Republic of Iraq
Ministry of Higher Education
and Scientific Research
University of Babylon
College of Engineering
Department of Electrical Engineering

Enhancement the Nonlinear Optical Impairment of the Fiber Transmission System Based on Deep Neural Network

A Dissertation

Submitted to the College of Engineering of the University of Babylon in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Engineering \ Electrical Engineering \ Electronic and Communications

By

Alaa Hammood Abed Jarah

Supervised by

Prof. Dr. Ibrahim A. Murdas

2022 A.D

بِسْمِ اللهِ الرَّحْمٰنِ الرَّحِيْمِ

﴿لِّئَلَّا يَعْلَمَ أَهْلُ الْكِتَابِ أَلَّا يَقْدِرُونَ عَلَىٰ شَيْءٍ مِّن فَضْلِ اللَّهِ ۙ وَأَنَّ الْفَضْلَ بِيَدِ اللَّهِ يُؤْتِيهِ مَن يَشَاءُ ۚ وَاللَّهُ ذُو الْفَضْلِ الْعَظِيمِ﴾

صدق اللهُّ العظيم

الآية 29 من سورة الحديد

Dedication

To my family
(My wife, Misk, Mohammed & Leen)

with love and appreciation

Supervisor Certification

We certify that this thesis entitled **"Enhancement the Nonlinear Optical Impairment of the Fiber Transmission System Based on Deep Neural Network"** was prepared by **(Alaa Hammood Abed Jarah AL-Khafaji)** under our supervision at the Department of Electrical Engineering, College of Engineering, University of Babylon, as a partial fulfillment of the requirement for the degree of Doctor of Philosophy in Electrical Engineering \ Electronic and Communication Engineering.

Signature:

Name: **Prof. Dr. Ibrahim A. Murdas**

Date: / / 2022

In view of the above recommendation, I am forward this thesis for discussion by the Examination Committee.

Signature:

Name: ***Asst. Prof. Dr. Shamam Fadhil Alwash***

(Head of Electrical Engineering Dept.)

Date: / / 2022

Examining Committee Certificate

We certify that we have read this thesis entitled **"Enhancement the Nonlinear Optical Impairment of the Fiber Transmission System Based on Deep Neural Network"**, and as an examining committee, examined the student **(Alaa Hammood Abed Jarah AL-Khafaji)**, in its contents and that in our opinion it meets the standard of a thesis for the degree of Doctor of Philosophy in Electrical Engineering\Electronic and Communication Engineering.

Signature:
Name: ***Prof. Dr. Ehab A. Hussein***
(Member)
Date: / / 2022

Signature:
Name: ***Prof. Dr. Thamir R. Saeed***
(Member)
Date: / / 2022

Signature:
Name: ***Prof. Dr.*** **Kasim Karam Abdalla**
(Member)
Date: / / 2022

Signature:
Name: ***Prof. Dr. Qais Kareem Omran***
(Member)
Date: / / 2022

Signature:
Name: ***Prof. Dr. Hadi T. Ziboon***
(Chairman)
Date: / / 2022

Signature:
Name: ***Prof. Dr. Ibrahim A. Murdas***
(Supervisor)
Date: / / 2022

Signature:
Name: ***Asst. Prof. Dr. Shamam Fadhil Alwash***
(Head of Electrical Engineering Dept.)
Date: / / 2022

Signature:
Name: ***Prof. Dr. Hatem Hadi Obeid***
(Dean of College of Engineering)
Date: / / 2022

Acknowledgment

I wish to thanks many people that support me during my Ph.D journey. I wish to express my deep gratitude and appreciation to my supervisor ***Dr. Ibrahim A. Murdas*** *for his valuable advice, encouragement and patience throughout the preparing of this dissertation.*

I would like to thank the faculty members of the Electrical Engineering Department at Babylon University for their assistance and encouragement.

Also, I want to thanks ***my family*** *for their patience and support me during the years of my study.*

Finally, a special thanks to ***my wife*** *for her support and patience through good and bad time in my Ph.D.*

Alaa Hammood

Abstract

The global request for high data traffic has grown dramatically over the past few years and the future demand will outpace the capacity of the current backbone of the Internet, where a fiber-optic systems are the central infrastructure. So, performance improvement is essential for these systems and makes them more flexible and scalable to meet this growing demand require better models of fiber optic channels, complicated optimization methods, and consideration of totally new approaches to transmission, which leads to increase cost and complexity. Therefore, Independent and efficient algorithms can be considered a good solution at present to help explore possible solutions for increasing data transfer rates and resource allocation.

The traditional algorithms relied on the Split-Step Fourier method (SSFM) and digital backpropagation (DBP) methods to find the approximate solution to the Non-Linear Schrödinger equation (NLSE), to process the most intrinsic challenge facing such systems which is the nonlinear impairments, but its implementation requires too many resources for signal processing as well as an accurate and high-level understanding of these systems. While Artificial Intelligence (AI) algorithms promise to provide the best solutions to the described future challenges by being utilized to determine and resolve the deficiencies in very complicated systems and complex environments without required to develop a hardware (physical) model.

This thesis applied machine learning based on deep neural network (DNN) as a supervised learning method in novel contexts for coherent optical communication systems to increase their capabilities. It is proposed to build an Nonlinear Compensation (NLC) technique to compensate for the non-linear impairments that occur in optical communication systems. The system is designed and fully simulated in real-time by using MATLAB App. (V.2020a)

with Optisystem App. (Ver.17.0.0) and necessary testing/optimization has been performed in three directions:

Firstly, on single-polarization, coherent optical systems with 4-16 QAM modulation shapes with and without multi subcarrier OFDM, transmitted over single-carrier Standard Single-Mode Fiber (SSMF) with 60Gbps for 3000Km link distance.

The results found without using OFDM of the proposed DNNs-NLC model outperforms in terms of Q-factor, BER, higher transmission distance with Q-factor equal to (~13.4 dB) at testing for 4QAM coherent optical BtB 60 Gb/s @ (100 x 30span) Km measurements.

With OFDM, the proposed DNNs-NLC model outperforms in terms of Q-factor, higher transmission distance compared with ANN, LE, and IVSTF-NLE by 0, 3, and 1 dB, respectively

Secondly, on single-polarization, coherent optical systems with 16QAM modulation shapes with multi subcarrier OFDM, multi-carrier wavelength division multiplexing (8 channel WDM) with 40 GHz frequency spacing, transmitted over 3000Km SSMF link distance with totally 480Gbps bitrate. The DNNs-NLC proposed model ability was evaluated based on BER, Q-factor and correct mapper decoder. The final result obtained of the Q-factor equal to (~9 dB) at testing and outperforms compared with ANN and VSTF-NLE by <1, and 2 dB respectively

Finally, on dual-polarization, coherent optical systems with 16QAM modulation shapes, multi-carrier wavelength division multiplexing (8 channel WDM) with 37.5 GHz (3nm) frequency spacing, transmitted over 3000Km SSMF link distance with totally 960Gbps bitrate.

The achieved results demonstrated the efficiency of the proposed design and the significant improvement in performance, mainly in data rate and correlation range compared with the similar works.

Table of Contents

List of Abbreviations

Abbreviation	Definition
ADC	Analogue to Digital Convertor
AE	Autoencoder
AI	Artificial Intelligence
APD	Avalanche Photo-Diode
ASE	Amplified Spontaneous Emission
ASK	Amplitude Shift Keying
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CD	Coherent Detection
CD-OOFDM	Coherent Detection Optical Orthogonal Frequency Division Multiplexing
COC	Coherent Optical Communication
CS	Constellation Shaping
CW	Continuous Waveform
DAC	Digital to Analogue Convertor
DBP	Digital Back-Propagation
DCF	Dispersion Compensation Fiber
DD-OOFDM	Direct Detection Optical Orthogonal Frequency Division Multiplexing
DEMUX	Demultiplexer
DGD	Differential Group Delay
DMT	Discrete Multitone
DNN	Deep Neural Network
DP	Dual Polarization
DPSK	Differential Phase Shift Keying
DQPSK	Differential Quadrature Phase Shift Keying

Abbreviation	Definition
DSP	Digital Signal Processing
DVB	Digital Video Broadcasting
E/O	Electrical to Optical Conversion
EDFA	Erbium-Doped Fiber Amplifier
EVM	Error Vector Magnitude
FBG	Fiber Bragg Grating
FDM	Frequency Division Multiplexing
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FKVNEs	Full Kernel Volterra Non-Linear Equalizers
FSK	Frequency Shift Keying
FSO	Free Space Optical
FTTH	Fiber-To-The-Home
FWHM	Full Width at Half Maximum
FWM	Four-Wave Mixing
GN	Extended Gaussian Noise
GVD	Group-Velocity Dispersion
HD-FEC	Hard Decision- Forward Error Correction
I	In-Phase
ICI	Inter-Carrier-Interference
IF	Intermediate Frequency
IFFT	Inverse Fast Fourier Transform
IM	Intensity Modulation
IM/DD	Intensity Modulation/Direct Detection
IoT	Internet Of Things
IPM	Iterative Polar Modulation
ISI	Intersymbol Interference

Abbreviation	Definition
ITU-T	International Telecommunication Union
LD	Laser Diode
LED	Light Emitting Diode
LMS	Least Mean Squares
LO	Local Oscillator
LOS	Line Of Sight
LPF	Low-Pass Filter
LReLU	Leaky Rectified Linear Units
LTE	Long Term Evolution
LTI	Linear Time-Invariant
MCM	Multi-Carrier Modulation
MIMO	Multiple-Input Multiple-Output
ML	Machine Learning
MLSE	Maximum-Likelihood Sequence Equalizer
MP	Memory Polynomial
MSE	Mean Squared Error
MUX	Multiplexer
MZM	Mach-Zehnder Modulator
NLC	Nonlinear Compensation
NLI	Non-Linear Interference
NLSE	Non-Linear Schrodinger Equation
NN	Neural Network
NRZ	Non-Return to Zero
O/E	Optical To Electrical Conversion
OC	Optical Carrier
OCDMA	Optical Code-Division Multiple-Access System
OFDM	Orthogonal Frequency Division Multiplexing

Abbreviation	Definition
OOK	On-Off Keying
OPC	Optical Phase Conjugation
OSNR	Optical Signal-To-Noise Ratio
OTR	Optical-To-RF Downconverter
P/S	Parallel To Serial Converter
PAPR	Peak To Average Power Ratio
PC	Phase Conjugation
PCA	Principal Component Analysis
PDM	Polarization-Division Multiplexing
PIN	P-N Diode
PRBS	Pseudorandom Binary Sequence
PSD	Power Spectral Density
PSK	Phase Shift Keying
Q	Quadrature
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
ReLU	Rectified Linear Unit
RF	Radio Frequency
RTO	RF-To-Optical Upconverter
RZ	Return To Zero
S/P	Serial To Parallel Converter
SAC	Spectral Amplitude Coding
SD- FEC	Soft Decision- Forward Error Correction
SDM	Space Division Multiplexing
SPECTS	Sequence Spectral Phase-Encoded Time Spreading
SSFM	Split-Step Fourier Method
SVM	Support Vector Machine

Abbreviation	Definition
TEQ	Turbo Equalization
VLSI	Very Large-Scale Integration
VNI	Visual Networking Index
VR-AR	Virtual Reality- Virtual Augmented
VSTF	Volterra Series Transfer Function
WDM	Wavelength-division multiplexing
WLAN	Wireless Local Area Networks
WSS	Wavelength-Selective Switch

List of Symbols

Symbol	Definition
A_{eff}	Fiber's Effective Core Area
b	Max Epochs Size of NN
B	Bitrate
B.w	Reference Bandwidth
B_e	Band-Width of The Electrical Signal,
B_m	Appearance of Modal Birefringence
c	Light Speed
C	Represents Shannon Capacity
C_r	Chirp Factor
D(λ)	Chromatic Dispersion Factor at The Operating Wavelength
E_s	Symbol Energy
G	Gain
h	Planck Constant
K	Represent The Number of The Classes
L	Fiber Length
L	Loss of The Cross Entropy
L_{DCF}	Length of the DCF
NF	Noise Figure
n_{spe}	Factor of The Spontaneous-Emission
p	Number of The Polarizations
Pn	Power Of Noise
Ps	Signal Power
P_{sig}	Average Signal Power
R_S	Symbol Rate
S_0	Zero Dispersion Slope
T	NN Target
T_{FWHM}	Gaussian Full Width at Half Maximum (FWHM) Pulse Width
T_{in}	RMS Pulse Width

Symbol	Definition
w	NN Weight
X	NN Input
y	NN Output
α	Fiber Loss
β	Phase Constant
β_2	2nd-Order Chromatic Dispersion Effects
β_{2x}	X-Polarization, 2nd-Order Chromatic Dispersion Effects
β_{2y}	Y-Polarization, 2nd-Order Chromatic Dispersion Effects
γ	Non-Linear Coefficient
ΔT,	Pulse Broadening
Δt_{max}	Maximum Allowable Pulse Spread
$\Delta\lambda$	RMS Spectral Width
θ_0	Laser Beam Divergence Angle
λ	Carrier Wavelength
λ_c	Center Wavelength
σ	Sigmoid Activation Function
Φ	Angle Between the Perpendicular to The Receiver Plane and The Transmitter Receiver Trajectory
A_{Lf}	Fiber Loss for L
$I(t)$	Modulates Real Part
L_a	Additional Loss
L_c	Coupler Loss, And
M	Power Margin
N_{ASE}	Power Spectral Density of ASE Noise
P	Average Power
P_{ASE}	ASE Noise Power
P_{ch}	WDM Channel Power
P_t	Transmitter Power
$Q(t)$	Modulates The Imaginary Part
S_r	Receiver Sensitivity

Symbol	Definition
T_s	Symbol Duration
n	Fiber Refractive Index
n_2	Fiber Nonlinear Kerr Coefficient
n_{sp}	Amplifier Spontaneous Emission Factor
σ^2_{ASE}	ASE Variance
σ^2_{NLI}	Nonlinear Variance
$\Delta\omega$	Pulse Spectral Width
Δf	Subcarriers Spacing
λ_0	Central Wavelength
μ	Spectral Efficiency
ν	Maxwell-Boltzman Distribution Free Parameter
v_p	Phase Velocity
σ_0,σ_1	The Variance of Current Noise in The Detector
σ_{DC}	The Variance of Dark Current
τ_{DGD}	Differential Group Delay
ϕ_{LO}	Phase Shift of Local Oscillator Signal
ϕ_R	Modulated Phase of Received Signal
ϕ_t	Angle Of Transmission
$\emptyset_1$,$\emptyset_2$	Phase In Upper and Lower Arm Of MZM

List of Figures

List of Tables

Chapter 1
Introduction

This thesis studies the implementation of Artificial Intelligent (AI) algorithms based on the Deep Neural Network (DNN) learning method for optical communication systems. The purpose of this research will be explained in-depth in this chapter, as well as the importance of applying Machine Learning (ML) for the future generation related to optical communication systems.

1.1 Overview

The Cisco Visual Networking Index (VNI) stated that the global need for high-speed communications had expanded considerably recently, with data beginning to dominate traffic [1]. Data traffic is triple between 2014 and 2020, mainly due to developing applications that consume bandwidth such as cloud services, online games, HD video, virtual- augmented reality (VR-AR), multimedia exchange through smartphones, high-quality real-time video transmission, and so on. VNI stated that over a million minutes of multimedia (video) content would be transmitted over the IP network each second in 2020; moreover, information technologies are one of the main pillars of modern society and economy from social networks to electronic commerce, and it became one of the economic resources which are distributed over the Internet through a massive global physical network capable of transmitting information at the speed of light. Fiber-optic communication technology, which carries data across thousands of kilometers, lies at the core of this global infrastructural network known as the backbone. To send data across continents and nations, the world currently significantly depends on the Internet; therefore, the global data traffic across the Internet has

approached the zettabyte threshold according to [2], and these demands will exceed the capacity of the current backbone of the Internet. Therefore, more attention has been given to how information is produced, transferred, and stored, and one of the remarkable achievements of the human race, which is a direct result of advances in science and engineering, is the development of telecommunication systems. From a historical perspective, the expansion of fiber-optic communication started with the invention of the first laser in 1960 [3]. Following that, in 1966 [4], the fiber was initially created, leading to the standard single-mode fiber (SSMF), which has lower losses and longer transmission lengths compared to the preceding guided transmission media. After that, throughout the 1970s and 1980s, the first low-attenuation silica fibers were created [5]. The first optical amplifiers (the late 1980s), which superseded electronic regeneration techniques and enabled all-optical, inexpensive amplification, constituted the backbone of current Wavelength Division Multiplexing (WDM) systems [3, 6-7]. Coherent modulation, which might be implemented on present fiber infrastructure with access to both polarization and quadrature of each optical carrier, has been developed since the late 2000s [8, 9]. Depending on the coherent modulation technology, the spectral efficiency of 17.3 bit/s/Hz has been indicated in the year 2018, which is over 40 times the efficiency reached in the late 1990s [10]. In addition, currently, the transmission rates of up to 65 Tb/s per SSMF have been proved across oceans, which corresponds to an increase of 65 times compared to the systems available in the early 2000s [11]. As a result, coherent modulation technology has led to an impressive expansion in data throughput during this decade. Despite the tremendous growth in the capabilities of optical fiber communications technology, but it faces many challenges posed by the continuous increase in the demand for bandwidth and communication services around the world [1] and driven by new applications that consume heavy bandwidth, such as the Internet of Things (IoT) which poses challenging

demand on the present infrastructure and available response times. Machine-to-machine communication will soon outnumber human data traffic [12]. Figure (1.1) illustration that a 5G optical network support various applications and shows that: to increase the access data capacity, it is necessary to increase

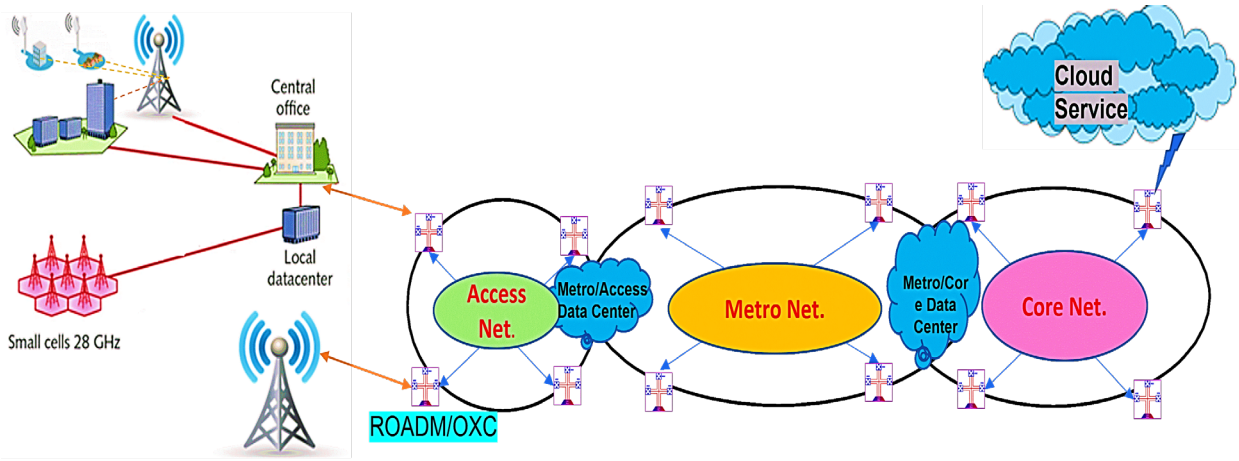


Figure (1.1): 5G fiber optic network consisting. ROADM: reconfigurable optical add/drop multiplexer. OXC: optical cross-connect [12]

the core and metro capacities.

The path towards achieving these challenging and stringent demands requires the future expansion of physical networks and several interdisciplinary tasks to be resolved. New digital signal processing (DSP) methods, self-monitoring systems, low-power optical devices, as well as adaptive and flexible routing mechanisms are required on one side [13]. On the other side, Artificial intelligence (AI) tools for optical signal processing are necessary to optimize the use of resources available in the optical network.

The triangle diagram in Figure (1.2) is shown the three major features to improving the optical transmissions systems in the decade of 2020s that have become the hot area for research and development [14].

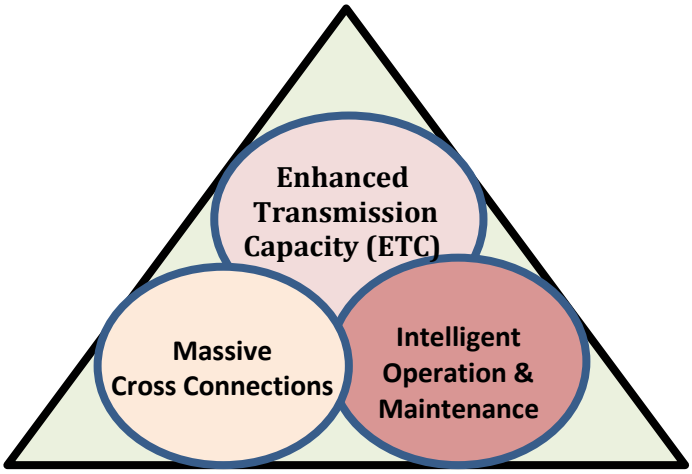


Figure (1.2): Main three features of improvement optical communications systems [14].

Excellently, these guidelines led to a significant development of optical communications. So, increasing the data rate and spectral efficiency for the fiber optic system within an SSMF requires completely utilizing the following five dimensions as shown in Figure (1.3) [14,15]

1. Type of modulation and detection technologies.
2. Polarization (single or dual-polarization).
3. Wavelength Division Multiplexing (WDM).
4. Space Division Multiplexing (SDM).
5. Core Signal Processing Algorithms.

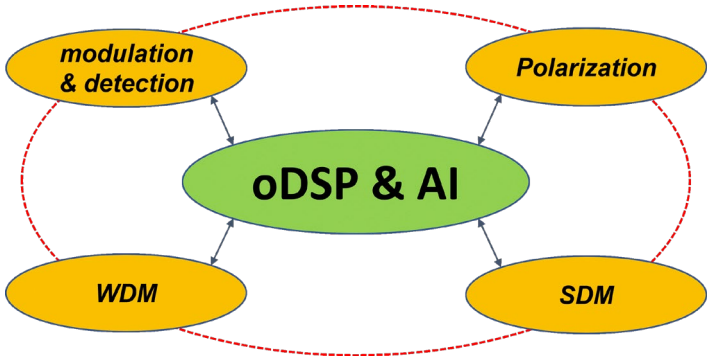


Figure (1.3): The Five Dimensions to Achieve ETC.

Unfortunately, the four physical dimensions shown above require an increase in optical launch power. However, high-power levels reduce the system performance efficiency and trigger the non-linear effects of SSMF known as Non-Linear Interference (NLI) [16]. As a result, linear and non-linear impairments in the fiber link are the main challenges standing against improving the capacity of optical systems. The linear impairments system modeling can be considered as a Linear Time-Invariant (LTI) channel, which can be characterized by its impulse response [16]. However, there is no absolute certainty in determining the NLI and calculating it using the approximate solution of Non-linear Schrodinger Equation (NLSE) via Digital Back Propagation (DBP) or the Split-step Fourier Method (SSFM), which is a widely utilized and effective approach for solving Manakov equation. So, this method has become known as Non-Linear Compensation (NLC) [17,18]. It will be discussed and studied in a brief detail in Chapter 2.

Alternatively, a new totally different approach uses Artificial Intelligence (AI) algorithms is used to identify and resolve optical impairments. AI allows optical communication to have a flexible statistical analysis regarding the complex systems with no dependence on certain models and has demonstrated high potential for improving non-linear compensation performance besides monitoring, traffic control, and signal design [19, 20].

1.2 Literature Review

Syed Tajammul Ahmad and Pradeep Kumar K, (2016) [24], proposed a neural network model based Non-Linear Equalizer (NLE) for coherent optical OFDM (CO-OFDM) systems. The weights of hidden layer neuron of the NLE model calculated based on K-means algorithm. The proposed model was provided up to 4 dB performance improvement in terms of Q-factor for 70-Gbps 16-QAM CO-OFDM transmission over 1000 km (10x100 km) fiber spam with Q=6.25 dB.

Elias Giacoumidis and et al., (2017) [77], Nonlinear effects are experimentally tackled, for the first time, in WDM-CO-OFDM by an artificial neural network (ANN)-based equalizer at 3200 km. For the middle 20-Gb/s channel ANN outperforms to Volterra-based equalization by -2-dB in Q-factor.

Mikko Närhi and et al., (2018) [23], developed a new method depending on ML algorithms and showed how ML could overcome the problems and limits of evaluating the time-domain features regarding the optical fiber modulation instability using simply spectrum intensity measurements. Unsupervised learning was also employed for classifying the noisy modulation instability spectra into subsets that corresponded to different temporal dynamic structures.

Jones Rasmus Thomas and et al., (2018) [25], by including a fiber channel model within two NNs, suggested a new constellation shaping approach for fiber optic communication systems. For the suggested design, experimental and simulation studies were undertaken, with performance gains of up to 0.12 bit/4D in the experiment case and 0.13 bit/4D in the simulation case.

Maximilian Schaedler and et al., (2019) [21], In a coherent DP-16QAM 88 Gbaud 600 Gbps optical system, suggested a DNN non-linear equalizer (DNN-NLE). The proposed design has been compared with full kernel Volterra non-linear equalizers (FKVNEs) and 5th order memory polynomial (MP). DNN-NLEs have been shown to reflect systematic non-linearities more accurately compared to 5th MP and FKVNEs. Depending on back-to-back measurements, it outperforms by 0.8 dB and 0.5 dB, respectively. Also, the author highlighted the performance gains of DNN-NLEs compared to the classic Volterra schemes. However, the complexities of the implementation were not compared.

Shaoliang Zhang and et al., (2019) [22], developed a new method for determining non-linear impairment that utilized machine learning algorithms to learn from the supplied data. After a 2800 km SSMF transmission link at 32 Gbaud signal data, researchers proposed a single-step, system-agnostic NLC algorithm depending on NN and showed a performance gain of roughly (~0.6) dB Q-factor enhancement in comparison to the single-step filtered DBP approach. Intra-channel cross-phase modulation (XPM) and Four-wave mixing (FWM) have been utilized as input features.

Wenlong Zhang and et al., (2019) [27] proposed and experimentally demonstrated a new NNs design and applied it to the CO-OFDM transmission system. The suggested design efficiently decreases the peak to average power ratio (PAPR) of CO-OFDM signals, according to the experimental results. The NNs were trained using the AWGN channel. The suggested design was proven in the experiment to have a considerable reduction of (~4) dB in BER.

1.3 Problem statement

Eliminate and reduce the linear and nonlinear optical effects in the Ultra-High Speed Dense Wavelength Division Multiplexing (DWDM) transmission system to increase the system performance against optical impairments.

1.4 Aim of the Thesis

1. The investigation of various machine learning (ML) algorithms for the compensation of different types of optical transmission impairments.

2. Design and performance evaluation of optical NLC model based on ML advanced algorithms under different optical turbulence conditions.

3. Optical System performance improvement by using new NLC structures modul based on ML algorithms tools.

1.5 Structure of the Thesis

This dissertation is divided into six chapters, and the organization of the chapters is detailed as follows: -

- **Chapter One** presents an introduction to optical impairments. Also, explain in detail the motivation of this research and demonstrate the significance of using Machine Learning (ML) for system performance improvement. In addition, this chapter includes a recent literature survey, the aim of the thesis, and the thesis structure.

- The fundamental constructing blocks regarding a coherent optical communication system are presented in **Chapter Two**. Furthermore, this chapter is a theoretical section that explains the general mathematical

model for many types of optical impairments and is relevant to the fiber optic communication system

- **Chapter Three** provides a thorough overview of all AI approaches used in optical systems, and provides a theoretical introduction to ML, focusing on DNNs.

- The design specifications for the suggested fiber optic system are presented in **Chapter Four**, which includes Dual Polarization CD-OOFDM) systems, Single Polarization Coherent Detection-OOFDM (SP-CD-OOFDM), and several modulation formats with WDM. This chapter also presents the suggested ML framework depending on DNNs that is applied to a designed fiber optic system.

- **Chapter Five** incorporates commercial software MATLAB coding and Optiwave photonic software (Optisystem-17.1) to give simulation findings of an ML-based receiver for coherent fiber-optic system performance study under different turbulence circumstances and transmission distances. The Q-factor improvement and BER are also compared in the simulation results analysis. Both received and transmitted power, as well as the Optical-Signal-to-Noise Ratio (OSNR), will be shown and discussed.

- **Chapter Six** presents the conclusion of the research results and summarizes the open challenges of the presenting methods. Also, it presents the suggestions directions for future works which can improve the performance of the whole technique.

Chapter 2
Fundamentals of Coherent Optical Communication (COC) System

2.1 Introduction

A fiber optic channel, an optical transmitter, and an optical receiver make up a simple COC system. A very brief introduction to the fundamental building blocks of a COC system and all theories and designs considerations related to it will be presented in this chapter. The concepts presented here are relevant to the primary subjects discussed in the following chapters of this study.

2.2 Coherent Optical Systems

The amplitude was utilized for encoding information in standard intensity-modulated/direct-detection (IM/DD) systems, whereas amplitude along with the phase and carrier frequency were employed for encoding data in coherent systems. In these systems, the receiver should be capable of identifying and accessing the frequency/phase regarding the transmitted carrier for determining which signals were transmitted. As a result, the usage of a Local Oscillator (LO) is essential. The simplified components related to a conventional COC system are shown in Figure 2.1. The Forward Error Correction (FEC) block is depicted in system's logical part, whereas the Digital Signal Processing (DSP) blocks for equalization and synchronization are exhibited in the electrical part. The optical level, which includes the fiber optic channel and a modulator at the transmitter and the coherent front-end at the receiver, each with its own LO, is the final component [28-30].

Coherent detection has two key benefits over IM/DD systems.

1. Using advanced coding and modulation systems was enabled

through multi-parameter scanning in the coherent detection system, which allowed designers to analyze the tradeoffs in terms of parameters of interest and select an optimal design for deployments through examining all available orthogonal dimensions to multiplex, encode, and transmit information.

2. Because of the entire knowledge of the electric field acquired, high spectrum efficiency modulation approach was reached, however coherent detection allows the receiver to minimize InterSymbol noise and efficiently adjust for fiber optic channel imperfections.

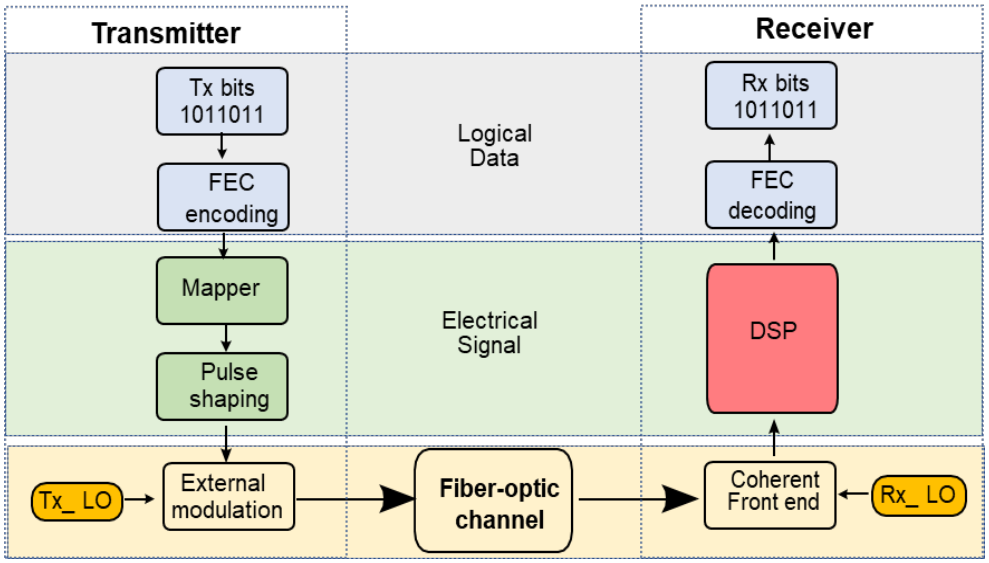


Figure (2.1): Typical COC system [28].

2.3 The Transmitter

The transmitter components regarding a COC system are shown in this section.

2.3.1 Optical Source

There are two types of photon-based optical source devices, which are: Laser Diode (LD) and Light-Emitting Diode (LED) [31-34].

2.3.1.1 Laser Diode (LD)

Light Amplification by Stimulated Radiation Emission is abbreviated to (LASER). Stimulated emissions occur when a photon strikes a stimulated electron, which causes it to fall and radiates a photon with an identical frequency and phase as incident photon during the transition. Figure (2.2) shows the fundamental diagram regarding a semi-conductor optical amplifier that is considered as the major idea behind the LASER technology. Prior to the transmission of the data signal across optical fiber, an optical carrier should optimally be continuous light-wave with constant frequency and amplitude. Unfortunately, there isn't a gadget that can create an ideal light source. The majority of optical carrier needs are met by steady lasers with linewidth values in MHz range. Direct modulated systems (for short-reach applications) and coherent detection systems (for long transmission distances) and both use such lasers [34].

The phase noise amount that contributed to carrier signal is governed by the linewidth of the lasers at the receiver and transmitter combined. Furthermore, the laser spectra have a broader spectral linewidth than a single spectral line [35].

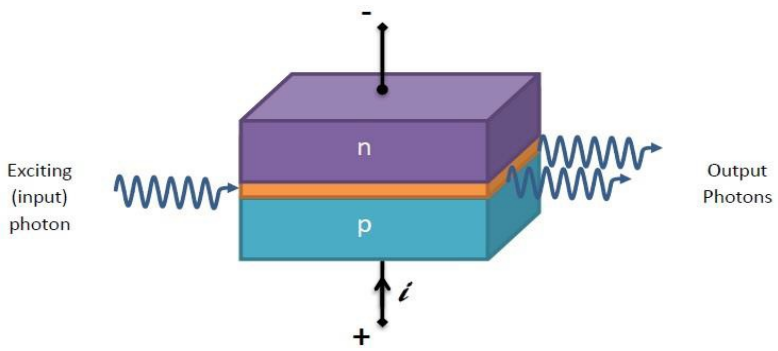


Figure (2.2): Laser Diode Diagram [34].

2.3.1.2 Light-Emitting Diode (LED)

When exposed to a forward biased voltage, the LED represents a semiconductor and optical source which emits optical beam. The LED's mechanism is a p-n connector. In the case where direct voltage has been applied to LED, electrons from n-side (i.e. anode) flow to p-side (i.e. cathode). When an electron reaches p-side hole, it releases the energy as photon. The radiated photon's wavelength is determined by energy gap between the materials of the semi-conductor. Figure (2.3) illustrates a block diagram of an LED. The comparison between LD and LED is shown in Table 2.1 [36].

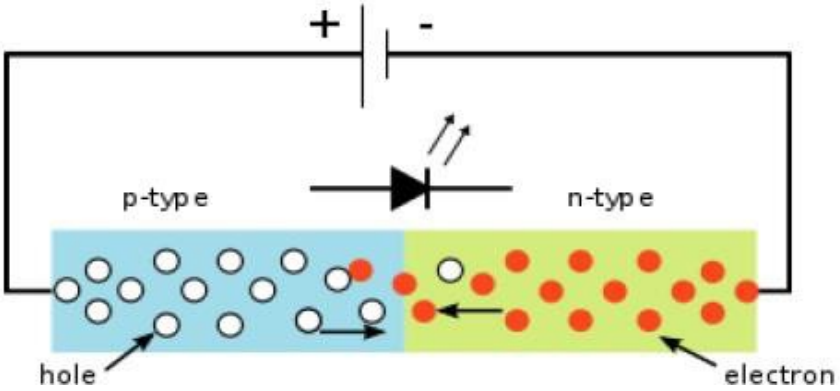


Figure (2.3): LED [36].

Table (2.1): Comparison for Laser Diodes and LEDs [36].

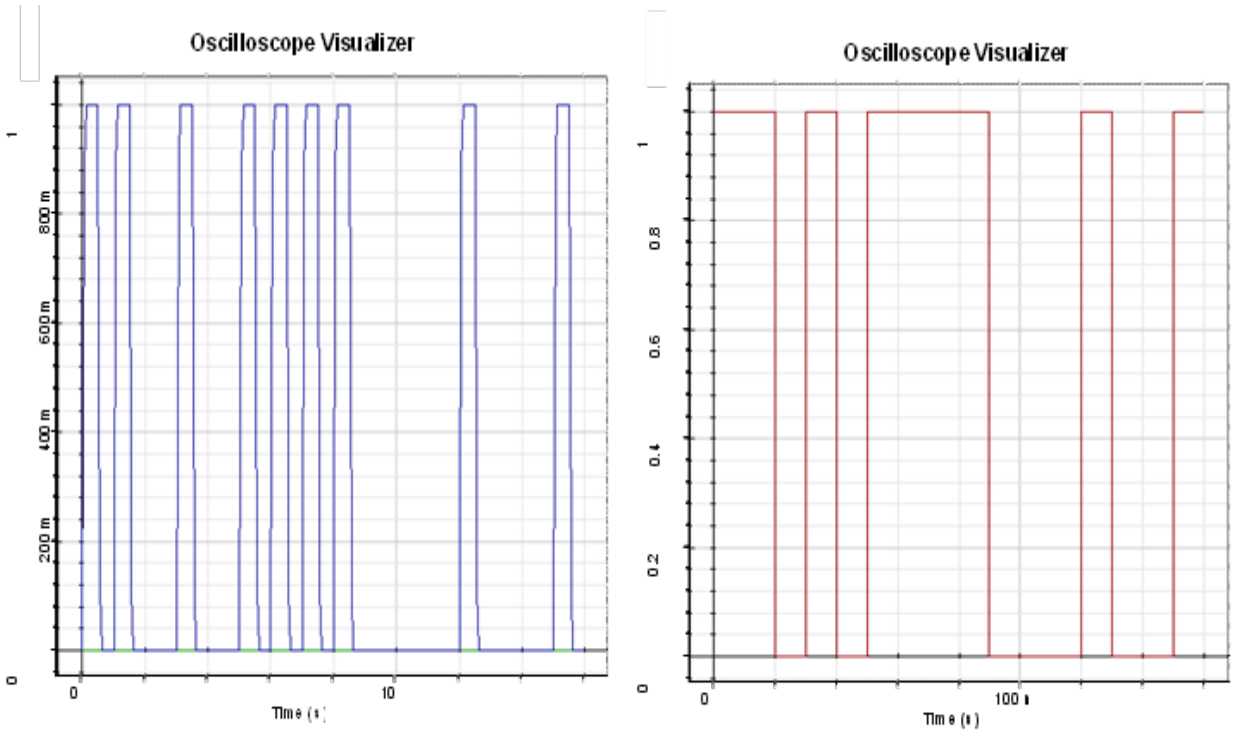
Character	LD	LED
Temperature Dependence	Varying temperature Dependence	Little
Spectral Width	Narrower (0.01°)	Wider (0.5 °)
Output Power	Higher	Lower
Lifetime	Medium lifetime	Long lifetime
Cost	Expensive	Cheap
Light Source	Coherent, no self-interference	Incoherent, self-interference

2.3.2 The Optical Formats

The conversion of an electrical signal into a bit stream is the initial stage in creation of a system of optical communication. The modulation format of the signal can be one of two types:

1. Non-return-to-Zero (NRZ)
2. Return-to-Zero (RZ)

Each one of the pulses represent bit 1 in RZ formats is shorter compared to the bit slot, while its amplitude returns to 0 prior to when the duration of the bit is over, as seen in Fig. (2.4a). The pulse stays on over the bit slot in the NRZ format, while its amplitude doesn't drop to zero between at least two consecutive 1 bits, as seen from Figure (2.4b). Therefore, pulse width differs based upon bit pattern, but in RZ format, it remains constant. Since the on-off transitions take place less frequently in NRZ format, the bandwidth related to the bit stream is around a factor of two smaller than in RZ format.



(a) RZ modulation format

(b) NRZ modulation format

Figure (2.4): Optical modulation formats [37].

2.3.3 The Optical Modulator

The modulation type of the transmitted signal, modulating the information on the optical carriers, has a critical function for the digital and optical systems of communication.

The phase, amplitude, and frequency of information carried by an optical carrier are modulated in optical communications systems in a similar way as they are in the radio frequency communications systems [38]. The in-phase (I) and quadrature (Q) components of optical signals carry information corresponding to phase and amplitude in coherent modulation systems [39, 40]. Two "Mach-Zehnder" modulators (MZM) are used in an I/Q modulator. MZMs work by splitting the light from a laser source into two spindles, each of which holds one MZM. In addition, MZMs are amplitude modulators with only one dimension. A total of two orthogonal signals are formed and thus 2Ds are modulated with the addition of a (π/2) phase shift in one of spindles, as shown in Figure (2.5). The given modal is replicated for both Y and X polarization in dual-polarization transmission.

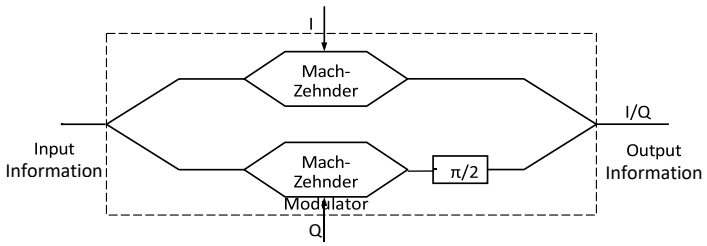


Figure (2.5): I/Q Optical modulator [38].

2.4 The Fiber-Optic Channel

Standard Single-Mode Fiber (SSMF) can be defined as a dielectric cylindrical waveguide made of glass or silica, with total internal reflection as its guiding basis. Within SSMF, the light propagation is affected by both nonlinear and linear factors, which should be considered jointly in order to create an appropriate fiber channel model. Figure (2.6) shows a

common transmission arrangement made up of fiber spans. The transmitted optical waveform is subjected to loss, attenuation, chromatic dispersion (CD), and non-linear effects across the span. Those optical effects are modeled using non-linear Schrödinger equation (NLSE). The NLSE has been considered as differential equation which indicates the propagation of light through SSMF [3, 16,41]:

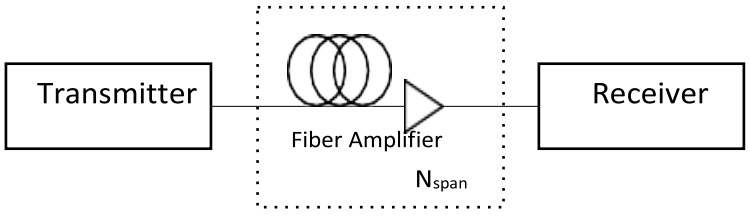


Figure (2.6): Optical fiber communication system with optical Amplifier [3].

The slow-varying complex-valued envelope of an optical carrier propagating in an SMF is represented by the Jones vector of the electrical field A(z, t) The NLSE equation can be written:

$$\frac{dA}{dz} = -\frac{\alpha}{2} A - j \frac{\beta_2}{2} \frac{d^2A}{dt^2} + j\gamma |A|^2 A \qquad (2.1a)$$

$$\gamma = 2\pi n_2/(\lambda_c A_{eff}) \qquad (2.1b)$$

In which β_2 and α represent the 2nd-order chromatic dispersion effects and fiber loss, respectively, where the 3rd-order CD effect is neglected. Also, the non-linear coefficient (γ) can be specified as the non-linear-index coefficient n_2, in which λ_c represent the center wavelength and A_{eff} represent the fiber's effective core area. Because of the differing propagation rates of light waves at various wavelengths, CD causes the broadening of the pulse in the time domain in a single channel and a walk-off over the channels of the WDM. The "nonlinear impairments" term in Kerr-Effect introduces non-linearities modeled in the fiber optic channel, like the self-phase modulation (SPM), cross-phase modulation (XPM), and Four-wave mixing (FWM) [3, 16]. For dual-polarization light waves, the matrices α, γ, β_2 are given by:

$$\alpha = \begin{bmatrix} \alpha_x & 0 \\ 0 & \alpha_y \end{bmatrix}, \gamma = \begin{bmatrix} \gamma_x & 0 \\ 0 & \gamma_y \end{bmatrix}, \beta_2 = \begin{bmatrix} \beta_{2x} & 0 \\ 0 & \beta_{2y} \end{bmatrix},$$

Regarding the majority of practical cases, the second order dispersion, attenuation, and non-linear coefficients are the same for both polarizations, i.e., $\beta_{2x} = \beta_{2y} = \beta_2$, $\alpha_x = \alpha_y = \alpha$, and $\gamma_x = \gamma_y = \gamma$. Also, in the case when only the combination regarding those effects is object of analysis, then the Manakov equations might be utilized for further simplify NLSE, and Eq. (2.1) become the next pair of coupled scalar NLSEs [16].

$$\frac{dA_x}{dz} = -\frac{\alpha}{2} A_x - j \frac{\beta_2}{2} \frac{d^2 A_x}{dt^2} + j \frac{8}{9} \gamma \left(|A_x|^2 + |A_y|^2 \right) A_x \qquad (2.2a)$$

$$\frac{dA_y}{dz} = -\frac{\alpha}{2} A_y - j \frac{\beta_2}{2} \frac{d^2 A_y}{dt^2} + j \frac{8}{9} \gamma \left(|A_x|^2 + |A_y|^2 \right) A_y, \qquad (2.2b)$$

In which $A_y(z,t)$ and $A_x(z,t)$ represent the pulses of the two orthogonal light wave polarizations modes. In addition, the dispersion of the polarization mode is not taken into account in (2.2). Over long fiber lengths, the factor 8/9 of expression is responsible for a mean value of interplay between the non-linearities and continuously rotating fields of polarization. The terms $|\ |^2$ in the two equations (2.1 and 2.2) indicate the dependence of light signal on its strength (i.e. its power). As a result, at high power, non-linear impairments predominate and performance is hampered. The typical characteristics for such SSMF systems are listed in Table (2.2) [42].

Table (2.2): Typical parameters of Corning SMF-28 [42]

Dispersion D	17 ps/(nm.km)
Fiber loss α (attenuation)	0.2dB/km
Carrier wavelength λ	1550 nm
Non-linear coefficient γ	1.3 $(W\ km)^{-1}$

The amplification noise is the final variable to be considered. The noise is caused by compensating for fiber losses through the amplification of optical signal at each span's end. In this study, the Erbium-Doped Fiber Amplifier (EDFA's) amplified spontaneous emission (ASE) is used, which is normally represented with an additive white Gaussian noise (AWGN) with the Power Spectral Density (PSD) [43].

$$n_{spe} = \frac{(NF-1).G}{2(G-1)}, \tag{2.3a}$$

$$\sigma^2_{\mathrm{ASE}} = B.w.\frac{n_{spe}.C.h.(G-1)}{\lambda_c}, \tag{2.3b}$$

$$OSNR = \frac{Psig}{\sigma^2_{\mathrm{ASE}}} \tag{2.4}$$

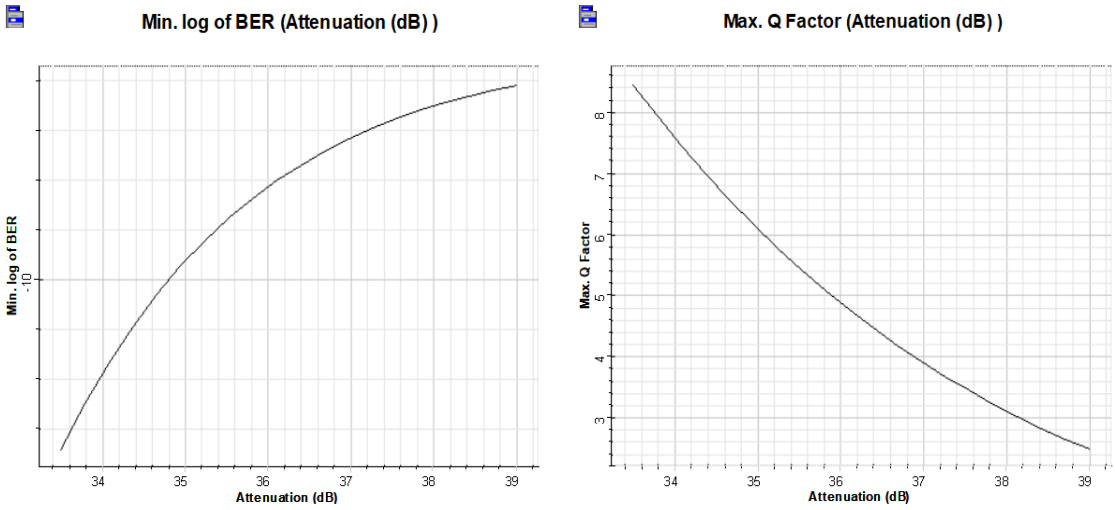
where n_{spe} represent the factor of the spontaneous-emission, h represents Planck constant, G represent the gain, C represents light speed, NF represent noise figure and in the range of 4 to 6 dB, $B.w$ represent the reference bandwidth and typically selected to be 12.5 GHz (0.1 nm), λ_c represent the center wavelength, and $\mathrm{P_{sig}}$ represent the average signal power over the two polarization.

2.4.1 Attenuation

The optical pulse transmitted by an optical fiber is attenuated, which is one of the most critical physical processes that limits efficient performance regarding optical system. It is caused by a variety of factors, the most common of which being Rayleigh scattering and material absorption [16]. The loss factor coefficient (α) was used to include the effect of power loss in the two equations (2.1 and 2.2). The effects of attenuation on optical data transmission with regard to bit error rate (BER) and Quality factor (Q) are depicted in Figure (2.7). The power loss is compensated through employing an optical amplifier, which is the most effective and frequent way. The optical

amplifier is positioned at the end of every one of the fiber spans in the WDM system, as shown in Figure (2.8), which provides an example of a span configuration and EDFA with SSMF. In equations (2.3) and (2.4), the noise effect of EDFA was discussed.

The signal will not be correctly detected in the case when the optical signal power at receiver is less compared with the receiver's sensitivity. The power margin is considered as the difference between the receiver sensitivity and the observed signal power. The system loss is compared to the difference between the receiver sensitivity and transmitter power using the budget equation (2.5):



(a) (b)

Figure (2.7): Attenuation effect on optical data transmission in terms of (a)BER and (b) Q-Factor.

$$P_t - S_r = AL_f + L_c + L_a + M, \qquad (2.5)$$

Where, P_t = transmitter power, AL_f = fiber loss, L_a = additional loss, S_r = receiver sensitivity, L_c = coupler loss, and M = power margin. According to the datasheet for Corning SMF-28 fiber, the additional known loss L_a = 0 dB, the total coupling loss L_C = 2 x 0.5 dB = 1 dB, and the system margin M = 6 dB [16].

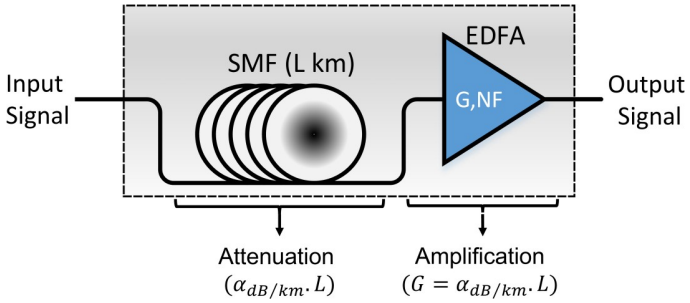


Figure (2.8): Example of a fiber span with an optical amplifier [16].

2.4.2 Chromatic Dispersion (CD)

Because of the linear effect which is known as the chromatic dispersion (CD), also known as group-velocity dispersion (GVD), or "dispersion," the frequency components regarding an optical signal traveling over the fiber will meet varying group velocities. The CD causes pulse broadening by producing a relative time delay among frequency components of a pulse travelling along the fiber. There is no modal form between various propagating modes, and dispersion is not totally eliminated. Material and waveguide dispersion are the most common types of dispersion. The index of refraction causes material dispersion, whereas the propagation constant causes the other. A time pulse will broaden through propagation in the case when the derivative of the group velocity in terms of frequency is non-zero, and the spectral components will arrive at various times. The influence of dispersion on pulse propagation in optical fibers is in the "linear" zone, see Figure (2.9), with the following key effects [3, 16]:

1. Dispersal induced pulse chirping.
2. Dispersal induced pulse broadening.
3. Dispersal induced pulse compression.

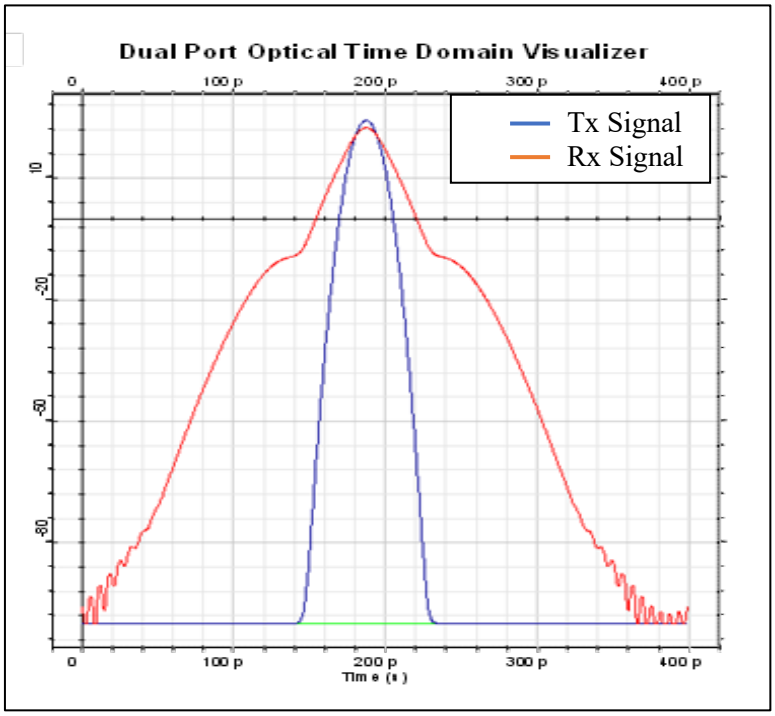


Figure (2.9): Effect of dispersion on optical pulse propagation [3, 16]

The pulse broadening ΔT, is associated with the derivative of phase constant β by:

$$\Delta T = L\,\Delta\omega\,(d^2\,\beta / d\omega^2) = L\,\beta_2\,\Delta\omega, \qquad (2.6)$$

In which $\Delta\omega$ represent the pulse spectral width and L represent the fiber length. In another form, might be written with regard to wavelength as:

$$\Delta T = L\,D_\lambda\,\Delta\lambda, \qquad (2.7)$$

$$D = -2\pi\,c\,\beta_2 / \lambda_c^2, \qquad (2.8)$$

Furthermore, anomalous dispersion occurs when a fiber's dispersion characteristic has a positive value, and higher frequency components move quicker. Normal dispersion occurs when the value is negative, and lower frequency components travel quicker, as seen in Figure (2.10) [3, 16].

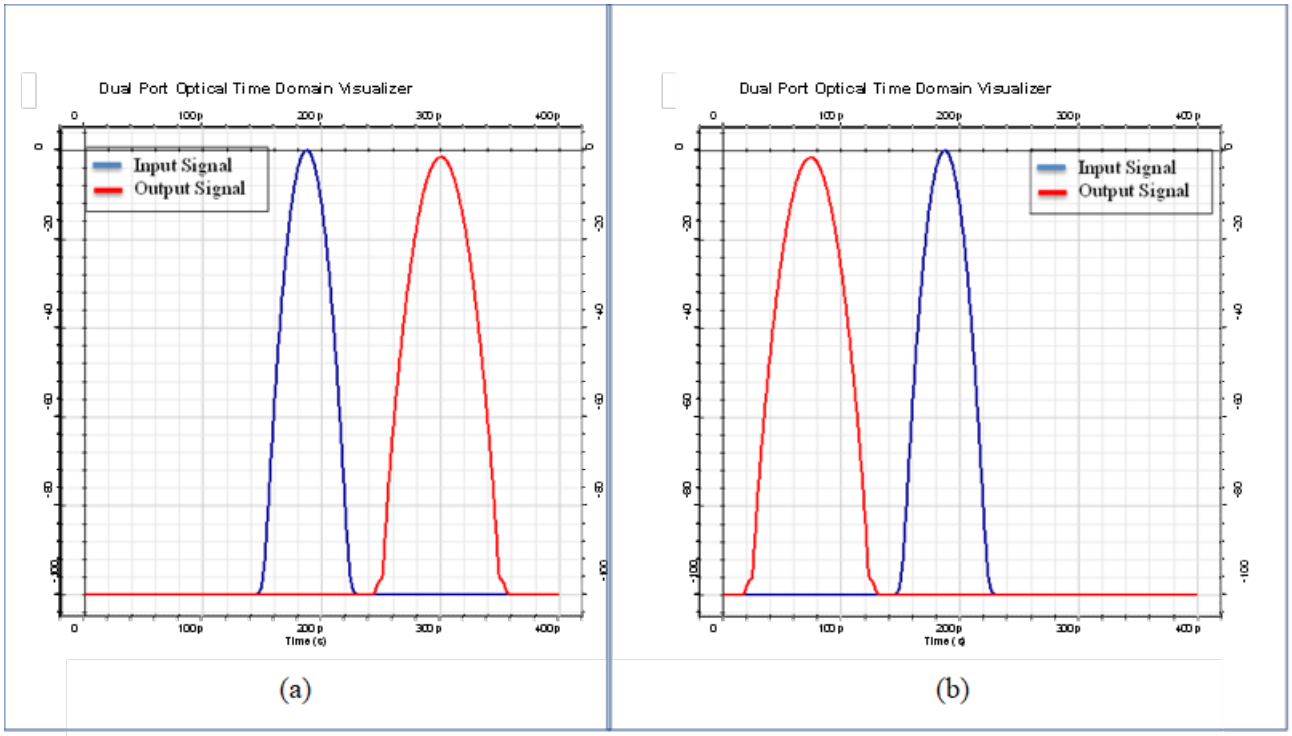


Figure (2.10): (a) Normal dispersion, (b) Anomalous dispersion

A variety of approaches were explored to eliminate this dispersal, including:

1. Fiber Bragg Grating (FBG).
2. Dispersion Compensation Fiber (DCF).
3. Optical Phase Conjugation (OPC).
4. Dispersion-Equalizing Filters.
    I. All-Pass Filters.
    II. Mach-Zehnder Filters.
    III. Gires-Tournois Filters.
5. Electronic-Equalizing Dispersion Compensation.

Now, the DCF will be discussed, according to their use in this work. The maximum allowable pulse spread (Δt_{max}) is given as: ($\Delta t_{max} = 1/4B$), where B is bit-rate. and the pulse spread due to chromatic dispersion is provided in eq (2.7) as $\Delta t = LD_{(\lambda)}\Delta_{(\lambda)}$, while the chromatic dispersion factor $D(\lambda)$ at the operating wavelength write as [3, 16]:

$$D(\lambda) = \frac{S_0}{4}\left[\lambda - \frac{\lambda_0^4}{\lambda^3}\right] \quad (2.9)$$

where S_0 = zero dispersion slope (ps/nm².km) and λ_0 = 0 dispersion wavelength (nm), and from the specification sheet for Corning SMF-28, the λ_0 = 1312 nm, and the S_0 = 0.09 ps/nm^2, with 1550 nm operating wavelength. In addition, $\Delta\lambda$ represent the RMS spectral width regarding the optical source. Chirped Gaussian pulse utilized for representing the output of a laser diode and it is specified via RMS pulse width (T_{in}) and the chirp factor (*C*). Furthermore, the T_{in} is associated with the Gaussian Full Width at Half Maximum (FWHM) pulse width (T_{FWHM}) as:

$$T_{FWHM} = 1.665T_{in} \quad (2.10)$$

Then, the RMS spectral width ($\Delta\lambda$) is obtained as:

$$\Delta\lambda = \frac{\lambda^2\sqrt{1+C^2}}{2\pi\, c\, T_{in}} \quad (2.11)$$

The maximum allowable dispersion-limited fiber length (L) can now find from eq (2.7). If the length of the fiber optic used in the designed system is greater than (L), then it is necessary to use DCF. Then, to calculate the length of the DCF (L_{DCF}), the equation (2.7) it becomes as below:

$$\Delta t = L\, D(\lambda)\, \Delta(\lambda) + L_{DCF}\, D(\lambda)_{DCF}\, \Delta(\lambda)_{DCF} \quad (2.11)$$

2.4.3 Polarization Mode Dispersion

The ideal SSMF will support two symmetry degenerated orthogonal polarization modes, which means, equal refractive indices ($n_x = n_y$) will be for both modes. However, in real fibers, small defects during the production process or differences in the basic shape due to bending or other external factors will break the symmetry of the fiber waveguide, causing the appearance of modal birefringence (B_m). The B_m is described by a difference between the refractive indices of the two orthogonal polarization modes ($n_x > n_y$, or $n_x < n_y$).

Therefore, the polarized light is going to be propagated at various speeds [3, 16].

$$B_m = |n_x - n_y| = \Delta n \quad (2.12)$$

which results in a difference in the propagation constants related to the polarization modes.

$$\Delta\beta = | \beta_x - \beta_y | = (\omega / c) \Delta n. \quad (2.13)$$

The difference in relative conditioned group delay ($\Delta\tau$) is referred to as the differential group delay (DGD) which is specified as

$$\Delta\tau = | \beta_{1x} - \beta_{1y} | L \quad (2.14)$$

The first-order polarization mode (PMD) is referred to as the DGD value. In the case when DGD doesn't have a constant frequency response over the desired bandwidth, the polarization dispersion should be calculated using higher-order PMD conditions. Yet, as a non-linear fiber model, the average effect regarding modal birefringence has been included in NLSE. In Kerr-Nonlinearity, Eq. (2.1) and Eq. (2.2) simulate the average effect of birefringence, yet not statistical modeling of PMD. Unfortunately, the PMD's time-varying effects are the primary drivers of an ISI, affecting optical communication system performance, particularly at high bitrate transmissions [3, 16, 44-46].

2.4.4 Kerr Nonlinearities

The presence of an electric field within the dielectric silica fibers will lead to creating a dielectric polarization field. The interactions between electrical field and dielectric will result in producing an equivalent field in the medium. This interaction represents a non-linear input field function and causes Kerr effect that is present in the silica fibers and responsible for NLI in the systems of optical communication [47]. The Kerr NLI is the complex interaction between linear and nonlinear interactions and apparent as data-dependent collective multiplicative distortions which is a function of noise that

is added by physical characteristics of the fiber medium, and the copropagating signals. A part of Kerr NLI results from the signal-noise and noise-noise interactions and it is stochastic in nature, while the second part, originates from signal-signal interactions and has a deterministic nature [48]. The Kerr NLI is included in the NLSE propagation models given in Eq. (2.1) and (2.2) by the term the nonlinear coefficient (γ), and usually specified in 1/(W.km). If there is no loss, dispersion, and nonlinear effects it will have the exact solution by NLSE propagation models. However, the dispersion and nonlinear effects are real and inevitable in fiber optic system and in this case, the exact solutions for NLSE are not known, generally [3, 16].

In coherent optical WDM systems, the NLI is modeled as additional AWGN source at receiver, and signal-to-noise ratio (SNR) realized by receiver is represented as [50, 51].

$$SNR_{Rx} = \frac{P_{sig}}{\sigma_{NLI}^2 + \sigma_{ASE}^2} \tag{2.15}$$

Figure (2.11) shows the classification of the NLI effects from the receiver point of view and Table (2.3) presents the effects of main fiber nonlinearity impairments in terms of bit rate and channel spacing. The intrachannel non-linear impacts are the distortions that are produced inside the bandwidth of a channel include SPM, XPM and FWM [16]. While the inter-channel nonlinear effects are the distortions from different channels in the case of WDM that take place (XPM, FWM). NLI effects caused by signal-signal interactions are deterministic, while on the other hand, NLI effects caused by signal-noise interactions are stochastic [52].

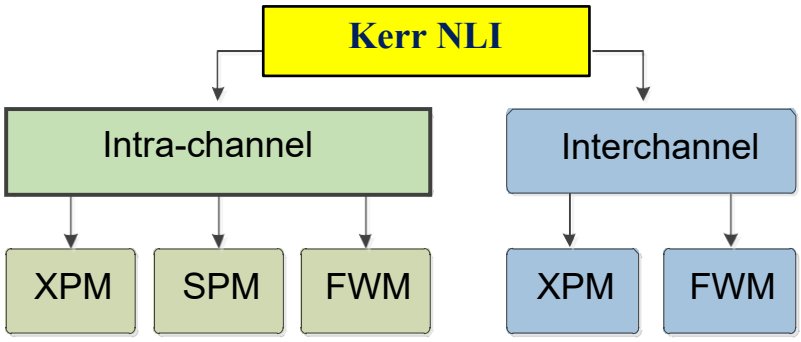


Figure (2.11): Classification of the NLI effect from the receiver point of view [16]

Table (2.3): NLI effect [16].

Types	XPM	SPM	XPolM	FWM
Bit rate (+)	+	+	+	no effects
Channel spacing (-)	+	no effects	+	+

* (+ increase), (- decrease)

2.5 Coherent Optical Receivers

In the present section, a short description of the coherent optical receiver will be presented, followed by an explanation of fundamental Nonlinear Impairment Compensation (NLC) based on DSP.

2.5.1 Coherent Optical Front-End

The main work for the coherent front-end of coherent receiver is to convert optical received signal to electrical domain [8]. Firstly, the received optical signal is directed to the local oscillator removing carrier from signal. Secondly, the resultant signal is detected through balanced photodiode devices, then as can be seen in Figure (2.12) the resulting baseband signal is converted into digital domain via the analog-to-digital converter (ADCs) [53]. As a result, the modulated (I, Q) components signal is recovered. Thereafter, the coherent optical receiver system has to perform a stream of the DSP tasks in an attempt of finally recovering transmitted data.

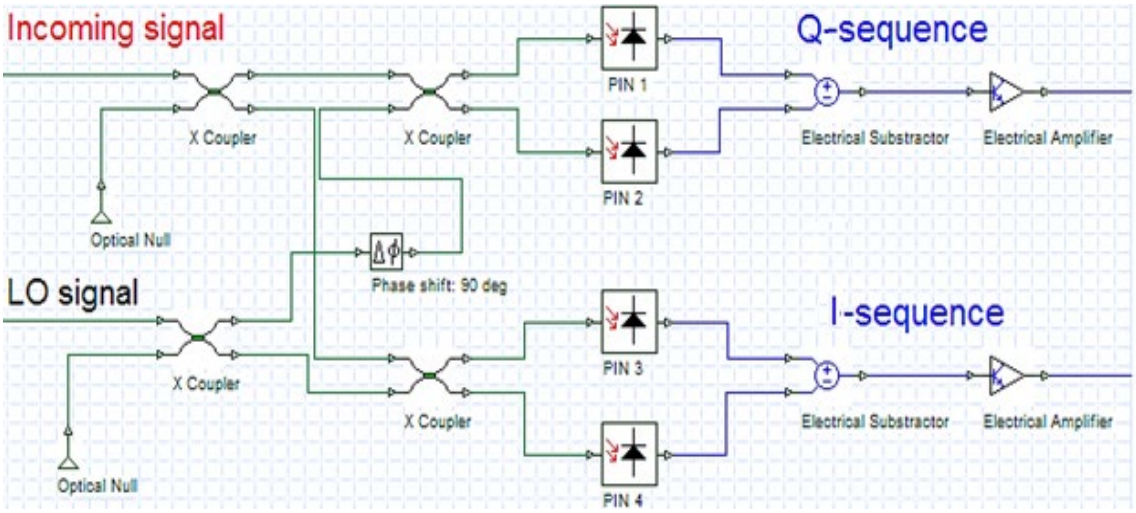


Figure (2.12): Diagram of the standard coherent optical front-end [3, 16].

2.5.2 Digital Signal Processing

The missions assigned to the DSP can be divided into five groups [3, 16 54, 55]:

1. Synchronization (timing recovery).
2. Channel equalization (chromatic dispersion compensation and compensation all residual linear dispersion effects).
3. Digital backpropagation (mitigation of nonlinearities effects NLI).
4. Carrier phase recovery (frequency offset compensation).
5. Symbol decision.

Note, the main objective of the studies presented in this thesis is to compensate the nonlinear impairments (NLI) of the optical transmission systems by using machine learning based on Deep Neural Networks (DNNs).

Based on the above, the next section will present a systematic comprehensive survey for the most effective Non-linear Compensation (NLC) methods that were used in order to reduce or eliminate nonlinear impairments in the fiber optic communications systems, which based upon the approximate

solution of the Non-linear Schrodinger Equation (NLSE) through the split-step Fourier Method (SSFM). In order to make a comparison between the results achieved previously by these techniques and the results achieved from this work proposed in this thesis, which will be discussed in (Chapter 5) in terms of performance and complexity.

2.5.3 NLC Methods based on DSP Algorithms

NLC can be defined as an effective approach and major technology to increase the optical transmission system performances. NLC methods may be carried out in the optical or the digital domains at the transmitter or the receiver parts and others are performed in optical link. Generally, they are implemented at receiver's part. The most commons of the NLC modules include the Volterra Series Transfer Function (VSTF), digital backpropagation (DBP), Turbo Equalizer (TEQ) and Maximum-Likelihood Sequence Equalizer (MLSE) [17]. Table (2.4) lists a detailed survey for the overall fiber NLC methods, and Figures (2.13), (2.14) show the NLC under same experiment parameters.

Table (2.4): Overall fiber NLC techniques [17].

NLC Techniques	Types/Locations	Non-linearity	Tx systems
Digital back propagation (DBP)	Digital/ Tx/Rx	Intra-sub-carrier	Nyquist/OFDM
Phase conjugation (PC)	Optical/ Rx/Link	Non-linear phase	Nyquist/OFDM
Volterra-based nonlinear equalizer (VNLE)	Digital/ Tx/Rx	Intra-sub-carrier	Nyquist/OFDM
Total-field digital back propagation (TF-DBP)	Digital/ Tx/Rx	Intra/Inter sub-carrier	Nyquist/OFDM
Inter-sub-carrier-interference canceler (INIC)	Digital/ Rx	Intra- and inter-sub-carrier	Nyquist/OFDM
Perturbation-based NLC	Digital/ Tx/Rx	Intra-sub-carrier/XPM	Nyquist/OFDM
Wiener-Hammerstein	Digital/ Rx	Intra-sub-carrier	OFDM
Non-linear Fourier transformation	Digital/ Tx/Rx	Intra- and inter-sub-carrier	Nyquist/OFDM
Support vector machine	Digital/ Rx	Intra-sub-carrier	Nyquist/OFDM
RF-pilot tones	Digital/ Rx	Non-linear phase shift	OFDM
Code-aided expectation-maximization algorithm	Digital/ Rx	Non-linear phase noise	Nyquist
Optical back propagation	Optical/ Link	Non-linear phase	Nyquist
Adaptive MLSD	Digital/ Rx	Non-linear phase noise	Nyquist
Electronic compensation method	Digital/ Rx	Non-linear phase noise	Nyquist
Non-linear polarization crosstalk canceller	Digital/ Rx	XPolM	Nyquist

In addition, the drawback of utilizing those techniques include [3, 16, 17, 20]:

1. Complicated computing and implementation.
2. Need for the NLSE model.
3. Susceptive to the mismatch of the model parameter in general.
4. Need for the non-linear signal processing.
5. Need high-complexity SSFM for solving the propagation of the light-wave.
6. Costly and more power-consuming.

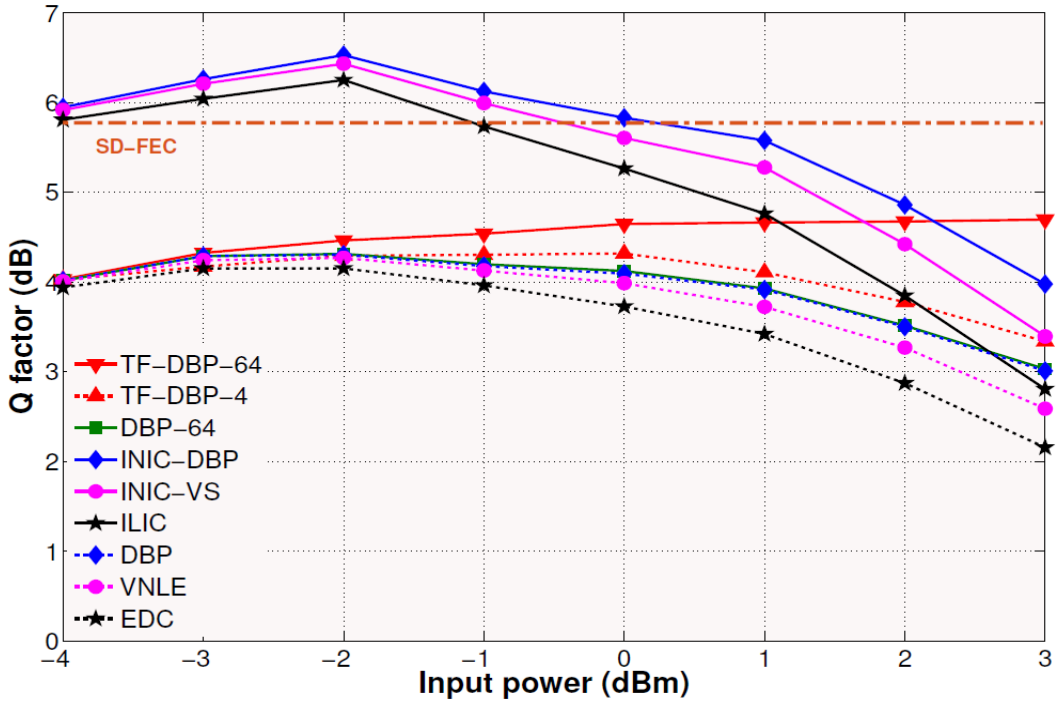


Figure (2.13): Q factor vs. lunch power for most effective NLC techniques under same experiment parameters [56].

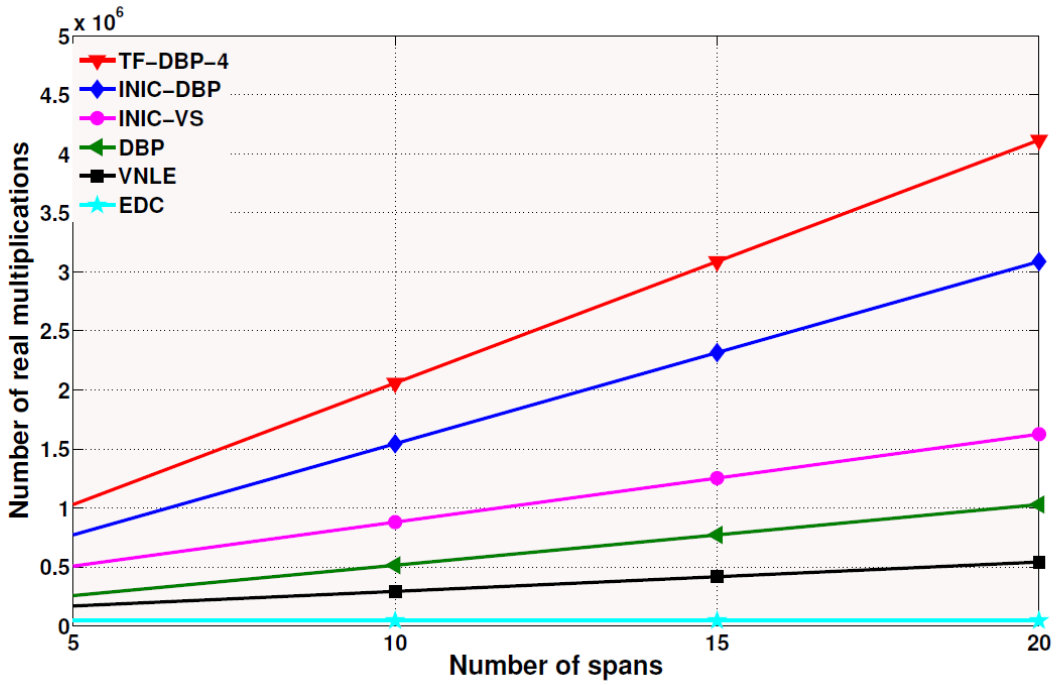


Figure (2.14): NLC complexity analysis under same experiment parameters [56].

2.6 Enhanced Fiber Transmission Capacity

As a reminder in the first chapter Figure (1.3), some of the main key technologies that have been used to enhance the fiber transmission capacity and increase the overall system capacity rate will be discussed in this section.

2.6.1 Modulation and Detection Technologies

In general, the light wave has three basic physical features which may be modified so as to carry data, its amplitude, polarization and phase. Therefore, there are 3 main common schemes of modulation and detection, one is an intensity modulation and direct detection (IM/DD) approach, and the other one is a phase shift keying scheme and the last is a coherent optical modulation scheme.

In IM/DD, on-off-keying (OOK) is used as optical modulation format where the intensity of the light turned off and on for representing the states of

"0" and "1" of the digital signal. OOK represents a simple transmitter and receiver implementation but it's very limited performance when compared with other modulation formats.

In the phase-shift keying, the differential quadrature-phase-shift keying (DQPSK) and differential binary-phase-shift keying (DBPSK) are commonly used as optical modulation formats. The idea is to utilize a phase difference between the neighbor symbols to represent states of the digital signal.

These basic technologies have been greatly developed to meet the demand for global data-traffic. Table (2.5) shows the technological progress and profits made during the last forty years in the optical fiber transmission system. [17].

Table (2.5): Development in optical system over the last forty years [17].

Year	Channel rate (Gb/s)	Modulation format	Link features	System features	Capacity /system (Gb/s)	Reach distance (Km)
1980	2.5	00K-NRZ	Single span	Direct modulation.	2.5 single ch.	100
1990	10	00K(RZ)	WDM-EDFA	HD-FEC	400 (40WDM)	1000
2000	10-40	DPSK QPSK	DWDM ROADM	DQPSK	1600 (40WDM)	1000-3000
2010	100	PND-QPSH	DWDM, WSS, ROADM	Coherent detection, oDSP, SD- FEC	8000 (80DWDM)	4000
2020	200-400	PDM, nQAM, CS, PAM4, DMT	DWDM, WSS, ROADM, EDFA	Advanced oDSP and AI.	16-32 Tb/s (Flexible-grid WDM)	1000-2000

Depending on the electric field in equation (2. 16) with the complex envelope, the amplitude may change and also the phase of the electrical signal [3, 16]. In contrast to OOK, which relied on changing the amplitude only, and also for the PSK, which modified the signal phase only.

$$\boldsymbol{E_t}(t) = Re[E_x e^{i\phi_x - i\omega t}\hat{x} + E_y e^{i\phi_y - i\omega t}\hat{y}]. \qquad (2.16)$$

PSK has no effects on the output power, but the optical carrier's phase. Then there's the logical next step, which is called the Quadrature Amplitude Modulation (QAM) which modulates the phase and amplitude in a simultaneous manner. QAM is a Higher multilevel modulation format, commonly used in numerous digital data radio communications, it was used to meet the demand for capacity [16, 57]. In advanced modulation schemes there might be 4 or even more than 1024 different states that a symbol can take (M), instead of BPSK and QPSK which actually correspond to the (2QAM) and (4QAM) respectively, also instead of only having two states such as 00K. Then, term bit is substituted by the symbol and the bit-rate (R_B) is substituted by symbol rate (R_S) as [3, 16]:

$$R_S = \frac{R_B}{\log_2 M} \tag{2.17}$$

For example, a system has (0, 0.5, 1, 1.50) would have M = 4 and lead to symbol rate that equals to half that of the regular format of the OOK. QAM can be expressed by the complex envelope as:

$$\boldsymbol{E}(t) = Re[(I(t) + iQ(t))e^{i\omega t}]. \tag{2.18}$$

where $Q(t)$ modulates the imaginary part and $I(t)$ modulates real part of the complex signal. At eliminating complex entry, it will become as:

$$\boldsymbol{E}(t) = I(t)\ cos\ \omega t - Q(t)\ sin\ \omega t. \tag{2.19}$$

The (t) signal is out of the phase by 90° and is referred to as the quadrature-phase carrier, whereas the other one is in-phase carrier. For the purpose of visualizing those different symbols, they're plotted on constellation or complex plane as shown in Figure (2.15).

High spectral efficiency is a very important benefit of the QAM format, it can be seen that symbol periods of the 4QAM system are twice as bit periods of the 00K system, and that indicates the fact that you need 50% of the signal's bandwidth. The spectral efficiency measures the number of the bits per sec. that may be transmitted for each Hertz as, given by Equation (2.20) [3].

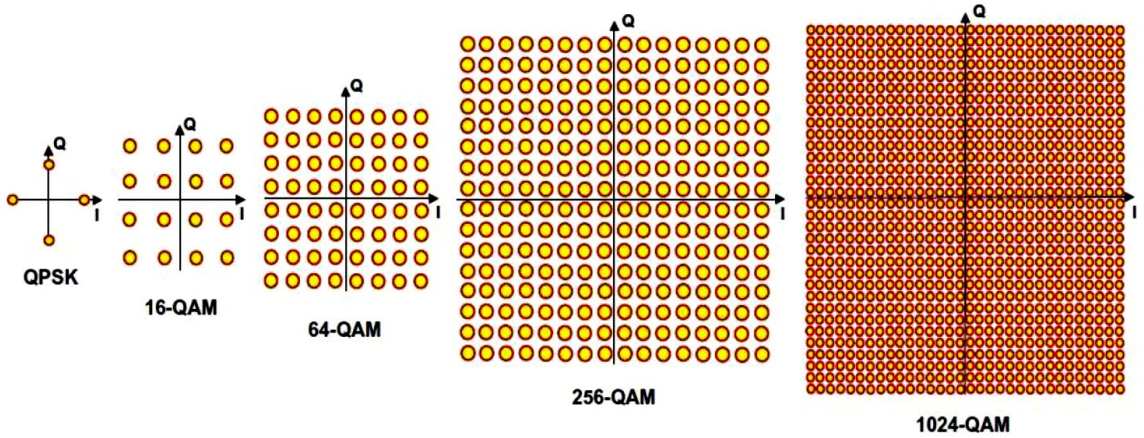


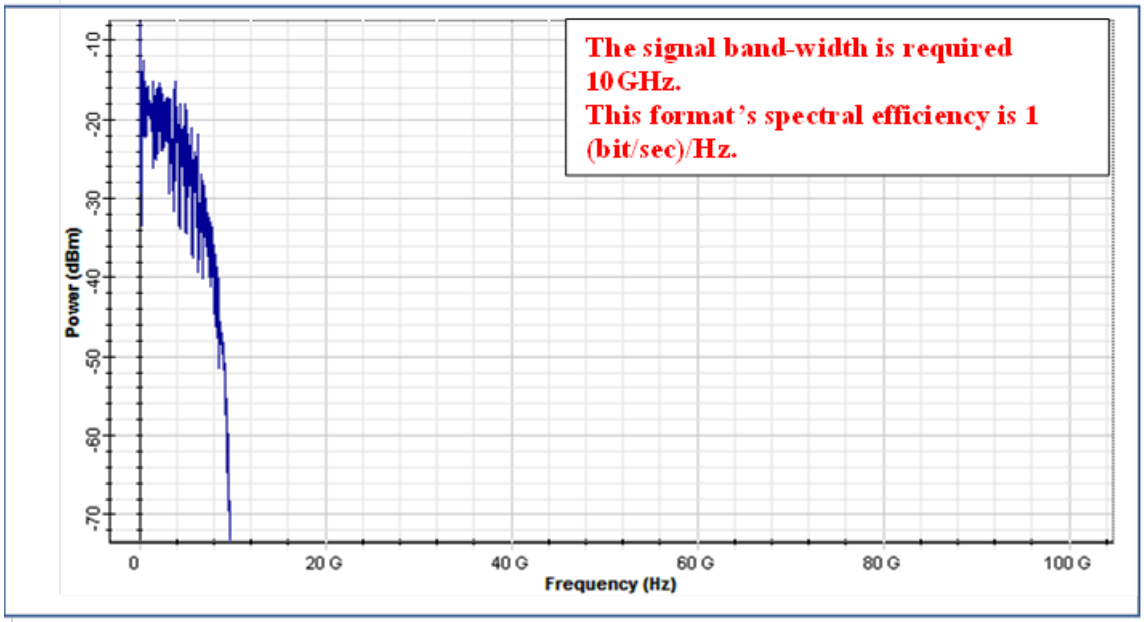
Figure (2.15): Examples of M-QAM Constellations [3,16].

$$Spectral\ Efficiency = Gross\ Bit\ Rate\ /\ Signal\ Bandwidth \qquad (2.20)$$

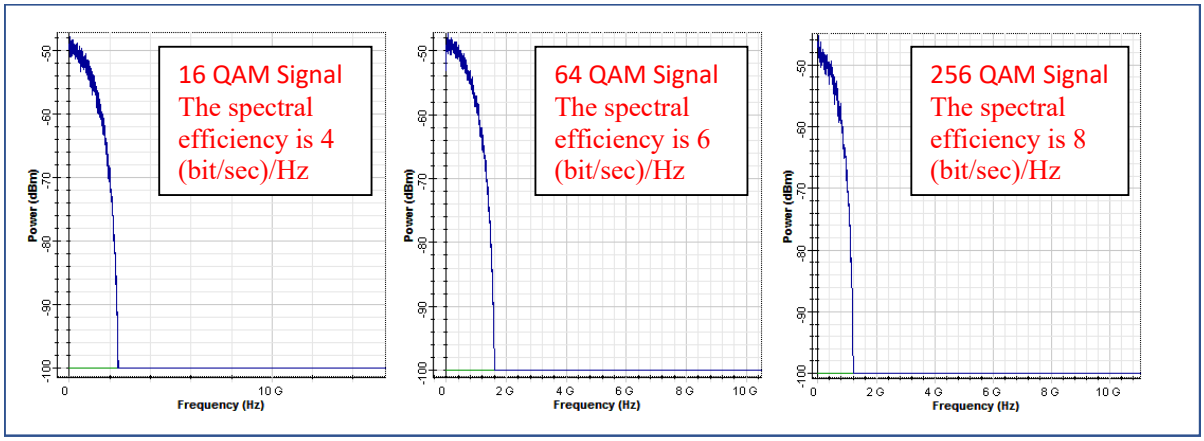
In addition, it can be seen in Figure (2.15) that, the distance between adjacent symbols has been greatly reduced with the increase in the number of symbols and this means that BER increases and a lot of SNR is required with the increase in the number of the symbols and it is understandable by Shannon's theory. Based on Shannon's limit of capacity, Spectral Efficiency (μ) may be represented by [3]:

$$C = Bw.\log_2(1 + SNR) \rightarrow \mu = \frac{C}{Bw}\log_2(1 + SNR) \qquad (2.21)$$

where C represents Shannon capacity and Bw represents the channel bandwidth. It evidently exhibits that increased spectral efficiency has to be paid by higher SNR.



(a)



(b)

Figure (2.16): Power spectrum of the detected 10Gbps [3]. (a)00K. (b)mQAM

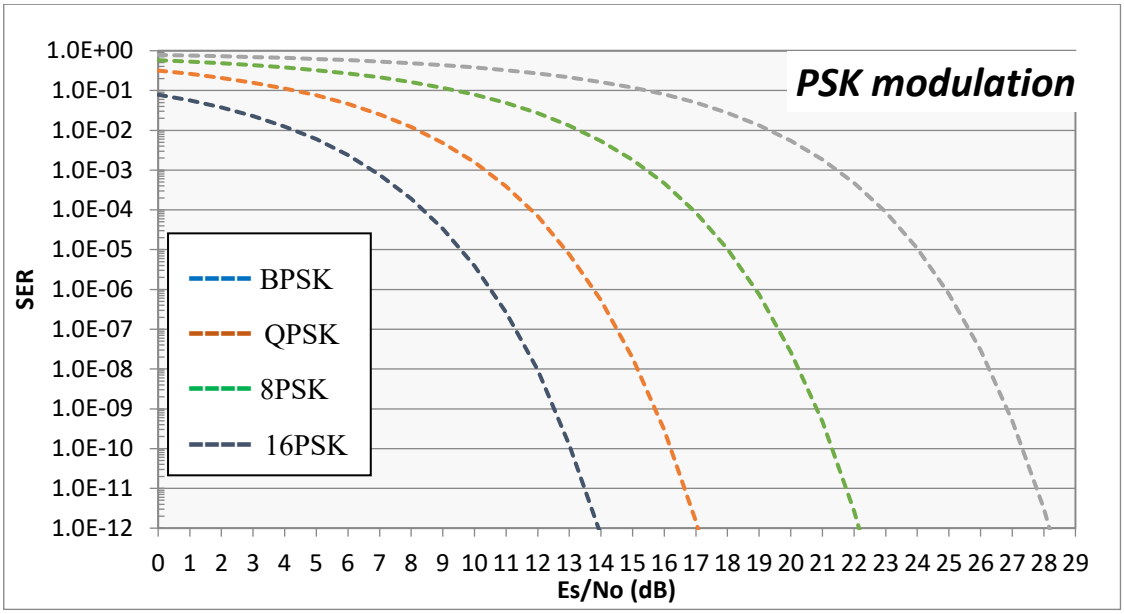
Figure (2.17) illustrates required SNR of MQAM compared to the QPSK and PAM modulations formats for a given BER. Calculations of QAM, PSK and PAM are based on the following eq. successively.

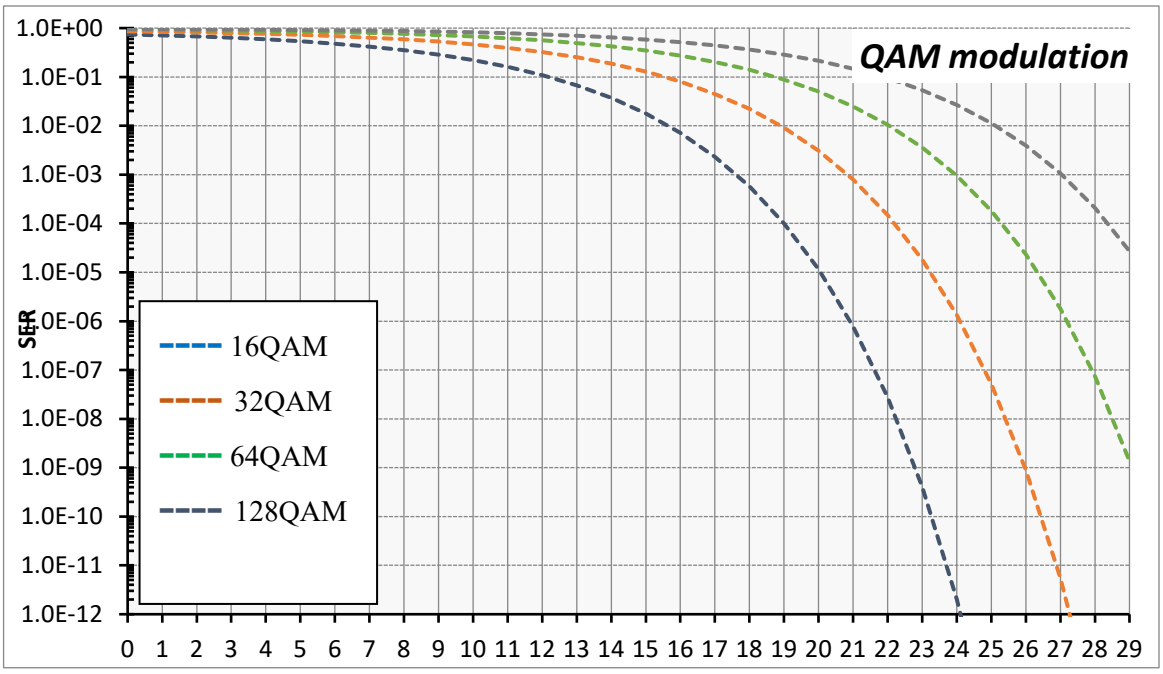
$$\mathrm{SER}_{QAM} = 2\left(1-\frac{1}{\sqrt{M}}\right) erfc\left(\sqrt{\frac{3Es}{2(M-1)No}}\right) - \left(1-\frac{2}{\sqrt{M}}+\frac{1}{M}\right) erfc^2\left(\sqrt{\frac{3Es}{2(M-1)No}}\right);$$

$$\mathrm{BER} = \frac{1}{log2(M)} \cdot SER \qquad \ldots\ldots\ldots (2.22)$$

$$SER_{PSK} = \text{erfc}\left(\sqrt{\frac{Es}{No}}\sin\left(\frac{\pi}{M}\right)\right) ; \text{BER} = \frac{1}{log_2(M)}.\ SER \qquad \ldots\ldots\ldots (2.23)$$

$$SER_{PAM} = \frac{M-1}{M}\cdot\text{erfc}\left(\sqrt{\frac{3Es}{(M2-1)\cdot No}}\right); \text{BER} = \frac{1}{log2(M)}\ .SER \qquad \ldots\ldots\ldots (2.24)$$





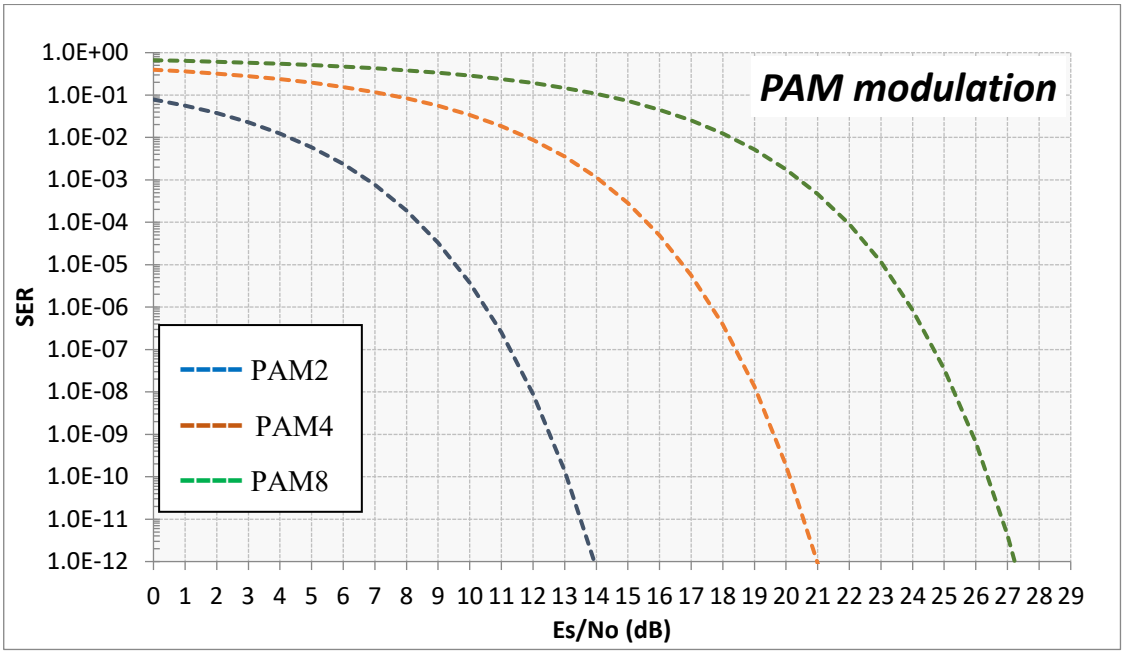


Figure (2.17): SER vs E_s/N_0 for QAM, PSK and PAM [3, 16].

2.6.1.2 Optical OFDM

The first OFDM principle was introduced by Chang [58]. OFDM has been the first standard of the Digital Audio Broadcasting (DAB) in 1995, and after that, it has become the most effective and significant scheme of modulation in more standards such as Wireless Local Area Networks (WLAN), Digital Video Broadcasting (DVB), wireless metropolitan area networks (WAN), and long-term evolution (LTE) that represents the 4-generation of the mobile communication technology. Due to benefits that occur when using the OFDM technique, such as, exact multi-parameter scanning provides the system designers with the ability of studying the tradeoffs in relation to the parameters of interest and to choose an optimal design for deployments. Also, it is a high-spectrum efficiency modulation technique able to eliminate InterSymbol interference (ISI). However, these advantages of OFDM in optical communications were not unknown until 2005 [59]. Figure (2.18) illustrates the basic optical OFDM block diagram of the transmitter and receiver system.

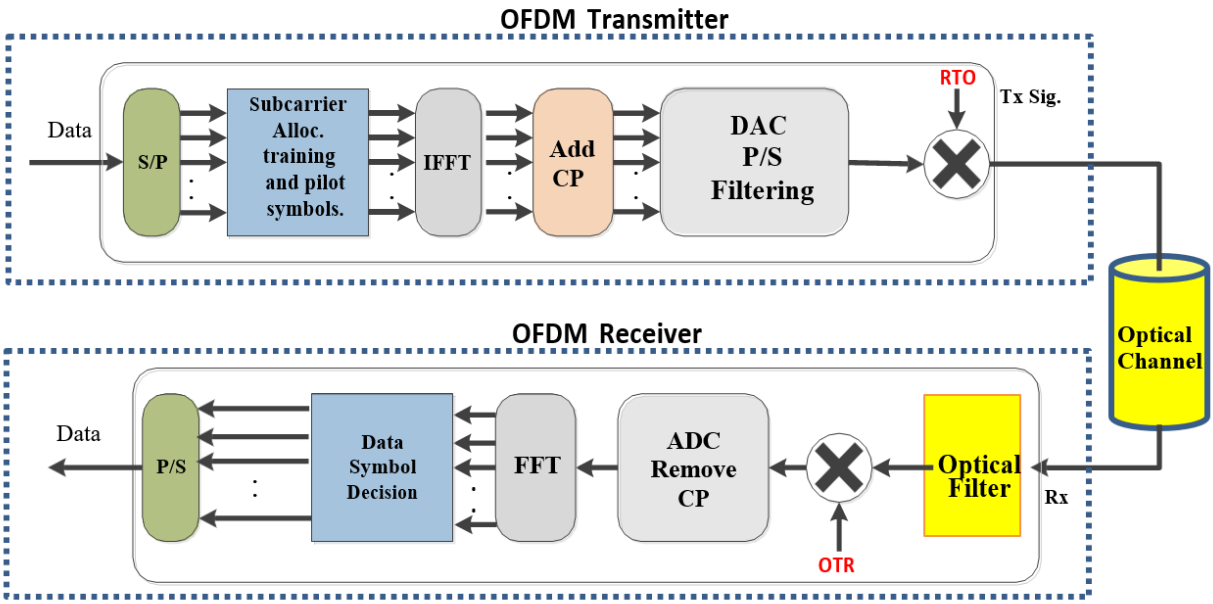


Figure (2.18): Basic Block Diagram of an Optical OFDM System [59].

So, depending on the RF to Optical Upconverter (RTO) technique the main interests of optical OFDM technique will be two main disciplines

1. Direct detection-Optical OFDM (DD-OOFDM), as shown in Figure (2.19).

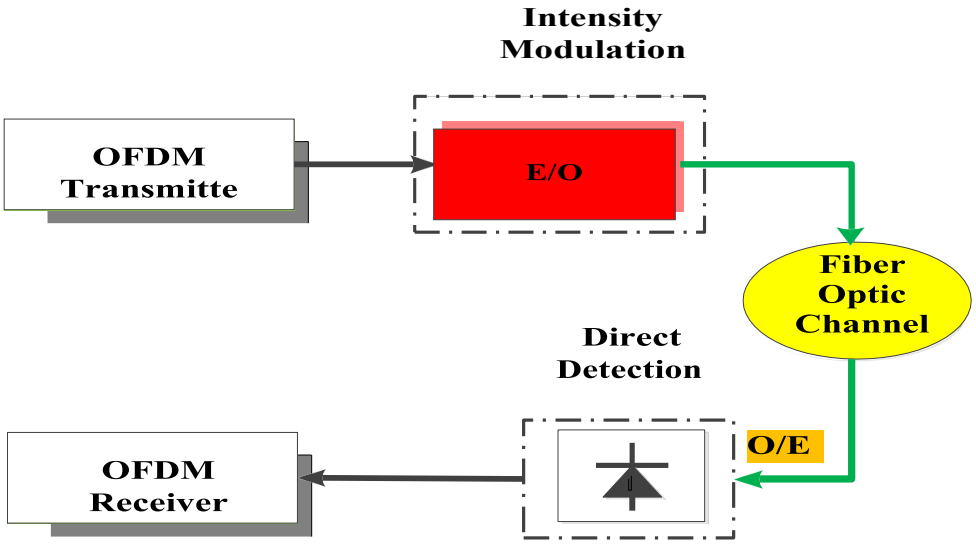


Figure (2.19): DD-OOFDM System Block Diagram [59].

2. Coherent Detection-Optical OFDM (CD-OFDM), as shown in Figure (2.20).

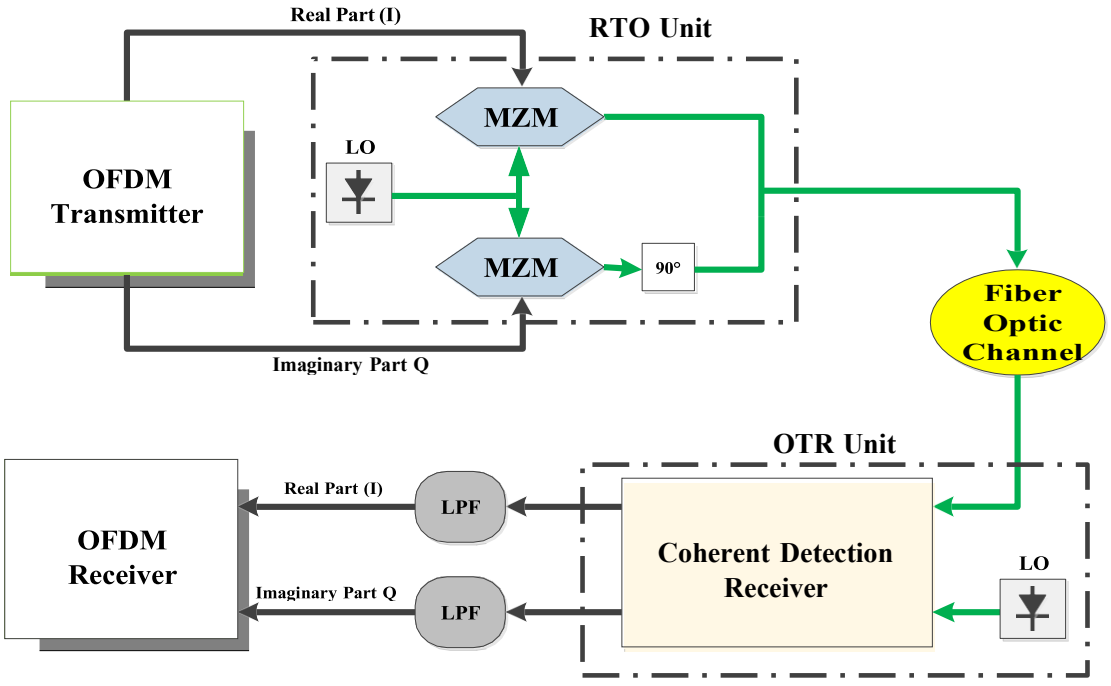


Figure (2.20): CD-OOFDM Block Diagram [59].

2.6.2 Wavelength Division Multiplexing (WDM)

WDM is a technique which enable of using various light wavelengths to transmitted data through the same fiber medium. Two or more of light wavelengths can be traveling over one fiber in the same time. Today, there are two types of WDM: Coarse WDM (CWDM) is defined with fewer than eight active wavelengths per fiber and used for short range transmission system. The other type is Dense WDM (DWDM), it defined with narrow wavelength spacing guard-band fits several channels over a single fiber optic and used for transmission system with more than eight active wavelengths per fiber [14]. Plus, the demands for high capacity, optical networks want to transfer for long distance to efficaciously support regional, metropolitan and wide area network applications. To achieve these requirements, optical transmission evolved from SMF single mode wavelength transmission system to multi-mode WDM system with flexible channel grids (FCG) and reconfigurable optical add/drop

multiplexers (ROADM). So, the average transmission capacity increased more than 30% per year over the last 30 years [14]. Figure (2.21) illustration the evolution of WDM capacity per fiber/carrier for the research as well as the commercial products. Introducing the next generation WDM communication systems to meet the growth of demands for high capacity are expected to operate at 32-64 Tb/s rate. Now, more technologies are the hot topic of research to study their potential implementation, in terms of costs and complexity and comes in the forefront using of AI algorithms.

The use of the multi-mode/core fibers optic had been suggested for the next WDM system generation [17] and produced the Space division multiplexing (SDM). The data rate will be increasing based on the number of modes/cores in fiber optic. In the recent years, some advances in this technique have been achieved principally for few (mode/core) fibers.

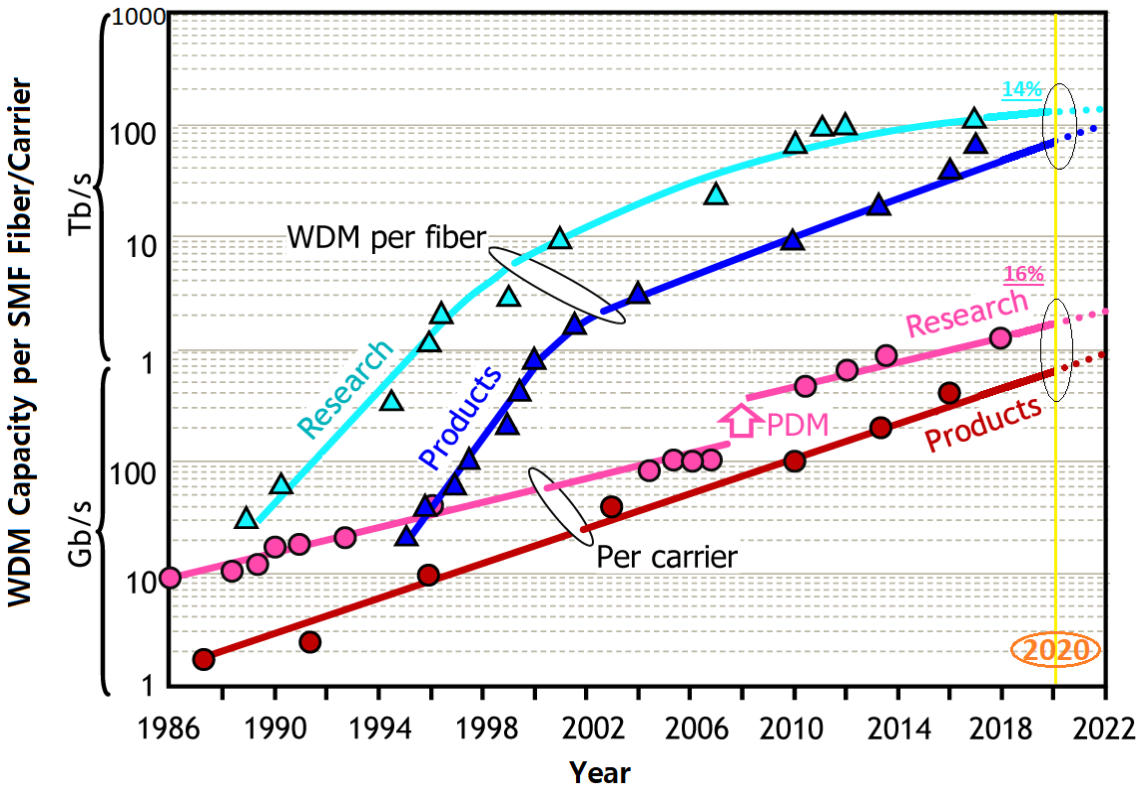


Figure (2.21): Illustration the evolution of the WDM capacity per fiber/carrier [15]

In addition, SDM still faces several challenges like the complexity of development of optical amplifier [15] and the practical implementation is very expensive because of needing for replace overall the already installed SMF by new MMF. Therefore, SMF will remain the technical of choice in the near future. So, the researchers now focusing on SSMF to increasing the transmission capacity and meet the growth of global traffic demands, so to achieve this goal used a multiplexing combined with the fiber non-linearity compensation (NLC) methods [60]. The principal of the super-channel approach is dividing the WDM channel into different sub-carriers with narrow wavelength and separated it by narrow spacing guard-band. the main advantages of super-channel are higher spectral efficiency, flexible to the architecture of the network, higher tolerance and lower cost. Furthermore, lower requirements for optical signal-to-noise ratio (OSNR) and digital-to-analog converters (DAC)/analog-to-digital converters (ADC) in comparison to single carrier mode [17]. In addition, there are two types of super-channel based on the orthogonality of subcarriers are: multiband (MB) orthogonal frequency-division multiplexing (MB-OFDM) as well as Nyquist WDM as shown in Figure (2.22).

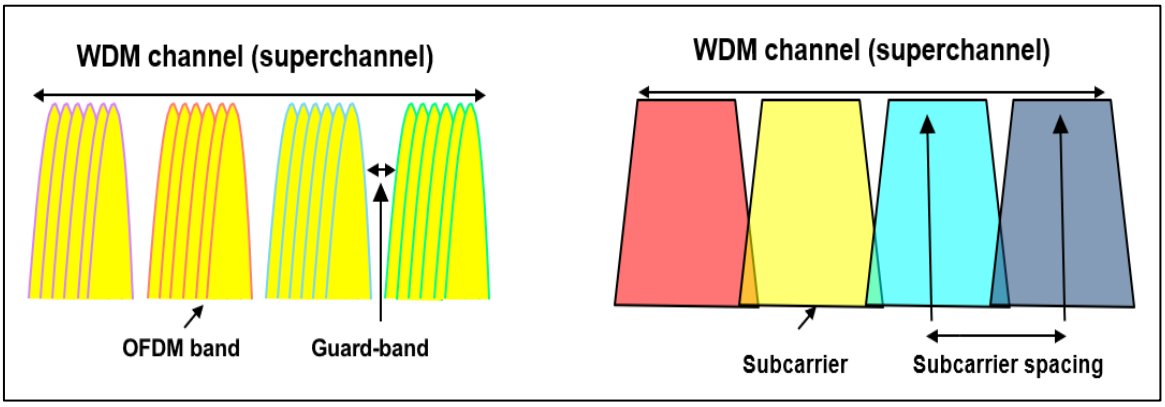


Figure (2.22): Super-channel WDM transmission spectrum [17]. (a): MB-OFDM. (b):Nyquist WDM

2.7 Optical System Performance Metrics

This section will present four main metrics for evaluating the optical communication system performance.

2.7.1 Signal-to-Noise Ratio (SNR)

Ratio between signal power and noise power in electrical domain within the signal band-width. It is a measure of the quality of a signal [41],

$$\text{SNR [dB]} = P_s/P_n, \tag{2.24}$$

Ps represents signal power and Pn is power of noise. SNR measured at the received part is called the Effective SNR, which comprises, noise contributions from the defects of the transmitter, noise from amplification, and non-linear effects which can be written as [41]:

$$\frac{1}{SNR_{eff}} = \frac{1}{SNR_{Tx}} + \frac{1}{SNR_{ASE}} + \frac{1}{SNR_{NLI}} \tag{2.25}$$

2.7.2 Optical Signal-to-Noise Ratio (OSNR)

The ratio of the signal to noise powers in optical domain in a 12.5GHz reference band-width. OSNR is in direct relation with electrical SNR [41],

$$\text{OSNR[dB]} = \text{SNR[dB]} + 10.(\log_{10}\frac{pB_e}{2B_0}) \tag{2.26}$$

B_e represents band-width of the electrical signal, p represents the number of the polarizations and B_o=12.5GHz represents reference band-width.

2.7.3 Bit Error Rate (BER)

The BER is a very beneficial measure for the description of the communication system quality. It simply represents the number of the error bits per transmitted bit sequence.

The main goal of the communication systems is keeping that value low at increasing bitrate value and keeping costs low at the same time. As can be seen from Equation (2.27), the exact calculations of BER is dependent upon the bit sequence's length, that means a simulation test with only (1000) bits sequence length will be have only a 0.10% likelihood of facing a 1-bit error. Considering that the simulation test results in a 1-error, BER will be incorrectly computed as 10^{-3}, which is obviously not the real BER. As a result, the longer the bit sequence, the higher the BER accuracy.

$$\text{BER} = \frac{\text{number of the bit errors}}{\text{bit sequence length}} \tag{2.27}$$

BER may be estimated as well by examining an eye chart for NRZ/RZ type scheme and a constellation chart can be used for advanced schemes.

2.7.4 Quality Factor (Q-Factor)

According to ITU-T [61], the measurement of the Q factor occupies an intermediate position between traditional optical parameters (which include the OSNR, power, and wavelength) and digital parameters based on BER. The Q factor is a comprehensive measure of optical channel quality in time domain through the analysis of optical pulse shape statistics and presents qualitative receiver performance description. The Q-factor is suggesting the minimal SNR that is needed for obtaining a certain value of the BER. A higher the bit rate, represents a higher required OSNR ratio. For example, the OC192 ((STM-64): 9953.28 Mbps) transmissions, OSNR has to be at least 27dB to 31dB in comparison with 18db to 21dB for OC-48 ((STM-16): 2488.32 Mbps) [61].

The mathematic relations to BER when the threshold is set to the optimum value as:

$$BER = \frac{1}{2} erfc\left(\frac{Q}{\sqrt{2}}\right) \cong \frac{1}{Q.\sqrt{2\pi}}\, e^{-\frac{Q^2}{2}} \tag{2.28}$$

$$erfc = \frac{1}{\sqrt{2}} \int_x^{\infty} e^{-\frac{\beta^2}{2}}\, d\beta \tag{2.29}$$

Q-factor can be written in terms of decibels as:

$$Q \text{ dB} = 20 \log_{10} Q \tag{2.30}$$

Chapter 3
Machine Learning (ML) Methods

3.1 ML Concepts

The applications of AI-based techniques to improve the performance of optical communication systems has recently received increased concern and has become a hot area under extensive research in the past years because of the difficulties when using the nonlinear equalizations and compensations techniques as well as the impressive results achieved when using AI [24].

Apart from traffic control, monitoring, and signal design, AI allows optical communication to have a flexible statistical analysis regarding the complicated systems with no dependence on certain models, and it has shown significant promise for increasing non-linearity compensation performance [19, 20, 49]. As shown in Figure (3.1), we briefly present a systematically comprehensive survey of the total AI approaches in terms of learning methods, statistical models, and decision-making methods which have been applied in many applications of optical transmission systems, the most important of which are analysis, statistics, and compensation for nonlinear disabilities to improve the overall system performance.

As a rule, ML was classified into three subfields, the first type is supervised: it works with the labeled dataset for training, the second type is unsupervised: this type uses an unlabeled dataset to generate a set of patterns known as features; and the last type is reinforcement: it describes a method for training and learning by rewarding itself for activities that prefer the desired result. The idea is to learn a black-box ML to be able to overcome situations where the use of computational theories is complex and intractable. A lot of ML algorithms which

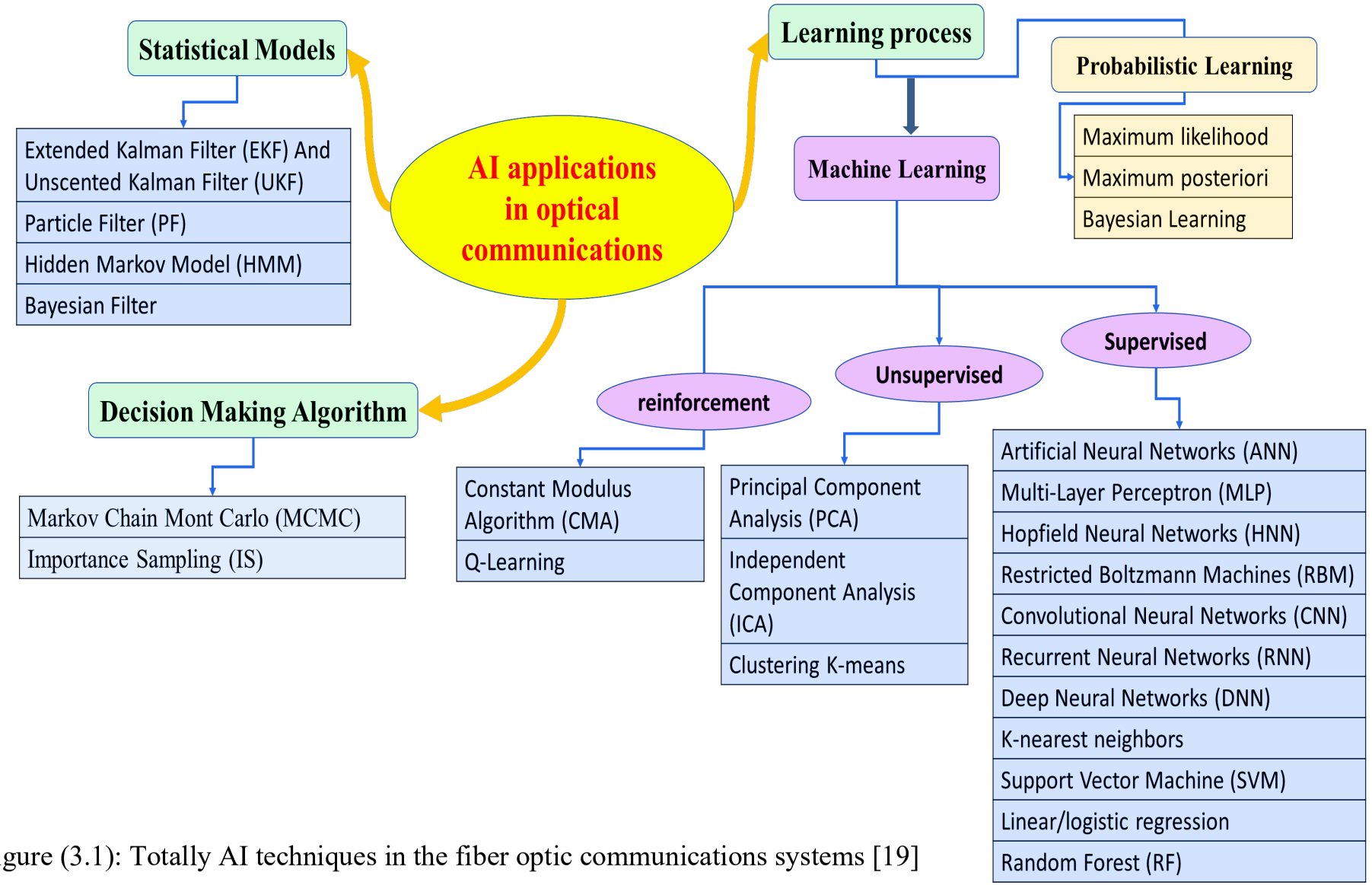


Figure (3.1): Totally AI techniques in the fiber optic communications systems [19]

exist and are widely used for optical communications systems as seen in Table (3.1), yet they didn't receive much attention as DNN due to the benchmark results achieved in DNN [21, 24]. Figure (3.2) show the path of ML and DNN applications in fiber optic communications system. In addition, Figure (3.3) illustrated only the key applications of ML in fiber-optic communications.

TABLE (3.1): ML that exists to optical communications [24].

Machine Learning	Applications
Kernel Density Estimation (KDE), Gaussian Mixture Model (GMM), Density Estimation Trees (DETs)	Linear impairment: CD, PMD laser amplitude and phase, carrier phase and polarization tracking and estimation
Principal Component Analysis (PCA) and Independent Component Analysis (ICA)	Nonlinearity: SPM, XPM, XPolM, FWM. OSNR monitoring, Modulation format recognition
Unscented Kalman's Filter (UKF), Particle Filters (PFs), Extended Kalman's Filter (EKF),	Polarization Recovery, Carrier Synchronization, Sym. Detection.
Neural Networks Techniques [(ANN), (MLP), (HNN), (CNN) and (RNN)]	phase recovery- nonlinear regression, classifies and identifies modulation formats.
Markov Chain MonteCarlo (MCMC), Importance Sampling.	Nonlinear equalization.
Support Vector Machine (SVM)	Nonlinear analysis, QoT estimator
Deep Learning (Deep Belief Networks (DBN))	Modulation design, OSNR monitoring, Phase recovery, nonlinear regression.

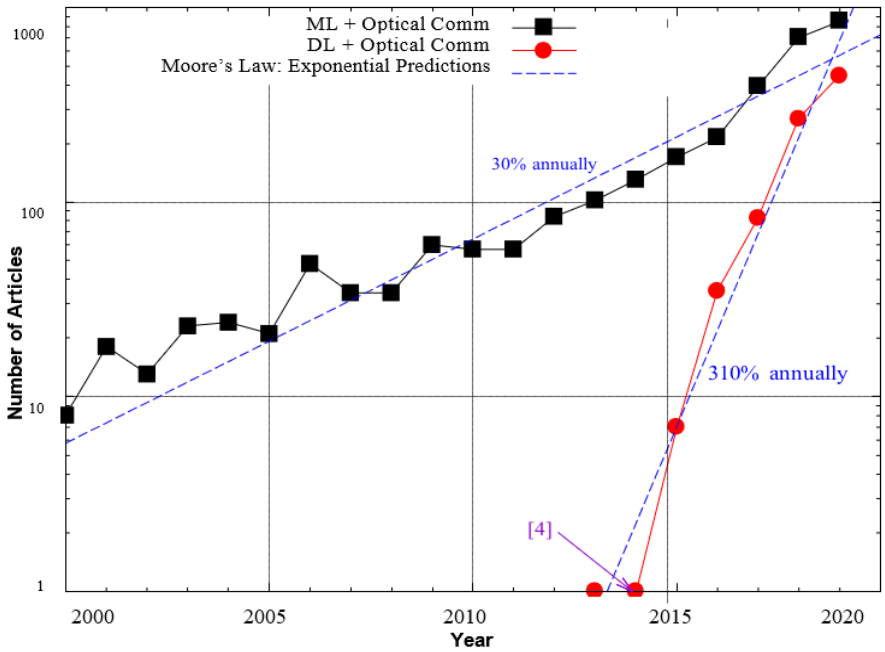


Figure (3.2): Machine learning /DL trend in the optic communication applications (key-word hits on GoogleScholar@2020) [24].

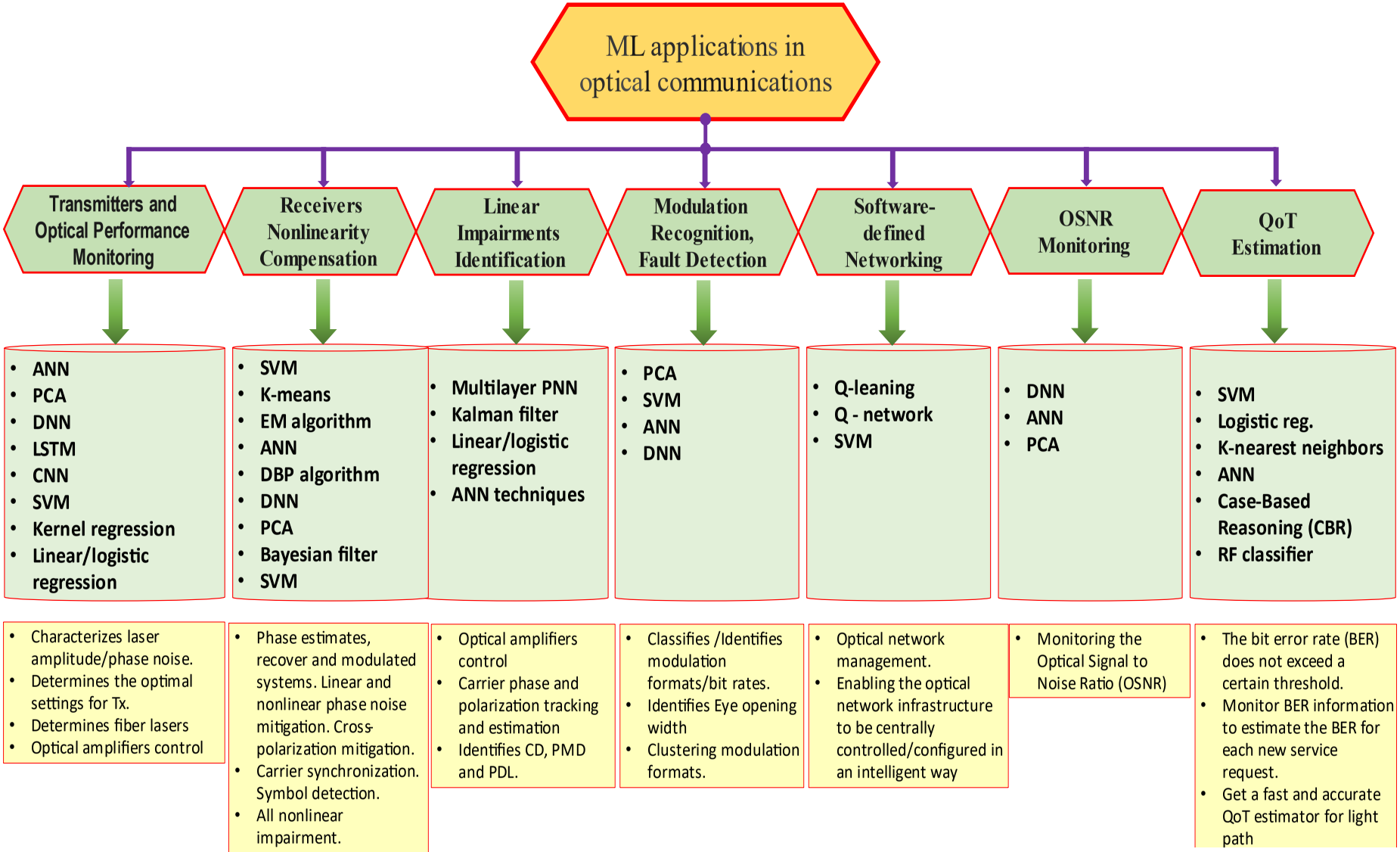


Figure (3.3): Key applications of ML in fiber-optic communications [19].

3.2 Learning Algorithms

ML is a branch of computer science which uses statistical approaches for enabling computers to learn from pieces of information or interactions with a virtual or real environment. It's usually divided to three sub-fields: Reinforcement, Unsupervised and Supervised [62].

3.2.1 Supervised learning

This type of the learning employs labeled data set for training the algorithm, which the model might then use to quickly provide the problem outcome. There are two sorts of problems that are addressed by supervised learning: regression problems and classification problems [62].

3.2.2 Unsupervised learning

The algorithm for unsupervised learning is the opposite of supervised learning. In unsupervised learning, which is self-organized learning that tries to explore and find a group of the descriptive patterns in data set that has been referred to as features, there is no clean and complete labelled data set. After that, only the features are utilized, and the original data set is ignored, yet the features of every one of the data are kept, which leads to a new data set with less degree of the redundancy. Unsupervised learning is used to solve two sorts of problems: clustering and association problems [62].

3.2.3 Reinforcement Learning

This type of the learning represents learning that is neither unsupervised nor supervised, here the algorithms learn to interact with the environment on their own by rewarding itself for actions that favor the desired result [62]. This

learning method is useful in the field of Robotics, Gaming [63]. Moreover, there are researchers who proposed this learning method in the field of telecommunication [64].

Figure (3.4) shows the basic structures of ML algorithms for unsupervised, supervised, and reinforcement learning algorithms.

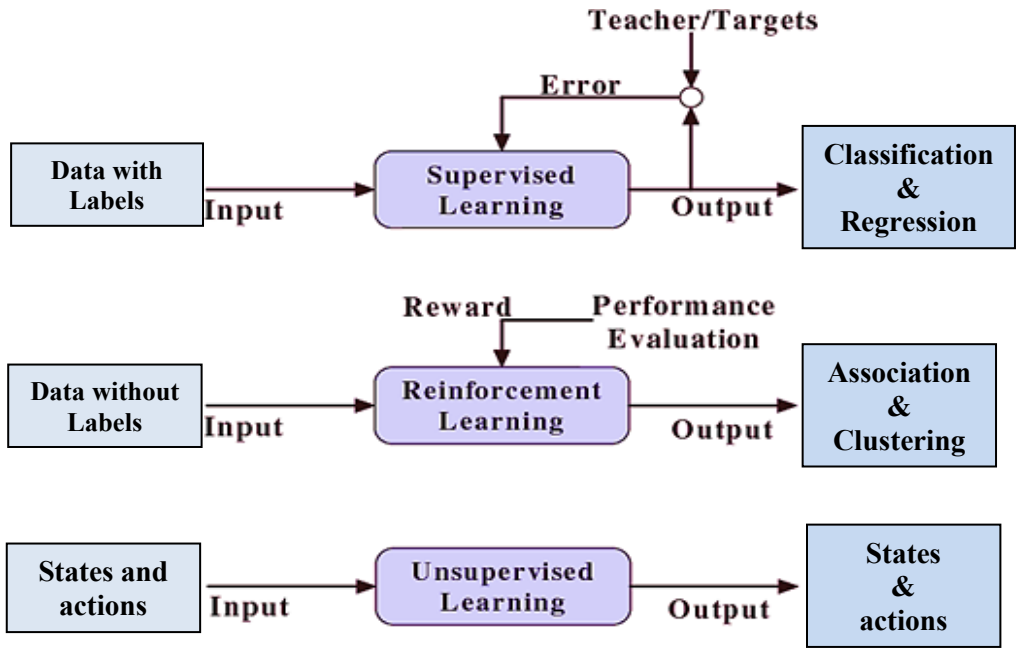


Figure (3.4): Basic structures of the ML algorithms: supervised, unsupervised and reinforcement [62].

3.3 Neural Network (NN)

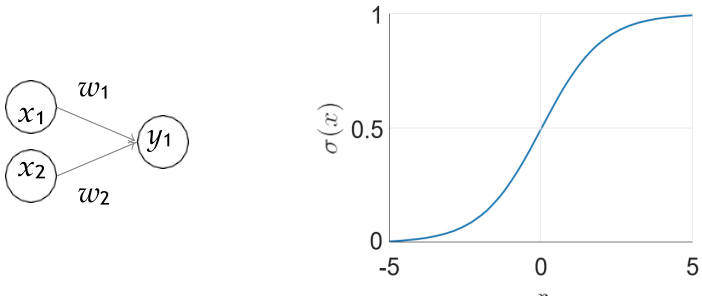
NNs are made up of many nodes, known as artificial neurons, that have several inputs, every one of them has been multiplied by the different weights to produce desired output [65]. The node's desired output is determined by the linear combination regarding the weighted inputs. The activation function specifies the desired output rule. An NN is created by connecting several nodes in a parallel manner. DNNs are created by linking multiple hidden layers in a series and using the outputs of previous layer as input for the following layer, resulting in a network that has a higher complexity. DNNs are suitable for very complex unknown relationships modeling [21, 24, 66].

3.3.1 Neural Network Structure

Figure (3.5) shows a single artificial neuron (node) with two inputs (x1 and x2). Each input is connected with weight (w1 & w2), a sigmoid activation function, and the output y_1, the output is related to the input in the following way [66]:

$$yn = \sigma(\sum_{i=n}^{N} w_i x_i) \tag{3.1a}$$

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{3.1b}$$



(a) (b)

Figure (3.5): (a) single neuron (b) sigmoid activation function.

A basic structure of NN is shown in Figure (3.6), with five nodes hidden layer, two nodes input layer, and two nodes output layer.

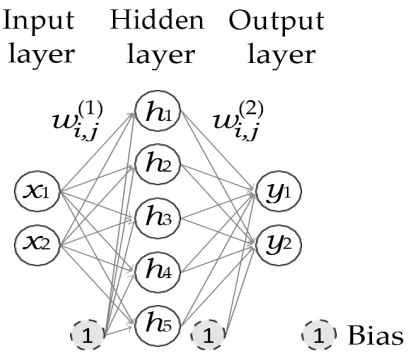


Figure (3.6): Basic neural network structure [66].

For example, the output is associated to the inputs for the system (shown in figure (3.6)) as follow:

$$h_i = \sigma\left(\sum_{j=1}^{2} x_j \, w_{i,j}^{1} + w_{i,0}^{1}\right) \tag{3.2a}$$

$$y_i = \left(\sum_{j=1}^{5} h_j \, w_{i,j}^{2} + w_{i,0}^{2}\right) \tag{3.2b}$$

This NN has a single hidden layer that includes five hidden nodes h_i, one input layer with two nodes x_i, and one output layer with two nodes y_i. The input layer is interconnected with hidden layer by weight values $w_{i,j}^{1}$ and hidden and output layers are connected by weight values $w_{i,j}^{2}$. The biases weights are $w_{i,0}^{1}$ and $w_{i,0}^{2}$, which are utilized for calculating the offset. Through the increase of the number of the hidden nodes, NNs have an ability for representing arbitrarily complicated functions. The computational complexity is determined by number of the hidden nodes, and NNs have been referred to as DNNs [65].

3.3.2 Training NNs

The neural network is supervised learning which means that it uses a labelled dataset as a group of the input vectors {x(n), n = 1,..., N}, to train an

algorithm and the model can learn from it to easily provide the corresponding result y(n) as group of the vectors that are referred to as the targets t(n). Training dataset samples are used to adjust the NNs weights, so that output vector y(n) will become comparable to target vector t(n) { y(n) ≈ t(n)} for each sample of the training dataset.

The first step towards this goal is to bring the error between the network outcomes and the intended target close to zero, and this is done by defining an appropriate "loss function". The regression and classification problems are the two basic categories of problems that are addressed by supervised learning algorithms. Cross-entropy (3.3a) is a common loss function for classification problems. MSE loss function (3.3b) is often suited for regression problems (3.3b) [65].

$$\boldsymbol{L_{cross-entropy}} = \frac{\boldsymbol{1}}{\boldsymbol{S}} \sum_{\boldsymbol{s=1}}^{\boldsymbol{S}} \left\{ - \sum_{\boldsymbol{o=1}}^{\boldsymbol{O}} \boldsymbol{t_o^s} \ \mathbf{log}(\boldsymbol{y_o^s}) \right\} \tag{3.3a}$$

$$\boldsymbol{L_{MSE}} = \frac{\boldsymbol{1}}{\boldsymbol{S}} \sum_{\boldsymbol{s=1}}^{\boldsymbol{S}} \left\{ \frac{\boldsymbol{1}}{\boldsymbol{2O}} \sum_{\boldsymbol{o=1}}^{\boldsymbol{O}} (\boldsymbol{t_o^s} - \ \boldsymbol{y_o^s})^{\boldsymbol{2}} \right\} \tag{3.3b}$$

where S represent the number of the samples in training dataset and *o* represent number of the output nodes. For classification tasks with a cross-entropy loss function, the loss function will minimize the loss for the given training samples and the output vector will be represented as a probability vector when training the NNs. This is usually achieved by optimizing the gradient ratio. With a gradient of the loss function, the current group of the weight values is updated arithmetically so that loss is reduced to the minimum. This method is called the backpropagation (BP) algorithm [65].

Furthermore, training samples are frequently divided to several of the batches, with loss function averaging over such batches and the entire sample of training still being utilized, yet the first weights update stage being

determined with first batch, the second weights are updated stage with 2nd batch, etc. Then, after all of the batches were utilized, the first batch is utilized again. The number of times NN was trained with all of the batches has been known as "epochs" and such scenario procedure has been known as the "stochastic gradient descent" (SGD) [67]. The weights are modified after many iterations with the weights update procedure, and the NN learns the relation in the training samples better. However, validations and tests are used to assess whether a trained NN is performing properly.

3.3.3 Tests and Validations of NN

The data-set has been divided to three sections: testing, training, and validation. By computing the loss, the validation and test groups have been utilized for estimating the performance of NNs [65].

3.3.4 Function of Activation

This function allows the NNs to learn nonlinear relationships between the input and output in the dataset though adding nonlinearity functions. There are many activation functions, but the most commonly used are the sigmoid and hyperbolic as showed in Figure (3.7) [65].

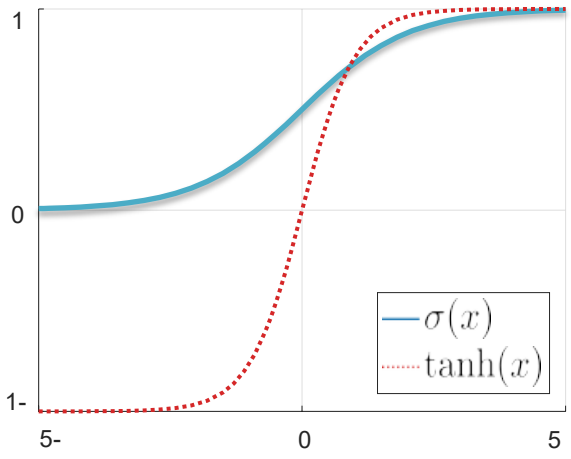


Figure (3.7): Hyperbolic and sigmoid functions of activation [65].

3.3.5 NNs Over-fitting and Regularization

Overfitting refers to a situation in which the learning algorithm's output matches the output of the training set so well that the learning algorithm might predict adequately despite the existence of noise in the training set alone. For example, if the training set's loss is reducing while the test set's loss is increasing, the NN is overfitting on the training set. Therefore, the NN model becomes extremely excellent at predicting or classifying data from the training set, yet not so good at classifying the data which has not been trained.

"Regularization" is considered as one of the most essential approaches to counter overfitting. The L2 and L1 regularization approaches [65] are based on the idea of lowering the complexity of the NN model through the addition of penalty term to loss function. L1 penalty aims at the reduction of weights' absolute values (3.4a), whereas L2 penalty aims at the reduction of weights' squared magnitudes (3.4b). The cost function (3.4c) is the sum of the loss function and a regularization penalty term, and it will be addressed in Section 3.5.

$$L_1\left(\theta = \left\{w_{i,j}^k\right\}\right) = \sum_{i,j,k} w_{i,j}^k \tag{3.4a}$$

$$L_2\left(\theta = \left\{w_{i,j}^k\right\}\right) = \sum_{i,j,k} \left|w_{i,j}^k\right|^2 \tag{3.4b}$$

$$C(\theta) = \ell(\theta) + \beta\, L(\theta) \tag{3.4c}$$

where β is a parameter that controls the regularization value and k is the layer number. Figure (3.8) illustrates the logistic sigmoid activation function with different parameters of (a), as see the sigmoid function becomes as a unit step function at $\{a \to \infty\}$.

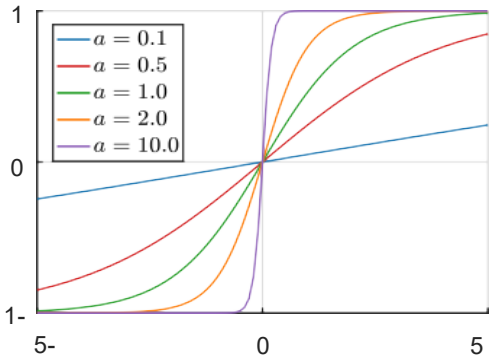


Figure (3.8): Logistic sigmoid activation function at different input scales [65].

3.4 Classification and Regression

NN is a supervised learning technique that may be used for dealing with two sorts of problems: regression and classification [62, 65]. This work focuses on the classification task that will be discussed and used in fiber optic communication system to compensation the fiber nonlinear impairments in the next Chapter.

NN is utilized for classifying the labeled dataset, for example, a QAM signal after transmission over the fiber channel, taking a training group (1,000) symbols that have been sampled evenly from QAM constellation. This data-set is made up of samples that were received and labels that assigned each sample to one of the four QAM symbol classes. The label is a 1-D vector for the QAM encoder. Figure (3.9) shows the example results from the NNs that are made up of a single input layer that has 2-nodes (in-phase (I)/quadrature (Q)), a single hidden layer with 28 nodes, and 1 output classification layer that comprises 4 nodes.

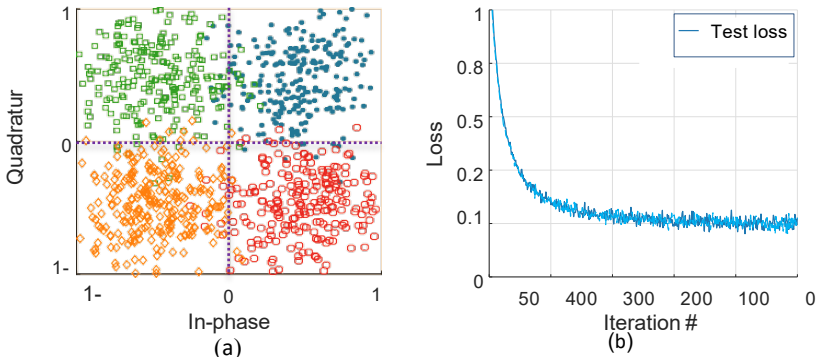


Figure (3.9): (a) Received QAM signal with four classes respective to QAM encoder symbol. (b) Loss function of the test set [65].

3.5 Deep Learning (DL)

Usually, DL means that there are multiple layers sequentially supposedly in the hidden layers. Generally, the NNs is called deep NNs if more than one hidden layer is used. The basic structure's design of a DNN is shown in Figure (3.10), where x represents nodes amount of an input layer, h represents nodes count of hidden layers and y denotes nodes amount of output layer. DNNs are convenient for modeling very complex unknown relationships and its capable of learning nonlinear equalization very efficiently where nonlinear equalization can be treated as a problem of supervised nonlinear in machine learning terms, so, the optical communication community embraces DNNs for a wide range of tasks same as other engineering and scientific disciplines [66, 68].

Many of the ideas that have influenced DL aren't connected to deep structures, yet rather help train them and enhance their performance [69]. As a result, the major approaches that contribute to DL and will be used in the following chapters will be presented in this part.

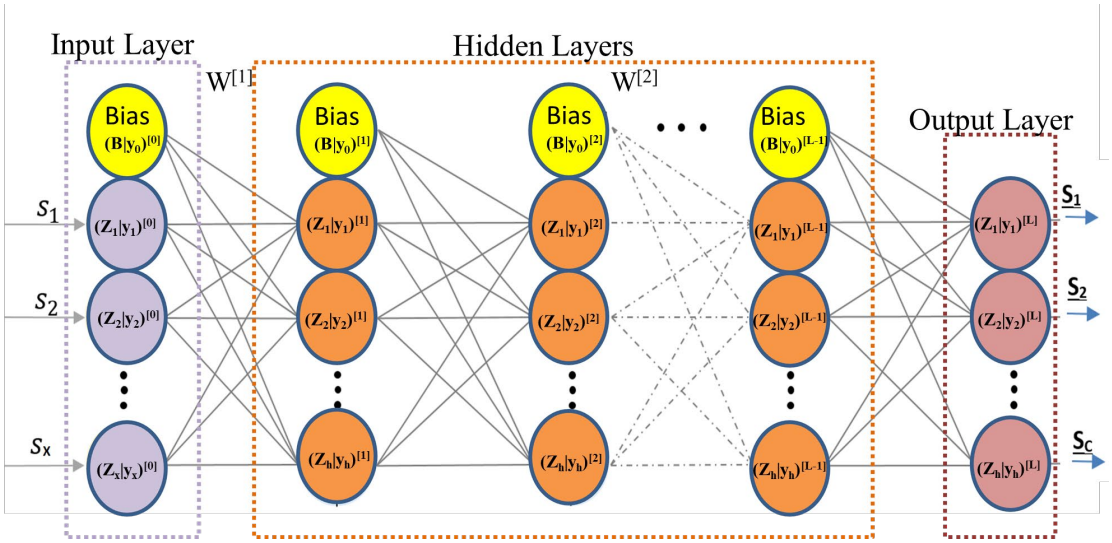


Figure (3.10): Basic structures of a DNNs [66].

Nonlinear activation function

The evanescence gradients problem in logistic hyperbolic and sigmoid tangent functions of activation causes the formation of the nonlinear activation function [66, 70, 71]. The gradients of activation function for every one of the nodes are given to previous layer then multiplied by gradients as in the backpropagation algorithm, this operation will be repeated a number of times and gradients will be accumulated in deep NNs architectures using the SGD algorithm. Due to the small final gradient, the network's training phase becomes slow and impractical. The leaky ReLU (LReLU) and ReLU activation functions have only two gradients, on the other hand Scaled Exponential Linear Unit (SELU) and Exponential Linear Unit (ELU) have more than two gradients as can be seen in Figure (3.11), preventing vanishing gradients and simplifying the training step.

3.5.2 Momentum

Momentum is an approach which introduces a memory effect by adding fraction of previous update vector to the newest gradient. This helps the training converge faster. Only advanced optimization techniques, like Adam optimization algorithm, are included in momentum [72].

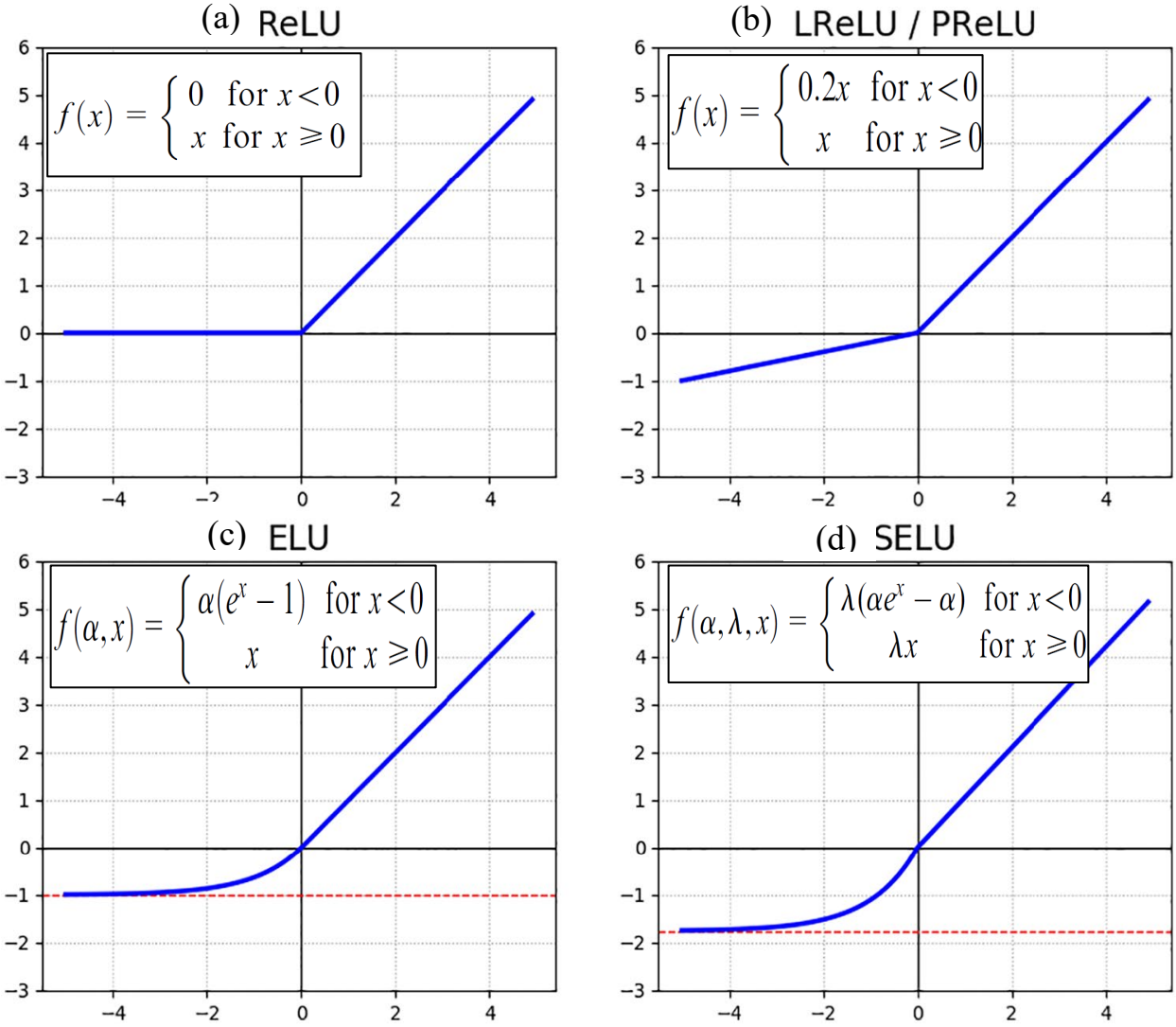


Figure (3.11): Plots of different activation functions. (a):ReLU, (b):LReLU, (c):Scaled Exponential Linear Unit (SELU) and (d):Exponential Linear Unit (ELU) functions of activation [71].

3.5.3 Automatic differentiation

Automatic differentiation can be defined as a numerical approach for computing the derivatives of numeric functions adequately [73], comparable to the backpropagation algorithm for differentiating NNs, yet more general. It iteratively determines the gradient of arbitrary deep mathematical structures using the chain rule, after that uses the gradient for optimization.

Chapter 4
Optical System Design

4.1 Introduction

The information presented in the previous three chapters will be used as a foundation in this chapter for the proposed design of NLC-DNNs with different types of optical communication systems, such as SP (QPSK, M-QAM, CD-OOFDM), DP (QPSK, M-QAM, CD-OOFDM) with and without the WDM technique. As well as a brief description of the main simulation tools and software that were used in this work.

4.2 Simulation Tools and Software.

1. MATLAB® (Ver. 2020a) from MathWorks, Inc. is a programming platform that has been specifically designed for the scientists as well as engineers for analyzing data, developing algorithms, and designing models, applications, and systems. MATLAB is a matrix-based language, allowing the most natural arithmetic mathematics expression.

2. OptiSystem (Ver.17.1.0) from Optiwave Systems Inc. is a programming platform that has been specifically designed for optical applications. It allows the optical telecom. engineers to plan, design, implement, simulate, and test with optimization a wide variety of optical applications within a modern optical system.

4.3 Single-Polarization Coherent System

The proposed Single-Polarization Coherent System is shows in Figure (4.1). Generally, the optical coherent system consists of three parts: transmitter with RF to optical unit (RTO) part, fiber-optic channel, and coherent receiver with optical to RF unit (OTR).

In the transmitter part, the length of the input data sequence has been set to 2^{16} and it has been built with Pseudorandom Binary Sequence (PRBS) for the generation of a bit sequence, fed after that into the advanced modulation system for producing M-Array sequence, which is serial-to-parallel converted as a two-dimensional (I & Q) complex signal each as blocks of data symbols comprising more than one bit for m-ary encoding. The CW generates a continuous wave modulated optical signal with 1550 nm laser source and the laser source power has been set to optimum lunching power according to [3] with thermal noise equal to $1x10^{-22}$ W/Hz, the responsivity is 1A/W, and dark current being 10 nA.

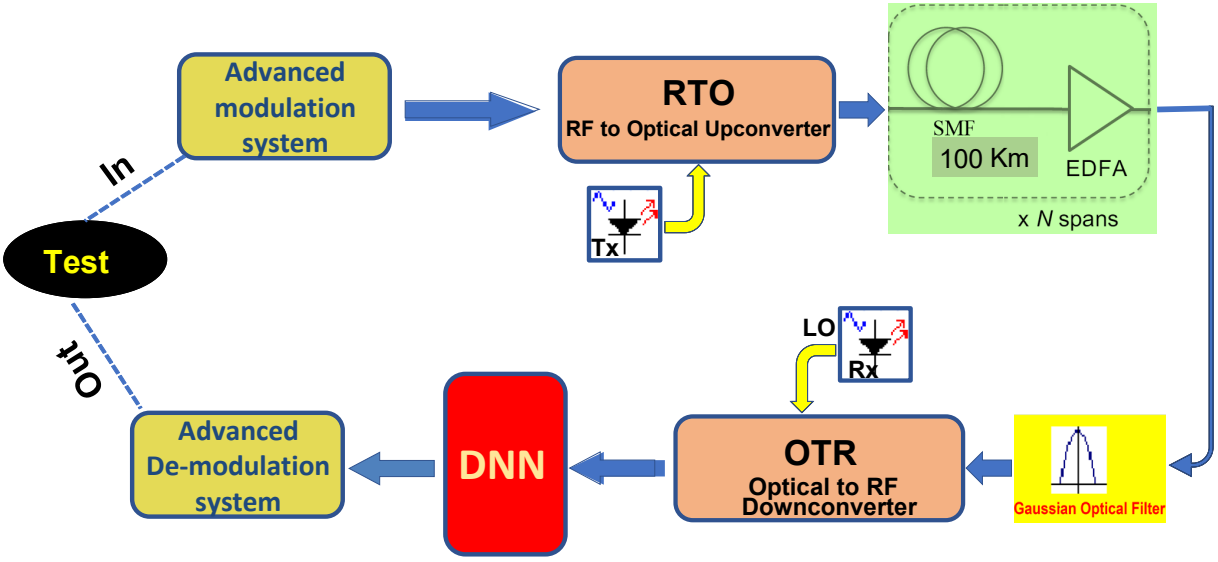
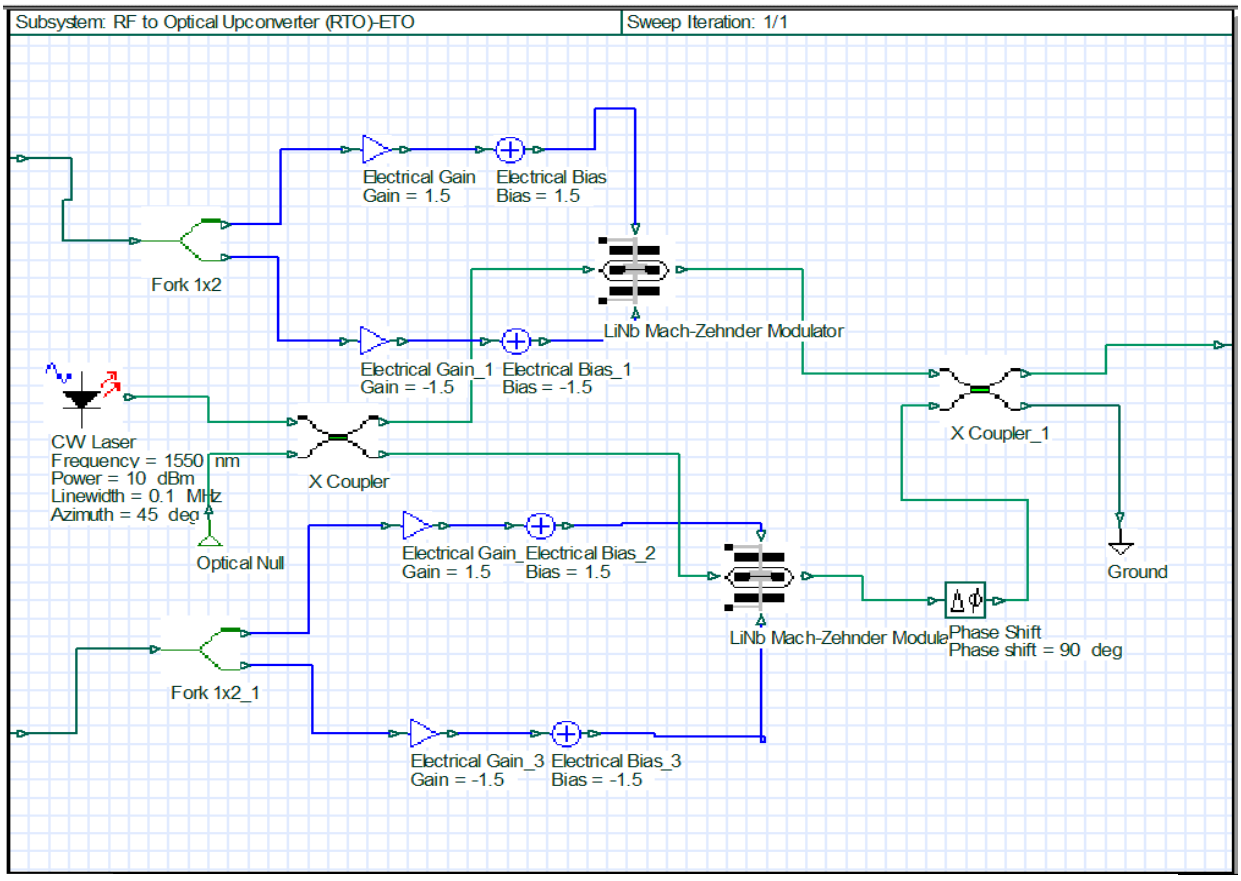


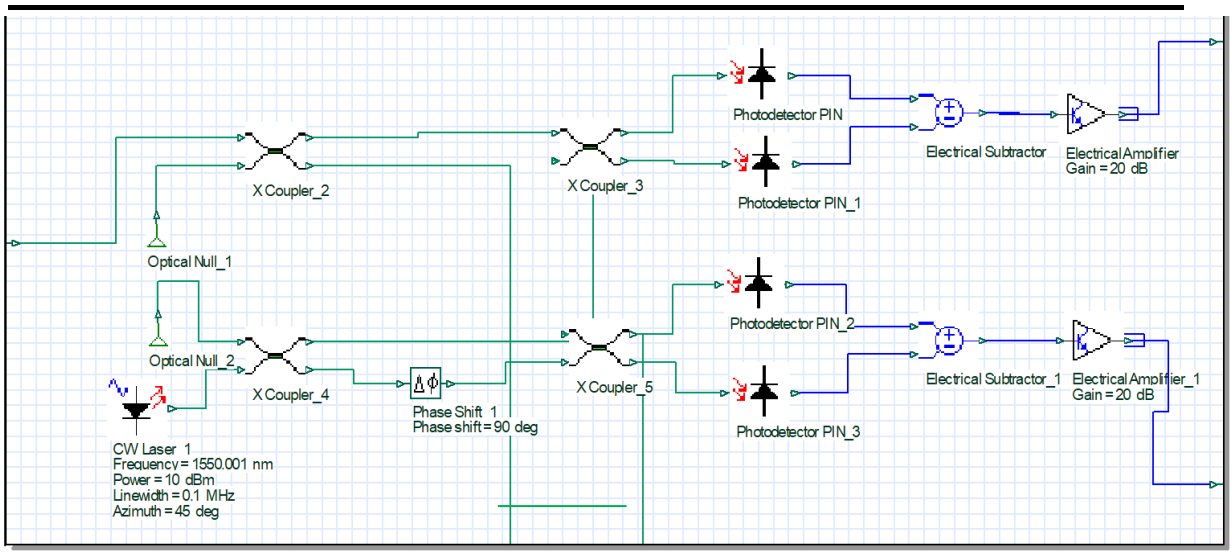
Figure (4.1): Block Diagram of Proposed SP-Coherent System

RTO unit consists of X-coupler, two dual-drive lithium Niubate Mach-Zehnder modulators (MZM), and an optical combining unit as shown in Figure (4.2 a), while, the OTR unit as shown in Figure (4.2 b) is consists of four X-couplers, one 90° phase shift, four PIN photodetectors with balanced noise detectors, and two electrical subtractors.

The work of the single-polarization coherent detection proposed system will be divided into three parts: transmission part, transmission media and receiver part as follows:



(a)



(b)

Figure (4.2): Simulation Setup for: (a) RTO, (b) OTR

Firstly, The transmission part: The input data sequence is built by using PRBS to generate a bit sequence with a length set to (2^{16}) and then, it is fed to M-QAM/QPSK sequence generator encoder in order to generate two parallel M-ary symbol sequences as in-phase (I) and quadrature (Q) components from

input binary signals. The output resulting signal from the M-ary encoder (in-phase (I)/quadrature (Q)) is fed to the RTO unit.

In the RTO unit, the Low Pass Cosine Roll off Filter has been utilized for each baseband M-ary (I/Q) signal component. After using electrical gain, each baseband (I/Q) signal component is split into two parts by using Fork device (1x2) to copies the electrical input signal into two output electrical signals which will be fed to electrical input ports of the MZM, in addition, the optical input port of the MZM will connect on CW Leaser. Finally, the optical output signals of the both MZMs are combined by a Power Combiner (2x1) device to combine the two (I/Q) MZM optical signals into the complex optical signal with amplified it to prepare for transmitted.

Secondly, the fiber optic transmission media: with main parameters as shown in Table (4,1).

Thirdly, the receiver part: The received optical signal is detected by the OTR unit with balanced noise detectors with the setting parameters shown in Table (4.2).

In the OTR unit, for the purpose of extracting components (I) & (Q) separately, the first X-Coupler has been utilized in order to divide the incoming received optical signal to two parts. In the same way, the 2^{nd} X-coupler has been utilized in order to divide LO signal into two parts with 90^0 phase shift in order to extract the optical modulated signal, and the output of 2nd X-coupler is mix with the output of the first X-coupler (optical data signal in branches (I) and (Q)) by using two pair X-coupler (3rd /4th). The outputs from the 3rd and 4th X-coupler are feeding to balanced detector PDs for the purpose of generating (I/Q) components of a base-band electrical M-ary signal after the subtraction of photo-currents output. Finally, the OTR unit output is (I/Q) signal has been fed to proposed DNN-ML block to compensation the optical nonlinear impairment. After that, to the mapper decoder to regenerates the transmitted binary data signal.

Table (4.1): Fiber Optic Link- Main Parameters Settings

Parameter	Value
Reference wavelength (λ0)	1550nm
Link distance (L)	3,000 km (30x100)
Attenuation (α)	0.2 dB/km
Dispersion (D)&GVD	20ps/(nm.km)
PMD coefficient	$0.50 ps/(km)^{-0.5}$
Nonlinear Factor	$2 m^2/W$
Differential group delay	0.20ps/km
β2 & β3	$-20 ps^2/km$, $-20 ps^3/km$
Slope of Dispersion	$0.075 ps/(nm^2.km)$
Birefringence type	Deterministic
Self-phase modulation	True
Effective area (Aeff)	$80 \mu m^2$
N2 (nonlinear index)	$2.6 X 10^{-20} m^2/W$
Nonlinear phase shift	5 mrad
Nonlinear propagator	Exponential
Raman self-shift time	15 fs
Nonlinear Polarization Fractional	0.2
Nonlinear Polarization Orthogonal	0.75

Table (4.2): RTO & OTR Main Parameters Settings

RTO	
Parameter	**Value**
RTO Format	Mach-Zehnder (MZ) modulator
MZ Extinction ratio	20dB
MZ Loss	5dB
MZ Modulation voltage	(2, -2) volt
MZ Bias voltage	0 volt
Optical Source	CW Laser
Wavelength	1550nm
Reference wavelength (λ0)	1550nm

CW Laser Linewidth	0.1 MHz
CW Laser Power	Sweep (0-30) dBm, setup to optimum
CW Laser initial phase	0
CW Laser polarization type	Azimuth
CW Laser polarization angle	45^0
OTR	
Parameter	**Value**
OTR Format	Photo Detector (PIN)
PIN-Responsivity	1 A/W
Optical & Electrical Amplifier	20dB/10dB
PIN-Thermal noise	1×10^{-22}W/Hz
PIN-Dark Current	10nA
Transmitter Optical Output Power (Pt)	Sweep (0-30) dB, setup to optimum
SNR	≤ (0-30) dB [74], setup to optimum
LO	CW Laser

4.3.1 M-QAM System

Quadrature amplitude modulation (QAM) is used to generate two parallel M-ary symbol sequences (which are I & Q) from the input binary signals. The proposed M-QAM system is described in Figures (4.3) with main parameters settings as shown in Table (4.3).

Table (4.3): M-QAM Main Parameters Settings

Parameter Name	Value	Description
Bit rate	60 Gbps	High bit rate
Symbol rate	Bit rate/h GBd	h= no. of bit/symbol
Mapper	4,16,64 QAM	Square constellation map
Sequence length	2^16 bits	Long seq. length gives accurate result
Samples per bit	4-8-16 Sample/bit	Sweep no. of sample/bit to gives accurate result
Guard Bits	100 bits	defines the number of bits to ignore before and after a sequence of bits
FEC	Gray code	
I-Q mapping	User-defined	Define the (I&Q) position to use as target and classes in ML

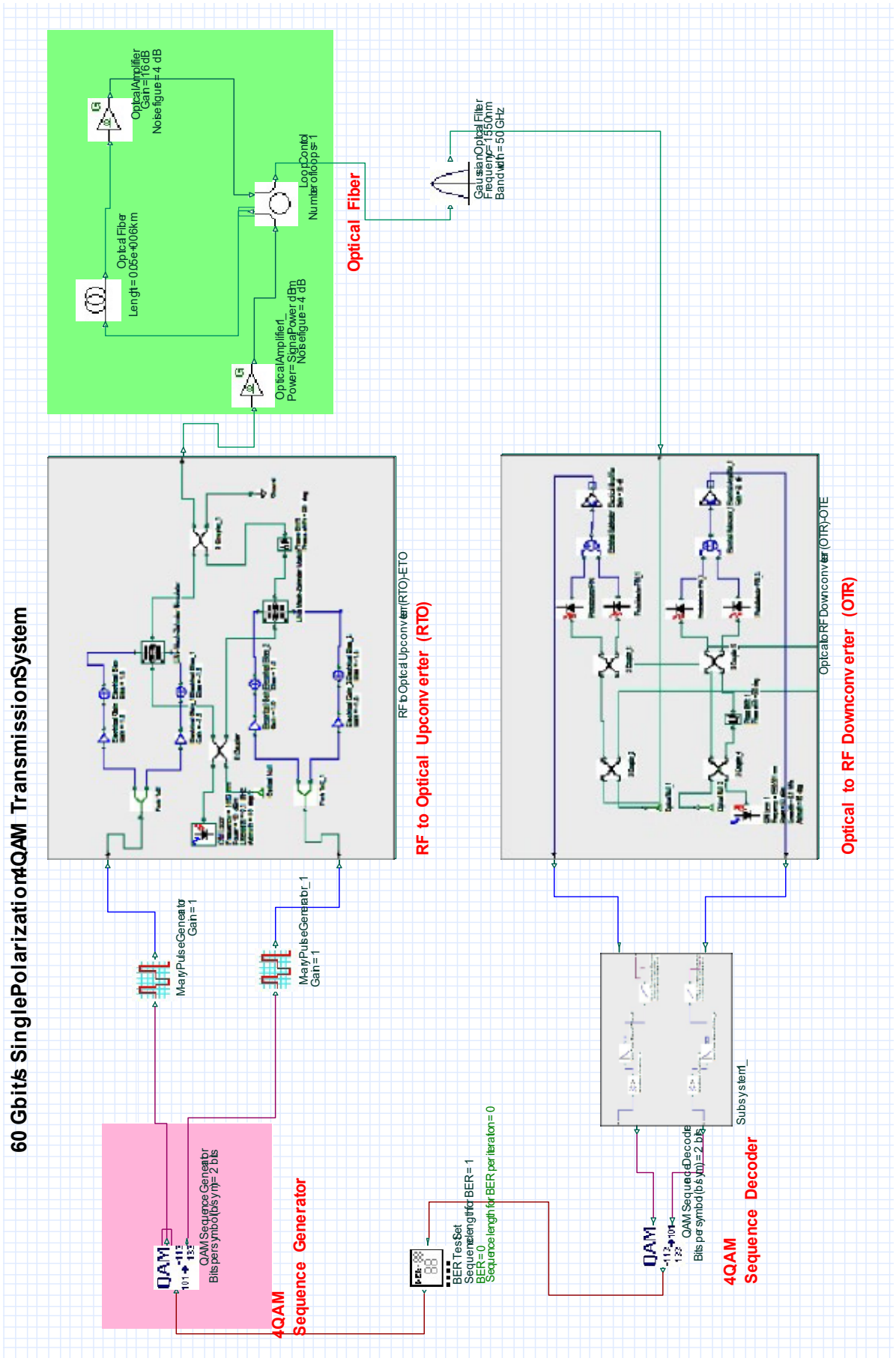


Figure (4.3): The Simulation Setup for the Proposed M-QAM System.

4.3.1.1 Proposed M-QAM IQ Maps and Targets

With QAM sequence generator, bit sequence has been split to two parallel sub-sequences, every one of which may be transmitted in two quadrature carriers in the case of building QAM modulator. Which has been accomplished with the use of serial to parallel converter. In the square QAM maps, can vary signal amplitude based on source symbols in the case of information transmission. For every port of the output, amplitude can take a value from the group of the amplitude values [3] as follow:

$$a_1 = (2i - 1 - M), i = 1,2, \ldots., M$$
$$M = 2^{h/2} \quad (4.1)$$

where M represents number of the potential binary digit sequences and h represents number of the bits for each symbol.

The proposed 4-16-64 QAM (I-Q) map and Classes are shows in Tables (4.4), (4.5) and (4.6) respectively. In addition, the proposed reference target and classes of (4-16-64) QAM is shown in Figures (4.4), (4.5) and (4.6) respectively

Table (4.4): 4-QAM I-Q map and Classes

Sequence	I	Q	Class
00	-1	-1	1
01	-1	1	2
10	1	-1	3
11	1	1	4

Table (4.5): 16-QAM I-Q map and Classes

Sequence	I	Q	Class
0000	-3	-3	1
0001	-3	-1	2
0010	-3	1	3
0011	-3	3	4
0100	-1	-3	5
0101	-1	-1	6
0110	-1	1	7
0111	-1	3	8
1000	1	-3	9

1001	1	-1	10
1010	1	1	11
1011	1	3	12
1100	3	-3	13
1101	3	-1	14
1110	3	1	15
1111	3	3	16

Table (4.6): 64-QAM I-Q map and Classes

Sequence	I	Q	Class	Sequence	I	Q	Class
000000	-7	-7	1	100000	1	-7	33
000001	-7	-5	2	100001	1	-5	34
000010	-7	-3	3	100010	1	-3	35
000011	-7	-1	4	100011	1	-1	36
000100	-7	1	5	100100	1	1	37
000101	-7	3	6	100101	1	3	38
000110	-7	5	7	100110	1	5	39
000111	-7	7	8	100111	1	7	40
001000	-5	-7	9	101000	3	-7	41
001001	-5	-5	10	101001	3	-5	42
001010	-5	-3	11	101010	3	-3	43
001011	-5	-1	12	101011	3	-1	44
001100	-5	1	13	101100	3	1	45
001101	-5	3	14	101101	3	3	46
001110	-5	5	15	101110	3	5	47
001111	-5	7	16	101111	3	7	48
010000	-3	-7	17	110000	5	-7	49
010001	-3	-5	18	110001	5	-5	50
010010	-3	-3	19	110010	5	-3	51
010011	-3	-1	20	110011	5	-1	52
010100	-3	1	21	110100	5	1	53
010101	-3	3	22	110101	5	3	54
010110	-3	5	23	110110	5	5	55
010111	-3	7	24	110111	5	7	56
011000	-1	-7	25	111000	7	-7	57
011001	-1	-5	26	111001	7	-5	58
011010	-1	-3	27	111010	7	-3	59
011011	-1	-1	28	111011	7	-1	60
011100	-1	1	29	111100	7	1	61
011101	-1	3	30	111101	7	3	62
011110	-1	5	31	111110	7	5	63
011111	-1	7	32	111111	7	7	64

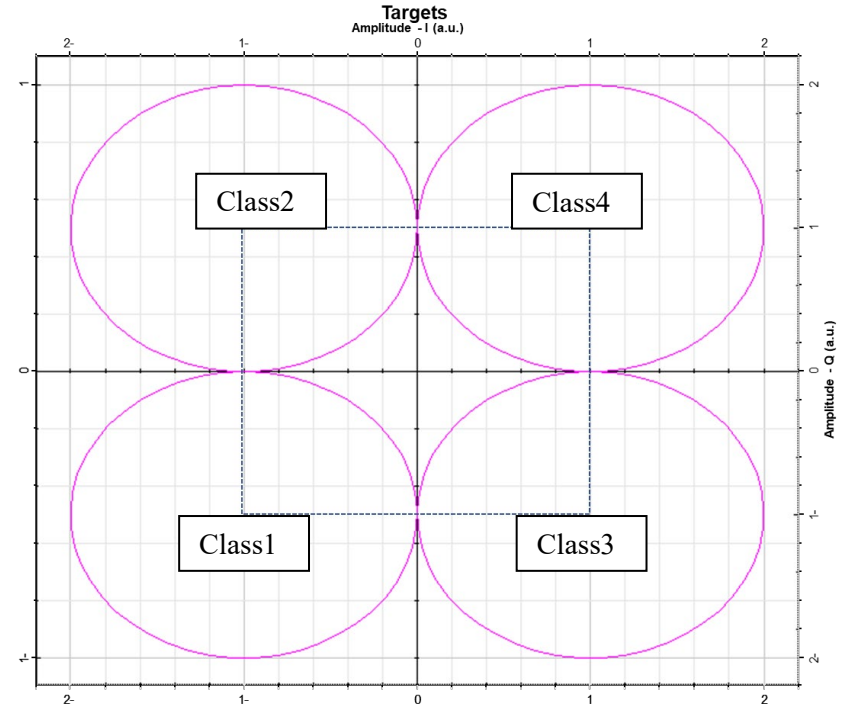


Figure (4.4): 4QAM Reference Target and Classes

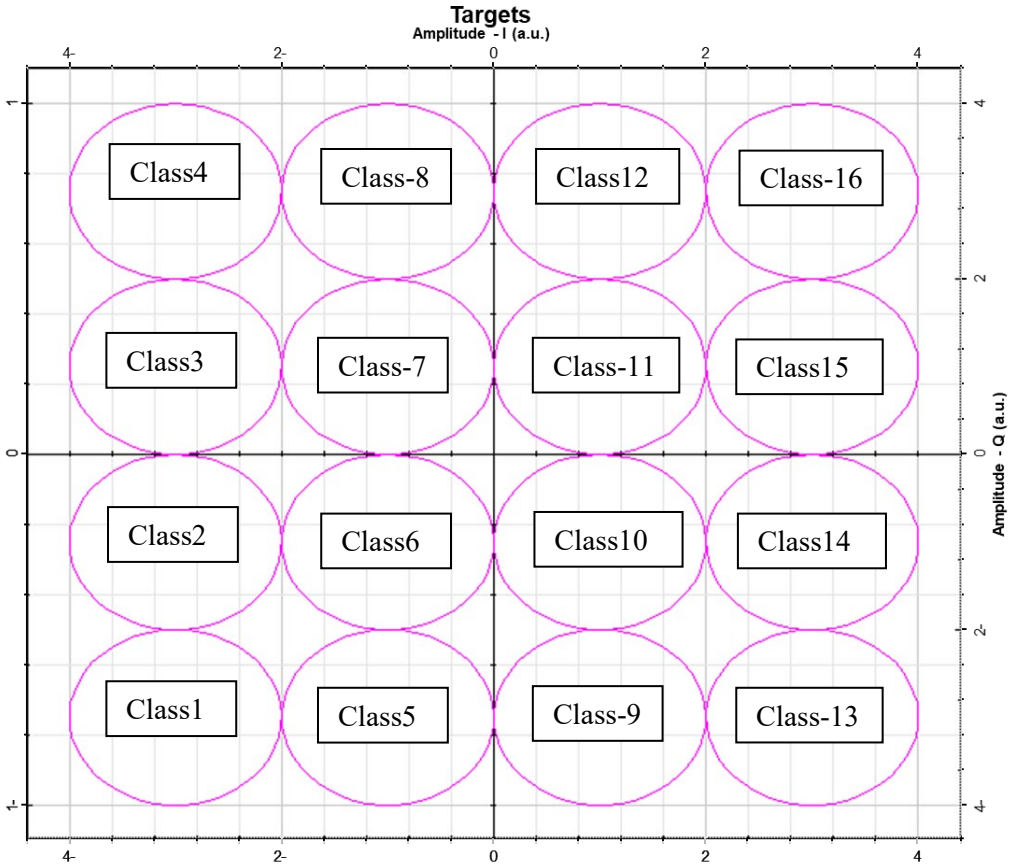


Figure (4.5): 16QAM Reference Target and Classes

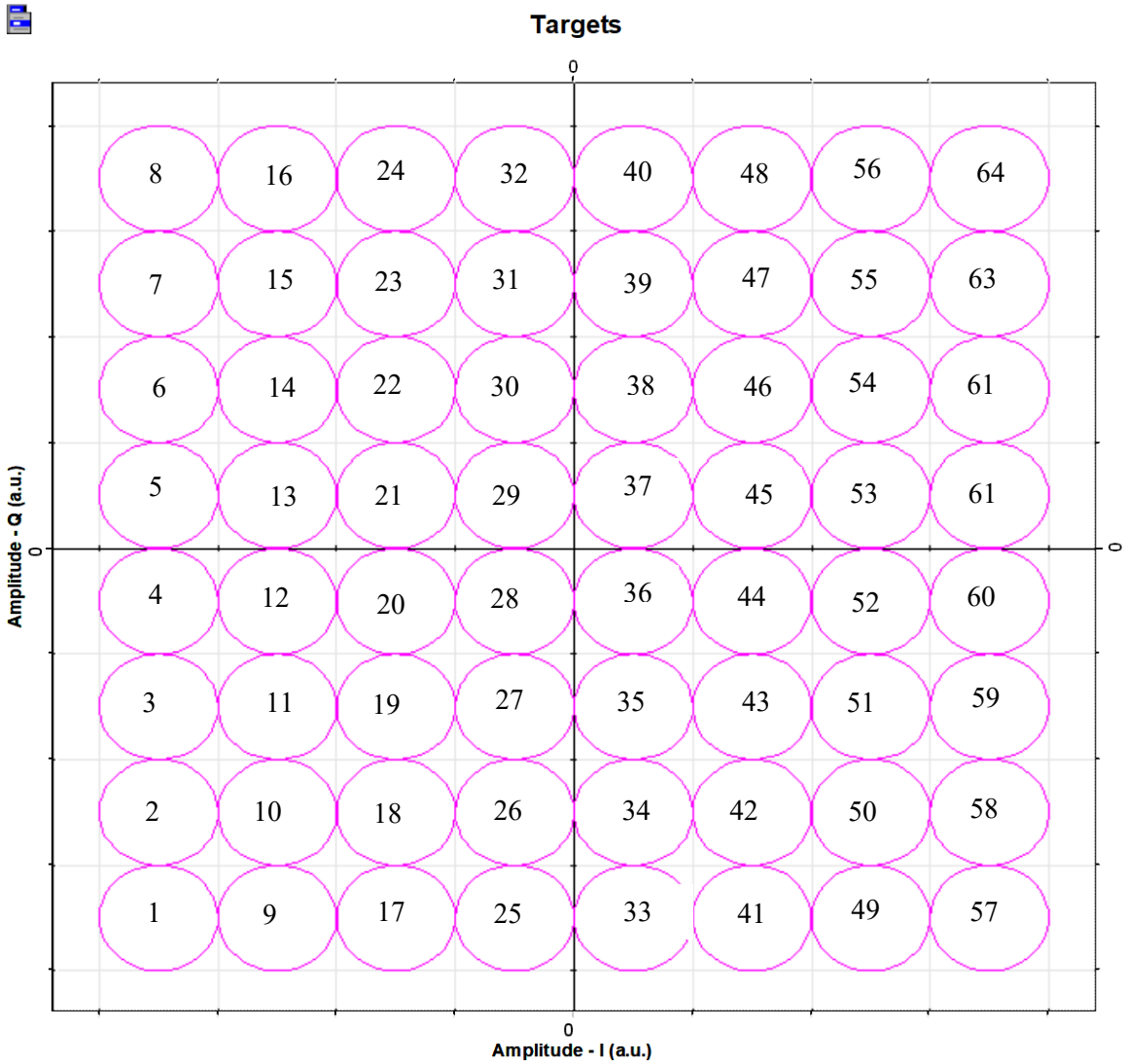


Figure (4.6): 64QAM Reference Target and Classes

4.3.2 QPSK Proposed System

Quadrature phase shift keying modulation (PSK) is used to generate two parallel M-ary symbol sequences from the signals of binary inputs. In the case of information transmission, signal phase can be varied based on source symbols. Phase values have been taken from a collection of the angles, as follow:

$$\varphi_i = \left(\frac{2\pi}{M}\ (i-1) + \ \theta\right), i = 1, 2, \ldots . M \qquad (4.2)$$

M represents number of the possible binary digit sequence, which has been estimated based on: $M= 2^h$, where h represents the number of the bits in each

symbol, and θ represents phase off-set. The in-phase and quadrature-channel will be having amplitudes based on:

$$I_i = \cos(\varphi_i), Q_i = \sin(\varphi_i) \qquad (4.3)$$

The proposed QPSK parameters is shows in Tables (4.7), also the QPSK (I-Q) map and Classes is shows in Tables (4.8). In addition, the proposed reference target and classes of QPSK is shown in Figure (4.7).

Table (4.7): QPSK Main Parameters.

Description	Default value	Comment
Bits per symbol	2 bits (b/sym)	Num of bits for each symbol, used in the coding
Phase off-set	45 deg, rad	Initial phase offset
FEC	Gray Code	Use gray coding
I-Q mapping	User-defined	I&Q position to use as target in ML

Table (4.8): QPSK I-Q map and Classes

Sequence	I	Q	Class
00	$1/\sqrt{2}$	0	1
01	0	$1/\sqrt{2}$	2
10	$-1/\sqrt{2}$	0	3
11	0	$-1/\sqrt{2}$	4

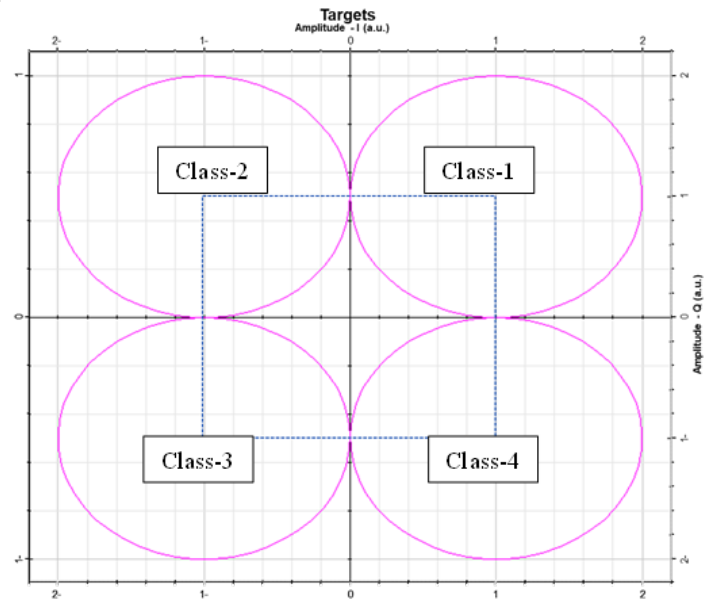


Figure (4.7): QPSK Reference Target and Classes

The proposed QPSK is shown in Figure (4.8). Note that, all system parameters are setting as presented in Tables (4.1) and (4.2). Here, the values that were explained only for the QPSK mapper. In addition, the OSNR is set to the optimum value as illustrated in Section (2.6.1) Figure (2.17).

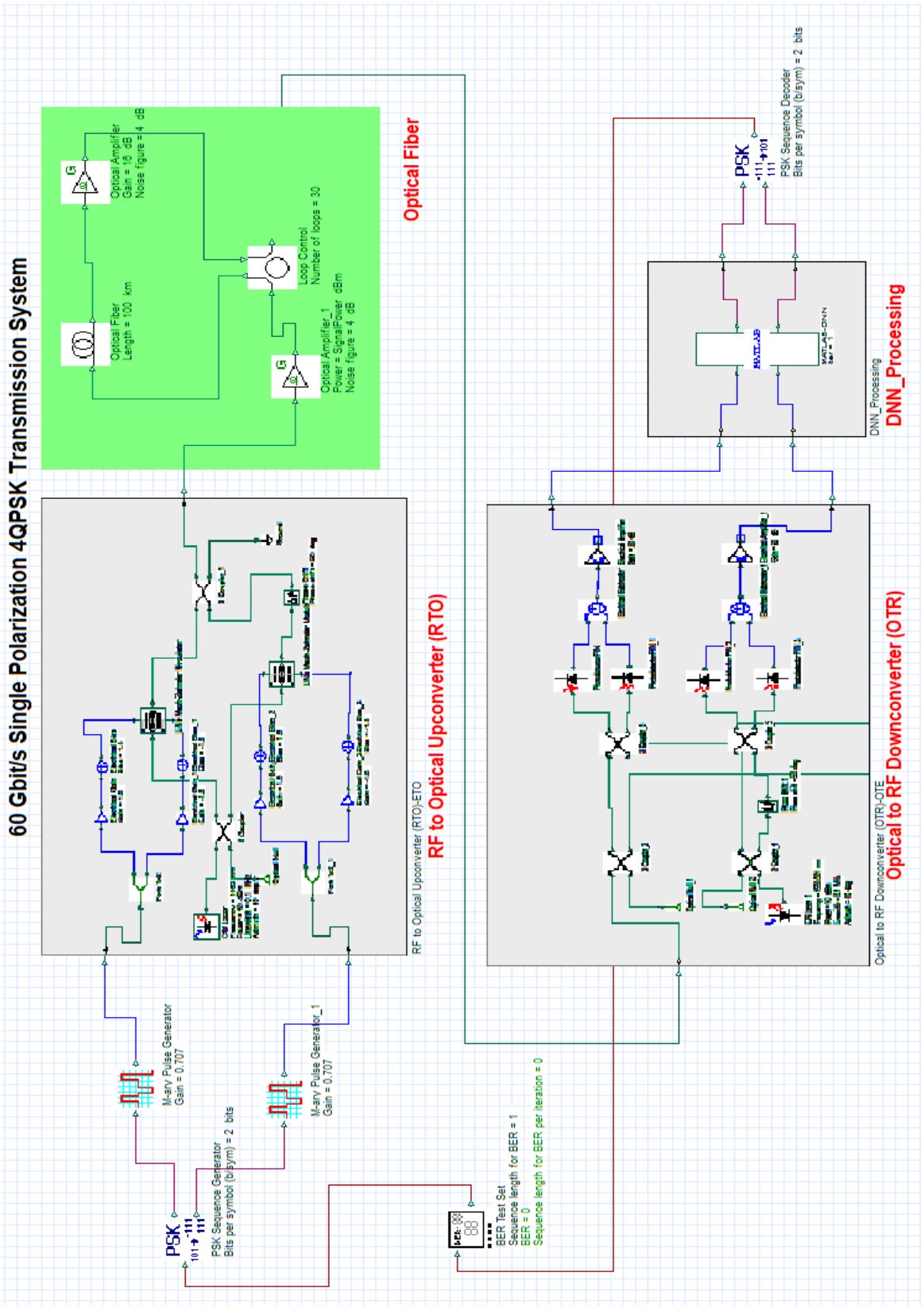


Figure (4.8): The proposed QPSK System

4.3.3 CD-OOFDM System

Figures (4.9) and (4.10) describe the proposed CD-OOFDM system. Also, Table (4.9) illustrates the main parameters settings of the proposed system. The coherent detection-OOFDM system has two additional units compared to direct detection namely RF to the optic (RTO) unit in transmitter part and the optic to RF (OTR) unit in receiver part.

RTO unit consists of X-coupler, two dual-drive lithium Niubate Mach-Zehnder modulators (MZM), and an optical combining unit as shown in Figure (4.2 a), while, the OTR unit as shown in Figure (4.2 b) is consists of four X-couplers, one 90° phase shift, four PIN photodetectors with balanced noise detectors, and two electrical subtractors.

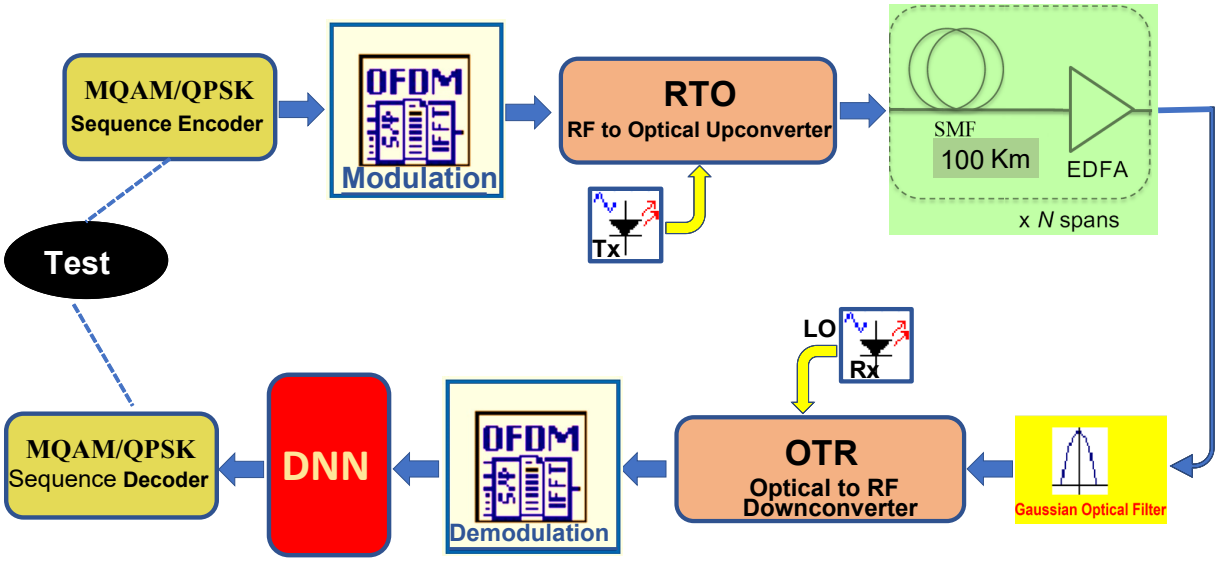


Figure (4.9): Block Diagram of Proposed CD-OOFDM System.

The work of the CD-OOFDM proposed system will be divided into three parts: transmission part, transmission media, receiver part as follows: **Firstly, The transmission part:** The input data sequence is built by using PRBS to generate a bit sequence with a length set to (2^{16}) and then, it is fed to M-QAM, QPSK sequence generator encoder to generate two parallel M-ary

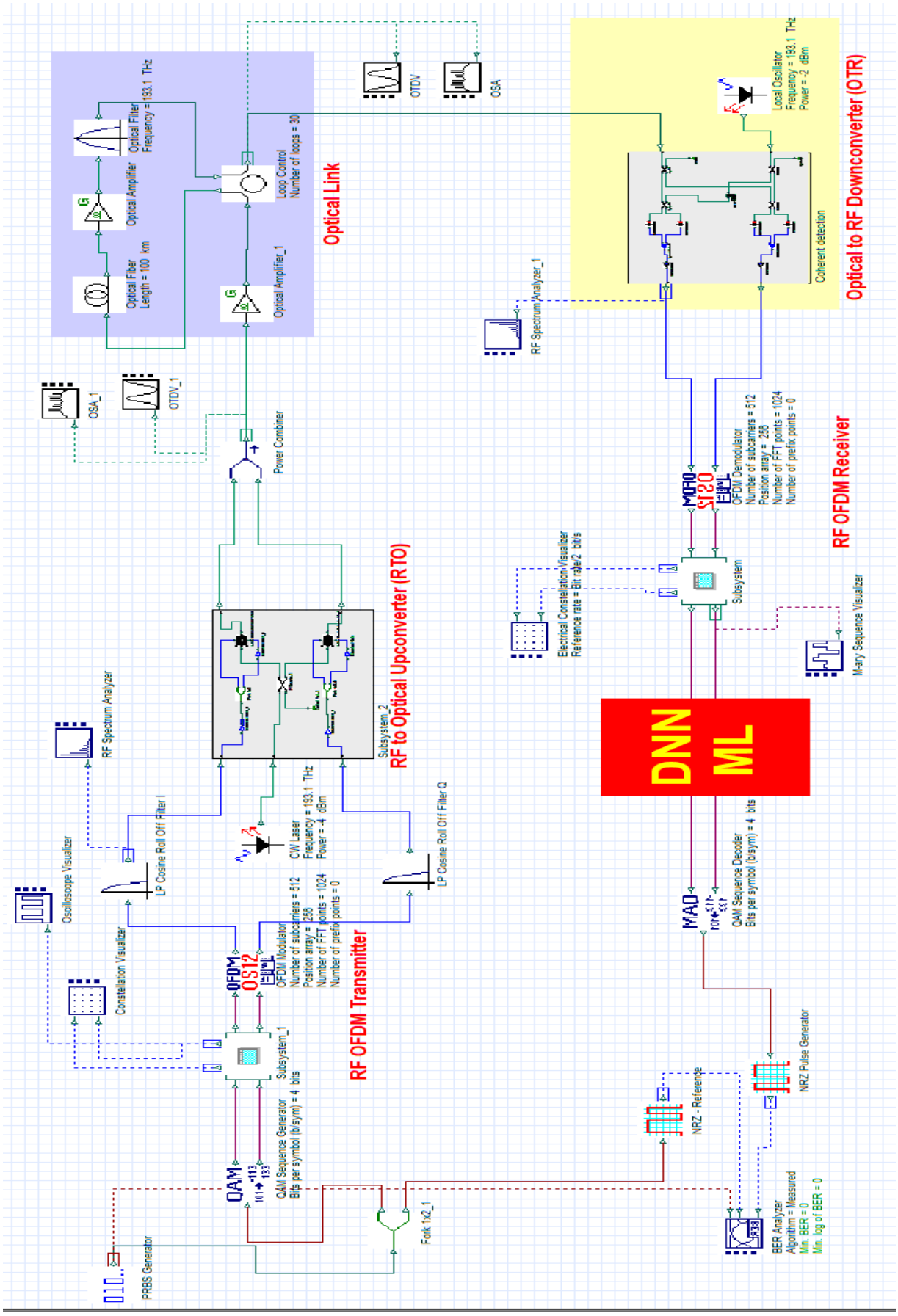


Figure (4.10): The Simulation Setup for the Proposed CD-OOFDM System.

sequences of symbols as in-phase (I) and quadrature (Q) components from input binary signals. after that, the output of the M-ary encoder is fed to an OFDM modulator with main parameters as illustrated in Table (4.9). then, the output resulting signal from the OFDM modulator (in-phase (I)/quadrature (Q)) is fed to the RTO unit. In the RTO unit, the Low Pass Cosine Roll Off Filter has been utilized for each baseband OFDM (I/Q) signal component with the setting parameters shown in Table (4.2). after using electrical gain, each baseband OFDM (I or Q) signal component is split into two parts by using Fork device (1X2) to copies the electrical input signal into two output signals will be fed into electrical input ports of MZM, in addition, the optical input port of the MZM will feed on CW Leaser. Finally, the optical output signals of both MZMs are combined by a Power Combiner (2x1) device to combine the two (I/Q) MZM optical input signals into the complex optical OFDM signal with amplified it to prepare for transmitted.

Secondly, the fiber optic transmission media: with main parameters as shown in Table (4.1).

Thirdly, the receiver part: The received optical signal is detected by the OTR unit with balanced noise detectors. In the OTR unit, for the purpose of extracting components (I) and (Q) separately, the first X-Coupler has been utilized in order to divide the incoming received OOFDM signal to two parts. Similarly, in order to extract the optical modulated signal, the second X-coupler has been utilized in order to divide the LO signal to two parts that have 90^0 phase shift and mix it with the output of the first X-coupler (OOFDM signal in branches (I) and (Q)) by using two pair X-coupler (3rd /4th). The outputs from the 3rd and 4th X-coupler are feeding to PDs of balanced detector for the generation of (I/Q) components of base-band electrical signal of the OFDM after the subtraction of photo-currents output. Finally, OTR unit output is (I/Q) signal fed to the OFDM Demodulator then to the proposed DNN-ML.

Table (4.9): SP-CD-OFDM Parameters Setup.

Parameter	Value
Users number	1
Polarization	Single Polarization
Bit Rate	60 Gbps
Modulation	M-QAM and QPSK
Coupling ratio of X-couplers	0.5
OFDM Subcarriers	512
Position array	256
IFFT number	1024
Cyclic prefix	100, Zero values
DAC Interpolation	Cubic
Pilot symbols	100 symbols for Carrier Phase Estimation
Training symbols	100 symbols for Channel Estimation

4.4 Dual-Polarization Coherent Detection System

The proposed dual-polarization coherent detection optical system block diagram is shown in Figures (4.11), (4.12) and (4.13). It's essentially constituted of two transmitter parts that operate in a parallel process, every one of them generating base-band signal (one for each one of the polarizations). The 1st transmitter here has been referred to as the X-polarization and the second transmitter as Y-polarization. The architecture of the transmitter part is described in the previous section in brief detail.

The two electrical baseband signals from both (X/Y) polarization of the transmitter part are fed directly to the RTO unit as the first input, while the 2nd input of the RTO unit is fed from the output of the polarization splitter (1x2) which divide the beam of the CW laser source into two beams with 90^0 degree orthogonal polarization.

The output signal from each one of the RTOs units represent one polarization, after that the two generated polarization x and y have been combined with the Polarization Combiner (2x1) device for the purpose of forming dual-polarization optical signal and then it is amplified and sent through the optical cable.

At receiver's side, the obtained optic signal has been split to two orthogonal components of polarization through a polarization splitter (1x2) optical device, also the LO beam is split as well to two orthogonal components of polarization through the polarization splitter (1x2). After that the outputs of the two-polarization splitter devices are fed to the OTR unit, which is described in section (4.3) in brief details.

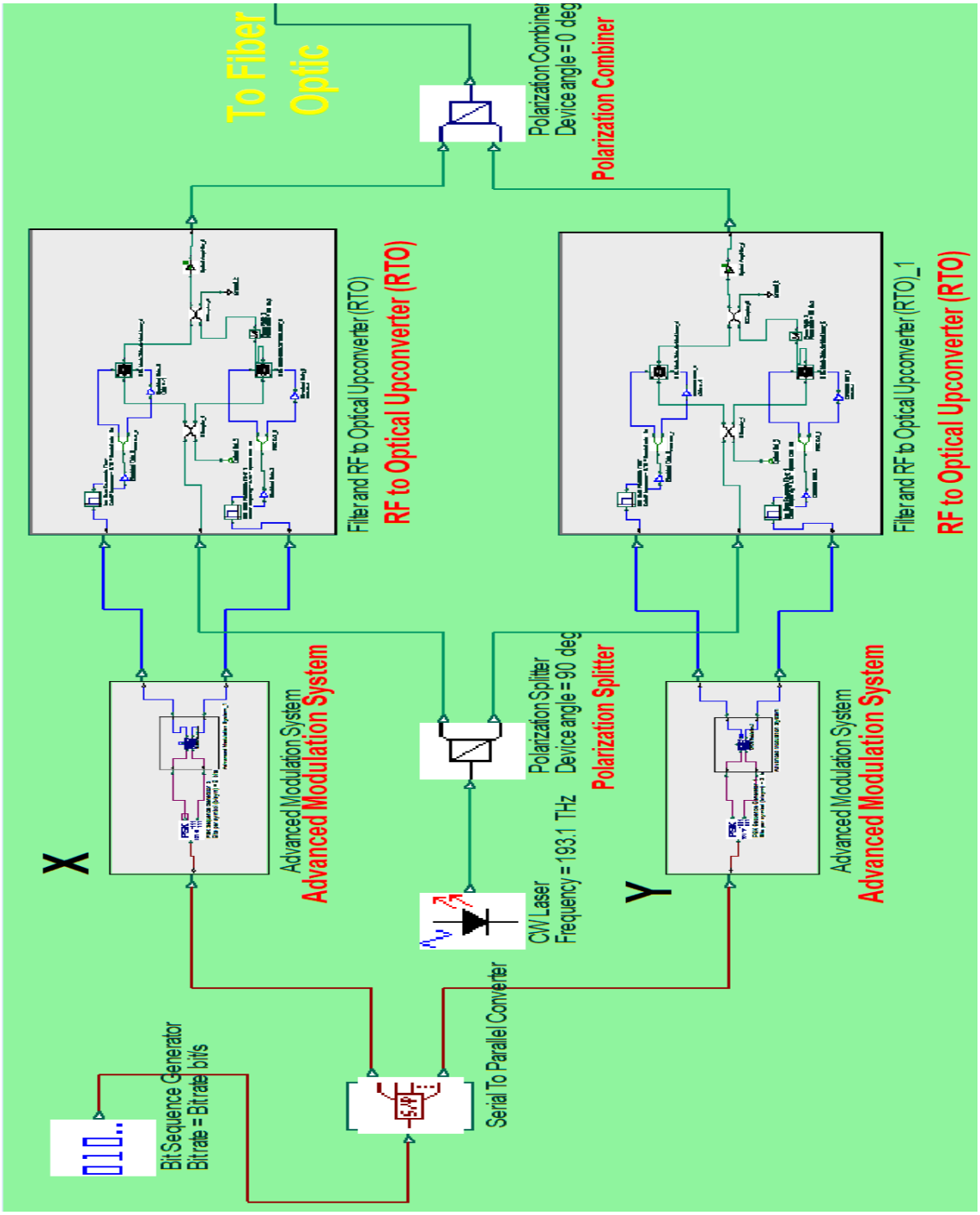


Figure (4.11): Simulation for Suggested DP-CD Transmitter.

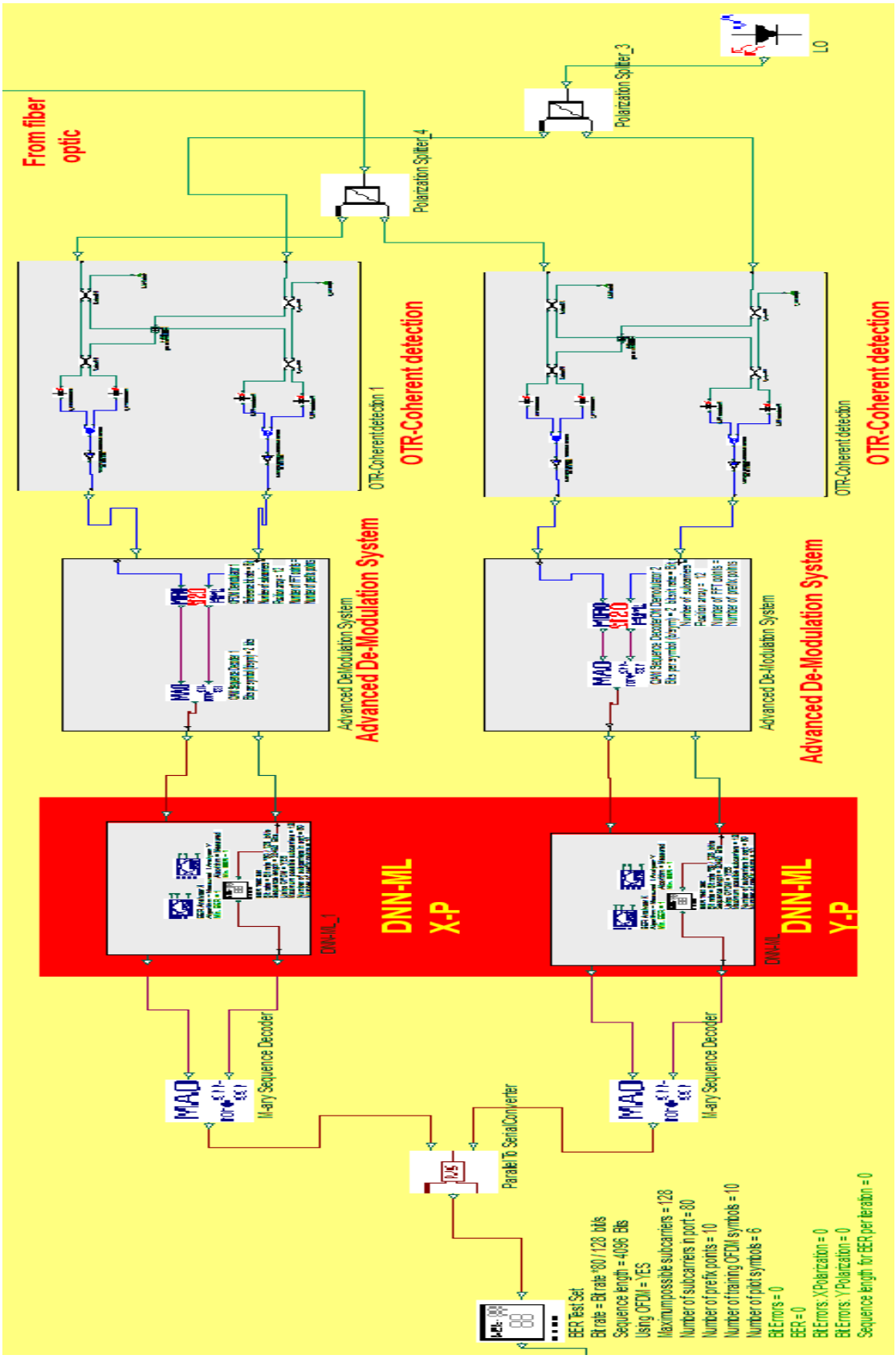
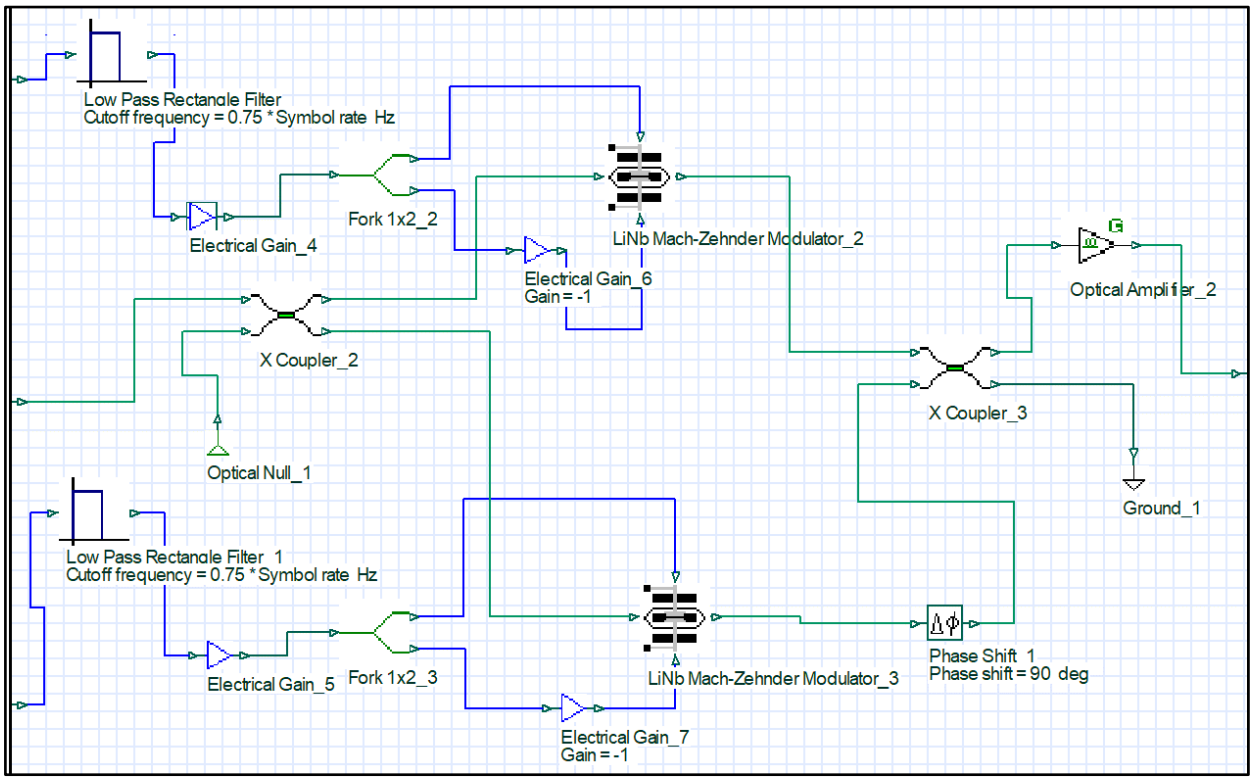
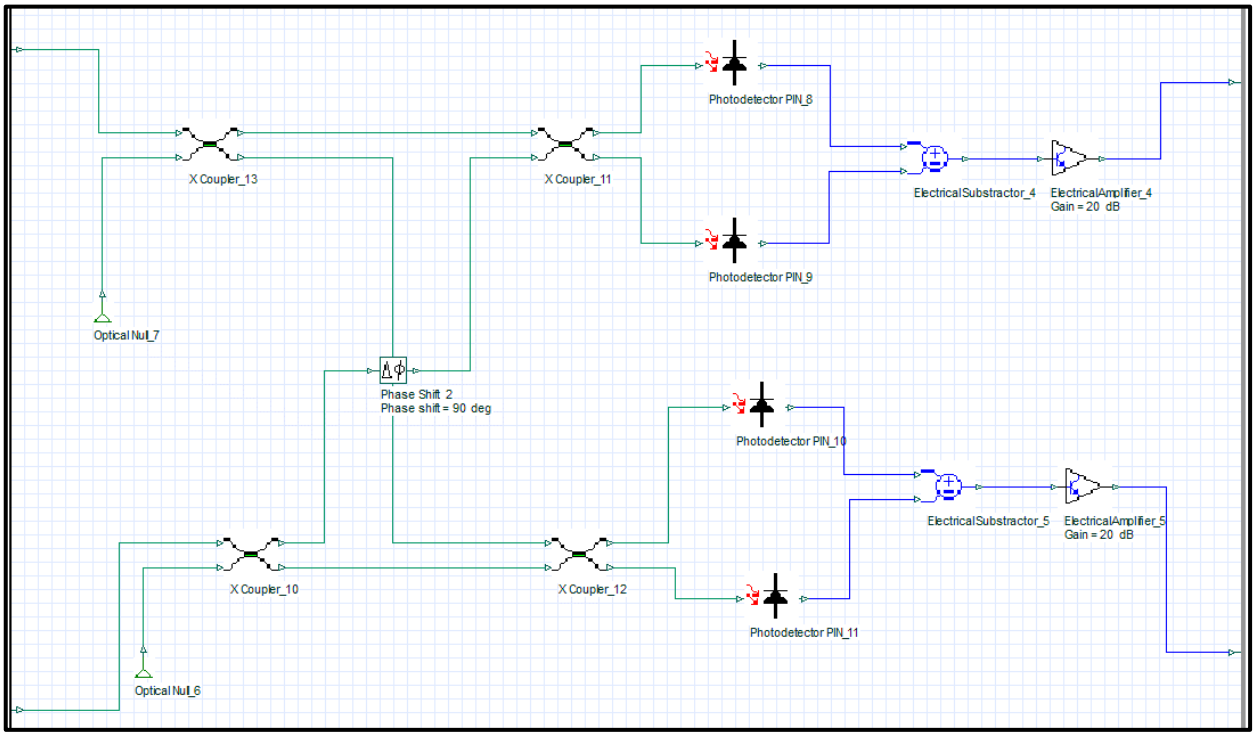


Fig. (4.12): Simulation of Suggested DP-CD Receiver.



(a) RTO Unit Simulation Setup for Each Polarization.



(b) OTR Unit Simulation Setup for Each Polarization.

Figure (4.13): Simulation Setup: (a) RTO. (b) OTR

Now, the output of the OTR unit is (I/Q) signal of every signal (X & Y polarization) is input to demodulator step and then to the proposed DNN-ML process to compensation the optical nonlinear impairment. Finally, mapper decoder scheme with parallel to serial converted are applied to produce the binary data signal.

4.4.1 M-QAM-DP-Coherent Detection System

Table (4.10) shows the setting of the main parameters for the proposed DP-CD-M-QAM system. In addition, the I/Q mapping that used as reference target and classes for the proposed DNN-Ml for each polarization is the same as for the single polarization which is described in section (4.3) in brief details.

Table (4.10): Main Parameters Setting of Suggested DP-CD-M-QAM.

Parameter Name	Values	Description
Bit rate	120 Gbps	High bit rate
Symbol rate	Bit rate/h GBd	h= no. of bit/symbol
Mapper	4,16,64 QAM	Square constellation map
Sequence length	2^16 bits	Long sequence length gives accurate result
Samples per bit	4-8-16 Sample/bit	Sweep no. of sample/bit to gives accurate result
Guard Bits	100 bits	Defines the number of bits to ignore before and after a sequence of bits
FEC	Gray code	
I-Q mapping	User-defined	User define the (I&Q) position to use as reference target and classes in ML

4.4.2 OFDM-DP-Coherent Detection System

The setting of the key parameters for suggested DP-CD-OOFDM system is shown in Table (4.11). In addition, the details of I/Q mapping that used with OOFDM as M-ary sequence encoder is the same which described in the previous section.

Table (4.11): Key Parameters of Suggested DP-CD-OOFDM System

Parameters	Values	Comments
Bit Rate	120Gbps	High date rate
Bit Rate for every one of the polarizations	60Gbps	For each modulation
Modulation	4, 16, 64 – QAM	
No. of Users	1	number of the users for OFDM system
Wavelength	1550nm	
Number of prefix points	100	Number of points used in the guard period
Guard Bits	100 bits	Defines the number of bits to ignore before and after a sequence of bits
Pilot symbols	100 symbols	for Carrier Phase Estimation
Training symbols	100 symbols	for Channel Estimation
OFDM FFT/IFFT	1024	Number of points used in the IFFT/FFT
OFDM Subcarriers	512	Number of subcarriers used in the transmission by each user
Channel	SMF-Fiber-Optic	Main Parameters Setting shown in Table (4.1)
LD/LO Power	Sweep (0-20) dBm	Set to optimum power
DAC Interpolation	Cubic	
Photodetector	PIN	Responsivity=1 A/W [3]
Coupling ratio of X-couplers	0.50	
Polarization types	Dual	X & Y directions
Dark Current	10nA	Thermal noise = 1×10^{-22}W/Hz

4.5 WDM - Coherent Detection System

In this section, WDM will be used with all of the proposed optical communication systems described in the previous sections. The diagram of suggested WDM has been illustrated in Figure (4.14) and Table (4.12) presented the main parameters setting of the Proposed WDM system

This part discussed designing multi-channel WDM system and all blocks that are needed for building a typical single-channel system have already been explained, and only small modifications have been required to extend the design to multi-channel systems. A multiplexer device should be added at transmitter part to combine all the optical sub-channels so that they can be transmitted over the optic fibers, respectively, a demultiplexer device should be added at receiver part which will be providing the separation of optical sub-channels in a frequency domain and it may be easier to analyze them separately.

One of the potential designs of WDM transmitter utilizing the SP-M-QAM modulation. It should be noted that an identical design may be utilized for the transmission of any required format of modulation provided that external modulators are accordingly substituted.

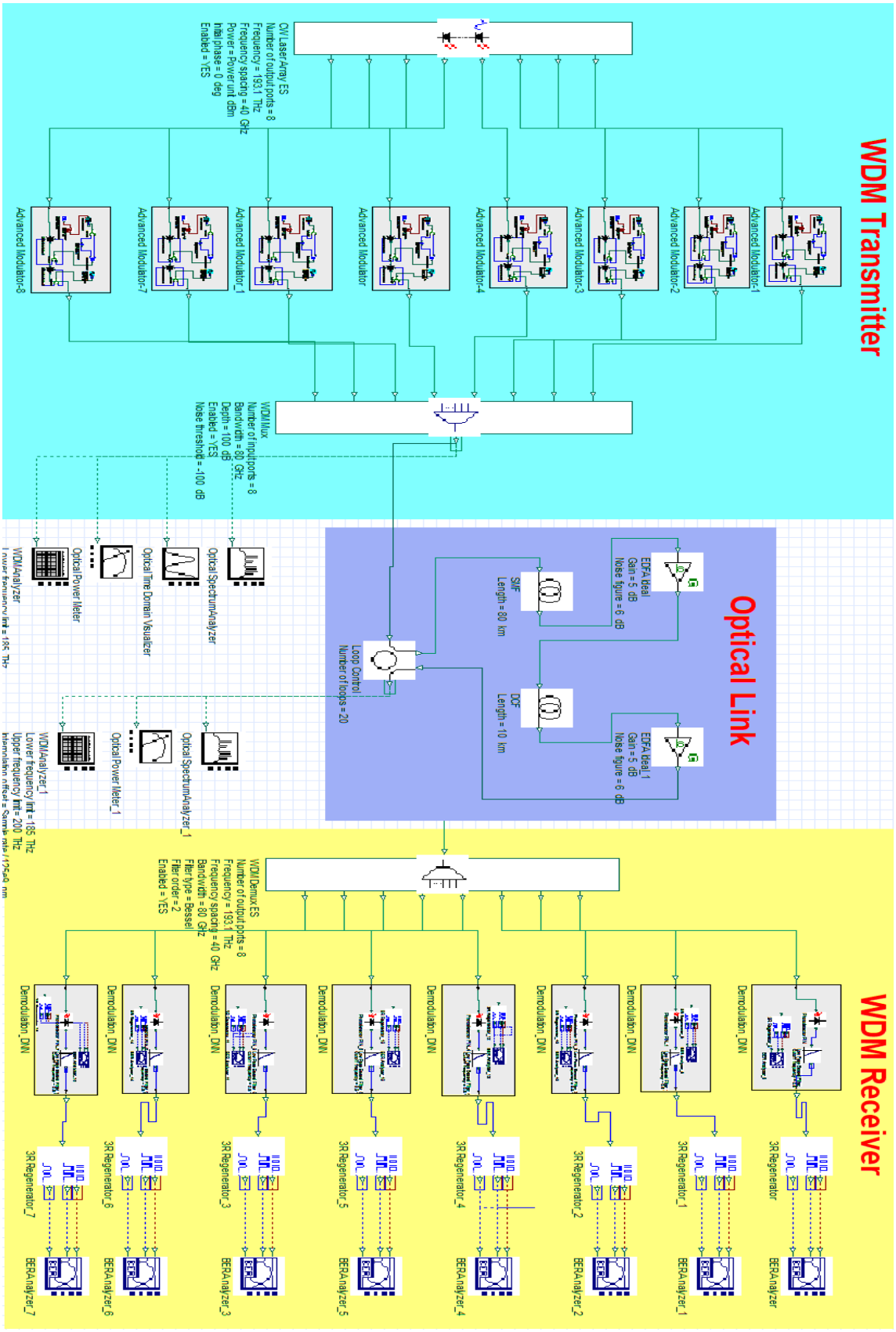


Figure (4.14): Simulation Setup for the Proposed WDM System.

Table (4.12): Main parameters setting of the Proposed WDM system

CW Laser Array Equally Spaced (ES)		
It has been considered as an equivalent to traditional CW Laser Array element. None-the-less, CW Laser Array equally spaced model is easier to setup for the system of WDM, due to the fact that it requires only initial laser emission frequency and spacing. The power of signal output has been identical for all of the signals of the output.		
Parameter	Value	Description
Optical Source	CW Laser ES	
Number of output ports	8	Sub-channel
Frequency	193.1 THz	First laser's emission frequency
Linewidth	10 MHz	
Initial phase	0	Phase of the first laser
Power	Sweep [-5,0,5,10] dBm	Set to optimum result
Frequency spacing	40 GHz	Frequency spacing between the neighboring lasers
Noise threshold	-100 dB	Minimum value for adaptation of noise bins
Modulation System		
Bit sequence generator	PRBS	Pseudo-Random
Bit sequence length	2^16	Long sequence length gives accurate result
Bitrate/channel/polarization	60Gbps	
Mapper constellation	4,16,64 QAM QPSK	Square constellation map
Polarization	Single & Dual polarization	
Samples per bit	4-8-16 Sample/bit	Sweep no. of sample/bit to gives accurate result
Guard Bits	100 bits	defines the number of bits to ignore before and after a sequence of bits
FEC	Gray code	
I-Q mapping	User-defined	Define the (I&Q) position to use as reference target and classes in ML as descried in previous sections

Parameter	Value	Description
OFDM	With & without	OFDM system as descried in previous sections
WDM Mux		
Number of input ports	8	
Bandwidth	Symbol rate	
Filter type	Bessel	3 Order of the function
Ripple	0 dB	
Filter Depth	100 dB	Maximum attenuation value for the filter

Input optical signals are filtered with the use of optical filter and combined in a single signal. The Bessel optical filter is used with 3rd order. The subsystem is illustrated in Figure (4.15).

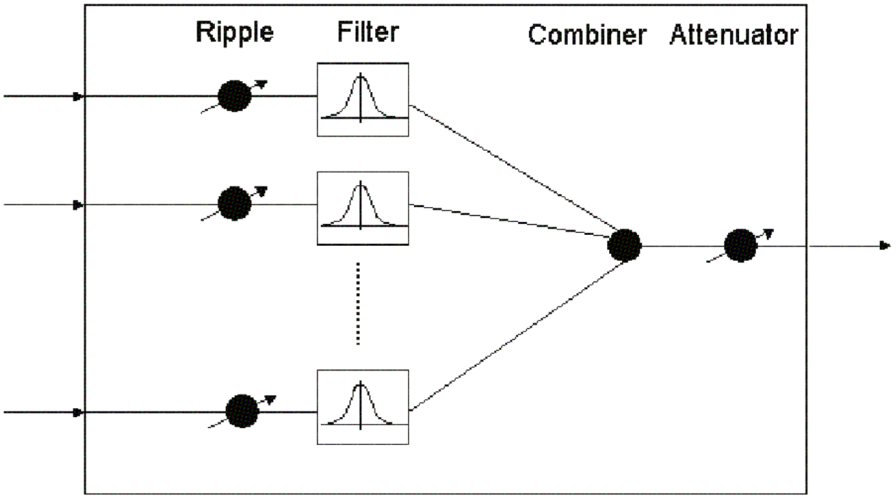


Figure (4.15): WDM subsystem [3]

The ripple, bandwidth, and depth of the filter determine the level of crosstalk for DEMUX and MUX components. When determining a specific channel's performance, these three criteria will specify the amount of the power from the adjacent channels that will serve as crosstalk terms. The most crucial parameter is depth, because it will have the greatest impact on the power levels of neighboring channels.

4.6 DNN-ML System Design

The proposed design of DNNs model is illustrated in Figure (4.16), as well as the main parameters for setting up the layers are shown in Table (4.13).

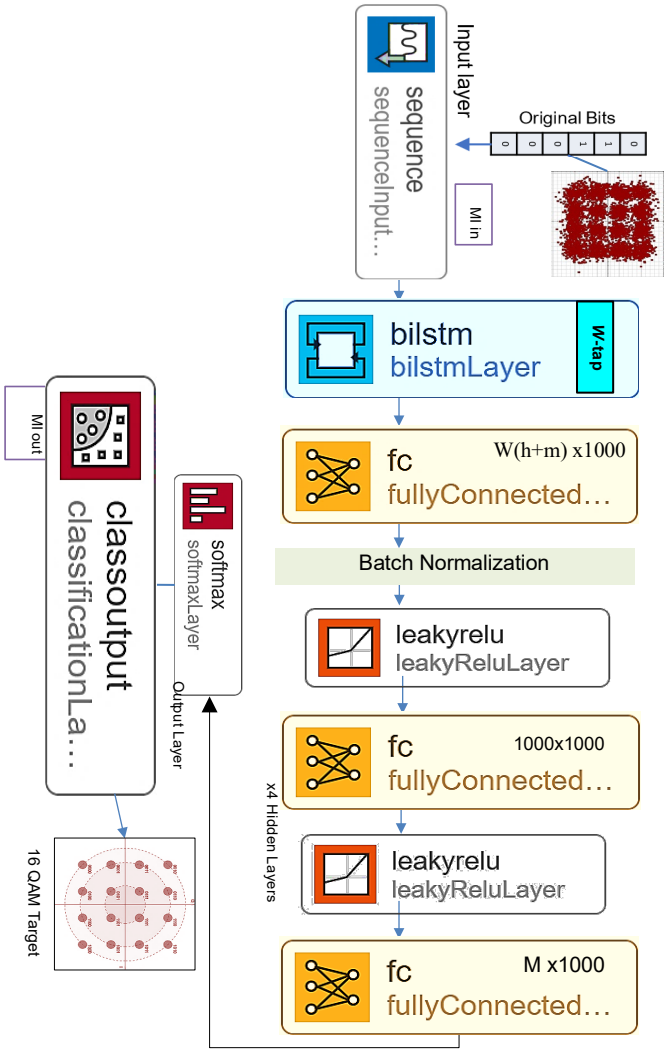


Figure (4.16). Proposed DNNs layers architecture

Input features: To create a working model of non-linear impairment, the DNN algorithm requires huge data, so, it is necessary to provide the DNN model with nonlinear impairment features in the data set to have sufficient nonlinearity compensation. Therefore, in order to get these nonlinear impairment features the launch power P_0 of the optical transmission system should be larger than the optimum channel power.

These features will be provided to the proposed DNN model through the intrachannel nonlinear impairment in the case of using single-channel SSMF, while the inter-channel nonlinear impairment features will be provided in the case of using WDM.

For remember, the intrachannel nonlinear impairment are the distortions produced inside the bandwidth of a channel and include intra-channel cross-phase modulation (IXPM), intra-channel 4-wave mixing (IFWM), intrachannel self-phase modulation (SPM), while the inter-channel nonlinear effects are the distortions from different channels in the case of WDM that take place (XPM, FWM). Also, keep in mind that the NLI effects caused by signal-signal interactions are deterministic nature, while on the other hand, NLI effects caused by signal-noise interactions are stochastic nature [52].

In addition, the DNN needs huge data to learn and it consumes a high amount of hardware resources to do the calculations which makes it suitable for future optical communications [22]. Therefore, the Single Nonbinary SoftMax Classification Algorithm can be used but does not perform well for higher-order mapper constellation because it needed to unrealistically hugely datasets for training. So, this proposed DNNs model used a multi-label SoftMax Classification Algorithm, which is more scalable for high-order modulation.

Finally, the size of the dataset used (128 x 262144) samples. The number 128 indicates that: the transmission system will be repeated 128 times for the same sequence input for the purpose of adequately training the NN and reach to find the perfect estimates of neurons' weights (W), where 80% of the dataset utilized for training, while 20% for validation and testing.

Before being fed into DNN, the dataset will be divided into imaginary and real parts. The non-linear activation function is utilized in hidden layers rather than the linear activation function because it achieves better performance than linear activation function as discussed in Sections (3.3.4 and

3.5.1).

As a result, and based on the study as shown in Figure (4.17), the Leaky ReLU is the best non-linear activation function according to this work and it will be used to maximize the profit gain of the proposed deep neural network model.

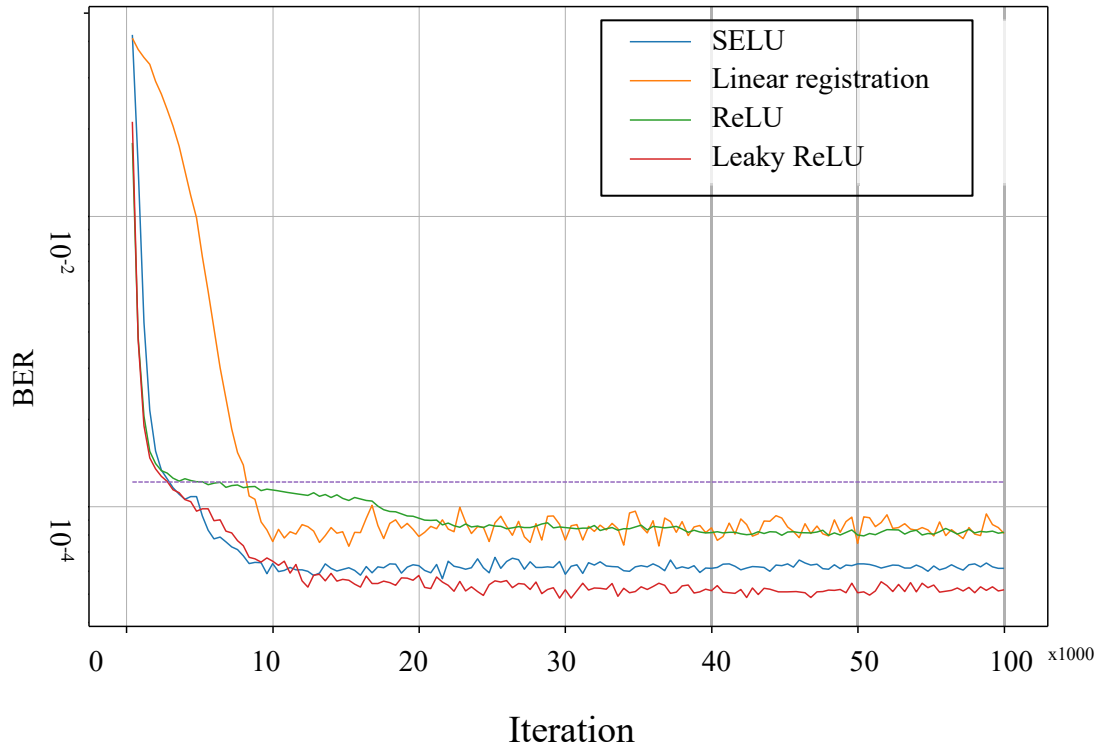


Figure (4.17): Train & Test with Different Activation Functions.

Table (4.13): Main DNN Layers parameters and Setup.

Layers Type	Description	Properties & Setup
Sequence Inputs	Sequence input layer it is the inputs sequence data to the network.	▪ Size of input =Sequence Length ▪ Number of inputs =2. (Number of Variables (Real and Imaginary Parts))
LSTM	An LSTM layer. In time series and sequence data, an LSTM layer learns long-term dependencies amongst the time steps. The layer results in the production of the additive interactions that might aid in the improvement of gradient flow across long training sequences.	▪ Number of hidden nodes per layer=100 ▪ Output Mode= sequence output the complete sequence. ▪ Activation function to update cell and hidden state= tan h. ▪ Activation function for applying gates= sigmoid. ▪ Function for the initialization of input weight values =Glorot initializer ▪ Function for the initialization of recurrent weight values = Glorot initializer ▪ Function for the initialization of bias = 'ones' ▪ Learning rate factor for input weight values= Output gate ▪ Learning rate factor for recurrent weight values= Output gate ▪ Learning rate factor for bias values=User define ▪ L2 regularization factor for input weight values=Output gate ▪ L2 regularization factor for recurrent weight values=Output gate ▪ L2 regularization factor for bias values=Output gate
BiLSTM	A Bi-directional LSTM BiLSTM layer is used for learning the bi-directional long-term relationships amongst the time steps in a time series or sequence of data. In the case where one wants the network to learn from whole time series at every one of the time steps, such dependencies might be advantageous.	

Layers Type	Description	Properties & Setup
LReLU	A Leaky Rectified Linear Unit layer, conducts threshold process, which multiplies any input value that is less than by some fixed scalar value.	▪ Scalar multiplier for input values = range (0.1-0.5) $f(x) = \begin{cases} x, & x \geq 0 \\ scale * x, & x < 0 \end{cases}$
Batch Normalization Layer	Every one of the inputs is normalized over a minibatch via layer of batch normalization. Use batch normalization layers for the purpose of speeding up NN training and decrease sensitivity to network initialization. Through the subtraction of minibatch average and dividing by minibatch standard deviation, this layer initially normalizes the activations of each input. The layer after that shifts and scales the input through a learnable value of offset β as well as a learnable scale factor γ.	▪ Function to initialize the input scale factors= 'ones' ▪ Function to initialize the input offsets= 'zeros' ▪ Scale Learning rate factor= 'ones' ▪ Offset Learning rate factor= 'ones' ▪ L2 regularization Scale factor= 'zeros' ▪ L2 regularization Offset factor= 'zeros' Epsilon — Constant to add to minibatch variances=1e-5
Fully connected	In this type of the layers, input is multiplied by some matrix of weights, and after that an addition of bias vector has been performed.	▪ Output size =Num. of Classes ▪ Input size= Hidden Size. ▪ Weight learns rate factor=1 ▪ Weight L2 factor=1 ▪ Bias Learn rate factor=1 ▪ Bias L2 factor=0 ▪ Function to initialize the weight='glorot' ▪ Function to initialize the bias='zeros'
SoftMax & Classification Output	The input is given a SoftMax function by a SoftMax layer. For the problems of the multiclass classification with the mutually-exclusive classes, a classification layer evaluates the cross-entropy loss. The number of the classes has been inferred from output size regarding the preceding layer.	▪ Output size=Num. of Classes. ▪ Loss function=cross-entropy.

The DNNs is trained with sigmoid activation functions in the input layer, a SoftMax & Classification Output layer, a LSTM layer, and a ReLU at hidden layer are used, which performs a threshold operation with cross-entropy loss function with ADAM optimization algorithm. ADAM learning algorithm with an initial learning rate of 0.001 and max epochs size of b = 1000, minibatch size is 100 and learn rate drop factor is 0.05.

4.6.1 DNN-Based Optical NLC

The proposed DNNs model is a supervised learning algorithm will be deal with classification problems, therefore the proposed optical NLC will be treated as a supervised nonlinear classification problem in machine learning terms. The proposed DNNs-ML-based optical NLC is shows in Figure (4.16)

The amount of neurons in input layer is dependent upon specific dimensions of the input data vector. In this work number of neurons in an input layer equal to two neurons. Additionally, there are no fixed rules for choosing the optimal number of the hidden layers or optimal number of the neurons for a given hidden layer in a DNN. Usually, the choice is made by experimentation, experience, and prior knowledge of the problem, this scenario is known as the hyperparameters of a DNN [66].

A classification DNNs learns by the training processes on the dataset as a function F(.): $F^X \rightarrow F^Y$ where X represents the input labels and Y represents the output classes. A single neuron is an autonomous processing unit defined by the input (x), Bias (B), weights (W), and the Activation Function σ(.). The activation function (AF) which applied to find the neuron desired output σ(z) = y. In addition, the Bias term for each neuron is included by some additional branch.

Every one of the neurons contains a number of inputs (associated with a weight) and one output. to obtain the neuron output, the neuron calculates the AF parameter by sum the input weight and the bias. This output of AF must be identical to the application and represent general dynamic range of a target signal.

Note that the AF in the input layer different from the AF in hidden layers and in output layer. In this work, the tanh activation function was utilized for updating cell and hidden state while a sigmoid function of activation was applied to gates for the input layer.

For classification problems, the activation function must be matched with the application, so the best common option of the function of activation in last layer is therefore a SoftMax activation function.

This activation function is different from activation functions which is utilized in hidden layers, which were used the Leaky ReLU as a nonlinear activation function according to the study presented in the previous section.

The notation of the DNNs design model in Figure (3.10) is as follows

$\mathbf{S} = [S_1 \ldots S_X]$ - denotes the input signal,

$\underline{\mathbf{S}} = [\underline{S_1} \ldots \underline{S_C}]$ –represents classified signal,

$W^{[L]}$ - represents weight matrix (which includes Bias weight)

$y_{(i)}$ - represents output of i^{th} neuron in L^{th} layer,

$z_{(i)}$ - denotes sum up input of the i^{th} neuron in L^{th} layer,

$\sigma(z_{(i)})^{[L]}$ - denotes activation function of the i^{th} neuron in the L^{th} layer.

Here, h denotes number of the hidden layers nodes and L denotes current layer and. The output may be expressed as follows:

$$y^{[0]} = S \tag{4.4}$$

$$z^{[L]} = W^{[L]}.\, y^{[L-1]} \tag{4.5}$$

$$y^{[L]} = \sigma^{[L]}.\, z^{[L]} \tag{4.6}$$

where the equations above denote the input layer with the next layer, which iteratively executes for successive layers to get output $y^{[L]}$.

The classified signal has been obtained at DNN model output and may be extracted as:

$$S_{C\ DNN-NLC} = y^{[L]}\ (S, W) \tag{4.7}$$

For the purpose of finding the appropriate estimation weight values for each neuron such that desired outputs $y^{[L]}$ should be the same as the target outputs S_C, and also, to evaluate the DNN model performance, the Loss function {L (S, C, W)} must be presented.

DNN-NLC Proposed Model: Optimization and Regularization

1. Define Initial Custom Weighted Layer: Experience, experimentation, and prior knowledge regarding the problem are typically used for defining the initial custom weighted layer depending on the hyperparameters of a DNN scenario. The biases and weights are after that initialized using specific functions stated in the layer attributes, as shown in Table (4.13), once the network has been trained.

2. Define a Forward Loss Function: Characterize and define loss between predictions Y and training targets T, the forward loss function that used with classifications problems is cross entropy loss function. Weighted cross entropy can be defined as measurement of the error between

prediction scores made by the network Y and targets of training T, weighted loss of the cross entropy is given by:

$$L = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K} w_i\, T_{ni} \log Y_{ni} \quad (4.8)$$

In which N represent the number of the observation's samples, K represent the number of the classes, and W represent a weight vector for every one of the classes.

The cross-entropy loss assesses how closely the network predictions match the target classification. The cross-entropy loss function has been estimated for the classification task with the use of the next formula:

$$Loss = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C}\left(T_{i,j}\log\left(X_{i,j}\right) + \left(1 - T_{i,j}\right)\log\left(1 - X_{i,j}\right)\right) \quad (4.9)$$

where here $X_{i,j}$ is the network response for a given class, $T_{i,j}$ is the target value of that classes, and C denotes total number of the classes.

3. Define a Backward Loss Function: Specify and define the loss function's derivative with regard to the predictions. The loss derivatives regarding input data and the learnable parameters are returned by backward Loss functions. In terms of the predictions Y, the derivative of the cross-entropy loss is:

$$\frac{\delta L}{\delta Y_i} = -\frac{1}{N}\frac{w_i T_i}{Y_i} \quad (4.10)$$

The input Y contains the network's predictions, while T contains the training targets. The derivative of the loss function in terms of the predictions Y is the output (dL/dY).

4. The Output Layer: this layer should match the application and represent the target signal's general dynamic range; in this case, it's a classification layer that is used for the calculation of cross-entropy losses for the

problems of the mutually exclusive classes. The layer of classification should come after SoftMax layer in most classification networks. SoftMax layer uses SoftMax activation function, and the classification layer uses the SoftMax function's values for assigning every one of the inputs to a K mutually exclusive class during network training.

SoftMax function is given by:

$$y_r(x) = \frac{\exp(a_r(x))}{\sum_{j=1}^{k} \exp\left(a_j(x)\right)} \tag{4.11}$$

where $0 \leq y_r \leq 1$ (probability vector), $a_r(x)$ is the output nodes of the neural network, k the number of classes and $\sum_{j=1}^{k} y_j = 1$

SoftMax is a continuously differentiable activation function. This enables the derivative of the loss function to be calculated in terms of each weight in the NN. This property allows the model to alter the weights for minimizing the loss function (model output close to the true values).

The major goal, according to the ML principle, is to select the DNN with the lowest loss estimates of W and the maximum probability of training data set. This is accomplished through the maximization of joint conditional probability density function, which is written as:

$$P(C_r|X,\theta) = \frac{P(X,\theta|C_r)P(C_r)}{\sum_{j=1}^{k} P\left(X,\theta\middle|C_{rj}\right)P(C_j)} \tag{4.12}$$

where θ is the DNN parameters vector, $0 \leq P(C_r|X,\theta) \leq 1$ & $P\left(C_j\middle|X,\theta\right)$ =1.

$$P(C_r|X,\theta) = \frac{\exp(a_r(x,\theta))}{\sum_{j=1}^{k} \exp\left(a_j(x,\theta)\right)} \tag{4.13}$$

$$a_r = \ln\left(P(X,\theta|C_r).P(C_r)\right) \tag{4.14}$$

where $P(X, \theta|C_r)$ represent conditional probabilities of a sample given class *r*, and $P(C_r)$ represent prior probability of a class.

Equations (4.13) and (4.14) indicates the final overall loss function of all samples in the dataset.

For the purpose of finding the weights, which decrease the cross-entropy function at highest PDF, the SGD Algorithm is used.

5. The Loss function's SGD Algorithm must be calculated with regard to parameters of DNN (θ). The standardized gradient descent algorithm updates parameters of a network (i.e. bias and weight values) for minimizing loss function through taking little steps in the negative gradient direction of loss function at every one of the iterations. The common backpropagation algorithm [74] is an efficient approach. In terms of the network parameters (θ), the back-propagation method approaches loss function derivatives from training dataset. It efficiently applies the chain rule to derivatives. By applying the chain rule, the loss function derivation with regard to weight values in L^{th}-layer down to 1^{st} layer as follows:

$$\theta_{\ell+1} = \theta_{\ell} - \alpha \nabla E(\theta_{\ell}) \tag{4.15}$$

In which $\alpha>0$ represent rate of learning, ℓ represent iteration number, θ represent DNN parameter vector, and E(θ) represent function of loss [74].

6. Stochastic Gradient Descent with the Momentum: Because SGD approach could oscillate along steepest descent path to optimal one, adding a term of the momentum to parameter update is a technique to lessen oscillations [75]. The alteration for the SGDM will be expressed as follows:

$$\theta_{\ell+1} = \theta_{\ell} - \alpha\nabla E(\theta_{\ell}) + \gamma(\theta_{\ell} - \theta_{\ell+1}) \quad (4.16)$$

In which γ specifies contribution regarding previous gradient step to current one.

Furthermore, for all parameters, the SGD algorithm and the SGDM use a single learning rate. Adam optimization techniques aim to enhance network training through utilizing rates of learning, which vary based on the parameter and might adjust to loss function that is automatically optimized [76]

7. Adam [76] uses a parameter update with added term of momentum. It keeps elementwise moving mean value of parameter gradients as well as their square values.

$$m_{\ell} = \beta_1 m_{\ell-1} + (1 - \beta_1)\,\nabla E(\theta_{\ell}) \quad (4.17)$$

$$v_{\ell} = \beta_2 v_{\ell-1} + (1 - \beta_2)[\nabla E(\theta_{\ell})]^2 \quad (4.18)$$

where β2 and β1 are respectively, Squared Gradient Decay Factor and Gradient Decay Factor. Adam used moving averages for updating parameters of the network as follows:

$$\theta_{\ell+1} = \theta_{\ell} - \frac{\alpha\, m_{\ell}}{\sqrt{v_{\ell}}+\epsilon} \quad (4.19)$$

where ε is the epsilon parameter was defined in Table (4.13)

In the case when the gradients throughout various iterations are comparable, utilizing gradient moving average allows parameter updates in order to gain momentum in specific direction. When gradients include a lot of noise, gradient's moving average become smaller, and parameter updates become smaller as well.

8. Gradient Clipping: The training has been considered unstable and might diverge in some iterations in the case when the gradients rise in magnitude exponentially. Gradient clipping stabilizes the training at

high rates of learning and in case of existence of the outliers, preventing gradient explosion [75]. Gradient clipping allows training the networks more quickly while having no effect on the learned task's accuracy. There are 2 gradient clipping types:

A. Gradient clipping depending on a threshold rescales the gradient but does not modify the gradient's direction.

B. Any partial derivative bigger than the threshold is clipped using value-based gradient clipping, causing the gradient to change direction arbitrarily. The behavior of value-based gradient clipping might be unpredictable, although small modifications don't cause the divergence of the network. The Gradient Threshold Method's 'absolute-value' value is employed in this study, with a value of (1-2) depending on the test outcome [75].

9. L_2 Regularization: One technique for reducing overfitting [75] is to add a term of regularization for weight values to loss function. With the regularization term, the loss function looks like this:

$$E_R(\theta) = E(\theta) + \lambda\Omega(w) \tag{4.20}$$

In which w represent the vector of weight, λ represent the factor of regularization (i.e. the coefficient), while the function of regularization Ω(w) is:

$$\Omega(w) = \frac{1}{2}\,(w.w^T) \tag{4.21}$$

Chapter 5
Simulation Results and Analysis

5.1 Introduction

In this chapter, advanced coherent optical systems are simulated and analyzed with the most important improvements made to reach the desired goal of reducing the nonlinear optical impairments using machine learning based on a deep neural network. This work was done with the help of the integration of OptiSystem Photonic Software and the MATLAB programming language to create a powerful and versatile tool. The OptiSystem Photonic Software component is used to create the custom optical system and then the found result data is called to MATLAB through which it was building the deep neural network to perform the processing on result data in real-time.

Figure (5.1) shows all the proposed designs of the optical system under study, analysis, and processing based on the proposed design of DNN-NLC. Two main types of constellation scheme formats M-QAM and QPSK are used in the simulation due to the high spectral efficiency of these types, which makes them give a high bit-rate capacity with a low BER as dicsussed in Chapter two.

The performance of the suggested DNNs-NLC model under different conditions of the optical systems has been studied and the results will be studied and analyzed as follows:

Firstly, the performance evaluation of the proposed DNN-NLC is examined using M-QAM-QPSK of constellation scheme formats without using the OFDM coherent detection system, and the study analysis is divided into two parts according to the system polarization (single and double).

Secondly, the performance of the proposed design of the DNN-NLC for the same optical systems above will be evaluated, but when coherent detection OFDM technology is used, also the study analysis is divided into two parts

according to the system polarization (single and double).

Finally, the performance evaluation of the proposed design of DNN-NLC when using multi- carrier WDM for single- and dual-polarization optical transmission systems will be studied and analyzed.

The results found within DNN-NLC will be compared with those of a conventional NLC based on DSP algorithms through calculating the end-to-end Q-factor and BER of the proposed optical transmission systems under study to identify the main role played by artificial intelligence technologies based on deep machine learning.

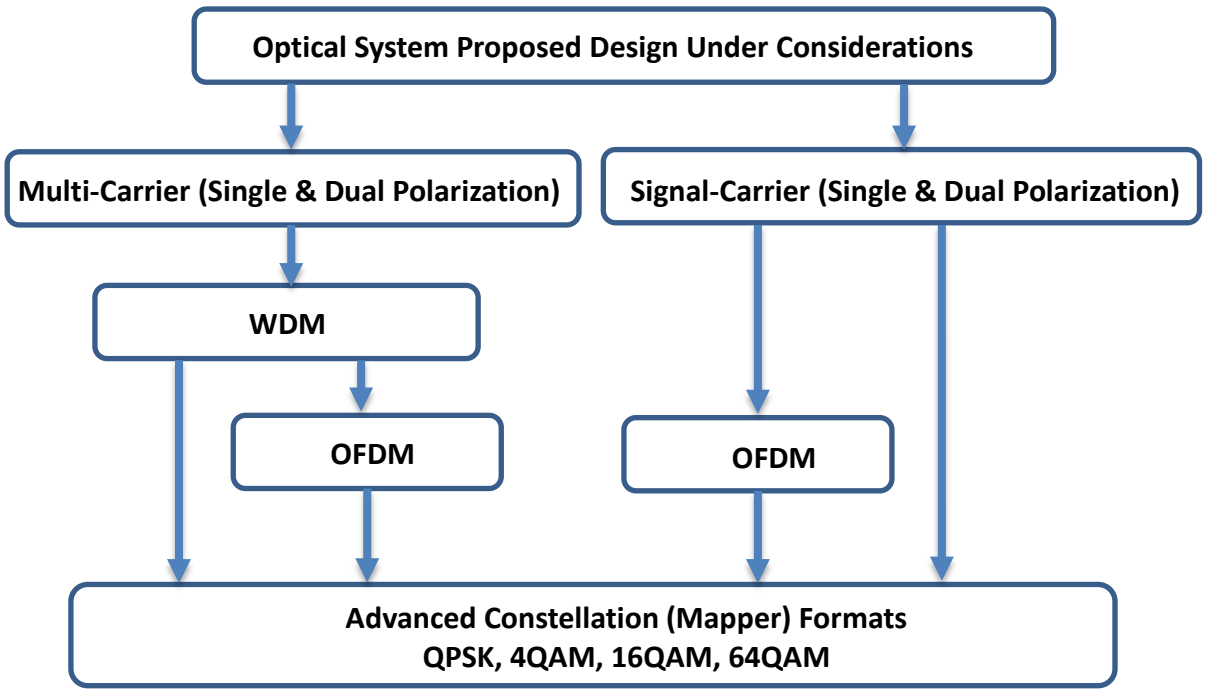


Figure (5.1): Optical Systems Under Study

5.2 SP- QPSK, M-QAM

In this Section, the evaluation of DNNs-NLC proposed model performance is examined using M-QAM, QPSK of advanced modulation schemes coherent detection single polarization system.

5.2.1 SP- QPSK & M-QAM: Results Analysis

When the power of fiber optic is low, the fiber channel characteristic will be treated as a linear mechanism, but at the high-level power the effects of nonlinearity impairments must be considered. The spectral broadening or pulse chirping occurs by the intrachannel SPM and XPM affects, this leads to increase the penalties for dispersion.

Here to evaluate the performance of the proposed DNN-NLC, it is necessary to have sufficient nonlinearity impairments through making the launch power P0 should be larger than the optimum channel power. The perfect power for the proposed designed SP-QPSK and 4QAM optical system as shown in Figure (5.2).

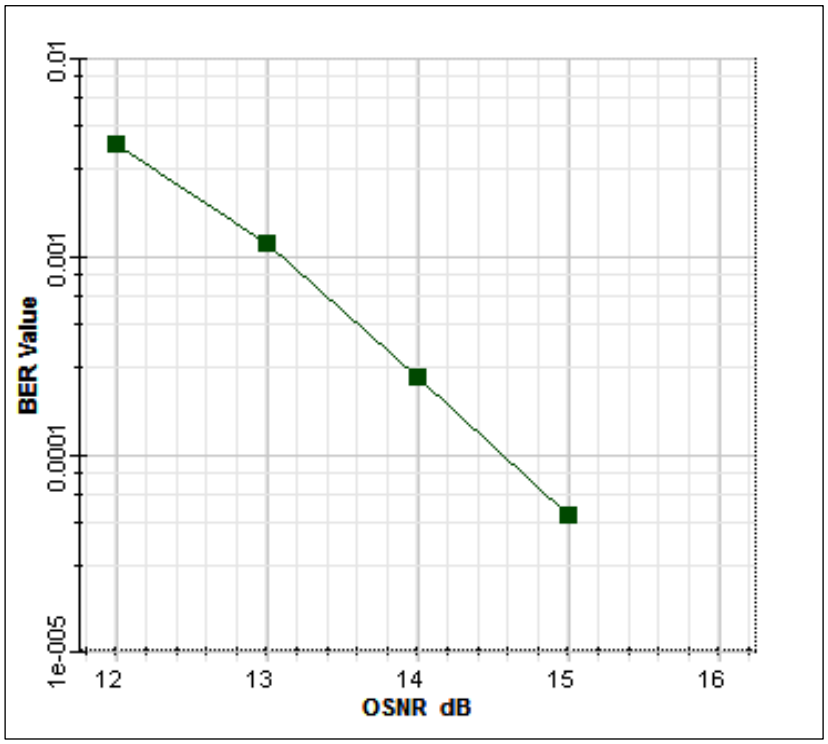
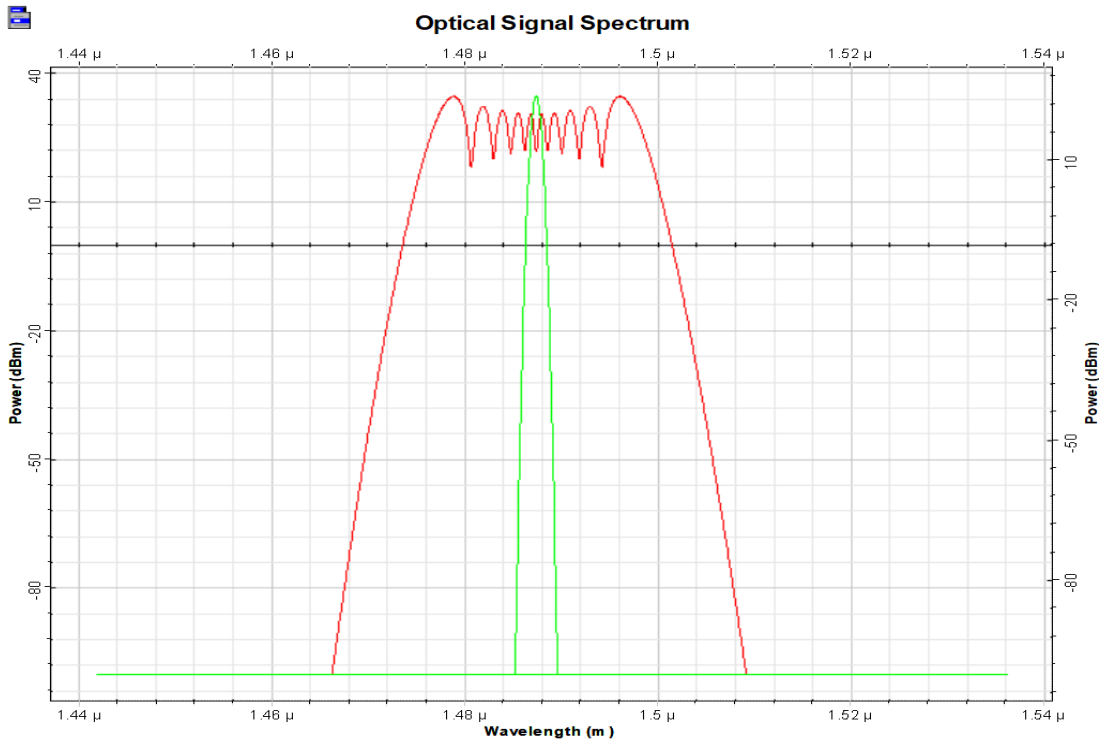
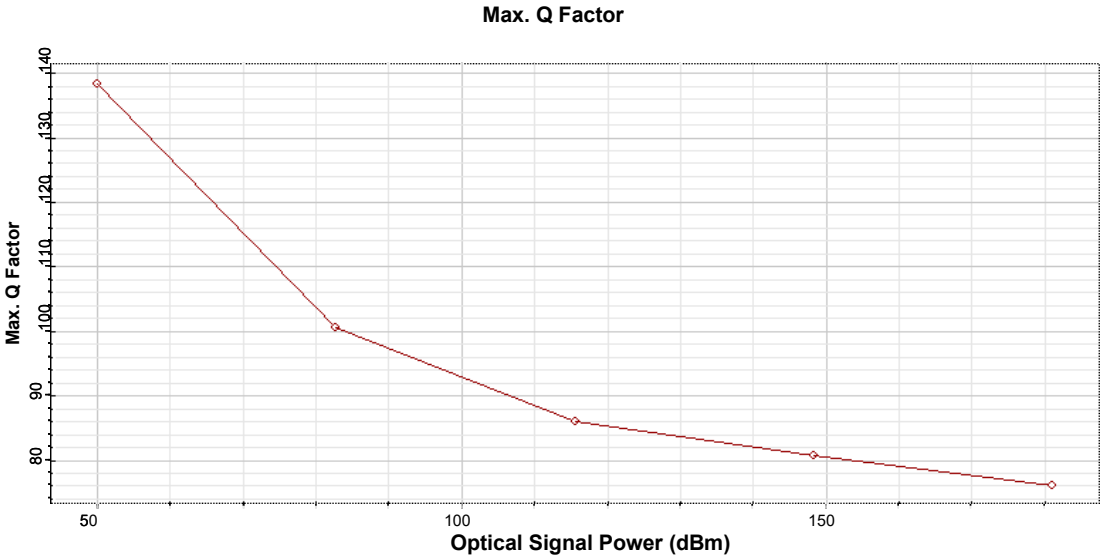


Figure (5.2): Optimum power for proposed SP-QPSK and 4QAM

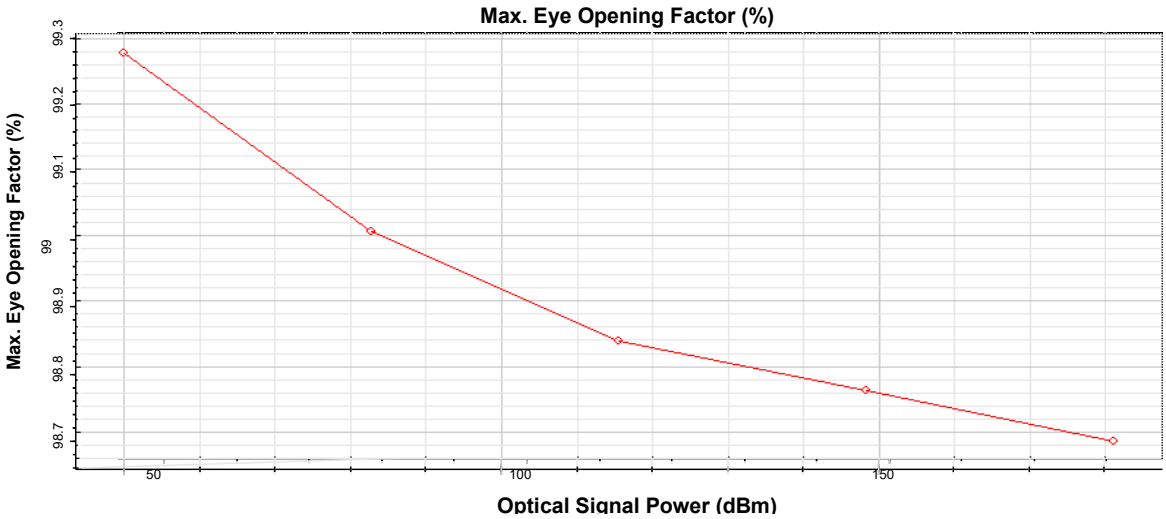
Therefore, will be work to increase the OSNR to raise the problems in the optical systems designed under test and the Figure (5.3) shows the effect of SPM in terms of spectrum broadening, Quality factor (Q), Eye-Opening Factor (EOF) and finally BER. In addition, Figure (5.4) shows the 4QAM constellations mapper at various levels of launch power.



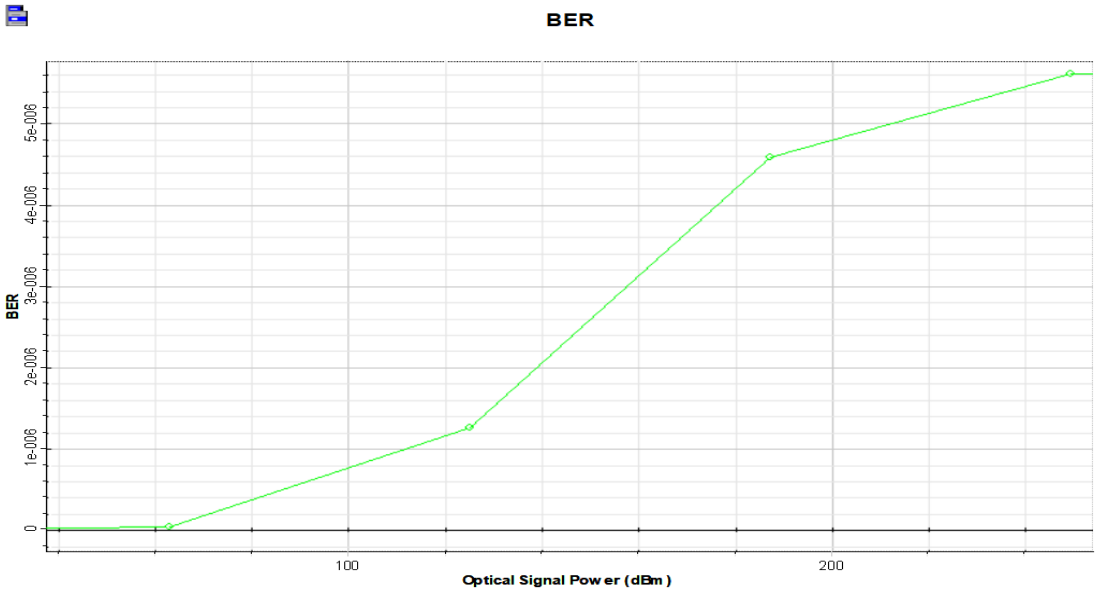
(a)



(b)



(c)



(d)

Figure (5.3): The effect of SPM as a spectrum broadening for the SP-QPSK-4QAM optical system. (a) Pulse Spectrum Broadening. (b) decrease the Max. Q Factor. (c) decrease the Max. Eye Opening Factor. (d) decrease the BER.

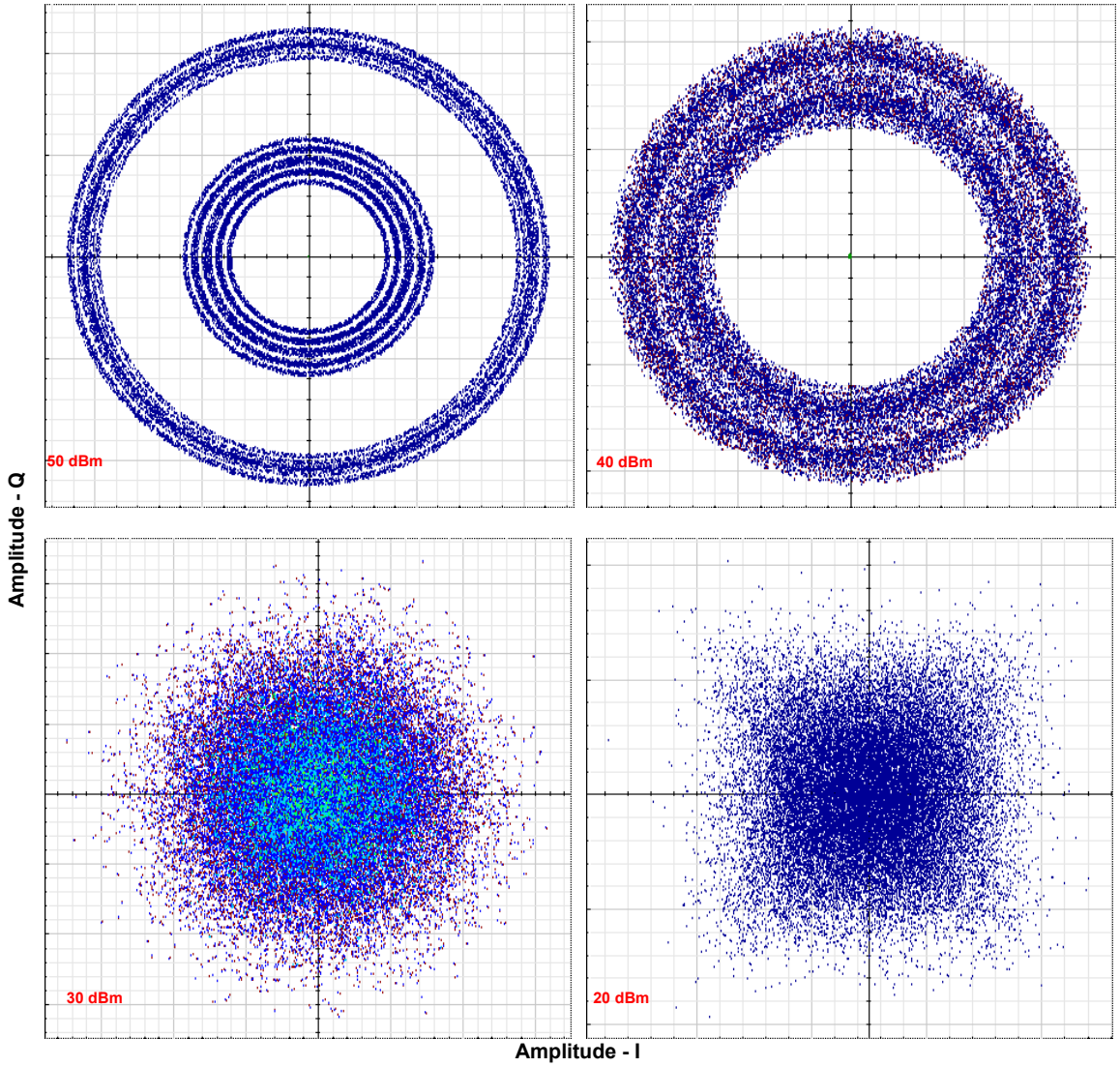


Figure (5.4): 4QAM constellations at various levels of launch power.

The proposed optical communications under test is Single-Channel SP-M-QAM @ 60 Gbps bitrate are sent over standard single mode fiber (SMF) towards coherent receiver side to demonstrate improvement after @3000 km standard. Also, will be consider N-spans of fiber links with 100km per-span. Assuming noise figure is 4dB and used Erbium-Doped Fiber Amplifiers (EDFA) to compensated span loss with all ASE noise added completely before the receiver side. EDFA is especially suited for performing the prompt analysis of performance of the amplifiers in the long-haul system. Also, used 3rd order optical filter with a Gaussian frequency transfer function at receiver side only with 100 dB Depth (maximum attenuation value for the filter). In addition, at

receiver, the incoming optical signal is coherently detected and the symbols are fed to a linear equalization (LE) for compensating Chromatic-Dispersion (CD) and channel distortions (linear dispersion). As a result of fiber non-linearity, residual distortions after LE will limits achievable rates of information. Figure (5.5) illustrates the optical spectrum analyzer for the received optical signal.

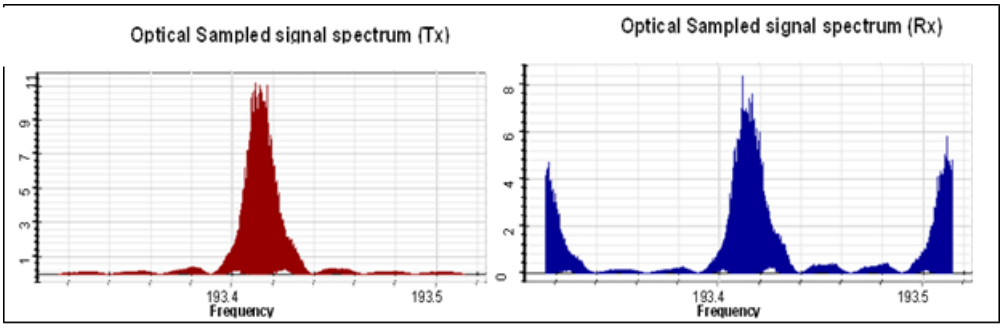


Figure (5.5): The Optical Spectrum Analyzer for the Tx & Rx 4QAM.

Figure (5.6) illustrates example of the residual optical distortion after (31) tap LE @ 30-span transmissions for 4-QAM constellation format. it is seen that constellation is distorted more seriously with an increase in the launch power as a result of fiber non-linearity.

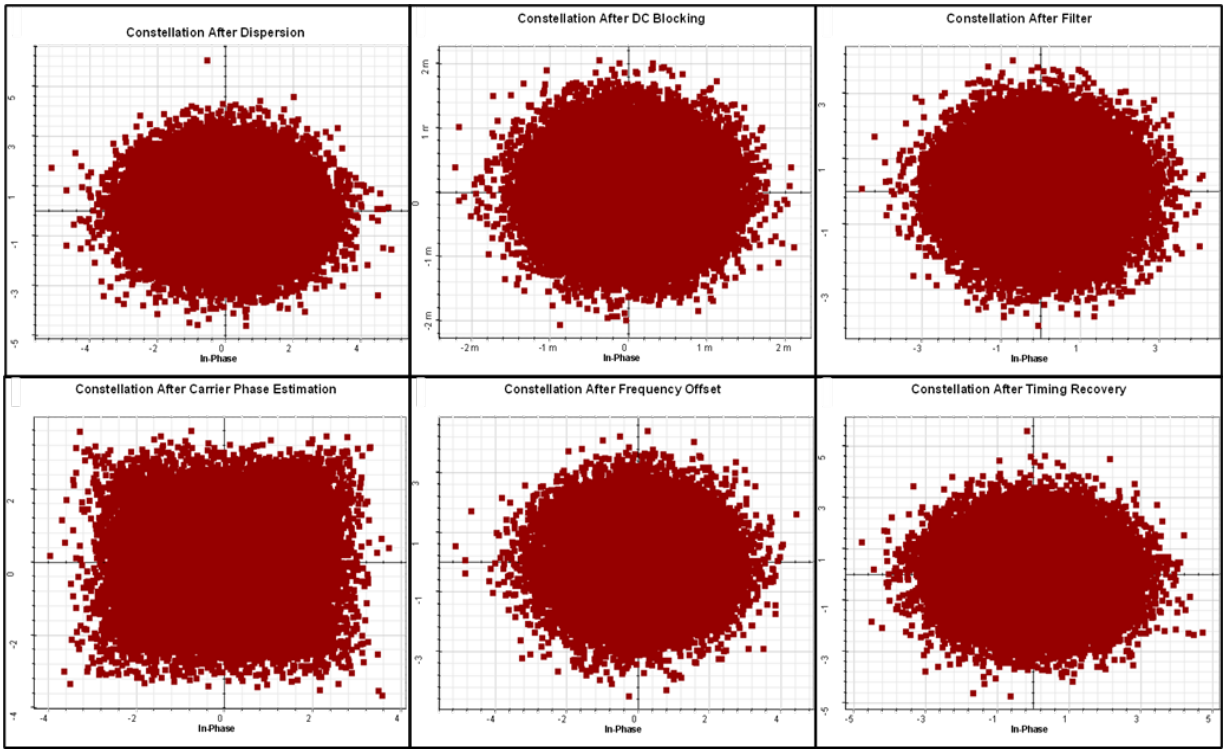


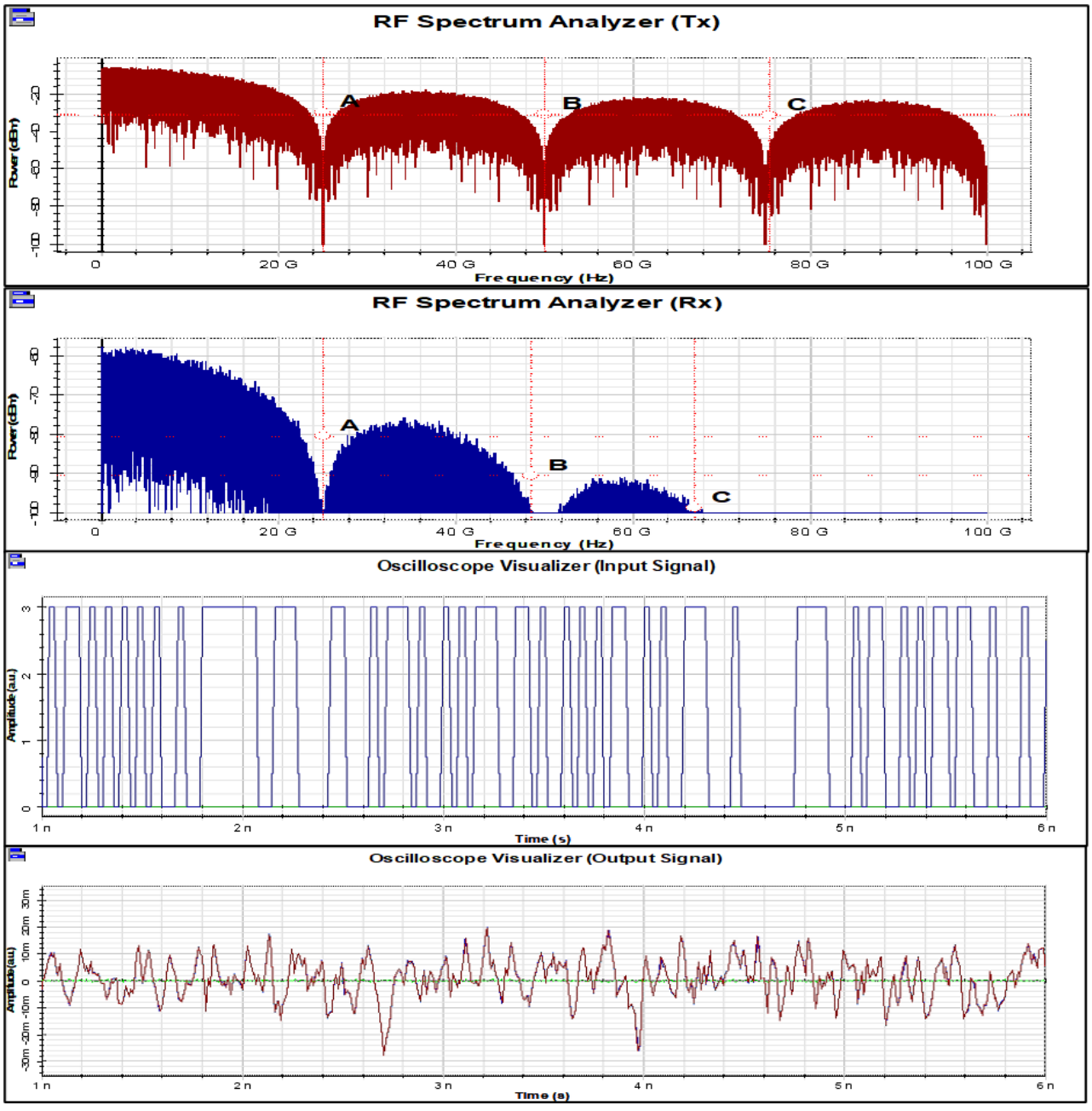
Figure (5.6): 4QAM constellation signal after 31 tap LE.

Depending on nonlinear Schrödinger equation the nonlinear parameter can be written as: $\gamma = 2\pi n_2 / \lambda A_{eff}$, with A_{eff} as the effective area, n_2 nonlinear index and λ the carrier wavelength. Two quantities can be defined to describe their relative importance they are the dispersion length (L_D), $L_D = T^2_0 / \beta_2$ and nonlinear length (L_{NL}), $L_{NL} = \frac{1}{\gamma P_0}$. The relative importance of the dispersion and the SPM effects is dependent upon the value of parameter N, ($N = L_D / L_{NL}$), When N >> 1, the SPM effects are dominant and in the case where N << 1, those of GVD dominate. For demonstrating their interplay, N=1 is chosen. The non-linear phase shift φ_{NL} is increased with the fiber length L and effective length L_{eff} for fiber of length L has been defined as $\{L_{eff} = [1-\exp(-\alpha L)] / \alpha\}$ and at make the fiber losses $\alpha = 0$, and $L_{eff} = L$. The maximal phase shift φ_{max} will occur and $\{\varphi_{max} = L_{eff}/L_{NL} = \gamma P_0 L_{eff}$, according to [3]. Table (5.1) and Figure (5.7) shows the simulation result after running the coherent 4QAM optical system, can be visualized, the Eye diagrams, RF spectrum analyzer with an oscilloscope visualizer for the transmitted/received signal with compare the input&output M-ary sequence.

Table (5.1): DNN results analysis and optimize for 4QAM optical

Parameter	Value
Tx : Min. Signal Power (dBm)	8.085147650425979
Tx : Min. Signal Power (W)	0.00643449940481577
Tx : Max. Signal Power (dBm)	8.085147650425979
Tx : Max. Signal Power (W)	0.00643449940481577
Tx : Frequency at Max. Signal Power (Hz)	193.414489032258e+012
Tx : Wavelength at Max. Signal Power (nm)	1550
Tx : Total Signal Power (dBm)	8.085147650425979
Tx : Total Signal Power (W)	0.00643449940481577
Tx : Min. SNR (dB)	24.35701703393499
Tx : Max. SNR (dB)	24.35701703393499
Tx : Frequency at Max. SNR (Hz)	193.414489032258e+012
Tx : Wavelength at Max. SNR (nm)	1550
Tx : Min. OSNR (dB)	27.37761694721462
Tx : Max. OSNR (dB)	27.37761694721462
Rx : Min. Signal Power (dBm)	-0.5008904679605624
Rx : Min. Signal Power (W)	0.0008910682167127443

Rx : Max. Signal Power (dBm)	-0.5008904679605624
Rx : Total Signal Power (dBm)	-0.5008904679605624
Rx : Min. Noise Power (dBm)	-14.28630817103477
Rx : Max. Noise Power (dBm)	-14.28630817103477
Rx : Total Noise Power (dBm)	-14.28630817103477
Rx : Min. SNR (dB)	13.7854177030742
Rx : Max. SNR (dB)	13.7854177030742
Rx : Min. OSNR (dB)	19.80601761635383
Rx : Max. OSNR (dB)	19.80601761635383
Min. Noise Figure (dB)	35.13649258213911
Max. Noise Figure (dB)	35.13649258213911



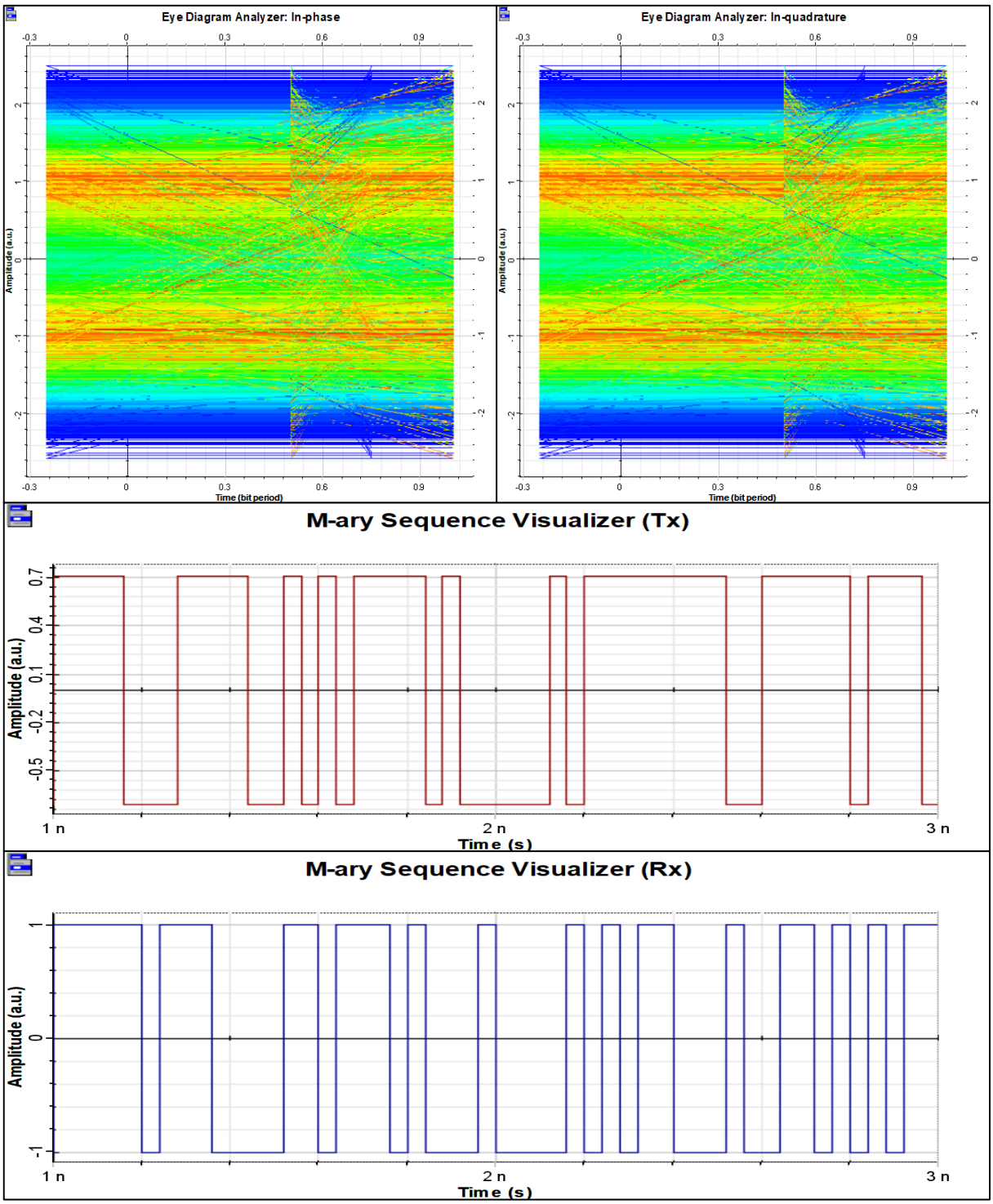


Figure (5.7): The simulation result of coherent 4QAM optical system after (31) tap LE @ 30-span transmissions

For the purpose of compensating the residual linear and non-linear distortion, the DNN-based NLC will be applied.

The DNNs-NLC ability was evaluated based on BER and correct constellation mapper decoder. Firstly, there are 2 bits per symbol (m=2) in SP-4QAM, leading to (2^2 = 4) classes to identify per symbol. The single Nonbinary SoftMax Classification Algorithm can be used but does not perform well for higher-order mapper constellation because it needed to unrealistically hugely datasets for training. So, will be used multi-label classification, which is more scalable for high-dimensional and high-order modulation. The carrier data after LE and dispersion compensation are including inter-symbol interference (ISI) with propagation distance. In addition, the optical nonlinear effects are deterministic (ISI) in this system as shown in Figure (5.7) The received data after LE and dispersion compensation are fid to the DNNs to trained by sigmoid activation function, SoftMax & Classification Output layer, an LSTM layer and A Leaky ReLU layer are used, which performs a threshold operation with ADAM optimization algorithm and cross-entropy loss function. The DNNs have 4 hidden layers with (16-128) hidden nodes. The size of the dataset used (128 x 262144) samples. The number 128 indicates that: the optical transmission system will be repeated 128 times for the same sequence input for the purpose of properly training the neural network and reach to find the perfect estimates of neurons' weights (W), where 80-85% of the dataset utilized for training and 15-20% for the testing and validation. The data-set will be separated to imaginary and real parts prior to being fed to the DNN.

Training: The aim of training is obtaining a sufficient estimation weight vector in other words, the output of final network Soft-Max layer has to be optimally similar to one-hot input vector. The NN has been trained by using Adam optimizer, mini-batch size is 250, which corresponds to 750000 iterations of the algorithm of optimization. It should be noted that in the majority of the cases, the convergence in loss and error rate of validation symbol of the trained models

has been obtained following considerably < 600000 iterations, which has been utilized as fixed termination criterion. Throughout the process of the training, the Standard deviation of σ 0.10 has been utilized to initialize weight matrices W. The vectors of the bias b have been initialized with the value of 0. Training validation has been carried out throughout the process of optimization every 5,000 iterations. Sufficient convergence of validation and corresponding Loss values has been reached. The trained model has been saved and after that, separately loaded for the testing that has been carried out over a group of various arbitrary input messages. Optical Visualizer results from the testing have been illustrated in figures over this section. Convergence of results has been confirmed, and for the mini-batch sizes of 500 and 100. When designing DNNs, the option of the hyper-parameters like number of the layers, activation functions, number of the nodes in the layer, mini-batch size, learning rate, and so on, is of a high importance, and in this work the choice was made by experimentation, experience, and prior knowledge of the problem with the aim of keeping the networks rather small and therefore, the efforts of the training controllable. More sufficient results in the terms of the efficiency as well as its trade-offs with the complexity may be reached with the well-modelled hyper-parameter set.

The efficiency for different hyperparameters is studied as shown in Figures (5.8) and (5.9). The performance of hidden units' number, over sampling and the size of dataset will be taken at two transmissions lengths (2000-3000) km at (20-30 spans).

Hidden units' number: Figure (5.8) (a and d) shows, at 64 hidden units the system is sufficient for neural network data modeling, but with 32 hidden units performance deteriorates, and no significant improvement of the performance has been accomplished with more than 128 of the hidden units.

Sampling rates: Figure (5.8) (b and e) shown, from 8 to 16 sample/symbol are

appropriate for capturing the information that has been provided by signal according to the transmission distance, and no noticeable enhancement of the performance is achieved with over 16 sampling rates.

Data set size: Figure (5.8) (c and f) shows, in more than 2^18 training data samples, will result in little performance improvement.

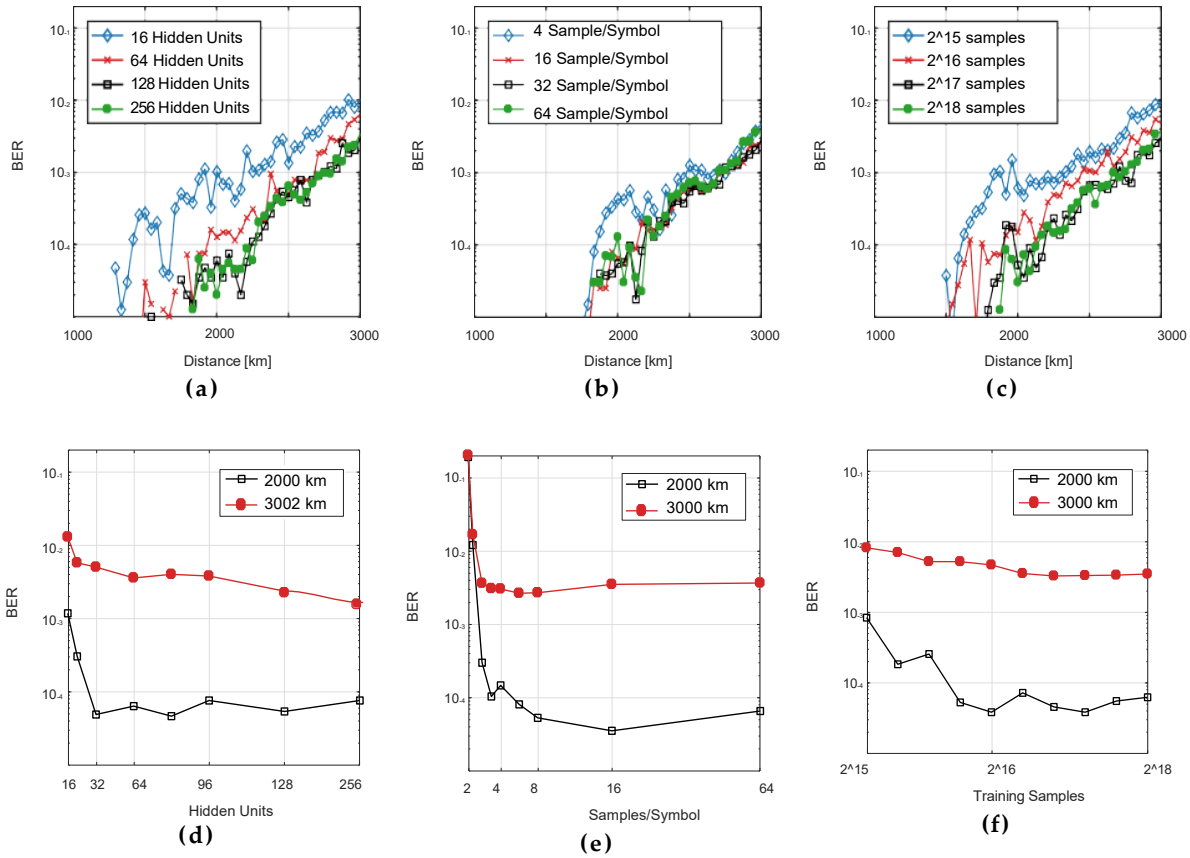


Figure (5.8): Deep NN receiver BER performance at a variety of the hyperparameters.

Figure (5.9) illustrates the learning process performance of the inphase and quadrature components with four various batch size value, related to numbers of iterations. It can be seen that the converge is fastest in the small mini-batch size at training, which has led to the rapid convergence but the final performance is somewhat worse, but the speed of the learning is a little slower, especially at the beginning when used a large training batch size, while giving a higher final performance with longtime training process.

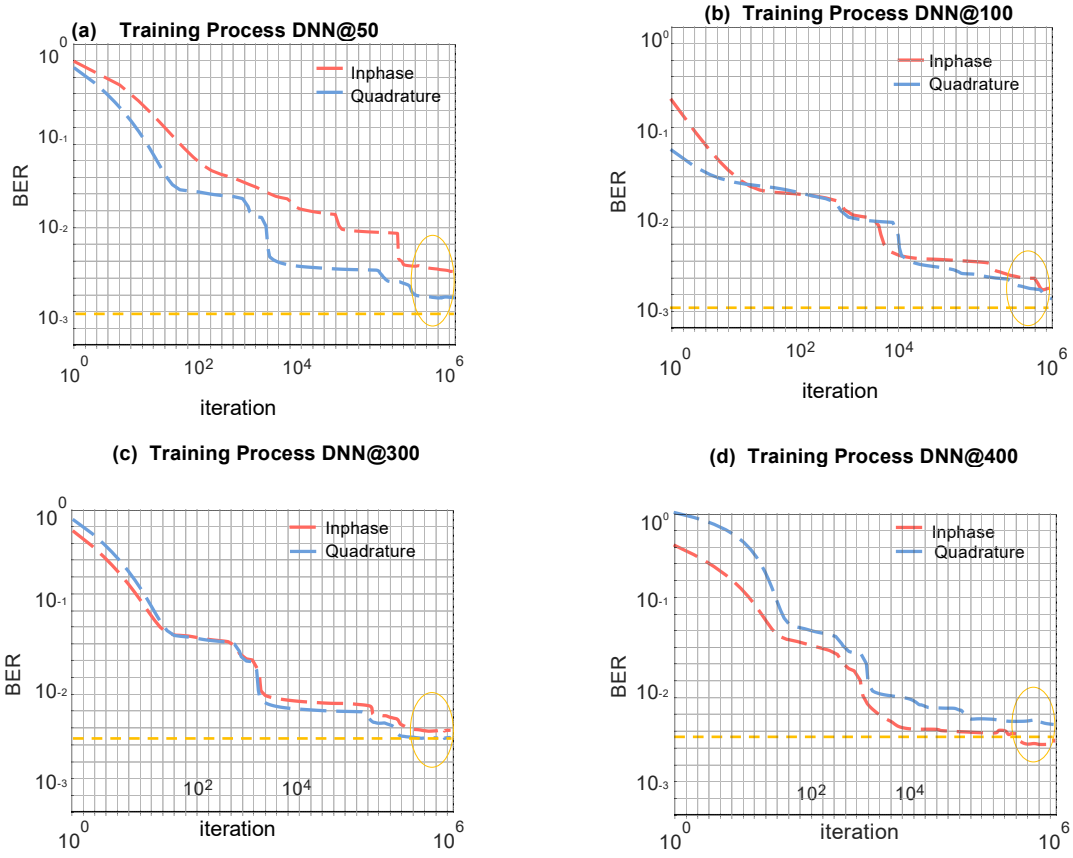


Figure (5.9): Training Process Performance for Various Batch Size Value.

The longtime with high consumption of computer resources at training process is one of the main drawbacks of the deep neural network model and it requires further refinement, optimization and research [69].

In addition, Tables (5.2), (5.3) and (5.4) shows the DNN-NLC model performance

Table (5.2)**:** DNN performance @ 10^6 iteration for different batch size.

Batch size	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
50	0.98**04**	0.**0196**	0.98**0**	0.**02**
100	0.98**59**	0.**0141**	0.98**30**	0.**017**
300	0.98**52**	0.**0141**	0.98**14**	0.**0186**
400	0.98**13**	0.**0185**	0.97**58**	0.**0242**
500	0.98**99**	0.**0180**	0.98**14**	0.**0186**

Table (5.3): DNN performance @ 100 batch size for different iterations.

iterations	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
10^2	0.**6**519	0.**3**181	0.**6**524	0.1236
10^3	0.**6**636	0.**30**24	0.**6**604	0.1028
10^4	0.**81**34	0.**1**900	0.**8**614	0.1078
10^5	0.9**80**4	0.**0196**	0.9**80**	0.**02**
10^6	0.989**3**	0.0**01**	0.9851	0.0**0**18

Table (5.4)**:** DNN performance @ (10^6 iteration) and (100 batch size) for different learning rate.

Learning rate	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
0.001	0.98**99**	0.0**01**	0.986	0.0**0**17
0.0001	0.9**528**	0.**02**37	0.900**3**	0.**0**469
0.0016	0.953**1**	0.**0**123	0.951**3**	0.**0**157
0.0004	0.9508	0.1178	0.930**6**	0.1703

As a result, the small mini-batch size at training results in the rapid convergence but the ultimate performance is somewhat worse, and when used a large size of the training batch, it resulted in slower convergence, however, a higher final performance. A valid tradeoff between the performance, convergence, and computation time has been obtained in the case of training the DNN model at started with small size of the batch and increased it after the initial convergence.

Whatever the case of the constellation, the number of layers, the number of the hidden nodes, the size of the DNN model depends on the order of the constellation.

Finally, the hyperparameters for training the proposed DNN model are:

Layers: 7

Hidden nodes/layer: 64-128

Activation function: ReLU

Learning rate: 0.001

Training batch size: Adaptive

Optimization approach: Adam

Mini-Batch size=100

Epochs=5000 (100 iteration per epoch)

Finally result shown in Figure (5.10),

for Training: Accuracy: 0.9893, loss: 0.001

for Testing: Accuracy: 0.9851, loss: 0.0028.

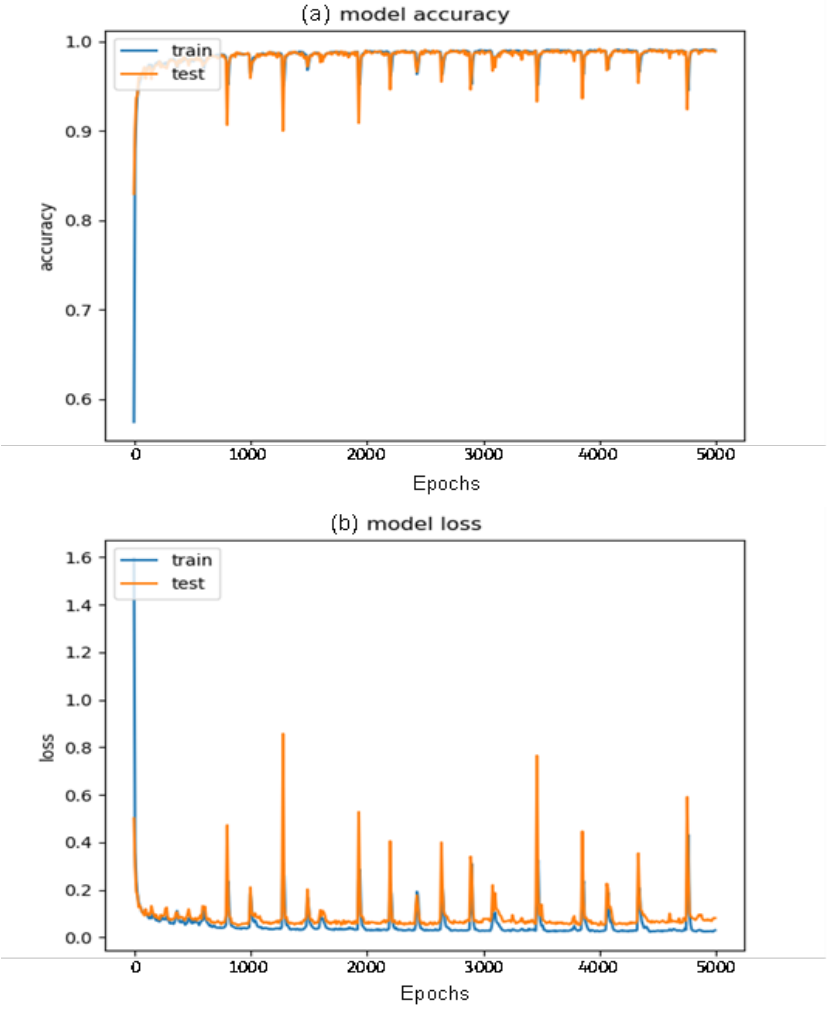


Figure (5.10): The Accuracy & Loss Curve of DNN Proposed Model.

5.2.2 Joint Optimization of DNN-NLC & OptiSystem Parameters

The results shown in Figure (5.11) after training and testing the network proved that the DNNs based on classification output layer is capable of classifying the received symbol for the reference target 4-QAM and QPSK and at iteration (200,000) the result of constellation mapper is very clear although the network is classified the symbols correctly, but the bit locations were still incorrect.

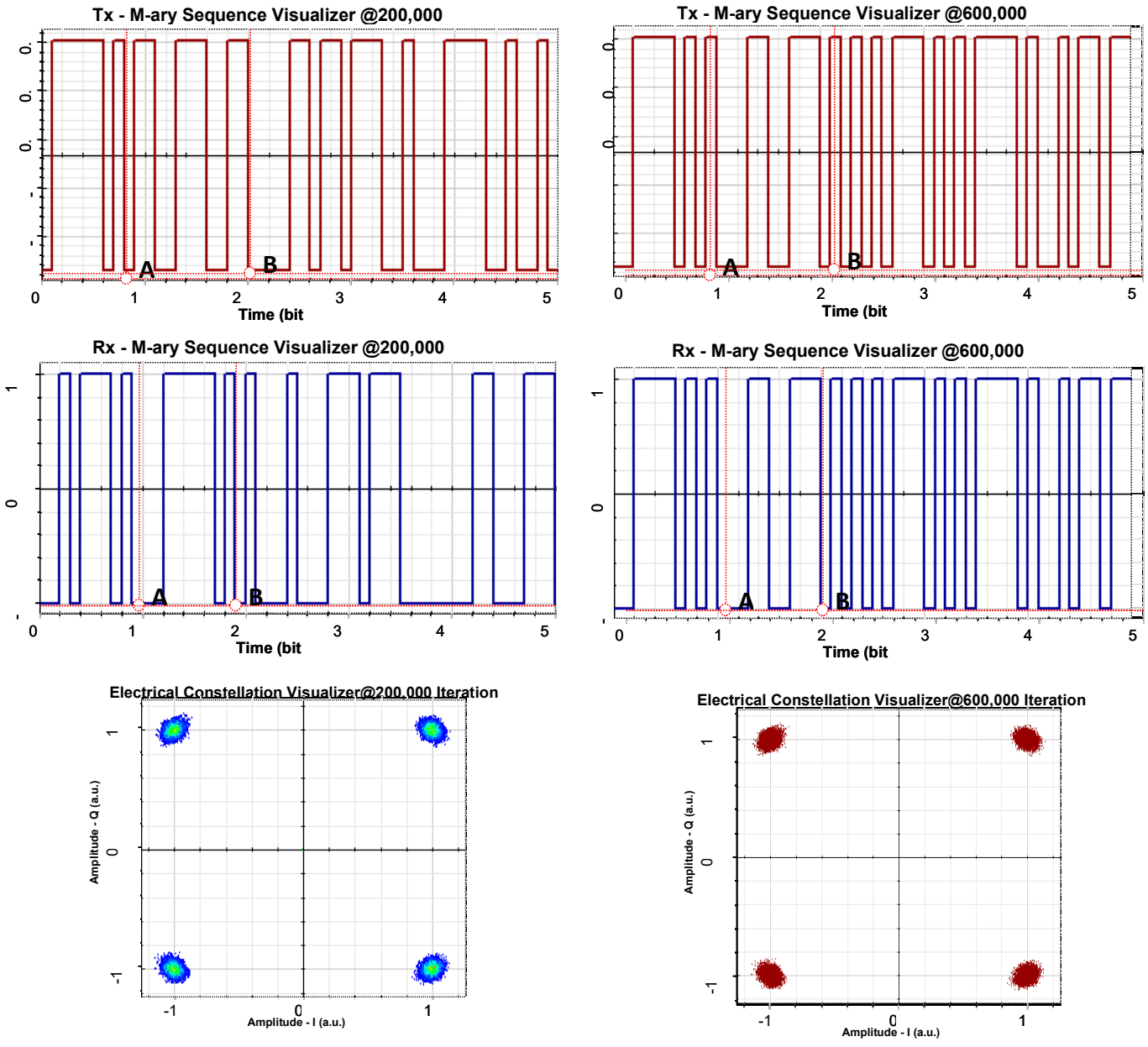


Figure (5.11): Sequence Visualizer testing after 200,000 &600,000 iterations.

So, when DNNs cost losses began to decrease about (~0) and the validation accuracy rises to nearly 100% @iteration (600,000). The Q-factor for the system at testing is equal to (~13.4 dB) as shown in Figure (5.12) for the one polarization 4QAM coherent optical BtB 60 Gb/s @ (100 x 30span) Km measurements.

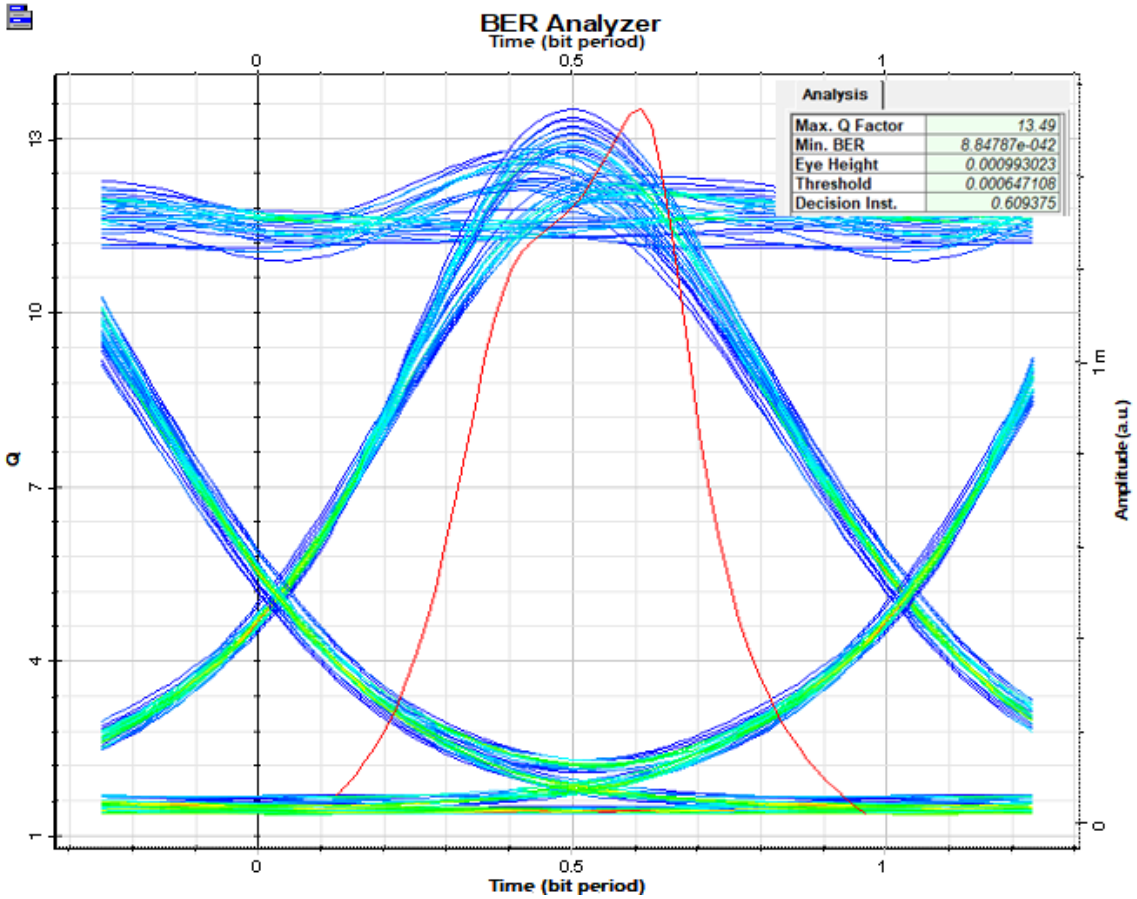


Figure (5.12): BER and Q-factor analyzer for the one polarization 4QAM coherent optical with 60 Gb/s bitrate @ (100 x 30span) Km measurements.

5.3 SP - CO_OFDM

The optical system model under study is 16-QAM, CO-OFDM, single-polarization, 120 Gbps bitrate is sent over fiber media towards coherent receiver side to demonstrate improvement after @3000 km standard single mode fiber (SMF) transmission.

N-span of fiber links will consider with 100 km per-span, the SMF has a nonlinear factor is i = 2 /W/km and dispersion parameter of D = 16 ps/nm/km, with an attenuation of 0.2 dB/km. we assuming the noise figure is 4 dB and used Erbium-Doped Fiber Amplifiers (EDFA) to compensated the span loss with all ASE noise added completely before the receiver side. EDFA is especially suited for performing the prompt performance analysis of the amplifiers in the long-haul systems. Also, used 3 order optical filter with a Gaussian frequency transfer function at receiver side only with 100 dB Depth (maximum attenuation value for the filter).

In addition, at the receiver, the incoming OFDM signal is coherently detected and after removing Cyclic Prefix (CP) and taking Fast Fourier Transform (FFT), the symbols are fed to a Linear equalizer (LE) to compensate the Chromatic-Dispersion (CD) and channel distortions (linear dispersion) based on the parameters of training & pilot symbols for the proposed CO-OFDM system, to estimate the Inter-Carrier-Interference (ICI) effects of Self-Phase Modulation (SPM), Four-wave-mixing (FWM) and Cross-Phase Modulation (XPM) due to its high Peak to Average Power Ratio (PAPR) [19]. The Kerr-mediate nonlinear process is deterministic (ISI) in multicarrier schemes like CO-OFDM. the residual distortion of SP-16QAM-CO_OFDM after LE for (30span *100Km) shown in Figure (5.13), which shows the more serious constellation deformation with increased the launch power (P_0) due to Kerr-nonlinearity. In addition, Figure (5.14) shows the signal visualizer of SP-16QAM-CO-OFDM system after (100 x 30span) SMF links and Table (5.5) optical system results analysis for 4QAM

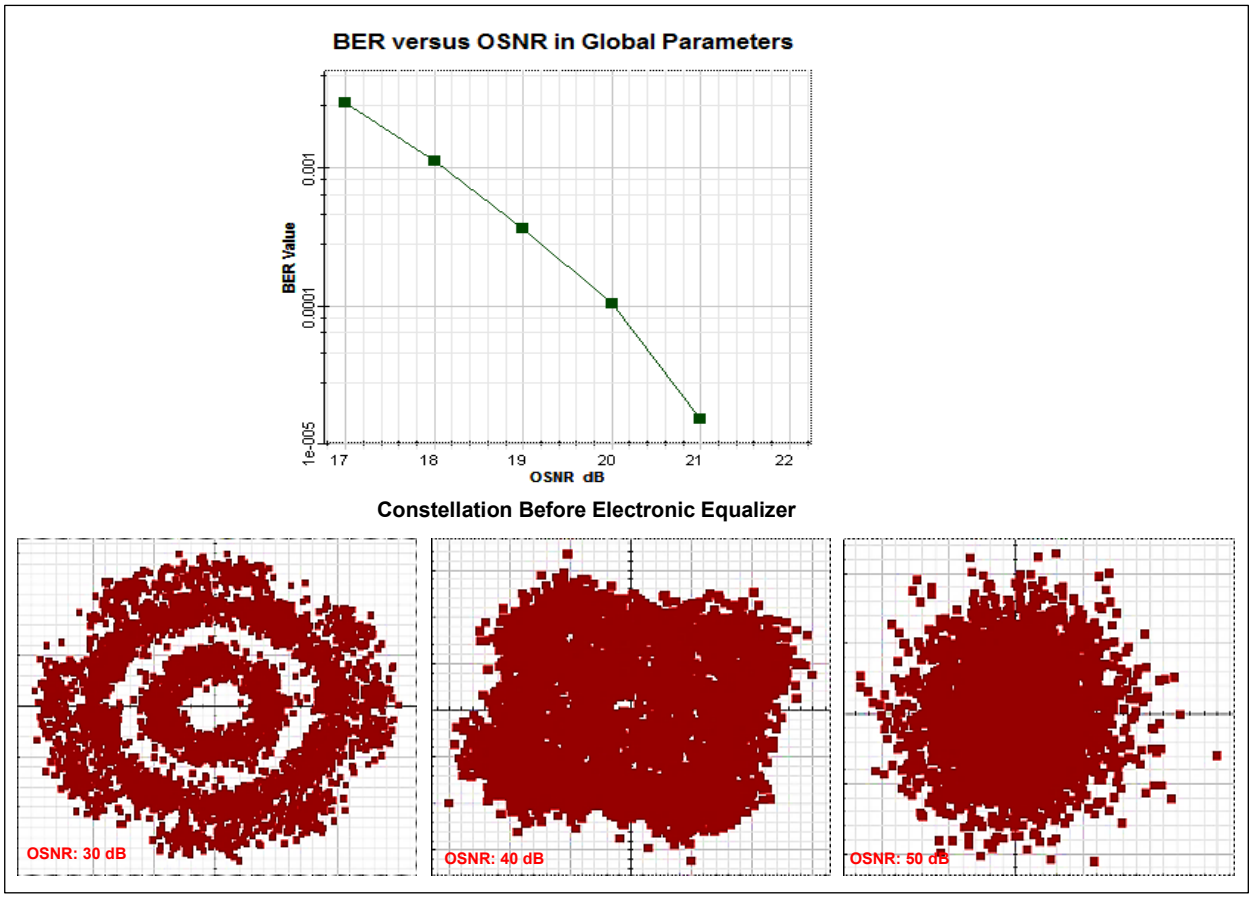
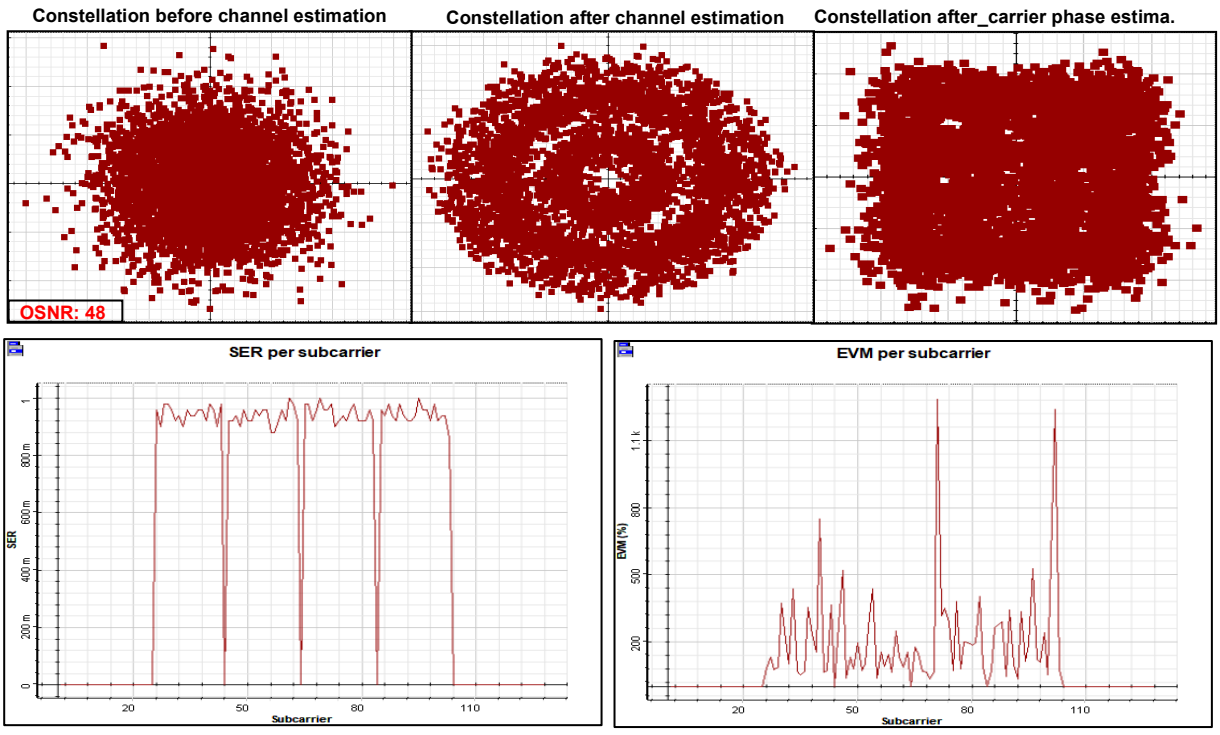


Figure (5.13): Residual distortion of SP-16QAM-CO-OFDM after LE for (100 x 30span) SMF links.



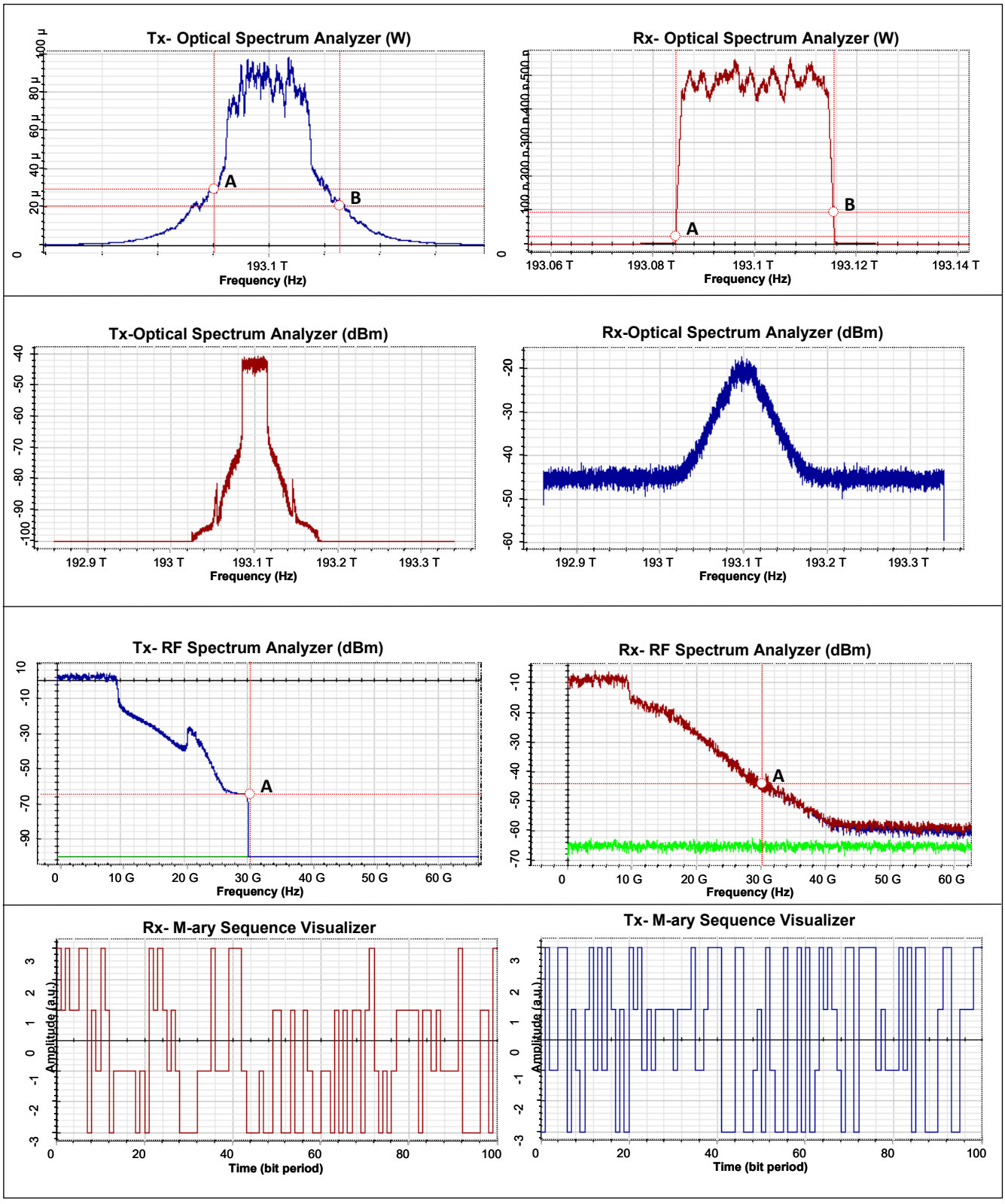


Figure (5.14): Visualizer of SP-16QAM-CO-OFDM system after (100 x 30span) SMF links.

Table (5.5): Optical System Results Analysis for 4QAM

Parameter	Value
Input : Min. Signal Power (dBm)	8.87979725499749
Input : Min. Signal Power (W)	0.00772644514301097
Input : Max. Signal Power (dBm)	8.87979725499749
Input : Max. Signal Power (W)	0.00772644514301097
Input : Total Signal Power (dBm)	8.87979725499749
Input : Total Signal Power (W)	0.00772644514301097
Input : Min. Noise Power (dBm)	-32.19087148526755
Input : Min. Noise Power (W)	0.6038274490368679e-006
Input : Max. Noise Power (dBm)	-32.19087148526755
Input : Max. Noise Power (W)	0.6038274490368679e-006
Input : Frequency at Max. Noise Power (Hz)	193.1e+012
Input : Wavelength at Max. Noise Power (nm)	1552.524381149663
Input : Total Noise Power (dBm)	-32.19087148526755
Input : Total Noise Power (W)	0.6038274490368679e-006
Input : Min. SNR (dB)	41.07066874026504
Input : Max. SNR (dB)	41.07066874026504
Input : Min. OSNR (dB)	47.88308111402091
Input : Max. OSNR (dB)	47.88308111402091
Output : Min. Signal Power (dBm)	-7.120942096983293
Output : Min. Signal Power (W)	0.0001940464894942533
Output : Max. Signal Power (dBm)	-7.120942096983293
Output : Max. Signal Power (W)	0.0001940464894942533
Output : Frequency at Max. Signal Power (Hz)	193.1e+012
Output : Wavelength at Max. Signal Power (nm)	1552.524381149663
Output : Total Signal Power (dBm)	-7.120942096983293
Output : Total Signal Power (W)	0.0001940464894942533
Output : Min. Noise Power (dBm)	-45.57995708949314
Output : Min. Noise Power (W)	27.66968984339347e-009
Output : Max. Noise Power (dBm)	-45.57995708949314
Output : Max. Noise Power (W)	27.66968984339347e-009
Output : Total Noise Power (dBm)	-45.57995708949314
Output : Total Noise Power (W)	27.66968984339347e-009
Output : Min. SNR (dB)	38.45901499250985
Output : Frequency at Min. SNR (Hz)	193.1e+012
Output : Wavelength at Min. SNR (nm)	1552.524381149663
Output : Max. SNR (dB)	38.45901499250985
Output : Min. OSNR (dB)	45.27142736626573
Output : Max. OSNR (dB)	45.27142736626573
PAPR	7.805024563427504
PAPR (dB)	8.923742741818105

5.3.1 DNN-NLC For SP - CO-OFDM

For the purpose of compensating the residual linear and non-linear distortion, the DNN-based NLC will be applied.

keep in mind, The DNN algorithm needs huge data to realize a working model of optical nonlinear weakness, also it is necessary to provide the DNN with nonlinear impairment features the firing launch power P_0 must be maximum and greater than optimum transmutation system power to obtain and have sufficient nonlinearity, these features are provided to the DNN through the intrachannel four-wave mixing (IFWM) and the first calculating of the intrachannel cross-phase modulation (IXPM) and self-phase modulation (ISPM).

Also, it consumes a high amount of hardware resources to do the calculations with longtime consumption at the training process which makes it suitable for future optical communications with further refinement, optimization, and research.

The proposed DNN-NLC model is treated as a supervised nonlinear problem. A DNNs learns by the training processes on the dataset. The neuron represents an independent processing unit which is determined by the input parameters, the Activation Functions (AF), the weights (W), and the biases, which applied to find the output required; each neuron contains a number of inputs (associated with a weight) and one output. to obtain the neuron output, the neuron calculates the AF parameter by sum the input weight and the bias. This output of AF must be identical to the application and represents the general effective functional framework of the target signals (differ from the AF in the hidden layers). The NLC signal has been obtained at the DNN output according to the input parameters and non-linear activation functions. So, for the purpose of evaluating the performance of DNNs, the loss function must be introduced to

find the perfect estimates neurons' weights (W), so that, the output of the DNNs model must be very close to the intended and desired target. The cross-entropy function was used as a loss function and to determine the error between the output of DNN model and the target values. The desired objective with the principle of machine learning is to find the minimum error value through finding the perfect estimates neurons' weights for the DNNs. This done by finding the highest result of the Joint Conditional Probability Density Function and the lowest value of the loss function. These values correspond to finding the ideal ML estimates neurons' weights (W) for the DNNs. So, in order to reduce the loss function and get the best performance of DNNs, the Gradient Descent Algorithm is applied to find the perfect estimates neurons' weights. This is the common and efficient method which was used, and called "Backpropagation Algorithm".

The Leaky ReLU non-linear activation function is used in the hidden layers because it achieves better performance than linear function [alaa]. Adam learning algorithm with an initial learning rate of 0.001 and initial max epochs size of b = 5000, initial mini batch size is 128, Standard deviation σ 0.1, and learn rate drop factor is 0.9.

The DNNs-NLC model's ability was evaluated based on BER and correct constellation mapper decoder. Firstly, there are 4 bits per symbol (m=4) in SP-16QAM, leading to (2^4 = 16) classes to identify per symbol. The received data after LE with channel, phase, and carrier estimation are including inter-symbol interference (ISI) with propagation distance and it fid to the DNNs to trained by sigmoid activation functions, SoftMax & Classification Output layer, an LSTM layer and a Leaky ReLU layer are used, which performs a threshold operation with ADAM optimization algorithm and the cross-entropy loss function. The DNNs have 4 hidden layers with (64-128) hidden nodes. The size of the dataset used (128 x 262144) samples, where 85% of the dataset utilized for training and

15% for the testing and validation. The data-set will be separated to imaginary and real parts prior to being fed to the DNN.

Training: The same scenario for training process that was previously shown and explained in Section (5.2) will follow in this optical transmission system.

The option of the hyperparameters, like number of the layers, activation functions, number of nodes in the hidden layer, learning rate, size of the mini-batch, and so on. is important, and in this work the choice was made by experimentation, experience, and prior knowledge of the problem with an aim to keep networks rather small and therefore, efforts of the training controllable.

The performance for different hyperparameters is studied as shown in Tables (5.6), (5.7) and (5.8).

Table (5.6): DNN performance @ 5000 Epoch for different batch size.

Mini-Batch Size	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
50	0.7608	0.2447	0.7014	0.3263
75	0.8007	0.071	0.8914	0.0874
100	0.9146	0.0218	0.9014	0.0328
128	0.9815	0.0173	0.9720	0.0240
256	0.9972	0.0140	0.9835	0.0215

Table (5.7): DNN performance @ 128 mini-batch size for different iterations.

Epoch	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
1000	0.**8**819	0.**3**181	0.**8**524	0.4236
2000	0.**8**600	0.**30**24	0.**8**004	0.1028
3000	0.9103	0.0225	0.9071	0.0342
4000	0.9704	0.0224	0.9401	0.0276
5000	0.9817	0.0183	0.9710	0.0270

Table (5.8): DNN performance @ (5000 Epoch, 128 batch size) for diff. learning rate.

Learning rate	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
0.001	0.9321	0.0200	0.9043	0.0203
0.0001	0.9788	0.0561	0.9309	0.0551
0.0016	0.9839	0.0196	0.9711	0.0216
0.0004	0.7917	0.3517	0.8056	0.4036

The performance of hidden units' number, over sampling number, the size of dataset, and the size of mini-batch size (MBS) will be studied at 5000 Epoch, 128 batch size and 0.001 learning rate as shonw in Figure (5.15)

Hidden units' number: Figure (5.15a) shows, at 96 hidden units the system is sufficient for neural network data modeling, but with 32and 64 hidden units the efficiency deteriorates, and no significant performance enhancement was accomplished with over 128 hidden units compared to complexity level that occurs.

Sampling rates: Figure (5.15b) shows, from 4 to 16 sample/symbol are appropriate for the capturing of information and no noticeable enhancement performance achieved with more than 16 sampling rates.

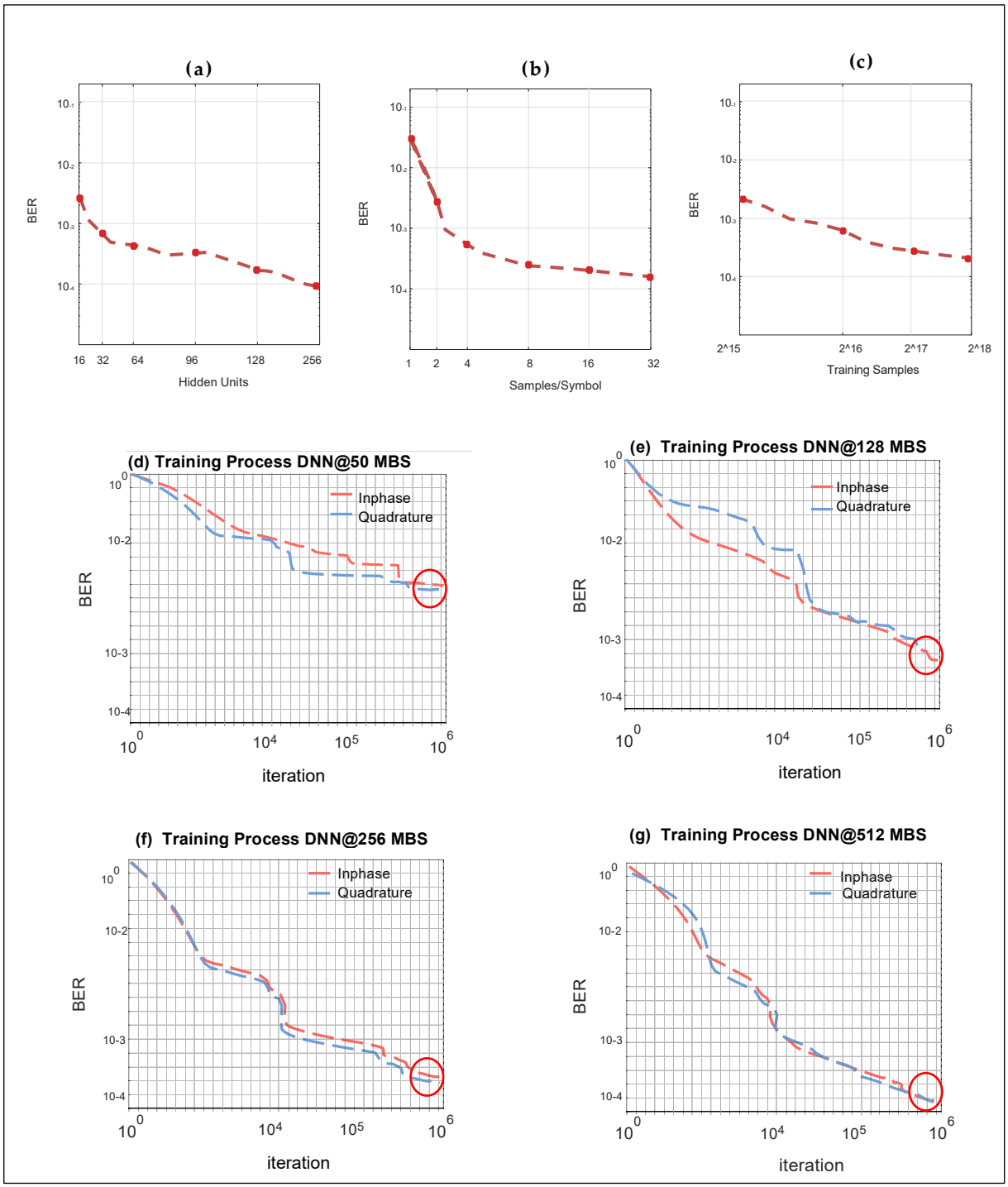


Figure (5.15): Training Process Performance for various DNN-NLC hyperparameter.

Data set size: Figure (5.15c) shows, in more than 2^18 training data samples, will result in little performance improvement.

As a result, the small mini-batch size at training leads to rapid convergence but the final performance is somewhat worse, and when used a large training mini batch size, it resulted in slower convergence with longtime training process but

higher final performance. A valid tradeoff between the performance, convergence, and computation time is obtained when training the DNN model at started with small size of the batch and increased it after the value of the initial convergence.

Finally, the hyperparameters for training the proposed DNN model are:

Layers: 7

Hidden nodes/layer: 64-128

Activation function: ReLU

Rate of Learning:0.0016

Size of the training batch: Adaptive (128-256)

Optimization approach: Adam

Mini-Batch size=128

Epochs=5000 (100 iteration per epoch)

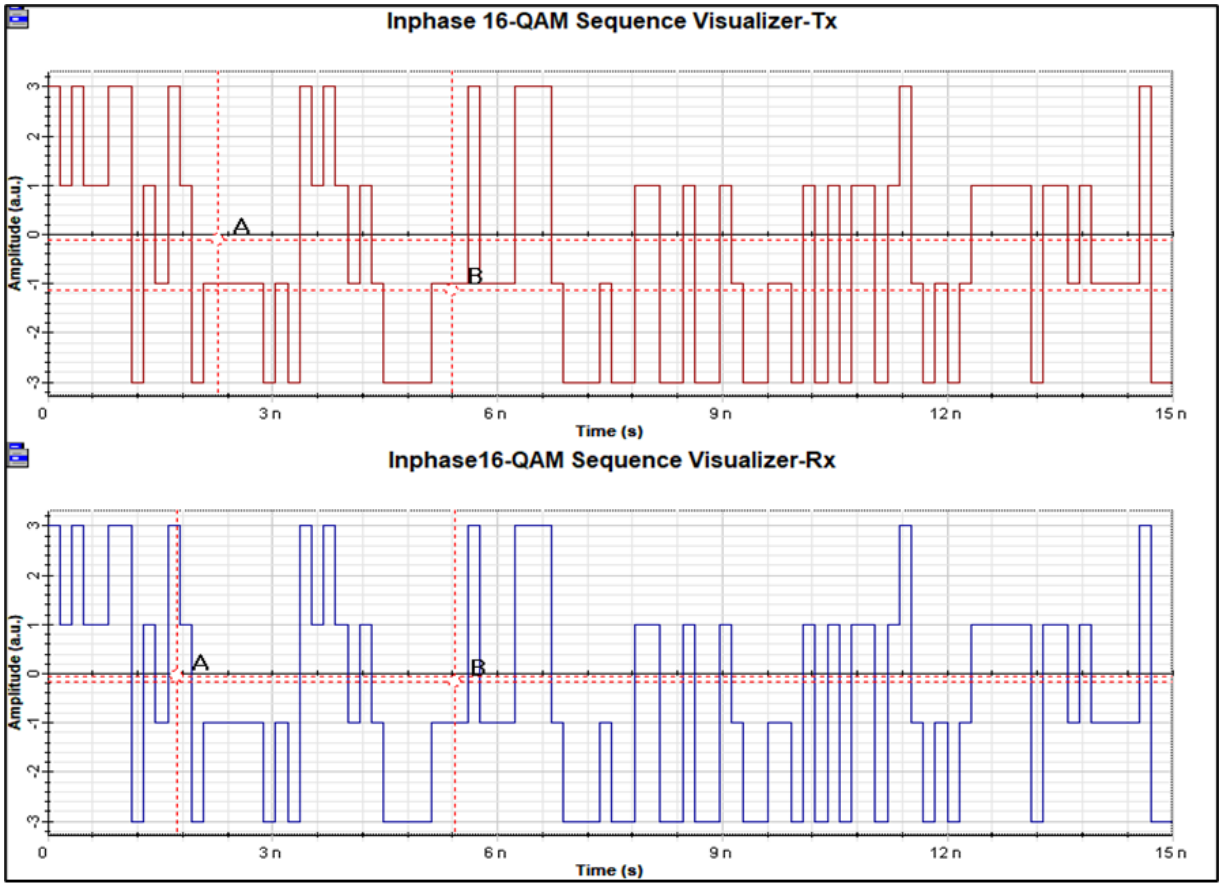
Finally result:

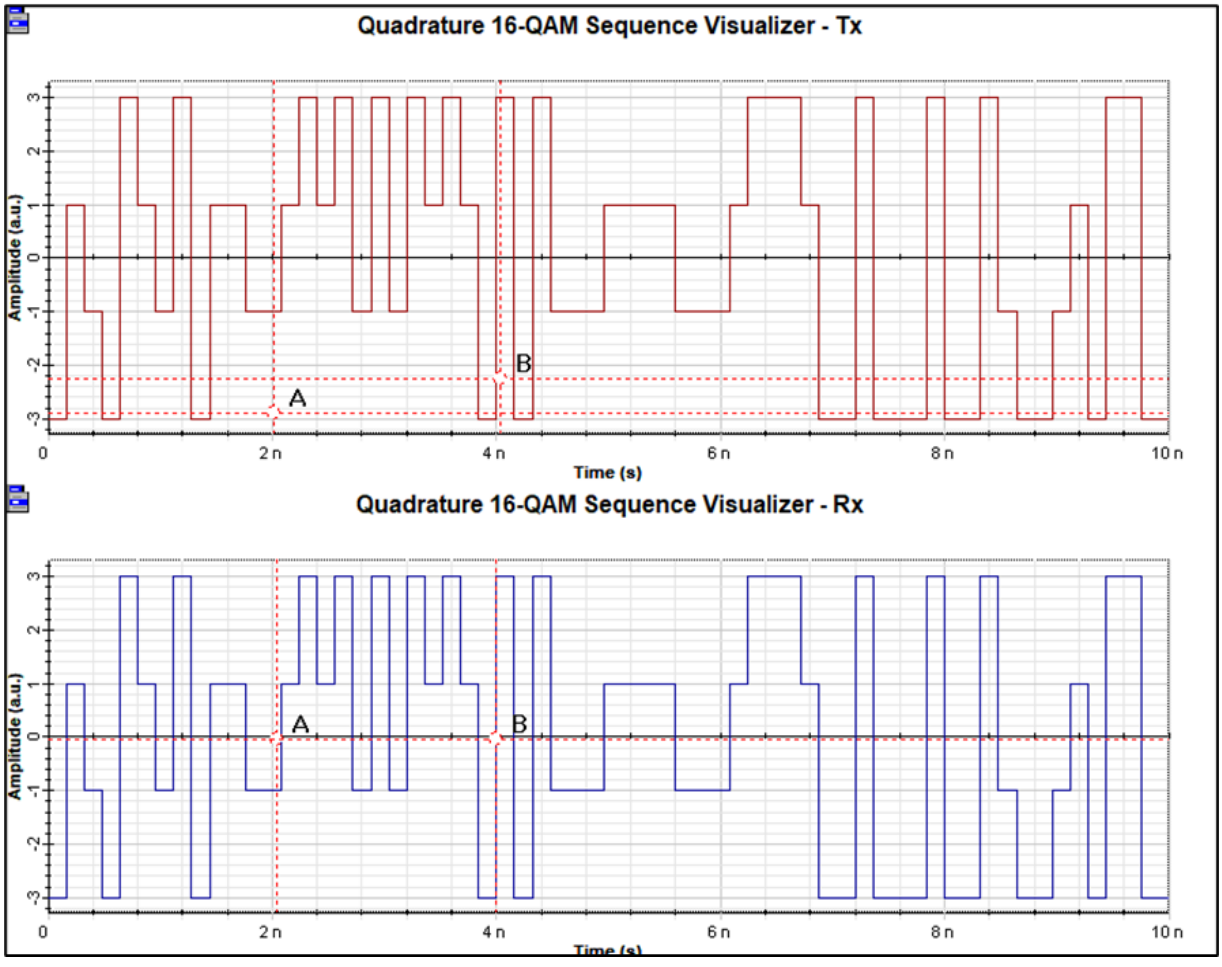
for Training: Accuracy: 0.9893, loss: 0.0186

for Testing: Accuracy: 0.9851, loss: 0.0206.

5.3.2 Joint Optimization of DNN-NLC & OptiSystem Parameters

The results shown after training and testing the network proved that the DNNs based on classification output layer is capable of classifying the received symbol for the reference target 16-QAM and the Q-factor visualizer testing for the optical system at testing is shown in Figure (5.16) and Figure (5.17) for the one polarization 16QAM coherent optical OFDM BtB 120 Gb/s @ (100 x 30span) Km link measurements.


Inphase 16-QAM Sequence Visualizer-Tx
Amplitude (a.u.)
Time (s)
A
B
Inphase16-QAM Sequence Visualizer-Rx
Amplitude (a.u.)
Time (s)
A
B


Quadrature 16-QAM Sequence Visualizer - Tx
Amplitude (a.u.)
Time (s)
A
B
Quadrature 16-QAM Sequence Visualizer - Rx
Amplitude (a.u.)
Time (s)
A
B

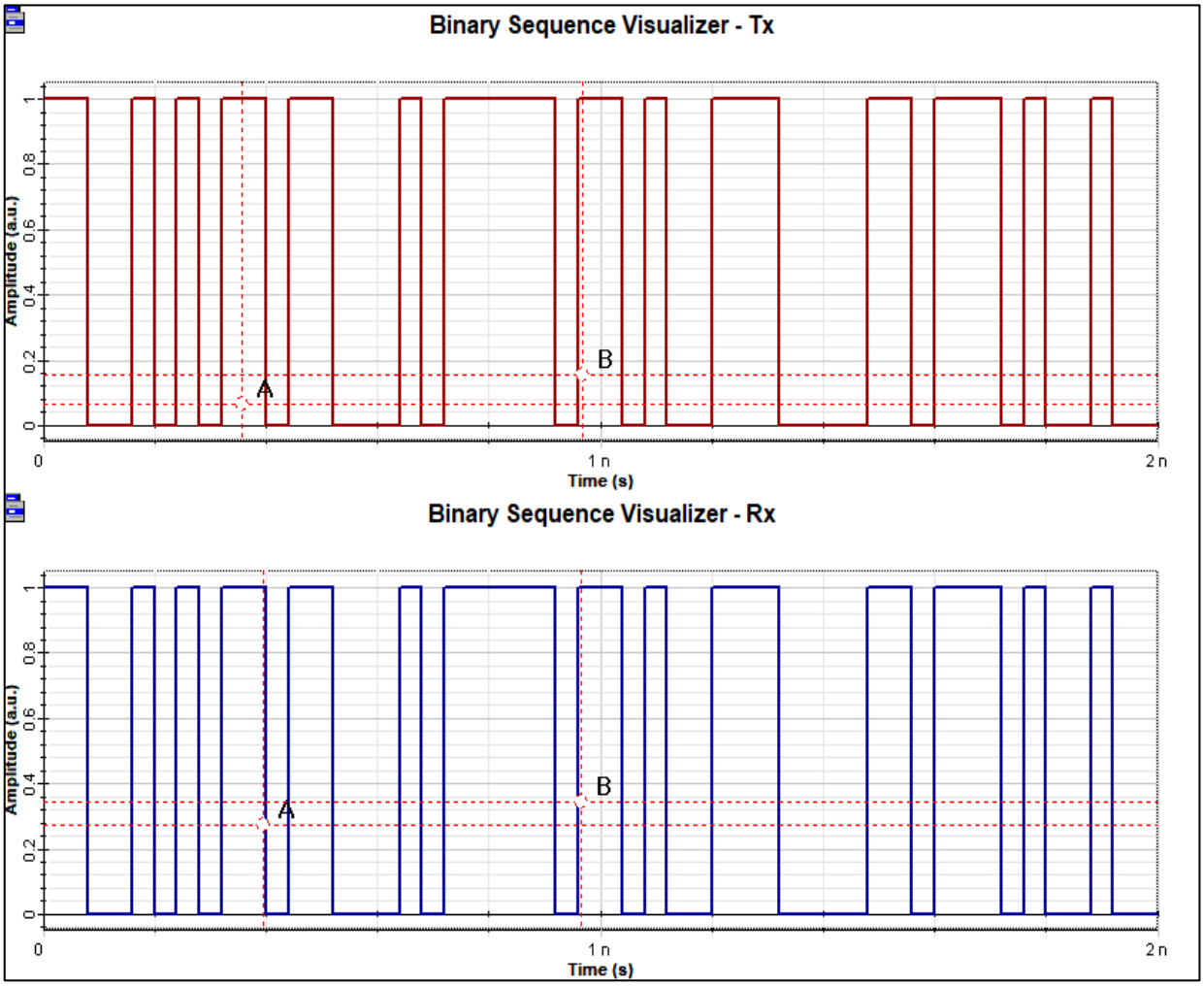


Figure (5.16): 16QAM-SP-CO-OFDM sequence visualizer testing the DNN-NLC model after 6000 Epochs (600,000 iterations) with Learning rate (0.0016) and training Mini-Batch size is 128.

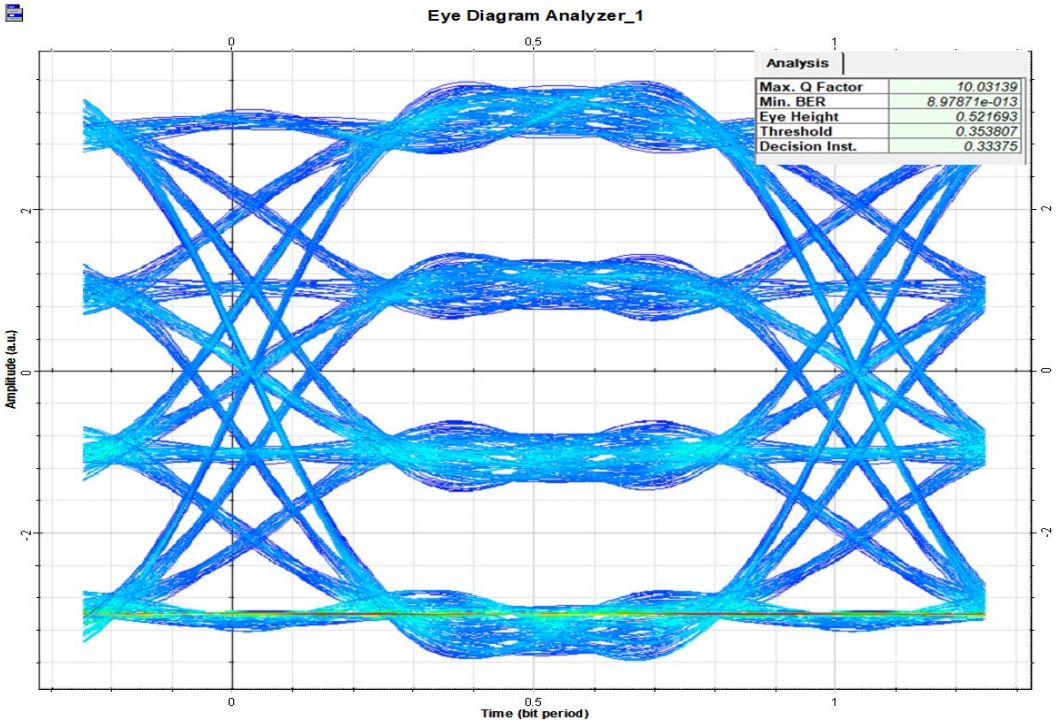


Figure (5.17): BER and Q-factor analyzer for the one polarization 16QAM coherent optical OFDM BtB 120 Gb/s @ (100 x 30span) Km measurements.

5.3.2 Summarization and Discussion

DNNs based NLC model has been proposed and designed for 120-Gbps, 16-QAM, CO-OFDM single-polarization optical systems. The obtained results showed that the neural network receiver learns near-optimal decision regions for 16-QAM constellation format and the proposed DNNs-NLC model outperforms in terms of low complexity, more robust, high Q-factor, higher transmission distance, and low signal delay on ANN, LE, and IVSTF-NLE by 0, 3, and 1 dB, respectively, according to the work in [27]

DNNs proved to be a robust nonlinearity compensation technique for coherent OOFDM and referring to the impressive results that have been achieved it should trigger to implementation of the nonlinear impairment compensation based on Deep Neural Networks in the next generation core networks.

5.4 SP - WDM - CO_OFDM

The optical system model under study is 8 channel WDM @ 60 Gbps bitrate per channel and 40GHz frequency-spacing are sent over standard single mode fiber (SMF) towards coherent receiver side to demonstrate improvement after @3000 km standard.

N-spans of fiber links with 100km per-span will be considered and assume the noise figure is 4dB, also, used Erbium-Doped Fiber Amplifiers (EDFA) to compensated spans losses with all ASE noise added completely before the receiver side. Optical filter with 3 order Gaussian frequency transfer function will be used at receiver side only with 100 dB Depth (maximum attenuation value for the filter). In addition, at the receiver, the incoming OOFDM signal is coherently detected and after removing Cyclic Prefix (CP) and taking Fast Fourier Transform (FFT), the symbols are fed to a linear equalization (LE) for compensating Chromatic-Dispersion (CD) and channel distortions (linear dispersion) based on parameters of OOFDM training & pilot symbols to estimate the channel ICI. As a result, the residual distortions after LE will limit achievable rates of information. Figure (5.18) illustrates the 8 channels optical spectrum analyzer for the received optical WDM (Demuxer 1x8). Figures (5.19) (5.20) and (5.21) shows an electrical spectrum analyzer with an oscilloscope visualizer for the transmitted/received 4-16QAM constellation signal it can be seen that constellation is distorted more seriously with an increase in launch power as a result of Kerr fiber nonlinearity. The effects of the XPM, SPM, and FWM are deterministic (ISI) in multicarrier schemes like CO-OFDM [18].

So, for the purpose of compensating the residual non-linear distortion, the DNN-based NLC will be introduced.

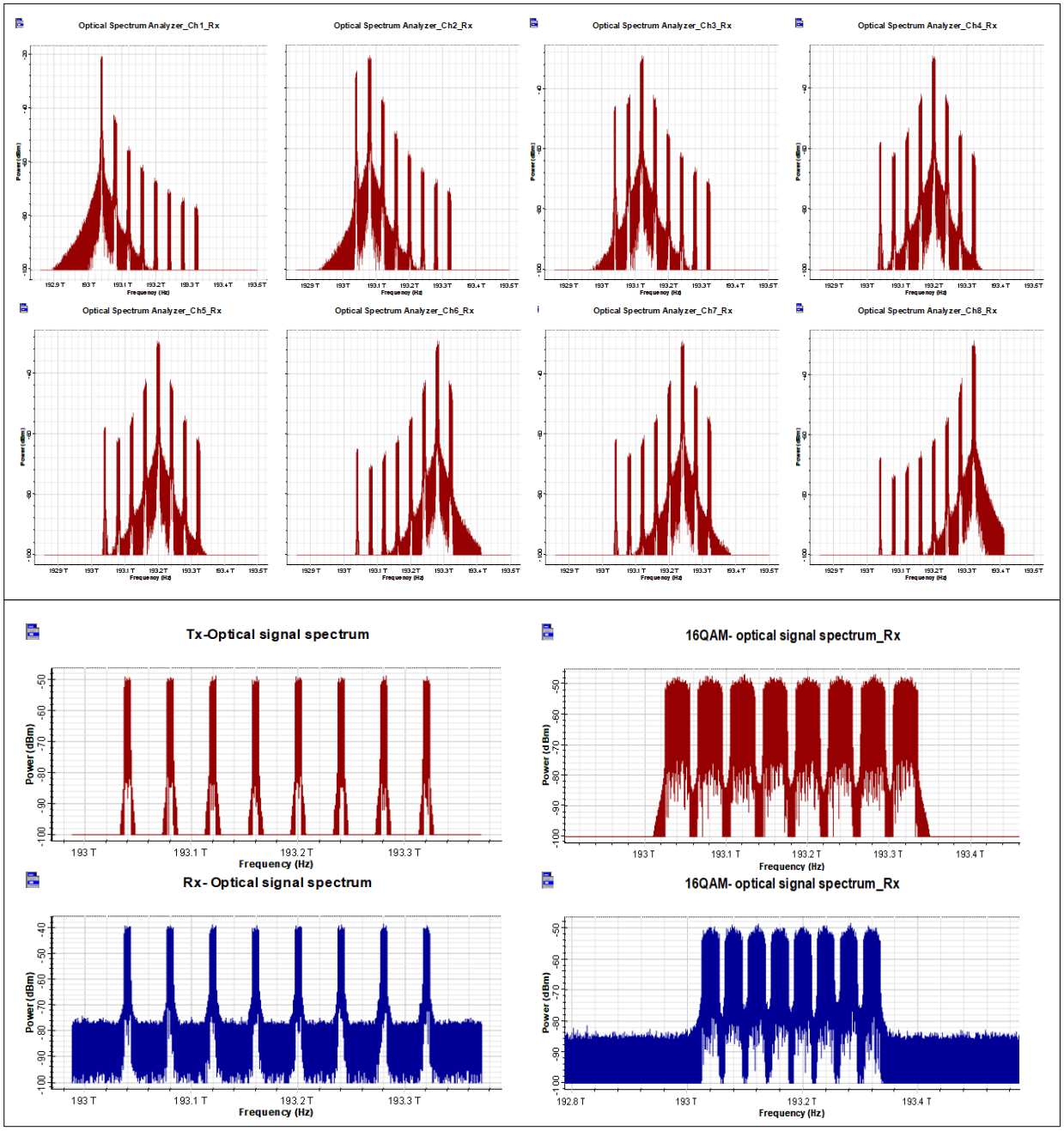


Figure (5.18): WDM (Demuxer 1x8) optical spectrum analyzer for the transmitted/received 4-16QAM constellation optical signal.

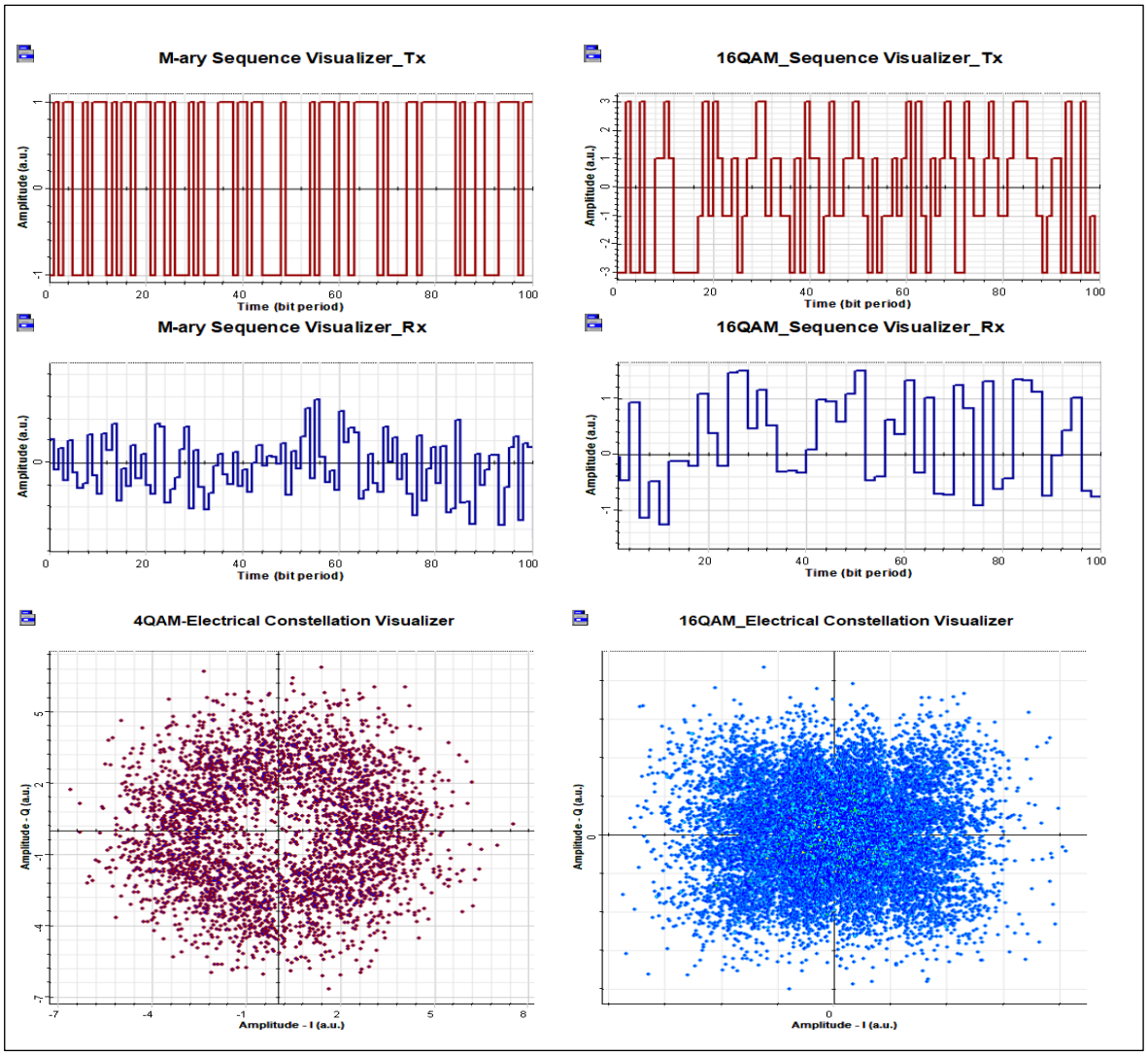
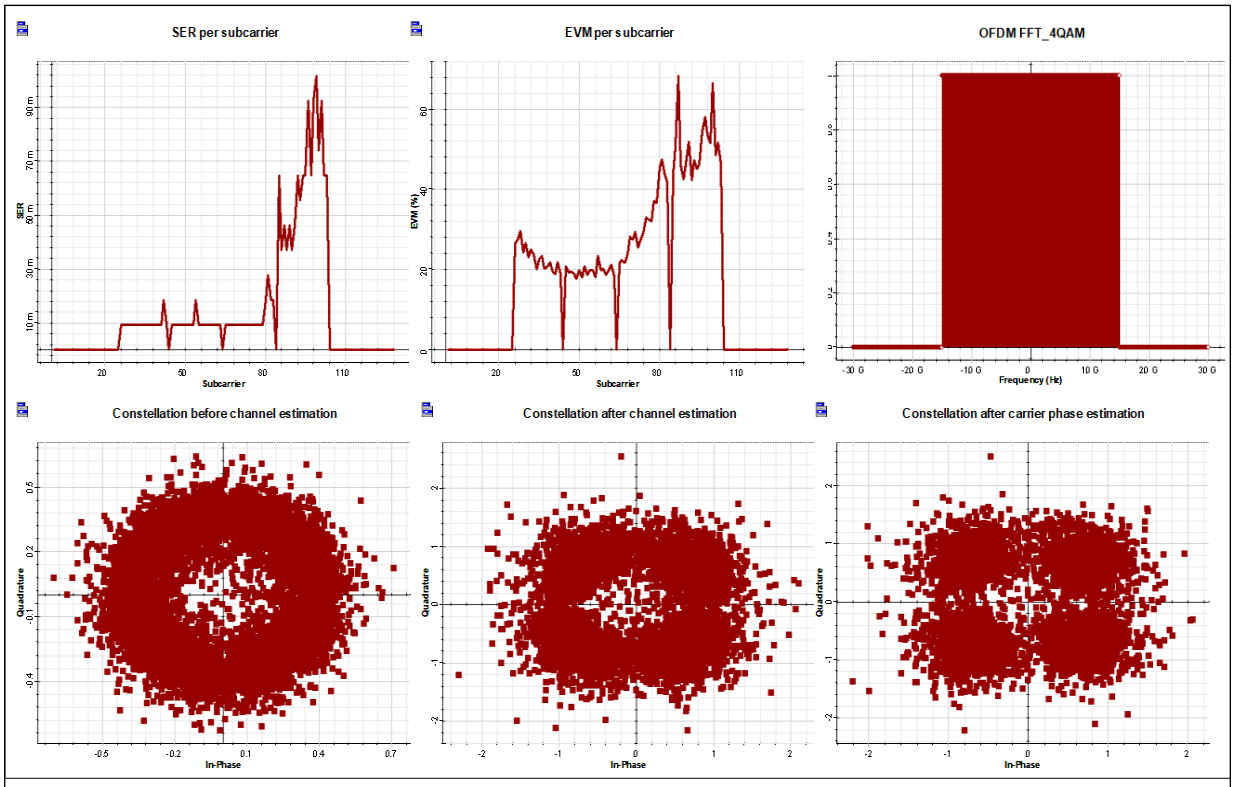
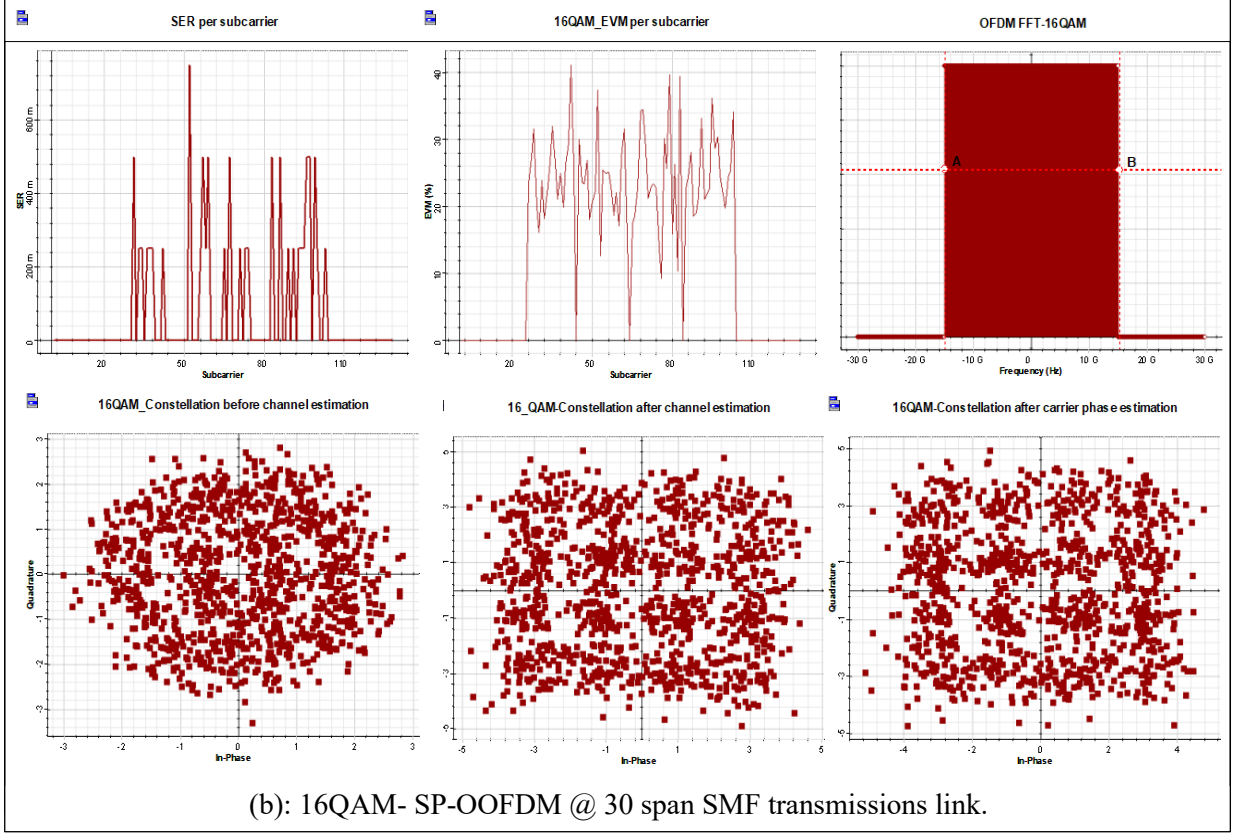


Figure (5.19): Electrical spectrum analyzer with an oscilloscope visualizer for the transmitted/received 4-16QAM constellation signal.



(a): 4QAM- SP-OOFDM @ 30 span SMF transmissions link.



(b): 16QAM- SP-OOFDM @ 30 span SMF transmissions link.

Figure (5.20): Spectrum analyzer for the received signal. (a): 4QAM constellation. (b): 16QAM constellation

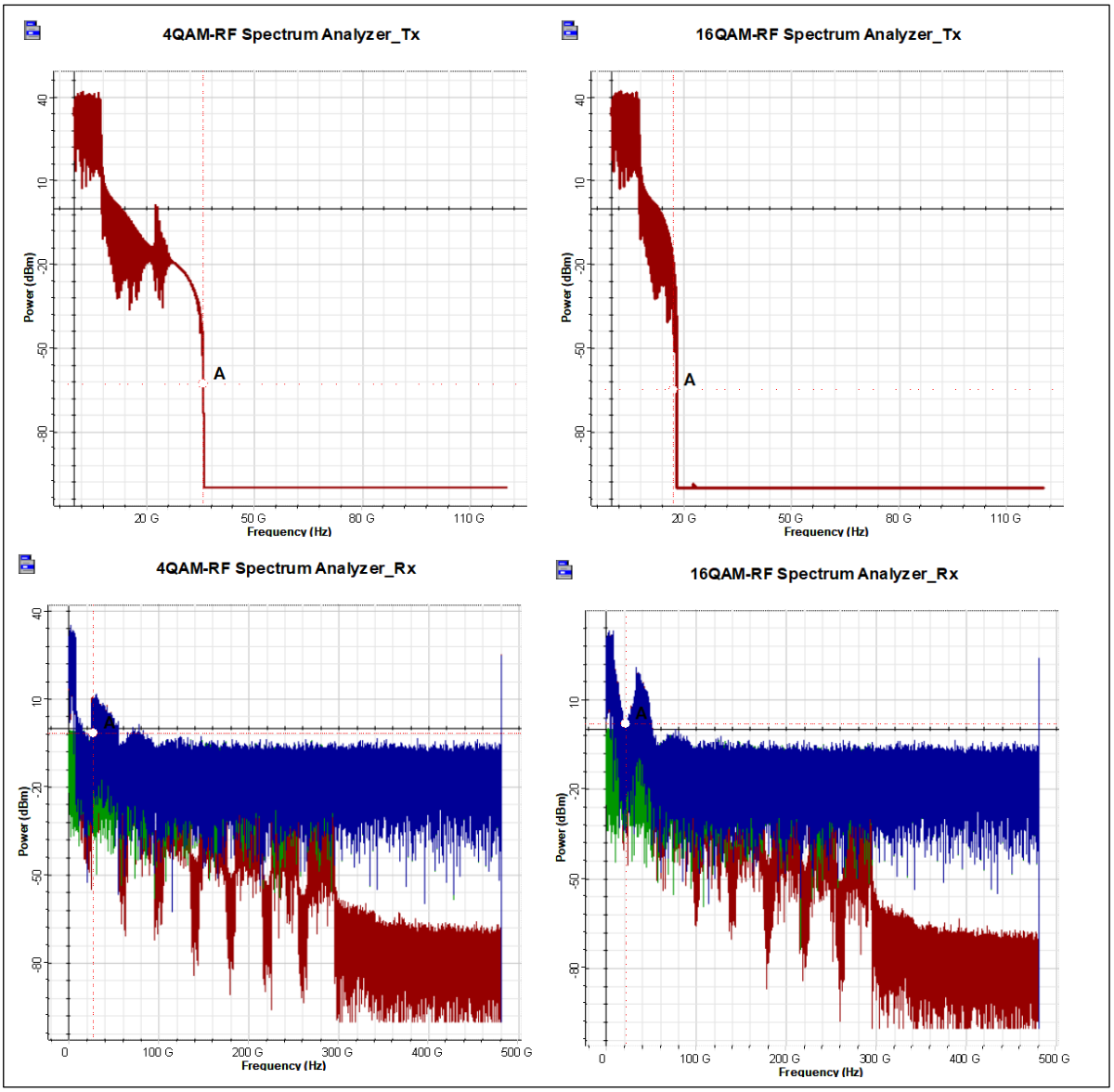


Figure (5.21): RF-Spectrum analyzer for the Transmitted and Received signal for 4-16QAM constellation

5.4 1 DNN-NLC For SP - WDM - CO_OFDM

The DNNs-NLC model's ability was evaluated based on BER and correct constellation mapper decoder. Firstly, there are 4 bits per symbol (m=4) in SP-16QAM, leading to (2^4 = 16) classes to identify per symbol. The received data after LE with channel, phase, and carrier estimation are including inter-symbol interference (ISI) with propagation distance and it fid to the DNNs to trained by sigmoid activation function, SoftMax & Classification Output layer, an LSTM layer and a Leaky ReLU layer are used, which performs a threshold operation

with ADAM optimization algorithm as well as cross-entropy loss function. DNNs have 5 hidden layers with (128-256) hidden nodes. The size of the dataset used (128 x 262144) samples, where 85% of the dataset utilized for training and 15% for the testing and validation. The dataset will be separated to imaginary and real parts prior to being fed to the DNN.

Training: The same scenario for training process that was previously shown and explained in Sections (5.2 and 5.3) will follow in this optical transmission system

The choice of hyperparameters like the number of the layers, activation function, number of nodes in the hidden layers, learning rate, size of the mini-batch, and so on. is significant, and in this work the choice was made by experimentation, experience, testing, and prior knowledge of the problem with the aim of keeping networks rather small and as a result, training efforts manageable.

The performance for different hyperparameters is studied for 16QAM-SP-CO-OFDM as shonw in Tables (5.9, 5.10 and 5.11).

Table (5.9): DNN performance @ 10^5 Epoch for different mini batch size.

Mini-Batch Size	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
50	0.7108	0.2447	0.6814	0.3263
75	0.8107	0.2110	0.8004	0.2974
100	0.9146	0.0918	0.9074	0.1028
128	0.9784	0.0677	0.9400	0.0911
256	0.9818	0.0490	0.9809	0.0686

Table (5.10): DNN performance @ 256 mini-batch size for different Epoch.

Epoch	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
10^2	0.**6**119	0.**3**181	0.**6**024	0.4236
10^3	0.**71**01	0.**30**24	0.**70**04	0.3628
10^4	0.8903	0.1025	**0.8001**	**0.2342**
10^5	0.9818	0.0490	0.9809	0.0686
10^6	0.9916	0.0230	0.9907	0.0455

Table (5.11): DNN performance @ (10^5 Epoch, 256 batch size) for diff. learning rate.

Learning rate	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
0.001	0.9818	0.0490	0.9809	0.0686
0.0001	0.9788	0.0551	0.9709	0.0651
0.0015	0.9739	0.0409	0.9351	0.0576
0.0005	0.7917	0.4517	0.8056	0.4366

The performance of hidden units' number, over sampling number, the size of dataset, and the size of mini-batch size (MBS) will be studied at 10^5 Epoch, 256 mini-batch size and 0.001 learning rate as shonw in Figure (5.22)

Hidden units' number: Figure (5.22a) shows, at 128 hidden units the system is sufficient for neural network data modeling, but with 16, 32, 64 and 96 hidden units the performance deteriorates.

A very high significant performance enhancement was accomplished with over 128 hidden units but the complexity and training time is very long.

Sampling rates: Figure (5.22b) shown, from 4 to 16 sample/symbol are appropriate for capturing information that is provided by the signal and no evident enhancement of the performance has been accomplished with more than

16 sampling rates.

Data set size: Figure (5.22c) shows, in more than 2^18 training data samples, will result in little performance improvement.

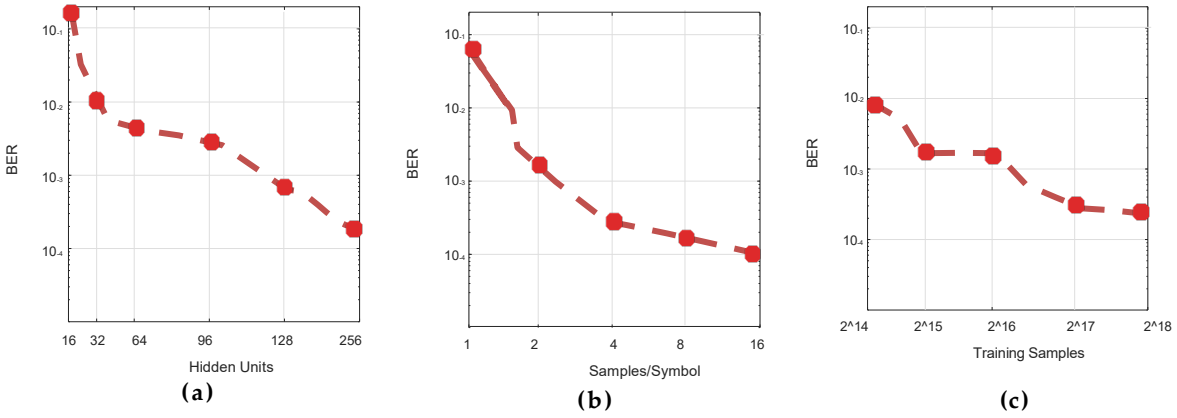


Figure (5.22): Training Process Performance for various DNN-NLC hyperparameter.

As a result, the small mini-batch size at training leads to rapid convergence but the final performance is somewhat worse, and when used a large training mini batch size, it resulted in slower convergence with longtime training process but higher final performance. A valid tradeoff between the performance, convergence, and computation time is obtained when training the DNN model at started with small batch size and increased it after the initial convergence value.

Finally, the hyperparameters for training the proposed DNN model are:

Layers: 7

Hidden nodes/layer: 256

Activation functions: ReLU

Learning rate: 0.001

Size of the training batch: Adaptive (128-256)

Optimization approach: Adam

Mini-Batch size= 256

Epochs=10^5

Finally result as shown in Figure (5.23): for Training: Accuracy: 0.9818, loss: 0.0490.

for Testing: Accuracy: 0.9809, loss: 0.0686.

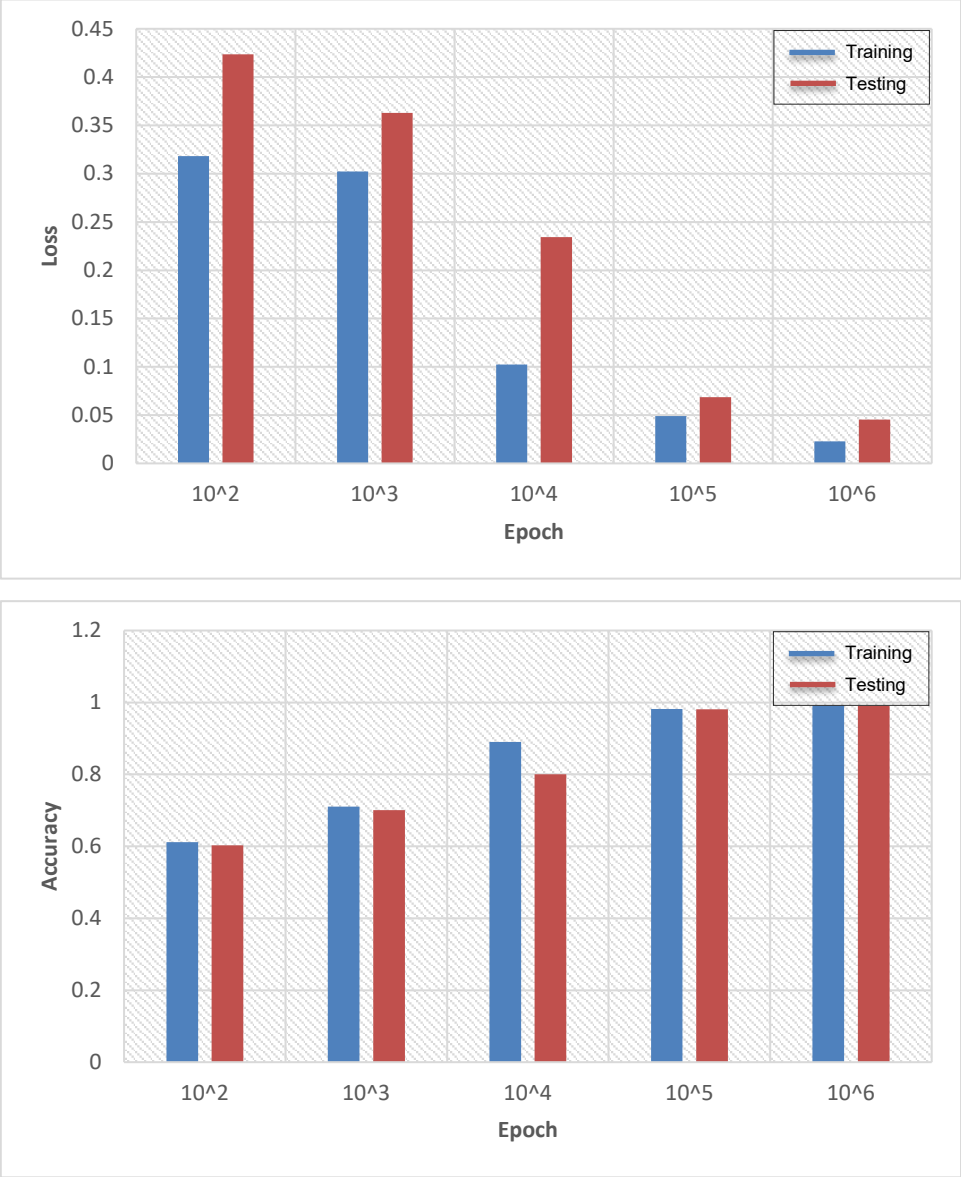
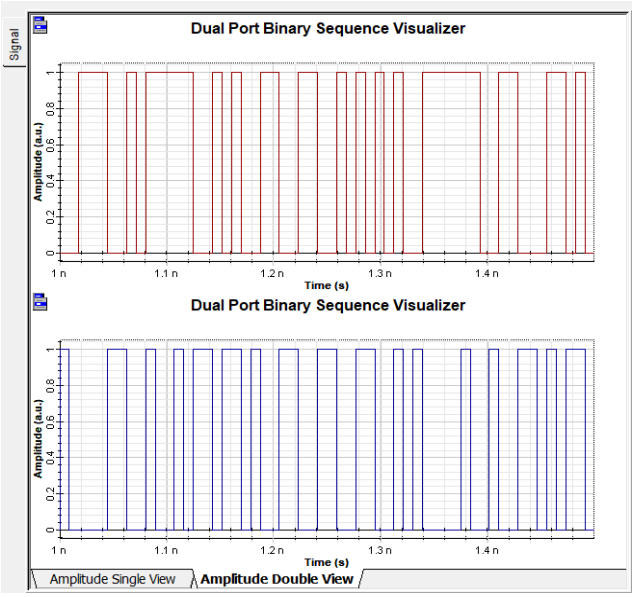


Figure (5.23): Accuracy & Loss Curve of DNN model.

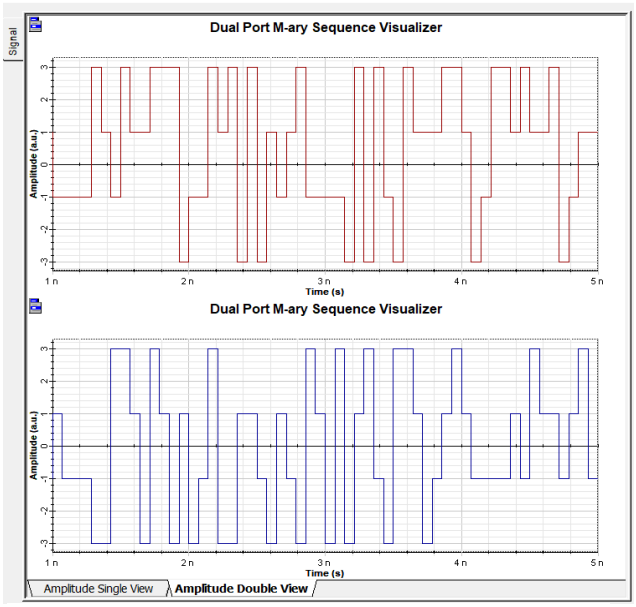
5.4.2 Joint Optimization of DNN-NLC & OptiSystem Parameters

The results shown in Figures (5.24 & 5.25) after training and testing the network proved that the DNNs based on classification output layer is capable of classifying the received symbol for the reference target 4&16-QAM and at iteration (200,000) the result of constellation mapper is very clear but the loss is high (~0.2) as shown in Figure (5.26a), although the network classified the

symbols correctly, but the bit locations were still incorrect So, when the DNNs losses began to decrease (~0) and the validation accuracy rises to nearly 100% @iteration (600,000). The final result of the Q-factor for the system at testing is equal to (~9 dB) @3000 Km link distance and 60 Gbps bitrate per channel.

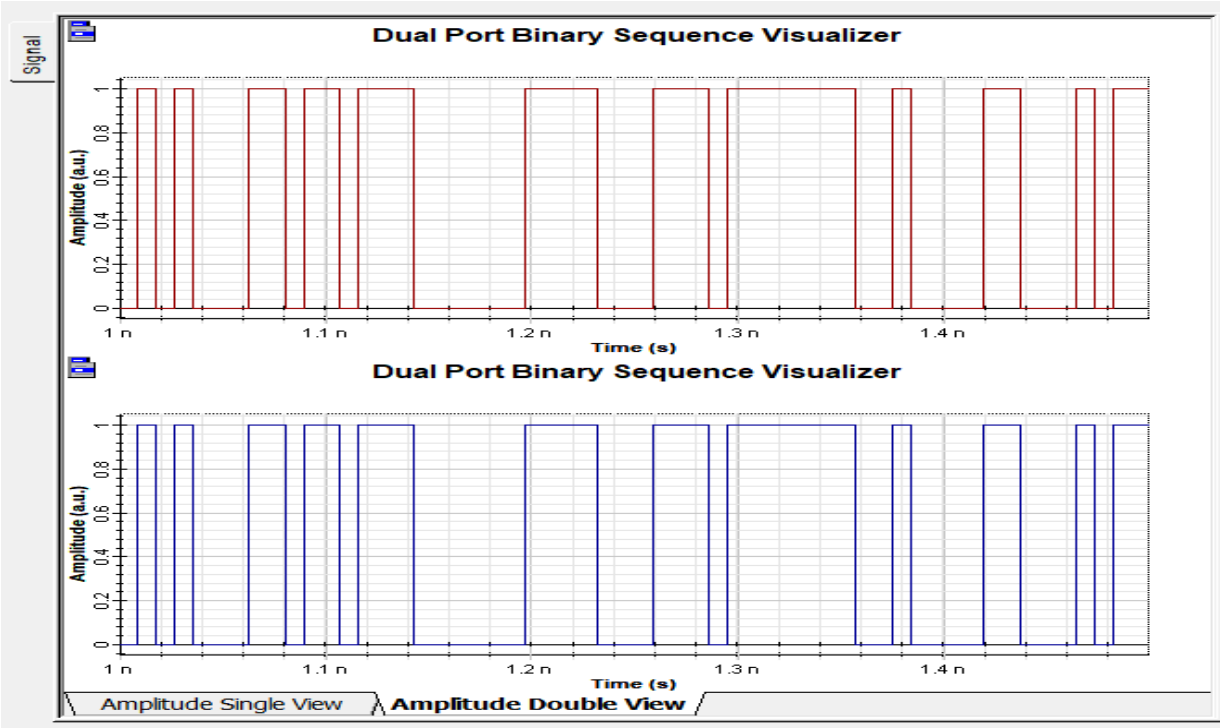


(a) 4QAM Binary Sequence Visualizer #Iteration @200,000



(b) 16QAM Binary Sequence Visualizer #Iteration @200,000

Figure (5.24): 16QAM Sequence Visualizer testing after : (a) 200000 & (b) 600000 iterations



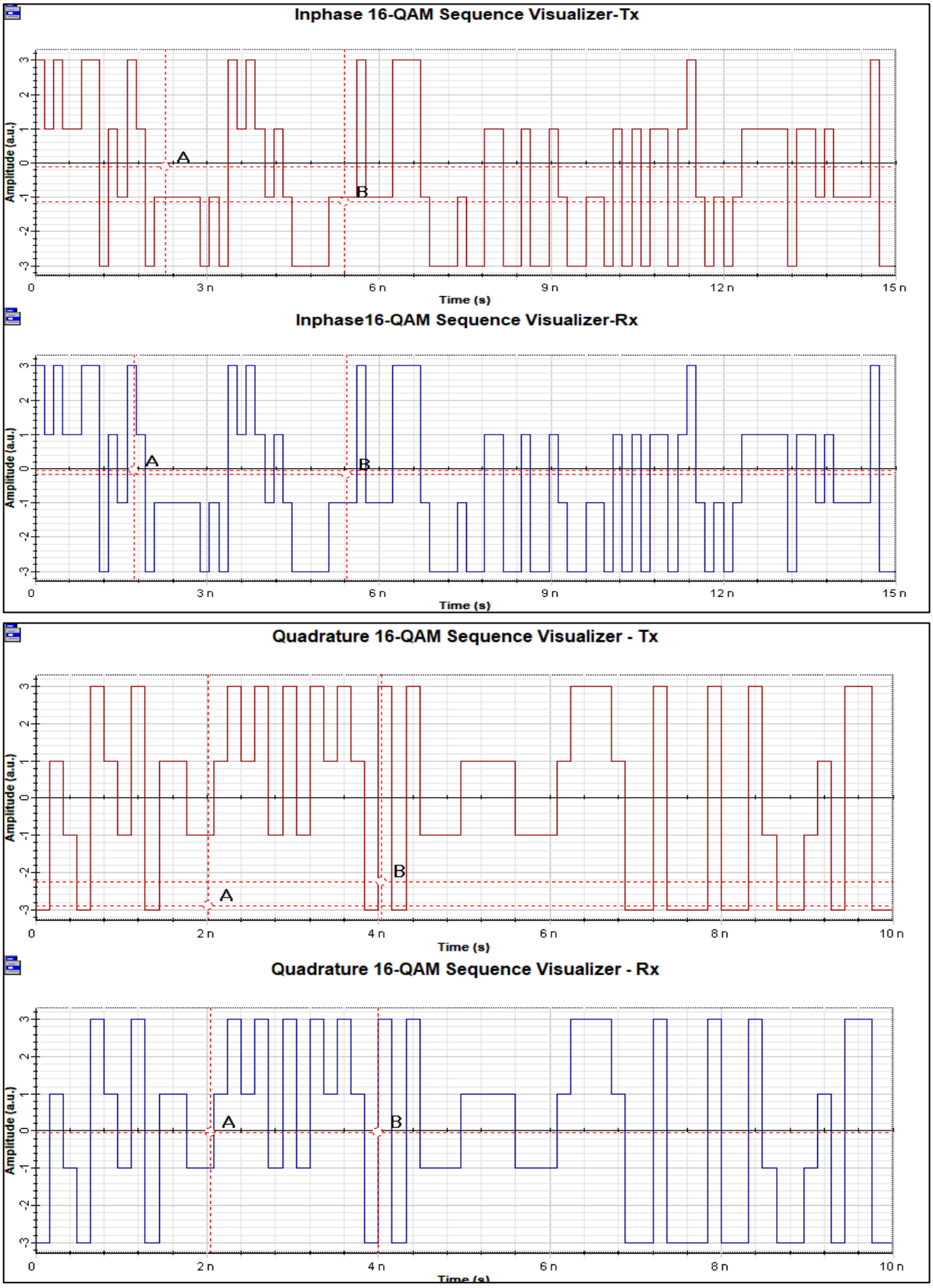


Figure (5.25): Final DNN-NLC result testing for 16QAM Sequence Visualizer testing.

5.4.3 Summarization and Discussion

The DNNs-NLC model ability was evaluated based on BER and correct constellation mapper decoder. The final result obtained of the Q-factor for the system at testing is equal to (~9 dB) @3000 Km link distance and 60 Gbps bitrate per channel for 8 WDM with 40GHz frequency-spacing.

The obtained results showed that: The proposed DNNs is successfully demonstrated to modify the QAM constellation and a Q-factor improvement less than 1 dB with respect to the artificial neural network ANN and more than 2 dB with respect to the VSTF-NLE according to the work in [77]. The work in [77] achieved a Q-factor value equal to ~9 dB @ 3200 km-20 Gbps (4x1 WDM). The proposed model outperforms in terms of low complexity, more robust, high Q-factor, higher transmission distance.

DNNs-NLC structures have been used for learning the formats of modulation that have been optimized for the lowest bit error rate over the amplified links. In addition to symbol constellation geometry. we have also optimized bit mapping. The learned formats of modulation are assessed experimentally and gains have improved the Q-factor with respect to the ANN and VSTF, it demonstrated for 4-16QAM-SP-CD-OOFDM. We experimentally demonstrated, for the first time, a DNN-NLC for 8 channel WDM at 3000 km with 40 GHz spacing in the "worst-case" WDM channel, DNN improved Q-factor and it can lead to effectively compensating a considerable fraction of non-linear distortions, which include the intra as well as inter-channel distortions.

5.5 DP - WDM – 16QAM

The optical system model under study is 8 channel WDM- DP-16QAM @ 120 Gbps bitrate per channel with (0.3 nm - 37.5 GHz) frequency-spacing are sent over standard single mode fiber (SMF) towards coherent receiver side to demonstrate improvement after @3000 km standard.

N-spans of fiber links with 100km per-span will be considered and assume the noise figure is 4dB, also, used Erbium-Doped Fiber Amplifiers (EDFA) to compensated spans losses with all ASE noise added completely before the receiver side. Optical filter with 3 order Gaussian frequency transfer function will be used at receiver side only with 100 dB Depth (maximum attenuation value for the filter).

Figure (5.26) illustrates the 8 channels optical spectrum analyzer for the received optical WDM (Demuxer 1x8). Figures (5.27) and (5.28) shows an electrical spectrum analyzer with an oscilloscope visualizer for the transmitted/received 16QAM signal for X and Y polarization (XP and YP), it can be seen that constellation is distorted more seriously with an increase in launch power as a result of Kerr fiber nonlinearity. The effects of the Cross-Phase Modulation (XPM), SPM, and FWM are deterministic (ISI) in multicarrier schemes.

So, for the purpose of compensating the residual non-linear distortion, the DNN-based NLC will be introduced.

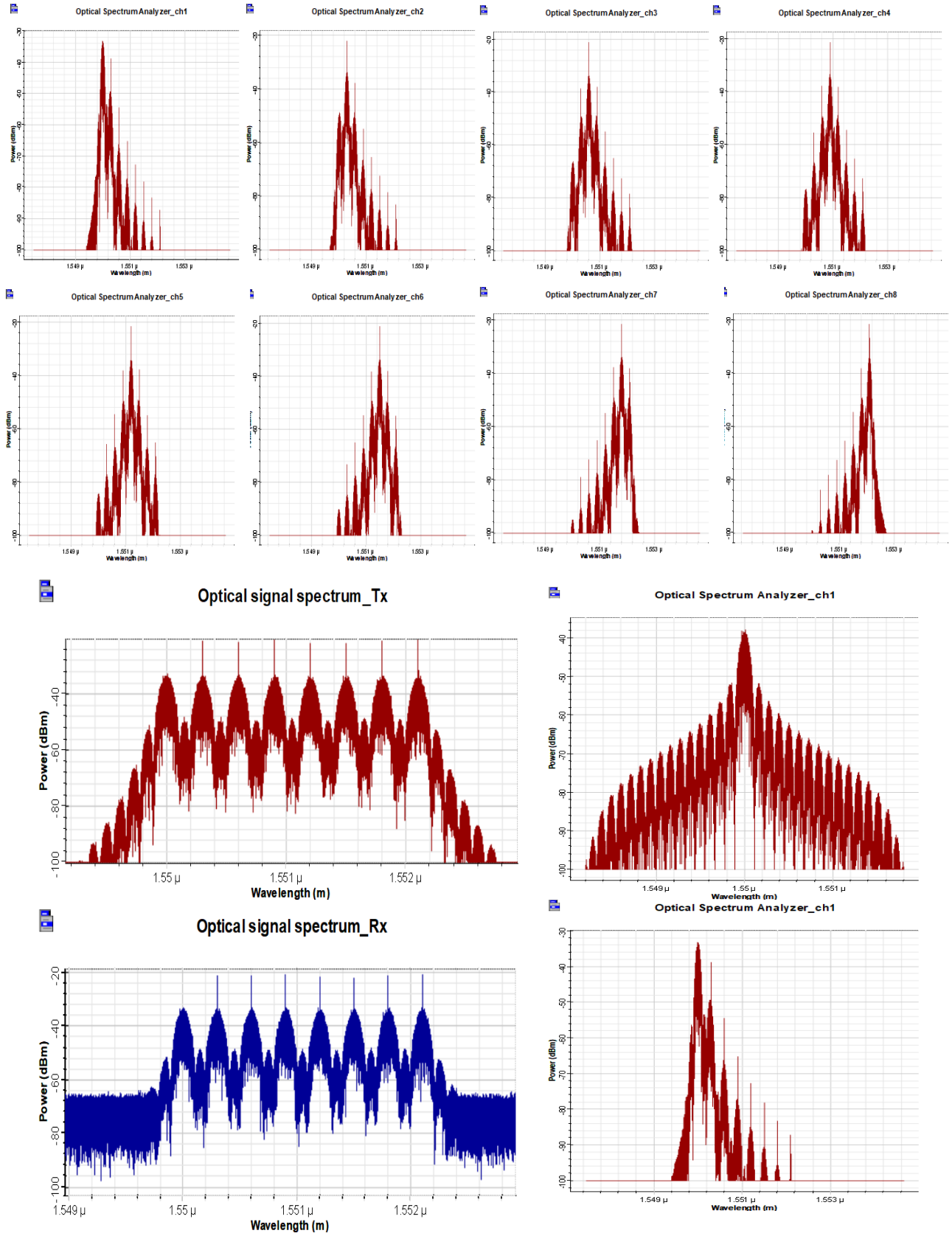


Figure (5.26): WDM (mixer 8x1) optical spectrum analyzer.

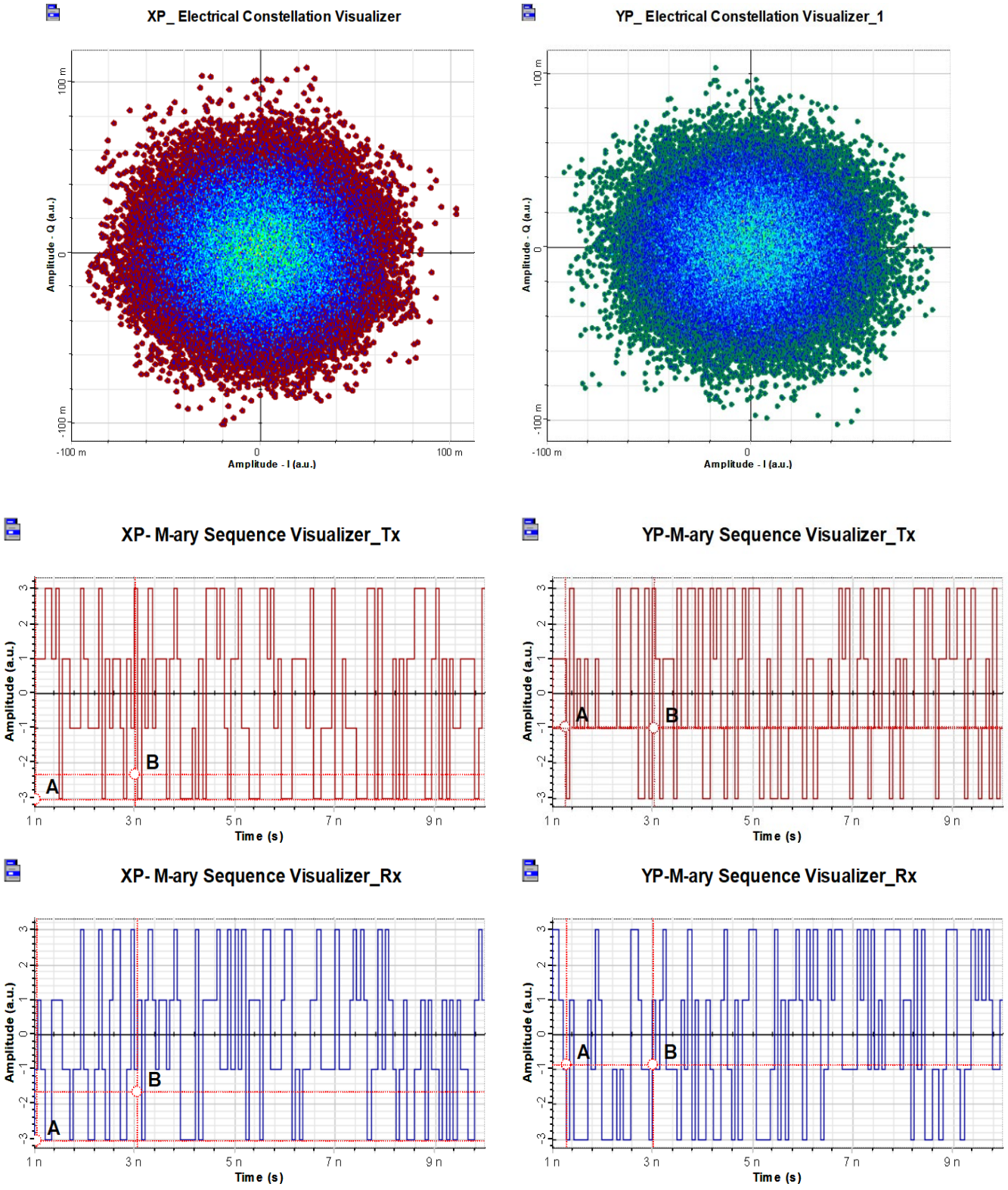


Figure (5.27): Spectrum analyzer for the received signal

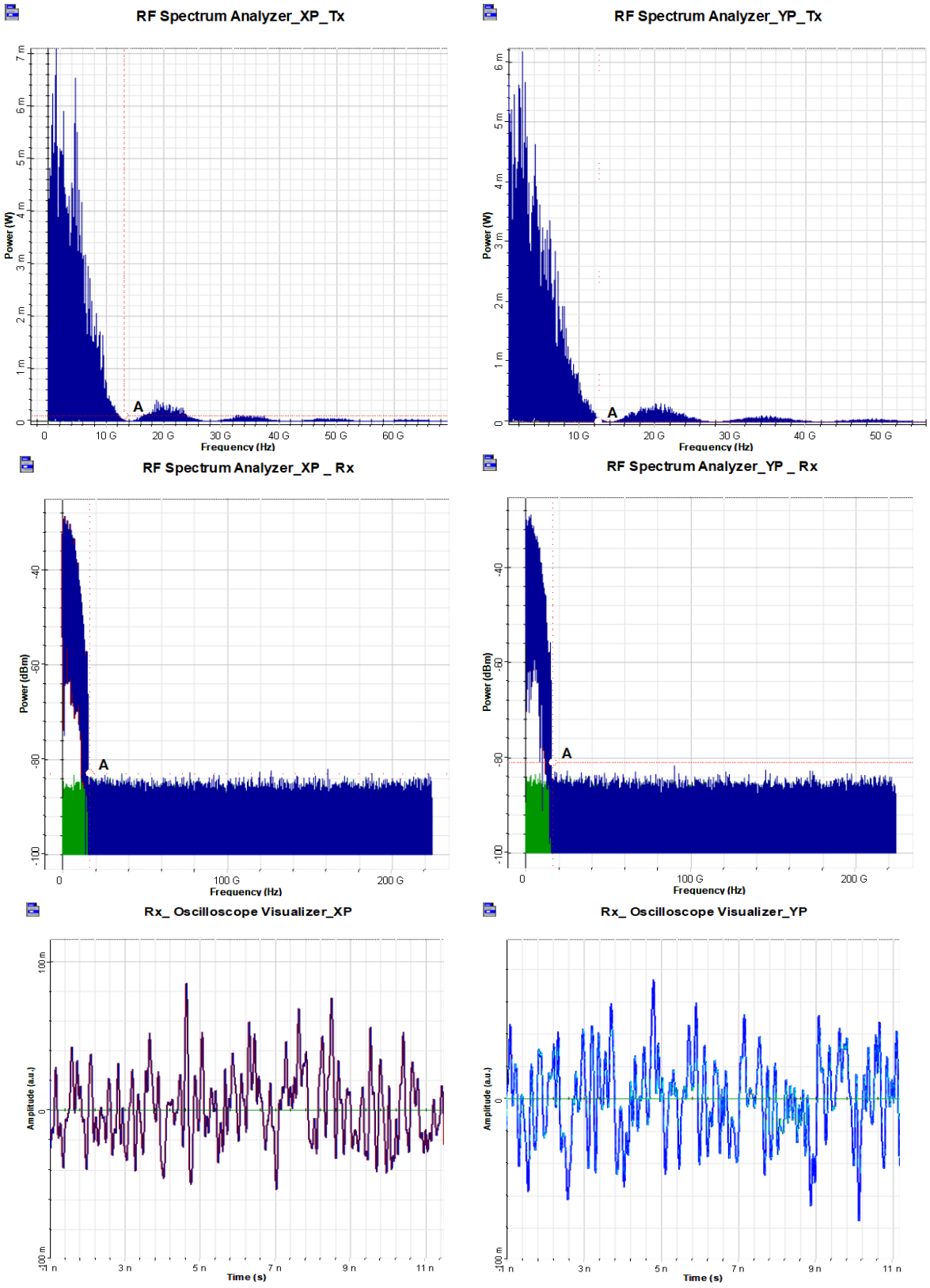


Figure (5.28): RF spectrum analyzer with an oscilloscope visualizer for the transmitted/received signal.

5.5.1 DNN-NLC For DP - WDM – 16QAM

The DNN-NLC weights and biases are training in many numerous epochs (iterations) and they are updated by the shifting of present values toward previous direction of gradient. So it will apply mini-batch back-propagation method, computing the cost function gradients with respect to network parameters, as well as stochastic gradient descent optimizer ADAM will be implemented in order to estimate and compute the perfect value for weights and biases which give the lower lost cost.

For the purpose of preventing the over-fitting, the size of the mini-batch is initially set to 1% of numbers of training symbols, respectively to ((2^16 Sequence length) / (4 Bit/Symbol * 2 polarization)) = 8192 symbols. The ADAM optimizer learning rate has been set to 10^2 initially and then increased to 10^4 which is the common value.

The design options for the DNN-NLCs are numerous and inter-related and there aren't any rules for number of the layers, activation functions, number of the nodes in the hidden layer, size of the mini-batch, rate of learning, and so on. to model a particular problem, predominantly, the choice was based on experimentation, experience, testing, and prior knowledge of the problem with the objective of keeping networks rather small, therefore, training efforts manageable.

However, the higher the number of the neurons and number of layers to which the deep network is connected, the more accurate the modeling capabilities but the complexity and training process will increase a lot in a way that prevents network management. Unfortunately, more complexity makes DNNs susceptible to overfitting.

In this study, the proposed DNNs-NLC model's ability was evaluated based on BER and correct constellation mapper decoder. Firstly, there are 4 bits per symbol per polarization (m=4), leading to (2^4 = 16) classes to identify per symbol. The received data is fid to the DNNs to trained by sigmoid activation

functions, SoftMax & Classification Output layer, an LSTM layer and a Leaky ReLU layer are used, which performs a threshold operation with cross-entropy loss function. The DNNs have (128-256) hidden nodes. The size of the dataset used (128 x 262144) samples, where 85% of the dataset utilized for training and 15% for the testing and validation. The dataset will be separated to imaginary and real parts prior to being fed to the DNN.

Training: The same scenario for training process that was previously shown will follow in this optical transmission system

The performance for different hyperparameters is studied for 16QAM-DP-CD-16QAM as shonw in Tables (5.12, 5.13, 5.14, 5.15, 5.16 and 5.17).

Table (5.12): DNN performance XP @ 10^6 Epoch for different mini batch size.

Mini-Batch Size	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
50	0.7990	0.2785	0.7685	0.2911
82	0.8801	0.2161	0.8557	0.2932
100	0.9039	0.2001	0.8887	0.4001
128	0.9242	0.1568	0.9053	0.2500
200	0.9522	0.1003	0.9196	0.2391
256	0.9702	0.0631	0. 9422	0.1088

Table (5.13): DNN performance XP @ 256 mini-batch size for different Epoch.

Epoch	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
10^2	0.6119	0.3181	0.6024	0.4236
10^3	0.8747	0.3479	0.8165	0.6096

10^4	0.9228	0.1751	0.9108	0.1847
10^5	0.9492	0.1214	0.9382	0.1310
5x10^5	0.9548	0.1055	0.9377	0.1334
10^6	0.9702	0.0631	0. 9422	0.1088

Table (5.14): DNN performance XP @ (10^6 Epoch, 256 batch size) for diff. learning rate.

Learning rate	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
0.01	0.92041	0.15919	0.91589	0.1698
0.001	0.9702	0.0631	0. 9422	0.1088
0.0015	0.9801	0.0471	0. 9556	0.0737
0.0001	0.9748	0.1040	0.9481	0.1102

Table (5.15): DNN performance YP @ 10^6 Epoch for different mini batch size.

Mini-Batch Size	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
50	0.8828	0.2304	0.8676	0.2581
82	0.9139	0.1757	0.8925	0.2619
100	0.9200	0.1783	0.9014	0.2534
128	0.9187	0.1527	0.9081	0.1859
200	0.9388	0.1221	0.9301	0.1357
256	0.9646	0.0404	0.9411	0.0827

Table (5.16): DNN performance YP @ 256 mini-batch size for different Epoch.

Epoch	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
10^2	0.6119	0.3181	0.6024	0.4236
10^3	0.8197	0.3479	0.8165	0.6096
10^4	0.9128	0.1751	0.9108	0.1847
10^5	0.9492	0.1214	0.9382	0.1391
5x10^5	0.9548	0.1055	0.9331	0.1314
10^6	0.9646	0.0404	0.9411	0.0827

Table (5.17): DNN performance YP @ (10^6 Epoch, 256 batch size) for diff. learning rate.

Learning rate	Training		Testing	
	Accuracy	Cost_Loss	Accuracy	Cost_Loss
0.01	0.9544	0.1079	0.9241	0.1759
0.001	0.9646	0.0404	0.9411	0.0827
0.0015	0.9525	0.1141	0.9521	0.1135
0.0001	0.9769	0.0142	0.9610	0.0801

Finally, the hyperparameters for training the proposed DNNs-NLC model are:

Layers: 7

Hidden units' number: 256

Learning rate: 0.0015 for X_Polarization

Learning rate: 0.0001 for Y_Polarization

Activation function: ReLU

Size of the training batch: Adaptive (128-256)

Optimization approach: Adam

Mini-Batch size= 256

Epochs=10^6

Finally result:

	Training		Testing	
	Accuracy	Loss	Accuracy	Loss
X_Polarization:	0.9801	0.0471	0. 9556	0.0737
Y_Polarization:	0.9769	0.0142	0.9610	0.0801

The performance of the number of the hidden units, the number of samples, the size of data set, the value of the batch size and its impact on the performance of proposed neural network model will be depended on the previous experiments and tests shown in this study.

5.5.2 Joint Optimization of DNN-NLC & OptiSystem Optical System Parameters

The results shown in Figure (5.29) after training and testing the network proved that the DNNs based on classification output layer

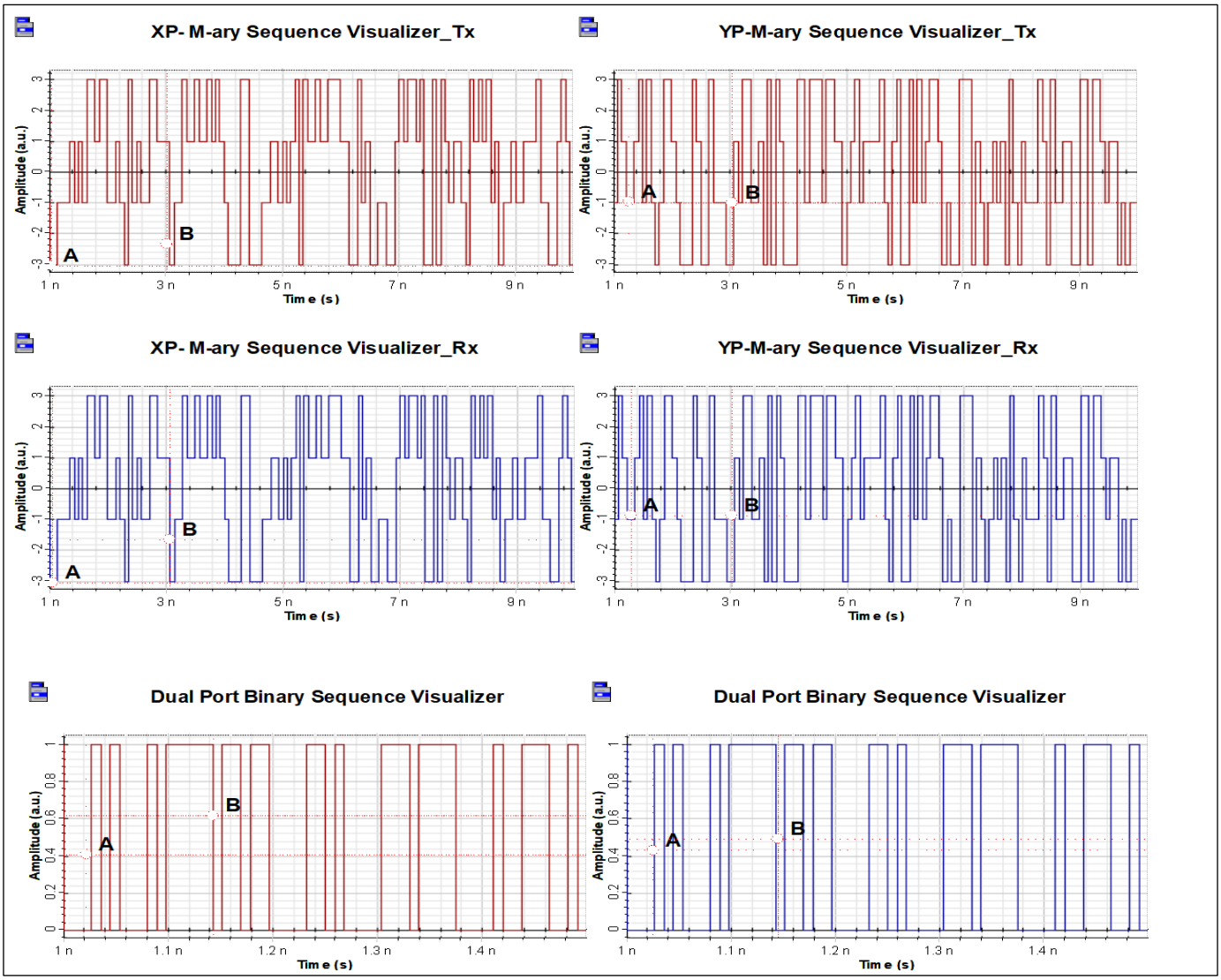
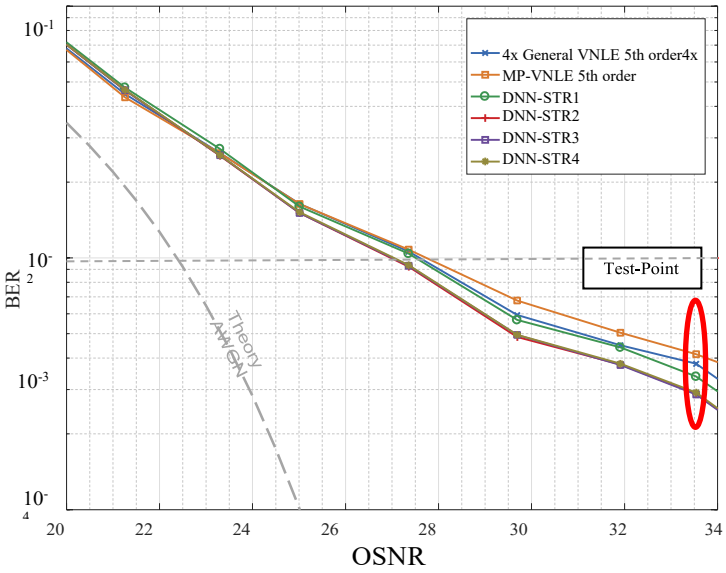


Figure (5.29): 16QAM-DP M-ary & binary sequence visualizer testing the proposed DNN-NLC model after 3000 Km SSMF.

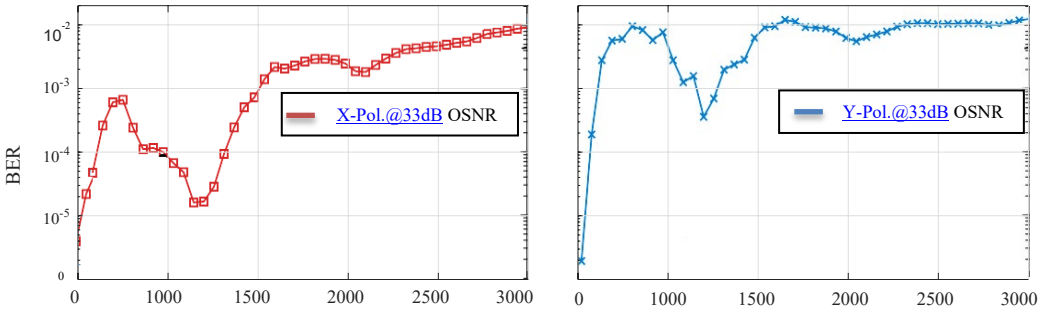
5.5.3 Summarization and Discussion

System-agnostic optical nonlinearity impairment compensation algorithm was proposed based on a DNN. The DNNs-NLC proposed model ability was evaluated based on BER and correct constellation mapper decoder. The final result was evaluated on coherent dual-polarization 16QAM - 960Gb/s (120

Gbaud) for all system - 120 Gbps bitrate (15 Gbaud) per channel, back-to-back measurements @3000 Km link distance for 8 WDM with (37.5GHz - 3nm) frequency-spacing as shown in Figure (5.30).



(a) Result of different DNN structure study in [21]



(b) Proposed DNN-NLC model performance

Figure (5.30): Summarization and discussion of simulation result for proposed DNN-NLC model:(a) Result of different DNN structure study in [21]. (b) Proposed DNN-NLC model.

Chapter6
Conclusions and Future Works

6.1 The Conclusion

There is a number of the conclusions that can be obtained from the presented thesis, which can be summarized as follows:

1- Based on the comprehensive study that has been prepared for the most important and advanced developments in the technologies of optical fiber communication systems by covering the techniques of modulation and detection, polarization, super-channel transmission WDM, SDM, and finally the advanced NLC techniques. where it focused directly and primarily on advanced NLC technologies and reviewed them in two directions:

The first side focused on classical computational methods based on DSP algorithms, while the second side reviewed the use of artificial intelligence techniques and their role in optical NLC and provided a survey of the existing research in the application of these techniques and describe its approaches in terms of learning, statistical, and decision-making methods, along with an overview of the challenges, gains and determining the appropriateness of these techniques when dealing as an NLC to improve the system capacity performance. According to results of the present comprehensive study, the following was the conclusion:

ML techniques have demonstrated the ability to find optimal solutions in highly complex models and environments without the need to develop a physical model, which makes them necessary to meet the challenges of optical transmission complexity and has achieved good benchmark results compared to the traditional conventional methods.

AI allows optical communication to have flexible statistical analysis of complicated systems without dependence on specific models besides, monitoring, traffic control and signal design.

The impressive and good results achieved when using deep learning algorithms are the motivation behind their use in this study, although there are many widely used ML algorithms for optical communication systems.

2- The study in this thesis has led to the proposal, design, and implementation of a Nonlinear Impairments Compensation (NLC) model based on Deep Neural Network (DNN). Theoretical analysis, simulation, test results, and necessary improvements for the proposed model were presented by studying the hyperparameter optimization and the following was the conclusion:

The hyperparameters of neural network has been independently optimized to demonstrate their impact on performance and complexity because the optimization of the hyperparameters is a multidimensional problem, and the methods for solving it depend on a grid search in hyperparameter space in order to obtain best performance but, grid search for every hyper-parameter space and training neural network is very infeasible. Four hyperparameters of the neural network which are important for its performance have been studied: The number of the hidden nodes, number of training samples, number of sample/bit, and the batch size and the following was the conclusion:

NN with infinite hidden nodes have the ability of modeling any functions, but when increase the number of the hidden nodes without knowing, performance efficiency will reach a point that no more information may be obtained from data. In addition, complexity and consumption of computer resources with the inability to manage network because of the long time spent on the learning process will be increased.

So, the least complex neural network was obtained by counting the number of the hidden nodes where performance converges.

The number of sample/bit and the number of training samples: They are in direct relationship with NN performance, and number of these parameters has to be minimal for the purpose of reaching lower complexity simultaneously, the rate of the sampling shouldn't be conflicting with sampling theorem, or else, the information will be lost.

The batch size: The small batch size at training led to rapid convergence but the final performance is somewhat worse, and when used a large size of the training batch, it resulted in slower convergence with longtime training process but higher ultimate performance. One of the valid trade-offs between the performance, convergence, and calculation time has been obtained when training proposed DNN model.

3- The NN receiver learns near-optimum decision areas for 4 and 16-QAM constellation format and the proposed DNNs-NLC model outperforms in terms of low complexity, more robust, high Q-factor, higher transmission distance, and low signal delay on ANN, LE, and IVSTF-NLE.

4- Work simulation results of using WDM-SP-CO-OFDM and WDM-DP 16QAM with the proposed NLC-DNNs model resulted in a significant improvement in all performance, especially in terms of data rate and link range.

6.2 Recommendations for Future Work

The DNNs-NLC proposed system model derivation, analytical studies of key system hyperparameters, mechanism limitation studies and experimental demonstration were conducted in-depth.

Through this investigation carried out in this thesis, many aspects would be extended that can be suggested as future works:

1- Other advanced modulation techniques such as SAC-OCDMA and SPECTS optic code division multiple access (O-CDMA) which are used in broadband access networks in Fiber to the home (FTTH) to provide various high-speed services to End users such as Internet, multimedia, telemedicine and distant learning could be considered in future work.

2- The use of artificial intelligence-based on DNN in an Optical Wireless Communication (OWC) System that classified into indoor such as directed LOS and non-directed LOS. Or outdoors Free-space optical (FSO) which includes terrestrial and deep-space optical links such as terrestrial FSO Link, inter-orbital Links (IOL), inter-satellite links (ISL) and satellite-to-airborne platforms to improve performance could be considered in future work.

Appendix A

WDM – Optical Power Meter Analyzer

Table (A.1): 16QAM-SP-CD-OOFDM-Power Meter Analyzer

Ch-1		Ch-2	
Tx : Frequency 1 (Hz)	193.04e+012	Tx : Frequency 2 (Hz)	193.08e+012
Tx : Wavelength 1 (nm)	1553.006931205968	Tx : Wavelength 2 (nm)	1552.685197845453
Tx : Signal Power 1 (dBm)	11.43825010595423	Tx : SignalPower 2 (dBm)	11.4391888643
Tx : Signal Power 1 (W)	0.01392595574143783	Tx : Signal Power 2 (W)	0.01392896626096874
Tx : Noise Power 1 (dBm)	-28.13718652906266	Tx : Noise Power 2 (dBm)	-28.19006268229133
Tx : Noise Power 1 (W)	1.5356119095835e-006	Tx : Noise Power 2 (W)	1.5170271830721e-006
Tx : SNR 1 (dB)	39.57350058412714	Tx : SNR 2 (dB)	39.57350058412714
Tx : OSNR 1 (dB)	43.3756130012432	Tx : OSNR 2 (dB)	43.3756130012432
Rx : Signal Power 1 (dBm)	-12.74438245219524	Rx : SignalPower 2 (dBm)	-12.50190909910172
Rx : Signal Power 1 (W)	53.15712924044e-006	Rx : Signal Power 2 (W)	56.209419647803e-006
Rx : Noise Power 1 (dBm)	-34.99023730338533	Rx : Noise Power 2 (dBm)	-34.99023730338533
Rx : Noise Power 1 (W)	0.3169394247012e-006	Rx : Noise Power 2 (W)	0.3169394247012e-006
Rx : SNR 1 (dB)	22.24585485119009	Rx : SNR 2 (dB)	22.4883282042836
Rx : OSNR 1 (dB)	26.04796726830615	Rx : OSNR 2 (dB)	26.29044062139967
Gain 1 (dB)	-24.18263255814947	Gain 2 (dB)	-23.94109796340172
Noise Figure 1 (dB)	-100	Noise Figure 2 (dB)	-100
Ch-3		**Ch-4**	
Tx : Frequency 3 (Hz)	193.12e+012	Tx : Frequency 4 (Hz)	193.16e+012
Tx : Wavelength 3 (nm)	1552.363597763049	Tx : Wavelength 4 (nm)	1552.042130875958
Tx : Signal Power 3 (dBm)	11.42370103473279	Tx : Signal Power 4 (dBm)	11.39364047496926
Tx : Signal Power 3 (W)	0.01387938118594513	Tx : Signal Power 4 (W)	0.01378364399191628
Tx : Noise Power 3 (dBm)	-28.13718652906266	Tx : Noise Power 4 (dBm)	-28.19006268229133

Tx : Noise Power 3 (W)	1.5356119095835e-006	Tx : Noise Power 4 (W)	1.5170271830721e-006
Tx : SNR 3 (dB)	39.56088756379544	Tx : SNR 4 (dB)	39.58370315726059
Tx : OSNR 3 (dB)	43.3629999809115	Tx : OSNR 4 (dB)	43.38581557437665
Rx : Signal Power 3 (dBm)	-12.59577970539349	Rx : Signal Power 4 (dBm)	-12.56158288323668
Rx : Signal Power 3 (W)	55.007515456047e-006	Rx : Signal Power 4 (W)	55.442360633376e-006
Rx : Noise Power 3 (dBm)	-34.98742608169354	Rx : Noise Power 4 (dBm)	-34.98742608169354
Rx : Noise Power 3 (W)	0.3171446512014e-006	Rx : Noise Power 4 (W)	0.3171446512014e-006
Rx : SNR 3 (dB)	22.39164637630005	Rx : SNR 4 (dB)	22.42584319845685
Rx : OSNR 3 (dB)	26.19375879341611	Rx : OSNR 4 (dB)	26.22795561557291
Gain 3 (dB)	-24.01948074012627	Gain 4 (dB)	-23.95522335820595
Noise Figure 3 (dB)	43.15912419582391	Noise Figure 4 (dB)	43.0937580819485
Ch-5		Ch-6	
Tx : Frequency 5 (Hz)	193.2e+012	Tx : Frequency 6 (Hz)	193.24e+012
Tx : Wavelength 5 (nm)	1551.720797101449	Tx : Wavelength 6 (nm)	1551.399596356862
Tx : Signal Power 5 (dBm)	11.43930494195278	Tx : Signal Power 6 (dBm)	11.43995071789226
Tx : Signal Power 5 (W)	0.01392933855750579	Tx : Signal Power 6 (W)	0.01393140994014505
Tx : Noise Power 5 (dBm)	-28.19771341292627	Tx : Noise Power 6 (dBm)	-28.13354986623487
Tx : Noise Power 5 (W)	1.5143583575123e-006	Tx : Noise Power 6 (W)	1.536897886542e-006
Tx : SNR 5 (dB)	39.63701835487905	Tx : SNR 6 (dB)	39.57350058412714
Tx : OSNR 5 (dB)	43.43913077199511	Tx : OSNR 6 (dB)	43.3756130012432
Rx : Signal Power 5 (dBm)	-12.4463964685604	Rx : Signal Power 6 (dBm)	-12.66417583314367
Rx : Signal Power 5 (W)	56.93251288927e-006	Rx : Signal Power 6 (W)	54.147999556753e-006
Rx : Noise Power 5 (dBm)	-34.98742608169354	Rx : Noise Power 6 (dBm)	-34.98461651398348
Rx : Noise Power 5 (W)	0.3171446516564e-006	Rx : Noise Power 6 (W)	0.3173498874335e-006
Rx : SNR 5 (dB)	22.54102961313313	Rx : SNR 6 (dB)	22.32044068083982
Rx : OSNR 5 (dB)	26.34314203024919	Rx : OSNR 6 (dB)	26.12255309795587
Gain 5 (dB)	-23.88570141051318	Gain 6 (dB)	-24.10412655103593
Noise Figure 5 (dB)	43.02214012177484	Noise Figure 6 (dB)	43.24548458529441

Ch-7		Ch-8	
Tx : Frequency 7 (Hz)	193.28e+012	Tx : Frequency 8 (Hz)	193.32e+012
Tx : Wavelength 7 (nm)	1551.078528559603	Tx : Wavelength 8 (nm)	1550.757593627147
Tx : Signal Power 7 (dBm)	11.42442626312458	Tx : Signal Power 8 (dBm)	11.39310896258647
Tx : Signal Power 7 (W)	0.01388169909745456	Tx : Signal Power 8 (W)	0.013781957043761
Tx : Noise Power 7 (dBm)	-28.13718652906266	Tx : Noise Power 8 (dBm)	-28.190062229133
Tx : Noise Power 7 (W)	1.5356119095835e-006	Tx : Noise Power 8 (W)	1.5170271821e-006
Tx : SNR 7 (dB)	39.63701835487905	Tx : SNR 8 (dB)	39.57350058412714
Tx : OSNR 7 (dB)	43.43913077199511	Tx : OSNR 8 (dB)	43.3756130012432
Rx : Signal Power 7 (dBm)	-12.60463772638133	Rx : Signal Power 8 (dBm)	-12.695954897938
Rx : Signal Power 7 (W)	54.89543493145e-006	Rx : Signal Power 8 (W)	53.7622039093e-006
Rx : Noise Power 7 (dBm)	-34.98461651398348	Rx : Noise Power 8 (dBm)	-34.984616398348
Rx : Noise Power 7 (W)	0.317349830005e-006	Rx : Noise Power 8 (W)	0.3173498005e-006
Rx : SNR 7 (dB)	22.37997878760216	Rx : SNR 8 (dB)	22.28938696500411
Rx : OSNR 7 (dB)	26.18209120471821	Rx : OSNR 8 (dB)	26.09149938212016
Gain 7 (dB)	-24.02906398950591	Gain 8 (dB)	-24.088338156585
Noise Figure 7 (dB)	-100	Noise Figure 8 (dB)	-100

Table (A.2): 4QAM-SP-CD-OOFDM-Power Meter Analyzer

Ch-1		Ch-2	
Tx : Frequency 1 (Hz)	193.04e+012	Tx : Frequency 2 (Hz)	193.08e+012
Tx : Wavelength 1 (nm)	1553.006931205968	Tx : Wavelength 2 (nm)	1552.6851975453
Tx : Signal Power 1 (dBm)	8.428425877725429	Tx : Signal Power 2 (dBm)	8.435258725429
Tx : Signal Power 1 (W)	0.006963740640675417	Tx : Signal Power 2 (W)	0.00696340675417
Tx : SNR 1 (dB)	5.07761057256892	Tx : SNR 2 (dB)	5.07761057256892
Tx : OSNR 1 (dB)	11.39310896258	Tx : OSNR 2 (dB)	11.39310896258
Rx : Signal Power 1 (dBm)	-3.227754557470081	Rx : Signal Power 2 (dBm)	-3.1816330985694
Rx : Signal Power 1 (W)	0.00047558101288308	Rx : Signal Power 2 (W)	0.00048065264684
Rx : SNR 1 (dB)	31.18251648033689	Rx : SNR 2 (dB)	31.228637954128
Rx : OSNR 1 (dB)	33.22371630689614	Rx : OSNR 2 (dB)	33.269837710053
Gain 1 (dB)	-8.040851074356159	Gain 2 (dB)	-7.9958983921381
Noise Figure 1 (dB)	-100	Noise Figure 2 (dB)	-100

Ch-3		Ch-4	
Tx : Frequency 3 (Hz)	193.12e+012	Tx : Frequency 4 (Hz)	193.16e+012
Tx : Wavelength 3 (nm)	1552.363597763049	Tx : Wavelength 4 (nm)	1552.04213875958
Tx : Signal Power 3 (dBm)	8.461000025429	Tx : Signal Power 4 (dBm)	8.39284756001
Tx : Signal Power 3 (W)	0.006963740640675417	Tx : Signal Power 4 (W)	0.00696340675417
Tx : SNR 3 (dB)	5.07761057256892	Tx : SNR 4 (dB)	11.39310896258
Tx : OSNR 3 (dB)	11.39310896258	Tx : OSNR 4 (dB)	2.0694929631618
Rx : Signal Power 3 (dBm)	-3.152995765980342	Rx : Signal Power 4 (dBm)	-3.1347998031261
Rx : Signal Power 3 (W)	0.00048383850035425	Rx : Signal Power 4 (W)	0.00048586993757
Rx : SNR 3 (dB)	31.25446394028165	Rx : SNR 4 (dB)	31.2726599031373
Rx : OSNR 3 (dB)	33.2956637668409	Rx : OSNR 4 (dB)	33.3138597269198
Gain 3 (dB)	-7.994225957571536	Gain 4 (dB)	-7.97709179203907
Noise Figure 3 (dB)	-100	Noise Figure 4 (dB)	-100
Ch-5		**Ch-6**	
Tx : Frequency 5 (Hz)	193.2e+012	Tx : Frequency 6 (Hz)	193.24e+012
Tx : Wavelength 5 (nm)	1551.720797101449	Tx : Wavelength 6 (nm)	1551.39959635662
Tx : Signal Power 5 (dBm)	8.91827364550011	Tx : Signal Power 6 (dBm)	8.1791823847561
Tx : Signal Power 5 (W)	0.00696340675417	Tx : Signal Power 6 (W)	0.00696340675417
Tx : SNR 5 (dB)	5.07761057256892	Tx : SNR 6 (dB)	5.07761057256892
Tx : OSNR 5 (dB)	11.39310896258	Tx : OSNR 6 (dB)	11.39310896258
Rx : Signal Power 5 (dBm)	-3.152074162633793	Rx : Signal Power 6 (dBm)	-3.1922367130218
Rx : Signal Power 5 (W)	0.000483941185171946	Rx : Signal Power 6 (W)	0.00047948649464
Rx : SNR 5 (dB)	31.2553855436282	Rx : SNR 6 (dB)	31.2124134803327
Rx : OSNR 5 (dB)	33.29658537018744	Rx : OSNR 6 (dB)	33.2536133069252
Gain 5 (dB)	-7.966676969424751	Gain 6 (dB)	-8.0059690958818
Noise Figure 5 (dB)	-100	Noise Figure 6 (dB)	-100
Ch-7		**Ch-8**	
Tx : Frequency 7 (Hz)	193.28e+012	Tx : Frequency 8 (Hz)	193.32e+012
Tx : Wavelength 7 (nm)	1551.078528559603	Tx : Wavelength 8 (nm)	1550.75759362147
Tx : Signal Power 7 (dBm)	8.4284258725429	Tx : Signal Power 8 (dBm)	8.4284258725429
Tx : Signal Power 7 (W)	0.00696340675417	Tx : Signal Power 8 (W)	0.00696340675417
Tx : SNR 7 (dB)	5.07761057256892	Tx : SNR 8 (dB)	5.07761057256892
Tx : OSNR 7 (dB)	11.39310896258	Tx : OSNR 8 (dB)	11.39310896258
Rx : Signal Power 7 (dBm)	-3.167264528716494	Rx : Signal Power 8 (dBm)	-3.1856492392081
Rx : Signal Power 7 (W)	0.00048225145555109	Rx : Signal Power 8 (W)	0.00048021427078

Rx : SNR 7 (dB)	31.237385664691	Rx : SNR 8 (dB)	31.219000954141
Rx : OSNR 7 (dB)	33.27858549125024	Rx : OSNR 8 (dB)	33.260200780566
Gain 7 (dB)	-8.009466956953016	Gain 8 (dB)	-8.0269509043292
Noise Figure 7 (dB)	-100	Noise Figure 8 (dB)	-100

Table (A.3): 16QAM-DP-8WDM- Power Meter Analyzer

Ch-1		Ch-2	
Tx : Frequency 1 (Hz)	193.15279814e+012	Tx : Frequency 2 (Hz)	193.19013919e+012
Tx : Wavelength 1 (nm)	1552.1	Tx : Wavelength 2 (nm)	1551.8
Tx : Signal Power 1 (dBm)	-0.56003069073009	Tx : Signal Power 2 (dBm)	-0.55826838109326
Tx : Signal Power 1 (W)	0.000879016304611	Tx : Signal Power 2 (W)	0.000879373070506
Tx : SNR 1 (dB)	99.4399693092694	Tx : Noise Power 2 (dBm)	-100
Tx : Noise Power 1 (dBm)	-100	Tx : Noise Power 2 (W)	12.4394901740e-015
Tx : Noise Power 1 (W)	9.0676576681e-015	Tx : SNR 2 (dB)	99.44173161890676
Tx : OSNR 1 (dB)	99.4399693092694	Tx : OSNR 2 (dB)	99.44173161890676
Rx : Signal Power 1 (dBm)	-0.56252112573037	Rx : Signal Power 2 (dBm)	-0.5557465090721
Rx : Signal Power 1 (W)	0.000878512382962	Rx : Signal Power 2 (W)	0.000879883855393
Rx : Noise Power 1 (dBm)	-30.494925501981	Rx : Noise Power 2 (dBm)	-30.472526325648
Rx : Noise Power 1 (W)	0.8922929251e-006	Rx : Noise Power 2 (W)	0.89690690481e-006
Rx : SNR 1 (dB)	29.93240437625078	Rx : SNR 2 (dB)	29.91677981657653
Rx : OSNR 1 (dB)	33.43488455959241	Rx : OSNR 2 (dB)	33.41925999991815
Gain 1 (dB)	-0.00249043499918	Gain 2 (dB)	0.002521872025278
Noise Figure 1 (dB)	23.98181756883528	Noise Figure 2 (dB)	23.99827893547906
Ch-3		**Ch-4**	
Tx : Frequency 3 (Hz)	193.22749468e+012	Tx : Frequency 4 (Hz)	193.26486462e+012
Tx : Wavelength 3 (nm)	1551.5	Tx : Wavelength 4 (nm)	1551.2
Tx : Signal Power 3 (dBm)	-0.56101212691061	Tx : Signal Power 4 (dBm)	-0.55651265999462
Tx : Signal Power 3 (W)	0.000878817683153	Tx : Signal Power 4 (W)	0.000879728646419
Tx : Noise Power 3 (dBm)	-100	Tx : Noise Power 4 (dBm)	-100
Tx : Noise Power 3 (W)	0	Tx : Noise Power 4 (W)	0
Tx : SNR 3 (dB)	99.43898787306189	Tx : SNR 4 (dB)	99.44348734000405
Tx : OSNR 3 (dB)	99.43898787306189	Tx : OSNR 4 (dB)	99.44348734000405
Rx : Signal Power 3 (dBm)	-0.56637912913667	Rx : Signal Power 4 (dBm)	-0.55585887775485

Rx : Signal Power 3 (W)	0.000877732313131	Rx : Signal Power 4 (W)	0.00087986108014
Rx : Noise Power 3 (dBm)	-30.479975764601	Rx : Noise Power 4 (dBm)	-30.474211464457
Rx : Noise Power 3 (W)	0.8953697620e-006	Rx : Noise Power 4 (W)	0.8965589567e-006
Rx : SNR 3 (dB)	29.91359663546275	Rx : SNR 4 (dB)	29.91835258670429
Rx : OSNR 3 (dB)	33.41607681880438	Rx : OSNR 4 (dB)	33.42083277004592
Gain 3 (dB)	-0.00536700220105	Gain 4 (dB)	0.00065378226816
Noise Figure 3 (dB)	23.99791177951545	Noise Figure 4 (dB)	23.99679582948837
Ch-5		Ch-6	
Tx : Frequency 5 (Hz)	193.30224901e+012	Tx : Frequency 6 (Hz)	193.33964787e+012
Tx : Wavelength 5 (nm)	1550.9	Tx : Wavelength 6 (nm)	1550.6
Tx : Signal Power 5 (dBm)	-0.55769382714431	Tx : Signal Power 6 (dBm)	-0.55660616856289
Tx : Signal Power 5 (W)	0.000879489415758	Tx : Signal Power 6 (W)	0.000879709704384
Tx : Noise Power 5 (dBm)	-100	Tx : Noise Power 6 (dBm)	-100
Tx : Noise Power 5 (W)	0	Tx : Noise Power 6 (W)	0
Tx : SNR 5 (dB)	99.44230617285568	Tx : SNR 6 (dB)	99.44339383143768
Tx : OSNR 5 (dB)	99.44230617285568	Tx : OSNR 6 (dB)	99.44339383143768
Rx : Signal Power 5 (dBm)	-0.56035784587483	Rx : Signal Power 6 (dBm)	-0.55596448251346
Rx : Signal Power 5 (W)	0.000878950090998	Rx : Signal Power 6 (W)	0.000879839691736
Rx : Noise Power 5 (dBm)	-30.4930045836007	Rx : Noise Power 6 (dBm)	-30.513155313187
Rx : Noise Power 5 (W)	0.8926876810e-006	Rx : Noise Power 6 (W)	0.88855531434e-006
Rx : SNR 5 (dB)	29.93264673772953	Rx : SNR 6 (dB)	29.95719083066263
Rx : OSNR 5 (dB)	33.43512692107116	Rx : OSNR 6 (dB)	33.45967101400427
Gain 5 (dB)	-0.00266401873048	Gain 6 (dB)	0.000641686052341
Noise Figure 5 (dB)	23.98055889348598	Noise Figure 6 (dB)	23.95634649086932
Ch-7		Ch-8	
Tx : Frequency 7 (Hz)	193.37706121e+012	Tx : Frequency 8 (Hz)	193.41448903e+012
Tx : Wavelength 7 (nm)	1550.3	Tx : Wavelength 8 (nm)	1550
Tx : Signal Power 7 (dBm)	-0.55923469838358	Tx : Signal Power 8 (dBm)	-0.63375877959245
Tx : Signal Power 7 (W)	0.000879177429501	Tx : Signal Power 8 (W)	0.000864219620804
Tx : Noise Power 7 (dBm)	-100	Tx : Noise Power 8 (dBm)	-100
Tx : Noise Power 7 (W)	12.402109807e-015	Tx : Noise Power 8 (W)	23.1763063957e-015
Tx : SNR 7 (dB)	99.44076530161645	Tx : SNR 8 (dB)	99.36624122040763
Tx : OSNR 7 (dB)	99.44076530161645	Tx : OSNR 8 (dB)	99.36624122040763
Rx : Signal Power 7 (dBm)	-0.56328193809383	Rx : Signal Power 8 (dBm)	-0.6353964722345

Rx : Signal Power 7 (W)	0.000878358491019	Rx : Signal Power 8 (W)	0.000863893796508
Rx : Noise Power 7 (dBm)	-30.496779215996	Rx : Noise Power 8 (dBm)	-30.501064580683
Rx : Noise Power 7 (W)	0.8919121465e-006	Rx : Noise Power 8 (W)	0.89103249367e-006
Rx : SNR 7 (dB)	29.93349727786472	Rx : SNR 8 (dB)	29.86566810842257
Rx : OSNR 7 (dB)	33.43597746120635	Rx : OSNR 8 (dB)	33.36814829176419
Gain 7 (dB)	-0.00404723971537	Gain 8 (dB)	-0.00163769200105
Noise Figure 7 (dB)	23.97650877429267	Noise Figure 8 (dB)	23.96899386979189

References

[1] C. V. N. Index, “Cisco visual networking index: Forecast and methodology, 2014-2019,” White paper, May 27, 2015.

[2] C. V. N. J. C. W. p. Index, “Cisco Global Cloud Index: Forecast and Methodology, 2015–2020”, White paper, June 6, 2016.

[3] G. P. Agrawal, Fiber-optic communication systems, 4th ed. Wiley, 2010.

[4] K. Kao and G. A. Hockham, “Dielectric-fibre surface waveguides for optical frequencies,” in Proceedings of the Institution of Electrical Engineers, IET, vol. 113, pp. 1151‑1158, 1966.

[5] T. Miya, Y. Terunuma, T. Hosaka, and T. Miyashita, “Ultimate low loss single-mode fibre at 1.55 μm,” Electronics Letters, vol. 15, no. 4, pp. 106‑108, 1979.

[6] E. Desurvire, Erbium-doped fiber amplifiers: principles and applications. Wiley New York, 1st eddition, March 1994.

[7] J. Zyskind and A. Srivastava, Optically amplified WDM networks. Academic press, 1st eddition, 2011.

[8] E. Ip, A. P. T. Lau, D. J. Barros, and J. M. Kahn, “Coherent detection in optical fiber systems,” Optics express, vol. 16, no. 2, pp. 753–791, 2008.

[9] S. J. Savory, “Digital coherent optical receivers: algorithms and subsystems,” IEEE Journal of Selected Topics in Quantum Electronics, vol. 16, no. 5, pp. 1164–1179, 2010.

[10] S. L. Olsson, J. Cho, S. Chandrasekhar, X. Chen, E. C. Burrows, and P. J. Winzer, “Record-High 17.3-bit/s/Hz Spectral Efficiency Transmission over 50 km Using Probabilistically Shaped PDM 4096-QAM,” in 2018 Optical Fiber Communications Conference and Exposition (OFC), IEEE, pp. 1–3, 2018.

[11] A. Ghazisaeidi, I. F. de J. Ruiz, R. Rios-Müller, L. Schmalen, P. Tran, P. Brindel, A. C. Meseguer, Q. Hu, F. Buchali, G. Charlet, J. Renaudier, "65Tb/s transoceanic transmission using probabilisticallyshaped PDM-64QAM," Proc. of the European Conference on Optical Communication (ECOC), postdeadline paper, 2016.

[12] P. J. Winzer and D. T. Neilson, "From scaling disparities to integrated parallelism: A decathlon for a decade," Journal of Lightwave Technology, vol. 35, no. 5, pp. 1099–1115, 2017.

[13] R. S. Tucker, "Green optical communications—part i: Energy limitations in transport," IEEE Journal of selected topics in quantum electronics, vol. 17, no. 2, pp. 245–260, 2011.

[14] X. Liu, "Evolution of Fiber-Optic Transmission and Networking toward the 5G Era", iScience, vol. 22, pp. 489–506, December 2019.

[15] J. Peter Winzer, T. David Neilson, And Andrew R. Chraplyvy, "Fiber-optic transmission and networking: the previous 20 and the next 20 years", OPTICS EXPRESS, vol. 26, no. 18, pp. 103-113, Sep 2018.

[16] G. P. Agrawal, Nonlinear Fiber Optics, Academic Press, 5th ed. 2012.

[17] A. Amari, Octavia A. Dobre, Ramachandran Venkatesan, O. S. Sunish Kumar, Philippe Ciblat, and Yves Jaouen, "A survey on fiber nonlinearity compensation for 400 Gbps and beyond optical communication systems", IEEE, vol. 19, pp. 3097 - 3113, June 2017.

[18] O. V. Sinkin, R. Holzlohner, J. Zweck, and C. R. Menyuk, "Optimization of the split-step fourier method in modeling optical-fiber communications systems," IEEE/OSA J. Lightw. Technol., vol. 21, no. 1, pp. 61–68, Jan. 2003.

[19] J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. Kumar Singh, A. Jukan, M. Chamania, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey", OSN. ScienceDirect, vol. 28, pp. 43 - 57, June 2018.

[20] R. Jones, “Machine Learning Methods in Coherent Optical Communication Systems,” Ph.D. Thesis submitted to Technical University of Denmark, Department of Photonics Engineering, December 2019.

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

[22] S. Zhang, F. Yaman, K. Nakamura, T. Inoue, V. Kamalov, L. Jovanovski, V. Vusirikala, E. Mateo, Y. Inada, and T. Wang, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks," Nature communications, vol. 10, no. 1, pp. 1-8, 2019.

[23] L. Salmela, M. Närhi, J. Toivonen, C. Billet, J. M. Dudley, and G. Genty, "Machine learning analysis of extreme events in optical fibre modulation instability," in Symposium on Future Prospects for Photonics: Sensing & Imaging, 2018.

[24] S. Tajammul Ahmad and P. Kumar K, " Radial Basis Function Neural Network Nonlinear Equalizer for 16-QAM Coherent Optical OFDM," DOI 10.1109, IEEE Photonics Technology Letters, 2016

[25] R. T. Jones, S. Gaiarin, M. P. Yankov, and D. Zibar, “Time-DomainNeural Network Receiver for Nonlinear Frequency Division Multiplexed Systems,” IEEE Photonics Technology Letters, vol. 30, no. 12, pp. 1079–1082, 2018.

[26] I. Aldaya, E. Giacoumidis, A. Tsokanos, M. Jarajreh, Y. Wen, J. Wei, G. Campuzano, M. L. F. Abbade, AND L. P. Barry, "Compensation of nonlinear distortion in coherent optical OFDM systems using a MIMO deep neural network-based equalizer," Optics Letters, Vol. 45, No. 20, pp. 5820-5824, 2020.

[27] W. Zhang, A. Hajomer, X. Yang, and W. Hu, "Neural Networks for PAPR Reduction in Optical OFDM Signal Transmission," International Conference on Transparent Optical Networks, ICTON 2019.

[28] R.J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, B. Goebel, “Capacity limits of optical fiber networks,” J. Light. Technol., vol. 28, no. 4, pp. 662–701, 2010.

[29] J. G. Proakis, Digital communications, 4th ed., McGraw-Hill, 2000.

[30] M. Seimetz, High-order modulation for optical fiber transmission, Springer, 2009.

[31] H. Brundage, "Designing a Wireless Optical Communication System", M.Sc. Thesis, Massachusetts Institute of Technology (MIT), February 2010.

[32] S. Hranilovic, Wireless Optical Communication Systems, Springer Publisher, 2005.

[33] J. Simpson, "A1 Mbps Underwater Communications System using LEDs and Photodiodes with Signal Processing Capability", Master Thesis, submitted to North Carolina State University, 2007.

[34] T. Pfau, S. Hoffmann, and R. Noé, “Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations,” Journal of Lightwave Technology, vol. 27, no. 8, pp. 989–999, 2009.

[35] M. H. Langaroodi, "Design and Performance of a 1550nm Free Space Optical Communications Link", M.Sc. Thesis, submitted to California State University, Northridge, December 2013.

[36] R. T., " Flexible and high data-rate coherent optical transceivers", PhD Thesis, submitted to Eindhoven University of Technology, 2017.

[37] Z. H. Peric, I. B. Djordjevic, S. M. Bogosavljevic, and M. C. Stefanovic, “Design of signal constellations for Gaussian channel by using iterative polar quantization,” in Electrotechnical Conference, 1998. MELECON 98., 9th Mediterranean, IEEE, vol. 2, 1998, pp. 866–869.

[38] X. Chen, "Design and Optimization of Optical Fiber Based PSK Demodulators for High-Bit-Rate Optical Networks", PhD Thesis, University of Polytechnic of Madrid, 2015.

[39] C. Cahn, “Combined digital phase and amplitude modulation communication systems,” IRE Transactions on Communications Systems, vol. 8, no. 3, pp. 150–155, 1960.

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

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

[42] J. M. Senior, M. Yousif Jamro, Optical Fiber Communications Principles and Practice, 3rd ed. Pearson Education Limited 2009.

[43] D. Marcuse, C. R. Menyuk, P. K. A. Wai, “Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence,” J. Light. Technol., vol. 15, no. 9, pp. 1735-1746, 1997.

[44] S. V. Manakov, “On the theory of two-dimensional stationary selffocusing of electromagnetic waves,” Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki, vol. 65, pp. 505-516, 1973.

[45] P. K. A. Wai, C. R. Menyuk, H. H. Chen, “Stability of solitons in randomly varying birefringent fibers,” Opt. Lett., vol. 16, no. 16, pp. 1231-1233, 1991.

[46] R. J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, B. Goebel, “Capacity limits of optical fiber networks,” J. Light. Technol., vol. 28, no. 4, pp. 662–701, 2010.

[47] R. Dar, M. Shtaif, M. Feder, “New bounds on the capacity of the nonlinear fiber-optic channel,” Opt. Lett., vol. 39, no. 2, pp. 398–401, 2014.

[48] A. Mecozzi, R. J. Essiambre, “Nonlinear Shannon limit in pseudolinear coherent systems,” J. Light. Technol., vol. 30, no. 12, pp. 2011-2024, 2012.

[49] F. Nadeem Khan, Q. Fan, C. Lu and A. Pak Tao L., "Machine learning methods for optical communication systems and networks," Optical Fiber Telecommunications VII, DOI: https://doi.org/10.1016/B978-0-12-816502-7.00029-4, PP 921-978, 2020.

[50] P. Poggiolini, G. Bosco, A. Carena, “The GN-model of fiber non-Linear propagation and its applications,” J. Light. Technol., vol. 32, no. 4, pp. 694-721, 2014.

[51] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, “Properties of nonlinear noise in long, dispersion-uncompensated fiber links,” Opt. Express, vol. 21, no. 22, pp. 25685-25699, 2013.

[52] E. P. da Silva, K. J. Larsen, D. Zibar, “Impairment mitigation in superchannels with digital backpropagation and MLSD,” Optics Express, vol. 23, no. 23, pp. 29493-2950, 2015.

[53] C. Laperle and M. O’Sullivan, “Advances in high-speed DACs, ADCs, and DSP for optical coherent transceivers,” Journal of Lightwave Technology, vol. 32, no. 4, pp. 629–643, 2014.

[54] X. Zhou and C. Xie, Enabling technologies for high spectral-efficiency coherent optical communication networks. John Wiley & Sons, 2016.

[55] E. Ip and J. M. Kahn, “Compensation of dispersion and nonlinear impairments using digital backpropagation,” Journal of Lightwave Technology, vol. 26, no. 20, pp. 3416–3425, 2008.

[56] A. Amari, O. A. D., R. Venkatesan, O. S. Sunish, P. Ciblat, and Y. Jaouen, “A survey on fiber nonlinearity compensation for 400 Gbps and beyond optical communication systems”, IEEE, vol. 19, pp. 3097 - 3113, June 2017.

[57] C.E. Shannon, "A Mathematical Theory of Communication", The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, 1948.

[58] Z. William Shieh and I. Djordjevic, OFDM for Optical Communications, Elsevier Press, 2010.

[59] K.d Alatawi, "High Data Rate Coherent Optical OFDM System for Long-Haul Transmission", M.Sc. Thesis, University of Denver, 2013.

[60] P. J. Winzer, “Scaling optical fiber networks: Challenges and solutions,” Opt. Photonics News, vol. 26(3), pp. 28–35, Sep 2015.

[61] ITU-T Recommendation G.976, Test methods applicable to optical fiber submarine cable systems, ITY-T, Jul 2007.

[62] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, pp. 1097–1105, 2012.

[63] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al., “Mastering the game of Go without human knowledge,” Nature, vol. 550, no. 7676, p. 354, 2017.

[64] F. A. Aoudia and J. Hoydis, “End-to-End Learning of Communications Systems Without a Channel Model,” arXiv preprint arXiv:1804.02276, 2018.

[65] C. M. Bishop, Neural networks for pattern recognition. Oxford university press, 1995.

[66] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning. MIT press Cambridge, 2016, vol. 1.

[67] D. J. MacKay and D. J. Mac Kay, Information theory, inference and learning algorithms. Cambridge university press, 2003.

[68] G. I., B. Y., C. A. Deep Learning; MIT Press: Cambridge, MA, USA, 2016; Chapter 5.

[69] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, “Deep learning of geometric constellation shaping including fiber nonlinearities,” European Conf. Optical Commun. (ECOC), Sep. 2018.

[70] G. E. Dahl, T. N. Sainath, and G. E. Hinton, “Improving deep neural networks for LVCSR using rectified linear units and dropout,” in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, IEEE, 2013, pp. 8609–8613.

[71] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” 2018.

[72] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[73] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, “Automatic differentiation in machine learning: a survey.,” Journal of machine learning research, vol. 18, no. 153, pp. 1–153, 2017.

[74] H. Nielsen, R. Theory of the backpropagation neural network. In Neural Networks for Perception; Academic Press: Cambridge, MA, USA, pp. 65–93, 1992.

[75] M. K. P. Machine Learning: A Probabilistic Perspective. The MIT Press, Cambridge, Massachusetts, 2012.

[76] K., Diederik, and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[77] E. Giacominis, et al., "Intra and inter-channel nonlinearity compensation in WDM coherent optical OFDM using artificial neural network based nonlinear equalization", Optical Fiber Communication Conference OSA Technical Digest (online) (Optical Society of America, 2017), paper Th2A.62.

الملخص

نما الطلب العالمي على حركة البيانات الضخمة بشكل كبير خلال السنوات الماضية وسيتجاوز الطلب المستقبلي قدرة العمود الفقري الحالي للإنترنت، حيث تشكل أنظمة الألياف الضوئية البنية التحتية المركزية له. لذلك، يعد تحسين الأداء أمرًا ضروريًا لهذه الأنظمة وجعلها أكثر مرونة وقابلة للتطوير لتلبية هذا الطلب المتزايد.

تتطلب هذه القيود نماذج أفضل لقنوات الألياف الضوئية وأنظمة الإرسال مع أفضل الطرق للتحسين والمعالجة، مما يؤدي هذا إلى زيادة التكلفة والتعقيد. لذلك، يمكن اعتبار الخوارزميات المستقلة والفعالة حلاً جيدًا في الوقت الحالي للمساعدة في تحسين كفاءة هذه الأنظمة.

اعتمدت الخوارزميات التقليدية على طريقة Split-Fourier (SSFM) تقسيم فورير وطرق الانتشار العكسي الرقمي (DBP) من خلال إيجاد الحل المقدر لـ "معادلة شرود نغر غير الخطية "(NLSE) ، لمعالجة التحدي الأكثر جوهرية الذي يواجه مثل هذه الأنظمة وهو الانحطاط غير الخطي ، لكن تنفيذه يتطلب الكثير من الموارد لمعالجة الإشارات بالإضافة إلى فهم دقيق وعالي المستوى لهذه الأنظمة. بينما تعد خوارزميات الذكاء الاصطناعي (AI) بتقديم أفضل الحلول للتحديات المستقبلية الموصوفة من خلال استخدامها لتحديد وحل أوجه القصور في الأنظمة المعقدة للغاية والبيئات المعقدة دون الحاجة إلى تطوير نموذج للأجهزة (المادية).

طبقت هذه الأطروحة التعلم الآلي على أساس الشبكة العصبية العميقة (DNN) كطريقة تعلم تحت الإشراف للأنظمة البصرية لرفع مستوى أدائها. اقترحت هذه الدراسة بناء تقنية NLC للتعويض عن الانحطاط غير الخطي الذي يحدث في أنظمة الاتصالات البصرية. تم تصميم النظام ومحاكاته بالكامل في الوقت الفعلي باستخدام تطبيق MATLAB. (V.2020a) مع تطبيق .Optisystem (الإصدار 17.0.0) وان الاختبار / التحسين تم إجراؤه في ثلاثة اتجاهات:

أولاً: في أنظمة بصرية متماسكة أحادية الاستقطاب مع 4-16 تشكيل QAM مع وبدون OFDM متعدد الموجات الحاملة الفرعية ، يتم نقله عبر الألياف أحادية الوضع القياسية أحادية الموجة (SSMF) بسرعة 60 جيجابت في الثانية لمسافة ارتباط تبلغ 3000 كيلو متر.

ثانيًا ، في الاستقطاب الأحادي ، الأنظمة البصرية المتماسكة مع تعديل QAM 16 مع ناقل فرعي متعدد OFDM وتعدد إرسال بتقسيم الطول الموجي 8 قنوات WDM مع تباعد تردد 40 جيجاهرتز ، يتم نقله عبر مسافة ارتباط في ألياف أحادية الوضع (SSMF) قياسية 3000 كيلو متر مع 480 جيجابت في الثانية تمامًا.

أخيرًا ، في الاستقطاب المزدوج ، والأنظمة البصرية المتماسكة مع تعديل QAM 16، وتعدد الإرسال بتقسيم الطول الموجي 8قنوات WDM مع تباعد تردد 37.5 جيجاهرتز (3 نانومتر) ، يتم نقله عبر مسافة ارتباط ألياف أحادية الوضع (SSMF) قياسية 3000 كيلو متر مع تمامًا 960 جيجابت في الثانية.

أظهرت النتائج المحققة كفاءة التصميم المقترح والتحسن الملحوظ في الأداء ، خاصة في معدل البيانات ومدى الارتباط مقارنة بالأعمال المماثلة.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل
كلية الهندسة
قسم الهندسة الكهربائية

تحسين الضعف البصري غير الخطي لنظام النقل الضوئي بالاعتماد على الشبكة العصبية العميقة

أطروحة
مقدمة الى قسم الهندسة الكهربائية - كلية الهندسة - جامعة بابل
كجزء من متطلبات نيل درجة الدكتوراه في الفلسفة في الهندسة / الهندسة الكهربائية / الكترونيك واتصالات

من قبل
علاء حمود عبد جر اح

بإشراف

الأستاذ
د. ابراهيم عبد الله مرداس

1443 هـ

2022 م