Machine Learning in Digital Signal Processing for Optical Transmission Systems

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Ingenieurswissenschaften (Dr.-Ing.)
der Christian-Albrechts-Universität

vorgelegt von

Maximilian Olivier Rainer Schädler

aus

Vaduz, Liechtenstein

Kiel 2022

1. Gutachter:	Prof. Dr.-Ing. Dipl.-Wirt. Ing. Stephan Pachnicke
2. Gutachter:	Prof. Dr.-Ing. Laurent Schmalen
3. Gutachter:	Prof. Dr.-Ing. Gerhard Schmidt
Tag der mündlichen Prüfung:	09. August 2022

Acknowledgements

This thesis was written during my time as researcher in the coherent optical communication group at the Huawei Munich Research Center.

Primarily, I would like to express my gratitude to my supervisor at CAU Kiel, Prof. Stephan Pachnicke, for giving me the opportunity to pursue this PhD, and together with my supervisor at Huawei, Dr. Maxim Kuschnerov, for the guidance, encouragement and many fruitful discussions over the last four years.

There are many people at Huawei who deserve my gratitude for all their support and discussions. Working with them has not only lead to several publications but also to friendships. My deepest gratitude goes to Dr. Maxim Kuschnerov, Dr. Fabio Pittlà and Dr. Christian Blümm for supporting me at my first steps as a researcher in fiber-optic communications. Moreover, I especially would like to thank Dr. Stefano Calabrò and Dr. Georg Böcherer for supporting me with their experience and expertise. Their great reviews and comments have helped to improve the quality of all papers dramatically.

The final words are for my family. Without your continuous support, neither my diploma nor this thesis would have been possible. I am deeply grateful that you are continuously encouraging me and that you always have my back. I cannot thank you enough for this.

Abstract

The future demand for digital information will exceed the capabilities of current optical communication systems, which are approaching their limits due to component and fiber intrinsic non-linear effects. Machine learning methods are promising to find new ways of leverage the available resources and to explore new solutions. Although, some of the machine learning methods such as adaptive non-linear filtering and probabilistic modeling are not novel in the field of telecommunication, enhanced powerful architecture designs together with increasing computing power make it possible to tackle more complex problems today.

The methods presented in this work apply machine learning on optical communication systems with two main contributions. First, an unsupervised learning algorithm with embedded additive white Gaussian noise (AWGN) channel and appropriate power constraint is trained end-to-end, learning a geometric constellation shape for lowest bit-error rates over amplified and unamplified links. In simulation and experimental studies the learned constellations improved performance compared to conventional constellation schemes. Second, supervised machine learning methods, especially deep neural networks with and without internal cyclical connections, are investigated to combat linear and non-linear inter-symbol interference (ISI) as well as colored noise effects introduced by the components and the fiber. On high-bandwidth coherent optical transmission setups their performances and complexities are experimentally evaluated and benchmarked against conventional digital signal processing (DSP) approaches. The results indicate that feed-forward deep neural networks (without internal cyclical connections) proved to reflect systematic non-linearities more accurately than common 5th-order Volterra non-linear equalizers (VNLEs). They either outperformed pruned VNLEs with equal complexity or achieved the same performance with less multipliers and hence lower complexity. In comparison to feed-forward architectures, recurrent architectures (with internal cyclical connections) compensated next to non-linear ISI as well as colored noise effects. They either matched or outperformed the reference DSP approaches, consisting of VNLEs accompanied symbol-spaced whitening filters and BCJR detectors. While feed-forward architectures are overall a promising candidate to replace VNLEs, the performance or respectively the complexity benefit of recurrent architectures is strongly related to the considered scenario.

This thesis shows how machine learning can be applied to optical communication systems. In particular, it is demonstrated that machine learning is a viable designing and DSP tool to increase the capabilities of optical communication systems.

Zusammenfassung

Die künftige Nachfrage nach digitalen Informationen wird die Kapazität der derzeitigen optischen Kommunikationssysteme übersteigen, die aufgrund der nichtlinearen Effekte von Komponenten und Fasern an ihre Grenzen stoßen. Methoden des maschinellen Lernens sind hier ein vielversprechender Ansatz, neue Wege und Lösungen zu finden, um die verfügbaren Ressourcen besser auszunutzen und die limitierenden Faktoren zu kompensieren. Obwohl einige dieser Methoden im Bereich der Telekommunikation nicht neu sind, wie zum Beispiel die adaptive nichtlineare Filterung und die probabilistische Modellierung, ermöglichen heutzutage neue leistungsstarke Architekturen in Verbindung mit zunehmender Rechenleistung die Lösung von Problemen höherer Komplexität.

Die in dieser Arbeit vorgestellten Methoden wenden maschinelles Lernen auf optische Kommunikationssysteme an und untersucht dabei zwei Hauptthemen. Zuerst wird das Erlernen von optimalen Konstellationsdiagrammen über eine verstärkte und unverstärkte Verbindung betrachtet. Dabei wird ein unüberwachter Lernalgorithmus mit eingebettetem AWGN-Kanal und entsprechender Leistungsbeschränkung Ende-zu-Ende trainiert. In Simulationen und experimentellen Studien zeigen die erlernten Konstellationen eine bessere Leistung als herkömmliche. Anschließend werden Methoden des überwachten maschinellen Lernens, insbesondere tiefe neuronale Netze mit und ohne interne zyklische Verbindungen, untersucht, um lineare und nichtlineare Intersymbolinterferenzen sowie farbige Rauscheffekte von den Komponenten und der Faser zu kompensieren. In kohärenten optischen Übertragungssystemen mit hoher Bandbreite werden die Leistung und Komplexität dieser neuronale Netze experimentell evaluiert und mit konventionellen DSP Ansätzen verglichen. Die Ergebnisse zeigen, dass vorwärtsgerichtete neuronale Netze (ohne interne zyklische Verbindungen) systematische Nichtlinearitäten genauer widerspiegeln als herkömmliche VNLEs. Entweder wurde bessere Leistung bei gleicher Komplexität oder gleiche Leistung mit geringerer Komplexität durch weniger Multiplikatoren erreicht. Im Vergleich zu vorwärtsgerichtete Architekturen kompensierten rekurrente Architekturen (mit internen zyklischen Verbindungen) neben der nichtlinearen Intersymbol-Interferenz auch die Auswirkungen von farbigem Rauschen. Die Referenzansätze, bestehend aus VNLEs mit Whitening-Filtern und BCJR-Detektoren, konnten reproduziert oder übertroffen werden. Während vorwärtsgerichtete Architekturen insgesamt vielversprechende Kandidaten sind, um VNLEs zu ersetzen, ist der Leistungs- beziehungsweise Komplexitätsvorteil von rekurrenten Architekturen stark vom betrachteten Szenario abhängig.

In dieser Arbeit wird die Anwendung maschinellen Lernens auf optische Kommunikationssysteme aufgezeigt. Insbesondere wird darauf eingegangen, dass maschinelles Lernen ein praktikables Werkzeug für den Entwurf und die digitale Signalverarbeitung ist, um die Kapazität optischer Kommunikationssysteme zu erhöhen.

Contents

Chapter 1

Introduction

The globalization of our society is largely attributed to the progress in communication technology [1]. In particular, the internet has evolved from a small network between research institutions to a global infrastructure with an increasing number of services and ever-growing demand on data rate. The number of network endpoints has thereby increased from a few millions in the 1990s to billions nowadays [2]. To support this demand of internet traffic, the internet backbones rely on fiber-optic communications technology. Optical fiber links offer, in contrast to wireless or even wired electrical links, low attenuation over long distances and a huge amount of available bandwidth.

In the past decades, the transmission capacity over optical fiber links has been enhanced by multiple technological breakthroughs [3]. In 1987, optical amplifiers, such as Erbium-doped fiber amplifiers (EDFAs), replaced electronic regeneration schemes as repeaters, which enabled pure optical and cost effective wavelength division multiplexing (WDM) systems. Furthermore, in 2008, coherent transceiver technology enabled access to both quadrature and polarization of each optical carrier, which opened up an additional dimension to transfer information. In combination with digital signal processing (DSP) and the increasing digital-to-analog converter (DAC) speed, state-of-the-art commercial coherent systems can reach transmission capacities of 400 Gbit/s per carrier.

However, as mentioned above, consumer demand for digital information is growing exponentially. In 2023 nearly two-thirds of the global population will have internet access and the number of devices connected to networks will increase from 18.4 (2018) to 29.4 billion [4]. Hence the number of connected devices will surpass the global population by more than three times. Machine-To-Machine (M2M) and Internet of Things (IoT) interconnections are thereby the fastest growing application type. They will constitute at least half of the global traffic and will probably outpace the traffic consumed by humans in the near future. Optical

communication systems have to keep up with this pace of service demand, while maintaining low cost and high energy efficiency. It is therefore inadvisable to increase the capacity by simply installing new optical transmission links exponentially. The limited capacity has to be overcome as well by supporting higher data rates per channels, i.e., increasing the symbol rates and the spectral efficiencies per carrier. However, higher symbol rates and richer signal constellations introduce several challenges.

1.1 Motivation

In optical long-haul transmission systems (>1000km), fiber's non-linearity caused by the Kerr effect [5] is arguably the most prominent effect that limits the achievable information rates of fiber-optic communication systems [6–8]. However, in optical transmission systems, distortions do not end with the transfer characteristics of the fiber itself. Numerous optical/electrical (O/E) components introduce additional impairments, which require effective compensation to attain highest data rates. The compensation is especially important, when high-bandwidth communications meet short reach, as reduced fiber lengths turn components into the dominate source of impairments. Typical use cases include data center interconnects (DCIs) with a reach of 80 km and metro core networks with hundreds of kilometers reach.

In such optical short-reach communication systems, the high data rate demand leads to symbol rates higher than the provided 3dB-bandwidth of the O/E components. This in turn leads to inter-symbol interference (ISI) and after equalization to colored noise effects due to noise enhancement. In addition to the linear impairments, the transmitted symbols are subject to non-linear distortions with memory effects arising from the non-linear transfer characteristics of the O/E components. Today's commercial coherent optical short-reach transceivers commonly compensation linear channel impairments and phase noise effects from the laser source. Non-linear compensation and colored noise effect reduction are no standard features yet, but will inventively become a key element of DSP to keep up with the every increasing data rates.

Several DSP-based techniques to compensate non-linear fiber as well as component effects are proposed in the literature. The most popular one for non-linear fiber compensation is the digital back-propagation (DBP) algorithm [9] based on the split-step Fourier method (SSFM) which approximates the non-linear Schrödinger equation (NLSE) [10]. Another popular approach against fiber as well as component non-linerities is the Volterra non-linear equalizer (VNLE) [11, 12]. By combining linear convolution and non-linear power series the VNLE

is capable of modeling time-invariant non-linear systems with finite fading memory. However, high computational complexity allied with limited benefits due to imprecise modeling remains a challenge for practical implementation of efficient DSP-based non-linear compensation. Here, machine learning methods, which have relished increased attention after the artificial intelligence (AI) winter [13], may be capable to explore new solutions.

1.2 Overview and Outline of Contribution

Although, some of the machine learning methods such as adaptive non-linear filtering and probabilistic modeling are not novel in the field of telecommunication, new powerful architecture designs together with increasing computing power make it possible to tackle more complex problems today. For instance, machine learning methods, especially deep neural networks (DNNs), have recently demonstrated excellent performance gains for the physical layer in various applications, such as channel equalization [14–21], optical performance monitoring [22–25] and end-to-end learning [26, 27].

The object of this thesis is to build upon the aforementioned contributions and to deepen the investigation of machine learning capabilities in DSP, focusing primarily on optical short reach communication and on a fair comparison with classical DSP algorithms in terms of performance and complexity. In particular, the thesis interpolates material from three journal papers [28–30] and four conference papers [31–34] published by the author within the duration of the Doctor of Philosophy (PhD) program. Their contents are related to two main topics, demonstrating that machine learning is a viable designing as well DSP tool for increasing the capabilities of optical communication systems.

Neural Network Assisted Geometric Shaping: To achieve maximum performance, it is essential to match the modulation to the actual transmit channel conditions, i.e., for optical short reach communication a constellation shape must take amplified and unamplified scenarios into account. In Chapter 4, novel auto-encoder structures based on deep neural networks are introduced to design modulation formats optimized for lowest bit error ratios (BERs) over amplified and unamplified links. Besides the geometry of the symbol constellation, also the bit mapping with and without label extension is optimized. The learned high-order modulation formats are numerically and experimentally evaluated.

Non-linear Compensation for Optical Transmission: In Chapter 5 two DNN architecture concepts are investigated to combat linear and non-linear ISI as well as colored noise effects. The first concept relies on feed forward one-dimensional convolutional neural

networks (1D-CNNs) as well known as feed forward time delay neural networks (TDNNs). Equal to conventional linear and non-linear equalizer designs, e.g., finite impulse response (FIR) filters and VNLEs, the memory effects of the channel and components are considered by adding time delayed versions of the observed channel output. This allows the TDNN to have a finite dynamic response to time series input data and to describe causal time-invariant non-linear systems with finite fading memory. In this context, particular attention is placed on hard and soft demapping and on the corresponding cost functions.

The second concept relies on bidirectional recurrent neural networks (BRNNs). In contrast to the TDNN, the BRNN architecture enables internal memory states corresponding to infinite impulse response (IIR) filters. For each concept we describe the main rationale behind the approach and compare its performance experimentally with classical DSP approaches to combat the strong impairments mentioned in Section 1.1.

1.3 Outline of the Thesis

In Chapter 2, an application-oriented introduction to a coherent optical transmission system is provided, including the description of the fiber-optic channel, the hardware components as well as the DSP blocks to transmit respectively to recover the desired information. Particular emphasis is placed on the linear and non-linear impairments caused by the components and fiber as well as the capabilities to compensate them by DSP. Furthermore, the design paradigm of joint optimization of high-order modulation and forward error correction (FEC), known as coded modulation, is reviewed, which lays the groundwork for the subsequent assumptions and discussions. Probabilistic shaping as a powerful technique to increase transmission performance is reviewed, too.

Chapter 3 covers an introduction into machine learning, focusing on general neural networks with and without internal cyclic connections. For each design the main concept is described as well as the corresponding forward and backward pass, and training process.

Chapter 4 and Chapter 5 contain the two outlined contributions in Section 1.2, namely Neural Network Assisted Geometric Shaping and Non-linear Compensation for Optical Transmission.

Each of the abovementioned chapters contains a concluding section that highlights the main findings. The most important aspects and potential future research directions of these individual remarks are compiled in Chapter 6.

Chapter 2

Coherent Optical Transmission Systems

Digital coherent optical transceivers use polarization, amplitude and phase to encode and decode information. It provides full knowledge of the optical field and hence enables the utilization of sophisticated modulation schemes, such as M-ary phase-shift keying and quadrature amplitude modulation (QAM). Moreover, the preserved phase information after optical-to-electrical conversion allows to compensate effectively for transmission impairments by means of DSP.

A simplified block diagram of an optical coherent transmission system is depicted in Fig. 2.1. The system consists of several electronic, electro-optic and optical components, which can be grouped into a digital, an electrical and an optical layer. The digital layer contains the

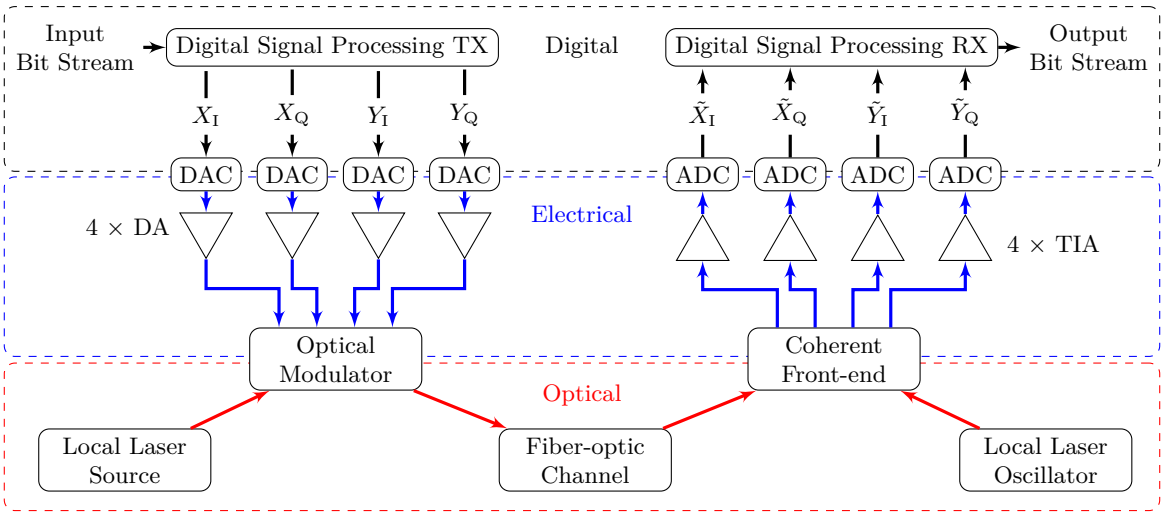


Figure 2.1: Block diagram of a coherent optical transmission system. The black dashed section denotes the digital, the blue section the electrical and the red section the optical part.

transmitter and receiver-side DSP. The electrical layer groups the DACs and driver amplifiers (DAs) at the transmitter, and the analog-to-digital converters (ADCs) and the transimpedance amplifiers (TIAs) at the receiver. The final layer is the optical level, composed of the optical modulator at the transmitter and the coherent front-end at the receiver, each with the respective laser, and finally in between the fiber-optic channel with filters, amplifiers, e.g., EDFAs, and the fiber itself. Note that, the DAC, ADC, optical modulator and the coherent front-end belong actually to more than one layer due to their task of conversion.

This chapter provides an application-oriented introduction to a coherent optical transmission system. The more important parts for the following chapters are descripted in more detail while a more comprehensive discussion on the other subjects can be found in the corresponding references. At first, the individual components of the transmitter are reviewed focusing on their linear and non-linear impairments. Secondly, next to the main deterministic linear fiber impairments like attenuation and chromatic dispersion, fiber non-linearities like self-phase modulation and cross-phase modulation are characterized. Finally, the components of the coherent receiver and the standard receiver-side DSP blocks, to recover the transmitted signal, are described.

2.1 Coherent Optical Transmitter

The role of the coherent optical transmitter is to modulate two orthogonal polarizations (conventionally referred to as horizontal and vertical polarization) of the optical carrier with four electrical real baseband signals and to transmit the corresponding modulated optical signal over the fiber. The data X_{I}, X_{Q} for the horizontal polarization and Y_{I}, Y_{Q} for the vertical polarization are thereby separately modulated and combined by a polarization beam combiner (PBC) to yield the polarization-multiplexed transmitted signal, as shown in Fig. 2.2.

In a real system, the amplitude and the envelope impulse response of the O/E components are not constant within their operational bandwidth, and the transmitted signal gets distorted in both amplitude and phase. This leads to strong linear channel dispersion and hence to ISI, especially in high-speed optical coherent transmission systems, where the symbol rates can be higher than the provided 3dB-bandwidth of the O/E components. In addition to these linear impairments, the transmitted symbols are subject to non-linear distortions with memory effects arising from the non-linear transfer characteristics of the O/E components.

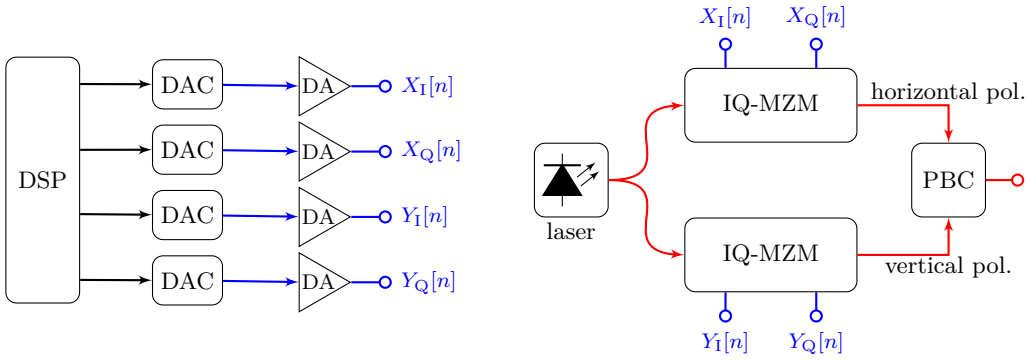


Figure 2.2: Transmitter for in-phase, quadrature and polarization multiplexing.

The sources of non-linearity may usually be characterized as memoryless. However, the non-linear O/E components are embedded in a network where linear filtering operation takes place. Consequently, the overall effect of the channel on the input signal is a non-linear mapping with memory [35]. Hence, even in a back-to-back (BtB) setup, linear and non-linear interdependencies in the received symbols after channel transmission occur.

In the following the transmitter components and their impairments are described in more detail. In case of DAC, DA and optical modulator the focus is on the origin of the non-linear behavior. Nevertheless, frequency responses of real components are shown as well.

2.1.1 Laser

In a coherent optical system a laser, acronym for light amplification of stimulated emission of radiation [36], provides the optical carrier. To enable high data rate optical transmission, the major requirements for such a laser are high optical power and narrow linewidth. The high power is required to compensate for the insertion loss of the optical modulator when high order modulations are used, while the narrow linewidth is required to keep the impact of the phase noise small. For instance, in a coherent 100-Gbit/s system [37], a linewidth of 1 MHz or less is required for both the transmitter light source and the local light source at the receiver. The effect of phase noise is caused by the random phase fluctuations in the laser cavity and can be described as a Wiener-Process [38, Eq. 2]. In a coherent modem, after reception, the carrier phase noise can be estimated and compensated via DSP by exploiting its auto-correlation. Section 2.3.2 on the receiver-side DSP chain includes a discussion on carrier phase recovery.

2.1.2 Digital-to-Analog Converter

The DACs convert the digital signal into the analog domain and have become essential elements in an optical transmission systems. Their advent has allowed the use of sophisticated DSP techniques at the transmitter to keep up with the demand of high data rates and to achieve high spectral efficiencies. Nowadays the sampling rates of high-speed DACs have reached 128 GSa/s for technologies based on Silicon-Germanium bipolar transistors [39–42] and 120 GSa/s for complementary metal-oxide semiconductor (CMOS) [41]. Today, electronic converters are a major performance limiter for high data rate communication.

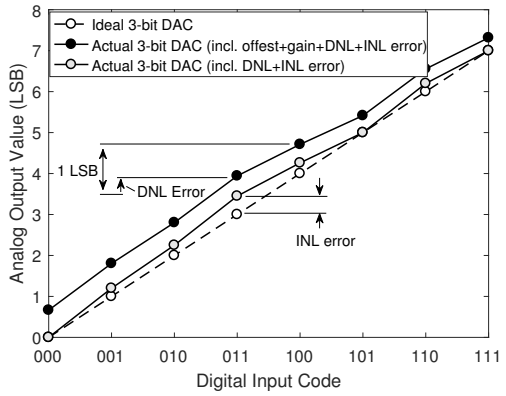


Figure 2.3: Non-linear impairment: Transfer function of a 3-bit DAC considering offset, gain and differential and integral non-linearity errors [43].

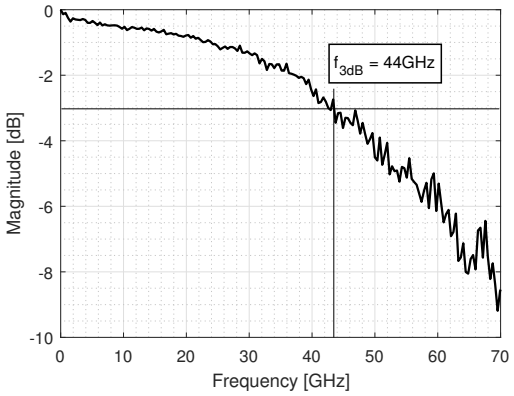


Figure 2.4: Linear impairment: Bandwidth limiting effects from the analog components lead to linear distortion, e.g. measured frequency response of Micram DAC 4 with 44 GHz 3dB bandwidth [44, Fig. bottom left p.3].

The DAC represent a limited number of discrete digital input codes by a corresponding number of discrete analog output values. The transfer function of an ideal N-bit DAC is therefore a series of 2^N equidistant discrete points as shown in Fig. 2.3. However, in practice statics distortions such as offset errors [43, Sec. 3.2], slope errors [43, Sec. 3.2], differential non-linearity (DNL) errors [43, Sec. 3.3] and integral non-linearity (INL) errors [43, Sec. 3.4] affect the conversion. While the differences in the offset and slope can be adjusted to zero by trimming, the non-equidistant discrete points due to DNL/INL cause non-linear signal distortions.

Next to the aforementioned non-linear impairments, the transmitted symbols are subject to linear distortions caused by bandwidth limitation. Fig 2.4 depicts the measured frequency response of the Micram DAC4 [44] with 44GHz 3dB-bandwidth which is used to perform the experiments in this thesis.

2.1.3 Driver Amplifier

The output amplitude of the DAC is typically limited to a range of 200 mVpp to 700 mVpp. Such voltages are too low to drive common optical modulators, which typically requiring $3 - 4$ Vpp (in future we may will see lower driving voltages). Electrical DA are therefore needed to amplify the DAC output signal and to achieve the input power required by the accompanied optical modulator.

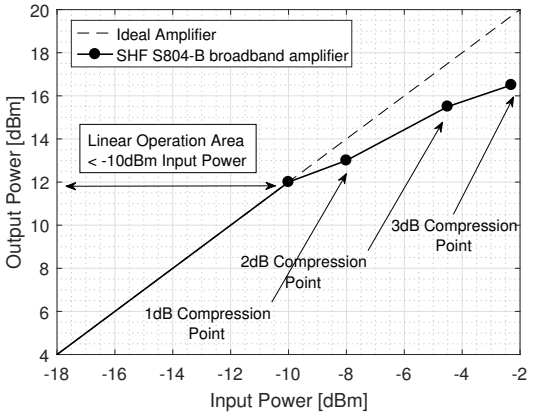


Figure 2.5: Non-linear Impairment: The gray curve represents the power transfer curve of an ideal amplifier, in which the slope is constant. The black curve represents the real SHF S804-B broadband amplifier [45], which can deliver the gain of 22dB only up to a certain input power level before it hits saturation.

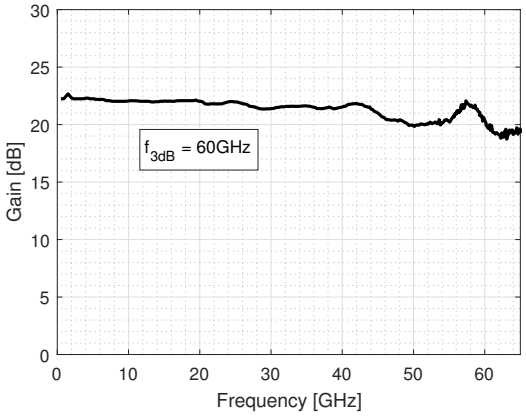


Figure 2.6: Linear Impairment: Bandwidth limiting effect leads to linear distortion, e.g. measured frequency response of a SHF S804-B broadband amplifier with 60 GHz 3dB bandwidth [45, Fig. center-left p.5].

Fig 2.5 illustrates the behavior of an ideal and real SHF S804-B broadband amplifier [45]. Above a certain input power level, the real amplifier gets into compression and the gain is actually less than for an input signal with lower power level. At a certain point, the output power will not get amplified, even if the input power is further increased, i.e., the amplifier is in saturation. To quantify this non-linear behaviour the datasheet provides the values of gain compression at 1 dB, 2 dB and 3 dB. For instance, the 1 dB compression point indicates the output power at which the output level is 1 dB less than that of an ideal (linear) amplifier. Non-linear regime operation leads to non-equidistant constellation points and also to intermodulation distortion (IMD) as discussed in [46].

In case of IMD, the spectrum at the output of the non-linear device will not only consist of the original signals but will also contain the sum and difference of the input signals along with their harmonics as shown in Fig. 2.7. While most of the harmonics can be removed

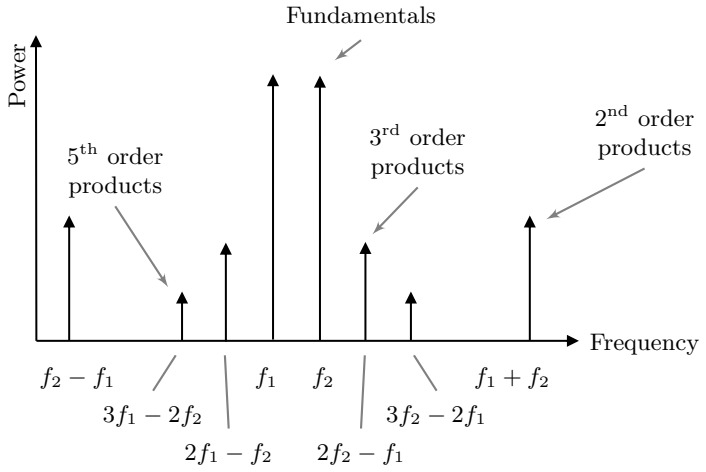


Figure 2.7: Intermodulation products and related frequency. The input test signal consists of two fundamentals frequency f_1 and f_2 while the output signal contains multiple intermodulation products when the amplifier is operated in the non-linear region.

by a filter, odd harmonics products, which fall very close to the fundamentals, will degrade the performance. In practice, the optimal operations point is a trade-off between required output power and non-linear effects.

As can be observed from Fig. 2.6, in comparison to the frequency response of the Micram DAC4 the frequency response of the SHF S804-B broadband amplifier is rather flat. Linear impairments caused by the limited bandwidth of the DA are therefore modest, even if high symbol rates are considered.

2.1.4 External Optical Modulator

The external Mach-Zehnder modulator (MZM) is an interferometric structure made from a material with strong electro-optic effect, such as Lithium niobate (LiNb03), Gallium arsenide (GaAs) or Indium phosphide (InP) [46]. Fig. 2.8 shows the general structure of an MZM. Note that, in coherent systems four of these MZMs are combined to modulate in-phase and quadrature components of both polarizations. Applying electric fields to the arms changes the optical path lengths, resulting in phase modulation. Combining two arms with opposite driving voltages $V_1(t) = V_{\text{in}}(t)$, $V_2(t) = -V_{\text{in}}(t)$ converts phase modulation into the desired intensity modulation.

The optical input $E_{\text{in}}(t)$ is split into the upper and lower modulator arms. Subsequently, the two fields are phase modulated with two phase shifters driven by the electrical signal

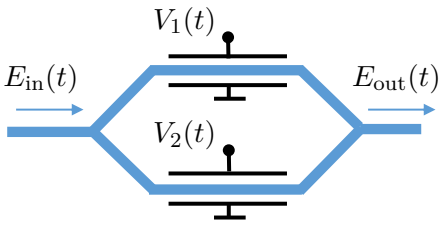


Figure 2.8: Structure of a MZM.

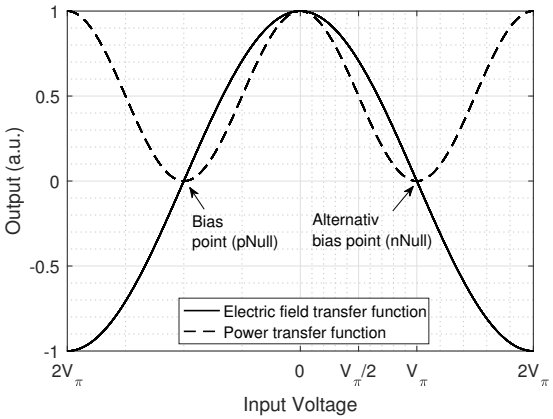


Figure 2.9: non-linear Impairment: MZM field and power non-linear transfer characteristics.

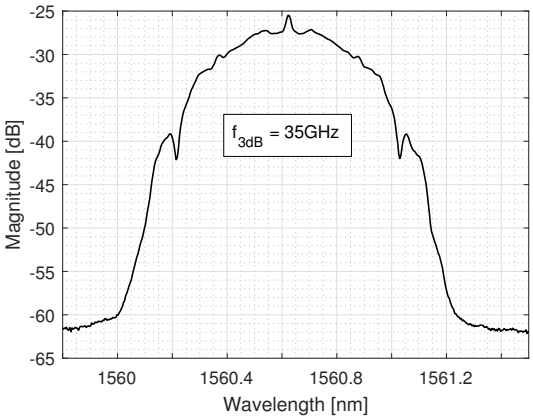


Figure 2.10: Linear Impairment: Bandwidth limiting effect of a Fujitsu FTM326422.

$V_{\text{in}}(t)$ and then recombined into the optical output $E_{\text{out}}(t)$. The corresponding output field is given by [47, Eq. 1],

$$E_{\text{out}}(t) = E_{\text{in}}(t)\frac{1}{2}\left(\exp\left(\frac{j\pi V_{\text{in}}(t)}{V_\pi}\right) + \varrho\exp\left(-\frac{j\pi V_{\text{in}}(t)}{V_\pi}\right)\right) \tag{2.1}$$

where V_π denotes the halfwave switching voltage. The parameter ϱ is a scaling factor between zero and one that accounts for a splitting imbalance and hence for an asymmetric device. It is related to the optical extinction ratio δ, defined as the ratio of maximum to minimum optical power, by $\varrho = (\sqrt{\delta}-1)/(\sqrt{\delta}+1)$ [47]. According to (2.1), the transfer function of an ideal device with infinite-extinction ratio is shown in Fig. 2.9. A typical modulator exhibits an optical extinction ratio in the range of 15 to 25 dB, which will results in a spurious phase shift in Fig. 2.9. The dashed line represents the output field, while the solid line represents the corresponding output power. In coherent systems the bias point is set to to the null point, in order to operate whenever possible in the linear regime [48]. However, if the output voltage swing of $V_{\text{in}}(t)$ exceeds the linear regime non-equidistant constellation points and IMDs occur in the output signal.

2.1.5 Digital Signal Processing Algorithms

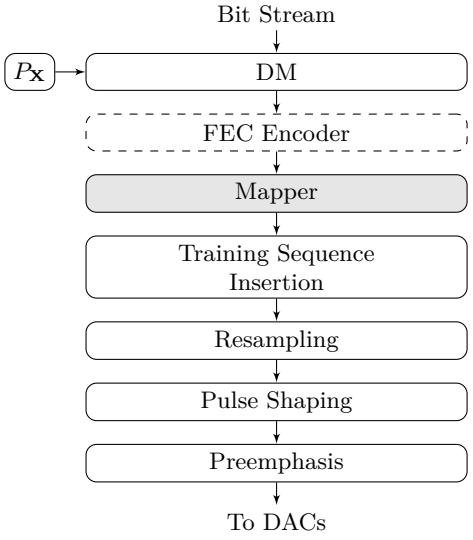


Figure 2.11: Block diagram of a coherent transmitter DSP chain. In the case of probabilistic shaping, the input distribution $P_{\mathbf{X}}$ is imposed on the FEC encoder input with a distribution matcher (DM). While the dotted blocks are omitted in the simulations and experiments in this thesis, the gray highlighted DSP block is replaced by a neural network, see Chapter 4.

Most of today's communication systems build on the design paradigm of bit-interleaved coded modulation (BICM) because of its powerful scheme to achieve highest data rates with general signal constellations. BICM was initially proposed by E.Zehavi in [49] and studied in information-theoretic terms by G.Caire in [50]. The key idea is to separate the actual coding from the modulation through an interleaving permutation [51]. In particular, as further outlined by AG i Fàbregas in [51], BICM takes advantage of the signal-space coding perspective, whilst allowing for the use of powerful families of binary codes, such as Turbo codes [52, 53] and low-density parity-check (LDPC) [54, 55] codes, with virtually any modulation format. The marginal performance loss (see Fig.4 in [50]) incurred by the BICM interface is thereby outweighed by the low complexity. In principle this loss can be reduced by using iterative decoding [56], however, at the cost of increasing latency and complexity. Further information on BICM can be found in [51, 57].

In this thesis, we assume non-iterative BICM schemes, which is the de-facto standard of modern communication systems. In the following, we briefly outline the DSP chain of the transmitter, including FEC, modulation schemes and constellations shaping. Note that the corresponding receiver-side DSP scheme is discussed in Section 2.3.2.

Forward Error Correction Encoder

Since Shannon's seminal landmark on the channel capacity of the additive white Gaussian noise (AWGN) channel [58], much research effort has been put into finding practical codes that approach this limit with reasonable complexity [59]. In the following, we will briefly discuss coding in general as well as linear block codes and the task of the FEC encoder. For more technical details on channel coding, the reader is referred to [60].

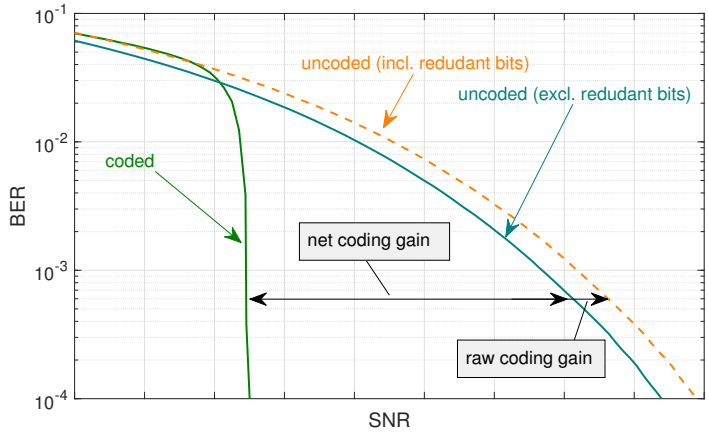


Figure 2.12: Coded and uncoded BER versus signal-to-noise ratio (SNR).

FEC is a common technique to control the errors in the data transmission over noisy channels. In principle, it is a method of improving the system bit error rate without requiring an increase in the SNR [61, Sec. II-C]. For instance, in fiber-optics the BER after FEC decoding, should be as low as 10^{-12} or 10^{-15} [62]. The key concept of FEC is to encode the signal in a redundant way by applying a channel code at the transmitter. This allows the receiver to detect a limited number of erroneous bits in the received signal and to correct these errors without the need of re-transmission. The channel codes we assume in this thesis are linear block codes. In a block code, a sequence of information bits of length k_c is encoded and redundancy is added, resulting in a codeword of length n_c [59]. Hence, such codes exhibits a rate of $R_\mathrm{C} = k_c/n_c$, or equivalently, a coding overhead (OH) of ${R_\mathrm{C}}^{-1} - 1$ [59]. If the information bits remain unchanged by the FEC encoder and only redundancy bits are added, the encoder is called systematic [59]. This approach is often used to reduce complexity.

The strength of the FEC is characterized by the coding gain, i.e., the difference in SNR at which the system operates with a specific BER with and without FEC. Note that, the serial addition of the redundant bits increases the bit rate. Hence, to maintain the same

performance, the required SNR increases by the ratio of the rate expansion, e.g, a 7% rate expansion requires a 0.3 dB increase in the SNR [61]. The coding gain is thus quoted typically as a net coding gain, which is obtained by subtracting the linear noise penalty associated with the expanded serial rate from the raw coding gain [61], as shown in Fig. 2.12.

Mapper

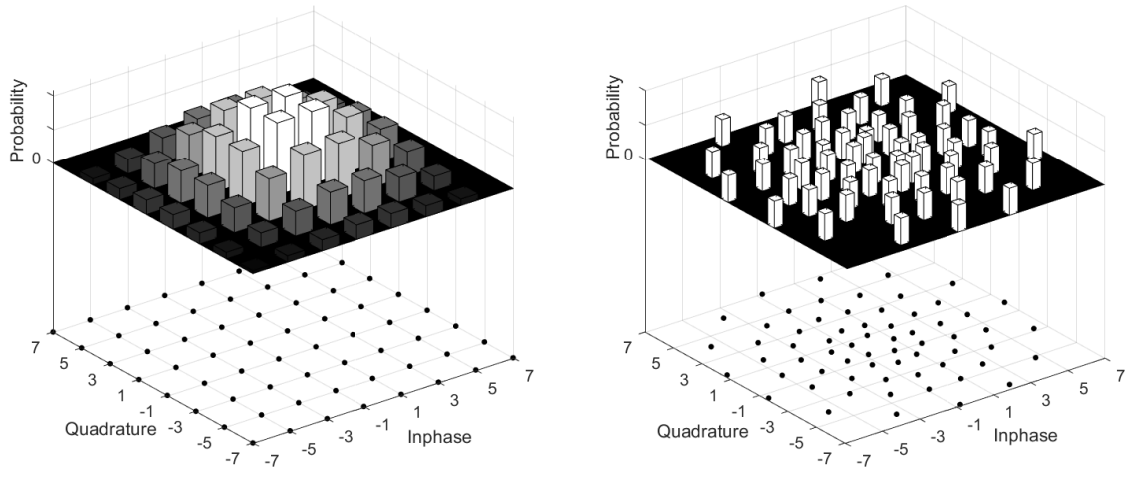
According to the modulation format in use, the encoded bits are mapped to the symbol constellation. The general purpose of a modulation format is to represent the encoded binary information in a appropriate form for transmission. The fundamental trade-off in the choice of modulation format is thereby: robustness to distortion while maintaining high spectral efficiency [59].

The most simplest modulation format represents a logical zero by the signal being absent and a logical one by presence of the signal. In the communication community this method is called on-off keying (OOK), with a maximum spectral efficiency of 1 bit/s/Hz. More advanced modulation formats, for instance two-dimensional (2D) modulations, such as QAM constellations, utilize the orthogonality of trigonometric functions to transmit information in the real part (in-phase) and imaginary part (quadrature) of the signal. Note that many 2D modulation formats are the results of two one-dimensional (1D) concatenated formats, e.g., squared (even) QAMs. In this case, the two dimension can be demodulated separately without any information loss at the receiver, however, this is only possible, if the channel impairments on the in-phase and quadrature components are not correlated.

Constellation Shaping

On the AWGN channel, discrete and equidistant constellations with uniform signaling result in a gap to capacity of 1.53 dB at high SNR [63, Sec. IV-B]. To compensate this performance loss, signal shaping [64] can be employed to mimic a Gaussian-like shape of the constellation and to obtain optimal signaling.

Signal shaping can be categorized into geometric and probabilistic shaping, as shown in Fig. 2.13. In geometric shaping, a non-uniformly spaced constellation with equiprobable symbols is used, whereas in probabilistic shaping, the constellation is fixed to a uniform grid with differing probabilities per constellation point [59]. While probabilistic shaping is able to close the gap to AWGN capacity, in geometric shaping a gap remains, if the cardinality of



(a) In probabilistic shaping, the constellation is fixed to a uniform grid with differing probabilities per constellation point.

(b) In geometric shaping, a non-uniformly spaced constellation with equiprobable symbols is used.

Figure 2.13: Constellation shaping.

constellation is finite. However, both techniques attract significant attention and are included in various standards, for instance in DVB-NGH [65], where geometric shaping is proposed. A comprehensive comparison of both geometric and probabilistic shaping in terms of their information theoretic achievable rates can be found in [66].

In this thesis, geometric shaping is applied for designing modulation formats and the approach of probabilistic shaping is assumed for label extension. In the case of probabilistic shaping we assume the probabilistic amplitude shaping (PAS) architecture of [67], which optimally intertwines shaping and coding in a capacity approaching and efficiently implementable way by concatenating a constant composition distribution matcher (CCDM) and an off-the-shelf FEC. The role of the DM is thereby to transform a uniformly distributed bit stream to a non-uniform bit stream. More details on the principle of CCDM are given in [68]. Note that in the case of probabilistic shaping a binary systematic FEC encoder is assumed, since it does not affect the information bits, hence the shaped DM output remains unchanged by the FEC encoder.

Further DSP algorithms

After the mapper, constant amplitude zero auto-correlation (CAZAC) training sequences and pilot symbols are inserted prior to the payload for data-aided frame detection, carrier

frequency offset estimation, channel impulse response estimation and carrier phase estimation at the receiver DSP [69]. The training symbols and payload data form together the DSP frame structure. The resulting frame is resampled to the sample rate of the DAC and shaped by a root-raised-cosine (RRC) filter. Pulse shaping is required to reduce the spectral width of the signal and hence to reduces the ISI in bandlimited channels. For a detailed description on pulse shaping the reader is refer to [70], where an elaborate explanation on the design of pulse shaping filters is provided. Finally, to compensate for transmitter impairments, a static preemphasis can be applied after pulse shaping.

2.2 Fiber-Optic Channel

Within the fiber-optic channel, the transmitted orthogonally polarized waveform is exposed to loss, dispersion and non-linear effects. Its evolution through a standard single-mode fiber (SSMF) can be modeled by the coupled NLSE [59, Eq. 3.4]

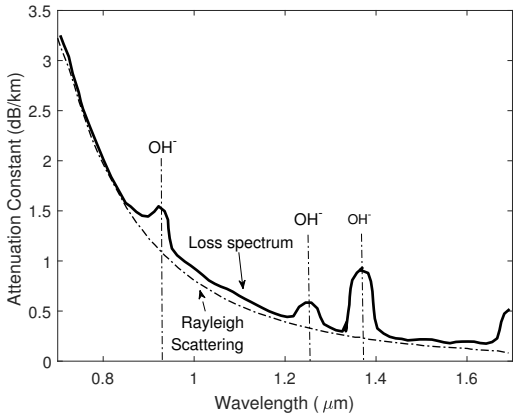
$$\frac{\partial A_x}{\partial z} = \underbrace{-\frac{\alpha}{2}A_x}_{\substack{\text{Linear Impairment:}\\ \text{Attenuation}}} - \underbrace{\beta_{1,x}\frac{\partial A_x}{\partial t}}_{\substack{\text{Linear Impairment:}\\ \text{Group delay}}} + \underbrace{j\frac{\beta_{2,x}}{2}\frac{\partial^2 A_x}{\partial t^2}}_{\substack{\text{Linear Impairment:}\\ \text{Chromatic disperison}}} + \underbrace{\frac{\beta_{3,x}}{6}\frac{\partial^2 A_x}{\partial t^2}}_{\substack{\text{Linear Impairment:}\\ \text{Third-order disperison}}} \underbrace{- j\gamma(|A_x^2| + \frac{2}{3}|A_y^2|)A_x}_{\substack{\text{Non-linear Impairment:}\\ \text{Kerr elctro-optic effect}}} \tag{2.2}$$

where $A(z,t)$ denote the pulse envelope, α the fiber loss, β_1 the group delay, β_2 the chromatic dispersion, β_3 third-order dispersion and γ the non-linearity of the waveguide. The corresponding couple NLSE of the y-polarization can be obtain by changing the indices. The first, second, third and fourth term on the right side represent the linear effects, whereas the fifth term describes the influence of the non-linearity. In the following, the linear (Section 2.2.1, 2.2.2, 2.2.3) and non-linear effects (Section 2.2.4) will be considered independently in order to describe their behavior.

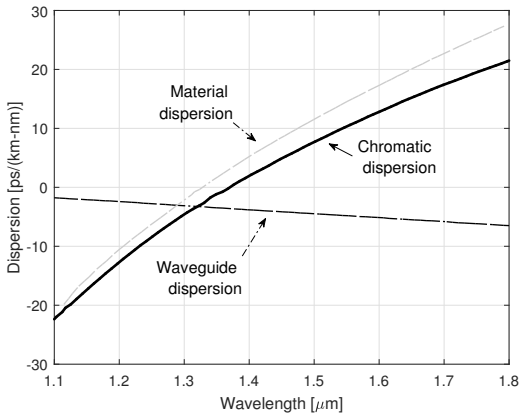
Note that the coupled NLSE is a simplification of the general non-linear Schrödinger equation (GNLSE) [71, Eq. 4.375]. Depending on the required channel model, e.g., if a more realistic fiber with randomly varying polarisation states and modal birefringence is necessary, the coupled NLSE must be further extended to fully account for such effects. As they play a subordinate role in this thesis, a detailed discuss is omitted here. The interested reader is referred to, e.g., [71, Eq. 4.392], for further details how to extend the coupled NLSE.

2.2.1 Attenuation

An optical pulse carried through an optical fiber channel is attenuated by material absorption and Rayleigh scattering [5], whereby the total attenuation depends on the selected wavelength and fiber in use as shown in Fig 2.14a. The attenuation constant is especially high near the resonance frequencies of the oscillators in the media, because the power of the electromagnetic field will be partly transformed into kinetic energy. For optical fibers these



(a) Attenuation in a fiber made of silica glass, the dashed line depicts the Rayleigh scattering (OH^- hydroxyl ions) [5, Fig 2.4].



(b) Material, waveguide, and chromatic dispersion in SSMF [72, Fig. 3.4].

Figure 2.14: Attenuation and chromatic dispersion.

oscillators are built by the molecules of the basic material (silica glass) and impurities in the fiber, mainly hydroxyl (OH^-) ions [5]. Hence, the effect occurs not only near the fundamental resonances of the molecules, but also in the vicinity of their harmonics. Common coherent optical communication systems operating therefore in the range of 1550 nm (C-band frequency spectrum), where a SSMF exhibits an minimum average attenuation coefficient of $\alpha_{\mathrm{dB}} = 0.2$ dB/km.

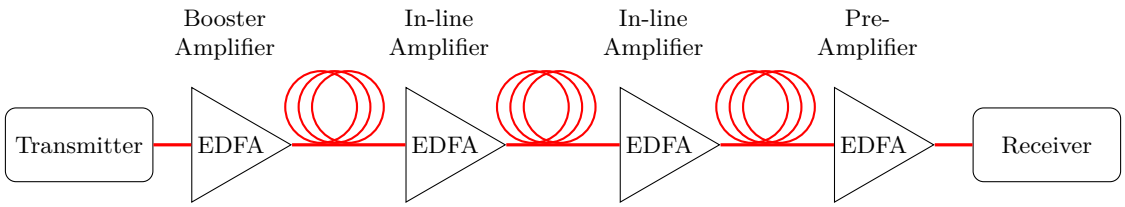


Figure 2.15: Booster, inline, and pre-amplifier EDFAs used in optical transmission line

The accumulated losses are compensated by optical amplifiers, e.g., EDFA, at the end of each span. However, the optical amplifiers introduce additive noise, in particular amplified spontaneous emission (ASE) noise during amplification. The amount of ASE noise is related to noise figure (NF) and the amplification factor of the EDFA.

2.2.2 Chromatic Dispersion

While the group delay and the third-order-dispersion term can be neglected by considering a reference frame and a long pulse duration, see [59, Sec. 3.1.2], the chromatic dispersion term cannot be omitted. In contrast to propagation in vacuum, the phase velocity in a medium is a function of the frequency or wavelength [5, Eq. 2.63]. If, for instance, a modulated optical carrier is injected into fiber the different frequency components will propagate with different phase velocities and will arrive at the output at different times. This alteration of the propagation time is the group-velocity dispersion (GVD) [5, Sec. 5.3].

In optical fibers the accumulated dispersion is called chromatic dispersion and it is the combination of material dispersion and waveguide dispersion as shown in Fig. 2.14b. As a consequence of GVD, the frequency components arrive out of phase at the output of the fiber and the duration of the resulting pulses is broadened. In high-speed optical transmission systems, which operate at minimum fiber attenuation (1550 nm), this will lead to ISI which has to be compensated at the receiver-side by DSP (see Section 2.3.2). A detailed derivation of the linear part of Eq. (2.2) and hence a mathematical description of the pulse broadening can be found in [5, Sec. 5.3].

2.2.3 Polarization-Mode Disperison

The degenerate nature of the orthogonally polarized modes holds only for an ideal SSMF with perfectly cylindrical core of uniform diameter [72]. Due to manufacturing imperfections, mechanical strain and stress-induced anisotropy caused by bending and vibrations, such a perfect circular shape with a constant refractive index along the fiber cannot be achieved [73]. Hence the mode degeneracy between the orthogonal polarization states is broken and the fiber acquires birefringence [72]. The fiber birefringence $\Delta n(\omega) = |n_f(\omega) - n_s(\omega)|$ [72, Eq. 2.2.40] causes a difference between the phase in each polarization mode [73, Eq. 2.19]

$$\Delta\beta(\omega) = |b_f(\omega) - b_s(\omega)| = \frac{w}{c}\Delta n(\omega) = \frac{w}{c}|n_f(\omega) - n_s(\omega)|, \tag{2.3}$$

where n_f is associated with the faster group velocity of the fast mode and n_s refers to the slow mode. This effect is called polarization mode dispersion (PMD). PMD can be modeled by considering concatenating S independent fiber segments with independent uniform

birefringence [69, Eq. 6]

$$\boldsymbol{H}_{\mathrm{PMD}}(\omega) = \prod_{s=1}^{S} \boldsymbol{R}(\phi_s^{'}) \boldsymbol{H}(\omega, \Delta\tau_s, \theta_s) \boldsymbol{R}(\phi_s^{''}) \tag{2.4}$$

where $\phi_s^{'}$ and $\phi_s^{''}$ denote the angles between the slow axis and the polarization state of the fast axis, respectively, θ_s the polarization phase and $\Delta\tau_s$ the differential group delay (DGD) between the two orthogonal fast and slow axes of the Hermitian birefringence matrix.

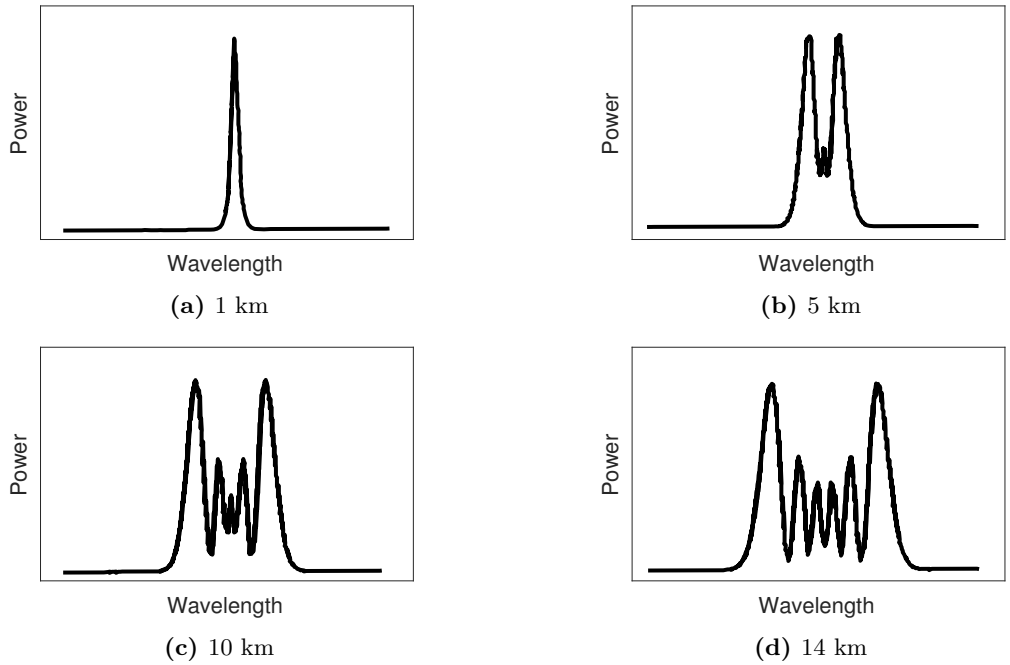
In general, because of its $\sqrt{L}$ dependence [72, Eq. 1.2.18], PMD-induced pulse broadening is relatively small compared with GVD effects. However, it becomes a limiting factor for high-speed communication systems and has to be continuously compensated at the receiver-side DSP due to its time-varying behavior [72].

2.2.4 Kerr Elctro-Optic Effect

A fiber-optic channel is a non-linear medium. Fiber non-linearities are generated by the Kerr effect and are covered by the fifth right term in Eq. (2.2). In 1875 [74], John Kerr discovered that the refractive index of a dielectric material changes its value in response to an external applied electric field. Furthermore, he showed that the induced change of the refractive index is directly proportional to the square of the external applied electric field.

In optical fiber communication the propagating light itself causes an instantaneous change in the electric field. Hence changes in the refractive index lead in turn to changes in the signal speed, which manifest themselves as small changes in the phase of the propagating wave [75].

By neglecting the linear parts in Eq. (2.2), the differential equation is directly solvable in time domain and the influence of the fiber non-linearity on a propagating pulse can be examined. According to the solution provided in [5, Sec. 5.5], it turns out that new frequency components are generated in the spectrum of a propagating pulse due to the non-linear behavior. Fig. 2.16 shows an example provided in [5, Sec. 5.5]. It depicts the spectral evolution of an initially unchirped optical pulse with a duration of 25 ps after a propagation distance of 1 km, 5 km, 10 km and 14 km in a SSMF without dispersion. Note that, if the pulse changes its spectrum due to its own intensity we observe self phase modulation (SPM), while the spectral broadening induced by other pulses, propagating at the same time in the waveguide, is called cross phase modulation (XPM). Both are, in principle, deterministic non-linear effects that can be compensated by inverting the NLSE (2.2). Since the bit pattern



(a) 1 km **(b)** 5 km **(c)** 10 km **(d)** 14 km

Figure 2.16: Spectral evolution of an unchirped optical Gaussian pulse with a duration of 25 ps at different distances in a SSMF [5, Fig. 4.20].

of the other channels is unknown, only SPM can be compensated with standard NLSE approaches at the receiver-side DSP (see Section 2.3.2).

2.3 Coherent Optical Receiver

Coherent detection allows a linear mapping of the optical signal into the electrical domain such that amplitude and phase of both polarizations are available in the electrical domain after demodulation by the optical front-end and detection by a row of photodiodes [76]. The full knowledge of the optical field enables the utilization of sophisticated modulation schemes, such as QAM, as well as the compensation of the O/E components and fiber impairments mentioned in the previous sections. In the following the optical front-end (Section 2.3.1) and the standard receiver-side DSP (Section 2.3.2) to recover the transmitted signal are described in more detail.

2.3.1 Optical Front-End

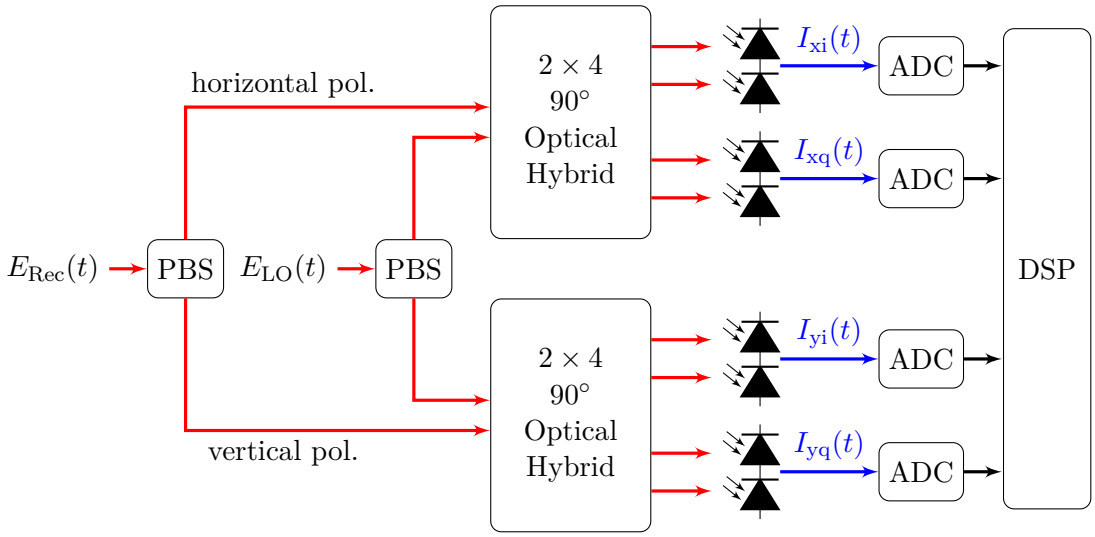


Figure 2.17: Coherent optical receiver consisting two polarization beam splitter (PBS), two 90° hybrids, four balanced photodiodes and four ADCs.

In Fig. 2.17 a high level block diagram of a coherent optical receiver is illustrated. The received optical signal is separated by a polarization beam splitter (PBS) into two orthogonal polarization components X and Y and fed to a 90° hybrid. In the 90° hybrid the signal is mixed with the local oscillator (LO) to separate the corresponding in-phase and quadrature components of each polarization. The LO consists of a LASER with a narrow linewidth and higher power compared to the received signal to enhance the receiver sensitivity [77]. The obtained optical signals are then transferred into the electrical domain by a row of balanced photodiodes. This is followed by four high-speed ADCs, which digitalize the baseband signals

within a certain range by a limited number of digital output codes [43]. Since the analog scale is continuous, while the digital codes are discrete, a quantization error is introduced. Note that quantization is indeed non-linear but can be modeled as AWGN.

2.3.2 Digital Signal Processing Algorithms

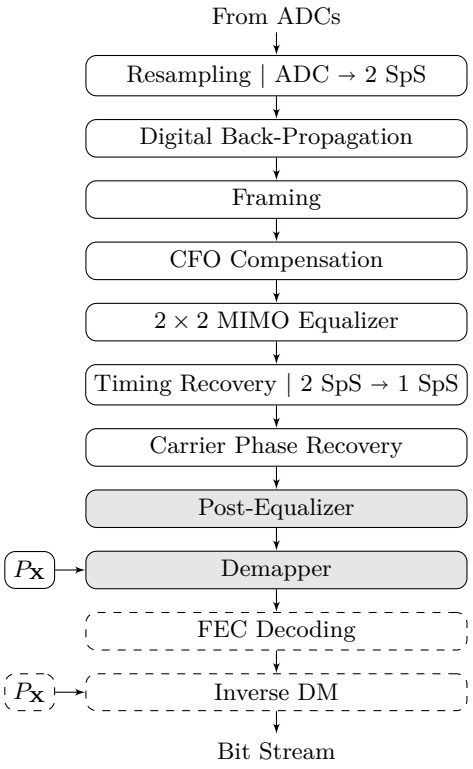


Figure 2.18: Block diagram of the coherent offline receiver DSP chain. The demapper and the CCDM requires knowledge about the input distribution $P_{\mathbf{X}}$ to correctly calculate the prior information respectively to inverse the shaping. While the dotted blocks are omitted in the simulation and experiments in this thesis, the gray highlighted DSP blocks are replaced by neural network, see Chapters 4 and 5.

Coherent detection recovers the entire optical field within the electronic bandwidth of the receiver and consequently DSP allows for efficient mitigation of the main degrading distortions arising from the optical fiber, component imperfections as well as clock and carrier recovery. Some of these effects are deterministic while others are random. This section describes the main standard DSP algorithms which are utilized to recover the received signals after transmission. The flow of the signal through the data-aided receiver DSP is depicted in Fig. 2.18.

Digital Back-Propagation

After resampling, DBP aims at inverting, via Eqs. (2.2), the optical fiber channel, i.e., the forward propagation [9]. However, an analytical solution does not exist. Only for isolated effects analytical solutions can be derived as it has been mentioned in the previous section to illustrate the phenomena of chromatic dispersion and fiber non-linearity. Nevertheless, an approximated solution can be obtained via the SSFM [78], a numerical finite-element method. Eq. (2.2) can be written in the following form

$$\frac{\partial A_x}{\partial z} = (\hat{L} + \hat{N})A_x \tag{2.5}$$

where $\hat{L} = j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2}$ is the linear and $\hat{N} = -j\gamma|A_x + A_y|^2$ the non-linear part. Following T. Schneider [5, Chap. 5.3 and 5.5] the solution of (2.5) reads as

$$A_x(z + \Delta z, t) = e^{\Delta z(\hat{L}+\hat{N})}A_x(z, t) \tag{2.6}$$

where z is the current position within the span, t the time and Δz the step size. If Δz is sufficiently short, the right part of (2.6) can be approximated by the SSFM, breaking the interaction of linear and non-linear distortion within this short step, yielding the solution

$$e^{\Delta z(\hat{L}+\hat{N})}A_x(z, t) \approx e^{\Delta z\hat{L}}e^{\Delta z\hat{N}}A_x(z, t). \tag{2.7}$$

These short steps are then cascaded and solve for the link length. The linear part of (2.7) provides the bulk chromatic dispersion compensation. Chromatic dispersion can be considered as a time-invariant linear impairment, unless the optical path is newly routed. Therefore, the chromatic dispersion can be estimated initially by blind algorithms [79–85] and afterwards compensated by a zero-forcing (ZF) frequency divison equalizer (FDE) [69, 84–86] per polarization, analytically expressed as [9, Eq. (7)]

$$\boldsymbol{A}(z + \Delta z, t) = \mathcal{F}^{-1}\{\mathcal{F}\{\boldsymbol{A}(z, t)\}H_{\mathrm{CD}}(w_k)\} \tag{2.8}$$

with $\boldsymbol{A} = [A_x, A_y]$ and $H_{\mathrm{CD}}(w_k)$ given by [69]

$$H_{\mathrm{CD}}(w_k) = \exp\Big(\, j\frac{D\lambda^2 w_k^2}{4\pi c}\Big)\,, \tag{2.9}$$

and where $D = \beta_2\Delta z$ in ps/nm is responsible for the accumulated dispersion of a signal after propagating through an optical fiber of length Δz in km and with group velocity dispersion β_2 in ps^{2}nm, λ in nm is the central wavelength of the signal, c is the speed of light in

vacuum. Since the frequency equalizer is discrete $w_k, k = 0, \ldots, N-1$ represents the discrete angular frequencies, where N is the length of the filter. One step frequency-domain linear compensation for the two polarizations requires thus the evaluation of four N-point complex fast Fourier transforms (FFTs) and $2N$ complex multiplications, which gives a complexity of $2N + 2N\log_2(N)$ complex multiplications or $8N + 8N\log_2(N)$ real multiplications [87]. Due to the overlap-save algorithm for blockwise frequency-domain filtering, the number of useful samples on both polarizations is reduced from $2N$ to $2(N - 2N_{\text{overlap}})$, where N_{overlap} denotes the number of overlapping samples. As a result, the required number of real multiplications per symbol for one linear step is [87, Eq. 7]

$$\text{mul}_{\text{CD-FDE}} = \frac{n_{\text{pol}} n_{\text{samp}}(8N + 8N\log_2(N))}{n_{\text{pol}}(N - 2N_{\text{overlap}})}. \tag{2.10}$$

where n_{pol} is the number of polarizations and n_{samp} is the oversampling ratio per polarization. The non-linear part of (2.7) provides the non-linear phase corrections. It is applied in time-domain and can be analytically expressed by the following complex exponential factor [9, Eq. (8)]

$$\boldsymbol{A}(z + \Delta z, t) = \boldsymbol{A}(z, t) e^{-j\gamma |\boldsymbol{A}(z,t)|^2 \Delta z \phi} \tag{2.11}$$

where $\phi \in [0, 1]$ is the compensation scaling factor to optimize the non-linear mitigation as proposed in [88]. One step time-domain non-linear compensation for the two polarizations requires $2 \times 10N$ real multiplications. Hence, the number of real multiplications per symbol for one non-linear step is [87, Eq. 8]

$$\text{mul}_{\text{Standard-DBP}} = \text{mul}_{\text{CD-FDE}} + \frac{n_{\text{pol}} n_{\text{samp}} 10N}{n_{\text{pol}} N}. \tag{2.12}$$

In essence, the DBP uses the SSFM to calculate the transmitted signal by simulated propagation of the received signal through a fiber with negated parameters with respect to the fiber link [89]. Note that, the optimal value of the step size Δz, in terms of performance, would tend to zero. However, this would require an infinite number of steps within the fiber channel (and hence high computational complexity). In practice it is common to consider only one step per span when non-linear fiber compensation is applied.

Carrier Frequency Offset Compensation

The optical signal can be detected at the receiver either in homodyne or intradyne configuration. In homodyne detection, the beam of a laser source is split between transmitter

and receiver, which ensures matched phase and frequency. However, this configuration is only possible in the lab. In intradyne detection, two laser sources are used, i.e., one at the transmitter and one at the receiver. In real systems, the two laser cannot be perfectly synchronized. This leads to a frequency offset in the order of ±1.8GHz, which has to be compensated in the DSP domain. In this thesis, we used a carrier frequency offset (CFO) compensation algorithm based on Schmidl&Cox [90] and implemented it according to [91].

2×2 MIMO Equalizer

While the prior FDEs of the DBP are dedicated to perform coarse compensation of the static chromatic dispersion, a shorter adaptive 2×2 multiple input multiple output (MIMO) FDE is dedicated to polarization demultiplexing, residual PMD compensation, as well as tracking of time-varying effects [69]. The equalizer matrix $\boldsymbol{W}(w_k)$ is obtained by using the minimum mean square error (MMSE) solution, which aims to minimize the expectation of the squared Euclidean distance of the equalized and transmitted signal, and maximizes the signal-to-interference-and-noise ratio (SINR). The equalizer matrix is given by [69, Eq. (18)]

$$\boldsymbol{W}(w_k) = \boldsymbol{H}^H(w_k)[\boldsymbol{H}(w_k)\boldsymbol{H}^H(w_k) + \boldsymbol{\delta}(w_k)]^{-1}, \tag{2.13}$$

where $\boldsymbol{H}(w_k)$ denotes the channel matrix and $\boldsymbol{\delta}(w_k)$ the noise autocorrelation matrix. Note that, the MMSE solution converges in the high SNR regime to the ZF solution and in the low SNR regime to the matched filter (MF) solution. It provides the best trade-off, in terms of performance, between ISI suppression and noise enhancement.

Since in a real system the channel matrix $\boldsymbol{H}(w_k)$ of (2.13) is not known, the receiver needs to retrieve it from the received data. In this thesis, training-aided channel estimation based on CAZAC sequences is applied [69, Sec. IV]. This approach requires periodical transmission of the training sequence to track the time-varying effects [92], such as PMD.

Timing Recovery

At this stage the coarse frequency offset between the transmitter LASER and LO has been removed from the received samples. However, as the transmitter and receiver operate with separate radio frequency (RF) clock sources, sampling points at the ADC are not aligned to the incoming data samples. This corresponds to small sampling frequency offsets and drifting sampling phase relations. A digital timing recovery (TR), either in time domain [93,

94], combined time and frequency domain [95], or pure frequency domain [96], estimates the sampling phase and resamples the signal to align the samples. In this thesis, the most common time domain method is used, the Gardner method [94], which operates at two samples per symbol.

Carrier Phase Recovery

The carrier phase recovery (CPR) tracks and compensates laser phase noise introduced by free-running transmitter and LO lasers. Laser phase noise is often modeled as a Wiener noise process [97, 98], which is a random walk of Gaussian variables with zero mean and variance proportional to the combined linewidth of the lasers [99]. Various approaches can be used to compensate this deterministic distortion, such as the Viterbi and Viterbi [100], which was used in the receiver DSP in the thesis for small constellations. For high-order constellations a pilot-aided CPR [101] was employed.

Post-Equalizer

A linear or non-linear equalizer is placed before the memoryless demapper to combat residual distortion, like residual ISI due to non-ideal channel characteristics. The equalizer attempts to reverse the residual effects before the demapper, which assumes an AWGN channel and hence ideal compensation of the impairments. A popular approach for non-linear equalization is shown in Fig. 2.19, namely the VNLE. The VNLE has proven to be very effective against component and fiber non-linearities [11, 12].

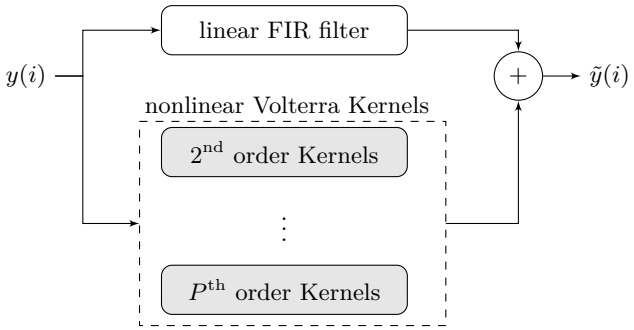


Figure 2.19: Blockdiagram of a Volterra non-linear equalizer.

By combining linear convolution and non-linear power series, the VNLE is capable to describe causal as well as non-causal time-invariant non-linear systems with finite fading memory. Let $y(i)$ and $\tilde{y}(i)$ represent a system with single input and single output, respectively, the p-th order non-causal discrete time Volterra series is given by [102, Sec. 4.2]

$$\tilde{y}(i) = f_{\text{VNLE}}(y(i), h_p) = \sum_{p=1}^{P} \sum_{s_1=-M_1}^{M_1} \cdots \sum_{s_p=-M_p}^{M_p} h_p(s_1, \cdots, s_p) \prod_{k=1}^{p} y(i - s_k), \tag{2.14}$$

where $h_p(s_1, \cdots, s_p)$ denotes the p-th order Volterra kernel, M_1 the symmetric memory length for the linear terms and M_2 to M_p the symmetric memory lengths for the non-linear terms of second order and higher. Generalizing [103, Table I] to an arbitrary order p, the relationship between the symmetric memory length and the number of equalizer kernels N_p is given by

$$N_p = \frac{1}{p!} \prod_{i=0}^{p-1} (2M_p + 1 + i). \tag{2.15}$$

The real-time complexity of a VNLE is defined by the number of kernels, see (2.14). Their number is directly connected to the required number of hardware multipliers, namely

$$\text{mul}_{\text{VNLE}} = \underbrace{N_1 + \sum_{i=2}^{P} (N_i +}_{\text{Kernels}} \quad \underbrace{(2M_i + 1) + 1}_{\text{Feature Matrix [104, Sec. IV-B]}} \quad). \tag{2.16}$$

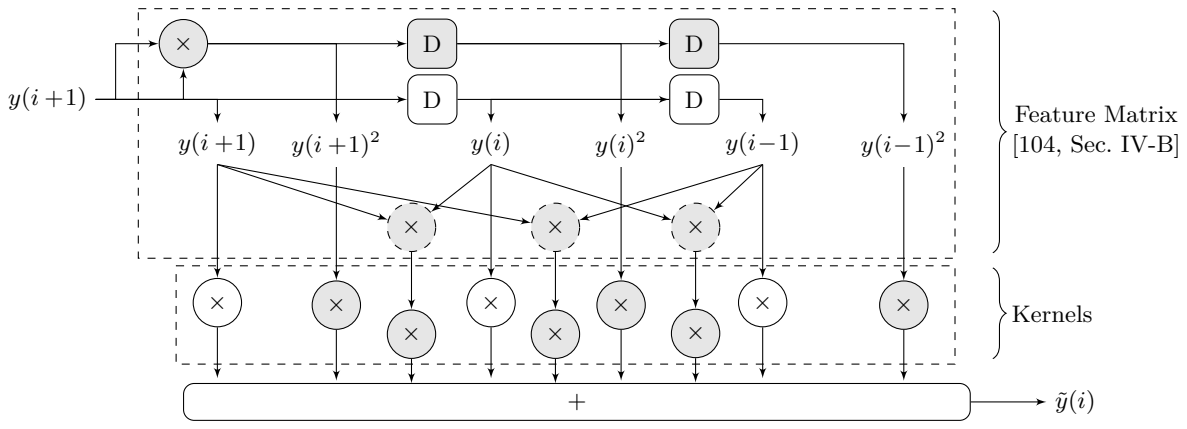


Figure 2.20: Signal flow diagram of a 2nd order Volterra equalizer with symmetric memory length of $M_1 = 1$ and $M_2 = 1$. The white blocks indicate the linear part (FIR filter), while the gray blocks indicate the non-linear part. The dashed multipliers denote thereby the generation of the cross-terms.

Eq. (2.16) is derived from the structure presented in [104, Sec. IV-B]. It considers terms that can be obtained by delaying other terms as well as the reuse of products of order k to compute products of order $k+1$.

Demapper

The a posteriori probability (APP) soft-demapper calculates soft-bits [105, Sec. 3]

$$\ell_i = \log \frac{\Pr(b_i = 0|y)}{\Pr(b_i = 1|y)}. \tag{2.17}$$

We can interpret the values of the soft-bits in the following way: the signs represent the hard decision, i.e., 0, if $\ell_i > 0$ and 1, if $\ell_i \leq 0$, and the absolute values $|\ell_i|$ indicate how confident the soft-demapper is about its decisions. Large $|\ell_i|$ indicate high confidence. With the aid of Bayes' rule, Eq. (2.17) can be expressed in terms of channel likelihoods and prior information [105, Eq. 7]

$$\ell_i = \underbrace{\log \frac{\sum_{x \in X(b_i=0)} p(y|x)}{\sum_{x \in X(b_i=1)} p(y|x)}}_{\text{channel likelihood}} + \underbrace{\log \frac{\Pr(b_i = 0)}{\Pr(b_i = 1)}}_{\text{prior information}}, \tag{2.18}$$

where the sets $X(b_i = j)$ contain all constellation symbols with the i-th bit equals j. The channel likelihood conditions on symbols and are independent of the source statistics. The prior information contains the information about the source statistics.

For independent and uniformly distributed symbol the prior information is equals zero. Further assuming an AWGN channel, the soft-bits in Eq. (2.18) can be calculated according to [106, Sec. 3.3]

$$\ell_i = \sum_{b \in \mathbb{B}} (-1)^b \log \sum_{x \in X(b_i=b)} \exp(-\frac{1}{N_0} ||y - x||^2), \tag{2.19}$$

where N_0 is the noise power spectral density. For binary phase shift keying (BPSK) transmission this further simplifies to

$$\ell_i = \frac{y}{N_0}. \tag{2.20}$$

A general alternative low complexity solution of Eq. (2.19) is given by applying the max-log approximation (MLA) [106, Sec. 3.3.3]

$$\ell_i = \frac{1}{N_0} \sum_{b \in \mathbb{B}} (-1)^b \min_{x \in X(b_i=b)} ||y - x||^2. \tag{2.21}$$

For non-uniformly distributed symbol levels, as in the case of probabilistic shaping, the prior information is non-zero. In this case the demapper requires knowledge about the applied input distribution at the transmitter to correctly calculate the soft-bits in Eq. (2.18).

In general, by using the reliability information contained in the soft demapper, soft-decision FEC can be more powerful than hard-decision schemes [59, 107], for instance, if low code rates are considered [108]. Since we can obtain hard-decision (HD) from soft decision (SD), mutual information can only decrease according to the data processing theorem.

2.4 Performance Metrics

This section introduces the main used metrics in this thesis for assessing the performance of optical communication systems.

2.4.1 Pre-FEC Bit-Error Ratio

For hard-decision coding schemes, the BER after decoding, denoted as post-FEC BER, is a deterministic function of the uncoded BER, denoted as pre-FEC BER, as long as burst errors are neglected [59]. This deterministic relation between BER before and after decoding has lead to the use of a hard-decision FEC limit, giving the maximum pre-FEC BER for which a desired post-FEC BER is not exceeded. The concept of a FEC limit makes the implementation of a hard-decision FEC encoding/decoding and hence time-consuming simulations obsolete.

2.4.2 Achievable Rate

In [62] the authors demonstrated, that the concept of a FEC limit is acceptable, if the decoder is based on hard-decisions and fed with bits. However, if BER based FEC limits are used in soft-decision FEC systems, the prediction of the post-FEC BER is imprecise, see [62, Sec. IV]. A more accurate post-FEC BER predictor is the achievable rate. The achievable rate indicates the number of bits per symbol that can be reliably transmitted through the channel. By [109, Eq. (100)], we can estimate from the transmitted bits b and the soft bits ℓ the achievable rate per polarization

$$R = m - \min_{s \geq 0} \sum_{i=1}^{n} \sum_{j=1}^{m} \log_2 \left[1 + \exp(-s(1 - 2b_{ij})\ell_{ij})\right] \tag{2.22}$$

where m is the number of bits per symbol, e.g., $m = 6$ for 64-QAM, where n is the number of transmitted symbols, and where we assumed uniformly distributed channel input symbols. The optimization over s is used for matching the softbits ℓ_{ij} to the 'true' probabilities so that $s\ell_{ij} \approx \log \frac{\Pr(B_{ij}=0|\boldsymbol{y}_i)}{\Pr(B_{ij}=1|\boldsymbol{y}_i)}$ where $\boldsymbol{y}_i$ are the channel outputs available to the demapper for estimating the bit B_{ij}. Without the optimization over s, the achievable rate estimate obtained by (2.22) can be too low, providing only a loose lower bound of the theoretical limit. See [109, Sec. V.C] for details. Note that (2.22) is called general mutual information (GMI) in [62, Eq. (32)].

2.4.3 Optical Signal-to-Noise Ratio

In optical transmission systems the optical signal-to-noise ratio (OSNR) is used to quantify the optical signal quality. The definition of OSNR is [7, Eq. 33]

$$\mathrm{OSNR} = \frac{P_{\mathrm{signal}}}{P_{\mathrm{ASE}}} = \frac{P_{\mathrm{signal}}}{2N_{\mathrm{ASE}}B_{\mathrm{ref}}} \tag{2.23}$$

where P_{signal} is the total average optical signal power summed over the two states of polarization, N_{ASE} is the spectral density of ASE noise and the reference bandwidth B_{ref} is usually 12.5 GHz, corresponding to a 0.1 nm resolution bandwidth of a optical spectrum analyzer (OSA) at 1550 nm carrier wavelength (193.4 THz carrier frequency). The definition of OSNR is in direct relationship to the SNR, and differs from it by a normalization factor as follows [7, Eq. 34]

$$\mathrm{OSNR} = \frac{pR_{\mathrm{s}}}{2B_{\mathrm{ref}}}\mathrm{SNR} \tag{2.24}$$

where $p = 1$ for a single polarization and $p = 2$ for a polarization-multiplexed signal. In the experimental setup the optical signal and noise powers of Eq. (2.23) are directly measured using the OSA.

Chapter 3

Artificial Neural Networks

Artificial neural networks (ANNs) were originally developed as mathematical models for information storage and organization in biological brains [110–112]. Although it is nowadays clear that ANNs have little resemblance to real biological neurons, they enjoy continuing popularity as non-linear models [113].

Many varieties of ANNs with different properties have been introduced over the last decades. One important distinction is thereby made between ANNs whose connections are acyclic, and those whose connections form cycles. Acyclic ANNs are referred to as feed-forward neural networks [113]. Typical examples of feed-forward neural networks include perceptrons [111], radial basis function networks [114] and Hofield nets [115]. The most widely used form of feed-forward neural network, and the one we focus on in this thesis and in Section 3.1, is the mulitlayer perceptron [116]. ANNs with cycles are referred to as feedback or recurrent neural networks (RNNs) [113]. They are discussed in Section 3.2. For both neural network architectures the forward and backward pass is described.

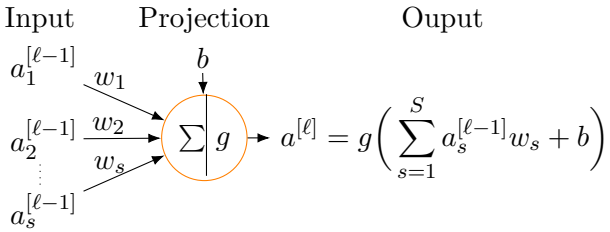


Figure 3.1: Single artificial neuron.

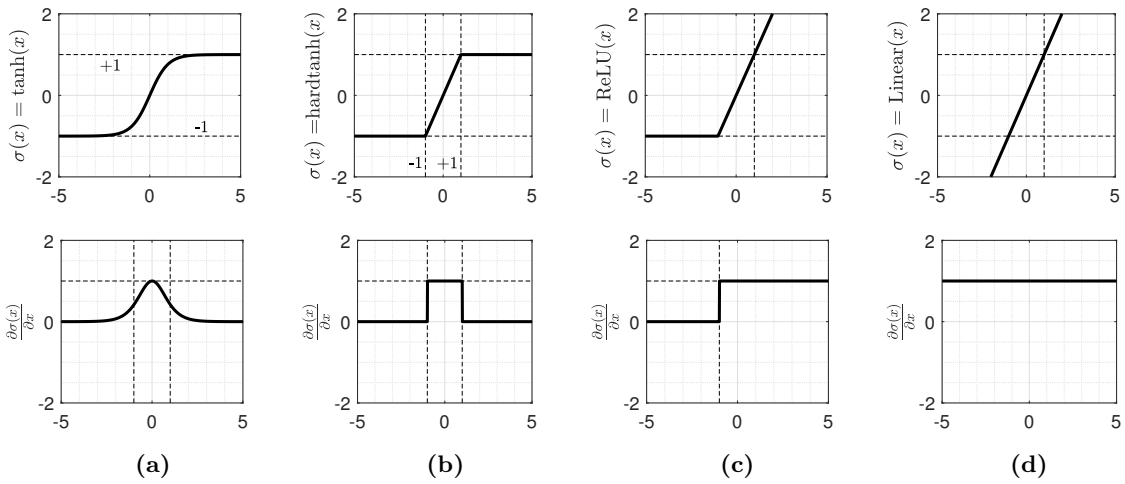


Figure 3.2: The used activation functions and the derivatives: **(a)** hyperbolic tangent (tanh) activation function **(b)** hard tanh (H-tanh) activation function **(c)** rectified linear unit (ReLU) activation function **(d)** linear activation function.

3.1 Multilayer Perceptron

Multilayer perceptrons computing structures are built from several layers of artificial neurons [117, Chap. 5]. A single artificial neuron is a processing unit with a number of inputs and one output, as shown in Fig. 3.1. Each input is associated with a weight. The neuron

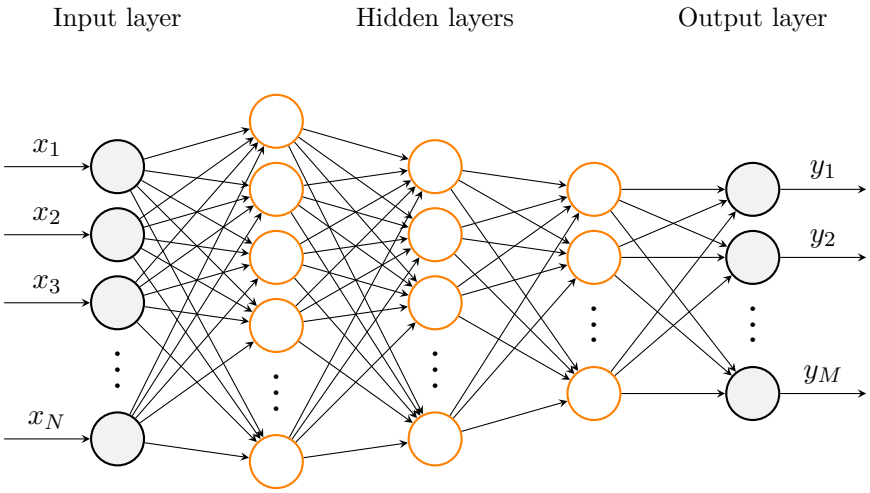


Figure 3.3: Example of a deep neural network (DNN) with three hidden layers. While the design of the input and output layer depends on the input and output dimension of the desired function, the design options of the hidden layers are numerous and interrelated.

first computes an activation by summing up the particular weighted inputs and a bias term. Secondly, an activation function $g(\cdot)$ is applied to obtain the neuron's output. The purpose of the activation function is to introduce non-linearity into the output of a neuron and to decide, whether a neuron should be activated or not. This non-linear transformation makes it capable to learn and perform more complex tasks. Fig. 3.2 depicts the activation functions and the derivatives used in this thesis. The derivatives are required for training, in particular for gradient descent, see Section 3.1.3. Interconnecting multiple neurons in parallel leads to a neural network with one hidden layer, a so-called shallow network. Multiple layers in series, where the output of the previous layer serves as input for the next layer, lead to a DNN [118] as shown in Fig. 3.3.

3.1.1 Forward Pass

Let $\boldsymbol{x} = [x_1, \ldots, x_N]$ and $\boldsymbol{y} = [y_1, \ldots, y_M]$ denote the input and output vectors with N input and M output neurons, respectively, the forward pass of a DNN with L-layers is given by

$$\boldsymbol{a}^{[0]} = \boldsymbol{x}, \tag{3.1}$$

$$\boldsymbol{z}^{[l]} = \boldsymbol{W}^{[l]}\boldsymbol{a}^{[l-1]} + \boldsymbol{b}^{[l]}, \tag{3.2}$$

$$\boldsymbol{a}^{[l]} = g(\boldsymbol{z}^{[l]}), \quad l = 1, \ldots, L \tag{3.3}$$

$$\boldsymbol{y} = \boldsymbol{a}^{[L]}, \tag{3.4}$$

where $\boldsymbol{z}^{[l]}$ denotes the activation, $\boldsymbol{a}^{[l]}$ the output vector of the l-th layer and $\boldsymbol{W}^{[l]}$ and $\boldsymbol{b}^{[l]}$ the weight matrices and bias vectors, respectively. Eq. (3.1) denotes the input layer and (3.2) as well as (3.3) are executed successively for layers $l = 1, 2, \ldots, L$ to obtain output $\boldsymbol{y}$ in (3.4). In the following, (3.1), (3.2), (3.3), and (3.4) will be referred to by

$$\boldsymbol{y} = f_{\mathrm{DNN}}(\boldsymbol{x}, \boldsymbol{W}^{[1]}, \ldots, \boldsymbol{W}^{[L]}, \boldsymbol{b}^{[1]}, \ldots, \boldsymbol{b}^{[L]}). \tag{3.5}$$

3.1.2 Depth of Multilayer Perceptron

Montúfar et al. (2014) [119] demonstrated that deep networks are more expressive and require far fewer neurons to represent a desired function than shallow networks. They showed that the number of decision regions of deep models grows exponentially in the depth and polynomially in the number of neurons in the hidden layers. Their growth is much faster than that of shallow models which grow only polynomially in the number of hidden units [120].

Consequently, deeper models need exponentially less parameters to reach the desired representation capacity. Note that, choosing a deeper model and hence a cascade structure matches as well better with the intertwined memory effects and non-linearities of the physical channel in optical communication. Deeper layers can reuse constructed conclusions from the lower layers in order to gradually build more complex functions [121]. Furthermore, in 2017, Montúfar et al. [122, Eq. 5] introduced an upper bound on the number of activation patterns of a multilayer perceptron. In a neural network the number of activation patterns correspond to the number of possible distinct input space regions that the neural network can distinguish. For fixed parameters, a network with n_0 inputs and $n_1, \ldots, n_L$ ReLUs in the hidden layers realizes at most

$$\prod_{l=1}^{L} \sum_{j=0}^{m_{l-1}} \binom{n_l}{j}, \qquad m_{l-1} = \min\{n_0, \ldots, n_{l-1}\} \tag{3.6}$$

activation patterns as one traverses the input space. The fundamental of the bound is based on the assumption that an arrangement of n ReLUs divides a d dimensional space (input) into at most $\sum_{j=0}^{d} \binom{n}{j}$ regions. In general, the bound provides an indication of the representation capacity and hence of the capacity of a given multilayer perceptron architecture.

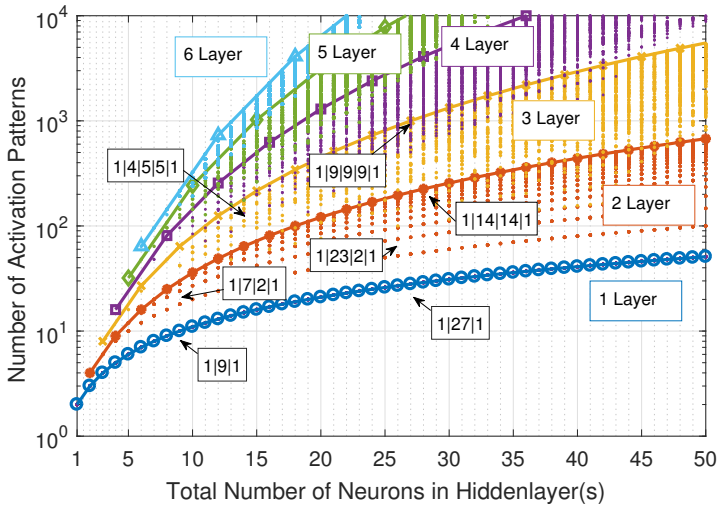


Figure 3.4: According to (3.6), number of activation patterns of a neural network with one input neuron and one to six hidden layers with various neurons and one ouput neuron. Some exemplary designs are labeled with their structures, where the numbers denote the number of neurons.

Fig. 3.4 shows the number of activation patterns related to the total number of neurons in the hidden layers of a shallow neural network and a deep neural network with up to six layers. Each small dot denotes an individual neural network architectural design. The number

of input and output neurons is set to one. Some exemplary designs are labeled with their structure, where the labels denote the number of neurons. As expected the representation capacity grows exponential within the number of hidden layers and that for each depth an architectural trade-off between complexity and representation capacity exists. The optimal architectures are located at the corresponding envelops and are the designs with equal neurons in the hidden layers. In theory they are therefore the preferred choices, however, in practice matched architecture with non-equal distributed neurons can outperform the equal distributed structures in terms of both complexity and performance.

3.1.3 Backward Pass

In this thesis, neural networks are trained in a supervised learning manner, building knowledge from a dataset comprised of inputs and desired outputs. In such a supervised learning manner the neural network learns and approximates the unknown underlying mapping function from inputs to outputs by tuning its parameters. This process involves using an optimization algorithm that searches through a space of possible parameters for a set of weights and biases which minimize a predefined cost function $\mathcal{L}$. However, training deep neural networks is very challenging. In [123] the authors have shown that neural networks are difficult to optimize because they are non-convex (or non-concave) and contain local minima, saddle points and a highly multidimensional space. Methods which solve these problems currently best are gradient-based optimization methods, e.g., gradient descent [124, Chap.2]

$$\boldsymbol{W}^{[l]}_{\text{new}} \leftarrow \boldsymbol{W}^{[l]}_{\text{old}} - \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{W}^{[l]}_{\text{old}}} \tag{3.7}$$

$$\boldsymbol{b}^{[l]}_{\text{new}} \leftarrow \boldsymbol{b}^{[l]}_{\text{old}} - \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{b}^{[l]}_{\text{old}}} \tag{3.8}$$

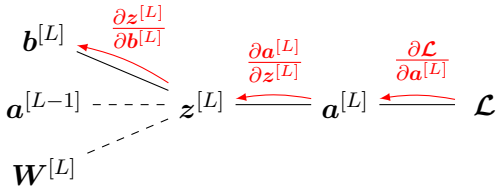


Figure 3.5: Simple computational graph of the last layer of a neural network to illustrate the backpropagation of error algorithm.

where the parameters receive an update proportional to the partial derivatives of the cost function with respect to the current network parameters in each iteration of training. An effective algorithm to compute these gradients is the backpropagation of error algorithm [118, Sec. 6.5]. It computes the particular gradients by iterating backward from the last layer to the first layer and by applying thereby the chain rule.

The derivation of a general cost function with respect to the weights and biases in the L-th layer can be expressed by applying the chain rule as follows

$$\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{W}^{[L]}} = \frac{\partial \boldsymbol{z}^{[L]}}{\partial \boldsymbol{W}^{[L]}} \frac{\partial \boldsymbol{a}^{[L]}}{\partial \boldsymbol{z}^{[L]}} \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{a}^{[L]}} = \boldsymbol{a}^{[L-1]} g'(\boldsymbol{z}^{[L]}) \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{a}^{[L]}}, \tag{3.9}$$

$$\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{b}^{[L]}} = \frac{\partial \boldsymbol{z}^{[L]}}{\partial \boldsymbol{b}^{[L]}} \frac{\partial \boldsymbol{a}^{[L]}}{\partial \boldsymbol{z}^{[L]}} \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{a}^{[L]}} = g'(\boldsymbol{z}^{[L]}) \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{a}^{[L]}}, \tag{3.10}$$

where $g'(\cdot)$ is the derivative of the activation function. The gradient of the previous layers, towards the first layer, are subsequently calculated by propagation backwards. The derivations for $l = L, \ldots, 2$ are defined as follows

$$\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{W}^{[l-1]}} = \frac{\partial \boldsymbol{z}^{[l-1]}}{\partial \boldsymbol{W}^{[l-1]}} \frac{\partial \boldsymbol{a}^{[l-1]}}{\partial \boldsymbol{z}^{[l-1]}} \frac{\partial \boldsymbol{z}^{[l]}}{\partial \boldsymbol{a}^{[l-1]}} \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{z}^{[l]}} = \boldsymbol{a}^{[l-2]} g'(\boldsymbol{z}^{[l-1]}) \boldsymbol{W}^{[l]} g'(\boldsymbol{z}^{[L]}) \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{a}^{[L]}}, \tag{3.11}$$

$$\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{b}^{[l-1]}} = \frac{\partial \boldsymbol{z}^{[l-1]}}{\partial \boldsymbol{b}^{[l-1]}} \frac{\partial \boldsymbol{a}^{[l-1]}}{\partial \boldsymbol{z}^{[l-1]}} \frac{\partial \boldsymbol{z}^{[l]}}{\partial \boldsymbol{a}^{[l-1]}} \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{z}^{[l]}} = g'(\boldsymbol{z}^{[l-1]}) \boldsymbol{W}^{[l]} g'(\boldsymbol{z}^{[L]}) \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{a}^{[L]}}. \tag{3.12}$$

From (3.11) and (3.12) it can be observed that, the computation of the gradients in the previous layer builds on the particular gradients of the post layer. Hence, some particular gradients of the expended chain rule expression have to be evaluated only once.

Training neural networks with gradient-based learning methods and backpropagation encounters the problem of vanishing gradient, i.e, the gradients in the early layers become extremely small and the network cannot learn the parameter within this layer effectively. The main cause of vanishing gradients are the non-linear activation functions. Many non-linear activation functions map their input into a very restricted output range. As a result, there are large regions of the input space which are mapped to a small range. In these regions of the input space, even a large change in the input will produce a small change in the output, hence the gradient is small. The problem of small gradients becomes enhanced in DNN, where layers are stacked on top of each other. Gradients smaller than 1 of deeper layers will get multiplied up to L times to compute gradients of the early layers in an L-layer

network. This reduces the value of the gradient for the initial layers even further. As a results, those layers are not able to learn properly. The depth of a DNN is thus limited and in practice the optimum performance is found in a trade off between number of hidden layers and the effect of vanishing gradients.

3.2 Recurrent Neural Networks

In the previous section we considered neural networks whose connections did not form cycles. If we relax this condition, and allow cyclical connections another type of ANN is obtained, the RNNs. A type in which some of the connections between neurons point backwards. The network connections form thus a directed graph along a temporal sequence, which allows the RNN to exercise temporal dynamic behavior [125, Chap.2.3]. Derived from feed-forward neural networks, which can only map from the input to the output vectors, RNNs take as their input not just the current input vector but also what they have perceived previously in time. The key point is that the recurrent connections allow a memory of previous inputs to persist in the network's internal state, and thereby influence the network's output [113]. It is therefore a specialized neural network for processing a sequence of values [126]. This makes them applicable to tasks such as handwriting [127] or speech recognition [128]. In this section, we provide the basic information of RNNs, which are used within this thesis. For additional information on RNNs, we refer the reader to the textbook of Graves (2012) [113, Chap. 3.2].

3.2.1 Forward Pass

The forward pass of a recurrent layer concur with the feed-forward layer, expect of activations arriving at the hidden layer are from both the current external input and the hidden layers' activations from a previous timestep [128]. With $\boldsymbol{y}(t)$ representing the input vector, the hidden state $\boldsymbol{h}(t)$ at time t is given by

$$\boldsymbol{h}(t) = g(\boldsymbol{W}\boldsymbol{y}(t) + \boldsymbol{U}\boldsymbol{h}(t-1) + \boldsymbol{b}). \tag{3.13}$$

where the hidden state of the previous time step $\boldsymbol{h}(t-1)$ is multiplied by its own hidden-state-to-hidden-state weight matrix $\boldsymbol{U}$, otherwise known as a transition matrix and similar to a Markov chain. The hidden-state-to-hidden-state weight matrix acts as filter and determines how much importance to accord to both the present input and the past hidden state. Because

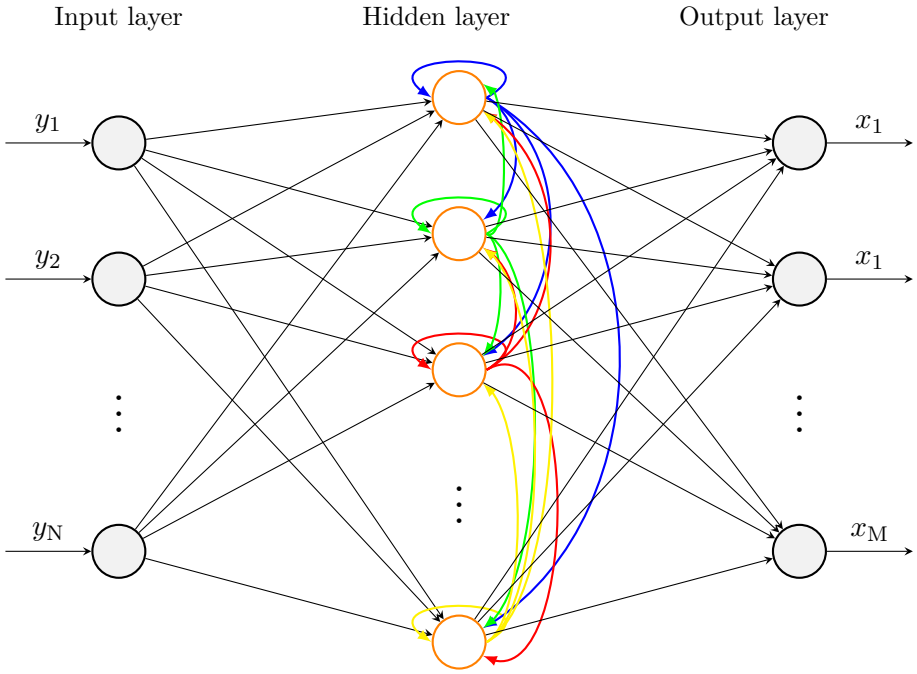


Figure 3.6: Example of recurrent neural network (RNN) with one self connected hidden layer. The colored arrows denote the additional connections of each particular neuron within the layer.

this feedback loop appears at every time step of the series, each hidden state contains traces not only of the previous hidden state, but also of all those that preceded $\boldsymbol{h}(t-1)$ for as long as memory persist.

The differentiable non-linear activation function $g(\cdot)$ is applied on the sum of the weight inputs and hidden states similar to a feed-forward layer. The complete sequence of the hidden activations can be calculated by starting at $t = 1$ and recursively applying (3.13), incrementing t at each step [128]. Note that for these initial values the hidden state matrix $\boldsymbol{h}(0)$ is required. In this thesis, the initial values are always set to zero. However, in [129] the authors have shown that the RNN stability and performance can be improved in some scenarios by using nonzero initial values.

3.2.2 Backward Pass

Just like feed-forward neural networks, RNNs are trained with gradient-based learning methods. Two well-known algorithms have been devised to efficiently calculate weight derivatives

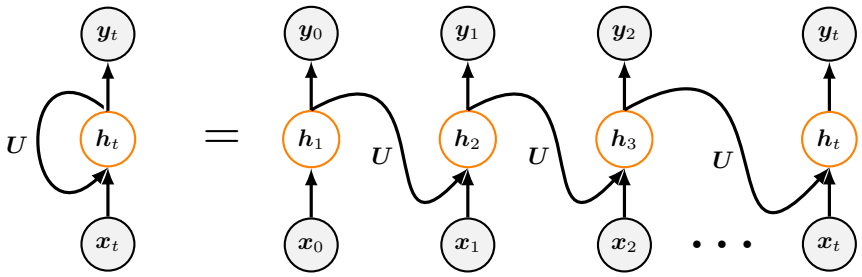


Figure 3.7: An unrolled recurrent neural network with one hidden layer. Each node represents a layer of network units at a single timestep. Note that, the same weights are reused at every timestep.

for RNNs, the real time recurrent learning (RTRL) [130] and the backpropagation through time (BPTT) [131]. In this thesis, BPTT is used, since it is conceptually simpler.

Like standard backpropagation, BPTT consists of a repeated application of the chain rule. The subtlety is that, for recurrent networks, the loss function depends on the activation of the hidden layer not only through its influence on the output layer, but also through its influence on the hidden layer at the next timestep [128]. A way to visualize this and to point out the functionality of BPTT is to consider the update graph. The update graph is formed by unfolding the network along the input sequence. Fig. 3.7 shows an unfolded RNN with one hidden layer. It can be observed that time is expressed by a well-defined ordered series of calculations, linking one time step to the next. Hence, RNNs can be simply considered as a nested composition of feed-forward layers with equal weights and the backpropagation algorithm explained in Section 3.1.3 can be reused to compute the corresponding derivatives of an RNN.

3.3 Optimization Algorithm

Various enhanced optimizers are researched within the last few couples of years to update the weights and biases of the neural network. One of them is the stochastic gradient descent (SGD) optimization algorithm, which iteratively updates the parameters of a model by moving them in the direction of the negative gradient of the loss evaluated on a minibatch.

However, the SGD has trouble navigating through ravines, i.e. areas where the surface curves much more steeply in one dimension than in another [132], which are quite common around local optima [124]. This issue can be overcome by determining the trend of the gradients,

namely by the momentum. Momentum [133] is a method that accelerates convergence towards the relevant direction and dampens oscillations to the irrelevant directions. Popular algorithms in this line of research are adaptive gradient methods, which update the stepsize on the fly according to the gradients received along the way. Theses algorithm have gained widespread use in high dimensional problems for their ability to converge robustly, without the need to fine-tune parameters [134].

In the following, we will outline two optimization algorithms which are used within this thesis to update the weights and biases of the neural network, namely RMSprop and ADAM. A greater overview of other gradient descent optimization algorithms is given by S. Ruder in [124].

3.3.1 RMSprop

RMSprop is an adaptive learning rate method proposed by Geoff Hinton. The main idea behind RMSprop is to divide the learning rate by an exponentially decaying average of squared gradients in order to hinder the search in the direction of oscillation.

$$\boldsymbol{W}_{\mathrm{t+1}}^{[l]} = \boldsymbol{W}_{\mathrm{t}}^{[l]} - \frac{\eta}{\sqrt{\boldsymbol{v}_t + \epsilon}} \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{W}_{\mathrm{t}}^{[l]}} \tag{3.14}$$

with

$$\boldsymbol{v}_t = \rho \boldsymbol{v}_{t-1} + (1-\rho) \left(\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{W}_{\mathrm{t}}^{[l]}} \right)^2 \tag{3.15}$$

where η is the initial learning rate and v_t the estimates of the uncentered variance of the gradients [124]. Hinton suggests ρ to be set to 0.9, while a good default value for the learning rate η is 0.001.

3.3.2 Adam

Adaptive Moment Estimation (Adam) [135] is another method that computes adaptive learning rates for each parameter. In addition to RMSprop, which stores an exponentially decaying average of past square gradients $\boldsymbol{v}_t$, Adam keeps as well an exponentially decaying average of the past gradients [124, Sec. 4.6],

$$\boldsymbol{W}_{\mathrm{t+1}}^{[l]} = \boldsymbol{W}_{\mathrm{t}}^{[l]} - \eta \frac{\boldsymbol{m}_t}{\sqrt{\boldsymbol{v}_t + \epsilon}} \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{W}_{\mathrm{t}}^{[l]}} \tag{3.16}$$

with

$$\boldsymbol{m}_t = \delta \boldsymbol{m}_{t-1} + (1-\delta)\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{W}_{\mathrm{t}}^{[l]}}. \tag{3.17}$$

where $\boldsymbol{m}_t$ is the estimate of the mean of the gradients. Default values for ρ and δ are 0.999 and 0.9, receptively.

Chapter 4

Neural Network Assisted Geometric Shaping

For a given transmit bandwidth, optical communication systems achieve highest data rates by increasing their spectral efficiency with higher modulation orders. To achieve maximum performance, it is essential to match the modulation to the actual transmit channel conditions. End-to-end learning is a machine learning method for designing high-order optimized modulations formats and to realize geometrical constellation shaping for various channel scenarios. It enables joint optimization of the mapper and demapper to learn optimal symbol constellations. The mapper and demapper are implemented as DNNs consisting of various hidden layers, that realize the functionality of so-called auto-encoders [26]. The DNN structure enforces dimension reduction and makes the auto-encoder map its input into an inphase (I) and quadrature quadrature (Q) transmit component. A limitation of the end-to-end design is that the required gradients of the cost function between the mapper and demapper must be known. In a real transmission system, the transfer function of the channel is not available in analytical form and the mapper cannot be optimized by backpropagation of the gradient. Therefore, the neural network based auto-encoders are usually trained offline by considering idealized channel models in additional layers [26, 136]. In [26] the authors use an auto-encoder to improve communication over an AWGN channel, while in [136] the auto-encoder is combined with the Gaussian noise (GN)-model to optimize the constellation design for amplified long-haul scenarios. In both cases the training of the auto-encoders relies on one-hot encoded vectors, which led to constellations' optimized links with respect to the symbol error ratio (SER).

In contrast, in this proposal, the bit stream is directly fed to the DNN mapper to jointly optimize the positions of the constellation points and the binary labeling and hence the BER. This is a significant enhancement, because the performance of a given constellation

geometry depends critically on the bit mapping. It is well known that a Gray-mapping or a quasi Gray-mapping is necessary in case of BICM. The approach also enables the realization of binary mappings with label extension [137] and the use of the binary cross-entropy as cost function.

We demonstrate the concept by designing optimized constellations both for amplified and unamplified optical links with and without label extension. In particular, we consider next generation DCI compatible coherent 800 Gb/s and 1 Tb/s systems using 92 GBd DP-32QAM and 82 GBd DP-128QAM with 15% FEC. The gain of the learned 32QAM and 128QAM modulation formats over standard cross-32QAM and cross-128QAM is confirmed by transmission experiments with offline processing.

4.1 Auto-Encoder and Decoder

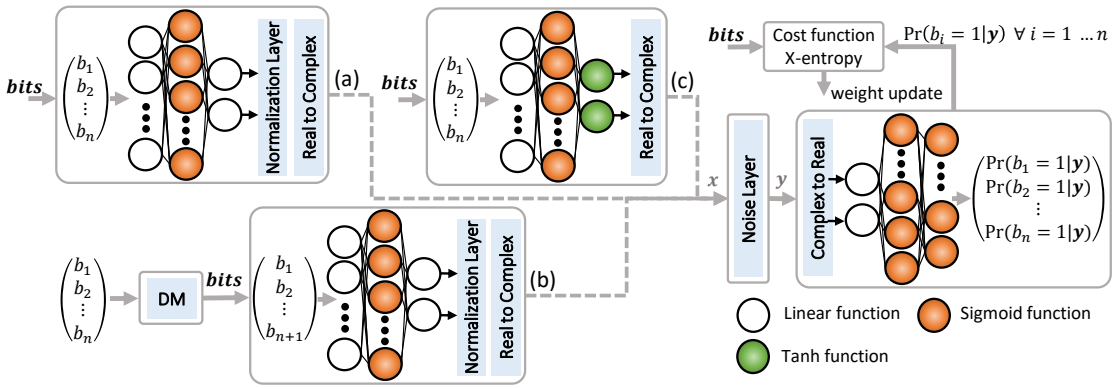
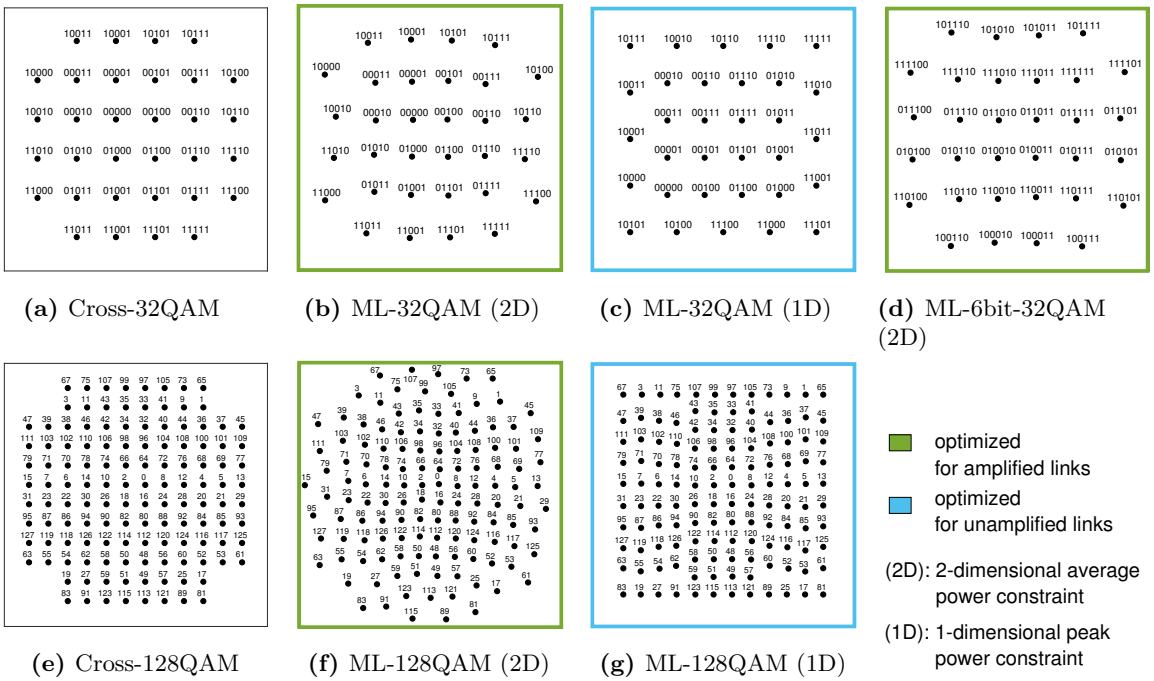


Figure 4.1: End-to-end communication as an autoencoder with various transmitters adapted for different use cases: (a) amplified link, (b) amplified link with label extension and distribution matcher (DM) and (c) unamplified link.

Fig. 4.1 shows the proposed DNN mappers for three different communication use cases, as well as a single DNN demapper, which is suitable for all scenarios. Version **(a)** represents the structure for communication systems using optical amplifiers within the link. In this scenario, the power budget applies to a two-dimensional constellation, i.e. to the I and Q components jointly. This makes it a two-dimensional average power constraint, which is met by a normalization layer. Version **(b)** enables geometric shaping with label extension for an amplified link. Label extension, introduced by Smith [137], is a technique where an additional bit is added as the least significant bit. The cardinality of the alphabet is thereby increased

and more codewords are available per symbol. The more labeling options, the higher the probability for perfect Gray-labeling of odd constellations, where normally perfect Gray mapping is no longer possible.

Version **(c)** represents the structure for optimizing constellation diagrams for unamplified links, i.e., a link without optical amplifiers between transmitter and receiver. Unamplified links with large link budget are common in datacom and short-reach telecom applications. For instance, within the 400ZR project, the Optical Internetworking Forum (OIF) has the objective to define a 40 km (ER) unamplified application code. It is therefore expected that a new proposal will be necessary soon for the upcoming 800ZR systems. However, such links require high transmitter optical power to increase link budget and performance. As described in Section 2.1.4, in a coherent optical transmitter, the power of the transmit laser is equally split between two orthogonal polarization planes X and Y and for each polarization plane between the I and Q components. Each tributary (XI, XQ, YI, YQ) is modulated via a Mach-Zehnder modulator that carves the desired signal shape by attenuating the laser. As a consequence, the maximum power per dimension is limited. If no optical amplifier follows, as in the case of an unamplified link, this calls for a one-dimensional peak power constraint rather than a two-dimensional average power constraint. In this scenario, it is therefore



(a) Cross-32QAM **(b)** ML-32QAM (2D) **(c)** ML-32QAM (1D) **(d)** ML-6bit-32QAM (2D)

(e) Cross-128QAM **(f)** ML-128QAM (2D) **(g)** ML-128QAM (1D)

Figure 4.2: Modulation formats.

important to minimize the peak-to-average power ratio (PAPR) for each one-dimensional signal component in order to achieve a higher transmit power and a correspondingly higher link budget.

The resulting optimized constellation schemes are plotted in Fig. 4.2. All auto-encoder structures are trained by minimizing the binary cross-entropy using the back-propagation linked with the RMSprop optimization algorithm [138]. Fig. 4.2a and 4.2e depict the common 32QAM and 128QAM cross schemes. Fig. 4.2b, 4.2d and 4.2f represent the learned constellations for amplified links using the structures (a) and (b) of Fig.4.1, while Fig. 4.2c and 4.2g show the learned constellations for unamplified cases using structure (c).

4.2 Numerical Study

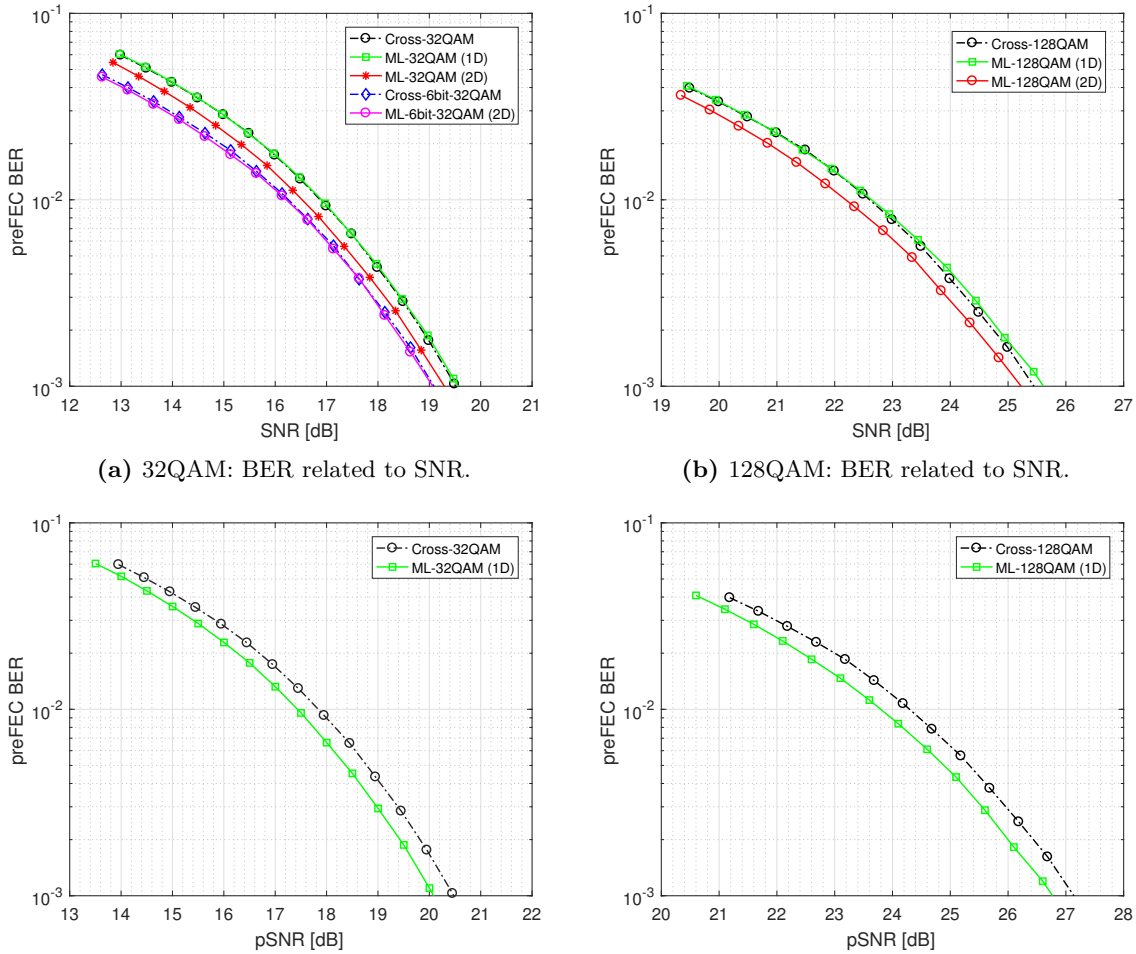
The proposed constellations schemes are first evaluated numerically on an AWGN channel. Fig. 4.3a and 4.3b show the BER as a function of the SNR, benchmarking the 32QAM and 128QAM constellation schemes for amplified links. It can be observed, that in comparison to conventional Cross-32QAM, an improvement of 0.45 dB and 0.90 dB is achieved at the FEC limit of 2×10^{-2} when ML-32QAM (2D) and ML-6bit-32QAM (2D) is employed, respectively. However, in comparison to Cross-32QAM with label extension, only a diminished gain of 0.1 dB is achieved by ML-6bit-32QAM (2D). This indicates that Cross-32QAM with label extension is nearly optimal, due to the Gray-labeling. Furthermore, in case of 128QAM, the proposed ML-128QAM (2D) exhibits a gain of ca. 0.4 dB in terms of SNR.

Figs. 4.3c and 4.3d show the BER as a function of the peak SNR (pSNR), benchmarking the 32QAM and 128QAM constellation schemes for unamplified links. The pSNR is a quality metric for constellation diagrams in combination with unamplified links. It takes the SNR value and hence the noise sensitivity as well as the peak power of the constellation into account. The definition of pSNR is given by [139]

$$\mathrm{pSNR} = \frac{\max\{|X|^2\}}{\sigma^2} \tag{4.1}$$

From Fig. 4.3a, it can be observed that the ML-32QAM (1D) constellation achieves in terms of noise sensitivity the same performance as Cross-32QAM. However, ML-32QAM (1D) has a one-dimensional PAPR of 3.52 dB, which is 0.5 dB lower than the PAPR of cross-32QAM. Hence, it outperforms Cross-32QAM by 0.5 dB in terms of pSNR in Fig. 4.3c. The same behavior is also observed if 128QAM is employed. ML-128QAM (1D) exhibits a 0.58 dB

lower one-dimensional PAPR than Cross-128QAM and hence outperforms it in terms of pSNR, due to an optimal tradeoff between noise robustness and PAPR.



(a) 32QAM: BER related to SNR.

(b) 128QAM: BER related to SNR.

(c) 32QAM: BER related to SNR.

(d) 128QAM: BER related to pSNR.

Figure 4.3: BER versus SNR for amplified links (2D) and pSNR for unamplified links (1D).

4.3 Experimental Investigation

A coherent single-carrier dual-polarization (DP) transmission system over a single-mode fiber (SMF) is employed to experimentally evaluate the performance of the proposed constellation schemes. The setup is shown in Fig. 4.4.

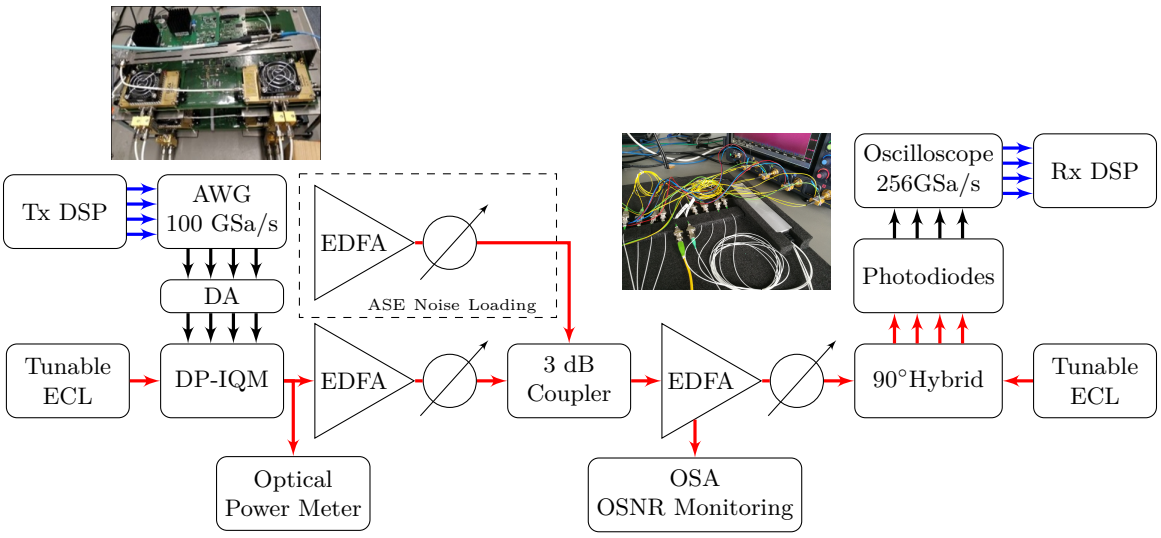


Figure 4.4: Schematic of optical back-to-back transmission system with ASE noise loading used for the experimental investigations. AWG: arbitrary waveform generator, DA: driver amplifiers, DP-IQM: dual polarization inphase-quadrature modulator, EDFA: Erbium-doped fiber amplifier, ECL: external cavity laser. The blue arrows indicate the logical part while the black and red arrows the electrical and optical part respectively. The photos show the DAC, 90°-hybrid, PDs and oscilloscope in use.

For a net bitrate of 800 Gb/s DP-32QAM without label extension, 920 Gb/s of pseudo random data including 15% overhead for FEC are transmitted at 92 GBd. For a net bitrate of 800 Gb/s with LE, using 6-bit labels instead of 5-bit labels, the baudrate has to be increased to $\mathrm{SE}_5/\mathrm{SE}_6 \cdot 92$ GBd $= 94.9$ GBd when using the same 15% overhead FEC code, due to the lower spectral efficiency of ML-LE-32QAM, as shown in Fig. 4.5. The spectral efficiencies are calculated as follows [66, Eq. 9]

$$\mathrm{SE}_m = H - m \cdot (1 - R_\mathrm{C}) \tag{4.2}$$

where H denotes the entropy, m the number of bits per symbol and R_C the FEC code rate. For a net bitrate of 1 Tb/s DP-128QAM, 1150 Gb/s of pseudo random data are transmitted at 82 GBd. The measurements were performed in a BtB configuration at 1550 nm with ASE

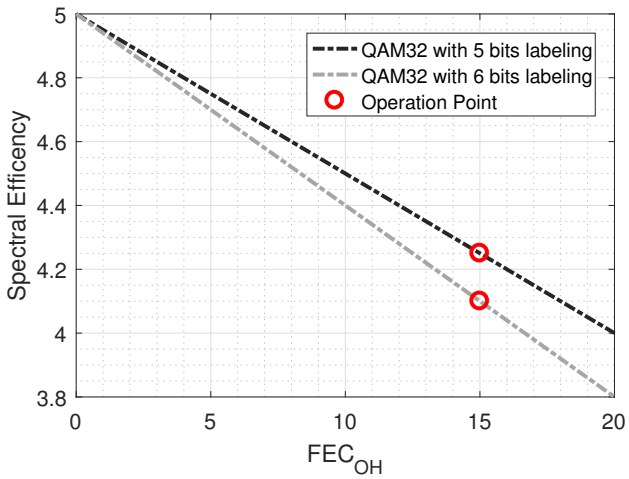
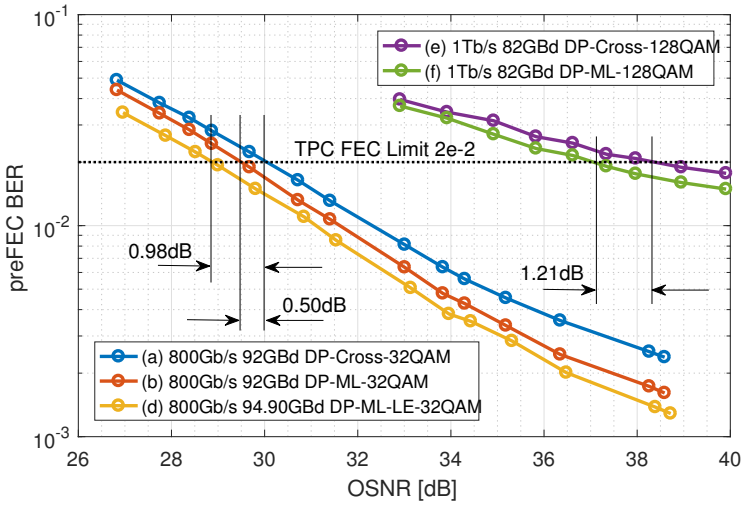


Figure 4.5: Spectral efficiency of 32QAM with and without label extension related to the assumed FEC overhead. The red circle indicate the considered operation point.

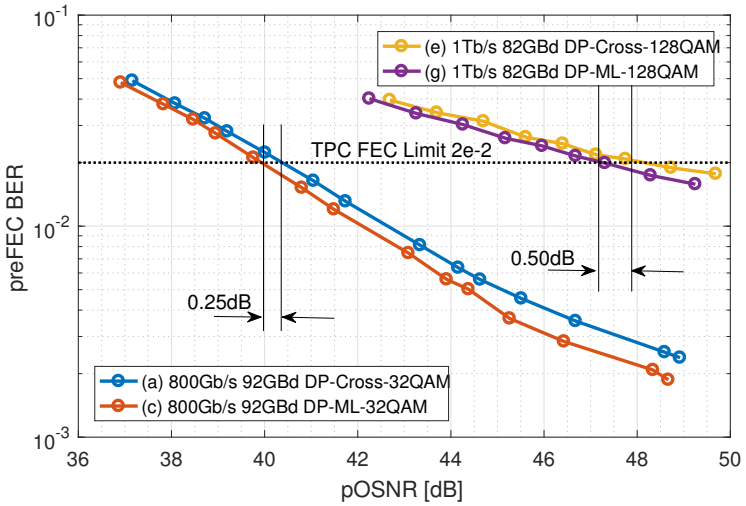
noise loading, in order to compare preFEC BERs at varying OSNR values. The electrical signals are generated by a 100 GSa/s Micram DAC with 40 GHz 3 dB-bandwidth and 4.5 ENOB. Subsequently, the signals are amplified by drivers (SHF S804A) which exhibit a 3 dB-bandwidth of 60 GHz. In the optical domain, an external cavity laser (ECL) source with 1 kHz linewidth and a wavelength of 1550 nm generates a continuous wave signal which is modulated by a DP-I/Q Modulator (Fujitsu-FTM7992HM-32 GHz). At the receiver side, the optical signal is combined with ASE noise generated by an EDFA and then amplified. After the 70 GHz photodiodes, the electrical signals are captured by a 110 GHz bandwidth real-time oscilloscope operating at 256 GSa/s. In order to evaluate the constellation schemes on unamplified links, the optical output power of the DP-I/Q Modulator is measured with an optical power meter.

4.3.1 Optical Back-to-Back Performance Evaluation

The performance of 800 Gb/s and 1 Tb/s over an optical BtB channel, using the constellation schemes of Fig. 4.2, is shown in Fig. 4.6. On the top, Fig. 4.6a shows the preFEC BER as a function of the OSNR, comparing the constellation schemes for amplified links. ML-32QAM (2D) outperforms the common Cross-32QAM by 0.50 dB in OSNR at the TPC FEC limit of $2 \cdot 10^{-2}$. A benefit of 1.0 dB in OSNR can be observed with ML-LE-32QAM (2D) including label extension. The drawback of label extension is the higher FEC throughput and hence the higher baudrate, which may imply larger complexity and power. Regarding the transmission



(a) preFEC BER related to OSNR.



(b) preFEC BER related to pOSNR.

Figure 4.6: Performance of 800 Gb/s and 1 Tb/s over an optical BtB channel using the constellation schemes of Fig. 4.2. On the top, the preFEC BER as a function of the OSNR is shown, comparing the constellation schemes for amplified links (2D). On the bottom, the preFEC BER as a function of the peak OSNR (pOSNR) is shown, comparing the common cross constellations and the learned constellation schemes for unamplified links (1D).

of 1 Tb/s, the proposed ML-128QAM (2D) outperforms the Cross-128QAM by 1.21 dB at the FEC limit. The benefit is emphasized by the high error floor. It is expected that in the linear regime the gain of ML-128QAM (2D) will settle in the range of 0.4 dB, as in the numerical study.

On the bottom, Fig. 4.6b shows the preFEC BER as a function of the peak OSNR (pOSNR), comparing the common cross constellations and the learned constellation schemes for unamplified links. The pOSNR is considered as follows,

$$\mathrm{pOSNR_{dB}} = \mathrm{OSNR_{dB}} + \mathrm{PAPR_{dB}} \tag{4.3}$$

$$= \mathrm{OSNR_{dB}} + 10\log_{10}\frac{\max\{|X|^2\}}{\mathbb{E}[|X|^2]}. \tag{4.4}$$

It takes the OSNR value and hence the noise sensitivity as well as the optical transmitter output power into account. The ML-32QAM (1D) exhibits a 0.50 dB lower analytical PAPR than Cross-32QAM having equal receiver sensitivity. Consequently, the ML-32QAM (1D) achieves a theoretical 0.5 dB higher transmit power and, in a dark fiber scenario, a correspondingly higher received power. However, in the high baudrate bandwidth-limited experimental setup, it can be observed that the benefit shrinks and only a diminished gain of 0.25 dB at the FEC limit can be achieved. Due to the lower baudrate of 82 GBd, the benefit of the 0.58 dB lower analytical PAPR of ML-128QAM (1D) is higher. A gain of 0.50 dB in peak OSNR at the FEC limit is achieved in comparison to the common Cross-128QAM constellation.

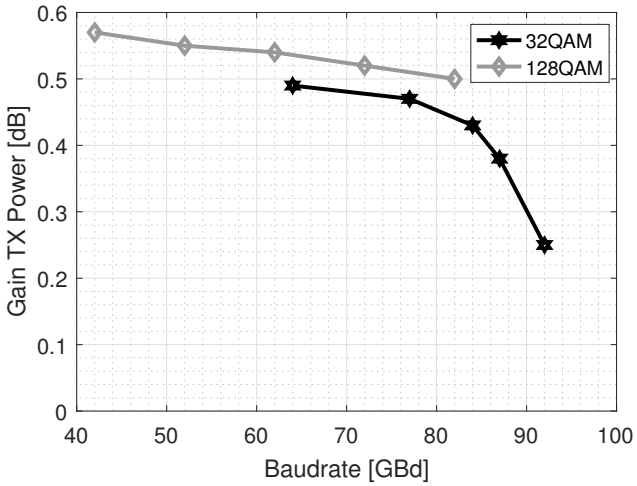


Figure 4.7: Transmission power gain of DP-ML-32QAM (1D) and DP-ML-128QAM (1D) related to DP-Cross-32QAM and DP-Cross-128QAM, respectively.

4.4 Summary

In this chapter, novel auto-encoder structures based on deep neural networks are introduce to design modulation formats optimized for lowest bit-error rate over amplified and unamplified links. Besides the geometry of the symbol constellation, the machine learning optimization method optimizes also the bit mapping with and without label extension. The learned modulation formats are experimentally evaluated and gains of up to 1.0 dB and 1.2 dB in OSNR and 0.25 dB and 0.50 dB in pOSNR are demonstrated at the FEC limit for 800 Gb/s DP-32QAM and 1 Tb/s DP-128QAM, respectively. The simulated and experimental results show improved performances, but also highlight the challenges in matching the offline learning architecture to the real system.

Chapter 5

Non-linearity Compensation for Optical Transmission

Figure 5.1: General communication black-box model.

In modern optical communication systems, SD-FEC and QAM schemes are key technologies for realizing high spectral efficiencies (SE) [62]. A core component of such an SD-FEC BICM [50] system is the soft-demapper, which provides, given the observation of the (symbol-level) channel output $y(i)$, bit-level reliability values $\ell(i) = [\ell_1(i), \ldots, \ell_m(i)]$, so-called conditional log-likelihood ratio (LLRs) [140, Chap. 16], suitable as SD-FEC decoding input. Their quality thus directly affects the overall system performance. Classical approaches to evaluate the L-values, as the analytical computation of the a-posteriori LLR [106, Sec. 3.3] or the less complex MLA [106, Sec. 3.3.3], assume ideal channel compensation of the channel impairments and thus AWGN. However, optical communication systems comprise bandwidth limited O/E components with non-linear transfer characteristics, which results in linear and non-linear ISI as well as colored noise effects after equalization. Their impact on the achievable rate is aggravated towards higher data rates, especially in high speed optical transmission systems where symbol rates exceed the provided 3 dB-bandwidth of the components. Therefore, computing the L-values under the AWGN assumption implies a performance penalty.

In recent years, machine learning methods, especially DNNs, have demonstrated excellent performance gains in various applications. In the optical communication community, DNNs have been applied in several publications on short-reach intensity modulation (IM)/direct detection (DD) systems to characterize O/E components [14–16] as well as on coherent

links to tackle fiber and component non-linearities [17–21]. In literature [17, 29, 33, 141], where machine learning is applied for L-values computation, the authors have shown that DNNs are a convenient tool to predict probabilities and hence to learn soft demodulation schemes.

This chapter builds upon the aforementioned contributions and investigates two DNN architecture concepts to combat these strong impairments. The first concept relies on feed forward 1D-CNNs, conveniently represented by feed forward TDNNs. Equal to conventional linear and non-linear equalizer designs, e.g., FIR filters and VNLEs, the memory effects of the channel, in particular of the components, are considered by adding time delayed versions of the observed channel output. This allows the feed forward designs to have a finite dynamic response to time series input data and to describe causal time-invariant non-linear systems with finite fading memory. In this context, special attention is placed on hard and soft demapping and on the corresponding cost functions. The second concept relies on BRNNs. In contrast to the TDNN, the BRNN architecture enables internal memory states corresponding to IIR filters. For each concept we describe the main rationale behind the approach and compare its performance experimentally with classical DSP approaches to combat linear and non-linear ISI as well as colored noise effects.

5.1 Non-linear Feed-Forward Equalizers

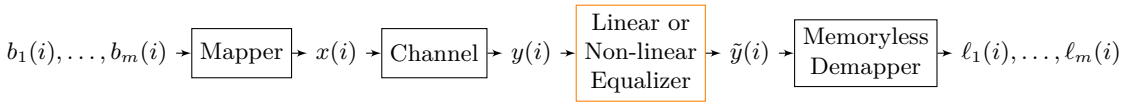


Figure 5.2: General communication model including a linear or non-linear equalizer followed by a memoryless demapper.

An approach to combat residual distortion is to extend the receiver DSP stack by a symbol-level linear/non-linear post-equalizer. The equalizer attempts to reverse the residual ISI effects before the memoryless demapper, which assumes an AWGN channel and hence ideal compensation of the impairments. In this section, a non-linear equalizer is realized by a TDNN and as reference by a VNLE. As discussed in Section 2.3.2, the VNLE is a popular approach, which has proven to be very effective against component non-linearities [11, 12]. In principle, it can be tailored to match any differentiable non-linear system by choosing a high enough polynomial order and memory depth [142]. However, excessive extension for optimal

performance can become a significant downside, as the architectural complexity of VNLEs increase exponentially. Here, machine learning may be capable to exploring solutions.

5.1.1 Time Delay Neural Network Equalizer

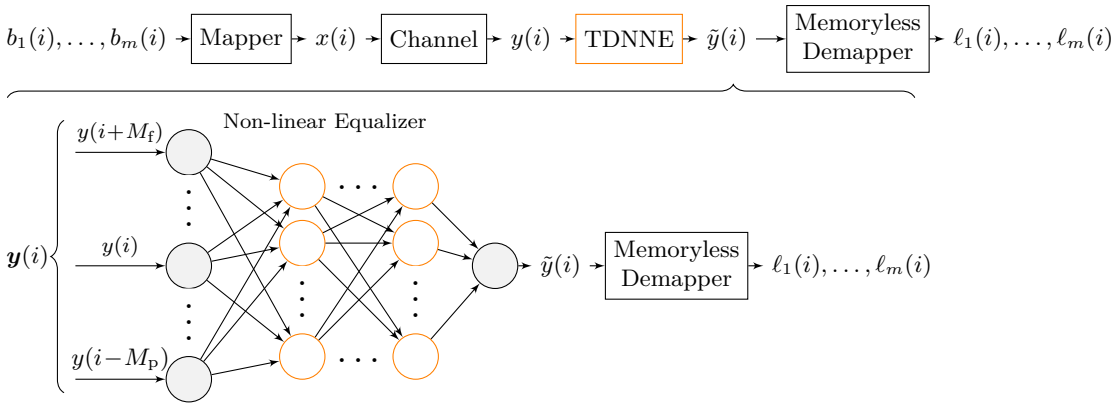


Figure 5.3: Block diagram of a general communication model including a TDNNE with multiple hidden layers and an accompanied memoryless demapper.

Fig. 5.3 shows the block diagram of a general communication model including a TDNN equalizer (TDNNE) with multiple hidden layers. Similar to the VNLE design, the memory of the channel and the components are considered by adding time delayed versions of the observed signal. This allows the TDNNE to have similar to the VNLE a finite dynamic response to time series input data and to describe causal time-invariant non-linear systems with finite fading memory. While the VNLE memory length of each particular order can be adjusted individually, the TDNNE memory length M is a single parameter. Let $\boldsymbol{y}(i) = [y(i-M_\mathrm{p}), \ldots, y(i), \ldots, y(i+M_\mathrm{f})]$ and $\tilde{y}(i)$ denote the delayed symbol-level input vector and scalar symbol-level output, respectively, the TDNNE with L-layers is given by

$$\boldsymbol{a}^{[0]}(i) = \boldsymbol{y}(i), \tag{5.1}$$

$$\boldsymbol{a}^{[l]}(i) = g(\boldsymbol{W}^{[l]}\boldsymbol{a}^{[l-1]}(i) + \boldsymbol{b}^{[l]}), \quad l = 1, \ldots, L \tag{5.2}$$

$$\tilde{y}(i) = \boldsymbol{a}^{[L]}(i) \tag{5.3}$$

where $\boldsymbol{a}^{[l]}$ denotes the output vector of the l-th layer and $\boldsymbol{W}^{[l]}$ and $\boldsymbol{b}^{[l]}$ the weight matrices and bias vectors, respectively. For the activation function g, we use the non-linear function

$g(x) = \tanh(x)$ for the hidden layers and the linear function $g(x) = x$ for the output layer. Eq. (5.1) denotes the input layer and (5.2) is executed successively for layers $l = 1, 2, \ldots, L$ to obtain the single output $\hat{x}(i)$ in (5.3). In the following, (5.1), (5.2), and (5.3) will be referred to by

$$\tilde{y}(i) = f_{\text{TDNNE}}(y(i), \boldsymbol{W}^{[1]}, \ldots, \boldsymbol{W}^{[L]}, \boldsymbol{b}^{[1]}, \ldots, \boldsymbol{b}^{[L]}). \tag{5.4}$$

5.1.2 Mean Square Error for Training Non-linear Equalizer

Before operations, the parameters of the VNLE and the TDNNE have to be identified, i.e., configured upon training data, in order to match the non-linearities of interest. The target of the equalizers during the training phase is to find the maximum-a-posterior (MAP) estimation of the parameters which gives the highest conditional probability of the target symbols $x(i)$ given the training data output $\tilde{y}(i)$, namely

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} P(x(i)|f_{\text{NLE}}(y(i), \boldsymbol{\theta})) \tag{5.5}$$

$$= \arg\max_{\boldsymbol{\theta}} P(x(i)|\tilde{y}(i)), \tag{5.6}$$

where θ represents the kernels or the weights and biases and $f_{\text{NLE}}(y(i), \boldsymbol{\theta})$ the corresponding non-linear function of the equalizer. A simple approach to the problem of determining the parameters is to follow [117, Sec. 5.2] and to make an analogy with the linear models for regression [117, Sec. 1.2.5]. In particular, assuming an Gaussian distribution of the error between the equalizer output and the target value with zero mean and variance σ^2, i.e., $X = \hat{Y} + Z$, $Z \sim \mathcal{N}(0, \sigma^2)$. The corresponding cost function can then be derived as follows,

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x(i) - f_{\text{NLE}}(y(i), \boldsymbol{\theta}))^2}{2\sigma^2}\Big) \tag{5.7}$$

$$\overset{(1)}{=} \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \frac{1}{2}\Big(x(i) - f_{\text{NLE}}(y(i), \boldsymbol{\theta})\Big)^2 \tag{5.8}$$

$$= \arg\min_{\boldsymbol{\theta}} \mathcal{L}(x(i), y(i), \boldsymbol{\theta}), \tag{5.9}$$

where in step (1) the quadratic cost function is obtained by applying the logarithm on (5.7). Consequently, minimizing the mean square error (MSE) corresponds to maximizing the a-posteriors of the weights. In the following it will be shown that the assumption of Gaussian noise, assuming the posterior being Gaussian, after non-linear MSE equalization is not valid. In Section 5.2 we are therefore going to revise criterion (5.7).

Since identifying discrete-time Volterra series from training signals can be considered as solving a set of linear equations, iterative as well as standard least-squares (LSs) approaches based on linear algebra are appropriate to solve (5.9). Iterative approaches like least mean square (LMS) [143, Sec. 3] or recursive least squares (RLSs) [144, Sec. III] are suitable, if channel dynamics call for frequent kernel update. Standard LS approaches like pseudo-inverses [145] are suitable, if a sufficiently large data set is available and if channel dynamics are nearly static. They invert matrices, built from transmitted and received training data [146]. It has been shown [147] that these matrices can be ill-conditioned and therefore become almost singular with growing VNLE complexity. As a result, the computation of its inverse is prone to large numerical errors. This property calls for a trade-off between matrix conditioning and the number of kernels for maximum performance in each particular scenario.

In comparison to the VNLE, only iterative training approaches are appropriate for the TDNNE when more than one hidden layer is considered. Under this framework, the iterative training is commonly based on gradient descent in combination with the backpropagation algorithm [148][118, Sec. 6.5] to minimize Eq. (5.9). The weights and biases are updated by shifting previous values towards the gradient descent of the iteratively calculated loss function. To move gently towards the global minimum an enhanced optimization algorithm such as the adaptive moment estimation (ADAM) [135] can be applied (see Chapter 3).

5.1.3 Volterra Series and Neural Networks

While VNLEs represent the solutions of non-linear differential systems based on its Volterra series and hence model non-linearities with polynomials, TDNNEs compute the solutions to a large class of general non-linear systems on basis of the non-linear activation functions. To depict the relation and the capability of replacing VNLEs by TDNNEs, D. I. Soloway et al. [149] introduced the approach of expanding the trained TDNNE into a Volterra series. This is possible, if the activation functions and hence their compositions are infinitely differentiable, e.g., tanh. The expansion enables a comparison of the linear and non-linear kernels of both equalizers based on the Volterra series. The dependency between the TDNNE parameters and the Volterra kernels is given at a specific input vector $\boldsymbol{y_0}$ by the partial derivations. For order one and two, the gradient and Hessian matrix are respectively

$$\boldsymbol{h}_1 = \nabla f_{\text{TDNNE}}(\boldsymbol{y}, \boldsymbol{W}^{[1]}, \ldots, \boldsymbol{W}^{[L]}, \boldsymbol{b}^{[1]}, \ldots, \boldsymbol{b}^{[L]})\Big|_{\boldsymbol{y}=\boldsymbol{y_0}}, \tag{5.10}$$

$$\boldsymbol{h}_2 = \text{Hess}\left[f_{\text{TDNNE}}(\boldsymbol{y}, \boldsymbol{W}^{[1]}, \ldots, \boldsymbol{W}^{[L]}, \boldsymbol{b}^{[1]}, \ldots, \boldsymbol{b}^{[L]})\right]\Big|_{\boldsymbol{y}=\boldsymbol{y_0}}, \tag{5.11}$$

where

$$\boldsymbol{h}_1 = [h_1(-M_\text{f}), \ldots, h_1(M_\text{p})], \tag{5.12}$$

$$\boldsymbol{h}_2 = \begin{bmatrix} h_p(-M_\text{f}, -M_\text{f}) & \ldots & h_p(-M_\text{f}, M_\text{p}) \\ \vdots & \ddots & \vdots \\ 0 & \ldots & h_p(M_\text{p}, M_\text{p}) \end{bmatrix}. \tag{5.13}$$

The obtained matrix which represents the 2nd order is an upper triangular matrix according to (2.14). Terms like $y(i)y(i-1)$ and $y(i-1)y(i)$ are identical, so that $h_2(0,1)$ and $h_2(1,0)$ can be combined to one unique kernel. Similarly, for order p, the kernel is

$$\boldsymbol{h}_p = \frac{1}{p!}\nabla^p f_{\text{TDNNE}}(\boldsymbol{y}, \boldsymbol{W}^{[1]}, \ldots, \boldsymbol{W}^{[L]}, \boldsymbol{b}^{[1]}, \ldots, \boldsymbol{b}^{[L]})\Big|_{\boldsymbol{y}=\boldsymbol{y_0}}. \tag{5.14}$$

Generalizing (5.12) and (5.13), we can represent the order p kernel by the set

$$\boldsymbol{h}_p = \{h_p(s_s, \ldots, s_p) \colon -M_\text{f} \leq s_1 \leq s_2 \leq \cdots \leq s_p \leq M_\text{p}\}. \tag{5.15}$$

5.1.4 Experimental Evaluation

This section outlines the measurement setup, as used to evaluate the non-linear compensation performance of the TDNNE and VNLE on basis of identical offline data. The quality of the obtained signal is evaluated in terms of hard-decision and soft-decision.

5.1.4.1 Measurement Setup

A coherent single carrier transmission system over an SMF is employed to experimentally evaluate the performance of the TDNNE. The performance is benchmarked against an VNLE on basis of dual polarization 88 GBd 16QAM BtB offline captures, where O/E component non-linearities dominate. The experimental setup is shown in Fig. 5.4. The measurements

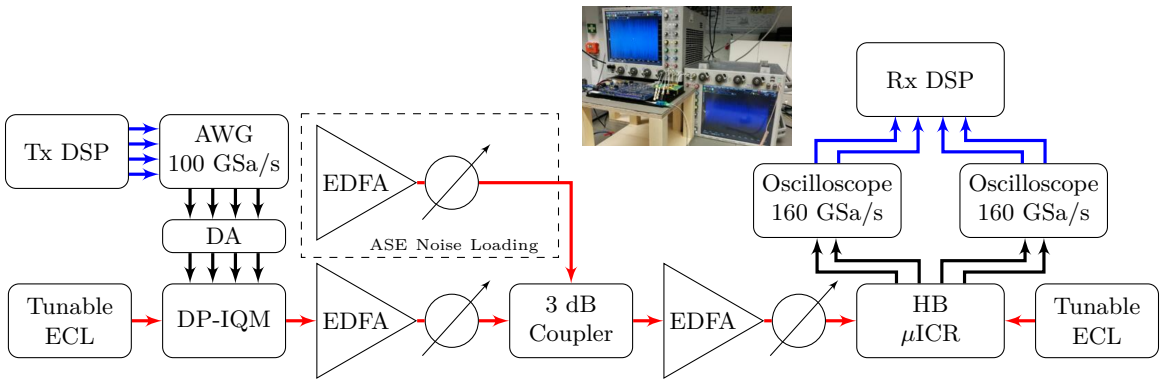


Figure 5.4: Schematic of optical back-to-back transmission system with ASE noise loading used for the experimental investigations. AWG: arbitrary waveform generator, DA: driver amplifiers, DP-IQM: dual polarization inphase-quadrature modulator, EDFA: Erbium-doped fiber amplifier, ECL: external cavity laser. The blue arrows indicate the digital part while the black and red arrows the electrical and optical parts respectively. The photo shows the high bandwidth (HB)-μICR and the two 2-channel oscilloscopes where one is upside down to ensure short RF connections.

were performed BtB at 1550 nm with ASE noise loading, in order to compare performances at varying OSNR levels.

The electrical signals are generated by four 100 GSa/s Micram DAC4 with 40 GHz bandwidth and 4.5 effective number of bits (ENOB). Subsequently, the signals are amplified by four SHF S804A DAs which exhibit 60 GHz 3 dB-bandwidth. In the optical domain, a tunable 100 kHz ECL source generates a continuous wave signal which is modulated by a LiNbO3 DP-I/Q Modulator (Fujitsu-FTM7992HM) with 32 GHz 3 dB-bandwidth. The modulated optical signal is then combined with ASE noise generated by an EDFA and amplified by an additional receiver-side EDFA. The optical receiver consists of a NeoPhotonics 64 GBd class-40 high bandwidth micro integrated coherent receiver (HB-μICR) with 40 GHz 3 dB-bandwidth. The obtained electrical signals are digitized using two 2-channel 160 GSa/s Keysight Infiniium real-time oscilloscopes with 63 GHz 3 dB-bandwidth where one is placed upside down to ensure short RF connections from the HB-μICR as shown on the photo in Fig. 5.4.

In order to compensate the potential mixed non-linearities and ISI effects from the O/E components, the receiver DSP stack in Fig. 5.5 includes next to the classical coherent signal recovery blocks, see Section 2.3.2, four VNLEs and four proposed TDNNEs. Both schemes operate independently on the real dimension with 1 sample per symbol (SpS) and consider four past and zero future memory symbols. Regardless of the type of non-linear equalizer in use, training on the particular non-linearities is essential before deployment. The train-

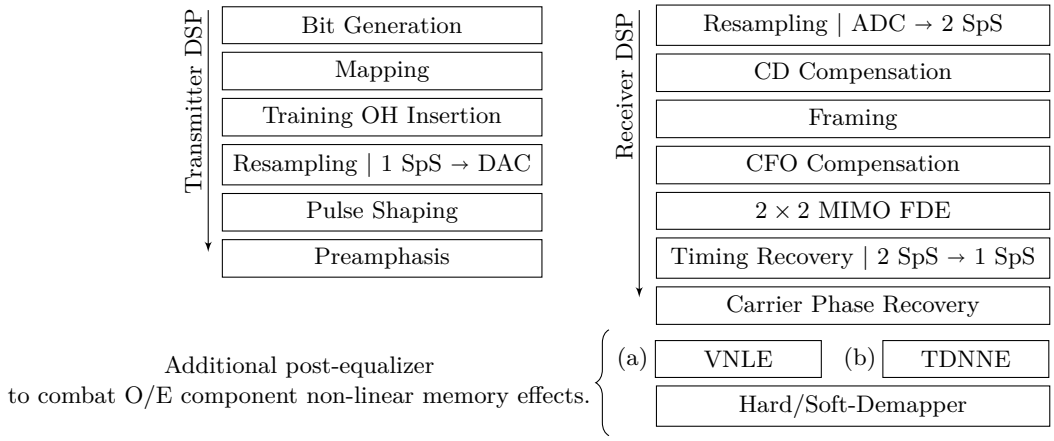
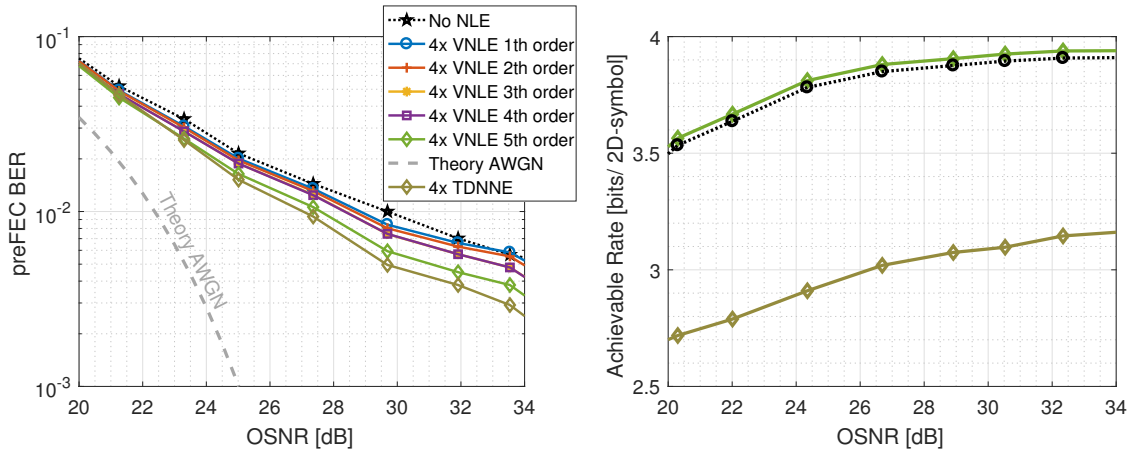


Figure 5.5: Block diagram of the coherent offline DSP stack with two options a) 4x Volterra non-linear equalizers (VNLE) and b) 4x time delay neural network equalizer for separate I and Q processing as well as for each polarization.

ing is done upon 50% of the payload of the first received frame per OSNR value, which consists of 66,444 symbols. To monitor the performance and possible overfitting, the non-linear equalizers are repetitively validated during the training phase on the remaining 50% of the payload. Once trained, the performances are evaluated on the second half of four new captured frames.

5.1.4.2 Optical Back-to-Back Performance Evaluation



(a) PreFec BER versus OSNR.

(b) Achievable rate versus OSNR.

Figure 5.6: Left: performance of optical back-to-back system 88 GBd DP-16QAM in terms of preFEC BER by applying trained linear as well as non-linear equalizers **Right:** performance in terms of achievable rate.

Fig. 5.6 depicts the performance improvement by applying VNLE architectures of different orders and a TDNNE architecture. On the left hand side, Fig. 5.6a shows the obtained pre-FEC BER. The gray dashed line depicts the theoretical upper limit for DP-16QAM at 88 GBd over an AWGN channel. It can be observed, that 1st, 3rd as well as 5th order kernels of the VNLE introduce the major benefit, while 2nd, 4th and higher orders, e.g. 6th and 7th, do not yield significant additional gain. This indicates that odd harmonics up to 5th orders are dominant in the particular optical BtB system.

Design	**Trainable Kernels/Weights**
VNLE 1st order	$5{\cdot}4 = \mathbf{20}$
VNLE 2nd order	$20{\cdot}4 = \mathbf{80}$
VNLE 3rd order	$55{\cdot}4 = \mathbf{220}$
VNLE 4th order	$125{\cdot}4 = \mathbf{500}$
VNLE 5th order	$251{\cdot}4 = \mathbf{1004}$
TDNNE	$(5{\cdot}10+10{\cdot}10+10{\cdot}10+10{\cdot}1){\cdot}4 = \mathbf{1040}$

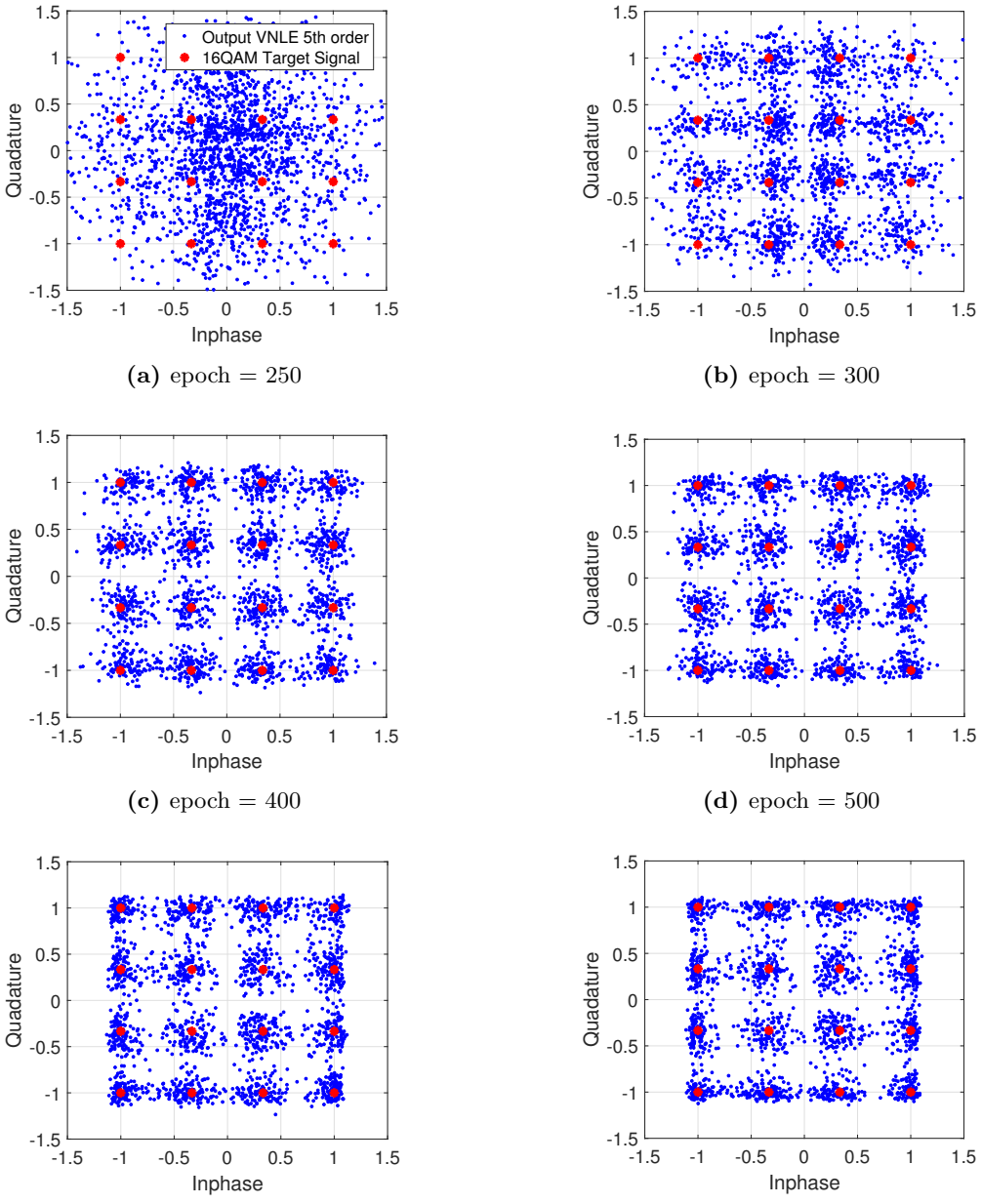
Table 5.1: Number of trainable kernels respectively weights of four VNLE and four TDNNE architectures.

For the TDNNE architecture different numbers of hidden layers and corresponding numbers of neurons per layer can be examined. To preserve a fair performance comparison a TDNNE architecture with nearly equal trainable weights as the best performing VNLE is chosen namely a design with three hidden layers with 10 neurons each, see Table 5.1. From Fig. 5.6a it can be observed that the TDNNEs outperform the 5th order VNLEs in terms of BER at a FEC limit of 10^{-2} by 0.5 dB OSNR. This behavior indicates that neural networks are capable to reflect systematic non-linearities with memory effects more accurate than 5th-order VNLEs.

In addition to the BER curves, Fig. 5.6a shows the corresponding performances in terms of achievable rates. It can be observed that the performance of the TDNNE is very poor, it even impairs the baseline performance. To get a better understanding of this behavior, Figs. 5.7 and 5.8 show the altered scatterplots of the 5th-order VNLE and the TDNNE output signal $\tilde{y}$ during the training phase. It can be observed, that the MSE trained non-linear equalizers have the ability of concentrating the constellation points and to saturated the possible outputs to the target 16QAM points, in order to reach a low value of the cost function respectively to reach the optimal MSE for non-linear equalization. However, the saturation causes non-Gaussian distributed noise in the output constellations and induces soft information loss which impairs the accompanied evaluation of the L-values by the soft-demappers.

This outcome is stronger when higher orders are consider in the VNLE, the deeper the TDNNE is designed and the smaller the constellation is chosen. In general, the training of a non-linear equalizer with respect to the MSE between transmitted x and received signal $\tilde{y}$ with 1 SpS can be considered as an approximated symbol-wise hard decision. While an improvement in BER can be achieved by using a hard-decision demapper a following post-FEC BER of a soft-decision FEC decoder would be suboptimal.

To illustrate the origin of the jail window constellation diagram, the optimal linear and non-linear equalizers using the MSE criteria are derived analytically in the Appendix A.1, considering a BPSK signal over an AWGN channel.

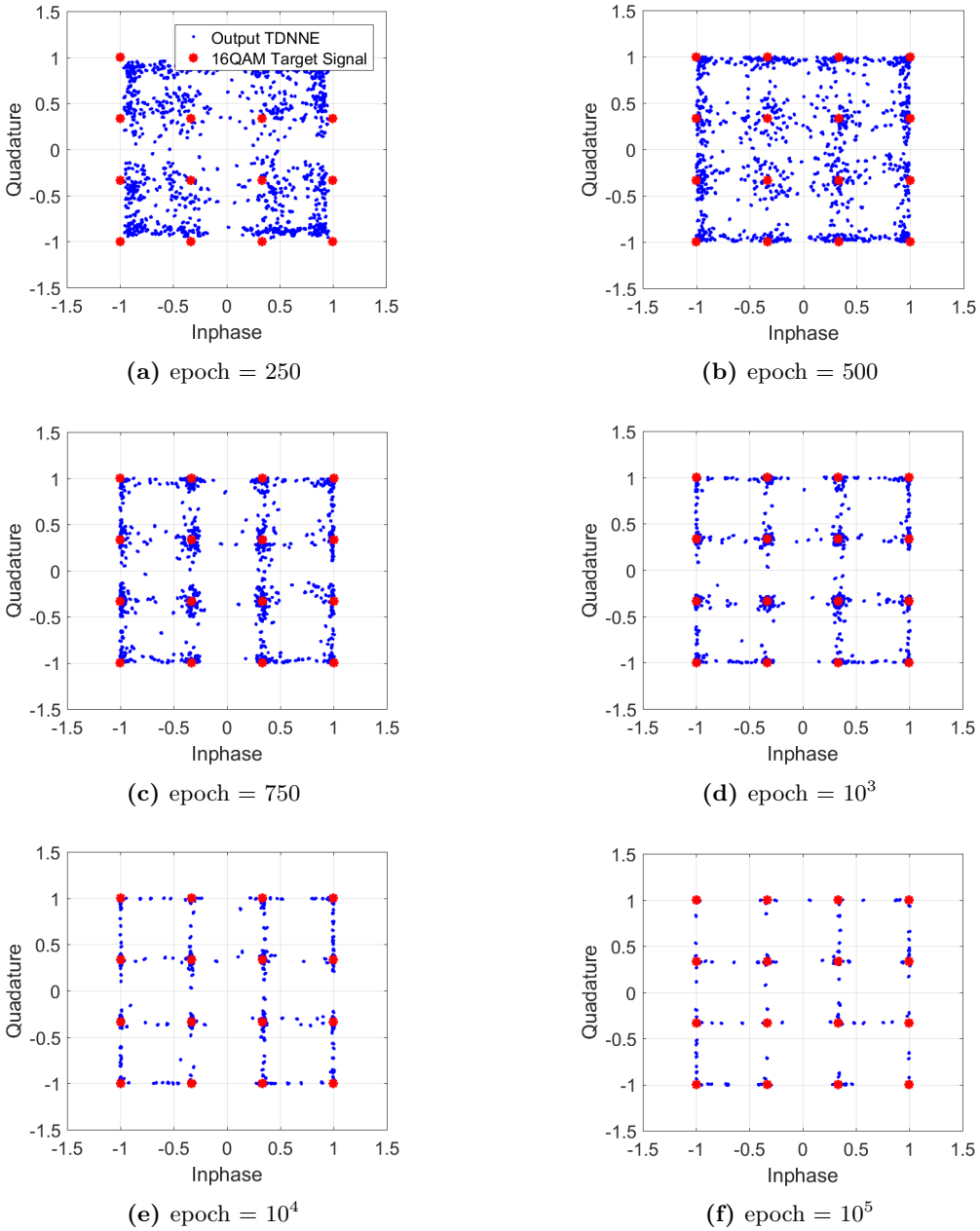


(a) epoch = 250 **(b)** epoch = 300

(c) epoch = 400 **(d)** epoch = 500

(e) epoch = 10^3 **(f)** epoch = 10^5

Figure 5.7: Scatterplots of 5th-order VNLE output $\tilde{y}$ during the learning process. The MSE trained non-linear VNLEs have the ability of concentrating the constellation points and to saturated the possible outputs to the target 16QAM points, in order to reach a low value of the cost function respectively to reach the optimal MSE for non-linear equalization. The saturation causes non-Gaussian distributed noise in the output constellations and induces soft information loss which impairs the accompanied evaluation of the L-values by the soft-demappers. This outcome is stronger when higher orders are consider in the VNLE.



(a) epoch = 250 **(b)** epoch = 500

(c) epoch = 750 **(d)** epoch = 10^3

(e) epoch = 10^4 **(f)** epoch = 10^5

Figure 5.8: Scatterplots of TDNNE output $\tilde{y}$ during the learning process. Similar to Fig. 5.7 the MSE trained TDNNEs have the ability of concentrating the constellation points and to saturated the possible outputs to the target 16QAM points, in order to reach a low value of the cost function respectively to reach the optimal MSE for non-linear equalization. The effect of saturation, which causes non-Gaussian distributed noise in the output constellations and induces soft information loss which impairs the accompanied evaluation of the L-values by the soft-demappers, is even stronger due to the cascade structure of neural networks. It converges towards a jail window. In general, the outcome is stronger the deeper the TDNNE is designed and the smaller the constellation is chosen.

5.2 Non-linear Feed-Forward Soft-Demapper

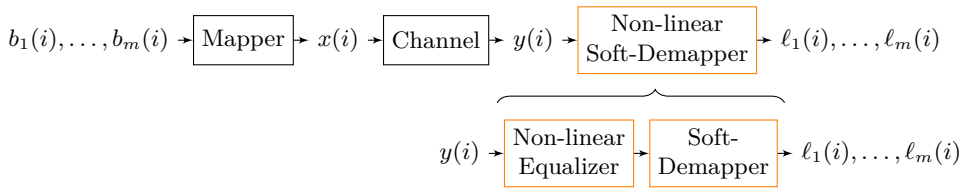


Figure 5.9: General communication black-box model.

The previous section indicated that a non-linear equalizer trained with respect to the MSE between the equalizer's output and the transmitted signal with 1 SpS each actually approximates symbol-wise hard-decision. While an improvement in BER in comparison to the VNLE could be achieved by using a hard-decision demapper a following post-FEC BER of a soft-decision FEC decoder would be suboptimal. To overcome this issue and to enable optimal soft-decision FEC decoding, a more appropriate cost function is needed to train the non-linear equalizer. In particular, the cost function should maximizes the achievable rate (2.22) (accurate post-FEC BER predictor, see Section 2.4.2) and hence the confidence of the soft-bits provided by the accompanied soft-demapper instead of minimizing the symbol-level MSE between the equalizer's output and the transmitted signal. To this end, as shown in Fig. 5.9, the non-linear equalizer and the soft-demapper are considered from now on as one DSP component acting as a non-linear soft-demapper equalizer. Hence they are now optimized jointly by using the bitwise cross equivocation (BCE) loss function

$$\mathcal{L}(b, \ell) = \log_2[1 + \exp(-(1 - 2b)\ell)] \tag{5.16}$$

where b is the transmitted bit and where ℓ is the soft-demapper output. In the appendix A.2 and A.3, the equivalence of (5.16) to the binary cross-entropy used in classic machine learning for binary classification is shown and also that this loss function is optimal, in the sense that a trained soft-demapper maximize the achievable rate (2.22). Note that (5.16) does not include the minimization over s that we have in (2.22). This allows to check, if training has been successful. For the soft bits output by the trained non-linear soft-demapper, the minimizing s in (2.22) should be equal to 1. Otherwise, further training is required.

5.2.1 General Volterra Non-linear Equalizer plus Soft-Demapper

From (5.16) it can be observed that only iterative training approaches are appropriate to minimize the BCE. This implies that the soft-demapper has to be differentiable to train

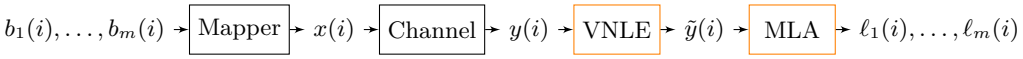


Figure 5.10: General communication model including a VNLE before the demapper.

the previous non-linear equalizer by gradient descent. For instance to optimize the VNLE with respect to the bitwise equivocation loss function (5.16) the soft-demapper can be realized by the MLA, which incurs virtually no loss in high SNR ranges. The MLA effectively

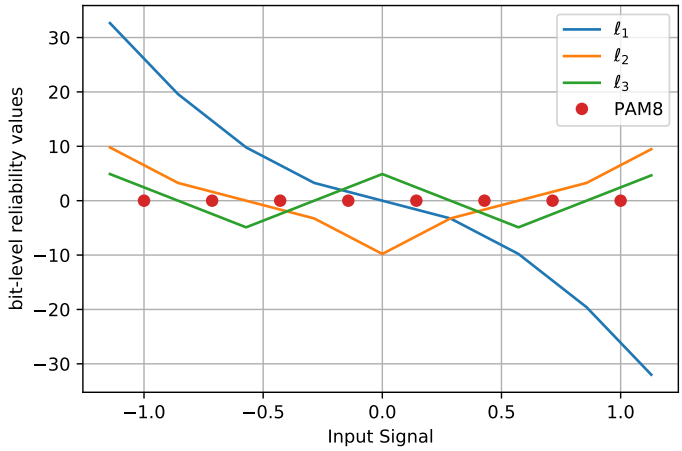


Figure 5.11: Piecewise differentiable linear approximations of the MLA soft-demapper considering PAM8.

uses piecewise differentiable linear approximations, whose slopes form the MLA parameters, as shown in Fig. 5.11. The MLA and the VNLE can then be optimized jointly by propagating the error backwards through the MLA and by using conventional gradient descent, minimizing the BCE (5.16).

5.2.2 Time Delay Neural Network Soft-Demapper

The designs of neural networks are numerous and interrelated. We could therefore either stick to the arrangement shown in Fig. 5.9 and discussed in Section 5.2.1, i.e., symbol-level non-linear equalizer plus subsequent differentiable soft-demapper, or use the TDNN directly as a non-linear TDNN soft-demapper (TDNN-SD). Hence combining the equalization and demapping tasks, as proposed in Fig. 5.12. To this end, we define m output units with linear

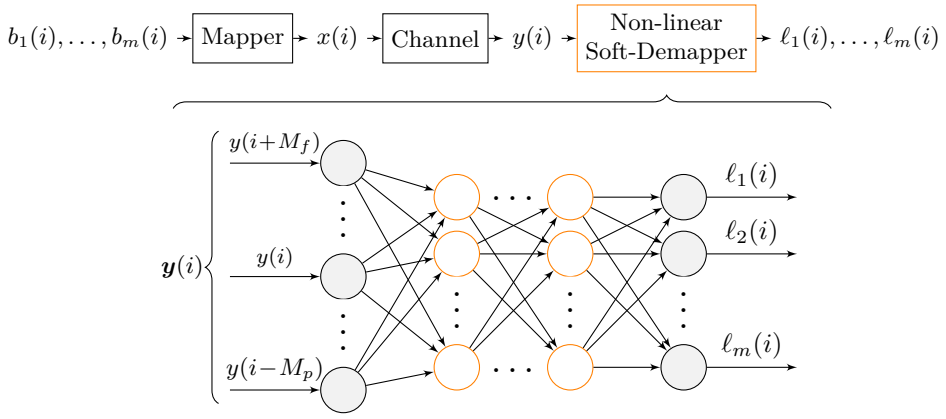


Figure 5.12: Time delay deep neural network for separate polarization processing. The architecture deploys a dual side memory structure where M_p and M_f denote the number of past and future memory taps and m is the number of bits per input symbol.

activation functions, to allow for negative and positive values, where m is the number of bits per input symbol. With $\boldsymbol{y}(i) = [y(i-M_\mathrm{p}),\ldots,y(i),\ldots,y(i+M_\mathrm{f})]$ and $\boldsymbol{\ell}(i) = [\ell_1(i),\ldots,\ell_m(i)]$ denoting the delayed signal input vector and soft-bits output vector, respectively, the TDNN-SD with L-dense layers is given by

$$\boldsymbol{a}^{[0]}(i) = \boldsymbol{y}(i), \tag{5.17}$$

$$\boldsymbol{a}^{[l]}(i) = g(\boldsymbol{W}^{[l]}\boldsymbol{a}^{[l-1]}(i) + \boldsymbol{b}^{[l]}), \quad l = 1,\ldots,L \tag{5.18}$$

$$\boldsymbol{\ell}(i) = \boldsymbol{a}^{[L]}(i) \tag{5.19}$$

where $\boldsymbol{a}^{[l]}$ denotes the output vector of the l-th layer and $\boldsymbol{W}^{[l]}$ and $\boldsymbol{b}^{[l]}$ the weight matrices and bias vectors, respectively. Eq. (5.17) denotes the input layer and (5.18) is executed successively for layers $l = 1,2,\ldots,L$ to obtain output $\boldsymbol{\ell}(i)$ in (5.19). For the activation function g, the non-linear function $g(x) = \tanh(x)$ is initially used while computationally efficient alternatives are discussed in Section 5.2.4.

5.2.3 Experimental Investigation

This section outlines the measurement setup, as used to evaluate the non-linear compensation capability of the proposed TDNN-SD. The performance and the computational complexity is benchmarked against a VNLE accompanied by an MLA soft-demapper on the basis of dual polarization 92 GBd 64QAM back-to-back (BtB) offline captures, where O/E component non-linearities dominate. For a detailed comparison of both approaches, we follow D. I.

Soloway et al. [149] and extract the linear and non-linear kernels from the trained TDNN-SD by Taylor expansion and compare with the Volterra kernels. Furthermore, several TDNN-SD architectures with different activation functions and representation capacity are compared in terms of performance and complexity. In general, the quality of the obtained soft bits is evaluated with the achievable rate (2.22).

5.2.3.1 Measurement Setup

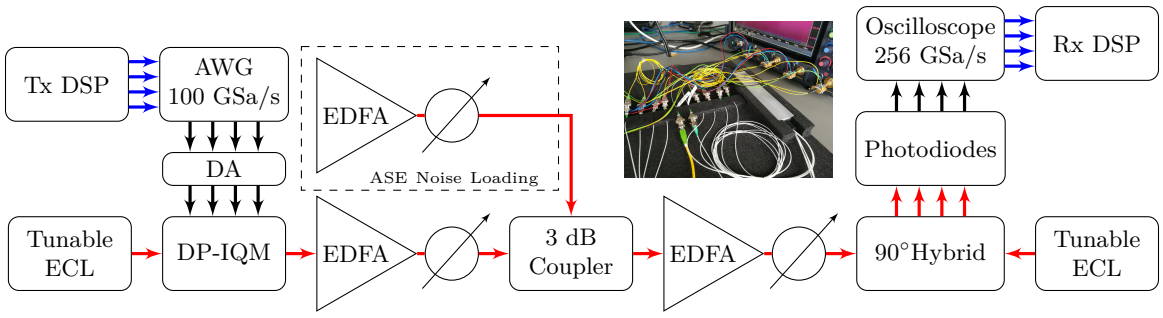
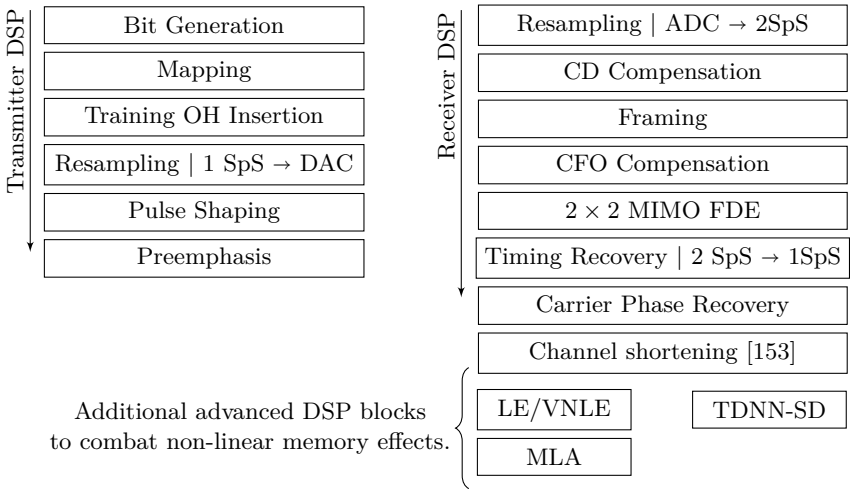


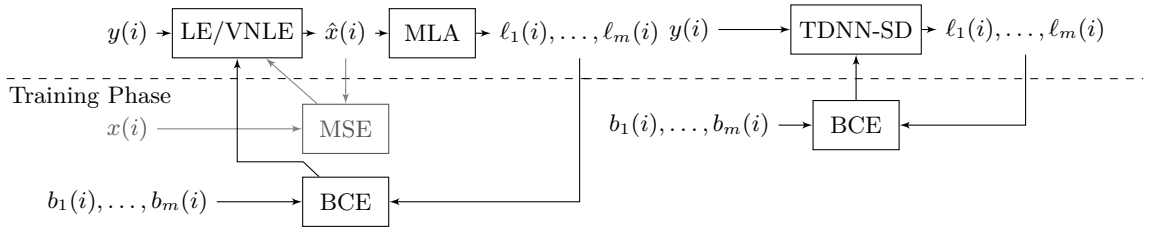
Figure 5.13: Schematic of optical back-to-back transmission system with ASE noise loading used for the experimental investigations. AWG: arbitrary waveform generator, DA: driver amplifiers, DP-IQM: dual polarization inphase-quadrature modulator, EDFA: Erbium-doped fiber amplifier, ECL: external cavity laser. The blue arrows indicate the digital part while the black and red arrows the electrical and optical part respectively. The photo shows the 90°-hybrid, PDs and oscilloscope in use.

A schematic of the coherent single carrier transmission system is shown in Fig. 5.13. The setup is optimized for maximum performance without any non-linear compensation and the measurements are performed BtB with ASE noise loading, in order to compare the achievable rate (2.22) at various OSNR values. The signal consists of a 92 GBd DP-64QAM with gross data rate of 1104 Gb/s. With 1% for training OH and assuming an FEC OH of 15% (e.g., "oFEC", [150, Table 9.1],[151]) and 20% (e.g., LDPC, [152, Sec. 3]), the corresponding net bit rate is 950 Gb/s and 912 Gb/s, respectively.

The signal output powers of the four 100 GSa/s Micram DACs with 40 GHz 3 dB-bandwidth and 6-bits nominal resolution are set to -6 dBm. Subsequently, the RF signals are amplified by four SHF S804A amplifiers with 22 dB gain and 60 GHz 3 dB-bandwidth. The amplifiers slightly operate in a non-linear region, which in turn results in intermodulation distortions. Their non-linear effects are mixed with potential non-linear distortions from a LiNbO3 DP-I/Q Modulator (Fujitsu-FTM7992HM-32 GHz) with a drive voltage of $\leq$4.2 Vpp. In the optical domain, two tunable 1 kHz ECLs are used at the transmitter and LO, respectively.



(a) Block diagram of the coherent offline DSP stack.



(b) LE/VNLE accompanied by a MLA soft-demapper trained with respect to the MSE (gray) or the BCE (black) loss function.

(c) TDNN-SD trained with respect to the BCE loss function.

Figure 5.14: Block diagram of the coherent offline DSP stack with two options a) 4x Linear equalizer (LE) / Volterra non-linear equalizers (VNLE) accompanied by a MLA soft-demapper and b) 4x time delay neural network soft-demapper (TDNN-SD) for separate I and Q processing as well as for each polarization.

The optically modulated signal is first amplified and then combined with the ASE noise generated by an EDFA. In comparison to the setup described in Section 5.1.4.1, the receiver consists of an optical 90°-hybrid and four 70 GHz balanced photodiodes instead of a NeoPhotonics 64 GBd class-40 HB-μICR with 40 GHz 3 dB-bandwidth. The photodiodes operate in the linear regime with 0 dBm optical input power. The electrical signals are digitized us-

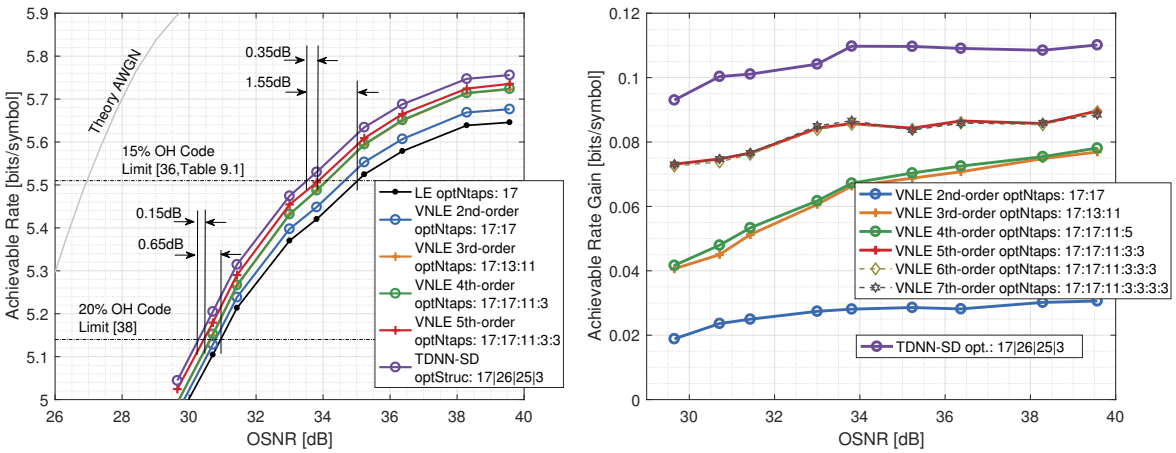
ing one Keysight Infiniium real-time oscilloscope including four 10-bits DACs operating at 256 GSa/s with 110 GHz 3 dB-bandwidth.

In order to compensate linear and non-linear ISI effects introduced from the O/E components, the receiver DSP stack in Fig. 5.14 includes next to the classical coherent signal recovery blocks, see Section 2.3.2, the stacked combination of four linear equalizers (LEs)/VNLEs plus MLA soft-decision demappers and the proposed TDNN-SDs. Both schemes operate on identical power normalized data and independently on each real dimension with 1 SpS. Note that, in a lab setup the linear part has due to pronounced cable reflections many more inputs than the non-linear part. To counterbalance this issue a linear channel shortening based on MMSE [153] is applied before the additional advanced DSP blocks to shorten the original linear channel. In general, channel shortening can be traced back to Falconer and Magee in 1974 [154], with the idea to filter the received signal with a prefilter such that the effective channel after filtering has much shorter duration with affordable complexity [155].

Regardless of the scheme in use, training on the particular linearities and non-linearities is essential before deployment. The training is done upon 50% of the payload of the first received frame per OSNR value, which consists of 66,444 symbols. In order to prevent overfitting on the training set, the performance is repetitively validated during the training phase on the remaining 50% of the payload, i.e. cross-validation. Once trained, the performance is evaluated on the second half of six new captured frames per OSNR value.

5.2.3.2 Optical Back-to-Back Performance Evaluation

Fig. 5.15a depicts the performance in terms of achievable rate (2.22) related to OSNR for the optical BtB 92 GBd 64QAM measurements by applying an LE, a VNLE up to 5th order kernels and a TDNN-SD. Kernel orders greater than 5th, e.g. 6th or 7th, do not yield any additional gain, see Fig. 5.15b, for readability, we therefore omit these results in Fig. 5.15a. The black dot-dashed lines denote the FEC limits for 15% and 20% FEC OH, respectively. All three architectures are optimized regarding achievable rate (2.22) performance and deploy a dual-side symmetric memory structure, i.e. previous as well as future symbols are taken symmetrically into account. The VNLE design e.g. 17:17:11:... stands for 17 linear memory taps (8 preceding symbols + the current symbol + 8 succeeding symbols) followed by 17 memory taps for the second order, followed by 11 memory taps for the third order and so on. The TDNN-SD design, e.g., 17|26|25|3, stands for 17 input neurons (again 8 preceding symbols + the current symbol + 8 succeeding symbols) followed by two hidden layers with 26 and 25 neurons, feeding into 3 output nodes.



(a) Performance in terms of achievable rate versus OSNR for the optical back-to-back system by applying optimized linear as well as non-linear equalizer.

(b) Non-linear compensation gains in bits/symbol related to OSNR. Blue, yellow, green, red, gold, gray represent the VNLE with different orders while purple represents the TDNN-SD.

Figure 5.15: optical back-to-back system 92 GBd DP-64QAM.

In addition to Fig. 5.15a, Fig. 5.15b plots the corresponding performance improvement compared to the LE at various OSNR values. It can be observed that 2nd, 3rd as well as 5th order kernels of the VNLE introduce the major benefit, while the 4th order does not yield significant additional gain. This indicates that odd harmonics dominate, which fits with the previous results in Section 5.1.4.1 and the theory in Section 2.1. The VNLE of 5th order improves the linear equalized baseline curve at higher OSNR areas up to ~ 0.09 bits/symbol. In lower OSNR ranges where ASE noise is the dominant distortion, the gain decreases slightly to ~ 0.07 bits/symbol. This behavior applies for all non-linear equalizer architectures.

For the TDNN-SD architecture different numbers of hidden layers and corresponding numbers of neurons per layer can be examined, in order to optimize the performance. An assessment of the architecture for 17 memory taps is given by Eq. (3.6). To release sufficient representation capacity for the equalization problem, in a first step a more complex design 17|26|25|3 is chosen. In a second stage, the complexity will be slightly reduced to determine the upper bound of the architecture complexity. It can be observed, that the appropriate TDNN-SD architecture outperforms the 5th order VNLE and improves the linear equalized baseline curve by up to ~ 0.11 bits/symbol and ~ 0.09 bits/symbol at lower OSNR values, respectively.

To highlight the impact of the proposed BCE (5.16) the obtained constellation diagrams after the LE, 3rd-orders VNLE and 5th-orders VNLE are depicted in Fig. 5.16, whereby

the corresponding achievable rates are listed in the corresponding Table 5.2. In case the equalizers are trained with respect to the MSE loss function (5.9) and independent of the accompanied MLA soft-decision demappers, the final constellations in the left column exhibit with increasing order non-Gaussian distributed noise, especially at the outer constellation points. This leads to suboptimal performance in terms of achievable rate, as described in the previous Section 5.1. Optimizing the equalizers and the MLA soft-decision demappers jointly with respect to the BCE (5.16) yields a constellation diagram with nearly Gaussian distributed noise and an achievable rate improvement of around 0.015 bits/QAM symbol for 5th-order VNLEs. This improvement confirms that the BCE loss function is the appropriate objective for maximizing the achievable rate.

Architecture:	MSE:	BCE	Diff.:
LE optNtaps: 17	**(a)** 5.409 bits/symbol	**(b)** 5.419 bits/symbol	0.010 bits/symbol
VNLE 3rd-order optNtaps: 17:13:11	**(c)** 5.474 bits/symbol	**(d)** 5.487 bits/symbol	0.013 bits/symbol
VNLE 5th-order optNtaps: 17:13:11:3:3	**(e)** 5.494 bits/symbol	**(f)** 5.509 bits/symbol	0.015 bits/symbol

Table 5.2

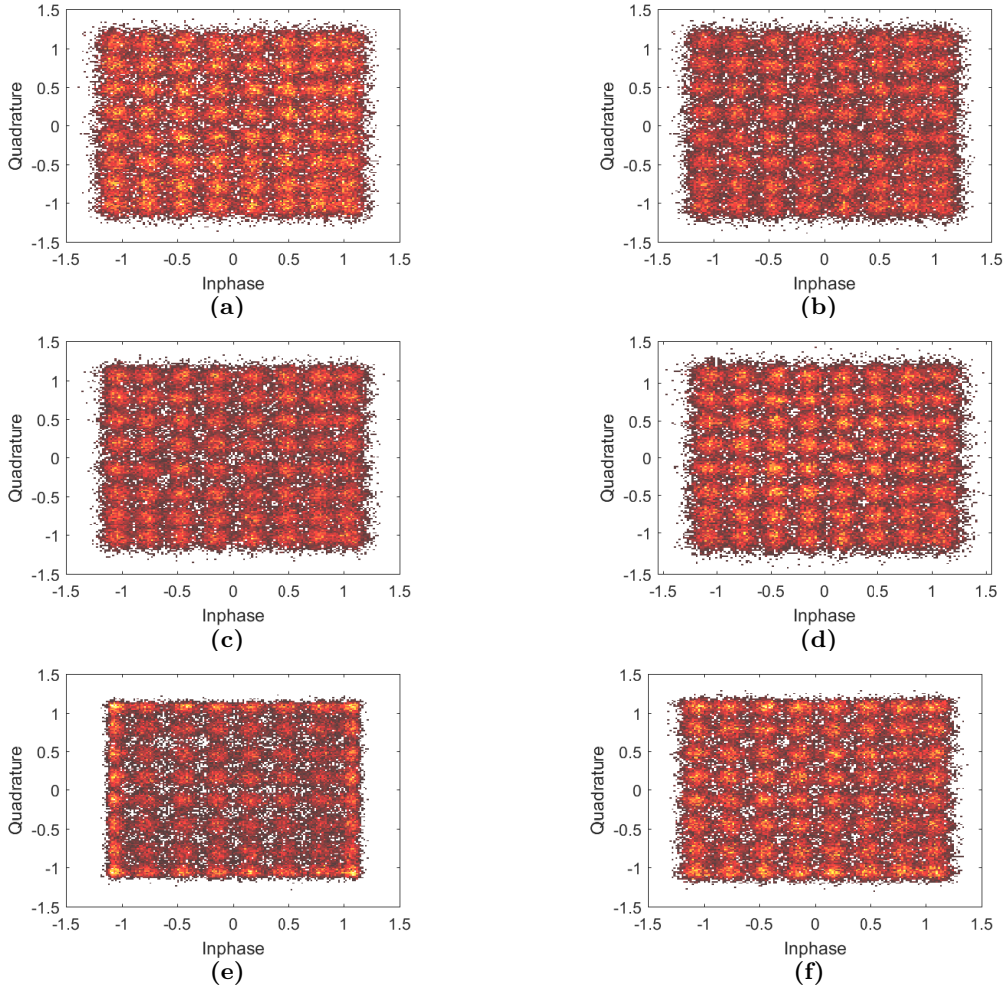
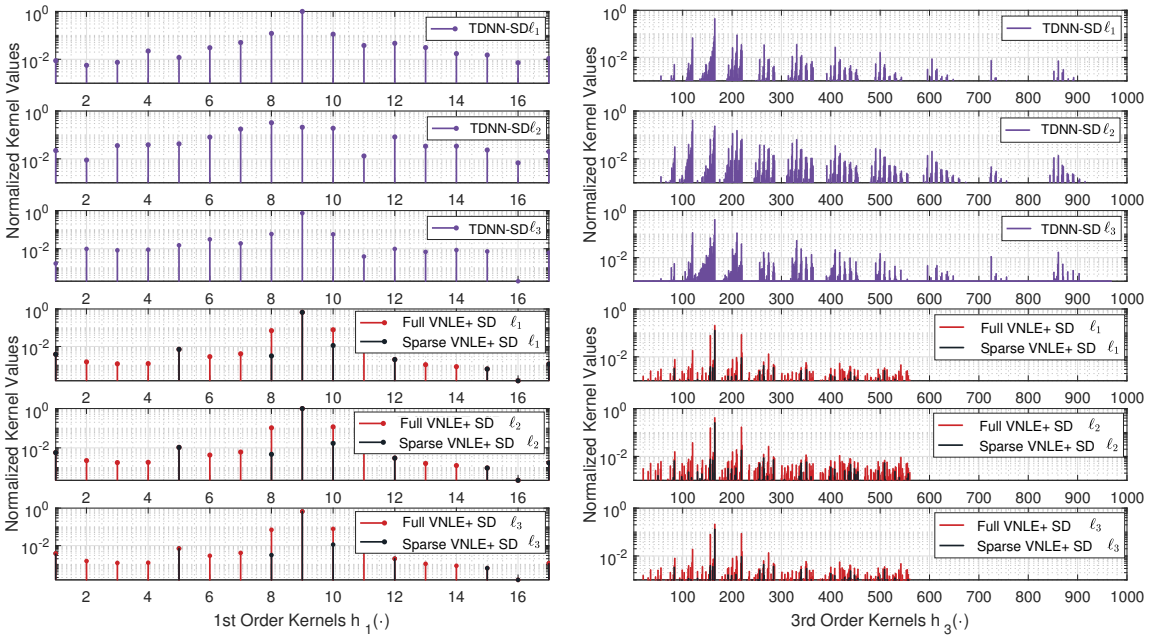


Figure 5.16: Impact on the constellation and hence on the achievable rate at 33.8 dB OSNR using the MSE or the BCE loss function for training the LE and the VNLE. Left column: constellations trained with respect to the MSE loss function (5.9) Right column: constellations trained with respect to the BCE loss function (5.16). The saturation on the outer points is best visible in Fig. 5.16e, similar to a jail window frame. In general, the effect is more pronounced (and/or more visible) on smaller constellations in combination with stronger non-linear equalizers.

5.2.3.3 Kernels of Volterra Non-linear Equalizer and Time Delay Neural Network Soft Demapper

The Volterra kernels provide a useful tool for analyzing the channel behavior. The first term in expression (2.14) represents the common finite linear impulse response of the system, while higher order Volterra kernels represent higher order impulse responses and hence describe the non-linear dynamic behavior. Following [149], as described in Section 5.1.3, the trained TDNN-SD can be expanded into a Volterra series and hence into comparable information. The kernels are expressed in terms of the parameters of the trained TDNN-SD by using expressions (5.10), (5.11) and (5.14) at $\mathbf{y_0} = \mathbf{0}$. In comparison to the customary considered symbol based VNLE kernels, the proposed TDNN-SD provides individual kernels corresponding to the soft outputs. To overcome this representation issue, the VNLE



(a) The linear kernels, or rather the linear finite impulse responses are extracted from the trained TDNN-SD by Taylor expansion and compared with the trained linear VNLE plus MLA kernels.

(b) The third order non-linear kernels, or rather the third order finite impulse responses are extracted from the trained TDNN-SD by Taylor expansion and compared with the trained third order non-linear VNLE plus MLA kernels.

Figure 5.17: Comparison of the unrolled extracted linear and third order non-linear kernels of the trained TDNN-SD with the unrolled linear and third order non-linear kernels of the trained VNLE plus MLA. The purple stems represent the bit-level TDNN-SD linear finite impulse response, while the red and black represent the full and sparse VNLE plus MLA linear finite impulse response, respectively.

kernels are extended with the particular weights of the MLA soft demapper in order to likewise represent the individual VNLE kernels for the soft bits. Fig. 5.17a depicts the in phase linear kernels from the VNLE of polarization x as well as the corresponding extracted linear kernels of the TDNN-SD. In addition, the corresponding kernels of the sparse VNLE are depicted, which we introduce in Section 5.2.4. The linear kernel values from TDNN-SD ℓ_1, TDNN-SD ℓ_3, VNLE ℓ_1 and VNLE ℓ_3 are very similar. Interestingly, the centered bit ℓ_2, which distinguishes between the inner and outer circles of the constellation points, exhibit higher distortions. While the TDNN-SD is able to adjust the kernels for each soft output individually by adapting the weights of the last layer separately, the VNLE does not exhibits that flexibility. The VNLE kernel values for each soft output only differ in the amplitude. An additional advantage of a TDNN-SD is shown in Fig. 5.17b. In comparison to the VNLE, where the number of high order taps is limited due to complexity, the TDNN-SD comprises all possible combinations of the input signal up to its input memory depth.

5.2.4 Comparison of Complexity

As part of the DSP of the optical transceiver, the non-linear equalizer is only deployable, if it can be implemented efficiently enough on an application-specific integrated circuit (ASIC). This section compares some complexity aspects of VNLEs plus MLAs versus TDNN-SD regarding the main ASIC resources, memory and logic cells on floating point level. Logic size can be well estimated in numbers of multipliers, as they are by far the most expensive logic blocks of VNLEs and TDNN-SD. In general, the complexity analysis of trainable equalizers includes two aspects, the training and the real-time execution aspect. However, in the case of optical communication, we can assume only insignificant changes of component non-linearities over time or slow processes like aging or temperature shifts. Thus, for a VNLE and a TDNN-SD the extraction of training data can be done offline, once after production on an (embedded) microcontroller. Therefore, what matters is the real-time complexity, i.e. the number of calculations that are done on each sample when executing the particular non-linear equalizer.

5.2.4.1 Complexity of Time Delay Neural Network Soft-Demapper

The real-time complexity of a TDNN-SD is defined by the number of layers, neurons and the activation function in use. As described in Section 5.2.2, the layers interconnect with a weight coefficient per connection. For implementation, those weights require memory space and a

hardware multiplier, just like the VNLE kernels. In comparison to the VNLE, an additional complexity part comes from the activation functions. The previously chosen tanh activation function could be implemented with the CORDIC algorithm [156]. However, it operates iteratively and thus, slow. A more sleek option is to linearly interpolate the tanh (I-tanh) and to use a look-up table (LUT) or to directly utilize a simpler activation function, e.g., a H-tanh or even a ReLU function. The ReLU is a linear function, which becomes non-linear by clipping. It provides the most simple activation function and requires for implementation only one hardware multiplier and one comparator for clipping. In turn, the H-tanh function is clipped on two positions, therefore, an additional comparator for clipping is required. The total required number of multipliers for a TDNN-SD with ReLU or H-tanh activation function is equal and defined as

$$\mathrm{mul}_{\text{TDNN-SD-ReLU/H- tanh}} = \underbrace{\sum_{i=1}^{L-1} s_i s_{i+1}}_{\text{Weights}} + \underbrace{\sum_{i=2}^{L-1} s_i}_{\text{Activation Function}} , \tag{5.20}$$

where L is the number of layers including input and output layers and where $\boldsymbol{s} = s_0|s_2|\ldots|s_L$ denotes the TDNN-SD design, i.e., s_i is the number of units in the i-th layer. For instance, the input layer has $s_1 = 2M + 1$ units and the output layer has $s_d = m$ units. The first term is related to the number of weights and the second term to the number of ReLU/H-tanh activation functions. If the slopes of the ReLU or H-tanh are equal to 1, no multipliers at all are needed for the activation function and the second term will be equal to zero. The complexity of a TDNN-SD with an I-tanh activation function depends on the number of linear interpolation points K, namely

$$\mathrm{mul}_{\text{TDNN-SD-Inter.Tanh}} = \underbrace{\sum_{i=1}^{L-1} s_i s_{i+1}}_{\text{Weights}} + \underbrace{(K-1)\sum_{i=2}^{L-1} s_i}_{\text{Activation function}} . \tag{5.21}$$

The first term is related once more to the number of weights, while the second term is related to the total number of multipliers which are required to generate the particular I-tanh activation functions. In particular, one interpolated activation function consists of $(K-1)$ ReLUs with a slope unequal to one or zero.

Fig. 5.18 shows the average achievable rate gain related to the number of multipliers when the previous chosen tanh activation function is replaced by an I-tanh, a H-tanh or a ReLU activation function. The average achievable rate gains are evaluated over the particular OSNR captures. The TDNN-SDs with ReLU and H-tanh activation functions are retrained,

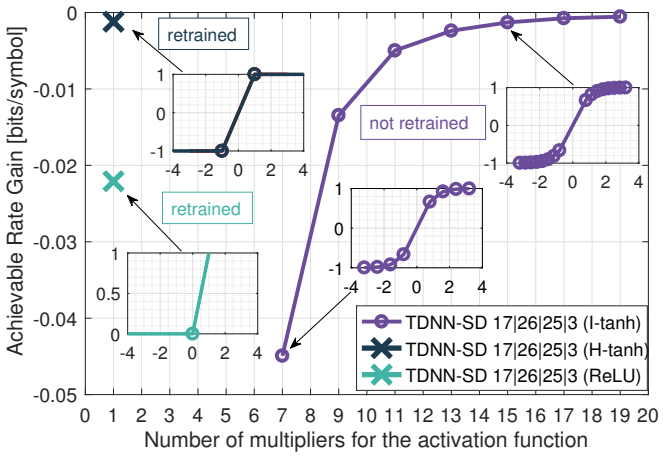


Figure 5.18: Average achievable rate performance gain in bits/symbol related to the number of required multipliers, when a linear interpolated tanh, a Hard-tanh or a ReLU activation function is applied. The three subplots illustrate the particular activation functions, where the purple circle indicate the interpolation points.

while the TDNN-SD with I-tanh activation function isn't retrained and hence contains the parameters of the TDNN-SD with tanh activation function. It can be observed that the performance of TDNN-SD with I-tanh decreases tremendously when less than 16 interpolation points and hence 15 multipliers are used. In contrast, the TDNN-SD with H-tanh achieves good results in performance and complexity, while the TDNN-SD with ReLU loses performance. Therefore, the best trade-off between performance and complexity is achieved with the H-tanh activation function.

5.2.4.2 Complexity Evaluation

Fig. 5.19 depicts the achievable rate gain of the full VNLE and the fully connected TDNN-SD architectures in relation to the LE versus the utilized number of multipliers at 33.8 dB OSNR (the behaviors are very similar for other OSNR values, see Fig. 5.15b). All architectures are optimized regarding complexity and achievable rate (e.g. the full VNLE architectures are optimized with respect to the number of linear and non-linear taps). The blue, yellow, green and red circled markers indicate individual VNLE architectures up to 5th order. It can be observed, that all four curves increase and decrease with the number of multiplications. The overall envelop indicates the optimized trade-off between complexity and performance. In general, it is more efficient to increase the degree of orders than increasing the particular

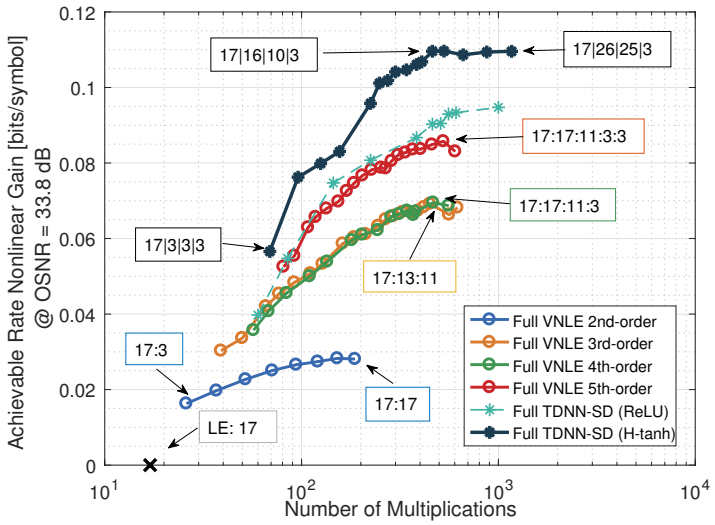


Figure 5.19: Non-linear compensation gains in bits/symbol related to number of multipliers for a 92-Gbaud 64QAM optical BtB system. The non-linear equalizer architectures are optimized regarding performance and complexity. Blue, yellow, green and red represent the VNLE architectures plus MLA with different orders while the cyan and dark blue represent the different TDNN-SD architectures with ReLU and H-tanh activation functions. Selected circled markers are labeled with their architecture.

memory sizes. However, as abovementioned, kernel orders greater than 5th do not yield any additional gain, but only increased complexity. The best performing VNLE architectures are labeled with their design.

The cyan and dark blue markers represent different TDNN-SD architectures with ReLU and H-tanh activation functions, respectively. The cyan line representing the TDNN-SD with ReLU activation function is dashed due to its limited performance and complexity trade-off and only plotted for completeness. The more interesting architectures, the TDNN-SD architectures with H-tanh activation functions, are indicated by the dark blue markers. It can be observed, that the introduced architecture 17|26|25|3 is very complex but outperforms the VNLE architectures. According to Eq. (3.6) in Section 3.1.2, the architecture 17|26|25|3 assumes that the network has to represent $\sim 10^{17}$ activation patterns for optimal equalization. On the one hand, if this assumption is not applicable, the complexity can be reduced without performance penalty. On the other hand, if complexity is reduced while the required number of activation patterns is essentially the TDNN-SD will exhibit insufficient capacity and will not be able to exactly equalize the signal anymore. Nevertheless, it will still approximate it with an error. In the current setup, by optimizing the TDNN-SD architecture regarding complexity and achievable rate (same as for the VNLE architectures)

the highest achievable rate with lowest complexity is achieved with the design 17|16|10|3. In comparison to the more complex 17|26|25|3 architecture no performance loss occurs, which indicates that in this scenario the network has to exhibit a representation capacity of at least $\sim 10^8$ activation patterns. If less neurons are used, the performance decreases immediately, which indicates that the required number of activation patterns is falling short of and the TDNN-SD exhibits insufficient representation capacity. Nevertheless, all TDNN-SD options outperform all optimized full VNLE architectures with equal complexity.

Pruning techniques such as the least absolute shrinkage and selection operator LASSO (ℓ1-norm) [157] for VNLEs or the gradual pruning algorithm [158, Sec. 3] for neural networks, have proven to reduce the set of kernels or the number of weights without significant performance loss. Fig. 5.20 shows the complexity reductions and the corresponding achievable rate gains of the sparse VNLE and the sparse TDNN-SD architectures. The filled circled markers represent full architectures, while the particular square markers represent sparse architectures with different regularization values or final sparsity values [158], respectively. It can be observed that on average the sparse VNLEs exhibit a 25% lower complexity than the full VNLEs without performance losses. The sparse TDNN-SDs exhibit a pruning factor of 20% for more complex architectures and up to 45% for lower complex architectures. A truncated single weight in a lower complexity architecture, with less neurons in the hidden layers, has a larger impact on the overall complexity.

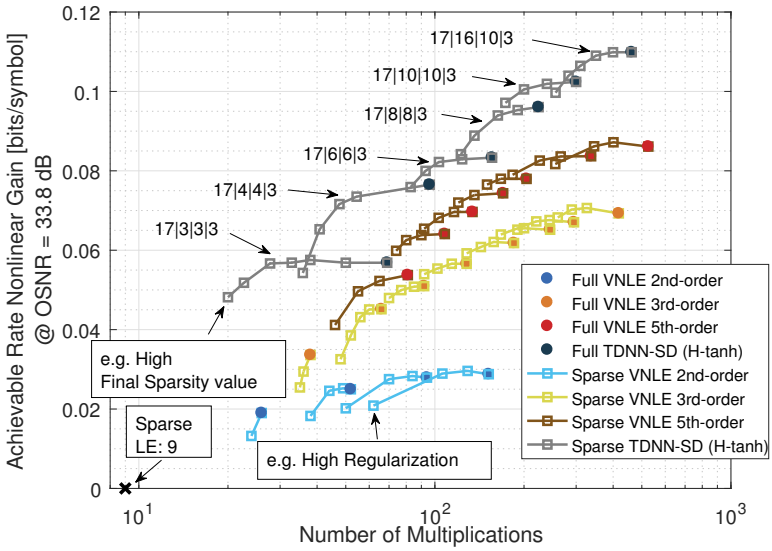


Figure 5.20: Non-linear compensation gains in bits/symbol related to number of multipliers. Complexity reduction of VNLE and TDNN-SD by L1-regularization and weight pruning.

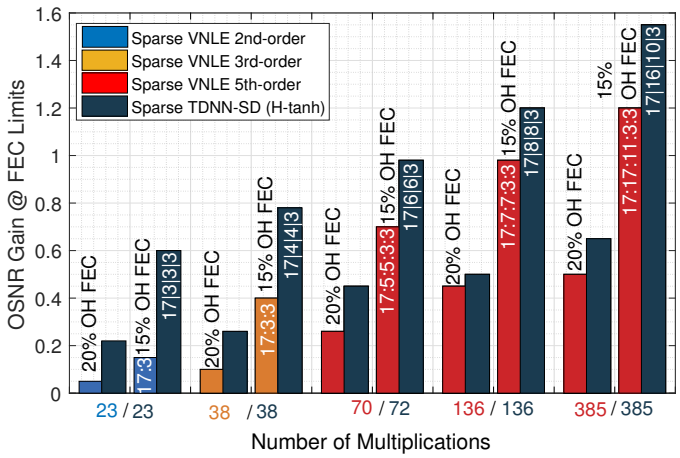


Figure 5.21: OSNR non-linear compensation gains in dB related to number of multipliers for a 92-Gbaud 64QAM optical BtB system. The labels of the architectures correspond to the 15% as well to the 20% OH FEC bars. The corresponding FEC requirements are shown in Fig. 5.15a.

Fig. 5.21 depicts a more detailed representation of the OSNR gains at the assumed FEC limits related to the number of multiplications. The particular gains depend on the assumed FEC OH and the corresponding FEC limit because of saturation, see Fig. 5.15a. Similar or even same complexity architectures are grouped. The sparse VNLE reaches the performance saturation and hence its highest OSNR gain with at least 385 multipliers. In contrary, the sparse TDNN-SD achieves similar performance with 136 multipliers. Its performance saturation and hence the highest OSNR gain is achieved with 385 multipliers. The VNLE architecture with equal complexity exhibits an OSNR penalty of 0.35 dB. In the low complexity regime the sparse VNLE performance decreases significantly more than the TDNN-SD performance.

5.2.5 Summary

High architectural complexity of Volterra non-linear equalizers has motivated investigations in non-linear equalizer alternatives based on deep neural networks. In optical coherent 92 GBd dual polarization 64QAM 950 Gb/s back-to-back measurements, where optical and electrical components non-linearities dominate, the proposed soft deep neural network equalizers proved to reflect systematic non-linearities more accurately than a 5th-order Volterra non-linear equalizer. They either outperform pruned Volterra non-linear equalizers by 0.35 dB in OSNR with equal complexity or achieve the same performance with 65% less

multipliers and hence lower complexity. In addition, we show that the deep neural network state-of-the-art cross-entropy cost function for classification problems is equivalent to bitwise cross equivocation, maximizing an achievable rate. It is therefore optimal for training DSP components acting as soft-demappers in modern communication systems with soft-decision FEC, e.g., training of non-linear equalizers should be done via a soft-demapper with respect to cross equivocation instead of mean square error.

5.3 Non-linear Recurrent Soft-Demapper

Feed-forward architectures whose inputs come from tapped delay lines, as considered in the previous section, can only model linear or non-linear FIR filters with memory at most equal to the number of the delayed inputs. In presence of colored noise effects in a bandwidth-limited channel, caused by the 2×2 MIMO equalizer due to noise enhancement, such architectures provide suboptimal performance.

In this section, we investigate as an alternative to feed-forward architectures with finite memory the compensation capability of recurrent architectures with infinite memory in presence of linear and non-linear ISI as well as colored noise effects. To assess its capacity, we extend the VNLE by a Forney [159] detector structure, which is able to handle colored noise and ISI effects optimally. The Forney detector structure comprises a symbol-spaced whitening filter (WF) and a channel-model-based Viterbi [160] or BCJR [161] detector.

5.3.1 General Volterra Non-linear Equalizer plus Whitening Filter and BCJR Detector

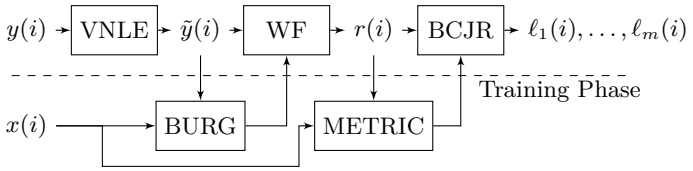


Figure 5.22: VNLE and whitening FIR filter (WF) accompanied by a BCJR.

A general equalizer targets to compensate the full signal response of the system. In strongly bandwidth limited systems, such as high speed optical coherent transmission systems, this behavior leads to strong amplification of the high signal frequencies and thus to colored noise effects after equalization. To combat this effect Forney [159] introduced a stack combination of a symbol-spaced FIR filter to whiten the noise and a Viterbi algorithm to remove the generated ISI effects, as shown in Fig. 5.22. This structure is also known as Forney's detector structure. The accompanied Viterbi algorithm is a recursive algorithm that produces the maximum likelihood message sequence of the observed signal $\boldsymbol{r}$ by determining an optimal state sequence on a trellis diagram of a hidden Markov model (HMM) [162]. However, the Viterbi algorithm computes hard-decisions, even if it is employing soft branch metrics, since a single path is selected to each state at each time. This results in an overall decision on an entire sequence of bits or symbols at the end of the algorithm [163, Chap. 14.3.1]. In [161]

the authors introduced as an alternative the BCJR algorithm, which is capable to compute soft outputs in the form of posterior probabilities for each of the message bits or symbols in a finite-state Markov system. We therefore considered the BCJR algorithm approach and implemented the corresponding max-log-MAP approximation [163, Chap. 14.3.15], which incurs virtually no loss in the SNR ranges considered in this work.

Before operation, the coefficients of the whitening filter and of the BCJR channel model have to be identified, i.e., configured upon training data, in order to match the statistical signal distribution. The simplest whitening filter has two taps with impulse response $1+\alpha D$, where D denotes the tap delay and α the weight. A longer filter impulse response would lead to a flatter noise spectrum, but would introduce a higher observation space and hence exponentially more states and transitions in the trellis, namely

$$N_{\text{States}} = C^{l-1}, \tag{5.22}$$

$$N_{\text{Transition}} = CC^{l-1} = C^{l}, \tag{5.23}$$

where C denotes the cardinality of the alphabet and l the length of the filter impulse response. To identify the coefficients, regardless of the filter length, the Burg algorithm [164] can be used. A detailed description of its usage is given in [165, Sec. II-E]. To achieve best performance, the branch metrics used by the BCJR detector should match the statistical distribution of the input signal $\boldsymbol{r}$. We considered a histogram based Gaussian channel model and pre-trained the mean μ_n and variance σ_n^2 of each possible transition individually before operation. Afterwards, the branch metrics are computed as

$$m_n = \frac{(r(i) - \mu_n)^2}{\sigma_n^2}, \quad n = 1, \ldots, C^l. \tag{5.24}$$

The real-time complexity of a WF and a max-log-MAP BCJR per symbol is defined by the number of states and the considered memory length. Their number is directly connected to the required number of hardware multipliers per symbol, namely

$$\text{mul}_{\text{WF+BCJR}} = \underbrace{(l-1)}_{\text{WF}} + \underbrace{2 \cdot 2C^l}_{\text{Branch metric}} \tag{5.25}$$

where the first term is related to the whitening FIR filter and the second to the BCJR. The anterior factor of 2 in the BCJR term refers to the forward and backward path, while the remaining part refers to the computational complexity of the branch metrics, see (5.24).

5.3.2 Bidirectional Recurrent Neural Network Soft-Demapper

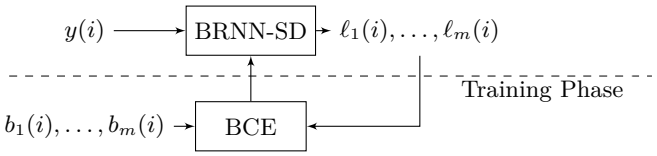


Figure 5.23: Bidirectional recurrent neural network soft-demapper (BRNN-SD).

Feed-forward neural network soft-demappers which inputs comprising current and historical captures, often implemented as tapped delay lines as shown in Section 5.2.2 are limited in functionality since they rely on a predetermined memory architecture and their outputs are independent of previous decisions. Hence, they can only model non-linear FIR filters with limited memory. A more powerful architecture can be implemented when cyclical connections are allowed, namely via RNNs [166], as shown in [167, 168]. The cyclical connections form internal memory states and enable dynamic temporal behaviors for time sequences. RNNs

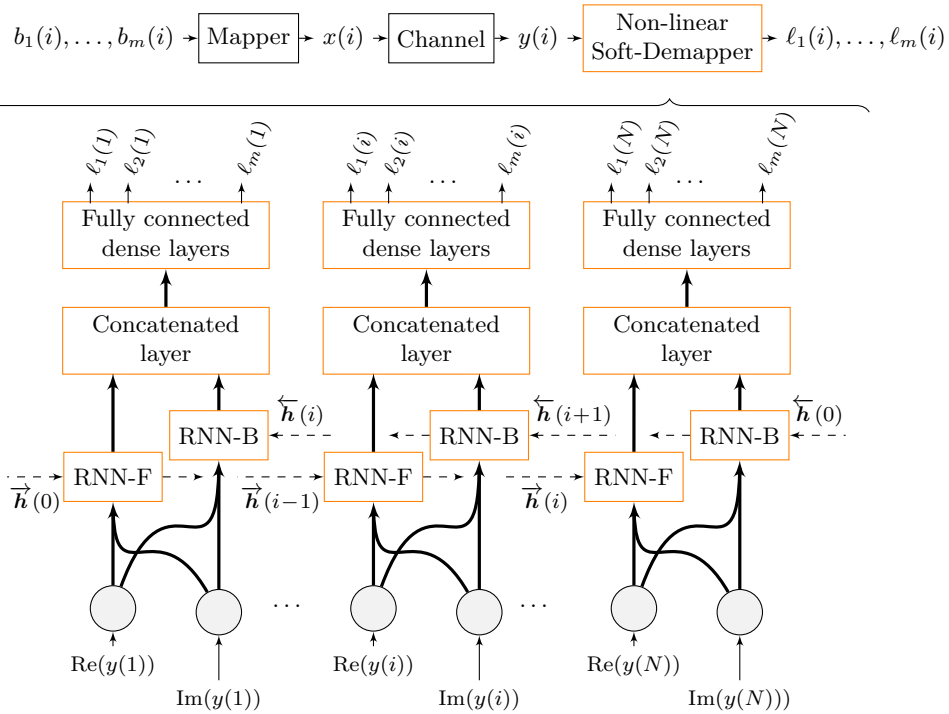


Figure 5.24: Bidirectional recurrent neural network soft demapper for joint I&Q processing. It consists of two independent RNNs, in order to predict the soft-bits of the sequence based on the symbol's past and future contexts, and one shared feed-forward neural networks with multiple dense layers to concatenate the outputs of the two parallel RNNs. In principle, the joint I&Q input options could learn as well phase impairments.

can therefore model non-linear IIR filters as shown in [169] and are therefore also related to hidden Markov models as shown in [170], which form the base model of BCJR detectors.

Fig. 5.24 illustrates the structure of the proposed BRNN soft demapper (BRNN-SD) for a sequence of length N, where N denotes the number of time steps [171]. This bidirectional architecture ensures that past as well as future symbols are taken into account to estimate the sequence of soft bits. The received symbol sequences are fed in the forward direction into one RNN-F unit and in backward direction (flipped) into an RNN-B unit [172]. The exchanging state vectors are denoted by $\overrightarrow{\boldsymbol{h}}$ and $\overleftarrow{\boldsymbol{h}}$, respectively, whereby $\overrightarrow{\boldsymbol{h}}(0)$ and $\overleftarrow{\boldsymbol{h}}(0)$ denote the initial states and are initialized with zeros. The outputs are then concatenated and fed into multiple fully connected dense layers terminated by a final linear layer of length equal to m, the number of bits per symbol. This architecture accounts for past and future symbols when estimating the soft-bits. With $\boldsymbol{y}(i) = [\mathrm{Re}(y(i)), \mathrm{Im}(y(i))]$ and $\boldsymbol{\ell}(i) = [\ell_1(i), \ldots, \ell_m(i)]$ representing signal input vector and soft-bits output vector, respectively, the BRNN-SD with U recurrent layers and L dense layers is given by

$$\overrightarrow{\boldsymbol{h}}^{[0]}(i) = \overleftarrow{\boldsymbol{h}}^{[0]}(i) = \boldsymbol{y}(i), \tag{5.26}$$

$$\overrightarrow{\boldsymbol{h}}^{[u]}(i) = g(\boldsymbol{W}_{\mathrm{F}}^{[u]} \overrightarrow{\boldsymbol{h}}^{[u]}(i-1) + \boldsymbol{W}^{[u]} \overrightarrow{\boldsymbol{h}}^{[u-1]}(i) + \boldsymbol{b}^{[u]}), \qquad u = 1, \ldots, U \tag{5.27}$$

$$\overleftarrow{\boldsymbol{h}}^{[u]}(i) = g(\boldsymbol{W}_{\mathrm{B}}^{[u]} \overleftarrow{\boldsymbol{h}}^{[u]}(i-1) + \boldsymbol{W}^{[u]} \overleftarrow{\boldsymbol{h}}^{[u-1]}(i) + \boldsymbol{b}^{[u]}), \qquad u = 1, \ldots, U \tag{5.28}$$

$$\boldsymbol{a}^{[0]}(i) = [\overrightarrow{\boldsymbol{h}}^{[U]}(i) \overleftarrow{\boldsymbol{h}}^{[U]}(i)], \tag{5.29}$$

$$\boldsymbol{a}^{[l]}(i) = g(\boldsymbol{W}^{[U+l]} \boldsymbol{a}^{[l-1]}(i) + \boldsymbol{b}^{[U+l]}), \qquad l = 1, \ldots, L \tag{5.30}$$

$$\boldsymbol{\ell}(i) = \boldsymbol{a}^{[L]}(i), \tag{5.31}$$

where $\boldsymbol{W}_{\mathrm{F}}^{[l]}$ denote the forward and $\boldsymbol{W}_{\mathrm{B}}^{[l]}$ the backward weight matrices. In comparison to the previous Section 5.1 joint I&Q processing is considered here, which is required for odd constellations. For the activation function g, we use, equal to the TDNN-SD, the non-linear H-tanh function for the hidden layers and the linear function for the output layer. The real-time complexity of the BRNN-SD per symbol is thus defined as

$$\mathrm{mul}_{\text{BRNN-SD-H-tanh}} = \underbrace{\sum_{i=0}^{U-1} 2(s_i s_{i+1} + s_{i+1}^2)}_{\text{Recurrent layers}} + \underbrace{2 s_U s_{U+1}}_{\text{Concatenate layer}} + \underbrace{\sum_{j=U+1}^{U+L} s_j s_{j+1}}_{\text{Dense layers}} + \underbrace{\sum_{k=1}^{U+L} s_k}_{\text{Activation Function}} \tag{5.32}$$

where $\boldsymbol{s} = s_0 | s_1 | \ldots | s_{U+L+1}$ denotes the BRNN-SD design, i.e., s_i is the number of units in the i-th layer. For instance, the input layer has $s_0 = 2$ units (inphase and quadrature) and the output layer has $s_{U+L+1} = m$ units. The first term is related to the number of

weights in the parallel recurrent layers and the second term to the number of weights in the concatenation layer. The third term is related to the number of weights in the subsequent dense layers. The fourth term to the number of H-tanh activation functions. If the slopes of the H-tanh are equal to 1, no multipliers at all are needed for the activation function and the fourth term will be neglectable.

5.3.3 Numerical Study

In this section, we first present a numerical study of the TDNN-SD and BRNN-SD on a simulated channel with colored noise only. Fig. 5.25 depicts the block diagram of the colored

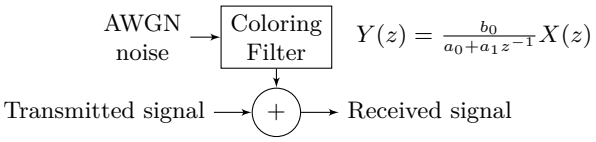


Figure 5.25: Block diagram of the colored noise channel.

noise channel. The filter coefficients are set to $a_0 = 1, a_1 = 0.5$ and $b_0 = 1$. The corresponding power spectral densitys (PSDs) before and after applying an appropriate WF with 2-taps are shown in Fig. 5.26-left. As expected, the noise enhancement is completely removed by the FIR filter due to the optimal architecture. The overall performance is shown in Fig. 5.26-right. The black line illustrates the baseline where no compensation technique is applied. The strong noise enhancement leads to a high number of errors. Adding the 2-tap WF and a BCJR detector improves the performance significantly. Note that, this architecture provides the optimal decoder in this scenario. It can be observed, that the BRNN-SD is

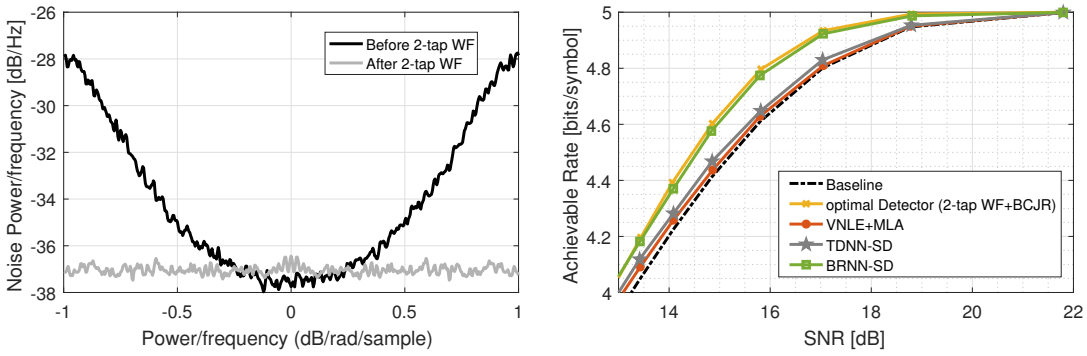


Figure 5.26: Left: Noise PSDs of a simulated 32-QAM signal with colored noise before and after whitening filter with 2-taps. **Right:** 32QAM performance in terms of achievable rate versus SNR for a simulated colored noise channel.

nearly capable of recovering the performance of the BCJR, while feed-forward architectures and hence common linear or non-linear equalizers fall short and are suboptimal in case of colored noise effects. One reason for this are the internal memory state of the BRNN-SD and hence the relation to the hidden Markov chain.

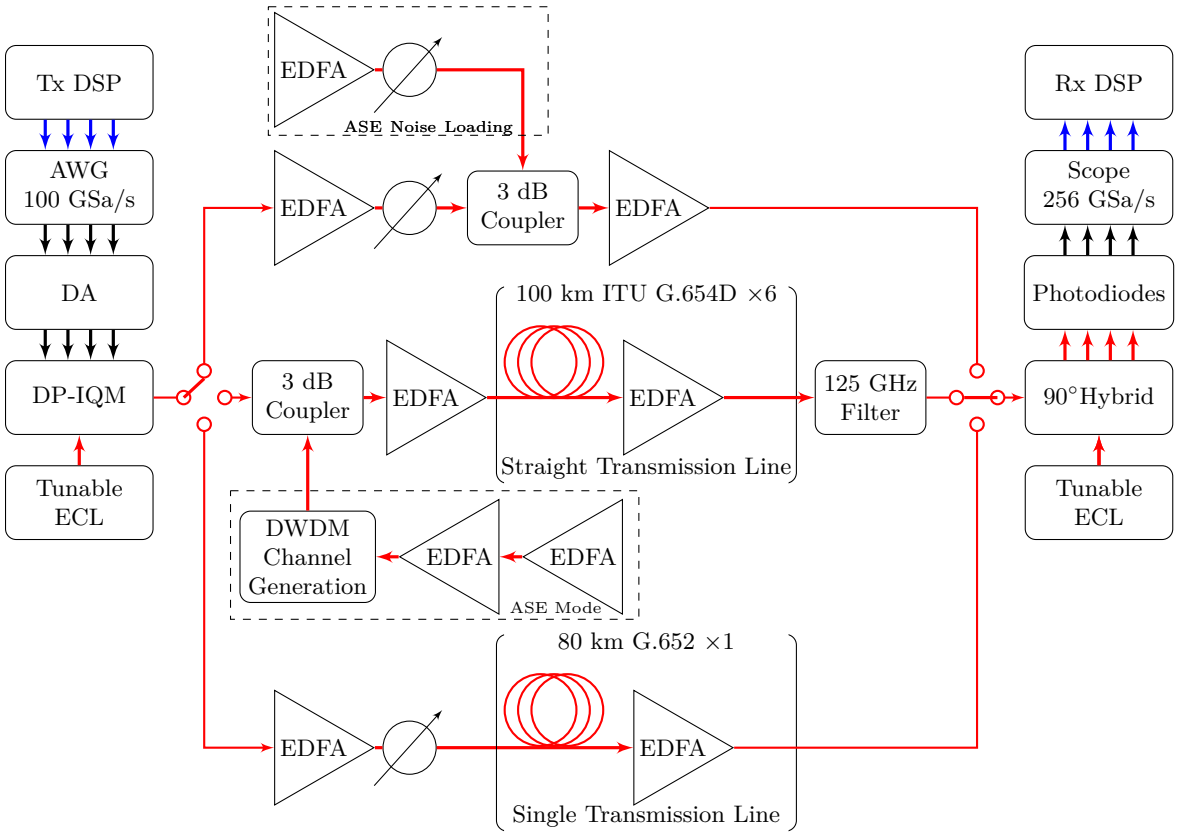
5.3.4 Experimental Investigations

The experimental evaluation of the proposed architectures is done on basis of identical offline data captured in the case of optical 92 GBd BtB transmission with ASE noise loading and 600 km 32-channel dense wavelength division multiplexing (DWDM) fiber transmission of coherent 96 GBd 800 Gb/s/λ dual polarization (DP)-32QAM. In such high speed optical coherent transmission systems, the data rate demand leads to symbol rates that are higher than the provided 3 dB-bandwidth of the O/E components. This results in ISI and after equalization to colored noise effects due to noise enhancement. Next to the aforementioned linear impairments, the transmitted symbols are subject to non-linear distortions with memory effects arising from the non-linear transfer characteristics of the O/E components. The non-linear effect is additionally enhanced, if long fiber is applied and hence non-linear interference arises in the link.

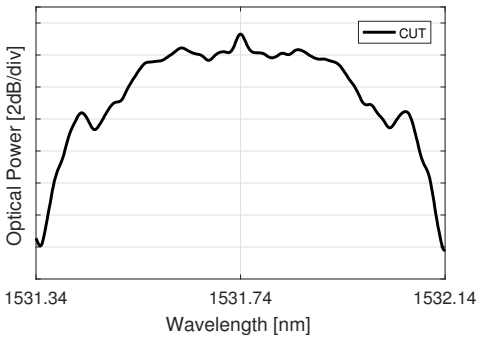
5.3.4.1 800G DWDM Experimental Setup

The experimental setup is shown in Fig. 5.27a including optical BtB with ASE noise loading (top path) and a straight transmission line with 600 km total fiber length (bottom path). In both cases the setup was optimized for maximum performance with linear coherent DSP. The channel under test (CUT) carries a 96 GBd DP-32QAM with gross data rate of 960 Gb/s. With 4.4% for training sequences and assuming an FEC OH of 15% (e.g., “oFEC”, [150, Table 9.1],[151]), the net bit rate is 800 Gb/s.

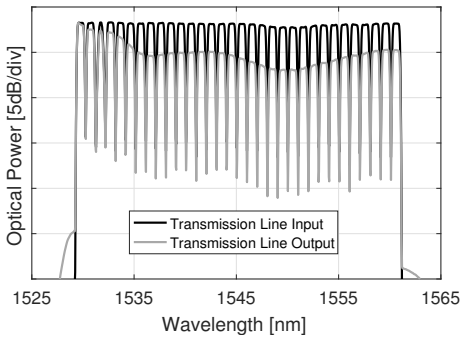
At the transmitter, the four 100 GSa/s Micram DACs with 40 GHz 3 dB-bandwidth generate a repeated pattern of 76800 samples, hence one frame, with output powers of -6 dBm. Subsequently, the RF signals are amplified by four SHF S804A amplifiers with 22 dB gain and 60 GHz 3 dB-bandwidth. The amplifiers operate in a slighly non-linear region, which in turn results in non-linear distortions. However, with this setting the best performance could be achieved. Their non-linear effects are mixed with potentially non-linear distortions from the LiNb03 DP-I/Q Fujitsu modulator with 32 GHz 3 dB-bandwidth and a drive voltage of $\leq$4.2Vpp.



(a) Schematic of the optical BtB system with ASE noise loading (top path), the DWDM transmission system including 600 km fiber (middle path) and the single transmission line including 80 km fiber (bottom path) used for the experimental investigations. The blue arrows indicte the digital part while the black and red arrows the electrical and optical part respectively.



(b) Optical spectrum of the 96 GBd DP-32QAM CUT with central wavelength at 1531.74 nm at the transmission line input. The bandwidth limiting effects can be observed.



(c) Optical DWDM spectrum at the input and output of the 600 km transmission line. After the transmission line a ripple of 5 dB can be observed due to the lack of gain filtering.

Figure 5.27: 800G DWDM experimental setup.

In the optical domain, two tunable 100 kHz ECLs are used at the transmitter and LO, respectively. The DWDM system is emulated by generating 31 channels loaded with noise and shaped with a 96 GHz RRC optical filter with 0.2 roll-off factor having central frequencies ranging from 192.095 THz (1529.774 nm) to 195.970 THz (1560.633 nm) on a 125 GHz grid. The channels are multiplexed together with the CUT by using a 3 dB coupler and sent to an EDFA acting as a booster. The DWDM signal is launched into a transmission line consisting of six spans, each of 100 km length. In use are pure-silica core fibers, with ultra low attenuation, compliant to ITU-T G.654D [173]. In comparison to G.652, G.654D fibers exhibit a larger effective area ($A_{\text{eff}} = 130\mu m^2$) and are thus in addition more robust against non-linearities [174]. The seven C-band EDFAs operate therefore with optimal input and output powers of 4 dBm and 20 dBm, respectively. Fig. 5.27b depicts the spectrum of the 96 GBd DP-32QAM CUT, whose wavelength is varied over the whole band. Fig. 5.27c illustrates the DWDM spectra at the input and output of the 600 km transmission line. The spectrum at the output of the booster amplifier is fairly flat with a ripple below 0.5 dB. After the transmission line, at the receiver amplifier, a ripple of 5 dB is observed due to the lack of gain filtering.

The receiver consists of an optical 90°-hybrid and four 70 GHz balanced photodiodes. The photodiodes operate in the linear regime with 0 dBm optical input power. The electrical signals are digitized using four 10-bits ADCs operating at 256 GSa/s with 110 GHz 3 dB-bandwidth.

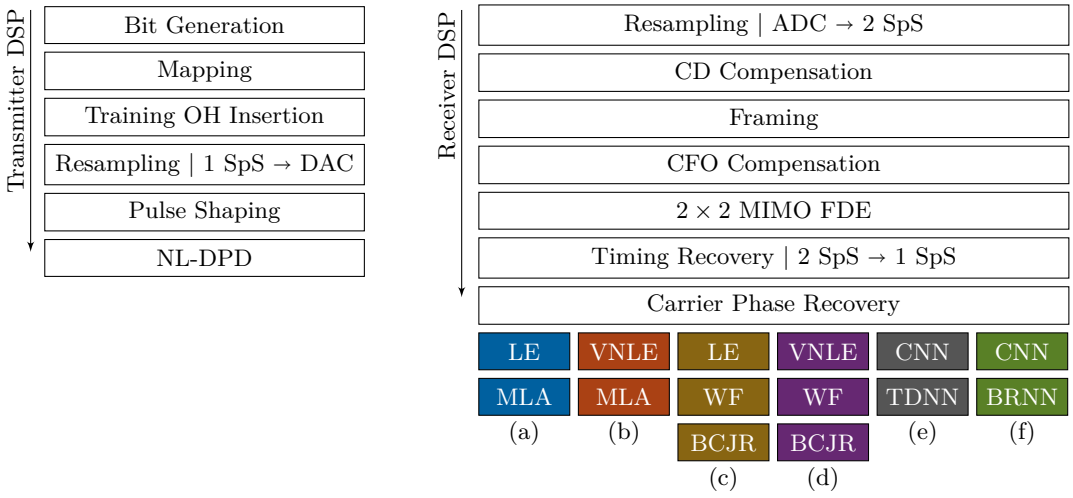


Figure 5.28: Block diagram of the coherent offline DSP stack. After the carrier phase recovery block, six architectures are selectable: **(a)** LE **(b)** VNLE **(c)** LE accompanied by 2-tap WF + BCJR **(d)** VNLE accompanied by a 2-tap WF + BCJR, **(e)** 1D-CNN plus a TDNN-SD and finally **(f)** 1D-CNN plus a BRNN-SD.

In order to compensate linear and non-linear ISI, as well as colored noise effects, the receiver DSP stack in Fig. 5.28, includes next to the classical coherent signal recovery blocks, see Chapter 2, several further options for enhanced signal processing, i.e., VNLE, WF + BCJR and the proposed TDNN-SD and BRNN-SD. After the carrier phase recovery block, six architectures are selectable: **(a)** LE **(b)** VNLE **(c)** LE accompanied by 2-tap WF + BCJR **(d)** VNLE accompanied by a 2-tap WF + BCJR, **(e)** 1D-CNN plus a TDNN-SD and finally **(f)** 1D-CNN plus a BRNN-SD. In comparison to Section 5.2, where linear channel shortening is applied to counterbalance the pronounced cable reflections in a lab setup, we state in this configuration the first linear layer as a separate neural network. In particular we consider an additional linear 1D-CNN layer, which is trained jointly with regard to BCE by gradient descent in order to optimize the achievable rate.

The performance of the proposed schemes is evaluated on the second half of eight captured frames, resulting in a total amount of 278,128 symbols. The first half of an additional captured frame, referred as training frame, is used for the training of VNLE, WF, BCJR channel model, TDNN-SD and BRNN-SD. This ensures that the data used for training is uncorrelated with the data used for evaluation. In order to monitor the learning process and to recognize possible noise overfitting the performance is repetitively cross-validated during the training phase on the second half of the training frame. The training frame is completely discarded in the evaluation phase.

5.3.4.2 Optical Back-to-Back Performance Evaluation

The overall performance of the individual DSP architectures is analysed on the optical BtB measurements with 92 GBd and shown in Fig. 5.29. In this scenario bandwidth limiting effects and O/E component non-linearities dominate. The black dashed line indicates the baseline where no additional equalizer after carrier phase recovery is applied. By switching on a complex LE (161 taps per real dimension) a gain of 0.61 dB is observed at the assumed FEC limit. A further gain of 0.67 dB is obtained when a VNLE of 5th order with linear and non-linear taps (161/11/7/3/3 taps of order 1/2/3/4/5 per real dimension) is used. The TDNN-SD considering 22 input neurons (11 input neurons per real dimension) followed by four hidden layers with 12 neurons each outperforms the VNLE by 0.15 dB. This performance improvement of 0.15 dB matches to the observed results in the previous Section 5.1, where a more detail comparison of VNLE and TDNN-SD is provided.

The combination of LE, WF + BCJR exceeds the performance of the VNLE and the TDNN-SD by another 0.37 dB and 0.22 dB, respectively. Hence, linear impairments caused by

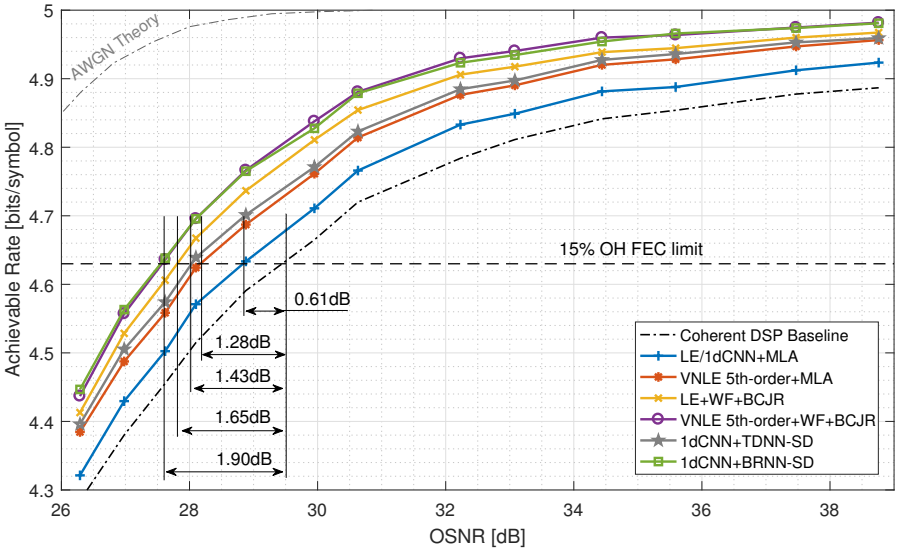


Figure 5.29: 92 GBd DP-32QAM optical BtB performance.

Dim.:	2-tap WF coeff.:
x_i	[1 0.32]
x_q	[1 0.28]
y_i	[1 0.41]
y_q	[1 0.45]

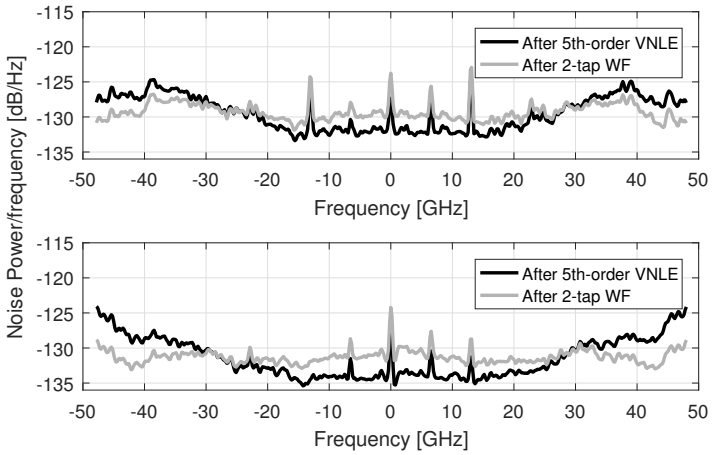


Figure 5.30: Noise PSDs of 92 GBd 32-QAM BtB before and after a whitening filter with 2-taps of x-polarization (top) and y-polarization (bottom). It is assumed that there is no residual ISI in the sequence after VNLE and the noise is calculated as the difference between the VNLE output and the transmitted complex symbols. The noise peaks originate from the 50 GHz clock and are located at 12.5 GHz and 6.25 GHz.

limited bandwidth have a stronger impact on the performance than non-linear impairments from the O/E components in the considered setup. Nevertheless, enabling altogether, i.e., VNLE 5th-order+WF+BCJR, the required OSNR is reduced to 27.6 dB resulting in a gain of 1.90 dB with respect to the baseline coherent DSP. Fig. 5.30 depicts the corresponding power spectral densities before and after the WF with 2-taps of x-polarization and y-polarization,

respectively. The spectrum is almost flat after the WF. The noise peaks originate from the 50 GHz clock and are located at 12.5 GHz and 6.25 GHz. The green curve in Fig. 5.29 indicates the performance of the BRNN-SD with two input neurons followed by four recurrent and two dense layers with 13 neurons each. It achieves the same performance as the stack combination of VNLE, WF + BCJR. This indicates that recurrent architectures are not only capable to handle mixed signal memory effects as feed-forward architectures do, but also compensate for colored noise effects.

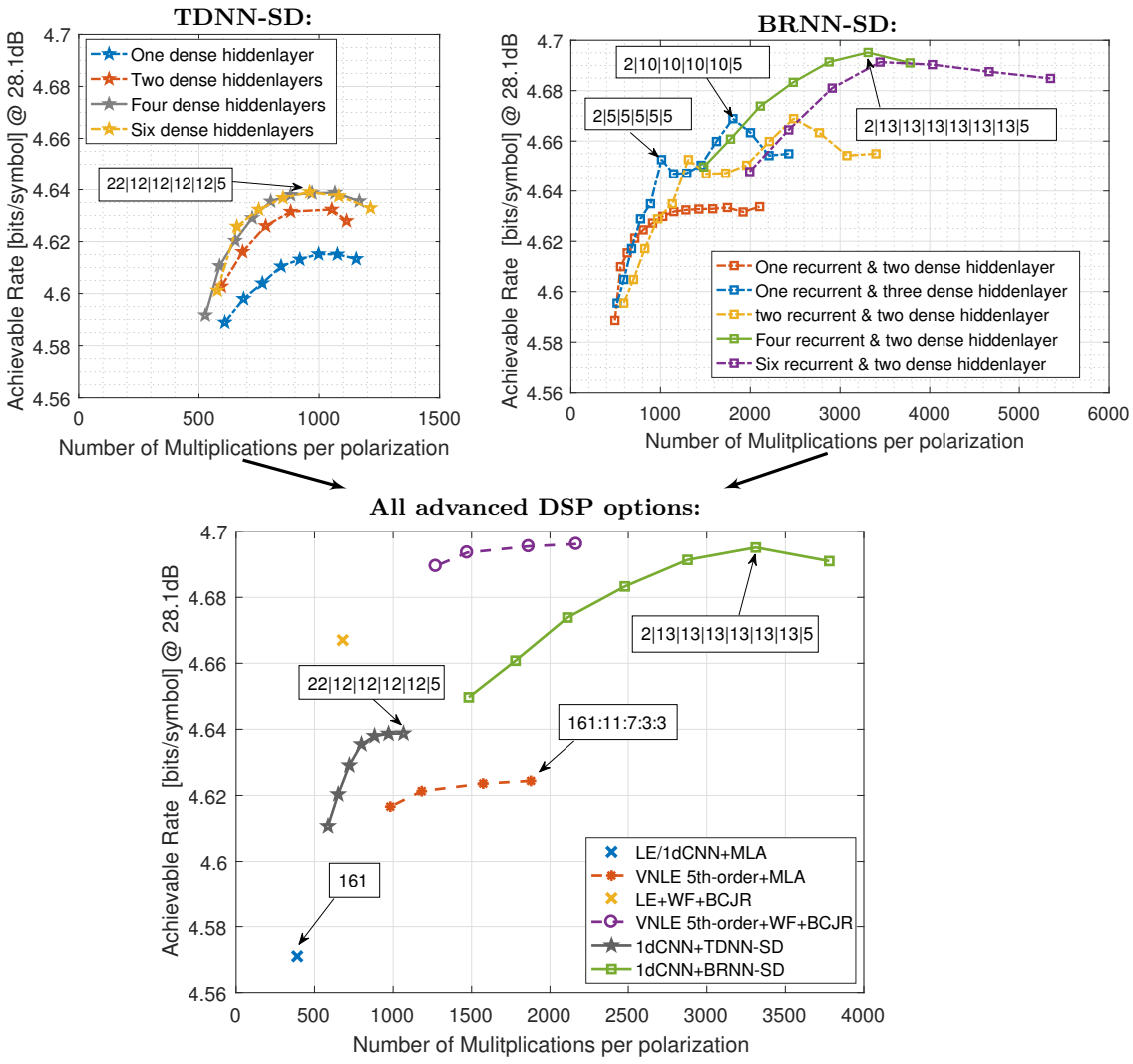


Figure 5.31: Achievable rate versus number of multiplications per polarization at 28.1 dB OSNR. The top figures depict the particular performance of different TDNN-SD and BRNN-SD designs, while the bottom figure compares the performance of all advanced DSP options.

Fig. 5.31 depicts the achievable rate gain of the particular applied DSP architectures in relation to the required number of real multipliers at 28.1 dB OSNR (closest average measurement point to FEC limit). The used architectures are labeled with their design. The red markers indicate individual VNLE architectures up to 5th order, while the star markers represent different TDNN-SD architectures with various numbers of neurons within four hidden layers. Higher order kernels respectively more hidden layers did not yield significant additional gain or complexity reduction, see top left figure. As expected and discussed in Section 5.1, all TDNN-SD options outperform all VNLE architectures with equal complexity. The more interesting DSP architectures, the WF+BCJR as well as BRNN-SD architectures, are indicated by the circled and square markers, respectively. It can be observed, that the recurrent architecture achieves the performance of the stack combination of VNLE+WF+BCJR with at least 3300 multipliers per polarization. If less neurons or hidden layers are assigned, its performance is deteriorated and the classical stack combination outperforms the BRNN-SD. The computational complexity gap between the classical stack combination and the BRNN-SD, in terms of real multiplications, is significant. However, it must be emphasized that the strength of the BRNN-SD is the non-linear compensation. In the considered scenario linear impairments are clearly dominant. A linear equalizer followed by a WF and a BCJR is in this scenario the logical choice and yields a good trade-off between complexity and performance.

5.3.4.3 Optical Transmission Performance Evaluation over 600 km Reach

The 800G 96 GBd DP-32QAM DWDM transmission results with reach of 600 km and captured at optimum launch power are shown in Fig. 5.32. In addition to the DSP architectures used on the optical BtB measurements, a DBP algorithm with 6-steps (1 step per span) is applied as a replacement for the chromatic dispersion FDE. The DBP is utilized to analyze the data and to compensate for possible fiber non-linearities, primarily SPM effects because XPM compensation needs knowledge of the other DWDM channels. The light and dark blue markers show the performance with and without DBP including an LE after the carrier phase recovery block. The results indicate that minor non-linear effects from the fiber are introduced in comparable measure into all the channels during propagation. Only channel 29 (fourth from the right) exhibits a higher non-linear impact from the fiber, which may be caused by a slight power profile imbalance. With its special role, channel 29 serves here as an indicator for BRNN-SD performance capabilities in systems with higher fiber non-linearities. For readability, the DBP is therefore only applied to the classical advanced DSP schemes in channel 29.

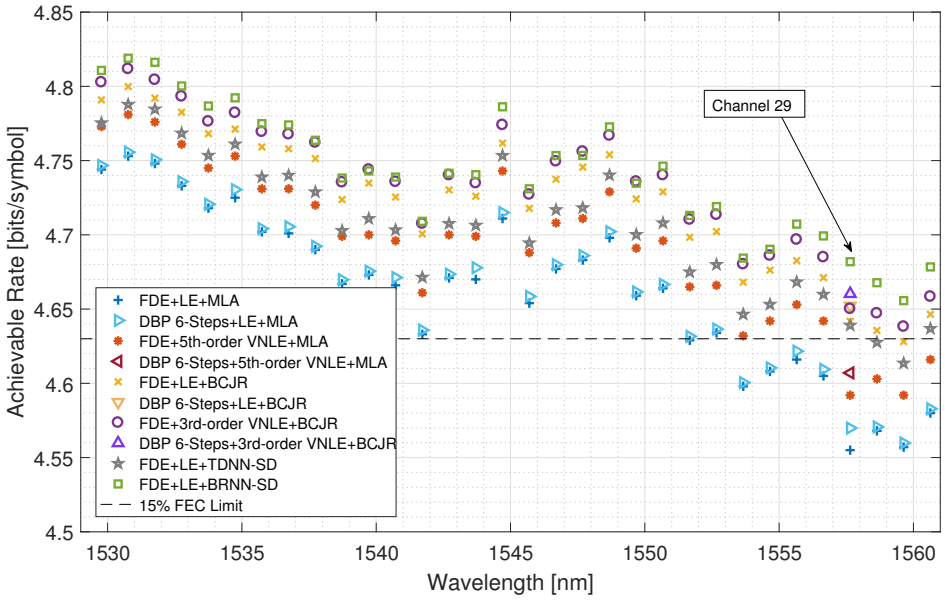


Figure 5.32: 800G 96 GBd DP-32QAM performance in terms of achievable rate for the 32-channel DWDM system over a 600 km G.654D fiber link at optimal launch power of 4 dBm. Note that, in the presence of stronger noise (wavelengths close to 1560 nm), accurate noise whitening is more important and that the higher noise level during training makes the training process more robust and reduces the generalization error.

Channel 1 to 32: The combination of LE, WF+BCJR exceeds the performance of the VNLE and the TDNN-SD, which indicates that colored noise effects caused by bandwidth limitation dominate. Nevertheless, compensating the non-linear effects from O/E components primarily by the VNLE improves the achievable rate and hence achieves best performance using conventional DSP schemes, see circled violet markers. However, Fig. 5.33 shows that the higher baudrate of 96 GBd instead of 92 GBd leads to stronger noise enhancement at high frequencies. In can be observed that the colored noise effect is not completely removed by the WF, even if the number of taps of the WF is increased to 4, which causes suboptimal performance.

The square green markers depict the performance of the BRNN-SD. In the 96 GBd DWDM scenario, where stronger colored noise effects and broader dispersed non-linear distortions occur [175], the BRNN-SD outperforms the best performing classical DSP combination of VNLE, and WF + BCJR. While in the BtB scenario with 92 GBd the classical algorithms can handle most of the impairments, in the case of transmission over fiber with 96 GBd, they cannot model all effects optimally. In comparison, the BRNN-SD is able to learn the

Dim.:	2-tap WF coeff.:
x_{i}	[1 0.32]
x_{q}	[1 0.28]
y_{i}	[1 0.41]
y_{q}	[1 0.45]

Dim.:	4-tap WF coeff.:
x_{i}	[1 0.25 -0.08 -0.06]
x_{q}	[1 0.21 -0.09 -0.03]
y_{i}	[1 0.31 -0.15 -0.03]
y_{q}	[1 0.34 -0.20 -0.03]

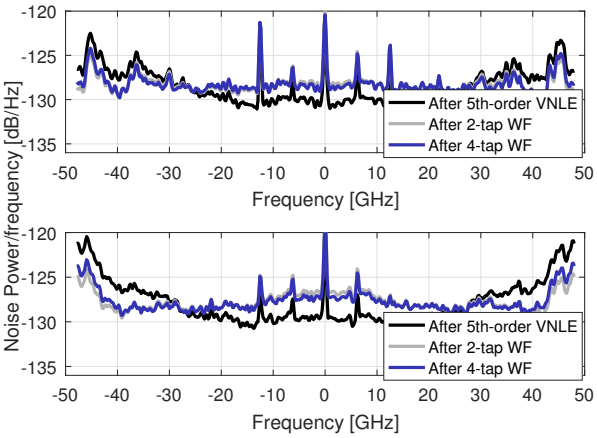


Figure 5.33: Noise PSDs of 96 GBd 32-QAM before and after a whitening filter with 2-taps of x-polarization (top) and y-polarization (bottom). It is assumed that there is no residual ISI in the sequence after VNLE and the noise is calculated as the difference between the VNLE output and the transmitted complex symbols. The noise peaks originate from the 50 GHz clock and are located at 12.5 GHz and 6.25 GHz.

channel and to adjust its model more accurately to strong colored noise effects as well as O/E component non-linearities arising at 96 GBd in a DWDM system.

Channel 29: On channel 29 the BRNN-SD achieves in comparison to the classical DSP combination of VNLE, and WF + BCJR the largest achievable rate gain. As mentioned above, channel 29 exhibits a higher non-linear impact from the fiber, as indicated by the DBP gain. The good BRNN-SD performance on channel 29 may indicate that recursive architectures are not only capable of recovering the performance of the BCJR, but may also recover the DBP gain, compensating for fiber non-linearities. For this, a more detailed study on a new set of data with a launch power sweep is provided in the next section, where SPM can be clearly distinguished.

5.3.4.4 Compensation of Self-Phase Modulation

The previous results in Fig. 5.32 indicated that BRNN-SD are not only able to compensate O/E component impairments but also fiber non-linearities. To examine this conjecture, the performance of the BRNN-SD is evaluated in the case of 80 km single-wavelength fiber transmission of coherent 80 GBd 800 Gb/s DP-64QAM using G.652 fiber. The reduced baudrate and the use of G.652 fiber turns non-linear effects from O/E components and fiber into the dominant source of distortions because band-limitation is less severe. In addition,

the deployment of DP-64QAM enables the division of inphase (I) and quadrature (Q) components and hence the comparison of independent and joint processing of I and Q. In case of independent processing four VNLEs/BRNN-SDs operate independently on the real dimension, while in case of joint processing I and Q are fed jointly into two VNLEs, respectively, one BRNN-SD per polarization as illustrated in Fig. 5.34.

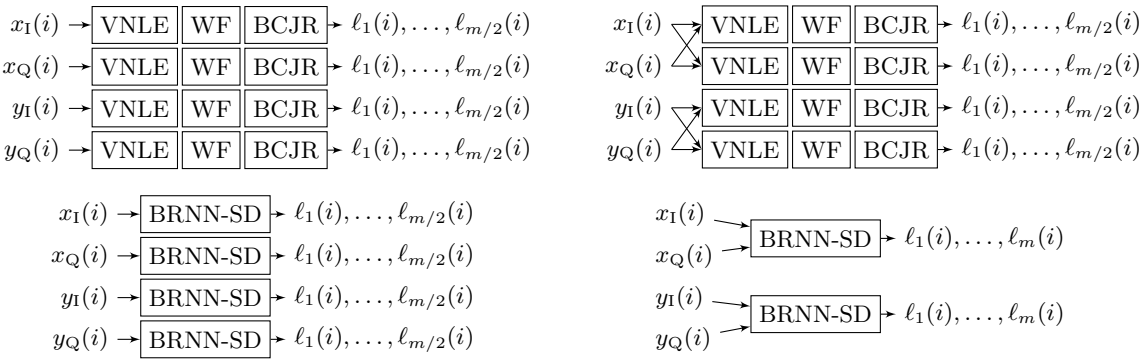


Figure 5.34: Independent (left) and joint (right) I and Q processing.

Fig. 5.35 displays the launch power sweep, i.e., the measured transmission performance in terms of achievable rate versus launch power. The small dots denote the performance of particular received frames, while the solid and dashed lines indicate the corresponding trends of

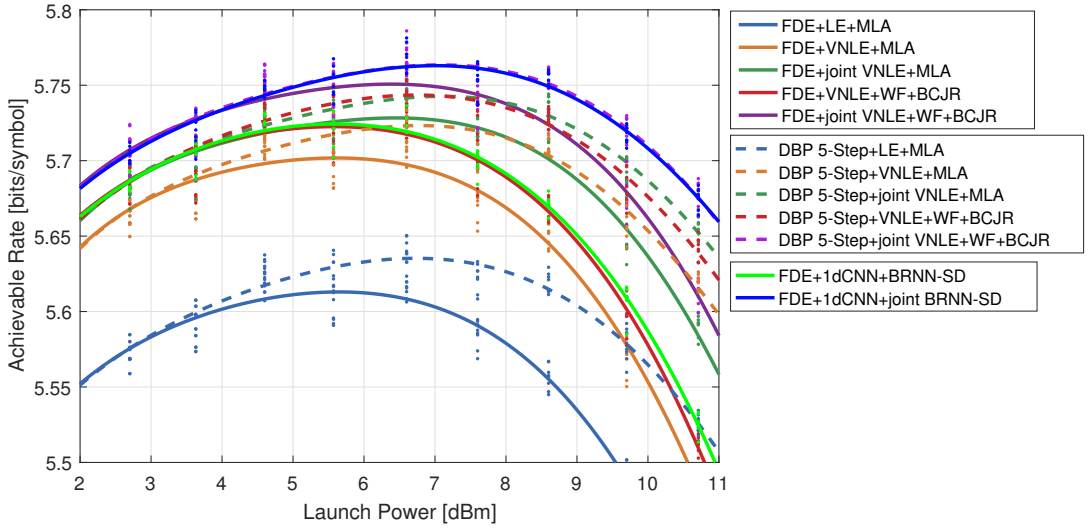
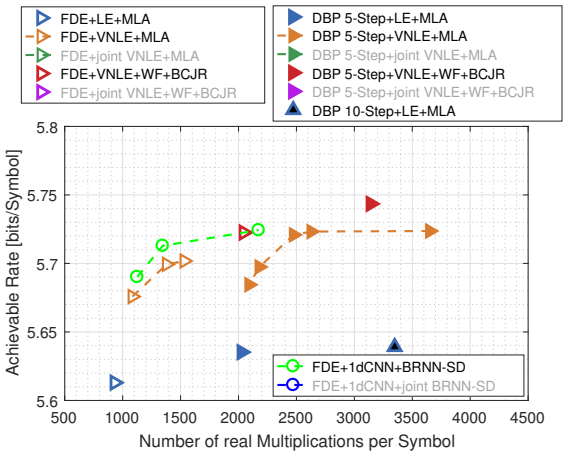
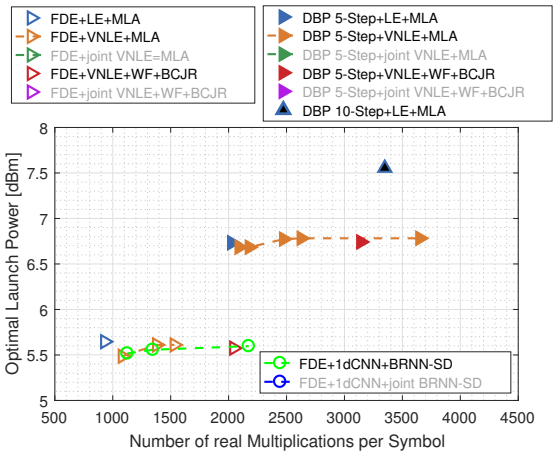


Figure 5.35: Single-wavelength 80 GBd DP-64QAM transmission over 80 km G.652 fiber. The small dots denote the performance of particular received frames, while the solid and dashed lines indicate the corresponding trends of the individual DSP architectures.

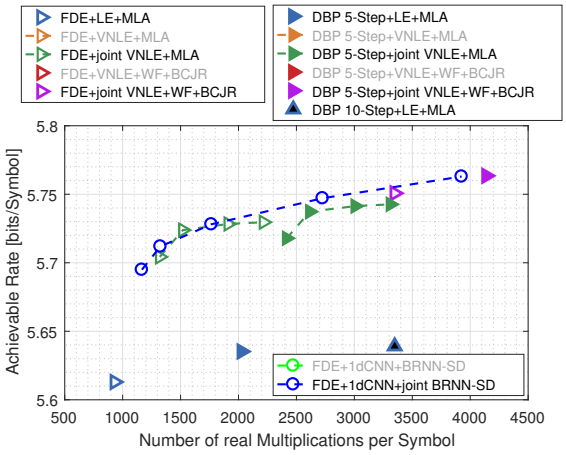
the individual applied DSP architectures. The solid lines represent the performance, when only chromatic dispersion is compensated by two complex FDEs with blocksize equal to 8192 and 25% overlap. The dashed lines represent the performance, when SPM is compensated additionally by a DBP algorithm. In general, a horizontal shift of the peak power to the right indicates the compensation of non-linear fiber effects while an equally distributed shift in vertical direction indicates linear and non-linear compensation of O/E components impairments. Fig. 5.36 depicts therefore an overall performance evaluation of the individual DSP algorithms in terms of maximum achievable rate and optimal launch power versus required number of real multiplications per polarization. The VNLE and BRNN-SD archi-



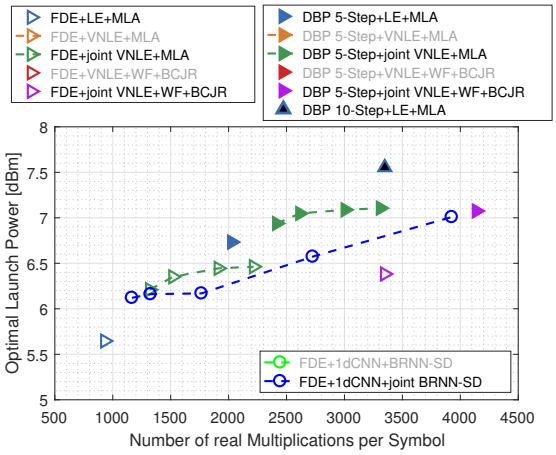
(a) Achievable rate related to number of real multiplications per symbol.



(b) Optimal launch power related to number of real multiplications per symbol.



(c) Achievable rate related to number of real multiplications per symbol.



(d) Optimal launch power related to number of real multiplications per symbol.

Figure 5.36: 80 GBd DP-64QAM optical transmission over 80km G.652 fiber.

tecture are optimized regrading complexity and achievable rate, i.e. the VNLE architectures are optimized with respect to the number of linear and non-linear taps, and the BRNN-SD to the number of neurons within the hidden layers. Particular architectures are labeled with their design.

Figs. 5.36a and 5.36b depict the maximum achievable rate, respectively the optimal launch power in case of independent I and Q processing. It can be observed, that the conventional stack combination, e.g., VNLE+WF+BCJR, as well as the BRNN-SD mainly compensate for component impairments, i.e., no launch power shift is obtained. A launch power shift is only obtained, if the prior DBP algorithm is employed, compensating SPM effects. As expected, the smaller the step size within the DBP, the higher the optimal launch power. However, the smaller the step size, the higher the computational complexity.

Figs. 5.36c and 5.36d depict the maximum achievable rate, respectively the optimal launch power in case of joint I and Q processing, where the compensation of non-linear phase noise correlated over time introduced by SPM is possible. In this case, the VNLE as well as the BRNN-SD are able to shift the optimal launch power by 0.8 dB, respectively 1.4 dB to the right. This indicates that the VNLE and the BRNN-SD, which operate at symbol rate, are not only capable to handle component impairments but also to a certain extend fiber non-linearities, which confirms the assumption in the previous 800G 96Gbd DP-32QAM measurements. However, in comparison to the DBP algorithm, whose fiber model is based on the NLSE, higher computational complexity is required to achieve the same optimal launch power, i.e., the well-kown analytical model outpaces the adaptive blackbox approach, especially in terms of complexity. Nevertheless, taking the component compensation into account, the BRNN-SD achieves with slightly lower computational complexity equal performance as the combination of classical stacked algorithms, consisting of a 5-step DBP, a joint VNLE, a WF and a BCJR detector. In particular, the BRNN-SD is able to model component impairments with less complexity while simultaneously compensating for SPM effects.

5.3.5 Summary

In this section a bidirectional recurrent neural network soft-demapper for bandlimited non-linear channels is proposed. In comparison to time delay neural networks soft-demapper, the recurrent architecture is able to handle colored noise effects caused by noise enhancement. In 92 GBd DP-32QAM optical back-to-back the proposed soft-demapper thus outperforms the time delay neural network soft-demapper and matches the performance of the reference DSP, consisting of a Volterra non-linear equalizer accompanied by a symbol-spaced whitening filter

and a BCJR detector. In 96 GBd DP-32QAM 800 Gb/s 32-channel DWDM transmissions over a 600 km G.654D fiber link the proposed approach even outperforms the reference DSP, due to stronger colored noise and component non-linearities within the received signal. However, in both considered setups linear impairments caused by limited bandwidth had a stronger impact on the performance than non-linear impairments from the components and fiber. In this cases, the BRNN is able to recover or slighly outperform the reference DSP, but is compared to classical algorithms computationally heavy in terms of multiplications.

In contrary, in a 80 GBd DP-64QAM transmission setup over a 80 km G.652 fiber link, where higher non-linear effects from the components as well as self-phase modulation effects from the fiber occur, the BRNN achieved with slightly less (5%) multipliers equal performance as the conventional stack combination, consisting of a 5-step digital back propagation algorithm, a joint VNLE, a WF and a BCJR detector. This confirms the assumption of concurrent SPM compensation and makes the BRNN a promising candidate for mid-term deployment in non-linearity-impaired and bandlimited high speed optical coherent transmission systems, especially in scenarios where no accurate analytic model is available and non-linear effects dominate.

Chapter 6

Conclusions and Outlook

The ever-increasing demand for digital information pushes towards new approaches. In this thesis, machine learning techniques for digital signal processing have been investigated and applied to coherent optical communication systems to increase their capacities. The main focus is set on next-generation high-bandwidth coherent short reach communication, where cost-efficiency and high capacity are the main parameters in the system design. In the following, the main results are summarized and an outlook on potential future research directions is presented.

Neural Network assisted Geometric Constellation Shaping In chapter 4, novel auto-encoder structures based on deep neural networks are introduced to design modulation formats optimized for lowest bit-error rate over amplified and unamplified links. Besides the geometry of the symbol constellation, the proposed method optimized also the bit mapping with and without label extension. The learned modulation formats are numerically and experimentally evaluated. In the numerical study performance gains of up to 0.9 dB and 0.4 dB in SNR and 0.5 dB and 0.58 dB in pSNR are obtained at the FEC limit for 32QAM and 128QAM, respectively. In contrary, in the experiment gains of up to 1.0 dB and 1.2 dB in OSNR and 0.25 dB and 0.55 dB in pOSNR are demonstrated at the FEC limit for 800 Gb/s DP-32QAM and 1 Tb/s DP-128QAM, respectively. The reported results show that machine learning techniques are a viable method for designing and optimizing complex optical communication systems.

Non-linear Compensation for Optical Transmission The trade-off between performance and high computational complexity of conventional compensation techniques have motivated investigations in alternatives based on DNNs. The study in Chapter 5 thus analyzed the capability of TDNNs and BRNNs for compensation of residual linear and non-linear impairments after the carrier phase recovery block.

In optical coherent 92 GBd DP-64QAM back-to-back measurements, where optical and electrical components non-linearities dominate, TDNNs proved to reflect systematic non-linearities more accurately than common 5th-order VNLEs. They either outperformed pruned VNLEs by 0.35 dB in OSNR with equal complexity or achieved the same performance with 65% less multipliers and hence lower complexity. In addition, we showed that the deep neural network state-of-the-art cross-entropy cost function for classification problems is equivalent to bitwise cross equivocation, maximizing an achievable rate. It is therefore optimal for training DSP components acting as soft-demappers in modern communication systems with soft-decision FEC, i.e., training of non-linear equalizers should be done via a soft-demapper w.r.t cross equivocation instead of mean square error.

Furthermore, in comparison to a feed-forward architecture, e.g., TDNN or VNLE, it has been shown that a BRNN architecture is capable to handle, next to non-linear ISI, colored noise effects caused by noise enhancement. In 92 GBd DP-32QAM optical back-to-back the BRNN thus outperformed the TDNN and matched the performance of the reference DSP, consisting of a VNLE accompanied by a symbol-spaced whitening filter and a BCJR detector. In 96 GBd DP-32QAM 800 Gb/s 32-channel DWDM transmissions over a 600 km G.654D fiber link, where stronger colored noise effects occur, the BRNN even outperformed the reference DSP. However, in both considered setups linear impairments caused by limited bandwidth had a stronger impact on the performance than non-linear impairments from the components and fiber. In this cases, the BRNN is able to recover or outperform the reference DSP, but is in comparison to the classical algorithm computationally heavy in terms of multiplications.

In contrary, in a 80 GBd DP-64QAM transmission setup over a 80 km G.652 fiber link, where higher non-linear effects from the components as well as SPM effects from the fiber occur, the BRNN achieved with slightly less (5%) multipliers equal performance as the stack combination of classical DSP algorithms, consisting of a 5-step DBP algorithm, a joint VNLE, a WF and a BCJR detector. This promotes the BRNN for mid-term deployment in non-linearity-impaired and bandlimited high speed optical coherent transmission systems, especially in scenarios where no accurate analytic model is available and non-linear effects dominate.

Future Research Directions

Neural Network assisted Geometric Constellation Shaping Both the simulated and experimental results show improved performances, but are as well highlighting the challenges in matching the offline learning architecture to the real system. In particular, the assumption of ideal channel compensation, i.e., AWGN is not valid in real systems. The unconsidered residual signal dependent impairments lead to a model discrepancy. Including further component and channel properties in the optimization chain like in [27] is therefore desirable, but faces challenges, like modeling of component and fiber imperfections together with their corresponding derivabilities. Nevertheless, the effects of such additional components and channel properties on the obtained constellation over amplified and unamplified links need to be further investigated to close the model discrepancy.

Non-linear Compensation for Optical Transmission In this thesis, the complexity aspects are estimated in number of multipliers, as they are by far the most expensive logic blocks. Future research may include a detailed study on low fixed-point representation and hence on quantization effects in combination with pruning. Reducing the power and the latency of neural networks is key if we want to integrate neural networks into edge modems with strict power and compute requirements.

Furthermore, due to the high quality of service requirements in optical transmission systems, i.e., post-FEC BER below 10^{-12} or 10^{-15}, the utilization of new digital signal processing (DSP) approaches in the data path of commercial products is very critical. DSP blocks based on neural networks must therefore be investigated in long term conditions to quantify their reliability and robustness against external effects, e.g., temperature changes. A first step, is real-time processing on field-programmable gate arrays (FPGAs).

Neural Network: Design and Trainability While the design of the input and output layers of a neural network depends on the input and output dimensions of the desired function, the design options of the hidden layer are numerous and interrelated. Regarding their size, general theoretical studies represent capacity upper bounds and state that minimally-sized neural networks down to a single hidden layer can represent a wide variety of models, when given appropriate parameters. However, those theorems do not describe the trainability of those parameters and are also mostly limited to neural networks whose connections are acyclic. In a real system, it is therefore currently still unclear, if the chosen neural network architecture design is optimal for a given task. More theoretical studies on the optimal architecture design and trainability are therefore required, including RNN, to displace hyper-parameter tuning techniques.

Neural Network: Activation functions In this thesis the most likely activation functions were applied. An interesting field for future research is to investigate customized activation function appropriate for optical communication, e.g., supporting optical impairment compensation. A possible approach is to use unsupervised learning techniques and to analyze the signal impairments in measurement data. These analytic algorithms operate without human bias or time constraints, computing every data combination to understand the data holistically and hence the hidden information. Such tailored activation functions may reduce the complexity of a neural network while maintaining its performance.

Neural Network: Information-theoretic perspective Information-theoretic methods have become the workhorse of several machine learning achievements over the past years [176], ranging from practical applications, e.g, the variational information bottleneck in representation learning [177], to theoretical investigation, e.g., the generalization bound induced by mutual information [178]. Specific investigations on machine learning in combination with optical communication from a information-theoretic perspective are however rare and are not discussed in this thesis, but will be mandatory and of pivotal importance to pave the way for neural networks in next-generation optical systems.

Appendix A

Loss Optimality

A.1 Optimal Linear and Non-linear Equalizer using Mean Square Error

In this section a simplified example based on BPSK is discussed to illustrated the origin of the jail window in Fig. 5.8 and also to emphasis the need of training a nonlinear equalizer w.r.t the bitwise cross equivocation (5.16) loss function. The example considered here is a simplification of the following manuscript [179, Chap. 5], which was recently published by G.Böcherer.

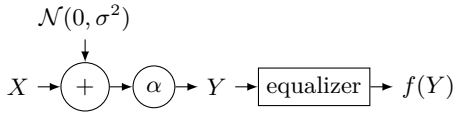


Figure A.1

Supposed $X \in \{-1, 1\}$ is transmitted and

$$Y = \gamma(X + \mathcal{N}(0, \sigma^2)) \tag{A.1}$$

is received, where σ^2 is the variance of the zero mean Gaussian noise source and γ the scaling factor. The corresponding zero-forcing equalizer is linear and realizes the function

$$f_{\text{zf}}(y) = \frac{1}{\gamma} y. \tag{A.2}$$

In practice a common objective for equalizers is to minimize the MSE

$$\mathbb{E}[(f(Y) - X)^2]. \tag{A.3}$$

Linear MSE Equalizer

A linear equalizer calculates $f_{\text{liner}}(y) = \alpha y$. By inserting $f_{\text{liner}}(y)$ into (A.3) and by evaluating the derivative w.r.t α and setting it equal to zero, we get [179, Eq. 5.21]

$$f_{\text{linear,mse}}(y) = \underbrace{\frac{1}{\lambda(1+\sigma^2)}}_{\alpha_{\text{MSE}}} y. \tag{A.4}$$

It can be observed, that the α which minimizes the MSE differs from the value in (A.2). However, with increasing SNR, the linear MSE equalizer approaches the zero-forcing equalizer.

Non-linear MSE Equalizer

In comparison to a linear equalizer, a non-linear equalizer has no restriction on its transfer function, in particular for each value of y, we can choose $f_{\text{non-linear}}(y)$ such that the MSE is minimized. Conditioned on $Y = y$, we write the MSE as [179, Eq. 5.22]

$$\mathbb{E}[(f_{\text{non-linear}}(y) - X)^2] = P_{X|Y}(-1|y)(f_{\text{non-linear}}(y) + 1)^2 + P_{X|Y}(1|y)(f_{\text{non-linear}}(y) - 1)^2. \tag{A.5}$$

By taking the derivative of the right-hand side w.r.t $f_{\text{non-linear}}(y)$ and setting it equal to zero, we get [179, Eq. 5.23]

$$f_{\text{non-linear,mse}}(y) = 2P_{X|Y}(-1|y) - 1. \tag{A.6}$$

The APP distribution $P_{X|Y}(-1|y)$ can be alternatively expressed as follows

$$P_{X|Y}(-1|y) = 1 - P_{X|Y}(0|y) = \frac{1}{1+\exp(-\ell(y))} = \text{sigmoid}(\ell(y)). \tag{A.7}$$

where

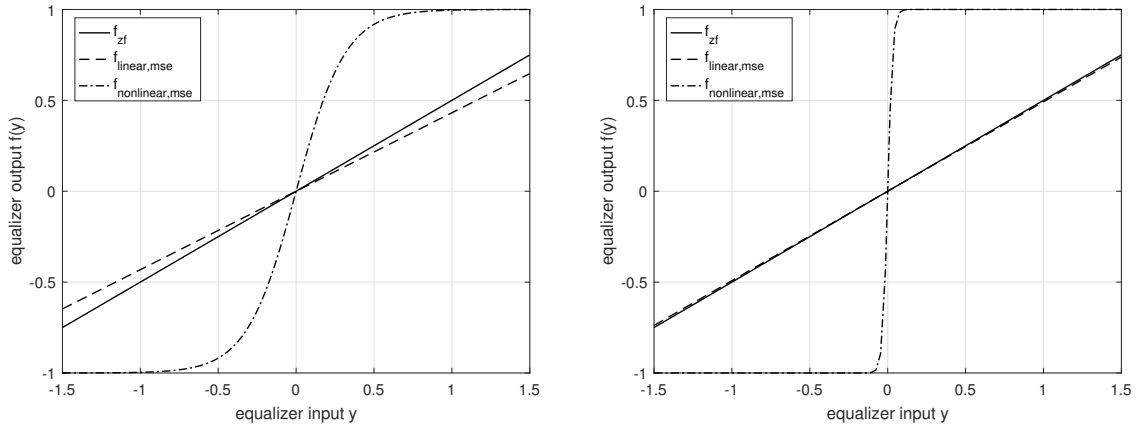
$$\ell(y) = \frac{2y}{\gamma\sigma^2}. \tag{A.8}$$

Combining (A.6), (A.7), (A.8) and considering $\tanh(x) = 2\text{sigmoid}(2x) - 1$, we finally get

$$f_{\text{non-linear,mse}}(y) = \tanh(\frac{y}{\gamma\sigma^2}) \tag{A.9}$$

Fig. A.2 shows the particular transfer functions of the zero-forcing, the linear and the non-linear MSE trained 1-tap equalizers. It can be observed that the optimal non-linear equalizer

trained w.r.t MSE exhibits saturation regions, resulting in the "jail window" shown in Fig. 5.8 in Section 5.1.

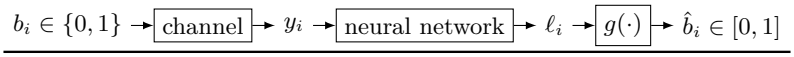


(a) Transfer functions of optimal, linear, and non-linear 1-tap equalizer considering BPSK over AWGN channel (8 dB SNR) and γ equals 2.

(b) Transfer functions of optimal, linear, and non-linear 1-tap equalizer considering BPSK over AWGN channel (18 dB SNR) and γ equals 2.

Figure A.2: Transfer functions of zero-forcing and linear as well as non-linear 1-tap equalizers trained w.r.t MSE at 8 and 18 dB SNR.

A.2 Soft-Demapping as Logistic Regression



soft-demapping loss (5.16)	logistic regression loss
$\mathcal{L}(b,\ell) = \begin{cases} \log(1+\exp(-\ell)), & \text{if } b=0 \\ \log(1+\exp(\ell)), & \text{if } b=1 \end{cases}$	$\mathcal{L}(b,\hat{b}) = \begin{cases} \log \frac{1}{\hat{b}}, & \text{if } b=0 \\ \log \frac{1}{1-\hat{b}}, & \text{if } b=1 \end{cases}$

Figure A.3: Bitwise soft-demapping as logistic regression.

We now show that our bitwise equivocation loss function $\mathcal{L}(b,\ell)$, defined in (5.16), that aims to maximize an achievable FEC rate is equivalent to the binary cross-entropy function $\mathcal{L}(b,\hat{b})$ applied to the output of a sigmoid activation function $\hat{b} = g(\ell)$, which is the common approach to binary classification ("logistic regression") in classic machine learning, see, e.g., [117, Sec. 4.3.2], [180, Sec. 8.3.1], [181, pp. 142–144]. We display the neural network and the two loss functions under consideration in Fig. A.3. Next we show that the two loss functions are equivalent. The sigmoid activation function [117, Sec. 2.4] is defined as

$$\hat{b} = \frac{1}{1+\exp(-\ell)}. \tag{A.10}$$

We now have

$$\mathcal{L}(b,\hat{b}) = \begin{cases} \log \frac{1}{\hat{b}}, & b=0 \\ \log \frac{1}{1-\hat{b}}, & b=1 \end{cases} \tag{A.11}$$

$$\overset{(A.10)}{=} \begin{cases} \log(1+\exp(-\ell)), & b=0 \\ \log(1+\exp(\ell)), & b=1 \end{cases} \tag{A.12}$$

$$= \log_2(1+\exp(-(1-2b)\ell)) \tag{A.13}$$

$$= \mathcal{L}(b,\ell) \tag{A.14}$$

which shows the equivalence of the loss functions.

A.3 Optimal Demapping with Neural Networks

For a trained neural network acting as soft-demapper, the neural network outputs ℓ and $\hat{b}$ are deterministic functions of the neural network input y. We define

$$q(0,y) = \hat{b}(y), \quad q(1,y) = 1-\hat{b}(y). \tag{A.15}$$

Note that since $0 \leq \hat{b}(y) \leq 1$, for each y, $q(\cdot,y)$ defines a distribution on $\{0,1\}$. We can now write the loss as

$$\mathcal{L}(b,\hat{b}(y)) = -\log_2 q(b,y). \tag{A.16}$$

For a sufficiently large number n of samples, the average loss is the cost that is minimized by training. We next derive an information-theoretic interpretation of the cost. To this end, we assume the channel output y takes values in a finite set $\mathcal{Y}$. Note that this assumption is in line with a practical setting, where y is discrete because of analog-to-digital conversion with finite resolution. For n samples, the cost is

$$\frac{1}{n}\sum_{i=1}^{n}[-\log_2 q(b_i,y_i)] \tag{A.17}$$

$$= \frac{1}{n}\sum_{b\in\{0,1\}}\sum_{y\in\mathcal{Y}}\sum_{i:b_i=b,y_i=y}[-\log_2 q(b,y)] \tag{A.18}$$

$$= \sum_{b\in\{0,1\}}\sum_{y\in\mathcal{Y}}[-\log_2 q(b,y)]\underbrace{\frac{\sum_{i:b_i=b,y_i=y}1}{n}}_{(\star)}, \tag{A.19}$$

where the term ($\star$) defines a joint distribution on $\{0,1\} \times \mathcal{Y}$. Let's introduce the random variable (RV) B for the channel input and the RV Y for the channel output observed by the neural network and define

$$P_{BY}(b,y) = \frac{\sum_{i:b_i=b,y_i=y} 1}{n}. \tag{A.20}$$

Then, we have

$$\frac{1}{n}\sum_{i=1}^{n}[-\log_2 q(b_i,y_i)] = \sum_{b\in\{0,1\}}\sum_{y\in\mathcal{Y}} P_{BY}(b,y)[-\log_2 q(b,y)] \tag{A.21}$$

$$= \sum_{y\in\mathcal{Y}} P_Y(y) \sum_{b\in\{0,1\}} P_{B|Y}(b|y)[-\log_2 q(b,y)] \tag{A.22}$$

$$\geq \sum_{y\in\mathcal{Y}} P_Y(y) \sum_{b\in\{0,1\}} P_{B|Y}(b|y)[-\log_2 P_{B|Y}(b|y)] \tag{A.23}$$

$$= \mathbb{H}(B|Y) \tag{A.24}$$

where (A.22) follows by Bayes' rule and (A.23) by the information inequality [182, Theorem 2.6.3] (recall that $q(\cdot,y)$ defines a distribution on $\{0,1\}$ for each y). We have equality in (A.23) if and only if

$$\hat{b}(y) = q(0,y) = P_{B|Y}(0|y) \tag{A.25}$$

$$1-\hat{b}(y) = q(1,y) = P_{B|Y}(1|y). \tag{A.26}$$

We next characterize the optimal ℓ. For $b=1$, we have

$$\frac{1}{1-\hat{b}(y)} = 1+\exp[\ell(y)] = \frac{1}{P_{B|Y}(1|y)} \tag{A.27}$$

$$\Rightarrow \exp[\ell(y)] = 1 - \frac{1}{P_{B|Y}(1|y)} = \frac{P_{B|Y}(0|y)}{P_{B|Y}(1|y)} \tag{A.28}$$

$$\Rightarrow \ell(y) = \log\frac{P_{B|Y}(0|y)}{P_{B|Y}(1|y)}. \tag{A.29}$$

(The same result is obtained when considering $b=0$). We conclude that with the considered loss functions, the neural network learns via $\hat{b}$ the APP distribution $P_{B|Y}$ and via ℓ, it learns the logarithmic APP ratio $\log\frac{P_{B|Y}(0|y)}{P_{B|Y}(1|y)}$, thereby achieving the minimum cost of $\mathbb{H}(B|Y)$.

List of Abbreviations

1D	one-dimensional
1D-CNN	one-dimensional convolutional neural network
2D	two-dimensional
ADAM	adaptive moment estimation
ADC	analog-to-digital converter
APP	posteriori probability
ASE	amplified spontaneous emission
ASIC	application-specific integrated circuit
AWGN	additive white Gaussian noise
BCE	bitwise cross equivocation
BER	bit error ratio
BICM	bit-interleaved coded modulation
BPSK	binary phase shift keying
BPTT	backpropagation through time
BRNN	bidirectional recurrent neural network
BRNN-SD	bidirectional recurrent neural network soft demapper
BtB	back-to-back
CAZAC	constant amplitude zero auto-correlation
CCDM	constant composition distribution matcher
CFO	carrier frequency offset
CMOS	complementary metal-oxide semiconductor
CPR	carrier phase recovery
CUT	channel under test

DA	driver amplifier
DAC	digital-to-analog converter
DBP	digital back-propagation
DCI	data center interconnect
DD	direct detection
DGD	differential group delay
DM	distribution matcher
DNL	differential non-linearity
DNN	deep neural network
DSP	digital signal processing
DWDM	dense wavelength division multiplexing
ECL	external cavity laser
EDFA	Erbium-doped fiber amplifier
ENOB	effective number of bits
FDE	frequency divison equalizer
FEC	forward error correction
FFT	fast Fourier transform
FIR	finite impulse response
FPGA	field-programmable gate array
GMI	general mutual information
GN	Gaussian noise
GNLSE	general non-linear Schrödinger equation
GVD	group-velocity dispersion
H-tanh	hard tanh
HB-μICR	high bandwidth micro integrated coherent receiver
HD	hard-decision
HMM	hidden Markov model
I	inphase
I-tanh	Interpolated tanh
IIR	infinite impulse response

IM intensity modulation
IMD intermodulation distortion
INL integral non-linearity
IoT Internet of Things
ISI inter-symbol interference

LDPC low-density parity-check
LE linear equalizer
LLRs log-likelihood ratios
LMS least mean squars
LO local oscillator
LS standard least-squares
LUT look-up table

M2M Machine-To-Machine
MAP maximum-a-posterior
MF matched filter
MIMO multiple input multiple output
MLA max-log approximation
MMSE minimum mean square error
MSE mean square error
MZM Mach-Zehnder modulator

NF noise figure
NLSE non-linear Schrödinger equation

O/E optical/electrical
OH overhead
OIF Optical Internetworking Forum
OOK on-off keying
OSA optical spectrum analyzer
OSNR optical signal-to-noise ratio

PAPR peak-to-average power ratio
PAS probabilistic amplitude shaping
PBC polarization beam combiner

PBS	polarization beam splitter
PhD	Doctor of Philosophy
PMD	polarization mode dispersion
PSD	power spectral density
pSNR	peak SNR
Q	quadrature
QAM	quadrature amplitude modulation
ReLU	rectified linear unit
RF	radio frequency
RLS	recursive least square
RNN	recurrent neural network
RRC	root-raised-cosine
RTRL	real time recurrent learning
SD	soft decision
SE	spectral efficiencies
SER	symbol error ratio
SGD	stochastic gradient descent
SINR	signal-to-interference-and-noise ratio
SMF	single-mode fiber
SNR	signal-to-noise ratio
SPM	self phase modulation
SpS	sample per symbol
SSFM	split-step Fourier method
SSMF	standard single-mode fiber
tanh	hyperbolic tangent
TDNN	time delay neural network
TDNN-SD	time delay neural network soft-demapper
TDNNE	time delay neural network equalizer
TIA	trans-impedance amplifier
TR	timing recovery
VNLE	Volterra non-linear equalizer

WDM	wavelength division multiplexing
WF	whitening filter
XPM	cross phase modulation
ZF	zero-forcing

List of Symbols

α	fiber attenuation
β_3	third-order disperion
β_2	phase constant of dispersion
Δn	fiber birefringence
$\Delta\tau_s$	differential group delay
γ	nonlinear fiber parameter
ω	angular frequency
σ^2	noise variance
θ_s	polarization phase
B_{ref}	reference bandwidth
$\mathcal{L}$	cost function
D	dispersion parameter
L	fiber length
N_0	power spectral density of white noise
N_{ASE}	spectral density of the ASE-noise in one polarization
P	power
R_{C}	code-rate
R_{s}	symbol rate
R	achievable rate
V_π	half-wave switching voltage of an MZM
c	speed of light
δ	extinction ratio

Bibliography

[1] M. Castells *et al.*, *Information technology, globalization and social development*, 114. UNRISD Geneva, 1999 (cit. on p. 1).

[2] M. K. Weldon, *The future X network: a Bell Labs perspective*. CRC press, 2016 (cit. on p. 1).

[3] P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: The previous 20 and the next 20 years", *Optics express*, vol. 26, no. 18, pp. 24 190–24 239, 2018 (cit. on p. 1).

[4] Cisco, *Cisco annual internet report (2018–2023) white paper*, [online]. Availabel: `https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/whitepaper-c11-741490.html`, accessed: 26-03-2021 (cit. on p. 1).

[5] T. Schneider, *Nonlinear optics in telecommunications*. Springer Science & Business Media, 2004 (cit. on pp. 2, 17–21, 24).

[6] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications", *Nature*, vol. 411, no. 6841, pp. 1027–1030, 2001 (cit. on p. 2).

[7] 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 (cit. on pp. 2, 32).

[8] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel", in *2014 IEEE International Symposium on Information Theory*, IEEE, 2014, pp. 2794–2798 (cit. on p. 2).

[9] A. Napoli *et al.*, "Reduced complexity digital back-propagation methods for optical communication systems", *Journal of lightwave technology*, vol. 32, no. 7, pp. 1351–1362, 2014 (cit. on pp. 2, 24, 25).

[10] V. G. Makhan'kov and O. K. O. Pashaev, "Nonlinear Schrodinger equation with noncompact isogroup", *Teoreticheskaya i Matematicheskaya Fizika*, vol. 53, no. 1, pp. 55–67, 1982 (cit. on p. 2).

[11] A Rezania, J. Cartledge, A Bakhshali, and W.-Y. Chan, "Compensation schemes for transmitter-and receiver-based pattern-dependent distortion", *IEEE Photon. Technol. Lett.*, vol. 28, no. 22, pp. 2641–2644, 2016 (cit. on pp. 2, 27, 56).

[12] J. C. Cartledge, "Volterra equalization for nonlinearities in optical fiber communications", in *Signal Processing in Photonic Communications*, 2017, SpTu2F–2 (cit. on pp. 2, 27, 56).

[13] J. Hendler, "Avoiding another AI winter", *IEEE Intelligent Systems*, vol. 23, no. 02, pp. 2–4, 2008 (cit. on p. 3).

[14] R. Rios-Müller, J. M. Estarán, and J. Renaudier, "Experimental estimation of optical nonlinear memory channel conditional distribution using deep neural networks", in *Proc. Opt. Fiber Commun. Conf.*, 2017, W2A–51 (cit. on pp. 3, 55).

[15] A. G. Reza and J.-K. K. Rhee, "Nonlinear equalizer based on neural networks for PAM-4 signal transmission using DML", *IEEE Photon. Technol. Lett.*, vol. 30, no. 15, pp. 1416–1419, 2018 (cit. on pp. 3, 55).

[16] P. Li, L. Yi, L. Xue, and W. Hu, "100Gbps IM/DD transmission over 25km SSMF using 20G-class DML and PIN enabled by machine learning", in *Proc. Opt. Fiber Commun. Conf.*, 2018, W2A–46 (cit. on pp. 3, 55).

[17] T. Koike-Akino, D. S. Millar, K. Parsons, and K. Kojima, "Fiber nonlinearity equalization with multi-label deep learning scalable to high-order DP-QAM", in *Signal Processing in Photonic Communications*, Optical Society of America, 2018, SpM4G–1 (cit. on pp. 3, 56).

[18] V. Kamalov *et al.*, "Evolution from 8QAM live traffic to PS 64-QAM with neural-network based nonlinearity compensation on 11000 km open subsea cable", in *Proc. Opt. Fiber Commun. Conf.*, Optical Society of America, 2018, Th4D–5 (cit. on pp. 3, 56).

[19] S. Zhang *et al.*, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks", *Nature communications*, vol. 10, no. 1, pp. 1–8, 2019 (cit. on pp. 3, 56).

[20] M. Schaedler, M. Kuschnerov, S. Pachnicke, C. Bluemm, F. Pittala, and X. Changsong, "Subcarrier power loading for coherent optical OFDM optimized by machine learning", in *Proc. Opt. Fiber Commun. Conf.*, 2019, M2H–3 (cit. on pp. 3, 56).

[21] O. Sidelnikov, A. Redyuk, and S. Sygletos, "Equalization performance and complexity analysis of dynamic deep neural networks in long haul transmission systems", *Optics Express*, vol. 26, no. 25, pp. 32 765–32 776, 2018 (cit. on pp. 3, 56).

[22] X. Wu, J. A. Jargon, R. A. Skoog, L. Paraschis, and A. E. Willner, "Applications of artificial neural networks in optical performance monitoring", *Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3580–3589, 2009 (cit. on p. 3).

[23] T. B. Anderson, A. Kowalczyk, K. Clarke, S. D. Dods, D. Hewitt, and J. C. Li, "Multi impairment monitoring for optical networks", *Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3729–3736, 2009 (cit. on p. 3).

[24] J. A. Jargon, X. Wu, H. Y. Choi, Y. C. Chung, and A. E. Willner, "Optical performance monitoring of qpsk data channels by use of neural networks trained with parameters derived from asynchronous constellation diagrams", *Optics Express*, vol. 18, no. 5, pp. 4931–4938, 2010 (cit. on p. 3).

[25] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, "Data-analytics-based optical performance monitoring technique for optical transport networks", in *Optical Fiber Communication Conference*, Optical Society of America, 2018, Tu3E–3 (cit. on p. 3).

[26] T. O'shea and J. Hoydis, "An introduction to deep learning for the physical layer", *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017 (cit. on pp. 3, 45).

[27] B. Karanov *et al.*, "End-to-end deep learning of optical fiber communications", *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, 2018 (cit. on pp. 3, 105).

[28] M. Schaedler, C. Bluemm, M. Kuschnerov, F. Pittalà, S. Calabrò, and S. Pachnicke, "Deep neural network equalization for optical short reach communication", *Applied Sciences*, vol. 9, no. 21, p. 4675, 2019 (cit. on p. 3).

[29] M. Schaedler, G. Boecherer, and S. Pachnicke, "Soft-demapping for short reach optical communication: A comparison of deep neural networks and Volterra series", *J. Lightw. Technol.*, 2021 (cit. on pp. 3, 56).

[30] M. Schädler *et al.*, "Recurrent neural network soft-demapping for nonlinear ISI in 800Gbit/s DWDM coherent optical transmissions", *J. Lightwave Technol.*, vol. 39, no. 16, pp. 5278–5286, 2021 (cit. on p. 3).

[31] M. Schaedler, F. Pittalà, G. Böcherer, C. Bluemm, M. Kuschnerov, and S. Pachnicke, "Recurrent neural network soft-demapping for nonlinear ISI in 800Gbit/s DWDM coherent optical transmissions", in *2020 European Conference on Optical Communications (ECOC)*, 2020, pp. 1–4. DOI: 10.1109/ECOC48923.2020.9333204 (cit. on p. 3).

[32] M. Schaedler *et al.*, "Neural network assisted geometric shaping for 800Gbit/s and 1Tbit/s optical transmission", in *Proc. Opt. Fiber Commun. Conf.*, IEEE, 2020, pp. 1–3 (cit. on p. 3).

[33] M. Schaedler, S. Calabrò, F. Pittalà, C. Bluemm, M. Kuschnerov, and S. Pachnicke, "Neural network-based soft-demapping for nonlinear channels", in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2020, pp. 1–3 (cit. on pp. 3, 56).

[34] M. Schaedler, F. Pittalà, S. Calabrò, G. Böcherer, C. Bluemm, and S. Pachnicke, "Recurrent neural network soft demapping for mitigation of fiber nonlinearities and ISI", in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2021, pp. 1–3 (cit. on p. 3).

[35] E. Biglieri, A. Gersho, R Gitlin, and T. Lim, "Adaptive cancellation of nonlinear intersymbol interference for voiceband data transmission", *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 5, pp. 765–777, 1984 (cit. on p. 7).

[36] S Takac and S Stojanović, "Characteristics of laser light", *Medicinski pregled*, vol. 52, no. 1-2, pp. 29–34, 1999 (cit. on p. 7).

[37] M. Seimetz, "Laser linewidth limitations for optical systems with high-order modulation employing feed forward digital carrier phase estimation", in *OFC/NFOEC 2008-2008 Conference on Optical Fiber Communication/National Fiber Optic Engineers Conference*, IEEE, 2008, pp. 1–3 (cit. on p. 7).

[38] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations", (cit. on p. 7).

[39] K. Schuh *et al.*, "Single carrier 1.2 Tbit/s transmission over 300 km with PM-64 QAM at 100 GBaud", in *Optical Fiber Communication Conference*, Optical Society of America, 2017, Th5B–5 (cit. on p. 8).

[40] F. Buchali *et al.*, "1.3-Tb/s single-channel and 50.8-Tb/s WDM transmission over field-deployed fiber", in *45th European Conference on Optical Communication (ECOC 2019)*, IET, 2019, pp. 1–4 (cit. on p. 8).

[41] F. Buchali *et al.*, "1.52 Tb/s single carrier transmission supported by a 128 GSa/s SiGe DAC", in *Optical Fiber Communication Conference*, Optical Society of America, 2020, Th4C–2 (cit. on p. 8).

[42] V. Bajaj, F. Buchali, M. Chagnon, S. Wahls, and V. Aref, "Single-channel 1.61 Tb/s optical coherent transmission enabled by neural network-based digital pre-distortion", in *2020 European Conference on Optical Communications (ECOC)*, IEEE, 2020, pp. 1–4 (cit. on p. 8).

[43] T. Instruments, "Understanding data converters", *Application report*, 1995 (cit. on pp. 8, 23).

[44] M. M. GmbH, *100 GS/s digital to analog converter systems data sheet*, [online]. `https://micram.net/vega-dac4-products-dac10000-family-brochure-2`, accessed: 21-10-2021 (cit. on p. 8).

[45] S. C. T. AG, *Datasheet SHF S804 B*, [online]. `https://www.shf-communication.com/wp-content/uploads/datasheets/datasheet_shf_s804_b.pdf`, accessed: 21-10-2021 (cit. on p. 9).

[46] B. H. Kolner and D. W. Dolfi, "Intermodulation distortion and compression in an integrated electrooptic modulator", *Applied Optics*, vol. 26, no. 17, pp. 3676–3680, 1987 (cit. on pp. 9, 10).

[47] S. Walklin and J. Conradi, "Effect of mach-zehnder modulator DC extinction ratio on residual chirp-induced dispersion in 10-Gb/s binary and AM-PSK duobinary lightwave systems", *IEEE Photonics Technology Letters*, vol. 9, no. 10, pp. 1400–1402, 1997 (cit. on p. 11).

[48] Z. Huang, X. Zhang, S. Zheng, X. Jin, and H. Chi, "Any bias point control of mach-zehnder electrooptic modulator and its applications in optimization of radio-over-fiber links", in *2011 International Topical Meeting on Microwave Photonics jointly held with the 2011 Asia-Pacific Microwave Photonics Conference*, IEEE, 2011, pp. 218–221 (cit. on p. 11).

[49] E. Zehavi, "8-psk trellis codes for a rayleigh channel", *IEEE Transactions on Communications*, vol. 40, no. 5, pp. 873–884, 1992 (cit. on p. 12).

[50] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation", *IEEE transactions on information theory*, vol. 44, no. 3, pp. 927–946, 1998 (cit. on pp. 12, 55).

[51] A. G. i Fabregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation", 2008 (cit. on p. 12).

[52] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes", in *Proceedings of ICC'93-IEEE International Conference on Communications*, IEEE, vol. 2, 1993, pp. 1064–1070 (cit. on p. 12).

[53] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes", *IEEE Transactions on communications*, vol. 44, no. 10, pp. 1261–1271, 1996 (cit. on p. 12).

[54] R. Gallager, "Low-density parity-check codes", *IRE Transactions on information theory*, vol. 8, no. 1, pp. 21–28, 1962 (cit. on p. 12).

[55] D. J. MacKay and R. M. Neal, "Near shannon limit performance of low density parity check codes", *Electronics letters*, vol. 32, no. 18, p. 1645, 1996 (cit. on p. 12).

[56] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding", in *1999 IEEE International Conference on Communications (Cat. No. 99CH36311)*, IEEE, vol. 2, 1999, pp. 858–863 (cit. on p. 12).

[57] A. Martinez, A. G. i Fabregas, G. Caire, and F. M. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective", *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2756–2765, 2009 (cit. on p. 12).

[58] C. E. Shannon, "A mathematical theory of communication", *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948 (cit. on p. 13).

[59] T. Fehenberger, "Analysis and optimization of coded modulation for nonlinear fiber-optic communication systems", Ph.D. dissertation, Technische Universität München, 2017 (cit. on pp. 13, 14, 17, 19, 30, 31).

[60] S. Lin and D. J. Costello, *Error control coding*, 4. Prentice hall Scarborough, 2001, vol. 2 (cit. on p. 13).

[61] I. P. Kaminow and T. Li, *Optical fiber telecommunications IV-B: systems and impairments*. Elsevier, 2002, vol. 2 (cit. on pp. 13, 14).

[62] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, "Replacing the soft-decision FEC limit paradigm in the design of optical communication systems", *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, 2015 (cit. on pp. 13, 31, 55).

[63] G. Forney, R Gallager, G Lang, F Longstaff, and S Qureshi, "Efficient modulation for band-limited channels", *IEEE journal on selected areas in communications*, vol. 2, no. 5, pp. 632–647, 1984 (cit. on p. 14).

[64] G. Böcherer and R. Mathar, "Matching dyadic distributions to channels", in *2011 Data Compression Conference*, IEEE, 2011, pp. 23–32 (cit. on p. 14).

[65] D. Gómez-Barquero, C. Douillard, P. Moss, and V. Mignone, "Dvb-ngh: The next generation of digital broadcast services to handheld devices", *IEEE Transactions on Broadcasting*, vol. 60, no. 2, pp. 246–257, 2014 (cit. on p. 15).

[66] F. Steiner and G. Böcherer, "Comparison of geometric and probabilistic shaping with application to ATSC 3.0", in *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*, VDE, 2017, pp. 1–6 (cit. on pp. 15, 50).

[67] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation", *IEEE Transactions on communications*, vol. 63, no. 12, pp. 4651–4665, 2015 (cit. on p. 15).

[68] P. Schulte and G. Böcherer, "Constant composition distribution matching", *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, 2015 (cit. on p. 15).

[69] F. Pittala, I. Slim, A. Mezghani, and J. A. Nossek, "Training-aided frequency-domain channel estimation and equalization for single-carrier coherent optical transmission systems", *J. Lightw. Technol.*, vol. 32, no. 24, pp. 4849–4863, 2014 (cit. on pp. 16, 20, 24, 26).

[70] M. Mazurczyk, "Spectral shaping in long haul optical coherent systems with high spectral efficiency", *Journal of lightwave technology*, vol. 32, no. 16, pp. 2915–2924, 2014 (cit. on p. 16).

[71] E. Voges and K. Petermann, *Optische Kommunikationstechnik: Handbuch für Wissenschaft und Industrie*. Springer-Verlag, 2019 (cit. on p. 17).

[72] G. P. Agrawal, *Fiber-optic communication systems*. John Wiley & Sons, 2012, vol. 222 (cit. on pp. 18–20).

[73] F. N. Hauske, *The Importance of Digital Signal Processing in High Speed Optical Receivers: Equalization, Impairment Compensation and Performance Monitoring*. Köster, 2013 (cit. on p. 19).

[74] J. Kerr, "A new relation between electricity and light: Dielectric media birefringent (second paper)", *Philosophical Magazine, S. 4*, vol. 50, no. 333, pp. 446–458, 1875 (cit. on p. 20).

[75] F. R. Kschischang, "Fiber-optic communication in the nonlinear frequency domain", 2014 (cit. on p. 20).

[76] F. N. Hauske, *The Importance of Digital Signal Processing in High Speed Optical Receivers*. Verlag Dr. Köster, Berlin, 2013 (cit. on p. 22).

[77] D. Van Den Borne, "Robust optical transmission systems: Modulation and equalization", 2008 (cit. on p. 22).

[78] P. Stoffa, J. T. Fokkema, R. de Luna Freire, and W. Kessinger, "Split-step Fourier migration", *Geophysics*, vol. 55, no. 4, pp. 410–421, 1990 (cit. on p. 24).

[79] R. A. Soriano, F. N. Hauske, N. G. Gonzalez, Z. Zhang, Y. Ye, and I. T. Monroy, "Chromatic dispersion estimation in digital coherent receivers", *Journal of Lightwave Technology*, vol. 29, no. 11, pp. 1627–1637, 2011 (cit. on p. 24).

[80] C. Malouin, P. Thomas, B. Zhang, J. O'Neil, and T. Schmidt, "Natural expression of the best-match search godard clock-tone algorithm for blind chromatic dispersion estimation in digital coherent receivers", in *Signal Processing in Photonic Communications*, Optical Society of America, 2012, SpTh2B–4 (cit. on p. 24).

[81] C. Malouin, M. Arabaci, P. Thomas, B. Zhang, T. Schmidt, and R. Marcoccia, "Efficient, non-data-aided chromatic dispersion estimation via generalized, FFT-based sweep", in *National Fiber Optic Engineers Conference*, Optical Society of America, 2013, JW2A–45 (cit. on p. 24).

[82] Q. Sui, A. P. T. Lau, and C. Lu, "Fast and robust blind chromatic dispersion estimation using auto-correlation of signal power waveform for digital coherent systems", *Journal of Lightwave Technology*, vol. 31, no. 2, pp. 306–312, 2012 (cit. on p. 24).

[83] N. Stojanovic, B. Mao, and F. Karinou, "Efficient and low-complexity chromatic dispersion estimation in coherent optical systems", in *2013 21st Telecommunications Forum Telfor (TELFOR)*, IEEE, 2013, pp. 153–156 (cit. on p. 24).

[84] M Kuschnerov, F. Hauske, K Piyawanno, B Spinnler, A Napoli, and B Lankl, "Adaptive chromatic dispersion equalization for non-dispersion managed coherent systems", in *Proc. Opt. Fiber Commun. Conf.*, 2009, OMT1 (cit. on p. 24).

[85] R. Corsini *et al.*, "Blind adaptive chromatic dispersion compensation and estimation for DSP-based coherent optical systems", *Journal of lightwave technology*, vol. 31, no. 13, pp. 2131–2139, 2013 (cit. on p. 24).

[86] I. Slim, A. Mezghani, L. G. Baltar, J. Qi, F. N. Hauske, and J. A. Nossek, "Delayed single-tap frequency-domain chromatic-dispersion compensation", *IEEE Photonics Technology Letters*, vol. 25, no. 2, pp. 167–170, 2012 (cit. on p. 24).

[87] Y. Gao, J. H. Ke, K. P. Zhong, J. C. Cartledge, and S. S.-H. Yam, "Assessment of intrachannel nonlinear compensation for 112 Gb/s dual-polarization 16QAM systems", *Journal of lightwave technology*, vol. 30, no. 24, pp. 3902–3910, 2012 (cit. on p. 25).

[88] D. S. Millar *et al.*, "Mitigation of fiber nonlinearity using a digital coherent receiver", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1217–1226, 2010 (cit. on p. 25).

[89] C. Fougstedt, M. Mazur, L. Svensson, H. Eliasson, M. Karlsson, and P. Larsson-Edefors, "Time-domain digital back propagation: Algorithm and finite-precision implementation aspects", in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, 2017, pp. 1–3 (cit. on p. 25).

[90] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM", *IEEE transactions on communications*, vol. 45, no. 12, pp. 1613–1621, 1997 (cit. on p. 26).

[91] M. Kuschnerov *et al.*, "Dsp for coherent single-carrier receivers", *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3614–3622, 2009 (cit. on p. 26).

[92] F. Pittalà, M. Msallem, F. N. Hauske, Y. Ye, I. T. Monroy, and J. A. Nossek, "Frequency domain training-aided channel estimation and equalization in time-varying optical transmission systems", in *IEEE Photonics Conference 2012*, IEEE, 2012, pp. 455–456 (cit. on p. 26).

[93] K. Mueller and M. Muller, "Timing recovery in digital synchronous data receivers", *IEEE transactions on communications*, vol. 24, no. 5, pp. 516–531, 1976 (cit. on p. 26).

[94] F. Gardner, "A BPSK/QPSK timing-error detector for sampled receivers", *IEEE Transactions on communications*, vol. 34, no. 5, pp. 423–429, 1986 (cit. on pp. 26, 27).

[95] M. Oerder and H. Meyr, "Digital filter and square timing recovery", *IEEE Transactions on communications*, vol. 36, no. 5, pp. 605–612, 1988 (cit. on p. 27).

[96] D Godard, "Passband timing recovery in an all-digital modem receiver", *IEEE Transactions on Communications*, vol. 26, no. 5, pp. 517–523, 1978 (cit. on p. 27).

[97] F. Munier, E. Alpman, T. Eriksson, A. Svensson, and H. Zirath, "Estimation of phase noise for QPSK modulation over AWGN channels", in *Proc. GigaHertz 2003 Symp*, Citeseer, 2003, pp. 4–5 (cit. on p. 27).

[98] E. Ip and J. M. Kahn, "Feedforward carrier recovery for coherent optical communications", *Journal of Lightwave Technology*, vol. 25, no. 9, pp. 2675–2692, 2007 (cit. on p. 27).

[99] T. Fehenberger, N. Hanik, T. A. Eriksson, P. Johannisson, and M. Karlsson, "On the impact of carrier phase estimation on phase correlations in coherent fiber transmission", in *2015 Tyrrhenian International Workshop on Digital Communications (TIWDC)*, IEEE, 2015, pp. 35–38 (cit. on p. 27).

[100] A. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission", *IEEE Transactions on Information theory*, vol. 29, no. 4, pp. 543–551, 1983 (cit. on p. 27).

[101] A. Spalvieri and L. Barletta, "Pilot-aided carrier recovery in the presence of phase noise", *IEEE transactions on communications*, vol. 59, no. 7, pp. 1966–1974, 2011 (cit. on p. 27).

[102] L. Guan, *FPGA-based Digital Convolution for Wireless Applications*. Springer, 2017 (cit. on p. 28).

[103] T. Wettlin, S. Pachnicke, T. Rahman, J. Wei, S. Calabro, and N. Stojanovic, "Complexity reduction of Volterra nonlinear equalization for optical short-reach IM/DD systems", in *Photonic Networks; 21th ITG-Symposium*, VDE, 2020, pp. 1–6 (cit. on p. 28).

[104] C. Tarver, A. Balatsoukas-Stimming, and J. R. Cavallaro, "Design and implementation of a neural network based predistorter for enhanced mobile broadband", in *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, IEEE, 2019, pp. 296–301 (cit. on pp. 28, 29).

[105] K. J. Layton, A. Mehboob, W. G. Cowley, and G. Lechner, "Improved demapping for channels with data-dependent noise", *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, pp. 1–12, 2018 (cit. on p. 29).

[106] L. Szczecinski and A. Alvarado, *Bit-interleaved coded modulation: fundamentals, analysis and design*. John Wiley & Sons, 2015 (cit. on pp. 29, 55).

[107] T. L. Koch, *Optical Fiber Telecommunications IIIB*. Elsevier, 2012 (cit. on p. 30).

[108] P. Massoud Salehi and J. Proakis, *Digital Communications*. McGraw-Hill Education, 2007 (cit. on p. 30).

[109] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems", *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, 2019 (cit. on p. 31).

[110] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943 (cit. on p. 33).

[111] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.", *Psychological review*, vol. 65, no. 6, p. 386, 1958 (cit. on p. 33).

[112] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation", California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985 (cit. on p. 33).

[113] A. Graves, "Supervised sequence labelling", in *Supervised sequence labelling with recurrent neural networks*, Springer, 2012, pp. 5–13 (cit. on pp. 33, 39).

[114] D. S. Broomhead and D. Lowe, "Radial basis functions, multi-variable functional interpolation and adaptive networks", Royal Signals and Radar Establishment Malvern (United Kingdom), Tech. Rep., 1988 (cit. on p. 33).

[115] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, 1982 (cit. on p. 33).

[116] C. M. Bishop *et al.*, *Neural networks for pattern recognition*. Oxford university press, 1995 (cit. on p. 33).

[117] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006 (cit. on pp. 34, 58, 109).

[118] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, 2. MIT press Cambridge, 2016, vol. 1 (cit. on pp. 35, 38, 59).

[119] G. Montúfar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks", in *Proc. Int. Conf. on Neural Information Processing*, 2014, pp. 2924–2932 (cit. on p. 35).

[120] A. R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function", *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 930–945, 1993 (cit. on p. 35).

[121] R. Pascanu, G. Montufar, and Y. Bengio, "On the number of response regions of deep feed forward networks with piece-wise linear activations", *arXiv preprint arXiv:1312.6098*, 2013 (cit. on p. 36).

[122] G. Montúfar, "Notes on the number of linear regions of deep neural networks", *Int. Conf. on Sampling Theory and Applications (SampTA)*, 2017 (cit. on p. 36).

[123] Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization", *arXiv preprint arXiv:1406.2572*, 2014 (cit. on p. 37).

[124] S. Ruder, "An overview of gradient descent optimization algorithms", *arXiv preprint arXiv:1609.04747*, 2016 (cit. on pp. 37, 41, 42).

[125] S. Dash, B. R. Acharya, M. Mittal, A. Abraham, and A. Kelemen, *Deep learning techniques for biomedical and health informatics*. Springer, 2020 (cit. on p. 39).

[126] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, [online]. `http://www.deeplearningbook.org`, accessed: 21-10-2021 (cit. on p. 39).

[127] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks", *Advances in neural information processing systems*, vol. 21, pp. 545–552, 2008 (cit. on p. 39).

[128] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks", in *2013 IEEE international conference on acoustics, speech and signal processing*, Ieee, 2013, pp. 6645–6649 (cit. on pp. 39–41).

[129] H. Zimmermann, R. Grothmann, A. Schaefer, and C. Tietz, "Identification and forecasting of large dynamical systems by dynamical consistent neural networks", *New Directions in Statistical Signal Processing: From Systems to Brain*, pp. 203–242, 2006 (cit. on p. 40).

[130] R. J. Williams and D. Zipser, "Experimental analysis of the real-time recurrent learning algorithm", *Connection science*, vol. 1, no. 1, pp. 87–111, 1989 (cit. on p. 41).

[131] ——, "Gradient-based learning algorithms for recurrent", *Backpropagation: Theory, architectures, and applications*, vol. 433, p. 17, 1995 (cit. on p. 41).

[132] R. Sutton, "Two problems with back propagation and other steepest descent learning procedures for networks", in *Proceedings of the Eighth Annual Conference of the Cognitive Science Society, 1986*, 1986, pp. 823–832 (cit. on p. 41).

[133] N. Qian, "On the momentum term in gradient descent learning algorithms", *Neural networks*, vol. 12, no. 1, pp. 145–151, 1999 (cit. on p. 42).

[134] R. Ward, X. Wu, and L. Bottou, "Adagrad stepsizes: Sharp convergence over nonconvex landscapes", in *International Conference on Machine Learning*, PMLR, 2019, pp. 6677–6686 (cit. on p. 42).

[135] D. Kingma and J Ba, "Adam: A method for stochastic optimization", in *Proc. Int. Conf. on Learning Representations*, 2015 (cit. on pp. 42, 59).

[136] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities", in *2018 European Conference on Optical Communication (ECOC)*, IEEE, 2018, pp. 1–3 (cit. on p. 45).

[137] J Smith, "Odd-bit quadrature amplitude-shift keying", *IEEE Transactions on Communications*, vol. 23, no. 3, pp. 385–389, 1975 (cit. on p. 46).

[138] S. Ruder, "An overview of gradient descent optimization algorithms", *arXiv preprint arXiv:1609.04747*, 2016 (cit. on p. 48).

[139] G. Böcherer, "Labeling non-square QAM constellations for one-dimensional bit-metric decoding", *IEEE Communications Letters*, vol. 18, no. 9, pp. 1515–1518, 2014 (cit. on p. 48).

[140] S. Lin and D. J. Costello, *Error control coding*, 2nd ed. Pearson Prentice Hall, 2004 (cit. on p. 55).

[141] M. Schaedler, F. Pittalà, G. Böcherer, C. Bluemm, M. Kuschnerov, and S. Pachnicke, "Recurrent neural network soft-demapping for nonlinear ISI in 800Gbit/s DWDM coherent optical transmissions", in *Proc. IET. Eur. Conf. Opt. Commun.)*, IEEE, 2020, pp. 1–4 (cit. on p. 56).

[142] M. Schetzen, "Theory of pth-order inverses of nonlinear systems", *IEEE Transactions on Circuits and Systems*, vol. 23, no. 5, pp. 285–291, 1976 (cit. on p. 56).

[143] A. Zaknich, *Principles of adaptive filters and self-learning systems*. Leipzig, Germany: Springer Science & Business Media, 2005 (cit. on p. 59).

[144] C. Xia and W. Rosenkranz, "Nonlinear electrical equalization for different modulation formats with optical filtering", *J. Lightw. Technol.*, vol. 25, no. 4, pp. 996–1001, 2007 (cit. on p. 59).

[145] G. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix", *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, vol. 2, no. 2, pp. 205–224, 1965 (cit. on p. 59).

[146] F. M. Ghannouchi, O. Hammi, and M. Helaoui, *Behavioral modeling and predistortion of wideband wireless transmitters*. John Wiley & Sons, 2015 (cit. on p. 59).

[147] R. Raich and G. T. Zhou, "Orthogonal polynomials for complex Gaussian processes", *IEEE transactions on signal processing*, vol. 52, no. 10, pp. 2788–2797, 2004 (cit. on p. 59).

[148] R. Hecht-Nielsen, "Theory of the backpropagation neural network", in *Neural networks for perception*, Elsevier, 1992, pp. 65–93 (cit. on p. 59).

[149] D. I. Soloway and J. T. Bialasiewicz, "Neural network modeling of nonlinear systems based on Volterra series extension of a linear model", in *Proc. IEEE International Symposium on Intelligent Control (ISIC)*, 1992, pp. 7–12 (cit. on pp. 59, 70, 76).

[150] Z. Jia and L. A. Campos, Eds., *Coherent Optics for Access Networks*. CRC Press, 2020 (cit. on pp. 70, 89).

[151] *Open ROADM MSA 3.01 W-Port Digital Specification (200G-400G)*, [online]. Availabel: `http://openroadm.org/download.html`, 2019 (cit. on pp. 70, 89).

[152] F. Buchali, A. Klekamp, L. Schmalen, and T. Drenski, "Implementation of 64QAM at 42.66 GBaud using 1.5 samples per symbol DAC and demonstration of up to 300 km fiber transmission", in *Proc. Opt. Fiber Commun. Conf.*, Optical Society of America, 2014, M2A–1 (cit. on p. 70).

[153] N. Al-Dhahir and J. M. Cioffi, "Efficiently computed reduced-parameter input-aided MMSE equalizers for ML detection: A unified approach", *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 903–915, 1996 (cit. on pp. 71, 72).

[154] D. D. Falconer and F. Magee Jr, "Adaptive channel memory truncation for maximum likelihood sequence estimation", *Bell System Technical Journal*, vol. 52, no. 9, pp. 1541–1562, 1973 (cit. on p. 72).

[155] S. Hu, *Channel shortening in wireless communication*. Lund University, 2017 (cit. on p. 72).

[156] J. Duprat and J.-M. Muller, "The CORDIC algorithm: New results for fast VLSI implementation", *IEEE Transactions on Computers*, vol. 42, no. 2, pp. 168–178, 1993 (cit. on p. 78).

[157] R. Tibshirani, "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996 (cit. on p. 81).

[158] M. Zhu and S. Gupta, "To prune, or not to prune: Exploring the efficacy of pruning for model compression", *Int. Conf. on Learning Representations*, 2018 (cit. on p. 81).

[159] G. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference", *IEEE Trans. Inf. Theory*, vol. 18, no. 3, pp. 363–378, 1972 (cit. on p. 84).

[160] J. Hagenauer, "A viterbi algorithm with soft-detection outputs and its applications", *Trans. Inf. Theory*, 1989 (cit. on p. 84).

[161] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (corresp.)", *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974 (cit. on p. 84).

[162] T Quach and M Farooq, "Maximum likelihood track formation with the viterbi algorithm", in *Proceedings of 1994 33rd IEEE Conference on Decision and Control*, IEEE, vol. 1, 1994, pp. 271–276 (cit. on p. 84).

[163] T. K. Moon, *Error correction coding: mathematical methods and algorithms*. John Wiley & Sons, 2020 (cit. on pp. 84, 85).

[164] J. P. Burg, "Maximum entropy spectral analysis, paper presented at the 37th annual international meeting, soc. of explor", *Geophys., Oklahoma City, Okla*, 1967 (cit. on p. 85).

[165] T. Wettlin, S. Calabrò, T. Rahman, J. Wei, N. Stojanovic, and S. Pachnicke, "DSP for high-speed short-reach IM/DD systems using PAM", *J. Lightw. Technol.*, vol. 38, no. 24, pp. 6771–6778, 2020 (cit. on p. 85).

[166] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures", *Neural networks*, vol. 18, no. 5-6, pp. 602–610, 2005 (cit. on p. 86).

[167] R. Gencay and T. Liu, "Nonlinear modelling and prediction with feedforward and recurrent networks", *Physica D: Nonlinear Phenomena*, vol. 108, no. 1-2, pp. 119–134, 1997 (cit. on p. 86).

[168] M. Sundermeyer, H. Ney, and R. Schlüter, "From feedforward to recurrent LSTM neural networks for language modeling", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 3, pp. 517–529, 2015 (cit. on p. 86).

[169] G. Kechriotis, E. Zervas, and E. S. Manolakos, "Using recurrent neural networks for adaptive communication channel equalization", *IEEE transactions on Neural Networks*, vol. 5, no. 2, pp. 267–278, 1994 (cit. on p. 87).

[170] A. Salaün, Y. Petetin, and F. Desbouvries, "Comparing the modeling powers of RNN and HMM", in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, IEEE, 2019, pp. 1496–1499 (cit. on p. 87).

[171] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997 (cit. on p. 87).

[172] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks", *Trans. Signal Process*, vol. 45, no. 11, pp. 2673–2681, 1997 (cit. on p. 87).

[173] F. Pittala *et al.*, "800ZR+ DWDM demonstration over 600km G. 654D fiber enabled by adaptive nonlinear TripleX equalization", in *Proc. Opt. Fiber Commun. Conf.*, 2020, M4K–5 (cit. on p. 91).

[174] Y. Kawaguchi, Y. Tamura, T. Haruna, Y. Yamamoto, and M. Hirano, "Ultra low-loss pure silica core fiber", *SEI Technical review*, vol. 80, pp. 50–55, 2015 (cit. on p. 91).

[175] P. Poggiolini *et al.*, "Analytical and experimental results on system maximum reach increase through symbol rate optimization", *Journal of Lightwave Technology*, vol. 34, no. 8, pp. 1872–1885, 2016 (cit. on p. 96).

[176] S. Yu, L. S. Giraldo, and J. Principe, "Information-theoretic methods in deep neural networks: Recent advances and emerging opportunities", in *International Joint Conference on Artificial Intelligence*, 2021, pp. 4669–4678 (cit. on p. 106).

[177] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, "Deep variational information bottleneck", *arXiv preprint arXiv:1612.00410*, 2016 (cit. on p. 106).

[178] A. Xu and M. Raginsky, "Information-theoretic analysis of generalization capability of learning algorithms", *arXiv preprint arXiv:1705.07809*, 2017 (cit. on p. 106).

[179] G. Böcherer, *Lecture notes on machine learning for communications*, [online]. Availabel: `http://georg-boecherer.de/mlcomm.pdf`, accessed: 17-01-2022 (cit. on pp. 107, 108).

[180] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012 (cit. on p. 109).

[181] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. Sevastopol, CA, USA: O'Reilly Media, 2019 (cit. on p. 109).

[182] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., 2006 (cit. on p. 111).