Doctoral School
of Computer Science,
Telecommunications
and Electronics of Paris

PhD Thesis

presented in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy from Télécom ParisTech

Specialization: **Communications and Electronics**

Mehrez Selmi

Advanced Digital Signal Processing tools for QAM-based optical fiber communications

Defense scheduled by the end of September 2011 with the (expected) committee

C. Laot	Reviewer
S. Savory	Reviewer
R. Freund	Examiner
J.-C. Simon	Examiner
S. Bigo	Examiner
P. Ciblat	Supervisor
Y. Jaouën	Supervisor
E. Pincemin	Invited member

Table of Contents

Acronyms List **v**

General Introduction **1**

1 Problem Statement **5**
1.1 Introduction . . . 5
1.2 State-of-the-art of optical transmission systems . . . 6
1.2.1 Increasing the transmission capacity of WDM systems . . . 7
1.2.2 Coherent detection . . . 9
1.2.2.1 Phase diversity coherent receiver . . . 10
1.2.2.2 Phase and polarization diversity receiver . . . 11
1.3 Signal model . . . 12
1.3.1 Single versus Multiple carrier transmission . . . 12
1.3.2 Propagation impairments . . . 13
1.3.2.1 Linear Impairments . . . 13
1.3.2.2 Nonlinear impairments . . . 17
1.3.2.3 Other (neglected) impairments . . . 18
1.3.3 Mathematical signal model . . . 19
1.3.4 Optical Signal to Noise Ratio (OSNR) . . . 20
1.4 State of the art of Digital Signal Processing . . . 21
1.4.1 Chromatic dispersion compensation . . . 23
1.4.2 PMD and residual CD compensation . . . 25
1.4.2.1 Constant Modulus Algorithm . . . 26
1.4.2.2 Radius Directed Equalizer . . . 27
1.4.2.3 Decision Directed Least Square . . . 28
1.4.3 Carrier Frequency Estimation . . . 28
1.4.4 Constant Phase Estimation . . . 29
1.5 Simulation Set-up . . . 30
1.6 Outline of the thesis . . . 31

2 Newton based adaptive equalization **33**
2.1 Introduction . . . 33
2.2 Adaptive equalizers . . . 33
2.2.1 Adaptive Data-Aided (A-DA) . . . 33
2.2.2 Adaptive Decision-Directed (A-DD) . . . 34
2.2.3 Adaptive Constant Modulus Algorithm (A-CMA) . . . 34
2.2.4 Adaptive Radius-Directed Equalization (A-RDE) . . . 34

2.3 Pseudo-Newton based adaptive equalizers . . . 35
2.4 Convergence analysis . . . 37
2.4.1 Convergence of the A-CMA . . . 37
2.4.2 Convergence of the AN-CMA . . . 38
2.4.3 Convergence of the A-RDE . . . 39
2.4.4 Convergence of the AN-RDE . . . 39
2.4.5 Impact of the filter length . . . 41
2.4.6 Impact of chromatic dispersion . . . 43
2.4.7 Impact of time-varying channels . . . 45
2.5 Computational Load . . . 46
2.6 Conclusion . . . 48

3 Block-wise Carrier Frequency Offset Estimation 49
3.1 Introduction . . . 49
3.2 Signal model . . . 49
3.3 State-of-the-Art . . . 50
3.3.1 Differential phase based method . . . 50
3.3.2 FFT maximization based method . . . 53
3.4 Proposed Carrier Frequency Offset estimate . . . 53
3.5 Estimation performance . . . 56
3.5.1 Tolerance to CD . . . 58
3.5.2 Tolerance to PMD . . . 59
3.5.3 Tolerance to phase noise . . . 61
3.5.4 Residual CFO compensation . . . 61
3.5.5 BER of the whole system . . . 64
3.6 Conclusion . . . 65

4 Block-wise Blind Equalization 67
4.1 Introduction . . . 67
4.2 Coherence time . . . 68
4.3 Block-wise algorithms . . . 69
4.4 Block-wise equalization performance . . . 75
4.4.1 Static channel case . . . 75
4.4.2 Non-static channel case . . . 80
4.5 Singularity issue . . . 83
4.5.1 State-of-the-Art . . . 84
4.5.1.1 Filter re-initialization based algorithm . . . 84
4.5.1.2 Penalized function based algorithm . . . 85
4.5.2 Proposed deflation based algorithm . . . 86
4.5.3 Simulation results . . . 87
4.6 Experimental results . . . 89
4.6.1 Experimental set-up . . . 90
4.6.2 Experimental Performance . . . 90
4.6.2.1 Single polarization transmission . . . 90
4.6.2.2 POLMUX transmission . . . 94
4.7 Conclusion . . . 98

Conclusion and Perspectives **99**

A Appendix related to Chapter 1 **101**
A.1 CMA derivations 101
A.2 Phase derivations 101

B Appendix related to Chapter 4 **103**
B.1 Derivations of Eq. (4.30) 103

Bibliography **105**

Acronyms List

Acronym	English Language
ADC	Analog to Digital Converter
ASE	Amplified Spontaneous Emission
BER	Bit Error Rate
CD	Chromatic Dispersion
CMA	Constant Modulus Algorithm
CW	Continuous Wave (Laser)
DCF	Dispersion Compensation Fiber
DD	Decision Directed
DFE	Decision Feedback Equalizer
DGD	Differential Group Delay
DSP	Digital Signal Processing
ECL	External Cavity Laser
EDFA	Erbium Doped Fiber Amplifier
FEC	Forward Error Correction
FFE	Feed Forward Equalizer
FIR	Finite Impulse Response
FO	Frequency Offset
FTTH	Fiber To The Home
GVD	Group Velocity Delay
I/Q	In-phase / Quadrature
IIR	Infinite Impulse Response
IM/DD	Intensity Modulation / Direct Detection
ISI	Inter Symbol Interference
LMS	Least Means Square
LO	Local Oscillator
MAN	Metropolitan Area Network
MSE	Minimum Square Error
NLS	Nonlinear Schrodinger
NRZ	Non Return to Zero
OFDM	Orthogonal Frequency Division Multiplexing
OOK	On-Off Keying
OSNR	Optical Signal To Noise Ratio
PBC	Polarization Beam Combiner
PBS	Polarization Beam Splitter
PDL	Polarization Dependent Loss

PMD	Polarization Mode Dispersion
PN	Pseudo Newton
POLMUX	Polarization Multiplexing
PSK	Phase Shift Keying
PSP	Principal State of Polarization
QAM	Quadrature Amplitude Modulation with M constellation points
RDE	Radius Directed Equalizer
RZ	Return to Zero
SMF	Single Mode Fiber
SPM	Self Phase Modulation
WDM	Wavelength Division Multiplexing
XPM	Cross Phase Modulation

General Introduction

Problem statement

The work presented in this PhD thesis has been produced thanks to the collaboration of the teams "Digital Communications" and "Optical Telecommunications" of the department "Communications and Electronics" (COMELEC) at Telecom ParisTech in the framework of a "Futur & Ruptures" grant supported by the Fondation Telecom and the Institut Telecom.

Fiber-optic communication systems have revolutionized the telecommunications industry and have played a major role in the advent of the information age since their introduction in the 1970's. Because of its advantages over electrical transmission, optical fibers have been largely deployed in core networks. During the last years, IPTV, HDTV, VoD, mobile broadband services and internet applications have boomed, causing saturation issues in the networks and leading to an increase in bandwidth demand. This pushed carriers to increase the capacity of WDM channels by introducing 100G and beyond systems.

With the new possibilities offered for the high speed digital circuits, coherent systems have attracted a lot of attention during the last years. Apart from the receiver sensitivity, the interest lies now in the increase of spectral efficiency as well as tolerance against dispersion effects and fiber nonlinearities. Those are today's the most limiting factors in ultra long haul communication systems.

Additionally in contrast to Intensity Modulation Direct Detection (IM-DD) systems or differential Phase Shift Keying (PSK) systems, the received electrical signal in coherent receiver is proportional to the electrical field vector of the optical signal. Therefore the system becomes linear, which means that all linear distortions like Chromatic Dispersion (CD) and Polarization Mode Dispersion (PMD) can theoretically be compensated without any losses and also non linear effects can be compensated very efficiently. Moreover, the access to both the phase and the amplitude of the signal makes the use of advanced modulation formats such as multilevel formats such as Quadrature PSK (QPSK) and Quadrature Amplitude Modulation (QAM), which can raise the spectral efficiency up to several bit/s/Hz, compared to only 1bit/s/Hz for IM-DD systems. However, these advantages have their price. A coherent receiver is much more complex than a simple direct detection receiver. High speed analog to digital converters (ADC) are needed to convert the received signal into the digital domain. And these ADCs must be interfaced with a digital signal processing unit, which performs polarization control,

equalization and finally the carrier and data recovery. To develop these components and algorithms for the next generation of optical communication systems, which can run at 100G and beyond is still a challenging issue.

The aim of this thesis is to develop robust Digital Signal Processing (DSP) tools specific to the optical channel. Those algorithms at least deal with the operations of carrier phase recovery, frequency offset estimation, equalization and the tracking of the variation of PMD. Knowing that those algorithms are to be implemented in circuits running with a speed of some tens of MHz, the optimization of those tools is essential. Moreover, the already proposed tools are well adapted for QPSK formats, and as higher order modulation formats are more sensitive to signal distortions, accurate estimators and robust equalizers are still required for such QAM formats. Our proposed algorithms should be tested using a simulation setup of an optical transmission system using coherent detection and validated with offline processing of real measurements in the framework of the European Network of Excellence (EURO-FOS).

Outline and contributions

In this section, we give the thesis outline and we mention the most important results. Basically, our system of interest is based on coherent PolMux transmission using QAM modulation for reaching more than 100Gbit/s per WDM channel.

In Chapter 1, we initially introduce the optical communications scheme, the signal model, the different existing algorithms for mitigating the linear impairments created by the fiber, and the phase distortions. This chapter highlights the fact that works have to be done for mitigating the linear impairments in the context of QAM modulations.

In Chapter 2, we develop new adaptive blind and decision-directed equalizers that are well adapted to QAM and that offers a better convergence speed with only a little extra computational load. The idea is to use non-constant step-size in the equalizer update equation. The approach deals with the Pseudo-Newton gradient-descent algorithm.

In Chapter 3, we focus on the carrier frequency offset (CFO) estimator. Here, a block-wise approach is involved. We have remarked that the existing CFO estimate really offer poor performance for high-order QAM modulation. Therefore, we have proposed new CFO estimate very adapted to QAM modulation that yields remarkable performance and enables the system to work without penalty.

In Chapter 4, we deal with the blind equalization implementing for mitigating the residual CD, the PMD. According to the fact that the channel is quasi-static in optical fibers (compared to the data rate), we propose to focus rather on block-wise implementation of the equalizers than on the adaptive ones (since there is nothing to adapt quickly!). We thus introduce new equalizers which are more precise and can be calculated much faster. Moreover, the new equalizers are modified in order to prevent the singularity phenomenon which yields the same equalizer output on both polarizations. Finally our block-wise approach is especially well adapted for burst communication.

Publications

The work presented in this manuscript has led to the following publications:

Journal paper

- M. Selmi, C. Gosset, M. Noelle, P. Ciblat, and Y. Jaouën : Blockwise Digital Signal Processing for PolMux QAM/PSK optical coherent systems, accepted for publication to IEEE Journal of Lightwave Technology.

Conference papers

- M. Selmi, P. Ciblat, Y. Jaouën and C. Gosset : A robust deflation based demultiplexing algorithm for QAM coherent optical systems, accepted for publication at European Conference on Optical Communications (ECOC), Geneva (Switzerland), September 2011.
- M. Selmi, P. Ciblat, Y. Jaouën, and C. Gosset : Complexity Analysis of block equalization approach for PolMux QAM coherent systems, OSA Signal Processing Workshop on Photonic Communications (SPPCOM), Toronto (Canada), June 2011.
- M. Selmi, P. Ciblat, Y. Jaouën and C. Gosset : Block versus Adaptive MIMO Equalization for Coherent PolMux QAM Transmission Systems, European Conference on Optical Communications (ECOC), Turin (Italy), September 2010.
- M. Selmi, Y. Jaouën, P. Ciblat and C. Gosset : Pseudo-Newton based equalization algorithms for QAM coherent optical systems, Optical Fiber Communication Conference (OFC), San Diego (USA), March 2010.
- M. Selmi, Y. Jaouën and P. Ciblat : Accurate digital Frequency Offset Estimator for coherent PolMux QAM transmission systems, European Conference on Optical Communications (ECOC), Vienna (Austria), September 2009.

Chapter 1

Problem Statement

1.1 Introduction

Fiber-optic communication systems present many advantages over electrical transmission using copper cable. Those advantages include low attenuation, long reach and high capacity. Modern telecom fibers exhibit attenuation coefficients below 0.2dB/km across a bandwidth of many THz compared to tens of dB/km of losses for a few hundred MHz bandwidth for typical coaxial cable [1], i.e., more than 10Tbit/ s of capacity can be transmitted for distances exceeding 10.000km. As a consequence, fiber-optics have been massively deployed in the core network since their introduction in the mid 1970's and nowadays they are a key enabler for broadband Internet access and services, and the main transporter of the Internet Protocol (IP) traffic.
According to CISCO [2], the global IP traffic grew by 45% during 2009, and will quadruple from 2009 to 2014 to reach 64 exabyte (1EB = 10^{18}bytes) per month. Overall, IP traffic will grow at a rate of 34% annually in the same period. This increase in the bandwidth demand is fueled by the boom of High Definition TV, Video-on-Demand, mobile broadband services and the introduction of new video based internet applications such as streaming, tele-working, network gaming and web conferencing.
To cope with the increase of bandwidth demand and solve the saturation issues in the networks, carriers are pushed to increase the capacity of the already installed optical networks by introducing 40Gbit/s and 100Gbit/s channels in wavelength-division-multiplexing (WDM) systems.
In this chapter, we review the structure of optical communication WDM systems, and explore the different options to increase the overall capacity of those systems (Section 1.2). Then, we review the state of the art of optical systems especially operating at 40 and 100Gbit/s using the advanced coherent detection scheme. In Section 1.3, the main channel impairments and system imperfections disturbing the coherent receiver are drawn as well as the mathematical signal model. The existing digital signal processing algorithms especially for systems employing phase modulation formats will be presented in Section 1.4. Section 1.5 is devoted to the simulation set-up used throughout the thesis. Finally, in Section 1.6, the main contributions of this thesis are roughly summarized.

1.2 State-of-the-art of optical transmission systems

Optical transmission systems have been evolving in the last three decades in terms of capacity, reach and the structure of networks. In the 1980's, single span fibers were used to transmit of $\sim$ 622Mbit/s. In the 1990's, the development of Erbium Doped Fiber Amplifier (EDFA) allowed the use of multi span fibers and the increase of the transmission distance. Moreover, the Wavelength Division Multiplexing (WDM) was introduced. In fact, multiplexing different channels in a single fiber where each channel is centered on a specific wavelength, ensures high transmission capacity and lowers the cost per transmitted bit. Such systems were operating at a rate of 2.5Gbit/s per WDM channel.

In the year 2000's, along with the continuous deployment of dense WDM (DWDM) systems, mesh based architectures for traffic were introduced in order to achieve better capacity efficiency. This was made possible through the use of intelligent network elements such as Optical cross-connects or the Reconfigurable Optical Add and Drop Multiplexers (ROADM). Typically, those systems operate at a rate of 10Gbit/s per channel and are widely used nowadays.
Telecom Fibers are used for a wide range of applications including the access (distances $<$ 100km), Metropolitan ($100-300$km), Regional ($300-1000$km) and (Ultra) long-haul ($>$1000km) terrestrial or submarine transmission. In the framework of this thesis, we focus on terrestrial long-haul transmission using optical amplification. For this kind of systems, the $\mathcal{C}$-band ranging from 1528.77nm and 1568.36nm is generally used. This $\mathcal{C}$-band corresponds to the minimum attenuation of the fiber (see Fig.1.2 and placed inside the EDFA bandwidth (see Fig.1.6).
In 2002, the International Telecommunication Union Standardization (ITU-T) recommendation G.694.1 introduced the spectral grid for DWDM applications. G.964.1 specified a grid having exactly 100GHz (about 0.8nm) spacing in optical frequency, with a reference frequency fixed at 193.10THz (1552.52nm). The spacing between channels is presented by B_m in Fig.1.1 that describes the configuration of a typical DWDM system working with NRZ On-Off Keying 10Gbit/s WDM channel. For practical purposes

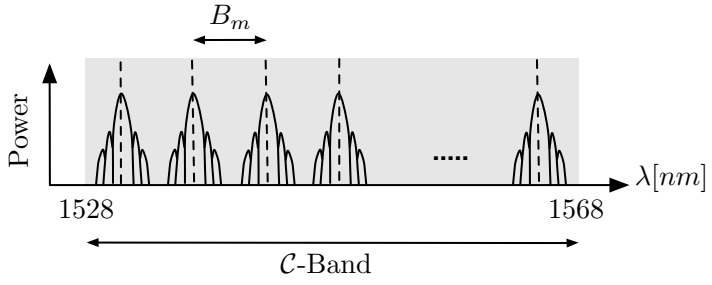


Figure 1.1: Configuration of an optical transmission system using DWDM channels

the grid has been extended to 50GHz and 25GHz spaced grids. Today, most deployed optically-routed mesh networks use 80 WDM channels centered on the 50GHz ITU grid, each channel is of capacity 10Gbit/s using an (Non-return-to-Zero) NRZ On-Off Keying (OOK) modulation format. The total capacity of the fiber link is therefore 0.8Tbit/s across 2000km. The spectral efficiency (SE) is defined as the net per-channel bit rate

R_B divided by the WDM channel spacing B_m

$$SE = \frac{R_B}{B_m} \tag{1.1}$$

Using the definition in Eq. (1.1), the spectral efficiency of most deployed systems today is 0.2bit/s/Hz.
There are different options to increase the total capacity of the already installed DWDM channels, those alternatives will be discussed in details in the next section.

1.2.1 Increasing the transmission capacity of WDM systems

Two main options can be considered to ensure higher transmission capacity:

- The first one concerns the use of extra bands, i.e, operating also outside the $\mathcal{C}$-band.
- The second one consists of operating at a fixed optical amplification bandwidth, i.e., increasing the transmission capacity inside the widely used $\mathcal{C}$-band.

We now discuss the first option: The optical bands are presented in Fig. 1.2 [3]. $\mathcal{C}$-band is widely used in telecommunication applications mainly because of the low losses and because it is included in the EDFA bandwidth. However, $\mathcal{L}$-band, ranging from

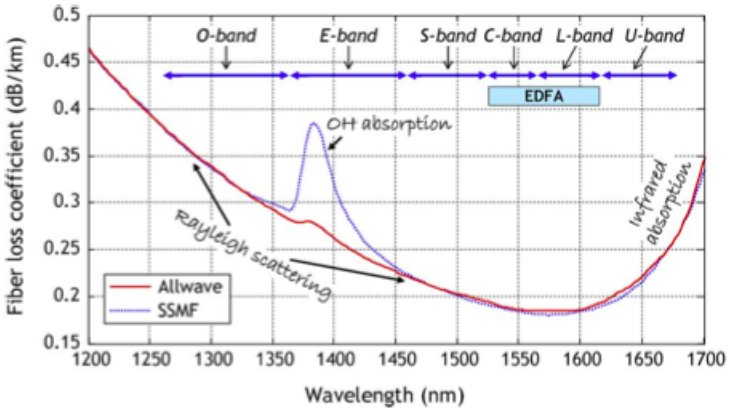


Figure 1.2: Fiber losses versus the wavelength for two types of fiber: Allwave and SSMF

1568.77nm to 1610.49nm , is sometimes used in commercial long-haul systems. Combining $\mathcal{C}$ and $\mathcal{L}$ bands, approximately 160 wavelength channels on the 50-GHZ ITU grid was reached in [4]. Some commercial systems using both $\mathcal{C}$ and $\mathcal{L}$ bands exist [5].
An alternative technique consists of distributed Raman amplification [6]. This solution is less attractive to carriers because it requires expensive investment in the optical components of the network. As a consequence, the first option is not a suitable choice and will move on the analysis of the second one.

The second option is actually a promising method to increase the capacity of the existing systems. Indeed, during the last 15 years, this capacity has increased more than three times, proportionally to the bandwidth demand [7]. This growth has been carried out through increasing the baud rate of the intensity modulated direct-detection

(IM-DD) signal and the number of the DWDM channels inside the $\mathcal{C}$-band. Another solution may consist in using multilevel intensity modulated systems. We discuss now these three alternatives:

- *Reducing the channel spacing inside $\mathcal{C}$-band*:
  Adding more channels raises two main problems: *coherent WDM crosstalk* and *filter narrowing* [1, 8].
  The *coherent WDM crosstalk* is due to interference of the received optical signal field and the residual optical field of a neighboring channel. This leads to signal distortion after WDM demultiplexing and system penalties.
  The *filter narrowing* is the fact that the concatenation of several multiplexing-demultiplexing filters in today's mesh routed optical networks narrows the overall (equivalent) filter and distorts the signal.

- *Using Multilevel intensity modulation*:
  M-ary Amplitude Shift Keying have showed poor performances for fiber-optic transmission so far, mainly due to important back-to-back receiver sensitivity penalty compared to binary OOK [1]. For example, 4-ASK produce a penalty of about 8dB.

- *Increasing the Baud rate of IM-DD*:
  The intensity modulation system is widely used because of its relative tolerance to the Amplified Spontaneous Emission (ASE) noise and the simplicity of the modulators and the receivers. The receivers use the Direct-Detection scheme with a threshold detector.
  Keeping the same NRZ-OOK modulation format and increasing the bit rate raises many problems: First, The spectrum of 40 and 100Gbit/s becomes too large and cannot fit into the 50GHz ITU grid. Second, the robustness against the propagation impairments such as the Chromatic Dispersion (CD) and the Polarization Mode Dispersion (PMD) decreases. Tolerance to CD is divided by a factor of 16 when increasing the bit rate to 40Gbit/s and by a factor of 100 for 100Gbit/s. Concerning PMD, only 2.5ps is tolerable for 40Gbit/s and 1ps for 100Gbit/s. Third, the sensitivity of the receiver which is lowered considerable inducing shorter transmission distances. The penalties due to the increase of the symbol rate are summarized in Table 1.1. Moreover, non linearity induced by Kerr effect induces

Data rate	CD [ps/nm]	PMD [ps]	OSNR penalty [dB]
10Gbit/s	1600	12	0
40Gbit/s	100	2.5	6
100Gbit/s	16	1	10

Table 1.1: CD, PMD tolerance and OSNR penalty (10Gbit/s) is taken as the reference for NRZ-OOK system using different data rates [9]

  more penalty for data rate of 40 and 100Gbit/s [9].

These three options are not convenient for the upgrade of the already installed WDM systems. Therefore the increase of the spectral efficiency of the system through the use

of advanced modulation formats and detection schemes is required. In Table 1.2, we will summarize the performance of a 40Gbit/s system using different modulation formats that fit into the 50GHz ITU spectral grid.

Modulation format	Duobinary	NRZ-DQPSK	RZ-DQPSK	POLMUX NRZ-QPSK
Bits/Symbol	1	2	2	4
DGD*/PMD*	7/2.3 ps	16/5.3 ps	20/6.5 ps	75/25 ps
CD	325 ps/nm	400 ps/nm	320 ps/nm	thousands ps/nm
Non linear effects	Slightly better	Better	Better	Better
gain in OSNR** (0.5nm)	-3dB	+2dB	+4dB	+5dB
50GHz grid	Yes	Yes	Yes	Yes
Mixed 10/40G				
complexity/cost/size	+	+++	++++	+++++

Table 1.2: performance of the 40Gbit/s using different modulation formats [9], *1-dB OSNR penalty @10^{-9}, **OSNR sensitivity in 1nm @10^{-9}, NRZ-OOK is the reference

The binary modulation (duobinary) format has the worst performance in comparison to DQPSK (RZ or NRZ) and Polarization Multiplexing (POLMUX)-NRZ-QPSK that both use multilevel modulation formats. POLMUX-NRZ-QPSK has the best spectral efficiency operating at 10Gbaud, but requires the use of the coherent detection explained in more details in the next section.

1.2.2 Coherent detection

Optical systems using a Direct detection receiver remained the widely used technique for the last four decades. This receiver, simple to operate, makes decisions based only on the intensity of the received signal as depicted in Fig 1.3-(a). Therefore, it allows only amplitude based modulation formats such as the OOK (which offer very poor performance when the number of states increases) and differential binary phase shift keying (DPSK) which suffer of the same drawback. Differential Quadrature PSK (DQPSK) can be obtained as well using more complicated receiver based on the coherent detection.

The basic idea behind coherent detection is depicted in Fig. 1.3-(b). It consists of combining the optical signal coherently with continuous-wave (CW) optical field before it falls on the photo-detector [10].

If we consider that the incident electric field can be written as follows for both the received signal $E_{rx}(t)$ and the LO signal $E_{LO}(t)$:

$$\begin{aligned} E_{rx}(t) &= \sqrt{P_{rx}(t)}e^{i.(\omega_{rx}t+\phi_{rx}(t))} \\ E_{LO}(t) &= \sqrt{P_{LO}(t)}e^{i.(\omega_{LO}t+\phi_{LO}(t))} \end{aligned} \tag{1.2}$$

where P_{rx} and P_{LO} are the power of the received signal and the LO respectively. The electric fields incident on the upper and lower photo-diodes of the balanced detector are

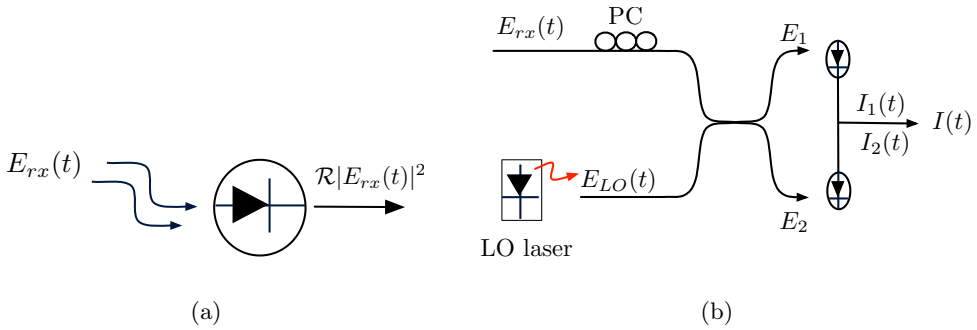


Figure 1.3: Block diagram of an optical receiver using: (a) Direct detection (b) Coherent detection

given as [11]

$$E_1 = \frac{1}{\sqrt{2}}(E_{rx}(t) + E_{LO}(t)), \tag{1.3}$$

$$E_2 = \frac{1}{\sqrt{2}}(E_{rx}(t) - E_{LO}(t)). \tag{1.4}$$

Therefore, the balanced detector output is given as

$$I(t) = I_1(t) - I_2(t) = 2\mathcal{R}\sqrt{P_{rx}(t)P_{LO}(t)}\cos(\omega_{IF}t + \phi_{rx}(t) - \phi_{LO}(t)) \tag{1.5}$$

with $\mathcal{R}$ is the responsivity of the photo-diode and $\omega_{IF} = \omega_{rx} - \omega_{LO}$ is the intermediate frequency. Since P_{LO} can be made much larger than P_{rx}, the power enhancement can exceed 20dB [10]. Although shot noise is also enhanced, the homodyne detection improves the signal-to-noise ratio (SNR) by a large factor.
Coherent detection attracted a lot of attention in the 1970s and the 1980s, proposal of heterodyne receivers for optical communications can be found in [12–14]. This attention is driven by the need to improve the system sensitivity. Indeed, coherent detection provides 4.3dB in noise tolerance over traditional direct detection for coherent BPSK versus OOK [7].
Most of the research activities were abandoned mainly because of the technical problem related to the stabilization of the frequency and phase [15] and after the invention of the EDFA. EDFAs offered an easy way to attain longer reach and a gain in OSNR and it was rapidly deployed in WDM networks.

1.2.2.1 Phase diversity coherent receiver

For optical coherent detection, the six-port 90^o optical hybrid mixes the incoming signal with the four quadrature states associated with the reference signal, emitted by the LO in the complex-field space. The optical hybrid then delivers the four light signals to two pairs of balanced detectors and is depicted in Fig. 1.4. Using the 90^o optical hybrid, we

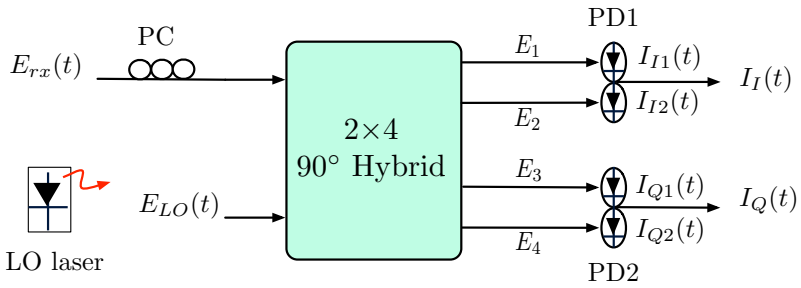


Figure 1.4: Block diagram of a phase diversity coherent receiver (Single Polarization)

obtain the four outputs E_1, E_2, E_3 and E_4 from the two inputs E_{rx} and E_{LO} as

$$\begin{aligned} E_1 &= \frac{1}{2}(E_{rx} + E_{LO}) \\ E_2 &= \frac{1}{2}(E_{rx} - E_{LO}) \\ E_3 &= \frac{1}{2}(E_{rx} + i.E_{LO}) \\ E_4 &= \frac{1}{2}(E_{rx} - i.E_{LO}) \end{aligned} \tag{1.6}$$

(1.7)

the output photo-currents for balanced photo-detectors can be expressed as

$$I_I(t) = I_{I1}(t) - I_{I2}(t) = 2\mathcal{R}\sqrt{P_{rx}(t)P_{LO}(t)}\cos(\omega_{IF}t + \phi_{rx}(t) - \phi_{LO}(t)) \tag{1.8}$$

$$I_Q(t) = I_{Q1}(t) - I_{Q2}(t) = 2\mathcal{R}\sqrt{P_{rx}(t)P_{LO}(t)}\sin(\omega_{IF}t + \phi_{rx}(t) - \phi_{LO}(t)) \tag{1.9}$$

The currents corresponding to the In-phase and Quadrature components are represented by I_I and I_Q respectively and can be given by

$$\begin{aligned} I_I(t) &= \mathcal{R}\sqrt{P_{LO}(t)P_{rx}(t)}\cos(\phi(t)) \\ I_Q(t) &= \mathcal{R}\sqrt{P_{LO}(t)P_{rx}(t)}\sin(\phi(t)) \end{aligned} \tag{1.10}$$

with $\phi(t) = \phi_{rx}(t) - \phi_{LO}(t)$. The current $I_I(t) + i.I_Q(t)$ allows to retrieve both the phase and the amplitude of the incident optical signal provided that we estimate the intermediate frequency. As a consequence, the use of advanced modulation formats such as Quadrature and Amplitude Modulations (QAM) and M-ary Phase Shift Keying (M-PSK) becomes possible.

1.2.2.2 Phase and polarization diversity receiver

In order to increase the spectral efficiency of the coherent systems, and utilize the polarization degree of freedom, multiplexing independent modulated data streams in the two orthogonal polarizations is used in 40Git/s to ensure a transmission at 10Gbaud and for the coherent 100Gbit/s. Those systems are denoted by Polarization Multiplexing (POLMUX). For POLMUX coherent system, a phase and polarization diversity receiver is used to map the phase and in-quadrature components of the incident signal contained in both polarizations. The block diagram of such a receiver in depicted in Fig. 1.5.

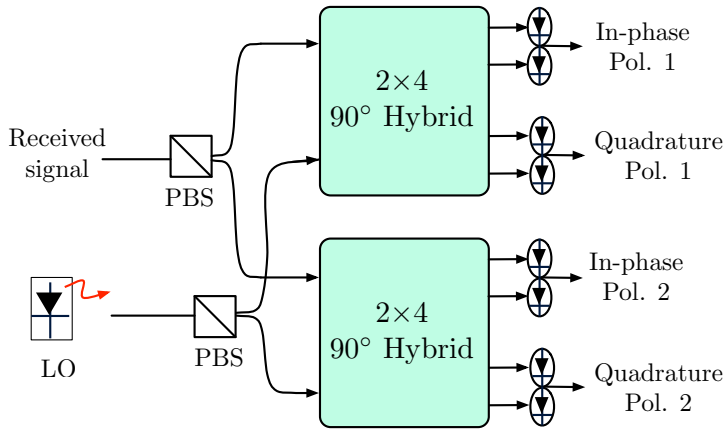


Figure 1.5: Block diagram of a phase and polarization diversity optical coherent receiver (POLMUX), LO and PSB stand for Local oscillator and Polarization Beam Splitter respectively

1.3 Signal model

In this section, we firstly motivate your choice of modulating the transmit signal by means of a single carrier modulation rather than of a multiple carrier modulation. Then, in Subsection 1.3.2, we introduce the various sources of degradation of the fiber. In Subsection 1.3.3, we summarize the impairments considered in this thesis, actually the linear ones, and we provide the mathematical signal model. Finally a discussion devoted to OSNR is drawn in Subsection 1.3.4.

Finally, in Subsection 1.4, we give the state-of-the-art related to the digital processing algorithms used for mitigating the mentioned (linear) impairments.

1.3.1 Single versus Multiple carrier transmission

At our target baud rate, the fiber propagation will introduce Inter-Symbol Interference. In the literature, two main approaches can be advocated to deal with. The first one is the so-called *single carrier* approach and it requires an equalization step at the receiver side, *i.e.*, a signal manipulation mitigating the Inter-Symbol Interference phenomenon. The second one is the so-called *multiple carrier* approach, also known as OFDM. In the sequel, we will discuss the state-of-the-art of these techniques in the framework of coherent optical communications.

Currently, only systems operating at 40 and 100Gbit/s are available. Those systems use the single carrier POLMUX QPSK modulation based on coherent detection method. Other advanced single carrier modulations were explored experimentally and they include 16-QAM [16, 17], 32-QAM [18] and even 256-QAM [19].
High order QAM formats increase the bit rate of the fiber-optic channel, but, at the expense of higher OSNR requirement at the receiver. Therefore, a combination of QAM and strong advanced Forward Error Correction(FEC) used to increase the sensitivity of the system is a suitable alternative for the near future systems operating at 400Gbit/s or 1Tbit/s [20].
The use of multiple carrier for optical communication has been a hot topic since 2006.

Indeed, the optical Orthogonal Frequency Division Multiplexing (OFDM) has been investigated as a potential candidate for 100Gbit/s Ethernet [21]. Unlike single carrier transmission systems, OFDM systems require additional overhead due to the cyclic prefix. This leads to a lower spectral efficiency in comparison to single-carrier transmission [22]. Nevertheless the Inter-Symbol Interference handling is easier since the equalization step is done subcarrier per subcarrier with one-tap (trivial) filter. However, the complexity of its transmitter (especially due to the FFT) and limited performance on periodically compensated dispersion maps makes the OFDM less attractive for 100Gbit/s Ethernet [20].
For all these reasons, **we focus in this thesis on single carrier transmission using advances modulation formats and POLMUX**.

1.3.2 Propagation impairments

Different effects contribute to the distortion of the transmitted signal. At the receiver side, the phase, amplitude and polarization of the signal are altered mainly due to the propagation effects and the subsystems and optoelectronic components imperfections. In this part, we will discuss in detail the propagation impairments that can be divided into the *linear* ones (in Section 1.3.2.1) and the *nonlinear* ones (in Secs. 1.3.2.2 and 1.3.2.3).

1.3.2.1 Linear Impairments

1.3.2.1.1 Attenuation and optical amplification The power of an optical signal propagating in an optical fiber is attenuated before it reaches the receiver. In the wavelength range $(1200-1700\text{nm})$, the attenuation has an overall tendency to increase with increasing wavelength, below 1300nm due to Rayleigh scattering. Secondly, it shows absorption peaks due to hydroxyl (OH^-) around 1400nm. Hence, this absorption peak can be eliminated by reducing the concentration of OH^- ions in the core of the fiber. *AllWave* Fiber of "OFS" has Zero Water Peak (ZWP) as presented in Fig. 1.2. Above 1600nm, the attenuation has a tendency to increase mainly due to SiO_2 absorption. The minimum absorption is obtained around 1550nm corresponding to the $\mathcal{C}$-band and it is $\sim 0.2\text{dB/km}$.
In order to ensure long haul-transmission, optical signal needs to be amplified after each span of $50-100\text{km}$.
The widely used type of amplifiers are the EDFA, The signal to be amplified and a pump laser are multiplexed into the doped fiber, and the signal is amplified through stimulated emission of photons generated by Erbium ions transitions. Besides decaying via stimulated emission, electrons in upper level energy can also decay by spontaneous emission, which occurs at random. So, photons are emitted spontaneously in all directions, and a proportion of those will be captured in the fiber. Those captured photons may interact with other dopant ions, and are therefore amplified by spontaneous emission. This initial spontaneous emission is therefore amplified in the same manner as the incident signals, resulting in *Amplified spontaneous Emission* Noise. This noise is considered one of the main limitations of optical communication systems. This noise has the characteristics of additive Gaussian noise.
The bandwidth of EDFAs are typically $\sim$5THz as presented in Fig.1.6. This bandwidth

limits the use of additional WDM channels. Independently of the polarization bases

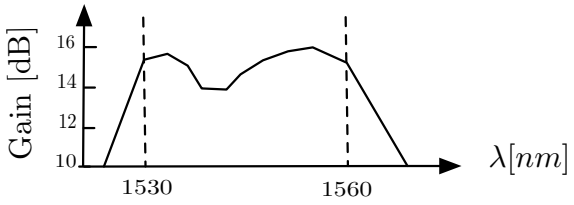


Figure 1.6: Gain Curve of a typical EDFA

chosen to represent signals, ASE noise is white and Gaussian on both components. For both polarizations POL.X and POL.Y, the noise can be expressed as follows:

$$n_p(t) = n_I^p(t) + i.n_Q^p(t) \tag{1.11}$$

Both the In-phase n_I and Quadrature n_Q component on each polarization denoted by p are "locally" white and their Power Spectrum Density (PSD) is denoted $N_0/2$. For periodically spaced discrete EDFAs, the noise PSD per state of polarization N_0, generated at the end of a transmission line composed of a chain of N_A amplifiers spaced by fiber spans of length L_A is given by

$$N_0 = N_A(e^{\alpha_p L_A} - 1)h\nu_s n_{sp} \tag{1.12}$$

where $h\nu_s$ is the photon energy and n_{sp} is the spontaneous emission factor ($n_{sp} \geq 1$) [3].

1.3.2.1.2 Chromatic Dispersion Chromatic dispersion (CD) is caused by a combination of waveguide and material dispersion. Waveguide dispersion can be controlled by careful design [23]. However, material dispersion is due to the wavelength (and thus frequency) dependency of the fiber refractive index. Indeed, the spectral components of the modulated signals travel at a different speeds in the fiber. This causes some wavelengths to arrive before others and therefore the signal pulse to broaden.
In single mode optical fibers, neglecting nonlinearities and polarization related effects, the optical field $E(z,t)$ propagates according to the following wave equation:

$$\frac{\partial E}{\partial z} = -\frac{\alpha_p(z)}{2}E - i\beta(\omega)E \tag{1.13}$$

where, z is the transmission distance, ω the angular frequency, $\alpha_p(z)$ is the fiber attenuation coefficient (we have $\alpha_p(z) = \alpha_0$ for a transmission fiber with uniform loss) and $\beta(\omega)$ is the propagation constant.

$$\beta(\omega) = n(\omega)\frac{\omega}{c} \tag{1.14}$$

where $n(\omega)$ is the fiber refractive index.
As the spectral components of the signal are close to the carrier angular frequency ω_0, the wavelength dependency of $\beta(\omega)$ can be approximated using a Taylor series expansion given by Eq. (1.15)

$$\beta(\omega) \simeq \beta_0 + \beta_1(\omega_0)(\omega - \omega_0) + \frac{\beta_2}{2}(\omega - \omega_0)^2 \tag{1.15}$$

where

$$\beta_i = \frac{d^i \beta(\omega)}{d\omega^i}|_{\omega=\omega_0} \tag{1.16}$$

and ω_0 is the angular reference frequency corresponding to the carrier of the transmitted signal.

The term β_0 represents a constant phase shift, β_1 corresponds to the speed at which the envelope of the pulse propagates, so, the group-velocity of the pulse v_g is defined as $\beta_1 = 1/v_g$, β_2 represents the group velocity delay (GVD) [in ps^2/km] defining the acceleration of the spectral components of the pulse.

In optical communications engineering, the dispersion parameter D at wavelength λ defined in Eq. (1.17) is widely used to characterize CD

$$D = -\frac{2\pi c}{\lambda^2}\beta_2 \tag{1.17}$$

where c is the velocity of the light, Neglecting the constant phase shift propagation delay in the fiber and the attenuation which do not really distort the signal, Eq. (1.13) can be solved in the frequency domain as follows:

$$\tilde{E}(z,\omega) = \tilde{E}(0,\omega)e^{i(\frac{\omega^2\beta_2 z}{2})} \tag{1.18}$$

where $\tilde{E}(z,\omega) = \int E(z,t)e^{i\omega}d\omega$ is the Fourier transform of the signal in the time domain. Using Eqs. (1.18) and (1.13), we obtain that the frequency channel response for the **CD** phenomenon is given as follows by

$$\tilde{\mathbf{C}}_{\mathrm{CD}}(\omega) = \begin{bmatrix} e^{-i\frac{\lambda^2\omega^2 DL_f}{4\pi c}} & 0 \\ 0 & e^{-i\frac{\lambda^2\omega^2 DL_f}{4\pi c}} \end{bmatrix} \tag{1.19}$$

with the fiber length L_f. The response is a matrix since we have considered both polarizations.

Uncompensated CD leads to pulse broadening, causing inter-symbol interference (ISI). Long-haul systems use Dispersion Compensation Fiber (DCF) to compensate CD optically. However inexact matching between the β_2 of transmission fiber and DCF dictates the need for residual dispersion compensation at high bit rates, typically 40Gbit/s or higher. In reconfigurable networks, data can be routed dynamically through different fibers, so the residual dispersion can be time-varying. This necessitates tunable dispersion compensators.

1.3.2.1.3 Polarization Mode Dispersion Polarization Mode Dispersion (PMD) has its origins in optical birefringence. In a perfect fiber, both orthogonal polarizations have the same group delay. However, in reality, fibers have some amount of asymmetry due to imperfections in the manufacturing process and/or mechanical stress on the fiber after manufacture. The asymmetry breaks the degeneracy of the orthogonally polarized modes, resulting in birefringence and a difference in the phase and group velocities of the two modes.

In first order PMD, the difference in group delay $\Delta\beta_1$ between the fast (f) and the slow (s) axis is given by:

$$\Delta\beta_1 = |\beta_{1,f} - \beta_{1,s}| = \frac{\omega}{c}|n_f - n_s| \tag{1.20}$$

where n_f (resp. n_s) is the fiber refractive index of the fast (resp. slow) axis. In the time domain, for a short section of fiber of length L, the differential group delay (DGD), denoted by τ_{DGD}, is defined as $\Delta\beta_1 \times L$ as illustrated in Fig.1.7. Even very small amounts of birefringence can cause evolution of the polarization state as light propagates through fiber. We refer to first-order PMD when the DGD is considered

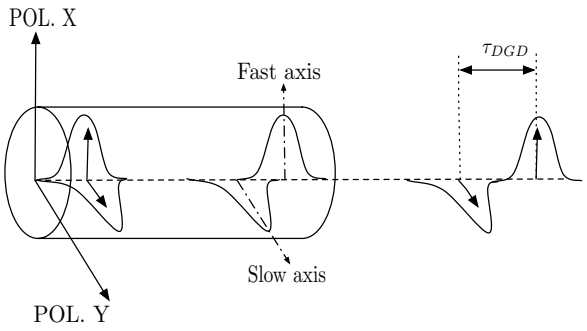


Figure 1.7: Illustration of the time domain effect of PMD in a short fiber

constant over wavelength (as done above), while high-order PMD refers to the fact that DGD changes with wavelength. DGD can be indeed considered constant across a single WDM channel but varying across multiple channels [1].
For first-order PMD, the frequency channel response for the **PMD** phenomenon is given [24, 25]

$$\tilde{\mathbf{C}}_{\mathrm{PMD}}(\omega) = \mathbf{R}_{\theta_1}\mathbf{D}_{\tau_{\mathrm{DGD}},\phi}(\omega)\mathbf{R}_{\theta_2}^{-1} \tag{1.21}$$

with the following birefringence diagonal matrix

$$\mathbf{D}_{\tau_{\mathrm{DGD}},\phi}(\omega) = \begin{bmatrix} e^{i(\omega\frac{\tau_{\mathrm{DGD}}}{2}+\phi)} & 0 \\ 0 & e^{-i(\omega\frac{\tau_{\mathrm{DGD}}}{2}+\phi)} \end{bmatrix} \tag{1.22}$$

associated with the differential group delay between the principle states of polarizations (PSP) τ_{DGD}. Moreover, we have

$$\mathbf{R}_{\theta} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \tag{1.23}$$

which represents the rotation of the reference polarization axis of the fiber's PSPs. Unlike CD which is quasi static, PMD (both DGD and PSP) is a time varying phenomenon. Actually, it fluctuates on a time scale of order of milliseconds [26,27]. As PMD is a time-varying phenomenon, it can be modeled by a random process. For instance, *Poole* [28], and *Foschini et al.* [29], showed that τ_{DGD} has a Maxwellian distribution, whose mean value $\mathbb{E}[\tau_{\mathrm{DGD}}]$ grows as the square root of fiber length.
In Single Mode Fibers (SMF), $\mathbb{E}[\tau_{\mathrm{DGD}}]$ is typically of order $0.1\mathrm{ps}/\sqrt{\mathrm{km}}$. Therefore PMD has no significant impact in systems using 10Gbit/s NRZ-OOK. However it becomes a limiting factor for systems operating at 40Gbit/s and higher, because τ_{DGD} can be a significant fraction of the symbol period. Uncompensated PMD can lead to system outage.

1.3.2.1.4 Polarization Dependent Loss In POLMUX based systems, Polarization Dependent Loss (PDL) means that the two signals associated with both orthogonal

polarizations reach the receiver with different OSNR. In fact, this is caused by slight polarization dependence in the optical components of the network such as the couplers, isolators and EDFAs. In long-haul transmission systems, the signal passes through a large number of those optical components and their PDL is accumulated. The PDL obviously degrades the quality of the signal and can induce penalty.

1.3.2.2 Nonlinear impairments

Throughout this thesis, we consider only the linear impairments generated by the transmission along the optical fiber. However, it is essential to understand the origin of non-linearities can be can be observed with data from experimental systems.

1.3.2.2.1 Kerr effect The propagation of an optical signal is described by the Non Linear Schrödinger Equation (NLSE). The evolution of an optical field $E(z,t)$ is given by:

$$\frac{\partial E}{\partial z} = \underbrace{-\frac{\alpha_p}{2}E}_{\text{Attenuation}} \underbrace{-i\frac{\beta_2}{2}\frac{\partial^2 E}{\partial t^2}}_{\text{CD}} + \underbrace{i\gamma|E|^2E}_{\text{Kerr effect}} \tag{1.24}$$

The first and the second term of the equation correspond respectively to the power loss and CD. Those two effects were already discussed in the previous subsections. Compared to Eq. (1.13), we have added the last term in the RHS of the previous equation.
The third term represents the non linear contribution, which is proportional to the power of the optical signal and the Kerr non linearity coefficient γ.
The Kerr effect is the most significant non-linear effect. It gives rise to a variation in the refractive index of the fiber $n(z,t)$, and this variation depends in the power of the signal $P(z,t) = |E(z,t)|^2$

$$n(z,t) = n_L + n_{NL}\frac{P(z,t)}{A_{eff}} \tag{1.25}$$

where n_L is the conventional refractive index, n_{NL} is the non linear index coefficient and A_{eff} is the optical mode effective area, i.e., the area of the fiber core.
The non linearity coefficient γ in $[\mathrm{W}^{-1}\mathrm{km}^{-1}]$ combines the nonlinear refractive index and the effective core area of the fiber in the following expression:

$$\gamma = \frac{2\pi}{\lambda}.\frac{n_{NL}}{A_{eff}} \tag{1.26}$$

For SMF-type fiber, typical values for γ are between 1 and $2\mathrm{W}^{-1}\mathrm{km}^{-1}$ depending on the current fiber. The variability is mostly due to A_{eff} that lies between 55 and $120\mu\mathrm{m}^2$.
By neglecting the GVD term (β_2), Eq. (1.24) can be solved in closed-form and the solution is given as follows:

$$E(z,t) = E(0,t)e^{-\frac{\alpha_p}{2}z}e^{-i\Phi_{NL}(z)} \tag{1.27}$$

where $\Phi_{NL}(z)$ is the phase shift caused by the Self Phase Modulation (SPM) phenomenon and takes the following form

$$\Phi_{NL}(z) = \gamma|E(0,t)|^2L_{eff} \tag{1.28}$$

where L_{eff} is the so-called effective length of the fiber. Because of the fiber attenuation, the majority of non-linearity occurs in the first part of the fiber or after the EDFAs. Thus this behavior can be translated in the effective length of the fiber used in Eq. (1.28).

$$L_{eff} = \frac{1 - e^{\alpha_p z}}{\alpha_p} \tag{1.29}$$

At the receiver side, the digital coherent receiver can measure $E(L_f, t)$, where L_f denotes the transmission distance. The measured complex amplitude will be distorted due to a combination of CD, PMD and non linearity [30].

1.3.2.2.2 Inter-channel impairments The non linear inter-channel impairments are related to the non linear effects caused by WDM neighboring channel, those including mainly the *Cross Phase Modulation* (XPM) and *Four wave Mixing* (FWM): XPM is the instantaneous change of the phase of a signal caused by the fluctuation of the optical power in a neighboring channel traveling nearly at the same group velocity.
FWM is a the process of inter-modulation that involves three different signals at different wavelengths, i.e., a fourth signal will be generated due to the interaction of the first three.

1.3.2.3 Other (neglected) impairments

In addition to the propagation impairments that degrade the quality of the received signal, optoelectronic subsystems may cause distortions of the signal as well. Those sources of the degradation of the signal include mainly the In-phase and Quadrature imbalance and the resolution of the processing circuits and the analog to digital converters (ADC).

1.3.2.3.1 In-phase and Quadrature imbalance Phase diversity receivers are vulnerable to imperfections of the optical hybrid, resulting in DC offsets, and errors in both amplitude and phase in the output photo-currents [31]. Other sources for the IQ imbalance include implementation imperfections such as incorrect bias points setting for the IQ and phase ports, imperfect splitting ratio of couplers, photo-diode responsivity mismatch and mis-adjustment of the polarization controllers [32].
All these sources of imperfections destroy the orthogonality between the received channels and degrade the performance of the coherent receiver. Hence, additional digital signal processing is required to compensate for this effect.

1.3.2.3.2 Digital circuits resolution The analog-to-digital conversion is carried out using two subsystems: a sampler converting the analog signal to analog discrete time signal and a quantizer transforming the analog discrete time signal into a finite set of values determined by the resolution of the ADC [25]
Advanced modulation formats are sensitive to the resolution of the ADC and the timing jitter of the sampler. Recent advances made available fast ADC such as the Fujitsu [33] ADC having a 8bits resolution at a rate of 56Gbaud, and the Altera Stratix [34] transceivers operating at 28Gbit/s for 40 and 100Gbit/s applications.

1.3.2.3.3 Laser phase noise Laser phase noise has its origins in spontaneous emission. In coherent systems, the most likely sources of phase noise are the LO or the transmitter laser with linewidth typically $\sim 1-5$MHz for Distributed Feedback (DFB) and $\sim 0.1-0.2$MHz for External cavity Lasers (ECL).
The laser phase noise is modeled by a Wiener process. Let $\varphi_{n,p}$ be the phase noise rotation at symbol n on polarization p. The term $\varphi_{n,p}$ is given by the following equation [35]:

$$\varphi_{n,p} = \sum_{m=-\infty}^{n} \phi_m \tag{1.30}$$

where ϕ_m's are independent and identically distributed random Gaussian variables with zero mean and variance

$$\sigma_\phi^2 = 2\pi\Delta\nu T_s \tag{1.31}$$

where $\Delta\nu$ is the sum linewidth of signal and LO lasers, and T_s is the symbol period.
In non-coherent systems, phase noise has no impact on the performance on the systems since decisions only relied on the intensity of the received signal. However, in coherent systems, information is coded in the phase of the signal. Thus, carrier synchronization is required at the receiver. This can be carried out using whether a Phase Locked Loop (PLL) or a Feed Forward (FF) Carrier synchronizer [36].

1.3.3 Mathematical signal model

In this section, we will model the transmit signal and the received signal. The link between the transmitted one and the received one only assume linear distortions (CD and PMD) and frequency mis-synchronization. The other impairments (PDL, Kerr effect, Inter-channel impairments, I/Q imbalance, laser phase noise, etc) are omitted in this model. Nevertheless, in Chapter 4, some experimental results are involved and thus have taken into account all these effects since these effects may occur on each experience made on the testbed.

The transmitted signal (in baseband) on polarization p is linearly modulated by a i.i.d. sequence of QAM/PSK symbols, denoted by $\{s_p(k)\}_k$, as follows

$$x_{a,p}(t) = \sum_k s_p(k) g_a(t - kT_s) \tag{1.32}$$

where T_s is the symbol period and $g_a(t)$ is the shaping filter and may be, for instance, a NRZ pulse.
The continuous-time received signal (in baseband) after the received filter can be written as follows

$$\mathbf{y}_a(t) = \left(\mathbf{C}_a(t) \star \mathbf{x}_a(t)\right) e^{2i\pi\delta f_a t} + \mathbf{b}_a(t) \tag{1.33}$$

with

- $\mathbf{y}_a(t) = [y_{a,1}(t), y_{a,2}(t)]^{\mathrm{T}}$ the bivariate received signal where $y_{a,1}(t)$ (resp. $y_{a,2}(t)$) is the received signal on X-polarization (resp. Y-polarization), and where the superscript $(.)^{\mathrm{T}}$ stands for the transposition operator.
- $\mathbf{x}_a(t) = [x_{a,1}(t), x_{a,2}(t)]^{\mathrm{T}}$ the bivariate transmitted signal where $x_{a,1}(t)$ (resp. $x_{a,2}(t)$) is the transmitted signal on X-polarization (resp. Y-polarization).

- $\mathbf{b}_a(t) = [b_{a,1}(t), b_{a,2}(t)]^{\mathrm{T}}$ the bivariate circularly-symmetric Gaussian noise with zero mean and variance N_0 per real dimension [37]. We also assume that the noise is white in time and in polarization. As it is circularly-symmetric [37], the In-phase and Quadrature components are independent and identically distributed (i.i.d.).
- the 2×2 MIMO channel whose the impulse response is given as follows

$$\mathbf{C}_a(t) = \begin{bmatrix} c_{a,1,1}(t) & c_{a,1,2}(t) \\ c_{a,2,1}(t) & c_{a,2,2}(t) \end{bmatrix}$$

  where $c_{a,p,p}(t)$ corresponds to the inter-symbol interference created by its own polarization (CD and PMD), and where $c_{a,p,q}(t)$ $(p \neq q)$ corresponds to the inter-polarization interference created by the first-order PMD phenomenon.
- δf_a is the continuous-time frequency offset between the LO and the transmit laser expressed in Hertz.
- $\star$ stands for the convolution product.

Notice that the subscript a stands for a continuous-time/analog signal.

We remind that $\tilde{\mathbf{C}}(\omega) = \int \mathbf{C}_a(t) e^{i\omega t} \mathrm{d}t$ is the Fourier transform of the continuous-time channel impulse response $\mathbf{C}_a(t)$. In our derivations or algorithm designs, we will assume that

$$\tilde{\mathbf{C}}(\omega) = \tilde{\mathbf{C}}_{\mathrm{CD}}(\omega)\tilde{\mathbf{C}}_{\mathrm{PMD}}(\omega) \tag{1.34}$$

where $\tilde{\mathbf{C}}_{\mathrm{CD}}(\omega)$ is given by Eq. (1.19) and $\tilde{\mathbf{C}}_{\mathrm{PMD}}(\omega)$ is given by Eq. (1.21)

1.3.4 Optical Signal to Noise Ratio (OSNR)

The amplified spontaneous emission (ASE) represents an optical source of Gaussian noise [1] which can be considered "white" across the signal spectrum. The Optical Signal to Noise Ratio (OSNR) is defined as the average optical signal power divided by the ASE power, measured in both polarizations and in a reference bandwidth fixed to 0.1nm corresponding to the resolution bandwidth of optical spectrum analyzers at 1550nm.
The OSNR is given by [3]:

$$\mathrm{OSNR} = \frac{P}{2N_{ASE}B_{ref}} \tag{1.35}$$

where P is the total average signal power summed over the two states of polarization, N_{ASE} is the spectral density of the ASE in one polarization and the reference bandwidth B_{ref} fixed to 0.1nm (12.5GHz).
In digital communications, the SNR of an additive white Gaussian noise channel (AWGN) is given by the following equation:

$$\mathrm{SNR} = \frac{E_s}{N_0} \tag{1.36}$$

where N_0 is the noise spectral density and E_s is the energy per symbol, $E = PT_s$ with P the average power of one modulated symbol. In the case of POLMUX, the OSNR and SNR are related

$$\mathrm{OSNR} = \frac{2R_s}{2B_{ref}}\mathrm{SNR} \tag{1.37}$$

where $R_s = 1/T_s$ is the symbol rate, and where N_{ASE} and N_0 are assumed to be equivalent.
An other SNR is very useful in digital communications. It is the so-called E_b/N_0, namely, the SNR per bit. We get

$$E_b/N_0 = \frac{\text{SNR}}{m} \tag{1.38}$$

where $m = \log_2(M)$ is the number of encoded bits per symbol.
Therefore, the relationship between the OSNR and $\text{SNR}_b = E_b/N_0$ is given by:

$$\text{OSNR} = \frac{R_b}{2B_{ref}}\text{SNR}_b \tag{1.39}$$

where R_b is the bit rate.

Assuming the use of Gray coding (presented in Fig.1.8 for 4-QAM and 16-QAM modulation formats), the BER of M-QAM constellation with coherent detection in AWGN channel is approximated as [36, 38]:

$$P_b^{M-\text{QAM}} \approx \frac{4}{m}\frac{\sqrt{M}-1}{\sqrt{M}}Q\left(\sqrt{\frac{3m\text{SNR}_b}{M-1}}\right) \tag{1.40}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-u^2/2}du = \frac{1}{2}\text{erfc}\left(\frac{x}{\sqrt{2}}\right)$$

is the Gaussian tail for real-valued Gaussian variable with zero-mean and unit-variance.

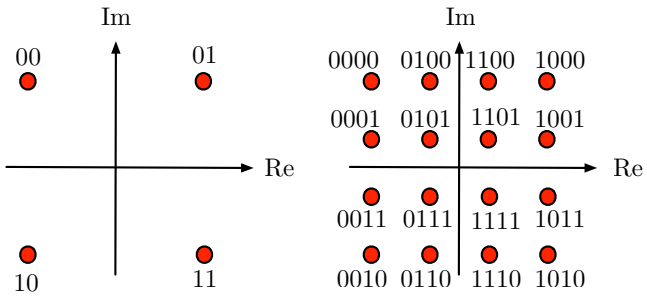


Figure 1.8: Constellation of QPSK and 16-QAM using Gray encoding

In Fig. 1.9, as an illustration, we have plotted the theoretical error probability (given by Eq. (1.40)) in AWGN channel and the simulated Bit Error Rate (BER) for various constellation sizes in a Back-to-Back context (with 14GBd data rate) versus the OSNR.

1.4 State of the art of Digital Signal Processing

After phase and polarization diversity receiver and anti-aliasing filtering, in order to satisfy Shannon's sampling theorem, the signal is sampled at twice the baud rate. Due to the oversampling, no information is lost, and we can omit timing synchronization step. We thus focus on $y_p(n) = y_{a,p}(nT_s/2)$ where we remind that p stands for the

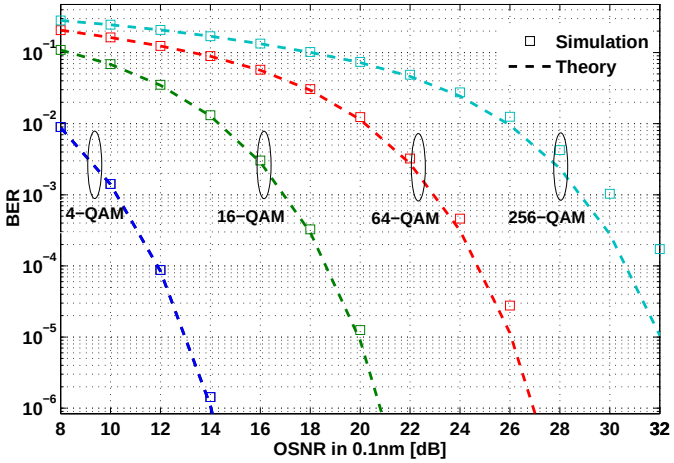


Figure 1.9: Theoretical and simulated BER for QAM versus OSNR

polarization p. In order to "work" at the symbol rate, we stack two consecutive received samples into a bivariate process as follows

$$\mathbf{y}_p(n) = [y_{a,p}(nT_s), y_{a,p}(nT_s + T_s/2)]^{\mathrm{T}}. \tag{1.41}$$

Before going further let us introduce the global filter: $h_{a,p,q}(t) = c_{a,p,q}(t) \star g_a(t)$. We assume that the dispersion time of the channel is roughly upper-bounded by $(K-1)T_s$ whatever the considered polarizations. The discrete-time received signal for the polarization p takes the following form

$$\begin{aligned} \mathbf{y}_p(n) &= e^{2i\pi\tilde{\varphi}_1 n} \sum_{k=0}^{K-1} \mathbf{h}_{p,1}(k) s_1(n-k) \\ &+ e^{2i\pi\tilde{\varphi}_1 n} \sum_{k=0}^{K-1} \mathbf{h}_{p,2}(k) s_2(n-k) \\ &+ \mathbf{b}_p(n) \end{aligned} \tag{1.42}$$

where

$$\mathbf{h}_{p,q}(n) = [h_{a,p,q}(nT_s), h_{a,p,q}(nT_s + T_s/2)]^{\mathrm{T}}, \tag{1.43}$$

$$\mathbf{b}_p(n) = [b_{a,p}(nT_s), b_{a,p}(nT_s + T_s/2)]^{\mathrm{T}} \tag{1.44}$$

and

$$\tilde{\varphi}_1 = \delta f_a T_s/2 \tag{1.45}$$

is the (normalized) discrete-time frequency offset. Notice, in our model, the constant phase offset is encompassed in the channel impulse response. Moreover, we will assume that the channel impulse response and the frequency offset is static over the entire observation window.

One of the goal of the digital signal processing in the framework of a coherent receiver is to compensate for the propagation linear and nonlinear impairments as well as the circuit imperfections. Even if a joint processing dealing with all the impairments will be optimal (but will also induce a huge computational load), it is worth splitting

the processing into several steps. Each step will only focus on one impairment which enables us to reduce the complexity [39].

As this thesis will only focus on mitigating the linear impairments, we introduce here the different steps associated with the linear impairments :

- The CD compensation: usually the step boils down to a "static" equalization of the CD by means of Zero-Forcing equalizer principle (see Section 1.4.1). This equalizer is usually implemented with analog device and is not modified at all.

- The PMD and residual CD compensation: although the first step, CD may be remained sine the fiber is not perfectly known. Therefore, residual CD still occurs and treated jointly with the PMD compensation. Usually as the PMD is assumed as a time-varying phenomenon, "adaptive" equalization is carried out (see Section 1.4.2). As the most important amount of CD has already been compensated for, the number of taps of this equalizer could be small. Nevertheless, in this thesis, we will propose to do this step by means of "static" equalization since the PMD is slow time-varying compared to the baud rate. More details and justifications will be given in Chapter 4.

- The Carrier Frequency Offset (CFO) compensation(see Section 1.4.3).

- The Constant Phase compensation(see Section 1.4.4).

In Fig. 1.10, we summarize the various steps that a receiver has to be carried out.

Notice that if non linear impairments such as the Kerr effect and the nonlinear phase noise have to be compensated for too, additional nonlinear DSP will be required.

In the sequel, we remind the main ways introduced in the literature for mitigating the four above-mentioned linear impairments.

1.4.1 Chromatic dispersion compensation

The impulse response of CD in the time domain is obtained by applying the inverse Fourier Transform to Eq. (1.19). We assume a line with length L_f. We obtain that

$$\mathbf{C}_{\mathrm{CD}}(t) = \begin{bmatrix} \sqrt{\frac{c}{iD\lambda^2 L_f}} e^{(-i\frac{t^2}{2\beta_2 L_f})} & 0 \\ 0 & \sqrt{\frac{c}{iD\lambda^2 L_f}} e^{(-i\frac{t^2}{2\beta_2 L_f})} \end{bmatrix}. \tag{1.46}$$

We are then enable to obtain the associated Zero-Forcing (ZF) equalizer which takes the following form

$$\mathbf{C}^{-1}_{\mathrm{CD}}(t) = \begin{bmatrix} \sqrt{\frac{ic}{D\lambda^2 L_f}} e^{(i\frac{\pi c}{D\lambda^2 L_f}t^2)} & 0 \\ 0 & \sqrt{\frac{ic}{D\lambda^2 L_f}} e^{(i\frac{\pi c}{D\lambda^2 L_f}t^2)} \end{bmatrix}. \tag{1.47}$$

The ZF equalizer approach is here optimal since $\|\mathbf{C}_{\mathrm{CD}}(\omega)\| =$ constant and so does not lead to noise enhancement.

If we would like to implement the equalizer with a discrete-time version, we have to sample it and to truncate the impulse response in order to bound the time duration.

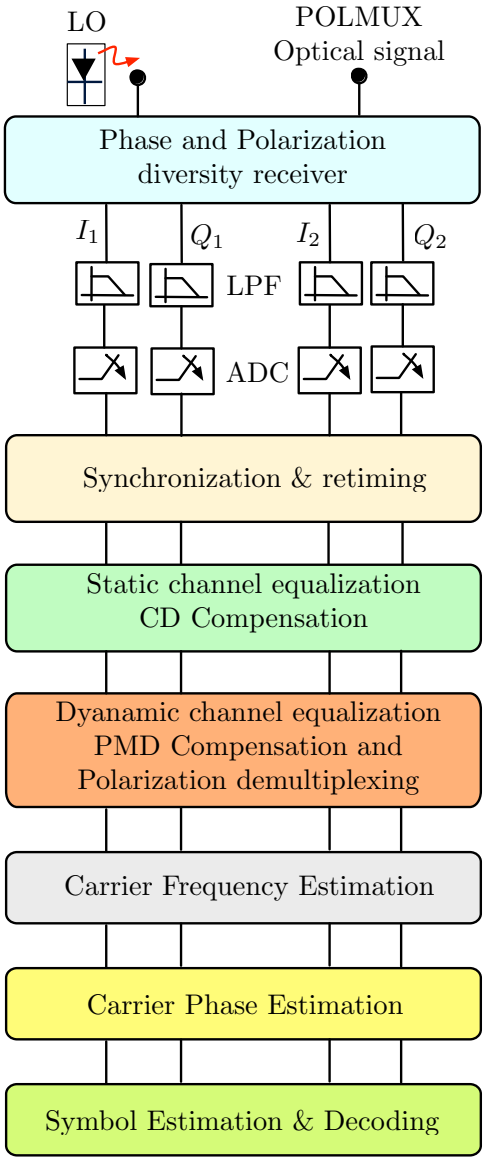


Figure 1.10: Structure of the digital signal processing for a coherent receiver, LPF stands for Low Pass Filter

Hence if we consider a sampling rate $1/T$ and an odd number of taps such that the total number of taps is N_{CD}, we have the following tap weights per polarization

$$c_k = \sqrt{\frac{icT^2}{D\lambda^2 L_f}} e^{(-i\frac{\pi cT^2}{D\lambda^2 L_f}k^2)} \tag{1.48}$$

with $k \in \{-\lfloor N_{\text{CD}}/2 \rfloor, \cdots, \lfloor N_{\text{CD}}/2 \rfloor$ and $N_{\text{CD}} = 2 * \lfloor |D|\lambda^2 L_f/(2cT^2) \rfloor + 1$ and where $\lfloor x \rfloor$ is the integer part of x rounded towards minus infinity. Using Eq. (1.48), when operating at the wavelength $\lambda = 1550$nm, and if we consider a transmission with symbol rate R_s Gbaud and 2 samples per symbol, then the number of taps per 1000ps/nm of CD will be given by $N_{\text{CD}} = 0.031\text{R}_\text{s}^2$. Hence, 4000km of transmission across an SMF fiber (with D = 17ps/nm/km) for a 10 Gbaud system requires less than 250 taps for full CD compensation [40].

1.4.2 PMD and residual CD compensation

The polarization dependent effects, in contrast to CD which may be considered constant, is a time-varying phenomenon. This is mainly due to the relatively rapid variation in the polarization states. Therefore the compensation scheme in existing works have been done in an "adaptive" manner [40]. In PolMux context, the PMD will lead to a 2×2 MIMO frequency-selective channel. Therefore the PMD equalizer also satisfies a 2×2 MIMO structure as depicted in Fig. 1.11. where $\mathbf{y}_p(n)$ is defined in Eq. (1.42).

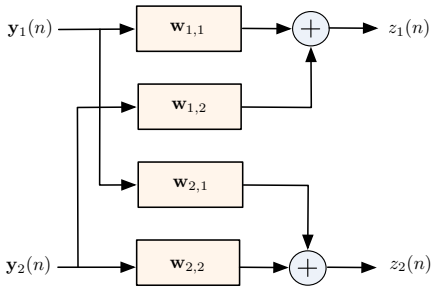


Figure 1.11: MIMO-FS Equalizer butterfly structure

Notice that the length of the FIR filters operating in this structure is generally small and thus it can just compensate for PMD and small amount of (residual) CD.

We remind that the received signal has been sampled at the twice the baud rate in order to satisfy Shannon-Nyquist sampling theorem. Therefore, in order to compensate for the channel impulse response, we have to introduce a $T_s/2$-Fractionally Spaced Equalizer (FSE). Let $z_p(n)$ be the scalar output of the FSE associated with the polarization p. We have

$$z_p(n) = \sum_{k=0}^{L-1} \left(\overline{\mathbf{w}_{p,1}(k)} \mathbf{y}_1(n-k) + \overline{\mathbf{w}_{p,2}(k)} \mathbf{y}_2(n-k) \right) \tag{1.49}$$

where $\{\overline{\mathbf{w}_{p,q}(k)}\}_{k=0,\cdots,L}$ is the filter of length L (notice that each coefficient $\overline{\mathbf{w}_{p,q}(k)}$ is a 1×2 vector, *i.e.*, corresponds to a filter with 2 inputs and 1 output) between the input polarization p and the output polarization q. The overline stands for the complex conjugation.

Eq. (1.49) can be re-shaped easily by means of matrices as follows

$$z_p(n) = \mathbf{w}_p^{\mathrm{H}} \mathbf{y}^{(L)}(n) \tag{1.50}$$

where

- $\mathbf{w}_p = [\mathbf{w}_{p,1}(0), \cdots, \mathbf{w}_{p,1}(L-1), \mathbf{w}_{p,2}(0), \cdots, \mathbf{w}_{p,2}(L-1)]^{\mathrm{T}}$,
- $\mathbf{y}^{(L)}(n) = [\mathbf{y}_1(n)^{\mathrm{T}}, \mathbf{y}_1(n-1)^{\mathrm{T}}, \cdots, \mathbf{y}_1(n-L+1)^{\mathrm{T}}, \mathbf{y}_2(n)^{\mathrm{T}}, \mathbf{y}_2(n-1)^{\mathrm{T}}, \cdots, \mathbf{y}_2(n-L+1)^{\mathrm{T}}]^{\mathrm{T}}$.
- the superscript $(.)^{\mathrm{H}}$ stands for conjugate transposition.

The filters $\mathbf{w}_{p,q}$ have $2L$ coefficients as the received signals have been sampled at twice the baud rate and therefore $\mathbf{w}_p$ has $4L$ coefficients. Moreover, the rate of the sequence

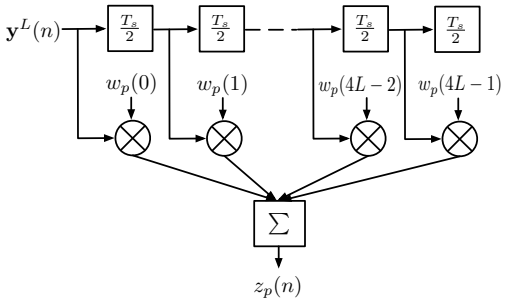


Figure 1.12: FIR equalizer filter structure ($w_p(n)$ denotes the n-th tap of $\mathbf{w}_p$)

$z_p(n)$ is $1/T_s$ as that of $s_p(n)$. The structure of the FIR equalizer filters associated with $z_p(n)$ is plotted in Fig. 1.12.

The ultimate aim of digital equalizer is to exhibit the filter $\mathbf{w}_p$ enabling us to have $z_p(n)$ close to $s_p(n)$.

Depending on whether we use training sequence to estimate the channel, several estimators can be used. Generally in single carrier based optical transmission, blind equalizers are used because of their simplicity and their tracking property. In fact, sending periodic training sequences decreases the spectral efficiency of the transmission systems. In the next paragraphs, we describe the most widely-spread adaptive blind equalizers cited in the literature:

- the Constant Modulus Algorithm (CMA),
- the Radius Directed Equalizer (RDE) which is an adaptation of CMA to QAM modulation formats with different amplitude levels,
- and finally the Decision-Directed equalizer (DD).

For the sake of simplicity, we introduce only the various criteria for one polarization (at the receiver side), namely, p. Similar works can be done for the other polarization.

1.4.2.1 Constant Modulus Algorithm

The estimation of the filter coefficients is based on the constant modulus criterion [41] which looks for the minimization of the following cost function

$$J_{\text{CMA}}(\mathbf{w}_p) = \mathbb{E}[J_{\text{CMA},n}(\mathbf{w}_p)] \tag{1.51}$$

with

$$J_{\text{CMA},n}(\mathbf{w}_p) = (|z_p(n)|^2 - R)^2 \tag{1.52}$$

and

$$R = \frac{\mathbb{E}[|s_p(n)|^4]}{\mathbb{E}[|s_p(n)|^2]} \tag{1.53}$$

In the context of adaptive algorithm, it is usual to implement the stochastic gradient descent algorithm version of the CMA to update the filter coefficients at each sample. Therefore, we have

$$\mathbf{w}_{p,n+1} = \mathbf{w}_{p,n} - \mu \nabla J_{\text{CMA},n}(\mathbf{w}_p)|_{\mathbf{w}_{p,n}} \tag{1.54}$$

$$\nabla J_{\mathrm{CMA},n} = (|z_p(n)|^2 - R)\overline{z_p(n)}\mathbf{y}^{(L)}(n) \tag{1.55}$$

where μ is the constant step-size parameter, $\nabla J_{\mathrm{CMA},n}(.)$ is the gradient at time n, and $\mathbf{w}_{p,n}$ is the equalizer at time n.

In Appendix A.1, we remind how to obtain mathematically Eqs. (1.54)-(1.55). Notice that the approach followed in this Appendix can be extended to the other stochastic gradient algorithms involved in this thesis.

1.4.2.2 Radius Directed Equalizer

The RDE criterion corresponds to an adaptation of CMA to QAM constellations [42,43] where the modulus of the constellation is not constant as shown in Fig. 1.13.

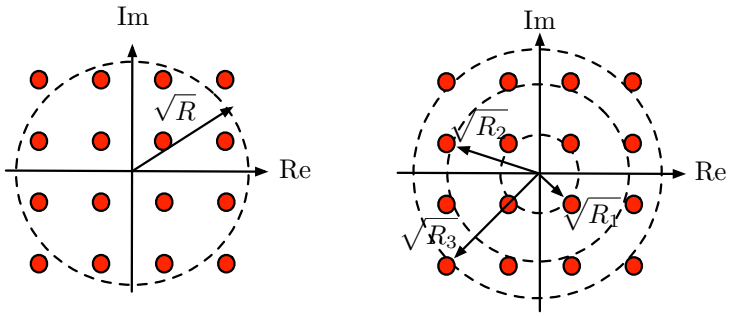


Figure 1.13: Radii used for the CMA (left) and RDE (right) equalization

In RDE equalization [25, 44], the cost function can be written as

$$J_{\mathrm{RDE}}(\mathbf{w}_p) = \mathbb{E}[J_{\mathrm{RDE},n}(\mathbf{w}_p)] \tag{1.56}$$

with

$$J_{\mathrm{RDE},n}(\mathbf{w}_p) = (|z_p(n)|^2 - R_0)^2 \tag{1.57}$$

where R_0 is given by:
if $|z_p(n)| < (\sqrt{R_1} + \sqrt{R_2})/2$ **then**
point belongs to group 1 and $R_0 = R_1$
else
if $|z_p(n)| > (\sqrt{R_2} + \sqrt{R_3})/2$ **then**
point belongs to group 3 and $R_0 = R_3$
else
point belongs to group 2 and $R_0 = R_2$
end if
end if

The stochastic gradient algorithm associated with this cost function is the same as in Eqs. (1.54)-(1.55) where R has to be replaced with R_0.

As an indication, in case of 16-QAM with $E_b = 1$, we actually obtain $R_1 = 0.2$, $R_2 = 1.0$, and $R_3 = 1.8$.

1.4.2.3 Decision Directed Least Square

The Decision Directed (DD) approach leads to the minimization of the following cost function

$$J_{\mathrm{DD}}(\mathbf{w}_p) = \mathbb{E}[J_{\mathrm{DD},n}(\mathbf{w}_p)] \tag{1.58}$$

with

$$J_{\mathrm{DD},n}(\mathbf{w}_p) = |z_p(n) - \hat{s}_p(n)|^2 \tag{1.59}$$

and $\hat{s}_p(n)$ the current decision of symbol $s_p(n)$.

Then the adaptive version (actually the stochastic gradient algorithm) is as follows

$$\mathbf{w}_{p,n+1} = \mathbf{w}_{p,n} - \mu \nabla J_{\mathrm{DD},n}(\mathbf{w}_p)|_{\mathbf{w}_{p,n}} \tag{1.60}$$

$$\nabla J_{\mathrm{DD},n}(\mathbf{w}_p) = \overline{(z_p(n) - \hat{s}_p(n))}\mathbf{y}^{(L)}(n) \tag{1.61}$$

1.4.3 Carrier Frequency Estimation

Carrier Frequency offset (CFO) is due to the frequency mismatch between the LO and the signal lasers. Thanks to the equalization procedure described in the previous section, we can now assume that CD and PMD were perfectly removed, *i.e.*, can be omitted. Therefore the (baud-rate) output of the equalizer on polarization p, already denoted by $z_p(n)$, can be written as follows

$$z_p(n) = s_p(n)e^{2i\pi(\varphi_{0,p}+n\varphi_1)} + b'_p(n) \tag{1.62}$$

where it remains two drawbacks:

- $\varphi_1 = \delta f_a T_s$ is the discrete-time (baud-rate) FO. The FO is independent of the polarization state of the received POLMUX signals.
- $\varphi_{0,p}$ corresponds to the constant phase. This constant phase occurs since the blind equalizer is only able to determine the filter up to a constant phase.

and where $b'_p(n)$ is the additive zero-mean circularly-symmetric complex-valued Gaussian noise.

In the optical communications literature, the most widely-spread carrier frequency estimator (CFE) has been presented for M-PSK modulation [45]. We assume that the CFO is constant or quasi-constant within an observation window. Let L_B be the number of observation samples (when sampled at the baud rate) The CFO is thus usually estimated through the following algorithm

$$\hat{\varphi_1} = \frac{1}{2\pi M}\angle \sum_{m=0}^{L_B-1} [z_p((n+1+m))\overline{z_p(n+m)}]^M \tag{1.63}$$

where $\angle(.)$ stands for the angle of complex-valued number. Elevating the signal to the M^{th}-power eliminates the M-PSK modulation. Therefore this CFO estimator offers excellent performance in the context of PSK modulation. This estimate is summarized in Fig. 1.14.

An extension to M-QAM can be done by replacing the term M with 4 (whatever the QAM size) in Eq. (1.63) However, because of a non constant modulus for the QAM modulation formats, the 4^{th} power of the received signal does not eliminate the modulation and only a rough estimate of the CFO is obtained. Moreover, as QAM formats are more sensitive to the CFO distortion, accurate CFO estimators are still required for this kind of modulation.

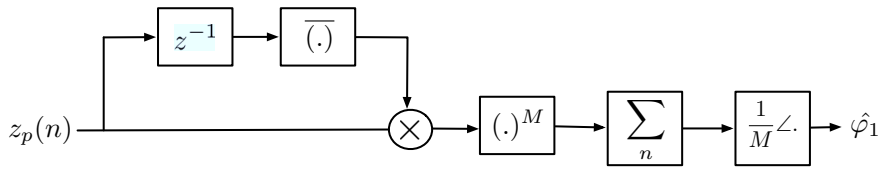


Figure 1.14: Frequency offset estimator for coherent receivers. (z^{-1} corresponds to one symbol delay)

1.4.4 Constant Phase Estimation

Constant phase estimation (CPE) associated with φ_0 is generally carried out after the compensation of the channel impairments as proposed here. Assuming a perfect equalization and CFO estimation (done in previous sections), the received signal can be written as:

$$v_p(n) = s_p(n)e^{2i\pi\varphi_{0,p}} + b''_p(n) \tag{1.64}$$

where $v_p(n) = z_p(n)^{-2i\pi\varphi_1 n}$ and $b''_p(n)$ is still an additive zero-mean circularly-symmetric complex-valued Gaussian noise.

The Viterbi-Viterbi algorithm (for PSK) is based on the elevation of the signal to the M-th power for M-PSK [46, 47] and to the 4-th power for any M-QAM [48]. Thus, we have

$$\hat{\varphi}_{0,p} = \frac{1}{Q}\angle\left(\frac{1}{L_B}\sum_{n=0}^{L_B-1} v_p(n)^Q\right) \tag{1.65}$$

where $Q = M$ for M-PSK and $Q = 4$ for M-QAM. The structure of this CPE is presented in Fig. 1.15. The performance offers by this estimate is generally good enough

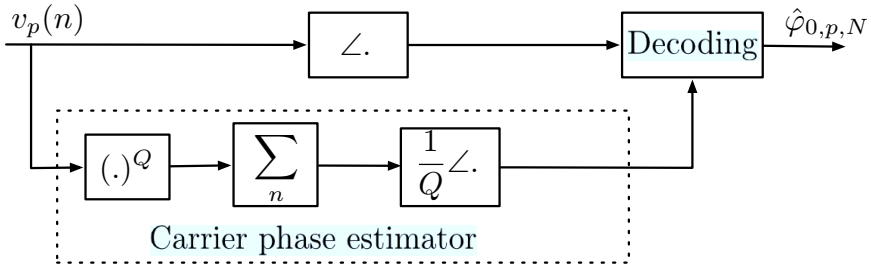


Figure 1.15: Carrier phase recovery for coherent systems

and so leads to significant BER improvement [47]. Notice that this previous algorithm can be improved, if necessary, by applying another (but more complicate) non-linear function to $v_p(n)$ depending on the OSNR value [49].

Once the blind phase estimator has worked, one can move to the Decision-Directed phase estimator which will provide better performance (if well initialized by the blind method). This DD estimator is described as follows

$$\hat{\varphi}_{0,p,\mathrm{DD}} = \angle\left(\frac{1}{N}\sum_{n=0}^{N-1} v_p(n)\overline{\hat{s}_p(n)}\right) \tag{1.66}$$

These estimates can be easily adapted when phase noise occurs, i.e. when the phase is not constant anymore. We remind that, in coherent optical communications, phase noise is sometimes considered as one of the most limiting transmission factors. It is due to the linewidth of the used lasers for the transmitter and the LO. Indeed, the linewidth of the widely used semiconductor distributed-feedback (DFB) lasers typically ranges from 100kHz to 10MHz which is some contexts can not be neglected at all.

The simplest way to adapt our estimates to phase noise is to consider an observation window L_B small enough in order to almost satisfy the constant phase assumption even in phase noise occurs. An other way is to find the stochastic gradient algorithm linked to the estimates given in Eqs. (1.15)-(1.66) respectively. Actually, concerning the blind estimate, this leads to

$$\hat{\varphi}_{0,p,n+1} = \hat{\varphi}_{0,p,n} + \mu_{\text{blind}} \Im[v_p(n)^Q e^{-2i\pi Q \hat{\varphi}_{0,p,n}}] \tag{1.67}$$

where μ_{blind} is the step size and $\hat{\varphi}_{0,p,n}$ the estimated phase at time n.

Concerning the DD estimate, its adaptive version takes the following form

$$\hat{\varphi}_{0,p,n+1,\text{DD}} = \hat{\varphi}_{0,p,n,\text{DD}} - \mu_{\text{DD}} \Im[v_p(n) \overline{\hat{s}_p(n)} e^{-2i\pi \hat{\varphi}_{0,p,n,\text{DD}}}] \tag{1.68}$$

where μ_{DD} is the step-size parameter as done in [43, 50].

In Appendix A.2, we provide the algebraic manipulations needed to translate the blockwise-version phase estimates into the adaptive ones.

1.5 Simulation Set-up

Our simulation set-up of the optical coherent system is as follows:

- a 112Gbit/s transmission is achieved by multiplexing both polarizations with 16-QAM modulated signals which corresponds to 14Gbaud transmission per polarization.
- The transmit shaping filter is a square root raised cosine filter with a roll-off factor equal to 1. This filter is used to reduce the bandwidth of the QAM pulse since rectangular pulses produce very large frequency spectrum.
- The ASE noise is loaded at the receiver before a 50GHz optical filter.
- A matched filter associated with the shaping filter is applied at the receiver side.
- The continuous received electrical signal is sampled at twice the baud rate. A fifth-order Bessel low-pass filter with a 3dB bandwidth equal to 80% of the symbol rate was used as anti-aliasing filter.

Concerning the propagation, except otherwise stated, we will only simulate the main linear channel impairments in fiber-optic transmission:

- CD modeled by Eq. (1.19),
- PMD modeled by Eq. (1.21),
- CFO,

- AWG noise.

At the receiver side, we will proceed into three steps (as shown in Fig. 1.16) to retrieve the transmitted symbols $s_p(n)$ from the noisy observations $\mathbf{y}_p(n)$ given the signal model of Eq. (1.42):

- the blind PMD/residual CD compensation through the evaluation of a MIMO linear fractionally spaced equalizer (FSE). By construction, our blind equalizer is robust to the presence of the CFO.

- the blind CFO estimation. We will see that our estimator performs better if it relies on the post-equalized signal instead of on the pre-equalized signal.

- the blind constant phase estimation. After PMD/residual CD and CFO compensations, as already seen, the constellation may be still rotated by a constant phase since the blind equalizer has phase ambiguity. Adaptive version of this phase estimator will be then able to manage the presence of the laser phase noise.

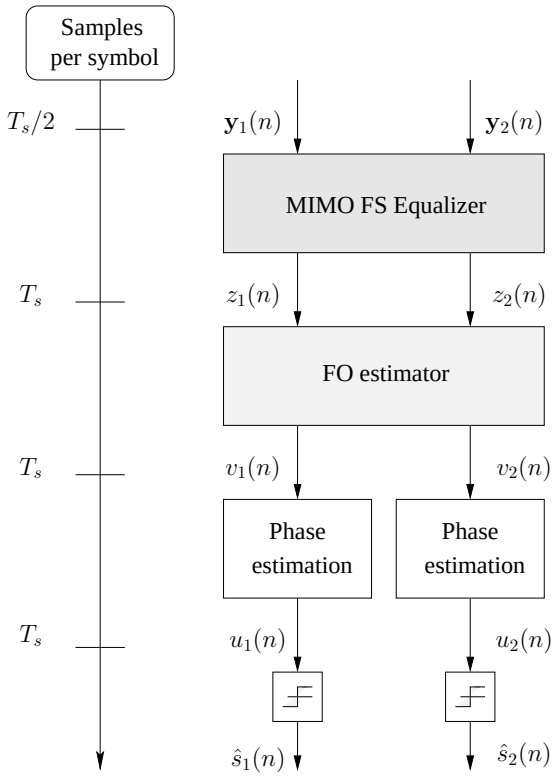


Figure 1.16: Receiver structure

1.6 Outline of the thesis

In this chapter, we have reviewed the state-of-the-art of WDM optical transmission systems. Then, we have discussed the different alternatives to increase the total capacity of fiber links. The promising option is to improve the spectral efficiency of the channel

using the coherent detection. A combination of coherent detection and advanced modulation formats allow to exploit all the degrees of freedom of the system, i.e., using both the amplitude and the phase and the two polarizations to transmit information. Moreover recent advances in digital circuits made the compensation of the main propagation impairments and the system imperfections possible in the electronic domain. Therefore coherent detection is a reasonable way to the increase of spectral efficiency.

In the state-of-the-art of DSP for coherent receivers, we have unveiled some drawbacks:

- rudimentary algorithms are used for blind equalization and demultiplexing. Those equalization algorithms based on the stochastic gradient algorithm suffer from slow convergence due to the use of a fixed step-size. In the context of demultiplexing the PolMux signals, they can lead to the same output on both polarizations which induces the loss of one data stream out of both.
- Moreover, the widely used CFO estimator based on the Viterbi-Viterbi algorithm is unsuitable for QAM formats that have stringent requirements concerning the residual CFO.

In this thesis, we develop new advanced DSP algorithms which are suitable for QAM and PolMux transmission. We especially focus on the improvement of equalizer performance and CFO estimate performance. More precisely, we introduce

- adaptive equalizers with non-fixed step-size. This leads to convergence speed improvement and also steady-state improvement at the expense of a slight additional complexity. This part is drawn in Chapter 2.
- new (block-wise) CFO estimate well adapted to QAM constellation. Chapter 3 is devoted to this topic.
- block-wise equalizers. This kind of approach is for the first time applied in optical communications. We will see, in Chapter 4, that the convergence speed is dramatically improved once again at the expense of a slight additional complexity. We also introduce a modified version of these equalizers in order to prevent singularity issue, *i.e.*, when the same data sequence is on both polarizations.

Chapter 2

Newton based adaptive equalization

2.1 Introduction

So far, the adaptive MIMO equalization is an essential part of a coherent digital processor to mitigate Inter-Symbol Interference caused by the residual CD and the polarization dependent effects (PDE) such as PMD. Many MIMO equalizers, such as the Data aided based algorithm, the Decision-Directed algorithm, the constant modulus algorithm [40, 41], the radius directed equalizer (RDE) initially developed in [32, 42], have been applied in optical communications by means of stochastic gradient descent algorithm with fixed step-size. Such an approach (constant step-size) unfortunately suffers from slow convergence.

In this chapter, we introduce variable step-size version of these gradient algorithms The variable step-size is calculated via the Pseudo-Newton method. The performance are investigated through Monte Carlo simulations. We have observed faster convergence and better steady-state performance.

This chapter is organized as follows: in Section 2.2, we remind the main MIMO equalizers. In Section 2.3, we introduce our new variable step-size versions of these algorithms. In Section 2.4, we analyze numerically the convergence speed of the proposed algorithms. Robustness to time-varying channels is also studied. In Section 2.5, we derive the computational load of the proposed methods compared to the constant step-size approach. Finally, concluding remarks are given in Section 2.6.

2.2 Adaptive equalizers

2.2.1 Adaptive Data-Aided (A-DA)

The Data-Aided (DA) equalization techniques relies on a known training sequence used to estimate the channel or directly the equalizer. This method assumes the synchronization is carried out successfully. Training sequences should be transmitted periodically in order to allow the receiver to estimate the channel or the equalizer. Data aided equalization thus reduces the spectral efficiency of the transmission.

The DA equalizer is based on the minimization of the following cost function

$$J_{\mathrm{DA}}(\mathbf{w}_p) = \mathbb{E}[J_{\mathrm{DA},n}(\mathbf{w}_p)] \tag{2.1}$$

with

$$J_{\mathrm{DA},n}(\mathbf{w}_p) = |z_p(n) - s_p(n)|^2. \tag{2.2}$$

We recall that this minimization can be implemented using the stochastic gradient descent algorithm as follows

$$\mathbf{w}_{p,n+1} = \mathbf{w}_{p,n} - \mu \nabla J_{\mathrm{DA},n}(\mathbf{w}_p)|_{\mathbf{w}_{p,n}} \tag{2.3}$$

with

$$\nabla J_{\mathrm{DA},n}(\mathbf{w}_p) = \overline{(z_p(n) - s_p(n))}\mathbf{y}^{(L)}(n). \tag{2.4}$$

2.2.2 Adaptive Decision-Directed (A-DD)

As already said in Chapter 1, the Decision-Directed equalization approach is based on a decision on the data. However, the DD approach is still a blind equalization algorithm since it does not rely on training sequence. Unlike DA, the cost function is so based on $\hat{s}_p(n)$ instead of $s_p(n)$. Thus, the equations describing the DD adaptive equalizer based on the stochastic gradient algorithm (denoted by A-DD) are given in Eqs. (1.60)-(1.61) and are the same as for the A-DA except the symbol replaced with the detected symbol.

As the A-DD computes the error function based on the decision on the current symbol, it is very sensitive to the presence of phase distortions (CFO, constant phase and phase noise).

2.2.3 Adaptive Constant Modulus Algorithm (A-CMA)

The CMA algorithm relies neither on the symbol decisions nor on a training sequence. This algorithm operates blindly and it is more robust to phase distortions (CFO, constant phase, phase noise) than the DD approach.

Once again, the algorithm, denoted by A-CMA, associated with fixed step-size has been described in previous chapter through Eqs. (1.54)-(1.55).

Despite of its robustness, the CMA nevertheless suffers from some drawbacks especially for high spectral efficiency system based on QAM modulation. Indeed, it is not suitable for M-QAM modulations (with $M \geq 8$) since such modulations do not satisfy constant modulus property. *Savory* [25] showed that the error function has a non zero optimum minimum . Since, the error does not tend to zero at the optimum, if a stochastic gradient method is used, the equalizer continues to adapt erroneously, introducing noise into the signal, and therefore degrading the performance.

For all these reasons, the CMA equalizer may be used for coarse equalization step (pre-convergence) in QAM context. In order to enhance the steady state, it can be followed by a fine equalization step carried out through the A-DD algorithm or the A-RDE algorithm.

2.2.4 Adaptive Radius-Directed Equalization (A-RDE)

The algorithm, denoted by A-RDE, associated with fixed step-size has been described in previous chapter through Eq. (1.57).

The Radius Directed equalization relies on the correct decision regarding the transmitted ring radius. As it has been explained by *Winzer et al.* [51], since the ring spacing in QAM constellations is generally smaller than the minimum symbol spacing, those

decisions may show a significant number of errors especially for strong noise. For this reason, the use of RDE alone to compensate for the residual CD/PMD may fail. Therefore, a pre-convergence using the A-CMA algorithm is required in general. In order to enhance the steady-state of the equalization step, the A-RDE can be applied after the pre-convergence.

All these fixed step-size based algorithms lead to a large number of iterations. As a consequence, the observation window for achieving convergence must be large enough. If the channel is not time-varying, this is not an issue. But if the channel is time-varying (before the convergence has been reached), or if the system has to be often re-initialized, a faster convergence speed may be required. In the sequel, we propose to improve the adaptive algorithm by changing the step-size at each instant. This will be done by using the classical Pseudo-Newton approach in the adaptive algorithm/stochastic approximation literature.

2.3 Pseudo-Newton based adaptive equalizers

The choice of the fixed step-size is a crucial task for gradient algorithm and arises from a trade-off between convergence speed and steady-state performance. High fixed step-size leads to fast convergence and poor steady-state performance whereas small fixed step-size leads to the contrary. To overcome this problem and so improve simultaneously the convergence speed and the steady-state performance, we propose to implement variable step-size approach, *i.e.*, replacing μ with $\boldsymbol{\mu}_n$ in Eq. (2.3) for DA, in Eq. (1.60) for DD and in Eq. (1.54) for CMA/RDE.

Thus, we get

$$\mathbf{w}_{p,n+1} = \mathbf{w}_{p,n} - \boldsymbol{\mu}_n \nabla J_n(\mathbf{w}_p)|_{\mathbf{w}_{p,n}} \tag{2.5}$$

where $\boldsymbol{\mu}_n$ is a 2×2 variable step-size.

To derive $\boldsymbol{\mu}_n$, we consider the Pseudo-Newton algorithm [52] that exploits the Hessian matrix of the cost function. Therefore the names for these algorithms introduced in this chapter will be AN-DA (AN stands for "Adaptive-Newton"), AN-DD, and AN-CMA/AN-RDE.

The Pseudo-Newton approach leads to

$$\boldsymbol{\mu}_n = \mu \mathbf{H}_n^{-1}(\mathbf{w}_{p,n}), \quad \text{with} \quad \mathbf{H}_n(\mathbf{w}_{p,n}) = \frac{\partial^2 J_n(\mathbf{w}_p)}{\partial \mathbf{w}_p^{\mathrm{T}} \partial \overline{\mathbf{w}_p}}\Big|_{\mathbf{w}_{p,n}}. \tag{2.6}$$

After simple algebraic manipulations, the Hessian matrix is given by:

$$\mathbf{H}_n(\mathbf{w}_{p,n}) = f_p(n)\mathbf{y}^{(L)}(n)\mathbf{y}^{(L)^{\mathrm{H}}}(n) \tag{2.7}$$

where $f_p(n)$ is specified for each algorithm in Table 2.1.

	A-DA	**A-DD**	**A-CMA/A-RDE**
$f_p(n)$	1	1	$2\|z_p(n)\|^2$

Table 2.1: $f_p(n)$ for A-DA, A-DD and A-CMA/A-RDE algorithms

The size of $\mathrm{H}_n(\mathbf{w}_{p,n})$ is of $4L \times 4L$. In order to obtain the inverse of the Hessian matrix given by Eq. (2.7), we will not directly inverse it since it raises two problems:

- Huge additional computational load since the matrix inversion complexity is of order $\mathcal{O}(L^3)$.
- High numerical instability since the instantaneous Hessian matrix value fluctuates strongly between two instants.

In order to fix these problems, one solution consists of using a recursive update of the Hessian $\mathbf{H}_n(\mathbf{w}_{p,n})$ incorporated a forgetting factor with past realizations of this matrix. The recursivity will reduce the complexity while the forgetting factor (as done for the so-called RMS algorithm where RMS stands for "Recursive Mean Square") will reduce the instability.

The update structure for the Hessian matrix is thus given as in Eq. (2.8).

$$\mathbf{H}_n(\mathbf{w}_{p,n}) = \lambda \mathbf{H}_{n-1}(\mathbf{w}_{p,n}) + (1-\lambda) f_p(n) \mathbf{y}^{(L)}(n) \mathbf{y}^{(L)^{\mathrm{H}}}(n) \tag{2.8}$$

where λ is a forgetting factor satisfying: $0 \leq \lambda \leq 1$ and assuming [52].
Note that for $\lambda = 0$, we obtain the instantaneous value of the Hessian matrix, i.e., the system has no memory about the past values of the Hessian and any problem has been solved, and for $\lambda = 1$, we just consider the previous value of the Hessian and thus iteratively the Hessian matrix associated with the first sample.

In a second step, the Inverse of the Hessian matrix can be obtained according to *Woodbury inversion lemma*. Let us remind this lemma hereafter. Let $\mathbf{A}, \mathbf{U}, \mathbf{C}$ and $\mathbf{V}$ be matrices of complex elements having sizes respectively $(n \times n)$, $(n \times k)$, $(k \times k)$ and $(k \times n)$, with n and k two positive integers.This lemma "says" that

$$(\mathbf{A} + \mathbf{UCV})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}.\mathbf{U}.(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}.\mathbf{U})^{-1}.\mathbf{V}.\mathbf{A}^{-1} \tag{2.9}$$

If we rewrite Eq. (2.8) as follows

$$\mathbf{H}_n(\mathbf{w}_{p,n}) = \lambda \mathbf{H}_{n-1}(\mathbf{w}_{p,n}) + \mathbf{y}(n)(1-\lambda) f_p(n) \mathbf{y}^{\mathrm{H}}(n) \tag{2.10}$$

and if we identify the different elements as follows

$$\begin{aligned} \mathbf{A} &= \lambda \mathbf{H}_{n-1}(\mathbf{w}_{p,n}) \\ \mathbf{U} &= \mathbf{y}(n) \\ \mathbf{C} &= (1-\lambda) f_p(n) \\ \mathbf{V} &= \mathbf{y}^{(L)^{\mathrm{H}}}(n) \end{aligned} \tag{2.11}$$

we obtain

$$\begin{aligned} \mathbf{H}_n^{-1}(\mathbf{w}_{p,n}) = {} & \lambda^{-1} \mathbf{H}_{n-1}^{-1}(\mathbf{w}_{p,n}) \\ & - \frac{\lambda^{-2} \mathbf{H}_{n-1}^{-1}(\mathbf{w}_{p,n}) \mathbf{y}^{(L)}(n) \mathbf{y}^{(L)^{\mathrm{H}}}(n) \mathbf{H}_{n-1}^{-1}(\mathbf{w}_{p,n})}{[(1-\lambda) f(n)]^{-1} + \lambda^{-1} \mathbf{y}^{(L)^{\mathrm{H}}}(n) \mathbf{H}_{n-1}^{-1}(\mathbf{w}_{p,n}) \mathbf{y}^{(L)}(n)} \end{aligned} \tag{2.12}$$

Note that both the inverse of μ and $1-\lambda$ are measures of the memory of the algorithms. A rule for selecting λ and μ can be as follows [52]

$$\lambda + \mu = 1 \tag{2.13}$$

The initialization of the Hessian matrix is done as follows

$$\mathbf{H}_0^{-1}(\mathbf{w}_{p,0}) = \delta \mathbf{Id} \tag{2.14}$$

with the identity matrix $\mathbf{Id}$ and a fixed positive number δ small enough so that the learning algorithm has reliable initial convergence.

2.4 Convergence analysis

We have considered the following channel: $1000ps/nm$ of CD, $50ps$ of DGD and a polarization rotation angle $\theta_1 = \theta_2 = \pi/4$ in Eqs. (1.19)-(1.21).
The equalizer length (whatever the considered algorithm to calculate it) is $L = 3$, *i.e.*, each $\mathbf{w}_{p,q}$ has 6 complex-valued coefficients. Such a length choice The choice of 6 is typical for real-time implementations where the number of taps should be small to decrease the complexity.

For the sake of simplicity (also as the CMA performance are not degraded by these impairments), we do not have neither phase noise nor CFO. In contrast, a constant phase estimate is necessary after equalizer due to phase ambiguity. Such a phase estimation is done through DD algorithm based on Eq. (1.66). The filter coefficients for both polarizations $\mathbf{w}_p$ are initialized with $\mathbf{w}_{p,0} = (0..010..0)$ and are then updated using the different algorithms and different step-size choices. The OSNR in 0.1nm is fixed to 20dB corresponding to nearly +3dB of the 17dB theoretical OSNR required to obtain a BER of 10^{-3} for 16-QAM.

2.4.1 Convergence of the A-CMA

The convergence of the A-CMA equalizer is presented Fig. 2.1, The BER is plotted versus the number of filter updates. The update is done at the baud rate. As the receive signal has been sampled at twice the baud rate, the number of filter updates corresponds only to half the number of samples.

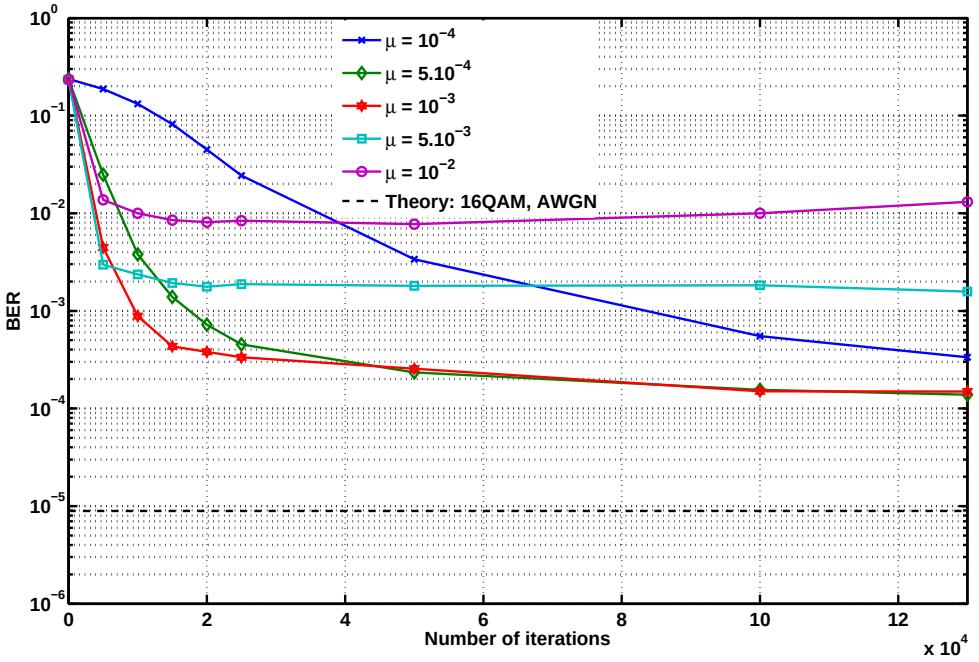


Figure 2.1: Convergence speed of the A-CMA equalizer for different step-sizes μ

The convergence of the equalizer relies largely on the choice of the step-size. On the one hand, a large step-size includes a fast convergence speed, however the steady state

associated with the BER is high. Going the filter coefficients update with too large step-size even degrades the performance as it can be seen for $\mu = 10^{-2}$. On the other hand, small step-size can ensure better BER at steady-state, but at the expense of a slow convergence.
The value of $\mu = 10^{-3}$ corresponds to the best trade-off between the convergence speed and the BER at steady-state. Actually, a BER of 10^{-3} is obtained within 10,000 iterations. For this reason, this value will be used for the rest of this thesis. Notice that the A-CMA equalizer is insensitive to the presence of reasonable phase noise.

2.4.2 Convergence of the AN-CMA

Keeping the value of $\mu = 10^{-3}$ obtained in the last section on the A-CMA, we now want to study the impact of the initialization of the Hessian algorithm (parameter δ) on the performance of the AN-CMA. This convergence is presented for different values of δ in Fig. 2.2.

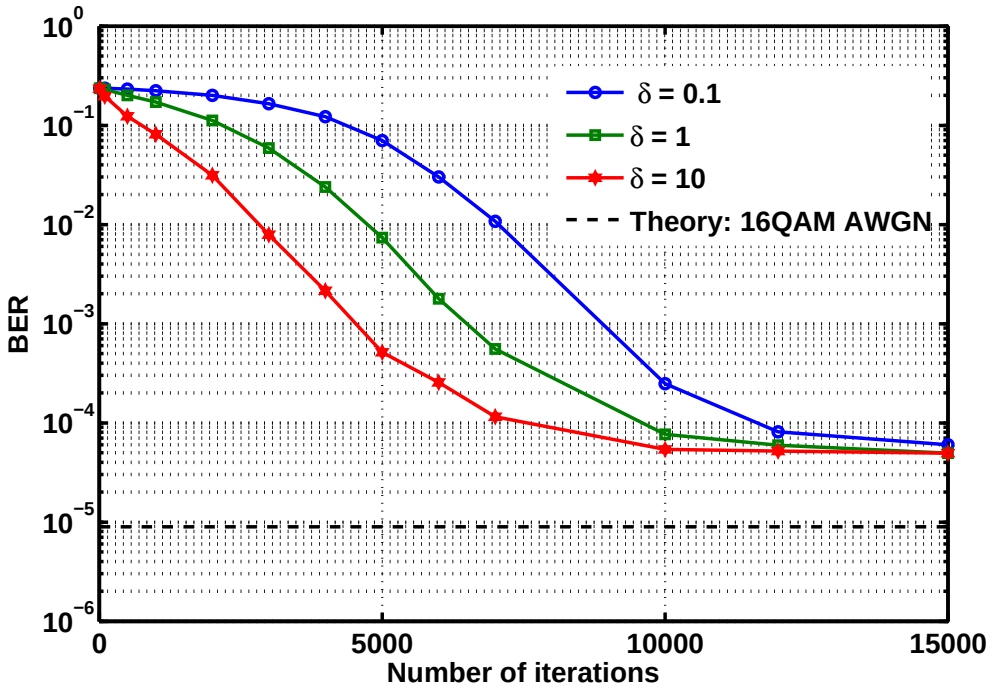


Figure 2.2: Convergence speed of the AN-CMA equalizer for different values of the Hessian matrix initialization δ ($\mu = 10^{-3}$).

The AN-CMA has a faster convergence speed for $\delta = 10$. For instance, a BER of 10^{-3} is now obtained with a sequence of only 5,000 symbols which corresponds to doubling the convergence speed in comparison to the A-CMA. At the steady-state, a BER $\sim 4.10^{-5}$, is obtained. This value should be compared to a BER slightly higher than 10^{-4} for the A-CMA. Once again, the AN-CMA equalizer is insensitive to the impact of phase noise.

2.4.3 Convergence of the A-RDE

As already said during the discussion of the A-RDE convergence, the main drawback of this method is connected to the fact that the right transmitted ring radius has to been accurately selected. Therefore to ensure robustness, a pre-convergence step using the A-CMA/AN-CMA is required. The A-RDE is thus used to enhance the steady-state BER.
To confirm that, for a given number of iterations, we assigned the first half for equalization using the A-CMA and the second half with the A-RDE equalizer. The performance of such a scheme is presented for different step-sizes in Fig. 2.3. Notice that the same step-size is used for both the coarse convergence step (A-CMA) and the fine convergence step phase (A-RDE).

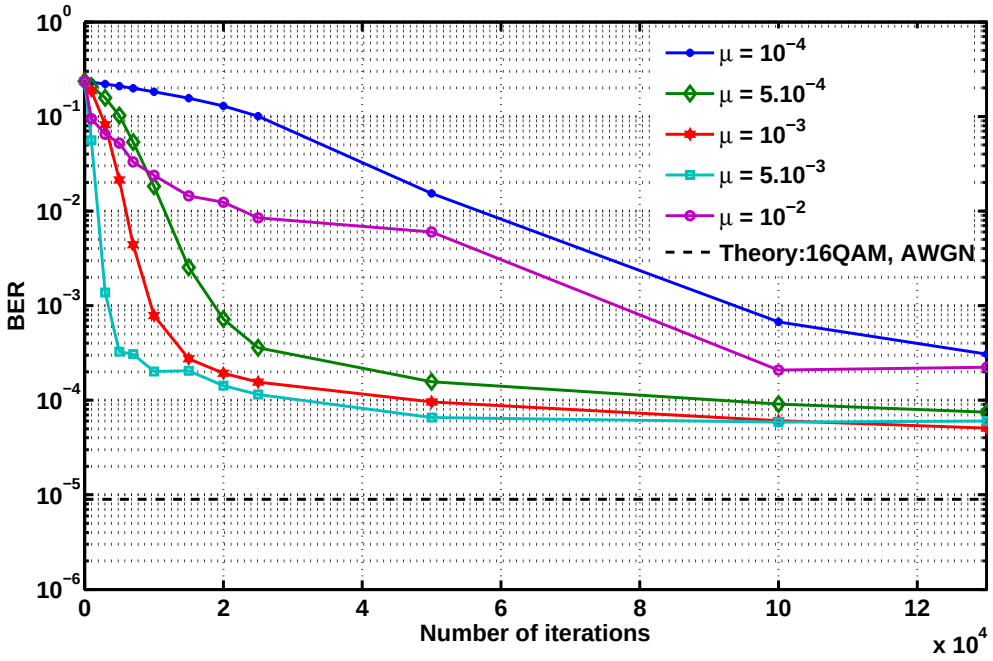


Figure 2.3: Convergence speed of the A-RDE equalizer for different step-sizes μ.

The algorithm tends to converge for nearly all the considered step sizes. This can be explained by the fact is A-RDE is well adapted to QAM modulations. Moreover, faster convergence is possible since once again the A-RDE offers a gain of factor 2 compared to the A-CMA. The steady-state is also improved and is similar to that obtained with the AN-CMA. Finally the value of $\mu = 5.10^{-3}$ represents the best value for the step-size, and therefore will be used to study the performance of AN-RDE.

2.4.4 Convergence of the AN-RDE

This method is a combination of an algorithm adapted to QAM formats (RDE) and a fast converging algorithm (AN approach). The performance are depicted in Fig. 2.4.

It is the best algorithm so far since the convergence is obtained in just 2,000 symbols and a steady-state BER $\sim 5.10^{-5}$ is ensured and is the closest to the theoretical limit.

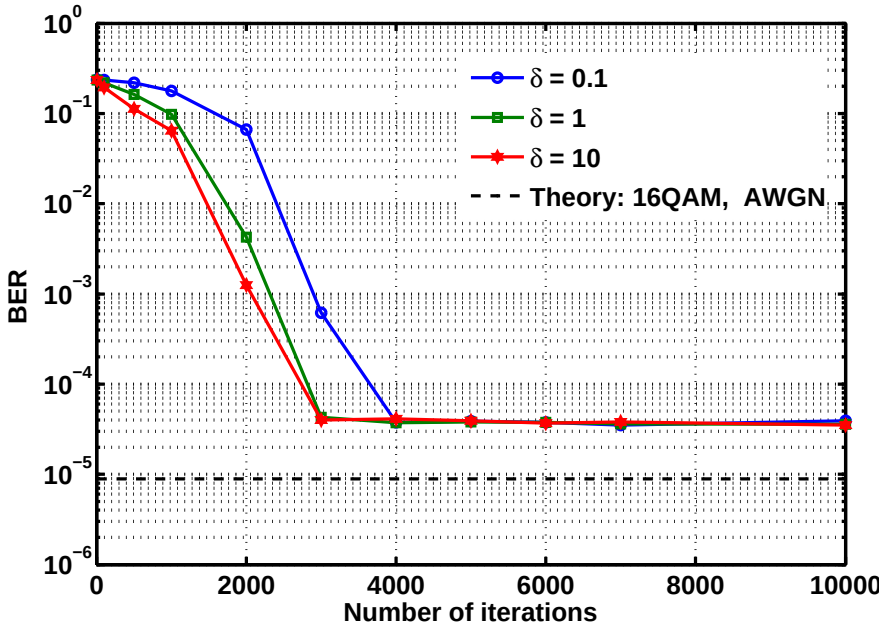


Figure 2.4: Convergence speed of the AN-RDE equalizer for different values of the Hessian matrix initialization δ ($\mu = 5.10^{-3}$).

In Fig. 2.5, we plot the equalizers on both polarizations (left= polarization X, right=polarization Y) after the initialization step ((a)-(b)), after the A-CMA ((c)-(d)), and after the AN-RDE ((e)-(f)) when 10,000 iterations have run.

We observe the equalizers do not exhibit similar taps which is logical since the steady-states are different for each considered algorithm.

In Fig. 2.6, we plot the 16-QAM constellations after A-CMA or AN-RDE followed by constant phase compensation for both polarizations (first row=polarization X, second row=polarization Y).

Once again, we observe that the AN-RDE leads to a better 16-QAM than the A-CMA. Indeed, the 16-QAM for AN-RDE offer narrower disks which leads to a higher robustness to additive noise and thus a gain in OSNR.

In Fig. 2.7, we plot the BER versus the OSNR the number of iterations for the different equalizers to 10.000.

The A-RDE/AN-RDE (especially the A-RDE) show worse performance in comparison to the A-CMA/AN-CMA equalizer at low OSNR. However, the A-RDE and AN-RDE become better at high SNR than the A-CMA and AN-CMA respectively. We also see that the Newton based equalizers have better steady-state in comparison to the gradient based ones using a fixed step-size which justifies the contribution done in this chapter. The AN-CMA and the AN-RDE have very close performance and also close to the theoretical limit.

Nevertheless, all our analysis is based on one CD value (which is one of the worst one) and on time-unvarying filter. Therefore, in the sequel, we study our An based algorithms for these two cases (CD in Section 2.4.6 and time-varying channel in Section 2.4.7)

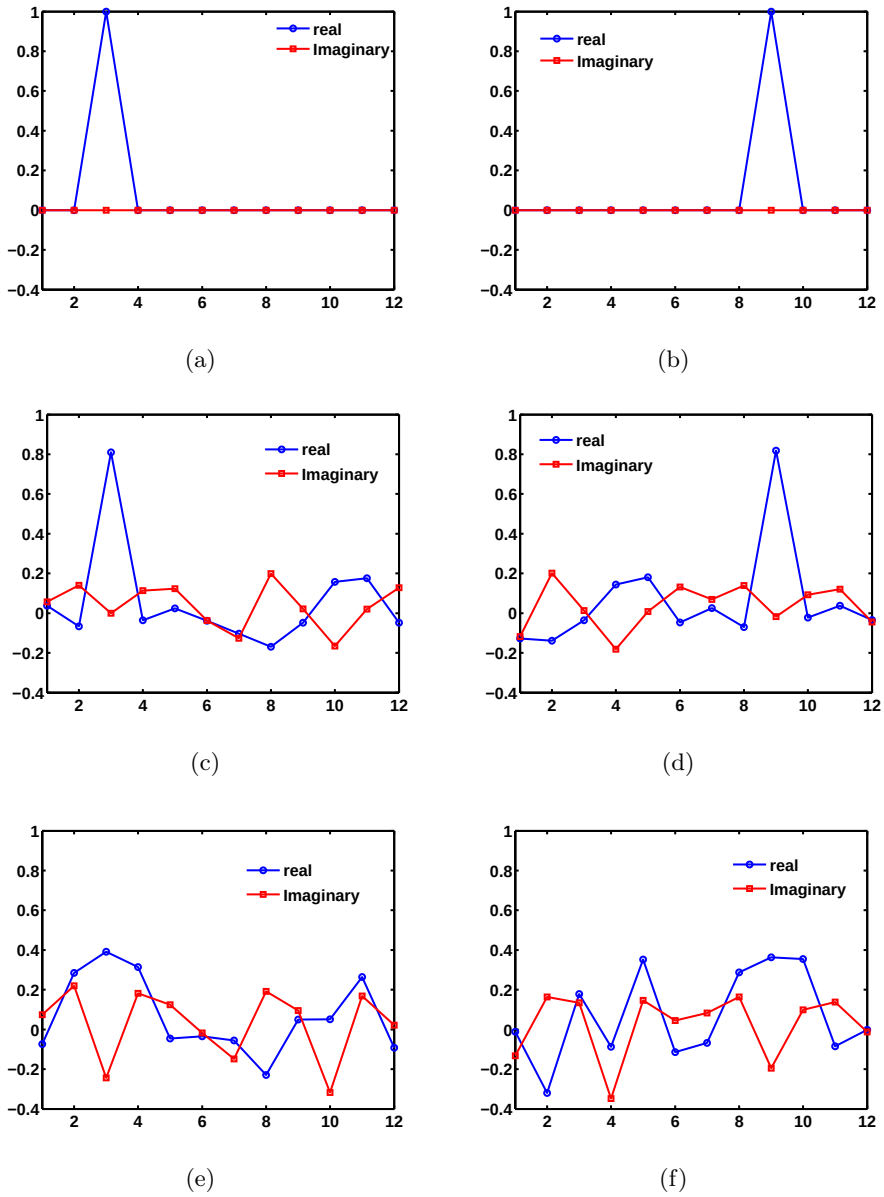


Figure 2.5: Real and imaginary parts of equalizers: (a) at initialization on Pol.X, (b) at initialization on Pol.Y, (c) after A-CMA on Pol.X, (d) after A-CMA on Pol.Y, (e) after AN-RDE on Pol.X, (f) after AN-RDE on Pol.Y. The number of iterations is 10,000.

2.4.5 Impact of the filter length

We study the impact of the filter length on the performances of the different algorithms: A-CMA, AN-CMA, A-RDE, and AN-RDE.

In Fig. 2.8, the convergence of the A-CMA is presented for different equalizer lengths. The best performance are obtained with equalizer lengths of $L = 3$, $L = 4$ and $L = 5$. The equalizer may either not converge or slowly converge when the equalizer filter is too small. Moreover the steady-state BER is higher. Increasing the equalizer length too much may also degrade the performance after convergence.

In Fig. 2.9, the convergence of the A-RDE is presented for different equalizer lengths.

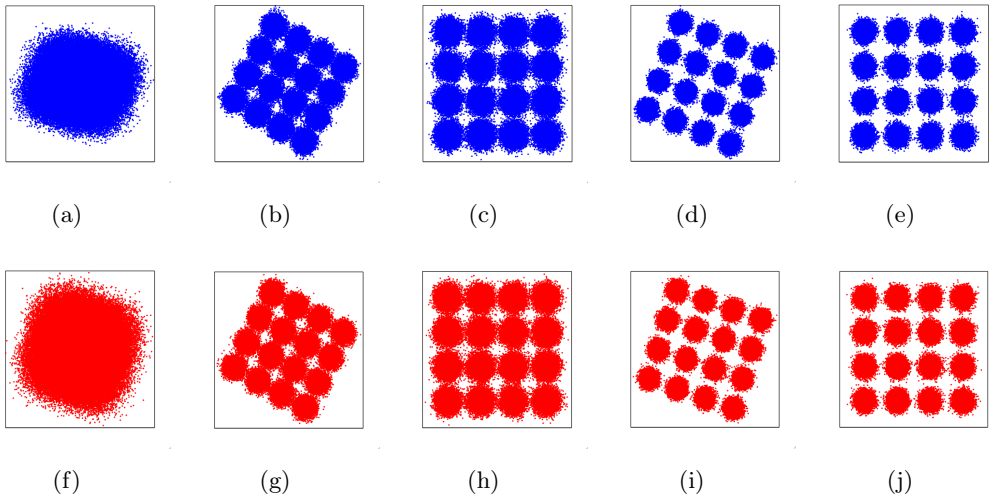


Figure 2.6: 16-QAM constellation: (a) at the receiver, (b) after A-CMA, (c) after A-CMA and constant phase compensation, (d) after AN-RDE, (e) after AN-RDE and constant phase compensation. The number of iterations is 10,000. First (resp. second) row corresponds to the Polarization X (resp. Y).

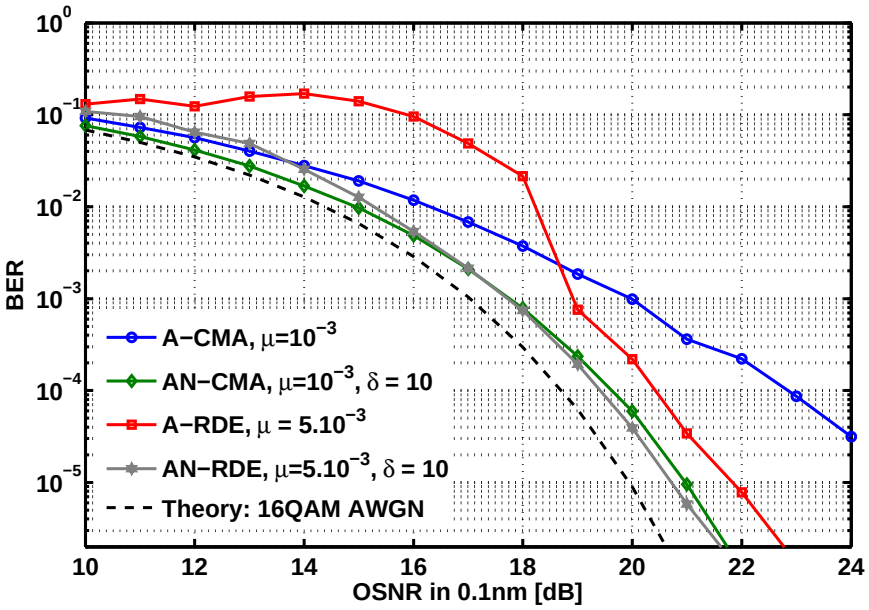


Figure 2.7: BER versus OSNR for different blind equalizers. The number of iterations is 10,000.

The convergence of the A-RDE is more sensitive to the equalizer length than the A-CMA. The speed of convergence is slower when using large filters.

In Figs. 2.10 and 2.11, the convergence of the AN-CMA and AN-RDE are analyzed with respect to different equalizer lengths respectively. The Newton based equalizers are less sensitive to the increase of the equalizer length. For $L > 2$, the equalizers have the same convergence speed and nearly similar steady states as seen in Fig. 2.10 and Fig. 2.11.

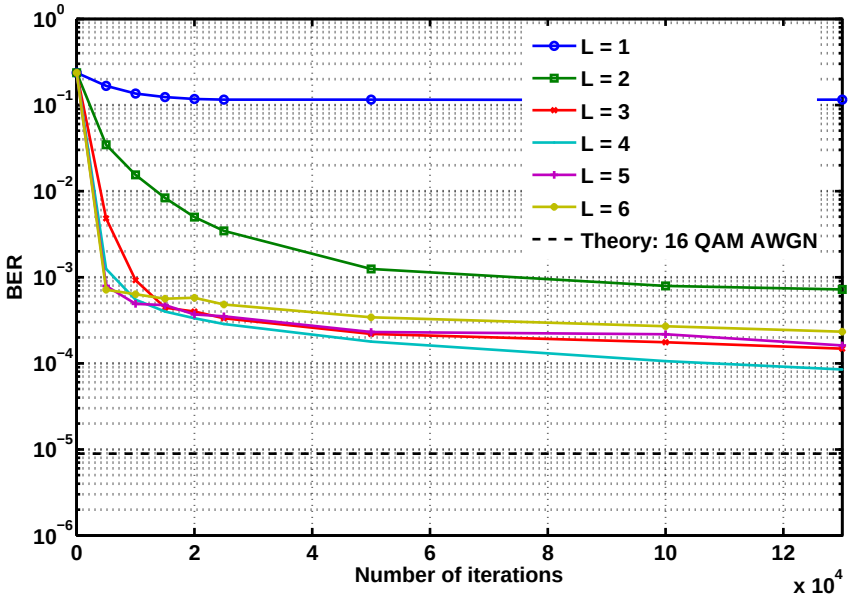


Figure 2.8: Convergence of the A-CMA for different filter lengths ($\mu = 10^{-3}$).

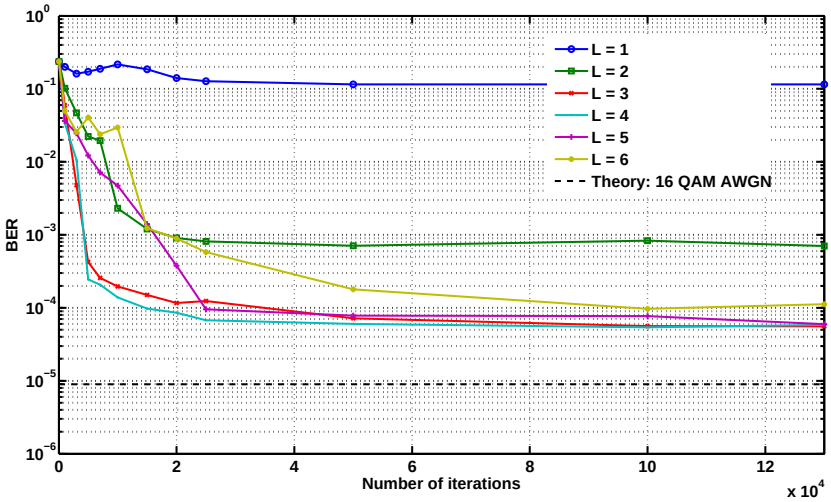


Figure 2.9: Convergence of the A-RDE for different filter lengths ($\mu = 5.10^{-3}$).

2.4.6 Impact of chromatic dispersion

Now we want to study the compensation of CD using the above described blind equalizers. We have considered 10,000 iterations for each equalizer and an equalizer length $L = 3$. In the case of A-RDE/AN-RDE equalization, the first half iterations (*i.e.*, 5000) are implemented for coarse convergence with A-CMA/AN-CMA and the second half for the A-RDE/AN-RDE. The BER versus CD is given in Fig.2.12.

BER is kept below 10^{-3} up to 1000ps/nm for the A-CMA, 1250ps/nm for the A-RDE and 1750ps/nm for both Newton based equalizers AN-CMA and AN-RDE.
For high CD values, the CMA based equalizers are better than the RDE ones. Indeed,

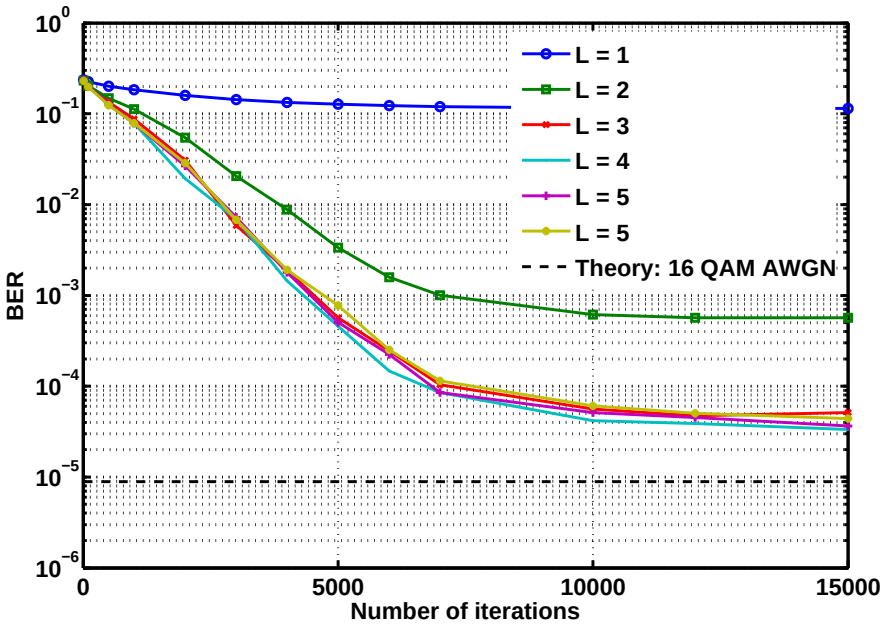


Figure 2.10: Convergence of the AN-CMA for different filter lengths ($\mu = 10^{-3}, \delta = 10$).

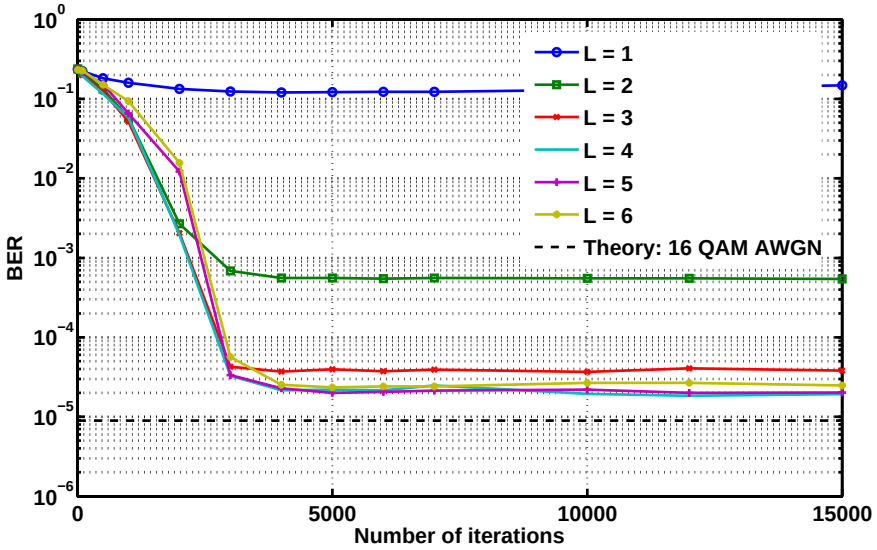


Figure 2.11: Convergence of the AN-RDE for different filter lengths ($\mu = 5.10^{-3}, \delta = 10$).

this can be explained by the fact that the first 5000 iterations of the CMA are not enough to ensure the coarse convergence of the algorithm before moving to the RDE. The used RDE takes decisions based on the wrong signal radii and therefore it induces errors to the estimation and degrades the performance. Once again, we observe that the AN approach proposed here provides relevant solution to the presence of high CD.

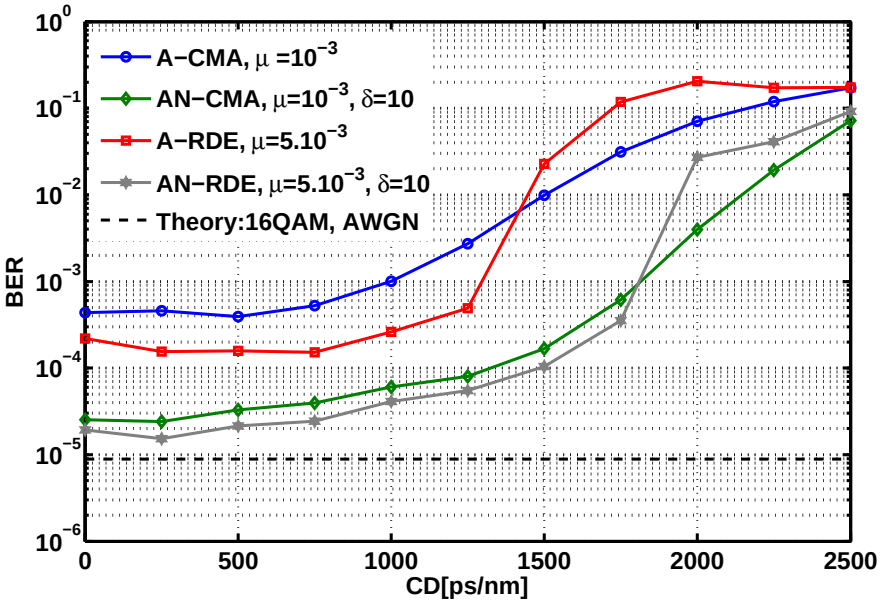


Figure 2.12: BER versus CD for different blind equalizers. The number of iterations is 10,000 symbols.

2.4.7 Impact of time-varying channels

The study of the tracking ability of the proposed equalizers is also important. Both RDE and CMA should be able to track relatively fast varying channel variations.

For the sake of simplicity, we only consider infinite polarization rotation modeled by the Jones matrix. Consequently, the residual CD is assumed to be null, and the PMD only gives rise to one time-varying rotation. The polarization mixing is thus instantaneous and does not lead to inter-symbol interference but just to inter-polarization interference. The channel impulse response at time t_0, denoted by $t \mapsto \mathbf{C}_{a,t_0}(t)$, can be written as [40]

$$\mathbf{C}_{a,t_0}(t) = \begin{bmatrix} \cos(\Omega t_0) & \sin(\Omega t_0) \\ -\sin(\Omega t_0) & \cos(\Omega t_0) \end{bmatrix} \delta(t) \tag{2.15}$$

where Ω is the rotation speed in rad/s.

The BER versus the rotation speed for an OSNR=20dB in 0.1nm is depicted in Fig. 4.14.

Firstly, the fixed step-size based adaptive equalizers show better tracking capacity than the AN approach. This is not a surprise since the Hessian matrix update uses the system memory through the forgetting factor λ. The Newton based equalization is thus well adapted for slowly time-varying channels. Nevertheless all the algorithms show a tracking capacity for a speed of channel variation of $\sim$ 0.5Mrad/s and except the AN-CMA, all the other algorithms can track variations larger than 1Mrad/s.
Those algorithms can thus be robust enough to track the maximum variation of the state of polarization of about some tens of radians in already installed systems [26].

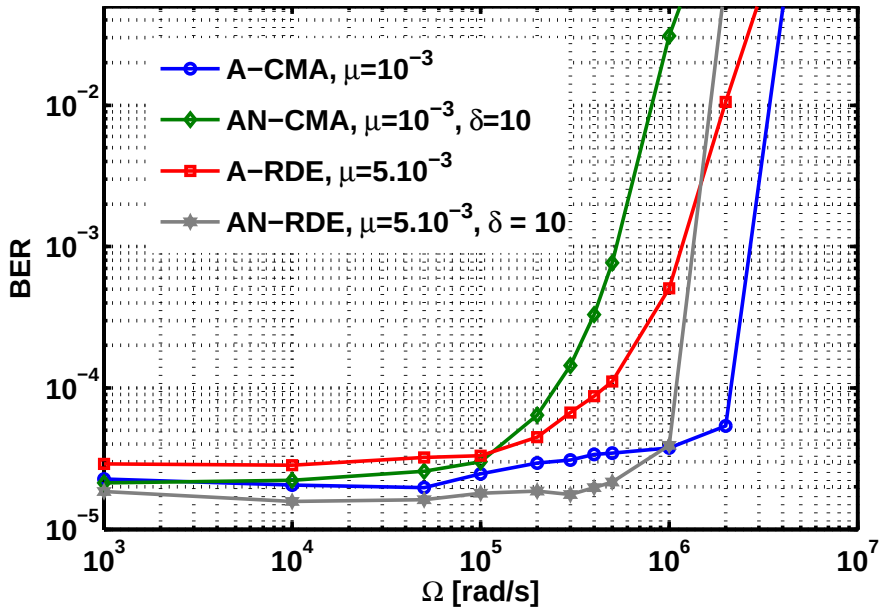


Figure 2.13: Tracking capacity for the different equalizers

2.5 Computational Load

The complexity of the equalization algorithms is an important issue for the implementation of real time coherent receivers. Due to an important number of iterations to reach convergence and the necessity to adapt often the filter taps, it is essential to use a small number of filter taps for the MIMO equalizer and to increase the convergence speed.

The implementation of a complex multiplication (flop) in digital circuits of two complex scalars w_p and y_p implies four real multiplications and two additions as depicted in Fig. 2.14.

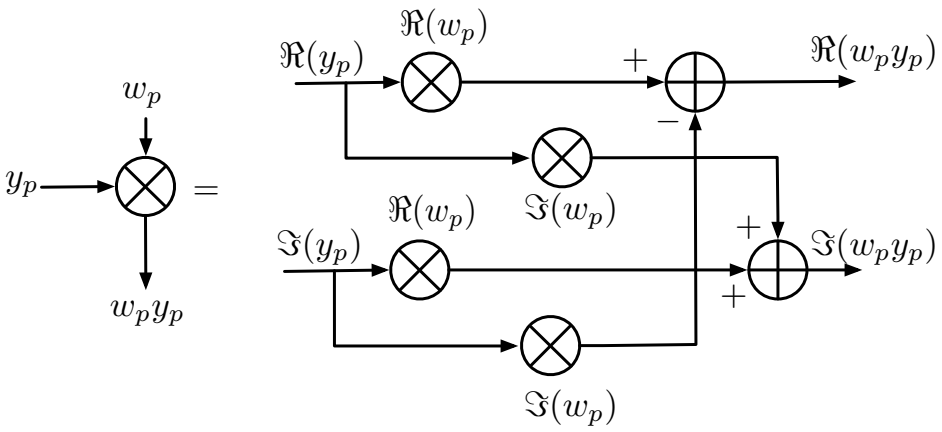


Figure 2.14: Complex-valued multiplication.

Therefore, it is costly in terms of calculation in comparison to real multiplications,

divisions, additions and subtractions. We thus neglected these last operations in the calculation of the computational load. The computational load of the fixed step-size based approach is $\mathcal{O}(4L)$ corresponding to vector multiplication while that of the Newton base one is $\mathcal{O}((4L)^2)$ corresponding to the calculation of the Hessian matrix. Now, we compare the computational complexity of the different algorithms described above. This complexity is given as the number of flops required to attain a target BER of 10^{-3}. We neglect the control operations determining the best radius to use so, the RDE and CMA has the same number of flops per iteration which corresponds to one filter update. However in practice the RDE is slightly more complex than the CMA.
Based on Fig. 2.15 that summarizes the performances of the different equalizers in terms of convergence, the number of iterations required for BER$\sim 10^{-3}$ is extracted and the total flops used to optimize the filter coefficients are summarized in Table. 2.2.

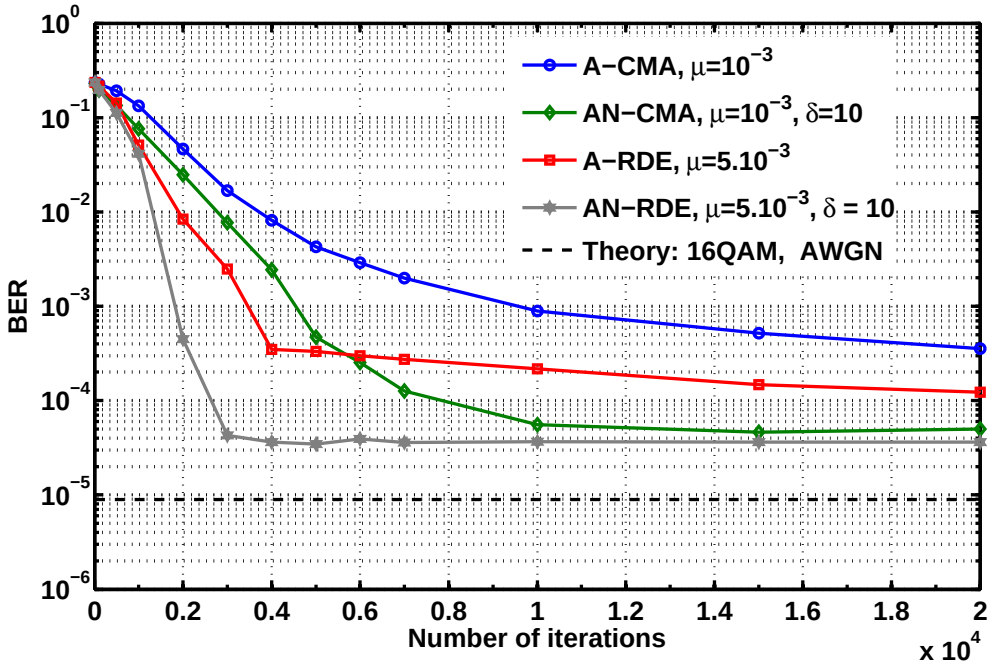


Figure 2.15: Speed of convergence of the different equalizers for their optimized parameters

Algorithm	A-CMA	AN-CMA	A-RDE	AN-RDE
Update equation per it. and pol.	$2(4L+1)$	$2(4L+1)$	$2(4L+1)$	$2(4L+1)$
Hessian eval. per it. and pol.	-	$4(4L)^2+4L+2$	-	$4(4L)^2+4L+2$
# it.	10000	5000	4000	2000
Total Flops ($\times 10^3$)	**520**	**3085**	**260**	**1232**

Table 2.2: Complexity for different equalizers.

As shown in previous Sections, the Newton based adaptive equalization doubles the speed of convergence of both the CMA and RDE equalizers. However, the computational complexity increases by a factor of $\sim$5.

2.6 Conclusion

In this chapter, we studied the adaptive equalizer using a fixed and variable step-sizes. The performance of the RDE was also addressed. The RDE algorithm, when used after a pre-convergence based on the classical CMA algorithm, improves the algorithm in terms of steady-state. Newton based adaptive equalization method relies on the calculation of the inverse of the Hessian matrix on a recursive ways, allows to have faster convergence speed and better BER at the steady-state. Convergence speed is actually improved by a factor of $\sim$2.
In the last part of this chapter, we addressed the computational load in terms of the number of flops (complex multiplication), the Newton based adaptive equalizer requires the calculation of a square matrix of size the FIR filter corresponding to one polarization, increases the complexity of the algorithm by a factor of $\sim$5.

Chapter 3

Block-wise Carrier Frequency Offset Estimation

3.1 Introduction

M-ary quadrature amplitude modulation (M-QAM) formats combined with coherent detection and digital signal processing (DSP) are promising candidates for the implementation of the next generation optical transmission systems. However, those modulation formats are more sensitive to phase errors than M-PSK formats.
These phase errors may correspond to constant phase offset, carrier frequency offset (CFO) and laser phase noise [53]. Several carrier frequency estimators have been already presented for QPSK based optical transmissions. These algorithms rely either on the phase difference between two adjacent received samples [45,54] or the maximization of the FFT of the fourth-power received samples. The performance of these algorithms are quite sufficient for QPSK based transmission. In contrast, when QAM modulation is used, the phase difference based algorithm fails and the FFT based algorithm does not provide accurate enough estimate. Indeed, QAM modulation is more sensitive to CFO than QPSK and thus requires very accurate estimate of the CFO.

In this chapter, we propose a new (block-wise) blind estimate for the CFO well suitable for QAM modulation.

This chapter is organized as follows: Section 3.2 is devoted to signal model and also performance metric definition. In Section 3.3, we remind the two CFO estimate commonly used in the "optical communications literature". In Section 3.4, we propose our new estimate based on both polarizations and the real spectrum maximization. We especially show, in Section 3.5 that the obtained performance are well adapted for real systems requirement using QAM modulation even when high CD, PMD or phase noise occur. This algorithm can be adapted to PSK formats as well.

3.2 Signal model

The Carrier Frequency offset (CFO) is due to the frequency mismatch between the LO and the signal lasers. Thanks to the equalization procedure described in the previous chapter, we can now assume that the inter-symbol interference raised by the residual CD and the PMD. Therefore the (baud-rate) output of the equalizer on polarization p,

already denoted by $z_p(n)$, can be written as follows

$$z_p(n) = s_p(n)e^{2i\pi(\varphi_{0,p}+n\varphi_1)} + b'_p(n) \tag{3.1}$$

where it remains two drawbacks:

- $\varphi_1 = \delta f_a T_s$ is the discrete-time (baud-rate) FO. The FO is independent of the polarization state of the received POLMUX signals.
- $\varphi_{0,p}$ corresponds to the constant phase. This constant phase occurs since the blind equalizer is only able to determine the filter up to a constant phase.

and where $b'_p(n)$ is the additive zero-mean circularly-symmetric complex-valued Gaussian noise.

The goal of this chapter is to exhibit accurate estimate of the "discrete" CFO φ_1.

To evaluate the accuracy of the estimate, the performance metric used will be the mean square error (MSE) defined as

$$\text{MSE} = \mathbb{E}[|\varphi_1 - \hat{\varphi}_{1,N}|^2] \tag{3.2}$$

where $\hat{\varphi}_{1,N}$ is the considered estimate when N received samples $z_p(n)$ are available on each polarization.

We assume the CFO φ_1 is constant within the observation window of length N.

3.3 State-of-the-Art

3.3.1 Differential phase based method

The widely used Carrier Frequency Offset estimate for optical communications was presented in [45] and is suitable for M-PSK modulation. It is defined as

$$\hat{\varphi}_{1,N} = \frac{1}{2\pi M}\angle\sum_{k=0}^{N-1}[z_p(k+1)\overline{z_p(k)}]^M \tag{3.3}$$

To understand the underlying idea associated with this estimate, we have to work on

$$Z(k) = [z_p(k+1)\overline{z_p(k)}]^M$$

After simple derivations using Eq. (3.1), we obtain that

$$Z(k) = [s_p(k+1)\overline{s_p(k)}]^M e^{2i\pi M\varphi_1} + e(n) \tag{3.4}$$

where $e(n)$ is an additive zero-mean noise depending on the symbols and on the noise. Notice that $e(n)$ vanishes in noiseless case.

When M-PSK, is used at the transmitter side, the modulated term in Eq (3.4) is eliminated and becomes constant equal to R^{2M} where R is the radius of the considered PSK. Finally, we have

$$Z(k) = R^{2M}e^{2i\pi M\varphi_1} + e(n). \tag{3.5}$$

Then standard ML phase estimator (assuming a Gaussian independent distribution for $e(n)$ even if not) is applied and leads to Eq. (3.3). Therefore this estimate offers excellent performance in the context of PSK. Modification can be done to adapt this algorithm

to QAM by considering $M = 4$ whatever the QAM size. But, the performance will be dramatically degraded. Indeed, due to the non-constant property of the QAM, the term $[s_p(k+1)\overline{s_p(k)}]^M$ can not be removed and so prevents a good estimate of the CFO φ_1.

Let us come back to the PSK case. The estimate of the CFO φ_1 can be viewed as a phase estimate based on the modified received signal $Z(k)$. It is well-known that the the MSE of a phase estimate decreases inversely proportional to the number of observation samples. Therefore the MSE of this CFO estimate decreases as $1/N$ which is not optimal at all. Same reasoning can be applied for the QAM extension with, in extra, a self noise phenomenon. As M-QAM is more sensitive to FO, designing more accurate estimators is still required.

In order to evaluate the performance of the differential phase based CFO, we present first, in Fig. 3.1, the MSE versus OSNR for the 14Gbaud PolMux QPSK-case with different sizes of the observation window. We generate an CFO randomly located between 0 and 3.5GHz. Note that due to the elevation to M^{th}-power, the range of this kind of estimators is $[-R_s/M, +R_s/M]$ where R_s is the baud rate of the transmission. The

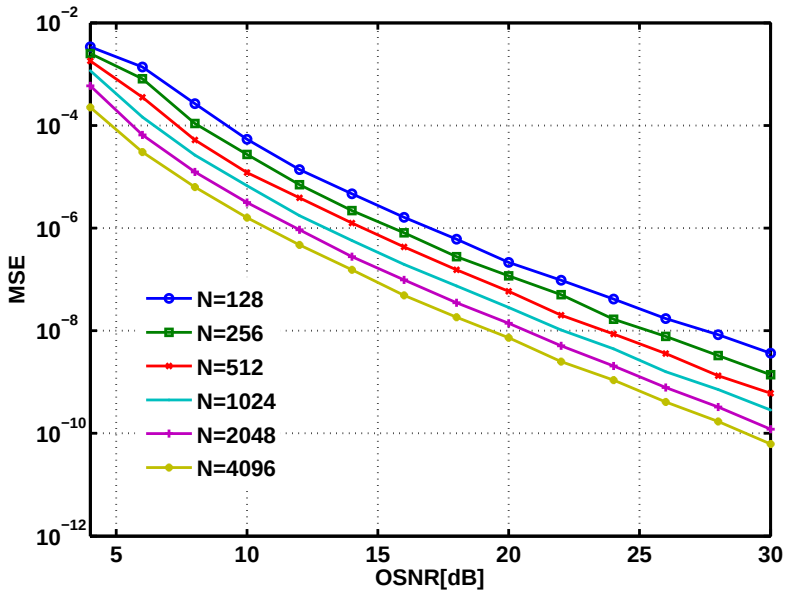


Figure 3.1: MSE versus OSNR using the differential phase based method for QPSK

MSE decays as the inverse of the OSNR and is below $\sim 10^{-6}$ as soon as OSNR is larger than 10dB and for an observation window of 4096. Notice also that the larger the observation window used is, the more accurate this kind of estimators is. The residual CFO can then be considered as a phase noise and can be managed using the Viterbi-Viterbi algorithm for the constant phase estimate reminded in Chapter 1 [47].

In the QAM case, we have to modify the estimate introduced in Eq. (3.3) accordingly. The MSE of the differential phase based method versus OSNR for different observation window duration for 16-QAM is depicted in Fig. 3.2. A rough estimate of the CFO is obtained since only an MSE of $\sim 10^{-5}$ can be obtained and a error floor occurs due to the self noise provoked by the modulated term in Eq. (3.4). Moreover the MSE equal

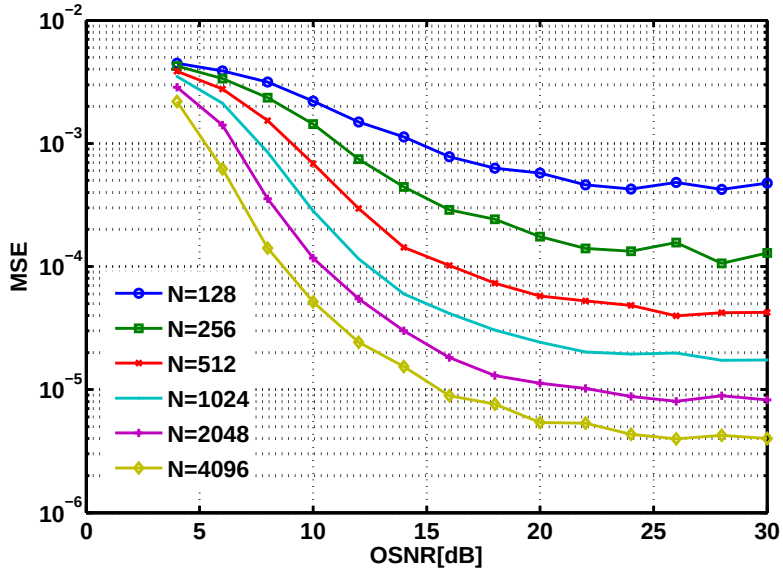


Figure 3.2: MSE versus OSNR using the differential phase based method for 16-QAM

to $\sim 4.10^{-6}$ is only reached for high OSNR, typically larger than 20dB.

In Fig. 3.3, we depict the MSE versus the size of the observation window N. We confirm that the MSE decreases as $1/N$ for both QPSK and QAM formats. The MSE is around $\sim 10^{-8}$ for QPSK and $\sim 10^{-5}$ for QAM when the observation window is more larger than $1,000$ symbols.

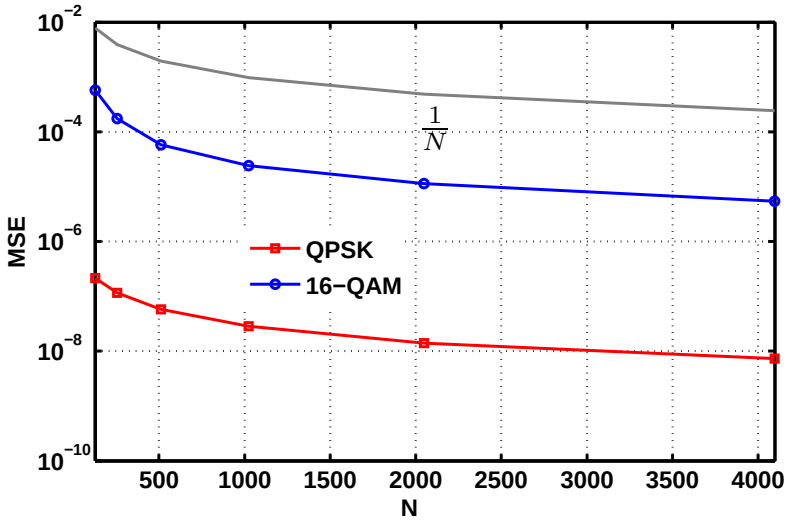


Figure 3.3: MSE versus N at OSNR = 20dB for QPSK and 16-QAM

Finally as QAM formats are more sensitive to the CFO, our numerical illustrations confirm that accurate CFO estimators have still to be developed.

3.3.2 FFT maximization based method

An other way to estimate the CFO obviously relies on the spectrum of the received signal in which a peak at the CFO value should be there. Let us introduce such a type of algorithm for M-PSK since the literature in "optical communications" only evokes spectrum maximization based method under PSK assumption [25, 55].

Once again, we elevate at the power M a function of the received signal. Let $\tilde{Z}(n) = z(n)^M$ for a M-PSK. Then

$$\tilde{Z}(n) = s_p(n)^M e^{2i\pi M(\varphi_{0,p}+\varphi_1 n)} + e'(n)$$

where $e'(n)$ is an additive zero-mean noise depending on the transmit symbols and the noise. When M-PSK is used, the term $s_p(n)^M$ becomes constant. Therefore, we have

$$\tilde{Z}(n) = R^M e^{2i\pi M(\varphi_{0,p}+\varphi_1 n)} + e'(n)$$

So we just would like to estimate a pure harmonic in additive noise. Without considering the dependence of the noise with the data, we shall do that by maximizing the spectrum. In order to reduce the complexity, only the N- FFT frequencies are considered. Thus

$$\hat{\varphi}_{1,N} = \frac{1}{N} \arg \max_{k\in\{0,\cdots,(N-1)/M\}} \left| \frac{1}{N} \sum_{n=0}^{N-1} z_p(n)^M e^{-2i\pi Mkn/N} \right|^2$$

The MSE of such an algorithm decreases as $1/N^2$ which is better than that of Section 3.3 but still not optimal [56].

In this Chapter, we will extend and improve such a approach by spectrum maximization: indeed, the tested frequencies will be modified by keeping the complexity reasonable. Then an two polarization based version will be considered. Finally QAM modulations (instead of PSK ones) will be taken into account.

3.4 Proposed Carrier Frequency Offset estimate

The construction of relevant blockwise blind estimators for the frequency offset in the context of either PSK or QAM modulations can be done by using the unique framework of the non-circularity [48, 49]. Indeed, due to rotation symmetry, it is well-known that for M-PSK, the term $\mathbb{E}[s_p(n)^Q] \neq 0$ with $Q = M$. For M-QAM, we have $\mathbb{E}[s_p(n)^Q] \neq 0$ with $Q = 4$. Then one can write $z_p(n)^Q$ as $\mathbb{E}[z_p(n)^Q] + e_p(n)$ where $e_p(n)$ is a zero-mean process that can be viewed as disturbing noise. Moreover as the noise $b'_p(n)$ is a circularly-symmetric Gaussian noise, we have that

$$\mathbb{E}[z_p(n)^Q] = \mathbb{E}[s_p(n)^Q] e^{2i\pi Q(\varphi_{0,p}+\varphi_1 n)}.$$

Consequently, we get

$$z_p^Q(n) = A_p e^{2i\pi Q(\varphi_{0,p}+n\varphi_1)} + e_p(n) \tag{3.6}$$

where $A_p = \mathbb{E}[s_p^Q(k)] \neq 0$ is a constant amplitude. The most important thing now is to remark that $z_p^Q(n)$ is actually a constant-amplitude complex exponential with frequency $Q\varphi_1$ disturbed by a zero-mean additive noise. One can thus deduce the

following frequency offset estimator based on the maximization of the sums of the periodograms associated with $z_p^Q(n)$ and $z_q^Q(n)$.

$$\hat{\varphi}_{1,N} = \frac{1}{Q} \arg \max_{\varphi \in [-\frac{1}{2}, \frac{1}{2})} (f_1(\varphi) + f_2(\varphi)) \tag{3.7}$$

where

$$f_p(\varphi) = \left| \frac{1}{N} \sum_{n=0}^{N-1} z_p(n)^Q e^{-2i\pi\varphi n} \right|^2 \tag{3.8}$$

with N the number of available samples.

When PSK is encountered, our algorithm is a natural extension of the so-called Viterbi-Viterbi algorithm [46, 57] by combining linearly the periodogram obtained on each polarization. When QAM is encountered, our algorithm is also a natural extension of an existing algorithm [48]. Notice that even if the same framework enables us to treat PSK and QAM together, the performance of these algorithms are constellation-dependent. Actually, PSK works better since $\mathbb{E}[s_p(n)^Q] = s_p^Q(n)$ whereas, for QAM, $\mathbb{E}[s_p(n)^Q] \neq s_p^Q(n)$ which leads to self-noise [56, 58].

The main issue now concerns the evaluation of the maximum in Eq. (3.7). Actually, in the "optical communications" literature, the maximization is done through the computation of a discrete-frequency spectrum (FFT). This FFT either has N points or has been zero-padded with αN points ($\alpha > 1$ is fixed once). Thanks to [56], the Mean Square Error (MSE) on the frequency offset decreases as $1/N^2$ for such algorithms implementation. As M-QAM is more sensitive to frequency offset, such MSE decreasing trend is not enough and more accurate estimator is required. Therefore we here propose to maximize the periodogram in different way. We compute the maximization of periodogram into two steps as follows

1. a coarse step which detects the maximum magnitude peak which should be located around the true frequency offset. This is carried out via a Fast Fourier Transform (FFT) of size N.

2. a fine step which inspects the cost function around the peak detected by the coarse step. This step may be implemented by a gradient-descent algorithm or the Newton algorithm [58].

Since [58], we know that the MSE associated with the algorithm carrying out the two steps decreases as $1/N^3$ and thus is significantly more accurate than the FFT based maximization.

In the second step, a Newton based gradient-descent algorithm is used, and the update equation is as follows

$$\hat{\varphi}_{1,N}^{\ell+1} = \hat{\varphi}_{1,N}^{\ell} + \mu \frac{f_1'(\hat{\varphi}_{1,N}^{\ell}) + f_2'(\hat{\varphi}_{1,N}^{\ell})}{|f_1''(\hat{\varphi}_{1,N}^{\ell}) + f_2''(\hat{\varphi}_{1,N}^{\ell})|}$$

where

$$f_p'(\varphi) = \frac{\partial f_p(\varphi)}{\partial \varphi} = 2\Re(g_p \overline{g_p'})$$

and

$$f_p''(\varphi) = \frac{\partial^2 f_p(\varphi)}{\partial \varphi^2} = 2\Re(g_p \overline{g_p''}) + |g_p'|^2$$

with

$$g_p = \sum_{n=0}^{N-1} z_p(n)^Q e^{-2i\pi\varphi n},$$

$$g'_p = \frac{\partial g_p}{\partial \varphi} = -2i\pi \sum_{n=0}^{N-1} n z_p(n)^Q e^{-2i\pi\varphi n},$$

and

$$g''_p = \frac{\partial^2 g_p}{\partial \varphi^2} = -4\pi^2 \sum_{n=0}^{N-1} n^2 z_p(n)^Q e^{-2i\pi\varphi n}.$$

A stop condition for the iteration can be given by a maximum number of iteration (≤ 50 per example) or satisfying the condition given by Eq. (3.9)

$$\|\hat{\varphi}_{1,N}^{\ell+1} - \hat{\varphi}_{1,N}^{\ell+1}\| \leq \varepsilon \tag{3.9}$$

with ε a small positive number.

As a conclusion, while the two-steps based maximization has been already used in "wireless communications" literature, the proposition of combining two periodograms of both ways (*i.e.* two polarizations in optical context or two antennas in wireless context) is new.

Fig. 3.4 and Fig. 3.5 illustrate our two steps procedure for the proposed frequency estimate. Indeed, in Fig. 3.4, we plot the FFT of the fourth power of the received signal since 16-QAM is employed. The sought CFO is equal to 1GHz, which means that a peak around 4GHz has to occur on the spectrum. We observe several peaks around

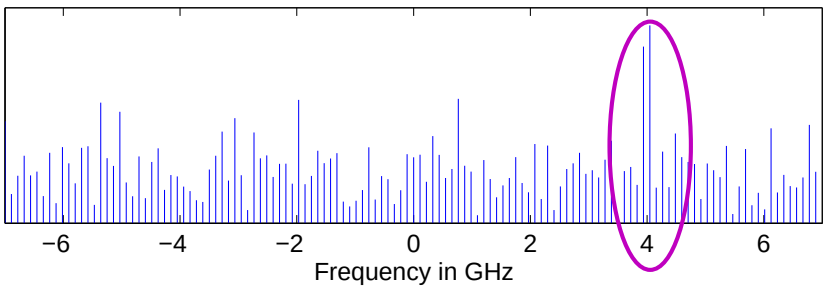


Figure 3.4: FFT of the fourth-power of the received signal for 16-QAM. Sought CFO is 1GHz.

4GHz but not at 4GHz since this frequency in our example does not correspond to a FFT frequency. Therefore, we would like to refine the estimate to catch the true peak at 4GHz. In Fig. 3.5, a zoom around the true point is performed and not only FFT frequencies are considered. We observe that the extra-step enables us to improve the estimation accuracy since we are now able to be closer to the sought CFO.

We remind that the range of our CFO estimate is $[-R_s/Q, R_s/Q]$ where Q is the considered power ($Q = 4$ for QAM). In the case of 14Gbaud POLMUX 16-QAM transmission, the range is thus equal to [-3.5, 3.5]GHz.

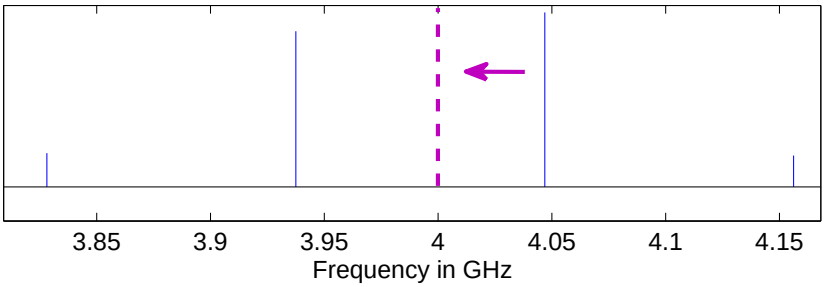


Figure 3.5: Zoom on the spectrum of the fourth power of the received signal for 16-QAM. Sought frequency is 1GHz.

3.5 Estimation performance

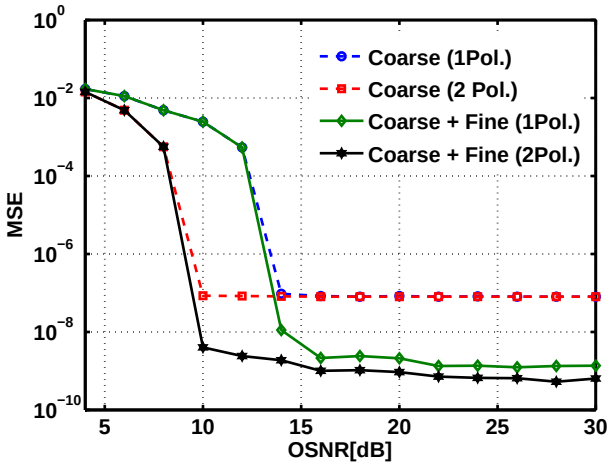
In order to evaluate the performance of the proposed block-wise carrier frequency offset estimator, we used the above-described model to generate the PolMux 16-QAM signal. Unless otherwise stated, we simulate a channel without CD and PMD (as explained in Section 3.3, we have assumed a perfect channel compensation). We then added CFO randomly chosen between 0 and 3.5GHz. The CFO is estimated using one of the 4 following methods:

- Coarse step based on one polarization (which corresponds to an already-existing algorithm described in Section 3.3.2),
- Coarse step based on both polarizations,
- Coarse and fine steps based on one polarization,
- Coarse and fine steps based on both polarizations.

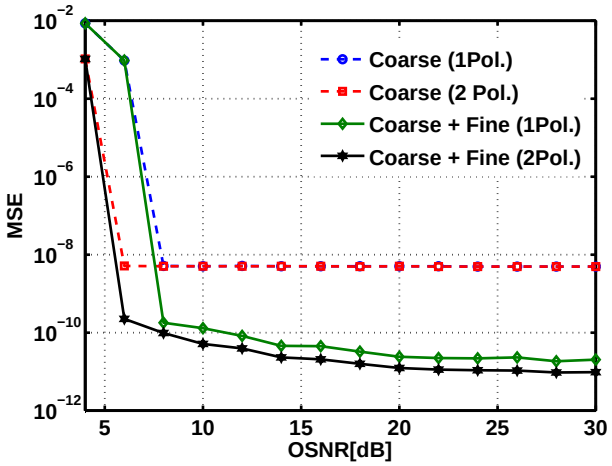
In Fig. 3.6, we plot the MSE versus OSNR when the FFT size (equivalently the observation window) is $N = 256, 1024$ and 4096. The most important gain in performance is due to the use of the fine step. For instance, a MSE below 10^{-11} (corresponding to some hundreds of kHz of residual FO) can be reached by using both polarizations and both steps as soon as $N = 1024$ and OSNR = 20dB. The outliers effect [56] observed at low OSNR is reduced thanks to the use of both polarizations. When $N = 1024$, at OSNR = 20dB (usual OSNR for 16-QAM transmission), the MSE is equal to 10^{-11} with our complete estimate, 10^{-8} with the FFT based estimate, and 10^{-4} with the differential phase based estimate (cf. Fig. 3.2). Notice that this corresponds respectively to a remaining CFO of 40kHz, 1MHz, and 100MHz. Consequently, our estimate strongly improves the CFO accuracy.

In Fig. 3.7, we plot the MSE versus N when OSNR =20dB. One can easily check that the MSE decays as $1/N^2$ for the methods based on the coarse step and $1/N^3$ for those based on the fine steps.

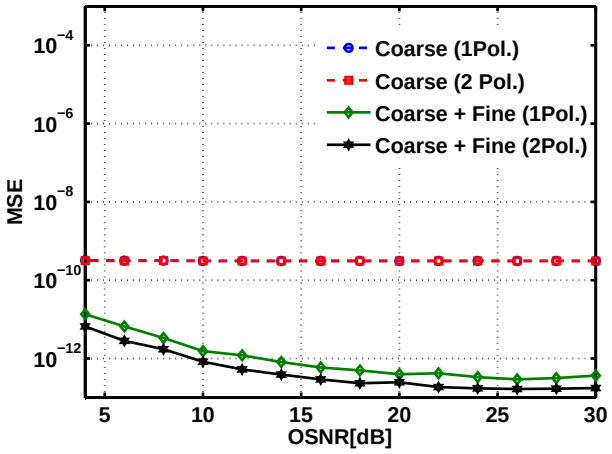
In Fig. 3.8, we plot the value of the residual CFO versus the true value of the sought CFO δf_a. The extrema for the x-label are chosen such that $4\varphi_1 = 4\delta f_a T_s$ correspond to two adjacent FFT frequencies k_0/N and $(k_0+1)/N$ with k_0 an arbitrary integer. In our example, since $N = 1024$, the gap on δf_a between two FFT frequencies is equal to 3.439MHz.



(a)



(b)



(c)

Figure 3.6: MSE versus OSNR (a) $N = 256$ (b) $N = 1024$ (c) $N = 4096$.

Thanks to the fine step, the residual CFO is insensitive to the location of the true CFO, and it is kept below 200kHz. In contrast, using the methods only based on the

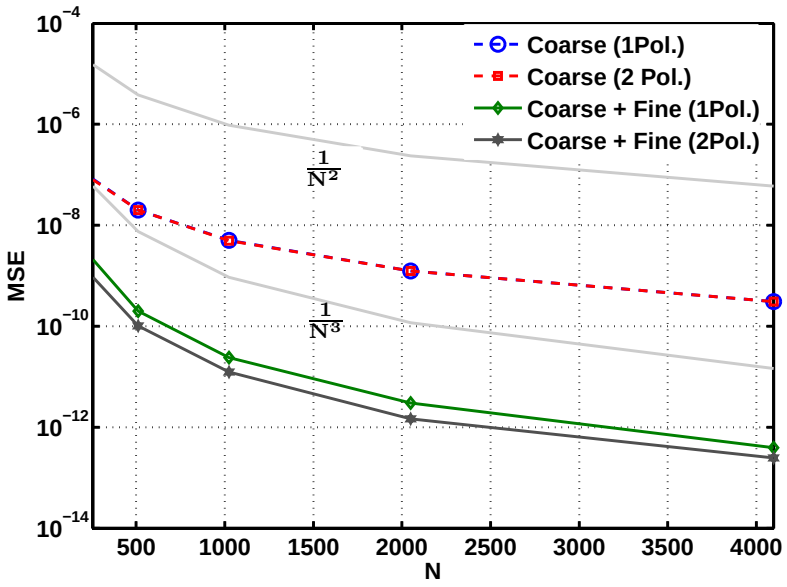


Figure 3.7: MSE versus N (OSNR = 20dB).

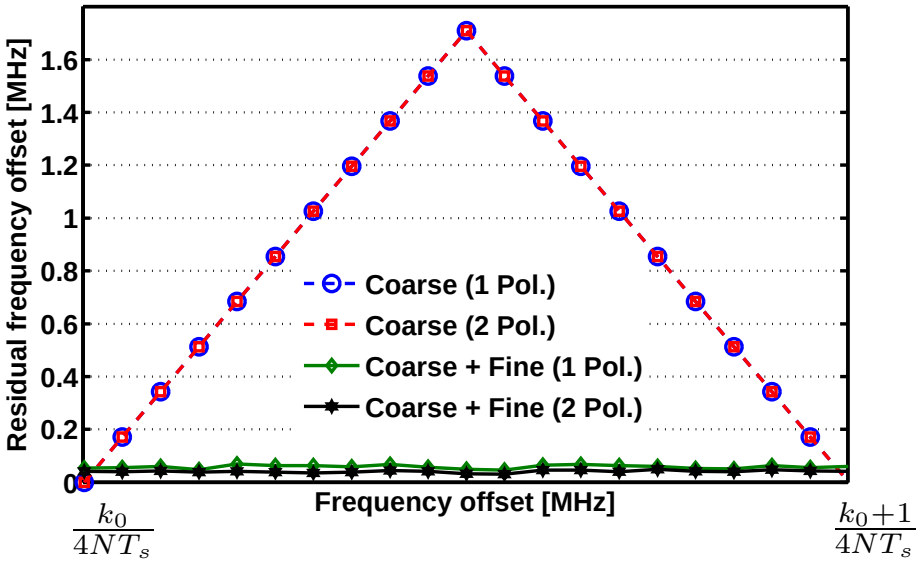


Figure 3.8: Residual CFO versus the initial CFO for different methods ($N = 1024$, OSNR = 20dB). For example,

coarse step often leads to larger residual CFO exceeding 1MHz.

3.5.1 Tolerance to CD

We would like know analyze the robustness of our estimate to the presence of residual CD whereas we have so far assumed that the CD is perfectly compensated for. The estimator described in previous section remains the same even if the signal model given

in Eq. (3.1) is slightly modified due to the presence of the residual CD.

We considered here OSNR = 20dB and $N = 1024$ and a 14Gbaud PolMux 16-QAM transmission . In Fig. 3.9, we plot the residual CFO versus the true CFO for different values of the residual CD when the proposed CFO estimate with both polarizations and fine step is carried out.

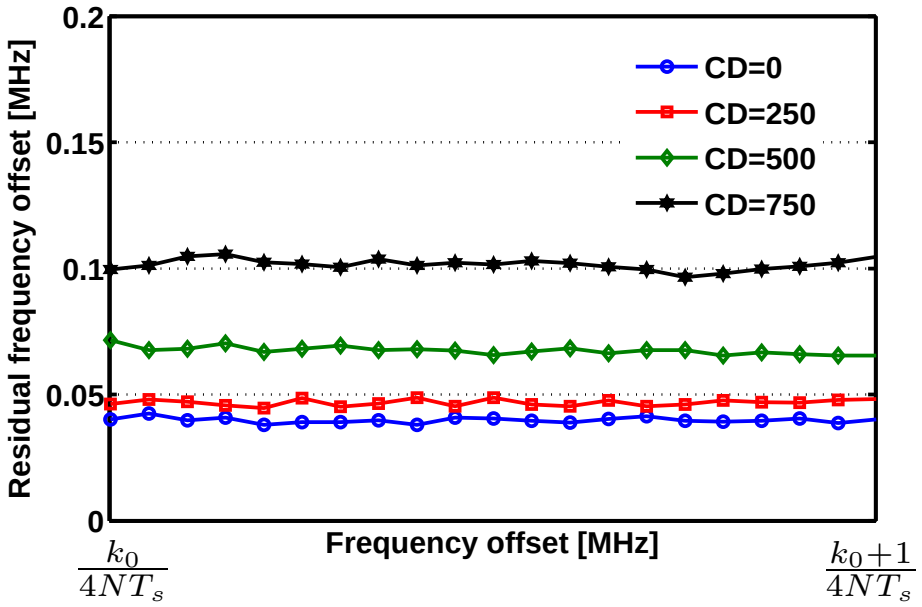


Figure 3.9: Residual CFO versus the initial CFO for different residual CDs using the proposed CFO estimate with both polarizations and fine step ($N = 1024$, OSNR = 20dB).

We remark that the proposed CFO estimate is robust to the presence of the residual CD since the residual CFO is kept below 150kHz up to 750ps/nm of CD.

3.5.2 Tolerance to PMD

Now, we would like to analyze the robustness of the proposed FO estimator against the PMD (assuming no residual CD). For the sake of simplicity, we consider the following channel filter frequency response (cf. [25, 59] or Ed. (1.21))

$$\tilde{\mathbf{C}}_a(\omega) = \mathbf{R}_\theta \mathbf{D}_{0,\phi}(\omega) \mathbf{R}_\theta^{-1} \tag{3.10}$$

with

$$\mathbf{R}_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}, \tag{3.11}$$

and

$$\mathbf{D}_{0,\phi}(\omega) = \begin{bmatrix} e^{+i\phi} & 0 \\ 0 & e^{-i\phi} \end{bmatrix}. \tag{3.12}$$

We thus omit the DGD since $\tau_{\text{DGD}} = 0$.

In Fig. 3.10, we plot MSE for CFO estimation versus ϕ and θ defined in Eqs. (3.11)-(3.12). For each channel realization, the MSE is averaged over 100 different values of CFO randomly chosen between 0 and 3.5GHz. We inspect four cases:

1. one polarization and fine step based estimator implemented **before** CMA equalization,

2. both polarizations and fine step based estimator implemented **before** CMA equalization,

3. one polarization and fine step based estimator implemented **after** CMA equalization,

4. both polarizations and fine step based estimator implemented **after** CMA equalization.

Notice that in order to avoid the singularity issue (the two output polarizations after CMA provide the same stream!), we implement the CMA proposed in [59] which handles the singularity issue [1].

When the CFO estimator is implemented before PMD compensation, *i.e.*, before CMA equalization, the CFO estimation fails when θ is between 30^o and 60^o. The failure probability is stronger when only one polarization is used as already seen for the outliers effect in Fig. 3.10. In contrast, the failure probability totally vanishes when the CFO estimator is implemented after the PMD compensation. Therefore, we advocate to equalize the received signal before to estimate the CFO which confirms the receiver structure described in Fig. 1.16.

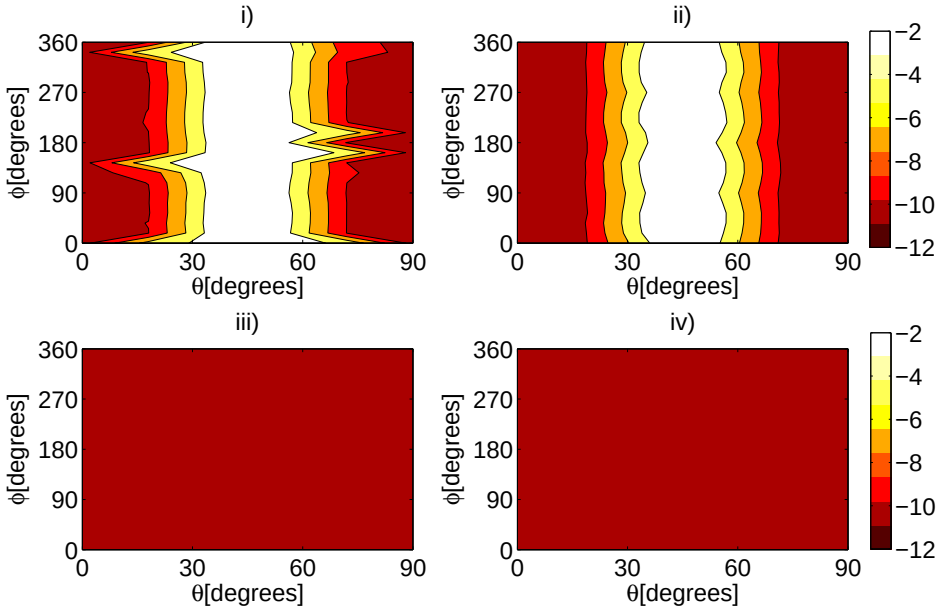


Figure 3.10: $\log_{10}$(MSE) ($N = 1024$, OSNR=20dB). i) one polarization before equalization, ii) both polarizations before equalization, iii) one polarization after equalization, and iv) both polarizations after equalization.

[1] Notice that we here implement the adaptive version of the CMA proposed in [59], while an adaptation to a blockwise version of this CMA is straightforward and done in Chapter 4.

In Fig. 3.11, we plot the received signal samples (at baud rate) before/after CMA equalization and before/after CFO estimation when $N = 1024$ and OSNR = 20dB for 16-QAM. Cd and PMD are assumed to be zero. The initial CFO is 2.2575GHz. We

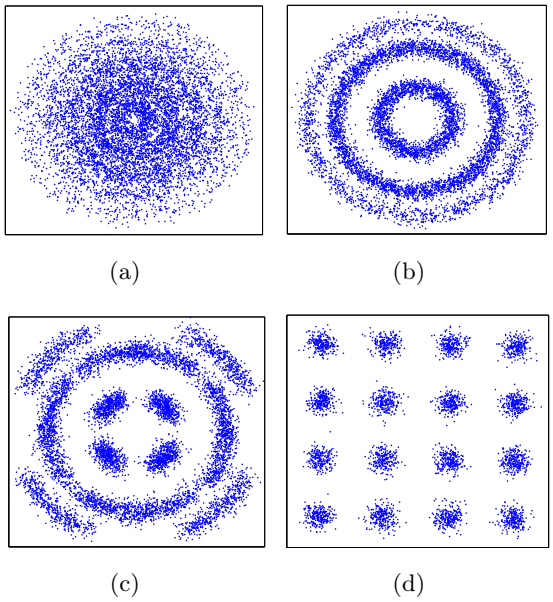


Figure 3.11: Received signal samples (at baud rate) (a) before CMA equalization (b) after CMA equalization (c) after CMA equalization and CFO estimation using only the coarse step (d) after CMA equalization and CFO estimation using fine step.

observe that the CMA equalizer followed by our CFO estimate enables us to retrieve the 16-QAM constellation properly.

3.5.3 Tolerance to phase noise

The phase noise is a major propagation effect that should be compensated for when using multilevel modulation formats since the receiver makes decision relying not only on the signal amplitude but also on the signal phase. Therefore any proposed CFO estimate has to tolerate phase noise.

In Fig. 3.12, we plot the residual CFO versus the initial CFO (taken between two adjacent FFT frequencies) for different values of the phase noise. The OSNR is fixed to 20dB, $N = 1024$, and only the CFO based on both polarizations and the fine step is considered. CD and PMD are assumed to be zero.

The residual CFO is sensitive to the phase noise level. For example, a residual CFO less than 2MHz is still ensured with a phase noise level of $\delta\nu T_s = 10^{-4}$. Such a phase noise level is close the maximum tolerable value in the state-of-the-art of the CFO estimators [32], [35] when PolMux 16-QAM modulation formats are considered.

3.5.4 Residual CFO compensation

After our CFO estimate, we remarked that a residual CFO still occurs. We would like to see if this residual CFO can be still compensated for by viewing it as a time-varying phase $(\hat{\varphi}_{1,N} - \varphi_1)n$. To compensate for this "phase noise", we will implement

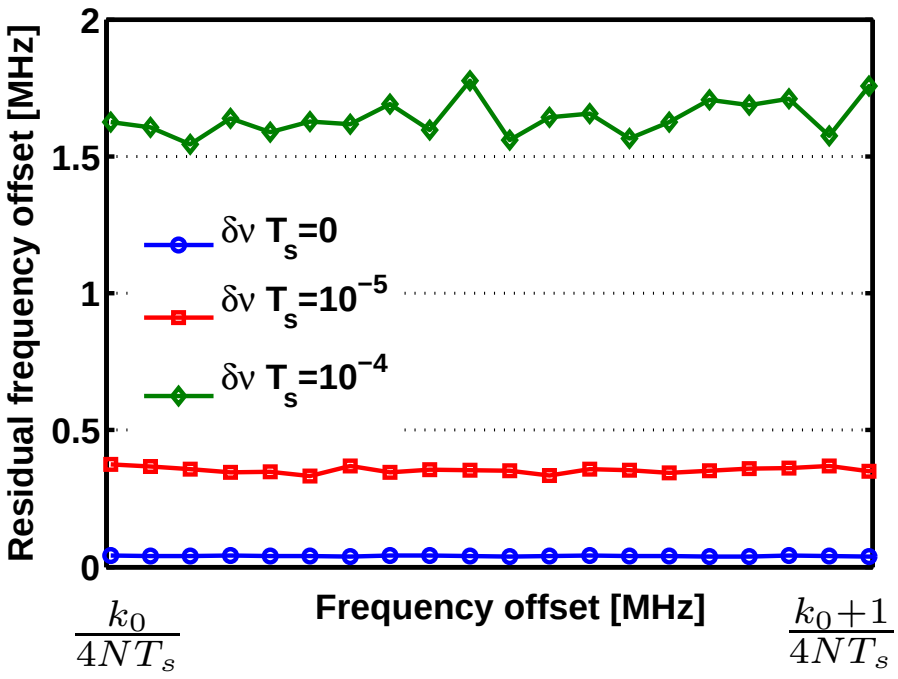


Figure 3.12: Residual CFO versus the initial CFO for different values of the phase noise using the proposed CFO estimate with both polarizations and fine step ($N = 1024$, OSNR = 20dB).

an additional phase tracking DD algorithm based on Eq. (1.68). In this section, CD and PMD are fixed to zero.

First of all, to designing the step-size of the DD phase adaptive algorithm, we plot in Fig. 3.13 the BER obtained with this algorithm versus OSNR in AWGN 16-QAM context. The phase and the CFO are assumed equal to zero.

We observe that even in absence of phase deviation, the phase compensation algorithm can degrade the performance if μ is too high. Indeed, due to noise, the 16-QAM points are not necessary well located and then the phase correction can be high (if μ is high) and goes to a wrong direction. Therefore μ has to be small enough to prevent this phenomenon. We will consider $\mu_{\text{DD}} = 10^{-2}$.

In Fig.3.14, we consider the BER versus OSNR when arbitrary residual CFO $\varphi_1 = \delta f_a T_s$ is considered. The values of φ_1 corresponds to 140kHz, 1.4MHz, and 14MHz respectively which are realistic values of residual CFO as viewed in previous subsections. The residual CFO is then compensated for through the DD adaptive phase algorithm. This extra phase estimate enables us to compensate for a residual CFO up to $140kHz$ without OSNR penalty at any reasonable target BER. Therefore coupling our block-wise CFO estimate (which provides typically a residual CFO of tens of kHz) with a DD adaptive algorithm for the residual one is a relevant scheme.

Finally, in Fig. 3.15, we plot the BER versus the initial CFO when different block-wise CFO estimates followed by a DD adaptive phase algorithm are carried out. We considered $N = 1024$, OSNR = 20dB, $\mu_{\text{DD}} = 10^{-2}$. Thanks to the fine step, the BER is is insensitive to the location of the initial CFO. In contrast, using the methods only

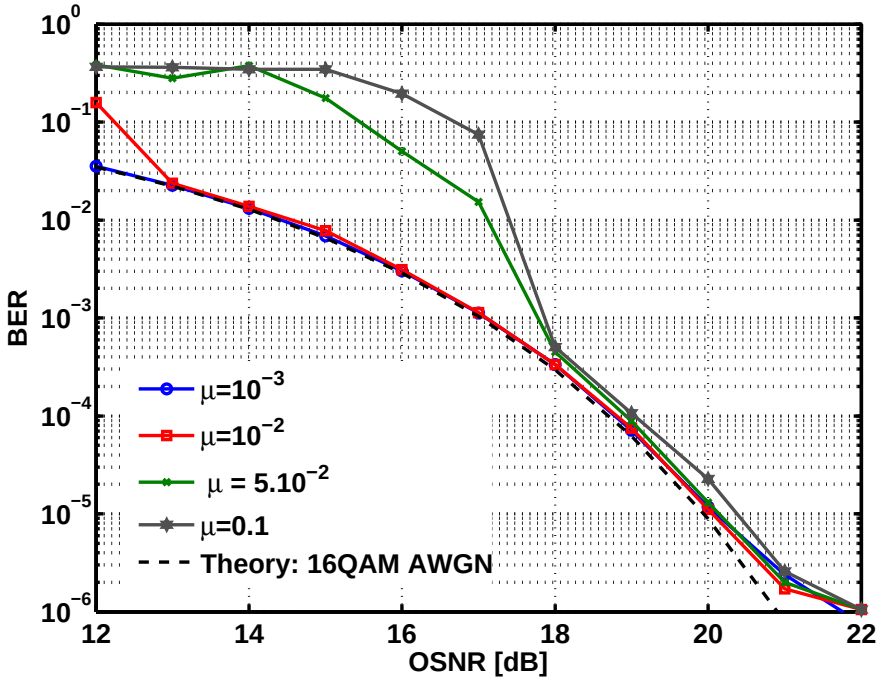


Figure 3.13: BER versus OSNR using DD adaptive phase estimate and various step-sizes μ_{DD}.

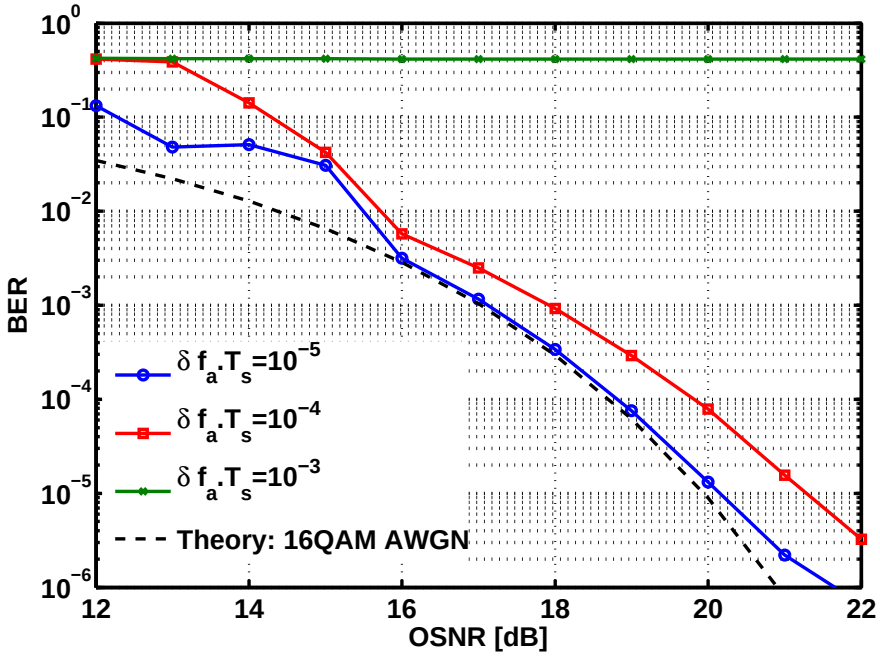


Figure 3.14: BER versus OSNR for DD adaptive phase algorithm when various residual CFO are used.

based on the coarse step often leads to a higher BER.

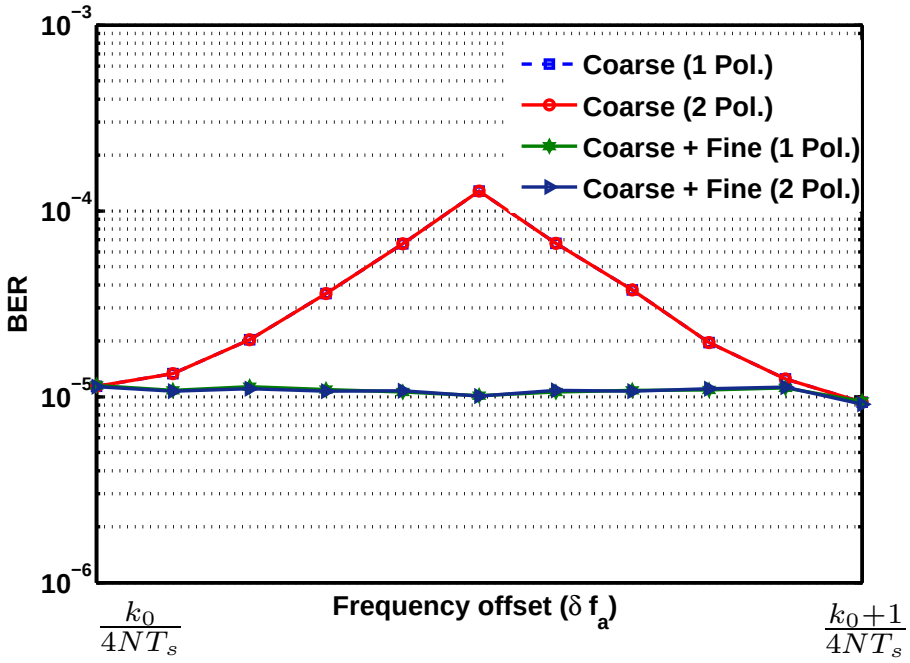


Figure 3.15: BER versus the initial CFO for different methods ($N = 1024$, OSNR = 20dB).

3.5.5 BER of the whole system

In order to study the impact of the accuracy of the used CFO estimator, we considered a POLMUX tranmission using 16-QAM 112Gbit/s. The channel exhibits a residual CD of 250ps/nm, a polarization rotation $\theta = \pi/4$, a value of DGD $\tau_{\text{DGD}} = 50ps$ and a total laser phase noise for both the transmission and the local oscillator lasers of $\delta\nu T_s = 10^{-5}$. CFO is generated randomly from 0 to 3.5GHz (corresponding to $R_s/4$).

In Fig. 3.16, we plot the BER of the above-mentioned system versus OSNR when three types of CFO estimates have been considered after residual CD and PMD compensation:

1. A differential based method with a size of block $N = 1024$ [45],
2. A spectrum maximization method based on the coarse step only with an FFT size of $N = 1024$, and using the signal contained on just one polarization,
3. A spectrum maximization method based on both the coarse and fine steps and using both polarizations, once again $N = 1024$ is the size of the FFT used.

The residual CD and PMD have been compensated for with an A-CMA with equalizer length $L = 3$ and fixed step-size $\mu = 10^{-3}$. Once the CFO is estimated and compensated for, we used a DD-based constant phase estimator as explained in the previous section with a fixed-step-size $\mu_{\text{DD}} = 10^{-2}$ to compensate for the residual CFO and the laser phase noise. The number of iterations is fixed to 10,000.

BER is kept below the FEC limit (2.10^{-3}) for an OSNR larger than 20dB, when using the third method. However, an error floor of $\sim 2.10^{-2}$ is obtained when using the

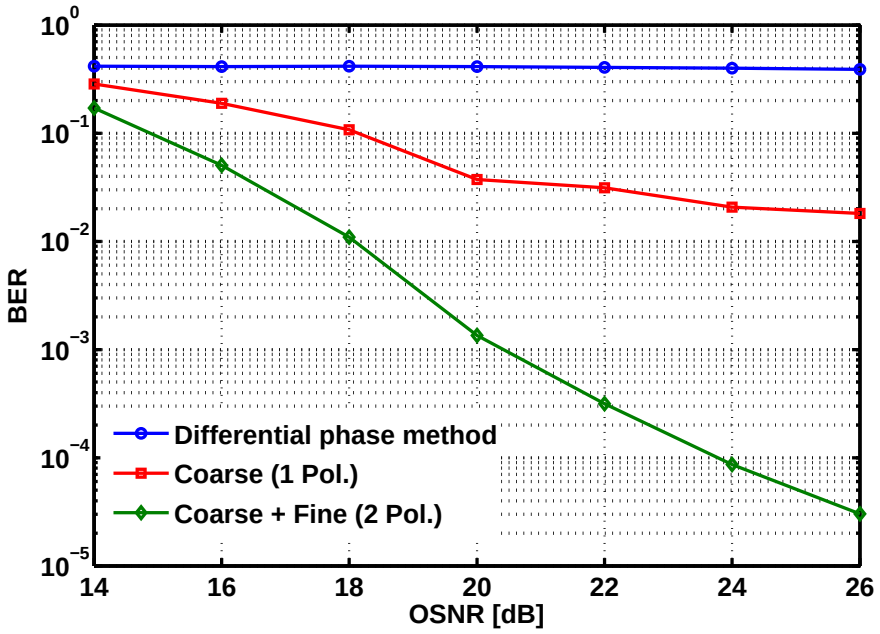


Figure 3.16: Performance of the transmission system using different methods for CFO estimation.

second method. Moreover, using the first method (due to insufficient accuracy of the CFO estimator), the receiver can not retrieve the transmitted sequence and the BER is kept too high regardless the considered OSNR.

3.6 Conclusion

In this chapter we presented a block-wise CFO estimate based on the spectrum maximization. This maximization relies on a coarse step via the Fast Fourier Transform of the fourth power received signal followed by a fine step based on a gradient-descent optimization. Thanks to the fine step, the residual CFO is insensitive to the location of the FO and it is kept small. Moreover the use of both polarizations reduces the outliers zone and still improves the estimation accuracy.
The proposed CFO estimate allows to achieve accuracy of some kHz, and it is robust to channel propagation impairments such as CD and laser phase noise. However, this estimator may fail if used before polarization demultiplexing. Our CFO estimate is well adapted for the multilevel modulation formats QAM and can be applied as well to the PSK formats.

Chapter 4

Block-wise Blind Equalization

4.1 Introduction

Coherent detection combined with multilevel modulation such as M-ary quadrature modulation (M-QAM) formats are one of the most relevant techniques to increase the spectral efficiency and reach higher bit rates [19, 60, 61]. Indeed, it has been shown that up to 400Gbit/s optical coherent transmission can be done by combining a real Analog to Digital Converter (ADC) with offline signal processing [20,62]. Nevertheless, only 112Gbit/s [63, 64] and 40Gbit/s [65] coherent transmission has been experimentally tested in real-time is now proposed in commercial products. Therefore coherent transmission is the leading candidate for the next generation optical transmission network at 100Gbit/s (also, called, 100Gbit Ethernet). However, due to the increase of the data traffic in a mid-term future, very high bit rate will be required (up to 1Tbit/s). To satisfy such a rate, the symbol rate and the constellation size have to be increased accordingly. Unfortunately, this ultra-high data rate transmission will be more sensitive to the various signal distortions generated by the optical fiber and the transmitter/receiver devices. Consequently the main challenge will be to develop digital signal processing algorithms counter-acting the propagation impairments (typically, the transmission distance is about several thousand kilometers) but compatible with electronic circuits complexity and speed.

Throughout this chapter, only the linear propagation impairments will be assumed. When polarization multiplexing (PolMux) is carried out, there are two kinds of interference: *i)* inter-symbol interference (ISI) associated with its own polarization due to the so-called (residual) chromatic dispersion (CD) and with the filtering effect at the transmitter and receiver sides, and *ii)* inter-polarization interference (IPI) due to the mixing of both polarizations given rise by the so-called polarization mode dispersion (PMD) [66, 67]. Another source of degradation concerns the phase errors which can be split into three categories: *i)* frequency offset, *ii)* constant phase offset, and *iii)* laser phase noise [68]. When the launched power is too high, some non-linear distortions such as those induced by the Kerr effect have to be taken into account as well [30, 69].

In the "signal processing" literature, numerous blind techniques have been developed for mitigating the ISI/IPI, the frequency offset, and the constant phase offset. We do not consider here training approaches for which a symbol sequence known both at the transmitter and receiver sides is periodically sent in order to estimate all the impairments parameters. Then, once those parameters are estimated, impairments are

mitigated using particular techniques. The description of these techniques is out of scope of this thesis.
In the "optical coherent receiver design" literature, the most widespread technique for the blind ISI/IPI compensation is the Constant Modulus Algorithm (CMA) [40] and its variants such as the Radius Directed Equalizer (RDE) [42] or the Multi Modulus Algorithm (MMA) (potentially followed by the Decision-Directed (DD) algorithm) [43]. For instance, these algorithms as implemented in [43] can compensate up to 1000ps/nm of CD in a 16-QAM coherent system, and lead to reach 100Gbit/s. Notice that all the above-mentioned algorithms belong to the set of the blind linear equalizers. So far, the blind ISI compensator has been implemented through adaptive algorithms, *i.e.*, the linear equalizer coefficients are updated as soon as one sample is incoming. Usually, for the sake of simplicity, the update equations are derived by means of the so-called stochastic gradient descent algorithm. Such a gradient-descent algorithm is carried out either with a constant step-size (as in [40]) or with a Hessian matrix based time-varying step-size (as done in [70] and Chapter 2).

Before going further, we remind that the propagation channel in optical communications is static over a large observation window since it varies very slowly compared to the symbol period. Indeed, the symbol period for 100Gbit/s QPSK systems is about 40ps whereas the coherence time of the channel is of order of a few milli-seconds [26, 36, 71]. Therefore, we propose here a new way (actually, a block-wise version) for implementing blind equalizers in optical communications.

The chapter is organized as follows: in Section 4.2, we calculate the coherence time of the optical channel in order to check the ability for the block-wise blind equalizer to be adapted to the optical channel. In Section 4.3, we introduce the block-wise blind equalizers. In Section 4.4, we analyze the performance of the proposed algorithms through simulations in depth. In Section 4.5, we focus on the singularity problem and we propose a new solution. In Section 4.6, experimental study is done thanks to real signals provided by Heinrich-Herz Institute (HHI) at Berlin. Finally concluding remarks are drawn in Section 4.7.

4.2 Coherence time

The *coherence time* denoted by T_{coh} is defined by the maximum delay for which the normalized MSE between two channel impulse responses is less than $\leq 1\%$. For the sake of simplicity, let us consider that the PolMux channel only exhibits an infinite polarization rotation. As a consequence, the channel impulse response at time t_0 is equal to (cf. Eq. (2.15))

$$\mathbf{C}_{a,t_0}(t) = \begin{bmatrix} \cos(\Omega t_0) & \sin(\Omega t_0) \\ -\sin(\Omega t_0) & \cos(\Omega t_0) \end{bmatrix} \delta(t). \tag{4.1}$$

In [26], it is showed that the speed of rotation Ω in already installed optical networks is about 26rad/s. Applying the definition, the coherence time is then given by

$$\| \mathbf{C}_{a,t_0}(t) - \mathbf{C}_{a,0}(t) \|^2 \leq 1\% \Rightarrow T_{coh} = 3.8ms \tag{4.2}$$

The coherence time is around few milliseconds. This should be compared to the duration of block of PolMux 16QAM 14Gbaud transmission, let us say containing 1000

symbols, which is of about 71ns. As a consequence, the fiber channel can be considered constant over few thousands blocks, and as stated in [72] "the impairments are of long duration, >100 blocks".
Consequently it is worth treating the data block-by-block rather than sample-by-sample. Therefore *the main contribution of this chapter is to propose an implementation of the blind equalizers in a block-wise way.* The main advantage of the block-wise approach compared to the sample-wise one is the convergence speed and the steady-state performance. Moreover, if bursty communications (with typical values of frame duration equal to a few micro-seconds) are considered, the first samples of the burst are enough to converge to an adequate equalizer whereas, as we will see later, the adaptive approach has not always converged at the burst end. In burst mode, the algorithm is usually initialized by a trivial equalizer at each burst beginning since the CD and the PMD can be strongly different and unknown for each burst since they depend on the wavelength routing and switching.
In addition, a lot of calculations can be also done in parallel and thus can be implemented with the current electronic devices. More precisely, *in this chapter, we introduce a block-wise version of the CMA and DD equalizer.* All the proposed algorithms work well for any PSK and QAM constellation, except BPSK when CMA is carried out (for more details, see [73]) . In the simulation part, 16-QAM is considered while, in the experimental part, 8-PSK is.
One of the main drawback of the block-wise algorithms may be its less ability to track the propagation channel variation. Nevertheless, given the quasi-static property of the optical fiber channel, we will see later that our approach is still robust to its time-variation, especially to its PMD variation.

4.3 Block-wise algorithms

In order to compensate for the channel impulse response, we introduce a $T_s/2$-fractionally spaced equalizer (FSE). We remind briefly the main notations needed in this chapter and already presented in Chapter 1.

Let $z_p(n)$ be the scalar output of the FSE associated with the polarization p. We have

$$z_p(n) = \sum_{k=0}^{L-1} \left(\overline{\mathbf{w}_{p,1}(k)} \mathbf{y}_1(n-k) + \overline{\mathbf{w}_{p,2}(k)} \mathbf{y}_2(n-k) \right) \tag{4.3}$$

where $\{\overline{\mathbf{w}_{p,q}(k)}\}_{k=0,\cdots,L}$ is the filter of length L (notice that each coefficient $\overline{\mathbf{w}_{p,q}(k)}$ is a 1×2 vector, *i.e.*, corresponds to a filter with 2 inputs and 1 output) between the input polarization p and the output polarization q.

Eq. (4.3) can be re-shaped easily by means of matrices as follows

$$z_p(n) = \mathbf{w}_p^{\mathrm{H}} \mathbf{y}^{(L)}(n) \tag{4.4}$$

where

- $\mathbf{w}_p = [\mathbf{w}_{p,1}(0), \cdots, \mathbf{w}_{p,1}(L-1), \mathbf{w}_{p,2}(0), \cdots, \mathbf{w}_{p,2}(L-1)]^{\mathrm{T}}$,
- $\mathbf{y}^{(L)}(n) = [\mathbf{y}_1(n)^{\mathrm{T}}, \mathbf{y}_1(n-1)^{\mathrm{T}}, \cdots, \mathbf{y}_1(n-L+1)^{\mathrm{T}}, \mathbf{y}_2(n)^{\mathrm{T}}, \mathbf{y}_2(n-1)^{\mathrm{T}}, \cdots, \mathbf{y}_2(n-L+1)^{\mathrm{T}}]^{\mathrm{T}}$.

- the superscript $(.)^{\mathrm{H}}$ stands for conjugate transposition.

Notice that the filters $\mathbf{w}_p$ have $4L$ coefficients as the received signals have been sampled at twice the baud rate.

We now would like to exhibit the filter $\mathbf{w}_p$ enabling us to have $z_p(n)$ close to $s_p(n)$. To do that, it is relevant to use the CMA criterion defined as the minimization of the following cost function [41].

$$J_p(\mathbf{w}_p) = \mathbb{E}[J_{p,n}(\mathbf{w}_p)] \tag{4.5}$$

with

$$J_{p,n}(\mathbf{w}_p) = (|z_p(n)|^2 - R)^2, \tag{4.6}$$

and

$$R = \frac{\mathbb{E}[|s_p(n)|^4]}{\mathbb{E}[|s_p(n)|^2]} \tag{4.7}$$

Here starts the main difference with the usual approach employed in coherent optical communications so far. Indeed, instead of implementing an adaptive version of this cost function, we decide to estimate the mathematical expectation of Eq. (4.5) given an observation block. Therefore we propose to minimize the following estimated cost function

$$\hat{J}_{p,N}(\mathbf{w}_p) = \frac{1}{N}\sum_{n=0}^{N-1} J_{p,n}(\mathbf{w}_p) \tag{4.8}$$

where N is the number of available quadrivariate samples $[\mathbf{y}_1(n)^{\mathrm{T}}, \mathbf{y}_2(n)^{\mathrm{T}}]$. Our purpose boils down to find the minimum of $\mathbf{w}_p \mapsto \hat{J}_{p,N}(\mathbf{w}_p)$. To do that, we suggest to use the (non-stochastic) gradient descent algorithm with optimal step size. If $\mathbf{w}_p^\ell$ is the estimated equalizer at the ℓ-th iteration (note that the data block is the same for each iteration), we have the following update relation [74] [75]

$$\mathbf{w}_p^{\ell+1} = \mathbf{w}_p^\ell - \mu^\ell \boldsymbol{\Delta}^\ell \tag{4.9}$$

with

$$\boldsymbol{\Delta}^\ell = \left.\frac{\partial \hat{J}_{p,N}(\mathbf{w})}{\partial \overline{\mathbf{w}}}\right|_{\mathbf{w}_p^\ell}.$$

One can check that

$$\boldsymbol{\Delta}^\ell = \frac{1}{N}\sum_{n=0}^{N-1}(|z_p(n)|^2 - R)\overline{z_p(n)}\mathbf{y}^{(L)}(n) \tag{4.10}$$

where $z_p(n)$ is calculated by inserting $\mathbf{w}_p^\ell$ into Eq. (4.4).

In order to find the optimal step size μ^ℓ at the ℓ-th iteration, we minimize the estimated cost function with respect to μ^ℓ, i.e.,

$$\mu^\ell = \arg\min_{\mu} \hat{J}_{p,N}(\mathbf{w}_p^\ell - \mu\boldsymbol{\Delta}^\ell). \tag{4.11}$$

The derivative of

$$\mu \mapsto \hat{J}_{p,N}(\mathbf{w}_p^\ell - \mu\boldsymbol{\Delta}^\ell)$$

is the following third-order polynomial function [74] [75]

$$P^\ell(\mu) = p_3^\ell\mu^3 + p_2^\ell\mu^2 + p_1^\ell\mu + p_0^\ell \tag{4.12}$$

where

$$
\begin{aligned}
p_3^\ell &= \frac{1}{N}\sum_{n=0}^{N-1} a_n^2, \\
p_2^\ell &= \frac{1}{N}\sum_{n=0}^{N-1} a_n b_n, \\
p_1^\ell &= \frac{1}{N}\sum_{n=0}^{N-1} (2a_n b_n + b_n^2), \\
p_0^\ell &= \frac{1}{N}\sum_{n=0}^{N-1} b_n c_n
\end{aligned}
$$

with

$$
\begin{aligned}
a_n &= |z_p(n)|^2, \\
b_n &= -2\Re(z_p(n)\overline{\delta}_n^\ell), \\
c_n &= (|z_p(n)|^2 - R), \\
\delta_n^\ell &= (\mathbf{\Delta}^\ell)^{\mathrm{H}}\mathbf{y}^{(L)}(n).
\end{aligned}
$$

Thanks to Eq. (4.12), we obtain in closed-form the roots of polynomial $P^\ell(.)$ and the real-valued root providing the minimum value of

$$\mu \mapsto \hat{J}_{p,N}(\mathbf{w}_p^\ell - \mu\mathbf{\Delta}^\ell)$$

will be the selected step size at the ℓ-th iteration. Finally, the architecture of the proposed block-wise equalizer is summarized in Fig. 4.1.

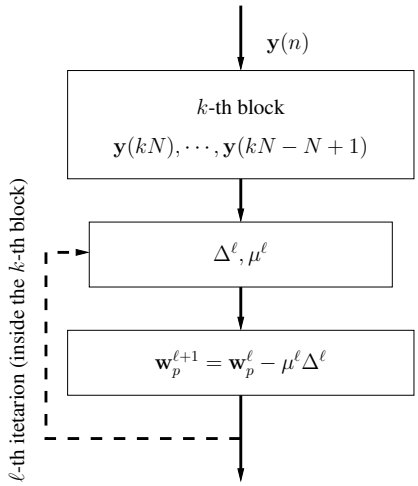


Figure 4.1: Structure of the proposed block-wise equalizer

The high complexity of the block-wise CMA with optimal step-size is mainly due to the numerical evaluations of the polynomial coefficients given in Eq. (4.12). To overcome this problem, as suggested in [76], a lot of these numerical evaluations can be done once (*i.e.*, we do not need to do them at each iteration, but just before the first

iteration). Indeed, one can observe that

$$\begin{aligned}
p_3^\ell &= \mathcal{T}_4(\boldsymbol{\Delta}^\ell, \boldsymbol{\Delta}^\ell, \boldsymbol{\Delta}^\ell, \boldsymbol{\Delta}^\ell), \\
p_2^\ell &= -3\Re(\mathcal{T}_4(\boldsymbol{\Delta}^\ell, \boldsymbol{\Delta}^\ell, \boldsymbol{\Delta}^\ell, \mathbf{w}_p^\ell)), \\
p_1^\ell &= -R\mathcal{T}_2(\boldsymbol{\Delta}^\ell, \boldsymbol{\Delta}^\ell) + 2\mathcal{T}_4(\mathbf{w}_p^\ell, \mathbf{w}_p^\ell, \boldsymbol{\Delta}^\ell, \boldsymbol{\Delta}^\ell) \\
&\quad + \Re(\mathcal{T}_4(\mathbf{w}_p^\ell, \boldsymbol{\Delta}^\ell, \mathbf{w}_p^\ell, \boldsymbol{\Delta}^\ell)) \\
p_0^\ell &= \Re(R\mathcal{T}_2(\mathbf{w}_p^\ell, \boldsymbol{\Delta}^\ell) - \mathcal{T}_4(\mathbf{w}_p^\ell, \mathbf{w}_p^\ell, \mathbf{w}_p^\ell, \boldsymbol{\Delta}^\ell))
\end{aligned}$$

with

$$\begin{aligned}
\mathcal{T}_2(\mathbf{a}_1, \mathbf{a}_2) &= \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{a}_1^{\mathrm{H}} \mathbf{y}^{(L)}(n)(\mathbf{y}^{(L)}(n))^{\mathrm{H}} \mathbf{a}_2, \\
\mathcal{T}_4(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) &= \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{a}_1^{\mathrm{H}} \mathbf{y}^{(L)}(n)(\mathbf{y}^{(L)}(n))^{\mathrm{H}} \mathbf{a}_2 \\
&\quad \times \mathbf{a}_3^{\mathrm{H}} \mathbf{y}^{(L)}(n)(\mathbf{y}^{(L)}(n))^{\mathrm{H}} \mathbf{a}_4
\end{aligned}$$

where $\mathbf{a}_s = [a_{s,0}, a_{s,1}, ..., a_{s,(4L-1)}]^{\mathrm{T}}$ for $s \in \{1, 2, 3, 4\}$ is a column vector of length $4L$ corresponding to the size of $\mathbf{y}^{(L)}(n)$. Actually, one can remark that

$$\mathcal{T}_2(\mathbf{a}_1, \mathbf{a}_2) = \mathbf{a}_1^{\mathrm{H}} \mathbf{R}_{\mathbf{y}} \mathbf{a}_2, \tag{4.13}$$

$$\mathcal{T}_4(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4) = \sum_{j,k,l,m} \mathbf{M}_{\mathbf{y}}(j,k,l,m)\overline{a_{1,j}} a_{2,k} \overline{a_{3,l}} a_{4,m} \tag{4.14}$$

where $\mathbf{R}_{\mathbf{y}}$ is the output covariance matrix given by

$$\mathbf{R}_{\mathbf{y}} = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{y}^{(L)}(n)(\mathbf{y}^{(L)}(n))^{\mathrm{H}} \tag{4.15}$$

and where the tensor $\mathbf{M}_{\mathbf{y}}$ is given by

$$\mathbf{M}_{\mathbf{y}}(j,k,l,m) = \frac{1}{N}\sum_{n=0}^{N-1} y_j(n)\overline{y_k(n)} y_l(n) \overline{y_m(n)} \tag{4.16}$$

where $y_s(n)$ is the s-th component of the vector $\mathbf{y}^{(L)}(n)$.

To calculate the polynomial coefficients, we essentially need to compute Eqs. (4.13) and (4.14). These computations can be strongly simplified by pre-computed $\mathbf{R}_{\mathbf{y}}$ and $\mathbf{M}_{\mathbf{y}}$ (via Eqs. (4.15) and (4.16)) which only depend on the data of the current block and are thus independent of the iteration.

Obviously the same derivations have to be done for the polarization q. Here we decide arbitrary to treat the ISI/IPI compensation on $z_p(n)$ and $z_q(n)$ separately which implies the minimization of the following both cost functions J_p and J_q. An alternative way is possible by minimizing the mixed function $J_p + J_q$. After extensive simulations, we have remarked that such an approach leads to similar results and thus is omitted hereafter. The block-wise approach has been introduced here by minimizing the CMA criterion. It is clear that this block-wise approach can be mimicked for other criteria, such as, the Decision-Directed (DD). For example, the DD is very useful when the blind compensation has converged in order either to track slight channel modification and to improve the estimation quality.

For instance, the block-wise DD equalizer carried out with the (non-stochastic) gradient algorithm using optimal step size is very simple to implement since we are able to exhibit closed-form expression for the optimal step size. Indeed, we have

$$\boldsymbol{\Delta}_{\mathrm{DD}}^{\ell} = \frac{1}{N}\sum_{n=0}^{N-1}(\overline{z_p(n) - \hat{s}_p(n)})\mathbf{y}^{(L)}(n) \tag{4.17}$$

where $\hat{s}_p(n)$ is the current decision on the symbol $s_p(n)$. Then minimizing the function $\mu \mapsto \hat{J}_{p,N,\mathrm{DD}}(\mathbf{w}_p^{\ell} - \mu\boldsymbol{\Delta}_{\mathrm{DD}}^{\ell})$ leads to

$$\mu_{\mathrm{DD}}^{\ell} = \frac{\sum_{n=0}^{N-1}\Re\{\overline{\delta_{n,\mathrm{DD}}^{\ell}}(z_p(n) - \hat{s}_p(n))\}}{2\sum_{n=0}^{N-1}|\delta_{n,\mathrm{DD}}^{\ell}|^2} \tag{4.18}$$

with $\delta_{n,\mathrm{DD}}^{\ell} = (\boldsymbol{\Delta}_{\mathrm{DD}}^{\ell})^{\mathrm{H}}\mathbf{y}^{(L)}(n)$.

An other way to implement the block-wise algorithms is the Pseudo-Newton one (as done in Chapter 2 in the adaptive context). Instead of choosing the optimal step-size, we select the Hessian matrix as step-size. This leads to the following algorithm when CMA criterion is considered.

$$\mathbf{w}_p^{\ell+1} = \mathbf{w}_p^{\ell} - \mu\mathbf{G}^{\ell}\boldsymbol{\Delta}^{\ell} \tag{4.19}$$

with $\mathbf{G}^{\ell}$ the inverse of the Hessian matrix calculated in a recursive manner and using the matrix inversion lemma. This matrix thus takes the following form

$$\mathbf{G}^{\ell} = \lambda^{-1}\mathbf{G}^{\ell-1} - \frac{\lambda^{-2}\mathbf{G}^{\ell-1}\mathbf{m}\mathbf{m}^{\mathrm{H}}\mathbf{G}^{\ell-1}}{\left(2(1-\lambda)\frac{1}{N}\sum_{n=0}^{N-1}|z_p^{\ell}(n)|^2\right)^{-1} + \lambda^{-1}\mathbf{m}^{\mathrm{H}}\mathbf{G}^{\ell-1}\mathbf{m}} \tag{4.20}$$

where

$$\mathbf{m} = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{y}^{(L)}(n), \quad 0 \leq \lambda \leq 1, \tag{4.21}$$

$\lambda + \mu = 1$, and $\mathbf{G}^0 = \delta\mathbf{Id}$ with $\delta > 0$.

In addition to the analysis of the performance of the proposed block-wise approaches (see Section 4.4), it is important to check that the complexity of the block-wise approaches is kept to reasonable values.

As the computational complexity of the real multiplications, divisions, additions and subtractions are negligible in comparison to complex multiplications. These operations are neglected in the computation in the calculation of the computational complexity. In addition, as the extraction of the third-order degree polynomial roots for the BO-CMA does not depend neither on the length of the equalizer nor on the block size, and can be done using some analytical methods such as the Cordano's formula, we have neglected it as well.

In the sequel, the considered algorithms are listed below:

- A-CMA: standard Adaptive CMA.
- AN-CMA: Adaptive Newton based CMA
- BF-CMA: Block-wise CMA with fixed step-size (here, the step-size is $\mu = 0.02$; for more details about this choice, see Section 4.4).

- BN-CMA: Block-wise CMA with variable step-size based on the Pseudo-Newton equalization approach.
- BO-CMA: Block-wise Optimal step-size based CMA.

In next Tables, we put the number of flops (complex multiplications) required for various algorithms to reach the same BER performance (in the simulation, the target BER was fixed to 2.10^{-3} without channel coding technique). The equalizer length is fixed to $L = 3$. In Table 4.1, we have plotted the results for the adaptive algorithms.

Adaptive		
Algorithms	A-CMA	AN-CMA
per it. and pol.	$2(4L+1)$	$4(4L)^2 + 12L + 4$
# it.	10000	5000
Total Flops ($\times 10^3$)	**520**	**3085**

Table 4.1: Complexity for various adaptive CMA.

In Table 4.2, we have plotted the complexity for the different block-wise algorithms when the pre-computation technique has not been employed. Concerning the BO-CMA, we have remarked (see Section 4.4) that the optimal step-size associated with the cost function $J_1(.)$ is almost identical to the optimal step-size associated with the cost function $J_2(.)$. Consequently, we only compute the polynomial once per iteration (either on $J_1(.)$ or on $J_2(.)$).

Block without pre-computation ($N = 1000$)			
Algorithms	BF-CMA	BN-CMA	BO-CMA
Update eq. (per it. and pol.)	$2N(4L+1)$	$2N(4L+1) + 4(4L)^2 + 4L + 2$	$2N(4L+1)$
Polynomial evaluation (per it.)	-	-	$4N(3L+1) + 4L$
# it.	40	35	25
Total Flops ($\times 10^3$)	**2080**	**1862**	**2300**

Table 4.2: Complexity for various CMA without pre-computation.

In Table 4.3, we have plotted the complexity for the different block-wise algorithms when the "pre-computation" trick has been used.

Block with pre-computation ($N = 1000$)			
Algorithms	BF-CMA	BN-CMA	BO-CMA
Pre-computation	$N[\binom{4L+3}{4} + \binom{4L+1}{2}]$		
Update eq. per it. and pol.	$(4L)^3 + 2.(4L)$	$(4L)^3 + 2.(4L)$	$(4L)^3 + 2.(4L)$
Hessian eval. per it. and pol.	-	$4(4L)^2 + 4L + 2$	-
Polynomial eval. per pol.	-	-	$5(4L)^4 + 2(4L)^2$
# it.	40	35	25
Total Flops ($\times 10^3$)	**1584**	**1607**	**4130**

Table 4.3: Complexity for various Block CMA with pre-computation.

The pre-computation technique is not worthy for BN-CMA and BO-CMA in our context. Actually, the pre-computation approach will be of interest if the number of iterations grows significantly. By the way, we remark that the BF-CMA (resp. BO-CMA)

is only three times (resp. eight times) more complex than the A-CMA but uses a much smaller set of samples and converges quite fastly with 40 (resp. 25) iterations. At the expense of a higher (but not unreasonable) complexity, the block-wise approaches thus converge with few samples and are especially well-adapted for burst mode transmission.

4.4 Block-wise equalization performance

This section deals with the performance evaluation of the block-wise algorithms through simulations. We firstly focus on the static channel impulse response along the entire observation window in Section 4.4.1. Then we move on non-static channel in Section 4.4.2. The channel is modeled as in Section 1.3.3.

4.4.1 Static channel case

Except otherwise stated, in order to evaluate the performance of our algorithms, we considered the following transmission channel: the chromatic dispersion $DL_f = 1000$ps/nm (such a DL_f value corresponds to a standard residual CD), the DGD delay $\tau_{\text{DGD}} = 50$ps, and the polarization rotation $\theta = \pi/4$. The OSNR (in 0.1nm) is set to 20dB. The equalizer length is fixed to $L = 3$.

We test our block-wise CMA algorithms by initializing each equalizer filter $\mathbf{w}_1$ and $\mathbf{w}_2$ with the filter $\mathbf{w}^0$ whose coefficients are 0 except the central one equal to 1. These equalizer filters are initialized with $\mathbf{w}^0$. Then, inside each block, the coefficients of these equalizer filters are updated according to Eq. (4.9). When we stop to update the filter, we apply the obtained equalizer filter to the entire considered block. The BER point of any figure is obtained by averaging 150 block trials.

In Fig. 4.2, we depict the BER of the BO-CMA versus the number of iterations for various block sizes N. The algorithm convergence is mostly obtained for a number of iterations larger than 25. We are able to obtain a BER equal to 10^{-3} (so just below the FEC limit) when the block sizes are larger than 1000. We obviously remark that the steady state of the BO-CMA is better for large block sizes.

We now move into the analysis of the BF-CMA. We firstly would like to find a convenient value for the fixed step-size μ ensuring the convergence. For doing that, we consider the block size $N = 1000$ and the number of iterations equal to 50. In Fig. 4.3, we plot the histogram (displayed in percent) of the best fixed step-size μ. The histogram is carried out in the following way: numerous channel realizations are simulated by considering the cumulative chromatic dispersion DL_f varying from 0 to 1250ps/nm, τ_{DGD} varying from 0 to 80ps, and the polarization rotation θ from 0 to π. For each channel realization, we select the μ between 10^{-1} and 10^{-3} leading to the best steady-state BER when the BF-CMA equalizer filters are initialized with $\mathbf{w}^0$. By looking at Fig. 4.3, we decide to choose definitively the fixed step-size (for the BF-CMA) equal to 0.02. Before comparing the performance of the BF-CMA and the BO-CMA, it is of interest to compare the fixed step-size $\mu = 0.02$ of the BF-CMA with the values of the optimal step-size revealed by the BO-CMA. We have remarked that the optimal step-sizes obtained through the BO-CMA for each polarization, *i.e.* $\mu_1^{\ell,\text{opt}}$ and $\mu_2^{\ell,\text{opt}}$, are quite identical as already-mentioned in Section 4.3. Consequently, we consider only one optimal step-size, denoted now by $\mu^{\ell,\text{opt}}$ (at the ℓ-th iteration). In Fig. 4.4, we plot the optimal step-size (of the BO-CMA) and the fixed step-size (of the BF-CMA) versus

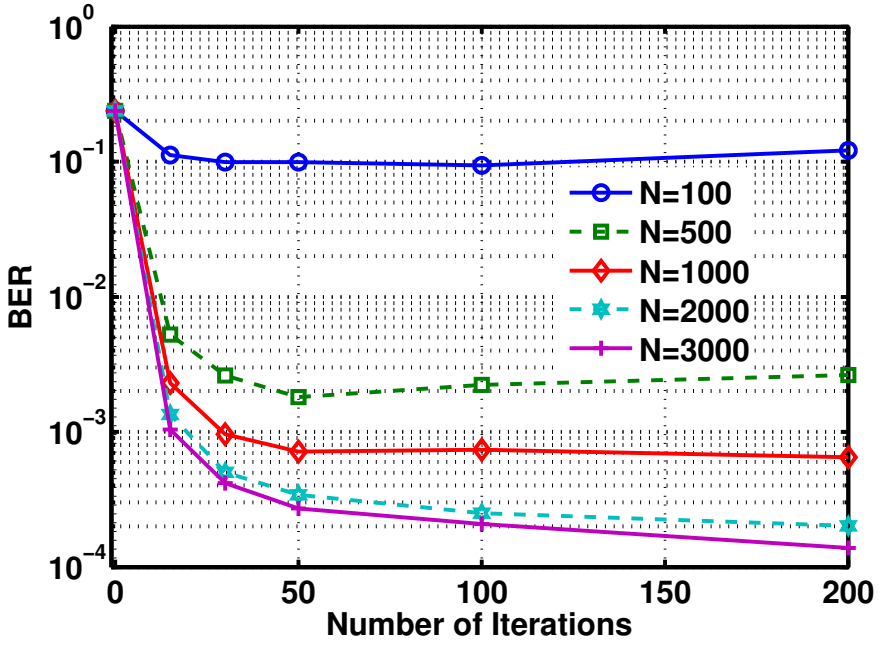


Figure 4.2: BER of the BO-CMA versus the number of iterations for various block sizes N (OSNR = 20dB, $DL_f = 1000$ps/nm, $\tau_{\text{DGD}} = 50$ps, $\theta = \pi/4$).

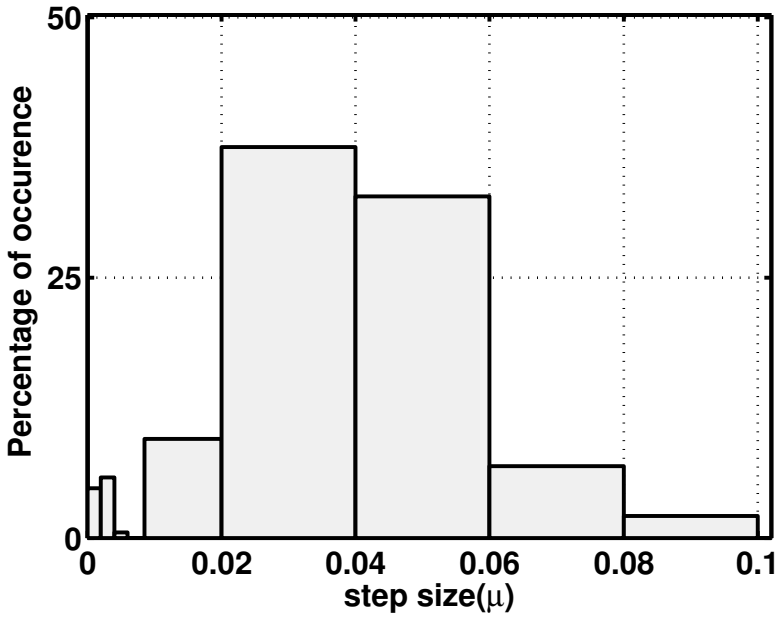


Figure 4.3: Histogram of the best fixed step-size μ for the BF-CMA ($N = 1000$, OSNR = 20dB, $DL_f \in [0, 1250]$ps/nm, $\tau_{\text{DGD}} \in [0, 80]$ps, $\theta \in [0, \pi]$).

the number of iterations when $N = 1000$. We observe that the optimal step-size is higher than the fixed one for the first iterations leading to a higher convergence speed. As for the last iterations, the optimal step-size is smaller than the fixed one in order to prevent the algorithm to oscillate too much around the final value.

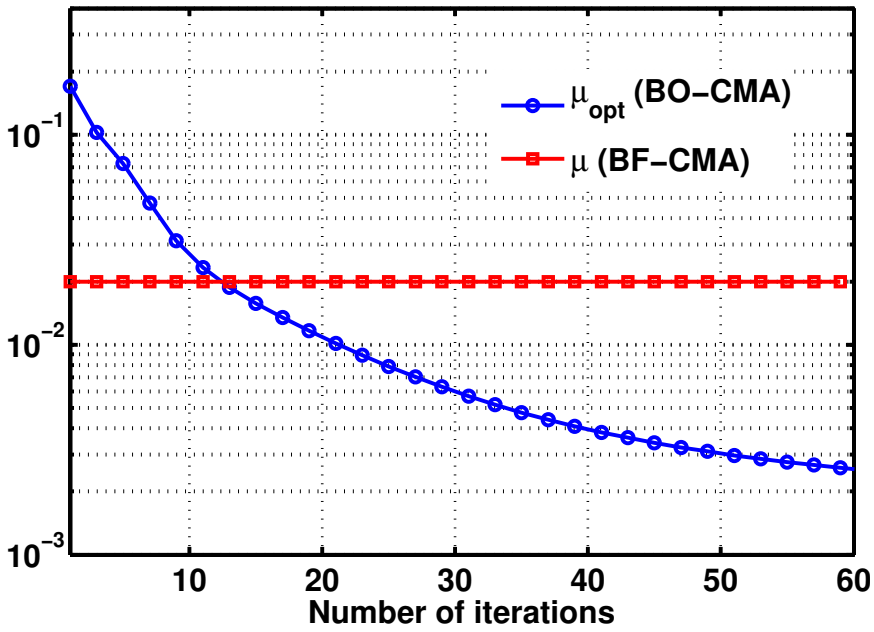


Figure 4.4: Optimal step size (BO-CMA) and fixed step-size (BF-CMA) versus the number of iterations ($N = 1000$, OSNR = 20dB, $DL_f = 1000$ps/nm, $\tau_{\text{DGD}} = 50$ps, $\theta = \pi/4$).

In Fig. 4.5, we then compare the convergence speed for the BO-CMA, the BN-CMA, and the BF-CMA versus the number of iterations when $N = 1000$. The BO-CMA is the fastest one since only 25 iterations were required to obtain a BER equal to 10^{-3} whereas 40 iterations are needed for the BF-CMA. However their steady-states are similar. Notice that the number of iterations used in Tables 4.1, 4.2, and 4.3 have been chosen according to Fig. 4.5.

For the sake of clarity and simplicity, we now only display performance associated with the BO-CMA. So far, we only compare block-wise CMA algorithms to each others. To inspect the real usefulness of block-wise CMA algorithms, we will compare them (actually, only the BO-CMA) to the well-known adaptive CMA (A-CMA). In Fig. 4.6, we plot the BER of the BO-CMA (with 50 iterations inside each block) and the A-CMA (with fixed step-size equal to 10^{-3}) versus the observation window length. Notice that, for the BO-CMA, the observation window length is identical to the block size N, whereas, for the A-CMA, the observation window length is identical to the number of iterations. Both algorithms are initialized with $\mathbf{w}^0$ at the beginning of the observation window. We show that the BO-CMA significantly improves the convergence speed since only 1000 symbols are necessary to reach the usual target BER (around 10^{-3}) instead of 10000 for the A-CMA.

Until now, we only looked the performance for one block transmission. Such an approach is of interest when we would like to analyze a transmission start. In the context of successive block transmission, it is clear that we have to look at the behavior of these algorithms when the initialization of the k-th block is provided by the equalizer filters obtained in the $(k-1)$-th block. Hereafter, the channel realization is still assumed to be the same whatever the considered block.

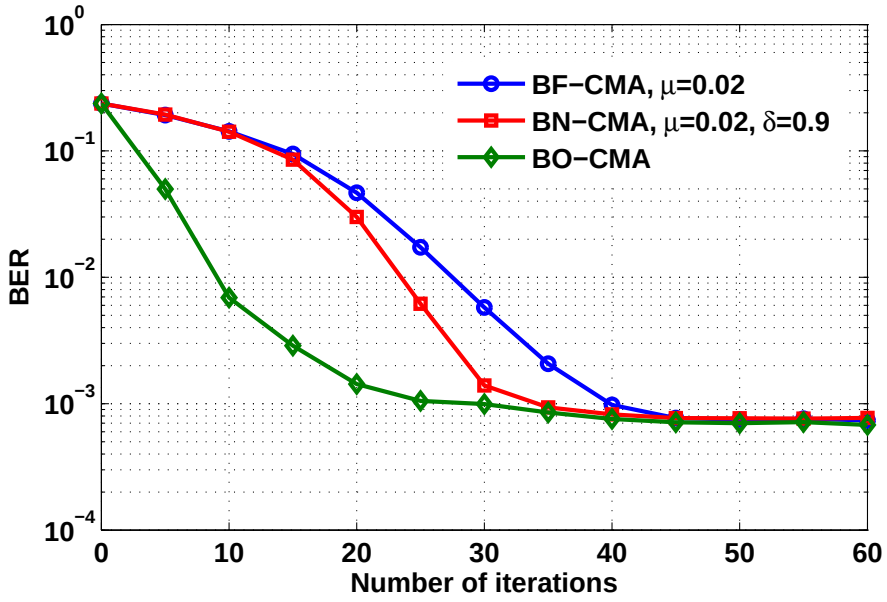


Figure 4.5: BER of the BO-CMA and the BF-CMA versus the number of iterations ($N = 1000$, OSNR = 20dB, $DL_f = 1000$ps/nm, $\tau_{\text{DGD}} = 50$ps, $\theta = \pi/4$).

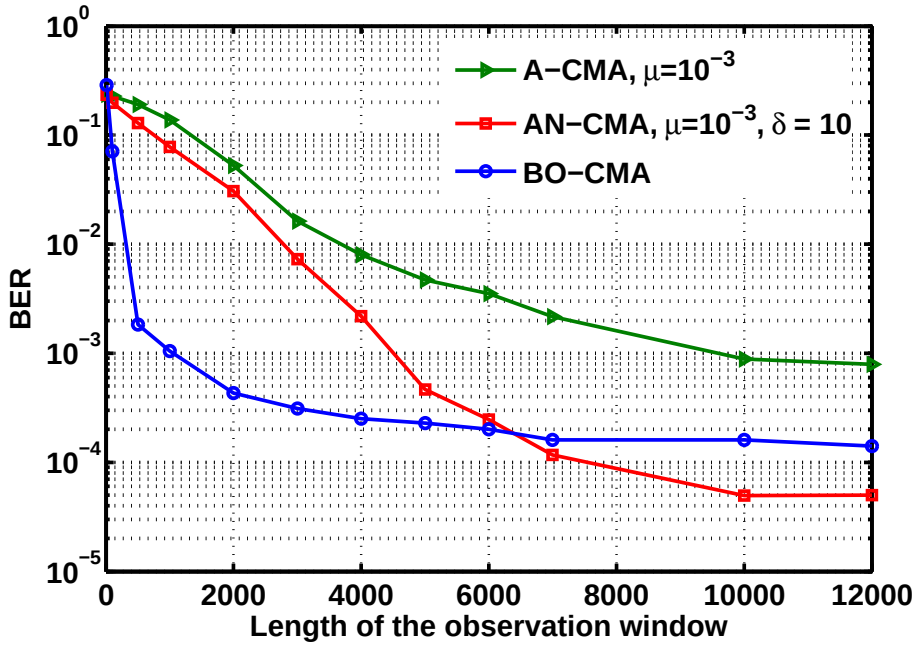


Figure 4.6: BER of the BO-CMA and the A-CMA versus the observation window length (OSNR = 20dB, $DL_f = 1000$ps/nm, $\tau_{\text{DGD}} = 50$ps, $\theta = \pi/4$). For the BO-CMA, the observation window length is identical to the block size N. For the A-CMA, the observation window length is identical to the number of iterations.

In Fig. 4.7, we plot the number of iterations versus the position of the block within the transmission flow. As the channel is static, we see that the number of iterations

decreases with respect to the block number. It makes sense since at the beginning of the transmission (corresponding to a transition phase), the algorithm has to learn more about the channel compared to the middle and to the end of the transmission. At the end of the transmission, the algorithm is already well-initialized and just has to update slightly the equalizer coefficients. So, the more block number is high, the less iteration number is needed. For a block size $N = 1000$, less than 10 iterations is necessary after the transition phase.

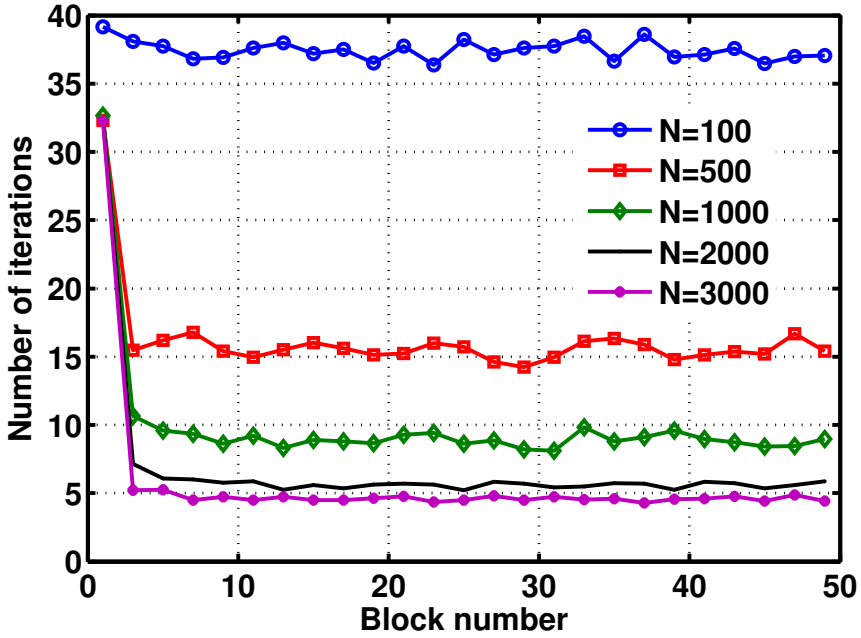


Figure 4.7: The number of iterations versus the block number for the BO-CMA, (OSNR = 20dB, DL_f = 1000ps/nm, τ_{DGD} = 50ps, $\theta = \pi/4$) when the proposed stopping condition is applied on the BO-CMA.

As the number of iterations depends on the block number, on the channel realization, it is worth developing a stopping criterion different from the number of iterations. We propose to stop the update when the term

$$\alpha_p^\ell = \frac{||\mathbf{w}_p^{\ell+1} - \mathbf{w}_p^\ell||}{||\mathbf{w}_p^\ell||} \tag{4.22}$$

is below a certain threshold. It is clear that if the steady-state is almost reached, the term α_p^ℓ will be very small. After extensive simulations not reported in this chapter, we found that a target BER of 10^{-3} is usually reached when α_p^ℓ is around 5.10^{-3}. Therefore, concerning the BO-CMA, we fix the threshold for α_p^ℓ to 5.10^{-3}. To be sure to stop the algorithm (even if it does not converge), we add a second constraint by fixing the maximum number of iterations to be equal to 40.

In Fig. 4.8, we plot the stopping numbers versus the number of iterations for both polarizations. We remark that the stopping numbers decrease rapidly and have almost the value for both polarizations.

In Fig. 4.9, we plot the BER for the BO-CMA versus the block size N when the BO-CMA applied on the k-th block is initialized by the equalizer filters provided by

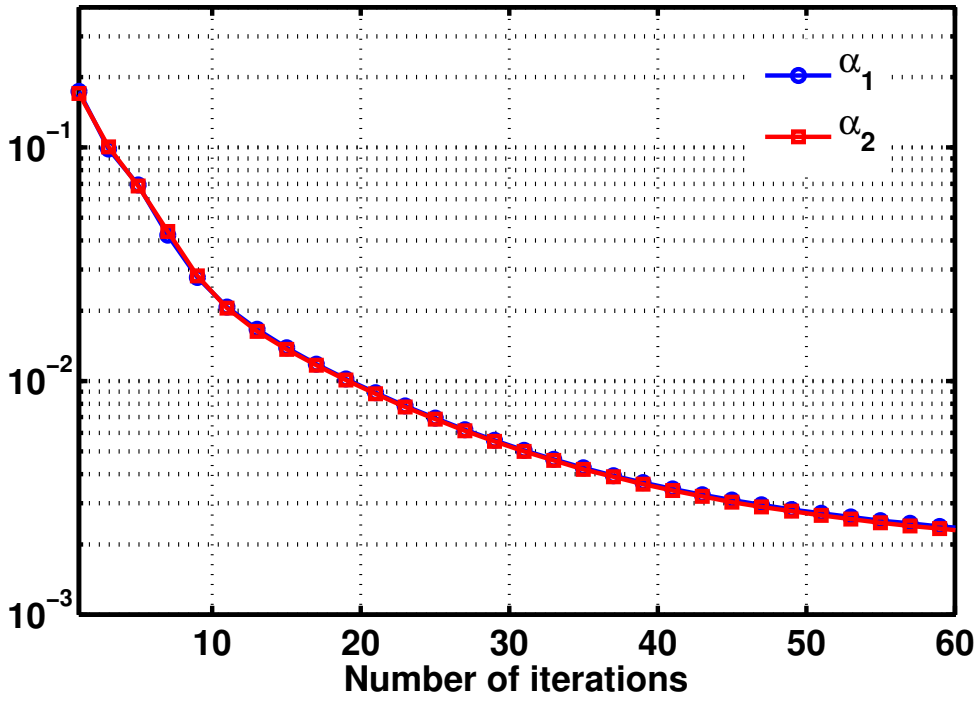


Figure 4.8: Evolution of α_1 and α_2 versus the number of iterations inside one block.

the $(k-1)$-th block and when the aforementioned stopping condition is considered. We have observed that when the block size is too small (*e.g.*, $N = 100$), the performance are poor. The reason is that the necessary number of iterations is then higher than 40. As soon as N is large enough, the stopping condition is well-designed and the performance in terms of BER are really good.

Herafter, we consider an observation window of 10000 which corresponds to the number of filter updates for the adaptive equalization case. For block-wise approach, we considered 10 blocks of size 1000 symbols each, and once again at the beginning of each block k, the equalization filter is initialized with the coefficients calculated at the block $(k-1)$.

We study the impact of the equalizer length of the overall performance of the system for both the adaptive and the block-wise equalization approaches. Using the A-CMA, equalizers filters with $L > 3$ offer the best system performance (see Fig. 4.10). The block-wise approaches offer the best performance with $L = 3$ (see Figs. 4.11 and 4.12).

Fig. 4.13 presents the BER versus the OSNR for the A-CMA, BO-CMA, and BF-CMA. The system has an OSNR penalty of 3dB at BER=10^{-3} for the A-CMA, in comparison to 2dB for the methods based on the block-wise approach. Note that the BF-CMA and the BO-CMA have the same performance. However, the BO-CMA requires less iterations inside each block to ensure the stopping condition $\alpha_p^\ell \leq 5.10^{-3}$.

4.4.2 Non-static channel case

In this subsection, we would like to analyze the ability of the BO-CMA to track channel time-variation. For the sake of simplicity, we only consider infinite polarization rotation

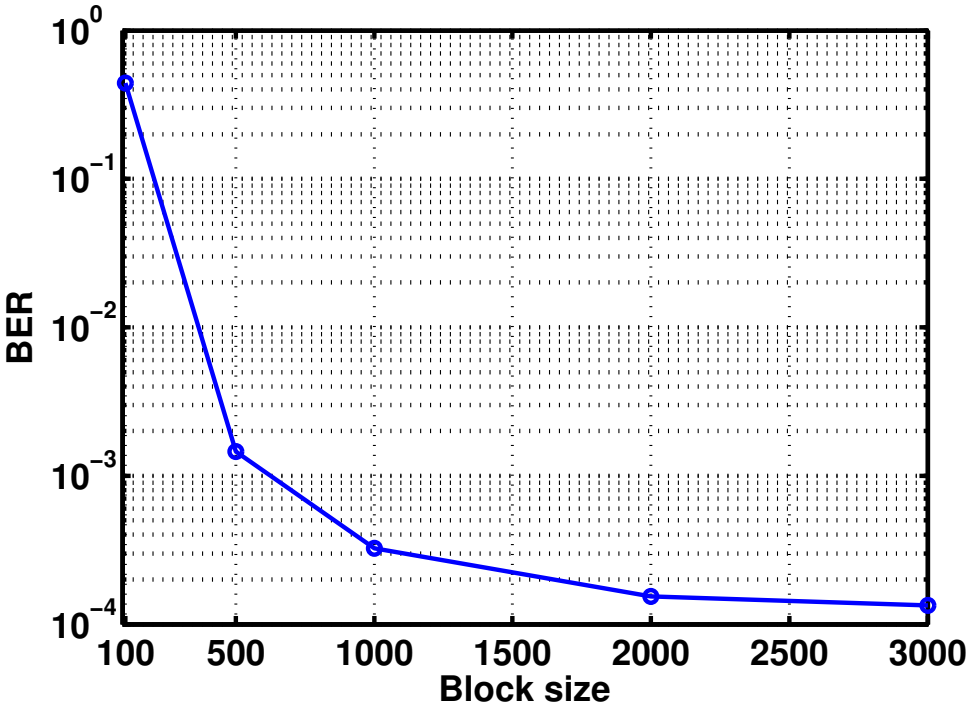


Figure 4.9: BER versus the block size N for the BO-CMA (OSNR = 20dB, DL_f = 1000ps/nm, $\tau_{\text{DGD}} = 50$ps, $\theta = \pi/4$).

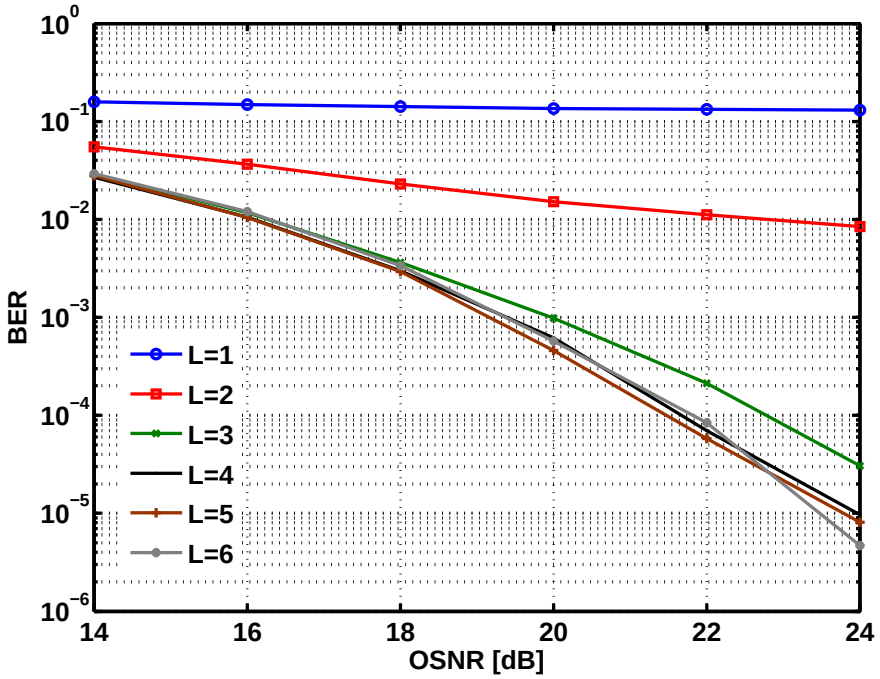


Figure 4.10: BER versus OSNR for various A-CMA equalizer lengths (the observation window is 10000 symbols, $DL_f = 1000$ps/nm, $\tau_{\text{DGD}} = 50$ps, $\theta = \pi/4$).

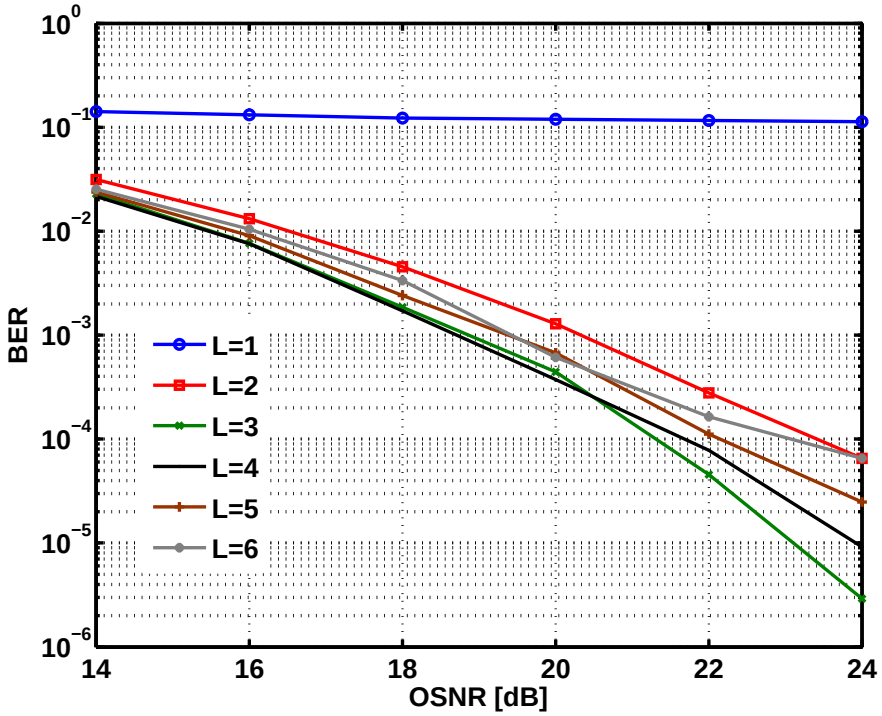


Figure 4.11: BER versus OSNR for various BF-CMA equalizer lengths (the observation window is 10000 symbols, $DL_f = 1000$ps/nm, $\tau_{\text{DGD}} = 50$ps, $\theta = \pi/4$, $\mu = 0.02$).

modeled by the Jones matrix. Consequently, the residual CD is assumed to be null, and the PMD only gives rise to one time-varying rotation. The polarization mixing is thus instantaneous and does not lead to inter-symbol interference (ISI) but just to inter-polarization interference (IPI). The channel impulse response at time t_0 is that defined in Eq. (2.15) following the Jones model.

In Fig. 4.14, the BER for the BO-CMA and the A-CMA is numerically evaluated versus the rotation speed Ω. We inspect several values of the block size N in the case of the BO-CMA. We remind that the BO-CMA algorithm for the k-th block is initialized with the equalizer filters provided by the $(k-1)$-th block, and the stopping condition on α_p^ℓ is applied. The tracking ability is better for small block sizes. Moreover the BO-CMA has better tracking ability than the A-CMA. For example, for $N = 1000$, a target BER of 10^{-3} is reached up to $\Omega = 3$Mrad/s while the A-CMA is unable to track variation above $\Omega = 1$Mrad/s. The steady-state is better for larger block sizes and for low rotation speed of polarization. Notice that the steady-states are different from those offered in previous figures since the channel is built differently and is easier for small equalizer lengths due to the absence of inter-symbol interference. Besides, the smaller the block size is, the better the track ability is. As a conclusion, the block-wise CMA approach is an very promising solution since it needs smaller observation window and it offers better tracking ability.

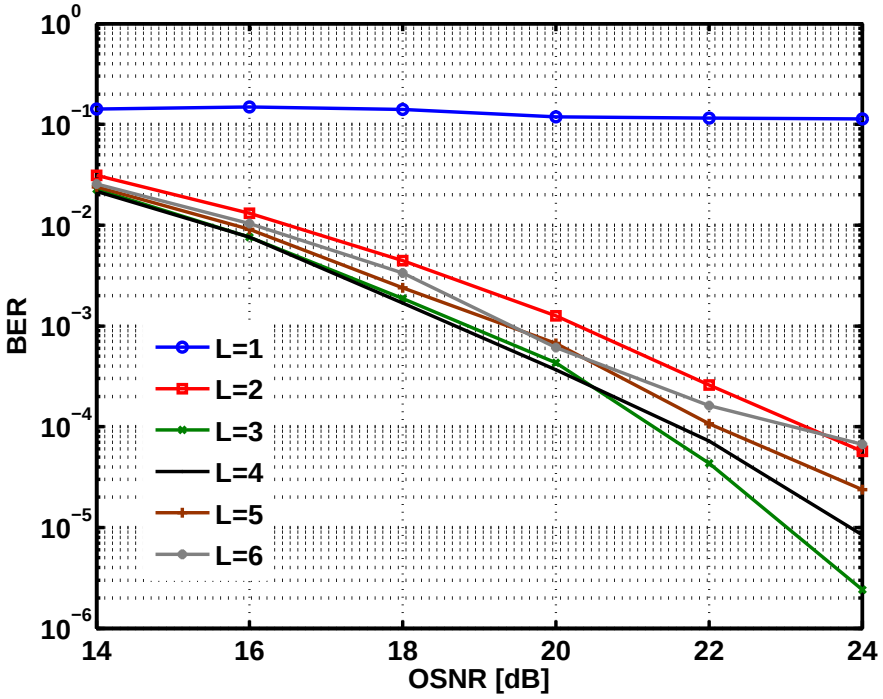


Figure 4.12: BER versus OSNR for various BO-CMA equalizer lengths (the observation window is 10000 symbols, $DL_f = 1000$ps/nm, $\tau_{\mathrm{DGD}} = 50$ps, $\theta = \pi/4$).

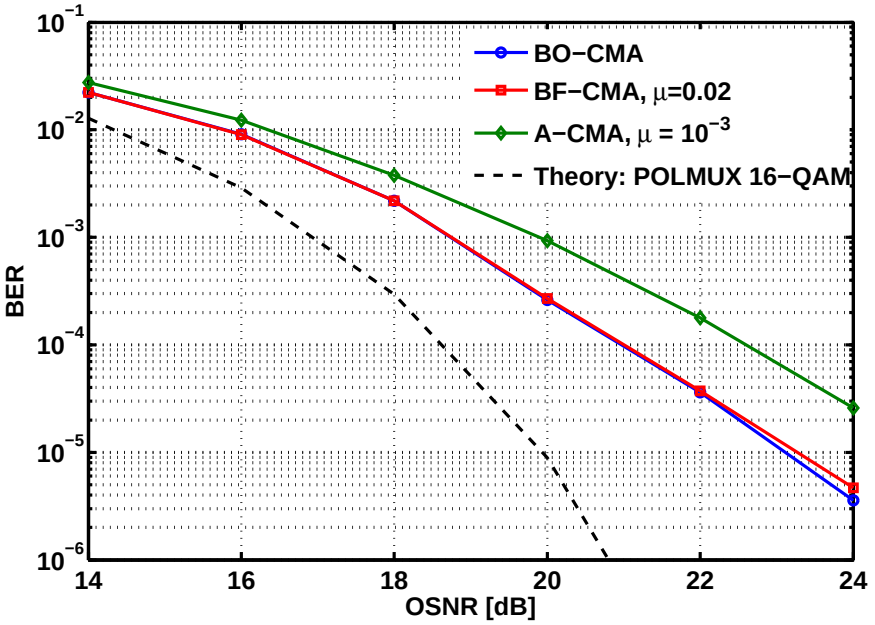


Figure 4.13: BER versus OSNR for the different equalizers (the observation window is 10000 symbols, $DL_f = 1000$ps/nm, $\tau_{\mathrm{DGD}} = 50$ps, $\theta = \pi/4$).

4.5 Singularity issue

The use of polarization multiplexing in coherent optical systems and multilevel modulation formats such as M-ary quadrature amplitude modulation QAM enables us to

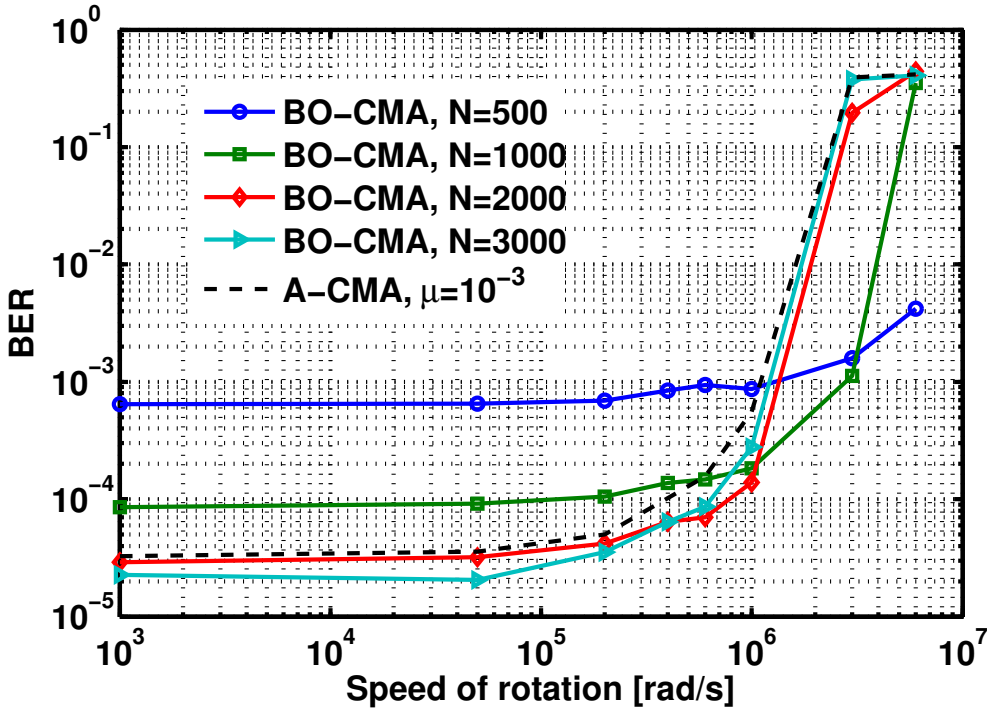


Figure 4.14: BER of the BO-CMA (with different block sizes N) and the A-CMA versus the rotation speed Ω of the polarization

increase the spectral efficiency and the transmission bit rate. The propagation linear impairments mitigation is ensured by digital signal processing (DSP), that compensates for residual Chromatic Dispersion (CD) and monitors polarization dependent effects such as Polarization Mode Dispersion (PMD) and Polarization Dependent Loss (PDL). Constant Modulus Algorithm (CMA) based Fractionally spaced equalization is widely used for Multiple-Input Multiple-Output (MIMO) equalization due to its simplicity and insensivity to the phase of the signal. However, it is well known that this equalizer may converge to the same output at both polarizations [59, 77, 78]. In order to solve this singularity issue, different methods were proposed which rely mainly on

i) the algorithm re-initialization after convergence by a well-chosen filter [59, 77],

ii) a cost function discriminating between the equalized signals [78],

iii) the complicated Independent Component Analysis (ICA) [79].

In Subsection 4.5.1, we mention more precisely the techniques currently used in optical communications which are based on approaches i) and ii). In Subsection 4.5.2, we propose to apply a well-known coming from the "blind source separation" community to our context. Finally Subsection 4.5.3 is devoted to simulations.

4.5.1 State-of-the-Art

4.5.1.1 Filter re-initialization based algorithm

The method exploits the fact that when DGD is the only source of distortion, the channel transfer matrix of the fiber is given by an unitary matrix in the frequency

domain. This means that

$$\tilde{\mathbf{C}}_{\mathrm{DGD}}(\omega) = \begin{bmatrix} u(\omega) & v(\omega) \\ -\overline{v(\omega)} & \overline{u(\omega)} \end{bmatrix}. \tag{4.23}$$

The inverse of the channel (actually the ZF equalizer) in the frequency domain, denoted by $\widetilde{\mathbf{W}}_{\mathrm{DGD}}(\omega) = \tilde{\mathbf{C}}^{-1}_{\mathrm{DGD}}(\omega)$, is also an unitary matrix and thus we have the following relationship between their components.

$$\begin{aligned} \widetilde{\mathbf{w}}_{2,2}(\omega) &= \overline{\widetilde{\mathbf{w}}_{1,1}(\omega)} \\ \widetilde{\mathbf{w}}_{2,1}(\omega) &= -\overline{\widetilde{\mathbf{w}}_{1,2}(\omega)}. \end{aligned} \tag{4.24}$$

Therefore, in the time domain, the impulse response of the ZF equalizer associated with the DGD, denoted by $\mathbf{W}_{a,\mathrm{DGD}}(t) = \mathrm{FT}^{-1}(\omega \mapsto \widetilde{\mathbf{W}}_{\mathrm{DGD}}(\omega))$ satisfies the following constraints

$$\begin{aligned} \mathbf{w}_{2,2}(t) &= \overline{\mathbf{w}_{1,1}(-t)} \\ \mathbf{w}_{2,1}(t) &= -\overline{\mathbf{w}_{1,2}(-t)}. \end{aligned} \tag{4.25}$$

This special relationship can be used to solve the singularity issue. Indeed, the algorithm works as follows:

1. Set the initial equalizer values $\mathbf{w}^0_{1,1} = [\cdots, 0, 1, 0, \cdots]$ and $\mathbf{w}^0_{1,2} = [\cdots, 0, 0, 0, \cdots]$.
2. Run CMA for output polarization 1 until reaching convergence.
3. Set the initial tap values for output polarization 2 according to Eq. (4.25) and values obtained in the previous step.
4. Continue CMA updating for both polarizations independently.

This relationship (and thus this algorithm) no longer holds if there are other sources of channel distortion such as CD and PDL which is the main drawback of this approach.

4.5.1.2 Penalized function based algorithm

The cost function of the CMA applied on both output polarizations can be [80]

$$J(\mathbf{w}_p, \mathbf{w}_q) = \mathbb{E}[(|z_1(n)|^2 - R)^2] + \mathbb{E}[(|z_2(n)|^2 - R)^2] \tag{4.26}$$

In order to avoid singularity, it has been proposed in [78] to add a penalty function ensuring deceleration between both output polarizations. This leads to

$$J_{\mathrm{pen.}}(\mathbf{w}_p, \mathbf{w}_q) = \mathbb{E}[(|z_1(n)|^2 - R)^2] + \mathbb{E}[(|z_2(n)|^2 - R)^2] + 2\sum_{p=1}^{2}\sum_{\delta=\delta_1}^{\delta_2} |r_{p,1-p}(\delta)|^2 \tag{4.27}$$

where the cross-correlation function between the equalizer outputs p and $q = 1 - p$ is given by

$$r_{p,q}(\delta) = \mathbb{E}[z_p(n)\overline{z_q(n-\delta)}] \tag{4.28}$$

Integers δ_1, δ_2 must be chosen such that the minimum and the maximum time shifts between the two signals lie within $[\delta_1, \delta_2]$.

Finally, the filter coefficients update is obtained as follows

$$\frac{\partial J_{\text{pen.}}}{\partial \overline{\mathbf{w}_p}} = 4(|z_p(n)|^2 - R)\overline{\mathbf{y}^{(L)}(n)} + 2\sum_{\delta=\delta_1}^{\delta_2} \overline{\hat{r}_{pq}(\delta) z_q(n-\delta)}\mathbf{y}^{(L)}(n)$$

$$\frac{\partial J_{\text{pen.}}}{\partial \overline{\mathbf{w}_q}} = 4(|z_q(n)|^2 - R)\overline{\mathbf{y}^{(L)}(n)} + 2\sum_{\delta=\delta_1}^{\delta_2} \hat{r}_{pq}(\delta)\overline{z_p(n)}\mathbf{y}^{(L)}(n)$$

where $\hat{r}_{pq}(\delta)$ is the empirical estimate of the correlation.

4.5.2 Proposed deflation based algorithm

Based on the principle of deflation introduced in [81] in the blind source separation context, we propose a new singularity-free algorithm suitable for QAM and PSK equalization. The algorithm extracts the two polarizations one by one. It is well suitable for "block" processing as done in [82] as well as for the standard sample-by-sample processing. It does not require a prior knowledge of the propagation channel characteristics unlike [59, 77].

The proposed deflation based CMA equalizer can be roughly described as follows: the idea behind deflation is to detect one transmit sequence (let us say on polarization 1 via any blind equalizer) and to re-construct the output component associated with this transmit sequence (this reconstruction can now be done since the transmit sequence on polarization 1 is known now and thus a Data-Aided-like channel estimator between polarization 1 and both output polarizations can be carried out). Then the contribution of the input polarization 1 on both output polarizations can be removed. As a consequence, it remains only the contribution of the input polarization 2 that can be detect once again by any blind equalizer. More precisely, we have

1. Set the initial equalizer values $\mathbf{w}_{1,1}^0 = [\cdots, 0, 1, 0, \cdots]$ and $\mathbf{w}_{1,2}^0 = [\cdots, 0, 0, 0, \cdots]$.

2. Run CMA until convergence is reached using the following stopping criterion

$$\frac{\|\mathbf{w}_1^{\ell+1} - \mathbf{w}_1^{\ell}\|}{\|\mathbf{w}_1^{\ell}\|} \leq \varepsilon \tag{4.29}$$

   where $\mathbf{w}_1^{\ell}$ is the equalizer obtained after the ℓ-th CMA iteration.

3. Calculate $z_1(n)$, compensate for frequency offset and phase (with any existing estimator), and make a decision $\hat{s}_1(n)$ on the transmit sequence $s_1(n)$.

4. Estimate the channel impulse responses $\mathbf{h}_{p,1} = [\mathbf{h}_{p,1}(0)^{\mathrm{T}}, \cdots, \mathbf{h}_{p,1}(K-1)^{\mathrm{T}}]^{\mathrm{T}}$ between input polarization 1 and output polarization p. Assuming white data sequence (even if not perfectly true), we have

$$\hat{\mathbf{h}}_{p,1} = \frac{1}{NP_0}\mathbf{T}_s^{\mathrm{H}}\mathbf{y}_p \tag{4.30}$$

   where P_0 is the mean transmit power,

$$\mathbf{y}_p = [\mathbf{y}_p(0)^{\mathrm{T}}, \mathbf{y}_p(1)^{\mathrm{T}}, \cdots, \mathbf{y}_p(N-1)^{\mathrm{T}}]^{\mathrm{T}}. \tag{4.31}$$

   and $\mathbf{T}_s$ is the following $2N \times 2K$ matrix

$$\mathbf{T}_s = \mathbf{S} \otimes \mathbf{I}_2 \tag{4.32}$$

with

$$\mathbf{S} = \begin{bmatrix} \hat{s}_1(0) & \hat{s}_1(-1) & \cdots & \hat{s}_1(-L_f+1) \\ \hat{s}_1(1) & \hat{s}_1(0) & \cdots & \hat{s}_1(-L_f) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{s}_1(N-1) & \hat{s}_1(N-2) & \cdots & \hat{s}_1(N-L_f) \end{bmatrix}. \tag{4.33}$$

and $\otimes$ is the Kronecker product.

Eq (4.30) is proven in Appendix B. We remind that the estimator given in Eq. (4.30) will perform well if $\mathbf{S}^{\mathrm{H}}\mathbf{S}$ is close to identity. Such an assumption holds here since $s_1(n)$ is an iid sequence.

5. Calculate the contribution of the input polarization 1 on the output polarizations 1 and 2. Thus

$$\hat{\mathbf{y}}_{p,1}(n) = \sum_{\ell=0}^{K-1} \hat{\mathbf{h}}_{p,1}(\ell)\hat{s}_1(n-\ell).$$

6. The previously-calculated contribution is subtracted at each polarization:

$$\tilde{\mathbf{y}}_p(n) = \mathbf{y}_p(n) - \hat{\mathbf{y}}_{p,1}(n). \tag{4.34}$$

Then run CMA until convergence is reached with initialization $\mathbf{w}_{2,2}^0 = [\cdots, 0, 1, 0, \cdots]$ and $\mathbf{w}_{2,1}^0 = [\cdots, 0, 0, 0, \cdots]$ in order to obtain the polarization 2.

The algorithm is summarized in Fig. 4.15.

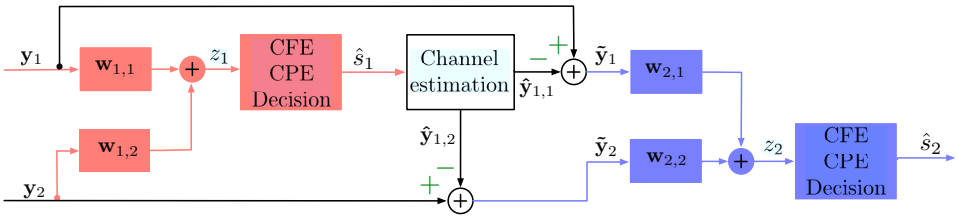


Figure 4.15: Structure of the deflation based equalizer.

4.5.3 Simulation results

A 112Gbit/s transmission is carried out by multiplexing both polarizations with 16-QAM leading to 14Gbaud transmission per polarization. At the transmitter and receiver sides, a square root raised cosine filters with a roll-off factor equal to 1 is used. A 5-th order Bessel electrical filter with a 3dB bandwidth equal to $0.8/T_s$ was included as the anti-aliasing filter. Neither phase noise nor frequency offset between the signal laser and local oscillator were considered. Finally the signal is sampled at twice the baud rate. The channel is a concatenation of CD, DGD, Polarization State Transformation (PST), and PDL. Therefore the channel frequency response is

$$\tilde{\mathbf{C}}(\omega) = \begin{bmatrix} \sqrt{1+\gamma} & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \times \begin{bmatrix} \cos(\theta) & \sin(\theta)e^{-i\phi} \\ -\sin(\theta)e^{i\phi} & \cos(\theta) \end{bmatrix} \times \begin{bmatrix} e^{i\omega\tau/2} & 0 \\ 0 & e^{-i\omega\tau/2} \end{bmatrix} \times e^{i\alpha\omega^2} \tag{4.35}$$

The attenuation γ is related to the PDL denoted by

$$\Gamma_{\text{dB}} = 10 \log_{10} \left(\frac{1+\gamma}{1-\gamma} \right). \tag{4.36}$$

In Figs. 4.16 and 4.17, we plot the BER versus θ and ϕ for the standard block (std.)-CMA [82] with optimal step size (on the left) and the deflation based block (def.)-CMA with optimal step-size proposed here (on the right). The block size is $N = 1000$. The equalizer length is $L = 3$. Two PDL values have been considered (0dB on Fig. 4.16 , 3dB on Fig. 4.17). One thousand blocks have been tested for each channel realization. The proposed deflation based algorithm enables us to cancel the singularity completely.

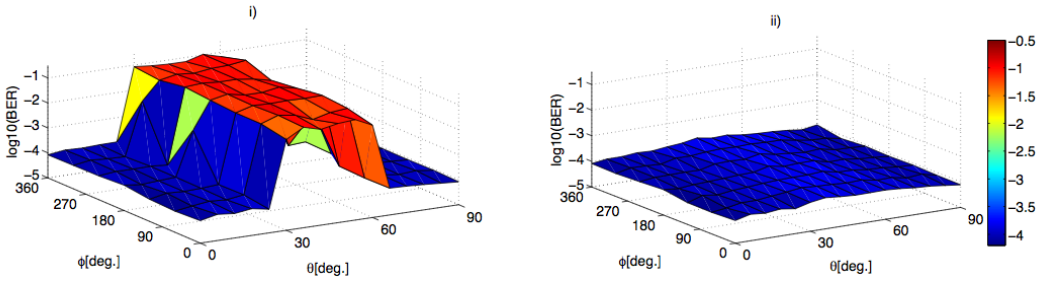


Figure 4.16: BER versus θ and ϕ (OSNR = 20dB, DL_f = 0ps/nm, τ_{DGD} = 0ps, $N = 1000$): iii) PDL = 0dB, std.-CMA, iv) PDL = 0dB, def.-CMA.

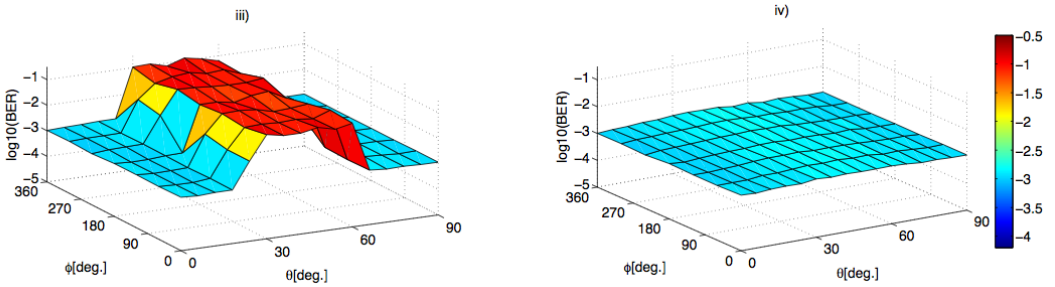


Figure 4.17: BER versus θ and ϕ (OSNR = 20dB, DL_f = 0ps/nm, τ_{DGD} = 0ps, $N = 1000$): iii) PDL = 3dB, std.-CMA, iv) PDL = 3dB, def.-CMA.

In Fig. 4.18, we display the BER versus residual CD for the std.-CMA and the def.-CMA for channel realizations exhibiting no singularity. Both CMA are applied on successive data block, and the CMA at block k is initialized by the equalizer obtained at block $k-1$. We show that the std.-CMA and the proposed CMA have the same performance and so any penalty is introduced in absence of singularity. In order to reach the stopping condition $\varepsilon = 5.10^{-3}$, In Fig. 4.19, we display the BER versus number of considered CMA iterations for the def. CMA and the std. CMA for channel realizations exhibiting no singularity. the number of iterations of both algorithms are very close. Nevertheless def.-CMA is a little more complex than std.-CMA due to the step 4. Step 4 has a complexity of $4NL_f$ = 12kflops whereas each iteration (for both CMA) needs

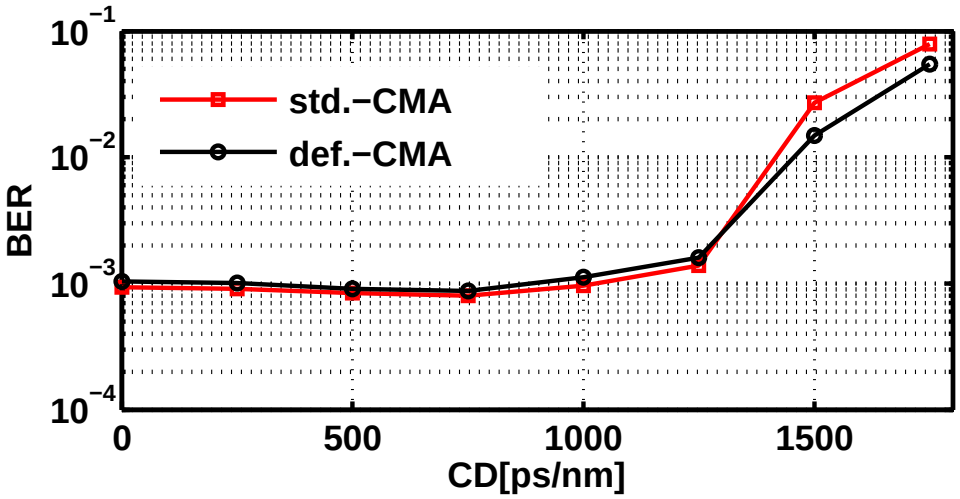


Figure 4.18: BER vs. residual CD (OSNR = 21dB, τ_{DGD} = 50ps, PDL = 3dB, $N = 1000$, $\theta = \phi = \pi/8$)

76kflops [83]. Moreover the def.-CMA may be applied in the first transmission block to ensure non-singularity and then one can move into the std.-CMA for the rest of blocks.

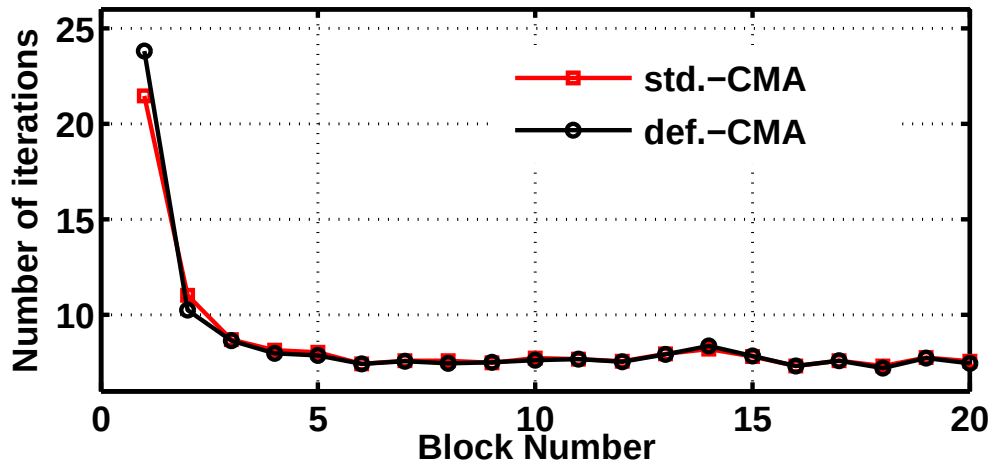


Figure 4.19: BER vs. #iterations of equalizers (OSNR = 21dB, DL_f = 500ps/nm, τ_{DGD} = 50ps, PDL = 3dB, $N = 1000$, $\theta = \phi = \pi/8$)

4.6 Experimental results

We will validate the proposed block-wise algorithms through experimental data. This enables us to investigate the effects that we did not take into account in our previous simulations, such as, non-linear effects or non-ideal signal generation. The experimental data have been obtained by using the testbed of the *Heinrich Hertz Institute (HHI)* in Berlin. In Section 4.6.1, we describe the experimental set-up. In Section 4.6.2, the experimental performance of the BO-CMA algorithm are analyzed and compared to the A-CMA for a POLMUX 8PSK transmission.

4.6.1 Experimental set-up

The experimental set-up, shown in Fig. 4.20, is based on an optical 8PSK transmitter at 10GBaud corresponding to a bit rate of 30Gbit/s. The optical modulated signal is then multiplexed in polarization by using a polarization beam splitter (PBS) and a polarization beam combiner (PBC). A delay line is inserted into one out of the two branches in order to decorrelate the two multiplexed streams. The total bit rate of the generated POLMUX 8PSK signal is thus 60Gbits/s. The transmission is performed through a recirculating loop which consists of one span of 80km of Standard Single Mode Fiber (SSMF) characterized by a cumulative dispersion of 1365ps/nm. The fiber loss is compensated for after each loop by using an Erbium-doped-fiber-amplifier (EDFA). A 5nm width filter is carried out in order to remove the out of band amplified spontaneous emission (ASE) noise. A second EDFA is used to control the injected power at the input of each span. At the receiver side, the POLMUX 8PSK signal is sent to a PBS whose outputs feed a 90^o hybrid device for each polarization. The same external cavity laser (ECL) is used for generating the 8PSK modulated signal and is shared by the local oscillator for both polarizations which implies that the frequency offset is zero. The spectral linewidth of the ECL is 100kHz which leads to a no significant phase noise level. The outputs of the two 90^o hybrid devices are converted with four balanced photo-diodes to generate the I and Q components for each polarization. Finally, these four signals are sampled by analog-to-digital converters at 50Gsamples/s which corresponds to 5 samples per symbol. The discrete-time data composed by 750000 samples, *i.e.*, 150000 symbols, are stored and processed offline. More details about the experimental set-up can be found in [84].

4.6.2 Experimental Performance

4.6.2.1 Single polarization transmission

In this subsection we consider the single polarization transmission. The DSP structure used to retrieve the transmitted sequence is the following one: The received signal has a rate of 5 samples per symbol, the signal is then resampled to obtain a rate of exactly 2 samples per symbol. Secondly, the CD or the residual CD is compensated for using FIR filters as explained in Chapter 1, and finally the Viterbi-Viterbi algorithm associated with the constant phase estimation is carried out to compensate for the phase noise of the laser using block of size 10.
Finally, the differentially encoded sequence is decoded and the BER is evaluated.

- *Back-To-Back performance*

The back-to-back performance is given in Fig. 4.21. For an OSNR higher than 14dB, the BER is kept below the FEC limit of 2.10^{-3}.

- *Inline CD compensation*

In this scenario, after each span of SSMF of length 80Km, the cumulative CD is compensated for using a Dispersion Compensation Fiber (DCF) of length 13km. Fig. 4.22 depicts the BER versus the transmit power at the input of the SSMF, and for different transmission distances ranging from 400 to 1200km.

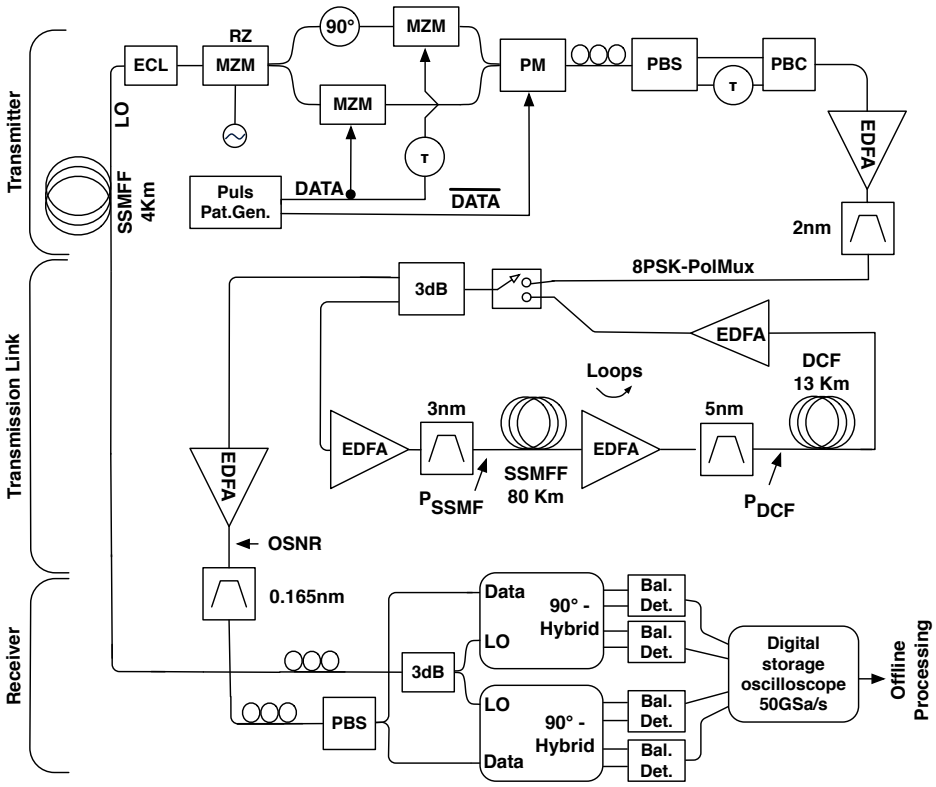


Figure 4.20: Measurement set-up of the POLMUX 8-PSK 60Gbit/s with offline digital signal processing. (PBC: Polarization Beam Combiner, PBS: Polarization Beam Splitter, MZM: Mach-Zehnder Modulator, ECL: External Cavity Laser, EDFA: Erbium Dopped Fiber Amplifier, PM: Polarization Multiplexing.)

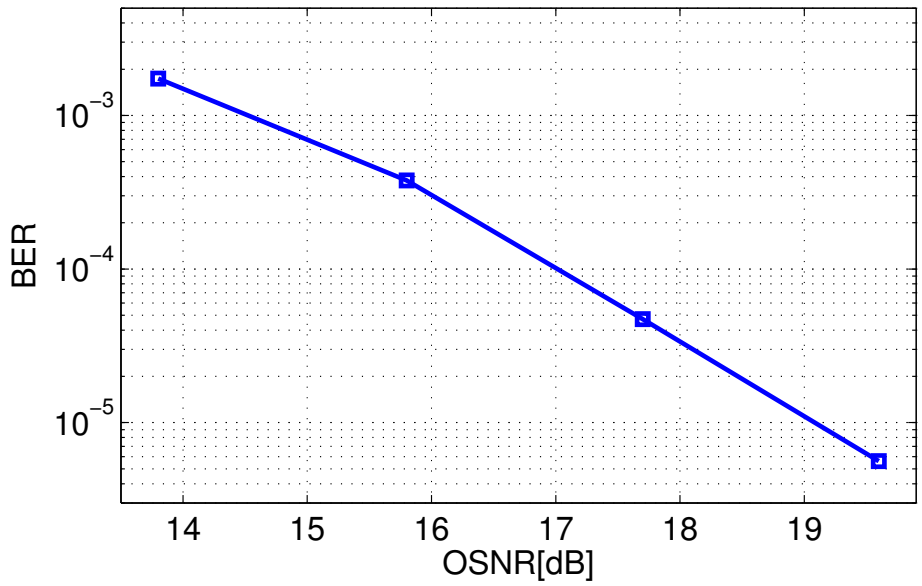


Figure 4.21: Back-to-Back performance for the 8-PSK.

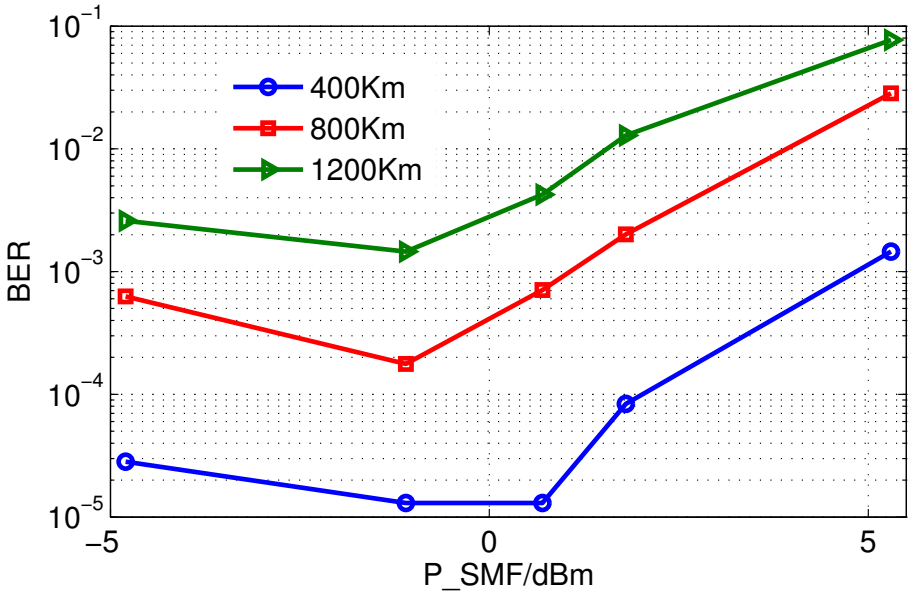


Figure 4.22: BER versus power at the input of the SMF fiber for different transmission distances.

The optimal input power at the SSMF is -1.1dBm. Below such input power value, the OSNR becomes low which degrades the performance of the system, and prevents to have a reasonable transmission distance. However, for high values of OSNR, the signal is degraded due to non-linear effects.
Fig. 4.23 shows the BER versus the transmission distance for an optimal transmission power at the input of SSMF. The system exhibits a BER lower than the FEC limit (2.10^{-3}) up to 1200km.

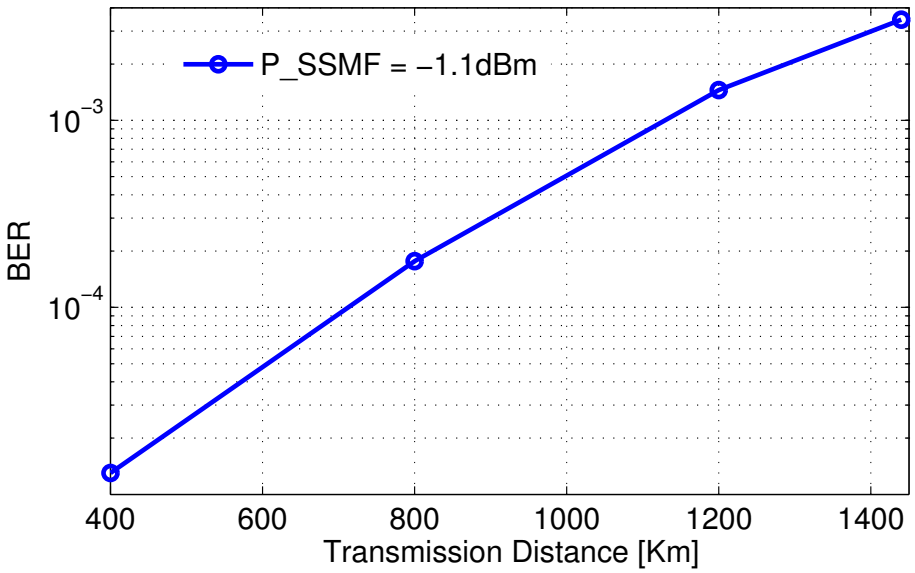


Figure 4.23: BER versus transmission distance for input power of -1.1dBm .

An illustration of the constellations of the signal before and after processing is presented in Fig. 4.24, for an input power at SSMF of -1.1dBm and a transmission distance of $L_f = 400$km.

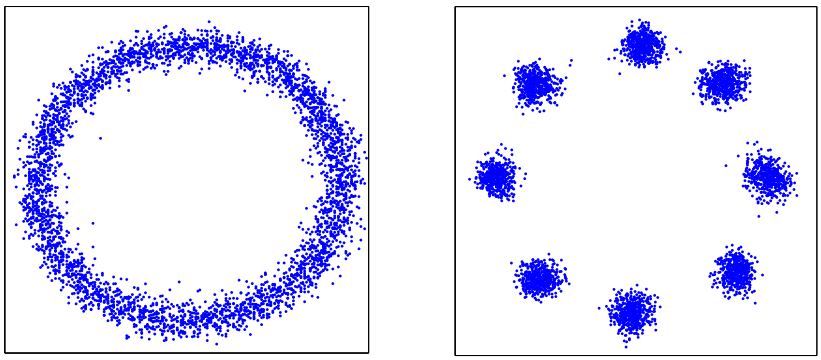


Figure 4.24: Constellations before and after processing using the Carrier phase recovery.

- *Offline CD compensation*

In this second scenario, all the cumulative CD is compensated for using an FIR filter whose coefficients are calculated according to Eq. (1.48) explicited in Chapter 1.
Fig. 4.25 depicts the BER versus the input power at the SSMF and for different transmission distances.

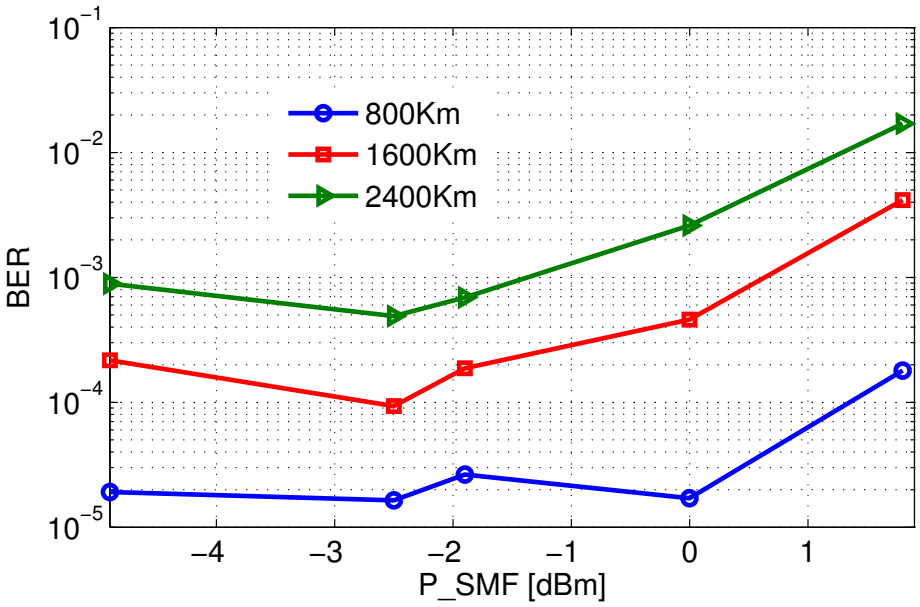


Figure 4.25: BER versus the input power for different transmission distances.

The optimal transmission distance is found at -2.5dBm. The BER versus the transmission distance for this optimal input power is given in Fig. 4.26. The BER is kept below the FEC limit up to 2400km.

This result confirms the fact that using coherent detection, in single carrier transmission, and neglecting the neighboring channel induced impairments, the uncompensated

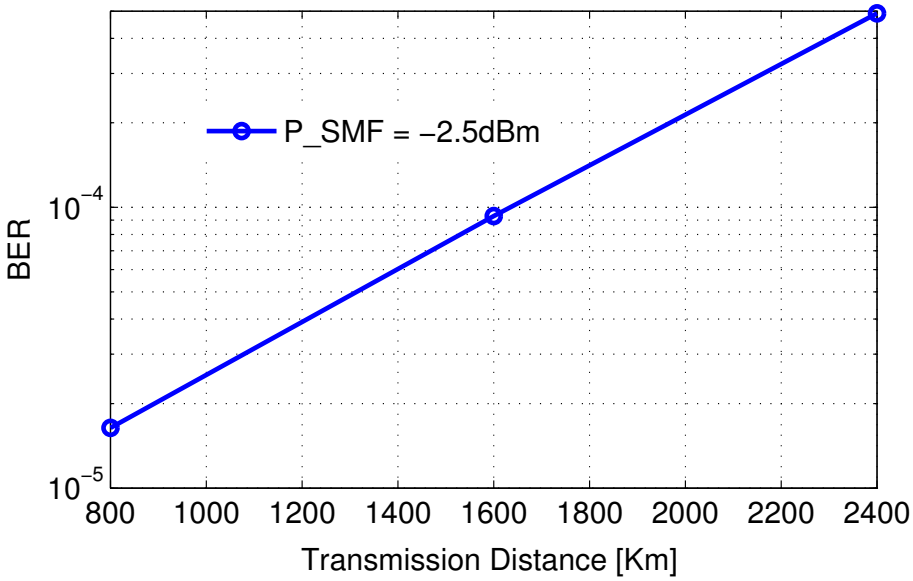


Figure 4.26: BER versus transmission distance of an input power at the SMF fiber -2.5dBm.

CD transmission is better than the compensated CD transmission scheme.
An illustration of the signal constellations before and after processing is given in Fig. 4.27 for an input power of -2.5dBm and a transmission distance of $L_f = 800$km.

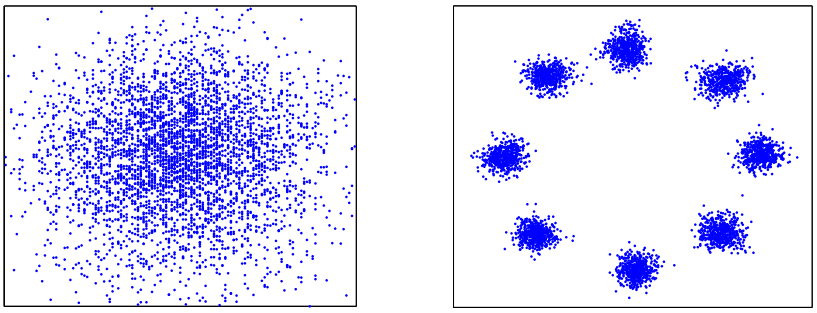


Figure 4.27: Constellations before and after processing compensation of CD in the electrical domain and carrier recovery.

An illustration of the compensating FIR for the accumulated CD over 800km of transmission is presented in Fig. 4.28.

4.6.2.2 POLMUX transmission

Except otherwise stated, we have considered a transmission over $L_f = 800$km, *i.e.*, 10 loops without inline CD compensation. The power at the input of each span was set to -0.9dBm. The cumulative CD (equal to 13650ps/nm for the 800km transmission) is partially compensated through a finite impulse response filter of length 512 [40], such as the residual cumulative CD is 1000ps/nm. On the one hand, this corresponds to practical situation when the fiber length is not perfectly known and, on the other

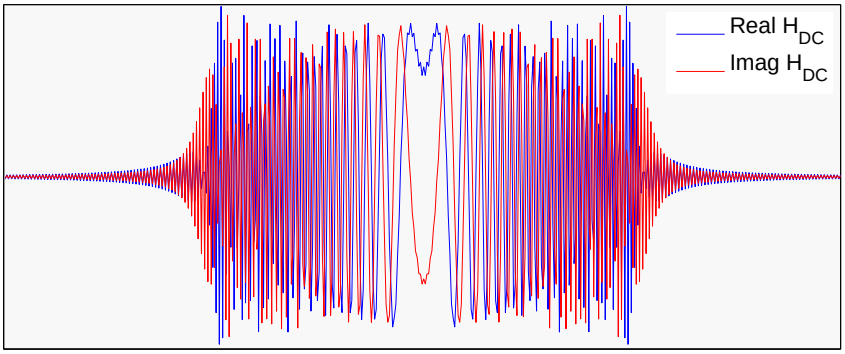


Figure 4.28: CD compensation filter coefficients real and imaginary parts.

hand, this enables us to exhibit the impact of residual CD on the proposed algorithms. The signal is then re-sampled in order to obtain exactly 2 samples per symbol. The proposed BO-CMA is finally used to compensate for the residual cumulative CD and the polarization dependent effects. As described in Section 1.3, we compute a $T_s/2$ FSE with $L = 3$, *i.e.*, $\mathbf{w}_1$ and $\mathbf{w}_2$ have 12 complex taps each. Furthermore, we have OSNR=23.7dB.

In Fig. 4.29, we plot the BER versus the number of iterations inside each block for different block sizes N. The BO-CMA is initialized with $\mathbf{w}^0$ and the BER is obtained

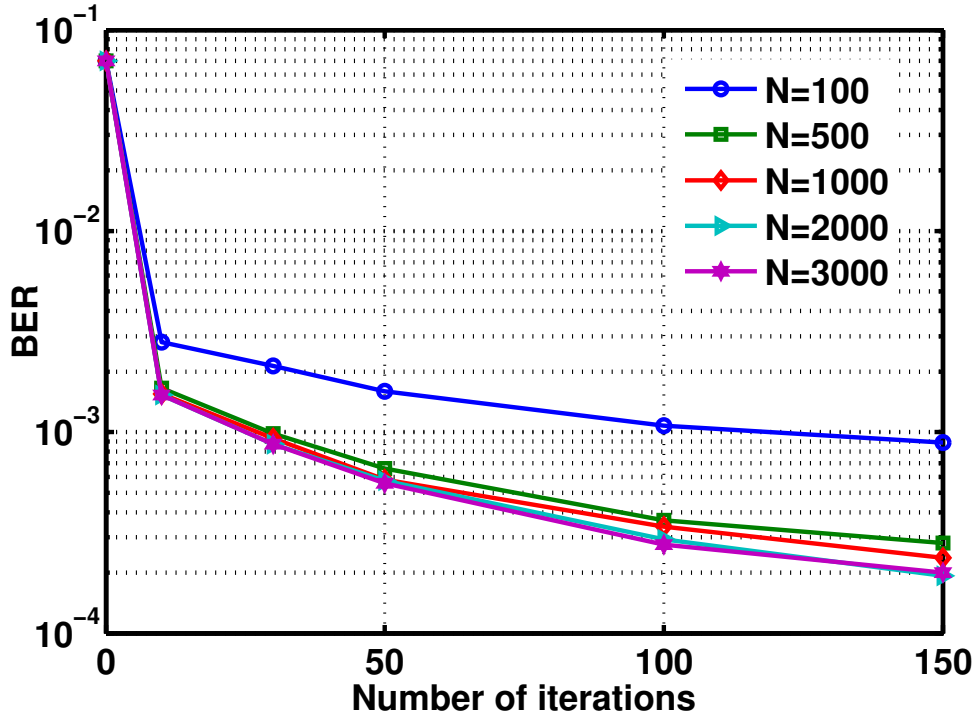


Figure 4.29: BER versus the number of iterations for various N. (8PSK, OSNR=23.7dB, residual CD of 1000ps/nm, transmission distance $L_f = 800$km.)

by averaging over at least 50 block observations. The target BER of 10^{-3} is obtained

with a reasonable number of iterations when data blocks are larger than 500. Unlike 16QAM (see Section 4.4.1), the BO-CMA with very small block size (*i.e.*, $N = 100$) offers a higher steady-state BER.

In Fig. 4.30, we plot the BER versus the observation window for BO-CMA, A-CMA and AN-CMA. We observe that, even for 8PSK, the A-CMA still needs about 5000 symbols (*i.e.*, 10000 samples) to converge.

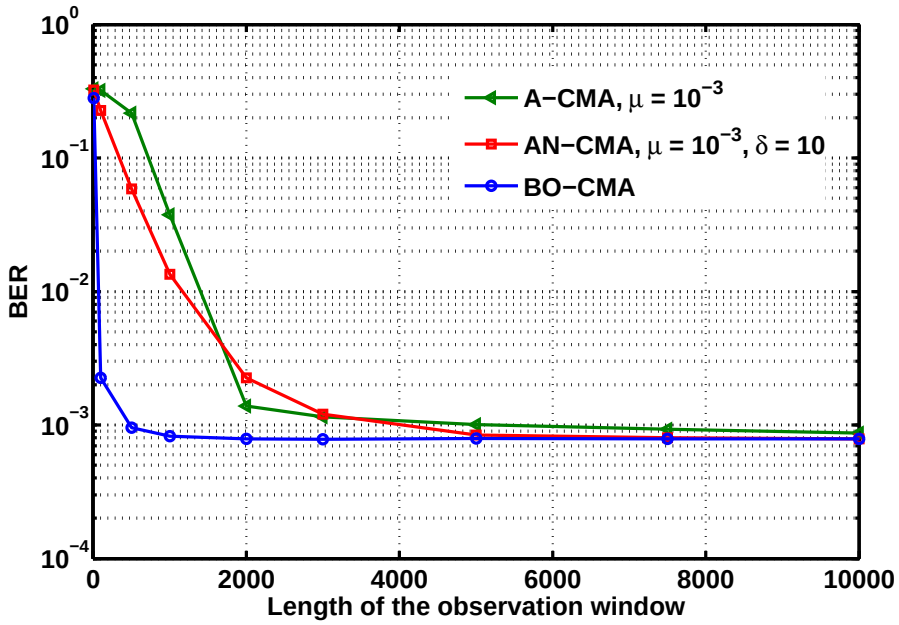


Figure 4.30: Speed of convergence for the different equalizers, $DL_f = 1000$ps/nm

In Fig. 4.31, we plot the BER versus the launched power at the input of each span for different block sizes N in order to study the influence of the intra-channel non-linear impairments. Along the data flow, the BO-CMA applied on the k-th block (of size N) is initialized with the equalizer provided by the $(k-1)$-th block (of size N). The equalizer provided by the BO-CMA (after a certain number of iterations) on the k-th block is only used on the k-th block. The number of iterations for each block is given by the stopping condition as explained in Section 4.4.1. Notice that the block size for the constant phase estimator has been fixed to 10 in order to be robust to the potential phase noise. As soon as the block size N is larger than 500, the steady-state of the BO-CMA is slightly better than that of the A-CMA (computed with $\mu = 10^{-3}$). Moreover, the BO-CMA is as robust to the non-linear effect as the A-CMA. In Fig. 4.32, we display the BER versus the residual CD. For the BO-CMA, we fix $N = 1000$. In order to handle high residual CD, the equalizer length is now increased to $L = 6$. We observe that the BO-CMA ensures slightly lower BER than the A-CMA. Even if the steady-state performance between the BO-CMA and the A-CMA are very close, we remind that the BO-CMA converges much faster than the A-CMA and thus is very well-adapted for bursty traffic mode as well as circuit mode.

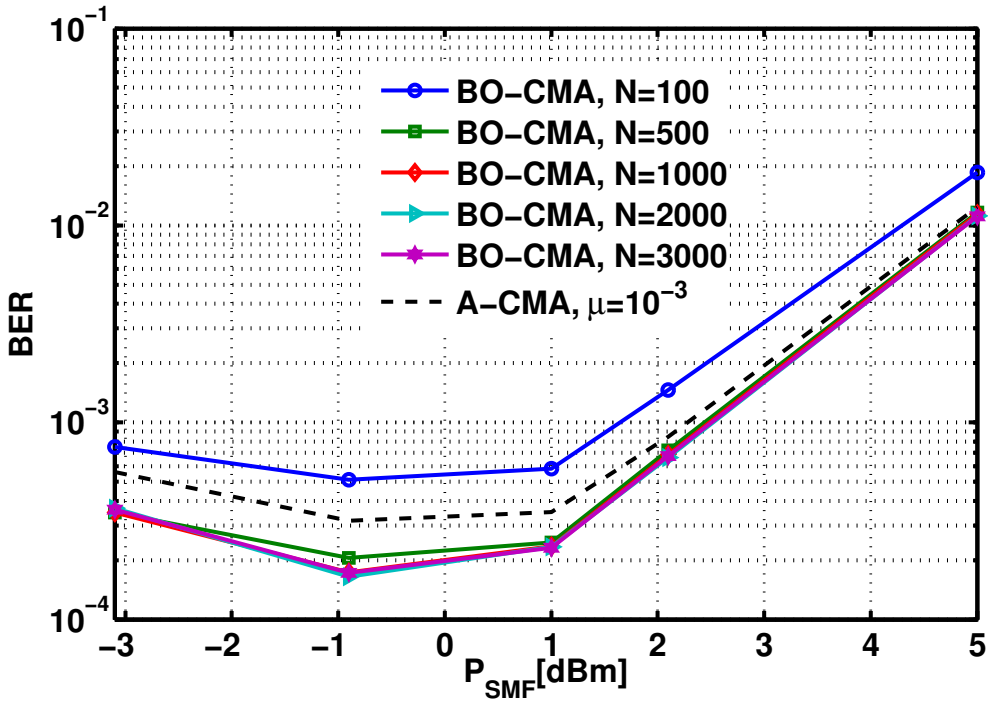


Figure 4.31: BER versus the power at the input of the SSMF for various N. (8PSK, residual CD of 1000ps/nm, transmission distance $L_f = 800$km.)

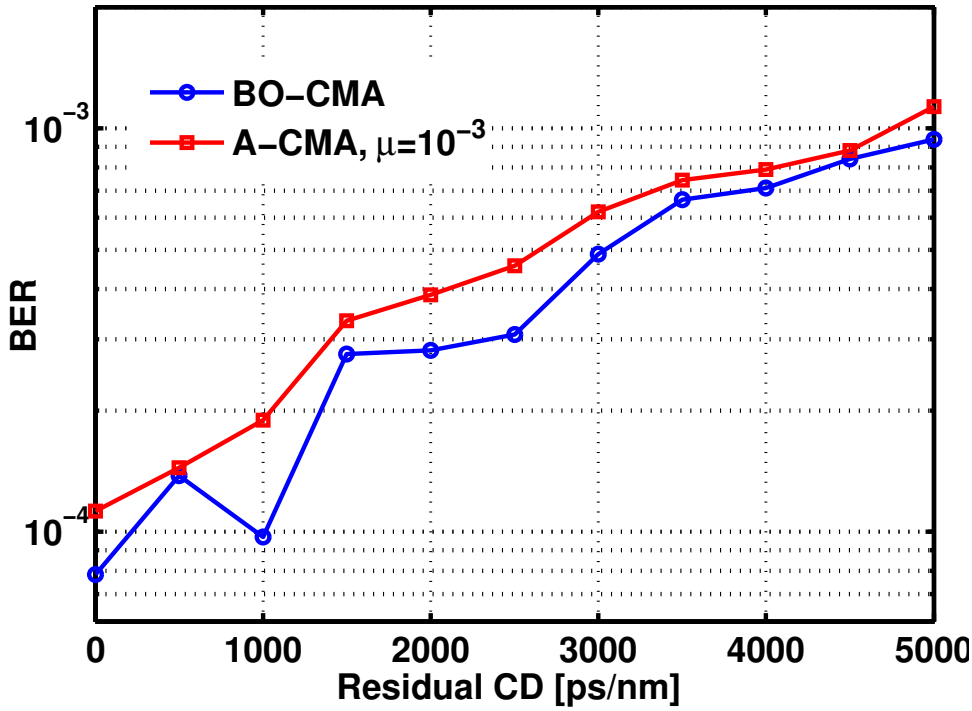


Figure 4.32: BER versus the residual CD (8PSK, OSNR=23.7dB, transmission distance $L_f = 800$km.)

4.7 Conclusion

In this chapter, we proposed The performance of different versions of block-wise CMA equalizer are investigated. Those versions include, fixed step size, or variable step size adapted using the inverse of the Hessian matrix or the optimal step-size. We showed that the observation window size required to converge for block-wise CMA approach is divided by ∼10 at the expense of an increase of the computational complexity by a factor ∼ 4. Moreover, the block-wise frequency offset estimation algorithm ensures low residual frequency offset. Finally, the block-wise digital signal processing enables us to relax the real-time implementation constraints on digital circuits running at some hundred of MHz and to offer a data throughput at a rate of tens of Gbaud. Those performances are validated through simulations and also through experimental data using a 60Gbit/s coherent optical system based on polarization multiplexing and RZ-8PSK modulation. Therefore the proposed algorithms are strong candidates for the next generation optical transmission systems.
In the last part of this chapter, we proposed a polarization demultiplexing algorithm that operates independently of the optical channel characteristics and that does not introduce OSNR penalty compared to the non singular case. this algorithm based on the deflation procedure induces a negligible increase of the computational complexity of the equalizer, and can be applied in the sample-by-sample equalization approach.

Conclusion and Perspectives

The work carried out in the thesis deals with the mitigation of the linear impairments in a PolMux QAM optical transmission. We have proposed several algorithms for improving the whole system. We remind here the main contributions.

- the adaptive equalizers have usually implemented through a constant step-size version of a gradient-descent algorithm. This unfortunately offers poor steady-state and slow convergence speed. We have proposed a non-constant step-size version of these algorithms. We have observed a significant improvement for the steady-state and the convergence speed at the expense of a slight additional complexity.

- the CFO due to mismatch between local oscillators can lead to very poor performance especially when high-oder QAM modulations are used. Here we have proposed a new algorithm based on the exploitation of both polarizations and the complete spectrum maximization. The introduced estimate yields impressive performance and almost enables to vanish the OSNR penalty due to the presence of the CFO.

- the last but not the least, we have developed block-wise version of adaptive equalizers. This modification actually modifies the paradigm in which we worked since the channel is viewed a constant over the processing duration. We have shown that the convergence speed is dramatically improved which enables to treat even small burst. Moreover the singularity issue is counter-acted through a modified version of the equalizers and is well adapted to singularity induced by PolMux as well as other types.

Evaluation of all the proposed algorithms has been done through extensive simulation and also some experimental data.

A number of issues, even in single-carrier context still deserves to be treated in the future. These problematics are listed below.

- a new way for increasing the data rate is to use simultaneously the different modes of the fiber. This induces additional degrees of freedom. Unfortunately, the different modes will interfere with each others. There, once again, a source separation/equalization issue occurs. This problem even if it is mathematically identical with the PolMux case is actually quite different since the number of modes is higher than two and the number of equalizer taps may be much larger.

- once the linear impairments are well treated, the main remaining problem is the nonlinear impairments such as nonlinear phase noise and Kerr effect. We would like to focus on the design of the receiver that takes into account the presence of such non-linear impairments, perhaps, by means of Volterra filters and other tools.

- in order to mitigate the PDL, Polar-Time codes can be added at the transmitter side. We thus do not send two independent streams of each polarization but a linear combination of these streams. The receiver has thus to be re-design to take into account the PDL filter matrix and also the structure of the Polar-Time codes. When the channel is frequency-selective and the transmission is single-carrier, the problem is far from being simple and works have to be done in this way.

Appendix A

Appendix related to Chapter 1

A.1 CMA derivations

As the filter $\mathbf{w}_p$ is complex-valued, the update equation is as follows

$$\mathbf{w}_{p,n+1} = \mathbf{w}_{p,n} - \mu \nabla J_{\text{CMA},n}(\mathbf{w}_p)|_{\mathbf{w}_{p,n}} \tag{A.1}$$

where

$$\nabla J_{\text{CMA},n}(\mathbf{w}_p)|_{\mathbf{w}_{p,n}} = \frac{\partial J_{\text{CMA},n}}{\partial \overline{\mathbf{w}}_p}|_{\mathbf{w}_{p,n}}.$$

The derivative function has to be done with respect to $\overline{\mathbf{w}}_p$ assuming the variable $\mathbf{w}_p$ fixed.

We remind that $z_p(n) = \overline{\mathbf{w}}_p^{\mathrm{T}} \mathbf{y}^{(L)}(n)$ and $J_{\text{CMA},n}(\mathbf{w}_p) = (|z_p(n)|^2 - R)^2$. Therefore, we obtain that

$$\nabla J_{\text{CMA},n} \propto (|z_p(n)|^2 - R)\overline{z_p(n)}\mathbf{y}^{(L)}(n). \tag{A.2}$$

A.2 Phase derivations

Concerning the blind algorithm, one can easily check that the (block-wise) algorithm is associated with the minimization of the following cost function

$$J_{\text{blind}}(\varphi) = \mathbb{E}[J_{\text{blind},n}(\varphi)] \tag{A.3}$$

with

$$J_{\text{blind},n}(\varphi) = |v_p(n)^Q - e^{2i\pi Q\varphi}|^2. \tag{A.4}$$

The adaptive version of the minimization of Eq. (A.3) can be done as follows

$$\hat{\varphi}_{0,p,n+1} = \hat{\varphi}_{0,p,n} - \mu_{\text{blind}} \frac{\partial J_{\text{blind},n}(\varphi)}{\partial \varphi}|_{\hat{\varphi}_{0,p,n}} \tag{A.5}$$

The derivative function $\partial J_{\text{blind},n}(\varphi)/\partial\varphi$ can be expressed as follows

$$\frac{\partial J_{\text{blind},n}(\varphi)}{\partial \varphi} \propto -\Im[v_p(n)^Q e^{-2i\pi\varphi}]$$

which eansl us to obtaine Eq. (1.67).

Concerning the DD algorithm, the estimate obtained in Eq. (1.66) is actually the result of the minimization of the following cost function

$$J_{\text{DD}}(\varphi) = \mathbb{E}[J_{\text{DD},n}(\varphi)] \tag{A.6}$$

with

$$J_{\mathrm{DD},n}(\varphi) = |v_p(n) - \hat{s}_p(n)e^{2i\pi\varphi}|^2. \tag{A.7}$$

This is actually a Least-Square (LS) criterion based on the received signal.

The (block-wise) minimization of this cost function leads to Eq. (1.66). In contrast, the (adaptive) minimization of this cost function can be done as follows

$$\hat{\varphi}_{0,p,n+1,\mathrm{DD}} = \hat{\varphi}_{0,p,n,\mathrm{DD}} - \mu_{\mathrm{DD}} \frac{\partial J_{\mathrm{DD},n}(\varphi)}{\partial \varphi}|_{\hat{\varphi}_{0,p,n,\mathrm{DD}}} \tag{A.8}$$

The derivative function $\partial J_{\mathrm{DD},n}(\varphi)/\partial\varphi$ can be easily derived and leads to Eq. (1.68).

Appendix B

Appendix related to Chapter 4

B.1 Derivations of Eq. (4.30)

The received signal at polarization p can be written as:

$$\mathbf{y}_p = \mathbf{T}_s \mathbf{h}_{p,1} + \mathbf{b}'_p \tag{B.1}$$

where $\mathbf{b}'_p$ is assumed to be an additive white Gaussian noise.
The maximum likelihood estimate is defined as

$$\hat{\mathbf{h}}_{p,1} = \arg\max_{\mathbf{h}_{p,1}} p(\mathbf{y}_p|\mathbf{h}_{p,1}) \tag{B.2}$$

where we have

$$p(\mathbf{y}_p|\mathbf{h}_{p,1}) \propto e^{-\frac{\|\mathbf{y}_p - \mathbf{T}_s\mathbf{h}_{p,1}\|^2}{2N_0}} \tag{B.3}$$

with N_0 the AWGN power spectrum density.

The maximization in Eq. B.2 boils down to the following minimization

$$\hat{\mathbf{h}}_{p,1} = \arg\min_{\mathbf{h}_{p,1}} \|\mathbf{y}_p - \mathbf{T}_s\mathbf{h}_{p,1}\|^2 \tag{B.4}$$

Then the solution of this minimization is given as

$$\hat{\mathbf{h}}_{p,1} = \mathbf{T}_s^{\#}\mathbf{y}_{p,1} = (\mathbf{T}_s^{\mathrm{H}}\mathbf{T}_s)^{-1}\mathbf{T}_s^{\mathrm{H}}\mathbf{y}_{p,1} \tag{B.5}$$

with $\mathbf{T}_s^{\#}$ is the Pseudo-Inverse of $\mathbf{T}_s$.

When the transmit sequence is i.d.d., we have $\mathbf{T}_s^{\mathrm{H}}\mathbf{T}_s = \mathbf{Id}$ which leads to Eq. (4.30).

Bibliography

[1] P. J. Winzer and R.-J. Essiambre, "Advanced modulation formats for high-capacity optical transport networks," *J. Lightwave Technol.*, vol. 24, pp. 4711–4728, Dec 2006.

[2] http://www.cisco.com

[3] R.-J. Essiambre, G. Kramer, P. Winzer, G. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Lightwave Technology, Journal of*, vol. 28, pp. 662 –701, feb.15, 2010.

[4] A. Gnauck, G. Charlet, P. Tran, P. Winzer, C. Doerr, J. Centanni, E. Burrows, T. Kawanishi, T. Sakamoto, and K. Higuma, "25.6-tb/s wdm transmission of polarization-multiplexed rz-dqpsk signals," *Lightwave Technology, Journal of*, vol. 26, pp. 79 –84, jan.1, 2008.

[5] http://www.kylia.com

[6] S. Namiki and Y. Emori, "Ultrabroad-band raman amplifiers pumped and gain-equalized by wavelength-division-multiplexed high-power laser diodes," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 7, pp. 3 –16, jan/feb 2001.

[7] K. Roberts, D. Beckett, D. Boertjes, J. Berthold, and C. Laperle, "100g and beyond with digital coherent signal processing," *Communications Magazine, IEEE*, vol. 48, pp. 62 –69, july 2010.

[8] T. Strasser and J. Wagener, "Wavelength-selective switches for roadm applications," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 16, pp. 1150 –1157, sept.-oct. 2010.

[9] E. Pincemin, "Challenges de la transmission wdm à ultra haut débit pour les réseaux de transport longue distance," tech. rep., France Telecom, Orange Labs, 23 Oct. 2008.

[10] G. P. Agrawal, *Fiber-Optic Communication Systems.* John Wiley & Sons, 3rd ed., 2002.

[11] K. Kikuchi, "Coherent optical communication systems," pp. 95 – 129, 2008.

[12] O. DeLange, "Wide-band optical communication systems: Part ii-frequency-division multiplexing," *Proceedings of the IEEE*, pp. 1683 – 1690, oct.

[13] T. Okoshi and K. Kikuchi, "Frequency stabilisation of semiconductor lasers for heterodyne-type optical communication systems," *Electronics Letters*, vol. 16, pp. 179 –181, 28 1980.

[14] F. Favre and D. le Guen, "High frequency stability of laser diode for heterodyne communication systems," *Electronics Letters*, vol. 16, pp. 709 –710, 28 1980.

[15] L. G. Kazovsky, G. Kalogerakis, and W. T. Shaw, "Homodyne phase-shift-keying systems: Past challenges and future opportunities," *Lightwave Technology, Journal of*, vol. 24, pp. 4876–4884, Dec. 2006.

[16] A. Gnauck, P. Winzer, C. Doerr, and L. Buhl, "10 x00d7; 112-gb/s pdm 16-qam transmission over 630 km of fiber with 6.2-b/s/hz spectral efficiency," pp. 1 –3, 22-26 2009.

[17] Y. Mori, C. Zhang, K. Igarashi, K. Katoh, and K. Kikuchi, "Unrepeated 200-km transmission of 40-gbit/s 16-qam signals using digital coherent optical receiver," pp. 1–2, July 2008.

[18] Y. Mori, C. Zhang, M. Usui, K. Igarashi, K. Katoh, and K. Kikuchi, "200-km transmission of 100-gbit/s 32-qam dual-polarization signals using a digital coherent receiver," in *Optical Communication, 2009. ECOC '09. 35th European Conference on*, pp. 1 –2, sept. 2009.

[19] M. Nakazawa, S. Okamoto, T. Omiya, K. Kasai, and M. Yoshida, "256-qam (64 gb/s) coherent optical transmission over 160 km with an optical bandwidth of 5.4 ghz," *Photonics Technology Letters, IEEE*, vol. 22, no. 3, pp. 185 –187, 2010.

[20] P. Winzer, A. Gnauck, S. Chandrasekhar, S. Draving, J. Evangelista, and B. Zhu, "Generation and 1,200-km transmission of 448-gb/s etdm 56-gbaud pdm 16-qam using a single i/q modulator," in *Optical Communication (ECOC), 2010 36th European Conference and Exhibition on*, pp. 1 –3, 2010.

[21] W. Shieh, X. Yi, Y. Ma, and Q. Yang, "Coherent optical ofdm: has its time come? (invited)," *J. Opt. Netw.*, vol. 7, pp. 234–255, Mar 2008.

[22] S. Jansen, "Optical ofdm, a hype or is it for real?," in *Optical Communication, 2008. ECOC 2008. 34th European Conference on*, p. 1, sept. 2008.

[23] *Understanding Optical Communications*. www.redbooks.ibm.com/redbooks/.

[24] E. Ip and J. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *Lightwave Technology, Journal of*, vol. 25, pp. 2033–2043, Aug. 2007.

[25] S. Savory, "Digital coherent optical receivers: Algorithms and subsystems," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 16, no. 5, pp. 1164 –1179, 2010.

[26] S. Salaun, F. Neddam, J. Poirrier, B. Raguenes, and M. Moignard, "Fast sop variation measurement on wdm systems are the opmdc fast enough?," in *Optical Communication, 2009. ECOC '09. 35th European Conference on*, pp. 1–2, Sept. 2009.

[27] H. Bulow, "Pmd mitigation techniques and their effectiveness in installed fiber," in *Optical Fiber Communication Conference, 2000*, vol. 3, pp. 110 –112 vol.3, 2000.

[28] C. D. Poole, "Statistical treatment of polarization dispersion in single-mode fiber," *Opt. Lett.*, vol. 13, no. 8, pp. 687–689, 1988.

[29] G. Foschini and C. Poole, "Statistical theory of polarization dispersion in single mode fibers," *Lightwave Technology, Journal of*, vol. 9, pp. 1439 –1456, nov 1991.

[30] K. Kikuchi, "Electronic post-compensation for nonlinear phase fluctuations in a 1000-km 20-gbit/s optical quadrature phase-shift keying transmission system using the digital coherent receiver," *Opt. Express*, vol. 16, no. 2, pp. 889–896, 2008.

[31] I. Roudas, M. Sauer, J. Hurley, Y. Mauro, and S. Raghavan, "Compensation of coherent dqpsk receiver imperfections," in *IEEE/LEOS Summer Topical Meetings, 2007 Digest of the*, pp. 19 –20, july 2007.

[32] I. Fatadin, S. Savory, and D. Ives, "Compensation of quadrature imbalance in an optical qpsk coherent receiver," *Photonics Technology Letters, IEEE*, vol. 20, pp. 1733 –1735, oct.15, 2008.

[33] http://www.fujitsu.com

[34] http://www.altera.com

[35] T. Pfau, S. Hoffmann, and R. Noé, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for m-qam constellations," *J. Lightwave Technol.*, vol. 27, no. 8, pp. 989–999, 2009.

[36] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, "Coherent detection in optical fiber systems," *Opt. Express*, vol. 16, no. 2, pp. 753–791, 2008.

[37] D. Tse and P. Viswanath, *Fundamentals of wireless communications*. Cambridge University Press, May, 2005.

[38] J. Proakis, *Digital Communications*. Mc-Graw-Hill, 2001.

[39] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express*, vol. 16, no. 2, pp. 804–817, 2008.

[40] S. Savory, "Compensation of fibre impairments in digital coherent systems," in *Optical Communication, 2008. ECOC 2008. 34th European Conference on*, pp. 1–1, Sept. 2008.

[41] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *Communications, IEEE Transactions on*, vol. 28, pp. 1867–1875, Nov 1980.

[42] H. Louchet, K. Kuzmin, and A. Richter, "Improved dsp algorithms for coherent 16-qam transmission," pp. 1–2, Sept. 2008.

[43] I. Fatadin, D. Ives, and S. Savory, "Blind equalization and carrier phase recovery in a 16-qam optical coherent system," *Lightwave Technology, Journal of*, vol. 27, pp. 3042–3049, Aug.1, 2009.

[44] M. Ready and R. Gooch, "Blind equalization based on radius directed adaptation," in *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, pp. 1699 –1702 vol.3, apr 1990.

[45] A. Leven, N. Kaneda, U.-V. Koc, and Y.-K. Chen, "Frequency estimation in intradyne reception," *Photonics Technology Letters, IEEE*, vol. 19, pp. 366–368, March15, 2007.

[46] A. Viterbi, "Nonlinear estimation of psk-modulated carrier phase with application to burst digital transmission," *Information Theory, IEEE Transactions on*, vol. 29, pp. 543 – 551, July 1983.

[47] D.-S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi, "Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation," *Lightwave Technology, Journal of*, vol. 24, pp. 12–21, Jan. 2006.

[48] Y. Wang, E. Serpedin, and P. Ciblat, "Optimal blind nonlinear least-squares carrier phase and frequency offset estimation for burst qam modulations," in *Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on*, vol. 2, pp. 1499 –1503 vol.2, 2001.

[49] Y. Wang, E. Serpedin, and P. Ciblat, "Optimal blind nonlinear least-squares carrier phase and frequency offset estimation for general qam modulations," *Wireless Communications, IEEE Transactions on*, vol. 2, no. 5, pp. 1040 – 1054, 2003.

[50] L. He, M. Amin, J. Reed, C., and R. Malkemes, "A hybrid adaptive blind equalization algorithm for qam signals in wireless communications," *Signal Processing, IEEE Transactions on*, vol. 52, no. 7, pp. 2058 – 2069, 2004.

[51] P. Winzer, A. Gnauck, C. Doerr, M. Magarini, and L. Buhl, "Spectrally efficient long-haul optical networking using 112-gb/s polarization-multiplexed 16-qam," *Lightwave Technology, Journal of*, vol. 28, pp. 547 –556, feb.15, 2010.

[52] W.-J. Zeng, X.-L. Li, and X.-D. Zhang, "Adaptive newton algorithms for blind equalization using the generalized constant modulus criterion," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 0, pp. 2805–2808, 2009.

[53] M. Seimetz, "Multi-format transmitters for coherent optical m-psk and m-qam transmission," vol. 2, pp. 225–229 Vol. 2, July 2005.

[54] L. Li, Z. Tao, S. Oda, T. Hoshida, and J. Rasmussen, "Wide-range, accurate and simple digital frequency offset compensator for optical coherent receivers," pp. 1 –3, 24-28 2008.

[55] M. Morelli and U. Mengali, "Feedforward frequency estimation for PSK : a tutorial review," *European Transactions on Telecommunications*, vol. 9, pp. 103–115, March 1998.

[56] P. Ciblat and M. Ghogho, "Blind nlls carrier frequency-offset estimation for qam, psk, and pam modulations: performance at low snr," *Communications, IEEE Transactions on*, vol. 54, pp. 1725–1730, Oct. 2006.

[57] U. Mengali and A. D'Andrea, *Synchronization techniques for digital receivers.* Plenum press, 1997.

[58] P. Ciblat, P. Loubaton, E. Serpedin, and G. Giannakis, "Performance analysis of blind carrier frequency offset estimators for noncircular transmissions through frequency-selective channels," *Signal Processing, IEEE Transactions on*, vol. 50, pp. 130 –140, Jan. 2002.

[59] C. Xie and S. Chandrasekhar, "Two-stage constant modulus algorithm equalizer for singularity free operation and optical performance monitoring in optical coherent receiver," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pp. 1 –3, 2010.

[60] X. Zhou and J. Yu, "Multi-level, multi-dimensional coding for high-speed and high-spectral-efficiency optical transmission," *Lightwave Technology, Journal of*, vol. 27, no. 16, pp. 3641 –3653, 2009.

[61] M. Seimetz, "High spectral efficiency phase and quadrature amplitude modulation for optical fiber transmission - configurations, trends, and reach," in *Optical Communication, 2009. ECOC '09. 35th European Conference on*, pp. 1 –4, 2009.

[62] C. Fludger, T. Duthel, D. van den Borne, C. Schulien, E.-D. Schmidt, T. Wuth, J. Geyer, E. De Man, K. Giok-Djan, and H. de Waardt, "Coherent equalization and polmux-rz-dqpsk for robust 100-ge transmission," *Lightwave Technology, Journal of*, vol. 26, pp. 64–72, Jan.1, 2008.

[63] K. Fukuchi, D. Ogasahara, J. Hu, T. Takamichi, T. Koga, M. Sato, E. de Gabory, Y. Hashimoto, T. Yoshihara, W. Maeda, J. Abe, T. Kwok, Y. Huang, K. Hosokawa, Y. Yano, M. Shigihara, Y. Ueki, Y. Saito, Y. Nomiyama, K. Kikuchi, A. Noda, S. Shioiri, M. Arikawa, T. Wang, and T. Tajima, "112gb/s optical transponder with pm-qpsk and coherent detection employing parallel fpga-based real-time digital signal processing, fec and 100gbe ethernet interface," in *Optical Communication (ECOC), 2010 36th European Conference and Exhibition on*, pp. 1 –3, 2010.

[64] C. Fludger, J. Geyer, T. Duthel, S. Wiese, and C. Schulien, "Real-time prototypes for digital coherent receivers," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pp. 1 –3, 2010.

[65] H. Sun, K.-T. Wu, and K. Roberts, "Real-time measurements of a 40 gb/s coherent system," *Opt. Express*, vol. 16, pp. 873–879, Jan 2008.

[66] E. Ip and J. Kahn, "Fiber impairment compensation using coherent detection and digital signal processing," *Lightwave Technology, Journal of*, vol. 28, no. 4, pp. 502 –519, 2010.

[67] J. Renaudier, G. Charlet, M. Salsi, O. Pardo, H. Mardoyan, P. Tran, and S. Bigo, "Linear fiber impairments mitigation of 40-gbit/s polarization-multiplexed qpsk by digital processing in a coherent receiver," *Lightwave Technology, Journal of*, vol. 26, no. 1, pp. 36 –42, 2008.

[68] M. Seimetz, "Performance of coherent optical square-16-qam-systems based on iq-transmitters and homodyne receivers with digital phase estimation," in *Optical Fiber Communication Conference, 2006 and the 2006 National Fiber Optic Engineers Conference. OFC 2006*, p. 10 pp., 2006.

[69] J. Renaudier, O. Bertran-Pardo, G. Charlet, M. Salsi, M. Bertolini, P. Tran, H. Mardoyan, and S. Bigo, "Investigation on wdm nonlinear impairments arising from the insertion of 100-gb/s coherent pdm-qpsk over legacy optical networks," *Photonics Technology Letters, IEEE*, vol. 21, no. 24, pp. 1816 –1818, 2009.

[70] M. Selmi, P. Ciblat, Y. Jaouen, and C. Gosset, "Pseudo-newton based equalization algorithms for qam coherent optical systems," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pp. 1 –3, 21-25 2010.

[71] 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," in *Optical Fiber Communication Conference, 1999, and the International Conference on Integrated Optics and Optical Fiber Communication. OFC/IOOC '99. Technical Digest*, vol. 2, pp. 83 –85 vol.2, 1999.

[72] M. Taylor, "Algorithms for coherent detection," in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pp. 1 –24, 2010.

[73] S. Houcke and A. Chevreuil, "Characterization of the undesirable global minima of the godard cost function: case of noncircular symmetric signals," *Signal Processing, IEEE Transactions on*, vol. 54, pp. 1917 – 1922, May 2006.

[74] V. Zarzoso and P. Comon, "Optimal step-size constant modulus algorithm," *Communications, IEEE Transactions on*, vol. 56, pp. 10–13, January 2008.

[75] L. Mazet, P. Ciblat, and P. Loubaton, "Fractionally spaced blind equalization: Cma versus second order based methods," pp. 231 –234, 1999.

[76] V. Zarzoso and P. Comon, "Optimal step-size constant modulus algorithm," 2004.

[77] L. Liu, Z. Tao, W. Yan, S. Oda, T. Hoshida, and J. C. Rasmussen, "Initial tap setup of constant modulus algorithm for polarization de-multiplexing in optical coherent receivers," in *Optical Fiber Communication Conference*, p. OMT2, Optical Society of America, 2009.

[78] A. Vgenis, C. Petrou, C. Papadias, I. Roudas, and L. Raptis, "Nonsingular constant modulus equalizer for pdm-qpsk coherent optical receivers," *Photonics Technology Letters, IEEE*, vol. 22, no. 1, pp. 45 –47, 2010.

[79] H. Zhang, Z. Tao, L. Liu, S. Oda, T. Hoshida, and J. Rasmussen, "Polarization demultiplexing based on independent component analysis in optical coherent receivers," in *Optical Communication, 2008. ECOC 2008. 34th European Conference on*, pp. 1 –2, sept. 2008.

[80] C. Papadias and A. Paulraj, "A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays," *Signal Processing Letters, IEEE*, vol. 4, pp. 178 –181, jun 1997.

[81] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: a second-order stable algorithm for the general case," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol. 47, pp. 1056 –1071, jul 2000.

[82] M. Selmi, P. Ciblat, Y. Jaouën, and C. Gosset, "Block versus adaptive mimo equalization for coherent polmux qam transmission systems," in *Optical Communication (ECOC), 2010 36th European Conference and Exhibition on*, pp. 1 –3, sept. 2010.

[83] M. Selmi, P. Ciblat, Y. Jaouen, and C. Gosset, "Complexity analysis of block equalization approach for polmux qam coherent systems," in *Signal Processing in Photonics Communications (SPPCom) Advanced Photonics Congress, 2011*, 2011.

[84] M. Seimetz, L. Molle, M. Gruner, D.-D. Gross, and R. Freund, "Transmission reach attainable for single-polarization and polmux coherent star 16qam systems in comparison to 8psk and qpsk at 10gbaud," in *Optical Fiber Communication - incudes post deadline papers, 2009. OFC 2009. Conference on*, pp. 1 –3, 2009.