Universitat Politècnica de Catalunya

Escola Tècnica Superior de
Telecomunicacions de Barcelona

Projecte Final de Carrera

Enginyeria de Telecomunicacions

A frequency approach on MIMO DSP for optical communication systems

A thesis performed at Bell Laboratories (Alcatel-Lucent)

Author:
Xavier Palou Garcia

Supervisors:
Dr. Sebastian Randel
Dr. Joan M. Gené Bernaus

March 2013

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Escola Tècnica Superior de Telecomunicacions de Barcelona
Enginyeria de Telecomunicacions

Abstract

A frequency approach on MIMO DSP for optical communication systems

by Xavier Palou Garcia

Spatial Division Multiplexing (SDM) has arisen during the last years as the most promising technique to overcome the increasing necessity of capacity of the communication systems while avoiding the theoretical capacity limit of the single-mode fibers. Great results have been obtained already using few-mode fibers and multiple-input multiple-output (MIMO) digital signal processing (DSP). Although these are still the first steps of this road, some possible obstacles have been glimpsed already. The one addressed in this thesis is the increasing computational complexity of the algorithms needed when more spatial modes are used and the difficulty of meeting real time requirements.

In this thesis, the current technology and its interesting results are discussed. Moreover, a new approach on the MIMO DSP used to recover the signal is explained and analyzed. This new approach switches from a symbol-wise equalizer in the time domain to a more efficient block-wise equalizer in the frequency domain. These two changes allow to reduce the overall complexity of the algorithm while keeping the performance of the system.

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Escola Tècnica Superior de Telecomunicacions de Barcelona
Enginyeria de Telecomunicacions

Resum

Apropament freqüencial al processament MIMO per a sistemes de comunicació òptics

per Xavier Palou Garcia

La multiplexació en espai (SDM) s'ha postulat durant els darrers anys com la tècnica més prometedora a l'hora d'afrontar la creixent necessitat de capacitat dels sistemes de comunicació alhora que s'evita el límit de capacitat teòric de les fibres monomode. Ja s'han obtingut molt bons resultats utilitzant fibres multimode (amb pocs modes) i tècniques de processament digital del senyal (DSP) de l'estil *multiple-input multiple-output* (MIMO). Tot i que es tracta dels primers passos en aquest nou camí, alguns dels possibles obstacles ja han aparegut. El problema que planteja aquesta tesi és la creixent complexitat computacional dels algoritmes necessaris per utilitzar més modes espaials i la dificultat de complir amb els requeriments del processament en temps real.

En aquesta tesi, es fa una ullada a la tecnologia actual i als interessants resultat que s'han obtingut. A més a més, es presenta una nova estrategia en el processament MIMO utilitzat per recuperar el senyal. Aquest nou algoritme canvia d'un filtre bit-a-bit en el domini del temps a una arquitectura de bloc en el domini freqüencial molt més eficient. Aquests dos canvis permeten reduir la complexitat de l'algoritme mantenint-ne les prestacions.

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Escola Tècnica Superior de Telecomunicacions de Barcelona
Enginyeria de Telecomunicacions

Resumen

Acercamiento frecuencial al procesado MIMO para sistemas de comunicación ópticos

por Xavier Palou Garcia

La multiplexación en espacio (SDM) se ha postulado durante los últimos años como la técnica más prometedora a la hora de enfrentarse a la creciente necesidad de capacidad de los sistemas de comunicación a la vez que se evita el límite de capacidad teórico de las fibras monomodo. Ya se han obtenido muy buenos resultados utilizando fibras multimodo (con pocos modos) y técnicas de procesado digital de la señal (DSP) del estilo *multiple-input multiple-output* (MIMO). Aunque se trata de los primeros pasos en este nuevo camino, algunos de los posibles obstáculos ya han aparecido. El problema que se afronta en esta tesis es la creciente complejidad computacional de los algoritmos necesarios para utilizar más modos espaciales y la dificultad de cumplir con los requerimientos del procesado en tiempo real.

En esta tesis, se repasa la tecnologia actual y los interesantes resultados que se han obtenido. Además, se presenta una nueva estrategia en el procesado MIMO utilizado para recuperar la señal. Este nuevo algoritmo cambia de un filtro bit-a-bit en el dominio del tiempo a una arquitectura de bloque en el dominio frecuencial mucho más eficiente. Estos dos cambios permiten reducir la complejidad del algoritmo manteniendo sus prestaciones.

Acknowledgements

First of all, I want to thank the two people who made possible the experience that ends with this thesis. Without the initiative and effort of Joan M. Gené and Peter Winzer, I would have never been able to visit the Bell Laboratories and discover their work.

I want to thank too all the people of Bell Labs in Crawford Hill for the great experience they made possible during the nine months I spent there. I felt participant of their work and not a mere visiting student. Specially, I want to mention Sebastian Randel and Roland Ryf who took care of me and taught me almost everything I have learned in this stay.

Finally, I want to thank all the people who supported me from home. My family, because they always believe in me and encourage me in my decisions, whatever they are. My closest friends, because they always show an undeserved confidence in my success and are there through the thick and the thin.

Contents

Abbreviations

BER	**B**it **E**rror **R**ate
BLAST	**B**ell Labs **LA**yered **S**pace **T**ime
BPD	**B**alanced **P**hoto**D**etector
CD	**C**hromatic **D**ispersion
CDMA	**C**ode-**D**ivision **M**ultiple **A**ccess
CMA	**C**onstant **M**odulus **A**lgorithm
CPE	**C**arrier **P**hase **E**stimation
CPR	**C**arrier **P**hase **R**ecovery
DA	**D**ata-**A**ided
DBBS	**D**e **B**rujin **B**it **S**equence
DD	**D**ecision-**D**irected
DFB	**D**istributed **F**eed**B**ack laser
DGD	**D**ifferential **G**roup **D**elay
DN-MZM	**D**ouble-**N**ested **M**ach-**Z**ehnder **M**odulator
DSF	**D**ispersion-**S**hifted **F**iber
DSP	**D**igital **S**ignal **P**rocessing
ECL	**E**xternal **C**avity **L**aser
EDFA	**E**rbium-**D**oped **F**iber **A**mplifier
FD	**F**requency **D**omain
FDE	**F**requency **D**omain **E**qualizer
FEC	**F**orward **E**rror **C**orrection
FMF	**F**ew-**M**ode **F**iber
I	**I**n-phase component
LMS	**L**east **M**ean-**S**quare
LO	**L**ocal **O**scillator
LP	**L**inearly **P**olarized

MIMO	**M**ultiple-**I**nput **M**ultiple-**O**utput
MMA	**M**ulti-**M**odulus **A**lgorithm
MMSE	**M**inimum **M**ean-**S**quare **E**rror
NZ-DSF	**N**on-**Z**ero **D**ispersion-**S**hifted **F**ibers
OFDM	**O**rthogonal **F**requency **D**ivision **M**ultiplexing
OBF	**O**ptical **B**andpass **F**ilter
PBS	**P**olarization **B**eam **S**plitter
PD-CRX	**P**olarization **D**iversity **C**oherent **R(X)**eceiver
PDM	**P**olarization **D**ivision **M**ultiplexing
PPG	**P**rogrammable **P**attern **G**enerator
PPM	**P**ulse **P**osition **M**ultiplexing
Q	**Q**uadrature component
QAM	**Q**uadrature **A**mplitude **M**odulation
QPSK	**Q**uadrature **P**hase **S**hift **K**eying
SDM	**S**patial **D**ivision **M**ultiplexing
SMF	**S**ingle **M**ode **F**iber
SNR	**S**ignal-to-**N**oise **R**atio
SPS	**S**amples **P**er **S**ymbol
TD	**T**ime **D**omain
TDE	**T**ime **D**omain **E**qualizer
TDM	**T**ime **D**omain **M**ultiplexing
VOA	**V**ariable **O**ptical **A**ttenuator
WDM	**W**avelength **D**ivision **M**ultiplexing

Chapter 1

Introduction

1.1 Evolution of optical communication systems

Since optical fiber was pointed as the technology that should allow a much greater capacity and longer reach for communication, a lot of new approaches have been introduced. It started with multi-mode fibers which allowed only short distances due to their high modal and chromatic dispersion. Already in the 1980s, single-mode fibers (SMF) were found as a much better solution because the absence of modal dispersion allowed wider bandwidth. The equipment needed much more precision due to the smaller core of the fiber and hence was more expensive, but the advantages overcame the weaknesses. Its introduction gave an important boost to the possibilities of the optical systems. Since then, multiple directions have been taken to get greater performances.

In the fiber design side, new types of single-mode fibers as Dispersion-shifted fibers (DSF) and Non-zero Dispersion-shifted fibers (NZ-DSF) were developed in order to minimize the chromatic dispersion. These fibers take advantage of the double nature of intramodal dispersion which includes both material and waveguide dispersion. Modifying the index profile of the fiber, waveguide dispersion can be offset and the overall zero-dispersion wavelength can be shifted closer to the minimum-loss window around 1550 nm. In this way, it is possible to have a fiber with both low loss and low dispersion.

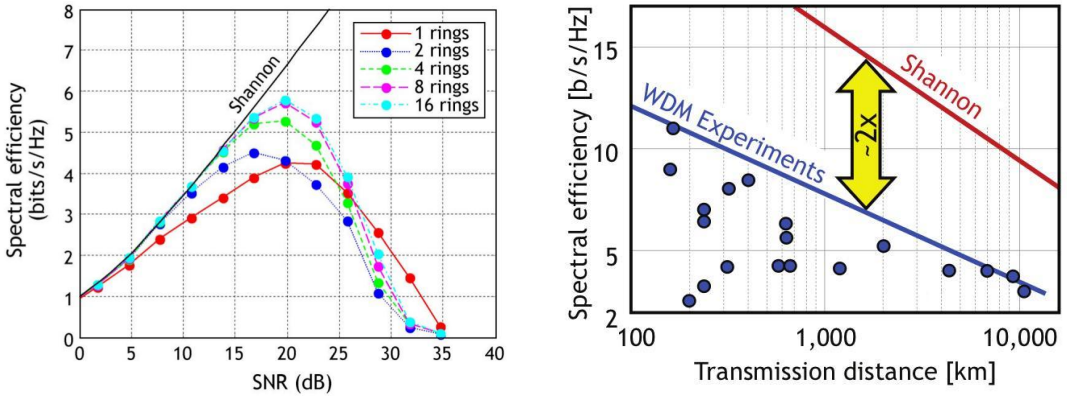
On the other hand, in the transmission side, techniques such as Wavelength Division Multiplexing (WDM) and Polarization Division Multiplexing (PDM) increased the number of parallel channels that could be transmitted in a single fiber [1]. The first one, allows the use of several optical carriers centered in different wavelengths generally within the C band (1530 to 1565 nm). Such signals are orthogonal and individual bit streams can be recovered from each frequency bin using optical bandpass filters or electronic filters following a coherent receiver. The second one, Polarization Division Multiplexing, was first introduced only to re-establish the orthogonality between neighboring frequency bins in WDM when energy leaks degraded its performance. Alternating the polarization

of the adjacent channels solved the problem. This technique was called Polarization Interleaving and finally evolved to true PDM which send an independent signal in each orthogonal polarization doubling the capacity of the system. It is important to note that PDM was not possible until the introduction of coherent receivers.

Other multiplexing formats that have been used to increase the channel capacity are Orthogonal Frequency Division Multiplexing (OFDM) and Code-Division Multiple Access (CDMA). The OFDM allows the avoided interleaving of WDM by using as frequency spacing between carriers the reciprocal of the symbol duration. With this technique, a receiver can separate the information transported in each carrier by multiplying the whole signal with a sine wave of the desired frequency. A similar approach is applied in CDMA, but using orthogonal sequences of pulses that differentiate the carriers.

Finally, advanced high-order modulation formats use the real and the imaginary part of the optical field in the Quadrature Amplitude Modulation (QAM). These modulation formats help in the improvement of the spectral efficiency of the channel [2].

However, not all have been good news in the last decades. All the mentioned techniques have enabled a great improvement in the channel capacity of the SMF, but this is coming to an end. In 2010, Essiambre and his co-authors showed in [3] that, in the optical channel case, the theoretical capacity limit is way below the general case set by Shannon in 1948 [4]. When the transmission distance increases, the power needed increases too. This greater power leads to the appearance of non-linear effects that penalize the total capacity achievable. In fact, the current SMF systems are getting very close to this limit, within a factor of 2 [5].



(a) (b)

FIGURE 1.1: (a) Theoretical optical spectral efficiency after a transmission over 2000 km in the case of uniform constellations with different number of rings. Figure from [3]. (b) Experimentally achieved WDM spectral efficiencies (blue) and nonlinear Shannon limit (red). Figure from [5].

It is clear that a new way has to be explored to give fresh air to the field of information transport technologies that sees how the traffic growth maintain a pace over an annual

50% since the year 2000 [2]. With all the other physical dimensions already exploited, only the space multiplexing is still available for this purpose. Of course, the first approach would be placing several parallel fibers in a single cable but that is not an economically neither energetically efficient solution because it means the installation of more and more fiber. The technique that could allow a significant capacity increment in an efficient way is the Spatial Division Multiplexing (SDM). The SDM consists in the use of fibers with more than one physical path, that is, several propagation modes allowed or several cores. As WDM, this technique is highly scalable as using M modes or cores means, ideally, multiplying the capacity of the system by a factor of M and, potentially, we can choose this M very big.

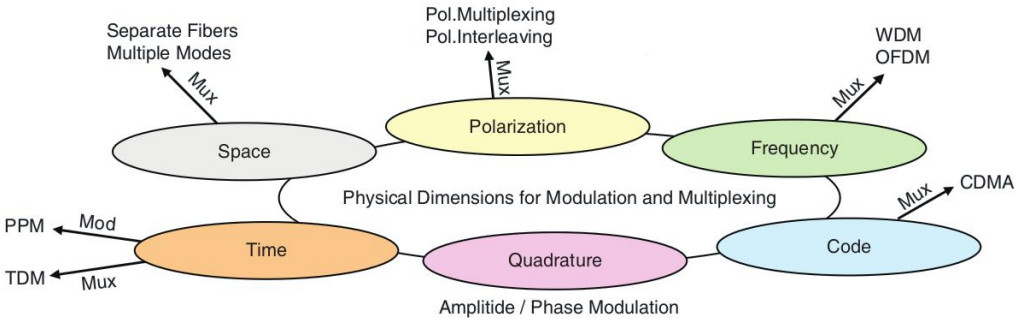


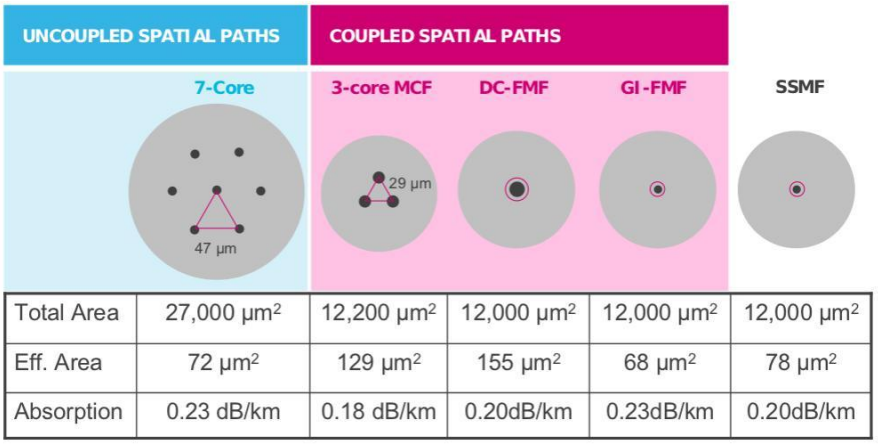
FIGURE 1.2: Known physcal dimensions that can be used for modulation and multiplexing in optical communications. Figure from [1].

1.2 Spatial Division Multiplexing

The SDM techniques, as stated above, consist in using not only one physical path within the fiber but many. This can be achieved by using muti-core fibers or by exciting more than one propagation mode in the fiber. In the first case, each one of the cores behaves as it was a normal single-mode fiber propagating information only in the fundamental mode. In the second case, the first several propagating modes of the fiber are excited in order to use each one of them as a parallel communication channel. This few-mode approach leads to a high crosstalk between the different SDM paths with big transmission penalties. When the penalties are unacceptable it is necessary the use of Multiple-Input Multiple-Output (MIMO) techniques -originally developed for wireless systems [6]- to separate the transmitted signals.

Between multi-core and multi-mode approaches there is a continuous set of possibilities. If the distance between the cores decreases, crosstalk appears between them. At some point, these crosstalk can achieve the same level than in a few-mode fibers and finally, the several cores fuse into only one with several modes excited. It is still difficult to say where is the optimum balance within this set of possibilities because there are a lot of factors involved, namely the fiber and transponders fabrication cost, the complexity of the MIMO digital signal processing, the overall performance achieved, etc. Some

experiments have been carried out with success in all cases: for low-crosstalk multi-core fibers [7, 8], for few-mode fibers [9–11], for highly coupled multi-core fibers [12] and even for hybrid multi-core fibers with a mix of single-mode cores and few-mode cores [13]. This thesis will be centered in the few-mode fiber case.



Total Area	27,000 µm²	12,200 µm²	12,000 µm²	12,000 µm²	12,000 µm²
Eff. Area	72 µm²	129 µm²	155 µm²	68 µm²	78 µm²
Absorption	0.23 dB/km	0.18 dB/km	0.20dB/km	0.23dB/km	0.20dB/km

FIGURE 1.3: Possible physical properties achievable today for different types of fiber in terms of core/mode configuration.

1.2.1 The few-mode fiber model

Neglecting the nonlinearities, a few-mode fiber (FMF) that allows the propagation of M modes can be described with the following linear matrix equation [1]

$$\mathbf{y} = \sqrt{E_0}\sqrt{L}\,\mathbf{H}\mathbf{x} + \mathbf{n} \tag{1.1}$$

represented schematically in Figure 1.4. Vectors $\mathbf{x}$ and $\mathbf{y}$ represent the input and output of the modes of the FMF, respectively. The propagation modes allowed in the channel correspond to a set of physical orthogonal modes which can include spatial and polarization components. Between them, there may be differential gain or loss and differential delay. However, in this brief mathematical description we are going to neglect the latter. We assume that the transmitter is able to address $M_T \leq M$ modes with an average signal energy per symbol period and mode of E_0. Hence, the total transmit energy across all modes is $M_T E_0$ per symbol and the vector $\mathbf{x}$ satisfy $\langle ||\mathbf{x}||^2 \rangle = M_T$. In the same way, the receiver is able to coherently extract $M_R \leq M$ modes. The SDM channel is described by an $M \times M$ matrix $\widetilde{\mathbf{H}}$. It can be normalized as $\widetilde{\mathbf{H}} = \sqrt{L}\,\mathbf{H}'$ by factoring out the mode-average propagation loss L, which can be calculated as

[1] In the whole document *italics* are used for scalars, **bold print lower-case** for vectors and **BOLD PRINT UPPER-CASE** for matrices.

$$L = \frac{1}{M} tr\{\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^{\dagger}\} = \frac{1}{M}\sum_{i=1}^{M}\widetilde{\lambda}_i; \tag{1.2}$$

where $tr\{.\}$ denotes the trace of a matrix, $\widetilde{\mathbf{H}}^{\dagger}$ is the transpose conjugate of the matrix $\widetilde{\mathbf{H}}$ and $\widetilde{\lambda}_i$ are the M eigenvalues of $\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^{\dagger}$. Our MIMO system, where we only excite and receive M_T and M_R modes respectively, is described by an $M_R \times M_T$ matrix $\mathbf{H}$ that span the corresponding subspace within the matrix $\mathbf{H}'$. Finally, $\mathbf{n}$ represents the circularly symmetric complex Gaussian noise with power spectral density N_0 per mode added at the receiver. Of course, the matrix $\mathbf{H}$ is just an instant representation of the channel. It belongs to an ensemble $\mathcal{H}$ of possible instantiations.

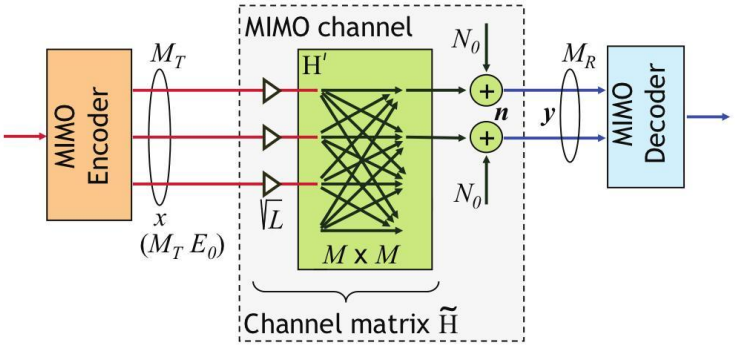


FIGURE 1.4: Representation of a linear MIMO system with $M = 4$, $M_T = 3$ and $M_R = 2$. Figure from [5].

Using this MIMO configuration, the capacity limit for our SDM channel when sending uncorrelated signals of equal power E_0 on all transmit modes is [5]

$$C = \sum_{i=1}^{m} \log_2\left(1 + \lambda_i \frac{E_0 L}{N_0}\right), \tag{1.3}$$

known as the open-loop Bell Labs Layered Space Time (BLAST) capacity. The λ_i are the $m \leq \min\{M_T, M_R\} \leq M$ non-zero eigenvalues of $\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^{\dagger}$, the rank of the MIMO system. The term

$$SNR = \frac{E_0 L}{N_0} \tag{1.4}$$

represents the mode-average signal-to-noise ratio (SNR) at the receiver when all M channel modes are addressed and detected ($M_T = M_R = M$). Again, the capacity described by Eq. (1.3) in an instantiation of the channel capacity as it depends on the instantiation $\mathbf{H}$ of the channel itself.

If we assume that the channel impairments attenuate and couple the modes, but without eliminating completely the propagation of one of them, the rank of the MIMO system

is given by $m = \min\{M_T, M_R\} \leq M$. In the absence of mode-dependent loss, the maximum capacity limit of the under-addressed FMF is m times that of the single-mode fiber C_s [5]

$$C \leq mC_s = m \log_2 (1 + SNR) . \tag{1.5}$$

This means that to reach the full MIMO capacity of a fiber that allows the propagation of M modes, we have to address and detect a complete orthonormal basis of modes of dimension M. For instance, in the FMF this basis could correspond to a set of guided linearly polarized propagation modes.

In general, in a waveguide, the wave equation and the boundary conditions lead to a set of solutions that determine the modes that can be propagated. This modes can be classified regarding its electromagnetic field pattern in the plane perpendicular to the propagation direction. This classification leads to four main types of modes:

- Transverse electric (TE): no electric field in the propagation direction.
- Transverse magnetic (TM): no magnetic field in the propagation direction.
- Transverse electromagnetic (TEM): neither electric nor magnetic field in the propagation direction.
- Hybrid (EH/HE): nonzero electric and magnetic fields in the propagation direction.

In the optical fiber case, all these modes are present. Nevertheless, when the difference between the refractive indexes of the core and the cladding is very small ($n_{core} \approx n_{clad}$), the longitudinal components of the modes are almost negligible in front of the transverse components. This leads to what is called the Weakly Guiding Fiber approximation in which the guided modes are almost TEM [14, 15]. This modes are called linearly polarized (LP) because the polarization directions x and y form orthogonal modes and can be used practically as a base for the modes in the fiber (PDM and SDM are closely related). The LP modes are combinations of the exact solution modes of the fiber (TE, TM and EH/HE) and follow the rule on Table 1.1. This relation can be better understood in the plots of Figure 1.5.

LP modes	Exact solution modes
LP_{0m}	HE_{1m}
LP_{1m}	TE_{0m}, TM_{0m}, HE_{2m}
LP_{lm}	$EH_{l-1,m}$, $HE_{l+1,m}$

TABLE 1.1: Correspondence between the exact solutions of the wave equation and the LP modes obtained with the Weakly Guiding Fiber approximation.

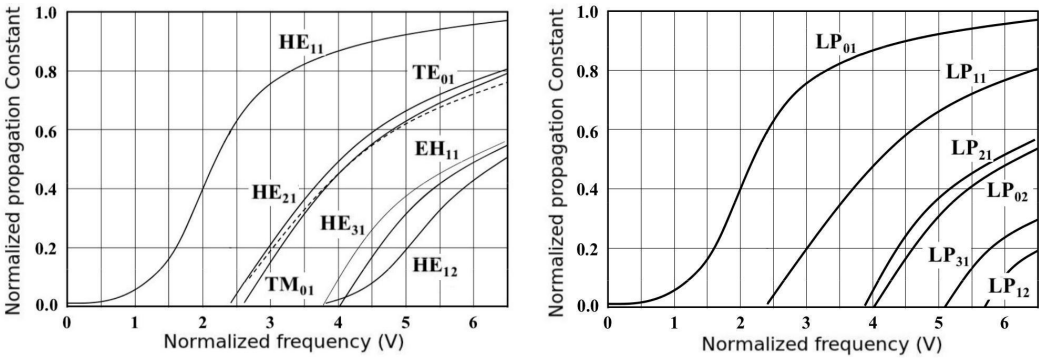


FIGURE 1.5: Cut-off normalized frequencies of the exact modes of the fiber and the LP modes. It is easy to see how the exact modes are grouped to form the LP modes.

The number of LP modes that can be propagated in a fiber is determined by a parameter of the fiber called normalized frequency. This parameter depends on the size of the core and on the refractive index profile of the fiber as

$$V = 2\pi \frac{a}{\lambda_0} \sqrt{n_{core}^2 - n_{clad}^2}, \tag{1.6}$$

where a is the radius of the core and λ_0 is the central frequency of the signal. Each LP mode has a cut-off normalized frequency below which it can not be propagated. Moreover, the LP modes are described by two subscripts l and m. The first one, l, indicates the number of cycles in the angular distribution and the second one, m, the number of nodes in the radial distribution.

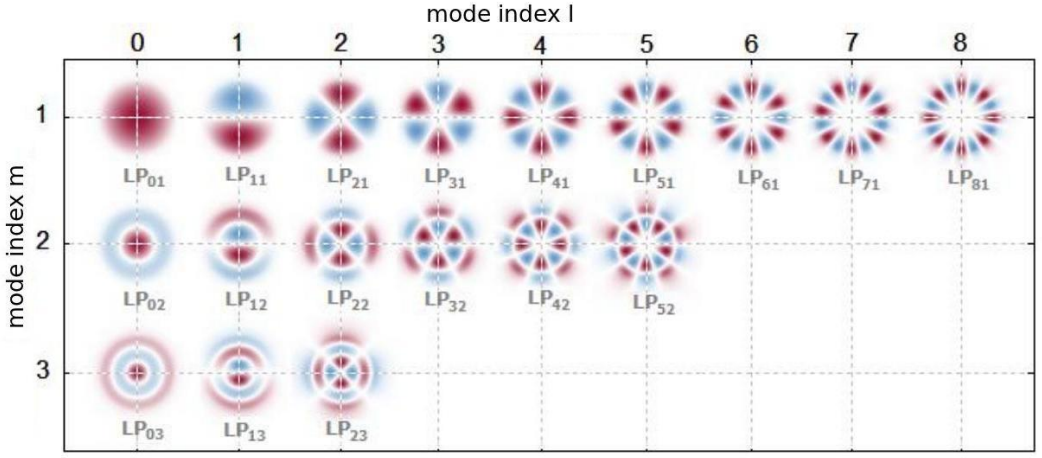


FIGURE 1.6: Theoretical transverse field profiles of the LP modes for a graded index FMF. The blue and red areas represent opposed phase regions.

Finally, it is important to note that all the LP modes, with the exception of the LP_{0m}, have a phase indeterminacy of $\pi/2l$. This means that the modes have a twofold degeneracy. The pairs of degenerated modes have the same propagation constants and can be used as two different paths in SDM.

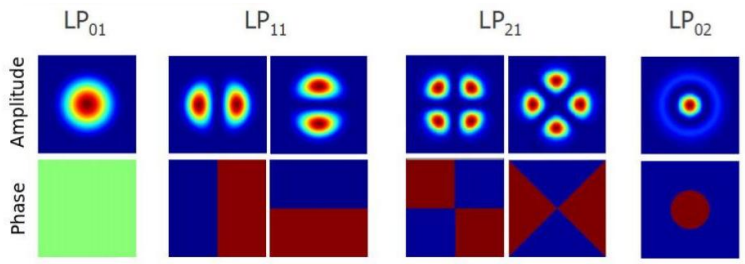


FIGURE 1.7: Theoretical intensity and phase profiles of the LP modes for a graded index FMF with 6 spatial modes. The degeneracy of LP_{11} and LP_{21} is shown.

1.3 Overview

The aim of this work is to analyze a new approach in the frequency domain for the MIMO DSP used to recover the transmit data in SDM experiments. Nevertheless, to understand the motivation and the ideas involved in this new algorithm is necessary to previously introduce the current paradigm. For that reason, in Chapter 2, the setup used in all the cited experiments (some of them done during this thesis [10, 11] and some before and after it) will be explained. Although the algorithm is mainly mathematical, some of the stages answer directly to physical impairments related to the optical setup and it is important to know a bit about it. In Chapter 3, the algorithm used in most of the already successful SDM experiments will be explained in detail. This code is based in an adaptive gradient algorithm in the time domain, in particular the standard Least Mean-Square algorithm (LMS). Some of the ideas used in this algorithm are the base for the development of the new approach developed for this thesis and explained in Chapter 4. On the other hand, the limitations of the algorithm observed in the experiments motivated the change of viewpoint in the new code. The main change respect to his predecessor is the fact that all the work is done in the frequency domain taking advantage of the high efficiency in the calculation of the FFT achieved by the FFT-in-the-West routine (FFTW) [16].

Chapter 2

The experimental setup

The configuration of the different experiments has changed slightly depending on the purpose of them. Moreover, the progressive better understanding of the SDM and the solutions to detected problems have added novelties to the setup too. Therefore, the description of the system will be as general as possible and indicating all the variations. In Figure 2.1 the most general setup possible is shown. Depending on the experiment, some of the functional blocks were not included.

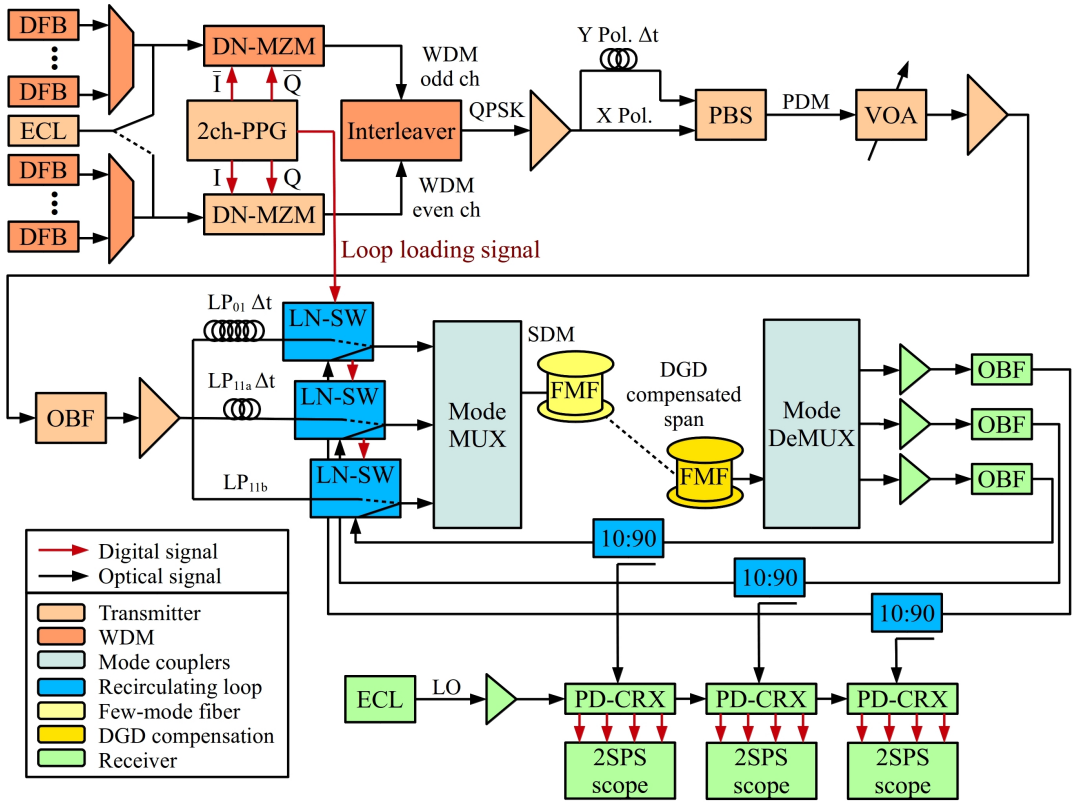


FIGURE 2.1: General SDM experimental setup for a 3-mode FMF transmission. The functional blocks are differentiated with colors.

2.1 The transmitter

All the experiments performed consisted, at least, in the transmission of QPSK signals using PDM and SDM, simultaneously. If N_m is the number of spatial modes guided by our FMF ($N_m = 3$ or $N_m = 6$, up to now. See Figure 1.7), $2N_m$ uncorrelated signals are needed to cover both the PDM and the SDM modes. Nevertheless, the system starts with only a 2-channel programmable pattern generator (PPG) that produces two independent De Brujin bit sequences (DBBS) used as the in-phase (I) and quadrature (Q) components of a QPSK signal. The speed of the PPG depends on the needs of the experiment but the most used value is 20 GBaud/s, that is, 40 Gb/s for the QPSK case. The length of the DBBS sequences is 2^p, where p is the order of the generating polynomial. Again, it depends on the experiment and will be discussed later. The two signals drive a double-nested $LiNbO_3$ Mach-Zehnder modulator (DN-MZM). As a light source, an external cavity laser (ECL) with a linewidth of 100 KHz is used.

To generate the $2N_m$ uncorrelated signals, the QPSK is delayed different number of symbols using extra fiber paths. First of all, a delayed copy is used for the orthogonally polarized component which is added to the original with a polarization beam splitter (PBS). The PDM signal is further split and delayed in N_m different amounts to generate the N_m modal components. To keep the resulting signals uncorrelated from the receiver equalizers point of view, the delays introduced between them have to be longer than the maximum modal dispersion the signal will suffer during the propagation. This conditions the length of the DBBS which has to be greater than the sum of all the delays. The value used for p ranges from 11 to 14, that is, the DBBS length ranges from 2048 symbols [17] to 16384 symbols [11].

To get suitable input power in the FMF, erbium-doped fiber amplifiers (EDFA) are introduced between the modulator and the PBS and after the PBS. Moreover, if the experiment requires the loading of additional noise to the transmission, it is generated with a variable optical attenuator (VOA), an optical bandpass filter (OBF) and, if needed, another EDFA.

In addition, in the most recent experiments [10, 11], the PDM and the SDM have been already combined with WDM. In this case, a grid of 50 GHz spaced distributed feedback lasers (DFB) is used to independently modulate the odd and the even frequency channels. That means using two DN-MZM driven with conjugated copies of the same I and Q DBBS signals. For the channel under test, the above mentioned ECL is introduced in the corresponding position in the grid. The two groups of channels are put together with an interleaver.

At this point, a PDM (and in some cases WDM) signal has been generated for each one of the spatial modes to be used. Nevertheless, they are still propagating inside single-mode fibers and have to be introduced in a FMF coupled in each of the fiber modes.

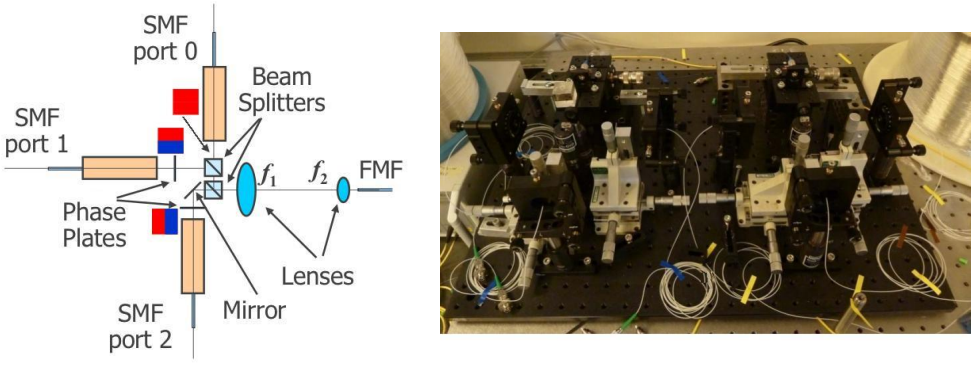
Nowadays, different techniques to do it based in different principles and with different performances coexist. It is not clear yet which of them will succeed, so some will be briefly presented.

2.1.1 Mode couplers

The goal of a mode coupler is to take the light from multiple incoming beams or SMFs and combine it such that every incoming signal is coupled to a different mode of the FMF. As the LP modes are orthogonal it should be possible to make it lossless. This is an analogy to a polarization beam splitter for PDM or an arrayed waveguide grating for WDM. All the techniques explained in the following can be used both as multiplexers and demultiplexers.

The first mode coupler used was based in the phase profile of the LP_{01} and LP_{11} modes. As seen in Figure 1.7, those profiles consist only of an homogenous phase for LP_{01} and two half planes with a constant phase difference of π for LP_{11}. Inserting in the beam path a thin glass plate with a phase pattern consisting of thickness differences, each one of the phase patterns can be created [18]. With these plates, it is not possible to create the amplitude pattern of the modes. However, because of the symmetry of the modes, the amplitude pattern obtained can be coupled only in the corresponding mode.

The plates were made of 0.5 mm Borosilicate glass with a 1.7 μm thickness difference etched with a photolithographic method. As each plate can only couple a mode at a time, beam splitters are required to combine the modes generated by several plates. The global mode conversion loss (around 9-10 dB in [17, 18]) comes basically from the beam splitters and the optical system that images the facet of the incident SMF on the facet of the FMF. For higher number of modes, the phase plates become more complicated. In addition, more beam splitters are used and the insertion loses increase.



(a) (b)

FIGURE 2.2: (a) Architecture of the phase plate 3-mode coupler. (b) Setup of the phase plate 3-mode coupler in the laboratory.

The second technique used is based in both the amplitude and the phase profile of the modes. When illuminating a FMF with a focused spot, the intensity and phase of the excited modes can be determined by the overlap between the spot and the mode profiles. In [19] it was shown that, for the case of the 3-mode fiber, it is possible to address the modes using the same three spots necessary in 3-core fiber experiments. The spots do not directly excite particular fiber modes, but a linear combination of them. By choosing the correct parameters for each beam, it is possible to eliminate the mode-dependent loss and the coupling process can be described by a unitary transformation of the excited fiber modes and thus can be inverted by MIMO DSP. This technique can be extended to excite a higher number of modes as it is developed in [20]. The number of necessary spots coincide with the number of spatial modes to be excited.

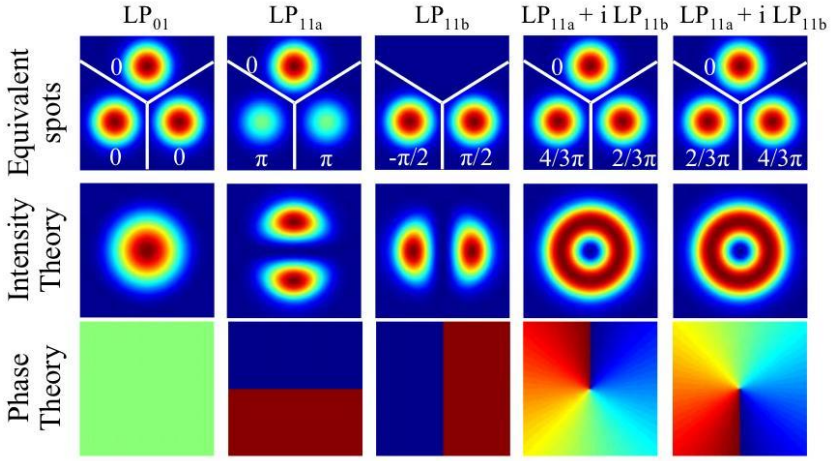


FIGURE 2.3: Amplitude and phase parameters for each spot (first row) needed to generate the theoretical amplitude and phase patterns of the LP modes (second and third row). Fourth and fifth column show an equivalent basis mode set for the representation of the degenerated LP_{11}. Figure from [19].

This principle can be put in practice in many different ways. In a first approach, the position of the spots is achieved by a mirror setting as the one shown in Figure 2.4(a). The lower insertion loses achieved with this method were below 4 dB for all three excited modes [19]. However, as in the phase plates case, the generalization to a higher number of modes is non-trivial. Another approach is the fabrication of a 3 dimensional optical circuit to guide the beams from a planar configuration to the desired spot positions. This kind of 3D waveguides (see Figure 2.4(b)) are inscribed be means of laser techniques and can be generalized for a higher number of modes. For example, in [11] it was used for a experiment with 6 spatial modes with an insertion loss of less than 6 dB.

It is important to note that in this kind of technique one key factor is the optimization of the spot configuration. The size of the spots and their distance to the center are critical. In [20], some geometrically symmetric configurations are studied for increasing number of spatial modes but other useful configurations could be found. Again, it is still very difficult to say what will succeed.

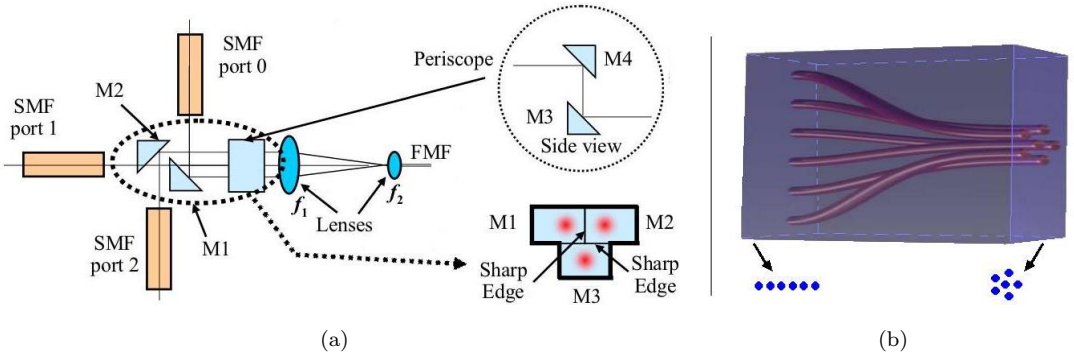


FIGURE 2.4: (a) Architecture of the mirror-based spot coupler for 3 modes. Figure from [19]. (b) Representation of a 3D waveguide for 6 modes.

Finally, the last approach is based on the photonic lantern [21]. In this technique, multiple single-mode fibers are fused together to form a new multi-mode fiber. This idea comes from astronomy and, although in the experiments [11] the insertion loses were close to 6 dB as in the 3D waveguide, it is expected to be able to lower them down to almost 0 dB. As the initial signals propagates in several SMF, the excited modes in the FMF are again orthogonal linear combinations of the LP modes separable by MIMO DSP.

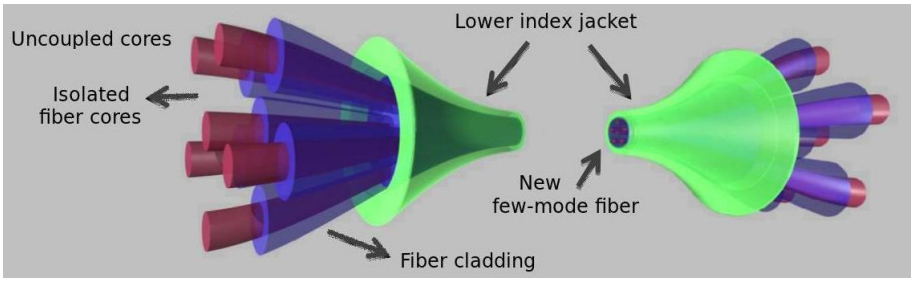


FIGURE 2.5: Representation of a photonic lantern to couple 6 SMF into a 6-mode FMF.

2.2 The fiber

The few-mode fibers used in all the experiments are very similar. In general, fibers with a graded-index profile and a loss around 0.2 dB/km are used. When only LP_{01} and LP_{11} modes are excited, the effective areas use to be of 64 μm^2 and 67 μm^2. In the last experiments, where LP_{21} and LP_{02} are allowed too, the effective areas grew to 90 μm^2 for LP_{01} and LP_{11}, to 120 μm^2 for LP_{21} and to 180 μm^2 for LP_{02}. The chromatic dispersion, important when the experiments include WDM, is usually around 18 ps/nm·km.

Nonetheless, the most valuable parameter in the SDM experiments is a small modal differential group delay (DGD). That is because the MIMO DSP has to recover the

information from all the modes, so the time window of the filters has to be, at least, as long as the accumulated modal dispersion. As mentioned above, the most usual transmission velocity is 20 GBaud/s, that is, every 50 ps of modal dispersion represent one more symbol to take into account in the equalizers. This is the reason why in most of the experiments, two or more spools of fiber with different lengths and DGDs are combined to get a longer fiber with as low as possible DGD. In some cases, it is possible to completely compensate the DGD as in [22]. However, a great diversity of fiber samples is needed, specially when the number of modes increase and so the number of DGDs to optimize.

Finally, to be able to increase the transmission distance without using more fiber spools, some of the experiments incorporate a recirculating loop. $LiNbO_3$ switches (LN-SW) placed in the SMF before the mode multiplexers control the loading and the closing of the loop. The loop consists, in this order, of a mode multiplexer, a DGD-compensated FMF span, a mode demultiplexer and an amplification stage for each SMF (EDFAs and filters to block the out-of-band noise and equalize the power in the loop). To extract the signal from the loop, 10:90 couplers are used. It is important to align accurately the loop in order to not introduce more delays between paths or mode-dependent loss which penalize dramatically the global performance.

2.3 The receiver

The signal taken from the mode demultiplexer (or from the 10:90 couplers if there is a loop circuit) is preamplified with EDFAs and, in some of the experiments, filtered to eliminate the out-of-band noise. Then, it is detected by polarization-diversity coherent receivers (PD-CRX) run by an ECL as a local oscillator (LO). This kind of receiver comprises two PBS to separate the two polarizations of the received signal (R) and the LO. Afterward, a 90° hybrid produces four interference signals from each polarization (R+LO, R-LO, R+jLO, R-jLO). Two balanced photodetector (BPD) use the first and the second pair of interference signals to produce the I and Q components, respectively. That means that for each propagation mode of the FMF, four real components are detected: the I and Q components of the two orthogonal polarizations (I_x, Q_x, I_y, Q_y).

These PD-CRXs work in intradyne configuration, which means that the angular frequency of the LO, ω_{LO} and the signal, ω_R are very close and the signal recovered by the BPD is close to the baseband ($\omega_{IF} = \omega_R - \omega_{LO}$). In this case, the photodetectors need a significantly small bandwidth with respect to the heterodyne case (ω_{IF} of tens or hundreds of MHz) and just slightly bigger than in the homodyne case ($\omega_{IF} = 0$). Moreover, digital DSP techniques can compensate the carrier frequency offset.

These waveforms are, finally, sampled by digital storage oscilloscopes. Although in a practical implementation the DSP should be performed in real time by an integrated

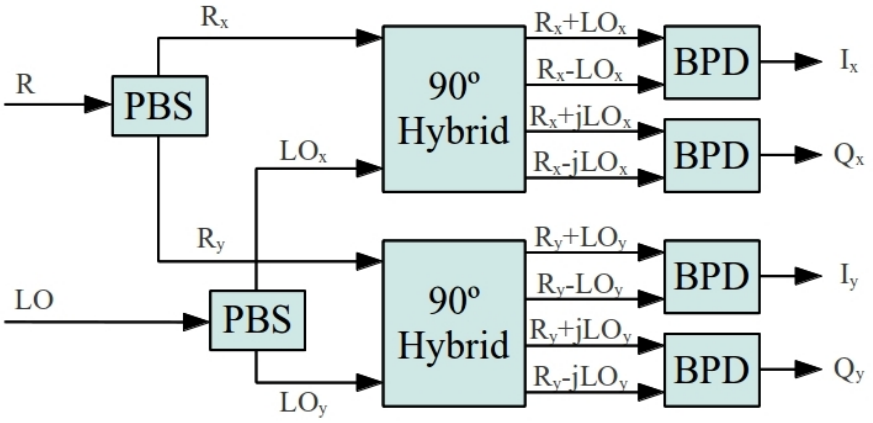


FIGURE 2.6: Block diagram of a polarization-diversity coherent receiver.

circuit, up to now, it is done off-line in a computer. This is the reason why the signal is stored in the scopes.

The desired sampling rate is always double the transmission rate so the data to process contains two samples per symbol (SPS) and can be completely restored (Nyquist samplig theorem). This means, for example, that for a sequence transmitted at 20 GBaud/s, the scopes have to work at 40 GHz. If the available scopes do no fit that specification, it is possible to digitally resample the data before the DSP.

Chapter 3

The Time-Domain equalizer

Once the signal detected in the coherent receivers is measured and stored by the digital oscilloscopes, it is the time for the DSP. As mentioned before, in all the experiments considered in this thesis the processing is done off-line. The main reason for this is the lack of the proper equipment and expertise to develop the electronics capable of the real-time DSP. In a first stage, the algorithms are programmed in Matlab for its convenience in vector and matrix calculations and its flexibility. In a second stage, when the algorithms are stable and working at their full capacity, they are translated to C language to take advantage of its speed.

The received signal consist of $4N_m$ streams of real data: the I and Q components ($\times 2$) of each polarization ($\times 2$) of each spatial mode ($\times N_m$). Nonetheless, the I and Q components are added from the beginning so the algorithm works with $2N_m$ complex vectors. As mentioned in Chapter 2, the signal is oversampled with a factor of two, that is, two samples per symbol are used. This signal is corrupted by different impairments that must be compensated: chromatic dispersion, frequency offset, AWGN noise, mode coupling, mode dispersion and random phase noise.

The chromatic dispersion and the frequency offset (basically from the offset in the central frequencies of the LO lasers of the transmitter and the receiver) will not be treated in depth in this thesis. In first versions of the code, these effects were compensated before starting the symbol-wise processing that simulates a real time DSP. It was done using the FFT of the stored data. That was not a good approach for a real-time application because the whole record of stored data was used. In real time, a big window would be needed to get good results and the delay would be unacceptable.

The value of the chromatic dispersion of a given fiber is previously known, so the trick to compensate it consist of a simple multiplication in the frequency domain by an appropriate factor. In the frequency version of the algorithm explained in Chapter 4 it can be incorporated in the filter. In the case of the frequency offset, the FFT of the signal was used to estimate the value of the offset and, afterward, it was compensated in the

time domain. In a real-time approach, the estimation can be done by time-averaging the phase error during a long window and applying the corresponding correction.

The main part of the algorithm corresponds to the adaptive FIR filters which use the Least Mean-Square method (LMS) for the adaptation. These filters are intended to reverse the coupling between the modes and to eliminate the noise accumulated in the signal during the propagation which can be taken as Additive White Gaussian Noise. That is why the filters follow a MIMO architecture combining all the received data. The length of the filters is taken to be able to compensate the maximum mode dispersion of the signal. This structure will be explained in detail in Sections 3.1 and 3.2.

Finally, the random phase error, which come from different sources like the instability of the transmitter and receiver lasers or the non-linearities of the fiber, is the most unknown effect. There are different techniques to estimate and compensate it. For QPSK signals, the most accepted method is the fouth-power algorithm which in our case is adapted to fit in the LMS filter structure. This will be shortly explained in Section 3.3.

To finish Chapter 3, the experimental results achieved with this algorithm will be recalled in Section 3.4 and its limitations will be pointed out in Section 3.5 to justify the need of the new point of view developed in Chapter 4.

3.1 The Least Mean-Square algorithm

As mentioned above, the filters used in the code are adaptive causal FIR filters using the Least Mean-Square algorithm for the adaptation. This algorithm is a descent gradient algorithm based in the Wiener optimum filter [23]. If the input of a system is $x(k)$ (with k the temporal index) and the filter goal is to replicate as close as possible a desired signal $d(k)$, the output $y(k)$ of the system will be

$$y(k) = \sum_{l=0}^{L-1} w_l \, x(k-l) \tag{3.1}$$

where L is the number of past samples used in the estimation of the current output and w_l are the coefficients of the filter. The error of the output with respect to the desired signal will be

$$e(k) = d(k) - y(k) \tag{3.2}$$

The theory of the Wiener optimum filter uses the minimum mean-square error (MMSE) criterion to find the optimal coefficients for the filter. That is, minimize the cost function

$$J(k) = E\{|e(k)|^2\} \tag{3.3}$$

with respect to the coefficients w_l. This process leads to an optimal vector of coefficients

$$\mathbf{w}_{opt} = \mathbf{R}^{-1}\mathbf{p} \tag{3.4}$$

where $\mathbf{R} = E\{\mathbf{x}^*(k)\mathbf{x}^T(k)\}$ is the autocorrelation matrix of the complex conjugate of the input vector, $\mathbf{x}^* = [x^*(k), x^*(k-1), \ldots, x^*(k-L+1)]^T$, and $\mathbf{p} = E\{d(k)\mathbf{x}^*(k)\}$ is its correlation with the desired signal sample $d(k)$.

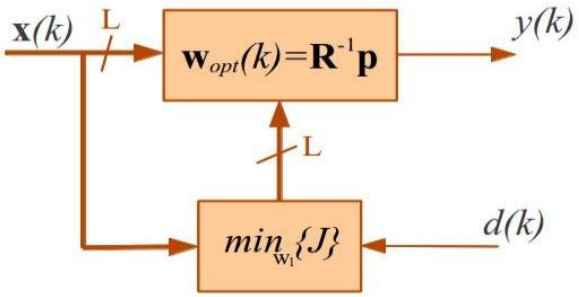


FIGURE 3.1: Structure of the causal FIR Wiener optimum filter.

However, this approach is not practical because the correlations $\mathbf{R}$ and $\mathbf{p}$ are not usually known and it is not easy to estimate them. Moreover, the solution implies inverting the matrix $\mathbf{R}$ which has a computational complexity of at least the square of the number of coefficients in the filter.

To avoid the calculation of the inverse matrix and regarding the fact that the overall problem is a minimization problem, it is possible to introduce the notion of descent gradient algorithm. This kind of algorithm solve the optimization of a differentiable function by iteratively following his direction of steepest slope which leads always to a local maximum/minimum. This direction corresponds to the gradient of the function with positive or negative sign for maximization or minimization, respectively. Moreover, an initial value for the coefficients must be provided.

For the case of the optimum filter, the gradient of the cost function $J(k)$ is

$$\nabla_{\mathbf{w}(k)} J(k) = 2(\mathbf{R}\mathbf{w}(k) - \mathbf{p}) \tag{3.5}$$

so the descent gradient algorithm is defined as

$$\begin{aligned} \mathbf{w}(0) &= \mathbf{w}_0 \\ \mathbf{w}(k+1) &= \mathbf{w}(k) - \mu \nabla_{\mathbf{w}(k)} J(k) \\ &= \mathbf{w}(k) - 2\mu(\mathbf{R}\mathbf{w}(k) - \mathbf{p}) \end{aligned} \tag{3.6}$$

where μ is the adaptation gain, a real-valued positive constant that controls the speed and convergence of the method [23]. The initial vector of coefficients $\mathbf{w}_0$ can be engineered to make the convergence faster, but usually it is enough with a normalized constant vector $\mathbf{w}_0 = [1/L, \ldots, 1/L]$.

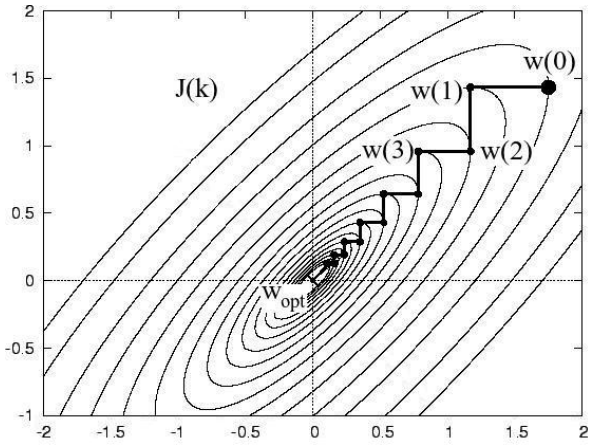


FIGURE 3.2: Evolution of a convex cost function $J(k)$ in relation with a two dimensional coefficient vector $\mathbf{w}(k)$ in a descent gradient algorithm.

On the other hand, to solve the problem of the estimation of the correlations $\mathbf{R}$ and $\mathbf{p}$, instantaneous approximations of their value are used. With this solution, the algorithm has a much lower computational complexity. This approximated method is called the Least Mean-Squares algorithm or the stochastic gradient algorithm. The instantaneous estimations used are

$$\hat{\mathbf{R}}(k) = \mathbf{x}^*(k)\mathbf{x}^T(k) \tag{3.7}$$

$$\hat{\mathbf{p}}(k) = d(k)\mathbf{x}^*(k) \tag{3.8}$$

Now, the gradient of the the cost function $J(k)$ is approximated by

$$\begin{aligned} \widehat{\nabla_{\mathbf{w}(k)}} J(k) &= 2\left[\mathbf{x}^*(k)\mathbf{x}^T(k)\mathbf{w}(k) - d(k)\mathbf{x}^*(k)\right] \\ &= -2\mathbf{x}^*(k)\left[d(k) - \mathbf{x}^T(k)\mathbf{w}(k)\right] \\ &= -2\mathbf{x}^*(k)e(k) \end{aligned} \tag{3.9}$$

Finally, applying Eq. (3.9) to Eq. (3.6), the algorithm takes the shape

$$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mu\mathbf{x}^*(k)e(k) \tag{3.10}$$

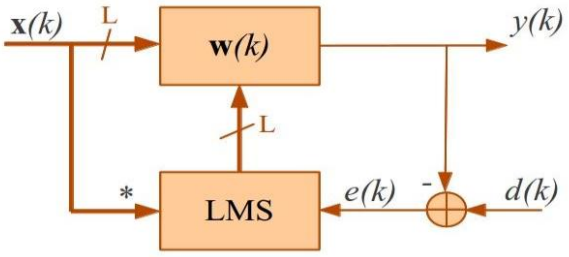


FIGURE 3.3: Structure of the LMS or stochastic gradient algorithm.

As already commented above, this method is iterative and needs a convergence time. In addition, for the calculation of $\mathbf{w}(k+1)$ it is necessary to know the value of the error $e(k)$ or, what is the same, it is necessary to know the value of the desired signal $d(k)$ to compare it with the output $y(k)$. In a real scenario, the desired signal in the receiver is, of course, the transmitted signal so it is not possible to know it in advance. Nevertheless, what is possible is to use training sequences of a given length which help the algorithm to converge. When the filter coefficients get close to the optimum, the outputs will be close to the desired signal making possible a correct guess of $d(k)$ in most of the cases. At that point, the training sequence can be stopped and substituted by real data unknown in the receiver.

With this in mind, the algorithm has two different stages. The first one, called Data-Aided stage (DA), uses as desired signal $d(k)$ a previously known training sequence which is feed to the equalizer to help the convergence. After a while (a fixed time depending a lot on the experiment), the algorithm switches to the Decision-Directed stage (DD) in which no information is given to the algorithm. After each iteration, the algorithm decides which constellation point may correspond to $y(k)$ depending on the Euclidean distance criterion. In this stage, therefore, the algorithm is a blind equalizer.

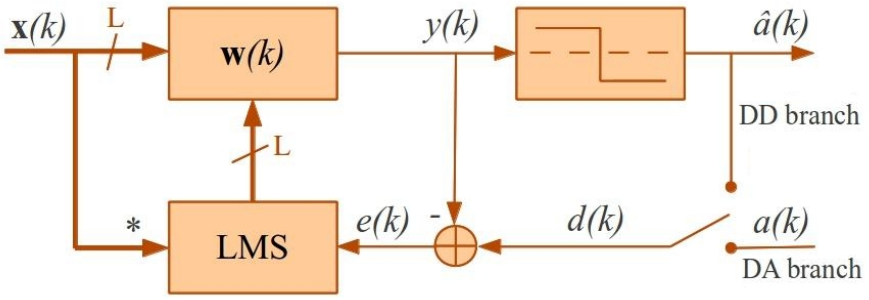


FIGURE 3.4: Structure of the LMS algorithm with two branches for the Data-Aided and the Decision-Directed stages.

Although there is a theory around the convergence speed of the gradient algorithms [23], after first theoretical approximations, the value of the adaptation gain and the duration of the DA stage is determined through a trial and error optimization process.

It is important to note that because of the randomness of the impairments of a real fiber (noise, mode coupling,...) the optimum coefficients $\mathbf{w}_{opt}$ are continuously and randomly changing. Nonetheless, if the system is stable, the solution will be always similar and, once the filter has converged close to those values with the DA stage, the LMS algorithm will be able to track the slight changes in the DD mode. This means that, after a good choice of the DA stage parameters, the algorithm is able to track the changes in the fiber indefinitely.

3.2 The Multiple-Input Multiple-Output architecture

All the development on Section 3.1 has been done taking into consideration a single input signal $\mathbf{x}(k)$ with which a single output $y(k)$ is generated trying to replicate as close as possible a single transmitted signal $d(k)$. In the SDM environment, as already discussed, there is an input and an output for each spatial and polarization mode, that is $2N_m$ inputs and outputs. Moreover, the output $y_i(k)$ reconstructing the i-th transmitted mode needs contributions from all the received signals to reverse the mode coupling occurred in the fiber. That is why a Multiple-Input Multiple-Output structure is needed to recover the transmitted data.

In Subsection 1.2.1, it was shown that the maximal capacity of a SDM system using as physical channel a FMF was achieved when a complete set of propagation modes was both addressed and detected. This means that the same number of inputs and outputs ($2N_m$) are involved. Nevertheless, it was pointed out too that the MIMO architecture was invented in the wireless field [6] and it can be used in many other fields. Therefore, it is fair to generalize the dimensions of the system and consider the possibility of transmitting N_t signals using different physical paths and receive $N_r \geq N_t$ signals. Again, an asymmetric structure may not make sense in the FMF environment, but it could in other fields where a MIMO structure can be implemented.

All this leads to a bank of filters like the one in Figure 3.5. The LMS filter explained in the previous Section is repeated N_t times (each row in the figure), once per transmitted signal to be recovered. In turn, each filter is divided in N_r sub-filters to calculate the contribution of each received signal to the output. This means that, the algorithm is practically a mesh of $N_r \times N_t$ causal FIR filters.

To be a little bit more clear in how the modes couple and how the dispersion between them occur, Figure 3.6 can help. First of all, it is possible to differentiate three areas.

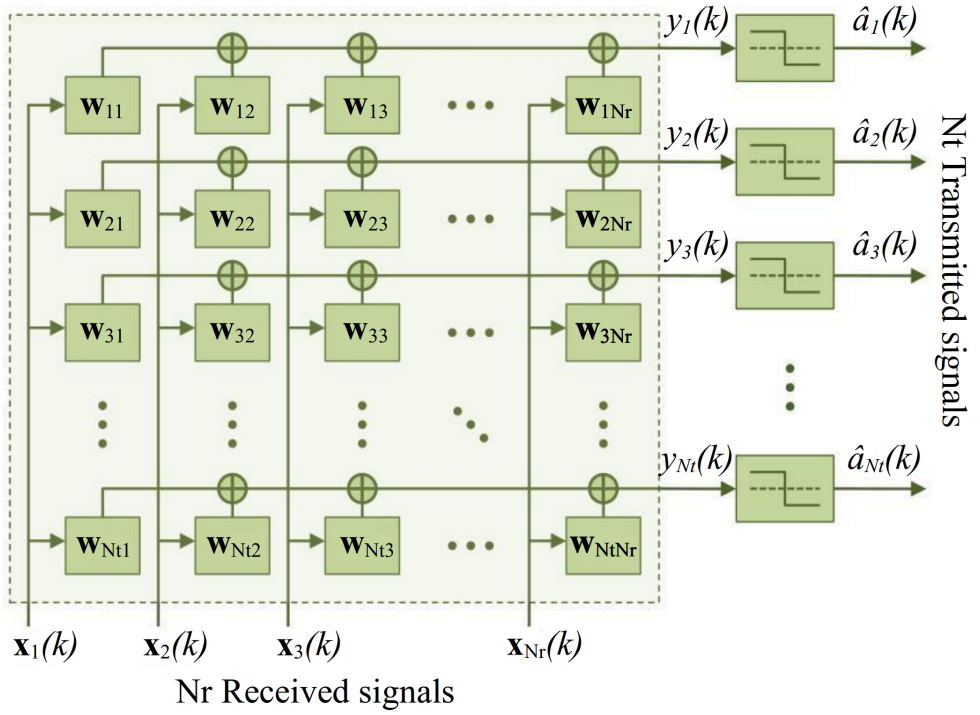


FIGURE 3.5: Generalized structure of the bank of filters needed to carry out the MIMO equalization. If we think in the LMS filters described in Section 3.1, the adaptation loop has not been represented for simplicity.

The yellow and the pink plots, correspond to the impulse responses between the sub-modes (PDM and degeneracy) of the LP_{01} and the LP_{11} modes, respectively. The blue regions show the impulse responses in the crossed paths between LP_{01} and LP_{11}.

The length of the impulse responses match the 2.6 ns dispersion of the span of fiber used in the corresponding experiment [24]. With this in mind and looking to the diagonal impulse responses (corresponding to the direct paths between transmitted and received modes), it is possible to see that for the LP_{01} and the LP_{11} modes the strong signal peak is in opposite extremes of the impulse response. This helps to understand the fact that the two modes travel with different group velocities. The low constant plateau along the impulse response indicates that the coupling between modes can occur at any moment during the propagation.

Another effect easy to understand is that the coupling within LP_{01} and LP_{11} is very strong because the impulse responses of the crossed polarization and/or degenerate sub-modes is almost the same as the impulse responses of the direct paths. For example, the impulse response between the transmitted LP_{01x} and the received LP_{01y} is almost equal to the impulse response between the same transmitted LP_{01x} and the correspondent received LP_{01x}. Conversely, between LP_{01} and LP_{11}, the coupling is much weaker and the impulse response is only a plateau without any significant peak.

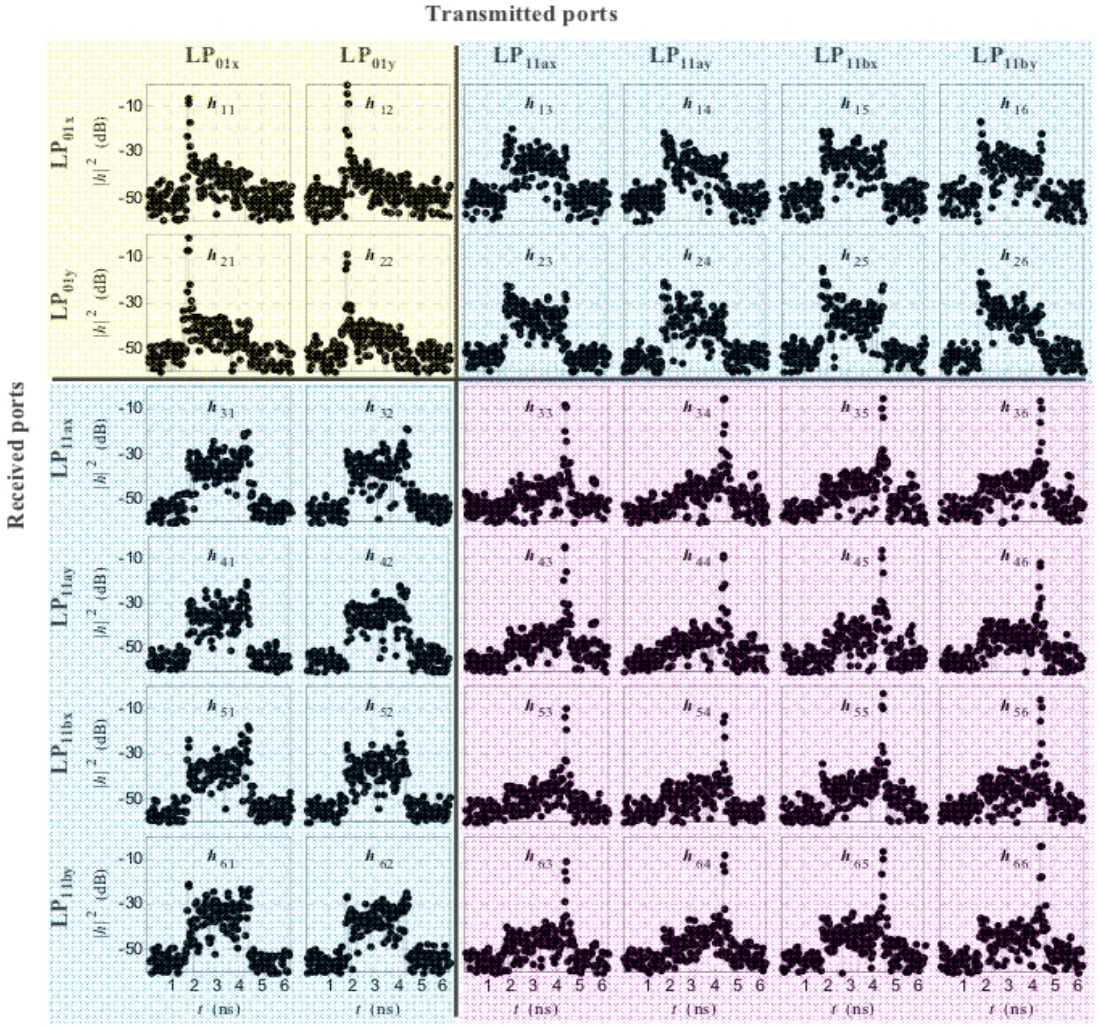


FIGURE 3.6: Square magnitude of the 6×6 impulse responses of the mode-to-mode physical channels of a 3-mode FMF with PDM estimated by the Least-Squares method. Figure from [24].

3.3 The carrier phase recovery

In phase-amplitude modulation formats like QPSK, one of the most important impairments is the phase noise. The fact that can be originated in almost any part of the communication system makes it difficult to characterize and so, difficult to fight. Anyway, the QPSK constellation presents a symmetry that helps a lot when it comes to estimate this phase error. The fourth-power algorithm take advantage of it.

The QPSK modulation format is usually classified as the simplest case of the Quadrature Amplitude Modulations. In fact, the usual next step is using 16-QAM constellations. Nevertheless, it comes from the family of the Phase Shift Keying modulations (PSK) where, although the constellation points can be represented in the complex plane, the important parameter is not the I or Q component, but the phase. In particular, in

the usual implementations, the points of the constellation are roots of the unity (or the negative unity). For instance, in the M-PSK

$$a_r = e^{j\frac{(2r+1)\pi}{M}} \tag{3.11}$$

with $r = 0, ..., M-1$. Because of this, taking the M-th power of any of the constellation symbols collapses it to

$$a_r^M = e^{j(2r+1)\pi} = e^{j\pi} = -1 \tag{3.12}$$

If now, the symbol is corrupted with some phase error $\Delta\theta$

$$(a_r e^{j\Delta\theta})^M = e^{j(\pi + M\Delta\theta)} \tag{3.13}$$

and from the phase $\pi + M\Delta\theta$ it is easy to recover the phase error. In the case of the QPSK signal, the most used one, it is necessary to take the fourth power to recover the phase error and thus the name of the algorithm.

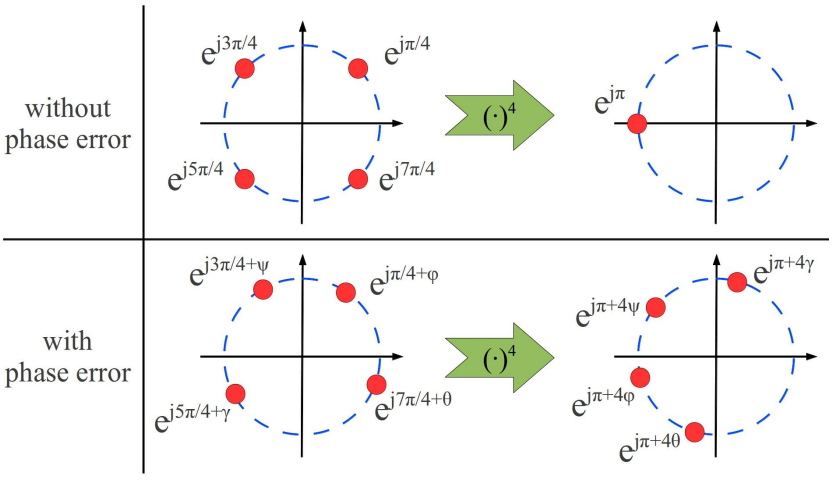


FIGURE 3.7: When the QPSK symbols are corrupted by phase error it is easy to estimate it with the fourth-power algorithm.

The fact of taking a power of a complex number, which in the phase represents a product and a 2π-module, introduces an ambiguity in the estimated phase error. For example, in the QPSK case, phase errors of θ and $\theta + \pi/2$ would lead to the same estimation θ which would be correct for the first case but not for the second. This $\pi/2$ or quadrant ambiguity can be solved by tracking the evolution of the estimation if the phase error is slow enough.

Finally, to make a smoother compensation of the phase error, a short time average of few tens of symbols is done. This introduces a delay in the processing, but it is acceptable.

3.4 Results

The algorithm explained until this point (or slightly simpler versions) has been used in all the experiments mentioned in the preceding. More or less the half of the work done for this thesis consisted in optimizing and minimally improving the algorithm and using it for the realization of the following two experiments.

3.4.1 Combined SDM and WDM transmission over 700 km of FMF

The experiment in [10] consisted in the transmission of a 20 GBaud/s QPSK signal combining both SDM and WDM. There were transmitted 34 WDM channels with a spacing of 50 GHz using a 3-mode FMF (LP_{01} and the twofold degenerated LP_{11}). Taking into account the PDM and the 2 bits/symbol of the QPSK, the aggregate line-rate was

$$\begin{aligned} R_b &= 34 \text{ WDM} \times 3 \text{ SDM} \times 2 \text{ PDM} \times 2 \text{ bits/symbol} \times 20 \text{ GBaud/s} \\ &= 8.16 \text{ Tb/s} \end{aligned} \tag{3.14}$$

As the bandwidth was 50 GHz $\times$ 34 WDM = 1.7 THz, the spectral efficiency of the link was 4.8 b/s/Hz.

The setup of the experiment used the spot couplers based in mirrors explained in Subsection 2.1.1. Moreover, it comprised a DGD-compensated recirculating loop with 70 km of Graded Index FMF obtained by the splicing of several fibers with different lengths and DGDs. The order of the spools was chosen to compensate most of the DGD after each spool pair. In this way, the maximum excursion of the DGD occurring during the whole span was 2 ns.

Spool Nr.	Length [km]	Total DGD [ns]
1	12.5	-2.00
2	25.0	1.64
3	12.5	-1.17
4	20.0	1.52
Total	70.0	-0.01

TABLE 3.1: Mode dispersion parameters of the different spools used in the DGD-compensated span.

The MIMO DSP used filters with 400 taps that, regarding the 2 samples per symbol rate and the 20 GBaud/s transmission, correspond to a 10 ns window. With this experimental setting was possible to recover the data after 700 km (10 loops) always with a maximum

observed bit error rate (BER) $< 10^{-2}$ which can be compensated with the state-of-the-art forward error correction (FEC) with a 20% overhead. Of course, the use of a FEC reduces the real spectral efficiency to 4.8 b/s/Hz$\cdot\frac{100\%}{120\%} = 4$ b/s/Hz.

These results can be seen in more detail in Figure 3.8. Each group of points corresponds to a different transmission distance. For the first two loops (140 km) almost no errors were found after the DSP in any of the PDM or SDM modes (6 columns of points of each distance) of any of the WDM channels. Increasing the distance, made the BER grow until, for 10 loops (700 km), it was just below the limit of 10^{-2}. This distance represented a record distance in a combined WDM/SDM transmission at the moment.

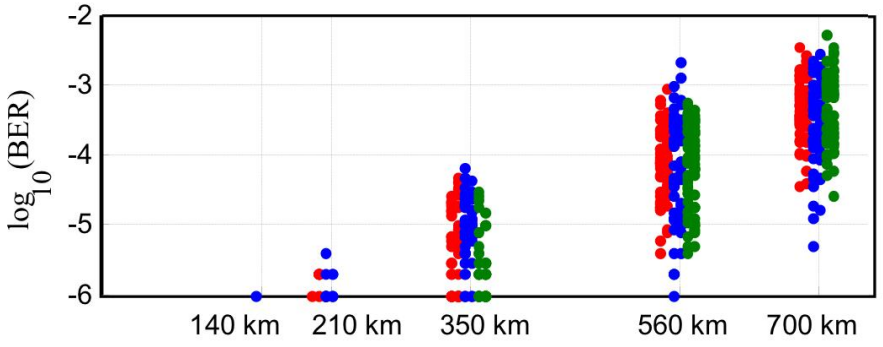


FIGURE 3.8: Each point column corresponds to the 34 WDM channels of each polarization and spatial mode. Moreover, the three spatial modes are differentiated with red, blue and green. Figure from [10].

3.4.2 12×12 MIMO transmission over 130 km of FMF

The experiment in [11] consisted in the first ever transmission using a FMF with 6 spatial modes, that is the usual LP_{01} and twofold degenerated LP_{11} plus the also twofold degenerated LP_{21} and the non-degenerated LP_{02}. Again, the transmitted signal was a 20 GBaud/s QPSK and SDM was combined with WDM (50 GHz channel spacing). In this case, the number of WDM channels was only 8.

$$\begin{aligned} R_b &= 8 \text{ WDM} \times 6 \text{ SDM} \times 2 \text{ PDM} \times 2 \text{ bits/symbol} \times 20 \text{ GBaud/s} \\ &= 3.84 \text{ Tb/s} \end{aligned} \tag{3.15}$$

The bandwidth was 50 GHz × 8 WDM = 400 GHz, so the spectral efficiency of the link was 9.6 b/s/Hz (taking into account the 20% FEC overhead accepted, 8 b/s/Hz).

As mode couplers this experiment used for the first time too the photonic lantern and the 3D-waveguide in the transmitter and the receiver sides, respectively. It was used a span of 65.3 km of Graded Index FMF that supported 6 spatial modes. In this case, the

maximum DGD in the span was reduced to 4.49 ns using 4 different spools (see Table 3.2). As shown in the previous experiment, it is possible to compensate completely the DGD but it requires a great diversity of fiber samples. In the FMF, the twofold degenerated modes travel with the same group velocity, so for three spatial modes, only one DGD (between LP_{01} and LP_{11}) has to be compensated. In the case of 6 spatial modes, there are three DGDs to compensate at the same time.

Spool Nr.	Length [km]	DGD_{01-11} [ns]	DGD_{01-21} [ns]	DGD_{01-02} [ns]
1	8.3	1.23	2.34	2.33
2	30.0	4.14	7.31	7.42
3	10.0	0.01	-1.75	-1.64
4	17.0	-0.90	-6.71	-7.01
Total	65.3	**4.49**	1.20	1.11

TABLE 3.2: Mode dispersion parameters of the different spools used in the 6-mode FMF span. The DGD between LP_{11}, LP_{21} and LP_{02} can be computed from the ones in this table. The maximum uncompensated DGD was between LP_{01} and LP_{11}.

As the accumulated DGD was growing a lot in each loop, in this experiment, the number of taps used for distance had to be increased. For the first loop (65.3 km), the usual 400 taps (10 ns) were enough. For the second loop (130.6 km) the number of taps had to be increased to 500 (12.5 ns). Finally, for the third loop (195.9 km) the filters needed 700 taps (17.5 ns) and even like this, the obtained BER exceeded the acceptable 10^{-2}. As it will be explained in the next Section, the problem was not only the mode dispersion but the convergence of the algorithm.

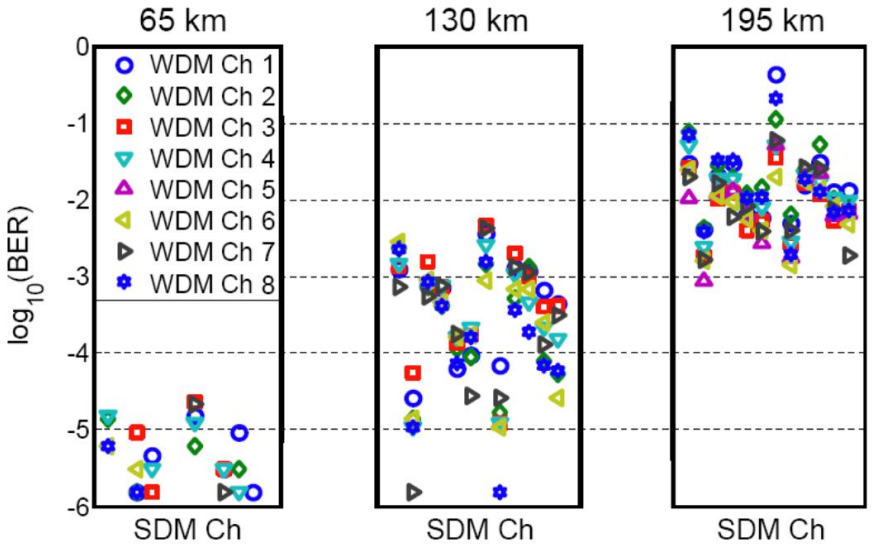


FIGURE 3.9: Each point column represents one of the PDM/SDM modes. For the third loop (195 km) the BER limit of 10^{-2} was exceeded. Figure from [11].

3.5 Limitations

As already mentioned, only the half of this thesis work was devoted to the LMS algorithm explained above. In the remainder of the time, some of the impairments of this code were detected and the new approach explained in Chapter 4 was developed. In this Section, the reasons that led to this change of viewpoint are presented.

3.5.1 Computational complexity

The experiment explained in Subsection 3.4.2 was the first that used FMF allowing the propagation of 6 spatial modes. With a higher number of modes the coupling options grow dramatically. If for the 3-mode FMF it is needed a 6×6 bank of LMS filters, for the 6-mode FMF it grows to a 12×12 structure. Of course, this affects directly the computational complexity of the algorithm because each one of the filters must be evaluated and adapted in every iteration. On the other hand, a higher number of modes also increases the difficulty of compensating the DGD. If some DGD is left, the modal dispersion grows linearly with the propagation distance and so does the number of symbols L in the windows of the equalizers.

This bad scaling of the algorithm can be easily seen calculating its computational complexity. Before starting, it is necessary to make some comments. Consider a general system with N_r received signals and N_t transmitted signals to recover. Regarding the MIMO structure it is easy to see that the calculations to reconstruct each of the transmitted signals (each row in Figure 3.5) are equivalent, so we can simplify and count the complexity to recover only the m-th transmitted signal $y_m(k)$ $(m = 1, \ldots, N_t)$.

Another consideration has to do with the sampling rate. It was already mentioned that to comply with the Nyquist sampling theorem two samples per symbol are taken. That means that the number of taps in our filters is not really L, but $2L$.

Finally, to calculate the overall complexity of the algorithm, only the complex multiplications are taken into account. This is because those are the most costly operations and constitute the major part of the algorithm. Moreover, the complexity analysis will only cover the LMS filter structure (see Figure 3.4) because it is what the frequency-domain equalizer can improve. The phase recovery will be omitted.

With all this in mind, we can divide the algorithm in three basic steps. The first one is the evaluation of the m-th filter output which consist in the scalar product between the input vector and the coefficient vector. As it has to be repeated for each input signal to calculate its contribution, it becomes

$$y_m(k) = \sum_{n=1}^{N_r} \mathbf{x}_n^T(k)\mathbf{w}_{mn}(k) \tag{3.16}$$

The computational complexity of this part is of $2LN_r$ complex multiplications. The second step of the algorithm is the computation of the error. For the m-th output, it is

$$e_m(k) = d_m(k) - y_m(k) \tag{3.17}$$

which does not include any complex multiplication, so it does not contribute to the computational complexity. Finally, every iteration ends with the adaptation of the filters. This is,

$$\mathbf{w}_{mn}(k+1) = \mathbf{w}_{mn}(k) + 2\mu \mathbf{x}_n^*(k) e_m(k) \tag{3.18}$$

for $n = 1, \ldots, N_r$. The only multiplications of this stage are the product of the conjugated input signal with the error: $2LN_r$ complex multiplications. In total, the complexity of the time-domain LMS equalizer for every transmitted signal, $y_m(k)$, is

$$C_{TDE} = 2LN_r + 2LN_r = 4LN_r \tag{3.19}$$

As already stated in the introduction of the present Section, the complexity is linear with the number of received modes N_r and with the length of the filters $2L$. Nevertheless, the latter is indirectly dependent on the number of modes too, so the scaling of the code is not linear as we could expect. In addition, it is important to remember that this complexity corresponds only to the recovery of the m-th transmitted signal and usually all of them will be computed. This means that another factor N_t will be always involved, although this one is inherent to the MIMO architecture and nothing can be done about it.

It is clear that the time-domain equalizer does not scale very well, but this is not its only problem. In the experiment of Subsection 3.4.2, it was explained that for the third loop, the 700-taps filters needed to cover all the modal dispersion did not work as expected. Taking a look to Figure 3.10, it can be seen that the length of the impulse response grows linearly with the distance. For the first loop, it is ~ 5 ns long (in fact, it should match the 4.49 ns of total DGD of the fiber). For the second loop it doubles to ~ 10 ns, and for the third loop it is ~ 15 ns.

Translating those times in terms of number of symbols (and taps) for a 20 GBaud/s (40 Gsamples/s) signal gives the following.

Loops	IR length	Symbols	Taps
1	~ 5 ns	~ 100	~ 200
2	~ 10 ns	~ 200	~ 400
3	~ 15 ns	~ 300	~ 600

Of course, as it is complicated to perfectly synchronize all the system, it is always advisable to take some more taps and make sure that the signal fits correctly in the filter. In fact, for the first loop, 400 taps were used as it was the usual value in the 3-mode FMF experiments and it was way enough. For the second loop, 500 taps were taken and, for the third, 700. Unfortunately, for the third loop, although all the LMS parameters were optimized, the convergence of the filter was very bad and so was the BER obtained.

For example, taking a big value for the adaptation gain μ led to the miss-convergence of the algorithm. Intuitively, it can be explained saying that the algorithm can deal only with certain amount of change per iteration. With so many taps, even small μ values exceed that allowed amount and the filter taps never converged. On the other hand, if the value of the adaptation gain was taken very small, the algorithm converged but so slowly that it was not acceptable.

In the end, the best performance was achieved with only 600 taps although, as stated above, it is very difficult that all the signal was fit in that filter length. This means that the time-domain algorithm, apart from a bad scaling, has a convergence problem.

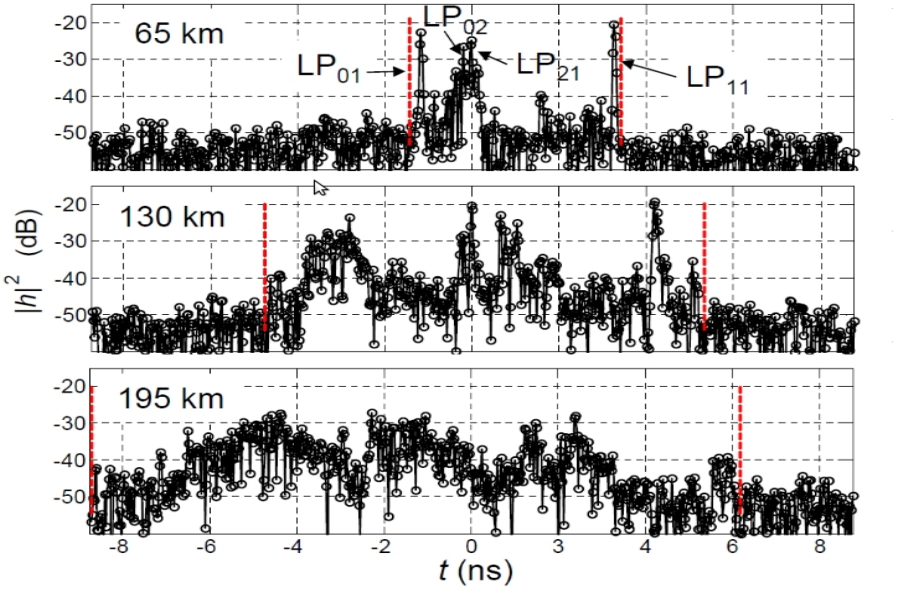


FIGURE 3.10: Impulse response of the 6-mode FMF for 65 km, 130 km and 195 km. With 6 spatial modes, it is not possible anymore to distinguish strong and weak coupling between modes by inspecting the impulse response of each physical path.

One idea to solve this problem may be to make a block-wise version of the adaptive LMS filter. This is to adapt the filter only once every block of M symbols. That would lighten the complexity of the algorithm and, although adapting less often should slow down the convergence, it may improve the quality of this convergence. This version was not tried, but the complexity is easy to compute. The evaluation of the filter would be done at each iteration anyway ($2LN_r$ complex multiplications) and the adaptation only once every M iterations ($2LN_r/M$ complex multiplications), so

$$C_{B-TDE} = 2LN_r + \frac{2LN_r}{M} = 2LN_r\frac{M+1}{M} \xrightarrow[M\to\infty]{} 2LN_r \tag{3.20}$$

After all, the complexity keeps the same dependency with N_r and L, so it scales in the same way. As we will see in Chapter 4, the frequency-domain equalizer is capable of much more.

3.5.2 Carrier phase recovery feedback loop

Another problem of the current algorithm is related with the phase noise. In the DD stage of the LMS filter, when it comes to decide which point of the constellation corresponds to a given output $y_m(k)$, if the phase noise is very big the decision can not be done properly. The LMS filter is able to remove the AWGN noise, the mode coupling and dispersion and may be able to compensate small amounts of phase noise, but the latter is not its job. The fact that the carrier phase recovery (CPR) comes after the decision penalize a lot the performance. Another possibility is to put the CPR before the filters, but then it is the phase recovery that does not work because of the presence of all the noise and mode coupling.

The implemented solution is halfway: to put the CPR just after the evaluation of the filters but before the decision. Nevertheless, as the decision is not done exactly with the output given by the filter, in order to adapt it properly it is necessary to make a phase correction in the LMS. That is why the CPR has to take a feedback-like shape as the shown in Figure 3.11. Because of the small average of the CPR, the whole algorithm suffers a delay, but again, as commented in Section 3.3, it is acceptable.

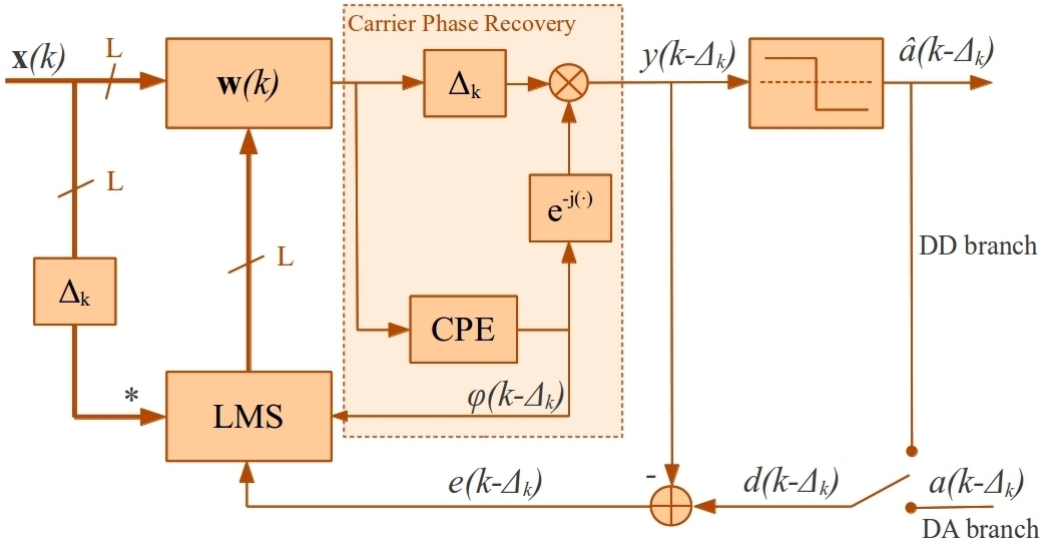


FIGURE 3.11: The CPR is placed within the adaptation loop of the LMS filter. The block CPE (carrier phase estimation) corresponds to the fourth-power algorithm.

Although this solution worked and the algorithm was used in all the experiments explained in the preceding, the fact of having a feedback loop inside the adaptation loop is not advisable. If the signal to be processed contains big phase errors the feedback loop can produce jumps in the error calculation of the LMS algorithm and make it unstable. The average in the CPR was included just to smooth this effect as much as possible but this was still not a good solution to the lack of phase information for the decision. As it will be seen in Chapter 4, it could be solved in an alternative way.

Chapter 4

The Frequency Domain and the Constant Modulus algorithm

In the present Chapter we will present solutions to the limitations described in Section 3.5. In the first Section we will discuss an efficient frequency version of the LMS algorithm that will allow a dramatic decrement of the computational complexity. In Section 4.2, the Constant Modulus algorithm and its variations will be introduced as the solution for the phase noise problem of Subsection 3.5.2. Finally, as the work of this thesis finished before showing the performance of the new algorithm with an experiment, a result obtained afterward by the research group will be explained in Section 4.3.

As in all the preceding, the mathematical development of this Chapter will be done in the maximum generalized possible way so it can be extended to any multiplexing and even modulation scenario that could be used in the future. The work in this thesis was limited to a certain period of time but was included in a continuous research to increase the capacity and performance of the coherent optical communication systems. This means that, in every moment, it is thought as an open and extendable work.

4.1 The Frequency-Domain LMS algorithm

The first question that may arise when facing the frequency adaptation of the time-domain (TD) LMS algorithm is why it can give such an improvement in the complexity without losing in performance. The answer is found in the equivalence of the time-domain linear convolution, a very costly operation, with the frequency-domain (FD) multiplication, much more simpler. As it will be seen, this is only an advantage in a block-wise structure where the logarithmic scalability of the FFTW algorithm [16] makes a difference.

However, before entering in the details, it is important to make a comment on how the oversampling of the signal affects the structure of the LMS algorithm and how it is translated to a block-wise frequency environment. In Subsection 4.1.1 it will be explained how this oversampling can be understood as yet another diversity dimension. After that, in Subsection 4.1.2, the main ideas on the adaptation to the FD will be explained. Finally, in Subsection 4.1.3 the computational complexity of the new FD algorithm will be calculated and compared with those on Section 3.5 for the TD equalizer.

4.1.1 Fractional space diversity

From the beginning, it was explained that the detected signals are sampled with a rate doubling the transmission speed. Again, this is to comply with Nyquist's sampling theorem. Of course, although the input of the system is oversampled, only one output per symbol is calculated. With this in mind, the contribution of the two samples taken in a symbol period can be thought almost in the same way as the contribution of the samples from different received modes. In the latter, different combinations of those contributions lead to the reconstruction of different transmitted signals. Instead, in the case of the oversampling, there is only one satisfactory combination. However, from the DSP point of view, the oversampling can be seen as another diversity dimension.

The implication of this interpretation is that Eqs. (3.16) and (3.18) can be rewritten as

$$y_m(k) = y_m^e(k) + y_m^o(k) \tag{4.1}$$

where

$$\begin{aligned} y_m^e(k) &= \sum_{n=1}^{N_r} \mathbf{x}_n^{eT}(k)\mathbf{w}_{mn}^e(k) \\ y_m^o(k) &= \sum_{n=1}^{N_r} \mathbf{x}_n^{oT}(k)\mathbf{w}_{mn}^o(k) \end{aligned} \tag{4.2}$$

and

$$\begin{aligned} \mathbf{w}_{mn}^e(k+1) &= \mathbf{w}_{mn}^e(k) + 2\mu\mathbf{x}_n^{e*}(k)e_m(k) \\ \mathbf{w}_{mn}^o(k+1) &= \mathbf{w}_{mn}^o(k) + 2\mu\mathbf{x}_n^{o*}(k)e_m(k) \end{aligned} \tag{4.3}$$

The subscripts e and o are referred to the even and odd elements of the original vectors. For example, $\mathbf{x}$ is divided in $\mathbf{x}^e = [x(k), x(k-2), x(k-4), \dots]^T$ and $\mathbf{x}^o =$

$[x(k-1), x(k-3), x(k-5), \dots]^T$. Therefore, each one of the $2L$-tap filters used to compute the m-th output of our system can be divided in two L-tap filters, one for the even samples and the other for the odd samples.

This change of structure may seem innocuous, but breaks with the mental idea of having the sampling device phase-synchronized with the transmitter, that is, having always one sample of each pair aligned with the center of the symbol and the other one aligned with the transition between symbols. That usually does not happen, the two samples are taken at arbitrary instants so both samples in each symbol period have the same value in the recovery of the data. And what is more important, this is not only a philosophical discussion. When switching from a TD symbol-wise algorithm to a FD block-wise algorithm it becomes important.

In the first case, each iteration produces one only output $y_m(k)$ and, therefore, one only error $e_m(k)$. This means that when it comes to adapt the filter taps it is clear which error value to use in Eq. (3.18). Even in the block-wise case, the adaptation of the filter taps could be done in many different ways, but it is clear that for every couple of input samples and every couple of filter taps, there is one only output and one only error.

When moving to the frequency domain, the first step is to calculate the FFT of the input block (suppose M symbols, ie. $2M$ samples). The resulting frequency input is a vector with $2M$ elements. After all the algorithm, independently of what it does, the output must be a vector with only M equalized samples trying to match the M transmitted symbols. From them, only M errors can be computed. This situation seems very similar to the TD scenario but now there is not anymore a clear way to relate pairs of input samples (or pairs of filter taps) with each output (or error) so it is not clear how to perform the adaptation.

To solve this problem, it may be possible to double our error samples in some inventive way (padding with zeros, repeating samples,...) but in the frequency domain those operations may not have a clear mathematical meaning. Instead, in our code, the even and the odd parts of the input signal are separated in advance and are managed like completely independent contributions to the output. In the next Subsection, the FD equalizer will be explained from the beginning as a filter to treat each of these oversampling parts of each polarization and modal received signal. This means that the input block will be just a 1 SPS signal and all the dimensions in the filter will correspond to the size of the block in symbols, M.

4.1.2 Overlap & Save method

The frequency-domain equalizer we are about to explain is based in the same mathematical algorithms as the time-domain equalizer. In fact, in Figure 4.1 it is easy to identify

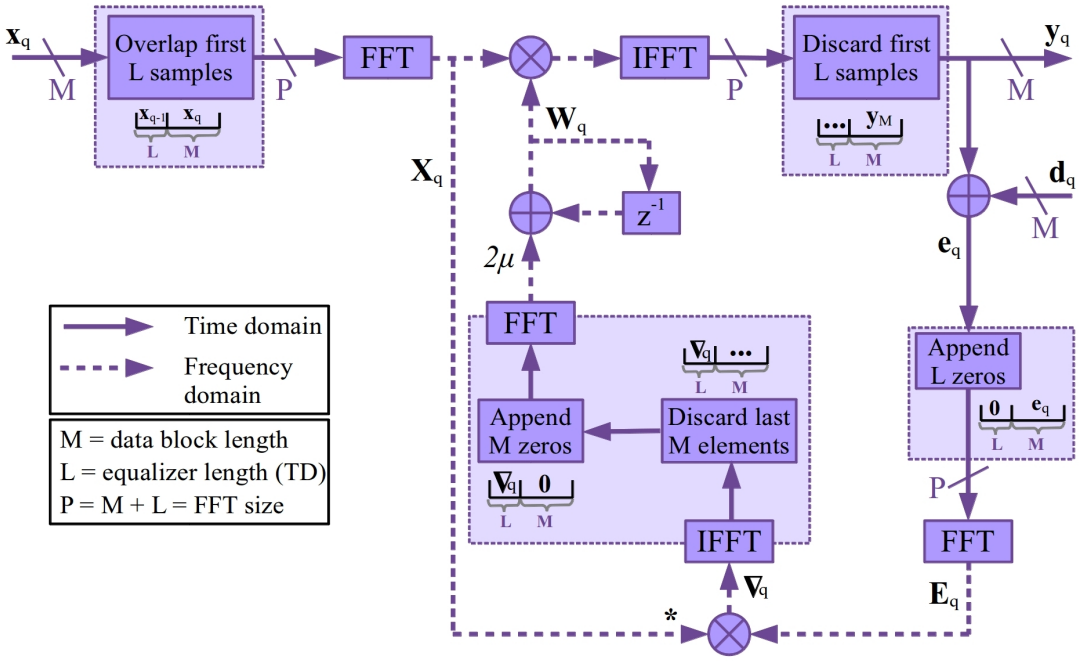


FIGURE 4.1: Block diagram of the FD equalizer for one of the oversampling parts of one polarization/spatial mode. The shadowed blocks indicate the necessary tricks to make the FD operations equivalent to the TD ones.

the LMS operations of Eqs. (3.16) and (3.18). That is evaluating the equalization, computing the error, multiplying it by the conjugate of the signal, scaling by the adaptation gain μ and adding the previous filter taps. What makes it special is the fact of doing the calculations in the FD because, as it will be shown, it is not straightforward to keep the equivalency of the operations.

The main trick used in the algorithm is called either Overlap & Save or Overlap & Discard method and helps in the calculation of discrete linear convolutions. To start, it will be analyzed how it works in the equalization step. In the TD, forgetting about the fact that the filter taps are adapted at each iteration, the calculation performed is:

$$\mathbf{y} = \mathbf{x} * \mathbf{w} = \mathcal{F}^{-1}\{\mathcal{F}\{\mathbf{x}\} \cdot \mathcal{F}\{\mathbf{w}\}\} \tag{4.4}$$

where $\mathbf{x}$, $\mathbf{w}$ and $\mathbf{y}$ are infinite vectors for the input, the filter taps and the output, respectively[1]. Moreover, $\mathbf{w}(k) = 0\ \forall k \notin [1, \ldots, L]$ as it is a FIR filter, $*$ is the linear convolution, $\mathcal{F}$ is the discrete-time Fourier transform and $\cdot$ denotes the element-wise

[1]In this Chapter, the notation is though. As in the preceding, *italics* is used for the scalars (including particular elements of a vector) and **bold print** is used for vectors. However, the lower-case and the UPPER-CASE denote that a variable belongs to the time domain or to the frequency domain, respectively (there are no matrices, only vectors). About subscripts and ordinals, a number in (brackets) represents the position of a scalar in a vector or the initial element of a vector (depending on the print used for the name of the variable, *italics* or **bold**); an $_{UPPER-CASE\ SUBSCRIPT}$ indicates the size of a vector; instead a $_{lower-case\ subscript}$ indicates the relation of the variable with a particular block iteration within the block-wise structure of the FDE.

multiplication. Instead, in the FD, a block $\mathbf{x}_P$ of the input signal (with $P = M + L$ and $M \gg L$) is transformed with the Fast Fourier Transform (FFT) and multiplied by a finite version of the filter $\mathbf{w}$, $\mathbf{w}_P = [\mathbf{w}_L, \mathbf{0}_M]$. This is

$$\mathbf{y}_P = IFFT\{FFT\{\mathbf{x}_P\} \cdot FFT\{\mathbf{w}_P\}\} = \mathbf{x}_P \otimes \mathbf{w}_P \tag{4.5}$$

where $\otimes$ is the circular convolution of two finite vectors. Of course, the linear and the circular convolution are not equivalent. However, if $\mathbf{x}_P$ is a segment of $\mathbf{x}$ corresponding to the q-th block, this is, starting at $n = qM$,

$$y_P(k) = y(n) \tag{4.6}$$

for $k \in [L+1, \ldots, P]$ and $n \in [qM + L + 1, qM + P]$. This is, for the last M samples of the output, the linear and the circular convolution coincide. Choosing each input block $\mathbf{x}_P$ starting at $n = qM - L$ (this means overlapping the first L samples with the previous block, ie. $\mathbf{x}_P = [x(qM - L), \ldots, x(qM - 1), x(qM), \ldots, x(qM + M - 1)]$) it is possible to take advantage of this fact. The corresponding outputs $\mathbf{y}_M = [y_P(L + 1), \ldots, y_P(P)] = [y(qM + 1), \ldots, y(qM + M)]$ can be concatenated to reproduce the desired linear convolution $\mathbf{y}$ from Eq. (4.4). This process can be better understood in Figure 4.2.

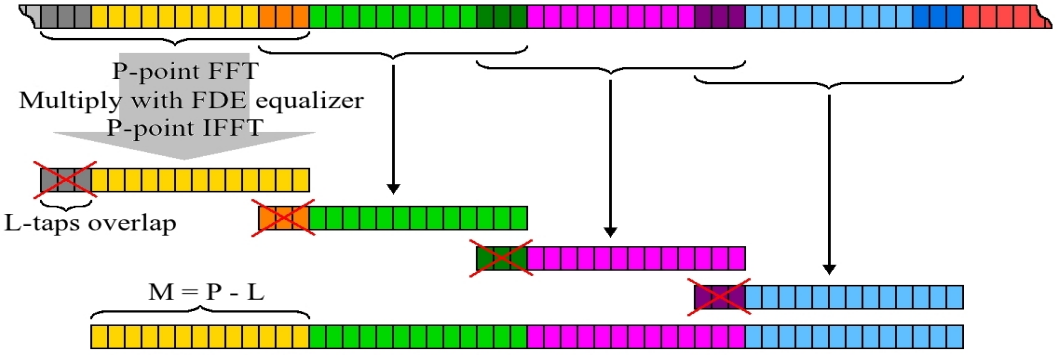


FIGURE 4.2: In each iteration, L extra overlapped samples are taken in the input but only the M output samples of interest are saved. The other L output samples are discarded. Hence the two names of the method.

It is important to remember that the length of the filter L is chosen to compensate all the modal dispersion occurred in the channel. On the other hand, to make the algorithm really efficient, $M \gg L$ and $P = M + L$ has to be a power of 2 in order to take advantage of the best performance of the FFTW algorithm. However, M can not be arbitrarily high because it determines how often the filter taps are updated and it has to be slower than the channel variation speed.

For example, in the first tests of this algorithm, data from the experiment in [11] was used. Remember that for the TDE, the filters had 400 taps for the first loop data and

500 taps for the second loop data. As it was in a 2 SPS scenario, for the FDE, 200 taps and 250 were taken, respectively. To comply with all the conditions, the values chosen for P were 2^9 and 2^{10}.

After each filtering iteration, it is the time to update the equalizer taps which, again, needs the use of the Overlap & Save method ideas [25, 26]. First of all, with the output $\mathbf{y}_M$ and the corresponding desired signal samples, the error vector $\mathbf{e}_M$ is computed. Considering a block-wise structure like the one needed in the FD equalizer, the TDE Eq. (3.18) after processing the q-th block would become

$$\mathbf{w}_{q+1} = \mathbf{w}_q + 2\mu \sum_{i=0}^{M-1} e(qM+i)\mathbf{x}^*(qM+i) = \mathbf{w}_q + 2\mu\nabla_q \tag{4.7}$$

where $\mathbf{x}(qM+i) = [x(qM+i), \ldots, x(qM+i-L+1)]^T$ and $\mathbf{w}_q$ refers to the filter taps in the TD for the q-th block. This means that the j-th element of the vector ∇_q can be written as

$$\nabla_q(j) = \sum_{i=0}^{M-1} e(qM+i)x^*(qM+i-j) \tag{4.8}$$

for $j = 0, \ldots, L-1$. It is important to note that ∇_q has only L elements in the TD. As it is the updating term for $\mathbf{w}_q$, it has the same structure and needs to be filled with M zeros. Eq. (4.8) shows that the elements of ∇_q are given by the cross correlation of the error sequence with the filter input. This operation can be done with a simple multiplication in the frequency domain but the two factors must be prepared. For the input, it is important to take into account all the elements present in Eq. (4.8). That is

$$\mathbf{x}_q^T = [x(qM-L), \ldots, x(qM-1), x(qM), \ldots, x(qM+M-1)] \tag{4.9}$$

In the case of the error vector $\mathbf{e}_M$, it is filled, with L zeros at the beginning to keep the time relation with the elements in $\mathbf{x}_q$

$$\mathbf{e}_q = [\mathbf{0}_L, e(qM), \ldots, e(qM+M-1)] \tag{4.10}$$

With these two vectors, it is possible to calculate ∇_q but, as in the equalization process, some elements of the output will be discarded. In this case, the components that correspond to the desired result are the first L

$$\nabla_q = \text{ first } L \text{ terms of } IFFT\{FFT\{\mathbf{e}_q\} \cdot FFT\{\mathbf{x}_q^*\}\} \tag{4.11}$$

Finally, the tap updating equation performed in the frequency domain is

$$\mathbf{W}_{q+1} = \mathbf{W}_q + 2\mu FFT\{[\nabla_q, \mathbf{0}_M]\} \tag{4.12}$$

Taking as initial tap vector in the FD $\mathbf{W}_0 = FFT\{\mathbf{w}_0\}$ where $\mathbf{w}_0$ has the last M components zero, Eq. (4.12) is an exact implementation of Eq. (4.7) in the frequency domain [25]. This means that the adaptation of the filter taps follows the same path as the block-wise TDE and, thus, the performance of the algorithm is the same. The only important change, as will be shown in the next Subsection, is its operational speed.

4.1.3 Complexity analysis

Now that the FD LMS algorithm has been explained in detail, it is time to compute its complexity and compare it with the TDE. To count the operations of the algorithm, the same criteria as in Section 3.5 will be followed. Moreover, the number of complex multiplications needed to compute an FFT (or its inverse) with size N being a power of 2 is $\frac{N}{2}\log_2(N)$. This corresponds to the algorithm in the library FFTW, which implements the Cooley-Tuckey method with radix-2 [16]. Finally, recalling that the computation of the error does not contribute to the complexity as defined herein (no complex multiplications involved), the FD LMS can be divided in two main parts: the equalization of the input samples and the updating of the filter taps.

The first of this parts starts with a size-P FFT. As the filter structure explained above is repeated because of the oversampling ($\times 2$), the PDM ($\times 2$) and the SDM ($\times N_r$) the total complexity is $(4N_r)\frac{P}{2}\log_2(P)$. After the FFT, the input samples are multiplied by the filter taps. This operation is an element-wise product of length-P vectors and again is repeated $4N_r$ times, so its total computational complexity is $(4N_r)P$. To finish this stage, the FFT has to be undone. However, here, instead of repeating it $4N_r$ times, as the FFT is a linear operation, it is more efficient to first add all the oversampling, PDM and SDM contributions and then do only one IFFT. Therefore, the complexity of this step is just $\frac{P}{2}\log_2(P)$. In summary, the complexity of the equalization is

$$C_{FDE\,equalization} = \left[(4N_r + 1)\frac{P}{2}\log_2(P) + (4N_r)P\right]\frac{1}{M} \tag{4.13}$$

where a factor $\frac{1}{M}$ has been added because each iteration gives M outputs at once.

The second part of the algorithm, the adaptation of the filter, starts with the transformation of the error vector. It is a length-P vector and is the same for all the oversampling, PDM and SDM branches so the complexity of this operation is only $\frac{P}{2}\log_2(P)$. Afterward, the error vector is element-wise multiplied by the input vector conjugated. With this operation, the algorithm separates again all the contributions, so the factor $4N_r$

appears again. This means a total of $(4N_r)P$ complex multiplications. The next step is the zero-padding of the vector ∇_q. As this is done in the TD, a FFT and an IFFT are performed for each filter. This means a total complexity of $2(4N_r)\frac{P}{2}\log_2(P)$. Finally, to finish the adaptation, it is necessary to multiply ∇_q by 2μ. As this is a constant, in a hardware implementation it can be performed much more efficiently than a simple complex multiplication so it will not be counted. In any case, its impact would be minimal. To sum up, the complexity of the filter tap adaptation is

$$C_{FDE\,adaptation} = \left[(8N_r + 1)\frac{P}{2}\log_2(P) + (4N_r)P\right]\frac{1}{M} \tag{4.14}$$

Adding the two terms, the total complexity of the FD LMS explained in this thesis is

$$C_{FDE} = \left[(6N_r + 1)\log_2(P) + (8N_r)\right]\frac{P}{M} \tag{4.15}$$

In Figure 4.3, the complexity of the TDE (and its block version) is compared with the complexity of the FDE. The plot has been obtained for the case of $N_r = N_t = 6$, with a FFT size of $P = 2^9$ and sweeping the size of the filter L until 400 taps (remember that the most used value is 200 taps). It is clear that the scalability of the FD algorithm is much better than its TD counterpart. Although it is also linear with the number of modes N_r, its relation with the size of the filter comes with the term $\frac{P}{M} = \frac{M+L}{M}$. When $L \ll M$, this factor is almost a constant. When L grows, the relation becomes less advantageous. However, if the channel admits a longer FFT size P, the advantage can be restored.

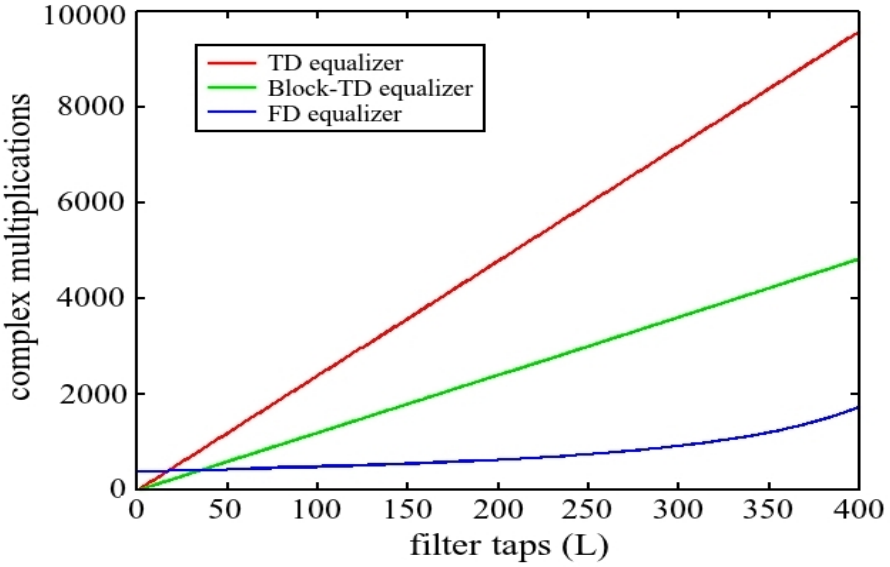


FIGURE 4.3: The FD equalizer is less efficient for very short filter sizes. However, for $L \approx 16$ and $L \approx 34$ it already overcomes the TD equalizer and the Block-TD equalizer, respectively. Usually, L is in the order of few hundreds.

4.2 The Constant Modulus algorithm

In Section 3.5, it was pointed out the problem caused by the phase noise during the filter tap adaptation. To compensate the lack of precision in the calculation of the error, the CPR had been placed within the adaptation loop. This solution meant having a double loop with a feedback and, although it was working with most of the laboratory signals (with controlled phase noise), a better solution was needed to avoid possible instabilities.

Analyzing the problem in depth, it is possible to see that what made that important the phase error is the fact of using the LMS as adaptation algorithm. As explained in Chapter 3, descent gradient algorithms are based in the minimization of a given cost function. However, this cost function can be derived from any imaginable error criterion. In fact, only two conditions are imposed. The function has to be well defined and differentiable all over its domain. In the case of our FIR filter, the cost function depends on its taps (the only parameters we have to describe the filter), so the domain is the L-dimensional complex linear space where the taps can take values.

With all this freedom, the purpose of the cost function is to define what parameters of the output are taken into account to improve the taps of the filter. Lots of criteria can be used and the LMS algorithm takes the most general possibility in which both the amplitude and the phase of the recovered symbol are compared with the desired symbol. Recall that the cost function, given in Eq. (3.3), is

$$J(k) = E\{|e(k)|^2\} \tag{4.16}$$

where $e(k) = d(k) - y(k)$. The fact that the phase is included in the definition of the error makes the algorithm very sensitive to the phase noise and forces to assume a solution like the one explained in Section 3.5. The key to solve this problem is to find another cost function less sensitive to the phase noise.

The first solution that was considered was the use of the Constant Modulus algorithm (CMA). As its name indicates, this method uses a cost function that compares the amplitude of the output signal with a constant modulus, a predefined value characteristic of the modulation constellation. This means changing the definition of the error $e(k)$ by the following expression [27]

$$e(k) = |y(k)|^2 - R^2 \tag{4.17}$$

with R depending on the constellation points as

$$R = \sqrt{\frac{E\{|a_k|^4\}}{E\{|a_k|^2\}}} \tag{4.18}$$

In this case, the adaptation equation becomes [27]

$$\mathbf{w}(k+1) = \mathbf{w}(k) - 4\mu\mathbf{x}^*(k)y(k)e(k) \tag{4.19}$$

In fact, the CMA is a well known and widely used algorithm. In optical communications itself, lots of systems combines QPSK signals with CMA in the receiver. This is a great choice because all the points in the QPSK constellation share the same amplitude and so the algorithm can compensate the noise easily. Afterward, a carrier phase recovery algorithm takes care of the phase error and the signal is recovered with a great performance.

Nevertheless, along with the SDM, another technique that should allow an increment of the channel capacity is the use of higher order modulation formats. For example, in the experiment of [11] the same transmission system was used with a 16-QAM signal. Unfortunately, the performance was not good enough to publish the results. Although the CMA also works with this type of modulations, its performance is not as good because the reference radius is not the radius of every point in the constellation but an average of them (see Figure 4.4). The better behavior of the LMS algorithm with multi-modulus constellations was the main reason to choose it.

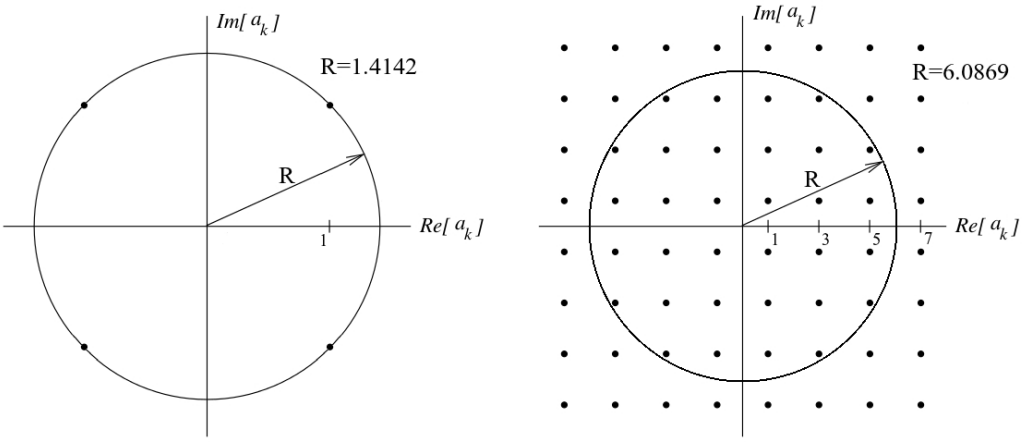


FIGURE 4.4: The CMA radius only matches the constellation points radii for QPSK. For bigger QAM constellations, the points have different modulus and the reference radius is an average.

From the basis of the CMA, a lot of multi-modulus versions have appeared [28, 29]. With the general name of Multi-Modulus Algorithm (MMA) they try to address the problem of extending the useful CMA to higher order modulation formats. Nonetheless, up to now, no one of this versions has reached the acceptance of CMA and the problem is still open. In our case, the MMA version that we tried was based in a previous decision about the ring to which the point in question belongs. This means that before calculating the error, the value of R is chosen from a closed set representing the different radii of the points in the constellation.

For instance, in the case of the 16-QAM, there are three rings of points with different radii. If we give to the I and Q components values from the set $\{-3, -1, 1, 3\}$, the four inner points have a radius $R_1 = \sqrt{2}$, the four outer points have a radius $R_3 = \sqrt{18}$ and the remaining eight points have an intermediate radius $R_2 = \sqrt{10}$. After filtering the input, the radius R_1, R_2 or R_3 is chosen depending on the modulus of the output and used instead of R in the classical CMA cost function.

$$e(k) = |y(k)|^2 - R_i^2 \tag{4.20}$$

where $i = 1, 2, 3$.

The performance of this MMA algorithm was not as good as expected. Once the filter taps are already close to the optimum, the algorithm tracks perfectly the variations of the signal. However, its initial convergence is very poor. During the first iterations, a big amount of wrong decisions about the radius are done and the convergence suffers a huge penalization. On the contrary, an interesting trick was found. Using the LMS cost function during the convergence stage and then switching to the MMA algorithm did work very well. That became the key to solve the CPR feedback loop problem.

As described in Section 3.1, the LMS algorithm implementation is divided in two stages. In the first one, the DA stage, a training sequence is used to help the filter taps converge to the Wiener optimum filter. After a while, the algorithm switches to the DD stage in which it becomes blind. In addition to the LMS, the CPR is included inside the adaptation loop to provide phase information to the decision and obtain a better error value. However, this phase information is only needed in the DD mode because in the DA stage the training sequence gives already a perfect match with the desired signal. This means that the feedback loop is only necessary in the DD stage or, what is the same, once the equalizer taps are already close to the optimum.

With this in mind, the final solution is a combination of the LMS algorithm and the CMA/MMA and, again, it is divided in two stages. In the Data-Aided mode, the standard LMS is used with the training sequence as desired signal. After a fixed period of time, the equalizer taps are close to the optimum and the algorithm switches to a DD stage. However, in this case, the blind adaptation is done using the CMA or the MMA depending on the modulation format of the system (QPSK or bigger QAM). Figures 4.5 and 4.6 show the final architecture of the algorithm. For simplicity, the block diagrams represent the case of the symbol-wise TDE but, of course, the new error calculation is straightforward adapted to the block-wise FD paradigm.

As most of the time the equalizer works in the DD mode, which now uses the CMA/MMA, it is fair to analyze if this change affects the overall computational complexity. Looking through the equations, the only change is found in the adaptation step (Eq. (4.19)). The

term $e(k)$ of the LMS, which did not contribute to the computational complexity because of its lack of multiplications, is substituted by a new error term $\tilde{e}(k) = y(k)e(k) = y(k)(|y(k)|^2 - R^2)$, which do involve multiplications. In particular, two multiplications, one to compute $|y(k)|^2$ and another to compute $e(k)y(k)$. Calculating R^2 is not necessary as it is a constant that can be saved in the memory.

In the block-wise structure of the FDE, this means that $2M$ additional multiplications are performed per iteration or, what is the same, 2 more multiplications are performed per output. This complexity increment is a constant, not a term linear with N_r or L, so it does not affect the scalability of the algorithm (neither in the TD nor in the FD) which is the main concern. The new computational complexity is

$$C_{FDE-CMA} = [(6N_r + 1)\log_2(P) + (8N_r)]\frac{P}{M} + 2 \tag{4.21}$$

As the phase information is not needed in any of the stages, the CPR feedback is eliminated and the conventional CPR explained in Section 3.3 is placed totally unlinked just after the equalizers. For modulation formats different than the classic QPSK, the CPR has to be changed because the fourth-power algorithm cannot take advantage anymore of the unity-root nature of the constellation. Nonetheless, this aspect is out of the scope of this thesis.

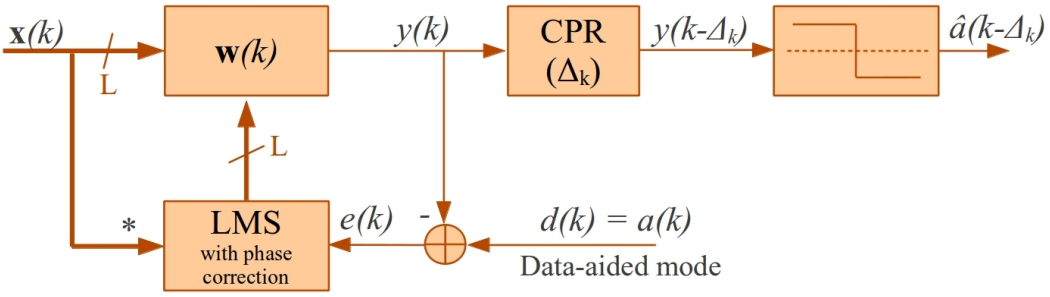


FIGURE 4.5: In the DA stage, the filter taps converge with the LMS algorithm using the phase information from the training sequence. The CPR feedback loop is eliminated.

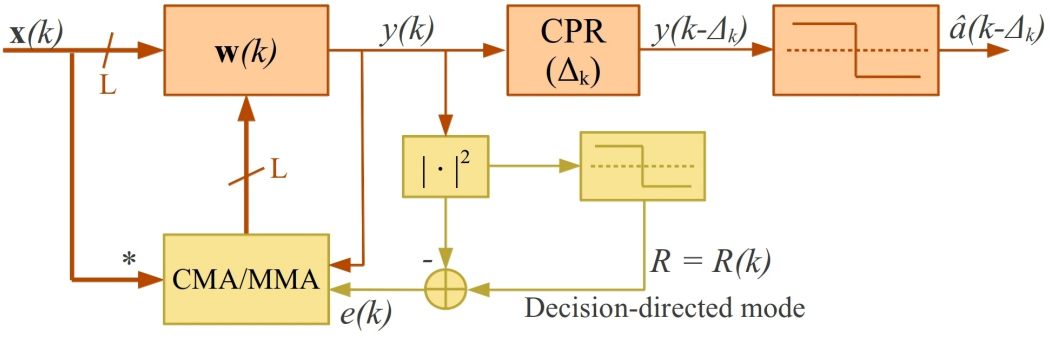


FIGURE 4.6: In the DD stage, the updating algorithm is the CMA or the MMA depending on the modulation format of the system. These methods do not need phase information so, again, the CPR feedback loop is eliminated.

4.3 Results

It was shown in Section 4.1 that the performance of the FDE is exactly the same as the block-wise TDE as all their equations are equivalent. On the other hand, the performance of the symbol-wise TDE and its block-wise counterpart are very similar [30]. They essentially minimize the same cost function and for wide-sense stationary signals their miss-adjustment with the Wiener optimum solution and their convergence speed is identical. The main difference is that the maximum value of the adaptation gain μ so that the algorithm is stable is scaled down by the block size M.

This means that although all the results cited in the preceding were obtained with the symbol-wise TDE explained in Chapter 3, the performance expected with the new FDE is similar. In fact, before the end of the work for this thesis, the first tests with the FDE algorithm using data from the experiment in [11] were performed with very promising results. As expected, the speed of the algorithm increased dramatically and, with just a slight parameter optimization, the BER was close to the obtained with the TDE. Unfortunately, the time limitation of the thesis did not allow further tests and new experiments that could prove the real advantage provided by the algorithm.

To compensate this lack of results, and taking advantage of the time inverted to write the present thesis, a later experiment performed by the research group will be cited. This experiment was performed using the new FD algorithm with the CMA/MMA mode explained in the thesis, so it is a consequence of this work.

4.3.1 32-bit/s/Hz spectral efficiency WDM transmission over 177 km few-mode fiber

The experiment in [31] consisted in the transmission of 20 GBaud/s 16-QAM signal combining SDM and WDM. For the SDM, 6 spatial modes were used (LP_{01}, LP_{02} and the twofold degenerated LP_{11} and LP_{21}). For the WDM, 32 channels were transmitted with a spacing of 25 GHz. Taking into account the PDM and the 4 bits/symbol of the 16-QAM signal, the aggregate line-rate was

$$\begin{aligned} R_b &= 32 \text{ WDM} \times 6 \text{ SDM} \times 2 \text{ PDM} \times 4 \text{ bits/symbol} \times 20 \text{ GBaud/s} \\ &= 30.72 \text{ Tb/s} \end{aligned} \tag{4.22}$$

With the half of the spacing between WDM channels used in previous experiments, the total bandwidth was 25 GHz $\times$ 32 WDM = 800 GHz. This means that the spectral efficiency is 38.4 b/s/Hz. Applying the usually accepted 20% overhead FEC, this is 32 b/s/Hz as stated in the publication title.

The mode couplers used in this experiment were based in the 3D-waveguide for both the transmitter and the receiver. A span of 59 km of Graded Index FMF supporting 6 spatial modes was used. It was build by splicing fiber spools with different modal DGD in order to minimize the overall DGD in the span (see Table 4.1 and Figure 4.7). The value of the maximum DGD of the span was, in the end, of 350 ps.

Spool Nr.	Length [km]	DGD_{01-11} [ns]	DGD_{01-21} [ns]	DGD_{01-02} [ns]
1	3.0	-1.60	-3.19	-3.26
2	30.0	3.17	8.56	8.56
3	19.3	-0.70	-3.14	-3.31
4	6.5	-0.67	-1.88	-1.82
Total	58.8	0.20	**0.35**	0.16

TABLE 4.1: Mode dispersion parameters of the different spools used in the 6-mode FMF span. The DGD between LP_{11}, LP_{21} and LP_{02} can be computed from the ones in this table. The maximum uncompensated DGD was between LP_{01} and LP_{21}.

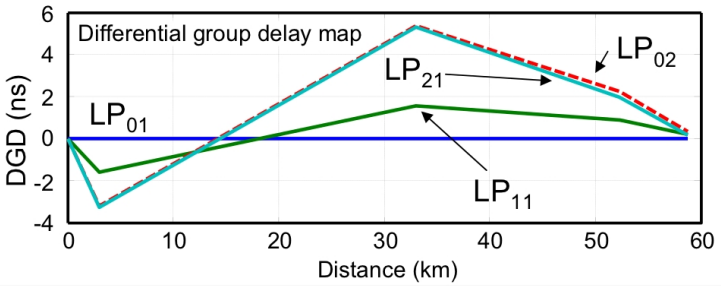


FIGURE 4.7: DGD map for the 59 km compensated FMF span. Figure from [31].

With this values, the maximal excursion in DGD (time difference between the slowest and the fastest possible path) was of 10.87 ns. However, the intensity impulse response of the span shown in Figure 4.8 (calculated as an average of the impulse responses of the 12×12 individual paths) has a dispersion plateau of only 9 ns. Because of this, the use of an equalizer window of 10 ns (200 symbols for the 20 GBaud/s signal transmitted) was enough to get good transmission performance.

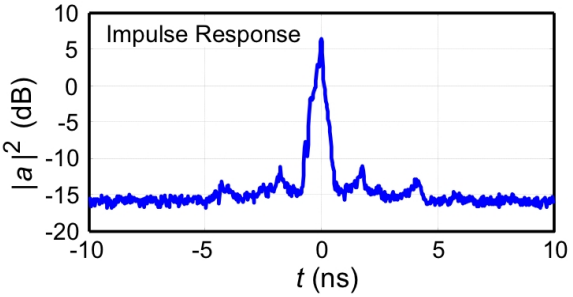
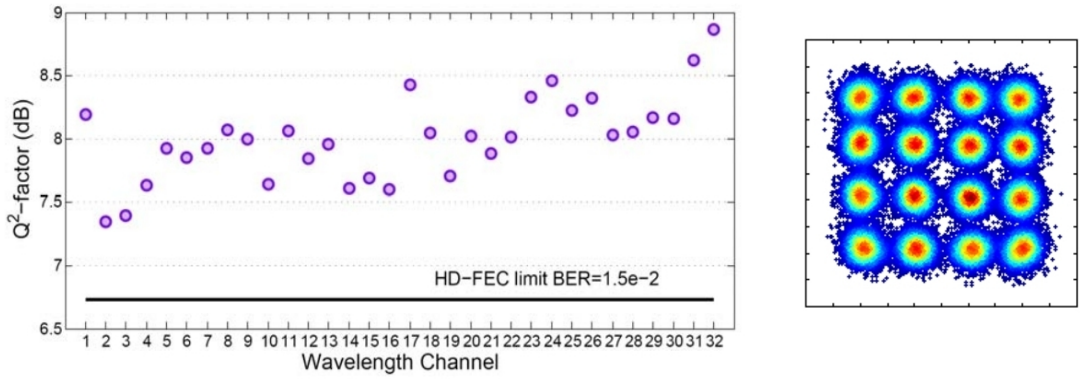


FIGURE 4.8: Intensity impulse response of the 59 km FMF span. The small peaks at the sides of the 1 ns wide central peak are due to the fiber splices. Figure from [31].

The maximum distance achieved without exceeding the limit BER of 10^{-2} allowed by the state-of-the-art FEC was 177 km, corresponding to three loops of the span. The

spectral efficiency of 32 b/s/Hz demonstrated is way over the $\sim$ 20 b/s/Hz predicted by the nonlinear Shannon capacity limit of a standard SMF for such a distance [3]. Moreover, it represents a record in a SDM transmission. Figure 4.9 shows the Q-factor of every WDM channel (from averaged BERs over the 12 PDM and SDM modes) and a typical constellation after the transmission for the 177 km experiment.



(a) (b)

FIGURE 4.9: (a) Q-factors for the 177 km transmission for all WDM channels obtained from averaging the BER over all 12 PDM and SDM modes. (b) Typical constellation after transmission. Figures from [31].

Chapter 5

Conclusions

The goal of this thesis was to show the potentiality of the Spatial Division Multiplexing using Few-mode Fibers as a possible solution to the capacity crunch of the single-mode fiber systems due to nonlinearities. First of all, it has been explained the overall paradigm of the SDM technique and two record experiments have been presented to show the results already achieved. Next, a new approach on the Multiple-Input Multiple-Output Digital Signal Processing used to undo the crosstalk between modes has been proposed showing that SDM has still a long way to go and its a promising technology.

In the first part of the thesis, the MIMO DSP has been analyzed to show how it is a key component of the SDM environment. It has been described as based in causal FIR filters adapted with the Least Mean-Square algorithm. Moreover, it has been shown how it is capable of compensating the modal coupling and dispersion occurred in the FMF. During the first part of the work for the thesis, some minor changes were introduced to the code used until that moment (for instance, the generalization from 3 spatial modes to an arbitrary number). Moreover, the code was used to process the sampled data of the two experiments given as examples and the results were analyzed.

In the first one, a 3-mode Graded Index FMF was used to transmit 20 GBaud/s QPSK signals through 3 SDM×2 PDM×34 WDM channels. An aggregate line-rate of 8.16 Tb/s and a spectral efficiency of 4 b/s/Hz were demonstrated over a maximum distance of 700 km, which at the moment represented a record distance for a combined SDM/WDM transmission.

In the second experiment presented, a GI-FMF with 6 spatial modes was used for the first time. Again the transmitted signals were 20 GBaud/s QPSK sequences. In this case, the number of WDM channels was only 8 and an aggregate line-rate (including PDM, SDM and WDM) of 3.84 Tb/s and a spectral efficiency of 8 b/s/Hz were achieved. The maximum transmission distance was 130 km. In addition, in this experiment, two new technologies to couple the modes into the FMF (photonic lantern and 3D-waveguide)

were demonstrated improving the insertion loss of the previous devices (phase plates) and showing that SDM is still a very open field of research.

Finally, during the experiments, some limitations of the MIMO DSP were detected. To solve them, the new frequency approach was proposed and implemented. A dramatic improvement in the computational complexity has been achieved with this new algorithm. Furthermore, the CMA/MMA method have been mixed with the LMS algorithm to solve the problems related with the phase error.

To show the performance of the new code, an experiment done by the same research group after the conclusion of this thesis work has been presented. This experiment has been performed with the algorithm developed and implemented in the thesis. A record spectral efficiency of 32 b/s/Hz over 177 km of 6-mode GI-FMF and using 32 WDM channels has been achieved. This figure is really important because it exceeds for the first time the theoretical capacity limit for the SMF case.

In conclusion, the SDM techniques along with the MIMO DSP have been presented as a very young but already highly promising way to keep tracking the traffic necessities. Of course, some problems are found in the path, but the results of the experiments keep going further and giving credit to the field. This thesis addressed and solved with success one of this problems, the computational scalability of the MIMO DSP algorithms.

List of publications

- M. Salsi, R. Ryf, G. Le Cocq, L. Bigot, D. Peyrot, G. Charlet, S. Bigo, N. K. Fontaine, M. A. Mestre, S. Randel, X. Palou, C. Bolle, B. Guan, and Y. Quiquempois, "A six-mode erbium-doped fiber amplifier", *European Conference and Exposition on Optical Communications (ECOC'12), Postdeadline papers, Th.3.A.6,* 2012.

- [11]: R. Ryf, N. K. Fontaine, M. A. Mestre, S. Randel, X. Palou, C. Bolle, A. H. Gnauck, S. Chandrasekhar, X. Liu, B. Guan, R.-J. Essiambre, P. J. Winzer, S. G. Leon-Saval, J. Bland-Hawthorn, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, L. Grüner-Nielsen, R. V. Jensen, and R. Lingle Jr., "12x12 MIMO transmission over 130-km few-mode fiber erbium-doped fiber amplifier", *Frontiers in Optics (FiO12), FW6C.4,* 2012.

- [10]: R. Ryf, M. A. Mestre, S. Randel, X. Palou, A. H. Gnauck, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, X. Jiang, and R. Lingle Jr., "Combined SDM and WDM transmission over 700-km few-mode fiber", *Proc. Optical Fiber Communications Conference (OFC/NFOEC13), OW1I.2,* 2013.

- N. K. Fontaine, R. Ryf, M. A. Mestre, B. Guan, X. Palou, S. Randel, Y. Sun, R. V. Jensen, and R. Lingle Jr., "Characterization of space-division multiplexing systems using a swept-wavelength interferometer", *Proc. Optical Fiber Communications Conference (OFC/NFOEC'13), OW1K.2,* 2013.

Bibliography

[1] P. J. Winzer, "Modulation and multiplexing in optical communication systems", *IEEE-LEOS Newsletter*, February 2009. `http://photonicssociety.org/newsletters/feb09/modulation.pdf`.

[2] R. W. Tkach, "Scaling optical communications for the next decade and beyond", *Bell Labs Technical Journal*, vol. 14, no. 4, pp. 3–9, 2010.

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

[4] C. E. Shannon, "A mathematical theory of communication", *The Bell Systems Technical Journal*, vol. 27, pp. 379–423, 623–656, July-October 1948.

[5] P. J. Winzer and G. J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems", *Optics Express*, vol. 19, no. 17, August 2011.

[6] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas", *Bell Labs Technical Joural*, vol. 1, no. 2, pp. 41–59, October 1996.

[7] J. Sajaguchi, Y. Awaji, N. Wada, A. Kanno, T. Kawanishi, T. Hayashi, T. Taru, T. Kobayashi, and M. Watanabe, "109-Tb/s (7x97x172-Gb/s SDM/WDM/PDM) QPSK transmission through 16.8-km homogeneous multi-core fiber", *Proc. Optical Fiber Communications Conference (OFC/NFOEC'11), PDPB6*, 2011.

[8] B. Zhu, T. F. Taunay, M. Fishteyn, X. Liu, S. Chandrasekhar, M. F. Yan, J. M. Fini, E. M. Monberg, F. V. Dimarcello, K. Abedin, P. W. Wisk, D. W. Peckham, and P. Dziedzic, "Space-, wavelength-, polarization-division multiplexed transmission of 56-Tb/s over a 76.8-km seven-core fiber", *Proc. Optical Fiber Communications Conference (OFC/NFOEC'11), PDPB7*, 2011.

[9] A. Li, A. Al Amin, X. Chen, and W. Shieh, "Reception of mode and polarization multiplexed 107-Gb/s CO-OFDM signal over a two-mode fiber", *Proc. Optical Fiber Communications Conference (OFC/NFOEC'11), PDPB8*, 2011.

[10] R. Ryf, M. A. Mestre, S. Randel, X. Palou, A. H. Gnauck, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, X. Jiang, and R. Lingle Jr., "Combined SDM and WDM transmission over 700-km few-mode fiber", *Proc. Optical Fiber Communications Conference (OFC/NFOEC'13), OW1I.2*, 2013.

[11] R. Ryf, N. K. Fontaine, M. A. Mestre, S. Randel, X. Palou, C. Bolle, A. H. Gnauck, S. Chandrasekhar, X. Liu, B. Guan, R.-J. Essiambre, P. J. Winzer, S. G. Leon-Saval, J. Bland-Hawthorn, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, L. Grüner-Nielsen, R. V. Jensen, and R. Lingle Jr., "12x12 MIMO transmission over 130-km few-mode fiber", *Frontiers in Optics (FiO'12), FW6C.4*, 2012.

[12] R. Ryf, R.-J. Essiambre, A. H. Gnauck, S. Randel, M. A. Mestre, C. Schmidt, P. J. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, T. Hayashi, T. Taru, and T. Sasaki, "Space-division multiplexed transmission over 4200-km 3-core microstructured fiber", *Proc. Optical Fiber Communications Conference (OFC/NFOEC'12), PDP5C.2*, 2012.

[13] D. Qian, E. Ip, M.-F. Huang, M.-J. Li, A. Dogariu, S. Zhang, Y. Shao, Y.-K. Huang, Y. Zhang, X. Cheng, Y. Tian, P. Nan Ji, A. Collier, Y. Geng, J. Liares, C. Montero, V. Moreno, X. Prieto, and T. Wang, "1.05Pb/s transmission with 109b/s/Hz spectral efficiency using hybrid single- and few-mode cores", *Frontiers in Optics (FiO'12), FW6C.3*, 2012.

[14] D. Gloge, "Weakly guiding fibers", *Applied Optics*, vol. 10, no. 10, pp. 2252–2258, October 1971.

[15] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, ch. 8. John Wiley & Sons, Inc., 1991.

[16] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3", *Proc. of the IEEE*, vol. 93, no. 2, pp. 216–231, 2005.

[17] S. Randel, R. Ryf, A. Sierra, P. J. Winzer, A. H. Gnauck, C. Bolle, R.-J. Essiambre, D. W. Peckham, A. McCurdy, and R. Lingle Jr., "6x56-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 6x6 MIMO equalization", *Optics Express*, vol. 19, no. 17, August 2011.

[18] R. Ryf, C. Bolle, and J. von Hoyningen-Huene, "Optical coupling components for spatial multiplexing in multi-mode fibers", *European Conference and Exposition on Optical Communications (ECOC'11), OSA Technical Digest, Th.12.B.1*, 2011.

[19] R. Ryf, M. A. Mestre, A. H. Gnauk, S. Randel, C. Schmidt, R.-J. Essiambre, P. J. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, X. Jiang, D. W. Peckham, A. McCurdy, and R. Lingle Jr., "Low-loss mode coupler for mode-multiplexed transmission in few-mode fiber", *Proc. Optical Fiber Communications Conference (OFC/NFOEC'12), PDP5B.5*, 2012.

[20] R. Ryf, N. K. Fontaine, and R.-J. Essiambre, "Spot-based mode coupler for mode-multiplexed transmission in few-mode fiber", *Photonics Society Summer Topical Meeting Series, IEEE*, pp. 199–200, July 2012.

[21] N. K. Fontaine, R. Ryf, S. G. Leon-Saval, and J. Bland-Hawthorn, "Evaluation of photonic lanterns for lossless mode-multiplexing", *European Conference and Exposition on Optical Communications (ECOC'12), OSA Technical Digest, Th.2.D.6*, 2012.

[22] S. Randel, R. Ryf, A. H. Gnauk, M. A. Mestre, C. Schmidt, R.-J. Essiambre, P. J. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, X. Jiang, and R. Lingle Jr., "Mode-multiplexed 620-GBd QPSK transmission over 1200-km DGD-compensated few-mode fiber", *Proc. Optical Fiber Communications Conference (OFC/NFOEC'12), PDP5C.5*, 2012.

[23] N. Benvenuto and G. Cherubini, *Algorithms for communication systems and their applications*. John Wiley and Sons, Inc., 2002.

[24] R. Ryf, S. Randel, A. H. Gnauck, C. Bolle, A. Sierra, S. Mumtaz, M. Esmaeelpour, E. C. Burrows, R.-J. Essiambre, P. J. Winzer, D. W. Peckham, A. McCurdy, and R. Lingle Jr., "Mode-division multiplexing over 96 km of few-mode fiber using coherent 6x6 MIMO processing", *Journal of Lightwave Technology*, vol. 30, no. 4, pp. 521–531, February 2012.

[25] E. R. Ferrara, "Fast implementation of LMS adaptive filters", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 474–475, August 1980.

[26] O. Zia-Chahabi, R. Le Bidan, M. Morvan, and C. Laot, "Efficient frequency-domain implementation of block-LMS/CMA fractionally spaced equalization for coherent optical communications", *IEEE Photonics Technology Letters*, vol. 23, no. 22, pp. 1697–1699, November 2011.

[27] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems", *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867–1875, November 1980.

[28] J. Yang, J.-J. Werner, and G. A. Dumont, "The multimodulus blind equalization and its generalized algorithms", *IEEE Journal on selected areas in Communications*, vol. 20, no. 5, pp. 997–1015, June 2002.

[29] K. Banović, M. A. S. Khalid, and E. Abdel-Raheem, "A configurable fractionally-spaced blind adaptive equalizer for QAM demodulators", *Digital Signal Processing*, vol. 17, no. 6, pp. 1071–1088, 2007.

[30] J. J. Shynk, "Frequency-domain and multirate adaptive filtering", *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14–37, January 1992.

[31] R. Ryf, S. Randel, N. K. Fontaine, M. Montoliu, E. Burrows, S. Chandrasekhar, A. H. Gnauck, C. Xie, R.-J. Essiambre, P. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, L. Grüner-Nielsen, R. V. Jensen, and R. Lingle Jr., "32-bit/s/hz spectral efficiency wdm transmission over 177-km few-mode fiber", *Proc. Optical Fiber Communications Conference (OFC/NFOEC'13), PDP5A.1*, 2013.