Equalization in Optical Fiber Communication Using Model-based Neural Networks

Mohanad Abu-Romoh

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2023IPPAT037

Thèse de doctorat

Equalization in Optical Fiber Communication Using Model-based Neural Networks

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom Paris

École doctorale n°626 de l'Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Réseaux, informations et communications

Thèse présentée et soutenue à Palaiseau, le 15/11/2023, par

MOHANAD ABU-ROMOH

Composition du Jury :

Cédric WARE Professeur, Télécom Paris, Palaiseau (LTCI)	Président / Examinateur
Christophe PEUCHERET Professeur, ENSSAT, Lannion (FOTON)	Rapporteur
Stéphane AZOU Professeur, ENIB, Brest (Lab-STICC)	Rapporteur
Vahid MEGHDADI Professeur, Université de Limoges, Limoges (XLIM laboratory)	Examinateur
Mansoor YOUSEFI Maitre de Conférence, Télécom Paris, Palaiseau (LTCI)	Directeur de thèse
Yves JAOUËN Professeur, Télécom Paris, Palaiseau (LTCI)	Co-directeur de thèse
Bernhard SPINNLER Docteur en ingénierie, Infinera GmbH, Allemagne	Invité

To my guiding lights,
Forever grateful.

Acknowledgements

Firstly, I want to express my deepest appreciation to my thesis director and main supervisor, Prof. Mansoor Yousefi, for his invaluable guidance and continuous support. His knowledge, wisdom, and patience played a crucial role in shaping this thesis and my overall growth as a researcher. I am also grateful to my co-supervisor, Yves Jaöuen, for his valuable insights, helpful feedback, and support throughout my research. His expertise in the field expanded my understanding and improved the quality of this thesis.

I want to extend my deepest appreciation to my industrial co-supervisors from Infinera, namely Antonio Napoli, Bernhard Spinnler, and Nelson Costa. Their guidance, industry perspective, and collaborative efforts were instrumental in connecting academia and industry. I am thankful for their constant support, insightful discussions, and provision of valuable resources that enhanced my research.

Additionally, I am deeply gratefully to my parents for their unwavering love, support, and belief in my abilities. Their sacrifices, encouragement, and understanding have been my driving force throughout this journey. I will always be indebted to them for their unwavering faith in me.

I would like to express my gratitude for the support received from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 813144. This funding has been invaluable in facilitating my research endeavors and advancing the field of optical communication.

Lastly, I want to express my gratitude to all the teachers, mentors, and colleagues who contributed to my academic and personal growth. Your guidance, knowledge sharing, and inspiring discussions shaped my research and helped me become a better researcher.

To everyone mentioned above and all others who supported me in various ways, I am sincerely grateful. Your belief in me, encouragement, and assistance played a vital role in my achievements. This thesis is a testament to our collective efforts, and I feel honored to have had such remarkable individuals in my life.

Contents

Acronyms

ADC analog-to-digital converter

AIR achievable information rate

ASE amplified spontaneous emission

ASK amplitude-shift keying

AWGN additive white Gaussian noise

BER bit-error-ratio

Bi-LSTM bi-directional long short-term memory

BPSK binary phase-shift keying

CD chromatic dispersion

CDC chromatic dispersion compensation

CFO carrier frequency offset

CMA constant modulus algorithm

CNLSE coupled nonlinear Schrödinger's equation

CNN convolutional neural network

CPE constant phase estimation

DBP digital back-propagation

DCF dispersion-compensating fiber

DD direct-detection

DGD differential group delay

DM dispersion-managed

DNN dense neural network

DSF dispersion-shifted fiber

DSP digital signal processing

DWDM dense wavelength-division multiplexing

EDBP enhanced digital back-propagation

EDFA erbium-doped fiber amplifier

EVM error vector magnitude

FD frequency-domain

FDBP filtered digital back-propagation

FFT fast Fourier transform

FIR finite frequency response

FSK frequency-shift Keying

FWM four wave mixing

GDBP generalized digital back-propagation

GVD group velocity dispersion

IFFT inverse fast Fourier transform

IM intensity modulation

ISI inter-symbol interference

LDBP learned digital back-propagation

LE linear equalization

LpS layers per span

LSTM long short-term memory

MIMO multiple input multiple output

MSE mean squared error

MZM Mach-Zehnder modulator

NDM non-dispersion-managed

NLSE nonlinear Schrödinger equation

NN neural network

NRZ non-return-to-zero

NZ-DSF non-zero dispersion-shifted fiber

OOK on-off keying

OSNR optical signal-to-noise ratio

PDM polarization-division multiplexing

PMD polarization mode dispersion

PN phase noise

PSD power spectral density

PSK phase-shift keying

QAM quadrature amplitude modulation

QPSK quadrature phase-shift keying

RDE radially directed equalizer

RM real multiplication

RMpS real multiplications per symbol

RNN recurrent neural network

RRC root-raised-cosine

SER symbol-error-ratio

SGD stochastic gradient descent

SMF single-mode fiber

SNR signal-to-noise ratio

$\mathbf{SNR_{eff}}$ effective signal-to-noise ratio

SOP state of polarization

SPM self-phase modulation

SSFM split-step Fourier method

StpS steps per span

TD time-domain

WDM wavelength-division multiplexing

XPM cross-phase modulation

List of Symbols

Symbol	Meaning
n_1	Refractive index of the core
n_c	Refractive index of the cladding
P_0	Launched optical signal power [W]
α	Attenuation coefficient
λ	Wavelength
f	frequency
ω	Angular frequency
c	Speed of light
$n_g(\omega)$	Frequency dependant group index
v_g	Group velocity
β	Mode-propagation constant
β_1	First-order dispersion parameter
β_2	Group velocity dispersion parameter
β_3	Third-order dispersion parameter
L_{fiber}	Fiber length
T_{CD}	Chromatic dispersion-induced pulse spreading
D	Dispersion coefficient
B_m	Strength of modal birefringence
A_{eff}	Effective core area
q_x, q_y, $\mathbf{q}$	Complex optical field envelope for x-polarization, y-polarization, and vectorial form, respectively
L_{eff}	Effective length
L_{span}	Span length
Φ_{NL}	SPM-induced nonlinear phase

Symbol	Meaning
Δf_{NL}	frequency nonlinear shift
h	Planck's constant
n_{sp}	Spontaneous emission factor
G	Amplification gain
S_{ASE}	Power spectral density of the ASE noise
L_B	Beat length
N_{seg}	Number of fiber segments for SSFM
δ_s	SSFM step size
δ_d	DBP step size
τ_i	DGD parameter for the i-th segment
E_b	Energy per bit
E_s	Energy per symbol
N_0	Noise spectral power
$\phi_i(t)$	Orthonormal basis function
$\vec{s}$	Signal vector
M	Modulation order
$\mathbf{x}$	Input vector
$\mathbf{w}$	Weight vector
b	Bias term
y , $\mathbf{y}$	Target output (scalar), Target output vector
$\hat{y}$, $\hat{\mathbf{y}}$	Predicted output (scalar), Predicted output vector
n	Number of examples
$\mathcal{D}$	Input-output dataset
m	Batch size
$\mathcal{B}_i$	i-th mini-batch
T_{max}	Maximum number of epochs
$\mathrm{sign}(\cdot)$	Sign function
$\mathrm{ReLU}(\cdot)$	ReLU function
$\tanh(\cdot)$	hyperbolic tangent function
$\sigma_l(\cdot)$	Activation function of the l-th layer
η	Learning rate

Symbol	Meaning
$L(\cdot)$	Loss function
$E(\cdot)$	Expected value
$\mathcal{X}$	Input space
$\mathcal{Y}$	Output space
∇L	Loss gradient
$\mathcal{H}$	Decision procedure
N_i,N_{h_k},N_o	Input layer size, k-th hidden layer size, and output layer size, respectively.
Ψ_l	Mapping from the $(l-1)$-th layer to the l-th layer.
$\mathbf{W}_l$	Weight matrix of the l-th layer
$\mathbf{b}_l$	bias vector of the l-th layer
x_t	RNN input at time step t
y_t	RNN output at time step t
h_t	RNN hidden state at time step t
i_t, f_t, c'_t, o_t, c_t	Input state, forget state, candidate cell state, output state, and cell state at time step t
$\mathbf{s}_x$, $\mathbf{s}_y$	Symbols transmitted from x and y-polarizations
$\mathbf{b}_x$, $\mathbf{b}_y$	Bits transmitted from x and y-polarizations
$p(\cdot)$	Pulse shape
T_s	Symbol duration time
N_s	Number of transmitted symbols
$\mathbf{F}$, $\mathbf{F}^{-1}$	DFT and IDFT matrices
$\mathrm{SNR}_{\mathrm{eff}}$	Effective Signal-to-Noise Ratio
EVM	Error Vector Magnitude
BER	Bit-Error Ratio
Q_{fac}	Q-factor based on BER

Abstract

To meet the growing demand for achievable information rates (AIR), the optical fiber network—the backbone of global communication infrastructure—must continually evolve to support faster data transmission speeds. However, this effort is hindered by distortions affecting the optical signals. Both linear and nonlinear effects, as well as noise, impact the signals propagating through optical fibers. These distortions critically compromise signal quality, especially at higher transmission rates. Linear effects, which broaden the pulse width, become more pronounced at higher transmission rates and can be effectively corrected through optical means, such as dispersion compensating fiber (DCF), or electronically, using digital signal processing (DSP) algorithms at the receiver. While DCFs were common in earlier network generations, they have been largely replaced by highly optimized DSP chains, now prevalent in commercial applications, capable of coherent equalization. The Kerr effect specifically emerges from the nonlinear optical behavior of fiber material, leading to changes in refractive index proportional to the light intensity. As signal power increases, this nonlinearity's impact intensifies, resulting in various impairments in optical fiber networks, including four-wave mixing, self-phase modulation, and cross-phase modulation.

The two primary limiters of AIR in optical fiber networks are nonlinear channel effects, originating from Kerr nonlinearity, and amplified spontaneous emission (ASE) noise. As the demand for AIR increases, these nonlinear effects pose a serious challenge, diminishing network efficiency. Equalizing these effects is challenging, requiring processing capabilities beyond those of traditional commercial systems. Digital backpropagation (DBP) is a theoretical method that can mitigate these issues by sequentially correcting linear and nonlinear effects across small segments, addressing each effect individually. Despite its promise, the computational demands of DBP limit its practical application in commercial settings.

Recent advances in machine learning have spurred numerous efforts to apply these techniques for nonlinearity mitigation in optical networks. Neural networks, in particular, have gained prominence due to this ability to learn from examples, making them well-suited for addressing nonlinear distortions. They can learn the channel model based on

pairs of transmitted and received signals, allowing them to approximate the inverse channel transfer function. Additionally, neural networks could offer a simpler alternative to traditional nonlinearity mitigation methods like DBP, even without precise knowledge of channel parameters.

Within the realm of neural network-based optical signal equalizers, two primary design strategies exist: model-agnostic and model-based approaches. Model-agnostic neural networks (NNs) are valued for their design flexibility, where the selection of hyperparameters is determined without specific assumptions about the data model. This process often involves using optimization techniques, like Bayesian optimization, to find the most effective configurations. Although model-agnostic approaches are lauded for their flexibility and adaptability, enabling their integration at various points within the receiver's DSP, they demand large models and extensive datasets to achieve optimal performance. In contrast, model-based approaches, grounded in existing models rather than arbitrary design, utilize the physical propagation model for structuring and initializing neural networks. This leads to more efficient designs. However, these approaches often require detailed knowledge of the system's parameters, which may not always be available or easy to estimate. They can also be less flexible in adapting to new data or scenarios that deviate from the modeled conditions, potentially limiting their applicability across different or evolving environments.

In this thesis, we specifically focus on the model-based neural network approach, with emphasis on a prominent model known as Learned Digital Backpropagation (LDBP). LDBP is a neural network method that adopts DBP as the foundational model for the neural network structure, effectively optimizing DBP parameters through neural networks. Notably, model-based methods have shown proficiency in many-to-many equalization, as opposed to the common one-to-one or many-to-one equalization observed in model-agnostic or black-box neural network equalizers. The ability to perform many-to-many equalization can potentially simplify complexity and is well-suited for real-time implementation in optical receivers. We introduce simplifications through "parameter sharing" to reduce trainable parameters. The parameter sharing method in neural networks involves reusing the same weights across multiple layers. This approach significantly reduces the total number of unique trainable parameters, leading to a more efficient and compact model that requires less computational resources to train and can generalize better to new data.

Moreover, we propose repurposing legacy Dispersion-Managed (DM) systems, which typically use basic modulation formats such as intensity modulation and direct detection (IM/DD), by integrating higher-order modulation formats like 16-QAM and 64-QAM. This work aims to boost data rates within these systems. Initially, we propose a flexible DBP implementation tailored to DM systems, featuring a fractional number of steps per span. Subsequently, LDBP is utilized at the receiver to equalize channel impairments in multi-channel, dual-polarization transmission systems by parameterizing the proposed

DBP and optimizing its parameters using stochastic gradient descent. Our comprehensive performance and complexity analysis shows that the proposed algorithms surpass both linear equalization and DBP across various transmission setups.

Résumé

Pour répondre à la demande croissante en débits d'information réalisables (AIR), le réseau de fibres optiques, épine dorsale de l'infrastructure de communication mondiale, doit évoluer continuellement pour supporter des vitesses de transmission de données plus rapides. Cet effort est entravé par des distorsions affectant les signaux optiques, compromettant gravement la qualité du signal, surtout à des taux de transmission élevés. Les effets linéaires et non linéaires, ainsi que le bruit, impactent les signaux se propageant dans les fibres optiques. Les effets linéaires, élargissant la largeur d'impulsion, deviennent plus prononcés à des taux de transmission plus élevés et peuvent être corrigés efficacement par des moyens optiques, tels que la fibre à compensation de dispersion (DCF), ou électroniquement, en utilisant des algorithmes de traitement du signal numérique (DSP) au récepteur. Tandis que les DCF étaient courants dans les générations précédentes de réseaux, ils ont été largement remplacés par des chaînes DSP hautement optimisées, désormais prévalentes dans les applications commerciales, capables d'une égalisation cohérente. L'effet Kerr, émergeant spécifiquement du comportement optique non linéaire du matériel de fibre, conduit à des changements de l'indice de réfraction proportionnels à l'intensité de la lumière. Avec l'augmentation de la puissance du signal, l'impact de cette non-linéarité s'intensifie, résultant en diverses altérations dans les réseaux de fibres optiques, incluant le mélange à quatre ondes, la modulation de phase auto-induite et croisée.

Les deux principaux limitateurs de AIR dans les réseaux de fibres optiques sont les effets de canal non linéaires, provenant de la non-linéarité de Kerr, et le bruit d'émission spontanée amplifiée (ASE). Avec l'augmentation de la demande pour AIR, ces effets non linéaires posent un défi sérieux, diminuant l'efficacité du réseau. Égaliser ces effets est difficile, nécessitant des capacités de traitement au-delà de celles des systèmes commerciaux traditionnels. La rétropropagation numérique (DBP) est une méthode théorique qui peut atténuer ces problèmes en corrigeant séquentiellement les effets linéaires et non linéaires à travers de petits segments, abordant chaque effet individuellement. Malgré sa promesse, les exigences computationnelles de DBP limitent son application pratique dans les contextes commerciaux.

Les avancées récentes dans l'apprentissage automatique ont incité de nombreux efforts

pour appliquer ces techniques à l'atténuation de la non-linéarité dans les réseaux optiques. Les réseaux de neurones, en particulier, ont gagné en importance grâce à leur capacité à apprendre à partir d'exemples, les rendant bien adaptés pour aborder les distorsions non linéaires. Ils peuvent apprendre le modèle de canal basé sur des paires de signaux transmis et reçus, leur permettant d'approximer la fonction de transfert inverse du canal. De plus, les réseaux de neurones pourraient offrir une alternative plus simple aux méthodes traditionnelles d'atténuation de la non-linéarité comme DBP, même sans connaissance précise des paramètres du canal.

Dans le domaine des égaliseurs de signal optique basés sur les réseaux de neurones, il existe deux stratégies de conception principales : les approches agnostiques au modèle et les approches basées sur le modèle. Les réseaux de neurones agnostiques au modèle sont appréciés pour leur flexibilité de conception, où la sélection des hyperparamètres se fait sans hypothèses spécifiques sur le modèle de données. Ce processus implique souvent l'utilisation de techniques d'optimisation, comme l'optimisation bayésienne, pour trouver les configurations les plus efficaces. Bien que les approches agnostiques au modèle soient louées pour leur flexibilité et adaptabilité, permettant leur intégration à différents points dans le DSP du récepteur, elles nécessitent de grands modèles et d'importants ensembles de données pour atteindre une performance optimale. En contraste, les approches basées sur le modèle, qui s'appuient sur des modèles existants plutôt que sur une conception arbitraire, utilisent le modèle de propagation physique pour structurer et initialiser les réseaux de neurones, menant à des conceptions plus efficaces. Cependant, ces approches requièrent souvent une connaissance détaillée des paramètres du système, qui peuvent ne pas toujours être disponibles ou faciles à estimer. Elles peuvent également être moins flexibles pour s'adapter à de nouvelles données ou à des scénarios qui s'écartent des conditions modélisées, limitant potentiellement leur applicabilité à travers différents environnements ou en évolution.

Dans cette thèse, nous nous concentrons spécifiquement sur l'approche des réseaux de neurones basée sur le modèle, en mettant l'accent sur un modèle important connu sous le nom de Rétropropagation Numérique Apprise (LDBP). LDBP est une méthode de réseau de neurones qui adopte le DBP comme modèle fondamental pour la structure du réseau de neurones, optimisant ainsi efficacement les paramètres du DBP à travers les réseaux de neurones. De plus, nous proposons de réaffecter les systèmes Gérés par Dispersion (DM) existants, utilisant typiquement des formats de modulation de base, en intégrant des formats de modulation d'ordre supérieur comme le 16-QAM et le 64-QAM. Ce travail vise à augmenter les débits de données au sein de ces systèmes. Initialement, nous proposons une mise en œuvre flexible du DBP adaptée aux systèmes DM. Ensuite, le LDBP est utilisé au récepteur pour égaliser les altérations de canal dans les systèmes de transmission multi-canaux à double polarisation, en paramétrant le DBP proposé et en optimisant ses

paramètres à l'aide de la descente de gradient stochastique. Nos analyses complètes de performance et de complexité montrent que les algorithmes proposés surpassent à la fois l'égalisation linéaire et le DBP dans diverses configurations de transmission.

CHAPTER 1

Introduction

In today's digital era, the vast majority of the world's data travels through optical fiber networks. The adoption of optical fibers, in place of their copper predecessor, has revolutionized the telecommunications industry, enabling the seamless transmission of vast amounts of data with unparalleled speed, reliability, and security. The inherent advantages of optical fibers have propelled them to become the backbone of global communication networks, supporting the increasing demands of our digital world. In fact, more than 99% of global Internet traffic now passes through these optical fiber networks, placing them under constant scrutiny as they strive to meet the ever-increasing demand. Moreover, the relentless advancements in technology and the advent of emerging applications like Internet of Things, artificial intelligence, and virtual reality further amplify the demand for faster and more robust data transmission. The seamless integration of these transformative technologies relies on the steadfast performance of optical fiber networks to support their data-intensive requirements.

Given the paramount significance of optical fiber networks in facilitating the transmission of the world's data, the burden falls upon these networks to rise to the occasion. To meet the ever-growing demands placed upon them, optical fiber networks must continually evolve and adapt. This involves deploying cutting-edge technologies, upgrading infrastructure, and optimizing network architectures. Additionally, ongoing research and development efforts are essential to explore innovative solutions that enhance the capacity, efficiency, and resilience of optical fiber networks.

The two primary limiters of the achievable information rate (AIR) in optical fiber networks are the nonlinear channel effects originating from the Kerr nonlinearity and the amplified spontaneous emission (ASE) noise. While the linear effects also impact the signal transmission through fiber, they can be corrected through simple linear equalization, in contrast to their nonlinear counterpart. The Kerr effect in optical fiber arises from the nonlinear optical response of the material, resulting in a change in the refractive index with the intensity of the light. As the power of transmitted signal is increased, the impact of

the fiber nonlinearity becomes more significant. The Kerr nonlinearity gives rise to several impairments in optical fiber networks, such as the four-wave mixing, self- and cross-phase modulation. The ASE noise, on the other hand, is a fundamental quantum effect resulting from spontaneous photon emission within amplifiers, and not reversible. As the demand for higher data rates increases, network operators require adopting higher-order modulation formats. This requires increasing transmission power, pushing communications further into the nonlinear regime. However, achieving spectrally-efficient transmission in the nonlinear regime is challenging.

The digital back-propagation (DBP) is an effective equalizer to mitigate the channel effects in optical fiber [90]. However, due to its high computational requirement, DBP is impractical for real-time operation [46]. Extensive research has been dedicated to reducing the computational burden associated with DBP, resulting in significant advancements such as the proposal of the folded-DBP [34, 69, 110, 111], filtered-DBP[27, 87], enhanced-DBP [91, 92], and generalized DBP [30]. With the rapid progress in the field of machine learning, the attention has recently focused on leveraging the advances in neural networks for the nonlinear equalization. These emerging approaches harness the capabilities of neural networks to learn and model the intricate nonlinear patterns present in the received signals. Recent studies have showcased promising outcomes, achieving comparable performance to the traditional methods such as DBP with reduced computational complexity [2, 12, 48, 49, 55].

The integration of the neural networks into fiber-optic communication systems is a noteworthy advancement, as it presents a promising alternative to the nonlinearity mitigation techniques that are often computationally demanding. By capitalizing on the learning capabilities of the neural networks, researchers aim to overcome the limitations of the traditional approaches and enhance the performance and reliability of the fiber-optic communication systems.

The main focus of this thesis is to investigate the potential of the model-based neural network equalizers, primarily the learned digital back-propagation (LDBP), for low-complexity mitigation of the linear and nonlinear effects in dual-polarization optical fiber transmission. The performance and complexity trade-off of the different equalizers is studied, applied to both dispersion-uncompensated links, as well the legacy dispersion-managed (DM) systems with inline optical dispersion compensation. Below, the contributions of this PhD work are described for each chapter.

Chapter 2 provides an overview of the optical fiber transmission systems relevant to this thesis. The chapter starts by describing the physical structure and characteristics of optical fiber. This sets the stage for the subsequent sections, where we explore the linear effects, such as the chromatic dispersion, and the nonlinear effects, such as self- and cross-phase modulation, that impact the propagation of optical signals in both single-

polarization and dual-polarization transmission. Proceeding further, we delve into the mathematical model of the propagating signal, characterized by the nonlinear Schrödinger equations, along with the vectorial form of the Schrödinger's equation averaged over fast-varying polarization mode dispersion (PMD). Continuing our exploration, we investigate the optical transmitter, starting with the advanced modulation formats, followed by an analysis of the Mach-Zehnder modulator (MZM), which converts the electrical to optical signals. We subsequently delve into the techniques for generating optical signals with higher-order modulations through the use of the MZM. The chapter then introduces the receiver, highlighting the equalization of the adverse channel effects. We will then describe in detail the linear DSP chain, comprising the CD compensation, MIMO equalization, carrier frequency offset estimation, and constant phase estimation. Finally, we present the LDBP for equalizing the linear and nonlinear effects within the channel.

In Chapter 3, we delve into the fundamentals of neural networks, which constitute an important part of this research. Our exploration begins with the mathematical model of the perceptron, which serves as the fundamental building block for constructing neural networks. We outline the mathematical model underlying the neural networks, introduce various architectures including the fully-connected, convolutional, and long short-term memory (LSTM) neural networks, and provide an exposition of the training algorithms. Furthermore, we provide literature examples on the neural network-based approaches to mitigating the effects of optical fiber channels. Through Chapters 2 and 3, we establish the essential background necessary for comprehending the subsequent chapters of the thesis.

In Chapter 4 , we compare the performance and complexity of LDBP and LDBP. We will highlight the problem with the training complexity of LDBP in long-haul transmission, and introduce a parameter sharing approach in order to simplify the training of LDBP, applied to the uncompensated fiber transmission links.

Chapter 5 represents the core contribution of this thesis, where we study the potential of LDBP in enhancing DM systems with coherent transmission, thereby doubling the achievable data rates. LDBP is a model based neural network that shares the same computational graph as the LDBP, except that, in the case of LDBP, the filter taps are treated as "free weights" that can be learned. By doing so, the model can be trained to capture the complex interactions in the signal propagating in the channel, while maintaining the same complexity as the LDBP. Despite their outdated nature, DM systems continue to be utilized in certain segments of optical fiber networks, primarily for carrying information at bitrates in the tens of Gbits/s range, using basic on-off keying (OOK) modulation. Our research demonstrates a substantial increase in achievable bitrates to the order of several hundreds of Gbits/s. This is made feasible by leveraging modern techniques like polarization-division multiplexing and higher order M-quadrature amplitude modulation (QAM) modulations, effectively harnessed through the implementation of the LDBP re-

ceiver.

Finally, Chapter 6 concludes the thesis and provides insights into potential areas for future research.

CHAPTER 2

Fiber-optic communication systems

An optical communication system has three main components: a transmitter, a receiver, and a transmission channel. The goal of fiber-optic communication systems is to carry the information reliably from the transmitter to the receiver through the optical fiber channel.

Contents

2.1 Introduction

Optical communication systems have revolutionized the way information is transmitted over long distances, offering high-speed and reliable data transfer. These systems comprise

three main components: the transmitter, the receiver, and the optical fiber transmission channel. The primary objective of fiber-optic communication systems is to ensure the seamless and efficient delivery of information from the transmitter to the receiver through the optical fiber channel.

Unlike the transmitter and receiver, we have limited control over the optical fiber components. While the fabrication of optical fibers can be improved, there are theoretical limits for the fiber parameters. For example, the lowest-ever transmission losses were achieved in 2018 by researchers at Sumitomo's Optical Communications Laboratory in Japan, with reported attenuation coefficients of 0.1419 dB/km at 1560 nm wavelength and 0.1424 dB/km at 1550 nm in a Ge-free silica-core optical fiber [99]. Prior to this, progress in reducing attenuation coefficients was relatively slow with values around 0.162 dB/km in 2010 [82], 0.154 dB/km in 2013 [53], and 0.152 dB/km in 2016-17 [104]. Even with the accelerated technological advancements, the attenuation of silica-glass cannot be brought to zero, since it is approaching its theoretical limit. The Chromatic Dispersion of a fiber is another negative impacting effect that can be changed by acting on the physical properties of the material. Various fiber types were developed to reduce dispersion, including dispersion-shifted fiber (DSF) with a dispersion parameter $D < 1$ ps/(nm.km) at a wavelength $\lambda = 1550$ nm, and non-zero dispersion-shifted fiber (NZ-DSF) with a D value ranging from 2 to 4 ps/(nm.km). The most commonly used fiber in networks is referred to as 'dispersion-unshifted' single-mode fiber (SMF). SMF exhibits low chromatic dispersion in the optical window around λ =1310 nm but experiences higher chromatic dispersion (CD) in the 1550 nm region, limiting transmission length without appropriate compensation. SMF is designed to minimize chromatic dispersion in the 1550 nm window, achieving $D \approx$17 ps/(nm.km), but it has zero dispersion somewhwer between 1525 nm and 1575 nm [19]. However, DSF has drawbacks, such as significant four wave mixing (FWM) generation, rendering communication systems incorporating dense wavelength-division multiplexing (DWDM) practically impossible. Additionally, randomly varying effects like polarization mode dispersion (PMD) and state of polarization (SOP) further pose challenges to communication systems. Overcoming these fiber-related challenges requires smart design and optimization of the transmitter and receiver.

This chapter aims to explore the effects of the fiber channel that influence the quality of optical signal transmission through optical fiber. We categorize these effects into three main groups: attenuation, linear effects, and nonlinear effects. We then explore the design aspects of optical fiber transmitters, as well as the modulation formats employed for optical fiber transmission. Furthermore, we examine the essential components of optical fiber receivers, focusing on the various DSP blocks that constitute modern coherent receivers widely utilized in contemporary communication systems. By covering these topics, we gain a comprehensive understanding of the key elements involved in achieving reliable and

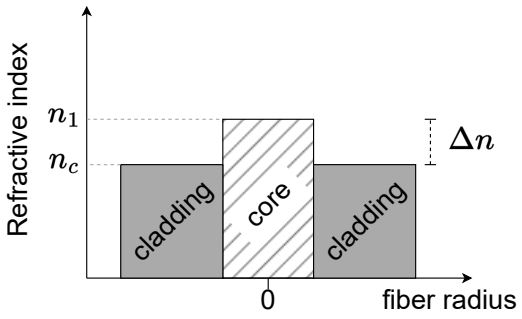


Figure 2.1: Refractive index profile.

high-performance optical fiber communication.

2.2 Optical Fiber Channel

Optical fibers are narrow strands made of silica glass which carry the optical signal from one end to the other. Step-index fibers consist of a core made of glass with a refractive index n_1 that is higher than the refractive index of the surrounding cladding layer n_c, as shown in Figure 2.1. This configuration traps light propagating inside the fiber by continuously bouncing the light against the boundary separating the core and cladding [6]. The refractive index difference can be achieved by either increasing the refractive index for the core silica through germanium-oxide (GeO_2) doping, or by decreasing the refractive index of the cladding through fluoride (F) doping. Additionally, the polymer coating is sometimes considered as a third layer. However, this layer does not interfere with the signal propagation; its primary purpose is to provide protection for the glass fiber against external abrasion, environmental factors, micro-bending loss, and stress corrosion or fatigue

Within these optical fibers, the light can travel according to one or several modes. Here, the term 'mode' denotes the unique routes or patterns that the light follows as it propagates within the fiber. These modes correspond to varied spatial arrangements of the light waves as they reflect and traverse along the fiber. The number of modes that a fiber can accommodate is dependent on several parameters, including the size of the core, the refractive index distribution, and the operational wavelength. This multiplicity of modes leads to the classification of optical fibers into two fundamental categories: single mode fibers and multi-mode fibers. Single mode fibers permit the propagation of a single signal mode, whereas multi-mode fibers supports multiple orthogonal modes. The capability of a fiber to support single or multi-mode propagation at a given wavelength primarily hinges on the difference in refractive index between the core and the cladding, denoted as $\Delta n = n_1 - n_c$, in conjunction with the diameter of the core. Conversely, a smaller refractive index disparity and a smaller core diameter result in a fiber that can support

fewer modes.

The principal impairments that occur during long-haul transmissions can be broadly categorized into two groups: linear and nonlinear effects. Linear transmission effects encompass phenomena such as chromatic dispersion that causes pulse spreading, attenuation leading to signal power loss, and polarization mode dispersion. On the other hand, nonlinear transmission effects arise due to interaction between the optical signal and fiber's nonlinear characteristics resulting in signal distortion and degradation. These nonlinear effects include phenomena such as self and cross-phase modulation, four-wave mixing, cross-polarization modulation, stimulated Raman and Brillouin scattering [83]. These impairments can greatly affect the quality of the transmission and must be carefully considered in the design and operation of long-haul optical networks. In the subsequent sections, we delve into the various impairments that affect long-haul transmissions.

2.2.1 Linear Transmission Effects

2.2.1.1 Attenuation

The power of the signal propagating through the fiber decreases exponentially with respect to z, which follows the equation:

$$P(z) = P_0 \cdot \exp(-\alpha z), \tag{2.1}$$

where P_0 is the launched power measure in [W], z is the distance in [km] and α is the attenuation coefficient measured in [/km]. The attenuation of optical fibers is often quantified in [dB/km] units, which is calculated using the following formula:

$$\alpha[\text{dB/km}] = 10 \log_{10}\big(\exp(\alpha[/\text{km}])\big) \approx 4.343 \cdot \alpha[/\text{km}]. \tag{2.2}$$

Expressing the attenuation in [dB/km] simplifies the calculation of loss, as the signal loses α [dB] of its power for each kilometer it travels. Typically, single mode fiber has a minimum loss around 0.2 dB/km for fibers operating in the at 1550nm.

Several factors contribute to fiber loss, but the primary causes are material absorption and Rayleigh scattering. Silica glass, which is commonly used in fiber optics, has electronic resonances in the ultraviolet and vibrational resonances in the far-infrared, but it does not absorb much light in the 0.5 to 2 μm wavelength range [6]. Despite the numerous attempts to reduce fiber losses since the invention of glass fiber as a means of data transmission, the ultimate limit on how low fiber attenuation can get is set by the intrinsic Rayleigh scattering. This loss, which varies with the wavelength as λ^{-4} and is most prominent at short wavelengths, is intrinsic to silica fiber and cannot be completely eliminated.

Doping the fiber also contributes to increased losses, however to a lesser extent. While both fluoride and Germanium-Oxide dopants can result in higher losses, using fluoride

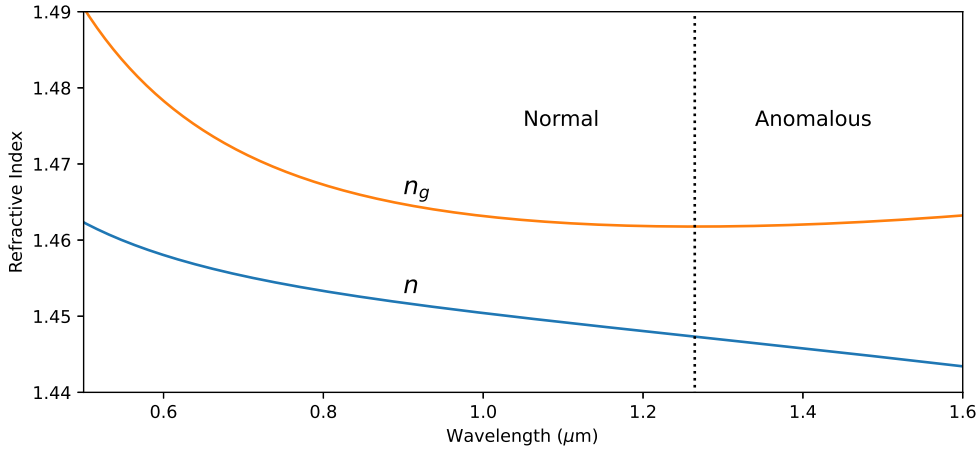


Figure 2.2: Wavelength-dependent refractive index n and group refractive index n_g for fused silica.

to dope the cladding rather than Germanium-Oxide in the core can lead to better attenuation characteristics as the majority of the signal's energy is propagated through the core. Additionally, other factors such as bending of the fiber and scattering of light at the core-cladding interface also impact fiber losses.

The current lowest-ever achieved transmission losses in a *Ge*-free silica-core optical fiber is reported in [99], where the authors used fluorine co-doping and a two-layered polymer coating.

2.2.1.2 Chromatic Dispersion

The refractive index of the transparent optical medium through which the optical signal propagates depends on the optical frequency, denoted as f. The propagation velocity of the signal is determined by the refractive index of the medium, causing different frequency components to travel at distinct velocities, represented as $c/n_g(\omega)$, where $\omega = 2\pi f$ is the angular frequency in $[rad/s]$, c is the speed of light in vacuum, measuring 2.9979×10^8m/s, and $n_g(\omega)$ is the frequency dependant group index, defined as:

$$n_g(\omega) = \frac{\partial}{\partial\omega}\big(\omega \cdot n(\omega)\big) = n(\omega) + \omega\frac{dn}{d\omega}. \tag{2.3}$$

The variation of the refractive index and group index for silica glass, in relation to the wavelength, are illustrated in Figure 2.2. This dependency of the refractive index on the angular frequency, also known as *material dispersion*, is an intrinsic characteristic of silica glass, which is a fundamental component of optical fibers. The group velocity v_g, corresponding to any spectral component of the wave propagating through the optical fiber, is given by:

$$v_g(\omega) = \frac{c}{n(\omega) + \omega\frac{dn}{d\omega}}. \tag{2.4}$$

Due to the varying propagation speeds of spectral components, a phase shift occurs among the different spectral components of a signal as it travels through the fiber, resulting in the phenomenon known as CD. The chromatic dispersion leads to a broadening of the pulse in the time domain, as illustrated in Figure 2.3. This dispersion causes interference between symbols and consequently impairs the propagation of short optical pulses, even in the linear regime where nonlinear effects are less pronounced.

When defining the wavelength-dependent refractive index of a transparent optical material, the Sellmeier equation [1] is commonly used:

$$n(\lambda) \approx \sqrt{1 + \sum_{j=1}^{m} \frac{A_j \lambda^2}{\lambda^2 - \lambda_j^2}}, \tag{2.5}$$

where λ_j is the wavelength corresponding to the resonance frequency ω_j, and A_j is a parameter determined experimentally by measuring the dispersion at various wavelengths and fitting the measurements to Eq. 2.5. For instance, when calculating the refractive index value for fused silica, as documented by Malitson [71], the equation becomes:

$$n(\lambda) \approx \sqrt{1 + \frac{0.6961663\lambda^2}{\lambda^2 - 0.0684043} + \frac{0.4079426\lambda^2}{\lambda^2 - 0.1162414} + \frac{0.8974794\lambda^2}{\lambda^2 - 9.896161}}, \tag{2.6}$$

which provides an accurate measurement for wavelengths ranging from 0.21 μm to 3.71 μm at 20 °C.

The effects of fiber dispersion can be effectively seen by considering the Taylor series expansion of the mode-propagation constant β around the central frequency ω of the pulse spectrum

$$\beta(\omega) = \beta_0 + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2 + \frac{1}{6}\beta_3(\omega - \omega_0)^3 + \dots, \tag{2.7}$$

where the term β_i is the i-th derivative of the propagation constant with respect to the angular frequency ω_0, such as

$$\beta_i = \frac{\partial^i \beta}{\partial \omega^i}\Big|_{\omega=\omega_0} \quad (i = 0, 1, 2, \dots). \tag{2.8}$$

Each of the constants β_i, characterizes different effects related to the propagation of optical signals in an optical fiber. The zero-order term, β_0, measured in $[km^{-1}]$, represents a common phase shift in the propagating signal. The first-order term, β_1, is measured in units of [ps/km] and is the inverse of the group velocity v_g, which represents the speed at which the overall energy of the signal propagates. The group index n_g is defined as the ratio of the speed of light c to the group velocity v_g, i.e., $n_g = c/v_g$. The second-order term β_2, measured in [ps^2/km], corresponds to the group velocity dispersion (GVD), and

[1] The equation in Sellmeier's original work [71] uses wavelengths. However, a modified version of the Sellmeier equation, which operates with optical frequency, is derived by replacing λ_j with ω_j and the constants A_j with B_j.

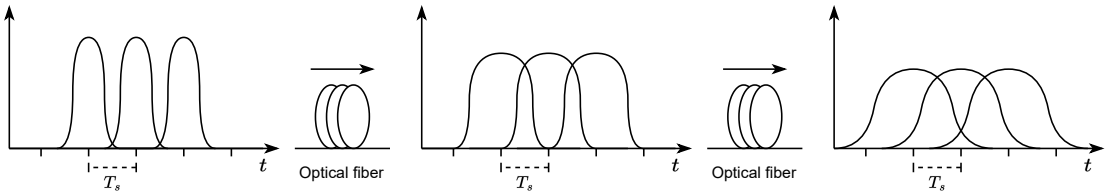


Figure 2.3: A schematic diagram of the pulse spreading due to chromatic dispersion. The attenuation due to transmission over fiber is neglected.

quantifies the change in velocities of different spectral components within a propagating pulse. Finally, the third-order term β_3 is measured in [ps^3/km], and represents the slope of the GVD and indicates the variation of GVD with respect to angular frequency.

When a pulse with a spectral width $\Delta\omega$ is launched into the fiber of length L, the pulse broadening for the fiber length is governed by

$$\Delta T_{\mathrm{CD}} = \frac{\partial T}{\partial \omega}\Delta\omega = \frac{\partial}{\partial \omega}\Big(\frac{L}{v_g}\Big)\Delta\omega = L\frac{\partial^2 \beta}{\partial \omega^2}\Delta\omega = L\beta_2\Delta\omega, \tag{2.9}$$

where T denotes the duration required for the pulse to propagate. This indicates that the severity of chromatic dispersion is primarily influenced by the coefficient β_2, the total length of the optical fiber, and the spectral width of the signal. Eq. (2.9) uses the definition $v_g = (\partial\beta/\partial\omega)^{-1}$.

When considering fiber-optic transmission systems, it is generally more prevalent to utilize the dispersion coefficient, D, and its slope, D'. These factors depend on the wavelength and are connected to β_2 and β_3 as follows:

$$D = \frac{\partial \beta_1}{\partial \lambda} = -\frac{2\pi c}{\lambda^2}\beta_2 \tag{2.10}$$

$$D' = \frac{\partial D}{\partial \lambda} = \frac{4\pi c}{\lambda^3}\left(\beta_2 + \frac{\pi c}{\lambda}\beta_3\right), \tag{2.11}$$

where D is measured in $[ps/(nm.km)]$ and D' is measured in $[ps/(nm^2.km)]$. Apart from the material dispersion we just discussed, another type of dispersion, called *waveguide dispersion*, arises due to the fiber's structure, which is determined by the transverse index profile. It affects the propagation of different modes, leading to variations in their group velocities and causing further broadening of the optical pulse. While material dispersion alone cannot be zero, it is possible to achieve zero dispersion at a specific wavelength, known as the zero-dispersion wavelength, where the influences of material and waveguide dispersion have equal magnitudes but with opposite signs.

2.2.1.3 Polarization Mode Dispersion

Single-mode fibers can support two signals polarized in orthogonal directions, x and y, which are represented by two orthogonally polarized modes in linear superposition. This

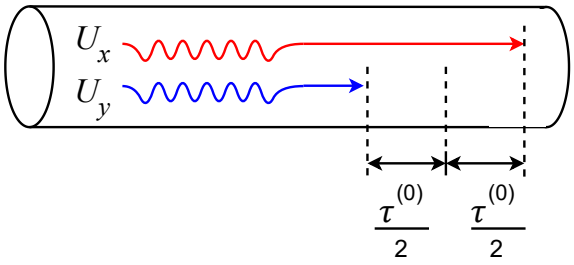


Figure 2.4: A diagram illustrating the effect of first order PMD (also referred to as DGD) in time-domain over a fiber with constant birefringence.

allows for any polarized optical wave to be decomposed into these two modes, each carrying independent information. However, the phenomenon of PMD may arise in systems utilizing dual-polarization transmission due to variations in the refractive index along the length of the fiber, which affects the propagation of each polarized mode differently. The PMD can be caused by a variety of factors, including material birefringence, random imperfections or bending of the fiber, mechanical stress, and temperature changes.

These factors lead to an asymmetry in the fiber, causing one polarized mode to propagate at a different velocity than the other, leading to the differential group delay (DGD). A diagram illustrating the effect of DGD is shown in Figure 2.4. In an ideal fiber, the properties of both modes are the same because of its cylindrical symmetry, but in real fibers, asymmetry causes the modes to have different properties, leading to PMD. The birefringence B_m and the difference in group velocity $\Delta\beta_1$ are linked together by the equation

$$B_m = \frac{c}{\omega_0} |\beta_{1x} - \beta_{1y}| = |n_x - n_y|, \tag{2.12}$$

where β_{1x} and β_{1y} are the inverse of the group velocities of each mode respectively along the x- and y- axes, and n_x and n_x is the respective refractive indices. The axes x and y are sometimes referred to with the slow and fast axes depending on the propagation velocity. The difference in propagation velocities across each polarization leads to pulse broadening due to PMD. In fibers with constant birefringence, the time delay due to DGD, denoted as ΔT_{DGD}, between the two polarization components propagating inside a fiber of length L can be estimated by

$$\Delta T_{\mathrm{DGD}} = \left| \frac{L}{v_{gx}} - \frac{L}{v_{gy}} \right| = L \left| \beta_{1x} - \beta_{1y} \right|, \tag{2.13}$$

where v_{gx} and v_{gy} are the group velocities of the signals propagating in the x- and y-polarizations, respectively. When a continuous wave signal is transmitted through an optical fiber at an angle to one of its axes, the velocity difference between the two axes induces a phase shift between the two polarizations. The polarization state of the light in this case oscillates from linear to elliptical, then to circular, and back to linear. The beat length, defined as $L_B = \lambda / B_m$, corresponds to the distance at which this cycle repeats once.

Birefringence effects and polarization orientation change randomly every approximately 1 cm in high-birefringence fibers, and up to tens of meters in low-birefringence fibers [8].

2.2.2 Nonlinear Transmission Effects

The Kerr effect is a nonlinear optical phenomenon characterized by a change in the refractive index of a material that is proportional to the intensity of an electric field within the material. This change in refractive index leads to a change in the phase of light passing through the material, resulting in nonlinear optical effects such as self-phase modulation, four-wave mixing, and soliton formation. The Kerr effect thus start to impact the transmission system at high transmitted power.

The change in refractive index with the intensity of the signal, $|q|^2$, can be modeled by:

$$\tilde{n}(\omega, |q|^2) = n(\omega) + \frac{n_2}{A_{\text{eff}}}|q|^2, \tag{2.14}$$

where A_{eff} is the effective core area of the fiber measured in $[\text{m}^2]$, $n(\omega)$ is the linear part of the refractive index given by Eq. (2.5), and n_2 is the nonlinear index in $[\text{m}^2/\text{W}]$. The typical values of these parameters are $n_2 = 2.5{-}3.2\times10^{-20}\,\text{m}^2/\text{W}$ and $A_{\text{eff}} = 50{-}120\,\mu\text{m}^2$. For SMF, $A_{\text{eff}} = 80\,\mu\text{m}^2$.

The channel effects discussed so far, including attenuation, chromatic dispersion, and nonlinearities, simultaneously impact the transmission of a signal along a fiber. We will denote the dual-polarized signal at time t and distance z with the vector $\mathbf{q}(z,t) = [q_x(z,t), q_y(z,t)]$. The propagation of a signal over a long, randomly birefringent transmission fiber is governed by the nonlinear Schrödinger equation (NLSE):

$$\frac{\partial \mathbf{q}(z,t)}{\partial z} = -\frac{\alpha}{2}\mathbf{q} - \frac{j\beta_2}{2}\frac{\partial^2 \mathbf{q}}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3 \mathbf{q}}{\partial t^3} + j\gamma\,|\mathbf{q}|^2\,\mathbf{q}, \tag{2.15}$$

where γ is the nonlinear coefficient which is defined as:

$$\gamma = \frac{8}{9}\frac{2\pi}{\lambda}\frac{n_2}{A_{\text{eff}}}. \tag{2.16}$$

The factor $8/9$ in the nonlinear coefficient in Eq. 2.16 is specific to randomly birefringent fibers, and changes to 1 in fibers with constant birefringence. The NLSE demonstrates that the strength of the Kerr effect is directly proportional to the optical signal power, as represented by $|\mathbf{q}|^2$. This means that the impact of the Kerr effect is strongest in the initial kilometers of the fiber transmission, where the optical power is high. As the optical signal attenuates inside the fiber, the effect of nonlinearity becomes negligible at the end of the fiber. It is useful to define an effective length L_{eff} normalized against the signal power. This formula is derived by integrating the normalized power along the entire length of the fiber.

$$L_{\text{eff}} = \int_0^L \exp\left(-\alpha z\right)dz = \frac{1-\exp\left(-\alpha L\right)}{\alpha}, \tag{2.17}$$

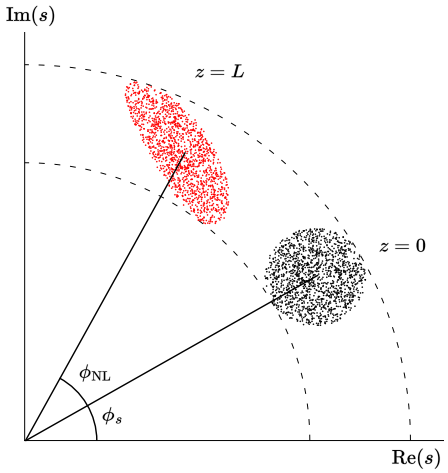


Figure 2.5: An illustration of the effect of self-phase modulation (SPM) on the received symbols constellation.

where L is the fiber length and α is the attenuation coefficient. The effective length of a 100 km fiber assuming an attenuation coefficient of 0.2 dB/km is calculated to be 21.5 km. However, for a 15 km fiber, the effective length is half of this value, demonstrating that approximately 50% of the nonlinearity in a 100 km fiber occurs in the first 15 km of the fiber transmission.

Nonlinear effects attributed to Kerr nonlinearity form a wide family of different adverse effects impacting the quality of the propagating signal. When different signals with different wavelengths propagate together in the fiber (as is the case in wavelength-division multiplexing (WDM)), the combined energy contribute to nonlinear impairments which impact all the signals. Depending on the source of the nonlinear impairment, Kerr effect is divided in two categories namely: inter-channel and intra-channel effects. Intra-channel refers to nonlinearities occurring within a wavelength channel itself whereas inter-channel effects originate from the interaction between different wavelength channels. Both types of effects can be in turn broken down into more elementary nonlinearities arising either from signal-signal interactions or from the interplay between the optical signal field and the noise, mainly amplified spontaneous emission (ASE).

2.2.2.1 Self-Phase Modulation

As the electromagnetic field propagates through the optical fiber, it interacts with the refractive index based on its instantaneous power, as described in Eq. 2.14. This interaction results in a nonlinear phase shift known as self-phase modulation (SPM). The impact of SPM can be studied using the NLSE, represented by Eq. 2.15, assuming a non-dispersive channel ($\beta_2 = \beta_3 = 0$). The simplified equation capturing the nonlinear components of the

NLSE is:

$$\frac{\partial q(t,z)}{\partial z} = -\frac{\alpha}{2}q + j\gamma\left|q\right|^2 q. \tag{2.18}$$

The solution for this reduced model is obtained as:

$$q(t,z) = q(t,0)\cdot\exp\left(\frac{-\alpha z}{2}\right)\exp\left(j\Phi_{\mathrm{NL}}(t,z)\right) \tag{2.19}$$

$$\Phi_{\mathrm{NL}}(t,z) = \gamma\cdot\frac{1-\exp\left(-\alpha z\right)}{\alpha}\left|q(t,0)^2\right|, \tag{2.20}$$

where $\Phi_{\mathrm{NL}}(t,z)$ represents the SPM-induced nonlinear phase. The effects of phase distortion generated by SPM are illustrated in Figure 2.5. The figure demonstrates that SPM distorts the initial circular noise cloud into an elongated ellipse as it propagates through the fiber with a length of L. Two dashed circular arcs indicate the range of amplitude fluctuation. The induced phase distortion adversely affects the coherent receiver's ability to accurately detect the transmitted signal's phase. Furthermore, the interaction between CD and SPM can introduce distortions in the signal's amplitude. Therefore, it is crucial to meticulously design the system to mitigate the influence of SPM. This involves operating within the linear regime by reducing the launched power to a lower level. The SPM-induced phase shift results in a nonlinear frequency shift, which is described by the equation

$$\Delta f_{\mathrm{NL}}(t) = -\frac{1}{2\pi}\frac{\partial\Phi_{\mathrm{NL}}}{\partial t}. \tag{2.21}$$

SPM can hence lead to the generation of new frequency components as the pulse propagates through the fiber, which prompts the need to consider SPM impact when designing optical filters.

2.2.2.2 Cross-Phase Modulation

cross-phase modulation (XPM) is a nonlinear optical effect that can occur in optical fibers. It arises when two or more optical signals simultaneously propagate through a medium and dynamically interact with each other. This interaction leads to a significant phase shift experienced by one or both of the signals involved in the SPM process. The mechanism of SPM can be understood by considering the interaction between the optical fields and the refractive index of the medium through which they are propagating. When two or more optical fields are present, they create a time-varying refractive index in the medium, which in turn affects the propagation of each of the fields. The resulting phase shift in one of the fields depends on the intensity and wavelength of the other fields, as well as the length of the fiber and the nonlinear coefficient of the medium.

Let us consider the WDM signal with a complex envelope denoted as $\mathbf{q}(0,t)$, which is launched into the fiber link by combining the signals from different WDM channels:

$$q(0,t) = \sum_n q_n(0,t)e^{-j\omega_n t}, \tag{2.22}$$

where $q_n(0,t)$ represents the signal propagating in the n-th channel of the WDM, centered at ω_n. The index $n = 0$ corresponds to the central channel, and ω_0 represents the central frequency. Under the same assumptions made in Eq. 2.18, the induced nonlinear phase due to SPM can be expressed as:

$$\Phi_{\mathrm{NL,XPM}}(z,t) = \gamma \cdot \frac{1-\exp(-\alpha z)}{\alpha} \sum_{n\neq 0} |q_n(0,t)|^2 . \tag{2.23}$$

Similar to SPM, SPM poses challenges to transmission systems as it generates "*inter-channel crosstalk*" when multiple signals propagate together. In the case of multi-channel transmission, the detrimental effects of XPM are more pronounced compared to SPM. This is because the receiver only perceives the signal from one channel and remains unaware of the other signals propagating in adjacent channels. However, it is worth noting that XPM can also have positive effects on optical communication systems, particularly when employed for wavelength conversion or signal regeneration purposes.

2.2.2.3 Four-Wave Mixing

FWM is a nonlinear phenomenon where two or more frequency components interact to generate new ones. Four-wave mixing in fibers shares a common origin with self-phase modulation and cross-phase modulation, all resulting from the Kerr nonlinearity, where the intensity-dependant refractive index of silica is modulated by different frequency components. When two co-propagating input frequency components f_1 and f_2 are present, a refractive index modulation occurs, resulting in the generation of two new frequency components ($f_3 = 2f_1 - f_2$ and $f_4 = 2f_2 - f_1$). Other frequency components can also be generated, however they are less commonly used due to difficulty in phase-matching, especially in fiber optic systems. Two scenarios are considered in the context of FWM. In the first scenario, where $f_1 \neq f_2$, the frequency of the fourth wave is determined by adding the frequencies of the first and second waves and then subtracting the frequency of the third wave, denoted as $f_4 = f_1 + f_2 - f_3$. This is referred to as 'non-degenerate FWM Generation.' In the second scenario, when $f_1 = f_2$, the frequency of the fourth wave is expressed as $f_4 = 2f_1 - f_3$. This unique condition results in a distinct phenomenon known as 'Degenerate FWM,' which, although similar to SPM and XPM, differs in the degeneracy of the involved waves.

2.2.3 Optical Amplification

As optical signals travel through the fiber, they experience losses due to absorption and scattering, which weakens the signal and reduces its quality. To overcome the attenuation and dispersion effects that occur in optical fiber transmission, optical signal amplification is necessary. This compensates for the losses by boosting the signal power in the optical

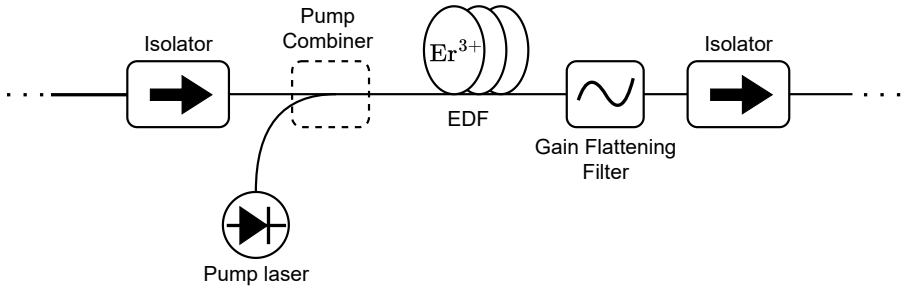


Figure 2.6: Schematic diagram of EDFA

domain, which is achieved through various methods, including erbium-doped fiber amplifiers (EDFAs), semiconductor optical amplifiers (SOAs), and Raman amplification. Fiber amplifiers are a type of optical amplifier that use optical fibers as a laser gain medium. They utilize glass fibers doped with rare earth ions such as erbium, neodymium, ytterbium, praseodymium, or thulium. These active dopants are provided with energy by pumping them with light from a laser, such as a fiber-coupled diode laser.

When it comes to long-haul optical fiber communications, EDFAs are by far the most adopted fiber amplifiers. EDFAs have the ability to effectively amplify light in the wavelength range of 1550 nm, where the optical fiber incurs the lowest loss. In EDFA, amplification is conducted through an Erbium-doped optical fiber which is pumped with light from the laser diodes. The pump light which propagates through the fiber core together with the signal to be amplified, which is typically 980 nm or 1480 nm in wavelength, is used to excite the Erbium ions (Er^{3+}) into an excitation state. By periodically amplifying the signal along the fiber, a significant loss of signal quality can be avoided.

EDFA Amplification is accompanied with ASE, which originates from the optical amplifier itself. Specifically, ASE is caused by the spontaneous emission of photons in the gain medium of the amplifier that are amplified by the same gain medium, resulting in amplified noise. This consecutively diminishes the signal-to-noise ratio. When using a series of cascaded lumped amplifiers, the accumulation of ASE noise can significantly increase. Each amplifier adds ASE noise, which is further amplified by the subsequent amplifiers in the link. If all amplifiers in the chain have the same gain $G = e^{\alpha L_A}$ and are spaced apart by the same distance L_A, the total ASE power for a chain of N_A amplifiers can be calculated as follows [5]:

$$P_{tot} = 2n_{sp}h\nu_0 N_A(G-1)\Delta\nu_0 = 2N_A S_{\text{ASE}}\Delta\nu_0. \tag{2.24}$$

Here, the factor 2 in the equation accounts for the ASE generated for the two orthogonal polarizations. S_{ASE} represents the power spectral density of the ASE noise, h is Planck's constant, $1 < n_{sp} \lesssim 1.5$ is the spontaneous emission factor, ν_0 is the carrier frequency, and $\Delta\nu_0$ is the bandwidth of the amplifier gain.

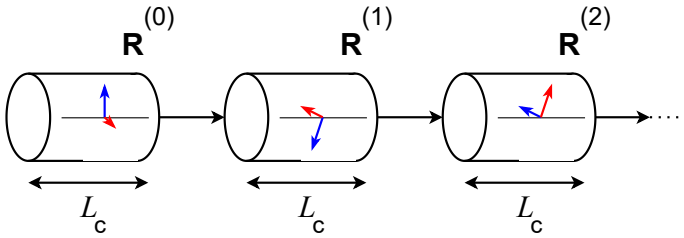


Figure 2.7: A diagram showing a fiber divided into a series of small birefringent segments coupled randomly.

2.2.4 Vectorial Form of the Nonlinear Schrödinger Equation

The equation which governs the evolution of the signals in two polarizations in optical fiber is the coupled nonlinear Schrödinger's equation (CNLSE) [73]

$$\frac{\partial q_x}{\partial z} + \beta_{1x}\frac{\partial q_x}{\partial t} + \frac{j\beta_2}{2}\frac{\partial^2 q_x}{\partial t^2} + \frac{\alpha}{2}q_x = j\gamma\left(|q_x|^2 + \frac{2}{3}|q_y|^2\right)q_x + \frac{\gamma}{3}q_y^2 q_x^* \exp(j\Delta\beta z), \quad (2.25a)$$

$$\frac{\partial q_y}{\partial z} + \beta_{1y}\frac{\partial q_y}{\partial t} + \frac{j\beta_2}{2}\frac{\partial^2 q_y}{\partial t^2} + \frac{\alpha}{2}q_y = j\gamma\left(|q_y|^2 + \frac{2}{3}|q_x|^2\right)q_y + \frac{\gamma}{3}q_x^2 q_y^* \exp(j\Delta\beta z), \quad (2.25b)$$

where q_x and q_y represent the complex envelopes of the signal in the x-polarization and y-polarization, respectively, as functions of time t and distance z, β_{1x} and β_{1y} are the first-order dispersion coefficients along each polarization, which are related to the beat length L_B with the formula $\Delta\beta = \beta_{1x} - \beta_{1y} = 2\pi/L_B$. The parameters β_2 and γ are the same for either polarization since they operate at the same wavelengths. The last term in (2.25a) and (2.25b) is caused by the coherent coupling between both polarizations, however for fiber length $L_{\text{fiber}} \gg L_B$ its contribution averages out to zero [8].

The first-order dispersion coefficients β_{1x} and β_{1y} are polarization-dependent, which results in DGD between the two polarizations. In practice, the direction of light polarization is not always maintained due to various factors such as fiber imperfections and twisting, which can introduce rapidly varying birefringence.

2.2.4.1 Numerical Resolution: Split Step Fourier Method

To numerically simulate the signal propagation in the forward direction, from the transmitter to the receiver, the commonly used method is the split-step Fourier method (SSFM) with distributed PMD [7]. The system of equations in (2.25) is solved within each span by discretizing the fiber span into N_{seg} segments of length δ_s. Within each segment i, a linear step, a PMD step, and a nonlinear step are performed iteratively from the transmitter end to the receiver end, as follows.

1. *Linear step*: Solves for the signal loss and CD in the frequency domain, by assuming

$\gamma = 0$. Considering only the terms which include α and β_2 in Eq. (2.25), we obtain

$$\hat{q}_{x/y}(z,\omega) \rightarrow \exp\left(-\frac{\alpha}{2}\delta_s + \frac{j\beta_2}{2}\omega^2\delta_s\right)\hat{q}_{x/y}(z,\omega), \tag{2.26}$$

where $\hat{q}_{x/y}(z,\omega)$ denotes the Fourier transform of the time domain signal $q_{x/y}(z,t)$.

2. *PMD step*: The PMD effects, including the DGD and random PSP, can be both modeled by applying the unitary Jones matrix $\mathbf{J}^{(i)}(\omega)$ to the signal vector $\hat{\mathbf{q}}(z,\omega) = [\hat{q}_x(z,\omega), \hat{q}_y(z,\omega)]^\top$ as follows

$$\hat{\mathbf{q}}(z,\omega) \rightarrow \mathbf{J}^{(i)}(\omega)\hat{\mathbf{q}}(z,\omega), \tag{2.27}$$

where $\mathbf{J}^{(i)}(\omega) = \mathbf{R}^{(i)}\mathbf{D}^{(i)}(\omega)$. The $\mathbf{R}^{(i)}$ here is a rotation matrix described as

$$\mathbf{R}^{(i)} = \begin{bmatrix} e^{j\frac{\phi_i}{2}}\cos\theta_i & e^{-j\frac{\phi_i}{2}}\sin\theta_i \\ -e^{j\frac{\phi_i}{2}}\sin\theta_i & e^{-j\frac{\phi_i}{2}}\cos\theta_i \end{bmatrix}, \tag{2.28}$$

where θ_i and ϕ_i, $i \in 1, 2, ..., N_{seg}$, are sequences of independent identically distributed (i.i.d.) random variables drawn from a uniform distribution from $[0, 2\pi)$. Furthermore, $\mathbf{D}^{(i)}(\omega)$ is the DGD matrix

$$\mathbf{D}^{(i)}(\omega) = \begin{bmatrix} e^{-j\omega\frac{\tau_i}{2}} & 0 \\ 0 & e^{j\omega\frac{\tau_i}{2}} \end{bmatrix}, \tag{2.29}$$

where the DGD parameters $(\tau_i)_{i=1}^{N_{seg}}$ are taken to be as i.i.d. random variables drawn from the normal probability distribution $\mathcal{N}(0, \tau\sqrt{\delta_s})$, where τ is the characteristic constant of the channel called the *PMD coefficient*.

3. *Nonlinear step*: Solves for the signal nonlinear effects by only considering the terms which include γ in Eq. (2.25)

$$q_{x/y}(z,t) \rightarrow \exp\left(j\gamma\delta_s\left(|q_{x/y}|^2 + \frac{2}{3}|q_{y/x}|^2\right)\right)q_{x/y}(z,t). \tag{2.30}$$

Upon performing the three aforementioned steps, the signal propagation is numerically simulated at a distance of $z + \delta_s$, resulting in the signal $q_{x/y}(z + \delta_s, t)$. This process is iteratively repeated for each segment until the end of the span. To simulate the entire optical fiber channel, amplification is applied at the end of each span to compensate for signal attenuation, and noise generated by ASE is added. The SSFM simulation continues until the signal has propagated through the entire optical fiber link.

2.2.4.2 Manakov Equation

One challenge in solving CNLSE is the rapidly changing SOP with a length scale of 0.3-100m, whereas PMD, chromatic dispersion, and Kerr nonlinearity have much larger length

scales spanning hundreds or thousands of kilometers. This changing SOP can cause the x-polarization and y-polarization signals to be indistinguishable due to rapid changes in the orientation of the axes of birefringence θ along the fiber. When the PMD is small but the birefringence varies rapidly, the Manakov equation can be used to describe signal propagation, which is given by [73]

$$\frac{\partial \mathbf{q}}{\partial z} = \left(-\frac{1}{2}\alpha - \beta_1(z)\frac{\partial}{\partial t} - j\beta_2\frac{1}{2}\frac{\partial^2}{\partial t^2} + j\frac{8}{9}\gamma||\mathbf{q}||^2 \right)\mathbf{q}. \tag{2.31}$$

Here, $\mathbf{q}(z,t) = [q_x(z,t), q_y(z,t)]^T$ is the Jones vector containing the propagating signals in both polarizations, and $\beta_1(z)$ describes the evolution of β_1 with distance. The Manakov equation is particularly well-suited for digital back-propagation (DBP) applications, especially when the evolution of the SOP along the fiber is unknown.

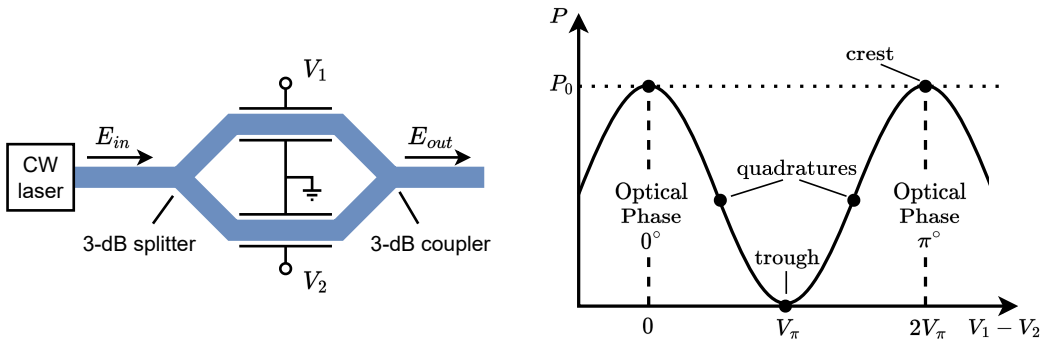
2.3 Optical Transmitters

The data carrying the information starts at the transmitter as a stream of bits, 1s and 0s, and are then encoded into a continuous signal using a modulation format. The continuous signal in the electrical domain must be transformed to the optical domain before being launched in the fiber. Some means of achieving this include: directly modulated Lasers (DML), electro-absorption modulators (EAM), Mach-Zehnder modulators (MZMs), and phase modulators. The MZM stands out as the most suitable option for practical long-haul high-speed transmission systems, making it the focus of this thesis.

Current optical fiber links are typically bi-directional, utilizing two fibers for two-way signal transmission. The combination of optical transmitters and receivers is usually housed in a single module known as a *transponder*.

2.3.1 Mach-Zehnder Modulators for OOK Modulation Format

The MZM is a type of external modulation component that is widely used in long-haul high-speed transmission systems. In this system, the Mach-Zehnder modulator is cascaded with a continuous wave (CW) distributed feedback (DFB) laser. The reasons why the MZMs stands out as a suitable option among all the other modulators, such as Directly Modulated Lasers (DML), Electro-Absorption Modulators (EAM), and Phase Modulators, are discussed in the following. First, the MZMs has a high electro-optical bandwidth. This means that it can support high data rates, which are essential for long-haul high-speed transmission systems. This high electro-optical bandwidth of the MZMs makes it a suitable choice for transmission rates of 40 Gb/s or higher, which have been demonstrated experimentally. On the other hand, Directly Modulated Lasers (DML) have a relatively small electro-optical bandwidth, which makes them less suitable for high-speed long-haul WDM systems. Another advantage of the MZMs is its low frequency chirp behavior.



(a) (b)

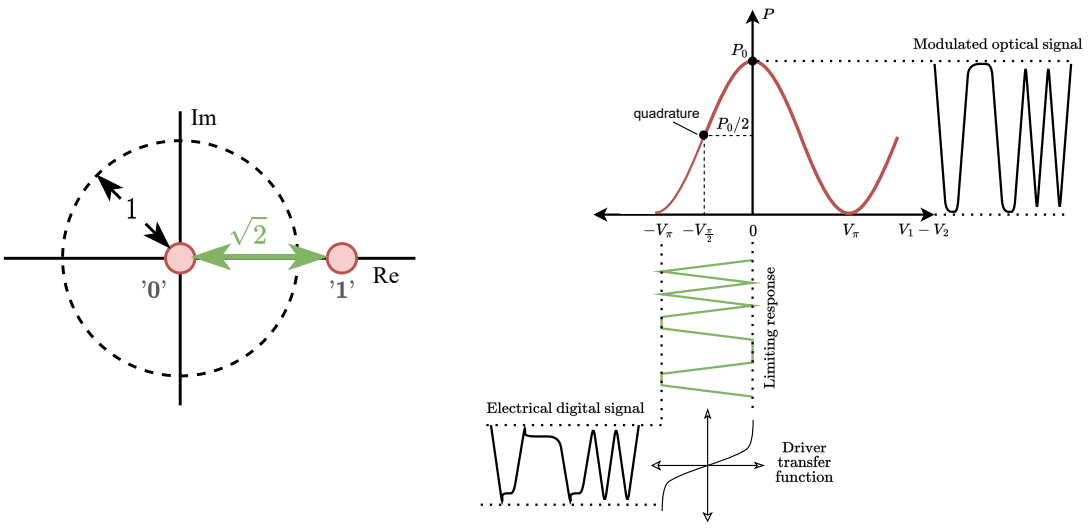
Figure 2.8: Mach-Zehnder modulator, (a) Schematic of the modulator circuit, (b) Relationship between the optical output and the voltage difference in a MZMs.

Chirp is an undesired phase modulation that results from the direct modulation of the pump current at high data rates. This results in a broadened spectrum and increased sensitivity to chromatic dispersion, which can degrade the signal quality. In contrast, the MZMs modulates the phase of the signal, which minimizes the impact of chirp on the signal quality.

MZMs were first proposed by Ernst Mach and Ludwig Zehnder and consist of a 3-dB coupler, a 3-dB splitter (sometimes referred to as a Y-junction), and two waveguides (arms) of equal length. The incoming light is divided in half and travels through each waveguide. The two waveguides are made of materials with strong electro-optic properties, such as $LiNbO_3$, GaAs, and InP. When electric fields are applied across the electro-optic material, this alters the optical path length, which induces a phase-shift in the optical signal based on the applied voltage as depicted in Fig. 2.8, leading to phase modulation. By adjusting the drive voltage, a specific phase difference between the signals in each arm can be achieved, which is then translated into amplitude fluctuations when the signals at the end of each arm interfere through the second 3-dB coupler. The input-output relationship of the MZM is determined by the voltages V_1 and V_2 applied to its arms, which modulate the output electric field, E_{out}, as a function of the input, E_{in}. This modulation is described by:

$$E_{out} \propto E_{in} \cos\left(\pi \frac{V_1 - V_2}{2V_\pi}\right), \tag{2.32}$$

where the phase shift in E_{out} is controlled by the difference $V_1 - V_2$, with V_π being the voltage that induces a π radian phase shift. This voltage difference also corresponds to a change from constructive interference to destructive interference. Both arms are often modulated jointly with $V_1 = -V_2$, the modulator is referred to as a "push-pull" configuration.



(a) (b)

Figure 2.9: non-return-to-zero (NRZ)-OOK: (a) Constellation Diagram with Normalized Power, (b) Modulated Optical Signal Generation via MZMs.

2.3.1.1 Intensity Modulation/Direct Detection Systems

When optical fiber systems first emerged, they relied on basic modulation formats based on intensity modulation (IM), such as non-return-to-zero (NRZ)-OOK, combined with direct-detection (DD) at the receiver. In the NRZ-OOK modulation format, binary data in its electrical form is directly transformed into an optical signal. This is accomplished by replicating the binary sequence as pulses of light in the optical domain. The "0"s in the electrical binary sequence are encoded as weak optical-field power, ideally with null power, while the "1"s are encoded with high power. This direct encoding of the binary data in the amplitude of the optical field makes OOK a type of amplitude-shift keying (ASK) modulation. The constellation diagram of an NRZ signal, as seen in Fig. 2.9, displays the possible symbols used by the modulation scheme as points in the complex plane, and the amplitude is null for "0" symbols and high for "1" symbols with a constant phase. In OOK modulation format, assuming that the signal average power is E, the Euclidean distance d_{min} between two points in the complex plane has a value of $\sqrt{2E}$. This Euclidean distance is closely related to the bit-error-ratio (BER), as they are linked by the equation

$$BER = Q\left(\sqrt{\frac{d_{min}^2}{2N_0}}\right), \tag{2.33}$$

where $N_0/2$ is the spectral density assuming additive white Gaussian noise (AWGN) and $Q(\cdot)$ is the Q-function described as

$$Q(x) = \frac{1}{2}\text{erfc}\left(\frac{x}{\sqrt{2}}\right). \tag{2.34}$$

The BER for OOK is hence

$$BER_{\text{OOK}} = SER_{\text{OOK}} = Q\left(\sqrt{\frac{E_b}{N_0}}\right), \tag{2.35}$$

where E_b is the energy per bit. For long-haul optical transmissions, the NRZ-OOK format is typically generated through a continuous-wave DFB laser and a MZM. The MZM's bias voltage is set at the quadrature point, which is positioned in the middle of the linear region of the MZM transfer function. The MZM's output is then driven by the modulating signal whose minimum and maximum levels are within this linear region as shown in Fig.2.9b. To ensure that the MZM operates within the linear region, the electrical drive signal must have a peak-to-peak amplitude of less than V_π. Beyond the quadrature point in either direction, the linear relationship weakens, and the transfer function enters the quadratic region. On the receiving end, a single photodiode converts the optical signal into an electrical signal, and an electrical circuit integrates the photocurrent over a specific time period. The system's receiver, equipped with a clock, then detects the pulses and determines the transmitted binary data through a threshold decision, which is represented as a slicer in Fig. 2.4-a.

2.3.2 I&Q Mach-Zehnder Modulator for Advanced Modulation Formats

The transmission of information at a high bit rate is one of the primary goals of communication systems. To achieve this goal, various modulation formats have been proposed that support coherent detection. In coherent detection, both the amplitude and phase information of the carrier signal are extracted, making it possible to recover the original information with high accuracy.

Modulation relies on the concept of the geometric representation of signals, which represents a set of M energy signals $[s_i(t)]$ as a linear combination of N orthonormal basis functions, where $N \leq M$. For a set of real-valued energy signals $s_1(t), s_2(t), \ldots, s_M(t)$ with duration T seconds, each signal is expressed as

$$s_i(t) = \sum_{i=1}^{N} s_{i,j}\phi_j(t), \quad \begin{cases} 0 \leq t \leq T \\ i = 1, 2, \ldots, M \end{cases} \tag{2.36}$$

where the coefficients $s_{i,j}$ are defined as

$$s_{i,j} - \int_0^T s_i(t)\phi_i(t)dt, \quad \begin{cases} i = 1, 2, \ldots, M \\ j = 1, 2, \ldots, N \end{cases} \tag{2.37}$$

The basis functions $\phi_j(t)$ are orthonormal, which means that all real-valued basis functions satisfy the condition

$$\int_0^T \phi_i(t)\phi_j(t)dt = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases} \tag{2.38}$$

The analysis above demonstrates that a finite energy signal can be uniquely represented as a set of orthonormal basis functions, and there is a one-to-one correspondence between the signal and the orthogonal basis constructing it. The coefficients for a N-dimensional vector, called the signal vector $\vec{s_i}$, is defined as:

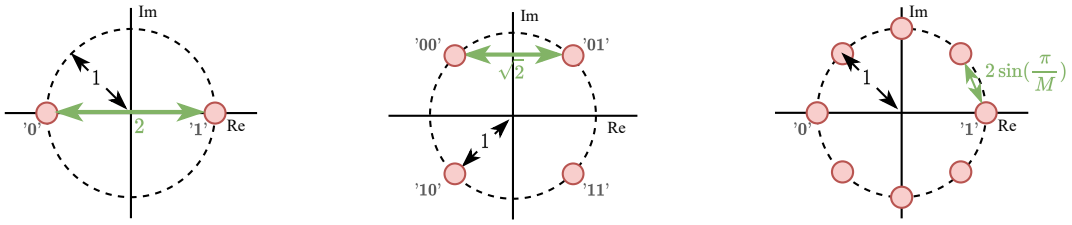
$$\vec{s_i} = \begin{bmatrix} s_{i,1} \\ s_{i,1} \\ \cdots \\ s_{i,N} \end{bmatrix}, \quad i = 1, 2, \ldots, M. \tag{2.39}$$

Thus, the use of an orthogonal basis set allows for efficient representation of finite energy signals in terms of a finite number of coefficients. The receiver uses a maximum likelihood procedure to determine the transmitted signal. It begins by finding the received signal vector through an inner product between the received signal waveform and each orthonormal basis. The procedure then assumes that the transmitted signal is the closest signal vector from a pre-defined set of constellation points.

The choice of modulation format depends on factors such as available bandwidth, channel conditions, and noise level. Popular modulation formats, including ASK, frequency-shift Keying (FSK), phase-shift keying (PSK), and quadrature amplitude modulation (QAM). Each modulation format has unique advantages and disadvantages, making it difficult to definitively state that one modulation format is superior or inferior to another. While higher modulation enables higher bitrates, it also increases susceptibility to noise and interference, requiring a higher signal-to-noise ratio (SNR) for reliable transmission. In general, no single modulation format is superior to all others in all aspects. The following sections discuss some of these modulation formats.

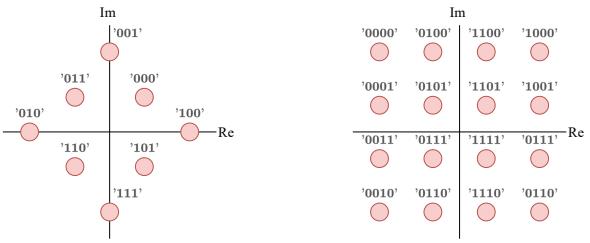
2.3.2.1 Phase-Shift Keying

PSK is a widely used modulation format in optical fiber communication. It involves transmitting a signal with a constant amplitude, with information encoded within a set of M phase levels, where each phase corresponds to a specific value from 0 to 2π. To understand the advantages of coherent detection and phase encoding, a direct comparison between binary PSK and OOK can be made, as both schemes transmit one bit of data in each symbol. The signal constellations of binary phase-shift keying (BPSK) and OOK, as shown in Fig.2.10 and Fig.2.9, illustrate the increased minimum distance d_{min} of BPSK compared to OOK, demonstrating BPSK's superior noise sensitivity. By normalizing the BPSK constellation diagram with respect to the square root of the average power of the optical signal, the increased minimum distance, d_{min}, in the case of BPSK compared to OOK, results in better noise sensitivity, as the BER decreases with the increase of d_{min}, according to (2.33) (the Q-function is a monotonically decreasing function). This is due to the fact that both



(a) (b) (c)

Figure 2.10: Constellation diagrams for (a) binary PSK, (b) quadrature PSK, and (c) 8-PSK modulation schemes, along with their minimum distance values.



(a) (b)

Figure 2.11: Two examples of M-QAM constellations with Grey coded bit sequences, (a) shows an 8-QAM constellation, and (b) shows a 16-QAM constellation.

symbols in BPSK carry the same energy, but with different phases (either 0 or π). Thus, the same Euclidean distance as OOK can be achieved with half the power. This leads to a 3-dB lower optical signal-to-noise ratio (OSNR) requirement to achieve the same BER as OOK. The symbol-error-ratio (SER), which is equivalent to the BER formula for BPSK, is given by the formula:

$$BER_{\text{BPSK}} = SER_{\text{BPSK}} = Q\left(\sqrt{\frac{2E_b}{N_0}}\right), \tag{2.40}$$

where E_b is the energy per bit and $N_0/2$ is the noise spectral density.

2.3.2.2 Quadrature Amplitude Modulation

In general, as the number of phases in M-PSK increases, the sensitivity to noise also increases rapidly. This is why values for M higher than 8-PSK are rarely used in practice. However, to encode more information in the signal while still maintaining a reasonable level of noise sensitivity, amplitude modulation can be used to add an extra dimension for information encoding. One such modulation technique is QAM, which combines changes in both amplitude and phase levels. Compared to PSK, QAM can achieve higher spectral efficiency and allows for transmitting more information per symbol. Therefore, it is a popular modulation technique used in modern communication systems.

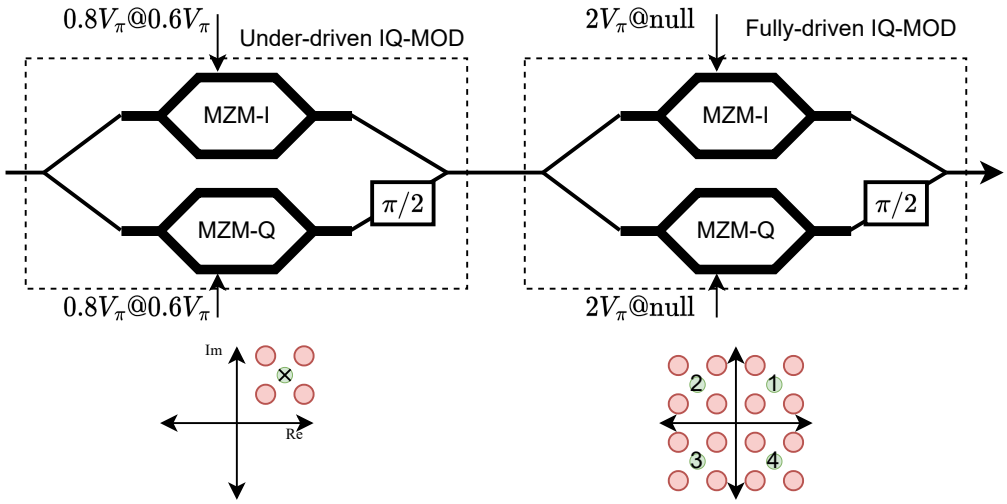


Figure 2.12: Principle of 16-QAM signal generation with tandem operation of two IQ-modulators

Fig. 2.11 shows the constellation points of 8-QAM and 16-QAM, and the encoded bits using Grey coding. The BER for M-QAM is approximately [86]

$$BER_{M\text{QAM}} \approx \frac{1}{\log_2(M)} SER_{M\text{QAM}} = \frac{4}{\log_2(M)} Q\left(1 - \frac{1}{\sqrt{M}}\right)\left(\sqrt{\frac{3E_s \cdot \log_2(M)}{N_0(M-1)}}\right), \tag{2.41}$$

where E_s is the energy of the symbol, and $N_0/2$ is the spectral density of the AWGN.

Generating an M-QAM signal typically requires the use of two MZMs operating in tandem. One MZM generates the in-phase component, while the other produces the quadrature component. These components are added together after applying a phase shift of $\frac{\pi}{2}$ to the quadrature component.

For $M > 4$, each MZM's driving signal is multi-level, comprising $\log_2 M$ levels. However, using multi-level driving signals for the MZMs introduces challenges that impede the generation of high-quality QAM signals: i) Imposing stringent bandwidth requirements on both the modulator and the electrical drivers when handling multi-level electrical driving signals. ii) Introducing high complexity associated with high-order optical integration. Therefore, an alternative to multi-level electrical driving signals is the use of binary electrical driving signals, which can simplify the generation of advanced optical modulation formats [70]. In this scenario, multiple binary-signal-driven IQ-MZMs are employed in tandem, as illustrated in Fig. 2.12.

2.3.3 Polarization-Division Multiplexing I&Q Modulator

polarization-division multiplexing (PDM) is a technique used in fiber optic transmission systems to increase the amount of data that can be transmitted. This is accomplished by exploiting the polarization property of light waves, which allows data to be sent simul-

taneously over the same physical medium (in this case, an optical fiber). Light can be polarized in two orthogonal directions, typically referred to as the vertical and horizontal states. PDM works by modulating the signal data onto two orthogonally polarized light waves, which are then combined and transmitted through the same fiber. At the receiving end, the light is split into its constituent polarization states, and the data is demodulated. Because the two states of polarization do not interfere with each other, the data streams can be separated cleanly at the receiving end, effectively doubling the transmission capacity of the fiber.

PDM has become particularly important with the advent of coherent optical fiber systems, which use a local oscillator and coherent detection to receive the signal. These systems can accurately separate the two polarization states, making PDM an effective way to increase data transmission capacity. In addition to being used on its own, PDM is often used in combination with other multiplexing techniques, such as WDM. The combination of WDM and PDM can further increase the capacity of fiber optic systems. For example, PDM-quadrature PSK is a commonly used format in long-haul transmission systems due to its high spectral efficiency and resilience to fiber nonlinearities

While PDM offers numerous advantages, it also comes with its own set of challenges. For instance, a phenomenon known as PMD can occur in optical fibers, which causes the different polarization modes to propagate at different speeds. This can lead to signal degradation and limit the overall system performance. PMD is a significant issue in older fibers but has been largely mitigated in newer, higher-quality fibers. Additionally, adaptive equalization techniques at the receiver can compensate for PMD to some extent.

2.4 Digital Coherent Receiver

The receiver's primary goal is to retrieve the transmitted bits. In essence, the receiver operates in reverse order compared to the transmitter, converting the optical signal to the electrical signal, and then processing the electrical signal to retrieve the transmitted bits. Since we generally have no control over the fiber channel connecting the transmitter and the receiver, the receiver has the additional task of mitigating the deterministic channel effects that impact the transmitted signal, a process known as channel equalization.

Optical receivers can be divided into two categories based on whether they use the signal's phase to decode information: direct detection and coherent detection. Direct detection receivers, widely used in earlier optical fiber transmission systems, are the simplest type of receivers. In this type of receiver, a photodiode produces an electrical signal proportional to the light's power, and the receiver's decision is based on a set of thresholds determined by the signal energy. This mode of reception is utilized in amplitude modulation techniques such as ASK, including its simpler variant, OOK.

Coherent detection, on the other hand, is fundamentally different from direct detection in that it retrieves both the amplitude and phase of the optical signal, or alternatively, the I- and Q-components. This is accomplished by using a local oscillator wave that matches the frequency of the oscillator at the transmitter. By mixing the optical signal and the local oscillator wave, an electrical signal is generated. The signal is then sampled at a rate of at least the Nyquist rate, and the resulting data can be processed by the digital signal processing block to compensate for channel impairments. As a result, coherent detection allows for a wider range of modulation formats, including mixed phase and amplitude modulation formats like M-QAM.

Coherent reception relies on compensating for the channel effects that affect the transmitted signal as it travels through the fiber channel. The DSP chain at the receiver is responsible for equalizing the channel effects, and it consists of multiple individual blocks, each designed to address a specific type of distortion or impairment. The DSP chain starts with matched filtering to reduce noise, followed by chromatic dispersion compensation to compensate for the dispersion effect, multiple input multiple output (MIMO) equalization to mitigate PMD, and carrier and frequency offset estimation to correct for phase and frequency offsets. The linear DSP can effectively perform equalization for linear effects, but nonlinear equalization may require the incorporation of blocks such as the DBP to jointly reverse the chromatic dispersion and nonlinear impairments. In the following sections, we will provide a concise description of each of these processing blocks.

2.4.1 Polarization Diversity Coherent Detector

The coherent optical detector consists of a local oscillator, a 90° optical hybrid, balanced photodiodes, and trans-impedance amplifiers. To achieve polarization diversity, a polarization beam splitter is added to split the oscillator's signal and the received signal into their polarization components, and a second 90° optical hybrid is used. The local oscillator generates a strong optical signal with a stable phase and frequency, which is mixed with the received optical signal in each 90° optical hybrid. An illustration of the polarization diversity coherent detector is shown in Fig. 2.13.

Assuming the emitted optical field $q_s(t)$ carries information on its two orthogonal polarizations, x and y, through polarization division multiplexing, where both fields are phase-modulated, the signal can be expressed as follows, neglecting any added AWGN:

$$\mathbf{q}_s(t) = \begin{bmatrix} q_{s,x}(t) \\ q_{s,y}(t) \end{bmatrix} = \begin{bmatrix} \sqrt{P_{s,x}} \exp(j\omega_{LO}t + j\phi_{s,x}(t) + j\phi_{\Delta\nu}(t)) \\ \sqrt{P_{s,y}} \exp(j\omega_{LO}t + j\phi_{s,y}(t) + j\phi_{\Delta\nu}(t)) \end{bmatrix}, \tag{2.42}$$

where $\omega_0 = 2\pi f_0$ is the central angular frequency of the laser, $P_{s,x}$ and $P_{s,y}$ are the powers of each polarized field, $\phi_{s,x}(t)$ and $\phi_{s,y}(t)$ are the modulated phases, and $\phi_{\Delta\nu}$ represents the phase distortion effecting both polarizations. The linearly polarized light of the local

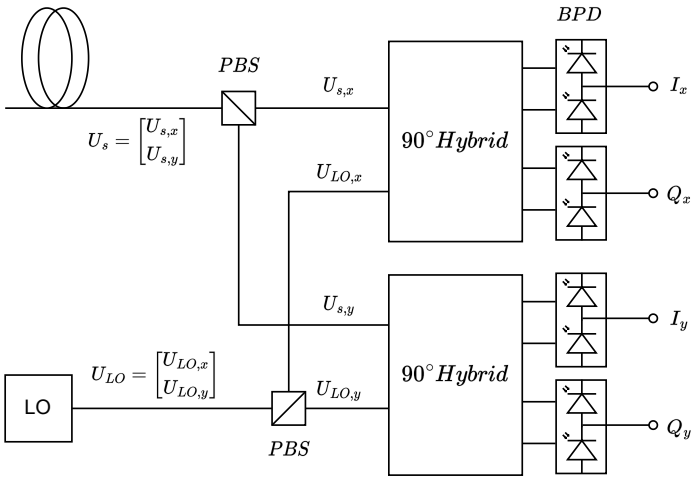


Figure 2.13: polarization diversity coherent detector.

oscillator is also split into two polarization components, and each beam is directed into a coherent mixer. A polarizer and quarter-wave plate are then used to transform the linear polarization, allowing each beam to be represented as:

$$\mathbf{q}_{LO}(t) = \frac{1}{2}\begin{bmatrix} \exp(j\frac{\pi}{2})q_{LO} \\ q_{LO} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \sqrt{P_{LO}}\exp(j\omega_{LO}t + j\frac{\pi}{2}) \\ \sqrt{P_{LO}}\exp(j\omega_{LO}t) \end{bmatrix}, \tag{2.43}$$

where P_{LO} is the power output of the oscillator. The diagram in Figure 2.13 shows that the signals from the optical hybrids are connected to balanced photodiodes. These photodiodes produce two currents, representing the I and Q components of the optical signal's polarization, respectively, which are given by:

$$I_x = I_{PD1} = R_{PD}\sqrt{P_{s,x}P_{LO}}\cos(\Delta\omega t + \phi_{s,x}(t) + \phi_{\Delta\nu}(t)) \tag{2.44}$$

$$Q_x = I_{PD2} = R_{PD}\sqrt{P_{s,x}P_{LO}}\sin(\Delta\omega t + \phi_{s,x}(t) + \phi_{\Delta\nu}(t)) \tag{2.45}$$

$$I_y = I_{PD3} = R_{PD}\sqrt{P_{s,y}P_{LO}}\cos(\Delta\omega t + \phi_{s,y}(t) + \phi_{\Delta\nu}(t)) \tag{2.46}$$

$$Q_y = I_{PD4} = R_{PD}\sqrt{P_{s,y}P_{LO}}\sin(\Delta\omega t + \phi_{s,y}(t) + \phi_{\Delta\nu}(t)), \tag{2.47}$$

where R_{PD} is the photodetector responsivity coefficient.

2.4.2 DSP-Based Equalization

The invention of digital coherent receiver revolutionized the design of optical transmission systems. Not only had digital coherent receivers contributed to improving the sensitivity of transmission systems, but they can also be used to overcome optical impairments such as CD and PMD [89]. There are four key subsystems that make up an optical receiver: Map the optical signal into a set of electrical signals, convert the analogue signal into a

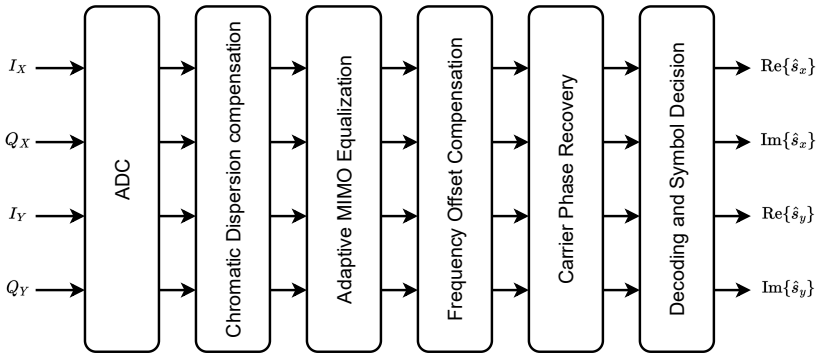


Figure 2.14: Key components of a traditional DSP chain.

digital signal, equalize and demodulate the signal, and apply error correction decoding and detection. Since we are mainly concerned with the third block in this thesis, we assume knowledge of the signal $\mathbf{q}(z,t) = [q_x(z,t), q_y(z,t)]^T$ in the digital domain.

The DSP chain, as depicted in Fig. 2.14, is designed to compensate for the linear effects within the channel, including chromatic dispersion, phase and frequency mismatch, and the effects arising from polarization mode dispersion. In the absence of nonlinear effects, the DSP algorithm can perform optimally, eliminating all linear effects.

2.4.2.1 Chromatic Dispersion Compensation

The effects of CD has been addressed earlier. Since CD is a polarisation-independent phenomenon, it can be compensated first before equalising and demultiplexing the received signal. This can be achieved by designing a filter that compensates for the residual amount of chromatic dispersion in the fiber. The value of dispersion can be accurately determined as it primarily relies on two factors, namely the dispersion parameter β_2 and the total length L of the fiber, which are often known by design. The transfer function of the chromatic dispersion compensation (CDC) filter is

$$H(\omega, L) = \exp\left(-j\frac{D\lambda^2}{4\pi c}\omega^2 L\right). \tag{2.48}$$

Here, D represents the dispersion coefficient of the fiber, λ is the wavelength, and c is the speed of light. The fraction $\frac{D\lambda^2}{4\pi c}$ can sometimes be replaced with the dispersion parameter $\beta_2/2$ for ease of expression.

While an analytical expression for this filter exists, it cannot be practically implemented due to its infinite duration. Therefore, finite frequency response (FIR) filters are used, whose length depends on the amount of CD to be compensated for. To recover the original signal from the dispersed signal as demonstrated in [88], the FIR filter with tap weights given by:

$$h_{cd}[k] = \frac{1}{\sqrt{\rho}}\exp\Big(-j\frac{\pi}{\rho}\Big[k - \frac{N-1}{2}\Big]^2\Big), \tag{2.49}$$

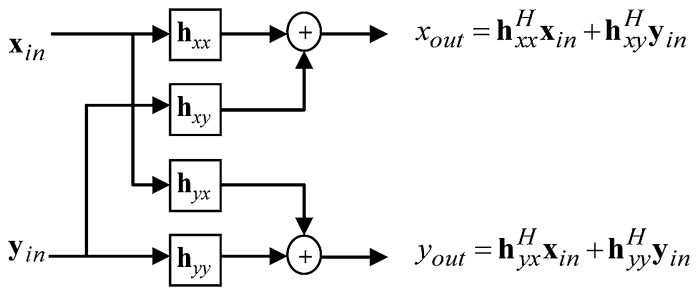


Figure 2.15: MIMO equalizer as shown in [88].

where $k \in [0, N-1]$, $N = \lfloor |\rho| \rfloor$ is the number of taps equal to the memory of the channel given by

$$\rho = 2\pi\beta_2 L_{total}/T_{ADC}^2, \tag{2.50}$$

where T_{ADC} is the sampling time of the analog-to-digital converter (ADC). In addition to the CD compensation, dynamic channel equalization is needed to compensate time-varying effects such as the state of polarization and PMD.

2.4.2.2 Multi-Input Multi-Output Equalization

After extracting and converting the signals for each polarization component into the electrical domain and compensating for time spreading caused by dispersion, the PMD effects need to be addressed. As PMD effects are time-varying, they require adaptive filters for dynamic processing. To perform adaptive equalization of PMD effects, a set of four complex-valued FIR filters is required, which together form a 2x2 MIMO matrix and perform the inverse of the Jones matrix of the dynamic channel in Eq. (2.27). The outputs of the MIMO equalizer are expressed as:

$$x_{\text{out}}[k] = \mathbf{h}_{xx}\, \mathbf{x}_{\text{in}}[k] \; + \mathbf{h}_{xy}\, \mathbf{y}_{\text{in}}[k] \tag{2.51}$$

$$y_{\text{out}}[k] = \mathbf{h}_{yx}\, \mathbf{x}_{\text{in}}[k] \; + \mathbf{h}_{yy}\, \mathbf{y}_{\text{in}}[k]\,, \tag{2.52}$$

where $\mathbf{h}_{xx}, \mathbf{h}_{xy}, \mathbf{h}_{yx}, \mathbf{h}_{yy}$ are vectors of length N representing the FIR taps, with $\mathbf{x}_{\text{in}}[k]$ and $\mathbf{y}_{\text{in}}[k]$ representing the sliding window of length $N+1$ over which the filter is convolved, expressed as

$$\mathbf{x}_{\text{in}}[k] = x_{\text{in}}(t = k' \cdot T_s, t = (k'+1) \cdot T_s, \ldots, t = (k'+N) \cdot T_s) \tag{2.53}$$

$$\mathbf{y}_{\text{in}}[k] = y_{\text{in}}(t = k' \cdot T_s, t = (k'+1) \cdot T_s, \ldots, t = (k'+N) \cdot T_s), \tag{2.54}$$

where T_s is the symbol period and $k' = k \cdot n_s - \frac{N}{2}$, where n_s is the oversampling rate of the input signals, measured with the number of samples per symbol. Following the MIMO equalizer, the output sequences are sampled at 1 sample/symbol.

A popular method for optimizing the FIR taps blindly and adaptively is called the constant modulus algorithm (CMA), which is typically used for phase modulations. Here these coefficients $\mathbf{h}_{xx}, \mathbf{h}_{xy}, \mathbf{h}_{yx}, \mathbf{h}_{yy}$ are updated, using the stochastic gradient method, which is described as

$$\mathbf{h}_{xx} \rightarrow \mathbf{h}_{xx} + \mu\varepsilon_x \mathbf{x}_{\text{out}} \cdot \mathbf{x}_{\text{in}}^* \tag{2.55}$$

$$\mathbf{h}_{xy} \rightarrow \mathbf{h}_{xy} + \mu\varepsilon_y \mathbf{y}_{\text{out}} \cdot \mathbf{x}_{\text{in}}^* \tag{2.56}$$

$$\mathbf{h}_{yx} \rightarrow \mathbf{h}_{yx} + \mu\varepsilon_x \mathbf{x}_{\text{out}} \cdot \mathbf{y}_{\text{in}}^* \tag{2.57}$$

$$\mathbf{h}_{yy} \rightarrow \mathbf{h}_{yy} + \mu\varepsilon_y \mathbf{y}_{\text{out}} \cdot \mathbf{y}_{\text{in}}^*, \tag{2.58}$$

where μ is a convergence parameter, and ε_x and ε_y are normalized error terms given by

$$\varepsilon_x = 1 - |x_{\text{out}}| \tag{2.59}$$

$$\varepsilon_y = 1 - |y_{\text{out}}|\,, \tag{2.60}$$

and assume a unit amplitude signals such as quadrature phase-shift keying (QPSK). However, for higher order M-QAM modulations, an enhancement called the radially directed equalizer (RDE) [89] is used conjointly with the CMA. It is worth mentioning that an adaptive equalizer that employs complex tap coefficients can additionally compensate for any remaining chromatic dispersion resulting from the misevaluating the amount of the dispersion in the link.

2.4.2.3 Carrier Frequency Offset Estimation

Accurate frequency estimation plays a crucial role in the coherent reception of optical signals. The presence of frequency offset causes the phase of the signal to change rapidly, making it challenging for constant phase estimation (CPE) to track the signal phase. To tackle this issue, CPE and carrier frequency offset (CFO) are treated as separate blocks in the DSP, handling two different but related tasks. A good CFO algorithm ensures that CPE works effectively by reducing the amount of phase CPE has to track, thereby improving carrier recovery efficacy. The impact of frequency mismatch and phase offset on signal detection can be expressed by the following equation:

$$x_{\text{sym}}[k] = x_{\text{MIMO}}[k] \exp(j[\phi[k] + 2\pi\Delta f k T_{sym}]), \tag{2.61}$$

Here, x_{MIMO} refers to the output samples of the MIMO system, x_{sym} represents the accurately detected symbols, Δf is the frequency offset, and $\phi[k]$ is nonlinear phase noise. The objective is to estimate the frequency mismatch Δf, which causes a continuous and linear phase drift. One popular method for CFO estimation is the Viterbi model proposed by Viterbi and Viterbi [103] for wireless transmission. The Viterbi phase estimator, shown in Fig. 2.16, is designed for M-PSK modulation and involves using a nonlinear function based

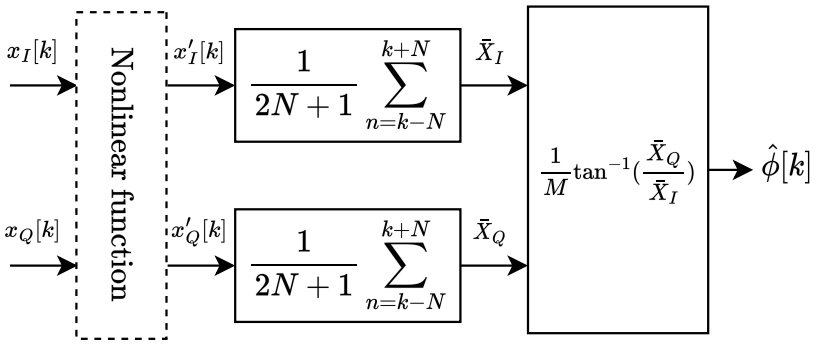


Figure 2.16: Viterbi-based phase estimator for M-PSK carriers

on the M-power of the symbol sequence. The nonlinear function transforms the symbol $x_I + jx_Q$ into $x'_I + jx'_Q$ using the rule:

$$\begin{aligned} x'_I + jx'_Q &= F(\rho_k)e^{j\Phi(\phi_x)} \\ \text{where, } \rho_k &= \sqrt{x_I^2 + x_Q^2} \\ \text{and, } \phi_x &= \tan^{-1}(x_Q/x_I) \end{aligned} \tag{2.62}$$

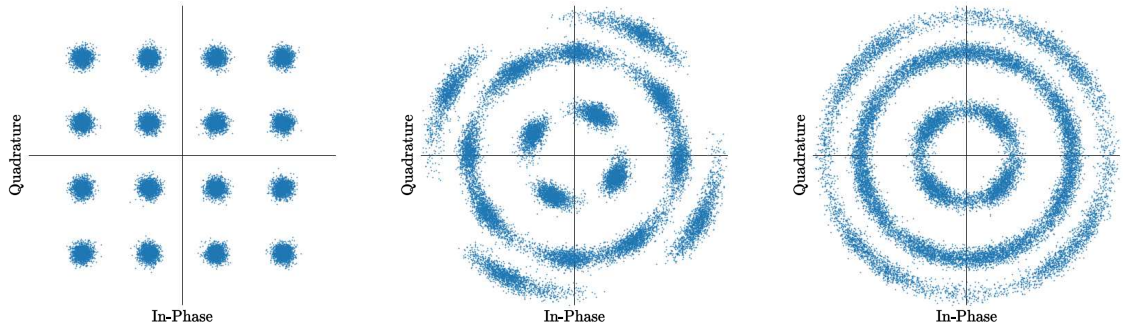
In this equation, $F(\rho) = \rho^M$ is an arbitrary nonlinear function, and $\Phi(\phi) = M\phi$ is a function effecting the phase which produces a multiple of M of the original phase. This ensures that all the points in the constellation are aligned, and the only phase variation is due to the phase drift resulting from frequency mismatch. The average value of the phase drift in the k-th symbol, $\hat{\phi}[k]$, is:

$$E\left[\hat{\phi}[k]\right] = 2\pi(\Delta f)kT_{sym}, \tag{2.63}$$

2.4.2.4 Phase Recovery

The transmission of optical signals depends on the precise frequency of the laser source. However, the light emitted by a single-frequency laser is practically not monochromatic, mainly due to technical imperfections such as excess noise from the pump source, vibrations of the laser resonator, or temperature fluctuations, as well as quantum noise, specifically associated with spontaneous emission in the gain medium. These factors result in laser phase noise (PN), which manifests as fluctuations in the optical phase. A common measure of the phase noise characteristics of a laser source is the linewidth of the laser $\Delta\nu$, which refers to the width of its optical spectrum, typically measured as a root-mean-square width. It is a measure of the spectral distribution of the laser's emitted electric field, with respect to frequency. A narrower linewidth corresponds to lower phase noise and improved performance. The Discrete-time Wiener model [72] describes the phase noise as

$$\phi_i = \phi_{i-1} + \sigma_{\text{PN}}\nu_i, \tag{2.64}$$



(a) No laser PN. (b) 100 kHz. (c) 200 kHz.

Figure 2.17: Effect of laser linewidth on phase noise and symbol constellation in 16-QAM modulation format during transmission of 24,000 symbols at a 10 GBaud rate. (a) Reference case with no laser phase noise. (b) Laser phase noise with a linewidth of 100 kHz. (c) Laser phase noise with a linewidth of 200 kHz.

where $\nu_i \sim N(0,1)$, and $\sigma^2_{\text{PN}} = 2\pi\Delta\nu T_s$ is the variance of the difference between two adjacent samples of phase noise. Here, $\Delta\nu$ corresponds to the laser linewidth, and T_s is the sample period. The power spectral density (PSD) of the complex exponential signal influenced by Wiener phase noise is known to follow a Lorentzian function, with a 3-dB linewidth equal to $\sigma^2_{\text{PN}}/(4\pi) = \Delta\nu T_s/2$. The impact of laser phase noise is therefore largely determined by the value of the product $\Delta\nu T_s$. Therefore, the lower the symbol rate, the narrower the laser linewidth should be to minimize this product and mitigate the impact of phase noise on the transmitted signal. Figure 2.17 illustrates the effects of laser phase noise on symbols modulated using 16-QAM. As the linewidth of the laser increases, the phase shift due to laser phase noise becomes larger, which causes the constellation points to merge into each other, resulting in a significant degradation in transmission quality.

The phase offset is composed of two main types of errors: phase error and phase noise. Phase error is caused by channel deterministic effects, while phase noise results from non-deterministic random effects from laser sources that are associated with the laser linewidth, as previously discussed. To compensate for these effects, a process called CPE needs to be performed. CPE is a method used to recover the phase offset θ, which is defined as follows:

$$\begin{bmatrix} x_{\text{sym}}[k] \\ y_{\text{sym}}[k] \end{bmatrix} = e^{-j\theta} \begin{bmatrix} x_{\text{MIMO}}[k] \\ y_{\text{MIMO}}[k] \end{bmatrix} \tag{2.65}$$

There have been several CPE algorithms proposed in the literature. For QPSK modulation, a potential method to remove the signal component of the phase involves using the fourth-order power of the signal, which aligns all the constellation points in one quadrant. To estimate the phase offset in a symbol block of size $N+1$, the Barycenter algorithm [15,

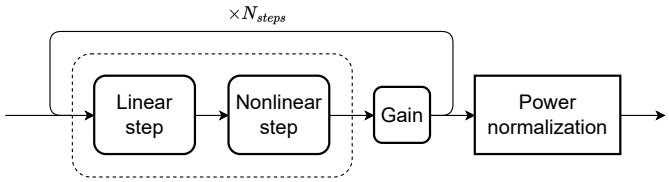


Figure 2.18: A diagram illustrating DBP using the Wiener model.

54] can be utilized. This algorithm can be modeled as follows:

$$\theta[k] = \frac{1}{N+1} \sum_{k=-N/2}^{N/2} w[k] \arg\left\{x^4_{\text{MIMO}}[k+n]\right\}. \tag{2.66}$$

Here, $w[k]$ represents a weighing function, and $\arg\{\cdot\}$ represents the argument of a complex number. A simplified form of CPE is non-blind recovery using pilot symbols, where the angle between the transmitted symbols and the pilots is calculated over a sliding window. This can be done by directly comparing the angles between the received symbols and the pre-known pilot symbols. The symbols in the window are then rotated according to the mean angle between the pilots and transmitted symbols. The output of CPE is then passed to the detection block, which assigns the equalized symbol to the nearest symbol in the constellation and demaps the symbol to its corresponding bits.

2.5 Digital Back Propagation

DBP is a popular method to effectively mitigate nonlinear impairments in optical fiber transmission systems by numerically solving the NLSE equation [16, 56, 78]. In contrast to SSFM, which simulates signal propagation in the forward direction, DBP equalizes the received signal by simulating signal propagation in the reverse direction, using fewer steps than SSFM in order to reduce the computational burden on the hardware.

The DBP block is placed before the MIMO equalizer in the DSP, which means that the input signals are subject to PMD variations in the optical fiber. However, it is important to note that the standard DBP algorithm is not designed to equalize random effects like PMD. Instead, it primarily addresses deterministic effects such as CD and nonlinear effects. In practical systems, the DBP is typically based on the propagation model in Eq. (2.31), which is averaged over the rapidly-varying SOP along the fiber due to the presence of PMD. DBP operates by dividing the optical channel into equal-length spatial segments known as "steps," denoted as δ_d. Within each step, we separately consider the dispersive and nonlinear channel effects. To tackle this, we split Eq. (2.31) into two distinct equations,

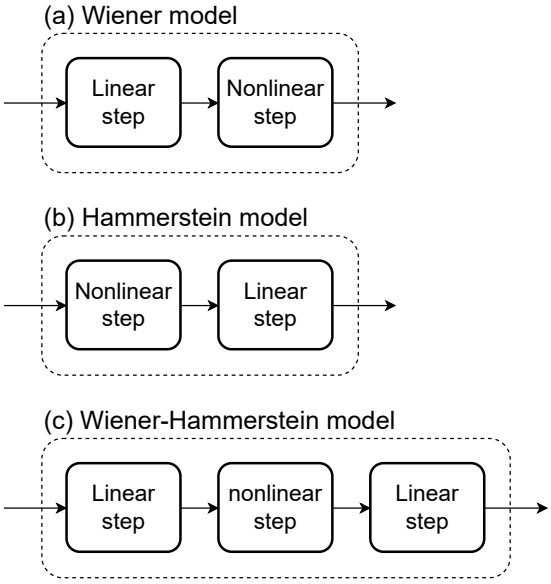


Figure 2.19: DBP order of operations in different nonlinear models [78]: (a) Wiener model, with a linear substep followed by a memoryless nonlinear substep; (b) Hammerstein model, with a memoryless nonlinear substep followed by a linear substep; (c) Wiener-Hammerstein model, consisting of a linear substep followed by a memoryless nonlinear substep, followed by a second linear substep.

one addressing linear effects and the other nonlinear effects

$$\frac{\partial \mathbf{q}}{\partial z} = \left(-j\beta_2 \frac{1}{2} \frac{\partial^2}{\partial t^2} \right) \mathbf{q}, \tag{2.67}$$

$$\frac{\partial \mathbf{q}}{\partial z} = \left(-\frac{1}{2}\alpha + j\frac{8}{9}\gamma ||\mathbf{q}||^2 \right) \mathbf{q}. \tag{2.68}$$

Upon solving the two equations, where (2.67) is solved in the frequency domain and (2.68) is relatively simpler to solve in the time domain, two partial solutions can be obtained:

$$\mathbf{q}(z+\delta_d, \omega) = \underbrace{\exp\left(-j\beta_2 \frac{1}{2}\omega^2 \delta_d \right)}_{L(\delta_d, \omega)} \mathbf{q}(z, \omega), \tag{2.69}$$

$$\mathbf{q}(z+\delta_d, t) = \underbrace{\exp\left(\frac{1}{2}\alpha\delta_d - j\frac{8}{9}\gamma ||\mathbf{q}(z,t)||^2 \delta_d \right)}_{N(\delta_d, t)} \mathbf{q}(z, t), \tag{2.70}$$

where the location z is measured relative to the receiver, with $z = 0$ corresponding to the receiver's location. The DBP is hence characterized by two sets of operators: the linear operator $L(\delta_d, \omega)$ and the nonlinear operator $N(\delta_d, t)$. The equalized signal can be obtained from the received signal by alternating between these two solutions along the length of the fiber in the backward direction (from the receiver to the transmitter).

2.6 Summary

Optical fibers made of silica have enabled transmission systems to exceed previously impossible speeds due to their low attenuation and noise, and high transmission speed. However, as demand for high-speed and bandwidth increases, limitations in optical fiber transmission systems have started to impact communications.

Some of the channel effects that impact the quality of optical signal transmission include dispersion, which spreads symbols in time and leads to inter-symbol interference (ISI) if no equalization measures are taken. Nonlinearities also become problematic, especially at higher signal powers. ASE noise generated by signal amplification needed for long-range transmission is another issue. Imperfections in components such as the laser source generate laser phase noise, which affects the quality of transmission. There are additional effects that impact the signal to varying degrees in different scenarios, whether it be high power, high speed, single-channel or WDM transmission. Researchers have proposed various solutions to mitigate these effects with varying degrees of success. Fast and efficient DSP algorithms enabled the utilization of available transmission bandwidth more effectively through coherent transmission and detection. DSP utilizes algorithms such as CD compensation, MIMO-based equalization, and CPE and CFO estimation to correct for linear channel effects. These processes allow for the use of higher-order modulations such as QPSK and M-QAM to transmit data at higher rates than ever possible. However, more advances in optical receivers are necessary to deal with nonlinearities as systems push towards achieving higher transmission speeds.

CHAPTER 3

Neural networks for equalization in optical fiber communication

Contents

3.1 Introduction

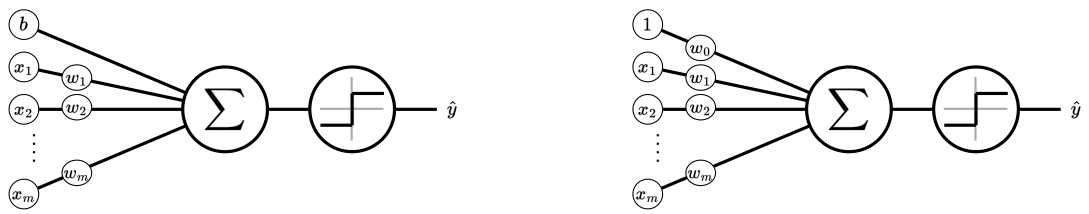
Neural networks are a type of machine learning algorithm modeled after the structure and function of the human brain. They consist of layers of interconnected "neurons" that

process information and can be trained to recognize patterns and make predictions. The origins of neural networks can be traced back to the 1940s and 1950s, when researchers first began exploring the idea of using mathematical models to simulate the behavior of neurons in the brain. Early work in this field was led by Warren McCulloch and Walter Pitts, who proposed the first mathematical model of an artificial neuron, called perceptron, in 1943. Their work laid the foundation for the development of neural networks and machine learning. The perceptron was first demonstrated by a machine implementation in 1958 by Frank Rosenblatt, a psychologist and computer scientist working at the Cornell Aeronautical Laboratory. His work in this field focused on the development of simple perceptrons, which is considered as one of the first models of a neural network, and it consists of a single layer of artificial neurons that can be trained to recognize patterns in data. The free parameters that control the output of the perceptron were adjusted manually in the earliest stages of its development. The perceptron was initially seen as a promising solution for a wide range of problems, such as image recognition, natural language processing, and even artificial intelligence. However, soon it was discovered that the perceptron algorithm was limited in its capability to solve problems that are not linearly separable, this led the researchers to develop other algorithms and architectures that can solve such problems. This realization led to a decline in interest in neural networks, and it wasn't until the late 1970s and early 1980s that the field began to see renewed growth. Some of the major breakthroughs in neural network research include the development of backpropagation algorithm in the 1980s, which greatly improved the ability of neural networks to learn from data. Also, the breakthrough of using deep neural networks with multiple fully connected layers which improved the accuracy of the predictions. This made it possible to train neural networks with multiple layers, which is known as multi-layer perceptron (MLP) and it is considered as the backbone of the fully connected neural networks. Additionally, the advent of high-performance computers and the availability of large amounts of data for training, made it possible to train large fully connected neural networks, allowing them to process and represent a wide range of information.

In the following subsections, we will present a detailed examination of various neural network architectures, including their mathematical models and underlying principles. We will also provide examples of their applications and their strengths and limitations. Additionally, we will delve deeper into the mathematical derivations and the optimization techniques used to train these architectures.

3.1.1 Mathematical model of the perceptron

The perceptron network processes input data, which is represented as a feature vector $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ with a dimensionality of d. These input values are then propagated from the input layer to the output node via edges, each carrying a weight assigned from



(a) (b)

Figure 3.1: Perceptron's mathematical model (a) with a bias node, and (b) with the bias node replaced by an input $x_0 = 1$ multiplied by a weight w_0.

the weight vector $\mathbf{w}$. In the output node, these weights are employed to compute a linear function. The sign of this calculation is subsequently used to make predictions for the binary class variable $\hat{y}$ associated with the input data, whose target output is y. To enhance prediction accuracy, it is a common practice to introduce a bias term b at the output node, which adjusts the output values. The prediction $\hat{y}$ is determined as follows [4]:

$$\hat{y} = \text{sign}\{\mathbf{w} \cdot \mathbf{x} + b\} = \text{sign}\{\sum_{j=1}^{d} w_j x_j + b\} \tag{3.1}$$

where the sign function maps a numeric value to either +1 or -1, which is suitable for classification into two distinct classes. The perceptron algorithm enables the classification of inputs into different categories by optimizing the weight vector $\mathbf{w}$ and the bias term b.

To make the derivation of training equations easier, the bias may be treated as a constant node multiplied by a weight as described in Fig. 3.1. The prediction error is represented by $E(\mathbf{w}, \mathbf{x}, y) = y - \hat{y}$, which is a value selected from $\{-2, 0, 2\}$. The aim of the perceptron algorithm is to reduce the number of errors in classifying the input data, which we describe mathematically as

$$\min_{\mathbf{w}} \; L(\mathbf{w}; (\mathbf{x}, y)) = \sum_{\{(\mathbf{x},y)\} \in \mathcal{D}} E(\mathbf{w}, \mathbf{x}, y)^2 = \sum_{\{(\mathbf{x},y)\} \in \mathcal{D}} (y - \text{sign}\{\sum_{j=1}^{d} w_j x_j\})^2, \tag{3.2}$$

where $L(\mathbf{w}; (\mathbf{x}, y))$ represents the loss function based on squared error, while $\mathcal{D}$ denotes the input-output dataset. The minimization of the expression is achieved by adjusting the weights $\mathbf{w}$ through a heuristic method known as gradient descent. This method employs a step-by-step approach to iteratively adapt the model parameters, ultimately reducing the error function and enhancing prediction accuracy.

It is important to note that the objective function specified in Eq. (3.2) is not a smooth function. To overcome this issue, the perceptron algorithm uses an approximation of the gradient of this objective function, which is intrinsically set to minimize the *perceptron criterion* for each example:

$$\nabla L = \sum_{\{(\mathbf{x},y)\} \in \mathcal{D}} (y - \hat{y}) \cdot \mathbf{x} \tag{3.3}$$

Following this, the update rule for each example is stated as

$$\mathbf{w} \leftarrow \mathbf{w} + \eta \cdot (y - \hat{y}) \cdot \mathbf{x} \tag{3.4}$$

where η is the learning rate. During the training process, the perceptron updates its weight vector $\mathbf{w}$ and bias term b using a set of correctly labeled examples using the following update rule.

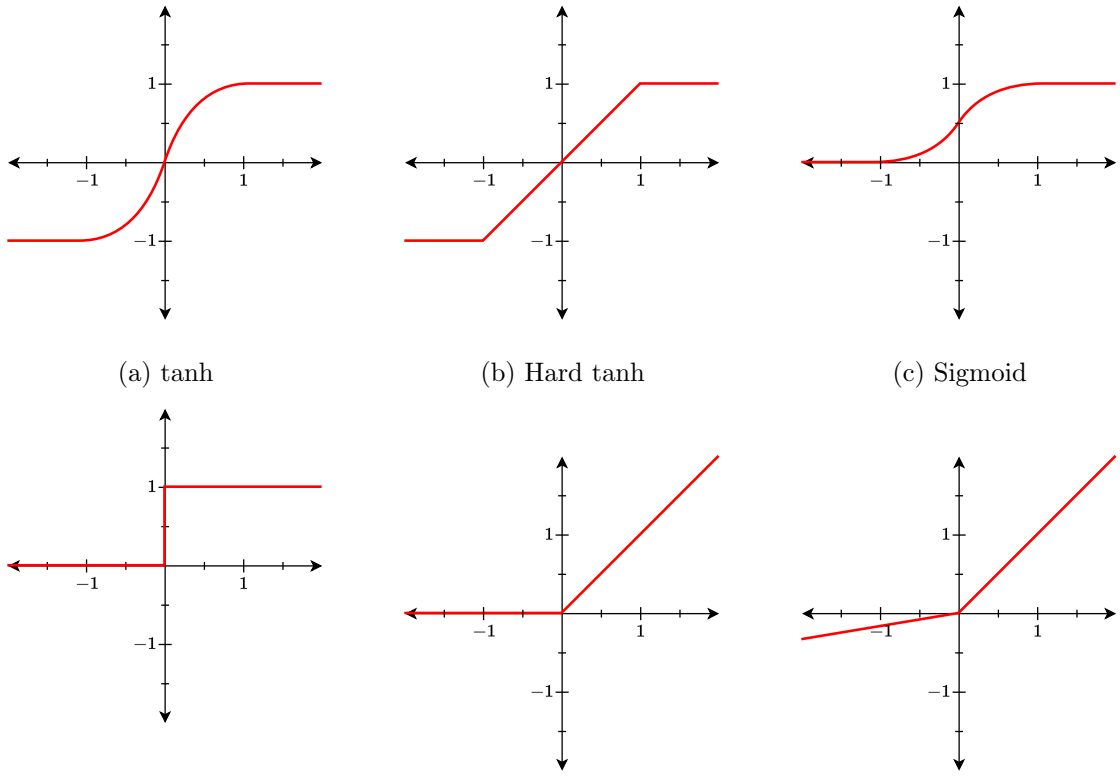
3.1.2 Activation functions

Neural networks comprise numerous perceptrons organized into discrete layers. For effective weight updates within neural networks, it is imperative that each perceptron function is differentiable to facilitate gradient calculation, which is an essential step for weight optimization using gradient-based optimization techniques. However, the 'sign' function employed in perceptrons lacks differentiability across its entire domain and can only accommodate binary decision boundaries. This limitation proves inadequate for addressing real-world problems that necessitate complex and continuous input-to-output mappings. Consequently, the 'sign' function is replaced by specialized nonlinear functions referred to as "*activation functions*".

The activation function serves as a vital element in introducing nonlinearity to the network. Without the activation function, the network collapses into a single layer, limited to performing linear transformations on the input. Hence, the activation function plays a crucial role in expanding the representational capacity of the neural network by allowing it to model a broader range of functions. Although functions like *tanh* and sigmoid were popular historically, their approximations and derivatives can impose computational overhead in hardware implementations. To address this issue, simpler activation functions such as ReLU and hard *tanh* have gained popularity in recent years. These functions are easy to compute and their derivatives are straightforward to calculate, making them more hardware-friendly. However, it's worth noting that the choice of activation function should ultimately depend on the specific requirements of the task at hand. The sigmoid function is particularly suitable for classification tasks that involve categorical data, as it maps the output to a range between 0 and 1. This allows for a straightforward binary decision on whether an input belongs to a class or not. the ReLU activation function, on the other hand, is often preferred for approximating continuous functions as it is not bounded and can more effectively handle a wider range of values.

3.1.3 The Input and Outputs of Neural Networks

The neural network training follows the supervised learning scenario, where a set of examples $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{1 \leq i \leq n}$ is provided to the neural network, each composed of a set of explanatory variables $\mathbf{x}_i \in \mathbb{R}^N$ as inputs, and a target $\mathbf{y}_i \in \mathbb{R}^M$ as outputs. The objective



(a) tanh (b) Hard tanh (c) Sigmoid

(d) Perceptron (e) ReLU (f) Leaky ReLU

Figure 3.2: Example of commonly used activation functions.

Name	Function $\sigma(x)$	Derivative $\sigma'(x)$
Sigmoid	$\frac{1}{1+e^{-x}}$	$\sigma(x)(1-\sigma(x))$
ReLU	$\max(0,x)$	$\begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$
tanh	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$	$1-\sigma^2(x)$
Hard *tanh*	$\max(\min(1,x),-1)$	$\begin{cases} 1, & \text{if } \lvert x\rvert < 1 \\ 0, & \text{if } \lvert x\rvert > 1 \end{cases}$
Leaky ReLU	$\begin{cases} x & x \geq 0 \\ ax & x < 0 \end{cases}$	$\begin{cases} 1, & \text{if } x \geq 0 \\ \alpha, & \text{if } x < 0 \end{cases}$
ELU	$\begin{cases} x & x > 0 \\ \alpha(e^x - 1) & x \leq 0 \end{cases}$	$\begin{cases} 1, & \text{if } x \geq 0 \\ \alpha e^x, & \text{if } x < 0 \end{cases}$

Table 3.1: Common activation functions and their derivatives

of the neural network is to forecast output values based on the input values provided. In this sense, the neural network's learning algorithm can be viewed as a mapping from an input dataset $\mathcal{D}$ to a decision procedure $\mathcal{H}$. The decision procedure $\mathcal{H}$ in this case is able to associate an output $\mathbf{y} \in \mathcal{Y}$ to any input $\mathbf{x} \in \mathcal{X}$. Then we have $\mathcal{H} : \mathbf{x} \in \mathcal{X} \longrightarrow \mathbf{y} \in \mathcal{Y}$ [22]. The inputs of the neural network learning can hence be either a whole data set $\mathcal{D}$, or a particular instance $\mathbf{x}$ for which we want to find the prediction y. The former represents a neural network operating in the **training phase**, and the later represents a neural network operating in the **inference phase**.

The input element $\mathbf{x}$ can belong to one of the following categories:

- *Vectorial*: where the input data is taken from a relational database and the descriptors are treated as dimensions of an input space, typically a vectorial space. This allows for the definition of a distance metric, which transforms the vectorial space into a normed vectorial space. This type of data representation is particularly useful for various mathematical techniques that have been developed for such spaces.

- *Non Vectorial*: where the number of elements within a given example is not fixed. This type of data is commonly found in areas such as natural language processing, where machine learning algorithms are tasked with analyzing documents with an undefined number of words or pages, or in image recognition, where the algorithm is expected to provide a descriptive text of undefined length based on the input image. Due to the nature of this type of data, it can be challenging to define appropriate distances, however, specialized techniques have been developed to accommodate this.

- *Structured data*: refers to data that can be analyzed based on its internal structure. It can be broken down into further categories, such as: sequential data (ordered elements), spatial data (dependencies between adjacent elements), graphical data (described as graphs), and relational data (complex structures like DNA snippets or texts).

3.2 Neural Networks Architectures

Multiple neural network architectures exist to perform the decision procedure $\mathcal{H} : \mathbf{x} \in \mathcal{X} \longrightarrow \mathbf{y} \in \mathcal{Y}$, with each architecture more suited to specific tasks. To model a wide range of functions, a neural network must possess two important elements. Firstly, it should have adjustable parameters $\mathbf{w} = \{w_i\}_{1 \leq i \leq n}$ that are optimized to minimize training error. Secondly, a neural network must contain nonlinear elements, commonly known as activation functions, to enable the so-called "expressive power" of the neural network. The neuron is the fundamental building block of a neural network, incorporating both adjustable parameters and nonlinear activation functions. Neurons are inspired by perceptrons, but they

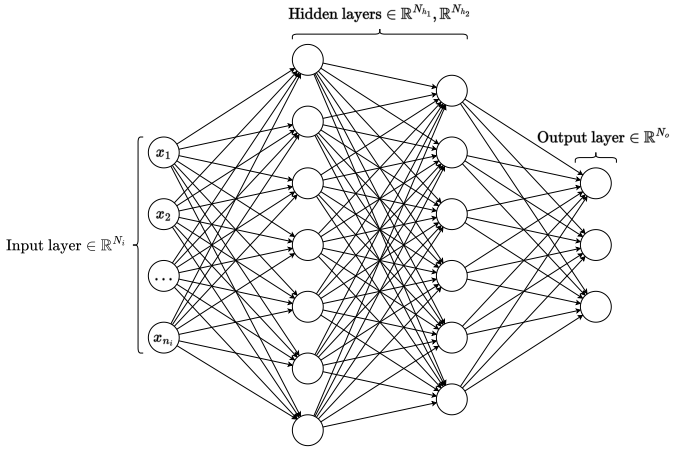


Figure 3.3: A diagram illustrating the structure of a Multilayer Perceptron (MLP) consisting of an input layer with N_i units, two hidden layers with N_{h_1} and N_{h_2} units, respectively, and an output layer with N_o units.

use smooth continuous activation functions instead of the unit step activation used by perceptrons. Each neuron comprises a weight vector $\mathbf{w}$ and bias b, which are adjusted during the training process to minimize the training error. The choice of activation function and hyperparameters such as the depth and width of the neural network is left to the designer during the initialization phase. The parameters $\mathbf{w}$, on the other hand, are usually initialized randomly and then optimized during training using the gradient descent algorithm, which is discussed in later sections.

Here, we discuss three common architectures: multi-layer perceptrons, convolutional neural networks, and recurrent neural networks.

3.2.1 Multi-layer perceptrons

Multi-layer perceptrons (MLPs), are named after the perceptron. MLPs are composed of multiple layers of interconnected neurons, where each neuron in one layer is connected to every neuron in the next layer. The flow of information in the network is unidirectional, moving from the input layer to the output layer through consequent hidden layers. Fig. 3.3 provides an illustration of the fully-connected neural network. Each layer l in the network, with d_l hidden units, takes input from the previous layer (denoted by $l-1$), and applies the mapping $\Psi_l : \mathbb{R}^{N_{h_{l-1}}} \mapsto \mathbb{R}^{N_{h_l}}$ to the input $\mathbf{x}_{l-1}$. This transforms the input into the output as shown below:

$$\begin{aligned} \mathbf{x}_l &= \Psi_l(\mathbf{x}_{l-1}) \\ &= \sigma_l(\mathbf{W}_l^T \mathbf{x}_{l-1} + \mathbf{b}_l), \end{aligned} \tag{3.5}$$

where $\mathbf{W}_l \in \mathbb{R}^{d_{l-1} \times d_l}$ is the weight matrix, $\mathbf{b}_l$ is the bias vector, and $\sigma_l(\cdot)$ is the

activation function, applied element-wise to its vector argument.

3.2.2 Convolutional neural networks

convolutional neural networks (CNNs) were first introduced by Yann LeCun in the late 1980s as a specialized type of neural network designed to process grid-like data. Originally developed for image recognition, CNNs have been applied to a broad range of tasks, including natural language processing and speech recognition. CNNs are particularly well-suited for processing sequential data using a series of convolutional layers to extract features, followed by fully-connected layers for classification. With their ability to learn features automatically from raw data, CNNs are especially effective at identifying spatial patterns and hierarchies of features in image data, enabling them to achieve state-of-the-art performance on many image classification benchmarks.

The mathematical function for a single convolutional layer in the CNN is expressed as

$$\mathbf{x}_{ij}^{(l+1)} = \sigma\left(\sum_{k=1}^{C^{(l)}}\sum_{p=1}^{F_l}\sum_{q=1}^{F_h} W_{kpq}\mathbf{x}_{(i+p-1)(j+q-1),k}^{(l)} + b_j\right), \tag{3.6}$$

where $\mathbf{x}_{ij}^{(l+1)}$ is the output of the j-th feature map at location (i, j) in the l-th layer, $C^{(l)}$ is the number of feature maps in the (l)-th layer, F_l and F_h are the filter dimensions of the convolutional layer, $W_{kpq}^{(l+1)}$ is the weight of the connection between the k-th feature map in the (l)-th layer and the j-th feature map in the l-th layer at the offset (p, q), and $b_j^{(l+1)}$ is the bias term for the j-th feature map in the l-th layer. The activation function $\sigma(\cdot)$ is applied element-wise to the output of each neuron. Although the output of a convolutional operation is a 2D array with a single feature map, convolutional layers have an additional parameter called the number of channels, which controls the number of feature maps produced by that layer and enables the network to learn multiple representations of the input data.

3.2.3 Recurrent neural networks

While fully-connected and convolutional neural networks are suitable for tasks with fixed input and output dimensions, they are not well-suited for tasks that involve sequential data that has time dependencies. Both types of neural networks consider each input to be independent of all previous inputs, which makes it difficult to model temporal dependencies between inputs. To address this issue, researchers developed recurrent neural networks (RNNs), which have connections between their hidden layers that allow them to maintain an internal state and process sequential data. RNNs include connections that feed information back into the network, allowing them to process sequential data with temporal dependencies. At each step in the sequence, the RNN updates its hidden state based on the

current input and the previous hidden state. The function of a simple recurrent network (SRN), also known as the Elman model [28], with a hidden state h_t at time step t, taking input x_t and outputting y_t can be stated as follows

$$h_t = \sigma_h(W_h h_{t-1} + W_x x_t + b_h) \tag{3.7}$$

$$y_t = \sigma_y(W_y h_t + b_y) \tag{3.8}$$

Where σ_h and σ_y are activation functions, W_h and W_x are weight matrices for the hidden state and input, respectively, W_y is a weight matrix for the output, and b_h and b_y are bias vectors. The output of the network at time step t depends not only on the input x_t, but also on the current hidden state h_t, which is updated based on the previous hidden state h_{t-1} and the current input x_t.

One of the key features of RNNs is their ability to handle variable-length inputs, as the internal state of the network can be adjusted to match the length of the input sequence. This makes them well-suited for tasks such as speech recognition or natural language processing, where the length of the input varies from one example to the next.

3.2.4 LSTM Networks

One limitation of standard RNNs is that they can struggle with learning long-term dependencies due to the vanishing gradient problem, where the gradient of the loss function becomes very small for long sequences, making it difficult for the network to update its weights. To address this issue, researchers have developed variants of RNNs such as long short-term memory (LSTM) and gated recurrent units (GRU), which use different types of gating mechanisms to selectively update the internal state and avoid the vanishing gradient problem. In an LSTM network, information is stored and retrieved through a set of memory cells and associated gates.

The LSTM network can be defined as follows. Let x_t be the input at time step t, and let h_t and c_t be the hidden state and cell state, respectively, at time step t. The LSTM network consists of the following equations:

$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i) \tag{3.9}$$

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f) \tag{3.10}$$

$$c'_t = \tanh(W_{xc'} x_t + W_{hc'} h_{t-1} + W_{cc'} c_{t-1} + b_g) \tag{3.11}$$

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + W_{co} c_{t-1} + b_o) \tag{3.12}$$

$$c_t = f_t c_{t-1} + i_t c'_t \tag{3.13}$$

$$h_t = o_t \tanh(c_t) \tag{3.14}$$

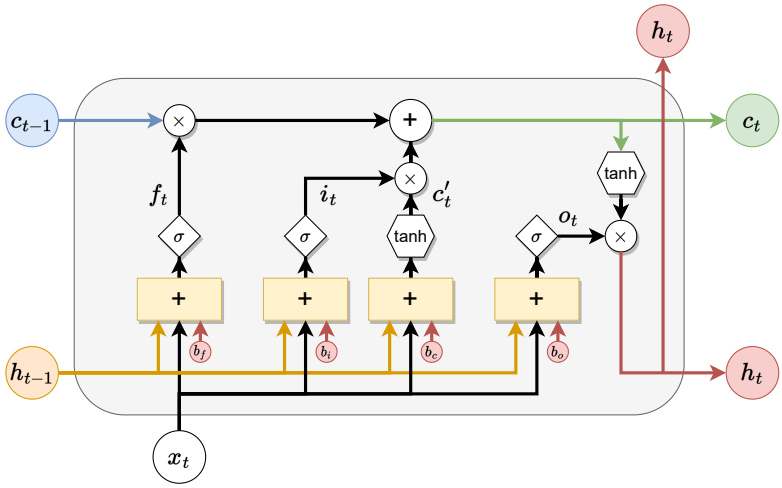


Figure 3.4: Diagram illustrating the detailed structure of the LSTM building block. The LSTM cell contains four interacting neural network layers that control the flow of information: the input gate, the forget gate, the output gate, and the cell memory state. The gates are controlled by sigmoid activation functions and the cell memory state is updated by a hyperbolic tangent activation function. This architecture allows the LSTM to selectively remember or forget information from previous time steps, while maintaining a stable gradient during training.

where σ is the sigmoid function, and i_t, f_t, c'_t, and o_t are the input state, forget state, candidate cell state, and output state, respectively. The weights W_{xi}, W_{hi}, W_{ci}, W_{xf}, W_{hf}, W_{cf}, $W_{xc'}$, $W_{hc'}$, $W_{cc'}$, W_{xo}, W_{ho}, W_{co}, and biases b_i, b_f, b_c, b_o of the neural network, are learned during training. In this model, the input gate i_t controls the flow of new information into the cell state c_t, the forget gate f_t controls the amount of old information that is retained in the cell state, and the output gate o_t controls the flow of information from the cell state to the output h_t. The cell gate c'_t is the candidate new memory content.

3.2.5 Loss functions and training algorithms

The objective of training a neural network is to optimize its parameters such that it minimizes a specified loss function. The loss function, denoted as L, is a mathematical function that measures the difference between the predicted output of the neural network and the true output. The goal is to minimize the loss function over the training data, so that the neural network can accurately predict the output for new inputs.

The choice of loss function depends on the type of task at hand. For regression tasks, where the goal is to predict a continuous output y, the mean-squared error (MSE) loss function is commonly used. The MSE is defined as:

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where n is the number of data points, y_i is the true output, and $\hat{y}_i$ is the predicted

output of the neural network.

For classification tasks, where the goal is to predict a discrete output, different loss functions are used depending on the nature of the problem. If the output classes are mutually exclusive, meaning that each input can belong to only one of several classes, then the categorical cross-entropy loss function is commonly used. The categorical cross-entropy is defined as:

$$L_{CE} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{C} y_{i,j}\log(\hat{y}_{i,j})$$

where C is the number of classes, $y_{i,j}$ is the true label of the i-th data point for the j-th class, and $\hat{y}_{i,j}$ is the predicted probability of the i-th data point for the j-th class.

On the other hand, if the output classes are independent of each other, such as in binary classification, then the binary cross-entropy loss function is used. The binary cross-entropy is defined as:

$$L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n} y_i\log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)$$

where $y_i, \hat{y}_i \in [0,1]$ are the true label and the predicted probability of the i-th data point, respectively.

3.2.5.1 Gradient descent

Gradient descent is a widely used optimization algorithm in neural network training. It is a first-order optimization method that iteratively updates the weights of the neural network to minimize the loss function L. The gradient descent algorithm works by computing the gradient of the loss function with respect to the weights of the neural network. The gradient represents the direction of the steepest ascent of the loss function. The weights are then updated by subtracting a small fraction, denoted as η, of the gradient (referred to as the learning rate) from the current weights, which are typically small positive values. The update rule for the weights of a neural network using gradient descent is as follows:

$$w_i \leftarrow w_i - \eta\frac{\partial L}{\partial w_i}, \tag{3.15}$$

where w_i is the i-th weight of the neural network. The partial derivative of the loss function with respect to the weights, $\frac{\partial L}{\partial w_i}$, is computed using the chain rule. This requires computing the partial derivative of the output of the neural network with respect to each weight, and then using those partial derivatives to compute the partial derivative of the loss function with respect to each weight. The gradient descent algorithm iteratively updates the weights of the neural network using the update rule above until the loss function reaches a minimum or a stopping criterion is met. The stopping criterion can be a maximum

number of iterations, a minimum value of the loss function, or a combination of both. Gradient descent has several variants, including stochastic gradient descent, batch gradient descent, and mini-batch gradient descent. In stochastic gradient descent, the gradient is computed using only one training example at a time, while in batch gradient descent, the gradient is computed using all training examples at once. Mini-batch gradient descent is a compromise between the two, where the gradient is computed using a small subset of the training examples.

3.2.5.2 Stochastic gradient descent

The stochastic gradient descent algorithm updates the parameter vector in the following way:

$$\mathbf{w} = \mathbf{w} - \eta \frac{1}{n} \sum_{i=1}^{n} \nabla_{\mathbf{w}} L(\mathbf{w}; x_i, y_i). \tag{3.16}$$

The algorithm randomly selects a batch of training examples and updates the parameter vector based on the gradient computed using only those examples. This allows the algorithm to make faster progress towards the solution compared to batch gradient descent.

Algorithm 1 Stochastic Gradient Descent Algorithm

Require: Training dataset $(\mathbf{x}_i, y_i)_{i=1}^{n}$, learning rate η, batch size m, maximum number of epochs $T_{\max}$.

Ensure: Learned parameters $\mathbf{w}$.

1: Initialize parameters $\mathbf{w}$ randomly.
2: **for** $t = 1$ to $T_{\max}$ **do**
3: Shuffle the training dataset.
4: **for** $i = 1$ to $\frac{n}{m}$ **do**
5: Randomly select a mini-batch $\mathcal{B}_i = (\mathbf{x}_j, y_j)_{j=1}^{m}$ from the training dataset.
6: Compute the gradient estimate $\mathbf{g}_i$ using the mini-batch:

$$\mathbf{g}_i = \frac{1}{m} \sum \nabla \mathbf{w} L(\mathbf{w}; \mathbf{x}_j, y_j) \quad \text{, for } (\mathbf{x}_j, y_j) \in \mathcal{B}_i.$$

7: Update the parameters using the gradient estimate:

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \mathbf{g}_i.$$

8: **end for**
9: **end for**

Here, the algorithm randomly selects mini-batches of size m from the training dataset, computes a gradient estimate based on the mini-batch, and updates the parameters using

the gradient estimate and a learning rate η. The algorithm repeats this process for a maximum number of epochs $T_{\max}$.

3.3 Neural network-based equalization in fiber-optic communications

In previous chapters, we discussed the interplay of five effects in a dual-polarization optical fiber transmission system: chromatic dispersion, polarization-mode dispersion, Kerr nonlinearity, amplified spontaneous emission (ASE), and component imperfections. The equalization at the receiver is performed by the DSP, which compensates for the deterministic linear channel effects and estimates the randomly varying effects such as PMD. However, the nonlinearity resulting from Kerr effect is often disregarded due to its high complexity, which limits the system performance at the high power regime.

Given that fiber nonlinear effects are distributed along the fiber's length, mitigating these nonlinear effects necessitates concurrent compensation for chromatic dispersion. Hence, the digital back-propagation (DBP) method remains a preferred choice as it effectively addresses both aspects incrementally. However, Many studies such as [27, 45, 58, 64, 95, 101], have highlighted the high complexity associated with DBP, which primarily arises from the substantial number of FFTs and IFFTs required to execute DBP. The complexity of NLC has ignited extensive research into developing techniques and DSP algorithms for addressing this challenge [18, 96]. Volterra series-based equalization [41], which precisely models and compensates for nonlinear effects introduced by the fiber channel, has also been introduced.

Neural network-based equalization is a promising solution for overcoming the challenges posed by the complex interplay of different effects in optical fiber transmission systems. Neural networks can model complex functions and estimate the inverse of the channel to perform channel equalization. However, utilizing neural networks for channel equalization presents several challenges. Firstly, neural network-based equalizers require careful and specialized customization to adapt to the unique nature of the problem. This is particularly challenging considering that state-of-the-art transponders, which utilize forward error correction (FEC), require a minimum bit-error-ratio (BER) of $10^{-2} - 10^{-3}$, a performance that far exceeds what is considered excellent for typical neural task, *i.e.* , image recognition. Secondly, the complexity of the neural network must be carefully managed to ensure efficient and implementable performance in real-world systems operating at high data rates. To address these challenges, neural networks must demonstrate higher efficiency in both learning and inference complexity compared to other fields [37].

Another aspect of neural networks that needs to be addressed is that neural network models need to be retrained when faced with dynamic effects such as randomly varying

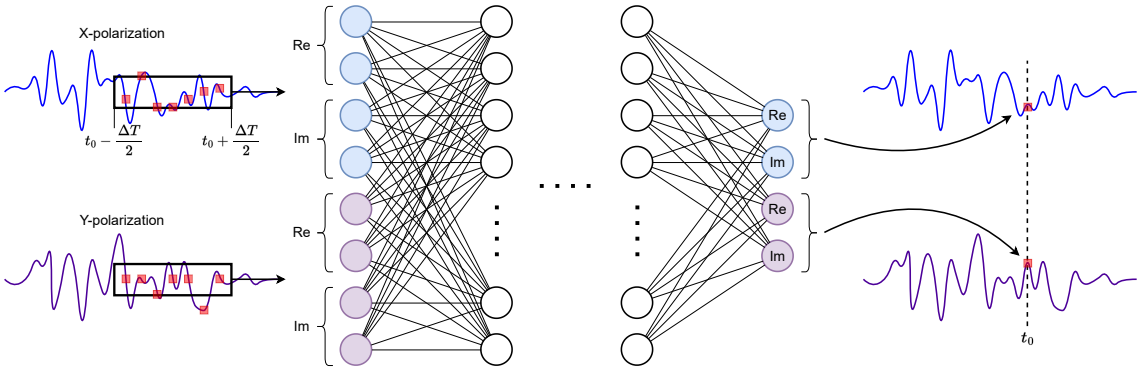


Figure 3.5: A schematic diagram illustrating an example of a multi-layer perceptron (MLP) equalizer that performs many-to-one equalization. The equalizer takes multiple symbols as input and outputs a single symbol that is extracted from both polarizations. Specifically, the input of the neural network is a vector formed by concatenating four signals, which are the real and imaginary parts of the X- and Y-polarizations. This approach allows the neural network to process dependencies between different signals, which enhances its ability to perform accurate equalization.

polarization mode dispersion (PMD) and laser noise. Retraining a neural network can be computationally expensive, which highlights the importance of combining the neural network equalizer with the receiver's DSP to mitigate the impact of rapidly changing effects such as randomly varying PMD and laser noise. To avoid the need of retraining the model for each instance of dynamic effects, the neural network can directly receive its input from the output of the DSP, which has already undergone equalization to accommodate these dynamic effects. Alternatively, the neural network can be placed behind the DSP, and joint training with the DSP can be performed [31]. Joint training allows the neural network to learn from the DSP processing, which enables a more efficient training and achieves better performance. In both cases, however, the input to the neural network entails temporal dependencies resulting from the impact of the coupling between the nonlinearity and channel dispersive effects. As a result, many-to-one equalization (such as the one shown in Fig. 3.5) is favored over one-to-one equalization even if the neural network is placed after the DSP.

In the following we provide a review of the recent state-of-the-art neural network-based equalizers in the literature. Specifically, we highlight two essential approaches: model-agnostic approaches and model-based approaches (also referred to as model-informed approaches). We will discuss the advantages and limitations of each approach and provide insights into the application of these approaches in various communication systems. Moreover, in Chapter 4 of this dissertation, we focus on utilizing the model-based approach to equalization in dispersion-managed (DM) systems.

3.3.1 Performance metrics

To assess the effectiveness of channel equalization, various metrics can be used, such as the error vector magnitude (EVM), effective signal-to-noise ratio ($\mathrm{SNR}_{\mathrm{eff}}$), and the Q-factor related to BER. It is worth noting that both neural network-based equalizers and conventional methods can be evaluated using these metrics.

The EVM measures the average deviation between the ideal received signal x_n and the actual received signal y_n. It is defined as follows:

$$\mathrm{EVM} = \left[\frac{\sum_{n=1}^{N} |y_n - x_n|^2}{\sum_{n=1}^{N} |x_n|^2} \right]^{\frac{1}{2}} \tag{3.17}$$

However, it is important to note that the $\mathrm{SNR}_{\mathrm{eff}}$ provides a reliable estimation of the SNR under the assumption that the error deviation follows a Gaussian distribution. The $\mathrm{SNR}_{\mathrm{eff}}$, representing the ratio of signal power to distortion power, is computed as follows:

$$\mathrm{SNR}_{\mathrm{eff}} = \frac{1}{\mathrm{EVM}^2} \tag{3.18}$$

The Q-factor is another metric for evaluating the quality of transmission and is computed using the BER. It is defined as follows:

$$Q\text{-factor} = \sqrt{2}\, \mathrm{erfc}^{-1}(2\ \mathrm{BER}) \tag{3.19}$$

where $\mathrm{erfc}(\cdot)$ is the complementary error function. It is important to mention that the BER can only be linked with SNR assuming that the error deviation follows a Gaussian distribution. In this case, all four metrics described above are interrelated and can be obtained from each other. Literature provides extensive information on the BER for various modulation schemes and signal constellations in digital communication systems [52].

3.3.2 Model-agnostic approaches

In model-agnostic approaches, designing the neural network typically involves a trial-and-error process. Users experiment with various types and hyperparameters to identify the most performing option, the one with the least complexity, or a suitable trade-off between the two. In this context, users enjoy greater flexibility and autonomy, allowing them to choose the neural network's type and structure to align with their specific needs and preferences. Notably, the neural network structure in these approaches does not directly replicate the physical model of the phenomena it simulates.

In [17], a fully connected neural network comprising two layers, each with 50 or 100 neurons, is introduced to address nonlinear effects in optical fibers. This neural network was integrated into the DSP chain, and its performance was evaluated at two positions within the chain: after multiple input multiple output (MIMO) equalization and after CPE, just before the symbol decoder.

The proposed system model operated at 32 GBaud, employing DP-16-QAM modulation. In each simulation setup, an ITU-T G.652 standard single-mode fiber (SMF) link with varying lengths $L \in 100, 200, 300, 400$ km was considered. Notably, the neural networks employed in all scenarios were initialized randomly, meaning they possessed no prior knowledge of the channel parameters. The architecture of the fully connected neural network utilized in this study is illustrated in Figure 3.7. The authors observed a Q-factor improvement ranging from 0.5 dB to 1 dB when employing the neural network as opposed to linear equalization. This gain was validated through both simulation and experimental results. Figure 3.8 depicts the performance gain achieved by the neural network when placed after CPE.

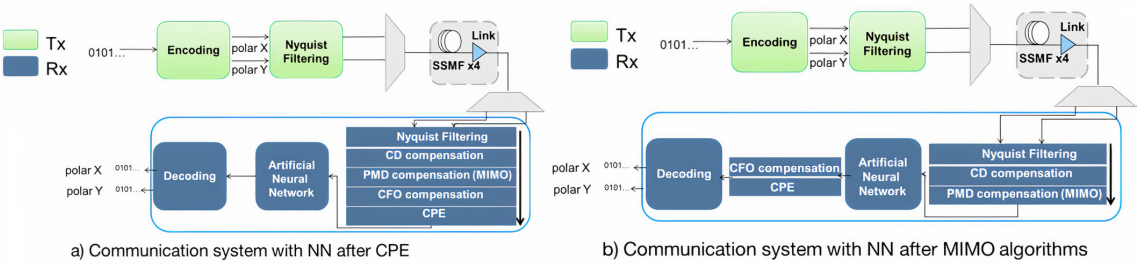


Figure 3.6: Simulated systems with neural networks (NNs) for nonlinear effects mitigation as shown in [17] for two cases: after CPE (a) and after MIMO (b).

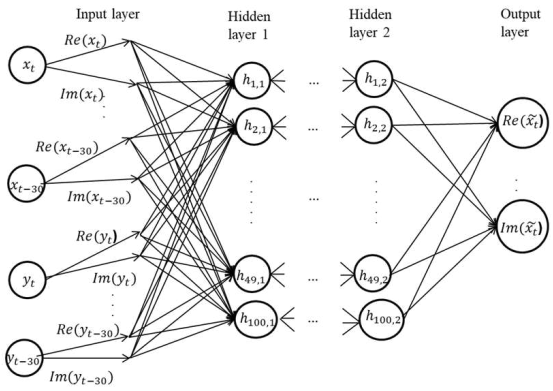


Figure 3.7: Artificial neural network design as shown in [17].

Considering the extensive range of hyperparameters that require optimization, more effective methods such as Bayesian hyperparameter optimization [84] can be used for fine-tuning neural network hyperparameters. This method has been applied in [36] to determine the hyperparameters for a novel complex-valued neural network design operating in both standard SMF and large effective-area fiber (LEAF) links, particularly in highly nonlinear regimes. Numerical and experimental tests were conducted, with results from DP-64QAM 32 GBd single-channel optical signal test cases indicating that this approach consistently outperforms standard DSP and conventional neural networks in both experimental and

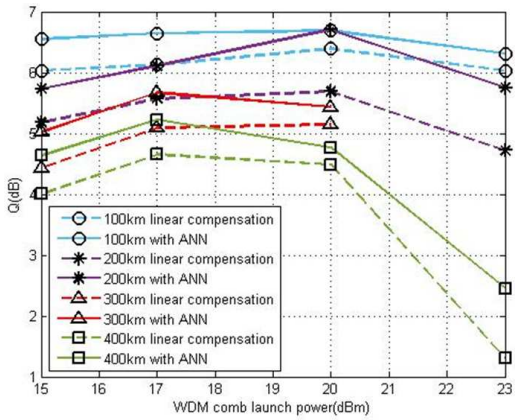


Figure 3.8: Simulated Q-factor vs. span input power of the WDM multiplex with experimental data, when the neural network is located after CPE as shown in [17].

numerical evaluations.

The studies in [18, 25, 32, 106] offer valuable insights into the performance comparisons among various model-agnostic NN models. However, the study presented in [39] stands out as a unique contribution, providing invaluable insights into the trade-offs between complexity and performance across different model-agnostic neural networks.

In [109], the authors introduced a fully-connected neural network, to learn from received data and compensate for nonlinear impairments without prior knowledge of the transmission system. Their system-agnostic neural network algorithm demonstrated approximately 0.6 dB in Q-factor improvement following signal transmission spanning a distance of 2800 km over standard SMF at a rate of 32 Gbaud.

In [24], the authors introduced the utilization of LSTM architectures for compensating fiber nonlinearities in digital coherent systems. They conducted numerical simulations involving both C-band and O-band transmission systems, considering single-channel and multi-channel 16-QAM modulation formats with polarization multiplexing. the authors conducted transmission simulations in both single-channel and multi-channel scenarios to evaluate the performance of a bi-directional long short-term memory (Bi-LSTM) neural network for compensating fiber nonlinearities in optical communication systems. In single-channel transmission, they utilized various equalization techniques, including FDE, FDE followed by Bi-LSTM, and DBP with varying numbers of steps per span. They compared performance at two different wavelengths, 1310 nm and 1550 nm. The results showed that Bi-LSTM significantly improved BER compared to linear equalization, especially in the C-band, where it outperformed DBP with 2 steps per span. In multi-channel transmission, Bi-LSTM was compared to FDE and DBP. Despite focusing only on the information from the central channel, Bi-LSTM demonstrated superiority over DBP, which only addressed intra-channel effects. This performance improvement was particularly pronounced in the C-band. The authors also analyzed the impact of several internal properties of the Bi-

LSTM network, such as the number of hidden units and the word length, on classification performance. They found that increasing the number of hidden units improved BER performance, and a word length exceeding 20 symbols was optimal. Additionally, the word length's significance related to the channel memory, determined by accumulated dispersion, was highlighted. In the C-band, where channel memory increased due to higher dispersion, the optimal word length surpassed 50 symbols.

In the paper [61, 62], the authors present an innovative approach where they implement an optical fiber communication system as an end-to-end deep neural network. This system encompasses the complete transceiver, including the transmitter, channel model, and receiver, enabling a joint optimization process. The study focuses on intensity modulation/direct detection (IM/DD) systems and demonstrates the capability to achieve BERs below the 6.7% hard-decision forward error correction (HD-FEC) threshold. NN techniques are applied to discover transmitter and receiver configurations that minimize the symbol error rate. The authors' proposed training method is validated through simulations, ensuring the development of robust and flexible transceivers capable of reliably transmitting signals over a wide range of link dispersions. Notably, the experimental results align with the simulation findings. Achieving information rates of 42 Gb/s below the HD-FEC threshold at distances exceeding 40 km, the end-to-end deep learning approach surpasses conventional IM/DD solutions using PAM2/PAM4 modulation and receiver equalization. The study also introduces a training method that enhances the system's robustness to variations in the link distance, offering flexibility in real-world implementations. The research signifies a pioneering step toward the utilization of end-to-end deep learning for optimizing optical fiber communication systems, with the potential for broader applications beyond IM/DD systems.

The paper [25] compares three bi-directional RNN models: Bi-LSTM, bi-GRU, and bi-Vanilla-RNN, as equalization techniques for mitigating fiber nonlinearities in digital coherent systems carrying polarization-multiplexed 16-QAM and 32-QAM signals. The study reveals that all three bi-RNN models significantly enhance BER performance when compared to linear equalization in a transmission experiment spanning 960 km. Among these models, bi-Vanilla-RNN stands out as it offers competitive performance while maintaining the lowest complexity. These models prove particularly effective in dispersion unmanaged systems and consistently outperform Volterra nonlinear equalizers. Furthermore, the research demonstrates that bi-Vanilla-RNN maintains its superiority even when dealing with more complex modulation formats and higher baud rates. The complexity analysis highlights that bi-Vanilla-RNN is notably less complex than the Volterra nonlinear equalizer, particularly when decoding multiple symbols simultaneously.

Applications of NNs in optical fiber communication extend beyond direct symbol equalization. In [59], the authors propose a novel receiver for nonlinear frequency division

multiplexed (NFDM) optical systems, utilizing neural networks (NNs) to adaptively detect NFDM signals amid losses and noise. Traditional methods, like the nonlinear Fourier transform (NFT) receiver, assume ideal conditions and struggle with real-world distortions. The NN receiver excels at handling these distortions, enabling reliable NFDM transmission over long distances. The study also investigates the impact of various hyperparameters on the NN receiver's performance. This work presents an improved alternative to traditional NFDM receivers for practical optical communication systems.

In [79], the authors introduce a novel approach called Co-LSTM (center-oriented long short-term memory network) for mitigating fiber nonlinearity in coherent optical communication systems. This method aims to overcome the high computational complexity associated with neural network-based equalization schemes. The authors conduct experiments using a ten-channel wavelength division multiplexing (WDM) transmission over 1600 km of standard SMF with 64 Gbaud polarization-division-multiplexed 16-QAM signals. Co-LSTM demonstrates a 0.51 dB improvement in Q^2-factor, which is comparable to the performance of DBP, while its computational complexity is significantly lower. Co-LSTM's complexity remains nearly constant regardless of the transmission distance, providing a distinct advantage over DBP.

Finally, to comprehensively address the issue of the performance-complexity tradeoff inherent in various model-agnostic neural network architectures, the authors in [39] have provided insights into the performance and complexity of several neural network types. These include CNN combined with biLSTM layers (CNN+biLSTM), CNN with MLP (CNN+MLP), biLSTM networks, three-layer perceptrons (MLP), and echo state networks (ESN). This extensive analysis encompasses both numerical simulations and practical experimental optical fiber transmission setups. For each setup, they optimized the hyperparameters of these networks using Bayesian optimization. Computational complexity expressions in terms of real multiplications per symbol were measured for each neural network type. One of their key findings is the remarkable performance of the CNN+biLSTM architecture, which outperforms other counterparts when computational complexity is not a constraint. This architecture delivers a 4.38 dB improvement in the Q-factor compared to conventional DSP. In scenarios where computational resources are limited, simpler architectures like the MLP prove to be the optimal choice.

This study, however, sheds light on a fundamental drawback of model-agnostic neural networks: their inherent complexity. While it does not provide direct comparisons with model-based approaches, it does offer a relevant comparison with DBP (with equivalent complexity to learned digital back-propagation (LDBP), as explained in later chapters). All neural networks discussed in the paper fall under the model-agnostic approach, and their complexity is shown to be orders of magnitude higher than that of DBP. For instance, DBP achieves a complexity of $10^{3.33}$ RMpS, whereas the complexity for CNN+biLSTM,

CNN+MLP, ESN, biLSTM, and MLP stands at $10^{7.52}$, $10^{6.89}$, $10^{4.93}$, $10^{7.23}$, and $10^{5.09}$ RMpS, respectively.

In the simulation results, DBP ranked as the third best performing method, offering 4.12 improvement in Q-factor compared to linear equalization (LE). This performance placed it behind only CNN+biLSTM and biLSTM, which achieved Q-factor improvements of 4.38 and 4.33 over LE, respectively. However, the modest improvement relative to DBP does not justify the significantly greater complexity associated with the other methods.

In experiments, however, since all the neural network-based models are inherently trainable, and the signals obtained from real-life experimental scenarios are influenced by various distortions and channel imperfections, these neural network methods displayed the capacity to generalize and adapt to these real-world complexities, ultimately outperforming the DBP. We hypothesize that enabling the DBP with the capability for training, similar to other neural network methods, could help it acquire the capacity to learn and adapt to real-life distortions. This, in turn, could potentially lead to improved performance in experimental setups while maintaining its low complexity. The LDBP effectively embodies this concept by adopting a model-based approach, which entails parameterizing a neural network with the DBP. This strategy enables the network to undergo training, effectively equipping it with the ability to capture and adapt to additional factors influencing the signal.

3.3.3 Model-based neural network equalizers

In contrast to the arbitrary design of model-agnostic neural networks, where design elements like architecture (shape, type, width, depth), and activation functions are often determined through a trial-and-error approach based on user experience, the model-based approach to neural network design alleviates the challenge of hyperparameter testing and selection. This is accomplished by deriving the neural network structure from the physical model of the phenomenon it aims to simulate.

One of the pioneering instances of the model-based approach can be found in the work of Sidelnikov et al. in [97]. In this work, they introduced a deep CNN (DCNN) characterized by alternating layers, roughly resembling the computational graph of SSFM. The activation function employed in their neural network is based on the enhanced SSFM approach. This function incorporates the nonlinear interactions between the considered symbols within the channel of interest, and its neighboring symbols from the surrounding channels.

Another prominent approach in the model-based neural network category is the LDBP proposed by Häger et al. in [12, 47–50]. The LDBP approach is based on the computational graph of the DBP, which consists of alternating between linear and nonlinear operators. The neural network's computational graph consists of a sequence of linear layers and non-

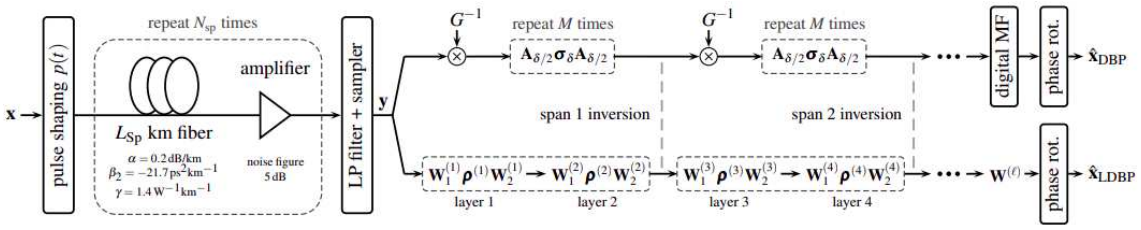


Figure 3.9: The system model shown in [49]. The top processing branch depicts DBP implemented via the SSFM, while the bottom branch shows the LDBP approach obtained by "unrolling" the SSFM. The LP and MF blocks represent the low-pass and matched filters, respectively. LDBP performs similarly to DBP in the absence of parameter optimization via deep learning and when the same number of steps/layers are used.

linear activation functions, which is identical to the corresponding computational graph of the DBP. This similarity allows us to use the DBP's computational graph as a blueprint for the neural network. The neural network implemented in LDBP is complex-valued and replaces the chromatic dispersion compensation (CDC) block in the DSP at the receiver. An illustration of the LDBP is shown in Fig. 3.9, where the upper branch shows the computational graph generated by the symmetric DBP with 2 steps/span, and the lower branch shows the proposed equivalent LDBP with similar graph, but the parameters in the linear layers and nonlinear functions are trainable and optimized by the neural network.

In the research presented in [65] and [47], the first simulations of LDBP were conducted for single-channel transmission, specifically focusing on 16-QAM modulated signals, and were carried out at a transmission rate of 10.7 Gbaud, and subsequently in [49], at a rate of 20 Gbaud. The scope of these simulations was later expanded to encompass multi-channel wavelength-division multiplexing (WDM) transmission, as detailed in [48]. Throughout these simulations, the optical fiber link spanned a total distance of 3200 km, divided into 32 spans, each with a length of $L_{sp} = 100$ km. Notably, LDBP consistently demonstrated a reduction in computational complexity while concurrently maintaining, and in some instances even enhancing, Q-factor performance. For instance, the performance gain of LDBP over DBP and linear equalization is demonstrated in 3.10, which is obtained by comparing the performance of LDBP with 1 step/span to DBP with 2 steps/span. The authors report that LDBP achieves comparable performance while utilizing only half the number of steps, resulting in a remarkable 50% reduction in complexity compared to DBP with similar performance. These compelling results strongly indicate that LDBP holds substantial promise as a more efficient alternative for mitigating nonlinear interference in long-haul optical communication systems.

The authors extended their model in subsequent studies [12, 50], to include PMD effects in their simulations. This expansion aimed to tackle the challenges associated with PMD in optical communication systems. In their simulation setup, the authors fine-tuned

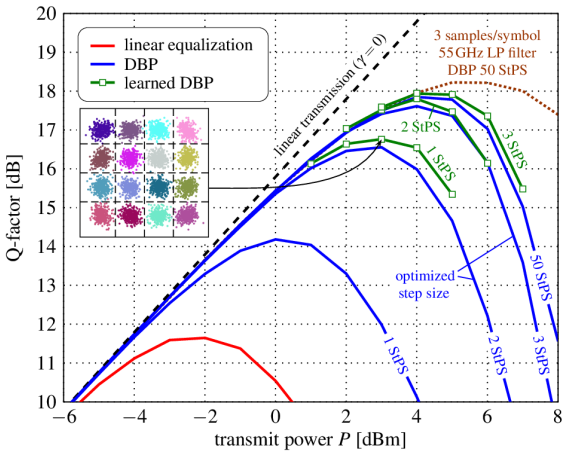


Figure 3.10: Simulation results comparing the proposed LDBP against DBP and linear equalization, as shown in [49].

parameters within the LDBP model to introduce PMD emulation at each simulation step, and labeled their method as "LDBP-PMD". The authors examined different parameterization options, including Free MIMO Filters, Free DGD Filters + Free Matrices, Free DGD Filters + SU(2)* Matrices, Lagrange Filters + Free Matrices, and Lagrange Filters + SU(2)* Matrices. Among these, "Lagrange Filters + SU(2)* Matrices" emerged as the method offering optimal balance between performance and complexity. Remarkably, LDBP-PMD demonstrated efficient convergence, delivering performance levels close to those in PMD-free scenarios, all without requiring prior knowledge of specific PMD realizations.

As a continuation of the previous task, which aimed to equip model-based approaches with the capability to perform adaptive equalization, the authors in [31] have taken a significant step forward by combining ML with adaptive DSP within a single trainable model called the generalized digital back-propagation (GDBP). Integrating adaptive DSP components as stateful Neural Network (NN) layers not only facilitates the utilization of standard ML training techniques but also enhances the performance of optical communication systems. The authors tested the GDBP in a multichannel scenario, employing a 7x288 Gb/s transmission experiment. Their results demonstrate that GDBP outperforms existing methods, including DBP, enhanced digital back-propagation (EDBP), filtered digital back-propagation (FDBP), and CDC, both in experimental and numerical evaluations over a transmission distance of 1125 km.

In [100], the authors propose a novel method for joint intra and inter-channel nonlinearity compensation in long-haul coherent optical communication systems. They introduce a deep-unfolded neural network architecture called DU-DBP-NPCC, which combines conventional DBP and nonlinear polarization crosstalk canceller (NPCC) into a single trainable model. Through their simulations of a 7-channel, 20-GBaud DP-16QAM, 3200-km co-

herent optical transmission system, DU-DBP-NPCC achieves a performance improvement of 1 dB and 0.36 dB in Q-factor compared to CDC and cascaded DBP-NPCC, respectively, at a launch power of -1 dBm/channel. DU-DBP-NPCC extends the maximum transmission reach by 28% (700 km) compared to CDC, demonstrating its effectiveness in long-haul communication systems. It also outperforms cascaded DBP-NPCC by 7% in terms of transmission distance. An analysis of computational complexity reveals that DU-DBP-NPCC offers approximately 26% lower computational complexity compared to conventional DBP-NPCC.

Finally, in [40], the authors focus on improving the performance of nonlinear frequency division multiplexing (NFDM) communication systems, which rely on the nonlinear Fourier transform (NFT). NFDM systems have shown remarkable progress but face challenges related to fiber loss and noise, as the NFT theory assumes a lossless transmission fiber. To address this, the authors propose a novel approach involving the replacement of the conventional NFT receiver with a time-domain NN-based symbol decisor. They experimentally validate this idea by developing a receiver architecture that combines a two-stage iterative carrier recovery with an NN-based symbol decisor. The experimental results demonstrate significant improvements, with a two-fold increase in transmission reach (approximately 1600 km) compared to a conventional NFT receiver (approximately 560 km) in a practical link scenario using erbium-doped fiber amplifiers (EDFAs).

3.3.4 Comparing model-agnostic and model-based approaches

While literature, to the best of our knowledge, lacks a specific study directly comparing model-based approaches and model-agnostic approaches, we can discern several key advantages and drawbacks associated with each of these approaches.

Model-based approaches offer several compelling advantages when designing neural networks. Firstly, they typically involve fewer trainable parameters, which not only reduces the risk of convergence-related issues but also mitigates overfitting concerns, enabling training with smaller datasets and saving time [51]. Moreover, model-based methods are rooted in the physical model of a phenomenon (e.g., DBP in the case of LDBP), which enables accurate parameter optimization to enhance accuracy and performance. Consequently, LDBP is expected to consistently outperform models like DBP. Additionally, model-based approaches have demonstrated the capability to perform many-to-many equalization, in contrast to many black-box neural network equalizers that typically perform one-to-one or many-to-one equalization. Many-to-Many equalization may contribute more effectively to reducing complexity and may be better suited for real-time implementation in optical receivers.

On the contrary, model-agnostic approaches, also known as black-box approaches, offer distinct advantages. Firstly, they provide greater flexibility in selecting hyperparameters

and designing the neural network. Secondly, deep neural networks are recognized as universal approximators, meaning that they can model any function with sufficient neural network width. Lastly, these approaches can be positioned at various points within the DSP chain, either at the receiver or even at the transmitter for pre-equalization, granting versatility in their integration into optical communication systems.

Indeed, while both approaches offer their respective advantages, it is crucial to acknowledge their associated limitations. Model-based approaches, for instance, often employ a lower number of parameters, potentially constraining their representational power. Overcoming this limitation typically requires precise and appropriate initializations to ensure effective convergence. Additionally, effectively implementing the model-based approach depends on having sufficient knowledge of the channel, including parameters like fiber properties and lengths. Furthermore, these approaches may face constraints concerning their placement within the receiver architecture. Take the example of LDBP, a model-based approach which must be situated in the DSP instead of the CDC block since DBP takes the place of CDC in DSP.

Lastly, model-agnostic approaches have their own limitations. They tend to create large and complex models, increasing their computational complexity [39]. These models demand extensive datasets for effective learning and to prevent overfitting. Since they lack grounding in specific physical phenomena or established theories, comprehending what the neural network is learning can be challenging. It remains uncertain whether the acquired knowledge is inherently linked to the channel it operates in or if it can be applied to different channels. Experiments have shown that such neural networks can even learn random number generator sequences and perform well in controlled experiments. However, their performance decreases in practical systems with truly random data. Furthermore, model-agnostic approaches typically perform one-to-one equalization.

Considering the advantages and limitations of each approach, the decision between a model-based and a model-agnostic approach ultimately depends on the user's specific needs. Both approaches have their merits and drawbacks, and there is no one-size-fits-all solution. The choice should be guided by the particular application at hand, ensuring it aligns with the strengths of the selected approach.

3.4 Summary

When it comes to channel equalization using neural networks, two main categories of approaches can be distinguished: model-agnostic and model-based neural networks. Model-agnostic neural networks do not rely on prior knowledge or assumptions about the underlying physical system, while model-based approaches incorporate known physical models of the channel into the neural network design.

Model-agnostic approaches, such as dense neural networks (DNNs), CNNs, and RNNs, have gained significant attention in recent years due to their ability to learn complex mappings between input and output signals without any assumptions about the underlying physics. These models are trained on large datasets of input-output pairs, and they are able to generalize to unseen data by learning patterns and correlations from the training data. Model-based approaches, on the other hand, take advantage of the known physical models of the channel to design neural networks that can perform channel equalization. For example, LDBP is a model-based approach that uses a physical model of the fiber channel to compensate for signal distortion caused by fiber dispersion and nonlinearity.

One advantage of model-agnostic approaches is their flexibility and ability to learn complex mappings without any prior knowledge of the physical system. However, these approaches may suffer from overfitting if the training data does not represent the full range of channel impairments, and they may not generalize well to unseen channels or scenarios. On the other hand, model-based approaches have the advantage of incorporating prior knowledge of the physical system into the design of the neural network, which can lead to better performance and more accurate channel equalization. However, these approaches are often limited by the accuracy of the physical model, and they may not perform well if the channel is significantly different from the model assumptions.

CHAPTER 4

Low-Complexity Learned Digital Back-Propagation for Dual-Polarization Systems

We propose the use of a parameter-sharing method applied in convolutional neural network to mitigate the effects of fiber transmission. This approach significantly reduces the number of trainable parameters by a factor of five compared to other equalization methods while improving Mean Squared Error (MSE) by 3.5 dB in comparison to digital back-propagation (DBP), while maintaining similar complexity.

Contents

4.1 Introduction

As the need for higher data rates continues to grow, so does the stress on optical fiber networks. To attain these higher data rates, the utilization of higher-order modulation formats becomes necessary. However, these advanced modulation formats demand increased transmitted power. As the transmission power increase, the distortions induced by Kerr

nonlinearity become a critical limiting factor for achievable information rates (AIRs). The mitigation of the nonlinear effects is a major challenge for network operators, one that is becoming increasingly necessary to address if they wish to elevate the transmission rates to meet today's demands.

Many effective methods for the mitigation of nonlinear distortions have been proposed, most notably, DBP, as found in multiple studies such as [23, 57, 74, 78, 81]. DBP excels in the realm of nonlinear equalization because it offers a step-by-step process to eliminate the nonlinearity at each step, which mirrors the model describing how the signal propagates. DBP, however, requires knowledge of fiber parameters, and can be computationally expensive in part due to potentially large number of spatial segments. Furthermore, several studies have indicated that DBP parameters require system-specific optimization [46, 90, 93, 94], making it challenging to find universally optimal parameters.

Neural networks provide a potentially less complex alternative [108]. Importantly, neural networks can address certain limitations of the model by training their parameters for channel equalization without requiring perfect knowledge of channel parameters. Neural network-based methods have the capability to learn the channel model from data, utilizing a set of examples for training.

In Chapter 3, we highlighted the use of neural network-based methods for equalizing optical signals. This has led to numerous equalization techniques, mainly falling into two categories: model-agnostic (black box) approaches and model-based (model-informed) approaches. Model-based approaches are systematic and generally more compact than their counterparts. One example of a model-based approach is the learned digital back-propagation (LDBP), proposed in [12], which combines the strengths of DBP, known for its high performance, with neural networks, recognized for their ability to learn the channel model from a set of training data. C. Häger introduced the LDBP as the pioneering model-based approach, paving the way for others in the same vein, such as Enhanced-DBP (EDBP)[91, 92] and Generalized-DBP (GDBP)[31].

The LDBP approach, which is model-based and inspired by the computational graph of the DBP, optimizes the linear filters within the DBP using the built-in neural network learning algorithm. When comparing LDBP to DBP, LDBP clearly outperforms it while maintaining the same level of complexity. In the pursuit of improving LDBP, several directions have been proposed: reducing the complexity of the activation function by replacing it with generic, easy-to-compute activation functions such as ReLU and tanh, or reducing the complexity of the linear operators. However, both approaches have proven to be challenging for the following reasons: Firstly, the activation function in LDBP, which is the same as the nonlinear function used in DBP, plays a crucial role in LDBP's correct operation, but generic activation functions like ReLU and tanh cannot capture the complex relationships between the signal and the nonlinear phase distortions affecting it, as well as

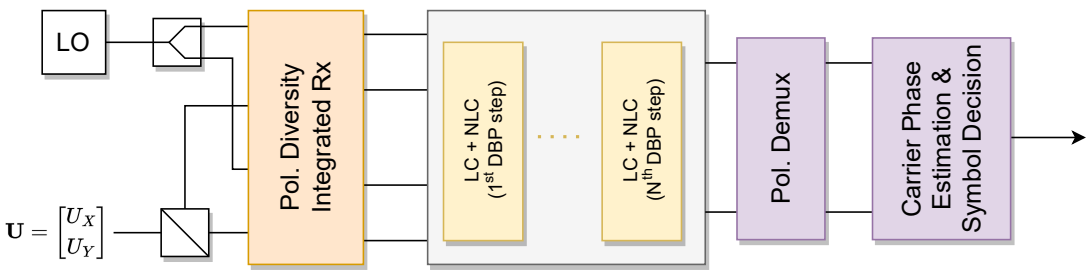


Figure 4.1: Block diagram of the coherent receiver with DBP used in our simulation. (LC=Linear compensation, and NLC = Nonlinear compensation)

the interactions between the two signals in the two polarizations. Secondly, the number of taps in the linear steps is proportional to the distance traveled and cannot be reduced beyond what DBP and LDBP already achieve. However, one potential simplification of LDBP, which we aim to explore, involves reducing the number of trainable parameters. This can be achieved by exploiting similarities among DBP steps, where similar steps are assigned identical weights, thus streamlining parameters into a single trainable layer. We refer to this method as 'parameter sharing' and explore its advantages in this study.

This chapter is divided into two sections: the first section is dedicated to optimizing the step size in DBP. In the second part of the chapter, we explore the LDBP, which represents another DBP optimization approach, leveraging neural networks for further enhancement in performance.

4.2 DBP step-size optimization

The DBP technique employs an iterative multi-step process to reverse the channel. Each step consists of a linear substep that equalizes chromatic dispersion, followed by a nonlinear substep that compensates for the nonlinearity introduced by Kerr effects, as described in Chapter 2. While DBP theoretically has the ability to completely counteract the nonlinear effects of optical fiber propagation, practical limitations exist due to the stochastic nature of certain impairments such as amplified spontaneous emission (ASE) and polarization mode dispersion [60]. However, in practical applications, the complexity of hardware implementation is limited, which directly limits the total number of DBP steps, which highlights the importance of step optimization in DBP under specific complexity constraint.

4.2.1 Simulated system model

We simulated an optical fiber transmission system with a symbol rate of 32 Gbaud using PM-16QAM modulation format. The noise figure of the erbium-doped fiber amplifier (EDFA) used in the simulation is 5.5 dB. The transmission fiber consists of 32 spans,

each measuring 72 km in length. The effect of polarization mode dispersion (PMD) are included in the simulation, where PMD coefficient is 0.05 ps/$\sqrt{\text{km}}$. At the receiver, the signal is converted into an electrical signal by the polarization diversity integrated receiver and sampled to 2 samples/symbol. it is then passed to either the chromatic dispersion compensation (CDC) block, or the DBP algorithm. The output is then processed by the DSP block. The DSP block chain consists of a polarization demultiplexer, carrier frequency offset (CFO), and constant phase estimation (CPE) block. The output of the DSP chain is sampled at 1 sample/symbol. A block diagram of the receiver is shown in Fig. 4.1.

The Manakov-PMD equation is widely used to simulate the signal propagation in the fiber in the presence of PMD effects

$$\frac{\partial \mathbf{q}}{\partial z} = -\frac{\alpha}{2}\mathbf{q} + \frac{j\beta_2}{2}\frac{\partial^2}{\partial t^2}\mathbf{q} - j\gamma\frac{8}{9}||\mathbf{q}||^2\mathbf{q} \tag{4.1}$$

The equation can be split into two subequations describing the linear effects and nonlinear effect, such as

$$\frac{\partial \mathbf{q}}{\partial z} = (\hat{D} + \hat{N})\mathbf{q} \tag{4.2}$$

where, $\mathbf{q} = [q_x, q_y]^T$ is the signal vector, $\hat{D} = \frac{j\beta_2}{2}\frac{\partial^2}{\partial t^2}$, $\hat{N} = -j\gamma\frac{8}{9}\mathbf{q}^{\mathbf{H}}\mathbf{q} - \frac{\alpha}{2}$,

The solution of the Manakov-PMD equation can thus be described as

$$\mathbf{q}(z+h,t) = \exp\left(h(\hat{D}+\hat{N})\right)\mathbf{q}(z,t) \approx \exp(h\hat{D})\exp(h\hat{N})\mathbf{q}(z,t) \tag{4.3}$$

The attenuation coefficient in the nonlinear operator $\hat{N}$ alters the signal intensity during fiber propagation, resulting in attenuation in the forward direction (for SSFM) or magnification in the backward direction (as in DBP). The effect of attenuation can be omitted by assuming that DBP operates on signal with normalized power, in which case, the transmission distance must be normalized. This can be performed by taking the effective nonlinear length described as

$$L_{\text{eff}} = \frac{1 - \exp(-\alpha L_{\text{span}})}{\alpha} \tag{4.4}$$

where L_{span} is the span length. Assuming the asymmetric Wiener-Hammerstein model, and solving for one span using 1 steps per span (StpS), the solution can be written as follows:

$$\mathbf{q}(z+L_{\text{span}},t) \approx \exp(\mu L_{\text{span}}\hat{D})\exp(\varepsilon L_{\text{eff}}\hat{N})\exp\left((1-\mu)L_{\text{span}}\hat{D}\right)\mathbf{q}(z,t). \tag{4.5}$$

Here, $\mu \in [0,1]$ is a parameter controlling the distance of the first linear substep. It can be seen that $\mu = 1$ corresponds to a Wiener model, $\mu = 0$ corresponds to a Hammerstein model, and $\mu = 0.5$ corresponds to a symmetric Wiener-Hammerstein model, as shown in Fig. 2.19. Another degree of freedom, $\varepsilon \in [0,1]$, has been introduced in equation 4.5 to improve the optimization process. It can be observed that when $\varepsilon = 0$, this corresponds to performing linear equalization.

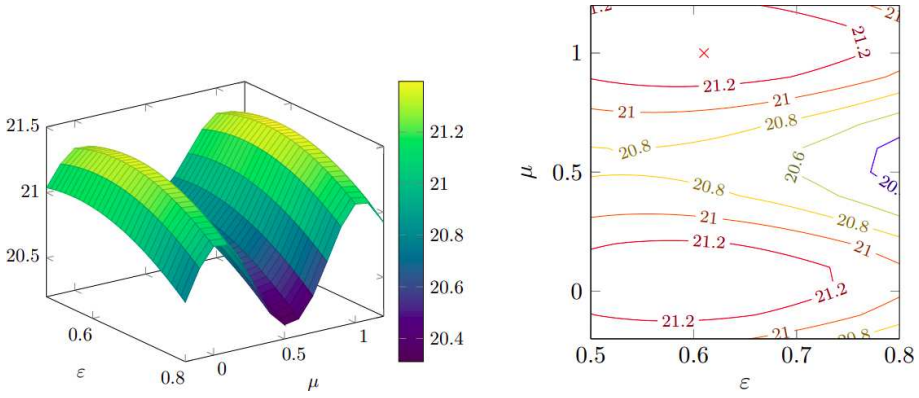


Figure 4.2: Performance of DBP measured in SNR_{eff} [dB] as a function of the parameters μ and ε, with a signal launch power of -2 dBm. The left plot shows the 3D surface, while the right plot presents a contour plot for easier visualization.

4.2.2 Simulation Results

In the following, we demonstrate the results obtained form numerical optimization of 1 StpS and 2 StpS DBP by optimizing the step sizes and the nonlinear coefficient.

4.2.2.1 Optimizing 1 step/span DBP

DBP with a StpS resolution of 1 is the most commonly utilized approach in industrial systems [13, 78]. Although higher resolutions may theoretically offer better results [49, 81], as the complexity of the system increases, the interactions between linear and nonlinear effects become more intricate and challenging to solve with a greater number of StpS. Therefore, increasing the StpS may not always guarantee improved performance, and system operators often opt for a 1 StpS resolution to avoid excessive complexity.

For the 1 StpS DBP, we assume that the position of the nonlinear step and the value of ε are the same for each span, which provides two degrees of freedom to optimize the DBP according to Eq. 4.5, namely μ and ε, where the PM-16QAM signal is launched at a launch power of -2 dBm. The equalization performance of the 1 StpS DBP is evaluated across various values of μ and ε through a grid search. Fig.4.2 shows the performance of the DBP as a function of μ and ε, measured by the effective SNR (SNR_{eff}), defined as:

$$\text{SNR}_{\text{eff}} = \frac{\sum_{n=1}^{N} |\hat{s}_{x,n} - s_{x,n}|^2}{\sum_{n=1}^{N} |s_{x,n}|^2}, \tag{4.6}$$

where $\hat{s}_{x,n}$ represents the estimated symbols, and $s_{x,n}$ represents the real symbols. The graph displays peak performance at $\mu = 1$ and $\varepsilon = 0.62$, indicated by the red cross symbol, with a peak performance of 21.4 dB. However, in the unoptimized case of $\varepsilon = 1$ and $\mu = 0.5$, which corresponds to equal lengths for the two linear half-steps within each DBP step, the DBP performance declines to 20.3 dB, a reduction of 1.1 dB compared to the optimized parameters.

4.2.2.2 Optimizing DBP with more than 1 step/span

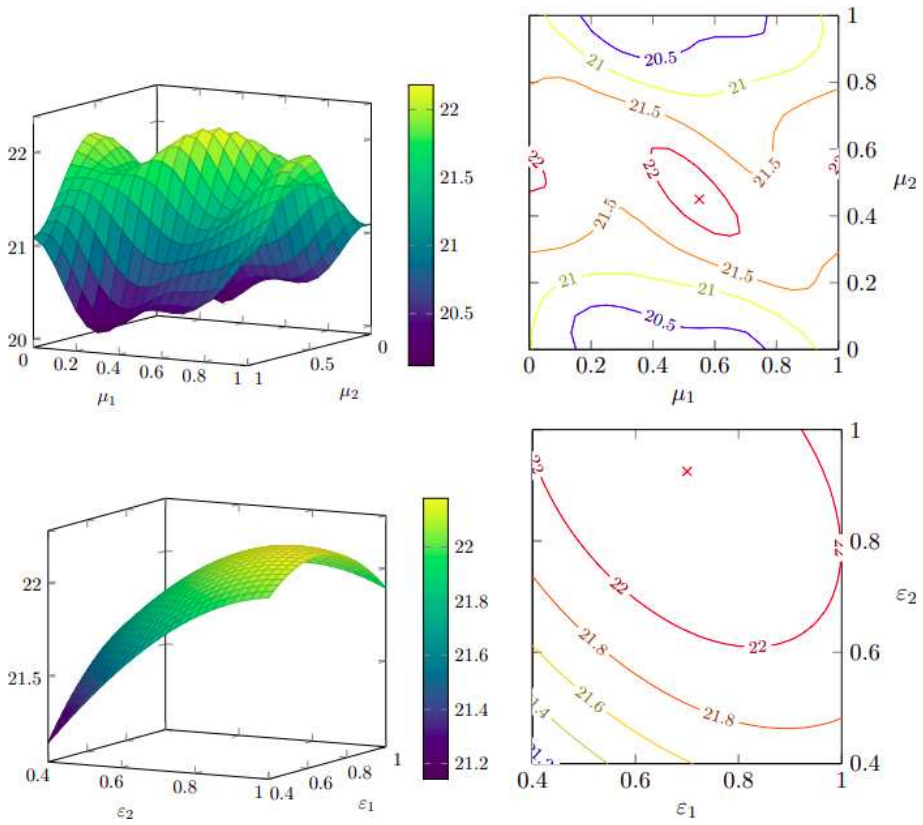


Figure 4.3: Performance of DBP measured in $\mathrm{SNR}_{\mathrm{eff}}$ [dB] shown as two 3D surface plots and their respective contour plots, as a function of parameters μ_1 and μ_2 (upper) and ε and ε_2 (lower). The red cross symbol indicates the location of the maximum performance in both diagrams.

In multi-step DBP, different optimization parameters can be assigned to each substep, denoted as $\mu_1, \mu_2, \ldots,$ and $\varepsilon_1, \varepsilon_2, \ldots$. Similar studies have explored the impact of this kind of optimization on DBP performance [9, 10, 90]. The evaluation of the 2 StpS DBP's equalization performance consists of two stages. Initially, a grid search is performed across various values of μ_1 and μ_2. The optimal combination of μ_1 and μ_2 is selected from this search. In the second stage, ε_1 and ε_2 are optimized through a similar grid search. The performance results are presented in Fig. 4.3, featuring two graphs. The first graph illustrates the performance variations with changes in μ_1 and μ_2. The peak performance, marked by the red cross symbol, is achieved at $\mu_1 = 0.52$ and $\mu_2 = 0.46$. The second graph in Fig. 4.3 depicts the performance while altering ε_1 and ε_2, with μ_1 and μ_2 fixed at their previously optimized values ($\mu_1 = 0.52$ and $\mu_2 = 0.46$). The peak performance is achieved at $\varepsilon_1 = 0.68$ and $\varepsilon_2 = 0.9$, reaching 22.2 dB. The optimal parameter values for DBPs with 1, 2, and 3 StpS are summarized in Table 4.1.

Finally, we present a comparison between unoptimized DBP and optimized DBP in Fig. 4.4. The PMD-aware DBP shown in the figure is an unrealistic DBP that assumes

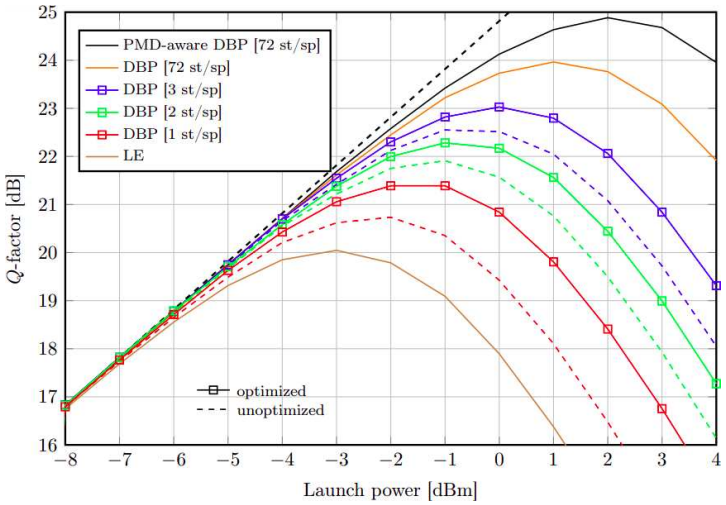


Figure 4.4: Comparison between different DBP equalizers.

as many DBP steps as there are steps in the split-step Fourier method (SSFM) used to simulate the signal propagation in the forward direction, and it also assumes knowledge of all the PMD values, phase rotations, and polarization mixing in each step in the SSFM. This way, the PMD-aware DBP is able to mitigate all the channel effects except for the ASE noise generated by the EDFAs, and hence represents the upper limit of what is achievable by different equalization techniques. We also simulated the max-precision DBP, which has as many steps per span as the SSFM but without the knowledge of PMD values. Additionally, we included a comparison of the linear equalization performance in Fig. 4.4 to provide a benchmark for the performance of the different DBP models.

For the 1 StpS DBP, the unoptimized version achieves a maximum peak performance of 20.8 dB at a launch power of -2 dBm. In contrast, the optimized DBP with the same StpS achieves a peak value of 21.4 dB, which occurs at a launch power between 1 dBm and 2 dBm. Similarly, for the 2 StpS DBP, the unoptimized version attains a maximum peak performance of 21.9 dB at a launch power of -1 dBm, while the optimized DBP with the same StpS achieves a higher peak value of 22.2 dB at the same launch power. In the case of the 3 StpS DBP, the unoptimized version reaches a maximum peak performance of 22.5 dB at a launch power of -1 dBm. Conversely, the optimized DBP with the same StpS achieves an even higher peak value of 23 dB, observed at a launch power of 0 dBm. Generally, by comparing the optimized and unoptimized DBP in the figure, the simulation results demonstrate the potential improvement of DBP through step-size and parameter optimization.

	μ_1	μ_2	μ_3	ε_1	ε_2	ε_3
1 StpS	1.000			0.619		
2 StpS	0.518	0.468		0.676	0.910	
3 StpS	0.304	0.352	0.278	0.516	0.972	1.162

Table 4.1: The optimal values of μ_i and ε_i for different DBP resolutions

4.3 Learned Digital Back-propagation

In the previous section, we have explored the optimization of DBP by manually adjusting a few key parameters, such as step-sizes and the nonlinear parameter. In the next section however, we take a different approach. Instead of manually fine-tuning a limited set of parameters, we harness the power of LDBP as a novel tool which enables us to optimize individual DBP filter taps and capture more intricate interactions between the linear and nonlinear channel effects than achievable by conventional DBP optimization methods.

In the following sections, we introduce the fundamental concept of LDBP and present our novel approach to reduce its complexity. The results presented in the subsequent sections have been published in [1] in the Signal Processing in Photonic Communications (SPPCom).

4.3.1 Simulated system model

We consider a single-channel transmission of a polarization division multiplexed (PDM) signal using the system illustrated in Fig. 4.5. Here, we make the assumption that the channel operates without considering the effects of PMD. The dual-polarization transmitted signal, denoted as $q_{x,y}(t,0)$, can be expressed as:

$$q_{x,y}(t,0) = \sum_{k=1}^{N_s} s_{x,y}^k p(t - kT_s) \tag{4.7}$$

In this equation, $p(t)$ represents the pulse shape as a function of time t, $\{s_{x,y}^k\}_k$ are transmitted symbols drawn from a constellation $\mathcal{S}$, N_s denotes the number of symbols, and T_s is the symbol period. We generate this signal using the 16-QAM format and launch it into the optical fiber. The propagation of the signal through the optical fiber is modeled by the vector nonlinear Schrödinger system:

$$\frac{\partial q_i}{\partial z} = \left[-\frac{\alpha}{2} - \frac{\mathrm{j}\beta_2}{2}\frac{\partial^2}{\partial t^2} + \mathrm{j}\gamma \left(|q_i|^2 + \frac{2}{3}|q_{\bar{i}}|^2 \right) \right] q_i \tag{4.8}$$

In this equation, $(i, \bar{i})$ can take values (x, y) or (y, x), z represents distance, and α, β_2, and γ stand for the loss, dispersion, and nonlinearity coefficients, respectively. Analyz-

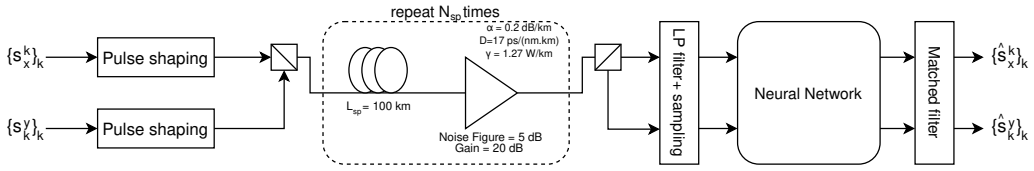


Figure 4.5: Block diagram of the end-to-end system model

ing the system, it becomes apparent that the signal propagating through the channel is primarily affected by chromatic dispersion and the Kerr nonlinearity originating from the signal itself. Since all these effects are deterministic (except for the noise generated from EDFAs), DBP equalization is practically capable of mitigating these effects, leaving only the noise that affects the signal. The DBP has been previously studied in Chapter 2. In the following, we will discuss the operation of the LDBP.

4.3.2 Mathematical Model of LDBP

When designing NNs, incorporating prior knowledge of the system's model can significantly expedite the training process and improve the convergence rate to a lower local minimum compared to black-box approaches [97]. The similarity between the SSFM and deep feed-forward NNs had been pointed out in the literature [12, 49], where both involve alternating between linear matrix multiplication and nonlinear element-wise operator. This similarity can be exploited to design a model-based NN with DBP as a blueprint, which allows for optimizing DBP parameters using the SGD algorithm [49].

The LDBP equalizer is a deep neural network model comprising a series of alternating layers Each layer in the LDBP is inspired by its equivalent step in the DBP, and it replicates both the linear and nonlinear components of the DBP step. Similarly to deep feed-forward NNs, the LDBP can be mathematically represented as a series of alternating linear operations, denoted as $\mathbf{A}^{(k)}$, and element-wise activation function, denoted as $\Phi^{(k)}$. These operations create a mapping from an input vector $\mathbf{u}$ to an output vector $\mathbf{v}$, as follows:

$$\mathbf{v} = \Phi^{(d_l)}(\mathbf{A}^{(d_l)}(\Phi^{(d_l-1)}(\dots \mathbf{A}^{(0)}(\mathbf{u})))), \tag{4.9}$$

where N_l represents the number of layers in the LDBP, and the superscript (k) denotes the index of the k-th layer of the LDBP. The function performed by the linear operator $\mathbf{A}^{(k)}$ varies depending on the neural network architecture. In the case of fully-connected NNs, it performs matrix multiplication. However, in the case of convolutional NNs, it performs convolution. When fully-connected NNs are considered, The k-th LDBP step linear operator $\mathbf{A}^{(k)}$ is a $d \times d$ matrix, which can be expressed as

$$\mathbf{A}^{(k)} = \mathbf{F}^{-1}\text{diag}(H_1, ..., H_N)\mathbf{F}. \tag{4.10}$$

Here, diag$(\cdot)$ represents a diagonal $d \times d$ matrix with diagonal elements $H_k = \frac{\alpha}{2}\delta - j\frac{\beta_2}{2}\omega_k^2\delta$

for $k = 1, 2, \ldots, d$, where $\omega_k = 2\pi f_k$ and f_k denotes the k-th discrete frequency. Additionally, $\mathbf{F}$ denotes the discrete Fourier transform (DFT) matrix

$$\mathbf{F} = \frac{1}{\sqrt{d}} \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & e^{j2\pi\frac{1}{d}} & \ldots & e^{j2\pi\frac{(d-1)}{d}} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{j2\pi\frac{(d-1)}{d}} & \ldots & e^{j2\pi\frac{(d-1)(d-1)}{d}} \end{bmatrix}. \tag{4.11}$$

The activation function for the k-th layer $\rho^{(k)} : \mathbb{C}^d \mapsto \mathbb{C}^d$ is modeled as $\rho^{(i)}(x) = xe^{-j\alpha_i|x|^2}$, where α_i is a trainable parameter.

The primary goal of equalization is to estimate the true output vector $\mathbf{v} \in \mathbb{C}^d$ as $\hat{\mathbf{v}} \in \mathbb{C}^d$ based on the input vector $\mathbf{u} \in \mathbb{C}^d$, while minimizing the error represented by the difference between the output and estimated vectors.

4.4 Simplifying LDBP with Parameter-Sharing

Let us consider the i-th layer in the LDBP, Each layer in the LDBP comprises two weight matrices, $W_1^{(i)}, W_2^{(i)} \in \mathbb{C}^{d\times d}$, where d is the dimension of the neural network weight matrices. During training, the weight matrices are constrained, such that they perform circular convolution with a symmetric filter of length $2K_m+1$, where K_m is a hyper-parameter. The matrix rows are a circularly shifted version of the vector $(u_{-K_m}, \ldots, u_{-1}, u_0, u_1, \ldots, u_{K_m}, 0, \ldots, 0)$, where $u_i \in \mathbb{C}$ and $u_{-i} = u_i$. The weight initializations are based on a pruned version of the matrix $\mathbf{A}_{\delta/2}$ in DBP (top branch in Fig. 3.9), where the final layer in LDBP $W^{(l)} \in \mathbb{C}^m$ performs match filtering and reduces the input's dimensionality to $m < n$ to account for the memory induced by dispersive channel effects. We note that the same DBP steps are repeated in each span. Thus, instead of joint training of all neural network layers, we train a few unique layers that are shared in network depth. With this parameter sharing method, we substantially reduce the number of model trainable parameters.

In CNNs, each layer maps the input vector $\vec{X}^{(l-1)}$ to an output vector $\vec{X}^{(l)} = \Phi(\vec{W}^{(l)} * \vec{X}^{(l-1)})$ by convolution multiplication with a kernel $\vec{W}^{(l)}$. The main idea of this paper is to exploit the similarity between the neural network function and the un-folding of the SSFM, since both involve alternating between linear and nonlinear operations in their functions. It is hence possible to initialize the deep CNN parameters to perform SSFM; This approach has been studied and is referred to as model-based neural networks. We extend the model-based design of neural networks by assuming that some of the training parameters can be shared between multiple layers. Fig.4.6 shows the proposed CNN equalizer with parameter sharing and the corresponding SSFM with $M = 2$ steps per span. Since the linear and nonlinear operations of SSFM (*i.e.* A_m and σ in Fig.4.6, respectively) are repeated in each span in the optical fiber, we are able to identify 4 layers with unique parameters,

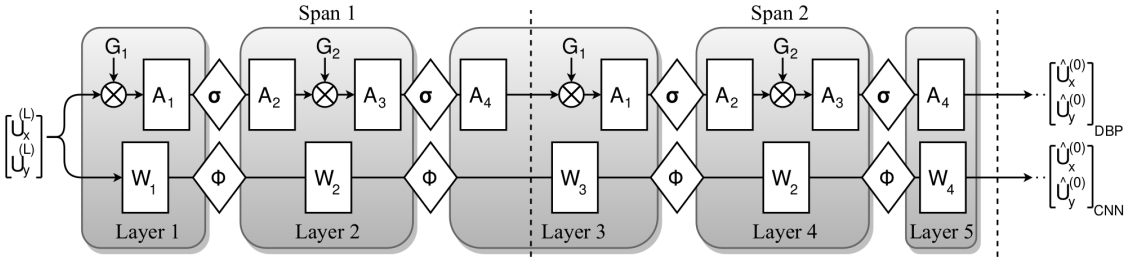


Figure 4.6: An illustration of the proposed neural network architecture. The top branch corresponds to asymmetric DBP, and the bottom branch to the proposed CNN. $q_{x,y}^{(0)}$ and $q_{x,y}^{(L)}$ refer to the signals at the transmitting and receiving ends of the fiber, where L is the number of segments.

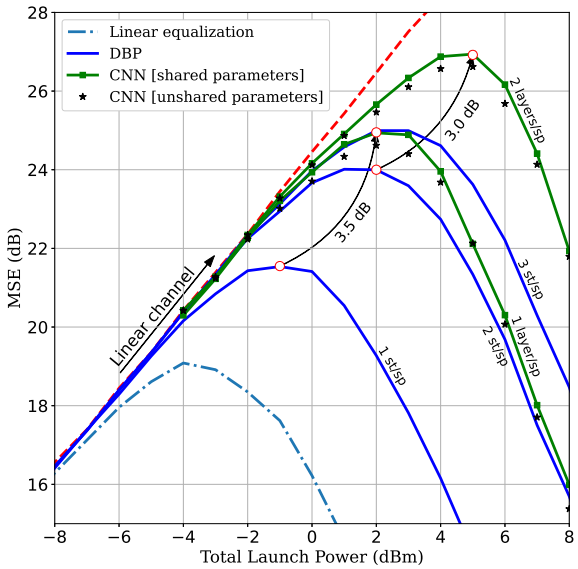


Figure 4.7: Comparison between compensation algorithms.

namely $W_1 \triangleq G_1A_1$, $W_2 \triangleq A_2G_2A_3$, $W_3 \triangleq A_4G_1A_1$ and $W_4 \triangleq A_4$. Overall, the number of uniquely defined layers is $M+2$ regardless of the number of spans in the optical link. The nonlinear activation function used in our CNN is $\Phi(\vec{X}) = \vec{X}e^{\mathrm{j}(||\vec{X}||^2+\frac{2}{3}||\vec{Y}||^2)}$ and $\Phi(\vec{Y}) = \vec{Y}e^{\mathrm{j}(||\vec{Y}||^2+\frac{2}{3}||\vec{X}||^2)}$ for each polarization. In our model, amplification is applied after each segment such that $\prod_{k=1}^{M} G_k = 1$. With this, we provide the neural network the ability to adjust the intensity of the signal, and hence, the nonlinear phase rotation without the need of trainable parameters in the activation function.

4.5 Results

In our simulations, we generate dual-polarized 16-QAM signals at 32 Gbaud using root raised cosine pulse-shaping (roll-off factor = 0.1). We set the fiber length L_f=1000 km, and the number of spans N_{sp}=10. Forward-propagation is simulated with 50 StpS, and data is initially sampled at a high sampling rate of 16 samples/symbol. On the receiver side, the signal is down-sampled to 2 samples/symbol before being processed by either CNN or

DBP algorithms. We use Keras API built on TensorFlow to implement and optimize the neural network parameters. During the training of the model, examples consisting of input-output pairs were randomly chosen from the set of launch powers $P \in \{0,1,2,3,4\}$ dBm. For each point in Fig.4.7, we generate 3300 input-output pairs for testing. We express the performance gain in terms of the normalized mean squared error (MSE) defined as $\frac{||X-\hat{X}||^2+||Y-\hat{Y}||^2}{||X||^2+||Y||^2}$. The proposed CNN with 1 layer/span achieved MSE of 25 dB at 2 dBm launch power which corresponds to 3.5 dB improvement compared to the best performance attained by DBP with optimized step-size and equivalent complexity. For 2 layers/span, the peak MSE value for the CNN is 27 dB caluclated at 5 dBm launch power, which is 3 dB higher than DBP with similar complexity. Comparing to LDBP (asterisks in Fig.4.7), a small gain in MSE is observed. We explain this gain by the neural networks ability to generalize better with fewer training parameters. In terms of training complexity, our model uses 3 uniquely defined layers comparing to 11 layers for LDBP, in the 1 layer/span setup, and 4 uniquely defined layers comparing to 21 layers for LDBP, in the 2 layers/span setup.

4.6 Summary

A parameter sharing method is proposed to reduce the training complexity of CNNs for equalization in optical fiber. The proposed approach yields 3 – 3.5 dB gain in MSE compared to optimized DBP with comparable complexity, and five-fold reduction in number of trainable parameters compared to LDBP at the same MSE.

CHAPTER 5

Learned Digital Back-Propagation for Dispersion Managed Systems

This chapter investigates the application of learned digital back-propagation (LDBP) for dual-polarization fiber-optic transmission in dispersion-managed (DM) links. The evaluation is performed in a simulated wavelength-division multiplexing (WDM) system at 256 Gbit/s per channel, covering a 2016 km link with 15% residual dispersion. Results indicate LDBP significantly enhances single-channel transmission by 6.3 dB compared to linear equalization (LE) and 2.5 dB compared to digital back-propagation (DBP). In WDM, Q-factor gains are 1.1 dB and 0.4 dB, respectively. Additionally, the complexity analysis favors the frequency-domain (FD) implementation of LDBP and DBP over the time-domain approach. This highlights LDBP's effectiveness in mitigating nonlinear effects in DM fiber-optic systems. The research and findings presented in this chapter have been published in the Optics Express journal in [3].

Contents

5.1 Introduction

In earlier generations of optical receivers, prior to the advent of coherent detection, dispersion management was utilized to mitigate chromatic dispersion optically. This technique involved using dispersion-compensating fiber (DCF) fibers with a strongly negated dispersion coefficient D of appropriate length, effectively reducing chromatic dispersion to zero at the receiver's end. A typical dispersion-managed (DM) system configuration consisted of the following components in each span: single-mode fiber (SMF), a first-stage erbium-doped fiber amplifier (EDFA), DCF, and a second-stage EDFA. The DCF, characterized by its high negative dispersion coefficient, was strategically integrated into the link to compensate for a specified percentage of dispersion.

5.1.1 History of Dispersion-Managed Systems

The advent of the EDFA in the late 1980s marked a significant milestone in long-distance fiber transmission[26, 75]. It eliminated the need for signal reconstruction and regeneration during transmission, effectively addressing the issue of signal attenuation in optical fibers. Additionally, the widespread adoption of dispersion-shifted fibers (DSFs) in the early 1990s, characterized by near-zero dispersion at the operating wavelength, successfully mitigated chromatic dispersion challenges [80]. However, it's important to note that DSF introduced its own set of challenges, particularly nonlinear distortions like Four-Wave Mixing (FWM), which becomes prominent at zero chromatic dispersion [44].

The earliest DM systems utilized a combination of NZDSFs with opposite dispersion signs to achieve a near-zero net chromatic dispersion (CD) value while maintaining high local CD values over specific distances to counteract phenomena such as four wave mixing (FWM) [20]. This approach offered effective compensation, to some degree, for both CD and nonlinearities and found widespread use in high-speed dense wavelength-division multiplexing (WDM) commercial systems. These systems primarily used intensity modulation (IM)/direct-detection (DD), in which information was encoded in the signal envelope, while the signal phase was disregarded. They typically achieved data rates of around 10 Gb/s per wavelength, which were later upgraded to 40 Gb/s.

However, between 2000 and 2010, with the introduction of coherent detection, IM/DD systems rapidly gave way to their more advanced counterpart. Coherent detection improves receiver sensitivity compared to IM/DD [68]. Coherent systems used advanced modulation formats and sophisticated digital signal processing algorithms, offering improved spectral efficiency and better performance compared to DM systems. The ability to capture both

the amplitude and phase information of the optical signal enabled higher data rates and longer transmission distances. This transformation revolutionized optical communication, enabling transmission capacities in the multi-terabit per second range and facilitating the deployment of advanced modulation formats such as quadrature amplitude modulation (QAM) to maximize spectral efficiency. As a result, DM systems became less prevalent, with coherent detection emerging as the de facto scheme, effectively replacing DM systems.

5.1.2 Advancement in Coherent Detection

The advent of coherent receivers enabled the complete compensation of accumulated CD in the electrical domain through digital signal processing (DSP) [89], eliminating the need for traditional DM systems. Consequently, the focus shifted from intricate optical configurations to advanced DSP algorithms capable of mitigating CD effects. Despite this, commercial DM links continue to operate today. This is primarily due to the significant cost associated with upgrading deployed DM solutions, particularly in long-haul and submarine fiber links. Ongoing research and development efforts in nonlinear mitigation offer opportunities to improve the performance and extend the lifespan of deployed DM systems. By leveraging these advancements, it becomes possible to enhance the capabilities of DM networks and adapt them to meet the evolving demands of modern optical communication.

One such advancement is digital back-propagation (DBP), a technique designed to counteract deterministic nonlinear effects by digitally simulating signal propagation at the receiver using reversed fiber parameters [57]. However, DBP's effectiveness is constrained by hardware limitations, particularly as data rates increase, resulting in exponential complexity growth. The practical complexity constraints of DBP limit its applicability. Furthermore, DBP struggles to accurately account for signal interactions between adjacent channels, making it less suitable for WDM scenarios. Considerable research efforts have been devoted to optimizing DBP algorithms to achieve lower complexity and better performance in various transmission scenarios [46, 90]. In the context of DM and non-dispersion-managed (NDM) systems, numerical optimization of channel parameters for DBP has been conducted [67]. Notably, the combination of transmitter-side DBP with frequency referenced carriers has demonstrated the potential to effectively double the transmission reach [102]. Additionally, optimal DBP step sizes for polarization-division multiplexing (PDM) transmission systems have been investigated [93].

To further enhance the performance of DBP, several variants have been proposed. One such variant is correlated-DBP, which takes into account the correlation between neighboring signal samples [66]. Another variant, dispersion folded-DBP, was originally designed for zero-residual dispersion but has been extended to accommodate any dispersion map [34, 69, 110, 111]. filtered digital back-propagation (FDBP) introduces a parameterized low-pass filter (LPF) in the nonlinear step to improve phase tracking [27, 87]. Additionally,

enhanced digital back-propagation (EDBP) considers interactions between channel signals and adjacent channels, leading to further performance enhancements [91, 92]. In the context of WDM systems, coupled-Channel EDBP focuses on achieving optimal cross-phase modulation (XPM) equalization [21].

More recently, the integration of neural networks (NNs) has emerged as a promising approach to enhance DBP. learned digital back-propagation (LDBP) utilizes a deep NN inspired by the split-step Fourier method (SSFM) computational graph, optimized using stochastic gradient descent (SGD)[2, 12, 48, 49, 55]. LDBP treats the LPF filter taps in DBP as free parameters optimized by the NN. This approach has been extended to dual-polarization transmission in NDM systems [12] and experimentally demonstrated with one layers per span (LpS) [48]. Additionally, generalized digital back-propagation (GDBP) employs a deep NN to parameterize DBP, combining NN training with adaptive digital signal processing [30]. The integration of these advanced techniques, such as optimized DBP algorithms and NN-based enhancements like LDBP and GDBP, holds great promise for significantly improving the performance and capabilities of optical communication systems.

5.1.3 Motivation and Objectives

Despite DM systems being mostly outdated, several optical fiber cables, especially submarine optic cables constructed prior to the year 2000, are still in deployment today, utilizing variations of optical CD compensation. Instead of restricting these systems to IM/DD schemes, this chapter explores repurposing legacy DM links for use in coherent transmission systems, thereby expanding their data-carrying capacities. To achieve this, we leverage recent successful advancements in coherent detection, specifically DBP and LDBP. For DBP, we propose a variant tailored for systems with arbitrary dispersion maps. We observe that the linear filters in the DBP steps benefit from inline dispersion management as they are required to mitigate less chromatic dispersion, requiring a smaller number of filter taps compared to linear dispersion map in NDM systems. This contributes to reducing DBP complexity by allowing us to use smaller filters or fewer steps. Indeed, one of our realizations in this study is that DM systems allow DBP to operate with less than 1 steps per span (StpS), going as low as 1/14 StpS. This DBP variant serves as the basis for the subsequent implementation of LDBP. We assess the performance of both DBP and LDBP in a realistic long-haul dual-polarization WDM DM transmission system, considering various channel effects including loss, CD, Kerr nonlinearity, polarization mode dispersion (PMD), amplified spontaneous emission (ASE) noise, and laser phase noise (PN).

We do not claim that DM systems are superior to NDM systems. Our objective here is solely to investigate the possibility of upgrading already-deployed DM systems, commonly used for IM/DD, to enable coherent transmission. To achieve this, we employ machine learning techniques to optimize DBP, specifically utilizing the Learned-DBP method [2,

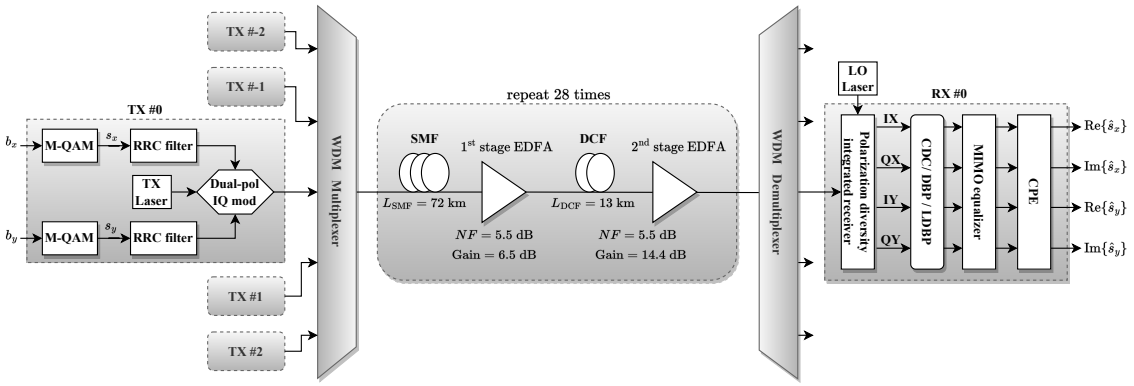


Figure 5.1: The DM-WDM optical fiber system.

12, 48, 49, 55].

Our results can be summarized as follows: In the context of single-channel transmission, the utilization of LDBP yields significant enhancements in terms of the effective signal-to-noise ratio (SNR_{eff}). Specifically, LDBP achieves a remarkable SNR_{eff} improvement of 6.3 dB and 2.5 dB compared to linear equalization (LE) and conventional DBP, respectively. For WDM transmission, LDBP further elevates the quality factor (Q-factor) by 1.1 dB and 0.4 dB compared to LE and DBP, respectively. To provide a comprehensive analysis, this research presents both time-domain (TD) and frequency-domain (FD) implementations of LDBP, denoted as TD-LDBP and FD-LDBP. Moreover, the impact of fiber parameter variations resulting from aging and laser PN on LDBP performance is thoroughly examined. Our investigations reveal that laser PN incurs a minor Q-factor penalty of less than 0.2 dB, as long as the laser linewidth remains below 200 kHz. Additionally, we demonstrate the robustness of LDBP by showcasing that its performance gains over DBP are consistently maintained even when the LDBP model is retrained with updated fiber parameters post-aging. These findings build upon the earlier work presented in [2] and provide valuable insights for enhancing data rates in DM systems through the utilization of coherent detection and LDBP techniques.

5.2 DBP Adaptation to DM Systems

5.2.1 System Model

In the polarization-multiplexed fiber transmission, two random bit streams with N_b bits, $\mathbf{b_x} = (b_x^{(1)}, b_x^{(2)}, \ldots, b_x^{(N_b)})$ and $\mathbf{b_y} = (b_y^{(1)}, b_y^{(2)}, \ldots, b_y^{(N_b)})$, $b_{x/y}^{(i)} \in \{0, 1\}$, are generated at the transmitter, then each is mapped into a sequence of N_s symbols $\mathbf{s_x} = (s_x^{(1)}, s_x^{(2)}, \ldots, s_x^{(N_s)})$ and $\mathbf{s_y} = (s_y^{(1)}, s_y^{(2)}, \ldots, s_y^{(N_s)})$, where $s_{x/y}^{(i)}$ are drawn from the constellation $\mathcal{S}$. The Gray mapping is used to map the bit stream to the symbols in the constellation. The baseband

signal of the polarization x is obtained by modulating $\mathbf{s_x}$ into the waveform

$$q_{x,n}(z=0,t) = \sum_{i=1}^{N_s} s_{x,n}^{(i)}\, p(t - iT_s), \tag{5.1}$$

where $q_{x,n}$ is the complex envelope of the signal in the WDM channel n as a function of distance z and time t, $p(\cdot)$ is the pulse shape, and T_s is the symbol period. $z = 0$ indicates that the signal is located at the transmitter. The equations for $q_{y,n}(0,t)$ are identical upon replacing x with y. The waveforms $q_{x,n}$ and $q_{y,n}$ are modulated into an optical signal $\mathbf{q}_n(0,t) = [q_{x,n}, q_{y,n}]$ using a dual-polarization Mach-Zehnder modulator. The complex envelope of the WDM signal $\mathbf{q}(0,t)$ launched in the fiber link is generated by adding the signals of the different WDM channels

$$\mathbf{q}(0,t) = \sum_{n} \mathbf{q}_n(0,t) e^{-jn\Delta\omega t}, \tag{5.2}$$

where $\Delta\omega$ is the frequency spacing between adjacent WDM channels. The WDM signal is launched through an optical fiber channel consisting of multiple spans, where each span includes a SMF, dual-stage EDFAs, separated by a DCF with a proper length to compensate for dispersion. DCF has a higher nonlinearity compared to SMF, so the first-stage DCF provides pre-DCF gain, but only to a power level that would not generate excessive nonlinear effects. The second EDFA in the span then amplifies the signal further to its original power level. The end-to-end channel with all components can be described by the interplay between CD, PMD and Kerr nonlinearity effects. The propagation of signals inside each span in the presence of PMD is governed by the coupled nonlinear Schrödinger's equation (CNLSE), modeling the interaction between the two states of polarization. The CNLSE was previously discussed in Chapter 2, particularly in Eq. 2.25. Furthermore, we consider a birefringent fiber, where the fiber length $\mathcal{L}$ is much larger than the beat length $L_B = 2\pi/(\beta_{0x} - \beta_{0y})$, and the coherent cross-polarization terms can be neglected [5, Eq. 6.1.11–12]. At the end of the span, an EDFA with gain G compensates for the fiber attenuation and introduces ASE noise. The noise $n(t)$ is a band-limited white circularly-symmetric complex Gaussian process with the auto-correlation function $E\big(n(t)n^*(t)\big) = \sigma_0^2 \delta_B(t - t')$, where $\delta_B(x) = B\text{sinc}(Bt)$, $\sigma_0^2 = \frac{1}{2}(G-1)Bh\nu_0\text{NF}$, where ν_0 is the carrier frequency, B is the signal bandwidth, h is Planck constant, and NF is the amplifier's noise figure.

In our experiments, we simulate the transmission system using Python, specifically employing the SSFM as detailed in chapter 2. Our simulations adopt a symmetric SSFM configuration, where the nonlinear step is applied in the middle of the two linear half-steps At the receiver, a polarization-diversity coherent receiver converts the optical signal to the electrical domain. A low-pass filter with the same bandwidth of the central WDM channel is applied to the signal, obtaining $\mathbf{q}_0(z = \mathcal{L}, t)$. The resulting signal is sampled at

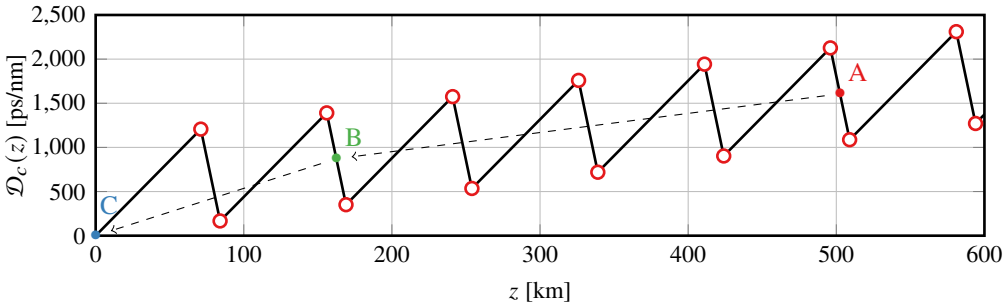


Figure 5.2: Dispersion map of the propagating optical signal over the first 7 spans, with two DBP steps shown. One DBP step is applied over 4 spans, while another is applied over 2 spans, with nonlinear operators applied at points A and B. Accumulated dispersion is fully compensated at point C.

2 samples/symbol, and translated into four electrical signals corresponding to the I and Q components of each polarization, which are then processed by the DSP chain [89]. The DSP chain consists of the following components.

1. *CD compensator (CDC)*, which reverses the CD as

$$\hat{q}_x(\mathcal{L},\omega) \rightarrow \exp\Big(-j\frac{\bar{\beta}_2}{2}\omega^2\mathcal{L}\Big)\hat{q}_x(\mathcal{L},\omega), \tag{5.3}$$

   where $\bar{\beta}_2$ is the mean group velocity dispersion in the link. According to Eq. (5.3), chromatic dispersion compensation (CDC) requires processing the signal in frequency domain using fast Fourier transform (FFT), followed by inverse fast Fourier transform (IFFT) to transform the signal back to the time domain. An alternative implementation of the CDC in the time domain can be realized using finite frequency response (FIR) filters, which we will discuss in Section 5.4.

2. *multiple input multiple output (MIMO) equalizer*, which compensates the time-varying PMD and the random state of polarization (SOP) in the channel. The adaptive equalization of both effects requires using a set of four real-valued FIR filters, which together preform the inverse of the Jones matrix of the dynamic channel in Eq. (2.27). The outputs of the MIMO equalizer are given by

$$x_{\text{MIMO,out}}[k] = \mathbf{h}_{xx}^H\,\mathbf{x}_{\text{in}}[k] + \mathbf{h}_{xy}^H\,\mathbf{y}_{\text{in}}[k] \tag{5.4a}$$

$$y_{\text{MIMO,out}}[k] = \mathbf{h}_{yx}^H\,\mathbf{x}_{\text{in}}[k] + \mathbf{h}_{yy}^H\,\mathbf{y}_{\text{in}}[k]\,, \tag{5.4b}$$

   where $\mathbf{h}_{xx}$, $\mathbf{h}_{xy}$, $\mathbf{h}_{yx}$, and $\mathbf{h}_{yy}$ are vectors of size ξ, representing the taps of the FIR filter, and $\mathbf{x}_{\text{in}}$ and $\mathbf{y}_{\text{in}}$ are sliding windows of the signal after CD compensation. These windows have a length of ξ, and they can be defined as $\mathbf{x}_{\text{in}}[k] = [x_{\text{in}}[k], x_{\text{in}}[k-1], \ldots, x_{\text{in}}[k-\xi]]$ and $\mathbf{y}_{\text{in}}[k] = [y_{\text{in}}[k], y_{\text{in}}[k-1], \ldots, y_{\text{in}}[k-\xi]]$. We employ the radially directed equalizer in conjunction with the constant modulus algorithm

(CMA) proposed in [89] to determine the filter taps of the MIMO equalizer, as detailed in Eq. 2.58. The output sequences of the MIMO are sampled at a rate of 1 sample per symbol.

3. *Carrier phase estimation (CPE)*: The last step in the DSP is to estimate the PN, ϕ_N, resulting from the phase fluctuations between the local oscillator at transmitter and receiver ends [85]. Within a single coherence time, the PN varies slowly compared to the signal, and can thus assumed to be constant. Therefore,

$$x_{\text{sym}}[k] = x_{\text{CPE,in}}[k] \exp(-j\phi_N), \tag{5.5}$$

where $x_{\text{CPE,in}}$ is the same as the output of the MIMO, $x_{\text{MIMO,out}}$.

The aforementioned optical fiber receiver only concerns equalizing the linear effects of the signal resulting from the dispersive channel and PMD.

5.2.2 Digital Back-Propagation

In Fig. 5.1, the DBP block takes the place of the CDC in the receiver and operates at the same sampling rate of 2 samples/symbol. DBP employs the SSFM with negative propagation parameters and larger spatial segments compared to the SSFM. This approach helps to mitigate the high complexity typically associated with a large number of segments in the SSFM, while maintaining accurate signal reconstruction. In practice DBP is limited to 3 or less StpS.

Let us assume that the time-sampled vector $\mathbf{U}^{(n)} = [\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}]^T$ represents a discretized version of the propagating dual-pol signal $\mathbf{q}(z,t)$. Here, $\mathbf{X}^{(n)} \in \mathbb{C}^N$ and $\mathbf{Y}^{(n)} \in \mathbb{C}^N$. The superscript (n) refers to the DBP step, with $n = 0$ representing the input to the DBP, and $n = N_d$ representing the output of the DBP. The linear step is represented by a matrix multiplication, given by

$$\mathbf{U}^{(n)} \rightarrow \mathbf{B}\mathbf{U}^{(n-1)} = \mathbf{W}^{-1}\text{diag}(e^{\delta_d H_1}, ..., e^{\delta_d H_n})\mathbf{W}\mathbf{U}^{(n-1)}, \tag{5.6}$$

where $\mathbf{B} \in \mathbb{C}^{N\times N}$ is a matrix representing the linear operator, $\mathbf{W}$ denotes the discrete Fourier transform matrix, $H_k = -j\beta_2\omega_k^2/2$, and $\omega_k = 2\pi f_k$, where f_k corresponds to the k-th discrete frequency. The nonlinear step is represented by the nonlinear transformation $K(\cdot)$

$$\mathbf{U}^{(n)} \rightarrow K\left(\mathbf{U}^{(n-1)}\right) = \mathbf{U}^{(n-1)} \exp\left(-j\gamma\varepsilon\delta_{\text{eff}}\frac{8}{9}\left(\mathbf{X}^{(n-1)} \odot \mathbf{X}^{(n-1)*} + \mathbf{Y}^{(n-1)} \odot \mathbf{Y}^{(n-1)*}\right)\right). \tag{5.7}$$

where $\odot$ denotes the Hadamard product. Additionally, a real-valued parameter $\varepsilon \in [0,1]$ is introduced, to accurately model the nonlinear effects. The optimization of this parameter will be discussed in the results section.

The DBP algorithm we use in our work allows for flexible selection of StpS values less than 1, enabling us to include multiple spans in a single DBP step, similar to what is considered in [42, 81, 92]. To model the accumulated dispersion in the fiber, we denote by $\mathcal{D}_c(z)$ the total accumulated dispersion inside the fiber as a function of distance z; see Fig. 5.2. The linear step is adjusted as follows: Let us assume $N_d + 1$ spatial steps, dividing a fiber of length $\mathcal{L}$ into spatially equal segments with step size $\delta_d = \mathcal{L}/N_d$, with the exception of the first and last steps where each has length $\delta_d/2$. Each step spans $[z_k, z_{k+1}]$, where $z_k = (k - \frac{1}{2})\delta_d$, $k \in \{1, \ldots, N_d\}$, $z_0 = 0$, $z_{N_d+1} = \mathcal{L}$. This configuration is similar to the Wiener-Hammerstein model in [78]. Within each step, we calculate the weighted-average dispersion coefficient, which is described by the following equation

$$\bar{D} = \frac{\mathcal{D}_c(z_k) - \mathcal{D}_c(z_{k-1})}{\delta_d}. \tag{5.8}$$

The equation Eq. (5.8) is essentially approximating the dispersion map between points z_k and z_{k+1} with a linear dispersion map with a $\bar{D}$ that is between the values of D for SMF and DCF. The power injected at the input of the DCF is set small enough to guarantee a quasi-linear transmission regime. As a consequence, the nonlinear step is performed with the coefficient $\bar{\gamma} = \gamma_{\text{SMF}}$ determined by the SMF. This approximation is accurate, as shown in the numerical simulations that will be presented in Sec. 5.3.2. The linear and nonlinear steps in the proposed DBP alternate until the algorithm spans over the entire optical link.

5.2.2.1 Time Domain and Frequency Domain Implementation of DBP

The conventional method for implementing DBP involves using FFT and IFFT for each step, which can be computationally demanding due to the numerous FFTs and IFFTs required. However, considering the relatively low accumulated dispersion at the receiver in DM systems, we are interested in investigating whether a time-domain implementation of DBP could provide a complexity advantage over the FD approach. In this time-domain implementation, we replace the parameter $\mathbf{B}$ in Eq. 5.6 by employing an FIR filter with complex-valued taps, denoted as $h_{\text{CDC}}(\delta_d)$. This filter performs circular convolution with the backpropagating signal to compensate for the dispersion introduced within a step of length δ_d. The DBP step in this case can be represented as [35]

$$\mathbf{q}(z + \delta_d, t) = (\mathbf{q}(z, t) * h_{\text{CDC}}(\delta_d)) \cdot \exp(\alpha\delta_d/2) \cdot \exp(-j\delta_{\text{eff}}\varepsilon\gamma||\mathbf{q}||^2), \tag{5.9}$$

where $h_{\text{CDC}} = (h_{-F}, \ldots, h_{-1}, h_0, h_1, \ldots, h_F)$, and each h_i for $i = 1, 2, \ldots, F$ denotes an individual tap. It is worth noting that the filter taps exhibit symmetry, such that $h_i = h_{-i}$. The minimum number of taps needed in the FIR filter to compensate for dispersion within a DBP step is mainly determined by the length of the impulse response, approximated using the formula provided in [98]

$$\tau_{CD}(\delta_d) = \frac{\lambda_c^2}{c}|D_{acc}|\Delta f, \tag{5.10}$$

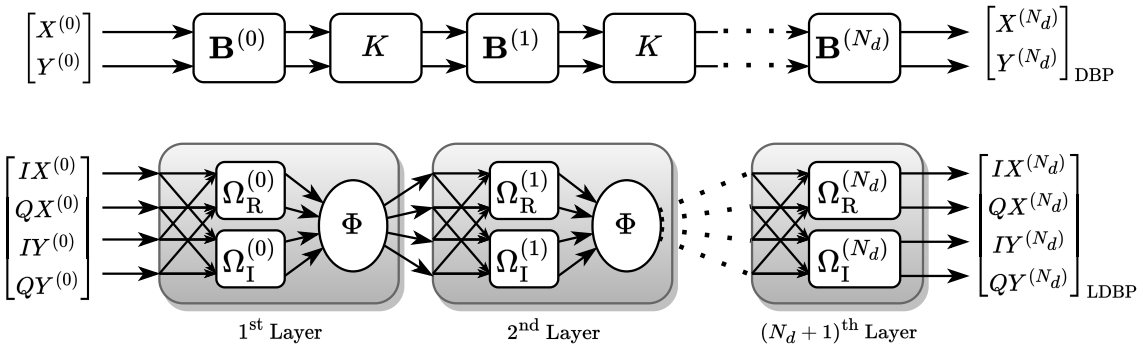


Figure 5.3: Block diagram of DBP (upper branch) and LDBP (lower branch) structures. The following symbols represent: IX $\coloneqq \mathrm{Re}\{\mathbf{X}\}$, QX $\coloneqq \mathrm{Im}\{\mathbf{X}\}$, IY $\coloneqq \mathrm{Re}\{\mathbf{Y}\}$, and QY $\coloneqq \mathrm{Im}\{\mathbf{Y}\}$.

where $|D_{acc}| = \bar{D}\,\delta_d$ is the total accumulated dispersion inside a step with length δ_d, Δf is the signal spectral width for a single channel, $\lambda_c = c/f_c$ is the carrier wavelength, and c is the speed of light. The channel impulse response length $\tau_{CD}(\delta_d)$ is measured in seconds, in which case, the number of filter taps for any step is

$$N_{CDC,\delta_d} = \left\lceil \frac{\tau_{CD}(\delta_d)}{T_s} n_s \right\rceil, \tag{5.11}$$

and $\lceil x \rceil$ denotes the smallest integer larger or equal to x, and n_s is the oversampling ratio.

5.3 Learned Digital Back-Propagation for DM systems

As discuss previously in chapter 4, LDBP is a neural network model based on the computational graph of the DBP. In our application for signal equalization, we specifically utilize the convolutional NN model. Within this model, the linear operator $\mathbf{A}^{(k)}(\mathbf{c})$ in each layer is defined as $\mathbf{A}^{(k)}(\mathbf{c}) = \mathbf{c} * \mathbf{\Omega}^{(k)} + \mathbf{b}^{(k)}$, where $\mathbf{\Omega}^{(k)} \in \mathbb{C}^m$ represents the convolutional filter, and $\mathbf{b}^{(k)} \in \mathbb{C}^N$ is the bias vector.

5.3.1 Architecture of LDBP

Fig. 5.3 depicts the block diagrams of LDBP and its blueprint DBP. The NN architecture of LDBP is a complex-valued network that comprises two real-valued networks operating jointly. Each of these networks contains $N_l = N_d$ layers and accepts four input vectors (IX, QX, IY, and QY). Each layer of the network consists of two parallel convolutional filters, Ω_R and Ω_I, corresponding respectively to the real and imaginary parts of the filter, and a nonlinear function Φ, which takes four input vectors and generates four output vectors. The number of non-zero weights in the convolutional filter is determined numerically, as will be described in the simulation setup section later. These layers perform the real-valued equivalent of the operation described in Eq. (5.9) in the time domain without the need for

FFT and IFFT. Biases are not utilized in our model and are set to zero. In our model, we do not use the biases and therefore set them to zero. The input dimension of LDBP is $[N_{ex}, N, 4]$, and its output dimension is $[N_{ex}, N - 2M, 4]$, where M represents the memory of the dispersive channel.

To train the LDBP, we simulated a five-channel WDM PMD-free transmission of a block of 2^{15} symbols at various launch powers. The signal was initially sampled at 16 samples per symbol duration for forward propagation using SSFM. However, after the WDM demultiplexer, the signal from the central channel was sampled at a sampling ratio of $n_s = 2$ samples per symbol, resulting in a received block consisting of 2^{16} samples. The NN operates in a sliding window fashion, with a window size of N sliding over the transmission block and advancing by $L \times n_s$, where the sliding factor L is an integer that determines the number of symbols shifted between examples. For LDBP training, we generated input-output pairs with a size of 1024 samples (512 symbols) for the input, and the corresponding output was 852 samples (426 symbols) long due to the dispersive channel effects (43 symbols on each side). We set the sliding factor to $L = 8$, generating $N_{ex} = 4021$ input-output pairs for training. To test LDBP, we generated 8 transmission blocks with PMD using the same overlapping and shifting technique as for training. However, this time, the sliding factor was set to $L = 426$ to ensure that each symbol in the transmission block was detected exactly once. The output of LDBP was then passed to the DSP to equalize PMD and dynamic channel effects.

We implement a symmetric DBP as the blueprint of LDBP, such that all layers are initialized with the corresponding parameters in a linear step of DBP at the full step size δ_d, except for the first and last layers which correspond to a half step $\delta_d/2$. The NN is trained by minimizing the mean squared error (MSE) loss function, using the *Adam* optimizer with a learning rate of 0.001. During training, 20% of the training examples were used for validation to monitor the LDBP's progress. The LDBP was trained for up to 75 epochs, with an early-stop condition triggered if the validation error did not decrease within 5 epochs. The best-performing epoch's weights were used in the final LDBP. After training at each launch power, we evaluate the performance of the LDBP by calculating the Q-factor using independently generated testing data.

5.3.2 Simulated System Setup and Performance Results

The performance results are based on the simulation of the transmission system shown in Fig. 5.1. All elements of the transmission system, including the transmitter, receiver, and channel, are simulated in Python, while the NN is implemented using the TensorFlow library. In this section, we present the performance results of DBP and LDBP for four different setups: (A), (B), (C), and (D). Setup (A) represents a single-channel transmission system with 16-QAM modulation, while setups (B) and (C) are WDM transmission

	# of channels	Modulation format	fiber coefficients
Setup (A)	1	16-QAM	α=0.2 dB/km, PMD = 0.05 ps/$\sqrt{\text{km}}$, D=17 ps/(nm.km), γ= 1.4/W/km
Setup (B)	5	16-QAM	Same as above
Setup (C)	5	64-QAM	Same as above
Setup (D)	5	16-QAM	α=0.24 dB/km, PMD = 0.3 ps/$\sqrt{\text{km}}$, D=17 ps/(nm.km), γ= 1.4/W/km

Table 5.1: Description of the simulated setups.

systems with 16-QAM and 64-QAM modulation formats, respectively. Setup (D) is also a WDM transmission using 16-QAM modulation, but it includes aging effects where a fiber channel undergoes aging. The aging study will be described in detail when we present the performance results for this setup. Table 5.1 provides a detailed description of each setup.

For all setups (A)–(D), the transmission symbol baud rate $B = 32$ GBaud for each channel using a root-raised-cosine (RRC) pulse-shape with a roll-off factor $\rho = 0.06$. The optical fiber link consists of $N_{sp} = 28$ spans, each span including an SMF and a DCF measuring 72 km and 13 km, respectively. The length of DCF is chosen such that it compensates for 85% of the CD in each span. An amplifier with gain $G_{\text{SMF}} = 6.5$ dB is applied at the end of the SMF, and a second amplifier with gain of $G_{\text{DCF}} = 14.4$ dB is applied after the DCF. The SMF parameters and PMD value for all setups are provided in Table 5.1. For WDM setups (B)–(D), the WDM channels are separated by a frequency spacing of 37.5 GHz, resulting in a guard band of 5.5 GHz between adjacent channels. The lasers used for these setups had a linewidth of 50 kHz. To avoid overestimation of nonlinear crosstalk, the data symbols of all WDM channels were intentionally made unsynchronized in terms of time, polarization state, and phase. At the receiver, an RRC filter with a bandwidth of $(1 + \rho)B$ is applied to filter out the adjacent channels, such that only the central channel is processed by the DBP or LDBP algorithms. The forward signal propagation using SSFM is simulated with 72 steps for each SMF and 13 steps for each DCF. The signals are sampled at a rate of 16 samples per symbol. At the receiver, the signal is downsampled to twice the symbol rate before being processed by either the CDC, DBP, or LDBP algorithms. Finally, the output of the CDC, DBP, or LDBP is downsampled to one sample per symbol and processed by the conventional DSP chain, which equalizes PMD effects and polarization mixing.

5.3.2.1 DBP Parameters Optimization

To optimize the performance of the DBP for all transmission scenarios (A)–(D), we select the parameter ε in Eq. 5.7 such that each DBP achieves the highest Q-factor at the optimal

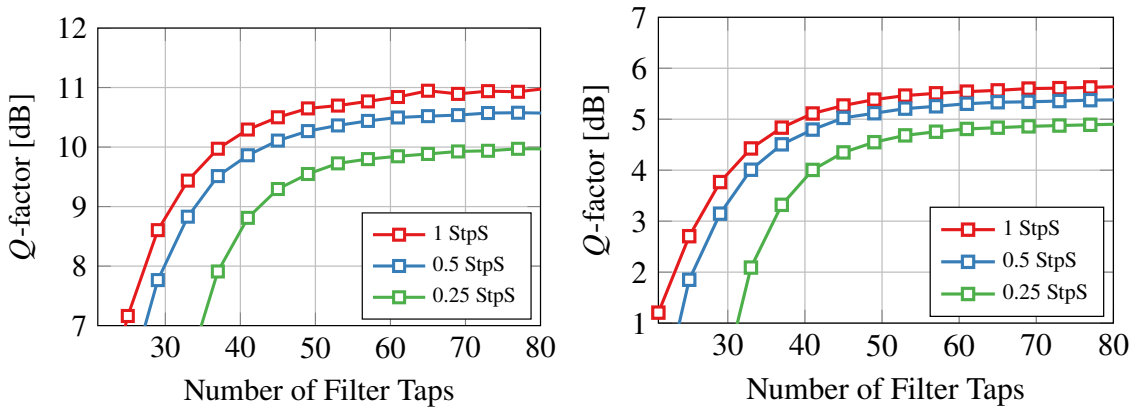


Figure 5.4: Achieved Q-factor by DBP at the optimal launch power (-2 dBm) for different values of StpS as a function of filter taps per step for (a) 16-QAM and (b) 64-QAM.

launch power. For a single channel transmission, we find that the value of ε is 1 for all values of StpS. For the WDM transmission of 5 channels (Set-ups B, C and D), we find that the optimal values are $\varepsilon = 0.85$ for DBP with 1 StpS, $\varepsilon = 0.75$ for DBP with 0.5 StpS, and $\varepsilon = 0.64$ for DBP with 0.25 StpS. It should be noted that $\varepsilon = 0$ corresponds to performing LE, and in this case, all DBP configurations perform similarly to LE, regardless of the number of StpS. The filter size in each DBP step varies depending on the number of StpS, Optimal values of F were determined for different DBP configurations by numerical optimization. Specifically, $F = 16$ was found to be optimal for both 1 StpS and 0.5 StpS DBP, $F = 24$ for 0.25 StpS, $F = 30$ for $\frac{1}{7}$ StpS, and $F = 36$ for $\frac{1}{14}$ StpS. These values represent the minimum required to ensure all DBP configurations outperform LE at all launch powers. The impact of filter width on performance is shown in Fig. 5.4, based on the Q-factor as the performance metric. The Q-factor is defined as follows

$$Q\text{-factor} = 20\log_{10}[\sqrt{2}\text{erfc}^{-1}(2\text{BER})], \tag{5.12}$$

where erfc$(\cdot)$ is the complementary error function.. It is worth noting that this finding agrees with previous literature, specifically [30], which reported that the filter width has a significant impact on the performance of DBP-based equalizers (see Fig.(9) in [30]).

5.3.2.2 Performance Comparison of DBP and LDBP in Single Channel Transmission

A single channel transmission is simulated in setup (A). We choose (SNR_{eff}) as a performance measurement for single channel transmission, which is defined as

$$\text{SNR}_{\text{eff}} = \frac{||\hat{\mathbf{s}}_{\mathbf{x}}||^2+||\hat{\mathbf{s}}_{\mathbf{y}}||^2}{||\mathbf{s}_{\mathbf{x}}-\hat{\mathbf{s}}_{\mathbf{x}}||^2+||\mathbf{s}_{\mathbf{y}}-\hat{\mathbf{s}}_{\mathbf{y}}||^2}, \tag{5.13}$$

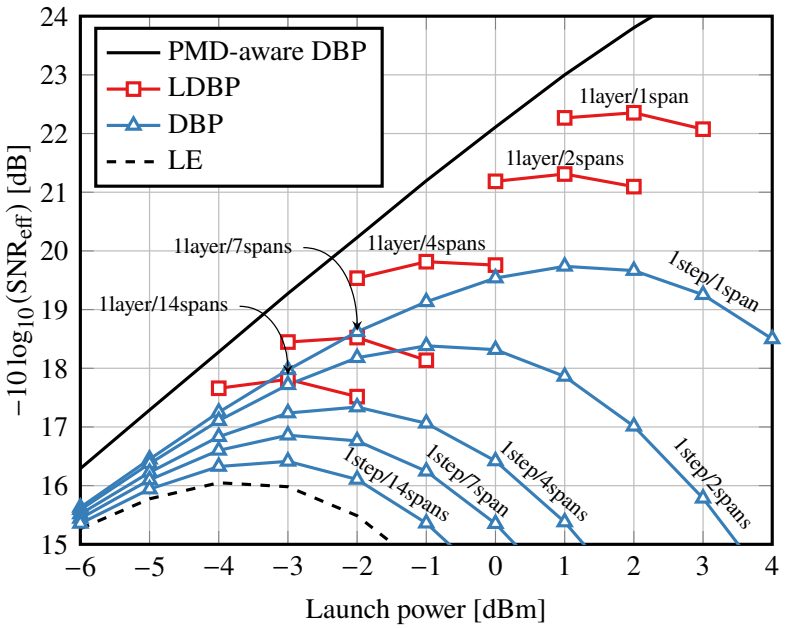
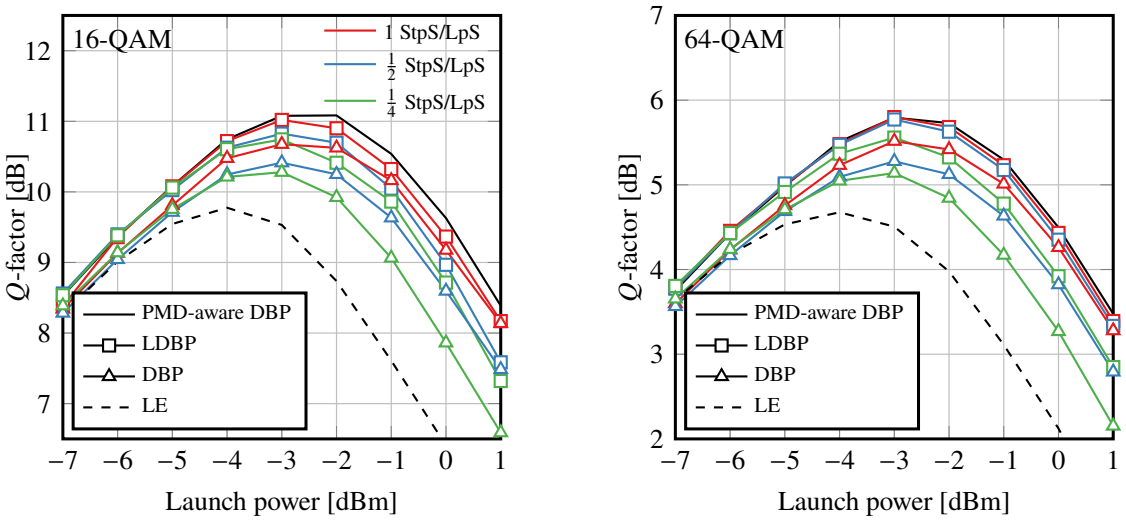


Figure 5.5: DBP performance for different StpS values in single channel transmission (setup A). The performance of LDBP is also shown for 3 launch powers around the peak $\mathrm{SNR_{eff}}$.

where $\hat{\mathbf{s}}_\mathbf{x}$ and $\hat{\mathbf{s}}_\mathbf{y}$ denote the estimated symbols at the output of the DSP, with true values $\mathbf{s_x}$ and $\mathbf{s_y}$, respectively. The choice of $\mathrm{SNR_{eff}}$ is motivated by the high reliability of signal transmission in a single channel, which results in an extremely low bit-error-ratio (BER). Accurately measuring BER becomes challenging when it is very low, as it requires a substantial number of training examples for precision. Under such circumstances, $\mathrm{SNR_{eff}}$ offers a more practical performance metric. The performance of the DM-adapted DBP and LDBP in setup (A) is depicted in Fig. 5.5 for varying signal launch powers. The PMD-aware DBP in Fig. 5.5 is a genie-aided model which assumes perfect knowledge of PMD and SOP across the fiber channel at the receiver and uses them in the back-propagation with the same step size δ_s used for SSFM. Despite its impracticality, the PMD-aware DBP model provides a useful upper bound on the performance that can be achieved by PMD-agnostic DBP and LDBP algorithms. Our simulations consider DBP and LDBP equalizers with a fractional number of StpS or LpS that is less than or equal to 1.

At the optimal launch power, LDBP with 1 LpS outperforms DBP with the same complexity by providing an SNReff of 22.3 compared to 19.8 dB, with the optimal launch powers differing by about 1 dBm between the two algorithms. The LE achieves its best performance at an SNReff value of 16 dB, which is achieved with a launch power of -4 dBm. Both DBP and LDBP outperform the LE with varying gains, with LDBP with 1 LpS achieving the highest gain of 6.3 dB and DBP with 1 StpS achieving a gain of 3.8 dB. The simulated LDBP with the least complexity has 2 layers (1 full step and 2 half steps) and 2 activation functions, and outperforms DBP with similar complexity by 1.4 dB and



(a) (b)

Figure 5.6: Achieved Q-factors for DBP and LDBP with different values of StpS and LpS for WDM transmissions over a 2016 km fiber with 32 Gbaud using (a) 16-QAM modulation, and (b) 64-QAM modulation.

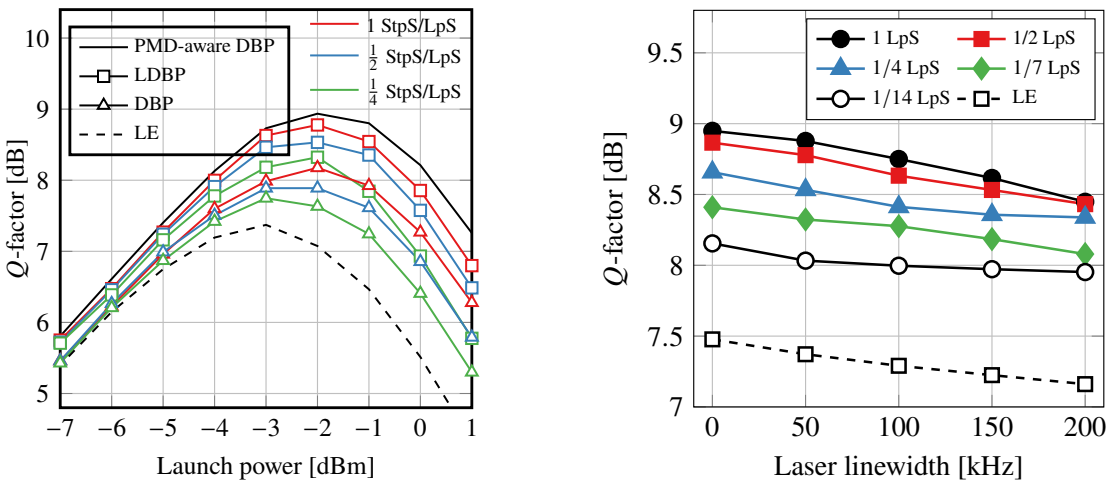
the LE by 1.8 dB.

5.3.2.3 Performance Comparison of DBP and LDBP in Multi-Channel Transmission

Setups (B)–(D) present multiple WDM transmission scenarios. In such cases, the nonlinearity affecting the received signal is dominated by the nonlinear interference introduced by the adjacent channels via XPM. Since only the signal from one single channel is fed to the receiver, the information in adjacent channels is unknown to the receiver. Therefore, the nonlinearity generated by adjacent channels impacts all equalizers and limits their performance in the nonlinear regime. The performance of the PMD-aware DBP equalizer in these setups can be characterized by a bell-shaped curve, as seen in Figures 5.6 and 5.7(b), in contrast to the straight line observed in the single-channel scenario (A) shown in Fig. 5.5.

For setups (B) through (D), we employ the previously described Q-factor. Simulation results for setups (B) and (C) are presented in Fig. 5.6. In these setups, the DBP and LDBP algorithms were simulated with varying numbers of total steps or layers ($M \in 7, 14, 28$).

In setup (B), LDBP with 1 StpS achieved a peak Q-factor of 11 dB at the optimal launch power of $P = -3$ dBm. Notably, the peak performance of LDBP with 1 StpS is comparable to that achieved by the PMD-aware DBP at the same launch power. When comparing LDBP to DBP with 1 StpS, LDBP outperformed DBP by 0.3 dB at the same launch power. Furthermore, LDBP with 0.5 StpS and 0.25 StpS achieved 10.8 dB and 10.7



(a) (b)

Figure 5.7: Achieved Q-factors for LDBP and DBP with different StpS and LpS values for Set-up (D) simulating transmission over a 2016 km aged fiber at 32 Gbaud. Figure (a) shows Q-factors across launch powers, and (b) displays the impact of laser PN on the peak Q-factor of LDBP with varying LpS values.

dB, respectively, both outperforming DBP with similar numbers of StpS by 0.4 dB.

In setup (C), The LDBP with 1 StpS achieved a peak Q-factor of 5.8 dB, which is 0.3 dB higher than the peak performance achieved by the DBP with a similar number of StpS. The LDBP with 0.5 StpS and 0.25 StpS achieved a peak Q-factor of 5.8 dB and 5.6 dB, respectively, both outperforming the DBP with a similar number of StpS by 0.5 dB. On the other hand, the LE achieved a peak Q-factor of 4.7 dB.

It is worth noting that the launch powers corresponding to the peak Q-factor in both Fig.5.5 and Fig.5.6 are shifted by 4 dB for WDM transmission compared to single-channel transmission. This is due to the XPM generated by the four adjacent channels, which dominates the nonlinearity affecting the received signal in WDM setups, unlike single-channel transmission, which only experiences self-phase modulation (SPM) from the same channel, as explained in [29].

5.3.2.4 Impact of Aging Effects on LDBP Performance

Optical fiber links deployed for intensity modulation and direct detection (IM-DD) communication, especially those utilizing dispersion management, are particularly susceptible to aging and poor maintenance due to being outdated systems [11, 14]. Changes in the fiber material occur as the fibers age, leading to the deterioration of their parameters over time. Additionally, yearly temperature variations can contribute to higher values of differential group delay (DGD). Additionally, fiber splicing, which is done during maintenance, can increase attenuation and necessitate higher amplification gain in the link, resulting in ASE

noise. To reflect these changes in a more realistic transmission scenario, we consider the impact of aging and maintenance-induced changes on the fiber link. It is worth noting that the CD coefficient β_2 and the nonlinearity coefficient γ are not affected by aging effects. While only these two coefficients are used to initialize the LDBP, the new attenuation coefficient α in aged fibers can cause a reduction in the effective length L_{eff}. Thus, retraining the NN with the updated parameters is necessary. Our goal is to examine how these changes affect the LDBP's performance relative to DBP, and compare the two methods in a realistic transmission scenario that incorporates aging and maintenance-induced changes in the fiber link.

The fiber parameters selected for setup (D) aim to simulate aging effects and consist of an attenuation coefficient of $\alpha = 0.24$ dB/km and a PMD coefficient of 0.3 ps/km$^{-1/2}$. The value of the attenuation coefficient after degradation is based on experimental findings reported in [14]. The simulation results for this particular setup are presented in Figure 5.7. To account for the changes in the fiber link, we retrained the model using data generated from simulations with the new parameters. The LDBP with 1 StpS achieved a peak Q-factor of 8.8 dB at a launch power of $P = -2$ dBm, which is 0.6 higher than the peak Q-factor of the DBP with similar complexity. Furthermore, we compared setup (D) with setup (B), which simulates a similar transmission scenario without aging effects. We observed an average drop in Q-factor of 2.2 dB across all graphs, but the Q-factor gain of LDBP over DBP was still maintained. This suggests that deploying the LDBP long-term is feasible with retraining of the model.

The existing DM systems deployed for IM/DD may not be optimized for coherent transmission. In fact, lasers used in IM/DD systems typically have higher laser PN than those used in coherent transmission, which can potentially impact the performance of receivers used for coherent detection [33]. In general, laser phase noise does not have a significant impact on the system, as the time interval over which the laser phase noise changes is much longer than the symbol period. However, to investigate its impact on the performance of LDBP for set-up (D), we choose to retrain the neural network using data that incorporates varying degrees of laser PN. While it is the primary role of the DSP following the LDBP to mitigate dynamic effects like laser PN, when the LDBP is exposed to examples of data affected by PN, it may be possible for the neural network to learn and reduce the impact of PN to some extent. The peak Q-factor achieved for different values of LpS and a range of laser linewidths is shown in Fig. 5.7(b). Since the input vector for LDBP has a width of 512 symbols, the deviation in PN for the laser linewidth used in the simulations is approximately 0.01 radian, which is too small to significantly impact the performance. The results indicate that the average drop in Q-factor is 0.1 dB for a laser linewidth of 100 kHz and 0.2 dB for a laser linewidth of 200 kHz.

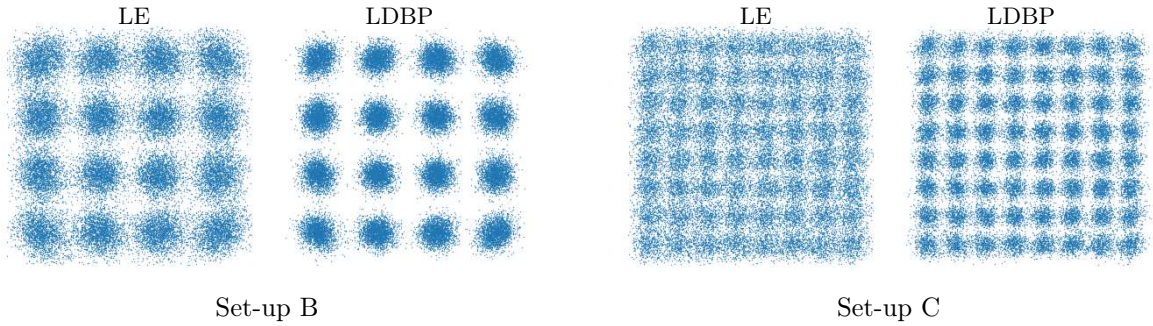


Figure 5.8: Constellation diagrams for LE and LDBP at their respective optimal launching powers for Set-ups B&C. These constellations were obtained at the end of the DSP prior to hard-decision.

5.4 Complexity of DBP and LDBP

We quantify the complexity of each step in terms of the number of real multiplications per symbol (RMpS), excluding additions. It is important to note that both DBP and LDBP exhibit the same complexity when considering the same number of StpS and LpS. Therefore, the complexity formulas derived for LDBP in this section are identical to those for DBP.

To efficiently compute the exponential function in the activation function, various approximation algorithms can be employed, such as the CORDIC algorithm [43, 76]. These algorithms utilize look-up tables and bit-shifts, eliminating the need for multiplications. By employing similar algorithms, the activation function can be computed efficiently, with each computation of the activation requiring only 9 real multiplications (RMs).

We make a distinction between the complexities of TD-LDBP and FD-LDBP and present their respective complexity formulas as follows:

FD-LDBP Complexity The total complexity of FD-LDBP per detected symbol, can be calculated in RMpS for a signal with input size N and sampled at rate n_s samples/symbol. The complexity is given by [97]

$$C_{\text{FD–LDBP}} = (N_d + 1)\left(4\frac{N(\log_2(N) + 1)n_s}{N - 2N_{CDC,\mathcal{L}} + 1}\right) + \frac{9}{2}N_d n_s. \tag{5.14}$$

TD-LDBP Complexity The complexity of TD-LDBP for each detected symbol is determined by the convolution of a kernel with size F and an input of size N, which involves $4FN$ RMs. This computation assumes a dilation of 1, stride of 1, and padding of $F - 1$, resulting in an output size that is the same as the input size for each convolutional layer. Therefore, the total complexity per detected symbol for TD-LDBP can be calculated as

$$C_{\text{TD–LDBP}} = (N_d + 1)\left(4\frac{(2F + 1)Nn_s}{N - 2N_{CDC,\mathcal{L}} + 1}\right) + \frac{9}{2}N_d n_s. \tag{5.15}$$

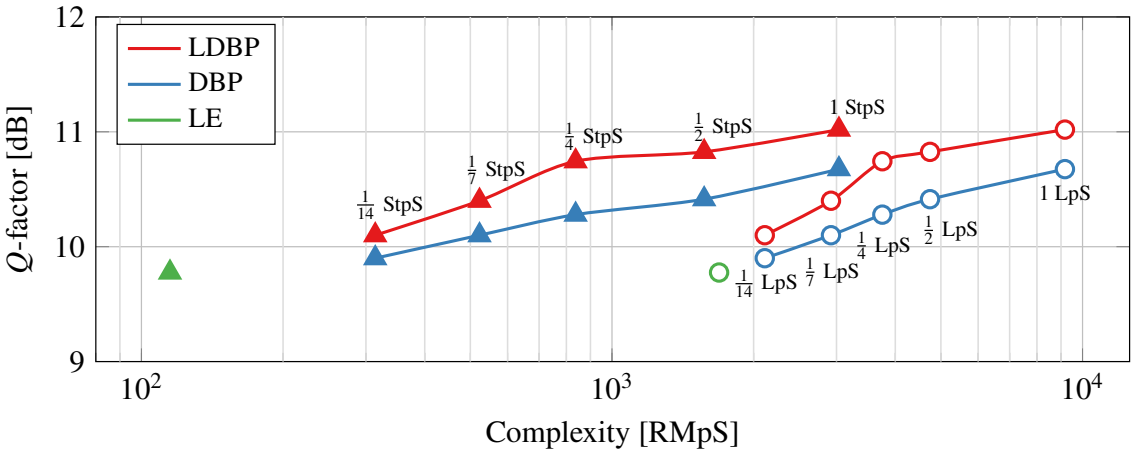


Figure 5.9: Trade-off between complexity and performance for different equalizers in both FD-LDBP (represented by triangles) and TD-LDBP (represented by circles).

Fig. 5.9 depicts the RMpS complexity of DBP and LDBP using both TD and FD implementations. The complexity of FD-LDBP (or FD-DBP) is primarily dependent on the number of steps involved in the algorithm, and can be approximated by 114 times the number of FFT/IFFT uses. The complexity of 1 LpS FD-LDBP, which uses 29 FFT/IFFT pairs, is approximately 3300 RMpS. On the other hand, TD-LDBP with the same number of LpS has a complexity of around 14000 RMpS. In fact, FD-LDBP is less computationally complex compared to TD-LDBP across all values of LpS. We initially hypothesized that, due to the lower accumulated dispersion in the DM system compared to NDM, smaller filters would be sufficient for each linear step in the DM system. However, our results indicate that TD implementation did not provide any complexity advantage over FD implementation. We hypothesize that this is due to the accumulation of truncation errors with each subsequent step in the DBP, resulting in a high overall error. As a result, larger filters are required to mitigate the truncation error and reduce its impact. We determined the number of filter taps required in each step through numerical simulations, as demonstrated by Fig. 5.4. Nevertheless, TD implementation restricts the neural network to a lower number of trainable parameters, which simplifies the training process and enhances training convergence even with smaller training datasets.

5.5 Summary

In this chapter, we have presented the LDBP approach for mitigating nonlinear effects in DM optical fiber transmission systems. LDBP is a DBP optimization method that leverages NNs training algorithms to optimize the DBP parameters. Our comparative study has shown that LDBP outperforms DBP, providing a significant gain in Q-factor, with an average improvement of 0.4 dB. The application of NNs to DM links is an important new

development, as it demonstrates the possibility of repurposing DM systems for coherent transmission. The results have significant implications for the fiber-optics industry, suggesting that data rates in conventional DM optical links can be substantially improved using modern, simple coherent receivers. Additionally, we demonstrated the complexity of both DBP and LDBP in the TD and FD implementations. TD-implementation was found to be more complex across all values of LpS. Overall, our findings highlight the potential of LDBP as an effective method for mitigating nonlinear effects in DM optical fiber transmission systems.

Conclusions and Perspectives

Conclusions

In today's communication landscape, where the demand for high data rates continues to surge, optical fiber network operators must address the challenge of nonlinearity–an inherent feature of optical fiber networks. This need arises because achieving and sustaining these high data rates necessitates the use of advanced modulation formats. However, these formats require increased transmission power, leading to distortions caused by Kerr nonlinearity that significantly affect achievable information rates (AIRs). Traditional methods for mitigating nonlinearity, like digital back-propagation (DBP), have proven impractical, with DBP rarely deployed in commercial systems to date. To tackle this challenge, optical network operators are exploring innovative tools, such as neural networks. Neural networks have demonstrated their effectiveness in various complex tasks across different applications in artificial intelligence, including image recognition and complex language models.

This work lies at the intersection of optical fiber communications, with a particular focus on coherent optical receiver design, and machine learning models, specifically neural networks. It explores the application of neural network models to mitigate nonlinear effects affecting optical signals transmitted through optical fiber channels.

This thesis begins with an introduction and a motivation of the problem in Chapter 1, followed by a comprehensive review in Chapter 2, delving into optical fiber physics, channel effects on signal propagation, optical signal generation, and the optical fiber receiver, including its various blocks for equalization and symbol detection. This section addresses the first aspect of the thesis title: optical fiber systems.

Chapter 3 discusses the second aspect of the thesis, focusing on the fundamentals of neural networks and highlighting two known approaches: the model-agnostic and the model-based approaches. This chapter also includes a brief literature review of recent advances in optical channel equalization using neural networks. Model-agnostic neural networks have been the subject of many studies in the literature on neural network equalizers. These models offer design and hyperparameter flexibility and are flexible with their position within the receiver architecture. However, they often exhibit unnecessary size and

high complexity, posing challenges for practical implementation.

Model-based neural networks draw inspiration from the physical model of signal propagation and tend to be more compact, performing well with careful initialization. We focus specifically on the LDBP model-based approach in Chapter 4. Here, we introduce the "parameter sharing" method, reducing the number of trainable parameters in the neural network, which enhances training and enables better convergence with smaller datasets. The LDBP with parameter sharing achieved a 3–3.5 dB reduction in mean squared error (MSE) compared to optimized DBP with comparable complexity and reduced the number of trainable parameters by a factor of five compared to conventional LDBP with the same MSE.

In Chapter 5, we explore the potential of LDBP in transitioning dispersion-managed (DM) systems from intensity modulation with direct detection (IM/DD) to coherent detection. This transition extends the AIR achievable in these systems. Our comparative study shows that LDBP outperforms DBP in DM systems, providing a significant Q-factor improvement, averaging 0.4 dB. This application of neural networks to DM links signifies a significant development, enabling the repurposing of DM systems for coherent transmission. We do not claim that DM systems surpass the commonly used non-dispersion-managed (NDM) systems. Our objective is to investigate the feasibility of upgrading deployed DM systems, primarily used for IM/DD, to enable coherent transmission. These findings hold significant implications for the fiber-optics industry, suggesting that data rates in traditional DM optical links can be substantially improved using modern, straightforward coherent receivers. Additionally, we analyzed the complexity of both DBP and LDBP using two techniques: time-domain (TD) and frequency-domain (FD) implementation. Contrary to our initial assumption, TD implementation was found to be more complex across all layers/spans. In summary, our findings underscore the potential of LDBP as an effective method for mitigating nonlinear effects in DM optical fiber transmission systems.

Perspectives

Despite the promising results of neural network-based approaches in the context of nonlinearity mitigation, particularly those achieved by learned digital back-propagation (LDBP), a notable limitation becomes evident: These methods are designed for processing signals from a single channel, making them well-suited for single-channel transmission scenarios but less suitable for multi-channel wavelength-division multiplexing (WDM) configurations. To grasp the gravity of this limitation, let us consider a scenario where 20 channels are co-propagating alongside the channel of interest in a WDM setup. In this context, most of the nonlinearity affecting the channel of interest is due to cross-phase modulation (XPM), caused by neighboring channels that co-propagate with the channel of interest, while a

minority of it results from self-phase modulation (SPM), caused by the signal itself.

The existing literature on neural network equalization has predominantly focused on addressing nonlinear impairments in single-channel transmission, yielding remarkable performance improvements compared to conventional equalization methods. For instance, studies have reported a 2.9 dB gain with CNN+biLSTM over linear equalization (LE) [39], and a 2 dB gain with complex-valued neural networks over LE [38]. However, there is a notable scarcity of literature addressing equalization in the context of WDM transmissions. The few studies that have explored this area, such as [30], have reported more modest performance gains, like the 0.4 dB improvement achieved with generalized digital back-propagation (GDBP) compared to LE [30].

There's a critical need for neural networks capable of jointly processing adjacent channel signals to model external nonlinear effects. However, as channel numbers increase, complexity grows exponentially. Innovative machine learning models, like the Siamese neural network (SNN) in [77], alleviate this complexity. SNNs use symbols from the main channel in one branch while extracting meaningful features from adjacent channels for input into other branches. LDBP can draw inspiration from the SNN by employing multiple input branches, each fed with input consisting of features extracted from adjacent channels to aid the decision-making process.

Another promising approach to mitigate fiber nonlinear impairments is the utilization of random forests, as proposed in [63, 105, 107]. In particular, [63] demonstrates its effectiveness in an experimental setting, specifically in a 120 Gb/s, 375 km, dual-polarization 64-quadrature amplitude modulation (QAM) optical fiber communication platform. Random forests divide constellation points into two regions using a specific decision threshold, then uses different random forest equalizers to compensate points in these distinct regions. This innovative strategy significantly reduces system complexity, as it relies only on comparisons, achieving a decision with 0 multiplications. Moreover, the random forest-based equalizer proves feasible for real-world optical fiber communication systems due to its simple design. LDBP can significantly benefit from innovative approaches like random forests for extracting meaningful features from adjacent channels with very low complexity. These features can then be used in the LDBP algorithm to make more informed decisions, enhancing its performance beyond analyzing signals solely from the channel of interest.

Similar innovations are key enablers of significantly reducing the complexity associated with ML-based equalization. They pave the way for less computationally demanding methods capable of simultaneously processing multiple signals, thereby enhancing the modeling of nonlinearity originating from adjacent channels without introducing additional complexity.

Bibliography

[1] Mohannad Abu-romoh, Nelson Costa, et al. "Low Complexity Convolutional Neural Networks for Equalization in Optical Fiber Transmission". In: *OSA Advanced Photonics Congress (AP) 2021*. Optical Society of America, 2021.

[2] Mohannad Abu-romoh, Nelson Costa, et al. "Learned Digital Back-Propagation for Dual-Polarization Dispersion Managed Systems". In: *Proc. Eur. Conf. Opt. Commun.* Basel, Switzerland, 2022, We1C.6. URL: https://opg.optica.org/abstract.cfm?URI%20=%20ECEOC-2022-We1C.6.

[3] Mohannad Abu-romoh, Nelson Costa, et al. "Equalization in dispersion-managed systems using learned digital back-propagation". In: *Opt. Continuum* 2.10 (Oct. 2023), pp. 2088–2105. URL: https://opg.optica.org/optcon/abstract.cfm?URI=optcon-2-10-2088.

[4] Charu C. Aggarwal. *Neural Networks and Deep Learning. A Textbook*. Cham: Springer, 2018, p. 497.

[5] G.P. Agrawal. *Fiber-Optic Communication Systems*. Wiley, 2021. URL: https://books.google.fr/books?id%20=%20RkwvEAAAQBAJ.

[6] Govind Agrawal. "Chapter 1 - Introduction". In: *Nonlinear Fiber Optics (Fifth Edition)*. Ed. by Govind Agrawal. Fifth Edition. Optics and Photonics. Boston: Academic Press, 2013, pp. 1–25. URL: https://www.sciencedirect.com/science/article/pii/B9780123970237000012.

[7] Govind Agrawal. "Chapter 2 - Pulse Propagation in Fibers". In: *Nonlinear Fiber Optics (Fifth Edition)*. Ed. by Govind Agrawal. Fifth Edition. Optics and Photonics. Boston: Academic Press, 2013, pp. 27–56. URL: https://www.sciencedirect.com/science/article/pii/B9780123970237000024.

[8] Govind P. Agrawal. "Chapter 6 - Polarization Effects". In: *Nonlinear Fiber Optics (Fifth Edition)*. Ed. by Govind Agrawal. Fifth Edition. Optics and Photonics. Boston: Academic Press, 2013, pp. 193–244. URL: https://www.sciencedirect.com/science/article/pii/B9780123970237000061.

[9] Rameez Asif, Chien-Yu Lin, et al. “Optimized digital backward propagation for phase modulated signals in mixed-optical fiber transmission link”. In: *Opt. Express* 18.22 (Oct. 2010), pp. 22796–22807. URL: https://opg.optica.org/oe/abstract.cfm?URI%20=%20oe-18-22-22796.

[10] Rameez Asif, Chien-Yu Lin, et al. “Digital Backward Propagation: A Technique to Compensate Fiber Dispersion and Non-Linear Impairments”. In: *Applications of Digital Signal Processing*. Ed. by Christian Cuadrado-Laborde. Rijeka: IntechOpen, 2011. Chap. 2. URL: https://doi.org/10.5772/25410.

[11] B. Bakhshi, L. Rahman, et al. “Impact of Fiber Aging and Cable Repair in an Installed 28-nm Transatlantic 96 x 10 Gb/s DWDM System”. In: *OFC/NFOEC Technical Digest. Opt. Fiber Commun. Conf., 2005.* Vol. 1. 2005, pp. 169–171.

[12] Rick M. Bätler, Christian Häger, et al. “Model-Based Machine Learning for Joint Digital Backpropagation and PMD Compensation”. In: *IEEE J. Lightw. Technol.* 39.4 (Feb. 2021), pp. 949–959.

[13] Bertold Ian Bitachon, Amirhossein Ghazisaeidi, et al. “Deep learning based digital backpropagation demonstrating SNR gain at low complexity in a 1200 km transmission link”. In: *Opt. Express* 28.20 (Sept. 2020), pp. 29318–29334.

[14] J. Bohata, J. Jaros, et al. “Long-Term Polarization Mode Dispersion Evolution and Accelerated Aging in Old Optical Cables”. In: *IEEE Photon. Technol. Lett.* 29.6 (Mar. 2017), pp. 519–522.

[15] M.-L. Boucheret, I. Mortensen, et al. “A new algorithm for nonlinear estimation of PSK-modulated carrier phase”. In: *3rd European Conference on Satellite Communications - ECSC-3, 1993.* 1993, pp. 155–159.

[16] J.-X. Cai, H. Zhang, et al. “200 Gb/s and Dual Wavelength 400 Gb/s Transmission over Transpacific Distance at 6.0 b/s/Hz Spectral Efficiency”. In: *J. Lightwave Technol.* 32.4 (Feb. 2014), pp. 832–839. URL: https://opg.optica.org/jlt/abstract.cfm?URI%20=%20jlt-32-4-832.

[17] Clara Catanese, Reda Ayassi, et al. “A Fully Connected Neural Network to Mitigate 200G DP-16-QAM Transmission System Impairments”. In: *OSA Advanced Photonics Congress (AP) 2020 (IPR, NP, NOMA, Networks, PVLED, PSC, SPPCom, SOF)*. Optical Society of America, 2020, SpTh3I.1. URL: http://www.osapublishing.org/abstract.cfm?URI%20=%20SPPCom-2020-SpTh3I.1.

[18] Clara Catanese, Ahmed Triki, et al. “A Survey of Neural Network Applications in Fiber Nonlinearity Mitigation”. In: *Proc. 21st Int. Conf. Transp. Opt. Netw.* Angers, France, 2019, pp. 1–4.

[19] Gildas Chauvel. *Dispersion in Optical Fibers*. 2008. URL: https://dl.cdn-anritsu.com/en-us/test-measurement/files/Technical-Notes/White-Paper/Disp_in_Opt_Fibers_PMD_CD.pdf.

[20] Andrew R Chraplyvy, Robert W Tkach, et al. *Optical fiber for wavelength division multiplexing*. US Patent 5,327,516. July 1994.

[21] S. Civelli, E. Forestieri, et al. "Coupled-Channel Enhanced SSFM for Digital Backpropagation in WDM Systems". In: *Proc. Opt. Fiber Conf.* San Francisco, CA, USA, 2021, pp. 1–3.

[22] Antoine Cornuéjols and Christel Vrain. "Designing Algorithms for Machine Learning and Data Mining". In: *A Guided Tour of Artificial Intelligence Research: Volume II: AI Algorithms*. Ed. by Pierre Marquis, Odile Papini, et al. Cham: Springer International Publishing, 2020, pp. 339–410. URL: https://doi.org/10.1007/978-3-030-06167-8_12.

[23] Cristian B. Czegledi, Gabriele Liga, et al. "Digital Backpropagation Accounting for Polarization-Mode Dispersion". In: *Opt. Express* 25.3 (Feb. 2017), pp. 1903–1915. URL: https://opg.optica.org/oe/abstract.cfm?URI%20=%20oe-25-3-1903.

[24] Stavros Deligiannidis, Adonis Bogris, et al. "Compensation of Fiber Nonlinearities in Digital Coherent Systems Leveraging Long Short-Term Memory Neural Networks". In: *Journal of Lightwave Technology* 38.21 (2020), pp. 5991–5999.

[25] Stavros Deligiannidis, Charis Mesaritakis, et al. "Performance and Complexity Analysis of Bi-Directional Recurrent Neural Network Models Versus Volterra Nonlinear Equalizers in Digital Coherent Systems". In: *Journal of Lightwave Technology* 39.18 (2021), pp. 5791–5798.

[26] E. Desurvire, J. R. Simpson, et al. "High-gain erbium-doped traveling-wave fiber amplifier". In: *Opt. Lett.* 12.11 (Nov. 1987), pp. 888–890.

[27] Liang B. Du and Arthur J. Lowery. "Improved Single Channel Backpropagation for Intra-Channel Fiber Nonlinearity Compensation in Long-Haul Optical Communication Systems". In: *Opt. Express* 18.16 (Aug. 2010), pp. 17075–17088. URL: https://opg.optica.org/oe/abstract.cfm?URI%20=%20oe-18-16-17075.

[28] Jeffrey L. Elman. "Finding Structure in Time". In: *Cognitive Science* 14.2 (1990), pp. 179–211. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog1402_1. URL: https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog1402_1.

[29] René-Jean Essiambre, Gerhard Kramer, et al. "Capacity Limits of Optical Fiber Networks". In: *IEEE J. Lightw. Technol.* 28.4 (Feb. 2010), pp. 662–701.

[30] Qirui Fan, Chao Lu, et al. “Combined Neural Network and Adaptive DSP Training for Long-Haul Optical Communications”. In: *IEEE J. Lightw. Technol.* 39.22 (Nov. 2021), pp. 7083–7091.

[31] Qirui Fan, Chao Lu, et al. “Combined Neural Network and Adaptive DSP Training for Long-Haul Optical Communications”. In: *IEEE J. Lightw. Technol.* 39.22 (Nov. 2021), pp. 7083–7091.

[32] Qirui Fan, Gai Zhou, et al. “Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning”. In: *Nature Communications* 11.1 (July 2020), p. 3694. URL: https://doi.org/10.1038/s41467-020-17516-7.

[33] Irshaad Fatadin, David Ives, et al. “Laser Linewidth Tolerance for 16-QAM Coherent Optical Systems Using QPSK Partitioning”. In: *IEEE Photon. Technol. Lett.* 22.9 (May 2010), pp. 631–633.

[34] Johannes Karl Fischer, Christian-Alexander Bunge, et al. “Equivalent Single-Span Model for Dispersion-Managed Fiber-Optic Transmission Systems”. In: *Journal of Lightwave Technology* 27.16 (2009), pp. 3425–3432.

[35] Christoffer Fougstedt, Mikael Mazur, et al. “Time-domain Digital Back Propagation: Algorithm and Finite-precision Implementation Aspects”. In: *Proc. Opt. Fiber Conf.* Los Angeles, CA, USA, 2017, pp. 1–3.

[36] P. J. Freire, V. Neskorniuk, et al. “Complex-Valued Neural Network Design for Mitigation of Signal Distortions in Optical Links”. In: *IEEE J. Lightw. Technol.* 39.6 (Mar. 2021), pp. 1696–1705.

[37] Pedro J. Freire, Antonio Napoli, et al. “Neural Networks-Based Equalizers for Coherent Optical Transmission: Caveats and Pitfalls”. In: *IEEE Journal of Selected Topics in Quantum Electronics* 28.4: Mach. Learn. in Photon. Commun. and Meas. Syst. (2022), pp. 1–23.

[38] Pedro J. Freire, Vladislav Neskornuik, et al. “Experimental Verification of Complex-Valued Artificial Neural Network for Nonlinear Equalization in Coherent Optical Communication Systems”. In: *2020 European Conference on Optical Communications (ECOC)*. 2020, pp. 1–4.

[39] Pedro J. Freire, Yevhenii Osadchuk, et al. “Performance Versus Complexity Study of Neural Network Equalizers in Coherent Optical Systems”. In: *IEEE J. Lightw. Technol.* 39.19 (Oct. 2021), pp. 6085–6096.

[40] Simone Gaiarin, Francesco Da Ros, et al. “Experimental Demonstration of Nonlinear Frequency Division Multiplexing Transmission With Neural Network Receiver”. In: *Journal of Lightwave Technology* 38.23 (2020), pp. 6465–6473.

[41] Yan Gao, Fan Zhang, et al. “Intra-channel Nonlinearities Mitigation in Pseudo-linear Coherent QPSK Transmission Systems via Nonlinear Electrical Equalizer”. In: *Opt. Commun.* 282.12 (2009), pp. 2421–2425.

[42] Ying Gao, Jian Hong Ke, et al. “Assessment of Intrachannel Nonlinear Compensation for 112 Gb/s Dual-Polarization 16QAM Systems”. In: *Journal of Lightwave Technology* 30.24 (2012), pp. 3902–3910.

[43] Mario Garrido, Petter Källström, et al. “CORDIC II: A New Improved CORDIC Algorithm”. In: *IEEE Trans. Circuits Syst. II: Express Briefs* 63.2 (Feb. 2016), pp. 186–190.

[44] S. J. Garth and C. Pask. “Four-photon mixing and dispersion in single-mode fibers”. In: *Opt. Lett.* 11.6 (June 1986), pp. 380–382.

[45] Elias Giacoumidis, Son T. Le, et al. “Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based nonlinear equalization”. In: *Opt. Lett.* 40.21 (Nov. 2015), pp. 5113–5116. URL: https://opg.optica.org/ol/abstract.cfm?URI%20=%20ol-40-21-5113.

[46] Kseniia Goroshko, Hadrien Louchet, et al. “Overcoming Performance Limitations of Digital Back Propagation Due to Polarization Mode Dispersion”. In: *Proc. 18th Int. Conf. Transp. Opt. Netw.* Trento, Italy, 2016, pp. 1–4.

[47] C. Häger and H. D. Pfister. “Deep Learning of the Nonlinear Schrödinger Equation in Fiber-Optic Communications”. In: *2018 IEEE International Symposium on Information Theory (ISIT)*. 2018, pp. 1590–1594.

[48] C. Häger and H. D. Pfister. “Physics-Based Deep Learning for Fiber-Optic Communication Systems”. In: *IEEE J. Sel. Areas Commun.* 39.1 (Jan. 2021), pp. 280–294.

[49] Christian Häger and Henry D. Pfister. “Nonlinear Interference Mitigation via Deep Neural Networks”. In: *Proc. Opt. Fiber Conf.* San Diego, CA, USA, 2018, pp. 1–3.

[50] Christian Häger, Henry D. Pfister, et al. “Model-Based Machine Learning for Joint Digital Backpropagation and PMD Compensation”. In: *Optical Fiber Communication Conference (OFC) 2020*. Optical Society of America, 2020, W3D.3. URL: http://www.osapublishing.org/abstract.cfm?URI%20=%20OFC-2020-W3D.3.

[51] Douglas M. Hawkins. “The Problem of Overfitting”. In: *Journal of Chemical Information and Computer Sciences* 44.1 (2004). PMID: 14741005, pp. 1–12. eprint: https://doi.org/10.1021/ci0342472. URL: https://doi.org/10.1021/ci0342472.

[52] S. Haykin. *Digital Communication Systems*. Wiley, 2013. URL: https://books.google.fr/books?id%20=%20YGZXAAAACAAJ.

[53] M. Hirano, T. Haruna, et al. “Record low loss, record high FOM optical fiber with manufacturable process”. In: *2013 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC)*. 2013, pp. 1–3.

[54] Sebastian Hoffmann, Ralf Peveling, et al. “Multiplier-Free Real-Time Phase Tracking for Coherent QPSK Receivers”. In: *IEEE Photonics Technology Letters* 21.3 (2009), pp. 137–139.

[55] Takashi Inoue, Ryosuke Matsumoto, et al. “Learning-based Digital Back Propagation to Compensate for Fiber Nonlinearity Considering Self-phase and Cross-phase Modulation for Wavelength-Division Multiplexed Systems”. In: *Opt. Express* 30.9 (Apr. 2022), pp. 14851–14872.

[56] E. Ip and J. M. Kahn. “Compensation of Dispersion and Nonlinear Impairments Using Digital Backpropagation”. In: *Journal of Lightwave Technology* 26.20 (2008), pp. 3416–3425.

[57] Ezra Ip and Joseph M. Kahn. “Compensation of Dispersion and Nonlinear Impairments Using Digital Backpropagation”. In: *IEEE J. Lightw. Technol.* 26.20 (Oct. 2008), pp. 3416–3425.

[58] Mutsam A. Jarajreh, Elias Giacoumidis, et al. “Artificial Neural Network Nonlinear Equalizer for Coherent Optical OFDM”. In: *IEEE Photonics Technology Letters* 27.4 (2015), pp. 387–390.

[59] Rasmus T. Jones, Simone Gaiarin, et al. “Time-Domain Neural Network Receiver for Nonlinear Frequency Division Multiplexed Systems”. In: *IEEE Photonics Technology Letters* 30.12 (2018), pp. 1079–1082.

[60] Ivan Kaminow, Tingye Li, et al. *Optical Fiber Communication B: Systems and Networks (Sixth ed.)* London: Academic Press, 2013.

[61] Boris Karanov, Mathieu Chagnon, et al. “End-to-End Deep Learning of Optical Fiber Communications”. In: *IEEE J. Lightw. Technol.* 36.20 (Oct. 2018), pp. 4843–4855.

[62] Boris Karanov, Vinícius Oliari, et al. “End-to-End Learning in Optical Fiber Communications: Experimental Demonstration and Future Trends”. In: *Proc. Eur. Conf. Opt. Commun.* Brussels, Belgium, 2020, pp. 1–4.

[63] Chao Li, Yongjun Wang, et al. “Ultra-low complexity random forest for optical fiber communications”. In: *Opt. Express* 31.7 (Mar. 2023), pp. 11633–11648. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-31-7-11633.

[64] Lei Li, Zhenning Tao, et al. “Implementation efficient nonlinear equalizer based on correlated digital backpropagation”. In: *2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference*. 2011, pp. 1–3.

[65] Mengke Lian, Christian Häger, et al. “What Can Machine Learning Teach Us about Communications?” In: *2018 IEEE Information Theory Workshop (ITW)*. 2018, pp. 1–5.

[66] Xiaojun Liang and Shiva Kumar. “Correlated Digital Back Propagation Based on Perturbation Theory”. In: *Opt. Express* 23.11 (June 2015), pp. 14655–14665.

[67] Chien-Yu Lin, Michael Holtmannspoetter, et al. “Compensation of Transmission Impairments by Digital Backward Propagation for Different Link Designs”. In: *Proc. Eur. Conf. Opt. Commun.* Turin, Italy, 2010, pp. 1–3.

[68] Richard A Linke and Alan H Gnauck. “High-capacity coherent lightwave systems”. In: *Journal of Lightwave Technology* 6.11 (1988), pp. 1750–1769.

[69] Xiang Liu, S. Chandrasekhar, et al. “Efficient Fiber Nonlinearity Mitigation in 50-GHz-DWDM Transmission of 256-Gb/s PDM-16QAM Signals by Folded Digital-back-propagation and Channelized FBG-DCMs”. In: *Proc. Opt. Fiber Conf.* San Francisco, CA, USA, 2014, pp. 1–3.

[70] Guo-Wei Lu, Mats Sköld, et al. “40-Gbaud 16-QAM transmitter using tandem IQ modulators with binary driving electronic signals”. In: *Opt. Express* 18.22 (Oct. 2010), pp. 23062–23069. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-18-22-23062.

[71] I. H. Malitson. “Interspecimen Comparison of the Refractive Index of Fused Silica∗,†”. In: *J. Opt. Soc. Am.* 55.10 (Oct. 1965), pp. 1205–1209. URL: https://opg.optica.org/abstract.cfm?URI%20=%20josa-55-10-1205.

[72] Silvio Mandelli. “Analysis of Wiener phase noise issues in optical transmission systems”. PhD thesis. Milano, Italy: Politecnico di Milano, Jan. 2016. URL: http://hdl.handle.net/10589/114642.

[73] D. Marcuse, C.R. Manyuk, et al. “Application of the Manakov-PMD Equation to Studies of Signal Propagation in Optical Fibers with Randomly Varying Birefringence”. In: *Journal of Lightwave Technology* 15.9 (1997), pp. 1735–1746.

[74] Eduardo F. Mateo, Fatih Yaman, et al. “Efficient Compensation of Inter-channel Nonlinear Effects via Digital Backward Propagation in WDM Optical Transmission”. In: *Opt. Express* 18.14 (July 2010), pp. 15144–15154.

[75] Robert J Mears, L Reekie, et al. “Low-noise erbium-doped fibre amplifier operating at 1.54 μm”. In: *Electronics Letters* 19.23 (1987), pp. 1026–1028.

[76] Pramod K. Meher, Javier Valls, et al. “50 Years of CORDIC: Algorithms, Architectures, and Applications”. In: *IEEE Trans. Circuits Syst. I: Regul. Pap.* 56.9 (Sept. 2009), pp. 1893–1907.

[77] Marina Melek. PhD thesis. UWSpace, University of Waterloo Library, Dec. 2021. URL: Mitigating%20Fiber%20Nonlinearity%20with%20Machine%20Learning.

[78] David S. Millar, Sergejs Makovejs, et al. “Mitigation of Fiber Nonlinearity Using a Digital Coherent Receiver”. In: *IEEE J. Sel. Top. Quantum Electron.* 16.5 (Sept. 2010), pp. 1217–1226.

[79] Hao Ming, Xinyu Chen, et al. “Ultralow Complexity Long Short-Term Memory Network for Fiber Nonlinearity Mitigation in Coherent Optical Communication Systems”. In: *Journal of Lightwave Technology* 40.8 (2022), pp. 2427–2434.

[80] L.F. Mollenauer, S.G. Evangelides, et al. “Long-distance soliton propagation using lumped amplifiers and dispersion shifted fiber”. In: *Journal of Lightwave Technology* 9.2 (1991), pp. 194–197.

[81] Antonio Napoli, Zied Maalej, et al. “Reduced Complexity Digital Back-propagation Methods for Optical Communication Systems”. In: *IEEE J. Lightw. Technol.* 32.7 (Apr. 2014), pp. 1351–1362.

[82] S. Ohnuki. “Further attenuation improvement of a pure silica core fiber with large effective area”. In: *Proc. SubOptic 2010*. 2010.

[83] Payal and Suresh Kumar. “Nonlinear Impairments in Fiber Optic Communication Systems: Analytical Review”. In: *Futuristic Trends in Network and Communication Technologies*. Ed. by Pradeep Kumar Singh, Marcin Paprzycki, et al. Singapore: Springer Singapore, 2019, pp. 28–44.

[84] Martin Pelikan, David E. Goldberg, et al. “BOA: The Bayesian Optimization Algorithm”. In: *Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation - Volume 1*. GECCO’99. Orlando, Florida: Morgan Kaufmann Publishers Inc., 1999, pp. 525–532.

[85] Timo Pfau, Sebastian Hoffmann, et al. “Hardware-Efficient Coherent Digital Receiver Concept With Feedforward Carrier Recovery for M -QAM Constellations”. In: *IEEE J. Lightw. Technol.* 27.8 (June 2009), pp. 989–999.

[86] Proakis. *Digital Communications 5th Edition*. McGraw Hill, 2007.

[87] Danish Rafique, Marco Mussolin, et al. “Compensation of Intra-Channel Nonlinear Fibre Impairments Using Simplified Digital Back-Propagation Algorithm”. In: *Opt. Express* 19.10 (May 2011), pp. 9453–9460. URL: https://opg.optica.org/oe/abstract.cfm?URI%20=%20oe-19-10-9453.

[88] Seb J. Savory. “Digital filters for coherent optical receivers”. In: *Opt. Express* 16.2 (Jan. 2008), pp. 804–817. URL: http://www.opticsexpress.org/abstract.cfm?URI%20=%20oe-16-2-804.

[89] Seb J. Savory. “Digital Coherent Optical Receivers: Algorithms and Subsystems”. In: *IEEE J. Sel. Top. Quantum Electron.* 16.5 (Sept. 2010), pp. 1164–1179.

[90] Bernhard Schmauss, Chien-Yu Lin, et al. “Progress in Digital Backward Propagation”. In: *Proc. Eur. Conf. Opt. Commun.* Amsterdam, Netherlands, 2012, pp. 1–3.

[91] Marco Secondini, Domenico Marsella, et al. “Enhanced Split-Step Fourier Method for Digital Backpropagation”. In: *Proc. Eur. Conf. Opt. Commun.* Cannes, France, 2014, pp. 1–3.

[92] Marco Secondini, Simon Rommel, et al. “Single-Step Digital Backpropagation for Nonlinearity Mitigation”. In: *Photonic Network Communications* 31.3 (June 2016), pp. 493–502.

[93] Jing Shao, Shiva Kumar, et al. “Digital Back Propagation With Optimal Step Size for Polarization Multiplexed Transmission”. In: *IEEE Photon. Technol. Lett.* 25.23 (Dec. 2013), pp. 2327–2330.

[94] Jing Shao, Shiva Kumar, et al. “Digital Back Propagation With Optimal Step Size for Polarization Multiplexed Transmission”. In: *IEEE Photonics Technology Letters* 25.23 (2013), pp. 2327–2330.

[95] Thomas Shun Rong Shen and Alan Pak Tao Lau. “Fiber nonlinearity compensation using extreme learning machine for DSP-based coherent communication systems”. In: *16th Opto-Electronics and Communications Conference*. 2011, pp. 816–817.

[96] Oleg Sidelnikov, Alexey Redyuk, et al. “Equalization Performance and Complexity Analysis of Dynamic Deep Neural Networks in Long Haul Transmission Systems”. In: *Opt. Express* 26.25 (Dec. 2018), pp. 32765–32776.

[97] Oleg Sidelnikov, Alexey Redyuk, et al. “Advanced Convolutional Neural Networks for Nonlinearity Mitigation in Long-Haul WDM Transmission Systems”. In: *IEEE J. Lightw. Technol.* 39.8 (Apr. 2021), pp. 2397–2406.

[98] Bernhard Spinnler. “Equalizer Design and Complexity for Digital Coherent Receivers”. In: *IEEE J. Sel. Top. Quantum Electron.* 16.5 (Sept. 2010), pp. 1180–1192.

[99] Yoshiaki Tamura, Hirotaka Sakuma, et al. “The First 0.14-dB/km Loss Optical Fiber and its Impact on Submarine Transmission”. In: *IEEE J. Lightw. Technol.* 36.1 (2018), pp. 44–49.

[100] Du Tang, Zhen Wu, et al. “Joint intra and inter-channel nonlinearity compensation based on interpretable neural network for long-haul coherent systems”. In: *Opt. Express* 29.22 (Oct. 2021), pp. 36242–36256. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-29-22-36242.

[101] Zhenning Tao, Liang Dou, et al. “Multiplier-Free Intrachannel Nonlinearity Compensating Algorithm Operating at Symbol Rate”. In: *Journal of Lightwave Technology* 29.17 (2011), pp. 2570–2576.

[102] E. Temprana, E. Myslivets, et al. “Two-Fold Transmission Reach Enhancement Enabled by Transmitter-Side Digital Backpropagation and Optical Frequency Comb-Derived Information Carriers”. In: *Opt. Express* 23.16 (Aug. 2015), pp. 20774–20783. URL: https://opg.optica.org/oe/abstract.cfm?URI%20=%20oe-23-16-20774.

[103] Andrew J. Viterbi and Audrey M. Viterbi. “Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission”. In: *IEEE Transactions on Information Theory* 29.4 (1983), pp. 543–551.

[104] H. Yamaguchi. “Ultra-low loss and large Aeff pure-silica core fiber advances”. In: *Proc. SubOptic 2016*. 2010.

[105] Guanqun Zhan, Anlin Yi, et al. “Fiber Nonlinearity Mitigation for mQAM Coherent Optical Communication System Utilizing Random Forest Algorithm”. In: *2022 Asia Communications and Photonics Conference (ACP)*. 2022, pp. 677–681.

[106] Fan Zhang, Xiansong Fang, et al. “Neural Network-based Fiber Nonlinearity Mitigation in High-speed Coherent Optical Transmission Systems”. In: *2022 Optical Fiber Communications Conference and Exhibition (OFC)*. 2022, pp. 1–3.

[107] Huale Zhang, Rui Cheng, et al. “Design of Intelligent Detection Algorithm for Optical Fiber Communication Fault Data Based on Improved Random Forest”. In: *2023 International Conference on Networking, Informatics and Computing (ICNETIC)*. 2023, pp. 282–288.

[108] S. Zhang, F. Yaman, et al. “Neuron-Network-Based Nonlinearity Compensation Algorithm”. In: *Proc. Eur. Conf. Opt. Commun.* Rome, Italy, 2018, pp. 1–3.

[109] Shaoliang Zhang, Fatih Yaman, et al. “Field and lab experimental demonstration of nonlinear impairment compensation using neural networks”. In: *Nature Communications* 10.1 (July 2019), p. 3033. URL: https://doi.org/10.1038/s41467-019-10911-9.

[110] Likai Zhu and Guifang Li. “Folded Digital Backward Propagation for Dispersion-Managed Fiber-Optic Transmission”. In: *Opt. Express* 19.7 (Mar. 2011), pp. 5953–5959. URL: https://opg.optica.org/oe/abstract.cfm?URI%20=%20oe-19-7-5953.

[111] Likai Zhu and Guifang Li. "Nonlinearity Compensation Using Dispersion-Folded Digital Backward Propagation". In: *Opt. Express* 20.13 (June 2012), pp. 14362–14370. URL: https://opg.optica.org/oe/abstract.cfm?URI%20=%20oe-20-13-14362.

ECOLE DOCTORALE

Titre : Égalisation dans la Communication par Fibre Optique en Utilisant des Réseaux Neuronaux Basés sur des Modèles

Mots clés : Fibre optique, Apprentissage automatique, Traitement du signal

Résumé : Répondre à la demande croissante de transmission de données à haute vitesse nécessite des solutions efficaces pour atténuer la non-linéarité dans les systèmes de communication optique. Les méthodes traditionnelles telles que la rétropropagation numérique (DBP) sont confrontées à d'importants défis computationnels, les rendant impraticables pour les systèmes du monde réel. Les modèles de réseaux neuronaux ont émergé comme une approche prometteuse pour résoudre ce problème. Deux approches principales existent pour la conception de réseaux neuronaux : les méthodes agnostiques au modèle et les méthodes basées sur le modèle. Les techniques agnostiques au modèle offrent de la flexibilité en termes de taille et d'hyperparamètres et peuvent être placées à différentes positions dans la chaîne de traitement du signal numérique (DSP) du récepteur. Cependant, elles exigent une taille substantielle et des données d'entraînement étendues pour fonctionner efficacement. En revanche, les approches basées sur le modèle emploient des réseaux neuronaux guidés par le modèle physique de la propagation du signal. Ces approches ont tendance à être plus compactes mais nécessitent une initialisation soigneuse pour une généralisation correcte. Une technique basée sur le modèle de premier plan est la rétropropagation numérique apprise (LDBP), qui optimise les paramètres DBP à l'aide de réseaux neuronaux. LDBP promet des performances améliorées ou une complexité réduite par rapport à DBP. Cette étude se concentre principalement sur LDBP, introduisant des simplifications par le biais du "partage de paramètres" pour réduire les paramètres entraînables. De plus, nous proposons de réaffecter les systèmes hérités à gestion de la dispersion (DM) en incorporant des formats de modulation d'ordre supérieur tels que 16-QAM et 64-QAM pour améliorer les débits de données au sein de ces systèmes. Une analyse approfondie des performances et de la complexité démontre que les algorithmes proposés surpassent l'égalisation linéaire et DBP dans divers systèmes de transmission.

Title : Equalization in Optical Fiber Communication Using Model-based Neural Networks

Keywords : Optical fiber, Machine learning, Signal processing

Abstract : Meeting the increasing demand for high-speed data transmission requires effective solutions for mitigating nonlinearity in optical communication systems. Traditional methods like Digital Backpropagation (DBP) face significant computational challenges, rendering them impractical for real-world systems. Neural network models have emerged as a promising approach to address this issue. Two primary approaches exist for designing neural networks: model-agnostic and model-based methods. Model-agnostic techniques offer flexibility in terms of size and hyperparameters and can be placed at various positions in the receiver's Digital Signal Processing (DSP) chain. However, they demand substantial size and extensive training data in order to operate effectively. In contrast, model-based approaches employ neural networks guided by the physical model of signal propagation. These approaches tend to be more compact but require careful initialization for proper generalization. One prominent model-based technique is Learned Digital Backpropagation (LDBP), which optimizes DBP parameters using neural networks. LDBP offers the promise of improved performance or reduced complexity compared to DBP. This study primarily focuses on LDBP, introducing simplifications through "parameter sharing" to reduce trainable parameters. Additionally, we propose repurposing the legacy Dispersion-Managed (DM) systems, by incorporating higher-order modulation formats such as 16-QAM and 64-QAM, to enhance data rates within these systems. A comprehensive analysis of the performance and complexity demonstrates that the proposed algorithms outperform linear equalization and DBP in various transmission systems.

Institut Polytechnique de Paris
91120 Palaiseau, France