

DESIGN AND OPTIMIZATION OF DSP
TECHNIQUES FOR THE MITIGATION OF
LINEAR AND NONLINEAR IMPAIRMENTS IN
FIBER-OPTIC COMMUNICATION SYSTEMS

DESIGN AND OPTIMIZATION OF DIGITAL SIGNAL
PROCESSING TECHNIQUES FOR THE MITIGATION OF
LINEAR AND NONLINEAR IMPAIRMENTS IN FIBER-OPTIC
COMMUNICATION SYSTEMS

BY

MAHMOUD M. T. MAGHRABI, M.Sc., B.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

© Copyright by Mahmoud M. T. Maghrabi, January 2021

All Rights Reserved

DOCTOR OF PHILOSOPHY (2021)
(Electrical & Computer Engineering)

McMaster University
Hamilton, Ontario, Canada

TITLE: Design and Optimization of Digital Signal Processing
Techniques for the Mitigation of Linear and Nonlinear
Impairments in Fiber-Optic Communication Systems

AUTHOR: Mahmoud M. T. Maghrabi
M.Sc.
Department of Engineering Mathematics
(Faculty of Engineering, Cairo University, Giza, Egypt)
B.Sc.
Department of Electronics and Electrical Communica-
tions Engineering
(Faculty of Engineering, Cairo University, Giza, Egypt)

SUPERVISOR: Prof. Mohamed H. Bakr
Prof. Shiva Kumar

NUMBER OF PAGES: xxvii, 172

Lay Abstract

This thesis proposes two powerful and computationally efficient digital signal processing (DSP)-based techniques, namely, artificial neural network nonlinear feed forward equalizer (ANN-NFFE) and adaptive digital back propagation (A-DBP) equalizer, for mitigating the induced distortions in short-reach and long-haul fiber-optic communication systems, respectively. The ANN-NFFE combats nonlinear impairments of direct-detected short-haul optical fiber communication systems, achieving compensation performance comparable to the benchmark performance obtained using maximum-likelihood sequence estimator with much lower computational cost. A novel adjoint sensitivity analysis (ASA) approach is proposed to significantly accelerate the sensitivity analyses of fiber-optic design problems. The A-DBP exploits a gradient-based optimization method coupled with the ASA algorithm to blindly compensate for the distortions of coherent-detected fiber-optic communication systems and networks, utilizing the minimum possible overhead of performed system simulations. The robustness and efficiency of the proposed equalizers are demonstrated using numerical simulations of varied examples extracted from practical optical fiber communication systems scenarios.

Abstract

Optical fibers play a vital role in modern telecommunication systems and networks. An optical fiber link imposes some linear and nonlinear distortions on the propagating light-wave signal due to the inherent dispersive nature and nonlinear behavior of the fiber. These distortions impede the increasing demand for higher data rate transmission over longer distances. Developing efficient and computationally non-expensive digital signal processing (DSP) techniques to effectively compensate for the fiber impairments is therefore essential and of preeminent importance. This thesis proposes two DSP-based approaches for mitigating the induced distortions in short-reach and long-haul fiber-optic communication systems.

The first approach introduces a powerful digital nonlinear feed-forward equalizer (NFFE), exploiting multilayer artificial neural network (ANN). The proposed ANN-NFFE mitigates nonlinear impairments of short-haul optical fiber communication systems, arising due to the nonlinearity introduced by direct photo-detection. In a direct detection system, the detection process is nonlinear due to the fact that the photo-current is proportional to the absolute square of the electric field intensity. The proposed equalizer provides the most efficient computational cost with high

equalization performance. Its performance is comparable to the benchmark compensation performance achieved by maximum-likelihood sequence estimator. The equalizer trains an ANN to act as a nonlinear filter whose impulse response removes the intersymbol interference (ISI) distortions of the optical channel. Owing to the proposed extensive training of the equalizer, it achieves the ultimate performance limit of any feed-forward equalizer. The performance and efficiency of the equalizer are investigated by applying it to various practical short-reach fiber-optic transmission system scenarios. These scenarios are extracted from practical metro/media access networks and data center applications. The obtained results show that the ANN-NFFE compensates for the received BER degradation and significantly increases the tolerance to the chromatic dispersion distortion.

The second approach is devoted for blindly combating impairments of long-haul fiber-optic systems and networks. A novel adjoint sensitivity analysis (ASA) approach for the nonlinear Schrödinger equation (NLSE) is proposed. The NLSE describes the light-wave propagation in optical fiber communication systems. The proposed ASA approach significantly accelerates the sensitivity calculations in any fiber-optic design problem. Using only one extra adjoint system simulation, all the sensitivities of a general objective function with respect to all fiber design parameters are estimated. We provide a full description of the solution to the derived adjoint problem. The accuracy and efficiency of our proposed algorithm are investigated through a comparison with the accurate but computationally expensive central finite-differences (CFD) approach. Numerical simulation results show that the proposed ASA algorithm has the same accuracy as the CFD approach but with a much lower computational cost.

Moreover, we propose an efficient, robust, and accelerated adaptive digital back

propagation (A-DBP) method based on adjoint optimization technique. Provided that the total transmission distance is known, the proposed A-DBP algorithm blindly compensates for the linear and nonlinear distortions of point-to-point long-reach optical fiber transmission systems or multi-point optical fiber transmission networks, without knowing the launch power and channel parameters. The NLSE-based ASA approach is extended for the sensitivity analysis of general multi-span DBP model. A modified split-step Fourier scheme method is introduced to solve the adjoint problem, and a complete analysis of its computational complexity is studied. An adjoint-based optimization (ABO) technique is introduced to significantly accelerate the parameters extraction of the A-DBP. The ABO algorithm utilizes a sequential quadratic programming (SQP) technique coupled with the extended ASA algorithm to rapidly solve the A-DBP training problem and optimize the design parameters using minimum overhead of extra system simulations. Regardless of the number of A-DBP design parameters, the derivatives of the training objective function with respect to all parameters are estimated using only one extra adjoint system simulation per optimization iterate. This is contrasted with the traditional finite-difference (FD)-based optimization methods whose sensitivity analysis calculations cost per iterate scales linearly with the number of parameters.

The robustness, performance, and efficiency of the proposed A-DBP algorithm are demonstrated through applying it to mitigate the distortions of a 4-span optical fiber communication system scenario. Our results show that the proposed A-DBP achieves the optimal compensation performance obtained using an ideal fine-mesh DBP scheme utilizing the correct channel parameters. Compared to A-DBPs trained using SQP algorithms based on forward, backward, and central FD approaches, the proposed

ABO algorithm trains the A-DBP with 2.02 times faster than the backward/forward FD-based optimizers, and with 3.63 times faster than the more accurate CFD-based optimizer. The achieved gain further increases as the number of design parameters increases. A coarse-mesh A-DBP with less number of spans is also adopted to significantly reduce the computational complexity, achieving compensation performance higher than that obtained using the coarse-mesh DBP with full number of spans.

To my wife, Marwa, and my little boy, Hamza

To the memory of my parents to whom I am indebted for as long as I live

To the memory of my little sisters, Mai & Manar, and my aunt, Somaia

To my beloved family and friends

Acknowledgements

All praise is due to Allah, the Almighty, for his uncountable blessings, among which is giving me the strength to complete this research work. I hope that he accepts my sincere intention of making this work useful and helpful for others.

Words cannot express my deep gratitude to my supervisors, Dr. Mohamed Bakr and Dr. Shiva Kumar. My sincere thanks and appreciation to the invaluable and endless support I received from them during the course of my work. I am really grateful for their time, innovative and insightful ideas, long discussions, and funding throughout the development of this work. Definitely, this work would have never been accomplished without their continuous guidance, enthusiasm, optimism, motivation, encouragement, and patience. Honestly, Dr. Bakr and Dr. Kumar made every effort to ensure maintaining a perfect, peaceful, quite, stress-free, and comfortable environment during all times of my Ph.D. pursuit. Moreover, they showed all aspects of understanding, care, and humanity in the social side of my life. It was really my fortunate, honor, and privilege to work under supervision of such great, dedicated, and distinguished professors. Dr. Bakr and Dr. Kumar, thanks for everything you did for me.

I would like to express my sincere thanks and appreciation to Dr. Jamal Deen and Dr. Xun Li, the members of my supervisory committee, for their precious comments,

suggestions, interesting ideas, valuable feedbacks, and useful scientific discussions during my research work. I would also like to thank my external examiner, Dr. John Cartledge with Queens University, for his constructive suggestions and insightful review of the dissertation. I am also grateful to Dr. James Reilly and Dr. Jiankang Zhang, the course instructors who taught me during my Ph.D. program, for their wonderful lectures. Rest in peace Dr. Zhang.

I sincerely acknowledge Ahmed Elsharabasy, Mohamed E. Fouda with University of California, Islam S. Hassan with Concordia University, and Omar A. El-Gendy with Purdue University, for their continuous encouragement, assistance, support, interesting ideas, and fruitful scientific discussions.

Moreover, for their faithful friendship and encouragement, it is a pleasure to thank my friends and colleagues at the Department of Electrical and Computer Engineering, Ehab Sayed, Ayman Negm, Mohamed Fathy, Laleh Seyyed-Kalantari, Rishad Arfin, Neda Khiabani, Yu Zhang, Xiaojun Liang, Jing Shao, Shovasis Biswas, Mahdi Naghshvarianjahromi, Elham Bidaki, Hytham Afifi, Mostafa Medra, Ahmed Darwish, Mohamed Hamouda, Khaled Ahmed, Ahmed Morra, Khaqan Majeed, Karim Osama, and Abdallah Shawky.

I would like also to thank the administrative team at the Department of Electrical and Computer Engineering, McMaster University, Cheryl Gies in particular, for the constant support and for providing all the convenience and assistance I have needed.

Special gratitude is dedicated to my beloved wife, Marwa Negm, who came to my life and brought lots of happiness. During the last two years and half of my Ph.D., it was impossible to accomplish this journey without her great support, patience, and compassion. My dear son, Hamza, thanks for all joy and happiness you brought to

our life. Your cute smile was and will always be the secret that makes hard times easily fly.

I am deeply grateful and indebted to my beloved parents who dedicated their lifetime to ensure that my sisters and I have an easy and comfortable life, and acquire the best education. I wish they were with me right now, witnessing the completion of this chapter in my life. However, I am pretty sure that they are now happy with my graduation as they always used to be with any achievement I make in my life. My mom and dad, making you happy and proud was and will always be the main motivation and inspiration that push me to work hard, pursuing academic excellence.

Last but not least, I would like to express my profound gratitude and appreciation to my family whose patience, continuous encouragement, and overseas support enabled me to complete this work. Special thanks to my uncle Maghrabi Taha, uncle Mohamed Khalil, my aunt Nadia Osman, my mother-in-law, Hanan Said, my cousin, Khaled Maghrabi, and my cousin, Abeer Mahmoud, for the incredible and continuous moral support I received from them throughout my Ph.D. journey.

Contents

Lay Abstract	iii
Abstract	iv
Acknowledgements	ix
Abbreviations	xxiii
1 Introduction	1
1.1 Evolution of Fiber-Optic Communication systems	2
1.2 Impairments in an Optical Fiber Channel	9
1.3 Electrical Compensation Techniques for Optical Fiber Distortions . .	15
1.4 Thesis Contribution	34
2 Dispersion Compensation of Direct Detected Optical Fiber Commu- nication Systems Using Artificial Neural Networks (ANNs)	41
2.1 Optical Communication System Model	44
2.2 ANN Equalizer	48
2.3 Results	55
2.4 Conclusion	68

3	An Adjoint Sensitivity Analysis Approach for the Nonlinear Schrödinger Equation	69
3.1	Matrix Representation of the NLSE	70
3.2	ASA for the NLSE	72
3.3	Adjoint Problem Solution	79
3.4	Results	83
3.5	Conclusion	93
4	Adaptive Digital Back Propagation Exploiting Adjoint-Based Optimization	95
4.1	Problem Formulation of a DBP Model	96
4.2	ASA for the Multi-Span DBP Model	106
4.3	Adjoint-Based Optimization Algorithm	119
4.4	Results	127
4.5	Conclusion	144
5	Conclusions and Future Work	146
5.1	Future work	151

List of Figures

1.1	Basic structure of a fiber-optic communication system (Senior and Jamro, 2009).	3
1.2	An illustration of pulse broadening due to chromatic dispersion of the fiber (Kalander and WANG, 2017).	10
1.3	An instance of signal degradation due to the chromatic dispersion of fiber. The transmitted data are 10 return-to-zero Gaussian pulses. The modulation format is on-off-keying, the data rate is 10 Gbps, and the fiber length is 140 km.	11
1.4	An illustration of nonlinear interactions in a single channel and WDM fiber-optic communication system (Kumar <i>et al.</i> , 2018).	13
1.5	Block diagram of a M -taps linear feed-forward equalizer (Momtaz and Green, 2010). T stands for a one bit time interval delay, C_i , $i = 1, 2, \dots, M$, are adjustable tap-weights, and M is the total number of taps. The signals $x(n)$ and $y(n)$ are the un-equalized input and equalized output at sampling time nT_b , respectively, where T_b is the inverse baud rate.	17
1.6	General architecture of a decision feedback equalizer.	19

2.1	Model of short-reach fiber optic communication system with direct photo-detection. DML: Directly modulated laser, EML: Electrically modulated laser, ADC: Analog to digital converter, BPF: Band pass filter, LPF: Low pass filter.	44
2.2	Basic configuration of the introduced ANN-NFFE.	49
2.3	Schematic for the generation of the complete set training data. K is the number of interfering symbols and $n = 2K + 1$ is the required number of equalizer taps. The arrows refer to the actual training data, used during the ANN weights adjustment, after discarding the redundant guard data.	52
2.4	The value of training objective function $f(\mathbf{W})$ versus the number of optimization iterates. The transmitted data are NRZ-OOK with raised-cosine pulse shaping. The data rate is 10 Gbps and the fiber length is 140 km.	59
2.5	Received BER, with and without the ANN-NFFE, versus OSNR. The transmitted data are NRZ-OOK with raised-cosine pulse shaping. The data rate is 10 Gbps and the fiber length is 140 km.	60
2.6	BER of the received equalized data versus OSNR. The transmitted data rate is 10 Gbps, and the modulation format is NRZ-OOK. Various transmission distances are considered and compared to the B2B case, when the ANN-NFFE is used for the equalization of the received signal. 61	
2.7	The required OSNR at received BER= 1×10^{-3} versus fiber optic length. A NRZ-OOK modulation format with raised-cosine pulse shaping is assumed.	62

2.8	BER of the received equalized data versus OSNR. The transmitted data rate = 10 Gbps, and the modulation format is RZ-OOK with 50% duty cycle. Various transmission distances are considered and compared to the B2B case, when the ANN-NFFE is used for the equalization process.	63
2.9	The required OSNR at received BER= 1×10^{-3} versus fiber optic length. The performance of the system with zero frequency chirping without equalization case is compared to the cases of frequency chirping with and without equalization. The data rate is 10 Gbps and the modulation format is RZ-OOK with 50% duty cycle. The parameters of the ANN-NFFE are $n = 5$, $m = 5$ at $L \leq 40$ km, and $n = 7$ and $m = 6$ at longer distances.	65
2.10	Received BER, with and without the ANN-NFFE, versus OSNR in case of transmission distance (a) 10 km, (b) 15 km, and (c) 20 km. The modulation format is NRZ-OOK with raised-cosine pulse shaping and the data rate is 28 Gbps.	66
2.11	The required OSNR at received BER= 1×10^{-3} versus fiber optic length. The data rate is 28 Gbps. A NRZ-OOK modulation format with raised-cosine pulse shaping is assumed. The parameters of the ANN-NFFE are $n = 5$, $m = 5$ at $L \leq 15$ km, and $n = 7$, $m = 6$ at longer distances.	67

3.1 The normalized ASA sensitivities of objective function (3.29) with respect to the parameters: $\beta_1, \beta_2, \beta_3$ and γ , for a sweep of L , as compared to the more computationally expensive CFD approach. The nominal fiber parameters are as given in Table 3.1. The modulation format is OOK with Gaussian pulse shaping, the baud rate is 10 Gbps, and the input pulse peak power is $P_0 = 2$ dBm. 85

3.2 The normalized ASA sensitivities of objective function (3.29) with respect to the parameters: L and P_0 , for a sweep of L , as compared to the more computationally expensive CFD approach. The nominal fiber parameters are as given in Table 3.1. The modulation format is OOK with Gaussian pulse shaping, the baud rate is 10 Gbps, and the input pulse peak power is $P_0 = 2$ dBm. 86

3.3 The normalized ASA sensitivities of objective function (3.29) with respect to the parameters: $\beta_1, \beta_2, \beta_3$ and α , for a sweep of L , as compared to the more computationally expensive CFD approach. The fiber parameters are as given in Table 3.3. The modulation format is 16 QAM with a root raised cosine pulse shaping, the baud rate is 28 Gbaud, and the average launch power is $P_{av} = 6$ dBm. 88

3.4 The normalized ASA sensitivities of objective function (3.29) with respect to the parameters: γ, P_{av} and L , for a sweep of L , as compared to the more computationally expensive CFD approach. The fiber parameters are as given in Table 3.3. The modulation format is 16 QAM with a root raised cosine pulse shaping, the baud rate is 28 Gbaud, and the average launch power is $P_{av} = 6$ dBm. 89

3.5	The normalized ASA sensitivities of objective function (3.31) with respect to the parameters: β_2, γ, L and P_0 , for a sweep of P_0 , as compared to the more computationally expensive CFD approach. The fiber parameters are as given in Table 3.4.	92
3.6	The normalized optical power of the transmitted and received pulses. The peak power of the launch pulse is: (a) 15 mW, i.e., less than the required peak power to form a soliton P_{sol} , and (b) 30 mW, i.e., greater than P_{sol}	93
4.1	The propagation in a single-span optical transmission fiber (forward propagation), and a single-span virtual DBP fiber (backward propagation). Tx: transmitter, Rx: receiver, TF: transmission fiber, DBP: digital back propagation.	100
4.2	Model of propagation in an M -span optical fiber transmission system. Tx: transmitter, Rx: receiver, FS: fiber span, Amp, amplifier.	101
4.3	Block diagram of a digital back propagation model for a M -span fiber-optic communication system. Rx: receiver, VFS: virtual fiber span, LE: loss element, DBP: digital back propagation.	101
4.4	Propagation model of an original DBP simulation, and its corresponding adjoint DBP simulation. Rx: receiver; DBP: digital back propagation; VFS: virtual fiber span; LE: loss element; AFS: adjoint fiber span.	111

4.5	Flow diagram of the modified symmetric SSFS algorithm used for solving the adjoint DBP problem.	113
4.6	Flow diagram of the proposed adjoint-based optimization algorithm.	125
4.7	Block diagram of the fiber-optic communication system considered in the simulation. Tx: transmitter; BPF: band pass filter; Rx: receiver; A/D: analog to digital converter; CDC: chromatic dispersion compensation; A-DBP: adaptive digital back propagation; MF: matched filter.	128
4.8	The value of the training objective function F in (4.52) versus the number of optimization iterates for the 4-span fiber-optic communication system scenario. Four optimization algorithms are considered for the training of the A-DBP, namely, the ABO algorithm, the SQP-CFD algorithm, SQP-FFD algorithm, and SQP-BFD algorithm.	132
4.9	BER versus average launch power for the 4-span fiber-optic communication system scenario. The equalization at the receiver is performed using the CDC unit only, the ideal DBP, the DBP, or the A-DBP. The number of transmitted symbols is 2^{17}	134
4.10	(a) Asymptotic number of conventional SSFS simulations required to train a A-DBP versus number of design parameters N ; (b) Asymptotic values of the simulation saving factor η achieved by the ABO algorithm versus design parameters number N . The ABO algorithm is compared to the SQP-CFD, SQP-FFD, and SQP-BFD. The number of training samples is 8192.	137

4.11	BER versus average launch power for the 4–span fiber-optic communication system scenario. The equalization at the receiver is performed using ideal DBP with fine-mesh and coarse-mesh models. The signal and system parameters are the same as that used for Fig. 4.9.	139
4.12	Total number of required SSFS steps versus average launch power for the 4–span fiber-optic communication system scenario. The equalization at the receiver is performed using ideal DBP with fine-mesh and coarse-mesh models. The number of transmitted symbols is 2^{17}	140
4.13	(a) BER, and (b) Quality factor versus average launch power for the 4-span fiber-optic communication system scenario. The equalization at the receiver is performed using the CDC unit only, the coarse-mesh ideal DBP, the coarse-mesh 2–virtual fiber spans A-DBP, the coarse-mesh 3–virtual fiber spans A-DBP, or the coarse-mesh 4–virtual fiber spans A-DBP. The number of transmitted symbols is 2^{17}	143
4.14	Total number of required SSFS steps versus average launch power for the 4-span fiber-optic communication system scenario. The equalization at the receiver is performed using the CDC unit only, the coarse-mesh ideal DBP, the coarse-mesh 2–virtual fiber spans A-DBP, the coarse-mesh 3–virtual fiber spans A-DBP, or the coarse-mesh 4–virtual fiber spans A-DBP. The number of transmitted symbols is 2^{17}	144

List of Tables

2.1	Simulation parameters	56
2.2	Mean photo-current versus OSNR	57
3.1	Simulation parameters of Example 1	84
3.2	The computational time for obtaining the sensitivity results using the CFD and the ASA approaches.	87
3.3	Simulation parameters of Example 2.	88
3.4	Fiber parameters used in Examples 3	91
4.1	Computational complexity of the conventional and modified symmetric split-step Fourier scheme (SSFS) algorithms.	118
4.2	Fiber simulation parameters.	128
4.3	Lengths of fiber spans used in the simulation of the 4–span fiber-optic communication system scenario.	129
4.4	Optimal design point solutions of the A-DBP in case of the 4-span fiber- optic communication system scenario. Four optimization algorithms are used to train the A-DBP, namely, the ABO algorithm, the SQP- CFD algorithm, the SQP-FFD algorithm, the SQP-BFD algorithm. The initial design point (in km) is $\mathbf{x}_0 = [85 \ 85 \ 85 \ 85]^T$	135

4.5	Computational cost of sensitivity calculations required to train the A-DBP for mitigating the distortions of the 4-span fiber-optic communication system scenario. The A-DBP is trained using ABO, SQP-CFD, SQP-FFD, and SQP-BFD algorithms.	136
4.6	Optimum values of the A-DBP design parameters for mitigating the 4-span fiber-optic communication system scenario. A coarse-mesh model is considered with a maximum allowed nonlinear phase change per SSFS step of 0.1 rad. The ABO is used to train A-DBP with 2-, 3-, and 4-virtual fiber spans.	141

Abbreviations

ABO	Adjoint-based optimization
A/D	Analog to digital
ADC	Analog to digital converter
A-DBP	Adaptive digital back propagation
AFS	Adjoint fiber span
ANN	Artificial neural network
ANN-NFFE	Artificial neural network nonlinear feed-forward equalizer
ASA	Adjoint sensitivity analysis
ASE	Amplified spontaneous emission
B2B	Back-to-back
BER	Bit error rate
BFD	Backward-finite differences
BFGS	Broyden–Fletcher–Goldfarb–Shanno

BPF	Bandpass filter
CD	Chromatic dispersion
CDC	Chromatic dispersion compensation
CFD	Central finite differences
C-RAN	Cloud radio access network
DBP	Digital back propagation
DCN	Data center network
DFE	Decision feedback equalizer
DML	Directly modulated laser
DSF	Dispersion-shifted fiber
DSP	Digital signal processing
EDFA	Erbium-doped fiber amplifier
EML	Externally modulated laser
EON	Elastic optical network
EVD	Eigenvalue decomposition
FD	Finite-difference
FDTD	Finite-difference time-domain
FFD	Forward-finite differences

FFE	Feed-forward equalizer
FFT	Fast Fourier transform
FWM	Four-wave mixing
IFWM	Intra-channel four-wave mixing
IM-DD	Intensity-modulation and direct detection
INLSE	Inverse nonlinear Schrödinger equation
ISI	Inter-symbol interference
IXPM	Intra-channel cross-phase modulation
KKT	Karush-Kuhn-Tucker
LAN	Local area network
LC	Linear compensation
LMS	Least-mean square
LPF	Low pass filter
MLSE	Maximum likelihood sequence estimator
MPSK	M -ray phase-shift keying
NLSE	Nonlinear Schrödinger equation
NN	neural network
NNE	Neural networks-based equalizer

NRZ	Non-return to zero
OEO	Optical-electrical-optical
OSNR	Optical signal to noise ratio
OOK	On-off keying
PAM	Pulse-amplitude modulation
QAM	Quadrature amplitude modulation
QP	Quadratic programing
QPSK	Quadrature phase-shift keying
RLS	Recursive least squares
Rx	Receiver
RZ	Return to zero
SMF	Single-mode fiber
SPM	Self-phase modulation
SQP	Sequential quadratic programing
SSFS	Split-step Fourier scheme
SSMF	Standard single-mode fiber
TLM	Transmission-line modeling
Tx	Transmitter

VA	Viterbi algorithm
VFS	Virtual fiber span
VSE	Volterra series-based equalizer
WDM	Wavelength-division multiplexing
XPM	Cross-phase modulation

Chapter 1

Introduction

A telecommunication system transmits information from one place to another, whether separated by a few kilometers or by transoceanic distances (Al-Amri *et al.*, 2016). It may but does not need to involve optics. The information is usually carried by an electromagnetic carrier wave whose frequency ranges between few kilohertz up to several hundred terahertz (Kaminow *et al.*, 2013). The traditional microwave wireless communication systems, for which the typical frequency is about 1 GHz, are utilized in many applications of our daily lives, including broadcast radio and television, wireless local area networks (LANs), and mobile phones (Weber, 2001). However, the data rates in most of the microwave systems are limited to few gigabits per second due to the lack of the available spectrum in the radio frequency range (Pozar, 2011). In contrast, an optical fiber communication system transmits data through light-waves, whose frequency carrier is very high- around 200 THz. It therefore has a much larger bandwidth and can support data rates up to 100 terabits per second and beyond (Malekiha, 2011).

Optical fibers are deployed in several telecommunication systems and applications.

According to the transmission distance, the fiber-optic communication systems can be classified into 2–main categories, namely, short-haul and long-haul optical fiber communication systems (Agrawal, 2012). While the transmission distance in short-haul optical communication systems ranges from a few meters to a few tens of kilometers, the long-haul systems can have transmission distances of thousands of kilometers. The initial deployment of fiber-optic was mainly for long-haul transmission system or fiber-optic submarine transmission (Liang, 2015). However, optical fibers are being considered recently in many promising short-reach applications. Metro and media access networks, data center networks (DCNs), and cloud radio access network (C-RANs) are instances among many other potential applications to short-haul fiber optic communication systems (Lam *et al.*, 2014; Andrews *et al.*, 2014).

1.1 Evolution of Fiber-Optic Communication systems

A basic fiber-optic communication system is shown in Fig. 1.1. It consists of a transmitter, followed by an optical fiber transmission channel, and then a receiver. The transmitter converts the electrical signal data into a lightwave modulated signal. The optical fiber carries the lightwave signal to the receiver. The receiver accepts the lightwave signal and converts it back into an electrical signal. The evolution of optical fiber communications is thus closely linked to the advent of technology breakthroughs in any of these 3–main components of the fiber-optic communication system (Yang, 2010; Shao, 2015).

The first breakthrough in optical fiber communications dates back to the laser

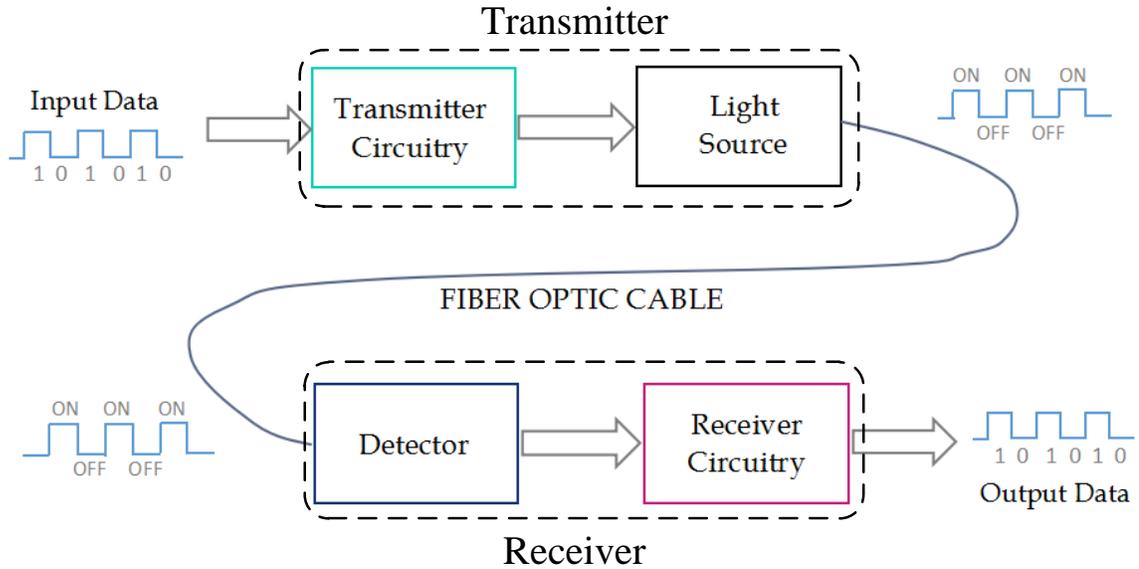


Figure 1.1: Basic structure of a fiber-optic communication system (Senior and Jamro, 2009).

invention in 1960, providing the coherent optical source required for transmitting data through lightwaves (Maiman, 1960). In 1966, the idea of using an optical fiber medium as a potential candidate for lightwave transmission was proposed, due to its capability of guiding the light in a manner similar to the guiding of electrons in a copper wire (Kao and Hockham, 1966). However, all studies around this idea remained theoretical during the 1960s. The main limitation at that time was the fact that all realized optical fibers were suffering from extremely high losses (over 1000 dB/km). Nevertheless, in the early 1970s, the fiber losses were brought down to less than 20 dB/km at a wavelength near 1 μm , by removing the fiber impurities using a novel fabrication technique (Keck *et al.*, 1973). Soon after, the low-loss fiber was realized in 1979, at the operating wavelength of 1.55 μm , where fiber losses could be significantly reduced to around 0.2 dB/km (Miya *et al.*, 1979).

The simultaneous feasibility of stable optical sources and low-loss optical fibers

led to an extensive research efforts and rapid development of optical fiber communication systems. According to a fundamental change in the theoretical limit of the lightwave system capacity (operating data rate) and/or the feasible transmission distance reach (the distance over which a lightwave system can transmit data without introducing errors), the progress of fiber-optic communications can be classified into five development generations (Agrawal, 2005; Kumar and Deen, 2014).

The first-generation has been investigated during 1970s, for which the GaAs semiconductor laser (Hayashi *et al.*, 1970), operating at a wavelength near $0.8 \mu\text{m}$, has been utilized as an optical source. The operating transmission rate was 45 Mb/s, and the repeater spacing was 10 km. Note that since signal is degraded during transmission, most lightwave systems require periodic regeneration of the optical signal through devices called repeaters. The repeater converts the optical signal into an electrical signal, processing that electrical signal and then regenerating an optical signal based on the processed electrical signal pattern. Although a repeater spacing of 10 km value may seem too small from a modem perspective, this was 10 times larger than the 1 km spacing required in coaxial systems at that time. The major limitation of the repeater spacing in this generation system comes from chromatic dispersion (CD) which causes significant pulse broadening and thus inter-symbol interference (ISI).

In the second-generation system, the repeater spacing could be increased considerably by shifting the operating wavelength of the lightwave system near to the $1.3 \mu\text{m}$ region. In this wavelength region, optical fibers exhibit minimum dispersion with fiber losses less than 0.5 dB/km. This realization led to a worldwide effort for the development of semiconductor lasers and detectors operating near the $1.3 \mu\text{m}$ region. In the early of 1980s, the first stable second-generation lightwave system

became available. However, the transmission data rate of this early system was limited by a threshold of 100 Mb/s, due to the intermodal dispersion of the multi-mode fibers (Green, 1996). In order to reduce the dispersion effect, a single mode-fiber (SMF) was then introduced (Yamada *et al.*, 1981). As opposed to multi-mode fibers, the SMF is designed to support the propagation of only one mode of light, exhibiting lower dispersion. This allows to have a higher system data rate and longer repeater distance. By 1987, advanced second-generation fiber-optic communication systems were commercially available. This generation system was characterized by operating at data rates of up to 1.7 Gb/s with a repeater spacing of about 50 km.

The main drawback of second-generation systems was the relatively high fiber losses (~ 0.5 dB/km) at the operating wavelength of $1.3 \mu\text{m}$, which limits the repeater spacing. It has been demonstrated with the realization of the low-loss fiber, whose typical loss is around 0.2 dB/km, that losses of silica fibers become minimum near $1.55 \mu\text{m}$ (Miya *et al.*, 1979). This fact motivated the development of a third-generation fiber-optic system with an operating wavelength of $1.55 \mu\text{m}$. However, the emergence of third-generation systems was considerably delayed because of the relatively large fiber dispersion in the wavelength region near $1.55 \mu\text{m}$. Moreover, the conventional InGaAsP semiconductor lasers could not be used as the coherent source of the system. This is due to the fact that the conventional InGaAsP semiconductor laser operating at $1.55 \mu\text{m}$ oscillates multiple longitudinal modes simultaneously, resulting in large pulse broadening. Many studies have been investigated during 1980s aiming to overcome these problems, either by using a dispersion-shifted fiber (DSF) designed to have minimum dispersion near the $1.55 \mu\text{m}$ region, or by limiting the laser spectrum to a single longitudinal mode. A dispersion-shifted single-mode silica

fiber was designed and fabricated in 1979 (Cohen *et al.*, 1979). The minimum chromatic dispersion of the DSFs, was shifted from 1.3 μm to the window of 1.55 μm , by controlling their waveguide dispersion and dopant-dependent material dispersion. It thus became feasible to have a transmission fiber with both of its low dispersion and low attenuation occurring at the 1.55 μm transmission window. In 1983, an InGaAsP/InP semiconductor laser with a narrow single-mode linewidth of 10 kHz was demonstrated (Wyatt and Devlin, 1983). By 1985, laboratory experiments indicated the possibility of transmitting information at bit rates of up to 4 Gb/s over distances in excess of 100 km (Gnauck *et al.*, 1985). Later in 1990, the DSFs in combination with distributed-feedback semiconductor lasers oscillating in a single longitudinal mode with narrow linewidths enabled the emergence of commercial third-generation fiber-optic communication systems (Agrawal, 2012). The operating transmission rate was 2.5 Gb/s, and the repeater spacing was up to 70 km. It has also been demonstrated in 1995 that such systems are capable of operating at a bit rate of up to 10 Gb/s (Agrawal, 2012).

The main drawback of third-generation systems was that the optical transmitted signal are regenerated periodically using optical-electrical-optical (OEO) repeaters, spaced apart typically by 60 – 70 km, because of the fiber losses. Such regenerating procedure in the electrical domain would not be effective for multichannel lightwave systems, as each single wavelength would require an OEO repeater, leading to excessive system complexity. Another limitation of using OEO repeaters was that due to the high data rate in optical fiber systems, high-speed electronic devices are therefore required, but it would be very hard and expensive to make extra-high speed electronics at that time. The idea that attracted the attention of many researchers during

1980s was introducing an optical amplification method to be utilized as an efficient alternative approach for the electrical amplification method. Optical amplifiers can amplify the signal directly in the optical domain without the need for optical-electrical or electrical-optical conversions. The first realization of optical amplification was in 1983, using semiconductor laser amplifiers (Mukai and Yamamoto, 1983). In 1986, the optical amplification was then realized using Raman amplifiers (Byron, 1986). Later in 1987, the most effective wideband optical amplification method was realized using erbium-doped fiber amplifiers (EDFAs) (Shaw and Digonnet, 1987). Once EDFAs have been commercially emerged, they widely replaced OEO repeaters due to their cost efficiency, especially for multichannel long-haul lightwave systems (Mears *et al.*, 1987; Desurvire *et al.*, 1987). The benefit of wideband optical amplification with low noise and high gain, provided by EDFAs, stimulated the development of transmitting signal through a single fiber channel using multiple carriers simultaneously, which can be implemented using a wavelength-division multiplexing (WDM) scheme (DeLange, 1970). In a WDM system, signals in different channels are modulated on separate wavelengths and simultaneously launch to the fiber by a multiplexer at the transmitter. The signals in different channels would then be demultiplexed at the receiver side. The WDM scheme therefore significantly increases the transmission capacity without the need for extra fiber channels.

The fourth-generation of fiber-optic communication systems made use of optical amplifiers and WDM for increasing the repeater spacing and the data rate, respectively. This technology breakthrough started a revolution and brought a new era to optical fiber communication systems. The system capacity could be doubled every 6 months reaching up to a bit rate of 10 Tb/s by 2001. Fiber losses could also be

compensated periodically using EDFAs, typically spaced apart by 60 to 80 km. The possibility of data transmission over 14,300 km at a bit rate of 5 Gb/s has been experimentally shown in 1991, using recirculating-loop configuration (Bergano *et al.*, 1991). By 1996, the transmission over 11,300 km at a bit rate of 5 Gb/s has been demonstrated using actual submarine cables (Otani *et al.*, 1995). Soon after, commercial transatlantic and transpacific cable systems became available, and submarine lightwave systems started to largely deploy worldwide.

To further increase the system capacity and the transmission reach, the fifth-generation of fiber-optic communication systems aims at extending the wavelength range over which a WDM system can operate simultaneously (Sharma *et al.*, 2013). Up to now, the modern WDM systems can handle hundreds of channels and they can expand a basic 10 Gb/s fiber-optic system to a total capacity of several Tb/s over a single fiber channel (Willner, 2019). Although WDM systems can greatly improve the capacity of optical fiber transmission systems through increasing the number of channels, the achievable feasible data rate is limited by the optical amplifier bandwidth and the induced nonlinear distortions of fiber. Optical amplifiers do not only compensate for fiber losses, but they also add a white Gaussian noise to the amplified signal. The main source of this noise is the amplified spontaneous emission (ASE) noise. While the amplifying gain of the EDFA is provided by the stimulated emission process, the ASE noise is due to the spontaneous emission process (Mears *et al.*, 1987). Increasing the transmitted signal power, i.e., the optical signal to noise ratio (OSNR), can limit the performance deterioration due to the noise. However, the transmitted signal power cannot be arbitrarily large, because of the fiber nonlinearity. In order to address the bandwidth and power limitation, the design of fifth-generation systems

need to simultaneously enhance the spectral efficiency and the power efficiency. The spectral efficiency can be increased using advanced spectrally efficient modulation format schemes. For instances M-ray phase-shift keying (MPSK) and quadrature amplitude modulation (QAM) schemes. The power efficiency can be improved by minimizing the required OSNR at a given level of bit error rate (BER). Therefore, the fifth-generation systems is also concerned with developing robust and powerful digital signal processing (DSP) and electrical equalization techniques to compensate for fiber distortions at the receiver side, which in turns minimizes the average launch power required for achieving a certain BER level (Willner, 2019). It should be clear that the more effective DSP techniques are available at the receiver side, the more feasible transmission distance reach can be achieved.

This thesis focuses on how to improve the power spectral efficiency of optical fiber systems. The focus will therefore be on efforts that have been made to develop reliable, stable, cost-efficient, and powerful digital equalization techniques that can be used to compensate for the linear and nonlinear distortions of the fiber. We first discuss the main sources of signal distortions in an optical fiber link in the next section. Then, a discussion of several methods developed for mitigating the fiber impairments is drawn in Section 1.3.

1.2 Impairments in an Optical Fiber Channel

While propagating through an optical fiber link, the transmitted signal is subjected to several sources of unavoidable distortions. These distortions are classified into linear and nonlinear impairments. The linear distortions are due to the loss, and chromatic dispersion (CD) of optical fibers, whereas the main source for nonlinear impairments

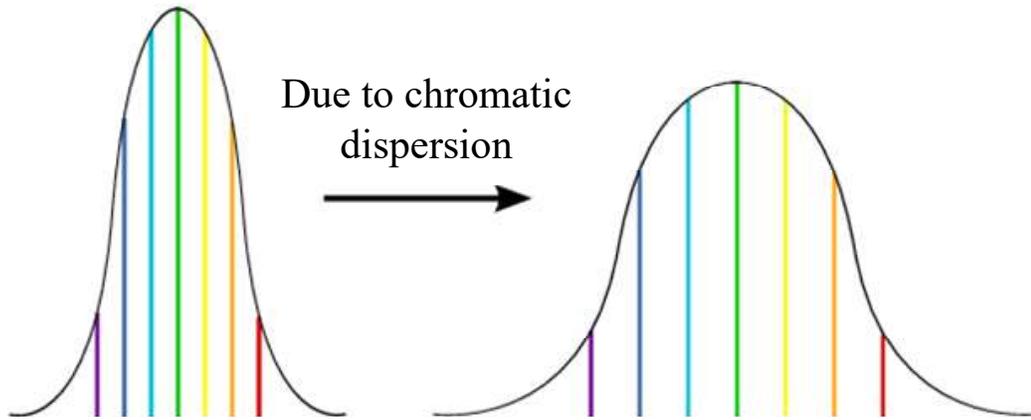


Figure 1.2: An illustration of pulse broadening due to chromatic dispersion of the fiber (Kalander and WANG, 2017).

is the optical Kerr effect (Kerr, 1875). Material absorptions, the Rayleigh scattering effect, and the splice effect are main reasons for the fiber loss (Ohashi *et al.*, 1992; Tsujikawa *et al.*, 2007; Pradhan *et al.*, 2003). Thanks to the realization of low-loss silica fiber, fiber losses can be made as low as 0.2 dB/km. However, signal attenuation becomes significant in long-haul transmission systems. In this case, fiber losses have to be compensated using EDFAs which though add ASE noise. The ASE noise in addition to other noises in the system (e.g. shot noise and thermal noise) represent an upper threshold to the overall system performance, setting the hard constraint of any practical fiber-optic communication system.

The fiber CD phenomenon is another major source for linear distortions in the optical communication system. Due to the dispersive nature of an optical fiber medium, its refractive index depends on frequency. As a result, each frequency component of an optical pulse propagates, into the fiber link, at different speed leading to pulse broadening, as illustrated in Fig. 1.2. When multiple pulses propagate together through

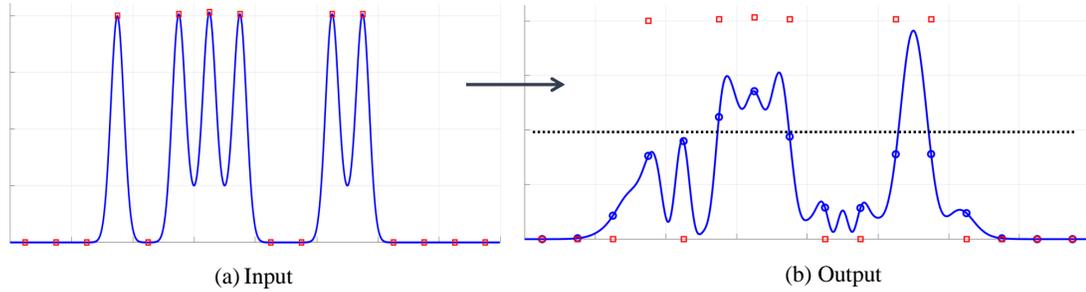


Figure 1.3: An instance of signal degradation due to the chromatic dispersion of fiber. The transmitted data are 10 return-to-zero Gaussian pulses. The modulation format is on-off-keying, the data rate is 10 Gbps, and the fiber length is 140 km.

the fiber, the pulse spreading causes each optical pulse to interfere with its neighbouring pulse, which is known as intersymbol interference (ISI) effect. As data rate or transmission distance increases, the ISI effect causes severe degradation to the received signal, resulting in bad data recovery and large bit error rate (BER) at the receiver. Figure 1.3 shows an instance of transmitting a sequence of 10-Gaussian pulses through a 140 km-optical fiber link, at a bit rate of 10 Gbps. Due to the CD effect, the received signal deteriorates significantly as compared to the transmitted signal, as declared in Fig. 1.3. Consider a half-level decision detector at the receiver, the recovered bit pattern is then ‘0011000000’ as opposed to the ‘1011100110’ original transmitted pattern, i.e., 4 bits out of 10 are recovered wrongly. The CD induced spectrum broadening, even without nonlinearity, is therefore very crucial for high data-rate transmission systems (> 2.5 Gb/s), and it limits the maximum feasible (error-free) transmission distance.

Another major origin of signal impairments in fiber-optic systems is the fiber nonlinearity effect. The optical fiber can only be approximated as a linear medium when the launch power is sufficiently low. In other words, the linearity assumption of fiber is only valid for short-haul optical fiber transmission systems, where launch

power can be low. As transmission distance increases, the use of EDFAs becomes mandatory, and the launch power must hence be increased to keep the OSNR high enough for combating the accumulated ASE noise added by the EDFAs, distributed along the transmission fiber link. When the launch power gets considerably high, the nonlinear behaviour of the optical fiber becomes significant, resulting in severe performance degradation. The fiber nonlinearity effects cannot therefore be ignored for long-haul optical fiber communication systems. Fiber nonlinearity also enforces an upper limit to the launch power of fiber-optic communication system, since it increases as the optical signal power increases.

Nonlinear effects in optical fibers are mainly due to the optical Kerr effect, which denotes to the refractive index dependence on the intensity of the propagated optical pulse (Kerr, 1875). This phenomenon originates from the fact that when the optical field intensity is comparable to the inter-atomic electric field of fiber, optical signal interacts with atoms of fiber, generating new frequency components (Boyd, 2019). The Kerr effect causes three types of nonlinear distortions in a single fiber-optic channel. They are self-phase modulation (SPM), intra-channel cross-phase modulation (IXPM), and intra-channel four-wave mixing (IFWM). Figure 1.4 illustrates these nonlinear effects (see for example channel 3).

The SPM represents the induced nonlinear phase shift added to an optical pulse due to its own intensity, causing a spectral pulse broadening. Note that the SPM interacts with the chromatic dispersion of the fiber. This interaction results in temporal pulse broadening in case of normal dispersion, or causes pulse compression in the anomalous dispersion case. The latter interaction (the SPM with the anomalous fiber dispersion) can be exploited to formulate an important phenomenon known as

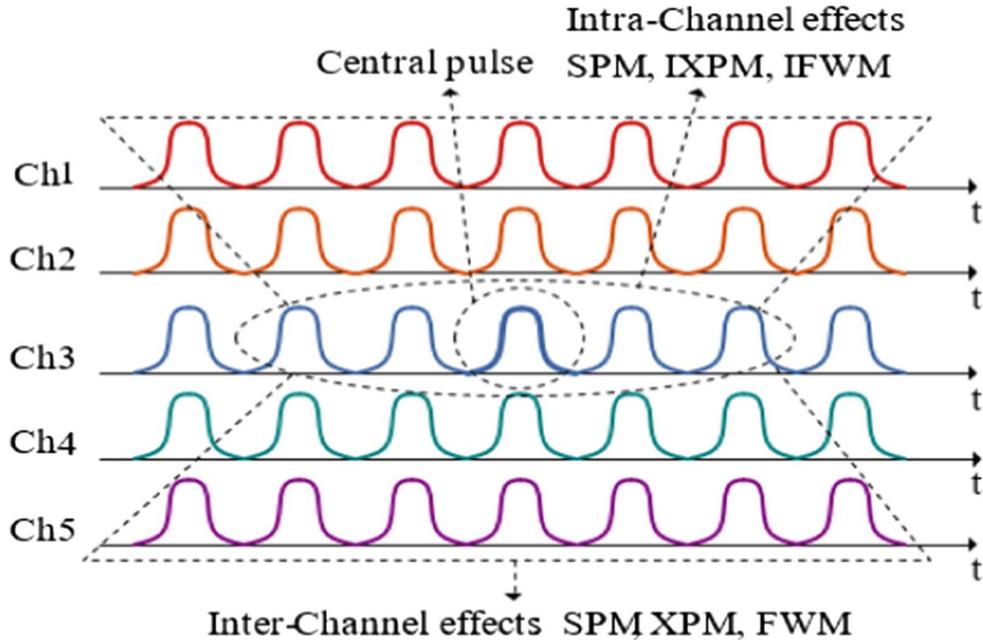


Figure 1.4: An illustration of nonlinear interactions in a single channel and WDM fiber-optic communication system (Kumar *et al.*, 2018).

an optical soliton (Hasegawa and Tappert, 1973; Hasegawa and Kodama, 1981). The optical soliton is a solitary light wave that maintains its shape while travelling through an anomalous optical fiber, due to the cancellation of the SPM and the dispersive effects of the fiber. The concept of the optical soliton formation is later discussed in more details in Chapter 3. The IXPM is the nonlinear phase shift added to an optical pulse because of the intensity dependence of its neighboring pulses. It causes asymmetric spectral broadening for the optical pulse, resulting in amplitude distortion in time-domain. The nonlinear interaction among the overlapped optical pulses of a same channel also generates ghost or echo pulses at new frequencies. This effect is known as the IFWM, and it represents the dominant penalties for high data-rate (40 Gbaud and above) optical fiber systems (Essiambre *et al.*, 1999). For instance, the interaction between 3 pulses centered at t_1 , t_2 , and t_3 leads to generating two

ghost pulses at $t_1 + t_2 - t_3$ and $-t_1 + t_2 + t_3$.

In case of a WDM system, the Kerr effect causes nonlinear interactions between not only pulses of the same channel but also pulses of different channels, as shown in Fig. 1.4. The latter interactions are called inter-channel nonlinear effects. These effects are divided into cross-phase modulation (XPM) and four-wave mixing (FWM). While the XPM corresponds to the nonlinear interaction of two WDM channels, the FWM corresponds to the nonlinear interactions of three WDM channels. The inter-channel XPM represents the modulation of the phase of a signal in one channel due to the power variance of a signal in another channel. Also, nonlinear interactions among signal channels of center frequencies f_1 , f_2 , and f_3 results in new frequency components at $f_1 + f_2 - f_3$ and $-f_1 + f_2 + f_3$ due to the scattering of the incident photons, which is known as the FWM effect (Tkach *et al.*, 1995). These new frequency components generated through FWM act as noise on channels centered at $f_1 + f_2 - f_3$ and $-f_1 + f_2 + f_3$.

Moreover, nonlinear interactions between signal and ASE noise lead to additional signal impairments, known as nonlinear phase noise. Due to the ASE noise, the signal instantaneous power fluctuates, which causes random change in the signal phase because of the nonlinear Kerr effect. This phenomenon is known as the Gordon-Mollenauer effect (Gordon and Mollenauer, 1990). The nonlinear phase noise could lead to significant performance degradation in the absence of fiber dispersion, limiting the usefulness of phase-modulation schemes (Mecozzi, 1994). However, in the presence of moderate to large dispersion, the effect of nonlinear phase noise is significantly reduced (Kumar, 2005).

This thesis focuses on studies of mitigating the impairments induced by a single fiber-optic channel. We therefore summarize, in the next section, the electrical equalization techniques developed for combating the linear and nonlinear distortions induced by both short-reach and long-haul optical fiber communication systems, with single channel carrier.

1.3 Electrical Compensation Techniques for Optical Fiber Distortions

Various compensation techniques have been proposed in literature for mitigating the linear and nonlinear distortions imposed by optical fiber communication systems. The impairments due to the optical fiber link can be equalized either by pre-compensating the signal at the transmitter side (Bülow *et al.*, 2008), or by using a compensation technique to equalize the received signal. The latter solution is known as post-compensation technique, and it is currently of preeminent interest, due to the fact that the exact pre-determination of the of transmission fiber link parameters represents a major challenge to utilize a pre-compensating scheme.

The post-compensation techniques can be classified into optical and electrical methods. While the optical compensation methods equalize the distorted signal in optical domain, the electrical compensation techniques do the equalization in the electrical domain, after the photo-detection step. Optical compensation techniques could provide more efficient equalization performance since they deal with the impairments in the domain where they actually occur. However, optical compensation methods severely suffer from the high cost required for designing accurate and complex optical

devices (Sarkis, 2009). Electrical compensation methods would not provide the same efficiency that could be acquired in the optical domain. Optical domain equalizers have drawn significant attention due to their ability to compensate for linear and nonlinear distortions in real time and due to their WDM compatibility (Kumar and Yang, 2011; Kumar and Shao, 2013; Bidaki and Kumar, 2019). However, as compared to optical equalizers, electrical compensation techniques offer superior performance in terms of speed, robustness, stability, and cost (Franceschini *et al.*, 2007). Moreover, electrical equalizers offer an adaptive equalization that traces dynamic linear distortions of the optical fiber (Bohn and Xia, 2009).

The first set of electrical equalizers, that have been utilized for compensating optical fiber impairments, were mimicked from solutions introduced earlier to problems of wireless communications systems. Instances of these solutions are linear feed-forward equalizer (FFE) (Curri *et al.*, 2004), decision feedback equalizer (DFE) (Wang and Kahn, 2004), and maximum-likelihood sequence estimator (MLSE) (Foggi *et al.*, 2006).

1.3.1 Feed-Forward and Decision Feedback Equalizers

The Linear FFE is the most popular and easy-to-implement electronic equalizer for compensating linear impairments of optical fibers. The linear FFE, which is also known as a digital transversal filter (or a tapped-delay-line), has the basic structure shown in Fig. 1.5. The input signal x is fed into a delay line composed of M taps. Each delay corresponds to one bit interval. The signal and its delayed versions are multiplied by adjustable equalization coefficients or tap-weights, $(C_i, i = 1, 2, \dots, M)$, and then summed up to generate the equalized output signal y . With the right choice

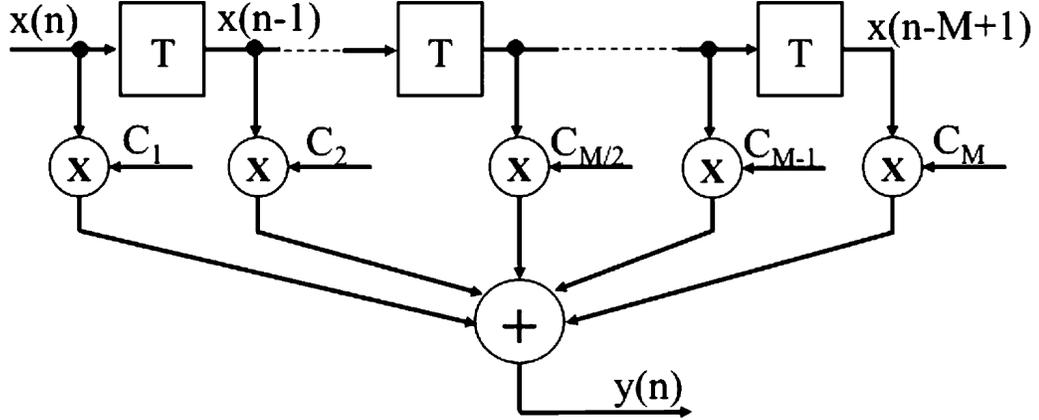


Figure 1.5: Block diagram of a M -taps linear feed-forward equalizer (Momtaz and Green, 2010).

T stands for a one bit time interval delay, C_i , $i = 1, 2, \dots, M$, are adjustable tap-weights, and M is the total number of taps. The signals $x(n)$ and $y(n)$ are the un-equalized input and equalized output at sampling time nT_b , respectively, where T_b is the inverse baud rate.

of the number of taps and the tap weights, we can use this equalizer to provide the inverse response of any linear channel. Thus, the FFE can be utilized to undo the linear distortions (e.g. chromatic dispersion effect) of the fiber.

Several algorithms can be applied to adaptively adjust and optimize the FFE tap-weights. Three of the most widely used algorithms are least-mean square (LMS) algorithm (Widrow and Kamenetsky, 2003), recursive least squares (RLS) algorithm (Cioffi and Kailath, 1984), and zero-forcing algorithm (Lucky, 1965). The LMS algorithm aims at minimizing the current mean square error between the desired and the received signal. Whereas, the RLS algorithm recursively finds the tap-weights that minimize the total weighted squared error between the desired signals and the received signals, i.e., it adjust the equalizer weights based on the total error computed from the beginning. In contrast, the zero-forcing algorithm seeks to minimize the worst-case error at the output of the equalizer. As compared to the FFE trained

using zero-forcing algorithm, the LMS and RLS based FFEs provide better performance especially when the noise in the system is slightly high. At the expense of more complexity and computational cost, the RLS algorithm provides smaller steady state error as opposed to the LMS algorithm. However, the LMS based linear FFEs are the most widely utilized in practical fiber-optic problems due to the fact that the LMS algorithm provides the best trade-off between performance and computational cost (Dhiman *et al.*, 2013).

The equalization performance of linear FFE can be slightly enhanced using the DFE. The basic idea of the DFE is to make use of previous decisions (previous equalized and estimated data) in attempting to estimate the current data. The general structure of a DFE consists of two linear FFE blocks, as shown in Fig. 1.6. The first FFE block consists of M -delayed taps, and is fed by the received distorted signal x , like the case in the normal FFE. On the other hand, the second FFE block (which is called decision feedback block) has N -delayed taps, and is fed by the estimated output signal of the decision circuit z . Therefore, this equalizer is formally denoted as a feed-forward decision feedback equalizer. However, for simplicity, we use the term DFE for the combination of the feed-forward and feedback blocks. The total equalized output signal y of the DFE equalizer is obtained by subtracting the feedback output from the feedforward output, as illustrated in Fig. 1.6. In other words, any trailing (post-cursor) linear distortion caused by previous symbols is reconstructed and then subtracted.

Thus, the leading linear distortions (e.g. precursor ISI) are removed in the same way as in the linear FFE using the feed-forward block. However, the post-cursor linear impairments are now removed through the feedback block using the output from the

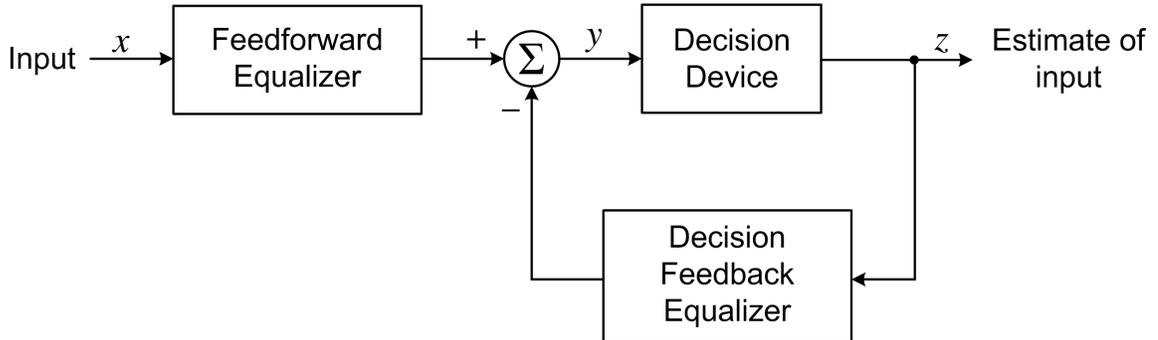


Figure 1.6: General architecture of a decision feedback equalizer.

decision circuit instead of the output from the forward delay line as the case in the linear FFE. This way of estimating the past symbols has the advantage that it is free of noise since it represents the precise transmitted symbol value, assuming that the past decision was correct. Actually, this is the main reason that the DFE outperforms the BER performance of the FFE. However, if it happened that a previous decision was wrong, the DFE equalizer would add more distortions to the output signal y , which in turns could cause further decision errors. This effect of passing wrong equalized decisions to the feedback process is known as error propagation (Bohn and Xia, 2009). Nevertheless, the advantages of the DFE outweigh its disadvantages, and in practice, DFEs slightly outperform the performance of FFEs at the expense relatively higher implementation complexity and computational cost (Watts, 2007).

Linear FFEs and DFEs are the simplest and fastest option amongst all other available electronic equalizers. Their computational cost represents the lower benchmark threshold for the computational complexity of electrical compensation techniques. The computational cost per symbol of both FFE and DFE scales linearly with the ISI span (i.e. the required number of taps). Also, both FFE and DFE can effectively compensate for the linear fiber distortions (e.g. CD dispersion), excluding the

nonlinear impairments induced in long-haul coherent optical fiber communication systems. In contrast, they fail to mitigate these linear fiber distortions for short-reach fiber-optic communication systems, despite the fact that nonlinear fiber distortions are negligible in short-reach optical fiber transmission systems. This failure is due to the fact that in practical short-haul optical fiber applications, the intensity of the optical carrier is modulated by the electrical information signal, and at the receiver a direct photo-detection system is usually used (Chen *et al.*, 2020). This is known as the intensity-modulation and direct detection (IM-DD) systems. The direct photo-detection receiver is a nonlinear detector, since it consists of a single photodiode that acts as a square-law detector, and correspondingly, the phase information of the optical signal is lost at the electrical domain. The linear optical distortions thus turn into nonlinear impairments in the electrical domain in the short-reach optical fiber communication systems. As a result, linear equalizers such as FFEs and DFEs fail to compensate for these nonlinear distortions, and the use of nonlinear equalizers becomes mandatory to mitigate the distortions induced in the IM-DD systems.

1.3.2 Maximum-Likelihood Sequence Estimator

The maximum likelihood sequence estimator (MLSE) is one of the most effective, but computationally expensive, nonlinear equalization techniques that can be used at the receiver to compensate for the linear and nonlinear impairments of fiber-optic communication systems. It provides the optimum equalization performance from the error probability point of view. Its basic idea is to consider all possible combinations of sequence stream, and the sequence for which the probability (to have been sent) is the highest, is decided to be the transmitted signal sequence (Agazzi *et al.*, 2005; Foggi

et al., 2006). In other words, the MLSE aims at comparing the received distorted (due to linear and nonlinear fiber impairments) and noisy signal with all possible distorted but noise-free received sequences, and then selecting the closest one. For sequences of length N bits, i.e., MLSE with memory size of N , this requires comparison with 2^N different noise-free sequences. The MLSE is thus equivalent to the problem of estimating the state of a discrete-time finite-state machine. When the symbols are M -ary and the MLSE memory size is N , the state machine has M^N states.

The MLSE technique requires knowledge of the channel characteristics. It therefore needs to be first trained by sending all possible sequence combinations and storing their corresponding received data in order to create a look-up table for the channel impairments. Once the look-up table has been established, actual unknown data can be transmitted. The MLSE algorithm then employs the Viterbi algorithm (VA) to search for the most probable path through a M^N -state trellis (Proakis, 2007). The MLSE shows high robustness performance for the nonlinear compensation of nonlinear fiber-optic impairments (Bohn and Xia, 2009). It has been demonstrated that the MLSE provides the superior BER performance over all other electrical equalization techniques utilized for fiber-optic communication systems (Weiss, 2003). It can therefore be used as a BER lower bound estimation of optical fiber systems. Despite that, it severely suffers from the high computational complexity. Its computational cost per symbol is exponentially proportional to the ISI span (i.e. the required memory size) (Bülow *et al.*, 2008; Bohn and Xia, 2009). This drawback significantly limits the possibility of high-speed transmission symbol rate.

Alternative Candidates to the MLSE

The high computational cost of the MLSE can be reduced, at the expense of lower BER performance, using other nonlinear compensation techniques. Volterra series-based equalizers (VSEs) have been investigated for nonlinear compensation of optical fiber nonlinear impairments (Schetzen, 2006). Until recently, the VSE has been utilized as a good candidate for reducing the computational complexity of the MLSE, offering a slightly lower BER performance as compared to the MLSE performance (Liu *et al.*, 2011). However, the computational complexity of VSEs is still reasonably high, since it needs an enormous amount of calculations. Its computational cost per symbol is of $O(N^3)$ to compensate for fiber nonlinear impairments, where N is the ISI span (Xia and Rosenkranz, 2007). We are therefore motivated to investigate for another nonlinear compensation technique that could provide the same performance of the VSE with much simpler implementation and lower computational complexity.

A neural network-based equalizer (NNE) could be a potential candidate offering lower computational complexity for the compensation of the distortions induced in the IM-DD fiber-optic systems. The neural network (NN) is a mathematical model that simulates the human neurons ability. Biological neurons fire when the input from all of its synapses exceeds a certain threshold. The mathematical neuron (node) of the NN imitates this biological behavior. Since it mimics the most complicated processing system in life (human brain), the NN is capable of performing any required linear or nonlinear mapping function from the input parameters space into the space of output responses (Bakr, 2013). Moreover, the model of a NN with single hidden layer has been mathematically proved to be universal (Hornik *et al.*, 1989).

Due to their promising performance, NNs have been widely used in various engineering fields and applications. For instance, NNs have been exploited to combat the uncompensated ISI effects of wireless communication systems (Zayani and Bouallegue, 2007; Burse *et al.*, 2010). Moreover, it has been demonstrated that NNEs could potentially improve the computational complexity required by VSEs (Owaki and Nakamura, 2018). While the computational cost of NNEs increases in proportion to the number of neurons in each layer, the VSE cost increases exponentially as the number of tapped delays and the order of the Volterra series increase (Schetzen, 2006). In particular, the computational cost per symbol of the NNE scales as $O(N^2)$, which is opposed to the $O(N^3)$ required by the VSE (Otsuka *et al.*, 2018).

1.3.3 Digital Back Propagation

The traditional direct detection receiver is not employed for long-haul fiber-optic transmission systems. In this case, a coherent photo-detection system is usually utilized at the receiver side. Unlike direction detection, the coherent detection technique makes it possible to obtain both the amplitude and phase information of the complex signals (Yamamoto and Kimura, 1981). The basic idea of the coherent receiver is to mix the received optical signal with a continuous wave signal from a coherent local oscillator in a way such that the in-phase and quadrature components of information signal can be detected using photo diodes (Basch and Brown, 1985). Since the phase information of transmitted signal is preserved in the coherent detected systems (i.e., the complex valued electrical field is fully detected at the receiver side), the amplitude, phase, and frequency of the optical carrier can all be utilized to carry the information. In other words, more advanced modulation schemes and modulation constellations,

e.g., QPSK and QAM, can be received as contrasted to the direct detection receivers. Therefore, the spectral efficiency can be further improved. Moreover, the preservation of the temporal phase enables more effective methods for the adaptive electronic compensation of transmission fiber impairments.

One of the most important advantage of coherent detected systems is that linear fiber distortions (e.g. chromatic dispersion effect) can effectively be compensated using the computationally non-expensive linear equalizers. This is due to the fact that the nonlinear impairments introduced by the square-law detection is absent in coherent receivers. However, since nonlinear fiber effects cannot be neglected in long-reach optical fiber communication systems, a nonlinear equalizer is still required to fully equalize the distortions. Digital back propagation (DBP) is the most commonly used technique to compensate for fiber impairment in long-reach optical fiber systems with coherent detection (Ip and Kahn, 2008; Li *et al.*, 2008). The basic idea of the DBP is to virtually transmit the received distorted signal through a virtual fiber whose loss, dispersion, and nonlinear parameters are the same in magnitude, as the actual transmission fiber, but with opposite signs, such that the distortion effects of the transmission link can be inverted, i.e., the transmitted signal can be recovered.

The signal propagation through an optical fiber link is fully described by a partial differential equation known as nonlinear Schrödinger equation (NLSE) (Agrawal, 2007; Kumar and Deen, 2014). The NLSE cannot be solved analytically, it is usually solved numerically using a method called split-step Fourier scheme (SSFS) (Sinkin *et al.*, 2003; Shao *et al.*, 2014; Deiterding and Poole, 2016). The NLSE and its solution technique are discussed in details later on in Chapter 3. The NLSE is also an invertible equation. The DBP therefore aims at numerically solving the inverted

NLSE to undo the induced fiber impairments. Particularly, the DBP is fed by the received optical signal after being converted into digital domain. Then, it applies the SSFS method for numerically solving the inverse NLSE backward in space, in order to undo the distortions and recover the transmitted signal. More details on DBP analysis are given in Chapter 4. Provided that the system is a noise-free and an enough small step-size is used in the SSFS algorithm, the DBP can fully compensate for all deterministic distortions due to dispersion and signal-signal nonlinear interactions. However, in practice, the noise exists and the noise-signal nonlinearity interactions cannot be compensated by the DBP. Also, the small SSFS step-size will increase the computational burden, even though it increases the DBP performance.

DBP was first introduced as a pre-compensation method to be utilized at the transmitter side of single optical fiber channel with direct detection system (Roberts *et al.*, 2006). The DBP is fed by the desired signal to be obtained at the receiver side. Notice that the desired signal at the receiver is the ideal signal with no noise (i.e. perfect constellations), which is already available at the transmitter. Pre-distortions are then added to the signal using the pre-compensator DBP. The pre-distortions imposed on the transmitted signal cancel with the distortions accumulated by the real transmission fiber link. The non-distorted transmitted signal could thus be recovered at the receiver if the system was noise-free. The implementation of a transmitter-based DBP was due to the fact that in the absence of coherent detection, manipulation of the field is only possible at the modulator. However, as previously mentioned, post-compensation techniques provide better flexibility over pre-compensation methods since it allows various adaptive equalization without the need for a feedback.

With coherent detection, employing a receiver-based DBP becomes applicable.

DBP was first implemented as a post-compensation technique in (Ip and Kahn, 2008; Li *et al.*, 2008). The DBP is inserted after a coherent detector to equalize the distorted received digital signal. In (Ip and Kahn, 2008), a non-iterative asymmetric SSFS algorithm is proposed for solving the inverse NLSE. It has been theoretically demonstrated that 3 samples per symbol are required for the DBP to achieve a good numerical accuracy (Ip and Kahn, 2008). It has also been shown that setting a step-size equal to the fiber span length achieves a reasonable performance provided that the symbol rate is 10 Gbaud or below. In (Li *et al.*, 2008), the dispersion step of the DBP is realized using finite impulse response filters. A parallel architecture is also designed for real-time implementation. It has been demonstrated using numerical simulations that 2 samples per symbol in the digital domain are sufficient to achieve significant nonlinearity compensation (Li *et al.*, 2008).

Exploiting the DBP to compensate for nonlinear effects of long-reach fiber-optic systems with coherent detection has been extensively studied in literature. DBPs based on coupled NLSE and total-field NLSE are utilized to fully compensate for the XPM and FWM effects in (Mateo *et al.*, 2008). A symmetric iterative SSFS algorithm is employed to solve the inverse NLSEs. The obtained results show that FWM distortions are relatively weak as compared to the XPM impairments if the fiber has high dispersion as in standard single-mode fibers (SSMF). It has also been demonstrated that the coupled NLSE-based DBP is sufficient to provide effective compensation for most nonlinearity effects in long-haul systems with large accumulated dispersion. In (Mateo and Li, 2009), an enhanced coupled NLSE based-DBP algorithm was proposed to fully compensate for the XPM effects and partially compensate for the FWM distortions. At the expense of slightly lower compensation performance, significant

reduction in the computational cost has been achieved as compared to the previous DBP scheme. Then, the computational complexity of the coupled NLSE based-DBP has been reduced by a factor of 4 through introducing an advanced SSFS method in (Mateo *et al.*, 2010).

In order to further reduce the computational cost of the DBP and to increase the potential for real-time implementations in practical systems, a filtered DBP scheme was proposed (Du and Lowery, 2010). Using the filtered DBP, it has been shown that the required number of DBP steps per fiber span can be reduced to less than one as compared to tens of steps needed by the conventional DBP. In the filtered DBP scheme, a low pass filter (LPF) is inserted in each DBP step to filter the signal before the nonlinearity compensation sub-step. The filtered DBP aims at enhancing the accuracy of the conventional DBP method while allowing the numerical step-size to be relatively large. In other words, the filtered DBP can provide compensation performance higher than the conventional DBP when a large step-size is used in both the two schemes. However, the optimal compensation performance obtained using the filtered DBP cannot exceed the conventional DBP optimum performance achieved when the step-size is reasonably small.

Note that the filtered DBP scheme reduces the required number of steps per span, as compared to the conventional DBP, but at the expense of increased computational complexity for each step. The filtered DBP therefore does not provide absolute reduction in the computational cost for all studied systems with any desired compensation performance level. In other words, the compensation performance of the filtered DBP is lower than that of the conventional DBP with small step-size, and it does not always guarantee achieving the desired resultant computational cost reduction benefit. For

instance, the resultant computational complexity (i.e. total number of required real multiplications) of the filtered DBP is compared with the conventional DBP resultant computational cost for different fiber-optic communication scenarios in (Gao *et al.*, 2012). It has been shown that, for single fiber-optic channels, the conventional DBP is preferred over the filtered DBP when large number of steps per span is considered. In other words, the conventional DBP requires lower resultant computational complexity as compared to the filtered DBP resultant computational cost when high compensation performance is required.

Moreover, as the data rate per channel increases, the DBP with the step-size being equal to the fiber span length does not provide a good performance, and it has been shown that the step-size has to be smaller than the fiber span length (Shao *et al.*, 2014). However, the virtual fiber may be divided into a few sections and lengths of each section are determined by optimizing the area mismatch between the ideal exponential profile and its stepwise approximation (Shao *et al.*, 2014).

One of the major limitations of conventional and filtered DBPs is the fact that they only compensate for deterministic linear and nonlinear distortions of the fiber, and inherently do not consider noise. In other words, the compensation performance of DBPs is ultimately limited by non-deterministic nonlinear effects, e.g., nonlinear signal-noise interactions between the transmitted signal and the amplified spontaneous emission (ASE) noise, which is known as the nonlinear phase noise (Essiambre *et al.*, 2010; Rafique and Ellis, 2011; Galdino *et al.*, 2017). In (Irukulapati *et al.*, 2014, 2016), stochastic DBP methods were proposed to compensate not only for deterministic linear and nonlinear impairments, but also to mitigate stochastic distortions due to the nonlinear phase noise. The basic idea of the stochastic DBP is exploiting the

Bayesian detection theory and factor graphs to formulate maximum a posteriori probability detectors that can mitigate data-pattern-dependent nonlinear impairments, by minimizing the error probability. It has been shown that for dispersion-managed links, stochastic DBP shows a significant increase in the system reach as compared to the conventional DBP. However, the stochastic DBP and the conventional DBP exhibit similar performance for dispersion uncompensated links (Irukulapati *et al.*, 2014).

It is clear that the compensation of fiber impairments using the DBP requires pre-determination of detailed physical fiber-optic channel parameters. Such information can only be available at the receiver when the case of a point-to-point transmission system is considered. Contrary, in case of a dynamic optical network, precise information of the physical fiber channel parametrization may not be available. For instance, transmitted carriers in elastic optical networks (EONs) may propagate through several different routes until reaching the same receiver point (Sharma and Kumar, 2017). These routes are determined according to the dynamic network configurations. Hence, it becomes challenging and very difficult to determine the fiber channel parameters in such scenarios (da Silva *et al.*, 2015). Moreover, the physical parameters of optical fiber links and in-line EDFAs are subject to slight variations, from their nominal values, due to environmental changes. The need for an adaptive DBP scheme that is capable of obtaining and tracking the optimal DBP parameters is therefore essential and of preeminent.

Some approaches have been proposed in literature for estimating the conventional DBP parameters without knowing the launch power and channel parameters. A semi-blind DBP method was first proposed in (Tanimura *et al.*, 2010). In this method, the dispersion and nonlinear parameters of the DBP are tuned according to information

derived from the received signal after detection. However, such approach requires significant delay and is not practical for high-speed transmission systems, since it needs prior knowledge of previously detected data that have been recovered using the computationally expensive forward error correction device. An adaptive DBP method which could only self-determine the fiber nonlinearity coefficient utilizing steepest descent algorithm is proposed in (Lin *et al.*, 2014). Starting from wrong nonlinear coefficient estimation, the DBP is trained, using the known transmitted data, to minimize the phase-noise variance of the received constellation. In (da Silva *et al.*, 2015), a modified adaptive DBP technique is proposed where the nonlinear fiber coefficient is optimized, exploiting a stochastic gradient algorithm, for minimizing the metric of error vector magnitude.

In order to reduce the number of required DBP steps, Jiang *et al.* extended the adaptive technique introduced in (Lin *et al.*, 2014), for the conventional DBP, to the filtered DBP scheme (Jiang *et al.*, 2016). Instead of the phase noise variance, the proposed approach aims at minimizing the Godard's error function utilizing the steepest descent algorithm. However, the optimization process in this algorithm is only limited to the nonlinear compensation parameter. In other words, the dispersion compensation parameter and the linear filter bandwidth are assumed to be known. In Ref. (Zhang *et al.*, 2016a), Zhang *et al.* proposed an adaptive filtered DBP method that is capable of optimizing both the nonlinear scaling factor and the filter bandwidth. However, this approach is limited to homogenous links, where the parameters are considered identical for all DBP steps. Zhang *et al.* then extended their method and proposed a multi-parameter adaptive filtered DBP method for heterogeneous optical fiber links (Zhang *et al.*, 2017). In the extended scheme, the steepest descent

algorithm is used to optimize the dispersion compensation, nonlinearity compensation, and filter bandwidth parameters for every DBP step, where the objective is to minimize the received noise variance. Zhou *et al.* recently proposed an adaptive DBP method based on the perturbation theory (Zhou *et al.*, 2020). Instead of utilizing the assisted filter approach, the computational complexity of the DBP is reduced in this scheme exploiting the perturbation theory (Liang and Kumar, 2014). Provided that the total accumulated chromatic dispersion is known, the proposed adaptive perturbation theory-based DBP method tends to optimize the nonlinear-step tap values using the Levenberg-Marquardt method.

So far, the common thing between all the currently available adaptive DBP methods is the fact that they all utilize a gradient-based optimization algorithm for estimating the channel parameters and adapting the DBP parameters. Any gradient-based optimization technique needs to evaluate the full gradient vector of the objective function (i.e. its derivatives with respect to all design parameters) with respect to the current design parameter vector, at least once per each optimization step, in order to determine the values of the next design parameter vector. Since the NLSE and its inverse equation could not be solved analytically, the required gradient information cannot be determined analytically. Rather, they are obtained numerically.

The classical approach to estimate the gradient information, which has been used in all aforementioned adaptive DBP methods, utilizes finite-difference (FD) approximations at the system response level. A total of extra N system simulations are required, using forward or backward finite-differences approaches, for the training of a DBP with N parameters. The more accurate central finite-differences (CFD)

approach requires $2N$ extra simulations. This overhead is prohibitive and impractical for such problems with large number of design parameters. For instance, if a fiber-optic network has 20 spans, the DBP at the receiver would need to know the values of as many as 80 parameters. Hence, the nonlinearity compensation at the receiver using DBP would require optimization of these parameters, which would be computationally expensive. However, this intensive computational cost can be significantly reduced through evaluating the sensitivities using an adjoint sensitivity analysis (ASA) approach (Bakr *et al.*, 2017).

The ASA method estimates the sensitivities of the desired objective function or response with respect to all design parameters using at most one extra system simulation, regardless of the design parameters number. A number of commercial software-tools, e.g., HFSS (hfs, 2013) and CST (cst, 2013), have started to implement ASA algorithms to limit the computational time necessary for the sensitivity calculations. Several ASA algorithms based on Maxwell's equations or wave equation have been first proposed for estimating the sensitivities of high frequency structures. The first set of these algorithms has been developed for frequency-domain simulation techniques, e.g., the finite element method (Park *et al.*, 1992, 1996; Webb, 2001, 2002; Igarashi and Watanabe, 2010), and the method of moments (Georgieva *et al.*, 2002).

The second set includes the ASA algorithms demonstrated for time-domain simulation techniques, e.g., the transmission-line modeling (TLM) method, and the finite-difference time-domain (FDTD) method. The time-domain ASA algorithms have the advantage of supplying wideband sensitivity analysis information using a single extra adjoint simulation. Different TLM-based ASA algorithms were introduced for sensitivity estimation of high frequency structures. These algorithms are dedicated to

problems with linear non-dispersive materials (Bakr and Nikolova, 2004; Negm *et al.*, 2014), linear dispersive materials (Ahmed *et al.*, 2012), or TLM-based nonlinear and non-dispersive materials (Bakr *et al.*, 2014).

Various FDTD-based ASA algorithms have also been proposed for problems with either linear non-dispersive materials (Chung *et al.*, 2000, 2001; Nikolova *et al.*, 2004; Swillam *et al.*, 2007), or linear dispersive materials (Zhang *et al.*, 2014, 2016b). However, these approaches cannot estimate the sensitivities of problems with nonlinear media. Moreover, all these algorithms are implemented utilizing the second-order space derivatives of the wave equation, which results in complex formulation and significantly increases the required memory storage. Bakr *et al.* developed the first FDTD-based ASA approach that is derived from first-order basic Maxwell's equations in (Bakr *et al.*, 2016, 2017). However, this algorithm was limited to only linear problems. The linear ASA approach of (Bakr *et al.*, 2016, 2017) is recently extended to general EM and photonic problems with nonlinear materials (Maghrabi *et al.*, 2020). The theory was also extended to adopt incident field excitation problems in addition to problems excited by current sources.

Moreover, some ASA approaches have been proposed in other fields. An ASA technique was proposed to estimate the sensitivities of switched reluctance motors in (Sayed *et al.*, 2018). ASA approaches have also been introduced for the linear time-independent Schrödinger equation (Swillam *et al.*, 2008), and for the linear time-dependent Schrödinger equation (Ayad *et al.*, 2015) to evaluate the sensitivities of semiconductor quantum structures in their static and dynamic cases, respectively. However, none of the existing ASA algorithms can be utilized to estimate the sensitivities for the fiber-optic systems governed by the general nonlinear Schrödinger

equation. We are therefore motivated to develop a general ASA approach for the NLSE and the inverse NLSE. Such approach could be used to significantly accelerate the training process of adaptive DBP schemes. Furthermore, an NLSE-based ASA algorithm could be exploited to accelerate the design optimization of any optical fiber system using gradient-based optimization algorithms. The optimal design of dispersion-managed coherent fiber-optic systems is an instance of such design optimization problems (Yang and Kumar, 2009).

1.4 Thesis Contribution

This thesis focuses on the mitigation of fiber-optic communication systems impairments using electronic compensation techniques. Particularly, the objective is to develop powerful and computationally efficient digital equalization techniques to compensate for the linear and nonlinear distortions induced by single-channel short-reach and long-haul optical fiber transmission systems and networks. This objective is achieved using two approaches.

The first approach, presented in Chapter 2, is proposed for mitigating the distortions of short-reach optical fiber communication systems with direct detection. Although fiber nonlinearity effects are neglected in short-haul fiber-optic transmission systems, the induced linear impairments turn into nonlinear, at the receiver, due to the square-law detection of direct photo-detectors. A nonlinear equalizer is therefore required at the DSP unit of the receiver to effectively combat these distortions. The MLSE provides the best compensation performance for such system. However, its computational cost is very expensive, which significantly limits the desired transmission of high-speed symbol rate.

Chapter 2 proposes a powerful, fast, and easy-to-implement artificial neural network nonlinear feed forward equalizer (ANN-NFFE). The proposed ANN-NFFE provides a slightly lower compensation performance, in terms of BER, as compared to that of MLSE. However, the computational cost of the ANN-NFFE is significantly lower than MLSE. While the ANN-NFFE requires computational cost per symbol growing linearly with the length of ISI span, the cost of MLSE is exponentially proportional to the ISI span. The proposed equalizer utilizes a trained ANN that acts as a nonlinear filter whose impulse response inverts the nonlinear response of the optical communication channel. At first, the parameters of the ANN are optimized and adopted using extensive training process considering all possible data combinations that will be later transmitted through the optical channel. Two possible application areas for the potential use of the proposed ANN-NFFE, namely, metro networks and data center networks (DCNs), are studied in this Chapter. Our simulation results show that the ANN-NFFE increases the CD tolerance and significantly extends the feasible system reach.

In the second approach, we aim at developing an efficient adaptive DBP scheme with fast adaption process to mitigate the nonlinear distortions of long-reach fiber-optic communication systems with coherent detection. The DBP method is widely used to compensate for the nonlinear fiber impairments of coherent optical fiber systems. Given that the signal propagation through an optical fiber link is described by the NLSE, the DBP algorithm tends to invert the fiber effects through applying the inverse NLSE on the received distorted signal. Although the conventional DBP scheme requires precise information of the fiber channel parametrization, a number of adaptive DBP schemes were proposed to combat the fiber impairments without

knowing the launch power and channel parameters. These schemes utilize a gradient-based optimization algorithm to adapt the DBP parameters. The required sensitivity information are estimated by applying the FD approach at the system level. However, the computational cost of the FD approach scales linearly with the number of parameters. This overhead is prohibitive and impractical, especially when considering all freedom degrees of the DBP designable parameters. Estimating the sensitivity information using the computationally efficient ASA approach could though significantly reduce the intensive computational cost of sensitivity analyses. Developing an ASA approach based on the NLSE is therefore essential to increase the potential for real-time implementations of the adaptive DBP in practical fiber-optic systems.

Chapter 3 proposes an efficient ASA approach for the NLSE. To our knowledge, this is the first time that an ASA approach is introduced for the general time-dependent NLSE. The proposed ASA algorithm could be exploited to significantly accelerate the required sensitivity calculations in any fiber-optic design problem. Regardless of the number of design parameters, our ASA approach estimates the sensitivities of a general objective function with respect to all design parameters using only one extra adjoint system simulation. This is contrasted with the classical FD approach whose computational cost depends on the number of designable parameters. We derive an adjoint nonlinear system of equations corresponding to the SSFS representation of the NLSE solution. A modified split-step Fourier scheme method is also introduced to solve the derived adjoint problem. The accuracy and efficiency of the proposed ASA algorithm are numerically investigated through optical fiber examples. The results show that the proposed ASA algorithm has the same accuracy as the CFD approach but with a much lower computational cost.

In Chapter 4, we propose a novel, powerful, and computationally efficient adaptive DBP (A-DBP) scheme, exploiting the NLSE-based ASA approach, to blindly compensate for the linear and nonlinear distortions of optical fiber transmission channels. The ASA introduced for the NLSE is extended to adopt the sensitivity analysis of the general multi-span DBP model, i.e., the extended ASA includes both the effect of the inverse nonlinear Schrödinger equation (INLSE) and the in-line amplifier inverse effect. An adjoint-based optimization (ABO) algorithm based on the sequential quadratic programming method is used to train and optimize the parameters of the A-DBP. All sensitivity calculations, required by the ABO algorithm, are estimated using the ASA approach for significant acceleration of the A-DBP training process. Regardless of the considered A-DBP design parameters number, full gradient information of the A-DBP training problem with respect to all parameters is obtained using only one extra adjoint DBP system simulation. The robustness and efficiency of the proposed A-DBP algorithm is investigated through applying it to mitigate the distortions induced in a 4-span fiber-optic communication system. Provided that the total transmitted distance is known, the proposed A-DBP could be rapidly trained, achieving the optimum compensation performance that is obtained using a conventional DBP with the correct parameters of the channel. The A-DBP training using the ABO algorithm is shown to be as accurate as the training utilizing a traditional CFD-based optimizer, but with much lower computational cost. To further reduce the computational complexity, we also demonstrate the compensation of the 4-span optical communication scenario using a coarse-mesh A-DBP model with less number of spans. The obtained results show that an optimized 2-span coarse-mesh A-DBP

model provides the best trade-off between equalization performance and computational cost.

The research work has resulted in the following publications:

Journal publications:

1. Mahmoud M.T. Maghrabi, Shiva Kumar, and Mohamed H. Bakr, “Nonlinear neural network equalizer for metro optical fiber communication systems,” *The Applied Computational Electromagnetics Society (ACES) Journal*, vol. 33, no. 10, pp. 1159-1160, 2018.
2. Mahmoud M. T. Maghrabi, Shiva Kumar, and Mohamed H. Bakr, “Dispersion compensation of fiber optic communication system with direct detection using artificial neural networks (ANNs),” *Optics Communications*, vol. 409, pp. 109-116, 2018.
3. Mahmoud M. T. Maghrabi, Mohamed H. Bakr, and Shiva Kumar, “Adjoint sensitivity analysis approach for the nonlinear Schrödinger equation,” *Optics letters*, vol. 44, no. 16, pp. 3940-3943, 2019.
4. Mahmoud M. T. Maghrabi, Mohamed H. Bakr, and Shiva Kumar, Atef Z. Elsherbeni, and Veysel Demir, “FDTD-based adjoint sensitivity analysis of high-frequency nonlinear structures,” *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 6, pp. 4727-4737, 2020.
5. Mahmoud M. T. Maghrabi, Shiva Kumar, and Mohamed H. Bakr, “Nonlinear

Schrödinger equation-based adjoint sensitivity analysis,” The Applied Computational Electromagnetics Society (ACES) Journal, vol. ?, no. ?, pp. ??, 2020.

6. Mahmoud M. T. Maghrabi, Mohamed H. Bakr, and Shiva Kumar, “Adaptive digital back propagation using adjoint sensitivity analysis-based gradient optimization,” to be submitted in Journal of Lightwave Technology, IEEE/OSA.

Conference publications:

1. Mahmoud M. T. Maghrabi, Mohamed H. Bakr, Shiva Kumar, Atef Z. Elsherbeni, and Veysel Demir, “Adjoint sensitivity of nonlinear structures using the FDTD method,” In the 5th Advanced Electromagnetic Symposium (AES 2017), Inchon, Korea, 2017.
2. Mahmoud M. T. Maghrabi, Shiva Kumar, and Mohamed H. Bakr, “Nonlinear neural network equalizer for metro optical fiber communication systems,” In 2018 International Applied Computational Electromagnetics Society Symposium (ACES 2018), Denver, Colorado, USA, IEEE, 2018.
3. Mahmoud M. T. Maghrabi, Shiva Kumar, and Mohamed H. Bakr, “ANN-based mitigation of optical fiber nonlinear distortions in data center networks,” In 2018 Photonics North (PN 2018), Montreal, Quebec, Canada, 2018.
4. Mahmoud M. T. Maghrabi, Mohamed H. Bakr, and Shiva Kumar, “Linear adjoint sensitivity analysis of the time-dependent Schrödinger equation,” In 2019 International Applied Computational Electromagnetics Society Symposium (ACES 2019), Miami, Florida, USA, IEEE, 2019.

5. Mahmoud M. T. Maghrabi, Mohamed H. Bakr, and Shiva Kumar, “Nonlinear Schrödinger equation-based adjoint sensitivity analysis,” In 2020 International Applied Computational Electromagnetics Society Symposium (ACES 2020) , Monterey, California, USA, IEEE, 2020.
6. Mahmoud M. T. Maghrabi, Mohamed H. Bakr, and Shiva Kumar, “Nonlinear adjoint sensitivity analysis of fiber-optic communication systems,” in 2020 Canadian Photonics Online Meetup (canPOM 2020), December 2020.

Chapter 2

Dispersion Compensation of Direct Detected Optical Fiber Communication Systems Using Artificial Neural Networks (ANNs)

In this chapter, we propose a powerful, fast, and easy-to-implement artificial neural network nonlinear feed-forward equalizer (ANN-NFFE). The ANN exploited in this work is a mathematical model that simulates the human neurons ability. Biological neurons fire when the input from all of its synapses exceeds a certain threshold. The mathematical neuron (node) of the ANN imitates this biological behavior. The ANN, because it mimics the most complicated processing system in life (human brain), is capable of performing any required linear or nonlinear mapping function from the input parameters space into the space of output responses (Bakr, 2013). Moreover,

the model of an ANN with single hidden layer has been mathematically proved to be universal (Hornik *et al.*, 1989). Due to their promising performance, ANNs have been widely used in various engineering fields and applications. Equalization of wireless communication channels is a particular instance of exploiting ANNs to combat uncompensated ISI effects (Burse *et al.*, 2010).

The proposed ANN-NFFE provides a slightly lower transmission system performance, in terms of bit error rate (BER), as compared to that of maximum likelihood estimator (MLSE). However, the computational cost of the ANN-NFFE is significantly lower than MLSE. While the ANN-NFFE requires computational cost per symbol growing linearly with the length of intersymbol interference (ISI) span, the cost of MLSE is exponentially proportional to the ISI span. The proposed equalizer utilizes a trained ANN that acts as a nonlinear filter whose impulse response inverts the nonlinear response of the optical communication channel. At first, the parameters of the ANN are optimized and adopted using extensive training process considering all possible data combinations that will be later transmitted through the optical channel.

We discuss, in this chapter, two possible application areas for the potential use of the ANN-NFFE (i) metro networks and (ii) data center networks (DCNs). These networks use directly modulated lasers and direct detection receivers. As the bit-rate increases, the transmission distance is mainly limited by the chromatic dispersion (CD). Our results show that the ANN-NFFE increases the CD tolerance and significantly extends the feasible transmission distance. To achieve a BER of 10^{-9} , our ANN-NFFE extends the transmission distance up to 200 km with at most 2 dB-OSNR penalty as compared to the back-to-back (B2B) transmission case.

Typically, the transmission distance of DCNs ranges from a few meters to a few tens of kilometers (Wei *et al.*, 2012). For baud-rates less than 10 Gbaud and transmission distances less than 2 km, direct modulation and on-off keying (OOK) are simple and cost effective (Lam *et al.*, 2014). The CD of the fiber is usually not a limiting factor for distances less than 2 km. However, as the link speed increases from 10 Gbps to 100 Gbps and 400 Gbps, novel modulation schemes and digital signal processing will be needed for DCNs (Lam *et al.*, 2014). Our results show that when considering the OOK modulation format and a baud-rate of 28 Gbaud, the ANN-NFFE extends the feasible transmission distance from 10 km to 20 km. Moreover, the ANN-NFFE might be used in transmission distances 10 km, in order to reduce the OSNR penalty for achieving a BER of 10^{-9} .

The chapter is organized as follows: The general structure of direct detected short-reach fiber optic communication system is described in Section 2.1. A mathematical model of the entire system is provided. Section 2.2 represents the configuration of the introduced ANN equalizer and its governing mathematical equations. The extensive training methodology used to adjust parameters of the ANN-NFFE is described. Section 2.3 shows simulation results of adapting the ANN-NFFE to compensate the dispersion imposed on non-return to zero (NRZ)- and return to zero (RZ)-on-off keying (OOK) modulated format transmitted data through short-reach/metro fiber optic communication systems. The typical bit-rate, used in such short-reach optical communication systems, 10 Gbps, is assumed. Furthermore, we investigate the performance of the ANN-NFFE when considering higher bit-rate, in order to follow up the current trend in the applications of short-reach optical communication systems (Sun *et al.*, 2015). Particularly, we study the case of 28 Gbps data communication rate

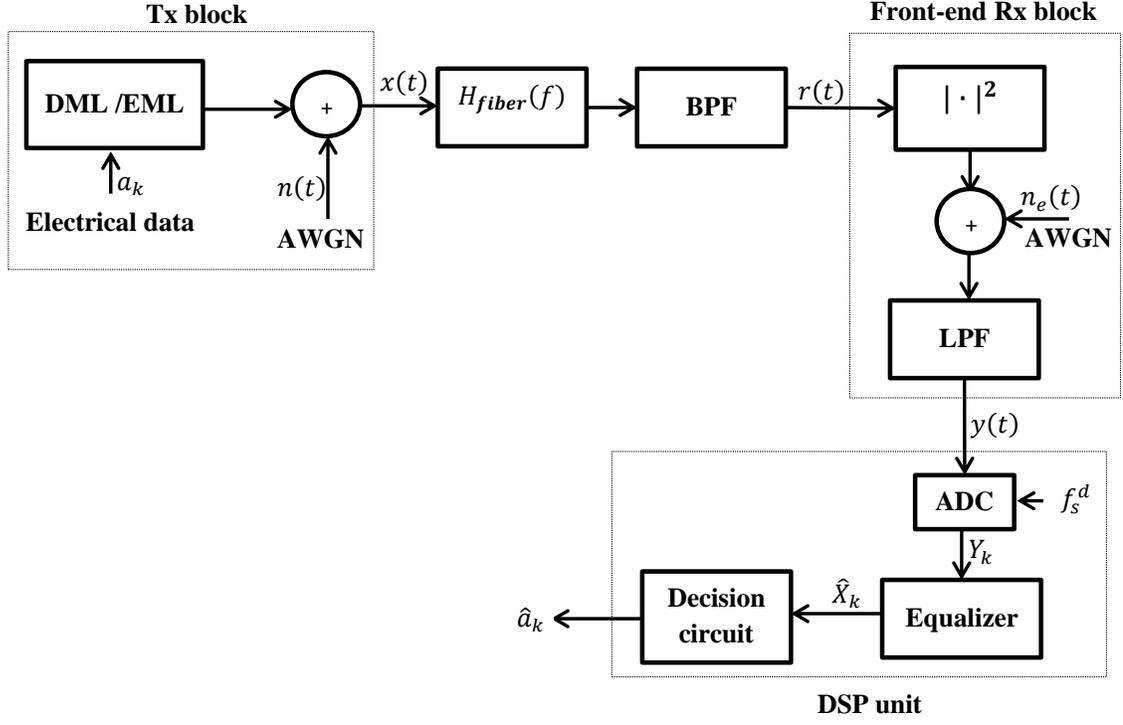


Figure 2.1: Model of short-reach fiber optic communication system with direct photo-detection.

DML: Directly modulated laser, EML: Electrically modulated laser, ADC: Analog to digital converter, BPF: Band pass filter, LPF: Low pass filter.

that has been recently used in metro fiber optic links and data center networks (Wei *et al.*, 2012; Lam *et al.*, 2014; Sun *et al.*, 2015). Finally, conclusions of the work are drawn in Section 2.4.

2.1 Optical Communication System Model

Throughout this chapter, our attention is focused on short-reach optical communication systems, e.g., metro and media access networks, and DCNs. The general model of an optical communication system with direct photo-detection is shown in Fig. 2.1. It is comprised of 3-main blocks, namely, transmitter (Tx) block, fiber-optic channel,

and receiver (Rx) block.

The Tx block consists of a directly modulated or externally modulated laser. For baud rates 10 Gbaud, the directly modulated laser (DML) is simple and cost effective. However, for higher baud rates, externally modulated lasers (EML) such as those using dual drive Mach-Zender modulator are utilized. Despite its relatively high cost, EML has the advantage of negligible induced frequency chirp waveform distortions, as compared to the DML (Krehlik, 2007). The noise $n(t)$ due to transmitter is assumed to be white and Gaussian. We consider the generalized transmitter model (Kumar and Deen, 2014) in which the output of the transmitter is given by

$$x(t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} a_k F(t - kT) + n(t), \quad (2.1)$$

where a_k are independent and identically distributed (i.i.d.) input data symbols. N is the total number of transmitted symbols, T^{-1} is the symbol rate and $F(t)$ is the complex pulse shape. For the case of DML, $F(t)$ is given as (Krehlik, 2007):

$$F(t) = \sqrt{P(t)} \exp(-i\phi(t)), \quad (2.2)$$

where $P(t)$ is the output power distribution and $i = \sqrt{-1}$. The phase $\phi(t)$ is introduced as a result of the DML-induced frequency chirping. It is expressed by the integral (Krehlik, 2007):

$$\phi(t) = \frac{\alpha}{2} \int_0^t \left(\frac{1}{P(t)} \frac{dP(t)}{dt} + K_v P(t) \right) dt, \quad (2.3)$$

where α is the line enhancement factor and K_v is the adiabatic chirp coefficient. The

first and second terms in (2.3) represent the transient chirp and the adiabatic chirp, respectively. On the other hand, for the case of EML, $F(t)$ is given as (Kumar and Deen, 2014):

$$F(t) = \sqrt{P(t)}. \quad (2.4)$$

The channel block is assumed as a linear single mode fiber (SMF) that is modeled by the transfer function (Agrawal, 2007; Kumar and Deen, 2014):

$$H_{fiber}(f) = \exp\left(\frac{i}{2}\beta_2(2\pi f)^2 L\right), \quad (2.5)$$

where β_2 is the group velocity dispersion coefficient and L is the optical fiber length. The output of the fiber passes through an optical bandpass filter (BPF) of bandwidth B_o . The signal at the receiver (Rx) front end is expressed as

$$r(t) = \left(x(t) * h_{fiber}(t)\right) * h_o(t), \quad (2.6)$$

where the operation $*$ denotes convolution and $h_{fiber}(t)$ and $h_o(t)$ are the impulse responses of the fiber and the optical BPF, respectively.

The receiver (Rx) block includes the front-end unit and the electronic digital signal processing (DSP) unit. The front-end unit consists of a photodetector which translates optical signal to electrical signal, and an electrical low pass filter (LPF) with 3 dB bandwidth B_e . The receiver noise $n_e(t)$ includes shot noise and thermal noise. The photo-detection is a direct detection, which is widely used in short-reach systems. It is preferred over coherent detection in these applications, due to its simplicity and low cost (Agrawal, 2007; Kumar and Deen, 2014). At the output of

the electrical LPF, the signal is given by

$$y(t) = \left(|r(t)|^2 + n_e(t) \right) * h_e(t), \quad (2.7)$$

where $h_e(t)$ is the impulse response of the electrical LPF.

It is clear that even after ignoring the nonlinear effect (Kerr effect) of the optical fiber channel, the linear ISI impairments imposed on the optical received signal $r(t)$ turn into nonlinear distortions in the electrical domain signal $y(t)$. This is due to the loss of the phase information of the received signal. Therefore, it is necessary to have a nonlinear electronic equalizer within the DSP unit. This equalizer should be capable of compensating for these impairments in order to detect or reconstruct the transmitted data effectively.

The DSP unit is comprised of an analog to digital converter (ADC), a dispersion compensation equalizer, and a decision circuit. The sampling rate at the ADC is $f_s^d = 1/T_s^d$, where T_s^d is the tap spacing required by the equalizer. The output samples of the ADC are denoted by $y_k = y(kT_s^d)$, $k = 0, 1, 2, \dots$. It is shown in (Franceschini *et al.*, 2007; Curri *et al.*, 2004), according to the Nyquist condition, that a sampling rate of two samples per symbol is sufficient for accurate reconstruction of time-continuous signals. The equalizer block is described in Section 2.2. Finally, the decision circuit classifies the equalized data \hat{X}_k to the nearest symbol \hat{a}_k , according to certain decision threshold/function.

In the next Section, the structure of our ANN-NFFE is introduced. We describe our proposed extensive training routine to determine the parameters of the equalizer. The generation process of the required training input-output data set is explained. We also describe the methodology of the ANN-NFFE to combat the ISI impairments

during the transmission of actual data.

2.2 ANN Equalizer

The objective of this work is to implement an inexpensive and simple equalizer. This equalizer should be able to effectively compensate for nonlinear distortions as well as to provide low computational cost. We suggest an equalizer that mainly consists of an artificial neural network (ANN) (Bakr, 2013). It represents a nonlinear mapping from the space of the input parameters $\mathbf{Y} \in \mathbb{R}^n$ to the space of output response $\hat{X} \in \mathbb{R}$, where \mathbf{Y} is the input vector of equalizer. The vector \mathbf{Y} represents the received distorted data and n is the required number of equalizer taps. The general configuration of the proposed ANN nonlinear feed forward equalizer (NFFE) is shown in Fig. 2.2. It is comprised of 2-layers. These two layers, the hidden and the output layer, have m -nodes and 1-node, respectively. Without loss of generality, we consider one single hidden layer. The universal approximation theorem (Hornik *et al.*, 1989) shows that an ANN with only one hidden layer is sufficient to model any nonlinear mapping, provided that the correct number of hidden layer nodes is used.

Weighted versions of the input vector \mathbf{Y} are applied to each node of the hidden layer. A scalar nonlinear activation function of the hidden layer $f_h(\cdot)$ is then applied on the summation of each hidden layer node, resulting in an output of the form:

$$z_j = f_h \left(\sum_{l=1}^n w_{jl}^h y_l \right), \quad j = 1, 2, \dots, m, \quad (2.8)$$

where w_{jl}^h is the weight assigned to the connection between l^{th} input y_l and j^{th} node of hidden layer.

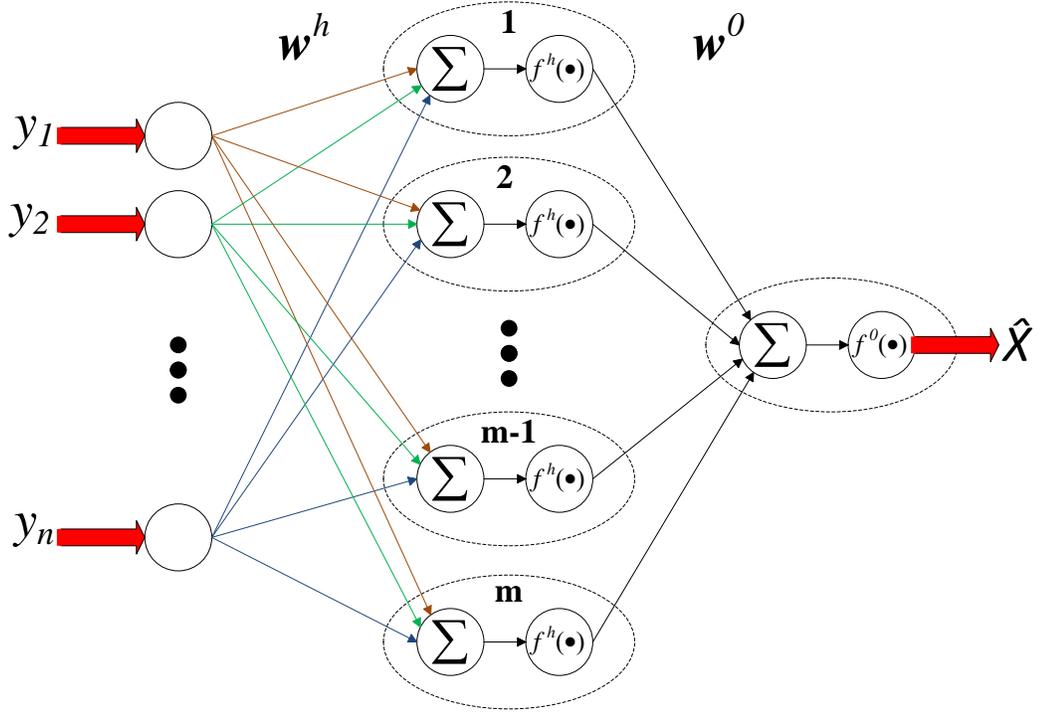


Figure 2.2: Basic configuration of the introduced ANN-NFFE.

Similarly, the hidden layer outputs z_j are assigned to different weights and their summation triggers an output layer activation function $f_0(\cdot)$, to produce the final ANN output \hat{X} :

$$\hat{X} = f_0 \left(\sum_{j=1}^m w_j^0 z_j \right), \quad (2.9)$$

where w_j^0 is the weight assigned to the connection between j^{th} node of the hidden layer and the output layer node. We define the weights vector $\mathbf{W} = [\mathbf{W}_h^T \mathbf{W}_0^T]^T$, where $\mathbf{W}_h \in \mathbb{R}^{nm}$ is the set of weights connecting the input parameters to the hidden layer nodes and $\mathbf{W}_0 \in \mathbb{R}^m$ is the set of weights connecting the hidden layer nodes to the output layer nodes.

In order simulate the required functional response of the equalizer, the ANN is first

trained by using a sufficient set of known input-output training data pairs. Through this training process, the weight values of the ANN-NFFE are optimized. Thereafter, the ANN-NFFE estimates the output response of inputs beyond the given training data, following the desired mapping.

The ANN-NFFE runs in two different modes, namely, the training mode and the transmission mode. In the training mode a set of known symbols is transmitted through the channel and the weights \mathbf{W} of the equalizer are adjusted such that the equalizer models the inverse response of the channel, i.e., $\hat{X}_k = X_k$, for all the transmitted training symbols. After the optimum values of weights have been set, the ANN-NFFE switches to the transmission mode where actual data (unknown to the receiver) are transmitted.

2.2.1 Training Mode

The training mode is essentially an optimization problem. This mode aims at achieving the optimal ANN weights \mathbf{W}^* that efficiently implement the required equalization mapping: $\mathbb{R}^n \rightarrow \mathbb{R}$. In other words, our goal in this mode is training the ANN equalizer to model the inverse response of the optical channel. The training process uses a finite set of N_{tr} input-output training pairs $\{\mathbf{Y}_k, X_k\}$, $k = 1, 2, \dots, N_{tr}$, where \mathbf{Y}_k is the k^{th} equalizer input, given by $\mathbf{Y}_k = \left[y_{-\frac{N_{tr}}{2}+k-1} \quad y_{-\frac{N_{tr}}{2}+k} \cdots y_{-\frac{N_{tr}}{2}+k+n-2} \right]^T$, and X_k is a delayed version, by time shift τ , of the k^{th} time instant transmitted symbol at the middle of each training scheme, i.e., $X_k = x(kT - \tau)$. The training step is thus cast as the optimization problem:

$$\underset{\mathbf{W}}{\text{minimize}} \quad f(\mathbf{W}) = \sum_{k=1}^{N_{tr}} \|\hat{X}_k(\mathbf{W}, \mathbf{Y}_k) - X_k\|_2^2, \quad (2.10)$$

where the error $E_k = \hat{X}_k(\mathbf{W}, \mathbf{Y}_k) - X_k$ represents the difference between the output of the ANN-NFFE and its corresponding desired output for the k^{th} training symbol. The square of the Euclidean norm is utilized in (2.10).

In order to have an efficient equalization performance, the length of the ANN-NFFE input vector $\mathbf{Y} \in \mathbb{R}^n$ as well as the required number of training data N_{tr} should be properly determined. Depending on the modulation format, the optical fiber length L , and its dispersion parameter β_2 ; the number of neighbouring symbols that may interfere with the current un-equalized symbol (i.e. ISI span) can be predicted. This determines values of n and N_{tr} . In practice, the number of equalizer taps may be estimated by performing a convergence analysis. First, we choose the minimum possible value $n = 3$, and train the ANN-NFFE, then we test the equalizer performance. Afterwards, we gradually increase n until the BER value saturates.

An instance of determining the number of training data N_{tr} is as follows: consider an M -ary modulated data transmitted through an optical communication system that introduces certain chromatic dispersion. As a result, assume that the number of interfering symbols is K . In this case, the length of equalizer input should be given as $n = 2K + 1$ and the number of training data samples should be $N_{tr} = M^n$. In other words $(n \times M^n)$ -different input training symbols and $(1 \times M^n)$ -corresponding output symbols are generated. However, such training does not include all possible patterns that might be received during transmission mode. The outer symbols in each pattern are still affected by the tail symbols of its successive neighbouring patterns. This results in receiving schemes not included within the M^n -patterns. For a complete training data set, we thus add K -guard symbols around the sides of each

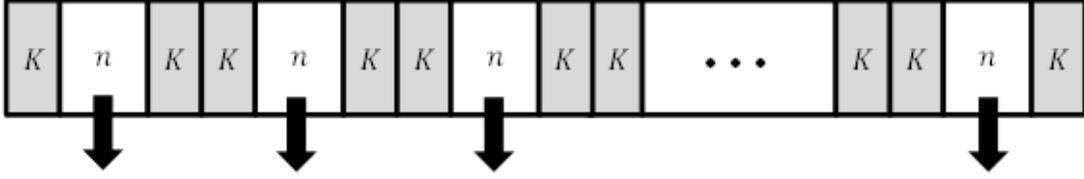


Figure 2.3: Schematic for the generation of the complete set training data. K is the number of interfering symbols and $n = 2K + 1$ is the required number of equalizer taps. The arrows refer to the actual training data, used during the ANN weights adjustment, after discarding the redundant guard data.

scheme, as shown in Fig. 2.3. In other words, we first generate $((n + 2K) \times M^{n+2K})$ -different symbols, then after transmitting them, we discard those guard symbols, as illustrated on Fig. 2.3. The final number of training schemes is thus $N_{tr} = M^{n+2K}$. A total of $(n \times M^{n+2K})$ -input training symbols and $(1 \times M^{n+2K})$ -corresponding output training symbols are utilized. This extensive training guarantees that all possible combination patterns that might be received during the transmission mode are taken into account.

Once the input-output training data have been chosen, the optimal ANN-NFFE weights \mathbf{W}^* are determined by solving the optimization problem (2.10). This problem is effectively solved using any gradient-based optimization technique. In this work, the Levenberg-Marquardt method is used (Bakr, 2013). The Levenberg-Marquardt is one of the most effective gradient-based algorithm that is widely used to solve training problems of ANNs (Hagan and Menhaj, 1994). It utilizes a step that is a combination between the steepest-decent step and the Gauss-Newton step. That provides the best trade-off between the global convergence and the convergence speed, achieved by the steepest-decent algorithm and the Gauss-Newton algorithm. Given a current weights vector $\mathbf{W}^{(i)}$, the update formula of the Levenberg-Marquardt algorithm is

given by (Bakr, 2013; Hagan and Menhaj, 1994):

$$\mathbf{W}^{(i+1)} = \mathbf{W}^{(i)} - (\mathbf{H}^{(i)} + \mu \mathbf{I})^{-1} \nabla f^{(i)}, \quad i = 0, 1, 2, \dots, \quad (2.11)$$

where $\nabla f^{(i)}$ and $\mathbf{H}^{(i)}$ are the gradient vector and the Hessian matrix of the objective function $f(\mathbf{W})$ at the current weights $\mathbf{W}^{(i)}$, respectively. μ is a positive scalar and \mathbf{I} is the identity matrix. The step (2.11) requires the first- and second-order derivatives information, $\nabla f^{(i)}$ and $\mathbf{H}^{(i)}$. This computational overhead may become formidable if these derivatives information are evaluated through finite differences. In order to accelerate the optimization process, the analytical derivatives of the objective function f with respect to \mathbf{W} are derived. Using the chain rule of differentiation (Bakr, 2013; Hagan and Menhaj, 1994),

$$\nabla f = \frac{\partial f}{\partial \mathbf{W}} = 2 \sum_{k=1}^{N_{tr}} \nabla \hat{X}_k(\mathbf{W}, \mathbf{Y}_k) E_k, \quad (2.12)$$

the gradient vector of the ANN output with respect to \mathbf{W} is $\nabla \hat{X} = [\nabla_h^T \hat{X} \quad \nabla_0^T \hat{X}]^T$, where $\nabla_h = [\partial/\partial w_{jl}^h]^T$, $j = 1, 2, \dots, m$, and $l = 1, 2, \dots, n$, and $\nabla_0 = [\partial/\partial w_j^0]^T$, $j = 1, 2, \dots, m$. The derivatives of \hat{X} with respect to the weights \mathbf{w}_{jl}^h and \mathbf{w}_j^0 are given by:

$$\frac{\partial \hat{X}}{\partial w_{jl}^h} = f'_0(z_j) w_j^0 f'_h(v_j) y_l, \quad (2.13)$$

$$\frac{\partial \hat{X}}{\partial w_j^0} = f'_0(z_j) f_h(v_j), \quad (2.14)$$

where $f'_h(\cdot)$ and $f'_0(\cdot)$ denote 1–st order derivative of the hidden and output activation

functions, respectively, and $v_j = \sum_{l=1}^n w_{jl}^h y_l$. The analytical formula of the hessian matrix \mathbf{H} is approximated as (Hagan and Menhaj, 1994):

$$\mathbf{H} = \mathbf{J}^T \mathbf{J}, \quad (2.15)$$

where the Jacobian matrix \mathbf{J} is given by

$$\mathbf{J} = \left[\nabla \hat{X}_1 \quad \nabla \hat{X}_2 \quad \nabla \hat{X}_3 \quad \dots \quad \nabla \hat{X}_{N_{tr}} \right]^T. \quad (2.16)$$

2.2.2 Transmission Mode

Once the optimal weights have been obtained, the equalizer is switched to the transmission mode where actual unknown data is transmitted. During this mode, the input vectors of the ANN-NFFE are given as

$$\mathbf{Y}_k = [y_{-K+k} \quad y_{-K+k+1} \quad \dots \quad y_k \quad \dots \quad y_{K+k-1} \quad y_{K+k}]^T, \quad k = 0, 1, 2, \dots, \quad (2.17)$$

where y_k is transmitted symbol at the k^{th} time instant, and K is the expected number of interfering symbols.

Corresponding to each input vector $\mathbf{Y}_k \in \mathbb{R}^n$, the ANN-NFFE produces a dispersion compensated version of the k^{th} time instant transmitted signal $\hat{X}_k \in \mathbb{R}$. Thereafter, the decision circuit estimates the corresponding transmitted symbol \hat{a}_k , which is assumed to be a reliable estimate of the actual transmitted symbol a_k .

The fiber dispersion value is subjected to adherent fluctuations due to variations of environmental conditions. However, these fluctuations occur at much slower rate than

the transmission data rate. Our ANN-NFFE can be simply modified to adaptively re-adjust its weights and trace these channel fluctuation. The required variation in the weights' values is small due to the limited effect of these environmental fluctuations on the whole channel response (Kumar and Deen, 2014). Therefore, the weights could be varied adaptively, in real time, after specific number of received symbols, using a small number of gradient optimization steps. In practice, only one optimization step is considered (Kumar and Deen, 2014). Furthermore, since we have no information about the transmitted signal, in this case, the output of decision circuit \hat{a}_k is used to calculate the current error of the training objective function, i.e.,

$$E_k = \hat{X}_k(\mathbf{W}, \mathbf{Y}_k) - \hat{a}_k. \quad (2.18)$$

2.3 Results

In this Section, we study the efficiency of the introduced ANN-NFFE by investigating its equalization performance for various scenarios of short-reach optical fiber communication systems with direct detection. Two modulation formats are considered, namely, non-return to zero (NRZ)-raised cosine pulse shaping on-off keying (OOK) and return to zero (RZ)-Gaussian pulse shaping OOK. The achieved results are compared to the results obtained using the standard equalization techniques (FFE, DFE and MLSE). The comparison shows that our equalizer achieves comparable equalization performance to the benchmark performance achieved by the maximum likelihood estimator (MLSE). Our equalizer, however, has significantly lower computational cost and complexity. Furthermore, we investigate the performance of our ANN-NFFE to mitigate the impairments due to fiber dispersion in DCNs.

Table 2.1: Simulation parameters

Parameter	Value
Wavelength, λ	1.55 μm
Dispersion coefficient, β_2	-21 ps^2/km
Modulation scheme	On-off keying (OOK)
Number of transmitted bits, N	16384
Sequence pattern	PRBS $2^{10} - 1$
Fiber simulation number of samples per bit, N_s	16
DSP simulation number of samples per bit, N_s^d	2
Optical filter	Gaussian band pass filter
Electrical filter	Gaussian low pass filter
Load resistance, R_L	1 $K\Omega$
Temperature	300 K
Hidden layer activation function, $f_h(\cdot)$	$\tanh(\cdot)$
Output layer activation function, $f_0(\cdot)$	1

For all results presented in this Section, the parameters of the communication channel are as summarized in Table 2.1. The number of transmitted bits and sequence pattern are selected such that the BER is stable and reliable. The effect of the DML frequency chirping (2.3) is neglected, unless otherwise is stated. The variances of shot noise and thermal noise are given by (Agrawal, 2007; Kumar and Deen, 2014):

$$\sigma_{shot}^2 = 2qIB_e, \quad (2.19)$$

$$\sigma_{thermal}^2 = 4K_BTB_e/R_L, \quad (2.20)$$

where q is the electron charge, I is the mean photo-current, B_e is the 3-dB bandwidth of the electrical LPF, K_B is the Boltzmann's constant, T is the temperature in Kelvin, and R_L is the load resistance. Note that the value of received photo-current I_r is not constant; it varies with the received OSNR. In our simulations, we keep the average

Table 2.2: Mean photo-current versus OSNR

OSNR (dB)	Mean photo-current, I (mA)
8.0	1.50
10.0	1.30
13.5	1.10
17.0	1.00
22.0	0.95

*The mean signal power launch to fiber = 1 mW .

signal power launched to the fiber fixed at 1 mW ; however, as the noise power increases (OSNR value decreases), the value of mean photo-current increases. This is due to the fact that the value of received photo-current is proportional to the transmitter's noise-noise beating term (Kumar and Deen, 2014), where the received photo-current is given by:

$$I_r = R|r(t) + n(t)|^2, \quad (2.21)$$

where $r(t)$ and $n(t)$ are the optical signal and the noise before the photo-detector, respectively. R is the responsivity of photo-detector. Therefore,

$$I = \langle I_r \rangle \propto \langle |r(t)|^2 \rangle + \langle |n(t)|^2 \rangle, \quad (2.22)$$

where $\langle \cdot \rangle$ denotes the mean. Table 2.2 tabulates values of mean photo-current I as a function of received OSNR.

The performance of ANN-NFFE, in the presence of noise, is evaluated through the bit error rate (BER). Under the Gaussian noise assumption, the BER is related

to the Quality Factor by (Agrawal, 2007; Kumar and Deen, 2014):

$$BER = \frac{1}{2} \operatorname{erfc} \left(\frac{Q}{\sqrt{2}} \right), \quad (2.23)$$

where $\operatorname{erfc}(\cdot)$ stands for the complementary error function. The Quality Factor Q is obtained by performing many Monte Carlo simulations of the entire system. It is calculated through the formula (Agrawal, 2007; Kumar and Deen, 2014):

$$Q = \frac{I_1 - I_0}{\sigma_1 + \sigma_0}, \quad (2.24)$$

where I_q and σ_q , $q = 0, 1$, are the mean and the standard deviation, respectively, of the received signals and q refers to the type of the transmitted bit ('0' or '1'). It should be noticed that the noise in direct-detection system is not Gaussian-distributed but it has a Chi-square distribution. However, as shown in (Humblet and Azizoglu, 1991), the difference in BER obtained by considering the noise to be Gaussian-distributed and Chi-square-distributed is small. We therefore approximate the noise probability density function by a Gaussian distribution.

2.3.1 10 Gbaud Fiber Optic System

We consider a metro optical fiber communication system with bit-rate of 10 Gbps. The bandwidth of the optical and electrical filters are set as $B_0 = 50$ GHz and $B_e = 7$ GHz, respectively. The transmitted data are modulated with NRZ-OOK format and the pulse shape is a raised cosine with roll-off factor of 0.6.

We first study communication scenario with fiber length $L = 140$ km. Due to pulse broadening caused by chromatic dispersion (CD), the BER at the receiver without

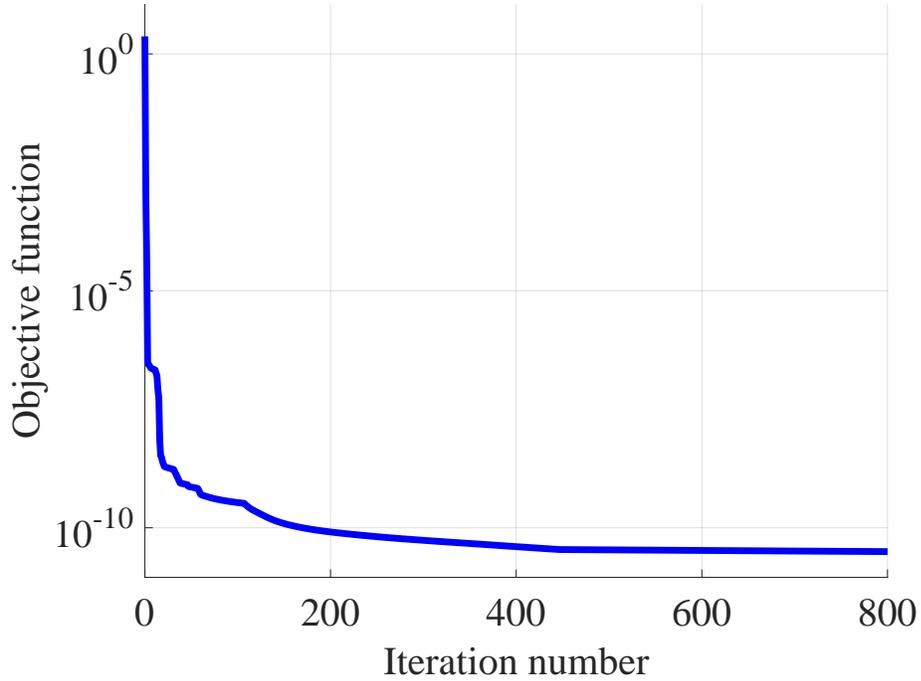


Figure 2.4: The value of training objective function $f(\mathbf{W})$ versus the number of optimization iterates. The transmitted data are NRZ-OOK with raised-cosine pulse shaping. The data rate is 10 Gbps and the fiber length is 140 km.

equalization and in the absence of noise is 4×10^{-2} . For this scenario, Fig. 2.4 shows the change of the objective function $f(\mathbf{W})$ with the number of optimization iterates, during the training process of the ANN-NFFE. The required number of equalizer taps and the number of the hidden layer nodes of the ANN-NFFE are $n = 7$ and $m = 6$, respectively. We denote such an equalizer configuration by ANN-NFFE{7,6}.

Figure 2.5 compares the BER, with and without equalization, versus swept OSNR values, where OSNR is the conventional received optical signal to noise ratio calculated in 0.1 nm bandwidth (Kumar and Deen, 2014; Agrawal, 2007). From Fig. 2.5, we see that a BER = 10^{-9} is achieved at $OSNR = 14.3$ dB and higher.

In Fig. 2.6, we further explore the performance of the ANN-NFFE for optical

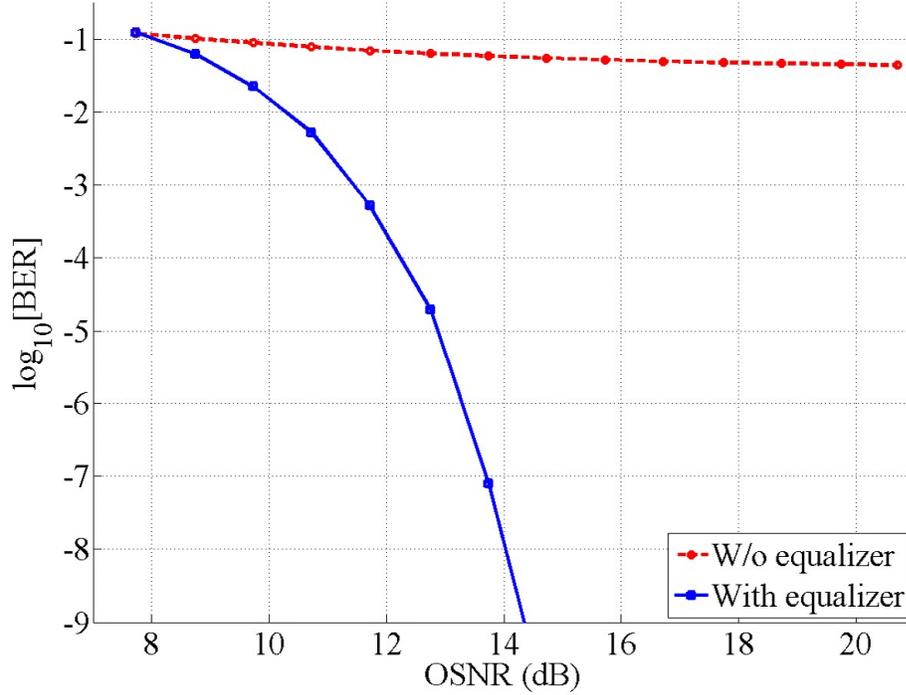


Figure 2.5: Received BER, with and without the ANN-NFFE, versus OSNR. The transmitted data are NRZ-OOK with raised-cosine pulse shaping. The data rate is 10 Gbps and the fiber length is 140 km.

fiber lengths $L = 180$ km, $L = 200$ km, and $L = 220$ km. In all these cases, the ANN-NFFE{7,6} is utilized. The values of BER, versus OSNR, are compared to back-to-back (B2B) transmission case. It is shown from Fig. 2.6 that the ANN-NFFE extends the transmission length up to 200 km while achieving $\text{BER} = 10^{-9}$ with at most 2 dB-OSNR penalty, as opposed to the B2B case. On the other hand, the 4 dB-OSNR penalty of the $L = 220$ km-case can be reduced by increasing the number of equalizer taps. However, in this case the computational cost would highly increase, as the ANN-NFFE computational cost per bit is proportional to mn .

In order to evaluate our ANN-NFFE, Fig. 2.7 plots the OSNR required to achieve

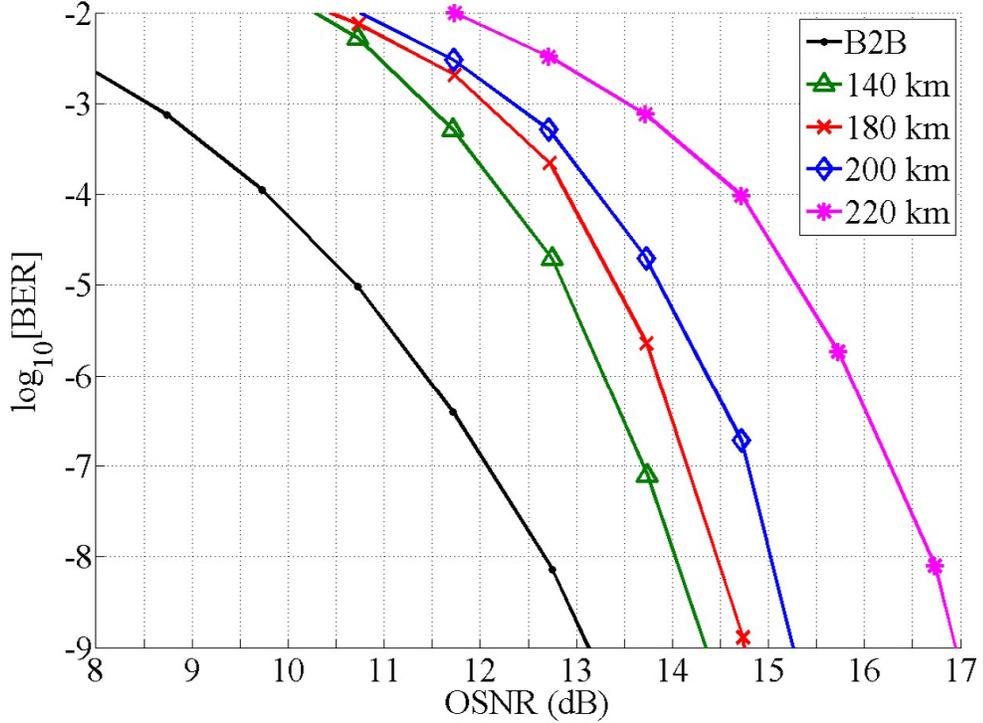


Figure 2.6: BER of the received equalized data versus OSNR. The transmitted data rate is 10 Gbps, and the modulation format is NRZ-OOK. Various transmission distances are considered and compared to the B2B case, when the ANN-NFFE is used for the equalization of the received signal.

the BER = 1×10^{-3} versus the length of optical fiber. We compare our ANN-NFFE to the linear feed forward equalizer (FFE) (Curri *et al.*, 2004), FFE with decision feedback equalizer (DFE) (Wang and Kahn, 2004), and maximum-likelihood sequence estimator (MLSE) (Foggi *et al.*, 2006). We compare the performance of ANN-NFFE{7,6} to FFE{7}, DFE{4,3}, MLSE{7} and MLSE{9}, where {·} refers to the number of FFE taps or the memory size of MLSE. In case of DFE, {·,·} denotes the numbers of its feed-forward and feed-backward taps, respectively. For a fair comparison, the same extensive training scheme, described in subsection 2.2.1, is used to train all equalizers. In Fig. 2.7, we specifically show the required OSNR for

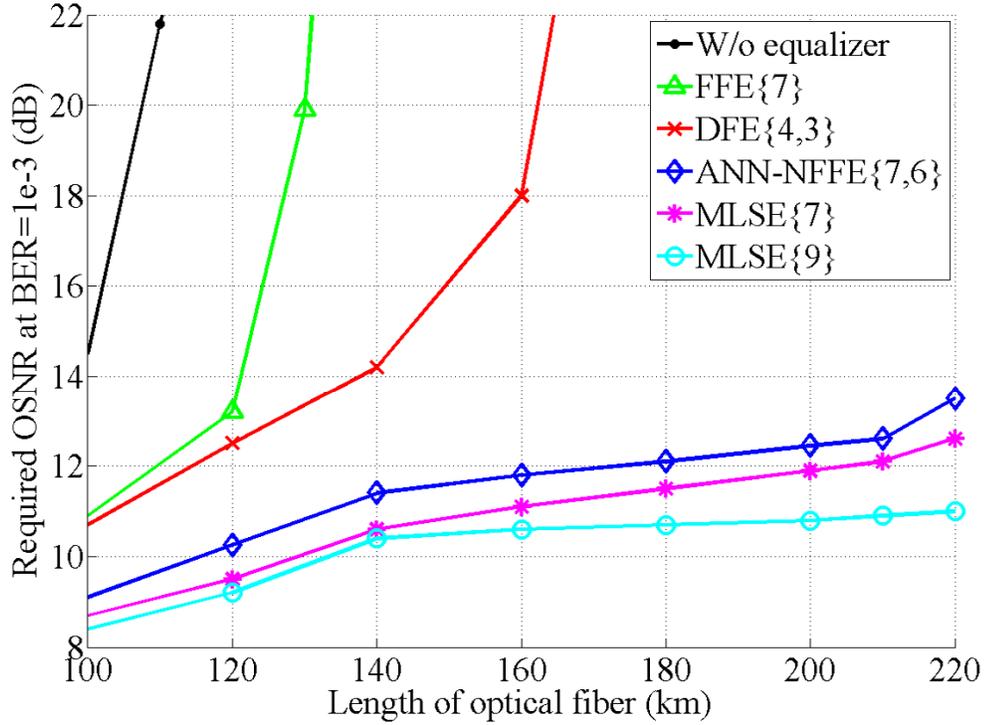


Figure 2.7: The required OSNR at received $\text{BER} = 1 \times 10^{-3}$ versus fiber optic length. A NRZ-OOK modulation format with raised-cosine pulse shaping is assumed.

BER of 10^{-3} , since the unequalized and the equalized data using FFE and DFE do not reach the level of $\text{BER} = 10^{-9}$ in most of the shown scenarios. As can be seen, FFE and DFE do not provide performance benefit since the channel (fiber channel + detection) is nonlinear. MLSE and ANN-NFFE provide performance benefit, with MLSE slightly outperforming ANN-NFFE. However, the computational cost of MLSE increases exponentially with the number of taps. The computational cost of MLSE scales as $\sim 2^n$, where n is the memory size of MLSE (Proakis, 2007). On the other hand, the computational cost of the ANN-NFFE scales as $\sim m \times n$. Particularly, in the scenarios considered in Fig. 2.7, MLSE{7} needs 128 multiplication operations per bit. Moreover, the required number of multiplication operations per bit soars to

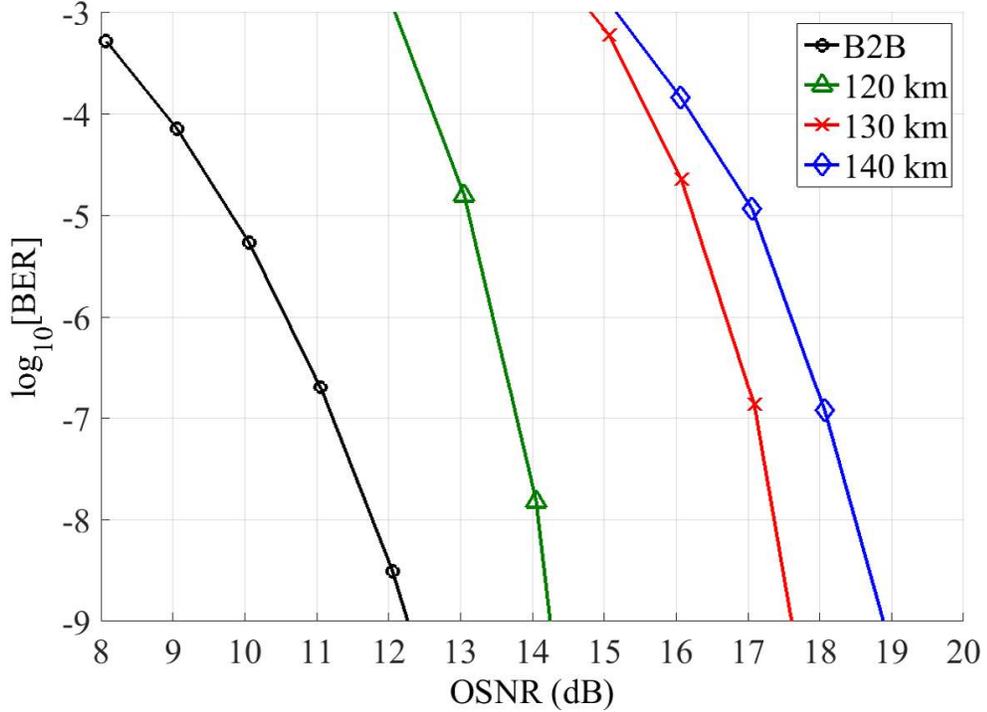


Figure 2.8: BER of the received equalized data versus OSNR. The transmitted data rate = 10 Gbps, and the modulation format is RZ-OOK with 50% duty cycle. Various transmission distances are considered and compared to the B2B case, when the ANN-NFFE is used for the equalization process.

512 in the case of MLSE{9}, whereas the ANN-NFFE{7,6} needs only 42 multiplication operations per bit. Hence, the ANN-NFFE provides a better trade-off between performance and computational cost.

In Fig. 2.8, we examine the ANN-NFFE performance for the case of return to zero (RZ) pulse with 50% duty cycle and Gaussian pulse shaping. For all the transmission distances compared in Fig. 2.8, the equalizer parameters are $n = 5$ and $m = 5$ at distances $L \leq 120$ km, and $n = 7$ and $m = 6$ at longer distances. From Fig. 2.8, it can be seen that as compared to the B2B case, to have a BER of 10^{-9} , the OSNR penalties are 2 dB and 5 dB for transmission distances 120 km and 140 km, respectively. As compared to the case of NRZ-OOK (Fig 2.6), the maximum achievable transmission

distance in the case of RZ-OOK is lower since the RZ pulse is more sensitive to CD due to its narrower width.

So far, we neglected the waveform distortion induced by the DML frequency chirping. This chirp distortion interacts with the fiber dispersion resulting in further BER performance degradation. We consider here the case of DML frequency chirping with the parameters: $\alpha = 2.4$ and $K_v = 7.9 \times 10^{12}$ Hz/W (Krehlik, 2007). The required OSNR to obtain the BER = 1×10^{-3} versus the optical fiber length is plotted in Fig. 2.9. In case of no equalization, we compare the performance of the 50% RZ-Gaussian pulse shaping system before and after taking into account the DML frequency chirping effect. It can be observed that the performance deteriorates dramatically due to the DML chirping. The feasible transmission distance is significantly limited to around 20 km as opposed to 100 km in the case of zero chirping. However, our ANN-NFFE can be adopted to extend the feasible transmission distance up to 50 km, as shown in Fig. 2.9. In this case, the ANN-NFFE parameters are $n = 5$ and $m = 5$ at $L \leq 40$ km, and $n = 7$ and $m = 6$ at longer distances.

2.3.2 28 Gbaud Fiber Optic System

In this subsection, we investigate the impact of dispersion in data center networks (DCNs) operating at 28 Gbps. Recently, this communication data rate has been used in fiber-optic links deployed in data center applications (Wei *et al.*, 2012; Lam *et al.*, 2014; Sun *et al.*, 2015), either in 28 Gbaud serial links or in 4×28 Gbaud parallel links, for implementing 100 Gbps data rate. Particularly, we consider DCNs with transmission distances of the order of a few tens of kilometers. In such scenarios, the communication configuration, described in Fig. 2.1, is still valid. Moreover, for the

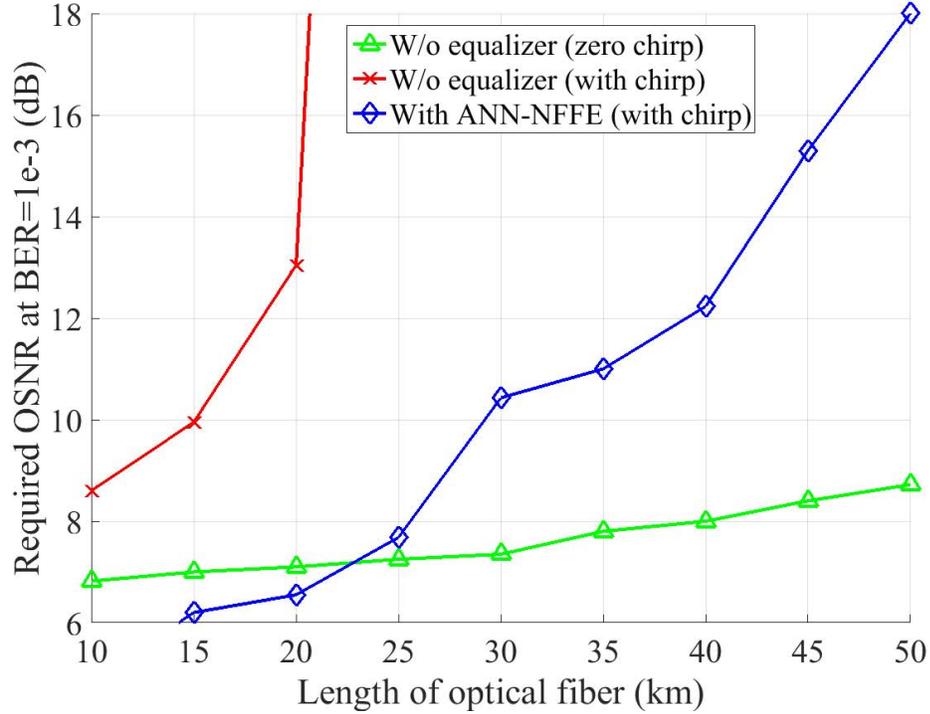


Figure 2.9: The required OSNR at received $\text{BER} = 1 \times 10^{-3}$ versus fiber optic length. The performance of the system with zero frequency chirping without equalization case is compared to the cases of frequency chirping with and without equalization. The data rate is 10 Gbps and the modulation format is RZ-OOK with 50% duty cycle. The parameters of the ANN-NFFE are $n = 5$, $m = 5$ at $L \leq 40$ km, and $n = 7$ and $m = 6$ at longer distances.

considered transmission distance range, the use of OOK modulation scheme is preferable (Wei *et al.*, 2012). The results obtained in this subsection are also applicable for metro networks operating at 28 Gbaud. The bandwidth of the optical and electrical filters are assumed as $B_0 = 100$ GHz and $B_e = 21$ GHz, respectively.

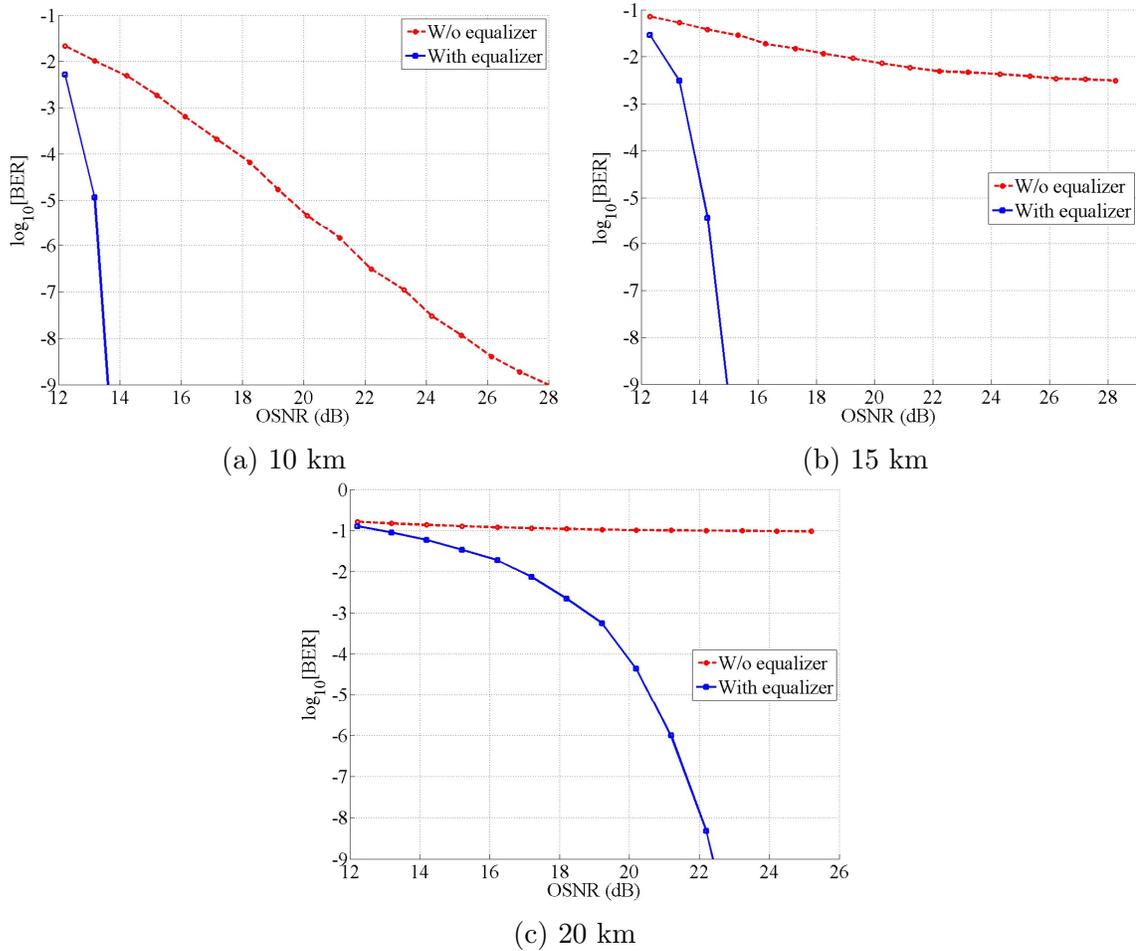


Figure 2.10: Received BER, with and without the ANN-NFFE, versus OSNR in case of transmission distance (a) 10 km, (b) 15 km, and (c) 20 km. The modulation format is NRZ-OOK with raised-cosine pulse shaping and the data rate is 28 Gbps.

Figure 2.10 depicts the received BER, with and without equalization, versus OSNR for transmission distances $L = 10$ km, 15 km and 20 km. The equalizer parameters are assumed as $n = 5$ and $m = 5$ at $L = 10$ km and $L = 15$ km, whereas $n = 7$ and $m = 6$ in case of $L = 20$ km. We notice that when the optical fiber length is less than 10 km, the ISI distortion does not limit the achievable transmission distance. However, the equalizer might still be used to reduce the OSNR penalty for achieving

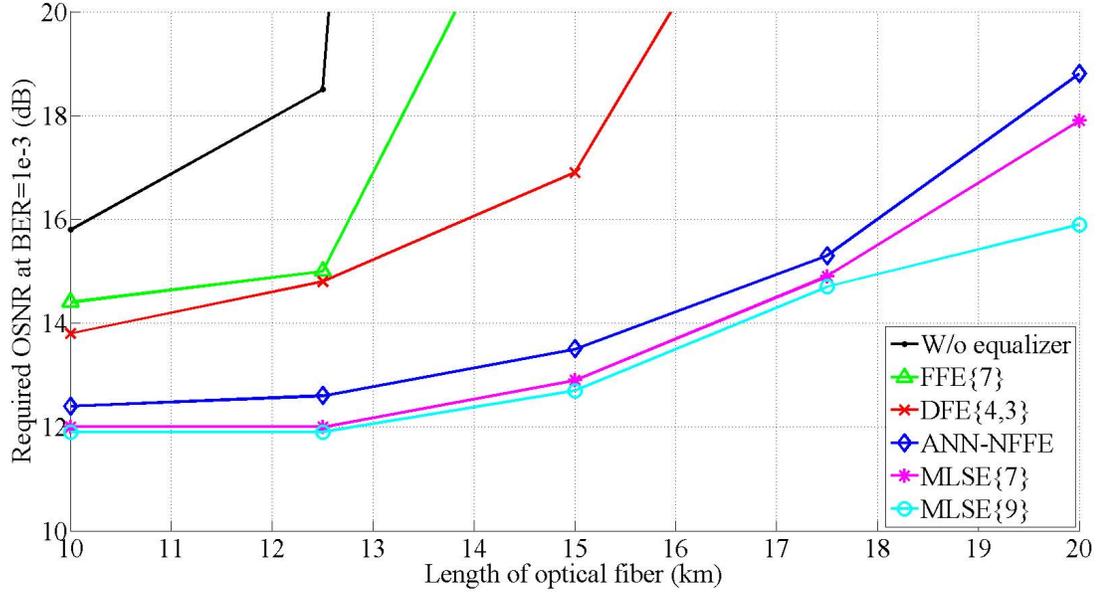


Figure 2.11: The required OSNR at received $\text{BER} = 1 \times 10^{-3}$ versus fiber optic length. The data rate is 28 Gbps. A NRZ-OOK modulation format with raised-cosine pulse shaping is assumed. The parameters of the ANN-NFFE are $n = 5$, $m = 5$ at $L \leq 15$ km, and $n = 7$, $m = 6$ at longer distances.

BER of 10^{-9} . A 14.4 dB reduction in OSNR can be achieved using the ANN-NFFE at $\text{BER} = 10^{-9}$, as shown in Fig. 2.10a. At longer distances, the distortions of fiber dispersion dramatically increase, prohibiting to have a good transmission performance without using an equalizer, even when the OSNR is high. In this case the use of digital equalizers such as ANN-NFFE becomes mandatory, in order to improve dispersion tolerance and to increase feasible distance limit. Figs. 2.10b and 2.10c show how the ANN-NFFE can mitigate fiber dispersion, extending the feasible transmission distance up to 20 km.

Figure 2.11 shows the OSNR required to obtain the $\text{BER} = 1 \times 10^{-3}$ versus the optical fiber length. The performance of the ANN-NFFE is compared to the performance of FFE{7}, DFE{4,3}, MLSE{7} and MLSE{9}. As can be seen, the ANN-NFFE

provides a performance close to that of MLSE with much lower computational cost. At most 42 multiplication operations per bit are required as opposed to 128 and 512 in cases of MLSE{7} and MLSE{9}, respectively.

2.4 Conclusion

This chapter proposed a computationally efficient nonlinear feed forward equalizer based on artificial neural networks. The introduced equalizer is used to effectively compensate for the chromatic dispersion-induced distortions in short-reach optical fiber communications systems with direct detection. The robustness and efficiency of the equalizer has been demonstrated through a number numerical examples extracted from practical short-haul optical fiber communication scenarios, e.g., metro optical fiber communication systems and data center networks. It has been shown that the proposed equalizer significantly reduces the computational cost, achieving a BER compensation performance comparable to the MLSE.

Chapter 3

An Adjoint Sensitivity Analysis Approach for the Nonlinear Schrödinger Equation

This chapter proposes, for the first time, an efficient adjoint sensitivity analysis (ASA) approach for the general time-dependent nonlinear Schrödinger equation (NLSE). Regardless of the number of design parameters, our proposed approach estimates the sensitivities of a general objective function with respect to all the design parameters using only one extra adjoint system simulation. We derive an adjoint nonlinear system of equations and develop a numerical technique to solve this adjoint problem. We also show the accuracy and efficiency of the proposed ASA algorithm through a number of optical fiber examples.

The chapter is organized as follows: Section 3.1 provides a mathematical matrix representation of the general complex NLSE in real space. In Section 3.2, we derive an adjoint system simulation corresponding to the original system simulation of the

NLSE. A full description of the introduced algorithm for solving the derived adjoint simulation is given in Section 3.3. Section 3.4 investigates the validity and efficiency of our approach through applying the introduced ASA algorithm to estimate the sensitivities of 2–optical fiber numerical examples operating at varied transmission bit rate and utilizing different modulation schemes. We also exploit the proposed ASA approach to verify the optical soliton phenomenon in an optical fiber communication system (Hasegawa and Tappert, 1973; Hasegawa and Kodama, 1981). All results obtained using our ASA approach are compared with the results of the accurate but computationally expensive central finite differences (CFD) approach. Finally, conclusions of the work are drawn in Section 3.5.

3.1 Matrix Representation of the NLSE

The light propagation in an optical fiber link is governed by the normalized nonlinear Schrödinger equation (NLSE), given by (Agrawal, 2007; Kumar and Deen, 2014):

$$-\frac{\beta_3}{6} \frac{\partial^3 u}{\partial t^3} + \frac{i\beta_2}{2} \frac{\partial^2 u}{\partial t^2} + \beta_1 \frac{\partial u}{\partial t} + \frac{\partial u}{\partial z} + \left(\frac{\alpha}{2} - i\gamma P_0 |u|^2 \right) u = u_{tx}(t) \delta(z), \quad (3.1)$$

where $u(z, t) = q/\sqrt{P_0}$ is a normalized signal of the complex envelope of the optical field $q(z, t)$, and P_0 is the peak power of the transmitted pulse. The signal $u_{tx}(t) = q_{tx}/\sqrt{P_0}$ is the normalized complex optical field envelope of the transmitter output and $i = \sqrt{-1}$. The parameter β_1 is the inverse group speed, β_2 and β_3 are the second- and third-order dispersion coefficients, respectively, α is the fiber loss coefficient, and γ is the nonlinear coefficient of the fiber. Note that in (3.1), we assume zero boundary

condition field, i.e. $u(0, t) = 0$, and the field excitation is given by the right hand side term $u_{tx}(t)\delta(z)$. Expressing u in terms of its independent real and imaginary parts ($u = u_{re} + iu_{im}$), substituting in (3.1), and separating the real and imaginary terms, we obtain:

$$\begin{aligned}
 & \begin{bmatrix} -\tilde{\beta}_3 & 0 \\ 0 & -\tilde{\beta}_3 \end{bmatrix} \begin{bmatrix} \frac{\partial^3 u_{re}}{\partial t^3} \\ \frac{\partial^3 u_{im}}{\partial t^3} \end{bmatrix} + \begin{bmatrix} 0 & -\tilde{\beta}_2 \\ \tilde{\beta}_2 & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial^2 u_{re}}{\partial t^2} \\ \frac{\partial^2 u_{im}}{\partial t^2} \end{bmatrix} + \begin{bmatrix} \tilde{\beta}_1 & 0 \\ 0 & \tilde{\beta}_1 \end{bmatrix} \begin{bmatrix} \frac{\partial u_{re}}{\partial t} \\ \frac{\partial u_{im}}{\partial t} \end{bmatrix} \\
 & + \begin{bmatrix} \partial/\partial z & 0 \\ 0 & \partial/\partial z \end{bmatrix} \begin{bmatrix} u_{re} \\ u_{im} \end{bmatrix} + \begin{bmatrix} \tilde{\alpha} & 0 \\ 0 & \tilde{\alpha} \end{bmatrix} \begin{bmatrix} u_{re} \\ u_{im} \end{bmatrix} + \begin{bmatrix} 0 & \tilde{\gamma} \\ -\tilde{\gamma} & 0 \end{bmatrix} \begin{bmatrix} u_{re} \\ u_{im} \end{bmatrix} \\
 & = \begin{bmatrix} Re \{u_{tx}\} \delta(z) \\ Im \{u_{tx}\} \delta(z) \end{bmatrix}, \tag{3.2}
 \end{aligned}$$

where $\tilde{\beta}_1 = \beta_1$, $\tilde{\beta}_2 = \beta_2/2$, $\tilde{\beta}_3 = \beta_3/6$, $\tilde{\alpha} = \alpha/2$, and $\tilde{\gamma} = \tilde{\gamma}(u_{re}, u_{im}) = \gamma P_0(u_{re}^2 + u_{im}^2)$. The computational domain is discretized to M equally-spaced spatial cells, with a step size of h . Writing (3.2) for all the discretization cells in the computational domain and approximating the derivatives using finite differences, Eq. (3.2) is rewritten for the entire domain as:

$$\mathbf{B}_3 \ddot{\mathbf{U}} + \mathbf{B}_2 \dot{\mathbf{U}} + \mathbf{B}_1 \mathbf{U} + \mathbf{K}_c \mathbf{U} + \mathbf{A} \mathbf{U} + \mathbf{\Gamma}(\mathbf{U}) \mathbf{U} = \mathbf{U}_{in}, \tag{3.3}$$

where $\mathbf{U} = [\mathbf{u}_{re}^T \ \mathbf{u}_{im}^T]^T$ is the system state vector, $\mathbf{U}_{in} = [Re \{u_{tx}\} \mathbf{e}_1^T \ Im \{u_{tx}\} \mathbf{e}_1^T]^T$ is the excitation vector, and $\mathbf{e}_1 = [1 \ 0 \ 0 \ \dots \ 0]^T$ is an elementary column vector, and \cdot denotes time derivative. The constant matrix \mathbf{K}_c approximates the spatial derivatives of (3.2). The system matrices are given as follows: $\mathbf{B}_1 = \tilde{\beta}_1 \mathbf{I}_{2M}$, $\mathbf{B}_2 =$

$\tilde{\beta}_2 \begin{bmatrix} \mathbf{0} & -\mathbf{I}_M \\ \mathbf{I}_M & \mathbf{0} \end{bmatrix}$, $\mathbf{B}_3 = -\tilde{\beta}_3 \mathbf{I}_{2M}$, and $\mathbf{A} = \tilde{\alpha} \mathbf{I}_{2M}$, where \mathbf{I}_M is an identity matrix of

size M . The matrix $\mathbf{\Gamma}(\mathbf{U}) = \begin{bmatrix} \mathbf{0} & \tilde{\gamma} \\ -\tilde{\gamma} & \mathbf{0} \end{bmatrix}$ is the nonlinear matrix, where $\tilde{\gamma}$ is a diagonal matrix whose j^{th} entry is given by: $\gamma P_0 [u_{re}^2(jh) + u_{im}^2(jh)]$, $j = 0, 1, \dots, M-1$. It can be shown that: \mathbf{B}_1 , \mathbf{B}_3 and \mathbf{A} are symmetric matrices, whereas \mathbf{B}_2 , \mathbf{K}_c and $\mathbf{\Gamma}$ are skew-symmetric matrices. Typically, Eq (3.3) is solved numerically using the split-step Fourier scheme (SSFS) method (Sinkin *et al.*, 2003; Shao *et al.*, 2014; Deiterding and Poole, 2016).

We derive next the adjoint simulation problem corresponding to the original problem (3.3).

3.2 ASA for the NLSE

We aim at estimating the sensitivities of an objective function of the form (Bakr *et al.*, 2017):

$$F = \int_{-T_m}^{T_m} \psi(\mathbf{x}, \mathbf{U}) dt, \quad (3.4)$$

where the kernel of the objective function integral ψ is a predefined function, $\mathbf{x} \in \mathbb{R}^N$ is the design parameters vector of the optical fiber, and T_m is half of the computational time window size. The analytic sensitivity of (3.4) with respect to the k^{th} parameter

x_k , $k = 1, 2, \dots, N$, is given by

$$\frac{\partial F}{\partial x_k} = \int_{-T_m}^{T_m} \frac{\partial^e \psi}{\partial x_k} dt + \int_{-T_m}^{T_m} \left(\frac{\psi}{\partial \mathbf{U}} \right)^T \frac{\partial \mathbf{U}}{\partial x_k} dt, \quad (3.5)$$

where $\partial^e / \partial x_k$ denotes the explicit dependence. The analytic expression in (3.5) cannot be evaluated unless the vector $\partial \mathbf{U} / \partial x_k$ is known for every time step. The classical approach for evaluating (3.5) involves repeatedly simulating perturbed structures, for each designable parameter x_k , $k = 1, 2, \dots, N$. However, this required overhead computational cost which scales linearly with the number of design parameters is significant and impractical. For example, in the case of DBP, the NLSE should be solved $2N$ times to evaluate the gradient $\partial \mathbf{U} / \partial x_k$, $k = 1, 2, \dots, N$.

Alternatively, an adjoint sensitivity analysis (ASA) approach can be derived to estimate the implicit derivative in (3.5) with respect to all the design parameters using only one extra system simulation. The derivation starts by differentiating the original NLSE system (3.3) with respect to the k^{th} parameter x_k and shifting the known terms to the right hand side to get:

$$\mathbf{B}_3 \frac{\partial^4 \mathbf{U}}{\partial x_k \partial t^3} + \mathbf{B}_2 \frac{\partial^3 \mathbf{U}}{\partial x_k \partial t^2} + \mathbf{B}_1 \frac{\partial^2 \mathbf{U}}{\partial x_k \partial t} + \mathbf{K}_c \frac{\partial \mathbf{U}}{\partial x_k} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x_k} + \frac{\partial(\mathbf{\Gamma} \bar{\mathbf{U}})}{\partial \mathbf{U}^T} \frac{\partial \mathbf{U}}{\partial x_k} + \mathbf{\Gamma} \frac{\partial \mathbf{U}}{\partial x_k} = -\mathbf{R}_k, \quad (3.6)$$

where $\bar{\mathbf{U}}$ is the nominal value of \mathbf{U} , treated as constant during the differentiation operation in (3.6). It is clear that all the system matrices of (3.3) are constants with time, except for the nonlinear coefficients matrix $\mathbf{\Gamma}(\mathbf{U})$. This dependency arises from the fact that the nonlinearity term, $\tilde{\gamma} = \gamma P_0 |u|^2$, is a function of the local field. It can be noted that \mathbf{U}_{in} is independent of the system design parameters x_k . The residue

vector \mathbf{R}_k corresponding to the k^{th} parameter is given by:

$$\mathbf{R}_k = \frac{\partial^e \mathbf{B}_3}{\partial x_k} \ddot{\mathbf{U}} + \frac{\partial^e \mathbf{B}_2}{\partial x_k} \dot{\mathbf{U}} + \frac{\partial^e \mathbf{B}_1}{\partial x_k} \dot{\mathbf{U}} + \frac{\partial^e \mathbf{A}}{\partial x_k} \mathbf{U} + \frac{\partial^e \boldsymbol{\Gamma}}{\partial x_k} \mathbf{U}. \quad (3.7)$$

Multiplying both sides of (3.6) by the yet-to-be determined temporal adjoint vector $\boldsymbol{\lambda}$ and integrating over the total simulation time, we have:

$$\int_{-T_m}^{T_m} \boldsymbol{\lambda}^T \left(\mathbf{B}_3 \frac{\partial^4 \mathbf{U}}{\partial x_k \partial t^3} + \mathbf{B}_2 \frac{\partial^3 \mathbf{U}}{\partial x_k \partial t^2} + \mathbf{B}_1 \frac{\partial^2 \mathbf{U}}{\partial x_k \partial t} + \left(\mathbf{K}_c + \mathbf{A} + \frac{\partial(\boldsymbol{\Gamma}\bar{\mathbf{U}})}{\partial \mathbf{U}^T} + \boldsymbol{\Gamma} \right) \frac{\partial \mathbf{U}}{\partial x_k} \right) dt = - \int_{-T_m}^{T_m} \boldsymbol{\lambda}^T \mathbf{R}_k dt, \quad (3.8)$$

where $\boldsymbol{\lambda}(z, t) = [\boldsymbol{\lambda}_{re}^T \ \boldsymbol{\lambda}_{im}^T]^T$ is the adjoint state vector. Integrating the first, second and third terms (mixed derivative terms) in (3.8) by parts yields:

$$\begin{aligned} & \boldsymbol{\lambda}^T \mathbf{B}_3 \frac{\partial^3 \mathbf{U}}{\partial x_k \partial t^2} \Big|_{-T_m}^{T_m} + \boldsymbol{\lambda}^T \mathbf{B}_2 \frac{\partial^2 \mathbf{U}}{\partial x_k \partial t} \Big|_{-T_m}^{T_m} + \boldsymbol{\lambda}^T \mathbf{B}_1 \frac{\partial \mathbf{U}}{\partial x_k} \Big|_{-T_m}^{T_m} + \\ & \int_{-T_m}^{T_m} \left[-\dot{\boldsymbol{\lambda}}^T \mathbf{B}_3 \frac{\partial^3 \mathbf{U}}{\partial x_k \partial t^2} - \dot{\boldsymbol{\lambda}}^T \mathbf{B}_2 \frac{\partial^2 \mathbf{U}}{\partial x_k \partial t} + \left(-\dot{\boldsymbol{\lambda}}^T \mathbf{B}_1 + \boldsymbol{\lambda}^T (\mathbf{K}_c + \mathbf{A} + \frac{\partial(\boldsymbol{\Gamma}\bar{\mathbf{U}})}{\partial \mathbf{U}^T} + \boldsymbol{\Gamma}) \right) \times \right. \\ & \left. \frac{\partial \mathbf{U}}{\partial x_k} \right] dt = - \int_{-T_m}^{T_m} \boldsymbol{\lambda}^T \mathbf{R}_k dt. \end{aligned} \quad (3.9)$$

The adjoint vector $\boldsymbol{\lambda}$ is selected to have a terminal value of $\boldsymbol{\lambda}(T_m) = \mathbf{0}$. Also, the vectors \mathbf{U} , $\partial \mathbf{U} / \partial t$ and $\partial^2 \mathbf{U} / \partial t^2$ have zero initial conditions regardless of the value of the parameter x_k , $k = 1, 2, \dots, N$. It follows that the first 3-terms in (3.9) vanish and we get:

$$\int_{-T_m}^{T_m} \left[-\dot{\lambda}^T \mathbf{B}_3 \frac{\partial^3 \mathbf{U}}{\partial x_k \partial t^2} - \dot{\lambda}^T \mathbf{B}_2 \frac{\partial^2 \mathbf{U}}{\partial x_k \partial t} + \left(-\dot{\lambda}^T \mathbf{B}_1 + \lambda^T \left(\mathbf{K}_c + \mathbf{A} + \frac{\partial(\Gamma \bar{\mathbf{U}})}{\partial \mathbf{U}^T} + \Gamma \right) \right) \times \frac{\partial \mathbf{U}}{\partial x_k} \right] dt = - \int_{-T_m}^{T_m} \lambda^T \mathbf{R}_k dt. \quad (3.10)$$

We repeat the integration by parts 2-times again until we have no more mixed derivative terms, and by taking into account that $\dot{\lambda}(T_m) = \mathbf{0}$ and $\ddot{\lambda}(T_m) = \mathbf{0}$, Eq. (3.10) becomes:

$$\int_{-T_m}^{T_m} \left(-\ddot{\lambda}^T \mathbf{B}_3 + \ddot{\lambda}^T \mathbf{B}_2 - \dot{\lambda}^T \mathbf{B}_1 + \lambda^T \left(\mathbf{K}_c + \mathbf{A} + \frac{\partial(\Gamma \bar{\mathbf{U}})}{\partial \mathbf{U}^T} + \Gamma \right) \right) \frac{\partial \mathbf{U}}{\partial x_k} dt = - \int_{-T_m}^{T_m} \lambda^T \mathbf{R}_k dt. \quad (3.11)$$

Notice that the second term on the right-hand side of (3.5) has the same form as the left-hand side of (3.11). Equating these terms, and taking the transpose of both two sides, we thus obtain the adjoint system problem, given as:

$$-\mathbf{B}_3 \frac{\partial^3 \lambda}{\partial t^3} - \mathbf{B}_2 \frac{\partial^2 \lambda}{\partial t^2} - \mathbf{B}_1 \frac{\partial \lambda}{\partial t} - \mathbf{K}_c \lambda + \mathbf{A} \lambda + \Gamma^\lambda(\mathbf{U}) \lambda = \mathbf{Q}_{in}^\lambda, \quad (3.12)$$

where \mathbf{Q}_{in}^λ is the adjoint excitation given by: $\mathbf{Q}_{in}^\lambda(t) = \psi / \partial \mathbf{U} = [\psi / \partial \mathbf{u}_{re}^T \quad \psi / \partial \mathbf{u}_{im}^T]^T$.

The nonlinear matrix of the adjoint problem Γ^λ is given as:

$$\Gamma^\lambda(\mathbf{U}) = \frac{\partial(\Gamma \bar{\mathbf{U}})^T}{\partial \mathbf{U}} - \Gamma = \begin{bmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & -\mathbf{a} \end{bmatrix}, \quad (3.13)$$

where \mathbf{a} , \mathbf{b} and \mathbf{c} are diagonal matrices whose j^{th} diagonal elements are given as follows:

$$a_{jj} = 2\gamma P_0 u_{re}(jh)u_{im}(jh), \quad (3.14a)$$

$$b_{jj} = \gamma P_0 [3u_{re}^2(jh) + u_{im}^2(jh)], \quad (3.14b)$$

$$c_{jj} = \gamma P_0 [u_{re}^2(jh) + 3u_{im}^2(jh)], \quad (3.14c)$$

where $j = 0, 1, \dots, M-1$, and M is the total number of discretization cells utilized in the original simulation (3.3). Equation (3.12) describes the evolution of the adjoint variable vector $\boldsymbol{\lambda}$ over time and space. It can be noticed that, excluding the nonlinear terms ($\mathbf{F}\mathbf{U}$ and $\mathbf{F}^\lambda\boldsymbol{\lambda}$), the adjoint problem (3.12) has a similar form to the original problem (3.3), but with different values of the system parameters (opposite signs of the β_1 , β_2 , and β_3 parameters) and different excitation vector. The adjoint excitation in (3.12) is the derivative of the objective function kernel ψ with respect to the original system state vector \mathbf{U} . It should also be clear that the adjoint problem is solved in a reversed spatial direction, $Z = L - z$, where L is the optical fiber length. It follows that the linear part of the adjoint system can be solved using the same numerical solver, i.e., the split-step Fourier scheme (SSFS), used to solve the original system (Sinkin *et al.*, 2003; Shao *et al.*, 2014; Deiterding and Poole, 2016). However, solving (3.12) for the nonlinear part requires a modification to the SSFS algorithm. We discuss later, in Section 3.3 the required modification to solve the adjoint problem.

Once the original and adjoint fields are evaluated for every time step in all spatial points of the computational domain, through solving (3.3) and (3.12), respectively,

the adjoint sensitivity of the objective function F , given by (3.4), with respect to the k^{th} parameter x_k , is obtained by:

$$\frac{\partial F}{\partial x_k} = \int_{-T_m}^{T_m} \frac{\partial^e \psi}{\partial x_k} dt - \int_{-T_m}^{T_m} \boldsymbol{\lambda}^T \mathbf{R}_k dt, \quad k = 1, 2, \dots, N. \quad (3.15)$$

where the residue vector \mathbf{R}_k is defined in (3.7). It should be clear that all explicit derivatives arising in \mathbf{R}_k are known. They can be evaluated using the definitions of the original system matrices \mathbf{B}_1 , \mathbf{B}_2 , \mathbf{B}_3 , \mathbf{A} and $\boldsymbol{\Gamma}$. For instance, assume that the current design parameter is the second dispersion coefficient (i.e. $x_k = \beta_2$), then all terms of (3.7) vanish except for the explicit derivative term of \mathbf{B}_2 , i.e., $\mathbf{R}_k = \partial^e \mathbf{B}_2 / \partial x_k = -0.5 \mathbf{I}_{2M}$.

3.2.1 Algorithm

The NLSE-based ASA approach given by (3.3)–(3.15) can be summarized by the following algorithm:

1. First, the kernel of the objective function ψ , given in (3.4), and the observation domain are defined. Note that the observation domain is a subset of the computational domain. Practically, in optical fiber simulation problems, the objective function is defined in terms of the received signal. Hence, the observation domain is typically defined as the output discretization cell.
2. The original NLSE simulation (3.3) is executed in the forward spatial direction z . During this step, the following data are calculated and stored at all time steps:

- (a) The adjoint excitation $\partial\psi/\partial\mathbf{U}$ in the observation domain.
- (b) The values of the original field \mathbf{U} at all cells of the entire computational domain.
- (c) The explicit derivatives of the system matrices, $\partial^e\mathbf{B}_1/\partial x_k$, $\partial^e\mathbf{B}_2/\partial x_k$, $\partial^e\mathbf{B}_3/\partial x_k$, $\partial^e\mathbf{A}/\partial x_k$, and $\partial^e\mathbf{\Gamma}/\partial x_k$, for $k = 1, 2, \dots, N$, corresponding to all perturbation cells, i.e., the cells whose material parameters' values are perturbed due to x_k , $k = 1, 2, \dots, N$, perturbation.
- (d) The components of $\dot{\mathbf{U}}$, $\ddot{\mathbf{U}}$ and $\ddot{\mathbf{U}}$ at all cells of the perturbation domain.

Note that the data stored in (a) and (b) will be used later in the adjoint simulation. The data stored in (c) and (d) will be utilized, in step 4, to calculate the residue vector \mathbf{R}_k required for evaluating the objective function sensitivities.

3. The adjoint simulation (3.12) is executed in the backward spatial direction Z . During the simulation, the values of the adjoint field $\boldsymbol{\lambda}$ in the perturbation domain are stored for all time steps.
4. Finally, the adjoint sensitivity of the objective function F is evaluated through the formula given in (3.15).

It is worth emphasizing that the computational CPU time gain achieved by the ASA approach is at the expense of memory storage. In other words, there is a trade-off between computational time and memory storage. As illustrated in the proposed algorithm, the ASA algorithm calculates and stores some data, during original and adjoint simulations, to be accessed and used later during the adjoint simulation to evaluate the adjoint sensitivities. For instance, the original and adjoint fields at

all perturbation cells need to be stored during original and adjoint simulations. In contrast, the FD approaches do not store these information. Therefore, the proposed ASA algorithm significantly reduces the computational time cost required by FD approaches, at the expense of using extra memory storage.

Next, we discuss the solution of the adjoint problem (3.12).

3.3 Adjoint Problem Solution

In order to develop a numerical solution method of the adjoint problem, we first rewrite its corresponding system of equations at a certain spatial grid point inside the computational domain as follows:

$$\frac{\partial}{\partial Z} \widehat{\boldsymbol{\lambda}} - \left(\widehat{\boldsymbol{D}} + \widehat{\boldsymbol{N}} \right) \widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\lambda}}_{in} \delta(Z, t), \quad (3.16)$$

where $\widehat{\boldsymbol{\lambda}}(Z, t) = [\lambda_{re} \ \lambda_{im}]^T$ is the vector of real and imaginary adjoint field in a certain grid point, $\widehat{\boldsymbol{\lambda}}_{in}(t) = [\partial\psi/\partial u_{re} \ \partial\psi/\partial u_{im}]^T$ is the vector of real and imaginary parts of the adjoint excitation. In practice, the objective function F is defined in terms of the received signal. We therefore assume that the adjoint excitation is non-zero only at $z = L$ or $Z = 0$, in other words, $\widehat{\boldsymbol{\lambda}}_{in} = \widehat{\boldsymbol{\lambda}}_{in} \delta(Z)$. The operators $\widehat{\boldsymbol{D}}$ and $\widehat{\boldsymbol{N}}$ represent the linear (propagation delay, dispersion, and loss) and the nonlinear effects of the adjoint simulation, respectively. They are given by:

$$\widehat{\boldsymbol{D}} = \begin{bmatrix} -\frac{\beta_3}{6} \frac{\partial^3}{\partial t^3} + \beta_1 \frac{\partial}{\partial t} - \frac{a}{2} & -\frac{\beta_2}{2} \frac{\partial^2}{\partial t^2} \\ \frac{\beta_2}{2} \frac{\partial^2}{\partial t^2} & -\frac{\beta_3}{6} \frac{\partial^3}{\partial t^3} + \beta_1 \frac{\partial}{\partial t} - \frac{a}{2} \end{bmatrix}, \quad (3.17)$$

$$\widehat{\mathbf{N}}(u) = \begin{bmatrix} -2\gamma P_0 u_{re} u_{im} & \gamma P_0 (3u_{re}^2 + u_{im}) \\ -\gamma P_0 (u_{re}^2 + 3u_{im}^2) & 2\gamma P_0 u_{re} u_{im} \end{bmatrix}. \quad (3.18)$$

Employing the symmetric split-step scheme (Sinkin *et al.*, 2003; Shao *et al.*, 2014; Deiterding and Poole, 2016), it can be shown that the solution of (3.16) is approximated as:

$$\begin{aligned} \widehat{\boldsymbol{\lambda}}\left((j+1)h, t\right) &= \exp\left(\frac{h}{2}\widehat{\mathbf{D}}\right) \exp(h\widehat{\mathbf{N}}) \exp\left(\frac{h}{2}\widehat{\mathbf{D}}\right) \widehat{\boldsymbol{\lambda}}(jh, t) + O(h^3), \\ j &= 0, 1, 2, \dots, M-1, \end{aligned} \quad (3.19)$$

where M is the number of discretization cells, h is the spatial step size, and $\widehat{\boldsymbol{\lambda}}(0, t) = \widehat{\boldsymbol{\lambda}}_{in}$. The first linear sub-step in (3.19) is realized by writing the equivalent complex differential equation of the operation $\exp(0.5h\widehat{\mathbf{D}})\widehat{\boldsymbol{\lambda}}(jh, t)$. The sub-step linear solution denoted by $\widehat{\boldsymbol{\lambda}}^l\left((j+1/2)h, t\right)$ is then obtained using a pair of fast Fourier transforms (FFTs) (Sinkin *et al.*, 2003; Shao *et al.*, 2014; Deiterding and Poole, 2016) as follows:

$$\begin{aligned} \lambda^l\left((j+\frac{1}{2})h, t\right) &= \mathcal{F}^{-1}\left\{\exp\left[\frac{h}{2}\left(i\frac{\beta_3}{6}\omega^3 - i\frac{\beta_2}{2}\omega^2 + i\beta_1\omega - \frac{\alpha}{2}\right)\right] \mathcal{F}\{\lambda(jh, t)\}\right\}, \\ j &= 0, 1, 2, \dots, M-1, \end{aligned} \quad (3.20)$$

where $\lambda(jh, t) = \lambda_{re} + i\lambda_{im}$ is the complex adjoint field at $Z = jh$, $i = \sqrt{-1}$, and ω is the angular frequency. The operations $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$ denote Fourier and inverse Fourier transformations, respectively. Once the complex linear sub-step solution $\lambda^l\left((j+1/2)h, t\right)$ is evaluated, the vector of the sub-step solution is

given as: $\widehat{\boldsymbol{\lambda}}^l((j + 1/2)h, t) = [\text{Re}\{\lambda^l\} \ \text{Im}\{\lambda^l\}]^T$. The nonlinear step solution $\widehat{\boldsymbol{\lambda}}^{nl}((j + 1/2)h, t)$ is then obtained by solving the equivalent differential equation of the operation $\exp(h\widehat{\mathbf{N}})\widehat{\boldsymbol{\lambda}}^l((j + 1/2)h, t)$, given by:

$$\frac{\partial}{\partial Z} (\widehat{\boldsymbol{\lambda}}^{nl}) = \widehat{\mathbf{N}}\widehat{\boldsymbol{\lambda}}^{nl}(Z, t). \quad (3.21)$$

Here $\widehat{\boldsymbol{\lambda}}^{nl}(0, t) = \widehat{\boldsymbol{\lambda}}^l((j + 1/2)h, t)$. Unlike the case in the conventional SSFS when solving the original NLSE problem (Sinkin *et al.*, 2003; Shao *et al.*, 2014; Deiterding and Poole, 2016), Eq. (3.21) cannot be solved directly since the nonlinear operator matrix $\widehat{\mathbf{N}}$ here is not a diagonal or an off-diagonal matrix. However, we can convert (3.21) to a decoupled system using the eigenvalue decomposition (EVD) of $\widehat{\mathbf{N}}$. It can be shown that $\widehat{\mathbf{N}}$ is factorized as:

$$\widehat{\mathbf{N}} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}, \quad (3.22)$$

where the eigenvalues diagonal matrix $\boldsymbol{\Lambda}$ and the eigenvectors' matrix \mathbf{P} are given by:

$$\boldsymbol{\Lambda} = -i\sqrt{3}\gamma P_0 |u|^2 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad (3.23a)$$

$$\mathbf{P} = \begin{bmatrix} i & -i \\ \frac{\sqrt{3}|u|^2 + 2iu_{re}u_{im}}{|u|^2 + 2u_{re}^2} & \frac{\sqrt{3}|u|^2 - 2iu_{re}u_{im}}{|u|^2 + 2u_{re}^2} \end{bmatrix}. \quad (3.23b)$$

Substituting the EVD of $\widehat{\mathbf{N}}$ in (3.21), and by defining $\widehat{\mathbf{W}} = [W_1 \ W_2]^T = \mathbf{P}^{-1}\widehat{\boldsymbol{\lambda}}^{nl}$,

Eq. (3.21) becomes:

$$\frac{\partial}{\partial Z} (\widehat{\mathbf{W}}) = \mathbf{\Lambda} \widehat{\mathbf{W}}(Z, t), \quad (3.24)$$

with $\widehat{\mathbf{W}}(0, t) = \mathbf{P}^{-1} \widehat{\boldsymbol{\lambda}}^l((j + 1/2)h, t)$. Owing to the fact that (3.24) is now a decoupled system of differential equations, its solution is obtained directly as follows:

$$\widehat{\mathbf{W}}((j + 1)h, t) = \exp(h\mathbf{\Lambda}) \widehat{\mathbf{W}}(jh, t), \quad j = 0, 1, 2, \dots, M - 1. \quad (3.25)$$

The nonlinear step solution vector is obtained as:

$$\widehat{\boldsymbol{\lambda}}^{nl}\left(\left(j + \frac{1}{2}\right)h, t\right) = [\lambda_{re}^{nl} \quad \lambda_{im}^{nl}]^T = \mathbf{P} \widehat{\mathbf{W}}((j + 1)h, t). \quad (3.26)$$

Once the nonlinear step is executed, the second linear sub-step is applied, on $\lambda^{nl}\left(\left(j + \frac{1}{2}\right)h, t\right) = \lambda_{re}^{nl} + \lambda_{im}^{nl}$, to obtain the adjoint field vector at the next grid point, as follows:

$$\begin{aligned} \lambda((j + 1)h, t) = \mathcal{F}^{-1} \left\{ \exp \left[\frac{h}{2} \left(i \frac{\beta_3}{6} \omega^3 - i \frac{\beta_2}{2} \omega^2 + i \beta_1 \omega - \frac{\alpha}{2} \right) \right] \times \right. \\ \left. \mathcal{F} \left\{ \lambda^{nl}\left(\left(j + \frac{1}{2}\right)h, t\right) \right\} \right\}, \quad j = 0, 1, 2, \dots, M - 1. \end{aligned} \quad (3.27)$$

Finally, the vector of the next grid point solution is given as:

$$\widehat{\boldsymbol{\lambda}}((j + 1)h, t) = [Re \{ \lambda \} \quad Im \{ \lambda \}]^T. \quad (3.28)$$

The algorithm of the adjoint problem solution is summarized by Eqs. (3.20), (3.25),

and (3.27).

The robustness and efficiency of the proposed ASA algorithm is investigated in next through three numerical examples.

3.4 Results

In all of the following examples, the step-size used for solving the original and adjoint simulation is chosen adaptively such that the maximum nonlinear phase per step does not exceed 0.01 rad. All simulations are conducted on an Intel Xeon CPU $\times 5670@2.93$ GHz. The central finite differences (CFD) estimates are obtained with perturbations of 2% of the nominal parameter values.

3.4.1 Example 1

We consider first an optical fiber communication system operating at 10 Gbps. The transmission fiber is a standard single-mode fiber (SSMF) whose parameters are as given in Table 3.1. The transmitted data have a Gaussian pulse shape and are modulated with a return to zero on-off keying (OOK) format with 50% duty cycle. The peak power of the transmitted pulse is $P_0 = 2$ dBm, and the number of transmitted bits is 32.

We aim at obtaining the sensitivities of an objective function of the form:

$$F = \int_{-T_m}^{T_m} |u_{rx} - u_{tx}|^2 dt, \quad (3.29)$$

where $u_{rx} = u(L, t)$ is the normalized complex envelope field of the received signal, L is the optical fiber length, $u_{tx} = u_{tx}(0, t)$ is the normalized transmitted complex

Table 3.1: Simulation parameters of Example 1

Parameter	Value
Inverse group speed, β_1	0.016 ps/km
Second-order dispersion coefficient, β_2	-21 ps ² /km
Third-order dispersion coefficient, β_3	0.1 ps ³ /km
Loss coefficient, α	0 dB/km
Nonlinear coefficient, γ	1.1 W ⁻¹ km ⁻¹

envelope field, and $T_m = 1.6$ ns. This objective function is a measure of the signal distortion due to fiber dispersion and nonlinear effects.

We utilize the proposed ASA algorithm to estimate the sensitivities of (3.29) with respect to the design parameters $\mathbf{x} = [\beta_1 \ \beta_2 \ \beta_3 \ \gamma \ L \ P_0]^T$.

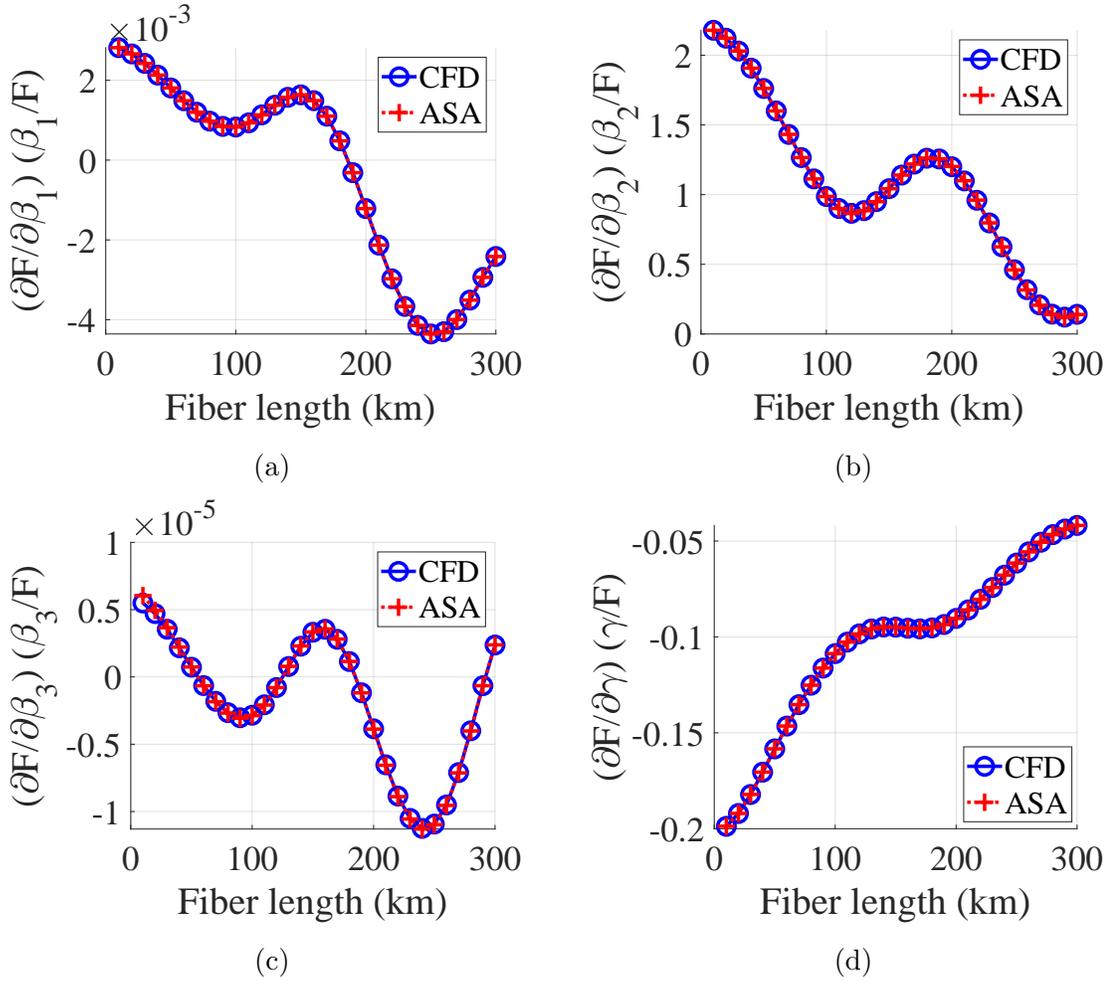


Figure 3.1: The normalized ASA sensitivities of objective function (3.29) with respect to the parameters: β_1 , β_2 , β_3 and γ , for a sweep of L , as compared to the more computationally expensive CFD approach. The nominal fiber parameters are as given in Table 3.1. The modulation format is OOK with Gaussian pulse shaping, the baud rate is 10 Gbps, and the input pulse peak power is $P_0 = 2$ dBm.

Figs. 3.1 and 3.2 show the obtained normalized sensitivities of (3.29) for a sweep of the optical fiber length L . It can be noticed that the error objective function (3.29) is more sensitive to the parameters β_2 , γ and P_0 than the parameter β_3 . This agrees with the physical fact that the impairments in an optical fiber communication system,

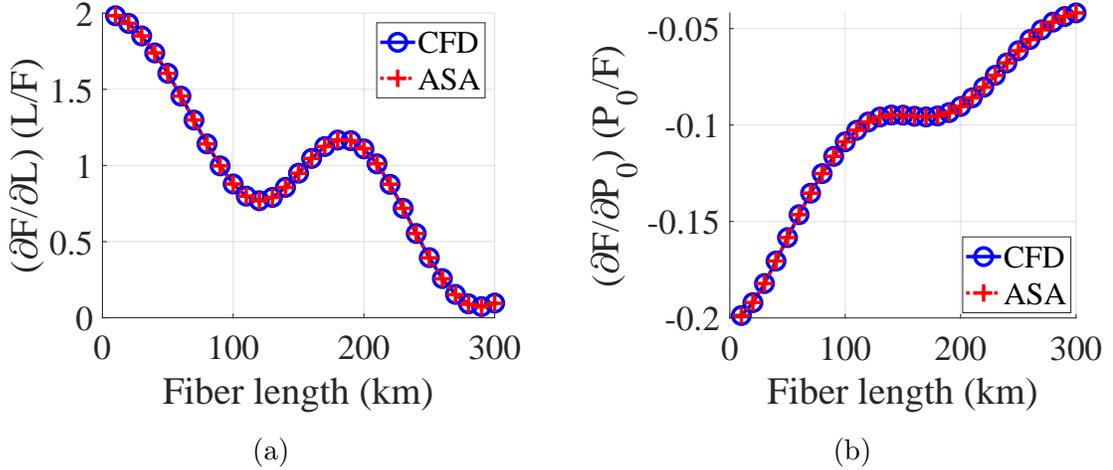


Figure 3.2: The normalized ASA sensitivities of objective function (3.29) with respect to the parameters: L and P_0 , for a sweep of L , as compared to the more computationally expensive CFD approach. The nominal fiber parameters are as given in Table 3.1. The modulation format is OOK with Gaussian pulse shaping, the baud rate is 10 Gbps, and the input pulse peak power is $P_0 = 2$ dBm.

at a data rate of 10 Gbps, is dominated by β_2 and γ rather than β_3 (Agrawal, 2007; Kumar and Deen, 2014). The sensitivities estimated using our ASA method are also compared, in Figs. 3.1 and 3.2, to the sensitivities obtained using the CFD method. As can be seen, a good match is achieved with CFD for all parameters. The ASA algorithm requires only one extra system simulation per gradient evaluation, while the CFD requires 12 extra simulations (2 simulations per design parameter). In other words, a total of 11 extra simulations per gradient calculation have been avoided, in this example, using the proposed ASA algorithm.

We define a time saving factor S_T as the ratio $T_{CFD} = T_{ASA}$, where T_{CFD} and T_{ASA} are the computational times of the CFD approach and our ASA approach, respectively. The value of S_T is calculated to be 4.4 for this example, as shown in Table 3.2. Note that the proportion of time saving achieved is less than the proportion

Table 3.2: The computational time for obtaining the sensitivity results using the CFD and the ASA approaches.

Example	T_{CFD} (sec)	T_{ASA} (sec)	Time saving factor $S_T = T_{CFD}/T_{ASA}$
1	10.74	2.45	4.38
2	178.7	32.8	5.45
3	0.83	0.27	3.05

of the saved simulations. This is due to the fact that the adjoint simulation performs extra mathematical operations to evaluate the adjoint nonlinear matrix \mathbf{F}^λ and the residue vector \mathbf{R}_k (see subsection 3.2.1), which is not involved in the original system simulation.

3.4.2 Example 2

We further investigate the performance of the proposed ASA algorithm in case of optical fiber systems operating at higher baud rate with higher level modulation schemes. We consider a single span SSMF link with the same nominal parameters given in Table 3.3. The input signal is 28 Gbaud, modulated with 16 quadrature amplitude modulation (QAM) format using a root raised cosine pulse shaping with a roll-off factor of 0.1. The average launch power is $P_{av} = 6$ dBm, and the number of transmitted symbols is 4096. We consider a relatively high average launch power in order to have a noticeable nonlinear effect. A noise-free amplifier is inserted at the end of the fiber span to compensate for the fiber loss.

Our target is to evaluate the sensitivities of the same error objective function defined in (3.29) with respect to all 7 design parameters: $\mathbf{x} = [\beta_1 \ \beta_2 \ \beta_3 \ \alpha \ \gamma \ L \ P_{av}]^T$.

Table 3.3: Simulation parameters of Example 2.

Parameter	Value
Inverse group speed, β_1	0.016 ps/km
Second-order dispersion coefficient, β_2	-21 ps ² /km
Third-order dispersion coefficient, β_3	0.1 ps ³ /km
Loss coefficient, α	0.2 dB/km
Nonlinear coefficient, γ	1.1 W ⁻¹ km ⁻¹

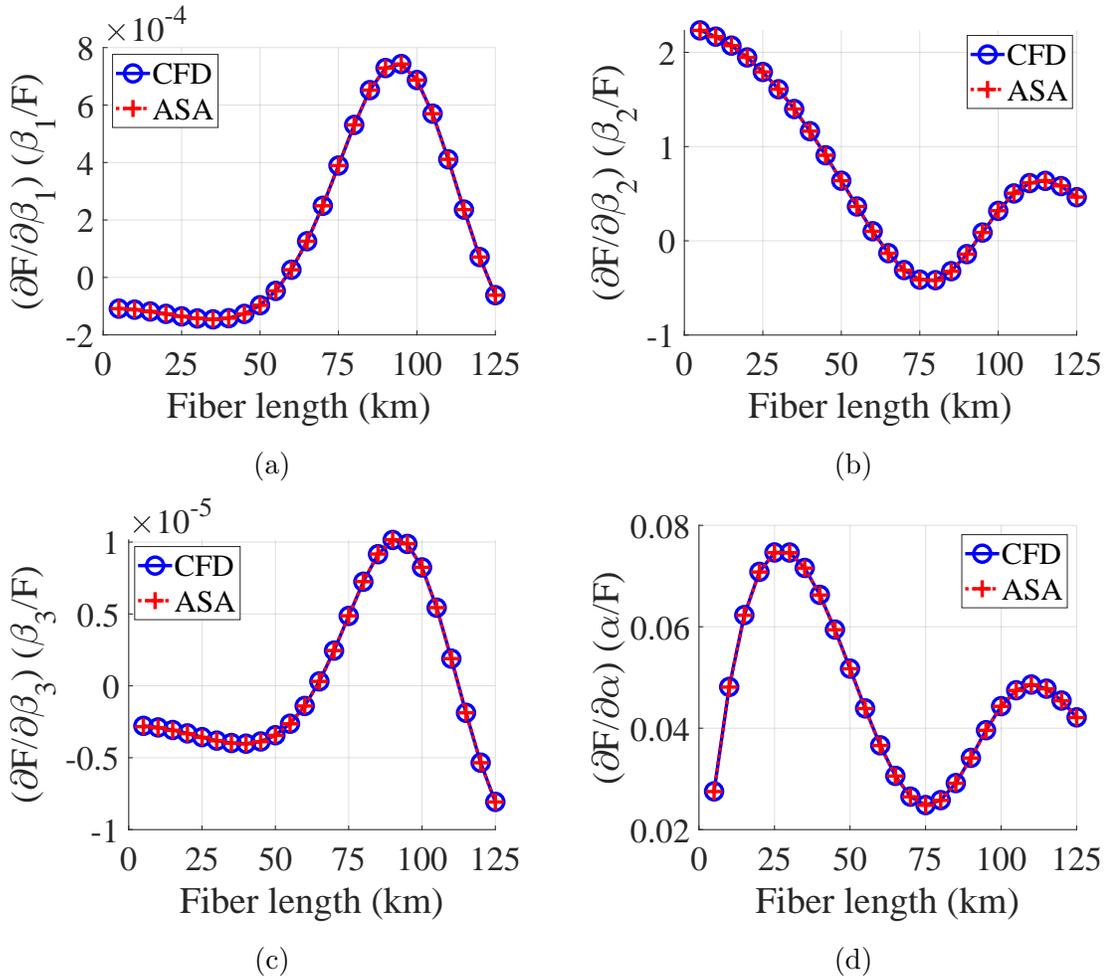


Figure 3.3: The normalized ASA sensitivities of objective function (3.29) with respect to the parameters: β_1 , β_2 , β_3 and α , for a sweep of L , as compared to the more computationally expensive CFD approach. The fiber parameters are as given in Table 3.3. The modulation format is 16 QAM with a root raised cosine pulse shaping, the baud rate is 28 Gbaud, and the average launch power is $P_{av} = 6$ dBm.

The sensitivities estimated using the ASA algorithm, for a sweep of the fiber length parameter L , are shown in Figs. 3.3 and 3.4. The results of the sensitivities estimated using the ASA approach are compared to the CFD results. Good agreement is observed between the ASA sensitivities and the corresponding CFD ones.

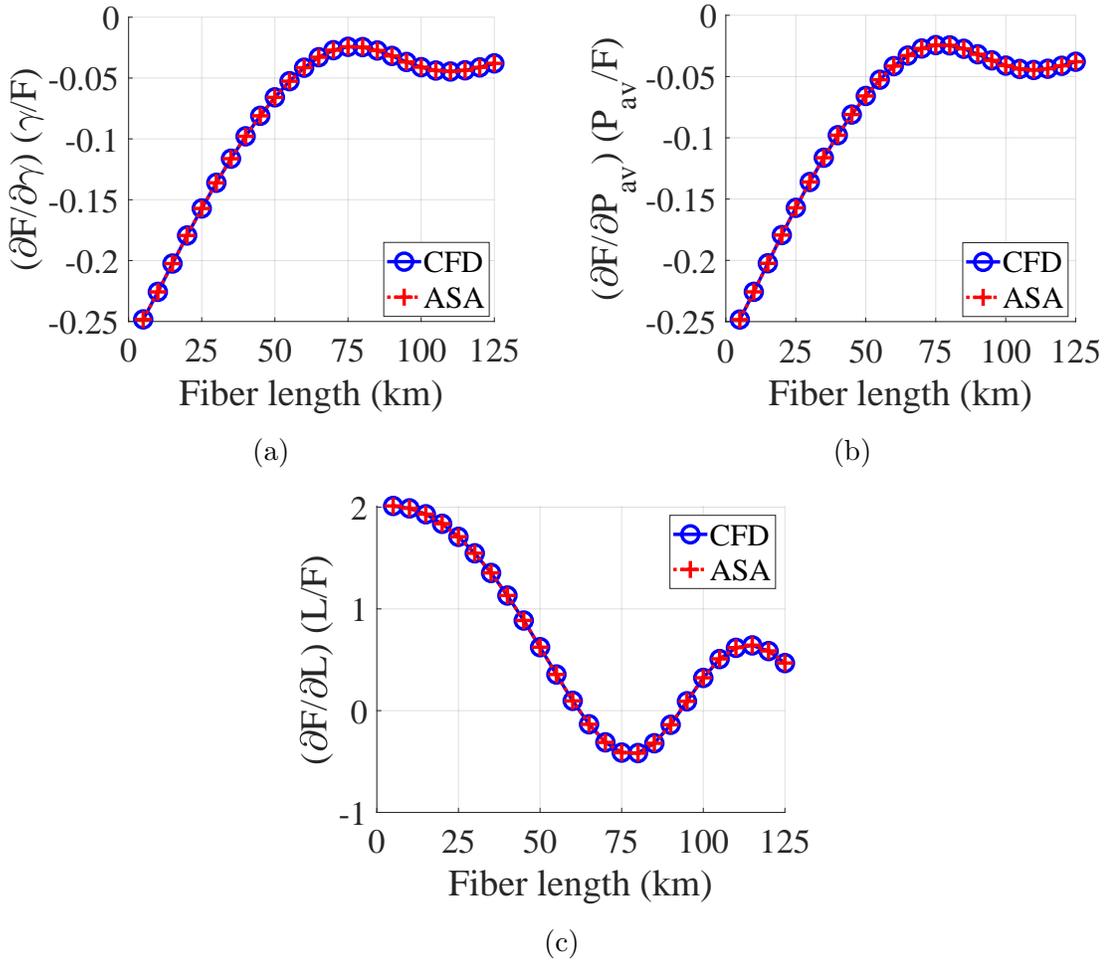


Figure 3.4: The normalized ASA sensitivities of objective function (3.29) with respect to the parameters: γ , P_{av} and L , for a sweep of L , as compared to the more computationally expensive CFD approach. The fiber parameters are as given in Table 3.3. The modulation format is 16 QAM with a root raised cosine pulse shaping, the baud rate is 28 Gbaud, and the average launch power is $P_{av} = 6$ dBm.

The ASA requires only one extra simulation while the CFD approach requires 14

extra simulations. In this example, the ASA algorithm is 5.4 times faster than the CFD approach in estimating the sensitivities (see Table 3.2). It is thus clear that our ASA approach becomes more efficient with the increase of the number of design parameters N . In other words, the saving computational time, achieved using the proposed ASA algorithm, gets increase as the number of design parameters increases.

3.4.3 Example 3

In this example, we exploit the proposed ASA approach to verify the optical soliton phenomenon in an optical fiber communication system. Optical soliton is a solitary light wave that maintains its shape while travelling through an anomalous optical fiber, due to the cancellation of nonlinear and dispersive effects of the fiber (Hasegawa and Tappert, 1973; Hasegawa and Kodama, 1981). It has been shown that the normalized fundamental soliton pulse propagating through an anomalous fiber has a hyperbolic secant shape of the form (Hasegawa and Tappert, 1973; Hasegawa and Kodama, 1981):

$$u(z, t) = \operatorname{sech}\left(\sqrt{\frac{-\gamma P_0}{\beta_2}} (t - \beta_1 z)\right) \exp\left(\frac{i\gamma P_0 z}{2}\right). \quad (3.30)$$

The above equation represents the normalized complex optical field envelope of a fundamental soliton that propagates without any change in its pulse shape. It only acquires a phase shift due to the propagation that is proportional to the pulse peak power P_0 and the nonlinear coefficient of the fiber γ . Assume that we launch the hyperbolic pulse (3.30) with full width at half maximum of 50 ps into a lossless optical fiber. The fiber has the parameters tabulated in Table 3.4. The analytic peak power required to form a soliton in such system settings is $P_{sol} = 23.7$ mW (Kumar

Table 3.4: Fiber parameters used in Examples 3

Parameter	Value
Inverse group speed, β_1	0 ps/km
Second-order dispersion coefficient, β_2	-21 ps ² /km
Third-order dispersion coefficient, β_3	0 ps ³ /km
Loss coefficient, α	0 dB/km
Nonlinear coefficient, γ	1.1 W ⁻¹ km ⁻¹
Fiber length, L	80 km

and Deen, 2014).

In order to numerically verify this peak power, we apply our proposed ASA algorithm to obtain the sensitivities of the following objective function:

$$F = \int_{-T_m}^{T_m} |u_{rx} - u_{ref}|^2 dt, \quad (3.31)$$

where $u_{ref} = u_{tx} \exp(i\gamma P_0 L/2)$, and $u_{tx} = u(0, t)$ with u as given by (3.30).

Fig. 3.5 depicts the normalized sensitivities of (3.31) estimated with respect to $\mathbf{x} = [\beta_2 \ \gamma \ L \ P_0]^T$ for a sweep of P_0 . As can be seen, a local minimum of the objective function (zero gradient value) is detected at $P_0 \approx P_{sol}$. Notice that when the dispersion balances the nonlinear effect at $P_0 = P_{sol}$ the objective function F and its derivative $\partial F/\partial P_0$ become zeros. For P_0 less than P_{sol} , the dispersion effect dominates over nonlinearity leading to pulse broadening as shown in Fig. 3.6a. The derivative $\partial F/\partial P_0$ becomes negative, indicating a decreasing objective function. On the other hand, for $P_0 > P_{sol}$, the nonlinear effect dominates dispersion, leading to pulse compression (see Fig. 3.6b). The derivative $\partial F/\partial P_0$ becomes positive indicating a rising objective function. It should also be clear that the peak values of sensitivities appearing around $P_0 = P_{sol}$ (see Fig. 3.5) are because of the normalization.

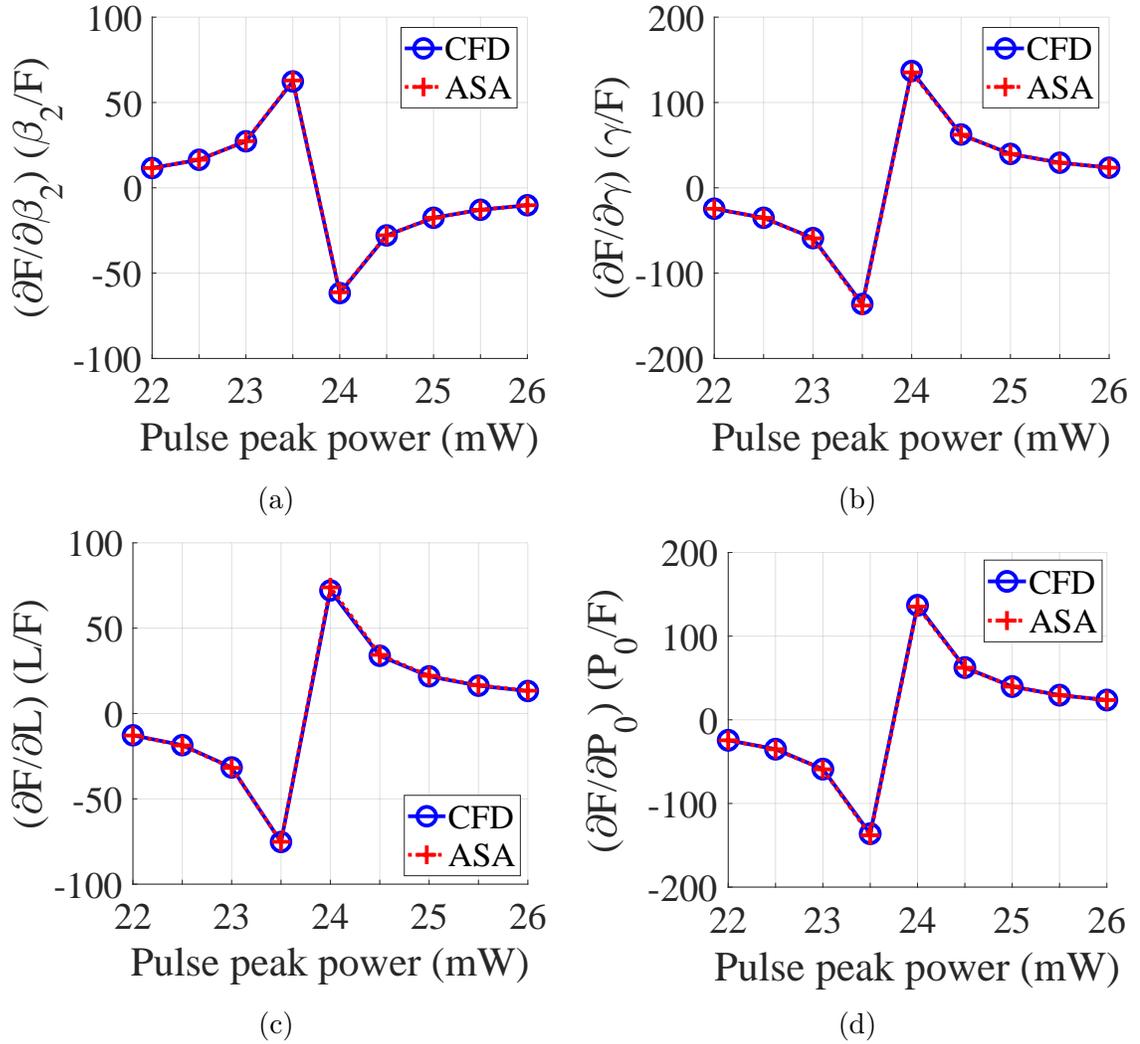


Figure 3.5: The normalized ASA sensitivities of objective function (3.31) with respect to the parameters: β_2 , γ , L and P_0 , for a sweep of P_0 , as compared to the more computationally expensive CFD approach. The fiber parameters are as given in Table 3.4.

Since the absolute value of the objective function F is very small near $P_0 = P_{sol}$, dividing by the small number of F causes peaks in the normalized sensitivities values, as shown in Fig. 3.5.

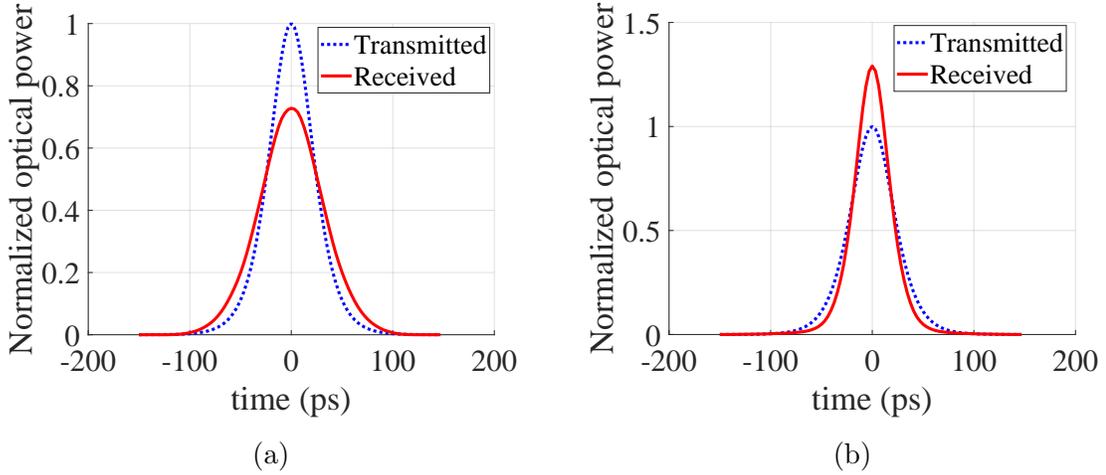


Figure 3.6: The normalized optical power of the transmitted and received pulses. The peak power of the launch pulse is: (a) 15 mW, i.e., less than the required peak power to form a soliton P_{sol} , and (b) 30 mW, i.e., greater than P_{sol} .

The estimated ASA sensitivities are also compared to the sensitivities obtained using the accurate CFD approach. As shown in Fig. 3.5, the sensitivities obtained using the ASA approach are well matched to the CFD estimates. All the ASA sensitivities shown in Fig. 3.5 require only one extra simulation as opposed to 8 extra simulations required by the CFD approach, i.e., the ASA algorithm prevents extra 7 simulations per gradient calculation. In this example, the ASA is 3.1 times faster than the CFD (see Table 3.2).

3.5 Conclusion

In this Chapter, we proposed a computationally efficient nonlinear adjoint sensitivity analysis approach for the general time-dependent nonlinear Schrödinger equation. In contrast to finite-difference (FD) approaches that work on the system response level, the basic idea of the proposed ASA approach is to store the original system state data

and utilize these information in the adjoint simulation to estimate the sensitivities with minimum number of system simulations. In other words, as compared to the FD approach, the ASA approach significantly reduces the computational time cost of the sensitivity analysis at the expense of extra memory storage usage. As compared to the computationally expensive central finite-difference approach, the proposed algorithm estimates accurate sensitivities of the desired objective function with respect to all the design parameters but with a much lower computational cost. The efficiency of the approach is illustrated through numerical examples of fiber-optic communication systems. It has been shown that the gain of using the proposed ASA algorithms becomes more significant as the number of design parameters increases.

Chapter 4

Adaptive Digital Back Propagation Exploiting Adjoint-Based Optimization

As an application to the nonlinear Schrödinger equation (NLSE)-based ASA approach introduced in Chapter 3, we might utilize our ASA approach to accelerate the parameters extraction of fiber-optic communication systems. In this Chapter, we apply our NLSE-based ASA approach for the sensitivity analysis of the digital back propagation (DBP) system. The ASA algorithm is extended to estimate the sensitivities for the multi-span DBP model, i.e., virtual backward fiber and inline loss element. The DBP-based ASA algorithm in addition to a gradient-based optimization approach are then exploited to develop a powerful and fast adjoint-based optimization (ABO) algorithm. The ABO algorithm is used to train and optimize the design parameters of the DBP in order to blindly compensate for the linear and nonlinear distortions of the optical fiber transmission channel. The proposed ABO algorithm is shown to effectively

optimize the compensation performance of the DBP in the case of a point-to-point fiber-optic transmission system or an optical fiber transmission network.

The chapter is organized as follows: In Section 4.1, we review the basic concept of digital back propagation and derive a general mathematical matrix representation required for the DBP simulation. In Section 4.2, we derive the general adjoint system simulation corresponding to the original system simulation of the DBP. A full description of the modified SSFS algorithm required for solving the derived adjoint simulation is also given. Moreover, Section 4.2 provides a complete computational complexity analysis of the modified SSFS algorithm. Section 4.3 provides a detailed description of the proposed ABO algorithm. In Section 4.4, we investigate the efficiency of the proposed ABO algorithm through applying it to train a DBP for blindly mitigating the distortions in a typical 4-span fiber-optic communication system. We also compare the performance of the ABO algorithm as compared to finite-difference-based optimization algorithms. The obtained results show that our ABO algorithm has the same accuracy as the central finite-difference-based optimizer with much lower computational complexity. Finally, conclusions of the work are drawn in Section 4.5.

4.1 Problem Formulation of a DBP Model

4.1.1 Single-Span DBP Model

Consider first a single-span optical fiber link with certain fiber dispersion, nonlinear and loss coefficients. The evolution of the normalized optical field envelope in the

fiber is described by the NLSE (Agrawal, 2007; Kumar and Deen, 2014):

$$\frac{\partial u}{\partial z} = \left[\hat{D} + \hat{N} \right] u, \quad (4.1)$$

where $u(z, t) = q/\sqrt{P_0}$ is the normalized signal of the complex optical field envelope $q(z, t)$, and P_0 is the peak power of the transmitted pulse. The operators \hat{D} and \hat{N} represent the linear (delay, dispersion, and loss) and the nonlinear effects of the optical fiber, respectively. They are given by:

$$\hat{D}(t) = \frac{\beta_3}{6} \frac{\partial^3}{\partial t^3} - \frac{i\beta_2}{2} \frac{\partial^2}{\partial t^2} - \beta_1 \frac{\partial}{\partial t} - \frac{\alpha}{2}, \quad (4.2)$$

$$\hat{N}(z, t) = i\gamma P_0 |u(z, t)|^2, \quad (4.3)$$

where β_1 is the inverse group speed, β_2 and β_3 are the second- and third-order dispersion coefficients, respectively. The parameter α is the fiber loss coefficient, γ is the nonlinear coefficient of the fiber, and $i = \sqrt{-1}$.

The analytical solution of (4.1) is obtained as follows:

$$u(L, t) = \hat{M}u(0, t), \quad (4.4)$$

where L is the fiber length and $u(0, t) = u_{tx}(t) = q_{tx}/\sqrt{P_0}$ is the normalized complex envelope of the transmitted field $q_{tx}(t)$. The fiber operator \hat{M} is given as:

$$\hat{M} = \exp \left(\int_0^L \left[\hat{D}(t) + \hat{N}(z, t) \right] dz \right). \quad (4.5)$$

Multiplying both sides of (4.4) by the inverse operator of \hat{M} yields (Agrawal, 2007;

Kumar and Deen, 2014):

$$u(0, t) = \hat{M}^{-1}u(L, t), \quad (4.6)$$

where

$$\hat{M}^{-1} = \exp\left(-\int_0^L [\hat{D}(t) + \hat{N}(z, t)] dz\right). \quad (4.7)$$

In (4.6), the signal $u(L, t)$ represents the received field which is distorted due to fiber linear and nonlinear effects. In a noise-free system, if we multiply the received field by the inverse fiber operator \hat{M}^{-1} , the linear and nonlinear fiber distortions can thus be completely removed. Eq. (4.6), with \hat{M}^{-1} as given by (4.7), is equivalent to solving the following partial differential equation (Agrawal, 2007; Kumar and Deen, 2014):

$$\frac{\partial v}{\partial Z} = -[\hat{D} + \hat{N}]v, \quad (4.8)$$

with the initial condition $v(Z = 0, t) = u(z = L, t)$. Notice that the equalized signal $v(Z, t)$ is solved in the reversed spatial direction $Z = L - z$. This equalization technique is therefore referred to as back propagation. The equation describing the evolution of the normalized field in a single-span DBP model, denoted as the inverse nonlinear Schrödinger equation (INLSE), is obtained by rewriting (4.8) as follows (Ip and Kahn, 2008; Li *et al.*, 2008):

$$\frac{\partial v}{\partial Z} = [\hat{D}_b + \hat{N}_b]v, \quad (4.9)$$

with $v(Z = 0, t) = u(z = L, t)$ and

$$\hat{D}_b(t) = -\frac{\beta_3}{6}\frac{\partial^3}{\partial t^3} + \frac{i\beta_2}{2}\frac{\partial^2}{\partial t^2} + \beta_1\frac{\partial}{\partial t} + \frac{\alpha}{2}, \quad (4.10)$$

$$\hat{N}_b(Z, t) = -i\gamma P_0 |v(Z, t)|^2. \quad (4.11)$$

It can be noticed that the INLSE (4.9) has the same form as the NLSE (4.1), but with different values of the parameters (opposite signs of β_1 , β_2 , β_3 , α and γ), and with different initial condition. The initial field of the INLSE problem is the received field $u(L, t)$ in electrical domain while the initial field in the NLSE problem is the transmitted optical field $u(0, t)$. The INLSE problem can therefore be assumed as a simulation of a virtual fiber. The INLSE is also solved in a spatial direction Z reversed to the direction z at which the forward NLSE is solved. It thus follows that the problem of the INLSE can be solved using the same numerical solver, i.e., the split-step Fourier scheme (SSFS) method, used for solving the NLSE problem. Fig. 4.1 illustrates the forward and backward propagations of a single-span optical transmission fiber and virtual DBP fiber, respectively.

As shown in Fig. 4.1, the digital signal after the DBP can be written as:

$$\begin{aligned} v_{out}(t) &= v(Z = L, t) = e^{i \int_0^L [\hat{D}_b(t) + \hat{N}_b(Z, t)] dZ} u(z = L, t) \\ &= e^{i \int_0^L [\hat{D}_b(t) + \hat{N}_b(Z, t)] dZ} \times e^{i \int_0^L [\hat{D}(t) + \hat{N}(z, t)] dz} u(z = 0, t) \\ &= e^{-i \int_0^L [\hat{D}(t) + \hat{N}(z, t)] dz} \times e^{i \int_0^L [\hat{D}(t) + \hat{N}(z, t)] dz} u(z = 0, t) \\ &= u(z = 0, t) = u_{tx}(t). \end{aligned} \quad (4.12)$$

Thus, the transmitted signal is fully recovered after the DBP if there is no noise.

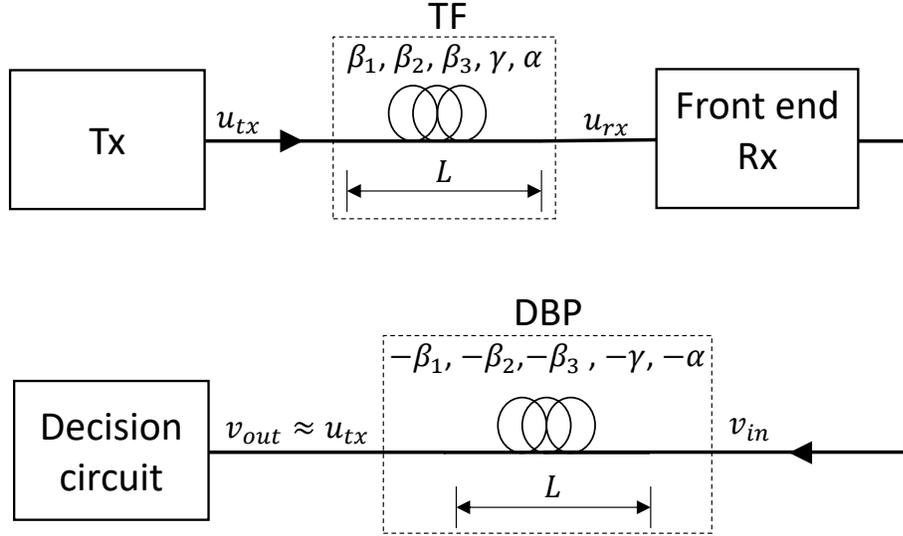


Figure 4.1: The propagation in a single-span optical transmission fiber (forward propagation), and a single-span virtual DBP fiber (backward propagation).

Tx: transmitter, Rx: receiver, TF: transmission fiber, DBP: digital back propagation.

4.1.2 Multi-Span DBP Model

For multi-span fiber-optic systems, Fig. 4.2 shows the typical propagation model in an M -span optical fiber system. In order to compensate for the fiber losses, an erbium-doped fiber amplifier (EDFA) is inserted at the end of each fiber span, with gain $\exp(g_j)$, $j = 1, 2, \dots, M$. The inserted EDFA not only compensates completely for the fiber span loss, but it also adds a white Gaussian noise $n_j(t)$, $j = 1, 2, \dots, M$, to the amplified signal. The main source of this noise is known as the amplified spontaneous emission (ASE) noise, having a noise figure N_f . Note that the amplifier gain is provided by the stimulated emission process, while the noise is due to the spontaneous emission process (Mears *et al.*, 1987).

To undo the propagation distortions in the multi-span system, we generalize the concept of the single-span DBP. The noisy amplifier with gain $\exp(g)$ is substituted

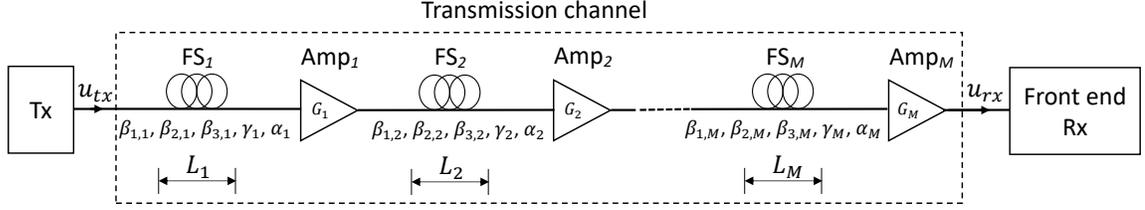


Figure 4.2: Model of propagation in an M -span optical fiber transmission system. Tx: transmitter, Rx: receiver, FS: fiber span, Amp: amplifier.

by a noise-free loss element with gain $\exp(-g)$ in the digital domain, and the real fiber with parameters $(\beta_1, \beta_2, \beta_3, \alpha, \gamma, L)$ is replaced by a virtual fiber having parameters $(-\beta_1, -\beta_2, -\beta_3, -\alpha, -\gamma, L)$, as shown in Fig. 4.3. Notice that the signal distortions due to the last fiber span in the fiber-optic link is compensated first in the digital domain. Whereas, the distortions due to the first fiber span is compensated using the last virtual span of the DBP. Although the DBP can compensate for the deterministic linear and nonlinear fiber distortions, it cannot undo the impact of the ASE noise and the nonlinearity-ASE coupling such as Gordon-Mollenauer phase noise (Gordon and Mollenauer, 1990). In other words, the transmitted electric field can be fully recovered after the DBP if the transmission system is free of noise.

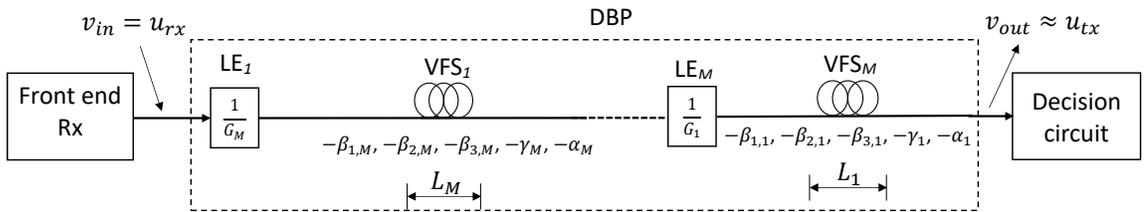


Figure 4.3: Block diagram of a digital back propagation model for a M -span fiber-optic communication system. Rx: receiver, VFS: virtual fiber span, LE: loss element, DBP: digital back propagation.

In order to mathematically represent a complete span of the DBP model shown

in Fig.4.3 (i.e. virtual backward fiber + inline attenuator), we modify Eq. (4.9) as follows:

$$\frac{\partial v}{\partial Z} - \left[\hat{D}_b + \hat{N}_b \right] v + \hat{G}v = v_{in} \delta(Z), \quad (4.13)$$

where the linear and nonlinear operators (\hat{D}_b and \hat{N}_b) are given in Eqs. (4.10) and (4.11), respectively. The attenuator operator \hat{G} is defined as:

$$\hat{G} = \begin{cases} 0, & \forall Z \neq jL_j \\ g_j, & \text{at } Z = jL_j \end{cases}, \quad (4.14)$$

where L_j and $G_j = \exp(g_j)$ are the fiber length and the amplifier gain, respectively, of the j^{th} span, and $j = 1, 2, \dots, M$. Note that in (4.13), we assume zero boundary condition, i.e., $v(0, t) = 0$, and the field excitation is given by the right-hand side term $v_{in} \delta(Z)$, where $v_{in} = u(z = L, t)$. Expressing the complex field $v(Z, t)$ in terms of its real and imaginary parts ($v = v_{re} + iv_{im}$), substituting in Eq. (4.13), and separating real and imaginary terms, we obtain:

$$\begin{aligned}
 & \begin{bmatrix} \tilde{\beta}_3 & 0 \\ 0 & \tilde{\beta}_3 \end{bmatrix} \begin{bmatrix} \frac{\partial^3 v_{re}}{\partial t^3} \\ \frac{\partial^3 v_{im}}{\partial t^3} \end{bmatrix} + \begin{bmatrix} 0 & \tilde{\beta}_2 \\ -\tilde{\beta}_2 & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial^2 v_{re}}{\partial t^2} \\ \frac{\partial^2 v_{im}}{\partial t^2} \end{bmatrix} + \begin{bmatrix} -\tilde{\beta}_1 & 0 \\ 0 & -\tilde{\beta}_1 \end{bmatrix} \begin{bmatrix} \frac{\partial v_{re}}{\partial t} \\ \frac{\partial v_{im}}{\partial t} \end{bmatrix} \\
 & + \begin{bmatrix} \partial/\partial Z & 0 \\ 0 & \partial/\partial Z \end{bmatrix} \begin{bmatrix} v_{re} \\ v_{im} \end{bmatrix} + \begin{bmatrix} \tilde{\alpha} & 0 \\ 0 & \tilde{\alpha} \end{bmatrix} \begin{bmatrix} v_{re} \\ v_{im} \end{bmatrix} + \begin{bmatrix} 0 & -\tilde{\gamma} \\ \tilde{\gamma} & 0 \end{bmatrix} \begin{bmatrix} v_{re} \\ v_{im} \end{bmatrix} \\
 & + \begin{bmatrix} \tilde{g} & 0 \\ 0 & \tilde{g} \end{bmatrix} \begin{bmatrix} v_{re} \\ v_{im} \end{bmatrix} = \begin{bmatrix} Re\{v_{in}\} \delta(Z) \\ Im\{v_{in}\} \delta(Z) \end{bmatrix}, \tag{4.15}
 \end{aligned}$$

where $\tilde{\beta}_1 = -\beta_{1,j}$, $\tilde{\beta}_2 = \beta_{2,j}/2$, $\tilde{\beta}_3 = \beta_{3,j}/6$, $\tilde{\alpha} = -\alpha_j/2$, and $\tilde{\gamma}(v) = \gamma_j P_0(v_{re}^2 + v_{im}^2)$. The parameter $\tilde{g} = g_j \delta\left(Z - \sum_{k=j+1}^M L_k\right)$, where j can be 1, 2, ..., or M . Assume that the j^{th} span is discretized into p_j spatial cells with a step size of h . Writing Eq. (4.15) for all cells of the span, and approximating the spatial derivatives using central-finite differences, yields:

where $\tilde{h} = 1/(2h)$, and the vectors \mathbf{v}_{re} and \mathbf{v}_{im} are the real- and imaginary-parts of all components of the complex field v at all span cells. It should be clear that the right-hand term (excitation vector) has a non-zero first component only at the 1st DBP span. Eq. (4.16) can be written for the whole computational domain as follows:

$$\mathbf{B}_3 \ddot{\mathbf{V}} + \mathbf{B}_2 \dot{\mathbf{V}} + \mathbf{B}_1 \mathbf{V} + \mathbf{K}_c \mathbf{V} + \mathbf{A} \mathbf{V} + \mathbf{\Gamma}(\mathbf{V}) \mathbf{V} + \mathbf{G} \mathbf{V} = \mathbf{V}_{in}(t), \quad (4.17)$$

where $\mathbf{V} = [\mathbf{v}_{re}^T \ \mathbf{v}_{im}^T]^T$ is the system state vector, containing all real and imaginary parts of the complex field v in the entire computational domain, and $\dot{\cdot}$ denotes time derivative. The vector $\mathbf{V}_{in}(t) = [Re\{v_{in}\} \mathbf{e}_1^T \ Im\{v_{in}\} \mathbf{e}_1^T]^T$ is the excitation vector, where $\mathbf{e}_1 = [1 \ 0 \ \dots \ 0]^T$ is the 1st elementary column vector of size $2p$, $p = \sum_{j=1}^M p_j$ is the total number of spatial cells in the computational domain, and p_j , $j = 1, 2, \dots, M$, is the number of cells in the j^{th} span. The constant matrix \mathbf{K}_c approximates the spatial derivatives. The system matrices are given as: $\mathbf{B}_1 = \begin{bmatrix} \tilde{\beta}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\beta}_1 \end{bmatrix}$, $\mathbf{B}_2 = \begin{bmatrix} \mathbf{0} & \tilde{\beta}_2 \\ -\tilde{\beta}_2 & \mathbf{0} \end{bmatrix}$, $\mathbf{B}_3 = \begin{bmatrix} \tilde{\beta}_3 & \mathbf{0} \\ \mathbf{0} & \tilde{\beta}_3 \end{bmatrix}$, and $\mathbf{A} = \begin{bmatrix} \tilde{\alpha} & \mathbf{0} \\ \mathbf{0} & \tilde{\alpha} \end{bmatrix}$, where $\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\beta}_3$, and $\tilde{\alpha}$ are all diagonal matrices whose l^{th} diagonal entry contains the corresponding values of $\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\beta}_3$, and $\tilde{\alpha}$, respectively, at the l^{th} cell of the computational domain, and $l = 1, 2, \dots, p$. The matrix $\mathbf{G} = \begin{bmatrix} \tilde{\mathbf{g}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{g}} \end{bmatrix}$ is the loss elements matrix, where $\tilde{\mathbf{g}}$ is a diagonal matrix given as follows:

$$\tilde{\mathbf{g}} = \sum_{j=1}^M g_j \mathbf{e}_{l_j} \mathbf{e}_{l_j}^T, \quad (4.18)$$

where \mathbf{e}_{l_j} is an elementary column vector of size $2p$ whose nonzero entry is $l_j =$

$1 + \sum_{k=1}^j p_{k-1}$ with the assumption that $p_0 = 0$. The matrix $\mathbf{F}(\mathbf{V}) = \begin{bmatrix} \mathbf{0} & \tilde{\gamma} \\ -\tilde{\gamma} & \mathbf{0} \end{bmatrix}$ is the matrix corresponding to the fiber nonlinearity, where $\tilde{\gamma}(\mathbf{V})$ is a diagonal matrix whose l^{th} entry is given by: $\gamma_l P_0 [v_{re}^2(lh) + v_{im}^2(lh)]$, $l = 0, 1, \dots, p-1$. Note that in the definition of system matrices, we assume that the computational domain is discretized with a fixed step-size h for mathematical convenience. However, the multi-span DBP matrix representation given in (4.17) is general and still valid for the varied step-size discretization case.

Eq. (4.17) represents the original simulation of the multi-span DBP system. Next, we derive an adjoint system simulation corresponding to (4.17). Using those 2 simulations (the original and adjoint simulations), the full gradient of a general objective function with respect to all DBP design parameters can be estimated.

4.2 ASA for the Multi-Span DBP Model

Our objective is to estimate the sensitivities of the following objective function (Bakr *et al.*, 2017):

$$F = \int_{-T_m}^{T_m} \psi(\mathbf{x}, \mathbf{V}) dt, \quad (4.19)$$

where $2 \times T_m$ is the total simulation time (i.e. time length of the normalized transmitted signal u_{tx}), $\mathbf{x} \in \mathbb{R}^N$ is the vector of design parameters which includes all the DBP design parameters, and \mathbf{V} is the system state vector. The analytic derivative

of (4.19) with respect to the k^{th} parameter x_k , $k = 1, 2, \dots, N$, is given by:

$$\frac{\partial F}{\partial x_k} = \int_{-T_m}^{T_m} \frac{\partial^e \psi}{\partial x_k} dt + \int_{-T_m}^{T_m} \left(\frac{\partial \psi}{\partial \mathbf{V}} \right)^T \frac{\partial \mathbf{V}}{\partial x_k} dt, \quad (4.20)$$

where $\partial^e / \partial x_k$ denotes the explicit dependence. The analytic expression in (4.20) cannot be evaluated unless the vector $\partial \mathbf{V} / \partial x_k$ is known for every time step. The classical approach for evaluating (4.20) involves repeatedly simulating perturbed structures, for each design parameter x_k , $k = 1, 2, \dots, N$. A total of extra N full DBP simulations are required for forward or backward differences. The more accurate central differences require $2N$ extra DBP simulations. This overhead can be significant even for a small number of design parameters, especially for such time intensive simulation problem. Notice that the computational complexity of a single-span DBP simulation run scales as $\sim M_p N_s \log_2(N_s)$, where N_s is the total number of time samples and M_p is the number of propagation steps per span (Ip and Kahn, 2008; Li *et al.*, 2008).

The alternative approach is to derive an adjoint sensitivity analysis (ASA) algorithm to estimate the implicit derivative of (4.20) with respect to all design parameters x_k , $k = 1, 2, \dots, N$, using only one extra adjoint simulation. Similar to the derivation steps shown in Chapter 3, we start by differentiating the original multi-span DBP system (4.17) with respect to the k^{th} parameter x_k , $k = 1, 2, \dots, N$, as follows:

$$\begin{aligned} \frac{\partial^e \mathbf{B}_3}{\partial x_k} \frac{\partial^3 \mathbf{V}}{\partial t^3} + \mathbf{B}_3 \frac{\partial^4 \mathbf{V}}{\partial x_k \partial t^3} + \frac{\partial^e \mathbf{B}_2}{\partial x_k} \frac{\partial^2 \mathbf{V}}{\partial t^2} + \mathbf{B}_2 \frac{\partial^3 \mathbf{V}}{\partial x_k \partial t^2} + \frac{\partial^e \mathbf{B}_1}{\partial x_k} \frac{\partial \mathbf{V}}{\partial t} + \mathbf{B}_1 \frac{\partial^2 \mathbf{V}}{\partial x_k \partial t} + \mathbf{K}_c \frac{\partial \mathbf{V}}{\partial x_k} + \\ \frac{\partial^e \mathbf{A}}{\partial x_k} \mathbf{V} + \mathbf{A} \frac{\partial \mathbf{V}}{\partial x_k} + \frac{\partial^e \mathbf{\Gamma}}{\partial x_k} \mathbf{V} + \frac{\partial (\mathbf{\Gamma} \bar{\mathbf{V}})}{\partial \mathbf{V}^T} \frac{\partial \mathbf{V}}{\partial x_k} + \mathbf{\Gamma} \frac{\partial \mathbf{V}}{\partial x_k} + \frac{\partial^e \mathbf{G}}{\partial x_k} \mathbf{V} + \mathbf{A} \frac{\partial \mathbf{G}}{\partial x_k} = \mathbf{0}. \end{aligned} \quad (4.21)$$

where $\bar{\mathbf{V}}$ is the nominal value of \mathbf{V} , treated as constant during the differentiation

operation in (4.21). It should be clear that all system matrices arising in (4.17) are constants with time, except for the nonlinear coefficients matrix $\mathbf{\Gamma} = \mathbf{\Gamma}(\mathbf{V})$. This dependency arises from the fact that the nonlinearity coefficient $\tilde{\gamma} = \gamma p_0 |v|^2$ is function of the local field value. Notice also that the normalized excitation vector \mathbf{V}_{in} is not a function of the DBP design parameters. All explicit derivatives appearing in (4.21) are known and they can be evaluated using the definitions of DBP system matrices: \mathbf{B}_1 , \mathbf{B}_2 , \mathbf{B}_3 , \mathbf{A} , $\mathbf{\Gamma}$, and \mathbf{G} . For example, if the second dispersion coefficient is the current design parameter (i.e. $x_k = \beta_2$), then $\partial^e \mathbf{B}_1 / \partial x_k = \mathbf{0}$, $\partial^e \mathbf{B}_2 / \partial x_k = -0.5 \mathbf{I}_{2m}$, $\partial^e \mathbf{B}_3 / \partial x_k = \mathbf{0}$, $\partial^e \mathbf{A} / \partial x_k = \mathbf{0}$, $\partial^e \mathbf{\Gamma} / \partial x_k = \mathbf{0}$, and $\partial^e \mathbf{G} / \partial x_k = \mathbf{0}$. Reorganizing (4.21) to shift all the known terms to the right-hand side, we get:

$$\begin{aligned} \mathbf{B}_3 \frac{\partial^4 \mathbf{V}}{\partial x_k \partial t^3} + \mathbf{B}_2 \frac{\partial^3 \mathbf{V}}{\partial x_k \partial t^2} + \mathbf{B}_1 \frac{\partial^2 \mathbf{V}}{\partial x_k \partial t} + \mathbf{K}_c \frac{\partial \mathbf{V}}{\partial x_k} + \mathbf{A} \frac{\partial \mathbf{V}}{\partial x_k} + \frac{\partial (\mathbf{\Gamma} \bar{\mathbf{V}})}{\partial \mathbf{V}^T} \frac{\partial \mathbf{V}}{\partial x_k} + \mathbf{\Gamma} \frac{\partial \mathbf{V}}{\partial x_k} + \\ \mathbf{G} \frac{\partial \mathbf{V}}{\partial x_k} = -\mathbf{R}_k, \end{aligned} \quad (4.22)$$

where \mathbf{R}_k is the residue vector corresponding to the k^{th} parameter x_k , given by:

$$\mathbf{R}_k = \frac{\partial^e \mathbf{B}_3}{\partial x_k} \ddot{\mathbf{V}} + \frac{\partial^e \mathbf{B}_2}{\partial x_k} \dot{\mathbf{V}} + \frac{\partial^e \mathbf{B}_1}{\partial x_k} \mathbf{V} + \frac{\partial^e \mathbf{A}}{\partial x_k} \mathbf{V} + \frac{\partial^e \mathbf{\Gamma}}{\partial x_k} \mathbf{V} + \frac{\partial^e \mathbf{G}}{\partial x_k} \mathbf{V}. \quad (4.23)$$

Multiplying both sides of (4.22) by the yet-to-be-determined temporal adjoint vector $\boldsymbol{\lambda}$, integrating over time, and exploiting integration by parts to have no mixed derivative terms in the kernel of integral, we obtain:

$$\int_{-T_m}^{T_m} \left(-\ddot{\boldsymbol{\lambda}}^T \mathbf{B}_3 + \ddot{\boldsymbol{\lambda}}^T \mathbf{B}_2 - \dot{\boldsymbol{\lambda}}^T \mathbf{B}_1 + \boldsymbol{\lambda}^T \left(\mathbf{K}_c + \mathbf{A} + \frac{\partial(\boldsymbol{\Gamma}\bar{\mathbf{V}})}{\partial \mathbf{V}^T} + \boldsymbol{\Gamma} + \mathbf{G} \right) \right) \frac{\partial \mathbf{V}}{\partial x_k} dt = - \int_{-T_m}^{T_m} \boldsymbol{\lambda}^T \mathbf{R}_k dt, \quad (4.24)$$

where $\boldsymbol{\lambda}(Z, t) = [\boldsymbol{\lambda}_{re}^T \ \boldsymbol{\lambda}_{im}^T]^T$ is the adjoint state vector. Note that to obtain (4.24), we assume the following zero boundary conditions: $\mathbf{V}(Z, -T_m) = \dot{\mathbf{V}}(Z, -T_m) = \ddot{\mathbf{V}}(Z, -T_m) = \mathbf{0}$, and $\boldsymbol{\lambda}(Z, T_m) = \dot{\boldsymbol{\lambda}}(Z, T_m) = \ddot{\boldsymbol{\lambda}}(Z, T_m) = \mathbf{0}$. Comparing the second term in (4.24) with the left-hand side (i.e. the implicit integral) of (4.20), we enforce that:

$$-\ddot{\boldsymbol{\lambda}}^T \mathbf{B}_3 + \ddot{\boldsymbol{\lambda}}^T \mathbf{B}_2 - \dot{\boldsymbol{\lambda}}^T \mathbf{B}_1 + \boldsymbol{\lambda}^T \left(\mathbf{K}_c + \mathbf{A} + \frac{\partial(\boldsymbol{\Gamma}\bar{\mathbf{V}})}{\partial \mathbf{V}^T} + \boldsymbol{\Gamma} + \mathbf{G} \right) = \left(\frac{\partial \psi}{\partial \mathbf{V}} \right)^T. \quad (4.25)$$

From the system matrices definitions in (4.16), it is clear that: \mathbf{B}_1 , \mathbf{B}_3 , \mathbf{A} , and \mathbf{G} are symmetric matrices, whereas \mathbf{B}_2 , \mathbf{K}_c , and $\boldsymbol{\Gamma}$ are skew-symmetric matrices. Taking the transpose of both two sides of (4.25), we thus obtain the adjoint system problem, given as:

$$-\mathbf{B}_3 \frac{\partial^3 \boldsymbol{\lambda}}{\partial t^3} - \mathbf{B}_2 \frac{\partial^2 \boldsymbol{\lambda}}{\partial t^2} - \mathbf{B}_1 \frac{\partial \boldsymbol{\lambda}}{\partial t} - \mathbf{K}_c \boldsymbol{\lambda} + \mathbf{A} \boldsymbol{\lambda} + \boldsymbol{\Gamma}^\lambda(\mathbf{V}) \boldsymbol{\lambda} + \mathbf{G} \boldsymbol{\lambda} = \mathbf{Q}_{in}^\lambda(t), \quad (4.26)$$

where \mathbf{Q}_{in}^λ is the adjoint excitation vector given by: $\mathbf{Q}_{in}^\lambda(t) = \partial \psi / \partial \mathbf{V} = [\partial \psi / \partial \mathbf{v}_{re}^T \ \partial \psi / \partial \mathbf{v}_{im}^T]^T$. It can be shown that the matrix $\boldsymbol{\Gamma}^\lambda$ corresponding to the fiber nonlinearity of the adjoint problem is given as:

$$\mathbf{\Gamma}^\lambda(\mathbf{V}) = \frac{\partial(\mathbf{\Gamma}\bar{\mathbf{V}})^T}{\partial\mathbf{V}} - \mathbf{\Gamma} = \begin{bmatrix} -\mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{a} \end{bmatrix}, \quad (4.27)$$

where \mathbf{a} , \mathbf{b} , and \mathbf{c} are diagonal matrices whose l^{th} diagonal elements are given as follows:

$$a_{ll} = 2\gamma P_0 u_{re}(lh) u_{im}(lh), \quad (4.28a)$$

$$b_{ll} = \gamma P_0 \left(3u_{re}^2(lh) + u_{im}^2(lh) \right), \quad (4.28b)$$

$$c_{ll} = -\gamma P_0 \left(u_{re}^2(lh) + 3u_{im}^2(lh) \right), \quad (4.28c)$$

where h is the spatial discretization step size, $l = 0, 1, \dots, p-1$, and p is the total number of the spatial discretization cells in the entire computational domain. Eq. (4.26) represents the adjoint problem corresponding to the original multi-span DBP system (4.17), where it describes the evolution of the adjoint variable vector $\boldsymbol{\lambda}$ over time and space. As opposed to the original problem (4.17), it can be noticed from the opposite sign of the spatial derivative matrix ($-K_c$) that the adjoint problem is solved in a reversed direction to the backward direction Z at which the original DBP problem (4.17) is solved. In other words, the adjoint DBP problem (4.26) is solved in the forward direction, $z = L_{tot} - Z$, where L_{tot} is the total fiber length.

Figure 4.4 depicts the adjoint DBP simulation (4.26) as opposed to the original DBP simulation (4.17). As shown, each adjoint fiber span (AFS) has the same absolute value of dispersion coefficients, but the signs are reversed as compared to the original DBP. Since the dispersion coefficients of the virtual fiber spans have their

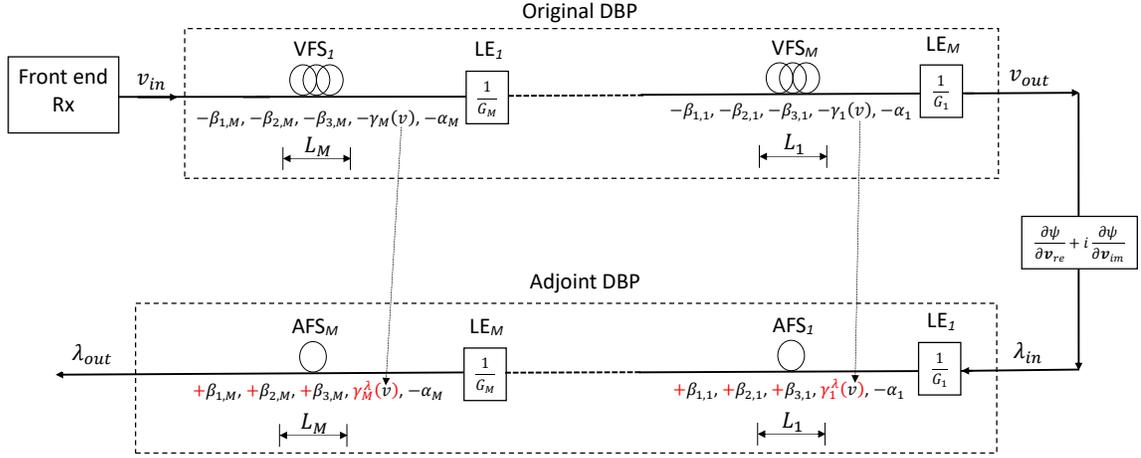


Figure 4.4: Propagation model of an original DBP simulation, and its corresponding adjoint DBP simulation.

Rx: receiver; DBP: digital back propagation; VFS: virtual fiber span; LE: loss element; AFS: adjoint fiber span.

signs opposite of the physical fiber spans, dispersion coefficients of the adjoint DBP spans have the same sign as that of the physical fiber spans. The gain coefficient ($-\alpha_j$) of each AFS is identical to its corresponding virtual fiber span (VFS) gain in the original simulation. All loss elements also remain the same in the adjoint simulation. The input of the adjoint problem λ_{in} is obtained by differentiating the kernel of the objective function integral ψ with respect to the original DBP field v . The nonlinear behavior of each AFS is different than its corresponding VFS. However, each AFS nonlinear coefficient is a function of the original field v (not the adjoint field λ) at its corresponding AFS in the original DBP simulation, as declared in Fig. 4.4.

Once the original DBP response \mathbf{V} and its corresponding adjoint response $\mathbf{\lambda}$ are evaluated for every time step, in all the computational domain, through solving Eqs. (4.17) and (4.26), respectively, the adjoint sensitivities of the objective function F

are evaluated through:

$$\frac{\partial F}{\partial x_k} = \int_{-T_m}^{T_m} \frac{\partial^e \psi}{\partial x_k} dt - \int_{-T_m}^{T_m} \boldsymbol{\lambda}^T \mathbf{R}_k dt, \quad k = 1, 2, \dots, N. \quad (4.29)$$

4.2.1 The Adjoint DBP Problem Solution

Excluding the adjoint amplifier term ($\mathbf{G}\boldsymbol{\lambda}$) arising in the adjoint multi-span DBP system (4.26), it is clear that the adjoint DBP problem (4.26) has a similar form to the adjoint problem of the NLSE (see Eq. (3.12) in Chapter 3), but here it is carried out for back propagation which means that the signs of dispersion and nonlinear coefficients are reversed. It thus follows that the adjoint DBP problem (4.26), without the adjoint amplifier effect, is solved using the same modified SSFS numerical algorithm introduced in Section 3.3 of Chapter 3.

The flow diagram, illustrated in Fig. 4.5, summarizes the modified SSFS algorithm utilized for solving the adjoint multi-span DBP problem (4.26). First, the distance parameter z is set to zero, the current span index is set as $j = 1$, and the initial adjoint field is calculated, according to the desired objective function as follows:

$$\lambda^l = \frac{\partial \psi}{\partial v_{re}} + i \frac{\partial \psi}{\partial v_{im}}, \quad (4.30)$$

where ψ is the integral kernel of the desired objective function, and $i = \sqrt{-1}$. The fields v_{re} and v_{im} are the real and imaginary parts of the output solution of the original DBP problem: $\mathbf{V}_{out} = [v_{re} \ v_{im}]^T$. Notice that, in practice, the objective function F is defined in terms of the DBP output signal \mathbf{V}_{out} . We therefore assume that the adjoint excitation $\mathbf{Q}_{in}^\lambda(t) = \partial \psi / \partial \mathbf{V}$ is nonzero only at the input of the adjoint DBP problem, i.e., at $z = 0$ or $Z = L_{tot}$, where L_{tot} is the total fiber length.

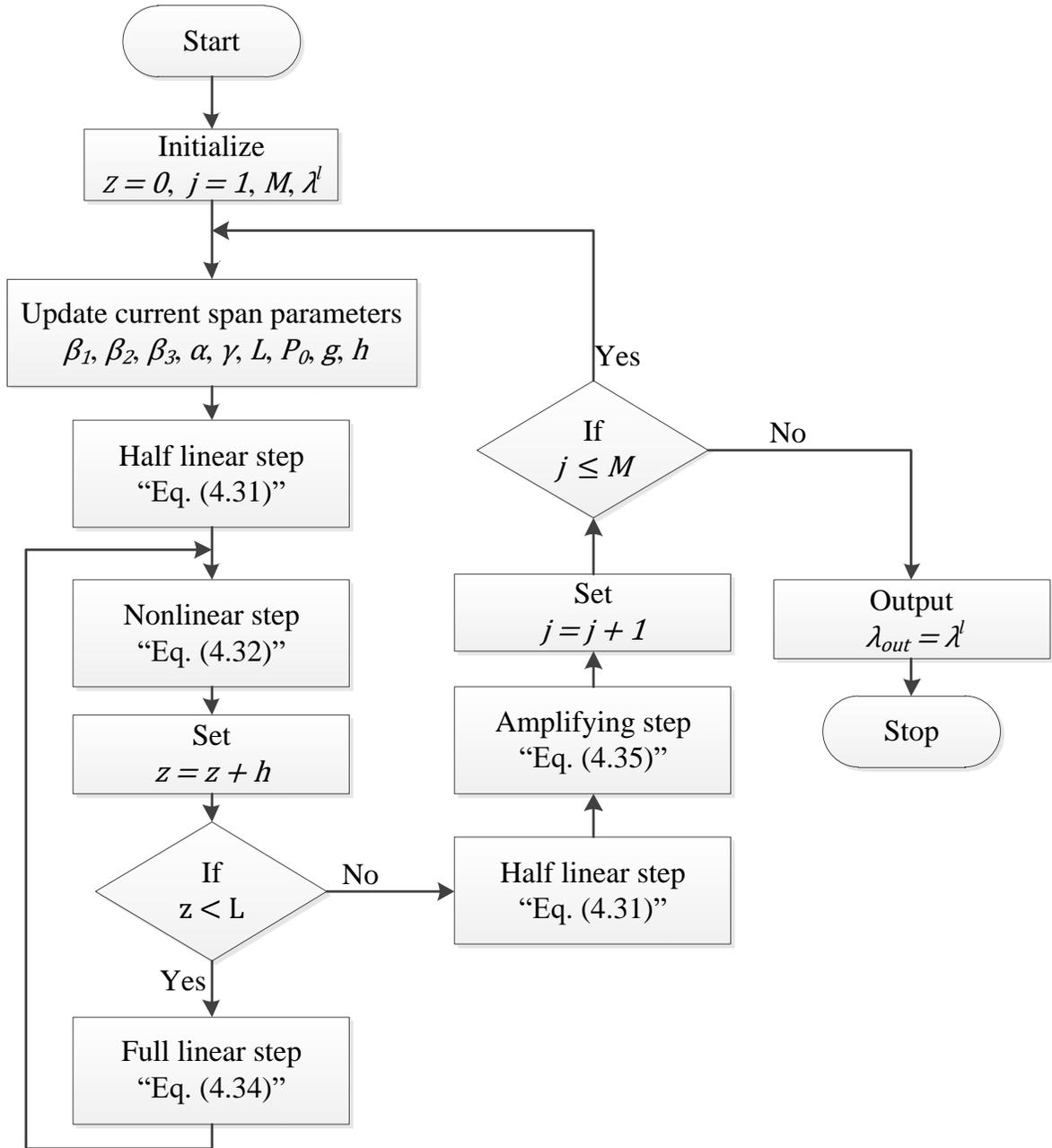


Figure 4.5: Flow diagram of the modified symmetric SSFS algorithm used for solving the adjoint DBP problem.

In the 2nd block (see Fig. 4.5), we update the parameters of the current adjoint DBP and amplifier j^{th} span, and determine the step size(s) utilized for the current span. Then, a first half linear step is performed in the 3rd block as follows:

$$A^l = \mathcal{F} \{ \lambda^l \}, \quad (4.31a)$$

$$\Lambda^l = A^l \times H(\omega, h/2), \quad (4.31b)$$

$$\lambda^l = \mathcal{F}^{-1} \{ \Lambda^l \}, \quad (4.31c)$$

$$\widehat{\boldsymbol{\lambda}}^l = \left[\text{Re} \{ \lambda^l \} \quad \text{Im} \{ \lambda^l \} \right]^T, \quad (4.31d)$$

where $H(\omega, z) = \exp \left(\left[-\frac{i\beta_3\omega^3}{6} + \frac{i\beta_2\omega^2}{2} - i\beta_1\omega + \frac{\alpha}{2} \right] z \right)$ is the linear transfer function of the adjoint DBP problem, and ω is the angular frequency. The operations $\mathcal{F} \{ \cdot \}$ and $\mathcal{F}^{-1} \{ \cdot \}$ denote Fourier and inverse Fourier transformations, respectively.

The nonlinear step executed in the 4th block includes the following operations:

$$\widehat{\boldsymbol{W}} = \boldsymbol{P}^{-1} \widehat{\boldsymbol{\lambda}}^l, \quad (4.32a)$$

$$\widehat{\boldsymbol{W}} = \exp \left(\boldsymbol{D}_\mu(v) h \right) \widehat{\boldsymbol{W}}, \quad (4.32b)$$

$$\widehat{\boldsymbol{\lambda}}^{nl} = \boldsymbol{P} \widehat{\boldsymbol{W}}, \quad (4.32c)$$

$$\lambda^l = \lambda_{re}^{nl} + i\lambda_{im}^{nl}, \quad (4.32d)$$

where $\widehat{\boldsymbol{\lambda}}^{nl} = \left[\lambda_{re}^{nl} \quad \lambda_{im}^{nl} \right]^T$, and the matrices \boldsymbol{P}^{-1} , \boldsymbol{D}_μ , and \boldsymbol{P} are given as:

$$\mathbf{P}^{-1}(v) = \begin{bmatrix} -\frac{v_{re}v_{im}}{\sqrt{3}|v|^2} - i\frac{1}{2} & \frac{(|v|^2+2v_{re}^2)}{2\sqrt{3}|v|^2} \\ -\frac{v_{re}v_{im}}{\sqrt{3}|v|^2} + i\frac{1}{2} & \frac{(|v|^2+2v_{re}^2)}{2\sqrt{3}|v|^2} \end{bmatrix}, \quad (4.33a)$$

$$\mathbf{D}_\mu(v) = i\sqrt{3}\gamma P_0|v|^2 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad (4.33b)$$

$$\mathbf{P}(v) = \begin{bmatrix} i & -i \\ \frac{\sqrt{3}|v|^2+2iv_{re}v_{im}}{|v|^2+2v_{re}^2} & \frac{\sqrt{3}|v|^2-2iv_{re}v_{im}}{|v|^2+2v_{re}^2} \end{bmatrix}, \quad (4.33c)$$

where $v = v_{re} + iv_{im}$ is the original DBP problem solution at the current spatial point. After applying the nonlinear step, the distance parameter is updated. If z is still less than the length of the current fiber span, a full linear step is performed as follows:

$$A^l = \mathcal{F} \{ \lambda^l \}, \quad (4.34a)$$

$$A^l = A^l \times H(\omega, h), \quad (4.34b)$$

$$\lambda^l = \mathcal{F}^{-1} \{ A^l \}, \quad (4.34c)$$

$$\hat{\lambda}^l = \left[\text{Re} \{ \lambda^l \} \quad \text{Im} \{ \lambda^l \} \right]^T. \quad (4.34d)$$

On the other hand, if z is larger than or equal the current fiber span length, the last half linear step is applied as given by Eq. (4.31). Then, an amplifier/loss step is performed as follows:

$$\lambda^l = \exp(g) \lambda^l. \quad (4.35)$$

Once the simulation of the current j^{th} span is completed, the span index j is updated.

For $j \leq M$, where M is total number of spans, the algorithm continues by updating the parameters of the new span and so on. Once the current j exceeds M , the final output of the adjoint DBP problem, $\lambda_{out} = \lambda^l$, is assigned, and the algorithm terminates.

4.2.2 Computational Complexity of the modified SSFS Algorithm

It should be clear that the computational complexity of each adjoint nonlinear step in the modified SSFS algorithm is larger than that in the standard SSFS algorithm due to the extra required diagonalization step, i.e., Eqs. (4.32a) and (4.32c). The adjoint linear step though has the same complexity of the linear step in the standard SSFS algorithm.

To analyze the complexity of the modified SSFS algorithm shown in Fig. 4.5, we consider a block size of N_s samples. One step of the adjoint linear step, Eq. (4.31) or Eq. (4.34), requires the evaluation of two N_s -point complex fast Fourier transforms (FFTs) and N_s complex multiplications. Since each complex FFT operation costs $0.5N_s \log_2(N_s)$ complex multiplications, the complexity of one linear step is approximately $N_s + N_s \log_2(N_s)$ complex multiplications (Spinnler, 2010; Gao *et al.*, 2012; Zhang *et al.*, 2015). Note that one complex multiplication involves 4-real multiplications (Alt and van Leeuwen, 1981). Therefore, the total number of real multiplications per one adjoint linear step is

$$4N_s \left(1 + \log_2(N_s) \right). \quad (4.36)$$

As for the first adjoint nonlinear sub-step, Eq. (4.32a), evaluating the components of the \mathbf{P}^{-1} matrix costs $9N_s$ –real multiplications, and multiplying by the vector $\widehat{\boldsymbol{\lambda}}^l$ requires $4N_s$ –real multiplication. Notice that the squared magnitude function evaluation requires two real multiplications, and the complexity of a real division operation is the same as that of a real multiplication (Alt and van Leeuwen, 1981). To calculate the components of the \mathbf{D}_μ matrix, extra $2N_s$ –real multiplications are needed. Multiplying \mathbf{D}_μ by the current step h needs $2N_s$ extra real multiplications. The exponential function $\exp(\mathbf{D}_\mu h)$ is then implemented with a lookup table, while the multiplication in $\exp(\mathbf{D}_\mu h) \times \widehat{\mathbf{W}}$ requires $2N_s$ more complex multiplications, or $8N_s$ more real multiplications. Thus, the total cost of sub-step (4.32b) is $12N_s$ –real multiplications. Evaluating the \mathbf{P} matrix components in sub-step (4.32c) requires extra $5N_s$ –real multiplications, and the multiplication of $\mathbf{P} \times \widehat{\mathbf{W}}$ needs extra $12N_s$ –real multiplications ($4N_s$ –real multiplications + $2N_s$ –complex multiplications). Therefore, each adjoint nonlinear step, Eq. (4.32), in the modified SSFS algorithm requires a total $42N_s$ –real multiplications.

Suppose the entire M –span adjoint DBP simulation is performed using K_{NLS} nonlinear steps. The total number of linear steps needed will thus be $K_{NLS} + M$, where the step length used in the first and last linear steps of each span is a half-step, i.e., $h/2$. However, all intermediate linear steps within a span use full length step h , as illustrated in Fig. 4.5. Also, the total number of real multiplications required by the amplifier step, Eq. (4.35), is $2N_s M$, where each amplifying step costs $2N_s$ –real multiplications. In total, the number of real multiplications for the modified

Table 4.1: Computational complexity of the conventional and modified symmetric split-step Fourier scheme (SSFS) algorithms.

Algorithm	Number of needed real multiplications
Conventional SSFS (Spinnler, 2010)	$4N_s (K_{NLS} + M) (1 + \log_2 (N_s)) + 10N_s K_{NLS} + 2N_s M$
Modified SSFS	$4N_s (K_{NLS} + M) (1 + \log_2 (N_s)) + 42N_s K_{NLS} + 2N_s M$

symmetric SSFS algorithm, given in Fig. 4.5, is:

$$4N_s (K_{NLS} + M) (1 + \log_2 (N_s)) + 42N_s K_{NLS} + 2N_s M, \quad (4.37)$$

where M , N_s , K_{NLS} are the number of spans, the number of samples, and the number of performed nonlinear steps, respectively. The computational complexity of the modified SSFS algorithm as compared to the complexity of the conventional SSFS algorithm are summarized in Table 4.1. As can be seen, an overhead of $32N_s$ -real multiplications per step are required by the modified SSFS algorithm, as opposed to the conventional SSFS algorithm, which is due to the extra diagonalization step needed within each adjoint nonlinear step.

In order to better compare the computational cost of the conventional SSFS method to the modified SSFS method cost, we define a Modified SSFS overhead parameter τ as the ratio of the modified SSFS cost over the conventional SSFS cost. In other words, the Modified SSFS overhead parameter τ , is given as:

$$\tau = 1 + \frac{32K_{NLS}}{4(K_{NLS} + M) (1 + \log_2 (N_s)) + 10K_{NLS} + 2M}. \quad (4.38)$$

In practical simulations, the number of performed nonlinear steps K_{NLS} is usually much larger than the number of spans M . We may therefore neglect the M -terms,

simplifying Eq. (4.38) to:

$$\tau = 1 + \frac{16}{7 + 2\log_2(N_s)}. \quad (4.39)$$

For example, with $N_s = 2 \times 2^{17}$, $M = 20$, and $K_{NLS} = 540$, using Eq. 4.38, we find $\tau = 1.36$, and using Eq. 4.39, we find $\tau = 1.37$. Thus, the computational overhead for the modified SSFS over the conventional SSFS is around 40%.

Next, we introduce the adjoint-based optimization algorithm proposed for accelerating the training process of the adaptive digital back propagation parameters.

4.3 Adjoint-Based Optimization Algorithm

In order to train and optimize the design parameters of an adaptive digital back propagation (A-DBP), a constrained nonlinear optimization problem has to be solved. A general formulation to this problem is given as:

$$\begin{aligned} & \min_{\mathbf{x}} F(\mathbf{x}, \mathbf{V}) \\ & \text{subject to} \\ & c_j(\mathbf{x}) \geq 0, \quad j = 1, 2, \dots, m, \end{aligned} \quad (4.40)$$

where $F(\mathbf{x}, \mathbf{V})$ is a general nonlinear objective function that measures the error between the current DBP output and the desired one, $\mathbf{x} \in \mathbb{R}^N$ is the DBP design parameters vector, and \mathbf{V} is the system state vector. The inequality constraints $c_j(\mathbf{x})$, $j = 1, 2, \dots, m$, might generally be nonlinear functions of the optimization variables \mathbf{x} , and m is the total number of constraints. Particularly, in the training

of the adaptive DBP, the constraints $c_j(\mathbf{x})$ are defined to restrict the variation of the optimization parameters within the practical ranges, and/or to guide the optimizer to the global optimum solution. It should be clear that the derivatives of each constraint function $c_j(\mathbf{x})$ with respect to the design parameters \mathbf{x} can be obtained analytically, using the definition of $c_j(\mathbf{x})$. Also, note that in the definition of the optimization problem (4.40), only inequality constraints are considered. However, an equal constraint $h(\mathbf{x}) = 0$ can simply be imposed in (4.40) as a couple of two inequality constraints given by (Rao, 2019): $h(\mathbf{x}) \geq 0$ and $-h(\mathbf{x}) \geq 0$.

The first order optimality condition of problem (4.40) is obtained utilizing the Lagrangian function (Rao, 2019). Although the Lagrangian function is applied on equality constraints, we may define an active set $\mathcal{A}(\mathbf{x})$ that contains all indices of the inequality constraints $c_j(\mathbf{x})$ that are active at the current design vector \mathbf{x} . In other words, \mathcal{A} includes all indices of $c_j(\mathbf{x})$ that are equal to 0 at the current \mathbf{x} . The active set \mathcal{A} is mathematically defined as (Rao, 2019):

$$\mathcal{A}(\mathbf{x}) = l : c_l(\mathbf{x}) = 0. \quad (4.41)$$

The Lagrangian function corresponding to problem (4.40) is thus formulated as follows:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\pi}) = F(\mathbf{x}, \mathbf{V}) - \sum_{l \in \mathcal{A}(\mathbf{x})} \pi_l c_l(\mathbf{x}), \quad (4.42)$$

where the Lagrangian multipliers π_l are positive nonzero scalars. The first order

Karush-Kuhn-Tucker (KKT) optimality conditions of Eq. (4.42) are then given by:

$$\nabla \mathcal{L}(\mathbf{x}, \boldsymbol{\pi}) = \begin{pmatrix} \nabla_{\mathbf{x}} F - \sum_{l \in \mathcal{A}} \pi_l \nabla_{\mathbf{x}} c_l \\ -\mathbf{C} \end{pmatrix} = \mathbf{0}, \quad (4.43)$$

where $\pi_l \neq 0, \forall l \in \mathcal{A}(\mathbf{x})$, and ∇ denotes gradient operation with respect to the vectors \mathbf{x} and $\boldsymbol{\pi}$, i.e., $\nabla = [\nabla_{\mathbf{x}}^T \ \nabla_{\boldsymbol{\pi}}^T]^T = [\partial/\partial x_1, \partial/\partial x_2, \dots, \partial/\partial x_N, \partial/\partial \pi_{i_1}, \partial/\partial \pi_{i_2}, \dots, \partial/\partial \pi_{i_{q(\mathbf{x})}}]^T$. The active constraint vector $\mathbf{c}(\mathbf{x})$ is a column vector that contains all active constraints at certain \mathbf{x} , i.e., $\mathbf{C}(\mathbf{x}) = [c_{i_1} \ c_{i_2} \ \dots \ c_{i_{q(\mathbf{x})}}]^T$, where $\mathcal{A}(\mathbf{x}) = \{i_1, i_2, \dots, i_{q(\mathbf{x})}\}$, and $q(\mathbf{x}) \leq m$ is the number of active constraints at \mathbf{x} . Starting from an initial guess \mathbf{x}_0 and $\boldsymbol{\pi}_0$, the optimal set $(\mathbf{x}^*, \boldsymbol{\pi}^*)$ that satisfies all constraints of (4.40) and at which the KKT conditions (4.43) holds can be obtained recursively using any well-established nonlinear optimization algorithm. In this work, we utilize the sequential quadratic programming (SQP) algorithm (Palomares and Mangasarian, 1976; Han, 1976; Nocedal and Wright, 2006).

SQP is one of the most effective methods for the numerical solution of nonlinear constrained optimization problems. The basic idea of the SQP method is to model the objective function by its local quadratic function approximation at the current design point \mathbf{x}_k , and to linearize the most violated constraint(s) at point \mathbf{x}_k (i.e. current active constraints) in order to construct a quadratic programming (QP) subproblem. The QP subproblem is then solved, using Newton or quasi-Newton method, to obtain a new solution iterate \mathbf{x}_{k+1} . The procedure is repeated iteratively till converging to a local optimum \mathbf{x}^* that minimizes the objective function and satisfies all problem constraints.

Before introducing a framework of the SQP method, we first recall the linear and

quadratic approximations of a scalar function $g(\mathbf{x})$ given, respectively, as (Nocedal and Wright, 2006):

$$g(\mathbf{x}) \approx g_k + \nabla g_k (\mathbf{x} - \mathbf{x}_k), \quad (4.44)$$

$$g(\mathbf{x}) \approx g_k + \nabla g_k (\mathbf{x} - \mathbf{x}_k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_k)^T \mathbf{H} g_k (\mathbf{x} - \mathbf{x}_k), \quad (4.45)$$

where ∇g_k and $\mathbf{H} g_k$ are the gradient vector and the Hessian matrix of the function $g(\mathbf{x})$ at \mathbf{x}_k , respectively, and g_k denotes $g(\mathbf{x}_k)$.

At a current iterate solution \mathbf{x}_k , a basic SQP algorithm determines the search direction of the next iterate by solving the following QP subproblem (Palomares and Mangasarian, 1976; Han, 1976; Nocedal and Wright, 2006):

$$\min_{\mathbf{s}_k} (\nabla_{\mathbf{x}} F_k)^T \mathbf{s}_k + \frac{1}{2} \mathbf{s}_k^T (\nabla_{\mathbf{x}\mathbf{x}}^2 \mathcal{L}_k) \mathbf{s}_k$$

subject to

$$c_l(\mathbf{x}_k) + (\nabla_{\mathbf{x}} c_l(\mathbf{x}_k))^T \mathbf{s}_k = 0, \forall l \in \mathcal{A}(\mathbf{x}_k), \quad (4.46)$$

where $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ is the search direction for the next iterate solution \mathbf{x}_{k+1} . The matrix $\nabla_{\mathbf{x}\mathbf{x}}^2 \mathcal{L}_k$ is the hessian matrix of the Lagrangian function $\mathcal{L}(\mathbf{x}, \boldsymbol{\pi})$, given in (4.42), with respect to the \mathbf{x} parameters only. This QP problem is obtained through replacing the Lagrangian function (4.42) by its local quadratic approximation at \mathbf{x}_k . The current active constraints are also replaced by their local linear approximations at current \mathbf{x}_k . Using the constraint given in (4.46) and by omitting the constant terms, it can be shown that the objective function provided in the QP subproblem (4.46) is

equivalent to the local quadratic approximation of the Lagrangian function $\mathcal{L}(\mathbf{x}, \boldsymbol{\pi})$ at \mathbf{x}_k (Nocedal and Wright, 2006).

In order to solve problem (4.46), we apply Newton’s method to solve the nonlinear system of equations given by the KKT conditions of (4.46), given as (Nocedal and Wright, 2006):

$$\begin{pmatrix} \nabla_{\mathbf{x}\mathbf{x}}^2 \mathcal{L}_k \mathbf{s}_k + \nabla_{\mathbf{x}} F_k - \mathbf{J}_k^T \boldsymbol{\pi}_{k+1} \\ \mathbf{J}_k \mathbf{s}_k + \mathbf{C}_k \end{pmatrix} = \mathbf{0}, \quad (4.47)$$

where \mathbf{J}_k is the Jacobian matrix of the active constraints $\mathbf{C}(\mathbf{x})$ at \mathbf{x}_k , defined by

$$\mathbf{J}(\mathbf{x})^T = \left[\nabla_{C_{i_1}}(\mathbf{x}), \nabla_{C_{i_2}}(\mathbf{x}), \dots, \nabla_{C_{i_{q(\mathbf{x})}}(\mathbf{x})} \right], \quad (4.48)$$

where $\mathcal{A}(\mathbf{x}) = \{i_1, i_2, \dots, i_{q(\mathbf{x})}\}$, and $q(\mathbf{x}) \leq m$. Reformulating Eq. (4.47), the search direction of the next iterate \mathbf{s}_k and the new Lagrangian multipliers $\boldsymbol{\pi}_{k+1}$ can thus be identified with the solution of

$$\begin{bmatrix} \nabla_{\mathbf{x}\mathbf{x}}^2 \mathcal{L}_k & -\mathbf{J}_k^T \\ \mathbf{J}_k & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{s}_k \\ \boldsymbol{\pi}_{k+1} \end{bmatrix} = \begin{bmatrix} -\nabla_{\mathbf{x}} F_k \\ -\mathbf{C}_k \end{bmatrix}. \quad (4.49)$$

Clearly, the solution of Eq. (4.49) requires the computation of the Lagrangian function’s Hessian $\nabla_{\mathbf{x}\mathbf{x}}^2 \mathcal{L}(\mathbf{x}_k, \boldsymbol{\pi}_k)$. However, for lower computational cost, this Hessian matrix might be replaced with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) approximation (Bakr, 2013). The BFGS method is one of the most robust quasi-Newton techniques developed for the Hessian matrix approximation, utilizing the first-order derivatives information. Suppose that the Hessian matrix $\nabla_{\mathbf{x}\mathbf{x}}^2 \mathcal{L}_k$ at the k^{th} iteration is replaced by the approximate Hessian matrix \mathbf{B}_k . A better solution \mathbf{x}_{k+1} is thus

determined by solving the quasi-Newton-KKT system of equations, given as:

$$\begin{bmatrix} \mathbf{B}_k & -\mathbf{J}_k^T \\ \mathbf{J}_k & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{s}_k \\ \boldsymbol{\pi}_{k+1} \end{bmatrix} = \begin{bmatrix} -\nabla_{\mathbf{x}} F_k \\ -\mathbf{C}_k \end{bmatrix}. \quad (4.50)$$

The approximation of the Hessian matrix at the new point \mathbf{x}_{k+1} is then obtained through the BFGS updating formula (Bakr, 2013):

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k} + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}, \quad (4.51)$$

where $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ and $\mathbf{y}_k = \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}_{k+1}, \boldsymbol{\pi}_{k+1}) - \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}_k, \boldsymbol{\pi}_k)$. In most Quasi-Newton methods, the initial Hessian matrix approximation \mathbf{B}_0 is set to an identity matrix (Bakr, 2013). It is clear that the evaluation of the Lagrangian function's gradient at the k^{th} iterate, $\nabla_{\mathbf{x}} \mathcal{L}_k = \nabla_{\mathbf{x}} F(\mathbf{x}_k) - \mathbf{J}_k^T \boldsymbol{\pi}_k$, requires to calculate the gradient of the objective function $F(\mathbf{x}, \mathbf{V})$, and the Jacobian of the current active constraints $\mathbf{J}(\mathbf{x})$ at \mathbf{x}_k . It is worth to emphasize that the Jacobian matrix is evaluated analytically using the definition of the constraints $C_j(\mathbf{x})$, $j = 1, 2, \dots, m$, given in (4.40). Contrarily, the gradient of $F(\mathbf{x}, \mathbf{V})$, with respect to \mathbf{x} , cannot be determined analytically due to the implicit dependence on the system state vector \mathbf{V} . However, we exploit our proposed ASA algorithm, described in Section 4.2, to estimate the gradient of F at each optimization iterate using only one extra adjoint simulation. The introduced optimization algorithm is therefore denoted as an adjoint-based optimization (ABO) algorithm. We summarize the ABO algorithm in the next subsection.

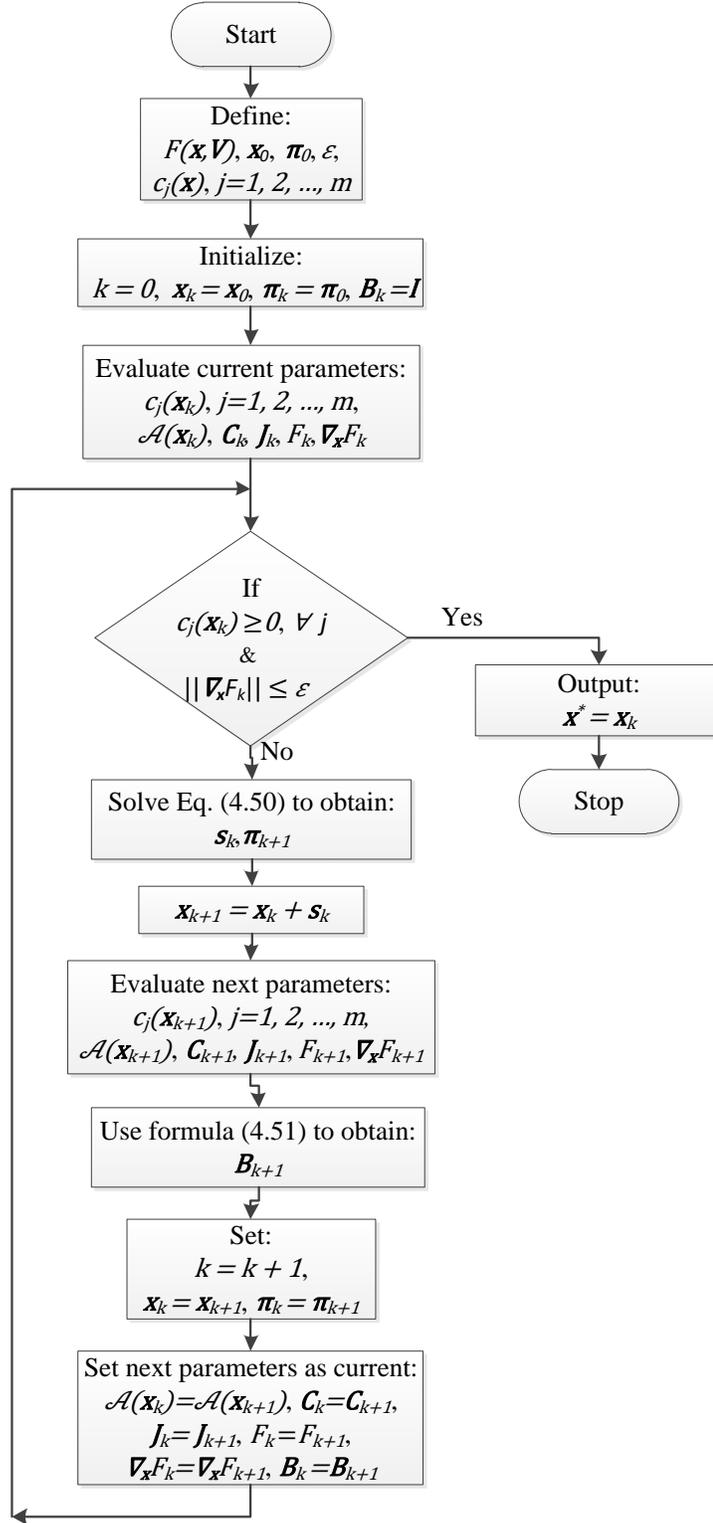


Figure 4.6: Flow diagram of the proposed adjoint-based optimization algorithm.

4.3.1 Algorithm

The ABO algorithm proposed for solving problem (4.40) is illustrated in Fig. 4.6. We first define the objective function $F(\mathbf{x}, \mathbf{V})$, the initial solution \mathbf{x}_0 , the initial Lagrangian multipliers $\boldsymbol{\pi}_0$, the stopping gradient criterion ε , and the problem constraints $c_j(\mathbf{x})$, $j = 1, 2, \dots, m$. Note that a feasible initial solution is assumed to be provided to the algorithm, i.e., a solution that satisfies the problem constraints. The parameters of the algorithm are then initialized. In step 3, we calculate the constraints $c_j(\mathbf{x})$, $j = 1, 2, \dots, m$, at the current \mathbf{x}_k in order to determine the current active constraints and update $\mathcal{A}(\mathbf{x}_k)$. The current active constraints vector \mathbf{C}_k and its Jacobian matrix \mathbf{J}_k are determined. Also, the original and adjoint DBP problems, (4.17) and (4.26), are performed to obtain the current objective function F_k and its gradient $\nabla_{\mathbf{x}} F_k$. If the current solution satisfies stopping criteria (satisfying all constraints and having a gradient norm value less than ε), we assign it to the optimal solution $\mathbf{x}^* = \mathbf{x}_k$, and terminate. Otherwise, we solve the quasi-Newton-KKT system of equations (4.50) to obtain the next search direction and determine the new solution set $(\mathbf{x}_{k+1}, \boldsymbol{\pi}_{k+1})$, where $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$. We then repeat step 3 to update the indices of the next active constraints set $\mathcal{A}(\mathbf{x}_{k+1})$. Also, we evaluate the next active constraints vector \mathbf{C}_{k+1} , the next Jacobian matrix \mathbf{J}_{k+1} , the next objective function F_{k+1} , and the next point gradient value $\nabla_{\mathbf{x}} F_{k+1}$. Then, the next Hessian matrix approximation \mathbf{B}_{k+1} is evaluated using the BFGS formula (4.51). The iteration index is incremented $k = k + 1$, and the values of all next algorithm parameters are set as the current values, i.e., $\mathcal{A}(\mathbf{x}_k) = \mathcal{A}(\mathbf{x}_{k+1})$, $\mathbf{C}_k = \mathbf{C}_{k+1}$, $\mathbf{J}_k = \mathbf{J}_{k+1}$, $F_k = F_{k+1}$, $\nabla_{\mathbf{x}} F_k = \nabla_{\mathbf{x}} F_{k+1}$, $\mathbf{B}_k = \mathbf{B}_{k+1}$, $\mathbf{x}_k = \mathbf{x}_{k+1}$, and $\boldsymbol{\pi}_{k+1} = \boldsymbol{\pi}_k$. If the stopping criteria are not satisfied yet, we again solve Eq. (4.50) for a better solution. The algorithm is

repeated until satisfying the stopping criteria test. Once the optimal solution is obtained (i.e. stopping conditions are satisfied), the final solution $\mathbf{x}^* = \mathbf{x}_k$ is assigned, and the algorithm terminates.

Next, we investigate the robustness and efficiency of the proposed ABO algorithm for the training of an adaptive DBP, through a number of numerical examples.

4.4 Results

For all results presented in this section, we consider a Monte-Carlo simulations of the single-channel fiber-optic communication system, shown in Fig. 4.7. The system is operating at 28 Gbaud, and the transmitted data have a root raised cosine pulse shaping with a roll-off factor of 0.1. The modulation format used in the simulation is a 16-quadrature amplitude modulation (16-QAM) format. The transmission channel is a multi-span optical fiber system with the same configuration illustrated in Fig 4.2. Each span consists of a standard single-mode fiber followed by an inline erbium-doped fiber amplifier (EDFA). The simulation parameters of the fiber are given in Table 4.2. The length of each fiber span is different, with values varying between 50 km and 125 km. This is the typical range of the fiber span lengths in North America (Kumar and Deen, 2014). The EDFA at the end of each fiber span has a gain that fully compensates for the respecting fiber loss, with a noise figure $N_f = 4.77$ dB. In our simulations, we particularly consider an 4–span optical fiber communication system scenario with fiber spans lengths as given in Table 4.3. The forward propagation in these scenarios are simulated using the split-step Fourier scheme (SSFS) algorithm, with an adaptive step-size guaranteeing that the maximum nonlinear phase per step does not exceed 0.01 radian.

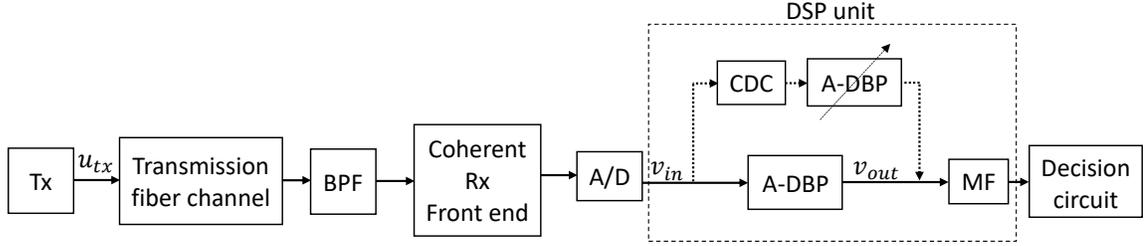


Figure 4.7: Block diagram of the fiber-optic communication system considered in the simulation.

Tx: transmitter; BPF: band pass filter; Rx: receiver; A/D: analog to digital converter; CDC: chromatic dispersion compensation; A-DBP: adaptive digital back propagation; MF: matched filter.

Table 4.2: Fiber simulation parameters.

Parameter	Value
Inverse group speed, β_1	0 ps/km
Second-order dispersion coefficient, β_2	-21 ps ² /km
Third-order dispersion coefficient, β_3	0 ps ³ /km
Loss coefficient, α	0.2 dB/km
Nonlinear coefficient, γ	1.1 W ⁻¹ km ⁻¹

A Gaussian band pass filter (BPF) with 50 GHz bandwidth is used before the coherent receiver. The front-end coherent receiver translates the in-phase and quadrature-phase components of the received optical signal into electrical signal components. An analog to digital (A/D) converter is utilized before the digital signal processing (DSP) unit, to reduce the sampling rate from 8 to 2 samples per symbol. The chromatic dispersion (CD) compensation block is used to estimate the total length of the transmission fiber link, L_{tot} . It is only applied during the training process of the adaptive digital back propagation (A-DBP). After being trained, the A-DBP compensates for the linear and nonlinear distortions induced by the fiber channel. The output of the A-DBP passes through a matched filter (i.e. a root raised cosine filter), to limit the noise. The symbol sequence after the filter is compared with the symbol sequence at

Table 4.3: Lengths of fiber spans used in the simulation of the 4–span fiber-optic communication system scenario.

Fiber span number	Length (km)
1	50
2	80
3	85
4	125

the transmitter and those symbols which have crossed the boundaries are counted as error symbols.

Unless otherwise is specified, the proposed ABO algorithm is used to train the A-DBP by solving the following A-DBP training problem:

$$\min_{\mathbf{x}} \int_{-T_m}^{T_m} |v_{out} - u_{tx}|^2 dt$$

subject to

$$\begin{cases} L_1 + L_2 + \dots + L_M = L_{tot}, \\ 50 \text{ km} \leq L_j \leq 125 \text{ km}, \quad j = 1, 2, \dots, M, \end{cases} \quad (4.52)$$

where v_{out} is the normalized output signal of the A-DBP, u_{tx} is the normalized transmitted signal in electrical domain, and T_m is half the window size used in the DSP unit. Notice that the number of samples per symbol considered in u_{tx} is reduced to the number of samples per symbol used in the DSP unit. In other words, u_{tx} is a noise-free back-to-back version of the normalized transmitted signal at the DSP unit. The design vector $\mathbf{x} = [L_1 \ L_2 \ \dots \ L_M]^T$ contains the lengths of all fiber spans, where

L_j is the fiber length of the j^{th} span. The training problem (4.52) aims at obtaining the optimal design vector \mathbf{x}^* that minimize the error between the equalized signal v_{out} (i.e. output signal of the A-DBP) and the desired output u_{tx} . Note that in order to reduce the estimation time, we assume that all fiber span parameters are known to the receiver, except for the length of each fiber. The proposed ABO algorithm is, however, general and can be used to estimate all fiber parameters as well as the average power launched to the fiber. In practice, most of the fiber links are standard single-mode fibers. Therefore, fiber span parameters other than the length (e.g. 2^{nd} order dispersion coefficient β_2 and nonlinear coefficient γ) are known and similar for all spans. Their values might only be subject to slight deviations, from nominal values given in Table 4.2, due to environmental changes.

It is clear that the objective function in (4.52) has a form similar to the general form of (4.19) which has been used to derive the adjoint sensitivity problem of the original DBP problem (4.17). Hence, our adjoint sensitivity analysis algorithm can be utilized to estimate the gradient of this objective function using one extra adjoint simulation. The equality constraint in (4.52) is denoted as the linear compensation (LC) constraint. It is introduced to guide the ABO to a global solution that minimizes both linear and nonlinear distortions of the fiber. Obviously, any DBP whose fiber span lengths satisfies the LC condition can effectively compensate for the linear distortions of the transmission fiber link. All these linear solutions (satisfying the LC condition) may be seen as local minima of problem (4.52) without the LC constraint. While solving it, the ABO might get stuck at any of these local minima points. We therefore introduce the LC constraint to guide the ABO algorithm to search for the global optimal solution (that minimizes both linear and nonlinear distortions) among

all possible linear solutions. Notice also that the constraints in problem (4.52) can be re-formulated to obtain the same general form of (4.40), as follows:

$$\min_{\mathbf{x}} F = \int_{-T_m}^{T_m} |v_{out} - u_{tx}|^2 dt$$

subject to

$$\left\{ \begin{array}{l} L_1 + L_2 + \dots + L_M - L_{tot} \geq 0, \\ -L_1 - L_2 - \dots - L_M + L_{tot} \geq 0, \\ L_j - 50 \text{ km} \geq 0, \quad j = 1, 2, \dots, M, \\ 125 \text{ km} - L_j \geq 0, \quad j = 1, 2, \dots, M. \end{array} \right. \quad (4.53)$$

The initial parameters of the ABO algorithm are given as: $\boldsymbol{\pi}_0 = [1 \ 1 \ \dots \ 1]^T$ and $\varepsilon = 10^{-12}$.

4.4.1 Validating the ASA Algorithm

We first investigate the accuracy and efficiency of the proposed ABO algorithm in training and optimizing the parameters of the A-DBP. We study the ability of the ABO to converge to the ideal DBP solution. We also compare our ABO algorithm to other finite-differences-based optimization algorithms in terms of convergence rate and computational complexity. Notice that the notation ideal DBP solution means a DBP that uses the exact parameters of the transmission channel, and its SSFS simulation is performed using a small step-size such that the DBP simulation has the same accuracy as the simulation of the forward propagation problem.

The ABO algorithm is applied to train a 4–span A-DBP system for mitigating the fiber impairments of the 4–span fiber communication system scenario. The virtual

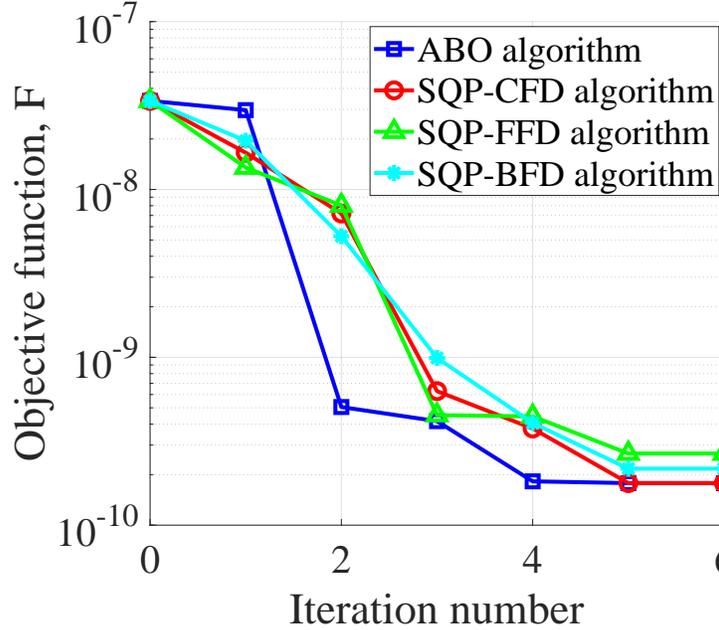


Figure 4.8: The value of the training objective function F in (4.52) versus the number of optimization iterates for the 4-span fiber-optic communication system scenario. Four optimization algorithms are considered for the training of the A-DBP, namely, the ABO algorithm, the SQP-CFD algorithm, SQP-FFD algorithm, and SQP-BFD algorithm.

fibers of the A-DBP has the same transmission fibers parameters given in Table 4.1. The number of known symbols used to train the A-DBP is 4096. The average launch power at the training process is $P_{av} = 7$ dBm. The total length of the channel (estimated using the CD compensation unit) is $L_{tot} = 340$ km. Initially, the 4 spans of the A-DBP are supposed to have an equal length of $L_{tot}/4$. In other words, the initial design point given to the ABO algorithm (in km) is $\mathbf{x}_0 = [85 \ 85 \ 85 \ 85]^T$. Clearly, this initial point is a feasible design point, i.e., it satisfies all the constraints of training problem (4.52).

Fig. 4.8 illustrates the evolution of the objective function in (4.53) with number of optimization iterates. The ABO algorithm terminates after 6 iterates with an

optimal solution $\mathbf{x}^* = [50 \ 79.92 \ 85.08 \ 125]^T$ km. These values are quite close to the actual fiber span lengths (See Table 4.3). The Euclidian norm of the objective function gradient at this point is $4.1 \times 10^{-16} < \varepsilon = 10^{-12}$. As can be seen, the ABO algorithm is capable of training the A-DBP, and successfully converges to the ideal DBP solution.

Next, 2^{17} data symbols are transmitted through the 4-span transmission system and pass through the trained A-DBP to retrieve the transmitted signal. Fig. 4.9 shows the obtained bit error rate (BER) in case of using the initial A-DBP or the trained A-DBP, as compared to the ideal DBP case, versus a sweep of the average launch power. The BER values are also compared to the BER obtained when a linear chromatic dispersion compensation (CDC) unit is only used at the receiver. As compared to the initial A-DBP, the trained A-DBP effectively mitigates the distortions at the higher launch power ranges, achieving as BER performance as the ideal DBP. The BER is found to be zero, using the trained A-DBP, for average fiber launch powers larger than -4 dBm and less than 14 dBm. For low launch power ($P_{av} \leq -8$ dBm), the fiber nonlinear distortion is negligible, and the distortions are mainly due to fiber CD effects and amplifiers' noise (Kumar and Deen, 2014). Therefore, the CDC unit or the initial A-DBP can provide a BER performance comparable to the trained A-DBP (as well as the ideal DBP) at this region. As the power slightly increases, the signal to noise ratio increases resulting in a BER reduction. However, at higher power levels ($P_{av} \geq -4$ dBm) the fiber nonlinearity becomes the dominant source of distortion, and the CDC cannot mitigate the impairments anymore. Contrarily, the trained A-DBP can still provide a BER performance benefit up to average launch power of 14 dBm. For very high launch power ($P_{av} > 14$ dBm), both the trained A-DBP and

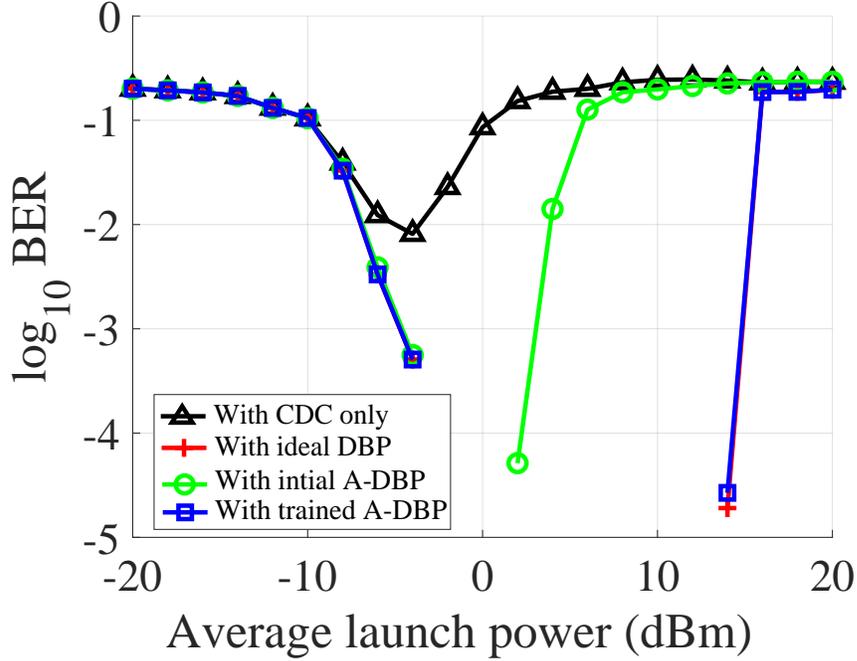


Figure 4.9: BER versus average launch power for the 4-span fiber-optic communication system scenario. The equalization at the receiver is performed using the CDC unit only, the ideal DBP, the DBP, or the A-DBP. The number of transmitted symbols is 2^{17} .

the ideal DBP fail to mitigate the distortions due to the severe signal-noise nonlinear interactions (Kumar and Deen, 2014).

Next, we re-train the initial A-DBP using the same optimization method (sequential quadratic programming (SQP) method). However, the gradient of the objective function is evaluated using the central-finite differences (CFD), the forward-finite differences (FFD), or the backward-finite differences (BFD) approach. These techniques are respectively denoted as SQP-CFD, SQP-FFD, and SQP-BFD. All the three algorithms (SQP-CFD, SQP-FFD, and SQP-BFD) terminate after 6-iterations with an objective function gradient norm of 1.19×10^{-16} , 2.9×10^{-14} , and 2.6×10^{-14} , respectively. The optimal design points obtained are given in Table 4.4. Owing to the

Table 4.4: Optimal design point solutions of the A-DBP in case of the 4-span fiber-optic communication system scenario. Four optimization algorithms are used to train the A-DBP, namely, the ABO algorithm, the SQP-CFD algorithm, the SQP-FFD algorithm, the SQP-BFD algorithm. The initial design point (in km) is $\mathbf{x}_0 = [85 \ 85 \ 85 \ 85]^T$.

Fiber span number	Fiber span’s optimum length (km)			
	ABO	SQP-CFD	SQP-FFD	SQP-BFD
1	50.00	50.00	51.85	50.00
2	79.92	80.01	78.21	78.74
3	85.08	84.99	85.57	86.26
4	125.00	125.00	124.37	125.00

proper definition of the training problem (4.52), all three optimizers converge to the ideal DBP solution as well as the ABO optimizer does. However, the solutions of the ABO and SQP-CFD algorithms are slightly better (more closer to the ideal solution) than the SQP-FFD and SQP-BFD solutions. This is due to the fact that the derivatives estimations of both the adjoint and CFD approaches are more accurate than the FFD and BFD estimations. Fig. 4.8 shows the objective function evolution in case of using the 3–finite-differences-based optimizers as compared to that of the ABO algorithm. Obviously, there is a slight difference between the routes each algorithm takes towards the optimal solution. This is due to the slight differences between the gradient estimations of each sensitivity technique.

Table 4.5 compares between the four optimizers (ABO, SQP-CFD, SQP-FFD, and SQP-BFD) in terms of the total number of SSFS simulations performed during the training process of the initial A-DBP. The total number of samples used during the training process is 4096×2 , where the number of training symbols is 4096, and number of samples per symbol used in the DSP unit is 2. Substituting in Eq. (4.39), the modified SSFS overhead parameter for these settings is $\tau \approx 1.48$. In other words, the

Table 4.5: Computational cost of sensitivity calculations required to train the A-DBP for mitigating the distortions of the 4–span fiber-optic communication system scenario. The A-DBP is trained using ABO, SQP-CFD, SQP-FFD, and SQP-BFD algorithms.

Algorithm	Number of iterations	Number of original DBP simulations	Number of adjoint DBP simulations	Total number of equivalent conventional SSFS simulations
ABO	6	6×1	6×1	14.88
SQP-CFD	6	$6 \times (1 + 2N)$	–	54
SQP-FFD	6	$6 \times (1 + N)$	–	30
SQP-BFD	6	$6 \times (1 + N)$	–	30

*The number of design parameters in this example is $N = 4$.

complexity of one adjoint DBP simulation run is approximately equivalent to $1.48 \times$ complexity of one original DBP simulation run. At each iterate of the ABO algorithm, we need to perform 1–original and 1–adjoint DBP simulations, to evaluate current objective function value and its gradient vector. This is equivalent to performing 2.48 conventional SSFS simulation runs. Contrarily, each SQP-CFD iterate requires $(2N + 1)$ –original DBP simulations, where N is the number of design parameters, to determine current objective function value and the gradient vector. Each of the SQP-FFD and SQP-BFD needs to run $(N + 1)$ –original DBP simulations every optimization iterate, to obtain the current objective function value and a less accurate (as compared to ABO and SQP-CFD algorithms) gradient vector estimation.

As shown in the last column of Table 4.5, our proposed ABO algorithm saves around 39–system simulations, as opposed to the SQP-CFD algorithm, to train the A-DBP with similar accuracy as that of the SQP-CFD. Moreover, with a better training accuracy than the SQP-FFD or SQP-BFD, our ABO algorithm prevents

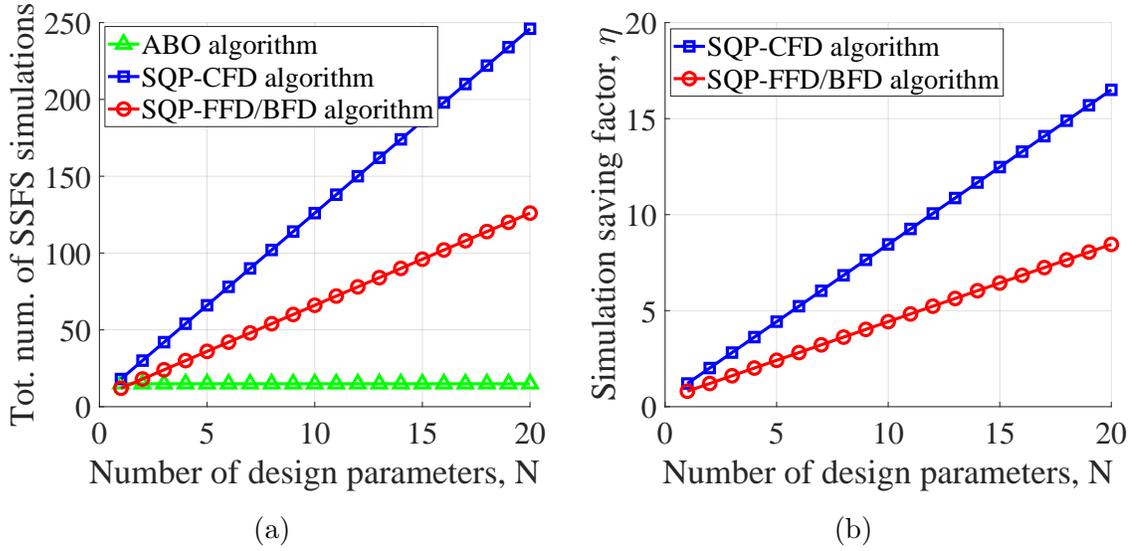


Figure 4.10: (a) Asymptotic number of conventional SSFS simulations required to train a A-DBP versus number of design parameters N ; (b) Asymptotic values of the simulation saving factor η achieved by the ABO algorithm versus design parameters number N . The ABO algorithm is compared to the SQP-CFD, SQP-FFD, and SQP-BFD. The number of training samples is 8192.

extra 15–system simulations required to train the 4–span A-DBP using the SQP-FFD or SQP-BFD algorithm. It should be emphasized that the number of extra simulations avoided by the ABO algorithm is expected to increase linearly as the number of parameters increases, as shown in Fig. 4.10a. This is due to the fact that the number of extra simulations required by the finite-differences-based optimizers (SQP-CFD, SQP-FFD, and SQP-BFD) is linearly proportional to the number of design parameters. Contrarily, the ABO optimizer always requires one extra adjoint simulation per iterate to evaluate the full gradient information whatever the number of parameters is.

In order to better clarify the gain achieved by the ABO algorithm in terms of the computational cost, we define a simulation saving factor η as the ratio of the total

number of conventional SSFS simulations required by a certain optimizer to train a A-DBP over the simulations' number needed by the ABO to train the same A-DBP. For this example, the value of η is calculated to be 3.63 and 2.02 for the SQP-CFD and SQP-FFD/BFD, respectively, when the number of design parameters = 4. In other words, our ABO algorithm is 3.63 times faster than the SQP-CFD in training the 4–span A-DBP. The asymptotic gain of applying the ABO algorithm to optimize the A-DBP parameters versus a sweep of the design parameters number is shown in Fig. 4.10b. Clearly, as the number of parameters increases, the gain achieved by the ABO algorithm, as opposed to other FD-based algorithms, soars.

4.4.2 Optimum Low-Complexity A-DBP

In this subsection, our objective is to reduce the computational complexity of the A-DBP. At the expense of a lower/worse equalization performance, we aim at accelerating the A-DBP equalization processing time. For a trained A-DBP or ideal DBP, we may replace the small step-size used in the SSFS simulation with a relatively larger step. Since the symmetric SSFS simulation has an error of order $O(h^3)$, a coarse-mesh A-DBP simulation with relatively larger step-size will not be identical to the inverse response of the fiber channel. As a result, the equalization performance of a fiber communication system with a coarse-mesh DBP model degrades as opposed to a fine-mesh DBP model. For the 4–span fiber-optic communication system scenario, Fig. 4.11 shows the BER performance of a coarse-mesh model of the ideal DBP as compared to the fine-mesh ideal DBP model. The adaptive step-size used in the coarse-mesh model is chosen to ensure that the maximum nonlinear phase change within each step is smaller than 0.1 radian, as opposed to 0.004 radian in

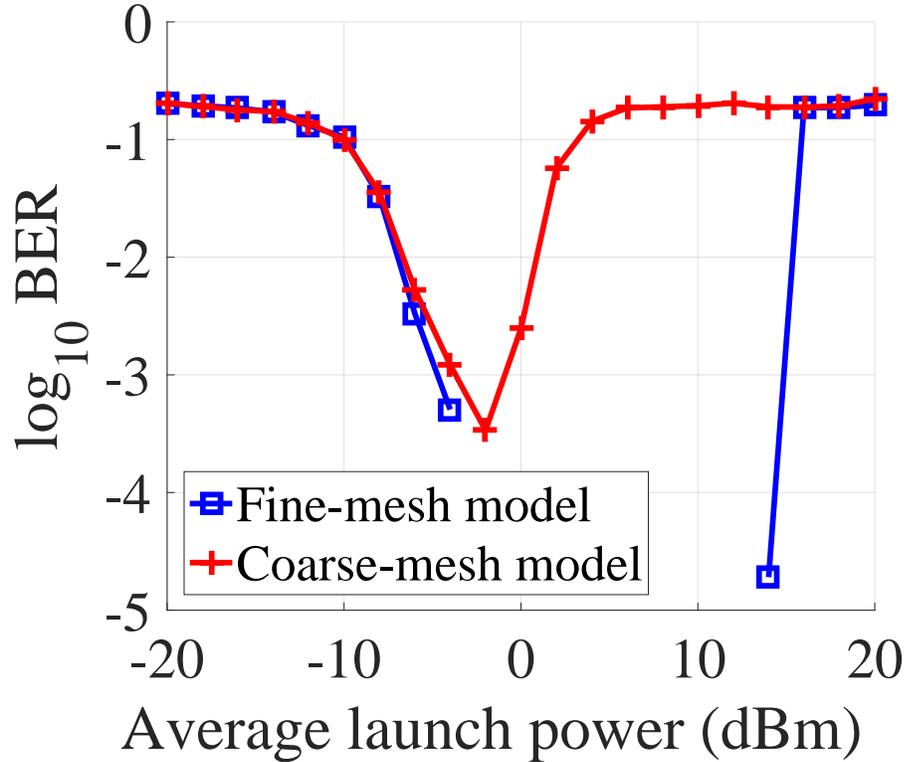


Figure 4.11: BER versus average launch power for the 4-span fiber-optic communication system scenario. The equalization at the receiver is performed using ideal DBP with fine-mesh and coarse-mesh models. The signal and system parameters are the same as that used for Fig. 4.9.

the fine-mesh case. As shown in Fig. 4.11, the BER performance of the coarse-mesh ideal DBP model deteriorates significantly at average launch power above 0 dBm. However, the coarse-mesh ideal DBP model requires only a total of 4-SSFS steps at $P_{av} = 0$ dBm, as compared to 53-SSFS steps required by the fine-mesh ideal DBP model at the same launch power.

In order to enhance the performance of the coarse-mesh ideal DBP model with a computational cost within the same range, we apply our ABO algorithm to train

Table 4.6: Optimum values of the A-DBP design parameters for mitigating the 4–span fiber-optic communication system scenario. A coarse-mesh model is considered with a maximum allowed nonlinear phase change per SSFS step of 0.1 rad. The ABO is used to train A-DBP with 2–, 3–, and 4–virtual fiber spans.

Fiber span number	Optimum parameters					
	2–span A-DBP		3–span A-DBP		4–span A-DBP	
	L (km)	γ ($\text{W}^{-1}\text{km}^{-1}$)	L (km)	γ ($\text{W}^{-1}\text{km}^{-1}$)	L (km)	γ ($\text{W}^{-1}\text{km}^{-1}$)
1	123.47	3.69	132.15	3.82	50.50	1.47
2	216.53	18.82	1.36	20.00	80.69	1.32
3	–	–	206.49	9.97	16.85	1.69
4	–	–	–	–	191.96	6.08

virtual fiber span. The A-DBP training problem is modified as follows:

$$\min_{\mathbf{x}} F = \int_{-T_m}^{T_m} |v_{out} - u_{tx}|^2 dt$$

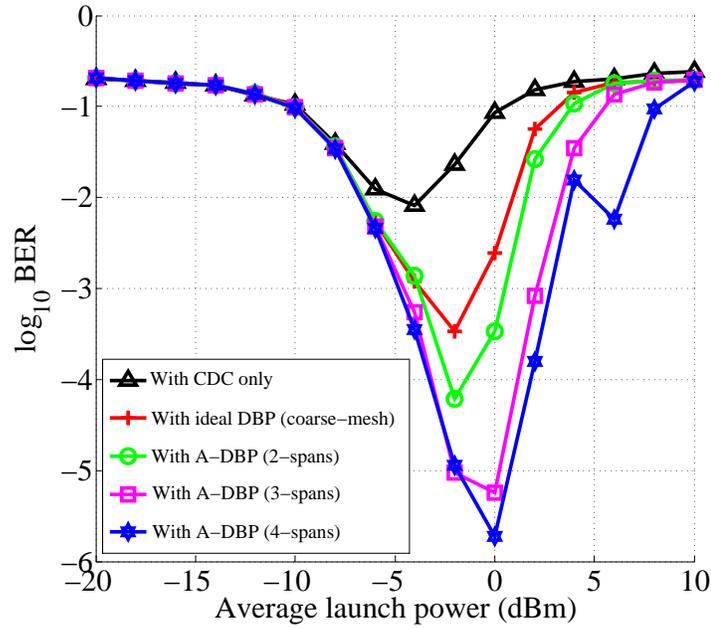
subject to

$$\begin{cases} L_1 + L_2 + \dots + L_M = L_{tot}, \\ 0 \text{ km} \leq L_j \leq L_{tot}, \quad j = 1, 2, \dots, M, \\ 0.001 \text{ W}^{-1}\text{km}^{-1} \leq \gamma_j \leq 20 \text{ W}^{-1}\text{km}^{-1}, \quad j = 1, 2, \dots, M, \end{cases} \quad (4.54)$$

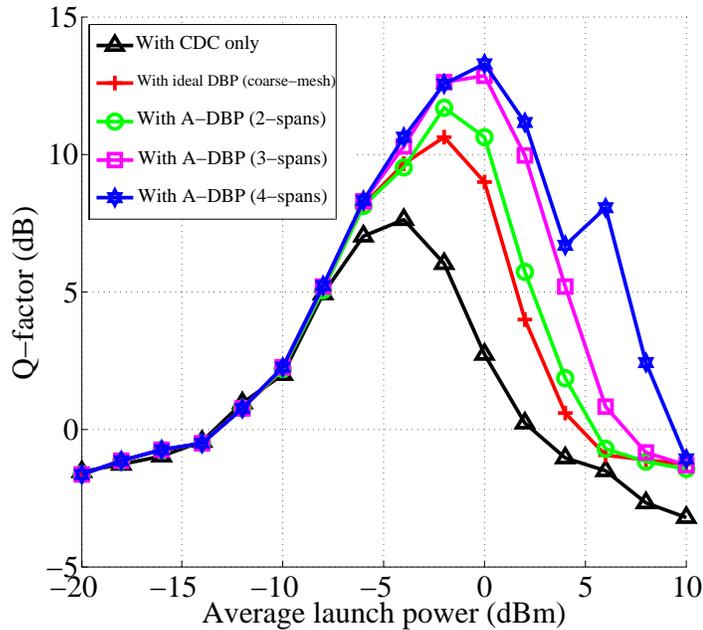
where $L_{tot} = 340$ km, and $M = 2, 3$, or 4. The average launch power at the training process is $P_{av} = 6$ dBm. The initial length and nonlinear coefficient of each span are set to $0.6 \text{ W}^{-1}\text{km}^{-1}$, and L_{tot}/M , respectively. The number of symbols used to train the A-DBP is 4096. The optimal design points of the 2–span, 3–span, and 4–span trained A-DBP models, obtained using the ABO algorithm, are given in Table 4.6. Notice that in this example we do not wish to converge to the actual channel parameters, since the coarse-mesh ideal DBP model (whose parameters are identical to actual channel parameters) does not provide the global optimum solution

anymore. We therefore relax the constraints of the training problem, owing to the fact that the A-DBP is a virtual fiber, which allows for non physical/practical values of the parameters. This relaxation extends the parameter space (i.e. the feasible region) at which the ABO algorithm searches for the optimal coarse-mesh A-DBP parameters, thus seeking to achieve the best possible compensation performance.

Fig. 4.13a shows the BER performance of the 4–span fiber-optic communication system as a function of fiber launch power for various cases of DBP and CDC only. The quality factors and the total number of SSFS steps required by each equalizer, versus a sweep of the launch power, are also compared in Figs. 4.13b and 4.14, respectively. It is worth emphasizing that the compensation performance of the 4-span A-DBP shows a peak at $P_{av} = 6$ dBm, since this is the average launch power at which the A-DBP is trained. As can be seen in Fig. 4.13b, all M –span trained A-DBP, with $M = 2, 3$, and 4, outperform the coarse-mesh ideal DBP. The 2–span A-DBP increases the quality factor by 1 dB with a lower computational cost; 2–SSFS steps as opposed to 4–SSFS steps required by the coarse-mesh ideal DBP at $P_{av} = -2$ dBm, for the 2–span A-DBP. A Q-factor increments of 2.2 dB and 2.7 dB are achieved respectively by the 3–span and 4–span A-DBPs. The 3–span and 4–span A-DBPs though require 2–extra SSFS step as compared to the coarse-mesh ideal DBP. Hence, the 2–span A-DBP provides a better trade-off between performance and computational cost.



(a)



(b)

Figure 4.13: (a) BER, and (b) Quality factor versus average launch power for the 4-span fiber-optic communication system scenario. The equalization at the receiver is performed using the CDC unit only, the coarse-mesh ideal DBP, the coarse-mesh 2-virtual fiber spans A-DBP, the coarse-mesh 3-virtual fiber spans A-DBP, or the coarse-mesh 4-virtual fiber spans A-DBP. The number of transmitted symbols is 2^{17} .

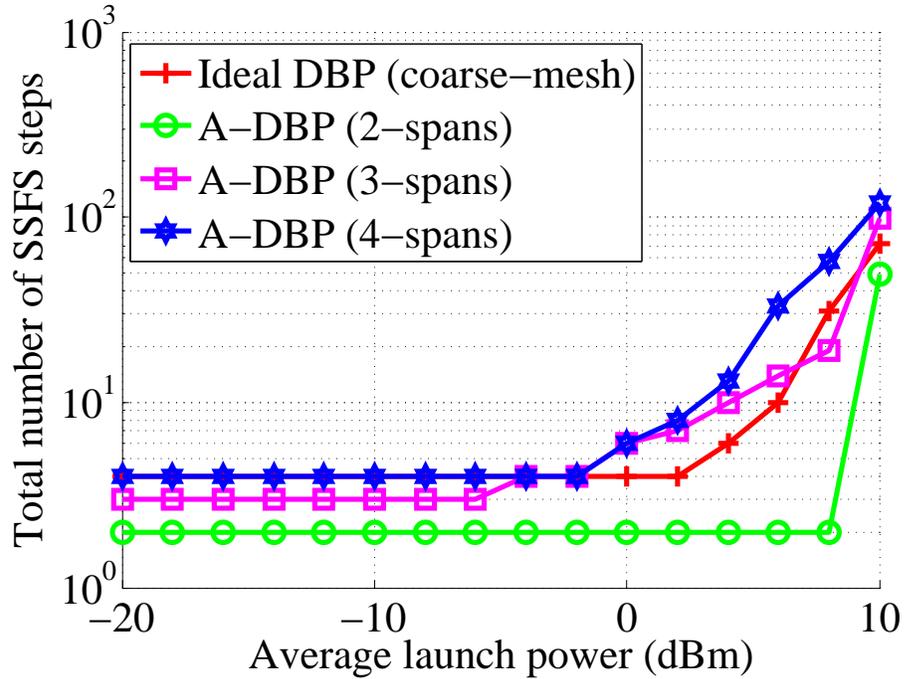


Figure 4.14: Total number of required SSFS steps versus average launch power for the 4-span fiber-optic communication system scenario. The equalization at the receiver is performed using the CDC unit only, the coarse-mesh ideal DBP, the coarse-mesh 2-virtual fiber spans A-DBP, the coarse-mesh 3-virtual fiber spans A-DBP, or the coarse-mesh 4-virtual fiber spans A-DBP. The number of transmitted symbols is 2^{17} .

4.5 Conclusion

A powerful and computationally efficient adaptive DBP method has been proposed to blindly mitigate the linear and nonlinear impairments-induced in long-reach optical fiber communication systems and networks. The proposed approach uses an adjoint-based optimization algorithm for which an ASA algorithm is utilized for sensitivity analysis. Regardless of the number of design parameters, the ABO algorithm obtains all the required sensitivity calculations per optimization iterate using only one extra adjoint system simulation. This is contrasted with conventional FD-based

optimizers whose required sensitivity calculations per iterate scales linearly with the number of parameters. To demonstrate the efficiency and reliability of the proposed A-DBP method, it was employed to blindly compensate for the distortions of a 4-span fiber-optic communication system scenario. Regardless of the initial parameters, the A-DBP showed good ability to converge to the optimal compensation performance achieved using the ideal DBP with fine-mesh and known channel parameters. It has also been shown that a coarse-mesh A-DBP with lower number spans could be used to significantly reduce the computational complexity, achieving compensation performance higher than that obtained using the coarse-mesh DBP with complete number of spans.

Chapter 5

Conclusions and Future Work

Optical fibers have revolutionized the telecommunication world. They form a main backbone of modern telecommunication systems, because of their low loss, large bandwidth, and robustness to electromagnetic interference. With the rapid development of fiber-optic communication systems, the demand for higher transmission data rate with longer transmission reach becomes vital and more essential. However, the light-wave propagation through the optical fiber suffers from severe and unavoidable linear and nonlinear distortions, due to the loss, chromatic dispersion, and optical Kerr effect. These impairments are the major barriers to further increase the capacity and feasible transmission reach of optical fiber systems. This thesis therefore focused on the development of reliable, stable and computationally efficient electrical equalization techniques to compensate for the linear and nonlinear distortions induced in optical fiber communication systems and networks.

We first studied the mitigation of short-reach fiber-optic communication systems with direct detection. Nonlinear distortions due to optical Kerr effect are negligible for short-haul optical fiber systems. Only linear distortions that are mainly due to

chromatic dispersion exist in such systems. However, direct detectors are usually used to detect received signals in the case of short-haul transmission systems. Due to the square-law detection of the direct detection system, the linear impairments in optical domain turn into nonlinear in the electrical domain. A nonlinear equalizer is therefore required at the digital signal processing (DSP) unit of the receiver to effectively compensate for the resultant nonlinear distortions. The maximum likelihood sequence estimator (MLSE) is one of the most effective candidates to offer an excellent compensation performance for such system. However, its computational cost is significantly high, which makes it impractical to be implemented in a DSP chip.

To reduce the computational complexity, at the expense of slightly lower compensation performance, Chapter 2 proposed a simple and computationally-efficient nonlinear feed forward equalizer (NFFE) based on artificial neural networks (ANN). The proposed equalizer is used to mitigate the distortions due to chromatic dispersion in metro optical fiber communication systems and data center networks, with direct photo-detection. Our ANN-NFFE equalizer is comprised of one ANN-hidden-layer which leads to superfast signal processing. An extensive training process for the equalizer is introduced which allows to reach the best possible FFEs performance. The proposed ANN-NFFE achieves a transmission system performance comparable to the MLSE with much lower computational cost. Its required cost per symbol is linearly proportional to the ISI span, in contrast to the exponential growth of the MLSE computational cost.

The efficiency of the proposed equalizer was demonstrated in two fiber-optic application areas. The first of which was 10 Gbps data rate communication system widely used in media and access networks. The obtained results show that the introduced

ANN-NFFE increases the chromatic dispersion tolerance and significantly extends the feasible transmission distance. It has been shown that the ANN-NFFE extends the transmission distance up to 200 km, achieving a BER of 10^{-9} with at most 2 dB-OSNR penalty in comparison to the back-to-back transmission case. The second application was 28 Gbps data rate short-reach fiber optic communication system, which is widely deployed in data center networks. Our results show that the ANN-NFFE significantly reduces the OSNR penalty for achieving a BER of 10^{-9} in case of transmission distances ≤ 10 km. Furthermore, it extends the feasible transmission distance beyond 10 km, allowing the reach up to 20 km with reasonable OSNR penalty as compared to the back-to-back performance.

We then studied combating the nonlinear distortions of long-haul fiber-optic communication systems with coherent detection. The digital back propagation (DBP) method is one of the most effective electronic techniques to mitigate the nonlinear fiber impairments of coherent fiber-optic systems. Since lightwave propagation within an optical fiber link is governed by the nonlinear Schrödinger equation (NLSE), the idea of DBP approach is to apply the inverse of the NLSE on the received signal to undo the distortion effects of the fiber. This requires precise information of the transmission channel parameters. However, such information is not available at the receiver end for optical fiber networks, due to the random network configurations and the unavoidable environmental changes. For such cases, an adaptive DBP scheme can be utilized to compensate for the fiber impairments without knowing the launch power and channel parameters. Initially at a training phase, a stream of known data is sent, and a gradient-based optimization algorithm is used to train the DBP, estimating the parameters of the channel. Then, the adaptive DBP can be used to recover

unknown data. In such adaptive DBP method, the required sensitivity information required during the training process are typically obtained using the traditional finite-difference (FD) approach. However, the computational cost of the FD approach scales linearly with the number of parameters. This overhead is prohibitive and prevents the real-time implementations of adaptive DBP methods in practical fiber-optic systems. We therefore developed a general adjoint sensitivity analysis (ASA) technique to significantly accelerate the estimation of sensitivity information required for any fiber-optic system. Then, as an application to the developed ASA approach, it was exploited to propose a computationally efficient adaptive DBP (A-DBP) method.

A novel, powerful, and easy-to-implement ASA algorithm based on the NLSE was first proposed in Chapter 3. Using only one extra adjoint simulation, the gradient of the desired objective function is estimated with respect to all the design parameters of the problem. A modified algorithm to the split-step Fourier scheme (SSFS) was developed for solving the adjoint problem. It has the same accuracy of the SSFS since it uses the same grid points distribution used in the original problem solution. The efficiency of the proposed ASA algorithm was demonstrated through the sensitivity analysis of three numerical fiber-optic examples. Good agreement is observed with the results obtained using the accurate, but computationally expensive, central finite-difference (CFD) approach. The results show that our proposed algorithm significantly accelerates the sensitivity analysis of any optical fiber design problems. For instance, it has been shown that the ASA requires only one extra simulation while the CFD approach requires 14 extra simulations, in case of design problems with 7 design parameters, i.e., the ASA could save extra 13 system simulations required by the CFD. Also, the ASA algorithm was 5.4 times faster than the CFD approach to

estimate the sensitivities in the mentioned example. This computational time saving would also increase as the number of design parameters increases.

We then proposed a computationally efficient A-DBP method exploiting our NLSE-based ASA approach in Chapter 4. An efficient ABO technique was introduced for accelerating the parameters extraction of the A-DBP. The ABO algorithm uses the sequential quadratic programming (SQP) optimization technique for solving the A-DBP training problem. It also utilizes a generalized ASA approach, based on the inverse nonlinear Schrödinger equation (INLSE), to rapidly evaluate the required sensitivity information at each optimization iterate. This ASA approach estimates the full gradient information of the desired objective function with respect to all A-DBP design parameters, using only one extra adjoint DBP simulation. Full details of the modified (SSFS) method required for solving the adjoint DBP simulation was presented. A computational complexity analysis for the modified SSFS algorithm was also performed. It has been shown that the computational complexity of the adjoint DBP simulation requires an overhead of $32N_s$ —real multiplications per step as compared to the computational cost of the conventional/original DBP simulation, where N_s is the number of processing samples. However, the overhead of the ASA approach is constant and does not depend on the number of design parameters. This is contrasted with the classical FD approaches whose computational complexity scales linearly with the design parameters number.

The efficiency and robustness of the proposed ABO algorithm have been demonstrated through a typical 4—span optical fiber communication system example. Provided that the total transmission distance is known, the ABO algorithm shows an excellent ability to predict the fiber channel parameters, and effectively trains the

A-DBP to converge to the ideal DBP solution, starting from random initial parameters. As compared to the A-DBP training using FD-based SQP algorithms, our ABO algorithm is 2.02 times faster than the backward/forward FD-based optimizers, and 3.63 times faster than the more accurate CFD-based optimizer. Having an accuracy as high as the CFD-based optimizer, the computational complexity gain of our ABO algorithm also increases linearly with the number of parameters. Moreover, we investigated the reduction of the A-DBP computational cost at the expense of slightly lower equalization performance. The proposed ABO algorithm has been applied to train a 2-, 3-, and 4-span A-DBP systems with coarse-meshing models. The obtained results show that an optimized 2-span coarse-mesh A-DBP model provides the best trade-off between equalization performance and computational cost. It provides a 1-dB quality factor increment as compared to the coarse-mesh ideal DBP equalizer, with half the computational cost.

Last but not least, we discuss some suggestions for future work in the next section.

5.1 Future work

This thesis provided theoretical analysis and presented simulation results of electrical compensation techniques for mitigating the optical fiber induced linear and nonlinear distortions. Experimental research is one of the most important future works to measure and validate the theoretical results obtained.

Other suggestions of future work directions are as follows:

- I. Using noisy data to train and adopt the parameters of the proposed artificial neural network nonlinear feed forward equalizer (ANN-NFFE). In Chapter 2, we utilized noise-free data to train the equalizer, which is the typical way for

training. However, optimizing the ANN weights using multiple sets of noisy data would be of more practical sense, and could further improve the compensation performance of the equalizer.

- II. Applying the proposed ANN-NFFE for the compensation of direct-detected short-reach optical fiber systems with advanced modulation formats. Although we investigated the proposed equalizer for the case of on-off keying (OOK) modulation formats, we believe the same algorithm of the proposed ANN-NFFE can be applied to higher modulation formats e.g., pulse amplitude modulation (PAM). The same internal structure of ANN-NFFE could be used. In this case, the equalizer would be applied on symbols rather than bits. The same concept and procedures would still apply. For instance, in case of PAM-4, the output of ANN-NFFE will vary between 4 states instead of 2 states (of OOK).

- III. Extending the proposed ANN-NFFE to compensate for nonlinear distortions of long-haul fiber-optic communication systems with coherent detection. For this case, the coherent receiver will be able to detect the complex optical field rather than the signal power, i.e., both in-phase and quadrature-phase information will be available at the DSP unit. In other words, the conversion of linear distortions into nonlinear impairments will be avoided. However, since nonlinear fiber distortions cannot be neglected for long-haul transmission systems, a nonlinear equalizer is still required to recover the transmitted data. We believe that the approach of the developed ANN-NFFE could be easily extended to compensate for the nonlinear distortions of coherent-detected long-reach optical fiber systems. Instead of using one ANN equalizer, we would need to train 2-ANNs to simultaneously recover the transmitted data. Both the two networks would

have the in-phase and quadrature phase signals as inputs. The 1st network would then be trained to equalize the in-phase signal (i.e. its output would be the in-phase part of the transmitted signal), while the 2nd network would output the equalized quadrature-phase signal part. Notice that this suggested scheme is necessary for efficient compensation performance, in order to effectively compensate for the distortions-induced due to the fiber and due to the coupling between in-phase and quadrature phase data as well. It should also be clear that the computation complexity of the suggested scheme will be the same as the computational cost of one ANN-NNE, since both the 2-ANN equalizers would be independent and would be processing in parallel.

IV. Developing an ASA approach for the nonlinear Schrödinger equation (NLSE) in the Hilbert space. To derive the NLSE-based ASA approach proposed in Chapter 3, we first wrote the complex NLSE in real domain, and all preceding steps were done in the real space. However, the theory could be more elegant if derivations were accomplished in the Hilbert space, i.e., all derivation steps are performed in the complex domain. Moreover, it would be promising to extend the theory and develop the ASA approach for the Manakov system of equations (describing the average polarization evolution of the coupled vector NLSEs), in order to generalize the proposed ASA approach for sensitivity analysis of polarization multiplexed fiber-optic systems.

V. Applying the developed NLSE-based adjoint sensitivity algorithm to accelerate the sensitivity calculations required in other practical optical fiber design problems, e.g., dispersion-managed coherent fiber optic systems. Also, the proposed ASA could also be applied for the sensitivity analysis of any other dynamic

systems governed by the NLSE. Although we focus on fiber-optic problems, the proposed ASA approach is general and applicable for the sensitivity analysis of all other dynamic systems governed by the linear or nonlinear Schrödinger equation, e.g., time-dependent quantum structures.

VI. Developing an adaptive filtered digital back propagation (DBP) scheme based on adjoint optimization. The A-DBP method introduced in Chapter 4 was based on the conventional DBP scheme. In other words, the ASA approach derived includes only the effects of the inverse NLSE and in-line loss elements. If we wish to reduce the number of steps per span, we may extend the theory to the assisted filter A-DBP scheme. In this case, an extended ASA approach needs to be derived to include the effects of the inverse NLSE, the in-line loss elements, and the low pass filter preceding each nonlinear step, in order to estimate the derivatives with respect to the design parameters of fiber, LPF, and in-line loss elements.

Bibliography

([Accessed 18 Nov. 2013]). <http://www.microwavejournal.com/articles/2779-parametrics-and-optimization-using-ansoft-hfss>.

([Accessed 23 Oct. 2013]). <https://www.cst.com/content/news/20105-timedomain2011web.pdf>.

Agazzi, O. E., Hueda, M. R., Carrer, H. S., and Crivelli, D. E. (2005). Maximum-likelihood sequence estimation in dispersive optical channels. *Journal of Lightwave Technology*, **23**(2), 749.

Agrawal, G. P. (2005). *Lightwave technology: telecommunication systems*. John Wiley & Sons.

Agrawal, G. P. (2007). *Nonlinear fiber optics*. Academic press.

Agrawal, G. P. (2012). *Fiber-optic communication systems*, volume 222. John Wiley & Sons.

Ahmed, O. S., Bakr, M. H., Li, X., and Nomura, T. (2012). A time-domain adjoint variable method for materials with dispersive constitutive parameters. *IEEE Transactions on Microwave Theory and Techniques*, **60**(10), 2959–2971.

- Al-Amri, M. D., El-Gomati, M., and Zubairy, M. S. (2016). *Optics in Our Time*. Springer International Publishing.
- Alt, H. and van Leeuwen, J. (1981). The complexity of basic complex operations. *Computing*, **27**(3), 205–215.
- Andrews, J. G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A. C., and Zhang, J. C. (2014). What will 5g be? *IEEE Journal on selected areas in communications*, **32**(6), 1065–1082.
- Ayad, M. A., Obayya, S. S., and Swillam, M. A. (2015). Modelling of quantum confinement in optical nanostructures. *Journal of Optics*, **18**(1), 015201.
- Bakr, M. (2013). *Nonlinear optimization in electrical engineering with applications in Matlab*. The Institution of Engineering and Technology, London.
- Bakr, M., Elsherbeni, A., and Demir, V. (2016). An fdtd-based adjoint sensitivity approach. In *Antennas and Propagation (APSURSI), 2016 IEEE International Symposium on*, pages 2021–2022. IEEE.
- Bakr, M., Demir, V., and Elsherbeni, A. (2017). *Adjoint Sensitivity Analysis of High Frequency Structures with MATLAB*. The Institution of Engineering and Technology (IET).
- Bakr, M. H. and Nikolova, N. K. (2004). An adjoint variable method for time-domain transmission-line modeling with fixed structured grids. *IEEE transactions on microwave theory and techniques*, **52**(2), 554–559.
- Bakr, M. H., Ahmed, O. S., El Sherif, M. H., and Nomura, T. (2014). Time domain

- adjoint sensitivity analysis of electromagnetic problems with nonlinear media. *Optics express*, **22**(9), 10831–10843.
- Basch, E. and Brown, T. (1985). Introduction to coherent optical fiber transmission. *IEEE Communications Magazine*, **23**(5), 23–30.
- Bergano, N. S., Aspell, J., Davidson, C. R., Trischitta, P. R., Nyman, B. M., and Kerfoot, F. W. (1991). Bit error rate measurements of 14000 km 5 gbit/s fibre-amplifier transmission system using circulating loop. *Electronics Letters*, **27**(21), 1889–1890.
- Bidaki, E. and Kumar, S. (2019). A raman-pumped dispersion and nonlinearity compensating fiber for fiber optic communications. *IEEE Photonics Journal*, **12**(1), 1–17.
- Bohn, M. and Xia, C. (2009). Electrical and optical equalization strategies in direct detected high-speed transmission systems. *AEU-International Journal of Electronics and Communications*, **63**(7), 526–532.
- Boyd, R. W. (2019). *Nonlinear optics*. Academic press.
- Bülow, H., Buchali, F., and Klekamp, A. (2008). Electronic dispersion compensation. *Lightwave Technology, Journal of*, **26**(1), 158–167.
- Burse, K., Yadav, R., and Shrivastava, S. (2010). Channel equalization using neural networks: a review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, **40**(3), 352–357.
- Byron, K. (1986). Simultaneous amplification and pulse compression in a single-mode optical fibre. *Electronics Letters*, **22**(24), 1275–1277.

- Chen, X., Antonelli, C., Mecozzi, A., Che, D., and Shieh, W. (2020). High-capacity direct-detection systems. In *Optical Fiber Telecommunications VII*, pages 419–441. Elsevier.
- Chung, Y.-S., Cheon, C., Park, I.-H., and Hahn, S.-Y. (2000). Optimal shape design of microwave device using fdtd and design sensitivity analysis. *IEEE Transactions on Microwave Theory and Techniques*, **48**(12), 2289–2296.
- Chung, Y.-S., Cheon, C., Park, I.-H., and Hahn, S.-Y. (2001). Optimal design method for microwave device using time domain method and design sensitivity analysis. ii. fdtd case. *IEEE transactions on magnetics*, **37**(5), 3255–3259.
- Cioffi, J. and Kailath, T. (1984). Fast, recursive-least-squares transversal filters for adaptive filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **32**(2), 304–337.
- Cohen, L., Lin, C., and French, W. (1979). Tailoring zero chromatic dispersion into the 1.5–1.6 μm low-loss spectral region of single-mode fibres. *Electronics Letters*, **15**(12), 334–335.
- Curri, V., Gaudino, R., Napoli, A., and Poggiolini, P. (2004). Electronic equalization for advanced modulation formats in dispersion-limited systems. *Photonics Technology Letters, IEEE*, **16**(11), 2556–2558.
- da Silva, E. P., Asif, R., Larsen, K. J., and Zibar, D. (2015). Nonlinear compensation with modified adaptive digital backpropagation in flexigrad networks. In *CLEO: Science and Innovations*, pages SM2M–5. Optical Society of America.

- Deiterding, R. and Poole, S. W. (2016). Robust split-step fourier methods for simulating the propagation of ultra-short pulses in single-and two-mode optical communication fibers. In *Splitting Methods in Communication, Imaging, Science, and Engineering*, pages 603–625. Springer.
- DeLange, O. E. (1970). Wide-band optical communication systems: Part iifrequency-division multiplexing. *Proceedings of the IEEE*, **58**(10), 1683–1690.
- Desurvire, E., Simpson, J. R., and Becker, P. (1987). High-gain erbium-doped traveling-wave fiber amplifier. *Optics letters*, **12**(11), 888–890.
- Dhiman, J., Ahmad, S., Gulia, K., *et al.* (2013). Comparison between adaptive filter algorithms (lms, nlms and rls). *International journal of science, engineering and technology research (IJSETR)*, **2**(5), 1100–1103.
- Du, L. B. and Lowery, A. J. (2010). Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems. *Optics express*, **18**(16), 17075–17088.
- Essiambre, R.-J., Mikkelsen, B., and Raybon, G. (1999). Intra-channel cross-phase modulation and four-wave mixing in high-speed tdm systems. *Electronics Letters*, **35**(18), 1576–1578.
- Essiambre, R.-J., Kramer, G., Winzer, P. J., Foschini, G. J., and Goebel, B. (2010). Capacity limits of optical fiber networks. *Journal of Lightwave Technology*, **28**(4), 662–701.
- Foggi, T., Forestieri, E., Colavolpe, G., and Prati, G. (2006). Maximum-likelihood

- sequence detection with closed-form metrics in optical systems impaired by gvd and pmd. *Journal of lightwave technology*, **24**(8), 3073.
- Franceschini, M., Bongiorni, G., Ferrari, G., Raheli, R., Meli, F., and Castoldi, A. (2007). Fundamental limits of electronic signal processing in direct-detection optical communications. *Journal of lightwave technology*, **25**(7), 1742–1753.
- Galdino, L., Semrau, D., Lavery, D., Saavedra, G., Czegledi, C. B., Agrell, E., Killey, R. I., and Bayvel, P. (2017). On the limits of digital back-propagation in the presence of transceiver noise. *Optics express*, **25**(4), 4564–4578.
- Gao, Y., Ke, J. H., Zhong, K. P., Cartledge, J. C., and Yam, S. S.-H. (2012). Assessment of intrachannel nonlinear compensation for 112 gb/s dual-polarization 16qam systems. *Journal of lightwave technology*, **30**(24), 3902–3910.
- Georgieva, N. K., Glavic, S., Bakr, M. H., and Bandler, J. W. (2002). Feasible adjoint sensitivity technique for em design optimization. *IEEE transactions on microwave theory and techniques*, **50**(12), 2751–2758.
- Gnauck, A., Kasper, B., Linke, R., Dawson, R., Koch, T., Bridges, T., Burkhardt, E., Yen, R., Wilt, D., Campbell, J., *et al.* (1985). 4-gbit/s transmission over 103 km of optical fiber using a novel electronic multiplexer/demultiplexer. *Journal of lightwave technology*, **3**(5), 1032–1035.
- Gordon, J. P. and Mollenauer, L. F. (1990). Phase noise in photonic communications systems using linear amplifiers. *Optics letters*, **15**(23), 1351–1353.
- Green, P. E. (1996). Optical networking update. *IEEE Journal on selected areas in communications*, **14**(5), 764–779.

- Hagan, M. T. and Menhaj, M. B. (1994). Training feedforward networks with the marquardt algorithm. *IEEE transactions on Neural Networks*, **5**(6), 989–993.
- Han, S.-P. (1976). Superlinearly convergent variable metric algorithms for general nonlinear programming problems. *Mathematical Programming*, **11**(1), 263–282.
- Hasegawa, A. and Kodama, Y. (1981). Signal transmission by optical solitons in monomode fiber. *Proceedings of the IEEE*, **69**(9), 1145–1150.
- Hasegawa, A. and Tappert, F. (1973). Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. i. anomalous dispersion. *Applied Physics Letters*, **23**(3), 142–144.
- Hayashi, I., Panish, M., Foy, P., and Sumski, S. (1970). Junction lasers which operate continuously at room temperature. *Applied Physics Letters*, **17**(3), 109–111.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, **2**(5), 359–366.
- Humblet, P. A. and Azizoglu, M. (1991). On the bit error rate of lightwave systems with optical amplifiers. *Journal of lightwave technology*, **9**(11), 1576–1582.
- Igarashi, H. and Watanabe, K. (2010). Complex adjoint variable method for finite-element analysis of eddy current problems. *IEEE Transactions on Magnetics*, **46**(8), 2739–2742.
- Ip, E. and Kahn, J. M. (2008). Compensation of dispersion and nonlinear impairments using digital backpropagation. *Journal of Lightwave Technology*, **26**(20), 3416–3425.

- Irukulapati, N. V., Wymeersch, H., Johannisson, P., and Agrell, E. (2014). Stochastic digital backpropagation. *IEEE Transactions on Communications*, **62**(11), 3956–3968.
- Irukulapati, N. V., Marsella, D., Johannisson, P., Agrell, E., Secondini, M., and Wymeersch, H. (2016). Stochastic digital backpropagation with residual memory compensation. *Journal of Lightwave Technology*, **34**(2), 566–572.
- Jiang, L., Yan, L., Chen, Z., Yi, A., Pan, Y., Pan, W., and Luo, B. (2016). Low-complexity and adaptive nonlinearity estimation module based on godard’s error. *IEEE Photonics Journal*, **8**(1), 1–8.
- Kalander, A. and WANG, R. (2017). *Time-Domain Digital Back Propagation for Optical Communication in 28 nm FD-SOI*. Master’s thesis.
- Kaminow, I., Li, T., and Willner, A. E. (2013). *Optical fiber telecommunications volume VIB: systems and networks*. Academic Press.
- Kao, K. and Hockham, G. A. (1966). Dielectric-fibre surface waveguides for optical frequencies. In *Proceedings of the Institution of Electrical Engineers*, volume 113, pages 1151–1158. IET.
- Keck, D., Maurer, R., and Schultz, P. (1973). On the ultimate lower limit of attenuation in glass optical waveguides. *Applied Physics Letters*, **22**(7), 307–309.
- Kerr, J. (1875). XI. a new relation between electricity and light: Dielectrified media birefringent. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **50**(332), 337–348.

- Krehlik, P. (2007). Directly modulated lasers in negative dispersion fiber links. *Opto-Electronics Review*, **15**(2), 71–77.
- Kumar, S. (2005). Effect of dispersion on nonlinear phase noise in optical transmission systems. *Optics letters*, **30**(24), 3278–3280.
- Kumar, S. and Deen, M. J. (2014). *Fiber Optic Communications: Fundamentals and Applications*. John Wiley & Sons.
- Kumar, S. *et al.* (2018). Nonlinear impairments in fiber optic communication systems: Analytical review. In *International Conference on Futuristic Trends in Network and Communication Technologies*, pages 28–44. Springer.
- Kumar, S. and Shao, J. (2013). Optical back propagation with optimal step size for fiber optic transmission systems. *IEEE Photonics Technology Letters*, **25**(5), 523–526.
- Kumar, S. and Yang, D. (2011). Optical backpropagation for fiber-optic communications using highly nonlinear fibers. *Optics letters*, **36**(7), 1038–1040.
- Lam, C. F., Liu, H., and Urata, R. (2014). What devices do data centers need? In *Optical Fiber Communication Conference*, pages M2K–5. Optical Society of America.
- Li, X., Chen, X., Goldfarb, G., Mateo, E., Kim, I., Yaman, F., and Li, G. (2008). Electronic post-compensation of wdm transmission impairments using coherent detection and digital signal processing. *Optics Express*, **16**(2), 880–888.
- Liang, X. (2015). *Analysis and Compensation of Nonlinear Impairments in Fiber-Optic Communication Systems*. Ph.D. thesis.

- Liang, X. and Kumar, S. (2014). Multi-stage perturbation theory for compensating intra-channel nonlinear impairments in fiber-optic links. *Optics express*, **22**(24), 29733–29745.
- Lin, C.-Y., Napoli, A., Spinnler, B., Sleiffer, V., Rafique, D., Kuschnerov, M., Bohn, M., and Schmauss, B. (2014). Adaptive digital back-propagation for optical communication systems. In *Optical Fiber Communication Conference*, pages M3C–4. Optical Society of America.
- Liu, L., Li, L., Huang, Y., Cui, K., Xiong, Q., Hauske, F. N., Xie, C., and Cai, Y. (2011). Intrachannel nonlinearity compensation by inverse volterra series transfer function. *Journal of Lightwave Technology*, **30**(3), 310–316.
- Lucky, R. W. (1965). Automatic equalization for digital communication. *Bell System Technical Journal*, **44**(4), 547–588.
- Maghrabi, M. M., Bakr, M. H., Kumar, S., Elsherbeni, A., and Demir, V. (2020). Fdtd-based adjoint sensitivity analysis of high frequency nonlinear structures. *IEEE Transactions on Antennas and Propagation*, **68**(6), 4727–4737.
- Maiman, T. H. (1960). Stimulated optical radiation in ruby. *nature*, **187**(4736), 493–494.
- Malekiha, M. (2011). *Analysis Of Nonlinear Effects And Their Mitigation In Fiber-Optic Communication Systems*. Ph.D. thesis.
- Mateo, E., Zhu, L., and Li, G. (2008). Impact of xpm and fwm on the digital implementation of impairment compensation for wdm transmission using backward propagation. *Optics Express*, **16**(20), 16124–16137.

- Mateo, E. F. and Li, G. (2009). Compensation of interchannel nonlinearities using enhanced coupled equations for digital backward propagation. *Applied optics*, **48**(25), F6–F10.
- Mateo, E. F., Yaman, F., and Li, G. (2010). Efficient compensation of inter-channel nonlinear effects via digital backward propagation in wdm optical transmission. *Optics Express*, **18**(14), 15144–15154.
- Mears, R. J., Reekie, L., Jauncey, I., and Payne, D. N. (1987). Low-noise erbium-doped fibre amplifier operating at 1.54 μm . *Electronics Letters*, **23**(19), 1026–1028.
- Mecozzi, A. (1994). Limits to long-haul coherent transmission set by the kerr nonlinearity and noise of the in-line amplifiers. *Journal of lightwave technology*, **12**(11), 1993–2000.
- Miya, T., Terunuma, Y., Hosaka, T., and Miyashita, T. (1979). Ultimate low-loss single-mode fibre at 1.55 μm . *Electronics Letters*, **15**(4), 106–108.
- Momtaz, A. and Green, M. M. (2010). An 80 mw 40 gb/s 7-tap t/2-spaced feed-forward equalizer in 65 nm cmos. *IEEE Journal of Solid-State Circuits*, **45**(3), 629–639.
- Mukai, T. and Yamamoto, Y. (1983). Optical direct amplification for fiber transmission. *Rev. Electr. Commun. Lab.*, **3**, 340.
- Negm, M. H., Bakr, M. H., Ahmed, O. S., Nikolova, N. K., and Bandler, J. W. (2014). Wideband second-order adjoint sensitivity analysis exploiting tlm. *IEEE Transactions on Microwave Theory and Techniques*, **62**(3), 389–398.

- Nikolova, N. K., Tam, H. W., and Bakr, M. H. (2004). Sensitivity analysis with the fdtd method on structured grids. *IEEE transactions on microwave theory and techniques*, **52**(4), 1207–1216.
- Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.
- Ohashi, M., Shiraki, K., and Tajima, K. (1992). Optical loss property of silica-based single-mode fibers. *Journal of Lightwave Technology*, **10**(5), 539–543.
- Otani, T., Goto, K., Abe, H., Tanaka, M., Yamamoto, H., and Wakabayashi, H. (1995). 5.3 gbit/s 11300 km data transmission using actual submarine cables and repeaters. *Electronics Letters*, **31**(5), 380–381.
- Otsuka, Y., Fukumoto, Y., Owaki, S., and Nakamura, M. (2018). Computational-complexity comparison of artificial neural network and volterra series transfer function for optical nonlinearity compensation. In *2018 23rd Opto-Electronics and Communications Conference (OECC)*, pages 1–2. IEEE.
- Owaki, S. and Nakamura, M. (2018). Xpm compensation in optical fiber transmission systems using neural-network-based digital signal processing. *IEICE Communications Express*, **7**(1), 31–36.
- Palomares, U. G. and Mangasarian, O. L. (1976). Superlinearly convergent quasi-newton algorithms for nonlinearly constrained optimization problems. *Mathematical Programming*, **11**(1), 1–13.

- Park, I.-h., Lee, B.-t., and Hahn, S.-y. (1992). Design sensitivity analysis for non-linear magnetostatic problems using finite element method. *IEEE transactions on magnetics*, **28**(2), 1533–1536.
- Park, I.-H., Kwak, I.-G., Lee, H.-B., Hahn, S.-Y., and Lee, K.-S. (1996). Design sensitivity analysis for transient eddy current problems using finite element discretization and adjoint variable method. *IEEE transactions on magnetics*, **32**(3), 1242–1245.
- Pozar, D. M. (2011). *Microwave engineering*. John wiley & sons.
- Pradhan, S., Arbulich, J., and Srihari, K. (2003). Splice loss optimization of single mode fiber and erbium doped fiber experiments and issues. In *ASME International Mechanical Engineering Congress and Exposition*, volume 37149, pages 479–486.
- Proakis (2007). *Digital Communications 5th Edition*. McGraw Hill.
- Rafique, D. and Ellis, A. D. (2011). 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.
- Rao, S. S. (2019). *Engineering optimization: theory and practice*. John Wiley & Sons.
- Roberts, K., Li, C., Strawczynski, L., O’Sullivan, M., and Hardcastle, I. (2006). Electronic precompensation of optical nonlinearity. *IEEE Photonics Technology Letters*, **18**(2), 403–405.
- Sarkis, C. (2009). Reach enhancement in both direct-detection and coherent detection optical fiber communication systems.

- Sayed, E., Bakr, M. H., Bilgin, B., and Emadi, A. (2018). Adjoint sensitivity analysis of switched reluctance motors. *Electric Power Components and Systems*, **46**(18), 1959–1968.
- Schetzen, M. (2006). The volterra and wiener theories of nonlinear systems, reprint ed.
- Senior, J. M. and Jamro, M. Y. (2009). *Optical fiber communications: principles and practice*. Pearson Education.
- Shao, J. (2015). *Analysis of Signal Propagation in Fiber Optic Links and Compensation Techniques for Propagation Impairments*. Ph.D. thesis.
- Shao, J., Liang, X., and Kumar, S. (2014). Comparison of split-step fourier schemes for simulating fiber optic communication systems. *IEEE Photonics Journal*, **6**(4), 7200515.
- Sharma, D. and Kumar, S. (2017). An overview of elastic optical networks and its enabling technologies. *Int. J. Eng. Technol.(IJET)*, **9**(3), 1643–1649.
- Sharma, P., Pardeshi, S., Arora, R. K., and Singh, M. (2013). A review of the development in the field of fiber optic communication systems. *International Journal of Emerging Technology and Advanced Engineering*, **3**(5), 113–119.
- Shaw, H. J. and Digonnet, M. J. (1987). Fiber optic amplifier. US Patent 4,674,830.
- Sinkin, O. V., Holzlohner, R., Zweck, J., and Menyuk, C. R. (2003). Optimization of the split-step fourier method in modeling optical-fiber communications systems. *Journal of lightwave technology*, **21**(1), 61–68.

- Spinnler, B. (2010). Equalizer design and complexity for digital coherent receivers. *IEEE Journal of Selected Topics in Quantum Electronics*, **16**(5), 1180–1192.
- Sun, Y., Shubochkin, R., and Zhu, B. (2015). New development in optical fibers for data center applications. In *SPIE OPTO*, pages 938705–938705. International Society for Optics and Photonics.
- Swillam, M. A., Bakr, M. H., Nikolova, N. K., and Li, X. (2007). Adjoint sensitivity analysis of dielectric discontinuities using fdtd. *Electromagnetics*, **27**(2-3), 123–140.
- Swillam, M. A., Bakr, M. H., Li, X., and Deen, M. J. (2008). Efficient sensitivity analysis of the time independent schrödinger equation with application to quantum lasers. *Optics communications*, **281**(17), 4459–4463.
- Tanimura, T., Hoshida, T., Tanaka, T., Li, L., Oda, S., Nakashima, H., Tao, Z., and Rasmussen, J. C. (2010). Semi-blind nonlinear equalization in coherent multi-span transmission system with inhomogeneous span parameters. In *Optical Fiber Communication Conference*, page OMR6. Optical Society of America.
- Tkach, R., Chraplyvy, A., Forghieri, F., Gnauck, A., and Derosier, R. (1995). Four-photon mixing and high-speed wdm systems. *Journal of Lightwave Technology*, **13**(5), 841–849.
- Tsujikawa, K., Tajima, K., Shiraki, K., and Sankawa, I. (2007). Method for predicting rayleigh scattering loss of silica-based optical fibers. *Journal of lightwave technology*, **25**(8), 2122–2128.
- Wang, J. and Kahn, J. M. (2004). Performance of electrical equalizers in optically

- amplified ook and dpsk systems. *Photonics Technology Letters, IEEE*, **16**(5), 1397–1399.
- Watts, P. M. (2007). *Chromatic dispersion compensation using electronic signal processing in high speed optical communication*. University of London, University College London (United Kingdom).
- Webb, J. (2001). Design sensitivities using high-order tetrahedral vector elements. *IEEE transactions on magnetics*, **37**(5), 3600–3603.
- Webb, J. P. (2002). Design sensitivity of frequency response in 3-d finite-element analysis of microwave devices. *IEEE transactions on magnetics*, **38**(2), 1109–1112.
- Weber, R. J. (2001). *Introduction to microwave circuits: Radio frequency and design applications*. IEEE.
- Wei, J., Ingham, J., Cunningham, D., Penty, R., and White, I. (2012). Performance and power dissipation comparisons between 28 gb/s nrz, pam, cap and optical ofdm systems for data communication applications. *Lightwave Technology, Journal of*, **30**(20), 3273–3280.
- Weiss, A. J. (2003). On the performance of electrical equalization in optical fiber transmission systems. *IEEE Photonics Technology Letters*, **15**(9), 1225–1227.
- Widrow, B. and Kamenetsky, M. (2003). On the efficiency of adaptive algorithms.
- Willner, A. (2019). *Optical fiber telecommunications*, volume 11. Academic Press.
- Wyatt, R. and Devlin, W. (1983). 10 khz linewidth 1.5 μm ingaasp external cavity laser with 55 nm tuning range. *Electronics Letters*, **19**(3), 110–112.

- Xia, C. and Rosenkranz, W. (2007). Nonlinear electrical equalization for different modulation formats with optical filtering. *Journal of Lightwave Technology*, **25**(4), 996–1001.
- Yamada, J., Machida, S., and Kimura, T. (1981). 2 gbit/s optical transmission experiments at 1.3 μm with 44 km single-mode fibre. *Electronics Letters*, **17**(13), 479–480.
- Yamamoto, Y. and Kimura, T. (1981). Coherent optical fiber transmission systems. *IEEE Journal of Quantum Electronics*, **17**(6), 919–935.
- Yang, D. (2010). *Analysis and Design of Long Haul Fiber-Optic Communication Systems*. Ph.D. thesis.
- Yang, D. and Kumar, S. (2009). Intra-channel four-wave mixing impairments in dispersion-managed coherent fiber-optic systems based on binary phase-shift keying. *Journal of Lightwave Technology*, **27**(14), 2916–2923.
- Zayani, R. and Bouallegue, R. (2007). Pre-distortion for the compensation of hpa nonlinearity with neural networks: Application to satellite communications. *Int. J. Comput. Sci. Netw. Secur*, **7**(3), 97–103.
- Zhang, F., Zhuge, Q., Qiu, M., Wang, W., Chagnon, M., and Plant, D. V. (2015). Xpm model-based digital backpropagation for subcarrier-multiplexing systems. *Journal of Lightwave Technology*, **33**(24), 5140–5150.
- Zhang, F., Zhuge, Q., Qiu, M., Chagnon, M., and Plant, D. V. (2016a). Blind adaptive xpm model based digital backpropagation for subcarrier-multiplexing systems. In

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

Zhang, F., Zhuge, Q., Qiu, M., Zhou, X., Sowailem, M. Y., Hoang, T. M., Xiang, M., and Plant, D. V. (2017). Blind adaptive digital backpropagation for fiber nonlinearity compensation. *Journal of Lightwave Technology*, **36**(9), 1746–1756.

Zhang, Y., Ahmed, O. S., and Bakr, M. H. (2014). Wideband fdtd-based adjoint sensitivity analysis of dispersive electromagnetic structures. *IEEE Transactions on Microwave Theory and Techniques*, **62**(5), 1122–1134.

Zhang, Y., Negm, M. H., and Bakr, M. H. (2016b). An adjoint variable method for wideband second-order sensitivity analysis through fdtd. *IEEE Transactions on Antennas and Propagation*, **64**(2), 675–686.

Zhou, J., Wang, Y., and Zhang, Y. (2020). Blind back-propagation method for fiber nonlinearity compensation with low computational complexity and high performance. *Optics Express*, **28**(8), 11424–11438.