distances.

A set of nine DP-WDM channels with fixed grid spacing and equal launch power per channel is to be transmitted over a certain link consisting of several spans. The missing knowledge of the component parameters is, again, represented following a heuristic approach with certain mean values and standard deviations based on realistic assumptions and margins.

The links are analyzed for various modulation formats, i.e., QPSK, 8-QAM and 16-QAM with coherent detection. The symbol rate is changed between 32, 64 and 69 Gbaud, while the channel spacing is set to 37.5 GHz for 32 Gbaud and 100 GHz for 64 and 69 Gbaud. As for the launch power per channel P_{L}, values between -3 and +3 dBm are assumed. As for the uncertain parameters for example, the span length L_S is chosen as a random length with a mean $\overline{L_{\mathrm{S}}}$ of 80 km and a standard deviation σ of 5 km. The EDFA output power $\overline{P_{\mathrm{EDFA}}} = P_{\mathrm{L,Total}}$; σ: 0.5 dB and its noise figure $\overline{NF}$: 5 dB; σ: 0.5 dB are chosen for each EDFA in the link accordingly. The linear fiber parameters α and D are set to be $\overline{\alpha}$: 0.2 dB/km; σ: 0.02 dB/km and $\overline{D}$: 17 ps/(nm·km); σ: 0.2 ps/(nm·km), respectively. The nonlinear coefficient γ is assumed to be 1.295 $(\mathrm{W{\cdot}km})^{-1}$.

Therefore, the parameters differ for each transmission, for each link and for each span. To generate the most balanced dataset possible for training the framework, simulations were performed over 1000 links with lengths drawn from a normal distribution with a maximum length of 3000 km.

Each span in these links consists of an SSMF followed by an EDFA with a flat gain characteristic in the C-band and an OSA with a resolution of 13 pm (according to the specifications of commercially available OSAs) at each intermediate node. Also, the number of intermediate nodes in each link is drawn from a normal distribution with up to 8 intermediate nodes. The propagation of the light through the links is calculated using the split-step Fourier method to achieve the most accurate values for the OSNR and for the spectrum. A total of 11 different channel scenarios are simulated, ranging from single channel transmission over the entire 9 channels to free adjacent channels.

To extract as much information as possible from the spectra of intermediate nodes while maintaining a reasonable size for saving, they are reduced to 10,000 sampling points. However, this makes them only marginally usable as features, because the number of trainable parameters in the first layer of the ANN depends directly on the number of features. Therefore, the dimensionality of the input data is reduced using a VAE leading

to smaller sized classifiers and QoT estimation structures. In addition, the VAE is used for the semi-supervised anomaly detection.

Since soft-failures per definition can result in hard failures over time, laser aging dependent failures are again the focus. For this purpose, failures of the transmit laser are emulated by reducing the launch power of one randomly selected channel by a random distribution with a mean value $P_{\mathrm{L}} - 1$ dBm and σ: 2 dBm. Moreover, a failure in an EDFA is emulated by increasing its noise figure artificially, since a degradation of the pump laser in output power controlled mode would only show in higher output noise. The flawed noise figure is drawn from a Gaussian random distribution with a mean value 6 dB and σ: 2 dB. The failure related dataset contains around 12,600 feature sets.

Due to the increased size of the obtained spectrum, the structure of the VAE is changed. The VAE is composed out of an encoder with an input layer with a size equal to the 10,000 points of the spectrum, followed by a fully-connected layer with 100 neurons, a batch normalization layer and another fully-connected layer with an output size of 10. These outputs are then processed in a sampling layer so that the output represents a normal distribution on the one hand and the input of the encoder on the other. The decoder is built to reverse the functions of the encoder.

Failure identification is performed by an SVM, which is able to distinguish between an EDFA failure and a transmitter laser failure based on the latent space of the VAE. The third stage is divided into two parts: The first classifier is able to determine in which span the error of the EDFA occurred, while the second SVM determines which of the channels of the WDM transmission is affected by a lower launch power. Each of the three SVMs used in the framework has been optimized for the kernel coefficient γ_{K} and the regularization parameter C using a grid search algorithm with 250 steps per variable. The features of the latent space are also used for the QoT estimation. The QoT estimator is the same as described in 4.2.

6.2.2 Simulation Results

The framework is tested on 1000 failures randomly distributed in an unseen dataset from the COST266 European network topology (Fig. 4.3). The failure detection using the reconstruction probability of the VAE with an optimized threshold, shows an F1-score of 0.989 whereas an AE failure detection mechanism based on a fixed threshold reaches only an F1-score of 0.913 [12]. The performances of the other stages are summarized in Figs. 6.2 and 6.3. They show the confusion matrices for failure identification and failure localization. The failure identification stage reaches an F1-score of 0.996 with a low false

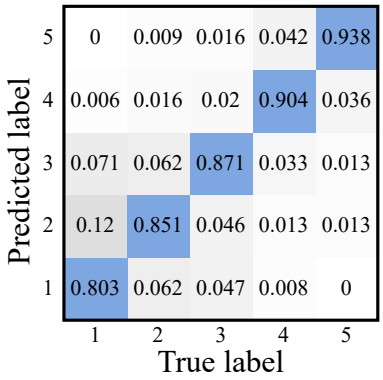


Figure 6.2: Confusion matrix of EDFA failure localization in the spans [23].

positive and false negative rate. However, since a supervised learning method is used, only previously seen classes can be labeled leading to other failures not being classified correctly. The framework shows to be capable of localizing the cause of the noise increase for an EDFA failure with an F1-score of 0.871. Also, the power drop in one channel is determined with an F1-score of 0.946. The performance of the QoT estimator is shown in Fig. 6.4. It can be seen that the predicted values are nearly on the actual GOSNR values. This is also represented in the R^2-score of 0.998 and an MAE of only 0.17 dB, which is about the same as a QoT estimator with manually selected features (see Section 4.3.2).

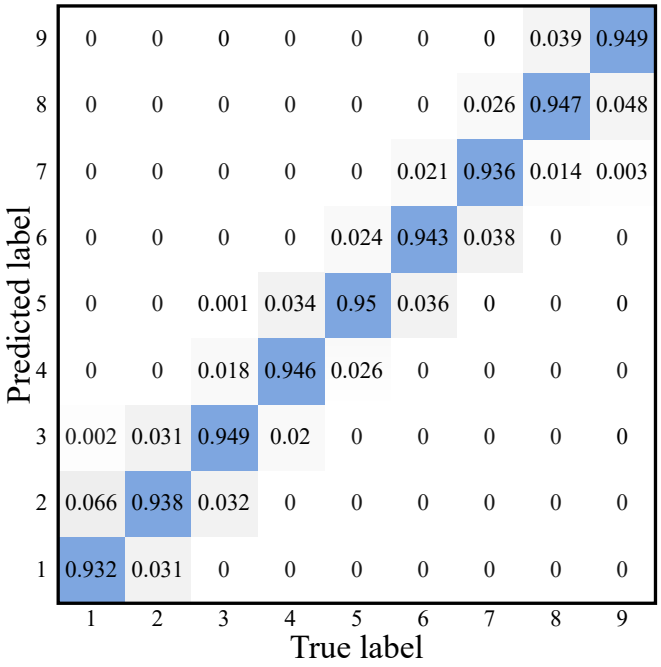


Figure 6.3: Confusion matrix of failure localization of the transmitter laser failure [23].

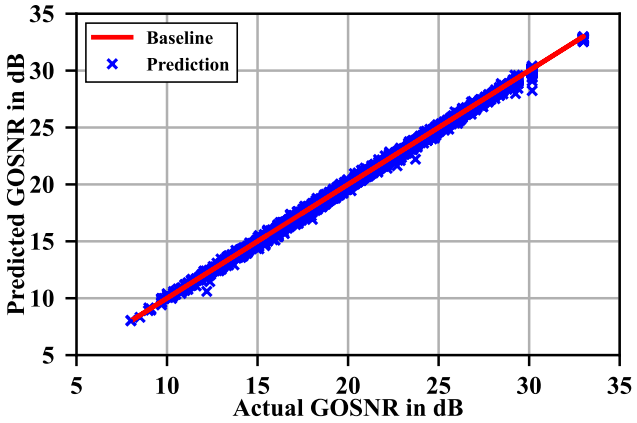


Figure 6.4: Predicted GOSNR over actual GOSNR for feature extraction with the VAE trained on simulation data; R^2-score: 0.998, MAE: 0.17 dB [22].

6.3 Experimental Validation

For experimentally validating the proposed framework, the experimental datasets for QoT estimation and soft-failure management from the previous chapters are combined and used. The presented results are based on our work published in [23].

6.3.1 QoT Estimation

To accurately evaluate the capabilities of the framework for estimatingt the QoT, the dataset is divided into three separate segments: 60% for training, 20% for validation, and the remaining 20% for testing. Following the training phase, the QoT estimator underwent exhaustive testing not only on the designated test data but also on data including soft-failures. The visual representation of the prediction outcomes is depicted in Fig. 6.5.

Remarkably, the analysis reveals minimal deviations from the baseline, with an R^2 score of 0.9846 and low MAE of only 0.29 dB. The noticeable concentration of GOSNR values between 16 to 23 dB and 25.5 to 28 dB is due to the different modulation formats used. Specifically, the higher values are mostly associated with samples modulated with QPSK, while the lower end of the scale is dominated by samples using 8-QAM and 16-QAM symbols.

Invariably, due to soft-failures in failure measurements, data points on the graph may deviate from established patterns. Detecting and understanding these data points provides a more in-depth view of the system performance and potential weaknesses. The following section will explore the soft-failure detection aspect in more detail, using the QoT metric.

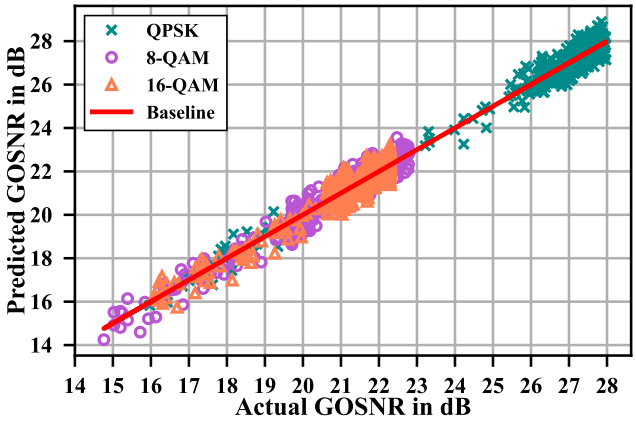


Figure 6.5: Predicted GOSNR by the estimator over actual GOSNR for all configurations in the dataset trained on experimental data; R^2-score: 0.9846, MAE: 0.29 dB [23].

6.3.2 Soft-Failure Detection

In this investigation, a technique that uses a QoT metric-based thresholding mechanism to detect soft-failures is employed. Initially, an acceptable range for fluctuations in the GOSNR is determined for the linear experiment over a distance of 265.2 km. This step is crucial in determining the appropriate threshold for each modulation format.

By analyzing the histogram of GOSNR occurrences in Fig. 6.6, certain characteristics

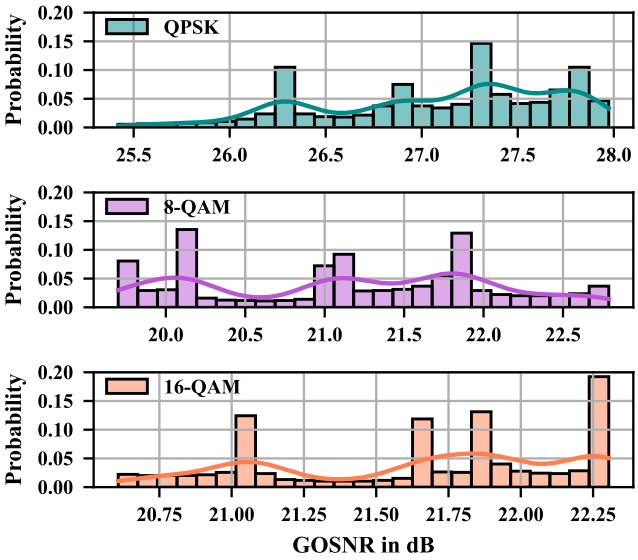


Figure 6.6: Histogram of the occurrence of GOSNR values in the dataset [23].

can be seen. At approximately 26.2 dB, a peak in the GOSNR distribution is present, corresponding to the QPSK modulation. This indicates a concentration of failure instances in this specific range. Similarly, such a peak for 8-QAM is located around 20.3 dB, while for 16-QAM it appears at approximately 21.1 dB.

By considering these specific thresholds as signs of failure for each modulation format an F1-score of 0.7135 can be achieved. This low score is due to the nature of emulated soft-failures, which may not always result in a significant decrease in GOSNR levels. Thus, the complex interplay between modulation format, fluctuations in GOSNR, and the detection of soft-failures results in the diverse results noted in this investigation. Comparing this approach to the detection capabilities of the VAE from the previous chapter, it becomes clear that the usage of the spectral data driven VAE is beneficial. This can reach nearly perfect F1-scores when relying on the Euclidean distance as mentioned in 5.2.2. The stages of soft-failure identification and localization remain the same as in the previous chapter, resulting in the overall framework performance summarized in Table 6.1.

6.4 Summary

In this chapter, a comprehensive framework for the simultaneous estimation of QoT and the detection, identification, and localization of soft-failures is presented, which is based on the analysis of optical spectra. Additionally, the inherent benefits of a soft-failure detection mechanism based on VAEs compared to an approach solely focused on QoT metrics are explored.

The accuracy of the proposed framework was confirmed through extensive testing. Data gathered from an experimental transmission setup with DP-QPSK at 32 GBaud, DP-8-QAM, and DP-16-QAM that covered up to 5 channels separated by 37.5 GHz intervals was used for testing purposes. The data for QoT estimation was generated using a recirculating loop that had 6 iterations of three spans. Each span consisted of 88.4 km of SSMF and three EDFAs. The data acquisition for soft-failures was done with a three-span

Table 6.1: Overall framework performance.

Framework Stage	Max. R2/F1-score
QoT Estimation	0.9846
Soft-Failure Detection	0.9941
Soft-Failure Identification	0.9943
Soft-Failure Localization	0.9939

setup. The soft-failure emulations included scenarios such as an increase in noise figure of the EDFA, frequency drift of the transmitting laser, reduction in transmit laser power, filter tightening, and filter shifting.

The results showed the effectiveness of the proposed VAE-based approach for robust QoT estimation. This was supported by an impressive R^2-score of 0.9846 and an MAE of 0.29 dB. The significance of the optical spectrum when estimating QoT and handling soft-failures was highlighted. The advantages and feasibility of this approach are apparent and relevant to real-life scenarios for enabling accurate and autonomous network management.

Chapter 7

Conclusions and Outlook

In this work, different machine learning based solutions for QoT estimation and soft-failure management were investigated. To enable reliable operation of the ML algorithms, data has to be obtained. Transparent OPM is desired for low-effort evaluations of the data. This can be achieved using OSAs to obtain the optical spectrum. The optical spectrum was used to train the novel ML-based solutions. The novel approaches were first developed using extensive simulations before being experimentally validated using data obtained through a comprehensive laboratory environment.

To obtain a reliable QoT estimate, the nonlinear impairments of an optical link must be evaluated. Several approaches have been developed, including the analytical GN model and full-fiber propagation simulation using the SSFM. While the GN model offers low computational time, the SSFM achieves higher accuracy but suffers from high computational complexity and thus long computation time. However, extensions of the GN model, such as the IGN and closed-form methods, approximate nonlinear perturbations to reduce computational time. Both analytical and numerical models suffer from uncertain component and system parameters that interfere with the seamless operation of the estimators. Properly trained ML algorithms, however, can cope with varying input parameters. Such ML-based QoT estimation combines high accuracy with fast computation. Training these algorithms can take time, but once trained, ML estimation is fast and suitable for real-time applications.

Several ML algorithms were evaluated for estimating QoT in uncertain parameter environments based on optical spectral data. The algorithms were trained on simulation data which followed a novel approach of heuristically distributed component parameters and tested on experimentally obtained data. It has been demonstrated that the use of spectral data improves the overall accuracy of QoT estimation. Additionally, recursive interpretation of optical spectra specific to each node in conjunction with RNNs further improves the estimation accuracy. Advanced ML techniques were used to reduce input features and select optimal features for QoT estimation. The VAE-LSTM approach performed better

than the LSTM-neural network (NN) solution with manual feature selection by reducing estimation outliers with the influence of a VAE. The VAE-LSTM structure achieved an MAE of only 0.147 dB with an R^2-score of 0.948, based on experimental data while being trained on simulation data. It has been demonstrated in the investigation of required OSA resolutions that using VAE-based spectral feature selection produces estimates with an MAE of less than 0.2 dB with 50 pm of OSA resolution. The proposed approaches exhibited high generalization capabilities due to the heuristically distributed training parameters and included spectral features.

Although the QoT and failures in optical networks are closely connected, designing an QoT estimator which can reliably react to failures is hard. Thus, recent work was focused on finding optimal solutions for detection, identifying and localizing failures in networks which degrade the QoT, i.e., soft-failures. Such a degradation in the QoT can arise from aging network components and can eventually lead to service disruptive hard-failures. Recently, various solutions for detection, identification and localization of soft-failures have been proposed including QoT metric based threshold systems, single-channel spectral analysis or extensive analysis of SDN specific monitoring data. However, those solutions rely on extensive monitoring data from each channel being available, which might be not applicable in a real-world scenario.

The usage of OSAs and the optical spectrum for soft-failure management was investigated based on experimentally emulated component failures. The emulated soft-failures were an increase in noise figure of the EDFA, a shift in the transmit laser frequency, a change in transmit laser power, tightening of the optical filter, and a shift in the filter characteristics. It has been shown that those causes can be detected, identified and localized using ML algorithms running on the optical spectrum. For the detection of faults the well-established usage of AEs has been used in the extension of a VAE showing a high accuracy. Several ML algorithms have been experimentally compared for their identification and localization capabilities. The combination of a VAE with an NN and their two-step training approach reached the best performance with F1-scores of 0.9973 for identification and 0.9982 for localization. Those investigations were carried out with all training data being available. In practice, however, data of soft-failures is very rare, which motivates the development of a solution being able to react accurately to soft-failures based on low amounts of training data. For this purpose, a GAN is integrated in the soft-failure management framework enabled by the generative capabilities of the VAE. Using advanced training mechanisms for GANs even unknown spectra were identified by the framework. The VAE-GAN approach reached outstanding soft-failure identifica-

tion performance on only 6% of the training data with an F1-score of 0.9912. For the localization the aforementioned VAE-NN hybrid reaches F1-scores above 0.8 at only 6% training data. Besides reaching a high accuracy, the proposed solutions all suffer from high complexity compared to low-complexity ML solutions like the SVM based methods.

Low complexity methods should be the future of research in both QoT estimation and soft-failure management. A hybrid approach combining analytical and ML solutions without relying on high amounts of training data could be a potential option. Additionally, a small amount of training data is crucial in this regard. Combinations of generative ML with well-established ML algorithms offer the best opportunity to achieve this objective. Network-wide operation of algorithms is desirable for the implementation of federated learning or transfer learning approaches. These approaches also facilitate the utilization of the same ML structure in multi-vendor networks, where component characteristics vary by domain. Furthermore, accurately modeling optical components in simulations can lead to even more precise machine learning algorithms by providing more realistic training conditions, especially for managing soft-failures where failures are rare and data is sparse.

Considering the growing field of ML algorithms in optical communications, the developed frameworks and their parts can pave the way towards highly accurate zero-touch self-management of future networks as long as the monitoring data from desirably low-cost OSAs is available.

Bibliography

[1] Cisco, "Cisco annual internet report (2018–2023) white paper." `https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf`, 2020. Online; Accessed: 22.11.2023.

[2] D. Rafique, T. Szyrkowiec, H. Grießer, A. Autenrieth, and J.-P. Elbers, "Cognitive assurance architecture for optical network fault management," *Journal of Lightwave Technology*, vol. 36, no. 7, pp. 1443–1450, 2017.

[3] Y. Pointurier, "Design of low-margin optical networks," *Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A9–A17, 2017.

[4] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B60–B71, 2021.

[5] J. Lu, G. Zhou, Q. Fan, D. Zeng, C. Guo, L. Lu, J. Li, C. Xie, C. Lu, F. N. Khan, *et al.*, "Performance comparisons between machine learning and analytical models for quality of transmission estimation in wavelength-division-multiplexed systems," *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B35–B44, 2021.

[6] K. Abdelli, H. Grießer, P. Ehrle, C. Tropschug, and S. Pachnicke, "Reflective fiber fault detection and characterization using long short-term memory," *Journal of Optical Communications and Networking*, vol. 13, no. 10, pp. E32–E41, 2021.

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

[8] D. Rafique and L. Velasco, "Machine learning for network automation: overview, architecture, and applications [invited tutorial]," *Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D126–D143, 2018.

[9] F. Musumeci, C. Rottondi, G. Corani, S. Shahkarami, F. Cugini, and M. Tornatore, "A tutorial on machine learning for failure management in optical networks," *Journal of Lightwave Technology*, vol. 37, no. 16, pp. 4125–4139, 2019.

[10] M. Furdek, C. Natalino, A. Di Giglio, and M. Schiano, "Optical network security management: requirements, architecture, and efficient machine learning models for detection of evolving threats," *Journal of Optical Communications and Networking*, vol. 13, no. 2, pp. A144–A155, 2021.

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

[12] K. Sun, Z. Yu, L. Shu, Z. Wan, and K. Xu, "Generalized soft failure identification enabled by digital residual spectrum and autoencoder," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, IEEE, 2021.

[13] H. Lun, M. Fu, X. Liu, Y. Wu, L. Yi, W. Hu, and Q. Zhuge, "Soft failure identification for long-haul optical communication systems based on one-dimensional convolutional neural network," *Journal of Lightwave Technology*, vol. 38, no. 11, pp. 2992–2999, 2020.

[14] C. Zhang, D. Wang, J. Jia, L. Wang, K. Chen, L. Guan, Z. Liu, Z. Zhang, X. Chen, and M. Zhang, "Potential failure cause identification for optical networks using deep learning with an attention mechanism," *Journal of Optical Communications and Networking*, vol. 14, no. 2, pp. A122–A133, 2022.

[15] K. S. Mayer, J. A. Soares, R. P. Pinto, C. E. Rothenberg, D. S. Arantes, and D. A. Mello, "Machine-learning-based soft-failure localization with partial software-defined networking telemetry," *Journal of Optical Communications and Networking*, vol. 13, no. 10, pp. E122–E131, 2021.

[16] K. Sun, Z. Yu, L. Shu, Z. Wan, H. Huang, and K. Xu, "Experimental demonstration of soft failure identification based on digital residual spectrum and machine learning," in *2021 International Conference on UK-China Emerging Technologies (UCET)*, pp. 81–84, IEEE, 2021.

[17] S. Aladin, A. V. S. Tran, S. Allogba, and C. Tremblay, "Quality of transmission estimation and short-term performance forecast of lightpaths," *Journal of Lightwave Technology*, vol. 38, no. 10, pp. 2807–2814, 2020.

[18] J. Müller, S. K. Patri, T. Fehenberger, H. Griesser, J.-P. Elbers, and C. Mas-Machuca, "Qot estimation using egn-assisted machine learning for multi-period network planning," *Journal of Optical Communications and Networking*, vol. 14, no. 12, pp. 1010–1019, 2022.

[19] A. D'Amico, E. London, B. Le Guyader, F. Frank, E. Le Rouzic, E. Pincemin, N. Brochier, and V. Curri, "Experimental validation of gnpy in a multi-vendor flex-grid flex-rate wdm optical transport scenario," *Journal of Optical Communications and Networking*, vol. 14, no. 3, pp. 79–88, 2022.

[20] L. E. Kruse, S. Kühl, and S. Pachnicke, "Exact component parameter agnostic qot estimation using spectral data-driven lstm in optical networks," in *Optical Fiber Communication Conference*, pp. Th1C–1, Optica Publishing Group, 2022.

[21] L. E. Kruse, S. Kühl, A. Dochhan, and S. Pachnicke, "Experimental investigation of spectral data enhanced QoT estimation," *Journal of Lightwave Technology*, vol. 41, no. 18, pp. 5885–5894, 2023.

[22] L. E. Kruse and S. Pachnicke, "Joint qot estimation and soft-failure localization using variational autoencoder," in *2023 International Conference on Optical Network Design and Modeling (ONDM)*, pp. 1–3, IEEE, 2023.

[23] L. E. Kruse, S. Kühl, A. Dochhan, and S. Pachnicke, "Experimental validation of machine learning-based joint failure management and quality of transmission estimation," *IEEE Photonics Journal*, vol. 15, no. 6, pp. 1–9, 2023.

[24] J. Shao, X. Liang, and S. Kumar, "Comparison of split-step fourier schemes for simulating fiber optic communication systems," *IEEE Photonics Journal*, vol. 6, no. 4, pp. 1–15, 2014.

[25] P. Poggiolini, "The gn model of non-linear propagation in uncompensated coherent optical systems," *Journal of Lightwave Technology*, vol. 30, no. 24, pp. 3857–3879, 2012.

[26] V. Curri, "Gnpy model of the physical layer for open and disaggregated optical networking," *Journal of Optical Communications and Networking*, vol. 14, no. 6, pp. C92–C104, 2022.

[27] Q. Zhuge, X. Zeng, H. Lun, M. Cai, X. Liu, L. Yi, and W. Hu, "Application of machine learning in fiber nonlinearity modeling and monitoring for elastic optical networks," *Journal of Lightwave Technology*, vol. 37, no. 13, pp. 3055–3063, 2019.

[28] C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *Journal of Optical Communications and Networking*, vol. 10, no. 2, pp. A286–A297, 2018.

[29] I. Sartzetakis, K. K. Christodoulopoulos, and E. M. Varvarigos, "Accurate quality of transmission estimation with machine learning," *Journal of Optical Communications and Networking*, vol. 11, no. 3, pp. 140–150, 2019.

[30] E. Seve, J. Pesic, and Y. Pointurier, "Associating machine-learning and analytical models for quality of transmission estimation: combining the best of both worlds," *Journal of Optical Communications and Networking*, vol. 13, no. 6, pp. C21–C30, 2021.

[31] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *Journal of Lightwave Technology*, vol. 34, no. 1, pp. 157–179, 2015.

[32] M. Seimetz, *High-order modulation for optical fiber transmission*, vol. 143. Springer, 2009.

[33] S. Ohlendorf, *Experimental demonstration of flexible modulation formats for optical data center interconnects.* Shaker Verlag, 2021.

[34] R. Koch, *Application of Machine Learning in Optical Communications.* PhD thesis, Kiel University, 2022.

[35] P. A. Höher, *Grundlagen der digitalen Informationsübertragung.* Springer, 2011.

[36] R. Rath, *Investigation of digital signal processing techniques for compensation of linear and nonlinear impairments in fiber-optic communication systems.* Shaker Verlag, 2020.

[37] S. Pachnicke, *Fiber-optic transmission networks: efficient design and dynamic operation.* Springer Science & Business Media, 2011.

[38] G. P. Agrawal, *Fiber-optic communication systems.* John Wiley & Sons, 2012.

[39] E. Säckinger, *Broadband circuits for optical fiber communication.* John Wiley & Sons, 2005.

[40] G. P. Agrawal, "Nonlinear fiber optics," in *Nonlinear Science at the Dawn of the 21st Century*, pp. 195–211, Springer, 2000.

[41] K. Hirabayashi, H. Tsuda, and T. Kurokawa, "Tunable liquid-crystal fabry-perot interferometer filter for wavelength-division multiplexing communication systems," *Journal of Lightwave Technology*, vol. 11, no. 12, pp. 2033–2043, 1993.

[42] M. Born and E. Wolf, "Principles of optics," *Cambridge University Press*, vol. 7, 1999.

[43] J. Stone and L. Stulz, "Pigtailed high-finesse tunable fibre fabry-perot interferometers with large, medium and small free spectral ranges," *Electronics letters*, vol. 23, no. 15, pp. 781–783, 1987.

[44] A. Sneh and K. M. Johnson, "High-speed continuously tunable liquid crystal filter for wdm networks," *Journal of Lightwave Technology*, vol. 14, no. 6, pp. 1067–1080, 1996.

[45] H. Tsang, M. W. Mak, L. Chan, J. Soole, C. Youtsey, and I. Adesida, "Etched cavity ingaasp/inp waveguide fabry-perot filter tunable by current injection," *Journal of Lightwave Technology*, vol. 17, no. 10, p. 1890, 1999.

[46] J. Ciosek, "Narrow-band interference filters with unconventional spacer layers," *Applied Optics*, vol. 39, no. 1, pp. 135–140, 2000.

[47] Y. Hibino, F. Hanawa, H. Nakagome, M. Ishii, and N. Takato, "High reliability optical splitters composed of silica-based planar lightwave circuits," *Journal of Lightwave Technology*, vol. 13, no. 8, pp. 1728–1735, 1995.

[48] S. Mino, K. Yoshino, Y. Yamada, T. Terui, M. Yasu, and K. Moriwaki, "Planar lightwave circuit platform with coplanar waveguide for opto-electronic hybrid integration," *Journal of Lightwave Technology*, vol. 13, no. 12, pp. 2320–2326, 1995.

[49] M. Kawachi, "Recent progress in silica-based planar lightwave circuits on silicon," *IEE Proceedings-Optoelectronics*, vol. 143, no. 5, pp. 257–262, 1996.

[50] K. Okamoto, "Recent progress of integrated optics planar lightwave circuits," *Optical and Quantum Electronics*, vol. 31, no. 2, pp. 107–129, 1999.

[51] T. Numai, S. Murata, and I. Mito, "1.5 μm tunable wavelength filter using a phase-shift-controlled distributed feedback laser diode with a wide tuning range and a high constant gain," *Applied physics letters*, vol. 54, no. 19, pp. 1859–1860, 1989.

[52] G. P. Agrawal and S. Radic, "Phase-shifted fiber bragg gratings and their application for wavelength demultiplexing," *IEEE photonics technology letters*, vol. 6, no. 8, pp. 995–997, 1994.

[53] R. Kashyap, *Fiber bragg gratings*. Academic press, 2009.

[54] G. Town, K. Sugden, J. Williams, I. Bennion, and S. Poole, "Wide-band fabry-perot-like filters in optical fiber," *IEEE Photonics Technology Letters*, vol. 7, no. 1, pp. 78–80, 1995.

[55] F. Bilodeau, K. Hill, B. Malo, D. Johnson, and J. Albert, "High-return-loss narrowband all-fiber bandpass bragg transmission filter," *IEEE Photonics Technology Letters*, vol. 6, no. 1, pp. 80–82, 1994.

[56] P. M. Becker, A. A. Olsson, and J. R. Simpson, *Erbium-doped fiber amplifiers: fundamentals and technology*. Elsevier, 1999.

[57] M. N. Islam, "Raman amplifiers for telecommunications," *IEEE Journal of selected topics in Quantum Electronics*, vol. 8, no. 3, pp. 548–559, 2002.

[58] L. G. Kazovsky, S. Benedetto, and A. E. Willner, *Optical fiber communication systems*. Artech House Optoelectronics L, 1996.

[59] D. Zhao, L. Xi, X. Tang, W. Zhang, Y. Qiao, and X. Zhang, "Digital pilot aided carrier frequency offset estimation for coherent optical transmission systems," *Optics express*, vol. 23, no. 19, pp. 24822–24832, 2015.

[60] S. J. Savory, "Digital coherent optical receivers: Algorithms and subsystems," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1164–1179, 2010.

[61] S. Haykin, "Signaling over awgn channels," *Digital Communication Systems; Knecht, J., Ed.; John Wiley and Sons: Hoboken, NJ, USA*, pp. 323–410, 2014.

[62] I. Fatadin, S. J. Savory, and D. Ives, "Compensation of quadrature imbalance in an optical qpsk coherent receiver," *IEEE Photonics Technology Letters*, vol. 20, no. 20, pp. 1733–1735, 2008.

[63] D. Wang, C. Lu, A. P. T. Lau, and S. He, "Adaptive chromatic dispersion compensation for coherent communication systems using delay-tap sampling technique," *IEEE Photonics Technology Letters*, vol. 23, no. 14, pp. 1016–1018, 2011.

[64] M. S. Faruk and S. J. Savory, "Digital signal processing for coherent transceivers employing multilevel formats," *Journal of Lightwave Technology*, vol. 35, no. 5, pp. 1125–1141, 2017.

[65] D. Chu, "Polyphase codes with good periodic correlation properties (corresp.)," *IEEE Transactions on information theory*, vol. 18, no. 4, pp. 531–532, 1972.

[66] G. Ren, Y. Chang, H. Zhang, and H. Zhang, "Synchronization method based on a new constant envelop preamble for ofdm systems," *IEEE Transactions on Broadcasting*, vol. 51, no. 1, pp. 139–143, 2005.

[67] M. Morsy-Osman, M. Chagnon, Q. Zhuge, X. Xu, M. E. Mousa-Pasandi, Z. A. El-Sahn, and D. V. Plant, "Training symbol based channel estimation for ultra-fast polarization demultiplexing in coherent single-carrier transmission systems with m-qam constellations," in *European Conference and Exhibition on Optical Communication*, pp. Mo–1, Optica Publishing Group, 2012.

[68] M. Selmi, Y. Jaouen, and P. Ciblat, "Accurate digital frequency offset estimator for coherent polmux qam transmission systems," in *2009 35th European Conference on Optical Communication*, pp. 1–2, IEEE, 2009.

[69] E. Ip and J. M. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *Journal of Lightwave Technology*, vol. 25, no. 8, pp. 2033–2043, 2007.

[70] O. Zia-Chahabi, R. Le Bidan, M. Morvan, and C. Laot, "Efficient frequency-domain implementation of block-lms/cma fractionally spaced equalization for coherent optical communications," *IEEE Photonics Technology Letters*, vol. 23, no. 22, pp. 1697–1699, 2011.

[71] T. Pfau and R. Noé, "Phase-noise-tolerant two-stage carrier recovery concept for higher order qam formats," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1210–1216, 2009.

[72] M. S. Faruk and K. Kikuchi, "Compensation for in-phase/quadrature imbalance in coherent-receiver front end for optical quadrature amplitude modulation," *IEEE Photonics Journal*, vol. 5, no. 2, pp. 7800110–7800110, 2013.

[73] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, 2010.

[74] G. Tzimpragos, C. Kachris, I. B. Djordjevic, M. Cvijetic, D. Soudris, and I. Tomkos, "A survey on fec codes for 100 g and beyond optical networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 209–221, 2014.

[75] M. A. Sluysk, "Open roadm msa 3.01 w-port digital specification (200g-400g)." `http://openroadm.org/download.html`, 2019. Online; Accessed: 02.03.2023.

[76] R. Schmogrow, B. Nebendahl, M. Winter, A. Josten, D. Hillerkuss, S. Koenig, J. Meyer, M. Dreschmann, M. Huebner, C. Koos, *et al.*, "Error vector magnitude as a performance measure for advanced modulation formats," *IEEE Photonics Technology Letters*, vol. 24, no. 1, pp. 61–63, 2011.

[77] L. E. Kruse and S. Pachnicke, "Experimental validation of exact component parameter agnostic qot estimation using spectral data-driven lstm over a 265.2 km ssmf link," in *Photonic Networks; 23th ITG-Symposium*, pp. 1–5, VDE, 2022.

[78] A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. Le Rouzic, J. Kundrát, G. Grammel, G. Galimberti, and V. Curri, "Gnpy: an open source application for physical layer aware open optical networks," *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C31–C40, 2020.

[79] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The gn-model of fiber non-linear propagation and its applications," *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 694–721, 2013.

[80] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *Journal of Lightwave Technology*, vol. 30, no. 10, pp. 1524–1539, 2012.

[81] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "Egn model of non-linear fiber propagation," *Optics express*, vol. 22, no. 13, pp. 16335–16362, 2014.

[82] D. J. Elson, G. Saavedra, K. Shi, D. Semrau, L. Galdino, R. Killey, B. C. Thomsen, and P. Bayvel, "Investigation of bandwidth loading in optical fibre transmission using amplified spontaneous emission noise," *Optics express*, vol. 25, no. 16, pp. 19529–19537, 2017.

[83] T. Richter, J. Pan, and S. Tibuleac, "Comparison of wdm bandwidth loading using individual transponders, shaped, and flat ase noise," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, pp. 1–3, IEEE, 2018.

[84] T. J. Xia, G. A. Wellbrock, M.-F. Huang, S. Zhang, Y.-K. Huang, D.-i. Chang, S. Burtsev, W. Pelouch, E. Zak, H. de Pedro, *et al.*, "Transmission of 400g pm-16qam channels over long-haul distance with commercial all-distributed raman

amplification system and aged standard smf in field," in *OFC 2014*, pp. 1–3, IEEE, 2014.

[85] C. Mazzali and H. L. Fragnito, "Recirculating loop for experimental evaluation of edfa saturated regime effects on optical communication systems," *IEEE transactions on microwave theory and techniques*, vol. 46, no. 3, pp. 253–257, 1998.

[86] S. Savory, G. Gavioli, V. Mikhailov, R. Killey, and P. Bayvel, "Ultra long-haul qpsk transmission using a digital coherent receiver," in *2007 Digest of the IEEE/LEOS Summer Topical Meetings*, pp. 13–14, IEEE, 2007.

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

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

[89] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[90] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[91] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.

[92] T. Mitchell, *Machine Learning*. McGraw Hill, 1997.

[93] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[94] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[95] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[96] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[97] B. Schölkopf, *Support vector learning*. PhD thesis, Oldenbourg München, Germany, 1997.

[98] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[99] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[100] B. De Ville, "Decision trees," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 6, pp. 448–455, 2013.

[101] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University*, 2012.

[102] S. B. Kotsiantis, "Decision trees: a recent overview," *Artificial Intelligence Review*, vol. 39, pp. 261–283, 2013.

[103] J. H. Friedman, "Stochastic gradient boosting," *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[104] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers of Computer Science*, vol. 14, pp. 241–258, 2020.

[105] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018.

[106] Z.-H. Zhou and Z.-H. Zhou, *Ensemble learning*. Springer, 2021.

[107] M. H. Sazli, "A brief review of feed-forward neural networks," *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, vol. 50, no. 01, 2006.

[108] C. M. Bishop *et al.*, *Neural networks for pattern recognition*. Oxford university press, 1995.

[109] M. Schädler, *Machine learning in digital signal processing for optical transmission systems*. PhD thesis, Kiel University, 2022.

[110] A. Graves and A. Graves, *Supervised sequence labelling*. Springer, 2012.

[111] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," *arXiv preprint arXiv:1801.01078*, 2017.

[112] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[113] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[114] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Gated feedback recurrent neural networks," in *International conference on machine learning*, pp. 2067–2075, PMLR, 2015.

[115] K. Abdelli, H. Grießer, and S. Pachnicke, "A machine learning-based framework for predictive maintenance of semiconductor laser for optical communication," *Journal of Lightwave Technology*, vol. 40, no. 14, pp. 4698–4708, 2022.

[116] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[117] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "Dbscan revisited, revisited: why and how you should (still) use dbscan," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.

[118] T. N. Tran, K. Drab, and M. Daszykowski, "Revised dbscan algorithm to cluster data with dense adjacent clusters," *Chemometrics and Intelligent Laboratory Systems*, vol. 120, pp. 92–96, 2013.

[119] D. P. Kingma, M. Welling, *et al.*, "An introduction to variational autoencoders," *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.

[120] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[121] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE signal processing magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[122] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, pp. 214–223, PMLR, 2017.

[123] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[124] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.

[125] M. J. Kearns, *The computational complexity of machine learning*. MIT press, 1990.

[126] S. Arora and B. Barak, *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[127] V. Pappakrishna, "Computational complexity of machine learning algorithms." `https://medium.com/datadailyread/computational-complexity-of-machine-learning-algorithms-16e7ffcafa7d`, 2022. Online; Accessed: 06.07.2023.

[128] P. Audenaert, "Cost266 network topology." `http://sndlib.zib.de/`, 2005. Online; Accessed: 18.09.2021.

[129] J. Leibrich, *Digitale Modulationsverfahren und Signalverarbeitungsalgorithmen für flexible optische Datenübertragung*. Shaker Verlag, 2018.

[130] L. E. Kruse, S. Kühl, and S. Pachnicke, "Generalizable qot estimation based on spectral data driven lstm in exact component parameter agnostic networks," in *2022 European Conference on Optical Communication (ECOC)*, pp. 1–4, IEEE, 2022.

[131] D. Wang, C. Zhang, W. Chen, H. Yang, M. Zhang, and A. P. T. Lau, "A review of machine learning-based failure management in optical networks," *Science China Information Sciences*, vol. 65, no. 11, p. 211302, 2022.

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

[133] L. Shu, Z. Yu, Z. Wan, J. Zhang, S. Hu, and K. Xu, "Dual-stage soft failure detection and identification for low-margin elastic optical network by exploiting digital spectrum information," *Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2669–2679, 2019.

[134] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special lecture on IE*, vol. 2, no. 1, pp. 1–18, 2015.

[135] L. E. Kruse, S. Kühl, A. Dochhan, and S. Pachnicke, "Experimental investigation of machine-learning-based soft-failure management using the optical spectrum," *Journal of Optical Communications and Networking*, vol. 16, no. 2, pp. 94–103, 2024.

[136] L. E. Kruse, S. Kühl, A. Dochhan, and S. Pachnicke, "Experimental demonstration of soft-failure management using variational autoencoder and gan on optical spectrum," in *2023 European Conference on Optical Communication (ECOC)*, pp. 1–4, IEEE, 2023.

[137] L. E. Kruse, S. Kühl, A. Dochhan, and S. Pachnicke, "Monitoring data augmentation of spectral information using vae and gan for soft-failure identification," in *Optical Fiber Communication Conference*, pp. M3I–4, Optica Publishing Group, 2024.

[138] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, 2016.

List of Abbreviations

ADC	analog-to-digital converter
Adam	adaptive moment
AI	artificial intelligence
AE	autoencoder
AOM	acousto-optic modulator
ASK	amplitude shift keying
ASE	amplified spontaneous emission
AWG	arbitrary waveform generator
AWGN	additive white Gaussian noise
ANN	artificial neural network
BATi	Boston Applied Technologies, Inc.
BER	bit error ratio
B2B	back-to-back
BPS	blind phase search
BW	bandwidth
CAZAC	constant amplitude zero auto-correlation
CD	chromatic dispersion
CFO	carrier frequency offset
CLT	classification tree
CNN	convolutional neural network
COI	channel of interest

CW continuous wave

DB database

DP dual-polarization

DAC digital-to-analog converter

DD direct-detection

DBSCAN density-based spatial clustering of applications with noise

DFE decision-feedback equalization

DML directly modulated laser

DSP digital signal processing

DTR decision tree regressor

EAM electro-absorption modulator

EDC electronic dispersion compensation

EDFA Erbium-doped fiber amplifier

EML electro-absorption modulated laser

ENOB effective number of bits

E/O electrical-to-optical

EQ equalizer

FEC forward error correction

FFE feed-forward equalization

FFT fast Fourier transform

FIR finite impulse response

FP Fabry-Perot

FP false positive

FN false negative

FWM four-wave mixing

FF-NN feed-forward neural network

GN Gaussian noise

GAN generative adversarial network

GNRF Gaussian noise model reference formula

GOSNR generalized optical signal to noise ratio

GRU gated recurrent unit

GSOP Gram-Schmidt orthogonalization procedure

HD hard decision

IFFT inverse fast Fourier transform

IIR infinite impulse response

IQ inphase-quadrature

IM intensity modulation

IM/DD intensity modulation/direct detection

ISI inter-symbol interference

k-NN k-nearest-neighbor

LCLF linear classifier

LMS least mean squares

LO local oscillator

LLM large-language model

LSTM long-short term memory

MAE mean absolute error

MIMO multiple input multiple output

ML machine learning

MLP multilayer perceptron

MMSE minimum mean square error

MSE mean square error

MZI Mach-Zehnder interferometer

MZM Mach-Zehnder modulator

NLI nonlinear interference

NLSE nonlinear Schrödinger equation

NN neural network

O/E optical-to-electrical

OOK on-off keying

OPM optical performance monitoring

OSNR optical signal-to-noise ratio

OSA optical spectrum analyzer

OADM optical add and dropp multiplexer

OTDR optical time-domain reflectometry

PAM pulse amplitude modulation

PBC polarization beam combiner

PBS polarization beam splitter

PD photodiode

PN phase noise

PMD polarization mode dispersion

PNC phase noise compensation

PolMux polarization multiplex

PM power meter

PS polarization scrambler

PRBS pseudo-random binary sequence

PRMS pseudo-random multilevel sequence

PSD power spectral density

PSK phase shift keying

QAM quadrature amplitude modulation

QPSK quadrature phase shift keying

QoT quality of transmission

RF radio frequency

RFC random forest classifier

ROF roll-off factor

ROP received optical power

ROADM reconfigurable optical add-drop multiplexer

RC raised cosine

RRC root-raised cosine

RBF radial basis function

RNN recurrent neural network

RTT round trip time

SBS stimulated Brillouin scattering

SD soft decision

SGD stochastic gradient descent

SDN software defined networking

SFD soft-failure detection

SFI soft-failure identification

SFL soft-failure localization

SNR signal-to-noise ratio

SOP state of polarization

SOA semiconductor optical amplifier

SPM self-phase modulation

SpS samples per symbol

SRS stimulated Raman scattering

SSB single-sideband

SSMF standard single mode fiber

SSFM split-step Fourier method

SVC support vector machine classifier

SVM support vector machine

SVR support vector regressor

TN true negative

TP true positive

TS training symbols

VAE variational autoencoder

VOA variable optical attenuator

WDM wavelength division multiplexing

WSS wavelength selective switch

XGB XGradientBoost

XPM cross-phase modulation

List of Symbols

A_{S}	amplitude of the optical signal
A_{eff}	effective area
A	optical envelope
B_{opt}	optical bandwidth
B_{ref}	reference bandwidth
C	regularization parameter in an SVM
D	dispersion parameter
E_{IQ}	correlation coefficient of the received signal
E_{I}	real part of the received signal
E_{Q}	imaginary part of the received signal
E_{S}	electrical field of the optical signal
F_D	cost function for dispersion compensation
F_{MSE}	cost function for the MSE
F_{N}	noise figure
F	PSD in the GN-model
G	gain
H_{MZM}	transfer function of the Mach-Zehnder modulator
H_{SSMF}	transfer function of the standard single mode fiber
L_{eff}	effective length
L_{s}	span length
L	length
M	order of the symbol constellation
N_{ASE}	noise power spectral density
N_{FFT}	FFT length
N_{S}	number of bits/sample
N_{TS}	number of training symbols

N	length of the data sequence
P_{ASE}	ASE noise power
P_{NLI}	nonlinear interference power
P_{N}	noise power
P_{Rx}	received power
P_{S}	signal power
P_{opt}	optical power
P_{LO}	local oscillator power
P_{in}	input power
P_{out}	output power
P_{sig}	power of the signal
P	power
Q	quality factor
R_{B}	bit rate
R_{S}	symbol rate
R_{PD}	photodiode responsitivity
S	dispersion slope
Φ_{LO}	phase of local oscillator
Φ_{sig}	phase of the signal
α_{RRC}	roll-off factor
α	attenuation coefficient
β	phase constant of dispersion
γ	nonlinear coefficient
λ	wavelength
$\mathbf{R}_{\mathrm{c}}$	compensation matrix
$\mathbf{R}_{xx}$	auto-correlation matrix of x
$\mathbf{R}$	rotation angle of the state of polarization
$\mathbf{W}$	weight matrix
$\mathbf{h}$	filter coefficients in vector notation
$\mathbf{r}_{xd}$	cross-correlation vector of x and d
$\mathbf{x}_{\mathrm{in}}$	input signal of the X-polarization of the equalizer in vector notation
$\mathbf{x}_{\mathrm{out}}$	output signal of the X-polarization of the equalizer in vector notation

$\mathbf{x}$	input vector
$\mathbf{y}_{\text{in}}$	input signal of the Y-polarization of the equalizer in vector notation
$\mathbf{y}_{\text{out}}$	output signal of the Y-polarization of the equalizer in vector notation
ω	angular frequency
ϕ_{p}	phase shift
ϕ_{r}	rotation angle of the polarizations
σ	standard deviation
h	Planck constant
$\varphi_{\text{NL,max}}$	maximum nonlinear phase shift
a_s	input s from the connected layer
b	bias
c	speed of light in vacuum
$d_{(i)}$	Euclidean distance of the ith sample
d	data symbol
f_{ADC}	sampling rate of the ADC
f_{DAC}	sampling rate of the DAC
f_{S}	signal frequency
f_{c}	carrier frequency
f	frequency
g	activation function of neural network layer
i	electrical current
l	SSFM step width
n_{sp}	spontaneous emission factor
n	refractive index
s	signal
t	time
u_{D}	modulator drive voltage
u_{bias}	modulator bias voltage
u	electrical voltage
w_s	neural network weights
$x_{(i)}$	ith sample
z	location

List of Figures

List of Tables