Title
Compensation of Nonlinear Optical Fiber Impairments Using Coding and Electronic Equalizer

Author
Taghavi Nasr Abadi, Zeinab

Publication Date
2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Compensation of Nonlinear Optical Fiber Impairments Using Coding and Electronic Equalizer

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering
(Communication Theory and Systems)

by

Zeinab Taghavi

Committee in charge:

Professor George C. Papen, Chair
Professor Philip E. Gill
Professor David A. Meyer
Professor Paul H. Siegel
Professor Stojan Radic

2011

The dissertation of Zeinab Taghavi is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2011

DEDICATION

To my mother and the soul of my father

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of many people. Many thanks to my adviser, professor George Papen, whose wisdom, guidance, support, patience, and enthusiasm helped me along the way. I also gratefully acknowledge the valuable guidance offered by Dr. Nikola Alic. I especially thank professor Stojan Radic for all his sincere support of my work. I greatly enjoyed working in the UCSD Photonics Lab under his supervision. I am grateful to professor Paul Siegel and late professor Jack Wolf from whom I learned the field of coding theory during the first two years of my research. I would like to thank my other committee members, professor Philip Gill, professor David Meyer, and professor Amin Vahdat for their helpful comments on my research.

Among my group mates in the STAR Lab and UCSD Photonics Lab, I particularly thank Dr. Evgeny Myslivets for his help during my experimental works. I am also grateful to my other group mates in the STAR Lab, Brian Butler, Panu Chaichanavong, Ismail Demirkan, Junsheng Han, Seyhan Karakulak, Henry Pfister, Joseph Soriaga, and Zheng Wu, and also to the current and past members of the UCSD Photonics Lab, Dr. Sanja Zlatanovic, Dr. Slaven Moro, Dr. Rui Jiang, Dr. Jose M. Chavez Boggio, James B. Coles, and Faezeh Gholami, who kindly created a fascinating environment for research.

I thank my wonderful friends in San Diego who have filled my life with so many unforgettable moments. In particular, I am grateful to Ghazaleh Rais-Esmaili, Ali Afsahi, Shadi Sagheb, Koohyar Minoo, Maryam Hosseinzadeh, Parham Minoo, Solmaz Kheiravar, Shahram Mahdavi, Simin Kazemi, Mohammad Reza Gharavi, Sara Motamedi, Ehsan Saberian, Elham Pirmoradian, Javad Kazemitabar, Raheleh Dilmaghani, Diba Mirza, Neda Nouri, Esra Vural, Hamed, Alireza, and Maryam Masnadi.

More personally, I am grateful for the support I have received from my family. I am thankful to my brother and his wife Hossein and Mona for being the closest friends. I owe everything I have achieved in my life to my mother, who was my first teacher and has provided me with unconditional love, guidance, and sacrifice throughout my life, and I dedicate this thesis to her. I would like to thank my uncle, Masoud, for being a great source of support in my life. The last but most importantly, I am grateful to my husband, Hamid, who has been my love and inspiration.

The text in Chapter 2, in part, is a reprint of the material as it appears in Zeinab Taghavi, Nikola Alic, and George Papen, "Maximum-Likelihood Detection and Constrained Coding on Optical Channels, " *Journal of Lightwave Technology*, volume 27, Issue 11, pp. 1469-1479 (2009). The dissertation author was the primary researcher and/or author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

VITA AND PUBLICATIONS

2001	Bachelor of Science, Electrical Engineering, Sharif University of Technology, Iran
2003	Master of Science, Electrical Engineering, Sharif University of Technology, Iran
2003-2011	Research Assistant, Electrical and Computer Engineering, University of California, San Diego, USA
2011	Doctor of Philosophy, Electrical and Computer Engineering, University of California, San Diego, USA

Zeinab Taghavi, Nikola Alic, George Papen, “Maximum-Likelihood Detection and Constrained Coding on Optical Channels,” *J. Lightw. Technol.*, vol. 27, no. 11, pp. 1469-79, Jun. 2009.

Zeinab Taghavi, Nikola Alic, George Papen, “Modulation Coding for Optical Channels,” Proc. of *IEEE Leos 2007-Summer Topicals*, Portland, Jul. 2007. *(Invited talk)*

ABSTRACT OF THE DISSERTATION

Compensation of Nonlinear Optical Fiber Impairments Using Coding and Electronic Equalizer

by

Zeinab Taghavi

Doctor of Philosophy in Electrical Engineering

(Communication Theory and Systems)

University of California San Diego, 2011

Professor George C. Papen, Chair

Ultra-high capacity fiber optic systems with data rates exceeding 100 Giga bits per second per fiber are currently being deployed with higher capacity systems in development. The requirement of a minimum energy per bit for reliable communication means that the power launched into a single fiber is now at a level where significant nonlinearites exist. Nonlinearities can also be produced for lower power intensity-modulated systems because of the square-law nature of sensors.

In order to maximize the information capacity, the combined channel that includes a combination of nonlinear impairments along with additional linear impairments must be mitigated. This mitigation can be achieved by a combination of modulation coding at the transmitter and equalization at the receiver. The development of these

techniques for nonlinear channels is significantly more complex than the corresponding techniques for linear channels because of the nature of the nonlinearity and the extremely high data rate. This rate limits the complexity of the equalization algorithm.

This thesis presents modulation coding and equalization techniques for several nonlinear fiber optic channels. We consider two classes of nonlinearity. The first arises from the combination of linear dispersion in an optical fiber and square-law sensing. The second arises from nonlinear propagation characteristics caused by a power-dependent index of refraction change called a Kerr nonlinearity.

A variety of nonlinear channel models can be constructed from these two fundamental forms of nonlinearity along with linear impairments. The dominant linear impairment is dispersion. One form occurs when the propagation characteristics for each mode depend on the frequency. A second form of dispersion arises because different polarization modes can have different propagation characteristics.

The research premise of this thesis is that a combination of modulation coding, sequence estimation (both single-user and multi-user) and nonlinear equalization based on heuristic algorithms can produce significant performance improvement relative to published techniques. We present several abstracted scenarios reflecting practical systems where a combination of these techniques is effective. We also describe situations where they are ineffective. These results lay the foundation for further work using these techniques to optimize specific nonlinear channels.

1 Introduction

1.1 Introduction

Digital communication using optical fiber has revolutionized both long distance trans-oceanic systems as well as short distance data communications systems. Optical fiber provides a small, inexpensive, low loss, and high bandwidth medium. Optical fiber is particularly advantageous for long-distance communications because of its exceptionally low loss compared to electrical cables. A typical fiber has less loss in 100 km ($\sim$ 20 dB) than a typical electrical cable in 100 m ($\sim$ 24 dB). For long distance communication, this property reduces the expense of adding repeaters for digital regeneration.

In the 1980's, telephone companies started to install regional fiber optic telecommunications networks throughout the world. This created the need to expand fiber's transmission capabilities through increasingly sophisticated modulation and compensation techniques. In 1990, Bell Labs transmitted a 2.5 Gbps signal over 7,500 km. Later in 1998, researchers transmitted 100 simultaneous optical signals, each at a data rate of 10 Gbps over a distance of about 400 km. In that experiment, dense wavelength-division multiplexing (DWDM technology), which allows multiple wavelengths to be combined into one optical signal, was used. The total data rate on one fiber in that experiment increased to 1 Tbps [2]. Recent laboratory systems constructed by Bell Labs have multiplexed 155 channels, each carrying 100 Gbps over 7000 km. This produces a total rate of 15.5 Tbps [3]. This technology is currently coming to the market. In March 2011, Alcatel-Lucent announced that they achieved a per channel transmission rate of 256 Gbps over a distance of 400 kilometers. This rate is more than twice the previous record. This system used 64-QAM modulation (Quadrature Amplitude Modulation) scheme [4].

1.2 Coherent and Non-Coherent Modulation

The transmission capacities of long-haul and ultra-long-haul fiber optic communications systems using modulation formats based on the intensity of the optical signal have been significantly increased by the introduction of erbium doped fiber amplifiers (EDFA), dense wavelength division multiplexing (DWDM), dispersion compensation, and forward error code technologies. This evolution started in the mid 1990s. Until late 1990's, practical systems used noncoherent modulation and detection. Noncoherent fiber optic systems are based on an intensity-modulated direct detection (IMDD) system using on/off-keying (OOK). A typical detector for this type of system is a photodiode that produces an electrical current that is proportional to the instantaneous power of the received optical signal. Therefore, information about the phase of the optical field is lost. The loss of the phase makes the detector nonlinear with respect to the optical signal. This format has been sufficient to address data rates up to 10 Gbps per channel. Intensity modulation is relatively inexpensive and straightforward to deploy.

In order to extend the reach and data capacity, several advancements have taken place over the past decade, including but not limited to: (1) adoption of a higher-order modulation formats that use both the phase and the amplitude, as opposed to using only intensity; (2) the concurrent development of optical coherent detection to preserve the phase information of the optical signal; and (3) progress in adaptive electrical equalization technology. The combination of these technologies can increase the spectral efficiency and robustness of signal transmission in the presence of noise and transmission impairments. The fact that coherent detection has the ability to distinguish the optical phase and polarization means that modulation formats can take full advantage of these additional degrees of freedom relative to using a single polarization and intensity modulation.

Using advanced modulation formats, the spectral efficiency (bit/s/Hz) can be increased by transmitting more data within the same optical bandwidth. In intensity modulated systems, the received electrical signal is proportional to the power of the optical signal because of the square-law characteristics of the optical detector. Therefore, linear channel impairments in the optical domain can produce a nonlinear dependence in the electrical domain. However, in coherent detection, the detected electrical signal is proportional to the optical signal, and the mitigation of linear transmission impairments in the electrical domain is more readily achieved [5].

1.3 Optical Fiber Channel Characteristics

Similar to virtually all other telecommunication systems, the transmitted optical signal changes as it propagates in the fiber. These changes can be either linear or nonlinear with respect to the optical signal. One of the basic changes is signal attenuation. The minimum loss in silica fibers is at a wavelength of approximately 1550 nm, and is about 0.2 dB/km. The operating regime of the current long-haul fiber optic system is in this wavelength band. The loss is compensated by using fiber amplifiers based on an optical fiber doped with erbium (EDFA) [6].

For a symbol transmission rate less than about 40 giga-symbols per second, there are three major signal impairments mechanisms in the fiber: (1) Group velocity dispersion (GVD), also known as chromatic dispersion (CD), (2) Nonlinear effects based on a third-order Kerr nonlinearity, and (3) Polarization mode dispersion (PMD). These effects degrade the signal and produce symbol errors at the receiver if they are not mitigated. An overview of these three characteristics is presented in this section.

1.3.1 Nonlinear Schrödinger Equation

We define the transmitted modulated signal in the fiber at a length z and at time t as $A(z,t)$.

$$A(z,t) = E(z,t)\exp(j2\pi f(t-\beta_1 z)),$$

where f is the carrier frequency, $E(z,t)$ is the envelope or low-pass equivalent signal, and β_1 is the group delay. It can be shown that the envelope $E(z,t)$ satisfies the nonlinear

Schrödinger equation [7]

$$\frac{\partial E}{\partial z} = -\frac{\alpha}{2}E + j\frac{\beta_2}{2}\frac{\partial^2 E}{\partial \tau^2} + j\gamma|E|^2 E. \tag{1.1}$$

In this equation, τ is the time measured in a reference frame moving at the group velocity defined as $(1/\beta_1)$, α is the fiber loss factor, and β_2 is the group velocity dispersion (GVD) parameter. The first term on the right side of the equation is the fiber loss. The second term in (1.1) is responsible for the chromatic dispersion. The third term is the fiber nonlinearity characterized by the nonlinear coefficient γ. This term governs the nonlinear signal propagation characteristic of fiber. Given this equation, the effect of dispersion and nonlinearity can be assessed. This equation is used for simulating nonlinear signal propagation in the fiber.

1.3.2 Dispersion

Group velocity dispersion (GVD) is a phenomenon in which the phase velocity of a wave depends on its frequency[1]. The physics of GVD has been carefully studied [7]. Group velocity dispersion is sometimes called chromatic dispersion (CD) to emphasize its wavelength-dependence. There are generally two sources of dispersion: (1) Material dispersion, whereby response of the glass that is used for the fiber depends on the frequency of the carrier, and (2) Waveguide dispersion, where the propagation characteristics of the fiber depend on the fraction of the field in core and cladding. If in (1.1) it is assumed that γ and α are equal to zero, then the effect of the fiber on the envelope

[1] In this thesis, we use the terms frequency and wavelength interchangeably.

E can be determined as follows

$$\mathcal{F}\{E\}(z = L, \omega) = \mathcal{F}\{E\}(z = 0, \omega)e^{\frac{j\beta_2 L\omega^2}{2}}, \tag{1.2}$$

where $\mathcal{F}$ stands for the Fourier transform of the envelope E and ω is the angular frequency. More details about the derivation of this equation will be given in later chapters. It can be seen that in the absence of any other effect, chromatic dispersion is an all-pass quadratic phase function. This frequency-dependent phase causes spreading of a transmitted pulse. The received pulses can then interfere producing intersymbol interference (ISI).

In a single-mode fiber, with moderate power levels of the transmitted signal such that the nonlinear term in (1.1) can be neglected, intersymbol interference caused by chromatic dispersion is the dominant source of signal degradation. One method to compensate chromatic dispersion is the use of dispersion compensation fiber (DCF). However, compensation in the electrical domain instead of in the optical domain is preferred, because of the cost of dispersion compensation fibers and flexibility of dispersion compensation using electrical signal processing techniques. These compensation methods form the basis of this thesis.

1.3.3 Kerr Nonlinearity

The Kerr effect is a nonlinear process that occurs when intense light propagates in a fiber. This high intensity modifies the refractive index. The power-dependent refractive index change leads to an effect called self-phase modulation (SPM). If there are different overlapping pulses in different frequencies or wavelength channels, then it can also lead to cross-phase modulation (XPM) where one wavelength channel modifies the propagation characteristics of other wavelength channels. In equation (1.1), if β_2 and α are set to zero, then the remaining term is generated by the Kerr nonlinearity. For this case, the received signal is

$$E(L,T) = E(0,T)e^{jL\gamma|E(0,T)|^2}. \tag{1.3}$$

As can be seen, the phase of the signal is distorted by a power-dependent phase that depends on the signal. This distortion causes the signal to spread in the frequency domain. In a single user channel, this effect is called self-phase modulation. The nonlinear distortion is a phase-only distortion if chromatic dispersion is negligible. When linear dispersion is not negligible, this leads to a coupled phase and amplitude distortion, that severely affects the system performance.

If more than one data stream is sent at different carrier frequencies that correspond to different channels, then the phase shift in (1.3) depends on both the power of the target pulse and the power of the neighboring wavelength channels.

In the presence of chromatic dispersion, there is no closed form solution for (1.1). Insight into the effect of the nonlinearity can be gained by determining the form of the interference term. Assume there are pulses E_n, for $n = 1, ..., N$, transmitted at frequencies f_n with powers P_n, respectively. The power of the interference term at channel i is proportional to

$$P_{interference,i} \propto \sum_{l,m,k} \eta_{l,k,m}\gamma^2 P_l P_k P_m \tag{1.4}$$

where, $f_i = f_l - f_k + f_m$, the parameter η is a function of the fiber loss, the dispersion, the propagation length, and frequency spacing between the channels. In this equation, if $l = k$ and $i = m$ then the effect is cross-phase modulation (XPM). If all indices are equal to i, then the effect is self-phase modulation (SPM). The remaining terms are called four-wave mixing (FWM). This effect is a inter-channel cross-talk effect. When the data transmission rate is more than 40 Gbps, the pulses in a single frequency channel can produce interference terms via four-wave mixing. This effect is called intra-channel interference.

In the presence of chromatic dispersion, pulses in different channels have different speeds. Therefore, these pulses walk off from each other, and as a result, the nonlinear interference from neighboring channels has memory. Most multi-channel WDM systems have tens of channels. In these systems, if the relative walk-off of the channels is large because of the channel-dependent group velocity, then the four-wave mixing terms are suppressed since these terms require three different channels to interact. For these systems, cross-phase modulation for which the two of the interacting channels are the same, is the dominant source of signal degradation. This effect causes distortion in both the phase and the amplitude.

1.3.4 Polarization Mode Dispersion (PMD)

Polarization mode dispersion (PMD) occurs because the propagation characteristics of fiber are polarization dependent. This means that the two orthogonal polarization states in the fiber can travel at slightly different speeds due to optical birefringence in the fiber. This effect causes random temporal spreading of the signal, which leads to inter-symbol interference. There are imperfections in a realistic fiber like elliptical cross sections, microbends, or microtwists that break the circular symmetry. In this case, because of birefringence in the fiber, the two polarizations travel at different group velocities, and the two polarization components of the signal will separate slowly. Birefringence changes randomly along fiber, which causes random coupling of the two polarizations. The pulse spreading effect can be modeled as a random walk. The mean polarization dependent time delay is called the differential group delay (DGD).

The nonlinear Schrödinger equation given in (1.1) is for a single polarization, single-user system. For a polarization-multiplexed system, we must use a coupled nonlinear Schrödinger equation [7]. Consider a system that has only polarization mode dispersion. For $\mathbf{E} = (E_x, E_y)$, E_x and E_y are the lowpass or complex baseband representation of the modulated signal in the parallel and perpendicular polarizations. This representation is called a Jones vector. Let $\mathbf{R} = (R_x, R_y)$ be the Jones vector of the output of the fiber. The input and the output of the channel are related by a linear transformation

$$\begin{aligned} \begin{bmatrix} R_x \\ R_y \end{bmatrix} &= l \begin{bmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{bmatrix} \begin{bmatrix} S_x \\ S_y \end{bmatrix} \\ &= lJ \begin{bmatrix} S_x \\ S_y \end{bmatrix}, \end{aligned} \tag{1.5}$$

where l is a real scalar describing the optical loss from the input to the output, and the polarization change due to fiber is described by a unitary Jones matrix J. In a fiber with polarization mode dispersion, the Jones matrix changes randomly in time.

Compensation of polarization mode dispersion can be accomplished in several ways. One method uses a polarization controller to split the output of the fiber into two principal polarizations. This method applies a delay to one output relative to the other to bring them back into alignment. These systems are expensive and complex. Another approach is to to use a polarization maintaining fiber (PM fiber). This type of fiber typically has a highly elliptical core. The symmetry in these fibers causes an input polarization along a principal axis to maintain its polarization at the output. Existing PM fibers have higher loss and cost relative to ordinary fibers. An extension of this idea is a single-polarization fiber in which only a single polarization state is allowed to propagate along the fiber.

A final method is to use an advanced modulation scheme with a reduced symbol rate relative to information rate. This reduced symbol rate is less sensitive to polarization mode dispersion. In this case, adaptive filters retrieve all the information carried by x and y polarization components of signal using two-input-two-output adaptive filter (electronic polarization demultiplexer), increasing the number of adaptive filter coefficients, also incorporates the equalization of intersymbol interference caused by polarization mode dispersion and residual chromatic dispersion.

1.4 Thesis Motivation and Organization

The goal of this thesis is to apply modern detection techniques to the electrically received signal to mitigate a combination of the linear and nonlinear impairments outlined in the earlier sections. Advanced electrical coded-modulation techniques can reduce the cost and increase the performance of next generation optical communications systems. This can be accomplished in several ways: (1) increase the channel capacity, (2) increase the length of fiber between consecutive relay stations, i.e. reduce the number of relay stations, and (3) decrease the cost of each relay station.

The research presented in this thesis addresses aspects of each these approaches. We develop techniques to mitigate the effects of the three signal impairment mechanisms using different channel models that represent both current systems as well as future systems.

Noncoherent receivers based on intensity modulations are discussed in the first three chapters. These systems are currently deployed. Coherent receivers are discussed in the Chapters 5 and 6.

To have a unified framework, we define a simplified discrete mathematical channel model that can be used for both types of systems. In this model, S is the discrete transmitted optical signal with a power P defined as $P = E\{|S|^2\}$ where $E\{.\}$ is the expected value. The term R is the received electrical signal. The optical noise N_0 is modeled as additive white gaussian noise mostly produced by erbium doped fiber amplifiers. The discrete model of chromatic dispersion is a linear filter D. Nonlinear

Table 1.1: Channel models used in this thesis. QPSK stands for quadrature phase shift keying. PolMux stands for polarization multiplexing.

Chapter	Modulation	Channel Model
2	Intensity	$R = \|S * D + N_0\|^2$
3	Intensity	$R_i = \|S_i + I_i * D^{-1} + N_0\|^2,\ i = 1...K$
4	Intensity	$R_i = \|S_i * D * D_{partial}^{-1} + I_i * D_{partial}^{-1} + N_0\|^2,\ i = 1...K$
5	PolMux-QPSK	$R_x = (S_x * J_{xx} + S_y * J_{xy}) * D + I_x + N_{0,x},$ $R_y = (S_x * J_{yx} + S_y * J_{yy}) * D + I_y + N_{0,y}$
6	QPSK	$R = S * D + I + N_0$

effects are modeled using I for the combined effect of self-phase modulation, cross-phase modulation, and four-wave mixing terms. Polarization effects are modeled using a Jones matrix

$$J = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{bmatrix}.$$

Table 1.1, lists the channel models used in this thesis.

1.4.1 Maximum-Likelihood Detection and Constrained Coding on a Single-User Channel

A method to enhance the performance of a communication channel is by matching the transmitted signal to the channel using coding and/or modulation. These types of codes are called modulation codes. A constrained code is a modulation code based on a finite state machine, commonly represented by a state diagram or trellis [8]. Constrained codes restrict transitions in a state diagram. An example of a constrained code is a code that limits the number of consecutive zero bits.

In Chapter 2, we consider an intensity modulation direct detection (IMDD) system. The detector is a photodiode, which is a square-law detector. The output of the noncoherent receiver has nonlinear intersymbol interference (ISI) caused by a combination of chromatic dispersion and square-law detection. The simplified discrete model of the received signal for this channel model is

$$R = |S * D + N_0|^2, \tag{1.6}$$

where $*$ is the convolution operator. In Chapter 2, we investigate using electrical compensation to mitigate the nonlinear intersymbol interference, instead of using more expensive and less flexible dispersion compensation fiber (DCF) as is often done in current systems. In addition, we design modulation codes to decrease the nonlinear intersymbol interference in order to increase the channel capacity.

A variety of nonlinear electrical compensation techniques in the electrical domain have been studied. We focus on maximum-likelihood sequence estimation (MLSE), which is the optimal detection and equalization method for minimizing the word error rate [9, 10]. One major contribution of this thesis is combination of this technique with a modulation code. This combined method is described in Chapter 2. In Chapters 5 and 6 coherent systems are analyzed using the same techniques.

1.4.2 Modulation Codes to Reduce Nonlinear Cross-Talk in a Dense Wavelength Division Multiplexing Channel

The nonlinearity in intensity-modulated systems is a result of the square-law detection process. Another form of nonlinearity occurs in the optical domain. In Chapter 3, we study modulation codes for current systems where both detection and fiber channel are nonlinear.

One method to reduce the cost of the optical communications system is to increase the length of fiber between relay stations. At each relay station, an amplifier, usually an erbium-doped fiber amplifier (EDFA), increases the signal power to compensate the attenuation of the signal. Optimum performance that produces a high output signal-to-noise ratio is achieved when the input signal power of the EDFA is above a minimum level. Therefore, to increase the length of a single span between each relay station, the launch power of signal at each span must increase. As described in Section 1.3.3, in dense wavelength division multiplexing (DWDM) systems, the power of interchannel interference is caused by self-phase modulation, cross-phase modulation, and four-wave mixing. This interference is proportional to the cube of the signal power, i.e., P^3 [7]. If the wavelength division multiplexing channel has K channels and assuming the dispersion is compensated by a dispersion compensation fiber, then the simplified discrete model of this fiber channel for channels $i = 1 : K$ can be written as

$$R_i = |S_i + I_i * D^{-1} + N_0|^2, \tag{1.7}$$

where $E\{|I^2|\} \propto P^3$ and D^{-1} is the inverse function of dispersion. The signal to interference plus noise for this system is $SINR = \frac{P}{N_0+\alpha P^3}$, where α is a scaler. When αP^3 is negligible with respect to N_0, then as the power P increases, the $SINR$ increases. However, when αP^3 is greater than N_0, then increasing the power P decreases the $SINR$ with a rate of P^2. The spectral efficiency of the communications channel has a direct relationship to the signal to interference plus noise ratio [11]. Therefore, increasing the launch power of the fiber beyond a certain point where the interference dominates the noise will reduce the spectral efficiency of the channel [12].

Applying a constrained code is one method to reduce the power of the nonlinear interference. Research shows that the constrained codes which avoid intra-channel four-wave mixing interference have capacity zero [13, 14, 15]. Since the inter- and intra-channel four-wave mixing have a similar structure, these results are applicable to inter-channel four-wave mixing interference. This means that there is not a constrained code that can completely eliminate four-wave mixing. In Chapter 3 we examine a method to reduce interference power using a constrained code recognizing that the elimination of nonlinear interference is not possible. If the level of interference is reduced, then the signal to interference and noise ration (SINR) increases and the capacity of each channel will increase.

1.4.3 Multiuser Receiver to Combat Nonlinear Cross-Talk

All of the techniques considered so far rely on the information in a single channel. If there is only a single channel and self-phase modulation is the dominant effect, then single-user maximum-likelihood sequence estimation shows great improvement in performance with respect to threshold detector [16]. In [17], the authors showed that if the nonlinear interference is week, then using the information of all of the channels in the detection process can improve the capacity of a wavelength division multiplexing system. In the ideal case, this capacity is equal to the capacity of a channel without interference. The idea of using this type of multiuser receiver was also suggested in [18]. It should be noted that the optimal multiuser receiver which detects all of the channels at the same time has a high complexity and it is not currently practical. Besides, in many cases, detecting the information of all of the channels in one receiver is not desirable.

In Chapter 4, we consider the channel model for $i = 1 : K$ is

$$R_i = |S_i * D * D_{partial}^{-1} + I_i * D_{partial}^{-1} + N_0|^2. \tag{1.8}$$

where $D_{partial}^{-1}$ is a partial-dispersion-compensation filter. We investigate the performance of multiuser maximum-likelihood sequence estimation on different systems with different chromatic dispersion parameters and levels of nonlinearity. We use partially dispersion compensation method to gain optimal performance. We also study the effect of laser noise on the system performance.

1.4.4 Maximum-Likelihood Equalization for Polarization Multiplexed Quadrature Phase Shift Keying Modulated Channel

The nonlinear distortion of an optical signal in a nonlinear regime is proportional to the power of the signal. One method to reduce this nonlinear distortion is to apply constant amplitude phase-shift keying (PSK) modulation instead of intensity modulation

[19].

The techniques discussed in Chapters 2-4 are based on the noncoherent modulation and detection. In these channels, linear impairments in optical domain like chromatic dispersion (CD) and polarization mode dispersion (PMD) are nonlinear after the signal is squared-law sensed. The spectral efficiency for a variety of detection and modulation methods in both the linear [20, 21, 22] and nonlinear regimes [12, 23] have been studied. Modulation formats based on noncoherent detection or differential modulation have a limited number of degrees of freedom available for the encoding of information [23]. These formats provide good power efficiency at a low spectral efficiency. Coherent detection based on modulation that uses both amplitude and phase provides additional degrees of freedom. This makes the system more robust to nonlinear effects. In addition, using polarization multiplexing (PolMux) modulation increases the spectral efficiency [24]. In Chapter 5, we consider a single-user, quadrature phase and polarization multiplexed modulation channel with a coherent receiver. The simplified channel model for this system may be written as

$$
\begin{aligned}
R_x &= (S_x * J_{xx} + S_y * J_{xy}) * D + I_x + N_{0,x}, \\
R_y &= (S_x * J_{yx} + S_y * J_{yy}) * D + I_y + N_{0,y},
\end{aligned} \tag{1.9}
$$

where the input and output components as well as the elements of the Jones matrix are defined in (1.5). Compensating for the impairments of coherent polarization multiplexing channel at the receiver in the digital signal processing (DSP) block is a challenge, especially when the nonlinearity is high. In a basic digital signal processing block at the receiver, the chromatic dispersion, the Kerr-nonlinearity, the polarization dependent effects, and carrier phase noise should all be compensated. In Chapter 5, we use maximum-likelihood sequence estimation to compensate for polarization division multiplexing as well as the nonlinearity.

1.4.5 Polynomial Fitted Equalizers for Nonlinear QPSK Modulated Channel

Low-complexity, high-performance compensation of nonlinear impairments is a fundamental research topic which has been addressed in the last decade. Backpropagation (BP) is a successful, but complex method that can compensate for the impairments in the coherent channel [25, 26, 27]. Backpropagation is based on the inverse of the standard numerical solution to the nonlinear Schrödinger equation. The complexity of the algorithm makes backpropagation currently impractical. Methods to reduce the complexity and the number of computation steps are essential for practical systems. In Chapter 6, we develop simplified practical techniques based on the following simplified channel model

$$R = S * D + I + N_0. \tag{1.10}$$

We present a novel compensation method based on backpropagation that has improved performance. This is achieved by fitting polynomials for both the linear and a nonlinear part of the compensation. We show an improvement in the output signal-to-noise ratio using a sampling rate of one and two samples per symbol.

1.4.6 Conclusion and Future Directions

In Chapter 7, we conclude our thesis, and describe some of the open problems identified by the research in this thesis.

2 Maximum-Likelihood Detection and Constrained Coding on Optical Channels

2.1 Introduction

In an optical communication system that uses a square-law detector, the combination of the square-law characteristic and intersymbol interference (ISI) produces a nonlinear channel model for the output electrical signal [28]. Recently, electrical-domain equalization methods have been investigated as an alternative to all-optical dispersion compensation, e.g., dispersion-compensating fibers (DCF). Several compensation methods have been proposed for systems using both the phase and amplitude of the signal, e.g., [29, 30, 31, 32, 33, 34]. For systems that do not use phase such as a typical noncoherent non-return-to-zero (NRZ) systems, effective compensation methods are more difficult to implement. It can be proved that no linear electrical equalizer can completely compensate this nonlinear channel even if the channel is noiseless [35, 36]. Therefore, the performance improvement achieved by electrical equalization techniques for nonlinear optical channels that do not use phase is inferior to equalization that does use phase such as all-optical compensation.

An effective solution to control the intersymbol interference is to use maximum-likelihood sequence estimation (MLSE), which is the optimal detection and equalization method for minimizing the word error rate (WER) [35]. It appears that maximum-likelihood sequence estimation was first suggested to compensate chromatic dispersion in optical fibers in [37] and [38]. The exact signal statistics for a long-haul link containing a single optical pre-amplifier, where the noise is added in the optical domain, was calculated and applied to maximum likelihood sequence estimation for electronic dispersion compensation in [9]. An approximation of the performance of such a system is analyzed in [39]. In [40], high complexity Viterbi algorithms are used to reach 1,040 km on a standard single-mode fiber at 10 Gbit/s.

In addition to improving detection, the overall system can be enhanced by matching the transmitted signal to the channel[1] using coding and/or modulation. These types of codes are called “modulation codes” or “line codes”. Usually, a modulation code is applied in conjunction with an error correction code (ECC), so that the coded data after an ECC is coded again by the modulation code block. Figure 2.1 shows a block diagram of a long-haul optical system using both ECC and a modulation code.

Modulation codes are widely implemented in magnetic and optical disc record-

[1] In this thesis, the term channel incorporates the transmitter, the fiber, and the optical sensor.

ing and most optical communication systems [41, 42, 43, 44]. While the ultimate goal of all types of modulation codes is to reduce the bit error rate, different codes use different design criteria to achieve this goal [8, 13, 14, 15].

Increasing the “distance” between the received points in the signal space is a common design criterion. In the absence of noise, the signal vectors corresponding to a sampled signal sequence form a constellation space. Given the channel statistics, a metric can be defined between each pair of signal vectors and it is commonly referred to as a distance[2]. Errors occur predominantly between vectors near in distance [35]. A distance-enhancing modulation code chooses and removes a subset of signal vectors to increase the minimum distance.

A constrained code is a modulation code based on a finite state machine, commonly represented by a state diagram or trellis [8]. Constrained codes restrict transitions in a state diagram. In [45] a constrained code was designed for an optical fiber channel to remove the isolated ”one” pattern that caused significant errors in a system using a threshold detector. In this Chapter, we extend this result and present a general constrained code design process for a noncoherent non-return-to-zero optical channel. In this method, we first characterize the most probable error events and then based on this analysis, we design distance enhancing constrained codes.

Distance enhancing constrained codes increase the distance by specifying a “forbidden list” of code strings whose omission ensures small distance error events do not occur [46, 47]. The technique of generating a constrained code using a “forbidden list”

[2]The metric only corresponds to a Euclidean distance when the noise is a additive Gaussian.

can be summarized in three steps [48]. First, the set of most probable input error events, S, for the specified channel is determined. Second, a list of forbidden patterns, Q, is chosen such that preventing them from being transmitted and received reduces the number of error events in set S. Finally, an efficient practical encoder and decoder for this constraint are constructed to avoid the set of patterns in Q. For this system, the sequence detector in the receiver is designed to incorporate the channel and code constraints. Therefore, the constraint can reduce the number of sequence detector states relative to the uncoded version.

The rest of the chapter is organized as follows. In the next section, a mathematical model for the long-haul optical channel is defined and maximum-likelihood sequence estimation at the receiver is characterized without coding. In the section 2.3, the design of distance-enhancing constrained codes based on a forbidden list of patterns for the nonlinear optical channel is described. In section 2.4, the performance of several codes is investigated with respect to key system parameters including effect of sampling shift and the number of states of the MLSE. Finally, the chapter concludes with prospects for application of these codes.

2.2 Optimal Detection for Optical Channel

In this section, we introduce the mathematical model of the optical channel, the transmitter, and the receiver.

2.2.1 Channel Model

The channel used is a noncoherent long-haul optical channel with a square-law detector. The transmitted signal is non-return-to-zero with a finite extinction ratio, ER, defined as $20\log_{10}(A_{max}/A_{min})$, where A_{max} is the maximum magnitude of an isolated mark (one) and A_{min} is the minimum magnitude of an isolated space (zero). For long-haul channels with optical amplifiers, amplified spontaneous emission (ASE) noise caused by amplifiers is the dominant noise. The received signal $r(t)$ is modeled as

$$r(t) = |[s_t(t) \star \ell(t) + n_o(t)] \star h(t)|^2 \star h_e(t),$$

where $\star$ is the convolution operator, $s_t(t)$ is the transmitted signal, $\ell(t)$ is the impulse response of the fiber channel, $n_o(t)$ is the complex additive white Gaussian optical amplifier noise (AWGN), $h(t)$ is the optical filter before detector, and $h_e(t)$ is the electrical filter. A block diagram of the channel model is shown in Figure 2.1.

The transmitted non-return-to-zero signal is

$$s_t(t) = max[A_{min}, min\{A_{max}, \mathbf{d} \star p_t(t)\}], \tag{2.1}$$

where $\mathbf{d}$ is the discrete transmitted bit sequence and $p_t(t)$ is the impulse response of the transmitter. A consequence of using non-return-to-zero signals is that, in general, the relationship between the transmitted sequence $\mathbf{d}$ and the transmitted waveform $s_t(t)$ is not linear even with respect to the optical field. A typical transmitted eye diagram is shown in Figure 2.2(a).

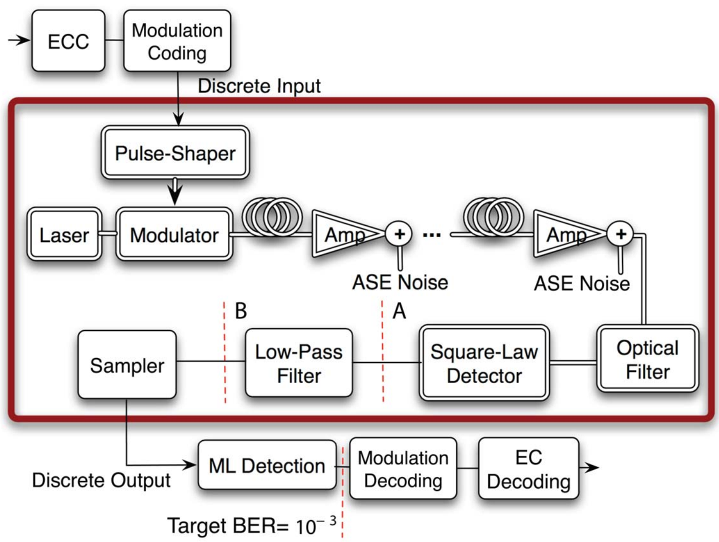
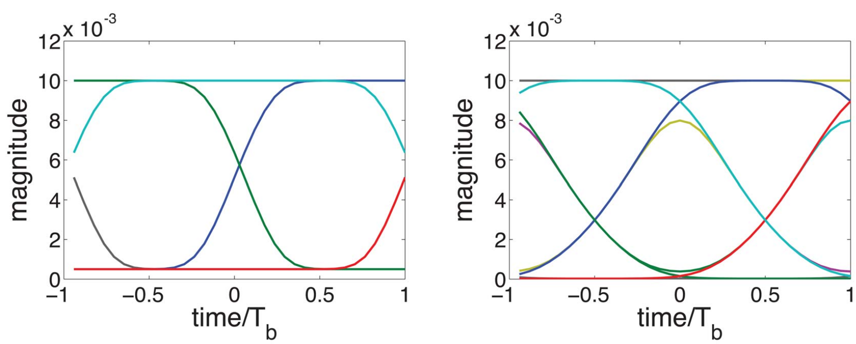


Figure 2.1: Long-haul optical channel including the transmitter and the receiver.



(a) (b)

Figure 2.2: Eye diagrams of a back-to-back system with $\xi = 0$: a) before the electrical filter, b) after the electrical filter.

The impulse response of the fiber channel, $\ell(t)$, which is the source of the intersymbol interference, is proportional to $\exp\left(jt^2/(2\beta_2 L)\right)$, where L is the propagation length and β_2 is the dispersion coefficient [7]. We now define the normalized dispersion index (NDI) as

$$\xi = -2\beta_2 L R_b^2.$$

The impulse response of the channel, $\ell(t)$, can then be written as

$$\ell(t) = \frac{e^{j\pi/4}\text{sign}(\xi)}{T_b\sqrt{\pi|\xi|}} \exp\left[-\frac{j}{\xi}\left(\frac{t}{T_b}\right)^2\right],$$

[49], where $R_b = 1/T_b$ is the symbol or baud rate with typical units of Gsym/s (10^9 symbols/second). The normalized dispersion index ξ is dimensionless and characterizes the intersymbol interference. The quadratic dependence of ξ on the symbol rate R_b implies that at a fixed fiber distance the intersymbol interference increases rapidly with respect to the symbol rate.

Similar to other communication channels, the memory length of a channel is defined as the number of neighboring samples affected by the interference. We define the memory length as the width of the infinite impulse response that contains 95% of the energy of the transmitted signal. The memory length is then a function of ξ. Table 2.1 provides a relationship between ξ, the memory length, and the corresponding length of a standard fiber for a given symbol rate.

In practical systems, $h_e(t)$ is usually chosen to be a lowpass filter. Here, we

Table 2.1: Relationship between ξ, the memory length, and the corresponding length of the fiber for a channel with the group velocity dispersion $D = 17$ ps/(nm.km) and symbol rate $R_b = 10.7$ Gsym/s.

ξ	0.5	0.85	1	1.5	2
Memory Length	2	3	4	5	6
L (km)	103	175	206	308	411

assume that it is a rectangular filter given by

$$h_e(t) = \begin{cases} 1 & -T_b < t < 0 \\ 0 & \text{Otherwise} \end{cases} .$$

The sensed electrical signal is convolved with this filter function and then sampled to produce

$$\begin{aligned} r_k &= r\left((k+\Delta)T_b\right) \\ &= \int_{(k+\Delta)T_b}^{(k+\Delta+1)T_b} \left|[s_t(t) \star \ell(t) + n_o(t)] \star h(t)\right|^2 dt, \end{aligned} \tag{2.2}$$

where k is an integer that represents the detected symbol and $-1 < \Delta < 1$ is the shift of sampling point with respect to the start of the symbol. This normalized range corresponds to the width of the convolution of the sensed electrical signal of an approximate width T_b with the rectangular filter function that also has a width T_b.

To produce a realistic channel model, we extracted the transmitted pulse shape, $(p_t(t))$, and optical filter, $(h(t))$, from the experimental setup in [1]. The pulse shape $p_t(t)$ is assumed to be raised cosine, and $h(t)$ is a Gaussian filter with time-bandwidth product $B_{3\,dB}.T_b = 3$, where $B_{3\,dB}$ is the 3-dB bandwidth of the filter. Figure 2.2 shows the simulated back-to-back eye diagram of the signal $r(t)$ for this system. The eye drawn in Figure 2.2(a) corresponds to what would be seen on an oscilloscope for a typical system without the presence of the electrical filter, i.e., point **A** in the Figure 2.1. The eye in Figure 2.2(b) is the eye after the electrical filter which is used in the detection process, i.e., point **B** in the Figure 2.1. The sampling point plays a vital role in the performance of this system [50]. The dependence of the performance of a system on the sampling shift, Δ, is investigated in section 2.4.

2.2.2 Optimal Detection: Maximum Likelihood Sequence Estimation

For a linear intersymbol interference channel, the Viterbi algorithm (VA) is the optimal maximum likelihood sequence estimation (MLSE) receiver when the noise is memoryless and the number of trellis states is large enough to span the memory. The optimal branch metrics of the corresponding trellis are calculated based on the statistics of the channel which is the negative logarithm of the conditional probability distribution function (pdf) of the sampled output of the channel, r_k, given a known transmission sequence, $\mathbf{d}$ [51],

$$-\log f_{r_k|\mathbf{d}}(r). \tag{2.3}$$

For our channel model with an optical noise-reducing filter before the sensor and a rectangular electrical filter at the receiver, (2.3) does not have a simple closed form. In [9, 52], and the references therein, the channel is modeled as the sum of parallel chi square channels.

The complexity of the Viterbi algorithm grows exponentially in the memory length and thus for long-haul channels with long memory lengths, it is prohibitively complex for high data rate transmission. A practical approach is to use a "reduced-state" Viterbi algorithm with the memory length of the corresponding trellis set to be less than the memory length of the channel. In contrast to a "full-complexity" Viterbi algorithm, the noiseless outputs for this reduced-state trellis corresponding to each outgoing branch from a state are not identical. The optimal detector for this reduced-state trellis uses branch metrics that are the logarithm of the average of the pdf over all the possible noiseless outputs [53]. An example of the implementation of a reduced-state Viterbi algorithm is presented in [40]. We implement our reduced state Viterbi algorithm including coding using 8 or less states. This choice is based on the fact that 4-state systems are realizable by state-of-the-art chips [54], and 8-state systems represent practical next generation systems.

Figure 2.3 shows the result of the simulated channel model described in this section using a 3-bit analog to digital converter (ADC) and a 4-state Viterbi algorithm. The curves are drawn for two choices of electrical filters: a system without electrical filter, and a system with a rectangular filter. The plots are compared to the results presented in [1]. The simulation results are derived by counting at least 1000 errors at the output. In these plots, it can be seen that the outputs of the simulation are typically within 2 dB of the experimental data[3]. The validation of the channel model to the experimental data

[3]The results presented in [1] were derived using a 0.2 nm OSA filter, while the paper stated that it used 0.1 nm filter[55]. Therefore, the curves taken from [1] are shifted by 3 dB to account for this difference.

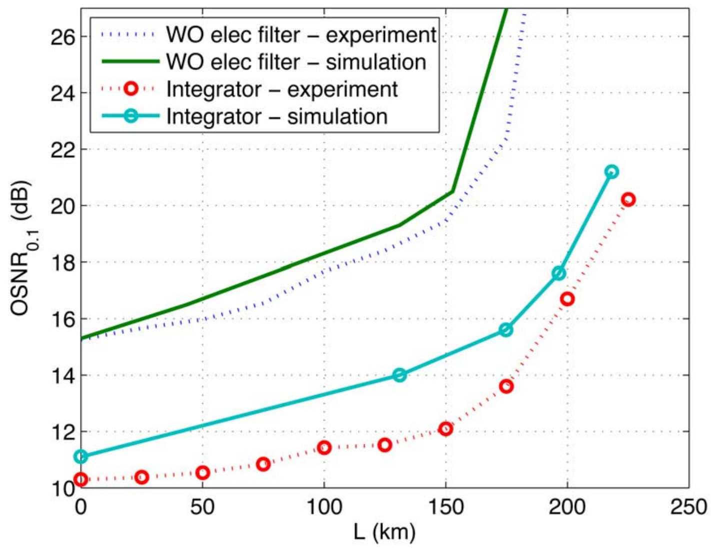


Figure 2.3: Comparison of simulation results based on the channel model in (2.2) and the experimental data presented at [1], using 4-state Viterbi algorithm and 3-bit ADC.

with no free parameters to within 2 dB gives us confidence that the simulated results of the effect of coding will be accurate. The results presented in this chapter are based on this model and an 8-bit ADC to isolate the effect of coding from the effect of quantizer. The effect of the output rate of quantizers on the performance of a system using MLSE at the receiver has been analyzed in [10].

2.3 Distance Enhancing Constrained Codes

In this section, we characterize error events caused at the output of the Viterbi algorithm, analyze them, and based on the dominant error patterns, we design constrained modulation codes. As described in Section 2.1, in order to design these codes, the most probable errors that occur when using a MLSE receiver must be identified. This process is called error event characterization.

In many linear communication channels, a concatenation of a partial-response (PR) equalizer and a sequence detector is used in the receiver. The first block equalizes the channel to a known channel which is not necessarily a channel with all the intersymbol interference removed, and thus is called a PR equalizer. The second block is then designed based on the known statistics of the equalized channel. Most of the analysis and code designs in the literature are based on known PR channel models. The main purpose of the equalization process is to have a deterministic output regardless of changes within the channel. The MLSE process can then be optimized for this deterministic channel. However, for a nonlinear channel this is not practical because no known equalizer can completely equalize the channel to a known channel model.[4] Therefore, we investigate the properties of MLSE on unequalized channels. Forms of a nonlinear equalizer will be presented in later chapters.

2.3.1 Error Characterization

For linear AWGN channels, there is an algorithm to characterize the error events, i.e., determining the pairs of input sequences resulting in a specific distance at the channel output [56]. This algorithm works in cases where the probability of an error sequence is a function of the absolute difference of the transmitted, $\mathbf{d}$, and detected, $\mathbf{d}'$, sequences, i.e., $|\mathbf{d} - \mathbf{d}'|$. For this class of channels, the Euclidean distance of the output sequences

[4] An equalizer can be used for a second purpose which is concentrating the energy of the signal spread by intersymbol interference to reduce the number of states for sequence estimation. In this case, the output of the equalizer is close to a known channel with a short memory length and the MLSE must tolerate some mismatch.

is a suitable metric for determining the most probable error events and is a function of only the difference $|\mathbf{d} - \mathbf{d}'|$. However, for the nonlinear optical channel model, the distance metric defined in (2.3) does not have a closed-form expression and is dependent on the ordered pair $(\mathbf{d}, \mathbf{d}')$ and not the difference $|\mathbf{d} - \mathbf{d}'|$. Specifically, the probability of sending $\mathbf{d}$ and detecting $\mathbf{d}'$ is not equal to the reverse occurrence. Therefore, the standard algorithm that is applicable to linear channels cannot be applied to our channel. The most straightforward approach for finding dominant error events is then a full scale numerical search.

Typical results of a numerical search for the most probable error events for the channel model in (2.2) are shown in Table 2.2 for $\xi = 0.85$ at bit error rate of 10^{-3}. The choice of bit error rate to be 10^{-3} is consistent with the the ECC threshold for fiber optic communication systems according to the ITU guidelines [57]. In this table, each row represents two sequences of bits, their difference, and the probability of sending one sequence and detecting the other. In the column showing the difference "n" represents no change, "+" represents transmitting a zero and detecting a 1, and "-" represents transmitting a 1 and detecting a zero. Examining the rows in the table, the probability of error events with equal differences are not necessarily equal. For example, there are 32 error events with the difference sequences $nn+nn$ or $nn-nn$. Among these, the error event in Row 2 $(\mathbf{d} \rightarrow \mathbf{d}') = (01010 \rightarrow 01110)$ has the probability of 4.3×10^{-5} $(= 10^{-4.37})$, while 22 of 32 have probabilities less than 10^{-5}. The other 10 error events are shown in bold in Table 2.2. In this table, the error events with lengths greater than 11 or with the bit error rate less than 10^{-5} are not listed. The sum of all these unlisted error events yields a bit error rate less than $4 \times 10^{-4} (= 10^{-3.4})$.

As described in section 2.2.1, the memory length of the channel is monotonically increasing with ξ. Therefore, the behavior of the error events changes as ξ changes, which occurs if either the fiber length or data rate change. In Table 2.3, probable error events are listed for four values of ξ. To avoid long lists, the error events are grouped based on their difference. Therefore, all error events with differences $\pm(\delta)$ are grouped under one error event class δ, since they contain same pairs of sequences. For example, all error events with the differences $\delta = nn + - + nn$ or $-\delta = -(nn + - + nn) = nn - + - nn$ are grouped under one error event class $\delta = nn + - + nn$. We again note that because the channel is nonlinear, all errors in one class are not equiprobable. The error probability for each class in this table is the total bit error probability caused by the corresponding error events. The classes which have a total probability of less than 10^{-5} are also not listed. The grouping of error events into classes implies that some classes listed in Table 2.3 are not present in Table 2.2, because the bit error rate of the total class is greater than 10^{-5}, while none of the individual error events within the class is greater than 10^{-5}.

For small values of ξ, shorter error events have a higher probability of occurrence. For these cases, the general error pattern is alternating plus and minus signs. This indicates that most of the sequences causing errors contain at least one of the patterns 101 or 010 corresponding to an isolated mark or space. These patterns are similar to the error patterns seen in the linear $(1 + D)$ PR channel [56].

Table 2.2: List of dominant error events generated by numerical search for $\xi = 0.85$, OSNR=13.9, Δ=0.5, and bit error rate=10^{-3}. The symbol "n" represents no change, "+" represents transmitting a zero and detecting a 1, and "-" represents transmitting a 1 and detecting a zero. $P = -log_{10}$(Bit Error Rate of the Error Event).

	$\mathbf{d}$	$\mathbf{d}'$	Difference	P
1	100100	101000	$nn + -nn$	4.34
2	01010	01110	$nn + nn$	4.37
3	1001001	1010101	$nn + - + nn$	4.38
4	01111	01011	$nn - nn$	4.45
5	11010	11110	$nn + nn$	4.50
6	11111	11011	$nn - nn$	4.58
7	1001010	1010110	$nn + - + nn$	4.63
8	01001	01101	$nn + nn$	4.66
9	100101001	101010101	$nn + - + - + nn$	4.66
10	10010	10110	$nn + nn$	4.68
11	100111	101011	$nn + -nn$	4.69
12	000100	001000	$nn + -nn$	4.69
13	01100	01000	$nn - nn$	4.71

Continued on next page

Table 2.2: Continued from previous page.

	$\mathbf{d}$	$\mathbf{d}'$	Difference	P
14	0001001	0010101	$nn + - + nn$	4.72
15	0101001	0110101	$nn + - + nn$	4.75
16	100101	101001	$nn + -nn$	4.76
17	1101001	1110101	$nn + - + nn$	4.76
18	01110	01010	$nn - nn$	4.82
19	10010100	10101000	$nn + - + -nn$	4.83
20	101001	100101	$nn - +nn$	4.85
21	11001	11101	$nn + nn$	4.91
22	110100	111000	$nn + -nn$	4.93
23	110111	111011	$nn + -nn$	4.95
24	11110	11010	$nn - nn$	4.97
25	1101010	1110110	$nn + - + nn$	4.96
	P> 5 or error length>11		Tot. Prob. $< 10^{-3.4}$	

For higher values of ξ, the next most common patterns, besides alternating plus and minus are $-+n-$, $+-n+$, $-n+-$, or $+n-+$. These patterns correspond to sequences containing the patterns 1001 or 0110 which are isolated double-mark or double-space.

2.3.2 Forbidden List of Patterns

There is no unique set of forbidden patterns to reduce the bit error rate to a desired level. However, some constraints will limit the selection process. The main issue is the capacity of the constraint, $\mathcal{C}$, which is an upper limit for the rate of the code, ρ, designed for the constraint [20]. The code rate is the amount of information sent for each transmitted binary symbol. The capacity of a finite state constraint can be calculated by the corresponding state diagram [20, 8]. Assume that the state diagram of a constraint has N states. The adjacency matrix of this diagram is $A = [a_{ij}]$, $i, j = 1...N$, where a_{ij} is the number of edges (transitions) starting at state i and ending at state j. The capacity of the constraint is equal to the $\mathcal{C} = \log_2 \lambda$, where λ is the maximum eigenvalue of A.

Another concern in the selection of a forbidden list is the size of the corresponding state diagram. To avoid receiving forbidden patterns, the Viterbi algorithm at the receiver should use a trellis based on the state diagram of the constraint. Therefore, restrictions in the number of states are translated to restrictions on the number and length of the forbidden patterns. For example, for a receiver with memory length 2, i.e., a 4-state trellis, the receiver can not process a received signal sequence longer than 3

Table 2.3: List of dominant error events generated by numerical search for: $\xi = 0.43$, OSNR=12.8, Δ=0.875; $\xi = 0.85$, OSNR=13.9, Δ=0.5; $\xi = 1.49$, OSNR=16.8, Δ=0; $\xi = 1.76$, OSNR=28.2, Δ=0. The target bit error rate is 10^{-3} for all channels and $P = -log_{10}$(Bit Error Rate of the Error Event Class).

ξ		Error Event Class	P
0.43	1	$nn+nn$	3.19
	2	$nn+-nn$	3.75
	3	$nn+-+nn$	4.16
	4	$nn+-+-nn$	4.55
	5	$nn+-+-+nn$	4.95
	P>5 or error length>11		Tot. Prob. = $10^{-4.75}$
0.85	1	$nn+nn$	3.52
	2	$nn+-nn$	3.65
	3	$nn+-+nn$	3.75
	4	$nn+-+-nn$	4.06
	5	$nn+-+-+nn$	4.14
	6	$nn+-+-+-nn$	4.42
	7	$nn+-+-+-+nn$	4.54
	P>5 or error length>11		Tot. Prob. = $10^{-4.45}$

Continued on next page

Table 2.3: Continued from previous page.

ξ		Error Event Class	P
1.49	1	$nn+nn$	3.64
	2	$nn+-nn$	4.01
	3	$nn+-n+-nn$	4.07
	4	$nn+n-+nn$	4.17
	5	$nn+-n+nn$	4.26
	6	$nn+-n+-n+nn$	4.42
	7	$nn+-n+-n+-nn$	4.48
	8	$nn+n-+n-+nn$	4.54
	9	$nn+-+n-+nn$	4.6
	10	$nn+-n+n-+nn$	4.74
	11	$nn+-n+-+nn$	4.75
	12	$nn+-+nn$	4.91
	P>5 or error length>12		Tot. Prob. = $10^{-3.49}$
1.76	1	$nn+-nn$	3.07
	2	$nn+-n+-nn$	3.93
	3	$nn+-n-+nn$	4.65
	P>5 or error length>11		Tot. Prob. = $10^{-5.6}$

symbols. Therefore, for this trellis, the maximum length of forbidden patterns can not exceed 3.

We use four different forbidden lists, shown in Table 2.4, for the optical channel in (2.2). Constraints A, B, and C are based on a 4-state Viterbi algorithm and support a channel with memory length 2. These constraints have capacities of 0.69, 0.79, and 0.879, respectively. Constraint A forbids patterns 101 and 010 from the whole sequence, while constraint B forbids these patterns starting at even positions in the sequence. For constraint C, the pattern 101 starting at even positions and 010 starting at odd positions are forbidden. Constraint D with capacity 0.916 is designed for Viterbi algorithm trellises supporting a memory length of 3, i.e., using eight states, and forbids the patterns 1010 and 0101 at the even positions of the sequence.

2.3.3 Method of Coding

For a given constraint, there are different encoding methods that map the information sequences into coded sequences. The coding rate is upper-bounded by the capacity imposed by the constraint. For systems working at high rates, e.g. optical systems, simplicity is a critical issue in designing the encoders and decoders. Block codes using look-up tables for encoding and decoding are relatively simple codes. However, finding high rate block codes based on a constraint can be difficult. Bit stuffing [58, 59] is another method of code mapping, which produces a coding rate close to the capacity of the constraint. The main drawback of this method is that it produces variable-length coded sequences, which causes error propagation at the output of the decoder and some practical problems in the implementation of the code. Solutions exist to overcome these problems by producing fixed rate bit stuffing encoders [60], or changing the order of the inner and outer codes for the encoders and decoders, i.e., the modulation code and the error correction code [61, 62, 63]. However, the cost of all of these solutions is the coding rate. To remove these issues from the analysis, we restrict our results to the error rate before the decoder, as is standard practice [8]. We note that not including the decoder can bias the error rates. As an example, if the target is set to be 10^{-3} after the decoder, for code *B* the OSNR increases less than 1 dB. For codes that use bit-stuffing, the decoder causes insertion-deletion errors.

Each of the forbidden lists of patterns illustrated in Table 2.4 used a different coding method. For constraints *B* and *D,* simple block codes of rates 3/4 and 8/9, respectively, designed in [64, 65], were used. These codes assume a non-return-to-zero inverted (NRZI) format while the transmission in our channel model is non-return-to-zero. Therefore, after the encoder, the format of the transmitted signal changes. The non-return-to-zero format discrete data sequence $\mathbf{d}^{NRZ}$can be generated from the non-return-to-zero-inverted format sequence $\mathbf{d}^{NRZI}$ using the rule $d_{k+1}^{NRZ} = d_k^{NRZ} + d_k^{NRZI} \pmod 2$. Block codes *B* and *D* are designed such that, in addition to distance-enhancing characteristics, they forbid transmitted runs of ones or zeros longer than 7 and 12 for non-return-to-zero data for the codes *B* and *D*, respectively. For codes *A* and *C,* bit-stuffing is used. The encoders for bit-stuffing add one extra 0 after the pattern 10 where 101 is unwanted and an extra 1 after the pattern 01 where 010 is unwanted. The codes designed for constraints *A* and *C* have the average rates of 0.66 and 0.856. For all codes listed in

Table 2.4: Forbidden lists of patterns and the corresponding coding schemes used for optical channel. Notation - Superscripts e and o: even and odd positions in the sequence, respectively; BS: bit-stuffing; LT: look-up table.

Code Name	Memory Length	Forbidden Patterns	Coding Rule	Constraint Capacity $\mathcal{C}$	Code Rate ρ
A	2	101 010	BS	0.69	0.66
B	2	010^e 101^e	LT	0.792	0.75
C	2	101^e 010^o	BS	0.879	0.856
D	3	0101^e 1010^e	LT	0.916	0.89

Table 2.4 the probabilities of the bits 0 and 1 at the output of the decoder are equal given they are equiprobable at the input.

2.4 Results

The dominant error events depend on several system parameters including ξ, the sampling time Δ, and the number of states used in the Viterbi algorithm. In this section, we investigate the dependence of error events on these three parameters based on the channel model described at (2.2.1).

2.4.1 Normalized Dispersion Index, ξ

Figures 2.4 - 2.7 show the behavior of the performance of the Viterbi algorithm and codes with respect to the changes in the parameter ξ. The graphs show the required optical signal-to-noise ratio (OSNR) to reach a bit error rate of 10^{-3} for different values of ξ. Figure 2.4 is the performance of the uncoded system using either a 4-, 8-, 16-, or 64-state Viterbi algorithm. For uncoded systems, the transmitted symbols are information bits and the units Gbit/s and Gsym/s are synonymous. In this figure, the performance of the reduced-state Viterbi algorithms is compared to the full-complexity Viterbi algorithm. For $\xi < 2.2$, 95% of the energy of the transmitted symbol is spread within the 6 neighboring symbols. Therefore, the Viterbi algorithm with 64 state can be considered a full-complexity Viterbi algorithm. As it can be seen, the 8-state receiver does not perform significantly better than the 4-state, while the performance of 16- and 64-state Viterbi algorithm at values of ξ higher than 1.8 shows significant improvement. The difference in the performance can be explained by examining the memory length as a function of ξ relative to the memory span for the Viterbi algorithm. For $\xi = 1.8$, 87% of the energy is within the 4 neighboring symbol periods while only 77% is within the 3 neighboring symbols. The additional 10% of energy collected by the 16-state Viterbi algorithm causes a significant improvement because this energy is not collected by the 8-state system and thus acts as additional interference producing a degradation in performance. For comparison, when $\xi = 0.9$, there is less than 2 dB difference in performance between a 4-, 8-, 16-, and 64-state Viterbi algorithm. In this case, the percentage of the collected energy in the neighboring 2, 3, 4, and 6 symbols are 86%, 94%, 98%, and greater than 99% and all four systems operate with approximately the same performance.

For each channel model with given parameters, there is an optimum sampling position, Δ. In section 2.4.3, it is shown that the optimum sampling positions vary with

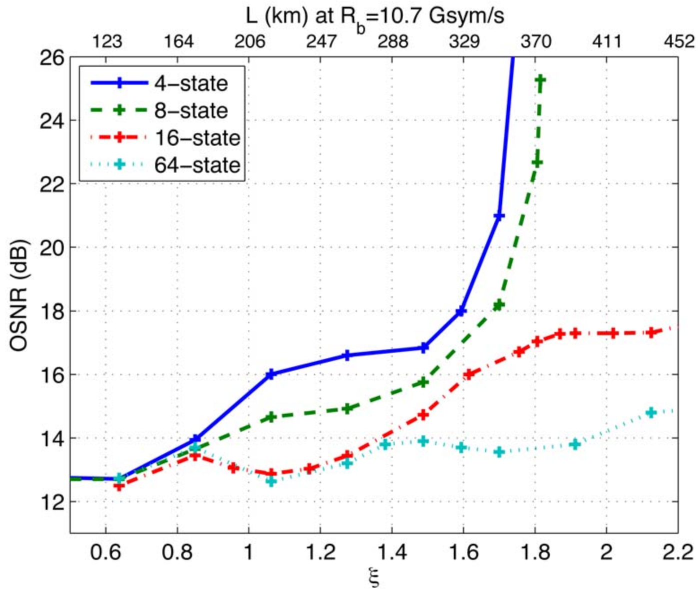


Figure 2.4: Required OSNR for 4-, 8-, 16-, and 64-state Viterbi algorithm at the receiver to reach a bit error rate of 10^{-3}. The x axis is relabeled at the top of the plot by the length of the corresponding fiber at the rate of R_b =10.7 Gsym/s using $D = 17$ ps/(nm.km). The simulations are run for the points and the lines connecting the points are only drawn to aid in viewing.

changes in the value of ξ. Although, increasing ξ increases the memory length and typically would increase the required OSNR, for some ranges of ξ the required OSNR does not increase significantly, or even decreases. This behavior is the result of small shifts in the optimum sampling positions while ξ increases and is consistent with other results presented in the literature (c.f. Figure 15 of [66]). Similar effects also occur for coded systems presented next.

Figures 2.5 and 2.6 show the performance of the uncoded and coded systems with a 4-state Viterbi algorithm using codes *A*, *B*, and *C* before the decoder. Figure 2.5 shows the comparison based on the assumption that the *symbol rates* are equal for all systems. It can be seen that the codes can produce significant improvement in the required OSNR at a fixed value of ξ. For example, this improvement can be as high as 5 dB for code *C* at ξ=1.7. The improvement in the performance can also be measured as the increase in the achievable fiber distance for a fixed OSNR. For example, using code *C* at OSNR=16 dB, the fiber distance increases by 53% or 120 km at $R_b = 10.7$ Gsym/s.

While all the curves in Figure 2.5 have the same symbol rate, their transmission information rates vary because the code rates are different. To compare systems with equal information rates, the symbol rates of each of the systems are rescaled by the code rate. For example, for the system using code *C,* with a code rate $\rho = 0.85$, the symbol rate is increased to $R_b = 10.7/0.85$=12.6 Gsym/s. Figure 2.6 rescales each of the curves in Figure 2.5 by the corresponding code rate so that the *information rate* is equal for all systems as a function of fiber distance. However, since the symbol rate is now different for each code and ξ depends on the symbol rate, there is no longer a fixed relationship between the fiber distance, *L,* and ξ as there was in Figure 2.5. It can be seen that if the system can tolerate the reduction in information rate by using codes, then for a fixed OSNR, significantly longer fiber lengths can be reached. However, if the reduction in information rate can not be tolerated and the system uses a higher symbol rate to achieve the same information rate, then for different fiber lengths, there is a significant reduction in performance. For the best cases, the coded systems show about a 1 dB improvement at specific distances. Code *A* shows an improvement in the range of fiber lengths less than 123 km, code *B* between 180 km and 230 km, and code *C* between 205 km and 250 km.

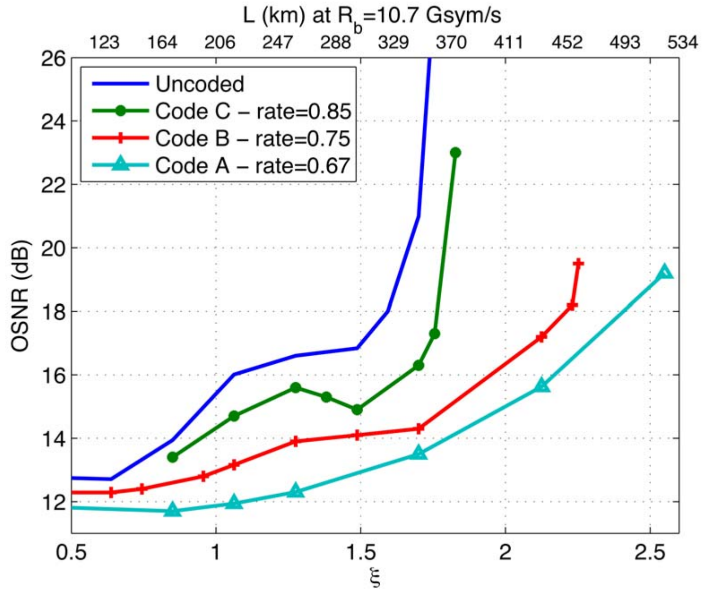


Figure 2.5: Comparison of uncoded and coded 4-state Viterbi detectors. The **symbol rate** of all of the systems is equal to 10.7 Gsym/s.

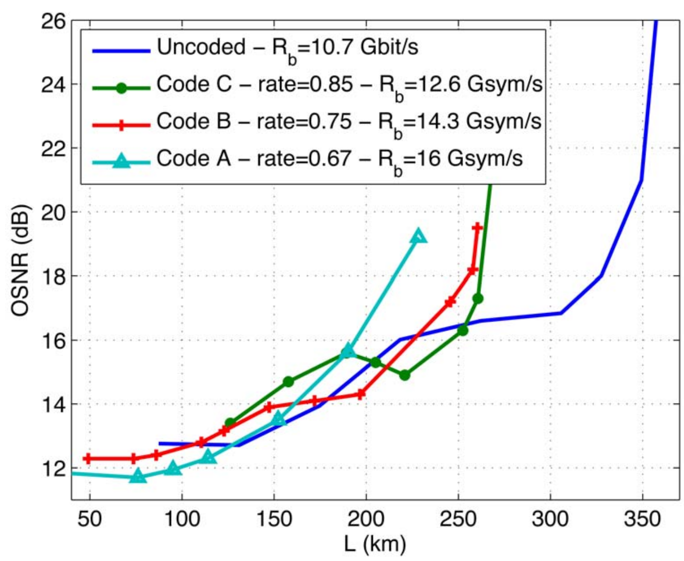


Figure 2.6: Comparison of uncoded and coded 4-state Viterbi detectors. The **information rate** of all of the systems is equal to 10.7 Gbit/s.

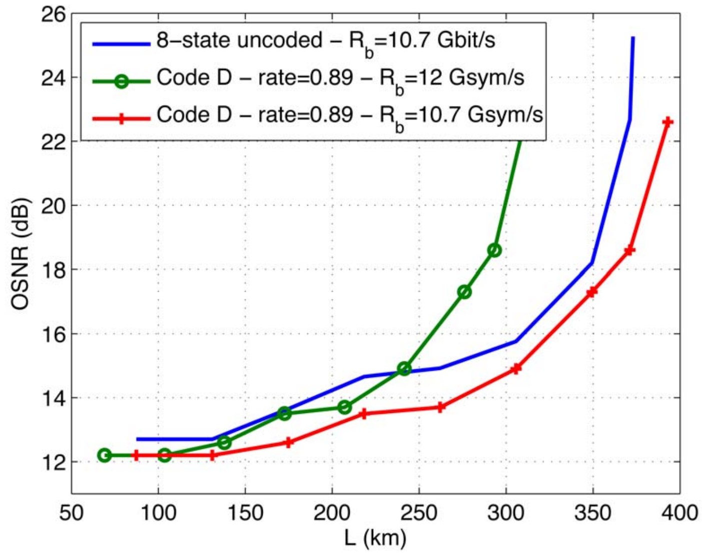


Figure 2.7: Comparison of uncoded and coded 8-state Viterbi detectors. The coded systems with $R_b = 10.7$ Gsym/s and $R_b = 12$ Gsym/s have the same symbol rate and information rate, respectively, as the uncoded system.

To study a higher rate constraint, code *D,* in Table 2.4, with rate 0.89 and longer forbidden patterns was tested. Figure 2.7 compares the performance of uncoded and coded 8-state trellis for both fixed symbol and information rates using code *D*. It can be seen that the coded system with the symbol rate of 12 Gsym/s enhances the performance of 8-state trellis up to 0.5 dB for fiber distances less than 245 km. Compared to the best performance of the three 4-state codes, code *D* performs at most 1 dB better, reaching the best improvement at the fiber distance of 210 km.

2.4.2 Number of States

Another parameter which has a significant effect on the distribution of errors and the effectiveness of the code is the number of states of the trellis. Table 2.5 lists the dominant error events of a system with $\xi = 1.49$ using a 16-state Viterbi algorithm at the receiver. Comparison with the corresponding list in Table 2.3 shows that the error events can be grouped in fewer classes for the 16-state Viterbi algorithm relative to the 4-state Viterbi algorithm. Consider that for $\xi = 1.49$, only 72% of the energy of a symbol can be captured from the two neighboring symbols. Therefore, a 4-state Viterbi algorithm treats the remaining 28% of the energy, which is spread over other symbols, as interference. Consequently, because the dominant error events lists are different for systems using different Viterbi receivers, the improvement in the performance of the systems is different. For example, the improvement of the system with a 16-state Viterbi algorithm using code *C* can be seen in the range of the optical distances of 140 km to 205 km with about 1 dB at the best case while the Figure 2.6 implies that the same code with the 4-state Viterbi algorithm improves the system performance for the fiber distances of 200 km to 250 km.

2.4.3 Optimum Sampling Point

As mentioned in section 2.2.1, an optimum sampling position, Δ_{opt}, exists at the receiver to achieve the best performance. The performance demonstrated in sections 2.3.1 and 2.4 are based on the optimum value of Δ. In this section, we determine the sensitivity of the performance as a function of Δ. Figure 2.8 plots the bit error rate versus Δ for several values of ξ. The OSNR for each value of ξ is chosen such that the bit error rate at the optimum sampling point (Δ_{opt}) is equal to 10^{-3}. In this graph, $\Delta = 0$

Table 2.5: List of dominant error events for $\xi = 1.49$, OSNR=14.5, Δ=0, and bit error rate=10^{-3} using a 16-state Viterbi detector. $P = -log_{10}$(Bit Error Rate of the Error Event Class).

	Error Event Class	P
1	$nnnn + -nnnn$	3.39
2	$nnnn + -n + -nnnn$	3.72
3	$nnnn + nnnn$	3.8
4	$nnnn + n - +nnnn$	4.55
5	$nnnn + -n + nnnn$	4.57
6	$nnnn + -n + - + nnnn$	4.85
	P>5 or error length>14	Tot. Prob. = $10^{-3.7}$

represents the sample r_k generated by the integration of the signal over the time period corresponding to the desired symbol, while $\Delta = 0.5$ represents the sample r_k generated by an integration interval that corresponds to half of the desired symbol and half of the next symbol. The performance is, therefore, symmetric about $\Delta = 0$ and the graph is shown for only $0 < \Delta < 1$.

The graph indicates that Δ_{opt} is a strong function of ξ. The optimum sampling point for lower values of ξ is at the cross-over point of the eye, $\Delta_{opt} = 0.5$, as shown in 2.9(a). The fact that the optimal sampling point is at the cross-over point is a direct consequence of the electrical filter which shifts the unfiltered eye shown in Figure 2.2 by a half symbol period for small values of ξ. As ξ increases, Δ_{opt} shifts to the center of the eye, $\Delta_{opt} = 1$. This effect can be intuitively justified by looking at Figure 2.9

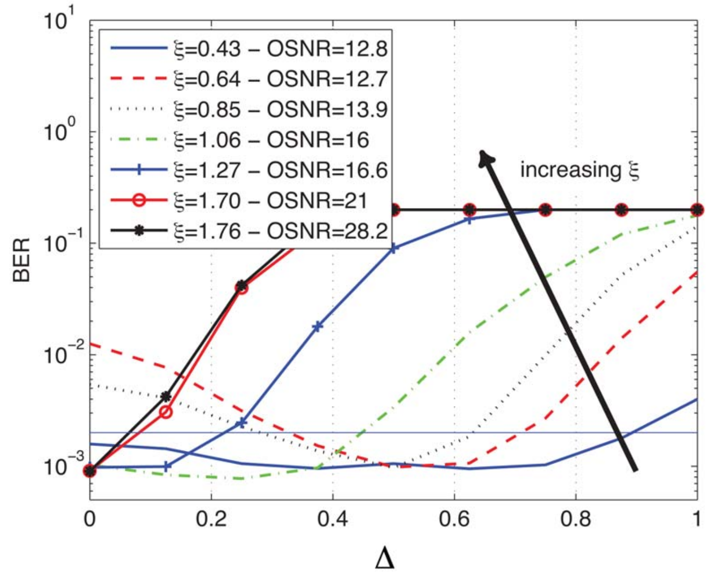
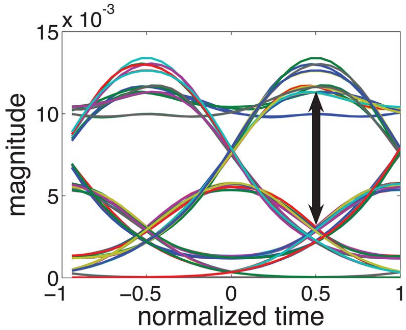


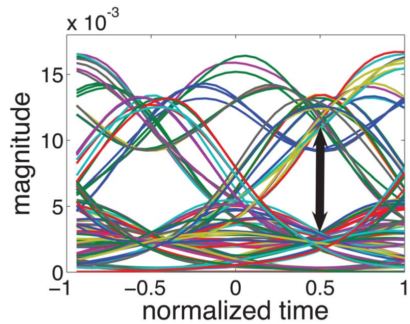
Figure 2.8: Bit error rate versus sampling shift, Δ, using 4-state Viterbi detector for different values of ξ. The thin horizontal line is at bit error rate=2×10^{-3}.

which illustrates the eye diagrams for different values of ξ after electrical filtering, i.e., point **B** in the Figure 2.1. In these eye diagrams, arrows show the widest opening in the eye. Comparing to the results shown Figure 2.8, these arrows are exactly located at the optimum sampling points of the corresponding systems. For example, the maximum opening shown in Figure 2.9(c), for $\xi = 1.70$ occurs at the normalized time corresponding to $\Delta = 0$, which in Figure 2.8, is the sampling position that produce the minimum bit error rate for the system.

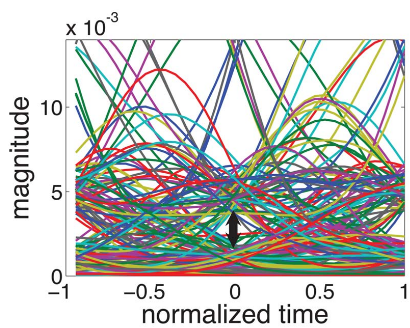
We now define the sampling range tolerance, ***TR***, $0 < TR < 2$, as the continuous range of Δ for which bit error rate is less than 2×10^{-3}. Large values of TR indicate more tolerance to sampling errors. For example, for $TR = 1.5$, the bit error rate is less than 2×10^{-3}, if the sampling occurs anywhere within 1.5 symbol intervals. The



(a) $\xi = 0.43$



(b) $\xi = 0.85$



(c) $\xi = 1.70$

Figure 2.9: Eye diagrams of the received signals after electrical filter (point **B** in Figure 2.1) for systems with different values of ξ. The arrows show widest opening of the eye.

reason the sampling range tolerance can be larger than a symbol interval is again a consequence of the fact that the width of the convolution of the sensed electrical signal over T_b with the rectangular filter function that also has a width T_b is at least $2T_b$. Figure 2.8 shows the largest value of the sampling range, $TR = 1.8$, occurs when the dispersion is low and $\xi = 0.43$. For larger values of ξ, *TR* decreases, as expected, and the bit error rate becomes more sensitive to the sampling point. When $\xi < 1.5$, there are only small changes in the ordering of the most probable error events when the sampling point shifts within the sampling tolerance range defined for bit error rate less than 2×10^{-3}. Therefore, for this range of parameters, the constrained codes studied are reasonably robust with respect to the sampling point.

Figure 2.10 shows TR for 4-state uncoded and coded systems operating at the same information rate. For these systems, the symbol rates are not equal and thus the timing tolerance is expressed in absolute time units using $TR \times T_b$. The sampling tolerance improves from approximately 40 ps for the uncoded system to at most 100 ps for coded systems at specific ranges. For example, at $L = 125$ km, using code B produces a two-fold improvement increasing $TR \times T_b$ from 35 ps to 80 ps. If the systems are operated at the same symbol rate, then the improvement is more dramatic. For example, code *B* increases *TR* to 2, 0.9, and 0.5 at ξ= 0.64, 1.27, and 1.70 respectively. For the uncoded system, the equivalent values determined from Figure 2.8 are 0.375, 0.4, and 0.1, respectively. For $\xi = 1.70$, this represents over a five-fold improvement in the sampling tolerance. These results show that, in addition to the distance-enhancing characteristic, the constrained codes can improve the system performance by increasing

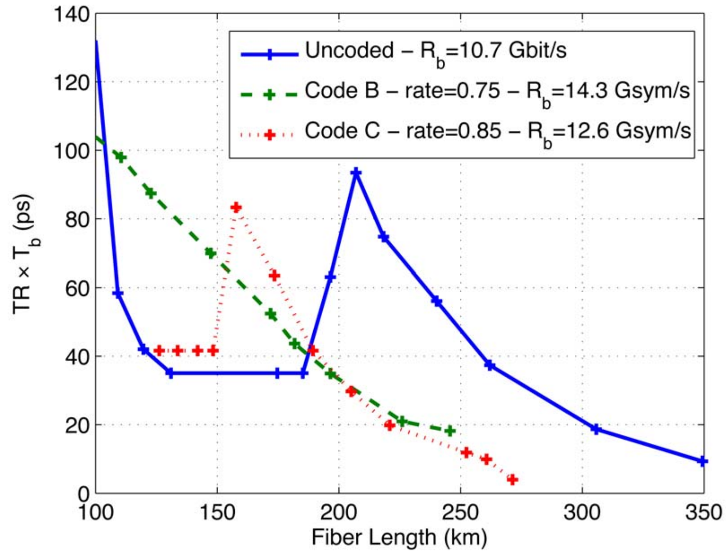


Figure 2.10: Comparison of sampling shift time tolerance for 4-state uncoded and coded systems.

the robustness of the detector with respect to the sampling errors. Intuitively, the use of these codes opens the eye more horizontally than vertically. This improves the tolerance of the system to sampling error and timing jitter. Timing recovery is an important issue which adds complexity to the receiver [67, 68, 69]. Our results indicate that using a constrained code can reduce the complexity of the timing recovery process for specific distances.

2.5 Summary and Conclusion

Detection based on a combination of a reduced-state sequence estimation algorithm and constrained coding can improve both the bit error rate, as well as the robustness of the system to sampling errors for a realistic long-haul fiber channel that has nonlinear intersymbol interference caused by the combination of linear dispersion and square-law detection. This is a low cost approach to combat intersymbol interference for systems of several hundred kilometers. Other approaches may be more applicable at longer distances. The generation of a forbidden list of patterns to incorporate into the code as well as several specific codes was presented. These codes can also incorporate other features such as run-length limits. The improvement depends on the code, the fiber distance, and the symbol rate. For systems with a fixed symbol rate, the coding gain can approach 5 dB relative to the uncoded system for a rate 0.85 code. This improvement can be useful for systems with a fixed symbol rate when the loss of 15% in the information rate is acceptable. In this case, these simple codes can provide longer reach. For fixed information rate systems with varying symbol rates, the quadratic dependence of normalized system memory ξ with respect to the symbol rate reduces the gain to the point where there is only a modest improvement in the error performance over specific distances. However, this improvement is achieved using a very simple encoder and decoder. The improvement with respect to sampling errors is more pronounced. The tolerance to sampling errors can be improved by a factor of 2 for systems operating at the same information rate and over 5 times for systems operating at the same symbol rate. This improves the timing recovery process and reduces the effect of jitter on the bit error rate. The two attributes of improving the bit error rate and increasing the tolerance to sampling errors make these codes attractive candidates for applications when one of these improvements is required. For a specific distance simultaneously achieving both a reduction in error rate and improving timing performance requires further investigation.

These codes are used in conjunction with sequence estimation and are compatible with existing electronic compensation technology. We believe this is the first error characterization and analysis of constrained codes for a nonlinear optical intersymbol interference channels. This work also provides a framework for further investigations in the area of modulation coding for optical channels.

Acknowledgment

The text in Chapter 2, in part, is a reprint of the material as it appears in Zeinab Taghavi, Nikola Alic, and George Papen, "Maximum-Likelihood Detection and Constrained Coding on Optical Channels, " *Journal of Lightwave Technology*, volume 27, Issue 11, pp. 1469-1479 (2009). The dissertation author was the primary researcher and/or author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

3 Modulation Codes to Reduce Nonlinear Cross-Talk in a Dense Wavelength Division Multiplexing Channel

3.1 Introduction

In the previous chapter, we used a channel model where the only fiber impairment was linear chromatic dispersion. Combining this linear impairment with the square-law sensing generated a nonlinear channel model. In reality however, a fiber channel can also have nonlinear propagation characteristics from a Kerr nonlinearity when the launch power is high.

High launch power occur in modern systems that use wavelength division multiplexing method (WDM). In this method, the frequency spectrum of the fiber is divided into orthogonal frequency bands, i.e. wavelength channels. In earlier generation systems, the frequency distance between neighboring channels was designed such that the nonlinear interference was minimized. However, in order to increase the capacity, the channel spacing must be reduced. The resulting modeling format is called dense wavelength division multiplexing (DWDM). For instance if the transmission rate in one channel is 10 Gbps, then a standard channel spacing is 50 GHz for a DWDM system, whereas earlier systems used 100 GHz spacing.

In DWDM systems both the intra-channel and inter-channel linear and nonlinear interference increase. Linear interference is present when the sampling rate of each channel is less than the Nyquist rate. Nonlinear interference is caused by the Kerr nonlinearity effect. In this effect, the interference power is proportional to the third power of the transmitted power of each channel. To mitigate this effect, this chapter presents a design method for new modulation codes. Given a discrete channel model, our goal is to design a code that can decrease the variance of nonlinear interference terms, while maintain the same overall channel capacity.

In the next section we review the nonlinear characteristic of fiber in details. The assumption that we are given a discrete mathematical model of fiber is generally too strong. In the Section 3.4, we propose a numerical simulation to estimate the parameters of the discrete mathematical model of fiber. In Section 3.5, we describe our method to design the constrained modulation code. In Section 3.6 we present our results based on the the proposed methods and we conclude at the end.

3.2 Nonlinear Impairments in DWDM System

In single-mode fibers, we define the transmitted modulated signal in the fiber at a length z and at time t as $A(z,t)$.

$$A(z,t) = E(z,t)\exp(j2\pi f(t-\beta_1 z)),$$

where f is the carrier frequency, $E(z,t)$ is the envelope or low-pass equivalent signal, and β_1 is the group delay. It can be shown that the envelope $E(z,t)$ satisfies the nonlinear Schrödinger equation [7]

$$\frac{\partial E}{\partial z} = -\frac{\alpha}{2}E + j\frac{\beta_2}{2}\frac{\partial^2 E}{\partial \tau^2} + j\gamma|E|^2 E. \tag{3.1}$$

In this equation, τ is the time measured in a reference frame moving at the group velocity defined as $(1/\beta_1)$, α is the fiber loss factor, and β_2 is the group velocity dispersion (GVD) parameter. The first term on the right side of the equation is the fiber loss. The second term in (3.1) is responsible for the chromatic dispersion. The third term is the fiber nonlinearity characterized by the nonlinear coefficient γ. This term governs the nonlinear signal propagation characteristic of fiber. Given this equation, the effect of dispersion and nonlinearity can be assessed.

To estimate the range of the fiber parameters for which each of these effects is important, we study a normalized version of (3.1). This normalization is obtained by scaling the time by the pulse width T,

$$\tau' = \tau/T,$$

and the low-pass amplitude by the peak power of the signal and the amount of the fiber loss. The normalized envelope is then

$$U(z,\tau') = \frac{E(z,\tau')}{\sqrt{P}\exp(-\alpha z/2)}.$$

Using this expression, (3.1) may be written as

$$\frac{\partial U}{\partial z} = j\frac{\operatorname{sgn}(\beta_2)}{4\pi L_D}\frac{\partial^2 U}{\partial \tau'^2} - \frac{\exp(-\alpha z)}{L_{NL}}|U|^2 U, \tag{3.2}$$

where

$$L_D = \frac{T^2}{2\pi|\beta_2|},$$
$$L_{NL} = \frac{1}{\gamma P}.$$

L_D and L_{NL} provide an estimate over the length of the fiber for which dispersion and nonlinearity, respectively, begin to be noticed. If we have a fiber channel with only linear dispersion, then when $L = L_D$, the signal time-width is twice the transmitted signal time-width [7]. Also, if we assume that the system only suffers from nonlinearity without dispersion, then when $L_{eff} = L_{NL}$, the maximum nonlinear phase shift is $|U(0,T)|^2$ radians, where $L_{eff} = (1-\exp(-\alpha L))/\alpha$ is the effective length of fiber (see (3.4)).

The behavior of the fiber can be categorized by the relative magnitude of the dispersion length L_D and the nonlinear length L_{NL}. We then define a ratio

$$\rho_L = \frac{L_D}{L_{NL}} = \frac{\gamma P T^2}{2\pi|\beta_2|},$$

that quantifies which effect is dominant. If $\rho_L \ll 1$ the dispersion-dominant regime is applicable, whereas for $\rho_L \gg 1$ the nonlinearity is dominant. When $L_D \ll L \ll L_{NL}$

the effect of GVD is negligible and the signal is only distorted by the nonlinearity. In this case, if we have a single channel, (3.2) can be simplified as

$$\frac{\partial U}{\partial z} = \frac{j\exp(-\alpha z)}{L_{NL}}|U|^2U. \tag{3.3}$$

A general solution of this equation is

$$U(L,T) = U(0,T)\exp[j\phi_{NL}(L,T)], \tag{3.4}$$

where

$$\phi_{NL}(L,T) = |U(0,T)|^2(L_{eff}/L_{NL}).$$

In (3.4), the signal phase is distorted by the signal power. This effect is called self-phase modulation (SPM) which is severe when $L_{eff} \gg L_{NL}$. If dispersion is included in the channel, there is no closed form solution for (3.2) and the combination of the dispersion and the nonlinear affects both the amplitude and the phase. In multiuser systems, the phase shift ϕ_{NL} is proportional to total power which includes the power of neighboring channels. The terms in the phase shift which are proportional to the power of one interfering channel are cross-phase modulation (XPM) terms and the terms depend on the product of the amplitude of two or more interfering channels are four-wave mixing (FWM) terms. In the presence of chromatic dispersion, signal streams in different channels have different speeds. Therefore, these streams walk off from each other, and as a result, the nonlinear interference from neighboring channels has memory. In most multi-channel WDM systems have tens of channels. In these systems, if the relative walk-off of the channels is large because of the channel dependent group velocity, then the four-wave mixing terms are suppressed since these terms require three different channels to interact. For these systems, cross-phase modulation for which the two of the interacting channels are the same, is the dominant source of degradation of signal. This effect causes distortion in both the phase and the amplitude.

In the next section, we develop a discrete mathematical model for a channel with this form of nonlinear interference.

3.3 Discrete Channel Model

In [17] an approximate discrete model of a weakly nonlinear fiber channel was proposed. It is based on a Voltera series expansion of the nonlinear Schrödinger equation given in [70, 71, 72]. To apply this result, we assume a weakly nonlinear channel where $L_{eff}/L_{NL} = \delta$, in which δ is small that terms of order $O(\delta^2)$ can be ignored from the resulting series.

Discretize $E(L, \tau)$, which is the continuous optical output of the fiber, to obtain a time-discrete optical signal denoted by $y_i(n)$, in which i is the user index and n is the discrete time. Let $x_i(n)$ be the discrete transmitted signal for user i at discrete time n. More precisely, $x_i(n)$ is the time-discretization of $E(0, \tau)$, which is the continuous optical signal at the entrance of the fiber. For a simplified model, we assume $x_i(n) = g_i d_i(n)$, where $d_i(n)$ is the channel-coded symbol sequence, $d_i(n) = 1$ at isolated mark bits, and zero at isolated space bits, and g_i is a complex constant for normalizing the input power and phase for the ith user.

Assume that a dispersion compensation fiber (DCF) is used; in that case, the received signal does not suffer from linear intersymbol interference (ISI), though, it sees the nonlinear interference from other channels as

$$y_i(n) = x_i(n) + I_i(n) + n_i(n),$$

where $I_i(n)$ is the discretization of inter-channel interference, and $n_i(n)$ is an additive Gaussian noise. Note that x_i here is exactly S_i in Section 1.4. The interference I includes self- and cross-phase modulation and four-wave mixing terms. In [17], it is shown that $y_i(n)$ can be written as

$$\begin{aligned} y_i(n) &= x_i(n) + \sum_{\substack{r,s,t,k,l,m \\ f_i = f_{i-k} - f_{i-l} + f_{i-m}}} \xi_{k,l,m}(r,s,t) x_{i-k}(n-r) x^*_{i-l}(n-s) x_{i-m}(n-t) \\ &\quad + n_i(n). \end{aligned} \tag{3.5}$$

Assuming the frequency channels are non-overlapping and equally distant, cross-talk coefficients $\xi_{k,l,m}$ are non-zero for $(i-k) - (i-l) + (i-m) = i$. The second term of the right-hand side of (3.5) can be grouped into three categories: (1) the term with $k = l = m = 0$, which is called self-phase modulation (SPM), (2) the terms with $k = l \neq m = 0$ or $m = l \neq k = 0$, which are called cross-phase modulation (XPM), and (3) the terms with $k \neq 0 \neq m$ which are called the four-wave-mixing (FWM). Self-phase and cross-phase modulation terms can be written as

$$\zeta_i(n) = \sum_{l,s} [\xi_{l,l,0}(s,s,0) + \xi_{0,l,l}(0,s,s)] |x_{i-l}(n-s)|^2 x_i(n). \tag{3.6}$$

$\xi_{l,l,0}(s,s,0)$'s are complex numbers. If we define

$$\nu_l(s) \triangleq [\xi_{l,l,0}(s,s,0) + \xi_{0,l,l}(0,s,s)],$$

then (3.6) can be written as

$$\zeta_i(n) = x_i(n) \sum_{l,s} \nu_l(s) |x_{i-l}(n-s)|^2.$$

In the channel models that for $s \neq 0$, $\nu_l(s) \neq 0$, the interference has memory caused by walk-off. The total signal $y_i(n)$ is equal to

$$y_i(n) = x_i(n) + \zeta_i(n) + \phi_i(n) + n_i(n), \tag{3.7}$$

where $\phi_i(n)$ is the FWM term.

When there is significant dispersion, the cross-talk coefficients involved in the FWM term are usually small[1]. Based on this observation, we neglect this term in our model.

If we use a noncoherent receiver, a simplified mathematical model of the received signal for user i at time n at the output of the photodiode is

$$\begin{aligned} r_i(n) &= |y_i(n)|^2 \\ &= |x_i(n) + \zeta_i(n) + n_i(n)|^2. \end{aligned}$$

Ignoring terms of order x_i^6, $r_i(n)$ can be simplified as

$$\begin{aligned} r_i(n) &\approx |x_i(n)|^2 + \sum_{s,l} \rho_l(s) |x_i(n)|^2 |x_{i-l}(n-s)|^2 + O(\rho^2 x_i^6) + \text{Noise Term} \\ &\approx |x_i(n)|^2 + \kappa_i(n) + O(\rho^2 x_i^6) + \text{Noise Term}, \end{aligned} \tag{3.8}$$

where $\rho_l(s)$ is the real part of $\nu_l(s)$ and is the nonlinear cross-talk. In the next section,

[1] It can be seen in Equation (92) of [17] that for more distant ω_1 and ω_2 or ω and ω_1, H' is smaller. Therefore, in Eq (93) of [17], if $k \neq l$ and $l \neq 0$, the amount under the integral decreases and therefore, $\xi_{k,l,m}$ is smaller than the coefficients $\xi_{l,l,0}$ or $\xi_{0,l,l}$.

we present a method to estimate the cross-talk coefficients $\rho_l(s)$. Note that r_i here is exactly R_i in Section 1.4.

3.4 Estimation of the Cross-Talk Coefficients

In [71, 72], a method to estimate the cross-talk coefficients based on the Voltera series was presented. However, the Voltera series does not produce accurate estimates of these coefficients for practical regimes we are interested in. Therefore, we generated the cross-talk coefficients using a commercial fiber channel simulating program[2].

Consider a noiseless two channel system. In this system, a continuous wave (CW), with constant power of P is sent through the central channel (channel 0) and a block of random data is sent through channel l_0. In this case, (3.8) simplifies to

$$r_i(n) \quad = \quad P + P\sum_{s,l}\rho_l(s)|x_l(n-s)|^2. \tag{3.9}$$

We average $r_i(n)$ over the values of n and name it m. Since the block of random data has been sent through channel l_0, the probability of $|x_l(n-s)|^2 = P$ is 1/2 and the probability of $|x_l(n-s)|^2 = 0$ is 1/2. Therefore m can be calculated as

$$m = E_n\{r_i(n)\}|_{|x_i(n)|^2=P} \quad = \quad P + P^2\rho_0(0) + P^2\sum_{s}\rho_{l_0}(s)/2.$$

Now for each value of s_0, we average $r_i(n)$ for the values of n for which $|x_{l_0}(n-s_0)|^2 = 0$ and name the average as m_{s_0}

$$m_{s_0} = E_n\{r_i(n)\}|_{|x_i(n)|^2=P\&|x_{l_0}(n-s_0)|^2=0} \quad = \quad P + P^2\rho_0(0) + P^2\sum_{s\neq s_0}\rho_{l_0}(s)/2.$$

[2] VPITransmissionMaker

Therefore, $\rho_{l_0}(s_0)$ can be estimated as

$$\rho_l(s_0) = 2(m - m_{s_0})/P^2. \tag{3.10}$$

The goal of this section is to use these estimated parameters to define the discrete mathematical model of the channel given in (3.8). Two examples of the calculation of the cross-talk coefficients $\rho_{l_0}(s_0)$ are presented in Section 3.6. In the next section, we describe a method to design a constrained modulation code based on the fitted discrete mathematical model to reduce the cross-talk interference.

When the information of the received signal from the neighboring channels is not accessible, the interference must be treated as noise. To suppress the cross-talk interference, different methods have been proposed. Using constant amplitude modulation limits the presence of interference to only the phase of the optical signal. As the interference is proportional to the instantaneous power of the signal, the constant amplitude makes the interference predictable and easy to remove. Therefore, phase modulation instead of intensity modulation can improve the performance of the system [19]. One method for suppressing inter-channel four-wave mixing (FWM) is by using unequally spaced frequency channels with some sacrifice of spectral efficiency. Alternating polarizations is another method to combat cross-phase modulation and four-wave mixing. Another method uses optical phase conjugation to exactly reverse the propagation direction and phase variation of the signal. All of these methods are complex and sacrifice spectral efficiency.

Compensation methods using digital signal processing (DSP) are less complex and expensive than optical methods. Multiuser detection is one method to combat performance degradation caused by inter-channel interference [17]. For this method, it is assumed that the detector has access to the received signals from the neighboring channels. As a result, the interference is known and can be compensated. However, multiuser detection is complex and not always practical because the receiver may not have access to the received signal from the neighboring channels.

3.5 Modulation Code Design to Mitigate the Cross-talk Interference

The received signal $r_i(n)$ defined in (3.8) suffers from interference from other users which reduces the signal to interference plus noise ratio $SINR$. Since the interference at bit zero and one are not equal, we define

$$\begin{aligned} SINR &= Pr\{d_i(n) = 1\} SINR|_{d_i(n)=1} + Pr\{d_i(n) = 0\} SINR|_{d_i(n)=0} \\ &= Pr\{d_i(n) = 1\} \frac{P}{N_0 + \sigma_I^2} + Pr\{d_i(n) = 0\} \frac{P}{N_0}, \end{aligned} \tag{3.11}$$

where N_0 is the power spectral density of the noise, σ_I^2 is the power of cross-talk interference, i.e., κ_i, in (3.8), and P is the power of signal $x_i(n)$ when $d_i(n) = 1$. In (3.11), it is assumed that the interference is negligible at bit zero, $d_i(n) = 0$. It should be noted that this definition of $SINR$ is not unique and our coding design method is not based on the specific definition of $SINR$.

Modulation coding is a method to avoid interference instead of compensating it. The goal is to increase the total capacity of the channel $C = C_d C_n$, where C_d is the coding capacity of the transmitted data stream and $C_n = B\log_2(1 + SINR)$ is the capacity of the noisy channel. We increase C by decreasing coding capacity C_d, but increasing channel capacity C_n. To increase C_n, we design code to increase the signal to interference plus noise ratio ($SINR$) by decreasing the interference power. This is done by controlling the transmitted bit sequence. Assume the target channel is channel 0 and that the data in all channels is independent and identical and time-invariant. The problem reduces to finding codes or modulation methods to minimize σ_I^2 where

$$\sigma_I{}^2 = E\{|\kappa_0(0) - \mu_I|^2\} \tag{3.12}$$

and

$$\mu_I = E\{\kappa_0(0)\}. \tag{3.13}$$

In (3.12) and (3.13), the expectation E is over the time- and channel-invariant random variables $x_l(s)$. For the systems that we are interested in, it is assumed that the FWM term, i.e., $\phi_i(n)$, is negligible. Therefore, the mean of the interference can be rewritten as

$$\begin{aligned}
\mu_I &= E\left\{\sum_{s,l} \rho_l(s)|x_0(0)|^2|x_{-l}(-s)|^2\right\} \\
&= \sum_{s,l} \rho_l(s) E\{|x_0(0)|^2|x_{-l}(-s)|^2\} \\
&= \sum_{s,l} \rho_l(s)\eta_l(s).
\end{aligned}$$

Let's assume that the power in each channel is equal to P. In addition for simplicity, we assume that the transmitted pulse shape is rectangular. Therefore, without loss of generality, we can assume that, $|x_l(s)|^2 = Pd_l(s)$, where $d_l(s) \in \{0, 1\}$ is the transmitted coded-bit. With these assumptions, $\eta_l(s)$ can be written as

$$\eta_l(s) = E\{|x_0(0)|^2|x_{-l}(-s)|^2\} = P^2 Pr\{d_0(0) = 1 \,\&\, d_{-l}(-s) = 1\}.$$

The variance of the interference is then equal to

$$\begin{aligned}
\sigma_I^2 &= E\left\{\left||x_0(0)|^2 \sum_{s,l} \rho_l(s)\, |x_{-l}(-s)|^2 - \mu_I\right|^2\right\} \\
&= E\left\{\left|\sum_{s,l} \rho_l(s)\left[|x_0(0)|^2|x_{-l}(-s)|^2 - \eta_l(s)\right]\right|^2\right\} \\
&= \sum_{\substack{(l,s)\\(k,r)}} \rho_l(s)\rho_k(r)\left[E\left\{|x_0(0)|^4\, |x_{-l}(-s)|^2\, |x_{-k}(-r)|^2\right\} - \eta_l(s)\eta_k(r)\right]. \\
&= P^4 \sum_{\substack{(l,s)\\(k,r)}} \rho_l(s)\rho_k(r)\left[Pr\left\{d_0(0) = 1 \,\&\, d_{-l}(-s) = 1 \,\&\, d_{-k}(-r) = 1\right\}\right. \\
&\quad \left. - Pr\{d_0(0) = 1 \,\&\, d_{-l}(-s) = 1\} Pr\{d_0(0) = 1 \,\&\, d_{-k}(-r) = 1\}\right].
\end{aligned} \tag{3.14}$$

The goal of the design of our constrained modulation code is to minimize the interference over mark signals, i.e., σ_l^2 given $d_0(0) = 1$. The main assumptions for this design is that we have excluded the second order inter-channel interference effects, chromatic dispersion is assumed to be fully compensated by dispersion compensation fiber (DCF), interference is not present at bit zero, and the signal has a rectangular shape with equal power in all channels. The designed code is for a single-user encoder and a single-user decoder and is time- and channel-invariant.

$$\begin{aligned}
\sigma_I^2|_{d_0(0)=1} &= P^4 \sum_{\substack{(l,s)\neq(0,0)\\(k,r)\neq(0,0)}} \rho_l(s)\rho_k(r)\left[Pr\left\{\, d_{-l}(-s)=1 \;\&\; d_{-k}(-r)=1\right\}\right. \\
&\quad \left. -Pr\{d_{-l}(-s)=1\}Pr\{d_{-k}(-r)=1\}\right] \\
&= P^4 \sum_{\substack{(l,s)\neq(0,0)\\(k,r)\neq(0,0)}} \rho_l(s)\rho_k(r)\left[Pr\left\{d_{-l}(-s)=1 \;\&\; d_{-k}(-r)=1\right\}\right. \\
&\quad \left. -Pr\{d_0(0)=1\}^2\right]. \qquad (3.15)
\end{aligned}$$

Define the probabilities $p(i)$ and $p_{s-r}(i|j)$ for $i,j \in \{0,1\}$ as follows

$$\begin{aligned}
p(i) &= Pr\{d_l(s)=i\} \\
&= Pr\{d_0(0)=i\}, \qquad (3.16)
\end{aligned}$$

$$\begin{aligned}
p_{s-r}(i|j) &= Pr\left\{d_l(s)=i \mid d_l(r)=j\right\} \\
&= Pr\left\{d_0(s-r)=i \mid d_0(0)=j\right\}. \qquad (3.17)
\end{aligned}$$

Using Bayes rule, we can write

$$\begin{aligned}
Pr\left\{d_{-l}(-s)=1 \;\&\; d_{-k}(-r)=1\right\} &= Pr\left\{d_{-l}(-s)=1 \mid d_{-k}(-r)=1\right\}\times \\
&\quad Pr\{d_{-k}(-r)=1\}. \qquad (3.18)
\end{aligned}$$

For $l \neq k$, it is assumed the bit sequences are independent. Therefore, we have

$$Pr\left\{d_{-l}(-s)=1 \;\&\; d_{-k}(-r)=1\right\} = \begin{cases} p_{s-r}(1|1)p(1), & l=k, \\ p(1)^2, & l\neq k. \end{cases} \qquad (3.19)$$

Using the probability definitions (3.16)-(3.19), σ_I^2 is then equal to

$$\sigma_I^2|_{d_0(0)=1} = P^4 \sum_{\substack{l\neq 0\\ s\neq 0\neq r}} \rho_l(s)\rho_l(r)p(1)\left[p_{s-r}(1|1)-p(1)\right]. \qquad (3.20)$$

In these equations, it is assumed that the coding is time- and channel-invariant.

The probabilities defined in (3.16)-(3.17) satisfy a set of constraints. The conditional equations are listed as follows,

$$\forall i, j \in \{0,1\} \forall p(i), p_s(i|j) \geq 0, \tag{3.21}$$

$$p(0) + p(1) = 1, \tag{3.22}$$

$$p(i) = p_s(i|0)p(0) + p_s(i|1)p(1), \tag{3.23}$$

$$p_s(i|j)p(j) = p_{-s}(j|i)p(i), \tag{3.24}$$

$$p_s(0|j) + p_s(1|j) = 1. \tag{3.25}$$

The goal is to minimize equation (3.20) given a set of constraints on the values of the probabilities. The values of conditional properties define the constrained coding Markov chain.

3.5.1 Markov Chain

To define a code and calculate its capacity, we define a finite state Markov chain. A Markov chain is a sequence of random variables S_t, for $t \geq 1$ with the Markov property, namely that, given the present state, the future and past states are independent.

$$Pr\{S_{t+1} = s|S_1 = s_1, S_2 = s_2, ..., S_t = s_t\} = Pr\{S_{t+1} = s|S_t = s_t\}.$$

In information theory, the entropy is a measure of the uncertainty associated with a random variable. The capacity quantifies the expected value of the information contained in a message. For a finite state Markov chain, the capacity is calculated as [11]

$$C = -\sum_{s,r} Pr\{S_t = s|S_{t-1} = r\}Pr\{S_{t-1} = r\}\log(Pr\{S_t = s|S_{t-1} = r\}). \quad (3.26)$$

In this chapter, for a given memory length M we define the state as

$$S_t = \{d_0(\tau)|t - M \leq \tau < t\}. \quad (3.27)$$

The probabilities $Pr\{S_t = s|S_{t-1} = r\}$ define the constrained code. These probabilities can be calculated based on probabilities $p(i)$ and $p_s(i|j)$.

3.5.2 Design of Constrained Code

To design the code, we solve the optimization problem

$$\mathrm{MIN}_p \ \sigma_I^2|_{d_0(0)=1} \quad = \quad \mathrm{MIN}_p \ P^4 \sum_{\substack{l \neq 0 \\ s \neq 0 \neq r}} \rho_l(s)\rho_l(r)p(1)\left[p_{s-r}(1|1) - p(1)\right], \quad (3.28)$$

given the following constraints:

1. The parameters p satisfy the equalities (3.22) to (3.25).

2. The capacity of the Markov chain defined as (3.27) is greater than a given value of C_0.

Satisfaction of the second constraint guarantees a minimum coding capacity for the designed code. Here we present two examples of the design of the code.

Example 1 *Markov chain with memory length of zero*

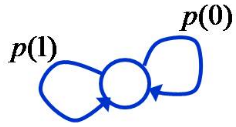


Figure 3.1: Markov chain model of memory length M=0.

We start with a simple code without memory. The memory length of the constraint, i.e., M is set to zero. In this case we have

$$p_{s-r}(i|j) = \begin{cases} p(i), & s \neq r, \\ 0, & s = r \ \& \ i \neq j, \\ 1, & s = r \ \& \ i = j. \end{cases} \tag{3.29}$$

As the result, the optimization problem will reduce to

$$\mathrm{MIN}_p \ \ P^4 \sum_{\substack{l \neq 0 \\ s \neq 0}} \rho_l(s)^2 p(1) \left[1 - p(1)\right], \tag{3.30}$$

with the conditions,

1. $p(1) + p(0) = 1$,

2. $C = -p(1) \log p(1) - P(0) \log p(0) > C_0$.

Figure 3.1 shows the simple 1-state Markov model of the code.

Example 2 *Markov chain with a memory length of one*

In this example, we assume we have one bit of memory, i.e., $M = 1$. In this case the optimization problem is as in (3.28). However, it should be noted that the only free variables in the minimization are $p(1)$, $p_1(1|1)$, and $p_1(1|0)$. All other variables can be

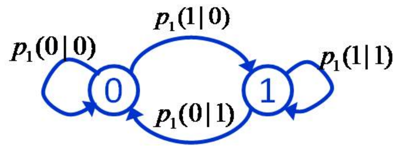


Figure 3.2: Markov chain model of memory length M=1.

calculated using equations (3.22)-(3.25) based on these three variables. The conditional Markov probabilities are $Pr\{S_t = i|S_{t-1} = j\} = p_1(i|j)$. Therefore, the capacity constraint for a memory length $M = 1$ is

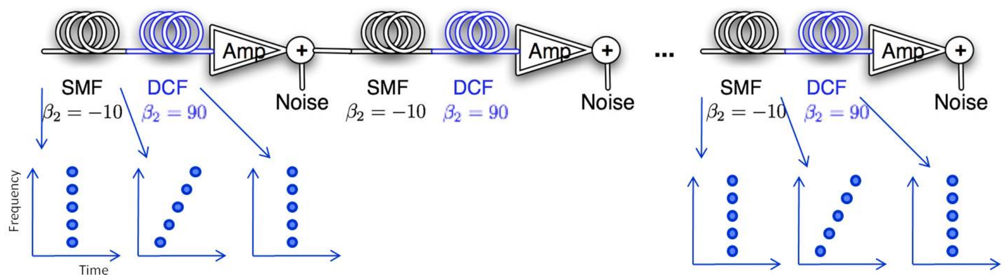
$$C = -\sum_{i,j=0,1} p_1(i|j)p(j)\log p_1(i|j) > C_0.$$

Figure 3.2 shows the 2-state Markov chain model of the code.

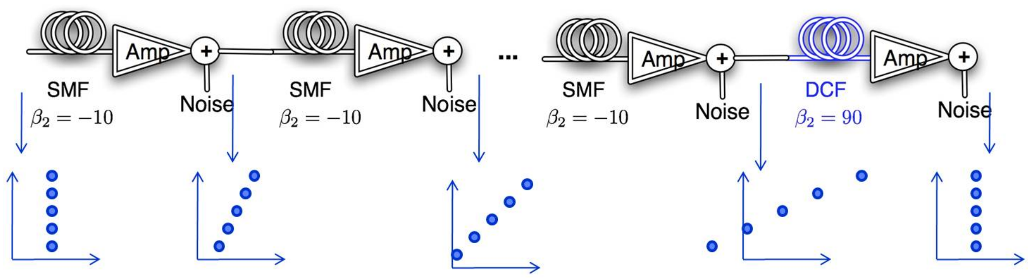
3.6 Results

We considered systems with two types of linear dispersion maps. The first one is with dispersion compensation fiber at each span of fiber and second is with one dispersion compensation fiber at the receiver. Figure 3.3 shows the diagram of the two types of dispersion maps. We estimated the nonlinear coefficient for the discrete channel model ρ for these two dispersion maps using the simulation tool VPITransmissionMaker.

Figures 3.4 and 3.5 show the result for the two dispersion maps given in Figure 3.3. For these simulations, it is assumed $\beta_2 = -10$ ps^2/km and $\lambda = 2$ for the single mode fiber, the length of each span of fiber is 50 km, seven frequency channels spaced 50 GHz apart, and the data transmission rate is 10 Gbps. The channels are numbered



(a) Block diagram of dispersion map 1 with dispersion compensation fiber (DCF) in each span of fiber.



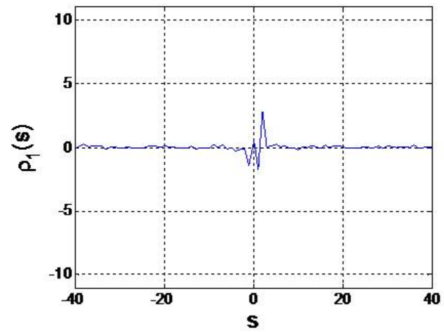
(b) Block diagram of dispersion map 2 with one dispersion compensation fiber (DCF) at the receiver.

Figure 3.3: Two Dispersion maps.

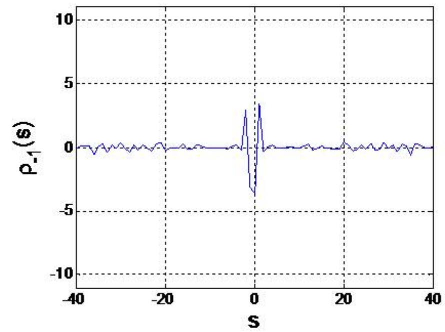
from -3 to 3 and the code is designed for the target channel 0.

We designed Markov model with memory length zero and one, based on Examples 1 and 2, for these two systems. Figures 3.6(a) and 3.6(b) show the normalized variance of interference versus the change in the probability of bit one $p(1)$. Black line in both figures is the result of calculation of variance using equations (3.30) and (3.28), given ρ values illustrated in Figures 3.4 and 3.5. To confirm the analysis we simulated the channels in VPITransmissionMaker with the designed code and calculated the variance of interference. Although the analytical and experimental results do not match completely, their behavior are similar. Therefore, these analytical results give a range of parameters that is close to the optimal result. In both figures, there is a hard constraint of a minimum code capacity of 0.9, and the improvement in the interference variance is 0.89; this is achieved by $p(1) = 1/3$.

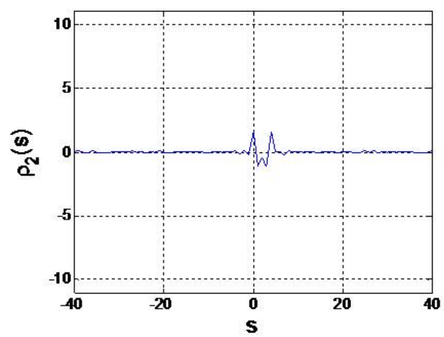
Figure 3.7 shows the level plots of code capacity and normalized interference variance versus Markov transition probabilities $p(1|1)$ and $p(1|0)$ for the code designed with memory length one for the dispersion maps in Figure 3.3. The optimum probabilities, computed in Example 2, occur at $p(1|1) = 2/3$ and $p(1|0) = 1/3$. This will maintain the code capacity to be not less than 0.9. The improvement in the interference variance is 0.83 for Figure 3.7(a) and 0.62 for Figure 3.7(b). We can see that for the dispersion map with dispersion compensation fiber at the end the code improves the performance significantly. Assuming a negligible noise power, $SINR$ is increased by 2



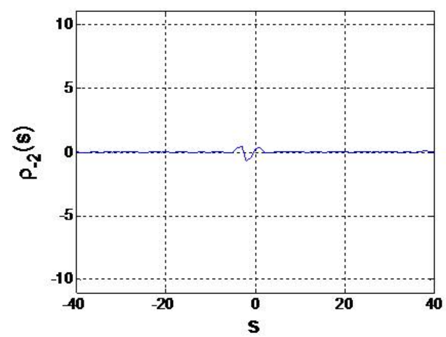
(a) Channel 1



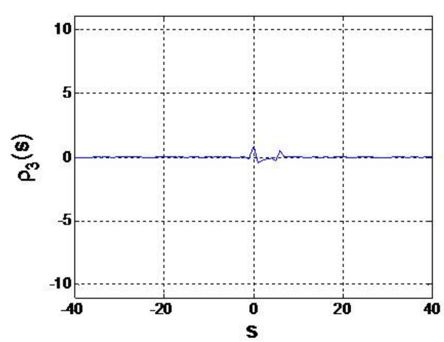
(b) Channel -1



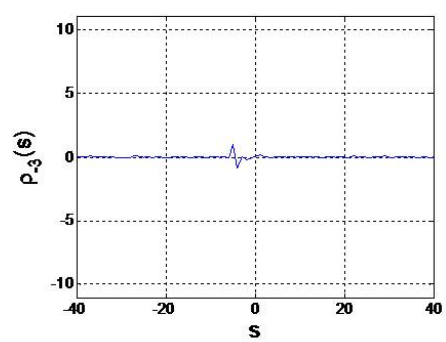
(c) Channel 2



(d) Channel -2

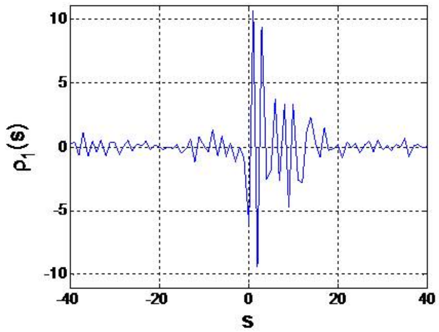


(e) Channel 3

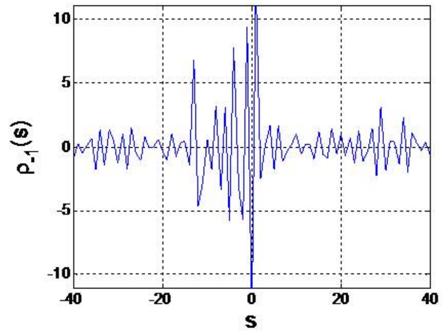


(f) Channel -3

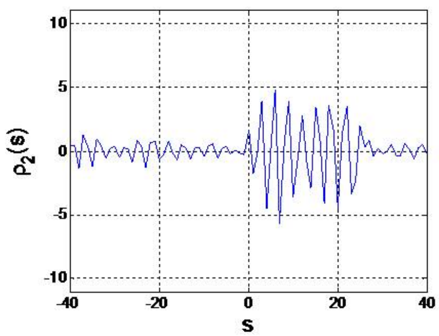
Figure 3.4: Cross-talk coefficients ρ calculated base on (3.10) for the dispersion map shown in Figure 3.3(a).



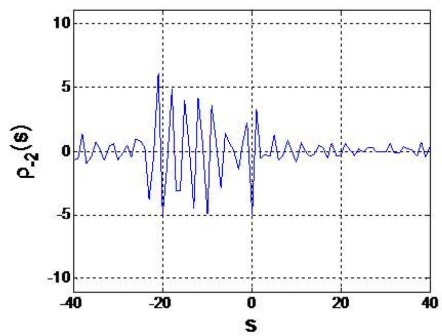
(a) Channel 1



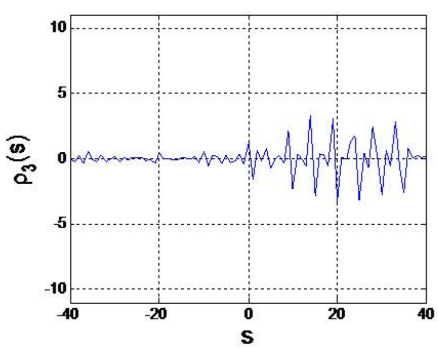
(b) Channel -1



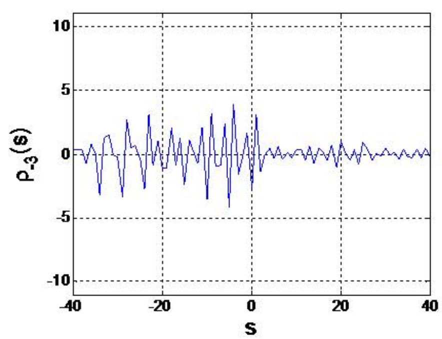
(c) Channel 2



(d) Channel -2

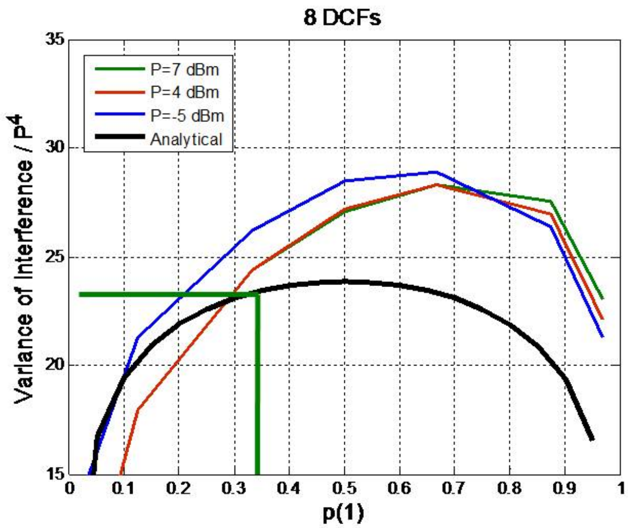


(e) Channel 3

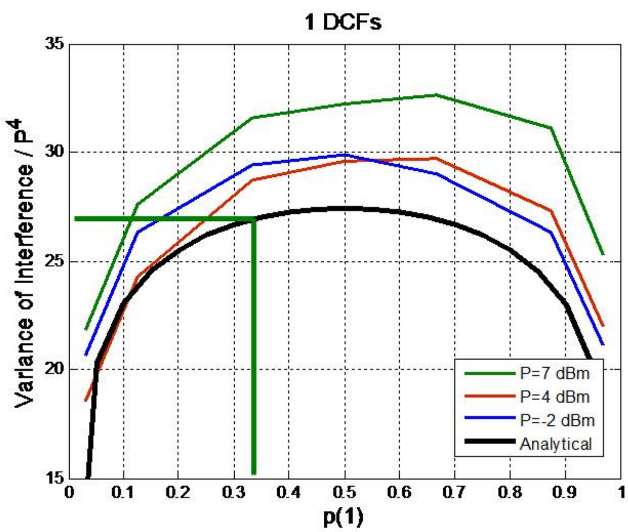


(f) Channel -3

Figure 3.5: Cross-talk coefficients ρ calculated base on (3.10) for the dispersion map shown in Figure 3.3(b).



(a) Dispersion map shown in Figure 3.3(a)



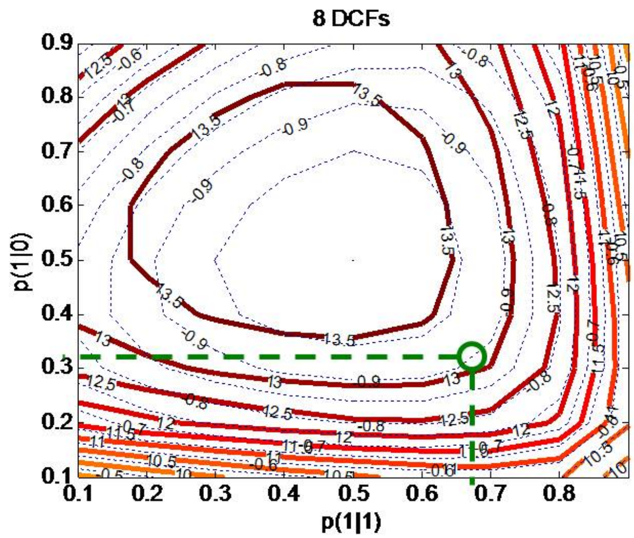
(b) Dispersion map shown in Figure 3.3(b)

Figure 3.6: Normalized variance of interference versus probability of bit one $p(1)$ for coding with memory length zero.

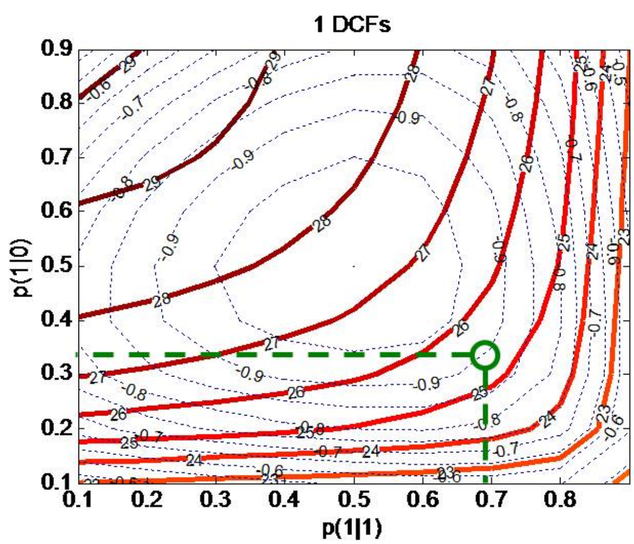
dB.

3.7 Summary and Conclusion

We presented a method to design modulation codes to decrease the self-phase and cross-phase modulation interference power in a dense intensity-modulated wavelength division multiplexing channel. Our approach was based on the design of constrained codes assuming that the transmitter and receiver have access to only one user. We designed finite-state Markov chains, with the transition probabilities optimized so that the power of the cross-talk for the transmitted coded signal is minimized. Our results show that the overall capacity as well as the interference variance improve. These results motivate the following future works: (1) solving more examples with different dispersion maps, (2) generalizing the optimization function, which is currently the cross-talk power in bits carrying signal 1, to complex function, (3) examining more complex modulation formats like 16-QAM, etc., (4) considering coherent receiver instead of noncoherent receiver, and (5) using multiuser coding and decoding.



(a) Dispersion map shown in Figure 3.3(a)



(b) Dispersion map shown in Figure 3.3(b)

Figure 3.7: Code capacity and normalized variance of interference versus conditional probabilities for coding with memory length one. Dashed lines are code capacity level plots. Solid red lines are the level plots of normalized interference variance.

4 Multiuser Receiver to Combat Nonlinear Cross-Talk

4.1 Introduction

In dense wavelength division multiplexing (DWDM) transmission systems, compensating nonlinear effects including self-phase modulation (SPM) and cross-phase modulation (XPM) is a challenging problem. The interference caused by these nonlinear effects is proportional to the third power of signal energy [7]. Therefore, increasing the launch power into the fiber decreases the spectral efficiency [12]. In contrast, if the launched power into the fiber channel is low so that the channel response is linear, then the spectral efficiency of channel increases logarithmically with respect to input power. As the power increases and nonlinear effects become dominant, the spectral efficiency decreases exponentially with power. The threshold power where fiber channel begins to exhibit nonlinear effects increases with increasing chromatic dispersion and with increasing channel spacing, because the channels walk off more quickly and this decreases the impact of cross-phase modulation. In [19, 23], phase-shift keying (PSK) modulation was suggested to mitigate these effects. For noncoherent modulation techniques based on both optical compensation as well as electronic compensation has been proposed [73]. In [74] higher order constellations are used to increase the capacity. In the last chapter, we studied modulation coding to combat this effect when the only information that is available is from the user's received signal. In this chapter, we study the compensation techniques based on the information in multiple channels. This is called multiuser detection.

When self-phase modulation is the dominant impairment in a DWDM system, a maximum likelihood sequence estimation (MLSE) produces a large improvement in performance with respect to threshold detectors [16] using only the information in one channel. However, when cross-phase modulation is significant, single-user MLSE fails to show any improvement [75]. In [17], the authors showed that using an optimal multiuser receiver, the capacity of the WDM system can be increased to reach the capacity of a single-user channel. The idea of using multiuser receiver was also suggested in [18, 76]. In these papers, analytical bounds on the performance of a multiuser MLSE receiver in a WDM system was derived.

It should be noted that the optimal multiuser receiver which detects all of the channels at the same time has a high complexity and it is not practical in current system configurations because detecting the information in all of the channels is not feasible. However, some sub-optimal receivers that are less complex can produce significant improvements. In this chapter, we use multiuser MLSE to achieve this goal. To keep the complexity as low as possible, we only investigate the performance of a two-user (TU) 16-state Viterbi algorithm (VA). We show that in the case which cross-phase modulation is the dominant effect, this receiver can provide a significant improvement in the performance with respect to a single-user receiver. In the next two sections, the optical channel model and the receiver used in this chapter is described. We then investigate the effect of having memory in time and cross channel dependency and the effect of laser phase noise on the performance of the receiver.

4.2 Memory in Time and Cross Channel Dependency

In Chapter 2, the system is assumed to have memory in time. In that case, the MLSE is the optimal receiver. In a multiple-access system, the interference depends on the cross talks from neighboring channels. In this chapter, we describe a multiuser MLSE approach suitable for multiple-access systems with interference that can have memory in time and dependency on neighboring frequency channels.

4.2.1 Setup

Figure 4.1 shows the schematic of the fiber optic channel model we used in this chapter. In our channel model, up to seven channels with spacing of 50 GHz ranging from 193.25 to 193.55 THz (1551.3 to 1548.91 nm) are used, each having individual lasers and pulse pattern generator using a Mach-Zehnder modulator (MZM). The modulation format is non-return-to-zero (NRZ) on-off keying (OOK) at the data rate of 10.7

Gbps. The total length of the transmission link is 500 km which is divided to N spans consisting of a length of fiber followed by an optical amplifier and a dispersion compensation fiber (DCF). The length of each span is changed from 50 km to 125 km, which corresponds to a range of 4-10 spans. Dispersion coefficients of the transmission fibers used are $D = 4,\ 8,$ and 16.9 ps/nm.km.

The dispersion coefficient for each fiber in each span is equal. The dispersion compensation fibers are chosen to have the dispersion coefficient of $D = -70$ ps/nm.km with their length match to the span such that the residual dispersion in each span is equal to 0. In this setup, amplification was provided by a gain-controlled erbium-doped fiber amplifier (EDFA). The power launched into each span is equal to the launch power of the system. This power is set such that the total input and output power of the first amplifier in each span is equal to -10 dBm and 5 dBm, respectively, to avoid nonlinear effects caused by the dispersion compensation fiber. The channels of interest were selected with a 50 GHz bandwidth tunable optical fiber.

We use a numerical simulation tool (VPItransmissionMaker) to simulate the setup. This software solves the nonlinear equation of the channel using a split-step method. The length of the sampled sequence is restricted to $2^{14} - 1$. Instead of having distributed sources of noise in each amplifier, the amplified spontaneous emission noise (ASE) are simulated by an additive Gaussian noise added once at the end of the link. The power of the Gaussian noise is calculated to be equal to the total power of the noise if it were distributed.

It is known that the other nonlinear effects modify the noise statistics so that

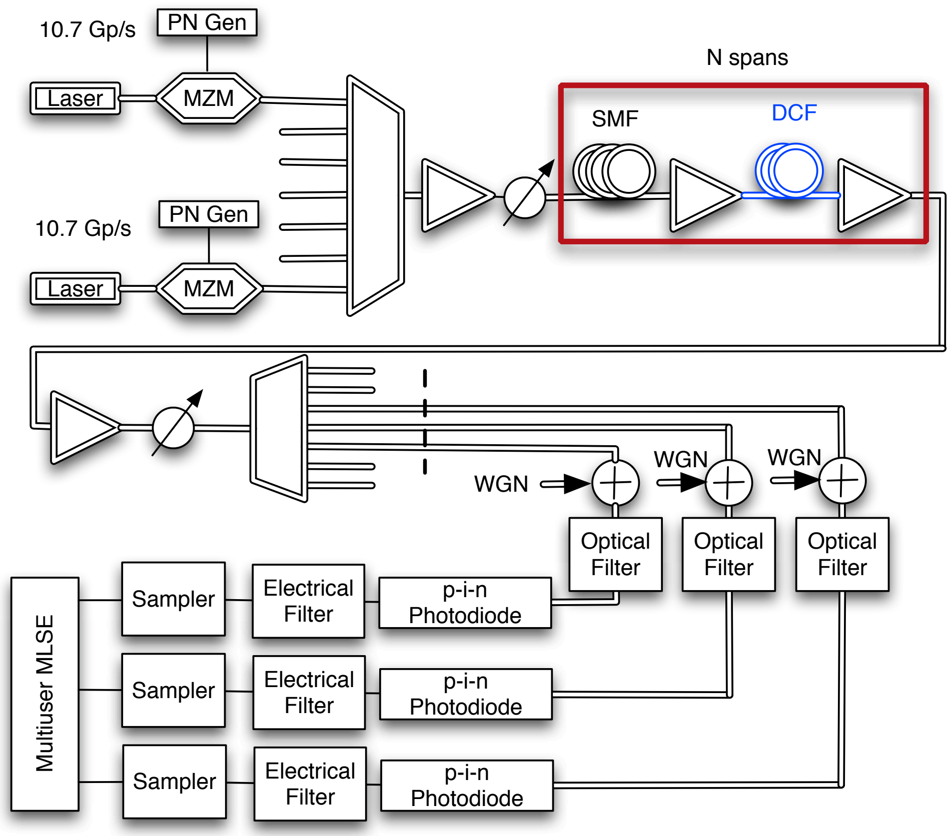


Figure 4.1: Setup of the dense wavelength division multiplexing system.

output distribution is non-Gaussian. We use a Gaussian approximation because it is tractable and is a reasonable starting point to evaluate the performance of the receiver algorithms.

Receiver

In this setup, the optical filter is designed to band-limit the noise at 10.7 GHz and the electrical filter is an integrator. In the receiver, the performance of two types of detectors are compared. The first detector is a simple single-user (SU) MLSE and the second is a two-user (TU) MLSE. For the two-user receiver, every pair of neighboring channels are detected in one receiver. The single-user receiver combats self-phase modulation, while two-user receiver is designed to combat cross-phase modulation caused by the neighboring frequency channel. Because of our concern about complexity, we only considered a 16-state two-user algorithm. In this type of MLSE, the states represent the last four bits of the two sequences of data in the neighboring channels (two bits per channel) at a given time. For each sample time, the detector has two inputs which are the noisy output symbols in the two neighboring channels. The branch metrics for the MLSE algorithms are calculated based on the joint probability density function (pdf) of the two received symbols from the two channels given three consecutive bits from each channel. At the end of the process, the two established sequences of bits are generated.

Table 4.1 illustrates the dependence length of 4 different single-user and two-user receivers, both in time and frequency channels. The dependence length in time is the corresponding number of neighboring bits in the transmitted bit sequence. The dependence length in frequency channels is the number of neighboring frequency channels, which incorporate in the detection of the output of one channel. Based on this definition, a 2^n-state single-user has a dependence length of n bits in time and no dependency in frequency channel, while a 2^n-state two-user Viterbi algorithm is a receiver with a dependence length in frequency channel of 1 and $n/2$ bits of dependence length in time.

The complexity of the receivers have direct relationship with the number of branches per step in the Viterbi algorithm. A 2^n-state two-user Viterbi algorithm has

Table 4.1: Comparison of the complexity of different receivers. SU and TU stand for single-user and two-user receivers, respectively.

Receiver	Dependence in Time	Dependence in Channel	Number of Branches per User
SU 4-State	2 bits	0	8
SU 16-State	4 bits	0	32
SU 32-State	5 bits	0	64
TU 16-State	2 bits	1 channel	32

2^{n+2} total branches, or 2^{n+1} branches per channel. Also, a 2^n-state single-user Viterbi algorithm has 2^{n+1} branches for one channel. Therefore, the complexity of a 16-state single-user and two-user Viterbi algorithm receivers are comparable.

4.2.2 Results

Table 4.2 shows the bit error rate of the three different systems using three types of fibers with different dispersion coefficients. The fiber link is a 500 km link with 4 spans of length 125 km. The average launch power of the link is equal to 12.8 dBm per channel. At this power level, the nonlinearity in the fiber shows a significant impact on bit error rate. The amount of additive noise injected to the system is calculated from $N_{se} = (N'_{se,1} + N_{se,2})N_{spans}$, where $N'_{se,1}$ is the equivalent power density of the amplified spontaneous emission noise generated by the first amplifier in each fiber span

Table 4.2: Comparison of the bit error rate of different systems with 3 choices of fiber dispersion coefficients (D) and 4 choices of receivers. SU and TU stand for single-user and two-user receivers, respectively.

D [ps/(nm.km)]	4	8	16.9
SU 4-State	1.2×10^{-1}	1.8×10^{-2}	8.2×10^{-4}
SU 16-State	4.9×10^{-2}	6.1×10^{-3}	2.4×10^{-4}
SU 32-State	7.4×10^{-3}	2.2×10^{-3}	1.5×10^{-4}
TU 16-State	3.2×10^{-4}	1.8×10^{-5}	1.7×10^{-5}

after passing through the second amplifier and $N_{es,2}$ is the power density of the amplified spontaneous emission noise generated by the second amplifier. The power densities of the noise of each amplifier are calculated by $N_{se} = hfn_{sp}(G-1)$, where h is Planck's constant, f is the central frequency, n_{sp} is the spontaneous emission factor, and G is the gain of the amplifier. The noise factor, F_N, of the amplifiers are assumed to be equal to 4, and the relation of F_N and n_{sp} is defined as $F_N = 1/G + 2n_{sp}(G-1)/G$. In Table 4.2, it can be seen that bit error rate at the end of a two-user receiver is at least one order of magnitude less than the bit error rate of a 32-state single-user receiver, which has higher complexity. This result shows that in these scenarios the effect of cross-phase modulation is more significant than self-phase modulation. Therefore multiuser receiver is more helpful than single-user MLSE receivers with long memory-length in time.

For the lengths of fiber spans other than 125 km, if only the amplified spontaneous emission noise from the amplifiers are considered in the simulation, the optical signal-to-noise ratio is high such that the bit error rate is lower than 10^{-6} and our simulation does not produce enough error samples to calculate the bit error rate with a high degree of statistical confidence. Therefore, to see the behavior of the receivers for different lengths of fiber spans, we adjust the noise level so that the bit error rate for these systems is higher than 10^{-6}. Figure 4.2 shows bit error rate versus span length for a fixed level of noise. In this figure, the length of each span is increasing from 50 km to 125 km. As the span length increases, the launch power must also increase so that the power at the end of each span is -10 dBm. When the span length is 50 km the required launch power is equal to -6 dBm/ch, and at the receiver the optical signal-to-noise ratio is about 8 dB, while for span length of 125 km, the average input power is equal to 12.8 dBm/ch and the optical signal-to-noise ratio is 6 dB.

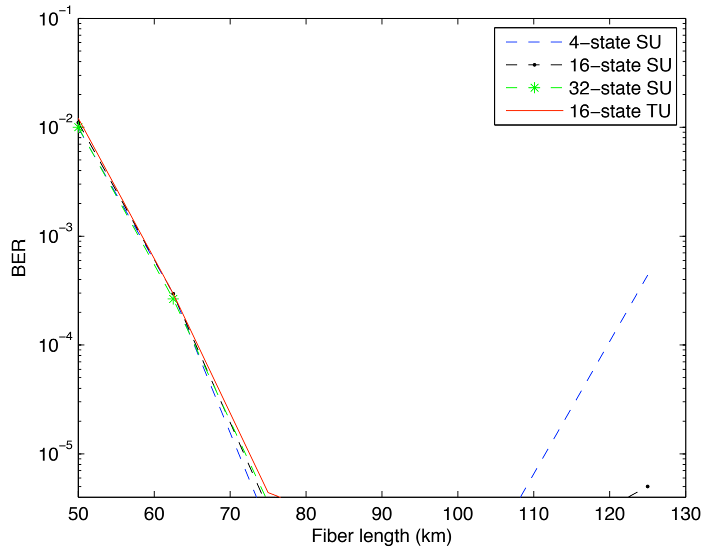
Every plot in Figure 4.2 contains three regimes: (1) Linear, where the launch power is low and noise is the dominant effect. As the length of fiber increases, the optical signal-to-noise ratio increases and therefore bit error rate decreases. (2) Uncharacterized intermediary, for which the optical signal-to-noise ratio is improved and the bit error rate can be low. (3) Nonlinear, where the launch power is so high that the Kerr nonlinearity effect is dominant. In this regime, as the fiber length increases, the power dependent nonlinearity increases, optical signal-to-noise ratio decreases, and the bit error rate increase. In all of these plots, 16-state two-user receiver outperforms the single-user receivers when power increases. At the fiber length of 125 km the improvement is more than 2 orders of magnitude. This means that for this set of conditions and power, the interference caused by cross-phase modulation is much stronger than the linear intersymbol interference and one effective method to combat that is to spread the memory of the receiver along the wavelength of the channels other than inside a channel. This is in agreement with the results presented in [17].

In low powers, the performance of all receivers is almost the same. As the power increases, the difference is more clear. In Figure 4.2(c), at the fiber length of 100 km, the 16-state two-user and the 4-state single-user, both with 2 bits of memory in time, have the same bit error rate which is slightly more than that of 16-state and 32-state single-user receiver. In this mode, the cross-phase modulation is not significant. However, at the length of 125 km, the performance of single-user receivers degrades while the two-user receiver provides the same bit error rate as the 100 km fiber length.

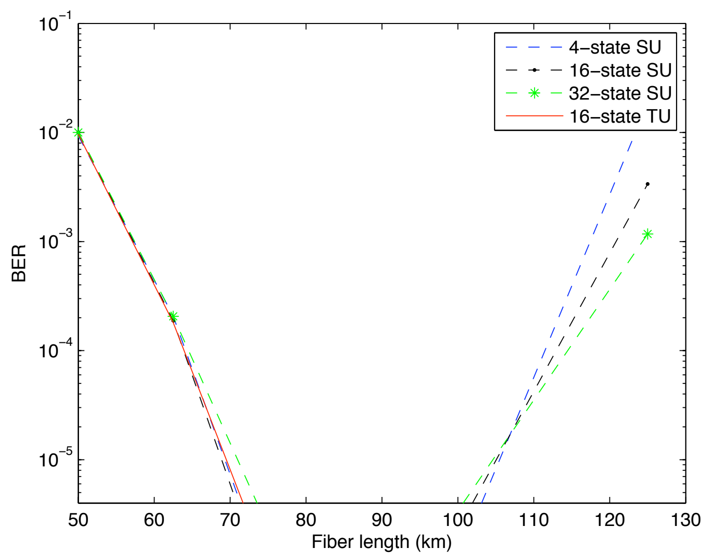
Another interesting phenomenon in the plots of Figure 4.2 is that by increasing the dispersion coefficient, the performance of the system at a high power level improves. The reason is that at a higher dispersion level, walk-off is more significant and cross-phase modulation averages out over more number of bits.

4.3 Effect of Laser Phase Noise

In this section, we investigate the effect of laser phase noise on the performance of the detection.

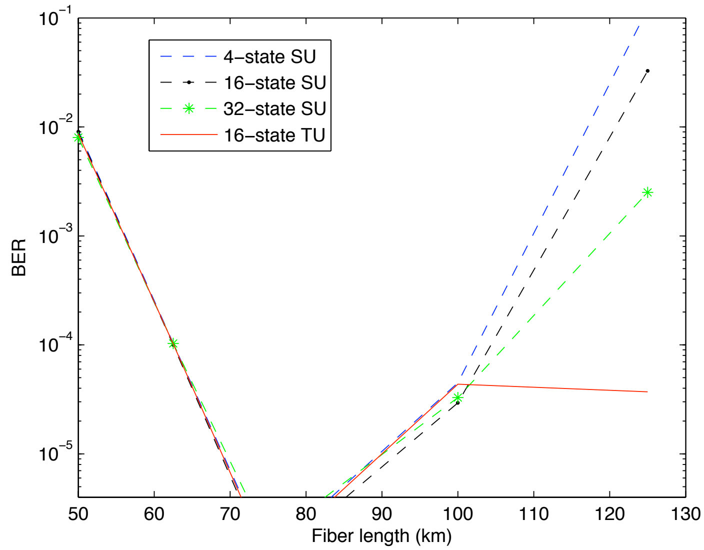


(a) $D = 16.9$ ps/(nm.km)



(b) $D = 8$ ps/(nm.km)

Figure 4.2: Comparison of bit error rate for different choices of fibers and receiver, given a fixed level of noise. The plots are cropped from below due to our inability to accurately measure the bit error rate.



(c) $D = 4$ ps/(nm.km)

Figure 4.2: Continued from previous page.

4.3.1 Laser Phase Noise

Phase noise is the random fluctuations in the phase of a waveform, caused by time domain instabilities. The output of a single-frequency laser is not perfectly monochromatic but rather exhibits some phase noise. Phase noise can be quantified by the power spectral density of the phase deviations, having units of $\mathrm{rad}^2/\mathrm{Hz}$ (or simply Hz^{-1}. This power spectral density often diverges for zero frequency, so that an r.m.s. value with integration down to zero frequency can not be specified. For a simple random-walk process, the specification of a coherence time or coherence length or of a linewidth value can be appropriate. The linewidth of a laser is typically specified as the width (typically the full width at half-maximum, FWHM) of its optical spectrum. More precisely, it is the width of the power spectral density of the emitted electric field in terms of frequency, wavenumber, or wavelength. The linewidth of typical-single frequency lasers can vary from 1 kHz to 100 MHz.

4.3.2 Setup

The setup to study the effect of laser phase noise on the performance of the receiver is the same as the setup in the last section with the following changes. The number of channels is reduced to three neighboring channels, ranging from 193.35 to 193.45 THz (1550.5 to 1549.7 nm). We keep individual lasers for each channel but instead of having individual pulse pattern generators, we use only one pattern generator for all channels. Therefore, a span of fiber is used to slide the channels with respect to each other for around 4 bits, such that the resulting three channels experience independent bit streams at each time. The total length of the transmission link is 180 km which is simulated using loops of a single span of fiber. We only consider the transmission fiber with the dispersion of $D = 4$ ps/nm.km. The power lunched into each span is equal to 13 dBm. In the last section it was assumed that the laser linewidth is zero. In this section, the simulation is repeated for two values of the laser linewidth: 1 kHz and 10 MHz.

We compared the performance of single-user, two-user, and three-user Viterbi receivers. For each of these receivers, the Viterbi algorithm considered the memory of two intra-channel neighboring bits. Therefore, single-user Viterbi includes $(2^2 =) 4$ states and $(2^{2+1} =) 8$ branches and detects the received signal of one channel. Two-user Viterbi algorithm is based on $(2^{2*2} =) 16$ states and $(2^{2*2+2} =) 64$ branches and detects

Table 4.3: Bit error rate (BER) of 4 different receivers for channel 2 applying lasers of linewidth 1 kHz. The BER of two other channels at 1550.5 nm and 1559.7 nm are lower than 10^{-6} which could not be measured accurately. ch stands for channels

	BER of channel 2 @ 1550.1 nm
single-user receiver	1.9×10^{-4}
two-user receiver (ch 1&2)	1.4×10^{-5}
two-user receiver (ch 2&3)	6.4×10^{-5}
three-user receiver (ch 1&2&3)	$\sim 1 \times 10^{-6}$

the received bit sequences of two channels. Finally, the three-user Viterbi algorithm utilizes $(2^{2*3} =)\, 64$ states and $(2^{2*3+3} =)\, 512$ branches and detects received bit sequences of three channels in one run of the Viterbi algorithm.

4.3.3 Results

When three channels are transmitted, the two side channels are impaired mostly from cross-phase modulation, while for the central channel four-wave mixing has also significant effects. Table 4.3 shows the bit error rate of 4-different receivers for the setup with a linewidth 1 kHz. The bit error rate for the channels 1 and 3 were essentially zero, which is not shown in the table. It can be seen that two-user receivers which are supposed to eliminate cross-phase modulation interference decreases the bit error rate of channel 2 about an order of magnitude and the three-user Viterbi also decreases the bit error rate about another order of magnitude, which is the compensation of four-wave mixing interference from the other two channels.

Table 4.4 demonstrates the bit error rate of receivers assuming the lasers have linewidth of 10 MHz. Let α be the ratio of the bit error rate of the two-user receiver over the bit error rate of single-user receiver for the same channel. Let β be the ratio of the bit error rate of the three-user receiver over the bit error rate of two-user receiver for the same user. The parameters α is equal to 0.13 and 0.61 and β is equal to 0.18 and 0.98 for channels 1 and 3, respectively. The smaller α and β, the more improvement multiuser receivers provide. Our results show that α is generally less than β. This means that the main part of multiuser receiver improvement comes from the two-user receiver rather than from the three-user receiver. This can be translated to the fact that the effect of cross-phase modulation is stronger than four-wave mixing. Also, we can see that the bit error rate improvement for the middle channel, channel 2, is not as significant as that for the side channels. The reason is that the main effect on the channel 2 is four-wave mixing. In contrast to cross-phase modulation in which the amount of interference is only proportional to the amplitude of the neighboring channel, the amount of interference in four-wave mixing is proportional to the amplitude and phase of the neighboring channels. In the presence of phase noise, phase shift between neighboring channels cannot be estimated by training. Therefore, without knowing the phase, the multiuser receiver is not effective in reducing the interference. From the results of two Tables 4.2 and 4.3 we can see that the phase noise caused by laser has strong effect on the effectiveness of the receiver when four-wave mixing is significant. On the other hand, when cross-phase modulation is dominant, multiuser receiver is effective. In the case where four-wave mixing is dominant, a solution to this problem is to estimate the

Table 4.4: Bit error rate (BER) of 4 different receivers for channel 2 applying lasers of linewidth 10 MHz. The BER of not shown are lower than 10^{-6} which could not be measured accurately. ch stands for channels

	BER of ch 1 @ 1549.7 nm	BER of ch 2 @ 1550.1 nm	BER of ch 3 @ 1550.5 nm
single-user receiver	5.6×10^{-5}	1.9×10^{-4}	6.7×10^{-5}
two-user receiver (ch 1&2)	7.1×10^{-6}	9.6×10^{-5}	-
two-user receiver (ch 2&3)	-	1.2×10^{-4}	4.1×10^{-5}
three-user receiver (ch 1&2&3)	1.3×10^{-6}	7.2×10^{-5}	4.0×10^{-5}

phase before the receiver.

4.4 Summary and Conclusion

The results presented in this section show that in a DWDM system, based on the strength of the nonlinearities which cause the inter-symbol interference (ISI) or inter-channel interference (ICI), single-user or multiuser receiver will be helpful. We showed, specifically in Table 4.2, that using a low complexity two-user receiver can reduce the bit error rate of a system with high power by 2 orders of magnitude using a simplified noise model. Therefore, it helps to reduce the number of amplifiers in a link. For example, in this case the length of fiber in each span can be extended to 125 km, or for a link of 500 km, the number of amplifiers can be reduced to 4. We also observed that in the case laser has severe phase noise, to keep the good performance, a phase estimator is necessary.

5 Maximum-Likelihood Equalization for Polarization Multiplexed Quadrature Phase Shift Keying (PolMux-QPSK) Modulated Channel

5.1 Introduction

In this thesis, all of the result up to this point was based on the optical intensity-modulated direct-detection (IMDD) channels. In these channels, the information is modulated onto the intensity of optical signal, and receiver detects the instantaneous power of the received signal. In these channels, the linear effect in optical domain like chromatic dispersion (CD) and polarization mode dispersion (PMD) are nonlinear on the received signal in electrical domain. Till recently, most systems were using binary modulation formats like IMDD which use on-off keying (OOK) or differential phase shift keying (DPSK). These modulations encode one bit per symbol. For dense wavelength-division multiplexing (DWDM), these modulation formats are able to achieve spectral efficiencies of 0.8 bit/s/Hz per polarization because of filtering constraints. Researchers have studied spectral efficiency limits for various detection and modulation methods in the linear [20, 21, 22] and nonlinear regimes [12, 23]. Noncoherent detection and differentially coherent detection limit the degrees of freedom available for the encoding of the information [23], but provide good power efficiency.

In coherent detection, the received electrical signal contains both phase and amplitude of the received optical signal. Therefore in systems with coherent detection, using phase modulation is attractive. Coherent detection allows to maximize the power and signal-to-noise ratio, since it is more robust to nonlinear phase effects. In addition, this detector allows the transmission of signals in two polarizations. Using polarization multiplexing (PolMux modulation) increases the achievable spectral efficiency [24]. If the output is sampled at the Nyquist rate of the transmitted signal, digital signal processing (DSP) can be used to compensate linear transmission impairments. A DSP-based receiver is the most flexible platform to implement adaptive algorithms that compensate time-varying transmission impairments, provide advanced forward error-correction coding, delay, and split digitized signal without degradation in signal quality. Improvements in semiconductor manufacturing technology has enabled the use of DSP-based techniques in coherent optical systems.

Compensating the impairments of a channel at the receiver of a coherent polarization multiplexed systems is a challenge, especially when there is a significant nonlinearity. In an ideal DSP block at the receiver, the linear chromatic dispersion, the nonlinear phase modulation, four-wave mixing, Kerr nonlinearity, polarization dependent effects, and carrier phase noise should be compensated. Digital equalization for linear polarization mode dispersion has been extensively studied [77, 78, 79, 80]. In [81], a least mean square technique by the help of training sequence is used for compensation. In [82], a decision-directed adaptive filter was proposed. When phase-shift keying modulation is used, a constant modulus algorithm (CMA) is a robust adaptive filter for demultiplexing and equalizing the polarization effects [83, 84, 85]. This method has a low complexity and is robust to both frequency offset and phase noise.

In [27, 86, 87], a DSP multistep linear and nonlinear equalization was studied. This compensation is called backpropagation. The backpropagation method is based on the inverse of the solution to the nonlinear Schrödinger equation using the split-step method. In this chapter and Chapter 6, we describe this method in detail.

In this chapter, we extend existing backpropagation techniques. This extension uses the Viterbi algorithm which is a maximum-likelihood sequence estimator (MLSE) in conjunction with backpropagation to compensate for the residual chromatic dispersion, self-phase modulation, and polarization mode dispersion. These residual effects were not fully compensated in backpropagation. We use a two-user Viterbi algorithm to detect the signals at both polarizations at the same time. We show that this method will help to improve the performance of the system.

5.2 Channel Model

The channel model used in this chapter differs from the earlier chapters because we wish to study if the detector can compensate for both cross-polarization nonlinear effects as well as linear effect. This channel model consists of 42.8 Gbps nonlinear transmission, using polarization multiplexed QPSK data at 10.7 GBaud with 4 bits per symbol. The total length of the transmission link is 60 km and is followed by one amplifier. The dispersion coefficients of the transmission is $D = 6.9$ ps/nm.km. The details of the transmitter, the receiver, and the DSP detector are described below.

5.2.1 Coherent Transmitter and Receiver

Figure 5.1 depicts the structure of the quadrature phase shift keying polarization multiplexed (QPSK-PolMux) transmitter. In this transmitter, the laser power is separated to two orthogonal polarizations by polarization beam splitter and fed into two phase modulators. Two bit streams are used for each quadrature component of phase shift keyed system.

At the receiver, the signal is mapped from the complete base band field in two polarization optical field into four electrical signals, corresponding to the in-phase and quadrature field components of the two polarizations. A coherent direct demodulation receiver is used to do that. In an direct demodulation optical receiver, the optical local oscillator has a frequency within the signal bandwidth but unlike the homodyne approach it is not phase locked to the original carrier. This eliminates the need for a

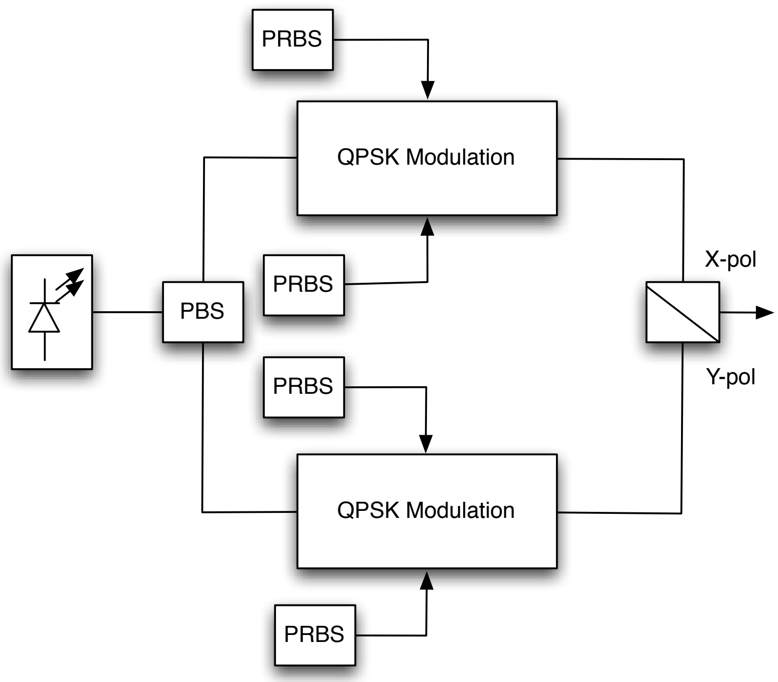


Figure 5.1: PolMux QPSK Transmitter. PBS stands for polarization beam splitter. PRBS stands for pseudo-random binary sequence.

complex control loop around the laser and a much narrower IF bandwidth is required as compared to the heterodyne approach. Figure 5.2 shows the structure of the receiver. In this receiver, signal is coupled with the local oscillator, and the output is equal to [85]

$$\begin{bmatrix} I_x \\ Q_x \\ I_y \\ Q_y \end{bmatrix} = \alpha \underbrace{\begin{bmatrix} Re\{E_x E_{lo}^*\} \\ Im\{E_x E_{lo}^*\} \\ Re\{E_y E_{lo}^*\} \\ Im\{E_y E_{lo}^*\} \end{bmatrix}}_{\text{coherent terms}} + \beta \underbrace{\begin{bmatrix} 2|E_x|^2 + 2|E_{lo}|^2 \\ 4|E_x|^2 + |E_{lo}|^2 \\ 2|E_y|^2 + 2|E_{lo}|^2 \\ 4|E_y|^2 + |E_{lo}|^2 \end{bmatrix}}_{\text{direct detection terms}},$$

where α, β are some constants. The direct detection local oscillator power can be removed using a DC block. If the ratio of the local oscillator to the signal is significantly larger than the signal-to-noise ratio the remaining direct detection terms can be ignored.

In practice, this ratio is in the region of 20 dB.

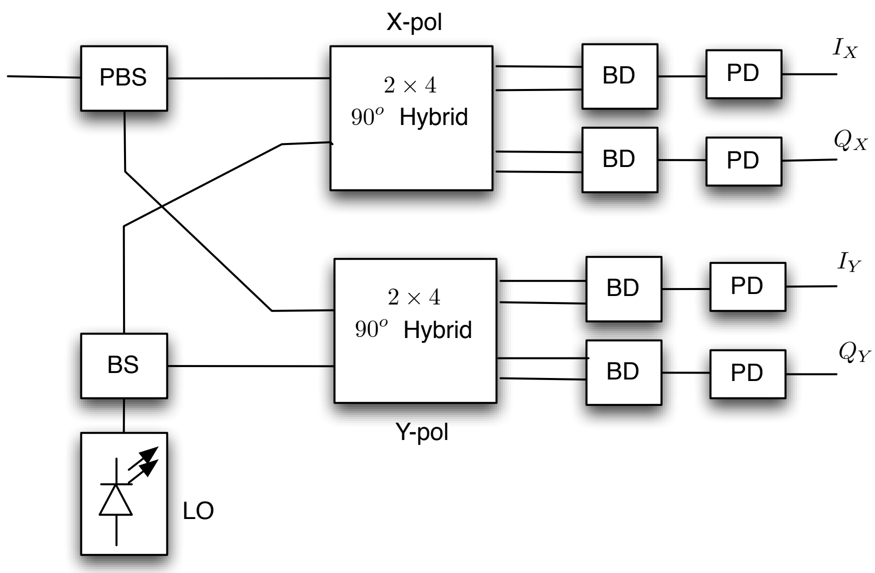


Figure 5.2: Polarization multiplexed coherent receiver. PBS stands for polarization beam splitter. BS stands for beam splitter. BD stands for balance detector. PD stands for photodiode.

5.2.2 Digital Detector

After signal is digitized, digital signal processing (DSP) is used to track the phase and polarization of the signal, which reduces the complexity of the receiver compared to an optical homodyne receiver. Figure 5.3 is the setup of channel model. The details of each DSP block follows.

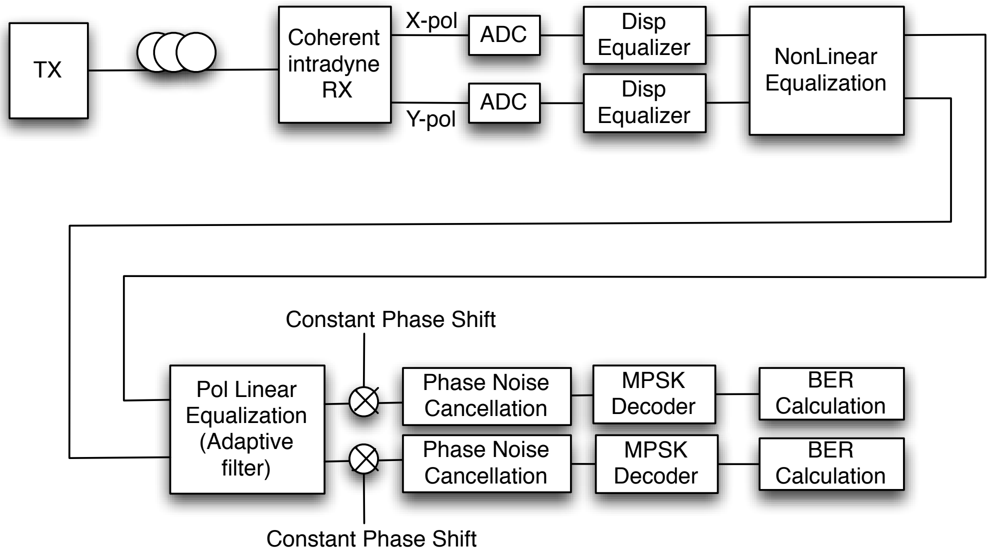


Figure 5.3: Block diagram of the transmitter and receiver of polarization multiplexed quadrature phase shift keying modulation. TX stands for transmitter. RX stands for receiver. ADC stands for analog to digital convertor. MPSK stands for M-ary phase shift keying. BER stands for bit error rate.

Polarization Demultiplexer and Polarization Mode Dispersion Equalizer

Polarization mode dispersion (PMD) is a linear distortion that occurs because the two orthogonal polarizations in the waveguide have different propagation characteristics due to optical birefringence in the fiber. This effect causes random temporal spreading of the signal, which leads to intersymbol interference, and thus an increased bit error rate. Polarization mode dispersion can be compensated if the state of polarization of the signal in the receiver estimated and transformed back to the state at the transmitter. The state of polarization is the relationship between the two vector components of the optical signal. If S_x and S_y are the complex representations of the modulated signal in the parallel and perpendicular polarization states respectively, then $\mathbf{S} = (S_x, S_y)$ is called the Jones vector representation of the polarization state.

In the transmitter, two synchronous data streams are modulated using orthogonal polarization components. The modulation format is arbitrary, but we will focus on quadrature phase shift keying. After transmission through fiber, the polarization state is transformed by random refringence effects within the fiber. If this transformation is linear, then polarization components of the received signal $\mathbf{E} = (E_x, E_y)$ are a linear combination of the original signals in two orthogonal polarization states. The output electrical field can then be related to the input electrical field by

$$\begin{aligned} \begin{bmatrix} E_x \\ E_y \end{bmatrix} &= l \begin{bmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{bmatrix} \begin{bmatrix} S_x \\ S_y \end{bmatrix} \\ &= lJ \begin{bmatrix} S_x \\ S_y \end{bmatrix}, \end{aligned} \tag{5.1}$$

where l is a real scalar describing the optical loss from the input to the output, and the polarization transformation is described by a unitary Jones matrix J. The random birefringence in the fiber causes the Jones matrix to be time varying.

Before the advent of DSP solutions, polarization demultiplexing was accomplished by using dynamic polarization controllers and polarization beam splitters. Recently, DSP approaches have been applied to polarization demultiplexing. The development of the techniques have been aided by an analogy between polarization division multiplexing and multiple-input multiple-output (MIMO) antenna wireless communications. As a result, algorithms for wireless MIMO channel estimation can be readily applied to polarization demultiplexing in optical polarization MIMO.

Ignoring the constant loss term l in (5.1), since J is a unitary matrix, this MIMO system, in theory, can transmit two synchronous channels without any penalty [88]. Because of environment variations, the polarization of the light wave in fiber generally drifts with time. The rate of this polarization drift is generally much slower than the transmission data rate. One way to estimate the Jones matrix J for the entire frame is by using a training sequence in the preamble of each frame to remove polarization cross-talk. Various channel estimation algorithms can be used to estimate J. Considering the high data rate used in optical communications, the least mean squares algorithm was chosen in [89] for its simplicity.

Estimation of the Jones matrix without relying on a training sequence is possible using the statistical properties of the transmitted symbols. This method is called blind estimation. For a coherent system with a constant-intensity modulation format such as QPSK, the Jones matrix can be estimated by constraining the modulus of the received signal to be constant. This method is called constant modulus algorithm (CMA) [83, 84, 85]. Without loss of generality, we assume that the constant modulus is unity. An estimate of the Jones matrix is obtained by minimizing the mean squared errors $\epsilon_x^2 = \mathcal{E}\left\{|1 - E_x^2|^2\right\}$ and $\epsilon_y^2 = \mathcal{E}\left\{|1 - E_y^2|^2\right\}$, where $\mathcal{E}$ denotes an average in time. The algorithm then works by forcing the gradient of the mean squared error with respect to the appropriate elements of the Jones matrix to be equal to zero

$$\frac{\partial \epsilon_x}{\partial J_{xx}} = 0, \qquad \frac{\partial \epsilon_x}{\partial J_{xy}} = 0, \qquad \frac{\partial \epsilon_y}{\partial J_{yx}} = 0, \qquad \frac{\partial \epsilon_y}{\partial J_{yy}} = 0.$$

To determine the optimal coefficients, the gradients are replaced by their instantaneous values. The algorithm results in an adaptive evaluation of matrix elements given in (5.1) with a convergence parameter μ

$$J_{xx} \rightarrow J_{xx} + \mu\epsilon_x \hat{S}_x.E'_x$$

$$J_{xy} \rightarrow J_{xy} + \mu\epsilon_x \hat{S}_x.E'_y$$

$$J_{yx} \rightarrow J_{yx} + \mu\epsilon_y \hat{S}_y.E'_x$$

$$J_{yy} \rightarrow J_{yy} + \mu\epsilon_y \hat{S}_y.E'_y$$

where, E' denotes the complex conjugate of E, and $\hat{S}$ is the demultiplexed signal.

Compensation of Kerr Nonlinearities and Chromatic Dispersion

The backpropagation method [27, 86] is an effective method to compensate the nonlinear propagation effects. In this method, the received signal is propagated through a simulated inverse channel. The channel is modeled using nonlinear Schrödinger equation (NLSE) as in (1.1). For a single polarization, single channel system, the equation used for the inverse propagation may be written as

$$\frac{\partial E}{\partial z} = (\mathcal{D} + \mathcal{N})E,$$

where E is the received optical signal, $\mathcal{D}$ is the differential operator for dispersion compensation and attenuation, and $\mathcal{N}$ is the nonlinear operator. These two operators are given by

$$\mathcal{D} = -\frac{j}{2}\beta_2\frac{\partial^2}{\partial t^2} + \frac{\alpha}{2}$$

$$\mathcal{N} = -j\gamma|E|^2$$

where α is the attenuation factor, β_2 is the group velocity dispersion parameter, and γ is the nonlinearity parameter. When these parameters are chosen to be exactly the negative of the values for the transmission fiber, the nonlinearity and the linear dispersion can be compensated through backward propagation.

The nonlinear Schrödinger equation is most commonly solved by using the split-step Fourier method [7]. The split-step Fourier method obtains an approximate solution by assuming that over a small distance h, the dispersive and nonlinear effects are decoupled. The propagation from z to $z+h$ is then treated in two steps. In the first step, it is assumed that operator $\mathcal{D}=0$ and only nonlinearity acts on the signal. In the second step, the assumption is that $\mathcal{N}=0$ and only dispersion is present. Combining these two steps, we can formally write the solution as

$$E(z+h,t)=\exp(h\mathcal{D})\exp(h\mathcal{N})E(z,t). \tag{5.2}$$

In this approach, each span of fiber is divided into sections of length h, and $\mathcal{N}$ and $\mathcal{D}$ are implemented sequentially

$$E_n(z+h,t)=\exp\left(-jh\gamma|E(z,t)|^2\right)E(z,t), \tag{5.3}$$

and

$$\mathcal{F}\{E\}(z+h,\omega)=\exp\left[-jh\left(\frac{\beta_2}{2}\omega^2+\frac{\alpha}{2}\right)\right]\mathcal{F}\{E_n\}(z+h,\omega), \tag{5.4}$$

where $\mathcal{F}\{E\}$ is the Fourier transform of E and ω is the angular frequency. In (5.3), the nonlinearity is compensated by applying the opposite sign of the power dependent phase shift. In (5.4), the compensator is the inverse of the quadratic phase all-pass filter that models the effect of dispersion. In most cases it is more efficient to implement the first step in time domain and the second step in frequency domain and use fast Fourier transform/inverse fast Fourier transform (FFT/IFFT) to transform between the time and the frequency domain.

The step size requirement for modeling the systems has been studied in [90]. For single channel detection however, it is not necessary to compute the nonlinear Schrödinger equation to a high degree of accuracy. The step size must be sufficiently small so that numerical errors are small compared to the impact of additive white Gaussian noise. The step size is related to the strength of the Kerr nonlinearity and the number of spans of fiber that include amplifiers. In almost all cases, we consider one step per span produces reasonable accuracy for a single channel system.

In a single channel system with polarization division multiplexing, the dispersion compensation can be applied to each polarization seperately. For nonlinear phase shift compensation (5.3) will change to [87]

$$\begin{aligned} E_{X,n}(z+h,t) &= \exp\left(-jh\left(a|E_X(z,t)|^2 + b|E_Y(z,t)|^2\right)\right) E_X(z,t), \\ E_{Y,n}(z+h,t) &= \exp\left(-jh\left(b|E_X(z,t)|^2 + a|E_Y(z,t)|^2\right)\right) E_Y(z,t). \end{aligned}$$

Usually, parameters a and b are calculated by exhaustive search. Figure 5.4 depicts the block diagram of this equalizer.

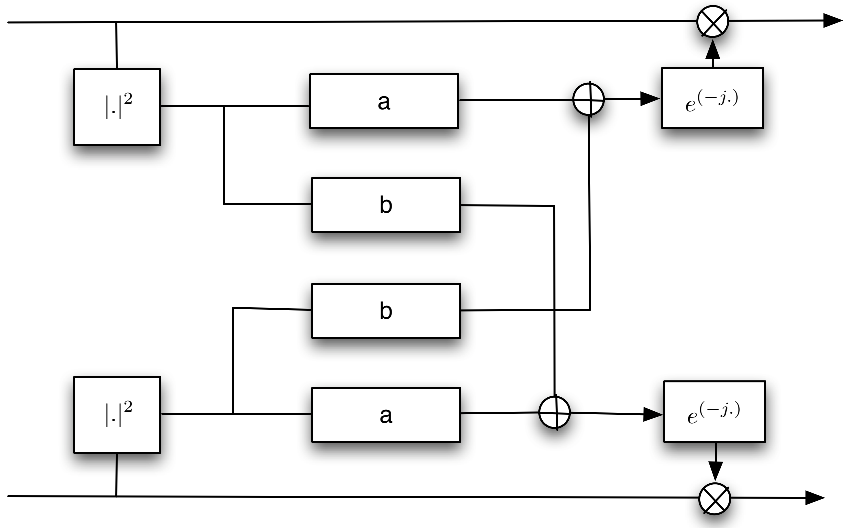


Figure 5.4: Polarization multiplexed nonlinear equalizer.

Digital Carrier-Phase Estimation

In high symbol rates, delays allowed in phase-locked loops in optical domain are small, and make them impractical [91]. Therefore, DSP-based phase estimation is the best choice. For a M-ary phase shift keying (MPSK) signal, most common method for carrier-phase estimation is to remove the data from signal by applying a power operator to the signal[92, 93]. Consider a quadrature phase-shift keying (QPSK) modulation. In the absence of noise and nonlinearities, the received signal can be represented as

$$E(t) = A\exp\{j[\theta_s(t) + \theta_c(t)]\},$$

where $\theta_c(t)$ is the optical carrier phase. This phase is the difference between the transmitter phase and the local oscillator phase. The term $\theta_s(t)$ is the data phase which takes four values $0, \pm\pi/2, \pi$ for quadrature phase shift keying. When the received signal is raised to the fourth power as shown in Figure 5.5, we obtain

$$E^4(t) = A^4 \exp\{j[4\theta_s(t) + 4\theta_c(t)]\} = A^4 \exp\{j4\theta_c(t)\},$$

because $\exp(j4\theta_s(t)) = 1$, i.e., the power operation removes the data leaving a term that is related to the phase difference between the signal and the local oscillator. The carrier phase can be computed and subtracted from the phase of the received signal to recover the data phase.

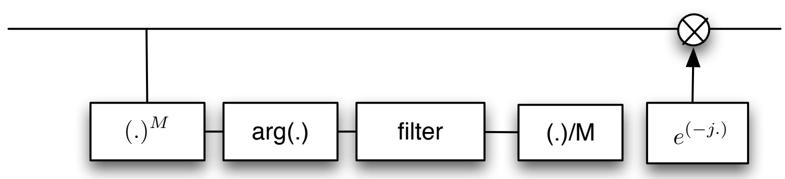


Figure 5.5: Block diagram of the digital carrier phase noise estimation.

5.3 Results

In this section, we applied the Viterbi algorithm to compensate the residual Kerr nonlinearity and the linear polarization cross-talk. We compared the resulting bit error rate, for threshold detection, and two-user Viterbi algorithm. The multiuser Viterbi algorithm is similar to the algorithm described in the last chapter. The only difference is that we consider the two polarization as the two users. In fact, we assume that in addition of compensation of nonlinearity, the effect of nonlinearity interference of two the polarizations on each other can be compensated using this detector. We used Viterbi algorithm with the memory length of 2 bits and 3 bits. The corresponding Viterbi algorithm have 16 states and 64 states, respectively. We repeat our simulation for different data transmission rates and launch power, to show the effect of our receiver over different regimes of dispersion and nonlinearity.

In the nonlinear regime, we observe that in the first few kilometers of fiber the nonlinearity has a strong effect on the signal. After this initial phase modulation, the power of the signal decreases and the linear chromatic dispersion has the strongest effect on the signal. As the dispersion compensation fiber, or any other linear equalization will not compensate for the nonlinearity, compensating the total span of the fiber will not result in the best bit error rate at the receiver. We therefore determine the optimum length of the compensation that will improve the bit error rate at the receiver in the presence of both linear and nonlinear effects. Finding the optimum value for the uncompensated length decreases the bit error rate without any additional complexity.

Figures 5.6-5.11, illustrate the bit error rate improvement versus the uncompensated length, L_c, of fiber with respect to the fully compensated system. In this setup, we just considered one span of fiber with the length of 60 km. Examining Figures 5.6, 5.9, and 5.9, when the dispersion is weak, i.e., $L_D/L = 5.3$, using a linear term that completely compensates for the span is not optimum. In all cases, $L_c = 15$ km. This is true for both threshold detection and the Viterbi algorithm. The improvement in the error rate of the threshold detection for these cases are between half to one order of magnitude. For 16-state Viterbi algorithm, when nonlinearity is weak to moderate, i.e., Figures 5.6 and 5.9, the improvement is about one order of magnitude, while for the case of high nonlinearity, Figure 5.9, the improvement is minimal. When dispersion is moderate, i.e., $L_D/L = 1.33$, in Figures 5.7, 5.10, and 5.13, the optimum uncompensated length, L_c, is between 2.5 km to 10 km. As the nonlinearity increases, L_c increases. However, the improvement of bit error rate is not significant with respect to the compensated case. For high dispersion, the optimum uncompensated length is 0 and the bit error rate increases rapidly, by increasing the uncompensated length.

5.4 Summary and Conclusion

In this chapter we showed that, Viterbi algorithm improves the performance by compensating the uncompensated residual effects of dispersion, nonlinearity, and polarization mode dispersion. Also, we showed that by optimizing dispersion compensation, the performance of the system can improve. Therefore, these improvement specially at high power help increase the transmitted power and transmission length of the system removing the need for optical amplifiers.

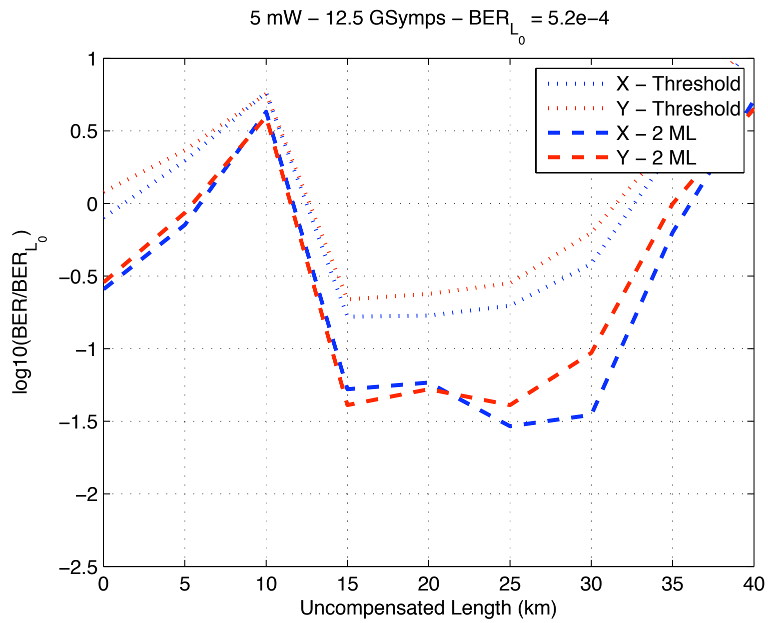


Figure 5.6: Lauch power = 5 mW, symbol rate = 12.5 Gsymps, bit rate = 50 Gbps $L_D/L_{NL} = 320km/154km = 2.07$, $L_D/L = 320km/60km = 5.3$, $L_{eff}/L_{NL} = 19km/154km = 0.12$; **weak dispersion, weak nonlinearity.**

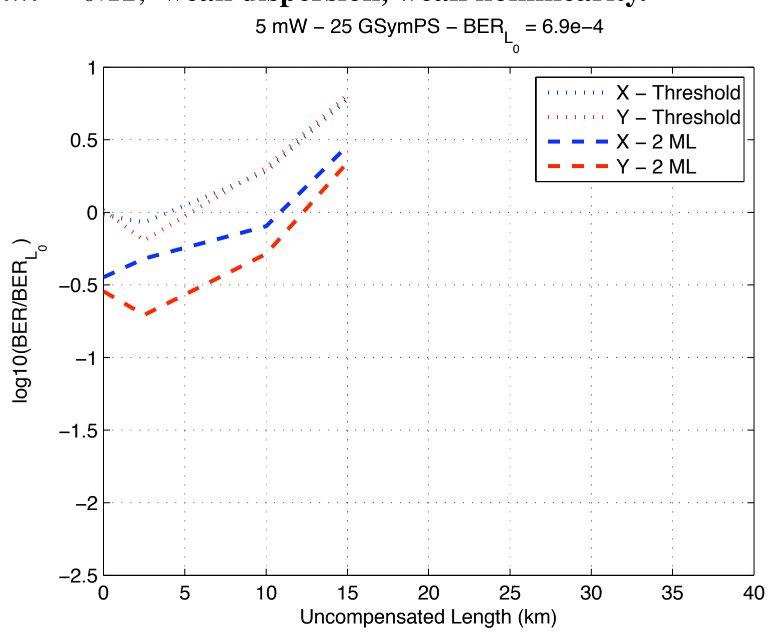


Figure 5.7: Lauch power = 5 mW, symbol rate = 25 Gsymps, bit rate = 100 Gbps $L_D/L_{NL} = 80km/154km = 0.52$, $L_D/L = 80km/60km = 1.33$, $L_{eff}/L_{NL} = 19km/154km = 0.12$; **moderate dispersion, weak nonlinearity.**

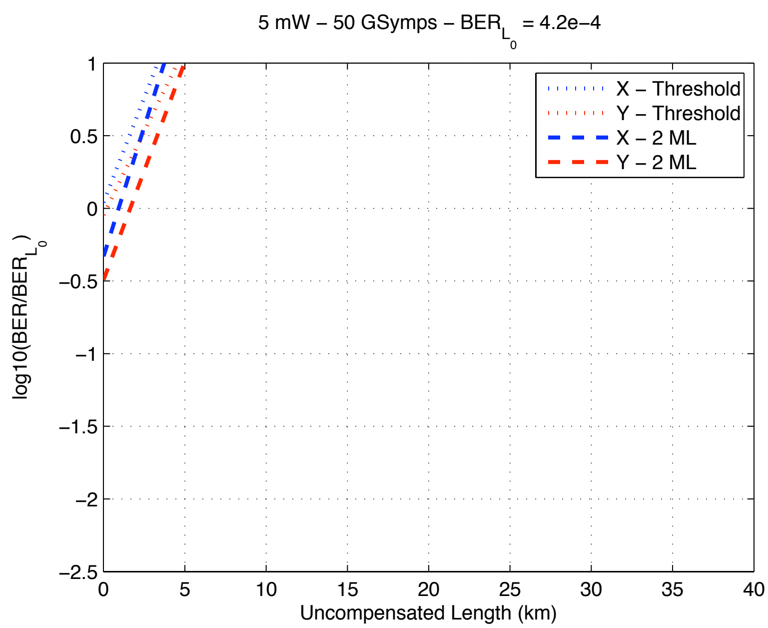


Figure 5.8: LauchPower = 5 mW, symbol rate = 50 Gsymps, bit rate = 200 Gbps $L_D/L_{NL} = 20km/154km = 0.13$, $L_D/L = 20km/60km = 0.33$, $L_{eff}/L_{NL} = 19km/154km = 0.12$; **high dispersion, weak nonlinearity.**

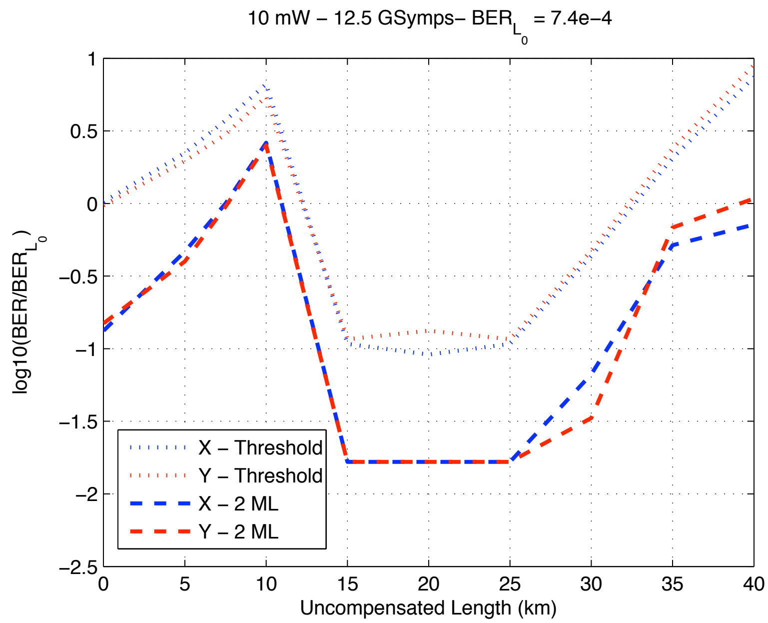


Figure 5.9: Lauch power = 10 mW, symbol rate = 12.5 Gsymps, bit rate = 50 Gbps $L_D/L_{NL} = 320km/77km = 4.16$, $L_D/L = 320km/60km = 5.3$, $L_{eff}/L_{NL} = 19km/77km = 0.25$; **weak dispersion, moderate nonlinearity.**

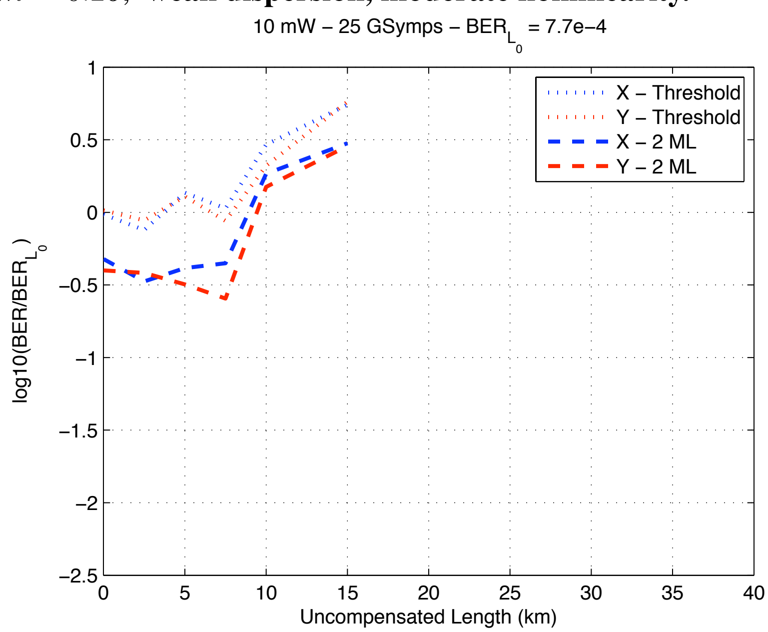


Figure 5.10: Lauch power = 10 mW, symbol rate = 25 Gsymps, bit rate = 100 Gbps $L_D/L_{NL} = 80km/77km = 1.04$, $L_D/L = 80km/60km = 1.33$, $L_{eff}/L_{NL} = 19km/77km = 0.25$; **moderate dispersion, moderate nonlinearity.**

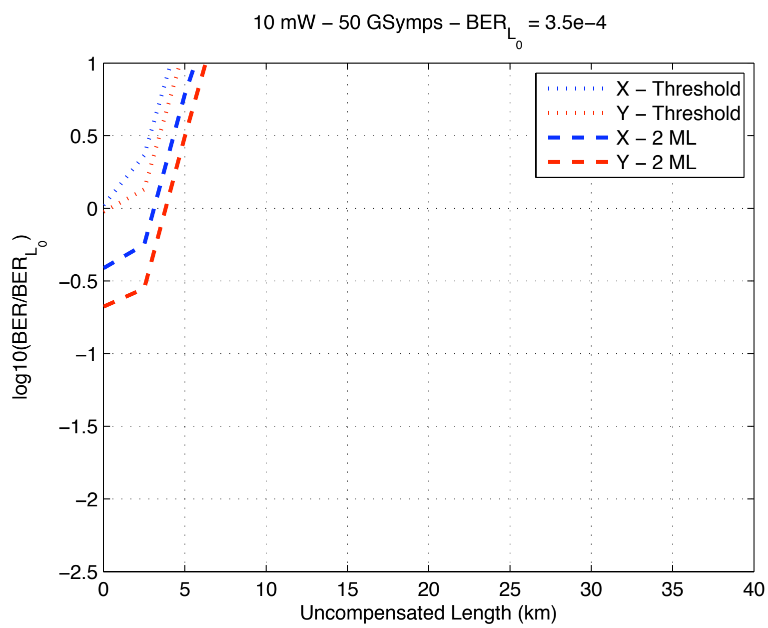


Figure 5.11: Lauch power = 10 mW, symbol rate = 50 Gsymps, bit rate = 200 Gbps $L_D/L_{NL} = 20km/77km = 0.26$, $L_D/L = 20km/60km = 0.33$, $L_{eff}/L_{NL} = 19km/77km = 0.25$; **high dispersion, moderate nonlinearity.**

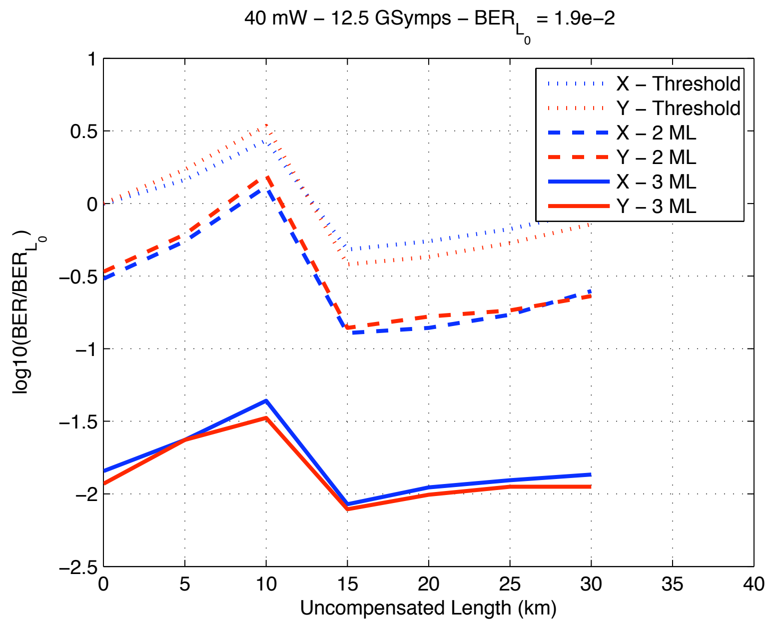


Figure 5.12: Lauch power = 40 mW, symbol rate = 12.5 Gsymps, bit rate = 50 Gbps $L_D/L_{NL} = 320km/19km = 16.8$, $L_D/L = 320km/60km = 5.3$, $L_{eff}/L_{NL} = 19km/19km = 1$; **weak dispersion, high nonlinearity.**

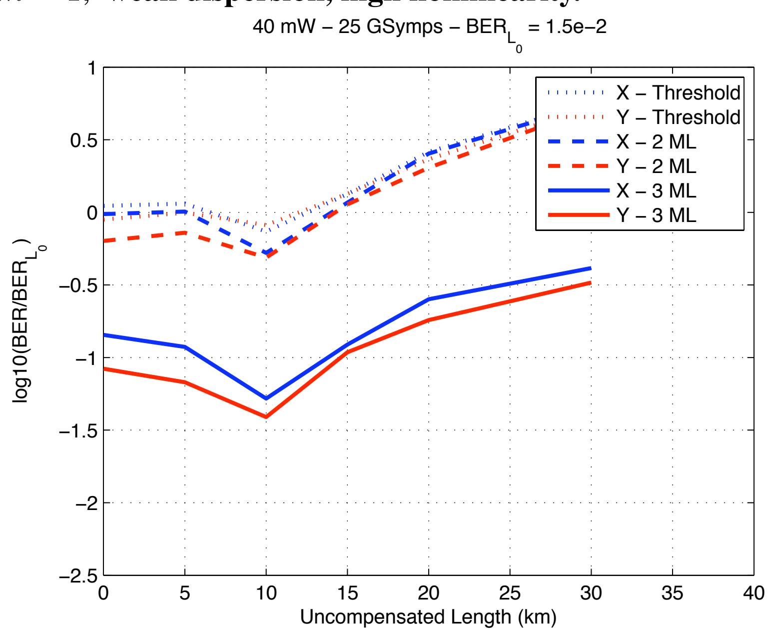


Figure 5.13: Lauch power = 40 mW, symbol rate = 25 Gsymps, bit rate = 100 Gbps $L_D/L_{NL} = 80km/19km = 4.2$, $L_D/L = 80km/60km = 1.33$, $L_{eff}/L_{NL} = 19km/19km = 1$; **moderate dispersion, high nonlinearity.**

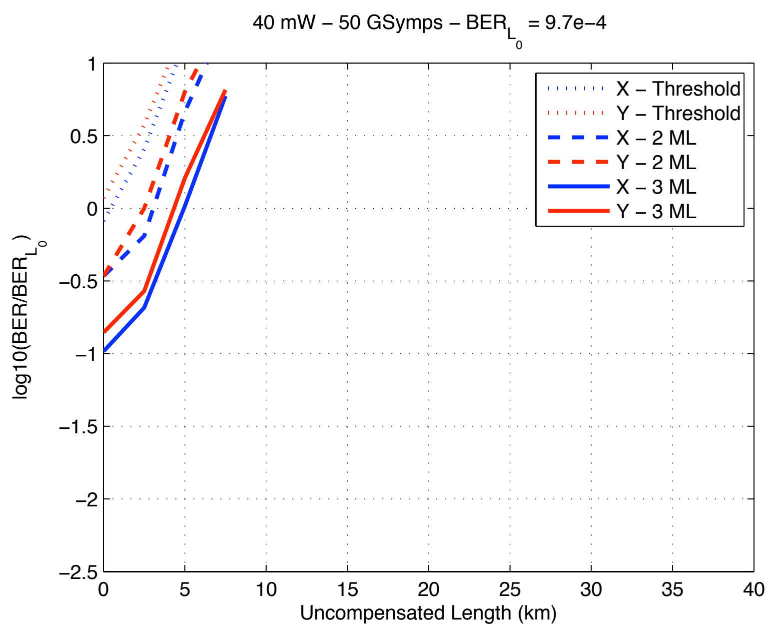


Figure 5.14: Lauch power = 40 mW, symbol rate = 50 Gsymps, bit rate = 200 Gbps $L_D/L_{NL} = 20km/19km = 1.05$, $L_D/L = 20km/60km = 0.33$, $L_{eff}/L_{NL} = 19km/19km = 1$; **high dispersion, high nonlinearity.**

6 Polynomial Fitted Equalizers for Nonlinear QPSK Modulated Channel

6.1 Introduction

The transmission rate in an optical fiber is mainly dependent on two factors: (1) compensation of impairments, and (2) high speed sampling devices. As described in last chapters, there are linear and nonlinear fiber impairments. Compensation of these impairments depends on available computational power and also on the compensation method. Therefore, low-complexity, high-performance compensation of impairments, particularly when the nonlinearity is high, is a fundamental research topic that has been addressed in the last decade. Backpropagation is a successful, but complex method that can compensate for the nonlinear impairments in the channel [86, 94]. Backpropagation solves an inverse nonlinear Schrödinger equation (NLSE) to estimate the transmitted signal. The complexity of this algorithm makes backpropagation currently inapplicable in current systems. We have described this method in detail in Section 5.2.2. Decreas-

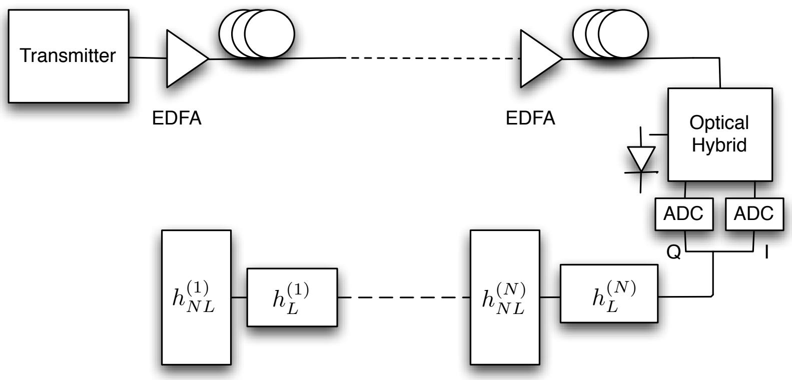


Figure 6.1: Block Diagram of Nonlinear Equalizer.

ing the complexity while maintaining the performance requires new design methods to reduce the required sampling rate of the signal. This sampling rate is major limitation in the state-of-the-art systems that process 25 giga symbols per second. Another approach is to find new methods to reduce the number of computational steps. It is known that for many practical systems, one compensation step per fiber span is sufficient to achieve near optimum performance [26]. In [95], the authors suggest using a filter for the nonlinear phase shift to avoid aliasing. This aliasing occurs because the nonlinear product terms contain frequency component higher than the fundamental data rate. Using these filters, they showed improvement in the performance of the detection using two samples per symbol. In the last chapter, we used maximum-likelihood estimation to mitigate the residual effects of linear dispersion, Kerr nonlinearity, and polarization mode dispersion after backpropagation and polarization demultiplexing methods. In this chapter, we investigate improving the performance of each block of backpropagation.

The use of FFT/IFFT transform methods for the linear compensation block means that the computational complexity is mostly decoupled from the duration of the impulse response. Therefore, the required number of computations is approximately proportional to the number of steps used and is also proportional to the oversampling factor. To reduce the computational complexity for the backpropagation method, the step size should increase and/or the sampling rate should decrease. In this chapter, we replace the analytical functions used in the last chapter that were based on the solution of the nonlinear Schödinger equation, with fitted curves. The resulting equalization algorithm is more efficient with respect to the receiver sampling rate and launch power. The design goal is to achieve the minimum output noise, i.e., maximum signal-to-noise ratio. In the next section, we describe this method and apply it to a system using the quadrature phase shift keying modulation. For this design, we only considered the signal transmitted in one polarization to avoid interaction of chromatic dispersion and fiber nonlinearity with the polarization mode dispersion.

6.2 Channel Model

The channel was numerically simulated using VPItransmissionMaker V8.5. We assume the fiber channel consists of N fiber spans each of which includes an optical erbium doped fiber amplifier (EDFA) to compensate for the signal attenuation. The transmission symbol rate is 25 Gsymbol/s (equivalent to 50 Gbps). Our modulation is quadrature phase shift keying (QPSK) and the receiver is coherent. The quadrature phase shift keying signal was generated by driving each arm of a Mach-Zehnder modulator (MZM) with a pseudorandom bit generator. The signal from the two modulators where then quadrature multiplexed giving a total bit rate of 50 Gbps. We used 80 km single-mode fiber (SMF) links for each span with no fiber dispersion compensation. The fiber is assumed to have an attenuation of 0.2 dB/km, a dispersion of 16 ps/nm/km, and nonlinearity factor of 1.3 /km/W.

A typical coherent receiver, which contains an optical local oscillator, optical hybrid to adjust the relative phase, and the two pairs of balanced photodiodes is used. After the receiver, the signal is passed through an integrating filter and an analog to digital converter (ADC) which samples the complex QPSK symbol stream at a rate of 50 Gsample/s. In this chapter, we assume to have only backpropagation equalizer and ignore the equalizers for other effects.

6.3 Discrete Signal Processing at the Receiver

For each fiber span, we design one linear equalization block and one nonlinear equalization block, denoted by $h_L^{(n)}$ and $h_{NL}^{(n)}$, respectively, for the nth block. Figure 6.1 shows a diagram of the compensation method. In the backpropagation method, the linear equalizer $h_L^{(n)}$ is the sampled quadratic-phase finite impulse response (FIR) dispersion compensation (DC) filter. The nonlinear equalizer $h_{NL}^{(n)}$ is the nonlinear phase estimation. In this chapter, we replace both with polynomial filters fitted to the linear and nonlinear response of the channel.

Throughout this chapter, we denote the input and output signals of the nth fiber

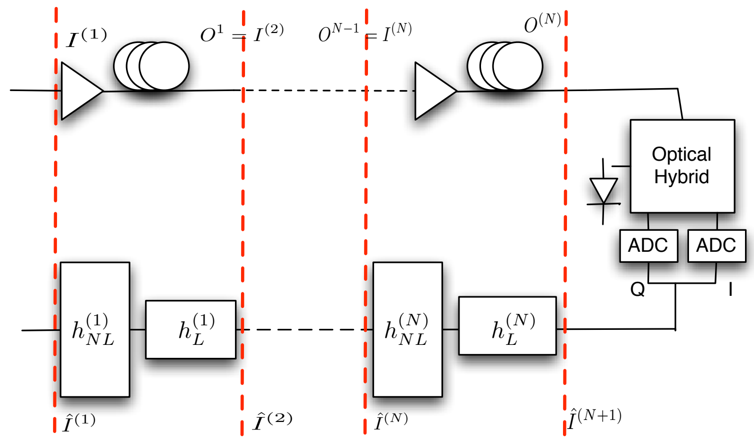


Figure 6.2: Fiber channel model and backpropagation block diagram.

span by $I^{(n)}$ and $O^{(n)}$, respectively. Also, we denote the ith sample of the input and output signals by I_i and O_i. For a backpropagation method with N steps of compensation, the input of the function $h_L^{(N)}$ is a vector of samples at the output of fiber span N, i.e., $O^{(N)}$. The output of this functional block is the input of the next functional block $h_{NL}^{(N)}$. The output of $h_{NL}^{(N)}$ is assumed to be the estimated input sample of the corresponding fiber span, i.e., $\hat{I}^{(N)}$. For $n < N$, the equalization sequence is as follows: $\hat{I}^{(n)} = h_{NL}^{(n)}(\hat{I}_L^{(n)})$, where $\hat{I}_L^{(n)} = h_L^{(n)}(\hat{I}^{(n+1)})$. After N steps we generate $\hat{I}^{(1)}$. We compare the performance of different realizations of backpropagation using different types of equalizers by their output signal-to-noise ratio (SNR), i.e., the SNR of $\hat{I}^{(1)}$.

To calculate the polynomial equalizer, we use known transmitted and received signal sequence for training. We design one block of compensation for each fiber span based on the output of all spans of fiber, $I^{(n)}$ for the known sequence. These values can be measured in a laboratory or can be calculated by simulation. Using these values, the filter can be calculated and applied to the output of the real fiber link. We used a MATLAB function 'lsqcurvefit' which solves nonlinear curve-fitting (data-fitting) problems to calculate coefficients of functions h's. 'lsqcurvefit' uses the trust region reflective optimization method to minimize square error of the output of the fitted filter. Figure 6.3, illustrates the block diagram of the curve fitting process. In the first step, we use the the input and output of the span n, to calculate the optimum coefficients for $h_L^{(n)}$ to minimize

$$\min_{h_L^{(n)}} E\{|I^{(n)} - h_L^{(n)}(O^{(n)})|^2\}.$$

After calculating the linear function, the nonlinear function is estimated as

$$\min_{h_{NL}^{(n)}} E\{|I^{(n)} - h_{NL}^{(n)}(\hat{I}_L^{(n)})|^2\},$$

where $\hat{I}_L^{(n)} = h_L^{(n)}(O^{(n)})$. We choose a polynomial form for both h_L and h_{NL}. We now discuss the specific form for each equalizer.

6.4 Linear Equalizer

In the backpropagation method, the linear part of the equalizer is an all-pass quadratic-phase, infinite impulse response (IIR) filter. To use it in a practical system, it should be truncated and sampled. If the signal is undersampled, then the sampled filter does not compensate the signal properly. In this case, calculating the function based on an optimization method can help to reduce the effect from both linear aliasing and nonlinearity. Since different spans of fiber are equal, we might expect that fitted response

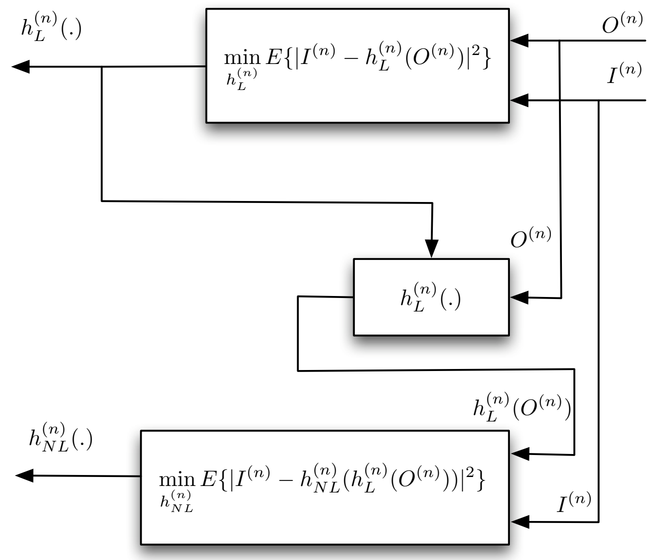


Figure 6.3: Block diagram of the calculation of the equalizers coefficients.

be the same for each span. However, if the degrees of freedom for the optimization of the equalizer is not large enough, then different spans can have different equalizer coefficients. In fact, in this case the equalizer for each span depends on the specific input. In this paper, we assumed that the number of taps of linear equalizer is 61. The linear dispersion compensation filter was implemented both in frequency domain with 2^{16} taps and in time domain with 61 taps. The performance of both these equalizers are approximately the same.

Figure 6.4 shows the result for using only linear equalizer for the compensation of channel. The fiber length is set to 10 spans of 80 km for a total length of 800 km, and the launch power is varied from -14 dBm to 7 dBm. The y-axis of the plot is the signal-to-noise ratio of the signal after passing through the set of linear equalizers. We assume that 10 blocks of linear equalizations with 61 taps are used. In this plot four different equalizers are compared: a one sample per symbol (1SPS) analytical linear quadratic-phase dispersion compensation, two samples per symbol (2SPS) analytical equalizer, and 1SPS and 2SPS fitted linear equalizers. Examining the figures, it can be seen that if a sampling rate of 1SPS is used and linear aliasing occurs, then the fitted linear equalizer is better than the analytical truncated finite impulse response equalizer by about 1 dB. Conversely, when two samples per symbol analytical equalizer is used, it works slightly better than the fitted linear equalizer by less than 1 dB. It can also be seen that as the launch power increases, the difference between the two graphs decreases because in that range of power, nonlinear effects are dominant and cannot be compensated by a linear technique. For this regime the two samples per symbol fitted linear equalizer is worse than the analytical equalizer. We believe this is because the number of training bits is not large enough. As a result, the calculated curve fit is biased to specific transmitted sequence instead of a more general sequence.

Figure 6.5 plots the absolute value of the coefficients of the linear equalizer for a -2 dBm launch power. The upper curve plots the first five spans while the lower plot shows that last five spans. In this figure, we can see that the coefficients for spans $n = 3, \ldots, 10$ contain local maxima that change as n increases. These peaks are indistinguishable from the main peak for the first two spans when $n = 1, 2$. After the second span, a local maximum appears. For $n = 3$, this occurs for a tap index of 5. For each subsequent span, the index where the peak occurs shifts. The value of this peak is smaller than the coefficients of taps indices 0, -1, and 1, but it is larger than all others. For filters $h_L^{(n)}$, $n = 3, \ldots, 10$, these peaks happen at indices $\pm p$ for p equal to 5, 5, 8,

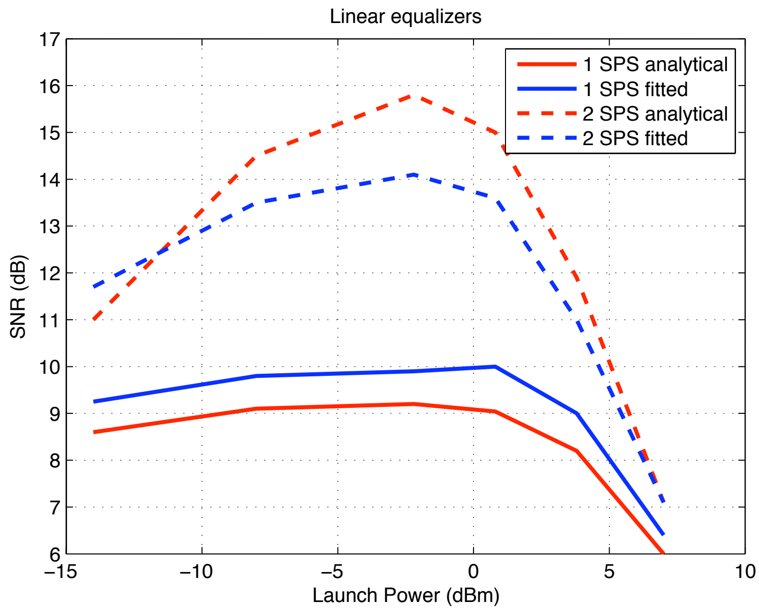
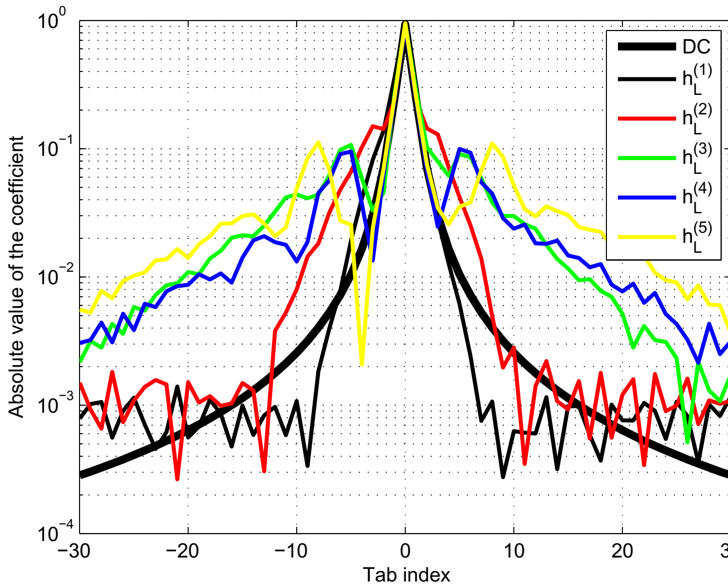


Figure 6.4: Comparison of the performance of the analytical linear equalizer for dispersion compensation and the fitted linear equalizer.

8, 11, 13, 13, 16, respectively. This suggests that in the 10th span, the signal has walked off 16 symbols. Therefore, the signal walks off 1.6 symbols per span of fiber.

6.5 Nonlinear Equalizer

In traditional backpropagation, the nonlinear equalizer is a memoryless power dependent phase shift, which is assumed to be identical for all spans of fiber. In this chapter, we modify the functional form for nonlinear equalizer. We select a two-term equalizer of type $h_{NL}^{(n)}(x) = \alpha_1^{(n)} x + \alpha_2^{(n)} x|x|^2$, where $\alpha_1^{(n)}$ and $\alpha_2^{(n)}$ are complex coefficients. This form is similar to the self-phase modulation discussed earlier. Another interpretation of this equation is that it is the truncated Taylor series of the function



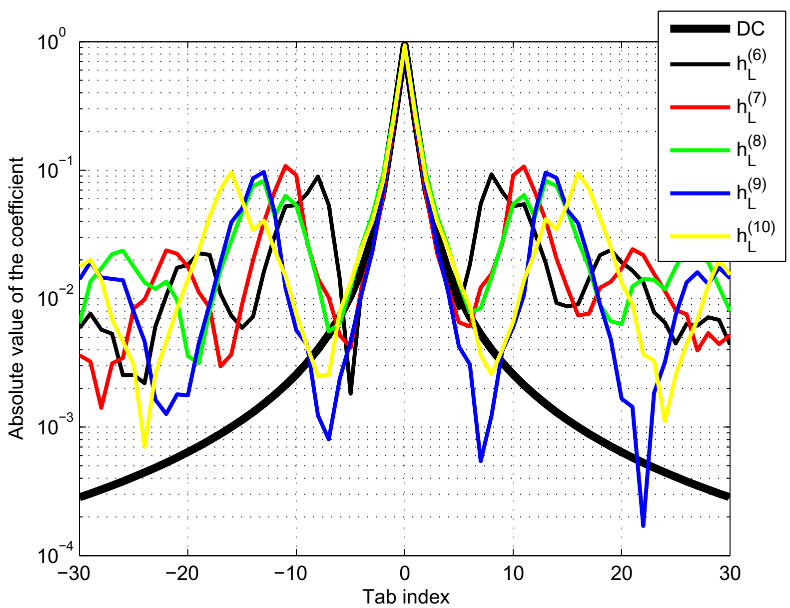


Figure 6.5: Amplitude of the coefficient of the polynomial of the fitted linear equalizer with the sample rate of one sample per symbol, calculated for equalizers $h_L^{(1)}, \ldots, h_L^{(10)}$, for the launch power of -2 dBm. DC stands for analytical equalizer.

$xe^{j\gamma|x|^2}$ which is the nonlinear phase shift equalizer.

Figure 6.6 shows a comparison of the analytical linear and nonlinear equalizers used in conventional backpropagation along with several fitted equalizers. In Figure 6.6(a) the sampling rate is one sample per symbol. As we discussed in the last section, we can see that the fitted linear equalizer outperforms the analytical linear equalizer. Also shown on the plot is the combination of the fitted nonlinear equalizer with both the fitted and analytical linear equalizers. The combination of the fitted linear and the nonlinear equalizers outperforms the combination of analytical linear equalizer and non-linear fitted equalizer by about 2 dB at a power level of 7 dBm.

In Figure 6.6(b), we compare the performance of the equalizers with the data sampled at a rate of 2 samples per symbol. In this figure, the best performance is for the combination of analytical linear equalizer and fitted nonlinear equalizer. Its improvement with respect to the analytical linear and nonlinear equalizers is between 3 and 4 dB. These results makes it clear that nonlinear compensation techniques must be tuned to the channel much more so than linear equalizers.

The coefficients of the fitted nonlinear equalizer, $h_{NL}^{(n)}(x) = \alpha_1^{(n)}x + \alpha_2^{(n)}x|x|^2$, has an interesting trend. The coefficients (α_1, α_2) for $n = 3, \ldots, 10$ are equal and α_2 is imaginary, but α_2 changes for $n = 1, 2$ to real values. This means that for $n = 1, 2$ these filters not only compensate for the nonlinear phase shift, but they also compensate for amplitude fluctuations. The largest improvement in the signal-to-noise ratio occurs when there is a decrease in the amplitude fluctuations. This compensation method is

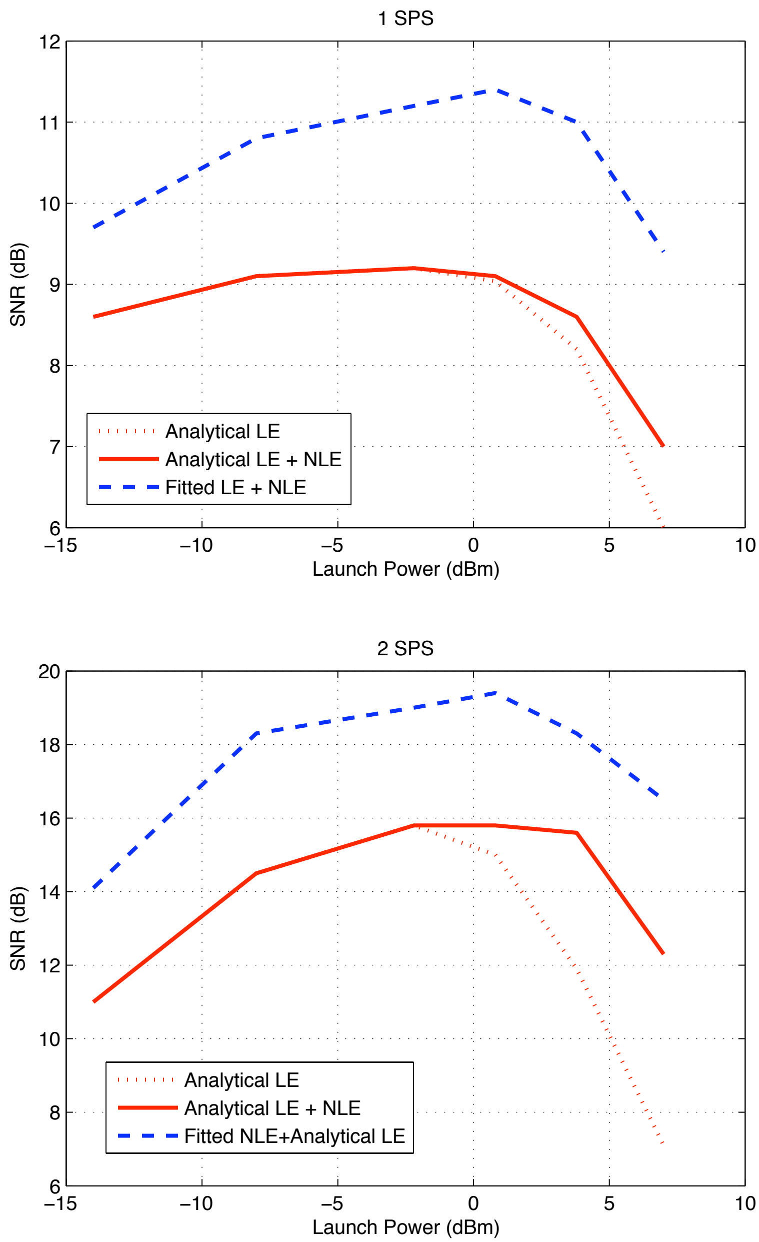


Figure 6.6: Comparison of the performance of an analytical backpropagation equalizer and the fitted linear plus nonlinear equalizers. LE stands for linear equalizer and NLE stands for nonlinear equalizer.

then effective when the systems used both amplitude and phase modulations as is the case for quadrature amplitude modulation.

6.6 Summary and Conclusion

Using a standard backpropagation algorithm as a baseline, we applied fitted linear and nonlinear equalizers instead of analytical equalizers. We showed that for a sampling rate of one sample per symbol, the combination of a fitted linear and a nonlinear equalizer, improves the signal-to-noise ratio by up to 2.5 dB. When signal is sampled at a rate of two samples per symbol, the combination of analytical linear equalizer and the fitted nonlinear equalizer improves the signal-to-noise ratio by up to 4 dB. This method shows that using fitted polynomials can improve the performance of backpropagation significantly without increasing the computational complexity.

7 Summary and Future Directions

The objective of this thesis was to develop techniques that increase the spectral efficiency and capacity of a fiber optic channel when the channel has nonlinear characteristics. To achieve that, we investigated different types of channels, modulations, and receivers. We used single-user detection in Chapters 2, 5, and 6, and wavelength division multiplexing in Chapters 3 and 4. The transmitted signals were modulated by intensity modulations and received by noncoherent receiver in Chapters 2-4 and modulated by quadrature phase shift keying and received by coherent receiver in Chapters 5 and 6. We also looked at polarization multiplexing modulation in Chapter 5.

We started with intensity modulation and a noncoherent receiver because it is widely deployed in current systems. We then considered coherent modulation because of the technology advancement in developing stable coherent receivers. Coherent modulation provides high spectral efficiency but is more expensive and complex relative to noncoherent modulation. Therefore, depending on the application and requirements, both types of systems will be used in the future.

As mentioned before in this thesis, there are three main impairments in optical fiber: (1) chromatic dispersion, (2) Kerr nonlinearity, and (3) Polarization mode dispersion. In Chapter 2, we used a modulation code with maximum-likelihood equalization to compensate for the chromatic dispersion for a nonlinear channel. We showed that our method improves the bit error rate by less than 1 dB, but improves the tolerance to the error in sampling position, i.e., jitter, by a factor of 2.

In Chapter 3, we proposed a method to design modulation codes to decrease the self-phase and cross-phase modulation interference power in a dense wavelength division multiplexing channel using intensity-modulation. We designed constrained codes assuming that the transmitter and receiver have access to only one user. The designed codes are based on a finite-state Markov chain, with the transition probabilities optimized so that the power of the cross-talk for the transmitted coded signal is minimized. We showed that the overall capacity as well as the interference variance improve. These results motivate the following future research: (1) solving more examples with different dispersion maps, (2) generalizing the optimization function, which is currently the cross-talk power in bits carrying signal one, to a complex function, (3) examining more complex modulation formats like 16-QAM, etc., (4) considering coherent receiver instead of noncoherent receiver, and (5) using multiuser coding and decoding.

In Chapter 4 multiuser maximum-likelihood sequence estimation was used to compensate the nonlinear cross-talk. We showed that in systems that are dominated by self-phase and cross-phase modulation, a two-user receiver improves the bit error rate. We also showed that it is hard to compensate for four-wave mixing especially when laser phase noise is present because of the nonlinear phase relationship between the channels.

One idea for future work is to estimate the phase before the receiver. This will help to estimate the four-wave mixing terms.

In Chapter 5, we used two-user maximum-likelihood estimation to compensate the residual effects of chromatic dispersion, Kerr nonlinearity, and polarization demultiplexing and polarization mode dispersion after estimating the signal by the backpropagation method. Depending on the value of dispersion and nonlinearity, we showed improvement in the bit error rate ranging from 0.5 dB to 1.5 dB. These results are for one span of fiber. Future works include the application to several spans.

In Chapter 6, we presented a method to improve the performance of the backpropagation method by designing a linear dispersion equalizer and nonlinear Kerr equalizer using curve fitting algorithms. We showed that for receivers using a sampling rate of one sample per symbol, the combined fitted linear and nonlinear equalizers improve the output signal to noise ratio by 2.5 dB. When the signal is sampled at two samples per symbol, the nonlinear equalizer improves the signal-to-noise ratio by near 4 dB. These equalizers are particularly suitable for more complex modulation formats such as QAM because they minimize signal to noise ratio in both phase and amplitude. This is a topic for future investigation.

7.1 Future Directions

General future directions of this research involve both code development and equalization techniques.

With respect to code development, both single-user and multiuser encoder and decoder may be useful for dense wavelength division multiplexing systems depending on the application. For example, consider a broadcast system in which there is one transmitter and several independent receivers, all sharing the same optical medium. The transmitter can encode all the transmitted signals such that the generated cross-talk in the shared medium is minimized. We call this a multiuser encoder. This code must be designed such that each individual receiver can decode without having access to the other users. This type of system may be useful for content distribution in a passive optical network (PON). Another example is when data is transmitted using more than one adjacent wavelength channel. In this case, both the transmitter and the receiver can process the information in all the channels. The transmitter can encode all of the signals together such that the generated cross-talk in the shared medium is minimized. At the receiver, the decoder uses all of these channels for signal detection. This is a multiuser encoder and decoder.

With respect to equalization, the methods presented in this thesis can be combined for a more general channel model. For example, a wavelength division multiplexing channel with polarization multiplexed quadrature phase modulation can use the methods presented in Chapters 4-6 to improve backpropagation in the presence of additional impairment such as polarization mode dispersion.

What is clear from current technology trends is that more power and thus stronger nonlinear effects will continue to be a significant issue for backbone fiber optic systems. Therefore, we expect the co-development of coding and equalization techniques for the nonlinear fiber optic channel to be an active research area for many years.

References

[1] N. Alic, G. C. Papen, S. Radic, and Y. Fainman, "Receiver Structure Trade-Offs in Equalized High-Speed Fiber-Optic Links," vol. 18, no. 17, pp. 1810 –2, Sep. 2006.

[2] [Online]. Available: http://www.fiber-optics.info/history/P1/

[3] "Alcatel-lucent bell labs announces new optical transmission record and breaks 100 petabit per second kilometer barrier." [Online]. Available: http://www.alcatel-lucent.com/wps/portal/newsreleases/detail?LMSG_CABI%NET=Docs_and_Resource_Ctr&LMSG_CONTENT_FILE=News_Releases_2009/News_Article_00%1797.xml&lu_lang_code=en

[4] S. Chandrasekhar and X. Liu, "Enabling Components for Future High-Speed Coherent Communication Systems," 2011, p. OMU5.

[5] S. Benedetto and E. Biglieri, *Principles of Digital Transmission*. Kluwer Academic Press, 1999.

[6] C. R. Giles and E. Desurvire, "Modeling Erbium-Doped Fiber Amplifiers," *J. Lightw. Tech.*, vol. 9, pp. 271–183, 1991.

[7] G. P. Agrawal, *Nonlinear Fiber Optics*. San Diego, CA: Academic Press, 2006, 4th ed.

[8] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for Digital Recorders," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2260 – 99, Oct. 1998.

[9] N. Alic, G. C. Papen, R. Saperstein, L. Milstein, and Y. Fainman, "Signal Statistics and Maximum Likelihood Sequence Estimation in Intensity Modulated Fiber Optic Links Containing a Single Optical Pre-amplifier," vol. 13, no. 12, pp. 4568 –79, Jun. 2005.

[10] N. Alic, "Information Processing for Improved Performance of Optical Networks," Ph.D. dissertation, Dept. Elect. and Comp. Eng., University of California, San Diego, CA, 2006.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 1991.

[12] P. Mitra and J. B. Stark, "Nonlinear Limits to the Information Capacity of Optical Fibre Communications," *Nature*, vol. 411, pp. 1027 – 30, 2001.

[13] N. Kashyap, P. H. Siegel, and A. Vardy, "Coding for the Optical Channel: the Ghost-Pulse Constraint," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 64 – 77, Jan. 2006.

[14] I. B. Djordjevic and B. Vasic, "Constrained Coding Techniques for the Suppression of Intrachannel Nonlinear Effects in High-Speed Optical Transmission," *J. Lightw. Tech.*, vol. 24, no. 1, pp. 411 – 9, Jan. 2006.

[15] V. Pechenkin and F. R. Kschischang, "Constrained Coding for Quasi-Linear Optical Data Transmission Systems," *J. Lightw. Tech.*, vol. 24, no. 12, pp. 4895 – 902, Dec. 2006.

[16] J. D. Downie, J. Hurley, and M. Sauer, "Behavior of MLSE-EDC With Self-Phase Modulation Limitations and Various Dispersion Levels in 10.7-Gb/s NRZ and Duobinary Signals," *IEEE Photon. Technol. Lett.*, vol. 19, no. 13, pp. 1017 – 19, Jul. 2007.

[17] M. H. Taghavi, G. C. Papen, and P. H. Siegel, "On the Multiuser Capacity of WDM in a Nonlinear Optical Fiber: Coherent Communication," *IEEE Trans. Inf. Theory*, vol. 19, no. 13, Nov. 2006.

[18] B. Xu and M. Brandt-Pearce, "Multiuser Detection for Square-law Receiver Under Gaussian Channel Noise With Applications to Fiber-Optic Communications," *Information Theory, IEEE Transactions on*, vol. 51, no. 7, pp. 2657 – 64, Jul. 2005.

[19] K. Ho and J. M. Kahn, "Channel Capacity of WDM Systems Using Constant-Intensity Modulation Formats," in *Proc. IEEE Optical Fiber Communications (OFC'02)*, Anaheim, CA, Mar. 2002, pp. 731 – 3.

[20] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379 – 423 and 623 – 656, Jul. and Oct. 1948.

[21] J. Geist, "Capacity and Cutoff Rate for Dense M-ary PSK Constellations," in *MILCOM '90*, Monterey, CA, USA, Sep./Oct. 1990, pp. 768 –70.

[22] K. P. Ho, "Exact Evaluation of the Capacity for Intensity-Modulated Direct-Detection Channels with Optical Amplifier Noises," *IEEE Photon. Technol. Lett.*, vol. 17, no. 4, pp. 858 –60, Apr. 2005.

[23] J. M. Kahn and K. Ho, "Spectral Efficiency Limits and Modulation/Detection Techniques for DWDM Systems," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 10, no. 2, pp. 259 – 72, Mar./Apr. 2004.

[24] C. Laperle, B. Villeneuve, Z. Zhang, D. McGhan, H. Sun, and M. O'Sullivan, "Wavelength division multiplexing (WDM) and polarization mode dispersion (PMD) performance of a coherent 40 Gbit/s dualpolarization quadrature phase shift keying (DP-QPSK) transceiver," in *Proc. OFC2007*, Anaheim, CA, 2007, p. PDP16.

[25] E. Ip and J. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Tech.*, vol. 26, no. 20, pp. 3416 –25, Oct. 2008.

[26] D. Millar, S. Makovejs, C. Behrens, S. Hellerbrand, R. Killey, P. Bayvel, and S. Savory, "Mitigation of fiber nonlinearity using a digital coherent receiver," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1217–26, Sept.-Oct. 2010.

[27] S. Savory, G. Gavioli, E. Torrengo, and P. Poggiolini, "Impact of interchannel nonlinearities on a split-step intrachannel nonlinear equalizer," *IEEE Photon. Technol. Lett.*, vol. 22, no. 10, pp. 673–675, May 2010.

[28] B. E. A. Saleh and M. I. Irshid, "Coherence and Intersymbol Interference in Digital Fiber Optic Communication Systems," *IEEE J. of Quantum Electron.*, vol. 18, no. 6, pp. 944–51, Jun. 1982.

[29] C. Fludger, T. Duthel, D. van den Borne, C. Schulien, E.-D. Schmidt, T. Wuth, E. de Man, G. Khoe, and H. de Waardt, "10 x 111 Gbit/s, 50 GHz Spaced, POLMUX-RZ-DQPSK Transmission over 2375 km Employing Coherent Equalisation," in *Proc. OFC/NFOEC'07*, Anaheim, CA, 2007, p. PDP22.

[30] S. Savory, G. Gavioli, R. Killey, and P. Bayvel, "Electronic Compensation of Chromatic Dispersion Using a Digital Coherent Receiver," *Opt. Express*, vol. 15, no. 5, pp. 2120–26, Mar. 2007.

[31] A. Lowery and J. Armstrong, "Orthogonal-Frequency-Division Multiplexing for Dispersion Compensation of Long-Haul Optical Systems," *Opt. Express*, vol. 14, no. 6, pp. 2079–84, Mar. 2006.

[32] I. B. Djordjevic and B. Vasic, "Orthogonal Frequency Division Multiplexing for High-speed Optical Transmission," *Opt. Express*, vol. 14, no. 9, pp. 3767–75, May 2006.

[33] W. Shieh and C. Athaudage, "Coherent Optical Orthogonal Frequency Division Multiplexing," *Electron. Lett.*, vol. 42, no. 10, pp. 587 – 9, May 2006.

[34] S. L. Jansen, I. Morita, N. Takeda, and H. Tanaka, "20-Gb/s OFDM Transmission Over 4,160-km SSMF Enabled by RF-Pilot Tone Phase Noise Compensation," in *Proc. OFC/NFOEC'07*, Anaheim, CA, 2007, p. PDP15.

[35] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[36] N. Alic, G. C. Papen, L. B. Milstein, P. H. Siegel, and Y. Fainman, "Performance Bounds of Maximum Likelihood Sequence Estimation in Intensity Modulated Fiber Optic Links," in *Proc. TIWC 2004*, Pisa, Italy, 2004, p. PDP15.

[37] J. H. Winters and R. D. Gitlin, "Electrical Signal Processing Techniques in Long-haul Fiber-Optic Systems," *IEEE Trans. Commun.*, vol. 38, no. 9, pp. 1439 – 53, Sep. 1990.

[38] J. H. Winters, R. D. Gitlin, and S. Kasturia, "Reducing the Effects of Transmission Impairments in Digital Fiber Optic Systems," *IEEE Commun. Mag.*, vol. 31, no. 6, pp. 68 –76, Jun. 1993.

[39] O. E. Agazzi, M. R. Hueda, H. S. Carrer, and D. E. Crivelli, "Maximum-Likelihood Sequence Estimation in Dispersive Optical Channels," *J. Lightw. Tech.*, vol. 23, no. 2, pp. 749 – 63, Feb. 2005.

[40] S. J. Savory, Y. Benlachtar, R. I. Killey, P. Bayvel, G. Bosco, P. Poggiolini, J. Prat, and M. Omella, "IMDD Transmission over 1,040 km of Standard Single-Mode Fiber at 10Gbit/s using a One-Sample-per-Bit Reduced-Complexity MLSE Receiver," in *Proc. OFC/NFOEC'07*, Anaheim, CA, 2007, p. OThK2.

[41] *IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks–Specific Requirements Part 3: Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*, IEEE 802.3-2005.

[42] *Fibre Channel Framing and Signaling - 3 (FC-FS-3)*, Project T11/1861-D Rev 0.30, 2007.

[43] *Data Interchange on Read-Only 120 mm Optical Data Disks (CD-ROM)*, Standard ECMA-130, 1996.

[44] *120 mm DVD - Read-Only Disk*, Standard ECMA-267, 2001.

[45] N. L. Swenson and J. M. Cioffi, "Sliding-block Line Codes to Increase Dispersion-limited Distance of Optical Fiber Channels," *IEEE J. Sel. Areas Comm.*, vol. 13, no. 3, pp. 485 – 98, Apr. 1995.

[46] R. Karabed and P. H. Siegel, "Coding for Higher Order Partial Response Channels," in *Proc. 1995 SPIE Int. Symp. Voice, Video, and Data Commun.*, vol. 2605, Philadelphia, PA, Oct. 1995, pp. 115 – 26.

[47] E. Soljanin, "On-Track and Off-Track Distance Properties of Class 4 Partial Response Channels," in *Proc. 1995 SPIE Int. Symp. Voice, Video, and Data Commun.*, vol. 2605, Philadelphia, PA, Oct. 1995, pp. 92–102.

[48] R. Karabed, P. H. Siegel, and E. Soljanin, "Constrained Coding for Binary Channels with High Intersymbol Interference," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1777 – 97, 1999.

[49] R. Gaudino, "Theoretical Limits for the Dispersion Limited Optical Channel," in *Optical Communication Theory and Techniques*, E. Forestieri, Ed. Springer US, 2005, pp. 29–36.

[50] A. Faerbert, S. Langenbach, N. Stojanovic, C. Dorschky, T. Kupfer, C. Schulien, J. Elbers, H. Wernz, H. Griesser, and C. Glingener, "Performance of a 10.7 Gb/s Receiver With Digital Equalizer Using Maximum Likelihood Sequence Estimation," in *ECOC 2004*, Stockholm, Sweden, Postdeadline paper PD, p. Th4.

[51] G. D. Forney, "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 3, pp. 363 – 78, May 1972.

[52] M. Keskinoz and B. V. K. V. Kumar, "Discrete Magnitude-Squared Channel Modeling, Equalization, and Detection for Volume Holographic Storage Channels," *Appl. Opt.*, vol. 43, no. 6, pp. 1368 – 78, Feb. 2004.

[53] D. E. Crivelli, H. S. Carrer, and M. R. Hueda, "On the Performance of Reduced-State Viterbi Receivers in IM/DD Optical Transmission Systems," in *ECOC 2004*, Stockholm, Sweden, p. We4.P.083.

[54] H. Bae, J. B. Ashbrook, J. Park, N. R. Shanbhag, A. C. Singer, and S. Chopra, "An MLSE Receiver for Electronic Dispersion Compensation of OC-192 Fiber Links," *IEEE J. Solid-Sate Circuits*, vol. 41, no. 11, pp. 2541 – 54, Nov. 2006.

[55] N. Alic, *Private communication*.

[56] S. A. Altekar, M. Berggren, B. E. Moision, P. H. Siegel, and J. K. Wolf, "Error-Event Characterization on Partial-Response Channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 241 – 7, Jan. 1999.

[57] *Series G: Transmission Systems and Media, Digital Systems and Networks - Digital Sections and Digital Line System - Optical Fibre Submarine Cable Systems - Forward Error Correction for High Bit-Rate DWDM submarine systems*, nternational Telecommunication Union Telecommunication Standardization Sector (ITU-T) Recommendation G.975.1, 2004.

[58] P. Lee, "Combined Error-Correcting/Modulation Recording Codes," Ph.D. dissertation, Dept. Elect. and Comp. Eng., University of California, San Diego, CA, 1988.

[59] P. Bender and J. K. Wolf, “A Universal Algorithm for Generating Optimal and Nearly Optimal Run-Length-Limited, Charge Constrained Binary Sequences,” in *Proc. 1993 IEEE Intl. Symp. on Inf. Theory*, San Antonio, TX, p. 6.

[60] Y. Sankarasubramaniam, “New Capacity-Approaching Codes for Run-Length-Limited Channels,” Ph.D. dissertation, School of Elec. and Comp. Eng., Georgia Inst. of Tech., Atlanta, GA, 2006.

[61] W. G. Bliss, “Circuitry for Performing Error Correction Calculations on Baseband Encoded Data to Eliminate Error Propagation,” *IBM Tech. Discl. Bull.*, vol. 23, pp. 4633 – 34, 1981.

[62] K. A. S. Immink, “A Practical Method for Approaching the Channel Capacity of Constrained Channels,” *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1389 – 99, Sep. 1997.

[63] J. L. Fan and A. R. Calderbank, “A Modified Concatenated Coding Scheme With Applications to Magnetic Data Storage,” *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1565 – 74, Jul. 1998.

[64] B. E. Moision, P. H. Siegel, and E. Soljanin, “Distance-Enhancing Codes for Digital Recording,” *IEEE Trans. Magn.*, vol. 34, no. 1, pp. 69 – 74, Jun. 1998.

[65] W. G. Bliss, “An 8/9 Rate Time-Varying Trellis Code for High Density Magnetic Recording,” *IEEE Trans. Magn.*, vol. 33, no. 5, pp. 2746 – 8, Sep. 1997.

[66] T. Foggi, E. Forestieri, G. Colavolpe, and G. Prati, “Maximum-Likelihood Sequence Detection With Closed-Form Metrics in OOK Optical Systems Impaired by GVD and PMD,” *J. Lightw. Tech.*, vol. 24, no. 8, pp. 3073 – 87, Aug. 2006.

[67] J. R. Barry, A. Kavcic, S. W. McLaughlin, A. Nayak, and W. Zeng, “Iterative Timing Recovery,” *IEEE Sig. Proc. Mag.*, vol. 21, pp. 89–102, Jan. 2004.

[68] W. Chung, W. A. Sethares, and C. R. Johnson, “Timing Phase Offset Recovery Based on Dispersion Minimization,” *IEEE Trans. on Sig. Proc.*, vol. 53, no. 3, pp. 1097 – 109, Mar. 2005.

[69] S. U. H. Qureshi, “Timing Recovery for Equalized Partial-Response Systems,” *IEEE Trans. Commun.*, vol. 24, no. 12, pp. 1326–30, Dec. 1976.

[70] G. C. Papen and R. E. Blahut, *Lightwave Communication Systems*, in preparation.

[71] K. V. Peddanarappagari and M. Brandt-Pearce, “Volterra Series Transfer Function of Single-Mode Fibers,” *J. Lightw. Tech.*, vol. 15, no. 12, pp. 2232 –41, Dec. 1997.

[72] B. Xu and M. Brandt-Pearce, “Comparison of FWM- and XPM-Induced Crosstalk Using the Volterra Series Transfer Function Method,” *J. Lightw. Tech.*, vol. 21, no. 1, pp. 40 –53, Jan. 2003.

[73] H. Bulow, F. Buchali, and A. Klekamp, "Electronic Dispersion Compensation," *J. Lightw. Tech.*, vol. 26, no. 1, pp. 158 – 67, Jan. 2008.

[74] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Tech.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[75] S. Chandrasekhar and A. H. Gnauck, "Performance of MLSE Receiver in a Dispersion-Managed Multispan Experiment at 10.7 Gb/s Under Nonlinear Transmission," *IEEE Photon. Technol. Lett.*, vol. 18, no. 23, pp. 2448 – 50, Dec. 2006.

[76] B. Xu and M. Brandt-Pearce, "Multiuser square-law detection with applications to fiber optic communications," *Communications, IEEE Transactions on*, vol. 54, no. 7, pp. 1289 – 98, Jul. 2006.

[77] C. Fludger, T. Duthel, D. van den Borne, C. Schulien, E.-D. Schmidt, T. Wuth, J. Geyer, E. D. Man, G. Khoe, and H. de Waardt, "Coherent Equalization and POLMUX-RZ-DQPSK for Robust 100-GE Transmission," *J. Lightw. Tech.*, vol. 26, no. 1, pp. 64–72, Jan. 2008.

[78] J. Renaudier, G. Charlet, M. Salsi, O. Pardo, H. Mardoyan, P. Tran, and S. Bigo, "Linear Fiber Impairments Mitigation of 40-Gbit/s Polarization-Multiplexed QPSK by Digital Processing in a Coherent Receiver," *J. Lightw. Tech.*, vol. 26, no. 1, pp. 36–42, Jan. 2008.

[79] T. Kupfer, J. Whiteaway, and S. Langenbach, "PMD Compensation using Electronic Equalization particular Maximum Likelihood Sequence Estimation," in *OFC/NFOEC 2007*, Mar. 2007, p. OMH1.

[80] S. J. Savory, A. D. Stewart, S. Wood, G. Gavioli, M. G. Taylor, R. I. Killey, and P. Bayvel, "Digital Equalisation of 40 Gbit/s per Wavelength Transmission Over 2480 km of Standard Fibre Without Optical Dispersion Compensation," in *ECOC 2006*, 2006, p. Th2.5.5.

[81] Y. Han and G. Li, "Coherent Optical Communication Using Polarization Multiple-Input-Multiple-Output," *Opt. Express*, vol. 13, no. 19, pp. 7527–34, Sep. 2005.

[82] D. E. Crivelli, H. S. Carter, and M. R. Hueda, "Adaptive Digital Equalization in the Presence of Chromatic Dispersion, PMD, and Phase Noise in Coherent Fiber Optic Systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 4, Dallas, TX,, Nov. 2004.

[83] K. Kikuchi, "Polarization-Demultiplexing Algorithm in the Digital Coherent Receiver," in *IEEE LEOS Summer Topicals*, Acapulco, Mexico, Jul. 2008.

[84] A. Leven, N. Kaneda, and Y.-K. Chen, "A real-time CMA-based 10 Gb/s polarization demultiplexing coherent receiver implemented in an FPGA," Anaheim, CA, p. OTuO2, Feb. 2008.

[85] S. Savory, "Digital Filters for Coherent Optical Receivers," *Opt. Express*, vol. 16, no. 2, pp. 804–17, Jan. 2008.

[86] X. Li, X. Chen, G. Goldfarb, E. Mateo, I. Kim, F. Yaman, and G. Li, "Electronic post-compensation of wdm transmission impairments using coherent detection and digital signal processing," *Opt. Express*, vol. 16, no. 2, pp. 880–8, Jan. 2008.

[87] T. Hoshida, T. Tanimura, S. Oda, T. Tanaka, H. Nakashima, Z. Tao, L. Li, L. Liu, W. Yan, and J. Rasmussen, "Recent progress on nonlinear compensation technique in digital coherent receiver," in *Proc. IEEE Optical Fiber Communications (OFC'10)*, San Diego, CA, May 2010, p. oTuE5.

[88] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From Theory to Practice: an Overview of MIMO Space-Time Coded Wireless Systems," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 281–302, 2003.

[89] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Wiley, 2003.

[90] O. V. Sinkin, R. Holzlöhner, J. Zweck, and C. R. Menyuk, "Optimization of the Split-Step Fourier Method in Modeling Optical-Fiber Communications Systems," *J. Lightw. Tech.*, vol. 21, no. 1, pp. 61–8, Jan. 2003.

[91] L. Kazovsky, "Balanced Phase-Locked Loops for Optical Homodyne Receivers: Performance Analysis, Design Considerations, and Laser Linewidth Requirements," *J. Lightw. Tech.*, vol. 4, no. 2, pp. 182 –95, Feb. 1986.

[92] K. Kikuchi, "Phase-Diversity Homodyne Detection of Multilevel Optical Modulation With Digital Carrier Phase Estimation," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, no. 4, pp. 563 –570, Jul.-Aug. 2006.

[93] G. Goldfarb and G. L., "BER Estimation of QPSK Homodyne Detection With Carrier Phase Estimation Using Digital Signal Processing," *Opt. Express*, vol. 14, no. 18, pp. 8043 – 53, Sep. 2006.

[94] W. Shieh, H. Bao, and Y. Tang, "Coherent optical ofdm: theory and design," *Opt. Express*, vol. 16, no. 2, pp. 841–59, Jan. 2008.

[95] L. B. Du and A. J. Lowery, "Improved Single Channel Backpropagation for Intrachannel Fiber Nonlinearity Compensation in Long-Haul Optical Communication Systems," *Opt. Express*, vol. 18, no. 16, pp. 17 075–17 088, Aug. 2010.