A LOW COMPLEXITY ADAPTIVE FRAMEWORK FOR THE RATE AND PERFORMANCE ENHANCEMENT OF PMD LIMITED FIBER OPTIC COMMUNICATION SYSTEMS

by

Turgut Mustafa Öktem

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Electrical & Computer Engineering

Koç University

November, 2006

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Turgut Mustafa Öktem

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Alper T. Erdoğan, Ph. D. (Advisor)

_____

Alper Demir, Ph. D.

_____

Alphan Sennaroğlu, Ph. D.

Date: _____

To my parents

ABSTRACT

In this thesis, an adaptive signal processing based framework for the improvement of fiber optical communication systems is presented. We use Polarization Division Multiplexing (PDM) technique in conjunction with the proposed blind source separation algorithms to increase the bandwidth efficiency of the system where two independent signals are transmitted over two orthogonal polarizations. We classify the PDM adaptive receiver design problem into two separate cases corresponding to low and high symbol rates. In the low symbol rate case, the main problem is to separate the transmitted sequences from their instantaneous mixtures. In the high symbol rate case, the goal is not only the separation of mixing in space but also the compensation of the mixing in time, i.e., the Inter Symbol Interference (ISI), caused mainly by the Polarization Mode Dispersion (PMD). Major constraints in the development of adaptive structures for these goals are the limitations imposed by the existing technology in terms of processing very high rate fiber communication signals. In fact, the approach proposed in this thesis is centered around the development of low complexity methods to address this issue. The proposed structures and algorithms are based on the special class of Blind Source Separation (BSS) algorithms exploiting magnitude boundedness of the digital communication signals. As shown in the thesis, the use of the Magnitude Bounded BSS (MB-BSS) approach enables the development of low complexity algorithms and the corresponding hardware structures to counteract the mixing of digital communication signals in both time and space caused by the communication medium. As the most challenging task is the design of adaptive structures for compensating convolutive mixtures corresponding to the high symbol rate, a special attention is given to this problem. For this purpose, the algorithmic extensions of the basic MB-BSS algorithm for instantaneous mixtures that make use of the novel paraunitary mapping approaches have been developed. Furthermore, the parametric versions of adaptive paraunitary MB-BSS algorithm are introduced for the purpose of reducing complexity and enabling all-analog compensators. Simulations are used to illustrate the convergence speeds and robustness of

the proposed algorithms whose results indicate that MB-BSS algorithm for both of the cases show nice convergence behaviors. We also present simulation results to demonstrate the successful tracking capability of the algorithm against time variations of the fiber channel. All simulation results indicate that MB-BSS based approach provides a promising framework for the rate and performance enhancement of optical communication systems which suffer from PMD.

ÖZETÇE

Bu tezde, fiber optik haberleşme sistemlerini geliştirmek için uyarlamalı sinyal işleme tabanlı bir sistem sunulmuştur. Sistemin bant genişliği verimini artırmak için, birbirinden bağımsız iki sinyalin iki dik polarizasyon üzerinden iletildiği Polarizasyon Bölüşümlü Çoğullama (PBÇ) tekniğini, Gözü Kapalı Kaynak Ayrıştırma (GKKA) algoritmalarıyla beraber kullanıyoruz. PBÇ uyarlamalı alıcı dizayn problemini düşük ve yüksek sembol hızlarına göre iki ayrı duruma sınıflandırıyoruz. Düşük sembol hız durumunda asıl problem, gönderilen dizilerin anlık karışımlarından ayrıştırılmalarıdır. Yüksek sembol hız durumunda ise, amaç sadece uzaydaki karışmayı ayrıştırmak değil, aynı zamanda zamandaki karışmayı, yani Polarizasyon Kipi Saçılması (PKS)'nın yolaçtığı simgelerarası karışmayı düzeltmektir. Bu amaçlar için uyarlamalı yapılar geliştirmede ana kısıt çok yüksek hızdaki fiber haberleşme sinyallerinin işlenmesindeki günümüz teknolojisinin getirdiği kısıtlamalardır. Nitekim bu amaçla, bu tezde öne sürülen yaklaşım, düşük karmaşıklığa sahip yöntemlerin geliştirilmesi merkezlidir. Önerilen yapılar ve yaklaşımlar genlik sınırlandırılmış sayısal haberleşme sinyalleri kullanan GKKA algoritmalarının özel bir sınıfı temellidir. Tezde izah edildiği üzere, genlik sınırlandırılmış GKKA yaklaşımı, sayısal haberleşme sinyallerinin haberleşme ortamı yüzünden hem zaman hem de uzayda karışmasını engellemek için düşük karmaşıklığa sahip algoritmaların ve buna karşılık gelen donanım yapılarının geliştirilmesine imkan sağlar. En zorlu görev yüksek sembol hızlarına karşılık gelen evrişimsel karışımları denkleştirmek için uyarlamalı yapılar dizayn etmek olduğundan, bu probleme özel bir önem verilmiştir. Bu amaçla, anlık karışımlar için olan temel genlik sınırlandırılmış algoritmanın yeni birimcilimsi eşleme yaklaşımları kullanan algoritmik uzantıları geliştirilmiştir. Buna ek olarak, algoritma karmaşıklığını azaltmak ve tümüyle analog denkleştiricileri mümkün kılabilmek için, genlik sınırlandırılmış GKKA algoritmasının parametrik sürümleri de tanıtılmıştır. Sonuçları genlik sınırlandırılmış GKKA algoritmasının her iki durumda da güzel yakınsama davranışları gösterdiğini belirten, önerilen algoritmaların yakınsama hızlarını ve sağlamlığını test eden benzetimler yapılmıştır. Ve ayrıca algoritmanın, fiber kanaldaki zamana bağlı

olan değişimleri başarılı bir şekilde izleme kapasitesine sahip olduğunu gösteren benzetim sonuçlarını da sunuyoruz. Tüm benzetim sonuçları göstermektedir ki, genlik sınırlandırılmış GKKA tabanlı yaklaşım, PKS'den zarar gören optik haberleşme sistemlerinin hız ve performans artırımı için ümit veren bir sistem sağlamaktadır.

ACKNOWLEDGMENTS

First, I would like to thank my supervisor Assistant Prof. Alper T. Erdoğan who has been a great source of knowledge for me. With his continuous help and patience, we have accomplished to complete this thesis work. I also acknowledge TUBITAK for their financial support for my thesis.

I am also grateful to the members of my thesis committee for critical reading of the thesis and for their valuable comments.

Many thanks to Tayyar Önal, Umut Aydemir, Salih Ertürk, Ahmet Aksoy, Ata Akatay, Hüseyin Çankaya, Erhan Yılmaz and Pınar Karabulut for their friendship and support.

Finally I thank my family for supporting me every time. They were always with me during the hard days of the research. All of my success during my whole life took place with their priceless support.

TABLE OF CONTENTS

List of Tables **xii**

List of Figures **xiii**

Nomenclature **xvi**

Chapter 1: Introduction **1**

1.1 Brief History of Fiber Optics 3

1.2 Fiber Optic Developments and Standards 4

1.3 Fiber Modes 6

1.3.1 Multi-Mode Fiber 7

1.3.2 Single-Mode Fiber 7

1.4 Dispersion in Single-Mode Fibers 8

1.4.1 Chromatic (Material) Dispersion 8

1.4.2 Polarization Mode Dispersion 9

1.5 Methods for Dispersion Compensation 12

1.5.1 Optical Dispersion Compensation 13

1.5.2 Electronic Dispersion Compensation 15

1.6 Thesis Outline and Contributions 17

Chapter 2: Fiber Channel: Physical and Simulation Models **19**

2.1 Physical Model 19

2.1.1 Transmitter Side 19

2.1.2 Fiber Link 20

2.1.3 Receiver Side 21

2.2 Simulation Model 22

2.2.1 Transmit and Receive Filters 22

2.2.2 Mathematical Description of PMD 23
2.2.3 Discrete-Time Equivalent Channel 25

Chapter 3: Blind Separation of Polarization Channels for Low Symbol Rates 30

3.1 Blind Source Separation 30
3.2 Blind Source Separation for Bounded Magnitude Sources (MB-BSS) 33
3.3 Parametric MB-BSS Algorithm 39
3.3.1 Parametrization of the Unitary Separator 40
3.3.2 Derivation of the Parametric MB-BSS 43
3.4 Simplified MB-BSS Algorithm 45
3.4.1 All-Analog Circuit Implementation of the Simplified Parametric MB-BSS 47
3.5 Simulation Results of the proposed Algorithms 49
3.5.1 Simulation Results of MB-BSS for Moving Window Approach 50
3.5.2 Simulation Results of Parametric MB-BSS for Moving Window Approach 50
3.5.3 Simulation Results of the Simplified Parametric MB-BSS 55

Chapter 4: Blind Separation and Dispersion Compensation for High Symbol Rates 57

4.1 Convolutive Blind Source Separation 58
4.2 Convolutive Blind Source Separation for Bounded Magnitude Sources 59
4.3 Mapping to Paraunitary Set 61
4.3.1 Method of Alternating Projections (MAP) Based Paraunitary Mapping 62
4.3.2 Paraunitary Mapping Approach on Complex Stiefel Manifold with Gradient Descent 63
4.3.3 Paraunitary Mapping Approach on Block Convolution Matrix Space . 64
4.4 Simulation Results for Paraunitary Mapping Algorithms 67
4.5 Simulation Results for Convolutive MB-BSS Algorithm 69
4.6 Parametric Convolutive MB-BSS Algorithm 69

4.7 Simulation Results for Parametric Convolutive MB-BSS Algorithm 72

Chapter 5: Tracking of Time-Varying Fiber Channels 74

5.1 Physical and Simulation Model for Time-Variations in the Fiber Channel . . 75

5.2 Tracking Performance of Proposed Blind Algorithms 75

5.2.1 Simulation Results for Instantaneous MB-BSS 75

5.2.2 Simulation Results for Convolutive MB-BSS 76

Chapter 6: Extensions and Future Work 80

Chapter 7: Conclusion 83

Appendix A: Derivation of Updates for the Parametric Instantaneous MB-BSS Algorithm 85

Appendix B: Derivation of Updates for the Parametric Convolutive MB-BSS Algorithm 87

Bibliography 91

Vita 97

LIST OF TABLES

1.1 Jones vector representation of polarization states. 10

LIST OF FIGURES

1.1 What was expected to happen 1
1.2 What really happened 2
1.3 Propagation of light through the fiber optic cable 2
1.4 A WDM system 4
1.5 Increase in the BL product through the fiber optic generations 6
1.6 DGD between two orthogonal modes 10
1.7 Long fiber model 12
1.8 Optical PMD compensator 14
1.9 Electronic PMD compensator 15
1.10 Overall fiber link and compensator for non-PDM transmission 16
1.11 Overall fiber link and compensator for PDM transmission 16

2.1 Receiver architecture for a coherent PDM system. HY: four-port optical hybrid 22
2.2 Channel simulation model 25
2.3 Mean and standard deviations of SISIR(dB) for data rates between 1Gsym/s and 55Gsym/s 29
2.4 Frequency of SISIR(dB) at 11 Gsym/s data rate over 100 simulations 29

3.1 Blind source separation setup 31
3.2 Frequency spectrum of fiber channel 33
3.3 All-analog hardware implementation of the overall simplified parametric MB-BSS system 48
3.4 Arithmetic operations unit in detail 49

3.5 MB-BSS simulation result for SNR = 10 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve. 51
3.6 MB-BSS simulation result for SNR = 20 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve. 52
3.7 Parametric MB-BSS simulation result for SNR = 10 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve. 53
3.8 Parametric MB-BSS simulation result for SNR = 20 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve. 54
3.9 Simplified parametric MB-BSS simulation result for SNR = 10 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve. 56

4.1 Block representation of an FIR filter 58
4.2 Comparison of convergence speeds of MAP and the proposed algorithms (a) Convergence curves for MAP and gradient descent on block convolution matrix space, (b) Convergence curves for MAP and gradient descent on Stiefel manifold. 68
4.3 Convolutive MB-BSS simulation result for SNR = 10 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve. 70
4.4 Convolutive MB-BSS simulation result for SNR = 20 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve. 71
4.5 Parametric convolutive MB-BSS simulation result for SNR = 20 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve. 73

5.1 Tracking performance of instantaneous MB-BSS algorithm for impulsive channel change (a) Initial and distorted channel spectrum of the 2^{nd} row 2^{nd} column element (b) Signal to Interference Ratio convergence curve. 77

5.2 Tracking capability of instantaneous MB-BSS for smooth channel variations (a) SIR convergence curve for the case that channel changes at every new window, (b) SIR convergence curve for the case that channel changes at every third window. 78

5.3 Tracking performance of convolutive MB-BSS algorithm for impulsive channel change (a) Initial and distorted channel spectrum of the 2^{nd} row 2^{nd} column element (b) Signal to Interference Ratio convergence curve. 79

6.1 Simulation result of simplified parametric MB-BSS for instantaneous case with multilevel signalling scheme (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve 81

NOMENCLATURE

PMD	Polarization Mode Dispersion
DGD	Differential Group Delay
CD	Chromatic Dispersion
PDM	Polarization Division Multiplexing
PDL	Polarization Dependent Loss
WDM	Wavelength Division Multiplexing
EDFA	Erbium Doped Fiber Amplifier
BSS	Blind Source Separation
DSP	Digital Signal Processor
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
LO	Local Oscillator
DBOR	Double Balanced Optical Receiver
ASE	Amplifier Spontaneous Emission
AWGN	Additive White Gaussian Noise
NA	Numerical Aperture
ADC	Analog to Digital Converter
MIMO	Multiple-Input Multiple-Output
ISI	Intersymbol Interference
SNR	Signal to Noise Ratio
SIR	Signal to Interference Ratio
FIR	Finite Impulse Response
IM/DD	Intensity Modulated/Direct Detection

Chapter 1

INTRODUCTION

Nowadays multimedia applications over the internet such as video streaming, live TV broadcasting and etc... are becoming widespread all over the world. These actions load a heavy burden on the telecommunications network. Therefore high bandwidth channels are required to tackle with the situation. A good candidate is to use light for data transmission.

It is not a new idea to use light to carry information. People have been using light for thousands of years to communicate with each other by means of smoke, fire, lighthouses and etc... However the use of light was limited because people used to think that light could only travel in straight lines. For this reason only line of sight (LOS) applications could be carried out at those times. In Boston 1870, an Irish physicist, John Tyndall gave a public demonstration of an experiment that light could travel in curved paths [1]. He filled a container with water and shone light into it through a glass plate. Then he pulled out the bung at the other side of the container in a dark room. Water gushed out and light shone out of the hole. It was thought that light would travel the way described in Figure 1.1,

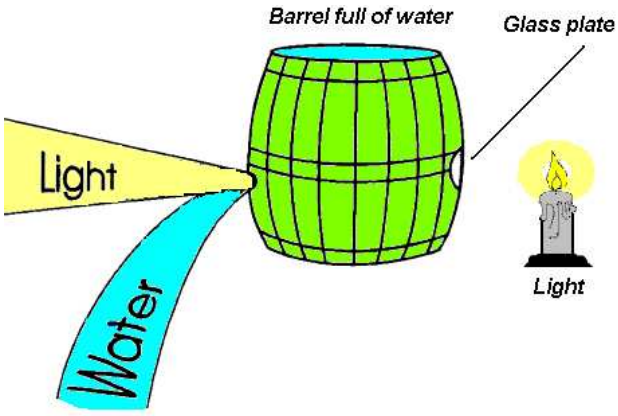


Figure 1.1: What was expected to happen

but instead, light stayed in water and followed the curved path as shown in Figure 1.2. With this experiment, scientists learned that light could be guided.

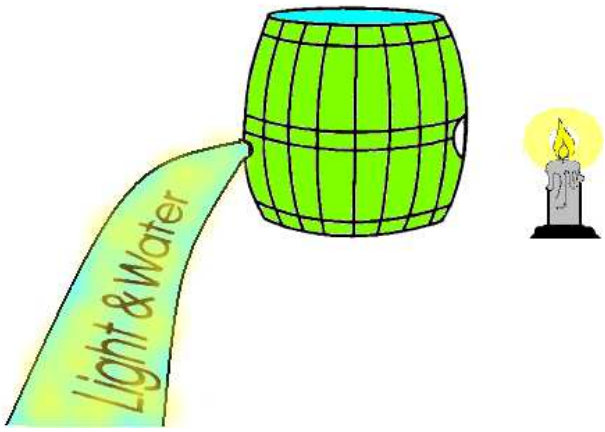


Figure 1.2: What really happened

The experiment carried by John Tyndall was a demonstration of total internal reflection which is the main principle to guide light through fiber glass. Total internal reflection is based on Snell's theory that light coming from the medium with greater refractive index reflects back to its medium if the angle between the ray and the surface normal is greater than critical angle. This critical angle can be easily calculated with Snell's law.

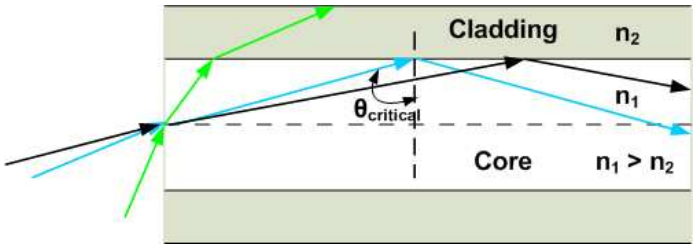


Figure 1.3: Propagation of light through the fiber optic cable

One can ask the question "what is the advantage of using optical communication systems?". One of the reasons is that it is free from electrical interference which is considered as the major source of noise arising from electrical connections. However the main advantage is the very huge bandwidth provided by optical communication systems. Optical communication systems offer a much wider bandwidth than the microwave communication systems

because optical carrier frequencies are typically 200 THz while the microwave carrier frequencies are about 1 GHz. As the bandwidth of the channel can be up to few percent of the carrier frequency, there are orders of magnitude difference with respect to bandwidth between microwave and optical communication systems [2]. Also there are more advantages in optical communication systems in terms of weight, size, material cost, information security and etc...[3].

1.1 Brief History of Fiber Optics

The idea of using glass fiber to carry speech signals was first originated by Alexander Graham Bell in 1880. But the development of optical communication technology mainly started in 1960s. Since 1960s huge amount of research has been focused on this area and many important inventions took place. Most important among them are the invention of lasers, low-loss optical fibers and erbium doped fiber amplifiers (EDFAs). These are the milestone events for the development of optical communication systems. Laser, as a coherent light source is a key element of the transmitter side of the channel. The low-loss optical fibers (0.2 dB/km at 1550nm) and EDFAs are also vital components of the optical communication systems. Before the invention of the EDFAs, electrical regenerators were used to be placed periodically along the fiber where the power of the optical signal fell below a certain threshold. The task of the regenerator was to convert the weak optical signal to electrical signal, amplify it and then convert it back to optical signal and retransmit it. The bit rate of the overall system was limited to the speed of the electronics used in this electro-optical conversion mechanism. Also during these conversions, considerable amount of noise was generated because electrical domain is a major source of noise. By the invention of the low-loss fibers the repeater spacing increased. This reduced the system cost and the overall noise of the system. However the speed limit due to the electronics was still a problem. EDFAs solved this problem by amplifying the optical signal in the optical domain, eliminating the need for electro-optical conversions. These EDFAs also produce noise called amplifier spontaneous emission (ASE), however this noise is small compared to the regenerator noise.

1.2 Fiber Optic Developments and Standards

Fiber optic technology is improving incredibly fast. In 1992, highest capacity of fiber optic channels was 2.5 Gb/s. Ever since the capacity is growing. Nowadays 40 Gb/s capacity can be achieved for a single channel. And by using wavelength division multiplexing (WDM) technology, Tb/s capacities are reached. In WDM technology, many signals in different wavelengths are multiplexed together and transmitted over the same channel. This obviously increases the system capacity but also the overall cost. Because for each signal there is a separate laser and at the receiver end of the fiber they all have separate photodetectors as can be seen in Figure 1.4. But it is absolutely worth.

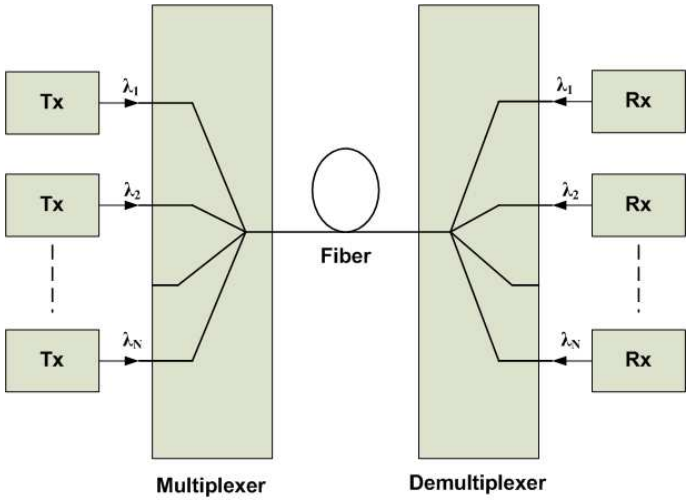


Figure 1.4: A WDM system

With this WDM technology, lightwave systems have reached to enormous capacities. Systems with more than 10 Tb/s capacity have been constructed with more than 250 WDM channels each operating at 40 Gb/s. The capacity of lightwave systems is open-ended and increases with every new technology. Future lightwave systems will force petabit per second speeds with enhanced laser and WDM technologies as proposed by Jeff D. Montgomery [4].

If we explain the evolution of fiber optic technology, we must begin with the first generation fiber optic systems. First generation fiber optic systems operated at a bit rate of 45 Mb/s and at wavelengths around 0.8 μm using GaAs semiconductor lasers between 1975-1980. They used repeaters for the amplification of the signal along the fiber cable, and

repeater spacings were approximately 10 km.

With second generation fiber optic systems, InGaAsP semiconductor lasers were begin to be used operating at 1.3 μm. Modal dispersion (explained later in the text) occurring due to multi-mode fibers was the biggest factor limiting the capacity. This problem was solved with the invention of single-mode fibers in 1981. Bit rates up to 1.7 Gb/s and repeater spacings up to 50 km were accomplished by 1987.

Third generation fiber optic systems operated at 1.55 μm. At this wavelength, fiber loss is minimum however dispersion is not. For this reason, dispersion shifted fibers (DSF) were began to be used to make the dispersion at 1.55 μm minimum. Repeater spacings up to 100 km and bit rates up to 2.5 Gb/s were reached.

Fourth generation systems had revolutionary improvements. EDFAs were invented and by this way, weak optical signals could be amplified in optical domain without any electro-optic conversions. As a result of this, repeaters were discarded. Data rate limitation problem due to the repeaters were solved. Furthermore WDM technology appeared in the fourth generation systems leading to huge capacity improvements. Capacities more than 10 Tb/s were accomplished.

Currently fifth generation systems are being developed. Research is going on to extend the wavelength ranges that WDM can operate. Also new dispersion compensation mechanisms are advancing.

To demonstrate the developments in the fiber optic generations, Figure 1.5 gives a good insight (The Figure is taken from [2].). In this Figure, bit rate distance product (BL product) versus year is plotted. Also the fiber generations are indicated on the graph. Bit rate distance product is a useful figure of merit in fiber optic communication systems. It is known that dispersion increases with the length of the fiber (We explain this issue later in detail.). For this reason reliable data transmission is limited by dispersion. As a result of this there is a trade-off between data rate and the length of the fiber. Therefore BL product is the most commonly used term to express the overall performance of the fiber link.

For the fiber optic standards, two of them are the most important. One of them is the Synchronous Optical Network (SONET). It was first developed by Bellcore and adopted by the American National Standards Institute (ANSI). Other one is the Synchronous Digital Hierarchy (SDH) developed by the International Telecommunications Union (ITU). These

two have different data rate formats. SONET uses optical carrier (OC) data rate format and SDH uses synchronous transport module (STM). To calculate the speed of OC-n, it is sufficient to multiply n with 51.8 Mb/s. Currently OC-1, OC-3, OC-12, OC-24, OC-48, OC-96, OC-192 and OC-768 are used. OC-768 corresponds to a bit rate of approximately 40 Gb/s.

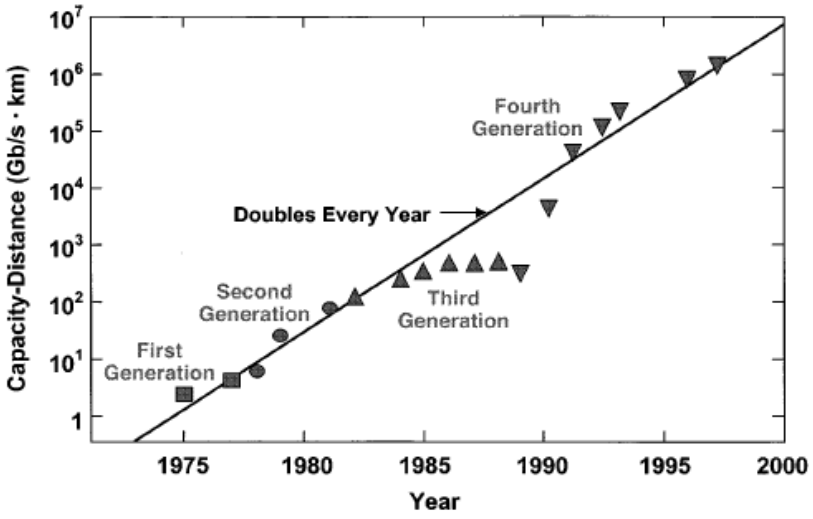


Figure 1.5: Increase in the BL product through the fiber optic generations

1.3 Fiber Modes

A very essential concept in fiber optic technology is the fiber mode propagating inside the fiber. Fiber modes determine the types of electromagnetic waves allowed to travel along the fiber. They are closely related to electromagnetic phenomenon. After some algebraic manipulations of Maxwell's electromagnetic equations along the fiber, the V parameter of the fiber is extracted. This V parameter determines the number of electromagnetic modes propagating inside the fiber. As V grows, then the number of modes increase. V parameter is given by:

$$V = k_0 a \sqrt{n_1^2 - n_2^2} \tag{1.1}$$

where $k_0 = 2\pi/\lambda$, a is the core radius, n_1 and n_2 are the refractive indices of the core and cladding of the fiber respectively as shown in Figure 1.3.

1.3.1 Multi-Mode Fiber

Multi-mode fibers have generally large core radii. Therefore it is easier to couple light into multi-mode fibers as they have a greater numerical aperture (NA). The core radius for them is typically between 25 μm and 30 μm whereas for a single-mode fibers, core radius is generally less than 5 μm. The use of multi-mode fibers is limited due to modal (multipath) dispersion. As its name suggests, there is a similar situation with a wireless communication system where multipath signals all arrive at different times. In multi-mode fibers, rays travel different paths of different lengths. Therefore their arrival times are different, which causes an input pulse to go out of the fiber broadened. This dispersion is a severe limitation for high bit rate transmission over multi-mode fiber. Although some research has been carried out to mitigate modal dispersion (such as graded index fibers), multi-mode fibers are not chosen for long-haul or high speed optical communications. They are mainly used for short range applications. Also recently some research is performed on multi-mode fiber optic channels to model them as a multiple input multiple output (MIMO) system [5]. The main underlying idea is to adapt the well known MIMO communication in wireless systems to multi-mode fiber optic communication systems by using multiple lasers and multiple photodetectors to enlarge the capacity by diversity techniques. With this new concept for multi-mode fibers, optimum usage of the multi-mode fiber capacity is aimed.

1.3.2 Single-Mode Fiber

When the V parameter discussed in Fiber Modes is below the cut-off V_c which is approximately 2.405, then only one mode will propagate through the fiber which is known as the fundamental mode or HE_{11} mode [6]. As there is only one mode present in single-mode fibers, modal dispersion does not take place which is the most difficult dispersion type to tackle with. Lack of modal dispersion makes the single-mode fiber suitable for high speed long-haul optical communication. However modal dispersion is not the only dispersion effect occurring in lightwave systems. There are many more dispersion types, and the aim of this thesis work is to electrically compensate one of these dispersions (polarization mode dispersion (PMD)) in single-mode fibers.

1.4 Dispersion in Single-Mode Fibers

Dispersion as a technical term, means broadening of a pulse at the output of a fiber due to some linear effects. The dispersion issue is an important problem in optical communications and precautions must be taken for it. If the broadening of the pulse is comparable with the bit period, then neighboring symbols interact with each other leading to intersymbol interference (ISI). Owing to ISI, reliable transmission of the symbols cannot be achieved. Therefore unless dispersion is compensated, high bit rate transmission is impossible. There are two major types of dispersion. One of them is chromatic dispersion (CD), and the other one is PMD.

1.4.1 Chromatic (Material) Dispersion

All optical pulses consist of a range of wavelengths. This range depends on the linewidth of the light source. Today with high quality, spectrally efficient lasers, wavelength of an optical pulse ranges from a fraction of a nm to 5 nm. Material dispersion originates from this multi-wavelength structure of the optical pulses because refractive index of the fiber is wavelength dependent. It varies with frequency as given by the famous Sellmeier equation. CD is an important effect because different spectral components associated with the pulse travel at different speeds given by c/n(ω) where c is the speed of light in vacuum and n(ω) is the frequency dependent refractive index of the single-mode fiber. Transfer function of a fiber of length L with only CD taken into consideration is given by:

$$H(\omega) = e^{jL\beta(\omega)} = e^{jL(\beta_0+\beta_1(\omega-\omega_0)+\beta_2(\omega-\omega_0)^2+...)}. \tag{1.2}$$

The mode propagation constant $\beta(w)$ was expanded around the central carrier frequency ω_0 where $\beta_m = (\frac{\partial^m \beta}{\partial \omega^m})_{\omega=\omega_0}$. The expansion around ω_0 is valid because $\Delta\omega << \omega_0$ where $\Delta\omega = \omega - \omega_0$. Above transfer function is an all-pass filter but causes waveform distortion due to its non-linear phase response. The first term in the expansion is just a phase shift, the second term is a time delay, however the other higher order terms are the sources for CD.

1.4.2 Polarization Mode Dispersion

PMD compensation is the main subject of this thesis, and therefore we will explain it in more detail. For bit rates beyond 10 Gb/s, PMD becomes the major limiting factor for optical transmission systems [7, 8]. Especially it causes severe problems in the fiber optic links deployed 10-20 years ago as they have been manufactured with old technologies. At the times they were installed, PMD was not a big concern as the bit rates were not so high. However as the bit periods are decreasing with the increasing bit rates, PMD is becoming one of the major obstacles in optical communications.

Polarization, one of the most important characteristics of the light waves, has been studied for centuries. An optical wave of arbitrary polarization can be represented as the superposition of two orthogonally polarized modes. Therefore a single-mode fiber actually supports two degenerate modes which are orthogonally polarized. In an ideal fiber these two modes are indistinguishable and have the same propagation constants due to the cylindrical symmetry of the waveguide. However in real fibers it is impossible to manufacture perfectly symmetrical fiber. Even if fibers are produced with perfect cylindrical symmetry, this symmetry breaks up because of environmental conditions in the deployed area such as stress, bending, twisting, temperature and etc... Another important thing is that, as these environmental conditions are not stable, PMD is not stationary. It changes all the time in a random fashion.

Loss of the circular symmetry gives rise to birefringence which means two orthogonal modes travel at different speeds because of the difference of their refractive indices. Birefringence is defined as the difference of propagation constants between the slow and fast modes:

$$\beta_s - \beta_f = \omega\,(n_s - n_f)\,/\,c = \omega\,\Delta\,n\,/\,c. \tag{1.3}$$

This birefringence is the source of PMD resulting in a differential group delay (DGD) between two orthogonal modes. As a result of DGD, slow mode is retarded $\Delta\tau$ seconds with respect to the fast mode as shown in Figure 1.6.

To see the difference between an ideal fiber and a birefringent fiber, first some important tools must be explained.

One of these useful tools is the complex Jones vector representation that is used to show

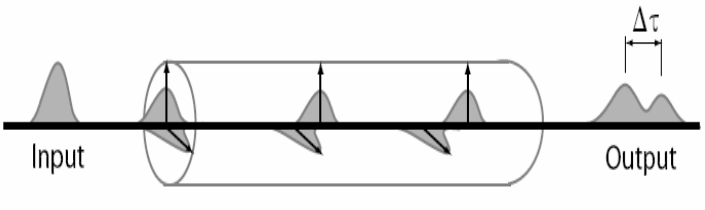


Figure 1.6: DGD between two orthogonal modes

the polarization state of a completely polarized light introduced by R. Clark Jones. Jones vector has two components and it is normalized (unit norm). For example $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ represents linear horizontal polarization state. You can see all the polarization states in Table 1.1.

Table 1.1: Jones vector representation of polarization states.

Polarization	*Jones vector*
linear horizontal →	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$
linear vertical ↑	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$
linear -45° ↘	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$
linear +45° ↗	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
circular left hand ↺	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}$
circular right hand ↻	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}$

There is also a Jones matrix representation. A Jones matrix is a 2 × 2 matrix, and we use it to express the transfer functions of the birefringent elements to model the PMD of long single-mode fibers which we will explain later. Now we can express the difference between an ideal fiber and a birefringent fiber with the help of Jones calculus:

The two orthogonal modes have the same propagation constants for an ideal fiber. There is no birefringence in ideal fibers. Therefore there is no DGD between the two modes and naturally no pulse broadening at the output. Also the state of polarizations do not change at the end of the fiber. This means that the polarization state of an electromagnetic wave is the same at the input and output of the fiber. If we represent the input state of polarization by the complex Jones vector $\mathbf{E_{in}}(\omega)$, then the 2×2 Jones transfer matrix $\mathbf{J}(\omega)$ of the ideal fiber and the output state of polarization $\mathbf{E_{out}}(\omega)$ is given for a fiber of length L by:

$$\mathbf{E_{out}}(\omega) = \mathbf{J}(\omega)\mathbf{E_{in}}(\omega), \tag{1.4}$$

where

$$\mathbf{J}(\omega) = \begin{pmatrix} e^{-j\beta L} & 0 \\ 0 & e^{-j\beta L} \end{pmatrix}. \tag{1.5}$$

But the situation is not the same for a birefringent fiber. The Jones transfer matrix is different because the two orthogonal polarization modes have different propagation constants leading to DGD. DGD and the Jones transfer matrix of the birefringent fiber with the same length is given by:

$$\mathbf{J}(\omega) = \begin{pmatrix} e^{-j\beta_x L} & 0 \\ 0 & e^{-j\beta_y L} \end{pmatrix}, \tag{1.6}$$

and

$$\Delta\tau = L\frac{d}{d\omega}(\beta_s - \beta_f). \tag{1.7}$$

As can be seen in (1.7), DGD for a short segment of fiber (polarization axis are fixed) is directly proportional to the length of the fiber L. If long-haul optical links are taken into account where the length of the single-mode fiber can exceed 10.000 km, then high data rate transmission would be impossible because of the accumulated PMD. Fortunately the situation is not the same for long fibers. This linear length dependence of DGD applies to fibers shorter than 100 - 1000 m. However, DGD does not accumulate along a long link of fiber in a linear fashion. While talking about long fibers, we model them as the concatenation

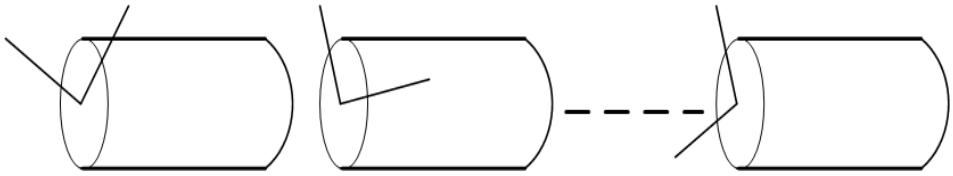


Figure 1.7: Long fiber model

of many short fiber sections whose polarization axis are fixed over their lengths as seen in Figure 1.7.

These small sections are all independent from each other and changing continuously (over milisecond time scales [9]) to emphasize the statistical random nature of PMD. Because of random variations in the perturbations along the fiber, the DGD in one section may either add to or subtract from another section of the fiber. As a result of this, mean DGD in long fibers accumulates in a random-walk-like fashion leading to a square-root length dependence [10, 11]. As PMD is a statistical process, it has an associated probability density function which is Maxwellian [12].

The D_p parameter known as the PMD parameter of the long fiber has the unit of $\text{ps}/\sqrt{km}$. It is in the range of 0.1-2 $\text{ps}/\sqrt{km}$ for the newly manufactured fibers. However, the fiber optic cables deployed in 1980s and 1990s have much greater D_p values. Because of the high D_p value in the old fibers, high speed optical communication over these fibers is a hard challenge. In order to realize high speed communications over fibers having high D_p values or over very long fibers having moderate D_p values, PMD compensation must be carried out.

1.5 Methods for Dispersion Compensation

The two orthogonal modes traveling along the fiber has lead to the idea of sending two independent symbols at the same time with two orthogonal polarizations. This concept is known as polarization division multiplexing (PDM) [13]. With this technique, bandwidth efficiency of the system is increased. Most commonly a single signal was transmitted and the received signal was demodulated at the receiver. Due to the birefringence, transmitted signal split into two orthogonal polarizations and traveled along the fiber, and because of PMD, random mode coupling (periodic exchange of energy between the modes) and DGD

took place resulting in a distorted received signal. PMD compensation methods tried to recover the symbol by using these two distorted polarizations. However in PDM, there is a different situation. Here two independent symbols are transmitted along the orthogonal polarizations which are mixed with each other and the compensation system must recover the two symbols.

In this thesis, we are trying to electronically compensate the PMD of a coherent optical channel using PDM. As PDM is a newly used transmission technique, the compensation methods to mitigate the PMD of these systems is a new and growing area of research. We will explore some of the proposed approaches very briefly. The PMD compensation techniques are divided into two major branches. One of them is the electrical compensation and the other one is the optical compensation, which will be explained in the following subsections. We should note that the common feature of optical and electronic PMD compensators is that both must be adaptive. Since PMD is a time varying effect changing in millisecond time scales, the compensator must be adaptive to track the variations in the channel. Otherwise it will be completely useless.

1.5.1 Optical Dispersion Compensation

There has been a considerable amount of research for the optical compensation of PMD. There are hundreds of papers written on this subject (See for example [14, 15, 16].). However as explained above, most of them are proposed for systems not using PDM transmission. They treat the fiber optic channel as a one input two output system, and try to recover the symbol from the two output polarization components [17, 18]. The optical compensation systems use optical components such as polarization controllers (PCs), delay lines, polarization maintaining fibers (PMF), polarization beam splitters (PBS) and etc...[14, 15]. These optical components are bulky and expensive. Also they may introduce polarization dependent loss (PDL) which can distort the linearity of the overall channel for PMD. A typical optical compensation system is performed before the photodetector as shown in Figure 1.8.

The general idea is to use adaptive feedback loops and send feedback signals to the optical compensator units based on some error criteria. Then the optical compensator is adjusted automatically or manually. The aim is to invert the system. Therefore the overall system (channel+compensator) would be the identity matrix or some scaled version of it.

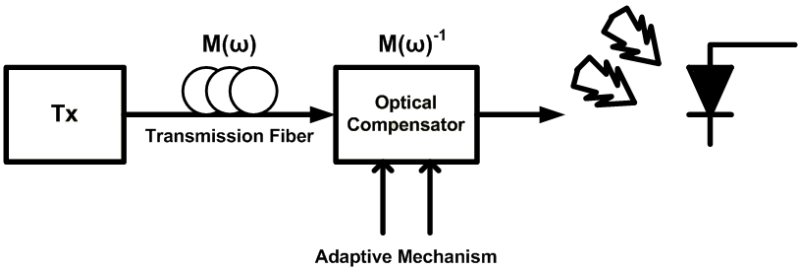


Figure 1.8: Optical PMD compensator

As an example for an optical compensator, we can consider the first order optical PMD compensator proposed in [19]. The optical first-order PMD equalizer demonstrated in this paper consists of a PC, a polarization splitter and combiner and a variable differential delay line. The delay line is adjustable to adapt the equalizer to different DGD values. Because DGD is not static due to PMD. Nearly all first-order PMD compensators are of this type. The received signal (before the photodetector) is split into two orthogonal polarizations by PBS, the fast polarization is retarded τ_c seconds (Delay time is dynamically adjusted according to the fiber DGD.), and finally the two polarizations are combined with a polarization beam combiner. PC is used here to transform the fast polarization component of the transmission fiber into the slow eigenstate of the equalizer. And this component is retarded with the help of the variable delay line. However this compensation is not very effective because the compensation can be achieved in a very small frequency region. Therefore more complicated designs with many PCs, delay lines are used to increase the degree of freedom (DOF) of the compensator. As a result, at the expense of increased cost and complexity, more flexible solutions can be obtained and higher-order PMD can be compensated. Interested readers can read [16] for further detail.

There are some proposed methods for PMD compensation of PDM systems. One of them is explained in [20]. In this method, polarization rotation is tracked along the fiber and compensated by automatic control methods. By this way, independent data transmitted over two polarizations can be recovered. However this is bulky and expensive.

1.5.2 Electronic Dispersion Compensation

In electronic compensation, as opposed to optical compensation, the processing takes place after the opto-electronic conversion as shown below:

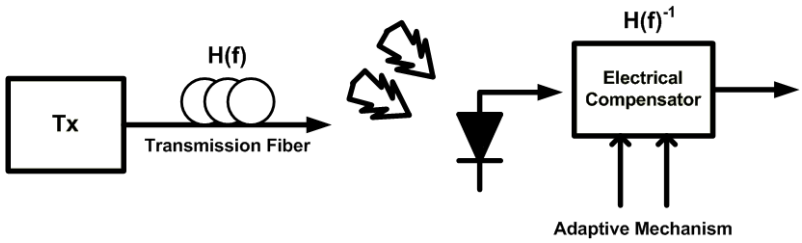


Figure 1.9: Electronic PMD compensator

There are two major types of opto-electronic conversion systems. One of them is the intensity modulated/direct detection (IM/DD) system. IM/DD receiver systems are the most basic receiver systems. They use only photodetectors. However the use of this system causes loss of precious phase information by square-law detection, which is needed for the linear compensation of the linearly modeled distortion of the fiber channel. Therefore complete compensation of PMD cannot be achieved with IM/DD systems. Therefore, as an alternative, a coherent receiver structure must be used. For a complete description of coherent receivers, you can look at [2]. The coherent receiver structure avoids the loss of phase information and enables the convenient linear modeling of the effective channel between the transmitter and the compensator. As a result, one can use linear compensation techniques, such as adaptive filters. For this reason, linear distortions such as CD and PMD can be effectively compensated by linear electrical equalizers for coherent detection systems [21]. Note that when we extract two orthogonal polarization components (for PDM systems), we use a more advanced coherent receiver structure. We explain the coherent receiver structure for PDM systems in detail in Chapter 2.

Depending on whether PDM is used or not, electronic compensator can take the form of a 2×1 vector (for systems without PDM) or a 2×2 matrix (for PDM systems). Because PDM systems treat the fiber channel as a 2×2 MIMO system (2 input and 2 output). Two orthogonal polarizations carry two independent signals. However for systems not using PDM, channel is modeled as 1×2 system. Because an input signal is splitted into two

orthogonal polarizations. At the receiver, the transmitted symbol is tried to be extracted from these two polarizations. We can visualize these two issues by the following Figures:

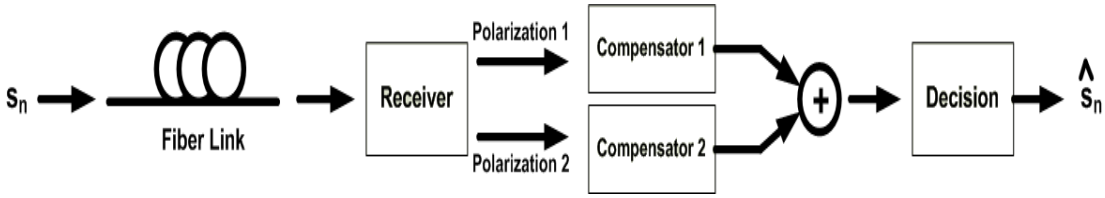


Figure 1.10: Overall fiber link and compensator for non-PDM transmission

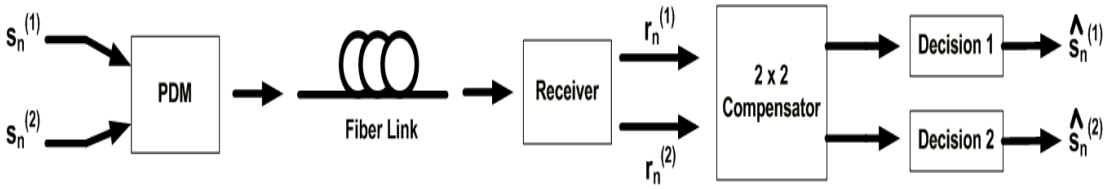


Figure 1.11: Overall fiber link and compensator for PDM transmission

Figure 1.10 shows the end-to-end block diagram for non-PDM transmission scheme. One symbol sequence is transmitted over the fiber link. It is then splitted into two orthogonal polarizations while traveling along the fiber. At the receiver, two compensators (Each of the compensators are scalars.) process the received orthogonal polarizations and try to recover the transmitted symbol. The outputs are summed and sent to the decision device.

Figure 1.11 shows the overall system implementing PDM transmission. Two independent digital communication signals are transmitted over two orthogonal polarizations and they travel along the fiber. The 2×2 compensator tries to extract two independent signals from the two mixture signals. Compensator outputs are sent to two separate decision devices. Finally the original source signals are detected by these devices.

One of the important issues about the adaptive compensator is whether it is training sequence based, i.e., supervised, or without using training i.e., unsupervised or blind. In the supervised scheme, when the receiver is connected to the line, the transmitter is required to send a known training sequence. This is undesirable in many scenarios especially if the particular channel is used for broadcasting or communication among several users.

Furthermore, after the initial "acquisition" phase, the equalizer needs to track the channel variations. Since the interruption of the information sequence, both is inconvenient and causes loss of bandwidth, it is more convenient to adapt the compensator using an unsupervised method, which will be the focus of our thesis, where we address the development of low complexity algorithms and structures for unsupervised compensation.

If we briefly survey the existing literature in this area: One of the earliest works on electronic compensation is by Winters and Gitlin [22]. In this reference, hardware, linear and decision feedback equalizers (DFE) are proposed where the use of decision directed sign least mean square (LMS) is proposed for low complexity unsupervised adaptation. Although the decision directed LMS is more appropriate for the tracking mode, it is not really appropriate for the initial training of the equalizer, as there are no reliable decisions available at the start.

There are various other electronic compensators proposed for dispersion compensation. In [21], linear equalizers are proposed and their performances are tested for coherent detection systems. In [23], MLSE receivers for IM/DD optical channels are introduced. [15] offers many methods and comparisons between them for electrical equalization, such as linear equalizers, DFEs and the combination of both.

Performance wise better approach is the use of maximum likelihood sequence estimator (MLSE) technique [24, 15]. However, the complexity of the MLSE algorithm is its major drawback making it unsuitable for a high-speed realtime application. Therefore, the results of [21, 15] provide more benchmarks about achievable performance.

If we compare electronic and optical compensation approaches: The electrical approach has lower size, complexity and cost compared to optical counterparts [15]. In addition, adaptation for both initial training and tracking is also more convenient with electrical compensation. Therefore, the current trend is to develop low cost-high performance electronic compensators integrated with the opto-electronic front-end.

1.6 Thesis Outline and Contributions

The organization of the thesis is as follows: The fiber channel, its physical and simulation models are introduced in Chapter 2. In Chapter 3, the blind algorithm is presented and its use in the PDM systems for low symbol rates are explained. Also simulation results

are given. In the next chapter we extend the concepts explained in Chapter 3 for high symbol rate systems. Chapter 3 and 4 are the main chapters of this thesis. As PMD is time-varying, the tracking phase is also a very important part and Chapter 5 explores the tracking performance of the adaptive blind algorithm. In Chapter 6, extensions and future work is presented. Finally Chapter 7 gives the conclusions of the overall work.

The contributions of this thesis work can be summarized as follows:

- The use of blind algorithms is a rather new concept in optical communications. There are a few research examples done so far. Also high computational complexities of general blind algorithms restrict their use in high speed applications. For this reason, we used BSS approach for Magnitude Bounded (MB-BSS) signals proposed in [25, 26] and its extensions for fiber communication applications. MB-BSS approach provides a suitable framework for development of low complexity blind structures, and in this thesis we demonstrated that PDM receivers designed based on MB-BSS have the desired low-complexity and high separation and dispersion compensation features.

- We have also developed parametric versions of the MB-BSS algorithm and used in this work. This parametrization opens the way of basic analog circuit designs to be used removing the need of digital signal processors (DSPs). DSP is mainly used for projection operations, and with the help of the parameterizations we eliminate the need of projections onto unitary and paraunitary spaces. We got successful results, especially for low symbol rate systems where ISI is negligible.

- The tracking performance of the algorithm was tested. As PMD is a time-varying phenomenon (in millisecond scales), the compensating algorithm must be adaptive and respond very quickly (in microseconds). This is the main difficulty of using blind algorithms, because it is a very hard task for blind update equations be completed in microsecond time scales. However as our algorithm has very low complexity, it can be used to track fast changing channels.

Chapter 2

FIBER CHANNEL: PHYSICAL AND SIMULATION MODELS

This chapter is devoted to explain the physical and simulation models of the single-mode fiber optic channel. The physical model describes the underlying natural phenomenon generating the propagation and reception of light in fiber optical communication system. Based on this physical model, an overall mathematical model describing the mapping between input and output signals is extracted. This mathematical model also forms a basis for the simulation model to be used for the characterization of the fiber channel and for the verification of the developed method. First of all, the physical fiber optic channel model is introduced, then the simulation model is explained.

2.1 Physical Model

A fiber optical communication link is composed of three major components: a transmitter, a fiber cable as the transmission medium and a receiver. At the transmitter side there are lasers and modulators. The modulated optical signal travels through the fiber optic channel which is composed of the single-mode fiber and periodically spaced EDFAs. At the receiver side there are opto-electronic transducers and demodulators. To increase the channel capacity with WDM, many signals of different wavelengths are multiplexed and transmitted over a single fiber. At the receiver side, these signals are demultiplexed and sent to different receiver branches, and in this thesis we concentrate on the design of adaptive receivers for each branch. For dispersion compensation and struggling against other fiber impairments, post-processing of the received signals are carried out by DSPs and/or analog filters. We will now look at each of the three components of a fiber optic link in more detail.

2.1.1 Transmitter Side

The transmitter side consists of encoders, modulators and light sources in general. An example for encoders is forward error correction (FEC). These are used in optical commu-

nication systems to maintain very low bit error rates (BER). The other two components are the most vital ones of the transmitter side. Modulator loads the information (the symbols) on the light source by modulating light based on these symbols. In optical communication systems, most common modulation scheme is On-Off Keying (OOK) which is a very simple scheme. As its name suggests, a light source is turned on or off based on the information bits. When a 1 is transmitted it is on, and when a 0 is transmitted it is off. As the light source, a laser is most commonly used. It is a very coherent source compared to light bulbs or light emitting diodes (LED). The laser technology has been a rapidly expanding field. Old lasers had linewidths of 50 MHz. Therefore they were not suitable for phase modulation. However today, lasers with linewidths of 1-5 MHz are commercially available which makes them suitable for phase modulation [21]. The other alternative modulation schemes that are typically employed are BPSK, 4-QAM, 16-QAM and M-PSK constellations. The constellation mapper outputs are used to modulate the transmit pulse sequence before modulating laser.

As introduced in Chapter 1, some transmitters can employ PDM scheme. With PDM, it is possible to send two independent information sequences over the two orthogonal polarizations. Considering the two receiver branches employed at a PDM receiver, the overall channel can be conveniently modeled with a 2×2 MIMO system. The PDM scheme provides better channel utilization and enables potential doubling of data rate. For example, if we use a 4-QAM signal constellation for each PDM input, we have a modulation efficiency of 4 bits per symbol which corresponds to a 40 Gb/s bit rate for a 10 Gsym/s symbol rate.

2.1.2 Fiber Link

The fiber link is the physical transmission medium. The modulated light is coupled into the fiber and travels along the fiber till it reaches the receiver. The fiber link is composed of the fiber cable and the optical amplifiers (OA) which are named as EDFAs. While traveling along the fiber, the signal strength decreases due to fiber losses. Therefore it must be amplified periodically along the fiber. Today typical optical amplifier repeating lengths are 50-100 km depending on the fiber quality and the wavelength of the optical signal. When WDM is not used, the mostly preferred wavelength is 1550 nm. Because at this wavelength fiber loss is minimum.

2.1.3 Receiver Side

The receiver architecture is more complicated for coherent optical systems than IM/DD systems. Coherent receiver structures are used to keep the phase information of the received signal after the opto-electric conversion. With IM/DD systems, the phase information is lost. To prevent this, coherent detection is used with the help of local oscillators. The received optical signal is combined with the local oscillator laser by beam combiners. Then the total signal is sent to the photodetector. Coherent detection with PDM is more complicated. In-phase and quadrature components of each of the two polarization components must be extracted now. The polarization of the local oscillator laser is also important and must be carefully designed, since it is combined with the corresponding polarization components of the received signals at the two branches. The overall reception process of the coherent receiver which is shown in Figure 2.1 can be described as follows:

1. The received optical signal is decomposed into its vertical and horizontal polarizations by PBSs.

2. Similarly the local oscillator laser source is decomposed into its vertical and horizontal polarizations.

3. Corresponding polarization components of the received optical signal and local oscillator are combined with the help of four-port optical hybrids.

4. The outputs of the optical hybrids pass through double balanced optical receivers (DBORs). Here opto-electronic conversion occurs and optical signal is converted into electrical current. The advantage of using DBORs is their ability to suppress relative intensity noise (RIN)[27].

5. The outputs of the photodetectors pass through trans-impedance amplifiers (TIA) to be converted to voltage. Finally TIA outputs are sampled at the symbol rate by the sampler.

All of the process is summarized visually by the following figure:

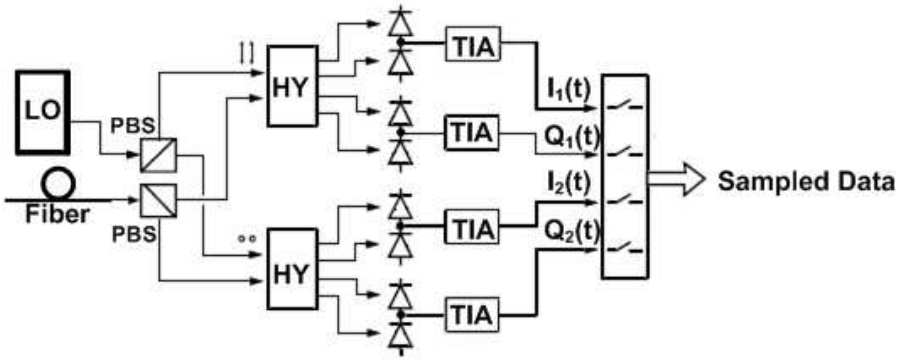


Figure 2.1: Receiver architecture for a coherent PDM system. HY: four-port optical hybrid

2.2 Simulation Model

As its name suggests, the simulation model enables the numerical experimentation of the overall fiber communication link in a computer environment. The simulation model consists of three parts. First part is the transmit filter used for pulse shaping. Second part is the fiber channel with PMD and third part is the receive filter matched to the transmit filter. First we obtain frequency domain transfer functions of all components and combine them to obtain the overall transfer function. Then a discrete time channel impulse response is extracted by fitting a finite impulse response (FIR) filter to the overall frequency spectrum.

We will first explain transmit and receive filters. Then the calculation of the channel PMD frequency response will be investigated. After all, we discuss fitting a discrete time causal FIR filter to the overall transfer function with least squares approximation.

2.2.1 Transmit and Receive Filters

Transmit filters are used in every communication systems for pulse shaping. The aim is to send the signals with suitable waveforms over the channel. Pulse shaping has several advantages. Some of them are listed below:

- Bandwidth limitation of the signal is achieved. The signal can be sent within the desired bandwidth. Bandwidth of the signal can be adjusted based on the data rate.

- Sending signals in suitable waveforms is an effective way to resist ISI.

It is desired that the transmitted signal waveform satisfies bandwidth constraints and can struggle against ISI. Some of the proposed waveforms for these issues are sinc pulses, raised cosine and square-root raised cosine pulses. Sinc pulses are more bandwidth efficient than other ones. However its tails die out more slowly in time domain. Small timing errors in sampling can result in severe ISI. Therefore in our model we used square-root raised cosine pulse filter at the transmitter and a matched filter (time reversed and conjugated filter) at the receiver. The aim of using a matched filter at the receiver is to maximize signal to noise ratio (SNR).

As we are designing a realizable system, time waveforms of the pulses must be finite. Therefore truncation operations will be carried out to satisfy finite time duration.

A raised cosine filter is constructed in frequency domain and has the following spectrum:

$$P(f) = \begin{cases} T & 0 \leq |f| < \frac{(1-\alpha)}{2T} \\ \frac{T}{2}\left[1 - sin(\frac{\pi T}{\alpha}(|f| - \frac{1}{2T}))\right] & \frac{(1-\alpha)}{2T} \leq |f| < \frac{(1+\alpha)}{2T} \\ 0 & \frac{(1+\alpha)}{2T} \leq |f| \end{cases} \tag{2.1}$$

Where $\frac{1}{T}$ is the data rate and α $(0 \leq \alpha \leq 1)$ is the roll-off parameter determining excess bandwidth. As can be seen the filter is bandlimited. The maximum bandwidth can be $\frac{1}{T}$ when $\alpha = 1$. Square-root raised cosine filter has the square-root of the above spectrum as its name suggests. From the frequency spectrum, a transmit filter of finite length in time domain can be easily designed with filter design techniques (such as equiripple filter design algorithm).

2.2.2 Mathematical Description of PMD

As explained in the previous chapter with Figure 1.7, PMD modeling of long single-mode fibers is accomplished by concatenation of many short fiber sections. To get realistic PMD statistics (Maxwellian distribution), more than 100 fiber sections are necessary. Most popular mathematical description of PMD is the 2 × 2 frequency dependent Jones Matrix. It indicates the equivalent baseband transmission properties of the channel PMD.

With the given insight, PMD is mathematically modeled as a concatenation of M polarization-maintaining fibers with varying group delays and rotations of the principal

axes [28, 19, 16]. Fiber frequency response $\mathbf{U}(\omega)$ is given by:

$$\mathbf{U}(\omega) = \prod_{\mathbf{i=1}}^{\mathbf{M}} \mathbf{D_i}(\omega)\mathbf{S_i} \tag{2.2}$$

where

$$\mathbf{D_i}(\omega) = \begin{pmatrix} e^{j\omega\frac{\Delta\tau_i}{2}} & 0 \\ 0 & e^{-j\omega\frac{\Delta\tau_i}{2}} \end{pmatrix}$$

and

$$\mathbf{S_i} = \begin{pmatrix} cos(\alpha_i)e^{j\phi_i} & sin(\alpha_i)e^{j\phi_i} \\ -sin(\alpha_i) & cos(\alpha_i) \end{pmatrix}.$$

- $\Delta\tau_i$ represents the group delay induced by the i^{th} section (each section has equal length). $\mathbf{S_i}$ is the scattering matrix and gives a frequency independent coordinate transformation of the principal axes.

- α_i and ϕ_i respectively denote the random polarization and phase angles, and are randomly generated with uniform distributions over (0,2π).

For the scattering matrix $\mathbf{S_i}$, there are more alternatives. Some authors use slightly different scattering matrices [29]. Total product of the matrices in (2.2) results in a frequency dependent unitary matrix $\mathbf{U}(\omega)$ where ω represents the angular frequency deviation from the carrier angular frequency ω_0. Therefore we have a baseband equivalent transfer function. Based on the scattering matrices used, $\mathbf{U}(\omega)$ may be a general unitary matrix or a structured unitary matrix in the following form [30, 31]:

$$\mathbf{U}(\omega) = \begin{pmatrix} u_1(\omega) & u_2(\omega) \\ -u_2^*(\omega) & u_1^*(\omega) \end{pmatrix} \quad \omega \in [\mathbf{0}, \mathbf{2}\pi) \tag{2.3}$$

satisfying the unitaryness constraint ($|u_1(\omega)|^2 + |u_2(\omega)|^2 = 1$ for all $\omega \in [0, 2\pi)$).

As a result, overall channel PMD transfer function is a paraunitary matrix (i.e.,unitary at each frequency). PDL may distort the unitary structure, however we neglect it in the thesis. Throughout the simulations in the thesis, we used a fiber cable of length 100 km having a PMD parameter (D_p) equal to 1 ps/$\sqrt{km}$.

2.2.3 Discrete-Time Equivalent Channel

Since the overall channel for the fiber link is bandlimited, we can model the effective channel within the transmission band in the sampled domain. The impulse response of the corresponding system can be obtained by fitting FIR impulse response spectrum to the transfer function in the transmit band. Although the overall channel is a 2 × 2 MIMO system, we describe the fitting procedure for one of its four scalar components.

The block diagram representation of the overall channel with transmit, receive filters and PMD channel is shown in the following figure. The components in the figure will be clarified soon.

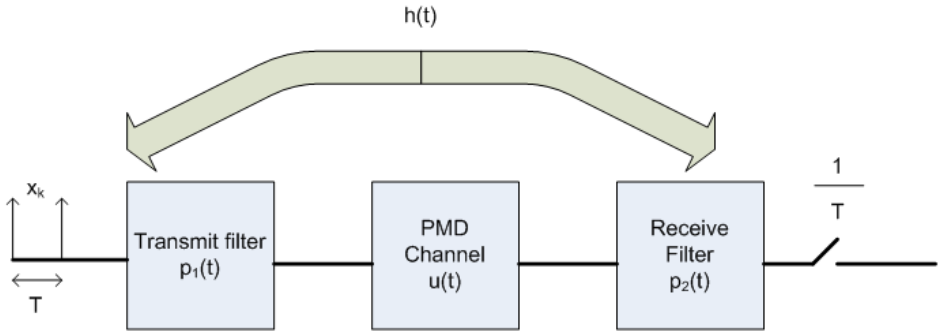


Figure 2.2: Channel simulation model

- Let $p_1(t)$ be the transmit filter and $p_2(t)$ be the receive filter where $p_2(t) = p_1^*(-t)$ (matched filter). Assume $u(t)$ be the impulse response of $U(w)$. Then the overall channel impulse response would be:

$$h(t) = p_1(t) \otimes u(t) \otimes p_2(t) = p(t) \otimes u(t) \tag{2.4}$$

  where $p(t) = p_1(t) \otimes p_2(t)$.

- As stated before, the combined receive transmit filter impulse response p(t) is selected as the raised cosine spectrum ($p_1(t)$ is a square-root raised cosine filter and $p_2(t)$ is time-reversed and conjugated version of $p_1(t)$. Therefore their combined spectrum is

raised cosine.) whose Continuous Time Fourier Transform (CTFT) is bandlimited to $\frac{1}{T}$.

$$P(f) = 0 \quad for \ \ |f| > \frac{1}{T} \tag{2.5}$$

where T is the symbol period. Therefore, the CTFT of the combined channel $h(t)$ is also bandlimited to $\frac{1}{T}$.

$$H(f) = P(f)U(f) = 0 \quad for \ \ |f| > \frac{1}{T}. \tag{2.6}$$

- As $H(f)$ is bandlimited to $\frac{1}{T}$, the system output $y(t)$ which is given by

$$y(t) = \sum_{k=-\infty}^{\infty} x_k h(t - kT) \tag{2.7}$$

would also be bandlimited to $\frac{1}{T}$.

- As total system is bandlimited to $\frac{1}{T}$, sampling at a rate of $\frac{2}{T}$ (Nyquist sampling rate) would be sufficient to represent this continuous signal. This sampling is carried out to obtain a discrete time equivalent channel. However sampling rate at the receiver is $\frac{1}{T}$ which is also the input symbol rate. Sampling the overall channel $h(t)$ at $\frac{2}{T}$ sampling rate gives the channel outputs z_n.

$$\begin{aligned} z_n &= y(n\frac{T}{2}) = \sum_{k=-\infty}^{\infty} x_k h(n\frac{T}{2} - kT) \\ &= \sum_{k=-\infty}^{\infty} x_k h((n-2k)\frac{T}{2}) \\ &= s_k \otimes g_k \end{aligned} \tag{2.8}$$

where s_k is the up-sampled version (with an up-sampling rate of 2) of x_k, i.e.,

$$s_k = \begin{cases} x_{k/2} & k \ \ even \\ 0 & k \ \ odd \end{cases} \tag{2.9}$$

and g_k are the sampled version of the continuous time impulse response $h(t)$ with a sampling rate of $\frac{2}{T}$, i.e.,

$$g_k = h(k\frac{T}{2}) \quad k \in Z. \tag{2.10}$$

Since $h(t)$ is bandlimited to $\frac{1}{T}$, the Discrete Time Fourier Transform (DTFT) of g_k has the following property that

$$G(w) = \sum_{k=-\infty}^{\infty} g_k e^{-jwk} = H(\frac{w}{T\pi}) \quad for \quad -\pi \leq \omega \leq \pi. \tag{2.11}$$

Now a causal FIR filter approximation is fit to the above frequency spectrum:

$$\sum_{k=0}^{L-1} g_k e^{-jwk} \approx H(\frac{w}{T\pi}) \quad for \quad -\pi \leq \omega \leq \pi \tag{2.12}$$

where L is the filter length. This approximation is a vital step because we cannot handle an infinite length filter to process the convolution. Therefore we make an approximation. The parameters g_k can be found by applying least squares to the system of equations obtained by evaluating above equation at different frequencies. Note that at the receiver, we sample at symbol intervals, if we call the symbol spaced samples o_k, we have

$$o_n = y(nT). \tag{2.13}$$

From (2.8) we know that $y(n\frac{T}{2}) = x_{k/2} \otimes g_k$, therefore

$$o_n = y(nT) = x_k \otimes g_{2k}. \tag{2.14}$$

Therefore, overall symbol spaced discrete equivalent channel would be equivalent to

$$h_k = g_{2k} \quad k = 0, 1, ..., \frac{L-1}{2}. \tag{2.15}$$

With the steps explained above, all four scalar components of the 2 × 2 MIMO discrete time impulse response of the overall channel are extracted. As PMD is a random phenomenon, each time we run the simulation, we get slightly different impulse responses. With the obtained discrete time impulse response, we can reflect the effective ISI of the channel as a function of symbol rate. For that purpose we introduce the Signal to ISI energy Ratio (SISIR) measure defined as:

$$SISIR = \frac{p^2}{\sum_{k=0}^{N-1} ||\mathbf{H}_k||_F^2 - p^2} \tag{2.16}$$

where

- $\{\mathbf{H}_k \, ; k = 0 \, , ..., \, N-1\}$ is the 2 × 2 discrete time channel,
- p is the frobenius norm of the main tap (i.e., the tap with the maximum frobenius norm):

$$p \;\; = \max_{k \in \{0,1,...,N-1\}} ||\mathbf{H}_k||_F. \tag{2.17}$$

In Figure 2.3, we plot SISIR as a function of symbol rate. Due to the random nature of PMD, SISIR is also random and therefore we plot its mean and its standard deviation neighborhood values.

The graph indicates that increasing data rates result in higher ISI (lower SISIR) as expected. This is because when we increase the data rate, bit periods decrease. And dispersion induced pulse broadening becomes comparable with bit periods. This causes ISI to increase with the increasing symbol rates.

Another plot (Figure 2.4) is shown to demonstrate the randomness of PMD. At a specific data rate (11 Gsym/s), 100 simulations were performed. And total number of occurrences of each SISIR value is given. As a result of this, the distribution of SISIR can be extracted.

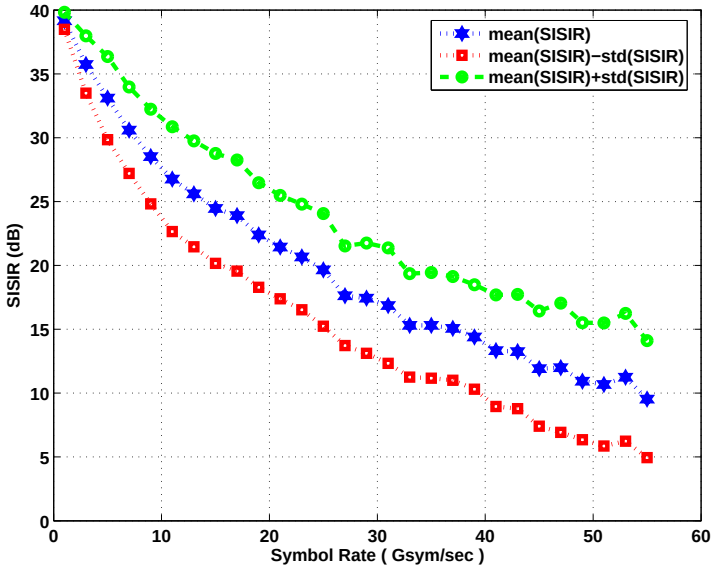


Figure 2.3: Mean and standard deviations of SISIR(dB) for data rates between 1Gsym/s and 55Gsym/s

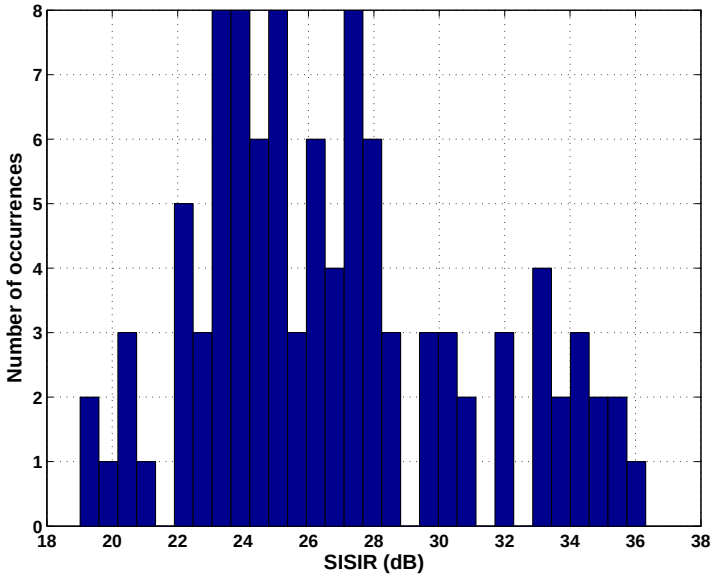


Figure 2.4: Frequency of SISIR(dB) at 11 Gsym/s data rate over 100 simulations

Chapter 3

BLIND SEPARATION OF POLARIZATION CHANNELS FOR LOW SYMBOL RATES

PMD as explained before, is a capacity limiting factor in optical communication systems. Especially it is more critical in systems using PDM transmission techniques. PMD induced coherent crosstalk makes the PDM systems much more sensitive to PMD [32]. There are some ways for struggling against PMD as we explained in preceding chapters. Polarization tracking techniques are used in PDM systems to observe the states of the independent polarization components. However this method requires optical components, hence it is expensive and bulky. Also principal states of polarization (PSP) methods are used against PMD. In this method, input state of polarization (SOP) is aligned to one of the PSPs. The SOP of a light does not change with frequency to first order when it is aligned along one of the PSPs of the fiber. In other words PSPs of the fiber are frequency independent to the first order. However in practice, a control signal must be fed back to the input from the receiver with an additional fiber and hardware which makes this method slow [14]. Therefore this is not practical too. As we elaborated in Chapter 1, electrical compensation techniques are better alternatives. Among the adaptive electrical filters probably the most popular one is the decision-directed LMS algorithm (DD-LMS). However DD-LMS cannot be used at the system startup as the decisions are not reliable at the startup. For a decision directed scheme, an eye-opening phase must be first completed before decision-directed algorithms are used. In this chapter, we will propose an alternative blind separation technique that can be used at the startup as it is not decision-directed. Before we describe our blind approach, we will provide some background about blind source separation.

3.1 Blind Source Separation

Blind signal processing is a rather new technique used in communication systems. Blind processing is used when no training signal is available.

In a conventional supervised adaptive scheme, the receiver training is carried out by using training sequences sent by the transmitter. The training sequence can be exploited by the receiver to identify the channel or for directly adjusting the adaptive receiver parameters such as adaptive equalizers or separators. Although the training based (i.e., supervised) adaptive filtering is currently being used in several communication applications, it has serious drawbacks. First of all the useful available bandwidth of the channel is wasted by training sequences. This reduces the overall capacity of the channel drastically. Furthermore there are some cases where the use of training sequences is not practical or very difficult to establish: The same communication link may be shared by several users and inclusion of a new user requires interruption in the service for other users. Furthermore, locating and synchronizing the training data location may be a nontrivial task especially for high-rate signals. As a result, unsupervised or blind adaptation of receivers is more preferable, which is indeed our approach in this thesis.

There are two main branches of blind signal processing. One of them is blind equalization (BE), and other one is blind source separation (BSS). In the broad sense, equalization is the process of counteracting the frequency selective effects of the channel, i.e. the mixing in time direction which is referred as ISI. In the BE, no training sequence is available, only outputs of the channel and some structural (and sometimes statistical) information about sources and channel is available.

Separation, on the other hand, is the process of extracting original signals from their mixtures. A general setup for separation problem is shown in Figure 3.1, where

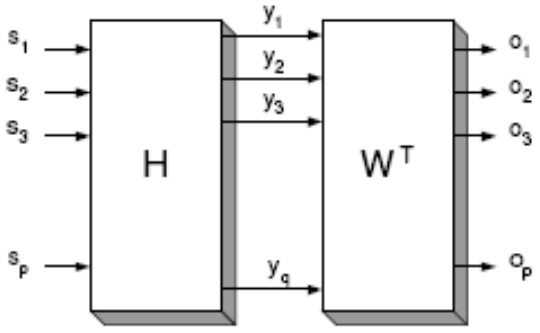


Figure 3.1: Blind source separation setup

- $s_1, ..., s_p$ are independent source signals,
- $\mathbf{H}$ is the unknown mixing channel,
- $y_1, ..., y_q$ are available mixtures.

The goal of separation is to come up with the separation system $\mathbf{W}^T$ such that its outputs are equal to or close to the original source signals. In the adaptive case, mixing channel $\mathbf{H}$ is unknown, and in the blind case no training signal is available. BSS algorithms generally exploit the independence of sources and some statistical or structural information about their values to obtain separators.

In PDM systems, since two independent polarization signals are transmitted and two different polarization signals are processed at the receiver, the corresponding $\mathbf{H}$ is a 2×2 MIMO system. Therefore, our goal in this thesis would be to devise adaptive algorithms and structures to obtain 2×2 separator system $\mathbf{W}^T$ (Transpose operation does not imply a special thing. We use it in MB-BSS algorithm that way for simplicity.).

We can classify BSS into two major fields: Instantaneous BSS and convolutive BSS. In *instantaneous* BSS, the mixing system $\mathbf{H}$ is memoryless, i.e. the mixtures at any time instants are functions of only the sources at the same time instant. In the fiber communications scenario, this refers to the case where there is no mixing in time (or ISI). Therefore, the instantaneous BSS scenario is applicable only if the symbol rate is low such that the frequency response of the channel within the transmission bandwidth can be assumed to be flat. According to Figure 3.2, which shows a typical frequency response of the 2×2 component of $\mathbf{H}$, this assumption seems to be valid for the frequency range (-20GHz,20GHz). Contrary to instantaneous BSS, in the convolutive BSS case, $\mathbf{H}$ has memory, therefore mixing happens in both time and space. In the fiber communication scenario, this refers to high symbol rate where the channel within the transmission band is frequency selective and create ISI. For the convolutive BSS case, the separator should also have memory to counteract against mixing in time.

In this chapter, we will consider low symbol rate scenario where channel can be considered as ISI free. High symbol rate case corresponding to convolutive BSS case will be addressed in Chaper 4.

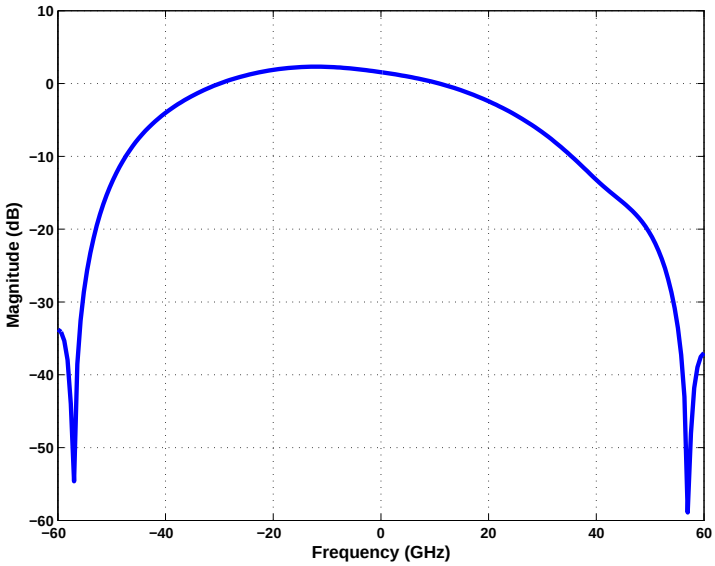


Figure 3.2: Frequency spectrum of fiber channel

Several blind algorithms have been proposed for these memoryless channels classified under this instantaneous blind source separation area [33]. These algorithms generally exploit the independence and the non-gaussianity of the original sources. The approaches are generally centered around the separation of the statistically independent sources (which is referred as Independent Component Analysis (ICA)) through the use of maximum likelihood [34, 35], maximization of non-gaussianity [36, 37, 38] and minimization of mutual information techniques [39, 40, 41].

3.2 Blind Source Separation for Bounded Magnitude Sources (MB-BSS)

The algorithm we use in this thesis is based on bounded magnitude property of digital communication sources hence it is named BSS for bounded magnitude sources (MB-BSS). All the algorithm is explained extensively in [26]. However as it is the heart of the thesis, we will also explain it here instead of just giving the reference.

The assumption about the boundedness of the source signals is realistic in many communication applications. For example, in multi-user digital communications applications, the signals are generated from the constellations which are bounded (QAM, QPSK and etc..). The MB-BSS algorithm is a gradient descent type (actually a subgradient descent type). A

convex cost function is minimized by using subgradient optimization techniques. Describing the BSS problem by the use of convex cost functions is a powerful way as they have only global minima and no saddle points. The overall BSS setup and the bounded magnitude sources algorithm are explained step by step below:

- The source signals are mixed with the memoryless MIMO system with transfer matrix **H**. **H** is a q × p matrix which implies that this mixing system has p inputs denoted by $s_1(k), s_2(k), ..., s_p(k)$ (bounded, mutually independent and complex symmetric) and q outputs denoted by $y_1(k), y_2(k), ..., y_q(k)$ i.e., at any instant k, we can write

$$\underbrace{\begin{bmatrix} y_1(k) \\ y_2(k) \\ \vdots \\ y_q(k) \end{bmatrix}}_{\mathbf{y(k)}} = \mathbf{H} \underbrace{\begin{bmatrix} s_1(k) \\ s_2(k) \\ \vdots \\ s_p(k) \end{bmatrix}}_{\mathbf{s(k)}}. \tag{3.1}$$

  In our case, the fiber channel is modeled as a 2 × 2 MIMO system as explained before. Therefore $q = p = 2$. **H** is a 2 × 2 matrix. **H** is the impulse response of the overall fiber channel. In the previous chapter, the impulse response was represented as an FIR filter with N taps (each tap being a 2 × 2 matrix). This is the general model. However with low symbol rates, the channel is frequency flat. Therefore it has one significant tap (main tap). And main tap is a unitary matrix as shown in the previous chapter ($\mathbf{HH}^H = \mathbf{I}$). Other taps can be neglected as their elements have very small magnitudes compared to the main tap. This can be verified by Figure 2.3. As can be seen in the plot, for data rates below 10 Gsym/s, SISIR is generally very high. Therefore other taps causing ISI can be neglected at these low rates. Low data rate fiber model is considered as ISI free. The only aim is to compensate the mixing of the polarization components by the fiber channel **H**.

- The purpose of BSS is to estimate the source signals from the observation sequences $\mathbf{y(k)}$ using a linear system with transfer matrix $\mathbf{W}^T$,i.e.,

$$\mathbf{z(k)} = \mathbf{W}^T\mathbf{y(k)}, \tag{3.2}$$

where $\mathbf{z(k)}=\begin{bmatrix} z_1(k) & z_2(k) & \dots & z_p(k) \end{bmatrix}^T$ contains the estimates of the source signals. $\mathbf{W}$ is obtained adaptively from the time samples (realizations) of $\mathbf{y(k)}$. No a priori knowledge of $\mathbf{H}$ and no training sequences are assumed.

- Generally $\mathbf{W}$ is decomposed into two operators $\mathbf{W} = \mathbf{W}_{pre}\mathbf{\Theta}$. $\mathbf{W}_{pre}$ is a q × p whitening matrix such that

$$\underbrace{\begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_p(k) \end{bmatrix}}_{\mathbf{x(k)}} = \mathbf{W}_{pre}^T\,\mathbf{y(k)} \tag{3.3}$$

is a white vector (elements are uncorrelated),i.e.,

$$E(\mathbf{x(k)x}^H\mathbf{(k)}) = \mathbf{I}. \tag{3.4}$$

The covariance matrix of a white vector as shown in the above equation may be an identity matrix, a scaled version of identity matrix or a diagonal matrix. In our case, we do not need a whitening matrix. The reason is that our channel is unitary. Therefore its outputs are already white as can be seen by:

$$E(\mathbf{y(k)y}^H\mathbf{(k)}) = E(\overbrace{\mathbf{Hs(k)}}^{\mathbf{y(k)}}\overbrace{\mathbf{s}^H\mathbf{(k)H}^H}^{\mathbf{y}^H\mathbf{(k)}}) = \mathbf{H}\overbrace{E(\mathbf{s(k)s}^H\mathbf{(k)})}^{\mathbf{I}}\mathbf{H}^H = \mathbf{I}. \tag{3.5}$$

The identity (or a scaled version of it) covariance matrix of the source signals $\mathbf{s(k)}$ comes from independence. As a result, we do not need a whitening matrix as we have a unitary fiber channel. In this case $\mathbf{W} = \mathbf{\Theta}$.

Not to carry out any whitening is another advantage of the BSS in fiber communications. It reduces the algorithm complexity drastically. Although the elements of $\mathbf{y(k)}$ are uncorrelated, they are not necessarily independent. Because being uncorrelated does not always guarantee independence. Therefore, our goal is to obtain a unitary

matrix $\mathbf{\Theta}$ which will decompose $\mathbf{y(k)}$ into the original source vector $\mathbf{s(k)}$ up to some permutation and phase ambiguities (These ambiguities are unavoidable in BSS.).

$$\mathbf{z(k)} = \mathbf{\Theta}^T\mathbf{y(k)} = \mathbf{\Phi P s(k)} \tag{3.6}$$

where $\mathbf{P}$ is a permutation matrix and $\mathbf{\Phi}$ is a diagonal matrix with unit-magnitude complex entries given as

$$\mathbf{\Phi} = \begin{bmatrix} e^{j\frac{\pi}{2}k_1} & 0 \\ 0 & e^{j\frac{\pi}{2}k_2} \end{bmatrix}. \tag{3.7}$$

$\mathbf{\Phi}$ represents the phase ambiguity and k_1 and k_2 are integers. For strictly real sources, the phase ambiguities would be multiples of π. In our thesis we use complex sources and phase ambiguities are multiples of $\pi/2$.

- In the previous step, BSS problem is formulated as finding a unitary $\mathbf{\Theta}$ matrix which converts the white fiber output vector $\mathbf{y(k)}$ to an independent vector $\mathbf{z(k)}$. The goal is to obtain $\mathbf{\Theta}$ in an iterative way. However there is a constraint that $\mathbf{\Theta}$ must be unitary. Therefore the problem is to obtain an optimizer solution under the unitary constraint:

$$\begin{aligned} &\text{maximize} && \mathcal{J}(\mathbf{z(k)}) \\ &\text{subject to} && \mathbf{\Theta}^H\mathbf{\Theta} = \mathbf{I}. \end{aligned}$$

Here $\mathcal{J}$ is the cost function to be optimized. In bounded magnitude sources algorithm the approach is:

$$\begin{aligned} &\text{minimize} && sup\ ||\mathcal{R}e\{\mathbf{z(k)}\}||_\infty \\ &\text{subject to} && \mathbf{\Theta}^H\mathbf{\Theta} = \mathbf{I}, \end{aligned}$$

which is the minimization of the supremum of the infinity norm of the real part of the output vector $\mathbf{z(k)}$ over the ensemble. There is no analytical solution found so far to the stated problem. Therefore iterative algorithms are used. However the proposed cost function is not differentiable. Fortunately, it is a convex cost function for which we can use subgradient methods.

In order to obtain a practical algorithm to minimize the cost function which can be used in real time applications, the evaluation of the maximum value of the infinity norm of $\mathbf{z(k)}$ should be limited to a finite window of time samples of z, i.e., $\{\mathbf{z(k)} : \mathrm{k} \in \{0, 1, ..., \Omega - 1\}\}$. Here we make a certain ergodicity assumption that these time samples reflect (or approximate) the ensemble behavior in terms of infinity norm of z. We will not further investigate the derivation of the algorithm. All derivations can be found in [26].

If we define

$$\mathbf{Y}= \left[\begin{array}{cccc} \mathbf{y(0)} & \mathbf{y(1)} & \dots & \boldsymbol{y(\Omega-1)} \end{array} \right],$$

as the matrix of input values in the window of interest, then for a given $\mathbf{\Theta}$, the corresponding outputs can be placed in a matrix $\mathbf{Z}$:

$$\mathbf{Z} = \left[\begin{array}{cccc} \mathbf{z(0)} & \mathbf{z(1)} & \dots & \boldsymbol{z(\Omega-1)} \end{array} \right] = \mathbf{\Theta}^T \, \mathbf{Y}.$$

Then update of the algorithm is given as:

$$\begin{aligned} \underline{\mathbf{\Theta}}^{(i+1)} &= \mathbf{\Theta}^{(i)} - \mu^{(i)} sign(\mathcal{R}e\{\mathbf{Z}^{(i)}_{m^{(i)},n^{(i)}}\})\bar{\mathbf{Y}}_{:,n^{(i)}}\mathbf{e}^T_{m^{(i)}} \\ \mathbf{\Theta}^{(i+1)} &= \mathcal{P}_{\mathcal{U}}\{\underline{\mathbf{\Theta}}^{(i+1)}\}, \end{aligned} \tag{3.8}$$

where

- $\mathbf{\Theta}^{(i)}$ is the value of $\mathbf{\Theta}$ at the i^{th} iteration,
- $\mathbf{Z}^{(i)}$ is the output matrix calculated based on $\mathbf{\Theta}^{(i)}$,
- $(m^{(i)}, n^{(i)})$ is the index for the maximum real component magnitude entry of $\mathbf{Z}^{(i)}$,
- $\mu^{(i)}$ is the step size at the i^{th} iteration,
- $\underline{\mathbf{\Theta}}^{(i+1)}$ is the unprojected version of the updated $\mathbf{\Theta}$,
- $\mathcal{P}_{\mathcal{U}}$ is the projection operator to the unitary matrix set,

- $\bar{\mathbf{Y}}_{:,n^{(i)}}$ is the n^{th} column of $\mathbf{Y}^{(i)}$ (which is $\mathbf{y(n)}$) with its elements complex conjugated,
- $\mathbf{e}_{m^{(i)}}$ is the m^{th} standard basis vector.

One important thing to note is the projection operation applied after each gradient update. This is a necessary step and it is enforced by the unitary constraint. At each iteration, $\mathbf{\Theta}$ matrix gets away from the unitary matrix set because of the subgradient update direction. Then the updated $\mathbf{\Theta}$ matrix is projected onto the unitary matrix set to realize the unitary constraint. The technique we use for $\mathcal{P}_{\mathcal{U}}$ in this thesis is minimum-distance orthogonalization. With this technique we try to find the projection operator which projects $\underline{\mathbf{\Theta}}^{(i+1)}$ to the minimum distance unitary matrix. The projection is given by (see [42] for proof):

$$\begin{aligned} \mathbf{\Theta}^{(i+1)} &= \mathcal{P}_{\mathcal{U}}\{\underline{\mathbf{\Theta}}^{(i+1)}\}, \\ &= \mathcal{P}_{\mathcal{U}}\{\mathbf{U}^{(i+1)}\mathbf{\Sigma}^{(i+1)}\mathbf{V}^{(i+1)^H}\} \\ &= \mathbf{U}^{(i+1)}\mathbf{V}^{(i+1)^H}, \end{aligned} \tag{3.9}$$

where $\mathbf{U}^{(i+1)}\mathbf{\Sigma}^{i+1}\mathbf{V}^{(i+1)^H}$ is the singular value decomposition (SVD) of $\underline{\mathbf{\Theta}}^{(i+1)}$.

The choice of the step size value is critical as it affects the convergence speed as well as the accuracy of the final result. Higher step size values are used for faster convergence speeds, however may increase the oscillations around the optimum solution. For more accurate results, step sizes need to be smaller. Mostly high step sizes are used at the first few iterations to get closer to the solution. Then step size is decreased to prevent big overshoots around the optimum point. The balance between convergence speed and accuracy requirements can be achieved by using dynamic step sizes with diminishing values. Various dynamic step size selection methods in subgradient optimization and their convergence behaviors have been studied extensively in [43, 44, 45, 46, 47]. Throughout the thesis we use different step size rules. Depending on the situation, constant step size can be used or a step size having the zero-limit and divergent sum (ZLDS) property ($\mu^{(i)} = \frac{\mu_0}{i+1}$) can be chosen.

And finally, for the real time implementation of the algorithm there are two different applications for the choice of the data window selection. One of them is the fixed window style. In this case, at every iteration of the algorithm the same data window is used. However the data window in this case is kept large. This approach has some drawbacks in terms of correlated noise enhancing and the difficulty in circuit implementation. The data window has noisy data which results from ASE noise or receiver thermal noise. These noise sources are represented as additive white Gaussian noise (AWGN) (zero mean noise source). If the maximum real magnitude component is detected at the same instance (say $\mathbf{z(k)}$) in consecutive iterations, then the same noise is added at every iteration. This enhances the noise and damages the algorithm. Also circuit implementation is more difficult for fixed window systems. The solution to the problem is using moving window approach. In this approach, each time a new data window is used (newly received data). As a result, correlated addition of noise is prevented and real time implementation of the hardware becomes easier. The simulation results will be shown for only the moving window approach. Fixed window approach also worked well. However moving window works better due to the reasons we have stated.

3.3 Parametric MB-BSS Algorithm

In this thesis, we have developed a parametric version of the MB-BSS algorithm. Parametric MB-BSS algorithm can lead to all-analog hardware implementation of the overall BSS setup or a complexity reduced version. With the parametric algorithm we get rid of the unitary constraint in the subgradient optimization problem. We convert the problem into an unconstraint one by parameterizing the unitary separator matrix. Optimization search is carried out on the unitary matrix set. Therefore there is no need to project onto the unitary matrix at each step. The removal of the unitary projection can make it possible to implement everything in analog domain. Because projection onto unitary matrix set requires complex algebraic equations (such as SVD) which can be carried out by DSPs. Even if digital implementation is used, total algorithm complexity will be reduced by getting rid of the unitary constraint. With parametric MB-BSS algorithm, we obtain two benefits. First of all, unitary projection is removed. This reduces the overall algorithm complexity.

Second one is that everything can be carried out in analog domain.

Analog implementation of the algorithm has several advantages over the digital counterpart. Hardware is simplified because analog to digital converters (ADCs) will not be present anymore. ADCs are important bottlenecks for the use of DSPs in high speed optical communications. At very high speeds, the precision of ADCs decrease a lot. Therefore quantization errors degrade the performance of the system. The current state of the art in ADCs is a 10-GSample/s, 5-bit ADC consuming 3.6 W [48]. And around 6-GSample/s, 6-bit ADCs are available. Also four ADCs are needed in the digital implementation (Two polarization signals each have in-phase and quadrature components.) which use much power. If analog filters are used, it is possible to process signals at higher rates with lower power consumption and also there will be no quantization errors. Also DSPs will not be used in the analog implementation. Therefore total cost of the BSS system will reduce.

3.3.1 Parametrization of the Unitary Separator

In the previous chapter we have stated that the 2×2 PMD transfer matrix is either a structured matrix in the form of (2.3) or a general unitary matrix depending on the scattering matrices used. In most of the literatures, transfer matrix is given as a structured unitary matrix as given in (2.3). Therefore the main tap of the impulse response of the overall channel (transmit filter + PMD channel + receive filter) is also a unitary matrix in one of the forms. We have developed parametric MB-BSS algorithm for both cases and obtained successful results. However we will explain the case for the general unitary case as it also covers the structured unitary matrices.

A general 2×2 unitary matrix $\mathbf{W}$ has the form of:

$$\mathbf{W} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{3.10}$$

All the components are complex numbers. However the elements are not arbitrary, they must satisfy some equations for the matrix to be unitary. We can list them as:

- The rows must have unit norm:

$$\begin{aligned} |a|^2 + |b|^2 &= 1 \\ |c|^2 + |d|^2 &= 1. \end{aligned} \tag{3.11}$$

It will be easier to write the elements in polar form because of the structure of the equations. These two equations enable us to write the components as:

$$\begin{aligned} a &= \cos(\alpha)e^{jx} \\ b &= \sin(\alpha)e^{jy} \\ c &= -\sin(\beta)e^{jz} \\ d &= \cos(\beta)e^{jt}. \end{aligned} \tag{3.12}$$

We can put arbitrary "-" or "+" in front of the coefficients. The reason why we have put "-" in front of $\sin(\beta)$ of "c" is that we tried to imitate the general structure of rotation matrices.

- The columns must have unit norm:

$$\begin{aligned} |a|^2 + |c|^2 &= 1 \\ |b|^2 + |d|^2 &= 1. \end{aligned} \tag{3.13}$$

If we compare this equation with (3.11), we immediately recognize that $|b| = |c|$ and $|a| = |d|$ meaning that $\alpha = \beta$. Therefore we rearrange (3.12) as:

$$\begin{aligned} a &= \cos(\alpha)e^{jx} \\ b &= \sin(\alpha)e^{jy} \\ c &= -\sin(\alpha)e^{jz} \\ d &= \cos(\alpha)e^{jt}. \end{aligned} \tag{3.14}$$

- The rows must be orthogonal:

$$ac^* + bd^* = 0. \tag{3.15}$$

If we plug in the values of "a, b, c, d" into (3.15), we obtain the following equation:

$$t = y + z - x. \tag{3.16}$$

Now the matrix is in the form of:

$$\mathbf{W} = \begin{bmatrix} \cos(\alpha)e^{jx} & \sin(\alpha)e^{jy} \\ -\sin(\alpha)e^{jz} & \cos(\alpha)e^{j(y+z-x)} \end{bmatrix}. \tag{3.17}$$

- The columns must be orthogonal:

We see that columns are already orthogonal as $ab^* + cd^* = 0$. Therefore orthogonality of columns are satisfied. As a result (3.17) is the final form of the 2×2 general unitary matrices. It has 4 real parameters denoted as α, x, y, and z.

There is another alternative parametrization technique for unitary matrices. The well-known theorem about unitary matrices say that any general 2×2 unitary matrix can be written as:

$$\mathbf{W} = e^{j\phi}\mathbf{U}, \tag{3.18}$$

where the exponential represents a global phase factor and $\mathbf{U}$ is a 2×2 unitary matrix with determinant +1. In Lie Algebra, $\mathbf{U}$ forms a special unitary group called as SU(2). Note that every unitary matrix has a determinant of either +1 or -1. The 2×2 unitary matrices having a determinant of +1 form the SU(2) group. The SU(2) group are the structured unitary matrices that we have mentioned before. They are in the form of:

$$\mathbf{U} = \begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix}, \tag{3.19}$$

where $|a|^2 + |b|^2 = 1$.

The **U** matrix can be parameterized by real parameters as we have done for a general unitary matrix. We again use the same techniques: Orthonormality of rows and columns. And a new constraint is the determinant being +1. With these operations, **U** is parameterized by 3 real parameters as expected. Because in (3.18), there is already a real parameter ϕ. With the 3 real parameters from the matrix **U**, we verify that any general 2×2 unitary matrix can be parameterized by 4 real parameters (ϕ, α, x and y):

$$\mathbf{W} = e^{j\phi} \overbrace{\begin{bmatrix} \cos(\alpha)e^{jx} & \sin(\alpha)e^{jy} \\ -\sin(\alpha)e^{-jy} & \cos(\alpha)e^{-jx} \end{bmatrix}}^{\mathbf{U}}. \tag{3.20}$$

The general 2×2 unitary matrix was parameterized by 4 real parameters in two different ways. Parametric BSS approaches using either of them showed approximately the same performances. Therefore we will only consider the case given by (3.20) in this chapter.

3.3.2 Derivation of the Parametric MB-BSS

The derivation of the parametric algorithm involves partial derivatives. We will use the chain rule to obtain the updates with respect to the 4 real parameters. First gradient is taken with respect to the unitary matrix as given in (3.8). Then differentiation with respect to the 4 real parameters are carried out. However we cannot use (3.8) directly. It must be modified because it involves complex numbers. Our 4 parameters are all real. Therefore derivation starts from the beginning of the algorithm. We transport everything in real number domain without any loss of precision. The derivation of the parametric MB-BSS is explained step by step below:

- The cost function is given as:

$$\begin{aligned} f(\mathbf{\Theta}) &= ||\mathcal{R}e\{vec(\mathbf{Z})\}||_\infty \\ &= ||\mathcal{R}e\{\underbrace{(\mathbf{Y}^T \otimes \mathbf{I})}_{A} \underbrace{vec(\mathbf{\Theta}^T)}_{\mathbf{b}}\}||_\infty \\ &= ||\mathcal{R}e\{\mathbf{Ab}\}||_\infty, \end{aligned} \tag{3.21}$$

where "vec" denotes the vectorization operation of the matrix and $\otimes$ denotes the Kronecker product. Vectorization of a matrix is simply placing the columns of the matrix on top of each other in order (first column is on the top). Here $\mathbf{\Theta}^T$ or $\mathbf{\Theta}$ is the unitary separator. The transpose does not change the unitaryness. A general unitary matrix will be used here as described by (3.20).

- Further simplification of the algorithm leads to:

$$\begin{aligned} ||\mathcal{R}e\{\mathbf{Ab}\}||_\infty &= ||\mathcal{R}e\{(\mathbf{A}_R + j\mathbf{A}_I)(\mathbf{b}_R + j\mathbf{b}_I)\}||_\infty \\ &= ||\mathbf{A}_R\mathbf{b}_R - \mathbf{A}_I\mathbf{b}_I||_\infty \\ &= ||\underbrace{\begin{bmatrix} \mathbf{A}_R & -\mathbf{A}_I \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} \mathbf{b}_R \\ \mathbf{b}_I \end{bmatrix}}_{\mathbf{r}}||_\infty \\ &= ||\underbrace{\mathbf{Kr}}_{\mathbf{s}}||_\infty = f(\mathbf{r}), \end{aligned} \tag{3.22}$$

where subscript "R" denotes the real part of the component while "I" denotes the imaginary part.

- A subgradient with respect to $\mathbf{r}$ is given as:

$$\underline{\mathbf{g}} = sign(s_i)\mathbf{K}_{i,:}^T \quad where \;\; |s_i| = ||\mathbf{s}||_\infty. \tag{3.23}$$

- Now chain rule is used to obtain the derivatives with respect to ϕ, α, x and y (which assumes that the cost function is differentiable at the current point). We will not explain the way for the 4 variables here as the method is the same for all. Just the case for α will be demonstrated in a brief way by skipping some details (The updates for the other 3 real parameters will be provided in the Appendix.). To obtain the derivative with respect to α, we use chain rule:

$$\frac{\partial f(\mathbf{r})}{\partial \alpha} = \underline{\mathbf{g}}^T . \frac{\partial \mathbf{r}}{\partial \alpha}. \tag{3.24}$$

In order to obtain $\frac{\partial \mathbf{r}}{\partial \alpha}$, first of all, the unitary separator matrix is vectorized to obtain $\mathbf{b}$. Then real and imaginary parts of the vector $\mathbf{b}$ is placed on top of each other to

construct the vector $\mathbf{r}$. These operations were all explained in preceding steps. If we skip the details, then the partial derivative vector is given by:

$$\begin{aligned}\frac{\partial \mathbf{r}}{\partial \alpha} = [\; & -\sin(\alpha)\cos(\phi + x) \quad -\cos(\alpha)\cos(\phi - y) \quad \cos(\alpha)\cos(\phi + y) \\ & -\sin(\alpha)\cos(\phi - x) \quad -\sin(\alpha)\sin(\phi + x) \quad -\cos(\alpha)\sin(\phi - y) \\ & \cos(\alpha)\sin(\phi + y) \quad -\sin(\alpha)\sin(\phi - x) \;]^T.\end{aligned} \tag{3.25}$$

- Final step is the update of the variable achieved as:

$$\alpha^{(i+1)} = \alpha^{(i)} - \mu^{(i)} \underline{\mathbf{g}}^T \cdot \frac{\partial \mathbf{r}}{\partial \alpha}, \tag{3.26}$$

where $\underline{\mathbf{g}}$ is given in (3.23) and $\frac{\partial \mathbf{r}}{\partial \alpha}$ in (3.25).

As with the normal algorithm, two different real time approaches (fixed window and moving window) are possible for the parametric MB-BSS algorithm. However as we stated before, moving window works better and is more advantageous from many aspects. Therefore we will only show the simulation results for the moving window approach (Fixed window approach also works well.).

3.4 Simplified MB-BSS Algorithm

Another important version of the MB-BSS algorithm is the simplified MB-BSS algorithm. The aim is to reduce the algorithm complexity even further and make the hardware implementation easier. Although the moving window is easier to construct than fixed window, it still requires that the data in the window of interest must be kept in a buffer for a while. Therefore memory must be added to the system which makes the whole system very complicated. Also operations must be carried out on this data window such as searching for the maximum magnitude real entry. All these are the drawbacks of the previous MB-BSS algorithms in a realtime hardware implementation. The simplified version brings great easiness to the overall hardware design. The simplified MB-BSS algorithm just processes the current data of the channel output. For this reason, it does not need any storage, hence

any hardware for memory. Also cooperation of the simplified MB-BSS algorithm with the parametric MB-BSS algorithm leads to a very important result: Overall system can be implemented in an all-analog fashion. Simplified version removes the need for memory and other complex operations that could only be carried out by DSPs. Parametric version removes the unitary projection operation which can only be done by DSPs. Therefore the application of the simplified algorithm with the parametric algorithm updates makes the all-analog implementation possible. The all-analog design will be introduced later in this chapter.

The simplified algorithm has the same subgradient updates given in (3.8). However it does not require any data window. It compares the current maximum magnitude real output of the separator with the target value of the absolute value of the real part of the input signal. If 4-QAM type constellations are used at the input, the absolute real output of the separator must be "1" in the convergence state. To summarize the algorithm:

- Assume that the received signal vector is $\mathbf{y}_i$ (channel output) at the instant. Then the separator output $\mathbf{z}_i$ is given as:

$$\mathbf{z}_i = \mathbf{\Theta}^T \, \mathbf{y}_i = \begin{bmatrix} a + jb \\ c + jd \end{bmatrix}. \tag{3.27}$$

- Let $K = \max\{|a|, |c|\}$ and let m be the index of this maximum. For example if $|a| > |c|$ then m = 1, otherwise m = 2. The simplified MB-BSS algorithm compares K with 1. If $K < 1$, no operation is performed. And next output is waited. If $K > 1$, then the separator is updated with the subgradient given by (3.8).

$$\begin{aligned} \underline{\mathbf{\Theta}}^{(i+1)} &= \mathbf{\Theta}^{(i)} - \mu^{(i)} V^{(i)} \, \bar{\mathbf{y}}_i \mathbf{e}_{m^{(i)}}^T \\ \mathbf{\Theta}^{(i+1)} &= \mathcal{P}_{\mathcal{U}}\{\underline{\mathbf{\Theta}}^{(i+1)}\}, \end{aligned} \tag{3.28}$$

where

$$V^{(i)} = \begin{cases} sign(a) & if \;\; m^{(i)} == 1 \\ sign(c) & if \;\; m^{(i)} == 2 \end{cases} \tag{3.29}$$

is defined as above. After the separator update, the next step is unitary projection indicated by (3.28).

The simplified MB-BSS algorithm can be easily implemented by DSPs and ADCs. The hardware design is straightforward. For the all-analog design, we must use the parametric version of this simplified algorithm as we explained before. Therefore parametric update with 4 real parameters will be used with this simplified algorithm (Therefore there will be no unitary projections.), where the simplified version of the parametric algorithm is obtained the same way. Again at every step, threshold comparison is carried out. If update is required, 4 real parameters are updated with the received signal vector $\mathbf{y}_i$. We will not show the details for it as the steps are the same. We will show the simulation results for the simplified parametric MB-BSS algorithm and the analog hardware design of it.

3.4.1 All-Analog Circuit Implementation of the Simplified Parametric MB-BSS

The research carried out for the simplified MB-BSS and parametric MB-BSS algorithms lead to the way for all-analog hardware implementation. All-analog hardware is designed for the simplified parametric MB-BSS algorithm (the combination of the simplified and parametric algorithms as explained before). The required operations are simple sine and cosine functions, adders and multipliers. Also comparators will be used for the threshold comparison. There is no complex operation such as SVD or matrix inverses. A complete hardware implementation is given in Figure 3.3. Reset switches are used once at the beginning of the algorithm to initialize the values of the 4 real parameters to 0. There are two important blocks in the design. One of them is the "Threshold Decision" block. It has two inputs which are the real outputs of the separator. This block determines whether an update is necessary or not by threshold comparison carried by analog comparator circuits. It has two outputs denoted by signZ and Data Select. If an update is not necessary, signZ is set to 0. Otherwise signZ indicates the sign of the maximum magnitude real output of the separator (which is $V^{(i)}$ given by (3.29)), and Data Select is the index of this maximum (which is $m^{(i)}$).

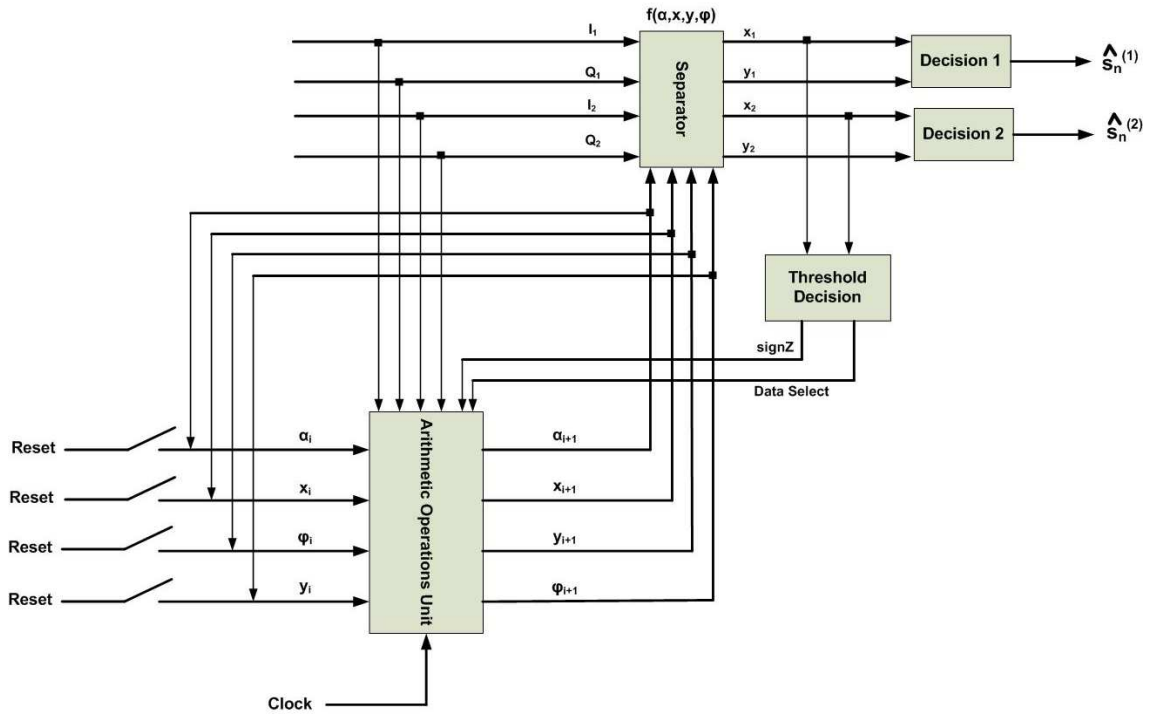


Figure 3.3: All-analog hardware implementation of the overall simplified parametric MB-BSS system

Other important block (Arithmetic Operations Unit) is the heart of the design. Therefore we will give an inside view for it. It is the place where the updates are calculated according to the simplified parametric MB-BSS equations. There are 8 coefficients in Figure 3.4 denoted by $p_1, p_2 ..., p_8$. Let us collect them in the $\mathbf{p}$ vector. The coefficients are adjusted according to the Data Select input.

$$\mathbf{p} = \begin{cases} [\ I_1 \quad I_2 \quad 0 \quad 0 \quad -Q_1 \quad -Q_2 \quad 0 \quad 0\ ] & if\ \ Data\ \ Select\ \ == 1 \\ [\ 0 \quad 0 \quad I_1 \quad I_2 \quad 0 \quad 0 \quad -Q_1 \quad -Q_2\ ] & if\ \ Data\ \ Select\ \ == 2 \end{cases} \tag{3.30}$$

We can see that, complex components are not needed in the overall design. All the sine, cosine, addition, multiplication and threshold comparison operations can be easily implemented with analog circuits.

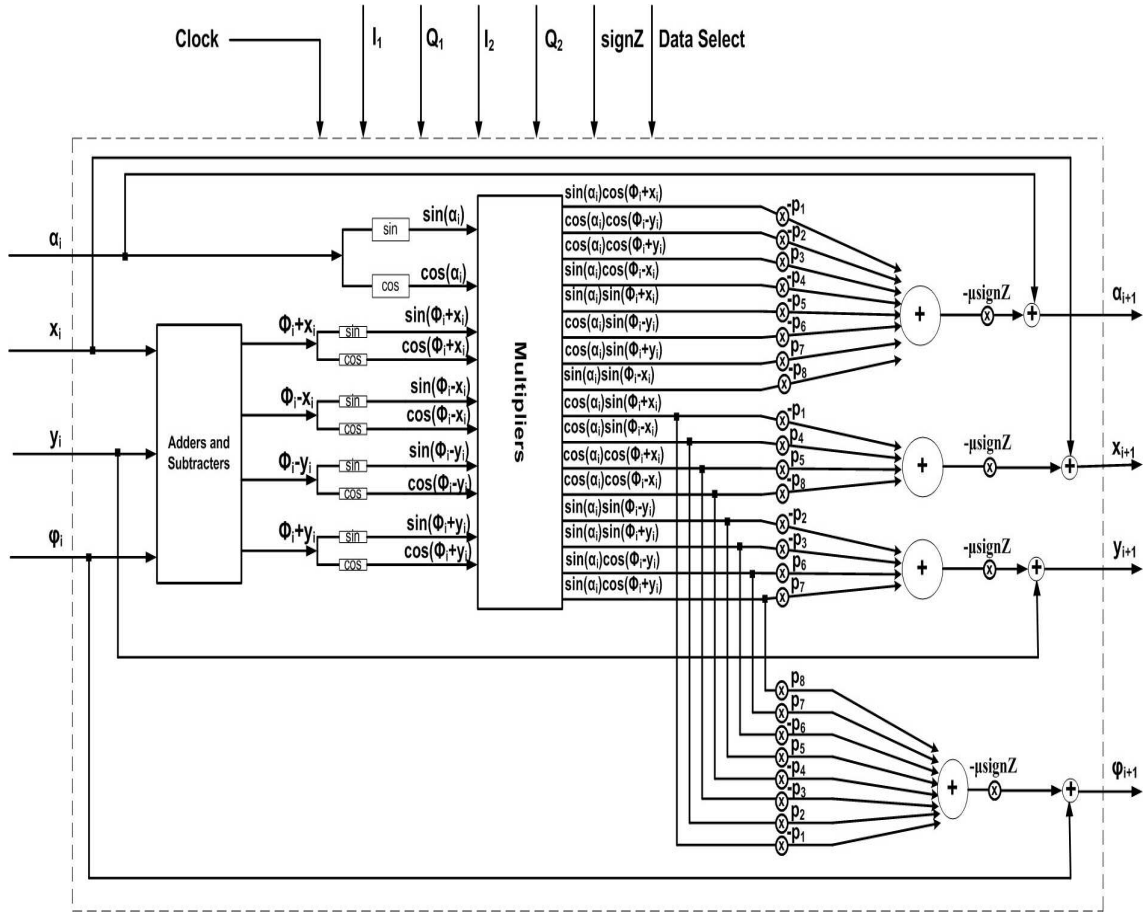


Figure 3.4: Arithmetic operations unit in detail

3.5 Simulation Results of the proposed Algorithms

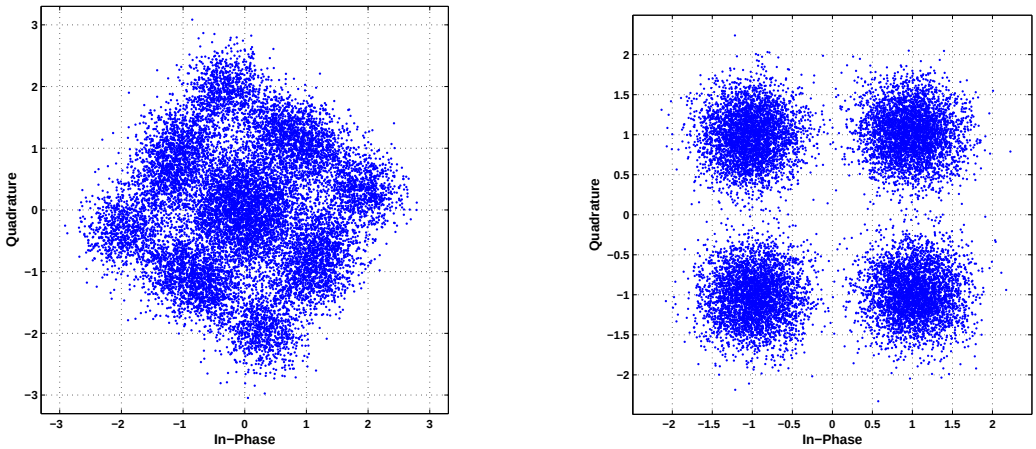
In this section, the simulation results of the proposed blind algorithms will be presented. There are three cases: MB-BSS algorithm, parametric MB-BSS algorithm and simplified parametric MB-BSS algorithm. Simulation results for the simplified parametric MB-BSS are the most important due to its all-analog implementation feature. For the simulations, different SNR values will be tested. However for the simplified parametric MB-BSS, we will only show the results for SNR = 10 dB. The reason is that, in realtime implementation, real SNR values are close to 10 dB for our case (For greater SNR values, algorithm shows better performances for sure.). The simulation results for the normal and parametric MB-BSS will be shown back to back. The results for the simplified parametric MB-BSS will be shown at the end.

3.5.1 Simulation Results of MB-BSS for Moving Window Approach

For the simulations, output of the fiber channel was corrupted by AWGN for two different SNR values. In every group there are 3 figures. First one is the scatter plot of the channel output and second one is the separator output. Third figure is the Signal to Interference Ratio (SIR) convergence curve (The plots are given for one polarization component.). SIR includes inter-source interference (interference due to the other polarization component), ISI and noise. However as told before ISI is very small because symbol rate is low. It is 5 Gsym/s in these simulations. The used constellations at the channel input is 4-QAM, and window lengths are 30. The ADCs were assumed to have a precision of 5 bits. At 5 Gsym/s, 6-bit precision ADCs are commercially available. However even 5-bit precision ADCs work well. We see that SIR reached nearly the maximum value in each cases (The maximum value is the SNR. It cannot be larger than SNR.). For the algorithm to converge, 100-150 iterations are necessary as seen in the plots. This shows that algorithm works very fast. Because 100 window means 3000 symbols and a total time of 0.6 μs at 5 Gsym/s. If we remember that PMD is a random phenomenon changing in millisecond time scales, the MB-BSS algorithm is able to converge much before channel changes. Also SIR and channel output figures show that BSS is absolutely successful to separate the polarization components from each other.

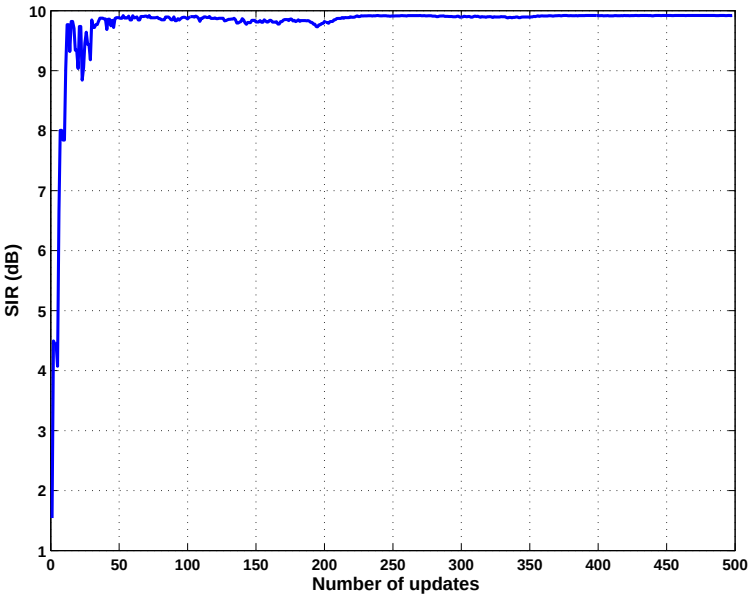
3.5.2 Simulation Results of Parametric MB-BSS for Moving Window Approach

Here we test the performance of the parametric MB-BSS algorithm. The simulation parameters are the same except the window length. Here window lengths were taken as 10. We see that parametric MB-BSS also converges fast. Approximately 500 iterations are enough for convergence which means 5000 symbols. And this lasts 1 μs at 5 Gsym/s data rate. Therefore parametric MB-BSS can be considered to converge a little bit slower than the normal MB-BSS. However there is not a great difference. On the other hand it has reduced complexity than the normal MB-BSS due to the removal of unitary projection. Therefore there is a trade-off between complexity reduction and convergence speed.



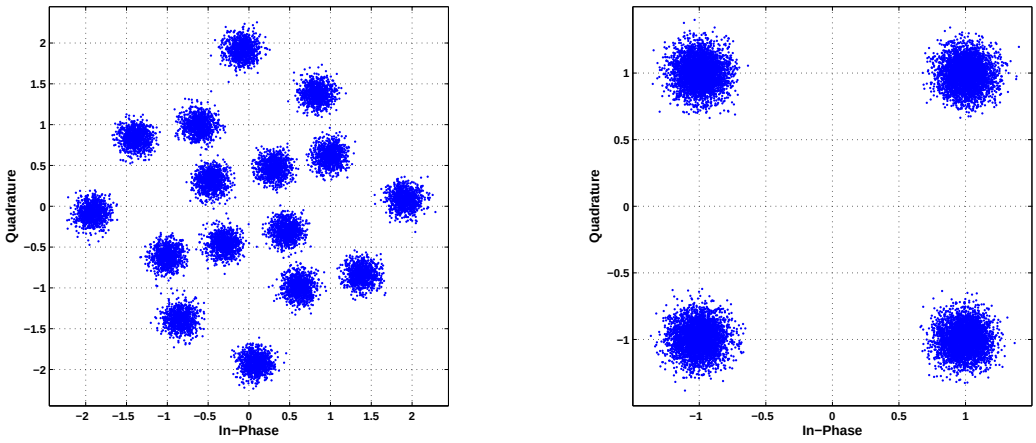
(a)

(b)



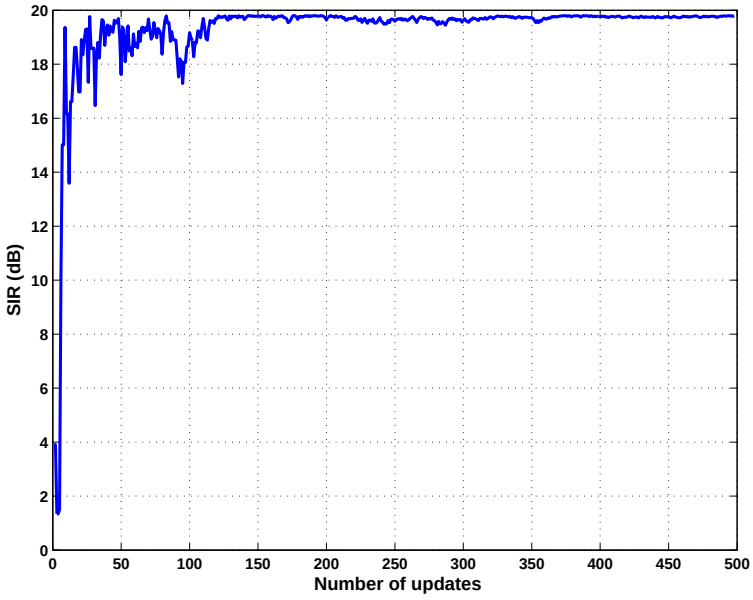
(c)

Figure 3.5: MB-BSS simulation result for SNR = 10 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve.



(a)

(b)



(c)

Figure 3.6: MB-BSS simulation result for SNR = 20 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve.

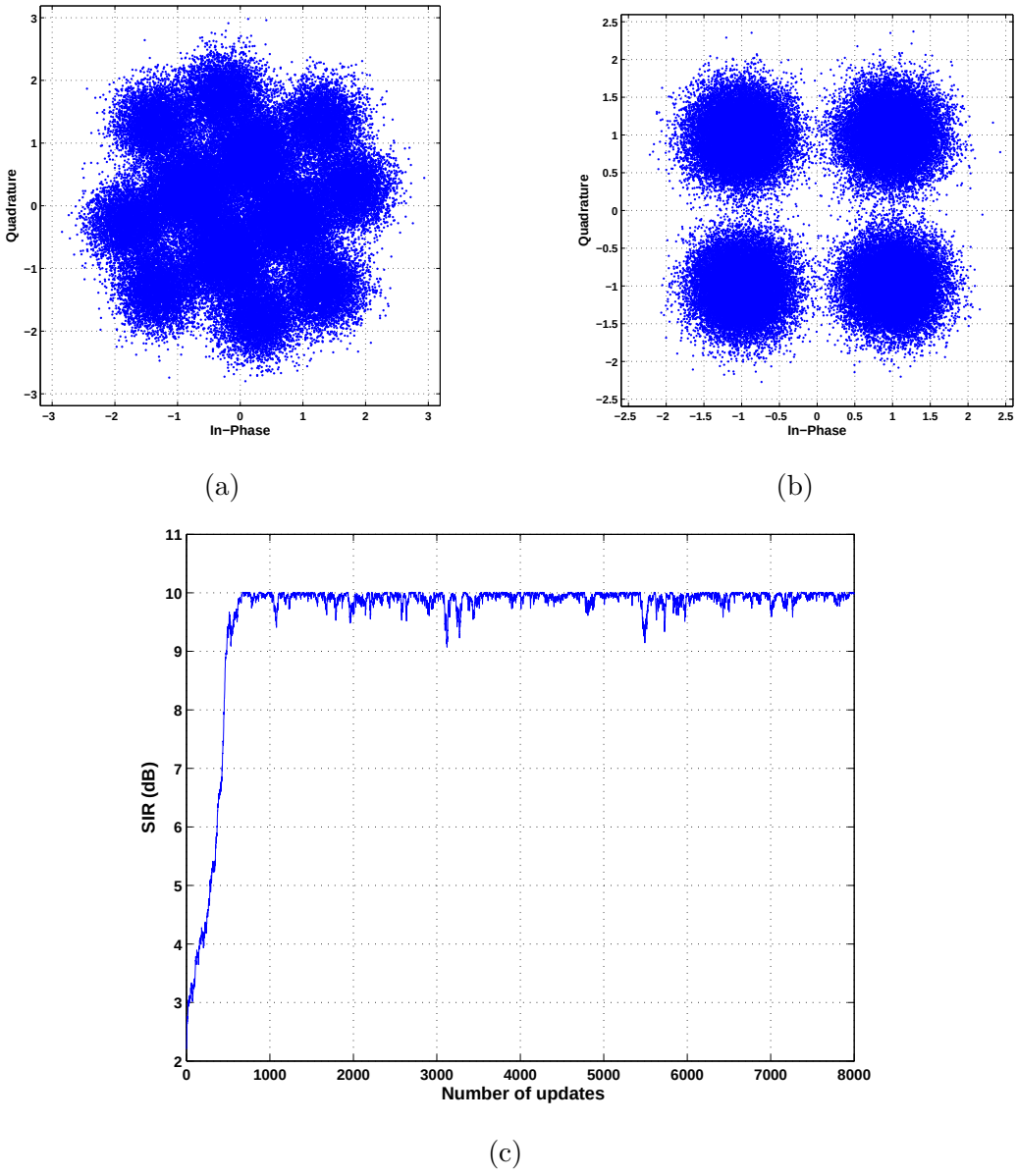


Figure 3.7: Parametric MB-BSS simulation result for SNR = 10 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve.

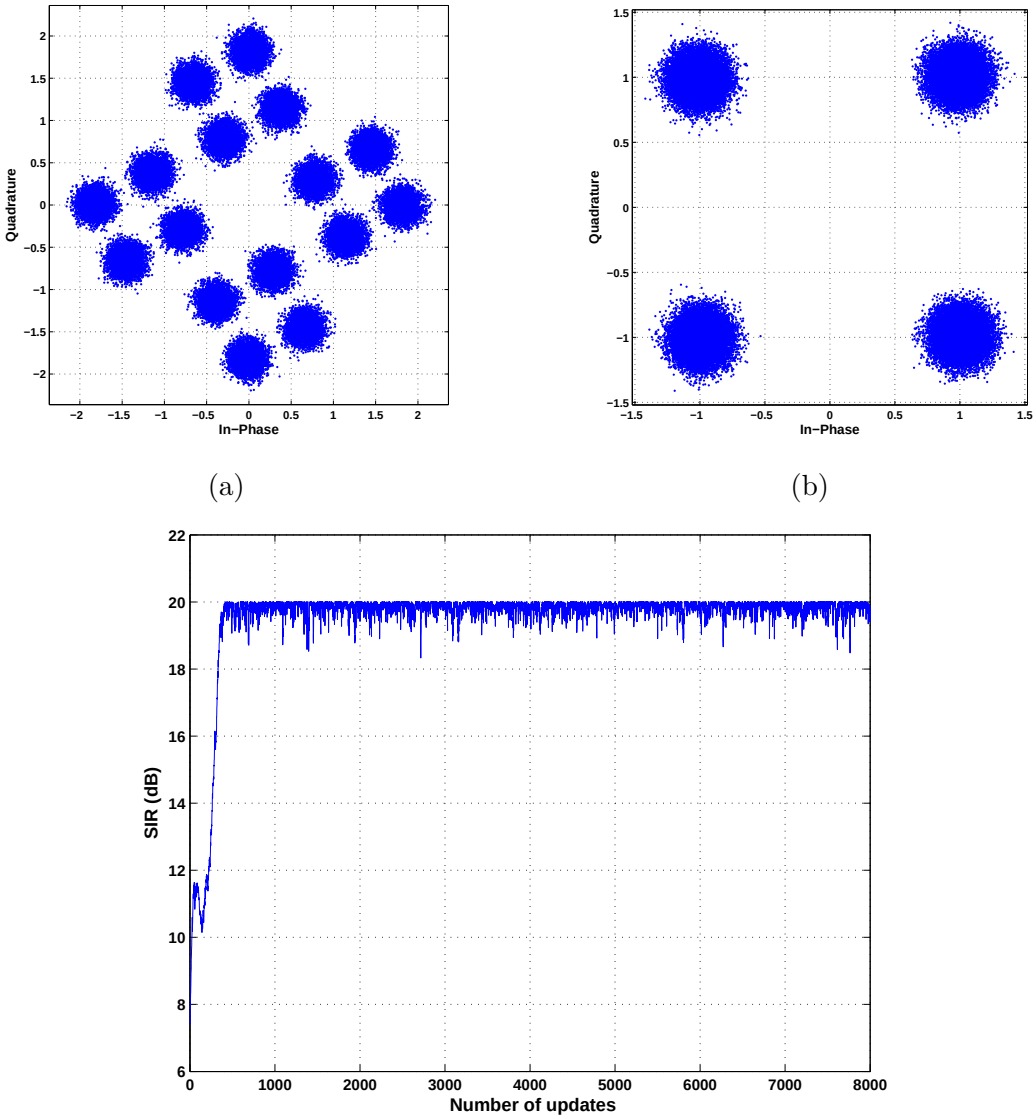
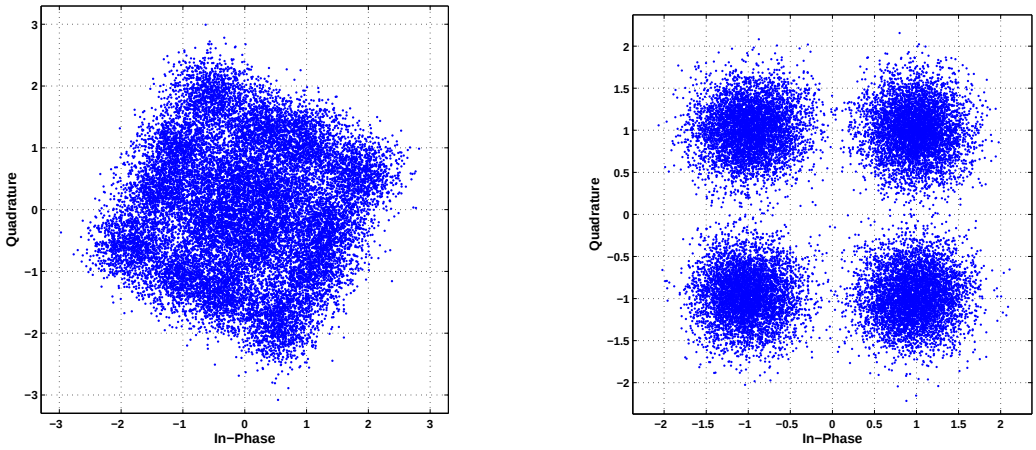


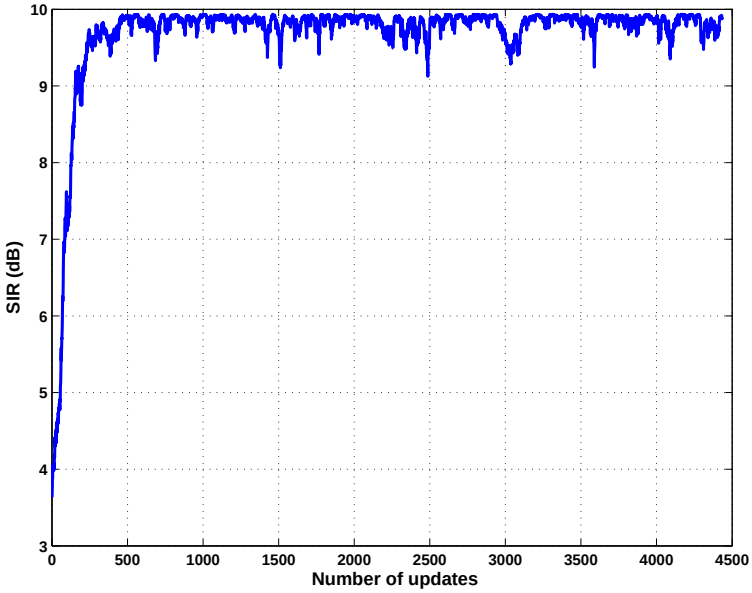
Figure 3.8: Parametric MB-BSS simulation result for SNR = 20 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve.

3.5.3 Simulation Results of the Simplified Parametric MB-BSS

Data rate is again 5 Gsym/s in this simulation. ADCs are not used as we realized the all-analog hardware design. We see that approximately 500 updates are sufficient for convergence. These updates occurred within 760 data duration (Note that there may not be an update at each step due to the threshold comparison.). This data duration corresponds to 0.152 μs at 5 Gsym/s. We can conclude that simplified parametric MB-BSS has also fast convergence behaviors. In addition to its fast convergence, its all-analog implementation feature makes this algorithm extremely useful for fiber optic communication systems.



(a) (b)



(c)

Figure 3.9: Simplified parametric MB-BSS simulation result for SNR = 10 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve.

Chapter 4

BLIND SEPARATION AND DISPERSION COMPENSATION FOR HIGH SYMBOL RATES

The subject of the previous chapter was designing BSS algorithms to extract originally transmitted polarization components at low symbol rates. At these data rates, channel can be assumed to be ISI free. The channel mixes the source signals in space by a linear transfer matrix. The aim is to separate the independent polarization components from each other by a unitary separator. We called this problem as instantaneous BSS. However at high data rates, the assumption about ISI free channel no longer holds. Because higher data rates necessitates the use of wider bandwidths (See Figure 3.2.). And in the wideband schemes the assumption about the frequency flatness of the channel would not be valid and the frequency-selective model of the channel would be more appropriate. The channel varies much over its entire frequency spectrum (The channel is not flat.). Because the channel has a much wider bandwidth now. Therefore this channel will introduce ISI in addition to polarization mixing.

Due to the convolutive nature of such channels, the receiver algorithms should counteract not only the mixing of polarization components in space domain, but also the mixing of polarization components in time domain (due to ISI). Therefore the unitary separator matrix used in the previous chapter is not enough to solve this problem. Instead a MIMO filter of finite order (such as FIR filters) must be used to counteract the additional time dispersion. A general view of an FIR filter of length N is shown in Figure 4.1.

Received signal $\mathbf{x(t)}$ (2×1 vector) is electrically filtered to produce the output signal $\mathbf{y(t)}$ (2×1 vector). This is an N-tap filter, each tap being a 2×2 matrix. The equation for $\mathbf{y(t)}$ is given by:

$$\mathbf{y(t)} = \sum_{k=0}^{N-1} \mathbf{W}_k \, \boldsymbol{x(t - k\Delta T)}, \tag{4.1}$$

where $\mathbf{W}_k$ are the tap weights and ΔT is the tap delay. Usually tap delay is taken as

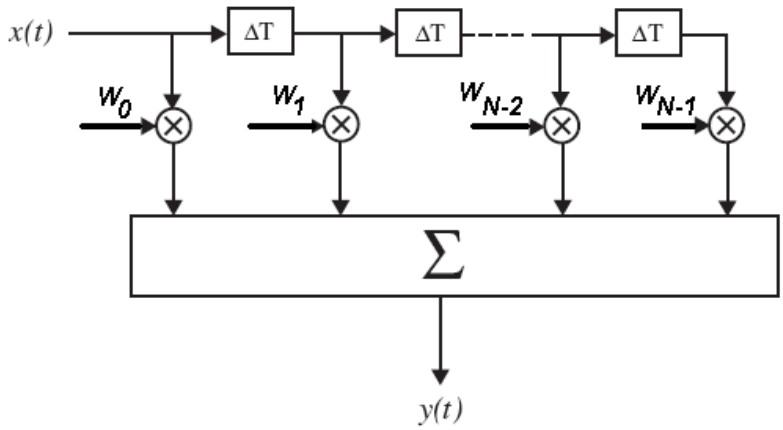


Figure 4.1: Block representation of an FIR filter

symbol period or a fraction of it (fractionally spaced equalizers (FSEs)). In this thesis, tap delays are taken as symbol periods.

Obtaining the separator transfer matrix $\mathbf{W(z)}$ blindly is a non-trivial task. In this chapter, we extend our approach based on MB-BSS to handle convolutive mixtures.

4.1 Convolutive Blind Source Separation

There are various existing approaches proposed for the blind deconvolution of convolutive MIMO channels. As an important result, it is shown in [49] that, the set of all equalizable MIMO channels consists of channels whose transfer functions can be decomposed as a product of an irreducible and a paraunitary transfer function. Therefore $\mathbf{H(z)}$ can be written as:

$$\mathbf{H(z)} \;=\; \mathbf{H}_I(z)\,\mathbf{H}_P(z), \tag{4.2}$$

where $\mathbf{H}_I(z)$ is the irreducible part (see [49] for definition), and $\mathbf{H}_P(z)$ is the paraunitary part where a $p \times p$ transfer matrix $\mathbf{T(z)}$ is paraunitary if $\mathbf{T(z)T}^*(z^{-*}) = \mathbf{I} \;\; \forall z$.

Note that such $\mathbf{T(z)}$ would be unitary on the unit circle, i.e, $\mathbf{T}(e^{j\omega})\mathbf{T}^*(e^{j\omega}) = \mathbf{I} \;\; \forall \omega$.

It is the extension of the instantaneous algorithm aimed to cover the convolutive BSS problem. The starting point of the novel convolutive BSS algorithm is (4.2). The main advantage is again the low computational complexity of the algorithm. The assumption

about the bounded magnitude sources is present again and forms the fundamental of the algorithm.

4.2 Convolutive Blind Source Separation for Bounded Magnitude Sources

In this section we will explain the convolutive MB-BSS algorithm that we use in this thesis. It is the extension of the instantaneous MB-BSS to handle the convolutive case. All the algorithm is explained extensively in [25]. However we will also explain it here. Because it forms the basis of this chapter. However we will skip some of the derivations. Interested readers can look in the reference to see the details. The overall convolutive BSS setup and the convolutive MB-BSS algorithm are explained step by step below.

- The source signals are mixed with the MIMO system with transfer matrix $\mathbf{H(z)}$. $\mathbf{H(z)}$ is a q $\times$ p matrix which implies that this mixing system has p inputs denoted by $s_1(k), s_2(k), ..., s_p(k)$ (bounded, mutually independent and complex symmetric) and q outputs denoted by $y_1(k), y_2(k), ..., y_q(k)$. In our case $q = p = 2$ since we have a 2 $\times$ 2 MIMO channel. In z-transform domain we can write

$$\underbrace{\begin{bmatrix} y_1(z) \\ y_2(z) \\ \vdots \\ y_q(z) \end{bmatrix}}_{\mathbf{y(z)}} = \mathbf{H(z)} \underbrace{\begin{bmatrix} s_1(z) \\ s_2(z) \\ \vdots \\ s_p(z) \end{bmatrix}}_{\mathbf{s(z)}}. \tag{4.3}$$

  Note that $\mathbf{H}$ is not a constant matrix in the convolutive case. It is a general polynomial function, i.e., $\mathbf{H(z)} = \sum_{k=0}^{L} \mathbf{H}_k \, z^k$, which is not a monomial. The dependence of the channel transfer matrix on "z" means that channel varies with frequency which reflects the frequency selective nature of it.

- As stated by (4.2), any equalizable transfer function can be written as a product of an irreducible and a paraunitary transfer function. The aim of the separator $\mathbf{W}^T(z)$ ($\mathbf{W(z)} = \mathbf{W}_{pre}(z)\mathbf{\Theta(z)}$) is to estimate the original independent source signals $\mathbf{s(z)}$ from the observations $\mathbf{y(z)}$. The purpose of $\mathbf{W}_{pre}(z)$ is the adaptive whitening of the observation signals. It compensates the irreducible part of the channel and reduces the

channel to a paraunitary transfer function. Note that our PMD frequency response $\mathbf{U(w)}$ is a unitary matrix for every frequency. The total frequency response with the transmit and receive filters distort the overall unitary structure a little. The combined spectrum of the transmit and receive filters does not introduce any distortion over a broad spectrum. However the distortions are mostly at the tails of the overall spectrum. Because aim of these filters is to limit the bandwidth. Therefore overall transfer function $\mathbf{H(z)}$ is close to paraunitary, i.e, $\mathbf{H}_I(z) = \mathbf{I}$.

As a conclusion, a whitening operator $\mathbf{W}_{pre}(z)$ is not required for our problem. Therefore $\mathbf{W(z)} = \mathbf{\Theta(}\boldsymbol{z}\mathbf{)}$, and overall transfer function of the channel $\mathbf{H(z)}$ is just equal to $\mathbf{H}_P(z)$. The channel output can be updated as:

$$\mathbf{y(z)} \;=\; \mathbf{H}_P(z)\,\mathbf{s(z)}. \tag{4.4}$$

- The convolutive MB-BSS tries to find a paraunitary $\mathbf{\Theta(}\boldsymbol{z}\mathbf{)}$ which will decompose $\mathbf{y(z)}$ as:

$$\mathbf{\Theta}^T(z)\,\mathbf{y(z)} \;=\; \mathbf{\Theta}^T(z)\,\mathbf{H}_P(z)\,\mathbf{s(z)} \;=\; \mathbf{\Lambda(}\boldsymbol{z}\mathbf{)}\;\mathbf{\Phi}\;\boldsymbol{P}\;\boldsymbol{s}(\boldsymbol{z}), \tag{4.5}$$

  where $\mathbf{\Phi}$ and $\mathbf{P}$ again represent the phase and permutation ambiguities respectively as in the instantaneous case. These inherent ambiguities are common in all types of BSS [50]. In (4.5), $\mathbf{\Lambda(}\boldsymbol{z}\mathbf{)}$ is a diagonal matrix with monomial entries corresponding to delays for the two different polarization components. It is given by:

$$\mathbf{\Lambda(}\boldsymbol{z}\mathbf{)} = \begin{bmatrix} z^{d_1} & 0 \\ 0 & z^{d_2} \end{bmatrix}. \tag{4.6}$$

- The aim of the convolutive MB-BSS is to obtain the paraunitary separator iteratively by minimizing a certain cost function. The cost function is the same as in the instantaneous case. The difference from the instantaneous case is that we now have a paraunitary constraint.

$$\begin{aligned} \text{minimize} \quad & sup\;||\mathcal{R}e\{\mathbf{z(k)}\}||_\infty \\ \text{subject to} \quad & \mathbf{\Theta}^H(z^{-*})\mathbf{\Theta}(z) = \mathbf{I}, \end{aligned}$$

where $\mathbf{z(k)}$ is the separator output. The problem is to obtain the FIR paraunitary matrix

$$\mathbf{\Theta}(\boldsymbol{z}) = \sum_{k=0}^{L-1} \mathbf{\Theta}_k z^k \tag{4.7}$$

adaptively from the channel outputs $\mathbf{y(z)}$. One important thing to note is that order of $\mathbf{\Theta}(\boldsymbol{z})$ (*which is* L) must be greater than or equal to the order of the paraunitary channel $\mathbf{H}_P(z)$. The approaches are similar as in the instantaneous case. Therefore we will not explain everything again. If we define the impulse response matrix for the paraunitary operator $\mathbf{\Theta}(\boldsymbol{z})$ as

$$\mathbf{\Upsilon} = \begin{bmatrix} \mathbf{\Theta}_0^T & \mathbf{\Theta}_1^T & \dots & \mathbf{\Theta}_{L-1}^T \end{bmatrix}^T,$$

and the input data matrix in the window of interest as:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y(0)} & \mathbf{y(1)} & \dots & \boldsymbol{y(\Omega-1)} \\ \mathbf{y(-1)} & \mathbf{y(0)} & \dots & \boldsymbol{y(\Omega-2)} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{y(-L+1)} & \mathbf{y(-L+2)} & \dots & \boldsymbol{y(\Omega-L)} \end{bmatrix},$$

then the corresponding separator outputs can be placed in a matrix $\mathbf{Z}$:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z(0)} & \mathbf{z(1)} & \dots & \boldsymbol{z(\Omega-1)} \end{bmatrix} = \mathbf{\Upsilon}^T\, \mathbf{Y}.$$

With these definitions the update of the algorithm is given as:

$$\begin{aligned} \underline{\mathbf{\Upsilon}}^{(i+1)} &= \mathbf{\Upsilon}^{(i)} - \mu^{(i)} sign(\mathcal{R}e\{\mathbf{Z}^{(i)}_{m^{(i)},n^{(i)}}\})\bar{\mathbf{Y}}_{:,n^{(i)}}\mathbf{e}^T_{m^{(i)}} \\ \mathbf{\Upsilon}^{(i+1)} &= \mathcal{M}_{\mathcal{P}}\{\underline{\mathbf{\Upsilon}}^{(i+1)}\}, \end{aligned} \tag{4.8}$$

where $\mathcal{M}_{\mathcal{P}}$ is the paraunitary mapping operator. This projection is not a simple operation as in the instantaneous case. There is no existing closed form solution for minimum distance projection to paraunitary set. Instead, we propose some alternative mapping techniques for the purpose of enforcing paraunitary constraint.

4.3 Mapping to Paraunitary Set

The enforcement of paraunitary constraint is the vital part of the convolutive BSS algorithms. In [25], a new method called "method of alternating projections (MAP)" is intro-

duced for paraunitary mapping. The algorithm is explained in detail in the same paper. However we will just explain the basic idea here. After that we will introduce our modifications to this algorithm, and then we will compare their performances in terms of convergence speed and algorithm complexity.

4.3.1 Method of Alternating Projections (MAP) Based Paraunitary Mapping

The MAP based paraunitary mapping method and our variations on it are based on the fundamental fact summarized by the following theorem [51]:

Theorem 1: A $p \times p$ transfer function ($p = 2$ in our case) $\mathbf{\Theta(z)} = \sum_{k=0}^{L-1} \mathbf{\Theta}_k z^k$ is paraunitary if and only if its $Lp \times (2L-1)p$ block convolution matrix

$$T_L(\mathbf{\Theta(z)}) = \begin{bmatrix} \mathbf{\Theta}_0 & \mathbf{\Theta}_1 & \mathbf{\Theta}_2 & \dots & \mathbf{\Theta}_{L\text{-}1} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{\Theta}_0 & \mathbf{\Theta}_1 & \dots & \mathbf{\Theta}_{L\text{-}2} & \mathbf{\Theta}_{L\text{-}1} & 0 & \dots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \dots & \mathbf{\Theta}_0 & \dots & \dots & \mathbf{\Theta}_{L\text{-}2} & \mathbf{\Theta}_{L\text{-}1} \end{bmatrix}$$

has orthonormal rows, i.e.,

$$T_L(\mathbf{\Theta(z)}) T_L(\mathbf{\Theta(z)})^H = \mathbf{I}. \tag{4.9}$$

According to this theorem, $\mathcal{M}_{\mathcal{P}}$ operator should force $T_L(\mathbf{\Theta(z)})$ to have orthonormal rows. If the following subsets are defined as:

$\boldsymbol{A}_1 = \{\mathbf{A} \in C^{Lp\times(2L-1)p} \ \ |\mathbf{A}\mathbf{A}^H = \mathbf{I}\}$, (Stiefel Manifold)

and

$\boldsymbol{A}_2 = \{\mathbf{A} \in C^{Lp\times(2L-1)p} \ \ |$ $\mathbf{A}$ is a $Lp \times (2L-1)p$ block convolution matrix with $p \times p$ blocks $\}$

then the set

$$\mathbf{A} = \mathbf{A}_1 \bigcap \mathbf{A}_2 \tag{4.10}$$

is the set of block convolution matrices with orthonormal rows which is the set of block convolution matrices corresponding to the paraunitary operators.

Since $\mathbf{A}$ is the intersection of $\mathbf{A}_1$ and $\mathbf{A}_2$, the mapping $\mathcal{M}_{\mathcal{P}}$ can be formulated based on the alternating projection procedure applied between sets $\mathbf{A}_1$ and $\mathbf{A}_2$. Since we do not have an explicit formula for projecting onto $\mathbf{A}$, we apply alternating projection between these sets to get closer to their intersection. Projection onto these sets have explicit formulas. Call $\mathcal{P}_{\mathbf{A}_1}$ as the projection onto set $\mathbf{A}_1$ and $\mathcal{P}_{\mathbf{A}_2}$ as the projection onto set $\mathbf{A}_2$. $\mathcal{P}_{\mathbf{A}_1}$ (unitary projection) was already given before in the previous chapter, and $\mathcal{P}_{\mathbf{A}_2}$ is a very simple process. Interested readers can see [25] for the implementation of $\mathcal{P}_{\mathbf{A}_2}$.

4.3.2 Paraunitary Mapping Approach on Complex Stiefel Manifold with Gradient Descent

The approach is different from MAP in the sense that instead of carrying out alternating projections between two sets, we move on one set by getting closer to the other set. As an alternative to MAP method described above, we propose two different methods for mapping to paraunitary matrices. These methods are based on gradient searches in one of the $\mathbf{A}_1$, $\mathbf{A}_2$ sets where the distance to the other set is defined as cost function. As the first alternative, we consider the search space as the Stiefel manifold (i.e. the set $\mathbf{A}_1$) and the cost function to be minimized is:

$$f(\mathbf{X}^{(i)}) = \|\mathbf{X}^{(i)} - \mathcal{P}_{\mathbf{A}_2}(\mathbf{X}^{(i)})\|_F^2. \tag{4.11}$$

With this cost function, the aim is to minimize the distance between $\mathbf{X}^{(i)}$ and the projection of $\mathbf{X}^{(i)}$ to the set $\mathbf{A}_2$. The optimization algorithm is realized on the complex Stiefel manifold (unitary matrix set, $\mathbf{A}_1$), and for this aim gradient descent or Newton algorithms can be used. However Newton is not preferable due to its much higher computational complexity than gradient descent. Therefore gradient descent is chosen for the minimization of this cost function. The steps in the algorithm are given below:

1. Gradient $\mathbf{D}_X$ calculation of the cost function:

$$\mathbf{D}_X^{(i)} = \nabla f(\mathbf{X}^{(i)}) = \mathbf{X}^{(i)} - \mathcal{P}_{\mathbf{A}_2}(\mathbf{X}^{(i)}). \tag{4.12}$$

2. Choosing a suitable step size μ : An alternative step size providing fast convergence speed known as relaxation step size rule

$$\mu^{(i)} = c\frac{f(\mathbf{X}^{(i)}) - f_{opt}}{\|\mathbf{D}_X^{(i)}\|_F^2} \tag{4.13}$$

can be used. Here "c" is a constant between 1 and 2, f_{opt} is the the optimum value of the function. In many problems f_{opt} is not known however in our case it is known and equal to 0. As can be seen in (4.13), numerator and denominator cancel each other in this problem. As a result $\mu^{(i)} = c$.

3. Updating $\mathbf{X}$: With the calculated $\mu^{(i)}$ and $\mathbf{D}_X^{(i)}$, $\mathbf{X}$ is updated:

$$\underline{\mathbf{X}}^{(i+1)} = \mathbf{X}^{(i)} - \mu^{(i)}\mathbf{D}_X^{(i)}. \tag{4.14}$$

4. As $\mathbf{X}$ moves away from Stiefel manifold due to the gradient descent direction, it is projected onto the Stiefel manifold again:

$$\mathbf{X}^{(i+1)} = \mathcal{P}_{\mathbf{A}_1}(\underline{\mathbf{X}}^{(i+1)}). \tag{4.15}$$

5. If $f(\mathbf{X}^{(i+1)}) < \epsilon$ or the maximum iteration number is reached, processing stops. Otherwise iterations continue again with the 1^{st} step.

If the computational complexity is considered, most of the computational load comes from the 4^{th} step. Because projection needs SVD which is a hard process. Algorithm complexity is approximately about $28L^3p^3$.

4.3.3 Paraunitary Mapping Approach on Block Convolution Matrix Space

In this alternative, the search is performed over the set of block convolution matrices and the goal is to minimize the distance to the Stiefel manifold. Therefore the cost function we are trying to minimize is:

$$f(\mathbf{X}^{(i)}) = \|\mathbf{X}^{(i)}\mathbf{X}^{(i)^H} - \mathbf{I}\|_F^2. \tag{4.16}$$

We provide both gradient descent and Newton approaches below, however we will provide simulation results only for the gradient descent.

Gradient Descent Method:

1. Gradient $\mathbf{D}_X$ calculation of the cost function:

$$\mathbf{D}_X^{(i)} = \nabla f(\mathbf{X}^{(i)}) = 2\,(\mathbf{X}^{(i)}\mathbf{X}^{(i)^H}\mathbf{X}^{(i)} - \mathbf{X}^{(i)}). \tag{4.17}$$

2. Choosing a suitable step size μ : A step size given in (4.13) can be used. Again f_{opt} is equal to 0.

3. Updating $\mathbf{X}$: With the calculated $\mu^{(i)}$ and $\mathbf{D}_X^{(i)}$, $\mathbf{X}$ is updated:

$$\underline{\mathbf{X}}^{(i+1)} = \mathbf{X}^{(i)} - \mu^{(i)}\mathbf{D}_X^{(i)}. \tag{4.18}$$

4. As $\mathbf{X}$ moves away from block convolution matrix space due to the gradient descent direction, it is projected onto the convolution matrix space again:

$$\mathbf{X}^{(i+1)} = \mathcal{P}_{\mathbf{A}_2}(\underline{\mathbf{X}}^{(i+1)}). \tag{4.19}$$

5. If $f(\mathbf{X}^{(i+1)}) < \epsilon$ or the maximum iteration number is reached, processing stops. Otherwise iterations continue again with the 1^{st} step.

Computational complexity of this algorithm is less than the previous one. It is approximately $6L^3p^3$. Most of the complexity is due to the gradient calculation this time. Because projection onto $\mathbf{A}_2$ ($\mathcal{P}_{\mathbf{A}_2}$) is an easy process.

Newton Method:

1. Gradient $\mathbf{D}_X$ calculation of the cost function:

$$\mathbf{D}_X^{(i)} = \nabla f(\mathbf{X}^{(i)}) = 2\,(\mathbf{X}^{(i)}\mathbf{X}^{(i)^H}\mathbf{X}^{(i)} - \mathbf{X}^{(i)}). \tag{4.20}$$

2. Hessian of the cost function must be calculated for Newton algorithm. However a second derivative results in a 4 dimensional matrix structure. Therefore we first vectorize the gradient calculated in the previous step. And then take the derivative with respect to vec($\mathbf{X}$).

$$\frac{\partial f(\mathbf{X}^{(i)})}{\partial vec(\mathbf{X})} = 2\ vec\ (\mathbf{X}^{(i)}{\mathbf{X}^{(i)}}^H\mathbf{X}^{(i)} - \mathbf{X}^{(i)}), \tag{4.21}$$

and

$$\begin{aligned}\frac{\partial^2 f(\mathbf{X}^{(i)})}{\partial vec^2(\mathbf{X})} \ = \ 2\,[\ \ &(\mathbf{X}^{(i)H}\mathbf{X}^{(i)}) \otimes \mathbf{I}_{Lp} \ + \ \mathbf{I}_{(2L-1)p} \otimes \ (\mathbf{X}^{(i)}\mathbf{X}^{(i)H}) \\ &+ \ (\mathbf{X}^{(i)} \otimes \mathbf{X}^{(i)H})_{(n)} \ - \ \mathbf{I}_{Lp\,(2L-1)p}\ \ ],\end{aligned} \tag{4.22}$$

where $(\mathbf{X}^{(i)} \otimes \mathbf{X}^{(i)H})_{(n)} \ = \ \mathbf{U}^T(\mathbf{X}^{(i)} \otimes \mathbf{X}^{(i)H})$.

The task of $\mathbf{U}$ is to rearrange the rows of the $(\mathbf{X}^{(i)} \otimes \mathbf{X}^{(i)H})$. For a complete description of $\mathbf{U}$, see [52]. As can be recognized, $\frac{\partial^2 f(\mathbf{X}^{(i)})}{\partial vec^2(\mathbf{X})}$ is an $Lp\ (2L-1)p \times Lp\ (2L-1)p$ matrix. Let us call it $\mathbf{B}^{(i)}$.

3. Newton direction is calculated:

$$\mathbf{S}^{(i)} \ = \ {\mathbf{B}^{(i)}}^{-1} vec(\mathbf{D}_X^{(i)}). \tag{4.23}$$

4. The Newton direction $\mathbf{S}^{(i)}$ is a vector now. To update $\mathbf{X}$, inverse vectorization is carried out on $\mathbf{S}^{(i)}$ to transform it into a matrix with suitable dimensions:

$$\mathbf{W}^{(i)} \ = \ ivec(\mathbf{S}^{(i)}). \tag{4.24}$$

5. Updating $\mathbf{X}$:

$$\underline{\mathbf{X}}^{(i+1)} = \mathbf{X}^{(i)} - \mathbf{W}^{(i)}. \tag{4.25}$$

6. As $\mathbf{X}$ moves away from block convolution matrix space due to Newton direction, it is projected onto the block convolution matrix space again:

$$\mathbf{X}^{(i+1)} = \mathcal{P}_{\mathbf{A}_2}(\underline{\mathbf{X}}^{(i+1)}). \tag{4.26}$$

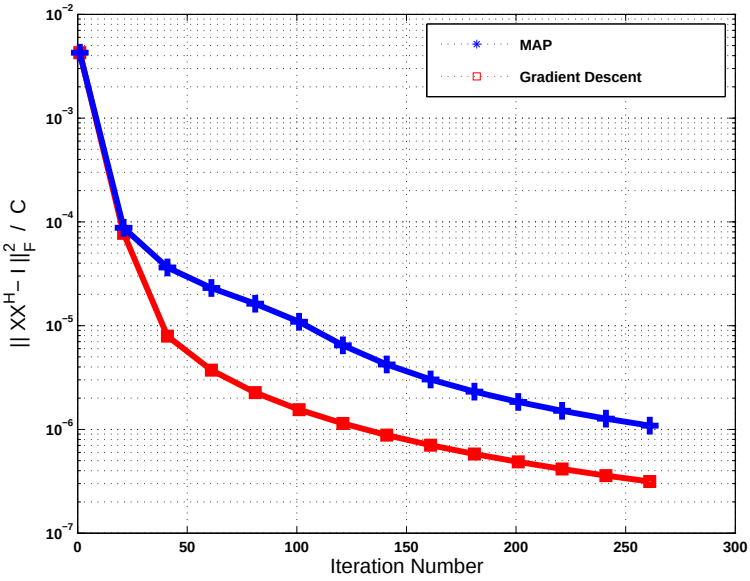
7. If $f(\mathbf{X}^{(i+1)}) < \epsilon$ or the maximum iteration number is reached, processing stops. Otherwise iterations continue again with the 1^{st} step.

As it is clear, this algorithm has a huge computational complexity. Therefore it is not practical to use it in a realtime application.

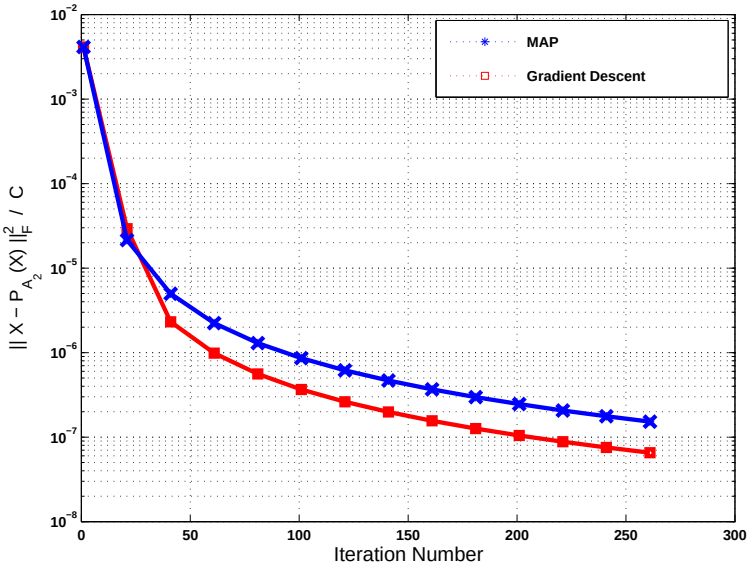
4.4 Simulation Results for Paraunitary Mapping Algorithms

In this section, we show the simulation results for the newly developed algorithms and compare them with MAP. To compare the algorithm complexities, MAP and optimization on the Stiefel manifold have approximately same complexities (about $28L^3p^3$). However optimization on the block convolution matrix space has a complexity of about $6L^3p^3$. Therefore from this aspect, optimization on block convolution matrix space with gradient descent is much more advantageous. For their convergence speeds, we compare them by using the following simulation plots. Gradient descent on Stiefel manifold and gradient descent on block convolution matrix space algorithms are not directly compared with each other. Because their cost functions are different. We compare each with MAP. And also the cost functions are normalized by the product of matrix dimensions which is C.

From the plots, we see that gradient descent type algorithms show better convergence behaviors. Among them gradient descent on block convolution matrix space shows the best performance from convergence speed and algorithm complexity aspects.



(a)



(b)

Figure 4.2: Comparison of convergence speeds of MAP and the proposed algorithms (a) Convergence curves for MAP and gradient descent on block convolution matrix space, (b) Convergence curves for MAP and gradient descent on Stiefel manifold.

4.5 Simulation Results for Convolutive MB-BSS Algorithm

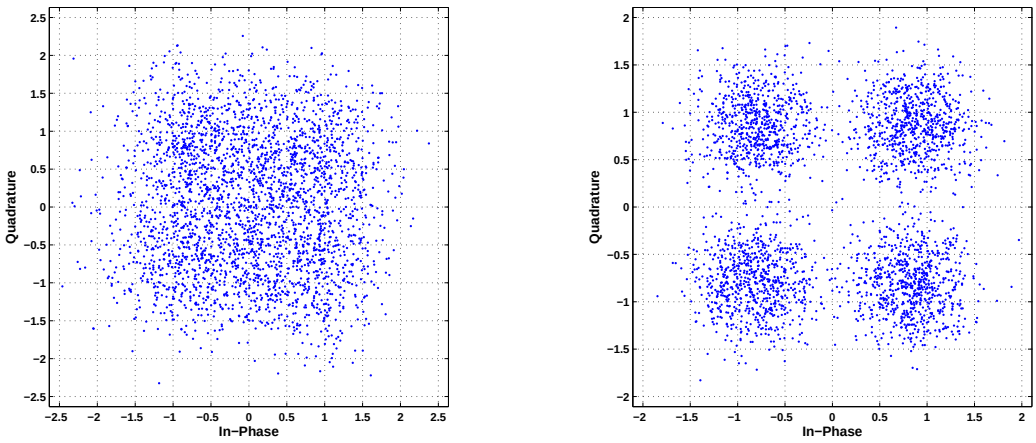
After explaining the paraunitary projection and demonstrating their simulation results, now the simulation results of the overall convolutive MB-BSS algorithm is given. The simulation consists of updating the FIR matrix and then projecting it onto the paraunitary set. For paraunitary mapping, optimization on the block convolution matrix space with gradient descent algorithm is used due to its advantages explained before. Data rate is 60 Gsym/s, ADC precision is taken as 6 bits, and separator filter length L is taken as 5. For the same reasons explained in the previous chapter, we will use moving window approach and show the simulations only for it. Window lengths are taken as 1000. For reducing algorithm complexity, paraunitary mapping is applied at every 5^{th} new window. Also paraunitary mapping is an iterative process and we have taken the number of paraunitary iterations as 100. The simulation results are shown in Figure 4.3 and Figure 4.4 (The results are for one polarization component again.):

From these figures, we can conclude that the convolutive MB-BSS algorithm succeeded in dispersion compensation and polarization separating as can be seen from the separator output figures. For the algorithm to converge, 300-400 iterations are enough depending on the SNR value as seen in the plots. Note that 400 iterations mean $400 \times 1000 = 400.000$ symbols and a total time of 6.67 μs at 60 Gsym/s which is a relatively short duration especially compared to the millisecond scale variations of the fiber channel due to PMD.

4.6 Parametric Convolutive MB-BSS Algorithm

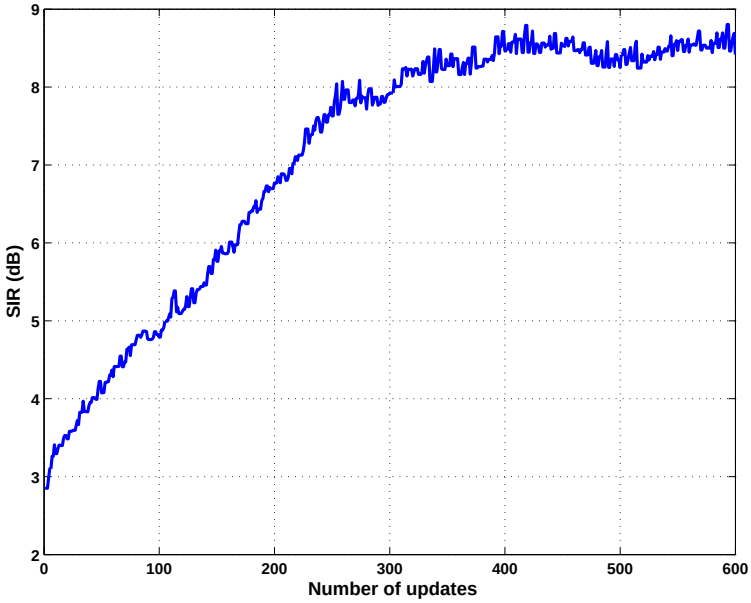
We have also developed a parametric version of the convolutive MB-BSS algorithm to get rid of the paraunitary constraint, since the paraunitary mapping operation is a difficult and time consuming process. To avoid the paraunitary mapping, we must move on the paraunitary surface. The parametrization of paraunitary transfer function is obtained based on the factorization proposed in [51] which states that any paraunitary operator $\mathbf{\Theta(z)}$ of order L can be written as:

$$\mathbf{\Theta(z)} = \mathbf{Q}(\mathbf{I} - \mathbf{v}_1\mathbf{v}_1^* + z^{-1}\mathbf{v}_1\mathbf{v}_1^*)(\mathbf{I} - \mathbf{v}_2\mathbf{v}_2^* + z^{-1}\mathbf{v}_2\mathbf{v}_2^*)...(\mathbf{I} - \mathbf{v}_{L-1}\mathbf{v}_{L-1}^* + z^{-1}\mathbf{v}_{L-1}\mathbf{v}_{L-1}^*), \tag{4.27}$$



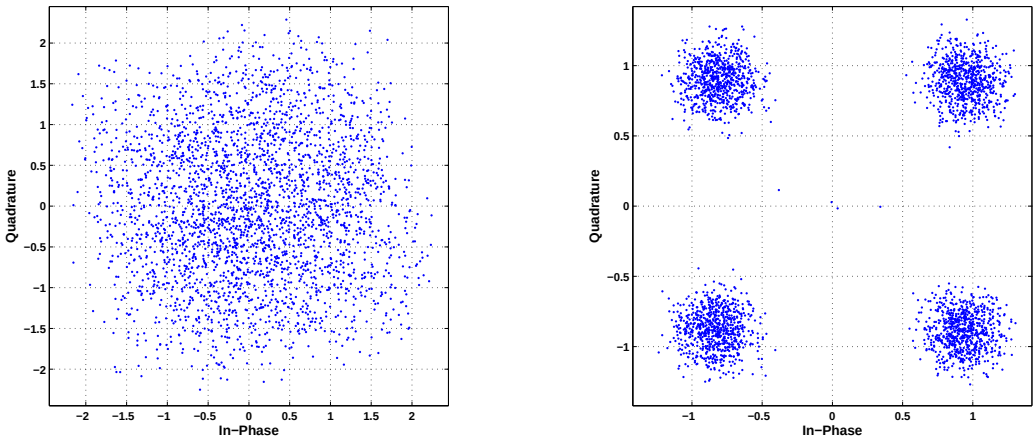
(a)

(b)



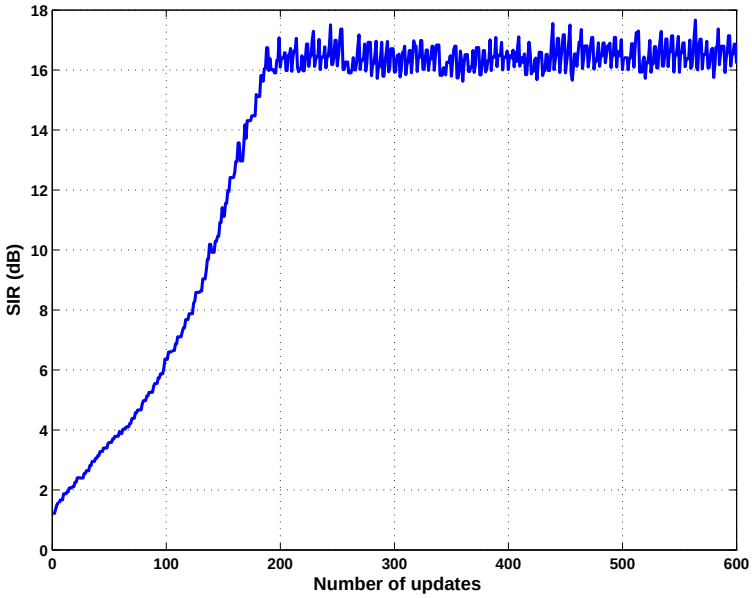
(c)

Figure 4.3: Convolutive MB-BSS simulation result for SNR = 10 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve.



(a)

(b)



(c)

Figure 4.4: Convolutive MB-BSS simulation result for SNR = 20 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve.

where v_i's are any general $p \times 1$ (p = 2 in our case) complex vectors with unit norm, and $\mathbf{Q}$ is any $p \times p$ unitary matrix. Here $\mathbf{Q}$ can be parameterized with 4 real parameters as we did in the previous chapter. For unit norm 2×1 complex $\mathbf{v}_i$ vectors we can write:

$$\mathbf{v}_i \;=\; \begin{bmatrix} \cos(\beta_i)e^{j\gamma_i} \\ \sin(\beta_i)e^{j\tau_i} \end{bmatrix}, \tag{4.28}$$

i.e., we need 3 real parameters β, γ and τ to parametrize each $\mathbf{v}_i$ vector. For simplicity we have taken the separator length L as 2. Therefore only first two components of the overall multiplication in (4.27) are present in the parametrization:

$$\boldsymbol{\Theta(z)} = \mathbf{Q}(\,\mathbf{I} \;-\; \mathbf{v}_1\mathbf{v}_1^* \;+\; z^{-1}\mathbf{v}_1\mathbf{v}_1^*). \tag{4.29}$$

Because even a filter length of 2 results in 7 real parameters (4 parameter from $\mathbf{Q}$, 3 parameter from the unit norm complex vector). Every unit norm complex vector adds 3 different parameters. The subgradient update is performed in a similar way we have done in the previous chapter by chain rule. First the subgradient with respect to $\boldsymbol{\Theta(z)}$ is carried out with (4.8). Next the derivative of $\boldsymbol{\Theta(z)}$ is taken with respect to the 7 real parameters. The parametrization implementation was completely explained in the previous chapter. We have a very similar case here. Therefore we will not explain the steps any further here. The details will be provided in the Appendix.

4.7 Simulation Results for Parametric Convolutive MB-BSS Algorithm

Plots are seen in the next page. The same setup values were used with the normal convolutive MB-BSS. Only difference is that data rate is slower in this simulation (30 Gsym/s) as the separator filter length (L) is 2. For higher data rates, greater separator lengths must be used. However even a filter length of 5 results in 16 real parameters ($4 \;+\; 3 \times 4$) meaning more processing time. There is a trade-off between. The result is given for SNR = 20 dB. Results indicate the convergence of the algorithm. However to make a more reliable comparison with the normal convolutive MB-BSS algorithm, data rates must be equal. For this reason, research is going on this subject to extend the parametric simulation with greater separator lenghts.

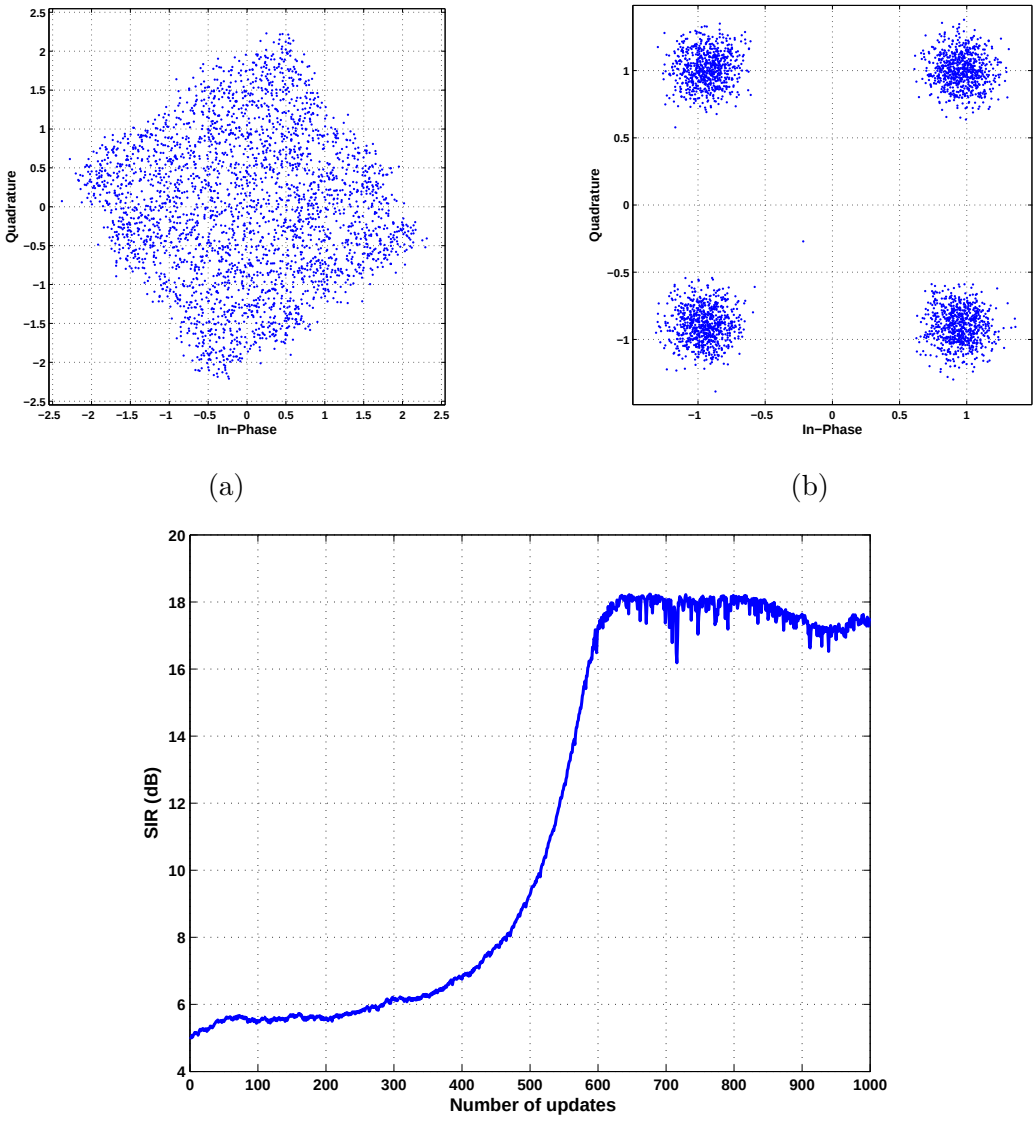


Figure 4.5: Parametric convolutive MB-BSS simulation result for SNR = 20 dB (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve.

Chapter 5

TRACKING OF TIME-VARYING FIBER CHANNELS

In this chapter, tracking performance of the instantaneous and convolutive MB-BSS algorithms will be explored. Tracking is an important issue in channel equalization algorithms. In fiber optic communications, channel is not stable all the time due to the time variations of PMD. Especially long-haul transoceanic fiber optic systems are more effected from environmental changes since they pass through many different countries, climates and oceans. Therefore PMD is not constant all the time. Instead it changes in millisecond time scales as we have stated before. For this reason, digital equalization systems used for fiber optic systems must be adaptive. They must be able to track the variations in the channel.

Digital equalization and/or separation process can be divided into two stages. First stage is the eye-opening stage. At the beginning of this stage, the separator outputs are completely corrupted such that no reliable decisions about transmitted symbols can be made based on them. At the end of the eye-opening stage, the separator outputs are close to original symbols transmitted and symbols can be more reliably detected by the decision device. The MB-BSS algorithms we introduced in the previous chapters are targeting the eye-opening stage and their non-decision directed nature and successful convergence.

The stage following the eye-opening stage is tracking. In this stage, the update of the adaptive parameter is slower than the eye-opening stage and the goal is to track variations in the channel. Since the decisions are more reliable at this stage, these decisions can also be exploited via use of a decision directed algorithm, such as LMS. However, our goal in this chapter would be to illustrate the tracking performance of the MB-BSS algorithm and to show that MB-BSS algorithm can also be used in the tracking mode successfully. Although switching to decision directed mode in the tracking stage is a logical choice, sticking with MB-BSS algorithm for both stages (potentially for reduced receiver complexity) is also a viable alternative.

5.1 Physical and Simulation Model for Time-Variations in the Fiber Channel

PMD was modeled as a concatenation of many polarization-maintaining fibers with varying group delays and rotations of the principal axes. To reflect the random nature of PMD, the rotation and phase angles of each section were chosen randomly over uniform distributions in (2.2). This simulation setup is run once to model the initial behavior of the channel. To include the dynamically changing nature, the initial channel is used. Because it is more probable that the phase and rotation angles for the next simulation will be correlated with their first values. Therefore to add a small variation to the channel, these rotation and phase angles are distorted a little around their initial values. The amount of these distortions determine the overall change in the channel. If large distortions are applied, new channel can be completely different from the first one. In such a condition, the algorithm would have to start from the beginning. In our tracking simulations, small distortions are applied. As a result of this, we can investigate how fast the proposed blind algorithms can track the changes.

5.2 Tracking Performance of Proposed Blind Algorithms

The simulations will be carried out for instantaneous MB-BSS and convolutive MB-BSS algorithms respectively. For the instantaneous MB-BSS, two cases will be explored. First case deals with impulsive channel change. In other words, the channel variation is not smooth, but changes suddenly at the 350^{th} window. Simulations will show the tracking capability for this sudden change. Second case deals with smooth channel variations. In other words, the channel change does not take place suddenly. It changes with every new window, or within 3 windows (Two cases will be explored.). Therefore this time the simulations will demonstrate the tracking capability of the algorithm for continuous channel changes. For the convolutive MB-BSS, only impulsive channel change will be explored.

5.2.1 Simulation Results for Instantaneous MB-BSS

For the simulation parameters, data rate is chosen as 6 Gsym/s, window length is 200, SNR is 20 dB and ADC precision is 6 bits. For the impulsive case, until the 350^{th} window, the channel is kept constant. After that, the channel transfer function is changed. Figure 5.1.a

reflects the change in the channel. From the SIR curve, it is obvious that instantaneous MB-BSS algorithm compensates the sudden change very quickly in just 22 (372-350) windows. This corresponds to a time duration of 0.73 μs.

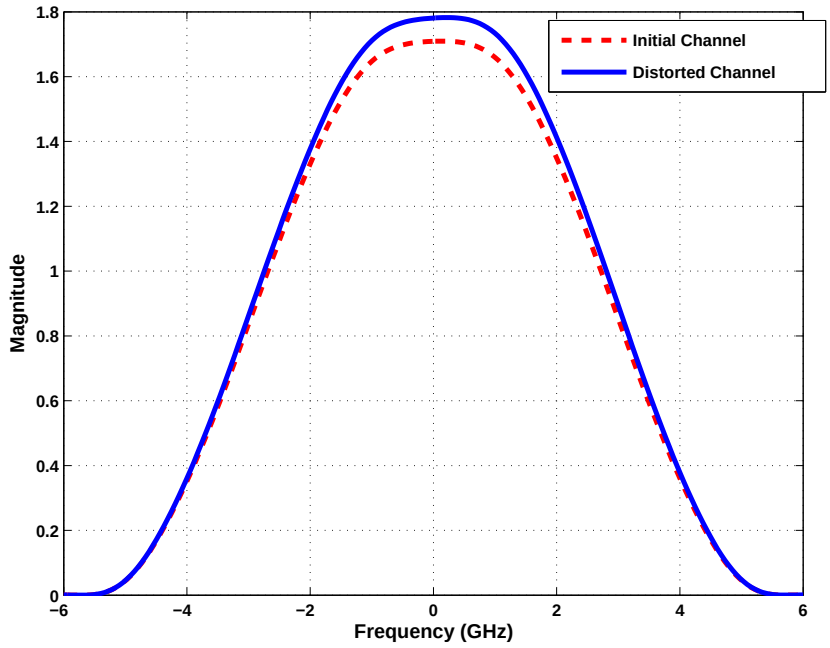
For the smooth variation case, we will apply channel change at every new window in the first simulation, and every third window at the other simulation. In reality, PMD does not change so fast. It changes in millisecond time scales. However we change it very often to show the tracking capability of the algorithm. To talk about the channel variations, we cannot show the initial and distorted channel plots this time since we have many new channels now. The maximum frobenius norm change normalized with the initial channels frobenius norm may give sufficient idea. If we assume that we have a total of S channels (including the initial channel), then

$$maxChange = \max_{i=2,3,...S} \frac{||\mathbf{H}_i - \mathbf{H}_1||_F}{||\mathbf{H}_1||_F} \tag{5.1}$$

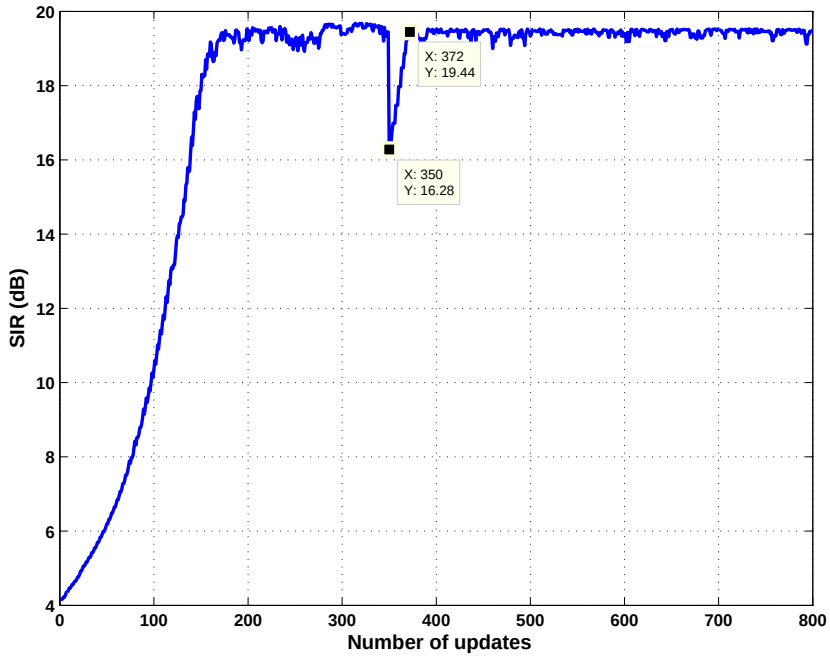
is a good figure of merit for the maximum channel variation magnitude. For the simulation in which channel changes with every window, maxChange is equal to 0.0634. For the other case, maxChange is 0.0612. The simulation results show the nice tracking capability of the algorithm in both cases. Oscillations are more in the more frequent channel variation case as expected. However it still manages to converge and track back the variations. Other one shows a better performance naturally as we have a more slowly varying channel this time.

5.2.2 Simulation Results for Convolutive MB-BSS

For the simulation parameters of the convolutive MB-BSS simulation, data rate is chosen as 60 Gsym/s and window length is 1000. The remaining are the same as the impulsive instantaneous simulation setup. Also the channel is kept constant till the 350^{th} window, and then a sudden change is applied. We see in Figure 5.3.b that 40 windows are enough for the algorithm. Therefore $40 \times 1000 = 40.000$ symbols (0.67 μs) are necessary to track back the sudden channel variation. These results show that convolutive MB-BSS algorithm tracks the impulsive channel variations very fast. If we remember that PMD changes in millisecond time scales, the tracking performance of these MB-BSS algorithms are highly satisfying. They can easily track the changes induced by PMD in very short time durations.

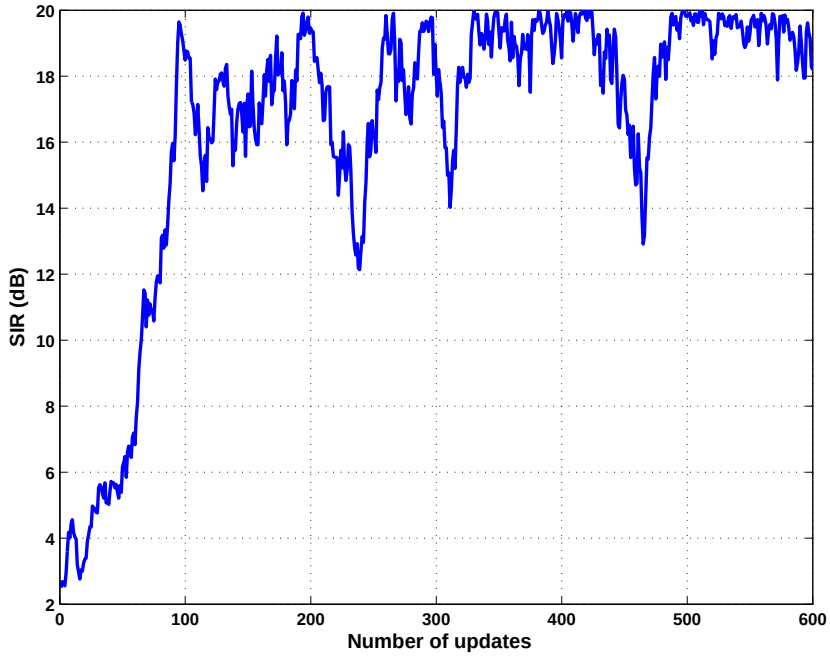


(a)

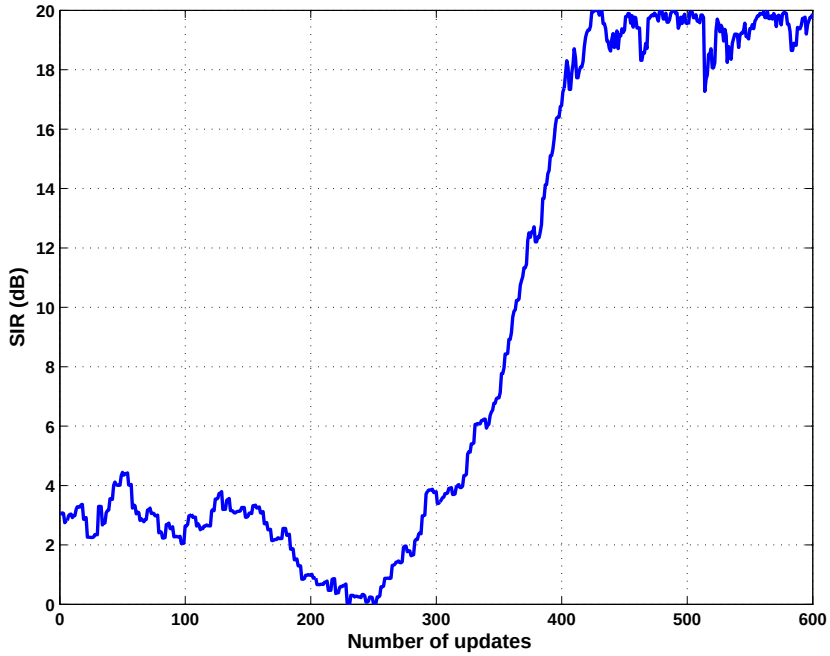


(b)

Figure 5.1: Tracking performance of instantaneous MB-BSS algorithm for impulsive channel change (a) Initial and distorted channel spectrum of the 2^{nd} row 2^{nd} column element (b) Signal to Interference Ratio convergence curve.

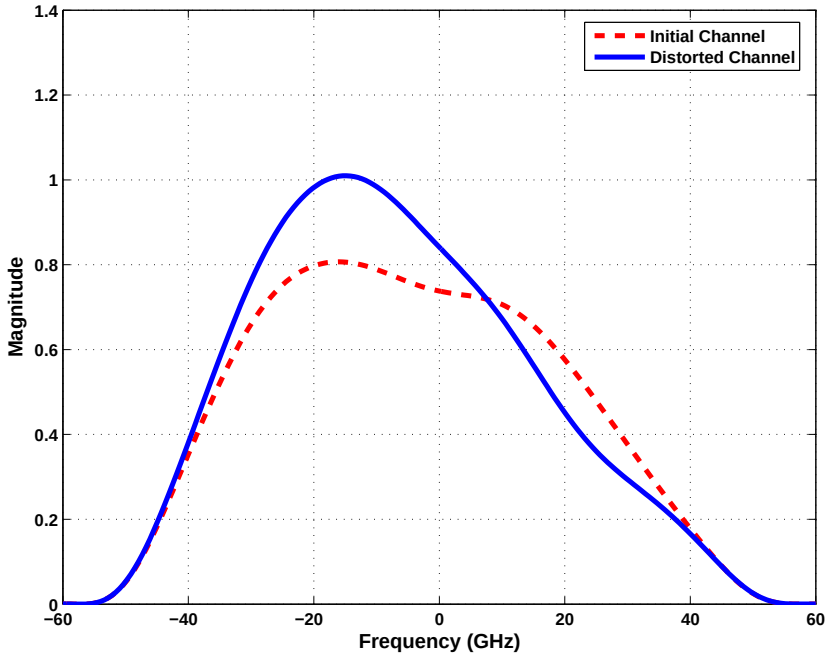


(a)

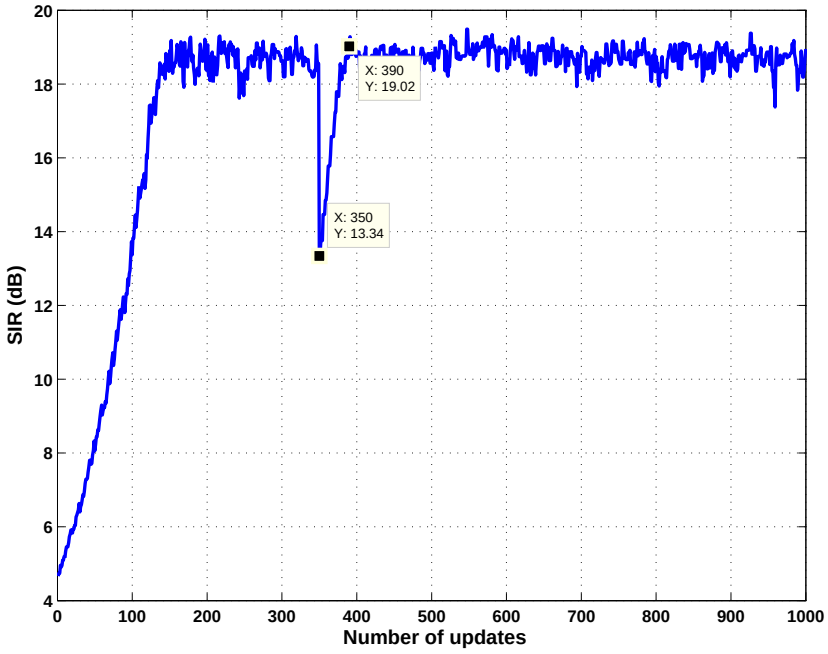


(b)

Figure 5.2: Tracking capability of instantaneous MB-BSS for smooth channel variations (a) SIR convergence curve for the case that channel changes at every new window, (b) SIR convergence curve for the case that channel changes at every third window.



(a)



(b)

Figure 5.3: Tracking performance of convolutive MB-BSS algorithm for impulsive channel change (a) Initial and distorted channel spectrum of the 2^{nd} row 2^{nd} column element (b) Signal to Interference Ratio convergence curve.

Chapter 6

EXTENSIONS AND FUTURE WORK

In the previous chapters, we introduced MB-BSS based framework for the adaptive receiver design for the systems employing PDM scheme. As the current approaches use simple constellations such as 2-PAM (or BPSK) or 4-QAM signalling schemes due to their ease of generation, all of our simulation examples involved the use of 4-QAM constellation. However, the MB-BSS framework applies perfectly to higher-level QAM signalling schemes such as 16-QAM, 64-QAM etc. in case these signalling schemes are chosen to be employed at the transmitters. This is due to the fact that MB-BSS scheme only relies on just independence and the magnitude-boundedness assumption of the sources [26].

In order to illustrate the performance of the proposed scheme for a multilevel signalling scheme, we carried out a simulation that both polarization modulators use 16-QAM constellation. Data rate is 5 Gsym/s, and SNR is 20 dB. ADCs are not used as we have accomplished the all-analog design. The modification needed for the simplified parametric MB-BSS to handle the multilevel signalling scheme is to alter the threshold value K. For 16-QAM constellation, the maximum absolute real entry is 3. Therefore for this type of input, K must be 3. With this small modification, the simulation results are shown in Figure 6.1. We see that approximately 500 updates are necessary for the algorithm to converge. These updates take place within 720 data duration (Note that there may not be an update at each step due to the threshold comparison.). This data duration corresponds to 0.144 μs at 5 Gsym/s. This result shows the speed of the algorithm.

For the future work, most of the research will be focused on the all-analog implementation of the convolutive MB-BSS algorithm. We have introduced the parametric version of the convolutive algorithm. However to overcome convergence problems of the parametric convolutive MB-BSS algorithm for high data rates, some effort must be spent for it (such as increasing filter length as explained before). After these problems are solved, all-analog implementation can be carried out as we have done for the instantaneous case. Another

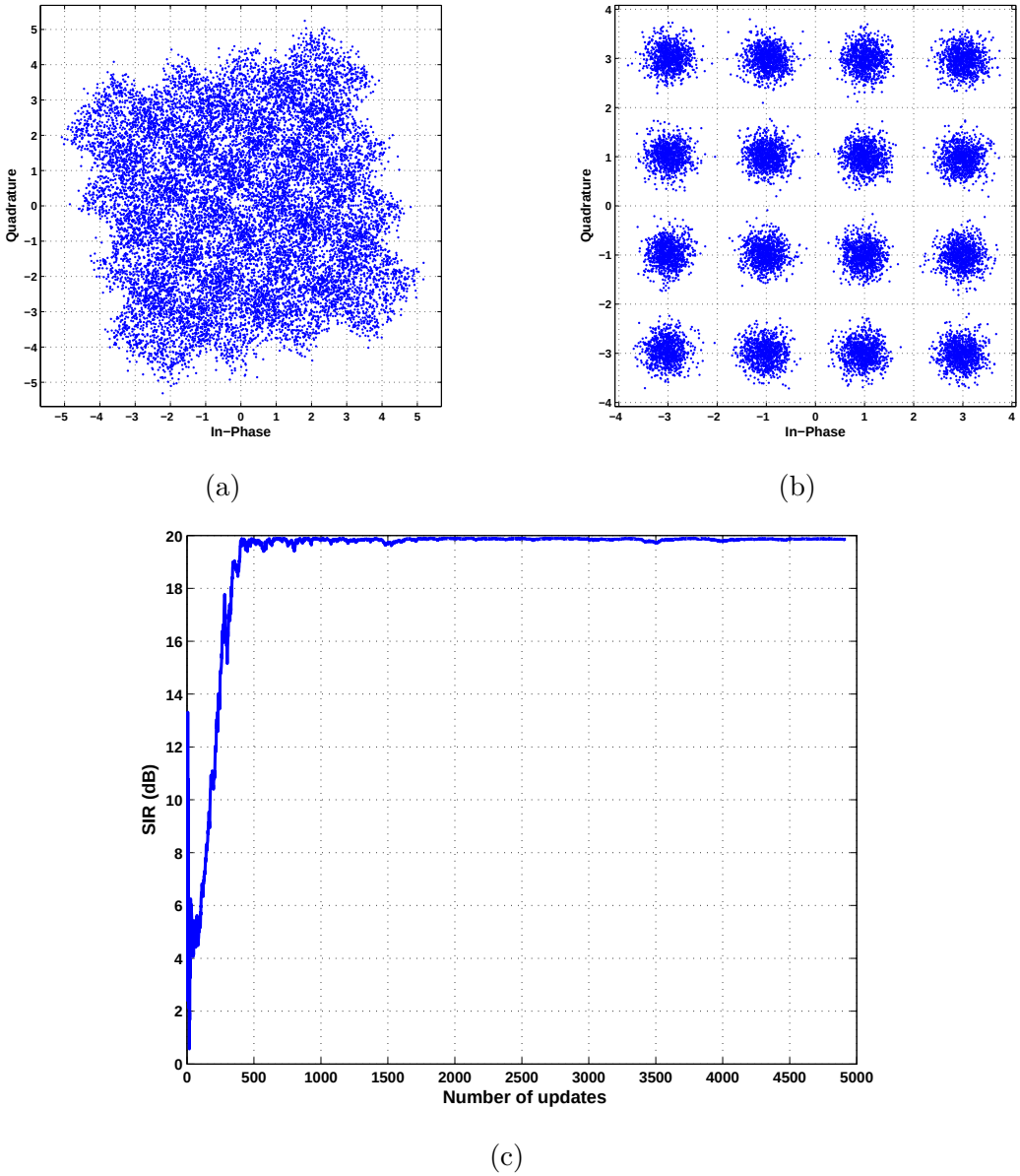


Figure 6.1: Simulation result of simplified parametric MB-BSS for instantaneous case with multilevel signalling scheme (a) The corrupted channel output, (b) The separator output (recovered sources) , (c) Signal to Interference Ratio convergence curve .

forcing factor for analog design of the convolutive MB-BSS is the limitation due to ADCs. Current technologies permit ADCs with 6-bit precision at speeds around 6 Gsym/s. On the other hand, at high data rates such as 60 Gsym/s, ADCs with 6-bit precisions are not available even with interleaved ADC systems and therefore all-analog implementations would be

more critical.

Another important future work topic is the more accurate modeling and simulation of the fiber channel variations. This would be important for both the development and the performance evaluation of algorithms designed for tracking the variations in the channel. For this purpose, we foresee the development of randomly varying paraunitary channel models based on different approaches. One possible approach is the stochastic process modeling for the parameters corresponding to the paraunitary channels. Another possible approach is based on the autoregressive moving average (ARMA) process modeling over the paraunitary manifold.

Chapter 7

CONCLUSION

In this thesis, we proposed a framework for the development of adaptive receivers for fiber optic systems employing PDM scheme. The use of PDM scheme promises potential doubling of the data transmission rate if the receiver successfully counteracts the space and time mixing caused by the PMD phenomenon. The MB-BSS based framework targets to achieve this goal, where the proposed algorithms could lead to increase in data rates not only due to the fact that polarization multiplexing can be done but also due to the fact that the symbol rate can be increased as the algorithms provide a means to counteract the corresponding increase in ISI.

Throughout the thesis, we considered various extensions of the MB-BSS algorithm applied to different scenarios. In the low-rate scheme where the ISI factor is negligible, instantaneous separation algorithm based on MB-BSS, is shown to be both successful and robust for the separation of the original polarization signals from their noisy mixtures. The robustness against noise can be further improved through the employment of the moving window scheme. Moreover, the algorithm can be simplified by using parametrization based scheme which eliminates algorithmic consideration of unitary constraints. Further simplification is achieved with the algorithm variation which replaces the maximum magnitude search over a given window by sample-by-sample threshold comparison. These simplifications lead to all-analog implementations eliminating need for ADC-DSP based processing.

In the high rate scheme, adaptive paraunitary filtering based on MB-BSS criterion is shown to be a viable scheme. Major obstacle in devising low complexity structures is the paraunitary constraint on the adaptive filter. For that purpose, we put forward some iterative schemes that relies on some iterative paraunitary mapping methods based on gradient over block convolution matrices and Stiefel manifold. Although these algorithms eliminate the problems caused by the paraunitary constraint, their implementation is possible only through use of DSP processing. We also offer a parametric adaptive paraunitary version of

the MB-BSS approach which can be potentially implemented with an all-analog structure eliminating the need for a DSP based adaptation.

Since adapting to the channel variations is a significant issue in terms of robustness, we tested the tracking performance of the proposed approach. Simulation results showed that MB-BSS algorithms can successfully track the channel variations without any performance loss and can even provide a means for fast recovery in case of abrupt channel changes. Therefore after the initial eye-opening period, the MB-BSS algorithm can continue to be used for the tracking stage, as an alternative to a decision directed scheme which would be non-robust against abrupt channel changes.

As a final word, we can conclude that our MB-BSS framework seems to be very promising for the development of adaptive structures for the future high data rate fiber optical communication schemes employing both PDM schemes and high symbol rates. Also we have to say that, the proposed approaches can compensate any type of linear distortion effects (such as CD). However, in this thesis we only concentrated on PMD.

Appendix A

DERIVATION OF UPDATES FOR THE PARAMETRIC INSTANTANEOUS MB-BSS ALGORITHM

In this part of the appendix, we complete the parameter updates given in Chapter 3 for the parametric instantaneous MB-BSS algorithm. We have provided the derivation for only one variable (α). Now we provide the results for the other 3 real variables. To begin with, the subgradient $\underline{\mathbf{g}}$ is the same for all 4 variables which is given by (3.23). The only difference comes from the partial derivatives with respect to the 4 real variables.

- For ϕ, the derivative is given as:

$$\frac{\partial f(\mathbf{r})}{\partial \phi} = \underline{\mathbf{g}}^T . \frac{\partial \mathbf{r}}{\partial \phi}. \tag{A.1}$$

where

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial \phi} = [\; & -\cos(\alpha)\sin(\phi+x) \quad \sin(\alpha)\sin(\phi-y) \quad -\sin(\alpha)\sin(\phi+y) \\ & -\cos(\alpha)\sin(\phi-x) \quad \cos(\alpha)\cos(\phi+x) \quad -\sin(\alpha)\cos(\phi-y) \\ & \sin(\alpha)\cos(\phi+y) \quad \cos(\alpha)\cos(\phi-x) \;]^T. \end{aligned} \tag{A.2}$$

- For x, the derivative is given as:

$$\frac{\partial f(\mathbf{r})}{\partial x} = \underline{\mathbf{g}}^T . \frac{\partial \mathbf{r}}{\partial x}. \tag{A.3}$$

where

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial x} = [\; & -\cos(\alpha)\sin(\phi+x) \quad 0 \quad 0 \quad \cos(\alpha)\sin(\phi-x) \\ & \cos(\alpha)\cos(\phi+x) \quad 0 \quad 0 \quad -\cos(\alpha)\cos(\phi-x) \;]^T. \end{aligned} \tag{A.4}$$

- For y, the derivative is given as:

$$\frac{\partial f(\mathbf{r})}{\partial y} = \underline{\mathbf{g}}^T \cdot \frac{\partial \mathbf{r}}{\partial y}. \tag{A.5}$$

where

$$\frac{\partial \mathbf{r}}{\partial y} = \begin{bmatrix} 0 & -\sin(\alpha)\sin(\phi - y) & -\sin(\alpha)\sin(\phi + y) & 0 \\ 0 & \sin(\alpha)\cos(\phi - y) & \sin(\alpha)\cos(\phi + y) & 0 \end{bmatrix}^T. \tag{A.6}$$

With these derivatives, the variables are updated with a suitable step size.

Appendix B

DERIVATION OF UPDATES FOR THE PARAMETRIC CONVOLUTIVE MB-BSS ALGORITHM

Here, we derive the update equations for the parametric convolutive MB-BSS algorithm. We have parameterized the paraunitary separator in Chapter 4 according to Vaidyanathan's theorem in (4.29) as:

$$\mathbf{\Theta}(\boldsymbol{z}) = \mathbf{Q}(\ \mathbf{I}\ -\ \mathbf{v}_1\mathbf{v}_1^*\ +\ z^{-1}\mathbf{v}_1\mathbf{v}_1^*) = \mathbf{\Theta_0} + \mathbf{\Theta_1}z^{-1}, \tag{B.1}$$

where

$$\mathbf{\Theta_0} = \mathbf{Q}(\ \mathbf{I}\ -\ \mathbf{v}_1\mathbf{v}_1^*) \tag{B.2}$$

and

$$\mathbf{\Theta_1} = \mathbf{Q}\mathbf{v}_1\mathbf{v}_1^*. \tag{B.3}$$

For this time for the general 2×2 unitary matrix $\mathbf{Q}$, we have used (3.17). Therefore

$$\mathbf{Q} = \begin{bmatrix} \cos(\alpha)e^{jx} & \sin(\alpha)e^{jy} \\ -\sin(\alpha)e^{jz} & \cos(\alpha)e^{j(y+z-x)} \end{bmatrix}, \tag{B.4}$$

and

$$\mathbf{v}_1\ =\ \begin{bmatrix} \cos(\gamma)e^{j\phi_1} \\ \sin(\gamma)e^{j\phi_2} \end{bmatrix}, \tag{B.5}$$

If we define $\mathbf{Y}$ as the input data matrix in the window of interest, and $\mathbf{Z}$ as the corresponding separator outputs as before, then the cost function can be written as:

$$\begin{aligned} f(\boldsymbol{M}) &= ||\mathcal{R}e\{vec(\mathbf{Z})\}||_\infty \\ &= ||\mathcal{R}e\{\underbrace{(\mathbf{Y}^T \otimes \mathbf{I})}_{A} \underbrace{vec(\mathbf{M})}_{\mathbf{b}}\}||_\infty \\ &= ||\mathcal{R}e\{\mathbf{Ab}\}||_\infty, \end{aligned} \tag{B.6}$$

where

$$\mathbf{M} = \begin{bmatrix} \boldsymbol{\Theta_0} & \boldsymbol{\Theta_1} \end{bmatrix}. \tag{B.7}$$

Further simplification of the algorithm leads to:

$$\begin{aligned} ||\mathcal{R}e\{\mathbf{Ab}\}||_\infty &= ||\mathcal{R}e\{(\mathbf{A}_R + j\mathbf{A}_I)(\mathbf{b}_R + j\mathbf{b}_I)\}||_\infty \\ &= ||\mathbf{A}_R\mathbf{b}_R - \mathbf{A}_I\mathbf{b}_I||_\infty \\ &= ||\underbrace{\begin{bmatrix} \mathbf{A}_R & -\mathbf{A}_I \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} \mathbf{b}_R \\ \mathbf{b}_I \end{bmatrix}}_{\mathbf{r}}||_\infty \\ &= ||\underbrace{\mathbf{Kr}}_{\mathbf{s}}||_\infty = f(\mathbf{r}), \end{aligned} \tag{B.8}$$

where subscript "R" denotes the real part of the component while "I" denotes the imaginary part. A subgradient with respect to $\mathbf{r}$ is given as:

$$\underline{\mathbf{g}} = sign(s_i)\mathbf{K}^T_{i,:} \quad where \quad |s_i| = ||\mathbf{s}||_\infty. \tag{B.9}$$

Now chain rule is used to obtain the derivatives with respect to the 7 real variables. This time, we will begin with the matrix $\mathbf{M}$. We will obtain its derivatives with respect to the 7 real variables, and from this point the derivatives of $\mathbf{r}$ will be obtained. The partial derivatives of $\mathbf{M}$ with respect to the 7 real variables are:

$$\frac{\partial \mathbf{M}}{\partial \alpha} = \left[\frac{\partial \mathbf{Q}}{\partial \alpha}(\mathbf{I} - \mathbf{v}_1\mathbf{v}_1^*) \qquad \frac{\partial \mathbf{Q}}{\partial \alpha}\mathbf{v}_1\mathbf{v}_1^*\right], \tag{B.10}$$

$$\frac{\partial \mathbf{M}}{\partial x} = \left[\frac{\partial \mathbf{Q}}{\partial x}(\mathbf{I} - \mathbf{v}_1\mathbf{v}_1^*) \qquad \frac{\partial \mathbf{Q}}{\partial x}\mathbf{v}_1\mathbf{v}_1^*\right], \tag{B.11}$$

$$\frac{\partial \mathbf{M}}{\partial y} = \left[\frac{\partial \mathbf{Q}}{\partial y}(\mathbf{I} - \mathbf{v}_1\mathbf{v}_1^*) \qquad \frac{\partial \mathbf{Q}}{\partial y}\mathbf{v}_1\mathbf{v}_1^*\right], \tag{B.12}$$

$$\frac{\partial \mathbf{M}}{\partial z} = \left[\frac{\partial \mathbf{Q}}{\partial z}(\mathbf{I} - \mathbf{v}_1\mathbf{v}_1^*) \qquad \frac{\partial \mathbf{Q}}{\partial z}\mathbf{v}_1\mathbf{v}_1^* \right], \tag{B.13}$$

$$\frac{\partial \mathbf{M}}{\partial \gamma} = \left[\mathbf{Q}(\mathbf{I} - \frac{\partial \mathbf{v}_1\mathbf{v}_1^*}{\partial \gamma}) \qquad \mathbf{Q}\,\frac{\partial \mathbf{v}_1\mathbf{v}_1^*}{\partial \gamma} \right], \tag{B.14}$$

$$\frac{\partial \mathbf{M}}{\partial \phi_1} = \left[\mathbf{Q}(\mathbf{I} - \frac{\partial \mathbf{v}_1\mathbf{v}_1^*}{\partial \phi_1}) \qquad \mathbf{Q}\,\frac{\partial \mathbf{v}_1\mathbf{v}_1^*}{\partial \phi_1} \right], \tag{B.15}$$

$$\frac{\partial \mathbf{M}}{\partial \phi_2} = \left[\mathbf{Q}(\mathbf{I} - \frac{\partial \mathbf{v}_1\mathbf{v}_1^*}{\partial \phi_2}) \qquad \mathbf{Q}\,\frac{\partial \mathbf{v}_1\mathbf{v}_1^*}{\partial \phi_2} \right], \tag{B.16}$$

where

$$\frac{\partial \mathbf{Q}}{\partial \alpha} = \begin{bmatrix} -\sin(\alpha)e^{jx} & \cos(\alpha)e^{jy} \\ -\cos(\alpha)e^{jz} & -\sin(\alpha)e^{j(y+z-x)} \end{bmatrix}, \tag{B.17}$$

$$\frac{\partial \mathbf{Q}}{\partial x} = \begin{bmatrix} j\cos(\alpha)e^{jx} & 0 \\ 0 & -j\cos(\alpha)e^{j(y+z-x)} \end{bmatrix}, \tag{B.18}$$

$$\frac{\partial \mathbf{Q}}{\partial y} = \begin{bmatrix} 0 & j\sin(\alpha)e^{jy} \\ 0 & j\cos(\alpha)e^{j(y+z-x)} \end{bmatrix}, \tag{B.19}$$

$$\frac{\partial \mathbf{Q}}{\partial z} = \begin{bmatrix} 0 & 0 \\ -j\sin(\alpha)e^{jz} & j\cos(\alpha)e^{j(y+z-x)} \end{bmatrix}, \tag{B.20}$$

$$\frac{\partial \mathbf{v}_1\mathbf{v}_1^*}{\partial \gamma} = \begin{bmatrix} -\sin(2\alpha) & \cos(2\alpha)e^{j(\phi_1-\phi_2)} \\ \cos(2\alpha)e^{j(\phi_2-\phi_1)} & \sin(2\alpha) \end{bmatrix}, \tag{B.21}$$

$$\frac{\partial \mathbf{v}_1\mathbf{v}_1^*}{\partial \phi_1} = \begin{bmatrix} 0 & \frac{1}{2}j\sin(2\alpha)e^{j(\phi_1-\phi_2)} \\ -\frac{1}{2}j\sin(2\alpha)e^{j(\phi_2-\phi_1)} & 0 \end{bmatrix}, \tag{B.22}$$

$$\frac{\partial \mathbf{v}_1\mathbf{v}_1^*}{\partial \phi_2} = \begin{bmatrix} 0 & -\frac{1}{2}j\sin(2\alpha)e^{j(\phi_1-\phi_2)} \\ \frac{1}{2}j\sin(2\alpha)e^{j(\phi_2-\phi_1)} & 0 \end{bmatrix}. \tag{B.23}$$

From these equations, we will be able to obtain the derivatives of $\mathbf{r}$. For example, we will obtain $\frac{\partial \mathbf{r}}{\partial \alpha}$ from $\frac{\partial \mathbf{M}}{\partial \alpha}$, the same way we obtain $\mathbf{r}$ from $\mathbf{M}$. Because all the operations taking place during this transformation are linear operations. Therefore we now have the

partial derivatives of $\mathbf{r}$ with respect to the 7 real variables. From here, we obtain the overall derivatives:

$$\frac{\partial f(\mathbf{r})}{\partial \alpha} = \underline{\mathbf{g}}^T \cdot \frac{\partial \mathbf{r}}{\partial \alpha}, \tag{B.24}$$

$$\frac{\partial f(\mathbf{r})}{\partial x} = \underline{\mathbf{g}}^T \cdot \frac{\partial \mathbf{r}}{\partial x}, \tag{B.25}$$

$$\frac{\partial f(\mathbf{r})}{\partial y} = \underline{\mathbf{g}}^T \cdot \frac{\partial \mathbf{r}}{\partial y}, \tag{B.26}$$

$$\frac{\partial f(\mathbf{r})}{\partial z} = \underline{\mathbf{g}}^T \cdot \frac{\partial \mathbf{r}}{\partial z}, \tag{B.27}$$

$$\frac{\partial f(\mathbf{r})}{\partial \gamma} = \underline{\mathbf{g}}^T \cdot \frac{\partial \mathbf{r}}{\partial \gamma}, \tag{B.28}$$

$$\frac{\partial f(\mathbf{r})}{\partial \phi_1} = \underline{\mathbf{g}}^T \cdot \frac{\partial \mathbf{r}}{\partial \phi_1}, \tag{B.29}$$

$$\frac{\partial f(\mathbf{r})}{\partial \phi_2} = \underline{\mathbf{g}}^T \cdot \frac{\partial \mathbf{r}}{\partial \phi_2}, \tag{B.30}$$

where the subgradient $\underline{\mathbf{g}}$ comes from (B.9). With these derivatives, the variables are updated with a suitable step size.

BIBLIOGRAPHY

[1] J. Crisp, *Introduction to Fiber Optics*. Newnes, 2001.

[2] G. P. Agrawal, *Fiber-Optic Communication Systems*. Wiley-Interscience, 2002.

[3] H. J. R. Dutton, *Understanding Optical Communications*. IBM, 1998.

[4] C. DeCusatis, *Fiber Optic Data Communication: Technological Trends and Advances*. Academic Press, 2002.

[5] A. R. Shah, R. C. J. Hsu, A. Tarighat, A. H. Sayed, and B. Jalali, "Coherent optical MIMO (COMIMO)," *Journal of Lightwave Technology*, vol. 23, August 2005.

[6] G. P. Agrawal, *Nonlinear Fiber Optics*. Academic Press, 2001.

[7] C. D. Poole and R. E. Wagner, "Phenomenological approach to polarization dispersion in long single mode fibers," *Electronics Letter*, vol. 22, pp. 1029–1030, 1986.

[8] A. O. Lima, I. T. L. Jr., T. Adalı, and C. R. Menyuk, "A novel polarization diversity receiver for pmd mitigation," *IEEE Photonics Technology Letters*, vol. 14, pp. 465–467, April 2002.

[9] H. Bulow, W. Baumert, H. Schmuck, F. Mohr, T. Schulz, F. Kuppers, and W. Weiershausen, "Measurement of the maximum speed of pmd fluctuation in installed field fiber," *Conf. Opt. Fiber Commun. Int. Conf. Integrated Opt. (OFC/IOOC)*, vol. 2, pp. 83–85, 1999.

[10] C. D. Poole, "Statistical treatment of polarization dispersion in single-mode fiber," *Optical Letters*, vol. 13, pp. 687–689, 1988.

[11] F. Curti, B. Daino, Q. Mao, and F. Matera, "Concatenation of polarization dispersion in single-mode fibers," *Electronics Letters*, vol. 25, pp. 290–292, February 1989.

[12] N. Gisin, R. Passy, J. C. Bishoff, and B. Perny, "Experimental investigations of the statistical properties of polarization mode dispersion in single mode fibers," *IEEE Photonics Technology Letters*, vol. 5, pp. 819–821, 1993.

[13] P. Hill, R. Olshansky, and W. Burns, "Optical polarization division multiplexing at 4 gb/s," *IEEE Photonics Technology Letters*, vol. 4, pp. 500–502, May 1992.

[14] H. Sunnerud, M. Karlsson, C. Xie, and P. A. Andrekson, "Polarization-mode dispersion in high-speed fiber-optic transmission systems," *Journal of Lightwave Technology*, vol. 20, pp. 2204–2219, December 2002.

[15] F. Buchali and H. Blow, "Adaptive pmd compensation by electrical and optical techniques," *Journal of Lightwave Technology*, vol. 22, pp. 1116–1126, April 2004.

[16] H. Sunnerud, C. Xie, M. Karlsson, R. Samuelsson, and P. A. Andrekson, "A comparison between different pmd compensation techniques," *Journal of Lightwave Technology*, vol. 20, pp. 368–378, March 2002.

[17] M. Secondini, E. Forestieri, and G. Prati, "Plc optical equalizer for chromatic and polarization-mode dispersion compensation based on mse control," *IEEE Photonics Technology Letters*, vol. 16, pp. 1173–1175, April 2004.

[18] C. K. Madsen, M. Cappuzzo, E. J. Laskowski, E. Chen, L. Gomez, A. Griffin, A. Wong-Foy, S. Chandrasekhar, L. Stulz, and L. Buhl, "Versatile integrated pmd emulation and compensation elements," *Journal of Lightwave Technology*, vol. 22, pp. 1041–1050, April 2004.

[19] T. Merker, N. Hahnenkamp, and P. Meissner, "Comparison of pmd-compensation techniques at 10 gbit/s using an optical first-order compensator and electrical tranversal filter," *Optical Commuincations*, pp. 135–141, August 2000.

[20] N.Walker and G.Walker, "Polarization control for coherent communications," *Journal of Lightwave Technology*, vol. 8, pp. 438–458, 1990.

[21] D. E. Crivelli, H. S. Carrer, and M. R. Hueda, "Adaptive digital equalization in the presence of chromatic dispersion, pmd, and phase noise in coherent fiber optic systems," *IEEE Communications Society Globecom*, pp. 2545–2551, 2004.

[22] J. H. Winters and R. D. Gitlin, "Electrical signal processing techniques in long-haul fiber-optic systems," *IEEE Transactions on Communications*, vol. 38, pp. 1439–1453, September 1990.

[23] O. E. Agazzi, M. R. Hueda, H. S. Carrer, and D. E. Crivelli, "Maximum-likelihood sequence estimation in dispersive optical channels," *Journal of Lightwave Technology*, vol. 23, pp. 749–763, February 2005.

[24] G. Katz, D. Sadot, and J. Tabrikian, "Electrical dispersion compensation equalizers in optical long-haul coherent-detection system," *ICTON 2005*, pp. 80–83, 2005.

[25] A. T. Erdogan, "A blind equalization approach for all equalizable MIMO convolutive channels," *submitted to IEEE Trans. on Signal Processing*, September 2005.

[26] A. T. Erdogan, "A simple geometric blind source separation method for bounded magnitude sources," *IEEE Trans. on Signal Processing*, vol. 54, pp. 438– 449, February 2006.

[27] W. van Etten and J. van der Plaats, *Fundamentals of Optical Fiber Communications*. Prentice Hall, 1991.

[28] W. Weiershausen, R. Leppla, F. Kuppers, and H. Scholl, "Polarization-mode dispersion in fibre transmission: theoretical approach, impact on systems, and suppression of signal-degradation effects," *Proceedings of ECOC99. 25th European Conference on Optical Communication, Nice, France*, vol. 2, pp. 130–132, September 1999.

[29] D. Marcuse, C. R. Menyuk, and P. K. A. Wai, "Application of the manakov-pmd equation to studies of signal propagation in optical fibers with randomly varying birefringence," *Journal of Lightwave Technology*, vol. 15, pp. 1735–1746, September 1997.

[30] G. D. VanWiggeren and R. Roy, "Transmission of linearly polarized light through a single-mode fiber with random fluctuations of birefringence," *Applied Optics*, vol. 38, pp. 3888–3892, June 1999.

[31] J. P. Gordon and H. Kogelnik, "Pmd fundamentals: Polarization mode dispersion in optical fibers," *PNAS*, vol. 97, pp. 4541–4550, April 2000.

[32] L. E. Nelson, T. N. Nielsen, and H. Kogelnik, "Observation of pmd-induced coherent crosstalk in polarization-multiplexed transmission," *IEEE Photonics Technology Letters*, vol. 13, pp. 738–740, July 2001.

[33] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley and Sons Inc., 2001.

[34] S. Amari, A. Cichocki, and H. H. Young, "A new learning algorithm for blind source separation," *In Advances in Neural Information Processing Systems, MIT Press Cambridge*, vol. 8, pp. 757–763, 1997.

[35] D.-T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. on Signal Processing*, vol. 45, pp. 1712 – 1725, July 1997.

[36] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach," *Signal Processing*, vol. 45, pp. 59–83, July 1995.

[37] C. Papadias, "Globally convergent blind source separation based on a multiuser kurtosis maximization criterion," *IEEE Trans. on Signal Processing*, vol. 48, pp. 3508–3519, December 2000.

[38] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[39] P. Common, "Independent component analysis- a new concept?," *Signal Processing*, vol. 36, pp. 287–314, April 1994.

[40] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483–1492, October 1997.

[41] D.-T. Pham, "Mutual information approach to blind separation of stationary sources," *IEEE Trans. on Signal Processing*, vol. 48, pp. 1935 – 1946, July 2002.

[42] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. on Signal Processing*, vol. 50, pp. 635 – 650, March 2002.

[43] A. Nedic and D. Bertsekas, "Convergence rate of incremental subgradient algorithms," *Stochastic Optimization: Algorithms and Applications, S.Ursayev and P.M. Pardalos, Editors*, 2000.

[44] J. Goffin, "On convergence rates of subgradient optimization methods," *Math. Programming*, vol. 13, pp. 329–347, May 1977.

[45] J. Goffin and K. Kiwiel, "Convergence of a simple subgradient level method," *Math. Programming*, vol. 85, no. 1, pp. 207–211, 1999.

[46] N. Z. Shor, "On the rate of convergence of the generalized gradient method," *Kibernatika*, vol. 4.

[47] H. Sherali, G. Choi, and C. Tuncbilek, "Variable target value method for nondifferentiable optimization," *Operations Research Letters*, vol. 26, pp. 1–8, February 2000.

[48] J. Lee, P. Roux, U.-V. Koc, T. Link, Y. Baeyens, and Y.-K. Chen, "A 5-b 10-GSample/s A/D converter for 10-Gb/s optical receivers," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1671–1679, October 2004.

[49] Y. Inouye and R.-W. Liu, "A system-theoretic foundation for blind equalization of an FIR MIMO channel system," *IEEE Trans. on CAS-I: Fund. Th. and Apps.*, vol. 49, pp. 425–436, April 2002.

[50] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley and Sons Inc., 2001.

[51] P. P. Vaidyanathan and Z. Doganata, "The role of lossless systems in modern digital signal processing: A tutorial," *IEEE Trans. on Educ.*, pp. 181–197, Aug. 1989.

[52] A. Graham, *Kronecker Products and Matrix Calculus with Applications*. Ellis Horwood Limited, 1981.

VITA

TURGUT MUSTAFA ÖKTEM was born in Ankara, Turkey on June 10, 1981. He received his B.Sc. degree in Electrical and Electronics Engineering from Bilkent University, Ankara, Turkey, in 2004. In September 2004, he joined the Electrical and Computer Engineering Department of Koç University, Istanbul, Turkey as a teaching and research assistant. His research interests are digital signal processing, convex optimization and fiber optic communications.