THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Capacity Analysis and Receiver Design in the Presence of Fiber Nonlinearity

Kamran Keykhosravi

CHALMERS
UNIVERSITY OF TECHNOLOGY

Communication Systems Group
Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2019

Capacity Analysis and Receiver Design in the Presence of Fiber Nonlinearity

Kamran Keykhosravi
ISBN 978-91-7905-137-2

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie Nr 4604
ISSN 0346-718X

This thesis has been prepared using LaTeX.

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

Printed by Chalmers Reproservice
Gothenburg, Sweden, May 2019

Abstract

The majority of today's global Internet traffic is conveyed through optical fibers. The ever-increasing data demands have pushed the optical systems to evolve from using regenerators and direct-direction receivers to a coherent multi-wavelength network. Future services like cloud computing and virtual reality will demand more bandwidth, so much so that the so called capacity-crunch is anticipated to happen in near future. Therefore, studying the capacity of the optical system is needed to better understanding and utilizing the existing fiber network.

The characterization of the capacity of the dispersive and nonlinear optical fiber described by the nonlinear Schrödinger equation is an open problem. There are a number of lower bounds on the capacity which are mainly obtained based on the mismatched decoding principle or by analyzing simplified channels. These lower bounds either fall to zero at high powers or saturate. The question whether the fiber-optical capacity has the same behavior as the lower bounds at high power is still open. Indeed, the only known upper bound increases with the power unboundedly.

In this thesis, we first study how the fiber nonlinear distortion is modeled in some simplified channels and what is the influence of the simplifying assumptions on the capacity. To do so, the capacity of three different memoryless simplified models of the fiber-optical channel are studied. The results show that in the high-power regime the capacities of these models grow with different pre-logs, which indicates the profound impact of the simplifying assumptions on the capacity of these channels.

Next, we turn our attention to demodulation and detection processes in the presence of fiber nonlinearity. We study a two-user memoryless network. It is shown that by deploying a nonlinearity-tailored demodulator, the performance improves substantially compared with matched filtering and sampling. In the absence of dispersion, we show that with the new receiver, unlike with matched filtering and sampling, arbitrarily low bit error rates can be achieved. Furthermore, we show via simulations that performance improvements can be obtained also for a low-dispersion fiber.

Then, we study the performance of three different dispersion compensation methods in the presence of inter-channel nonlinear interference. The best performance, in terms of achievable information rate, is obtained by a link with inline per-channel dispersion compensation combined with a receiver that compensates for inter-channel nonlinearities. Finally, the capacity analysis is performed for short-reach noncoherent channel, where the source of nonlinearity is not the fiber but a square-law receiver. Capacity bounds are established in the presence of optical and thermal noises. Using these bounds we show that optical amplification is beneficial at low signal-to-noise ratios (SNRs), and detrimental at high SNRs. We quantify the SNR regime for each case for a wide range of channel parameters.

Keywords: Achievable rate, channel capacity, fiber optics, information theory, nonlinearity mitigation, noncoherent optics, dispersion compensation, perturbation theory.

List of Publications

This thesis is based on the following publications:

[A] **K. Keykhosravi**, G. Durisi, and E. Agrell, "Bounds on the per-sample capacity of zero-dispersion simplified fiber-optical channel models," submitted to *Entropy*, May 2019.

[B] **K. Keykhosravi**, M. Tavana, E. Agrell, and G. Durisi, "Demodulation and detection schemes for a memoryless optical WDM channel," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 2994–3005, Jul. 2018.

[C] **K. Keykhosravi**, M. Secondini, G. Durisi, and E. Agrell, "How to increase the achievable information rate by per-channel dispersion compensation," *J. Lightw. Technol.*, vol. 37, no. 10, pp. 2443–2451, May. 2019.

[D] **K. Keykhosravi**, E. Agrell, M. Secondini, and M. Karlsson, "When to use optical amplification in noncoherent transmission: An information-theoretic approach," submitted to *IEEE Trans. Commun.*, May 2019.

Journal articles by the author, not included in this thesis

[E] **K. Keykhosravi**, H. Rastegarfar, and E. Agrell, "Multicast scheduling of wavelength-tunable, multiqueue optical data center switches," *J. Opt. Commun. Netw.*, vol. 10, no. 4, pp. 353–364, Apr. 2018.

[F] **K. Keykhosravi**, H. Rastegarfar, N. Peyghambarian, and E. Agrell., "Overcoming the switching bottlenecks in wavelength-routing, multicast-enabled architectures," submitted to *J. Lightw. Technol.*, Dec. 2018.

Conference papers by the author, not included in this thesis

[G] **K. Keykhosravi**, H. Rastegarfar, and E. Agrell, "Multicast scheduling for optical data center switches with tunability constraints," in *Proc. IEEE International Conference on Computing, Networking and Communications (ICNC),* Silicon Valley, Jan. 2017, pp. 308–312.

[H] H. Rastegarfar, **K. Keykhosravi**, K. Szczerba, E. Agrell, L. LaComb, and M. Glick, "Optical circuit granularity impact in TCP-dominant hybrid data center networks," in *Proc. IEEE International Conference on Computing, Networking and Communications (ICNC),* Silicon Valley, Jan. 2017, pp. 318–322.

[I] **K. Keykhosravi** and E. Agrell, "A novel demodulation scheme for a memoryless optical interference channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT),* Aachen, Germany, June 2017.

[J] **K. Keykhosravi**, G. Durisi, and E. Agrell, "A tighter upper bound on the capacity of the nondispersive optical fiber channel," in *Proc. European Conference on Optical Communication (ECOC),* Gothenburg, Sweden, Sep. 2017.

[K] H. Rastegarfar, **K. Keykhosravi**, E. Agrell, N. Peyghambarian, "Wavelength reuse for scalable multicasting: A cross-layer perspective," in *Proc. Optical Fiber Communication Conf. (OFC),* San Diego, CA, Mar. 2018, paper W2A.20.

[L] M. Tavana, **K. Keykhosravi**, V. Aref, and E. Agrell, "A low-complexity near-optimal detector for multispan zero-dispersion fiber-optic channels," in *Proc. European Conference on Optical Communication (ECOC),* Roma, Italy, Sep. 2018.

[M] **K. Keykhosravi**, H. Rastegarfar, and E. Agrell, "Scalable interconnection scheme for data center multicast applications," in *Proc. Int. Conf. Photonics in Switching and*

Computing (PSC), Limassol, Cyprus, Sep. 2018.

Acknowledgments

First and foremost, I would like to thank my supervisor Prof. Erik Agrell, who introduced me to the world of optics, and my cosupervisor Prof. Giuseppe Durisi, who helped me to gain a deeper understanding about information theory. I have had many fruitful discussions with them and I have developed many new skills. I am grateful for the time they have spent to guide me and give me valuable feedbacks. I also owe Prof. Marco Secondini a dept of gratitude for his guidance and sharing his vast knowledge with me, during my visit to Pisa. Also, I would like to thank Dr. Houman Rastegarfar for introducing me to the field of optical network and for all the motivation and encouragement I received from him.

Many thanks to Agneta, Natasha, and Rebecca for directing the administrative work. A special thanks goes to my office-mate and friend, Alireza, for his supports and all the good memories. Thanks to all the seniors for creating such a friendly environment in the communication group. I am grateful to be a part of the E2-force group, where I have learned about many different aspects of optical communications from Chayan, Cristian, Li, Alireza, Arni, Andreas, Morteza, and Shen. Also, I am thankful for the guidance that I have received from the FORCE group, specially from Prof. Magnus Karlsson and Prof. Pontus Johannisson. I would also like to thank all of my colleagues, former and current, in communication system group with whom I have had many valuable memories, in particular, Mohamad Ali, Sina, Mohamad Hossein, Yasaman, Nima, Naga, Wei, Behrooz, Canan, Amina, Themis, Rahul, Keerthi, Anver, Jesper, Sven, Chao, Johan, Nil, Gabriel, Christopher, Katharina, Jessica, Erik, Markus, Nill, Chouaib, Roman, and Fredrik. I have used their guidance, experiences, and expertise in many instances during these years. Furthermore, I would like thank all my Iranian friends both in Chalmers and in Pisa, for their companionship in the past four years; in particular, Maryam, Ahad, Abbas, Sadegh, Fatemeh, Parastoo & Abolfazl, Samar & Ramin, Mehrzad & Farzaneh, Ebrahim, Masoud, Bita, Fazele, Anis, Ali, Sepideh, Marziye, Tahmoores, Taha, Zahra, Navid, Fatemeh, Neda, Prastoo, Elnaz, Mona, Pegah, Raheb & Hoda, Nastaran & Saeed, Elham, Amir, Sobhan, Mohammad, Mehdi, Nima, Roham, Nooshin, Aidin & Maryam, and Pantea & Ramin.

It would not have been possible for me to pursue my studies without the support of my parents and my brother, to whom I am indebted for their sacrifices. Last but not least, I would like to thank my loving wife Parisa who was always on my side, supported me, and encouraged me.

Kamran Keykhosravi
Gothenburg, April 2019

Acronyms

AIR:	achievable information rate
AR:	auto-regressive
ASE:	amplified spontaneous emission
AWGN:	additive white Gaussian noise
CDM:	per-channel dispersion-managed
COI:	channel of interest
DC:	dispersion compensation
DCF:	dispersion compensating fiber
DBP:	digital backpropagation
DM:	dispersion-managed
DSP:	digital signal processor
EDFA:	erbium-doped fiber amplifier
FBG:	fiber Bragg grating
FFT:	fast Fourier transform
FWM:	four-wave mixing
iid:	independent and identically distributed
LP:	logarithmic perturbation
LPC:	logarithmic perturbative channel
MAP:	maximum a posteriori
MNC:	memoryless NLS channel
NDM:	non-dispersion-managed
NLS:	nonlinear Schrödinger

OPC:	optical phase conjugation
pdf:	probability distribution function
PSD:	power spectral density
RHS:	right hand side
ROADM:	reconfigurable optical add-drop multiplexers
RP:	regular perturbation
RPC:	regular perturbative channel
SER:	symbol error rate
SMF:	single-mode fiber
SNR:	signal-to-noise ratio
SPM:	self-phase modulation
SSF:	split-step Fourier
WDM:	wavelength-division multiplexing
XPM:	cross-phase modulation

Contents

Part I

Overview

CHAPTER 1

Introduction

1.1 Overview

Today's modern society relies on fast and reliable communication systems that connect the world. The massive amount of data produced every day is transmitted almost exclusively through the optical fibers that connect cities, countries, and continents. The two most significant advantages of optical fibers compared to other communication media are providing enormous bandwidth and having extremely low loss, which allows the fiber to transfer data at high rates to far destinations. Comparing to coaxial cables, the loss is 2–3 orders of magnitude smaller and the bandwidth is about 4000 times larger. Fibers are also immune to the electro-magnetic and radio-wave interference, resistant to lighting strikes, and are lighter and smaller than copper cables.

The optical-fiber communication system was born by manufacturing low-loss fibers in 1970 [1]. The history of the optical communication is divided to three eras [2, Sec. 2]. In the first era, from 1977–1993, on–off keying and direct-detection was used to transmit the data through the fiber. Also, at this stage, to compensate for the fiber loss, regenerators were deployed along the fiber, which recover the data in the electric domain and retransmit it over the fiber. With the introduction of the optical amplifiers in the late 1980s [3] the second era began. Transmitting multiple wavelength-devision multiplexed (WDM) channels over a single fiber became possible as all the channels could be amplified with a single optical amplifier. Finally, we are now in a new phase where using modern digital signal processors makes it possible to move from direct detection to coherent transmission systems.

With all the aforementioned advances, the optical communication systems were able to

cope with the increasing data demand in the past five decades, which is mainly attributed to the exponential growth of the Internet. The call for higher data rates is expected to grow even faster in the future as bandwidth-hungry applications like Internet of things, cloud computing, and virtual reality are beginning to emerge. It is anticipated that the current optical systems will meet their limit in the near future, an incident that is referred to as the capacity crunch.

To avoid the capacity crunch, it is essential to exploit all the potentials of the optical fiber. In today's fiber-optical transmissions, the coding, modulation, and demodulation algorithms are adopted from wireless systems. However, the optical fiber is a nonlinear medium because the intensity of the light changes the refractive index of the fiber core. Since the transmission system is designed for the linear channel, the nonlinearity is treated as noise. This so called nonlinear noise increases with the intensity of the signal such that the performance of the system degrades after an optimal input power, causing the achievable rates with commonly used transmission schemes to have a maximum at a certain power. An interesting open question is whether this peaky behavior of the achievable rate is fundamental or is merely because of deploying suboptimal algorithms. Information-theoretic analysis of the capacity is a standard way to obtain insights about the answer.

After about five decades from the birth of optical fibers and seven decades from the dawn of information theory [4], the characterization of the capacity of the fiber-optical channel remains an open problem. The main question is whether the capacity increases to infinity or is bounded. To answer this question, either an ever-increasing lower bound or a bounded upper bound on the capacity are required. However, no such bounds are established as yet. All the lower bounds either vanish or saturate at high powers where the only known upper bound [5] increases indefinitely with power. To obtain approximations of the capacity, many simplified models have been developed and bounds were established on their capacity (see [6, 7] for two recent reviews).

1.2 Thesis Outline

This thesis revolves around the influence of nonlinear effects on the performance of fiber-optical communication systems and methods to compensate for these effects. It is divided into two parts, where the first one includes some background and general information about the topic and the second comprises four papers. The rest of the first part is organized as follows. First, we study the nonlinear Schrödinger (NLS) equation that governs the propagation of a signal through a fiber in Chapter 2. The main goal of Chapter 2 is to describe the continuous-time optical channels to justify the models used in the appended papers. Also, a detailed description of the split-step Fourier (SSF) method, which is a common tool to simulate the fiber-optical channel is provided. In Chapter 3 different notions of capacity are investigated for continuous- and discrete-time channels, and some information-theoretic tools to bound the capacity are provided. Some

optical channel models, including the channel models used in the papers, are studied in Chapter 4. A review on the nonlinearity and dispersion compensation techniques is provided in Chapter 4.7.2. Finally, a summary of the contributions of the papers is provided in Section 6.

In Part II, we first investigate the accuracy of some existing channels that model the fiber nonlinearity (Paper A). We study the influence of some simplifying assumptions, used in developing simple optical channels, on capacity. We show that these assumptions have a profound influence on the capacity of optical models at high powers, and therefore the capacity results based on these models are not accurate in the high-power regime.

In Paper B, the matched filtering and sampling (MFS) which is the common demodulation method in optical systems is revisited. Multiple demodulation and detection schemes are proposed for a two-user memoryless WDM network. A significant gain in performance was observed with the optimal demodulation method, with which the symbol error rate (SER) decreases to zero with power, whereas with MFS the SER goes to approximately one in the high-power regime. Furthermore, the SER is evaluated for dispersive fibers via simulation, where some performance gain are obtained.

In Paper C, we consider the inter-channel nonlinear interference in WDM systems. We compare the influence of these distortions on the performance of three optical links: dispersion-managed (DM), non-dispersion-managed (NDM), and per-channel dispersion-managed (CDM). It was shown that CDM links obtain the best performance if the effects of inter-channel interference is compensated for at the receiver.

Finally, Paper D studies the capacity of noncoherent systems, where nonlinearity originates from the square-law receiver. We consider a more general model than previous studies by taking into account both optical and thermal noises. By using our capacity bounds we determine the power regimes where optical amplification is beneficial or detrimental to the capacity.

1.3 Notation

The following notation is used in this thesis. The complex zero-mean circularly symmetric Gaussian distribution is denoted by $\mathcal{CN}(0,\sigma^2)$, where σ^2 is the variance; also the real-valued normal distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu,\sigma^2)$. All logarithms are in base two. Vectors are denoted by underlined letters. The kth element of a vector (e.g., $\underline{x}$) is indicated by subscript k (x_k) and a vector that is composed of $x_1,\ldots,x_k$ is indicated by the superscript k ($\underline{x}^k$). The Euclidean norm is denoted by $\|\cdot\|$. We use boldface letters to show random quantities. The inner product between two complex functions $f(t)$ and $g(t)$ is defined as $\langle f,g\rangle = \int_{-\infty}^{\infty} f(t)g^*(t)\,\mathrm{d}t$, where $(\cdot)^*$ denotes complex conjugation. $H(\cdot)$ and $h(\cdot)$ represent the entropy and the differential entropy functions, respectively. The mutual information between two random variables $\mathbf{x}$ and $\mathbf{y}$ is denoted by $I(\mathbf{x};\mathbf{y})$. The imaginary unit is represented by $j=\sqrt{-1}$. We show with $p_{\mathbf{x}}$ the probability distribution (mass) function of the continuous (discrete)

random variable $\mathbf{x}$. Sets are indicated by upper-case script letters. For example, we indicate the alphabet set of a random variable (e.g., $\mathbf{x}$) by its corresponding upper-case script letter ($\mathcal{X}$). We use $\mathcal{R}(x)$, $|x|$, and $\angle x$ to denote the real part, the absolute value, and the phase of a complex number x. In the context of matrix operations, $|\cdot|$ and $(\cdot)^T$ represent the determinant and the transpose operators, respectively. The real line and the complex plane are represented by $\mathbb{R}$ and $\mathbb{C}$, respectively, and $*$ indicates the convolution operator. Finally, for two functions $q(x)$ and $r(x)$, we write $q(x) = \mathrm{o}(r(x))$ if $\limsup_{x\to 0} |q(x)/r(x)| = 0$ and we write $q(x) = \mathcal{O}(r(x))$ if $\limsup_{x\to 0} |q(x)/r(x)| < \infty$. All logarithms are in base two.

Notation inconsistencies:

Here we list the inconsistencies between the aforementioned notation and the notation used in some of the appended papers.

- In Paper B and D the probability mass function of a discrete random variable $\mathbf{x}$ at x is denoted by $\Pr(\mathbf{x} = x)$. Also, in Paper B the probability density function of a continuous random variable $\mathbf{x}$ at x is represented by $f_{\mathbf{x}}(x)$.
- In Paper C the random variables are not bold.

CHAPTER 2

Fiber-Optical Transmission Systems

The purpose of this chapter is to connect the discrete-time optical channel models used in the appended papers to the underlying physical channel. Also, it provides the reader with some general information about optical channel modeling. Furthermore, a comprehensive description of the SSF method, which is a standard tool to simulate the fiber-optical channel is included.

Optical core networks, which are information highways connecting cities, countries, and continents, carry a vast amount of Internet data. To connect two nodes via these networks, reconfigurable optical add-drop multiplexers (ROADMs) are used to add or drop a channel at each switching point. Fig. 2.1 shows a simple schematic of how WDM channels are added and dropped at different nodes of the network. Therefore, through propagation of a signal, the channels on neighboring wavelengths, which introduce interference to the signal, may change at each switch.

The standard single-mode fiber (SMF) has an extremely low loss. However, for long transmission distances (> 100 km), it should be compensated for. Optical amplifiers boost the signal energy while adding noise to it. Apart from the loss and noise, there are two main impairments associated with fibers, namely, dispersion and nonlinearity. They are mainly caused by the dependence of the refractive index on wavelength and electromagnetic field, respectively.

This chapter begins by describing the signal propagation through the fiber in Section 2.1. The effects of dispersion, nonlinearity, and loss will be studied separately in Sections 2.1.1, 2.1.2, and 2.1.3. The effects of optical amplification is added to the model in Section 2.2. Furthermore, in Section 2.3.1 nonlinearity and loss are investigated together. A comprehensive description of the SSF method is presented in Section 2.3.

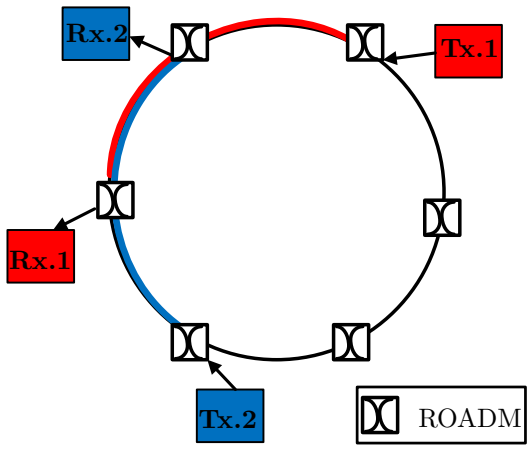

Figure 2.1: An schematic of an optical network, where switching is performed by ROADMs.

Signal spectral broadening is discussed in Section 2.4. Finally, in Section 2.5, WDM systems are described briefly.

2.1 Signal Propagation in Fibers

An optical signal can be described by a slowly varying envelope of the light's electromagnetic field. Therefore, its propagation through the fiber is governed by Maxwell's equations. Under some assumptions, which are discussed later, these equations lead to the Manakov equation [8]

$$\frac{\partial \underline{\mathbf{a}}}{\partial z} + \underbrace{j\frac{\beta_2}{2}\frac{\partial^2 \underline{\mathbf{a}}}{\partial t^2}}_{\text{Dispersion}} - \underbrace{j\gamma\frac{8}{9}\left\|\underline{\mathbf{a}}\right\|^2 \underline{\mathbf{a}}}_{\text{Nonlinearity}} + \underbrace{\frac{\alpha}{2}\underline{\mathbf{a}}}_{\text{Attenuation}} = 0. \tag{2.1}$$

Here, $\underline{\mathbf{a}} = [\mathbf{a}_x(z,t), \mathbf{a}_y(z,t)]$, where $\mathbf{a}_x(z,t)$ and $\mathbf{a}_y(z,t)$ are complex baseband signals propagating in the x and y polarizations at time t and location z, respectively. β_2 is the group-velocity dispersion parameter and α is the attenuation constant. The parameter $\gamma = 2\pi n_2/(\lambda A_{\text{eff}})$ is the nonlinear coefficient, where n_2 is the nonlinear refractive index, A_{eff} is the effective area of the fiber core, and λ is the wavelength in vacuum.

We note that (2.1) describes the signal propagation in a SMF without any amplification and hence without noise. The effects of optical amplifiers are added to this equation in Section 2.2. If the signal is transmitted in only one polarization, (2.1) changes to the NLS equation by replacing the random vector $\underline{\mathbf{a}}$ by the random variable $\mathbf{a}$ and also the

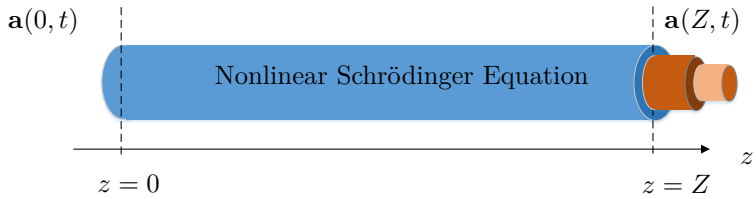

Figure 2.2: An schematic representation of signal propagation through the optical fiber governed by NLS equation.

vector norm $\|\cdot\|$ by the absolute value $|\cdot|$. As we solely consider single polarization in this thesis, we use the NLS equation to describe the signal propagation in fibers. It reads

$$\frac{\partial \mathbf{a}}{\partial z} + j\frac{\beta_2}{2}\frac{\partial^2 \mathbf{a}}{\partial t^2} - j\gamma|\mathbf{a}|^2\mathbf{a} + \frac{\alpha}{2}\mathbf{a} = 0. \tag{2.2}$$

We note that the NLS equation is an excellent model for fibers and its validity has been checked by many experiments. However, it also should be noted that many assumptions have been considered in the derivation of this equation; here we list the most important ones.

- The signal's spectral width Δf is much smaller than the central frequency f_0:[1] $\Delta f / f_0 \ll 1$.
- The wavelength region $0.5 - 2\mu$m is utilized for transmission.
- The fiber-loss is low.
- The refractive index of the core and cladding is invariant of spatial coordinates.
- The nonlinear response is instantaneous.
- The higher order nonlinearities and dispersion are weak.

Therefore, the NLS equation may be inaccurate at very high powers or at very low or high wavelengths. Also, it does not cover all the impairments of the fiber. More information about these assumption can be found in [9, Ch. 2].

For the general case of NLS equation, no analytical solution has been found as yet. However, some numerical methods can be used to solve this equation with arbitrary accuracy. We discuss one of these methods in detail in Section 2.3.

2.1.1 Chromatic Dispersion

In a SMF, each frequency component of a signal propagates with a different velocity through the fiber. Therefore, some components of the signal arrive at the receiver sooner than others, which results in the broadening of the signal in time. Specifically, a pulse

[1] The relation between signal frequency f_0 and wavelength λ is $f_0 = c/\lambda$, where c is the speed of light.

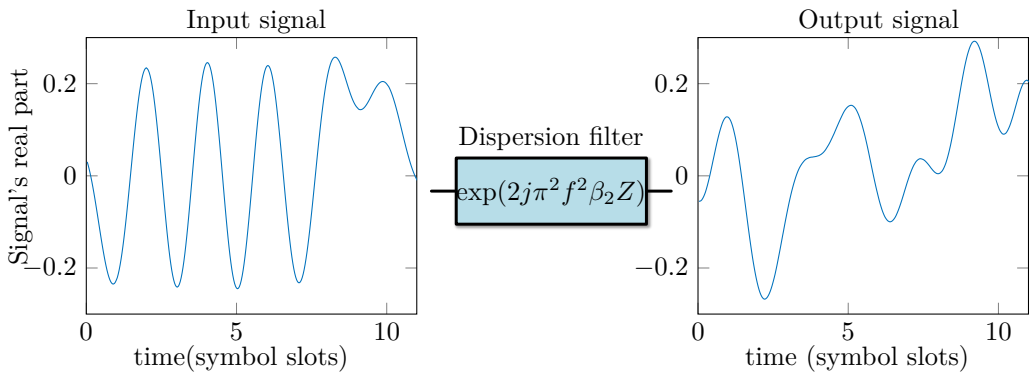

Figure 2.3: Effect of dispersion on a signal after a distance $Z > 0$.

with spectral width Δf broadens in time after propagating through a fiber with length Z by approximately $2\pi Z\beta_2\Delta f$.

The effect of the dispersion on the transmitted signal can be studied by neglecting the nonlinearity ($\gamma = 0$) and the attenuation ($\alpha = 0$). Doing so, we obtain from (2.2)

$$\frac{\partial \mathbf{a}}{\partial z} + j\frac{\beta_2}{2}\frac{\partial^2 \mathbf{a}}{\partial t^2} = 0. \tag{2.3}$$

To solve (2.3), one can apply the Fourier transform to both sides of the equality to obtain

$$\frac{\partial \mathbf{a}(z,f)}{\partial z} = 2j\pi^2 f^2 \beta_2 \mathbf{a}(z,f). \tag{2.4}$$

Solving (2.4), the signal at position z can be expressed in the frequency domain as

$$\mathbf{a}(z,f) = \mathbf{a}(0,f)e^{2j\pi^2 f^2 \beta_2 z}. \tag{2.5}$$

Therefore, the chromatic dispersion (CD) can be modeled by a linear all-pass filter with frequency response $H(z,f) = \exp(2j\pi^2 f^2 \beta_2 z)$. The corresponding impulse response is

$$h(z,t) = \frac{\exp(jt^2/(2\beta^2 z))}{\sqrt{2j\pi\beta_2 z}}. \tag{2.6}$$

If not compensated for, dispersion deteriorates severely the performance of the optical systems. Fig. 2.3 illustrates the effect of dispersion on a signal after propagation. It can be seen that the signal becomes completely distorted. Two categories of dispersion compensation (DC) techniques exist. The first one is to compensate the dispersion in the optical domain and the second is by utilizing digital signal processors at the receiver in the electrical domain. We will elaborate on these techniques in Section 5.1.

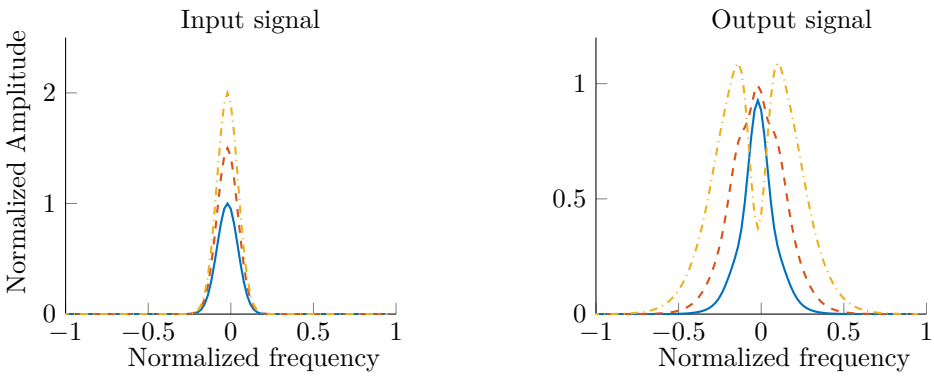

Figure 2.4: Effect of nonlinearity on three pulses with different amplitudes in the frequency domain after propagation.

2.1.2 Kerr Nonlinearity

The presence of Kerr nonlinearity is the main difference between the optical communication channels and the linear wireless ones. Its origin comes from the dependence of the refractive index on the electromagnetic field. Since the fiber core has a very small area, the intensity of the optical power launched into the fiber becomes very high. This high intensity changes the physical features of the glass like its refractive index. The effects of Kerr nonlinearity can be studied by neglecting dispersion $\beta_2 = 0$ and loss $\alpha = 0$ in (2.2) to obtain

$$\frac{\partial \mathbf{a}}{\partial z} = j\gamma|\mathbf{a}|^2\mathbf{a}. \tag{2.7}$$

This equation has the solution

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

Therefore, the nonlinearity appears as a phase distortion that is a function of the signal power. The effects of the nonlinearity distortion can be studied in the frequency domain. Fig. 2.4 shows the evolution of three Gaussian pulses in the frequency domain after propagation. It can be seen that the spectrum of the pulse broadens and this spectrum broadening increases with the amplitude of the signal. Many methods for compensating nonlinearity exist, some of which are presented in Section 5.2.

2.1.3 Fiber Loss

The fiber core is an extremely transparent glass. Its loss is approximately 0.2 dB/km at wavelength 1.55μm which is mainly due to a fundamental loss mechanism, Rayleigh scattering, and exists in all fibers [10, Sec. 2.5.3]. Neglecting dispersion and nonlinearity,

the solution to (2.2) becomes

$$\mathbf{a}(z,t) = \mathbf{a}(0,t)\exp(-\alpha z/2) \tag{2.9}$$

which indicates that the power decreases with the rate $\exp(-\alpha z)$. The power loss is often stated in terms of dB/km which can be calculated from α by $\alpha_{\mathrm{dB}} = 10\alpha \log_{10}(e)$.

2.1.4 Beyond the NLS Equation

As mentioned before, although the NLS equation covers the important features of signal propagation over fiber, it does not describe all the impairments. Here, we briefly list some of these additional phenomena.

- *Stimulated Raman scattering* is an important nonlinear effect that occurs at high powers. It transfers a portion of an optical field's energy to a lower frequency. This feature of the fiber is utilized to build distributed Raman amplifiers, where an optical beam is transmitted at higher frequencies than the signal to boost its energy continuously during propagation. This effect can also severely deteriorate the performance of WDM systems in the high-power regime by introducing inter-channel interference.
- *Stimulated Brillouin scattering* is a nonlinear process similar to the Raman effect. In this process some of the energy of an optical field is transferred to lower frequencies and propagates forward or backward through the fiber. The frequency range affected by Brillouin scattering is much smaller than that of Raman scattering [11, Sec. 5.2.3].

2.2 Optical Amplifications

For long-haul transmission, in order to compensate for the loss, the signal needs to be amplified. Traditionally, this process was performed by installing costly regenerators that brought the optical signal to the electrical domain and then retransmitted it. Now this task is done by optical amplifiers, which are more economical. Moreover, optical amplification is the key enabling technology to WDM systems. A single amplifier can boost the energy of optical signals in a wide spectrum of wavelengths, while with the old systems, one regenerator was needed for each WDM channel. The optical amplification is based on *stimulated emission*, where a copy of a photon is generated by an excited atom. This process is always accompanied with *spontaneous emission*, where an excited atom emits a photon randomly. Therefore, optical amplification always comes at the price of increased noise level. By considering the optical amplification, the signal propagation through the fiber can be described by the statistic NLS equation as

$$\frac{\partial \mathbf{a}}{\partial z} + j\frac{\beta_2}{2}\frac{\partial^2 \mathbf{a}}{\partial t^2} - j\gamma|\mathbf{a}|^2\mathbf{a} + \frac{\alpha - g(z)}{2}\mathbf{a} = \mathbf{n}. \tag{2.10}$$

Table 2.1: Amplifier parameters

Parameter	Symbol	Value
Amplifier Gain	G	$e^{\alpha Z_{\text{sp}}}$
Span length	Z_{sp}	80–100 km
Noise figure	F_n	4–7 dB
Spontaneous emission factor	n_{sp}	1–2.5
Number of spans	K	≥ 1
Planck's constant	h	$6.626 \cdot 10^{-34}$ Js
Carrier frequency	ν	$1.911 \cdot 10^{14} - 1.961 \cdot 10^{14}$ Hz

Here, $\mathbf{n}$ is the amplified spontaneous emission (ASE) noise, which is circularly symmetric white Gaussian, and $g(z)$ is the gain profile. Next, we present two commonly used amplification schemes and study the noise and $g(z)$ for them.

2.2.1 Lumped Amplifiers

In this amplification scheme the fiber is divided into multiple spans and an erbium-doped fiber amplifier (EDFA) is used at the end of each span to boost the signal's energy. The amplification noise, $\mathbf{n}(z,t)$ has the autocorrelation of

$$\mathrm{E}[\mathbf{n}(z,t)\mathbf{n}(z',t')^*] = (G-1)n_{sp}h\nu\delta(t-t')\delta(z-z')\sum_{k=1}^{K}\delta(z-kZ_{\text{sp}}). \tag{2.11}$$

Here, $\delta(\cdot)$ is the Dirac delta function, $h\nu$ is the optical photon energy, and n_{sp} is the spontaneous emission factor. It is also common to report the noise figure $F_n = (1 + 2n_{sp}(G-1))/G$ instead of n_{sp} [12, Eq. (7.2.15)]. The rest of the parameters are presented in Table 2.1 with their typical range of values. The gain profile can be described as

$$g(z) = \alpha Z_{\text{sp}}\sum_{k=1}^{K}\delta(z-kZ_{\text{sp}}). \tag{2.12}$$

If we assume that the signal bandwidth, W, is constant during propagation, the total noise variance is

$$P_N = 2K(G-1)n_{sp}h\nu W. \tag{2.13}$$

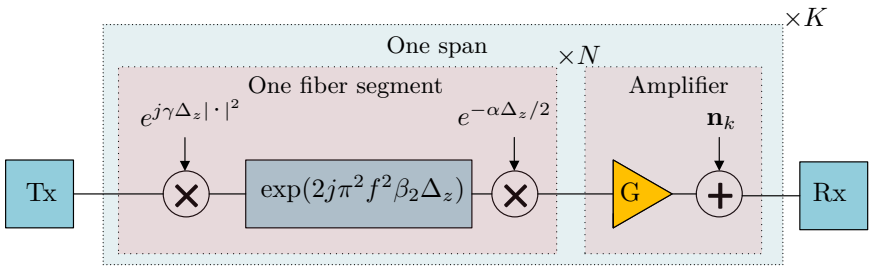

Figure 2.5: Split-step Fourier method for lumped amplification with K spans. Each fiber span is divided to N segments. $G = e^{\alpha Z_{\text{sp}}}$ is the amplifier power gain.

2.2.2 Distributed Amplifiers

Unlike lumped amplification, in the distributed one the transmission loss is continuously compensated for along propagation. Therefore, the signal power level remains almost constant. In this scheme, a pump wave is transmitted at higher frequencies than that of the signal. The pump wave copropagates with the signal and through the Raman effect, it gives the signal a portion of its energy. The gain profile for the ideal Raman amplification is $g(z) = \alpha$ and the noise autocorrelation is

$$\mathrm{E}[\mathbf{n}(z,t)\mathbf{n}(z',t')^*] = \alpha n_{sp} h\nu\delta(z-z')\delta(t-t'). \tag{2.14}$$

The total noise variance is

$$P_N = 2\alpha Z n_{sp} h\nu W \tag{2.15}$$

where Z is the total fiber length. We note that, in practice, the gain profile is not a constant and it decreases as the distance from the amplifier increases.

2.3 Split-Step Fourier Method

For special classes of input signals, such as solitons [9, Ch. 5], the lossless NLS equation ((2.2) with $\alpha = 0$) has an analytical solution. However, a general solution for the NLS equation in (2.2) has not been found. The evolution of the signal can be tracked by means of numerical methods, among which the most famous one is the SSF method.

2.3.1 Lumped Amplification

In the SSF method for lumped amplified systems, a fiber span with length Z_{sp} is split into N segments with length $\Delta z = Z_{\text{sp}}/N$. It is assumed that the linear and nonlinear operators in the NLS equation act independently in each segment. The output of the nth segment, $\mathbf{a}_n = \mathbf{a}(n\Delta z, t)$, can be obtained by applying three operators to the output

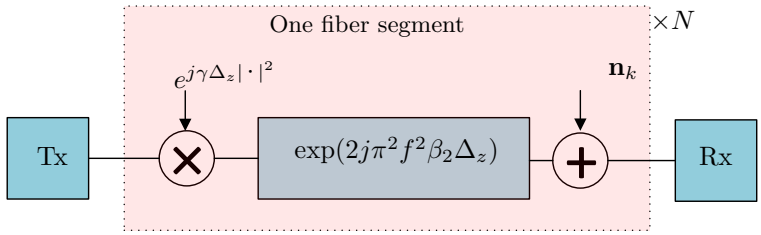

Figure 2.6: Split-step Fourier method for distributed amplification with N segments.

of its previous segment $\mathbf{a}_{n-1}$: *i)* a nonlinear operator as in (2.8) *ii)* a linear operator as in (2.5), and *iii)* an attenuation operator as in (2.9), i.e.,

$$\mathbf{a}_n = e^{-\alpha\Delta z/2}\left[h(\Delta z, t) * \left[\mathbf{a}_{n-1}\exp(j\gamma\Delta z|\mathbf{a}_{n-1}|^2)\right]\right]. \tag{2.16}$$

Here, $h(z,t)$ is the impulse response of the dispersion filter defined in (2.6).

In practice, to implement the SSF method, the input signal is sampled much faster than the Nyquist rate so that the effects of spectrum broadening, caused by the channel nonlinearity, is correctly captured. Moreover, to reduce the complexity of the method, the convolution in (2.16) is calculated in the frequency domain using fast Fourier transform (FFT).

For lumped amplification, the effect of the amplifier can be captured by multiplying the signal by $\sqrt{G} = \exp(\alpha Z_{\text{sp}}/2)$ and adding the ASE noise to the signal. A model for the SSF method for lumped amplification is provided in Fig. 2.5.

Next, we focus on the effects of nonlinearity and loss. If dispersion is ignored, we have $|\mathbf{a}_n|^2 = \exp(-n\alpha\Delta z)|\mathbf{a}_0|$. Therefore, the output of the first span can be described by

$$\mathbf{a}_N = \mathbf{a}_0 e^{-\alpha Z_{\text{sp}}/2}\prod_{n=0}^{N-1}\exp\left(j\gamma\Delta z e^{-n\alpha\Delta z}|\mathbf{a}_0|^2\right) \tag{2.17}$$

$$= \mathbf{a}_0 e^{-\alpha Z_{\text{sp}}/2}\exp\left(j\gamma|\mathbf{a}_0|^2\Delta z\frac{1-e^{-\alpha Z_{\text{sp}}}}{1-e^{-\alpha\Delta z}}\right). \tag{2.18}$$

Using $\lim_{\Delta z\to 0}\Delta z/(1-\exp(-\alpha\Delta z)) = 1/\alpha$, the right-hand side (RHS) of (2.18) in the limit $\Delta z \to 0$ becomes

$$\mathbf{a}_0 e^{-\alpha Z_{\text{sp}}/2}\exp(j\gamma Z_{\text{eff}}|\mathbf{a}_0|^2) \tag{2.19}$$

where Z_{eff} is the effective length of the fiber with length Z_{sp} and is defined as

$$Z_{\text{eff}} = \frac{1-\exp(-\alpha Z_{\text{sp}})}{\alpha}. \tag{2.20}$$

Note that $Z_{\text{eff}} \leq z$, which indicates that the nonlinear phase shift is smaller than in the lossless case in (2.8). A channel model similar to the one in (2.19) is obtained in Section 4.3 and is studied in Paper A.

2.3.2 Distributed Amplification

For ideal distributed amplification, the fiber loss is compensated for completely and ASE noise is added to the signal continuously during propagation. One step of the SSF method with distributed amplification can be described as

$$\mathbf{a}_{n+1} = h(\Delta z, t) * [\mathbf{a}_n \exp(j\gamma \Delta z |\mathbf{a}_n|^2)] + \mathbf{n}_n \tag{2.21}$$

where $\mathbf{n}_n$ is the amplifier noise and has the autocorrelation of

$$\mathrm{E}[\mathbf{n}_n(t)\mathbf{n}_{n'}(t')^*] = \frac{\alpha n_{sp} h\nu}{N} \delta(n-n')\delta(t-t') \tag{2.22}$$

where N is the number of SSF steps. A model for the SSF method for distributed amplification is illustrated in Fig. 2.6.

2.3.3 A Cookbook for the SSF Method

Although (2.16) and (2.21) describe the SSF method completely, they do not give a straightforward description of how to implement this method. In this section, we provide step-by-step algorithms to implement the SSF method for lumped and distributed amplifications (with and without dispersion compensating fiber (DCF)).

Algorithm 1: SSF method for lumped amplification

Data: Input sample vector $\underline{x}_{\text{in}}$ of length L; sampling time Δ_t; fiber parameters: nonlinearity γ, dispersion β_2, attenuation α, span length Z_{sp}, number of spans K; SSF segment length Δ_z ($N = Z_{\text{sp}}/\Delta_z$ should be an integer); noise parameters: F_n, h, and ν as described in Table 2.1.

$\underline{x} \leftarrow \underline{x}_{\text{in}}$
$P_N \leftarrow (e^{\alpha Z_{\text{sp}}} - 1) n_{sp} h\nu / \Delta_t$
$d_\ell \leftarrow \exp\Big(j\Delta_z 2\beta_2 \left(\pi/\left(L\Delta_t\right)\right)^2 \left(L/2 - |L/2 - \ell|\right)^2\Big)$ for $\ell = 0, 1, \ldots L-1$
for $k \in \{1, 2, \ldots, K\}$ **do**
  for $n \in \{1, 2, \ldots, N\}$ **do**
    $x_\ell \leftarrow x_\ell \exp(j\gamma\Delta_z |x_\ell|^2)$ for $\ell \in \{0, 1, \ldots, L-1\}$
    $\tilde{\underline{x}} \leftarrow \text{FFT}(\underline{x})$
    $\tilde{x}_\ell \leftarrow \tilde{x}_\ell d_\ell$ for $\ell \in \{0, 1, \ldots, L-1\}$
    $\underline{x} \leftarrow \text{iFFT}(\tilde{\underline{x}})$
    $x_\ell \leftarrow x_\ell \exp(-\alpha\Delta_z/2)$ for $\ell \in \{0, 1, \ldots, L-1\}$
  if *DCF is used* **then**
    $\tilde{\underline{x}} \leftarrow \text{FFT}(\underline{x})$
    $\tilde{x}_\ell \leftarrow \tilde{x}_\ell / (d_\ell)^N$ for $\ell \in \{0, 1, \ldots, L-1\}$
    $\underline{x} \leftarrow \text{iFFT}(\tilde{\underline{x}})$
  $x_\ell \leftarrow x_\ell e^{\alpha Z_{\text{sp}}/2}$ for $\ell \in \{0, 1, \ldots, L-1\}$
  Draw L iid samples from $\mathcal{CN}(0, P_N)$ to generate z_ℓ for $\ell \in \{0, 1, \ldots, L-1\}$
  $x_\ell \leftarrow x_\ell + z_\ell$ for $\ell \in \{0, 1, \ldots, L-1\}$
if *Electronic DC is used* **then**
  $\tilde{\underline{x}} \leftarrow \text{FFT}(\underline{x})$
  $\tilde{x}_\ell \leftarrow \tilde{x}_\ell / d_\ell^{KN}$ for $\ell \in \{0, 1, \ldots, L-1\}$
  $\underline{x} \leftarrow \text{iFFT}(\tilde{\underline{x}})$
return $\underline{x}$

The FFT function $\tilde{\underline{x}} = \text{FFT}(\underline{x})$ is used to calculate the discrete-Fourier transform, which is defined as

$$\tilde{x}_\ell = \sum_{m=0}^{L-1} x_m \exp(-2j\pi\ell m/L) \text{ for } \ell \in \{0, 1, \ldots, L-1\} \tag{2.23}$$

and its corresponding inverse transform, iFFT, is

$$x_\ell = 1/L \sum_{m=0}^{L-1} \tilde{x}_m \exp(2j\pi\ell m/L) \text{ for } \ell \in \{0,1,\ldots,L-1\}. \tag{2.24}$$

In Algorithm 1, we neglect the nonlinearity and the loss of DCF, whose effects are typically negligible. Next, we present the SSF algorithm for ideal distributed amplification.

Algorithm 2: SSF method for distributed amplification

Data: Input sample vector $\underline{x}_{\text{in}}$ of length L; sampling time Δ_t; fiber parameters: nonlinearity γ, dispersion β_2, attenuation α; total length Z; SSF segment length Δ_z ($N = Z/\Delta_z$ should be an integer); noise parameters: n_{sp}, h and ν as described in Section 2.2.2.

$\underline{x} \leftarrow \underline{x}_{\text{in}}$
$P_N \leftarrow \alpha\Delta_z n_{sp} h\nu/\Delta_t$
$d_\ell \leftarrow \exp\left(j\Delta_z(2\beta_2)\left(\pi/\left(L\Delta_t\right)\right)^2\left(L/2-|L/2-\ell|\right)^2\right)$ for $\ell = 0,1,\ldots L-1$
for $n \in \{1,2,\ldots,N\}$ **do**
  $x_\ell \leftarrow x_\ell \exp(j\gamma\Delta_z|x_\ell|^2)$ for $\ell \in \{0,1,\ldots,L-1\}$
  $\underline{\tilde{x}} \leftarrow \text{FFT}(\underline{x})$
  $\tilde{x}_\ell \leftarrow \tilde{x}_\ell d_\ell$ for $\ell \in \{0,1,\ldots,L-1\}$
  $\underline{x} \leftarrow \text{iFFT}(\underline{\tilde{x}})$
  Draw L iid samples from $\mathcal{CN}(0,P_N)$ to generate z_ℓ for $\ell \in \{0,1,\ldots,L-1\}$
  $x_\ell \leftarrow x_\ell + z_\ell$ for $\ell \in \{0,1,\ldots,L-1\}$
if *Electronic DC is used* **then**
  $\underline{\tilde{x}} \leftarrow \text{FFT}(\underline{x})$
  $\tilde{x}_\ell \leftarrow \tilde{x}_\ell/d_\ell^N$ for $\ell \in \{0,1,\ldots,L-1\}$
  $\underline{x} \leftarrow \text{iFFT}(\underline{\tilde{x}})$
return $\underline{x}$

Both Algorithm 1 and Algorithm 2 can be adapted to simulate polarization-multiplexed systems. To do so, all the steps should be performed for both the x polarization and for the y polarization. Furthermore, the nonlinear operations (Step 6 in Algorithm 1 and Step 5 in Algorithm 2) should be replaced with

$$x_\ell \leftarrow x_\ell \exp(j\gamma\Delta_z(|x_\ell|^2 + |y_\ell|^2)) \text{ for } \ell \in \{0,1,\ldots,L-1\} \tag{2.25}$$

where y_ℓ denotes the ℓth sample of the signal propagated over the y polarization.

2.3.4 Precision of the SSF Method

As was mentioned, the SSF method is a standard approach to approximate the solution of the NLS equation. In this method, an optical fiber of length Z is modeled by a cascade of N segments of length $\Delta_z = Z/N$. The input to the SSF method is a complex vector

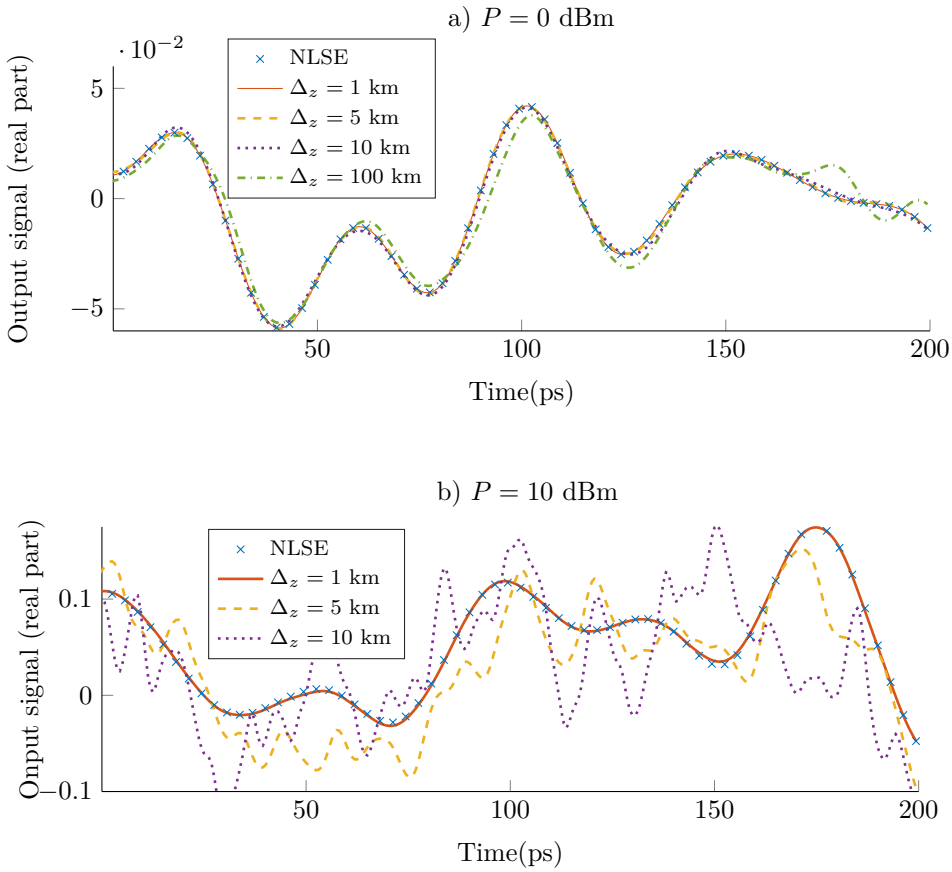

Figure 2.7: Output of the SSF method for different values of Δ_Z for a) $P = 0$ dBm and b) $P = 10$ dBm.

of length L containing samples taken every Δ_t seconds from the baseband input signal of duration $T = L\Delta_t$. The accuracy of the SSF method depends on the value of Δ_t and Δ_z. The approximation obtained by the SSF method gets precise in the limits $\Delta_t \to 0$ and $\Delta_z \to 0$.

To maintain a fixed level of accuracy, Δ_t and Δ_z should decrease as the input power P increases [13, Sec. II-A]. The value of Δ_z should be chosen proportional to the nonlinear length $Z_{NL} = 1/(\gamma P)$ and the dispersion length $Z_D = 1/(|\beta_2|W^2)$, with W being the maximum signal bandwidth during propagation. In general, the signal spectrum broadens through propagation due to the fiber nonlinearity. The bandwidth does not increase monotonically with the propagation distance (see Fig. 2.10). However, in general, the expansion of the signal bandwidth increases with the input power. Therefore, the sam-

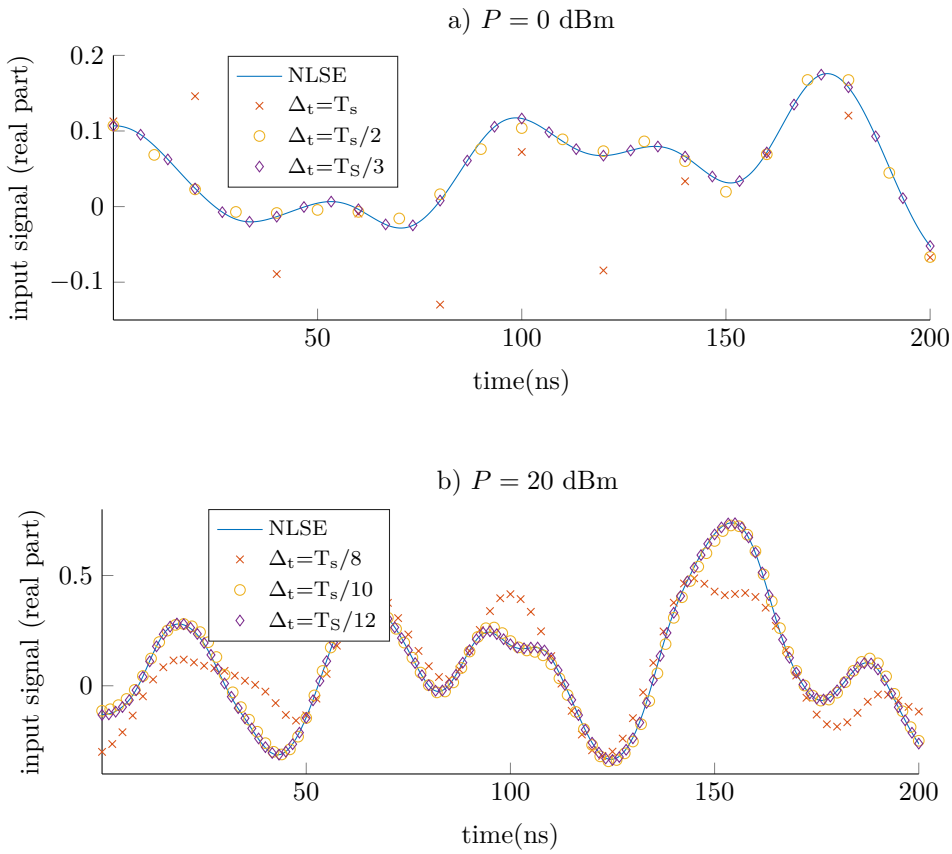

Figure 2.8: Output of the SSF method for different values of Δ_Z for a) $P = 0$ dBm and b) $P = 10$ dBm.

pling rate should increase (Δ_t should decrease) with the input power to maintain a fixed accuracy level of the SSF method. The topic of spectral broadening is further discussed in Section 2.4.

To examine the accuracy of SSF method for different values of Δ_z and Δ_t and for different input powers, we conduct the SSF simulation for an example of optical fiber system and sketch the output signal (as in [14, Fig. 2]). A 300-km lossless standard SMF is considered with typical parameters listed in Table 2.2. The exact solution of the NLS equation is calculated by choosing a very small Δ_z ($\Delta_z = 1$ m) and taking 32 samples per symbol. Root-raised cosine pulse shaping with roll-off factor of 0.2 is considered.

Fig. 2.7 depicts the output of the SSF algorithm for different values of Δ_z and two input powers. It can be seen in Fig. 2.7 (a) that with $P = 0$ dBm the output of the SSF method is almost equal to the NLS solution for $\Delta_z = 10$ km but not for $\Delta_z = 100$ km.

Table 2.2: Amplifier parameters

Parameter	Symbol	Value
Fiber length	Z	300 km
Dispersion	β_2	-21.7 ps^2/km
Nonlinearity	γ	1.27 (Wkm)$^{-1}$
Symbol rate	R_s	50 Gbaud

With $P = 10$ dBm, it can be seen in Fig. 2.7 (b) that smaller values of Δ_z are required for the SSF method to converge to the solution of the NLS equation; the output of the SSF method for $\Delta_z = 5$ km is completely different from the NLS solution. In Fig. 2.8, we examine the accuracy of the SSF method for different values of sampling time Δ_t. For $P = 0$ dBm it can be seen that by taking two samples per symbol one can obtain samples of the NLS solutions accurately via SSF method. However, to achieve a fair level of accuracy for $P = 20$ dBm, 10 samples per symbol are required.

Next, we evaluate the minimum number of samples per symbols that are required to approximate the NLS solution via the SSF method accurately. We define the error of the SSF method as the energy of the difference of the SSF output and the NLS solution normalized by the input energy. If the error is less than one percent, then the SSF method is deemed accurate. We consider a 10-Gbaud single fiber-optical channel, where the rest of parameters are specified in Table 2.2.

Fig. 2.9 (a) demonstrates the minimum required number of samples per symbol vs. the input power for three pulse shapes and two modulation formats. This figure shows that the sample rate must rapidly grow with the input power for $P > 0$ dBm. It also can be seen that the required number of samples is not just a function of the input power, but it depends also on the modulation and the pulse shape. Fig. 2.9 (b) depicts the required sampling rate vs. the transmission distance for $P = 10$ dBm with root-raised cosine pulse shaping and Gaussian input distribution. It can be seen that the required number of samples increases with the distance in the first 100 km and then it saturates.

2.4 Spectral Broadening

As was mentioned in the previous section, because of the fiber nonlinearity, the signal bandwidth broadens during propagation. The bandwidth expansion has a direct relation with the input power. Nonetheless, as yet, there is no analytical result that expresses the relation between the input power and the bandwidth of the received signal. Actually, the input power may not be the sufficient statistics to determine the propagating signal's bandwidth. Pulse shaping has an effect on the signal's spectrum during propagation. Soliton pulses [15, Ch. 5] can propagate through the fiber without suffering from spectral

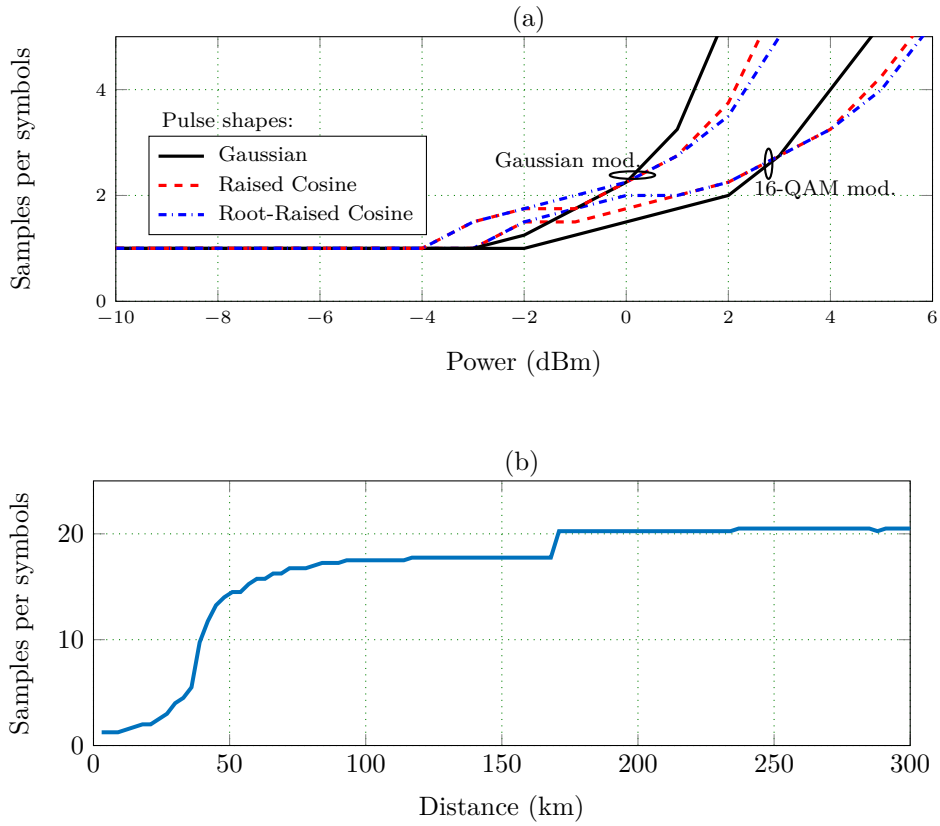

Figure 2.9: Minimum number of samples per symbol needed to reach less than 1% error in the SSF method vs. a) input power and b) distance.

broadening.

Fig. 2.10 depicts the maximum signal bandwidth as a function of the input power and the propagation distance. A single channel with 10 Gbaud symbol rate is simulated; the rest of the channel parameters can be found in Table 2.2. It can be seen in Fig. 2.10 (a) that for powers less than 0 dBm the signal bandwidth does not expand during the propagation, as the nonlinear effects are negligible; for higher powers, the bandwidth grows rapidly with P. Fig. 2.10 (a) depicts the signal bandwidth as a function of the propagation distance for $P = 10$ dBm. As it can be seen, the signal's bandwidth first increases at the beginning, reaches a maximum, and then saturates. It is evident from both Fig. 2.10 (a) and Fig. 2.10 (b) that the spectrum broadening does not only depend on input power but also other statistics of the signal.

As mentioned, currently there is no analytical expression that describes the relation between the propagating signal bandwidth, the statistics of the input signal, and the

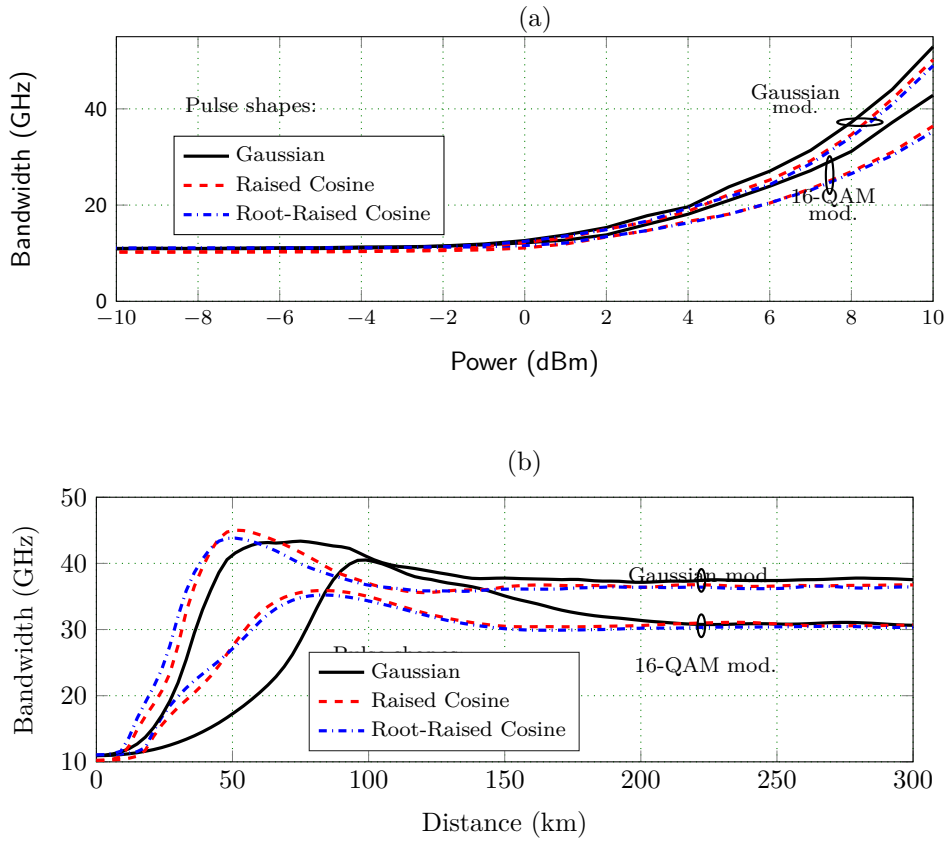

Figure 2.10: Bandwidth (99%) of the propagating signal vs. a) input power and b) distance.

channel parameters. As a conjecture, one may use the Hamiltonian invariant of the NLS equation as a starting point to tackle this open problem. For the lossless NLS equation (Eq. (2.2) with $\alpha = 0$), the Hamiltonian invariant [16] (see also [17, Eq. (12c)]) indicates that the value of

$$\beta_2 \int_{-\infty}^{\infty} \left| \frac{\partial a(z,t)}{\partial t} \right|^2 \mathrm{d}t + \gamma \int_{-\infty}^{\infty} |a(z,t)|^4 \, \mathrm{d}t \tag{2.26}$$

is constant during the propagation (is independent of z). The first term of (2.26) is proportional to the mean square of the signal bandwidth and the second term to the fourth moment of the signal. This suggest a relation between the signal's fourth moment and its bandwidth.

2.5 WDM Systems and Impairments Related to It

WDM systems consist of multiple channels at different carrier wavelengths copropagating through a single fiber. The number of channels is usually limited by the bandwidth of the optical amplifier used in these systems, which for EDFAs is roughly 5 THz. Therefore, around 100 channels with bandwidth of 50 GHz can be transmitted. Due to the Kerr effect these channels interfere with each other. Spectrum guard bands are applied to reduce the crosstalk between channels at the cost of reducing spectral efficiency.

To better understand the effects of nonlinearity distortions on the signal, it is customary to categorize them into signal–signal and signal–noise distortions. The former is caused by interaction between multiple frequency components of the signal and the latter by interactions between signal and noise. Independently, one can classify the nonlinear effects as intra-channel and inter-channel distortions. The former originate from the interactions between the frequency components within the bandwidth of the channel of interest (COI), while in the latter category, at least one of the components (signal or noise) belongs to the neighboring channels. Therefore, the nonlinear impairments of WDM systems can be classified into four categories: intra-channel/inter-channel signal–signal/signal–noise. The intra-channel signal–signal distortion is also known as self-phase modulation (SPM). The inter-channel signal–signal interference is further classified into cross-phase modulation (XPM) and four-wave mixing (FWM). In the following, we elaborate upon SPM, XPM, and FWM. For a more comprehensive discussion on this topic, see [18] and references therein.

To model the effect of the nonlinear crosstalk, we consider a WDM channel with two wavelengths. The signal propagation through these channels in the fiber can be described by the pair of coupled equations [15, p. 274].

$$\frac{\partial \mathbf{a}_1}{\partial z} + j\frac{\beta_{21}}{2}\frac{\partial^2 \mathbf{a}_1}{\partial t^2} - j\gamma_1(|\mathbf{a}_1|^2 + 2|\mathbf{a}_2|^2)\mathbf{a}_1 + \frac{\alpha}{2}\mathbf{a}_1 = 0 \tag{2.27}$$

$$\frac{\partial \mathbf{a}_2}{\partial z} + j\frac{\beta_{22}}{2}\frac{\partial^2 \mathbf{a}_2}{\partial t^2} - j\gamma_2(|\mathbf{a}_2|^2 + 2|\mathbf{a}_1|^2)\mathbf{a}_2 + \frac{\alpha}{2}\mathbf{a}_2 + jd\frac{\partial \mathbf{a}_2}{\partial t} = 0. \tag{2.28}$$

Here, β_{2i} is the group velocity dispersion at wavelength $i \in \{1, 2\}$, γ_i is the nonlinear parameter, and $\mathbf{a}_i$ corresponds to the ith baseband signal. The attenuation constant α is assumed equal for both channels. The parameter d is the group velocity mismatch

$$d = \frac{1}{v_{g2}} - \frac{1}{v_{g1}} \tag{2.29}$$

where v_{gi} is the group velocity. In (2.27) and (2.28), the time is measured according to a reference frame moving with the first signal. The term $j\gamma_1|\mathbf{a}_1|^2$ in (2.27) corresponds to the SPM, that is the modulation of signal's phase by its amplitude. The term $j\gamma_1|\mathbf{a}_2|^2$ in (2.27) is known as XPM and captures the interference between the copropagating signals. It can be seen that the phase shift caused by XPM is twice as large as that caused by SPM.

The phase distortions caused by SPM and XPM, although they do not change the signal's amplitude, expand the signal's spectrum. The only exception a perfect rectangular pulse shape, where the phase distortions are constant during one symbol period.

Another important impairment in WDM systems is the FWM, where three signal at different frequencies f_1, f_2, and f_3 combine with each other to create a signal at $f_4 = f_1 + f_2 - f_3$. With many signals propagating in WDM, FWM can be a major source of distortion. This effect can be mitigated by proper channel spacing [19].

CHAPTER 3

Channel Capacity

In this chapter, we study 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. The capacity and the spectral efficiency of the continuous-time additive white Gaussian noise (AWGN) channel are defined in Section 3.3. Section 3.4 provides some information-theory tools to bound capacity. Finally, we define the achievable information rate (AIR) in Section 3.5 and provide a detailed description on how to compute it for some auxiliary channels.

3.1 Differential Entropy and Mutual Information

Consider two random vectors $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ that take values in $\mathbb{R}^L$ and are distributed according to the joint probability distribution function (pdf) $p_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x},\underline{y})$. The differential entropy of $\underline{\mathbf{x}}$ is defined as [20, Sec. 8]

$$h(\underline{\mathbf{x}}) = -\int_{\mathbb{R}^L} p_{\underline{\mathbf{x}}}(\underline{x}) \log\big(p_{\underline{\mathbf{x}}}(\underline{x})\big)\, \mathrm{d}\underline{x}. \tag{3.1}$$

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

$$h(\underline{\mathbf{x}},\underline{\mathbf{y}}) = -\int_{\mathbb{R}^L}\int_{\mathbb{R}^L} p_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x},\underline{y}) \log\Big(p_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\underline{x},\underline{y})\Big)\, \mathrm{d}\underline{x}\, \mathrm{d}\underline{y}. \tag{3.2}$$

Moreover the mutual information between the random vectors $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ is defined as

$$I(\underline{\mathbf{x}};\underline{\mathbf{y}}) = h(\underline{\mathbf{y}}) + h(\underline{\mathbf{x}}) - h\big(\underline{\mathbf{x}},\underline{\mathbf{y}}\big) \tag{3.3}$$

$$= h(\underline{\mathbf{y}}) - h(\mathbf{y}|\underline{\mathbf{x}}) \tag{3.4}$$

where

$$h(\mathbf{y}|\underline{\mathbf{x}}) = h(\mathbf{y},\underline{\mathbf{x}}) - h(\underline{\mathbf{x}}) \tag{3.5}$$

is the conditional entropy. The mutual information can be written in terms of relative entropy as

$$I(\underline{\mathbf{x}};\underline{\mathbf{y}}) = D(p_{\mathbf{x},\mathbf{y}} \parallel p_{\mathbf{x}}p_{\mathbf{y}}) \tag{3.6}$$

where the the relative entropy between two pdfs $p(\underline{w})$ and $q(\underline{w})$ is defined as

$$D(p \parallel q) = \int p(\underline{w}) \log\left(\frac{p(\underline{w})}{q(\underline{w})}\right) \mathrm{d}\underline{w}. \tag{3.7}$$

The relative entropy is always nonnegative and serves as a distance metric between two distributions [20, Thm. 8.6.1]. Let now $\underline{\mathbf{x}}$ take value in $\mathbb{C}^L$. We have

$$h(\underline{\mathbf{x}}) = h(\underline{\mathbf{x}}^r, \underline{\mathbf{x}}^i) \tag{3.8}$$

where $\underline{\mathbf{x}}^r$ and $\underline{\mathbf{x}}^i$ are the random vectors representing the real and imaginary parts of $\underline{\mathbf{x}}$, respectively. Also, the mutual information between two complex random vectors can be defined similarly as in (3.3).

3.2 Capacity of Discrete-Time Channels

A discrete-time memoryless channel can be described by a conditional pdf $p_{\mathbf{y}|\mathbf{x}}(y|x)$, where $\mathbf{x}$ and $\mathbf{y}$ are complex random variables. The capacity $\mathcal{C}^{\mathrm{DMC}}$ of this channel is the maximum rate in bits per channel use at which the information can be transferred through the channel with arbitrary low error probability. By Shannon's channel coding theorem [4], the capacity under the average power constraint $P > 0$, can be calculated as

$$\mathcal{C}^{\mathrm{DMC}} = \sup_{p_{\mathbf{x}}(x)} I(\mathbf{x};\mathbf{y}) \qquad \text{bits per channel use} \tag{3.9}$$

where the supremum is taken over all input distributions $p_{\mathbf{x}}(x)$ such that $\mathrm{E}\big[|\mathbf{x}|^2\big] \leq P$.

Similarly, for a discrete-time block-memoryless communication channel described by the conditional pdf $p_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y} \mid \underline{x})$, the capacity, $\mathcal{C}^{\mathrm{DBC}}$, under the average power constraint $P > 0$, is

$$\mathcal{C}^{\mathrm{DBC}} = \sup_{p_{\underline{\mathbf{x}}}(\underline{x})} I(\underline{\mathbf{x}},\underline{\mathbf{y}}) \qquad \text{bits per block.} \tag{3.10}$$

Here, the supremum is taken over all input distributions $p_{\underline{\mathbf{x}}}(\underline{x})$ such that $\mathrm{E}\big[||\underline{\mathbf{x}}||^2\big] \leq LP$, where L is the input vector length, which coincides with the block size. Alternatively, we have

$$\mathcal{C}^{\mathrm{DBC}} = \frac{1}{L} \sup_{p_{\underline{\mathbf{x}}}(\underline{x})} I(\underline{\mathbf{x}}, \underline{\mathbf{y}}) \qquad \text{bits per channel use.} \tag{3.11}$$

3.3 Capacity of the Continuous-Time AWGN Channel

The capacity of a continuous-time channel in bits per second is the maximum of the average number of information bits that can be transmitted through the channel during one second at an arbitrarily low error probability. The maximum is over all modulation, pulse shaping, demodulation, detection, and coding schemes. For the complex AWGN channel, the input–output relation can be written as $\mathbf{y}(t) = x(t) + \mathbf{n}(t)$, where $\mathbf{n}(t)$ is a complex white Gaussian process with power spectral density (PSD) N_0 W/Hz and $x(t)$ is the complex input signal with two-sided bandwidth W Hz. The capacity, $\mathcal{C}^{\mathrm{AWGN}}$, of this channel under the power constraint $\lim\limits_{T\to\infty} 1/T \int_0^T |x(t)|^2 \,\mathrm{d}t \leq P$ is [4] (see also [20, Eq. (9.62)])

$$\mathcal{C}^{\mathrm{AWGN}} = W \log\left(1 + \frac{P}{WN_0}\right) \qquad \text{bits per second.} \tag{3.12}$$

For a continuous-time channel, spectral efficiency is measured in terms of bits per seconds per Hertz by dividing its capacity to its bandwidth, W. For the AWGN channel the spectral efficiency, SE^{AWGN} , is

$$SE^{\mathrm{AWGN}} = \log\left(1 + \frac{P}{WN_0}\right) \qquad \text{bits per second per Hertz.} \tag{3.13}$$

3.4 Some Information-Theoretic Tools for Bounding Capacity

In this section, we provide some standard information-theoretic tools to upper- and lower-bound the capacity. Also we introduce the Blahut–Arimoto algorithm, which can be used to find the capacity-achieving input distribution for discrete-time memoryless channels with finite-cardinality alphabets.

3.4.1 Maximum Entropy

Among all real random vectors $\underline{\mathbf{x}} \in \mathbb{R}^L$ with a fixed nonsingular correlation matrix $\mathsf{R}(\mathbf{x}) = \mathrm{E}\big[\underline{\mathbf{x}}\underline{\mathbf{x}}^T\big]$, the joint Gaussian distribution has maximum differential entropy [20, Thm. 8.6.5], i.e.,

$$h(\mathbf{x}) \leq \frac{1}{2} \log\Big((2\pi e)^L \det \mathsf{R}(\underline{\mathbf{x}})\Big) . \tag{3.14}$$

Using Hadamard's inequality [20, Thm. 17.9.2] and Jensen's inequality [20, Sec. 2.6] we can further upper-bound (3.14) as [5]

$$h(\underline{\mathbf{x}}) \leq \frac{L}{2}\log\left(\frac{2\pi e}{L}\operatorname{Tr}\left[\mathsf{R}(\underline{\mathbf{x}})\right]\right) \tag{3.15}$$

$$= \frac{L}{2}\log\left(\frac{2\pi e}{L}\mathrm{E}\left[\|\underline{\mathbf{x}}\|^2\right]\right). \tag{3.16}$$

3.4.2 Polar Coordinate System

The differential entropy of a complex random variable $\mathbf{x}$ can be computed in polar coordinates as [21, Lemma 6.16]

$$h(\mathbf{x}) = h(|\mathbf{x}|, \angle(\mathbf{x})) + \mathrm{E}[\log|\mathbf{x}|]\,. \tag{3.17}$$

Here, $\angle(\mathbf{x})$ denotes the phase of $\mathbf{x}$. Furthermore, [21, Lemma 6.15]

$$h\left(|\mathbf{x}|^2\right) = h(|\mathbf{x}|) + \mathrm{E}[\log|\mathbf{x}|] + \log 2. \tag{3.18}$$

Using (3.17) and (3.18), we can upper-bound $h(\mathbf{x})$ as

$$h(\mathbf{x}) = h(|\mathbf{x}|) + h(\angle(\mathbf{x})\,||\mathbf{x}|\,) + \mathrm{E}[\log|\mathbf{x}|] \tag{3.19}$$

$$= h\left(|\mathbf{x}|^2\right) - \log 2 + h(\angle(\mathbf{x})\,||\mathbf{x}|\,) \tag{3.20}$$

$$\leq h\left(|\mathbf{x}|^2\right) + \log\pi. \tag{3.21}$$

In the last step, we used that $h(\angle(\mathbf{x})\,|\,|\mathbf{x}|) \leq h(\angle(\mathbf{x})) \leq \log(2\pi)$.

3.4.3 Entropy Power Inequality

The entropy power inequality [20, Theorem 17.7.3] is a powerful tool for lower-bounding the entropy of the sum of two independent random variables/vectors. Let $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ be two real independent L-dimensional random vectors, we have that

$$2^{\frac{2}{L}h(\underline{\mathbf{x}}+\underline{\mathbf{y}})} \geq 2^{\frac{2}{L}h(\underline{\mathbf{x}})} + 2^{\frac{2}{L}h(\underline{\mathbf{y}})} \tag{3.22}$$

3.4.4 Duality Bound

The duality bound [21–23] upper-bounds the mutual information between two random variables $\mathbf{x}$ and $\mathbf{y}$ with density $p_{\mathbf{x},\mathbf{y}}$ and alphabets $\mathcal{X}$ and $\mathcal{Y}$ as

$$I(\mathbf{x};\mathbf{y}) \leq \mathrm{E}\left[\log\left(\frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})}{q_{\mathbf{y}}(\mathbf{y})}\right)\right] \tag{3.23}$$

$$= -\mathrm{E}[\log(q_{\mathbf{y}}(\mathbf{y}))] - h(\mathbf{y}|\mathbf{x}) \tag{3.24}$$

where $q_{\mathbf{y}}$ is any arbitrary distribution on $\mathcal{Y}$ and the expectation in (3.24) is with respect to the distribution $p_{\mathbf{y}}$. Eq. (3.24) can be proved by using (3.4) and

$$-\mathrm{E}[\log(q_{\mathbf{y}}(\mathbf{y}))] = \mathrm{E}\left[\log(\frac{p_{\mathbf{y}}(\mathbf{y})}{q_{\mathbf{y}}(\mathbf{y})})\right] - \mathrm{E}[\log(p_{\mathbf{y}}(\mathbf{y}))] \tag{3.25}$$

$$= D(p_{\mathbf{y}} \parallel q_{\mathbf{y}}) + h(\mathbf{y}) \tag{3.26}$$

$$\geq h(\mathbf{y}) \tag{3.27}$$

where (3.27) follows from the non-negativity of the relative entropy. One can use the duality bound together with the method of Lagrange multipliers to upper-bound the capacity under the power constraint P as [21]

$$\mathcal{C} = \max_{p_{\mathbf{x}}} I(\mathbf{y};\mathbf{x}) \tag{3.28}$$

$$\leq \max_{p_{\mathbf{x}}} -\mathrm{E}[\log(q_{\mathbf{y}}(\mathbf{y}))] - h(\mathbf{y}|\mathbf{x}) \tag{3.29}$$

$$\leq \max_{p_{\mathbf{x}}} -\mathrm{E}[\log(q_{\mathbf{y}}(\mathbf{y}))] - h(\mathbf{y}|\mathbf{x}) + \lambda\left(P - \mathrm{E}[|\mathbf{x}|^2]\right) \tag{3.30}$$

$$= \max_{p_{\mathbf{x}}} \mathrm{E}_{\mathbf{x}}\left[-\mathrm{E}_{\mathbf{y}|\mathbf{x}}[\log(q_{\mathbf{y}}(\mathbf{y}))] - h(\mathbf{y}|\mathbf{x}) + \lambda\left(P - |\mathbf{x}|^2\right)\right] \tag{3.31}$$

$$\leq \max_{x} -\mathrm{E}_{\mathbf{y}|\mathbf{x}}[\log(q_{\mathbf{y}}(\mathbf{y}))] - h(\mathbf{y}|\mathbf{x}=x) + \lambda\left(P - |x|^2\right) \tag{3.32}$$

where, $\lambda \geq 0$ is the Lagrange multiplier. Therefore, a capacity upper bound can be obtained by choosing the distribution $q_{\mathbf{y}}$ and minimizing (3.32) over λ.

3.4.5 Blahut–Arimoto Algorithm

For a memoryless channel, if mutual information can be calculated for every input distribution $p_{\mathbf{x}}$ (e.g., for discrete-time memoryless channels with finite input and output alphabets), then the capacity can be numerically evaluated via Blahut–Arimoto algorithm [20, Sec. 10.8]. This method was presented by Arimoto [24] and Blahut [25] independently. The algorithm can be used to find the capacity-achieving distribution in an iterative fashion. If the input or the output has a continuous alphabet, then one can estimate the capacity by discretization. The constrained capacity can also be computed with this method [25, Fig. 4].

3.5 AIR: Definition and Calculation

In this section, first the AIR is defined for channels with memory and then algorithms to calculate the AIR are provided for some auxiliary channels.

3.5.1 Definition of AIR

Consider a generic discrete-time communication channel defined by an input alphabet $\mathcal{X}^L$, output alphabet $\mathcal{Y}^L$, and a conditional probability function $p_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})$. In the absence

of an optimal receiver for this channel, one can deploy a mismatched receiver that is optimized for an auxiliary channel $q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})$. The AIR determines an information rate that can be transferred reliably through the true channel $p_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}$ by using the mismatched receiver optimized for the auxiliary channel $q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}$ for a fixed input distribution $p_{\underline{\mathbf{x}}}(x)$. It is defined as

$$I_q(\underline{\mathbf{x}},\underline{\mathbf{y}}) = \lim_{L\to\infty}\frac{1}{L}\mathrm{E}\left[\log\frac{q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{\mathbf{y}}|\underline{\mathbf{x}})}{q_{\underline{\mathbf{y}}}(\underline{\mathbf{y}})}\right] \tag{3.33}$$

$$= \lim_{L\to\infty}\frac{1}{L}\mathrm{E}\Big[\log q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{\mathbf{y}}|\underline{\mathbf{x}})\Big] - \lim_{L\to\infty}\frac{1}{L}\mathrm{E}\Big[\log q_{\underline{\mathbf{y}}}(\underline{\mathbf{y}})\Big] \tag{3.34}$$

$$\tag{3.35}$$

where L is the length of $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ and

$$q_{\underline{\mathbf{y}}}(\underline{y}) = \int_{\mathcal{X}^L} q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})p_{\underline{\mathbf{x}}}(\underline{x})\,\mathrm{d}\underline{x}. \tag{3.36}$$

Similarly, if $\mathcal{X}$ is a discrete set, we have

$$q_{\underline{\mathbf{y}}}(\underline{y}) = \sum_{\underline{x}\in\mathcal{X}^L} q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})p_{\underline{\mathbf{x}}}(\underline{x}). \tag{3.37}$$

The maximization of the AIR in (3.33) over all $q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}$ results in $I(\underline{\mathbf{x}};\underline{\mathbf{y}})$, and $q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}} = p_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}$ achieves this maximum. Further maximization over the input distribution $p_{\underline{\mathbf{x}}}$ results in the channel capacity. Therefore, the AIR calculated for any arbitrary auxiliary channel $q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}$ and any input distribution $p_{\underline{\mathbf{x}}}$ establishes a lower bound on the capacity. Also, roughly speaking, the tightness of this lower bound depends on how close is the auxiliary channel to the real one and the input distribution to the capacity-achieving one.

One can estimate the value of (3.33) with two long sequences $\underline{x}$ and $\underline{y}$ sampled from $p_{\mathbf{x},\mathbf{y}}$ as [26, Eq. (46)]

$$I_q(\underline{\mathbf{x}},\underline{\mathbf{y}}) \approx \frac{1}{L}\log q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x}) - \frac{1}{L}\log q_{\underline{\mathbf{y}}}(\underline{y}). \tag{3.38}$$

Therefore, the AIR can be estimate with the following three steps [26].

1. Generate a "very long" input vector $\underline{x}$ and sample an output vector $\underline{y}$ according to the true channel law $p_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})$.
2. Calculate $(1/L)\log(q_{\underline{\mathbf{y}}}(\underline{y}))$ and $(1/L)\log(q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x}))$.
3. Estimate the AIR via (3.38).

For the fiber-optical channel the output vector $\underline{y}$ in the first step can be obtained using the SSF method. Next, we study the calculation of the two terms in Step 2. Specifically, we will see how the AIR can be calculated effectively for independent and identically distributed (iid) input distributions and *a)* for AWGN auxiliary channels and *b)* for an auxiliary phase-noise channel with memory.

3.5.2 Auxiliary AWGN Channel With Gaussian Input

In this section we study the calculation of (3.38) when the auxilary channel is an AWGN one and the input distribution is an iid circularly-symmetric complex Gaussian. With the AWGN auxiliary channel, the input–output relation can be described by the memoryless channel $\mathbf{y}_\ell = \alpha \mathbf{x}_\ell + \mathbf{n}_\ell$, where $\mathbf{n}_\ell \sim \mathcal{CN}(0, \sigma_n^2)$. A natural choice of α and σ_n^2 is

$$\alpha = \sigma_{xy}/\sigma_x^2 \tag{3.39}$$

$$\sigma_n^2 = \frac{1}{L}\sum_{\ell=1}^{L} |y_\ell - \alpha x_\ell|^2 \tag{3.40}$$

$$= \sigma_y^2 - |\alpha|^2 \sigma_x^2 \tag{3.41}$$

where

$$\sigma_{xy} = \frac{1}{L}\sum_{\ell=1}^{L} y_\ell x_\ell^* \tag{3.42}$$

$$\sigma_x^2 = \frac{1}{L}\sum_{\ell=1}^{L} |x_\ell|^2 \tag{3.43}$$

$$\sigma_y^2 = \frac{1}{L}\sum_{\ell=1}^{L} |y_\ell|^2. \tag{3.44}$$

For the AWGN auxiliary channel, we have

$$q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x}) = \prod_{\ell=1}^{L} \frac{1}{\pi\sigma_n^2} \exp\left(-\frac{|y_\ell - \alpha x_\ell|^2}{\sigma_n^2}\right) \tag{3.45}$$

Therefore,

$$\frac{1}{L}\log q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x}) = -\log(\pi\sigma_n^2) - \frac{\log e}{L\sigma_n^2}\sum_{\ell=1}^{L} |y_\ell - \alpha x_\ell|^2 \tag{3.46}$$

$$= -\log(\pi e \sigma_n^2) \tag{3.47}$$

where in (3.47) we used (3.40).

Moreover, with $p_{\mathbf{x}} \sim$ iid $\mathcal{CN}(0, \sigma_x^2)$, $q_{\mathbf{y}}$ becomes an iid circularly symmetric Gaussian distribution with variance $|\alpha|^2\sigma_x^2 + \sigma_n^2$, which by (3.41) is equal to σ_y^2. Therefore,

$$-\frac{1}{L}\log q_{\underline{\mathbf{y}}}(\underline{y}) = \log(\pi e \sigma_y^2). \tag{3.48}$$

Substituting (3.47) and (3.48) into (3.38), we obtain

$$I_q(\underline{\mathbf{x}}, \underline{\mathbf{y}}) \approx \log\left(\frac{\sigma_y^2}{\sigma_n^2}\right) \tag{3.49}$$

$$= \log\left(\frac{\sigma_y^2 \sigma_x^2}{\sigma_y^2 \sigma_x^2 - |\sigma_{xy}|^2}\right). \tag{3.50}$$

Equation (3.50) is established in [6, Eq. (6)] and is used in Paper C to calculate the AIR based on the Gaussian auxiliary channel.

3.5.3 Auxiliary AWGN Channel With Discrete Input

In this section, we consider a more practical case, where the input takes value from a finite alphabet set $\mathcal{X}$. Compared with Section 3.5.2, the main difference here is that now $q_{\mathbf{y}}$ is not a Gaussian distribution and (3.48) does not hold. The auxiliary channel is $\mathbf{y}_\ell = \mathbf{x}_\ell + \mathbf{n}_\ell$ with $n_\ell \sim \text{iid}\, \mathcal{CN}(0, \sigma_n^2)$. In order to use this auxiliary channel, one should scale the output of the demodulator properly. For example, it can be scaled by the factor $1/\alpha$, where α is defined in (3.39).

A natural choice of the parameter σ_n^2 is

$$\sigma_n^2 = \frac{1}{L}\sum_{\ell=1}^{L} |y_\ell - x_\ell|^2. \tag{3.51}$$

Following the same calculation as in the previous chapter (with $\alpha = 1$) we have

$$\frac{1}{L} \log q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x}) = -\log(\pi e \sigma_n^2). \tag{3.52}$$

Furthermore, assuming that $q_{\underline{\mathbf{x}}}$ is iid, we have

$$-\frac{1}{L}\log\Big(q_{\underline{\mathbf{y}}}(\underline{y})\Big) = -\frac{1}{L}\log\left(\prod_{\ell=1}^{L} q_{\mathbf{y}_\ell}(y_\ell)\right) \tag{3.53}$$

$$= -\frac{1}{L}\sum_{\ell=1}^{L}\log\left(\sum_{x_\ell \in \mathcal{X}} p_{\mathbf{x}}(x_\ell) q_{\mathbf{y}|\mathbf{x}}(y_\ell|x_\ell)\right) \tag{3.54}$$

where $q_{\mathbf{y}|\mathbf{x}}$ can be found in (3.45) with $\alpha = 1$ and $L = 1$.

3.5.4 Auxiliary Channel With Auto-Regressive Phase-Noise Process

In this section, we review the method proposed in [26, 27] to compute the AIR for an auxiliary channel with phase noise described as

$$\mathbf{y}_\ell = h_0 \mathbf{x}_\ell e^{j\theta_\ell} + \mathbf{n}_\ell \tag{3.55}$$

where $\mathbf{n}_\ell \sim$ iid $\mathcal{CN}(0, \sigma_n^2)$, h_0 is channel coefficient. Here, $\mathbf{x}_\ell$ and $\mathbf{y}_\ell$ indicate the input and the output of the channel and belong to the alphabet sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. The parameter θ_ℓ is an real auto-regressive (AR) phase-noise process of order one described by

$$\theta_\ell = \theta_{\ell-1} + \mathbf{z}_\ell \qquad \mod 2\pi \tag{3.56}$$

where $\mathbf{z}_\ell \sim$ iid $\mathcal{CN}(0, \sigma_z^2)$. The process θ_ℓ can be also regarded as a discrete-time Wiener process. To calculate the AIR, we begin by evaluating the first term on the RHS of (3.38). We have

$$\frac{1}{L}\log q(\underline{y}|\underline{x}) = \frac{1}{L}\log \prod_{\ell=1}^{L} q(y_\ell|\underline{x}, \underline{y}^{\ell-1}) \tag{3.57}$$

$$= \frac{1}{L}\sum_{\ell=1}^{L} \log q(y_\ell|\underline{x}, \underline{y}^{\ell-1}) \tag{3.58}$$

$$= \frac{1}{L}\sum_{\ell=1}^{L} \log q(y_\ell|\underline{x}^{\ell}, \underline{y}^{\ell-1}). \tag{3.59}$$

Furthermore,

$$q(y_\ell|\underline{x}^{\ell}, \underline{y}^{\ell-1}) = \int_0^{2\pi} q(y_\ell, \theta_\ell|\underline{x}^{\ell}, \underline{y}^{\ell-1})\,\mathrm{d}\theta_\ell. \tag{3.60}$$

We continue by calculating

$$q(y_\ell, \theta_\ell|\underline{x}^{\ell}, \underline{y}^{\ell-1}) = \int_0^{2\pi} q(y_\ell, \theta_\ell, \theta_{\ell-1}|\underline{x}^{\ell}, \underline{y}^{\ell-1})\,\mathrm{d}\theta_{\ell-1} \tag{3.61}$$

$$= \int_0^{2\pi} q(\theta_{\ell-1}|\underline{x}^{\ell-1}, \underline{y}^{\ell-1}) q(\theta_\ell|\theta_{\ell-1}) q(y_\ell|x_\ell, \theta_\ell)\,\mathrm{d}\theta_{\ell-1} \tag{3.62}$$

where (3.62) follows from the Markov property of the channel. Based on the channel model (3.55), we have

$$q(\theta_\ell|\theta_{\ell-1}) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \sum_{m=-\infty}^{\infty} e^{\frac{-(\theta_\ell - \theta_{\ell-1} + 2m\pi)^2}{2\sigma_z^2}} \tag{3.63}$$

and also

$$q(y_\ell|x_\ell, \theta_\ell) = \frac{1}{\pi\sigma_n^2} \exp\left(\frac{-|y_\ell - h_0 x_\ell e^{\theta_\ell}|^2}{\sigma_n^2}\right). \tag{3.64}$$

Finally, we have

$$q(\theta_{\ell-1}|\underline{x}^{\ell-1},\underline{y}^{\ell-1}) = \frac{q(y_{\ell-1},\theta_{\ell-1}|\underline{x}^{\ell-1},\underline{y}^{\ell-2})}{q(y_{\ell-1}|\underline{x}^{\ell-1},\underline{y}^{\ell-2})}. \tag{3.65}$$

To calculate (3.65) one can use the equations (3.60) and (3.62) by replacing ℓ with $\ell-1$. This leads to an iterative process for calculating (3.57). The process is initialized by taking $q(\theta_0)$ as uniform distribution over the interval $[0, 2\pi)$. Then the process of calculating (3.57) can be described as follows.

Algorithm 3: Calculating $\frac{1}{L}\log q(\underline{y}|\underline{x})$

Data: Input vector: $\underline{x}$ of length L; output vector $\underline{y}$; parameters: σ_n^2, σ_z^2, h_0

$q(\theta_0|\underline{x}^0,\underline{y}^0) \leftarrow \frac{1}{2\pi}$

for $\ell \in \{1,2,\ldots,L\}$ **do**

  $q(y_\ell,\theta_\ell|\underline{x}^\ell,\underline{y}^{\ell-1}) \leftarrow \int_0^{2\pi} q(\theta_{\ell-1}|\underline{x}^{\ell-1},\underline{y}^{\ell-1})q(\theta_\ell|\theta_{\ell-1})q(y_\ell|x_\ell,\theta_\ell)\,\mathrm{d}\theta_{\ell-1}$

  $q(y_\ell|\underline{x}^\ell,\underline{y}^{\ell-1}) \leftarrow \int_0^{2\pi} q(y_\ell,\theta_\ell|\underline{x}^\ell,\underline{y}^{\ell-1})\,\mathrm{d}\theta_\ell$

  $q(\theta_\ell|\underline{x}^\ell,\underline{y}^\ell) \leftarrow q(y_\ell,\theta_\ell|\underline{x}^\ell,\underline{y}^{\ell-1})/q(y_\ell|\underline{x}^\ell,\underline{y}^{\ell-1})$

return $\frac{1}{L}\sum_{\ell=1}^{L}\log q(y_\ell|\underline{x}^\ell,\underline{y}^{\ell-1})$

Next, we calculate the second term on the RHS of (3.38) that is $\frac{1}{L}\log q(\underline{y})$. This term can be calculated in the same fashion as $\frac{1}{L}\log q(\underline{y}|\underline{x})$. These calculations are presented in the following algorithm.

Algorithm 4: Calculating $\frac{1}{L}\log q(\underline{y})$

Data: Input distribution: $p_{\mathbf{x}}(x)$; output vector $\underline{y}$ of length L; parameters: σ_n^2, σ_z^2, h_0

$q(\theta_0|\underline{y}^0) \leftarrow \frac{1}{2\pi}$

for $\ell \in \{1,2,\ldots,L\}$ **do**

  $q(y_\ell,\theta_\ell|\underline{y}^{\ell-1}) \leftarrow \int_0^{2\pi} q(\theta_{\ell-1}|\underline{y}^{\ell-1})q(\theta_\ell|\theta_{\ell-1})q(y_\ell|\theta_\ell)\,\mathrm{d}\theta_{\ell-1}$

  $q(y_\ell|\underline{y}^{\ell-1}) \leftarrow \int_0^{2\pi} q(y_\ell,\theta_\ell|\underline{y}^{\ell-1})\,\mathrm{d}\theta_\ell$

  $q(\theta_\ell|\underline{y}^\ell) \leftarrow q(y_\ell,\theta_\ell|\underline{y}^{\ell-1})/q(y_\ell|\underline{y}^{\ell-1})$

return $\frac{1}{L}\sum_{\ell=1}^{L}\log q(y_\ell|\underline{y}^{\ell-1})$

In Algorithm 4, we have that

$$q(y_\ell|\theta_\ell) = \int_{\mathcal{X}} q(y_\ell, x_\ell|\theta_\ell)\,\mathrm{d}x_\ell \tag{3.66}$$

$$= \int_{\mathcal{X}} p_{\mathbf{x}}(x_\ell) q(y_\ell|\theta_\ell, x_\ell)\,\mathrm{d}x_\ell \tag{3.67}$$

where $p_{\mathbf{x}}$ is the input distribution. If $\mathcal{X}$ is discrete then

$$q(y_\ell|\theta_\ell) = \sum_{x_\ell \in \mathcal{X}} p_{\mathbf{x}}(x_\ell) q(y_\ell|\theta_\ell, x_\ell). \tag{3.68}$$

3.5.5 AIR Calculation Using Particle Method

In general, the integrals in Algorithm 3 and Algorithm 4 cannot be calculated analytically. Here, we study the particle approach presented in [27] to calculate the AIR numerically. With this method, an arbitrary pdf, say $p_{\mathbf{v}}(v)$, is represented by a set of particles $\{v^{(m)}\}_1^M$ and a set of weights $\{\omega^{(m)}\}_{m=1}^M$ such that $\sum \omega^{(m)} = 1$ and for any function $g(\,\cdot\,)$

$$\mathrm{E}_{\mathbf{v}}[g(\mathbf{v})] \approx \sum_{i=1}^{M} \omega^{(m)} g(v^{(m)}). \tag{3.69}$$

We use the notation $p_{\mathbf{v}}(v) \equiv \{v^{(m)}, \omega^{(m)}\}$ to indicate the particles and the weights corresponding to the pdf $p_{\mathbf{v}}(v)$.

To evaluate the AIR based on (3.38), we first calculate the term $\frac{1}{L}\log q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})$. Based on (3.59), this term is equal to $1/L \sum_\ell^L \log \lambda_\ell$, where

$$\lambda_\ell = q(y_\ell|\underline{x}^\ell, \underline{y}^{\ell-1}) \tag{3.70}$$

$$= \int_0^{2\pi} q(\theta_\ell|\underline{x}^\ell, \underline{y}^{\ell-1}) q(y_\ell|x_\ell, \theta_\ell)\,\mathrm{d}\theta_\ell \tag{3.71}$$

$$= \mathrm{E}_{\theta_\ell}\left[q(y_\ell|x_\ell, \theta_\ell)|\underline{x}^\ell, \underline{y}^{\ell-1}\right] \tag{3.72}$$

$$\approx \sum_{m=1}^{M} \omega_\ell^{(m)} q(y_\ell|x_\ell, \theta_\ell^{(m)}) \tag{3.73}$$

where the particle set $\{\theta_\ell^{(i)}\}_1^M$ and the weight set $\{\omega_\ell^{(m)}\}_1^M$ are the particle representation of $p(\theta_\ell|x^\ell, y^{\ell-1})$ which is equal to $p(\theta_\ell|x^{\ell-1}, y^{\ell-1})$. To proceed, in order to calculate $\lambda_{\ell+1}$ in the next step, we need to find the particle representation of $p(\theta_{\ell+1}|x^\ell, y^\ell)$. Assuming that

$$q(\theta_\ell|x^\ell, y^\ell) \equiv \{\tilde{\theta}_\ell^{(m)}, \tilde{\omega}_\ell^{(m)}\} \tag{3.74}$$

by (3.56), we have that

$$q(\theta_{\ell+1}|x^\ell, y^\ell) \equiv \{\tilde{\theta}_\ell^{(m)} + z_\ell^{(m)}, \tilde{\omega}_\ell^{(m)}\}. \tag{3.75}$$

To find $\tilde{\theta}_\ell^{(m)}$ and $\tilde{\omega}_\ell^{(m)}$, for any arbitrary function $g(\theta_\ell)$ we have that

$$E_{\theta_\ell}\left[g(\theta_\ell)|x^\ell, y^\ell\right] = \int_0^{2\pi} g(\theta_\ell) q(\theta_\ell|\underline{x}^\ell, \underline{y}^\ell)\,\mathrm{d}\theta_\ell \tag{3.76}$$

$$= \int_0^{2\pi} g(\theta_\ell) q(\theta_\ell|\underline{x}^\ell, \underline{y}^{\ell-1}) \frac{q(y_\ell|\underline{x}^\ell, \underline{y}^{\ell-1}, \theta_\ell)}{q(y_\ell|\underline{x}^\ell, \underline{y}^{\ell-1})}\,\mathrm{d}\theta_\ell \tag{3.77}$$

$$= E_{\theta_\ell}\left[g(\theta_\ell)\frac{q(y_\ell|\underline{x}^\ell, \underline{y}^{\ell-1}, \theta_\ell)}{q(y_\ell|\underline{x}^\ell, \underline{y}^{\ell-1})}|x^\ell, y^{\ell-1}\right] \tag{3.78}$$

$$\approx \sum_{m=1}^{M} \omega_\ell^{(m)} g(\theta_\ell^{(m)}) \frac{\beta_\ell^{(m)}}{\lambda_\ell} \tag{3.79}$$

where λ_ℓ is defined in (3.70) and

$$\beta_\ell^{(m)} = q(y_\ell|\underline{x}^\ell, \underline{y}^{\ell-1}, \theta_\ell^{(m)}) \tag{3.80}$$

$$= q(y_\ell|x_\ell, \theta_\ell^{(m)}). \tag{3.81}$$

From (3.79) we obtain

$$\tilde{\theta}_\ell^{(m)} = \theta_\ell^{(m)} \tag{3.82}$$

$$\tilde{\omega}_\ell^{(m)} = \omega_\ell^{(m)} \frac{\beta_\ell^{(m)}}{\lambda_\ell}. \tag{3.83}$$

Based on equations (3.73), (3.75), (3.82), and (3.83), the term $\frac{1}{L}\log q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})$, can be evaluated recursively. We initialize the recursion by taking q_{θ_1} to be the uniform distribution over $[0, 2\pi)$.

Algorithm 5: Calculating $\frac{1}{L}\log q_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\underline{y}|\underline{x})$ using particle approach

Data: Input vector: $\underline{x}$ of length L; output vector $\underline{y}$; parameters: σ_n^2, σ_z^2, h_0

Draw M samples uniformly from $[0, 2\pi)$ to produce $\{\theta_1^{(i)}\}_1^M$

$\omega_1^{(m)} \leftarrow 1/M \quad$ for $m \in \{1, \ldots, M\}$

for $\ell \in \{1, 2, \ldots, L\}$ **do**

  $\lambda_\ell \leftarrow \sum_{m=1}^{M} w_\ell^{(m)} q(y_\ell|x_\ell, \theta_\ell^{(m)})$

  $\omega_{\ell+1}^{(m)} \leftarrow \omega_\ell^{(m)} q(y_\ell|x_\ell, \theta_\ell^{(m)})/\lambda_\ell \quad$ for $m \in \{1, \ldots, M\}$

  $\theta_{\ell+1}^{(m)} \leftarrow \theta_\ell^{(m)} + z_\ell^{(m)} \quad$ for $m \in \{1, \ldots, M\}$

return $\frac{1}{L}\sum_{\ell=1}^{L} \log \lambda_\ell$

The calculation of the second term on the RHS of (3.38), $\frac{1}{L}\log q(\underline{y})$, can be performed with the same method. We consider two cases. If $\mathcal{X}$ is a finite set with small cardinality, the following algorithm describes the steps to calculate $\frac{1}{L}\log q(\underline{y})$.

Algorithm 6: Calculating $\frac{1}{L}\log q(\underline{y}|\underline{x})$ using particle approach

Data: Input distribution: $p_{\mathbf{x}}(x)$; output vector $\underline{y}$ of length L; parameters: σ_n^2, σ_z^2, h_0

Draw M sample uniformly from $[0, 2\pi)$ to produce $\{\theta_1^{(m)}\}_1^M$

$\omega_1^{(m)} \leftarrow 1/M \quad \text{for } i \in \{1, \ldots, M\}$

for $\ell \in \{1, 2, \ldots, L\}$ **do**

  $\gamma_\ell = \sum_{x \in \mathcal{X}} p(x) q(y_\ell | x, \theta_\ell^{(m)})$

  $\lambda_\ell \leftarrow \sum_{m=1}^{M} w_\ell^{(m)} \gamma_\ell$

  $\omega_{\ell+1}^{(m)} \leftarrow \omega_\ell^{(m)} \gamma_\ell / \lambda_\ell \quad \text{for } m \in \{1, \ldots, M\}$

  $\theta_{\ell+1}^{(m)} \leftarrow \theta_\ell^{(m)} + z_\ell^{(m)} \quad \text{for } m \in \{1, \ldots, M\}$

return $\frac{1}{L} \sum_{\ell=1}^{L} \log \lambda_\ell$

If the set $\mathcal{X}$ is not finite or is very large, one can modify Algorithm 6 to represent $p(x)$ with particles. The algorithm is presented as follows.

Algorithm 7: Calculating $\frac{1}{L}\log q(\underline{y}|\underline{x})$ using particle approach

Data: Input distribution: $p_{\mathbf{x}}(x)$; output vector $\underline{y}$; parameters: σ_n^2, σ_z^2, h_0

Draw M sample uniformly from $[0, 2\pi)$ to produce $\{\theta_1^{(m)}\}_1^M \quad \text{for } i \in \{1, \ldots, M\}$

$\omega_1^{(m)} \leftarrow 1/M$

for $\ell \in \{1, 2, \ldots, L\}$ **do**

  Draw M samples from $p_{\mathbf{x}}(x)$ to produce $\{x_\ell^{(m)}\}_1^M$

  $\lambda_\ell \leftarrow \sum_{i=1}^{M} w_\ell^{(m)} q(y_\ell | x_\ell^{(m)}, \theta_\ell^{(m)})$

  $\omega_{\ell+1}^{(m)} \leftarrow \omega_\ell^{(m)} q(y_\ell | x_\ell^{(m)}, \theta_\ell^{(m)}) / \lambda_\ell \quad \text{for } m \in \{1, \ldots, M\}$

  $\theta_{\ell+1}^{(m)} \leftarrow \theta_\ell^{(m)} + z_\ell^{(m)} \quad \text{for } m \in \{1, \ldots, M\}$

return $\frac{1}{L} \sum_{\ell=1}^{L} \log \lambda_\ell$

Finally, we note that in order to stabilize the numerical algorithms based on the particle approach, one should "resample" the particles after a few iterations. This procedure is explained in detail in [27].

CHAPTER 4

Optical Channel Models

In this chapter, we investigate the input–output relation of a discrete-time channel obtained by applying linear modulation and MFS demodulation to a linear and a nonlinear channel. Also, we derive the channel models considered in the appended papers. Finally, we review some known results on the capacity of optical channel models.

4.1 Modulation and Demodulation for Linear and Nonlinear Channels

In this section, we study two discrete-time channels based on applying a linear modulation and a MFS demodulation on *i)* the AWGN channel and *ii)* a nonlinear optical channel.

4.1.1 Matched Filtering and Sampling for a Linear Channel

Consider the complex-valued linear AWGN channel, where the input–output relation is described by

$$\mathbf{y}(t) = x(t) + \mathbf{n}(t). \tag{4.1}$$

Here, $\mathbf{n}(t)$ is a zero-mean circularly symmetric complex Gaussian noise with PSD N_0. With linear modulation, we have

$$x(t) = \sum_{\ell} x_\ell p(t - \ell T) \tag{4.2}$$

where x_ℓ are complex numbers and represent the information symbols generated by the source. The symbol period is given by T. Moreover, $p(t)$ is a real-valued function of

time t and represents the pulse shape which is assumed to satisfy the orthonormality condition

$$\int_{-\infty}^{\infty} p(t)p(t-lT)\,\mathrm{d}t = \begin{cases} 1 & l=0 \\ 0 & l\neq 0 \end{cases} \tag{4.3}$$

for all integers l. The transmitted power can be computed as

$$\lim_{T'\to\infty} \frac{1}{2T'} \int_{-T'}^{T'} |x(t)|^2\,\mathrm{d}t = \frac{1}{T} \lim_{N\to\infty} \frac{1}{2N} \sum_{\ell=-N}^{N} |x_\ell|^2. \tag{4.4}$$

To detect the transmitted symbols, the receiver performs demodulation to obtain a complex number per each time interval T. It is well-known that MFS is the optimal demodulation method for the linear AWGN channel. The output of this demodulator is obtained by calculating the inner product between the received signal and the time-shifted pulse shape, i.e.,

$$y_\ell = \int_{-\infty}^{\infty} y(t)p(t-\ell T)\,\mathrm{d}t \tag{4.5}$$

for all integers ℓ. The output y_ℓ can also be obtained by passing the signal through a filter with impulse response $p(-t)$ and sampling at time instances $t=\ell T$. After normalizing by $1/\sqrt{T}$, the resulting discrete-time channel is

$$\mathbf{y}_\ell = \mathbf{x}_\ell + \mathbf{n}_\ell \tag{4.6}$$

where $\mathbf{n}_\ell \sim \mathrm{iid}\,\mathcal{CN}(0, N_0/T)$.

4.1.2 Matched Filtering and Sampling for a Nonlinear Channel

For a linear channel, MFS provides a sufficient statistics for the detection process [28, Thm. 26.4.1], i.e., given the set $\{\mathbf{y}_\ell\}$, the received signal $\mathbf{y}(t)$ becomes independent of the transmitted signal. However, this is not the case for nonlinear channels such as a fiber. Here, we study the impact of the fiber nonlinearity on the performance of MFS. We assume that the pulse shape $p(t)$ is zero outside the interval $[0,T)$ and only consider the effects of nonlinearity on the signal. In this case, the output of the demodulator for the first symbol is (see (2.8))

$$y_0 = x_0 \int_0^T p^2(t) e^{j\gamma Z|x_0|^2 p^2(t)}\,\mathrm{d}t + \mathbf{n}. \tag{4.7}$$

where $\mathbf{n}$ is a complex Gaussian noise and Z is the fiber length. The value of y_0 depends on the shape of $p(t)$. If $p(t)$ is a rectangular pulse shape, i.e., $p(t)=1$, $0\le t\le T$, we obtain from (4.7) that $y_0 = x_0\exp(j\gamma Z|x_0|^2)+\mathbf{n}$. Therefore, the effects of nonlinearity appears as a phase shift in the discrete channel. In general, if $p(t)$ is not rectangular, the nonlinearity effects both the phase and the amplitude of the received signal. At high

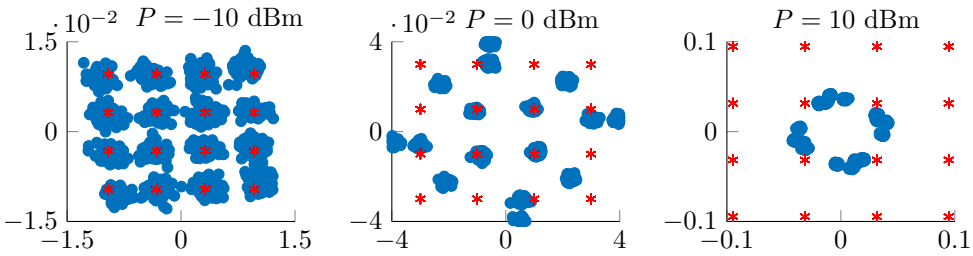

Figure 4.1: Scatter plots of the output of the MFS demodulator in the presence of Kerr nonlinearity described in (4.7).

powers, where $|x_0|^2 \gg 1$, the phase of the integrand in (4.7) changes quickly with time, which scales down the result of the integral, i.e., $|y_0| < |x_0|$.

Fig. 4.1 is the scatter plot of the output of MFS demodulator described in (4.7). It can be seen that at the input power of -10 dBm, the effects of nonlinearity is minor. The nonlinear distortion manifests itself as a phase noise in the discrete-time channel at $P = 0$ dBm. By increasing the power to 10 dBm, it can be seen that the nonlinearity affects both the amplitude and phase of the output. Therefore, MFS is not a suitable demodulation method for the fiber channel at high powers.

4.2 Perturbation Theory

For many nonlinear systems, a linearization around a working point can provide an approximation of the system's output. With the fiber-optical channel, this approach leads to perturbative channel models. The main assumption here is that the effects of nonlinear distortion are weak. Specifically, the solution to the NLS equation, with initial condition $\mathbf{a}(0,t)$, is approximated by

$$\mathbf{a}(z,t) \approx \mathbf{a}^L(z,t) + \Delta\mathbf{a}(z,t) \tag{4.8}$$

where $\mathbf{a}^L(z,t)$ is the solution to the NLS equation with $\gamma = 0$ and $\Delta\mathbf{a}(z,t)$ is a nonlinear perturbation. This approach of approximating the solution is commonly referred to as the regular perturbation (RP) method.

There are three main approaches for calculating $\Delta\mathbf{a}(z,t)$. The first one is to insert $\mathbf{a}^L(z,t) + \Delta\mathbf{a}(z,t)$ into the NLS equation and to neglect the nonlinear terms that include $\Delta\mathbf{a}(z,t)$ [29–31]. The second is to write the solution as a power series of the nonlinear parameter γ, i.e.,

$$\mathbf{a}(z,t) = \sum_{i=0}^{\infty} \gamma^i \mathbf{a}_i(z,t) \tag{4.9}$$

and then inserting (4.9) into the NLS equation to find out the signals $\mathbf{a}_i(z,t)$ [32, 33]. by equating coefficients of γ^0 one obtains $\mathbf{a}_0(z,t) = \mathbf{a}^L(z,t)$. The third method is to

use Volterra series, which approximate the NLS channel by a transfer function in a multi-dimensional frequency domain [34,35]. All these methods result in roughly similar channel models. In [33] it has been shown that the order n solution of the second method coincides with the order $2n+1$ result of Volterra series.

Now, we consider the channel model derived in [32, Sec. 3], where the second method is used to approximate the NLS equation. Let $\mathbf{b}_i(z,t) = \mathbf{a}_i(z,t)\exp(\alpha z/2)$ for all z. The first-order perturbative term, which is a cubic function of the optical field, can be calculated as [32, Eq. (12)]

$$\mathbf{b}_1(z,t) = j\int_0^z (|\mathbf{b}_0(z,t)|^2\mathbf{b}_0(z,t)) * h(z-\zeta,t)e^{-\alpha\zeta}\,\mathrm{d}\zeta \tag{4.10}$$

where $h(\,\cdot\,,\cdot\,)$ is defined in (2.6). Here, we consider the memoryless case when $\beta_2 = 0$, which is the case treated in Paper A. If the channel memory is neglected, the linear solution, $\mathbf{b}_0(z,t)$, can be obtained by setting $\beta_2 = \gamma = 0$ in (2.2) as $\mathbf{b}_0(z,t) = \mathbf{a}(0,t)$. The first perturbative term $\mathbf{b}_1$ in the memoryless case can be calculated by substituting $h(z,t) = \delta(t)$ into (4.10) to obtain

$$\mathbf{b}_1(z,t) = j|\mathbf{a}(0,t)|^2\mathbf{a}(0,t)\int_0^z e^{-\alpha\zeta}\,\mathrm{d}\zeta \tag{4.11}$$

$$= jZ_{\text{eff}}|\mathbf{a}(0,t)|^2\mathbf{a}(0,t) \tag{4.12}$$

where Z_{eff} is defined in (2.20). Therefore, the first-order perturbation theory approximates the solution of the memoryless NLS equation as

$$\mathbf{a}(z,t) \approx \left(\mathbf{a}(0,t) + j\gamma Z_{\text{eff}}|\mathbf{a}(0,t)|^2\mathbf{a}(0,t)\right)e^{-\alpha z/2}. \tag{4.13}$$

Another approach to obtain perturbative channel models is to use the logarithmic perturbation (LP) method to obtain a faster convergence [32,36]. This method can be viewed as applying the RP technique to the logarithm of the field. The normalized solution $\mathbf{b}(z,t) = e^{\alpha z/2}\mathbf{a}(z,t)$ is written as

$$\mathbf{b} = \mathbf{b}_0\exp\left(j\gamma\sum_{i=0}^{\infty}\gamma^i\mathbf{b}_i^{\text{lp}}\right) \tag{4.14}$$

where $\mathbf{b}_0 = \mathbf{a}^L\exp(\alpha z/2)$ and $\mathbf{b}_i^{\text{lp}}$ can be found by substituting (4.14) into the NLS equation. Doing so, one can show that [36, Eq. (10b)]

$$\mathbf{b}_0^{\text{lp}} = \frac{1}{\mathbf{b}_0}\int_0^z \left(|\mathbf{b}_0|^2\mathbf{b}_0\right) * h(z-\zeta,t)e^{-\alpha\zeta}\,\mathrm{d}z. \tag{4.15}$$

For the memoryless case, $h(z,t) = \delta(t)$, we obtain

$$\mathbf{b}_0(z,t) = \mathbf{a}(0,t) \tag{4.16}$$

$$\mathbf{b}_0^{\text{lp}}(z,t) = Z_{\text{eff}}|\mathbf{a}(0,t)|^2. \tag{4.17}$$

Therefore, the first-order LP results in the approximate solution

$$\mathbf{a}(z,t) \approx \mathbf{a}(0,t)e^{-\alpha z/2}e^{j\gamma Z_{\text{eff}}|\mathbf{a}(0,t)|^2}. \tag{4.18}$$

Observe that (4.13) and (4.18) are equal up to a first-order linearizion. A combination of regular and logarithmic approach is used in [37] to obtain a more accurate approximation.

4.3 A Memoryless Optical Channel

In this section we neglect the effects of dispersion to obtain a memoryless channel model. If the dispersion is set to zero, i.e., $\beta_2 = 0$ in (2.2), we get

$$\frac{\partial \mathbf{a}}{\partial z} - j\gamma|\mathbf{a}|^2\mathbf{a} + \frac{\alpha}{2}\mathbf{a} = 0. \tag{4.19}$$

This equation can be solved analytically to obtain [9, Sec. 4.1.1]

$$\mathbf{a}(z,t) = \mathbf{a}(0,t)e^{-\alpha z/2}e^{j\gamma Z_{\text{eff}}|\mathbf{a}(0,t)|^2} \tag{4.20}$$

where Z_{eff} is defined in (2.20).

To obtain a channel model for memoryless lumped amplified systems, we can use (4.20) to describe the signal propagation over a single span, before amplification. At the end of each span, an amplifier multiplies the signal by $\exp(\alpha Z/2)$ and adds a Gaussian noise $\mathbf{n}(t)$, as described in Section 2.2.1. We denote the signal at the end of the kth span and after amplification by $\mathbf{a}_k = \mathbf{a}(kZ_{\text{sp}}, t)$ for $k = 0, \ldots, K$. The output of the memoryless channel model, $\mathbf{a}_K$, can be obtained by iterating the following step for $k = 0, \ldots, K-1$

$$\mathbf{a}_{k+1} = \mathbf{a}_k e^{j\gamma Z_{\text{eff}}|\mathbf{a}_k|^2} + \mathbf{n}_k \tag{4.21}$$

where $\mathbf{n}_k(t)$ is the noise added by the kth amplifier.

A similar channel model can be developed for a distributed amplification system by letting the number of spans N go to infinity for a fixed system length, Z. The channel output, $\mathbf{a}_N$ can be obtained by iterating the following equation for $n = 0, \ldots, N$

$$\mathbf{a}_{n+1} = \mathbf{a}_n e^{j\gamma Z/N|a_n|^2} + \mathbf{n}_n \tag{4.22}$$

where $\mathbf{n}_n$ describes the amplifier ASE noise discussed in (2.22). This model accurately describes the nondispersive NLS channel when $N \to \infty$. A discrete-time version of (4.22) is studied in Paper A.

4.4 Channel Models in Paper A

As was illustrated in Section 4.1.1, the effects of nonlinearity prevents the development of a tractable discrete-time model attained by linear filtering and sampling at the receiver. To obtain discrete-time models from nonlinear continuous channels, in Paper A,

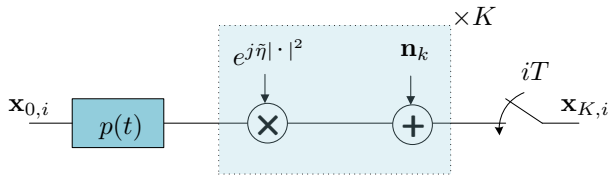

Figure 4.2: Discrete-time memoryless optical channel model with K spans of lumped amplification.

we assume that samples are taken at symbol rate at the receiver. It is assumed that the noise is band-limited because of the use of inline optical filters. Using these assumptions, the input–output relation of the discrete-time channels resembles that of the continuous-time ones. This method of developing discrete-time optical channels has been used extensively throughout the literature (see [38–40] for example). The importance of designing demodulation methods that are compatible with the nonlinear nature of the optical fiber is highlighted in Paper B.

In Paper A, three discrete-time optical channels were considered, which have been developed using different methods. All these models represent the same fiber-optical channel under different assumptions. Although the simplifying assumptions used in these methods are only valid in the low-power regime, it is not uncommon to use these models to draw conclusions at high powers. We analyze the capacity of these models for the special case of zero dispersion to study the effects of the simplifying assumptions on the capacity. The results indicate that although these models represent the same physical channel, their capacities are drastically different in the high-power regime. Specifically, their capacities grow according to different pre-logs. Here, the capacity pre-log is defined as $\lim_{P\to\infty} \mathcal{C}/\log P$, where P is the average input power and $\mathcal{C}$ is the capacity of the channel.

All the three models in Paper A describe the same physical optical system and share the same set of parameters. A distributed-amplified optical system with length Z is considered. The loss is assumed to be completely compensated for by the amplifiers. The discrete-time model is obtained from the continuous-time channel using the sampling receiver. The three considered channels can be found in the following list.

1. *Regular perturbative channel (RPC)* is a discrete-time distributed-amplified system with length Z, which is based on the RP method. The signal–noise interaction is neglected and all the noise is added at the receiver. Therefore, the channel model can be obtained from (4.13) by letting $\alpha \to 0$ and $Z_{\text{eff}} = Z$ as

$$\mathbf{y} = \mathbf{x} + j\eta|\mathbf{x}|^2\mathbf{x} + \mathbf{n} \tag{4.23}$$

where the nonlinearity is quantified by $\eta = Z\gamma$. The amplification noise is denoted by $\mathbf{n}$, which is complex circularly symmetric Gaussian with zero mean

and variance P_N defined in (2.15).

2. *Logarithmic perturbative channel (LPC)* can be obtained from the LP channel in (4.18). It reads
$$\mathbf{y} = \mathbf{x}\exp(j\eta|\mathbf{x}|^2) + \mathbf{n} \tag{4.24}$$
where η and $\mathbf{n}$ are the same as those used in RPC.
3. *Memoryless NLS Channel (MNC)* is a discrete-time model of the continuous-time channel presented in (4.22). It can be described as a concatenation of K segments. Denoting the input of the MNC by $\mathbf{x}_0$, its output $\mathbf{x}_K$ can be obtained by iterating the following equation for $k = 0, \dots, K-1$
$$\mathbf{x}_{k+1} = \mathbf{x}_k \exp(j\tilde{\eta}|\mathbf{x}_k|^2) + \mathbf{n}_k. \tag{4.25}$$
Here, $\tilde{\eta} = \eta/K$ and $\mathbf{n}_k$ are independent complex circularly symmetric Gaussian noise with zero mean and variance P_N/K. Fig. 4.2 illustrates the input–output relation in this model, where $\mathbf{x}_{0,i}$ and $\mathbf{x}_{K,i}$ denote the ith input and output symbols, respectively. This channel describes the signal propagation in the memoryless NLS channel (MNC) and has been studied extensively in the literature (see [38, 39, 41] for some examples).

4.5 Channel Model in Paper B

In Paper B, we consider a simplified model of a two-user WDM network, which has been introduced in Section 2.5. If the dispersion and group velocity mismatch in (2.27)–(2.28) are neglected ($d = \beta_{21} = \beta_{22} = 0$), these equations have the analytical solution

$$\mathbf{a}_1(Z,t) = \mathbf{a}_1(0,t)\exp(-\frac{\alpha}{2})\exp\big(j\phi_1^{\mathrm{nl}}\big) \tag{4.26}$$
$$\mathbf{a}_2(Z,t) = \mathbf{a}_2(0,t)\exp(-\frac{\alpha}{2})\exp\big(j\phi_2^{\mathrm{nl}}\big) \tag{4.27}$$

where

$$\phi_1^{\mathrm{nl}} = \gamma_1 Z_{\mathrm{eff}}\left(|\mathbf{a}_1(0,t)|^2 + 2|\mathbf{a}_2(0,t)|^2\right) \tag{4.28}$$
$$\phi_2^{\mathrm{nl}} = \gamma_2 Z_{\mathrm{eff}}(\underbrace{|\mathbf{a}_2(0,t)|^2}_{\mathrm{SPM}} + \underbrace{2|\mathbf{a}_1(0,t)|^2}_{\mathrm{XPM}}) \tag{4.29}$$

where Z_{eff} is defined in (2.20). In Paper B, we consider a lumped amplified system with K spans, where a simplified model is obtained from (4.26) and (4.27) by multiplying the phase shifts, ϕ_1^{nl} and ϕ_2^{nl}, by K and adding all the amplification noise at the receiver.

A discrete-time network based on this model is studied in [42–44]. We show that by using nonrectangular pulse shaping and proper demodulation technique, the XPM distortion can be effectively eliminated. However, when MFS is used as a demodulation scheme, because of SPM and XPM distortions, the error probability increases with power in the high-power regime.

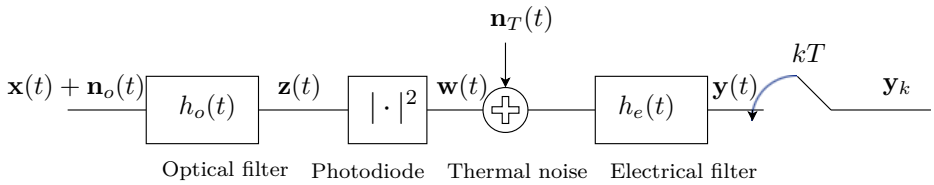

Figure 4.3: A schematic of a noncoherent optical receiver.

4.6 Noncoherent Optical Channel

Noncoherent transmission systems are often used in short-haul optical systems, wireless optical communication, and data center networks. These systems are attractive since they reduce the complexity and the cost of data transmission. Since the propagation distance in noncoherent systems is short, the nonlinear distortion caused by the fiber is negligible. The effects of dispersion can also be neglected if the fiber length is much smaller than the dispersion length. Otherwise the dispersion is compensated for either by DCFs or by predistorting the signal at the transmitter.

Fig. 4.3 depicts a noncoherent receiver with optical and electrical filters. The received signal is corrupted by the white circularly-symmetric Gaussian noise $\mathbf{n}_o(t)$, whose PSD is N_o. The source of this noise in the case of short-haul fiber links is either optical amplification or the cross-talk from neighboring channels; in the case of wireless optics it is the ambient light [45]. After filtering the signal in the optical domain, $\mathbf{z}(t) = \mathbf{x} * h_o(t) + \mathbf{n}_o * h_o(t)$ is obtained. The output of the photodiod is then $\mathbf{w}(t) = R|\mathbf{z}(t)|^2$. Here, $R = \zeta q/(h\nu) \approx 1.25\zeta$ (A/W) is the photodetector responsivity [10, Eq. (4.1.3)], where $\zeta \leq 1$ is quantum efficiency and q is the electron charge. Then, the thermal noise $\mathbf{n}_T(t)$ is added to the signal, which is a real white Gaussian noise with PSD $N_T = 2k_bT/R_L$, where $k_b = 1.38 \cdot 10^{-23}$ (J/K) is the Boltzman's constant, T is the temperature in Kelvin, and R_L is the receiver resistance. The origin of the thermal noise is the random movement of electrons, which is present in any conductor. By adding the thermal noise to $\mathbf{w}(t)$ and passing it through the electrical filter, we obtain

$$\mathbf{y}(t) = \left[R|\mathbf{x} * h_o(t) + \mathbf{n}_o * h_o(t)|^2\right] * h_e(t) + \mathbf{n}_T * h_e(t). \tag{4.30}$$

After electrical filtering $\mathbf{y}_k$ is obtained by sampling at the symbol rate.

4.6.1 Channel Model in Paper D

To go from the continuous-time channel (4.30) to a discrete-time one, we consider linear modulation

$$\mathbf{x}(t) = \sum_{\ell} \mathbf{x}_\ell p(t - \ell T). \tag{4.31}$$

Here, $\mathbf{x}_\ell$ is the kth symbol, T is the symbol time, and $p(t)$ is the pulse shape that is band-limited to W Hz. Also, we assume that $p(t)$ satisfies the Nyquist criterion, i.e.,

$$p(\ell T) = \begin{cases} p(0), & \ell = 0 \\ 0, & \ell \neq 0. \end{cases} \tag{4.32}$$

For a general case of $h_o(t)$ and $h_e(t)$, the resulting discrete-time channel between $\mathbf{y}_k$ and $\mathbf{x}_k$ has memory. In Paper D we consider a special choice of the two filters, which removes the channel memory, namely

$$h_o(t) = W \operatorname{sinc}(Wt) \tag{4.33}$$

$$h_e(t) = 2W \operatorname{sinc}(2Wt) \tag{4.34}$$

where $\operatorname{sinc}(t) = \sin(\pi t)/(\pi t)$. In other words, we apply brick-wall filters with bandwidths W and $2W$. Since $h_o(t)$ is band-limited to W, $\mathbf{w}(t) = R|\mathbf{x} * h_o(t) + \mathbf{n}_o * h_o(t)|^2$ is band-limited to $2W$ and passes through the electrical filter without distortion, i.e., $\mathbf{w} * h_e(t) = \mathbf{w}(t)$. Therefore, the received signal is not distorted by optical or electrical filters. By sampling the received signal at the symbol rate the following discrete-time channel model is obtained

$$\mathbf{y} = |\mathbf{x} + \mathbf{n}_o|^2 + \mathbf{n}_T. \tag{4.35}$$

Here, $\mathbf{n}_o \sim \mathcal{CN}(0, 2N_o W R)$ and $\mathbf{n}_{T,k} \sim \mathcal{N}(0, 4N_T W)$. In Paper D we study the capacity of the channel model (4.35) under the power constraint $\mathrm{E}[|\mathbf{x}|^2] \leq P$.

4.7 Known Results on the Capacity of Fiber-Optical Channel Models

In this section, we present a literature review on the capacity studies of coherent and noncoherent optical transmission systems.

4.7.1 Results on coherent optical transmission

The fiber-optical system described by the NLS equation is a continuous-time channel with memory. There are two main difficulties in studying the spectral efficiency of this channel. The first is to obtain a discrete-time model. When linear modulation and demodulation are applied, because of the nonlinear distortion, the output of the demodulator will be a nonlinear function of all the input symbols, which is hard to formalize. Some efforts to obtain an approximation of the discrete-time channel have been made, which have lead to the perturbative models (see Section 4.2). The second is that because of nonlinearity, the spectrum of the signal changes as it propagates through the fiber, which makes it difficult to define the channel bandwidth and apply the sampling theorem.

A number of lower bounds on the capacity of a variety of optical channel models have been proposed, many of which are based on mismatched decoding [46] (see for example [47–52]). On the other hand, only one upper bound is known as yet [5], which indicates that the capacity, in bits per channel use, of a discrete-time NLS channel cannot exceed that of the AWGN one. Here, "channel use" is referred to the elements of the sample vector described in Section 4.7.1; this indicates that the upper bound gets looser as the sampling time Δ_t becomes smaller. A recent review on the capacity results is available in [6].

There are different optical communication systems studied in the literature. An optical system can be lumped- or distributed-amplified and single- or dual-polarized. Also, WDM systems or a single channel can be considered. Furthermore, different models can be obtained using various simplifying assumptions. All these models may be studied with or without memory (dispersion). Here, we categorize some capacity results based on different types of channel models. We consider the three categories mentioned in [7], namely, perturbative models, Gaussian-noise models, and the memoryless NLS model. In addition, we also review some capacity results based of the SSF method and AIR computation based on experimental data.

Perturbative models As discussed in Section 4.2, in this class of optical channels, the nonlinear distortion is modeled either by an additive perturbative term or by a signal-dependent phase distortion. The former is referred to as RP and the latter as LP. A channel model based on first-order RP theory was developed in [30] and lower bounds on its capacity were derived analytically. In [50] an LP channel model was derived and analytical lower bounds on the capacity of this channel were obtained by calculating AIR for a mismatched receiver. A novel multivariate RP channel model is derived in [53], in which the effects of signal–signal interaction as well as signal–noise interaction are taken into account. Using LP channel models, AIRs were evaluated in [49, 54–56] to illustrate the effects of XPM mitigation techniques on the capacity lower bounds. In [35] the Volterra series is used to derive a WDM channel model. It is shown that if joint processing is possible, the effects of nonlinear distortion on the capacity region is minimal. Authors in [57] study the capacity of a WDM system, where the effects of FWM is modeled by a RP term. It is shown that different assumptions on the behaviour of the interfering channels have a profound impact on the capacity.

Gaussian-noise models In this class of optical channel models, the nonlinear distortion is treated as an additive Gaussian noise, whose variance is determined based on signal power and channel parameters (see for example [58–60]). In [61, 62], the impact of FWM distortion in WDM systems was modeled as a Gaussian noise. The mismatched-decoding capacity lower bound based on this model has a maximum and then decreases with power. By considering the cross-phase distortions as an additive noise, in [48] the sum capacity of WDM networks was lower-bounded by the capacity of an equivalent AWGN channel

with the same covariance matrix. This lower bound goes to zero as the input power grows large. In [63], the capacity of the Gaussian-noise model presented in [58, 59, 64] was studied and the maximum value of the rates achievable based on these models was calculated. In [65] it was shown that the capacity of a finite-memory Gaussian-noise channel is nondecreasing.

Memoryless NLS model Unlike the general case of the NLS equation, the input–output relation of a MNC can be derived in closed form [40, 41, 66]. Using this closed-form channel law, lower bounds on capacity have been derived to prove that the capacity increases to infinity with power [40, 66]. In [67] a lower bound on the capacity of this channel has been derived via machine learning. An upper bound on the capacity was established in [68], which together with lower bounds allow characterizing the capacity. Receiver and constellation design for this channel is studied in literature extensively (see for example [69–72]). All the aforementioned works on the capacity of the MNC assume sampling receiver, which has infinite bandwidth. By calculating the autocorrelation function of the channel output, the capacity with band-limited receivers was upper-bounded in [73].

Split-step Fourier model The SSF method has been used in the literature with two different approaches to obtain bounds on the capacity. The first is to use the SSF method to obtain a channel model, whose input and output are complex vectors [5, 74]. In [5] it was proven that the capacity of the SSF channel in bits per sample is upper-bounded by $\log(1 + \mathrm{SNR})$, where SNR is the signal-to-noise ratio. If one sample is taken from each symbol, this upper bound coincides with the capacity of an equivalent linear AWGN channel. However, at high powers, where more than one sample per symbol is required to assure the accuracy of the SSF method, the upper bound diverges from the capacity of the linear channel. The second is to conduct Monte Carlo simulations to obtain channel statistics in order to evaluate lower bounds on the capacity (see for example, [18, 51, 75, 76]). In [51] different mismatched lower bounds based on different receivers have been evaluated using SSF simulation. An auxiliary backward channel is used in [76] to obtain a lower bound that is evaluated by the SSF method.

AIR with experimental data An strong aspect of AIR is that it can be used to derive lower bounds on the capacity of physical channels using experimental data. The main benefit of this method is that all the imperfections of the physical channel are taken into consideration. This method has been used in [77–80] with circularly-symmetric complex Gaussian distribution as auxiliary channel.

4.7.2 Results on noncoherent optical transmission

In this section, we review some of the existing work on channel modeling and capacity evaluation of the noncoherent optical system shown in Fig. 4.3. Three categories are

considered based on the studied channel models, namely, with memory and no electrical noise, memoryless without electrical noise, memoryless without optical noise. In paper D we study the memoryless noncoherent channel in the presence of both electrical and optical noises. To the best our knowledge this is the first work to consider the effect of both noises on the capacity.

Channel models with memory and without electrical noise Many studies investigated the derivation of the discrete-time input–output relation of the channel in Fig. 4.3 with no electrical noise and arbitrary choices of optical and electrical filters to calculate the BER (see for example [81–83]). In [84] a review of these methods is provided and their complexity is compared. A compact formula for the channel model with memory can be found in [84, Eq. (72)]. To the best of our knowledge, the capacity of this channel has not been studied to date.

Memoryless models without thermal noise This channel model is (4.35) with $\mathbf{n}_T = 0$, i.e.,

$$\mathbf{y} = |\mathbf{x} + \mathbf{n}_o|^2. \tag{4.36}$$

In [85, Sec. 11.2] this channel is regarded as intensity-modulation direct-detection optical channel. Some bounds on the capacity as well as capacity approximations are derived. The capacity is studied in [86] and numerically evaluated in [87]. Most of the studies on the capacity of this channel can be found in the context of phase-noise channel [88–91], that is $\mathbf{y} = \mathbf{x}e^{j\theta} + \mathbf{n}$, where $\mathbf{y}$ and $\mathbf{x}$ are input and output of the channel and $\theta \sim$ uniform$[0, 2\pi)$ and $\mathbf{n} \sim \mathcal{CN}(0, P_N)$ are independent noise sources. Since with the phase-noise channel the amplitude of $\mathbf{y}$ is sufficient statistics, it has the same capacity as (4.36). Upper bounds on the capacity are derived using duality technique [21] in [88, 91] and [89] (memoryless case) as well as some lower bounds. In [91], it is proved that the capacity-achieving distribution is discrete with infinite number of mass points. In [88] it is shown that the half normal distribution asymptotically achieves the capacity and that the capacity behaves as $\mathcal{C} = (1/2)\log(\mathrm{SNR}/2)$ at high powers. In [92] the authors show that, if enough samples are taken from the received signal, the capacity of the direct-detection channel is more than the capacity of the equivalent coherent channel minus one bit.

Memoryless models without optical noise This channel model is (4.35) with $\mathbf{n}_o = 0$, i.e.,

$$\mathbf{y} = |\mathbf{x}|^2 + n_T. \tag{4.37}$$

In [85, Sec. 11.2.3] this channel is regarded as thermal noise dominated intensity-modulation direct-detection optical channel and its capacity is estimated. A continuous-time version of (4.37) is presented in [45, Eq. (1)] for wireless infrared communication with strong ambient or thermal noise. The capacity of this channel with multiple subcarriers has

been upper-bounded in [93]. In [94] asymptotically tight bounds are derived on the capacity of the continuous-time bandwidth-limited channel. Constellation design is performed in [95]. The capacity of (4.37) is tightly bounded in low- and high-power regimes in [96]. Bounds on the capacity of multiple-input multiple-output [97], multiple-input single-output [98,99], and input-dependent noise [100] intensity-modulated optical channels have been derived, where (4.37) can be regarded as a special case of these studies.

CHAPTER 5

Mitigation of Fiber Impairments

In this chapter we study different methods to compensate for the effects of CD and fiber Kerr nonlinearity in Section 5.1 and Section 5.2, respectively.

5.1 Chromatic Dispersion Compensation

The CD distorts the signal severely after a short distance [10, Sec. 7.1]. It can be compensated for either in the optical or in the electrical domain. With the latter method, the effects of CD are compensated for at the receiver via digital signal processor by applying a filter with frequency response $\exp(-2j\pi^2 f^2 \beta_2 Z)$, where Z is the fiber length (see [101] for a recent work on this topic). The optical-domain compensation of CD is performed either through the whole spectrum via DCFs [10, Sec. 7.4] or locally within the bandwidth of each channel via fiber Bragg grating [102]. We compare these three methods in terms of AIR in Paper C.

5.2 Nonlinearity Mitigation Methods

The fiber nonlinearity is the central impairment of optical transmission systems and limits the achievable data rate. Here, we list a number of fiber-nonlinearity mitigation techniques that are implemented in the optical or digital domains, and also, we add our contribution at the end.

The inverse scattering transform [103, 104] is among the first and yet most powerful nonlinearity mitigation techniques. It is based on the fact that some pulses, which

are called solitons, can propagate through the lossless NLS channel without any distortion. Specifically, during the propagation of a soliton, the nonlinear effects are effectively canceled by the linear ones. Developing a communication system based on soliton transmission has received a great deal of attention in the recent years [105–109] under the name of nonlinear Fourier transform. The performance of this approach can be significantly affected by random transmission effects such as polarization-mode dispersion and the fiber loss in lumped amplified systems [110].

Digital backpropagation (DBP) is another important nonlinearity compensation approach which is based on the SSF method. If the effects of noise are ignored, the deterministic distortion of the NLS channel can be completely compensated for using the SSF method with inverse parameters $(-\beta_2, -\gamma, -\alpha)$ at the receiver, or transmitter, or both [111–116]. DBP only compensates for the deterministic signal–signal distortions. To mitigate the effects of signal–noise interaction, stochastic DBP can be implemented [117]. DBP suffers a high computational complexity. For the systems with inline DC, the complexity of DBP can be reduced by implementing folded DBP [118, 119].

Optical phase conjugation (OPC) was first suggested in [120] as a method for DC, where the phase of the signal is inverted at the middle of the transmission line. Here, we provide a brief justification of this method. The origin of chromatic dispersion is the difference between the group velocities of the high and low frequency components of the signal. After phase conjugation, the high-frequency components change their place with the low-frequency ones and therefore they travel with the same average speed, which mitigates the dispersion. In [121] it was pointed out that OPC can also be used as an effective way of compensating nonlinearity. Since then, OPC has received attention as an effective approach to reduce the effects of fiber nonlinearity [122–125].

Nonlinearity-tailored detection techniques use an approximation of the optical input–output joint pdf in order to implement detectors based on the maximum a posteriori or the maximum likelihood criteria [126–128]. In this approach the signal is sampled at the receiver after a low-pass filter and the sample vector is fed to a Viterbi algorithm for detection. In [129] it is shown that by setting the bandwidth of the receiver filter larger than that of transmitted signal and oversampling the received signal, higher rates can be achieved compared to matched filtering.

XPM compensation by equalization: The XPM distortion varies slowly with time since it is the aggregation of the interference of many channels. Exploiting this property, the XPM can be mitigated by deploying adaptive or turbo equalization [130–132].

Our contribution in Paper B is to introduce a novel SPM and XPM mitigation method by using a novel demodulation technique. This method exploits the temporal correlation of the aforementioned distortions during one time slot, which has not been considered as yet. In one time slot, the distortion generated by SPM and XPM are correlated and dependent on the pulse shape. This property can be exploited to differentiate these distortions from noise and effectively mitigate their effect.

CHAPTER 6

Summaries of the Appended Papers

In this chapter we provide a summary of the four appended papers.

6.1 Paper A

In Paper A we investigate the accuracy of two perturbative models by studying their capacity for the zero-dispersion case. These models are the *regular perturbative channel (RPC)* and the *logarithmic perturbative channel (LPC)*. We also study the *memoryless NLS channel (MNC)* that is described in Section 4.4. Comparing their capacity, three different capacity pre-logs are established for these models: 3 for RPC, 1 for LPC, and 1/2 for MNC. This shows that the two perturbative channels, RPC and LPC, are grossly inaccurate in the high-power regime. Therefore, care should be exercised in interpreting the high-power results that have been established using these models.

Contributions: Kamran Keykhosravi (KK) derived the analytical bounds, numerically evaluated them, analyzed the results, and wrote the paper. Erik Agrell (EA) formulated the problem and contributed to the analysis. Giuseppe Durisi (GD) contributed to the information-theoretic analyses. All authors reviewed and revised the paper.

Context: Section 4.4.

6.2 Paper B

In Paper B we show that the effects of XPM can be effectively compensated for by exploiting the time coherence of the XPM distortion during one symbol period. This

method of XPM mitigation is introduced for a two-user simplified memoryless network. A new demodulation scheme named maximum matching is introduced. Also, the optimal receiver, based on maximum a posteriori (MAP) detection is developed. The performance of these two receivers was compared with the MFS method, using the SER as a metric. Unlike with MFS, the SER with maximum matching and MAP receiver goes to zero at high powers.

Contributions: KK formulated the problem, derived the analytical results, conducted the simulations, analyzed the results, and wrote the paper. All authors reviewed and revised the paper.

Context: Section 4.5.

6.3 Paper C

In Paper C we conduct a comparison between three WDM links: CDM, NDM, and DM. All inband signal–signal distortions are compensated for via DBP and therefore, XPM is the dominant impairment. The performance is assessed based on AIR with two different auxiliary channels: AWGN and AR phase-noise channel (see (3.55)). We show that if the effects of XPM are compensated for (with AR auxiliary channel), CDM links outperform the NDM ones, however, if not (with AWGN auxiliary channel) NDM prevails. The DM link has the worst performance in both cases. These results are explained by resorting to the frequency-resolved LP model [50] and deriving the autocorrelation function of the XPM phase noise.

Contributions: KK derived the analytical results, conducted the simulations, analyzed the results, and wrote the paper. MS formulated the problem and contributed in conducting the simulations and analyzing the results. All authors reviewed and revised the paper.

Context: Sections 3.5 and 4.5.

6.4 Paper D

In Paper D we investigate the capacity of the noncoherent channel (4.35). We derive an upper bound using the duality technique (see Section 3.4.4) and we establish a lower bound. A capacity estimation is also made via Blahut–Arimoto algorithm. Our upper and lower bounds improve on the results of [91] for the special case of $\mathbf{n}_T = 0$, and [96] for the special case of $\mathbf{n}_o = 0$. Our bounds also characterize the capacity of the general case with both noises for the first time. Our results suggest that optical preamplification is beneficial at low received signal powers and detrimental at high powers. The boundary between these two regions is determined for a wide range of channel parameters. Furthermore, it is illustrated that in some ranges of channel parameters, applying optical amplification with a finite gain is optimal.

Contributions: KK derived the capacity bounds and numerically evaluated them, analyzed the results, and wrote the paper. EA proposed the research problem. MS contributed in numerical evaluation of the bounds. MK provided expertise in physical relevance of the channel model. All authors reviewed and revised the paper.

Context: Section 4.6.1.

Bibliography

[1] F. Kapron, D. Keck, and R. Maurer, "Radiation losses in glass optical waveguides," *Appl. Phys. Lett.*, vol. 17, no. 10, pp. 423–425, Nov. 1970.

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

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

[4] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423/623–656, Jul.–Oct. 1948.

[5] G. Kramer, M. I. Yousefi, and F. R. Kschischang, "Upper bound on the capacity of a cascade of nonlinear and noisy channels," in *IEEE Info. Theory Workshop (ITW)*, Apr-May 2015.

[6] M. Secondini and E. Forestieri, "Scope and limitations of the nonlinear Shannon limit," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 893–902, Apr. 2017.

[7] E. Agrell, G. Durisi, and P. Johannisson, "Information-theory-friendly models for fiber-optic channels: A primer," in *IEEE Info. Theory Workshop (ITW)*, 2015.

[8] P. Wai, C. R. Menyuk, and H. Chen, "Stability of solitons in randomly varying birefringent fibers," *Opt. Lett.*, vol. 16, no. 16, pp. 1231–1233, Aug. 1991.

[9] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed. New York: Academic Press, 2007.

[10] ——, *Fiber-optic communication systems*, 3rd ed. New York, NY: John Wiley, 2002.

[11] G. C. Papen and R. E. Blahut, *Lightwave Communications*. Cambridge University Press, 2019.

[12] G. P. Agrawal, *Fiber-Optic Communication Systems*, 4th ed. John Wiley & Sons, 2010.

[13] Q. Zhang and M. I. Hayee, "Symmetrized split-step Fourier scheme to control global simulation accuracy in fiber-optic communication systems," *J. Lightw. Technol.*, vol. 26, no. 2, pp. 302–316, Jan. 2008.

[14] E. Agrell, "Capacity bounds in optical communications," in *Proc. European Conference on Optical Communication (ECOC)*, Gothenburg, Sweden, Sep. 2017.

[15] G. P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. San Diego, Calif.: Academic Press, 2001.

[16] A. Shabat and V. Zakharov, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media," *Soviet physics JETP*, vol. 34, no. 1, pp. 62–69, Jan. 1972.

[17] M. Desaix, D. Anderson, and M. Lisak, "Accuracy of an approximate variational solution procedure for the nonlinear Schrödinger equation," *Physical review A*, vol. 40, no. 5, p. 2441, Sep. 1989.

[18] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity limits of information transport in fiber-optic networks," *Phys. Rev. Lett.*, vol. 101, no. 16, p. 163901, Oct. 2008.

[19] F. Forghieri, R. Tkach, A. R. Chraplyvy, and D. Marcuse, "Reduction of four-wave mixing crosstalk in WDM systems using unequally spaced channels," *IEEE Photon. Technol. Lett.*, vol. 6, no. 6, pp. 754–756, Jun. 1994.

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

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

[22] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[23] A. Lapidoth and S. M. Moser, "Convex-programming bounds on the capacity of flat-fading channels," in *Proc. IEEE Int. Symp. Inform. Theory*, Washington, D. C., USA, Jun. 2001, p. 52.

[24] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[25] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[26] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcăicá, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[27] J. Dauwels and H. Loeliger, "Computation of information rates by particle methods," *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 406–409, Jan. 2008.

[28] A. Lapidoth, *A foundation in digital communication.* Cambridge University Press, 2009.

[29] R. Holzlohner, V. Grigoryan, C. Menyuk, and W. Kath, "Accurate calculation of eye diagrams and bit error rates in optical transmission systems using linearization," *J. Lightw. Technol.*, vol. 20, no. 3, pp. 389–400, Mar. 2002.

[30] A. Meccozzi and R.-J. Essiambre, "Nonlinear Shannon limit in pseudolinear coherent systems," *J. Lightw. Technol.*, vol. 30, no. 12, pp. 2011–2024, Jun. 2012.

[31] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, and J. C. Rasmussen, "Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate," *J. Lightw. Technol.*, vol. 29, no. 17, pp. 2570–2576, Sep. 2011.

[32] E. Forestieri and M. Secondini, "Solving the nonlinear Schrödinger equation," in *Optical Communication Theory and Techniques.* New York, NY: Springer, 2005.

[33] A. Vannucci, P. Serena, and A. Bononi, "The RP method: A new tool for the iterative solution of the nonlinear Schrödinger equation," *J. Lightw. Technol.*, vol. 20, no. 7, pp. 1102–1112, Jul. 2002.

[34] B. Xu and M. Brandt-Pearce, "Comparison of FWM and XPM-induced crosstalk using the Volterra series transfer function method," *J. Lightw. Technol.*, vol. 21, no. 1, p. 40, Jan. 2003.

[35] M. H. Taghavi, G. C. Papen, and P. H. Siegel, "On the multiuser capacity of WDM in a nonlinear optical fiber: Coherent communication," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5008–5022, Nov. 2006.

[36] E. Ciaramella and E. Forestieri, "Analytical approximation of nonlinear distortions," *IEEE Photon. Technol. Lett.*, vol. 17, no. 1, pp. 91–93, Jan. 2005.

[37] M. Secondini, E. Forestieri, and C. R. Menyuk, "A combined regular-logarithmic perturbation method for signal-noise interaction in amplified optical systems," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3358–3369, 2009.

[38] L. Beygi, E. Agrell, M. Karlsson, and P. Johannisson, "Signal statistics in fiber-optical channels with polarization multiplexing and self-phase modulation," *J. Lightw. Technol.*, vol. 29, no. 16, pp. 2379–2386, Aug. 2011.

[39] C. Häger, L. Beygi, E. Agrell, P. Johannisson, M. Karlsson, and A. Graell i Amat, "A low-complexity detector for memoryless polarization-multiplexed fiber-optical channels," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 368–371, Feb. 2014.

[40] M. I. Yousefi and F. R. Kschischang, "On the per-sample capacity of nondispersive optical fibers," *IEEE Trans. Inform. Theory*, vol. 57, no. 11, pp. 7522–7541, Nov. 2011.

[41] A. Mecozzi, "Limits to long-haul coherent transmission set by the Kerr nonlinearity and noise of the in-line amplifiers," *J. Lightw. Technol.*, vol. 12, no. 11, pp. 1993–2000, Nov. 1994.

[42] H. Ghozlan and G. Kramer, "Models and information rates for multiuser optical fiber channels with nonlinearity and dispersion," *IEEE Trans. Inform. Theory*, vol. 63, no. 10, pp. 6440–6456, Oct. 2017.

[43] ——, "Interference focusing for mitigating cross-phase modulation in a simplified optical fiber model," in *Proc. IEEE Int. Symp. Inform. Theory*, Austin, Tx., Jun. 2010, pp. 2033–2037.

[44] ——, "Interference focusing for simplified optical fiber models with dispersion," in *Proc. IEEE Int. Symp. Inform. Theory*, Saint Petersburg, Russia, Jul–Aug 2011, pp. 376–379.

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

[46] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[47] L. Wegener, B. Povinelli, A. Green, P. Mitra, J. Stark, and P. Littlewood, "The effect of propagation nonlinearities on the information capacity of WDM optical fiber systems: Cross-phase modulation and four-wave mixing," *Physica D: Nonlinear Phenomena*, vol. 189, no. 1, pp. 81–99, Feb. 2004.

[48] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, Jun. 2001.

[49] M. Secondini and E. Forestieri, "Analytical fiber-optic channel model in the presence of cross-phase modulation," *IEEE Photon. Technol. Lett.*, vol. 24, no. 22, pp. 2016–2019, Nov. 2012.

[50] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, Dec. 2013.

[51] E. Forestieri and M. Secondini, "The nonlinear fiber-optic channel: Modeling and achievable information rate," in *Progress In Electromagnetics Research Symposium (PIERS)*, Prague, Czech Republic, Jul. 2015.

[52] M. Secondini and E. Forestieri, "The limits of the nonlinear Shannon limit," in *Proc. Optical Fiber Communication Conf. (OFC)*, Anaheim, CA, Mar. 2016.

[53] M. Sorokina, S. Sygletos, and S. Turitsyn, "Ripple distribution for nonlinear fiber-optic channels," *Opt. Express*, vol. 25, no. 3, pp. 2228–2238, Feb. 2017.

[54] D. Marsella, M. Secondini, E. Agrell, and E. Forestieri, "A simple strategy for mitigating XPM in nonlinear WDM optical systems," in *Proc. Optical Fiber Communication Conf. (OFC)*, Los Angeles, CA, Mar. 2015.

[55] M. Secondini and E. Forestieri, "On XPM mitigation in WDM fiber-optic systems," *IEEE Photon. Technol. Lett.*, vol. 26, no. 22, pp. 2252–2255, Nov. 2014.

[56] 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.

[57] E. Agrell and M. Karlsson, "Influence of behavioral models on multiuser channel capacity," *J. Lightw. Technol.*, vol. 33, no. 17, pp. 3507–3515, Sep. 2015.

[58] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, "Analytical modeling of nonlinear propagation in uncompensated optical transmission links," *IEEE Photon. Technol. Lett.*, vol. 23, no. 11, pp. 742–744, Jun. 2011.

[59] P. Poggiolini, G. Bosco, A. Carena, V. Curri, and F. Forghieri, "A simple and accurate model for non-linear propagation effects in uncompensated coherent transmission links," in *International Conference on Transparent Optical Networks (ICTON)*, Stockholm, Sweden, Jun. 2011, p. We.B1.3.

[60] P. Johannisson and M. Karlsson, "Perturbation analysis of nonlinear propagation in a strongly dispersive optical communication system," *J. Lightw. Technol.*, vol. 31, no. 8, pp. 1273–1282, Apr. 2013.

[61] C. Kurtzke, "Kapazitätsgrenzen digitaler optischer Übertragungssysteme (Capacity bounds of digital optical transmission systems)," Ph.D. dissertation, Technical University of Berlin, Germany, 1995, in German. [Online] http://dx.doi.org/10.14279/depositonce-5080.

[62] A. Splett, C. Kurzke, and K. Petermann, "Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities," in *Proc. European Conference on Optical Communication (ECOC)*, Montreux, Switzerland, 1993, pp. 41–44.

[63] G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri, "Analytical results on channel capacity in uncompensated optical links with coherent detection," *Opt. Express*, vol. 19, no. 26, pp. B440–B451, Dec. 2011.

[64] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1524–1539, May 2012.

[65] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2862–2876, Aug. 2014.

[66] K. S. Turitsyn, S. A. Derevyanko, I. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Phys. Rev. Lett.*, vol. 91, no. 20, p. 203901, Nov. 2003.

[67] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning," in *Proc. European Conference on Optical Communication (ECOC)*, Rome, Italy, Sep. 2018.

[68] K. Keykhosravi, G. Durisi, and E. Agrell, "A tighter upper bound on the capacity of the nondispersive optical fiber channel," in *Proc. European Conference on Optical Communication (ECOC)*, Gothenburg, Sweden, Sep. 2017.

[69] A. P. T. Lau and J. M. Kahn, "Signal design and detection in presence of nonlinear phase noise," *J. Lightw. Technol.*, vol. 25, no. 10, pp. 3008–3016, Oct. 2007.

[70] M. Tavana, K. Keykhosravi, V. Aref, and E. Agrell, "A low-complexity near-optimal detector for multispan zero-dispersion fiber-optic channels," in *Proc. European Conference on Optical Communication (ECOC)*, Sep. 2018.

[71] A. S. Tan, H. Wymeersch, P. Johannisson, E. Agrell, P. Andrekson, and M. Karlsson, "An ML-based detector for optical communication in the presence of nonlinear phase noise," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011.

[72] C. Häger, A. Graell i Amat, A. Alvarado, and E. Agrell, "Design of APSK constellations for coherent optical channels with nonlinear phase noise," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3362–3373, Aug. 2013.

[73] G. Kramer, "Autocorrelation function for dispersion-free fiber channels with distributed amplification," *IEEE Trans. Inform. Theory*, vol. 64, no. 7, pp. 5131–5155, Jul. 2018.

[74] M. I. Yousefi, "The asymptotic capacity of the optical fiber," *arXiv preprint:1610.06458*, 2016.

[75] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," *J. Lightw. Technol.*, vol. 23, no. 11, pp. 3755–3763, Nov. 2005.

[76] N. V. Irukulapati, M. Secondini, E. Agrell, P. Johannisson, and H. Wymeersch, "Improved lower bounds on mutual information accounting for nonlinear signal–noise interaction," *J. Lightw. Technol.*, vol. 36, no. 22, pp. 5152–5159, Nov. 2018.

[77] R. Maher, D. Lavery, D. Millar, A. Alvarado, K. Parsons, R. Killey, and P. Bayvel, "Reach enhancement of 100% for a DP-64QAM super-channel using MC-DBP," in *Proc. Optical Fiber Communication Conf. (OFC)*, Los Angeles, CA, USA, Mar. 2015, paper Th4D–5.

[78] D. Millar, R. Maher, D. Lavery, T. Koike-Akino, A. Alvarado, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen, S. J. Savory *et al.*, "Transceiver-limited high spectral efficiency Nyquist-WDM systems," in *Proc. Optical Fiber Communication Conf. (OFC)*, Los Angeles, CA, USA, Mar. 2015, paper Th2A–13.

[79] T. A. Eriksson, T. Fehenberger, P. A. Andrekson, M. Karlsson, N. Hanik, and E. Agrell, "Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments," *J. Lightw. Technol.*, vol. 34, no. 9, pp. 2256–2266, May 2016.

[80] A. Leven, F. Vacondio, L. Schmalen, S. Brink, and W. Idler, "Estimation of soft FEC performance in optical transmission experiments," *IEEE Photon. Technol. Lett.*, vol. 23, no. 20, pp. 1547–1549, Oct. 2011.

[81] J.-S. Lee and C.-S. Shim, "Bit-error-rate analysis of optically preamplified receivers using an eigenfunction expansion method in optical frequency domain," *J. Lightw. Technol.*, vol. 12, no. 7, pp. 1224–1229, Jul. 1994.

[82] E. Forestieri, "Evaluating the error probability in lightwave systems with chromatic dispersion, arbitrary pulse shape and pre-and postdetection filtering," *J. Lightw. Technol.*, vol. 18, no. 11, p. 1493, Nov. 2000.

[83] M. Kac and A. J. Siegert, "On the theory of noise in radio receivers with square law detectors," *J. Applied Phys.*, vol. 18, no. 4, pp. 383–397, Apr. 1947.

[84] E. Forestieri and M. Secondini, "On the error probability evaluation in lightwave systems with optical amplification," *J. Lightw. Technol.*, vol. 27, no. 6, pp. 706–717, Mar. 2009.

[85] K.-P. Ho, *Phase-Modulated Optical Communication Systems*. Springer, 2005.

[86] A. Mecozzi and M. Shtaif, "On the capacity of intensity modulated systems using optical amplifiers," *IEEE Photonics Technology Letters*, vol. 13, no. 9, pp. 1029–1031, Sep. 2001.

[87] K.-P. Ho, "Exact evaluation of the capacity for intensity-modulated direct-detection channels with optical amplifier noises," *IEEE Photon. Technol. Lett.*, vol. 17, no. 4, pp. 858–860, Apr. 2005.

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

[89] G. Durisi, "On the capacity of the block-memoryless phase-noise channel," *IEEE Commun. Lett.*, vol. 16, no. 8, pp. 1157–1160, Aug. 2012.

[90] R. Nuriyev and A. Anastasopoulos, "Capacity and coding for the block-independent noncoherent AWGN channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 866–883, Mar. 2005.

[91] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2257–2270, Oct. 2004.

[92] A. Mecozzi and M. Shtaif, "Information capacity of direct detection optical transmission systems," *J. Lightw. Technol.*, vol. 36, no. 3, pp. 689–694, Feb. 2018.

[93] R. You and J. M. Kahn, "Upper-bounding the capacity of optical IM/DD channels with multiple-subcarrier modulation and fixed bias using trigonometric moment space method," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 514–523, Feb. 2002.

[94] S. Hranilovic and F. R. Kschischang, "Capacity bounds for power-and band-limited optical intensity channels corrupted by Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 784–795, May 2004.

[95] J. Karout, E. Agrell, K. Szczerba, and M. Karlsson, "Optimizing constellations for single-subcarrier intensity-modulated optical systems," *IEEE Trans. Inform. Theory*, vol. 58, no. 7, pp. 4645–4659, Jul. 2012.

[96] A. Lapidoth, S. M. Moser, and M. A. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4449–4461, Oct. 2009.

[97] S. M. Moser, M. Mylonakis, L. Wang, and M. Wigger, "Asymptotic capacity results for MIMO wireless optical communication," in *Proc. IEEE Int. Symp. Inform. Theory*, Aachen, Germany, Jun. 2017, pp. 536–540.

[98] S. M. Moser, L. Wang, and M. Wigger, "Asymptotic high-SNR capacity of MISO optical intensity channels," in *IEEE Info. Theory Workshop (ITW)*, Kaohsiung, Taiwan, Nov. 2017, pp. 86–90.

[99] ——, "Capacity results on multiple-input single-output wireless optical channels," *IEEE Trans. Inform. Theory*, vol. 64, no. 11, pp. 6954–6966, Nov. 2018.

[100] S. M. Moser, "Capacity results of an optical intensity channel with input-dependent Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 58, no. 1, pp. 207–223, Jan. 2012.

[101] A. Sheikh, C. Fougstedt, A. Graell i Amat, P. Johannisson, P. Larsson-Edefors, and M. Karlsson, "Dispersion compensation FIR filter with improved robustness to coefficient quantization errors," *J. Lightw. Technol.*, vol. 34, no. 22, pp. 5110–5117, Nov. 2016.

[102] K. Hill and G. Meltz, "Fiber Bragg grating technology fundamentals and overview," *J. Lightw. Technol.*, vol. 15, no. 8, pp. 1263–1276, Aug. 1997.

[103] P. D. Lax, "Integrals of nonlinear equations of evolution and solitary waves," *Communications on pure and applied mathematics*, vol. 21, no. 5, pp. 467–490, Sep. 1968.

[104] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, "Korteweg-de Vries equation and generalizations. VI. methods for exact solution," *Communications on pure and applied mathematics*, vol. 27, no. 1, pp. 97–133, Jan. 1974.

[105] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear Fourier transform, part I: Mathematical tools," *IEEE Trans. Inform. Theory*, vol. 60, no. 7, pp. 4312–4328, Jul. 2014.

[106] ——, "Information transmission using the nonlinear Fourier transform, part II: Numerical methods," *IEEE Trans. Inform. Theory*, vol. 60, no. 7, pp. 4329–4345, Jul. 2014.

[107] ——, "Information transmission using the nonlinear Fourier transform, part III: Spectrum modulation," *IEEE Trans. Inform. Theory*, vol. 60, no. 7, pp. 4346–4369, Jul. 2014.

[108] M. I. Yousefi and X. Yangzhang, "Linear and nonlinear frequency-division multiplexing," in *Proc. European Conference on Optical Communication (ECOC)*, Dusseldorf, Germany, Sep. 2016.

[109] S. T. Le, V. Aref, and H. Buelow, "Nonlinear signal multiplexing for communication beyond the Kerr nonlinearity limit," *Nature Photonics*, vol. 11, no. 9, p. 570, Sep. 2017.

[110] R. Dar and P. J. Winzer, "Nonlinear interference mitigation: Methods and potential gain," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 903–930, Apr. 2017.

[111] R.-J. Essiambre and P. J. Winzer, "Fibre nonlinearities in electronically predistorted transmission," in *Proc. European Conference on Optical Communication (ECOC)*, vol. 2, Glasgow, UK, Sep. 2005, paper Tu3.2.2.

[112] G. Goldfarb, M. G. Taylor, and G. Li, "Experimental demonstration of fiber impairment compensation using the split-step finite-impulse-response filtering method," *IEEE Photon. Technol. Lett.*, vol. 20, no. 22, pp. 1887–1889, Nov. 2008.

[113] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416–3425, Oct. 2008.

[114] R. Dar and P. J. Winzer, "On the limits of digital back-propagation in fully loaded WDM systems," *IEEE Photon. Technol. Lett.*, vol. 28, no. 11, pp. 1253–1256, Jun. 2016.

[115] K. Roberts, C. Li, L. Strawczynski, M. O'Sullivan, and I. Hardcastle, "Electronic precompensation of optical nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 18, no. 2, pp. 403–405, Jan. 2006.

[116] L. B. Du, D. Rafique, A. Napoli, B. Spinnler, A. D. Ellis, M. Kuschnerov, and A. J. Lowery, "Digital fiber nonlinearity compensation: Toward 1-Tb/s transport," *IEEE Signal Process. Mag.*, vol. 31, no. 2, pp. 46–56, Mar. 2014.

[117] N. V. Irukulapati, H. Wymeersch, P. Johannisson, and E. Agrell, "Stochastic digital backpropagation," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3956–3968, Nov. 2014.

[118] L. Zhu and G. Li, "Folded digital backward propagation for dispersion-managed fiber-optic transmission," *Opt. Express*, vol. 19, no. 7, pp. 5953–5959, Mar. 2011.

[119] L. B. Du and A. J. Lowery, "Channelized chromatic dispersion compensation for XPM suppression and simplified digital SPM compensation," in *Proc. Optical Fiber Communication Conf. (OFC)*, San Francisco, CA, USA, Mar. 2014.

[120] A. Yariv, D. Fekete, and D. M. Pepper, "Compensation for channel dispersion by nonlinear optical phase conjugation," *Opt. Lett.*, vol. 4, no. 2, pp. 52–54, Feb. 1979.

[121] R. A. Fisher, B. Suydam, and D. Yevick, "Optical phase conjugation for time-domain undoing of dispersive self-phase-modulation effects," *Opt. Lett.*, vol. 8, no. 12, pp. 611–613, Dec. 1983.

[122] S. Le, M. McCarthy, S. Turitsyn, I. Phillips, D. Lavery, T. Xu, P. Bayvel, and A. Ellis, "Optical and digital phase conjugation techniques for fiber nonlinearity compensation," in *Opto-Electronics and Communications Conference (OECC)*, Shanghai, China, 2015.

[123] A. Ellis, S. Le, M. McCarthy, and S. Turitsyn, "The impact of parametric noise amplification on long haul transmission throughput," in *International Conference on Transparent Optical Networks (ICTON)*, Budapest, Hungary, Jul. 2015.

[124] A. Ellis, M. McCarthy, M. Al-Khateeb, and S. Sygletos, "Capacity limits of systems employing multiple optical phase conjugators," *Opt. Express*, vol. 23, no. 16, pp. 20 381–20 393, Aug. 2015.

[125] P. Kaewplung and K. Kikuchi, "Simultaneous cancellation of fiber loss, dispersion, and effect in ultralong-haul optical fiber transmission by midway optical phase conjugation incorporated with distributed Raman amplification," *J. Lightw. Technol.*, vol. 25, no. 10, pp. 3035–3050, Oct. 2007.

[126] Y. Cai, D. G. Foursa, C. R. Davidson, J.-X. Cai, O. Sinkin, M. Nissov, and A. Pilipetskii, "Experimental demonstration of coherent MAP detection for nonlinearity mitigation in long-haul transmissions," in *Proc. Optical Fiber Communication Conf. (OFC)*, San Diego, CA, Mar. 2010.

[127] G. Bosco, I. N. Cano, P. Poggiolini, L. Li, and M. Chen, "MLSE-based DQPSK transmission in 43 Gb/s DWDM long-haul dispersion-managed optical systems," *J. Lightw. Technol.*, vol. 28, no. 10, pp. 1573–1581, Mar. 2010.

[128] A. Rezania and J. C. Cartledge, "Transmission performance of 448 Gb/s single-carrier and 1.2 Tb/s three-carrier superchannel using dual-polarization 16-QAM with fixed LUT based MAP detection," *J. Lightw. Technol.*, vol. 33, no. 23, pp. 4738–4745, Dec. 2015.

[129] G. Liga, A. Alvarado, E. Agrell, M. Secondini, R. I. Killey, and P. Bayvel, "Optimum detection in presence of nonlinear distortions with memory," in *Proc. European Conference on Optical Communication (ECOC)*, Valencia, Spain, Sep.–Oct. 2015.

[130] L. Li, Z. Tao, L. Liu, W. Yan, S. Oda, T. Hoshida, and J. C. Rasmussen, "Nonlinear polarization crosstalk canceller for dual-polarization digital coherent receivers," in *Proc. Optical Fiber Communication Conf. (OFC)*, San Diego, CA, Mar. 2010.

[131] I. B. Djordjevic, L. L. Minkov, L. Xu, and T. Wang, "Suppression of fiber nonlinearities and PMD in coded-modulation schemes with coherent detection by using turbo equalization," *J. Opt. Commun. Netw.*, vol. 1, no. 6, pp. 555–564, Nov. 2009.

[132] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: modeling and mitigation," *J. Lightw. Technol.*, vol. 33, no. 5, pp. 1044–1053, Mar. 2015.