THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Phase Noise and Polarization Effects in Fiber-Optic Communication Systems: Modeling, Compensation, Capacity, and Sensing

Mohammad Farsi

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Göteborg, Sweden, 2024

Phase Noise and Polarization Effects in Fiber-Optic Communication Systems: Modeling, Compensation, Capacity, and Sensing

Mohammad Farsi
ISBN 978-91-8103-051-8

Acknowledgements, dedications, and similar personal statements in this thesis, reflect the author's own views.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5509
ISSN 0346-718X
This thesis has been prepared using LaTeX, Tikz, and PgfPlots.

Department of Electrical Engineering
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Phone: +46 (0)31 772 1000
www.chalmers.se

Printed by Chalmers Reproservice
Göteborg, Sweden, May, 2024

To my loving wife, Neda, and my everliving brother Abbas.

Phase Noise and Polarization Effects in Fiber-Optic Communication Systems

Mohammad Farsi
Department of Electrical Engineering
Chalmers University of Technology

Abstract

High data rate optical communications are susceptible to phase noise and state of polarization (SOP) perturbations. The dynamic nature of phase noise and SOP fluctuations requires a comprehensive investigation of their effects on the channel capacity and the development of robust and efficient communication technologies. This thesis unravels phase and polarization challenges in optical communication systems by characterizing polarization drift channels, introducing polarization tracking algorithms, utilizing polarization data for fiber sensing, and investigating capacity implications.

The SOP drifts at a much slower rate than typical transmission rates in buried or underwater fibers. Thus, we first characterize the capacity of the block-constant polarization drift channel under an average power constraint and imperfect channel knowledge. An achievable information rate is derived, showing strong dependence on the channel estimation technique. A novel data-aided channel estimator is proposed, enforcing the unitary constraint, and its superior performance is validated through Monte Carlo simulations. However, in aerial fibers, the SOP drift does not follow the block-constant assumption and can drift quickly over time. Hence, the next contribution involves investigating the robustness of polarization tracking algorithms in the presence of fast SOP drift and polarization-dependent loss. Novel tracking algorithms are proposed, showing a higher tolerance to SOP drift compared to the gradient descent-based algorithms without the need for parameter tuning. Thereafter, we explore the application of polarization for fiber sensing by proposing a physics-based learning approach. The proposed approach shows lower sensitivity to additive noise compared to previous inverse scattering methods.

Next, we turn our attention to phase noise and investigate the capacity of a discrete-time multiple-input-multiple-output channel with correlated phase noises originating from electro-optic frequency combs (EO-comb). We derive capacity bounds and show that the multiplexing gain is $M-1$ where M is the number of channels. Moreover, a constant gap between the bounds is observed in the high signal-to-noise ratio regime, which vanishes for the special case of $M=2$. Finally, we study optimal pilot placement for channels impaired by phase noise from EO-combs. Contrary to regular multichannel systems, it is demonstrated that allocating the first and last channels for pilots is optimal under a fixed pilot overhead.

Keywords: Capacity, duality bound, electro-optic frequency comb, fiber sensing, mismatched decoding, phase noise, polarization dependent loss, polarization drift.

List of Papers

This thesis is based on the following publications:

[A] **Mohammad Farsi**, Magnus Karlsson, and Erik Agrell, "Capacity Bounds under Imperfect Polarization Tracking," *IEEE Transactions on Communications*, vol. 70, no. 11, pp. 7240–7249, Nov. 2022.

[B] **Mohammad Farsi**, Christian Häger, Magnus Karlsson, and Erik Agrell, "Polarization Tracking in the Presence of PDL and Fast Temporal Drift," *IEEE Journal of Lightwave Technology*, vol. 40, no. 19, pp. 6408–6416, Oct. 2022.

[C] **Mohammad Farsi**, Christian Häger, Magnus Karlsson, and Erik Agrell, "Learning to Extract Distributed Polarization Sensing Data from Noisy Jones Matrices," in *Proc. Optical Fiber Communication Conference (OFC)*, San Diego, CA, Mar. 2024, Paper Th2A.

[D] **Mohammad Farsi**, Hamdi Joudeh, Gabriele Liga, Alex Alvarado, Magnus Karlsson, and Erik Agrell, "On the Capacity of Correlated MIMO Phase-Noise Channels: An Electro-Optic Frequency Comb Example," Submitted to *IEEE Transactions on Information Theory*, May 2024.

[E] **Mohammad Farsi**, Magnus Karlsson, and Erik Agrell, "Pilot Distributions for Phase Noise Estimation in Electro-Optic Frequency Comb Systems," in *Proc. European Conference on Optical Communications (ECOC)*, Glasgow, Scotland, Oct. 2023, Paper We.A.7.1.

Other publications by the author, not included in this thesis, are:

[F] **Mohammad Farsi**, Christian Häger, Magnus Karlsson, and Erik Agrell, "Improved Polarization Tracking in the Presence of PDL," in *Proc. European Conference on Optical Communication (ECOC)*, Basel, Switzerland, Sept. 2022, Paper We4C.3.

[G] Aditya S. Rajasekaran, Monireh Vameghestahbanati, **Mohammad Farsi**, Hamid Saeedi, and Halim Yanikomeroglu, "Resource allocation-based PAPR analysis in uplink SCMA-OFDM systems," *IEEE Access*, vol. 7, pp. 162803–162817, 2019.

Acknowledgments

This thesis marks the most rewarding four years of my life, and I owe its completion to the continuous support of many individuals whom I deeply appreciate.

First and foremost, I want to express my sincere appreciation to my main supervisor, *Prof. Erik Agrell*, for his unwavering support and guidance throughout these years. *Erik*, it has been a pleasure collaborating with you, and your approach to research has left a lasting impression on me. You have always been patient and supportive of my over-ambitious deep research dives. I am also thankful to *Prof. Magnus Karlsson* for his guidance in the field of optical communications, which has significantly enriched my understanding of this domain. *Magnus*, I appreciate your unwavering positivity and your efforts to promote my work within the community.

Special thanks to *Prof. Alex Alvarado* for generously sharing his knowledge during my visit to Eindhoven, and to *Prof. Christian Häger* for introducing me to machine learning and its applications to fiber sensing. *Christian*, working with you was a delight because of your unconditional support, and knowing you were only a Slack message away!

Thanks to the seniors and students in the FORCE group for their insights during our project meetings. My sincere appreciation goes to all my colleagues, past and present, in the COMSYS group. I am grateful to the seniors for fostering a friendly environment. Thank you, *Alex*, for all the good discussions and hanging out, and also for having me as your teaching assistant. Special thanks to *Fredrik* and *Erik* for perfectly managing more than 50 nerds in the group. *Javad*, you've been not only a colleague but also a lifelong friend, and I've cherished every conversation we've had since our bachelor days. *Morteza*, our spirited discussions may make people think we are enemies, but they are proof of our friendship. *Moein*, our shared journey as Ph.D. students has been filled with enjoyable discussions, games, and travels. Your positivity has been contagious, and I appreciate your companionship. To *Jinxiang* and *Deekshith*, I have cherished our time as officemates and colleagues. And to all my other colleagues, thank you for making this journey enjoyable. Special thanks to *Nima, Shen, Ekaterina*, *Navid*, *Mohammadhosein*, *Mohammad Nazari*, *Ehsan*, *Hoang*, *Mehdi*, *Karl*, *Roman*, and *José* for your acquaintance.

Furthermore, I extend my thanks to my Iranian friends within and outside of Chalmers for their companionship over the past four years. Especially *Shokofeh, Moein, Zahra, Mostafa, Maryam, Mohammadreza, Parisa, Kamran, Samin, Hadi*, and *Narges*.

I acknowledge the unwavering support of my parents and brothers, to whom I owe a debt of gratitude for their sacrifices. Finally, my heartfelt thanks go to my loving wife, *Neda*. None of this would have been possible without you. You are my longest-lasting friend and my biggest supporter. Thank you for everything!

Mohammad Farsi
Göteborg, May, 2024.

Financial support

This work was funded by the Knut and Alice Wallenberg Foundation, grant No. 2018.0090. I would also like to acknowledge the Ericsson Research Foundation and The Solveig and Karl G. Eliasson Memorial Foundation for funding my research travels. Moreover, the computations were enabled by resources provided by Chalmers e-Commons.

Acronyms

Φ-OTDR	phase-sensitive optical time-domain reflectometry
AIR	achievable information rate
ASE	amplified spontaneous emission
AWGN	additive white Gaussian noise
BPS	blind phase search
CD	chromatic dispersion
CMA	constant modulus algorithm
CW	continuous wave
DAC	digital to analog convertor
DAS	distributed acoustic sensing
DD	decision-directed
DDLMS	decision-directed least mean squares
DGD	differential group delay
DOFS	distributed optical fiber sensing
DP	dual-polarization
DSP	digital signal processing
EDFA	erbium-doped fiber-amplifier
EKF	extended Kalman filter
EKS	extended Kalman smoother
EO-comb	electro-optic frequency comb
GD	gradient descent
GMI	generalized mutual information
i.i.d.	identically and independently distributed
IQ	in-phase-quadrature
ISA	inverse scattering algorithm
JSAC	joint sensing and communication
LMS	least mean square
LO	local oscillator
LS	least-square error
MCT	multi-core transmission
MI	mutual information
MIMO	multiple-input multiple-output
MMA	multi-modulus algorithm
MMSE	minimum mean square error
MMT	multi-mode transmission
MSE	mean square error
MZM	Mach-Zehnder modulator
NLS	nonlinear Schrödinger
pdf	probability density function
PDL	polarization-dependent loss
PMD	polarization-mode dispersion
QAM	quadrature amplitude modulation
QPSK	quadrature phase-shift keying
RDE	radially directed equalizer
RF	radio frequency
SDM	space-division multiplexing
SER	symbol error rate
SNR	signal-to-noise ratio
SOP	state of polarization
SSAC	separate sensing and communication
SW-Kabsch	sliding window Kabsch
SW-LS	sliding window least squares
WDM	wavelength-division multiplexing

Contents

Part I

Overview

CHAPTER 1

Background

Today global connectivity is no longer a dream; it is a reality that came true through modern, robust, and efficient communication systems. The majority of the global internet traffic is carried by fiber optic networks thanks to their exceptional bandwidth capacity and low signal attenuation characteristics. The compact and lightweight design of optical fibers differentiates them from traditional copper cables, making them the preferred choice for long-distance and high-speed data transmissions. In addition, optical fibers are resilient to electromagnetic and radio wave interference and lightning strikes.

Optical fibers, thin strands made of glass or plastic, were invented in the late 20th century for communications purposes. The invention of the laser [1], and the low-loss fiber [2] made commercial fiber-optic communication possible. Subsequently, the invention of the erbium-doped fiber-amplifier (EDFA) [3] extended the reach to longer distances, and coherent detection accompanied by the use of digital signal processing (DSP) further increased the spectral efficiency of fiber-based communication systems. The invention of wavelength-division multiplexing (WDM) systems [4] enabled the use of different light wavelengths to simultaneously transmit multiple data streams over a single optical fiber, significantly increasing the network's capacity. However, internet traffic shows an exponential growth rate due to several factors such as ever-increasing connected devices, bandwidth-intensive applications like streaming video and cloud computing, and the advanced 5G and beyond networks. Consequently, even with recent technological advances in optical communication systems, it is predicted that we are rapidly approaching the so-called capacity crisis [5], [6] where current optical communication systems will reach their capacity limit.

The space-division multiplexing (SDM) technique led to achieving higher data rates

by extending the data-carrying dimensions into the spatial domain. There are two general approaches in SDM [5] namely multi-mode transmission (MMT) and multi-core transmission (MCT). In MMT, the information is encoded onto different modes (spatial eigenstates) of a multimode fiber. This technique can potentially handle up to hundreds of strongly coupled spatial channels, which requires powerful DSP to decouple the channels [7]. On the other hand, in MCT the throughput is increased by transmitting the data through a multicore fiber. The MCT technique requires much simpler DSP than MMT, as the cores operate almost independently and can be processed individually. However, the scalability of this approach is currently limited, as the maximum number of cores that can be packed within a single fiber cladding while ensuring manageability is currently below 50.

In addition to SDM, utilization of the phase and polarization of light has further improved the performance of optical systems. Information is conveyed in two orthogonal polarizations of light. These systems are known as dual-polarization (DP) [8] and can double the throughput compared to a single-polarization transmission. However, the improved spectral efficiency comes at the cost of reduced tolerance to impairments, such as additive noise, nonlinearities, laser phase noise, or polarization drift. Therefore, future higher-order modulation formats require more powerful DSP to mitigate these impairments accurately.

Optical fibers are resilient to electromagnetic interference and highly sensitive to external environmental factors (e.g., temperature, strain, pressure, vibration, etc.), making them favorable for use as sensors. Fiber sensing finds widespread applications across industries, including structural health monitoring [9], environmental monitoring [10], and medical diagnostics. Additionally, fiber sensing technology plays a critical role in seismic monitoring, oil and gas exploration, and infrastructure integrity assessment, further highlighting its versatility and importance in diverse fields [11]. Recently, the phase and polarization of the telecom signal have been utilized for environmental sensing purposes [12]–[14]. Interestingly, the phase and polarization information is already available in modern coherent transceivers and therefore can be used for sensing at no additional cost and without requiring expensive instruments.

In this thesis, we study the impact of phase and polarization on optical communication systems, focusing on analyzing their influence on information capacity and designing DSP techniques. In addition, we examine how polarization information can be utilized in distributed sensing applications.

1.1 Thesis Outline

This thesis consists of two parts: the first part provides background information for the second part, which contains the included publications. The first part is organized as follows: Chapter 2 offers an overview of modern optical communication systems and common signal impairments during transmission. Chapter 3 defines the channel capac-

ity and introduces tools for bounding the channel capacity. Chapter 4 introduces the channel models used in Papers A-E. Chapter 5 discusses DSP tools to compensate for transmission impairments such as phase noise and polarization drift, and techniques for channel estimation tracking. Chapter 6 briefly explores fiber-based environmental sensing and state-of-the-art techniques for communication and sensing applications. Finally, Chapter 7 summarizes the appended publications and suggests future research directions.

1.2 Notation

Throughout Part I, we use the following notational conventions. Column vectors are denoted by underlined letters $\underline{x} = (x_0, \ldots, x_{M-1})$, and the M-dimensional vector of ones is denoted by $\mathbf{1}_M = (1, 1, \ldots, 1)$. Matrices are denoted by uppercase Roman letters X, and the M-dimensional identity matrix is denoted by I_M. Bold-face letters $\mathbf{x}$ are used for random quantities and their corresponding nonbold counterparts x for their realizations. An M-tuple or a column vector of $(x_{m+1}, \ldots, x_{m+M})$ is denoted by $\{x_i\}_{i=m+1}^{m+M}$ or $\{x_i\}_{m+1}^{m+N}$ whenever it is clear from the context. Random processes are considered ordered sequences and indicated inside braces, i.e., $\{\mathbf{x}_k\} = \{\mathbf{x}_k\}_1^\infty$ is a random process.

The log refers to the base-2 logarithm. The argument (phase) of a complex value is denoted by $\angle\underline{x} \in [-\pi, \pi)$. The $\mathrm{wrap}_\pi(\theta)$ function wraps θ into $[-\pi, \pi)$ and is defined as $\mathrm{wrap}_\pi(\theta) = \mod_{2\pi}(\theta + \pi) - \pi$. Moreover, $\theta \stackrel{\pi}{+} \phi$ denotes $\mathrm{wrap}_\pi(\theta + \phi)$. We denote Hadamard's (component-wise) product by $\circ$. Absolute value is denoted by $|\cdot|$, Euclidean norm by $\|\cdot\|$, and Frobenius norm by $\|\cdot\|_\mathrm{F}$. Whenever a scalar function is applied to a vector, e.g., $|\underline{x}|$, $\angle\underline{x}$, $\max(\underline{x})$, etc., it stands for applying the function to each element of the vector. Whenever inequalities are applied to a vector, e.g., $\underline{x} > c$, it stands for applying the inequalities to each element of the vector, i.e., $\underline{x} > c \iff x_i > c,\ \forall i$.

Probability density functions (pdfs) are denoted by $f_\mathbf{x}(x)$ and conditional pdfs by $f_{\mathbf{y}|\mathbf{x}}(y|x)$, where arguments or subscripts may sometimes be omitted if clear from the context. The expectation over random variables is denoted by $\mathbb{E}[\cdot]$. Sets and distributions are indicated by uppercase calligraphic letters, e.g., $\mathcal{X}$. The cardinality of the set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. The uniform distribution on the range $[a, b)$ is denoted by $\mathcal{U}[a, b)$. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. We denote the standard zero-mean complex circularly symmetric Gaussian distribution for a scalar by $\mathcal{CN}(0, 1)$ and for an M-dimensional vector by $\mathcal{CN}(\underline{0}, \mathrm{I}_M)$.

Notational Inconsistencies

We warn the reader regarding the following notational inconsistencies between the thesis overview given in Part I and the appended papers in Part II:

- The input vector is denoted by $\underline{\mathbf{x}}_k$ throughout the entire thesis, except in Paper A and Paper B where it is used as the channel output vector.

- The variable M denotes the channel dimensions across the entire thesis, except in Paper B where it represents the number of constellation points.
- The variable ν denotes the exponent parameter in the generalized mutual information formula throughout the thesis, but in Paper B, it denotes the sliding stride of the equalization window.
- Constellation sets are denoted by $\mathcal{C}$ in Part I, whereas in Papers A and B, they are denoted by $\mathcal{S}$, and in Papers D, by $\mathcal{X}$.
- The $(\dagger)$ sign denotes the conjugate transpose operator in this thesis, except in Paper E where it denotes the pseudoinverse.
- The variable γ represents the nonlinearity parameter in Part I, while in Paper D, it signifies the truncation parameter for truncated probability distributions.
- Throughout this thesis, $p(\cdot)$ denotes the probability mass function, except in paper B where it denotes the probability density function.
- The variable μ denotes the step size in adaptive equalizers throughout the thesis, except in Paper D where it denotes the scaling parameter of gamma distributions.
- The polarization-dependent loss matrix $\mathbf{\Gamma}_n$ is defined differently in the thesis overview and in Paper C.
- In Part I, $g(z)$ denotes the amplifier gain function, whereas in Paper D, it denotes a function consisting of differential entropy terms.

CHAPTER 2

Fiber-Optic Communication Systems

Optical communication is the core of modern telecommunication systems, offering reliable and efficient data transmission over long distances. Unlike conventional electrical transmission methods, which are limited by factors such as signal attenuation and electromagnetic interference, optical communication takes advantage of the unique properties of light to achieve high data transfer rates and reliability.

The optical transmission works by guiding the light beam through an optical fiber. The optical fibers have a core made of silica surrounded by a cladding, where the core has a higher refractive index than the cladding to ensure that light signals remain confined within the core [15, Ch. 2]. Optical fibers come in various types to meet different transmission requirements. Single-mode fibers (SMFs), designed to carry a single mode of light, are the preferred choice for long-haul transmission due to their lower signal attenuation. Few-mode and multimode fibers because of their larger core diameter support multiple modes of light, making them suitable for scenarios where increased data capacity and spatial multiplexing are required. In the beginning, the few-mode and multimode fibers were used for short-haul applications. In the last decade, however, tremendous progress has been reported in terms of transmission reach, bandwidth, and the number of fiber modes used for transmission [16].

Although fibers are low-loss mediums, signal attenuation can still occur over long transmission distances. To mitigate this attenuation, optical amplifiers are inserted along the fiber optic link at regular intervals. These amplifiers effectively recover the optical signal strength, resulting in reliable data transmission over longer distances. However, the amplification process distorts the signal by adding white Gaussian noise, commonly known as amplified spontaneous emission (ASE) noise. Besides the additive noise of

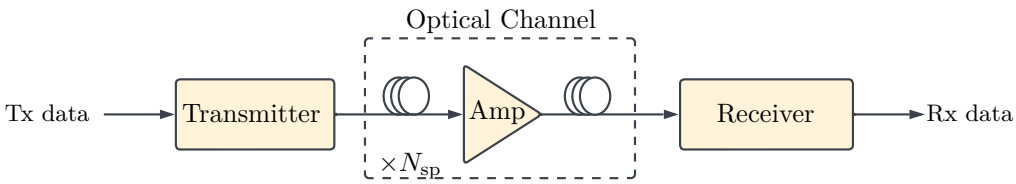

Figure 2.1: A typical fiber-optic long-distance link includes a transmitter, N_{sp} spans of optical fiber with amplification, and a coherent receiver.

the amplifiers, transmission can be affected by stochastic and deterministic impairments, which become more destructive as the link length increases.

This thesis focuses mainly on long-haul transmission, where the channel conditions can significantly impact the data transmission performance. In this chapter, we provide an overview of various transmission impairments in fiber-optic communication without delving into mathematical details. We also explore the deterministic and random effects that influence signal propagation in fiber-optic channels by studying the signal propagation equations.

In Fig. 2.1, a typical fiber-optic link is depicted. The transmitter converts the input bits into an optical waveform, which then travels through the fiber and gets amplified to reach the receiver. At the receiver, an estimate of the transmitted bits is generated based on the received optical waveform. The following sections describe the underlying elements of this system and their effect on the received signal.

2.1 The Optical Transmitter

Fig. 2.2 shows a typical scheme of a coherent optical transmitter for a single wavelength and a single polarization. The light source serves as the carrier for transmitting the information and is typically a laser or light-emitting diode. Then, the light enters the in-phase-quadrature (IQ) modulator where the data signal gets encoded into one polarization of the light. The data signals are created by converting information data into waveforms using the digital to analog convertor (DAC) module. Data transmission over both polarizations is enabled by utilizing a polarization splitter to divide the light into two orthogonal polarizations. Each polarization is independently modulated using IQs, and subsequently, the two modulated polarizations are combined using a polarization combiner.

2.1.1 The laser

The laser is a key component in coherent communication systems as it functions both as a carrier in the transmitter and as a local oscillator (LO) in the receiver. Commonly, diode lasers due to their efficiency and reliability are favored to be used at both the trans-

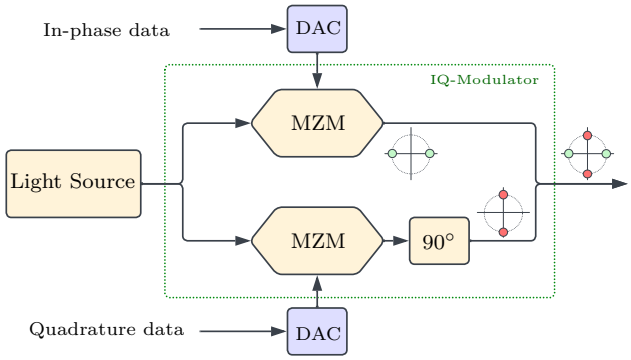

Figure 2.2: An optical transmitter consists of a laser, digital-to-analog converters, and an in-phase-quadrature modulator made of parallel Mach-Zehnder modulators.

mitter and the receiver. The lasers' efficiency relies on several factors such as wavelength stability, launch power, and linewidth.

The laser's launched power determines the amplifier (or repeater) spacing. Increasing the launch power extends the spacing between the amplifiers, saving costs by reducing the number of required amplifiers. However, the nonlinear effects of the fiber limit the extent to which the launch power can be raised. Typically, launched power is denoted in dBm units, with 1 mW acting as the reference level.

The laser linewidth determines the spread of optical frequencies around the laser's nominal output frequency, which directly impacts the phase noise of the laser. Thus, narrower linewidth implies more stable frequency emission and larger linewidth indicates higher frequency deviation. The impact of the laser phase noise on the performance of optical transmission systems will be discussed later in Section 2.3.4.

2.1.2 The Modulator

The IQ modulators are commonly used to modulate the data on the optical field. A Mach-Zehnder modulator (MZM) is a device commonly used for modulating optical signals, and it consists of two arms with a phase modulator in each arm. The MZM can control the interference between the two arms by modulating the phase of the light in each arm. To perform IQ modulation, parallel MZMs are used to independently control the in-phase and quadrature components of the optical signal [17]. Fig. 2.2 depicts an MZM-based IQ modulator.

2.1.3 Modulation Formats

In telecommunication systems, a modulation format, commonly known as constellation, refers to a method describing how bits are encoded onto a carrier wave. Modulation for-

mats can be designed to modify certain aspects of the carrier wave, such as its amplitude, frequency, or phase, to encode the desired information into the carrier wave.

A single-polarization constellation is typically a set $\mathcal{C} = \{c_1, c_2, \ldots, c_{|\mathcal{C}|}\}$ consisting of complex, zero-mean elements, with $|\mathcal{C}|$ denoting the number of constellation points. Generally, having more points in $\mathcal{C}$ means better spectral efficiency because each point in the constellation represents more bits. However, this improvement comes with a downside, as the received signal becomes more sensitive to distortions after transmission.

In modern fiber-optic communications, common modulation formats include phase-shift keying, and DP quadrature amplitude modulation (QAM). Here, DP means the same constellation is used for both polarizations of the optical signal. Hence, a DP constellation is referred to the set $\mathcal{C}^2 = \{\underline{c}_1, \ldots, \underline{c}_{|\mathcal{C}^2|}\}$ with its elements being complex zero-mean 2-dimensional vectors.

In more complex systems, such as those involving multiple dimensions, constellation points extend beyond single or dual polarizations. In an M-dimensional constellation, denoted as $\mathcal{C}^M = \{\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_{|\mathcal{C}^M|}\}$, each constellation point is represented as an M-dimensional vector. These vectors are complex, and zero-mean. The cardinality of $\mathcal{C}^M$ corresponds to the number of constellation points in the M-dimensional space.

2.2 Signal Propagation in Fibers

In this section, we briefly explain the underlying mechanism of signal propagation in the fiber.

2.2.1 The Nonlinear Schrödinger Equation

The behavior of an optical signal can be characterized by the well-known Maxwell's equations. These equations describe how electric and magnetic fields interact and propagate through space and provide a theoretical framework for understanding the behavior of optical signals in optical fiber communication systems.

The nonlinear Schrödinger (NLS) equation is a partial differential equation describing the slow-varying envelope of the single-polarized electromagnetic field of light as [15, Ch. 2]

$$\frac{\partial \mathbf{E}}{\partial z} + \underbrace{\frac{\alpha}{2}\mathbf{E}}_{\text{Attenuation}} + j\underbrace{\frac{\beta_2}{2}\frac{\partial^2 \mathbf{E}}{\partial t^2}}_{\text{Dispersion}} - j\underbrace{\gamma \mathbf{E}\,|\mathbf{E}|^2}_{\text{Nonlinearity}} = \mathbf{0}. \tag{2.1}$$

Here, $\mathbf{E}$ represent the complex baseband signal propagating in the x or y polarizations at time t and location $0 \leq z \leq L_{\mathrm{f}}$, with L_{f} denoting the total length of the fiber. Additionally, α represents the attenuation parameter, β_2 denotes the group-velocity dispersion coefficient, and γ denotes the Kerr nonlinearity coefficient. Moreover, t represents the retarded time relative to the center of the propagating signal, rather than absolute time.

Note that (2.1) corresponds to the unamplified signal propagation in an SMF and

does not account for additive noise. The validity of the NLS equation is confirmed through numerous experiments, making it a reliable model for light propagation in optical fibers. However, its accuracy depends on several critical assumptions, including single-polarization transmission, narrow spectral width relative to the central frequency, specific wavelength ranges for transmission, low fiber loss, spatially invariant refractive indices, instantaneous nonlinear response, and weak higher-order nonlinearities and dispersion. Therefore, the equation may not be accurate under extreme conditions of very high powers or wavelengths and does not consider intrinsic fiber impairments. Further details on these assumptions can be found in [18, Ch. 2].

2.2.2 The Manakov–PMD Equation

The NLS equation (2.1) applies to perfectly circular optical fibers under ideal conditions, assuming no birefringence. However, real-world optical fibers may exhibit birefringence that varies throughout the link. The birefringence causes different polarizations of light to travel at different speeds, a phenomenon commonly known as polarization-mode dispersion (PMD)[1].

Neglecting the nonlinear PMD terms which are usually negligible in communication applications, the PMD-Manakov equation takes PMD into account and is given by [19, Eq. (68)]

$$\frac{\partial \underline{\mathbf{E}}}{\partial z} + \frac{\alpha}{2}\underline{\mathbf{E}} - \Delta\beta_1 \frac{\partial \underline{\mathbf{E}}}{\partial t} + j\frac{\beta_2}{2}\frac{\partial^2 \underline{\mathbf{E}}}{\partial t^2} - j\gamma\frac{8}{9}\underline{\mathbf{E}}\,\|\underline{\mathbf{E}}\|^2 = \underline{\mathbf{0}}. \tag{2.2}$$

Here, $\underline{\mathbf{E}} = (\mathbf{E}_{\mathrm{x}}(z,t), \mathbf{E}_{\mathrm{y}}(z,t))$, where $\mathbf{E}_{\mathrm{x}}(z,t)$ and $\mathbf{E}_{\mathrm{y}}(z,t)$ represent the complex baseband signals propagating in x and y polarizations. Moreover, the third term on the left-hand side of (2.2) accounts for PMD and in contrast to the NLS equation (2.1), the nonlinear term incorporates a factor of 8/9 [20]. The term $\Delta\beta_1 = (\beta_{1,\mathrm{x}} - \beta_{1,\mathrm{y}})$ represents the group delay between polarizations x and y, characterized by propagation constants $\beta_{1,\mathrm{x}}$ and $\beta_{1,\mathrm{y}}$, respectively. The PMD coefficient of the fiber is defined as $D_{\mathrm{PMD}} = 2\sqrt{2L_{\mathrm{c}}}\Delta\beta_1$, where L_{c} denotes the correlation length of the fiber and is typically within the range of 10 to 100 meters. The correlation length L_{c} determines the length over which the two polarizations stay correlated. For modern fibers, PMD coefficient values typically range between 0.04 and 0.1 $\mathrm{ps}/\sqrt{\mathrm{km}}$, with potential increases to several $\mathrm{ps}/\sqrt{\mathrm{km}}$ for older fiber generations.

The PMD-Manakov equation (2.2) describes signal propagation without any additive noise (no amplification). The statistical PMD-Manakov equation can be obtained by a straightforward modification: substituting the zero vector on the right-hand side of (2.2) with a circularly symmetric and Gaussian vector $\underline{\mathbf{n}}$.

While PMD has been mitigated to varying degrees in modern optical fiber designs, it can still be present in commercial systems utilizing standard SMF, dispersion-compensating fiber, non-zero dispersion shifted fiber, polarization-maintaining fiber, and high-birefringence

[1] The PMD will be presented in more detail in Section 2.3.6.

Table 2.1: Typical amplifier parameters

Parameter	Symbol	Value
Amplifier Gain	G	$e^{\alpha L_{\text{sp}}}$
Carrier frequency	v	$191.1 - 196.1$ THz
Number of spans	N_{sp}	≥ 1
Planck's constant	h	$6.626 \cdot 10^{-34}$ Js
Span length	L_{sp}	$80 - 100$ km
Spontaneous emission factor	n_{sp}	$1 - 2.5$

fiber. For instance, as we will see in Paper C, the presence of PMD could be useful for fiber sensing applications.

2.3 Transmission Impairments

This section provides an overview of the main transmission impairments caused by either hardware imperfections or fiber's physical properties. While this thesis is mainly focused on the phase and polarization effects, it is useful to consider other sources of impairments in fiber-optic systems such as dispersion and nonlinearities for completeness.

2.3.1 Attenuation and Additive Noise

The standard SMFs exhibit a wavelength-dependent transmission loss with a minimum around 0.2 dB/km at approximately 1550 nm. This loss mostly originates from the Rayleigh scattering phenomenon that exists in all fibers [15, Sec. 2.5.3]. Unless compensated, this loss limits the transmission reach in long-haul fibers by making the signal undetectable to the receiver.

We can investigate the effect of attenuation on the signal by disregarding dispersion and nonlinearities, i.e., setting $\beta_2 = 0$ and $\gamma = 0$. This enables us to solve the NLS equation (2.1) as

$$\mathbf{E}(z,t) = \mathbf{E}(0,t)\exp(-\alpha z/2), \tag{2.3}$$

showing that the signal decays exponentially along the fiber due to the fiber loss.

Early solutions involved installing transceivers in multiple locations to detect and repeat the signal. However, this technique was expensive and incompatible with elastic

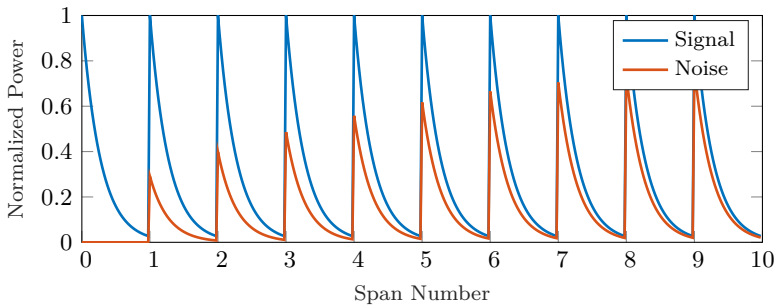

Figure 2.3: A qualitative illustration of the amplification effect on signal and added noise as a function of the number of spans.

optical networks since the repeaters were configured for a fixed location, wavelength, baud rate, modulation format, code, and other transmission parameters. The invention of the EDFA [3] and the Raman amplifier [21] enabled the transmission over very long distances by simultaneously amplifying the signal in the optical domain. While the amplification mitigates the fiber loss, it also generates additive ASE noise, which degrades signal-to-noise ratio (SNR) and limits the achievable transmission reach [22].

By considering optical amplification, the signal propagation through the fiber can be described by the statistical NLS equation as

$$\frac{\partial \mathbf{E}}{\partial z} + \frac{\alpha - g(z)}{2}\mathbf{E} + j\frac{\beta_2}{2}\frac{\partial^2 \mathbf{E}}{\partial t^2} - j\gamma \mathbf{E}\left|\mathbf{E}\right|^2 = \mathbf{n}, \tag{2.4}$$

where $\mathbf{n}$ models the additive ASE noise which is circularly symmetric and Gaussian, and $g(z)$ is the gain profile. The amount of additive noise and the gain profile depend on the amplification method. There are mainly two methods:

- **Lumped Amplification:** It is done by dividing the fiber in N_{sp} spans and inserting an EDFA at the end of each span. This leads to

$$g(z) = \alpha L_{\text{sp}} \sum_{n=1}^{N_{\text{sp}}} \delta(z - nL_{\text{sp}}),$$

  where $\delta(\cdot)$ is the Dirac delta function. Fig. 2.3 shows a qualitative description of the Lumped amplification effect on signal and added noise as a function of the number of spans. With lumped amplification in place, the autocorrelation function of the additive noise can be written as

$$\mathbb{E}[\mathbf{n}(z,t)\mathbf{n}(z',t')] = n_{\text{sp}}(G-1)h\nu\delta(t-t')\delta(z-z')\sum_{n=1}^{N_{\text{sp}}}\delta(z - nL_{\text{sp}}), \tag{2.5}$$

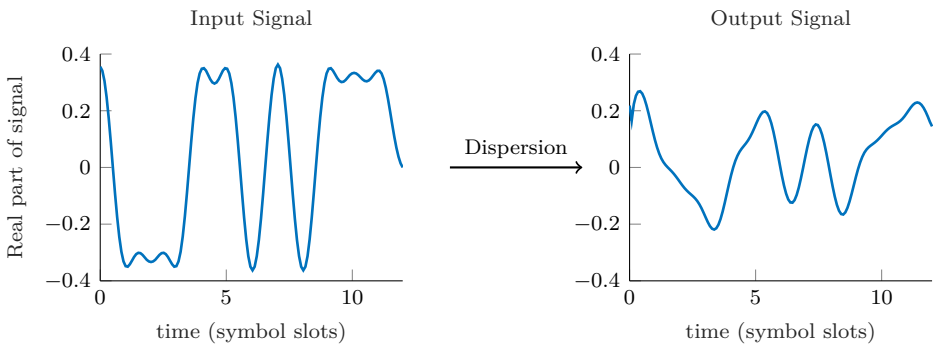

Figure 2.4: The impact of dispersion on a signal after propagating the signal.

where the parameters are defined in Table 2.1.

- **Distributed Amplification:** In contrast to lumped amplification, distributed amplification continuously compensates for transmission loss throughout propagation, maintaining a nearly constant signal power level. In this method, a pump wave is transmitted at frequencies higher than those of the signal. The pump wave co-propagates with the signal, transferring a portion of its energy to the signal through the Raman effect. Ideally, Raman amplification leads to $g(z) = \alpha$ and

$$\mathbb{E}[\mathbf{n}(z,t)\mathbf{n}(z',t')] = \alpha K_{\mathrm{T}} h v \delta(t-t')\delta(z-z'), \tag{2.6}$$

  where K_{T} is the photon occupancy factor, and it is approximately 1.13 for Raman amplification.

  The ideal Raman amplification, where each pump photon efficiently converts into a corresponding signal photon without any losses or noise, is not achievable due to several factors. The Raman amplification gain is wavelength-dependent, causing efficiency degradation outside a specific bandwidth. The stimulated Raman scattering phenomenon sets a threshold on the pump power, beyond which some pumps are lost due to scattering processes instead of being converted into signal power. Moreover, high pump power also invokes nonlinear effects in the fiber, such as self-phase modulation and four-wave mixing, which degrade amplification efficiency by introducing crosstalk [23, Ch. 3.3.1]. Despite their limitations, modern Raman amplification systems still offer substantial gains and noise reduction, which are vital for long-distance optical communication.

2.3.2 Chromatic Dispersion

The optical fiber is a dispersive medium where the group velocity is wavelength dependent which causes different wavelengths of light to travel at different speeds, a phenomenon

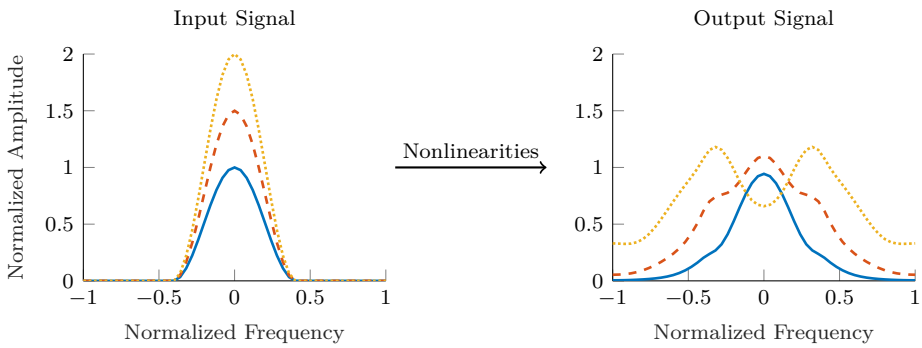

Figure 2.5: The spectrum-broadening impact of nonlinearities is shown for three pulses with different amplitudes in the frequency domain after propagation.

commonly known as chromatic dispersion (CD) or group velocity dispersion. In the time domain, the CD leads to pulse broadening since different spectral components of the light arrive at different times. This spreading effect can degrade the quality of the transmitted signal, leading to intersymbol interference and ultimately reducing the transmission distance and data rate of optical communication systems.

If we neglect the nonlinearity, we can investigate the effect of CD. The frequency domain solution of (2.1) would be

$$\mathbf{E}(z, f) = \mathbf{E}(0, f) \exp(j2\beta_2\pi^2 f^2 z) \exp(-\alpha z/2), \tag{2.7}$$

where $\mathbf{E}(z, f)$ represents the Fourier transform of $\mathbf{E}(z, t)$. Thus, the CD behaves as an all-pass filter, altering only the phase of the spectrum while leaving the amplitude unaffected. Fig. 2.4 shows that the dispersion completely distorts the signal after propagation.

2.3.3 Kerr Nonlinearity

Kerr nonlinearity serves as one of the key distinguishing factors between optical communication channels and linear wireless channels. Kerr nonlinearity in fibers originates from the fact that the intensity of the electromagnetic field impacts the core and the cladding refractive indices. The Kerr nonlinearity causes effects such as self-phase modulation, cross-phase modulation, and four-wave mixing which play an important role in the performance of the long-haul transmission systems.

Neglecting the CD and the fiber loss by setting $\beta_2 = 0$ and $\alpha = 0$, we can solve (2.1) as follows

$$\mathbf{E}(z, t) = \mathbf{E}(0, t) \exp(j\gamma \left|\mathbf{E}(0, t)\right|^2 z). \tag{2.8}$$

Thus, the Kerr nonlinear effects introduce an intensity-dependent phase shift while keeping the signal amplitude unaltered in the time domain. Consequently, the signal spectrum

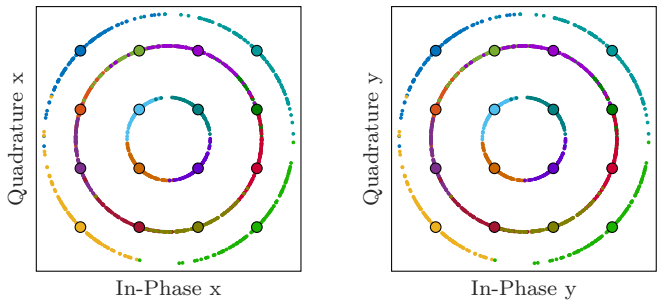

Figure 2.6: Phase noise impact on 16-DP-QAM constellations. Colors represent the received vectors corresponding to a certain transmitted point. Larger dots denote the transmitted constellation points and the smaller dots denote their corresponding received vectors. Both x and y polarizations experience the same phase noise.

broadens as a function of the signal intensity. Fig. 2.5 shows an example of the nonlinear effects in the frequency domain. It can be seen that the spectrum of the signal broadens as the input power increases.

2.3.4 Carrier Frequency Offset and Phase Noise

Coherent optical communication systems achieve high spectral efficiency by modulating the phase and amplitude of the optical field in both polarizations. To access phase and amplitude, the incoming modulated wave is down-converted to the baseband using a continuous wave (CW) optical field. The required CW can be sourced from a separate laser at the receiver, transmitted as a pilot tone orthogonal to the data, or on a separate frequency or spatial channel. The preferred method involves obtaining the intended CW from a laser at the receiver, known as the LO, to maintain spectral efficiency [24]. Ideally, the receiver LO must match the transmitter carrier frequency, but in reality, the carrier and the LO are not phase-locked which results in frequency and phase mismatch. While the phase offset rotates the transmitted constellation points with a constant phase, the frequency offset introduces a linear phase rotation to the phase of the received symbols.

As mentioned earlier, the linewidth $\Delta\nu$ of a laser refers to the spectral width of its optical output, indicating the range of frequencies over which its optical power is distributed. A perfect laser produces a perfect sinusoidal carrier wave which means its optical spectrum is a delta function, i.e., $\Delta\nu = 0$. In reality, however, the spectrum of a laser is broader which comes from the phase fluctuations of the optical field known as laser phase noise. Phase noise in a laser represents instability in the optical phase of its output, arising from factors such as spontaneous emission, cavity length fluctuations, thermal effects, and frequency-to-phase conversion processes [25, Ch. 7.6]. The relationship between linewidth and phase noise is direct: a laser with higher phase noise

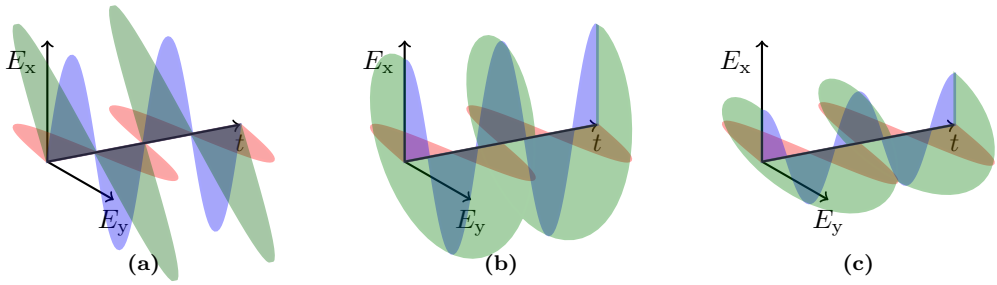

Figure 2.7: (a) Linear polarization when $|E_x| = |E_y|$ and the relative phase is zero. (b) Circular polarization with $|E_x| = |E_y|$ and relative phase $\pi/2$. (c) Elliptical polarization with $|E_x| \neq |E_y|$ and relative phase $\pi/2$.

typically exhibits a broader linewidth, while a laser with lower phase noise tends to have a narrower linewidth.

At the constellation level, the phase noise results in random rotations of the received constellations. Fig. 2.6 shows the effect of phase noise on 16-DP-QAM transmission. In Sections 4.5 and 5.4, we delve more into the concept of phase noise, exploring its modeling and various compensation techniques.

2.3.5 State of Polarization Drift

The state of polarization (SOP) of light describes the orientation of the electric field vector as light propagates through the fiber. Light is an electromagnetic wave, and its electric field oscillates perpendicularly to the direction of propagation. The state of polarization specifies the direction of this oscillation in relation to a chosen reference axis.

The electric field E can be analyzed by breaking it down into two perpendicular components denoted as E_x and E_y. Depending on their respective amplitudes and relative phase differences, various types of polarization states can emerge. These include linear polarization, wherein the electric field oscillates along a fixed direction, circular polarization, where the electric field traces a circular path, and elliptical polarization, which entails an elliptical trajectory for the electric field vector. Fig. 2.7 shows the three aforementioned types of polarization states.

The E_x and E_y components, representing the electric fields of the x and y polarizations, can be individually modulated in both phase and amplitude. In an ideal scenario, light traveling through a straight fiber would maintain the same SOP in which it was initially launched. However, in reality, fiber optic cables possess asymmetrical cores as a result of manufacturing processes, mechanical/thermal stresses encountered during and after installation, and environmental changes. These imperfections disrupt the orthogonality between the E_x and E_y components, leading to a phenomenon known as SOP drift. Because of the stochastic fluctuations in environmental conditions affecting the fiber,

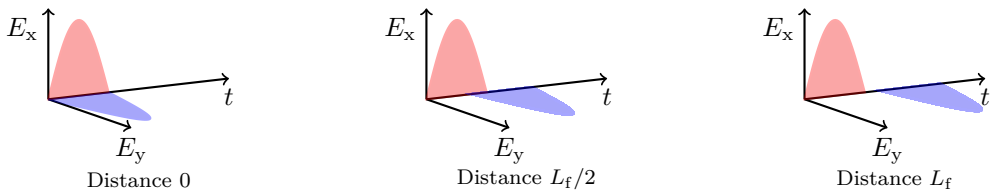

Figure 2.8: Illustration of the PMD effect on x and y polarizations. Two pulses, each having equal power in the x and y polarizations, propagate through a short fiber of length L_f. The pulses are depicted at three different propagation distances, where the time separation between them increases as the distance grows.

the geometry of fibers undergoes random variations both longitudinally and temporally, causing the SOP drift to be a random process. The drift in the SOP leads to a time-variant power exchange between the two polarizations, disrupting their orthogonality and consequently limiting the overall performance of the system.

Experimental measurements indicate that the drift in the SOP can exhibit vastly different rates depending on the type of fiber and its environmental conditions. For buried fibers, the SOP drift may occur over extended periods, spanning hours or even days [26]. In contrast, aerial fibers experience much faster SOP drift, with timescales measured in microseconds [27]–[29]. For example, field observations of an aerial fiber link revealed that the SOP drift rate could reach up to 5.1 Mrad/s during lightning strikes [30]. In Sections 4.2, 4.3, and 5.3 we will delve deeper into the modeling of the drift in the SOP, and explore tracking schemes with a focus on their ability to tolerate rapid drifts.

2.3.6 Polarization-Mode Dispersion

Due to the imperfections in the fiber's geometry (asymmetry, stress, bending), the refractive index differs for different polarizations, a phenomenon known as birefringence. The birefringence causes different polarization states of light to travel at different speeds [18, Ch. 2.2], which is known as PMD.

Fig. 2.8 depicts the PMD effect on the x and y polarizations of the signal. As the light travels through the fiber, the two polarizations get separated in time due to the different propagation speeds. The amount of this time separation is commonly known as differential group delay (DGD). In practical terms, PMD limits the transmission capacity and quality of optical communication systems by causing distortion and spreading of optical signals. It becomes particularly significant in high-speed optical transmission systems where pulses are packed closely in time.

As the impact of PMD on the signal differs based on polarization, and as mentioned previously polarization drifts over time, PMD becomes a time-varying impairment. To mitigate PMD, various techniques are employed, including using fibers with low PMD

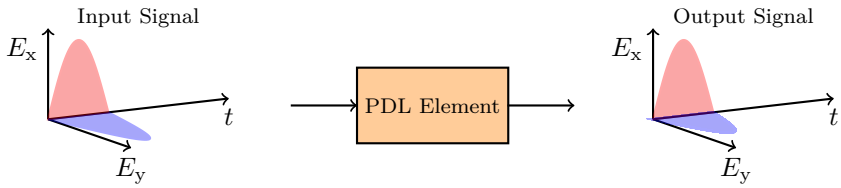

Figure 2.9: Qualitative impact of a PDL element resulting in a power imbalance between the two different polarizations.

characteristics, employing PMD compensators, or using DSP techniques to mitigate its effects. From a DSP perspective, PMD induces frequency-dependent crosstalk between polarizations, which can be mitigated by adaptive equalizers.

2.3.7 Polarization-Dependent Loss

Polarization-dependent loss (PDL) refers to the phenomenon where the loss of transmitted light through an optical system, such as a fiber optic network or a photonic device, varies depending on the polarization state of the light. In other words, the loss experienced by light traveling through the system differs based on whether its polarization aligns with the preferred axis of the system or not. PDL can occur due to asymmetries in the optical components, such as isolators, multiplexers, and couplers [31]. Fig. 2.9 depicts the effect of PDL on two different polarizations. In this particular example, the signal in x-polarization remained unchanged, while the signal in y polarization experienced attenuation. Note that the amount of power imbalance and its orientation depends on the preferred axis of the PDL element.

High levels of PDL can degrade the performance of optical communication systems by inducing power imbalances and disrupting the orthogonality between polarizations [32]. Consequently, mitigating the PDL effects is important for maintaining the reliability and efficiency of optical networks and devices. Additionally, when random SOP drift is combined with PDL, the resulting power imbalances fluctuate over time, which makes the polarization tracking more challenging.

In the literature, PDL has typically been characterized by concatenating many randomly oriented PDL elements along the fiber, with its statistical properties extensively examined in [33]–[35], which describe the aggregated PDL using a Maxwellian distribution. In [36] and [37] the effects of average PDL on the optical link's performance are investigated. Furthermore, the interaction between PDL and other phenomena such as PMD and Kerr nonlinearity has been investigated in [38] and [39], respectively. In Sections 4.4 and 5.3, we provide more details into the modeling and compensation methods of PDL.

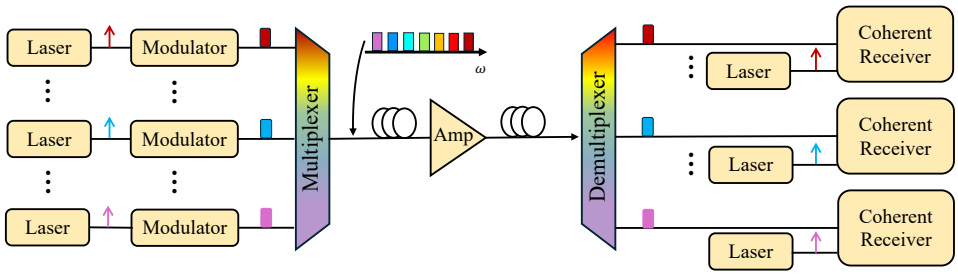

Figure 2.10: A standard WDM system, each channel comprises an optical transceiver that operates autonomously, separate from all other co-propagating channels. Each WDM channel requires two lasers, one at the transmitter side and another at the receiver end.

2.4 Coherent Wavelength Division Multiplexing Systems

The coherent WDM systems through utilizing multiple wavelengths enable simultaneous transmission of data over a single optical fiber. Thus, WDM systems effectively increase the throughput by expanding transmission bandwidth and enhancing efficiency. Neglecting all the impairments and the nonlinearities and by only considering additive white Gaussian noise (AWGN), the capacity of a bandlimited channel with power constraint ρ and noise variance σ^2 is denoted by the formula expressed in bits per second, $W \log(1 + \rho/\sigma^2)$, where W represents bandwidth [40, Ch. 9.3]. Considering the capacity formula, one can grasp why WDM efficiently enhances throughput. Notably, capacity increases linearly with bandwidth and logarithmically with power ρ. Therefore, doubling throughput requires increasing the SNR from ρ/σ^2 to ρ^2/σ^4. Alternatively, this increase in throughput can be achieved by doubling bandwidth and only doubling power. Similarly, modulating both polarizations enables doubling the throughput by merely doubling the transmitted power.

Today, state-of-the-art experiments may employ hundreds of wavelengths within a single fiber core [41], requiring large stacks of lasers in transmitters. Furthermore, the growing popularity of coherent intradyne links [42], [43], which require LO lasers and DSP in receivers, highlights the importance of coherent sources not only in transmitters but also in receivers.

Fig. 2.10 shows the principle of a WDM system. For each WDM channel, a pair of transmitters and receivers independent of all the other channels are used. The fiber and the amplifier are shared between all the channels resulting in cost and power consumption reduction.

2.4.1 Electro-Optic Frequency Combs

An electro-optic frequency comb (EO-comb) is a device that generates a spectrum of equally spaced optical frequencies, typically spanning a broad range, using electro-optic

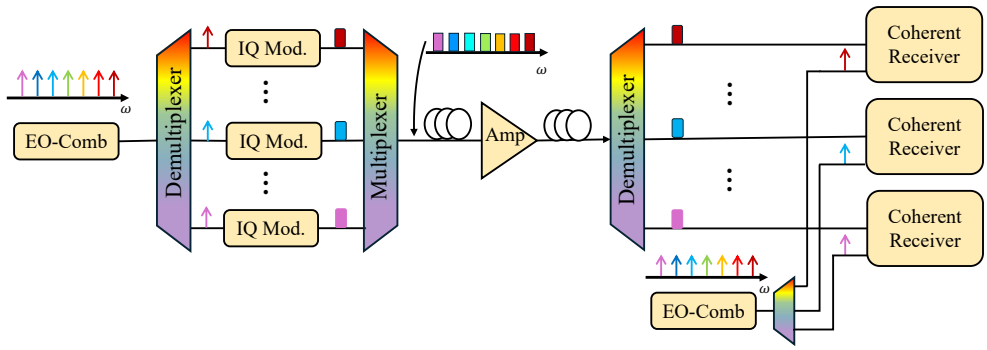

Figure 2.11: A comb-based WDM system: All WDM channels are fed by a common comb source at the transmitter, while the local oscillator lines at the receiver side are generated from receiver combs. Modulation and receiver processing are typically conducted independently for each wavelength channel.

modulation techniques. This technology relies on the nonlinear optical effects induced by applying an electrical signal to an electro-optic modulator, which modulates the phase or intensity of an optical carrier wave. The resulting optical spectrum consists of a series of closely spaced frequency components, similar to the teeth of a comb, hence the name "frequency comb" [44]. EO-combs have applications in various fields such as spectroscopy, telecommunications, and frequency metrology, where their precise and stable optical frequencies are highly beneficial.

2.4.2 Frequency-Comb Based WDM System

As illustrated in Fig. 2.10, each WDM channel typically requires a dedicated laser, resulting in a large stack of lasers. However, this bulky setup could potentially be replaced by an EO-comb both at the transmitter and the receiver, leading to cost reduction and more efficient packaging, given that lasers (e.g., external cavity lasers and tunable lasers) are known for their high expense and bulkiness. Utilizing a single light source results in a strong phase correlation among the comb lines. This correlation can be leveraged to enhance phase-noise tolerance [45] or simplify DSP required to compensate the phase noise at the receiver [46].

There are also challenges with introducing combs to replace the stack of lasers in WDM systems. The most important one is the limited line power inherent in a multi-wavelength source, which poses optical SNR limitations on comb-based transmitters. The insufficient optical SNR reduces the transmission reach, especially for higher-order modulation formats. Various comb technologies have shown potential to replace free-running lasers in WDM systems. Examples include Kerr effect-based micro-ring resonators [47], mode-locked lasers [48], and parametric mixers [44]. However, these technologies are beyond the scope of this thesis.

The principle of a WDM system employing frequency combs is depicted in Fig. 2.11 where the stack of lasers is replaced with EO-comb at both ends. In Section 4.5 we provide a channel model for the phase noise of such systems. Moreover, in Section 5.4 we discuss the phase noise compensation techniques which are more suitable to EO-comb-based systems.

CHAPTER 3

Channel Capacity

This chapter introduces the concept of channel capacity. Differential entropy and mutual information are defined in Section 3.1. Section 3.2 provides the definition of the capacity of discrete-time channels. Section 3.3 provides some information-theory tools to bound the channel capacity. Finally, Section 3.4 defines the achievable information rate (AIR) and provides a detailed description of how to compute it for some auxiliary channels. All the definitions are given in vector form, as the scalar expressions are special cases of vector expressions.

3.1 Entropy, Relative Entropy, and Mutual Information

Definition 3.1 (Entropy)**.** *Consider an arbitrary vector, M-dimensional vector $\underline{\mathbf{x}}$ taking values from the discrete alphabet $\mathcal{X}^M$ and distributed according to the probability mass function $p_{\underline{\mathbf{x}}}(\underline{x})$. The entropy of $\underline{\mathbf{x}}$ is defined as*

$$H(\underline{\mathbf{x}}) = -\sum_{\underline{x}\in\mathcal{X}^M} p_{\underline{\mathbf{x}}}(\underline{x}) \log\left(p_{\underline{\mathbf{x}}}(\underline{x})\right). \tag{3.1}$$

The entropy quantifies the information content of a random variable. Moreover, it is nonnegative $H(\underline{\mathbf{x}}) \geq 0$ and when $\mathcal{X}^M$ is finite, we have $H(\underline{\mathbf{x}}) \leq \log|\mathcal{X}^M|$.

Definition 3.2 (Differential entropy)**.** *Consider two arbitrary vectors, $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$, with values in $\mathbb{R}^M$, and distributed according to the joint pdf $f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x},\underline{y})$. The differential en-*

tropy of $\underline{\mathbf{x}}$ *is defined as [40, Sec. 8]*

$$h(\underline{\mathbf{x}}) = -\int_{\mathbb{R}^M} f_{\underline{\mathbf{x}}}(\underline{x}) \log\left(f_{\underline{\mathbf{x}}}(\underline{x})\right) \mathrm{d}\underline{x}. \tag{3.2}$$

Similarly, the joint differential entropy of $\underline{\mathbf{x}}$ *and* $\underline{\mathbf{y}}$ *is*

$$h(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = -\int_{\mathbb{R}^M}\int_{\mathbb{R}^M} f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x}, \underline{y}) \log\left(f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x}, \underline{y})\right) \mathrm{d}\underline{x}\mathrm{d}\underline{y}. \tag{3.3}$$

The differential entropy is a measure of the information content of a random quantity and is defined over continuous alphabets. Unlike the entropy, the differential entropy can be either positive or negative with $h(\underline{\mathbf{x}}) = -\infty$ corresponding to a deterministic vector.

Definition 3.3 (Conditional entropy)**.** *Suppose* $\underline{\mathbf{x}}$ *and* $\underline{\mathbf{y}}$ *are* M*-dimensional real random vectors. The conditional entropy of* $\underline{\mathbf{x}}$ *given* $\underline{\mathbf{y}}$ *is defined as*

$$h(\underline{\mathbf{x}}|\underline{\mathbf{y}}) = -\int_{\mathbb{R}^M}\int_{\mathbb{R}^M} f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x}, \underline{y}) \log\left(\frac{f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x}, \underline{y})}{f_{\underline{\mathbf{y}}(\underline{y})}}\right) \mathrm{d}\underline{x}\mathrm{d}\underline{y}. \tag{3.4}$$

It is a fact that conditioning decreases entropy, expressed as $h(\underline{\mathbf{x}} \,|\, \underline{\mathbf{y}}) \leq h(\underline{\mathbf{x}})$, with equality only when $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ are independent. Moreover, the chain rule can be proved from (3.2)–(3.4) as

$$\begin{aligned} h(\underline{\mathbf{x}}, \underline{\mathbf{y}}) &= h(\underline{\mathbf{x}}) + h(\underline{\mathbf{y}} \,|\, \underline{\mathbf{x}}) \\ &= h(\underline{\mathbf{y}}) + h(\underline{\mathbf{x}} \,|\, \underline{\mathbf{y}}). \end{aligned} \tag{3.5}$$

Definition 3.4 (Mutual information)**.** *The mutual information (MI) between random vectors* $\underline{\mathbf{x}}$ *and* $\underline{\mathbf{y}}$ *can be defined as*

$$\begin{aligned} I(\underline{\mathbf{x}}; \underline{\mathbf{y}}) &= h(\underline{\mathbf{y}}) - h(\underline{\mathbf{y}} \,|\, \underline{\mathbf{x}}) \\ &= h(\underline{\mathbf{x}}) - h(\underline{\mathbf{x}} \,|\, \underline{\mathbf{y}}). \end{aligned} \tag{3.6}$$

MI measures the amount of information shared between two random variables, indicating the degree to which knowledge about one variable reduces uncertainty about the other. An important feature of the MI is that no clever manipulation of $\underline{\mathbf{y}}$ or $\underline{\mathbf{x}}$ can increase the MI. This property is known as the data processing inequality and states that if $\underline{\mathbf{x}}$–$\underline{\mathbf{y}}$–$\underline{\mathbf{z}}$ form a Markov chain[2], then $I(\underline{\mathbf{x}}; \underline{\mathbf{z}}) \leq \min\{I(\underline{\mathbf{x}}; \underline{\mathbf{y}}), I(\underline{\mathbf{y}}; \underline{\mathbf{z}})\}$.

Definition 3.5 (Relative entropy)**.** *The relative entropy between two pdfs* $f_{\underline{\mathbf{x}}}(\underline{x})$ *and*

[2] In a Markov chain, denoted as $\underline{\mathbf{x}}$–$\underline{\mathbf{y}}$–$\underline{\mathbf{z}}$, the pair $(\underline{\mathbf{z}}, \underline{\mathbf{x}})$ are independent when $\underline{\mathbf{y}}$ is given, that is $f_{\underline{\mathbf{z}}|\underline{\mathbf{x}},\underline{\mathbf{y}}} = f_{\underline{\mathbf{z}}|\underline{\mathbf{y}}}$.

$q_{\underline{\mathbf{x}}}(\underline{x})$ *is a nonnegative function and defined as*

$$D_{\mathrm{KL}}(f_{\underline{\mathbf{x}}}(\underline{x}) \parallel q_{\underline{\mathbf{x}}}(\underline{x})) = \int f_{\underline{\mathbf{x}}}(\underline{x}) \log\left(\frac{f_{\underline{\mathbf{x}}}(\underline{x})}{q_{\underline{\mathbf{x}}}(\underline{x})}\right) \mathrm{d}\underline{x}. \tag{3.7}$$

Another way to represent the MI is to define it in terms of relative entropy also known as Kullback–Leibler divergence as [40, Thm. 8.6.1]

$$I(\underline{\mathbf{x}};\underline{\mathbf{y}}) = D_{\mathrm{KL}}\left(f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x},\underline{y}) \parallel f_{\underline{\mathbf{x}}}(\underline{x}) f_{\underline{\mathbf{y}}}(\underline{y})\right). \tag{3.8}$$

Lastly, considering a complex vector $\underline{\mathbf{x}}$ with values in $\mathbb{C}^M$, we define its differential entropy as

$$h(\underline{\mathbf{x}}) = h(\underline{\mathbf{x}}_{\mathrm{r}}, \underline{\mathbf{x}}_{\mathrm{i}}), \tag{3.9}$$

where $\underline{\mathbf{x}}_{\mathrm{r}}$ and $\underline{\mathbf{x}}_{\mathrm{i}}$ represent the real and imaginary components of the complex random vector $\underline{\mathbf{x}}$. Analogously, all previously mentioned metrics—conditional entropy, joint entropy, MI, and relative entropy—can be similarly defined for complex vectors.

3.2 Channel Capacity

Channel capacity refers to the maximum rate at which information can be reliably transmitted through a communication channel, considering noise and limitations inherent to the channel. The concept, often expressed in bits per second, represents the upper limit of data transfer that can be achieved with minimal error. Channel capacity theory, pioneered by Claude Shannon [49], provides a framework for optimizing communication systems to approach this theoretical limit.

Definition 3.6 (Capacity of a memoryless channel)**.** *The channel capacity for a memoryless channel with input* $\underline{\mathbf{x}}$ *and output* $\underline{\mathbf{y}}$ *can be defined as [49]*

$$C(\rho) = \sup I\big(\underline{\mathbf{x}};\underline{\mathbf{y}}\big), \tag{3.10}$$

where the supremum is over all probability distributions on $\underline{\mathbf{x}}$ *that satisfy the average-power constraint* $\mathbb{E}[\|\underline{\mathbf{x}}\|^2] \leq \rho$.

The following definition outlines the capacity of a channel with memory that maintains information stability. Loosely speaking, in an information-stable channel the input that maximizes MI and its corresponding channel output behave ergodically [50].

Definition 3.7 (Capacity of a channel with memory)**.** *Consider an information-stable discrete-time complex channel with memory with random input vector* $\underline{\mathbf{x}}_k$ *and random output vector* $\underline{\mathbf{y}}_k$ *where* k *is the time index. Then, the capacity under an average power*

constraint $\rho > 0$ is given by [50]

$$C(\rho) = \lim_{K\to\infty} \sup \frac{1}{K} I\big(\{\underline{\mathbf{x}}_k\}_1^K; \{\underline{\mathbf{y}}_k\}_1^K\big), \tag{3.11}$$

where the supremum is over all probability distributions on $\{\underline{\mathbf{x}}_k\}_1^K$ that satisfy the average-power constraint

$$\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left[\|\underline{\mathbf{x}}_k\|^2\right] \le \rho. \tag{3.12}$$

3.3 Information-Theoretic Tools for Bounding Capacity

This section presents conventional information-theoretic tools for establishing upper and lower bounds on capacity.

3.3.1 Maximum Entropy

For any real random vector $\underline{\mathbf{x}} \in \mathbb{R}^M$ with a fixed nonsingular correlation matrix $\mathrm{R}_{\underline{\mathbf{x}}} = \mathbb{E}[\underline{\mathbf{x}}\underline{\mathbf{x}}^T]$, it is known that the joint Gaussian distribution achieves the maximum differential entropy [40, Thm. 8.6.5]

$$h(\underline{\mathbf{x}}) \le \frac{1}{2}\log\left((2\pi e)^M \det \mathrm{R}_{\underline{\mathbf{x}}}\right). \tag{3.13}$$

3.3.2 Entropy in Polar Coordinates

It is useful to define the differential entropy of a complex random variable in polar coordinates as [51, Lemma 6.16]

$$\begin{aligned} h(\mathbf{x}) &= h(|\mathbf{x}|, \angle\underline{\mathbf{x}}) + \mathbb{E}[\log|\mathbf{x}|] \\ &= h(|\mathbf{x}|) + h(\angle\underline{\mathbf{x}}\,|\,|\mathbf{x}|) + \mathbb{E}[\log|\mathbf{x}|], \end{aligned} \tag{3.14}$$

where $\angle\underline{\mathbf{x}}$ is the phase of $\mathbf{x}$. Moreover, the entropy of a squared magnitude of a random variable can be written as [51, Lemma 6.15]

$$h(|\mathbf{x}|^2) = h(|\mathbf{x}|) + \mathbb{E}[\log|\mathbf{x}|] + \log 2. \tag{3.15}$$

Combining (3.14) and (3.15), we can write

$$h(\mathbf{x}) = h(|\mathbf{x}|^2) + h(\angle\underline{\mathbf{x}}\,|\,|\mathbf{x}|) - \log 2. \tag{3.16}$$

In case $\mathbf{x}$ is circularly symmetric, (3.16) can be simplified to

$$h(\mathbf{x}) = h(|\mathbf{x}|^2) + \log\pi, \tag{3.17}$$

which follows from that the amplitude $|\mathbf{x}|$ and the phase $\angle\underline{\mathbf{x}}$ are independent and $\angle\underline{\mathbf{x}} \sim \mathcal{U}[-\pi, \pi)$.

3.3.3 Duality Bound

The duality bound provides an upper bound on the MI. It is particularly useful when the true output distribution is not analytically well-behaved or hard to compute.

Theorem 3.1 (Duality bound)**.** *The MI between $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ with joint pdf $f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x}, \underline{y})$ can be upper bounded as [51]–[53]*

$$\begin{aligned} I(\underline{\mathbf{x}}; \underline{\mathbf{y}}) &\leq \mathbb{E}\left[\log\left(\frac{f_{\underline{\mathbf{y}}\,|\,\underline{\mathbf{x}}}(\underline{y}\,|\,\underline{x})}{q_{\underline{\mathbf{y}}}(\underline{y})}\right)\right] \\ &= -\mathbb{E}[\log q_{\underline{\mathbf{y}}}(\underline{y})] - h(\underline{\mathbf{y}}|\underline{\mathbf{x}}), \end{aligned} \tag{3.18}$$

where $q_{\underline{\mathbf{y}}}(\underline{y})$ is any arbitrary distribution on $\underline{\mathbf{y}}$ and the bound is tight for $q_{\underline{\mathbf{y}}}(\underline{y}) = f_{\underline{\mathbf{y}}}(\underline{y})$. Moreover, the expectation is with respect to the marginal distribution $f_{\underline{\mathbf{y}}}(\underline{y})$.

By employing the duality bound in conjunction with the Lagrange multiplier method, we can derive upper bounds on the capacity of a memoryless channel subject to the average power constraint ρ, expressed as

$$\begin{aligned} C(\rho) &= \sup_{f_{\underline{\mathbf{x}}}} I(\underline{\mathbf{x}}; \underline{\mathbf{y}}) \\ &\leq \sup_{f_{\underline{\mathbf{x}}}} -\mathbb{E}[\log q_{\underline{\mathbf{y}}}(\underline{y})] - h(\underline{\mathbf{y}}\,|\,\underline{\mathbf{x}}) \\ &\leq \sup_{f_{\underline{\mathbf{x}}}} -\mathbb{E}[\log q_{\underline{\mathbf{y}}}(\underline{y})] - h(\underline{\mathbf{y}}\,|\,\underline{\mathbf{x}}) + \lambda(\rho - \mathbb{E}[\|\underline{\mathbf{x}}\|^2]) \\ &= \sup_{f_{\underline{\mathbf{x}}}} \mathbb{E}_{\underline{\mathbf{x}}}\left[-\mathbb{E}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}[\log q_{\underline{\mathbf{y}}}(\underline{y})] - h(\underline{\mathbf{y}}\,|\,\underline{\mathbf{x}}) + \lambda(\rho - \|\underline{\mathbf{x}}\|^2)\right] \\ &\overset{\text{(a)}}{\leq} \sup_{\underline{x}}\left[-\mathbb{E}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}=\underline{x}}[\log q_{\underline{\mathbf{y}}}(\underline{y})] - h(\underline{\mathbf{y}}\,|\,\underline{\mathbf{x}} = \underline{x}) + \lambda(\rho - \|\underline{x}\|^2)\right], \end{aligned} \tag{3.19}$$

where $\lambda \geq 0$ is the Lagrange multiplier and (a) follows as the expectation is replaced with deterministic $\underline{x}$ that maximizes inside the expectation subject to $\|\underline{x}\|^2 \leq \rho$. Therefore, an upper bound on the capacity can be determined by choosing any distribution $q_{\underline{\mathbf{y}}}$ on $\underline{\mathbf{y}}$ and any $\lambda \geq 0$. To enhance the tightness of the upper bound, the parameters of $q_{\underline{\mathbf{y}}}$ and λ can be optimized jointly. We utilized (3.19) in Paper D to upper-bound the MI of a memoryless channel by selecting the auxiliary distribution $q_{\underline{\mathbf{y}}}$ to be circularly symmetric with the squared magnitude of $\underline{\mathbf{y}}$ following a gamma distribution. Interested readers are referred to Paper D.

In Paper D, we employed (3.19) to establish an upper bound on the MI of a memoryless

channel. This was achieved by selecting the auxiliary distribution $q_{\underline{\mathbf{y}}}$ to be circular symmetry, with the squared magnitude of $\underline{\mathbf{y}}$ following a gamma distribution. For further details, interested readers are encouraged to refer to Paper D.

3.4 Mismatched Decoding

An AIR is the rate at which reliable communication is possible over a specific channel, considering a given decoding rule that may not be optimal. This situation arises organically when imprecise channel measurements lead to the receiver employing maximum-likelihood decoding based on an incorrect channel law. Alternatively, intentional design choices may lead the receiver to adopt a suboptimal decoding rule for the sake of simplifying implementation [54].

A common way to compute mismatched AIR between the input $\underline{\mathbf{x}}$ and the output $\underline{\mathbf{y}}$ with joint distribution of $f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x},\underline{y})$ is to evaluate

$$I_q(\underline{\mathbf{x}};\underline{\mathbf{y}}) = \int f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x},\underline{y}) \log\left(\frac{q_{\underline{\mathbf{y}}\,|\,\underline{\mathbf{x}}}(\underline{y}\,|\,\underline{x})}{q_{\underline{\mathbf{y}}(\underline{y})}}\right) \mathrm{d}\underline{x}\mathrm{d}\underline{y}, \tag{3.20}$$

for suitable choices of the auxiliary channel law $q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}\,|\,\underline{x})$ and the input distribution $f_{\underline{\mathbf{x}}}(\underline{x})$. Note that $I_q(\underline{\mathbf{x}};\underline{\mathbf{y}}) \leq I(\underline{\mathbf{x}};\underline{\mathbf{y}})$ with equality when $q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x}) = f_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})$.

In its simplest form, the problem involves a memoryless channel with finite input and output alphabets, using a single-symbol decoding rule where each symbol is decoded independently based solely on the information available from that particular symbol. Despite its simplicity, the mismatch capacity—defined as the highest AIR—is not known for many channels. Addressing this problem could provide a solution to the longstanding challenge of computing the zero-error capacity[3] of a channel, as demonstrated in [55].

3.4.1 LM Rate

The pioneering studies on mismatched decoding, notably by Hui [56] and Csiszár and Körner [57], employed constant-composition random coding. They derived an AIR for mismatched discrete memoryless channelss (DMCs), commonly referred to as the LM rate. The LM stands for lower bound on mismatched decoding MI.

Definition 3.8. *The LM rate can be used as a lower bound on the MI* (3.6) *between complex vectors* $\underline{\mathbf{x}}$ *and* $\underline{\mathbf{y}}$ *with joint pdf of* $f_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x},\underline{y})$ *as following [54, Eq. (11)]*

$$I(\underline{\mathbf{x}};\underline{\mathbf{y}}) \geq I_{\mathrm{LM}}(\underline{\mathbf{x}},\underline{\mathbf{y}}) = \sup_{\nu\geq 0, a(\cdot)} \mathbb{E}_{\underline{\mathbf{x}},\underline{\mathbf{y}}}\left[\log \frac{\tilde{q}(\underline{\mathbf{x}},\underline{\mathbf{y}})^{\nu} e^{-a(\underline{\mathbf{x}})}}{\mathbb{E}_{\underline{\mathbf{x}}'}\left[\tilde{q}(\underline{\mathbf{x}}',\underline{\mathbf{y}})^{\nu} e^{-a(\underline{\mathbf{x}})}\right]}\right], \tag{3.21}$$

[3] The zero-error capacity is defined as the maximum rate at which information can be reliably communicated over a noisy channel with zero probability of error.

where $\tilde{q}(\underline{x},\underline{y}) = e^{-d(\underline{x},\underline{y})}$ *is the decoding metric and* $d(\underline{x},\underline{y})$ *is any function that satisfies* $\mathbb{E}_{\underline{\mathbf{x}},\underline{\mathbf{y}}}\left[d(\underline{\mathbf{x}},\underline{\mathbf{y}})\right] \leq \infty$. *Moreover,* $a(\underline{x})$ *belongs to the class of functions that are integrable with respect to* $f_{\underline{\mathbf{x}}}(\underline{x})$ *that is* $\mathbb{E}_{\underline{\mathbf{x}}}[|a(\underline{\mathbf{x}})|] \leq \infty$.

By choosing $\tilde{q}(\underline{x},\underline{y}) = f_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})$ and letting $a(\underline{x}) = 0$, the LM rate can achieve the true MI (3.6). It is important to mention that the decoding metric is not a pdf; however, it can be chosen to be.

3.4.2 Generalized Mutual Information

The generalized mutual information (GMI) extends MI application to scenarios where the joint distribution has tails that do not decay rapidly. It is a versatile metric that finds utility in various fields such as machine learning, signal processing, and neuroscience, enabling a more nuanced understanding of the relationships within complex systems.

Definition 3.9 (GMI)**.** *The GMI is a lower bound on the MI and can be defined as a special case of LM rate by setting* $a(\underline{x}) = 0$ *[54, Eq. (12)] which gives*

$$I_{\mathrm{GMI}}(\underline{\mathbf{x}},\underline{\mathbf{y}}) = \sup_{\nu \geq 0} \mathbb{E}_{\underline{\mathbf{x}},\underline{\mathbf{y}}}\left[\log \frac{\tilde{q}(\underline{\mathbf{x}},\underline{\mathbf{y}})^{\nu}}{\mathbb{E}_{\underline{\mathbf{x}}'}\left[\tilde{q}(\underline{\mathbf{x}}',\underline{\mathbf{y}})^{\nu}\right]}\right]. \tag{3.22}$$

Here, $\tilde{q}(\underline{x},\underline{y}) = \exp(-d(\underline{x},\underline{y}))$ *is known as decoding metric and can be any function that satisfies* $\mathbb{E}_{\underline{\mathbf{x}},\underline{\mathbf{y}}}\left[d(\underline{x},\underline{y})\right] \leq \infty$.

The GMI (3.22) is bounded by the LM rate (3.21), and the latter can remain strictly higher even after optimizing the input distribution. This observation is utilized in [54] to establish the achievability of (3.21) in scenarios involving general alphabets.

The GMI could be useful to design suboptimal mismatched receivers by optimizing an auxiliary channel when the optimal receiver is absent. The choice of the decoding metric $\tilde{q}(\underline{x},\underline{y})$ gives birth to different auxiliary channels.

Example 3.1. Consider a memoryless channel $\underline{\mathbf{y}} = \mathrm{H}\underline{\mathbf{x}} + \underline{\mathbf{z}}$, where $\underline{\mathbf{z}} \sim \mathcal{CN}(\underline{0}, \sigma_z^2 \mathrm{I}_n)$. Assuming that the receiver has access only to the estimated channel matrix $\hat{\mathrm{H}}$, a natural choice for the decoding metric could be $\tilde{q}(\underline{x},\underline{y}) = q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x}) \sim \mathcal{CN}(\hat{\mathrm{H}}\underline{x}, \sigma_z^2 \mathrm{I}_n)$ that is

$$\tilde{q}(\underline{x},\underline{y}) = \frac{1}{\pi^n \sigma_z^{2n}} \exp\left(-\frac{\left\|\underline{y} - \hat{\mathrm{H}}\underline{x}\right\|^2}{\sigma_z^2}\right). \tag{3.23}$$

Substituting (3.23) into (3.22) and optimizing over ν will result in a lower bound on the MI. This has been utilized in Paper A to derive a lower bound on the MI of a unitary channel in the presence of channel estimation error.

•

CHAPTER 4

Channel Models

A channel model mathematically relates the received signal to the transmitted signal. The model may include the transmission impairments as well as the transmitter and the receiver elements. Channel models play a crucial role in simulating, testing, and verifying the transmission schemes and designing efficient DSP algorithms before experimental trials. This can significantly decrease the cost and energy required for running real-world experiments. Moreover, the optimality of the designed DSP algorithms highly depends on the accuracy of the channel model. Often, there is a trade-off between the accuracy of the model and its practicality as usually, the most accurate models are very complex to be analyzed.

Channel models are especially useful for deriving fundamental limits imposed by the transmission medium, such as channel capacity. As performing capacity analysis often involves solving very complex optimizations, the simplest nontrivial channel models are commonly used to derive AIR and gain insights regarding the optimal transmission schemes.

This chapter provides mathematical models that describe how various transmission impairments impact the electric field of the signal propagating through optical fibers. In particular, Section 4.2 describes the block-constant SOP drift channel, Section 4.3 models a channel impaired by fast SOP drift and PDL. Section 4.4.2 combines DGD with SOP scramblers along with PDL to define a channel model that is useful for polarization sensing applications. Then, changing the focus on phase noise models, Section 4.5 provides a model for the correlated phase noise observed in WDM systems employing EO-combs.

4.1 Linear Modulation, Matched Filtering, and Sampling

Linear modulation refers to techniques to modulate information to a carrier signal by linearly varying its amplitude, frequency, or phase. The dual-polarized transmitted signal $\underline{\mathbf{E}}(0,t)$ over the transmission medium can be described as

$$\underline{\mathbf{E}}(0,t) = \sum_k \underline{\mathbf{x}}_k p_\mathrm{s}(t-kT) \tag{4.1}$$

where $\underline{\mathbf{x}}_k = (\mathbf{x}_{\mathrm{x},k}, \mathbf{x}_{\mathrm{y},k}) \in \mathcal{C}^2$ are the information symbols with an average energy of $\mathbb{E}[\|\underline{\mathbf{x}}_k\|^2]$, the real-valued pulse shape is denoted by $p_\mathrm{s}(t)$, T is the symbol duration, and $k \in \mathbb{Z}$ is the discrete-time index. The pulse shape is usually chosen such that it forms an orthonormal basis for any $k, k' \in \mathbb{Z}$, i.e.,

$$\int_{-\infty}^{\infty} p_\mathrm{s}(t-kT)p_\mathrm{s}(t-k'T)\mathrm{d}t = \begin{cases} 1, & k = k'. \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

We use matched filtering and sampling techniques to extract the discrete symbols at the distance L_f from the continuous time received signal $\underline{\mathbf{E}}(L_\mathrm{f}, t)$. Here is how it works:

- **Matched filtering:** The received signal is passed through a matched filter, which has an impulse response matched to the pulse shape used at the transmitter. This matched filter enhances the desired signal while suppressing noise and interference, maximizing the SNR.
- **Sampling:** After the signal has been filtered by the matched filter, the output of the filter is sampled at regular intervals.

Thus, by matched filtering and sampling we can extract the discrete received symbols as

$$\underline{\mathbf{y}}_k = \int_{-\infty}^{\infty} \underline{\mathbf{E}}(L_\mathrm{f}, t)p_\mathrm{s}(t-kT)\mathrm{d}t. \tag{4.3}$$

Example 4.1. By neglecting all fiber transmission impairments except for ASE noise, we can model the fiber optic channel as a memoryless complex AWGN channel. The received symbols after filtering and sampling can then be expressed as

$$\underline{\mathbf{y}}_k = \underline{\mathbf{x}}_k + \underline{\mathbf{z}}_k, \tag{4.4}$$

where $\underline{\mathbf{z}}_k = (\mathbf{z}_{\mathrm{x},k}, \mathbf{z}_{\mathrm{y},k}) \sim (\underline{\mathbf{0}}, \sigma_z^2 \mathrm{I}_2)$ models the accumulated complex ASE noise with

$$\mathbf{z}_{\mathrm{x},k} = \int_{-\infty}^{\infty} \underbrace{\left(\int_0^{L_\mathrm{f}} \mathbf{n}_\mathrm{x}(z,t)\mathrm{d}z \right)}_{\text{Accumulated noise}} p_\mathrm{s}(t-kT)\mathrm{d}t \tag{4.5}$$

and similarly for $\mathbf{z}_{\mathrm{y},k}$. Here, $\mathbf{n}_{\mathrm{x}}(z,t)$ denotes the amplification noise for polarization x at time t and location z with its autocorrelation function depending on the amplification method introduced in Section 2.3.1. Referring to equations (2.5) and (2.6), for lumped amplification, the noise variance is expressed as $\sigma_z^2 = n_{\mathrm{sp}}(G-1)hvN_{\mathrm{sp}}W$, while for distributed amplification, it is denoted as $\sigma_z^2 = \alpha K_{\mathrm{T}} hv L_{\mathrm{f}} W$, with W representing the signal bandwidth. •

4.2 Block-Constant SOP Drift Model

In this section, we consider a 2-dimensional complex channel that is impaired by block-constant SOP drift and ASE noise. We assume the channel remains constant during a transmission block of length K with changes occurring randomly and independently between these blocks. This assumption arises from observations in scenarios such as buried and underwater fibers, where the SOP drifts at a considerably slower rate compared to the typical transmission rates encountered in optical links [26], [58].

The selection of the block length K depends on both the specific application and the rate of channel drift. We also assume that PMD is negligible, and all channel impairments, such as nonlinearities and CD, are ideally mitigated, except for SOP fluctuation and ASE noise. In many studies, the fluctuation of the SOP is often represented as either constant randomly selected rotations [59], [60] or deterministic cyclic/quasi-cyclic changes [61]–[63]. While widely accepted in the literature [64], [65], the assumption of independence between random blocks is somewhat idealized and may not fully reflect real-world conditions.

Here, we establish the channel model for $M \geq 2$ dimensions, with $M = 2$ representing the block-constant SOP drift channel. This is because the generalized version of the channel model has been utilized in Paper A to derive the capacity of block-constant channels in the presence of channel estimation error covering both general multiple-input multiple-output (MIMO) and SOP drift channels.

Given the independence of different blocks, we focus exclusively on modeling the symbols within a single transmission block hereafter. The transmitted signal in each channel at time $k = 0, \ldots, K-1$ is represented by an M-dimensional random complex vector $\underline{\mathbf{x}}_k$, and when $M = 2$ it indicates DP, single-mode, single-core transmission. The elements of the vector $\underline{\mathbf{x}}_k$ are chosen from a set of complex zero-mean constellation points $\mathcal{C}^M$. Following filtering and resampling of the received signals to one sample per symbol, we can represent the vector of received samples $\underline{\mathbf{y}}_k$ as

$$\underline{\mathbf{y}}_k = \mathbf{H}\underline{\mathbf{x}}_k + \underline{\mathbf{z}}_k. \tag{4.6}$$

Here, the $M \times M$ matrix $\mathbf{H}$ represents a block-constant MIMO channel, and the complex ASE noise samples at time k are denoted by $\underline{\mathbf{z}}_k \sim \mathcal{CN}(\underline{0}, \sigma_z^2 \mathrm{I}_M)$, and are independent of $\underline{\mathbf{x}}_k$. The block-constant SOP drift channel model can be obtained by setting $M = 2$ and

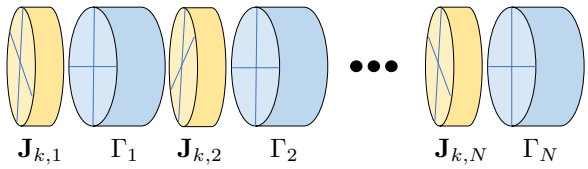

Figure 4.1: At the time index k, the physical model of a fiber can be considered as a concatenation of N sections. Each section contains a PDL component denoted as Γ_n, along with a drift element for the SOP represented as $\mathbf{J}_{k,n}$. This figure is borrowed from Paper B.

enforcing $\mathbf{H}$ to be a unitary matrix, i.e., $\mathbf{H}\mathbf{H}^\dagger = \mathbf{H}^\dagger\mathbf{H} = \mathrm{I}_2$.

4.3 PDL and SOP Drift Model

In this section, we consider DP transmission impaired by PDL, SOP drift, and ASE noise at the receiver. We assume that the PMD and nonlinearities are negligible and the CD is ideally compensated. To represent the channel, we merge the SOP drift model from [66] with concatenated PDL elements.

Following matched filtering and sampling, where each received signal corresponds to one sample per symbol, the vector of received complex samples $\underline{\mathbf{y}}_k$ can be formulated as

$$\underline{\mathbf{y}}_k = \mathbf{H}_k\underline{\mathbf{x}}_k + \underline{\mathbf{z}}_k. \tag{4.7}$$

Here, the 2-dimensional random vector $\underline{\mathbf{x}}_k$ denotes the transmitted symbols, $\mathbf{H}_k$ represents a 2×2 complex channel matrix modeling both SOP drift and PDL, and $\underline{\mathbf{z}}_k \sim \mathcal{CN}(0, \sigma_z^2\mathrm{I}_2)$ represents the ASE noise.

The physical model at time k is illustrated in Fig. 4.1, depicting a link including N sections. Each section is composed of an SOP element $\mathbf{J}_{k,n}$ and a PDL element Γ_n, where n denotes the section index. The dual-polarized channel matrix $\mathbf{H}_k$ merging PDL with SOP drift can be described as

$$\mathbf{H}_k = \Gamma_n\mathbf{J}_{k,N}\cdots\Gamma_1\mathbf{J}_{k,1} = \prod_{n=1}^{N}\Gamma_n\mathbf{J}_{k,n}, \tag{4.8}$$

where Γ_n represents a 2×2 diagonal matrix with positive real values, characterizing the polarization power imbalance resulting from PDL. In the following, we describe how the SOP and PDL elements can be modeled.

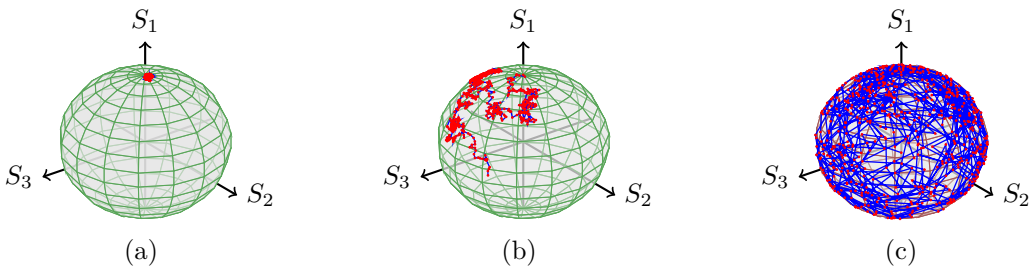

Figure 4.2: The evolution of SOP on the Poincaré sphere in a time period of 1 microsecond is depicted. The symbol rate is 28 Gbaud (i.e., the symbol duration $T = 3.57 \cdot 10^{-11}$ s) and $N = 20$. (a) $\Delta \mathrm{p}_{\mathrm{tot}} \cdot T = 10^{-8}$ s, (b) $\Delta \mathrm{p}_{\mathrm{tot}} \cdot T = 10^{-6}$, and (c) $\Delta \mathrm{p}_{\mathrm{tot}} \cdot T = 10^{-4}$ s.

4.3.1 Modeling SOP Drift

Various forms of polarization drift models have been suggested and applied in the literature. In [59], [60], a block-constant assumption is utilized. We used a similar block-constant model in Section 4.2 and consequently in Paper A to derive the capacity bounds on a unitary channel with imperfect channel estimation. However, the block-constant assumption does not hold for aerial fibers where rapid drift occurs, necessitating a dynamic and stochastic model instead of a static one.

Here, we model the SOP as a dynamic stochastic model as proposed in [66]. We consider $\mathbf{J}_{k,n}$ as a complex random 2×2 matrix that belongs to the special unitary group of degree two, denoted by $SU(2)$, that is

$$\mathbf{J}_{k,n}\mathbf{J}_{k,n}^{\dagger} = \mathbf{J}_{k,n}^{\dagger}\mathbf{J}_{k,n} = \mathrm{I}_2, \tag{4.9}$$

$$\det \mathbf{J}_{k,n} = 1. \tag{4.10}$$

Typically, complex 2×2 matrices have eight degrees of freedoms (DOFs), that is the real and imaginary components of the four elements in the matrix. However, unitary matrices belonging to the special unitary group $SU(2)$ are limited to three DOFs. Thus, we can model the SOP drift matrix $\mathbf{J}_{k,n}$ as a random walk on a unit sphere using a matrix exponential parameterized by three variables as [66]

$$\mathbf{J}_{k+1,n} = \exp(-j\underline{\boldsymbol{\alpha}}_{k,n} \cdot \overrightarrow{\sigma})\mathbf{J}_{k,n}, \tag{4.11}$$

where $\exp(\cdot)$ is the matrix exponential and

$$\underline{\boldsymbol{\alpha}}_{k,n} \sim \mathcal{N}(\underline{0}, \sigma_{\mathrm{p}}^2 \mathrm{I}_3). \tag{4.12}$$

Here, $\sigma_{\mathrm{p}}^2 = 2\pi\Delta \mathrm{p} T$, where $\Delta \mathrm{p}$ is referred to as the *polarization linewidth*, governing the rate of SOP drift, and T denotes the symbol duration. Finally, $\overrightarrow{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is a

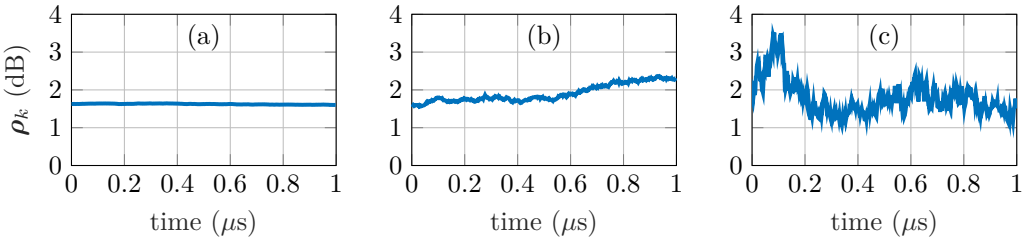

Figure 4.3: The evolution of aggregated PDL ratio $\boldsymbol{\rho}_k$ in dB for different SOP drifts in a time period of 1 microsecond is plotted. The symbol rate is 28 Gbaud (i.e., the symbol duration $T = 3.57 \cdot 10^{-11}$) and $N = 20$. (a) $\Delta \mathrm{p}_{\mathrm{tot}} \cdot T = 10^{-8}$, (b) $\Delta \mathrm{p}_{\mathrm{tot}} \cdot T = 10^{-6}$, and (c) $\Delta \mathrm{p}_{\mathrm{tot}} \cdot T = 10^{-4}$.

tensor of the Pauli spin matrices [67, eq. (2.5.19)]

$$\sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix}. \tag{4.13}$$

The total polarization linewidth scales with N and can be defined as

$$\Delta \mathrm{p}_{\mathrm{tot}} = N \cdot \Delta \mathrm{p}. \tag{4.14}$$

The initial matrices $\mathbf{J}_{0,1}, \ldots, \mathbf{J}_{0,N}$ are randomly selected from the set of all 2×2 special unitary matrices $SU(2)$. This ensures that the SOP achieves a uniform distribution over the Poincaré sphere [66]. Fig. 4.2 demonstrates the evolution of SOP on the Poincaré sphere over a one-microsecond window under three different SOP drift speeds.

4.3.2 Modeling PDL

The PDL matrices could potentially modeled as time-varying matrices, but to simplify matters, we assume that each PDL component can be adequately described by a deterministic matrix that remains constant over time and can be described as[34]

$$\Gamma_n = \begin{bmatrix} \sqrt{1+\gamma_n} & 0 \\ 0 & \sqrt{1-\gamma_n} \end{bmatrix}, \tag{4.15}$$

where $0 \leq \gamma_n \leq 1$ represents the PDL ratio for each section. In an extreme scenario where $\gamma_n = 1$, only one active polarization persists, while in the opposite extreme with $\gamma_n = 0$, both polarizations experience no PDL.

For the channel (4.8), we define the section-wise PDL in dB as

$$\varphi_n = 10 \log_{10} \left(\frac{1+\gamma_n}{1-\gamma_n} \right), \tag{4.16}$$

and the aggregated PDL ratio at time k as

$$\boldsymbol{\rho}_k = \frac{\|\boldsymbol{\lambda}_k^{\max}\|^2}{\|\boldsymbol{\lambda}_k^{\min}\|^2}, \tag{4.17}$$

where $\boldsymbol{\lambda}_k^{\max}$ and $\boldsymbol{\lambda}_k^{\min}$ are the singular values of $\mathbf{H}_k$. The average aggregated PDL for K symbol transmissions is defined in dB as

$$\bar{\rho} = 10\log_{10}\left(\mathbb{E}_{\mathbf{H}}\left[\frac{1}{K}\sum_{k=0}^{K-1}\boldsymbol{\rho}_k\right]\right). \tag{4.18}$$

Although each PDL component Γ_n is assumed to remain constant, $\boldsymbol{\rho}_k$ will vary over time due to SOP drift. Fig. 4.3 demonstrates how $\boldsymbol{\rho}_k$ is influenced by SOP drift. The graph illustrates PDL evolution over a one-microsecond window under three different SOP drift speeds. In Section 5.3, we will further explore the existing and proposed channel tracking algorithms designed to effectively manage channels affected by both SOP drift and PDL.

The model presented in this section is used in Paper B to develop algorithms that are resilient to fast SOP drift in the presence of PDL.

4.4 Concatenated PDL and DGD Model

Here, we consider a DP optical transmission system, which accounts for PDL, SOP scramblers (rotations), and DGD resulting from PMD. The SOP scrambling is considered a deterministic phenomenon, unlike the SOP drift introduced in Section 4.3, which is modeled as a random process. We also assume negligible nonlinearities and perfect compensation for CD. Specifically, we adopt a channel model similar to the one utilized in [68] which characterizes the channel response at time index k and frequency ω_i as a concatenation of N sections, each including elements of SOP scramblers, PDL, and DGD. Thus, the 2×2 complex-valued Jones matrix is defined as follows

$$\mathrm{H}(\omega_i, \underline{\Theta}_k) = \prod_{n=1}^{N} \Gamma_n \mathrm{R}\left(\phi_{n,k}, \psi_{n,k}\right) \mathrm{T}(\omega_i). \tag{4.19}$$

Here, n denotes the section index, and to avoid confusion with the matrix multiplication ordering, the product operator is defined as $\prod_{n=1}^{N} A_n = A_N A_{N-1} \cdots A_1$. The PDL elements Γ_n are defined as in (4.15) and are assumed to remain constant over time for all n. Moreover, the SOP scramblers are described as

$$\mathrm{R}(\phi_{n,k}, \psi_{n,k}) = \begin{bmatrix} \cos\phi_{n,k} & je^{j\psi_{n,k}}\sin\phi_{n,k} \\ je^{-j\psi_{n,k}}\sin\phi_{n,k} & \cos\phi_{n,k} \end{bmatrix}, \tag{4.20}$$

where $\phi_{n,k}$ and $\psi_{n,k}$ are rotation angles. Here, the rotation parameters are considered to be static when no environmental disturbance is present. Finally, the DGD elements are modeled as

$$\mathrm{T}(\omega_i) = \begin{bmatrix} e^{j\omega_i \tau/2} & 0 \\ 0 & e^{-j\omega_i \tau/2} \end{bmatrix}, \tag{4.21}$$

where τ is the DGD per section.

In general, τ can be modeled as random and possibly dynamic. For instance, in [69], the DGD elements are modeled as Gaussian random variables with a specific mean and variance. Here, the DGD elements are assumed to be deterministic and constant over time for simplicity. Finally, we define $\underline{\Theta}_k = \{\gamma_n, \phi_{n,k}, \psi_{n,k}\}_{n=1}^{N}$ to denote all channel parameters with γ_n denoting the PDL ratio parameter and $\phi_{n,k}$, $\psi_{n,k}$ denoting SOP scrambling parameters.

4.4.1 Modeling the Time-Dependency of the SOP Rotation Parameters

The channel model (4.19) is used in Paper C for sensing purposes. To enable this model for sensing, one needs to define a relation between the environmental perturbations with the time-dependency of the model parameters. As an early step, in [13], it has been demonstrated that dynamic polarization fluctuations induced by strain or pressure result in relative birefringence fluctuations. However, various disturbances can impact the fiber without a proper model describing how they impact the parameters in the mathematical model. This presents an open and undiscovered opportunity to develop such models, which could simplify fiber sensing study through simulations and enable the use of various data-driven algorithms.

Here, acknowledging the lack of a model for translating the environmental disturbances into the SOP drift, we expand upon [68] and assume that the SOP is static when the fiber is undisturbed and dynamic under disturbances. Specifically, we assume that environmental disturbances introduce temporal fluctuations in the rotation angles $\boldsymbol{\phi}_{n,k}$ and $\boldsymbol{\psi}_{n,k}$ which are as Wiener processes, characterized by

$$\boldsymbol{\phi}_{n,k} = \boldsymbol{\phi}_{n,k-1} + \boldsymbol{\Phi}_{n,k}, \tag{4.22}$$

$$\boldsymbol{\psi}_{n,k} = \boldsymbol{\psi}_{n,k-1} + \boldsymbol{\Psi}_{n,k}. \tag{4.23}$$

Here, $\boldsymbol{\Phi}_{n,k}$ and $\boldsymbol{\Psi}_{n,k}$ are normally distributed with mean zero and variances $\sigma^2_{n,k}$ and $\varrho^2_{n,k}$ respectively, representing the intensity of environmental fluctuations. Larger values of $\sigma^2_{n,k}$ and $\varrho^2_n(k)$ indicate more substantial environmental changes. Modeling the rotation parameters according to (4.23) makes (4.20) a special case of SOP drift model (4.11). Unlike the SOP drift model in (4.11), the SOP scrambler model in (4.20) does not generate a uniform distribution over the Poincaré sphere. It is important to note that the model (4.19) assumes a timescale for the dynamic polarization perturbation significantly

slower than the DGD parameters τ.

4.4.2 Modeling the Noisy Channel Estimate

In the coherent receivers, the Jones matrix (4.19) is typically estimated through adaptive equalization (see Section 5.2). Therefore, we assume that we have only access to a noisy version of the Jones matrix

$$\tilde{\mathbf{H}}(\omega_i; \underline{\Theta}_k) = \mathrm{H}(\omega_i; \underline{\Theta}_k) + \mathbf{Z}_k(\omega_i), \tag{4.24}$$

where $\mathrm{H}(\omega_i; \underline{\Theta}_k)$ is defined in (4.19) and $\mathbf{Z}_k(\omega_i)$ is a 2×2 matrix with zero-mean independent complex normal elements with variance σ_{Z}^2 modeling the estimation error, i.e.,

$$\mathbb{E}[\mathbf{Z}_k(\omega_i) \circ \mathbf{Z}_k(\omega_i)^\dagger] = \sigma_{\mathrm{Z}}^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \tag{4.25}$$

where $(\circ)$ is the element-wise product. The estimation noise variance σ_{Z}^2 depends on various factors such as the choice of estimation algorithm and the channel SNR, where better algorithms and higher SNR lead to lower σ_{Z}^2.

In Section 6.2, we elaborate on the potential applications of this model in fiber sensing and discuss the algorithms developed for this purpose.

4.5 EO-Comb Phase-Noise Channel

In this section, we initially establish the relationship between the phase and frequency of comb lines in an EO-comb. Subsequently, we present a MIMO channel model that details the input-output relationship of a WDM system utilizing EO-combs at both the transmitter and receiver sides.

4.5.1 Phase and Frequency in EO-combs

Let us consider a common configuration of an EO-comb displayed in Fig. 4.4. Consider a CW laser source oscillating at the frequency ω^{c} with the phase noise $\boldsymbol{\theta}^{\mathrm{c}}(t)$ and the amplitude of A. Thus, the electrical field of the CW laser is given by

$$E_{\mathrm{c}}(t) = A \exp j(\omega^{\mathrm{c}} t + \boldsymbol{\theta}^{\mathrm{c}}(t)). \tag{4.26}$$

This laser is then coupled with a phase modulator that is driven by a radio frequency (RF) source operating at frequency ω^{r} [46], [70]. The electrical field of the normalized RF source is given by

$$E_{\mathrm{r}}(t) = \sin(\omega^{\mathrm{r}} t + \boldsymbol{\theta}^{\mathrm{r}}(t)), \tag{4.27}$$

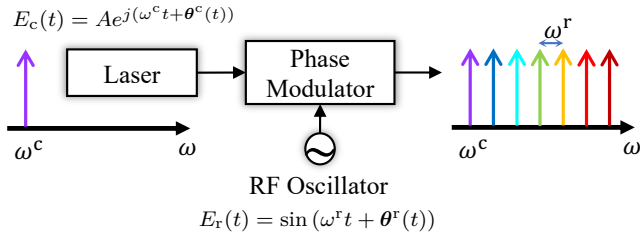

Figure 4.4: A common configuration of an EO-comb involves a CW laser operating at frequency ω^{c} with phase noise $\boldsymbol{\theta}^{\mathrm{c}}(t)$, coupled to a phase modulator driven by a radio frequency (RF) oscillator at frequency ω^{r} with phase noise $\boldsymbol{\theta}^{\mathrm{r}}(t)$. This modulation process yields the generation of M comb lines distinguished by a central frequency ω^{c} and a frequency spacing of ω^{r}. This figure is borrowed from Paper D.

where $\boldsymbol{\theta}^{\mathrm{r}}(t)$ denotes the phase noise of the RF oscillator. Consequently, the output optical field $E_{\mathrm{comb}}(t)$ after the phase modulator can be expressed as

$$E_{\mathrm{comb}}(t) = E_{\mathrm{c}}(t) \cdot e^{jV\sin(\omega^{\mathrm{r}}t+\boldsymbol{\theta}^{\mathrm{r}}(t))}, \tag{4.28}$$

where V is the modulation index of the phase modulator. By using the Jacobi-Anger expansion[4] [71, Ch. 17, Eq. (17.1.7)] we can write [72]

$$\begin{aligned} E_{\mathrm{comb}}(t) &= E_{\mathrm{c}}(t) \sum_{m=-\infty}^{\infty} J_m(V) j^n e^{jm(\omega^{\mathrm{r}}t+\boldsymbol{\theta}^{\mathrm{r}}(t))} \\ &= \sum_{m=-\infty}^{\infty} J_m(V) j^n e^{j(\omega^{\mathrm{c}}+m\omega^{\mathrm{r}})} e^{j(\boldsymbol{\theta}^{\mathrm{c}}(t)+m\boldsymbol{\theta}^{\mathrm{r}}(t))t}, \end{aligned} \tag{4.29}$$

where $J_m(V)$ denotes m-th order Bessel function. Therefore, the output field consists of discrete frequencies (comb lines)

$$\omega_m = \omega^{\mathrm{c}} + m\omega^{\mathrm{r}}, \tag{4.30}$$

and the net phase noise at each comb line is a linear combination of the CW laser and RF oscillator phase noises and can be written as [72]

$$\boldsymbol{\theta}_m(t) = \boldsymbol{\theta}^{\mathrm{c}}(t) + m\boldsymbol{\theta}^{\mathrm{r}}(t). \tag{4.31}$$

By cascading multiple phase modulators and also amplitude modulators, the comb lines can be altered to provide a relatively flat amplitude over a large bandwidth (number of lines) [70]. However, this cascading technique will not affect the phase noise, which will still be given by (4.31). For a finite modulation index V and sufficiently large m, however, the amplitudes will eventually fall off, limiting the practical EO-combs to employ $50-100$

[4] $e^{jx\sin(\theta)} = \sum_{m=-\infty}^{\infty} j^m J_m(x) e^{jm\theta}$

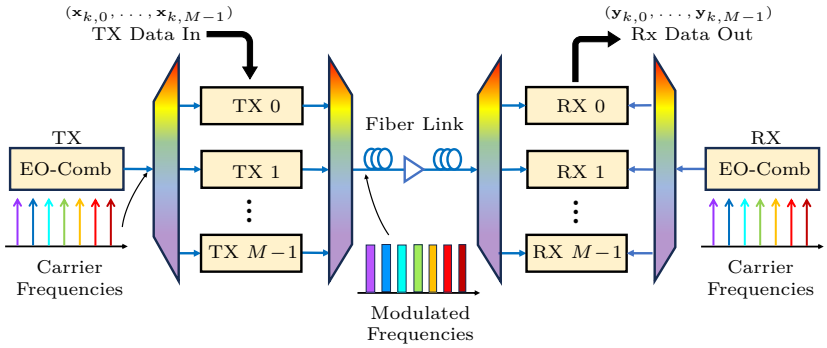

Figure 4.5: The structure of a comb-based WDM link employing free-running (uncorrelated) EO-combs at both the transmitter and receiver. This figure is borrowed from Paper D.

frequency lines.

4.5.2 Channel Model

As mentioned earlier, in Section 2.4, EO-combs can replace the transmitter and receiver laser arrays in WDM systems as illustrated in Fig. 4.5. The EO-combs employed at both sides are considered to be independent, commonly referred to as "free-running EO-combs" [73]. In such systems, the combined phase noise arises from the summation of transmitter and receiver phase noises. In practice, numerous comb lines are commonly employed which directs our attention towards scenarios with $M \geq 2$, where M represents the number of WDM channels.

We assume a single-polarization M-dimensional MIMO transmission in the presence of CW laser and RF oscillator phase noises, along with ASE noise at the receiver. Additionally, we assume negligible PMD and nonlinearities, and perfectly compensated CD. The EO-comb channel model can be expressed as

$$\underline{\mathbf{y}}_k = e^{j\underline{\boldsymbol{\theta}}_k} \circ \underline{\mathbf{x}}_k + \underline{\mathbf{z}}_k, \tag{4.32}$$

Here, k denotes the discrete time index, $\underline{\mathbf{x}}_k = \{\mathbf{x}_{k,m}\}_{m=0}^{M-1}$ represents the M-dimensional input vector, while $\underline{\boldsymbol{\theta}}_k = \{\boldsymbol{\theta}_{k,m}\}_{m=0}^{M-1}$ denotes the phase-noise process. Additionally, the additive noise $\underline{\mathbf{z}}_k = \{\mathbf{z}_{k,m}\}_{m=0}^{M-1}$ follows a complex normal distribution $\mathcal{CN}(\underline{\mathbf{0}}, \sigma_z^2 \mathrm{I}_M)$ and is independent for all k and m.

The channel (4.32) can potentially represent a variety of wireless MIMO links and optical MIMO channels. For example, when $\underline{\boldsymbol{\theta}}_k = \boldsymbol{\theta}_k \cdot \mathbf{1}_M$ and $\boldsymbol{\theta}_k$ follows a Wiener process, it describes the scenario of common phase noise as investigated in [74]. Furthermore, if we assume $\underline{\boldsymbol{\theta}}_k = (\boldsymbol{\theta}_{k,0}, \ldots, \boldsymbol{\theta}_{k,M-1})$ with independent and stationary $\boldsymbol{\theta}_{k,m}$, it corresponds to the model examined in [75, Model B3], where only receiver phase noise is considered, and the channel matrix is an identity matrix. Here, we introduce the EO-comb phase-

noise model, where two independent phase-noise sources are present and the phase noise of each channel is a linear combination of these two independent sources. As a result, the EO-comb phase-noise model fills the gap between the models proposed in [74] and [75].

The EO-comb phase-noise of channel index $m \in \{0, \dots, M-1\}$ at time k is obtained by sampling the continuous time phase noise (4.31) and can be written as [5] [72]

$$\boldsymbol{\theta}_{k,m} = \boldsymbol{\theta}_k^{\mathrm{c}} \mathbin{\overset{\pi}{+}} m\boldsymbol{\theta}_k^{\mathrm{r}}, \tag{4.33}$$

where $\boldsymbol{\theta}_k^{\mathrm{c}}$ and $\boldsymbol{\theta}_k^{\mathrm{r}}$ are the combined (transmitter and receiver) phase noises induced by the CW lasers and RF sources, respectively.

We introduce the convention c/r to prevent repeating similar equations multiple times. The phase-noise sources are modeled as

$$\begin{cases} \boldsymbol{\theta}_k^{\mathrm{c/r}} = \boldsymbol{\Delta}_k^{\mathrm{c/r}} \mathbin{\overset{\pi}{+}} \boldsymbol{\theta}_{k-1}^{\mathrm{c/r}}, & \text{if} \quad k = 1, 2, \dots \\ \boldsymbol{\theta}_k^{\mathrm{c/r}} \sim \mathcal{U}[-\pi, \pi), & \text{if} \quad k = 0, \end{cases} . \tag{4.34}$$

Here, $\boldsymbol{\Delta}_k^{\mathrm{c/r}} \sim \mathcal{WN}(0, \sigma_{\mathrm{c/r}}^2)$ with variances $\sigma_{\mathrm{c/r}}^2 = 2\pi B_{\mathrm{c/r}}/R_{\mathrm{s}}$ where R_{s} represents the symbol rate. Moreover, $B_{\mathrm{c/r}} > 0$ denote the linewidths of the CW laser and RF oscillator, respectively.

The processes $\{\boldsymbol{\theta}_k^{\mathrm{c/r}}\}$ and $\{\boldsymbol{\theta}_{k,m}\}$ are stationary due to that the initial phases are $\boldsymbol{\theta}_0^{\mathrm{c/r}} \sim \mathcal{U}[-\pi, \pi)$. Also, the identically and independently distributed (i.i.d.) assumption on $\{\boldsymbol{\Delta}_k^{\mathrm{c/r}}\}$ implies that $\{\boldsymbol{\theta}_k^{\mathrm{c/r}}\}$ are Markov processes [74] and consequently the process $\{\underline{\boldsymbol{\theta}}_k\}$ becomes a multivariate Wiener process.

In Paper D, we employ the concepts of duality and escape-to-infinity (refer to Section 3.3) to investigate the capacity of the channel (4.32)–(4.34). In Paper E, we explore the optimal pilot placement for EO-comb based WDM systems, assuming the presence of the channel (4.32)–(4.34).

[5] $\theta \mathbin{\overset{\pi}{+}} \phi$ denotes $\mathrm{wrap}_\pi(\theta + \phi)$ where $\mathrm{wrap}_\pi(\theta) = \mod_{2\pi}(\theta + \pi) - \pi$.

CHAPTER 5

Digital Signal Processing

In this chapter, we discuss estimation, compensation, and tracking techniques applicable to the channel models introduced in Chapter 4. First, static estimators that are suitable for block-constant channels are discussed in Section 5.1. Thereafter, in Section 5.2, adaptive equalization techniques relevant to dynamic channels are discussed. In Section 5.3, tracking algorithms specially designed for channels impaired by SOP drift and PDL are presented. Finally, Section 5.4 discusses methods for phase noise compensation and tracking, as well as pilot placement strategies for channels impaired by correlated EO-comb phase noise.

5.1 Static Estimators

In this section, we present channel estimation techniques typically applied in scenarios where the parameter requiring estimation remains static and stable over time. Consequently, these methods are particularly suitable for estimating block-constant channels like the one introduced earlier in Section 4.2. Additionally, many types of block-constant wireless channels can be effectively estimated using simple static methods, including the commonly encountered block-fading channels [52].

Here, we concentrate on data-aided (pilot-based) channel estimation methods where at each transmission block, a sequence of known symbols (pilots) with length $L_\mathrm{p} \geq M$ is transmitted with M representing the number of channels. These pilots form an $M \times L_\mathrm{p}$ matrix of complex-valued symbols $\mathrm{D} = (\underline{d}_0, ..., \underline{d}_{L_\mathrm{p}-1})$ with $\underline{d}_i$ being known M-dimensional pilot vectors for all $i \in \{0, \ldots, L_\mathrm{p} - 1\}$. Figure 5.1 depicts a typical

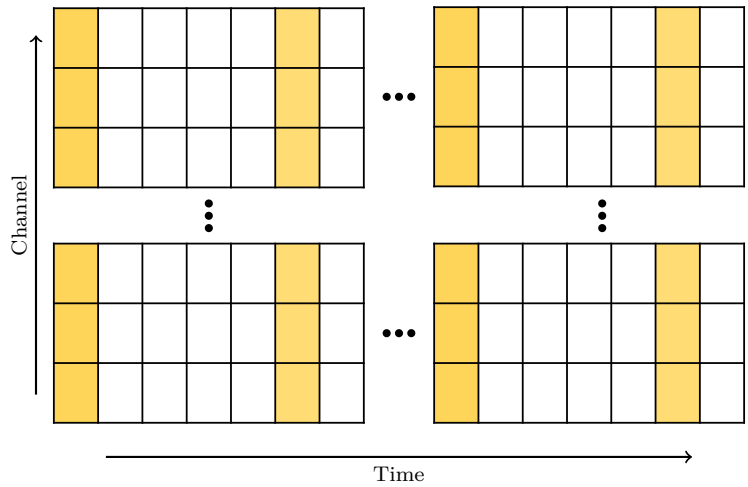

Figure 5.1: Typical pilot placement in a MIMO channel.

transmission block in a MIMO channel. Subsequently, the receiver attempts to estimate the unknown channel matrix using the provided pilots.

In the following subsections, we introduce pilot-aided techniques to estimate memoryless MIMO-AWGN channels similar to (4.6) with one difference that the channel is not necessarily unitary as relaxing the unitary condition allows for considering PDL as well. Considering the channel model in (4.6), and known pilots transmission, we can form the complex matrix of received symbols as $\mathbf{Y} = (\underline{\mathbf{y}}_0, ..., \underline{\mathbf{y}}_{L_\mathrm{p}-1})$ and the complex noise matrix as $\mathbf{Z} = (\underline{\mathbf{z}}_0, ..., \underline{\mathbf{z}}_{L_\mathrm{p}-1})$. Here, $\mathbf{Z}$ is an $M \times L_\mathrm{p}$ matrix made of i.i.d. noise vectors, each assumed to be $\mathcal{CN}(\underline{0}, \sigma_z^2 \mathrm{I}_M)$. We can write the matrix form as

$$\mathbf{Y} = \mathbf{H}\mathrm{D} + \mathbf{Z}, \tag{5.1}$$

where $\mathbf{H}$ is an $M \times M$ random channel matrix assumed to remain constant during a transmission block.

5.1.1 Least Squares Algorithm

The least-square error (LS) estimator derives the channel estimate by minimizing the squared error between the desired (pilot) and received signals. For MIMO-AWGN channels, the LS optimization problem can be simplified to [76]

$$\hat{\mathbf{H}}_\mathrm{LS} = \arg\min_{\hat{\mathbf{H}}} \left\| \mathbf{Y} - \hat{\mathbf{H}}\mathrm{D} \right\|_\mathrm{F}^2. \tag{5.2}$$

Given the matrix of pilots D and the matrix of received symbols $\mathbf{Y}$, the solution of (5.2) is given by [76], [77, Ch. 8]

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \mathbf{Y}\mathrm{D}^{\dagger}(\mathrm{D}\mathrm{D}^{\dagger})^{-1}. \tag{5.3}$$

The optimal choice of pilot symbols when the total transmission power is constrained has been extensively studied in the literature [78]–[80]. Let us consider the following power constraint on pilot transmission

$$\|\mathrm{D}\|_{\mathrm{F}}^{2} \leq L_{\mathrm{p}}\rho, \tag{5.4}$$

where $\rho \geq 0$. Thus, the optimal pilot matrix D that satisfies (5.4) and minimizes the channel estimation error, can be obtained as follows

$$\bar{\mathrm{D}} = \arg\min_{\mathrm{D}} \mathbb{E}\left[\left\|\mathbf{H} - \hat{\mathbf{H}}_{\mathrm{LS}}\right\|_{\mathrm{F}}^{2}\right], \qquad \text{s.t} \qquad \|\mathrm{D}\|^{2} \leq L_{\mathrm{p}}\rho. \tag{5.5}$$

It can be shown that the optimal pilot matrix $\bar{\mathrm{D}}$ has orthogonal rows, that is [78]

$$\bar{\mathrm{D}}\bar{\mathrm{D}}^{\dagger} = \frac{L_{\mathrm{p}}\rho}{M}\mathrm{I}_{M}. \tag{5.6}$$

Thus, any pilot matrix with orthogonal rows, each having the same norm $\sqrt{\rho L_{\mathrm{p}}/M}$, is optimal [79], [80]. It's worth noting that numerous pilot sequences satisfy these conditions.

Substituting (5.6) into (5.3) we obtain

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \frac{M}{\rho L_{\mathrm{p}}}\mathbf{Y}\bar{\mathrm{D}}^{\dagger} = \mathbf{H} + \frac{M}{\rho L_{\mathrm{p}}}\mathbf{Z}\bar{\mathrm{D}}^{\dagger}. \tag{5.7}$$

Define the estimation error matrix as $\mathbf{E} = \mathbf{H} - \hat{\mathbf{H}}_{\mathrm{LS}}$. Then, it has been shown that the error correlation matrix is [76]

$$\begin{aligned} \mathrm{R}_{\mathbf{E}} &= \mathbb{E}\left[\mathbf{E}^{\dagger}\mathbf{E}\right] \\ &\overset{(a)}{=} \frac{M^{3}}{\rho^{2}L^{2}}\bar{\mathrm{D}}\bar{\mathrm{D}}^{\dagger} \\ &\overset{(b)}{=} \frac{M^{2}\sigma_{z}^{2}}{\rho L_{\mathrm{p}}}\mathrm{I}_{M} = \frac{M}{\eta L_{\mathrm{p}}}\mathrm{I}_{M}, \end{aligned} \tag{5.8}$$

where (a) follows since $\mathbb{E}[\mathbf{Z}^{\dagger}\mathbf{Z}] = M\sigma_{z}^{2}\mathrm{I}_{L_{\mathrm{p}}}$, (b) follows from (5.6), and and $\eta = \rho/(M\sigma_{z}^{2})$ denotes the SNR per channel. It is evident that the estimation error of the LS algorithm decreases with increasing SNR and pilot length L_{p}.

5.1.2 Minimum Mean Square Error Algorithm

Let us obtain a linear estimator that minimizes the estimation error in the mean square error (MSE) sense. The solution of minimum mean square error (MMSE) estimator can be expressed in the following general form [81]

$$\hat{\mathbf{H}}_{\mathrm{MMSE}} = \mathbf{Y}\mathbf{A}^{*}, \tag{5.9}$$

where

$$\mathbf{A}^{*} = \arg\min_{\mathbf{A}} \mathbb{E}_{\mathbf{H},\mathbf{Z}} \left[\|\mathbf{H} - \mathbf{Y}\mathbf{A}\|_{\mathrm{F}}^{2} \right] \tag{5.10}$$

minimizes the MSE.

The solution of the MMSE problem is given [82] as

$$\mathbf{A}^{*} = \left(\mathrm{D}^{\dagger} \mathrm{R}_{\mathbf{H}} \mathrm{D} + \sigma_z^2 M \mathrm{I}_{L_{\mathrm{p}}} \right)^{-1} \mathrm{D}^{\dagger} \mathrm{R}_{\mathbf{H}}, \tag{5.11}$$

where $\mathrm{R}_{\mathbf{H}} = \mathbb{E}[\mathbf{H}^{\dagger}\mathbf{H}]$ is the channel correlation matrix [6]. Thus, the MMSE solution is

$$\hat{\mathbf{H}}_{\mathrm{MMSE}} = \mathbf{Y} \left(\mathrm{D}^{\dagger} \mathrm{R}_{\mathbf{H}} \mathrm{D} + \sigma_z^2 M \mathrm{I}_{L_{\mathrm{p}}} \right)^{-1} \mathrm{D}^{\dagger} \mathrm{R}_{\mathbf{H}}. \tag{5.12}$$

Let $\mathbf{E} = \mathbf{H} - \hat{\mathbf{H}}_{\mathrm{MMSE}}$. Then, the error correlation matrix is [82]

$$\mathrm{R}_{\mathbf{E}} = \mathbb{E}[\mathbf{E}^{\dagger}\mathbf{E}] = \left(\mathrm{R}_{\mathbf{H}}^{-1} + \frac{\mathrm{D}\mathrm{D}^{\dagger}}{\sigma_z^2 M} \right)^{-1}. \tag{5.13}$$

Note that, unlike the LS estimation, the MMSE requires knowledge of both the channel and noise correlation matrices. In practice, these matrices are often estimated, which can introduce errors affecting the performance of the MMSE estimator. Furthermore, in low SNRs, MMSE provides better estimation than the LS for any valid R_H, and as the SNR increases, it converges towards the performance of the LS estimator when using the same pilot matrix for both algorithms [78], [82].

The optimal pilot matrix can also be investigated from the MMSE viewpoint. The optimal pilot selection follows the water-filling principle where the pilot power allocated to each channel is carefully selected considering factors such as SNR and the channel correlation matrix $\mathrm{R}_{\mathbf{H}}$. For a deeper understanding of the optimal pilot matrix design, interested readers are referred to [82, Sec. V].

The estimated channel matrices $\hat{\mathbf{H}}_{\mathrm{LS}}$ and $\hat{\mathbf{H}}_{\mathrm{MMSE}}$ can be used at the receiver for channel equalization and symbol detection. For instance, channel equalization using MMSE estimation can be done by forming $\hat{\mathbf{Y}} = \hat{\mathbf{H}}_{\mathrm{MMSE}}^{-1}\mathbf{Y}$. Then, symbol detection can be done on $\hat{\mathbf{Y}}$. A less computationally expensive way to use the estimated channel matrix

[6] This matrix differs from the traditional definition of the channel correlation matrix, which is typically expressed as $\mathbb{E}[\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^{\dagger}]$. Here, the $\mathrm{vec}(\cdot)$ operator stacks the columns of a matrix vertically.

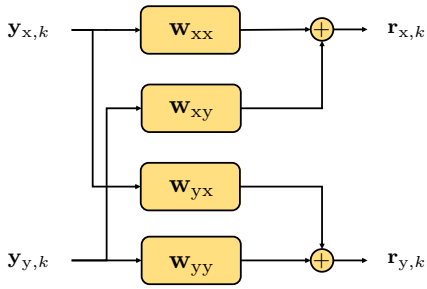

Figure 5.2: A general scheme for 2×2 feed-forward adaptive filters where $\mathbf{y}_{\mathrm{x}}, \mathbf{y}_{\mathrm{y}}$ denote the received signal in x and y polarizations, $\mathbf{w}_{\mathrm{xx}}, \mathbf{w}_{\mathrm{xy}}, \mathbf{w}_{\mathrm{yx}}, \mathbf{w}_{\mathrm{yy}}$ denote the equalizer weights, and $\mathbf{r}_{\mathrm{x}}, \mathbf{r}_{\mathrm{y}}$ denote the equalized signal in both polarizations.

for symbol detection is to use $\hat{\mathbf{H}}_{\mathrm{LS}}$ or $\hat{\mathbf{H}}_{\mathrm{MMSE}}$ to modify the constellations and hence modify the decision regions.

5.2 Adaptive Equalization

Static estimators perform well under the block-constant assumption. In practice, however, certain transmission impairments show dynamic characteristics, causing the channel response to vary over time. In this section, we present dynamic equalization techniques where a filter is continuously adjusted at the receiver to minimize a specific error function [83]. The choice of the filter and error function depends on the channel model under investigation. A dynamic equalizer tries to reduce the error caused by the combination of transmission impairments (e.g., phase noise and SOP drift) and the residual distortions that are not mitigated by the static filtering stage. The general schematic for feed-forward equalizers is shown in Figure 5.2.

Depending on the channel's memory, the equalization can be done in time or frequency domains. For instance, when the memory is large, frequency domain equalization can significantly reduce the computational complexity. Here, we focus on the time-domain DP channel defined as

$$\underline{\mathbf{y}}_k = \mathbf{H}_k \underline{\mathbf{x}}_k + \underline{\mathbf{z}}_k, \tag{5.14}$$

where k is the time index and $\mathbf{H}_k$ is a random 2×2 complex channel matrix. The DP received vector is denoted by $\underline{\mathbf{y}}_k = (\mathbf{y}_{\mathrm{x},k}, \mathbf{y}_{\mathrm{y},k})$, the transmitted vector by $\underline{\mathbf{x}}_k = (\mathbf{x}_{\mathrm{x},k}, \mathbf{x}_{\mathrm{y},k})$, and the noise vector by $\underline{\mathbf{z}}_k = (\mathbf{z}_{\mathrm{x},k}, \mathbf{z}_{\mathrm{y},k}) \sim \mathcal{CN}(\underline{0}, \sigma_z^2 \mathrm{I}_2)$. The subscripts denote the x and y polarizations.

The equalization is done by utilizing a finite impulse response matrix with one tap per

polarization[7] as

$$\underline{\mathbf{r}}_k = \mathbf{W}_k \underline{\mathbf{y}}_k, \tag{5.15}$$

where $\underline{\mathbf{r}}_k = (\mathbf{r}_{\mathrm{x},k}, \mathbf{r}_{\mathrm{y},k})$ denotes the equalized vector and $\mathbf{W}_k$ denotes the 2×2 complex-valued equalization matrix. The estimated channel matrix $\hat{\mathbf{H}}_k$ can be defined as

$$\hat{\mathbf{H}}_k = \mathbf{W}_k^{-1}. \tag{5.16}$$

The equalization problem can be solved by selecting an optimization cost function. Here, we focus on equalizers that minimize the expected equalization error in the MSE sense with the cost function

$$\epsilon_k = \mathbb{E}\left[\underline{\mathbf{e}}_k^{\dagger} \underline{\mathbf{e}}_k\right], \tag{5.17}$$

where $\underline{\mathbf{e}}_k = (\mathbf{e}_{\mathrm{x},k}, \mathbf{e}_{\mathrm{y},k})$ is the equalization error vector. The selection of the $\underline{\mathbf{e}}_k$ and hence cost function depends on the availability of known pilots, availability of feedback signal, allowed computational complexity, the memory of the channel, etc. This equalization problem can be solved using stochastic gradient descent and the update rule is given as [84]

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \mu \nabla_{\mathbf{W}_k} \epsilon_k, \tag{5.18}$$

where μ denotes the step size also known as the convergence parameter, and $\nabla_{\mathbf{W}_k}$ denotes the gradient with respect to the matrix $\mathbf{W}_k$.

5.2.1 Least Mean Squares Algorithm

The LS estimator requires all pilots simultaneously, and in scenarios with a high number of pilots and channels, computing the estimated channel can introduce delays. This issue can be elevated by using an iterative equalization algorithm. The least mean square (LMS) algorithm is an adaptive gradient search method that iteratively updates a weight vector at each sampling instant to minimize the MSE [85], [86].

The integration of the LMS algorithm into the adaptive equalization framework, as outlined in Section 5.2, is seamless. For the LMS algorithm, the equalization error is selected as

$$\underline{\mathbf{e}}_k^{\mathrm{LMS}} = \underline{d}_k - \underline{\mathbf{r}}_k, \tag{5.19}$$

where $\underline{\mathbf{r}}_k$ is defined in (5.15) and $\underline{d}_k$ is the desired symbols (pilots known to the receiver) at time k. Then, replacing (5.19) into (5.17) and computing the gradient, we get

$$\nabla_{\mathbf{W}_k} \epsilon_k = -\underline{\mathbf{e}}_k^{\mathrm{LMS}} \underline{\mathbf{y}}_k^{\dagger}. \tag{5.20}$$

[7] It is rather straightforward to extend the equalizers to accommodate more taps which is useful when the channel has memory. However, such equalizers are not utilized in the appended papers.

Replacing (5.20) into (5.18) the LMS update rule becomes

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu \underline{\mathbf{e}}_k^{\mathrm{LMS}} \underline{\mathbf{y}}_k^{\dagger}. \tag{5.21}$$

In case the LMS algorithm is applied to a channel that remains constant over time like the block-constant channel in (4.6), [86] demonstrated that as k approaches infinity, using (5.16), the solution of LMS algorithm converges to the optimal solution of the LS algorithm given in (5.3).

When the LMS algorithm is employed on a channel that remains constant over time, such as the block-constant channel described in (4.6), [86] demonstrated that as k tends to infinity, the solution of the LMS algorithm converges to the optimal solution of the LS algorithm, as given in (5.3).

Although optical channels in the presence of polarization impairments are not constant over time, the LMS algorithm can still be employed for tracking the channel. Moreover, a well-known trade-off exists between misadjustment and convergence speed, where a smaller step size μ leads to reduced misadjustment but slower convergence [86]. The optimization of step size has been investigated in [87] and [88], proposing a straightforward variable step size algorithm that provides both rapid tracking and minimal misadjustment.

5.2.2 Constant Modulus Algorithm

The constant modulus algorithm (CMA) was originally proposed as a blind equalizer for constant modulus constellations such as quadrature phase-shift keying (QPSK) [89]. The CMA and its variants such as modified CMA [90] and multi-modulus algorithm (MMA) [91] are the most popular adaptive equalizers in the field of optical communications. The CMA as its name suggests aims to minimize the MSE between a given reference level $\underline{a}_{\mathrm{ref}}$ and the equalized symbols $\underline{\mathbf{r}}_k$. Thus, the error vector of the CMA algorithm is defined as [89], [90]

$$\underline{\mathbf{e}}_k^{\mathrm{CMA}} = \underline{a}_{\mathrm{ref}} - \left|\underline{\mathbf{r}}_k\right|^2, \tag{5.22}$$

where $\left|\underline{\mathbf{r}}_k\right| = \left(\left|\mathbf{r}_{\mathrm{x},k}\right|, \left|\mathbf{r}_{\mathrm{y},k}\right|\right)$.

Note that $\underline{a}_{\mathrm{ref}}$ denotes the reference level for both polarizations and is a constant vector regardless of the number of amplitude levels of the transmitted constellations. Thus, the CMA works well for constant modulus constellations like DP-QPSK but its performance reduces for higher order constellations such as general DP-QAM. With a performance penalty, one can use the CMA for general QAM constellations by selecting $\underline{a}_{\mathrm{ref}}$ as[8] [90]

$$\underline{a}_{\mathrm{ref}} = \frac{\mathbb{E}[|\underline{c}|^4]}{\mathbb{E}[|\underline{c}|^2]}, \tag{5.23}$$

[8] $|\underline{c}| = \left(|c_{\mathrm{x}}|, |c_{\mathrm{y}}|\right)$ as mentioned in Section 1.2.

where $\underline{c} \in \mathcal{C}^2$ are the DP constellation points. Replacing (5.22) into (5.17) and computing the gradient, we get

$$\nabla_{\mathbf{W}_k} \epsilon_k = -\left(\underline{\mathbf{e}}_k^{\mathrm{CMA}} \circ \underline{\mathbf{r}}_k\right) \underline{\mathbf{y}}_k^{\dagger}, \tag{5.24}$$

where $(\circ)$ denotes the element-wise product. Finally, substituting (5.24) into (5.18) gives the estimated equalization matrix for CMA.

CMA implementation in a fully blind mode is generally reliable. However, due to the unconstrained nature of the equalizer outputs, there is a chance of convergence to the same output in both polarizations, resulting in a singular Jones matrix. Nonetheless, numerous established methods exist to effectively address such singularity issues [32], [92], [93].

As mentioned, the CMA is not suitable for multi-level constellations, but modifications in the error function can make it more suitable. For example, the radially directed equalizer (RDE) [94] can be obtained by modifying the error function to

$$\underline{\mathbf{e}}_k^{\mathrm{RDE}} = Q_{\mathrm{r}}(|\underline{\mathbf{r}}_k|^2) - |\underline{\mathbf{r}}_k|^2, \tag{5.25}$$

where $Q_{\mathrm{r}}(\cdot)$ is a function that quantizes the radius of each polarization based on the number of possible rings of the utilized constellation scheme. The probability of each ring is different, and thus the error function of the RDE can be modified to obtain probabilistic RDE [95].

5.2.3 Decision-Directed Least Mean Squares Algorithm

The CMA, with its single-level $\underline{a}_{\mathrm{ref}}$, is not well suited to multi-level constellations. The RDE, on the other hand, addresses this limitation by expanding the number of reference levels using a decision-directed (DD) approach. To further elevate the dependency on modulation formats, a fully DD scheme known as decision-directed least mean squares (DDLMS) is employed in [96], where the most likely decided symbol

$$\hat{\underline{d}}_k = \arg\min_{\underline{c} \in \mathcal{C}^2} \|\underline{\mathbf{r}}_k - \underline{c}\|^2, \tag{5.26}$$

replaces the fixed reference levels. Moreover, unlike the LMS algorithm that requires the complete knowledge of desired symbols $\underline{d}_k$ at the receiver, the DDLMS algorithm can be considered blind. This saves throughput compared to LMS as the pilot symbols required in LMS contain redundant information.

Assume that the transmitted signal at time k is a 2-dimensional random vector $\underline{\mathbf{x}}_k$, which takes on values from a set $\mathcal{C}^2$. Then, the error vector of the DDLMS is defined as [96]

$$\underline{\mathbf{e}}_k^{\mathrm{DDLMS}} = \hat{\underline{d}}_k - \underline{\mathbf{y}}_k. \tag{5.27}$$

Consequently, the DDLMS algorithm aims to reduce the variance between the detected symbol and the decided symbol.

The convergence speed and accuracy of the DDLMS algorithm are highly dependent on the symbol error rate (SER)[9] of the decision symbols $\underline{\hat{d}}_k$. In general, the SER is affected by the SNR as it determines how well the constellation points are separated. while DDLMS shows robust performance at higher SNRs, it may not be reliable at lower SNRs.

Since the initial decisions are more likely to be wrong (high SER), the blind startup of the DDLMS is not reliable, especially at low SNRs. Therefore, to ensure reliable convergence, either the pilot-based LMS algorithm or the CMA are employed for the pre-convergence phase, then switched to the DDLMS algorithm. While CMA and DDLMS share similar mathematical models, DDLMS has an extra feedback latency due to its DD circuitry [97].

5.3 SOP Drift Compensation

In this section, we first introduce the Kabsch algorithm [98] which is useful to estimate the block-constant SOP drift channel introduced in Section 4.2. Then, we briefly introduce the Czegledi algorithm [99] which is designed to track the fast SOP drift channels e.g., the model introduced in Section 4.3 when PDL and PMD are negligible. Finally, we briefly introduce the proposed algorithms in Paper B, which, unlike the Kabsch and Czegledi algorithms, account for the PDL as well as fast SOP drift.

The proposed algorithms do not compensate for PMD which does not reflect realistic fiber channels where PMD typically is present. However, these algorithms can be used in a hybrid fashion, as demonstrated in [100], where PMD compensation is done in the frequency domain, while SOP tracking is performed in the time domain.

5.3.1 The Kabsch Algorithm

With negligible nonlinearities, PMD, and PDL, the SOP drift channel models introduced in Sections 4.2 and 4.3 can be regarded as unitary channels. However, both blind estimation algorithms (e.g., CMA, RDE, and MMA) and pilot-aided estimation algorithms (e.g., LS and LMS) face a common issue: their optimization problems are not tailored for unitary channel estimation. Consequently, their solutions are suboptimal when the channel matrix is unitary, i.e., $\mathbf{H}^\dagger\mathbf{H} = \mathbf{H}\mathbf{H}^\dagger = \mathrm{I}_M$ where $M \geq 1$ denotes the number of channels.

Here, we incorporate the unitary constraint of the channel into the estimation problem

[9] The SER quantifies the probability of symbol errors.

described by (5.2) and formulate it as

$$\min_{\hat{\mathbf{H}}} \left\| \mathbf{Y} - \hat{\mathbf{H}}\mathrm{D} \right\|^2 \quad \text{s.t.} \quad \hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger = \mathrm{I}_M, \tag{5.28}$$

where D denotes the pilot matrix with orthogonal rows defined in (5.6). The Kabsch algorithm [98] provides the optimal solution to this problem as

$$\hat{\mathbf{H}}_{\text{Kabsch}} = \mathbf{U}\mathbf{V}^\dagger. \tag{5.29}$$

Here, $\mathbf{U}\Sigma\mathbf{V}^\dagger$ represents the singular value decomposition of $\mathbf{Y}\mathrm{D}^\dagger$. In simple words, the Kabsch algorithm gives the closest unitary matrix to the LS solution. In the case of the DP channel. i.e., $M = 2$, the singular value decomposition is deployed only on a two-by-two matrix, making it less computationally complex than for higher M.

Unlike blind estimators, the Kabsch algorithm is independent of the modulation format because it estimates the channel from the pilot symbols. This is also useful in scenarios where the pilots and transmitted data have different modulation formats. However, complete dependence on pilot symbols reduces the applicability of the Kabsch algorithm. This issue is partly addressed in [101] where the DD symbols are utilized instead of pilots. As previously mentioned, it is better to use DD-based equalizers in combination with pilot-based or blind algorithms as they perform poorly during the pre-convergence phase.

5.3.2 The Czegledi Algorithm

The Czegledi algorithm [99] is designed to track both the carrier phase and SOP drift jointly utilizing a DD-based architecture. The Czegledi algorithm shows a fast convergence rate and high tolerance to fast SOP drifts, and it is modulation format independent. In the following, we briefly describe the Czegledi algorithm while discarding its carrier phase estimation part. We refer the reader to [99] for the complete description of the algorithm and its performance analysis.

The Czegledi algorithm is designed for the channel model (4.8) in Section 4.3 when no concatenation $N = 1$ is done and PDL is neglected, which gives $\mathbf{H}_k = \mathbf{J}_{k,1}$. Using (4.11), the channel matrix can be simplified to

$$\mathbf{H}_k = \exp\left(-j\underline{\boldsymbol{\alpha}}_{k,1} \cdot \overrightarrow{\sigma}\right) \mathbf{H}_{k-1}, \tag{5.30}$$

which is then used in (4.7) to define the input–output relation.

Assume that the 2-dimensional random transmitted vector $\underline{\mathbf{x}}_k$ takes on values from a set $\mathcal{C}^2$ with complex, zero-mean, and equiprobable DP constellation points. Assuming that $\mathbf{H}_k$ does not change significantly over a symbol duration, the most likely estimate of the transmitted symbol can be obtained using the previous estimate of the channel

$\hat{\mathbf{H}}_k$ and employing the minimum Euclidean distance criterion as [99, Eq. (16)]

$$\hat{\underline{\mathbf{x}}}_k = \arg\min_{\underline{c}\in\mathcal{C}^2} \left\| \hat{\mathbf{H}}_k^{-1}\underline{\mathbf{y}}_k - \underline{c} \right\|^2, \tag{5.31}$$

where $\underline{\mathbf{y}}_k = \mathbf{H}_k\underline{\mathbf{x}}_k + \underline{\mathbf{z}}_k$ denotes the DP vector of received symbols as defined in (4.7).

Then, the estimated channel at time $k+1$ is updated as

$$\hat{\mathbf{H}}_{k+1} = \exp(-j\hat{\underline{\boldsymbol{\alpha}}}_{k,1}\cdot\overrightarrow{\sigma})\hat{\mathbf{H}}_k, \tag{5.32}$$

where $\overrightarrow{\sigma}$ is the Pauli tensors (4.13) and $\hat{\underline{\boldsymbol{\alpha}}}_{k,1}$ is the estimate of the innovation vector $\underline{\boldsymbol{\alpha}}_{k,1}$ defined in (4.12). The estimate of the innovation can be calculated by solving the following optimization problem

$$\hat{\underline{\boldsymbol{\alpha}}}_{k,1} = \arg\min_{\underline{\alpha}\in\mathbb{N}^3} \left\| \left(\exp(-j\underline{\alpha}\cdot\overrightarrow{\sigma})\hat{\mathbf{H}}_k\right)^{-1}\underline{\mathbf{y}}_k - \hat{\underline{\mathbf{x}}}_k \right\|^2, \tag{5.33}$$

Using the gradient descent method, the innovation vector can be estimated as

$$\hat{\underline{\boldsymbol{\alpha}}}_{k,1} = -2\mu_\alpha\mathcal{R}\left(j\left(\hat{\mathbf{H}}_k^{-1}\underline{\mathbf{y}}_k - \hat{\underline{\mathbf{x}}}_k\right)^\dagger \hat{\mathbf{H}}_k^{-1}\overrightarrow{\sigma}\underline{\mathbf{y}}_k \right), \tag{5.34}$$

where $\mu_\alpha \geq 0$ is the tracking step size and $\mathcal{R}(\cdot)$ extracts the real part of a complex scalar or vector. The tracking step size controls the speed of the convergence, the tracking accuracy, and the tolerance of the algorithm to fast changes.

5.3.3 The Proposed Algorithms

As discussed earlier, the Kabsch algorithm [92], [98] performs well for block-constant channels but is not suitable when the channel shows dynamic behavior as the one described in Section 4.3. In Paper B, we propose the sliding window Kabsch (SW-Kabsch) algorithm, an adaptive extension of the Kabsch algorithm tailored to handle SOP drift channels. Our SW-Kabsch algorithm demonstrates superior tolerance to rapid SOP drift compared to existing benchmarks.

While the Czegledi algorithm [99] tracks SOP drift, it does not consider PDL. Therefore, in Paper B, we also introduce the sliding window least squares (SW-LS) algorithm, which addresses both SOP drift and PDL effects.

When the PDL is negligible, SW-Kabsch shows greater tolerance to SOP drift compared to SW-LS and gradient descent (GD)-based algorithms (such as the Czegledi algorithm, CMA, and their variations). However, when considering PDL, the channel loses its unitary property, and SW-LS shows the highest SOP drift tolerance. Additionally, unlike GD-based algorithms, which require adjustment of step sizes, the proposed algorithms need no further parameter tuning, making them promising choices for memoryless fast-drifting optical systems. Further details about the algorithm can be found

in Paper B.

5.4 Phase-Noise Compensation

The presence of phase noise limits the use of high-order modulation formats and hence limits the transmission throughput. The problem of phase noise compensation has been extensively studied in both wireless [102]–[105] and optical domains [106]–[109]. In this section, we focus mainly on tracking the correlated phase noise encountered in EO-comb-based WDM systems. Thus, we begin with introducing the blind phase search (BPS) algorithm, a popular blind phase recovery algorithm. Thereafter, we introduce a Kalman-based filter designed for correlated phase noise tracking. Finally, we discuss how the phase noise correlation can influence the pilot placement strategy.

5.4.1 Blind Phase Search Algorithm

The BPS algorithm [110] is a blind algorithm to estimate the carrier phase noise by rotating the received signal $B > 0$ times and selecting the rotation that gives the lowest Euclidean distance to the closest constellation point. The BPS workflow is presented in the following.

Let

$$\phi_b = \frac{2\pi b}{B}, \qquad b = 0, 1, \ldots, B-1 \tag{5.35}$$

denote the test phases and $\Phi = \{\phi_0, \ldots, \phi_{B-1}\}$ denote the set of all the test phases. For constellations with four-fold symmetry, e.g., QAM, the test phases can be limited to the range of $[0, \pi/2]$. Also, let the M-dimensional random transmitted vector $\underline{\mathbf{x}}_k$ take on values from a set of complex, zero-mean, and equiprobable DP constellation points, denoted as $\mathcal{C}^2$.

At the first step of the BPS algorithm, the received signal $\underline{\mathbf{y}}_k$ is rotated with all the test phases simultaneously and the error function is calculated using the Euclidean distance to the closest constellation point as

$$\mathbf{e}_{k,b} = \left\| e^{-j\phi_b} \underline{\mathbf{y}}_k - \underline{\hat{\mathbf{x}}}_{k,b} \right\|^2, \tag{5.36}$$

where

$$\underline{\hat{\mathbf{x}}}_{k,b} = \arg\min_{\underline{c} \in \mathcal{C}^2} \left\| e^{-j\phi_b} \underline{\mathbf{y}}_k - \underline{c} \right\|^2 \tag{5.37}$$

is the closest constellation point to $e^{-j\phi_b}\underline{\mathbf{y}}_k$.

Then, to suppress the impact of the additive noise, the error function is averaged over

a window of length $L_{\mathrm{BPS}} \geq 1$

$$\bar{\mathbf{e}}_{k,b} = \sum_{|k'| \leq L_{\mathrm{BPS}}/2} \mathbf{e}_{k-k',b}. \tag{5.38}$$

Finally, the estimated phase is chosen as

$$\phi_b^* = \arg \min_{\phi_b \in \mathbf{\Phi}} \bar{\mathbf{e}}_{k,b} \tag{5.39}$$

The size of the window L_{BPS} in the BPS algorithm is dependent on the speed of phase noise variation, imposing a trade-off between noise level and phase noise adaptability. The accuracy of the BPS algorithm grows linearly with the number of test phases B, leading to increased computational complexity. As a result, for higher-order modulation schemes like 64-QAM, the algorithm's applicability is constrained by computational complexity. To address this issue, multiple adaptations and extensions of the BPS algorithm have emerged, often employing a multi-stage approach for phase estimation [110]–[112], combining coarse and precise estimation steps with different algorithms.

5.4.2 Extended Kalman Filter Algorithm

The extended Kalman filter (EKF) [83, Ch. 7.2] serves as a powerful tool in estimation and prediction tasks, particularly in nonlinear systems. By extending the principles of the traditional Kalman Filter to nonlinear models through linearization, the EKF provides an efficient means of estimating the state of a dynamic system based on noisy measurements. The use of EKF for phase noise tracking involves initializing parameters, predicting the next state based on a model accounting for phase noise, correcting estimates using the received data (decision-directed or pilots), and iteratively refining phase noise parameters for accurate tracking in dynamic systems.

The iterated EKF, also known as extended Kalman smoother (EKS) [83, Ch. 4.7], is an enhanced version of the EKF. It involves iteratively updating the EKF using linearization at predicted phase noise values. By repeating this process, initially based on predicted phase noise and then on updated estimates, we progressively refine our understanding of the phase noise, resulting in improved estimation accuracy.

The EKS estimator for correlated phase noise tracking in optical communication systems is introduced in [109, Algorithm 1]. This method is utilized in Paper E to track the phase noise $\underline{\boldsymbol{\theta}}_k$ within the EO-comb phase noise model described by (4.32)–(4.34). For more detailed information on implementing the algorithm, interested readers are referred to [83, Ch. 7.4] and [109, Algorithm 1].

5.4.3 Pilot-Symbol Placement

In general, the choice of where to place pilot symbols can significantly affect how well pilot-aided estimation algorithms perform [113], [114]. Previous research has shown that spreading pilots out evenly in time and across channels, similar to a grid, works best for systems with multiple channels [115]. Additionally, [116], [117] have demonstrated that the best pilot placement for space-division multiplexing systems depends on how much correlation there is between channels. However, there has not been much investigation into pilot placement for systems using EO-combs, where channels are completely correlated.

In Paper E, we investigate how phase correlation in EO-comb channels (see Section 4.5) affects the pilot placement strategy. We analyze different methods for placing pilots optimally. Unlike in a typical multichannel system, we find that placing pilots on the outer comb lines (carrier frequencies) is optimal for certain estimation techniques. Further details can be found in Paper E.

CHAPTER 6

Polarization-Based Fiber Sensing

The optical fibers serve not only as mediums for data transmission but also as sensors. It has been shown that optical fiber sensors are highly sensitive to external physical factors such as acoustic vibration, electric and magnetic fields, temperature, strain, humidity, and pressure. They are immune to electromagnetic interference and can withstand high temperatures and corrosive chemicals making them suitable for various harsh environments. The compactness and flexibility of optic fibers accompanied by their lightweight offer versatility in applications with space constraints and portability needs [118].

The application environments range from dangerous scenarios where there are radioactive, chemical, and other industrial-based hazards to more common and simple uses. Various industries and fields have benefited from fiber sensors, including healthcare systems, biomedical fields, food and water quality monitoring, and standard product manufacturing in the industry. The immunity to electromagnetic interference along with less frequent calibration needs, makes them ideal for remote monitoring and sensing applications [119]–[121] such as underwater sensing which enables monitoring of oil and gas infrastructure, seismic activity, and oceanographic parameters, ensuring the safety of underwater structures, and understanding marine ecosystems.

Distributed optical fiber sensing (DOFS) stands out among various sensor types by enabling the interrogation of numerous points along a single fiber, offering significant advantages over traditional node-type sensors for long-distance measurements [120]–[123]. The DOFS techniques can primarily be classified into two categories (separate and joint sensing and communication), which will be discussed in the subsequent subsections.

6.1 Separate Sensing and Communication

The separate sensing and communication (SSAC) techniques assign dedicated fibers or time and frequency channels for sensing. Among these, the methods utilizing backscattered light, such as Rayleigh [123]–[125], Brillouin [126], and Raman [127], have gained considerable attention over recent decades due to their ability to balance measurement accuracy and sensing distance. These techniques exploit different intrinsic fiber characteristics, such as length, attenuation, and breakpoints [128], as well as a wide range of physical parameters including temperature [129], strain [130], pressure [131], seismic waves [132], and refractive index [133], with high resolution.

Recently, a bidirectional method that relies on phase detection through ultra-low phase noise laser interferometry was proposed in [12]. This method involves transmitting CW pulses from each end of a fiber cable. When vibrations occur, the phases detected by coherent receivers at both ends can be correlated to pinpoint the location nearest to the incidence.

In the following, we provide a brief introduction to two of the most popular SSAC methods, followed by a discussion of their limitations.

6.1.1 Distributed Acoustic Sensing

Distributed acoustic sensing (DAS) [129] as a SSAC technique has become increasingly popular for geophysical applications, leveraging intrinsic Rayleigh backscattering to transform long stretches of fiber-optic cables into thousands of acoustic sensors. The term “acoustic” indicates that DAS is primarily used to detect sound waves or acoustic vibrations propagating through the medium surrounding the optical fiber, such as in seismic monitoring, underwater detection, and infrastructure monitoring applications. Offering high spatial resolution and sensitivity to external disturbances, DAS is widely used to monitor various geomechanical responses, including micro-fracturing within the earth and oceanic processes [134]. As fiber optic cables are widely deployed in urban areas, DAS is progressively being integrated into seismic detection systems, crucial for earthquake early warning and potentially integral to smart-city sensing networks [135], [136].

However, its spatial range is limited to approximately 150 km due to fiber attenuation, requiring the use of high-power lasers. Therefore, DAS is most effective when deployed on dark fibers without active telecommunication channels to minimize crosstalk and avoid potential traffic disruptions [14]. However, long-haul submarine cables typically have 2 to 8 pairs of fiber. This means that the unused (dark) fiber strands, which are needed for DAS, are rare [13].

6.1.2 Phase-Sensitive Optical Time-Domain Reflectometry

Phase-sensitive optical time-domain reflectometry (Φ-OTDR) [133] is another common backscattering-based DOFS technique which provides the basis for DAS. The Φ-OTDR technique has been successfully applied in many fields, including seismic wave detection (underground and undersea), traffic flow monitoring, and geological monitoring.

In Φ-OTDR, the inherent fiber characteristic of Rayleigh backscattering is utilized to measure dynamic strains (vibrations/intrusions) along the fiber. Although the sensing distance of Φ-OTDR is generally limited to a few tens of kilometers, it offers remarkable advantages of high sensitivity, high accuracy, fast response, and fully distributed measurements.

Initially, Φ-OTDR only utilized dedicated or idle undersea fiber cables as the sensing medium to detect seismic waves [137], [138]. In recent years, Φ-OTDR has been integrated into coherent optical communication networks by WDM to enable both data transmission and distributed vibration detection [139].

Both DAS and Φ-OTDR rely on the detection of backscattered light, but with different applications. The Φ-OTDR is mostly used for high-resolution sensing applications such as structural health monitoring [140], reservoir exploration [141], and environmental monitoring [142], [143]. On the other hand, DAS is mainly used for detecting acoustic disturbances with applications such as pipeline monitoring [129] and seismic monitoring [144].

6.2 Joint Communication and Sensing

Joint sensing and communication (JSAC) techniques involve methods that allow for both communication and sensing applications at the same time. Implementing JSAC techniques can potentially elevate challenges such as cost, scalability, and compatibility that are commonly faced in the SSAC techniques.

Recently, some sensing methods that depend on the phase and SOP of the forward transmission have been proposed [14], [132], [145]. Modern optical transceivers need to estimate the SOP and phase to demodulate the data from the received signal. Therefore, unlike SSAC techniques such as DAS, the SOP and phase are already accessible in the receiver without requiring a dark fiber, dedicated frequency channel, or special hardware.

In [146], the SOP of a terrestrial cable is monitored span-by-span to detect construction vehicles and hence avoid damaging the cable. Another study detected a magnitude 7.7 earthquake using in-service telecom signals transmitted through a submarine cable [13], [14].

The prior studies on utilizing polarization information for sensing, often relied on assumptions such as having access to multiple optical paths [12], [145], [147], reading outputs from each span [146], or utilizing loop-back configurations [148]. A question yet to be answered is how much location information about external environmental events

can be extracted by polarization sensing at the end of a single fiber.

In the upcoming subsections, we introduce techniques that extract spatially resolved polarization data directly from the estimated Jones matrix at the output of the DSP at the receiver. We briefly describe how these techniques can be used to obtain polarization information from the noisy estimate of the Jones matrix $\tilde{\mathbf{H}}(\omega_i; \underline{\Theta}_k)$ (4.24) introduced in Section 4.4.2.

6.2.1 Inverse Scattering Algorithm

The use of an inverse scattering algorithm (ISA) to extract the DGD profile from the overall impulse response (i.e., the time-domain Jones matrix) is explored. The ISA technique was further developed in [68] to handle PDL, allowing for the recovery of both DGD and PDL profiles. The ISA is theoretically accurate for the scenarios discussed in the cited works, under the assumption that the overall impulse response is completely known and free from noise.

The ISA algorithm analytically calculates the values of the unknown parameters $\underline{\Theta}_k$ (4.19) containing the rotation, PDL, and DGD parameters for each section and is performed iteratively in the time domain by first taking the inverse Fourier transform of the noisy estimate of the Jones matrix $\tilde{\mathbf{H}}(\omega_i, \Theta_k)$ (4.24). Then it calculates the unknown parameter values for the last section N, i.e., γ_N, $\phi_N(k)$, and $\psi_N(k)$, and subsequently employs these values to construct the equivalent total channel response for sections 1 to $N-1$, effectively removing the last section from the model. By repeating these steps, all the subsequent parameters will be obtained. Note that in the ISA algorithm proposed in [68], the DGD parameters τ are assumed to be the same and known for all sections.

6.2.2 The Proposed Algorithm

In Paper C, we introduce a physics-based machine learning method that directly characterizes the propagation model (4.19) and simultaneously optimizes all its associated parameters.

We represent the set of trainable parameters as $\hat{\underline{\Theta}}_k = \{\hat{\gamma}_n, \hat{\phi}_n(k), \hat{\psi}_n(k)\}_{n=1}^{N}$ and define the frequency-averaged cost function as

$$\mathcal{L}(\hat{\underline{\Theta}}_k) = \sum_{i=1}^{L} \left\| \tilde{\mathbf{H}}(\omega_i; \underline{\Theta}_k) - \mathrm{H}(\omega_i; \hat{\underline{\Theta}}_k) \right\|^2, \tag{6.1}$$

where L denotes the number of frequency samples. This loss function is then minimized iteratively using a gradient-descent optimizer. Here, we only have access to the overall noisy channel response $\tilde{\mathbf{H}}(\omega_i; \underline{\Theta}_k)$ without knowing its underlying parameters $\underline{\Theta}_k$.

The proposed algorithm exhibits better resilience to additive channel estimation noise when compared to the ISA algorithm. For further details, please refer to Paper C.

It is important to note that estimating the underlying parameters $\underline{\Theta}_k$ is not the end

goal. The ultimate goal is to recognize and interpret the perturbations in the estimated parameters as environmental disturbances and then draw conclusions about their potential physical causes (such as earthquakes, seismic waves, or traffic disturbances). However, there are numerous challenges even in these initial stages. For instance, establishing a clear link between a specific parameter perturbation and its cause requires a descriptive model, which may not always be available for various types of disturbances. Furthermore, in Section 4.4.1, we assumed that environmental changes only affect the SOP rotation elements, and modeled their impact on the parameters using Wiener processes. We made this simplifying assumption due to the lack of realistic models, but there is no reason to believe that it holds true in an actual fiber-optic system. This highlights the critical need for developing realistic and practical models that accurately describe the effects of environmental changes on the SOP of the transmitted light beam.

6.3 Comparing SSAC and JSAC Techniques

Here, we compare the SSAC and JSAC techniques at an abstract level. We also provide Table 6.1 which summarizes the subsequent points.

- **Compatibility with current networks:** Most of the SSAC techniques depend on Rayleigh backscattering, which poses compatibility issues with current telecom cables because of isolators at the output of inline amplifiers. Even if future submarine cables are reconfigured to support these techniques, achieving functionality over thousands of kilometers remains challenging due to the limited power of Rayleigh scattering and the low SNR of the received backscatter [145]. On the other hand, the JSAC techniques based on SOP or phase are entirely compatible with the current optical networks and require no adaptation.

- **Scalability:** The scalability of a technique refers to its ability to expand and be applied on a large scale efficiently and cost-effectively. In general the JSAC techniques are easy to scale while the SSAC ones require special hardware or dedicated bandwidth or fibers and are costly which reduces their scalability.

- **Sensing versatility:** Without question the SSAC techniques provide more powerful tools for detecting a wide variety of fiber stressors e.g., temperature, strain, vibration, etc.

- **Localization:** The SSAC techniques provide more accurate localization due to their distributed nature. Specific to DAS, it is capable of painting a comprehensive picture of strain along different spatial sections of fiber. Such information can be used to derive information about the conditions of the physical placement of the fiber (e.g. aerial vs. terrestrial), and the dynamics of its surrounding environment. On the other hand, JSAC techniques do not rely on back-scattering; instead, they integrate mechanical disturbance along the fiber length, which typically does not

Table 6.1: Comparison of JSAC and SSAC Techniques

Criteria	JSAC	SSAC
Compatibility	Minimal or no adaptation required.	Compatibility issues due to Rayleigh backscattering.
Scalability	Easy to scale.	Require special hardware or dedicated bandwidth/fibers.
Sensing Versatility	Limited.	More powerful in detecting various fiber stressors.
Localization	Limited in single-link setups; may require multiple links.	More accurate localization due to distributed nature; capable of providing comprehensive strain picture.
Interference	Interference-free.	Interference between communication and sensing signals can occur, affecting reliability and accuracy.
Throughput Constraints	Minimal impact on communication throughput.	Poses bandwidth constraints, reducing the nominal throughput of the communication system.
Cost & Security	Generally lower cost and fewer security concerns.	Costly equipment and security concerns related to eavesdropping and service disruption.

allow localization in a single link setup. For instance, the localization capability of the ISA and the proposed algorithms in a single link setup is still under question. However, simultaneous measurements on spatially diverse links or synchronization of multiple transmitter/receiver pairs on the same links can aid in localizing the disturbance source through interferometric measurements [12], [13], [145].

- **Signal interference:** While the JSAC techniques are interference-free, interference can occur between communication and sensing signals for SSAC methods, particularly affecting the reliability and accuracy of both functions. For example, in integrated systems like WDM-based Φ-OTDR, the transmission quality may degrade because the strong probe used in Φ-OTDR can easily trigger non-linear effects [149].
- **Throughput constraints:** The SSAC methods pose bandwidth constraints, especially when aiming to support both high-speed communication and high-resolution sensing simultaneously on the same fiber. These methods typically allocate either

an entire fiber cable or a portion of the frequency and time grids, consequently reducing the nominal throughput of the communication system [149].

- **Cost and security concerns:** The SSAC techniques using costly equipment like ultra-stable laser sources or specialized repeaters with built-in geophysical sensors raise concerns over cost, security, and data policies. Installing sensing equipment alongside telecommunications gear also raises security issues related to eavesdropping and service disruption [13].

CHAPTER 7

Summary of included papers

This chapter provides a summary of the included papers.

7.1 Paper A

"Capacity bounds under imperfect polarization tracking"

In this paper, motivated by the stability of SOP in buried and undersea fibers over long periods, we investigated the capacity of the block-constant polarization drift channel when polarization demultiplexing is not perfect. We developed an AIR using the mismatched decoding method and found that a unitary estimation of the channel leads to a more accurate AIR as the SOP channel behaves as a unitary matrix. Additionally, we introduced a data-aided version of the Kabsch algorithm and demonstrated through simulations that it achieves a higher AIR compared to the LS estimation algorithm.

Contributions: Magnus Karlsson (MK) proposed the channel model. Mohammad Farsi (MF) and Erik Agrell (EA) formulated the problem. MF derived the bounds, performed the simulations, and wrote the paper. EA also helped with the derivations and results validation. All authors reviewed and revised the paper.

Context: Sections 3.4, 4.2, 5.1.

7.2 Paper B

In Paper B, we remove the block-constant assumption of the channel from Paper A and extend the channel model to incorporate both SOP drift and PDL. We propose two polarization tracking algorithms, namely SW-Kabsch and SW-LS. Unlike GD-based techniques that require step size adjustments, these algorithms require no parameter tuning, rendering them promising choices for memoryless fast-drifting optical systems. We compare the performance and complexity of the proposed algorithms with state-of-the-art methods in the literature. Our analysis demonstrates that when PDL is negligible, the channel can be treated as a unitary matrix and SW-Kabsch, which ensures a unitary estimate of the channel, exhibits superior SOP drift tolerance compared to SW-LS and GD-based methods. However, when considering PDL, the power imbalance between the channels makes the channel nonunitary, resulting in SW-LS displaying the highest SOP drift tolerance.

Contributions: MF proposed the algorithms, performed the simulations, and wrote the paper. EA, MK, and Christian Häger (CH) helped with the problem formulation and results validations. All authors reviewed and revised the paper.

Context: Sections 4.3, 5.1, 5.2, 5.3.

7.3 Paper C

"Learning to Extract Distributed Polarization Sensing Data from Noisy Jones Matrices"

In this paper, we extended the model in Paper B to account for DGD and utilized the model to extract the polarization information from noisy estimates of the Jones matrices at the receiver. Our investigation showed that the performance of the existing ISA method quickly drops in the presence of the channel estimation noise. Thus, we proposed a novel physics-based learning strategy leveraging the proposed model that simultaneously optimizes all the underlying model parameters accounting for per section PDL, rotations, and DGD. We showcased its effectiveness in accurately recovering the location of time-varying polarization perturbations using the noisy channel estimates. However, we acknowledged persistent challenges, particularly concerning cases of model overparameterization which needs further exploration.

Contributions: MK and CH suggested the channel model, while MF devised the algorithm, conducted simulations, and wrote the paper. EA, MK, and CH contributed mathematical expertise, aiding in problem formulation and results validation. All authors participated in reviewing and revising the manuscript.

Context: Sections 4.4, 6.1, 6.2.

7.4 Paper D

"On the Capacity of Correlated MIMO Phase-Noise Channels: An Electro-Optic Frequency Comb Example"

In Paper D, we examined how a MIMO channel's capacity is affected by correlated phase noises from EO-combs. We established both lower and upper bounds determining the channel capacity for various values of SNR and phase-noise parameters. Additionally, we derived upper and lower capacity bounds for high-SNR scenarios, resulting in the multiplexing gain (pre-log) of $M-1$, where M is the number of channels (comb lines). Intuitively, the loss of one complex channel is the result of sacrificing two real dimensions to gain full knowledge about two unknown phase noise sources. Therefore, a feasible transmission strategy could involve using the phase of two channels (which carry no data) to estimate and eliminate phase noises from the other channels. Additionally, a constant gap, determined by the number of channels M, persists between the upper and lower bounds at high SNRs. When $M = 2$, this gap disappears, allowing for a characterization of the high-SNR capacity.

Contributions: MK developed the channel model and collaborated with EA and MF to define the problem. MF derived the bounds, conducted simulations, and authored the paper. EA contributed to the proof of multiple lemmas. EA and Hamdi Joudeh (HJ) contributed to the validation of the proofs and interpretation of the results. EA, Alex Alvarado (AA), Gabriele Liga (GL), and HJ offered mathematical insights and assisted in structuring the paper. All authors participated in reviewing and refining the manuscript.

Context: Sections 3.2 , 3.3, 3.4, 4.5.

7.5 Paper E

"Pilot Distributions for Phase Noise Estimation in Electro-Optic Frequency Comb Systems"

In this study, we explored the pilot placement issue in MIMO channels impacted by correlated phase noise from EO-combs and additive ASE noise. Through theoretical analysis and simulations, we demonstrated that for certain phase estimators, placing pilots at the outer channels (first and last) is optimal, in contrast to standard WDM channels where pilots are distributed across all channels. Furthermore, we conducted Monte Carlo simulations to compare various pilot distributions in terms of their effect on bit error rate performance when tracking correlated phase noise originating from EO-combs.

Contributions: MK introduced the channel model, while MF and EA conceptualized the problem. MF developed the theoretical framework, conducted simulations, and wrote the paper. All authors contributed to reviewing and revising the manuscript.

Context: Sections 4.5, 5.2.

7.6 Future Work

Regarding Papers A and B, a possible future direction is to adjust the studied channel models to account for PMD. This adaptation introduces memory in the channel and complicates the capacity analysis and the DSP algorithm required for channel tracking. To deal with the memory, algorithms in Papers B could be altered by increasing the equalization window and replacing the closed-form solutions with adaptive filters. It might be worthwhile to take advantage of that various transmission impairment drift in different time scales and employ the Bayesian filtering framework (e.g., Kalman filter, particle filter) to get more robust channel tracking algorithms.

There are multiple potential research directions regarding Paper C. For instance, there is a lack of mathematical models describing the impact of environmental perturbation on the phase and polarization of the transmitted telecommunication signal. Such models, if found, significantly improve the remote simulation, testing, and verification of future sensing algorithms. It also enables exploiting sophisticated learning algorithms to detect, classify, and predict environmental perturbations. Moreover, the results in Paper C are far from practical scenarios, as we only considered a concatenation of five sections. Another potential future work involves extending the learning model in Paper C to account for higher numbers of sections and elevating the overparameterized cases.

While the capacity bounds on the correlated EO-comb phase noise channel in Paper D are relatively tight in the high-SNR regime, they need to be improved for low SNRs. Novel information theoretic tricks are required to achieve this. Another interesting direction would be identifying scenarios where correlated phase noise could appear and using the framework employed in Paper D to obtain capacity bounds. Moreover, the results in Paper D and E suggest that a practical transmission system employing EO-combs could potentially utilize the inherent phase noise correlation of these systems to design simpler yet more effective DSP algorithms.

References

[1] T. H. Maiman, "Stimulated optical radiation in ruby," *Nature*, vol. 187, no. 4736, pp. 493–494, 1960.

[2] K. C. Kao and G. A. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," in *Proc. Instit. Elect. Eng.*, vol. 113, 1966, pp. 1151–1158.

[3] R. J. Mears, L. Reekie, I. Jauncey, and D. N. Payne, "Low-noise erbium-doped fiber amplifier operating at 1.54 μm," *Electron. Lett.*, vol. 19, no. 23, pp. 1026–1028, 1987.

[4] W. Tomlinson, "Wavelength multiplexing in multimode optical fibers," *Appl. Opt.*, vol. 16, no. 8, pp. 2180–2194, 1977.

[5] D. J. Richardson, J. M. Fini, and L. E. Nelson, "Space-division multiplexing in optical fibres," *Nature photon.*, vol. 7, no. 5, pp. 354–362, 2013.

[6] J. M. Kahn and D. A. Miller, "Communications expands its space," *Nature photon.*, vol. 11, no. 1, pp. 5–8, 2017.

[7] N. K. Fontaine, R. Ryf, H. Chen, *et al.*, "30× 30 MIMO transmission over 15 spatial modes," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2015, Paper Th5C–1.

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

[9] T. Wu, G. Liu, S. Fu, and F. Xing, "Recent progress of fiber-optic sensors for the structural health monitoring of civil infrastructure," *Sensors*, vol. 20, no. 16, 2020, Paper 4517.

[10] L. S. Goh, K. Onodera, M. Kanetsuna, K. Watanabe, and N. Shinomiya, "Constructing an optical fiber sensor network for natural environment remote monitoring," in *Proc. 17th Asia Pacific Conf. Commun.*, 2011, pp. 208–212.

[11] C. Pendão and I. Silva, "Optical fiber sensors and sensing networks: Overview of the main principles and applications," *Sensors*, vol. 22, no. 19, 2022, Paper 7554.

[12] G. Marra, C. Clivati, R. Luckett, *et al.*, "Ultrastable laser interferometry for earthquake detection with terrestrial and submarine cables," *Science*, vol. 361, no. 6401, pp. 486–490, 2018.

[13] Z. Zhan, M. Cantono, V. Kamalov, *et al.*, "Optical polarization-based seismic and water wave sensing on transoceanic cables," *Science*, vol. 371, no. 6532, pp. 931–936, 2021.

[14] M. Cantono, J. C. Castellanos, S. Batthacharya, *et al.*, "Optical network sensing: Opportunities and challenges," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2022, Paper M2F–1.

[15] G. P. Agrawal, *Fiber-optic communication systems*, 4th Ed. John Wiley & Sons, 2012.

[16] B. J. Puttnam, G. Rademacher, and R. S. Luis, "Space-division multiplexing for optical fiber communications," *Optica*, vol. 8, no. 9, pp. 1186–1203, 2021.

[17] K.-P. Ho and H.-W. Cuei, "Generation of arbitrary quadrature signals using one dual-drive modulator," *J. Lightwave Technol.*, vol. 23, no. 2, pp. 764–770, 2005.

[18] G. P. Agrawal, *Nonlinear fiber optics*, 3th Ed. Springer, 2002.

[19] C. R. Menyuk and B. S. Marks, "Interaction of polarization mode dispersion and nonlinearity in optical fiber transmission systems," *J. Lightwave Technol.*, vol. 24, no. 7, pp. 2806–2826, 2006.

[20] D. Marcuse, C. Manyuk, and P. K. A. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," *J. Lightwave Technol.*, vol. 15, no. 9, pp. 1735–1746, 1997.

[21] R. Stolen and E. Ippen, "Raman gain in glass optical waveguides," *Appl. Phys. Lett.*, vol. 22, no. 6, pp. 276–278, 1973.

[22] M. Secondini and E. Forestieri, "Scope and limitations of the nonlinear Shannon limit," *J. Lightwave Technol.*, vol. 35, no. 4, pp. 893–902, 2016.

[23] C. Headley and G. P. Agrawal, *Raman amplification in fiber optical communication systems*. Academic press, 2005.

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

[25] K. Petermann, *Laser diode modulation and noise*. Springer Science & Business Media, 2012, vol. 3.

[26] M. Karlsson, J. Brentel, and P. A. Andrekson, "Long-term measurement of PMD and polarization drift in installed fibers," *J. Lightwave Technol.*, vol. 18, no. 7, pp. 941–951, Jul. 2000.

[27] D. Waddy, P. Lu, L. Chen, and X. Bao, "Fast state of polarization changes in aerial fiber under different climatic conditions," *IEEE Photon. Technol. Lett.*, vol. 13, no. 9, pp. 1035–1037, Sep. 2001.

[28] J. Wuttke, P. Krummrich, and J. Rosch, "Polarization oscillations in aerial fiber caused by wind and power-line current," *IEEE Photon. Technol. Lett.*, vol. 15, no. 6, pp. 882–884, Jun. 2003.

[29] P. M. Krummrich, D. Ronnenberg, W. Schairer, D. Wienold, F. Jenau, and M. Herrmann, "Demanding response time requirements on coherent receivers due to fast polarization rotations caused by lightning events," *Opt. Express*, vol. 24, no. 11, pp. 12 442–12 457, May 2016.

[30] D. Charlton, S. Clarke, D. Doucet, *et al.*, "Field measurements of SOP transients in OPGW, with time and location correlation to lightning strikes," *Opt. Express*, vol. 25, no. 9, pp. 9689–9696, May 2017.

[31] B. T. Kirby, D. E. Jones, and M. Brodsky, "Effect of polarization dependent loss on the quality of transmitted polarization entanglement," *J. Lightwave Technol.*, vol. 37, no. 1, pp. 95–102, 2019.

[32] C. Xie and S. Chandrasekhar, "Two-stage constant modulus algorithm equalizer for singularity free operation and optical performance monitoring in optical coherent receiver," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Mar. 2010, Paper OMK3.

[33] N. Gisin, "Statistics of polarization dependent losses," *Opt. Commun.*, vol. 114, no. 5-6, pp. 399–405, 1995.

[34] A. Mecozzi and M. Shtaif, "The statistics of polarization-dependent loss in optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 14, no. 3, pp. 313–315, 2002.

[35] A. Galtarossa and L. Palmieri, "The exact statistics of polarization-dependent loss in fiber-optic links," *IEEE Photon. Technol. Lett.*, vol. 15, no. 1, pp. 57–59, 2003.

[36] M. Shtaif, "Performance degradation in coherent polarization multiplexed systems as a result of polarization dependent loss," *Opt. Express*, vol. 16, no. 18, pp. 13 918–13 932, 2008.

[37] T. Duthel, C. Fludger, J. Geyer, and C. Schulien, "Impact of polarisation dependent loss on coherent POLMUX-NRZ-DQPSK," in *Proc. Opt. Fiber Commun. (OFC)*, 2008, Paper OThU5.

[38] M. Shtaif and O. Rosenberg, "Polarization-dependent loss as a waveform-distorting mechanism and its effect on fiber-optic systems," *J. Lightwave Technol.*, vol. 23, no. 2, p. 923, 2005.

[39] N. Rossi, P. Serena, and A. Bononi, "Polarization-dependent loss impact on coherent optical systems in presence of fiber nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 26, no. 4, pp. 334–337, 2014.

[40] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, Jul. 2006.

[41] G. Rademacher, R. S. Luís, B. J. Puttnam, *et al.*, "1.53 Peta-bit/s C-band transmission in a 55-mode fiber," in *Proc. Europ. Conf. Opt. Commun. (ECOC)*, 2022, paper Th3C.3.

[42] F. Derr, "Coherent optical QPSK intradyne system: Concept and digital receiver realization," *J. Lightwave Technol.*, vol. 10, no. 9, pp. 1290–1296, 1992.

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

[44] V. Ataie, E. Temprana, L. Liu, *et al.*, "Ultrahigh count coherent WDM channels transmission using optical parametric comb-based frequency synthesizer," *J. Lightwave Technol.*, vol. 33, no. 3, pp. 694–699, 2015.

[45] M. Mazur, A. Lorences-Riesgo, J. Schröder, P. A. Andrekson, and M. Karlsson, "High spectral efficiency PM-128QAM comb-based superchannel transmission enabled by a single shared optical pilot tone," *J. Lightwave Technol.*, vol. 36, no. 6, pp. 1318–1325, 2018.

[46] L. Lundberg, M. Karlsson, A. Lorences-Riesgo, *et al.*, "Frequency comb-based WDM transmission systems enabling joint signal processing," *Appl. Sciences*, vol. 8, no. 5, 2018, Paper 718.

[47] P. Marin-Palomo, J. N. Kemal, M. Karpov, *et al.*, "Microresonator-based solitons for massively parallel coherent optical communications," *Nature*, vol. 546, no. 7657, pp. 274–279, 2017.

[48] J. N. Kemal, J. Pfeifle, P. Marin-Palomo, *et al.*, "Multi-wavelength coherent transmission using an optical frequency comb as a local oscillator," *Opt. Express*, vol. 24, no. 22, pp. 25 432–25 445, 2016.

[49] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379-423/623–656, Jul. 1948.

[50] S. Verdú *et al.*, "A general formula for channel capacity," *IEEE Trans. Info. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[51] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Info. Theory*, vol. 49, no. 10, pp. 2426–2467, 2003.

[52] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 437–453, 2006.

[53] A. Lapidoth, "On phase noise channels at high SNR," in *Proc. IEEE Info. Theory Workshop (ITW)*, 2002.

[54] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Info. Theory*, vol. 46, no. 7, pp. 2315–2328, 2000.

[55] I. Csiszar and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Info. Theory*, vol. 41, no. 1, pp. 35–43, 1995.

[56] J. Y. N. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, Mass. Inst. Technol. (MIT), Cambridge, MA, 1983.

[57] I. Csiszár and J. Korner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Info. Theory*, vol. 27, no. 1, pp. 5–12, 1981.

[58] C. T. Allen, P. K. Kondamuri, D. L. Richards, and D. C. Hague, "Measured temporal and spectral PMD characteristics and their implications for network-level mitigation approaches," *J. Lightwave Technol.*, vol. 21, no. 1, pp. 79–86, Jan. 2003.

[59] K. Kikuchi, "Polarization-demultiplexing algorithm in the digital coherent receiver," in *Proc. IEEE/LEOS Summer Topical Meetings*, 2008, Paper MC2.2.

[60] I. Roudas, A. Vgenis, C. S. Petrou, *et al.*, "Optimal polarization demultiplexing for coherent optical communications systems," *J. Lightwave Technol.*, vol. 28, no. 7, pp. 1121–1134, 2009.

[61] F. Heismann and K. Tokuda, "Polarization-independent electro-optic depolarizer," *Opt. Lett.*, vol. 20, no. 9, pp. 1008–1010, 1995.

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

[63] N. J. Muga and A. N. Pinto, "Adaptive 3-D Stokes space-based polarization demultiplexing algorithm," *J. Lightwave Technol.*, vol. 32, no. 19, pp. 3290–3298, 2014.

[64] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Time varying ISI model for nonlinear interference noise," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, Mar. 2014, Paper W2A.62.

[65] R. Dar, M. Shtaif, and M. Feder, "New bounds on the capacity of the nonlinear fiber-optic channel," *Opt. Lett.*, vol. 39, no. 2, pp. 398–401, Jan. 2014.

[66] C. B. Czegledi, M. Karlsson, E. Agrell, and P. Johannisson, "Polarization drift channel model for coherent fibre-optic systems," *Scientific Reports*, vol. 6, no. 1, 2016, Paper 21217.

[67] J. N. Damask, *Polarization Optics in Telecommunications*. NewYork, NY, USA: Springer, 2005.

[68] R. Noé, B. Koch, D. Sandel, and V. Mirvoda, "Polarization-dependent loss: New definition and measurement techniques," *J. Lightw. Technol.*, vol. 33, no. 10, pp. 2127–2138, 2015.

[69] R. M. Bütler, C. Häger, H. D. Pfister, G. Liga, and A. Alvarado, "Model-based machine learning for joint digital backpropagation and PMD compensation," *J. Lightw. Technol.*, vol. 39, no. 4, pp. 949–959, 2021.

[70] A. J. Metcalf, V. Torres-Company, D. E. Leaird, and A. M. Weiner, "High-power broadly tunable electrooptic frequency comb generator," *IEEE J. Sel. Topics in Quantum Electron.*, vol. 19, no. 6, pp. 231–236, 2013.

[71] A. A. Cuyt, V. Petersen, B. Verdonk, H. Waadeland, and W. B. Jones, *Handbook of continued fractions for special functions*. Springer Science & Business Media, 2008.

[72] A. Ishizawa, T. Nishikawa, A. Mizutori, *et al.*, "Phase-noise characteristics of a 25-GHz-spaced optical frequency comb based on a phase-and intensity-modulated laser," *Opt. Express*, vol. 21, no. 24, pp. 29 186–29 194, 2013.

[73] L. Lundberg, M. Mazur, A. Mirani, *et al.*, "Phase-coherent lightwave communications with frequency combs," *Nature Commun.*, vol. 11, no. 1, pp. 201–212, 2020.

[74] G. Durisi, A. Tarable, C. Camarda, and G. Montorsi, "On the capacity of MIMO Wiener phase-noise channels," in *Proc. Info. Theory Applic. Workshop (ITA)*, 2013.

[75] S. Yang and S. Shamai, "On the multiplexing gain of discrete-time MIMO phase noise channels," *IEEE Trans. Info. Theory*, vol. 63, no. 4, pp. 2394–2408, 2017.

[76] M. Biguesh and A. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Sig. Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[77] S. M. Kay, "Fundamentals of statistical signal processing: Estimation theory," in Englewood Cliffs: NJ: Prentice-Hall, 1993.

[78] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Info. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[79] T. L. Marzetta, "Blast training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. Annu. Allerton Conf. Commun., Control, and Comput.*, Citeseer, vol. 37, 1999, pp. 958–966.

[80] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans.s Sig. Process.*, vol. 51, no. 5, pp. 1351–1366, 2003.

[81] N. Jindal and A. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates," *IEEE Trans. Info. Theory*, vol. 49, no. 11, pp. 2895–2909, 2003.

[82] M. Biguesh and A. B. Gershman, "Downlink channel estimation in cellular systems with antenna arrays at base stations using channel probing with feedback," *EURASIP J. Advanc. Sig. Process.*, no. 9, pp. 1–10, Sep. 2004.

[83] S. Särkkä and L. Svensson, *Bayesian filtering and smoothing*. Cambridge university press, 2023, vol. 17.

[84] S. Qureshi, "Adaptive equalization," *Proc. of the IEEE*, vol. 73, no. 9, pp. 1349–1387, 1985.

[85] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proc. of the IEEE*, vol. 4, no. 63, pp. 719–720, Apr. 1975.

[86] B. Widrow, "Adaptive filters," in *Aspects of Network and System Theory*, R. E. Kalman and N. DeClaris, Eds., New York, NY: Holt, Rinehart and Winston, Apr. 1971, pp. 563–587.

[87] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans Acoust. Speech Sig. Process.*, vol. 33, no. 1, pp. 222–230, 1985.

[88] N. Bershad, "On the optimum gain parameter in LMS adaptation," *IEEE Trans. Acoust. Speech Sig. Process.*, vol. 35, no. 7, pp. 1065–1068, 1987.

[89] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.

[90] K. N. Oh and Y. O. Chin, "Modified constant modulus algorithm: Blind equalization and carrier phase recovery algorithm," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1, 1995, pp. 498–502.

[91] J. Yang, J.-J. Werner, and G. Dumont, "The multimodulus blind equalization and its generalized algorithms," *IEEE J. on Sel. Areas Commun.*, vol. 20, no. 5, pp. 997–1015, Jun. 2002.

[92] K. Kikuchi, "Performance analyses of polarization demultiplexing based on constant-modulus algorithm in digital coherent optical receivers," *Opt. Express*, vol. 19, no. 10, pp. 9868–9880, May 2011.

[93] L. Liu, Z. Tao, W. Yan, S. Oda, T. Hoshida, and J. C. Rasmussen, "Initial tap setup of constant modulus algorithm for polarization de-multiplexing in optical coherent receivers," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2009, Paper OMT2.

[94] M. Ready and R. Gooch, "Blind equalization based on radius directed adaptation," in *Proc. Int. Conf. Acoust. Speech Sig. Processesing (ICASSP)*, vol. 3, Apr. 1990, pp. 1699–1702.

[95] D. Lavery, M. Paskov, R. Maher, S. J. Savory, and P. Bayvel, "Modified radius directed equaliser for high order QAM," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Sep. 2015, Paper We4D5.

[96] R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," *The Bell System Technical Journal*, vol. 45, no. 2, pp. 255–286, 1966.

[97] M. S. Faruk and S. J. Savory, "Digital signal processing for coherent transceivers employing multilevel formats," *J. Lightwave Technol.*, vol. 35, no. 5, pp. 1125–1141, 2017.

[98] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 32, no. 5, pp. 922–923, Sep. 1976.

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

[100] Z. Zheng, N. Cui, H. Xu, *et al.*, "Window-split structured frequency domain Kalman equalization scheme for large PMD and ultra-fast RSOP in an optical coherent PDM-QPSK system," *Opt. Express*, vol. 26, no. 6, pp. 7211–7226, 2018.

[101] H. Louchet, K. Kuzmin, and A. Richter, "Joint carrier-phase and polarization rotation recovery for arbitrary signal constellations," *IEEE Photon. Technol. Lett.*, vol. 26, no. 9, pp. 922–924, May 2014.

[102] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: A unifying theory and numerical methods for characterisation," in *Proc. 35th Annu. Design Automat. Conf.*, 1998, pp. 26–31.

[103] A. Patapoutian, "On phase-locked loops and Kalman filters," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 670–672, 1999.

[104] P.-O. Amblard, J.-M. Brossier, and E. Moisan, "Phase tracking: What do we gain from optimality? Particle filtering versus phase-locked loops," *J. Sig. Process.*, vol. 83, no. 1, pp. 151–167, 2003.

[105] Q. Zou, A. Tarighat, and A. H. Sayed, "Compensation of phase noise in OFDM wireless systems," *IEEE Trans. Sig. Process.*, vol. 55, no. 11, pp. 5407–5424, 2007.

[106] J. Jignesh, B. Corcoran, C. Zhu, and A. Lowery, "Unscented Kalman filters for polarization state tracking and phase noise mitigation," *Opt. Express*, vol. 24, no. 19, pp. 22 282–22 295, 2016.

[107] T. Yang, C. Shi, X. Chen, *et al.*, "Linewidth-tolerant and multi-format carrier phase estimation schemes for coherent optical M-QAM flexible transmission systems," *Opt. Express*, vol. 26, no. 8, pp. 10 599–10 615, 2018.

[108] V. Rozental, D. Kong, B. Corcoran, D. Mello, and A. J. Lowery, "Filtered carrier phase estimator for high-order QAM optical systems," *J. Lightwave Technol.*, vol. 36, no. 14, pp. 2980–2993, 2018.

[109] A. F. Alfredsson, E. Agrell, and H. Wymeersch, "Iterative detection and phase-noise compensation for coded multichannel optical transmission," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5532–5543, 2019.

[110] T. Pfau and R. Noé, "Phase-noise-tolerant two-stage carrier recovery concept for higher order QAM formats," *IEEE J. Sel. Topics in Quantum Electron.*, vol. 16, no. 5, pp. 1210–1216, 2009.

[111] X. Zhou, "An improved feed-forward carrier recovery algorithm for coherent receivers with M-QAM modulation format," *IEEE Photon. Technol. Lett.*, vol. 22, no. 14, pp. 1051–1053, 2010.

[112] J. Li, L. Li, Z. Tao, T. Hoshida, and J. C. Rasmussen, "Laser-linewidth-tolerant feed-forward carrier phase estimator with reduced complexity for QAM," *J. Lightwave Technol.*, vol. 29, no. 16, pp. 2358–2364, 2011.

[113] M. D. Larsen, G. Seco-Granados, and A. L. Swindlehurst, "Pilot optimization for time-delay and channel estimation in OFDM systems," in *Proc. IEEE Int. Conf. on Acoust. Speech Sig. Process. (ICASSP)*, 2011, pp. 3564–3567.

[114] M. Simko, P. S. Diniz, Q. Wang, and M. Rupp, "Adaptive pilot-symbol patterns for MIMO OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4705–4715, 2013.

[115] A. F. Alfredsson, E. Agrell, M. Karlsson, and H. Wymeersch, "Pilot distributions for joint-channel carrier-phase estimation in multichannel optical communications," *J. Lightwave Technol.*, vol. 38, no. 17, pp. 4656–4663, 2020.

[116] A. F. Alfredsson, E. Agrell, H. Wymeersch, and M. Karlsson, "Pilot distributions for phase tracking in space-division multiplexed systems," in *Proc. Europ. Conf. Opt. Commun. (ECOC)*, 2017, Paper P1.SC3.48.

[117] E. Agrell, A. Alfredsson, B. J. Puttnam, R. S. Luís, G. Rademacher, and M. Karlsson, "Modulation and detection for multicore superchannels with correlated phase noise," in *Proc. CLEO: Sci. Innovat.*, 2018, Paper SM4C–3.

[118] R. Narayanaswamy and O. S. Wolfbeis, *Optical sensors: industrial, environmental and diagnostic applications*. Springer, vol. 1.

[119] A. Rogers, "Distributed optical-fibre sensing," *Measurement Science Technol.*, vol. 10, no. 8, R75, 1999.

[120] S. J. Russell, K. R. Brady, and J. P. Dakin, "Real-time location of multiple time-varying strain disturbances, acting over a 40-km fiber section, using a novel dual-Sagnac interferometer," *J. Lightwave Technol.*, vol. 19, no. 2, pp. 205–213, 2001.

[121] Q. Liu, Z. He, and T. Tokunaga, "Sensing the Earth crustal deformation with nano-strain resolution fiber-optic sensors," *Opt. Express*, vol. 23, no. 11, A428–A436, 2015.

[122] M. R. Fernández-Ruiz, M. A. Soto, E. F. Williams, *et al.*, "Distributed acoustic sensing for seismic activity monitoring," *APL Photon.*, vol. 5, no. 3, 2020.

[123] P. Jousset, T. Reinsch, T. Ryberg, *et al.*, "Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features," *Nature Commun.*, vol. 9, no. 1, pp. 2509–2520, 2018.

[124] M. K. Barnoski, M. D. Rourke, S. Jensen, and R. Melville, "Optical time domain reflectometer," *Appl. Opt.*, vol. 16, no. 9, pp. 2375–2379, 1977.

[125] F. Muñoz and M. A. Soto, "Enhancing fibre-optic distributed acoustic sensing capabilities with blind near-field array signal processing," *Nature Commun.*, vol. 13, no. 1, 2016, Paper 4019.

[126] Y. Mizuno, N. Hayashi, H. Fukuda, K. Y. Song, and K. Nakamura, "Ultrahigh-speed distributed Brillouin reflectometry," *Light Sci. Appl.*, vol. 5, no. 12, 2016, Paper e16184.

[127] J. Li and M. Zhang, "Physics and applications of Raman distributed optical fiber sensing," *Light Sci. Appl.*, vol. 11, no. 1, 2022, Paper 128.

[128] K.-I. Aoyama, K. Nakagawa, and T. Itoh, "Optical time domain reflectometry in a single-mode fiber," *IEEE J. Quantum Electron.*, vol. 17, no. 6, pp. 862–868, 1981.

[129] M. A. Soto, J. A. Ramirez, and L. Thévenaz, "Intensifying the response of distributed optical fibre sensors using 2D and 3D image restoration," *Nature Commun.*, vol. 7, no. 1, pp. 1–11, 2016.

[130] D. Zhou, Y. Dong, B. Wang, *et al.*, "Single-shot BOTDA based on an optical chirp chain probe wave for distributed ultrafast measurement," *Light Sci. Appl.*, vol. 7, no. 1, 2018, Paper 32.

[131] L. Qiu, D. Ba, D. Zhou, Q. Chu, Z. Zhu, and Y. Dong, "High-sensitivity dynamic distributed pressure sensing with frequency-scanning φ-OTDR," *Opt. Lett.*, vol. 47, no. 4, pp. 965–968, 2022.

[132] A. Mecozzi, M. Cantono, J. C. Castellanos, V. Kamalov, R. Muller, and Z. Zhan, "Polarization sensing using submarine optical cables," *Optica*, vol. 8, no. 6, pp. 788–795, 2021.

[133] G. Bashan, H. H. Diamandi, Y. London, E. Preter, and A. Zadok, "Optomechanical time-domain reflectometry," *Nature Commun.*, vol. 9, no. 1, p. 2991, 2018.

[134] J. B. Ajo-Franklin, S. Dou, N. J. Lindsey, *et al.*, "Distributed acoustic sensing using dark fiber for near-surface characterization and broadband seismic event detection," *Scientific Reports*, vol. 9, no. 1, 2019, Paper 1328.

[135] F. Huot, E. Martin, Z. Spica, and B. Biondi, "Distributed acoustic sensing (DAS) for large-scale urban monitoring and seismic hazard mitigation using preexisting telecommunication infrastructure," in *Proc. Geophys. Smart City Devel.*, 2019.

[136] Z. Li, Z. Shen, Y. Yang, E. Williams, X. Wang, and Z. Zhan, "Rapid response to the 2019 Ridgecrest earthquake with distributed acoustic sensing," *AGU Advances*, vol. 2, no. 2, e2021AV000395, 2021.

[137] E. F. Williams, M. R. Fernández-Ruiz, R. Magalhaes, *et al.*, "Distributed sensing of microseisms and teleseisms with submarine dark fibers," *Nature Commun.*, vol. 10, no. 1, 2019, Paper 5778.

[138] A. Sladen, D. Rivet, J. P. Ampuero, *et al.*, "Distributed sensing of earthquakes and ocean-solid Earth interactions on seafloor telecom cables," *Nature Commun.*, vol. 10, no. 1, 2019, Paper 5777.

[139] Y.-K. Huang and E. Ip, "Simultaneous optical fiber sensing and mobile front-haul access over a passive optical network," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, 2020, Paper Th1K–4.

[140] W. R. Habel and K. Krebber, "Fiber-optic sensor applications in civil and geotechnical engineering," *Photon. Sensors*, vol. 1, pp. 268–280, 2011.

[141] I. Ashry, Y. Mao, B. Wang, *et al.*, "A review of distributed fiber–optic sensing in the oil and gas industry," *J. Lightwave Technol.*, vol. 40, no. 5, pp. 1407–1431, 2022.

[142] Z. Pan, K. Liang, Q. Ye, H. Cai, R. Qu, and Z. Fang, "Phase-sensitive OTDR system based on digital coherent detection," in *Proc. Asia Commun. Photon. Conf. (ACP)*, 2011, Paper 83110S.

[143] S. Shaheen, K. Hicke, and K. Krebber, "Phase-sensitive optical time domain reflectometry based on geometric phase measurement," *Scientific Reports*, vol. 13, no. 1, 2023, Paper 2682.

[144] N. J. Lindsey, E. R. Martin, D. S. Dreger, *et al.*, "Fiber-optic network observations of earthquake wavefields," *Geophy. Res. Lett.*, vol. 44, no. 23, pp. 11 792–11 799, 2017.

[145] E. Ip, Y.-K. Huang, G. Wellbrock, *et al.*, "Vibration detection and localization using modified digital coherent telecom transponders," *J. Lightwave Technol.*, vol. 40, no. 5, pp. 1472–1482, 2022.

[146] J. E. Simsarian and P. J. Winzer, "Shake before break: Per-span fiber sensing with in-line polarization monitoring," in *Proc. Opt. Fiber Commun. (OFC)*, 2017, Paper M2E.6.

[147] J. C. Castellanos, Z. Zhan, V. Kamalov, *et al.*, "Optical polarization-based sensing and localization of submarine earthquakes," in *Proc. Opt. Fiber Commun. (OFC)*, San Diego, California, 2022.

[148] D. Charlton, S. Clarke, D. Doucet, *et al.*, "Field measurements of SOP transients in OPGW, with time and location correlation to lightning strikes," *Opt. Express*, vol. 25, no. 9, pp. 9689–9696, 2017.

[149] H. He, L. Jiang, Y. Pan, *et al.*, "Integrated sensing and communication in an optical fibre," *Light Sci. Appl.*, vol. 12, no. 1, 2023, Paper 25.