Mitigating Fiber Nonlinearity with Machine Learning

by

Marina Melek

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Joel Zylberberg
                        Assistant Professor, Dept. of Physics and Astronomy, York University

Supervisor(s):   David Yevick
                        Professor, Dept. of Physics and Astronomy, University of Waterloo

Internal Member: Kostadinka Bizheva
                        Professor, Dept. of Physics and Astronomy, University of Waterloo

Internal Member: Jeff Z.Y. Chen
                        Professor, Dept. of Physics and Astronomy, University of Waterloo

Internal-External Member: Oleg Michailovich
                        Associate Professor, Dept. of Electrical and Computer Engineering, University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Nowadays, optical communication transmission is based mainly on optical fiber networks. Increasing demands for higher-capacity systems are hampered by signal distortions due to nonlinear effects of the commercial optic fibers. Different techniques have been proposed to reverse and mitigate this noise effect on the transmitted signal such as the digital backpropagation (DBP), the Volterra nonlinear compensation, the advanced modulation transmission, and perturbation pre-compensation techniques. While these techniques achieve good results they are too complicated for practical industrial implementation and add more complexity overhead on the system.

This thesis is focused on investigating the merits of optical fiber mitigation using Artificial Intelligence (AI) techniques instead of analytical methods. Different AI techniques combined with perturbation-based nonlinear compensation method are used to predict the added nonlinear noise to a 16-Quadrature Amplitude Modulation (QAM) propagating signal. A MATLAB simulation program has been used to model the propagation of the signal and generate the transmitted data.

The AI simulations have been employed using Python on dual-polarization single-channel systems using single-stage AI techniques such as Neural Network (NN) at receiver or transmitter side and Siamese neural network (SNN), or two-stage AI techniques. In the two-stage method, different supervised classifiers have been used at the receiver side such as multi-layer perceptrons (MLP), decision tree, AdaBoosting, GBoosting, random forest, and extra trees while NN is placed at the transmitter. Additionally, different complexity reduction techniques have been applied to the proposed systems to achieve more practical performance in industrial environment applications.

For the first time, a nonlinear-compensation robustness study is applied to the proposed AI techniques by detecting the performance of each technique while changing the single-mode fiber's nonlinear coefficient value. Moreover, empirical equations are developed to represent the system's Q-factor enhancement achieved using each of the proposed techniques as a function of the fiber nonlinear coefficient and the data features.

Acknowledgements

Dedication

This thesis is dedicated to my family and friends.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI Artificial Intelligence iv, 4, 6, 7, 19, 38, 39, 41, 82, 87

AIR achievable information rate 31, 32

ANN artificial neural networks 39, 40

ASE-noise amplified spontaneous emission noise 4, 8, 14, 28, 31, 82

AWGN additive White Gaussian noise 33, 40, 41

CD chromatic dispersion 4, 35, 39, 82

CDC chromatic dispersion compensation 30, 35, 45, 53, 63

DBP digital backpropagation iv, 27–30, 43, 44, 66, 82, 83, 88

DGD Differential Group Delay 12

DSP digital signal processing 3, 6, 27, 28, 32, 40, 41, 87

EDFA Erbium-Doped-Fiber-Amplifiers 3, 4, 14, 38, 39

FFNN Feed forward Neural Network 6, 7, 23, 42, 43, 52, 59, 87, 88

FFT fast Fourier transform 28, 30, 69

GS Geometric shaping 32

GVD group velocity dispersion 11

MB Maxwell-Boltzmann 33

List of Symbols

Chapter 1

Introduction

1.1   Fiber Optic Communications

In optical communication, light replaces the electric current in standard communication systems. The channel, that is used in optical communication to carry signals to their destinations, consists of optical fibers because of their low attenuation, leading to higher bandwidth than the electrical transmission as shown in Figure 1.1. Therefore, optical networks are considered the best economic solution for transmitting large amounts of information, especially in computer networks [55].

Since 1950, researchers have been working toward finding solutions for improving the telecommunication systems' capacity limit using optics. However, their challenge was that neither an optically coherent source nor a suitable propagation medium was available. In 1960, the invention of laser solved the first problem, and then the focus was directed toward finding methods for introducing laser sources in optical communication [3].

During that time, most engineers ignored optical fibers for telecommunication applications because of their high losses, since 90% of the light that enters the fiber is lost during propagation after a few meters. However, in telecommunication, the signal has to be transported over at least a few kilometers. In 1966, it was suggested that removing the impurities from the silica glass, used in manufacturing optical fibers, could drastically minimize these losses. These low-losses fibers could be the best medium for optical communication [54]. The American company Corning, which specializes in materials science, adopted this challenge. In 1970, a breakthrough happened when three scientists from Corning indicated that they could reduce fiber losses near 630 nm wavelength to below
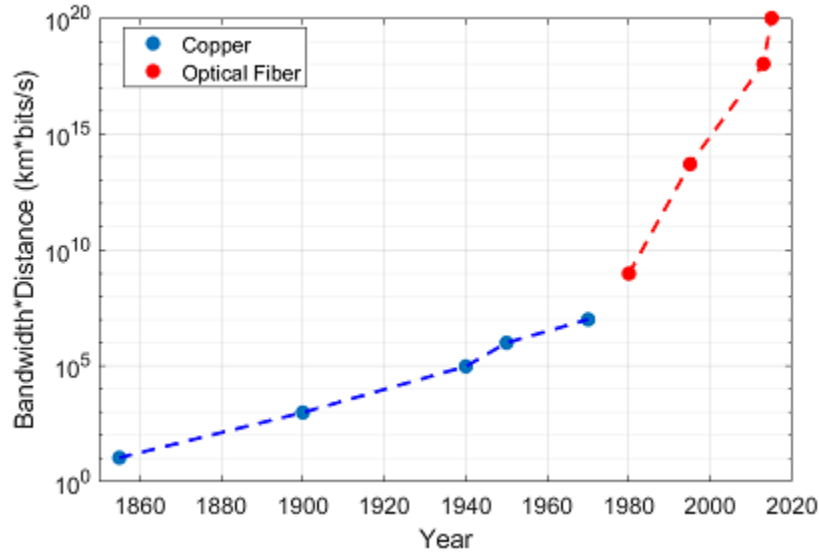
Figure 1.1: Growth in bandwidth-distance product across all telecommunications during 1840-2015 from Agrawal, 2016 [3]

20 dB/km [56]. Two years later, this group could produce a fiber with only 4 dB/km loss. They achieved that by replacing the dopant inside the core of the fiber from titanium to germanium. Several industrial laboratories started competing after that for achieving the minimum fiber losses. In 1979, another group from Japan could produce an optical fiber with 0.2 dB/km loss in the infrared wavelength region near 1.55 μm [75]. This loss value was the closest to the Rayleigh scattering phenomenon limit and is similar to that exhibited currently in modern fibers. Rayleigh scattering is the linear scattering of light at particles with dimensions smaller than the wavelength of the light.

1.1.1 Generations of Optical Fibers

First-generation optical fibers were designed to propagate data at a speed of 45 Mbit/s with GaAs semiconductor lasers that emit light with a wavelength near 850 nm. The problem with operating around that wavelength was having 3 dB/km loss, therefore repeaters were needed to regenerate the optical signal every 10 km. In 1980, commercial systems were installed for this generation after extensive laboratory experiments and several successful field trials.

Fiber optics' second generation communication systems became available in the early

1980s, but their speed was initially limited to 100 Mbit/s because of multimode fibers dispersion. To overcome this problem, single-mode fibers were used. By 1987, the operating bit rates of the second-generation commercialized systems were up to 1.7 Gbit/s with about 50 km repeater spacing. However, at their operating wavelength of 1.3 μm, fiber loss was the main limitation in the repeater spacing of the second-generation systems.

In 1990, optical fiber third-generation communication systems operating at 2.5 Gbit/s became commercially available, and their capacity was soon expanded to 10 Gbit/s. The combination of using dispersion-shifted fibers and a single longitudinal laser propagation mode could achieve the best transmission performance. Moreover, system designers were strongly urging the use of three novel ideas: implementing periodic dispersion compensation for controlling fiber dispersion, using periodic optical amplifiers for mitigating fiber losses, and adopting Wavelength division multiplexing (WDM) for improving the system capacity. The WDM technique proved to be promising in increasing the system capacity because it uses multiple lasers with slightly different wavelengths that allowed multiple data streams to be simultaneously propagated through the same optical fiber. The spacing between the frequency of two neighboring channels had to be carefully chosen to be as small as possible, while it should not be smaller than each channel bandwidth.

The fourth generation of optical fiber communication systems was combined with WDM systems to increase the transmission capacity. This generation of optical fibers used optical amplifiers to increase the distance between optical repeaters. In most WDM systems, fiber losses are periodically compensated every $60-80$ km using Erbium-Doped-Fiber-Amplifiers (EDFA).

For commercial WDM systems, the International Telecommunication Union (ITU) reserved a group of fixed frequencies using 50 GHz channel spacing. This group of frequencies lies in the wavelength region near 1550 nm where minimum fiber losses are located. The conventional band (C band) contains a wavelength range of $1530 - 1570$ nm, in which most of the commercial WDM systems operate. The S and L bands, which stand for the short- and long-wavelength sides of the C band respectively, are also used in limited conditions[35].

Although the research on coherent optical receivers was mostly concentrated in the late 1980s and early 1990s, no significant progress has been made until 2007. After the Dot-Com bubble evolution in 2000, studying coherent systems grew again due to the commercial need for the reduction of the cost per bit, which can be achieved using lower bandwidth components. Consequently, the digital coherent receiver emerged, which led to doubling the capacity and spectral efficiency of optical transmission by facilitating the detection of dual-polarization (in-phase and quadrature) components of the signal. Moreover, it facilitated

3

the use of digital signal processing (DSP) algorithms, which allows the compensation of the linear impairments of the optical channel while the signal is in the digital domain. In this case, a significant cost reduction could be achieved by digitally compensating unlimited amounts of chromatic dispersion (CD) and polarization mode dispersion (PMD) as the dispersion-compensating modules, and optical filter compensators could be eliminated, and the number of the EDFAs per link could be halved.

Probably the most important advantage of coherent detection is the ability to detect higher-order modulation formats such as polarization-division multiplexed quadrature phase-shift keying (PDM-QPSK) [83], polarization-division multiplexed eight phase shift keying (PDM-8PSK) [112], polarization-division multiplexed 8 quadrature amplitude modulation (PDM-8QAM) [113], PDM-16QAM [41], and PDM-36QAM [114]. PDM-QPSK is a format where every two bits of certain polarization of an optical signal can trigger in-phase 'I' and quadrature-phase 'Q' modulators, 90° separated from the in-phase, leading to selecting one of four possible carrier phase shifts corresponding to this combination of bits which is called a symbol. The distribution of the data symbols on the I-Q graph is called constellation, as shown in figure 1.2. In PDM-8PSK modulation, each symbol consists of three bits that are used to select one of the 8-phase shift states ($\pi/4$ separated) for the signal modulation. However, QAM modulation utilizes both amplitude and phase components to be able to provide high levels of spectrum usage efficiency. Using QAM, many different points can be used, where each is having defined values of phase and amplitude. For example, in 8-QAM modulation, symbols of three data bits are represented in eight modulation states using four-phase angles on 90° boundaries and two amplitudes. However, the denser the constellation diagrams which represent higher-order modulation formats are, the more EDFAs circularly symmetric Gaussian noise is generated along the propagation link. Even though this could be improved by increasing the launch power per wavelength channel to enhance the signal-to-noise ratio at the receiver, the transmission would still be limited by nonlinear distortions due to the Kerr effect. This means that the fiber refractive index and eventually the phase delay of the signal propagating through it is proportional to the square of the light intensity, this severely impacts the higher-order modulation formats. So, with the continuous demand for higher speed communication as shown in figure 1.3, the most pressing question of current optical communications research is: how can the capacity of the current optical transport network be maximized by effectively detecting the nonlinear noise and trying to reverse it? Therefore, the research presented in this thesis is mainly concerned with investigating the use of AI in mitigating optical fiber nonlinearity.

Beyond this chapter, the thesis is organized as follows:

Chapter 2 presents the propagation theory of the optical channel and describes the lin-
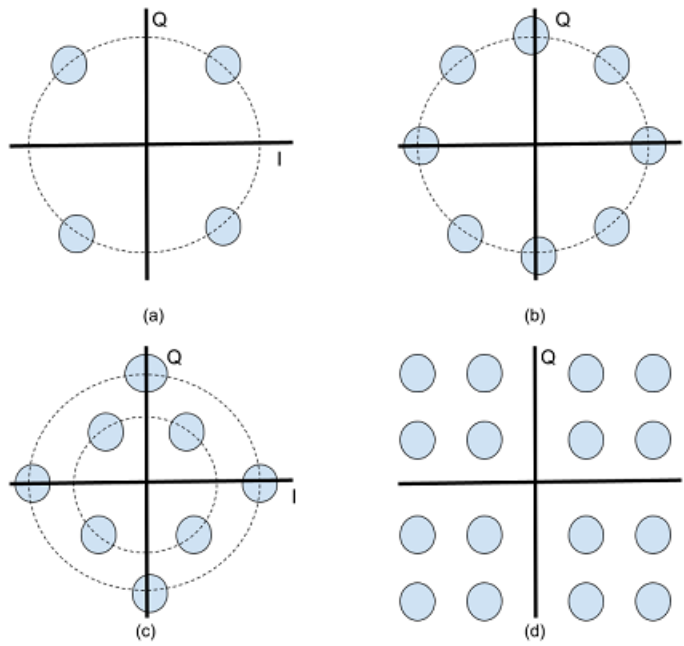
Figure 1.2: Constellation diagram of (a) QPSK, (b) 8PSK, (c) 8QAM, (d) 16QAM
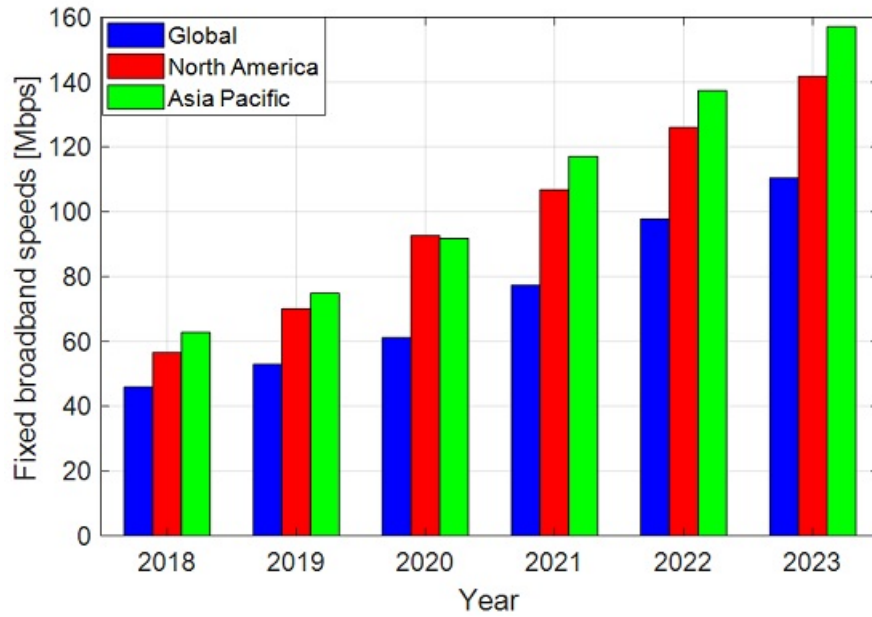
Figure 1.3: Fixed broadband speeds since 2018 and the expected growth till 2023 [22]

ear transmission distortions associated with it such as attenuation, CD, PMD, and amplified spontaneous emission noise (ASE-noise) as well as transmission nonlinear distortions, such as Self-phase modulation (SPM), Intra cross-phase modulation (IXPM), and intra four-wave mixing (IFWM). Moreover, this chapter describes the split-step Fourier method, which is considered the most prominent numerical model of optical fiber transmission. As well as introduces a background about the Machine learning (ML) models, types, and how to implement them.

Chapter 3 describes the different receiver-based DSP algorithms and equalizers used in optical communication such as digital back-propagation, Volterra-based nonlinear compensator, advanced modulation, and perturbation-based pre-compensation technique. Additionally, the chapter presents a literature review for the mitigation of fiber nonlinearity using AI.

Chapter 4 investigates the combination of using neural networks and the perturbation-based technique in mitigating the optical fiber nonlinearity. Moreover, it introduces complexity reduction techniques such as principal component analysis (PCA) and weight trimming techniques.

Chapter 5 proposes the use of neural networks at the transmitter side of the optical communication system and adds another stage of classifiers at the receiver. The investigated classifiers are Decision tree, Boosting, Random forest, Extra trees, and multi-layer perceptron (MLP).

Chapter 6 introduces Siamese Neural Networks (SNN) in combination with the perturbation-based technique to predict the nonlinear noise added to the signal during propagation.

Chapter 7 compares the performance of the AI techniques at the receiver and transmitter sides proposed in the previous chapters under more aggressive nonlinear noise. It also introduces general empirical equations for each technique's performance as a function of the nonlinear coefficient.

Finally, Chapter 8 draws conclusions. It is based on the results obtained in chapters 4, 5, 6, and 7, and it suggests potential research points for future work.

1.2   Research Contributions

The following authentic contributions to the optical communications field have been made in the course of conducting this research:

In Chapter 4, a cascaded complexity reduction technique was proposed using PCA and weight trimming. This technique helps in reducing the number of inputs to the Feed forward Neural Network (FFNN) and the overall complexity of the system. Here, after training the FFNN, the inputs (features) corresponding to weights below a certain threshold are eliminated and do not contribute to the new training cycle. Then PCA is added as another dimensionality reduction technique to eliminate further inputs without affecting the overall performance.

Chapter 5 introduces, for the first time, the two-stage technique in mitigating the fiber nonlinearity. The FFNN was placed at the transmitter side to implement a pre-compensation of the signal using the perturbation-based technique features, and different classifiers were applied at the receiver side for post-compensation.

Chapter 6 demonstrates the benefits of using SNN for the first time in mitigating the fiber nonlinearity. Previously the SNN was used mostly for comparison applications such as face recognition, object tracking, and image similarity detection. In this thesis, we applied the SNN to a totally different area.

Chapter 7 describes a detailed simulation study focused on investigating the effect of changing the fiber nonlinear coefficient on different AI techniques' performance. Since the choice of the proper AI for a system is dependent on the system parameters, this thesis introduces a more in-depth vision on the choice of the best technique according to the nonlinear noise of the system, in addition to providing approximate equations that describe the behavior of each technique at different nonlinear coefficient values.

Chapter 2

Fiber Properties and Machine Learning

2.1 Introduction

This chapter consists of two parts. The first one presents a detailed background on the theory of wave propagation through optical fiber, the factors affecting the propagation, and the well-known numerical method used for solving the equation of propagation. The second part illustrates in detail different types of ML models and explains the proper way to train machine learning models and evaluate their performance.

2.2 Theory of Propagation Through Fiber

This section presents the field propagation of optical signals through optical fibers, which is modeled by the Nonlinear Schrodinger equation (NLSE) and are affected by different impairments. These impairments could be linear with respect to the amplitude of signal fields such as attenuation, chromatic dispersion, PMD, and ASE-noise. Some other impairments have nonlinear behavior with respect to the signal field amplitude instigated by the Kerr effect. In this section, we will discuss these kinds of impairments.

2.2.1 Linear Effects

This section focuses in detail on the sources and the effects of linear impairments on the optical signal field propagating through optical fibers.

Attenuation

Attenuation is one of the linear impairments that affect light traveling through an optical fiber and lead to signal power loss over distance. This loss in power depends on the wavelength of the light and the absorption of the fiber material. In ultra-low-loss fibers which have a wider low-loss window in silica glass fibers, as shown in figure 2.1, the highest attenuation is at short wavelengths while the lowest loss occurs at the 1550 nm wavelength, which is the commonly used wavelength for long-distance transmissions [58]. For an optical fiber with an optical source of power $P(0)$, the optical power after length L is expressed as [58],

$$P(L) = P(0)10^{(-\alpha(\lambda)L/10)}, \tag{2.1}$$

where α is the attenuation coefficient, and λ is the wavelength of the optical signal.

Additional sources of power loss in fibers during propagation are scattering loss, the absorption of impurities in the fiber, and radiative losses, which occur when optical fiber bends with a finite radius of curvature [103]. If the bend radii are large compared to the fiber diameter, macro-bending losses take place as shown in figure 2.2. Macro-bending loss varies with wavelength and bend radius. However, if the bends are random microscopic fluctuations with small curvature radii along the fiber axis, micro-bending losses occur. These fluctuations sources are either manufacturing nonuniformities or nonuniform pressure applied on the fiber during cabling. Consequently, the power is coupled among modes in the fiber, and in the case of single-mode fiber, power is coupled from the guided fundamental mode to higher-order modes from where the power dissipates through normal loss and scattering or refraction in the coating. The loss associated with these small radii bending can be modeled as [82],

$$\alpha = (R/a)^{-2} \frac{V^4}{32\Delta^2}(0.65 + 1.62V^{-1.5} + 2.88V^{-6})^6, \tag{2.2}$$

$$V^2 = (ka)^2(n_1^2 - n_2^2) \tag{2.3}$$

where a is the core radius, R is the radius of curvature of the fiber, k is the propagation constant in free space, and n_1 and n_2 are the core and cladding refractive index, respectively.
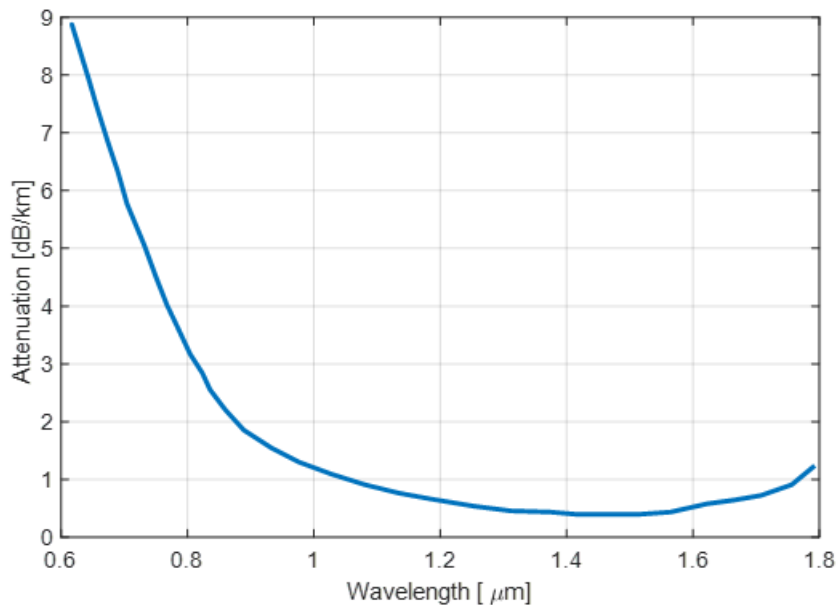
Figure 2.1: Standard mode fiber loss profile as a function of wavelength in μm

The fiber attenuation limits the maximum usable fiber length according to the minimum acceptable power at the receiver. A typical value of α for standard single mode fibres (SMF) at 1550 nm is around 0.046 km^{-1}, which corresponds to 0.2 dB/km, as $\alpha_{dB} = 10 \log_{10}(e)\alpha \approx 4.34\alpha$. Taking only the loss into consideration, the simplest equation for the propagation of an optical field A through an optical fiber can be expressed as,

$$\frac{\partial A}{\partial z} + \frac{\alpha}{2} A = 0. \tag{2.4}$$

Chromatic Dispersion

Chromatic dispersion, another linear impairment that affects light propagation in optical fiber, is a phenomenon in which different spectral components of a pulse travel at different velocities. Therefore, a modification is needed for equation 2.4 with the propagation constant β, which can be expanded as a Taylor series around the carrier frequency and truncated after the 3rd term. This extension leads to the linear part of the Schrödinger equation (LSE) [3],
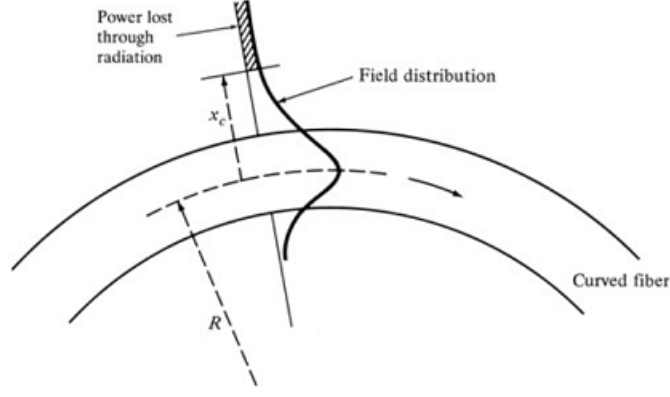
Figure 2.2: Loss in optical fiber caused by macro-bending[58]

$$\frac{\partial A}{\partial z} + \frac{\alpha}{2} A + \beta_1 \frac{\partial A}{\partial t} + \frac{i}{2}\beta_2 \frac{\partial^2 A}{\partial t^2} - \frac{1}{6}\beta_3 \frac{\partial^3 A}{\partial t^3} = 0. \tag{2.5}$$

Hence A is the field of the propagating optical signal, β_1 is related to the group velocity v_g, which identifies the speed of the optical pulse envelope propagating along the fiber as follows,

$$\beta_1 = \frac{1}{v_g} = \frac{1}{c}\left(n + \omega\frac{dn}{d\omega}\right), \tag{2.6}$$

where **c** is the speed of light in a vacuum, n is the linear refractive index, and ω is the optical frequency. The group velocity dispersion (GVD) β_2 leads to the broadening of the pulse while propagating through the fiber, where,

$$\beta_2 = \frac{d\beta_1}{d\omega} = \frac{1}{c}\left(2\frac{dn}{d\omega} + \omega\frac{d^2n}{d\omega^2}\right). \tag{2.7}$$

Consequently, the dispersion parameter D can be determined using β_1 and β_2 as follows,

$$D = \frac{d\beta_1}{d\lambda} = -\frac{2\pi c}{\lambda^2}\beta_2, \tag{2.8}$$

and the GVD slope parameter β_3 is expressed as,

$$\beta_3 = \frac{d\beta_2}{d\omega} = \frac{1}{c}(3\frac{d^2n}{d\omega^2} + \omega\frac{d^3n}{d\omega^3}). \tag{2.9}$$

For simplicity, the influence of the GVD slope can be neglected considering a standard SMF. Thus, using retarded time frame $T = t - z/v_g$ in equation 2.5 leads to the formula,

$$\frac{\partial A}{\partial z} + \frac{\alpha}{2}A + \frac{i}{2}\beta_2\frac{\partial^2 A}{\partial T^2} = 0. \tag{2.10}$$

Chromatic dispersion has two main sources; first, the material dispersion (DM) arises from the frequency-dependence of the material used to make optical fiber. Therefore, different frequency components of the signal's field travel at different speeds. Second, the waveguide dispersion is determined by fiber design characteristics like core radius and core-cladding index difference. In single mode optical fibers, the signal travels partially in the core and partially in the cladding, and the total mode field diameter changes with wavelength. Since the refractive index is different in the core than in the cladding, a change in mode field diameter also results in a change in average dispersion index and, therefore, signal velocity. Waveguide dispersion is the ratio of velocity change to wavelength change caused by this effect [62]. This phenomenon can be utilized to shift the zero-dispersion wavelength to around 1550 nm, where fiber loss is the lowest, as shown in figure 2.3.

Polarization Mode Dispersion

Since SMFs support the transmission of two orthogonal polarization modes, PMD is raised as another source of linear impairments. PMD occurs due to random birefringence generated from the mechanical stress on the fiber. Birefringence is a phenomenon in which the refractive index of the medium material depends on the polarization of the propagating field. Consequently, birefringence can be expressed as,

$$B_m = |\beta_x - \beta_y|\frac{\lambda}{2\pi} = |n_x - n_y|, \tag{2.11}$$

where n_x and n_y are the effective refractive index of the orthogonal modes and β_x and β_y are the corresponding propagation constant. Therefore, two pulses of the same power propagating in orthogonal modes, one along the fast axis and the other along the slow axis, would arrive at different times at the receiver. This difference leads to a pulse spreading and a Differential Group Delay (DGD) $\Delta\tau$, over length L,

$$\Delta\tau = |\frac{L}{v_{gx}} - \frac{L}{v_{gy}}|, \tag{2.12}$$

where v_{gx} and v_{gy} are the group velocity of the signal in x and y direction, respectively. Because the perturbations creating the birefringence are due to intrinsic factors such as
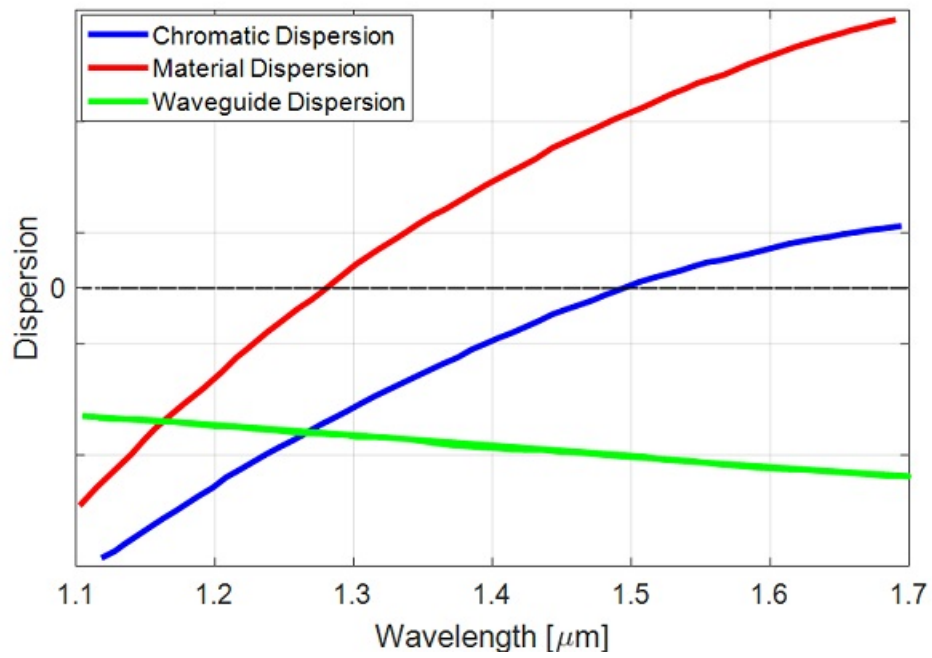
Figure 2.3: Optical fiber chromatic dispersion is the sum of material and waveguide dispersion

fiber core geometric irregularities or internal stress or external factors such as bending, twisting or environmental factors such as temperature, PMD varies randomly in space and time along the fiber. This results in a time-dependent random fluctuation in PMD values at the fiber output. To account for these changes, statistical estimates are required. The mean value of the DGD $< \Delta \tau >$ is a good method of characterizing PMD for long fiber lengths (> 0.1 km) according to the formula [58, 2],

$$< \Delta \tau >= \text{PMD} \sqrt{L}, \tag{2.13}$$

where PMD is the polarization mode dispersion of the fiber and typically varies between 0.01 and 0.5 ps/$\sqrt{\text{km}}$.

Amplified Spontaneous Emission Noise

Erbium-Doped-Fiber-Amplifiers (EDFA)s are used to compensate for the power loss associated with the signal propagation through fiber. These amplifiers are considered as the main source of noise during propagation, which limits the maximum allowed distance for propagation. Signal amplification in EDFA is based on the population inversion technique, which leads to stimulated emission; this process is also associated with spontaneous emission, which generates ASE-noise [27]. ASE-noise can be modeled as independently and identically distributed Gaussian random processes with zero means, which impacts real and imaginary parts of the optical signal; therefore, it is hard to predict and mitigate this kind of noise.

The noise power generated by the EDFA over a bandwidth B can be expressed as,

$$P_{ASE} = 2n_{sp} \left(G - 1 \right) hfB, \tag{2.14}$$

where G is the gain of the amplifier, h is Planck's constant, f is the carrier frequency, and n_{sp} is the spontaneous emission factor;

$$n_{sp} = NF.\frac{G}{2\left(G - 1 \right)}. \tag{2.15}$$

The noise figure is defined as the ratio between input signal-to-noise ratio (SNR) and output SNR,

$$NF = \frac{SNR_{input}}{SNR_{output}} = \frac{P_{sig}/P_N}{G.P_{sig}/(P_{ASE} + P_N)}. \tag{2.16}$$

2.2.2 Nonlinear Effects

This section focuses in detail on the nonlinear impairments that affect the optical signal field propagation through standard SMF.

Sources of Nonlinearity

The homogeneous equation 2.10 only represents the linear impairments. Therefore, extending it to include the nonlinear propagation effects, results in NLSE with the form,

$$\frac{\partial A}{\partial z} + \frac{\alpha}{2}A + \frac{i}{2}\beta_2\frac{\partial^2 A}{\partial T^2} = i\gamma\left|A\right|^2 A. \tag{2.17}$$

Hence γ is the nonlinearity coefficient, which is expressed as,

$$\gamma = \frac{n_2\omega}{cA_{eff}}, \tag{2.18}$$

where A_{eff} is the effective fiber's core area, and n_2 is the nonlinear part of the refractive index of the optical fiber represented as

$$n_t = n + n_2|A|^2. \tag{2.19}$$

The Kerr effect, as introduced in the previous chapter, can be shown in the previous equation because of the nonlinear part of the refractive index of the optical fiber, which depends on the signal intensity.

SPM is the nonlinear interaction that happens between the optical signals in the same channel. Therefore, if we assume SMF propagation and neglect chromatic dispersion, equation 2.17 solution will have the form,

$$A(z,T) = A(0,T)\exp\left(-\tfrac{\alpha}{2}z\right)\exp\left(i\phi_{SPM}\right), \tag{2.20}$$

$$\phi_{SPM}(z,T) = \gamma L_{eff}|A|^2. \tag{2.21}$$

The phase shift ϕ_{SPM} occurring due to SPM is proportional to the optical intensity (see the upper part of figure 2.4), where L_{eff} is,

$$L_{eff} = \frac{1 - \exp\left(-\alpha L\right)}{\alpha}. \tag{2.22}$$

15
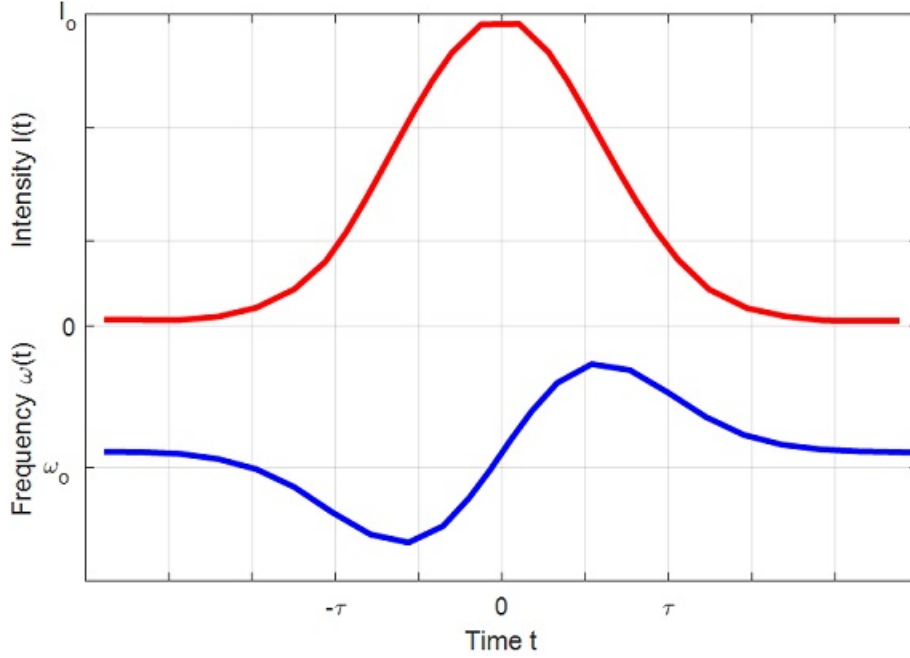
Figure 2.4: Self-phase modulation

This phase shift introduces carrier frequency fluctuation, referred to as chirp, as shown in the lower part of figure 2.4. The chirp generates new frequency components so that the leading edge shifts to lower frequencies ("redder" wavelengths), and the trailing edge to higher frequencies "bluer") while the very peak of the pulse is not shifted. The interaction of SPM with dispersion results in pulse distortion due to the time-dependence of the nonlinear phase shift.

We decompose the optical field A into three interacting field components to obtain insight into the SPM behavior in the situation of significant pulse overlap during transmission in SMF networks without fiber dispersion compensating. Hence, A_0, A_1, and A_2 describe pulses of the same wavelength channel in the time domain, with $\Delta\beta$ describing the difference in propagation constant between them due to the nonlinear phase shift associated with each pulse. Accordingly, equation 2.17 can be separated into three coupled equations. For example, in A_0:

$$\frac{\partial A_0}{\partial z} + \frac{\alpha}{2}A_0 + \frac{i}{2}\beta_2\frac{\partial^2 A_0}{\partial T^2} = i\gamma|A_0|^2 A_0 + 2i\gamma(|A_1|^2 + |A_2|^2)A_0 + i\gamma\sum_{l,m\neq 0} A_l A_m A_{l+m}^* \exp\left(i\Delta\beta z\right).$$

$$(2.23)$$

16

This mathematical representation shows clearly that the intra-channel self-phase modulation (ISPM) term $(i\gamma|A_0|^2 A_0)$, intra-channel cross-phase-modulation (IXPM) term $(2i\gamma(|A_1|^2 + |A_2|^2)A_0)$, and intra-channel four-wave-mixing (IFWM) terms $(i\gamma \sum_{l,m \neq 0} A_l A_m A_{l+m}^* \exp{(i\Delta\beta z)})$ depend on which pulses are involved when generating the contribution to the nonlinear phase shift.

Moving to a more practical case where the signal is transmitted in a single wavelength channel but using two orthogonal multiplexed polarization, Equation 2.17 can be as follows [3];

$$\frac{\partial A_X}{\partial z} + \frac{\alpha}{2}A_X + \frac{i}{2}\beta_2 \frac{\partial^2 A_X}{\partial T^2} = i\gamma(|A_X|^2 + \frac{2}{3}|A_Y|^2)A_X + \frac{i\gamma}{3}A_X^* A_Y^2 \exp{(-2i\Delta\beta z)}, \quad (2.24)$$

$$\frac{\partial A_Y}{\partial z} + \frac{\alpha}{2}A_Y + \frac{i}{2}\beta_2 \frac{\partial^2 A_Y}{\partial T^2} = i\gamma(|A_Y|^2 + \frac{2}{3}|A_X|^2)A_Y + \frac{i\gamma}{3}A_Y^* A_X^2 \exp{(-2i\Delta\beta z)}, \quad (2.25)$$

where A_X and A_Y correspond to the optical fields in X- and Y-polarization. The propagation constant difference,

$$\Delta\beta = \beta_x - \beta_y = (2\pi/\lambda)B_m = 2\pi/L_B, \quad (2.26)$$

is related to the fiber birefringence. If the fiber length L is much longer than the beat length L_B, the last term in equations 2.24 and 2.25 changes sign often and its contribution averages out to zero. The beat length describes the length required for the polarization to rotate 360 degrees. Therefore, in highly birefringent fibers, L_B typically around 1 cm, the four-wave-mixing term can often be neglected [2]. We can reduce the nonlinear phase shift incident on the X-polarization to:

$$\phi_{NL}(z, T) = \gamma L_{eff}(|A_X|^2 + \frac{2}{3}|A_Y|^2), \quad (2.27)$$

where the nonlinear phase shift ϕ_{NL} now splits into two parts: the SPM, as discussed earlier, and a second term representing the XPM that depends on the intensity profile of the orthogonal polarization.

2.2.3 Manakov Equation

The Coupled NLSE shown in equations 2.24, 2.25 assumes that the modeled fibers are perfectly circular and neglect the effect of randomly varying birefringence along the fiber. Therefore, for simplicity, the Manakov equation is applied instead of the coupled NLSE, as

17

it averages the field over random polarization fluctuations and has been able to accurately model propagation over the length scales crucial for optical communications [73]. The Manakov equation also assumes that both polarizations suffer equally from fiber nonlinearity because the birefringence scatters the polarization state on a much smaller length scale than the nonlinear length. Consequently, the NLSE, without the four-wave mixing nonlinear term, takes the form,

$$\frac{\partial A_x}{\partial z} + \frac{\alpha}{2} A_x + i\frac{\beta_2}{2}\frac{\partial^2 A_x}{\partial T^2} = i\gamma\frac{8}{9} \left(|A_x|^2 + |A_y|^2\right) A_x \tag{2.28}$$

$$\frac{\partial A_y}{\partial z} + \frac{\alpha}{2} A_y + i\frac{\beta_2}{2}\frac{\partial^2 A_y}{\partial T^2} = i\gamma\frac{8}{9}(|A_y|^2 + |A_x|^2) A_y \tag{2.29}$$

2.2.4 Numerical Techniques

The NLSE presented in equation 2.17 is a nonlinear partial differential equation that is difficult to solve analytically except in the cases in which the inverse scattering method [85] can be employed. Therefore, a numerical approach must be employed to understand the nonlinear effects in optical fiber propagation. A commonly used technique is the split-step Fourier method (SSFM) [30].

Split-Step Fourier Method

In this method, Manakov equations, shown in 2.28 and 2.29, can be written using the following operator notation,

$$\frac{\partial A}{\partial z} = (\hat{\mathcal{L}} + \hat{\mathcal{N}})A, \tag{2.30}$$

where $\hat{\mathcal{L}}$ and $\hat{\mathcal{N}}$ are linear and nonlinear operator with the following definitions:

$$\hat{\mathcal{L}} \triangleq -\frac{i}{2}\beta_2\frac{\partial^2}{\partial T^2} - \frac{\alpha}{2}, \tag{2.31}$$

$$\hat{\mathcal{N}} \triangleq i\gamma\frac{8}{9}|A|^2. \tag{2.32}$$

Dispersion and nonlinearity, in general, work together along the length of the fiber. The SSFM finds an approximation solution by assuming that the dispersive and nonlinear effects act independently while propagating the optical field across a small distance h. To illustrate, the propagation from z to $z + h$ is carried out in two steps. In the first step, the nonlinearity acts alone, and $\hat{\mathcal{L}} = 0$ in equation 2.30, while in the second step, dispersion
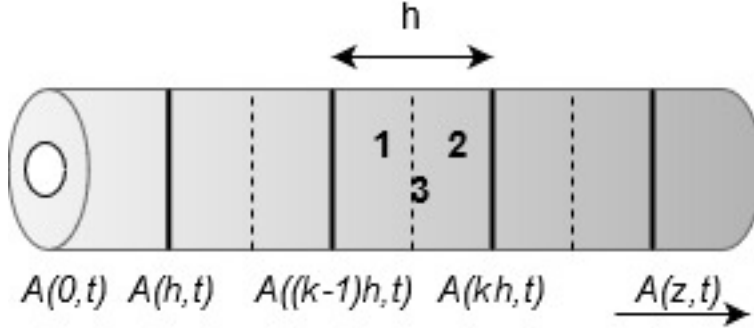
Figure 2.5: Split-step Fourier method Schematic illustration

acts alone, and $\hat{\mathcal{N}} = 0$ in equation 2.30. The solution to the Manakov equation under these conditions, at step $z = kh$, is given by [88]:

$$
\begin{aligned}
A(kh, T) &\approx \exp\left(\hat{\mathcal{L}}h + \hat{\mathcal{N}}h\right)A((k-1)h, T) = \\
&\exp\left(\hat{\mathcal{L}}\tfrac{h}{2}\right)\exp\left(\hat{\mathcal{N}}h\right)\exp\left(\hat{\mathcal{L}}\tfrac{h}{2}\right)A((k-1)h, T),
\end{aligned}
\tag{2.33}
$$

where $A((k-1)h, T)$ is the field solution for the previous step, and k is an integer number. The exponential operator $\exp(\hat{\mathcal{L}}h)$ can be evaluated in the Fourier domain using the formula,

$$
\exp(\hat{\mathcal{L}}h)B(z, T) = \mathcal{F}_T^{-1}\exp(\hat{\mathcal{L}}(-i\omega)h)\mathcal{F}_T B(z, T)
\tag{2.34}
$$

where $\mathcal{F}_T$ denotes the Fourier-transform operation, $\hat{\mathcal{L}}(-i\omega)$ is obtained from equation 2.31 by replacing the operator $\partial/\partial T$ by $-i\omega$, and ω is the frequency in the Fourier domain. The evaluation of equation 2.34 becomes straightforward and relatively fast. For this reason, the split-step Fourier method can be faster by up to two orders of magnitude compared with most finite-difference schemes. Figure 2.5 shows the steps of SSFM implementation. In the first and the second part (i.e., point 1 and 2), dispersion is calculated. While in the center of the step at point 3, nonlinearity is calculated. The above steps are continuously repeated till the end of the fiber length L to obtain the final field solution $A(L, T)$ [2].

2.3 Machine Learning Background

ML is considered as a subfield of AI in which computational algorithms are used to turn empirical data into usable models. Moreover, ML's goal is to increase the machines' ability for automatic learning through experience rather than being explicitly programmed. The

algorithms of ML are exceptionally helpful in areas where deploying explicitly written algorithms is unfeasible and impractical, especially with high-speed performance systems [21].

ML systems can be divided into three main types depending on the way the system is learning: reinforcement learning, unsupervised learning, and supervised learning. Reinforcement learning uses a trial and error feedback system to dynamically train the model in an interactive environment. Unsupervised ML models seek to identify previously unknown patterns and draw inferences from training data without any knowledge about pre-existing labels. Finally, supervised ML models use the training data labels to identify their complex patterns and eventually use these patterns to make decisions or predictions on new unseen data [102].

2.3.1 Supervised Machine Learning

The general goal of any supervised ML method is to use the features in the training data to model a target or outcome variable. To achieve this goal, a mathematical transformation is used to map the input features to the output variable, which is called the target function of the ML algorithm. The target function in most of the ML models is affected by the optimization of at least one model parameter, for example the number of trees in a random forest model, as will be discussed later.

Model Selection

The type of the target output variable (labels) is one of the main variables that influence the choice of the appropriate ML design. For instance, if the target variable is categorical (divided into categories) a model can be trained to classify an unseen observation to the correct category of the target variable. ML models employed in this setup seek to solve a ML classification problem. On the other hand, if the target variable is continuous, a model could be trained to predict the exact value of the target variable. ML models employed in this setup seek to solve a ML regression problem. The outputs of these classification and regression models can either be deterministic or probabilistic. Here, the ML model output could be the exact category or the value, or could generate probabilities for all possible outputs.

20

Figure 2.6: Input matrix to ML models with M samples or observations (rows) and N features (columns).

Feature Space

The feature space is the typically high-dimensional space containing the input information (features) for a ML model. The used features can be raw features that are used directly as input to the model, or mathematically manipulated from collected raw features before adding to the model. ML algorithms have the privilege of dealing with both homogeneous feature space and heterogeneous feature space that contains several input types such as binary, categorical, continuous.

Training Data

In any ML model, it is crucial to generate and collect adequate and balanced training cases. In addition to the ground truth label data which is used in supervised ML models for classification or regression problems. During the implementation of the ML model, the data input of the model is in matrix form as shown in figure 2.6 where the rows represent different samples of the data while the columns represent the dimensions of the samples (features). If the number of features increases compared to the number of samples, it becomes difficult for the ML model to achieve a generalized optimal solution to

the problem. In this case, the model enters an overfitting stage where it performs extremely well on the training data while its performance is poor on new data points.

On the other hand, if the model implemented is overly simple, under-fitting occurs in which the model performs poorly on both training and new data. In this case, the ML is incapable of learning the full complexity of the data. Therefore, researchers try to optimize their ML model by achieving an optimal balance between a model that 'under-fits' the training data and a model that overfits the training data (the so-called bias trade-off).

It should be noted that for most practical problems, there is no rule of thumb for deciding the proper number of training data needed to achieve optimal ML models training. Furthermore, choosing the number of samples required is not only related to the number of features, and the number of trainable model parameters but also related to the data complexity.

Another challenge related to the training data in ML models is the class imbalance. This issue arises from the lack of some training data and/or an inadequate representation of the data labels' variability. Consequently, some classes appear more frequently in the training set than others and the model becomes biased. More precisely, the model prediction accuracy will be higher for the more frequent classes than the infrequent ones that might even be completely disregarded in some cases [46].

Testing data

In order to evaluate the performance of the used ML model, after it has been trained, a data set independent of the training data is needed. The data used for testing are unlabeled, and it provides a final, real-world check of an unseen data set and compares the model's output with the actual samples' labels to confirm that the ML algorithm was trained effectively.

Model Evaluation

After selecting the ML model, different optimization techniques are used to minimize the error for each training example during the learning process. This error is called the loss function. Depending on the type of learning task, loss functions can be divided into two major categories: regression losses such as Mean Square Error, Mean Absolute Error (L1 Loss), and Mean Bias Error and classification loss such as Hinge Loss, and Cross Entropy Loss. An optimal metric should be selected in order to access the model's performance and efficiency properly. Some of the commonly used metrics to evaluate classification problems are accuracy, sensitivity, specificity, F1-score, and the area under the receiver operating

characteristic curve (ROC-AUC), while in regression problems correlation and the root mean squared error are used.

A well-known example for supervised machine learning is the Feedforward neural network FFNN. A FFNN consists of a group of neurons in which each neuron has a group of weighted inputs, a bias, and one output as shown in figure 2.7. The neuron's target function is given by,

$$y = f(\sum_{k=1}^{n} w_k x_k + b_k) \tag{2.35}$$

where x_k is the input from k-th neuron, w_k is the weight of the k-th input, b_k is the neuron's bias, and f is the activation function. There are several choices for the activation function such as sigmoid, tanh, and ReLU. In FFNNs, a layer is a collection of neurons that operate together at a certain depth. Every NN has an input and an output layer. Any intermediate layer is called a hidden layer. If the number of hidden layers is 2 or more and each layer is composed of relatively many neurons, then the network is referred to as a Deep Neural Network (DNN) as shown in figure 2.8. During the training of the NNs the input weights and biases are iteratively updated using optimization algorithms until a minimum mean squared error (MSE),

$$MSE = \frac{\sum_{1}^{n} (y - \hat{y})^2}{n} \tag{2.36}$$

where n, y, and $\hat{y}$ represent the batch size, the observed, and predicted output values respectively, is attained between the target of the training data and the NN output.

2.3.2   Unsupervised Machine learning

As discussed before, the unsupervised ML goal is to explore the data and find some structure within it. The data used to train the unsupervised ML algorithm is unlabeled data, which means it is applied on data that has no historical labels in the model. The unsupervised ML is mainly divided into two parts: Clustering and dimensionality reduction.

The clustering process starts by grouping together the similar entities, then using the grouped data to form clusters. The goal of such an unsupervised machine learning technique is to find similarities in the data points and group similar data points together and predicate the cluster to which the new data should belong.

Dimensionality Reduction is another type of Unsupervised ML model. Its goal is to remove the unwanted data from the input and consequently reduce the dimensions of the data. This technique also results in removing any undesirable or less effective features of
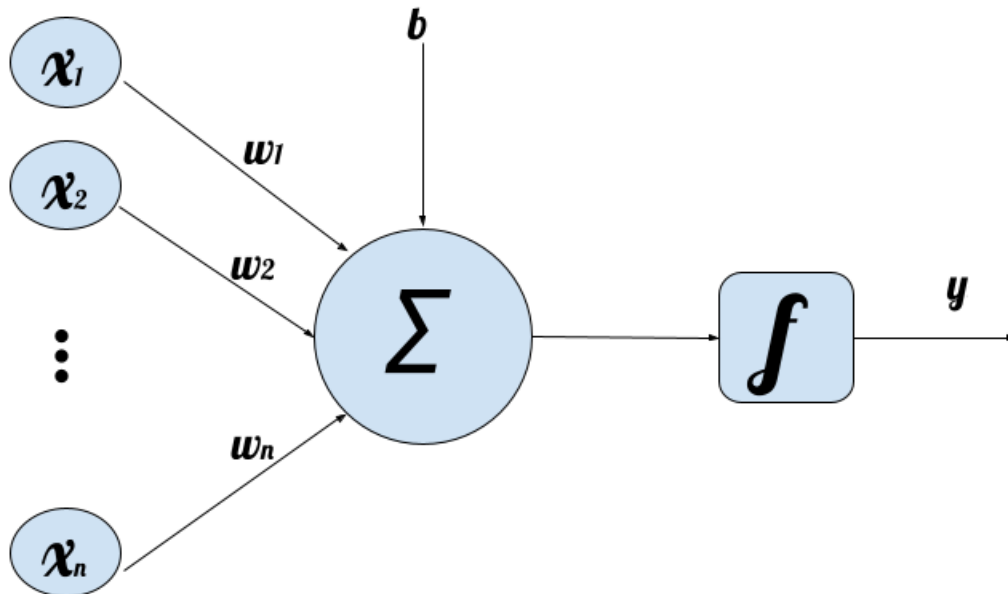
Figure 2.7: The Neuron structure

the data. It relates to the process of converting a set of data having large dimensions into data that have the same information with small sizes. In ML, there are different dimensionality reduction algorithms applied for different applications, such as PCA.

PCA is a feature extraction or dimension reduction approach that turns a data set from a d-dimensional space into a new coordinate system of p dimension, where $p < d$. The principal components (PCs) are the new variables that form the new coordinate system in which the variance under projection on these axes is maximal. Supposing that all the centered observations are stacked into the columns of a nxd matrix X and each column corresponds to a d-dimensional feature and there are n samples. The algorithm's goal is to find a linear combination of the columns of the matrix X with maximum variance. Such linear combinations are given as,

$$\sum_{j=1}^{j=d} u_j x_j = Xu \tag{2.37}$$

where u is a vector of constants $u_1, u_2, \ldots, u_d$. The variance of any such linear combination is given by,

$$var(u^{\mathrm{T}}X) = u^{\mathrm{T}}Su \tag{2.38}$$
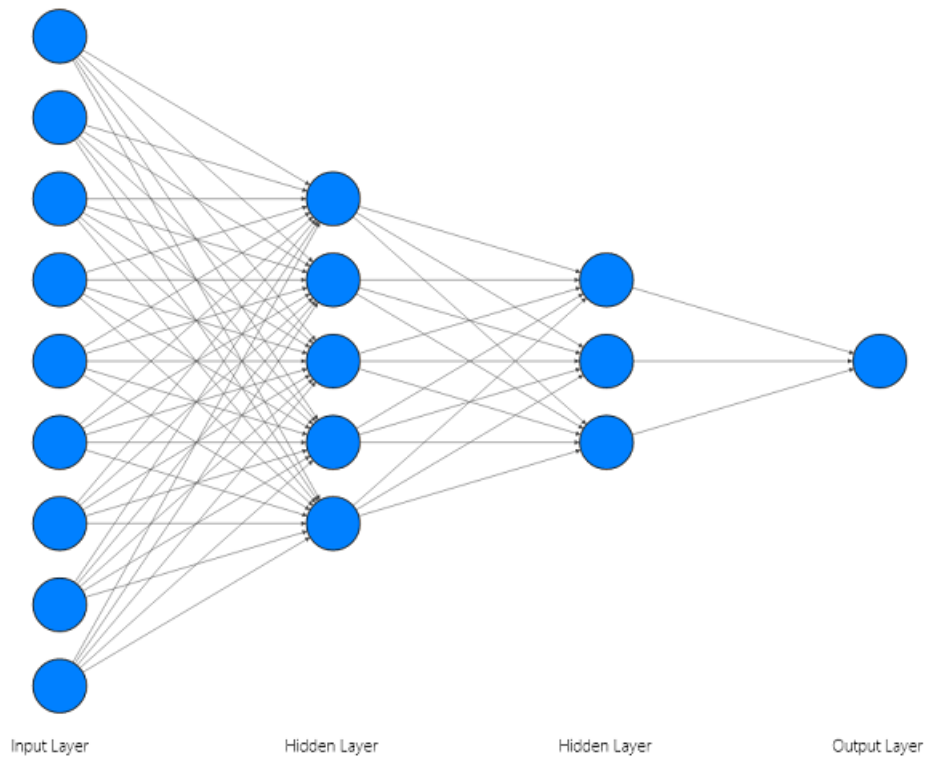
Figure 2.8: The Feed Forward Neural Network Structure

where S is the dxd sample covariance matrix of X and T denotes transpose. Clearly $var(u^{\mathrm{T}}X)$ can be maximized by obtaining a u matrix which maximizes the quadratic form $u^{\mathrm{T}}Su$. An additional limitation must be applied for this issue to have a well-defined solution, and the most common restriction is working with unit-norm vectors where $u^{\mathrm{T}}u = 1$. To solve this optimization problem a Lagrange multiplier λ is introduced where,

$$L(u, \lambda) = u^{\mathrm{T}}Su - \lambda(u^{\mathrm{T}}u - 1) \tag{2.39}$$

Differentiating with respect to the vector u, and equating to the null vector, produces the equation,

$$Su = \lambda u \tag{2.40}$$

Thus u is the eigenvector, and λ is the corresponding eigenvalue, of the covariance matrix S. This shows that the first principal component is given by the eigenvector with the largest associated eigenvalue of the sample covariance matrix S.

Similarly, any $d \times d$ real symmetric matrix, such as a covariance matrix S, has exactly d real eigenvalues, $\lambda_k(k = 1, \ldots, d)$, and their corresponding eigenvectors can be defined to form an orthonormal set of vectors u_k. In PCA, we select the p dominating eigenvectors based on the percentage of the original data's variance that must be preserved.

2.4   Summary

This chapter covered a background for both the propagation through optical fiber as a NLSE and the ML algorithms. For the optical fiber propagation part, both the sources and the effects of the linear and nonlinear impairments affecting the propagation through SMF have been discussed. Moreover, the split-step Fourier method has been defined as the most common numerical method used to model the propagation through optical fibers. Also, the Manakov equation has been introduced as a reliable approximation for coupled NLSE. In the second part of the chapter, types of ML algorithms have been discussed and the factors affecting the model selection and training. In the next chapter, a literature review will be presented about the different digital signal processing techniques that are used to estimate the received signal after propagation and mitigate the linear and nonlinear perturbations added to the signal.

Chapter 3

Fiber Digital Signal Processing Literature Review

3.1 Introduction

This chapter presents a comprehensive literature survey of different techniques that have been proposed to reconstruct the signal propagated through an optical fiber. Some of these methods are model-driven, where the known analytical model representation for the propagating channel is used with some changes in the fiber parameters. On the other hand, other methods use a data-driven approach where the received data is used to construct the inverse model of the channel and predict the transmitted signal without any previous knowledge of the parameters of the system or the link.

3.2 Digital Back-propagation

As discussed in the previous chapter, the forward propagation of the optical field through an optical fiber can be solved numerically by applying the SSFM to the Manakov equation.

Temporal pulse broadening and dispersive self-phase modulation compensation have been studied since the end of the 1970s [31, 111]. However, the first application of DBP that employed SSFM as a universal post-compensation scheme based on coherent detection and DSP for an optical signal propagating through an optical fiber was only advanced in 2008[63]. The algorithm is effectively able to fully reconstruct the signal and reverse

the effects of signal-signal nonlinear interference (NLI) [51, 15, 50]. In this algorithm, a modified Manakov equation is used to reconstruct the transmitted signals by inverting the signs of the dispersion and nonlinear terms in the forward propagation version shown in equations 2.28 and 2.29. Consequently, the corresponding linear ($\hat{\mathcal{L}}$) and nonlinear ($\hat{\mathcal{N}}$) operators are modified to the form,

$$\hat{\mathcal{L}} \triangleq \frac{i}{2}\beta_2\frac{\partial^2}{\partial T^2} - \frac{\alpha}{2}, \tag{3.1}$$

$$\hat{\mathcal{N}} \triangleq -i\gamma\frac{8}{9}|E|^2, \tag{3.2}$$

and the numerical solution for the modified equations can be expressed as,

$$E(z+h,T) \approx \exp{(\hat{\mathcal{L}}\frac{h}{2})}\exp{(\int_z^{z+h}\hat{\mathcal{N}}(s)ds)}\exp{(\hat{\mathcal{L}}\frac{h}{2})}E(z,T), \tag{3.3}$$

$$\int_z^{z+h}\hat{\mathcal{N}}(s)ds \approx \frac{h}{2}(\hat{\mathcal{N}}(z) + \mathcal{N}(z+h)) \tag{3.4}$$

where h is the step size. The integral for $\hat{\mathcal{N}}(s)$ is generally approximated to equation 3.4 by the trapezoidal rule, and two iterations are often used [51, 2]. The accuracy of the iterative SSFM improves by increasing the number of iterations or the number of steps per span, where ideally the best performance is achieved when the step size is identical to the step size used in forwarding propagation.

In practice, to run the SSFM faster, equation 3.3 over M successive steps can be calculated using the expression:

$$E(L,T) \approx \exp{(-\hat{\mathcal{L}}\frac{h}{2})}(\prod_{m=1}^{M}\exp{(\hat{\mathcal{N}}h)}\exp{(\hat{\mathcal{L}}h)})\exp{(\hat{\mathcal{L}}\frac{h}{2})}E(0,T), \tag{3.5}$$

where $L = Mh$ is the total fiber length, and the integral in equation 3.4 was approximated with $\hat{\mathcal{N}}h$. Thus, except for the first and last dispersive steps, all intermediate steps can be carried over the whole segments' length h.

Despite the DBP theoretical beneficial effects and its accurate solution for a reconstructed optical field, many factors limit the performance of this method, such as NLI arising from the interaction between the signal pulses, ASE-noise [81], PMD [36], and DSP massive complexity at the receiver [67]. DBP complexity depends on the fast Fourier transform (FFT) block-size and the total number of propagation steps N_{step} with the order of $O(N_{step}N_{FFT}log_2(N_{FFT}))$, taking into consideration that $N_{FFT} = N_{sym}N_{sp}$, where N_{sp} is the number of samples per symbol and N_{sym} is the total number of symbols fed to each FFT block [44].
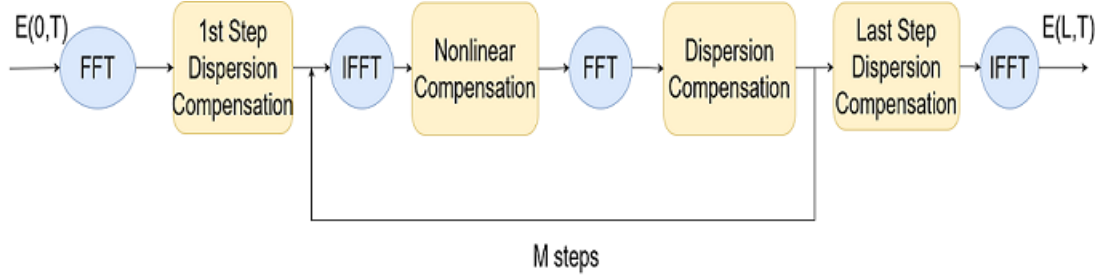
Figure 3.1: Block Diagram of Digital Back-propagation

3.3 Volterra Nonlinear Compensation

The Volterra series is a commonly used numerical tool in modeling and mitigating nonlinear phenomena [84]. In this tool, a truncated nth-order polynomial is combined with a memory effect in a series of convolution integrals to achieve its purpose. It was firstly proposed by Peddanarappagari and Brandt-Pearce for the modeling of SMF transmission systems [79] by solving the NLSE in the frequency-domain. This was achieved by enabling the extraction of the Volterra series transfer function (VSTF) which is a set of nth-order nonlinear transfer functions which model the input-output relationship. The frequency domain VSTF for NLSE, retaining only the first five kernels, can be expressed as follows,

$$
\begin{aligned}
E(\omega, z) = {} & K_1(\omega, z)E(\omega) + \\
& \int\int K_3(\omega_1, \omega_2, \omega - \omega_1 + \omega_2, z)E(\omega_1)E^*(\omega_2)E(\omega - \omega_1 + \omega_2)d\omega_1 d\omega_2 + \\
& \int\int\int\int K_5(\omega_1, \omega_2, \omega_3, \omega_4, \omega - \omega_1 + \omega_2 - \omega_3 + \omega_4) \\
& E(\omega_1)E^*(\omega_2)E(\omega_3)E^*(\omega_4)E(\omega - \omega_1 + \omega_2 - \omega_3 + \omega_4)d\omega_1 d\omega_2 d\omega_3 d\omega_4
\end{aligned}
\tag{3.6}
$$

where $K_1(\omega, z)$ is the linear transfer function, $K_3(\omega_1, \omega_2, \omega_3, z)$, and $K_5(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, z)$ are the third and fifth-order nonlinear transfer functions (Volterra kernels) of an optical fiber of length z, respectively.

Guiomar et al. proposed the use of the frequency domain VSTF as a powerful tool to compensate for the fiber nonlinearity in signal-polarization optical transmission instead of the DBP [44]. This was established by inverting the 3rd order nonlinear transfer function obtained from its forward propagation version, equation 3.6, but with the opposite sign for fiber parameters $(-\alpha, -\beta_2, -\gamma)$; this method is called an inverse VSTF (IVSTF). This proposed method can also be extended to polarization multiplexed signals propagating in multi-span fiber where the x-polarization signal for the nonlinear compensated optical field in the frequency domain, $\widetilde{E_x}^{NL}$, is given by [15]

29

$$\widetilde{E_x}^{NL}(\omega_n, z - L) = -i\frac{8}{9}\xi\gamma K_1(\omega_n, L)\sum_{m=1}^{N}\sum_{k=1}^{N} K_3(\omega_n, \omega_k, \omega_m)\widetilde{E_x}(\omega_{n+m-k}, z)$$
$$*[\widetilde{E_x}(\omega_k, z)\widetilde{E_x}^*(\omega_m, z) + \widetilde{E_y}(\omega_k, z)\widetilde{E_y}^*(\omega_m, z)], \tag{3.7}$$

Hence $\widetilde{E_x}$ is the x-polarization frequency-domain received signal, L is the step size of the IVSTF (multiple of the span length), N is the FFT block size, $0 < \xi \leq 1$ is a free optimization parameter, and ω_n is the angular frequency placed at an index n in the FFT block. The multi-span linear kernel, K_1, represents the attenuation and chromatic dispersion as

$$K_1(\omega_n, z) = \exp\left(\frac{\alpha}{2}L_s - \frac{i\beta_2}{2}\omega_n^2 z\right) \tag{3.8}$$

and the multi-span 3rd order nonlinear kernel, K_3, is given by

$$K_3(\omega_n, \omega_k, \omega_m) = \frac{1 - \exp\left(\alpha L_s - i\beta_2(\omega_k - \omega_n)(\omega_k - \omega_m)L_s\right)}{-\alpha + i\beta_2(\omega_k - \omega_n)(\omega_k - \omega_m)}F(\omega_n, \omega_k, \omega_m) \tag{3.9}$$

where $F(\omega_n, \omega_k, \omega_m)$ is the multi-span phased array factor represents the nonlinearity coherent accumulation between fiber spans:

$$F(\omega_n, \omega_k, \omega_m) = \exp\left(-i\frac{\beta_2(\omega_k - \omega_n)(\omega_k - \omega_m)}{2}(L - L_s)\right)\frac{\sin\left(\beta_2(\omega_k - \omega_n)(\omega_k - \omega_m)\frac{L}{2}\right)}{\sin\left(\beta_2(\omega_k - \omega_n)(\omega_k - \omega_m)\frac{L_s}{2}\right)} \tag{3.10}$$

The nonlinear equalized optical field, $\widetilde{E_x}^{NL}$, is then added to the chromatic dispersion compensation (CDC) signal, resulting in the output optical field after each IVSTF step as

$$\widetilde{E_x}^{eq}(\omega_n, z - L) = K_1(\omega_n, L)\widetilde{E_x}(\omega_n, z) + \widetilde{E_x}^{NL}(\omega_n, z - L), \tag{3.11}$$

$$\widetilde{E_y}^{eq}(\omega_n, z - L) = K_1(\omega_n, L)\widetilde{E_y}(\omega_n, z) + \widetilde{E_y}^{NL}(\omega_n, z - L), \tag{3.12}$$

The paper [44] shows that a 3rd-order truncated IVSTF could achieve higher performance than DBP at a low sampling rate (2 samples per symbol). This is due to avoiding the repeated transition between the time and frequency domains. However, the complexity overhead of the IVSTF numerical implementation is a major challenge. The complexity of IVSTF depends on the number of operations per sample, which is $O(N^2)$ due to the double summation performed in equation 3.7. Therefore, the choice of the FFT block size is critical.

30

3.4   Advanced Modulation Transmission

One of the main stages in optical communication is digital modulation, which transforms a digital sequence to an analog waveform to be reliably transmitted over a communication channel. In this process, a sequence of binary bits is converted to a sequence of symbols that is unique to each modulation format, pulse shaping, and modulating a high-frequency carrier [80]. Changing the modulation format of the system directly affects the achievable information rate (AIR) of the system, as the AIR is a metric that indicates the amount of information, bits per symbol, that can be reliably transmitted through the channel. Moreover, it determines the upper and lower bounds of optical fiber channel capacity [5]. AIR is also computed from the mutual information (MI) between the channel input sequence (constellation points) X_1^K and channel output sequence Y_1^K of length K [65] as a function of both the entropy $\mathcal{H}(X_1^K)$ and the conditional entropy $\mathcal{H}(X_1^K|Y_1^K)$,

$$\mathcal{I}(X;Y) = \lim_{K\to\infty} \frac{1}{K}\mathcal{I}(X_1^K;Y_1^K) = \lim_{K\to\infty} \frac{1}{K}[\mathcal{H}(X_1^K) - \mathcal{H}(X_1^K|Y_1^K) \tag{3.13}$$

where $\mathcal{H}(\mathcal{X})$ is the entropy function,

$$\mathcal{H}(X) = -\sum_x P_X(x)logP_X(x), \tag{3.14}$$

where $P_A(a) = P(A = a)$ is the probability mass function of a random variable A takes the discrete value a. $\mathcal{H}(X|Y)$ is the average uncertainty about X after observing a second random variable Y [43],

$$\mathcal{H}(X|Y) = \sum_y P_Y(y)[-\sum_x P_{X|Y}(x|y)logP_{X|Y}(x|y)], \tag{3.15}$$

As can be noticed from equation 3.13, MI expression contains both the entropy, in which its value sets an upper limit on the AIR, and the spectral efficiency and the conditional entropy that represents the received signal quality. Therefore, AIR is highly affected by the modulation alphabet X. For example, constellation alphabets that decrease nonlinear interference noise (NLIN), which consists of the nonlinear interference effects between signal-signal, signal-ASE-noise, and ASE-ASE, consequently enhance the signal-to-noise-plus-interference ratio, also known as the effective signal-to-noise ratio (SNR). This results in uncertainty reduction $\mathcal{H}\left(X_1^K|Y_1^K\right)$. Due to the constraints associated with the construction of such constellations design, contradictions may emerge due to the reduction in the entropy $\mathcal{H}\left(X_1^K\right)$. Generally, a constellation's design includes both the positions of the points in the I/Q plane that referred to as "Geometric Shaping" and their probabilities which is known as "Probabilistic Shaping."

3.4.1 Geometric Shaping

In the Geometric shaping (GS) method, the probability of the constellation points is uniform; however, the position of the constellation points in the Euclidean space is changed compared with the conventional cubic structures, which are widely used in communication systems. One of the leading research papers on GS for optical fiber systems was in 2009 by Freckmann et al. [32]. In their work, they proposed and optimized ring constellations for single polarized signal transmission. Their main idea was to limit the peak-to-average ratio of the power and consequently mitigate the fiber non-linearity by restricting high-energy symbols in the constellation. A similar constellation design was described and investigated in [64] where a total of 256 symbols were distributed on a ring constellation. The ring radii and the density of the symbols distributed on each ring were optimized using an iterative polar quantization method. Several other papers studied GS by adding constraints on the allowed multidimensional sequences. Both polarization and time slots were dimensions that were studied [74]. Moreover, 4-D constellations were designed using multi-sphere distributions in [106] to be applied in practical nonlinear channels. These constellations have a rotationally invariant and discrete amplitude, while their phases are continuous and uniform [57].

GS has some serious practical drawbacks, such as (i) It is hard to find a simple solution to allocate the location of the points of the GS constellation under arbitrary channel conditions; (ii) The irregular GS constellation points add more complexity to coherent DSP to achieve a robust signal recovery prior to decoding stage.

3.4.2 Probabilistic Shaping

Probabilistic shaping (PS) is a method that targets increasing the MI that can be achieved by optimizing the probability mass function (PMF) of the input symbols, $P_X(x)$, under the constraint of limited average power [28]. This leads to reducing the entropy $\mathcal{H}(X)$, as shown in equation 3.14, and thus maximizing the format spectral efficiency.

Finding the optimum distribution is not always possible, and sometimes the channel itself is mathematically complicated, e.g., nonlinear optical fiber channel [4]. However, the benefit from the non-uniform PMF can be shown near capacity-achieving systems in which the AIR is more affected by the receiver SNR than the entropy. Therefore, Benjamin Smith et al. in [90] performed PS using a trellis shaping method and simulated near-capacity performance. Moreover, 4D PMF was considered in [9] in which the points with smaller multidimensional amplitude, I/Q dimensions in each polarization, appear with higher probability.

In an additive White Gaussian noise (AWGN) channel case, the input Gaussian distribution maximizes the MI between the input and output of the channel and results in the Shannon capacity formula [86] which is a theoretical limit to the maximum data rate that can be transmitted in a channel with a given bandwidth. Usually, the source is considered with discrete symbols or messages. Therefore, a discrete Gaussian-like distribution for the input data, namely the Maxwell-Boltzmann (MB) distribution, should be used to optimize the MI in the AWGN channel with nonlinear distortion [49]. Consequently, PMF optimization was investigated in [29], where the PMF was drawn from the MB family, where any constellation point x, sampled from a constellation of cardinality M as a random variable X, is therefore transmitted with a probability,

$$P_x(X = x) = \frac{1}{\sum_{j=1}^{M} \exp(-v|x_j|^2)} \exp\left(-v|x|^2\right). \tag{3.16}$$

Hence v is the shaping factor. For $v = 0$, the probability distribution becomes uniform, and for $v > 0$, the constellation points will have a non-uniform probability. Increasing v causes the inner points of the constellation to have a higher probability than the outer constellation points. The PMF can be adjusted to the channel effective SNR by carefully optimizing the scaling parameter v. An experimental demonstration of combining the selection of MB PMFs with a convolutional low-party check code was presented in [13]. The gain was achieved experimentally by rate-matching the independent identically distributed input binary data to the specific MB PMF for different AIRs that were identical. However, different schemes have been suggested for the implementation of PS; the one that has had significant attention in the optical communications community is probabilistic amplitude shaping (PAS). The PAS applies coding and shaping in an independent structure with separate optimizations, which reduces the problem of burst errors and shaping distortion that are present in other realizations of PS [19].

These advanced constellations are challenging because they require non-conventional equalization and/or phase noise recovery. Therefore, pilot symbols (group of symbols is sent before the signal) at a rate of $1-2\%$ are needed for both purposes [39, 109]. Moreover, PMFs with memory may be used to increase the gain. Optimization of such PMFs is challenging, and accordingly, the receiver optimal processing becomes exponentially complex for systems with high spectral efficiency.

3.5 Perturbation Based Pre-compensation

The main idea of this algorithm is first to evaluate the intra-channel nonlinearity perturbation and then to remove the perturbation from the received signal to regenerate the transmitted waveform. The perturbation analysis technique evaluates an approximated solution in the time-domain to the Manakov equation, where the propagating total field is expressed as,

$$E(z,t) = E_0(z,t) + \Delta E(z,t) \tag{3.17}$$

where $E_0(z,t)$ is the linear propagation solution and $\Delta E(z,t)$ is the perturbation caused by the nonlinear effects. Therefore, the perturbations $\Delta E(z,t)$ for each polarization should satisfy,

$$\frac{\partial \, \Delta E_x(t,z)}{\partial z} + i\frac{\beta_2}{2}\frac{\partial^2 \, \Delta E_x(t,z)}{\partial t^2} = i\gamma \, \frac{8}{9}(|E_{x,0}(t,z)|^2 + |E_{y,0}(t,z)|^2) \, E_{x,0}(t,z), \tag{3.18}$$

$$\frac{\partial \, \Delta E_y(t,z)}{\partial z} + i\frac{\beta_2}{2}\frac{\partial^2 \, \Delta E_y(t,z)}{\partial t^2} = i\gamma \, \frac{8}{9}(|E_{y,0}(t,z)|^2 + |E_{x,0}(t,z)|^2) \, E_{y,0}(t,z). \tag{3.19}$$

Applying Fourier transform on these equations,

$$\frac{\partial \, \Delta E_x(\omega,z)}{\partial z} - i\frac{\omega^2\beta_2}{2}\Delta E_x(\omega,z) = i\gamma \, \frac{8}{9}(\int e^{(-i\omega t)}E_{x,0}E_{x,0}^*E_{x,0}dt + \int e^{(-i\omega t)}E_{y,0}E_{y,0}^*E_{x,0}dt), \tag{3.20}$$

$$\frac{\partial \, \Delta E_y(\omega,z)}{\partial z} - i\frac{\omega^2\beta_2}{2}\Delta E_y(\omega,z) = i\gamma \, \frac{8}{9}(\int e^{(-i\omega t)}E_{y,0}E_{y,0}^*E_{y,0}dt + \int e^{(-i\omega t)}E_{x,0}E_{x,0}^*E_{y,0}dt). \tag{3.21}$$

Solving these equations in the frequency domain for z=L,

$$\Delta E_x(\omega,z) = \exp(i\omega^2\beta_2 L/2)\int_0^L F_x(\omega,z)\exp(-i\omega^2\beta_2 z/2)dz, \tag{3.22}$$

where

$$F_x(\omega,z) = i\gamma\frac{8}{9}[\int \exp(-i\omega t)E_{x,0}E_{x,0}^*E_{x,0}dt + \int \exp(-i\omega t)E_{y,0}E_{y,0}^*E_{x,0}dt] \tag{3.23}$$

And

$$\Delta E_y(\omega,z) = \exp(i\omega^2\beta_2 L/2)\int_0^L F_y(\omega,z)\exp(-i\omega^2\beta_2 z/2)dz, \tag{3.24}$$

where

$$F_y(\omega,z) = i\gamma\frac{8}{9}[\int \exp(-i\omega t)E_{y,0}E_{y,0}^*E_{y,0}dt + \int \exp(-i\omega t)E_{x,0}E_{x,0}^*E_{y,0}dt] \tag{3.25}$$

Applying the inverse Fourier transform with assuming a Gaussian shape for six transmitted optical pulses $\sqrt{P_0}X_{m/l/n}\exp(-(t-T_{m/n/l})^2/2\tau^2)$ and $\sqrt{P_0}Y_{m/l/n}\exp(-(t-T_{m/n/l})^2/2\tau^2)$ at three timings T_m, T_l, T_n, analytical expressions for the perturbation expansion exists in the form of the exponential integral function [68, 98].

$$\Delta E_x(t+T_m-T_l+T_n,L) = i\frac{8}{9}\gamma P_0^{3/2}[X_nX_l^*X_m+Y_nY_l^*X_m]\exp(\frac{-t^2}{6\tau^2})$$

$$\int_0^L dz \frac{1}{\sqrt{1+2i\beta_2 z/\tau^2+3(\beta_2 z/\tau^2)^2}}\exp\left(-\frac{3[2t/3+T_m-T_l][2t/3+T_n-T_l]}{\tau^2(1+3i\beta_2 z/\tau^2)} - \frac{(T_n-T_m)^2}{\tau^2[1+2i\beta_2 z/\tau^2+3(\beta_2 z/\tau^2)^2]}\right),$$
(3.26)

And

$$\Delta E_y(t+T_m-T_l+T_n,L) = i\frac{8}{9}\gamma P_0^{3/2}[Y_nY_l^*Y_m+X_nX_l^*Y_m]\exp(\frac{-t^2}{6\tau^2})$$

$$\int_0^L dz \frac{1}{\sqrt{1+2i\beta_2 z/\tau^2+3(\beta_2 z/\tau^2)^2}}\exp\left(-\frac{3[2t/3+T_m-T_l][2t/3+T_n-T_l]}{\tau^2(1+3i\beta_2 z/\tau^2)} - \frac{(T_n-T_m)^2}{\tau^2[1+2i\beta_2 z/\tau^2+3(\beta_2 z/\tau^2)^2]}\right),$$
(3.27)

where P_0 is the pulse peak power at the launch point, $X_{m/n/l}$, $Y_{m/n/l}$ are the symbol complex amplitude imposed by data modulation, and $m/n/l$ denote symbol indices. If the current symbol (symbol of interest) is zero, then $l = m + n$.

This approach has shown promising results for intra-channel fiber nonlinearities pre-compensation [25] and post-compensation [78, 40] as well. In typical high-capacity transmission systems, especially ones without inline CDC, the pulse spreading due to CD is usually much larger than the symbol duration τ (i.e., $\beta_2 z \gg \tau^2$). Since the pre-distortion is assumed to operate at the symbol rate, only the perturbation value at $t = 0$ is considered and can be simplified as [98]

$$\Delta E_x = P_0^{\frac{3}{2}}\sum_{m,n}\left(X_nX_{m+n}^*X_m+Y_nY_{m+n}^*X_m\right)C_{mn} \tag{3.28}$$

$$\Delta E_y = P_0^{\frac{3}{2}}\sum_{m,n}\left(Y_nY_{m+n}^*Y_m+X_nX_{m+n}^*Y_m\right)C_{mn} \tag{3.29}$$

The corresponding nonlinear perturbation coefficient C_{mn} is given by

$$C_{mn} = j\frac{8}{9}\frac{\gamma\tau^2}{\sqrt{3}\,|\beta_2|}\,\mathrm{E}_1(-j\frac{mnT^2}{\beta_2 L}) \qquad m\neq 0, n\neq 0, \tag{3.30}$$

$$C_{mn} = j\frac{8}{9}\frac{\gamma\tau^2}{\sqrt{3}\,|\beta_2|}\frac{1}{2}\,\mathrm{E}_1\left(\frac{(n-m)^2\,T^2\tau^2}{3\,|\beta_2|^2\,L^2}\right) \qquad m\ or\ n=0, \tag{3.31}$$

$$C_{00} = j\frac{8}{9}\frac{\gamma\tau^2}{\sqrt{3}\,|\beta_2|}\int_0^L dz\frac{1}{\sqrt{\frac{\tau^4}{3\beta_2^2}+z^2}} \tag{3.32}$$

where τ is the pulse width, T is the inverse of symbol rate, and L is the transmission distance, and $E_1(.)$ is the exponential integral function [1]:

$$E_1(z) = \int_z^\infty \frac{\exp(-t)}{t}dt \qquad (|arg\ z| < \pi) \tag{3.33}$$

Equation 3.30 represents the coefficients of the intra-channel four-wave mixing (IFWM) terms, equation 3.31 shows the intra-channel cross-phase modulation (IXPM) terms, and equation 3.32 shows the intra-self phase modulation term [61]. To compensate for the ignored fiber attenuation in the previous derivation, the $\mathbf{C_{mn}}$s are scaled by a factor L_{eff}/L_{span}, where L_{eff} is the effective length of each span and L_{span} is the span length.

To reduce this pre-compensation technique's computational and development complexity, different techniques are proposed, such as aggressive quantization of the expansion coefficients [115] and the use of symmetric electronic dispersion compensation (SEDC) and root-raised-cosine (RRC) pulse shaping [37]. The reduction of complexity in SEDC results from two simplifications: 1) there are no real parts of the coefficients $\text{Re}[C_{mn}]$, and 2) the imaginary parts of the coefficients $\text{Im}[C_{mn}]$ are computed based on half of the link length $L/2$, as shown in figure 3.2. Moreover, using RRC pulse shape reduces the number of terms needed in the summations due to the reduction of the dispersion-induced pulse spreading.

The perturbation-based technique is helpful, as it can be used to pre-compensate accumulated intra-channel fiber nonlinearities with only one computation step for the entire link and can be calculated using one sample per symbol [68, 98]. However, the complexity of the system is still high due to the large number of complex multiplications included in calculating the perturbation and C_{mn} coefficients, as described in chapter 8. Moreover, potential improvements in system performance are still needed to extend the algorithm to account for inter-channel nonlinearities.

3.6 Artificial Intelligence in Telecommunication

Over the past decades, extensive research has been conducted on optical communication networks and systems to improve the configuration and operation of network devices, monitor the optical performance, recognize the modulation format, mitigate the fiber non-linearities, and estimate the quality of transmission (QoT).
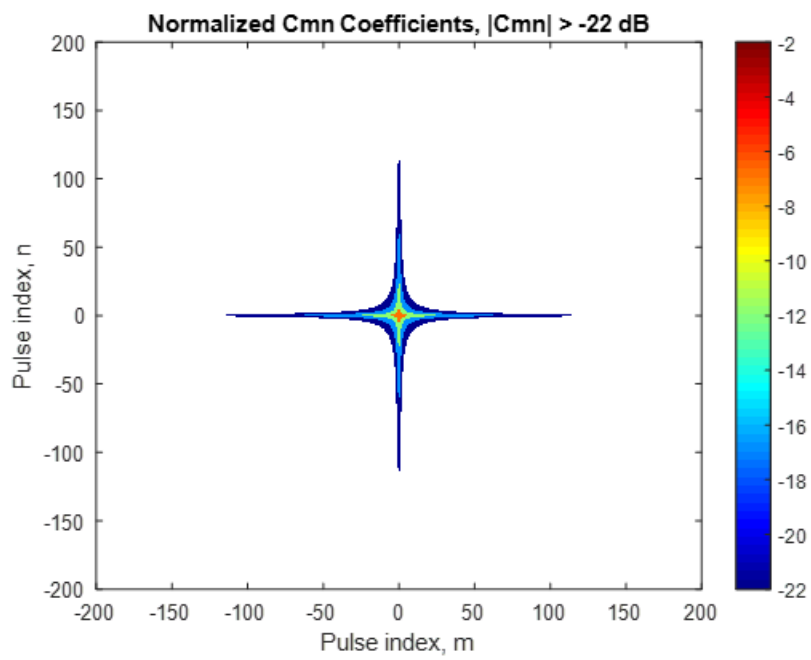
Figure 3.2: Normalized $\mathrm{Im}[C_{mn}(L/2)]$ coefficients for using RRC pulse shaping and SEDC in standard single-mode fiber.

3.6.1 Transmitters Operation and Characterization

The principle role of AI techniques is to facilitate statistical modeling of individual optical components by including the underlying physics in the characterization and demodulation algorithms, especially when the deterministic approach has an impractical computational overhead. Therefore, the improvement tools based on supervised and unsupervised learning algorithms are having an accurate and promising performance.

With the aim to increase spectral efficiency using advanced modulation formats, the need for robust synchronization tools for carrier frequency and phase becomes a necessity. In order to solve this problem, in conventional time-domain approaches, a combination of coherent detection and digital coherent receiver is needed [26, 59]; however, the estimated phase noise can be compromised, in more advanced modulation formats, with acceptable measurement noise. Various AI algorithms have been proposed to overcome these problems. For example, Zibar et al. [116] presented a combination framework of Bayesian filtering with expectation maximization (EM) parameter estimation to precisely identify the amplitude and phase noise of the used laser. In contrast to the conventional time-domain case, the results shown in the paper demonstrate that the proposed AI technique accurately estimates the phase noise even under large measurement noise.

Other applications for the use of AI techniques in transmitters optimization include the work by Brunton et al. [12] in which they propose genetic ML algorithms in combination with adaptive control techniques to help mode-locked fiber lasers in achieving a self-tuning mechanism. Brunton then extended the work on self-tuning with other groups [7, 93] by using deep learning and reinforcement learning techniques.

3.6.2 Erbium-Doped Fiber Amplifiers Operation

ML techniques have been introduced to efficiently solve a variety of challenges related to the EDFA operation within optical fiber propagation. To illustrate, Huang et al. [48] defined a regression problem using historical data to identify the effect of power excursion in multi-span EDFA networks on the channel performance using supervised machine learning with a radial basis function. The proposed technique could minimize the power disparity among channels by accurately providing the system with recommended methods on channel add/drop. This study has been extended in [47] taking the advantage of flex grid networks to enhance the system spectral efficiency by re-optimizing the assignment of the spectrum to active connections using dynamic defragmentation. This was achieved by using cascaded stages of AI algorithms; first, a regression model is used to identify

the impact of a given sub-channel, then a logistic regression is employed to determine if the contribution will result in an improvement or deterioration in the discrepancy among post-EDFA powers. The results show that applying adaptive and dynamic adjustments of pre-EDFA sub-channel powers was able to significantly decreasing post-EDFA power discrepancy. Additionally, Barboza et al. [6] proposed using MLP neural network to autonomously adjust the amplifiers' operating point in a cascaded EDFA system. The adjustment of the amplifiers' operating point optimizes the link performance. It helps ensure precise predefined input and output power levels achieving a minimum noise figure and discrepancy in the frequency response of the transmission system. Moreover, in [110] Yankov et al. introduced a cascade model of a ML-based EDFA gain. This model could predict the output power profile of a different number of spans and span lengths systems in real-time.

3.6.3   Performance Monitoring

Adapting the performance of the link parameters that change continuously with time, such as Optical signal to noise ratio (OSNR), PMD, CD, and nonlinearity sources, is one of the crucial challenges in telecommunication networks control and management. Estimating these parameters allows real-time diagnoses for the network and consequently takes immediate actions against any failure. These actions could be repairing the damage, or changing the traffic to a non-optimal route, or adjusting the system compensators or equalizers. AI techniques have been proposed to monitor these parameters[24].

For instance, Wu et al. [108] studied performance monitoring using artificial neural networks (ANN). The parameters extracted from the system's eye diagram are simultaneously employed to monitor accumulated system parameters. The eye diagram is a visual representation to check signal performance and integrity. This is achieved by repetitively sampling a digital signal from a receiver and applying it to the vertical input, while the data rate is represented at the horizontal sweep. On the other hand, in Szafraniec et al.'s [94] study, the Kalman filter has been presented as a tracking estimator for carrier phase and polarization as well as an estimator for the first-order PMD. Kalman filter is an algorithm that estimates some unknown variables given the measurements observed over time using state-space techniques. Therefore, it has been applied to a wide range of tracking and navigation problems. Additionally, Ming Chieng Tan et al. [96] could determine simultaneously OSNR, PMD, and CD, independently of the system bit-rate and modulation format, using PCA pattern recognition which is applied on asynchronous delay-tap plots. Takahito Tanimura et al. [97] focused on monitoring the channel OSNR. This was accurately achieved using an ANN, which consisted of at least 5 layers and was trained with

$400,000$ samples. Thrane et al. [100] also proposed a system that could perform OSNR monitoring and a modulation format classification for advanced modulated systems (up to 64-QAM). An ANN algorithm was used in estimating the OSNR, while a support vector machine (SVM) algorithm was used for the modulation format classification. Both algorithms were developed sequentially using features extracted from the eye diagram of the received signal, the expected OSNR, and the modulation format of the system. Nevertheless, the study's promising results for monitoring the system OSNR and classification of the modulation format ignored linear and nonlinear impairments of optical fiber and only considered the AWGN.

3.6.4 Mitigation of Fiber Nonlinearity using Artificial Intelligence

ML techniques have been recently incorporated into DSP to accurately detect the received symbol and efficiently mitigate the fiber nonlinearities. To illustrate, in the paper [42], the authors proposed a cognitive digital receiver that used clustering algorithms to recognize the receiving signal format, QPSK/8-PSK/16-QAM, without the need to receive prior pilot symbols. In addition, a combination of state-space models, Bayesian filtering, and expectation-maximization (EM) was described in [117]. These techniques aim to incorporate the channel and optical components underlying physics in the signal processing algorithms' representation. This approach results in an inclusive system improvement, including cross-polarization mitigation, carrier synchronization, and optimal symbol detection. However, this technique is not practical for dynamic optical networks because EM depends on the transmission link parameters.

Furthermore, Wang et al. [105] studied the mitigation of nonlinear phase noise (NLPN) added to M-ary phase-shift keying (M-PSK) based coherent optical system using a nonlinear SVM classifier. The algorithm's goal was to generate nonlinear decision boundaries that allow bypassing the errors introduced on the M-PSK constellation by the nonlinear impairments. This technique resulted in improvements both in the maximum transmission distance and dynamic range of the launched power. The main drawback of this technique is that it adds a significant complexity overhead because SVM is basically a binary classifier, and many SVMs would be needed to deal with higher-order modulation formats. On the other hand, Danshi Wang et al. [104] and Torres et al. [101] proposed a k-nearest neighbors-based detector as a multi-class classification that is capable of classifying multiple kinds of data simultaneously. This method was demonstrated on a 16-QAM coherent transmission system, and maximum transmission distance and nonlinear tolerance improvements were demonstrated. The computational complexity of this machine learning-based

40

demodulator is highly reduced and transparent with respect to the nonlinearity source; however, it is not practical for long haul transmission systems in which the AWGN is high, and consequently, the constellation points become more diffused with each other.

Another area of research is working on using AI techniques to reverse the propagation model effect instead of classifying the received constellations. For instance, neural networks were employed to evaluate digital back-propagation in [45] to estimate the received symbols. The number of the hidden layers in this method depends on the number of steps and number of spans of the link; therefore, it is impractical for multiple spans long transmission. Dynamic deep neural networks (DDNN) were also separately introduced by Oleg Sidelnikov et al. [89] to simplify nonlinearity mitigation in both 1 and 5 channel single-polarization systems. In this technique, they fed the neural network with a number of delayed taps of the propagating symbols, which depends on the transmission distance. Unfortunately, this technique leads to a large-sized network and requires a different neural network for each polarization, which must be retrained for each launch power.

AI is a data-driven technique where the used algorithm gains knowledge about the system from the data. Therefore, it sometimes has better performance than model-driven techniques such as IVSTF. In [53], the authors demonstrate how the artificial neural network nonlinear equalizers (ANN-NLE) outperform linear equalization and IVSTF-NLE.

In the following chapters, we introduce the use of different ML algorithms in combination with the perturbation analysis to solve the Manakov equation. This technique can achieve better performance in reconstructing the transmitted signal than the classification techniques and with lower complexity, especially for long-haul transmission systems.

3.7   Summary

This chapter covered a literature review for the different DSP techniques that have been proposed to reverse the effect of the optical fiber on the transmitted signal. Analytical and numerical techniques have been presented and their drawbacks in practical transmission. In the following chapters, we will introduce AI techniques that can achieve significant nonlinearity mitigation with practical complexity.

Chapter 4

Nonlinearity Mitigation Using Neural Network at Receiver

4.1   Introduction

This chapter presents Feedforward Neural Network FFNN as an efficient supervised ML alternative to solve coupled NLSE which describes the propagation of the optical signal through optical fiber as described in chapter 2. In our method, we combine the perturbation-based analytical solution with the FFNN to generate a near-optimal classifier for the noisy received symbols. Moreover, the FFNN can adapt the nonlinearity inherent in the activation function, as will be described in the chapter, to the degree of nonlinear amplitude and phase noise. Moreover, this activation function can in turn be implemented through a lookup table (LUT) which reduces more the computational complexity relative to analytic procedures.

4.1.1   The Communication System

In order to solve the coupled NLSE and mitigate the nonlinearity added to the propagating signal through optical fiber using FFNN, training and testing data needed to be generated. The system shown in figure 4.1 is used to generate the data. A simulated data set composed of two blocks of 2^{17} amplitude-modulated symbols of data, generated according to the parameters in Table A1, was transmitted through a fiber link with parameters close to the ITU specifications for SMF [52] as in Table A2. After the transmission, the signals are
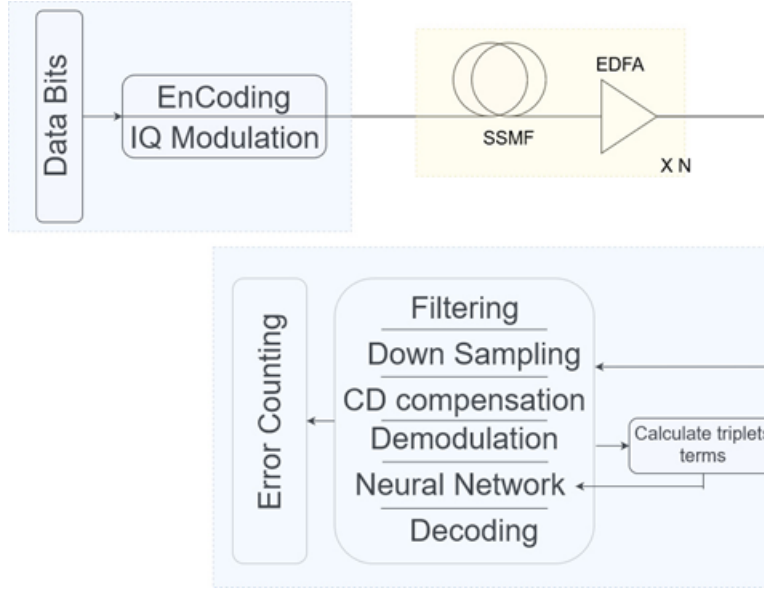
Figure 4.1: Block diagram for the data path and triplet calculations

coherently detected, demodulated, and passed through a FFNN programmed in scikit-learn that is trained to estimate the amplitude of the nonlinear distortions of the transmitted symbols. After the nonlinear distortion is mitigated by the network, the symbols are decoded to bits. The system performance is finally evaluated by comparing the received and transmitted bits, and the quality factor (Q-factor) is evaluated according to the standard formula.

$$Q = \sqrt{2}\,\mathrm{erfc}^{-1}(2\ \mathrm{BER}) \tag{4.1}$$

in which erfc^{-1} is the inverse Complementary Error Function and BER is the bit error rate.

4.2 Results

4.2.1 The Proposed Feed Forward Neural Network Design

As discussed in the literature survey, NN has been used to classify the received constellation and predict the transmitted symbols by training the NN using one or a group of delayed received symbols [16, 77]. Moreover, NNs achieved some improvement in mitigating nonlinearity by imitating DBP technique [45, 89] or just predicting the nonlinear
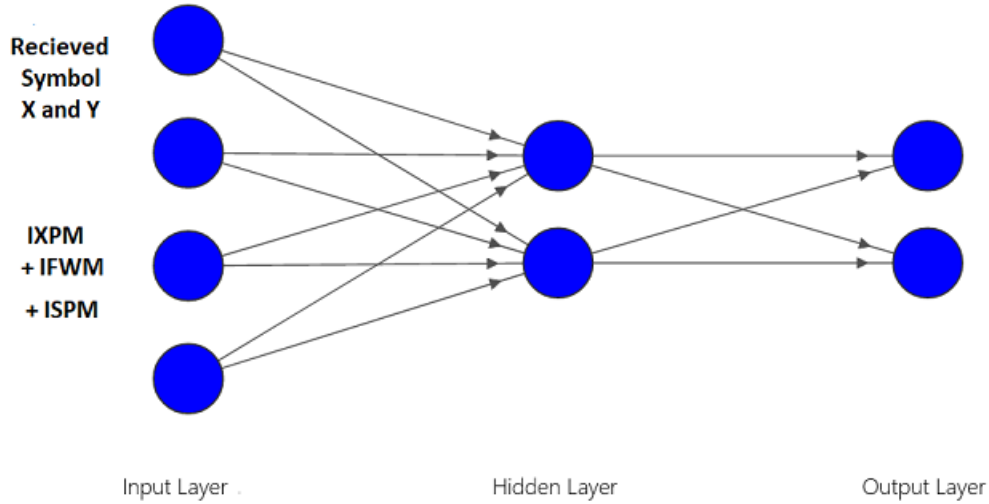
Figure 4.2: The Feed Forward Neural Network Structure

noise using DNN. Despite the improvement, the number of the hidden layers needed in both designs were impractical. As in the case of DBP, the achieved gain in Q-factor is 0.2 dB better than the conventional DBP with 2 steps per span; however, the number of the hidden layers depends on the number of steps per span and number of spans of the link. Also, for the prediction case, using dynamic deep neural networks, the inputs of the network are simply a window of the symbol of interest with a group of its neighboring symbols. Consequently, the NN input does not have enough features, and the optimum design for the NN becomes complex with a large width and more hidden layers [89]. The achieved enhancement for a link range between 1500 km and 2700 km is 0.9 dB to 1 dB, using a two hidden layers NN with 16 neurons in each layer, where each BER point is calculated by averaging the error rate over 15 signal block transmissions. This is an added complexity to the system to reduce the noise effect on the data before being used in the NN. Therefore, we propose the combination of the perturbation-based technique (see section 3.5) with the FFNN to achieve better performance with a smaller network structure to save resources. This can be achieved by adding the SPM triplets terms $(X_n X_{m+n}^* X_m, Y_n Y_{m+n}^* X_m, Y_n Y_{m+n}^* Y_m, X_n X_{m+n}^* Y_m)$, shown in equations 3.28, 3.29, to the input of the NN with the symbol of interest and the corresponding co-polarized symbol as shown in figure 4.2.

We first investigate the influence of the number of triplet terms (ISPM + IXPM + IFWM) on the Q-factor of the received data for different activation functions. The calcu-

lations are performed on about $130,000$ data symbols for each polarization at the required launched power. In the training cycle, the "Adam" algorithm for first-order gradient-based optimization of stochastic objective functions [60] is employed to reach the minimum MSE. In addition, the data symbols are launched with a power of 2 dB above the optimum launch power to increase the amount of nonlinearity captured by the NN. The network is trained with $80,000$ (60%) of the symbols and is then tested with the remaining symbols to ensure that the inputs are not overfitted. For other launch powers, rather than retraining the NN, a scaling factor is applied to the previously trained NN outputs. The transmitted symbol is then predicted by subtracting the scaled perturbations from the received symbol according to,

$$Tx_{symbol} = Rx_{symbol} - \alpha * (\text{NN})_{output} \tag{4.2}$$

Here Tx_{symbol},Rx_{symbol} are the predicted transmitted and received symbols, α is the power scaling factor, and $(\text{NN})_{output}$ signifies the output perturbations from the NN. To enhance the NN performance, the input and the target data are transformed by subtracting their mean and scaling their variance to unity before entering the NN. In the NN inputs, the optimum number of triplet terms, with corresponding magnitudes of $20\log(C_{mn}/C_{00})$ is larger than a certain threshold, are employed, an example for the interaction coefficient C_{mn} between the symbols m and n [98] for a -22 dB threshold is shown in figure 4.3.

Moreover, to determine the optimum activation functions, we display in figure 4.4 the behavior of Relu, sigmoid, and tanh activation functions for the hidden layer neurons. Figure 4.5 additionally demonstrates that, as the number of triplets terms increases, the Q-factor of the system increases up to a certain limit, after which the performance instead deteriorates as a result of overfitting. Further, it indicates that the Relu activation function yields superior results, reaching a 1 dB Q-factor enhancement for the optimum number of triplets, namely 2445. In comparison, the corresponding numbers for the sigmoid function are 0.7 dB and 1193 triplets. The figure compares the results to a system subject only to CDC. This behavior was expected, since using the proposed design for the NN with Relu activation function yields a function that combines the target functions of the used neurons as in equation 2.35. This resulting function is very close to equations 3.28, and 3.29 which represent the nonlinear distortion using the perturbation based nonlinear compensation technique. Also, Relu activation functions consider more system nonlinearity than simple linear activation functions would.

After determining the optimum activation function and triplet number, we consider the dependence of the performance on the number of input symbol window terms (input taps) centered around the symbol of interest. Figure 4.6 shows that adding more terms than the
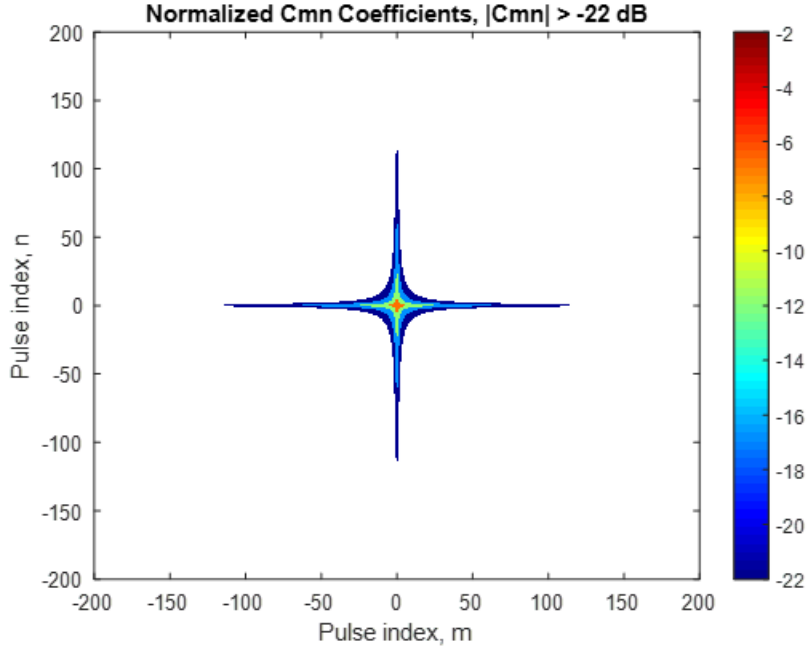
Figure 4.3: The triplets employed for different symbols around the symbol of interest

symbol of interest (on which the perturbation terms are based) to the input of the NN does not improve the performance or noticeably alter the Q-factor. However, removing terms results in performance degradation.

4.3   Discussion

The proposed NN design reduces the computational complexity compared to previous designs firstly by replacing the complex multiplications in 3.28, 3.29 with real multiplications and LUT, and secondly by employing only a single 2 node hidden layers. As a result, the number of real multiplications, N_{mult} needed to compute the NN output in the execution stage is given by,

$$
\begin{aligned}
N_{mult} &= N_{inputs} * N_{hiddenlayernodes} + N_{hiddenlayernodes} * N_{nodes(output)} + 2 \\
&= N_{inputs} * 2 + 2 * 2 + 2 = 2 * N_{inputs} + 6
\end{aligned}
\tag{4.3}
$$

where the additive constant 2 at the end of the first expression is associated with the scaling factor. Hence, the computational overhead varies only with the number of inputs,
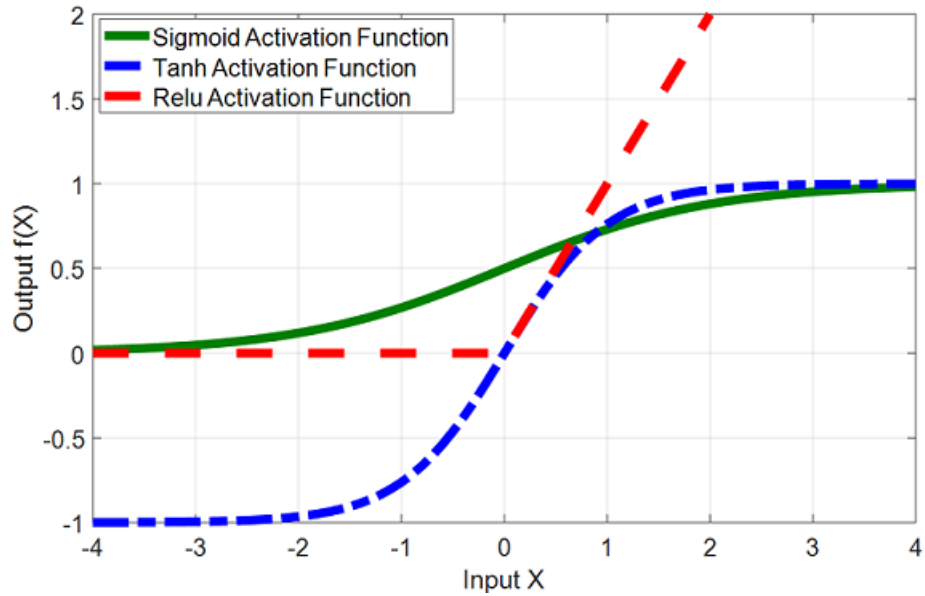
Figure 4.4: The Three activation functions examined in the NN design
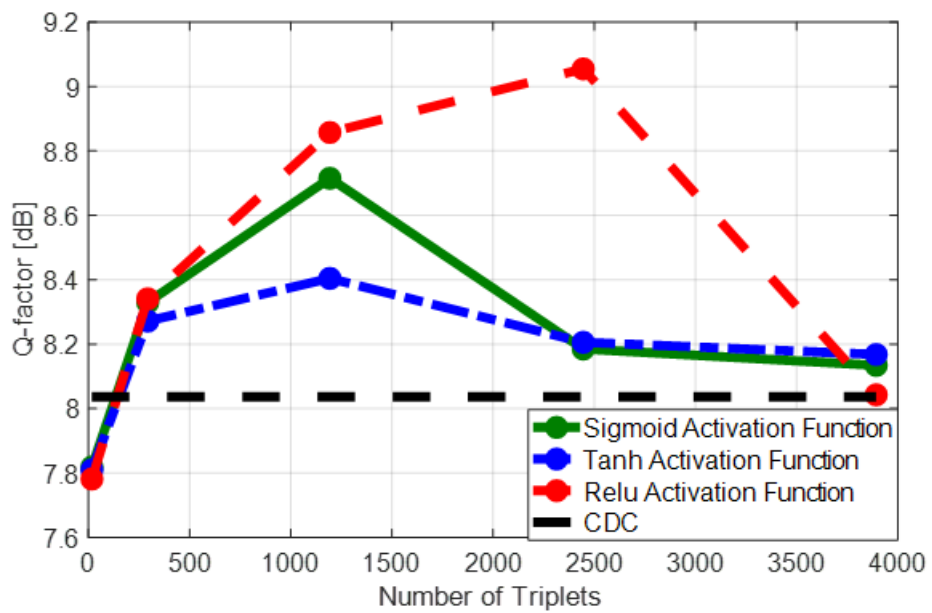
Figure 4.5: Q-factor as a function of the number of triplets for different activation functions

47
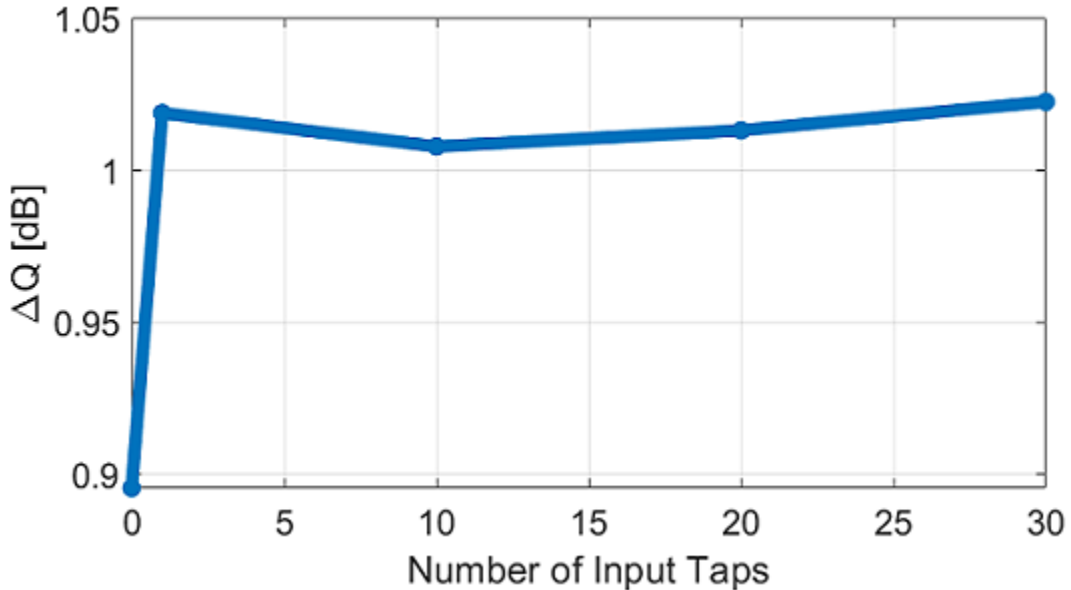
Figure 4.6: The variation of the Q factor enhancement with the number of symbols in the input window

N_{inputs}, which can be decreased by selecting only the most significant NN inputs.

One strategy is to perform an initial training cycle (offline) with all inputs present, as discussed earlier, yielding weights between each input node and the hidden node in the first layer. The inputs corresponding to weights below a certain threshold can then be eliminated in a process called weight trimming. Figures 4.5 and 4.6 demonstrate the effect of changing the threshold in dB on the performance of the NN after the appropriate number of inputs entering the NN are eliminated at each threshold value. At a threshold of -32 dB, the number of inputs is reduced by 1475 or, 30% while the Q-factor decreases by about 0.14 dB. This loss is insignificant in view of the greater ease of implementation of the simplified design.

Additionally, the dimensionality of the input data and hence the number of NN nonlinear compensator (NN-NLC) inputs can be reduced through PCA [8] as shown schematically in figure 4.9. Indeed, figure 4.10 illustrates the difference in the system performance at different launch power without any nonlinear compensation (CDC) and with the two cascaded dimensionality reduction methods (weight trimming and PCA) followed by a NN. The figure shows that the optimum launch power for the CDC system is at -1 dBm, where the system performance is primarily affected by linear noise and nonlinear noise below and
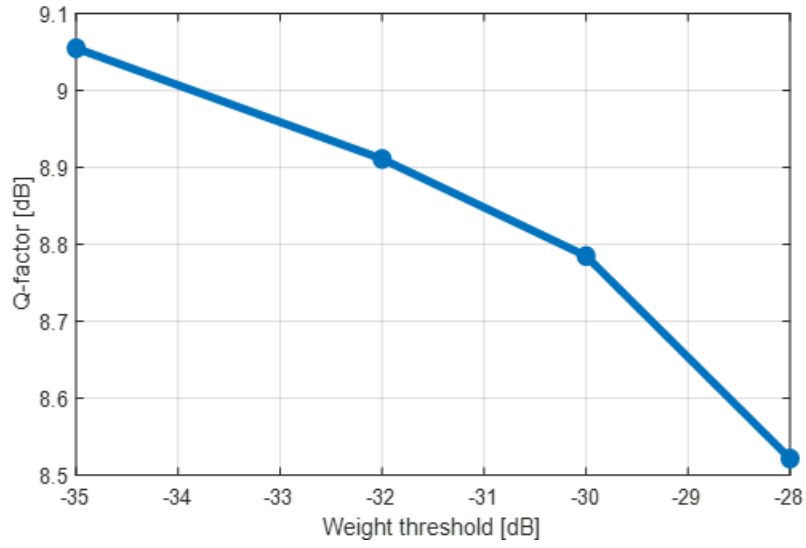
Figure 4.7: The Q factor as a function of the weight trimming threshold
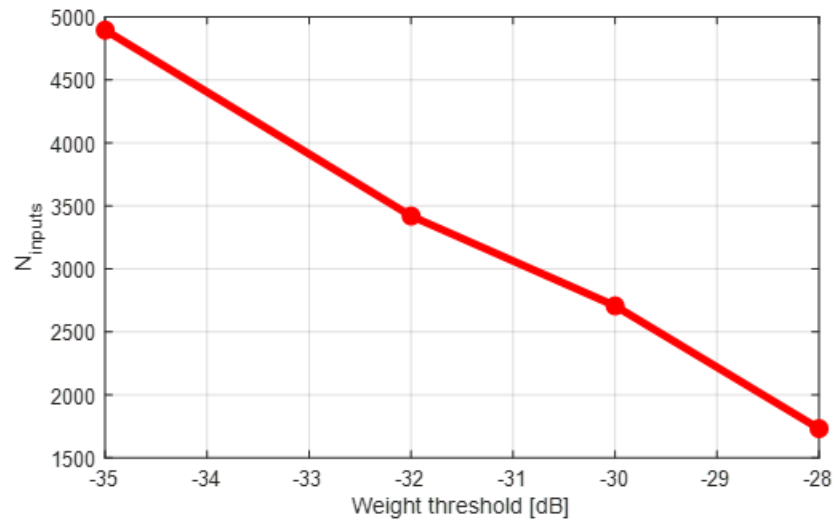
Figure 4.8: The number of inputs corresponding to each weight threshold

49

Figure 4.9: Block diagram for the complexity reduction technique
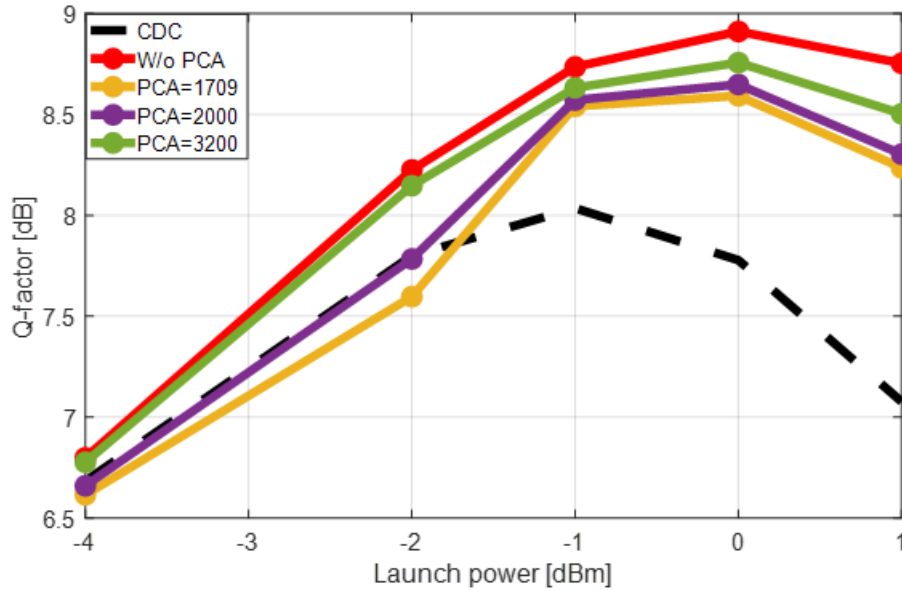
Figure 4.10: The Q factor as a function of the launch power at the receiver and transmitter

above this threshold power value, respectively. The figure also shows that the PCA can effectively decrease the number of inputs to 3200 (35% of the inputs are removed) without significantly altering the output performance. Eliminating further inputs, however, substantially affects the Q-factor such that for a PCA with 1709 inputs (50% of the initial reference input features), the enhancement decreases from 1 dB to 0.5 dB.

4.4   Summary

This chapter presented an optimum FFNN design which is a promising AI technique to solve the NLSE and predict the nonlinear perturbations add on the transmitted symbols through single-mode optical fibers. We have also presented the effect of adding the SPM

triplets terms on the Q-factor of the system. Moreover, since the purpose of the study is to reduce the complexity overhead, we have proposed weight trimming and PCA as two cascaded methods for further complexity reduction to the system.

Chapter 5

Two-Stage Nonlinearity Mitigation

5.1 Introduction

In this chapter, we propose a novel technique that employs AI to mitigate optical fiber nonlinearity. In this method, we use the optimum FFNN design proposed in the previous chapter and investigate the effect of moving the trained FFNN from the receiver to the transmitter side. The main advantage of changing the position of the FFNN to the transmitter side is that another equalizer can be employed at the receiver. Later in this chapter, we investigate the effect of adding classifiers of different complexity at the receiver side such as decision trees, boosting techniques, random forests, extra trees, and multi-layer perceptron classifiers.

5.1.1 The Communication System

The same system that was described in section 4.1.1 is used to generate the data; however, as shown in figure 5.1 after the data encoding, a shift in the data symbols by the negative non-linear noise imposed during propagation. This noise is predicted by the NN placed at the transmitter. After propagation, the signals are coherently detected, demodulated, and fed into the classifier which assigns the received symbol to one of the 16-QAM classes. Finally, the symbols are decoded, and the Q-factor is evaluated using equation 4.1.
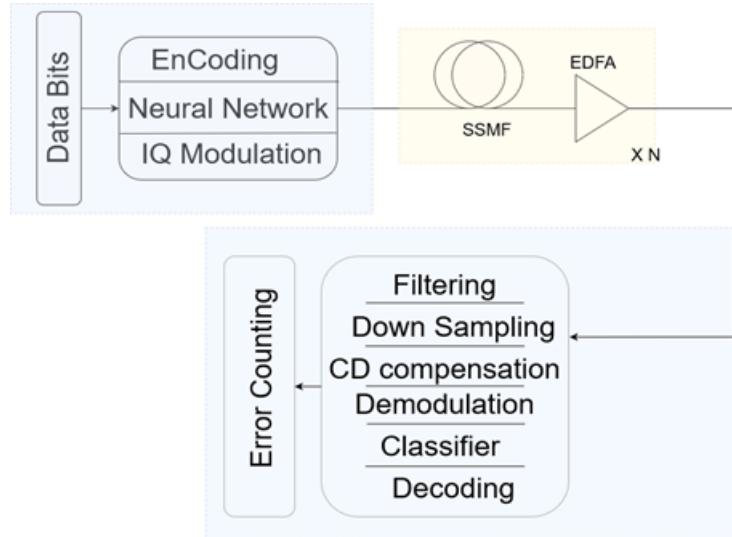
Figure 5.1: Block diagram for the data path

5.2 Results

5.2.1 Neural Network Pre-compensation

As presented in the previous chapter, the optimum NN design consists of 1 hidden layer with 2 neurons and each neuron has "Relu" activation function. The inputs are the symbol of interest, the co-polarization symbol at the same time slot, and the triplet terms with magnitudes larger than a threshold, $20 \log(C_{mn}/C_{00})$, equal -22 dB. The NN training cycle is employed at the receiver side on 80,000 received data symbols, of one polarization, at 2 dB above the optimum launch power [23]. This trains the NN on symbols with pronounced nonlinear noise features. During the execution stage, this trained NN can be used either on the receiver side for the post compensation of the received data as in the last chapter, or the transmitted side where the input to the trained NN is the transmitted data. In this case, the transmitted symbols are input into the NN, while the shift in the symbol position from its optimal position is employed as the output variable. To pre-compensate the signal before passing through the link, we subtract this shift from the symbol position before modulation. This procedure is applied on both polarizations with the same trained NN. Figure 5.2 displays the Q-factor with an NN employed as a nonlinear compensator (NLC) at the receiver and transmitter side for different launch powers compared to the system subject only to CDC. While the differences in the two sets of NN results are negligible,

53
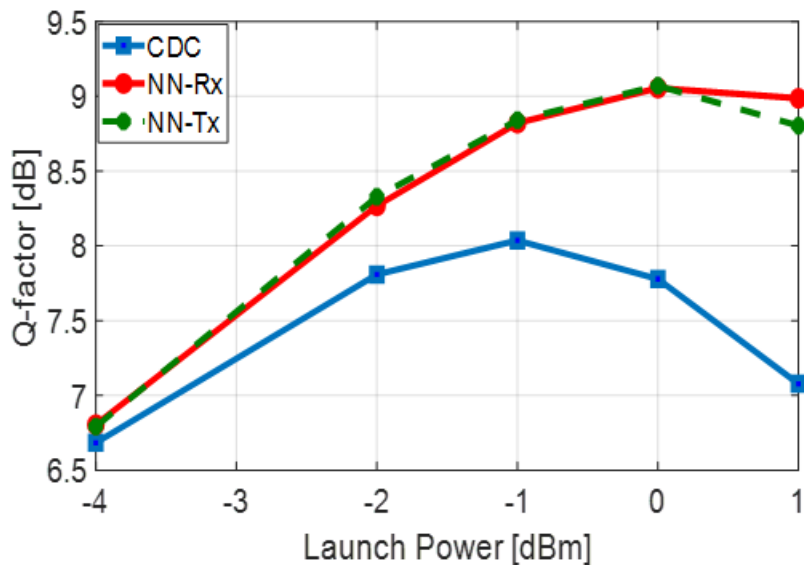
Figure 5.2: The Q factor as a function of the launch power at the receiver and transmitter

positioning a NN at the transmitter side enables the addition of a second classification stage at the receiver. This can improve performance without data overfitting.

5.2.2 The Receiver Classifier

Classification predictive modeling is the process of approximating a mapping function (f) from the input variables (x) to a set of discrete output variables (y). In the present context, the input variables consist of the real and the imaginary parts of the symbol of interest, and the co-polarized symbol propagating at the same time slot. The complex output symbols are then mapped to classes with labels 1 to 16 in the training stage of the classifier.

Since the data is already pre-compensated by a NN before propagation, a strong classifier that can process low amplitude nonlinear noise is required. The training employs around 80,000 data symbols launched at the optimum power. A NN classifier with an optimum dimension of one hidden layer with 4 neurons, each with a 'Relu' activation function is employed at the receiver side in figure 5.3 [23], which indicates that the classifier delivers a slight 0.03 dB Q-factor enhancement compared to a NN at only the transmitter. The smallest Q-factor enhancement is obtained from a Decision Tree classifier indicating that the NN and Decision Tree algorithms cannot adjust to the noise properties of the train-
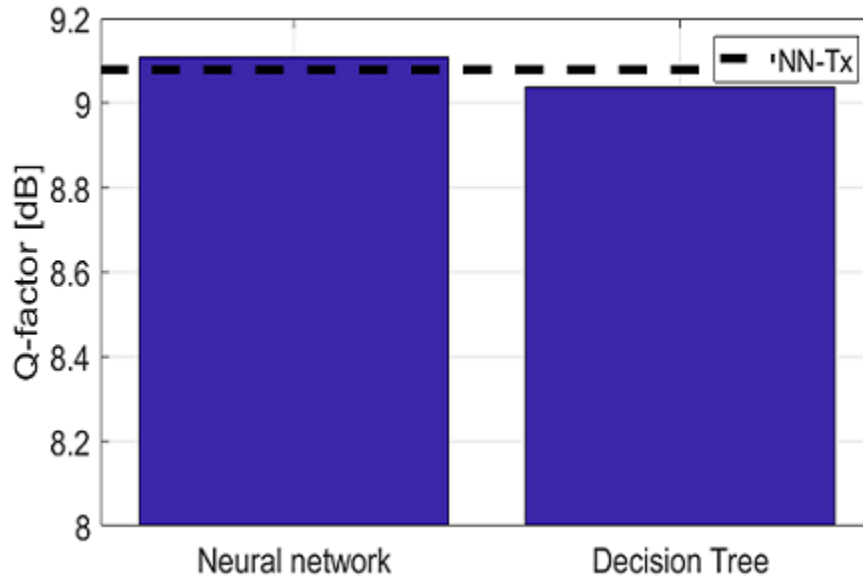
Figure 5.3: The Q-factor for different receiver classifiers

ing data. The classifier performance can however be improved with ensemble methods. While this requires implementing several models simultaneously, a combination of ensemble models and decision/regression trees is still computationally advantageous because of the efficiency of the standard tree-growing algorithm [38].

Boosting

Boosting is an ensemble technique that implements classifiers (learners) sequentially, such that subsequent classifiers learn from the errors of preceding classifiers. Adaptive boosting and gradient boosting [10] are considered below.

Adaptive Boosting

Adaptive boosting (AdaBoosting) supplies the initial weak classifier with a training set for which each input variable is equally likely to be selected [34]. The classifier performance is evaluated by computing the classification error rate ϵ_t, defined as the ratio of incorrect to correct classifications. This yields the weight updating parameter, β_t,

$$\beta_t = \epsilon_t/(1 - \epsilon_t), \tag{5.1}$$
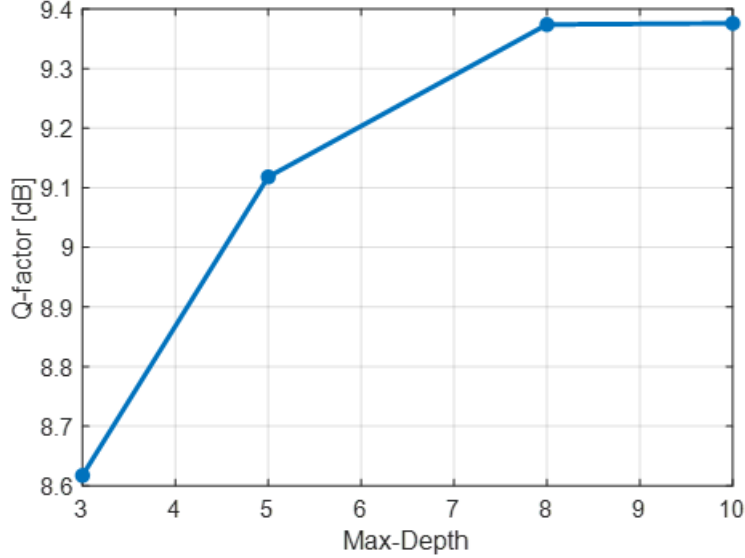
Figure 5.4: The Q-factor as a function of the tree maximum depth in AdaBoosting

which constitutes a measure of confidence in the classifier with a small β_t corresponding to high confidence. The probability distribution of the input variables selected for the next classifier is then computed by multiplying the weight of each input variable by β_t, which reduces the weight when the input is correctly identified by the previous classifier. After again normalizing the sum of the weights to unity, the process is repeated until either the preset number of classifiers are implemented or $\epsilon_t < 0.5$. The final output is then

$$F(x) = argmax \sum_{t:h_t(x)=y} W_t,$$

$$W_t = \log(1/\beta_t)$$

(5.2)

in which W_t are the weights used for the classifier's output, t is the classifier number, x is the input variable and y is the observed output [87]. Here the boosting algorithm is applied to the decision tree classifier for different maximum depths, where the depth is defined as the length of the longest path from a root to a leaf, as shown in figure 5.4. As the trees expand, the Q-factor is seen first to increase and then to level off. At a depth of 8, a 0.3 dB Q-factor enhancement can be achieved relative to the Q-factor of a system with only a NN at the transmitter. However, boosting the support vector classifier is not as effective as boosting the decision tree, as in figure 5.6.

Gradient Boosting

Gradient boosting learns a boosting classifier incrementally. The mapping function $f()$ is expressed as a linear combination of weak classifiers h as,

$$f(x) = \sum_{j=1}^{t} (\alpha_j h_j(x; \theta_j)), \qquad (5.3)$$

where α_j is a real-valued weight and $f(.)$ is constructed in a greedy manner by iteratively requiring the parameters θ_j and weight α_j of a weak classifier to minimize an augmented loss function given by [91],

$$\text{Loss} = \sum_{i=1}^{N} l(y_i, f(x_i)) = \sum_{i=1}^{N} \exp(-y_i f(x_i)), \qquad (5.4)$$

In figure 5.5 the performance of gradient boosting for the present example is seen to depend on the maximum depth of the decision tree classifiers with an optimum depth of 3. Figure 5.6 further indicates that gradient boosting is less effective than adaptive boosting with a Q-factor difference of 0.2 dB relative to a single NN at the transmitter, while the additional required computational time is observed to be substantial from Table 5.1.

Random Forest

The random forest ensemble method employs tree classifiers running in parallel, such that each tree depends on the values of an independently sampled random vector with the same distribution for all trees in the forest. In particular, a margin function, defined as,

$$mg(x, y) = av_t \, \mathrm{I}(h_t(x) = y) - max_{j \neq y} av_t \, \mathrm{I}(h_t(x) = j), \qquad (5.5)$$

where x is the input variable and y is the observed output, The indicator function $\mathrm{I}(.)$ quantifies the degree to which the average number of votes at x, y for the correct class exceeds the average vote for any other class. The larger the margin, the higher the confidence in the classification. Figure 5.6 demonstrates that the random forest performance is comparable to that of Ada-boosting applied to the decision tree classifier at the maximum depth of 8. However, the Ada-boosting training phase is nearly three times longer than that of the random forest, c.f. Table 5.1.
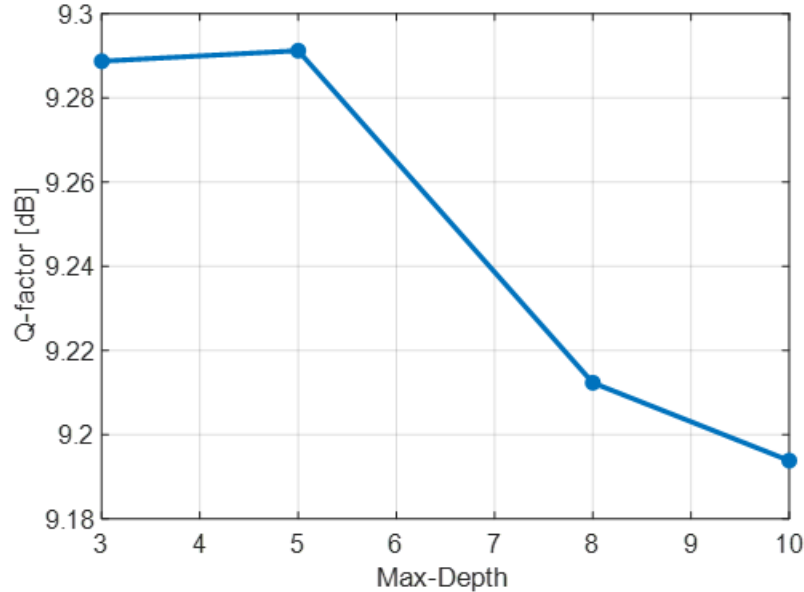
Figure 5.5: The Q-factor as a function of the tree maximum depth in gradient boosting

Extremely Randomized Trees

An extremely randomized, extra trees classifier constructs the decision tree in a more efficient fashion than the random forest classifier. As well, it typically exhibits improved performance for noisy data. Here, an ensemble of the un-pruned decision or regression trees is generated with a standard top-down procedure. Unlike the two previous tree-based ensemble methods, however, nodes are split at random cut-points and the entire learning sample data is employed to grow the trees. Each decision tree in the forest of the extra trees is thus built from the original training sample. At each test node, each tree is then provided with a random sample of features from which the decision tree selects the optimal feature to split the data based on specific mathematical criteria, typically the Gini index,

$$Gini_{index} = 1 - \sum_{1}^{C} (pi)^2, \tag{5.6}$$

where p_i is the probability of each class, and C is the total number of classes. Since features are sampled randomly, multiple de-correlated decision trees are generated [38]. Figure 5.6 demonstrates that the extra trees method performs similarly to the random forest and
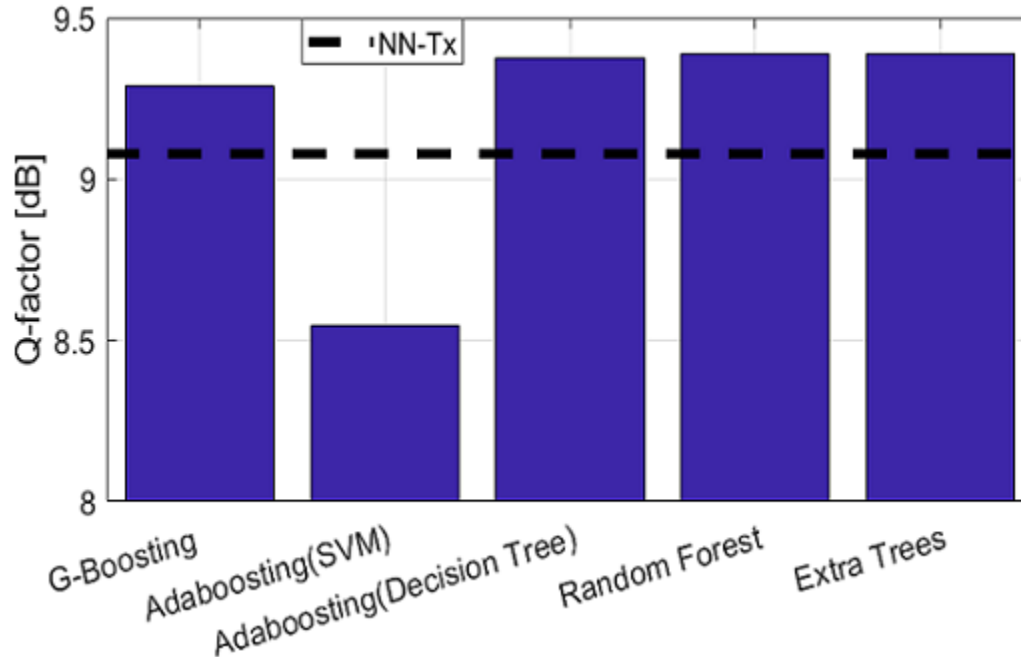
Figure 5.6: The Q-factor for different receiver classifiers employing ensemble methods

decision tree Ada-boosting. However, the required computational training is 47% less than the random forest method and 81% less than AdaBoosting as indicated in Table 5.1.

Classifier	Gradient Boosting	AdaBoosting	Random Forest	Extra Trees
Time (sec)	527	190.4	67.9	35.7

Table 5.1: ALGORITHMIC COMPUTATION TIME

5.3   Summary

In this chapter, we have introduced a new method for mitigating the optical fiber nonlinearity and enhance the performance of propagation through SMF. This method is based on having pre-and post-compensation for the nonlinear noise. The pre-compensation is at the transmitter side using regression FFNN in combination with SPM perturbation triplets terms. While the post-compensation stage is employed using a classifier at the receiver.

The study shows that the extra trees classifier has the optimum performance with an overall 1.3 dB improvement in the Q-factor and least computational time which is 47% less than the random forest method and 81% less than AdaBoosting.

Chapter 6

Customized Siamese Neural Network

6.1 Introduction

In this chapter, we apply the Siamese Neural Network SNN to fiber nonlinearity mitigation. SNNs form a class of artificial neural networks (ANNs) that have been successfully adapted to a wide variety of problems [107, 18, 14, 66, 99] including object tracking, face recognition, and image similarity detection [11, 20, 95]. A SNN contains two identical ANN branches, each of which is applied to a different data point. The two branches share weights and bias values that are updated during training. A subsequent network layer then maps the input pairs to latent variables that evaluate the similarity of the pair.

Our SNN implementation mitigates the nonlinear perturbations of the propagating signal described by equations 3.18, 3.19. In particular, the symbols of interest in both polarizations are the input into one of the two input ANN branches while the most significant SPM noise terms such as the ISPM, IXPM, and IFWM terms ($X_n X^*_{m+n} X_m$, $Y_n Y^*_{m+n} X_m, Y_n Y^*_{m+n} Y_m, X_n X^*_{m+n} Y_m$) are simultaneously input into the second ANN input branch. The two ANNs are then merged into a single ANN, or algebraic "averaging" layer, which evaluates the real and imaginary values of the perturbations added to the signal. An alternative design that employs three ANNs as input branches, one for the data symbols and the other two for the triplets (ISPM, IXPM, IFWM), is also analyzed below.

6.1.1 The Communication System

As shown in figure 6.1, the system transmitter and propagation are the same as in chapter 4. While at the receiver, the propagated signal is coherently detected, demodulated, and
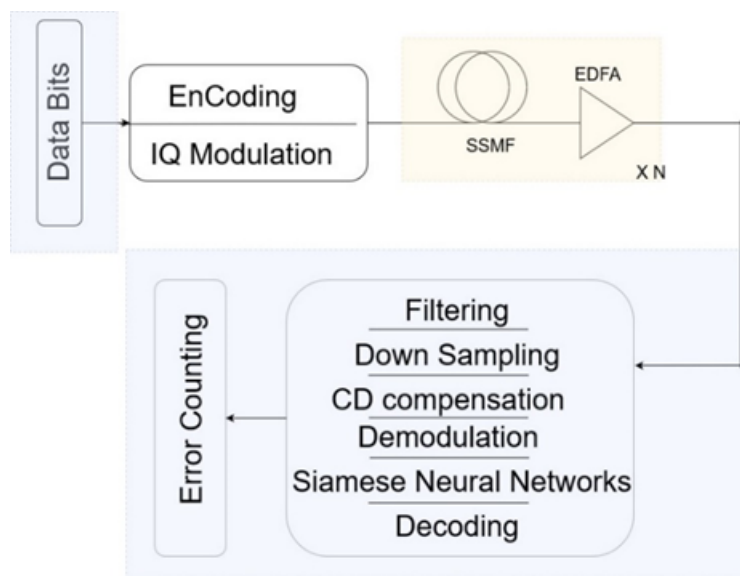
Figure 6.1: Siamese neural network based transmitter and receiver

passed through SNN code written in Keras/TensorFlow that is trained to estimate the amplitude of the nonlinear distortions of the transmitted symbols. After determining the amplitude of the distortions, the nonlinear distortion is mitigated, and the symbols are decoded. Comparing the received and transmitted bits yields the system performance given by the standard formula for the Q-factor shown in equation 4.1.

6.2 Results

6.2.1 The Siamese Neural Network Design

To investigate the performance of the two SNN configurations mentioned above, 130,000 data symbols for each polarization, an "Adam" optimizer, and MSE loss function were employed. The network was trained on the first 80,000 data symbols for a single polarization launched at 2 dB above the optimum launch power to increase the nonlinear distortion of the signal. The remaining $50,000$ symbols are used for validation which also suppressed overfitting. The trained network was further applied to orthogonally polarized training symbols and to symbols at other signal powers. To adapt the trained model to different launch powers, the output of the previously trained SNN is simply multiplied by a scaling

factor. The predicted transmitted symbols are then obtained by subtracting the scaled perturbations from the received symbol according to,

$$Tx_{symbol} = Rx_{symbol} - \alpha * (\text{SNN})_{output} \tag{6.1}$$

In Equation 6.1, Tx_{symbol} and Rx_{symbol} denote the predicted transmitted and actual received symbols, α is the power scaling factor, and $(\text{SNN})_{output}$ represents the SNN predicted nonlinear noise.

Two branches SNN

The first design employs the SNN architecture of figure 6.2 in which the symbol of interest and its corresponding orthogonal polarization symbol are input into one branch of the SNN while all triplets with $\mathbf{C_{mn}}$ that fulfill the requirement $20 \log(C_{mn}/C_{00}) > -22$ dB are input into the second branch. The two branches are subsequently connected to either a third NN or to an algebraic layer terminated by an output layer of two neurons associated with the real and imaginary parts of the noise perturbations acquired by the signal during propagation through the fiber transmission medium. If the two input branches are connected to a third NN, the dependence of the Q-factor on the first, second, and third NN dimensions is presented in figure 6.4. Evidently, changing the number of neurons in the first layer significantly affects the system's performance. The maximum achievable Q-factor is 8.9 dB, compared to 8 dB if only CDC is present. This occurs when each of the first, second, and third NNs possess 1 hidden layer of 2 neurons, a "RELU" activation function is employed for the neural network layers and the output neurons implement a "linear" activation function. An example of the neural network structure used to implement SNN is given in figure 6.3 below. If an "averaging" algebraic layer, which averages its input elements, is instead employed at the output the predicted system output becomes

$$Output = (Output_{NN1} + Output_{NN2})/2 \tag{6.2}$$

In which $Output_{NN1}$ and $Output_{NN2}$ are the output of the first and second NN, respectively, as indicated in figure 6.2. In this case, the Q-factor enhancements of figure 6.5 are obtained when the sizes of the first and second NNs are varied. The optimal performance, which is 0.04 dB less than that obtained above using a NN as an output layer, was achieved when the two branches implemented a single hidden layer with 2 neurons.

Three branches SNN

To determine if reducing the number of inputs entering the second branch in design A improves the SNN performance, the three-branch SNN architecture displayed in figure
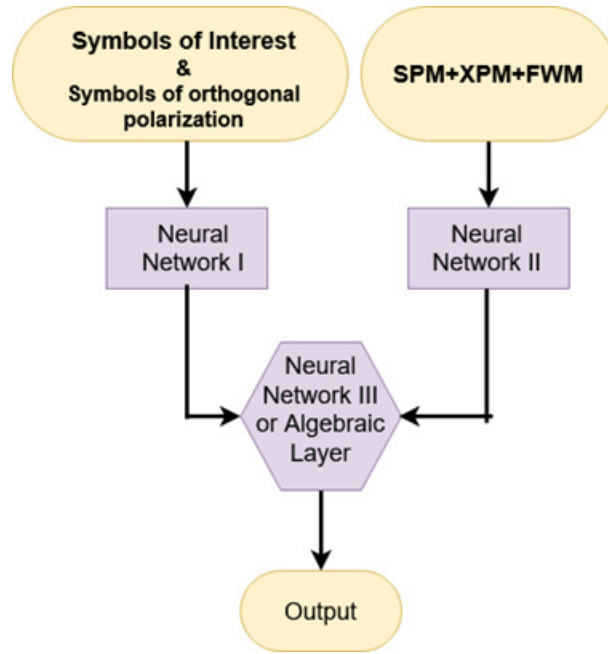
Figure 6.2: The Siamese NN "Design A" structure applied to each polarization
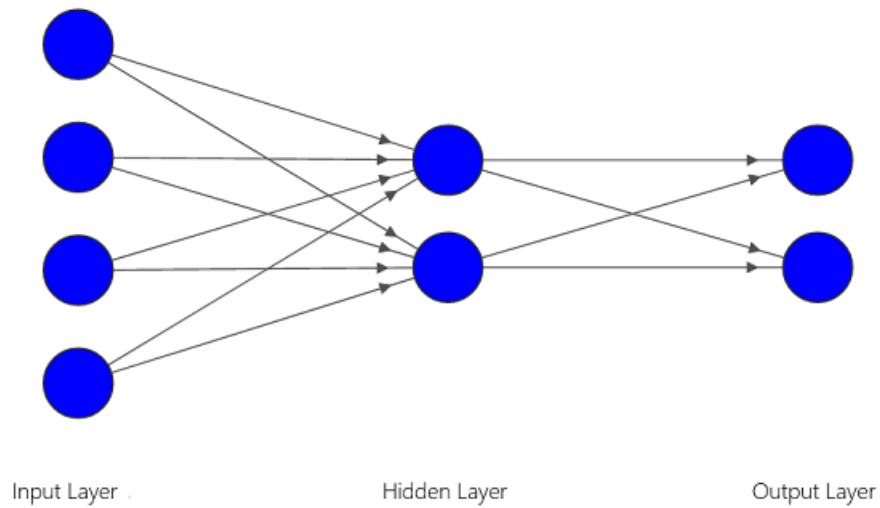
Figure 6.3: The architecture of ANN that constitutes each SNN branch
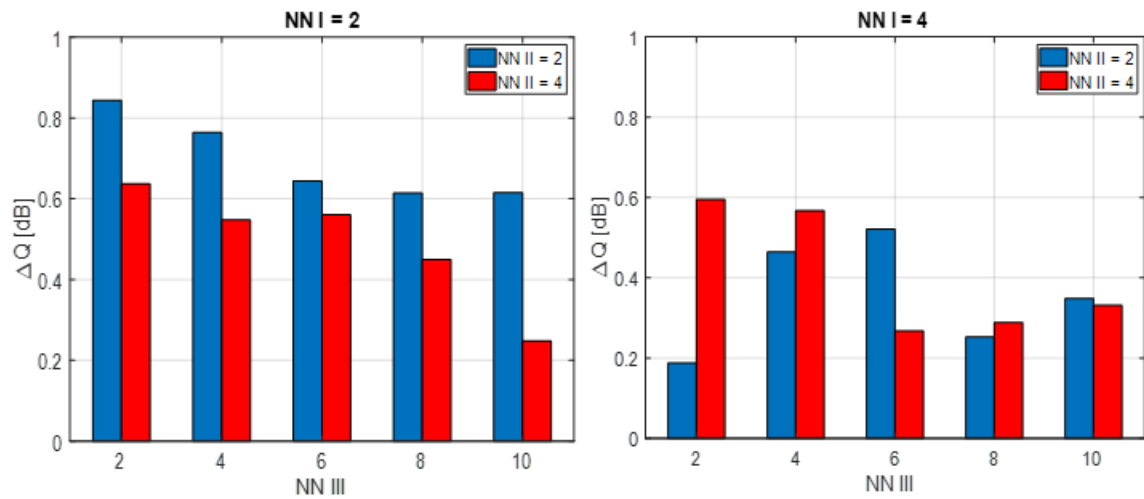
Figure 6.4: The dependence of the Q-factor on the three NNs sizes
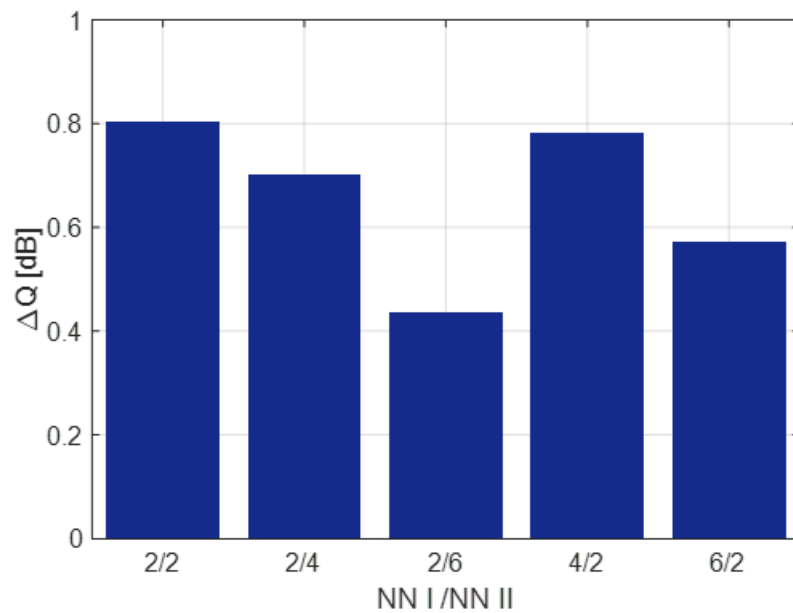
Figure 6.5: The Q-factor enhancement for different first and second NN sizes

6.6 was implemented. The symbols of interest in the first SNN branch, as well as the two groups of triplets in the second and third branches, each possess separate weights and biases that are simultaneously updated during training. However, the distribution of the inputs among the three branches is a further metaparameter. In our proposal, the inputs and the dimension of the first branch are identical to that of the previous two branch designs while the relative number of triplets employed in each of the other two branches are optimized for each pair of dimensions of the second and third NN. The triplets that are inserted into each of the two branches are ordered by $(m = -N_m/2, \ldots, 0, 1, 2, \ldots, N_m/2), n = (-N_m/2, \ldots, 0, 1, 2, \ldots, N_m/2)$ together with the constraint that, $20 \log(C_{mn}/C_{00}) > -22$ dB as shown in figure 4.3. Additionally, since the SNN inputs must be real, the real and imaginary parts of the triplets are input separately, yielding 4890 distinct values. For input branches consisting of a single 2 neuron hidden layer with a "Relu" activation function with outputs that are processed by a second two neuron hidden layer with a "linear" activation function, figure 6.7 demonstrates that the best result is achieved when 1496 triplets are input into the second branch while the remaining 3394 triplets are inserted into the third branch. The figure also establishes that the optimum ratio between the number of coefficients in each of the two branches is dependent on the sizes of the hidden layers. However, the 0.85 dB improvement in the Q-factor obtained with the three-branch architecture is nearly identical to that of the less resource-intensive two branch SNN (design A) as shown in figure 6.8.

6.2.2 Computational Efficiency

The efficiency of the proposed technique relative to the DBP can be estimated from the required number of real multiplications. For the DBP, this is given in terms of the number of propagation steps $N_{span}N_{stpsp}$ and the FFT size, N, by

$$C_{DBP} = 4N_{span}N_{stpsp}\Big[\frac{(N(log_2 N + 1)n_s)}{(N - N_D + 1)\log_2 M}\Big] + n_s, \tag{6.3}$$

$$N_D = \frac{(n_s\tau_D)}{T} \tag{6.4}$$

Here N_{span} denotes the number of spans, N_{stpsp} the number of steps per span, n_s the over-sampling factor, M the order of modulation format, τ_D the dispersive channel impulse response and T the symbol duration [89, 33]. Evidently, from equation 6.3, the computational time is of order $O(\log_2 N)$ times the number of longitudinal steps and that is therefore large for long fiber lengths. In contrast, the number of multiplications in the SNN
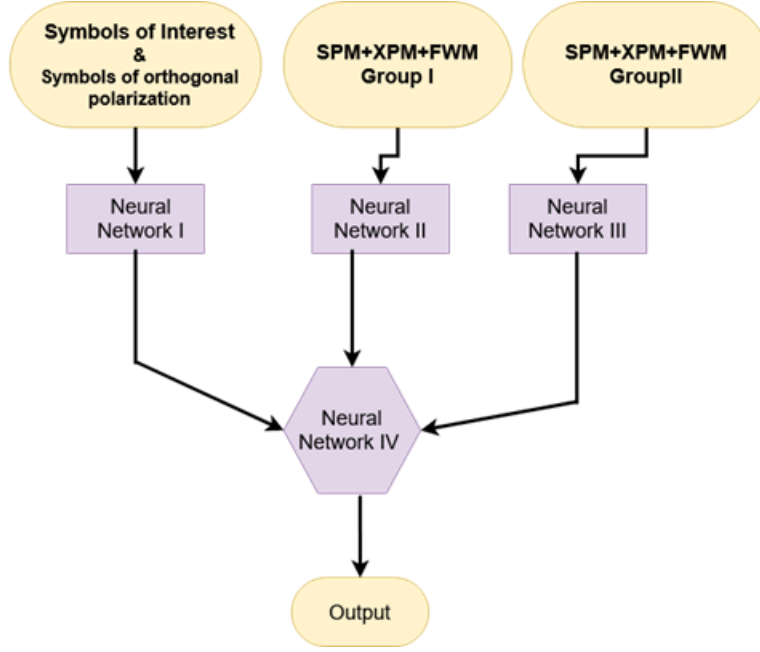
Figure 6.6: The Siamese NN "Design B"

depends on the number of inputs to each NN branch. Since the SNN is typically trained offline, the effective computation time can be approximated by that associated with the application of trained ANN input branches with a single hidden layer, which is from the discussion of [89]

$$C_{SNN}^{predict} = \frac{N_{ps}(\sum_{i=1}^{i=K} N_{inputs_i} N_{nodes_i} + (\sum_{i=1}^{K} N_{nodes_i})N_{outNN_1} + \sum_{m=1}^{S} N_{outNN_m} N_{outNN_{m+1}})}{(N_{ps} + N_{ts} + N_{vs})\log_2 M)}$$

(6.5)

in which N_{inputs_i} and N_{nodes_i} denote the number of inputs and the number of neurons in input branch I, respectively, NN_{outNN_m} is the number of neurons in the m:th hidden layer of the output NN, and K and S are the number of branches of the SNN and the output layers, respectively. N_{ps}, N_{ts}, and N_{vs} are the number of transmitted, training and validation symbols, respectively. Accordingly, the computational requirements of the SNN are proportional only to the number of inputs and are therefore less than those of the DBP for long-haul optical fiber systems.
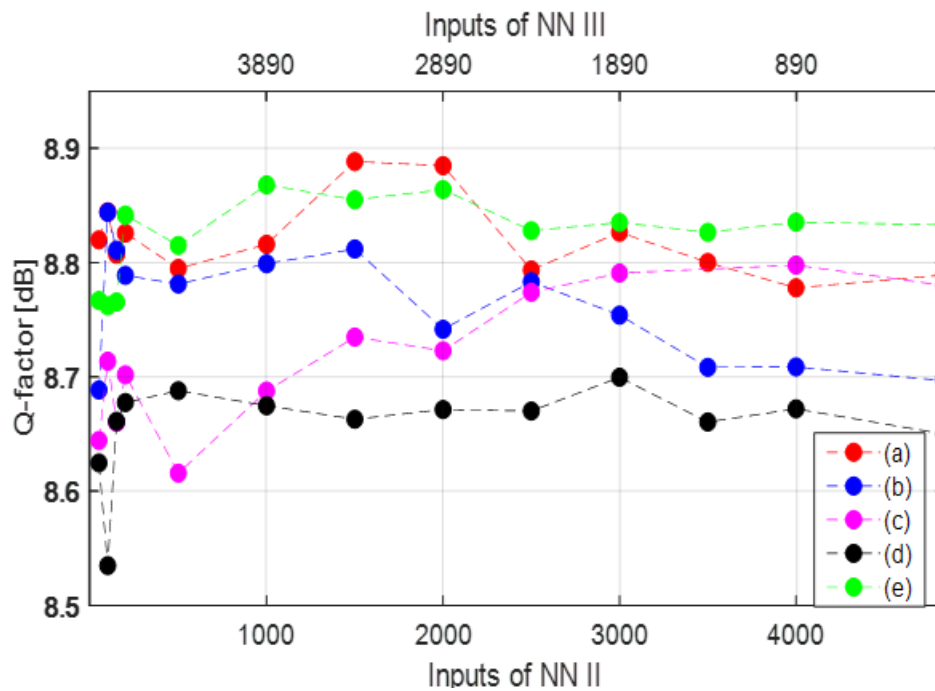
Figure 6.7: The dependence of the Q-factor on the number of inputs to the second and third SNN branches for a size of the NN I, NNII, NN III, and output NN are(a) 2,2,2,2 (b) 2,4,2,2 (c) 2,2,4,2 (d) 2,4,4,2 (e)2,2,2,4
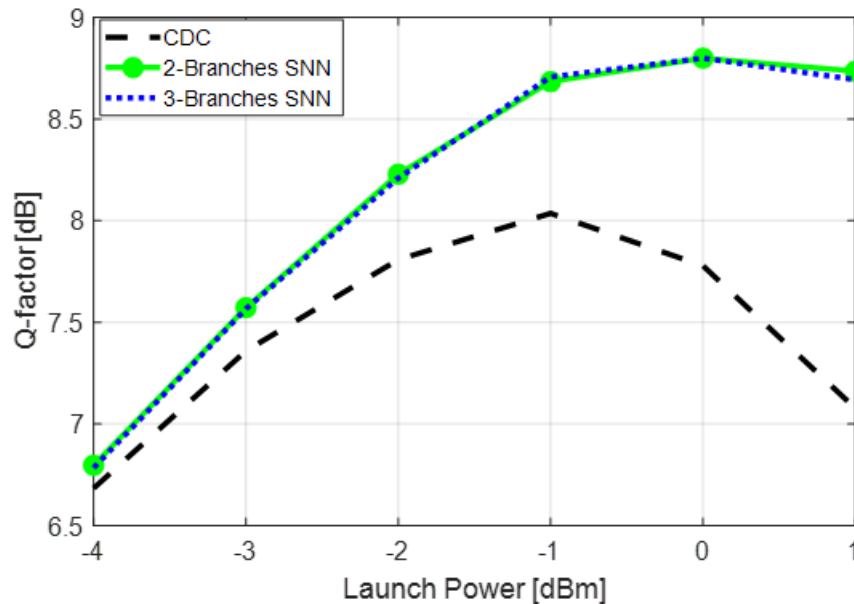
Figure 6.8: The optimum performance for the different launch powers

6.2.3 Principal Component Analysis Pre-processing

To reduce the complexity of the SNN, its inputs can be preprocessed by a PCA stage before the SNN. Since the computational requirements of the SNN are proportional to the number of quantities input into each NN branch, the PCA not only reduces complexity by decreasing the number of inputs but can also order the inputs according to their importance within each NN or set of branches [8]. The weights assigned to the inputs in each branch then ideally become comparable, such that the inputs within each branch similarly affect the output of the calculation. Indeed, figure 6.9 demonstrates that incorporating the PCA into the two-branch design of the previous section reduces the computational complexity such that a 0.76 dB enhancement is achieved with half of the original number of inputs.

6.3 Summary

In this chapter, we have shown that since SNNs do not incorporate operators such as FFT that require an extensive number of algebraic operations, they can potentially mitigate fiber nonlinearity with less computational overhead than corresponding procedures based on
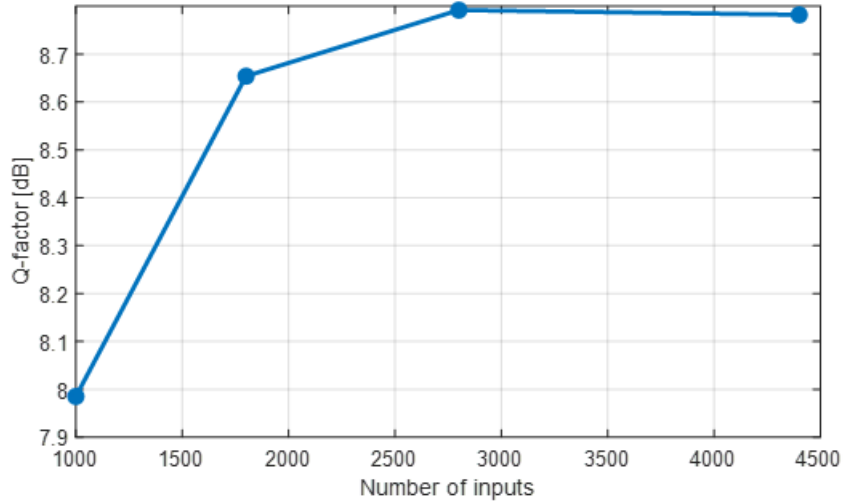
Figure 6.9: The Q-factor of a 2-branch Siamese NN with an additional PCA input stage

numerical back-propagation or kernel methods. Accordingly, this chapter has analyzed two input branch SNN designs as well as a three-input branch SNN. The Q-factor enhancement obtained when each of these network configurations was applied to the determination of the nonlinear noise accumulated during single frequency channel propagation in optical fiber was 0.85 dB. However, including a PCA stage before each input branch enabled a reduction in the number of inputs in the two branch SNN such that a 0.75 dB enhancement could be attained with half the input data.

Chapter 7

Non-linear Noise Compensation Study

7.1 Introduction

In previous chapters, we investigated the degree to which nonlinear noise prediction and compensation can be enhanced by including nonlinear perturbation terms ($X_n X_{m+n}^* X_m$, $Y_n Y_{m+n}^* X_m$, and $Y_n Y_{m+n}^* Y_m, X_n X_{m+n}^* Y_m$) at the receiver (chapter 4,6), and transmitter sides (chapter 5). This approach was also applied to an AI technique in which the fiber nonlinearity was compensated at both a transmitter stage through a NN and a second receiver stage employing various classifier strategies (chapter 5). Since the ML performance improvement depends on the characteristics of the dataset and, in particular, the signal-to-noise ratio, we here examine the performance of the previously introduced techniques at high nonlinear noise levels. This provides additional insight into the robustness of different AI techniques to nonlinear noise and aids in determining the most appropriate compensation technique for a given set of fiber properties. The results of this analysis could potentially reduce the cost and performance requirements of nonlinear optical fibers in optical communication system applications.

7.2 Results

7.2.1 One stage AI Techniques

Neural Networks

To determine the NN performance for different fiber nonlinearities, the nonlinear noise generated by fiber nonlinearity is quantified by the parameter γ. As shown in figures 7.1 and 7.2, the NN performance decreases with increasing γ for $\gamma > 2$ $W^{-1}km^{-1}$, even when NN complexity increased by increasing the number of triplets as input to the NN. Therefore, to describe the NN performance mathematically, we fit a curve to the average of all the results for different thresholds with the same γ value. The R-squared values associated with the curve fit are 96.7% and 92% when the NN is placed at the receiver and transmitter sides, respectively. As shown in figures 7.1 and 7.2, since the curves are similar in shape, they can be parameterized by the function of γ and the threshold magnitude θ according to,

$$\Delta Q = \frac{\theta}{21.6} \exp\left(-0.03\gamma\right), \tag{7.1}$$

when the NN is placed at the receiver side and

$$\Delta Q = \frac{\theta}{21.6} \exp\left(-0.024\gamma\right), \tag{7.2}$$

for a NN at the transmitter. These curves nearly coincide with the data and are nearly identical in the two cases, as is evident from figure 7.1, where the values obtained from the above formulas (solid lines) and the optimal fit (dotted lines) are compared with the data for 25 dB and 20 dB threshold values, respectively.

Siamese Neural Networks (SNN)

As shown in chapter 6, two designs of the SNN were employed at the receiver side of the system to mitigate the fiber nonlinearity with the aid of the nonlinear perturbation terms. To investigate the performance of the proposed SNN under different signal-to-noise ratios, the proposed SNN designs were also applied to systems with different values of γ, as shown in figures 7.3 and 7.4. While the Q-factor decreases monotonically with γ similarly to systems in which a standard NN is applied, either at the transmitter or the receiver sides, the magnitude of the Q-factor is less than that of the NN for all γ which is consistent with
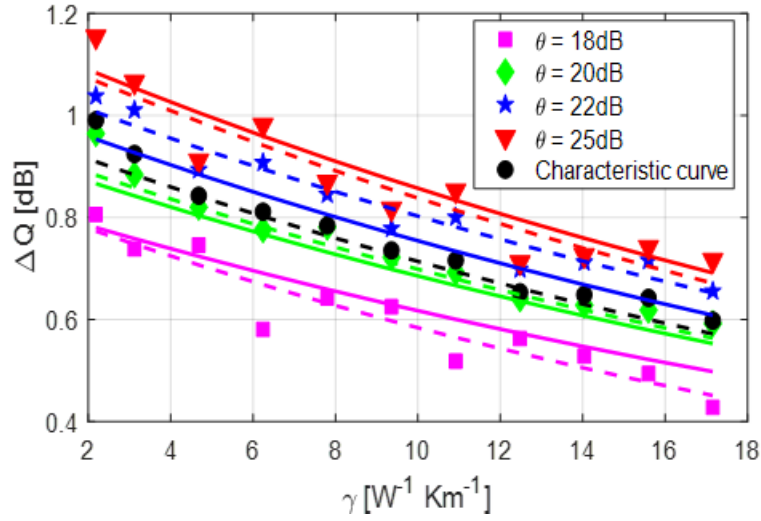
Figure 7.1: The performance of NN at the receiver side versus the nonlinearity coefficient. The dotted lines are the optimal algebraic curves while the solid lines are the characteristic equation representation.
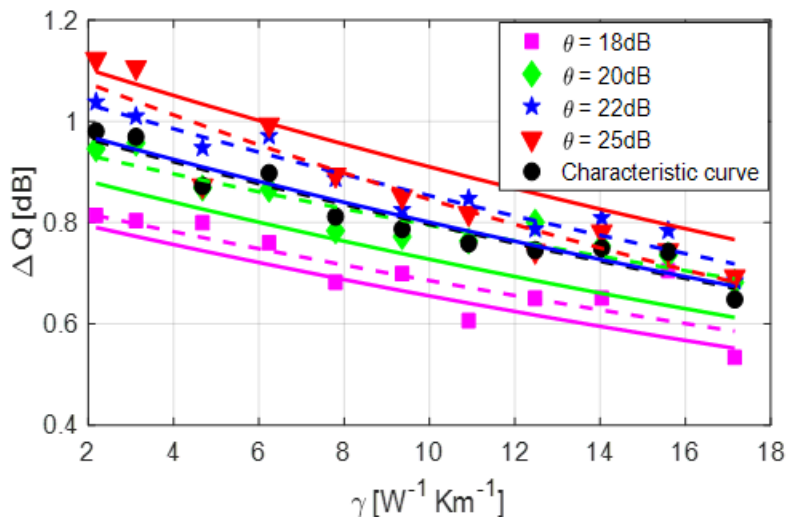
Figure 7.2: The performance of NN at the transmitter side versus the nonlinearity coefficient. The dotted lines are the optimal algebraic curves while the solid lines are the characteristic equation representation.
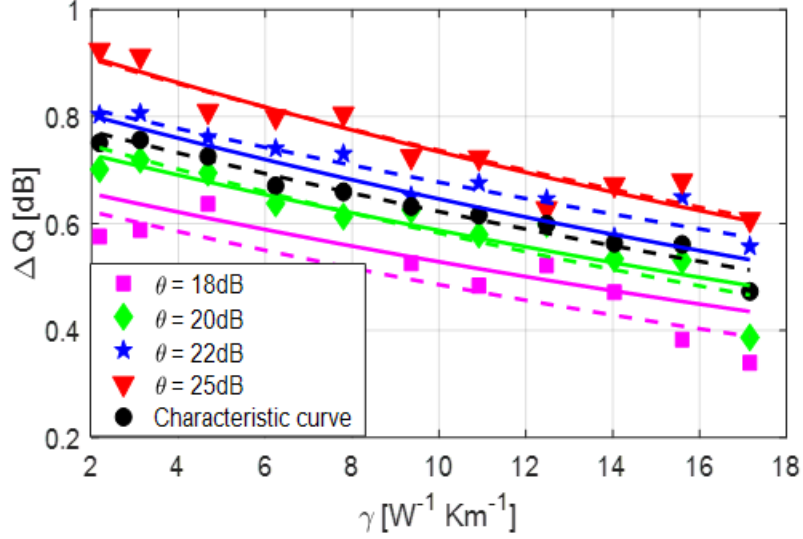
Figure 7.3: The Q value improvement associated with the SNN for two branches at the receiver side as a function of the nonlinearity coefficient. The dotted lines are the optimal algebraic curves while the solid lines are the characteristic equation representation.

the results of chapter 6. Indeed, for either SNN implementation, the improvement in the Q factor is approximately given by

$$\Delta Q = \frac{\theta}{26} \exp\left(-0.026\gamma\right), \tag{7.3}$$

which is clearly smaller than the corresponding quantities in equations 7.1 and 7.2 since the exponents in the equations are nearly identical. The R-squared values associated with these fits to the averaged curves are 93% and 96%, respectively.

7.2.2 Two-stage AI Techniques

Decision Tree

In section 5.2.2, a system with $\gamma = 1.4$ W^{-1}km^{-1} achieved a smaller Q-factor enhancement when a decision tree was employed as a classifier at the receiver compared to an optimized system with a NN placed at the transmitter. Figure 7.5 demonstrates that a decision tree, in fact, exhibits slightly improved performance relative to the NN implantation as γ
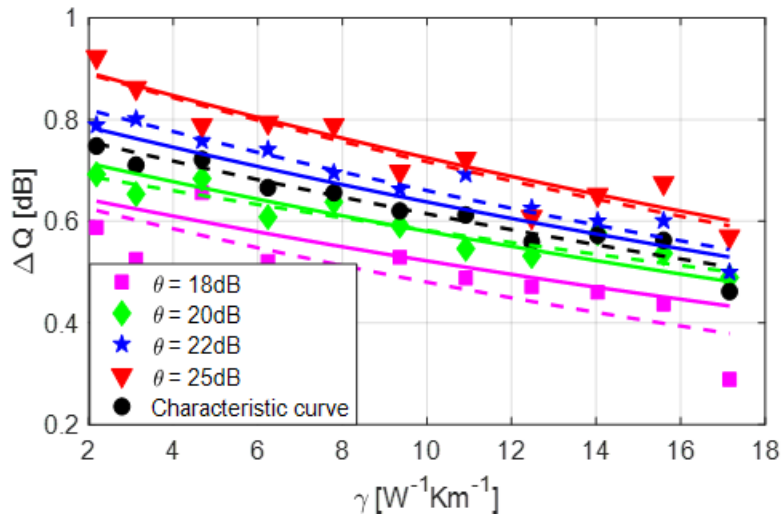
Figure 7.4: The Q value improvement associated with the SNN for three branches at the receiver side as a function of the nonlinearity coefficient. The dotted lines are the optimal algebraic curves while the solid lines are the characteristic equation representation.

increases. Accordingly, a decision tree is an ineffective classifier for nonlinearity mitigation. The continuous curve shown in figure 7.5 corresponds to the formula

$$\Delta Q = \frac{\sqrt{\theta}}{4.4} \gamma^{-0.14}, \tag{7.4}$$

indicating that the Q-factor enhancement is described by a negative power of γ with a 90% R-squared value.

Boosting

To improve the decision tree results, the classifier performance can be enhanced through boosting, as demonstrated for AdaBoosting in chapter 5. As shown in figure 7.6, a two-stage AI technique with AdaBoosting at the receiver yields improved performance relative to an uncompensated system for $\gamma < 9$ W^{-1}km^{-1}. However, the probability of misclassified data increases with γ. This negatively affecting performance, since the accuracy of the weight updating parameter depends on the ratio of incorrect to correct classifications, therefore it degrades rapidly as the noise increases [87].

Figure 7.5: Two-stage AI technique performance with a NN at the transmitter and a decision tree at the receiver as a function of the nonlinearity coefficient. The dotted lines represent the optimal fit while the solid lines are generated from an algebraic formula.

Strong Gboosting classifiers, which minimize the classification error by combining several weak classifiers have proved effective in compensating high nonlinear noise levels [91]. Indeed, Figure 7.7 demonstrates that the system performance enhancement decreases with γ, although for small γ the performance of G-boosting is less than that of Adaboosting for which the Q-factor enhancement can be approximated by

$$\Delta Q = 2.6 \exp\left(-0.089\gamma\right) - 1, \tag{7.5}$$

with a fitting parameter of $R^2 = 97.3\%$ which is almost independent of the threshold level. The corresponding G-boosting Q-factor enhancement in figure 7.7 is approximated with $R^2 = 95\%$ as,

$$\Delta Q = \frac{\sqrt{\theta}}{3.3}\gamma^{-0.171}, \tag{7.6}$$

Thus AdaBoosting is most advantageous for low fiber nonlinearity and limited computational resources as the number of triplets can be considerably reduced compared to G-boosting. Two-stage G-boosting technique was further found to be preferable to an isolated NN over the range of γ examined.

76
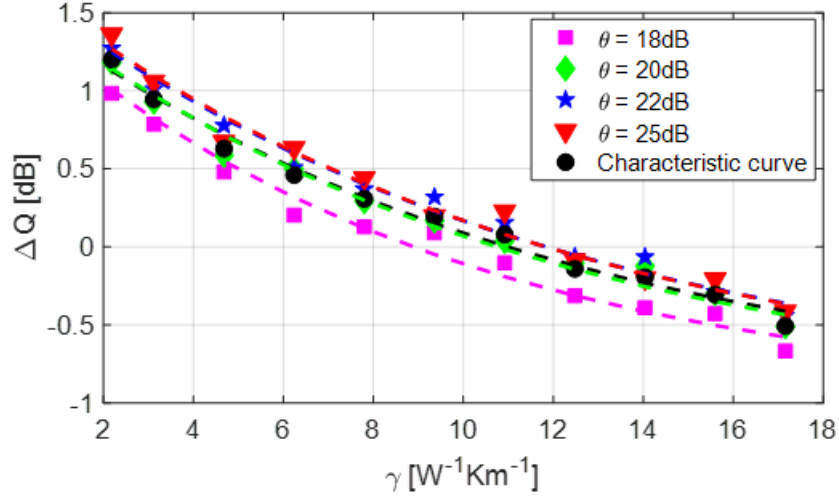
Figure 7.6: Two-stage AI performance with a NN at the transmitter and AdaBoosting at the receiver as a function of the nonlinearity coefficient. The dotted lines are the algebraic approximation for the results.
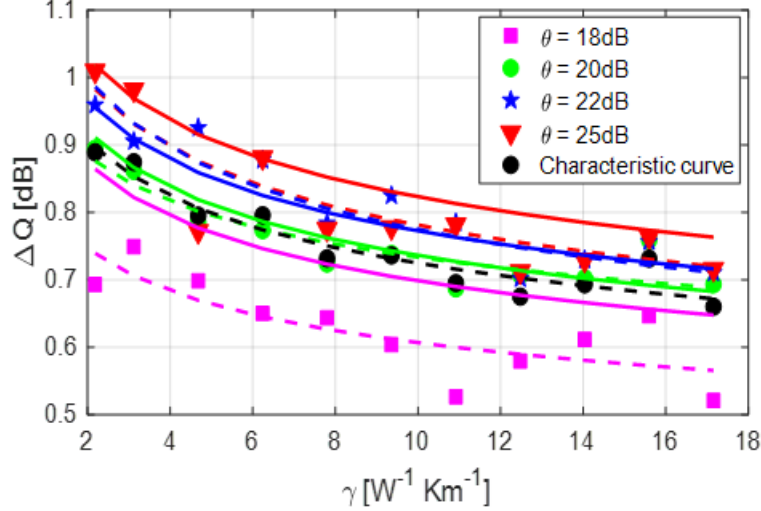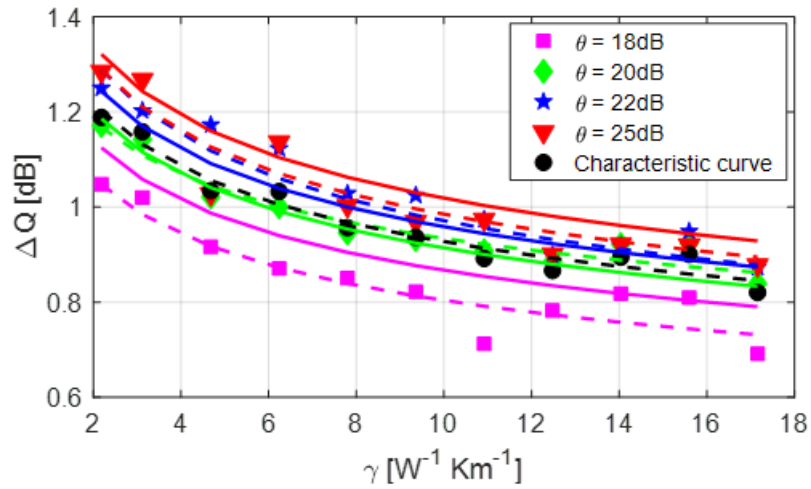

Figure 7.7: Two-stage AI performance with a NN at the transmitter and G-boosting at the receiver as a function of the nonlinearity coefficient. The dotted lines are the algebraic approximation for the results.

Random Forest and Extra trees

Although the decision tree technique yields only a limited improvement in the Q-factor relative to the results of the previous section, figures 7.8 and 7.9 demonstrate that the tree ensembles associated with the random forest and extra trees techniques enable significant further improvements. Moreover, figures 7.8 and 7.9 show that the random forest and extra trees techniques compensate for nonlinear noise more effectively than competing algorithms, especially at high levels of nonlinearity while the computational time is typically limited as indicated in chapter 5. As shown in the previous procedures, to model the dependence of the Q-factor enhancement in the random forest and extra trees methods on γ, an average is performed over of all the results for different thresholds with the same γ value. Unlike the previous methods, the slope of the curves in the random forest and extra trees procedures is approximately given as the following function of the threshold value,

$$\Delta Q = \frac{\sqrt{\theta}}{3.8}\gamma^{\frac{2.94}{\theta}}, \tag{7.7}$$

Figures 7.8 and 7.9 compare equation 7.7 (solid lines) to the optimum fit to the results (dotted lines). The agreement is especially pronounced for the 25 dB and 20 dB threshold curves.

Multi-layer perceptron (MLP) classifier

Employing a MLP classifier containing a single 4 neuron hidden layer with a 'Relu' activation function at the receiver, trained as indicated at the beginning of this section, yields the curves in figure 7.10. While this architecture is near-optimum, the Q-factor improvement is effectively identical to that of a standard NN placed at the receiver end. Further, the equation

$$\Delta Q = \frac{\theta}{21.6}\exp\left(-0.019\gamma\right), \tag{7.8}$$

that parameterizes the curves is almost identical to that associated with a transmitter side NN. This is identical to the result in chapter 5, which predicted a 0.03 dB enhancement for $\gamma = 1.4$ W^{-1}km^{-1}. Indeed, from figure 7.11, which displays the averaged results for the Q-factor enhancements associated with the AI configurations analyzed in this chapter, it is evident that all techniques perform nearly identically for small γ but differ increasingly for larger values of γ. This figure further establishes that the most appropriate AI technique for the system under consideration is a two-stage structure, with either a random forest or extra trees at the receiver side.
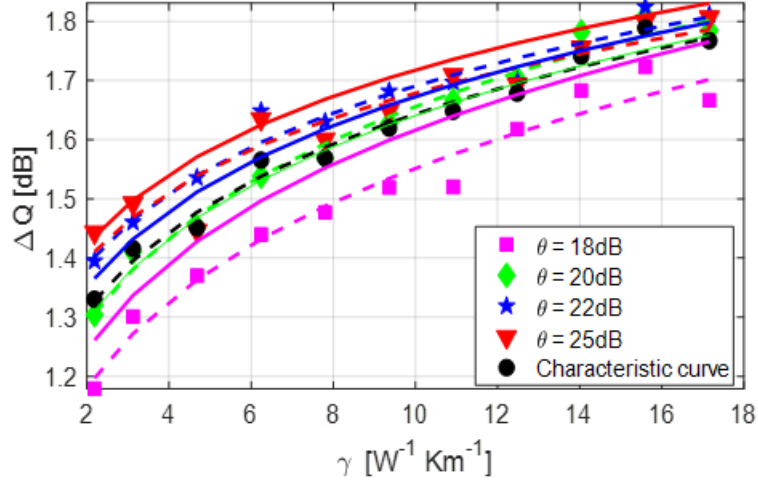
Figure 7.8: The Q-factor improvement as a function of γ for a two-stage AI technique with a NN at the transmitter side and a random forest at the receiver side. The dotted lines represent the best fit while the solid lines correspond to the algebraic approximation.
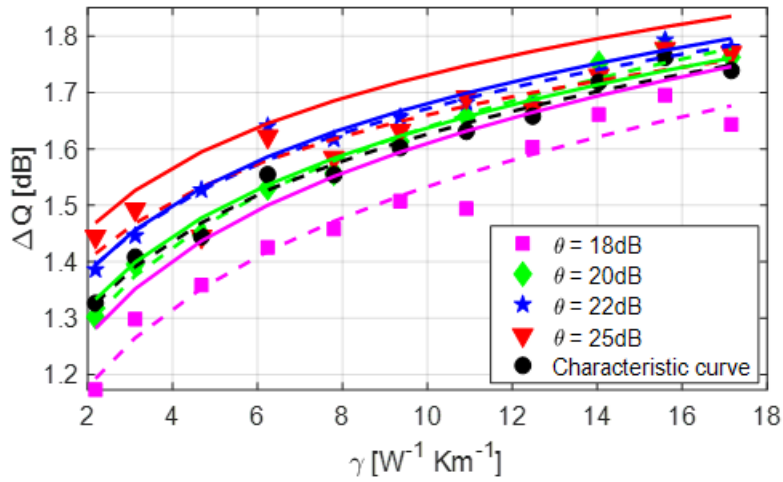
Figure 7.9: The Q-factor improvement as a function of γ for a two-stage AI technique with a NN at the transmitter side and extra trees at the receiver side. The dotted lines represent the best fit while the solid lines correspond to the algebraic approximation.
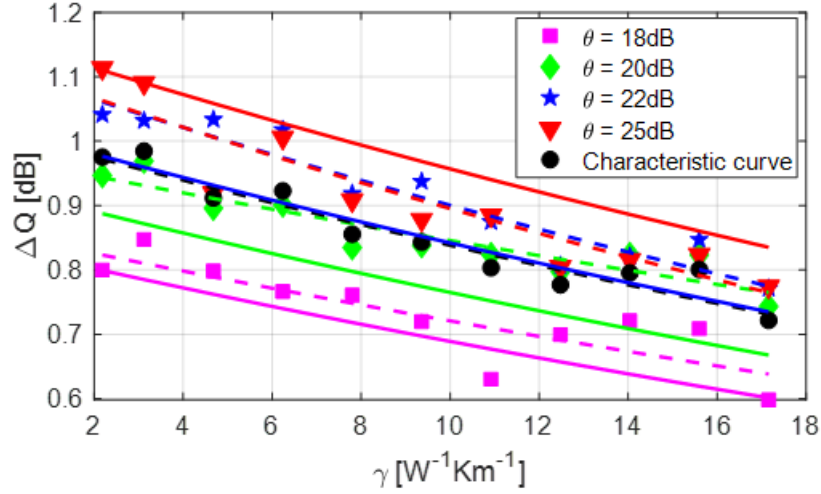
Figure 7.10: The performance for a transmitter side NN and a receiver side MLP as a function of the nonlinearity coefficient. The dotted lines are the best fit to the data while the solid lines are generated with equation 7.8
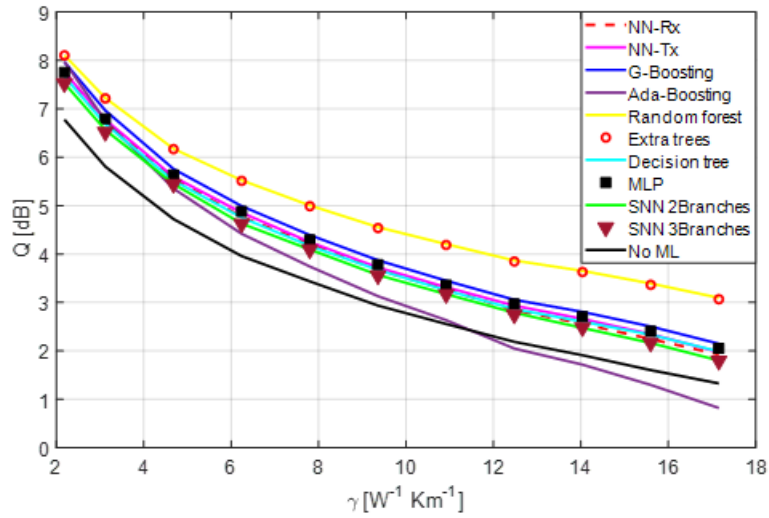
Figure 7.11: The system Q-factor improvement for different AI techniques

7.3 Summary

The results, in this chapter, demonstrate that a NN can be employed either at the receiver or the transmitter side over a wide range of nonlinear noise levels with identical Q-factor enhancements. Employing instead a SNN at the receiver side leads to slightly reduced performance. On the other hand, two-stage AI classifiers such as extra trees and random forests at the receiver can significantly compensate for high levels of nonlinear noise, while decision trees do not afford any noticeable advantage over the standard NN procedure. Additionally, Ada-boosting improves performance for small nonlinear coefficients, even if a reduced number of triplets are used as input into the transmitter NN, but its effectiveness decreases rapidly with nonlinearity. For each topology, an empirical algebraic equation was generated for the system Q-factor enhancement in terms of the triplet selection threshold and γ. The enhancement associated with each technique at any nonlinear noise value can therefore be rapidly estimated, which should be useful in communication system design.

Chapter 8

Conclusions and Future Work

This thesis demonstrates that significant Q-factor improvements can be achieved in high-capacity optical communications by mitigating fiber nonlinearity. These improvements can be achieved with less complexity by applying AI techniques compared to the conventional DBP technique and analytical solutions such as Volterra nonlinear compensators and perturbation-based compensation techniques. A variety of AI techniques and methods were examined to explore the trade-off between computational efficiency and Q-factor improvement. The results demonstrate the robustness of AI techniques, as they can achieve better system Q-factor at high nonlinear coefficient values compared to the Perturbation based compensation method. The proposed AI techniques can achieve an enhancement in the range between 0.85 dB and 1.3 dB in the system Q-factor compared to a system without any nonlinear compensator. Two stage AI method with Random forests and Extra trees classifiers at the receiver side have shown better performance and robustness than others.

A simulation program based on MATLAB has been used to model the propagation of a 16-QAM dual-polarization optical system through 3200 km optical fiber of 0.2 dB/km attenuation, 17 ps/km.nm dispersion, 1.4 $W^{-1}km^{-1}$ nonlinear coefficient, and 4.5 dB noise figure. The MATLAB simulation was also used to generate the 32 GB transmitted and received data. The simulation program is able to model a variety of transmitter side functions such as generating bits, encoding, and modulating the signal; channel parameters such as ASE-noise, CD, Kerr nonlinearity, and PMD using SSFM; as well as receiver side functions such as filtering, down-sampling, CDC, demodulation, and decoding. The AI techniques proposed were modeled using python language on scikit-learn, keras, and tensor-flow platforms.

In the previously proposed methods, the NNs were used to mitigate optical fiber non-linearity by simplifying the evaluation of DBP method [45, 89] or predicting the nonlinear noise using a Dynamic deep neural network. Despite the improvement, the number of the hidden layers needed in both designs were impractical. As in the case of solving DBP, the achieved gain in Q-factor is 0.2 dB better than the conventional DBP with 2 steps per span; however, the number of the hidden layers depends on the number of steps per span and number of spans of the link. Also, for the prediction case, using dynamic deep neural networks at the receiver side, the inputs of the network are simply a window of the received symbol of interest with a group of its neighboring symbols. Consequently, the NN input does not have enough features, and the optimum design for the NN becomes complex with a large width and more hidden layers [89]. The achieved enhancement for a link range between 1500 km and 2700 km is 0.9 dB to 1 dB, using a NN with two 16 neuron hidden layers, but each BER point is calculated by averaging the error rate over 15 signal block transmissions. While this reduces the effect of noise in the data input to the NN, it increases the system complexity and decreases the response time. In our model, the SPM, Intra XPM, and Intra four-wave mixing triplets terms were added as extra inputs to the FFNN with the data symbols. The optimum system configuration, based on the symbol of interest and its orthogonal polarization symbol, achieved 1 dB enhancement with only 2 neurons in one hidden layer, Relu activation function, and 2445 triplets. In order to reduce the complexity of the system, a two cascaded complexity reduction technique has been applied using PCA and weight trimming. Using these methods, we could achieve the same performance with only 65% of the inputs. Another advantage to this method over previously mentioned methods is that we are using a scaling factor to use the same trained NN for different channel's launch powers and to predict the orthogonal polarization data.

Subsequently, the NN was moved from the receiver to the transmitter to pre-compensate the nonlinear distortion, before signal propagation to reduce the computational overhead in triplets calculations as it can be implemented using LUT. Moreover, this transition adds another opportunity to improve the overall performance by using an additional AI stage. We simulated several classifying techniques at the receiver. These included decision trees, boosting, random forests, extra trees, and MLP. Also, comparing the computational time, as an indication of the complexity, of the two-stage compensation classifiers was performed. AdaBoosting, random forest, and extra trees achieved up to 1.3 dB enhancement in the system Q-factor; however, extra trees have optimal performance due to its least computational time which is 47% less than the random forest method and 81% less than AdaBoosting.

In an attempt to explore a different AI topology to achieve better performance, SNN was used for the first time in this field. Two designs of SNN were explored: two branches

and three branches SNN. When the typically used two-branches topology yielded limited Q-factor improvements, 0.85 dB, the three-branches topology was explored. Unfortunately, the three-branches architecture did not improve the performance much despite its increased complexity. Complexity reduction has been applied to reduce the number of SNN inputs for the two-branches design using PCA. Adding PCA to the system achieved 0.75 dB enhancement in Q-factor with only 50% of the inputs used in the original design.

To compare the robustness of the above-mentioned compensation AI methods against increasing levels of nonlinearities from the optical fiber was examined. The calculations were performed for multiple nonlinear coefficients. For each nonlinear coefficient value, NNs with several computational complexities were examined by changing their input sizes (cut-off thresholds/number of triplets). The study concluded that two-stage AI techniques with extra trees and random forests are the most robust techniques against nonlinear noise. Moreover, AdaBoosting performance is independent of the number of triplets, but it deteriorates the system performance at high nonlinear coefficient values. We also introduced an empirical equation for each technique that represents an approximation for the performance of the technique under different nonlinear noise and computational complexity. This is important to efficiently evaluate the performance of different techniques on different system parameters, thus saving time and resources.

Without actual implementation on physical links or additional extensive simulations of other techniques, the proposed AI techniques and previous analytical solutions can be meaningfully compared by analyzing their relative computational complexity. This is especially true in the case of long-haul optical transmission. For example, in DBP, the complexity for the evaluation of each sample is in the order of $O(N_{step}N_{FFT}\log_2(N_{FFT}))$. This is a huge number because it depends on the total number of propagating steps, which increases with the cable length. It also depends on the total number of symbols fed to each FFT block. Furthermore, the Volterra compensation technique is applied to the signal in the frequency domain, and therefore, the complexity is still dependent on the size of the FFT block used, with an order of $O(N_{FFT}^2)$. Similarly, the perturbation-based nonlinear compensator (PB NLC) technique works on the symbols level. Consequently, there is no need to apply FFT, but the system complexity still remains high due to the large number of complex multiplications included in calculating the perturbation and C_{mn} coefficients. On the other hand, using the AI techniques highly reduced the required computational complexity compared to the DBP and Volterra techniques. For instance, in the case of NN at the transmitter side, the complexity is in the order of $O(N_{inputs})$ where N_{inputs} is the number of inputs to the neural network. This is a significant improvement over any of the previous techniques. Also, the number of complex multiplications needed to construct the NN inputs is lower than that of the PB NLC case, as shown in figure 8.1.
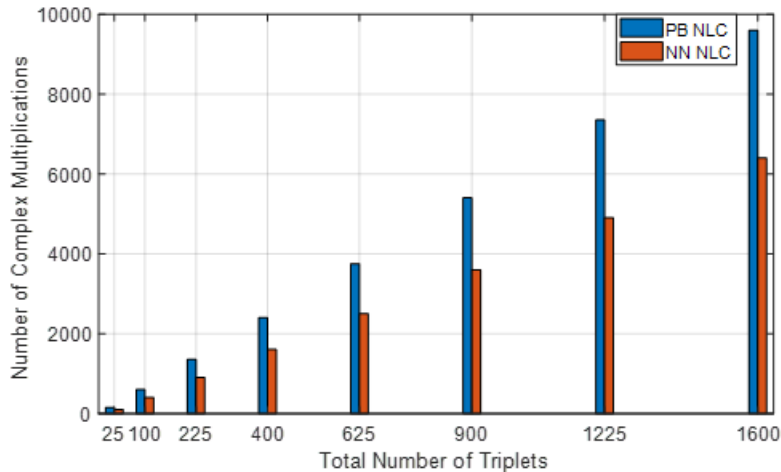
Figure 8.1: Comparing the number of complex multiplications needed in Perturbation based compensation method versus neural network

In addition to the reduced complexity, the proposed AI techniques provide additional improvements over traditional methods from the performance point of view (Q-factor). Applying the perturbation-based technique on our system achieves Q-factor enhancement of 1.2 dB compared to 1 dB enhancement using only NN at the transmitter or receiver side, and 1.3 dB in two-stage techniques using AdaBoosting, random forest, and extra trees classifiers at the receiver side. Also, unlike the AI techniques, using the same number of triplets, the PB NLC technique is vulnerable to any changes in nonlinear coefficient values as shown in figure 8.2. The performance of the PB NLC significantly drops as the nonlinear coefficient slightly increases more than $\gamma = 2$ W^{-1}km^{-1} and decrease the system performance compared to its performance without any compensators at $\gamma = 3$ W^{-1}km^{-1}. It can be noticed that the Q-factor of the system reaches zero at $\gamma = 5.8$ W^{-1}km^{-1}. This is because the number of triplets used in the PB NLC calculations is not sufficient to calculate the newly added nonlinear noise and so mislead the systems' decision about the transmitted symbol. Increasing the number of triplets means adding more complexity to the system, and more computational resources are needed. In contrast, the AI techniques are more robust against the change in nonlinear coefficient values, as they can achieve good enhancement in the system Q-factor till $\gamma = 17$ W^{-1}km^{-1}.

Practically, this robustness of the proposed AI techniques is helpful. First, it gives us insight into the behavior of these techniques in WDM systems since the nonlinear noise is more aggressive in multichannel systems, leading to a decrease in the system's performance.
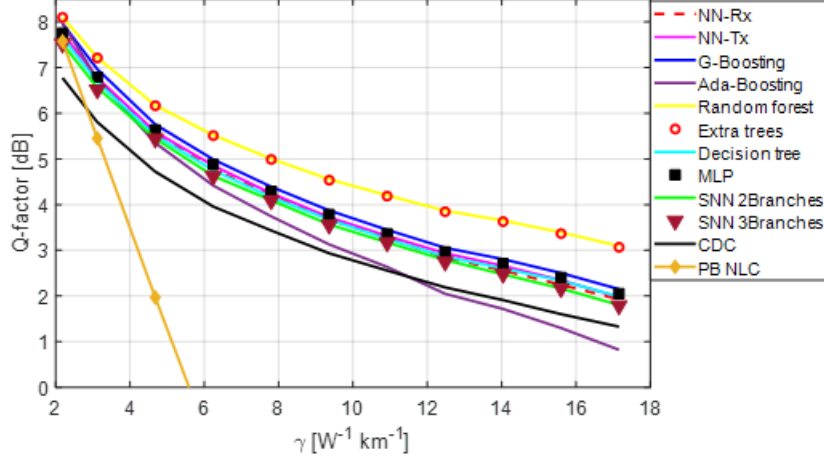
Figure 8.2: Comparing the robustness of Perturbation based compensation method versus neural network against fiber nonlinearity

Second, it adds more tolerance in fiber optics manufacturing and reduces the system's cost by achieving good performance with lower quality fibers.

As shown in this thesis, AI techniques have a promising performance in mitigating fiber nonlinearity. To practically implement such techniques, different chips and accelerators have been released lately for this purpose, such as IBM Telum Processor, Google Tensor Processing Units (TPU), Intel Ampere AI chip, and GraphCore Intelligence Processing Units (IPU). Most of these accelerators have software on which the user can run their AI codes. In the future, the proposed algorithms can be implemented on such chips, trained offline then added to the transmitter or the receiver side during operation, which is an added advantage for using AI techniques rather than other analytical methods.

The AI techniques used in this thesis are data-driven models, where the optimum design and parameters are exclusive to the data characteristics and features generated from this system. Changing the channel or signal parameters will lead to a new optimum design and a new set of curves and equations, but it is expected that these would exhibit similar behavior.

In the future, the techniques developed in this thesis could be extended to the mitigation of nonlinear distortions in WDM systems. A more comprehensive study can be implemented to examine the performance of the proposed techniques to compensate for nonlinearities in signals propagating through fibers with different numbers of channels.

Also of interest is the effect of changing the nonlinear coefficient value, PMD, and the NN complexity on the performance, as well as confirming the proposed empirical equations for each technique in chapter 7 on the WDM systems.

Another area of research could be enhancing the AI techniques' performance by adding more features to the NN inputs, which represent the inter XPM effect and inter four-wave mixing. Taking into consideration the effect of the walk-off between the neighboring channels, as well as the high computational overhead added in extending the investigation to the WDM systems. A further avenue of exploration could be finding a low complexity regression AI technique to compute the pre-compensation technique triplets before they are added to the NN. This will significantly reduce the complexity of the proposed techniques both in SMF and WDM cases. Additionally, the compensation performance of deep learning methods for large data sets, both with and without the addition of the triplet terms, can be examined. A further topic of interest would be to combine different compensation methods, such as applying AI nonlinearity mitigation methods to PS modulated systems. Since the probability mass function of the symbols is non-uniform, the data set would be unbalanced, which would affect the AI performance. Therefore, system training should be implemented on balanced data sets, but the performance should be evaluated using the PS data sets.

As a comparison for all DSP techniques introduced in the literature survey and proposed in the thesis, table 8.1 presents the performance and complexity of each technique. In the table, advantages are highlighted in green and drawbacks in red. From the table, it can be easily concluded that the use of FFNN at the receiver or transmitter side, together with a two-stage AI technique employing random forest and extra trees classifiers at the receiver, outperforms the other techniques with lower complexity.

Technique	Performance	Complexity per symbol
Analytical Solutions		
Digital back-propagation DBP	State of the art with appropriate number of samples per symbol	Extremely-high $O(N_{step}N_{FFT}\log_2(N_{FFT}))$
Volterra nonlinear compensation	Good Higher than DBP with low sampling rate (2 samples/symbol)	High $O(N_{FFT}^2)$
Advanced Modulation Format	Under some conditions can reach theoretical Shannon capacity formula.	Challenging • Finding the optimum distribution is not always possible. (PS) • Requires non-conventional equalization and/or phase noise recovery.
Perturbation based (PB NLC)	• Acceptable performance in single channel optical fiber. • Do not account for inter-channel nonlinearities	High due to the large number of complex multiplications in calculating the perturbation and C_{mn} coefficients
Artificial Intelligence Solutions		
Neural Networks FFNN	• Very close to PB NLC performance using the same number of triplets. • More robust than PB NLC under any change in the fiber nonlinearity therefore a promising performance is expected for it in WDM system.	Low complexity in • Structure: $O(N_{inputs})$ • Data preprocessing: Number of complex multiplication lower than PB NLC using the same number of triplets.
Technique	Performance	Complexity per symbol
Two-stage AI — Decision tree	Worse than using the NN only.	Time complexity $O(n^2)$ [92] n is the number of attributes[1].
Two-stage AI — Multi-layer perceptron MLP	Almost the same as NN.	Depend on the size of the NN

	GBoosting	• Better than using only NN. • Achieve enhancement at high nonlinearity	Training time is the highest among used techniques.
	AdaBoosting (Decision tree)	• Better than GBoosting. • Least robust against high changes in nonlinearity • Least dependent on the Tx NN complexity.	Training time lower than GBoosting but higher than Extra trees and Random forest.
	Random forests	• The best performance among proposed techniques. • The most robust against changes in fiber nonlinearity.	Training time slightly higher than Extra trees.
	Extra Trees	• The best performance among proposed techniques. • The most robust against changes in fiber nonlinearity	The least training time.
Siamese Neural Networks SNN		Less than the NN performance	Slightly higher than the NN.

Table 8.1: Summary of Presented Typologies in the Thesis

1

---

[1]An attribute that best partitions the training data is chosen as the splitting attribute for the root, and the training data are then partitioned into disjoint subsets satisfying the values of the splitting attribute.

References

[1] Milton Abramowitz and Irene A Stegun. Handbook of mathematical functions with formulas, graphs, and mathematical tables. national bureau of standards applied mathematics series 55. tenth printing. 1972.

[2] Govind P Agrawal. Nonlinear fiber optics. In *Nonlinear Science at the Dawn of the 21st Century*, pages 195–211. Springer, 2000.

[3] Govind P Agrawal. Optical communication: its history and recent progress. In *Optics in our time*, pages 177–199. Springer, Cham, 2016.

[4] Erik Agrell, Alex Alvarado, Giuseppe Durisi, and Magnus Karlsson. Capacity of a nonlinear optical channel with finite memory. *Journal of Lightwave Technology*, 32(16):2862–2876, 2014.

[5] Alex Alvarado, Tobias Fehenberger, Bin Chen, and Frans MJ Willems. Achievable information rates for fiber optics: Applications and computations. *Journal of Lightwave Technology*, 36(2):424–439, 2018.

[6] Erick de A Barboza, Carmelo JA Bastos-Filho, Joaquim F Martins-Filho, Uiara C de Moura, and Juliano RF de Oliveira. Self-adaptive erbium-doped fiber amplifiers using machine learning. In *2013 SBMO/IEEE MTT-S International Microwave & Optoelectronics Conference (IMOC)*, pages 1–5. IEEE, 2013.

[7] Thomas Baumeister, Steven L Brunton, and J Nathan Kutz. Deep learning and model predictive control for self-tuning mode-locked lasers. *JOSA B*, 35(3):617–626, 2018.

[8] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. *Deep learning*, volume 1. MIT press Massachusetts, USA:, 2017.

[9] Lotfollah Beygi, Erik Agrell, Joseph M Kahn, and Magnus Karlsson. Rate-adaptive coded modulation for fiber-optic communications. *Journal of lightwave technology*, 32(2):333–343, 2013.

[10] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[11] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a" siamese" time delay neural network. *Advances in neural information processing systems*, 6:737–744, 1993.

[12] Steven L Brunton, Xing Fu, and J Nathan Kutz. Self-tuning fiber lasers. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(5):464–471, 2014.

[13] Fred Buchali, Fabian Steiner, Georg Böcherer, Laurent Schmalen, Patrick Schulte, and Wilfried Idler. Rate adaptation and reach increase by probabilistically shaped 64-qam: An experimental demonstration. *Journal of Lightwave Technology*, 34(7):1599–1609, 2016.

[14] Davide Cangelosi, Simone Pelassa, Martina Morini, Massimo Conte, Maria Carla Bosco, Alessandra Eva, Angela Rita Sementa, and Luigi Varesio. Artificial neural network classifier predicts neuroblastoma patients' outcome. *Bmc Bioinformatics*, 17(12):83–93, 2016.

[15] John C Cartledge, Fernando P Guiomar, Frank R Kschischang, Gabriele Liga, and Metodi P Yankov. Digital signal processing for fiber nonlinearities. *Optics express*, 25(3):1916–1936, 2017.

[16] Grigorios Charalabopoulos, Peter Stavroulakis, and A Hamid Aghvami. A frequency-domain neural network equalizer for ofdm. In *GLOBECOM'03. IEEE Global Telecommunications Conference (IEEE Cat. No. 03CH37489)*, volume 2, pages 571–575. IEEE, 2003.

[17] T-K Chiang, Noboyuki Kagi, ME Marhic, and Leonid G Kazovsky. Cross-phase modulation in fiber links with multiple optical amplifiers and dispersion compensators. *Journal of Lightwave Technology*, 14(3):249–260, 1996.

[18] Davide Chicco and Cristina Rovelli. Computational prediction of diagnosis and feature selection on mesothelioma patient health records. *PloS one*, 14(1):e0208737, 2019.

[19] Junho Cho and Peter J Winzer. Probabilistic constellation shaping for optical fiber communications. *Journal of Lightwave Technology*, 37(6):1590–1607, 2019.

[20] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 539–546. IEEE, 2005.

[21] Rishabh Choudhary and Hemant Kumar Gianey. Comprehensive review on supervised machine learning algorithms. In *2017 International Conference on Machine Learning and Data Science (MLDS)*, pages 37–43. IEEE, 2017.

[22] Cisco. Cisco annual internet report (2018–2023) white paper, Mar 2020.

[23] Edson Porto da Silva, Metodi P Yankov, Francesco Da Ros, Toshio Morioka, and Leif K Oxenløwe. Perturbation-based fec-assisted iterative nonlinearity compensation for wdm systems. *Journal of Lightwave Technology*, 37(3):875–881, 2018.

[24] Zhenhua Dong, Faisal Nadeem Khan, Qi Sui, Kangping Zhong, Chao Lu, and Alan Pak Tao Lau. Optical performance monitoring: A review of current and future technologies. *Journal of Lightwave Technology*, 34(2):525–543, 2015.

[25] Liang Dou, Zhenning Tao, Lei Li, Weizhen Yan, Takahito Tanimura, Takeshi Hoshida, and Jens C Rasmussen. A low complexity pre-distortion method for intra-channel nonlinearity. In *2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference*, pages 1–3. IEEE, 2011.

[26] Thomas Duthel, Georg Clarici, Chris RS Fludger, Jonas C Geyer, Christoph Schulien, and Stefan Wiese. Laser linewidth estimation by means of coherent detection. *IEEE Photonics Technology Letters*, 21(20):1568–1570, 2009.

[27] Desurvire Emmanuel and MN Zervas. Erbium-doped fiber amplifiers: principles and applications, 1994.

[28] Tobias Fehenberger, Alex Alvarado, Georg Böcherer, and Norbert Hanik. On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel. *Journal of Lightwave Technology*, 34(21):5063–5073, 2016.

[29] Tobias Fehenberger, Georg Böcherer, Alex Alvarado, and Norbert Hanik. Ldpc coded modulation with probabilistic shaping for optical fiber systems. In *Optical Fiber Communication Conference*, pages Th2A–23. Optical Society of America, 2015.

[30] Robert A Fisher and W Bischel. The role of linear dispersion in plane-wave self-phase modulation. *Applied Physics Letters*, 23(12):661–663, 1973.

[31] Robert A Fisher, BR Suydam, and D Yevick. Optical phase conjugation for time-domain undoing of dispersive self-phase-modulation effects. *Optics letters*, 8(12):611–613, 1983.

[32] T Freckmann, R-J Essiambre, PJ Winzer, GJ Foschini, and G Kramer. Fiber capacity limits with optimized ring constellations. *IEEE Photonics Technology Letters*, 21(20):1496–1498, 2009.

[33] Pedro J Freire, Yevhenii Osadchuk, Bernhard Spinnler, Antonio Napoli, Wolfgang Schairer, Nelson Costa, Jaroslaw E Prilepsky, and Sergei K Turitsyn. Performance versus complexity study of neural network equalizers in coherent optical systems. *arXiv preprint arXiv:2103.08212*, 2021.

[34] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[35] Kiyoshi Fukuchi, Tadashi Kasamatsu, Masao Morie, Risato Ohhira, Toshiharu Ito, Kayato Sekiya, Daisaku Ogasahara, and Takashi Ono. 10.92-tb/s (273× 40-gb/s) triple-band/ultra-dense wdm optical-repeatered transmission experiment. In *Optical Fiber Communication Conference*, page PD24. Optical Society of America, 2001.

[36] Guanjun Gao, Xi Chen, and William Shieh. Influence of pmd on fiber nonlinearity compensation using digital back propagation. *Optics Express*, 20(13):14406–14418, 2012.

[37] Ying Gao, John C Cartledge, Abdullah S Karar, Scott S-H Yam, Maurice O'Sullivan, Charles Laperle, Andrzej Borowiec, and Kim Roberts. Reducing the complexity of perturbation based nonlinearity pre-compensation using symmetric edc and pulse shaping. *Optics express*, 22(2):1209–1219, 2014.

[38] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.

[39] Amirhossein Ghazisaeidi, Ivan Fernandez de Jauregui Ruiz, Rafael Rios-Muller, Laurent Schmalen, Patrice Tran, Patrick Brindel, Alexis Carbo Meseguer, Qian Hu, Fred Buchali, Gabriel Charlet, et al. 65tb/s transoceanic transmission using

probabilistically-shaped pdm-64qam. In *ECOC 2016-Post Deadline Paper; 42nd European Conference on Optical Communication*, pages 1–3. VDE, 2016.

[40] Amirhossein Ghazisaeidi, Ivan Fernandez de Jauregui Ruiz, Laurent Schmalen, Patrice Tran, Christian Simonneau, Elie Awwad, Bogdan Uscumlic, Patrick Brindel, and Gabriel Charlet. Submarine transmission systems using digital nonlinear compensation and adaptive rate forward error correction. *Journal of Lightwave Technology*, 34(8):1886–1895, 2016.

[41] Alan H Gnauck, Peter J Winzer, Sethumadhaven Chandrasekhar, Xiang Liu, Benyuan Zhu, and David W Peckham. 10× 224-gb/s wdm transmission of 28-gbaud pdm 16-qam on a 50-ghz grid over 1,200 km of fiber. In *Optical Fiber Communication Conference*, page PDPB8. Optical Society of America, 2010.

[42] Neil Guerrero Gonzalez, Darko Zibar, and Idelfonso Tafur Monroy. Cognitive digital receiver for burst mode phase modulated radio over fiber links. In *36th European Conference and Exhibition on Optical Communication*, pages 1–3. IEEE, 2010.

[43] Robert M Gray. *Entropy and information theory*. Springer Science & Business Media, 2011.

[44] Fernando P Guiomar, Jacklyn D Reis, António L Teixeira, and Armando N Pinto. Mitigation of intra-channel nonlinearities using a frequency-domain volterra series equalizer. *Optics express*, 20(2):1360–1369, 2012.

[45] Christian Häger and Henry D Pfister. Nonlinear interference mitigation via deep neural networks. In *2018 Optical Fiber Communications Conference and Exposition (OFC)*, pages 1–3. IEEE, 2018.

[46] Haibo He and Edwardo A Garcia. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9):1263–1284, 2009.

[47] Yishen Huang, Patricia B Cho, Payman Samadi, and Keren Bergman. Dynamic power pre-adjustments with machine learning that mitigate edfa excursions during defragmentation. In *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3. IEEE, 2017.

[48] Yishen Huang, Wiem Samoud, Craig L Gutterman, Cedric Ware, Mounia Lourdiane, Gil Zussman, Payman Samadi, and Keren Bergman. A machine learning approach for dynamic optical channel add/drop strategies that minimize edfa power excursions.

In *ECOC 2016; 42nd European Conference on Optical Communication*, pages 1–3. VDE, 2016.

[49] Hiroki Iimori, Razvan-Andrei Stoica, and Giuseppe Thadeu Freitas de Abreu. Constellation shaping for rate maximization in awgn channels with non-linear distortion. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5. IEEE, 2017.

[50] Ezra Ip. Nonlinear compensation using backpropagation for polarization-multiplexed transmission. *Journal of lightwave technology*, 28(6):939–951, 2010.

[51] Ezra Ip and Joseph M Kahn. Compensation of dispersion and nonlinear impairments using digital backpropagation. *Journal of Lightwave Technology*, 26(20):3416–3425, 2008.

[52] ITU. itu-t recommendation database, 2021.

[53] Mutsam A Jarajreh, Elias Giacoumidis, Ivan Aldaya, Son Thai Le, Athanasios Tsokanos, Zabih Ghassemlooy, and Nick J Doran. Artificial neural network nonlinear equalizer for coherent optical ofdm. *IEEE Photonics Technology Letters*, 27(4):387–390, 2014.

[54] K Charles Kao and George A Hockham. Dielectric-fibre surface waveguides for optical frequencies. In *Proceedings of the Institution of Electrical Engineers*, volume 113, pages 1151–1158. IET, 1966.

[55] K. Charles Kao and George A Hockham. Dielectric-fibre surface waveguides for optical frequencies. *IEE proceedings. Part J, Optoelectronics*, 133(3):191–198, 1986.

[56] FP Kapron, Donald B Keck, and Robert D Maurer. Radiation losses in glass optical waveguides. *Applied Physics Letters*, 17(10):423–425, 1970.

[57] Johnny Karout, René-Jean Essiambre, Erik Agrell, and Antonia Tulino. Achievable rates of multidimensional multisphere distributions. In *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3. IEEE, 2017.

[58] Gerd Keiser. *Optical communications essentials*. McGraw-Hill Education, 2003.

[59] Kazuro Kikuchi. Characterization of semiconductor-laser phase noise and estimation of bit-error rate performance with low-speed offline digital coherent receivers. *Optics Express*, 20(5):5291–5302, 2012.

[60] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[61] Shiva Kumar, John C Mauro, Srikanth Raghavan, and Dipak Q Chowdhury. Intrachannel nonlinear penalties in dispersion-managed transmission systems. *IEEE Journal of Selected Topics in Quantum Electronics*, 8(3):626–631, 2002.

[62] David Large and James Farmer. Chapter 4 - linear fiber-optic signal transportation. In David Large and James Farmer, editors, *Broadband Cable Access Networks*, The Morgan Kaufmann Series in Networking, pages 81–126. Morgan Kaufmann, Boston, 2009.

[63] Xiaoxu Li, Xin Chen, Gilad Goldfarb, Eduardo Mateo, Inwoong Kim, Fatih Yaman, and Guifang Li. Electronic post-compensation of wdm transmission impairments using coherent detection and digital signal processing. *Optics Express*, 16(2):880–888, 2008.

[64] TH Lotz, X Liu, S Chandrasekhar, PJ Winzer, H Haunstein, S Randel, S Corteselli, B Zhu, and DW Peckham. Coded pdm-ofdm transmission with shaped 256-iterative-polar-modulation achieving 11.15-b/s/hz intrachannel spectral efficiency and 800-km reach. *Journal of Lightwave Technology*, 31(4):538–545, 2012.

[65] Niklas Lüdtke, Stefano Panzeri, Martin Brown, David S Broomhead, Joshua Knowles, Marcelo A Montemurro, and Douglas B Kell. Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks. *Journal of The Royal Society Interface*, 5(19):223–235, 2008.

[66] Valerio Maggio, Marco Chierici, Giuseppe Jurman, and Cesare Furlanello. Distillation of the clinical algorithm improves prognosis by multi-task deep learning in high-risk neuroblastoma. *PloS one*, 13(12):e0208924, 2018.

[67] Eduardo Mateo, Likai Zhu, and Guifang Li. Impact of xpm and fwm on the digital implementation of impairment compensation for wdm transmission using backward propagation. *Optics Express*, 16(20):16124–16137, 2008.

[68] Antonio Mecozzi, C Balslev Clausen, and Mark Shtaif. Analysis of intrachannel nonlinear effects in highly dispersed optical pulse transmission. *IEEE Photonics Technology Letters*, 12(4):392–394, 2000.

[69] Marina M Melek and David Yevick. Machine learning two stage optical fiber nonlinearity mitigation. *Journal of Modern Optics*, 67(12):1072–1077, 2020.

[70] Marina M Melek and David Yevick. Nonlinearity mitigation with a perturbation based neural network receiver. *Optical and Quantum Electronics*, 52(10):1–10, 2020.

[71] Marina M Melek and David Yevick. Fiber nonlinearity mitigation with a perturbation based siamese neural network receiver. *Optical Fiber Technology*, 66:102641, 2021.

[72] Marina M Melek and David Yevick. Machine learning compensation of fiber nonlinear noise. *in review,Journal of Optics Communications*, 2021.

[73] CR Menyuk. Application of multiple-length-scale methods to the study of optical fiber transmission. *Journal of Engineering Mathematics*, 36(1):113–136, 1999.

[74] David S Millar, Toshiaki Koike-Akino, Sercan Ö Arık, Keisuke Kojima, Kieran Parsons, Tsuyoshi Yoshida, and Takashi Sugihara. High-dimensional modulation for coherent optical communications systems. *Optics express*, 22(7):8798–8812, 2014.

[75] Terenuma Miya, Y Terunuma, Tatsuya Hosaka, and Tadakazu Miyashita. Ultimate low-loss single-mode fibre at 1.55 μm. *Electronics Letters*, 15(4):106–108, 1979.

[76] Moshe Nazarathy, Jacob Khurgin, Rakefet Weidenfeld, Yehuda Meiman, Pak Cho, Reinhold Noe, Isaac Shpantzer, and Vadim Karagodsky. Phased-array cancellation of nonlinear fwm in coherent ofdm dispersive multi-span links. *Optics express*, 16(20):15777–15810, 2008.

[77] Shotaro Owaki, Yuta Fukumoto, Takahide Sakamoto, Naokatsu Yamamoto, and Moriya Nakamura. Experimental demonstration of spm compensation based on digital signal processing using a three-layer neural-network for 40-gbit/s optical 16qam signal. *IEICE Communications Express*, 7(1):13–18, 2018.

[78] Tomofumi Oyama, Hisao Nakashima, Shoichiro Oda, Tomohiro Yamauchi, Zhenning Tao, Takeshi Hoshida, and Jens C Rasmussen. Robust and efficient receiver-side compensation method for intra-channel nonlinear effects. In *OFC 2014*, pages 1–3. IEEE, 2014.

[79] Kumar V Peddanarappagari and Maıté Brandt-Pearce. Volterra series transfer function of single-mode fibers. *Journal of lightwave technology*, 15(12):2232–2241, 1997.

[80] John G Proakis and M Salehi. Digital communications, mcgraw-hill. *Inc., New York,*, 1995.

[81] Danish Rafique and Andrew D Ellis. Impact of signal-ase four-wave mixing on the effectiveness of digital back-propagation in 112 gb/s pm-qpsk systems. *Optics express*, 19(4):3449–3454, 2011.

[82] J. Saijonmaa and D. Yevick. Beam-propagation analysis of loss in bent optical waveguides and fibers. *J. Opt. Soc. Am.*, 73(12):1785–1791, Dec 1983.

[83] M Salsi, H Mardoyan, P Tran, C Koebele, E Dutisseuil, G Charlet, and S Bigo. 155× 100gbit/s coherent pdm-qpsk transmission over 7,200 km. In *2009 35th European Conference on Optical Communication*, volume 2009, pages 1–2. IEEE, 2009.

[84] Martin Schetzen. The volterra and wiener theories of nonlinear systems. 1980.

[85] A Shabat and V Zakharov. Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Soviet physics JETP*, 34(1):62, 1972.

[86] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[87] Durga L Shrestha and Dimitri P Solomatine. Experiments with adaboost. rt, an improved boosting scheme for regression. *Neural computation*, 18(7):1678–1710, 2006.

[88] Saroja V. Siddamal, R. M. Banakar, and B. C. Jinaga. Split step method in the analysis and modeling of optical fiber communication system. In Srija Unnikrishnan, Sunil Surve, and Deepak Bhoir, editors, *Advances in Computing, Communication and Control*, pages 254–261, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[89] Oleg Sidelnikov, Alexey Redyuk, and Stylianos Sygletos. Equalization performance and complexity analysis of dynamic deep neural networks in long haul transmission systems. *optics express*, 26(25):32765–32776, 2018.

[90] Benjamin P Smith and Frank R Kschischang. A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications. *Journal of Lightwave Technology*, 30(13):2047–2053, 2012.

[91] Jeany Son, Ilchae Jung, Kayoung Park, and Bohyung Han. Tracking-by-segmentation with online gradient boosting decision tree. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3056–3064, 2015.

[92] Jiang Su and Harry Zhang. A fast decision tree learning algorithm. In *Aaai*, volume 6, pages 500–505, 2006.

[93] Chang Sun, Eurika Kaiser, Steven L Brunton, and J Nathan Kutz. Deep reinforcement learning for optical systems: A case study of mode-locked lasers. *Machine Learning: Science and Technology*, 1(4):045013, 2020.

[94] Bogdan Szafraniec, Todd S Marshall, and Bernd Nebendahl. Performance monitoring and measurement techniques for coherent optical systems. *Journal of Lightwave Technology*, 31(4):648–663, 2012.

[95] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.

[96] Ming Chieng Tan, Faisal Nadeem Khan, Waled Hussein Al-Arashi, Yudi Zhou, and Alan Pak Tao Lau. Simultaneous optical performance monitoring and modulation format/bit-rate identification using principal component analysis. *Journal of Optical Communications and Networking*, 6(5):441–448, 2014.

[97] Takahito Tanimura, Takeshi Hoshida, Jens C Rasmussen, Makoto Suzuki, and Hiroyuki Morikawa. Osnr monitoring by deep neural networks trained with asynchronously sampled data. In *2016 21st OptoElectronics and Communications Conference (OECC) held jointly with 2016 International Conference on Photonics in Switching (PS)*, pages 1–3. IEEE, 2016.

[98] Zhenning Tao, Liang Dou, Weizhen Yan, Lei Li, Takeshi Hoshida, and Jens C Rasmussen. Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate. *Journal of Lightwave Technology*, 29(17):2570–2576, 2011.

[99] Roland Thiolliere, Ewan Dunbar, Gabriel Synnaeve, Maarten Versteegh, and Emmanuel Dupoux. A hybrid dynamic time warping-deep neural network architecture for unsupervised acoustic modeling. In *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[100] Jakob Thrane, Jesper Wass, Molly Piels, Julio CM Diniz, Rasmus Jones, and Darko Zibar. Machine learning techniques for optical performance monitoring from directly detected pdm-qam signals. *Journal of Lightwave Technology*, 35(4):868–875, 2016.

[101] Jhon J Granada Torres, Andrea Chiuchiarelli, Varghese A Thomas, Stephen E Ralph, Ana M Cárdenas Soto, and Neil Guerrero González. Adaptive nonsymmetrical demodulation based on machine learning to mitigate time-varying impairments. In *2016 IEEE Avionics and Vehicle Fiber-Optics and Photonics Conference (AVFOP)*, pages 289–290. IEEE, 2016.

[102] Lucas Lo Vercio, Kimberly Amador, Jordan J Bannister, Sebastian Crites, Alejandro Gutierrez, Matthew Ethan MacDonald, Jasmine Moore, Pauline Mouches, Deepthi Rajasheka, Serena Schimert, et al. Supervised machine learning tools: a tutorial for clinicians. *Journal of Neural Engineering*, 2020.

[103] S Walker. Rapid modeling and estimation of total spectral loss in optical fibers. *Journal of lightwave technology*, 4(8):1125–1131, 1986.

[104] Danshi Wang, Min Zhang, Meixia Fu, Zhongle Cai, Ze Li, Yue Cui, and Bin Luo. Knn-based detector for coherent optical systems in presence of nonlinear phase noise. In *2016 21st OptoElectronics and Communications Conference (OECC) held jointly with 2016 International Conference on Photonics in Switching (PS)*, pages 1–3. IEEE, 2016.

[105] Danshi Wang, Min Zhang, Ze Li, Yue Cui, Jingdan Liu, Yang Yang, and Hongxiang Wang. Nonlinear decision boundary created by a machine learning-based classifier to mitigate nonlinear phase noise. In *2015 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2015.

[106] Feng Wang, Guijun Hu, and Zhaoxi Li. A novel four dimensional constellation shaping with non-uniform signaling for long-haul fiber-optic communication. *Optics Communications*, 486:126755, 2021.

[107] Sebastian J Wetzel, Roger G Melko, Joseph Scott, Maysum Panju, and Vijay Ganesh. Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks. *Physical Review Research*, 2(3):033499, 2020.

[108] Xiaoxia Wu, Jeffrey A Jargon, Ronald A Skoog, Loukas Paraschis, and Alan E Willner. Applications of artificial neural networks in optical performance monitoring. *Journal of Lightwave Technology*, 27(16):3580–3589, 2009.

[109] Metodi P Yankov, Francesco Da Ros, Edson P da Silva, Søren Forchhammer, Knud J Larsen, Leif K Oxenløwe, Michael Galili, and Darko Zibar. Constellation shaping for wdm systems using 256qam/1024qam with probabilistic optimization. *Journal of Lightwave Technology*, 34(22):5146–5156, 2016.

[110] Metodi P Yankov, Uiara Celine de Moura, and Francesco Da Ros. Power evolution prediction and optimization in a multi-span system based on component-wise system modeling. *arXiv preprint arXiv:2009.05348*, 2020.

[111] Amnon Yariv, Dan Fekete, and David M Pepper. Compensation for channel dispersion by nonlinear optical phase conjugation. *Optics Letters*, 4(2):52–54, 1979.

[112] Jianjun Yu, Xiang Zhou, Ming-Fang Huang, Yin Shao, Dayou Qian, Ting Wang, Milorad Cvijetic, Peter Magill, Lynn Nelson, M Birk, et al. 17 tb/s ($161\times$ 114 gb/s) polmux-rz-8psk transmission over 662 km of ultra-low loss fiber using c-band edfa amplification and digital coherent detection. In *2008 34th European Conference on Optical Communication*, pages 1–2. IEEE, 2008.

[113] Xiang Zhou, Jianjun Yu, Ming-Fang Huang, Yin Shao, Ting Wang, Peter Magill, Milorad Cvijetic, Lynn Nelson, Martin Birk, Guodong Zhang, et al. 32tb/s ($320\times$ 114gb/s) pdm-rz-8qam transmission over 580km of smf-28 ultra-low-loss fiber. In *2009 Conference on Optical Fiber Communication*, pages 1–3. IEEE, 2009.

[114] Xiang Zhou, Jianjun Yu, Ming-Fang Huang, Yin Shao, Ting Wang, Lynn Nelson, Peter Magill, Martin Birk, Peter I Borel, David W Peckham, et al. 64-tb/s ($640\times$ 107-gb/s) pdm-36qam transmission over 320km using both pre-and post-transmission digital equalization. In *National Fiber Optic Engineers Conference*, page PDPB9. Optical Society of America, 2010.

[115] Qunbi Zhuge, Michael Reimer, Andrzej Borowiec, Maurice O'Sullivan, and David V Plant. Aggressive quantization on perturbation coefficients for nonlinear pre-distortion. In *OFC 2014*, pages 1–3. IEEE, 2014.

[116] Darko Zibar, Luis Henrique Hecker de Carvalho, Molly Piels, Andy Doberstein, Julio Diniz, Bernd Nebendahl, Carolina Franciscangelis, Jose Estaran, Hansjoerg Haisch, Neil G Gonzalez, et al. Application of machine learning techniques for amplitude and phase noise characterization. *Journal of Lightwave Technology*, 33(7):1333–1343, 2015.

[117] Darko Zibar, Molly Piels, Rasmus Jones, and Christian G Schäeffer. Machine learning techniques in optical communication. *Journal of Lightwave Technology*, 34(6):1442–1452, 2015.

APPENDICES

Appendix A

System and Link parameters

Parameters	Values
Symbol Rate	32 GB
Modulation Format	16QAM
Number of Polarization Waves	2
Pulse Shape	Root raised cosine
Roll-off factor	0.01

Table A.1: TRANSMISSION PARAMETERS OF THE SYSTEM MODEL

Parameters	Values
Fiber Attenuation (α)	0.2 dB/km
Fiber Dispersion (D)	17 ps/(km.nm)
Fiber Nonlinearity (γ)	1.4 $W^{-1}km^{-1}$
EFDA Noise Figure (NF)	4.5 dB
Simulation Wavelength (λ)	1550 nm
Span Length	80 km
Link Length	3200 km

Table A.2: LINK PARAMETERS OF THE SYSTEM MODEL