CHALMERS

Pilot-Aided Joint Phase and Polarization Tracking Algorithm

POOYAN SADEGHI

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2017

Master's thesis EX016/2017

Pilot-Aided Joint Phase and Polarization Tracking Algorithm

POOYAN SADEGHI

CHALMERS

Department of Electrical Engineering
Communication and Antenna Systems Group
Chalmers University of Technology
Gothenburg, Sweden 2017

Pilot-Aided Joint Phase and Polarization Tracking Algorithm

POOYAN SADEGHI

Examiner: Erik Agrell, Department of Electrical Engineering
Supervisor: Cristian B. Czegledi

Master's Thesis EX016/2017
Department of Electrical Engineering
Communication and Antenna Systems Group
Chalmers University of Technology
412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2017

Abstract

The availability of fast and reliable advanced communication systems made it possible to transfer massive amounts of data over thousands of kilometers. Long-haul fiber-optical links have a unique role in this achievement by using coherent transmission. Coherent systems encode the data onto both the phase and amplitude of the optical carrier, making it possible to use higher-order modulation formats and achieve improved spectral efficiency at the cost of a reduced tolerance against channel impairments. Digital signal processing methods have been exploited to mitigate the channel impairments, which can be deterministic or stochastic.

In this work we propose a pilot-aided algorithm to compensate jointly for the drifts of the carrier phase noise and state of the polarization, accommodating for an arbitrary modulation format. The algorithm is based on a channel model that accurately describes fiber propagation and it uses a sequence of known symbols at the receiver, named pilots, to estimate the channel and recover the carrier phase and state of polarization. The proposed pilot-aided algorithm originates from a blind algorithm that uses differential encoding. Simulation results compare both algorithms and show that the proposed algorithm has a better convergence rate and tolerance to the considered channel impairments, at the cost of a slight overhead. Furthermore, the state of polarization and phase ambiguities, which the blind algorithm suffers from, have also been resolved in the pilot-aided algorithm.

Keywords: coherent transmission, phase noise, polarization drift, pilot-aided, joint tracking, joint processing.

Acronyms

AWGN	additive white Gaussian noise
CS	cycle slip
DOF	degrees of freedom
DSP	digital signal processing
IM-DD	intensity-modulation-direct-detection
OOK	on-off keying
PM	polarization-multiplexed
QAM	quadrature-amplitude modulation
QPSK	quadrature phase-shift keying
MSE	mean squared error
SER	symbol error rate
SOP	state of polarization
SNR	signal-to-noise ratio

Contents

Chapter 1

Introduction

Choosing these days between the variety of information we receive at every moment is much harder than accessing them. One might say that is due to the availability of fast and reliable advanced communication systems, which made it possible to access massive amount of data by just a click. *Optical fibers* have a unique role in this achievement by transferring data over thousands of kilometers between continents or several meters between data-centers in a fast, reliable, and power-efficient manner. This is possible because of inimitable specification of optical fibers that have both enormous bandwidth and low power loss compare to all other transmissions media and make it the unbeatable candidate for long-haul/high-data rate communications.

On the other side, there are ever increasing demands for data and video services and applications over *Internet*, i.e., real-time video, online gaming; and even further, new coming services like Internet of things demands for more Internet traffic globally [1]. However, optical fibers are the backbone of Internet and have been deployed almost everywhere, but to be able to satisfy this growing demand, it is a necessity to aim for higher-performance optical communication systems from all aspects.

1.1 Background and History

We may address back the idea of transferring data over the light to the very past when the Lord "appeared to him [*Moses*] in flames of fire". But in recent years the bright idea of Charles Kao and George Hockham in 1966 of using "A dielectric fiber with a refractive index higher than its surrounding region" to transfer energy at optical frequencies [2] can be seen as the emerge of era of *optical communications*. This followed by laboratory experimenting of $<20\,\mathrm{dB/km}$ optical fiber loss in 1970 [3].

The commercial deployment of optical fibers happened in the 1980s and 1990s, meanwhile the first transcontinental optical fiber link named TAT-8 was installed over Atlantic ocean to connect USA to Europe in 1988 [4]. TAT-8 was a two-fiber-pair system where each pair could transfer $280\,\mathrm{Mb/s}$ [4]. Almost at the same time [5,6], the invention of erbium-doped fiber amplifiers caused a revolutionary increment in data rate of optical systems by leveraging wavelength-division multiplexing to transmit multiple channels over a single fiber each with a different optical carrier.

Until recently, the commercial optical systems were mostly using simple binary

modulation format known as on-off keying (OOK). This is a rather spectral inefficient modulation, which could transfer up to 10 Gb/s bit rate per wavelength channel [7, section 7] , but does not require complex transmitter and receiver.

While the demands for higher data rates increase, the optical fiber bandwidth is considered a limited source, so there is a great interest towards the deployment of more complex and higher order modulation formats than OOK, which can improve spectral efficiency and increase data throughput. This purpose has been achieved by using digital signal processing (DSP) to implement state-of-the-art coherent transceivers, which can modulate data onto both the in-phase and the quadrature components of the optical carrier in the transmitter and use coherent detection at the receiver side [8].

The spectral efficiency can be improved even further by using two orthogonal polarizations of the light, known as polarization multiplexing. In these systems, an independent data stream can be sent on each polarization, which makes data throughput doubled compared to single-polarization transmission. Here, a data symbol has in-phase and quadrature components in each polarization, which can be considered as a point in a four-dimensional space rather than two points in a two-dimensional space, as it is in the case of single-polarization. This inherent four-fold dimensionality of the light field will permit polarization-multiplexed (PM) systems to construct optical channels with four degrees of freedom (DOF) [9].

By using coherent detection at the receiver side, PM-binary phase-shift keying (PM-BPSK), PM-quadrature phase-shift keying (PM-QPSK), and PM-16-quadrature-amplitude modulation (PM-16-QAM) are commercialized, allowing for 50 Gb/s, 100 Gb/s, and 200 Gb/s, respectively, per channel transmission at a symbol rate of 25 Gbaud, and it is expected to achieve higher data rates by deploying higher order modulation formats [7, section 6]. But using higher order modulation imposes less tolerance against channel impairments. DSP can also be used here to mitigate these impairments, such as nonlinearities, chromatic dispersion, additive noise, and laser phase noise. To be able to compensate for these impairments, which can be deterministic or stochastic, we need to have a channel model for optical fibers to describe light propagation through, as accurate as possible.

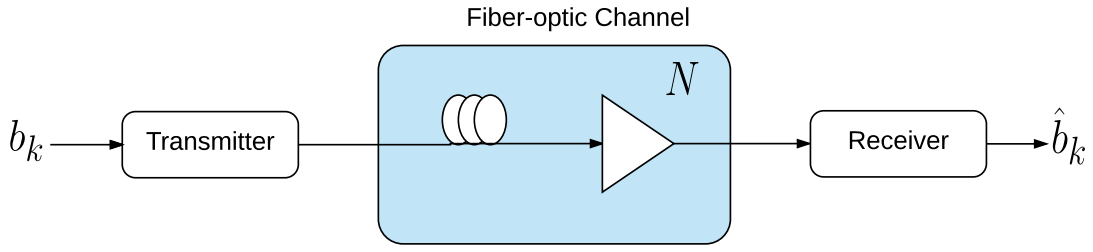

Figure 1.1: Basic schematic of a fiber-optic link

A basic optical fiber link is depicted in Figure 1.1. The transmitter converts the bits b_k to an optical signal to be transmitted over the optical channel. The link

consists of N spans, each including a fiber span and an optical amplifier to transfer the optical signal over the desired distance. The receiver decodes the received optical signal, where the output $\hat{b}_k$ will be an estimation of the transmitted bit.

1.2 Coherent Optical Communication Systems

Coherent systems are pioneer technologies that had a great influence on optical communication systems to reach to a better performance. They improved the spectral efficiency, increased receiver sensitivity, and enabled communications over longer distances compared to intensity-modulation-direct-detection (IM-DD) techniques [10, 11].

IM-DD systems modulate the data at the transmitter by changing the intensity of the laser light, where the information is contained. The detection is done at the receiver by a photo-detector, which generates an electrical current proportional to the power of the received optical signal [12,13]. IM-DD suffers from limited spectral efficiency and further, from nonlinearity in the receiver, which makes it hard to correct linear channel impairments. This nonlinearity comes from the fact that the generated electrical current by the photo-diode is proportional to the square of the received signal magnitude (square law principle) [14].

Coherent systems have some considerable advantages over IM-DD techniques. Besides better sensitivity, which is mainly due to using local oscillator and balanced photo-detectors in the receiver, they also have better spectral efficiency. Coherent detection made it possible to use amplitude/phase modulations in both polarizations of the light. This allows for multilevel modulation formats such as PM-QPSK and PM-16-QAM, which as a result gives several times better spectral efficiency compare to IM-DD [15]. Furthermore, contrary to IM-DD, coherent detection is a linear detection technique, where the generated electrical signal in the photo-detector is proportional to the electrical field vector of the optical signal. This linearity means that linear impairments and distortions such as chromatic dispersion or polarization mode dispersion, can be compensated effectively [16, 17].

Coherent detection technologies have been investigated in the academic area since the 1980s, mostly due to their better sensitivity than IM-DD for long distance transmission [18, 19]. However, they were shadowed for a period of 20 years and remained there mainly for two reasons. First, it was hard to implement such an phase-locked loop to lock the phase and frequency of the received signal, with available analog carrier recovery technologies. Second, by the invention of erbium-doped fiber amplifier and wavelength division multiplexing, which improved the sensitivity of the receivers considerably [5, 6], the main reasons for using coherent detection were solved and they became commercially less interesting. But as the genius prediction of John Barry and Edward Lee in 1990, "coherent techniques may someday revolutionize optical communication." [20], coherent detection came back to the table because of the demands for higher order modulation formats to improve the spectral efficiency. Furthermore, by technologies development in the field of electronics, coherent technologies could help to recover phase and frequency and also to compensate for channel impairments with the availability of powerful DSPs and high sampling rate analog-to-digital converters implemented in modern application-

specific integrated circuits [8, 21, 22].

The coherent systems are so advantageous that someone may think why it has not been implemented earlier, actually their advantages come with some expenses. The coherent transmitter and receiver are much more complex than their identical blocks in IM-DD techniques; they require powerful DSP and high sampling-rate analog-to-digital converters, which are components that can limit the transmission baud rate in the coherent transmitter [10, 14].

1.3 Contributions

In this report, a pilot-aided algorithm for joint tracking of the phase and polarization has been proposed. The algorithm uses a sequence of known symbols named *pilot packet* to estimate and compensate for phase noise and drift of the state of the polarization. For the first time, without using differential coding, both phase and polarization ambiguities are removed by using this algorithm, compared to the other known tracking algorithms, which suffer from at least one of the ambiguities.

1.4 Thesis Organization

The remainder of the report is organized as follows. Chapter 2 gives a brief overview of the fiber-optic channel and coherent optical fiber transmission. In Chapter 3, the system model is presented. Chapter 4, first, known algorithms for tracking phase and polarization are presented and a blind algorithm which jointly tracks phase and polarization is discussed. Thereafter, the proposed pilot-aided algorithm, as the extension of the blind algorithm, is proposed. Next, the calculations of the design parameters of the algorithm are presented and an implementable form of the algorithm is given. In Chapter 5, the performance of the proposed algorithm are evaluated and compared to the blind algorithm. Finally, Chapter 6, concludes this thesis.

Chapter 2

Coherent Optical Fiber Transmission

The total reflection phenomenon is the physical concept which optical fiber communication is perched on. An optical fiber consists of a cylindrical core and a covering cladding, both usually made of high-quality silica glass. The difference in the refractive index of the core and cladding (which has a smaller refractive index) will confine the light beam into the core.

There are two main types of optical fibers, multi-mode optical fiber and single-mode optical fiber, regarding the different optical modes the light can propagate through the fiber, which are solutions to Maxwell's equations [23, Ch.2]. A multi-mode optical fiber uses a larger core than single-mode optical fiber. This allows more modes to propagate and also the use of cheaper transmitter and receiver components. The core of single-mode fiber is smaller, however, it is proper for long-haul transmission and higher performance links.

In this thesis we are interested in long-haul transmission, where the optical signal propagation is highly affected by the channel impairments. So, in Section 2.1, we discuss briefly some of the channel impairments which will be considered in our channel model and continue with explaining the coherent transceiver structure in Section 2.2.

2.1 The Optical Fiber Channel

In this thesis we have considered three channel impairments, additive noise, phase noise, and polarization state drift, each explained briefly in the following without considering mathematical details.

The optical fiber has a much lower attenuation compared to other mediums like copper wires, but still when the optical signal propagates through the fiber, its power attenuates. Due the fact that the receiver requires a minimum signal power, the attenuation becomes a limiting factor for long distance transmission. To overcome this problem, as shown in Figure 1.1, the optical signal needs to be amplified periodically throughout the entire link. The optical amplifiers are the source of noise which degrade the signal-to-noise ratio (SNR). The accumulation of noise generated at each amplifier can be modeled as additive white Gaussian

noise (AWGN) [23, section 7.6].

Coherent systems modulate the optical signals in the phase and amplitude of the optical field in both polarizations of the light. Two independent lasers, one in the transmitter and one in the receiver are used in coherent systems to be able to access both phase and amplitude of the optical signal. The transmitter laser generates the optical carrier wave and the receiver laser is used as a reference to demodulate the received optical signal. These lasers are not perfect and synchronized, so they have an active phase difference, which results in phase noise. This phase noise causes random rotation of the received PM constellation, which induces disastrous errors in the detected symbols, if it is not compensated for. The spectral shape of the practical lasers is wider than a perfect delta function. This imperfection is the source of the phase noise [24]. We can quantify the phase noise by the linewidth sum $\Delta\nu$ of both transmitter and receiver lasers multiplied by the symbol period T. The dimensionless parameter $\Delta\nu T$ will be a measure for the phase noise sensitivity of the system.

Similar to radio waves, light is also an oscillating electromagnetic field. This means the propagation of the light can be characterized as a sinusoidal wave of the electrical field $\mathbf{E}$, where the orientation of this field is called polarization. The simplest type of polarization is linear polarization, where the field $\mathbf{E}$ oscillates in a single fixed plane and can be decomposed to two orthogonal components, E_x and E_y which are also sinusoidal wave forms. The polarization of the field $\mathbf{E}$ can change with time, causing the amplitude and the phase of both E_x and E_y to change. This phenomenon is called *drift* of the state of polarization (SOP). Coherent fiber-optic transmission modulates data independently in both E_x and E_y components, which are called X and Y polarizations. The changes in the physical environment, which the fiber goes through, will result in a random drift of the SOP. The SOP drift can be seen as a 4D rotation of the 4D PM constellation. To compensate for this stochastic impairment, it is desired to have a mathematical model for it, then design a special block in the receiver to recover the SOP.

2.2 Coherent Transmission

The schematic of a coherent transmitter is shown in Figure 2.1 [15]. A polarization beam splitter divides the laser light into two branches corresponding to X and Y polarizations. Thereafter each branch is split into two branches corresponding to the In-Phase and Quadrature components. Modulation in each branch is done by a Mach-Zehnder modulator, then a $\pi/2$ polarization rotation is applied to the signal to form the X-polarization and Y-polarization, respectively. The two X and Y polarized signals are combined in a polarization beam combiner and then injected to the fiber.

A block diagram of a coherent receiver is shown in Figure 2.2. The direction of the arrows show the signal flow from the optical signal to data symbols. The receiver consists of different blocks that compensate for the impairments and further estimation and detection block. The scope of this thesis includes only three blocks of the shown figure, SOP tracking, carrier phase recovery and symbol estimation/decoding blocks.

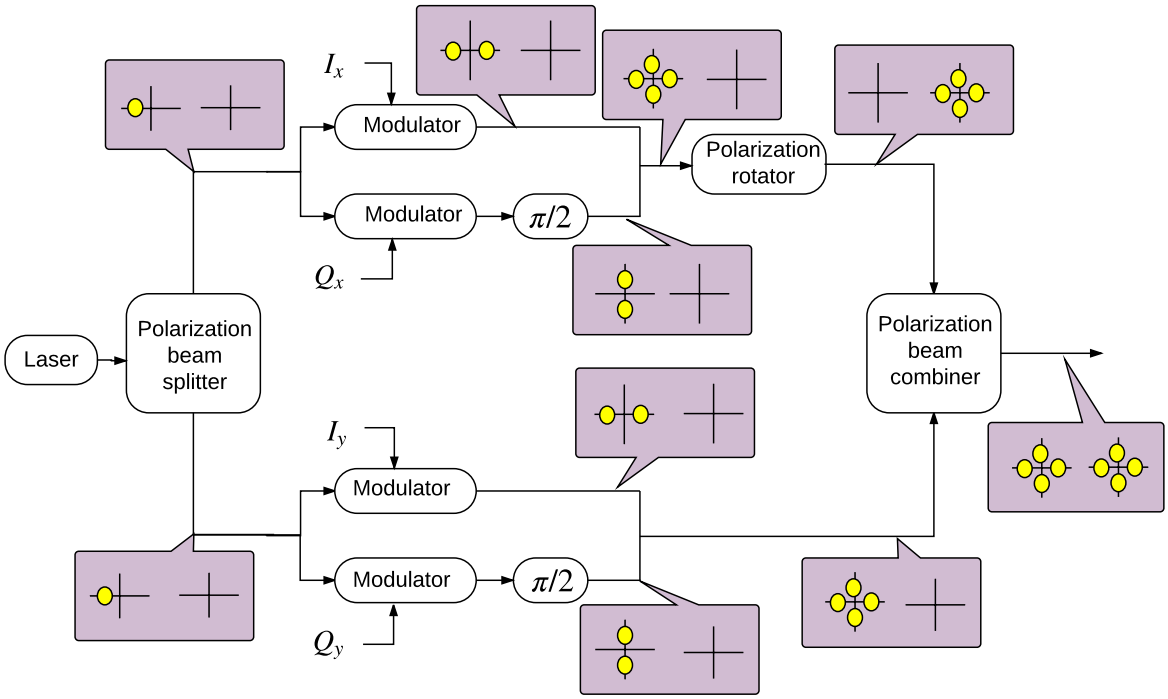

Figure 2.1: Schematic of an optical coherent transmitter [15].

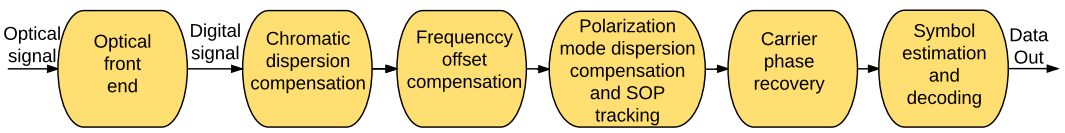

Figure 2.2: Block diagram of an optical coherent receiver

Chapter 3

System Model

This chapter, which is a duplication of the system model described in [25], presents the channel model used in this thesis to describe analytically the propagation of the electrical field conveying the optical signal in the presence of additive noise, phase nose, and drift of the SOP.

The optical signal and its propagation can be represented by three different formalisms: Jones formalism, Stokes formalism, and 4D formalism. In Section 3.1, the Jones formalism has been exploited to mathematically describe the optical signals. Section 3.2 describes jointly both the phase noise and the SOP drift based on a channel model proposed in [26].

The following notation conventions will be used through the thesis : column vectors are denoted by bold lower case (e.g., $\mathbf{u}$) and matrices by bold upper case (e.g., $\mathbf{U}$), except the Pauli matrices $\boldsymbol{\sigma}_i$ and the electrical field Jones vector $\mathbf{E}$, to be consistent with the literature. The conjugate transpose is written $\mathbf{u}^{\mathrm{H}}$, the 2×2 identity matrix as $\mathbf{I}_2$ and the expectation operator as $\mathbb{E}[\cdot]$. The dot operation is the inner product of two vectors with the same size, except in the cases $\boldsymbol{\alpha} \cdot \vec{\boldsymbol{\sigma}}$ and $\hat{\mathbf{H}}_k^{-1} \cdot \vec{\boldsymbol{\sigma}}$, which are defined $(\alpha_1, \alpha_2, \alpha_3) \cdot (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3) = \alpha_1\boldsymbol{\sigma}_1 + \alpha_2\boldsymbol{\sigma}_2 + \alpha_3\boldsymbol{\sigma}_3$ and $\hat{\mathbf{H}}_k^{-1} \cdot (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3) = \hat{\mathbf{H}}_k^{-1}\boldsymbol{\sigma}_1 + \hat{\mathbf{H}}_k^{-1}\boldsymbol{\sigma}_2 + \hat{\mathbf{H}}_k^{-1}\boldsymbol{\sigma}_3$, respectively, where $\vec{\boldsymbol{\sigma}} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3)$ is the tensor of the three Pauli matrices. The absolute value is denoted by $|\cdot|$ and the Euclidean norm by $\|\cdot\|$.

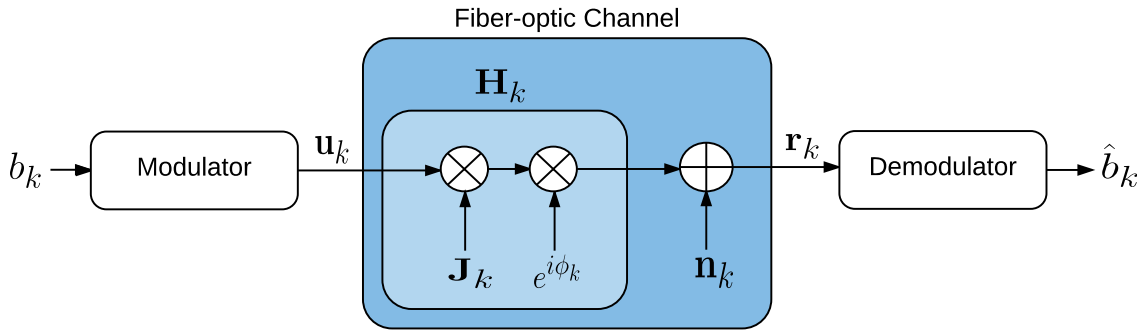

Figure 3.1: A channel model which considers phase noise, SOP drift and AWGN.

3.1 Representation and Propagation of Optical Signals

The coherent optical signal consists of two polarizations, each include two quadrature parts, and can be described by a *Jones vector*, as a function of propagation distance z and time t

$$\mathbf{E}(z,t) = \begin{pmatrix} E_x(z,t) \\ E_y(z,t) \end{pmatrix}. \tag{3.1}$$

This is a two-dimensional complex-valued vector, where E_x and E_y are baseband signals denoting the the X and Y field components. E_x and E_y are complex signals and consist of In-phase and Quadrature parts. $\mathbf{E}(0,t)$ is the signal sent into the transmission medium generated by linearly modulating the information symbols $\mathbf{u}_k \in \mathbb{C}^2$

$$\mathbf{E}(0,t) = \sum_k \mathbf{u}_k p(t-kT), \tag{3.2}$$

with a real-valued electrical pulse $p(t)$ for the symbol index $k \in \mathbb{Z}$, at symbol intervals T. The transmitted symbols $\mathbf{u}_k$ are discrete 2×1 complex vectors drawn uniformly and independently from a constellation $\mathscr{C} = \{\mathbf{c}_1, \mathbf{c}_2 \dots \mathbf{c}_M\}$ with M finite points. The average energy of the constellation is the average energy of the symbols $\|\mathbf{u}_k\|^2$ and equals

$$E_s = \frac{1}{M} \sum_{k=1}^{M} \|\mathbf{c}_k\|^2. \tag{3.3}$$

The discrete received symbol at distance L

$$\mathbf{r}_k = \int_{-\infty}^{\infty} \mathbf{E}(L,t) p(t-kT) dt, \tag{3.4}$$

is achieved by matched filtering and sampling with period T of the electrical field $\mathbf{E}(L,t)$.

In this thesis we only consider phase noise, SOP drift, and AWGN as the impairments. Therefore, the relation between the propagated symbol $\mathbf{r}_k$ and the transmitted symbol $\mathbf{u}_k$ can be described as

$$\mathbf{r}_k = \mathbf{H}_k \mathbf{u}_k + \mathbf{n}_k, \tag{3.5}$$

where the phase noise and the SOP drift is modeled as complex 2×2 matrix $\mathbf{H}_k \in \mathbb{C}^{2\times 2}$. The additive noise $\mathbf{n}_k \in \mathbb{C}^2$ is a complex zero mean Gaussian random variable with independent components in both X and Y polarizations. The total discrete-time noise power $\|\mathbf{n}_k\|^2$ is finite and equal to N_0, i.e., $\mathbb{E}[\mathbf{n}_k \mathbf{n}_k^{\mathrm{H}}] = N_0 \mathbf{I}_2$, resulting in a noise variance of $N_0/2$ per polarization [9]. This channel model is shown in Figure 3.1, and can replicate a 2×2 multiple-input multiple-output system used widely in wireless communications [27], even though here the inputs and outputs are signals from two independent polarizations of light.

The so-called *Jones matrix* $\mathbf{H}_k$ is a unitary matrix that does not diminish the power of input signal during propagation and can be characterized by

$$\mathbf{H}_k^{\mathrm{H}}\mathbf{H}_k = \mathbf{H}_k\mathbf{H}_k^{\mathrm{H}} = \mathbf{I}_2, \tag{3.6}$$

$$\det \mathbf{H}_k = 1. \tag{3.7}$$

Any unitary matrix $\mathbf{H}_k$ can be expressed in the form

$$\mathbf{H}_k = \begin{pmatrix} a & b \\ -e^{i\theta}b^* & e^{i\theta}a^* \end{pmatrix}, \ |a|^2 + |b|^2 = 1, \ \det \mathbf{H}_k = e^{i\theta}, \tag{3.8}$$

where a, b, and θ are real numbers. The matrix $\mathbf{H}_k$ depends on four parameters (the phase of a, the phase of b, the relative amplitude between a and b, and the angle θ), which gives *four* DOFs. The condition in (3.6) constrains the DOFs to be *four*, otherwise in general, a complex 2×2 matrix has eight DOFs, i.e., the real and imaginary parts of the four elements.

Structuring $\mathbf{H}_k$ like (3.8), will help to define it as a *exponential matrix* function $H(\phi, \boldsymbol{\alpha})$ with four parameters, i.e., ϕ modeling phase noise, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ modeling the SOP drift,

$$H(\phi, \boldsymbol{\alpha}) = \exp(-i(\boldsymbol{\alpha} \cdot \vec{\boldsymbol{\sigma}} + \phi\mathbf{I}_2)), \tag{3.9}$$

where $\vec{\boldsymbol{\sigma}} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3)$ is a tensor of the Pauli spin matrices [28]

$$\boldsymbol{\sigma}_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \ \boldsymbol{\sigma}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \boldsymbol{\sigma}_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{3.10}$$

The vector $\boldsymbol{\alpha}$ can be presented as a unit vector multiplied by its amplitude $\boldsymbol{\alpha} = \|\boldsymbol{\alpha}\|\mathbf{a}$. For ease of notation we call $\theta = \|\boldsymbol{\alpha}\|$ and the elements of the unit vector $\mathbf{a} = (a_1, a_2, a_3)$ represent the vectors direction on a unit sphere.

The two phenomena, phase noise ϕ and SOP drift $\boldsymbol{\alpha}$ modeled by $H(\phi, \boldsymbol{\alpha})$ in (3.9) can be factorized in two separate exponential terms

$$H(\phi, \boldsymbol{\alpha}) = e^{-i\phi} \exp(-i\boldsymbol{\alpha} \cdot \vec{\boldsymbol{\sigma}}), \tag{3.11}$$

where we define

$$J(\boldsymbol{\alpha}) = \exp(-i\boldsymbol{\alpha} \cdot \vec{\boldsymbol{\sigma}}) \tag{3.12}$$

that strictly models the SOP drift. Then, (3.5) can be written like

$$\mathbf{r}_k = e^{-i\phi_k}\mathbf{J}_k\mathbf{u}_k + \mathbf{n}_k. \tag{3.13}$$

Similarly to $\mathbf{H}_k$, $\mathbf{J}_k$ is also a unitary matrix, so it can be presented by as a matrix function $J(\boldsymbol{\alpha})$. However, it has an extra constraint $\det \mathbf{J}_k = 1$, this constraint reduces the number of DOFs to *three*, i.e., the elements of $\boldsymbol{\alpha}$.

The function $H(\phi, \boldsymbol{\alpha})$ and $J(\boldsymbol{\alpha})$ can be written as

$$H(\phi, \boldsymbol{\alpha}) = (\cos\phi - i\sin\phi)(\mathbf{I}_2\cos\theta - i\boldsymbol{a}\cdot\vec{\boldsymbol{\sigma}}\sin\theta), \tag{3.14}$$

$$J(\boldsymbol{\alpha}) = \mathbf{I}_2\cos\theta - i\boldsymbol{a}\cdot\vec{\boldsymbol{\sigma}}\sin\theta, \tag{3.15}$$

after expanding (3.9) and (3.12) into the Tailor series

$$\exp\mathbf{A} = \sum_{k=0}^{\infty}\frac{1}{k!}\mathbf{A}^k, \quad \forall\,\mathbf{A}\in\mathbb{C}^{n\times n} \tag{3.16}$$

defined in [29, p. 165] and using $(\boldsymbol{\alpha}\cdot\vec{\boldsymbol{\sigma}})^2 = \theta^2\mathbf{I}_2$.

Both $\mathbf{H}_k$ and $\mathbf{J}_k$ being unitary matrices, their inverses can be calculated by conjugate transposing which in this case, is the same as negating ϕ and $\boldsymbol{\alpha}$, so

$$H(\phi, \boldsymbol{\alpha})^{-1} = H(\phi, \boldsymbol{\alpha})^{\mathrm{H}} = H(-\phi, -\boldsymbol{\alpha}), \tag{3.17}$$

$$J(\boldsymbol{\alpha})^{-1} = J(\boldsymbol{\alpha})^{\mathrm{H}} = J(-\boldsymbol{\alpha}). \tag{3.18}$$

3.2 Phase and Polarization Drift Channel Model

We model the phase noise ϕ_k as a Wiener process [30, 31]

$$\phi_k = \dot{\phi}_k + \phi_{k-1}, \tag{3.19}$$

where $\dot{\phi}_k$ is the *innovation* of ϕ_k at each time instance k. The phase innovation $\dot{\phi}_k$ is a random variable drawn independently, from a real zero-mean Gaussian distribution

$$\dot{\phi}_k \sim \mathcal{N}(0, \sigma_\nu^2), \tag{3.20}$$

with the variance $\sigma_\nu^2 = 2\pi\Delta\nu T$, where $\Delta\nu$ is sum of the linewidths of transmitter and receiver lasers.

The accumulated phase noise ϕ_k at time k is the summation of all $\dot{\phi}_1, \ldots, \dot{\phi}_k$ added to the initial phase ϕ_0. The initial phase ϕ_0, which states the initial phase difference between the transmitter and receiver lasers, can be considered a uniformly distributed random variable in the interval $[0, 2\pi)$. Therefore, ϕ_k is the summation of k Gaussian random variable terms that becomes a Gaussian distributed random variable with mean ϕ_0 and variance $k\sigma_\nu^2$.

The SOP of the received symbol will be different from the transmitted one due to the possible imperfections along the fiber cable. The SOP is changing over the time and is a randomly dynamic process. Similar to the phase noise ϕ_k, the time evolution of the SOP drift can be described by a dynamic model. One proposed model which consider $\mathbf{J}_k$ as the sequence of random Jones matrices is [26]

$$\mathbf{J}_{k+1} = J(\dot{\boldsymbol{\alpha}}_k)\mathbf{J}_k, \tag{3.21}$$

where $J(\dot{\boldsymbol{\alpha}}_k)$ defined in (3.12) is the matrix that expresses the *innovation* of the SOP (cf. $\dot{\phi}_k$ in (3.19)).

The random parameter in $J(\dot{\boldsymbol{\alpha}}_k)$ is $\dot{\boldsymbol{\alpha}}$, which has three elements. These three elements are drawn independently, at each time instance k, from a real zero-mean Gaussian distribution

$$\dot{\boldsymbol{\alpha}}_k \sim \mathcal{N}(0, \sigma_\rho^2 \, \mathbf{I}_3), \tag{3.22}$$

with variance $\sigma_\rho^2 = 2\pi\Delta\rho T$, where $\Delta\rho$ is defined as the *polarization linewidth* [26], which states the speed of the SOP drift, similar to the concept of the laser linewidth describing the phase noise, cf. (3.20).

The initial state of the $\mathbf{J}_0 = J(\boldsymbol{\alpha}_0)$ is defined in a way to ensure that $\mathbf{J}_0\mathbf{u}$ is distributed uniformly over all possible SOP for a chosen $\mathbf{u}$ [26]. This means, the vector $\boldsymbol{\alpha}_0 = \theta\mathbf{a}$ should be formed from a 4D unit vector such $(\cos\theta, a_1 \sin\theta, a_2 \sin\theta, a_3 \sin\theta)^{\mathrm{T}} = \mathbf{g}/\|\mathbf{g}\|$, where $\mathbf{g} \sim \mathcal{N}(0, \sigma_\rho^2 \, \mathbf{I}_4)$.

The matrix $H(\phi, \boldsymbol{\alpha})$ in (3.14) combines the effects of both phase noise and SOP drift in one equation. Analogously to (3.21), the update of $\mathbf{H}_k$ can be expressed as

$$\mathbf{H}_{k+1} = H(\dot{\phi}_k, \dot{\boldsymbol{\alpha}}_k)\mathbf{H}_k, \tag{3.23}$$

where the phase innovation $\dot{\phi}_k$ and randomly evolving vector $\dot{\boldsymbol{\alpha}}_k$ are defined in (3.20) and (3.22).

Chapter 4

Polarization and Phase Tracking Algorithms

In this chapter different algorithms for the recovery of the phase and SOP introduced in Chapter 3 are discussed. Four available algorithms used to compensate drifts and estimate the phase noise and SOP drift are the blind phase search algorithm, the constant modulus algorithm, the multiple modulus algorithm, and the Kabsch algorithm [17, 30, 32, 33] . The *blind phase search algorithm* estimates and compensates for the phase noise separately in each polarization [30]. This algorithm only considers the phase noise and the additive noise. The *constant modulus algorithm* [17] and the *multiple modulus algorithm* [32] compensate for the drift of the SOP based on the same principle. The constant modulus algorithm was designed for constellations with constant modulus, like phase-shift keying, while a multiple modulus algorithm is adopted for multiple-modulus constellation. The *Kabsch algorithm* [33] jointly tracks the phase and SOP, can be used for arbitrary constellation. This algorithm and the blind phase search algorithm are blind to the phase ambiguity, therefore differential coding should be used to remove this ambiguity.

In Section 4.1, an overview of two algorithms that jointly recover both phase and SOP are presented. Section 4.2 introduces in detail the *Blind Joint Polarization and Phase Tracking Algorithm*. This algorithm is presented in [25]. Thereafter, in Section 4.3, the *Pilot-Aided Joint Polarization and Phase Tracking Algorithm* is proposed and discussed in details. This algorithm, which is the main focus of this thesis, is an extension of the Blind Algorithm [25], using known symbols to recover the phase and SOP.

4.1 Joint Polarization and Phase Tracking Algorithms

In this section, based on the model proposed in [26], which jointly models the stochastic behaviour of the phase noise and SOP drift, two model-based algorithms for jointly recovering these channel impairments are presented in Section 4.2 and Section 4.3, respectively. The first algorithm discussed in section 4.2, proposed in [25]. This algorithm compensate jointly for both the phase noise and SOP drift

using a non-data aided decision-directed architecture, so as referred in the name is a "Blind" algorithm.

The proposed "pilot-aided" algorithm, which is the main focus of this thesis, is discussed in Section 4.3. The pilot-aided algorithm uses *pilot packets* to estimate the channel matrix. However, tracking the channel along the pilot intervals is done exactly at the same way in the *blind algorithm*.

The blind algorithm uses *differential coding* [34, Sec. 2.6.1] independently in each polarization. Differential coding, which encodes information in the phase difference of two consecutive symbols, removes the phase ambiguity, hence prevents disastrous error caused by cycle slip (CS) or ambiguities. CS is a discontinuity in measuring phase of the received signal and happens when receiver losses the carrier phase lock. During a CS, all data symbols afterwards will be erroneous and could not be decoded correctly. The pilot-aided algorithm does not use differential encoding, so if a CS occurs during the data packet, it will not be detectable and track of the channel will be lost until the next pilot packet is received. Applying differential decoding to the blind algorithm comes at a cost of SNR penalty compared to the pilot-aided algorithm, which has a lower spectral efficiency due to the overhead imposed by the pilot symbols. This trade-off between SNR and spectral efficiency has been investigated in Chapter 5.

The blind algorithm suffers from polarization ambiguities, which means the transmitted symbol in one polarization, may be received in another polarization and detected in error on the receiver side. The polarization ambiguity is removed in the pilot-aided algorithm, by exploiting pilot symbols in each polarization. In this report, we have intentionally neglected the effect of *polarization ambiguity* on the performance of the blind algorithm.

4.2 Blind Joint Polarization and Phase Tracking Algorithm

It should be noted that this section is essentially identical to [25]. In this algorithm, it is supposed $\mathbf{H}_k$ does not change considerably over a symbol duration, so the transmitted symbol can be estimated from $\hat{\mathbf{H}}_k^{-1}\mathbf{r}_k$ based on the previous estimated of the channel matrix $\hat{\mathbf{H}}_k$ using the minimum Euclidean distance criterion [35, section 2.5]

$$\hat{\mathbf{u}}_k = \underset{\mathbf{c}\in\mathscr{C}}{\arg\min} \left\| \hat{\mathbf{H}}_k^{-1}\mathbf{r}_k - \mathbf{c} \right\|^2. \tag{4.1}$$

Thereafter $\hat{\mathbf{H}}_k$ is updated as

$$\hat{\mathbf{H}}_{k+1} = H(\hat{\phi}_k, \hat{\boldsymbol{\alpha}}_k)\hat{\mathbf{H}}_k\,, \tag{4.2}$$

where $\hat{\phi}_k$ and $\hat{\boldsymbol{\alpha}}_k$ are the estimates of $\dot{\phi}_k$ and $\dot{\boldsymbol{\alpha}}_k$. Here, an error function

$$e_k = \left\| \left(H(\phi, \boldsymbol{\alpha})\hat{\mathbf{H}}_k\right)^{-1}\mathbf{r}_k - \hat{\mathbf{u}}_k \right\|^2, \tag{4.3}$$

is defined, which is the Euclidean distance between the decoded symbol $\hat{\mathbf{u}}_k$ and the updated estimation $\hat{\mathbf{H}}_{k+1}\mathbf{r}_k$. This error function e_k, is minimized with respect to ϕ and $\boldsymbol{\alpha}$, i.e.,

$$[\hat{\phi}_k, \hat{\boldsymbol{\alpha}}_k] = \arg\min_{\phi,\,\boldsymbol{\alpha}} e_k\,, \tag{4.4}$$

by applying the gradient descent [36, section 9.3] method only for one iteration to achieve $\hat{\phi}_k$ and $\hat{\boldsymbol{\alpha}}_k$, so that

$$\hat{\phi}_k = -\mu_\phi \mathcal{R}e\left(\frac{\partial\, e_k}{\partial \phi}\bigg|_{\substack{\phi=0 \\ \boldsymbol{\alpha}=[0,0,0]^{\mathrm{T}}}} \right) \tag{4.5}$$

$$= -2\mu_\phi \mathcal{R}e\big(i(\hat{\mathbf{H}}_k^{-1}\mathbf{r}_k - \hat{\mathbf{u}}_k)^{\mathrm{H}}\hat{\mathbf{H}}_k^{-1}\mathbf{r}_k\big), \tag{4.6}$$

$$\hat{\boldsymbol{\alpha}}_k = -\mu_{\boldsymbol{\alpha}} \mathcal{R}e\left(\nabla_{\boldsymbol{\alpha}} e_k\bigg|_{\substack{\phi=0 \\ \boldsymbol{\alpha}=[0,0,0]^{\mathrm{T}}}} \right) \tag{4.7}$$

$$= -2\mu_{\boldsymbol{\alpha}} \mathcal{R}e\big(i(\hat{\mathbf{H}}_k^{-1}\mathbf{r}_k - \hat{\mathbf{u}}_k)^{\mathrm{H}}\hat{\mathbf{H}}_k^{-1} \cdot \vec{\boldsymbol{\sigma}}\mathbf{r}_k\big). \tag{4.8}$$

μ_ϕ and $\mu_{\boldsymbol{\alpha}}$ are positive tracking step size of the phase and SOP parameters, respectively, which determine the convergence speed of the algorithm and also the accuracy of the channel tracking. These parameters have been calculated heuristically [25] as

$$\mu_\phi = \frac{\sqrt{\Delta\nu T c}}{E_s}\,, \tag{4.9}$$

$$\mu_{\boldsymbol{\alpha}} = \frac{\sqrt{\Delta\rho T c}}{E_s}\,, \tag{4.10}$$

where c is a constant dependent on the modulation format. The value of c for PM-QPSK, PM-16-QAM, PM-64-QAM, and PM-256-QAM, has been given in [25, Table 1]. The tracking step sizes μ_ϕ and $\mu_{\boldsymbol{\alpha}}$ are chosen to be different values because their drift time are very different, typically ($\sim 1\,\mu$s) for the phase noise and ($\sim 1\,$ms) for the SOP.

It should be noted that the partial derivatives in (4.5) and (4.7) are evaluated at $\phi = 0$, $\boldsymbol{\alpha} = [0, 0, 0]^{\mathrm{T}}$ due to zero mean value of the innovation parameters $\dot{\phi}_k$ and $\dot{\boldsymbol{\alpha}}_k$, resulting in no preferred direction of $\hat{\phi}_k$ and $\hat{\boldsymbol{\alpha}}_k$.

4.3 Pilot-Aided Joint Polarization and Phase Tracking Algorithm

The pilot-aided algorithm uses a known sequence of symbols at the receiver named *pilots*, to estimate the channel matrix. Pilot symbols are inserted at the beginning of each data packet. The stream of transmitted symbols is shown in Figure 4.1. The transmitted pilot packet $\mathbf{U}_{\mathrm{P}i}$ consists of L known symbols, and the transmitted data packet $\mathbf{U}_{\mathrm{D}i}$ consists of K symbols, i.e.,

$$\mathbf{U}_{\mathrm{P}i} = \{\mathbf{u}_{\mathrm{P}i,1}, \mathbf{u}_{\mathrm{P}i,2} \ldots \mathbf{u}_{\mathrm{P}i,l} \ldots \mathbf{u}_{\mathrm{P}i,L}\}, \tag{4.11}$$

$$\mathbf{U}_{\mathrm{D}i} = \{\mathbf{u}_{\mathrm{D}i,1}, \mathbf{u}_{\mathrm{D}i,2} \ldots \mathbf{u}_{\mathrm{D}i,k} \ldots \mathbf{u}_{\mathrm{D}i,K}\}, \tag{4.12}$$

where i is the index of i^{th} packet in the symbol stream. Respectively, in the receiver, the received pilot packet $\mathbf{R}_{\text{P}i}$ and the received data packet $\mathbf{R}_{\text{D}i}$ are

$$\mathbf{R}_{\text{P}i} = \{\mathbf{r}_{\text{P}i,1}, \mathbf{r}_{\text{P}i,2} \dots \mathbf{r}_{\text{P}i,l} \dots \mathbf{r}_{\text{P}i,L}\} \tag{4.13}$$

$$\mathbf{R}_{\text{D}i} = \{\mathbf{r}_{\text{D}i,1}, \mathbf{r}_{\text{D}i,2} \dots \mathbf{r}_{\text{D}i,k} \dots \mathbf{r}_{\text{D}i,K}\}. \tag{4.14}$$

Here, we define the parameter *pilot rate* $R = L/K$ to be used later as the indication of the spectral efficiency of the algorithm.

At the receiver, the decoding block will check how many pilot symbols in pilot packet $\mathbf{R}_{\text{P}i}$ are erroneous, if at least 20% of pilot symbols are in error, we assume that a CS occurred, and the channel needs to be re-estimated. Otherwise, it will keep tracking of the channel in the same way explained in Section 4.2, no matter it is along the pilot or the data packet. It should be noted that at the beginning of transmission, the channel is completely unknown to the receiver and the first pilot sequence $\mathbf{R}_{\text{P}1}$ is used to estimate the channel matrix.

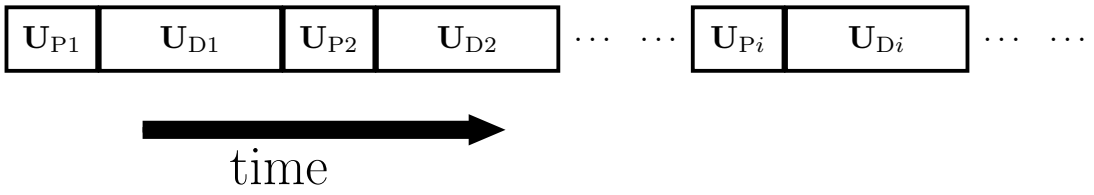

Figure 4.1: The stream of transmitted symbols.

4.3.1 Channel Estimation Using Pilot Packet $\text{U}_{\text{P}i}$

To estimate the channel over the received pilot packet $\mathbf{R}_{\text{P}i}$, we need to define an *average error function*, which is minimized by applying the gradient descent method for $n = 0, 1, \dots$ iterations [36, section 9.3]. At the n^{th} iteration, the average error function is

$$\overline{e}_{\mathbf{P}i}^{(n)} = \frac{1}{L} \sum_{l=1}^{L} \left\| \left(H(\phi, \boldsymbol{\alpha}) \hat{\mathbf{H}}_{\mathbf{P}i}^{(n)} \right)^{-1} \mathbf{r}_{\mathbf{P}i,l} - \mathbf{u}_{\mathbf{P}i,l} \right\|^2, \tag{4.15}$$

for $i = 1, 2, \dots$ and $l = 1, 2, \dots, L$,where $\mathbf{r}_{\mathbf{P}i,l}$ is the received pilot symbol corresponding to $\mathbf{u}_{\mathbf{P}i,l}$. Here, we do not want to estimate the real channel matrix $\mathbf{H}_{\mathbf{P}i,l}$ corresponding to each pilot symbol $\mathbf{u}_{\mathbf{P}i,l}$, but instead, define $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ as an overall estimation of the channel matrix over the entire packet. The estimate $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ is calculated, such that the average error function $\overline{e}_{\mathbf{P}i}^{(n)}$ is minimized with respect to ϕ and $\boldsymbol{\alpha}$, i.e.,

$$[\hat{\phi}_{\mathbf{P}i}^{(n)}, \hat{\boldsymbol{\alpha}}_{\mathbf{P}i}^{(n)}] = \arg\min_{\phi, \boldsymbol{\alpha}} \overline{e}_{\mathbf{P}i}^{(n)}. \tag{4.16}$$

Therefore, after applying the gradient descent method, $\overline{e}_{\mathbf{P}i}^{(n)}$ is minimized by

$$\hat{\phi}_{\mathbf{P}i}^{(n)} = -\mu_{\mathrm{P}} \left(\left. \frac{\partial\, \overline{e}_{\mathbf{P}i}^{(n)}}{\partial \phi} \right|_{\substack{\phi=0 \\ \boldsymbol{\alpha}=[0,0,0]^{\mathrm{T}}}} \right) \tag{4.17}$$

$$= -2\, \frac{\mu_{\mathrm{p}}}{L} \sum_{l=1}^{L} \mathcal{R}\!e\Big(j \big((\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)})^{-1} \mathbf{r}_{\mathbf{P}i,l} - \mathbf{u}_{\mathbf{P}i,l} \big)^{\mathrm{H}} (\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)})^{-1} \mathbf{r}_{\mathbf{P}i,l} \Big), \tag{4.18}$$

$$\hat{\boldsymbol{\alpha}}_{\mathbf{P}i}^{(n)} = -\mu_{\mathrm{P}} \left(\left. \nabla_{\boldsymbol{\alpha}}\, \overline{e}_{\mathbf{P}i}^{(n)} \right|_{\substack{\phi=0 \\ \boldsymbol{\alpha}=[0,0,0]^{\mathrm{T}}}} \right) \tag{4.19}$$

$$= -2\, \frac{\mu_{\mathrm{p}}}{L} \sum_{l=1}^{L} \mathcal{R}\!e\Big(i \big((\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)})^{-1} \mathbf{r}_{\mathbf{P}i,l} - \mathbf{u}_{\mathbf{P}i,l} \big)^{\mathrm{H}} (\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)})^{-1} \vec{\boldsymbol{\sigma}} \mathbf{r}_{\mathbf{P}i,l} \Big). \tag{4.20}$$

The estimate of the channel matrix is updated as

$$\hat{\mathbf{H}}_{\mathbf{P}i}^{(n+1)} = H(\hat{\phi}_{\mathbf{P}i}^{(n)}, \hat{\boldsymbol{\alpha}}_{\mathbf{P}i}^{(n)}) \hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}. \tag{4.21}$$

The parameters $\hat{\phi}_{\mathbf{P}i}^{(n)}$ and $\hat{\boldsymbol{\alpha}}_{\mathbf{P}i}^{(n)}$ in (4.18) and (4.20) have been evaluated at $\phi = 0$, $\boldsymbol{\alpha} = [0,0,0]^{\mathrm{T}}$ for each iteration of the gradient descent method. The derivations of (4.18) and (4.20) are given in the Appendix. In the next section, it is explained how $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ will be used for tracking the channel during the data packet $\mathbf{U}_{\mathrm{D}i}$.

The parameter μ_{P}, similar to μ_{ϕ} and $\mu_{\boldsymbol{\alpha}}$, is the positive tracking step size, which determines the convergence speed and the accuracy of the algorithm to reach to the minimum value of $\overline{e}_{\mathbf{P}i}^{(n)}$. In contrast with Section 4.2, for both noise and SOP parameters in (4.18) and (4.20), the step size is μ_{P}, to maintain the simplicity of the algorithm by including less parameters. The value of μ_{P}, which gives the fastest convergence rate, is calculated by simulations in Chapter 5.

The behaviour of the average error function $\overline{e}_{\mathbf{P}i}^{(n)}$ in (4.15) is unknown. Therefore, to calculate the number of iterations n, we need to define a stopping criterion when applying the gradient descent method. The stopping criterion can be defined based one the difference in the value of $\overline{e}_{\mathbf{P}i}^{(n)}$ after each iteration of the gradient descent method

$$\left\{ \left(\overline{e}_{\mathbf{P}i}^{(n-m)} - \overline{e}_{\mathbf{P}i}^{(n-m-1)} \right) < t \,\middle|\, m \in \big\{0, 1, \ldots M\big\} \right\}. \tag{4.22}$$

Parameters M and t are two fixed parameters in the algorithm to ensure the decrement of $\overline{e}_{\mathbf{P}i}^{(n)}$ after one iteration, has not been accidental and is small enough to stop further iteration. This allows us to use a flexible number of iteration n in the algorithm instead of a fixed number, which will be tested in Chapter 5.

It should be noted that in the pilot-aided algorithm, we did not consider the evolution of the matrix channel $H(\phi, \boldsymbol{\alpha})$ during the pilot packet. However, we will investigate the influence of the fast evolving channel $H(\phi, \boldsymbol{\alpha})$ during the pilot packet on the algorithm performance in Chapter 5.

4.3.2 Channel Tracking

In the case that less than 20% of the decoded pilot symbols in the pilot packet $\mathbf{U}_{\mathrm{P}i}$ are in error, we suppose the channel does not require to be re-estimated. Therefore, the channel matrix during the received pilot packet $\mathbf{U}_{\mathrm{P}i}$ or the received data packet $\mathbf{U}_{\mathrm{D}i}$ will be updated exactly as described in Section 4.2. In this case, we use $\hat{\mathbf{H}}_k$ as the estimation of both matrices, for ease of notation.

The proposed algorithm is explained in an implementable way in Section 4.3.6.

4.3.3 Steady State Behaviour

If the pilot-aided algorithm proposed in Section 4.3.1, always re-estimates the channel matrix at every received pilot packet $\mathbf{R}_{\mathrm{P}i}$, it can affect the performance of the algorithm in the steady state regime. It means, if the channel is being tracked properly along the data packet $\mathbf{R}_{\mathrm{D}(i-1)}$, using the pilot packet $\mathbf{R}_{\mathrm{P}i}$ to re-estimate the channel for the next data packet $\mathbf{R}_{\mathrm{D}i}$, will disturb the performance. The disturbance can be seen as a jump in the symbol error rate (SER), at the beginning of the received data packet $\mathbf{R}_{\mathrm{D}i}$. To remove this effect, as mentioned previously, the number of erroneous received pilot symbols $\mathbf{r}_{\mathbf{P}i,l}$ should be checked, e.g., if 20% of the received pilot symbols are in error, it can be decided that the track of channel is lost. Therefore, the channel re-estimation needs to be done based on (4.15) to (4.21), otherwise the channel tracking can be continued like (4.3) to (4.8).

The disturbing effect and the proposed treatment are investigated in Chapter 5.

4.3.4 Length of The Pilot Packet

It is ideal to have the shortest possible pilot packet, which in this case is *two* symbols. As explained in Section 3.1, $H(\phi, \boldsymbol{\alpha})$ has *four* DOFs, therefore to resolve it uniquely, at least it is needed to form a linear matrix equation like

$$\left[\mathbf{r}_{\mathbf{P}i,1}, \mathbf{r}_{\mathbf{P}i,2}\right] = H(\phi, \boldsymbol{\alpha})\left[\mathbf{u}_{\mathbf{P}i,1}, \mathbf{u}_{\mathbf{P}i,2}\right], \quad \mathbf{r}_{\mathbf{P}i,l}, \mathbf{u}_{\mathbf{P}i,l} \in \mathbb{C}^2, \tag{4.23}$$

to give four independent equations, where obviously

$$\mathbf{u}_{\mathbf{P}i,1} \neq \gamma \mathbf{u}_{\mathbf{P}i,2}, \quad \gamma \in \mathbb{C}, \tag{4.24}$$

and suppose $H(\phi, \boldsymbol{\alpha})$ does not change over a symbol duration.

The matrix $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ is the estimation of $H(\phi, \boldsymbol{\alpha})$, and can be achieved by minimizing $\overline{e}_{\mathbf{P}i}^{(n)}$ as explained in (4.15) to (4.21), even with *two* pilot symbols (4.23). However, this estimation would be highly affected by the additive noise, when the pilot packet length L is rather small. Let's consider propagation equation (3.5) along the pilot packet in the form of

$$\mathbf{R}_{\mathrm{P}i} = H(\phi, \boldsymbol{\alpha})\mathbf{U}_{\mathrm{P}i} + \mathbf{N}_{\mathrm{P}i}, \tag{4.25}$$

where

$$\begin{aligned} \mathbf{R}_{\mathrm{P}i} &= \left[\mathbf{r}_{\mathbf{P}i,1}, \mathbf{r}_{\mathbf{P}i,2}, \ldots, \mathbf{r}_{\mathbf{P}i,L}\right], \\ \mathbf{U}_{\mathrm{P}i} &= \left[\mathbf{u}_{\mathbf{P}i,1}, \mathbf{u}_{\mathbf{P}i,2}, \ldots, \mathbf{u}_{\mathbf{P}i,L}\right], \\ \mathbf{N}_{\mathrm{P}i} &= \left[\mathbf{n}_{\mathbf{P}i,1}, \mathbf{n}_{\mathbf{P}i,2}, \ldots, \mathbf{n}_{\mathbf{P}i,L}\right], \end{aligned} \tag{4.26}$$

and suppose $H(\phi, \boldsymbol{\alpha})$ is constant during these L symbols. Then, the minimum-variance unbiased estimator for linear model [37, Ch. 4] will be

$$\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)} = (\mathbf{U}_{\mathbf{P}i}\,\mathbf{U}_{\mathbf{P}i}^{\mathrm{H}})^{-1}\,\mathbf{U}_{\mathbf{P}i}\,\mathbf{R}_{\mathbf{P}i}^{\mathrm{H}}, \tag{4.27}$$

assuming $\mathbf{U}\,\mathbf{U}^{\mathrm{H}}$ is invertible, hence the covariance matrix will be

$$\mathbf{C}_{\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}} = \frac{N_0}{L}\mathbf{I}_2. \tag{4.28}$$

The smaller covariance matrix means the more accurate estimation, so as seen in (4.28) by increasing the length of the pilot packet L, the estimation becomes more accurate. However, it will be more affected by the phase noise and the SOP drift due to the evolution of the channel matrix $H(\phi, \boldsymbol{\alpha})$ over the interval of L pilot symbols (LT). The proper pilot length L will be investigated in Chapter 5.

It should be mentioned that the minimized *average error function* $\overline{e}_{\mathbf{P}i}^{(n)}$ (4.15), is not a proper indication of the accuracy of the estimated channel matrix $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ in (4.21). In fact, in the algorithm explained in (4.15) to (4.21), $\hat{\phi}$ and $\hat{\boldsymbol{\alpha}}$ are found in such a way that minimize $\overline{e}_{\mathbf{P}i}^{(n)}$, but $\overline{e}_{\mathbf{P}i}^{(n)}$ is blind against the ambiguity that happens due to not resolving $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ uniquely. Therefore, minimizing $\overline{e}_{\mathbf{P}i}^{(n)}$, does not mean that the mean squared error (MSE) between $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ and $H(\phi, \boldsymbol{\alpha})$

$$\begin{aligned} e_{\mathrm{MSE}} &= \mathbb{E}\left\{ \left\| H(\phi, \boldsymbol{\alpha}) - \hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}) \right\|_{\mathbb{F}}^2 \right\} \\ &= \mathrm{tr}\{\mathbf{H}_{\mathrm{dif}}\mathbf{H}_{\mathrm{dif}}^{\mathrm{H}}\}, \end{aligned} \tag{4.29}$$

is also minimized, where $\mathbf{H}_{\mathrm{dif}} = H(\phi, \boldsymbol{\alpha}) - \hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$, and $\|\cdot\|_{\mathbb{F}}$ is the Frobenius matrix norm. In Chapter 5, we will check the MSE of the channel estimation $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$, because it offers a better perspective about the proper length of the pilot packet L.

4.3.5 Choosing The Pilot Symbols

The symbols in the constellation diagram with the highest energy level, are the proper candidates to be chosen as the pilot symbol $\mathbf{u}_{\mathbf{P}i,l}$. These symbols are located at the corners of the constellations diagrams in X and Y polarizations, which have the lowest probability to be received incorrectly among all other symbols in the constellation diagrams. The pilot packet $\mathbf{U}_{\mathrm{P}i}$ with length L, is constructed from a combination of these symbols.

To find the proper combination of the pilot symbols, we shall consider a minimum mean squared error estimator, which estimates a channel optimally when the pilot symbols are orthogonal [38]. Even though in the channel estimation method in Section 4.3.1 we do not use minimum mean squared estimator, but still exerting the summation in (4.15) implies that we try to find $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ by averaging over the pilot packet. Therefore, it can have a positive influence on the accuracy of the estimation, if we can choose pilot symbols to be orthogonal. The fact that the constellation diagrams in X and Y polarizations provide only 4 independent bases

means, it can not exist more than 4 orthogonal pilot symbols in the pilot packet. Therefore, we construct the pilot packet $\mathbf{U}_{\mathrm{P}i}$ in (4.26) with length $L > 4$, such as

$$\mathbf{U}_{\mathrm{P}i} = \left[\mathbf{u}_{\mathbf{P}i,1}, \mathbf{u}_{\mathbf{P}i,2}, \ldots, \mathbf{u}_{\mathbf{P}i,L}\right], \tag{4.30}$$

where

$$\mathbf{u}_{\mathbf{P}i,l+1} = \mathbf{u}_{\mathbf{P}i,(\mathrm{l} \bmod 4)+1}, \quad l = 0, \ldots, L-1 \tag{4.31}$$

and

$$\mathbf{u}_{\mathbf{P}i,l} \,.\, \mathbf{u}^{\mathrm{H}}_{\mathbf{P}i,q} = 0, \quad \begin{cases} l, q \in \{1,2,3,4\} \\ l \neq q \end{cases} \tag{4.32}$$

e.g., for $L = 7$

$$\mathbf{U}_{\mathbf{P}i} = \left[\mathbf{u}_{\mathbf{P}i,1}, \mathbf{u}_{\mathbf{P}i,2}, \mathbf{u}_{\mathbf{P}i,3}, \mathbf{u}_{\mathbf{P}i,4}, \mathbf{u}_{\mathbf{P}i,5}, \mathbf{u}_{\mathbf{P}i,6}, \mathbf{u}_{\mathbf{P}i,7}\right]. \quad \begin{cases} \mathbf{u}_{\mathbf{P}i,1} = \mathbf{u}_{\mathbf{P}i,5} \\ \mathbf{u}_{\mathbf{P}i,2} = \mathbf{u}_{\mathbf{P}i,6} \\ \mathbf{u}_{\mathbf{P}i,3} = \mathbf{u}_{\mathbf{P}i,7} \end{cases}$$

In this part, similar to the average energy of the constellation E_s in (3.3), we need to define the average energy of the pilot packet $\mathbf{U}_{\mathbf{P}i}$, i.e.,

$$E_{\mathbf{P}} = \frac{1}{L} \sum_{l=1}^{L} \|\mathbf{u}_{\mathbf{P}i,l}\|^2, \tag{4.33}$$

which is used in Chapter 5, where we calculate the step size μ_{P}.

4.3.6 Algorithmic Summary

Algorithm 1: Proposed Pilot-Aided algorithm

$\hat{\mathbf{H}}_1^{-1} := \mathbf{I}_2$ // initialize the channel
$i := 1$ // the i^{th} packet, see Fig.(4.1)
wait for $\mathbf{U}_{\mathrm{P}i}$ // the i^{th} pilot packet
for $k = 1$ **to** $L + K$ **do** // symbol index in the i^{th} packet
  wait for $\mathbf{r}_{i,k} = \begin{cases} \mathbf{r}_{\mathrm{P}i,k} & \text{if } k \leq L \\ \mathbf{r}_{\mathrm{D}i,k-L} & \text{if } k > L \end{cases}$
  Output: $\hat{\mathbf{u}}_{i,k} := \arg\min_{\mathbf{c} \in \mathscr{C}} \left\| \hat{\mathbf{H}}_k^{-1} \mathbf{r}_{i,k} - \mathbf{c} \right\|^2$ // detect the symbol
  if $k = L$ ***and*** $\sum_{l=1}^{L} \mathrm{I}\big(\hat{\mathbf{u}}_{i,l} \neq \mathbf{u}_{\mathrm{P}i,l}\big) \geq \lfloor 0.2L \rfloor$ **then**
    // check end of pilot packet and check channel re-estimation
    $\overline{e}_{\mathbf{P}i}^{(0)} := 0$
    $(\mathbf{H}_{\mathbf{P}i}^{(0)})^{-1} := \mathbf{I}_2$
    $a := 0$
    $\mathrm{n} := 0$
    while $a < M$ **do**
      $\mathrm{n} := \mathrm{n} + 1$
      $\hat{\phi}_{\mathbf{P}i}^{(n)}$ // from (4.18)
      $\hat{\boldsymbol{\alpha}}_{\mathbf{P}i}^{(n)}$ // from (4.20)
      $(\mathbf{H}_{\mathbf{P}i}^{(n)})^{-1} := (\mathbf{H}_{\mathbf{P}i}^{(n-1)})^{-1} H(-\hat{\phi}_{\mathbf{P}i}^{(n)}, -\hat{\boldsymbol{\alpha}}_{\mathbf{P}i}^{(n)})$ // channel re-estimation
      $a := a + \operatorname{sgn}\Big(t - \big(\overline{e}_{\mathbf{P}i}^{(n)} - \overline{e}_{\mathbf{P}i}^{(n-1)}\big)\Big)$ // from (4.22)
    end
    $\hat{\mathbf{H}}_k^{-1} := (\mathbf{H}_{\mathbf{P}i}^{(n)})^{-1}$ // re-estimated channel
  else
    $\hat{\phi}_k$ // from (4.6)
    $\hat{\boldsymbol{\alpha}}_k$ // from (4.8)
    $\hat{\mathbf{H}}_{k+1}^{-1} := \hat{\mathbf{H}}_k^{-1} H(-\hat{\phi}_k, -\hat{\boldsymbol{\alpha}}_k)$ // updated channel
  end
end
$\hat{\mathbf{H}}_1^{-1} := \hat{\mathbf{H}}_{L+K+1}^{-1}$ // initial channel estimation for the $(i+1)^{\text{th}}$ packet
$i := i + 1$
goto 3

The proposed pilot-aided algorithm in Section 4.3, is presented in an easily implementable way in this section, without considering the detailed calculations.

The algorithm is summarized in Algorithm 1, by using more relaxed notations than Section 4.3. In fact, for the implementation of the algorithm with the simulation tools, e.g., MATLAB, we do not need to use the precise notation, while for the mathematical rendering of the algorithm, exploiting the very accurate notation is a necessity to avoid any ambiguity. However, in this section whenever the notations are different with the formulas in Section 4.3, it is denoted and clarified. The algorithm uses several constant parameters during its execution, for ease of

Table 4.1: List of constant parameters in Algorithm 1

Parameter	L	K	M	t	μ_ϕ	$\mu_{\boldsymbol{\alpha}}$	μ_{P}
Equation	(4.11)	(4.12)	(4.22)	(4.22)	(4.5)	(4.7)	(4.17)

implementation these parameters have been listed in Table 4.1.

The decoded symbol $\hat{\mathbf{u}}_{i,k}$ is the regularly output of the algorithm. The algorithm is initialized for the first time with setting the channel matrix to be $\hat{\mathbf{H}}_1^{-1} = \mathbf{I}_2$, then the loop in the algorithm starts. The loop, represented by counter i, which is a representation of the transmitted symbol stream shown in Figure 4.1. The following steps in the algorithm are taken for every received packet i, which consists of the received pilot packet $\mathbf{R}_{\mathbf{P}i}$ and the received data packet $\mathbf{R}_{\mathbf{D}i}$ in (4.13) and (4.14), respectively. The receiver does not discriminate between a received pilot symbol $\mathbf{r}_{\mathbf{P}i,l}$ or a received data symbol $\mathbf{r}_{\mathbf{D}i,k}$, when deciding the estimated symbol $\hat{\mathbf{u}}_{i,k}$. Therefore, we can consider the symbol index k as a number between $1 < k < L + K$ to be used as the common index for both pilot and data symbols. This, also allows us to use $\mathbf{r}_{i,k}$ instead of $\mathbf{r}_{\mathrm{P}i,k}$ (for $k \leq L$) and $\mathbf{r}_{\mathrm{D}i,k-L}$ (for $k > L$). However, the receiver is aware of the pilot symbols $\mathrm{u}_{\mathrm{P}i,l}$ inside the pilot packet $\mathbf{U}_{\mathrm{P}i}$, and at the end of each received pilot packet, where $k = L$, it will check the number of erroneous received pilot symbols. If the number of erroneous received pilot symbols is more than a threshold, decided to be 20% of the length of pilot packet L, the receiver will detect that the track of channel is lost, and needs to be re-estimated based on the formulas (4.15) to (4.22). In the Algorithm 1, the indicator function $\mathrm{I}(x)$ has been exploited for this purpose.

Every time the channel requires to be re-estimated, the receiver needs to initialize some parameters, which are $\overline{e}_{\mathbf{P}i}^{(0)}$, $(\mathbf{H}_{\mathbf{P}i}^{(0)})^{-1}$, n, and a. The initial state for the average error function in (4.15) is $\overline{e}_{\mathbf{P}i}^{(0)} = 0$, and $(\mathbf{H}_{\mathbf{P}i}^{(0)})^{-1} = \mathbf{I}_2$ is the initial inverse estimation of the channel matrix, during the pilot packet $\mathbf{U}_{\mathbf{P}i}$ in (4.21). $n = 0$, where n is the number of iterations of the gradient descent method defined as a flexible number. This means, the gradient descent method will be applied to minimize the average error function $\overline{e}_{\mathbf{P}i}^{(n)}$ in (4.15) until the condition in (4.22) is satisfied. The condition is satisfied when for M continuous iterations of the gradient descent method, the decrement in the average error function $\overline{e}_{\mathbf{P}i}^{(n)}$ becomes less than the threshold t, where M and t are designing parameters. To implement this condition in the Algorithm 1, a parameter a and the sign function $\mathrm{sgn}(x)$ are used. The inverse estimation of the channel matrix $(\mathbf{H}_{\mathbf{P}i}^{(n)})^{-1}$ achieved by minimizing average error function $\overline{e}_{\mathbf{P}i}^{(n)}$, is used as the inverse channel estimation $\mathbf{H}_k^{-1} = (\mathbf{H}_{\mathbf{P}i}^{(n)})^{-1}$ for the first symbol in the data packet, where the symbol index is $k = L + 1$.

In the case that the number of erroneous received pilots are less than 20% of the length of the pilot packet L, there is no need to re-estimate the channel, and the channel tracking is done as stated in formulas (4.2) to (4.8).

It should be noted that the algorithmic summery of the blind algorithm has been presented in [25].

Chapter 5

Analysis and Results

In this chapter, in Section 5.1, the parameters of the pilot packet $\mathbf{U}_{\mathrm{P}i}$ exploited in the pilot-aided algorithm in Section 4.3, are investigated. Thereafter, in Section 5.2, we will investigate the performance of the algorithm and compare it with the blind tracking algorithm described in Section 4.2. The simulations have been done for PM-QPSK, PM-16-QAM, PM-64-QAM, and PM-256-QAM modulation formats at the symbol rate of 28 Gbaud.

5.1 Pilot Packet Parameters

In this section the parameters, which are used to estimate the overall channel matrix $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ during the pilot packet are investigated by simulations. These parameters are the step size μ_{P}, the number of iterations n, and the length of the pilot packet L.

5.1.1 Step Size and Number of Iterations

In Figure 5.1, we examined the value of $\overline{e}_{\mathbf{P}i}^{(n)}$ in (4.15) after $n = \{5, 20, 60\}$ iterations of the gradient descent method as explained in Section 4.3.1, for a range of $0.002 < \mu_{\mathrm{P}} < 2$, for two different levels of the phase noise and the SOP drift. Moreover, as explained in Section 4.3, instead of considering a fixed number of iterations, we also consider n to be flexible so that satisfies the criterion given in (4.22), for $t = 10^{-5}$ and $M = 5$. The estimated channel $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ acquired after n iterations of the gradient descent method, is considered as the initial channel estimation for the the data packet $\mathbf{R}_{\mathrm{Di}}$. In these simulations, the initial estimate of the channel during the pilot packet, is considered $\hat{\mathbf{H}}_{\mathbf{P}i}^{(0)} = \mathbf{I}_2$.

The settings for this experiment for PM-QPSK, PM-16-QAM, PM-64-QAM, and PM-256-QAM, respectively, are as follows: SNR, which is the quantification of AWGN, is set to $\{13.86, 21.19, 27, 43, 33.52\}$ dB, such that a SER $= 10^{-3}$ will be achievable by this SNR in the AWGN scenario. The accumulated laser linewidth $\Delta\nu$ is chosen such that $\Delta\nu \cdot T = \{3.6 \cdot 10^{-5}, 0.36 \cdot 10^{-5}, 0.18 \cdot 10^{-5}, 0.04 \cdot 10^{-5}\}$ (corresponding to $\Delta\nu = \{1000, 100, 50, 10\}$ kHz at 28 Gbaud) and the polarization linewidth $\Delta\rho$ is set such that $\Delta\rho \cdot T = \{3.57 \cdot 10^{-8}\}$ (corresponding to $\Delta\rho = 1$ kHz at 28 Gbaud). The length of the pilot packet is $L = 10$, and the results are obtained by averaging over $5 \cdot 10^3$ transmissions of the pilot packets.

The figure shows that by increasing n, $\overline{e}_{\mathbf{P}i}^{(n)}$ tends to its minimum value for a wider range of μ_{P}. However, it is obvious that the flexible n defined in (4.22), behaves as effective as the case, where n is fixed and equal to 60, but for a narrower range of μ_{P}. It must be mentioned, even if n is increased by several orders, however with smaller amount of μ_{P}, $\overline{e}_{\mathrm{P}i}^{(n)}$ is minimized, but still there will be an error floor, which by a very low tolerance is the same as seen in Figure 5.1. The level of this error floor is dependent to the AWGN noise and the level of the phase and polarization noise, as shown in continuous and dashed lines in the figure.

From practical point of view, converging to the minimum value of $\overline{e}_{\mathbf{P}i}^{(n)}$ by fewer number of iterations is desirable, considering the amount of processing required in DSP. Therefore, in Figure 5.2, where the settings are the same as in Figure 5.1, respectively, we examine the average amount of the required iterations n in the flexible case to converge to the error floor in Figure 5.1 when $\mu_{\mathrm{P}} = \{0.4, 0.6, 0.8\}$, for PM-QPSK, PM-16-QAM, PM-64-QAM, and PM-256-QAM. The dashed lines in the figure show the average number of required iterations n; when $\mu_{\mathrm{P}} = 0.6$, on average with $n = 9$ iterations the in (4.22) can be satisfied, for all constellations. This is a significant reduction in the number of required iterations n to minimize $\overline{e}_{\mathrm{P}i}^{(n)}$. As n will be the smallest when $\mu_{\mathrm{P}} = 0.6$, this is offering us to choose it as the proper step size, which shall be normalized to $\mu_{\mathrm{P}} = 0.6/E_{\mathbf{P}}$, where $E_{\mathbf{P}}$ is the average energy of the pilot packet, defined in (4.33).

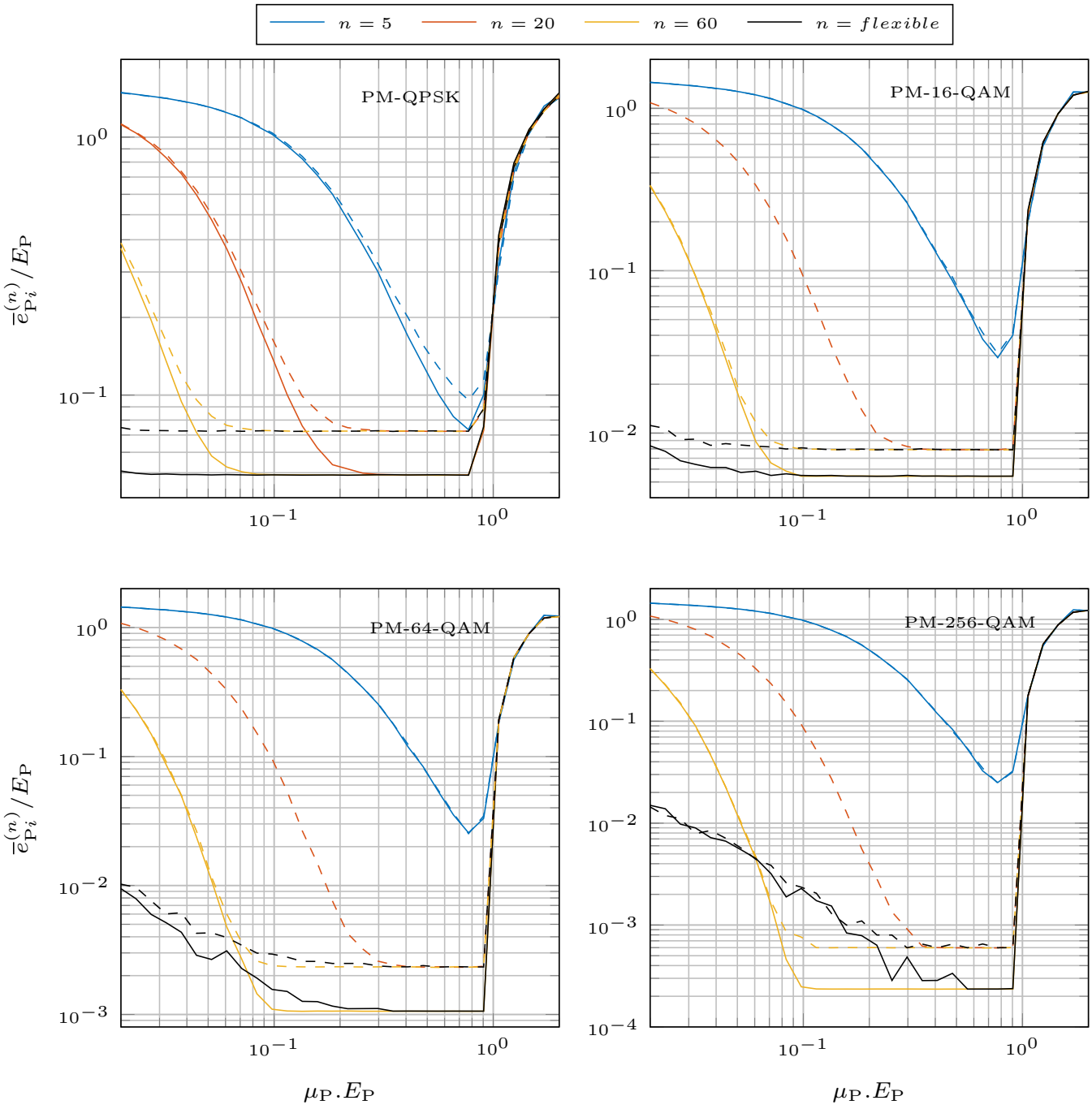

Figure 5.1: For different number of iterations of the gradient descent method, and also for flexible number of iterations, the achievable level of $\overline{e}_{\mathbf{P}i}^{(n)}$ is shown for different step sizes μ_{P}. Dashed lines correspond to the case that both $\Delta\nu \cdot T$ and $\Delta\rho \cdot T$ are 100 *times* higher than the continuous lines, to evaluate the effect of those impairments on the achievable level of $\overline{e}_{\mathbf{P}i}^{(n)}$.

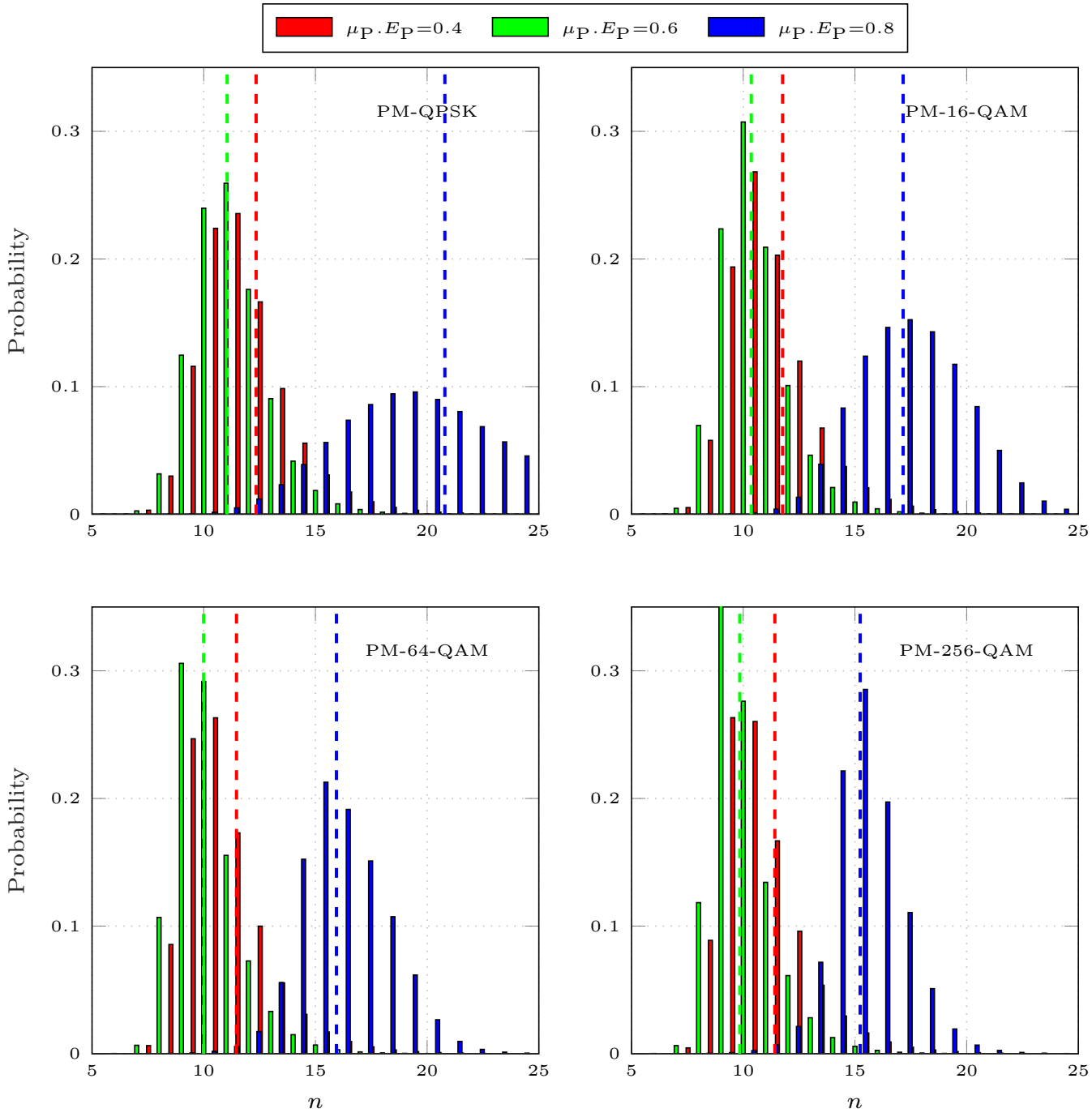

Figure 5.2: In the case that the number of iterations n of the gradient descent method, is flexible, it is shown, which step size $\mu_P = \{0.4, 0.6, 0.8\}$, gives the fastest convergence to the minimum of $\overline{e}_{Pi}^{(n)}$. The dashed lines shows the average number of required iterations n to converge to the minimum of $\overline{e}_{Pi}^{(n)}$.

5.1.2 Length of Pilot Packet

In this section, we investigate the proper length of the pilot packet, by looking at both $\overline{e}_{\mathrm{P}i}^{(n)}$ and MSE between $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ and $H(\phi, \boldsymbol{\alpha})$. As explained in Section 4.3.4, MSE is a better than $\overline{e}_{\mathrm{P}i}^{(n)}$, for choosing the proper length of the pilot packet.

In Figure 5.3, $\Delta\nu \cdot T$, $\Delta\rho \cdot T$, and SNR settings, as well as the number of realizations of pilot packets are the same as in Figure 5.1. $\overline{e}_{\mathrm{P}i}^{(n)}$ and MSE are drawn for different pilot packet lengths L. Both curves in Figure 5.3 (left), show when $L = 2$, $\overline{e}_{\mathrm{P}i}^{(n)}$ gets its lowest amount. Here, the channel estimation is achieved by using two pilot symbols, which makes the estimation more affected by AWGN noise, than the case that the pilot packet length is larger. The influence of the AWGN noise on the channel estimation is seen better in Figure 5.3 (right), where for $L = 2$, MSE gets its highest value in the red curve. This is coincident with what explained in (4.28), which for low or non-varying channel matrix $H(\phi, \boldsymbol{\alpha})$ over the interval of L pilot symbols(AWGN scenario), increasing the length of the pilot packet makes the the estimation of the channel matrix $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ more accurate.

The blue curves in Figure 5.3, show the case that both phase noise and SOP drift of the channel are 100 times higher than the red curves. It can be seen from the blue curves that by increasing the length of the pilot packet, both $\overline{e}_{\mathrm{P}i}^{(n)}$ and MSE increase. This can be explained by the fact that the channel matrix $H(\phi, \boldsymbol{\alpha})$ changes significantly from the pilot symbol one to the pilot symbol $L = 100$, due to rather high phase noise and SOP drift. Therefore, if the channel state changes considerably during the L pilot symbols, the channel estimation will be inaccurate. Later in Section 5.2.1, the consequences of choosing different pilot packet length L on the performance of algorithm will be examined.

In Figure 5.4, we investigate the influence of choosing a very long pilot packet on the SER, by averaging over 10^5 transmissions. In the red curve, the length of the pilot packet $L = 200$, and the channel matrix $H(\phi, \boldsymbol{\alpha})$ is not varying during the pilot packet. Hence, the channel estimate $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ is very close to the real channel matrix, resulting in no penalty on the SER at the very beginning of the data packet. In the blue and green curves, where $L = 200$ and $L = 500$, respectively, the phase and polarization noise are considered in the channel matrix. It is obvious from the curves that there is a penalty in theSER at the beginning of the data packet, although it converges very fast to SER $= 10^{-3}$ due to the ability of the algorithm to track the channel matrix. This test is done only to give a better understating of how the averaged channel estimation $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}$ works, and how it is affected by the length of the pilot packet L, and the considered channel impairments. In the real implementation of the algorithm, surely such long pilot packets are not interesting.

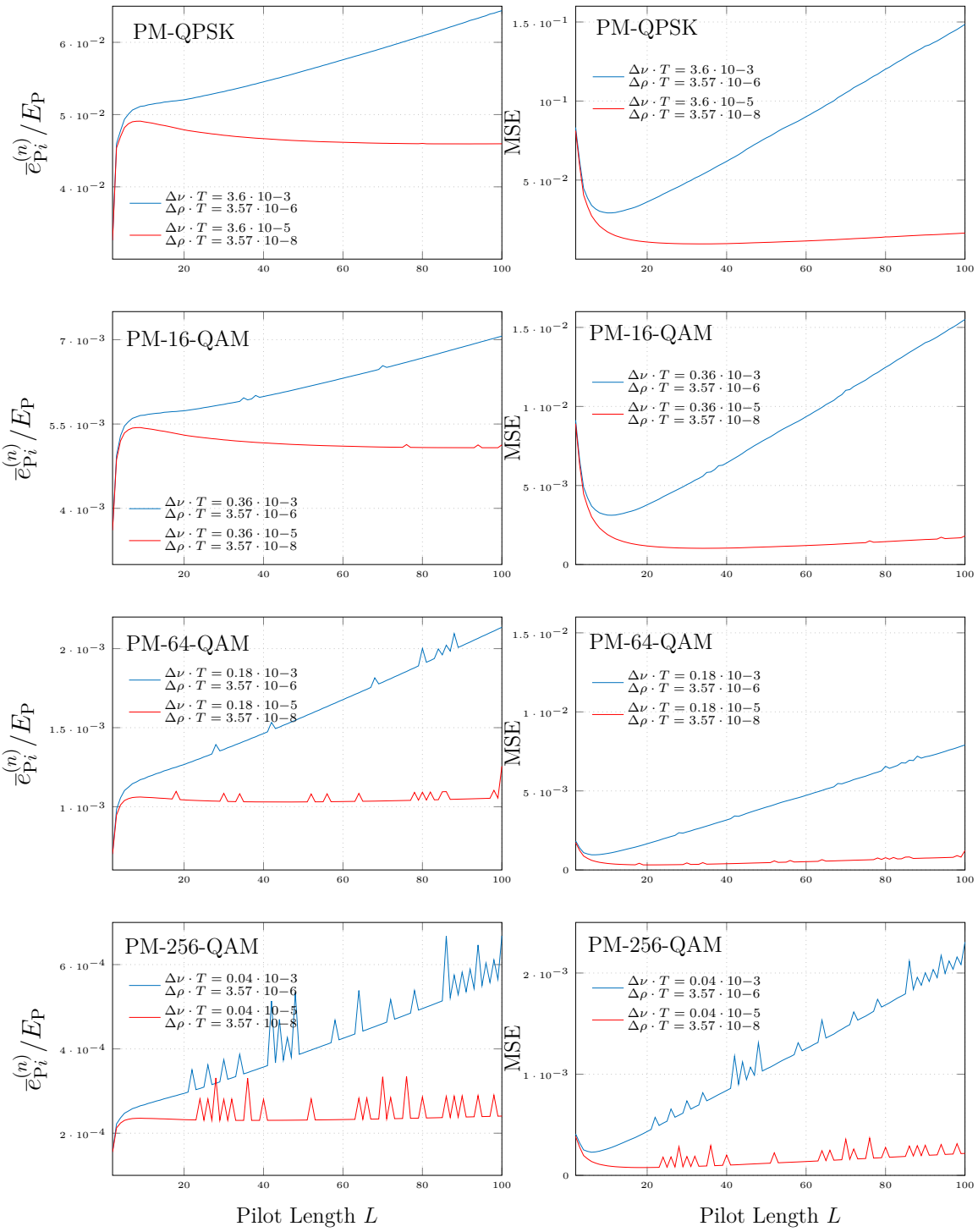

Figure 5.3: The figure shows the behaviour of both $\overline{e}_{\mathrm{P}i}^{(n)}$ (left) and MSE (right) when the pilot length increases, for two different scenarios of the phase and polarization noise. The parameter L, which minimizes the MSE, is the proper length of the pilot packet.

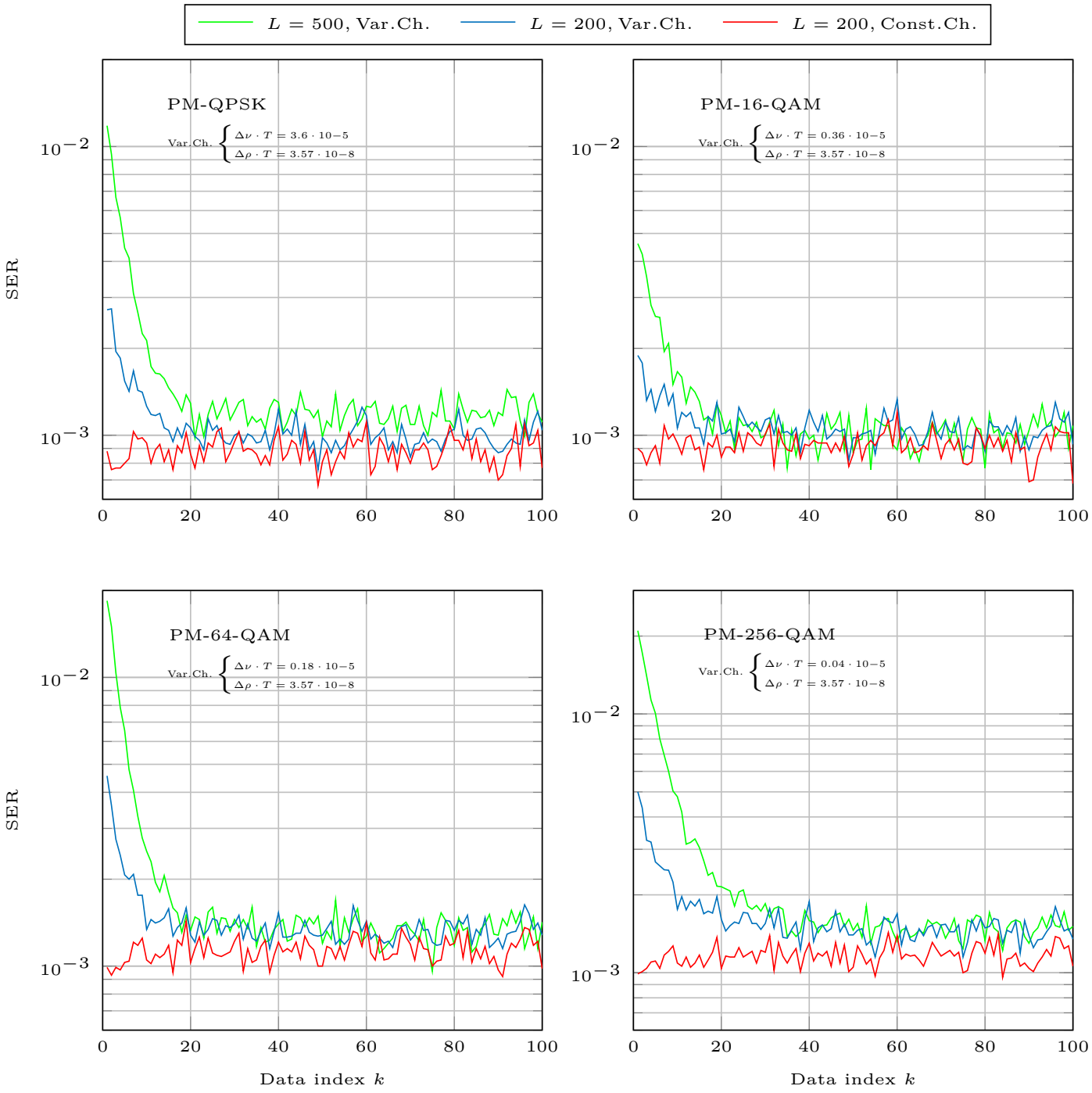

Figure 5.4: The effect of choosing a rather long pilot packet L on the convergence of the pilot-aided algorithm is shown for two scenarios, in the presence of the phase and polarization noise, and when the channel matrix is constant during the pilot packet.

5.2 Performance of Proposed Algorithm

In this section, the performance of the pilot-aided algorithm in Section 4.3 is compared with the blind algorithm in Section 4.2. In Section 5.2.1 the convergence of the pilot-aided algorithm is evaluated, in Section 5.2.2 the performance of the pilot-aided algorithm in the steady state regime is discussed. Section 5.2.3 investigates the tolerance of the pilot-aided algorithm against the phase and polarization noise, and further examines the occurrence of CS for different pilot rate R and pilot length L. Section 5.2.4 considers the sensitivity of the algorithm against the additive Gaussian noise.

5.2.1 Convergence

In Figure 5.5, the SNR is set such that SER $= 10^{-3}$ is achieved in the steady state regime. The SER is obtained by averaging over 10^5 transmissions of pilot and data packets, and the initial state of the estimated channel matrix is set to be $\hat{\mathbf{H}}_{\mathbf{P}}^{(0)} = \mathbf{I}_2$. The convergence rate of the pilot-aided algorithm for pilot packets with different length, is investigated. As can be seen in the figure, when L increases, due to the better estimation of the channel matrix at the beginning of the data packet, the algorithm converges faster. Furthermore, this also can be seen that for the higher order modulations, the convergence rate is less dependent to the length of the pilot packet L, due to employing the higher SNR.

Even though, it not is shown here, but it should be mentioned that the blind tracking algorithm using *differential encoding*, shows a very slower convergence rate compared to the pilot-aided algorithm. This is quite expected, as the data transmission starts when there is no information about the initial channel state $\mathbf{H}_0$, and the algorithm should track the channel in the way explained in Section 4.2. This can take *several thousand* symbols to converge to SER $= 10^{-3}$. The convergence rate of the blind algorithm has been investigated in [25].

5.2.2 Steady State Performance

In Figure 5.6, the performance of the pilot-aided algorithm in the steady state regime, is investigated, the settings are the same as in Section 5.2.1. The length of the pilot packet is $L = 10$, and the length of the data packet is $K = 400$, which implies that at the beginning of each 400 received data symbols, there is a packet of 10 pilot symbols. The total number of received data symbols is 1200.

The initial channel state at the beginning of each data packet is re-estimated by using the pilot packets, then during the data packets, the channel is tracked, the result shown in Figure 5.6 (left). The figure shows average SER versus data symbol k. It is obvious that there is an increase in the SER at the beginning of each data packet, which is not a desirable result. Inserting a pilot packet between the data packets should not influence the performance of the system in the steady state regime. However, this effect is unavoidable for the first data packet, where there is no information about the initial channel matrix. As proposed in Section 4.3.3, the track of the channel matrix during the pilot packet should be kept at the same way done during the data packet. If the number of erroneous received pilot symbols is

less then 20% of the pilot packet length, it can be supposed that a CS did not occur in the previous data packet, and the current estimation of the channel is accurate enough to continue with. Otherwise, the channel needs to be re-estimated by using the pilot packet. By applying this assumption to the pilot-aided algorithm, the problem is solved as can be seen in Figure 5.6 (right).

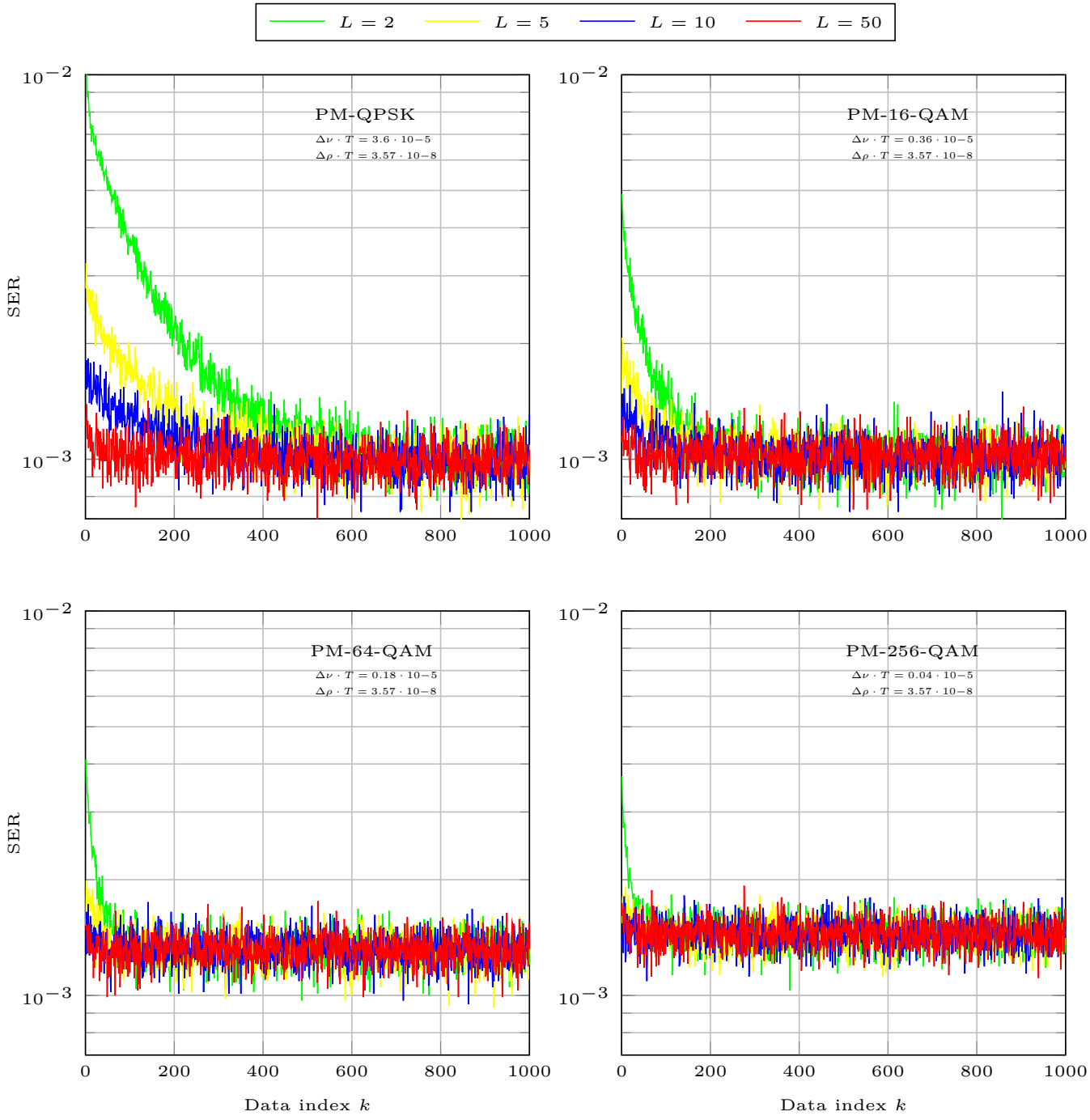

Figure 5.5: The convergence rate of the pilot-aided algorithms is shown for different pilot lengths L.

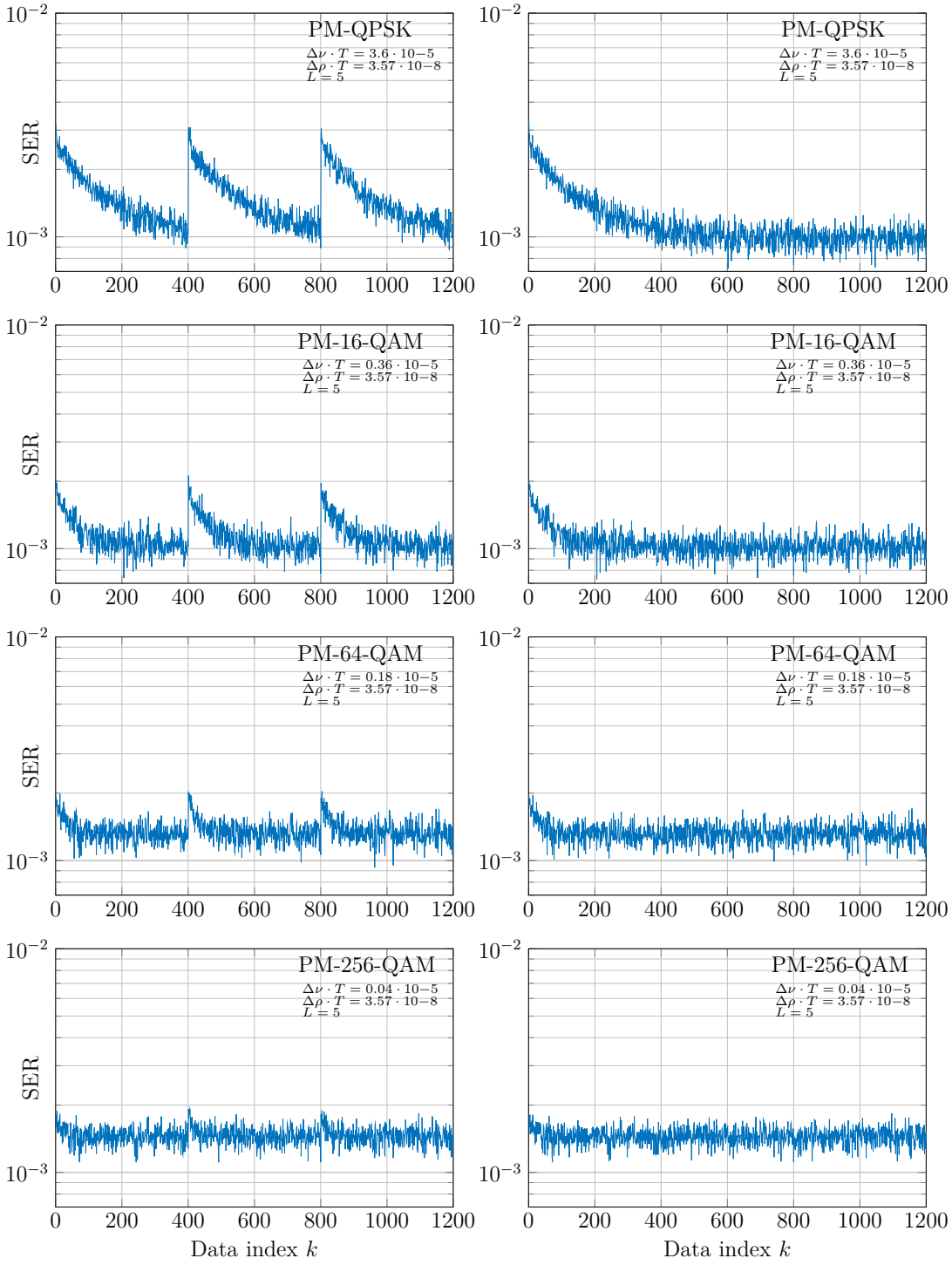

Figure 5.6: The behaviour of the algorithm is shown when using every received pilot packet to re-estimate the channel matrix (left), and when re-estimates the channel matrix only if more than 20% of the received pilot symbols in a pilot packet are erroneous (right).

5.2.3 Noise Tolerance and CS Probability

An important feature of tracking algorithm is the tolerance to the phase and polarization noise. In this section, we will test the ability of the algorithm to track the phase noise and the SOP drift simultaneously, for pilot lengths $L = \{5, 10, 50\}$, and pilot rates $R = \{0.01, 0.001\}$, and also compare them with the blind algorithm. In Figure 5.7, SER is obtained by averaging over 10^5 realizations, where at each realization the initial channel matrix is randomly generated. In this simulation, we suppose the initial channel matrix $\mathbf{H}_0$ is known to the receiver, i.e., $\hat{\mathbf{H}}_\mathbf{0} = \mathbf{H}_0$ for the blind algorithm, while for the pilot-aided algorithm $\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)} \approx \mathbf{H}_0$, for the n that satisfies the stopping criterion defined in (4.22). The SNR is set to achieve a SER $= 10^{-3}$ in an AWGN scenario for the blind algorithm.

As the results show in Figure 5.7, for lower order modulations, the pilot-aided algorithm shows a better tolerance against lower amount of the phase and polarization noise, compared to the blind algorithm. However, for the higher order modulations, there is not a considerable difference between two algorithms, due to employing a higher SNR. But when the phase and polarization noise are increased, the blind algorithm shows a better tolerance against those impairments than the pilot-aided algorithm for all considered modulation formats. This behaviour is expected, because in the blind algorithm, in which *differential encoding* has been implemented, the CS is mitigated, so it is less sensitive to the higher amount of the phase noise and the SOP drift.

It is also understood from the figure that in the pilot-aided algorithm, the tolerance against the phase and polarization noise decreases by increasing the data packet length K, where $K = L/R$.

The length of the pilot interval K, or in the other words, the data packet length has a direct relation with the occurance of a CS, this is investigated in Figure 5.8. The settings in the figure are the same as in Figure 5.7. The probability of a CS occurrence over 10^5 transmissions is calculated for the given range of the phase noise and the SOP drift. As can be seen in the figure, this probability increases by increasing the data packet length $K = L/R$.

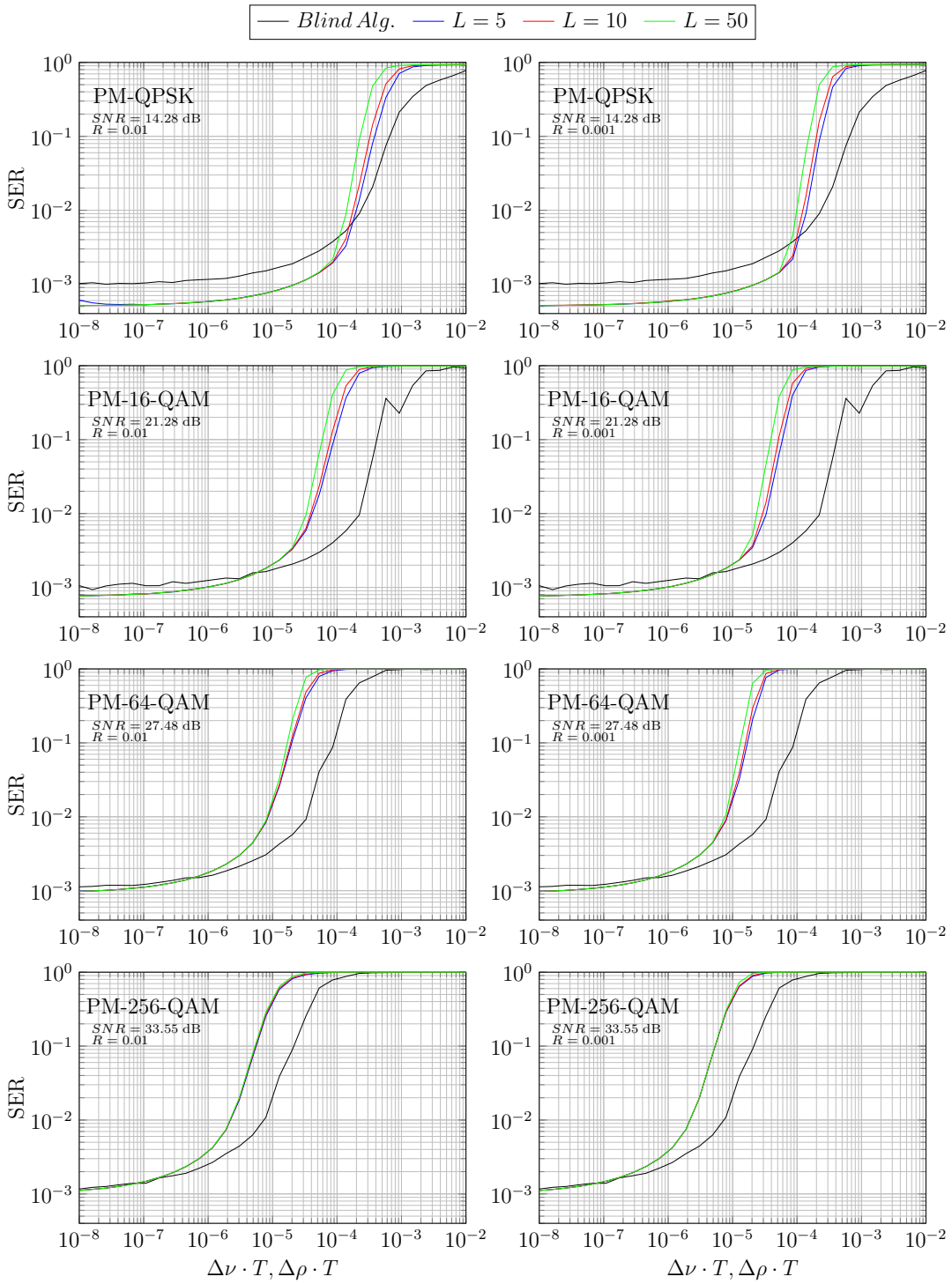

Figure 5.7: Tolerance of the blind and pilot-aided algorithms against the phase and polarization noise is shown for different pilot lengths L, and for two different pilot rates R. The figure shows by increasing the data packet length $K = L/R$, the tolerance against the phase and polarization noise drops down. This can be seen also that the pilot-aided algorithm is more sensitive to the high phase and polarization noise than the blind algorithm, which uses differential encoding.

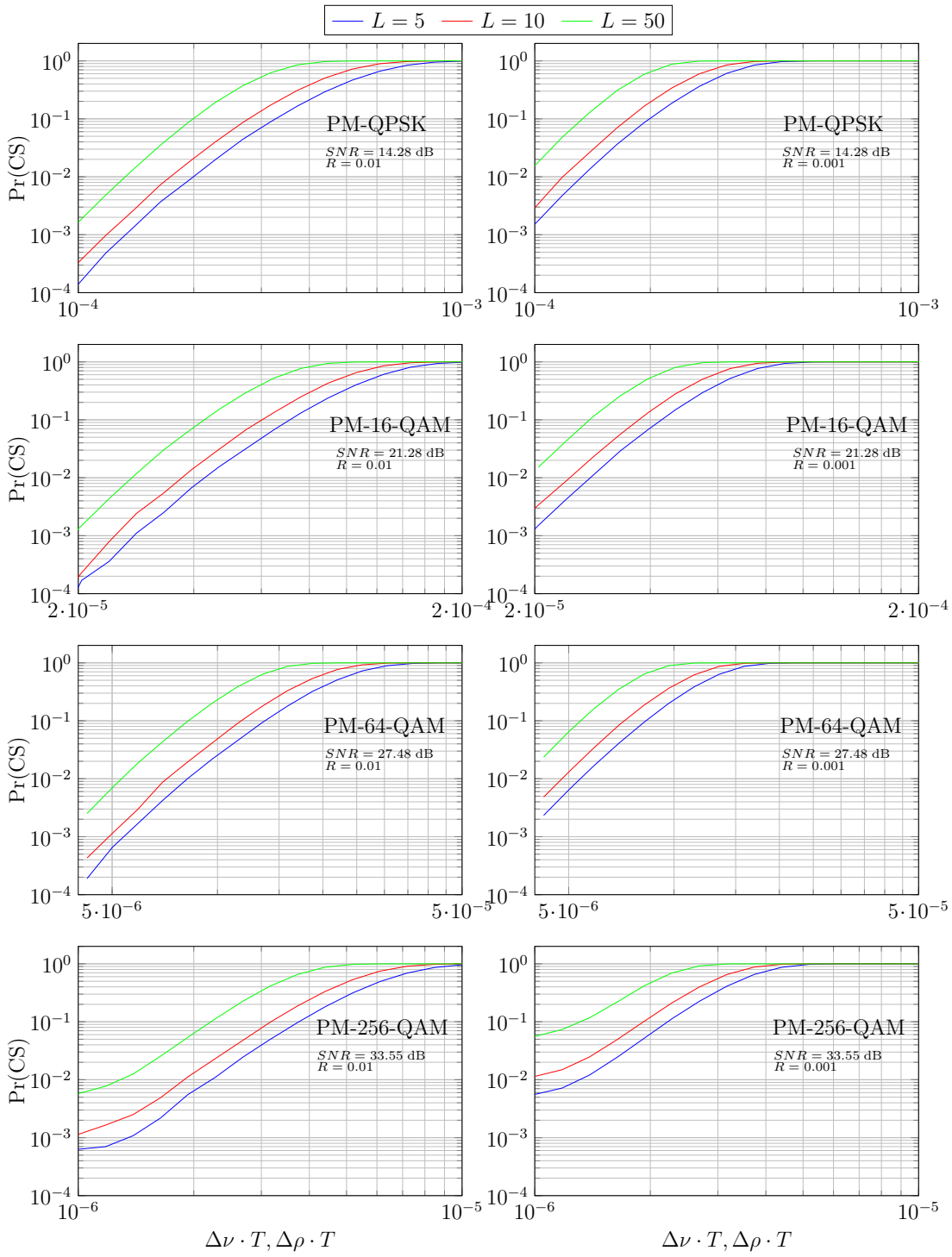

Figure 5.8: The CS probability in 10^5 transmissions of one pilot and one data packet is shown for two different pilot rates R, and different pilot lengths L, for the given range of the phase and polarization noise.

5.2.4 Additive-Noise Sensitivity

Figure 5.9 shows the SER as a function of SNR. It can be seen in the figure that the blind algorithm compared to the pilot-aided algorithm, has a penalty on the SNR at SER $= 10^{-3}$, induced by *differential encoding*. For the pilot-aided algorithm, the average symbol energy E_s is adjusted for the pilot rates $R = \{0.01, 0.001\}$, so that

$$\mathrm{SNR} = \mathrm{SNR}_k - 10\log_{10}(1 - R), \tag{5.1}$$

where SNR_k implies the dedicated energy to the data symbols. The figure also shows for the given pilot length $L = 5$, for both $R = \{0.01, 0.001\}$, the additive-noise sensitivity is homogeneous. Even though, it is not examined in this report, but it can be expected that for the bigger pilot rates R, we would observe a deterioration in the SER compared to those shown in Figure 5.9. For comparison, in Figure 5.9, the AWGN scenario is drawn, it can be seen for the higher order modulations the performance of both blind and pilot aided-algorithm is almost the same as the AWGN only scenario.

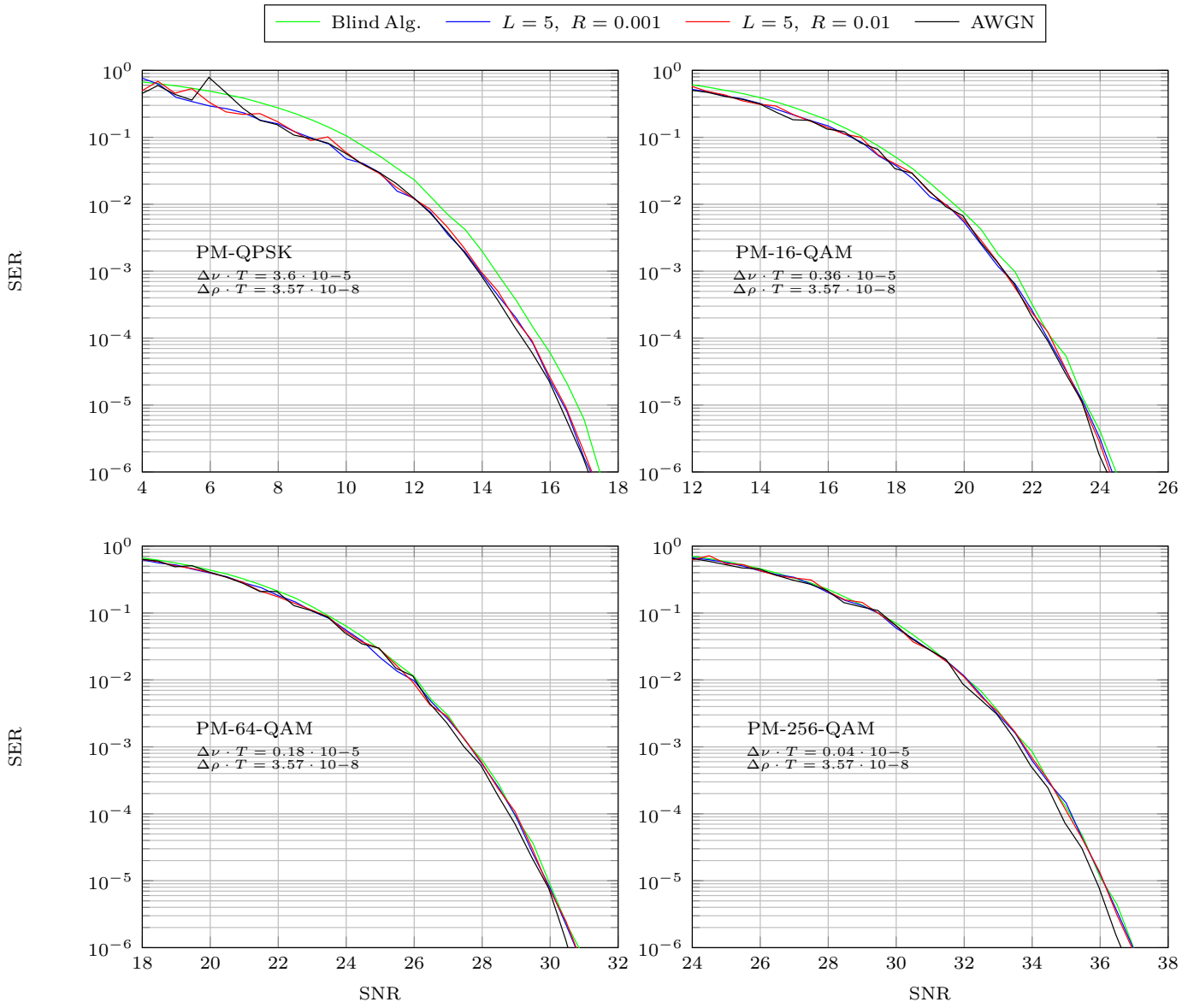

Figure 5.9: The additive-noise sensitivity of both algorithms is compared with each other and also with the AWGN only scenario, where in the pilot-aided algorithm for a given pilot packet length $L = 5$, the pilot rate is $R = \{0.01, 0.001\}$. This means the average symbol energy E_s should be adjusted based on the R in each case.

Chapter 6

Conclusions

This work presents a pilot-aided joint phase and polarization tracking algorithm. This algorithm uses pilot packets inserted between data packets with the rate $R = L/K$. The pilot packet detects if the track of channel is lost, and re-estimate the channel matrix for the next data packet, using the gradient descent method. The proposed pilot-aided algorithm is based on the blind joint phase and polarization tracking algorithm in [25], so a part of this thesis is to compare the performance of both algorithms.

The pilot-aided algorithm has some advantages over the blind one listed below.

- By choosing the proper pilot packet length L, the convergence rate of the pilot-aided algorithm tends to zero, while in the blind one, after several thousands symbols the algorithm converges.
- The SNR penalty induced due to employing differential coding in the blind algorithm, is eliminated in the pilot-aided one.
- The blind algorithm suffers from the polarization ambiguity, while it is removed in the pilot-aided. The polarization rotation is detected by the pilot packet.
- The pilot-aided algorithm has a better tolerance against the phase noise and the SOP drift when they are rather low.

The main drawback of the pilot-aided algorithm is its complexity compared to the blind one, which is a simple and easily implementable algorithm. Even though the hardware complexity of the pilot-aided algorithm is not calculated in this thesis, but this algorithm demands more settings and operations than the blind algorithm. The other drawback of the pilot-aided algorithm is the reduced spectral efficiency by the factor $1/(1 + R)$, compared to the blind algorithm.

Appendix A

Appendix

In this appendix, we derive the gradient of the average error function (4.15) with respect to ϕ and $\boldsymbol{\alpha}$. This derivation is independent of the number of iterations n and the packet index i, so for ease of notation

$$\overline{e}_{\mathrm{P}i}^{(n)} = \frac{1}{L}\sum_{l=1}^{L}\left\|\left(H(\phi,\boldsymbol{\alpha})\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}\right)^{-1}\mathbf{r}_{\mathbf{P}i,l} - \mathbf{u}_{\mathbf{P}i,l}\right\|^2, \tag{A.1}$$

is used instead of (4.15) in this derivation.

$$\begin{aligned}
\nabla_{\boldsymbol{\alpha}}\,\overline{e}_{\mathrm{P}i}^{(n)} &= \nabla_{\boldsymbol{\alpha}}\left(\frac{1}{L}\sum_{l=1}^{L}\left\|\left(H(\phi,\boldsymbol{\alpha})\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}\right)^{-1}\mathbf{r}_{\mathbf{P}i,l} - \mathbf{u}_{\mathbf{P}i,l}\right\|^2\right) && \text{(A.2)}\\
&= \frac{1}{L}\sum_{l=1}^{L}\nabla_{\boldsymbol{\alpha}}\left\|\left(H(\phi,\boldsymbol{\alpha})\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}\right)^{-1}\mathbf{r}_{\mathbf{P}i,l} - \mathbf{u}_{\mathbf{P}i,l}\right\|^2\\
&= \frac{1}{L}\sum_{l=1}^{L}\mathcal{R}\mathrm{e}\Bigg(2\Bigg(\left(H(\phi,\boldsymbol{\alpha})\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}\right)^{-1}\mathbf{r}_{\mathbf{P}i,l} - \mathbf{u}_{\mathbf{P}i,l}\Bigg)^{\mathrm{H}}\\
&\qquad\qquad \cdot\nabla_{\boldsymbol{\alpha}}\left(\left(H(\phi,\boldsymbol{\alpha})\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}\right)^{-1}\mathbf{r}_{\mathbf{P}i,l} - \mathbf{u}_{\mathbf{P}i,l}\right)\Bigg) && \text{(A.3)}\\
&= \frac{1}{L}\sum_{l=1}^{L}2\mathcal{R}\mathrm{e}\Bigg(\left(\left(\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}\right)^{-1}H(-\phi,-\boldsymbol{\alpha})\mathbf{r}_{\mathbf{P}i,l} - \mathbf{u}_{\mathbf{P}i,l}\right)^{\mathrm{H}}\left(\hat{\mathbf{H}}_{\mathbf{P}i}^{(n)}\right)^{-1}\\
&\qquad\qquad \cdot\nabla_{\boldsymbol{\alpha}}H(-\phi,-\boldsymbol{\alpha})\mathbf{r}_{\mathbf{P}i,l}\Bigg), && \text{(A.4)}
\end{aligned}$$

The equation (A.4) stands because, $\nabla_{\mathbf{x}}\|\mathbf{y}\|^2 = \nabla_{\mathbf{x}}(\mathbf{y}^{\mathrm{H}}\mathbf{y}) = 2\mathcal{R}\mathrm{e}(\mathbf{y}^{\mathrm{H}}\nabla_{\mathbf{X}}\mathbf{y})$ for any $\mathbf{y}\in\mathbb{C}^n$ and $\mathbf{x}\in\mathbb{R}^m$. The partial derivatives of $H(-\phi,-\boldsymbol{\alpha}) = e^{i\phi}J(-\boldsymbol{\alpha})$ from (3.14)

with respect to ϕ and $-\boldsymbol{\alpha}$ are,

$$\frac{\partial\, H(-\phi,-\boldsymbol{\alpha})}{\partial\phi} = ie^{i\phi}J(-\boldsymbol{\alpha}), \tag{A.5}$$

$$\frac{\partial\, H(-\phi,-\boldsymbol{\alpha})}{\partial\alpha_i} = e^{i\phi}\,\frac{\partial\, J(-\boldsymbol{\alpha})}{\partial\alpha_i} \tag{A.6}$$

$$= e^{i\phi}\,\frac{\partial\,(\mathbf{I}_2\cos\theta\,+\,i\boldsymbol{a}\,.\,\vec{\boldsymbol{\sigma}}\sin\theta)}{\partial\alpha_i} \tag{A.7}$$

$$= e^{i\phi}\Big(-\mathbf{I}_2\alpha_i\sin\theta\,+\,i\Big(\frac{\boldsymbol{\sigma}_i}{\theta}\,-\,\frac{\alpha_i}{\theta}\boldsymbol{a}\cdot\vec{\boldsymbol{\sigma}}\Big)\sin\theta + i\boldsymbol{a}\cdot\vec{\boldsymbol{\sigma}}\,\alpha_i\cos\theta\Big), \tag{A.8}$$

where (A.8) stands because $\partial\theta/\partial\alpha_i = \alpha_i/\|\boldsymbol{\alpha}\| = \alpha_i$. The derivatives in (A.5) and (A.8) should be evaluated at $\phi = 0$ and $\boldsymbol{\alpha} = [0,0,0]^{\mathrm{T}}$, respectively which results in

$$\frac{\partial\, H(-\phi,-\boldsymbol{\alpha})}{\partial\phi} = i\mathbf{I}_2, \qquad \frac{\partial\, H(-\phi,-\boldsymbol{\alpha})}{\partial\alpha_i} = i\boldsymbol{\sigma}_i\,. \tag{A.9}$$

By substituting (A.9) in (A.4), the expression for $\nabla_{\boldsymbol{\alpha}}\,\overline{e}_{\mathrm{P}i}^{(n)}$ is acquired which is substituted in (4.17) and (4.19) to obtain (4.18) and (4.20), respectively. It should be noted that this appendix is a duplication of the appendix of [25] justified for the summation case.

Bibliography

[1] "The zettabyte era–trends and analysis," Cisco Visual Networking, Tech. Rep., 2013.

[2] K. Kao and G. A. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," *Proceedings of the IEEE*, vol. 113, no. 7, pp. 1151–1158, 1966.

[3] F. Kapron, D. Keck, and R. Maurer, "Radiation losses in glass optical waveguides," *Applied Physics Letters*, vol. 17, no. 10, pp. 423–425, 1970.

[4] S. Abbott, "Review of 20 years of undersea optical fiber transmission system development and deployment since TAT-8," in *in Proc. European Conference on Optical Communication (ECOC)*, Brussels, Belguim, Sept. 2008, p. Mo.4.E.1.

[5] R. Mears, L. Reekie, I. Jauncey, and D. Payne, "Low-noise erbium-doped fibre amplifier operating at 1.54μm," *Electronics Letters*, vol. 23, no. 19, pp. 1026–1028, Sept. 1987.

[6] E. Desurvire, J. R. Simpson, and P. Becker, "High-gain erbium-doped traveling-wave fiber amplifier," *Optics Letters*, vol. 12, no. 11, pp. 888–890, 1987.

[7] E. Agrell, M. Karlsson, A. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton *et al.*, "Roadmap of optical communications," *Journal of Optics*, vol. 18, no. 6, p. 063002, 2016.

[8] H. Sun, K.-T. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system," *Optics Express*, vol. 16, no. 2, pp. 873–879, 2008.

[9] E. Agrell and M. Karlsson, "Power-efficient modulation formats in coherent transmission systems," *Journal of Lightwave Technology*, vol. 27, no. 22, pp. 5115–5126, Nov. 2009.

[10] T. Pfau, S. Hoffmann, O. Adamczyk, R. Peveling, V. Herath, M. Porrmann, and R. Noé, "Coherent optical communication: Towards realtime systems at 40 Gbit/s and beyond," *Optics express*, vol. 16, no. 2, pp. 866–872, 2008.

[11] S. J. Savory, G. Gavioli, R. I. Killey, and P. Bayvel, "Electronic compensation of chromatic dispersion using a digital coherent receiver," *Optics Express*, vol. 15, no. 5, pp. 2120–2126, 2007.

[12] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265–298, 1997.

[13] C. B. Czegledi, M. R. Khanzadi, and E. Agrell, "Bandlimited power-efficient signaling and pulse design for intensity modulation," *IEEE Transactions on Communications*, vol. 62, no. 9, pp. 3274–3284, 2014.

[14] G. Bennett, K.-T. Wu, A. Malik, S. Roy, and A. Awadalla, "A review of high-speed coherent transmission technologies for long-haul DWDM transmission at 100G and beyond," *IEEE Communications Magazine*, vol. 52, no. 10, pp. 102–110, 2014.

[15] A. M. Joshi, S. Datta, and A. Crawford, "Multilevel modulation formats push capacities beyond 100 Gbit/s," *Laser Focus World*, vol. 48, no. 2, pp. 58 – 62, 2012.

[16] G. Goldfarb and G. Li, "Chromatic dispersion compensation using digital IIR filtering with coherent detection," *IEEE Photonics Technology Letters*, vol. 19, no. 13, pp. 969–971, 2007.

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

[18] B. Glance, "Polarization independent coherent optical receiver," *Journal of Lightwave Technology*, vol. 5, no. 2, pp. 274–276, 1987.

[19] F. Favre, L. Jeunhomme, I. Joindot, M. Monerie, and J. C. Simon, "Progress towards heterodyne-type single-mode fiber communication systems," *IEEE Journal of Quantum Electronics*, vol. 17, no. 6, pp. 897–906, 1981.

[20] J. R. Barry and E. A. Lee, "Performance of coherent optical receivers," *Proceedings of the IEEE*, vol. 78, no. 8, pp. 1369–1394, 1990.

[21] C. Fludger, J. Geyer, T. Duthel, and C. Schulien, "Digital signal processing-From simulation to silicon," in *Proc. European Conference on Optical Communication (ECOC)*, Torino, Italy, Sept. 2010, p. Tu.5.A.1.

[22] M. G. Taylor, "Coherent detection method using DSP for demodulation of signal and subsequent equalization of propagation impairments," *IEEE Photonics Technology Letters*, vol. 16, no. 2, pp. 674–676, 2004.

[23] G. P. Agrawal, *Fiber-Optic Communication Systems.* Hoboken, NJ: John Wiley & Sons, 2010.

[24] K. Petermann, *Laser Diode Modulation and Noise.* Dordrecht, The Netherlands: Kluwer Academic Publishers, 1988.

[25] C. B. Czegledi, E. Agrell, M. Karlsson, and P. Johannisson, "Modulation format independent joint polarization and phase tracking for coherent receivers," *Journal of Lightwave Technology*, vol. 34, no. 14, pp. 3354–3364, 2016.

[26] C. B. Czegledi, M. Karlsson, E. Agrell, and P. Johannisson, "Polarization drift channel model for coherent fibre-optic systems," *Nature Scientific Reports*, vol. 6, 2016.

[27] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on selected areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[28] J. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, pp. 4541–4550, 2000.

[29] R. Bellman, *Introduction to Matrix Analysis*. New York, NY: McGraw-Hill, 1960.

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

[31] M. Tur, B. Moslehi, and J. Goodman, "Theory of laser phase noise in recirculating fiber-optic delay lines," *Journal of lightwave technology*, vol. 3, no. 1, pp. 20–31, 1985.

[32] H. Louchet, K. Kuzmin, and A. Richter, "Improved DSP algorithms for coherent 16-QAM transmission," in *Proc. European Conference on Optical Communication (ECOC)*, Brussels, Belgium, Sept. 2008, p. Tu.1.E.6.

[33] ——, "Joint carrier-phase and polarization rotation recovery for arbitrary signal constellations," *IEEE Photonics Technology Letters*, vol. 26, no. 9, pp. 922–924, 2014.

[34] M. Seimetz, *High-Order Modulation for Optical Fiber Transmission*. Heidelberg, Germany: Springer, 2009.

[35] J. B. Anderson, *Digital Transmission Engineering*. John Wiley & Sons, 2006.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

[37] S. M. Kay, *Fundamentals of Statistical Signal Processing, volume I: Estimation Theory*. Prentice Hall PTR, 1993.

[38] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE transactions on signal processing*, vol. 54, no. 3, pp. 884–893, 2006.