TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Nachrichtentechnik

Information-Theoretic Aspects of Fiber-Optic Communication Channels

Bernhard Walter Göbel

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr.-Ing. E. Steinbach
Prüfer der Dissertation:	1. Univ.-Prof. Dr.-Ing. N. Hanik
	2. Univ.-Prof. Dr. sc. techn. G. Kramer

Die Dissertation wurde am 16.06.2010 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 21.10.2010 angenommen.

"Our humble art, on the other hand, is not a dweller in gilded monuments to amusement with crimson curtains lit by gas lamps; ours is not caparisoned in sumptuous costumes, painted in alluring colors, festooned with oil and tempera canvas flats. No, our stage is a humble thread of metal stretched beneath the dark and cold ocean wave; ours is a cord lying on the frigid and muddy bottom of the sea. Our art manifests itself in the tiny galvanic flickers of light in a squalid, cramped, and dark room; a flicker with no more nuance or affection than on or off, positive or negative, left or right, dot or dash, yes or no. But that binary essence is about to join continents, unite nations and unseat tyrants through the spread of truth; it will permit monarchs to converse with presidents, mother with son, lover with beloved, you with me, no matter where on earth either of us may dwell; we shall converse as intimately or as grandly as we might right here, sitting face-to-face."

John Griesemer, *Signal & Noise*

Preface

I thank Professor Norbert Hanik for giving me the opportunity to work on this thesis as his assistant at TUM's Institute for Communications Engineering (LNT). I have always felt privileged in this position, and I thoroughly enjoyed the freedom and the possibilities he granted me in research, teaching and numerous other activities. His advice, support, encouragement and patience helped me and my research work prosper.

I am grateful to Professor Gerhard Kramer for acting as the co-referee of my dissertation and for his valuable and extensive advice that helped improve important parts of it.

Working at LNT is a highly rewarding thing to do, partly because of the pleasant social atmosphere. Among the many colleagues whose company I enjoyed in and out of the office and on many trips abroad, I would like to mention my office-mate Leonardo Coelho as well as Florian Breyer, Stephan Hellerbrand and Christian Kuhn. Professor Günter Söder and Johanna Weindl have helped me improve this thesis by carefully proofreading the manuscript. Over the years, I have had the pleasure of working with a number of excellent students. Among them was Maxim Kuschnerov, some of whose results are included in Chapter 4.

I was lucky and honored to spend last year's summer at Bell Laboratories on Crawford Hill. I am thankful to René-Jean Essiambre for his invitation and his warm hospitality. Working with and learning from him, Peter Winzer, Gerard Foschini and others was an exciting experience for me. (So was playing football with the Bell Labs staff right next to the famous horn antenna, although I am afraid that the excitement was on my side only.) Parts of Chapters 3 and 6 contain results of our joint research project.

I am fortunate to have a family that has encouraged and supported me at all times. This work is dedicated to my parents in deep gratitude for roots and wings. The final word is for Johanna. It was a pleasure to share three of my LNT years with you, and it is a gift to share my life with you.

München, June 2010 Bernhard Göbel

Contents

Zusammenfassung

Diese Arbeit behandelt informationstheoretische Grenzen der Datenübertragung über faseroptische Kanäle. Eine detaillierte Diskussion der wechselwirkenden physikalischen Effekte ermöglicht die Formulierung abstrahierter Kanalmodelle. Durch Zerlegung der Transinformation in Polarkoordinaten wird eine informationstheoretische Methode entwickelt, die sich zur Analyse partiell kohärenter Kanäle eignet. Ein Kanalmodell im Frequenzbereich erlaubt die getrennte Analyse miteinander wechselwirkender nichtlinearer Effekte. Mit diesem Modell wird die Kapazität des faseroptischen Kanals für verschiedene Szenarien abgeschätzt. Die höchste Kapazität erreicht ein Kanal, der nur durch Rauschen und durch die nichtlineare Wechselwirkung von Signal und Rauschen gestört wird. Für die Abschätzung der Kapazität von Wellenlängenmultiplex-Systemen wird zusätzlich ein phänomenologisches Kanalmodell im Zeitbereich entwickelt, das bestehende Ergebnisse bestätigt.

Abstract

This work addresses the information-theoretic limits of communication over fiber-optic channels. A detailed discussion of the interacting physical effects provides the background for a formulation of abstract channel models. An information-theoretic method is developed that decomposes the mutual information in terms of polar coordinates. This method is useful for the analysis of partially coherent channels. A channel model in the frequency domain allows the separate analysis of interacting nonlinear effects. The capacity of the fiber-optic channel is estimated with this model for various scenarios. The maximum capacity is reached by a channel that is impaired only by noise and by the nonlinear interaction of signal and noise. A phenomenological time-domain channel model is developed to obtain an additional estimate of the capacity of wavelength division multiplexing systems. This model confirms existing results.

1

Introduction

Digital communications has become an ubiquitous part of everyday life. Broadband connections to the internet are available in most parts of the developed world. Access networks transport data on copper lines, wireless links and – increasingly – optical fibers. The backbone network, which connects cities, regions, countries and continents, consists entirely of optical fibers. This network has become one of the key drivers of economic growth, in developed countries as well as in emerging regions of the world. Perhaps more importantly, the cultural and political progression of our world evolves more and more around the internet; fiber-optic communications has changed the world and continues to do so.

This change is reflected in the growing demand for capacity. According to market research [Tel10], the used international capacity has increased 22-fold since 2002. The compound bandwidth is predicted to continue its growth at an annual rate of 40 percent in the next five years. Unlike for example the capacity of the mobile communications channel, the capacity of optical fibers can grow infinitely. The available bandwidth can always be increased by deploying additional fibers. This option is economically unfeasible, of course; the telecommunications market is driven by technological advance and, consequently, falling prices per unit capacity (i. e. bits per second).

In the past thirty years, technological innovation has enabled an exponential growth (approximately 10-fold every four years) of the capacity-distance product (in bit/s · km) of optical fibers [Des06]. This "optical Moore's law" cannot persist. The amount of data that can be reliably transmitted over any channel, including optical fibers, is limited by the channel's capacity. This is one of the key statements of information theory, the discipline that Claude Shannon gave birth to in 1948 [Sha48].

The fiber-optic channel is characterized by an incomparably large bandwidth and low attenuation on the one hand but also by dispersion and nonlinear effects on the other. The nonlinear interaction of signal and noise from optical amplifiers during propagation presents an additional difficulty. What is the capacity of such a channel? The physics of light propagation in fibers are too complicated for a simple answer to this question. Moreover, the capacity of fiber-optic communication channels depends on a large variety of system parameters but also on how practical system constraints render physical effects fundamental, i. e. purely random or irrevertible. The motivation and aim of this thesis is to gain and provide insight into the fundamental limits that information theory sets to communication over optical fibers. In lieu of simple answers, methods and models are developed that can be applied to obtain a capacity estimate of a given specific channel. The results that are obtained for a variety of different scenarios suggest that the data rates of today's fiber-optic communication systems have by far not approached their theoretic limits as closely as those of for instance wireless or copper-based communication systems.

The thesis is organized as follows:

Chapter 2 introduces the physical background of light propagation in optical fibers, starting from Maxwell's equations and ending at the nonlinear Schroedinger equation. Material polarization is discussed as the origin of (almost) all linear and nonlinear fiber effects. An effective cubic susceptibility is calculated to include the effect of the randomly varying light polarization in the scalar propagation equations. The chapter's second half introduces the most important components of a fiber-optic communication system. Optical fibers are discussed from a system's standpoint. The principles of distributed and lumped optical amplification are explained.

Fundamental terms of information theory are reviewed in **Chapter 3**. The channel capacity as the maximum rate of reliable communication over any channel is introduced and discussed for various channel types. In the second part of this chapter, a method for the decomposition of mutual information in terms of polar coordinates is developed and discussed. Among other useful applications, this method can be applied to partially coherent channels such as the nonlinear fiber-optic channel in both time and frequency domain. To support the discussion of partially coherent channels, a short introduction to directional statistics is given. Finally, the discussion is extended to noncoherent channels such as optical direct-detection systems.

Chapter 4 deals with polarization mode dispersion (PMD). After a review of PMD fundamentals, the effect of PMD on the channel capacity of fiber-optic communication systems is considered in the absence of fiber nonlinearities. If the channel is considerably nonlinear, the interplay of PMD and fiber nonlinearities can lead to cross-polarization modulation, which causes fading and intereference between polarization-multiplexed signals.

The limits that fiber nonlinearities set on the channel capacity of fiber-optic communication systems transporting a single signal are considered in **Chapter 5**. A frequency-domain model is developed that allows separating the relevant nonlinear effects and obtaining an analytical solution for the capacity. Finally, the interaction of signal and noise

from optical amplification is included in the discussion. The case where nonlinear interactions of the signal with itself can be removed (either at the transmitter or at the receiver) is treated separately.

Chapter 6 treats the propagation of multiple channels in a system using wavelength division multiplexing. In view of the situation in transparent optical networks, it is assumed that intra-channel nonlinearities can be compensated for, while inter-channel nonlinearities remain as fundamental impairments. Two approaches for calculating the capacity in this scenario are considered. The first is based on a published numerical study that determines the capacity in the time domain. A phenomenological channel model with only two free parameters is proposed which permits a rapid calculation of the capacity using the decomposition method of mutual information introduced in Chapter 3. The second approach extends the frequency-domain model developed in Chapter 5. The nonlinear interaction of signal and noise is included in this discussion.

Finally, **Chapter 7** summarizes the results and suggests open research problems that are related to the content of this thesis.

Notes on the notation:

This thesis covers topics from such different disciplines as information theory, communications theory and optical communications. The notation was chosen such that it diverts as little as possible from the standard notation of those fields.

In consequence, different notations are used for similar cases depending on the context. For instance, electrical fields are denoted by capital letters (e. g. $E(t)$), whereas small letters are used for arbitrary time-domain signals (e. g. $a(t), s(t)$). Captical letters denote the Fourier transforms of these signals ($A(f), S(f)$). Unless they represent continuous-time physical signals, random variables are denoted by capital letters, their realizations by small letters.

In other cases, the same notation denotes different meanings. The power spectral density (PSD) of a signal $x(t)$ is denoted by $\Phi_x(f)$. The same symbol is used in Chapter 4 to denote the PSD of the transmit signal in x-polarization. The symbol P_s denotes either the total average signal power, the total average signal power per polarization (Chapter 4) or the total average signal power per WDM channel (Chapter 6).

In any case, the correct meaning is always clear from the context. Additionally, Appendix B contains an exhaustive list of symbols for reference.

2

Introduction to fiber-optic communications

A certain level of abstraction is necessary for defining information-theoretic channel models. However, any such model needs to be justified on the lowest level of abstraction, the physical level. In this chapter, the fiber-optic channel is reviewed on this lowest level.

This chapter is organized as follows. Section 2.1 starts from Maxwell's equations to derive the nonlinear Schroedinger equation (NLSE) that governs light propagation in optical single-mode fibers. The origin of linear and nonlinear effects during the propagation of light in optical fibers is reviewed and the role of light polarization is discussed. Analytical and numerical solutions of the NLSE are presented. In the remainder of the chapter, the most important components that are part of optical communication systems as well as the impairments they produce are introduced. In practice, some of these components are mature, others pose severe practical system limitations. In the derivation of fundamental conclusions and limits, it is sensible to differentiate between systematic impairments (such as dispersion, nonlinear fiber effects, noise and square-law photodetection) and impairments that have their origin in the degree of development of the components that are available today. This thesis is limited to the analysis of systematic impairments; therefore, the description is limited to the properties of "ideal" components. Section 2.2 introduces optical single-mode fibers on a system level. The parameters characterizing optical fibers are specified, and the most important effects and impairments are discussed. Section 2.3 introduces optical amplifiers, which are the dominant source of noise in optical communication systems. Other important components are reviewed in Section 2.4. Section 2.5 briefly summarizes the main results of this chapter.

2.1 Light wave propagation in single-mode fibers

2.1.1 The general wave equation

Light is an electromagnetic wave and as such, its propagation in any medium is governed by Maxwell's equations [Agr01]:

$$\nabla \times \overrightarrow{\boldsymbol{E}} = -\frac{\partial \overrightarrow{\boldsymbol{B}}}{\partial t}, \tag{2.1}$$

$$\nabla \times \overrightarrow{\boldsymbol{H}} = \overrightarrow{\boldsymbol{J}} + \frac{\partial \overrightarrow{\boldsymbol{D}}}{\partial t}, \tag{2.2}$$

$$\nabla \cdot \overrightarrow{\boldsymbol{D}} = \rho_f, \tag{2.3}$$

$$\nabla \cdot \overrightarrow{\boldsymbol{B}} = 0. \tag{2.4}$$

$\overrightarrow{\boldsymbol{E}}$ and $\overrightarrow{\boldsymbol{H}}$ are electric and magnetic field vectors, respectively, $\overrightarrow{\boldsymbol{D}}$ is the electric flux density field vector, $\overrightarrow{\boldsymbol{B}}$ is the magnetic flux density vector and ∇ denotes the Nabla operator. The electric current density vector $\overrightarrow{\boldsymbol{J}}$ as well as the electric charge density ρ_f equal zero in a dielectric medium such as optical fibers. The flux densities are related to the field vectors through the material equations

$$\overrightarrow{\boldsymbol{D}} = \varepsilon_0 \overrightarrow{\boldsymbol{E}} + \overrightarrow{\boldsymbol{P}}, \tag{2.5}$$

$$\overrightarrow{\boldsymbol{B}} = \mu_0 \overrightarrow{\boldsymbol{H}} + \overrightarrow{\boldsymbol{M}}, \tag{2.6}$$

where ε_0 is the vacuum permittivity, μ_0 is the vacuum permeability, and $\overrightarrow{\boldsymbol{P}}$ and $\overrightarrow{\boldsymbol{M}}$ are the electric and magnetic material polarizations, respectively. For a nonmagnetic medium such as optical fibers, $\overrightarrow{\boldsymbol{M}} = 0$.

By taking the curl of (2.1), we obtain an expression for the propagation of the electric field:

$$\begin{aligned} \nabla \times \nabla \times \overrightarrow{\boldsymbol{E}} &= -\nabla \times \frac{\partial \overrightarrow{\boldsymbol{B}}}{\partial t} \overset{(2.6)}{=} -\mu_0 \frac{\partial (\nabla \times \overrightarrow{\boldsymbol{H}})}{\partial t} \overset{(2.2)}{=} -\mu_0 \frac{\partial^2 \overrightarrow{\boldsymbol{D}}}{\partial t^2} \\ &\overset{(2.5)}{=} -\mu_0 \varepsilon_0 \frac{\partial^2 \overrightarrow{\boldsymbol{E}}}{\partial t^2} - \mu_0 \frac{\partial^2 \overrightarrow{\boldsymbol{P}}}{\partial t^2}. \end{aligned} \tag{2.7}$$

Using the identity

$$\nabla \times \nabla \times \overrightarrow{\boldsymbol{E}} = \nabla (\nabla \cdot \overrightarrow{\boldsymbol{E}}) - \Delta \overrightarrow{\boldsymbol{E}} \tag{2.8}$$

together with (2.3) and (2.5), we can write (2.7) as

$$\Delta \overrightarrow{\boldsymbol{E}} + \nabla \left(\frac{1}{\varepsilon_0} \nabla \cdot \overrightarrow{\boldsymbol{P}} \right) - \frac{1}{c_0^2} \frac{\partial^2 \overrightarrow{\boldsymbol{E}}}{\partial t^2} = \mu_0 \frac{\partial^2 \overrightarrow{\boldsymbol{P}}}{\partial t^2}, \tag{2.9}$$

where $c_0 = 1/\sqrt{\varepsilon_0 \mu_0}$ is the light velocity in vacuum and $\Delta \overrightarrow{\boldsymbol{E}} = \nabla^2 \overrightarrow{\boldsymbol{E}}$ denotes the Laplace operator.

Equation (2.9) is known as the *general wave equation*.

2.1.2 Material polarization

The interaction of light and matter is included in (2.9) through the material polarization $\overrightarrow{\boldsymbol{P}}$. Applying an external electric field to a dieletric medium leads to a shift of the internal carriers. This creates a dipole moment which itself is the source of electromagnetic radiation. On a macroscopic scale, the summation over all dipole moments per unit volume yields the material polarization $\overrightarrow{\boldsymbol{P}}$ [Sch04]. The incident electric field $\overrightarrow{\boldsymbol{E}}$ can be related to the material polarization $\overrightarrow{\boldsymbol{P}}$ through the material's susceptibility χ [Agr01] by

$$\overrightarrow{\boldsymbol{P}} = \varepsilon_0 \chi \overrightarrow{\boldsymbol{E}}. \tag{2.10}$$

In isotropic and linear media, the susceptibility χ is a scalar; optical fibers are different from such media in two aspects. For materials with a non-isotropic structure, $\overrightarrow{\boldsymbol{P}}$ can have a direction different from that of $\overrightarrow{\boldsymbol{E}}$; in this case, the (linear) susceptibility is a 2^{nd}-rank tensor denoted by $\overleftrightarrow{\boldsymbol{\chi}}$. When the incident field power is very high, the reset force of the shifted carriers becomes nonlinear [Sch04]; in this case, the susceptibility is a function of $\overrightarrow{\boldsymbol{E}}$, and $\overrightarrow{\boldsymbol{P}}$ can be described by a power series expansion [Han95]:

$$\overrightarrow{\boldsymbol{P}} = \overrightarrow{\boldsymbol{P}}^{(1)} + \overrightarrow{\boldsymbol{P}}^{(2)} + \overrightarrow{\boldsymbol{P}}^{(3)} + \ldots, \tag{2.11}$$

where $\overrightarrow{\boldsymbol{P}}^{(n)}$ is related to $\overrightarrow{\boldsymbol{E}}$ through the n^{th}-order material susceptibility $\overleftrightarrow{\boldsymbol{\chi}}^{(n)}$. Optical fibers are composed of silicon dioxide (SiO_2); because of the inversion symmetry of SiO_2 molecules, even-order $\overleftrightarrow{\boldsymbol{\chi}}^{(n)}$ are zero. The fifth- and higher-order susceptibilities are very small and can be neglected, as the power required to excite material polarization of fifth order would physically destroy the fiber [Hel77]. Therefore, (2.11) can be written as

$$\overrightarrow{\boldsymbol{P}} = \overrightarrow{\boldsymbol{P}}^{(1)} + \overrightarrow{\boldsymbol{P}}^{(3)}. \tag{2.12}$$

Linear material polarization

Each cartesian component $P_i^{(1)}$, $i = x, y, z$, of the linear material polarization at position $\overrightarrow{\boldsymbol{r}}$ is related to $\overrightarrow{\boldsymbol{E}}$ through [Han95]

$$P_i^{(1)}(\overrightarrow{\boldsymbol{r}}, t) = \varepsilon_0 \cdot \int_0^\infty \sum_{j=x,y,z} \chi_{ij}^{(1)}(\tau) \cdot E_j(\overrightarrow{\boldsymbol{r}}, t-\tau)\, \mathrm{d}\tau. \tag{2.13}$$

The so-called Einstein summation convention allows a compact notation of (2.13) by implying summation over all possible values of indices common to all factors of a product:

$$P_i^{(1)}(\overrightarrow{\boldsymbol{r}}, t) = \varepsilon_0 \cdot \int_0^\infty \chi_{ij}^{(1)}(\tau) \cdot E_j(\overrightarrow{\boldsymbol{r}}, t-\tau)\, \mathrm{d}\tau. \tag{2.14}$$

By taking the Fourier transform of (2.14), the linear polarization in frequency domain can be written as

$$P_i^{(1)}(\overrightarrow{\boldsymbol{r}}, f) = \varepsilon_0 \cdot X_{ij}^{(1)}(f) \cdot E_j(\overrightarrow{\boldsymbol{r}}, f), \tag{2.15}$$

where

$$X_{ij}^{(1)}(f) = \int_0^\infty \chi_{ij}^{(1)}(t) \cdot e^{-\jmath 2\pi f t}\,\mathrm{d}t. \tag{2.16}$$

In matrix-vector notation, (2.15) can be expressed as [Sch04]

$$\begin{pmatrix} P_x^{(1)}(\overrightarrow{\boldsymbol{r}}, f) \\ P_y^{(1)}(\overrightarrow{\boldsymbol{r}}, f) \\ P_z^{(1)}(\overrightarrow{\boldsymbol{r}}, f) \end{pmatrix} = \varepsilon_0 \cdot \begin{pmatrix} X_{xx}^{(1)}(f) & X_{xy}^{(1)}(f) & X_{xz}^{(1)}(f) \\ X_{yx}^{(1)}(f) & X_{yy}^{(1)}(f) & X_{yz}^{(1)}(f) \\ X_{zx}^{(1)}(f) & X_{zy}^{(1)}(f) & X_{zz}^{(1)}(f) \end{pmatrix} \cdot \begin{pmatrix} E_x(\overrightarrow{\boldsymbol{r}}, f) \\ E_y(\overrightarrow{\boldsymbol{r}}, f) \\ E_z(\overrightarrow{\boldsymbol{r}}, f) \end{pmatrix}. \tag{2.17}$$

All linear fiber effects have their origin in the first-order material susceptibility tensor $\overleftrightarrow{\boldsymbol{X}}^{(1)}$. A separation of these effects is obtained by expressing $\overleftrightarrow{\boldsymbol{X}}^{(1)}$ through the relative permittivity tensor of the material:

$$\begin{aligned} \overleftrightarrow{\boldsymbol{X}}^{(1)} &= \overleftrightarrow{\boldsymbol{\varepsilon}}_r - \overleftrightarrow{\mathbf{1}} = \overleftrightarrow{\boldsymbol{\varepsilon}}_{r,r} - \jmath \cdot \overleftrightarrow{\boldsymbol{\varepsilon}}_{r,i} + \overleftrightarrow{\boldsymbol{\varepsilon}}_{r,\Delta} - \overleftrightarrow{\mathbf{1}} \\ &= \underbrace{\begin{pmatrix} \varepsilon_{r,r} & 0 & 0 \\ 0 & \varepsilon_{r,r} & 0 \\ 0 & 0 & \varepsilon_{r,r} \end{pmatrix}}_{\text{chromatic dispersion}} - \jmath \cdot \underbrace{\begin{pmatrix} \varepsilon_{r,i} & 0 & 0 \\ 0 & \varepsilon_{r,i} & 0 \\ 0 & 0 & \varepsilon_{r,i} \end{pmatrix}}_{\text{loss}} \\ &\quad + \underbrace{\begin{pmatrix} \varepsilon_{r,\Delta,xx} & \varepsilon_{r,\Delta,xy} & \varepsilon_{r,\Delta,xz} \\ \varepsilon_{r,\Delta,yx} & \varepsilon_{r,\Delta,yy} & \varepsilon_{r,\Delta,yz} \\ \varepsilon_{r,\Delta,zx} & \varepsilon_{r,\Delta,zy} & \varepsilon_{r,\Delta,zz} \end{pmatrix}}_{\text{birefringence, mode coupling}} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned} \tag{2.18}$$

The first term in (2.18) causes the effect of *material dispersion*, which is the main contributor to *chromatic dispersion* in standard single-mode fibers [Agr02].[1] The quantity $\varepsilon_{r,r}$ is equal to the squared refractive index at position $\{x, y\}$:

$$\varepsilon_{r,r} = n^2(f, x, y) = n^2(f, r) \tag{2.19}$$

with $r = \sqrt{x^2 + y^2}$; the second equality is justified by the rotation-symmetric geometry of an ideal fiber. The dependence of the refractive index n on the wavelength λ is depicted in Figure 2.1.

The second term in (2.18) is imaginary and governs the linear attenuation. Material absorption limits the light propagation in the ultra-violet range ($\lambda < 0.4\,\mu\mathrm{m}$) through electronic resonances and in the infra-red range ($\lambda > 2\,\mu\mathrm{m}$) through molecular vibrations [Agr01]. In the wavelength range used for telecommunications, however, the scattering of light at local density variations and material impurities is the main cause of attenuation. This effect is called *Rayleigh scattering*; it scales with the fourth power of the light wave's frequency, so that scattering (and thus attenuation) increases with

[1] The other contributor to chromatic dispersion is *waveguide dispersion*, which has its origin in the frequency-dependence of the transversal field distribution (2.37). By adapting a fiber's refractive index profile, the waveguide dispersion can be modified in order to design dispersion-shifted, -compensating or -flattened fibers (cf. Section 2.2.1) [Agr02].

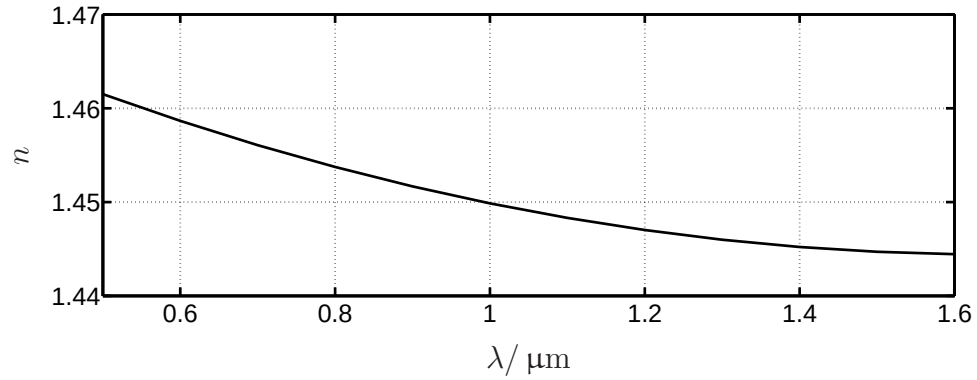

Figure 2.1: Refractive index versus wavelength for fused silica (after [Agr01]).

decreasing wavelength [Sch04]. The attenuation curve of silica fibers exhibits a broad minimum around 1.55 µm, so that $\varepsilon_{r,i}$ can be regarded constant in many cases. As the local density variations in the fiber have a homogeneous character on a larger scale, $\varepsilon_{r,i}$ is considered independent of the spatial coordinates. The attenuation (in dB/km) versus wavelength is depicted in Figure 2.5.

In a perfectly cylindrical fiber, only the three diagonal elements of $\overleftrightarrow{\boldsymbol{X}}^{(1)}$ are non-zero and $X^{(1)}_{xx}(f) = X^{(1)}_{yy}(f) = X^{(1)}_{zz}(f)$. In this case, all elements of $\overleftrightarrow{\boldsymbol{\varepsilon}}_{r,\Delta}$ are zero. Real fibers deviate from this ideal because of fabrication imperfections or external influences such as pressure or torsion. These imperfections, modeled by the tensor $\overleftrightarrow{\boldsymbol{\varepsilon}}_{r,\Delta}$, lead to birefringence, i. e. unequal propagation velocities of the cartesian components of a light wave, and mode coupling, i. e. coupling of the cartesian components of a light wave. Random birefringence and mode coupling are the cause of *polarization mode dispersion* (PMD), which is discussed in Chapter 4. By allowing the elements of $\overleftrightarrow{\boldsymbol{\varepsilon}}_{r,\Delta}$ to have an imaginary component, polarization-dependent loss (PDL) can be modeled as well.

Cubic material polarization

The third-order polarization $\overrightarrow{\boldsymbol{P}}^{(3)}$ is related to the eletrical field through the cubic nonlinear susceptibility tensor $\overleftrightarrow{\boldsymbol{\chi}}^{(3)}$ by

$$P_i^{(3)}(\overrightarrow{\boldsymbol{r}},t) = \varepsilon_0 \cdot \int_0^\infty \int_0^\infty \int_0^\infty \chi^{(3)}_{ijkl}(\tau_n,\tau_p,\tau_q) \cdot E_j(\overrightarrow{\boldsymbol{r}},t-\tau_n) E_k(\overrightarrow{\boldsymbol{r}},t-\tau_p) E_l(\overrightarrow{\boldsymbol{r}},t-\tau_q)\,\mathrm{d}\tau_n \mathrm{d}\tau_p \mathrm{d}\tau_q, \tag{2.20}$$

where the Einstein summation convention was used again.

The cubic material polarization in the frequency domain is obtained from the Fourier

transform of (2.20) as

$$P_i^{(3)}(\overrightarrow{\boldsymbol{r}}, f) = \varepsilon_0 \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X_{ijkl}^{(3)}(f_n, f_p, f_q) \cdot E_j(\overrightarrow{\boldsymbol{r}}, f_n) \cdot E_k(\overrightarrow{\boldsymbol{r}}, f_p) \cdot E_l(\overrightarrow{\boldsymbol{r}}, f_q) \cdot \delta(f - f_n - f_p - f_q)\,\mathrm{d}f_n\,\mathrm{d}f_p\,\mathrm{d}f_q, \quad (2.21)$$

where

$$X_{ijkl}^{(3)}(f_n, f_p, f_q) = \int_0^{\infty}\int_0^{\infty}\int_0^{\infty} \chi_{ijkl}^{(3)}(t_n, t_p, t_q) \cdot e^{-\jmath 2\pi f_n t_n} \cdot e^{-\jmath 2\pi f_p t_p} \cdot e^{-\jmath 2\pi f_q t_q}\,\mathrm{d}t_n\,\mathrm{d}t_p\,\mathrm{d}t_q. \quad (2.22)$$

It is apparent from (2.21) that any combination of four frequencies that satisfies $f = f_n{+}f_p{+}f_q$ contributes to the material polarization $P_i^{(3)}(\overrightarrow{\boldsymbol{r}}, f)$. For this reason, the nonlinear process that generates cubic material polarization is known as *four-wave mixing* (FWM) or *four-photon mixing*. Relevant to optical communications are frequency combinations of the form $f = f_n{+}f_p{-}f_q$, as these generate FWM products at frequencies within or near the original signal band through $X_{iijj}^{(3)}(f_n, f_p, -f_q)$.

The cubic material polarization can be separated into contributions from electrons and nuclei. While the electrons' response to an incident light wave is instantaneous [Agr01] and can be modeled by a single scalar parameter σ, the contribution of the heavier nuclei is described by two independent time-varying response functions $a(t)$ and $b(t)$ [Hel77].

$\overleftrightarrow{\boldsymbol{X}}^{(3)}$ is a 4^{th}-rank tensor with $3^4 = 81$ elements. However, the number of non-zero components is reduced to 21 by material symmetries, so that only the following elements remain [Han95, Hel77]:

$$X_{iijj}^{(3)}(f_n, f_p, -f_q) = \frac{\sigma + 2A(f_p - f_q) + B(f_n - f_q)}{6 \cdot \varepsilon_0}, \quad (2.23)$$

$$X_{ijij}^{(3)}(f_n, f_p, -f_q) = \frac{\sigma + 2A(f_n - f_q) + B(f_p - f_q)}{6 \cdot \varepsilon_0}, \quad (2.24)$$

$$X_{ijji}^{(3)}(f_n, f_p, -f_q) = \frac{\sigma + B(f_n - f_q) + B(f_p - f_q)}{6 \cdot \varepsilon_0}, \quad (2.25)$$

$$X_{iiii}^{(3)}(f_n, f_p, -f_q) = \frac{3\sigma + 2A(f_n - f_q) + 2A(f_p - f_q) + 2B(f_n - f_q) + 2B(f_p - f_q)}{6 \cdot \varepsilon_0}, \quad (2.26)$$

where $i, j \in \{x, y, z\}$ with $i \neq j$. $A(f)$ and $B(f)$ are the Fourier transforms of $a(t)$ and $b(t)$ respectively. Figure 2.2 shows $A(\Delta f)$ and $B(\Delta f)$ as a function of the frequency difference for SiO_2 [Hel77].

The cubic nonlinear material susceptibility tensor $\overleftrightarrow{\boldsymbol{X}}^{(3)}$ is responsible for all important nonlinear fiber effects with the exception of Brillouin scattering. The Kerr effect, responsible for degradations through self-phase modulation (SPM), cross-phase modulation (XPM) and four-wave mixing (FWM), is caused by $\Re\{\overleftrightarrow{\boldsymbol{X}}^{(3)}\}$, whereas $\Im\{\overleftrightarrow{\boldsymbol{X}}^{(3)}\}$

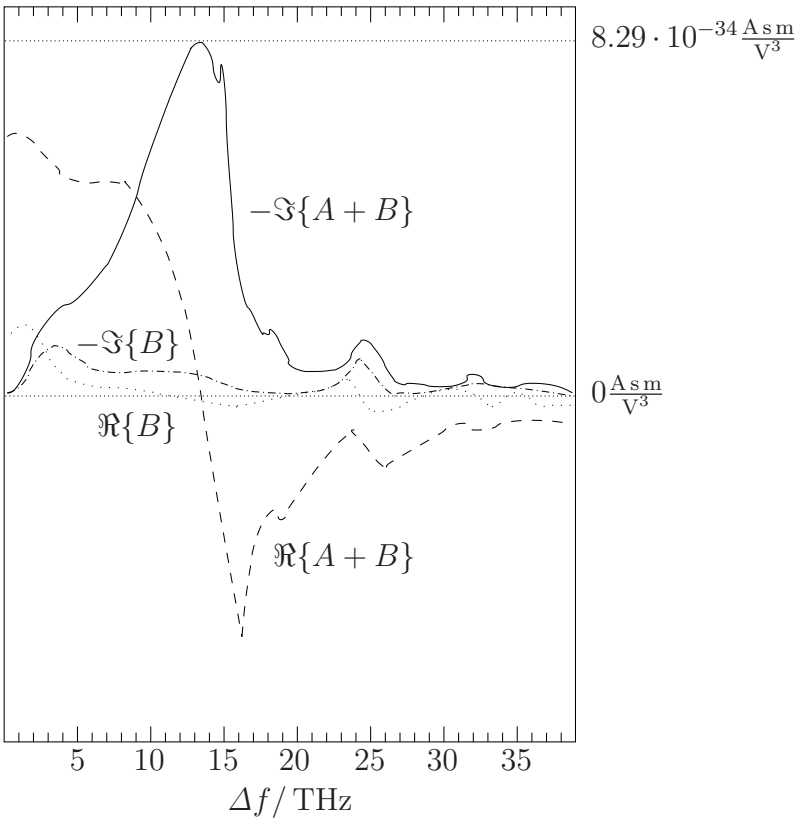

Figure 2.2: Fourier transforms $A(\Delta f)$ and $B(\Delta f)$ of the time-dependent nonlinear susceptibility functions $a(t)$ and $b(t)$ governing the nuclear contribution to the third-order polarization (after [Hel77]).

governs the power exchange between light waves through the inelastic process of Raman scattering [Han95].

In the analysis of these effects, two scenarios need to be distinguished. The maximum of $-\Im\{A+B\}$ depicted in Figure 2.2 lies at $\Delta f = 13.2\,\mathrm{THz}$ [Agr01, Han95]. At this frequency difference, the efficiency of the Raman process is maximized, while it can be neglected for frequency differences smaller than 1 THz. In contrast, distortions due to XPM$^{\mathcal{S}}$, FWM$^{\mathcal{S}}$ and FWM[2] decline with increasing frequency spacing between interacting waves; this will be explained in detail in Section 6.2.1.

The standardized spacing of channels in a dense wavelenth division multiplexing (DWDM) system is up to 100 GHz [ITU02], so that the Raman contribution can be neglected in the analysis of intra-channel effects and interactions between neighboring channels, and Equations (2.23)-(2.26) can be approximated by assuming $\Delta f \to 0$:

$$X^{(3)}_{iijj} = \frac{\sigma + 2A(0) + B(0)}{6 \cdot \varepsilon_0}, \tag{2.27}$$

[2]These terms are introduced in Section 2.2.2.

$$X^{(3)}_{ijij} = \frac{\sigma + 2A(0) + B(0)}{6 \cdot \varepsilon_0} = X^{(3)}_{iijj}, \tag{2.28}$$

$$X^{(3)}_{ijji} = \frac{\sigma + 2B(0)}{6 \cdot \varepsilon_0}, \tag{2.29}$$

$$X^{(3)}_{iiii} = \frac{3\sigma + 4A(0) + 4B(0)}{6 \cdot \varepsilon_0}. \tag{2.30}$$

Table 2.1 lists numeric values for the parameters determining the cubic nonlinear susceptibility. From these values, the nuclear contribution to the nonlinear cubic material susceptibility is calculated to lie between 10.3% and 28.7%.

Table 2.1: Numeric values for the cubic susceptibility parameters in $\mathrm{A\,s\,m\,V^{-3}}$(taken from [Han02]).

σ	$2.7 \cdot 10^{-33}$
$A(0)$	$4.67 \cdot 10^{-34}$
$B(0)$	$1.55 \cdot 10^{-34}$
$\max\limits_{\Delta f} \lvert\Im\{A(\Delta f) + B(\Delta f)\}\rvert$	$8.29 \cdot 10^{-34}$
$\max\limits_{\Delta f} \lvert\Im\{A(\Delta f)\}\rvert$	$7.8 \cdot 10^{-34}$
$\max\limits_{\Delta f} \lvert\Im\{B(\Delta f)\}\rvert$	$5 \cdot 10^{-35}$

It must be noted that in general evaluating the material polarization requires a quantum-mechanical approach [Agr01]. The results presented in this section were derived under the assumption that the shift of (light) electrons and (heavy) nuclei can be separated in the analysis of the susceptibility functions (Born-Oppenheimer approximation) [Han95, Hel77]. This assumption can be made when the optical frequencies used are far away from the material's resonance frequencies. In optical fibers, this is the case in the wavelength range from 0.5 µm to 2 µm, in which optical communication systems operate [Agr01].

2.1.3 The nonlinear Schroedinger equation

The general wave equation (2.9) governs the propagation of light in optical fibers but is not well suited for the practical analysis of this propagation. For that purpose, a simplified equation called nonlinear Schroedinger equation (NLSE) is used. In this section, the NLSE is derived, starting from Equation (2.9).

Fiber modes

A first step is made by observing that the contributions of $\overrightarrow{\boldsymbol{P}}^{(3)}$ as well as of $\overleftrightarrow{\varepsilon}_{r,i}$ and $\overleftrightarrow{\varepsilon}_{r,\Delta}$ to the material polarization $\overrightarrow{\boldsymbol{P}}$ are small compared to that made by $\overleftrightarrow{\varepsilon}_{r,r}$. These small

contributors are therefore neglected at first and added later as perturbative terms. With (2.12), (2.15) and (2.18), the material polarization is simplified to

$$\overrightarrow{\boldsymbol{P}} = \varepsilon_0 \cdot (\varepsilon_{r,r} - 1) \cdot \overrightarrow{\boldsymbol{E}}, \tag{2.31}$$

and the wave equation becomes

$$\Delta \overrightarrow{\boldsymbol{E}} + \nabla \left(\frac{1}{\varepsilon_0} \nabla \cdot \left(\varepsilon_0 (\varepsilon_{r,r} - 1) \cdot \overrightarrow{\boldsymbol{E}} \right) \right) = \frac{\varepsilon_{r,r}}{c_0^2} \cdot \frac{\partial^2 \overrightarrow{\boldsymbol{E}}}{\partial t^2}. \tag{2.32}$$

Using (2.3) with (2.5), (2.32) can be written as

$$\Delta \overrightarrow{\boldsymbol{E}} + \nabla \left(\frac{1}{\varepsilon_{r,r}} (\nabla \varepsilon_{r,r}) \cdot \overrightarrow{\boldsymbol{E}} \right) = \frac{\varepsilon_{r,r}}{c_0^2} \cdot \frac{\partial^2 \overrightarrow{\boldsymbol{E}}}{\partial t^2}. \tag{2.33}$$

In general, Equation (2.33) is solved numerically. For the special case of a step-index fiber, the refractive index $n(r)$ is independent of the spatial coordinates in the fiber's core and cladding, so that the relative permittivity's gradient is zero:

$$\nabla \varepsilon_{r,r} = 0. \tag{2.34}$$

Therefore, for step-index fibers, (2.33) reduces to the well-known wave equation of a homogeneous transparent medium (*homogeneous wave equation*), which can be solved analytically [Agr01]:

$$\Delta \overrightarrow{\boldsymbol{E}} - \frac{\varepsilon_{r,r}}{c_0^2} \cdot \frac{\partial^2 \overrightarrow{\boldsymbol{E}}}{\partial t^2} = 0. \tag{2.35}$$

Equation (2.33) can be solved to yield solutions of the form [Han95]

$$\overrightarrow{\boldsymbol{E}}(\overrightarrow{\boldsymbol{r}}, t) = E \cdot \overrightarrow{\boldsymbol{F}}(x, y) \cdot e^{\jmath\, 2\pi f_0 t} \cdot e^{-\jmath\, \beta(f_0) z}, \tag{2.36}$$

where E is the amplitude (in V/m) of the electrical field at frequency f_0, $\overrightarrow{\boldsymbol{F}}(x, y)$ describes the transversal field distribution and $\beta(f_0)$ is the field's propagation constant at f_0. Each solution of the wave equation is called a *mode*. For a given transversal profile of the refractive index and frequency f_0, the number of modes supported by the fiber and their respective modal distributions $\overrightarrow{\boldsymbol{F}}(x, y)$ and propagation constants $\beta(f_0)$ can be calculated [Agr01].

In single-mode fibers, the difference between the refractive indices in the fiber's core and cladding is small. Such *weakly guiding* fibers support only two orthogonal fundamental modes above a certain cutoff wavelength. For typical values of single-mode fibers – $n_{\text{core}} - n_{\text{cladding}} \approx 0.005$ and core radius $r_{\text{core}} = 4\,\mu\text{m}$ – the cutoff wavelength $\lambda_{\text{cutoff}} \approx 1.2\,\mu\text{m}$ [Agr01]. The fundamental modes are approximately linearly polarized, so

$$\overrightarrow{\boldsymbol{F}}_x(x, y) = \begin{pmatrix} F(x, y) \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \overrightarrow{\boldsymbol{F}}_y(x, y) = \begin{pmatrix} 0 \\ F(x, y) \\ 0 \end{pmatrix}. \tag{2.37}$$

The transversal field distribution $F(x, y)$ can be obtained numerically for arbitrary refractive index profiles. For a certain class of fibers with a power-law refractive index profile, $F(x, y)$ can be approximated by a Gaussian distribution of the form [Mar78]

$$F(r) \approx e^{-(r/r_{\text{eff}})^2}, \tag{2.38}$$

where $r = \sqrt{x^2 + y^2}$ and the effective fiber core radius r_{eff} is a wavelength-dependent fitting parameter that shall be discussed later in this chapter. The approximation (2.38) applies to step- and graded-index fibers but is invalid for fibers with a more complex refractive index profile such as dispersion-shifted fibers.

Nonlinear wave propagation

To obtain the guided modes' transversal field distribution, the terms $\overleftrightarrow{\varepsilon}_{r,i}$ and $\overleftrightarrow{\varepsilon}_{r,\Delta}$ contributing to $\overrightarrow{\boldsymbol{P}}^{(1)}$ as well as $\overrightarrow{\boldsymbol{P}}^{(3)}$ have been neglected. These relatively small terms have a perturbative effect on the light wave's propagation but do not significantly change the modes' transversal field distribution. Therefore, the guided modes still satisfy (2.33), implying $\nabla \cdot (\varepsilon_{r,r} \overrightarrow{\boldsymbol{E}}) = 0$. Using this condition, we can write the general wave equation as

$$\Delta \overrightarrow{\boldsymbol{E}} + \nabla \left(\frac{1}{\varepsilon_{r,r}} (\nabla \varepsilon_{r,r}) \cdot \overrightarrow{\boldsymbol{E}} \right) - \frac{1}{c_0^2} \frac{\partial^2 \overrightarrow{\boldsymbol{E}}}{\partial t^2} = \mu_0 \frac{\partial^2 \overrightarrow{\boldsymbol{P}}}{\partial t^2}. \tag{2.39}$$

The perturbative terms contained in the right-hand side of (2.39) modify the wave amplitude along the propagation, so that the amplitude E in (2.36) becomes z-dependent. A real-valued light wave of arbitrary polarization can be written as

$$\overrightarrow{\boldsymbol{E}}(\overrightarrow{\boldsymbol{r}}, t) = \frac{1}{2} \cdot \left(E_x(z) \cdot \overrightarrow{\boldsymbol{F}}_x(x, y) + E_y(z) \cdot \overrightarrow{\boldsymbol{F}}_y(x, y) \right) \cdot e^{\jmath 2\pi f_0 t} \cdot e^{-\jmath \beta(f_0) z} + \text{ cc}, \tag{2.40}$$

where cc stands for the expression's complex conjugate.

Inserting (2.40) into (2.39) yields (in complex notation) [Han95]

$$\begin{aligned} &- \jmath \cdot \beta(f_0) \cdot \left(\frac{\partial}{\partial z} E_x(z) \cdot \overrightarrow{\boldsymbol{F}}_x(x, y) + \frac{\partial}{\partial z} E_y(z) \cdot \overrightarrow{\boldsymbol{F}}_y(x, y) \right) e^{\jmath (2\pi f_0 t - \beta(f_0) z)} \\ &= \mu_0 \frac{\partial^2}{\partial t^2} \left(-\jmath \cdot \varepsilon_0 \cdot \varepsilon_{r,i} \cdot \overrightarrow{\boldsymbol{E}}(\overrightarrow{\boldsymbol{r}}, t) + \varepsilon_0 \cdot \overleftrightarrow{\varepsilon}_{r,\Delta} \cdot \overrightarrow{\boldsymbol{E}}(\overrightarrow{\boldsymbol{r}}, t) + \overrightarrow{\boldsymbol{P}}^{(3)} \right). \end{aligned} \tag{2.41}$$

The perturbed field's amplitude changes very slowly compared to the light wave's oscillations. Therefore, in obtaining (2.41), terms involving the second-order derivative of the field amplitudes with respect to z have been neglected (so-called *slowly-varying wave approximation*) [Han95]. Additionally, since the transversal field distribution is unchanged by the perturbative terms, those parts of the result that correspond to (2.33) equal zero and do not appear in (2.41).

A major simplification is achieved by reverting the vectorial equation (2.41) to a scalar description. The resulting model governs the propagation of identically linearly polarized

light waves in a perfectly cylindrical fiber or, likewise, the propagation of light exciting one of the fundamental modes of a polarization-maintaining fiber. The randomly varying birefringence and mode coupling in real fibers, modeled through $\overleftrightarrow{\varepsilon}_{r,\Delta}$, entail two important consequences:

- ▷ Birefringence, i. e. polarization-dependent propagation constants, and mode coupling, i. e. the exchange of power between light waves with orthogonal polarizations, lead to different frequency-dependent propagation constants. This linear effect is known as polarization mode dispersion (PMD). The modeling of PMD and its consequences on the capacity of fiber-optic communication systems are discussed in Chaper 4.
- ▷ As seen from Equations (2.23)–(2.26), the state of polarization determines the strength of nonlinear effects. Conversely, the polarization of a light wave can be affected by power present at another frequency through the third-order susceptibility. Therefore, a complete model of light propagation in optical fibers requires two coupled differential propagation equations [Han95]. In the simpler scalar model, it is possible to account for the interplay of light polarization and fiber nonlinearities by introducing an effective nonlinear susceptibility. This approach will be discussed in Section 2.1.6. The effect of the interaction between PMD and nonlinearities on optical communication systems is reviewed in Section 4.4.

To reduce the vectorial propagation equation (2.41) to a scalar description, we assume a light wave exciting only one of the fundamental modes (2.37) and consider only that cartesian component of the field which coincides with the excited mode (the index specifying the cartesian component is dropped in the following):

$$E(\overrightarrow{\boldsymbol{r}},t) = \frac{1}{2}\cdot E(z)\cdot F(x,y)\cdot e^{\jmath\, 2\pi f_0 t}\cdot e^{-\jmath\,\beta(f_0)z} + \text{ cc.} \tag{2.42}$$

Neglecting imperfections in the fiber geometry by omitting the term $\overleftrightarrow{\varepsilon}_{r,\Delta}$, (2.41) becomes

$$-\jmath\cdot\beta(f_0)\cdot\frac{\partial}{\partial z}E(z)\cdot F(x,y)\cdot e^{\jmath\,(2\pi f_0 t-\beta(f_0)z)} = \mu_0\frac{\partial^2}{\partial t^2}\left(-\jmath\cdot\varepsilon_0\cdot\varepsilon_{r,i}\cdot E(\overrightarrow{\boldsymbol{r}},t)+P^{(3)}\right). \tag{2.43}$$

The cubic polarization term $P^{(3)}$ in (2.43) can be evaluated assuming $N+1$ monochromatic, i. e. unmodulated, waves co-propagating at frequencies $f_0, f_1, f_2, \ldots, f_N$, where each wave is described by

$$E^{(n)}(\overrightarrow{\boldsymbol{r}},t) = \frac{1}{2}\cdot E^{(n)}(z)\cdot F^{(n)}(x,y)\cdot e^{\jmath\, 2\pi f_n t}\cdot e^{-\jmath\,\beta(f_n)z} + \text{ cc.} \tag{2.44}$$

The total electric field composed of $N+1$ waves of the form (2.44) is inserted into (2.20). Using the Fourier transform (2.22) and evaluating the result at f_0 yields

$$P^{(3)}(\overrightarrow{\boldsymbol{r}},t)\Big|_{f_0} = \frac{\varepsilon_0}{8}\cdot\sum_{n,p,q}X^{(3)}_{xxxx}(f_n,f_p,f_q)\cdot F^{(n)}(x,y)\cdot F^{(p)}(x,y)\cdot F^{(q)}(x,y)\cdot$$
$$E^{(n)}(z)\cdot E^{(p)}(z)\cdot E^{(q)}(z)\cdot e^{\jmath\, 2\pi f_0 t}\cdot e^{-\jmath\,(\beta(f_n)+\beta(f_p)+\beta(f_q))z}, \tag{2.45}$$

where the summation considers all combinations of three frequencies $f_n, f_p, f_q \in \{\pm f_0, \pm f_1, \ldots, \pm f_{N-1}, \pm f_N\}$ that satisfy $f_0 = f_n + f_p + f_q$.

Inserting (2.42) and (2.45) into (2.43) yields

$$\begin{aligned}
&-\jmath \cdot \beta(f_0) \cdot \frac{\partial}{\partial z} E(z) \cdot F(x,y) \cdot e^{\jmath(2\pi f_0 t - \beta(f_0) z)} \\
&= \frac{\jmath}{2} k_0^2 \varepsilon_{r,i} \cdot E(z) \cdot F(x,y) \cdot e^{\jmath(2\pi f_0 t - \beta(f_0) z)} \\
&\quad - \frac{k_0^2}{8} \cdot \sum_{n,p,q} X^{(3)}_{xxxx}(f_n, f_p, f_q) \cdot F^{(n)}(x,y) \cdot F^{(p)}(x,y) \cdot F^{(q)}(x,y) \\
&\qquad \cdot E^{(n)}(z) \cdot E^{(p)}(z) \cdot E^{(q)}(z) \cdot e^{\jmath 2\pi f_0 t} \cdot e^{-\jmath(\beta(f_n)+\beta(f_p)+\beta(f_q))z},
\end{aligned} \tag{2.46}$$

where $\partial^2/\partial t^2 e^{\jmath 2\pi f_0 t} = -(2\pi f_0)^2 \cdot e^{\jmath 2\pi f_0 t}$ was used and $k_0 = 2\pi f_0 \sqrt{\mu_0 \varepsilon_0}$ is the wavenumber of light in vacuum. Multiplying (2.46) with $\jmath \cdot e^{-\jmath(2\pi f_0 t - \beta(f_0) z)}/\beta(f_0)$ yields

$$\begin{aligned}
\frac{\partial}{\partial z} E(z) \cdot F(x,y) = &-\frac{\alpha}{2} \cdot E(z) \cdot F(x,y) \\
&- \jmath \frac{k_0^2}{8\beta(f_0)} \cdot \sum_{n,p,q} X^{(3)}_{xxxx}(f_n, f_p, f_q) \cdot F^{(n)}(x,y) \cdot F^{(p)}(x,y) \cdot F^{(q)}(x,y) \\
&\qquad \cdot E^{(n)}(z) \cdot E^{(p)}(z) \cdot E^{(q)}(z) \cdot e^{-\jmath \Delta\beta \cdot z},
\end{aligned} \tag{2.47}$$

where the attenuation coefficient $\alpha = k_0^2 \cdot \varepsilon_{r,i}/\beta(f_0)$ was introduced and $\Delta\beta = \beta(f_n) + \beta(f_p) + \beta(f_q) - \beta(f_0)$.

Equation (2.47) describes a propagating electromagnetic wave with field distribution $F(x,y)$. It can be simplified further by reducing the model to that of an equivalent plane wave propagating within an effective area. For this purpose, (2.47) is multiplied by

$$\frac{F(x,y)}{\iint F^2(x,y)\mathrm{d}x\,\mathrm{d}y}, \tag{2.48}$$

where $F(x,y)$ is assumed to be real. Integrating the resulting equation over the entire fiber cross-section, one obtains

$$\begin{aligned}
\frac{\partial}{\partial z} E(z) = &-\frac{\alpha}{2} \cdot E(z) - \jmath \frac{k_0^2}{8\beta(f_0)} \cdot \sum_{n,p,q} X^{(3)}_{xxxx}(f_n, f_p, f_q) \\
&\cdot \frac{\iint F(x,y) F^{(n)}(x,y) F^{(p)}(x,y) F^{(q)}(x,y)\mathrm{d}x\,\mathrm{d}y}{\iint F^2(x,y)\mathrm{d}x\,\mathrm{d}y} \cdot E^{(n)}(z) E^{(p)}(z) E^{(q)}(z) \cdot e^{-\jmath \Delta\beta \cdot z}.
\end{aligned} \tag{2.49}$$

In a last step, (2.49) is multiplied by

$$\sqrt{\frac{\iint F^2(x,y)\mathrm{d}x\,\mathrm{d}y}{A_{\mathrm{eff}}}}, \tag{2.50}$$

where

$$A_{\text{eff}} = \frac{\sqrt{\iint F^2(x,y)\mathrm{d}x\,\mathrm{d}y}\sqrt{\iint F^{(n)^2}(x,y)\mathrm{d}x\,\mathrm{d}y}\sqrt{\iint F^{(p)^2}(x,y)\mathrm{d}x\,\mathrm{d}y}\sqrt{\iint F^{(q)^2}(x,y)\mathrm{d}x\,\mathrm{d}y}}{\iint F(x,y)F^{(n)}(x,y)F^{(p)}(x,y)F^{(q)}(x,y)\mathrm{d}x\,\mathrm{d}y} \tag{2.51}$$

is known as the *effective fiber core area*. Using the following substitution for the field amplitudes,

$$\hat{E}(z) = E(z)\cdot\sqrt{\frac{\iint F^2(x,y)\mathrm{d}x\,\mathrm{d}y}{A_{\text{eff}}}} \tag{2.52}$$

and

$$\hat{E}^{(n)}(z) = E^{(n)}(z)\cdot\sqrt{\frac{\iint F^{(n)^2}(x,y)\mathrm{d}x\,\mathrm{d}y}{A_{\text{eff}}}}, \tag{2.53}$$

(2.49) becomes

$$\begin{aligned}\frac{\partial}{\partial z}\hat{E}(z) = &-\frac{\alpha}{2}\cdot\hat{E}(z)\\ &-\jmath\frac{k_0^2}{8\beta(f_0)}\cdot\sum_{n,p,q}X^{(3)}_{xxxx}(f_n,f_p,f_q)\cdot\hat{E}^{(n)}(z)\cdot\hat{E}^{(p)}(z)\cdot\hat{E}^{(q)}(z)\cdot e^{-\jmath\Delta\beta\cdot z}.\end{aligned} \tag{2.54}$$

The power of the electromagnetic wave described by (2.54) is calculated as [Han95]

$$P = \frac{1}{2Z}\cdot|\hat{E}|^2\cdot A_{\text{eff}}, \tag{2.55}$$

where $Z = Z_0/n_{\text{core}}$ is the wave impedance, $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ being the impedance of free space.

The *effective core area* A_{eff} is weakly dependent on the interacting waves' frequencies. When the Gaussian field approximation (2.38) is valid, (2.51) can be written as

$$A_{\text{eff}}(f_n,f_p,f_q,f_0) = \frac{\pi}{4}\cdot\left(r^{(q)}_{\text{eff}}\cdot r_{\text{eff}}\cdot\frac{r^{(n)^2}_{\text{eff}}+r^{(p)^2}_{\text{eff}}}{r^{(n)}_{\text{eff}}\cdot r^{(p)}_{\text{eff}}}+r^{(n)}_{\text{eff}}\cdot r^{(p)}_{\text{eff}}\cdot\frac{r^{(q)^2}_{\text{eff}}+r^2_{\text{eff}}}{r^{(q)}_{\text{eff}}\cdot r_{\text{eff}}}\right). \tag{2.56}$$

In the case where the spacing between the interacting waves is small, the difference between the effective fiber core radii is negligible and (2.56) becomes

$$A_{\text{eff}} = \pi\cdot r^2_{\text{eff}}. \tag{2.57}$$

A heuristic description of the frequency-dependence of r_{eff} has been reported in [Mar78] (numerical values taken from [Han95]):

$$r_{\text{eff}}(v) = r_{\text{core}}\cdot\left(0.65+\frac{1.619}{v^{1.5}}+\frac{2.879}{v^6}\right), \tag{2.58}$$

where $v = 2\pi r_{\text{core}}\sqrt{n_{\text{core}}^2 - n_{\text{cladding}}^2}/\lambda$ is the normalized frequency and r_{core} is the fiber core radius. For typical values of standard single mode fibers at $\lambda = 1.55\,\mu\text{m}$, $r_{\text{eff}} \approx 5\,\mu\text{m}$ [Han95, Agr01].

For any three frequencies $\pm f_n, \pm f_p, \pm f_q$, the sum term in (2.54) yields 48 combinations. Not all of these are relevant for optical communication systems, as only products of the form $f_0 = f_n + f_p - f_q$ affect frequencies f_0 within or near the original signal band. Therefore, a further simplification is obtained from considering only these frequency combinations. For one particular case, e. g. $f_0 = f_1 + f_2 - f_3$, six out of the original 48 combinations remain. Equation (2.44) describes a real-valued signal, so that the amplitude at the negative frequency is the complex conjugate of that at the positive frequency:

$$E^{(-n)}(z) = E^{(n)^*}(z). \tag{2.59}$$

Applying this conjugation to one of the amplitudes, the summation can be further restricted to only positive frequencies. This reduces the number of cases from six to two, so that a factor 3 is introduced to account for the omitted products. Finally, (2.54) becomes

$$\begin{aligned}\frac{\partial}{\partial z}\hat{E}(z) = &-\frac{\alpha}{2}\cdot\hat{E}(z)\\ &-\jmath\,\frac{k_0^2}{8\beta(f_0)}\cdot 3\cdot X^{(3)}_{xxxx}\cdot\sum_{n,p,q}\hat{E}^{(n)}(z)\cdot\hat{E}^{(p)}(z)\cdot\hat{E}^{(q)^*}(z)\cdot e^{-\jmath\,\Delta\beta\cdot z}\end{aligned} \tag{2.60}$$

with the *phase mismatch*

$$\Delta\beta = \beta(f_n) + \beta(f_p) - \beta(f_q) - \beta(f_0). \tag{2.61}$$

In (2.60), a sufficiently close spacing of the interacting waves is assumed, so that the cubic susceptibility can be replaced by its approximation (2.30).

NLSE in frequency domain

The model of co-propagating monochromatic waves is now extended to govern the propagation of a modulated carrier

$$E(z,t) = \hat{E}\cdot s(z,t)\cdot e^{\jmath\,2\pi f_c t}\cdot e^{-\jmath\,\beta(f_c)z}, \tag{2.62}$$

where $s(z,t)$ denotes the slowly varying complex envelope. Replacing the amplitude $\hat{E}(z)\cdot e^{-\jmath\,\beta(f)z}$ in (2.60) by the Fourier transform of (2.62),

$$E(z,f) = \hat{E}\cdot S(z, f - f_c)\cdot e^{-\jmath\,\beta(f_c)z}, \tag{2.63}$$

the propagation equation for the spectral component at frequency f is obtained as [Han95]

$$\begin{aligned}&\frac{\partial}{\partial z}S(z,f-f_c) + \jmath\cdot\left(\beta(f) - \beta(f_c)\right)\cdot S(z,f-f_c) = -\frac{\alpha}{2}\cdot S(z,f-f_c) -\\ &\jmath\,\frac{k_0^2}{8\beta(f)}\cdot 3\cdot X^{(3)}_{xxxx}\cdot|\hat{E}|^2\cdot\int_0^\infty\int_0^\infty\int_0^\infty S(z,f_n-f_c)\cdot S(z,f_p-f_c)\cdot S^*(z,f_q-f_c)\cdot\\ &\delta\left(f-(f_n+f_p-f_q)\right)\,\mathrm{d}f_n\mathrm{d}f_p\mathrm{d}f_q.\end{aligned} \tag{2.64}$$

The signal bandwidth of $S(z,f)$ is very small compared to the carrier frequency f_c, which is on the order of 200 THz. Therefore, a useful simplification of (2.64) is to set

$$\frac{k_0^2}{8\beta(f)} = \frac{\mu_0\varepsilon_0(2\pi f)^2}{8\beta(f)} \approx \frac{\mu_0\varepsilon_0(2\pi f_c)^2}{8\beta(f_c)} \tag{2.65}$$

and $\alpha = \alpha(f) \approx \alpha(f_c)$. In addition, the phase constant $\beta(\omega)$ is developed into a Taylor series about $\omega_c = 2\pi f_c$:

$$\beta(\omega) \approx \beta(\omega_c) + \beta_1 \cdot (\omega - \omega_c) + \frac{\beta_2}{2} \cdot (\omega - \omega_c)^2 + \frac{\beta_3}{6} \cdot (\omega - \omega_c)^3 + \ldots, \tag{2.66}$$

where

$$\beta_m = \frac{\mathrm{d}^m \beta(\omega)}{\mathrm{d}\omega^m}\Big|_{\omega=\omega_c}. \tag{2.67}$$

Considering the usable bandwidth of optical fibers, the fourth and higher-order terms in this expansion are generally negligible. The cubic term can be neglected in many practical cases as well; however, it must be included when the signal is located near the zero-dispersion wavelength of the fiber, i. e. when $\beta_2 \approx 0$. With these approximations, (2.64) becomes

$$\begin{aligned}
&\left(\frac{\partial}{\partial z} + \jmath \cdot \beta_1 \cdot (\omega - \omega_c) + \jmath \cdot \frac{\beta_2}{2} \cdot (\omega - \omega_c)^2 + \jmath \cdot \frac{\beta_3}{6} \cdot (\omega - \omega_c)^3\right) \cdot S(z, f - f_c) = \\
&-\frac{\alpha}{2} \cdot S(z, f - f_c) - \jmath \frac{\mu_0\varepsilon_0(2\pi f_c)^2}{8\beta(f_c)} \cdot 3 \cdot X^{(3)}_{xxxx} \cdot |\hat{E}|^2 \cdot \int_0^\infty \int_0^\infty \int_0^\infty \\
&S(z, f_n - f_c) \cdot S(z, f_p - f_c) \cdot S^*(z, f_q - f_c) \cdot \delta\left(f - (f_n + f_p - f_q)\right) \,\mathrm{d}f_n \mathrm{d}f_p \mathrm{d}f_q.
\end{aligned} \tag{2.68}$$

Equation (2.68) is called the *nonlinear Schrödinger equation* (NLSE) in the frequency domain.

NLSE in time domain

Applying an inverse Fourier transform to (2.68) yields [Han95]

$$\begin{aligned}
&\left(\frac{\partial}{\partial z} + \beta_1 \frac{\partial}{\partial t} - \jmath \cdot \frac{\beta_2}{2} \frac{\partial^2}{\partial t^2} - \frac{\beta_3}{6} \frac{\partial^3}{\partial t^3}\right) \cdot s(z,t) = \\
&-\frac{\alpha}{2} \cdot s(z,t) - \jmath \frac{\mu_0\varepsilon_0(2\pi f_c)^2}{8\beta(f_c)} \cdot 3 \cdot X^{(3)}_{xxxx} \cdot |\hat{E}|^2 \cdot |s(z,t)|^2 \cdot s(z,t).
\end{aligned} \tag{2.69}$$

Equation (2.69) is the *nonlinear Schrödinger equation* (NLSE) in the time domain. This term is used because of the NLSE's similarity to the Schrödinger equation in quantum mechanics.

The nonlinearity contained in the right-hand side of (2.69) can be attributed to the *Kerr effect*, i. e. to a power-dependence of the refractive index, resulting in a power-dependent

phase rotation of the signal. This becomes more obvious in a different formulation of the NLSE (2.69):

$$\left(\frac{\partial}{\partial z} + \beta_1 \frac{\partial}{\partial t} - \jmath \cdot \frac{\beta_2}{2}\frac{\partial^2}{\partial t^2} - \frac{\beta_3}{6}\frac{\partial^3}{\partial t^3} + \frac{\alpha}{2}\right) \cdot s(z,t) = -\jmath \cdot \gamma \cdot P(z,t) \cdot s(z,t). \tag{2.70}$$

In (2.70),

$$P(z,t) = \frac{1}{2Z} \cdot |\hat{E}|^2 \cdot |s(z,t)|^2 \cdot A_{\text{eff}} \tag{2.71}$$

is the signal power (cf. (2.55)) and

$$\gamma = \frac{\mu_0 \cdot \varepsilon_0 \cdot (2\pi f_c)^2}{8\beta(f_c)} \cdot 3 \cdot X^{(3)}_{xxxx} \cdot \frac{2Z}{A_{\text{eff}}} = \frac{2\pi f_c}{c_0 \cdot A_{\text{eff}}} \cdot \underbrace{\frac{2Z_0 \cdot 3 \cdot X^{(3)}_{xxxx}}{8 \cdot n_{\text{core}}^2}}_{n_2} \tag{2.72}$$

is called *nonlinear coefficient* [Sch04] or *nonlinear parameter* [Agr01]. It has units of $\mathrm{W}^{-1}\,\mathrm{m}^{-1}$. In (2.72),

$$\beta(f_c) \approx \frac{2\pi f_c}{c_0} \cdot n_{\text{core}} \tag{2.73}$$

was used. The nonlinear parameter γ is proportional to the *nonlinear index coefficient* [Agr01] or *nonlinear Kerr constant* [Sch04] n_2, measured in units of $\mathrm{m}^2 \,/\, \mathrm{W}$. The total refractive index observed in the presence of the Kerr effect is

$$n_{\text{total}} = n_{\text{core}} + n_2 \cdot \frac{P}{A_{\text{eff}}}. \tag{2.74}$$

In practice, different fiber types are characterized by the nonlinear parameters γ, n_2 and A_{eff}. Section 2.2.1 discusses these parameters as well as the linear coefficients governing attenuation and dispersion for various fiber types.

As mentioned above, β_3 is usually neglected unless β_2 is near zero. An additional simplification of the NLSE comes from observing the signal propagation relative to a reference frame moving along the fiber with the signal's group velocity

$$v_g = \frac{1}{t_g} = \frac{1}{\beta_1}. \tag{2.75}$$

By making the substitution

$$t' = t - z/v_g = t - \beta_1 \cdot z, \tag{2.76}$$

the NLSE becomes

$$\left(\frac{\partial}{\partial z} - \jmath \cdot \frac{\beta_2}{2}\frac{\partial^2}{\partial t'^2} + \frac{\alpha}{2}\right) \cdot s(z,t') = -\jmath \cdot \gamma \cdot P(z,t') \cdot s(z,t'). \tag{2.77}$$

The literature (see e. g. [Agr02]) frequently defines a normalized signal

$$a(z,t) = \sqrt{\frac{A_{\text{eff}}}{2Z}} \cdot \hat{E} \cdot s(z,t), \tag{2.78}$$

such that $|a|^2 = P$. The NLSE can then be written as

$$\left(\frac{\partial}{\partial z} - \jmath \cdot \frac{\beta_2}{2}\frac{\partial^2}{\partial t'^2} + \frac{\alpha}{2}\right) \cdot a(z,t') = -\jmath \cdot \gamma \cdot |a(z,t')|^2 \cdot a(z,t'). \tag{2.79}$$

The corresponding frequency-domain NLSE governing the propagation of $A(z,f)$ (the Fourier transform of $a(z,t)$) can be obtained by performing a Fourier transform to (2.79). Alternatively, (2.68) can be modified using (2.72), (2.76), (2.78), $f_p = f - f_n + f_q$ and the substitution $f' = f - f_c$ to read

$$\begin{aligned}&\left(\frac{\partial}{\partial z} + \jmath \cdot \frac{\beta_2}{2} \cdot \omega'^2 + \frac{\alpha}{2}\right) \cdot A(z,f') = \\ &-\jmath \cdot \gamma \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(z,f'_n) \cdot A^*(z,f'_q) \cdot A(z,f' - f'_n + f'_q)\,\mathrm{d}f'_n \mathrm{d}f'_q.\end{aligned} \tag{2.80}$$

The prime signs denoting shifted time and frequency are dropped in the following. To avoid confusion, this is noted throughout the thesis where necessary.

2.1.4 Raman scattering

Raman scattering is an inelastic scattering process, in which photons of a higher frequency f_P are absorbed by the material's molecules (SiO_2 for fibers), exciting them from their ground state to a higher energy state. Upon return to a lower energy state, the molecule emits a photon with energy $h \cdot f_S$, where $f_S < f_P$. The effect was discovered by and named after Indian physicist C. Raman. Allegedly, it was the beautiful blue color of the Mediterranean sea during an extended ship travel in 1924 that inspired Raman to investigate the scattering of light in fluids and gases. Raman received the Nobel prize in 1930 for his discovery. Unknown to Raman, the effect named after him had been predicted theoretically as early as 1923 by A. Smekal. Its experimental demonstration was described in 1928 by Raman and, simultaneously and independently, by G. Landsberg and L. Mandelstam [Sch04].

Figure 2.3 illustrates the energy exchange of light and matter during Raman scattering. A photon of the so-called *pump wave* with energy $h \cdot f_P$ is absorbed by a molecule, exciting it from its ground state (G in Figure 2.3) to a virtual state (V), which for the molecule is a forbidden energy state [Han95]. Therefore, simultaneously to the destruction of the pump photon, a new (so-called *Stokes*) photon of energy $h \cdot f_S$ is emitted and the molecule falls down to a lower intermediate energy level (I), from where it returns to its ground state relatively quickly ($\tau_I \approx 10^{-12}\,\mathrm{s}$). The difference energy $h \cdot (f_P - f_S) = h \cdot \Delta f$ remains in the material as vibrations. Unlike the effects governed by the real part of the material susceptibility, Raman scattering is an inelastic process in the sense that an energy transfer between photon and molecule takes place.

Stimulated Raman scattering (SRS) denotes the case where the emission of the Stokes photon is stimulated by an already existing Stokes photon. In this case, the newly emitted

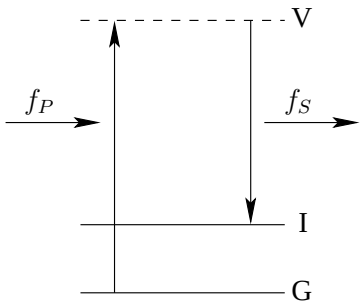

Figure 2.3: Energy diagram illustrating Raman scattering.

photon is identical to the stimulating photon in frequency, phase, polarization and direction. In practice, SRS has great relevance for optical amplification (see Section 2.3.2), where signals at $f_c = f_S$ are amplified using a pump wave at f_P, and wavelength division multiplexing (WDM) systems, where signal power is transferred from higher-frequency to lower-frequency channels through SRS.

Much less frequently than the down-conversion of a pump to a Stokes wave, the inverse process occurs. A molecule in state I absorbs a pump photon and, upon emission of an *anti-Stokes* photon of energy $h \cdot (f_P + \Delta f)$, returns to its ground state.

The evolution of the pump and Stokes wave is described by two coupled differential equations, which are derived from the propagation equation (2.60) in the following. In the frequency relation $f_0 = f_n + f_p - f_q$ assumed by (2.60), we set

$$f_0 = f_n = f_S \qquad \text{and} \qquad f_p = f_q = f_P. \tag{2.81}$$

Obviously, $\Delta\beta = \beta(f_S) + \beta(f_P) - \beta(f_P) - \beta(f_S) = 0$, so Raman scattering is always phase matched. With (2.81), the propagation equation for the Stokes wave at f_S is

$$\frac{\partial}{\partial z}\hat{E}^{(S)}(z) = -\frac{\alpha(f_S)}{2} \cdot \hat{E}^{(S)}(z) - \jmath\,\frac{k_0^2}{8\beta(f_S)} \cdot 6 \cdot X^{(3)}_{xxxx} \cdot \hat{E}^{(S)}(z) \cdot |\hat{E}^{(P)}(z)|^2. \tag{2.82}$$

To obtain an expression for the evolution of the wave's intensity during propagation along the fiber using

$$\frac{\partial}{\partial z}|\hat{E}^{(S)}(z)|^2 = \hat{E}^{(S)}(z) \cdot \frac{\partial}{\partial z}\hat{E}^{(S)^*}(z) + \hat{E}^{(S)^*}(z) \cdot \frac{\partial}{\partial z}\hat{E}^{(S)}(z), \tag{2.83}$$

the propagation equation of $\hat{E}^{(S)^*}(z)$ must be calculated. Using

$$f_0 = f_n = -f_S, \tag{2.84}$$

$$\hat{E}^{(S)}(-f_S) = \hat{E}^{(S)^*}(f_S), \tag{2.85}$$

$$X^{(3)}_{xxxx}(-f_S, f_P, -f_P) = X^{(3)^*}_{xxxx}(f_S, f_P, -f_P), \tag{2.86}$$

$$\beta(-f_S) = -\beta(f_S) \quad \text{and} \quad \alpha(-f_S) = \alpha(f_S) \tag{2.87}$$

yields

$$\frac{\partial}{\partial z}\hat{E}^{(S)^*}(z) = -\frac{\alpha(f_S)}{2}\cdot\hat{E}^{(S)^*}(z) + \jmath\,\frac{k_0^2}{8\beta(f_S)}\cdot 6\cdot X^{(3)^*}_{xxxx}\cdot\hat{E}^{(S)^*}(z)\cdot|\hat{E}^{(P)}(z)|^2. \tag{2.88}$$

Inserting (2.82) and (2.88) into (2.83) results in

$$\begin{aligned}\frac{\partial}{\partial z}|\hat{E}^{(S)}(z)|^2 =& -\alpha(f_S)\cdot|\hat{E}^{(S)}(z)|^2 \\ &+ 6\cdot\frac{k_0^2}{8\beta(f_S)}\cdot 2\cdot\Im\{X^{(3)}_{xxxx}(f_S,f_P,-f_P)\}\cdot|\hat{E}^{(S)}(z)|^2\cdot|\hat{E}^{(P)}(z)|^2.\end{aligned} \tag{2.89}$$

With (2.55), $Z=\sqrt{\mu_0/\varepsilon_0}/n_{\text{core}}$, $k_0=2\pi f_S\sqrt{\mu_0\varepsilon_0}$, and $\beta(f_S)\approx 2\pi f_S\cdot\sqrt{\mu_0\varepsilon_0}\cdot n_{\text{core}}(f_S)$, one obtains a propagation equation for the Stokes wave power:

$$\begin{aligned}\frac{\partial}{\partial z}P^{(S)}(z) =& -\alpha(f_S)\cdot P^{(S)}(z) \\ &+ \underbrace{3\cdot\frac{2\pi f_S\cdot\mu_0}{n_{\text{core}}(f_S)\cdot n_{\text{core}}(f_P)}\cdot\Im\{X^{(3)}_{xxxx}(f_S,f_P,-f_P)\}}_{G_R(f_P,f_S)}\cdot\frac{P^{(P)}(z)}{A_{\text{eff}}}\cdot P^{(S)}(z).\end{aligned} \tag{2.90}$$

$G_R(f_P,f_S)$ is the *Raman gain spectrum*, whose shape has been obtained from measurements, e. g. [SI73, SLJ84, MBG+96]. It is related to the nonlinear susceptibility functions depicted in Figure 2.2 through (2.90) and (2.26) for linearly co-polarized pump and Stokes waves. Various analytical expressions for the Raman gain spectrum and the corresponding Raman impulse response $h_R(t)$ have been reported [WK82, BW89, SGTH89, HC02, LA06]. Essentially, all these approaches are based on phenomenological expressions whose parameters are chosen such to provide an optimum agreement with the measured data. The particularly precise model of Hollenbeck and Cantrell describes $h_R(t)$ as a weighted superposition of damped harmonic oscillator functions reflecting the molecules' vibrational modes [HC02].

The Raman gain spectrum obtained from this model is depicted in Figure 2.4. It can be seen that the Raman gain is maximum for a frequency shift $\Delta f\approx 13.2\,\text{THz}$, where it reaches the value [SLJ84]

$$G_{R,\max} = 1.86\cdot 10^{-13}\frac{\text{m}}{\text{W}}\cdot\frac{532\,\text{nm}}{\lambda_P} = 9.895\cdot 10^{-20}\frac{\text{m}^2}{\text{W}}\cdot\frac{f_P}{c_0}. \tag{2.91}$$

Hence, the amplitude of G_R scales linearly with the pump wave frequency. The maximum value of $-\Im\{A+B\}$ given in Figure 2.2 has been obtained from (2.91) using the definition of G_R in (2.90) and $n_{\text{core}}(f_S)\approx n_{\text{core}}(f_P)\approx 1.51$ [Han95]. The scalar model used in the derivation of (2.90) implies linearly co-polarized pump and Stokes waves. When this is not the case, the Raman gain is reduced by a factor of 2 as discussed in Section 2.1.6. For $\lambda_P = 1550\,\text{nm}$, (2.91) yields $G_{R,\text{eff},\max}\approx 3.2\cdot 10^{-14}\,\text{m}\,/\,\text{W}$. A slightly different value of

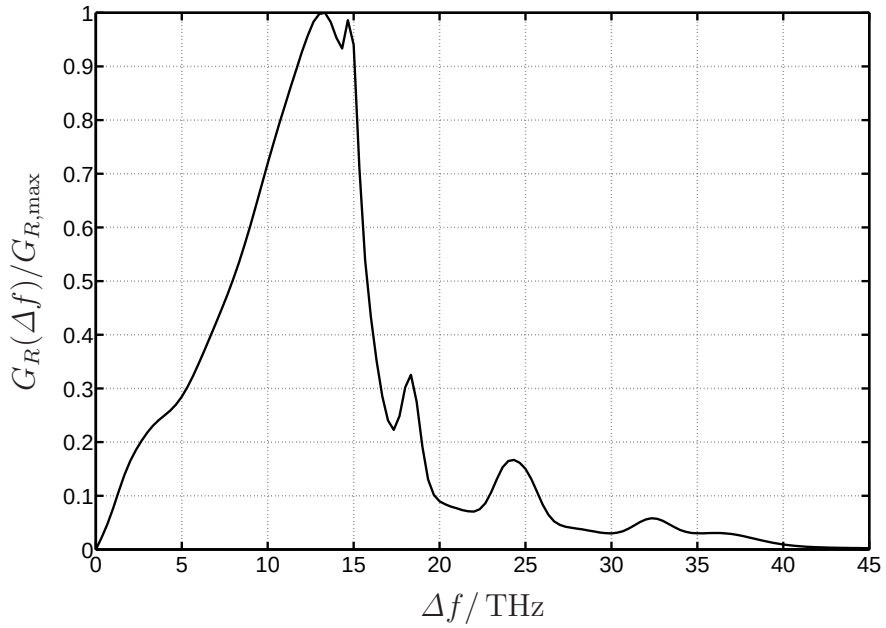

Figure 2.4: Normalized Raman gain spectrum over $\Delta f = f_P - f_S$.

$G_{R,\text{eff},\max} \approx 3.1 \cdot 10^{-14}\,\text{m}\,/\,\text{W}$ was reported in [MBG+96] for a standard single-mode fiber. In general, the Raman gain spectrum and its maximum value vary among different fiber types, depending on the fiber core doping and the fiber geometry [Sch04, Bro04].

Raman scattering reaches its maximum efficiency for $\Delta f = 13.2\,\text{THz}$. In view of this large difference, the frequency dependence of the parameters α (see Figure 2.5), n_{core} (see Figure 2.1) and A_{eff} used in (2.90) must be taken into account. A_{eff} is obtained from (2.51). Under the Gaussian field approximation, A_{eff} is given by (2.56) with (2.58). For the case of one pump and one Stokes wave, this simplifies to

$$A_{\text{eff}} = \frac{\pi}{2} \cdot \left(r_{\text{eff}}^{(S)^2} + r_{\text{eff}}^{(P)^2} \right). \tag{2.92}$$

From (2.90), the evolution of the Stokes wave intensity $I^{(S)} = P^{(S)}/A_{\text{eff}}$ is

$$\frac{\partial}{\partial z} I^{(S)}(z) = -\alpha(f_S) \cdot I^{(S)}(z) + G_R(f_P, f_S) \cdot I^{(P)}(z) \cdot I^{(S)}(z). \tag{2.93}$$

Analogously to the derivation of (2.93), the equation governing the pump wave intensity can be calculated as

$$\pm \frac{\partial}{\partial z} I^{(P)}(z) = -\alpha(f_P) \cdot I^{(P)}(z) - \frac{f_P}{f_S} \cdot G_R(f_P, f_S) \cdot I^{(S)}(z) \cdot I^{(P)}(z), \tag{2.94}$$

where the minus sign on the left hand side applies to the case of a counter-propagating pump wave, i. e. traveling in the direction of decreasing z.

In contrast to stimulated Raman scattering, *spontaneous Raman scattering* describes the spontaneous emission of Stokes light. When the pump power exceeds a certain threshold,

spontaneously emitted Stokes light becomes amplified, leading to significant conversion from pump power to Stokes power. This *critical pump power* $P_{\text{critical,R}}$ is defined as that value of the pump wave input power at which the entire pump power is converted to Stokes power. It has been determined as [Smi72]:

$$P_{\text{critical,R}} \approx 16 \cdot \frac{A_{\text{eff}}}{L_{\text{eff}} \cdot G_{R,\max}}, \tag{2.95}$$

where $L_{\text{eff}} = \left(1 - e^{-\alpha(f_P) \cdot L}\right) / \alpha(f_P)$ is the *effective fiber length* (2.142) introduced in Section 2.2.3. In obtaining (2.95), *pump depletion* (i. e. the second term on the right hand side of (2.94)) has been neglected. Even so, the result for $P_{\text{critical,R}}$ is accurate, as confirmed by [AY78], where (2.93) and (2.94) are solved analytically for the case $\alpha_s = \alpha_p$. For counter-propagating pump and Stokes wave, the critical power is increased by approximately 25%, so the numerical factor 16 in (2.95) becomes 20 [Smi72].

For a long fiber, $L_{\text{eff}} \to 1/\alpha(f_P)$. As an example, the critical power for a sufficiently long standard single-mode fiber (SSMF) with $\lambda_P = 1550\,\text{nm}$, $\alpha = 0.2\,\text{dB}\,/\,\text{km}$ and $A_{\text{eff}} = 80(\mu\text{m})^2$ is $P_{\text{critical,R}} \approx 0.92\,\text{W}$. This value is doubled when the polarization of the pump wave is randomly scrambled (cf. Section 2.1.6).

As signal power levels are usually well below the values obtained for $P_{\text{critical,R}}$, spontaneous Raman scattering is not a critical issue in optical communication systems and is not considered in this thesis.

2.1.5 Brillouin scattering

Brillouin scattering (named after Léon Brillouin) has its origin in the reflection of light at material density fluctuations. These density fluctuations propagate through the medium at the speed of sound and can be regarded as acoustic waves [Sch04]. A light wave is partly scattered at the periodic density fluctuations of the acoustic wave. This scattering process is maximized when the scattered light waves interfere constructively [Han95]. As the acoustic wave propagates at the speed of sound, the scattered light experiences a frequency shift through the Doppler effect. As in the case of Raman scattering, original and scattered light wave are denoted by the terms *pump* and *Stokes* wave, and their frequencies are f_P and f_S, respectively.

Caused by thermal motion, numerous acoustic waves of different wavelength and direction occur in the fiber. Consequently, a very small fraction of light is constantly scattered into all directions through *spontaneous Brillouin scattering*. When a Stokes wave propagating in opposite direction to the pump wave builds up through spontaneous Brillouin scattering from a pump with high power (or if Stokes light is coupled into the fiber at the opposite end, e. g. in a fiber that is used for bi-directional communication), the interference of pump and Stokes waves will amplify the acustic wave through the effect of *electrostriction* [VP02]. This case is called *stimulated Brillouin scattering* (SBS). Because increasing the pump power beyond a certain threshold leads to a total reflection of the light wave, SBS limits the maximum light power that can be used for transmission over optical fibers [Han95].

Analogously to (2.93) and (2.94), the evolution of the SBS pump and Stokes wave intensities can be described as

$$\frac{\partial}{\partial z} I^{(S)}(z) = \alpha(f_S) \cdot I^{(S)}(z) - G_B(f_P, f_S) \cdot I^{(P)}(z) \cdot I^{(S)}(z) \tag{2.96}$$

and

$$\frac{\partial}{\partial z} I^{(P)}(z) = -\alpha(f_P) \cdot I^{(P)}(z) - G_B(f_P, f_S) \cdot I^{(S)}(z) \cdot I^{(P)}(z), \tag{2.97}$$

where the inverted sign in (2.96) reflects the Stokes wave's counter-propagation [Han95]. The *Brillouin gain spectrum* $G_B(f_P, f_S)$ can be shown to have a Lorentzian shape [Agr01]. Compared to Raman scattering, the bandwidth of SBS is very small; typical values lie in the range of 10 MHz to 30 MHz [VP02]. The Brillouin gain is independent of the pump frequency, its bandwidth scales with λ_P^{-2}. The SBS efficiency is maximized at $\Delta f = f_P - f_S \approx 11\,\text{GHz}$, where it reaches a value of $G_{B,\max} \approx 5 \cdot 10^{-11}\,\text{m}\,/\,\text{W}$ for co-polarized pump and Stokes waves [Agr01]. This value is halved when the pump wave's polarization is completely scrambled [Agr01, Han95]. Precise numerical values characterizing the Brillouin gain function have been obtained experimentally for various fiber types [NTR97, YDT02].

As for Raman scattering, a critical power $P_{\text{critical,B}}$ can be specified for Brillouin scattering above which the spontaneous scattering process becomes significantly stimulated. Various definitions for $P_{\text{critical,B}}$ exist [Sch04]; Smith defines $P_{\text{critical,B}}$ as the pump power level at which the backscattered Stokes wave power is equal to the pump power and obtains [Smi72]

$$P_{\text{critical,B}} \approx 21 \cdot \frac{A_{\text{eff}}}{L_{\text{eff}} \cdot G_{B,\max}}. \tag{2.98}$$

For standard single-mode fibers (SSMF) at $\lambda = 1.55\,\mu\text{m}$, $P_{\text{critical,B}} \approx 3\,\text{mW}$. As light propagation beyond the critical power is effectively suppressed through SBS, this sets a tight upper limit on the maximum launch power that can be used for transmission. However, the power threshold given by (2.98) implies a monochromatic pump wave. When the pump wave is modulated, the Brillouin gain reduces as the data rate (i. e. the spectral width of the pump signal) increases. This relation is not observed for modulation schemes with a discrete DC component, i. e. with a Dirac delta at $f = 0$ in the power spectral density (PSD) function, so that the launch power of these modulation schemes is more severely limited by SBS [Han95]. Signals that lack a discrete DC component – e. g. on-off keying (OOK) signals with carrier-suppressed return-to-zero (CSRZ) pulses or duobinary-coded signals – may still contain arbitrarily long sequences of identical amplitude levels, so that SBS can occur temporarily. This can be avoided by use of DC-free line coding schemes such as alternate mark inversion (AMI). Other means of raising the critical power include intentionally increasing the transmit laser linewidth (so-called *SBS suppression*) through frequency dithering and adapting the fiber parameters, e. g. increasing the core diameter [Sch04].

Because the SBS gain reaches its maximum at $\Delta f \approx 11\,\text{GHz}$ and the gain bandwidth is small, Brillouin scattering does not induce any crosstalk between WDM channels. If the maximum power in each WDM channel is below the critical value, SBS will not occur.

Impairments caused by SBS are not considered in this thesis. It is assumed that the precautions mentioned above are employed in a practical system so that SBS does not occur. From a theoretic point of view, SBS defines upper limits on the PSD of the transmit signal. The analysis of other nonlinear fiber effects such as XPM, FWM or SRS caused by the cubic material polarization yields optimum (in the sense of maximum channel capacity) transmit power levels that are below the SBS threshold. Therefore, SBS plays a subordinate role in fiber-optic communication systems that have been optimized according to information-theoretic results.

2.1.6 Effect of light polarization on the nonlinear propagation

As discussed in Section 2.1.3, two orthogonal fundamental modes are able to propagate in a single-mode fiber. In deriving the scalar NLSE, an ideal geometry of the fiber was assumed, so that the fundamental modes have identical propagation properties. In practice, however, all fibers exhibit some *modal birefringence*, i. e. a slight difference between the effective refractive indices n_x and n_y, which has its origin in imperfections during fabrication or in external stress such as pressure or torsion [Agr01]. Two coupled differential equations are therefore required to completely describe light propagation in a single-mode fiber [Han02].

Therefore, before the scalar model is used for the analysis of the fiber-optic channel, its validity in the presence of random polarization changes needs to be discussed. Primarily, birefringence and mode-coupling lead to the linear effect of *polarization mode dispersion* (PMD), which is discussed in Chapter 4. Changes in the light polarization due to mode coupling lead to a continuous power exchange between a light wave's x- and y-componenent. These cartesian components of the propagating waves determine the strength of fiber nonlinearities, caused by the cubic susceptibility (2.23)–(2.26) or (2.27)–(2.30), respectively. Therefore, an averaged *effective cubic susceptibility* $X_{\text{eff}}^{(3)}$ is expected. In this section, the value of $X_{\text{eff}}^{(3)}$ is derived for different cases.

For a short segment of fiber whose length is smaller than the *correlation length* L_{corr} (cf. Section 2.2.3), a constant birefringence profile can be assumed. In such a fiber segment (or likewise in *polarization-maintaining fibers* (PMF) which have an intentionally strong birefringence profile that is kept constant over the entire fiber length), the *state of polarization* (SOP) of a light wave coupled into the fiber will evolve periodically and return to its initial SOP after the *beat length* L_B (cf. Section 2.2.3, (2.151)). Each such fiber has two orthogonal SOP's called *eigenpolarizations* in which light waves maintain their SOP's during propagation along the fiber. These eigenpolarizations are described in Jones space [GK00] as [Han02]

$$\overrightarrow{\boldsymbol{e}}_{\psi} = \frac{1}{\sqrt{1+|c_{\psi}|^2}} \cdot \begin{pmatrix} 1 \\ c_{\psi} \end{pmatrix} \tag{2.99}$$

and

$$\overrightarrow{\boldsymbol{e}}_{\chi} = \frac{1}{\sqrt{1+|c_{\chi}|^2}} \cdot \begin{pmatrix} 1 \\ c_{\chi} \end{pmatrix}. \tag{2.100}$$

On the Poincaré sphere, the eigenpolarizations are two points located on opposite sides of the surface. During propagation along the beat length of the fiber, the SOP of a light wave evolves as a circle around one of the eigenpolarizations on the Poincaré sphere's surface. Causes that lead to linear eigenpolarizations (located on the equator of the Poincaré sphere) include an elliptical fiber core, bending of the fiber and transversal pressure. For linear eigenpolarizations, $c_\psi, c_\chi \in \mathbb{R}$. Torsion of the fiber leads to circular eigenpolarizations (located on the Poincaré sphere's poles); in that case, $c_\psi = \pm\jmath$ and $c_\chi = \mp\jmath$. Generally, the eigenpolarizations can be elliptical and $c_\psi, c_\chi \in \mathbb{C}$ [Han02].

To determine the effective cubic susceptibility $X^{(3)}_{\text{eff}}$, the fiber link is decomposed into segments of length L_c. Within each segment, constant eigenpolarizations are assumed. The eigenpolarizations of successive segments are independent random variables. Because the SOP of a light wave evolves rapidly according to the birefringence in each segment, the SOP at the input of each segment can be considered completely random as well. As light propagates through many such segments, its SOP draws a random trajectory on the Poincaré sphere, resulting in an averaged cubic susceptibility. Because the *effective length* L_{eff} along which fiber nonlinearities are active (cf. Section 2.2.3) is much larger than the correlation length L_{corr}, an ergodicity argument allows to obtain $X^{(3)}_{\text{eff}}$ as an average over an ensemble of random fiber segments rather than as an average over the entire fiber length.

Small frequency spacing

The first case considers four co-propagating waves at frequencies f_0, f_n, f_p and f_q with identical input SOP $\overrightarrow{\boldsymbol{e}}$. When a sufficiently close spacing of the frequencies is assumed, all four waves undergo identical polarization changes, i. e. the four waves are in a common SOP at every point along the fiber length (even though this SOP changes rapidly and randomly during propagation). Neglecting rapidly oscillating terms that will not yield significant nonlinear contributions, Hanik [Han02] obtains the following propagation equations for the wave amplitude at f_0 and for the orthogonal component at the same frequency (which is zero at the fiber input):

$$\frac{\partial}{\partial z}\hat{E}(z) = -\frac{\alpha}{2}\cdot\hat{E}(z) - \jmath\,\frac{k_0^2}{8\beta(f_0)}\cdot 6\cdot e^{-\jmath\,\Delta\beta\cdot z}\hat{E}^{(n)}(z)\cdot\hat{E}^{(p)}(z)\cdot\hat{E}^{(q)^*}(z)\cdot \left(D_\parallel\cdot\left(\left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^*_\chi\right|^4 + \left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^*_\psi\right|^4\right) + 4\cdot D_\perp\cdot\left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^*_\chi\right|^2\cdot\left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^*_\psi\right|^2\right), \quad (2.101)$$

$$\frac{\partial}{\partial z}\hat{E}_\perp(z) = -\frac{\alpha}{2}\cdot\hat{E}_\perp(z) - \jmath\,\frac{k_0^2}{8\beta(f_0)}\cdot 6\cdot e^{-\jmath\,\Delta\beta\cdot z}\hat{E}^{(n)}(z)\cdot\hat{E}^{(p)}(z)\cdot\hat{E}^{(q)^*}(z)\cdot (D_\parallel - 2D_\perp)\cdot\left(\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^*_\chi\right)\cdot\left(\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^*_\psi\right)\cdot\left(\left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^*_\chi\right|^2 - \left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^*_\psi\right|^2\right). \quad (2.102)$$

Using the ellipticity ε and inclination ϕ that define a SOP, the vector products are calculated as

$$\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^*_\chi = \cos(\varepsilon-\varepsilon_\chi)\cdot\cos(\phi-\phi_\chi) + \jmath\cdot\sin(\varepsilon+\varepsilon_\chi)\cdot\sin(\phi-\phi_\chi) \quad (2.103)$$

and

$$\overrightarrow{\boldsymbol{e}} \cdot \overrightarrow{\boldsymbol{e}}_{\psi}^{*} = \cos(\varepsilon + \varepsilon_\chi) \cdot \sin(\phi - \phi_\chi) - \jmath \cdot \sin(\varepsilon - \varepsilon_\chi) \cdot \cos(\phi - \phi_\chi). \tag{2.104}$$

In (2.101) and (2.102), the coupling parameters $D_\parallel$ and $D_\perp$ are related to the cubic material susceptibility by

$$D_\parallel = \frac{1}{(1+|c_\chi|^2)^2} \cdot \left(X_{iiii}^{(3)} \cdot \left(1 + |c_\chi|^4\right) + 4X_{iijj}^{(3)} \cdot |c_\chi|^2 + X_{ijji}^{(3)} \cdot \left(c_\chi^2 + c_\chi^{*\,2}\right)\right) \tag{2.105}$$

and

$$\begin{aligned} D_\perp =& \frac{1}{(1+|c_\chi|^2) \cdot (1+|c_\psi|^2)} \\ & \cdot \left(2X_{iiii}^{(3)} + X_{iijj}^{(3)} \cdot \left(|c_\chi|^2 + |c_\psi|^2 - 2\right) + X_{ijji}^{(3)} \cdot \left(c_\chi c_\psi + c_\chi^* c_\psi^*\right)\right). \end{aligned} \tag{2.106}$$

In (2.101) and (2.102), the polarization of the light waves $\overrightarrow{\boldsymbol{e}}$ as well as the fiber's eigenpolarizations $\overrightarrow{\boldsymbol{e}}_\chi$, $\overrightarrow{\boldsymbol{e}}_\psi$ are random variables. To obtain the effective susceptibility, the expectation value of the propagation equations is calculated. It is assumed that all polarizations occur with equal probability; this corresponds to a uniform distribution on the Poincaré sphere surface. Values for $\{\varepsilon, \phi\}$ such that the SOP is uniformly distributed on the Poincaré sphere are obtained from transforming random variables u, v uniformly distributed in $[-1, 1]$ as

$$\varepsilon = \frac{\arccos(u)}{2} - \frac{\pi}{4} \quad \text{and} \quad \phi = \frac{\pi}{2} \cdot v. \tag{2.107}$$

When $\{\varepsilon, \phi\}$ are chosen as in (2.107), the expectation of (2.101) and (2.102) over $\{\varepsilon, \phi\}$ is independent of the fiber's eigenpolarizations:

$$\begin{aligned} \mathcal{E}_{\varepsilon,\phi}\left\{\frac{\partial}{\partial z}\hat{E}(z)\right\} =& -\frac{\alpha}{2} \cdot \hat{E}(z) - \jmath\, \frac{k_0^2}{8\beta(f_0)} \cdot 6 \cdot e^{-\jmath\, \Delta\beta \cdot z} \\ & \cdot \hat{E}^{(n)}(z) \cdot \hat{E}^{(p)}(z) \cdot \hat{E}^{(q)^*}(z) \cdot \underbrace{\frac{2}{3} \cdot \left(D_\parallel + D_\perp\right)}_{X_{\text{eff}}^{(3)}}, \end{aligned} \tag{2.108}$$

$$\mathcal{E}_{\varepsilon,\phi}\left\{\frac{\partial}{\partial z}\hat{E}_\perp(z)\right\} = 0. \tag{2.109}$$

The expectation value of the component orthogonal to the incident field is zero. Therefore, it can be concluded that in a sufficiently long fiber no nonlinear polarization rotation is created and a scalar description of the wave propagation is valid. In such a scalar approach, however, information about the SOP of the propagating light is lost.

When the eigenpolarizations are orthogonal, the sum of the coupling parameters is constant:

$$D_\parallel + D_\perp = X_{iiii}^{(3)} + X_{iijj}^{(3)}. \tag{2.110}$$

It is interesting to remark that this sum is independent of the cubic susceptibility component $X^{(3)}_{ijji}$. Taking (2.27) and (2.30) and the numerical values given in Table 2.1 yields

$$X^{(3)}_{iijj}/X^{(3)}_{iiii} \approx 0.358; \tag{2.111}$$

the effective cubic susceptibility becomes

$$X^{(3)}_{\text{eff}} \approx 0.905 \cdot X^{(3)}_{iiii}. \tag{2.112}$$

Using the approximation $X^{(3)}_{iijj}/X^{(3)}_{iiii} \approx 1/3$ yields

$$X^{(3)}_{\text{eff}} \approx \frac{8}{9} \cdot X^{(3)}_{iiii}, \tag{2.113}$$

a value that is frequently found in the literature [Agr01, MM06].

The above calculation extends to the analysis of any four frequencies $f_0 = f_n + f_p - f_q$, so that the propagation of multiple waves or of a signal with a continuous spectrum in a real fiber can be described in a scalar approach by replacing the nonlinearity parameter ($X^{(3)}_{iiii}$, γ or n_2) with the corresponding effective parameter ($X^{(3)}_{\text{eff}}$, γ_{eff} or $n_{2,\text{eff}}$) obtained by multiplication with the numerical factor given in (2.112) or (2.113).

When the incident light waves are not identically polarized, $X^{(3)}_{\text{eff}}$ can be further reduced below the value given in (2.112) [Han02]. In these cases, however, the orthogonal component does not vanish and a scalar approach is invalid. Assigning different SOP's to different spectral components of a signal as a means of reducing the overall nonlinear impairment remains an interesting open research problem that is not considered in this thesis.

Large frequency spacing

The case of a large frequency spacing between interacting waves, e. g. in a WDM system, requires a separate treatment. Because the fiber's birefringence is frequency-dependent, the assumption that the interacting waves undergo identical polarization changes along the fiber does no longer hold. In this scenario, FWM is effectively suppressed because of the waves' phase mismatch, and the dominating effect is cross-phase modulation (XPM), i. e. the interaction of two waves with $f_0 = f_0 + f_n - f_n$ (cf. Section 2.2.2).

Analoguously to (2.101) and (2.102), Hanik [Han02] derives propagation equations for a wave at f_0 with SOP $\overrightarrow{\boldsymbol{e}}$ interacting with a wave at f_n with SOP $\overrightarrow{\boldsymbol{e}}^{(n)}$ through XPM:

$$\begin{aligned}\frac{\partial}{\partial z}\hat{E}(z) = &-\frac{\alpha}{2}\cdot\hat{E}(z) - \jmath\,\frac{k_0^2}{8\beta(f_0)}\cdot 6\cdot\hat{E}(z)\cdot\left|\hat{E}^{(n)}(z)\right|^2\cdot \\ &\Big(D_{\parallel}\cdot\Big(\left|\overrightarrow{\boldsymbol{e}}^{(n)}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\chi}\right|^2\cdot\left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\chi}\right|^2 + \left|\overrightarrow{\boldsymbol{e}}^{(n)}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\psi}\right|^2\cdot\left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\psi}\right|^2\Big) + \\ &D_{\perp}\cdot\Big(\left|\overrightarrow{\boldsymbol{e}}^{(n)}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\psi}\right|^2\cdot\left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\chi}\right|^2 + \left|\overrightarrow{\boldsymbol{e}}^{(n)}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\chi}\right|^2\cdot\left|\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\psi}\right|^2\Big)\Big),\end{aligned} \tag{2.114}$$

$$\frac{\partial}{\partial z}\hat{E}_{\perp}(z) = -\frac{\alpha}{2}\cdot\hat{E}_{\perp}(z) - \jmath\,\frac{k_0^2}{8\beta(f_0)}\cdot 6\cdot\hat{E}(z)\cdot\left|\hat{E}^{(n)}(z)\right|^2\cdot (D_{\parallel}-D_{\perp})\cdot\left(\left|\overrightarrow{\boldsymbol{e}}^{(n)}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\psi}\right|^2-\left|\overrightarrow{\boldsymbol{e}}^{(n)}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\chi}\right|^2\right)\cdot\left(\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\chi}\right)\cdot\left(\overrightarrow{\boldsymbol{e}}\cdot\overrightarrow{\boldsymbol{e}}^{*}_{\psi}\right). \quad (2.115)$$

As before, the effective cubic susceptibility is obtained by calculating the expectation value of (2.114) and (2.115). In contrast to the case of waves with a small frequency spacing, the expectation is calculated over uniformly distributed SOPs $\overrightarrow{\boldsymbol{e}}$ and $\overrightarrow{\boldsymbol{e}}^{(n)}$ characterized by independent sets of ellipticity and inclination angles (obtained as in (2.107)). The result is independent of the fiber's eigenpolarizations:

$$\mathcal{E}_{\varepsilon,\phi,\varepsilon^{(n)},\phi^{(n)}}\left\{\frac{\partial}{\partial z}\hat{E}(z)\right\} = -\frac{\alpha}{2}\cdot\hat{E}(z) - \jmath\,\frac{k_0^2}{8\beta(f_0)}\cdot 6\cdot\hat{E}(z)\cdot\left|\hat{E}^{(n)}(z)\right|^2\cdot\underbrace{\frac{1}{2}\cdot\left(D_{\parallel}+D_{\perp}\right)}_{X^{(3)}_{\text{eff}}}, \quad (2.116)$$

$$\mathcal{E}_{\varepsilon,\phi,\varepsilon^{(n)},\phi^{(n)}}\left\{\frac{\partial}{\partial z}\hat{E}_{\perp}(z)\right\} = 0. \quad (2.117)$$

Again, the expectation value of the component orthogonal to the incident field is zero, so that a scalar approach is valid.

In principle, the effective cubic susceptibility is frequency-dependent as seen in Figure 2.2. However, for channel spacings at which XPM is a relevant effect (cf. Chapter 6), this frequency dependence of $\Re\{X^{(3)}\}$ can be neglected and one can use (2.110) and (2.111) to write

$$X^{(3)}_{\text{eff}} \approx 0.68\cdot X^{(3)}_{iiii}. \quad (2.118)$$

With the approximation $X^{(3)}_{iijj}/X^{(3)}_{iiii} \approx 1/3$ one obtains

$$X^{(3)}_{\text{eff}} \approx \frac{2}{3}\cdot X^{(3)}_{iiii}. \quad (2.119)$$

Because the SOPs $\overrightarrow{\boldsymbol{e}}$ and $\overrightarrow{\boldsymbol{e}}^{(n)}$ evolve randomly (relative to each other), the effective cubic susceptibility (2.118) is valid for any two input polarizations of the interacting waves.

For Raman and Brillouin scattering, analoguous calculations assuming a random polarization evolution of pump and Stokes waves lead to [Han95, Agr01]

$$G_{\text{R/B,eff}} = \frac{1}{2}\cdot G_{\text{R/B}}. \quad (2.120)$$

Table 2.2 summarizes the factors $X^{(3)}_{\text{eff}}/X^{(3)}_{iiii}$ and $G_{\text{R/B,eff}}/G_{\text{R/B}}$, respectively, used for a scalar description of wave propagation in a randomly birefringent fiber.

Table 2.2: Summary of effective nonlinear parameters for various cases.

Effect	Properties	Factor
SPM, XPM, FWM	small spacing; identical input polarization of interacting waves; identical SOP evolution	$0.905 \approx 8/9$
XPM	large spacing; any input polarization of interacting waves; random relative SOP evolution	$0.68 \approx 2/3$
SRS, SBS	random relative SOP evolution of pump and Stokes wave	1/2

2.1.7 Solving the nonlinear Schroedinger equation

Closed-form solutions to the NLSE (2.79) and (2.80) exist only in the special cases where either attenuation, dispersion or the cubic susceptibility can be neglected. A general analytical solution to the NLSE can be obtained by a Volterra series expansion. Several algorithms exist for the numerical simulation of the wave propagation described by the NLSE [Agr01]; all numerical results in this thesis were obtained by using the *split-step Fourier method*.

Analytical solutions

When dispersion can be neglected, i. e. $\beta_2 = 0$, the NLSE (2.79) can be written as

$$\frac{\partial}{\partial z} \cdot a(z,t) = -\frac{\alpha}{2} \cdot a(z,t) - \jmath \cdot \gamma \cdot |a(z,t)|^2 \cdot a(z,t) \tag{2.121}$$

and has the time-domain solution

$$a(z,t) = a(0,t) \cdot \exp\left(-\frac{\alpha}{2} \cdot z\right) \cdot \exp\left(-\jmath \cdot \gamma \cdot |a(0,t)|^2 \cdot L_{\text{eff}}\right), \tag{2.122}$$

where $L_{\text{eff}} = \left(1 - e^{-\alpha \cdot L}\right)/\alpha$ is the effective fiber length (2.142) introduced in Section 2.2.3. In this case, the signal aquires a power-dependent phase shift along the effective length, whereas the amplitude is not changed.

When the fiber is treated as a linear channel, i. e. $\gamma = 0$, it is useful to write the NLSE (2.80) in the frequency domain as

$$\frac{\partial}{\partial z} A(z,f) = -\frac{\alpha}{2} \cdot A(z,f) - \jmath \cdot \frac{\beta_2}{2} \cdot \omega^2 A(z,f). \tag{2.123}$$

The frequency-domain solution is

$$A(z,f) = A(0,f) \cdot \exp\left(-\left(\frac{\alpha}{2} + \jmath \cdot \frac{\beta_2}{2} \cdot \omega^2\right) \cdot z\right). \tag{2.124}$$

When the fiber attenuation can be neglected, i. e. $\alpha = 0$, the NLSE (2.79) has solutions called *solitons* [Agr02]. In the anomalous dispersion regime, i. e. when β_2 is negative, solitons are hyperbolic-secant-shaped pulses whose envelope either remains unchanged during

propagation (*fundamental soliton*) or changes periodically (*higher-order solitons*). The undistorted propagation of a fundamental soliton can be understood qualitatively by the following consideration. Due to the non-constant group delay caused by chromatic dispersion, different spectral components of a pulse propagate at different group velocities; as a consequence, pulses broaden. At the same time, pulses can be compressed by an interplay of nonlinearities and dispersion. Because of fiber nonlinearities, the increasing power at the leading edge of a pulse causes a red shift, i. e. a frequency decrease, whereas the trailing edge is blue-shifted (cf. (2.122)). When the fiber exhibits anomalous dispersion, the group delay decreases with increasing frequency, so that the blue-shifted trailing edge of the pulse propagates faster than the red-shifted leading edge, resulting in pulse compression. In fundamental soliton pulses, these two effects counteract each other exactly. Because the balance of chromatic dispersion and fiber nonlinearities is disrupted by the fiber attenuation, practical soliton systems require frequent optical amplification [Agr02].

Volterra series expansion

Nonlinear time-invariant causal systems with real-valued input $x(t)$ and output $y(t)$ can be described by the *Volterra series* [Unb98]

$$y(t) = \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) \cdot x(t-\tau_1) \cdots x(t-\tau_n) \, \mathrm{d}\tau_1 \cdots \mathrm{d}\tau_n, \tag{2.125}$$

where h_n is called the n^{th}*-order Volterra kernel*. The Volterra series is named after Italian mathematician Vito Volterra who developed it in 1887. The Volterra series can be regarded as an extension of the Taylor series to systems with memory. In fact, when $h_n = a_n \cdot \delta(\tau_1) \cdots \delta(\tau_n)$ the Volterra series becomes a Taylor series, describing a nonlinear memoryless system. If $h_n \equiv 0$ for all $n \geq 2$, the Volterra series turns into the convolutional integral of input function and impulse response of a linear time-invariant (LTI) system.

When the input function $x(t)$ is a modulated passband signal with carrier frequency f_c, even-order terms ($n = 2, 4, \ldots$) produce out-of-band signals at $f = 0$ and at integer multiples of $2f_c$. In systems where such out-of-band terms do not appear (as in optical communication systems), all even-order Volterra kernels can be set to zero. The Volterra series (2.125) describes the system response to a real-valued input signal. When the signal is represented by a complex equivalent baseband notation, a modified Volterra series can be used which lacks even-order terms [BBC87]. The equivalent baseband Volterra series in the frequency domain is obtained by a Fourier transform. A detailed derivation can be found in [Fel03]; the result can be written as

$$\begin{aligned} A(z,f) =& H_1(z,f)\cdot A(f)+ \\ & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_3(z,f_n,f_q,f)\cdot A(f_n)A^*(f_q)A(f-f_n+f_q)\,\mathrm{d}f_n\mathrm{d}f_q+ \\ & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_5(z,f_n,f_p,f_q,f_r,f)\cdot A(f_n)A^*(f_p)A(f_q)\cdot \\ & \qquad A^*(f_r)A(f-f_n+f_p-f_q+f_r)\,\mathrm{d}f_n\mathrm{d}f_p\mathrm{d}f_q\mathrm{d}f_r+\ldots, \end{aligned} \tag{2.126}$$

where $A(f)=A(z=0,f)$.

The solution of the NLSE (2.80) with a Volterra series has first been described by Peddanarappagari and Brandt-Pearce [PBP97]. They use the generic NLSE formulation

$$\begin{aligned} \frac{\partial}{\partial z}A(z,f) =& G_1(f)\cdot A(z,f)+ \\ & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_3(f_n,f_q,f)\cdot A(z,f_n)A^*(z,f_q)A(z,f-f_n+f_q)\,\mathrm{d}f_n\mathrm{d}f_q, \end{aligned} \tag{2.127}$$

where G_1 and G_3 are the *linear dispersion kernel* and the *fiber nonlinearity kernel*, respectively. By comparison with (2.80), these are

$$G_1(f) = -\frac{\alpha}{2} - \jmath\cdot\frac{\beta_2}{2}\cdot\omega^2 \tag{2.128}$$

and

$$G_3 = -\jmath\cdot\gamma. \tag{2.129}$$

When needed, β_1 or higher-order terms of the expansion (2.66) can be included in G_1. Likewise, G_3 can be used to model higher-order nonlinear effects such as self-steepening or Raman scattering that arise from the frequency-dependence of $X^{(3)}$ [PBP97].

To solve the NLSE, (2.126) is inserted into (2.127). By ordering and comparison of coefficients with an identical number of integrals, linear differential equations for the Volterra kernels are obtained. Using the boundary conditions $H_1(z=0,f)=1$ and $H_n(z=0,f)=0, n>1$, unique solutions to these differential equations are obtained:

$$H_1(z,f) = e^{G_1(f)\cdot z} = \exp\left(-\left(\frac{\alpha}{2}+\jmath\cdot\frac{\beta_2}{2}\cdot\omega^2\right)\cdot z\right), \tag{2.130}$$

$$\begin{aligned} H_3(z,f_n,f_q,f) =& G_3(f_n,f_q,f)\cdot\frac{e^{\left(G_1(f_n)+G_1^*(f_q)+G_1(f-f_n+f_q)\right)z}-e^{G_1(f)z}}{G_1(f_n)+G_1^*(f_q)+G_1(f-f_n+f_q)-G_1(f)} \\ =& \jmath\cdot\gamma\cdot\exp\left(-\left(\frac{\alpha}{2}+\jmath\frac{\beta_2}{2}\omega^2\right)z\right)\cdot\frac{e^{-(\alpha+\jmath\beta_2(\omega_n-\omega)(\omega_n-\omega_q))z}-1}{\alpha+\jmath\beta_2(\omega_n-\omega)(\omega_n-\omega_q)}. \end{aligned} \tag{2.131}$$

The (rather complex) solution for H_5 is given in [PBP97]; kernels of higher order have not been reported. Even-order kernels are zero because of the pass-band characteristic of the optical channel on the one hand (see above), and because of the absence of even-order fiber nonlinearities on the other. When the square-law characteristic of the photodiode in a direct-detection receiver is included in the Volterra series, only even-order kernels remain.

In the analysis of fiber links where α and β_2 are not constant, e. g. when inline dispersion compensation and optical amplification are considered, H_3 can be defined by an integral over the link length [Lou06].

The Volterra series solution, usually including only the first and third kernel, is used in many analytical studies of fiber-optic communication systems. Its accuracy depends on how nonlinear the system behaves. The error incurred by neglecting terms of order n and higher is proportional to $P_{\max}^{n/2}$, where

$$P_{\max} = \max_t |a(0,t)|^2 \tag{2.132}$$

is the peak power of the optical field [PBP97]; when the only nonlinear kernel taken into account is H_3, the error is proportional to $P_{\max}^{5/2}$. Feldhaus derives an upper limit of 10 mW for the input power below which the error of the Volterra series truncated after the third-order term remains small [Fel03]. This value is supported by numerical simulations comparing the Volterra series with the split-step Fourier method [Fel03, XBP03]. Therefore, the (cubic) Volterra series is not of much practical use for systems with high input powers or with many WDM channels.

Split-step Fourier method

Although the Volterra series approach can be used for a numerical evaluation of the NLSE and several finite-difference methods have been proposed [Agr01], the algorithm most commonly used for solving the NLSE is the *symmetrized split-step Fourier* (SSF) method.

The idea of the SSF algorithm is to subdivide the fiber into differential segments of length $\mathrm{d}z$. In each segment, the linear and nonlinear fiber effects are assumed to act independently, so that they can be evaluated using (2.122) and (2.124). To improve the accuracy, the nonlinear operation is placed in the middle between two linear operations each acting on a fiber length $\mathrm{d}z/2$ (symmetrized SSF). The solution for a differential fiber segment of length $\mathrm{d}z$ can then be written as

$$a(z+\mathrm{d}z,t) \approx \mathcal{F}^{-1}\Big(\mathcal{F}\Big(\mathcal{F}^{-1}\left(\mathcal{F}\left(a(z,t)\right)\cdot e^{-\left(\frac{\alpha}{2}+\jmath\cdot\frac{\beta_2}{2}\cdot\omega^2\right)\frac{\mathrm{d}z}{2}}\right)\cdot e^{-\jmath\cdot\gamma\cdot\left(|a(z,t)|^2+|a(z+\mathrm{d}z,t)|^2\right)\cdot\frac{\mathrm{d}z}{2}}\Big)\cdot e^{-\left(\frac{\alpha}{2}+\jmath\cdot\frac{\beta_2}{2}\cdot\omega^2\right)\cdot\frac{\mathrm{d}z}{2}}\Big), \tag{2.133}$$

where the optical power determining the nonlinear phase shift stems from the trapezoidal approximation

$$\int_z^{z+\mathrm{d}z} |a(\zeta,t)|^2 \mathrm{d}\zeta \approx \frac{\mathrm{d}z}{2}\cdot\left(|a(z,t)|^2+|a(z+\mathrm{d}z,t)|^2\right). \tag{2.134}$$

Obviously, this term is difficult to evaluate as $|a(z+\mathrm{d}z,t)|^2$ is unknown at first. Therefore, $|a(z+\mathrm{d}z,t)|^2$ is set to $|a(z,t)|^2$ at first and the calculation is iteratively repeated until the result converges (cf. Algorithm 2.1), which is usually the case after one or two iterations.

The choice of the step size $\mathrm{d}z$ presents a trade-off between accuracy and computational complexity. It is common to use a variable step size that increases as the optical power is attenuated during propagation. Various criteria for determining the step size have been proposed [SHZM03]. A common method to obtain $\mathrm{d}z$ is to use to the maximum *nonlinear phase shift*

$$\phi_{\mathrm{NL,max}} = \max_t \gamma \cdot |a(z,t)|^2 \cdot \mathrm{d}z = \gamma P_{\max} \mathrm{d}z \tag{2.135}$$

so that $\mathrm{d}z = \phi_{\mathrm{NL,max}}/(\gamma P_{\max})$. Commonly used values for $\phi_{\mathrm{NL,max}}$ range from 10^{-3} to 10^{-2} to keep the global error small [SHZM03].

Algorithm 2.1 summarizes the implementation of the symmetrized split-step Fourier method.

Algorithm 2.1 Symmetrized split-step Fourier method.

$\varepsilon = 10^{-12}$, $\phi_{\mathrm{NL,max}} = 0.01$, $z = 0$
while $z < L_{\mathrm{fiber}}$ **do**
  $P_{\max} = \max_t |a(z,t)|^2$, $\mathrm{d}z = \phi_{\mathrm{NL,max}}/(\gamma P_{\max})$, $z = z + \mathrm{d}z$
  if $z > L_{\mathrm{fiber}}$ **then**
    $\mathrm{d}z = L_{\mathrm{fiber}} - (z - \mathrm{d}z)$, $z = L_{\mathrm{fiber}}$
  end if
  $a_1' = \mathcal{F}^{-1}\left(\mathcal{F}(a) \cdot e^{-\left(\frac{\alpha}{2} + \jmath \cdot \frac{\beta_2}{2} \cdot \omega^2\right) \cdot \frac{\mathrm{d}z}{2}}\right)$
  $a_1 = a_1' \cdot e^{-\jmath \cdot \gamma \cdot |a_1'|^2 \cdot \mathrm{d}z}$
  $a_1 = \mathcal{F}^{-1}\left(\mathcal{F}(a_1) \cdot e^{-\left(\frac{\alpha}{2} + \jmath \cdot \frac{\beta_2}{2} \cdot \omega^2\right) \cdot \frac{\mathrm{d}z}{2}}\right)$
  $\mathrm{err} = 1$
  while $\mathrm{err} > \varepsilon$ **do**
    $a_2 = a_1' \cdot e^{-\jmath \cdot \gamma \cdot \left(|a|^2 + |a_1|^2\right) \cdot \frac{\mathrm{d}z}{2}}$
    $a_2 = \mathcal{F}^{-1}\left(\mathcal{F}(a_2) \cdot e^{-\left(\frac{\alpha}{2} + \jmath \cdot \frac{\beta_2}{2} \cdot \omega^2\right) \cdot \frac{\mathrm{d}z}{2}}\right)$
    $\mathrm{err} = \max_t \left(\frac{|a_1 - a_2|^2}{|a_1|^2}\right)$
    $a_1 = a_2$
  end while
  $a(z,t) = a_1$
end while

2.2 Optical fibers

2.2.1 Parameters

The linear and nonlinear propagation properties of fibers depend on the concentration of dopants in the material, the refractive index profile, the effective area and, of course, the

wavelength of the propagating field. The various types of single-mode fibers are typically characterized by measureable parameters.

For a specific wavelength λ_c, the fiber's chromatic dispersion is specified by the *dispersion parameter* D and the *slope parameter* S which are related to β_2 and β_3 (cf. (2.67)) through

$$D = \frac{\mathrm{d}t_g(\lambda)}{\mathrm{d}\lambda}\bigg|_{\lambda=\lambda_c} = -2\pi\frac{c_0}{\lambda_c^2}\cdot\beta_2 \tag{2.136}$$

and

$$S = \frac{\mathrm{d}^2t_g(\lambda)}{\mathrm{d}\lambda^2}\bigg|_{\lambda=\lambda_c} = 4\pi\frac{c_0}{\lambda_c^3}\cdot\beta_2 - \left(\frac{2\pi c_0}{\lambda_c^2}\right)^2\cdot\beta_3. \tag{2.137}$$

Standard single-mode fibers (SSMF) have a dispersion value of around 17 ps / nm / km at their attenuation minimum at $\lambda = 1.55\,\mu\mathrm{m}$. Their zero-dispersion wavelength is approximately 1.3 µm. *Dispersion-shifted fibers* (DSF) have their zero-dispersion wavelength at 1.55 µm. Because waves interacting through FWM are fully phase-matched in DSFs (see below), this fiber type is not used in modern lightwave systems and is replaced by *non-zero dispersion-shifted fibers* (NZDSF) which have a dispersion value larger than zero but smaller than SSMFs. *Dispersion-compensating fibers* (DCF) exhibit normal dispersion (i. e. $\beta_2 > 0 \Leftrightarrow D < 0$) at 1.55 µm. These fibers are used for optical dispersion compensation. Figure 2.5 depicts the attenuation and dispersion curves for these typical fiber types.

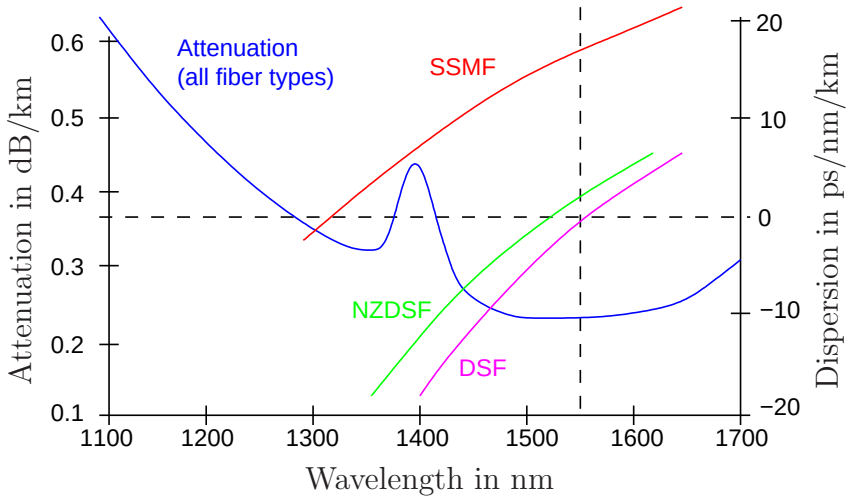

Figure 2.5: Attenuation and dispersion curves of various single-mode fibers (after [KLW08a]).

When the fiber is birefringent, different modes have different group velocities, leading to a *differential group delay* (DGD) between two polarizations. Due to the stochastic character of PMD, the DGD increases linearly with the square root of the distance (see Chapter 4). The *PMD parameter* (in units of $\mathrm{ps}/\sqrt{\mathrm{km}}$) of a fiber characterizes this feature. New fibers can have PMD parameters smaller than $0.1\,\mathrm{ps}/\sqrt{\mathrm{km}}$ [VP02].

Fiber nonlinearities are specified by the nonlinear index coefficient n_2 and the effective fiber core area A_{eff}. These parameters are combined in the nonlinear coefficient γ (cf. (2.72)). Taking the cubic susceptibility value in Table 2.1, $n_{\text{core}} = 1.445$, $A_{\text{eff}} = 80(\mu\text{m})^2$ and $\lambda_c = 1.55\,\mu\text{m}$ yields $n_2 = 2.7 \cdot 10^{-20}\,\text{m}^2\,/\,\text{W}$ and $\gamma \approx 1.37\,(\text{W km})^{-1}$. It must be noted that n_2 varies considerably among different fibers as a consequence of the doping of the silica core with other materials such as GeO_2; measured values for n_2 lie in the range $2.2 \ldots 3.9 \cdot 10^{-20}\,\text{m}^2\,/\,\text{W}$ [Agr01]. Additionally, the value of n_2 differs depending on which nonlinear effect is used for measuring it. As the light polarization changes randomly along the fiber, measurements yield an effective value $n_{2,\text{eff}}$ that is reduced by a factor of 8/9 (cf. Section 2.1.6) compared with the value expected for polarization-maintaining fibers. Agrawal recommends using a value of $n_{2,\text{eff}} = 2.6 \cdot 10^{-20}\,\text{m}^2\,/\,\text{W}$ for pulses wider than 1 ns [Agr01]; this corresponds to $n_2 = 2.9 \cdot 10^{-20}\,\text{m}^2\,/\,\text{W}$.

For the numerical simulations in this thesis, the fiber parameters specified in Table 2.3, taken from [DBDB02, VP02], have been used unless stated otherwise. The nonlinear parameters γ and n_2 of Table 2.3 need to be weighted with the appropriate factors given in Table 2.2 to account for random polarization changes.

Table 2.3: Typical parameters of different fiber types at $\lambda = 1.55\,\mu\text{m}$.

Type	α in $\frac{\text{dB}}{\text{km}}$	D in $\frac{\text{ps}}{\text{nm km}}$	S in $\frac{\text{ps}}{\text{nm}^2\,\text{km}}$	A_{eff} in $(\mu\text{m})^2$	n_2 in $\frac{\text{m}^2}{\text{W}}$	γ in $\frac{1}{\text{W km}}$
SSMF	0.2	17	0.057	80	$2.9 \cdot 10^{-20}$	1.47
DSF	0.2	0.0	0.07	50	$2.9 \cdot 10^{-20}$	2.35
NZDSF+	0.2	2.8	0.07	53	$2.9 \cdot 10^{-20}$	2.22
NZDSF-	0.2	−2.3	0.07	57	$2.9 \cdot 10^{-20}$	2.06
DCF	0.5	−80	−0.15	20	$2.9 \cdot 10^{-20}$	5.88

2.2.2 Fiber effects

As detailed in Section 2.1, light propagation in optical fibers is affected by the linear and cubic material susceptibility. None of the resulting impairments is genuinely stochastic or irrevertible, so that any information-theoretic analysis of the fiber-optic communication channel must include a discussion about which impairments to consider quasi-stochastic. This discussion requires a clear distinction between the different effects.

Linear fiber effects comprise attenuation, chromatic dispersion and polarization mode dispersion. Although the attenuation in optical fibers is small and has a very broad minimum, transmission over long distances requires optical signal amplification. Amplified spontaneous emission (ASE) is the main source of noise in optically amplified transmission systems (cf. Section 2.3).[3] While all other effects can – in principle – be inverted, noise

[3]Rayleigh scattering is not only the cause of attenuation, but also a source of interference through

is a stochastic impairment. Therefore, the channel capacity of fiber-optic communication channels is ultimately limited by attenuation.

Chromatic dispersion requires sophisticated electrical or optical equalization in practical systems, but poses no theoretic limit to the transmission of data. However, because of the interaction with nonlinear fiber effects, dispersion must be considered in the analysis of nonlinear impairments. In addition, optical dispersion compensation using DCFs suffers from the additional attenuation and nonlinearity of the DCFs.

Polarization mode dispersion is a random effect although when the PMD properties of a fiber change slowly with time, they can be measured and PMD equalization can be employed [Agr02]. The information-theoretic implications of PMD are treated in Chapter 4. As chromatic dispersion, PMD interacts with fiber nonlinearities; this interaction is briefly discussed in Section 4.4.

Nonlinear effects, which have their common origin in the cubic material susceptibility, can be further divided into Raman scattering and the Kerr effect, i. e. the power dependence of the refractive index, which leads to the interaction of four waves whose frequencies are related by $f_0 = f_n + f_p - f_q$ as stated in (2.60).

- When $f_0 \neq f_n \neq f_p \neq f_q$, the interaction is termed *four-wave mixing* (FWM). The case when $f_n = f_p$ so that $f_0 = 2f_n - f_q$ is called *degenerate FWM*. The phase mismatch (2.61) of FWM depends on the dispersion difference between the interacting waves; when D is large or the waves are spaced far apart, the exponential term in (2.60) oscillates rapidly; in this case, the nonlinear contribution cannot build up efficiently. In contrast, when D is small (or even zero as in DSFs), the nonlinear contribution builds up coherently.

- *cross-phase modulation* (XPM) denotes the case where $f_p = f_q$ which implies $f_n = f_0$. For XPM, the phase mismatch (2.61) is always $\Delta\beta = 0$, and (2.60) becomes

$$\frac{\partial}{\partial z}\hat{E}(z) = -\frac{\alpha}{2}\cdot\hat{E}(z) - \jmath\,\frac{k_0^2}{8\beta(f_0)}\cdot 6\cdot X_{\text{eff}}^{(3)}\cdot\hat{E}(z)\cdot\sum_{p|f_p\neq f_0}\cdot\left|\hat{E}^{(p)}(z)\right|^2. \tag{2.138}$$

  Hence, XPM causes a nonlinear phase rotation of the wave at f_0 which is proportional to the power of the waves co-propagating at other frequencies. The value of $X_{\text{eff}}^{(3)}$ depends on the frequency spacing between the interacting waves and ranges between the values given in Table 2.2. Because XPM is always phase matched, the effect does not depend on the frequency spacing between the interacting waves.

- Finally, there is the nonlinear interaction of the wave at $f_0 = f_0 + f_0 - f_0$ with itself; this effect is called *self-phase modulation* (SPM). The propagation equation for this case becomes

$$\frac{\partial}{\partial z}\hat{E}(z) = -\frac{\alpha}{2}\cdot\hat{E}(z) - \jmath\,\frac{k_0^2}{8\beta(f_0)}\cdot 3\cdot X_{\text{eff}}^{(3)}\cdot\hat{E}(z)\cdot\left|\hat{E}(z)\right|^2, \tag{2.139}$$

double Rayleigh scattering. As this effect can be limited by inserting optical isolators along the link, it is not considered as a fundamental impairment in this thesis [EKW+10].

so SPM describes a nonlinear phase shift of a wave which is proportional to the wave's own power. As XPM, SPM is always phase matched.

In the context of WDM systems, the terms SPM, XPM and FWM are used somewhat differently than explained above. To avoid confusion, we introduce the notation SPM$^{\mathcal{S}}$, XPM$^{\mathcal{S}}$ and FWM$^{\mathcal{S}}$ to denote WDM effects.[4] SPM$^{\mathcal{S}}$ refers to the temporally instantaneous nonlinear phase shift by $\gamma P(z,t)$ in the NLSE (2.79). When the term SPM$^{\mathcal{S}}$ is used in this meaning, the effect includes SPM, XPM and FWM as defined above. This "group" of nonlinear effects that are all covered by the NLSE (2.79) are sometimes referred to as *intra-channel* nonlinearities. In contrast, the *inter-channel* effects XPM$^{\mathcal{S}}$ and FWM$^{\mathcal{S}}$ describe the nonlinear phase rotation in one WDM channel induced by the power present in another and the mixing of the signals in four WDM channels, respectively. SRS also belongs to this group of inter-channel effects. In Section 6.2.1, coupled differential equations governing the simultaneous propagation of WDM channels will be derived, and the effects SPM$^{\mathcal{S}}$, XPM$^{\mathcal{S}}$, FWM$^{\mathcal{S}}$ and SRS will be discussed in this context.

Occasionally, the terms *intra-channel XPM* (iXPM) and *intra-channel FWM* (iFWM) can be found in the literature [EMR99]. They desribe the nonlinear interaction of overlapping pulses in the time domain through XPM and FWM, respectively [Sch04]. These terms are not used in this thesis; the effects they describe are of course included in the above description of fiber nonlinearities.

2.2.3 Length scales

In the analysis of the various effects that influence the propagation of light in optical fibers, it is very helpful to define charateristic length scales to quantify the range in which certain impairments are effective. Several length-scale arguments are implied in the derivation of the NLSE; a rigorous discussion of optical fiber transmission based on length scales can be found in [Men99].

A fiber segment of length L has an *effective fiber length*

$$L_{\text{eff}} = \frac{\int_0^L P(z)\mathrm{d}z}{P(0)}. \tag{2.140}$$

The general definition of (2.140) is important in the context of distributed amplification, cf. Section 2.3.2. For a passive fiber with

$$P(z) = P(0) \cdot e^{-\alpha \cdot z}, \tag{2.141}$$

(2.140) is reduced to

$$L_{\text{eff}} = \frac{1 - e^{-\alpha \cdot L}}{\alpha} \tag{2.142}$$

and specifies the distance at which the initial light power is attenuated by a factor of e. L_{eff} is a measure of the fiber length along which nonlinearities are effective (cf. e. g. (2.122)

[4] The letter $^{\mathcal{S}}$ is used to refer to a WDM system effect.

and (2.95)). For a SSMF with $\alpha = 0.2\,\mathrm{dB\,km^{-1}} = 0.02 \cdot (\log_{10} e)^{-1}\,\mathrm{km^{-1}} \approx 0.046\,\mathrm{km^{-1}}$, the effective fiber length tends towards $L_{\mathrm{eff}} = \alpha^{-1} \approx 21.7\,\mathrm{km}$ for long L. In optically amplified multi-span systems, the total effective length is obtained as the sum of the spans' individual effective lengths. In a multispan system with N_{spans} identical spans of length L_{span} and lumped amplification, the effective length of the entire link is [Sch04]

$$L_{\mathrm{eff,total}} = N_{\mathrm{spans}} \cdot \frac{1 - e^{-\alpha \cdot L_{\mathrm{span}}}}{\alpha}. \tag{2.143}$$

Two length scales that can give information about whether nonlinearities or dispersion are dominant effects in a particular system are the *dispersion length*

$$L_{\mathrm{D}} = \frac{1}{|\beta_2| \cdot B^2} \tag{2.144}$$

and the *nonlinear length*

$$L_{\mathrm{NL}} = \frac{1}{\gamma \overline{P}}. \tag{2.145}$$

In (2.144) and (2.145), B is the (arbitrarily defined) bandwith of the signal $a(z,t)$ and $\overline{P} = \mathcal{E}_t\left\{|a(0,t)|^2\right\}$ is the average signal power. The NLSE can be expressed as a function of L_{D} and L_{NL}. To do so, it is useful to introduce a dimensionless unattenuated signal by substituting

$$a(z,t) = u(z,t) \cdot e^{-\frac{\alpha}{2} z} \cdot \sqrt{\overline{P}}, \tag{2.146}$$

and a dimensionless time

$$\tau = t \cdot B, \tag{2.147}$$

where t refers to the reference time defined in (2.76) and its prime sign has been dropped to improve readability. Using (2.146) and (2.147), the NLSE (2.79) can be written as

$$\frac{\partial}{\partial z} u(z,\tau) = \jmath \cdot \frac{\operatorname{sgn}(\beta_2)}{2L_{\mathrm{D}}} \frac{\partial^2 u(z,\tau)}{\partial \tau^2} - \jmath \cdot \frac{e^{-\alpha z}}{L_{\mathrm{NL}}} \cdot |u(z,\tau)|^2 \cdot u(z,\tau). \tag{2.148}$$

The impact of fiber nonlinearities and dispersion can be assessed using this equation. The dispersion term, i. e. the first term on the right-hand side of (2.148), vanishes for long dispersion lengths; as seen from (2.144), this is the case when either the fiber's dispersion or the signal's bandwidth is small. The nonlinear term, i. e. the second term on the right-hand side, vanishes for long nonlinear lengths; from (2.145), this occurs when the nonlinear parameter γ is small or when the average signal power is low. Generally, four propagation regimes can be distinguished for a fiber of length L [Agr01, Sch04]:

- ▷ $L \ll L_{\mathrm{D}}$ and $L \ll L_{\mathrm{NL}}$: Neither dispersion nor nonlinearities act significantly on the signal.
- ▷ $L \ll L_{\mathrm{D}}$ and $L \gtrsim L_{\mathrm{NL}}$: Nonlinear effects dominate over chromatic dispersion. In good approximation, the fiber channel can be described by (2.121).

- ▷ $L \gtrsim L_{\mathrm{D}}$ and $L \ll L_{\mathrm{NL}}$: The propagation is dominated by chromatic dispersion. The fiber can be considered a linear channel as in (2.123).
- ▷ $L \gtrsim L_{\mathrm{D}}$ and $L \gtrsim L_{\mathrm{NL}}$: In this case, nonlinearities and dispersion act together and their interaction must be considered.

The length scale of the efficiency of FWM is given by the *coherence length* [Sch04]

$$L_{\mathrm{coh}} = \frac{\pi}{|\Delta\beta|} \tag{2.149}$$

which determines the oscillation length of the exponential term in (2.60). When the phase mismatch of the waves interacting through FWM is large, L_{coh} is small and the rapid oscillation averages out the FWM product of the slowly varying amplitudes. In contrast, when L_{coh} is large, the FWM products add up coherently to yield a significant distortion. In fused silica at $\lambda = 1.55\,\mu\mathrm{m}$, $L_{\mathrm{coh}} = 2.8\,\mathrm{m}$ for a channel spacing of $\Delta f = 1000\,\mathrm{GHz}$, $L_{\mathrm{coh}} = 286\,\mathrm{m}$ for $\Delta f = 100\,\mathrm{GHz}$, $L_{\mathrm{coh}} = 1144\,\mathrm{m}$ for $\Delta f = 50\,\mathrm{GHz}$ and $L_{\mathrm{coh}} = 2864\,\mathrm{km}$ for $\Delta f = 1\,\mathrm{GHz}$ [Sch04].

The effect of XPM$^{\mathcal{S}}$ between two WDM channels at λ_0 and λ_p is affected by the *walk-off length*

$$L_{\mathrm{W}} = \frac{1}{B \cdot |t_g(\lambda_0) - t_g(\lambda_p)|}, \tag{2.150}$$

where $t_g(\lambda)$ denotes the wavelength-dependent group delay and B is the signals' bandwidth. L_{W} is large when B is small or when the channels are closely spaced. In this case, a temporal section of one WDM channel (e. g. a pulse of width $1/B$) acts on a temporal section of another WDM for a significant length of the fiber; hence, XPM$^{\mathcal{S}}$ causes signal-dependent distortions. In contrast, when L_{W} is small, the signals' group delay difference is large, so that the XPM$^{\mathcal{S}}$-induced nonlinear phase rotation is averaged and becomes proportional to the average power in the modulating channel.

It is more common in the literature to relate L_{D} and L_{W} to the pulse width instead of the signal bandwidth and L_{NL} to the pulse power. The above expressions are of course equivalent, but have been defined with the intention of greater generality, as signals do not necessarily consist of pulse sequences but can be arbitrary band-limited waveforms in general.

A length scale that characterizes a constantly birefringent fiber is the *polarization beat length*

$$L_{\mathrm{B}} = \frac{\lambda}{\Delta n} = \frac{\lambda}{|n_x - n_y|} = \frac{2\pi}{\Delta\beta} = \frac{2\pi}{|\beta_x - \beta_y|}. \tag{2.151}$$

As mentioned in Section 2.1.6, the SOP of a light wave evolves periodically in such a fiber over L_{B}. In (2.151), $\Delta\beta$ is the difference between the propagation constants of the two orthogonal fundamental fiber modes and Δn is the corresponding refractive index difference; they are related through $\beta = 2\pi f n/c_0$. The beat lengths of standard fibers are typically $\sim 1 \ldots 10\,\mathrm{m}$, giving $\Delta n \sim 10^{-7}$ [GM05]. PMFs have beat lengths of a few mm, whereas *spun fibers* can have beat lengths of more than 100 m [VP02].

A statistic measure that characterizes long fibers which are subject to random birefringence and mode coupling is the *correlation length* L_{corr}. It is defined as the length after which a fraction of e^{-2} of the light power has been transferred from the initial SOP to the orthogonal SOP on average [GM05]. The correlation length is useful to determine whether a deterministic approach (assuming constant birefringence; $L \ll L_{\text{corr}}$) or a stochastic approach (assuming random distortions; $L \gg L_{\text{corr}}$) is appropriate in the modeling of PMD (cf. Section 4.1). The correlation length varies between a few meters and about 1 km; it depends strongly on the way the fiber is deployed [GM05].

2.3 Optical amplification

Despite the low attenuation in optical fibers, long-haul lightwave systems such as links bridging transoceanic distances cannot be operated without periodic signal amplification. For that reason, optical long-haul systems required expensive electrical regeneration until the 1990s, when *erbium-doped fiber amplifiers* (EDFA) became available and amplification could be achieved without optical-electrical conversion [KL02b, Ch. 5].

Today, EDFAs are widely deployed in optical communication systems. Raman amplifiers play a vital role in theoretical and experimental reseach, but are not yet commercially successful [Bro04]. A third type of optical amplifiers are *semiconductor optical amplifiers* (SOAs) [Agr02].

2.3.1 Erbium-doped fiber amplifiers

All optical amplifiers are based on the same principle. The material is excited by a pump wave, and incoming light is coherently amplified by stimulated emission. An EDFA consists of a fiber that is doped with erbium ions (Er^{3+}); other rare earth elements such as holmium, neodymium, samarium, thulium and ytterbium can be used to build amplifiers that operate at different wavelengths [Agr02]. As depicted in Figure 2.6, the erbium ions exhibit energy levels that allow them to absorb light at 980 nm and 1480 nm. Ions

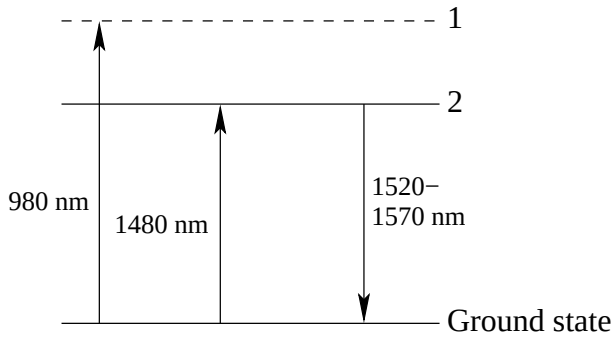

Figure 2.6: Energy levels of erbium ions.

that are excited to the higher energy level (1) decay nonradiatively to the metastable first excited state (2) within $\tau_1 \approx 1\,\mu\text{s}$. The time constant of this metastable state

is $\tau_2 = 10\,\text{ms}$ [KK97b, Ch. 2]. Light that is coupled into the erbium-doped fiber (in which a sufficiently large population of erbium ions must be excited) is amplified through *stimulated emission*: incident photons stimulate the excited ions to return to the ground state upon emission of a photon of identical frequency, phase and polarization. However, ions also return to their ground state spontaneously, thereby emitting photons of random phase and polarization; this *spontaneous emission* becomes amplified upon propagation along the doped fiber, a process that is known as *amplified spontaneous emission* (ASE).

ASE is the main source of noise in optically amplified lightwave communication systems. The power spectral density (PSD) of ASE noise can in good approximation be modeled as *additive white Gaussian noise* (AWGN) [Agr02, GLW63, GWL63]. The one-sided PSD per polarization is given by

$$N_{\text{ASE}} = h \cdot f_S \cdot (G-1) \cdot n_{\text{sp}}, \tag{2.152}$$

where G is the amplifier's gain, h is Planck's constant and f_S is the signal frequency. For typical signal bandwidths of optical communication systems, the frequency dependence of N_{ASE} can be neglected. The parameter n_{sp} is the *spontaneous emission factor*; it is related to the erbium ion population in the ground state N_{ground} and in the higher states N_{excited} by [Agr02]

$$n_{\text{sp}} = \frac{N_{\text{excited}}}{N_{\text{excited}} - N_{\text{ground}}}. \tag{2.153}$$

When total population inversion is achieved, $n_{\text{sp}} = 1$.

The spontaneous emission factor can not be directly observed, so the optical amplifier is characterized by its *noise figure* F_n. This parameter is defined as the ratio of the *electrical* signal-to-noise ratio (SNR) before the optical amplifier to the electrical SNR after the amplifier. A calculation assuming an ideal receiver delivers [Agr04]

$$F_n = 2 \cdot n_{\text{sp}} \cdot \frac{G-1}{G} + \frac{1}{G}. \tag{2.154}$$

In obtaining (2.154), a photodiode with maximum quantum efficiency was assumed and thermal noise was neglected. The noise figure is lower bounded by $F_n = 2$; this value is obtained when $G \gg 1$ and $n_{\text{sp}} = 1$, i. e. for total population inversion. Hence, the (electrical) SNR is reduced by 3 dB even for an ideal amplifier. The reason for this reduction lies in the shot noise process generated during photodetection. In practice, EDFAs achieve noise figures in the range $4 \ldots 8\,\text{dB}$ [Agr02].

Inserting (2.154) into (2.152) yields

$$N_{\text{ASE}} = \frac{1}{2} \cdot h \cdot f_S \cdot G \cdot F_n - \frac{h \cdot f_S}{2} \approx \frac{1}{2} \cdot h \cdot f_S \cdot G \cdot F_n, \tag{2.155}$$

where the second (approximate) equality holds for $G \gg 1$.

In multispan systems, the total ASE noise power is the sum of the ASE noise powers produced by each amplifier along the link. When all N_{spans} spans have identical fiber lengths

L_{span} and identical gains $G = e^{\alpha \cdot L_{\text{span}}}$, the one-sided ASE noise PSD per polarization becomes

$$N_{\text{ASE,total}} = \frac{1}{2} \cdot h \cdot f_S \cdot G \cdot F_n \cdot N_{\text{spans}} = \frac{1}{2} \cdot h \cdot f_S \cdot G \cdot F_n \cdot \frac{\alpha \cdot L_{\text{total}}}{\ln G}. \tag{2.156}$$

The optimum amplifier spacing (in the sense of minimum ASE noise) for such systems is found by setting $\partial N_{\text{ASE,total}}/\partial G = 0$ which yields $G = e$ or $L_{\text{span,opt}} = 1/\alpha$.

The gain of optical amplifiers depends on the signal and pump powers. When the signal power is low and the pump maintains population inversion, the observed gain is known as the *small signal gain*. When the input increases and G is reduced to half the small signal gain (i. e. reduced by 3 dB), the amplifier reaches *saturation*. Because of the large time constant τ_2 of the metastable energy level, the gain dynamics of EDFAs is slow even when the amplifier is in saturation. Therefore, the gain of an EDFA responds to the average incident power and is thus transparent to variable data rates and modulation formats [KK97b, Ch. 2]. Another important advantage of EDFAs is that their gain is polarization independent [KK97b, Ch. 2].

2.3.2 Distributed amplification

EDFAs provide *lumped amplification*, meaning that the attenuation that accumulates along the transmission fiber is compensated in an EDFA module containing a few (10. . . 20) meters of highly doped fiber. Although *distributed amplification* can be achieved by (lightly) doping the transmission fiber itself, such distributed EDFAs are not used commercially [Agr02]. In contrast, stimulated Raman scattering (SRS, cf. Section 2.1.4) is widely (although not commercially) used for distributed amplification. For Raman amplification, a high-power pump wave is coupled into the transmission fiber; this can be done in forward, backward or both directions. The silica material absorbs the pump power and the signal wave is amplified through stimulated emission during propagation along the fiber. The gain of Raman amplification depends on the frequency spacing between pump and Stokes (signal) wave and is depicted in Figure 2.4. The evolution of signal and pump wave is governed by (2.93) and (2.94), where the sign in (2.94) is determined by the propagation direction of the pump wave.

In contrast to EDFAs, Raman amplifiers are polarization sensitive. The reason lies in the polarization dependence of the cubic material susceptibility given by (2.23)-(2.26). The gain that is observed when the pump's polarization is orthogonal to the signal's is one magnitude below that for co-polarization [Bro04]. However, when the fiber length along which amplification through SRS occurs is sufficiently long (as is the case in practice), the polarization randomization discussed in Section 2.1.6 provides an effective gain (2.120) for both polarizations. This averaging effect is even more effective for counter-propagating signal and pump. Other means to reduce polarization-dependent gain (PDG) include the use of two orthogonally polarized pumps. Hence, PDG is not an issue in Raman-amplified systems [Bro04].

When the signal wave is amplified by a single co- or counter-propagating pump wave, the evolution of the signal intensity is governed by the differential equation (2.93). Neglecting

pump depletion, i. e. setting $G_R = 0$ in (2.94), (2.93) can be easily solved. The result, expressed through the signal power $P^{(S)} = I^{(S)} A_{\text{eff}}$, is [Agr04]

$$P^{(S)}(z) = P^{(S)}(z=0) \cdot \exp\left(-\alpha_S \cdot z + \frac{G_R(f_P, f_S) \cdot P^{(P)}(z=0) \cdot L_{\text{eff}}}{A_{\text{eff}}}\right), \tag{2.157}$$

where $L_{\text{eff}} = \left(1 - e^{-\alpha_P \cdot L}\right)/\alpha_P$ is the effective fiber length (2.142) and $\alpha_S = \alpha(f_S)$, $\alpha_P = \alpha(f_P)$.

Raman amplification provides distributed gain, where the *local gain* is given by

$$G(z) = \frac{G_R(f_P, f_S) \cdot P^{(P)}(z)}{A_{\text{eff}}}. \tag{2.158}$$

To make the scheme comparable to lumped amplification, the *on-off gain* is defined as

$$G_{\text{on-off}} = \frac{P^{(S)}(L)}{P^{(S)}(0) \cdot e^{-\alpha_S L}} = \exp\left(\frac{G_R(f_P, f_S) \cdot P^{(P)}(0) \cdot L_{\text{eff}}}{A_{\text{eff}}}\right). \tag{2.159}$$

When the amplifier compensates the loss of a fiber of length L, $G_{\text{on-off}} = \alpha \cdot L$. An *effective noise figure* $F_{n,\text{eff}}$ is introduced as the noise figure that a (fictitious) lumped amplifier delivering $G_{\text{on-off}}$ and producing the same amount of noise would exhibit [Bro04]. This effective noise figure of Raman amplifiers can lie below one (or zero in decibel units), showing that a lumped amplifier (whose noise figure is lower-bounded by 3 dB in the high-gain region) can never outperform distributed amplification in terms of noise performance [Bro04, Agr04],[KL02a, Ch. 5]. On the other hand, the higher signal power levels maintained by distributed amplification can increase fiber nonlinearities (cf. (2.140)).

ASE noise in Raman amplifiers has its origin in spontaneous Raman scattering (cf. Section 2.1.4). Because of its distributed nature, it is sensible to describe the ASE noise process by its differential autocorrelation function [Agr04] or by its differential power spectral density (per polarization)

$$N_{\text{ASE}}(z) = n_{\text{sp}} \cdot h \cdot f_S \cdot G(z), \tag{2.160}$$

where

$$n_{\text{sp}} = \left(1 - \exp\left(-\frac{h \cdot \Delta f}{k_B \cdot T}\right)\right)^{-1} \tag{2.161}$$

is the *spontaneous scattering factor* which depends on the phonon population in the vibrational state (and thus on the temperature T) [Agr05]. At $\Delta f = |f_P - f_S| = 13.2\,\text{THz}$ and $T = 290\,\text{K}$, $n_{\text{sp}} \approx 1.13$.

The differential evolution of ASE power is described by [PW02]

$$\frac{\partial}{\partial z} P_{\text{ASE}}(z) = (G(z) - \alpha_S) \cdot P_{\text{ASE}}(z) + n_{\text{sp}} \cdot h \cdot f_S \cdot G(z) \cdot B_{\text{opt}}, \tag{2.162}$$

where B_{opt} is the optical bandwidth of the Raman amplifier or of the optical filter. The local gain $G(z)$ depends on the pumping scheme (single or multiple co- or counter-propagating pumps) and the solution of the differentional equation (2.162) requires integration over $G(z)$ [PW02, Agr05].

In comparison to a counter-propagating pump, when the pump co-propagates with the signal, the noise performance is improved but fiber nonlinearities are enhanced because of the increased maximum signal power. It can be shown that the best performance in terms of the *optical signal-to-noise ratio* (OSNR) for a fixed nonlinear phase rotation ϕ_{NL} is obtained by maintaining a constant signal power along the link, i. e. when the local gain equals the local loss [PW02]. Such a scheme can be achieved by cascaded pumps. In this case, the ASE PSD (per polarization) of the entire link simplifies to

$$N_{\mathrm{ASE}} = n_{\mathrm{sp,av}} \cdot h \cdot f_S \cdot \alpha_S \cdot L, \tag{2.163}$$

where $n_{\mathrm{sp,av}}$ is an averaged value that depends on the various pumps and their spectral distance from the signal [PW02].

The physics behind optical amplification and the engineering challenges that the design of optical amplifiers implies fills volumes [Des02a, DBDB02]. Since the invention of EDFAs in 1987, enormous progress has been made in the development of amplifiers [KL02a]. Amplifiers with a spectrally wide flat gain are available using multiple pump lasers at different wavelengths and gain-flattening optical filters [Bro04, Agr02, KL02a]. In this thesis, amplifiers are considered ideal components in the sense that they provide a spectrally flat and signal-independent gain and produce AWGN with PSD (2.155) or (2.160). The difference of noise in optically amplified communication systems to that in conventional communications is three-fold:

- ▷ In multi-span systems, noise accumulates, resulting in a total PSD that is the sum of the PSDs of all amplifiers along the link. The accumulated noise power depends on the (lumped) amplifier spacing. For identical spans, $L_{\mathrm{span,opt}} = 1/\alpha$.
- ▷ ASE noise propagates with the signal (except in the case of a single-span system with lumped amplification) and interacts with it through the fiber nonlinearities. This interaction is discussed in Chapter 5.
- ▷ In direct-detection (DD) systems, the incoming light (which is signal and additive noise) is squared, transforming the ASE into a non-additive non-Gaussian noise process. DD is discussed in Section 3.6.3.

2.4 Other components

2.4.1 Photodetection

The conversion of light into electrical current at the receiver is achieved by semiconductor photodiodes. A photon with energy hf generates an electron-hole pair if the photon energy exceeds the semiconductor's bandgap energy. The photocurrent i_P is proportional to the optical power P, i. e.

$$i_P = R \cdot P = R \cdot |a|^2, \tag{2.164}$$

where R is the photodiode's *responsivity* (in A / W), and the normalized signal $a(z,t)$ in units of $\sqrt{\text{W}}$ was defined in (2.78). Hence, the photodiode is a square-law device that delivers the squared absolute value of the optical signal. The implications this has for DD receivers are discussed in Section 3.6.3. To recover the electric field of the light wave, coherent receiver structures must be used. Such receivers combine the incident lightwave with light from a local laser so that the photodiode produces a beat term that mixes the optical signal into the electrical baseband. Such coherent receiver structures are introduced in Chapter 4.

Owing to the particle nature of light, photodetection causes shot noise. This noise process sets the *quantum limit of photodetection*, i. e. the minimum number of photons per bit required to achieve a given bit error rate (BER). This limit can be as low as 9 photons per bit for a BER of 10^{-9} for coherent receivers and phase-shift keying (PSK) modulation [Agr02]. In addition, the electric receiver circuit produces thermal noise. Both shot and thermal noise can be neglected in the analysis of optically amplified lightwave systems, in which ASE noise dominates.

Photodiodes are characterized by parameters such as their quantum efficiency, dark current, rise time and bandwidth. These parameters are not considered in this thesis. The responsivity can range from $R = 0.4\,\text{A} / \text{W}$ in p-i-n photodiodes to $R > 100\,\text{A} / \text{W}$ for avalanche photodiodes (APD) [Agr02].

2.4.2 Lasers

The light source in optical communication systems are semiconductor lasers (*light amplification by stimulated emission of radiation*). They are based on the same principle of stimulated emission as optical amplifiers. A resonator structure reflects part of the light, so that a coherent light wave can build up inside the laser cavity if the pump power is sufficient to compensate for the light that exits the laser and for internal loss. When the reflection is not wavelength-selective, several longitudinal modes are emitted by the laser. To obtain *single-longitudinal mode* (SLM) lasers, frequency-selective reflectors are used; these can be grating structures that reflect light by means of *Bragg diffraction* or external cavities [Agr02].

Because of phase fluctuations, the light wave generated by a laser is not perfectly monochromatic. Commercial *distributed feedback* (DFB) lasers have linewidths in the MHz range [Agr02]; *external cavity lasers* (ECL) can have linewidths as small as 100 kHz [JMST09]. The demands on the laser linewidth depend on the data rate and the modulation format of the signal and are not satisfied by commercially available lasers in all cases [Sei08]. In this thesis, the impact of the nonzero laser linewidth is not considered.

2.4.3 Modulators

At high data rates, lasers are not directly modulated because of the unwanted frequency chirp they produce [Agr02]. Instead, external modulators are used to modulate the monochromatic laser light.

In a *Mach-Zehnder modulator* (MZM), schematically depicted in Figure 2.7, unmodulated laser light with amplitude $\hat{E}_{\text{in}}$ is split into two branches. A waveguide material

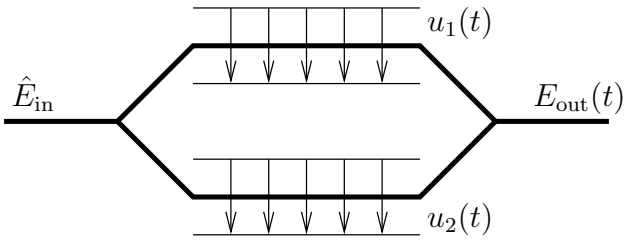

Figure 2.7: Structure of a dual-drive Mach-Zehnder modulator.

showing a strong electro-optic effect is used, commonly $LiNbO_3$ [Agr02]. As voltages $u_1(t)$ and $u_2(t)$ are applied to the branches, the refractive index of the material changes and a phase shift is induced. The complex envelope of the recombined branches can be written as [KWM+05]

$$E_{\text{out}}(t) = \hat{E}_{\text{in}} \cdot \cos\left(\frac{\pi \cdot (u_1(t) - u_2(t))}{2V_\pi}\right) \cdot e^{\jmath \cdot \theta \cdot \frac{\pi \cdot (u_1(t) - u_2(t))}{2V_\pi}}, \tag{2.165}$$

where V_π denotes the voltage difference $u_1(t) - u_2(t)$ that generates a phase shift of π between the two branches and

$$\theta = \frac{u_1(t) + u_2(t)}{u_1(t) - u_2(t)} \tag{2.166}$$

is the *asymmetry factor*. To generate pure amplitude modulation, the MZM is driven in *push-pull* configuration by selecting $u_1(t) = -u_2(t) = u(t)/2$, i. e. $\theta = 0$, resulting in

$$E_{\text{out}}(t) = \hat{E}_{\text{in}} \cdot \cos\left(\frac{\pi \cdot u(t)}{2V_\pi}\right). \tag{2.167}$$

The dual-drive MZM reaches every point in the complex plane and can – in principle – be used to generate arbitrary waveforms. However, in practice, *triple MZM* structures are less error-prone [WFBP06]. A triple MZM consists of two separate MZMs in push-pull configuration for the signal's in-phase and quadrature component, respectively, and a phase modulator that imprints a $\pi/2$ phase shift onto the output of the quadrature branch's MZM. The resulting signal is described by

$$E_{\text{out}}(t) = \frac{\hat{E}_{\text{in}}}{2} \cdot \left(\cos\left(\frac{\pi \cdot u_I(t)}{2V_\pi}\right) - \jmath \cos\left(\frac{\pi \cdot u_Q(t)}{2V_\pi}\right)\right). \tag{2.168}$$

Two parameters characterizing practical external modulators are the *extinction ratio* and the *modulation bandwidth*. They are not considered in this thesis, so MZMs are modeled as ideal elements with nonlinear transfer function (2.167).

2.4.4 Other optical components

Optical couplers can be used to split or combine optical signals as depicted in Figure 2.8. An optical coupler with a power splitting ratio of 1/2 (*3-dB coupler*) is modeled as [Agr04]

$$\underbrace{\begin{pmatrix} E_{\text{out},1}(t) \\ E_{\text{out},2}(t) \end{pmatrix}}_{\boldsymbol{E}_{\text{out}}} = \underbrace{\frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & \jmath \\ \jmath & 1 \end{pmatrix}}_{\boldsymbol{H}} \cdot \underbrace{\begin{pmatrix} E_{\text{in},1}(t) \\ E_{\text{in},2}(t) \end{pmatrix}}_{\boldsymbol{E}_{\text{in}}}. \tag{2.169}$$

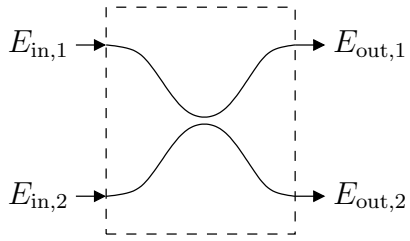

Figure 2.8: Optical coupler.

If this is mathematically more convenient, it is common to use different coupler transfer matrices such as (the factor of $1/\sqrt{2}$ omitted) [Ho05a, EKW+10]

$$\begin{pmatrix} 1 & \jmath \\ 1 & -\jmath \end{pmatrix}, \quad \begin{pmatrix} 1 & \jmath \\ -1 & \jmath \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \tag{2.170}$$

Other components of optical communication systems, such as *optical filters*, *polarization controllers*, *polarization beam splitters*, *multiplexers* and *optical hybrids* are introduced in the text where they appear.

2.4.5 Electrical components

The aim of optical communication systems is to transmit information which is encoded and stored in electrical signals. Therefore, the electrical transmitter and receiver are crucial components of lightwave systems. Frequently, the limited speed and bandwidth of the electrical components are the limiting factor in the design of optical communication systems.

This thesis does not consider limitations set by electrical components. The exponential development in the performance of electronic equipment described by Moore's law renders the persistence of such limitations in the future unlikely. Noise generated by electrical components is of course an exception. However, as mentioned before, shot noise generated during photodetection as well as thermal noise originating from the electric receiver circuit are subordinate noise sources compared to ASE noise and are therefore neglected in this thesis. In optical communication systems that are not optically amplified, they are of course the limiting factor.

2.5 Summary

This introductory chapter is intended to introduce the fundamentals of optical communication systems on a physical level. A focus is put on the propagation of electro-magnetic waves in optical fibers. In summary, the main points of this chapter that later chapters will build upon are:

- ▷ Light propagation in optical fibers is governed by Maxwell's equations. The nonlinear Schroedinger equation (NLSE) describes the propagation of a signal's envelope in a nonlinear dispersive fiber. To derive the NLSE, small terms of the material susceptibility are treated as perturbative terms.
- ▷ With the exception of Brillouin scattering, all relevant fiber effects have their origin in the material susceptibility χ. Due to the high signal intensities in single-mode fibers, the third order in a series expansion of χ must be taken into account. Even-order terms equal zero because of material symmetries in SiO_2.
- ▷ The linear material susceptibility causes attenuation, chromatic dispersion, birefringence and mode coupling. The cubic material susceptibility is responsible for the Kerr effect, causing self-phase modulation (SPM), cross-phase modulation (XPM) and four-wave mixing (FWM), and for Raman scattering. The relation of all these effects to the susceptibility function was derived. The frequency-dependence of $\overleftrightarrow{\boldsymbol{X}}^{(3)}$ can be neglected in many cases.
- ▷ Stimulated Brillouin scattering limits the peak power in fiber-optic systems. Because of its narrow gain, it does not cause WDM channel interference.
- ▷ A light wave's polarization is not changed through nonlinear effects *on average*. Therefore, a scalar description of the light propagation in optical fibers is valid. The cubic susceptibility is reduced to an effective value by random polarization rotation. This value depends on the frequency spacing between interacting waves.
- ▷ The NLSE has a closed-form solution only for few special cases. A Volterra series expansion can be used for a formal solution; its third-order term represents an approximate expression for the distortions imposed by fiber nonlinearities in the presence of dispersion. The split-step Fourier method is used for a numerical solution of the NLSE.
- ▷ Nonlinear fiber effects were introduced on a system level. Characteristic length scales were defined.
- ▷ Photodiodes represent a square-law element in direct-detection receivers, i. e. only the power of the electrical field is received and information about the phase is lost.
- ▷ The capacity of lightwave systems is ultimately limited by fiber loss and the ASE noise from optical amplifiers used to compensate it. In the presence of ASE noise, shot and thermal noise play a subordinate role and can be neglected.

3

Elements of information theory

The optical fiber channel introduced in Chapter 2 offers an incomparably large usable bandwidth and a very small attenuation but impairs the signal through nonlinearities. How much data can we reliably transmit over such a channel? This type of question is answered by *information theory*, a field that builds upon Claude Shannon's 1948 publication *A mathematical theory of communication* [Sha48]. Shannon showed that there is a maximum rate (the *channel capacity* C) at (or below) which communication is possible with arbitrarily small error probability. For the band-limited channel with *additive white Gaussian noise* (AWGN), Shannon derived

$$C = W \cdot \log_2 \left(1 + \frac{P_s}{W N_0}\right), \tag{3.1}$$

where W is the channel bandwidth, N_0 the power spectral density of the equivalent baseband channel and P_s the average signal power.

In a very simplified model that neglects all nonlinear effects, we can consider the fiber-optic channel with optical amplification as an AWGN channel. Because of the fiber's large bandwidth, it is tempting to let $W \to \infty$, which transforms (3.1) into

$$\lim_{W \to \infty} C = \frac{P_s}{N_0 \ln 2}. \tag{3.2}$$

Using (2.163) to approximate N_0, we obtain a capacity-distance product (per polarization and normalized to the input power) of $C \cdot L / P_s = (n_{\mathrm{sp,av}} \cdot h \cdot f \cdot \alpha)^{-1} \approx 2.2 \cdot 10^{20}\,\mathrm{bit/s} \cdot \mathrm{km} / \mathrm{W}$ (for $n_{\mathrm{sp,av}} \approx 1.13$, $f \approx 193\,\mathrm{THz}$ and $\alpha = 0.2\,\mathrm{dB} / \mathrm{km}$). Although it must be noted that C converges to (3.2) only for unrealistically long distances (N_0 large) or large bandwidths W, the numerical result nevertheless gives an idea of the enormeous channel capacity of optical fibers.

The first part of this chapter (Sections 3.1–3.2) provides the background on information theory that is required for a more substantiated analysis of the fiber-optic communication channel. In Section 3.1, basic terms of information theory are introduced. Section 3.2 contains a discussion of the channel capacity and provides capacity results for some important channel types.

In the second part of this chapter (Sections 3.3–3.6), a very intuitive and useful decomposition method for mutual information is developed. This method is derived and illustrated in Section 3.3. In Section 3.4, the decomposition method is applied to the AWGN channel with Gaussian input and phase-modulated input as well as for discrete input constellations. Section 3.5 contains a compact introduction to directional statistics. The reviewed terms and methods are required for the treatment of phase noise. Section 3.6 discusses channels that are only partially able to convey information encoded in the phase of the signal. Such channels are said to be *partially coherent*. Examples include systems affected by phase noise as well as optical communication systems with direct-detection receivers. Section 3.7 summarizes the results of this chapter.

3.1 Fundamentals

3.1.1 Entropy

Even before 1948, information was regarded (e. g. by R. V. L. Hartley and R. A. Fisher) as the randomness of a random variable. In this historical context, Shannon developed his information theory as a probabilistic concept. Shannon defines the *entropy* of a random variable X with realizations x from the support $\mathcal{X}$ as

$$H(X) = \mathcal{E}_X \left\{ \log \frac{1}{P(X)} \right\} = -\sum_{x \in \mathcal{X}} P(x) \log P(x), \tag{3.3}$$

where $P(X)$ is the probability mass function (PMF) of X, i. e. $X \sim P(x)$. It is common to use either the natural logarithm, leading to a result in *nats*, or the binary logarithm, which yields the entropy in units of *bits*, short for *binary digits*.[1] The entropy can be seen as a measure of the average uncertainty in the random variable; it is the number of bits required on average to describe the random variable [CT91]. Shannon chose the term *entropy* because of the similarity of (3.3) with the entropy in thermodynamics [Sha48]. His work fueled the discussion of connections between physics and information theory and fundamental physical limits for the storage, representation and processing of information. Significant contributions to this field were made by L. Brillouin [Bri56], who is well-known in optical communications for the scattering effect named after him (Section 2.1.5). Entropy satisfies $H(X) \geq 0$; it is maximized for uniformly distributed random variables. In that case, $H(X) = \log |\mathcal{X}|$ ($|\mathcal{X}|$ being the cardinality of $\mathcal{X}$), a quantity sometimes referred to as *Hartley entropy* [Har28].

[1] Shannon attributes the term *bits* to J. W. Tukey [Sha48].

The *entropy rate* of a discrete stochastic process, $X_1, X_2, \ldots$, is defined as

$$H(\boldsymbol{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n). \tag{3.4}$$

If the process is stationary, (3.4) can be written as [CT91]

$$H(\boldsymbol{X}) = \lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2}, X_{n-3}, \ldots). \tag{3.5}$$

If $X_1, X_2, \ldots$ are independent and identically distributed (i. i. d.) random variables, the entropy rate is calculated from (3.3). Entropy acquires its significance from Shannon's *source coding theorem*, which states that the data rate of a source cannot be reduced by data compression to a value below the source's entropy rate.

The *joint entropy* of two random variables X, Y is defined as

$$H(X,Y) = -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P(x,y) \log P(x,y), \tag{3.6}$$

the *conditional entropy* is defined as

$$H(X|Y) = -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P(x,y) \log P(x|y). \tag{3.7}$$

Similarly to (3.3), (3.6) and (3.7), the (joint, conditional) *differential entropy* of continuous random variables $X \sim p(x), Y \sim p(y)$ are defined as

$$h(X) = -\int_{\mathcal{X}} p(x) \log p(x) \mathrm{d}x, \tag{3.8}$$

$$h(X,Y) = -\int_{\mathcal{X}} \int_{\mathcal{Y}} p(x,y) \log p(x,y) \mathrm{d}x \mathrm{d}y, \tag{3.9}$$

$$h(X|Y) = -\int_{\mathcal{X}} \int_{\mathcal{Y}} p(x,y) \log p(x|y) \mathrm{d}x \mathrm{d}y. \tag{3.10}$$

Differential entropy is less intuitive than the entropy of a discrete random variable. In particular, it can become negative. For example, the differential entropy of a complex Gaussian random variable, given in (3.24), becomes negative for $P_s < 1/(\pi e)$.

3.1.2 Mutual information

Mutual information is the amount by which the uncertainty about a random variable can be reduced by observing another random variable:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}. \end{aligned} \tag{3.11}$$

Mutual information is symmetric, i. e. $I(X;Y)=I(Y;X)$, and non-negative, i. e. $I(X;Y) \geq 0$, with equality if and only if X and Y are statistically independent. The mutual information between continuous random variables has the same properties. It is defined as

$$I(X;Y) = h(X) - h(X|Y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \mathrm{d}x\mathrm{d}y. \tag{3.12}$$

Conditional mutual information between discrete random variables is defined as

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \sum_{z\in\mathcal{Z}} P(x,y,z) \log \frac{P(x,y|z)}{P(x|z)P(y|z)}. \tag{3.13}$$

Similary, the conditional mutual information between continuous (or between continuous and discrete) random variables can be defined. The *chain rule of information* is stated in terms of the conditional mutual information as [CT91]

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, X_{i-2}, \ldots, X_1). \tag{3.14}$$

The mutual information between discrete random processes,

$$I(\boldsymbol{X};\boldsymbol{Y}) = \lim_{n\to\infty} \frac{1}{n} I(X_1, X_2, \ldots; Y_1, Y_2, \ldots), \tag{3.15}$$

is called the *information rate* between the processes [ALV+06].

An important property of mutual information is given by the *data processing inequality*. Let X, Y, Z be random variables that form a Markov chain in that order; this is the case if and only if X and Z are conditionally independent given Y. Then [CT91]

$$I(X;Y) \geq I(X;Z). \tag{3.16}$$

The important implication of (3.16) is that it is impossible to infer any additional information about X (e. g. a channel input) by manipulation of Y (e. g. the channel output): $I(X;g(Y)) \leq I(X;Y)$. To satisfy (3.16) with equality, the function $g : Z = g(Y)$ must be injective, so that $I(X;Y|Z) = 0$.

A useful measure of the "distance" between probability distributions is given by the *Kullback Leibler (KL) distance* or *relative entropy* or *divergence*. For PMFs $P(X)$ and $Q(X)$, it is defined as

$$D(P\|Q) = \sum_{x\in\mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}. \tag{3.17}$$

The KL distance is always non-negative, $D \geq 0$, with equality if and only if $P = Q$. However, as it is not symmetric and does not satisfy the triangle equation, it is not a true distance measure. It is instructive that the mutual information between X and Y (3.11) can be expressed as the KL distance between the joint PMF $P(x,y)$ and the product of the marginal distributions $P(x)P(y)$, i. e. the joint PMF under statistical independence of X and Y, as $I(X;Y) = D(P(x,y)\|P(x)P(y))$.

3.2 Channel capacity

3.2.1 Introduction and calculation

In contrast to the physical channel description of Chapter 2, it suffices for information-theoretic purposes to characterize a *channel* with time-discrete input $\boldsymbol{X}$ and output $\boldsymbol{Y}$ by the conditional PDF $p(\boldsymbol{Y}|\boldsymbol{X})$. This definition incorporates a variety of practical channel types: memory-less channels with i. i. d. input can be characterized by $p(y|x)$; continuous-time channels (so-called *waveform channels*, of which optical fibers are an example) can be discretized in time by means of the sampling theorem; discrete channels with finite input and output alphabet $\mathcal{X}$ and $\mathcal{Y}$ are described by the conditional PMF $P(\boldsymbol{Y}|\boldsymbol{X})$.

Shannon showed [Sha48] that the *channel capacity*

$$C = \max_{p(\boldsymbol{X})} I(\boldsymbol{X};\boldsymbol{Y}) \tag{3.18}$$

is the maximum rate at which communication is possible with arbitrarily small error probability. Shannon offered an intuitive geometrical proof (*sphere packing*) of this *channel coding theorem* in 1949 [Sha49]. The capacity of a discrete-time channel is usually given in units of *bits per channel use* or *bits per symbol*. If the channel is continuous-time, the number of samples per time permitted by the sampling theorem determines a capacity value in units of bits/s. Dividing this value by the channel bandwidth yields a capacity value in units of bits/s/Hz. If the channel bandwidth equals the symbol rate, the capacity in bits/s/Hz is identical to the capacity in bits per symbol. In this thesis, the unit of capacity will be clear from the context. If the input symbols are i. i. d. and the channel does not introduce memory (i. e. statistical dependencies among the symbols), (3.18) becomes

$$C = \max_{p(X)} I(X;Y). \tag{3.19}$$

The maximization in (3.18) or (3.19) is performed over all possible input distributions $p(\boldsymbol{X})$ or $p(X)$, usually under some constraint such as an average power constraint, $\mathcal{E}\{|X|^2\} \leq P_s$, or a peak power constraint. Although $I(X;Y)$ is a concave function of $p(x)$ [CT91], the maximum can be difficult to find because of the large number of possible distributions satisfying the given constraint(s). An analytical solution may be found using the method of Lagrange multipliers or the Kuhn-Tucker conditions. For numerical input optimization, the iterative Blahut-Arimoto algorithm may be used [CT91]. In general, there is no closed form solution for the capacity [CT91].

In many cases, e. g. for channels with additive noise, $H(Y|X)$ is independent of $p(x)$, so that the capacity-achieving input distribution can be found by maximizing the entropy $H(Y)$. The entropy of a random variable $X \sim p(X)$ satisfying m constraints of the form

$$\int_{\mathcal{X}} p(x) f_i(x) \mathrm{d}x = c_i, \quad 1 \leq i \leq m, \tag{3.20}$$

is maximized by the input distribution

$$p(x) = e^{\lambda_0 + \sum_{i=1}^{m} \lambda_i f_i(x)}, \tag{3.21}$$

where $\lambda_0, \ldots, \lambda_m$ are chosen such that (3.20) and $\int_{\mathcal{X}} p(x)\mathrm{d}x = 1$ are satisfied [CT91].

The following maximum-entropy distributions appear later in this chapter:

- ▷ The Gaussian distribution maximizes the entropy of a continuous random variable under an average power (variance) constraint.
- ▷ The Maxwell-Boltzmann distribution maximizes the entropy of a discrete random variable under an average power (variance) constraint. This distribution varies from a uniform to a Gaussian distribution.
- ▷ The uniform distribution maximizes the entropy of an unconstrained (circular) random variable.
- ▷ The von Mises distribution maximizes the entropy of a circular random variable under a circular variance constraint.

These four distributions are all special cases of (3.21).

3.2.2 Important channel types

The additive white Gaussian noise channel

The time-discrete additive white Gaussian noise (AWGN) channel with complex-valued input X and output Y is described by

$$Y = X + N, \tag{3.22}$$

where N is a sampled zero-mean circularly symmetric complex-valued white Gaussian noise process with variance per real dimension σ_n^2, denoted by $N \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma_n^2)$:

$$p(n) = \frac{1}{2\pi\sigma_n^2} \cdot \exp\left(-\frac{\Re\{n\}^2 + \Im\{n\}^2}{2\sigma_n^2}\right). \tag{3.23}$$

If the input is unconstrained, the capacity of this channel is infinitely large. It is sensible to assume an average power constraint, i. e. $\mathcal{E}\{|X|^2\} \leq P_s$. In that case, a Gaussian input $X \sim \mathcal{N}_{\mathbb{C}}(0, P_s)$ maximizes the differential entropy (3.8), which becomes [Sha48, CT91]

$$h(X) = \log(\pi e P_s). \tag{3.24}$$

The capacity of the AWGN channel can be calculated as

$$\begin{aligned} C &= \max_{p(x)} h(Y) - h(Y|X) = \max_{p(x)} h(Y) - h(X+N|X) \\ &= \max_{p(x)} h(Y) - h(N) = \log_2(\pi e(P_s + 2\sigma_n^2)) - \log_2(\pi e 2\sigma_n^2) \\ &= \log_2\left(1 + \frac{P_s}{2\sigma_n^2}\right). \end{aligned} \tag{3.25}$$

Equation (3.25) yields a result in bits per symbol or bits per channel access. It is remarkable that this result depends only on the *signal-to-noise ratio* (SNR) $P_s/(2\sigma_n^2)$.

According to the sampling theorem, the number of independent samples per time unit that a continuous-time channel can carry is given by its bandwidth W. Hence, the capacity of the complex-valued band-limited waveform AWGN channel (3.1) is obtained from (3.25) by multiplication with the channel bandwidth W, yielding a result in bits/s. Divided by W, the waveform channel capacity in bits/s/Hz is given by (3.25). This differential expression is useful for channels with frequency-dependent SNR (discussed later in this section) or channels whose bandwidth is not explicitly stated.

In practical systems, the channel input is usually constrained to a discrete set of input symbols (the *constellation*). The capacity is then given by

$$C = \max_{P(X)} I(X;Y) = \max_{P(X)} \sum_{x\in\mathcal{X}} P(x) \int_{\mathcal{Y}} p(y|x) \log \frac{p(y|x)}{p(y)} \mathrm{d}y. \tag{3.26}$$

Under an average power constraint, the capacity-achieving input distribution is a Maxwell-Boltzmann distribution [KP93].

The Shannon limit (3.25) as well as the capacity curves of uniform M-ary input constellations, $M = 2, \ldots, 10$, are shown in Figure 3.1. At high SNRs, the curves saturate

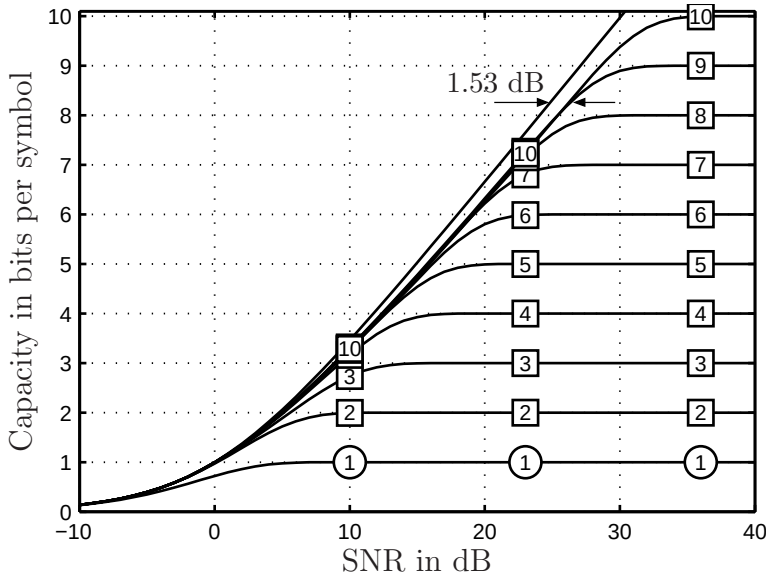

Figure 3.1: Capacity of AWGN channel over SNR in dB, $10\log_{10}(P_s/(2\sigma_n^2))$, with Gaussian input (unmarked line) and uniform M-QAM input (lines marked with numbers $\log_2 M$); $M=2$: BPSK, $M=8$: rectangular 8-QAM.

at $\log_2 M$, which is the entropy $H(X)$ of the input X. It can be observed that uniform discrete input constellations with a sufficient number of points M approach the Shannon limit closely for low SNRs. With increasing SNR, a gap opens up between uniform and Gaussian inputs. By replacing the uniform input distribution with a "Gaussian-like"

distribution – a process known as *signal shaping* –, this gap can be closed; it is therefore known as the *shaping gain*. The shaping gain tends towards a maximum value of $10\log_{10}(\pi e/6) \approx 1.53\,\text{dB}$ for large SNRs [FU98].

The discrete memoryless channel (DMC)

The capacity of a channel with discrete input and continuous output (3.26) is determined by the conditional probability density function $p(y|x)$. In contrast, the capacity of the memoryless discrete-input discrete-output channel is governed by the conditional probability mass function $P(y|x)$. As an example for a DMC, consider a system whose receiver processes the output samples of the AWGN channel (3.22) in order to recover the transmitted symbols. To minimize the symbol error probability in that case, the receiver will decide for the symbol that has the minimum Euclidean distance to the received sample. As a consequence of the data processing inequality (3.16), this quantization (known as *hard decision*) of the AWGN channel output Y will inevitably destroy information and decrease the channel capacity. In contrast, *soft-decision* receivers process the continuous output samples Y, e. g. as an input to iterative error correcting codes [Kra09]. Figure 3.2 illustrates the capacity of this examplary DMC for 4-QAM and 16-QAM input constellations. The significant capacity loss induced by hard decision can be clearly observed.

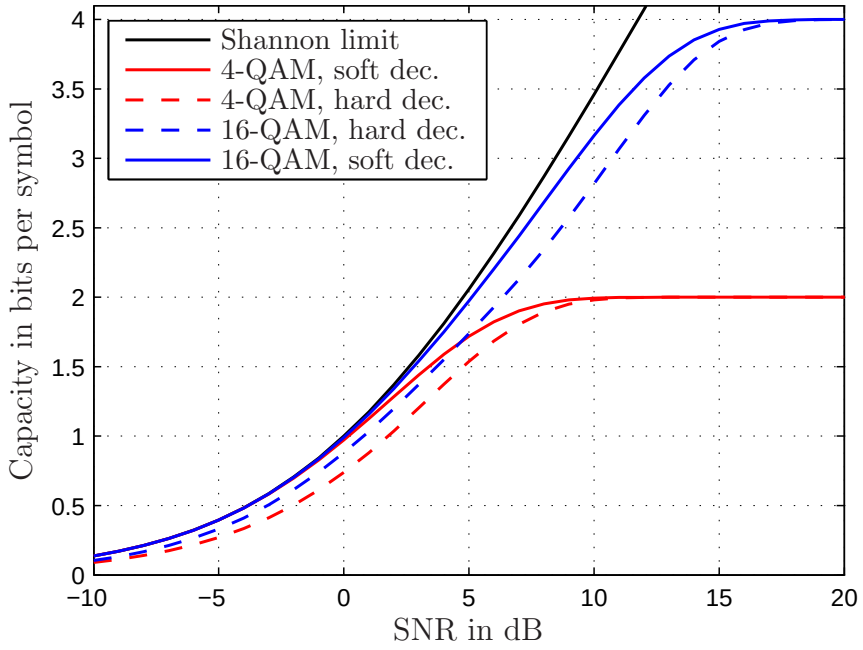

Figure 3.2: Comparison of channel capacities of AWGN channel with soft-decision (solid lines) and hard-decision (dashed lines) receiver for 4-QAM (red) and 16-QAM (blue) input constellations. The top line (black) indicates the Shannon limit.

Channels with memory

The capacity calculation is significantly complicated by channels that introduce statistical dependencies between the transmitted symbols. To find the capacity of such channels, (3.15) is maximized over all possible input processes $p(\boldsymbol{X})$.

An important special case of a waveform channel with memory is the Gaussian channel with memory. This channel can be characterized as a linear time-invariant (LTI) system with transfer function $H(f)$ followed by additive white or colored Gaussian noise with power spectral density (PSD) $\Phi_n(f)$. The frequency-dependent SNR of this channel is $\Phi_x(f) \cdot |H(f)|^2 / \Phi_n(f)$, where $\Phi_x(f)$ denotes the input signal power spectral density. The channel capacity can be found by decomposing the channel into many parallel flat, i. e. frequency-independent, sub-channels, which leads to [CT91]

$$C = \int_{-W/2}^{W/2} \log_2 \left(1 + \frac{\Phi_x(f) \cdot |H(f)|^2}{\Phi_n(f)} \right) \mathrm{d}f. \tag{3.27}$$

The capacity-achieving input distribution is Gaussian $\mathcal{N}_{\mathbb{C}}(0, P_s)$ [CT91]. The input PSD must be chosen such that more power is allocated to frequencies where the noise PSD is small. In an illustrative analogy, this procedure is known as *water filling*. The capacity-achieving input PSD can be expressed as

$$\Phi_x(f) = \frac{(\rho - \Phi_n(f))^+}{|H(f)|^2}, \tag{3.28}$$

where the function $(\,\cdot\,)^+$ returns the positive part of its argument, i. e.

$$(\rho - \Phi_n(f))^+ = \begin{cases} \rho - \Phi_n(f), & \rho \geq \Phi_n(f), \\ 0, & \rho < \Phi_n(f), \end{cases} . \tag{3.29}$$

This ensures $\Phi_x(f) \geq 0$; no power is allocated to frequencies whose noise spectrum $\Phi_n(f)/|H(f)|^2$ exceeds a certain threshold ρ. In (3.28), the constant $\rho = \Phi_x(f) \cdot |H(f)|^2 + \Phi_n(f)$ is chosen such that the average power constraint

$$\int_{-W/2}^{W/2} \Phi_x(f) \mathrm{d}f = P_s \tag{3.30}$$

is satisfied.

In many cases, the capacity of a channel with memory is higher than that of a comparable memoryless channel ("memory increases capacity") [Wol67, EKW+10]. However, there are cases in which memory decreases capacity [Arn03]. In general, the effect of memory on the channel capacity depends on the exact channel model and the normalization of the channel's transfer function or impulse response [XP03]. In fiber-optic communication systems, the main source of memory is the fiber's chromatic dispersion (cf. Section 2.2.2). Neglecting nonlinear fiber effects, the fiber channel can be characterized by the transfer function (2.130), from which $|H(f)|^2 = \text{const.}$ is easily derived. Considering the white

PSD of ASE noise (2.152), it can be seen that (3.27) and (3.1) are equivalent. Hence, the memory introduced by chromatic dispersion does not (by itself) affect the channel capacity.

A numerical approach to calculating the information rate (3.15) was proposed in [AL01] and [PSS01]. The idea is to make use of the *Shannon-McMillan-Breiman theorem*,

$$\lim_{n\to\infty} -\frac{1}{n}\log P(x_1, x_2, \ldots, x_n) = H(X) \tag{3.31}$$

for any ergodic source X [CT91]. The channel, characterized by $P(\boldsymbol{Y}|\boldsymbol{X})$, can be described by a finite state model and represented in a trellis diagram. Using

$$\log P(x_1, \ldots, x_n) = \sum_{i=1}^{n} \log P(x_i|x_{i-1}, \ldots, x_1), \tag{3.32}$$

the entropy rates $H(\boldsymbol{Y})$ and $H(\boldsymbol{Y}|\boldsymbol{X})$ required for computation of the information rate (3.15) can be calculated by computing the forward sum-product recursion of the BCJR algorithm [ALV+06, BCJR74, BC02]. In a series of papers, this numerical approach has been applied by Djordjevic et al. and Franceschini et al. to estimate the information rate of direct-detection optical communications in the presence of chromatic dispersion and filtering [FFRB06, FBF+07], fiber nonlinearities [DVIG05, IDV07], analog-to-digital conversion [DAPR07] and PMD [MDB+07].

Channels with feedback

The capacity of channels with a feedback link, i. e. an ideal and instantaneous link from receiver to transmitter, was considered by Shannon in 1956 [Sha56]. He showed that feedback does not increase the capacity of the AWGN channel. The situation is different for the Gaussian channel with colored noise, where feedback can increase capacity by up to one bit (for the complex-valued Gaussian channel) or a factor of two [CT91]. Nevertheless, feedback links are not considered in this thesis, as their implementation in high-speed optical transport networks is doubtful.

Time-variant channels

The transfer function H of a linear time-variant channel changes with time and can therefore be regarded as a random variable. The instantaneous realization $H(t)$ is called a *channel state*. As the capacity (3.27) is a function of H, $C = C(H)$, it becomes a time-variant random variable with PDF $p(C)$ itself. To determine a single capacity value for such channels, it is common to differentiate two cases [BPS98].

The *ergodic capacity*,

$$\overline{C} = \mathcal{E}_C\{C(H)\} = \mathcal{E}_H\{C(H)\} = \int_H p(H)C(H)\mathrm{d}H, \tag{3.33}$$

can be reached if the channel is ergodic, i. e. if it reveals a large number of channel states within the duration of one code word, or if both receiver and transmitter have perfect *channel state information* (CSI).

If the channel is *non-ergodic* (its time variance is slow relative to the length of a code word) and transmitter CSI cannot be assumed, the transmitter will send data at a certain constant rate. At the receiver, this data is declared lost whenever the instantaneous capacity falls below the transmission rate and the data cannot be decoded error-free (*outage*). The *x% outage capacity* C_x is defined as the value at which the capacity's cumulative distribution function (CDF) reaches $\text{Prob}(C < C_x) = x/100$:

$$\int_{H:0\leq C(H)\leq C_x} p(H)\mathrm{d}H = \int_{C=0}^{C_x} p(C)\mathrm{d}C = \frac{x}{100}. \tag{3.34}$$

The fiber-optic communication channel is largely time-invariant with the exception of polarization mode dispersion. The effect of PMD on the channel capacity is discussed in Chapter 4.

Multiple-input multiple-outpout (MIMO) channels

The channel with multiple transmitters and receivers can be written in matrix-vector notation as

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}. \tag{3.35}$$

Single-mode fibers support two orthogonal fundamental modes. As these modes are subject to mode coupling (cf. (2.18)), SMFs can be modeled as a 2×2 MIMO system with

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}. \tag{3.36}$$

This model is used in Chapter 4 to discuss the capacity of polarization-multiplexed (PolMux) systems in the presence of PMD.

In general, the capacity of an $N_T \times N_R$ MIMO channel with AWGN, no channel information at the transmitter (neither CSI nor information about the channel distribution) and full receiver CSI is given by [GJJV03, Wit05]

$$C = C(\boldsymbol{H}) = \log\det\left(\boldsymbol{I}_{N_R} + \frac{P_s}{N_T P_n}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right), \tag{3.37}$$

where $\boldsymbol{I}_{N_R}$ is the $N_R \times N_R$ identity matrix, $P_s = \sum_{i=1}^{N_T} \mathcal{E}\{|X_i|^2\}$ is the combined signal power of all inputs, $P_n = \mathcal{E}\{|N_i|^2\}$ is the noise power per output and $\boldsymbol{H}^{\mathrm{H}}$ denotes the Hermitian transpose of $\boldsymbol{H}$. Equation (3.37) implies that the transmit power is divided equally among all transmitters, i. e. $\mathcal{E}\{|X_i|^2\} = P_s/N_T$, and that the transmitted signals X_i are mutually uncorrelated Gaussian random variables.

3.3 A polar decomposition of mutual information

The information encoded in complex-valued signals has two degrees of freedom which are commonly taken to be the signal's two quadratures – its real and imaginary parts.

Alternatively, the signal can be decomposed into its polar coordinates – amplitude and phase. Historically, the first digital modulation constellations with two degrees of freedom were a combination of one-dimensional amplitude modulation (AM) and phase modulation (PM) [Cah60]. Quadrature amplitude modulation (QAM), i. e. amplitude modulation of two orthogonal carriers, was not described until 1962, with the most significant progress in understanding made in the 1970s [HWK00].

The decomposition of complex-valued signals into their real and imaginary parts is the method of choice when the sub-channels transporting them have identical form and noise statistics. In particular, this is the case for the AWGN channel e. g. with circularly symmetric Gaussian or square QAM input. In contrast, the "old-fashioned" AM-PM view can be useful when physical effects act differently on the different sub-channels. Examples are systems that clip the amplitude (e. g. nonlinear amplifiers) or systems that introduce phase noise (e. g. phase-locked loops or certain nonlinear optical fiber effects). However, even for channels that introduce equal impairments to the signal's quadratures (such as the complex-valued AWGN channel), the AM-PM view may be preferable if this facilitates the input description, for instance for ASK-PSK modulation schemes.

Decomposing signals using polar coordinates motivates decomposing the mutual information between the channel input and output using polar coordinates. We choose a decomposition that results in four terms: two partial channels with one degree of freedom each (an amplitude and a phase channel) and two mixed terms that govern the exchange of mutual information across the sub-channels.

Consider a channel with complex-valued input

$$X = X_{\shortparallel} \cdot e^{\jmath X_{\sphericalangle}}, \quad X_{\shortparallel} \in [0, \infty), X_{\sphericalangle} \in [-\pi, \pi) \tag{3.38}$$

and output

$$Y = Y_{\shortparallel} \cdot e^{\jmath Y_{\sphericalangle}}, \quad Y_{\shortparallel} \in [0, \infty), Y_{\sphericalangle} \in [-\pi, \pi), \tag{3.39}$$

where the notation $X_{\shortparallel}, Y_{\shortparallel}$ (amplitudes) and $X_{\sphericalangle}, Y_{\sphericalangle}$ (phase angles) reminds us of what parts of the signal these variables refer to. (We use lower-case fonts $x_{\shortparallel}$ to denote a realization and calligraphic fonts $\mathcal{X}_{\shortparallel}$ to denote the support of the random variable $X_{\shortparallel}$.)

The mutual information $I(X;Y)$ between this channel's input and output can be expanded by repeatedly applying the chain rule of mutual information (3.14) as

$$\begin{aligned} I(X;Y) &= I(X_{\shortparallel}, X_{\sphericalangle}; Y_{\shortparallel}, Y_{\sphericalangle}) \\ &= I(X_{\shortparallel}; Y_{\shortparallel}, Y_{\sphericalangle}) + I(X_{\sphericalangle}; Y_{\shortparallel}, Y_{\sphericalangle} | X_{\shortparallel}) \\ &= \underbrace{I(X_{\shortparallel}; Y_{\shortparallel})}_{\text{Amplitude term}} + \underbrace{I(X_{\sphericalangle}; Y_{\sphericalangle} | X_{\shortparallel})}_{\text{Phase term}} + \underbrace{I(X_{\shortparallel}; Y_{\sphericalangle} | Y_{\shortparallel})}_{\text{Mixed term I}} + \underbrace{I(X_{\sphericalangle}; Y_{\shortparallel} | X_{\shortparallel}, Y_{\sphericalangle})}_{\text{Mixed term II}}. \end{aligned} \tag{3.40}$$

The expansion (3.40) can be interpreted as decomposing the complex-valued channel with two degrees of freedom (amplitude and phase) into two sub-channels with one degree of freedom each.

The first sub-channel, represented by the amplitude term of the mutual information

$$I(X_{\shortparallel};Y_{\shortparallel}) = \int_{\mathcal{X}_{\shortparallel}} p(x_{\shortparallel}) \int_{\mathcal{Y}_{\shortparallel}} p(y_{\shortparallel}|x_{\shortparallel}) \log \frac{p(y_{\shortparallel}|x_{\shortparallel})}{p(y_{\shortparallel})} \mathrm{d}y_{\shortparallel} \mathrm{d}x_{\shortparallel} \tag{3.41}$$

conveys only the amplitude of the signal and is unaffected by impairments such as phase noise.

The second sub-channel is characterized by the phase term of the mutual information

$$\begin{aligned} I(X_{\triangleleft};Y_{\triangleleft}|X_{\shortparallel}) &= \int_{\mathcal{X}_{\shortparallel}} p(x_{\shortparallel}) I(X_{\triangleleft};Y_{\triangleleft}|x_{\shortparallel}) \mathrm{d}x_{\shortparallel} \\ &= \int_{\mathcal{X}_{\shortparallel}} p(x_{\shortparallel}) \iint_{\mathcal{X}_{\triangleleft},\mathcal{Y}_{\triangleleft}} p(x_{\triangleleft}, y_{\triangleleft}|x_{\shortparallel}) \log \frac{p(x_{\triangleleft}, y_{\triangleleft}|x_{\shortparallel})}{p(x_{\triangleleft}|x_{\shortparallel}), p(y_{\triangleleft}|x_{\shortparallel})} \mathrm{d}x_{\triangleleft} \mathrm{d}y_{\triangleleft} \mathrm{d}x_{\shortparallel} \\ &= \int_{\mathcal{X}_{\shortparallel}} p(x_{\shortparallel}) \underbrace{\int_{\mathcal{X}_{\triangleleft}} p(x_{\triangleleft}|x_{\shortparallel}) \int_{\mathcal{Y}_{\triangleleft}} p(y_{\triangleleft}|x_{\triangleleft}, x_{\shortparallel}) \log \frac{p(y_{\triangleleft}|x_{\triangleleft}, x_{\shortparallel})}{p(y_{\triangleleft}|x_{\shortparallel})} \mathrm{d}y_{\triangleleft} \mathrm{d}x_{\triangleleft}}_{I(X_{\triangleleft};Y_{\triangleleft}|x_{\shortparallel})} \mathrm{d}x_{\shortparallel} \\ &= \mathcal{E}_{X_{\shortparallel}} \left\{ I(X_{\triangleleft};Y_{\triangleleft}|X_{\shortparallel} = x_{\shortparallel}) \right\}, \end{aligned} \tag{3.42}$$

where $\mathcal{E}_X \{f(X=x)\}$ denotes the expectation of $f(X)$ with respect to the random variable X that takes on the values x. Eq. (3.42) can be paraphrased in words as the information that can be obtained about the *input phase* by observing the *output phase*, given that the *input amplitude* is already known. This term is significantly affected by phase noise, but agnostic to amplitude distortions such as clipping as long as the input amplitude is known.

After separating the complex-valued channel into an amplitude and a phase part, the two mixed terms (I and II) in (3.40) yield the "cross information" between these two sub-channels. Mixed term I represents the amount of information about the *input amplitude* that can be drawn from the *output phase* in addition to what has already been learnt about the *input amplitude* by observing the *output amplitude*. Finally, mixed term II yields the information about the *input phase* that can be obtained from observation of the *output amplitude* given the *input amplitude* and the *output phase*.

The polar decomposition of mutual information can be helpful in understanding the characteristics of the channel input, e. g. concerning symbol constellations, and transmission impairments. Moreover, the decomposition significantly simplifies the computation of the mutual information in cases where the mixed terms are zero or negligibly small. The computation of $I(X;Y)$ then reduces to evaluating the conditional probability densities in (3.41) and (3.42), which are often known. Even if the mixed terms do not vanish, the two main terms yield a lower bound on the mutual information (and can hence be used to get a lower bound on capacity).

We remark that maximizing the mutual information over all input distributions, as is required to obtain the channel capacity, cannot be performed separately for the four decomposition terms. For instance, maximizing the amplitude term over all possible input amplitude distributions and then maximizing the phase term over all input phase

distributions given the already determined input amplitude distribution does not yield the capacity (cf. Section 3.4.1).

3.4 Decomposition of the AWGN channel

The decomposition (3.40) is now applied to the AWGN channel (3.22) with the power constraint $\mathcal{E}\{|X|^2\} \leq P_s$. Analytical expressions are derived for the AWGN channel with average power constraint (Gaussian input) and with constant power constraint (phase-modulated input). In addition, decomposition results for discrete ASK/PSK and QAM constellations are presented.

3.4.1 Gaussian input

Amplitude term

The first term in the decomposition is $I(X_{\text{II}};Y_{\text{II}}) = h(Y_{\text{II}}) - h(Y_{\text{II}}|X_{\text{II}})$. The capacity of the AWGN channel (3.22) with average power constraint is maximized by $X \sim \mathcal{N}_{\mathbb{C}}(0, P_s)$ [CT91]. Since $N \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma_n^2)$, the channel output is Gaussian distributed, $Y \sim \mathcal{N}_{\mathbb{C}}(0, P_s + 2\sigma_n^2)$. Then, $Y_{\text{II}} = \sqrt{\Re\{Y\}^2 + \Im\{Y\}^2}$ follows a Rayleigh distribution with parameter $(P_s + 2\sigma_n^2)/2$ [Pro95]:

$$p(y_{\text{II}}) = \frac{2y_{\text{II}}}{P_s + 2\sigma_n^2} \cdot \exp\left(-\frac{y_{\text{II}}^2}{P_s + 2\sigma_n^2}\right). \tag{3.43}$$

The differential entropy of the output amplitude in bits is [CT91]

$$h(Y_{\text{II}}) = \frac{1}{2}\log_2(P_s + 2\sigma_n^2) + \frac{1}{\ln 2} + \frac{\gamma}{2\ln 2} - 1, \tag{3.44}$$

where $\gamma \approx 0.577$ is the *Euler constant*.

Calculating $h(Y_{\text{II}}|X_{\text{II}})$ requires knowledge of $p(y_{\text{II}}|x_{\text{II}})$, which is a Ricean distribution [Pro95]:

$$p(y_{\text{II}}|x_{\text{II}}) = \frac{y_{\text{II}}}{\sigma_n^2} \cdot \exp\left(-\frac{x_{\text{II}}^2 + y_{\text{II}}^2}{2\sigma_n^2}\right) \cdot \mathrm{I}_0\left(\frac{x_{\text{II}}y_{\text{II}}}{\sigma_n^2}\right), \tag{3.45}$$

where $\mathrm{I}_0(.)$ is the modified Bessel function of the first kind with order zero. It can be seen that for $x_{\text{II}} = 0$, the Ricean distribution turns into a Rayleigh distribution; for $P_s = 0$, (3.43) and (3.45) are equal. Using the general form (3.45) of the conditional PDF, the integration required to calculate $h(Y_{\text{II}}|X_{\text{II}})$ is intractable. A significant simplification is obtained when the channel's signal-to-noise ratio (SNR) $P_s/(2\sigma_n^2)$ is large. In this limit of large arguments of the Bessel function ($x_{\text{II}}y_{\text{II}}/\sigma_n^2 \gg 1$), we can use $\mathrm{I}_0(z) \to e^z/\sqrt{2\pi z}$ [AS72]. The Ricean PDF (3.45) then turns into the Gaussian PDF

$$p(y_{\text{II}}|x_{\text{II}}) \approx \frac{1}{\sigma_n\sqrt{2\pi}} \cdot \exp\left(-\frac{(y_{\text{II}} - x_{\text{II}})^2}{2\sigma_n^2}\right). \tag{3.46}$$

In deriving (3.46), we dropped a factor $\sqrt{y_{\shortparallel}/x_{\shortparallel}}$ which decays to 1 asymptotically with increasing SNR. With (3.46), the conditional differential entropy can be calculated as

$$h(Y_{\shortparallel}|X_{\shortparallel}) \approx \frac{1}{2}\log_2(2\pi e\sigma_n^2). \tag{3.47}$$

Finally, using (3.47) and (3.44), an asymptotic approximation for the amplitude term is

$$\begin{aligned} I(X_{\shortparallel};Y_{\shortparallel}) &= h(Y_{\shortparallel}) - h(Y_{\shortparallel}|X_{\shortparallel}) \\ &\approx \frac{1}{2}\log_2\left(1+\frac{P_s}{2\sigma_n^2}\right) - \frac{1}{2}\log_2\pi + \frac{1+\gamma}{2\ln 2} - 1 \\ &\approx \frac{1}{2}\log_2\frac{P_s}{2\sigma_n^2} \underbrace{-\frac{1}{2}\log_2\pi + \frac{1+\gamma}{2\ln 2} - 1}_{\approx -0.69}, \qquad P_s \gg 2\sigma_n^2. \end{aligned} \tag{3.48}$$

Phase term

The phase term $I(X_{\sphericalangle};Y_{\sphericalangle}|X_{\shortparallel}) = h(Y_{\sphericalangle}|X_{\shortparallel}) - h(Y_{\sphericalangle}|X_{\sphericalangle},X_{\shortparallel})$ is calculated similarly. For any input amplitude $x_{\shortparallel}$, the output phase is uniformly distributed in $[-\pi,\pi)$, so the first conditional entropy is easily found to be

$$\begin{aligned} h(Y_{\sphericalangle}|X_{\shortparallel}) &= -\int_{\mathcal{X}_{\shortparallel}}\int_{\mathcal{Y}_{\sphericalangle}} p(x_{\shortparallel},y_{\sphericalangle})\log_2 p(y_{\sphericalangle}|x_{\shortparallel})\mathrm{d}y_{\sphericalangle}\mathrm{d}x_{\shortparallel} \\ &= -\int_{\mathcal{X}_{\shortparallel}} p(x_{\shortparallel}) \underbrace{\int_{-\pi}^{\pi} p(y_{\sphericalangle}|x_{\shortparallel})\log_2 p(y_{\sphericalangle}|x_{\shortparallel})\mathrm{d}y_{\sphericalangle}}_{-\log_2(2\pi)}\,\mathrm{d}x_{\shortparallel} \\ &= \log_2(2\pi). \end{aligned} \tag{3.49}$$

Similarly, we can write

$$h(Y_{\sphericalangle}|X_{\sphericalangle},X_{\shortparallel}) = -\iint_{\mathcal{X}_{\shortparallel},\mathcal{X}_{\sphericalangle}} p(x_{\shortparallel},x_{\sphericalangle})\cdot \underbrace{\int_{-\pi}^{\pi} p(y_{\sphericalangle}|x_{\shortparallel},x_{\sphericalangle})\log_2 p(y_{\sphericalangle}|x_{\shortparallel},x_{\sphericalangle})\mathrm{d}y_{\sphericalangle}}_{-h(Y_{\sphericalangle}|x_{\shortparallel},x_{\sphericalangle})}\,\mathrm{d}x_{\shortparallel}\mathrm{d}x_{\sphericalangle}. \tag{3.50}$$

The conditional entropy $h(Y_{\sphericalangle}|x_{\shortparallel},x_{\sphericalangle})$ is not affected by the constant phase shift $x_{\sphericalangle}$, so that we can assume $x_{\sphericalangle}=0$ without loss of generality and write the conditional phase PDF as [Mid60, Pro95, AB93]

$$\begin{aligned} p(y_{\sphericalangle}|x_{\shortparallel},x_{\sphericalangle}=0) =& \frac{1}{2\pi}\exp\left(-\frac{x_{\shortparallel}^2}{2\sigma_n^2}\right) \\ &+ \frac{x_{\shortparallel}\cos y_{\sphericalangle}}{2\sqrt{\pi 2\sigma_n^2}}\cdot\exp\left(-\frac{x_{\shortparallel}^2(\sin y_{\sphericalangle})^2}{2\sigma_n^2}\right)\cdot\operatorname{erfc}\left(-\frac{x_{\shortparallel}\cos y_{\sphericalangle}}{\sqrt{2\sigma_n^2}}\right). \end{aligned} \tag{3.51}$$

The PDF (3.51) is periodic with a period of 2π; integrating it over any contiguous 2π interval yields one. Such *circular* PDFs are reviewed in Section 3.5.

If the channel SNR is low and we have $x_{\shortparallel}^2 \ll 2\sigma_n^2$, the phase becomes uniformly distributed. On the other hand, when $x_{\shortparallel}^2 \gg 2\sigma_n^2$, (3.51) can be approximated by the Gaussian PDF [Pro95]

$$p(y_\triangleleft|x_{\shortparallel}, x_\triangleleft = 0) \approx \frac{x_{\shortparallel}}{\sqrt{2\pi}\sigma_n} \cdot \exp\left(-\frac{y_\triangleleft^2}{2\sigma_n^2/x_{\shortparallel}^2}\right). \tag{3.52}$$

With this approximation, the inner entropy integral in (3.50) can be approximated as

$$h(Y_\triangleleft|x_{\shortparallel}, x_\triangleleft) \approx \frac{1}{2} \cdot \log_2\left(2\pi e \cdot \frac{\sigma_n^2}{x_{\shortparallel}^2}\right), \qquad x_{\shortparallel}^2 \gg 2\sigma_n^2, \tag{3.53}$$

and the entropy (3.50) becomes

$$\begin{aligned} h(Y_\triangleleft|X_\triangleleft, X_{\shortparallel}) &= \int_{\mathcal{X}_\triangleleft} p(x_\triangleleft) \cdot \int_{\mathcal{X}_{\shortparallel}} p(x_{\shortparallel}|x_\triangleleft) \cdot h(Y_\triangleleft|x_{\shortparallel}, x_\triangleleft)\mathrm{d}x_{\shortparallel}\mathrm{d}x_\triangleleft \\ &= \int_{-\pi}^{\pi} p(x_\triangleleft)\mathrm{d}x_\triangleleft \cdot \int_0^{\infty} p(x_{\shortparallel}) \cdot h(Y_\triangleleft|x_{\shortparallel}, x_\triangleleft)\mathrm{d}x_{\shortparallel} \\ &\approx \int_0^{\infty} \frac{x_{\shortparallel}}{P_s/2} \cdot \exp\left(-\frac{x_{\shortparallel}^2}{P_s}\right) \cdot \frac{1}{2}\log_2\left(2\pi e \cdot \frac{\sigma_n^2}{x_{\shortparallel}^2}\right)\mathrm{d}x_{\shortparallel} \\ &= \frac{1}{2}\log_2\frac{2\sigma_n^2}{P_s} + \frac{1+\gamma}{2\ln 2} + \frac{1}{2}\log_2\pi, \qquad P_s \gg 2\sigma_n^2. \end{aligned} \tag{3.54}$$

The separation of the integrals (second equality) is possible because $h(Y_\triangleleft|x_{\shortparallel}, x_\triangleleft)$ is independent of $x_\triangleleft$. In the same line, we used $p(x_{\shortparallel}|x_\triangleleft) = p(x_{\shortparallel})$ (which is a Rayleigh distribution).

Finally, the decomposition phase term can be approximated from (3.49) and (3.54):

$$\begin{aligned} I(X_\triangleleft; Y_\triangleleft|X_{\shortparallel}) &= h(Y_\triangleleft|X_{\shortparallel}) - h(Y_\triangleleft|X_\triangleleft, X_{\shortparallel}) \\ &\approx \log_2(2\pi) - \frac{1}{2}\log_2\frac{2\sigma_n^2}{P_s} - \frac{1+\gamma}{2\ln 2} - \frac{1}{2}\log_2\pi \\ &= \frac{1}{2}\log_2\frac{P_s}{2\sigma_n^2} + \frac{1}{2}\log_2\pi - \frac{1+\gamma}{2\ln 2} + 1, \qquad P_s \gg 2\sigma_n^2. \end{aligned} \tag{3.55}$$

Mixed terms

For the AWGN channel with Gaussian input, *mixed term I* in the decomposition is always zero. To prove this, observe that $p(y_\triangleleft) = p(y_\triangleleft|x_{\shortparallel}) = p(y_\triangleleft|x_{\shortparallel}, y_{\shortparallel}) = 1/(2\pi)$ in any 2π interval. Then, we obtain the conditional entropies

$$\begin{aligned} h(Y_\triangleleft|Y_{\shortparallel}) &= -\int_0^{\infty} p(x_{\shortparallel}) \underbrace{\int_{-\pi}^{\pi} p(y_\triangleleft|x_{\shortparallel}) \log_2 p(y_\triangleleft|x_{\shortparallel})\mathrm{d}y_\triangleleft}_{-\log_2(2\pi)} \mathrm{d}x_{\shortparallel} = \int_0^{\infty} p(x_{\shortparallel})\mathrm{d}x_{\shortparallel} \cdot \log_2(2\pi) \\ &= \log_2(2\pi) \end{aligned} \tag{3.56}$$

and

$$h(Y_\sphericalangle|Y_{\shortparallel}, X_{\shortparallel}) = -\iint_0^\infty p(x_{\shortparallel}, y_{\shortparallel}) \underbrace{\int_{-\pi}^{\pi} p(y_\sphericalangle|x_{\shortparallel}, y_{\shortparallel}) \log_2 p(y_\sphericalangle|x_{\shortparallel}, y_{\shortparallel}) \mathrm{d}y_\sphericalangle}_{-\log_2(2\pi)} \mathrm{d}x_{\shortparallel}\mathrm{d}y_{\shortparallel}$$

$$= \iint_0^\infty p(x_{\shortparallel}, y_{\shortparallel})\mathrm{d}x_{\shortparallel}\mathrm{d}y_{\shortparallel} \cdot \log_2(2\pi) = \log_2(2\pi), \tag{3.57}$$

and so

$$I(X_{\shortparallel}; Y_\sphericalangle|Y_{\shortparallel}) = h(Y_\sphericalangle|Y_{\shortparallel}) - h(Y_\sphericalangle|Y_{\shortparallel}, X_{\shortparallel}) = 0. \tag{3.58}$$

Mixed term II, $I(X_\sphericalangle; Y_{\shortparallel}|X_{\shortparallel}, Y_\sphericalangle)$, reaches its (numerically calculated) maximum value of approximately 0.08 bits/symbol at $10\log_{10}(P_s/(2\sigma_n^2)) = 1\,\mathrm{dB}$ and tends to zero for large SNRs.

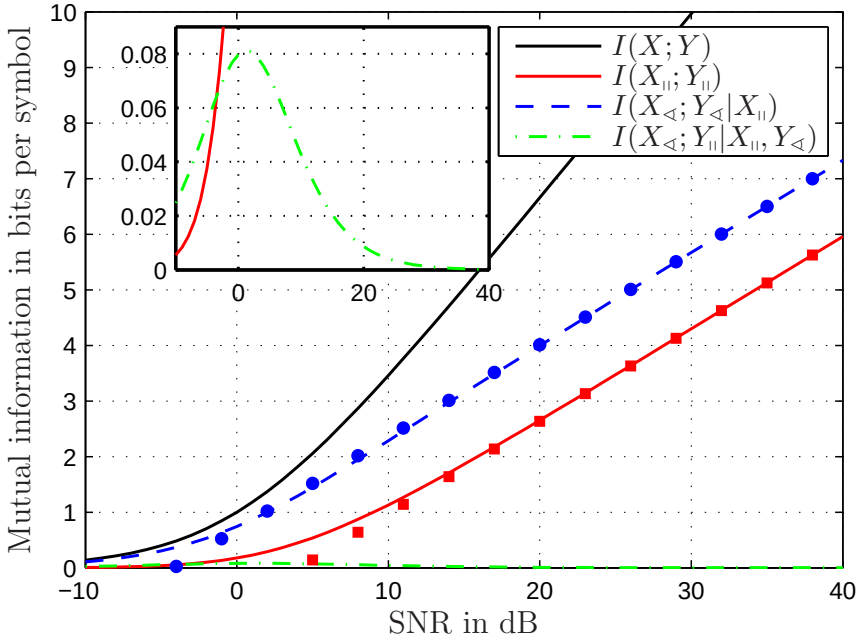

Figure 3.3: Mutual information decomposition terms as a function of SNR in dB, $10\log_{10}(P_s/(2\sigma_n^2))$, for the AWGN channel with Gaussian input. Lines show numerical results, markers correspond to analytical approximations (3.48) and (3.55). The inset shows the magnified curve of mixed term II.

The results of the decomposition for the AWGN channel with Gaussian input are shown in Figure 3.3. The depicted curves were obtained from numerical integration of the mutual information integrals; markers indicate the analytical approximations (3.48) and (3.55). Observe that the amplitude and phase terms are the main contributors to the channel capacity, whereas mixed term II (shown in the inset) is negligibly small. It can be seen that the analytical approximations are accurate at SNRs of approximately 15 dB and higher. At high SNRs, both mixed terms are (exactly or near) zero and the amplitude and phase terms add up to the full capacity, as expected from (3.48) and (3.55).

As mentioned earlier, it is noteworthy that the complex Gaussian input, which maximizes $I(X;Y)$, does not maximize the single decomposition terms independently. The amplitude term $I(X_{\shortparallel};Y_{\shortparallel})$, for instance, is maximized by a "half-Gaussian" rather than a Rayleigh distribution at large SNRs, see Section 3.6.3.

3.4.2 Phase-modulated input

The terms *constant-intensity*, *constant-envelope* or *ring* modulation are used in the literature to characterize the input of a system which encodes information only in the phase of the transmitted signal. Results on the capacity of constant-intensity channels in the presence of AWGN have been reported over a period of 50 years, e. g. [Bla53, Bla87, Gei90, AB93, HK02]. The capacity of a channel constrained to constant intensity ("continuous PSK") is an upper limit on the rates achievable with discrete PSK constellations.

An important detail in the definition of phase-modulated AWGN channels is whether the receiver has access to amplitude and phase of the received signal or to the phase only. Although it has been observed [AB93] that both capacities are equal in the limit of large SNRs, evaluating the capacity difference at lower SNR values has remained an open problem.

Performing a polar decomposition (3.40) of the phase-modulated AWGN channel is the key to shed light on this question. As no information is encoded in $X_{\shortparallel}=\sqrt{P_s}=\text{const.}$, the amplitude term and the mixed term I of the decomposition equal zero.

As expected for a phase-modulated system, the phase term conveys the greatest share of the transmitted information. In the absence of amplitude modulation, this term can be written as $I(X_{\sphericalangle};Y_{\sphericalangle}|X_{\shortparallel}=\sqrt{P_s}) = h(Y_{\sphericalangle}) - h(Y_{\sphericalangle}|X_{\sphericalangle})$. The capacity-achieving input distribution is uniform in $[-\pi,\pi)$ [Bla87]. Hence, $Y_{\sphericalangle}$ is uniformly distributed, too, and $h(Y_{\sphericalangle}) = \log_2(2\pi)$. To calculate $h(Y_{\sphericalangle}|X_{\sphericalangle})$, the entropy integral has to be solved for the conditional phase PDF (3.51). An asymptotic approximation can be found for large SNRs, where (3.51) can be replaced by its Gaussian approximation (3.52). The conditional differential entropy $h(Y_{\sphericalangle}|X_{\sphericalangle})$ then approaches (3.53), and the decomposition phase term becomes

$$\begin{aligned} I(X_{\sphericalangle};Y_{\sphericalangle}|X_{\shortparallel}=\sqrt{P_s}) &= h(Y_{\sphericalangle}) - h(Y_{\sphericalangle}|X_{\sphericalangle}) \\ &\approx \log_2(2\pi) - \frac{1}{2}\cdot\log_2\left(2\pi e\cdot\frac{\sigma_n^2}{P_s}\right) \\ &= \frac{1}{2}\cdot\log_2\left(\frac{4\pi}{e}\cdot\frac{P_s}{2\sigma_n^2}\right) \\ &\approx \frac{1}{2}\cdot\log_2\frac{P_s}{2\sigma_n^2} + 1.1 \text{ bits}, \quad P_s \gg 2\sigma_n^2. \end{aligned} \tag{3.59}$$

Hence, the capacity of the phase-modulated AWGN channel is approximately 1.1 bits/symbol larger than half that of the AWGN channel with Gaussian input for large SNRs.

Finally, mixed term II $I(X_{\sphericalangle};Y_{\shortparallel}|X_{\shortparallel}=\sqrt{P_s},Y_{\sphericalangle})$ represents the (small) amount of information that can be gained by receiving the signal amplitude and phase rather than the

phase only. Figure 3.4 shows the decomposition terms as a function of SNR; the phase term markers indicate the asymptotic approximation (3.59), which is accurate at SNRs greater than 15 dB.

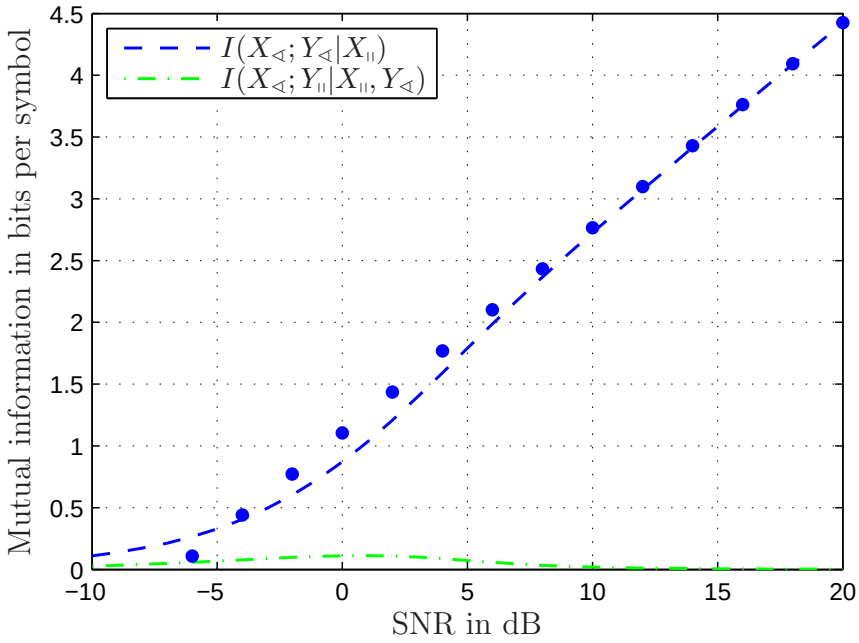

Figure 3.4: Mutual information decomposition terms as a function of SNR in dB, $10\log_{10}(P_s/(2\sigma_n^2))$, for the AWGN channel with constant-intensity (continuous ring) input. Lines show numerical results, markers correspond to analytical approximation (3.59).

Ho and Kahn [HK02] have used the constant-intensity channel model to calculate the effect of four-wave mixing on the capacity of phase-modulated optical WDM systems. However, since chromatic dispersion converts phase modulation into amplitude modulation in optical fibers, this approach is limited to dispersion-shifted fibers (cf. Section 6.1).

3.4.3 Discrete input constellations

In practical communication systems, the input consists of points from a discrete alphabet rather than of continuous values. Performing the polar decomposition for these discrete inputs is useful in two ways:

- ▷ The decomposition can help to adapt constellations to certain channel characteristics. For example, it may be beneficial for channels impaired by strong phase noise to re-arrange the points of a constellation in a way that the amplitude term is increased at the expense of the phase term. While the overall capacity may be hardly affected in the absence of phase noise, an increased capacity is obtained in the presence of phase noise. An example for this situation can be found in [BRHF02], where 8-PSK is compared to 8-OOK-PSK (7-PSK plus a point at the origin) and 8-star-QAM in the presence of fading and phase noise. The decomposition could

help to accelerate this search for good constellations and possibly to make it more systematic.

▷ When determining the mutual information numerically, the computational complexity can be significantly reduced by calculating the amplitude and phase terms rather than the full mutual information. This approach requires both mixed terms to be negligibly small.

In the following, decomposition results are given for some exemplary modulation schemes.

Modulation using one degree of freedom

As examples of modulation schemes where either amplitude or phase are modulated, Figure 3.5 shows the decomposition of *on-off keying* (OOK), i. e. $X \in \{0, 1\}$, and *phase-shift keying* (PSK) with $M = 16$ phase levels.

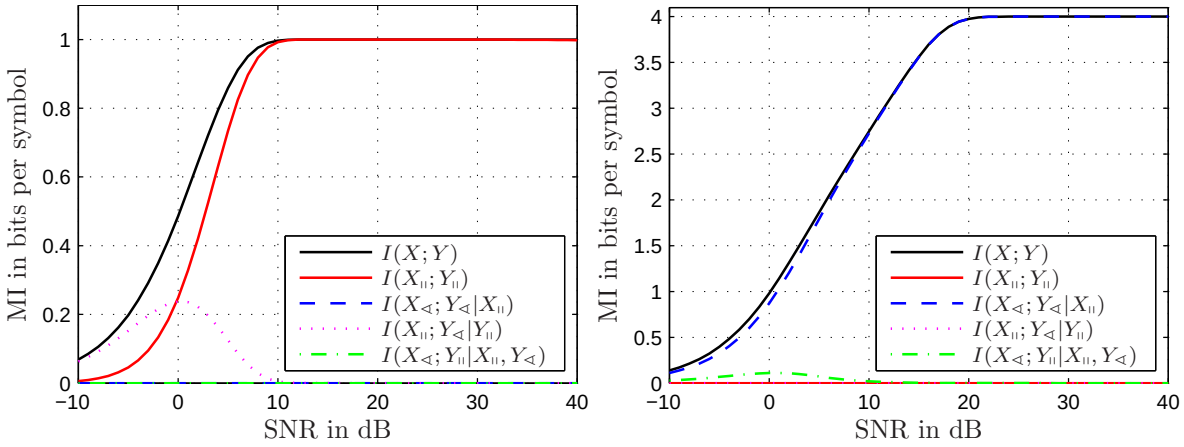

Figure 3.5: Polar decomposition of mutual information (MI) for OOK (left) and 16-PSK (right).

As the input phase carries no information with OOK, the phase term and the mixed term II are zero. The amplitude term yields the amount of information available when only the signal amplitude is received and processed. An example for such a system is the direct-detection receiver used in optical communication systems, where the photodiode responds to the incident light power [Agr02]. Receivers that have access to the full signal (amplitude and phase) can extract additional information about the input amplitude from the output phase. This information gain is reflected in the mixed term I (dotted line). In the optical communications example, this gain can be obtained by upgrading an optical OOK system from direct to coherent detection. At SNRs larger than 10 dB, all the information is contained in the received amplitude, so that receiving the signal phase does not yield any additional information.

For the PSK input, the amplitude term and the mixed term I are zero. A phase-only receiver captures most of the available information; the (rather small) gain that is obtained

from additional amplitude reception is mirrored in the mixed term II (dash-dotted line).

Combined ASK-PSK modulation

The simultaneous digital modulation of both amplitude and phase was first proposed in 1960 [Cah60]. Examples for this type of constellation, which later became known as *Type I* or *star-QAM* constellation, are shown in Figure 3.6.[2] The constellations depicted in the left column are combinations of 4 amplitude levels and 4, 8 and 16 phase levels, respectively. The constellations shown in the right column are modifications of these ASK/PSK schemes, where an additional phase offset was introduced between adjacent amplitude levels, thus increasing the minimum distance between neighboring constellation points.

The decomposition results shown in Figure 3.7 illustrate the capacity gain obtained from the phase offset. As the joint amplitude PDF $p(x_{\shortparallel}, y_{\shortparallel})$ remains unaffected by the phase offset, the amplitude term (red line) is equal for both constellations (compare plots on the left and on the right side of Figure 3.7). Similarly, the conditional joint phase PDF $p(x_{\triangleleft}, y_{\triangleleft}|x_{\shortparallel})$ only experiences a constant shift for amplitude levels with phase offset, which does not change the decomposition phase term (blue line). The capacity gain achieved by the phase offset is reflected in the increase of the mixed term I (magenta line; cf. top left and top right plots); this gain decreases with increasing number of phase levels (top to bottom).

By letting the number of phase levels go to infinity, the constellation turns into continuous concentric rings and the mixed term I tends towards zero. Such modulation schemes with a discrete number of amplitude levels and continuous phase angles (so-called *ring modulation*) were used in an extensive numerical study to estimate the channel capacity of fiber-optic channels [EKW+10] (see Section 6.2.2). As for the constellations discussed above (and most other constellations), the mixed term II (green line) is negligibly small but non-zero for ASK/PSK constellations.

QAM

The polar decomposition results for M-QAM constellations with $M = 4, 16, 64, 256, 512, 1024$ are shown in Figure 3.8. It can be seen that the amplitude and phase terms saturate at $H(X_{\shortparallel})$ (3.3) and $H(X_{\triangleleft}|X_{\shortparallel})$ (3.7), respectively. For instance, 16-QAM has three distinct amplitude levels with four or eight distinct phase levels each, so the decomposition terms tend towards

$$H(X_{\shortparallel}) = \frac{2}{4}\log_2 4 + \frac{1}{2}\log_2 2 = 1.5 \text{ bits} \tag{3.60}$$

and

$$H(X_{\triangleleft}|X_{\shortparallel}) = \frac{2}{4}\cdot \log_2 4 + \frac{1}{2}\log_2 8 = 2.5 \text{ bits}. \tag{3.61}$$

[2] A concise summary of the history of digital modulation schemes can be found in [HWK00]. For a comprehensive discussion of the capacities and error rates of various discrete modulation schemes on the AWGN channel, the reader is referred to [EKW+10, App. C] and [KAvWW03].

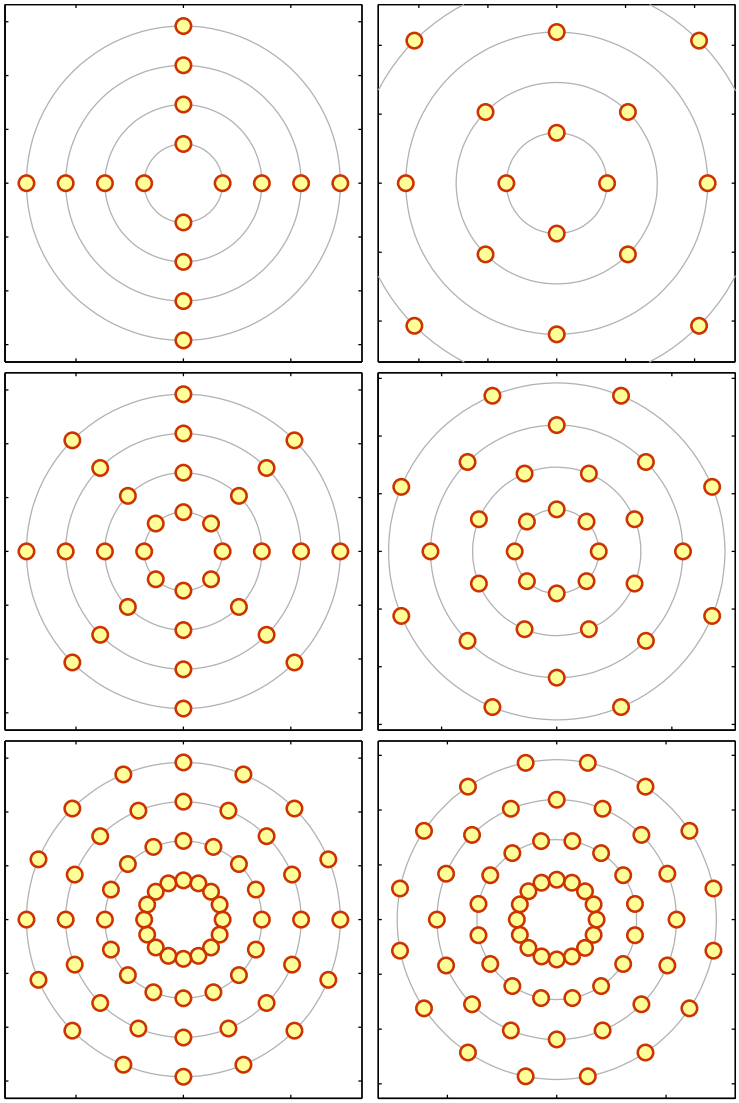

Figure 3.6: Combined ASK/PSK signal constellations with 4 phase levels (top), 8 phase levels (center) and 16 phase levels (bottom) without (left) and with (right) phase shift. The number of amplitude levels is always 4.

Among the considered QAM constellations, 4-QAM is a special case in the sense that its mixed term I is zero; being a PSK format, its decomposition resembles that of 16-PSK depicted in Figure 3.5. For $M > 4$, QAM constellations exhibit a significant mixed term I, so that in the analysis of this modulation scheme, the mutual information may not be approximated by the sum of the amplitude and phase terms only. Again, mixed term II is non-zero but negligibly small for all QAM constellations.

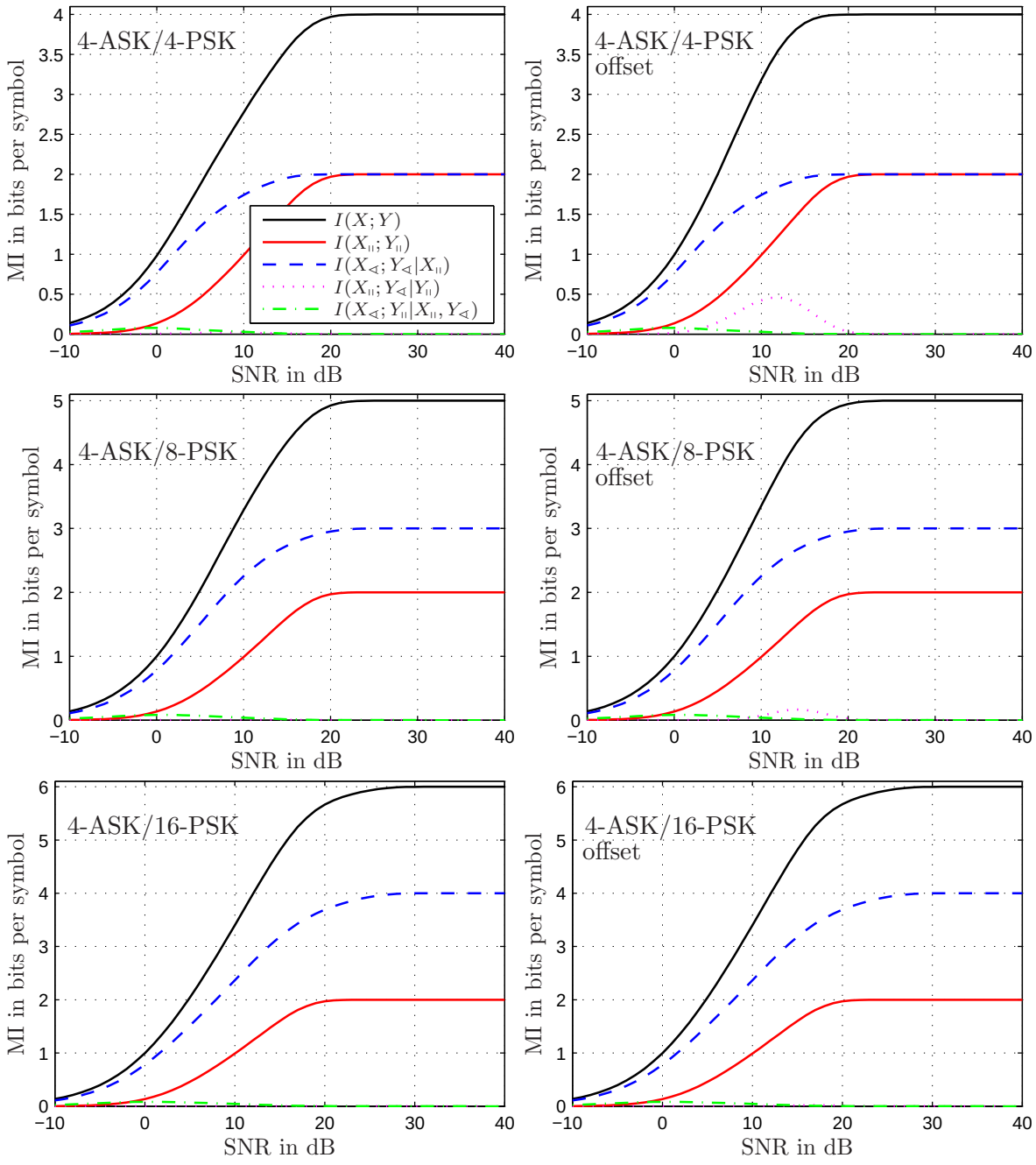

Figure 3.7: Polar decomposition of mutual information (MI) for ASK/PSK constellations depicted in Figure 3.6 without phase offset (left) and with phase offset (right).

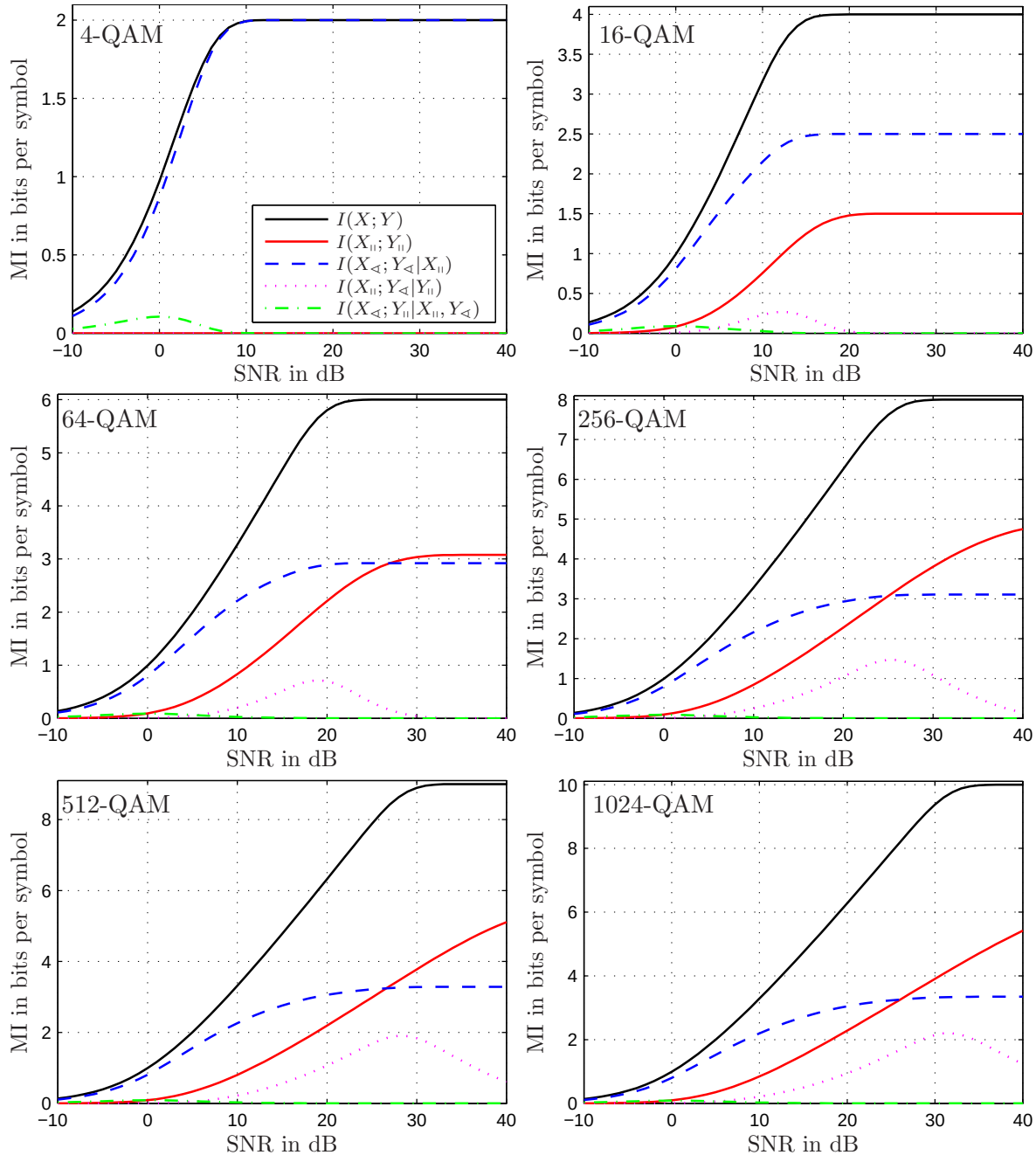

Figure 3.8: Polar decomposition of mutual information (MI) for M-QAM constellations with (from top left to bottom right) $M = 4, 16, 64, 256, 512, 1024$.

3.5 Introduction to directional statistics

This section reviews results from directional statistics that are useful for understanding phase noise and other circular random processes. Random variables such as phase angles or points on a spherical surface cannot be treated with "conventional" statistical methods. (E.g., the average wind direction calculated from two measurements of 358° and 2° is not 180°.) The field that deals with such *directional* (in contrast to *linear*) random variables is known as *directional statistics* [Mar72].

3.5.1 Trigonometric moments

We restrict our review to one-dimensional directional (or circular) random variables, e. g. phase angles. Such a random variable Θ is defined on an arbitrary interval of length 2π and has a periodic probability density function (PDF) that satisfies

$$\int_{c-\pi}^{c+\pi} p(\theta)\mathrm{d}\theta = 1, \quad c \in \mathbb{R}. \tag{3.62}$$

To ensure that the statistical moments of the directional random variable are invariant under a rotation of the coordinate system, the *trigonometric moments* are calculated from $e^{\jmath\,\Theta}$ rather than from Θ. The i^{th} trigonometric moment $m^{\circ}_{\Theta,i}$ of Θ is defined as [Fis96]

$$m^{\circ}_{\Theta,i} = \int_{-\pi}^{\pi} \left(e^{\jmath\,\theta}\right)^i p(\theta)\mathrm{d}\theta. \tag{3.63}$$

The first trigonometric moment can be calculated as

$$m^{\circ}_{\Theta,1} = \int_{-\pi}^{\pi} e^{\jmath\,\theta} p(\theta)\mathrm{d}\theta = \rho^{\circ}_{\Theta} \cdot e^{\jmath\,\mu^{\circ}_{\Theta}}, \tag{3.64}$$

where ρ°_{Θ} is the *resultant length* and μ°_{Θ} is the *mean direction* of Θ [Mar72]. The i^{th} *central trigonometric moment* is calculated as the i^{th} trigonometric moment of $\Theta - \mu^{\circ}_{\Theta}$.

To quantify the *concentration* (or, inversely, the *dispersion*) of a circular random variable Θ, it is common to define the *circular variance* as [Mar72, Fis96]

$$V^{\circ}_{\Theta} = 1 - \left|\mathcal{E}\left\{e^{\jmath\,\Theta}\right\}\right| = 1 - \rho^{\circ}_{\Theta}. \tag{3.65}$$

Clearly, the circular variance is maximized if Θ is uniformly distributed ($V^{\circ}_{\Theta} = 1$) and minimized for a constant Θ ($V^{\circ}_{\Theta} = 0$). It must be noted that the *circular standard deviation* is *not* defined as $\sqrt{V^{\circ}_{\Theta}}$, but as [Fis96]

$$\sigma^{\circ}_{\Theta} = \sqrt{-2\ln(1 - V^{\circ}_{\Theta})} = \sqrt{-2\ln\rho^{\circ}_{\Theta}}. \tag{3.66}$$

3.5.2 Circular distributions

An example for a circular distribution has been introduced above in (3.51), which describes the probability density of the phase angle of a complex phasor corrupted by complex-valued AWGN. This distribution ranges from a uniform distribution (in any 2π interval) for small SNRs to a Gaussian distribution for large SNRs. Middleton gives a series expansion of (3.51) [Mid60, § 9.2-2] which has been applied in the context of systems with phase noise (cf. references given in [Ho05a, Appendix 4.A]).

Wrapped Gaussian distribution

Another important circular distribution is the *wrapped Gaussian* distribution [KJ82, Bah06]:

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \sum_{k=-\infty}^{\infty} \exp\left(-\frac{(\theta - \mu - 2\pi k)^2}{2\sigma^2}\right). \tag{3.67}$$

This distribution occurs when a linear random variable $X \sim \mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$ is "wrapped" around a circle, i. e. $\Theta = X \mod 2\pi$. Various phase noise processes (generated by SPM, XPM and XPM$^{\mathcal{S}}$) with a wrapped Gaussian distribution are introduced and discussed in Chapters 5 and 6.

The mean direction μ_Θ°, resultant length ρ_Θ° and circular variance V_Θ° of a wrapped Gaussian random variable can be calculated as [Bah06]

$$\mu_\Theta^\circ = \mu \mod 2\pi, \quad \rho_\Theta^\circ = e^{-\frac{1}{2}\sigma^2} \tag{3.68}$$

and

$$V_\Theta^\circ = 1 - e^{-\frac{1}{2}\sigma^2}. \tag{3.69}$$

The wrapped Gaussian approaches a uniform distribution for large σ and can be approximated by a Gaussian distribution for small σ as shown in Figure 3.9.

Von Mises distribution

While the wrapped Gaussian distribution shares some of the properties of the linear Gaussian distribution [vBv$^+$08], it does not maximize the entropy for a given (circular) variance. This condition is met by the *von Mises* distribution (after Austrian mathematician R. von Mises) [KJ82, Fis96]

$$p(\theta) = \frac{\exp(\kappa \cos(\theta - \mu))}{2\pi\, \mathrm{I}_0(\kappa)}, \tag{3.70}$$

where μ is the circular mean (and is usually called the *centrality parameter*), κ is the *concentration parameter* and $\mathrm{I}_0(.)$ is the modified Bessel function of the first kind with order zero. In engineering, the von Mises distribution is known as the *Tikhonov* distribution (after V. I. Tikhonov) [dA07]; it appears in the description of the phase error of phase-locked loops [Vit63].

The circular variance is calculated using (3.65) with (3.64) as

$$\begin{aligned} V_\Theta^\circ &= 1 - \rho_\Theta^\circ \\ &= 1 - \left| \int_{-\pi}^{\pi} p(\theta) \cdot e^{\jmath\,\theta} \mathrm{d}\theta \right| \\ &= 1 - \frac{1}{2\pi\,\mathrm{I}_0(\kappa)} \left| \int_{-\pi}^{\pi} e^{\kappa\cos\theta} (\cos\theta + \jmath\,\sin\theta) \mathrm{d}\theta \right| \\ &= 1 - \frac{1}{\pi\,\mathrm{I}_0(\kappa)} \int_0^{\pi} e^{\kappa\cos\theta} \cos\theta \mathrm{d}\theta = 1 - \frac{\mathrm{I}_1(\kappa)}{\mathrm{I}_0(\kappa)}. \end{aligned} \tag{3.71}$$

To obtain (3.71), we use the modified Bessel functions of the first kind of order n defined as (see [AS72])

$$\mathrm{I}_n(\kappa) = \frac{1}{\pi} \int_0^{\pi} e^{\kappa\cos x} \cos(nx) \mathrm{d}x. \tag{3.72}$$

The differential entropy is calculated as

$$\begin{aligned} h(\Theta) &= \int_{-\pi}^{\pi} \frac{e^{\kappa\cos\theta}}{2\pi\,\mathrm{I}_0(\kappa)} \ln \frac{2\pi\,\mathrm{I}_0(\kappa)}{e^{\kappa\cos\theta}} \mathrm{d}\theta \\ &= \frac{\ln(2\pi\,\mathrm{I}_0(\kappa))}{2\pi\,\mathrm{I}_0(\kappa)} \cdot \int_{-\pi}^{\pi} e^{\kappa\cos\theta} \mathrm{d}\theta - \frac{k}{2\pi\,\mathrm{I}_0(\kappa)} \cdot \int_{-\pi}^{\pi} e^{\kappa\cos\theta} \cos\theta \mathrm{d}\theta \\ &= \ln(2\pi\,\mathrm{I}_0(\kappa)) - k \cdot \frac{\mathrm{I}_1(\kappa)}{\mathrm{I}_0(\kappa)}, \end{aligned} \tag{3.73}$$

where (3.72) was used twice in the last equality.

Among all linear distributions that satisfy an average power (or variance) constraint $\mathcal{E}\{|X|^2\} \leq P_s$, the Gaussian distribution maximizes the differential entropy $h(X)$ [CT91]. Similarly, one can ask for the circular distribution $p(\theta)$ that maximizes $h(\Theta)$ under a circular variance constraint $V_\Theta^\circ \leq A$. Without loss of generality, we assume $\mu_\Theta^\circ = 0$ which means that $\mathcal{E}\left\{e^{\jmath\,\Theta}\right\}$ is a non-negative real number and that $\mathcal{E}\{\sin\Theta\} = 0$. We can thus write the circular variance constraint as

$$\begin{aligned} V_\Theta^\circ &\overset{(3.65)}{=} 1 - |\mathcal{E}\left\{e^{\jmath\,\Theta}\right\}| \\ &= 1 - \int_{-\pi}^{\pi} p(\theta)\cos\theta \mathrm{d}\theta - \jmath \underbrace{\int_{-\pi}^{\pi} p(\theta)\sin\theta \mathrm{d}\theta}_{=0} \\ &= 1 - \mathcal{E}\{\cos\Theta\} \leq A. \end{aligned} \tag{3.74}$$

To prove that the von Mises distribution (3.70) maximizes the differential entropy under the circular variance constraint (3.74), we calculate the Kullback-Leibler distance (3.17)

between the von Mises distribution $p(\theta)$ and an arbitrary other distribution $q(\theta)$:

$$\begin{aligned} D(q\|p) &= \int_{-\pi}^{\pi} q(\theta) \ln \frac{q(\theta)}{p(\theta)} \mathrm{d}\theta \\ &= \underbrace{\int_{-\pi}^{\pi} q(\theta) \ln q(\theta) \mathrm{d}\theta}_{-h(q)} - \int_{-\pi}^{\pi} q(\theta) \ln p(\theta) \mathrm{d}\theta \\ &= -h(q) - \int_{-\pi}^{\pi} q(\theta) \ln \frac{e^{\kappa \cos\theta}}{2\pi \, \mathrm{I}_0(\kappa)} \mathrm{d}\theta \\ &= -h(q) + \ln(2\pi \, \mathrm{I}_0(\kappa)) - \kappa \cdot \int_{-\pi}^{\pi} q(\theta) \cos\theta \mathrm{d}\theta \\ &= -h(q) + \ln(2\pi \, \mathrm{I}_0(\kappa)) - \kappa \cdot \underbrace{\mathcal{E}_{\Theta \sim q(\theta)} \{\cos \Theta\}}_{\geq 1 - A} \\ &\leq -h(q) + h(p), \end{aligned} \tag{3.75}$$

where $h(q)$ denotes the differential entropy $h(\Theta)$ of a random variable $\Theta \sim q(\theta)$ and where κ is chosen to satisfy $1 - A = \mathrm{I}_1(\kappa)/\mathrm{I}_0(\kappa)$. Recall that $D(q\|p) \geq 0$ with equality if and only if $p = q$ [CT91]. Hence, we find that

$$h(p) \geq h(q) \tag{3.76}$$

with equality if and only if $q = p$.

A different path to get to the same result is to note that the von Mises distribution is a special case of the maximum entropy distribution (3.21). With the constraint (3.74), the maximum entropy distribution with coefficients $\lambda_0 = -\ln(2\pi \, \mathrm{I}_0(\kappa))$ and $\lambda_1 = \kappa$ transforms into (3.70). Barakat finds the same result using Lagrange multipliers [Bar87]. Observe that the von Mises distribution becomes uniform for large circular variance (small κ) and approaches a Gaussian distribution with variance $\sigma^2 = 1/\kappa$ when the circular variance is small (κ large) [KJ82]. Figure 3.9 shows the wrapped Gaussian PDF for $\mu = 0$ (i. e. $\mu_\Theta^\circ = 0$) and various values of σ.

Because of its maximum entropy property, the von Mises distribution is often considered to be the circular analogue of the linear normal distribution. Hence, it is sometimes referred to as the *circular normal distribution*; to avoid confusion with the wrapped Gaussian distribution, it is advisable not to use this term. The wrapped Gaussian and the von Mises distribution have a very similar shape [Bar87], see Figure 3.9. In practice, one often uses whichever is more convenient [Fis96].

Truncated Gaussian distribution

Suppose now for the sake of argument that the phase constraint is the usual second-order constraint $\mathcal{E}\{\Theta^2\} \leq A$, where the expectation is performed over the interval $[-\pi, \pi)$. Suppose further that we wish to maximize the entropy PDF over all PDFs with $\mathcal{E}\{\Theta\} = 0$ (the latter constraint is made to simplify the discussion). Consider the *truncated* Gaussian

distribution

$$p(\theta) = \frac{\lambda}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\theta^2}{2\sigma^2}\right), \quad -\pi \leq \theta < \pi, \tag{3.77}$$

where λ is a scaling constant that ensures (3.62) is valid, and σ^2 is chosen so that $\mathcal{E}\{\Theta^2\} = A$. We compute

$$\begin{aligned} h(\Theta) &= \int_{-\pi}^{\pi} -p(\theta) \ln p(\theta) \mathrm{d}\theta \\ &= \frac{1}{2} \ln\left(\frac{2\pi\sigma^2}{\lambda^2}\right) + \frac{1}{2\sigma^2} \cdot \underbrace{\int_{-\pi}^{\pi} p(\theta)\theta^2 \mathrm{d}\theta}_{=\mathcal{E}\{\Theta^2\}=A} \\ &= \frac{1}{2} \ln\left(\frac{2\pi\sigma^2}{\lambda^2}\right) + \frac{A}{2\sigma^2}. \end{aligned} \tag{3.78}$$

We further have

$$\begin{aligned} D(q\|p) &= \int_{-\pi}^{\pi} q(\theta) \ln \frac{q(\theta)}{p(\theta)} \mathrm{d}\theta \\ &= -h(q) - \int_{-\pi}^{\pi} q(\theta) \ln p(\theta) \mathrm{d}\theta \\ &= -h(q) + \frac{1}{2} \ln\left(\frac{2\pi\sigma^2}{\lambda^2}\right) + \frac{1}{2\sigma^2} \cdot \int_{-\pi}^{\pi} q(\theta)\theta^2 \mathrm{d}\theta \\ &= -h(q) + \frac{1}{2} \ln\left(\frac{2\pi\sigma^2}{\lambda^2}\right) + \frac{1}{2\sigma^2} \cdot \underbrace{\mathcal{E}_{\Theta\sim q(\theta)}\left\{\Theta^2\right\}}_{\leq A} \\ &\leq -h(q) + h(p). \end{aligned} \tag{3.79}$$

Using $D(q\|p) \geq 0$ with equality if and only if $q = p$, we find that a truncated Gaussian distribution maximizes entropy.

Figure 3.9 shows the PDFs for the truncated Gaussian distribution for $\mathcal{E}\{\Theta\} = 0$ and various values of σ (wrapped and truncated Gaussians) and $\kappa = 1/\sigma^2$ (von Mises). We remark that the physical meaning of our second-order constraint is unclear, but the same can be said for the circular variance constraint. It is interesting, however, that maximum entropy considerations lead to either a von Mises distribution or a truncated Gaussian distribution. Two interesting problems are whether the wrapped Gaussian distribution is maximum-entropy under some natural circular constraint, and whether the wrapped Gaussian has other natural "normal" properties [vBv+08].

3.6 Partially coherent channels

In Section 3.4, the transmitted phase was corrupted by AWGN. If the signal is impaired by phase noise (in addition to AWGN), the channel is only partially able to convey

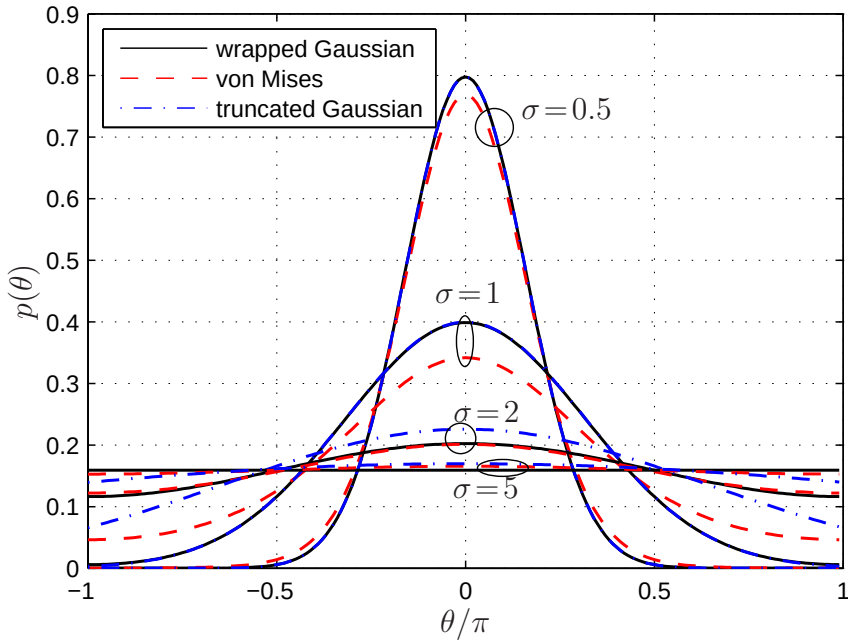

Figure 3.9: Wrapped and truncated Gaussian and von Mises PDFs with $\mu = 0$ and various values of σ and $\kappa = 1/\sigma^2$.

phase information even in the absence of AWGN. Such channels are called *partially coherent* [KSS04].[3]

Because of their phase noise impairment, partially coherent channels suggest themselves as a natural application of the polar decomposition method. The earliest information-theoretic results on channels with reduced degrees of freedom, e. g. transmitters or receivers that are limited to amplitude modulation (AM) or phase modulation (PM), date back to 1953 [Bla53]. Some time later, partially coherent channels became an important research topic in the context of phase jitter induced by phase demodulation [Vit63]. Good modulation schemes for such channels were presented in [FGW73]. To this date, little is known about the capacity-achieving input for partially coherent channels [KSS04].

Partially coherent channels can be described in continuous-time form by

$$y(t) = x(t) \cdot e^{\jmath\theta(t)} + n(t), \tag{3.80}$$

where $n(t) \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma_n^2)$ is a complex-valued AWGN process and $\theta(t)$ models the phase noise process.[4] We can differentiate various types of phase noise appearing in communication systems:

[3]The term *partially coherent* was introduced to communications engineering by A. Viterbi in 1965 [Vit65]. Viterbi possibly adopted the term from physical optics, where it characterizes the temporal or spatial correlation of electrical fields that are neither *coherent* (fully correlated) nor *incoherent* (uncorrelated) [BW99]. In communication and information theory, the term *noncoherent* (rather than *incoherent*) is used to refer to channels that are entirely unable to transmit any phase information.

[4]For the sake of consistency with the convention used in this thesis, continuous-time signals are denoted by small letters (e. g. $x(t)$) in this section, even though they denote random variables.

- ▷ The carrier itself as well as the local oscillator used for demodulation can have random noise fluctuations. This type of phase noise is particularly relevant in lightwave communication systems, where the laser phase performs a random walk (Brownian motion). The nonzero laser linewidth broadens the signal spectrum, so that spectrally sensitive operations (filtering, sampling) require special attention. "Classical" references on laser phase noise system aspects include [Sal85, FGV88, FVG89, SS89, DSS91] and many references therein.

- ▷ Another type of correlated phase noise emerges in communication systems where the carrier phase is imperfectly tracked at the receiver (e. g. in a phase-locked loop [FGW73, KSS04]). In this case, samples from the the phase noise process $\theta(t)$ are usually assumed to have a von Mises (Tikhonov) distribution (3.70).

- ▷ Uncorrelated (white) phase noise can be used to model the nonlinear effect of cross-phase modulation (XPM$^{\mathcal{S}}$) in multi-channel fiber-optic communication systems. In this case, the phase noise samples follow a wrapped Gaussian distribution (3.67) as explained in Section 6.2.2.

- ▷ Signal-dependent phase noise is produced in fiber-optic communication systems by the nonlinear effect of self-phase modulation (SPM$^{\mathcal{S}}$), see Section 6.2.1.

In general, all types of phase noise are capable of broadening the spectrum of the transmitted signal $x(t)$ [Ho05a]. This spectral broadening is the major obstacle in transforming (3.80) into a discrete-time channel model. Filtering (and sampling) a signal whose spectrum is broadened by phase noise can result (1) in signal distortions and energy loss when the filter is narrow [FGV88, DSS91] and (2) in an increased captured noise power when the filter bandwidth is wide (see [DSS91] and references therein). These effects can be neglected when the spectral broadening is moderate, which is the case for strongly correlated phase noise processes. Filtering and sampling at the symbol rate is then possible and leads to discrete-time channel models that have independent and identically distributed (i. i. d.) signal and noise samples, but correlated phase noise samples (see e. g. [PSSG00]). To obtain a discrete-time channel model with uncorrelated phase noise samples, the presence of an ideal interleaver and de-interleaver can be assumed (e. g., [KSS04]). It is then possible to transform (3.80) into the discrete-time form

$$Y = X \cdot e^{\jmath\,\Theta} + N, \tag{3.81}$$

in which the phase noise time samples Θ_i are assumed to be i. i. d. and statistically independent of X. In Section 3.6.2, we discuss partially coherent channels with white phase noise. There is no spectral broadening on these channels, so discretization by filtering and sampling is possible. However, the discrete channel model must be modified to account for an effect we call *spectral loss*.

Since the phase angle of AWGN is uniformly distributed, the order in which phase noise

and AWGN act on the transmitted signal is irrelevant:

$$\begin{aligned} Y &= (X+N)\cdot e^{\jmath\,\Theta} \\ &= X\cdot e^{\jmath\,\Theta} + N\cdot e^{\jmath\,\Theta} \\ &= X\cdot e^{\jmath\,\Theta} + N', \end{aligned} \tag{3.82}$$

where $N' \sim \mathcal{N}_{\mathbb{C}}\left(0, 2\sigma_n^2\right)$ has the same distribution as N.

The circular PDF $p(Y_{\sphericalangle})$ can be obtained by circular convolution [GG95] of (3.51) with $p(\Theta)$. In numerical experiments, it is usually more efficient to multiply the PDFs' discrete Fourier transforms (DFT) and perform an inverse DFT (IDFT) to obtain the final result [JBS00]. In particular, when the phase noise has a wrapped Gaussian distribution, the DFT of (3.51) can be multiplied with the DFT of the "unwrapped" Gaussian (which is again Gaussian). The following IDFT will implicitly "wrap" the resulting PDF so that it maintains its periodicity with 2π.

In the following discussion of partially coherent channels, the term *SNR* refers to the power ratio of signal and *additive* noise, $P_s/(2\sigma_n^2)$.

3.6.1 Input optimization and information rate calculation

The capacity-achieving input distribution for the partially coherent channel (3.81) is not Gaussian [HBF02]. It has been shown that the optimal distribution is circularly symmetric [HBF02], i. e. uniform in phase, and has discrete amplitude levels [HBF03, KSS04]. In other words, the capacity-achieving input distribution for the partially coherent channel consists of a number of continuous rings; the number, radii and probabilities of these rings are subject to optimization. Interestingly, the shaping gain that can be achieved by using non-equiprobable input symbols rather than a uniform distribution was shown to be significantly larger than the maximum shaping gain of 1.53 dB for the AWGN channel [HBF02]. Therefore, input optimization for the partially coherent channel may be more rewarding than for the AWGN channel, but remains an open problem.

The polar decomposition is very useful for the analysis of partially coherent channels when both mixed terms are small or can be neglected. As mentioned in Section 3.4, this is the case for AWGN channels with Gaussian or ring inputs. As the amplitude term $I(X_{\shortparallel};Y_{\shortparallel})$ is not affected by phase noise, it suffices to re-calculate the phase term in the presence of phase noise. The conditional phase PDF $p(y_{\sphericalangle}|x_{\shortparallel}, x_{\sphericalangle}=0)$ is obtained numerically or, where possible, analytically from a circular convolution [GG95] of (3.51) with the phase noise PDF, usually (3.67) or (3.70).

Figure 3.10 shows the decomposition results for the AWGN channel with Gaussian input with additional phase noise. The phase noise has a wrapped Gaussian distribution with parameter σ as shown in Figure 3.9.[5] The phase noise parameter values are $\sigma = 0, 0.5, 1, 2$.

[5] We remind the reader that σ_n^2 denotes the AWGN's variance per dimension, whereas σ is used as the parameter of the wrapped Gaussian distribution. Note that this distribution's circular variance is given by (3.69); it is not equal to σ^2.

For large σ, the circular variance goes to one and the wrapped Gaussian distribution becomes uniform. In this case, no information can be transmitted in the signal phase and the phase term tends to zero. An interesting observation can be made when σ is small (but nonzero). In this case, the phase term increases with increasing SNR, but tends towards a constant value asymptotically. At the point when the phase term reaches this asymptote, the SNR is so large that the effect of AWGN can be neglected and phase noise becomes the dominant impairment for the signal phase. As the effect of phase noise is independent of the SNR, increasing the SNR further does not bring any additional gain in the phase term. Therefore, as the SNR increases, the contribution to the total capacity $I(X;Y)$ of the phase term (which remains constant) becomes increasingly small compared to that of the amplitude term (which rises logarithmically with the SNR, cf. (3.48)). This statement is valid for any (arbitrarily low) phase noise variance. The phase noise variance determines the SNR at which the phase term contribution to the total capacity can be neglected. Figure 3.10 shows the amplitude term and the phase terms for $\sigma = 0, 0.5, 1, 2$ and the respective total capacities. The (very small) contribution of the mixed term II was neglected.

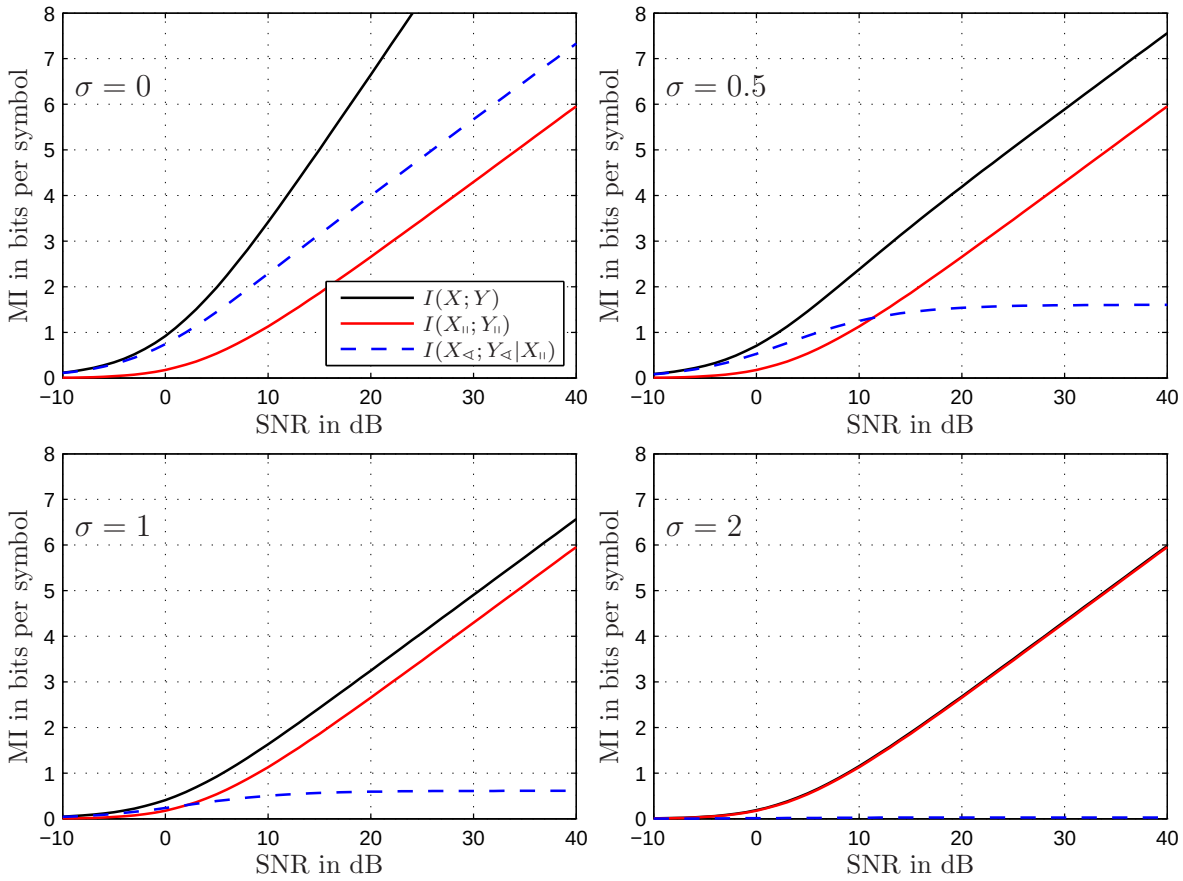

Figure 3.10: Polar decomposition of mutual information (MI) for an AWGN channel with Gaussian input with additional phase noise ($\sigma = 0, 0.5, 1, 2$). Mixed term II is negligible.

3.6.2 Spectral loss induced by white phase noise

As discussed above, certain types of phase noise induce spectral broadening. If the phase noise process $\theta(t)$ is white, i. e. if it is temporally uncorrelated, a related but qualitatively different effect occurs which we call *spectral loss*.

To describe this effect, we use the continuous-time channel model (3.80). We next derive the power spectral density (PSD) $\Phi_y(f)$ of $y(t)$. We assume that $x(t)$ and $n_{\mathrm{PN}}(t) = e^{\jmath\,\theta(t)}$ are stationary, ergodic, and statistically independent random processes. The autocorrelation function (ACF) $\varphi_y(\tau)$ of $y(t)$ is [OL02]

$$\begin{aligned}\varphi_y(\tau) &= \mathcal{E}\left\{x(t)\cdot e^{\jmath\,\theta(t)}\cdot x^*(t+\tau)\cdot e^{-\jmath\,\theta(t+\tau)}\right\} + \mathcal{E}\left\{n(t)\cdot n^*(t+\tau)\right\}\\ &= \mathcal{E}\left\{x(t)\cdot x^*(t+\tau)\right\}\cdot\mathcal{E}\left\{e^{\jmath\,\theta(t)}\cdot e^{-\jmath\,\theta(t+\tau)}\right\} + \varphi_n(\tau)\\ &= \varphi_x(\tau)\cdot\varphi_{n_{\mathrm{PN}}}(\tau) + \varphi_n(\tau),\end{aligned} \tag{3.83}$$

where $\mathcal{E}\{.\}$ denotes the ensemble average. In calculating the ACF $\varphi_{n_{\mathrm{PN}}}(\tau)$ of $n_{\mathrm{PN}}(t)$, we assume for simplicity that the phase noise follows a wrapped Gaussian distribution (3.67) with parameter σ. Since $\theta(t)$ and $\theta(t+\tau)$ are independent samples of a Gaussian random process, their sum or difference $\theta'(t) = \theta(t) \pm \theta(t+\tau) \sim \mathcal{N}_{\mathbb{R}}(0, 2\sigma^2)$ for $\tau \neq 0$. The autocorrelation function $\varphi_{n_{\mathrm{PN}}}(\tau)$ of the phase noise process $n_{\mathrm{PN}}(t)$ is

$$\begin{aligned}\varphi_{n_{\mathrm{PN}}}(\tau) &= \mathcal{E}\left\{e^{\jmath\,\theta(t)}\cdot e^{-\jmath\,\theta(t+\tau)}\right\}\\ &= \begin{cases} 1, & \tau = 0,\\ e^{-\sigma^2}, & \tau \neq 0,\end{cases}\end{aligned} \tag{3.84}$$

where the last result (for $\tau \neq 0$) is the resultant length (3.68) of an ergodic (wrapped) Gaussian random variable θ' with zero mean and variance $2\sigma^2$:

$$\mathcal{E}\left\{e^{\jmath\,(\theta(t)-\theta(t+\tau))}\right\} = \mathcal{E}\left\{e^{\jmath\,\theta'}\right\} = e^{-\sigma^2}, \quad \tau \neq 0. \tag{3.85}$$

The piecewise defined ACF (3.84) can be written as

$$\varphi_{n_{\mathrm{PN}}}(\tau) = e^{-\sigma^2} + \lim_{B\to\infty}(1 - e^{-\sigma^2})\cdot\mathrm{sinc}\,(B\tau)\,, \tag{3.86}$$

where $\mathrm{sinc}\,(x) = \sin(\pi x)/(\pi x)$. By the Wiener-Khinchin theorem [OL02], the PSD $\Phi_{n_{\mathrm{PN}}}(f)$ of $n_{\mathrm{PN}}(t)$ is

$$\begin{aligned}\Phi_{n_{\mathrm{PN}}}(f) &= \mathcal{F}\left(\varphi_{n_{\mathrm{PN}}}(\tau)\right)\\ &= e^{-\sigma^2}\delta(f) + \lim_{B\to\infty}(1 - e^{-\sigma^2})\cdot\frac{1}{B}\cdot\mathrm{rect}\left(\frac{f}{B}\right),\end{aligned} \tag{3.87}$$

where

$$\mathrm{rect}\,(f) = \left\{\begin{array}{ll} 1, & |f| < \frac{1}{2},\\ \frac{1}{2}, & |f| = \frac{1}{2},\\ 0, & |f| > \frac{1}{2}\end{array}\right\} = \mathcal{F}\left(\mathrm{sinc}\,(t)\right) \tag{3.88}$$

is the *rectangular function* [GG95]. Finally, the PSD $\Phi_y(f)$ of $y(t)$ is calculated using (3.83) and (3.87) as

$$\begin{aligned}\Phi_y(f) &= \Phi_x(f) \star \Phi_{n_{\mathrm{PN}}}(f) + \Phi_n(f) \\ &= e^{-\sigma^2}\,\Phi_x(f) + \lim_{B\to\infty} \Phi_x(f) \star (1 - e^{-\sigma^2}) \frac{\mathrm{rect}\,(f/B)}{B} + \Phi_n(f). \end{aligned} \tag{3.89}$$

The $\star$ sign denotes convolution. Equation (3.89) explains the spectral effect of phase noise: The original PSD $\Phi_x(f)$ is preserved in shape, but attenuated by a factor $e^{-\sigma^2}$. The remaining signal power (a fraction of $1 - e^{-\sigma^2}$) is spread over the entire spectrum from $-\infty$ to $+\infty$ through convolution with a rectangular function of infinite width and zero height. Because the fraction of power that leaks outside the original spectrum has arbitrarily low power in any given finite band, it does not appear as spectral interference. Hence, the effect is called *spectral loss* (in contrast to *spectral broadening*).

A remarkable feature of (3.89) is its simplicity: the original PSD $\Phi_x(f)$ is not broadened. One can therefore discretize the partially coherent channel with white Gaussian phase noise by means of filtering and sampling at the symbol rate. However, (3.81) must be adapted to account for spectral loss. The complete discrete-time model for this channel can be written as

$$Y = X \cdot e^{-\frac{1}{2}\cdot\sigma^2 + \jmath\,\Theta} + N, \tag{3.90}$$

where the factor $1/2$ accounts for the amplitude attenuation.

If the phase noise distribution is not (wrapped) Gaussian, the same calculation will lead to qualitatively similar results, with the value of the ACF (3.84) at $\tau \neq 0$ determining the spectral loss factor.

For numerical simulations of phase noise, (3.89) has an important implication. Due to the spectral loss effect, the output signal $y(t)$ has infinite bandwidth and is therefore generally undersampled in numerical simulations with finite bandwidth. Inevitably, the numerical simulation of phase noise will create aliasing inside and outside the original signal band through convolution of $\Phi_x(f)$ with a rectangular function of finite width and nonzero height. To keep this aliasing effect small in numerical simulations, it is necessary to oversample $x(t)$ by a sufficiently large factor and to filter the spurious out-of-band noise.

Figure 3.11 shows the decomposition of the AWGN channel with Gaussian input with phase noise and spectral loss for $\sigma = 0.5, 1$. Compared to the corresponding subplots in Figure 3.10, the curves are shifted by $10 \cdot \log_{10}(\exp(-\sigma^2))$, i. e., by $-1.1\,\mathrm{dB}$ and $-4.3\,\mathrm{dB}$.

The channel model with phase noise and spectral loss introduced in this section can be applied to estimate the channel capacity of fiber-optic WDM systems (Section 6.2.5).

3.6.3 Noncoherent channels

The *noncoherent* channel is the limiting case of the partially coherent channel (3.81) when Θ is distributed uniformly in $[-\pi, \pi)$. As the phase is completely randomized,

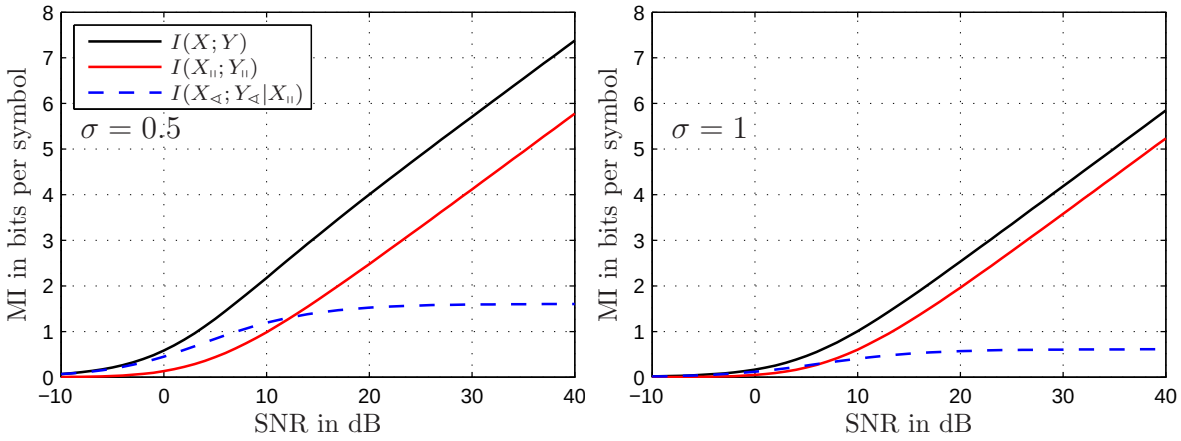

Figure 3.11: Polar decomposition of mutual information (MI) for an AWGN channel with Gaussian input with additional phase noise ($\sigma = 0.5, 1$) and spectral loss. Mixed term II is negligible.

the output phase $y_{\sphericalangle}$ carries no information and the phase term and the mixed term I of the polar decomposition are zero. The fact that the mixed term II is zero, too, is a consequence of $p(y_{\shortparallel}|x_{\shortparallel}, y_{\sphericalangle}) = p(y_{\shortparallel}|x, y_{\sphericalangle})$. The only information that can be transmitted over the noncoherent channel is, therefore, represented by the amplitude term $I(X_{\shortparallel}; Y_{\shortparallel})$.

The noncoherent channel is fully equivalent to the fiber-optic channel with optical amplification and direct detection (cf. Section 2.4.1). This channel can be described as[6]

$$Y = |X + N|. \tag{3.91}$$

Here, $y_{\sphericalangle}$ is constant, so phase term and mixed term I are zero. In the absence of conditioning upon $y_{\sphericalangle}$, mixed term II is also zero since $p(y_{\shortparallel}|x_{\shortparallel}) = p(y_{\shortparallel}|x)$. Hence, the direct-detection (DD) channel is completely characterized by $I(X_{\shortparallel}; Y_{\shortparallel})$ and is thus equivalent to the noncoherent channel.

A way to improve the spectral efficiency of practical noncoherent fiber-optic channels is differential modulation, see e. g. [KAvWW03, WE06].

A related but different situation occurs for channels that obey $Y = |X|^2 + N$. Channels of this kind are found in a variety of optical communication scenarios, with different statistics for N. For example, in thermal-noise limited receivers N is a Gaussian process, but other noise statistics can be found for channels limited by (multiplied) shot noise or by large amounts of optical background noise, both in fiber and free-space optical communications. For a discussion of optical intensity channels with AWGN, see e. g. [HK04]. While the phase term and the mixed term II are zero in this case as for the noncoherent and DD channels, the mixed term I can be larger than zero. Similarly, when the channel input is constrained to real-valued amplitude modulation, i. e. when the

[6]The photocurrent (2.164) is proportional to the optical power, so that the channel model should be stated as $Y = |X+N|^2$. However, as the square law is injective, the data processing inequality (3.16) is satisfied with equality, so the two channel models are equivalent.

channel model is $Y = X + N$, $\mathcal{X} = [0, \infty)$, the mixed term I can be larger than zero. The decomposition of the AWGN channel with OOK modulation discussed in Section 3.4.3 is an example.

Input optimization and information rate calculation

The conditional PDF $p(y_{\text{II}}|x_{\text{II}})$ of the noncoherent channel is Ricean (3.45), so the mutual information $I(X_{\text{II}}; Y_{\text{II}})$ is calculated along the lines of the amplitude term calculation in Section 3.4.1. The difficulty in finding the capacity of the noncoherent channel lies in finding the optimum input distribution $p(x_{\text{II}})$. Similar to the partially coherent channel, it is known that the optimum input distribution $p(x)$ is not Gaussian [CR01], i. e. the optimum $p(x_{\text{II}})$ is not a Rayleigh distribution (3.43). Rather, the capacity-achieving input is discrete [KSS04]. By numerical optimization, Ho found an optimum input (for the optical DD channel) that has a discrete probability mass at $x_{\text{II}} = 0$ and a continuous exponential profile at $x_{\text{II}} > 0$ [Ho05b]. At low SNRs, this distribution collapses to two discrete points at $x_{\text{II}} = 0$ and at $x_{\text{II}} > 0$, i. e. OOK, confirming a result reported in [KSS04].

An analytical approximation to the noncoherent channel's capacity is available in the limit of large SNRs. In this case, the Ricean distribution $p(y_{\text{II}}|x_{\text{II}})$ can be approximated by a Gaussian, and the capacity-achieving input distribution is a *positive normal* or *half-Gaussian* distribution [Bla53]

$$p(x_{\text{II}}) = \begin{cases} \sqrt{\frac{2}{\pi P_s}} \cdot \exp\left(-\frac{x_{\text{II}}^2}{2P_s}\right), & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{3.92}$$

In a derivation analogous to that of (3.48), the capacity is found to be [Bla53]

$$I(x_{\text{II}}; y_{\text{II}}) \approx \frac{1}{2} \log_2 \left(\frac{P_s}{2\sigma_n^2}\right) - \frac{1}{2}, \quad P_s \gg 2\sigma_n^2, \tag{3.93}$$

which is $(\log_2 \pi - (1 + \gamma)/(\ln 2) + 1)/2 \approx 0.19$ bits higher than the mutual information (3.48) that results from a Rayleigh-distributed input. The same result was found in an analysis of optical DD systems [MS01a]. Signal shaping methods for the optical DD are discussed in [MK08].

Noise statistics of direct-detection receivers with electrical filtering

Using the channel model (3.91), the conditional PDF $p(y_{\text{II}}|x_{\text{II}})$ is a Ricean distribution. If the square-law channel model $Y = |X + N|^2$ is assumed, $p(y_{\text{II}}|x_{\text{II}})$ is a noncentral (scaled) χ^2 distribution [MP92, Pro95]. As discussed above, both channel models have the same capacity.

The problem is significantly complicated by placing a linear filter after the square-law device. This situation occurs in optical direct-detection (DD) systems, where electrical filtering after the photodiode is performed to limited the electrical receiver noise [Agr02]. The samples' PDF at the output of the filter is unknown; finding the distribution of a filtered non-Gaussian random process is a non-trivial problem [CFS86].

An approach that has been successfully applied to the bit error rate calculation in optical DD systems is to expand the signal into a Karhunen-Loève series [LS94, CMG+09]. The output sample with unknown distribution can then be expressed as a summation of independent random variables, each obeying a noncentral (scaled) χ^2 distribution. Analytical expressions for the PDF of such a summation exist [CMLB05], but the usual approach is to multiply the moment-generating functions of the series terms' PDFs and obtain the resulting PDF from numerical inverse Laplace transformation [CMG+09]. For practical purposes, histogram-based methods or approximate PDFs [HA91, Lei07] are used.

The problem of determining the channel capacity of optical DD systems in the presence of electrical filtering is not pursued further in this thesis.

3.7 Summary

The question for the fundamental limits of communication over fiber-optic channels can be answered using methods from information theory. In the first part of this chapter, fundamental concepts of information theory were briefly introduced and reviewed. Channel capacity is the maximum rate of transmission over a channel with arbitrarily low error probability. The capacities of some important channel models were discussed for later reference.

After this review of existing results, a decomposition method for mutual information and its applications were proposed and discussed. In summary, the main results of this second part of this chapter are:

- ▷ A polar decomposition of mutual information into an amplitude term, a phase term and two mixed terms represents a tool for the analysis of how information is conveyed over a channel. It is useful for the analysis of channel impairments and modulation schemes and can be used for rapid numerical capacity calculations.

- ▷ The decomposition was performed for the AWGN channel with continuous and discrete input. Analytical approximations were found for the AWGN channel with Gaussian input. Existing results on phase-modulated AWGN channels were complemented with results from the polar decomposition.

- ▷ Fundamental concepts of directional statistics were reviewed.

- ▷ Partially coherent channels are channels with AWGN and additional phase noise. The decomposition amplitude term of such channels is not affected by phase noise. In contrast, the decomposition phase term is upper bounded by phase noise. Nonlinear fiber-optic channels impaired by SPM, XPM and XPM$^{\mathcal{S}}$ are modeled as partially coherent channels in Chapters 5 and 6.

- ▷ A property of partially coherent channels with white phase noise that we call spectral loss was derived and discussed. Effectively, this loss decreases the SNR; hence, the amplitude term is indirectly affected by phase noise, too. Spectral loss must be

taken into account in the analysis of channels impaired by phase noise as well as in their numerical simulation.

- ▷ Noncoherent channels are equivalent to optical direct-detection channels. Only the amplitude term remains from the decomposition of such channels. Asymptotically, the capacity of noncoherent channels is 0.5 bits per symbol lower than half the capacity of the AWGN channel with Gaussian input.

4

Capacity of PMD-impaired channels

In the derivation of the nonlinear Schroedinger equation in Section 2.1, the vectorial propagation equation was reduced to a scalar form by assuming that both orthogonally polarized fundamental modes have identical propagation constants. In real single-mode fibers, which lack perfect circular symmetry, this assumption is never fully justified. The effect that results from birefringence and mode coupling is known as *polarization mode dispersion* (PMD) [GM05]. The impairments caused by PMD vary randomly with both wavelength and time [KL02b, Ch. 15]. Apart from amplified spontaneous emission (ASE) noise from optical amplification, PMD can be regarded as the only really stochastic and time-variant effect in fiber-optic communication systems.

PMD became an important resarch topic after the invention of the Erbium-doped fiber amplifier, when all-optical links could be made long enough for the group delay difference between two polarizations to become significant [KK97a, Ch. 6]. The statistical properties of PMD were investigated and PMD compensation methods were proposed, either based on optical compensators or electrical equalization [GM05]. The recent renewed interest in coherent detection [KLW08a, Ch. 3] has changed the view on PMD. On the one hand, fast digital signal processing (DSP) in combination with signaling schemes such as orthogonal frequency division multiplexing (OFDM) has made the compensation of chromatic dispersion and PMD relatively simple. On the other hand, coherent receivers (which are polarization-sensitive) need to receive two orthogonal polarizations in order to avoid signal fading. This additional hardware effort facilitates the transmission and reception of two independent orthogonally polarized signals (*polarization division multiplexing*, PDM), so that PMD can be understood as a multi-user interference problem in

state-of-the-art fiber-optic communication systems.

The discussion of PMD in this chapter concentrates on systems with coherent detection. We will show that PMD does not degrade the channel capacity when optimum coherent receivers are used. In contrast, PMD was found to degrade the (numerically calculated) information rate of systems with direct detection [MDB+07]. Following a brief review of PMD basics in Section 4.1, coherent receiver structures for single and dual polarizations are introduced in Section 4.2. The effect of PMD on the channel capacity is discussed in Section 4.3 for these receivers. As described in Section 2.1.6, the effect of light polarization on fiber nonlinearities can be taken into account even in a scalar propagation model by using an effective value for the nonlinear material susceptibility. However, the interplay of PMD and fiber nonlinearities leads to fading and interference in PDM systems. Section 4.4 illuminates this aspect. The results of this chapter are summarized in Section 4.5.

Polarization-dependent loss (PDL) is not considered in this thesis. Although PDL (as every other loss phenomenon) does degrade the capacity, this degradation was shown to be moderate [NMS09].

4.1 Introduction to polarization mode dispersion

Optical fibers generally support a number of modes, i. e. solutions (2.36) of the wave equation, with distinct field distributions $\overrightarrow{\boldsymbol{F}}(x, y)$ and propagation constants $\beta(f_0)$. The modes' different group velocities and the power exchange between them lead to transmission impairments known as *modal dispersion* and *mode coupling*. Single-mode fibers support only two orthogonal fundamental modes (above the cutoff wavelength). In ideal single-mode fibers, these are indistinguishable or *degenerate* [KK97a, Ch. 6]. In contrast, real fibers exhibit a certain amount of anisotropy (due to intrinsic geometrical asymetries or due to external stress) that causes the orthogonally polarized fundamental modes to have different group velocities. The resulting modal dispersion, in this context known as *birefringence*, and the power exchange between the fundamental modes (mode coupling) lead to polarization mode dispersion. Both birefringence and mode coupling are linear effects modeled by $\overleftrightarrow{\varepsilon}_{r,\Delta}$ in the first-order material susceptibility tensor $\overleftrightarrow{\boldsymbol{X}}^{(1)}$ (2.18). If the diagonal elements $\varepsilon_{r,\Delta,ii}$ of the tensor $\overleftrightarrow{\boldsymbol{\varepsilon}}_{r,\Delta}$ are complex-valued, their imaginary parts cause PDL.

PMD is usually characterized as a "doubly-ergodic" phenomenon [VP02] in the sense that similar statistical properties are obtained from averaging over an ensemble of fibers, over different wavelengths of the same fiber, and over time. However, it has been pointed out that the time variance of the fiber strongly depends on its deployment and environmental conditions [KL02b, Ch. 15]. A realistic fiber channel model that takes these conditions into account is the *hinge* model, in which few "hot spots" or hinges (such as huts, bridges, etc.) dominate the PMD statistics of the fiber [BFBT06]. An implication of this model is that different wavelengths do exhibit different statistics, so that e. g. outage calculations [KL02b, Ch. 15] need to be performed for each channel separately. In contrast to the hinge model, the channel model introduced in this section is rather generic

but widely used. The properties of PMD that are relevant for an information-theoretic discussion of this effect are inherent to every PMD channel model, so that the simplest available model is the best choice for this purpose.

4.1.1 Short fiber segments

In a fiber segment whose length L_c is much smaller than the correlation length L_{corr} defined in Section 2.2.3, the birefringence can be assumed as constant. We restrict the following discussion to fiber segments with linear eigenpolarizations, i. e. $c_\psi, c_\chi \in \mathbb{R}$ in (2.99), (2.100). In fact, most perturbations of the fiber geometry except fiber twist, e. g. external stress, core ellipticity or bending, lead to linear birefringence. More importantly, however, the PMD statistics of a long fiber, which is modeled as a concatenation of short fiber segments, do not depend on these segments' eigenpolarizations [VP02]. Assuming linearly birefringent segments is therefore no limitation.

We assume that the linearly birefingent fiber segment's eigenpolarizations are aligned with the x and y axes of the coordinate system. These axes are called the *slow axis* (larger group delay) and *fast axis* (smaller group delay). Light that is linearly polarized in either axis' direction maintains its polarization during propagation and is not impaired by PMD. In contrast, if the light wave's power is split between both eigenpolarizations, e. g. if the wave is linearly polarized at a 45° angle, the differential refractive index $\Delta n = |n_x - n_y|$ causes a relative delay of the slow axis. The light polarization undergoes periodic changes along the fiber; one full period is called the beat length (2.151). In the assumed short fiber segment, the eigenpolarizations experience only the described differential group delay; no higher orders of dispersion are present. Hence, a Taylor series expansion of $\Delta\beta(\omega) = (2\pi\Delta n)/\lambda = \omega/c \cdot \Delta n$ about ω_c can be truncated after the linear term and written in equivalent baseband notation as

$$\begin{aligned}\Delta\beta(\omega) &\approx \Delta\beta(\omega_c) + \omega \cdot \frac{\mathrm{d}\Delta\beta(\omega)}{\mathrm{d}\omega}\Big|_{\omega=\omega_c} \\ &= \frac{\xi + \omega \cdot \Delta t_g}{L_c}.\end{aligned} \tag{4.1}$$

In (4.1), Δt_g is the *differential group delay* (DGD) in units of seconds or picoseconds. The PMD parameter of such short constantly birefringent fiber segments, $\Delta t_g / L_c$, has units of s / m or ps / km, so the DGD grows linearly with the fiber length. The random PMD phase angle ξ is frequently omitted in the literature. Its importance in the context of first-order PMD in long fibers is pointed out in the next section.

4.1.2 Long fibers

When the fiber length $L \gg L_{\text{corr}}$, the assumption of constant eigenpolarizations is no longer valid. Instead, the birefringence axes randomly change their orientation, leading to mode coupling. Long fibers are therefore modeled as a concatenation of N short fiber segments of length L_c with constant but random and independent eigenpolarizations [KK97a, Ch. 6] as depicted in Figure 4.1. Hence, a fiber channel devoid of attenuation, chromatic

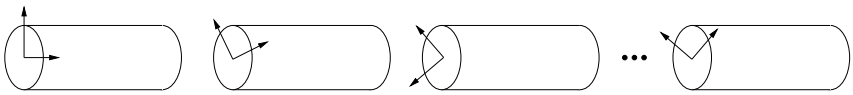

Figure 4.1: Fiber segments of length L_c with random orientation of eigenpolarizations.

dispersion and nonlinear effects is governed by [Men06]

$$\overrightarrow{\boldsymbol{E}}_{\text{out}}(f) = \boldsymbol{H}(f) \cdot \overrightarrow{\boldsymbol{E}}_{\text{in}}(f), \tag{4.2}$$

where

$$\boldsymbol{H}(f) = \prod_{i=1}^{N} \boldsymbol{R}(-\theta_i) \cdot \boldsymbol{U}_i(f) \cdot \boldsymbol{R}(\theta_i). \tag{4.3}$$

It must be noted that the fiber segments of (4.3) act on the signal in their reverse order, i. e. segment N is the first segment of the link. In (4.3), the matrix

$$\boldsymbol{R}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{4.4}$$

rotates the coordinate system by a random angle θ_i uniformly distributed in $[0, 2\pi)$, which aligns the signal coordinates with the segment's eigenpolarizations. Using (4.1), each segment's transfer matrix $\boldsymbol{U}_i(\omega)$ can be written as

$$\boldsymbol{U}_i(f) = \begin{pmatrix} e^{\jmath\,(\xi_i + 2\pi f \Delta t_{g,i})/2} & 0 \\ 0 & e^{-\jmath\,(\xi_i + 2\pi f \Delta t_{g,i})/2} \end{pmatrix}, \tag{4.5}$$

where $\Delta t_{g,i}$ and ξ_i are i. i. d. random variables.

For a sufficiently large number N of fiber segments, the state of polarization (SOP) at the fiber output assumes a uniform distribution over the Poincaré sphere regardless of the input SOP and the segments' eigenpolarizations [VP02]. This simplifies the modeling of long fiber links, because it is unnecessary to know the exact physical perturbations and their locations along the link. We may use fiber segments with arbitrary perturbations; composing the entire link of linearly birefringent segments is the most convenient choice.[1]

Although the pulse propagation is severely randomized on a long link composed of many fiber segments, two SOPs exist for every frequency in which light propagates without distortions from PMD to first order [GM05]. Poole and Wagner, who were the first to describe this phenomenon, called these SOPs *principal states of polarization* (PSP) [PW86]. In the absence of PDL, the PSPs are orthogonal. For each pair of input PSPs, there is a corresponding pair of orthogonal polarization states at the fiber output [GM05].

The group delay difference between both PSPs is the *differential group delay* (DGD) Δt_g. Unlike the DGD of short fiber segments described in Section 4.1.1, Δt_g is proportional to the square root of the fiber length of long links due to mode coupling [KK97a, Ch. 6].

[1] The reason that the hinge model [BFBT06] leads to qualitatively different conclusions lies in the limited number N of hinges assumed along the link.

The PMD parameter of long fibers is therefore given in units of $\mathrm{s}/\sqrt{\mathrm{m}}$. Fibers produced today have PMD parameters of less than $0.05\,\mathrm{ps}/\sqrt{\mathrm{km}}$ [GM05]. The number obtained by multiplying the PMD parameter with the square root of the fiber length yields the expected DGD, i. e. the mean $\overline{\Delta t_g}$ of the random variable Δt_g, which follows a Maxwellian distribution [GM05].

The PSPs vary with frequency, which corresponds to a frequency dependence of the DGD. To first-order in frequency, i. e. for narrow-band signals, the DGD between the PSPs can be assumed constant, and the channel can be modeled as a single constantly birefringent fiber as described in Section 4.1.1. The resulting effect is known as *first-order PMD*.

The difference between modeling the delay between random eigenpolarizations in short fiber segments and modeling first-order PMD in long fibers is the relation of the fiber length L to the correlation length L_{corr}. In the latter case, where $L \gg L_{\mathrm{corr}}$, the phase angle ξ in (4.1) becomes a random variable uniformly distributed in $[0, 2\pi)$. The importance of this random PMD phase angle was pointed out early [WMC91] and can be explained by a length-scale argument [MM06].

First-order PMD is an accurate model if the PSPs remain reasonably constant within the spectral width of the considered signal. If this is not the case, higher-order terms of the Taylor series expansion of the frequency-dependent DGD must be taken into account. Second-order PMD, for instance, considers a linear frequency dependence of Δt_g [GM05]. The analysis of PMD becomes very involved beyond the first order. In contrast, numerical simulations that use a sufficiently large number N of fiber segments produce an "all-order" PMD model. An indication of the tolerable spectral width covered by the first-order PMD model is given by the bandwidth B_{PSP} of the principal states. This bandwidth was determined both experimentally and analytically to be inversely proportional to the mean DGD:

$$B_{\mathrm{PSP}} = c_{\mathrm{PSP}}/\overline{\Delta t_g}. \tag{4.6}$$

Depending on the definition of the bandwidth, reported values for the proportionality constant c_{PSP} include $1/4$ [JNK99, GM05], $\sqrt{8}/\pi$ [KB99] and 0.64 [SMN00]. Consider the example of a fiber with $\overline{\Delta t_g} = 1.25\,\mathrm{ps}$ (e. g. PMD parameter $0.05\,\mathrm{ps}/\sqrt{\mathrm{km}}$ and length $L = 625\,\mathrm{km}$). The resulting bandwidths of the PSPs are 200 GHz, 720 GHz and 512 GHz. The normalized correlation function of the PMD vector at $B_{\mathrm{PSP}}/2$ drops to 0.89, 0.31 and 0.5, respectively, confirming that first-order PMD is sufficient to characterize the propagation of signals in a 100-GHz grid [KL02b, Ch. 15].

4.2 Coherent receiver structures

Before optical amplifiers became available, coherent receivers were widely used in optical communications because of their sensitivity [Agr02]. The recent renewed interest in coherent detection is motivated in part by the demand for increasing data rates that require the detection of spectrally efficient modulation schemes. Coherent receivers grant full access to the optical field, allowing modulation of both quadratures [ILBK08]. On the

other hand, advances in electronic digital signal processing (DSP) have made it possible to build coherent receivers with relatively low (optical) complexity [KLW08b, Ch. 3].

Three types of coherent receivers that can be differentiated are *homodyne*, *heterodyne* and *intradyne* receivers. All employ the same principle of operation. The incoming signal is combined with an unmodulated laser beam, called the *local oscillator* (LO). As the resulting light wave is transformed into an electric current in the photodiode, the signal-LO beat term mixes the signal either down to the baseband (homodyne detection) or to an intermediate frequency (heterodyne detection). The term intradyne detection is used if the intermediate frequency is smaller than the signal's bandwidth. In the case of heterodyne or intradyne detection, a second (electrical) local oscillator is required to downconvert the signal to the baseband. The advantage is that carrier and phase tracking algorithms can be implemented in the electrical domain. Most modern coherent receivers use intradyne detection with a free-running local oscillator (i. e. one whose phase drifts) and phase recovery implemented in DSP [KLW08b, Ch. 3]. In contrast, homodyne receivers require (rather complex) optical phase-locked loops [KLW08b, Ch. 3]. However, from a theoretical point of view, it is convenient to assume homodyne receivers with local oscillators matched to the signal carrier in frequency and phase.

Coherent receivers are usually designed using pairs of balanced photodiodes to suppress the DC components of photodetection [KLW08b, Ch. 3]. Figure 4.2 shows such a balanced coherent receiver. The polarization of the incoming signal $E_s(t)$ must be controlled to

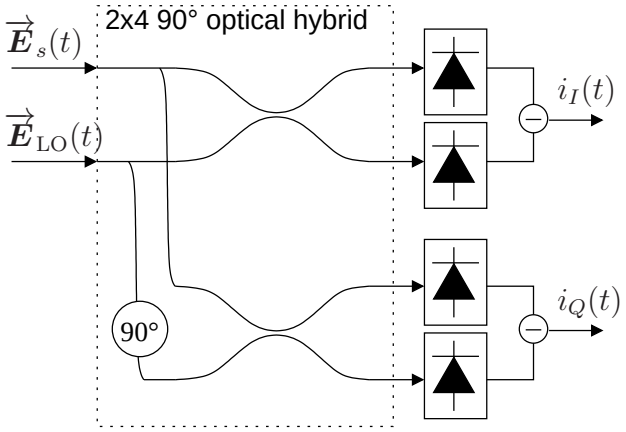

Figure 4.2: Single-polarization phase-diversity homodyne balanced coherent receiver.

match that of the local oscillator. The transfer matrix of the 3-dB optical coupler is

$$\frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{4.7}$$

We assume the local oscillator to be polarized in x-direction, i. e.

$$\vec{E}_{\mathrm{LO}}(t) = \begin{pmatrix} E_{\mathrm{LO}}(t) \\ 0 \end{pmatrix} = c_{\mathrm{norm}} \cdot \begin{pmatrix} a_{\mathrm{LO}}(t) \\ 0 \end{pmatrix}, \tag{4.8}$$

where the normalization constant c_{norm} ensures that $a_{\text{LO}}(t)$ has units of $\sqrt{W}$, cf. (2.78). The field components in x-direction entering the four photodiodes are given by

$$E_{1,x}(t) = E_{s,x}(t) + E_{\text{LO}}(t), \tag{4.9}$$
$$E_{2,x}(t) = E_{s,x}(t) - E_{\text{LO}}(t), \tag{4.10}$$
$$E_{3,x}(t) = E_{s,x}(t) + \jmath\, E_{\text{LO}}(t), \tag{4.11}$$
$$E_{4,x}(t) = E_{s,x}(t) - \jmath\, E_{\text{LO}}(t). \tag{4.12}$$

Hence, the current produced by the first photodiode is

$$i_1(t) = \frac{R}{4} \cdot \left(|a_{s,x}(t)|^2 + |a_{\text{LO}}(t)|^2 + |a_{s,y}(t)|^2 + 2 \cdot \Re\{a_{s,x}(t) \cdot a^*_{\text{LO}}(t)\} \right), \tag{4.13}$$

where R is the photodiode's responsivity introduced in (2.164). The other photocurrents are calcuated in a similar fashion. Finally, the two currents at the receiver output represent the signal's demodulated inphase and quadrature components:

$$i_I(t) = R \cdot \Re\{a_{s,x}(t) \cdot a^*_{\text{LO}}(t)\} \tag{4.14}$$

and

$$i_Q(t) = R \cdot \Im\{a_{s,x}(t) \cdot a^*_{\text{LO}}(t)\}. \tag{4.15}$$

Similar and equivalent single-polarization coherent receiver designs can be found e. g. in [KL02b, ILBK08, EKW+10]. Ho proposes a slightly different design [Ho05a, Fig. 3.4] which uses a circularly polarized local oscillator to demodulate the signal's I/Q components into different polarizations that are separated using a polarization beam splitter (PBS). The incoming signal must then be linearly polarized at a 45° angle.

Unlike direct-detection receivers, coherent receivers are polarization-sensitive. In the worst case of orthogonally polarized signal and local oscillator, the signal-LO beat term vanishes. On the other hand, it is straightforward to extend the single-polarization design shown in Figure 4.2 to dual polarizations. Such a receiver is shown in Figure 4.3. The

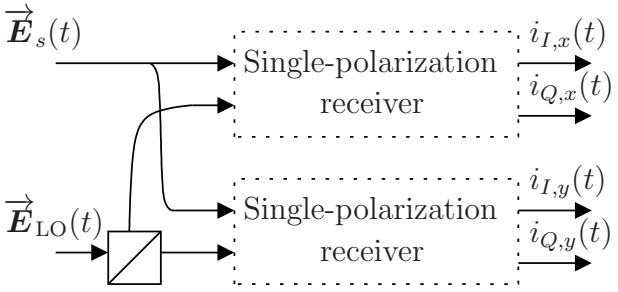

Figure 4.3: Dual-polarization coherent receiver.

local oscillator is linearly polarized at 45° and split by a PBS. Hence, the part of the incoming field that is polarized in x-direction is received by the upper single-polarization receiver, the y-component by the lower one. This receiver can be used to transmit two independent signals in orthogonal polarizations, a technique known as *polarization division multiplexing* (PDM or PolMux).

4.3 Impact of PMD on the channel capacity

When fiber nonlinearities can be neglected and coherent reception is used, the fiber-optic channel is linear and its capacity (3.27) can be calculated from the power transfer function (PTF) $|H(f)|^2$. Chromatic dispersion is an allpass filter function (cf. (2.124)) that does not affect the PTF; it can therefore be neglected. The remaining effects that determine the capacity are attenuation, ASE noise and PMD. The time variance of the channel that results from the temporal changes of PMD needs to be taken into account (cf. Section 3.2.2).

4.3.1 Channel transfer functions

The channel transfer matrix (4.3) for first-order PMD is

$$\boldsymbol{H}(f) = \begin{pmatrix} H_{xx} & H_{xy} \\ H_{yx} & H_{yy} \end{pmatrix}, \tag{4.16}$$

where the matrix elements are calculated using (4.4) and (4.5) as

$$H_{xx} = \cos^2(\theta) \cdot e^{\jmath\,(\xi+2\pi f \Delta t_g)/2} + \sin^2(\theta) \cdot e^{-\jmath\,(\xi+2\pi f \Delta t_g)/2}, \tag{4.17}$$

$$H_{xy} = \cos(\theta)\sin(\theta) \cdot e^{\jmath\,(\xi+2\pi f \Delta t_g)/2} - \cos(\theta)\sin(\theta) \cdot e^{-\jmath\,(\xi+2\pi f \Delta t_g)/2}, \tag{4.18}$$

$$H_{yx} = H_{xy}, \tag{4.19}$$

$$H_{yy} = \sin^2(\theta) \cdot e^{\jmath\,(\xi+2\pi f \Delta t_g)/2} + \cos^2(\theta) \cdot e^{-\jmath\,(\xi+2\pi f \Delta t_g)/2}. \tag{4.20}$$

It is important to note that $\boldsymbol{H}$ is unitary, i. e. $\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} = \boldsymbol{I}_2$.

We can assume the transmit signals to be linearly polarized in x- and y-direction, so that

$$\overrightarrow{\boldsymbol{E}}_{\text{in, single pol.}}(t) = \begin{pmatrix} E_{\text{in}}(t) \\ 0 \end{pmatrix} \tag{4.21}$$

is the electrical field at the transmitter for a single-polarization system, and

$$\overrightarrow{\boldsymbol{E}}_{\text{in, dual pol.}}(t) = \begin{pmatrix} E_{\text{in,x}}(t) \\ E_{\text{in,y}}(t) \end{pmatrix} \tag{4.22}$$

describes the input of a PolMux system.

At the input to the coherent receiver, the polarization of the signal is controlled to match that of the local oscillator. If the polarization controller is fast enough to track the varying polarization and phase of the signal, PMD can be completely mitigated. However, the PMD parameters can change on times scales on the order of milliseconds [KBA00], so it is sensible to assume that the automatic polarization control loop cannot respond to the signal's polarization changes instantaneously. Instead, we assume that the polarization is set such that the maximum signal intensity is captured on average. Introducing an additional rotation matrix $\boldsymbol{R}(\theta')$ before the receiver and maximizing the PTF with respect to θ' yields $\theta' = 0$ [Kus06]. Hence, local oscillator and transmit signal should be identically

polarized (e. g. in x-direction as assumed above) for the single-polarization system. When two PolMax signals are transmitted in x- and y-polarization, the local oscillator should be linearly polarized at an angle of 45°.

The transfer function that governs the PolMux channel is given by (4.16), that of the single-polarization system is given by (4.17). The PTF in this case becomes

$$|H(f)|^2 = |H_{xx}(f)|^2 = 1 - \frac{1}{2} \cdot \sin^2(2\theta) \cdot (1 - \cos(\xi + 2\pi f \Delta t_g)). \tag{4.23}$$

It can be seen from (4.23) that $0 \leq |H(f)|^2 \leq 1$. The PTF is periodic in f with a period of $2\pi/\Delta t_g$. This value is much larger than the bandwidth of the PSPs (4.6). Hence, a signal for which the first-order PMD model is valid encounters a relatively constant spectral slice of the PTF, whose value in this specific signal band depends on the random PMD phase angle ξ. Assuming that θ and ξ are uniformly distributed in $[0, 2\pi)$, the PTF's expectation is independent of the DGD Δt_g and can be calculated as

$$\mathcal{E}_{\theta,\xi,\Delta t_g} \left\{ |H(f)|^2 \right\} = 0.75. \tag{4.24}$$

4.3.2 Noise statistics

The effect of PMD on optical ASE noise requires a separate discussion. ASE noise is unpolarized; it can be modeled as AWGN with identical power spectral density (PSD) in each polarization (see Section 2.3). In most cases, ASE noise propagates along the fiber, either because distributed amplification is used or because multiple lumped amplifiers are cascaded in a multi-span system. To show that any type of ASE noise can be replaced by a lumped source of unpolarized AWGN noise at the receiver, it suffices to show that the statistical properties of ASE noise remain unchanged upon propagation through a single fiber segment with transfer function (4.16). The ASE PSD at the output of the fiber segment is given by

$$\begin{aligned}
\overrightarrow{\boldsymbol{N}}_{\mathrm{ASE,out}}(f) &= \begin{pmatrix} |H_{xx}|^2 & |H_{xy}|^2 \\ |H_{yx}|^2 & |H_{yy}|^2 \end{pmatrix} \cdot \begin{pmatrix} N_{\mathrm{ASE,x}}(f) \\ N_{\mathrm{ASE,y}}(f) \end{pmatrix} \\
&= \begin{pmatrix} |H_{xx}|^2 & |H_{xy}|^2 \\ |H_{yx}|^2 & |H_{yy}|^2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot N_{\mathrm{ASE}} \\
&= \begin{pmatrix} |H_{xx}|^2 + |H_{xy}|^2 \\ |H_{yx}|^2 + |H_{yy}|^2 \end{pmatrix} \cdot N_{\mathrm{ASE}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot N_{\mathrm{ASE}},
\end{aligned} \tag{4.25}$$

where the last equality holds because $\boldsymbol{H}(f)$ is unitary. Hence, PMD does not change the statistics of ASE noise, which can be modeled as lumped unpolarized AWGN in any scenario (single or multiple fiber segments, first- or higher-order PMD).

4.3.3 Capacity results for first-order PMD

As explained in Section 3.2.2, the capacity of time-variant channels is specified in terms of their ergodic or outage capacity. The temporal changes of PMD are too slow to assume

that the channel reveals its ergodicity during the codeword length [KBA00]. We will therefore determine the 10% outage capacity C_{10} (3.34) from the probability distribution of the instantaneous capacity.

Because the parameters θ, ξ and Δt_g are random variables, the PTF and the resulting channel capacity are random variables, too. Both θ and ξ are uniformly distributed in $[0, 2\pi)$. This makes the distribution of Δt_g irrelevant, as the argument of the exponential function is uniformly distributed in any case. A remarkable consequence is that the channel capacity is independent of the DGD.

Without any channel state information (CSI), the PSD of the transmit signal should be flat. A feedback loop that provides the transmitter with information on the current PMD properties of the channel seems unrealistic in high-speed optical networks. As we shall see below, it is more promising to optimize the receiver design to achieve capacity. We assume in the following that both $N_{\mathrm{ASE}}(f)$ and $\Phi(f)$ are constant within the signal band and replace $\Phi(f)/N_{\mathrm{ASE}}(f)$ by the SNR P_s/P_n where appropriate.

The instantaneous capacity (in bits/s/Hz) of the single-polarization receiver can be calculated using (3.27) and (4.23) as

$$\begin{aligned} C &= \log_2\left(1 + \frac{\Phi_x(f)\cdot|H_{xx}(f)|^2}{N_{\mathrm{ASE}}(f)}\right) \\ &= \log_2\left(1 + \frac{P_s}{P_n}\cdot|H_{xx}(f)|^2\right), \end{aligned} \tag{4.26}$$

where $\Phi_x(f)$ is the PSD of the transmit signal (in x-polarization), $P_s = \int_{-\infty}^{\infty}\Phi_x(f)\mathrm{d}f$ is the signal power and P_n is the ASE noise power in one polarization.

If two indendent signals with equal PSDs $\Phi_x(f) = \Phi_y(f) = \Phi(f)$ are transmitted (PolMux), PMD leads to fading and interference between the signals. Both sub-channels have the same instantaneous capacity which can be calculated as

$$\begin{aligned} C &= \log_2\left(1 + \frac{\Phi_x(f)\cdot|H_{xx}(f)|^2}{\Phi_y(f)\cdot|H_{xy}(f)|^2 + N_{\mathrm{ASE}}(f)}\right) \\ &= \log_2\left(1 + \frac{|H_{xx}(f)|^2}{|H_{xy}(f)|^2 + N_{\mathrm{ASE}}(f)/\Phi(f)}\right) \\ &= \log_2\left(1 + \frac{|H_{xx}(f)|^2}{|H_{xy}(f)|^2 + P_n/P_s}\right). \end{aligned} \tag{4.27}$$

Since the channel is time-variant, the capacities (4.26) and (4.27) are random variables. Their probability distributions were numerically calculated assuming uniformly distributed random variables θ and ξ. From these distributions, the 10% outage capacities were calculated as in (3.34). Figure 4.4 shows these 10% outage capacities. It can be seen that the capacity of the single polarization system is significantly reduced compared to that of an AWGN channel. The capacity of each sub-channel of the PolMux system is lower than that of the single-polarization system due to the additional interference from

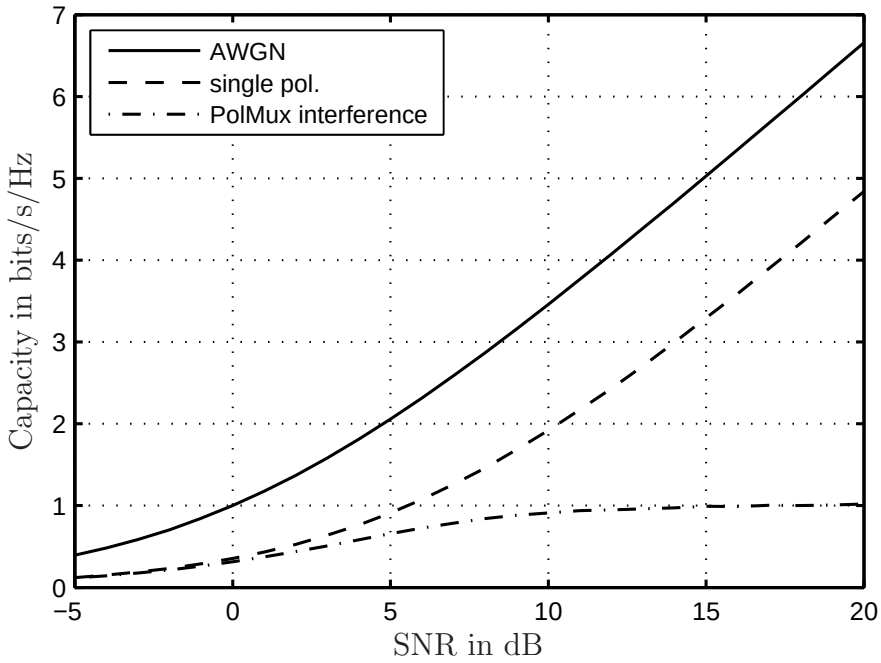

Figure 4.4: 10% outage capacities of linear fiber channel over SNR in dB, $10\log_{10}(P_s/P_n)$, with first-order PMD (single polarization and PolMux).

the orthogonally polarized sub-channel. As this interference increases with the SNR, the capacity converges to a (low) value.

Obviously, both coherent receivers do not deliver satisfactory performance. A significant improvement can be achieved by regarding the PolMux system as a multi-user system, in which detection, equalization and decoding are performed jointly for both signals. The system can then be seen as a 2×2 MIMO system with signal power P_s in each sub-channel (i. e. $2P_s$ total signal power) and noise power P_n per channel output (i. e. $2P_n$ total noise power). Using (3.37), the capacity of this MIMO system is

$$\begin{aligned} C &= \log_2 \det\left(\boldsymbol{I}_2 + \frac{2P_s}{2P_n}\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\right) = \log_2 \det\left(\boldsymbol{I}_2 + \frac{P_s}{P_n}\boldsymbol{I}_2\right) \\ &= \log_2\left(1+\frac{P_s}{P_n}\right)^2 = 2\cdot\log_2\left(1+\frac{P_s}{P_n}\right), \end{aligned} \tag{4.28}$$

which is twice the capacity of an AWGN channel without PMD shown in Figure 4.4. Hence, the capacity of the linear fiber-optic channel can be doubled (if the total transmit power is doubled, too) using PolMux and joint detection independent of the PMD of the fiber.

4.3.4 Extension to higher-order PMD

In order to model higher orders of PMD, the number N of fiber segments in the model (4.3) is increased. The average signal energy (4.24) that the single-polarization receiver

captures decreases from 0.75 to 0.5 as N increases [Kus06]. An interesting result is that the ergodic capacity of the all-order PMD channel (i. e. $N \to \infty$) asymptotically (i. e. for $P_s/P_n \to \infty$) approaches a value $1/\ln(2)$ bits/s/Hz (or 1 nats/s/Hz) smaller than that of an AWGN channel [GKH07].

It must be noted, however, that a detailed analysis of suboptimum receivers for higher PMD orders is of limited value. As discussed in Section 4.3.3, a PolMux system that employs MIMO signal processing can achieve twice the single-polarization AWGN capacity. MIMO detection for coherent PolMux systems was first proposed in 2005 [HL05], and such receivers have already been successfully implemented in practical lab experiments [JMST08, JMST09].

For higher PMD orders, the channel transfer matrix (4.3) remains unitary. Consequently, the ASE noise PSD remains white (see Section 4.3.2), the optimum input PSD remains flat, and the channel capacity is given by (4.28).

4.4 The interplay of PMD and fiber nonlinearities

In the discussion of PMD so far, nonlinear fiber effects were neglected. When the power of the propagating optical field is sufficiently high, however, fiber nonlinearities must be taken into account and cannot be treated separately from PMD, as both effects interact along the link. Recent experiments, e. g. [XWP+08, BPRC+09], confirm that PMD compensation schemes optimized for the linear regime perform poorly when the system is operated at higher powers.

A complete description of the light propagation in single-mode fibers is given by the vectorial propagation equation (2.41). From there, the coupled NLSE, the Manakov-PMD equation and the Manakov equation can be derived [MM06]. These equations govern the propagation of arbitrarily polarized fields in single-mode fibers. This propagation can be simulated numerically by combining the split-step Fourier Algorithm 2.1 with the approach outlined in Section 4.1.2. This method is known as the *coarse-step* method [MMW97].

The varying polarization state of the propagating field has an averaging effect on the cubic susceptibility. This leads to an effective cubic susceptibility $X_{\text{eff}}^{(3)}$. The derivation of $X_{\text{eff}}^{(3)}$ outlined in Section 2.1.6 shows that no signal energy is transferred to the orthogonal polarization due to nonlinear effects *on average* (see (2.109) and (2.117)). This observation allows us to reduce the vectorial propagation equations to a scalar approach. It does not imply, however, that the state of polarization (SOP) of a propagating light wave is not affected by nonlinear interaction with co-propagating waves. Light waves do alter each other's SOPs in a random manner. The resulting SOPs are therefore random variables, and the meaning of (2.109) and (2.117) is that these random variables have zero mean.

The nonlinear polarization rotation becomes apparent in WDM systems as *cross-polarization modulation* (XPolM) [KS06, PWS06]. Through this effect, rapid SOP changes in one WDM channel are induced by the co-propagating channels. These polarization

fluctuations occur on time scales on the order of a single symbol period [KS06]. PMD compensators or the MIMO processing algorithms mentioned in Section 4.3 are unable to track these rapid SOP changes. XPolM therefore introduces fast fading in single-polarization systems; in PolMux systems, XPolM leads to fading and interference between the sub-channels.

The impairments due to XPolM depend on numerous system parameters such as fiber parameters (attenuation, dispersion, PMD parameter), link design (amplification scheme, dispersion map) and system parameters (channel spacing, number of channels, power per channel, data rate and format). Predicting these impairments requires prohibitively extensive numerical simulations. A more promising approach is to view XPolM as a random distortion that can be quantified statistically. Winter et al. [WBSP09] have developed a statistical theory of XPolM that delivers the probability distribution of a probe channel's SOP as a function of the before-mentioned system parameters. In an application of their theory to PolMux systems, the authors show that the resulting interference term can be modeled as a random variable with Gaussian distribution under certain circumstances [WSP10]. The variance V_{xt} of this crosstalk (xt) term is given as [WSP10, (9)]

$$V_{\mathrm{xt}} = P_s/(2V'), \tag{4.29}$$

where P_s is the average optical power in the interfering sub-channel. The auxiliary parameter V' is related to the variance parameter V introduced by Winter [WBSP09] through

$$V = -2\ln\left(\coth(V') - 1/V'\right). \tag{4.30}$$

Whether XPolM sets fundamental limits to the data transmission in fiber-optic communication systems depends on the particular configuration of these systems. The analysis of nonlinear fiber effects performed in Chapter 6 delivers optimum input powers that maximize the channel capacity (see e. g. Figure 6.11). To determine whether XPolM is significant at these optimum power levels, we will now derive the interference variance (4.29) as a function of the channel power of a simplified scenario using the theory of [WBSP09]. After some algebra, the parameter V can be written as

$$V = \frac{2}{3}\cdot\gamma_{\mathrm{eff}}^2\cdot P_s^2\cdot z\cdot\sum_m L_{\mathrm{W},m}, \tag{4.31}$$

where $\gamma_{\mathrm{eff}} = 8/9\cdot\gamma$ is the effective nonlinear fiber parameter and $L_{\mathrm{W},m}$ is the walk-off length (2.150) of Channel m and the probe channel. The summation in (4.31) is over all WDM channels co-propagating with the probe channel. The following assumptions and simplifications were made in the derivation of (4.31):

1. An optimal distributed optical amplification scheme was assumed in which the local gain equals the local loss at every point along the link (cf. Section 2.3.2). While such a scheme maximizes the SNR for a given nonlinear phase rotation [PW02], it increases nonlinearities (compared to lumped amplification) because of the constantly high power level along the link. In the given context, in which ASE noise is not considered, this assumption corresponds to a worst-case scenario.

2. The optical link has no inline optical dispersion compensation.

3. The co-propagating channels transport (arbitrarily modulated) NRZ pulses. The autocovariance function [WBSP09, (40)] of the power in two interacting channels can then be replaced by the simpler expression [WBSP09, (41)].

4. The fiber has a low PMD parameter, e. g. $0.1\,\mathrm{ps}/\sqrt{\mathrm{km}}$, so the SOP decorrelation length [WBSP09, (53)] is very long compared to the walk-off length $L_{\mathrm{W},m}$. In that case and in combination with the previous assumption, the SOP autocovariance function [WBSP09, (46)] can be neglected (i. e. set to 1). In general, the effect of XPolM is smaller in fibers with larger PMD parameters [WBSP09]. Assuming a low-PMD fiber therefore corresponds to a worst-case scenario.

5. The link length z is very long compared to the walk-off length $L_{\mathrm{W},m}$. In combination with the two previous assumptions, this turns the autocovariance function [WBSP09, (40)] into a constant.

The interference term's variance (4.29) was calculated using (4.31) for a system with $z = 2000\,\mathrm{km}$ of SSMF (fiber parameters given in Table 2.3). The probe channel is the center channel of 21 WDM channels co-propagating at a channel spacing of $B = 100\,\mathrm{GHz}$. The symbol rate in each channel is $1/B$. Figure 4.5 shows the normalized crosstalk variance as a function of the power per WDM channel. It can be seen that the variance remains

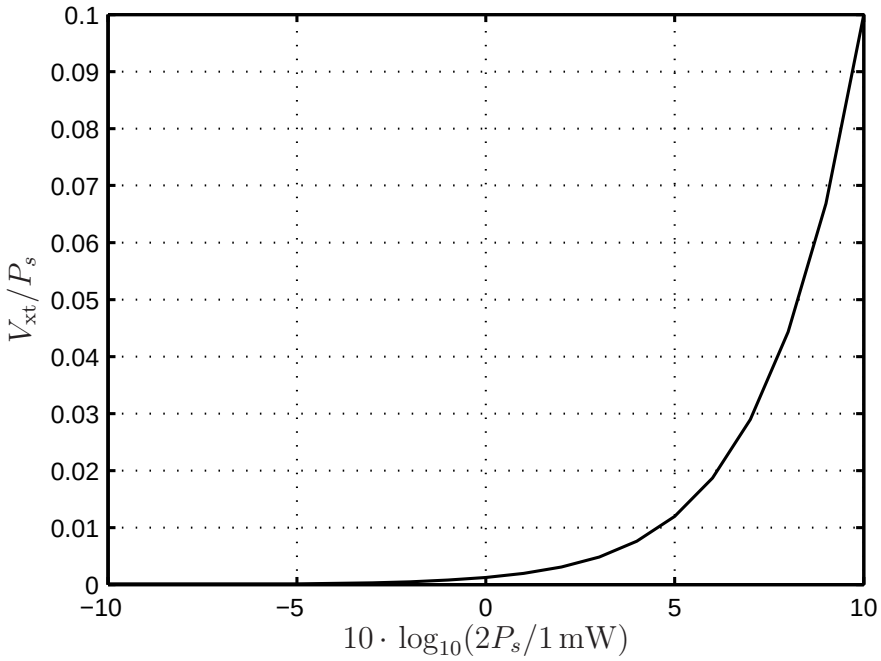

Figure 4.5: Normalized variance of XPolM-induced interference term over signal power per WDM channel in dBm.

low at moderate power levels. For comparison, the capacity-maximizing input power for a 2000-km link is $-2.6\,\mathrm{dBm}$, cf. Figure 6.10. If a system is operated at power levels for

which the curve shown in Figure 4.5 indicates significant XPolM-induced crosstalk, it is advisable to perform a more detailed analysis using the method of [WBSP09].

4.5 Summary

Single-mode fibers provide two orthogonal channels for light propagation, namely the two orthogonally polarized fundamental modes. These channels are coupled linearly (through PMD) as well as in a nonlinear way through the cubic susceptibility tensor $\overleftrightarrow{\boldsymbol{X}}^{(3)}(f)$. It is possible to neglect this vectorial nature of light propagation in optical fibers (and the following chapters will do so), but a discussion of the linear and nonlinear effects related to light polarization must precede this simplification and is therefore provided in this chapter. In summary, the main points and results of this discussion are:

- ▷ Polarization-mode dispersion (PMD) as a linear fiber effect was briefly introduced. PMD is a random effect that varies with time and wavelength.
- ▷ Channel models for short fiber segments and for long fibers were introduced. The concept of principal states of polarization was explained. The bandwidth of the principal states determines whether a first-order PMD channel model is valid.
- ▷ Single- and dual-polarization coherent receiver structures were introduced.
- ▷ In the absence of polarization-dependent loss, the fiber channel transmission matrix $\boldsymbol{H}$ is unitary. As one consequence, PMD does not change the statistical properties of ASE noise.
- ▷ The outage capacity of coherent fiber-optic communication systems in the presence of PMD was calculated. PMD leads to fading in single-polarization receivers and reduces the capacity significantly. In PolMux systems, uncompensated PMD leads to sub-channel interference that sets an upper bound on the channel capacity.
- ▷ If the channel is seen as a 2×2 MIMO system, the capacity is twice that of an AWGN channel for any amount and order of PMD. This is due to $\boldsymbol{H}$ being unitary. The transmitter does not require channel state information, and the optimum transmit PSD is flat.
- ▷ In the nonlinear regime, the interaction of nonlinearities and polarization can be allowed for by the effective cubic susceptibility $X_{\text{eff}}^{(3)}$. In WDM PolMux systems, XPolM can generate rapidly varying sub-channel interference. A simplified statistical model was derived to estimate the variance of this interference term. A case study showed that XPolM introduces only negligibly small interference at the capacity-maximizing power level (derived in Chapter 6).

5

Nonlinear propagation of a single field

The effect of fiber nonlinearities that impair the propagation of a single field (as opposed to a number of WDM channels) on the channel capacity are discussed in this chapter. Whether (and to what extend) fiber nonlinearities are "fundamental" and hence have an effect on the channel capacity depends significantly on our view of the system and its physical effects. For this reason, Section 5.1 opens the chapter with some introductory remarks. Afterwards, Section 5.2 introduces the system model used for the channel capacity discussion. The analysis of impairments is performed separately for the effects of self- and cross-phase modulation (SPM, XPM) (Section 5.3) and four-wave mixing (FWM) (Section 5.4). The interaction of signal and ASE noise through these nonlinear effects is treated in Section 5.5. Finally, the main results of this chapter are summarized in Section 5.6.

5.1 Introductory remarks

5.1.1 The character of nonlinear impairments

Noise generated by optical amplifiers (and, less dramatic, electrical shot and thermal noise) and the polarization evolution discussed in Chapter 4 are the only truly random effects in fiber-optic communication systems. Nevertheless, the literature that discusses "Shannon limits" of the fiber-optic channel usually considers the Kerr effect to be the fundamental impairment of this channel (apart from additive noise). However, the distortions caused by fiber nonlinearities are by no means fundamental. The nonlinear

Schroedinger equation (NLSE) (2.79) is invertible. The idea of designing optical equalizers that invert the linear and nonlinear kernels of the Volterra series expansion solution of the NLSE (Section 2.1.7) is sketched in [PBP97]. In a more practical approach, *electronic predistortion* (or *backpropagation*) has been proposed and applied in numerous papers including [EWW+06, RLS+06, WFBP06, HWHP07, IK08]. In this method, the signal is propagated backwards along the fiber (from $z = L$ to $z = 0$), resulting in a distorted waveform that, when transmitted, is received as the original signal. Backpropagation is easily implemented using the SSF Algorithm 2.1 and replacing z by $-z$ or, equivalently, the fiber parameters α, β_2 and γ by their negatives. For predistortion of PDM signals, backpropagation has also been applied to the coupled NLSE [YL09, MG09, Ip10].

It is pragmatic to view the impairments that originate from the material nonlinearity as interference. The entire amount of interference that a fiber-optic communication system is unable to remove is reducing this system's channel capacity in practice. To what extend the resulting limit is "fundamental" is then a question of practical system design.

The interference that fiber nonlinearity creates can be categorized in terms of signals propagating in different frequency bands (as in a WDM system). Interference that has its origin in the nonlinear interaction of the signal in one band (or channel) with one or more out-of-band signals (or other channels) are discussed in Chapter 6. The term *inter-channel effects* is usually used to refer to this class of nonlinearities. In contrast, the nonlinear interaction of the spectral components of the signal within the same band, treated in this chapter, leads to *intra-channel effects*. In practice, the latter are much less "fundamental", as they are easily removed e. g. by the mentioned predistortion approach, whereas inter-channel nonlinearities mitigation is complicated by the limited access that receivers of one particular channel have to the co-propagating channels in transparent optically routed networks [EKW+10].

Both classes of nonlinearities can be further sub-divided into interactions of the signal with itself (or other signals) and with co-propagating ASE noise. The latter can lead to a parametric gain of the noise power and to a modification of the noise power spectral density (PSD). Because nonlinear signal-ASE interaction results in really fundamental distortions, it requires a separate discussion. In this chapter, Section 5.5 serves this purpose.

5.1.2 Literature review

Comparatively few publications discuss the information-theoretic effect of fiber nonlinearities. A good overview is given by the review articles [KH04, EZC10, EKW+10]. In the following, those contributions that consider intra-channel nonlinearities are briefly summarized.

Tang considers the case of a dispersionless fiber, for which the NLSE has the explicit solution (2.122). Using a formula derived by Pinsker [Pin63], he finds approximate analytical expression for the channel capacity of single- and multi-span links [Tan01b, Tan01a]. Turitsyn et al. have criticized the use of the Pinsker formula [TDYT03]. For the same

dispersion-less fiber, they obtain results by direct evaluation of (3.18). They argue that the random nonlinear phase rotation eventually renders the channel noncoherent and obtain the asymptotic (i. e. in the limit of large SNRs) capacity (3.93) for a Gaussian input. This is in contrast to their (and Tang's) result obtained using Pinsker's formula; these results suggest that $C \to 0$ asymptotically.

The channel model becomes significantly more complicated if the fiber has dispersion and therefore introduces memory to the channel. Authors that have considered this case have used so-called *perturbative approaches* which consider nonlinear effects as small perturbations. In this scenario (which implies low signal powers), the channel can be described by a Volterra series expansion (2.127) that is truncated after the first nonlinear term. This term is then considered as an additive perturbation term in an AWGN channel model. Using this approach, Louchet determines capacity results taking into account both inter- and intra-channel nonlinearities [Lou06]. Tang applies the Volterra series method to multi-span systems [Tan02, Tan06]. In his calculation, the output signal of each fiber span serves as the input signal to the next span. Hence, new terms beyond the third order must be neglected at every span and the resulting model is confined to very low power levels. Narimanov and Mitra use a very similar perturbation approach and obtain an expression for the capacity that equals the AWGN capacity decreased by two terms [NM02]. The authors give a physical explanation of these terms' orgins and relate them to nonlinear out-of-band power loss (cf. Section 5.4) and noise enhancement (cf. Section 5.5).

Numerical calculations of the information rate of the nonlinear fiber-optic channel using the Shannon-McMillan-Breiman theorem (3.31) are presented in [DVIG05, IDV07]. The authors assume binary input constellations for their simulation. Other specific signaling schemes for which capacity results have been reported include OOK and code division multiple access [WN05] as well as orthogonal frequency division multiplexing (OFDM) [GFCH08, HM10].

5.1.3 A frequency-domain view on capacity

It is difficult to compare the reported results, mainly because of the different models, assumptions and simplifications used by the various authors. One statement that most authors explicitly agree to is that the time-domain NLSE is too difficult to use as a basis for an information-theoretic discussion unless significant simplifications are made (such as e. g. no dispersion or very weak nonlinearities). Therefore, no attempt to add any contributions to this time-domain discussion is made in this chapter. Instead, the proposed methodology to estimate the channel capacity is based on the analysis of the nonlinear processes in the frequency domain. Despite the popularity of the time-domain NSLE, it can be argued that the frequency domain is the "natural domain" of light propagation in optical fibers. This is obvious for linear effects, i. e. attenuation and dispersion, which can be modeled as a linear filter function $H(f)$. In contrast, fiber nonlinearities are usually seen as a time-domain phenomenon, namely the instantaneous power-dependent phase rotation invoked by the Kerr effect (see e. g. (2.70)). This dependence of the refractive index on the incident field strength (2.74), however, is a phenomenological description. It is physically rooted

in the interaction of combinations of four spectral components through the cubic material polarization (2.21) as discussed in Section 2.1.2. Rather than the time-domain NLSE, we therefore use propagation equation (2.60) as the starting point of our discussion.

It can be argued that the results obtained from this frequency-domain discussion are not universally applicable. As an example, the capacity obtained for dispersion-shifted fibers (DSF) tends to zero rapidly with increasing SNR, as we will see later. This result is in contrast to the above-mentioned asymptotic capacity (3.93) found by [TDYT03].[1] Moreover, electronic predistortion is particularly easily implemented in the absence of dispersion [Low07]. On the other hand, there are practical communication systems to which the frequency-domain approach to capacity calculation directly applies, e. g. OFDM systems [GFCH08]. More importantly, the approach introduced in this chapter is easily extended to the analysis of WDM systems. The critical reader who does not endorse the fundamental character of the impairments caused by intra-channel effects may therefore view this chapter as an introduction to Chapter 6.

5.1.4 Line coding schemes for fiber nonlinearity reduction

Apart from the electronic predistortion discussed in Section 5.1.1, which requires knowledge of the link parameters, several line coding algorithms that reduce fiber nonlinearities have been proposed. Such schemes do not change the channel capacity, but are one way (in addition to error-correcting codes) to get closer to it.

In multi-carrier systems (e. g. OFDM), certain phase relations of the sub-carriers favor the generation of nonlinear distortions [Han95]. By encoding the corresponding data patterns, fiber nonlinearities can therefore be mitigated. However, criteria for the dependence of the sub-carriers' phases and the resulting nonlinear distortions are difficult to formulate. An efficient line coding scheme that operates in the frequency domain has therefore not been reported. Instead, good results can be achieved by applying algorithms that reduce the signal's peak-to-average power ratio (or, more generally, the signal power variance) in the time domain [GFCH08, GHHH09, HGH09, GHH10].

In high-speed optical pulse transmission, certain bit patterns are known to favor the generation of so-called *ghost pulses* through intra-channel FWM [Sch04]. Several constrained coding schemes have been proposed to encode and avoid these bit patterns [VRD+04, DV06, KSV06, PK06, BDV07, STT07, TAP09].

5.2 System model

We consider the propagation of an arbitrary band-limited signal $E(z,t)$ (2.62) with Fourier transform $E(z,f)$ (2.63). The signal's bandwidth is B. The equation that governs this signal's propagation in the frequency domain is the NLSE (2.68). The triple integral in

[1] Although (3.93) is the capacity of a *linear* noncoherent channel, the nonlinear channel considered by [TDYT03] approaches the same capacity in the limit where strong nonlinear phase noise renders the channel noncoherent but does not affect the amplitude.

this equation ensures that all frequency combinations that lead to SPM, XPM and FWM, explained in Section 2.2.2, are included. The analysis of these nonlinear fiber effects is much simpler if we consider the propagation of a number of discrete spectral components, governed by (2.60). We will therefore use the following spectrally discrete description of $E(z, f)$ for the analysis of nonlinear impairments. Let $E^{(1)}(z), E^{(2)}(z), \ldots, E^{(N_{\mathrm{sc}})}(z)$ be N_{sc} samples of $E(z, f)$ at frequencies $f_1, f_2, \ldots, f_{N_{\mathrm{sc}}}$ with constant sampling interval $\Delta f = B/N_{\mathrm{sc}}$. To avoid aliasing, we ask that $E(z, t)$ is limited in time to $1/\Delta f = N_{\mathrm{sc}}/B$. To obtain the original spectrally continuous (and unlimited in time) signal, we let $N_{\mathrm{sc}} \to \infty$. As we will see later, the nonlinear impairments converge relatively fast as N_{sc} increases. This allows us to obtain results that apply to general arbitrary signals from a discussion of the nonlinear interaction of discrete spectral components.

This approach can be seen as a decomposition of the channel into N_{sc} parallel narrowband sub-channels with a common average power contraint. If each sub-channel can be modeled as an AWGN channel, the total channel capacity is given by (3.27).

To derive a propagation equation for the discrete spectral components, we start at (2.60). The frequency condition used in this chapter is

$$f_m = f_n + f_p - f_q, \tag{5.1}$$

where $1 \leq n, p, q \leq N_{\mathrm{sc}}$. The spectral broadening induced by FWM is reflected in the condition $2 - N_{\mathrm{sc}} \leq m \leq 2N_{\mathrm{sc}} - 1$. To obtain a notation that is convenient for the problem, we first define the normalized spectral component

$$X^{(n)}(z) = \sqrt{\frac{A_{\mathrm{eff}}}{2Z}} \cdot E^{(n)}(z) \cdot \Delta f \tag{5.2}$$

in units of $\sqrt{\mathrm{W}}$, cf. (2.55). Using (5.2) and (2.72), (2.60) can be written as

$$\begin{aligned} \frac{\partial}{\partial z} X^{(m)}(z) = &- \frac{\alpha}{2} \cdot X^{(m)}(z) \\ &- \jmath\, \gamma \cdot \sum_{n,p,q} X^{(n)}(z) \cdot X^{(p)}(z) \cdot X^{(q)^*}(z) \cdot e^{-\jmath\, \Delta\beta \cdot z} \end{aligned} \tag{5.3}$$

with phase mismatch

$$\Delta\beta = \beta(f_n) + \beta(f_p) - \beta(f_q) - \beta(f_m). \tag{5.4}$$

We ask that the signal $E(z, t)$ is designed such that $X^{(n)}, 1 \leq n \leq N_{\mathrm{sc}}$, are i. i. d. random variables.[2] This implies that $E(z, t)$ has a white PSD. If P_s denotes the average signal power, each spectral component has average power

$$P^{(n)} = \mathcal{E}\left\{|X^{(n)}|^2\right\} = \frac{P_s}{N_{\mathrm{sc}}}, \quad 1 \leq n \leq N_{\mathrm{sc}}. \tag{5.5}$$

As $N_{\mathrm{sc}} \to \infty$, the power of each spectral component goes to zero.

[2] OFDM is an example for a transmission scheme that satisfies this condition.

The signal bandwidth B is assumed to be sufficiently small so that the random polarization rotation can be accounted for by using $\gamma_{\text{eff}} = 8/9 \cdot \gamma$ (see Section 2.1.6, Table 2.2). To separate the effects of SPM, XPM and FWM, we use the frequency conditions discussed in Section 2.2.2 and write (5.3) as

$$\begin{aligned}\frac{\partial}{\partial z}X^{(m)}(z) = &- \underbrace{\frac{\alpha}{2} \cdot X^{(m)}(z)}_{\text{attenuation}} \\ &- \underbrace{\jmath\,\gamma_{\text{eff}} \cdot |X^{(m)}(z)|^2 \cdot X^{(m)}(z)}_{\text{SPM}} \\ &- \underbrace{\jmath\,\gamma_{\text{eff}} \cdot 2 \cdot \sum_{\substack{n=1 \\ n\neq m}}^{N_{\text{sc}}} |X^{(n)}(z)|^2 \cdot X^{(m)}(z)}_{\text{XPM}} \\ &- \underbrace{\jmath\,\gamma_{\text{eff}} \cdot \sum_{\substack{n,p,q \\ q\neq n, q\neq p}} X^{(n)}(z) \cdot X^{(p)}(z) \cdot X^{(q)^*}(z) \cdot e^{-\jmath\,\Delta\beta \cdot z}}_{\text{FWM}}. \end{aligned} \tag{5.6}$$

The summation in the FWM mixing term is over all $\{n, p, q\}$ that satisfy $m = n + p - q$ except if $q = n$ or $q = p$. The three nonlinear effects are discussed separately in the next two sections.

5.3 Self- and cross-phase modulation

It is obvious from (5.6) that SPM induces a phase rotation of $X^{(m)}(z)$ proportional to its own power $|X^{(m)}(z)|^2$. Neglecting the XPM and FWM terms, (5.6) can be solved in closed form, cf. (2.122). The SPM-induced phase shift is

$$\phi_{\text{SPM}} = \gamma_{\text{eff}} \cdot |X^{(m)}(z = 0)|^2 \cdot L_{\text{eff}}, \tag{5.7}$$

where L_{eff} is the effective fiber length (2.140). This deterministic phase rotation can be undone at either the receiver or the transmitter. SPM is therefore not considered as a relevant effect for the channel capacity.[3]

To isolate the effect of XPM, we neglect the SPM and FWM terms in (5.6). The solution

[3] An effect that we ignore in the following is that SPM translates FWM noise into a non-deterministic phase shift, similar to the interaction of SPM and ASE treated in Section 5.5. It is shown later that the magnitudes of FWM and ASE noise power are comparable at the optimum input power level. Consequently, the phase noise that the two noise sources cause through SPM is comparable, too. The interaction of ASE and SPM is shown to have a negligible effect in Section 5.5. We expect the same result for the interaction of FWM noise and SPM.

of the resulting differential equation is

$$X^{(m)}(z) = X^{(m)}(z=0) \cdot \exp\left(-\frac{\alpha}{2} \cdot z\right) \cdot \exp\left(-\jmath\, 2\gamma_{\text{eff}} \sum_{\substack{n=1 \\ n\neq m}}^{N_{\text{sc}}} |X^{(n)}(z=0)|^2 \cdot L_{\text{eff}}\right)$$
$$= X^{(m)}(z=0) \cdot \exp\left(-\frac{\alpha}{2} \cdot z - \jmath\, \phi_{\text{XPM}}\right). \tag{5.8}$$

Hence, XPM causes a random phase rotation by ϕ_{XPM} that is proportional to the combined power of the co-propagating spectral components. As these are i. i. d. random variables, the sum term ϕ_{XPM} has a Gaussian distribution for large N_{sc} as a consequence of the central limit theorem. The mean phase shift due to XPM,

$$\mathcal{E}\left\{\phi_{\text{XPM}}\right\} = (N_{\text{sc}} - 1) \cdot \frac{2 \cdot \gamma_{\text{eff}} \cdot P_s \cdot L_{\text{eff}}}{N_{\text{sc}}} \approx 2 \cdot \gamma_{\text{eff}} \cdot P_s \cdot L_{\text{eff}}, \tag{5.9}$$

is deterministic and can be removed as discussed above.

The variance V_{XPM} of ϕ_{XPM} is the sum of the variances of the individual terms in the summation. To find an expression for V_{XPM}, we assume that the spectral components have a Gaussian distribution, $X^{(n)} \sim \mathcal{N}_{\mathbb{C}}\left(0, P_s/N_{\text{sc}}\right)$, $1 \leq n \leq N_{\text{sc}}$. The power $|X^{(n)}|^2$ in each sub-channel then follows an exponential distribution with parameter N_{sc}/P_s [Pro95]. The variance of this (exponentially distributed) random variable is

$$\mathcal{E}\left\{|X^{(n)}|^4\right\} - \mathcal{E}\left\{|X^{(n)}|^2\right\}^2 = P_s^2/N_{\text{sc}}^2. \tag{5.10}$$

Finally, the variance V_{XPM} of ϕ_{XPM} becomes

$$V_{\text{XPM}} = (N_{\text{sc}} - 1) \cdot 4 \cdot \gamma_{\text{eff}}^2 \cdot L_{\text{eff}}^2 \cdot \frac{P_s^2}{N_{\text{sc}}^2} \approx \frac{\left(2\gamma_{\text{eff}} L_{\text{eff}} P_s\right)^2}{N_{\text{sc}}}, \tag{5.11}$$

where the approximation holds for large N_{sc}. Figure 5.1 shows $V_{\text{XPM}}/(2\gamma_{\text{eff}} L_{\text{eff}} P_s)^2$ as a function of N_{sc}. It can be seen that $V_{\text{XPM}} \to 0$ as $N_{\text{sc}} \to \infty$. Because of (5.5), this behavior is expected for any arbitrary distribution of $X^{(n)}$.

It is not immediately obvious from (5.11) and Figure 5.1 that XPM does not affect the capacity as $N_{\text{sc}} \to \infty$. Even though V_{XPM} scales with P_s^2/N_{sc} (and hence decreases with $1/N_{\text{sc}}$), the signal power per sub-channel (5.5) decreases at the same rate in N_{sc}, but scales with P_s/N_{sc}. Last, the noise power per sub-channel, $N_{\text{ASE}} \cdot B/N_{\text{sc}}$, decreases with $1/N_{\text{sc}}$.

To show that the channel capacity does in fact converge to that of a linear AWGN channel in the presence of XPM-induced phase noise as $N_{\text{sc}} \to \infty$, a numerical example is used. A fiber link of $L = 2000\,\text{km}$ standard single-mode fiber (SSMF) without optical dispersion compensation and with ideal distributed amplification (see Section 2.3.2) is assumed. This link setup is similar to that used by Essiambre et al. for capacity calculations [EFKW08, EKW+10]. The relevant system parameters are specified in Table 5.1.

The mutual information (in bits/s/Hz or bits/symbol) of a single sub-channel was calculated using the polar decomposition method derived in Section 3.3. All sub-channels

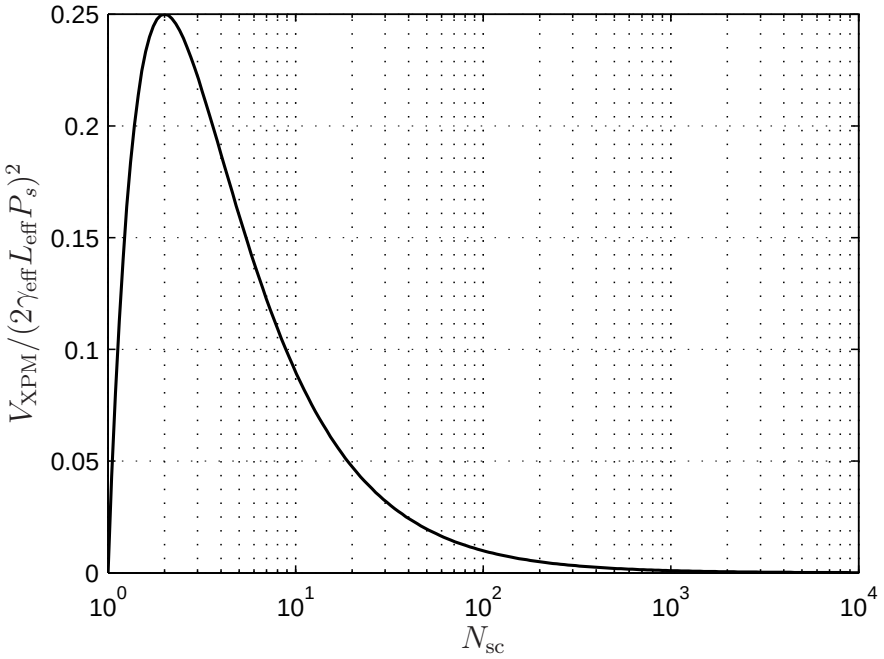

Figure 5.1: Normalized variance of XPM-induced phase noise over the number of sub-channels.

Table 5.1: System parameters used for numerical calculations.

System		
λ_S	1550 nm	
B	100 GHz	
Fiber		
L	2000 km	
α	0.2 dB / km	Table 2.3
γ_{eff}	1.3/ W / km	Tables 2.3, 2.2
Distributed amplification		
L_{eff}	2000 km	(2.140)
N_{ASE}		(2.163)
$\Delta f = f_P - f_S$	13.2 THz	
T	290 K	
$n_{\mathrm{sp,av}} = n_{\mathrm{sp}}$	≈ 1.13	(2.161)

(including our channel of interest) have complex-valued Gaussian inputs. Both decomposition mixed terms were neglected, cf. Section 3.4. As the XPM-induced phase noise samples in the sub-channels are strongly correlated, spectral loss (Section 3.6.2) was not taken into account.[4] Figure 5.2 shows the results of the described numerical calculation of mutual information.

[4]In the currently discussed scenario of sub-channels in the frequency domain, the effect we call "spectral loss" is in fact a temporal loss.

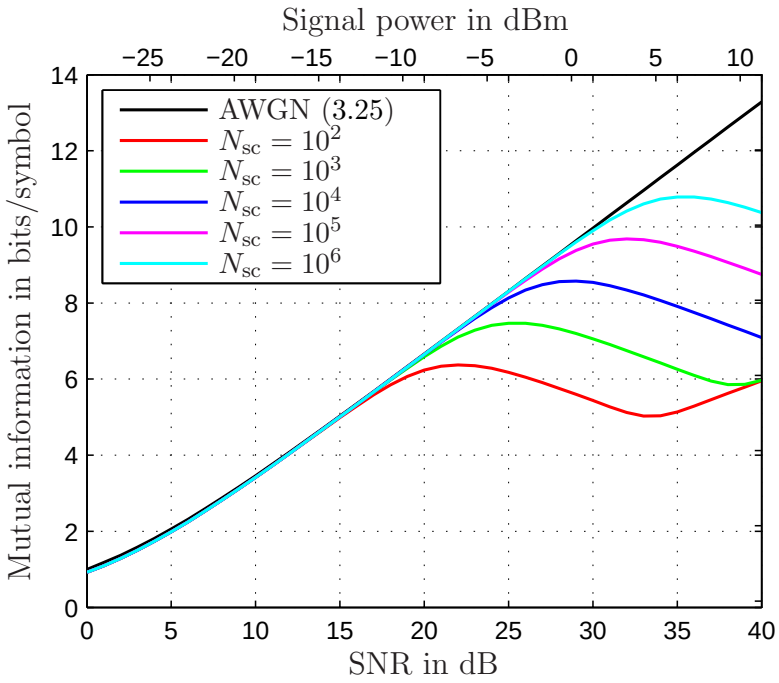

Figure 5.2: Mutual information of the channel described in Table 5.1 for N_{sc} co-propagating sub-channels with Gaussian input over SNR in dB, $10\log_{10}(P_s/P_{ASE})$. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\text{mW})$.

It can be observed that XPM significantly decreases the mutual information when the signal power is large. However, it also becomes clear that the mutual information approximates the channel capacity (3.25) of the AWGN channel as $N_{sc} \to \infty$. Among the curves shown in Figure 5.2, the case with the strongest phase noise ($N_{sc} = 10^2$, red line) allows an interesting observation. Between 22 dB and 33 dB SNR, the mutual information decreases, whereas it grows logarithmically with the SNR outside this range. The reason for this behavior is that the decomposition phase term (which is increasingly affected by the power-dependent phase noise) decreases faster than the amplitude term (which is unaffected by phase noise) grows. The phase term eventually tends to zero, while the amplitude term obeys (3.48) for large SNRs.

We conclude that XPM-induced phase noise can cause severe capacity reductions for practical fiber-optic communication systems that use multi-carrier modulation schemes (such as OFDM) at high power levels. As discussed above, the phase noise variance V_{XPM} is directly proportional to the power variance (5.10) of the co-propagating signals multiplied by $N_{sc} - 1$ and hence inversely proportional to N_{sc}. Therefore, increasing the number of sub-carriers or using modulation schemes with a low power variance (e. g. PSK) are means of decrasing the XPM impairments in multi-carrier systems. For the system model lined out in Section 5.2, however, XPM does not affect the channel capacity.

5.4 Four-wave mixing

Four-wave mixing (FWM) fundamentally differs from SPM and XPM in three aspects:

- ▷ The efficiency of FWM depends on the phase mismatch (5.4) of the interacting waves. A large phase mismatch leads to rapid oscillations of the term $e^{-\jmath\,\Delta\beta\cdot z}$ in (5.6). In contrast, SPM and XPM are always phase-matched.
- ▷ FWM leads to an energy exchange between interacting waves. As a consequence, the power spectral density (PSD) functions of the propagating signal and noise are modified.
- ▷ New waves at frequencies outside of the original spectrum can be generated. FWM leads to spectral broadening.

5.4.1 Estimate of the FWM power

Through FWM, three waves at f_n, f_p, f_q generate an additive signal at $f_m = f_n + f_p - f_q$ that we will call a *FWM product*. The total number of such FWM products in a system with N_{sc} spectral components amounts to $(N_{\mathrm{sc}}^3 - N_{\mathrm{sc}}^2)/2$. This result as well as the number of FWM products that fall on a specific frequency f_m are calculated in Appendix A. An estimate of the power of a single FWM product is given in the following.

By neglecting the SPM and XPM terms, (5.6) can be solved for each specific frequency combination. The power of the resulting FWM product is [HJKM78],[KL02b, Ch. 13]

$$P_{\mathrm{FWM}}^{(npq)} = (d_F \cdot \gamma_{\mathrm{eff}})^2 \cdot P^{(n)} \cdot P^{(p)} \cdot P^{(q)} \cdot \eta \tag{5.12}$$

with the *FWM efficiency*

$$\eta = \left| \frac{1 - \exp(-(\alpha + \jmath\,\Delta\beta) L_{\mathrm{span}})}{\alpha + \jmath\,\Delta\beta} \right|^2 \cdot \frac{1 - \cos(N_{\mathrm{spans}} \cdot \Delta\beta \cdot L_{\mathrm{span}})}{1 - \cos(\Delta\beta \cdot L_{\mathrm{span}})}, \tag{5.13}$$

where N_{spans} is the number of (identical) fiber spans of length L_{span}. The total link length is $L = N_{\mathrm{spans}} \cdot L_{\mathrm{span}}$. The *degeneracy factor* is $d_F = 1$ for degenerate FWM products and $d_F = 2$ for non-degenerate FWM products.[5]

The second factor in (5.13) extends the result to multiple optically amplified spans [Sch91]. In this context, the coherence length (2.149) plays an important role. If the signal is amplified at multiples of $2L_{\mathrm{coh}}$, i. e. $L_{\mathrm{span}} = 2NL_{\mathrm{coh}}$, $N \in \mathbb{N}$, the FWM product builds up constructively. In contrast, the FWM efficiency is minimized if $L_{\mathrm{span}} = (2N-1)L_{\mathrm{coh}}$. However, this observation cannot be used for link optimization, as the coherence length is different for every FWM product.

[5] As explained in Section 2.2.2, non-degenerate products satisfy $n \neq p, n \neq q, p \neq q$. For a given set $\{n, p, q\}$, two identical non-degenerate FWM products appear at $f_m = f_n + f_p - f_q = f_p + f_n - f_q$, cf. Appendix A. This doubles the amplitude of the resulting FWM term and a factor of $2^2 = 4 = d_F^2$ must be introduced in (5.12).

We will only consider links that completely compensate the fiber loss by optical amplification. For unamplified single-span links or multi-span links that do not compensate the loss in the last span, an additional factor $e^{-\alpha L_{\text{span}}}$ must be included in (5.12).

Using the Taylor series expansion (2.66), the phase mismatch (5.4) can be approximated as a function of the dispersion parameter (2.136) and slope parameter (2.137) as [SBW87, Sch04]

$$\begin{aligned}\Delta\beta &\approx \frac{2\pi\lambda_q^2}{c_0} \cdot \Delta f_{nq} \cdot \Delta f_{pq} \cdot \left(D + \frac{\lambda_q^2}{2c_0}(\Delta f_{nq} + \Delta f_{pq}) \cdot S \right) \\ &= \frac{2\pi c_0}{f_q^2} \cdot \Delta f_{nq} \cdot \Delta f_{pq} \cdot \left(D + \frac{c_0}{2f_q^2}(\Delta f_{nq} + \Delta f_{pq}) \cdot S \right), \end{aligned} \tag{5.14}$$

where $\Delta f_{nq} = |f_n - f_q|$. The introduced calculation of FWM products can be extended to links with variable dispersion and slope parameters, e. g. dispersion-managed systems that use dispersion-compensating fibers (DCF) [IT95].

Equation (5.12) is an estimate of the true FWM power in two aspects. First, pump depletion is neglected, i. e. the spectral components are depleted by fiber loss only, but not by the power exchange through FWM. Second, the FWM products do not produce new "second-order" FWM products themselves. Both assumptions are valid if the total power generated by FWM is small compared to the signal power [Sch04]. The accuracy of our FWM model is discussed in Section 5.4.3.

Dispersion-shifted fibers

If the optical signal is transmitted at the zero-dispersion wavelength, $\Delta\beta \approx 0$. In practice, this is the case if dispersion-shifted fibers (DSF) are used. Assuming lumped amplification after every span, (5.12) becomes

$$P_{\text{FWM}}^{(npq)} = (d_F \gamma_{\text{eff}} N_{\text{spans}} L_{\text{eff}})^2 \cdot \left(\frac{P_s}{N_{\text{sc}}} \right)^3, \tag{5.15}$$

where all spectral components are assumed to have equal power (5.5). The effective fiber length L_{eff} is given by (2.142), i. e. $N_{\text{spans}} L_{\text{eff}}$ corresponds to the total effective fiber length (2.143).

Optimum distributed amplification

For a link with optimum distributed amplification, i. e. local gain equals local loss, the FWM power can be calculated with $\alpha = 0$. In this case, (5.12) becomes

$$P_{\text{FWM}}^{(npq)} = (d_F \gamma_{\text{eff}})^2 \cdot \left(\frac{P_s}{N_{\text{sc}}} \right)^3 \cdot \frac{2(1 - \cos(\Delta\beta L))}{\Delta\beta^2}. \tag{5.16}$$

This expression illustrates the above-mentioned significance of the coherence length (2.149). If the link length L is a multiple of $2L_{\text{coh}}$, the FWM product vanishes completely. In contrast, P_{FWM} is maximum if L is an odd multiple of L_{coh}.

Finally, we can obtain the FWM power for $\alpha = 0$ and $\Delta\beta = 0$ from either (5.15) or (5.16) as

$$P_{\mathrm{FWM}}^{(npq)} = (d_F \gamma_{\mathrm{eff}} L)^2 \cdot \left(\frac{P_s}{N_{\mathrm{sc}}}\right)^3 . \tag{5.17}$$

5.4.2 Capacity calculation

It was shown in Section 5.3 that SPM and XPM do not degrade the channel capacity as the number of spectral components $N_{\mathrm{sc}} \to \infty$. Therefore, FWM remains as the only nonlinear impairment to capacity.

FWM noise and capacity calculation

The single FWM products add up to form an additive random process. Because of the large number of independent FWM products, given by (A.4) and (A.9), the resulting additive random variable can be assumed to have a Gaussian distribution due to the central limit theorem. We will refer to this additive Gaussian random process as *FWM noise*.

The number of FWM products (and hence the FWM noise power) is different for each frequency; it is largest at the center of the signal band and decreases towards the edges of the spectrum (see e. g. Figure A.1). Hence, each spectral component (or sub-channel) has a different channel capacity. Both ASE noise and FWM noise are additive with a Gaussian distribution. Therefore, each sub-channel can be seen as an AWGN channel (where the noise is partly signal-dependent). The optimum input distribution is Gaussian, $X^{(m)} \sim \mathcal{N}_{\mathbb{C}}\left(0, P^{(m)}\right)$. The channel capacity is then given by (3.1), which we can write as

$$\begin{aligned} C^{(m)} &= \Delta f \cdot \log_2\left(1 + \frac{P^{(m)}}{\Delta f \cdot N_{\mathrm{ASE}} + P_{\mathrm{FWM}}^{(m)}}\right) \\ &= \frac{B}{N_{\mathrm{sc}}} \cdot \log_2\left(1 + \frac{P_s}{B \cdot N_{\mathrm{ASE}} + N_{\mathrm{sc}} \cdot P_{\mathrm{FWM}}^{(m)}}\right), \end{aligned} \tag{5.18}$$

where $P_{\mathrm{FWM}}^{(m)}$ denotes the total FWM noise power at f_m that is obtained by summing up the powers $P_{\mathrm{FWM}}^{(npq)}$ of all FWM products that satisfy $m = n + p - q$. The second equality implies that all sub-channels have identical average signal power (5.5). This does not maximize the total channel capacity as we will discuss in the next paragraph. A lower bound to the total channel capacity can be given as

$$C \geq \sum_{m=1}^{N_{\mathrm{sc}}} C^{(m)}. \tag{5.19}$$

As $N_{\mathrm{sc}} \to \infty$, the summation becomes an integral over the signal bandwidth B.

It must be noted that by treating the spectral components as independent sub-channels, we ignore any information that may be contained in the correlation of the FWM noise

between the sub-channels. In this respect, too, the obtained capacity values are a lower bound to the actual capacity. The idea to apply decorrelation techniques to the FWM noise is not pursued in this thesis, but might yield higher and more precise capacity estimates.

Optimum input power spectral density

The PSD $\Phi_{\mathrm{FWM}}(f)$ of the FWM noise is non-white. Its exact shape depends on the used fiber type and the system parameters (cf. Sections 5.4.3 and 5.4.4), but it is always maximum at the center of the signal band and falls off towards the band's edges. Hence, the nonlinear SNR

$$\frac{P^{(m)}}{\Delta f \cdot N_{\mathrm{ASE}} + P_{\mathrm{FWM}}^{(m)}} \tag{5.20}$$

is a variable function of m.

The total channel can therefore be regarded as an additive non-white Gaussian noise channel with average power constraint $P_s = \sum_m P^{(m)}$. The capacity of such a channel is given by (3.27). It is maximized if the input signal's PSD is matched to the noise PSD according to (3.28) (water-filling).

The additional difficulty with the nonlinear fiber channel is that the noise PSD is not independent from the input signal PSD. The input signal's PSD could be optimized by water-filling to match the expected FWM noise PSD $\Phi_{\mathrm{FWM}}(f)$. This would in turn alter $\Phi_{\mathrm{FWM}}(f)$. The water-filling procedure could then be repeatedly applied until the result converges. Such an *iterative water-filling* approach would maximize the channel capacity, but is not pursued further in this thesis. As Section 5.4.4 shows, the frequency dependence of $\Phi_{\mathrm{FWM}}(f)$ in SSMFs is moderate.

The idea to spectrally shape the input signal so that it matches the expected FWM noise in order to reduce the BER in optical OFDM systems is patented [KERR09].

Required number of spectral components

Calculations of the FWM noise power spectral density and the capacity (5.19) deliver stable results very quickly as N_{sc} increases. In practice, a relatively low number of spectral components, e. g. $N_{\mathrm{sc}} = 100 \ldots 1000$, allows us to obtain reliable results for the system model of Section 5.2 which asks $N_{\mathrm{sc}} \to \infty$. Similar observations were reported for practical optical OFDM systems, for which fiber nonlinearities are found to be largely independent of the number of subcarriers [LAR07, LWP07]. A qualitative explanation for this independence of the system performance from N_{sc} can be derived from (5.18) as follows. The argument inside the logarithm is independent from N_{sc} if $P_{\mathrm{FWM}}^{(m)}$ scales with $1/N_{\mathrm{sc}}$. From (5.5) and (5.12) we see that the power $P_{\mathrm{FWM}}^{(npq)}$ of a single FWM product scales with $1/N_{\mathrm{sc}}^3$. Hence, the number of FWM products must be proportional to N_{sc}^2. The number $N_{\mathrm{FWM,NDG}}$ of non-degenerate FWM product, given in (A.9), satisfies this condition. Finally, we note that – unless N_{sc} is very small – $N_{\mathrm{FWM,NDG}}$ dominates the number of degenerate FWM products (A.4) (see e. g. Figure A.1).

5.4.3 Phase-matched FWM

As mentioned above, the FWM process is nearly phase-matched, i. e. $\Delta\beta \approx 0$, if light propagates at the zero-dispersion wavelength. The FWM efficiency and hence the FWM power are then maximum, so that a link that consists of DSFs is a worst-case scenario. This makes the DSF an inferior channel in practice. However, it is a very useful object for the following discussion of fiber nonlinearities. We assume a DSF with parameters given in Table 2.3. We will neglect the (non-zero) dispersion slope, so that $\Delta\beta = 0$ exactly.

FWM noise power spectral density

Each FWM product now has identical power given by (5.15). The total FWM power is obtained from a simple multiplication of (5.15) with the number of degenerate (A.3) and non-degenerate FWM products (A.8). As $N_{\mathrm{sc}} \to \infty$, the resulting discrete spectrum approaches the continuous FWM power spectral density. As mentioned before, this convergence is very fast; for the results that follow, we have used $N_{\mathrm{sc}} = 1000$.

Figure 5.3 depicts the FWM noise PSD $\Phi_{\mathrm{FWM}}(f)$ normalized to $(N_{\mathrm{spans}}^2 L_{\mathrm{eff}}^2 P_s^3)/B$. Because

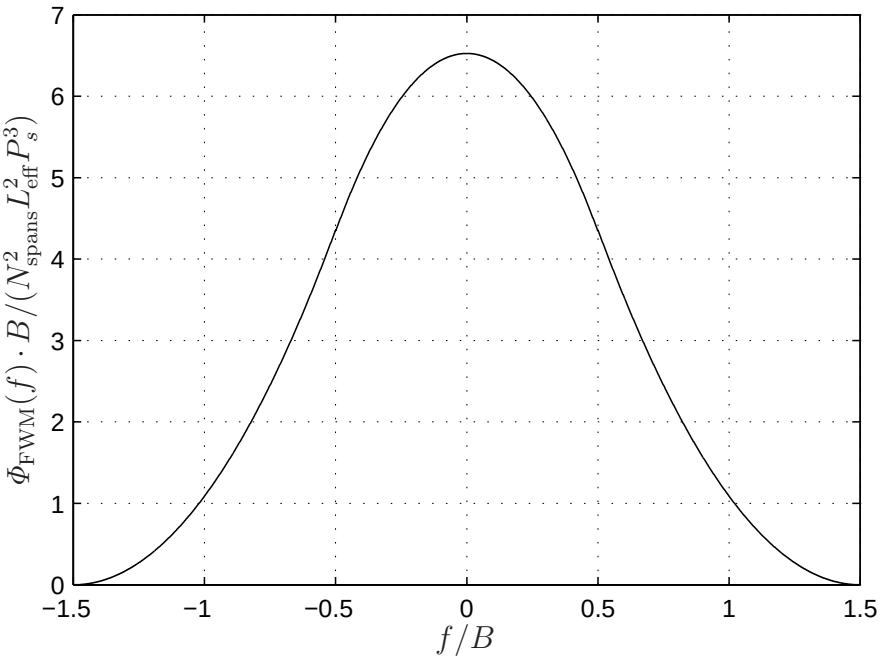

Figure 5.3: Normalized PSD of FWM noise over normalized frequency. The normalized PSD has units of $(\mathrm{W\,km})^{-2}$.

of the chosen normalization, the depicted result is rather general and can be used to calculate the FWM noise impairment for different signal powers, link lengths and amplification schemes. From the normalized PSD (and from (5.15)), we see that FWM scales quadratically with distance and and cubically with power. Scaling the FWM impairment with a secondary parameter such as the mean nonlinear phase shift $\phi_{\mathrm{NL}} = \gamma_{\mathrm{eff}} P_s N_{\mathrm{spans}} L_{\mathrm{eff}}$ is therefore not feasible.

Figure 5.3 shows that new signals are generated outside of the original bandwidth (which is $|f| \leq B/2$). The (one-sided) 3-dB bandwidth of the FWM noise PSD is approximately $0.64 \cdot B$. The ratio of in-band to out-of-band FWM noise power is approximately 3 dB. In other words, one third of the total FWM power is generated outside of the original spectrum. This power appears as interference in neighboring WDM channels. Simultaneously, it is lost to the original band. It can be argued that this loss effectively reduces the usable signal power and thus decreases capacity even for systems that are able to mitigate all nonlinear effects that occur inside the original signal band [NM02]. As pump depletion is neglected in our calculations, we do not take this loss into account. This is not a critical restriction as the absolute amount of signal power that leaks out of the original band is small; it is further reduced if the fiber has non-zero dispersion.

Accuracy of the FWM noise model

As mentioned in Section 5.4.1, the estimate of the FWM noise power becomes inaccurate at high power levels for two reasons. On the one hand, pump depletion is neglected, so that both the signal power and the FWM noise power are overestimated at high input power levels. On the other hand, sufficiently powerful FWM products produce "second-order" FWM products themselves. These second- and higher-order FWM products are not captured by our model. At very large power, the actual PSD can be significantly broader than that shown in Figure 5.3.

To gain confidence in our FWM model, it is necessary to demonstrate that the discussed restrictions do not significantly affect the calculation of the channel capacity. This can be achieved by means of a numerical simulation of the fiber link. Again, we use the parameters of Table 5.1, but assume a DSF with parameters given in Table 2.3 and $\Delta\beta = 0$. We use the closed-form solution (2.122) of the NLSE, so Algorithm 2.1 is not required for the simulation of this channel. For the numerical simulation, each spectral component is modulated with a 4-QAM symbol. The received signal constellation at an SNR of 10 dB is shown in Figure 5.4. The constellation plot does not show ASE noise. The phase rotation is mainly due to XPM ($\mathcal{E}\{\phi_{\mathrm{SPM}}\} \ll \mathcal{E}\{\phi_{\mathrm{XPM}}\}$), whereas the circularly symmetric additive noise is due to FWM.

The actual values of the received average signal power $P_{s,\mathrm{rx}}$ and of the FWM noise variance can be estimated from the received QAM symbols. Figure 5.5 shows the received power as a function of the transmitted power. It can be seen that FWM does not significantly deplete the signal power for low power levels ($10\log_{10}(P_s/1\,\mathrm{mW}) < 14\,\mathrm{dBm}$). At larger transmit power levels, however, the received signal power is significantly reduced by FWM. Eventually, all signal power is converted to FWM noise. Figure 5.5 allows us to specify power regions for which the FWM model without pump depletion is valid. It must be noted again, however, that the considered scenario maximizes FWM because of the large nonlinear coefficient and zero phase-mismatch of DSFs and because of the distributed amplification scheme.

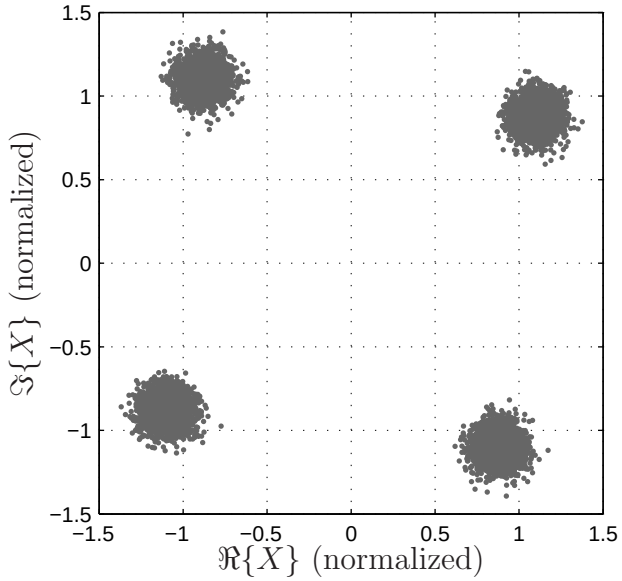

Figure 5.4: Received signal constellation of the numerically simulated DSF link described in the text at 10 dB SNR, corresponding to $10\log_{10}(P_s) \approx -18.8\,\text{dBm}$. The plot does not show ASE noise.

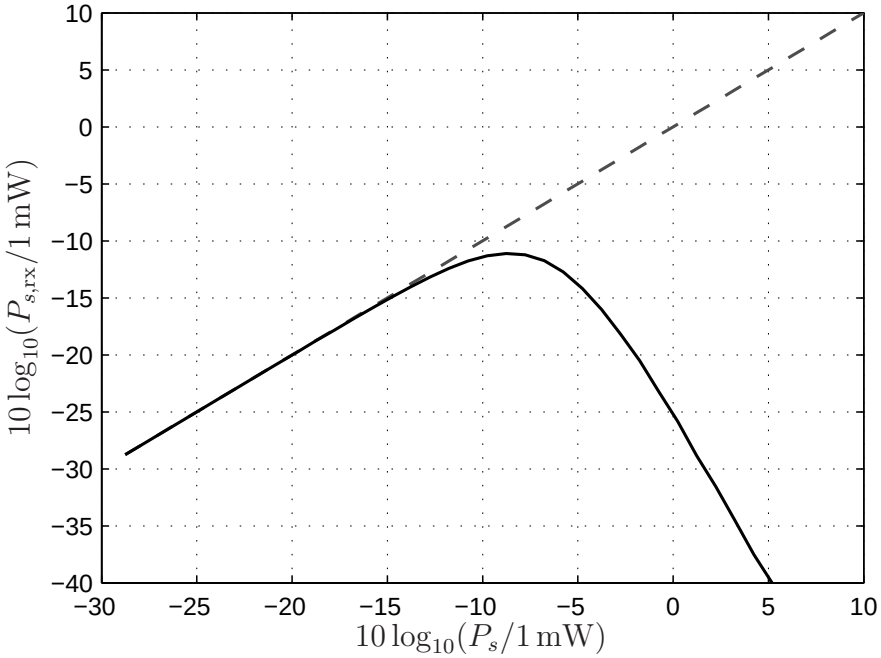

Figure 5.5: Received average signal power in dBm over average transmitted signal power in dBm. The angle bisector (dashed line) is shown for reference.

As we have seen, pump depletion leads to a systematic overestimation of the signal power in the highly nonlinear regime. This in turn leads to an overestimation of the generated FWM noise. We can observe this effect in Figure 5.6, which depicts the ratio of the

in-band FWM noise power

$$P_{\mathrm{FWM,in-band}} = \int_{-B/2}^{B/2} \Phi_{\mathrm{FWM}}(f)\mathrm{d}f, \tag{5.21}$$

to the signal power P_s in dB.[6] At an SNR of approximately 22 dB this ratio exceeds

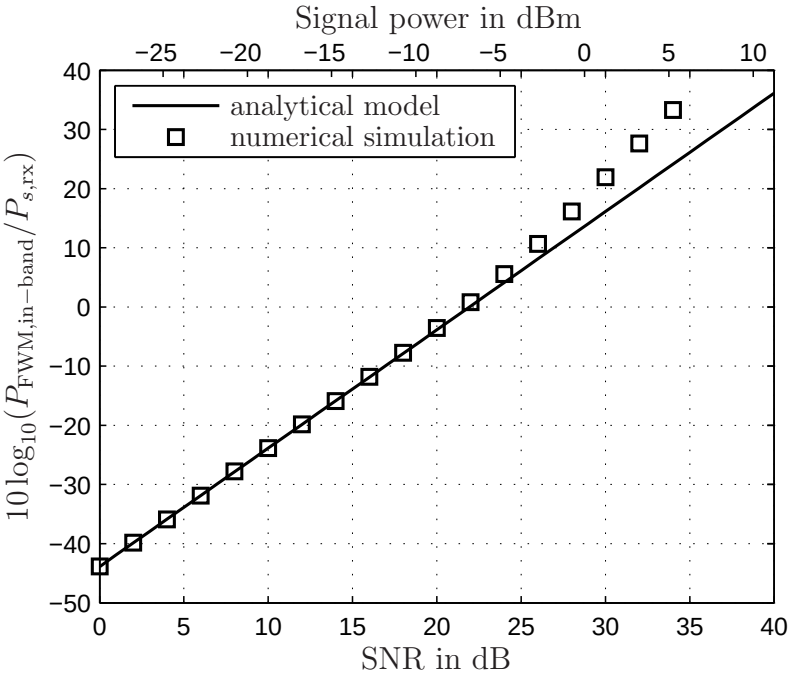

Figure 5.6: Ratio of in-band FWM noise power to received signal power in dB over SNR in dB. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$.

0 dB, i. e. our model predicts an amount of FWM noise power that exceeds the total transmit power that was coupled into the fiber. This is an obvious violation of the energy conservation principle.

Figure 5.6 also shows the same power ratio as obtained from the numerical simulation. Luckily, we can observe that the systematic overestimation of both signal power and FWM noise power is similar, so that the actual power ratio is very close to the analytically predicted one even at very large SNRs. This observation is particularly important because the ratio of signal power and FWM noise power determines the capacity (5.18) when the SNR is large and ASE noise plays a subordinate role.

In conclusion, we find that the analytical FWM model is applicable for capacity calculations even at power levels for which it is not physically valid. More importantly, however, it is fully valid (in the sense that pump depletion is negligible) at those power levels at which the channel capacity reaches its maximum, as we shall see next.

[6] Since the ratio of total to in-band FWM noise power is approximately 1.5, the ratio of $P_{\mathrm{FWM,total}}$ to P_s is approximately 1.76 dB higher than that depicted in Figure 5.6.

Comparison of amplification schemes

Next, we use the FWM noise PSD of Figure 5.3 to calculate the channel capacity of a given link. A capacity value in units of bits/s/Hz can be obtained by calculating (5.19) and dividing the result by the bandwidth B. As a first step, the capacities achievable with different amplification schemes are compared for a link with $L = 2000\,\text{km}$ of DSF and $B = 100\,\text{GHz}$ bandwidth. The parameters of the optimum distributed amplification scheme are given in Table 5.1. For lumped amplification, a noise figure of $F_n = 2$ is assumed. From a comparison of (5.15) and (5.17), it is expected that distributed amplification causes more FWM noise because of the constantly high signal power. Similarly, a system with lumped amplification is expected to produce higher FWM noise power if the span length is shorter, as this increases the total effective fiber length given by $N_{\text{spans}}L_{\text{eff}}$. On the other hand, a comparison of (2.156) and (2.163) shows that ideally distributed amplification produces less ASE noise than lumped amplification.

Figure 5.7 shows the capacity of the described system with distributed amplification and lumped amplification with $L_{\text{span}} = 50\,\text{km}$ and $L_{\text{span}} = 100\,\text{km}$. It can be seen that

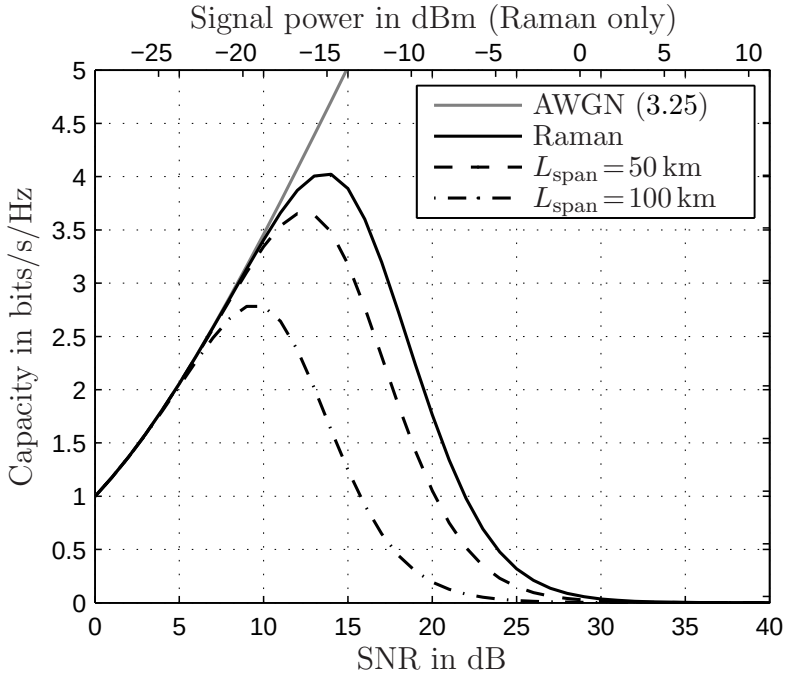

Figure 5.7: Channel capacity in bits/s/Hz of 2000-km DSF link with distributed and lumped amplification over SNR in dB, $10\log_{10}(P_s/P_{\text{ASE}})$. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\text{mW})$, for the distributed amplification setup.

decreasing the fiber length per span from $L_{\text{span}} = 100\,\text{km}$ to $L_{\text{span}} = 50\,\text{km}$ increases the maximum capacity by almost one bit for lumped amplification. A further decrease of the span length (the optimum length in terms of ASE noise is $L_{\text{span,opt}} = 1/\alpha \approx 21.7\,\text{km}$, see Section 2.3.1) yields only marginal capacity gains. Despite the higher FWM noise power, distributed amplification yields the largest capacity due to the low ASE noise power. The

top x-axis in Figure 5.7 shows the average signal power P_s in dBm for this case of Raman amplification. Although the capacity curves peak at similar SNR values, the signal power values are quite different for the three depicted cases due to the different ASE noise levels. The system with distributed amplification achieves the maximum capacity at an SNR of 13.6 dB, which corresponds to an input power of $10\log_{10}(P_s/1\,\mathrm{mW}) = -15.2\,\mathrm{dBm}$. The systems with lumped amplification have optimum input powers of $-10.9\,\mathrm{dBm}$ ($L_{\mathrm{span}} = 50\,\mathrm{km}$) and $-5.9\,\mathrm{dBm}$ ($L_{\mathrm{span}} = 100\,\mathrm{km}$). In conclusion, the maximum capacity can be obtained by using ideal distributed amplification. This was expected from the discussion in Section 2.3.2. However, if the system needs to be operated at signal power levels above those optimum for the Raman case, it can be beneficial to use lumped amplification and optimize the amplifier spacing with respect to both ASE and FWM noise generation.

Channel capacity as a function of distance

We have seen that the channel capacity as a function of the signal power exhibits a distinct maximum. At power levels below this point, the channel is quasi-linear and its capacity is limited by ASE noise. Beyond the optimum signal power, FWM becomes the dominant impairment and any further SNR increase reduces the channel capacity. The optimum capacity value and the corresponding optimum input power depend on the link length and on the used amplification scheme. Figure 5.8 shows the maximum capacity as a function of the link length (left x-axis, black line). Also shown is the input power at which the capacity is maximized (right x-axis, red line). The results assume DSF with ideal Raman amplification. As expected, the capacity decreases with increasing link length because of

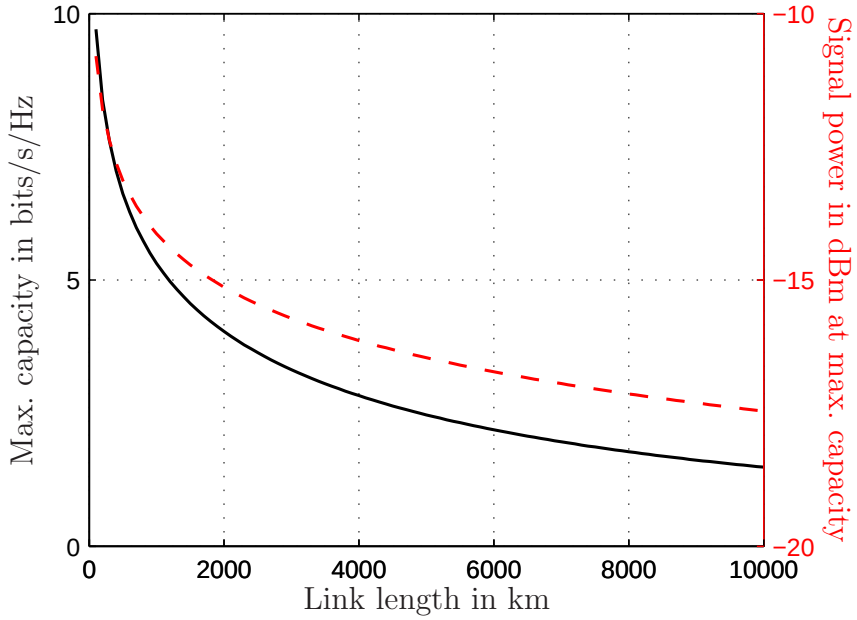

Figure 5.8: Maximum channel capacity in bits/s/Hz (black solid line) and optimum signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$, (red dashed line) over DSF length in km. Ideal distributed amplification is used.

an increase in ASE and FWM noise. The optimum signal power decreases as well. In short links, high power levels can be used to increase the SNR without generating too much nonlinear impairments. As the link length increases, this advantage of high signal power is removed by the increase in FWM noise. Consequently, the optimum input power continuously decreases.

5.4.4 FWM in standard single-mode fibers

Due to the non-zero dispersion in standard single-mode fibers (SSMFs), the phase mismatch (5.4) is a variable function of the interacting frequencies. While the number of FWM products falling on each frequency is still the same as in the DSF case, the power (5.12) of each individual FWM product varies. The power of each FWM product and of the total FWM noise process is decreased by phase mismatch, so that SSMFs suffer less from FWM than DSFs do.

Power spectral density of FWM noise

The FWM noise power must be calculated separately for each FWM product in order to find the FWM noise PSD. The resulting PSD's shape and magnitude depend on the system parameters B, N_{spans}, L_{span} and on the amplification scheme. A normalization such as the one used in Figure 5.3 is therefore not generally possible in the SSMF case.

The FWM noise PSD was calculated for an SSMF link with parameters given in Table 5.1. The SSMF parameters are taken from Table 2.3. For lumped amplification, $F_n = 2$ and $L_{\text{span}} = 50\,\text{km}$ was assumed. The resulting PSD functions for link lengths of 500 km, 1000 km and 2000 km are shown in Figure 5.9.

In comparison to Figure 5.3, we note that the FWM noise PSD is significantly different in DSFs and SSMFs. Because of the different scaling, the magnitude difference is not immediately obvious. The maximum value depicted in Figure 5.3 scaled to the x-axis of Figure 5.9 is $\Phi_{\text{FWM}}(f=0)/P_s^3 \approx 2.6 \cdot 10^{-4}\,\text{W}^{-2}\,\text{Hz}^{-1}$. The corresponding SSMF value shown in Figure 5.9 is $\Phi_{\text{FWM}}(f=0)/P_s^3 \approx 5.3 \cdot 10^{-7}\,\text{W}^{-2}\,\text{Hz}^{-1}$.

By comparing the PSD functions of different link lengths, we observe that the magnitude does not increase quadratically with L or L_{eff} as in the phase-matched case. This is in accordance with (5.12) and (5.16). Figure 5.10 shows the total FWM noise power as a function of link length for ideal Raman amplification. The derivative of the depicted function with respect to L shows that the FWM noise power increases slightly faster than linearly with L for the first 400 km, but slower than linearly for $L > 400\,\text{km}$. This point is revisited at the end of this section.

The FWM noise PSD is considerably more confined to the original signal band. For the 2000-km link with distributed amplification, we obtain a (one-sided) 3-dB bandwidth of approximately $0.49 \cdot B$. The ratio of in-band to out-of-band FWM noise power is approximately 26.5 dB. In contrast to the DSF channel, power leakage from the original spectrum is almost negligible.

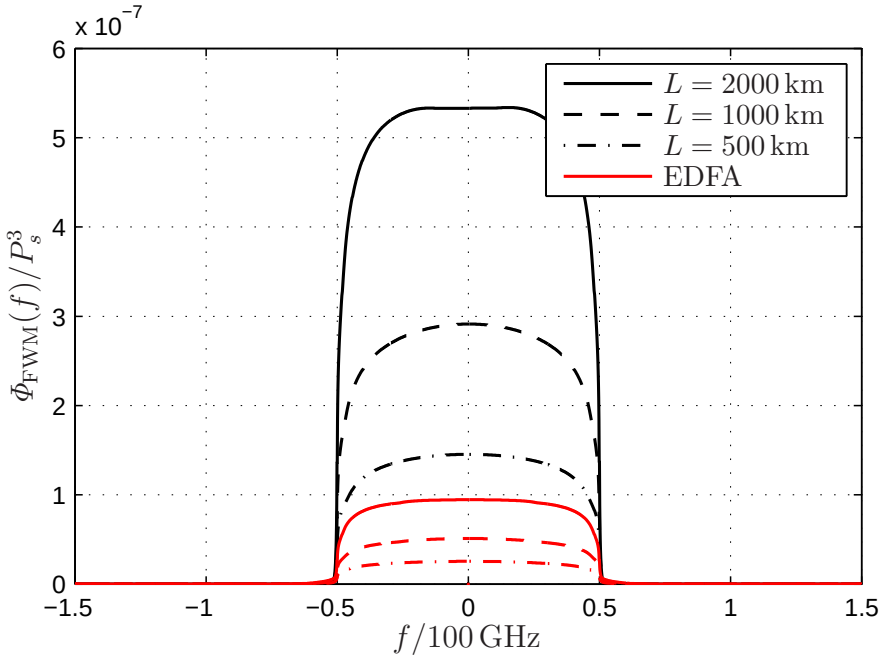

Figure 5.9: Normalized PSD of FWM noise for link lengths of $L = 500\,\mathrm{km}, 1000\,\mathrm{km}, 2000\,\mathrm{km}$ and distributed (black) or lumped (red) amplification. The normalized PSD has units of $\mathrm{W}^{-2}\,\mathrm{Hz}^{-1}$.

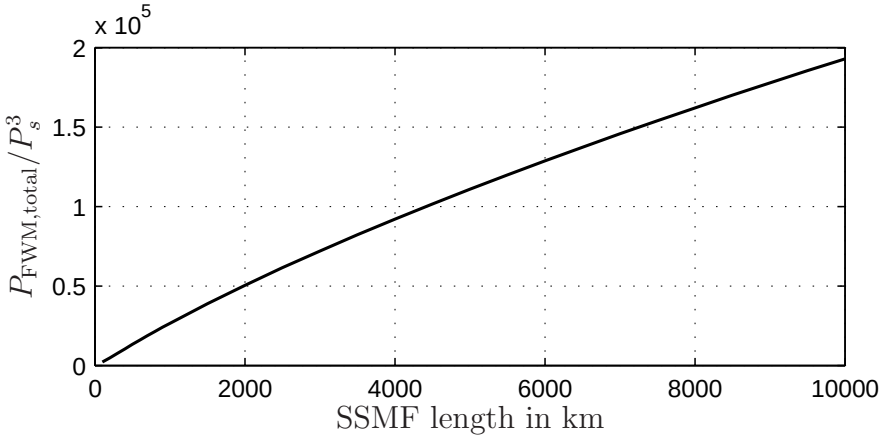

Figure 5.10: Total FWM noise power normalized to P_s^3 as a function of link length L.

Because of the less efficient FWM noise generation, the analytical FWM model remains physically plausible for a larger SNR range than in the DSF case. The ratio of in-band FWM noise power to signal power for the 2000-km Raman link lies approximately 26.6 dB below the line depicted in Figure 5.6. A power ratio of 0 dB is reached at approximately 35 dB SNR, corresponding to an approximate average signal power of 6.2 dBm.

Channel capacity of SSMF links

The channel capacity of the SSMF can be obtained as explained in Section 5.4.3. The results for the 2000-km link are shown in Figure 5.11.

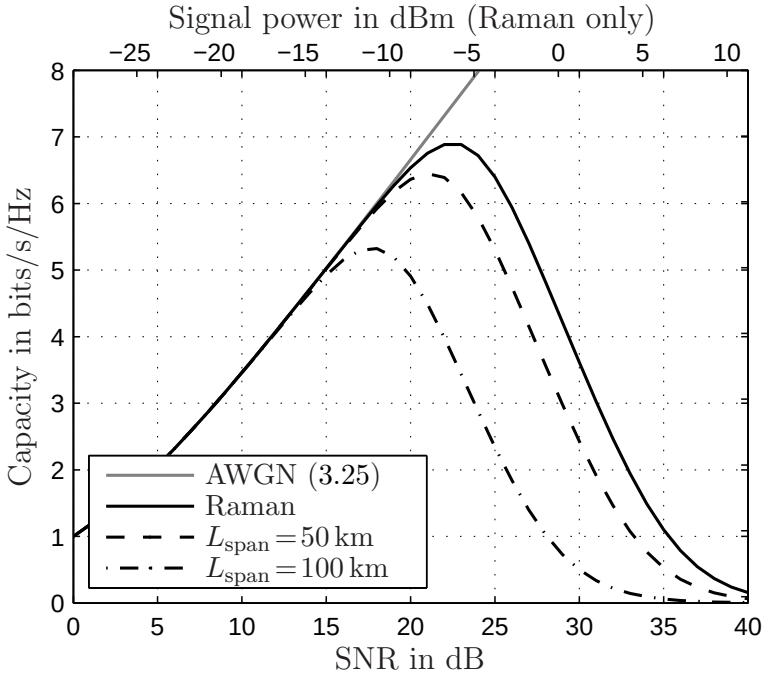

Figure 5.11: Channel capacity in bits/s/Hz of 2000-km SSMF link with distributed and lumped amplification over SNR in dB, $10\log_{10}(P_s/P_{\text{ASE}})$. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\text{mW})$, for the distributed amplification setup.

In comparison to the DSF results of Figure 5.7, we can see that the maximum capacity value is increased. This capacity gain is due to chromatic dispersion reducing the FWM efficiency, but also due to the reduced γ_{eff} of SSMF. The maximum capacity value of $C \approx 6.9\,\text{bit/s / Hz}$ is achieved with ideal Raman amplification. The capacity is maximized at an SNR of approximately 22.5 dB, corresponding to an approximate average signal power of $-6.25\,\text{dBm}$. Similar to the DSF channel, the maximum capacity (6.4 bit/s / Hz) achievable with lumped amplification lies in the same SNR region (21 dB), but at a higher signal power ($-1.9\,\text{dBm}$).

Pump depletion is much less of an issue for the SSMF channel. The ratio of FWM noise power to signal power at 22.5 dB SNR (where the capacity peaks) is approximately $-25.5\,\text{dB}$.

The investigated link is identical to the one used for the analysis of XPM that lead to the results shown in Figure 5.2. Although XPM is not a fundamentally relevant effect as discussed in Section 5.3, it is interesting to relate the results obtained separately for XPM and FWM. By comparison of Figures 5.2 and 5.11, we observe that $N_{\text{sc}} \gtrsim 1000$ is sufficient to effectively suppress XPM at the capacity-maximizing SNR. This number

of sub-carriers is realistic for practical multi-carrier systems. We can conclude that such systems should be operated at the optimum SNR as obtained from the analysis of FWM. At this SNR, XPM is negligible if the number of sub-carriers is sufficiently large (but still realistic).

Scaling of capacity and optimum signal power with distance

Finally, the capacity as a function of link length is discussed. With ideal Raman amplification, we obtain the results shown in Figure 5.12. As expected, the maximum capacity

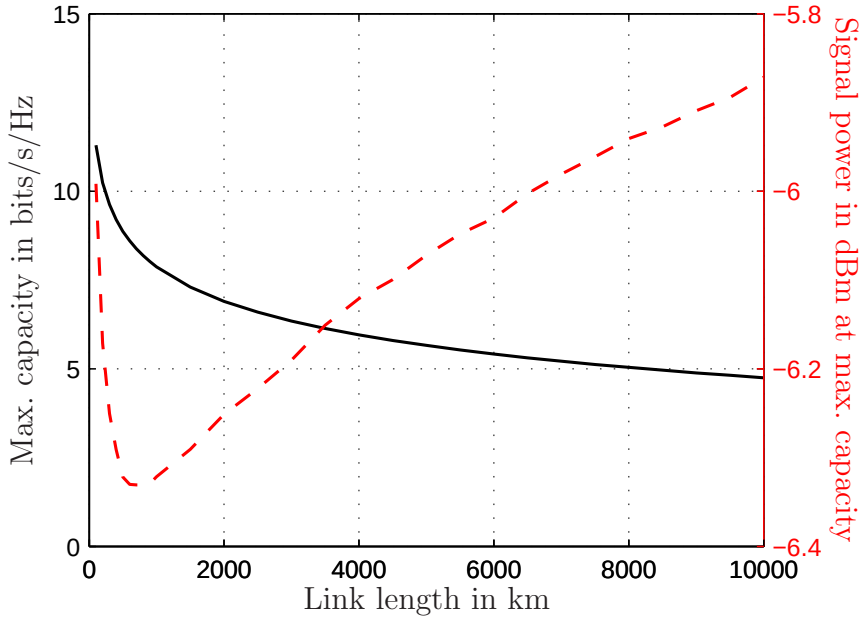

Figure 5.12: Maximum channel capacity in bits/s/Hz (black solid line) and optimum signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$, (red dashed line) over SSMF length in km. Ideal distributed amplification is used.

decreases with increasing fiber length on account of accumulating ASE and FWM noise. The optimum average signal power at which the capacity is maximized, however, does not decrease as it did for the DSF (see Figure 5.8). Instead, it exhibits a minimum in the $L = 600 \ldots 800\,\mathrm{km}$ region and shows a slight increase with distance after that point. The reason for this different behavior lies in the length dependence of the FWM noise power in different fiber types.

The capacity is largest when the nonlinear SNR (5.20), i. e. the quotient of signal power and the sum of ASE and FWM noise power, is maximized. At this point, the power levels of ASE and FWM noise will always have comparable magnitude. (If FWM was dominant, the nonlinear SNR could be increased by decreasing the signal power; vice versa, if ASE was dominant, the nonlinear SNR would increase with increasing P_s.) In the DSF, the FWM noise grows quadratically with L, whereas the ASE noise power grows linearly with L. Therefore, as L grows, the ratio of ASE and FWM noise powers becomes

imbalanced (in favor of FWM) and the optimum P_s decreases in order to maximize the nonlinear SNR and hence the capacity. The length dependence of the FWM noise power in SSMFs, shown in Figure 5.10, is qualitatively different. It grows faster than linearly for $L \lesssim 400\,\mathrm{km}$ and slower than linearly for $L \gtrsim 400\,\mathrm{km}$. This is reflected in the optimum average signal power depicted in Figure 5.12. Beyond its minimum, P_s must increase with distance to restore the balance of ASE and FWM noise.

5.5 Nonlinear signal-noise interaction

5.5.1 Channel capacity including nonlinear signal-noise interaction

So far, we have considered AWGN (from optical amplification) and the nonlinear interaction of signal components due to the Kerr effect as the two capacity-limiting physical effects. Because the ASE noise is added to the signal on the link and co-propagates with it, the nonlinear interaction of signal and noise must also be discussed.

We can immediately conclude from (5.11) that the signal-noise interaction due to XPM is not relevant for the channel capacity. The variance of the XPM-induced phase noise is increased by ASE noise, but tends to zero nevertheless as $N_{\mathrm{sc}} \to \infty$.

The situation is different for SPM. The SPM-induced phase shift (5.7), which is deterministic in the absence of ASE, becomes a random variable if noise is taken into account. We can write it as

$$\phi_{\mathrm{SPM}} = \gamma_{\mathrm{eff}} \cdot \int_0^L P^{(m)}(z)\mathrm{d}z, \tag{5.22}$$

where $P^{(m)}(z)$ is the total power at frequency f_m at fiber length z. With the noise field variable $N^{(m)}(z)$, we have

$$\begin{aligned} P^{(m)}(z) &= |X^{(m)}(z) + N^{(m)}(z)|^2 \\ &= \underbrace{|X^{(m)}(z)|^2}_{\text{signal}-\text{signal}} + \underbrace{2\Re\{X^{(m)}(z) \cdot N^{(m)^*}(z)\}}_{\text{signal}-\text{ASE}} + \underbrace{|N^{(m)}(z)|^2}_{\text{ASE}-\text{ASE}}. \end{aligned} \tag{5.23}$$

The signal-signal beat term in the summation is responsible for the deterministic phase shift (5.7). The ASE-ASE beat term introduces a random phase shift (with central (scaled) χ^2 distribution). However, at SNRs that are large enough for fiber nonlinearities to appear, this term is very small compared to the signal-ASE beat term and is therefore neglected in the following. For a given realization x of $X^{(m)}(z)$, the signal-ASE term has a Gaussian distribution with variance $2 \cdot |x|^2 \cdot \mathcal{E}\left\{|N^{(m)}(z)|^2\right\}$.

We will assume ideal Raman amplification in the following, but results for other amplification schemes can be obtained in a similar way. At every differential fiber piece along the link at location z, an ASE noise sample is added to the signal and co-propagates with it for the remaining link length $L - z$. All noise samples are i. i. d., so their variances must be added to determine the variance V_{SPM} of ϕ_{SPM}. This differential summation can be

formulated as an integral and leads to the solution

$$V_{\mathrm{SPM}} = \gamma_{\mathrm{eff}} \cdot 2 \cdot |x|^2 \cdot n_{\mathrm{sp,av}} \cdot h \cdot f_S \cdot \alpha \cdot B \cdot \int_0^L (L-z)^2 \mathrm{d}z$$
$$= \frac{2}{3} \cdot \gamma_{\mathrm{eff}} \cdot |x|^2 \cdot \underbrace{n_{\mathrm{sp,av}} \cdot h \cdot f_S \cdot \alpha \cdot B \cdot L}_{P_{\mathrm{ASE}}} \cdot L^2. \tag{5.24}$$

A simple numerical example indicates that the effect of this phase noise on the capacity is not dramatic. In Section 5.4.4, the capacity for the 2000-km SSMF link with Raman amplification was found to peak at a signal power of $-6.25\,\mathrm{dBm}$. If we insert this value as $|x|^2$ into (5.24), we get $V_{\mathrm{SPM}} \approx 1.4 \cdot 10^{-3}$.

Signal and noise also interact through FWM. ASE increases the power of each FWM product, which is still given by (5.12). However, the power of each participating spectral component in (5.12) is increased by the average ASE noise power at the respective frequency. Since the FWM efficiency and thus the normalized FWM noise PSD depend on the link length (see Section 5.4.4), a length-averaging approach such as (5.24) is not easily possible for FWM. Instead, we assume a worst-case scenario in which all the ASE noise is added at the transmitter and propagates with the signal for the entire link length. The FWM noise power then scales with

$$(P_s + P_{\mathrm{ASE}})^3 = \underbrace{P_s^3}_{\text{signal}-\text{signal}} + \underbrace{3P_s^2 P_{\mathrm{ASE}} + 3P_s P_{\mathrm{ASE}}^2}_{\text{signal}-\text{ASE}} + \underbrace{P_{\mathrm{ASE}}^3}_{\text{ASE}-\text{ASE}}. \tag{5.25}$$

The total capacity (5.19) including nonlinear signal-noise interactions is calculated from the individual capacity values of the spectral components as before. In the absence of phase noise, each sub-channel was modeled as an AWGN channel with capacity given by (5.18). Due to the presence of phase noise, this model has to be slightly modified. The sub-channels are now partially coherent (3.81). The phase noise has a wrapped Gaussian distribution (3.67) whose parameter σ^2 equals (5.24). Because the phase noise is small, the capacity-achieving input distribution is nearly Gaussian, see Section 3.6.1. Using the decomposition method developed in Chapter 3, the influence of SPM (including phase noise and the resulting spectral loss) and FWM on the channel capacity can be determined. One detail that requires attention in the calculation of both amplitude and phase term of the decomposition is that each amplitude level $x_{\shortparallel} = |x|$ experiences a different degree of spectral loss because the phase noise variance (5.24) is a function of $|x|^2$.

Figure 5.13 shows the channel capacity of the 2000-km link with ideal Raman amplification taking signal-ASE interaction through SPM and FWM into account (solid line). The depicted markers are taken from Figure 5.11 and show the result without nonlinear signal-noise interactions for comparison. Mixed term II was neglected in the calculation; this explains the small offset visible at low SNRs, cf. Figure 3.3. A difference between the results with and without signal-ASE interactions is hardly visible. We can conclude that the channel capacity is limited by ASE noise and by nonlinear signal-signal interactions, but not by nonlinear interactions of signal and noise. Consequently, the AWGN channel

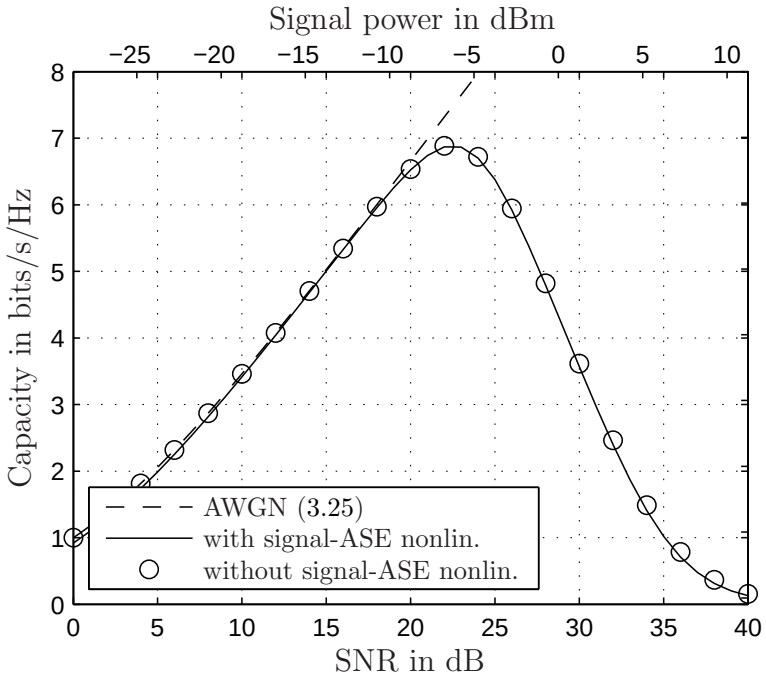

Figure 5.13: Channel capacity in bits/s/Hz of 2000-km SSMF link with ideal distributed amplification over SNR in dB, $10\log_{10}(P_s/P_{\mathrm{ASE}})$ with (solid line) and without (markers) nonlinear signal-ASE interaction. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$.

model used for each sub-channels remains valid and the optimum input distribution is Gaussian.

The described numerical calculations were repeated for fiber lengths between $L=500\,\mathrm{km}$ and $L=10000\,\mathrm{km}$, both for distributed and lumped amplification (with $L_{\mathrm{span}}=50\,\mathrm{km}$ and $L_{\mathrm{span}}=100\,\mathrm{km}$). In all cases, nonlinear signal-ASE and ASE-ASE interaction remained negligibly small compared to the two capacity-limiting processes, ASE noise and nonlinear signal-signal interaction.

5.5.2 Channel capacity with compensation of fiber nonlinearities

The results reported so far were obtained under the assumption that the nonlinear interaction of the signal with itself imposes an irreversible impairment. However, as discussed in Section 5.1.1, it is possible in principle to compensate these nonlinear distortions, e. g. using electronic predistortion. Even if a system employs such a compensation scheme to remove all nonlinear signal-signal interactions, the nonlinear interaction of signal and ASE noise (and of ASE noise with itself) remains. Although these signal-ASE (and ASE-ASE) interactions are weak as shown in Section 5.5.1, they are the only truly fundamental nonlinear impairment that remains in a fiber-optic communication system that employs all conceivable equalization mechanisms.

The discussion of Section 5.5.1 is easily modified to the case where only signal-ASE and ASE-ASE interactions remain. Electronic predistortion removes the constant phase shift induced by SPM and XPM; this constant phase shift does not affect the capacity anyway. The variable SPM part described in Section 5.5.1 is due to signal-ASE and ASE-ASE interaction and remains as it is. In contrast, FWM is significantly reduced by fiber nonlinearity compensation. The FWM noise considered in Section 5.5.1 scales with $(P_s + P_{\mathrm{ASE}})^3$. The compensation of signal-signal nonlinearities removes the term P_s^3 in (5.25), all other terms remain.

Taking SPM-induced phase noise and spectral loss as well as signal-ASE and ASE-ASE FWM noise into account, we obtain the channel capacity depicted in Figure 5.14 for the 2000-km SSMF link with ideal Raman amplification (solid line). The capacity curve still

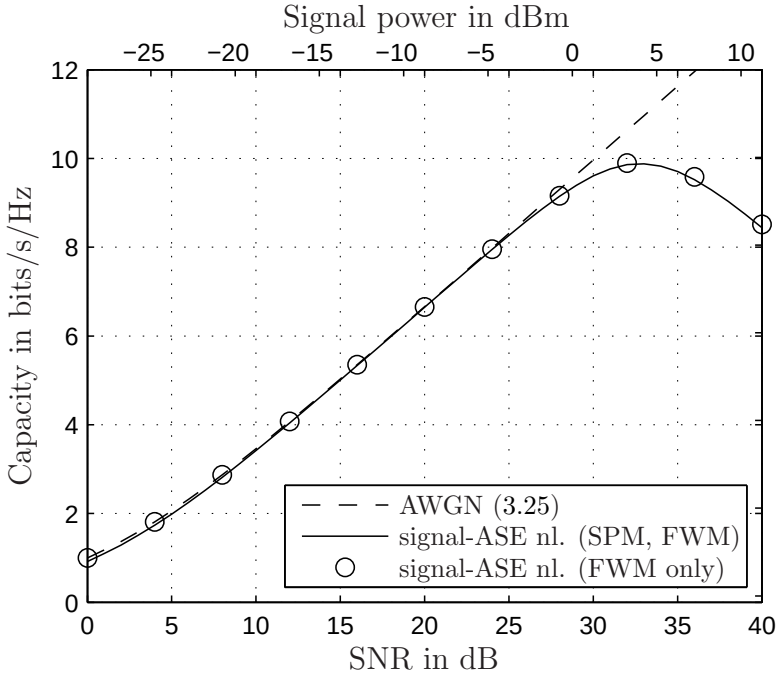

Figure 5.14: Channel capacity in bits/s/Hz of 2000-km SSMF link with ideal distributed amplification over SNR in dB, $10\log_{10}(P_s/P_{\mathrm{ASE}})$ impaired only by nonlinear signal-ASE interaction. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$.

has a distinct maximum that marks the boundary between the regions in which the channel is limited by ASE noise and by fiber nonlinearities, respectively. However, the maximum capacity increases to almost 10 bit/s / Hz; the optimum input power is approximately 4.2 dBm. The result does not change if SPM is neglected and only signal-ASE and ASE-ASE interaction through FWM is considered (markers shown in Figure 5.14). Hence, the effect of SPM is negligible compared to FWM. This implies that the AWGN channel model used for each sub-channels remains valid and the optimum input distribution is Gaussian.

The capacity curve shown in Figure 5.14 matches a published result [EKW+10, Fig. 36(4)] obtained for the same channel remarkably well. Both curves achieve the maximum ca-

pacity at the same SNR value of approximately 33 dB. In [EKW+10], a 16-ring input constellation was used. On the AWGN channel with an SNR of 33 dB, this constellation achieves approximately 0.8 bit/s / Hz less capacity than a Gaussian input [EKW+10, Fig. 16]. This explains the small mismatch (< 1 bit/s / Hz) between the two maximum capacity values for the nonlinear channel. Details of the model used in [EKW+10] are explained in Section 6.2.2.

The main conclusion from Figure 5.14 is that ASE noise and nonlinear signal-ASE and ASE-ASE interactions limit the capacity even if all other fiber effects are perfectly equalized. For any link, there is a maximum capacity and an optimum input power level that achieves this capacity.

The capacity calculation was repeated for larger signal bandwidths ($B = 300$ GHz, $B = 500$ GHz). The resulting capacities (in bits/s/Hz) were not significantly different from those shown in Figure 5.14 for $B = 100$ GHz (less than 0.1 bit/s / Hz difference between $B = 100$ GHz and $B = 500$ GHz at the capacity-maximizing SNR). The effect of signal-ASE interaction through SPM increases (relative to FWM) with larger bandwidths, but remains very small (approximately 0.1 bit/s / Hz difference between SPM/FWM and FWM only for $B = 500$ GHz at the capacity-maximizing SNR).

Finally, the capacity as a function of the link length is shown in Figure 5.15 for $B = 500$ GHz.

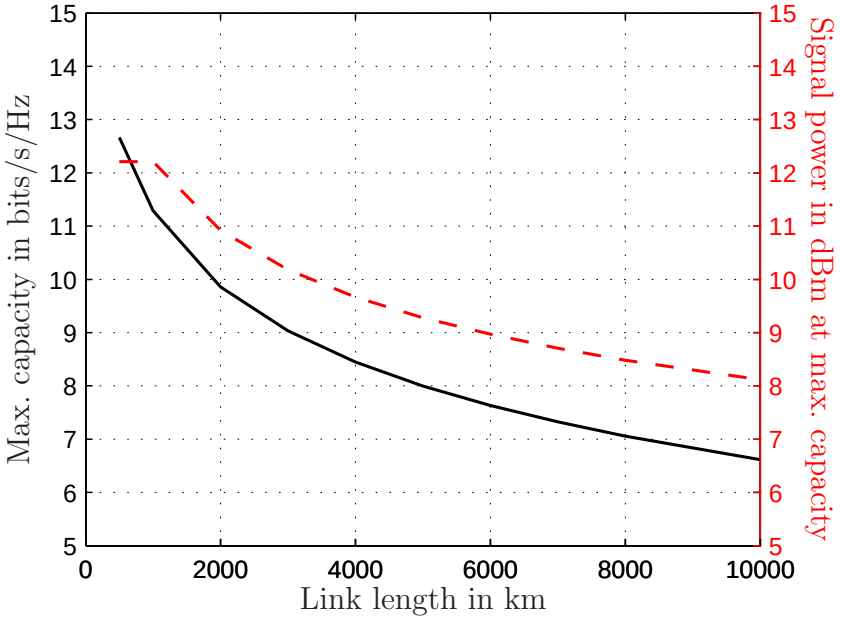

Figure 5.15: Maximum channel capacity in bits/s/Hz (black solid line) and optimum signal power in dBm, $10 \log_{10}(P_s/1\,\text{mW})$, (red dashed line) over SSMF length in km considering only signal-ASE (and ASE-ASE) interaction through FWM. Ideal distributed amplification is used. The signal bandwidth is $B = 500$ GHz.

Because of the five-fold bandwidth increase, the optimum input power (right y-axis) is increased by approximately 7 dB compared to the case where $B = 100\,\text{GHz}$ (cf. top x-axis in Figure 5.14).

As mentioned in Sections 5.4.2 and 5.5.1, the actual channel capacity values lie slightly higher than the estimates shown in Figures 5.14 and 5.15 for two reasons. First, the optimum input PSD is non-flat due to FWM. More importantly, nonlinear signal-ASE and ASE-ASE interactions through FWM are overestimated by assuming that the full ASE noise power propagates along the entire link length.

In obtaining the results presented in this section, the nonlinear signal-ASE (and ASE-ASE) interaction with out-of-band ASE noise was not considered. The factor by which this simplification reduces the resulting FWM noise power is easily calculated by comparing FWM spectra for different bandwidths. For a signal with bandwidth $B = 100\,\text{GHz}$, the FWM noise power is increased by approximately 11% if FWM products involving out-of-band spectral components are taken into account. However, optical filters can be employed to prevent out-of-band ASE noise from leaking into the signal spectrum through FWM. In dense WDM systems that efficiently utilize the available bandwidth of the fiber, the neighboring bands will be occupied with co-propagating signals anyway.

5.6 Summary

We considered the propagation of a single field along an optically amplified fiber link. Such a field can be a single signal, but it can also comprise several co-propagating WDM channels. The impact of fiber nonlinearities on the capacity of the fiber channel was discussed in this chapter. This discussion takes the interplay of nonlinearities, chromatic dispersion and ASE noise into account. After some important introductory remarks and a review of the existing literature, the following main results were obtained:

- ▷ The channel model used in this chapter was introduced. In the proposed model, the signal is represented by a number N_{sc} of signal samples in the frequency domain. The fiber effects SPM, XPM and FWM were separated.
- ▷ SPM introduces a deterministic phase shift and has no effect on the channel capacity.
- ▷ XPM induces phase noise. The variance of this phase noise reduces the channel capacity (through the phase noise itself and through spectral loss). As the number N_{sc} of spectral components (or sub-carriers in a practical multi-carrier system) grows, the phase noise variance decreases. In the proposed channel model, XPM does not affect the capacity.
- ▷ FWM leads to a spectral energy transfer. The resulting FWM noise is modeled as AWGN. Its variance and power spectral density were derived. The underlying model neglects pump depletion and the generation of higher-order FWM products. A discussion of the model's accuracy indicates that the model is valid nevertheless at the capacity-maximizing signal power and beyond.

- ▷ The channel capacity was calculated for DSF and SSMF links. Because of phase-mismatch, the capacity is higher in SSMFs. The capacity is limited by ASE noise at low power levels and by FWM noise at high power levels. It has a distinct optimum between the two regions, yielding an optimum input signal power.
- ▷ Lumped and ideal distributed amplification were compared. The maximum capacity is achieved with ideal Raman amplification. On a 2000-km SSMF link, the maximum capacity is 6.9 bit/s / Hz.
- ▷ The capacity and the optimum input power as functions of the link length L were reported. The optimum power decreases with L for DSFs and increases for SSMFs. This reflects the dependency of the FWM noise power on L.
- ▷ Finally, the interaction of signal and ASE noise through fiber nonlinearities was included in the discussion. A numerical calculation showed that the effect of nonlinear signal-ASE interaction on the capacity is negligible compared to signal-signal interactions.
- ▷ The proposed capacity calculation method was extended to systems that are able to mitigate signal-signal nonlinearities. The resulting capacity is close to that of an AWGN channel for a wide SNR range, but is still bounded by nonlinear signal-ASE and ASE-ASE interactions to almost 10 bit/s / Hz for the 2000-km Raman link. It matches a published result obtained with a different approach very well.

6

Capacity limits of WDM systems

Modern fiber-optic communication links split the enormous available bandwidth into smaller frequency bands (usually of 50 or 100 GHz bandwidth, cf. [ITU02]). The nonlinear effects that occur in such wavelength division multiplexing (WDM) systems are divided into intra-channel and inter-channel effects (see Sections 2.2.2 and 5.1.1).

The approach taken in Chapter 5 assumes that none or all of these effects can be mitigated. In optical WDM networks, however, little is known to the transmitter or receiver of one particular (*probe*) channel about co-propagating (*pump*) WDM channels. The reason is that these channels originate and terminate at different network nodes. Moreover, channels can be dropped and added at every network node along the link so that the signals that impair the probe channel can change along the link.

This chapter discusses the capacity limits of such optical WDM systems. This discussion assumes that intra-channel effects can be removed, e. g. by electronic predistortion, while inter-channel effects (and signal-ASE interactions) remain as fundamental impairments. Section 6.1 briefly reviews the literature that considers this scenario. Section 6.2 introduces a phenomenological time-domain channel model to determine the capacity of optical WDM links. The frequency-domain method introduced in the last chapter is extended and applied to WDM systems in Section 6.3. Finally, Section 6.4 summarizes this chapter.

6.1 Literature review

Mitra et al. were among the first to consider the capacity of WDM systems [MS01b, SMS01, WPG+04]. They identified XPM$^{\mathcal{S}}$ as the limiting inter-channel effect (and later included FWM$^{\mathcal{S}}$ in [WPG+04]) and formulated a phenomenological channel model for the fiber-optic channel.[1] This model is essentially an AWGN channel with multiplicative Gaussian noise. The central shortcoming of these contributions is that the physical justification of the proposed channel model remains largely unclear. This makes it rather difficult to determine the accuracy of the results as well as the system configurations that they are limited to. The same criticism applies to the model proposed by Desurvire [Des02b, Des02c].

Ho and Kahn have considered constant-intensity modulation schemes [HK02]. In the absence of random XPM$^{\mathcal{S}}$-induced phase noise, they identify FWM$^{\mathcal{S}}$ as the main impairment. Since chromatic dispersion converts phase modulation into amplitude modulation, this approach is limited to dispersion-shifted fibers (DSFs).

Essiambre et al. have obtained more generally valid results on the channel capacity by evaluating extensive numerical simulations of the fiber-optic channel [EFKW08, EFW+08, EFWK08, EKFW08, EFWK09, FEW+09, EFKW10, EKW+10]. This approach is discussed in Section 6.2.2.

Finally, Taghavi et al. have proposed to calculate the capacity of WDM systems using multi-user information theory [TPS06]. This approach requires a joint receiver that has access to all co-propagating WDM channels. It is not surprising that each WDM channel can achieve the capacity of a linear fiber channel in this scenario.

6.2 Time-domain analysis of the channel capacity

6.2.1 Nonlinear WDM effects

The nonlinear Schroedinger equation (NLSE) (2.70) governs the propagation of an arbitrary field $a(z,t)$ along the fiber. If this field comprises N WDM channels, it can be written as

$$a(z,t) = \sum_{m=1}^{N} a^{(m)}(z,t) \cdot e^{-\jmath 2\pi \cdot mW \cdot t}, \tag{6.1}$$

where W denotes the channel spacing bandwidth. We use (2.78) to insert (6.1) into (2.70) and obtain a set of N coupled differential equations. The propagation equation for

[1] The letter $^{\mathcal{S}}$ denotes a WDM system effect. See Sections 2.2.2 and 6.2.1 for an explanation of the terms SPM$^{\mathcal{S}}$, XPM$^{\mathcal{S}}$ and FWM$^{\mathcal{S}}$.

channel m is

$$\begin{aligned}\frac{\partial}{\partial z}a^{(m)} = &\underbrace{-\frac{\alpha}{2}a^{(m)}}_{\text{attenuation}} \underbrace{-\beta_1^{(m)}\frac{\partial}{\partial t}a^{(m)}}_{\text{walk-off}} \underbrace{+\jmath\cdot\frac{\beta_2^{(m)}}{2}\frac{\partial^2}{\partial t^2}a^{(m)}+\frac{\beta_3^{(m)}}{6}\frac{\partial^3}{\partial t^3}a^{(m)}}_{\text{chromatic dispersion}} \\ &\underbrace{-\jmath\,\gamma_{\text{eff}}|a^{(m)}|^2\cdot a^{(m)}}_{\text{SPM}^{\mathcal{S}}} \\ &\underbrace{-\jmath\,\gamma_{\text{eff}}\cdot 2\cdot\sum_{\substack{n=1\\n\neq m}}^{N}|a^{(n)}|^2\cdot a^{(m)}}_{\text{XPM}^{\mathcal{S}}} \\ &\underbrace{-\jmath\,\gamma_{\text{eff}}\cdot\sum_{\substack{n,p,q\\q\neq n,q\neq p}}a^{(n)}a^{(p)}a^{(q)^*}\cdot e^{-\jmath\,\Delta\beta\cdot z}}_{\text{FWM}^{\mathcal{S}}},\end{aligned} \tag{6.2}$$

where $a^{(m)}$ stands for $a^{(m)}(z,t)$. The phase mismatch term associated with FWM$^{\mathcal{S}}$ is given by (5.4).

The similarity of (6.2) with (5.6) motivated the use of the terms SPM$^{\mathcal{S}}$, XPM$^{\mathcal{S}}$ and FWM$^{\mathcal{S}}$ to denote the interaction of WDM channels. It is important to emphasize the difference between the equally named effects. All nonlinear interactions (through SPM, XPM and FWM) of spectral components belonging to channel m are subsumed by the SPM$^{\mathcal{S}}$ term. Similarly, the XPM$^{\mathcal{S}}$ term contains nonlinear interactions through XPM and FWM.

Just as SPM$^{\mathcal{S}}$ causes a phase modulation proportional to the instantaneous power, XPM$^{\mathcal{S}}$ rotates the phase proportionally to the power in all co-propagating WDM channels. XPM$^{\mathcal{S}}$ is always phase-matched; it does not become less efficient with increasing channel separation. However, two interacting channels experience a group velocity difference in fibers with chromatic dispersion. This channel *walk-off* has an averaging effect on the XPM$^{\mathcal{S}}$-induced phase modulation, so that the distortions due to XPM$^{\mathcal{S}}$ decline with increasing channel separation [KK97a, Ch. 8]. The length scale governing this behavior is the walk-off length (2.150). Finally, FWM$^{\mathcal{S}}$ is responsible for an energy exchange between WDM channels. Analogously to FWM, FWM$^{\mathcal{S}}$ products can appear at occupied frequencies (especially in WDM systems with equally spaced channels) or at new frequencies. FWM$^{\mathcal{S}}$ is effectively suppressed by the phase mismatch in WDM system (unless dispersion-shifted fibers are used).

At a large frequency spacing between interacting waves, γ_{eff} reduces from $8/9\cdot\gamma$ to $2/3\cdot\gamma$ as discussed in Section 2.1.6. Nevertheless, we will use $\gamma_{\text{eff}} = 8/9\cdot\gamma$ throughout this chapter. As explained above, the closest WDM neighbors generate the strongest inter-channel nonlinearities (XPM$^{\mathcal{S}}$ and FWM$^{\mathcal{S}}$). Since modern single-mode fibers have very low PMD parameters, the bandwidth (4.6) of the principal states of polarization is larger than the bandwidth of a single WDM channel. It is therefore reasonable to assume that the random changes of light polarization of neighboring WDM channels are strongly

correlated. While this may lead to results that slightly overestimate the effect of inter-channel nonlinearities, a more precise treatment would require knowledge of the exact transition of γ_{eff} from 8/9 to 2/3 as a function of the frequency separation. This question is left open for future research.

The one inter-channel effect that is not considered in this chapter is stimulated Raman scattering (SRS), see Section 2.1.4. If the bandwidth occupied by the WDM channels is very large, the Raman gain depicted in Figure 2.4 must be taken into account. SRS leads to an energy transfer from channels at higher frequencies to channels at lower frequencies. To maximize the capacity of the entire system, an iterative water-filling scheme (cf. FWM, Section 5.4.2) is required to find the optimum transmit power allocation.

6.2.2 Phenomenological channel model

Review of published results

Essiambre et al. have obtained results for the channel capacity of fiber-optic networks by evaluating numerical simulations of the channel [EKW+10]. The reported results stand out from the other contributions summarized in Section 6.1 in the sense that they are not limited to specific fiber effects or system configurations.

The system configuration is summarized in [EKW+10, Tables I-V]. In short, five densely packed WDM channels (channel bandwidth $B = 100\,\text{GHz}$, channel spacing 102 GHz) are propagated along a standard single-mode fiber (SSMF) link with ideal Raman amplification (local gain equals local loss) and ideal (rectangular) optical filtering. The central channel acts as the probe channel. Simulating more than five WDM channels does not change the results significantly [EKW+10], presumably due to walk-off. Distributed ASE is generated in the simulation and co-propagated with the signal. In each channel, sinc-shaped pulses are modulated with ring constellations (cf. Section 3.4.3). Hence, each channel's power spectral density (PSD) is rectangular. Intra-channel nonlinearities (except signal-ASE and ASE-ASE interactions) are removed by electronic predistortion. Links with different dispersion maps and link lengths are considered. We discuss dispersion-managed links and links without inline dispersion compensation separately in Section 6.2.3 and 6.2.4.

A numerical approach has the advantage of greater confidence in the results but at the expense of computational effort and complexity rendering it very challenging for parameter space optimization (such as for instance for the constellation optimization performed in [FEW+09]). In the following, we try to escape this predicament by formulating a phenomenological channel model which, while requiring a few observations made in the full numerical simulations, allows a fast scaling of fiber-optic channel capacities to other system operating regions.

Phenomenological channel model

Figures 6.1(a)-(c) display the received signal after (numerically simulated) transmission for a ring constellation with $r = 4$ rings (2000-km link as described in Section 6.2.3).[2] The constellations are "backrotated" to a reference phase as described in [EKW+10]. The input powers per channel are −15 dBm, −6 dBm and 3 dBm. The observed signals

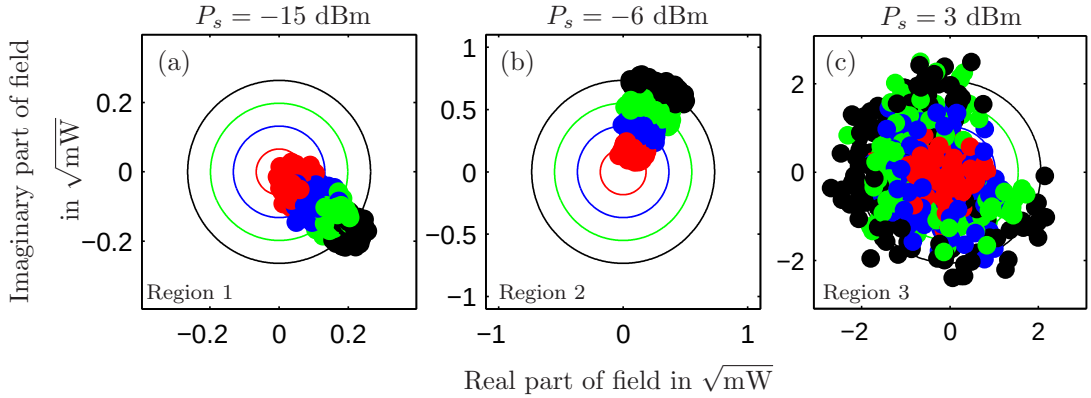

Figure 6.1: Typical received signals (four-ring constellation) for each SNR region.

show characteristic differences for each of the three SNR regions. For very low SNRs (Region 1), the channel is essentially linear and ASE noise is the dominant impairment. In this regime, all the transmitted points that belong to a given ring in the signal space are received as a "cloud" with circular symmetry (Figure 6.1(a)). The average phase rotation visible in Figures 6.1(a)-(c) is due to fiber nonlinearities. It can easily be removed at the receiver and consequently does not affect capacity. In the following, we will therefore consider the received (backrotated) phase as a zero-mean random variable. As signal powers increase, fiber nonlinearities become apparent (Region 2, SNR$\gtrsim$15 dB), producing large phase distortions that appear as kidney-shaped clouds (Figure 6.1(b)). The phase noise causing this angular spread can be attributed to XPM$^{\mathcal{S}}$ [EKW+10] as the signal phase is modulated by the time- and distance-dependent power levels $P^{(n)}(z,t)$ ($n \neq m$, where m is the channel considered) of the co-propagating WDM channels. A third regime (Region 3, SNR $\gtrsim$ 23 dB) of transmission is observed at very large signal powers, where a very strong nonlinear (NL) scattering of the received signal occurs in addition to AWGN and nonlinear phase noise (Figure 6.1(c)).

In Region 1, the channel can be modeled as an AWGN channel. The channel capacities using ring constellations grow logarithmically with the SNR (see Figure 6.3, or [EKW+10, Fig. 16]).

With increasing signal power, the distortion becomes clearly non-Gaussian; an AWGN channel model can no longer predict the correct fiber channel capacity. As each WDM channel carries independent data, the power levels $P^{(n)}(z,t)$ in the co-propagating WDM channels can be considered as independent random variables. We assume all channels to

[2]The depicted results were produced by R.-J. Essiambre and are used with permission.

have identical signal power P_s and data format, so that all $P^{(n)}(z,t)$ have identical mean and variance. For the link with ideal distributed amplification considered in [EKW+10], the mean remains constant and is given by P_s. In contrast, the variance σ_{P}^2 of the random variables $P^{(n)}(z,t)$ is not constant. It evolves along z depending on the dispersion map and is proportional to P_s^2. The problem of how to determine an "effective" σ_{P}^2 is discussed separately in Sections 6.2.3 and 6.2.4.

The probe channel's phase is XPM$^{\mathcal{S}}$-modulated by the sum of the powers $P^{(n)}(z,t)$. As expected from the central limit theorem, the resulting nonlinear phase Θ_{PN} will (asymptotically, i. e. for large n) assume a Gaussian PDF with mean and variance proportional to P_s and σ_{P}^2 (where σ_{P}^2 in turn is proportional to P_s^2), respectively. As mentioned before, the average phase rotation does not impact capacity and we consider Θ_{PN} as a zero-mean random variable. Figure 6.2 shows an an examplary histogram of Θ_{PN} ($r=8$, $10\log_{10}(P_s)=-2\,\mathrm{dBm}$, link setup as discussed in Sec. 6.2.3).

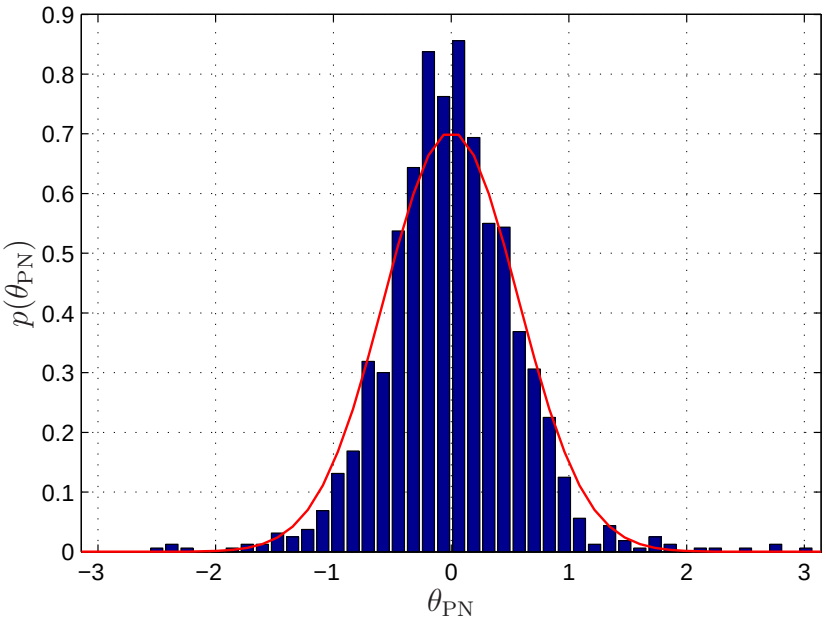

Figure 6.2: Histogram of phase angle Θ_{PN} of received samples (blue bars) and fitted Gaussian PDF (red line) for $10\log_{10}(P_s)=-2\,\mathrm{dBm}$.

The phenomenological fiber channel model in Region 2 is given by

$$Y = X \cdot e^{\jmath\,\Theta_{\mathrm{PN}}} + N, \tag{6.3}$$

where $\Theta_{\mathrm{PN}} \sim \mathcal{N}_{\mathbb{R}}\left(0, c_{\mathrm{R2}} \cdot \sigma_{\mathrm{P}}^2\right)$. Hence, the fiber channel is modeled as partially coherent (see Section 3.6) with a wrapped Gaussian phase noise distribution (3.67). The capacity in Region 2 is determined by the SNR P_s/P_{ASE} and the phase noise variance $c_{\mathrm{R2}} \cdot \sigma_{\mathrm{P}}^2$. The phase noise parameter c_{R2} (in units of W^{-2}) incorporates system specifications such as the number of WDM channels, the channel walk-off due to chromatic dispersion, the dispersion map and the link length. The estimation of c_{R2} is described in Section 6.2.3.

In anticipation of the capacity calculation details given in Section 6.2.3, Figure 6.3 depicts capacity values for various ring constellations obtained from the full numerical model (FNM) of [EKW+10], from an AWGN channel model (as applicable in Region 1) and from the phenomenological channel model (6.3) applicable in Region 2. The curves shown

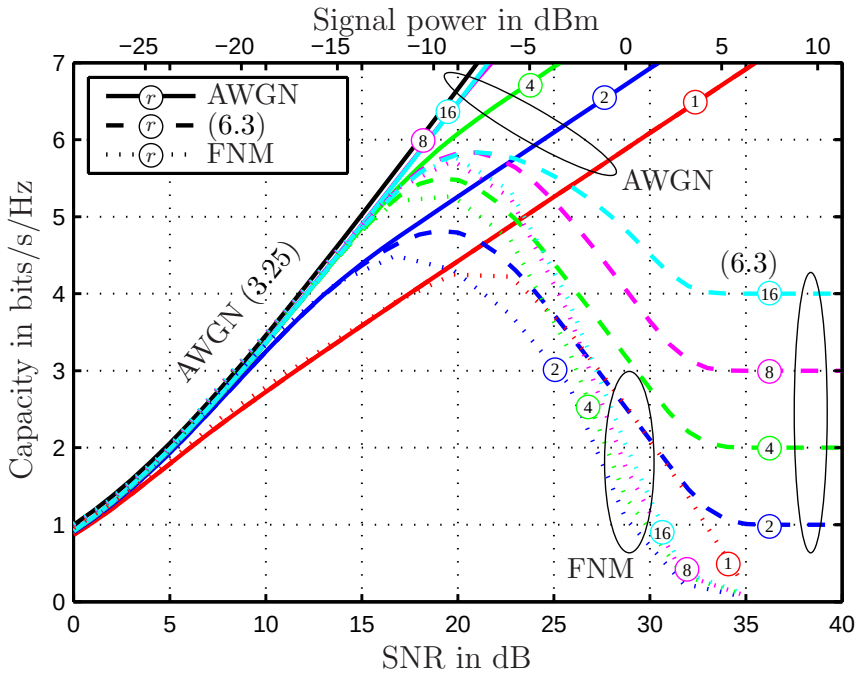

Figure 6.3: Channel capacity in bits/s/Hz of dispersion-managed 2000-km link over SNR in dB, $10\log_{10}(P_s/P_{\mathrm{ASE}})$, as obtained from AWGN channel model, partially coherent channel model (6.3) and full numerical model (FNM) [EKW+10] for various ring constellations. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$.

in Figure 6.3 (and in all other figures of this section) were obtained using the polar decomposition method of Section 3.3. The (very small) mixed term II was neglected in all cases. In obtaining the depicted capacity values, spectral loss was not taken into consideration.

It can be observed that the channel model (6.3) predicts the capacity reasonably well for the SNR range we call Region 2. This region includes the point of maximum capacity, at which the channel is weakly nonlinear, i. e. only few decibels above the linear regime (Region 1) [EKW+10]. Hence, the single parameter c_{R2} is sufficient to predict the capacity values in the SNR regions where C is maximum.

In regions of very high SNRs (Region 3, SNR $\gtrsim$ 23 dB), the channel (6.3) can no longer convey any of the information encoded in the signal phase; it becomes noncoherent. At this point, the phase's circular variance tends to 1 and its distribution approaches the uniform distribution. In contrast, the information encoded in the r amplitude levels remains unaffected by phase noise, so that the capacity of the r-ring constellation tends towards $\log_2(r)$ for very large SNR. This effect can be observed in Figure 6.3, where the dashed lines saturate at $\log_2(r)$ bits/s/Hz. In order to rectify this disagreement with

existing results from [EKW+10] (where $C \to 0$ for very large SNRs), a third parameter is required to characterize the channel completely.

In Region 3, we observe a very strong nonlinear scattering of the received signal occurs in addition to AWGN and nonlinear phase noise (Figure 6.1(c)). To account for this effect, we heuristically introduce a power-dependent complex AWGN process $N_{\mathrm{NL}} \sim \mathcal{N}_{\mathbb{C}}\left(0, c_{\mathrm{R3}} \cdot P_s^c\right)$, where the value of c determines the SNR level at which this process becomes apparent. We obtained a good agreement with the existing results for $c=3$, but did not optimize this parameter. The parameter c_{R3} has units of W^{1-c}, i. e. W^{-2} here. The explicit fiber channel model in Region 3 is

$$Y = X \cdot e^{\jmath\, \Theta_{\mathrm{PN}}} + N + N_{\mathrm{NL}}. \tag{6.4}$$

The physical cause for N_{NL} could be PM-to-AM conversion due to chromatic dispersion. Furthermore, the efffect of spectral loss derived in Section 3.6.2 might contribute to N_{NL} as discussed in Section 6.2.5.

6.2.3 Dispersion-managed links

Capacity results for dispersion-managed links with variable residual dispersion per span (RDPS) are reported in [EKW+10]. Here, we confine the discussion to the link setup used in [EFKW08]. In addition to the technical details listed in Section 6.2.2, the link uses a dispersion map with $-1050\,\mathrm{ps}\,/\,\mathrm{nm}$ pre-compensation and $20\,\mathrm{ps}\,/\,\mathrm{nm}$ residual dispersion per span. Results for link lengths of 500 km [EFWK09], 1000 km [EFWK08] and 2000 km [EFKW08] of SSMF are reported. We will concentrate on the 2000-km link in the following.

Calculation of effective power variance σ_{P}^2

The phenomenological channel models (6.3) and (6.4) rely on accurate estimates of the phase noise variance $c_{\mathrm{R2}} \cdot \sigma_{\mathrm{P}}^2$. This variance could be estimated directly from numerical simulation results at a meaningful SNR value (in Region 2). As σ_{P}^2 scales quadratically with P_s, the phase noise variance could then be scaled to different SNR values. This approach, however, requires separate numerical simulations for different data formats (i. e. constellations).

In order to make the phenomenological channel model scalable to various input constellations, we wish to estimate σ_{P}^2 analytically. A surprisingly simple estimate for the given dispersion-managed link is to calculate σ_{P}^2 as the power variance of the input constellation, e. g. $\sigma_{\mathrm{P}}^2 = 0$ for a 1-ring constellation and $\sigma_{\mathrm{P}}^2 \approx 0.57 \cdot P_s^2$ for $r=4$.

Although this approach appears overly simple, it delivers a much better agreement with observations made in numerical channel simulations than e. g. calculating σ_{P}^2 from the power variance of the actually propagating waveforms (taking pulses and dispersion into account) does. A possible explanation could be the lowpass character of $\mathrm{XPM}^{\mathcal{S}}$ (see e. g. [LWR02]) which entails two consequences:

1. The power variance of the signal waveform (which is a modulated sequence of sinc-pulses) approaches the power variance of the constellation when its spectrum is lowpass-filtered. For highly variant constellations, e. g. ring constellations with large r, this has a negligible effect. However, when the power variance is low, the effect of filtering is significant. In particular, the single-ring constellation produces very little phase noise through XPM$^{\mathcal{S}}$ in numerical simulations (see e. g. [FEW^{+}09, Fig. 3]). This is reflected in the high capacity of this modulation scheme at large SNRs, i. e. when XPM$^{\mathcal{S}}$ is strong.

2. The broadband waveforms propagating in the WDM channels are transformed by chromatic dispersion relatively quickly. However, only a lowpass-filtered version of each signal spectrum induces phase noise through XPM$^{\mathcal{S}}$. These "effective" narrowband signals are less susceptible to dispersion. As the given dispersion-managed link has only a limited amount of accumulated dispersion, the effect of dispersion can be neglected. In contrast, the effect of dispersion must be taken into account for uncompensated links as discussed in Section 6.2.4.

Parameter estimation

As mentioned above, the channel parameters $c_{\mathrm{R}2}$ and $c_{\mathrm{R}3}$ incorporate system specifications such as the number of WDM channels, the channel walk-off due to chromatic dispersion, the dispersion map and the link length. It seems difficult to calculate these parameters analytically. Instead, they are estimated from existing capacity results in a two-stage process.

First, $c_{\mathrm{R}2}$ is estimated from an existing capacity value in Region 2 ($\mathrm{SNR} = 20 \pm 2\,\mathrm{dB}$ here) for any constellation with more than one ring. To find $c_{\mathrm{R}2}$, the capacity of channel (6.3) is determined using the polar decomposition of mutual information introduced in Section 3.3. In a gradient search algorithm, $c_{\mathrm{R}2}$ is varied until the resulting capacity equals the given value. For the given dispersion-managed link, this estimation method delivers $c_{\mathrm{R}2} \approx 7.75 \cdot 10^5\,\mathrm{W}^{-2}$.

In the second step, $c_{\mathrm{R}2}$ is fixed and $c_{\mathrm{R}3}$ is determined from a second existing capacity value in Region 3. We use a point of capacity below half of the peak capacity of Region 2 ($\mathrm{SNR} = 29\,\mathrm{dB} \pm 2\,\mathrm{dB}$ here) as a reference capacity value. As before, the parameter $c_{\mathrm{R}3}$ is varied until the capacity of (6.4) (again calculated using the polar decomposition method) matches this reference capacity value. For the given link, we obtain $c_{\mathrm{R}3} \approx 8.2 \cdot 10^4\,\mathrm{W}^{-2}$.

This parameter estimatation method requires two capacity values obtained from the full numerical simulation. These values must be taken at SNR values that are representative for Regions 2 and 3, respectively. Once both parameters are estimated for a given link, they can be used to produce capacity results for all SNR values and input constellations.

Numerical capacity results

Using the estimated values for $c_{\mathrm{R}2}$ and $c_{\mathrm{R}3}$, we calculated the channel capacities for ring constellations (uniform ring probability, equidistant radii) from 1 to 16 rings and for

various values of SNR; the resulting curves are shown in Figure 6.4. The results match

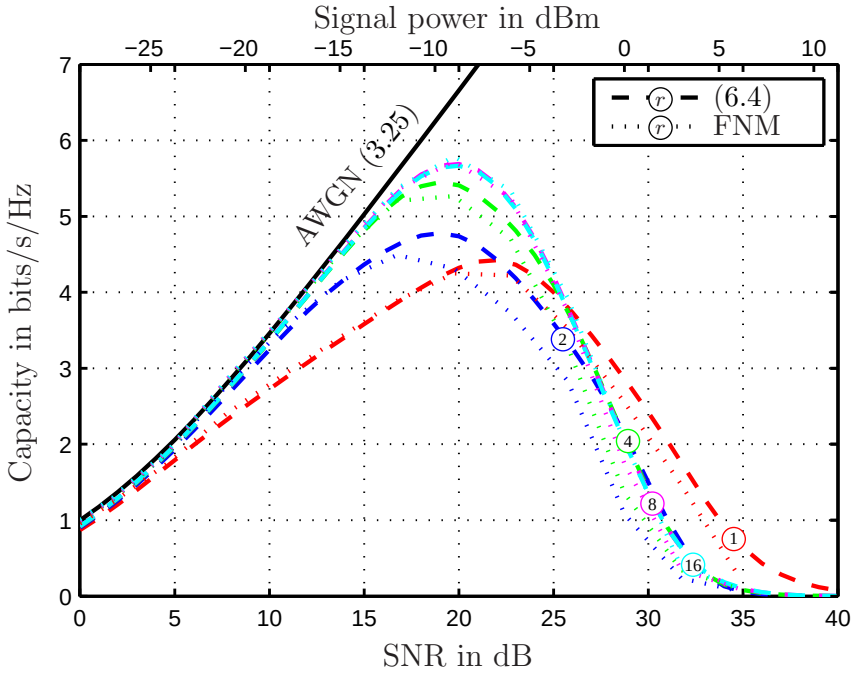

Figure 6.4: Channel capacity in bits/s/Hz of dispersion-managed 2000-km link over SNR in dB, $10\log_{10}(P_s/P_{\mathrm{ASE}})$, as obtained from phenomenological channel model (6.4) and full numerical model (FNM) [EKW+10] for various uniform ring constellations. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$.

the capacity values reported in [EFKW08] for all values of r and SNR within few tenths of bits/s/Hz, so that they are within the numerical uncertainty of a few tenths of bits/s/Hz reported in [EFKW08].

The proposed phenomenological channel may help understand the somewhat counter-intuitively good performance of the 1-ring constellation in Region 3. This constellation is scarcely affected by phase noise (low power variance), so it is only impaired by AWGN N and nonlinear AWGN N_{NL}, and outperforms the multi-ring constellations in Region 3 where phase noise is very strong.

Input optimization

Ring constellations with a sufficient number of rings approach the Shannon limit very closely [EKW+10, Fig. 16]. However, they are not the optimum input constellation for either an AWGN channel or a partially coherent channel. In [FEW+09], the ring radii and probabilities were numerically optimized to increase the capacity. This optimization leads to a small capacity increase of 0.5 bits/s/Hz for the 2-ring constellation. Using the same phenomenological channel model and parameters, i. e. the same $c_{\mathrm{R2}}, c_{\mathrm{R3}}$, we performed an optimization of the ring ratios and probabilities of the input constellation (see [FEW+09] for details of the optimization method).

Figure 6.5 depicts the capacity results for the optimized ring inputs. As expected, the capacity values as well as the optimum ring ratios and probabilities match those reported in [FEW+09] (to within a few tenths of bits/s/Hz). This demonstates the ability of the phenomenological channel model to predict capacity results for arbitrary modulation schemes (once c_{R2} and c_{R3} have been estimated from two capacity values of one particular modulation scheme).

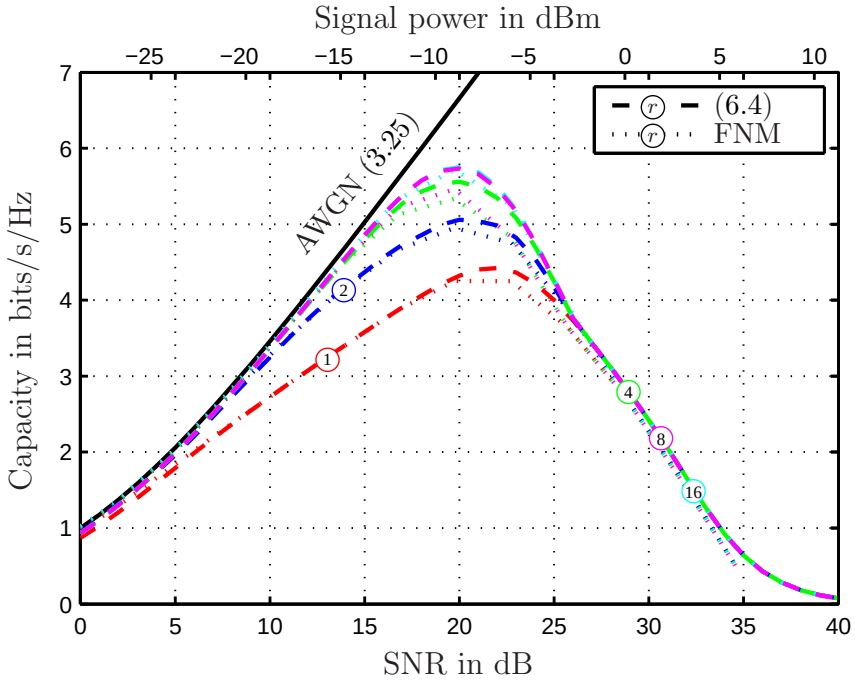

Figure 6.5: Channel capacity in bits/s/Hz of dispersion-managed 2000-km link over SNR in dB, $10\log_{10}(P_s/P_{\mathrm{ASE}})$, as obtained from phenomenological channel model (6.4) and full numerical model (FNM) [FEW+09] for various optimized ring constellations. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$.

The estimation process as well as the calculation of capacity curves such as those shown in Figure 6.4 take only a few seconds on a regular personal computer. One potential application of the phenomenological model is therefore the fast screening of spectrally efficient and robust modulation schemes to maximize the capacity of future fiber-optic WDM transmission systems. Due to the heuristic nature of the approach, one would always want to perform a closer examination of the identified modulation schemes using the full numerical channel simulation. In that sense, the proposed method does not intend to replace the approach of [EKW+10], but rather provide a tool for rapid capacity estimation in the parameter space around a given capacity evaluation.

6.2.4 Uncompensated links

The results reported in Section 6.2.3 were obtained for a link with inline compensation of chromatic dispersion. The used dispersion map with moderate residual dispersion per

span (20 ps / nm) can reduce fiber nonlinearities in conventional fiber-optic communication systems. However, the capacity is increased by reducing the amount of inline dispersion compensation [EKW+10, Fig. 34]. This is even more so if dispersion-compensating fibers are used, as such fibers have relatively large attenuation and nonlinearity parameters, see Table 2.3.

The XPM$^{\mathcal{S}}$-induced nonlinear phase noise Θ_{PN} of the phenomenological channel model (6.4) has a Gaussian distribution whose variance scales with the power variance σ_{P}^2, $\Theta_{\mathrm{PN}} \sim \mathcal{N}_{\mathbb{R}}\left(0, c_{\mathrm{R2}} \cdot \sigma_{\mathrm{P}}^2\right)$. As discussed in Section 6.2.3, a good agreement with existing results is achieved by calculating σ_{P}^2 from the signal constellation for dispersion-managed links with low accumulated dispersion. This is different for uncompensated links. In the absence of inline dispersion compensation, chromatic dispersion transforms the waveforms in each WDM channel into zero-mean Gaussian random processes with variance P_s relatively fast. The waveforms, and hence the XPM$^{\mathcal{S}}$-induced phase noise, become independent of the WDM channels' data formats for long links. Therefore, σ_{P}^2 scales with the power variance of the Gaussian random process. The power has a central (scaled) χ^2 distribution with variance P_s^2, so we have $\Theta_{\mathrm{PN}} \sim \mathcal{N}_{\mathbb{R}}\left(0, c_{\mathrm{R2}} \cdot P_s^2\right)$.

Capacity results for the 16-ring constellation for uncompensated links with $L = 500$, 1000, 2000, 4000, 8000 km are reported in [EKW+10, Fig. 35]. In the phenomenological channel model, different parameter sets $\{c_{\mathrm{R2}}, c_{\mathrm{R3}}\}$ are expected for each fiber length. The parameter estimation procedure described in Section 6.2.3 is therefore repeated for each link length separately. As before, c_{R2} is estimated from a single capacity result in the region of maximum capacity and c_{R3} from a single result in the highly nonlinear regime.

For the given link lengths, the resulting values are $c_{\mathrm{R2}} \approx 10200$, 23500, 43800, 44300, 77400 W^{-2} and $c_{\mathrm{R3}} \approx 1020$, 1910, 3770, 4130, 7025 W^{-2}. A simple, e. g. linear, relationship of $c_{\mathrm{R2}}, c_{\mathrm{R3}}$ and L cannot be observed. Length scaling of the obtained results therefore remains an open problem.

Figure 6.6 shows the results obtained from the phenomenological channel model for the various link lengths and a 16-ring input constellation. The results from the full numerical model (FNM) of [EKW+10, Fig. 35] are shown for comparison. We observe an excellent agreement between the results. With every two-fold increase in distance, the SNR which maximizes the capacity decreases by approximately 3 dB. At the same time, the ASE noise power increases by 3 dB. Hence, the signal power that maximizes the capacity remains approximately constant [EKW+10].

Finally, we use the parameter set $\{c_{\mathrm{R2}}, c_{\mathrm{R3}}\}$ found for the 2000-km link (green line in Figure 6.6) to obtain results for various input constellations. Figure 6.7 shows the results obtained from the phenomenological channel model for the 2000-km link and for $r = 1, 2, 4, 16$.

It can be seen that the phenomenological channel model does not reproduce the FNM results for $r = 1$ and $r = 2$ as accurately it did in the case of the dispersion-managed link (Section 6.2.3). As the power variance of the constellation decreases, (6.4) tends to underestimate the capacity. The reason for this disagreement lies in the estimation of

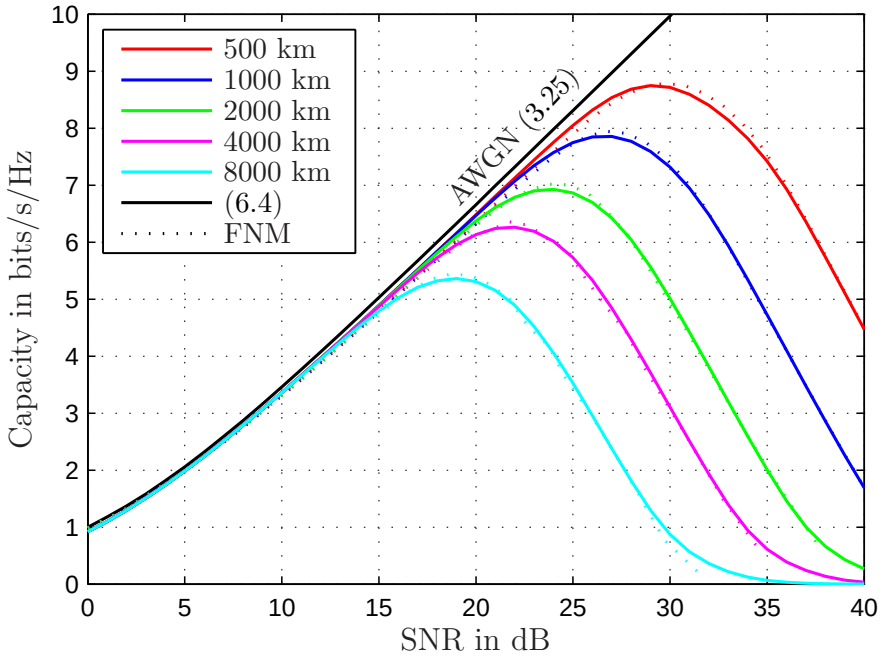

Figure 6.6: Channel capacity in bits/s/Hz over SNR in dB, $10\log_{10}(P_s/P_{\mathrm{ASE}})$, as obtained from phenomenological channel model (6.4) and full numerical model (FNM) [EKW+10] for uncompensated link with lengths $L = 500, 1000, 2000, 4000, 8000$ km and a 16-ring input constellation.

σ_{P}^2 outlined above, which assumes a Gaussian distribution of the propagating waveforms. However, the low-variance signals cause considerably less phase noise through XPM$^{\mathcal{S}}$ for a significant fiber length before dispersion eventually renders them Gaussian. To increase the accuracy of the results for low-variance constellations, σ_{P}^2 could be estimated (or measured) specifically for the format at hand.

6.2.5 Effect of spectral loss

As we have seen in Section 6.2.2, two nonlinear effects are necessary to obtain a good agreement of the phenomenological channel model with the numerical results of [EKW+10]. The existence of nonlinear phase noise can be attributed to XPM$^{\mathcal{S}}$. In contrast, a physical origin of the nonlinear impairment N_{NL} is missing. If the model (6.4) is accurate, a possible explanation for the noise N_{NL} is the conversion of XPM$^{\mathcal{S}}$-induced phase noise to amplitude noise through chromatic dispersion. Another candidate explanation is the spectral loss effect.

As discussed in Section 3.6.2, (6.3) is incomplete in that it neglects spectral loss. Including spectral loss, the phenomenological model (6.3) becomes

$$Y = X \cdot \exp\left(-\frac{1}{2}\cdot\sigma^2 + \jmath\,\Theta_{\mathrm{PN}}\right) + N. \tag{6.5}$$

By using the channel model (6.5) rather than (6.4), we attempt to replace the heuristic

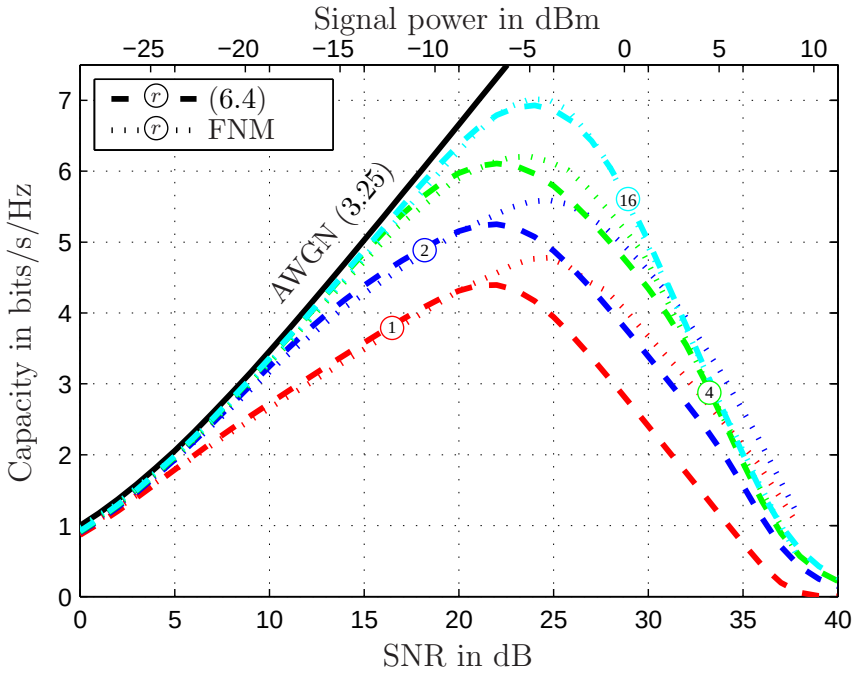

Figure 6.7: Channel capacity in bits/s/Hz of uncompensated 2000-km link over SNR in dB, $10\log_{10}(P_s/P_{\mathrm{ASE}})$, as obtained from phenomenological channel model (6.4) and full numerical model (FNM) for various ring constellations. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$.

additive noise term N_{NL} by the non-heuristic multiplicative term $\exp(-\sigma^2/2)$.

Figure 6.8 shows the decomposition results for ring constellations with $r = 1, 4, 16$ for the uncompensated 2000-km link. The same parameter $c_{\mathrm{R2}} \approx 4.4 \cdot 10^4$ W^{-2} as above was used.

Observe that the spectral loss attenuates the received signal power with increasing transmit signal power (or increasing SNR). The overall effect is similar to that of N_{NL} in (6.4) and the capacity tends towards zero for large SNRs. Spectral loss thereby lets us reduce the two-parameter $(c_{\mathrm{R2}}, c_{\mathrm{R3}})$ channel model (6.4) to a single-parameter (c_{R2}) channel model (6.5).

It must be noted that while the FNM results can be qualitatively reproduced when spectral loss is taken into account, the curves of Figure 6.8 do not match the original results [EKW+10] quite as well as those obtained from the phenomenological model (6.4). This is best observed from the 16-ring results, where (6.5) overestimates the capacity in the highly nonlinear regime, i. e. at high signal power levels (compare dashed and dotted lines in Figure 6.8 in the SNR range from 25 dB to 35 dB).[3] Thus, spectral loss should be considered only *one* further explanation for the capacity behavior caused by the fiber nonlinearities. At the same time, the model (6.5) with XPM$^{\mathcal{S}}$ phase noise Θ_{PN} and spectral

[3]The results' mismatch for $r = 1$ and $r = 4$ is less meaningful. It has its reason in the imprecise estimate of σ_{P}^2 as discussed in Section 6.2.4.

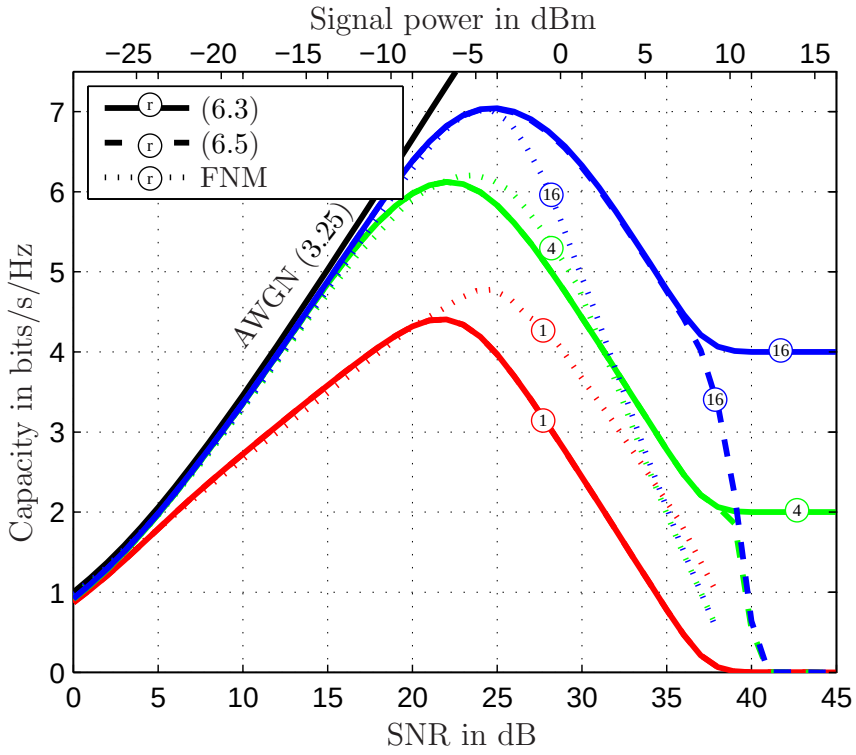

Figure 6.8: Channel capacity in bits/s/Hz of uncompensated 2000-km link over SNR in dB, $10\log_{10}(P_s/P_{\mathrm{ASE}})$, as obtained from phenomenological channel models (6.3) (without spectral loss (SL)) and (6.5) (with SL) for various uniform ring constellations. Top x-axis shows signal power in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$.

loss σ^2 does show the same qualitative capacity behavior as the FNM. A potential further refinement is to include amplitude-to-phase conversion through chromatic dispersion.

6.3 Frequency-domain analysis of the channel capacity

The method discussed in the previous section yields reliable estimates for the capacity of an idealized fiber-optic communication channel operated in the time domain. The original method [EKW+10] requires numerical simulations of the channel, the proposed phenomenological channel model is heuristic.

A strictly analytical (i. e. neither numerical nor heuristic) way to estimate the channel capacity is provided by the method introduced in Chapter 5. We will now extend this method to the analysis of WDM systems. The system model introduced in Section 5.2 remains unchanged. However, we change our view on the co-propagating spectral components $X^{(n)}$ defined by (5.2). Instead of assuming a number N_{sc} of co-propagating spectral components, we now have $N_{\mathrm{ch}} \cdot N_{\mathrm{sc}}$ co-propagating components. We assume that the number N_{ch} of WDM channels is odd. The N_{sc} spectral components in the center of this field (i. e. $X^{(N_{\mathrm{sc}}\cdot(N_{\mathrm{ch}}-1)/2+1)}, \ldots, X^{(N_{\mathrm{sc}}\cdot(N_{\mathrm{ch}}+1)/2)}$) represent the probe channel.

FWM products (including SPM and XPM) that satisfy $\{n,p,q\} \in [N_{\mathrm{sc}} \cdot (N_{\mathrm{ch}}-1)/2+1, N_{\mathrm{sc}} \cdot (N_{\mathrm{ch}}+1)/2]$ represent intra-channel nonlinearities. In combination, these products generate $\mathrm{SPM}^{\mathcal{S}}$ in the time domain. As in Section 6.2, we assume that these intra-channel nonlinearities are perfectly compensated. Every FWM product (including XPM) that involves at least one spectral component from outside the probe channel's band contributes to the inter-channel effects $\mathrm{XPM}^{\mathcal{S}}$ and $\mathrm{FWM}^{\mathcal{S}}$. These FWM products satisfy $m=n+p-q \in [N_{\mathrm{sc}} \cdot (N_{\mathrm{ch}}-1)/2+1, N_{\mathrm{sc}} \cdot (N_{\mathrm{ch}}+1)/2]$ and n or p or $q \notin [N_{\mathrm{sc}} \cdot (N_{\mathrm{ch}}-1)/2+1, N_{\mathrm{sc}} \cdot (N_{\mathrm{ch}}+1)/2]$. Such products are not removed by individual electronic predistortion of each WDM channel and remain as capacity-degrading impairments.

To find the resulting inter-channel FWM noise PSD, two spectra are calculated. The FWM noise PSD for $N_{\mathrm{ch}} \cdot N_{\mathrm{sc}}$ sub-carriers contains all inter- and intra-channel FWM products. The FWM noise PSD for N_{sc} sub-carriers, which can be taken from Chapter 5, contains only intra-channel FWM. The frequency spacing of the sub-carriers in both cases is Δf, so that the entire WDM bandwidth is $N_{\mathrm{ch}} \cdot B$. Subtracting the intra-channel FWM noise PSD from the WDM FWM noise PSD yields a PSD that contains only inter-channel FWM. The resulting FWM noise PSD for $N_{\mathrm{ch}}=5$ and $B=100\,\mathrm{GHz}$ is shown in Figure 6.9. In comparison to Figure 5.9, we observe a significant reduction of the FWM noise PSD

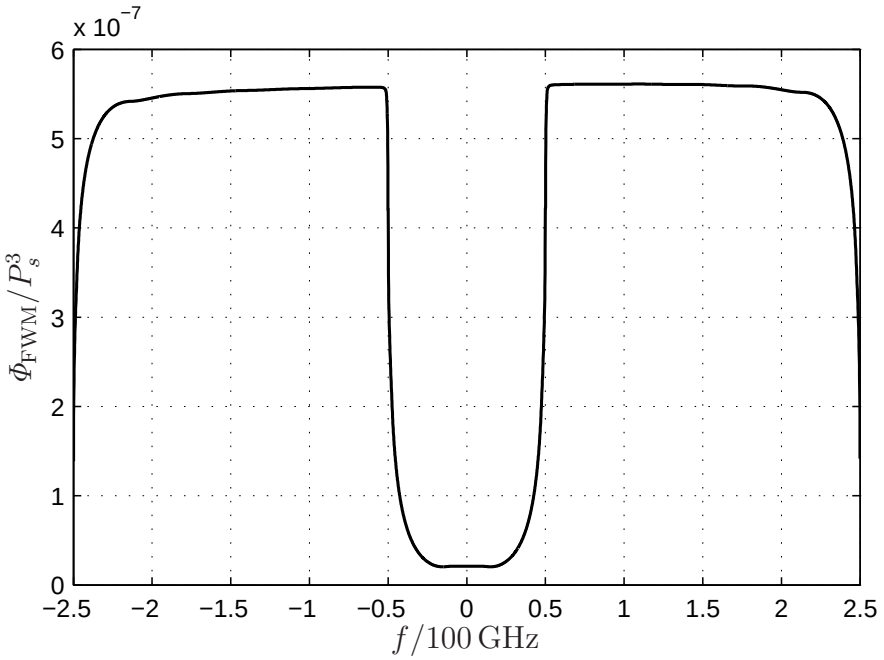

Figure 6.9: Normalized PSD of FWM noise in a WDM system with five channels for $L=2000\,\mathrm{km}$ and ideal Raman amplification. The normalized PSD has units of $\mathrm{W}^{-2}\,\mathrm{Hz}^{-1}$.

in the probe channel in the center of the depicted spectrum. The effect of the remaining (inter-channel) FWM noise on the capacity of WDM probe channel is discussed in the following sections.

Contrary to intra-channel FWM, the impairment of inter-channel FWM is stronger at the edges of the probe channel and smaller in its center. As the FWM noise spectrum

is non-flat, an iterative water-filling procedure as proposed in Section 5.4.2 is required to maximize the capacity. Such a scheme would take some signal power away from the band's edges and concentrate it in the center of the channel. In effect, this would establish *soft guard bands* between the WDM channels. This topic is left open for future research and guard bands between the WDM channels are not considered in the following.

6.3.1 Effect of inter-channel nonlinearities in WDM systems

We will neglect nonlinear signal-ASE interactions in this section and add them to the discussion in Section 6.3.3.

XPM was shown to have no effect on the channel capacity in Section 5.3. This result is valid for WDM systems, too. In fact, the variance of XPM-induced phase noise converges to zero even faster for WDM systems because of the larger number of interacting spectral components.

To determine the effect of FWM on the capacity, we proceed as in Section 5.4 but use the FWM noise PSD shown in Figure 6.9. For all calculations that follow, $N_{\mathrm{sc}} = 1000$, $N_{\mathrm{ch}} = 5$ and $B = 100\,\mathrm{GHz}$ was used. The results obtained for $N_{\mathrm{ch}} = 3$ were very similar, so that it is reasonable to assume that considering more than five channels does not change the capacity results at all. This result can be explained by the large phase mismatch of FWM between spectral components with very large frequency separation.[4]

The channel capacity results for a link with $L = 2000\,\mathrm{km}$ SSMF with ideal distributed and lumped amplification are shown in Figure 6.10. The system parameters used for calculating the depicted curves are the same as in Section 5.4.4.

The only impairments that degrade the capacity in this system model are ASE noise and inter-channel FWM noise from the co-propagating WDM channels. This FWM noise has lower power than the intra-channel FWM noise considered in Chapter 5. The values depicted in Figure 6.10 are therefore higher than those shown in Figure 5.11. Qualitatively, the results for the WDM and the single channel scenario are similar. The largest capacity ($C \approx 8.0\,\mathrm{bit/s\,/\,Hz}$) is obtained with ideal Raman amplification. The optimum SNR is approximately 26.15 dB, corresponding to $10\log_{10}(P_s/1\,\mathrm{mW}) = -2.6\,\mathrm{dBm}$. Lumped amplification performs worse at any span length. The maximum capacity at $L_{\mathrm{span}} = 50\,\mathrm{km}$ ($100\,\mathrm{km}$) is $C \approx 6.9\,\mathrm{bit/s\,/\,Hz}$ ($5.5\,\mathrm{bit/s\,/\,Hz}$) at an SNR of approximately 22.7 dB (18.2 dB), corresponding to an average input power of $-0.2\,\mathrm{dBm}$ ($-2.3\,\mathrm{dBm}$).

The capacity curve of the system using Raman amplification can be compared to the corresponding result reported in Section 6.2.4. In Figure 6.6, the green curve ($L = 2000\,\mathrm{km}$) peaks at approximately $7\,\mathrm{bit/s\,/\,Hz}$. Hence, the maximum capacity values of the two models differ by around $1\,\mathrm{bit/s\,/\,Hz}$; the optimum SNR value of the time-domain model is approximately 2 dB lower than that of the frequency-domain model. The numerical uncertainty of the time-domain model (a few tenths of bits/s/Hz [EKW+10]) is not an

[4] In [EKW+10], the required (and sufficient) number of WDM channels is $N_{\mathrm{ch}} = 5$, too. Here, the reason lies in the walk-off between the channels.

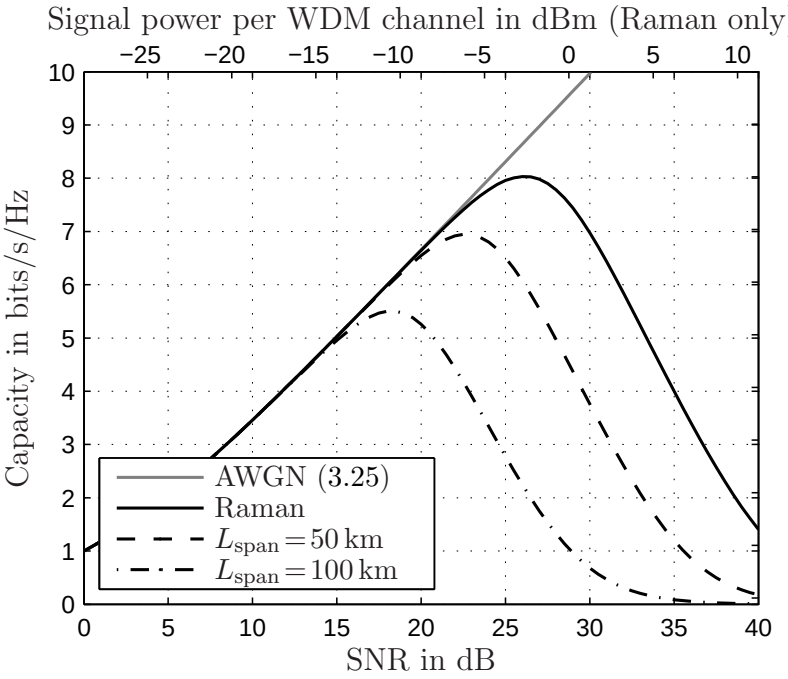

Figure 6.10: Channel capacity in bits/s/Hz of 2000-km SSMF link with distributed and lumped amplification over SNR in dB, $10\log_{10}(P_s/P_{\text{ASE}})$. Top x-axis shows signal power per WDM channel in dBm, $10\log_{10}(P_s/1\,\text{mW})$, for the distributed amplification setup.

adequate explanation for the observed capacity difference. Moreover, only a very small fraction of the capacity difference can be attributed to the different input constellations (16-ring and Gaussian, respectively), cf. [EKW+10, Fig. 16].

A more satisfactory explanation for the observed capacity gap lies in the effect that inter-channel fiber nonlinearities have in the two models. Both models incorporate the same nonlinearities. In the time-domain model, they appear mainly as XPM$^{\mathcal{S}}$-induced phase noise, while the frequency-domain model is impaired by FWM-induced AWGN. From the decomposition shown in Figure 3.3, we know that the phase carries more information than the amplitude does ($I(X_{\shortparallel};Y_{\shortparallel}) > I(X_{\sphericalangle};Y_{\sphericalangle}|X_{\shortparallel})$).[5] The XPM$^{\mathcal{S}}$-induced phase noise mainly impairs the phase term, i. e. the dominant sub-channel. In contrast, FWM affects both sub-channels, but impairs the dominant sub-channel less (relative to XPM$^{\mathcal{S}}$) and thus allows a larger sum capacity. A deeper understanding of the time-domain model (and of the equivalence of nonlinear effects in the time and frequency domain) is required to corroborate this explanation.

It is important to note that none of the two discussed capacity calculation methods is "more correct" than the other. Ideally, both methods should deliver identical capacity results. The capacity estimates are therefore expected to converge with a further refinement of the two models. As outlined in Section 5.4.2, a refined frequency-domain model

[5]The decomposition of the AWGN channel with Gaussian input and with a 16-ring constellation yields nearly identical values at SNRs up to 25 dB.

would take the non-white FWM PSD as well as the (potential) correlation of FWM noise between sub-channels into account. In turn, the time-domain model could deliver more precise (and higher) capacity estimates by a systematic input optimization. Because of the channel's memory, the optimized input might consist of correlated symbols.

In their current formulation, i. e. in the absence spectral shaping, decorrelation and input optimization measures, the models mimic "practical" fiber-optic communication systems. The observed difference therefore suggests that fiber-optic channels can achieve higher "practical" capacities when the information is encoded in the frequency domain, e. g. using OFDM.

6.3.2 Capacity results for different link lengths

The capacity calculation was repeated for different link lengths. With ideal Raman amplification, we obtain the results shown in Figure 6.11. A channel capacity of $C \approx 6.5\,\mathrm{bit/s\,/\,Hz}$ can be achieved on a 10000-km link.

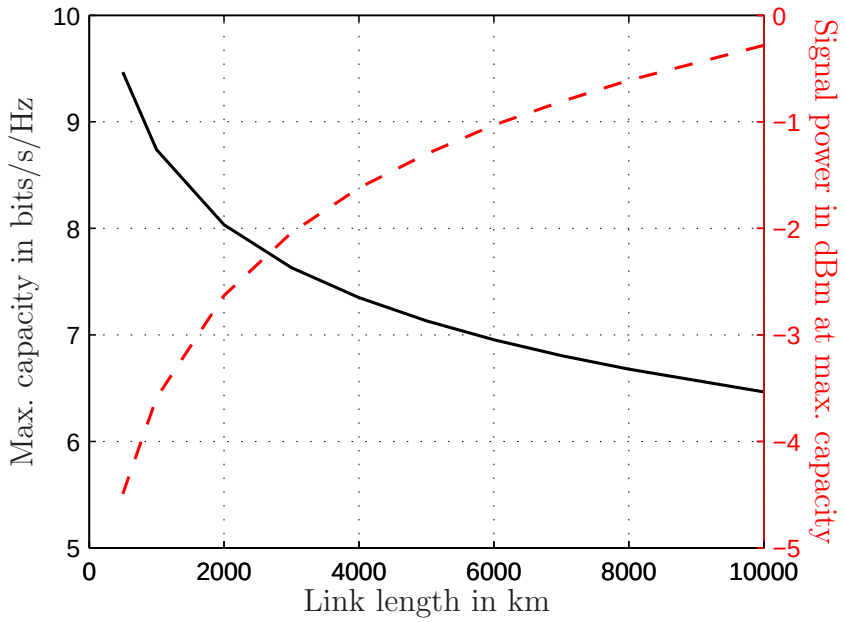

Figure 6.11: Maximum channel capacity in bits/s/Hz (black solid line) and optimum signal power per WDM channel in dBm, $10\log_{10}(P_s/1\,\mathrm{mW})$, (red dashed line) over SSMF length in km. Ideal distributed amplification is used.

At all link lengths, the capacities are slightly higher in the frequency-domain model than in the time-domain model (cf. Figures 6.11 and 6.6) as discussed in the previous section.

The optimum input power shown in Figure 6.11 increases with the fiber length. As explained in Section 5.4.4, the reason for this behavior lies in the length-dependence of the FWM noise power which grows slower than linearly with distance. This observation

is a second disagreement with the results of the time-domain model. It might have its reason in the same qualitative differences between the nonlinear effects in both models discussed in Section 6.3.1.

6.3.3 Nonlinear signal-noise interaction in WDM systems

Finally, we include nonlinear signal-ASE and ASE-ASE interactions in the capacity calculations. As signal-ASE interactions through SPM are small compared to those through FWM (see Section 5.5.2), we consider only FWM in the following. Hence, each subchannel is again modeled as an AWGN channel.

As in Section 5.5, we obtain a lower bound on the capacity by assuming that the entire ASE noise power propagates with the signal along the full link length. Instead of scaling the FWM noise PSD shown in Figure 6.9 with the average signal power per channel P_s, we scale it with $(P_s + P_{\mathrm{ASE}})^3$ (5.25).

Electronic predistortion of each individual WDM channel removes only nonlinear signal-signal interactions. The nonlinear interactions of signal and noise remain. In addition, we therefore need to take into account the FWM noise that has its origin in the nonlinear intra-channel signal-ASE (and ASE-ASE) interaction discussed in Section 5.5.2.

The channel capacity calculation was repeated including FWM noise from signal-ASE interaction. The resulting capacity decrease is very small. For distributed amplification, it amounts to less than 0.01 bit/s / Hz for $L = 500$ km, 0.045 bit/s / Hz for $L = 2000$ km and 0.4 bit/s / Hz for $L = 10000$ km. At $L = 10000$ km, the optimum SNR is reduced by 1 dB. It must be noted again that the actual capacity decrease is smaller than these values due to the distributed generation of noise that was neglected in the calculation.

We can conclude that nonlinear signal-ASE interaction has a negligible effect on the channel capacity for medium distances and a small effect in very long links. This result is in agreement with the observations reported in [EKW+10].

6.4 Summary

The effect that fiber nonlinearities have on the capacities of the individual channels in an optical WDM network were discussed in this chapter. This discussion assumes that intra-channel nonlinearities can be removed (e. g. by electronic predistortion), whereas inter-channel nonlinearities remain as fundamental impairments. After a review of the existing literature, the following main results were obtained:

- ▷ The nonlinear system effects SPM$^{\mathcal{S}}$, XPM$^{\mathcal{S}}$ and FWM$^{\mathcal{S}}$ were introduced using coupled differential equations that describe the propagation of WDM channels. These intra- (SPM$^{\mathcal{S}}$) and inter-channel (XPM$^{\mathcal{S}}$, FWM$^{\mathcal{S}}$) effects are related to SPM, XPM and FWM.

- ▷ A phenomenological channel model (PCM) was proposed that is able to reproduce

capacity results of a published numerical study. The PCM models the optical channel in the time domain as an AWGN with additional power-dependent phase noise and additive noise. Only two existing capacity results are required to estimate the two free parameters of the PCM. For a given link, the same set of parameters can be used to calculate the capacity for different SNR values and input constellations.

- ▷ The PCM was used to obtain capacity results for dispersion-managed as well as uncompensated links. The maximum channel capacity for a 2000-km link with ideal distributed amplification is approximately 7 bit/s / Hz, obtained in the absence of inline dispersion compensation.

- ▷ Spectral loss was discussed as one possible physical explanation for the power-dependent additive noise term.

- ▷ The frequency-domain model introduced in Chapter 5 was extended to the analysis of inter-channel nonlinearities in a WDM system. The channel capacity can be calculated from the FWM power spectral density in the absence of intra-channel FWM. No more than five WDM channels have to be considered for this calculation at $B = 100$ GHz.

- ▷ A capacity value of around 8 bit/s / Hz was found for the 2000-km link. The channel capacity obtained with lumped amplification does not exceed 7 bit/s / Hz. The time-domain and frequency-domain models deliver (slightly) different results. A possible reason for this (small) difference lies in the unequal distribution of information between the sub-channels represented by the amplitude and phase terms of the decomposition of mutual information.

- ▷ The maximum capacity and the optimum input power per WDM channel were calculated for different link lengths. The capacity ranges from 9.5 bit/s / Hz for a 500-km link to 6.5 bit/s / Hz for $L = 10000$ km. The optimum input power increases with distance.

- ▷ Finally, the interaction of signal and noise was included in the capacity calculation. Both signal-ASE and ASE-ASE interaction through intra- and inter-channel nonlinearities have to be considered. The results confirm that the nonlinear interaction of signal and noise has a very small effect.

7

Conclusions

How much information can we transmit over an optical single-mode fiber with arbitrarily low probability of error? Multiple answers to this question are suggested in this thesis; a single valid answer does not exist. The channel capacity of a fiber-optic communication system depends on a large variety of system parameters but also on how practical system constraints render physical effects fundamental, i. e. purely random or irrevertible. In any case, the answers given in this thesis indicate that today's practical fiber-optic communication systems are much further away from the Shannon limit than for instance wireless or copper-based communication systems. Improved system components as well as advanced signal processing methods, line coding and channel coding schemes will help close the gap to the Shannon limit in the future.

In the effort of increasing the data rate of future optical systems, it is important to know how far away from the fundamental limit the system operates. The methods introduced in this thesis can be used to answer this question individually for each specific system setup. The main contributions of this thesis can be summarized as follows:

- ▷ The linear and nonlinear processes that affect the signal propagation in optical fibers were derived rigorously from physical equations. In the development and formulation of (inevitably abstract) information-theoretic channel models, the implicit assumptions and simplifications remain transparent.

- ▷ A method of decomposing the mutual information in terms of polar coordinates was developed. This method essentially splits the channel into an amplitude and a phase part and yields insight into how information is conveyed over the channel. The decomposition was applied to the AWGN channel with Gaussian as well as non-optimum inputs. An application of the decomposition method to partially coherent

channels reveals a property of such channels that we call "spectral loss". Three different nonlinear effects (SPM, XPM and XPM$^{\mathcal{S}}$) render the fiber-optic channel partially coherent so that the proposed (general) decomposition method has an important application to fiber-optic channels.

- ▷ Polarization-mode dispersion (PMD) was discussed as a fading phenomenon. It was shown that coherent receivers should detect the light in two orthogonal polarizations separately. PMD as a linear effect does not reduce the MIMO capacity. If fiber nonlinearities are considered, cross-polariztion modulation leads to fast polarization rotations and hence interference between polarization-multiplexed signals. The variance of this interference term was calculated in a simple numerical example to show that the effect of XPolM is very small at optimum power levels. (These optimum power levels were determined in Chapters 5 and 6.)

- ▷ A frequency-domain model of the nonlinear fiber-optic communication channel was proposed for capacity calculation. In this model, the nonlinear effects of SPM, XPM and FWM can be treated separately. The same model was used to calculate the capacity for different fiber types (DSF, SSMF), amplification schemes (Raman, EDFA), nonlinear distortions (intra- or inter-channel effects) and signal-ASE interaction. In all cases, the capacity is limited by ASE noise at low power levels and by FWM noise at high power levels. It has a distinct optimum between the two regions, yielding an optimum input signal power.

- ▷ SPM and XPM were shown to have no fundamental effect on the capacity. The nonlinear effect that limits the channel capacity is FWM. An approximate expression for the power of a single FWM product was shown to be sufficiently accurate for capacity calculations even at very high signal power levels. The number of FWM products at a given frequency was determined analytically. FWM generates AWGN so that each spectral sub-channel can be modeled as an AWGN channel. The length dependence of FWM was discussed. As this function is (at most) quadratic in z, secondary parameters such as the mean nonlinear phase shift cannot be used to obtain generally valid scaling laws for capacity.

- ▷ The FWM power is reduced by chromatic dispersion; the highest capacity is therefore achieved with SSMF. Ideal distributed (Raman) amplification enables higher capacity values than ideal lumped (EDFA) amplification in general.

- ▷ The most fundamental capacity limit was obtained by considering only nonlinear signal-ASE interactions. At large bandwidths, the resulting capacity for a 2000-km link is approximately 10 bit/s / Hz. This value is limited by signal-ASE interaction through FWM, whereas ASE-ASE interaction and signal-ASE interaction through SPM play a subordinate role.

- ▷ The frequency-domain model was extended to the analysis of WDM systems, and a maximum capacity of approximately 8 bit/s / Hz was found for the 2000-km link. In WDM systems, inter-channel signal-signal interactions dominate over signal-ASE interactions.

- ▷ Based on observations made from a published numerical study [EKW+10], a simple phenomenological time-domain channel model was proposed. The capacity of the proposed channel model can be rapidly calculated using the polar decomposition method. Spectral loss is identified as an important effect in this model. The results agree very well with those of the published numerical study and show only small differences to the results of the frequency-domain model. A result from the decomposition method was used to provide a potential explanation for this difference.

In short, the main contribution of this thesis is that – for the first time – an approach for estimating the channel capacity was proposed that is neither approximate nor heuristic nor numerical. This approach is complemented by the decomposition of mutual information that provides insight as well as a rapid numerical calculation tool.

There are a number of interesting questions and problems that remain open for future research:

- ▷ The total channel capacity of an optical fiber is very high due to the large usable bandwidth. If this bandwidth is fully occupied, channels at high frequencies will lose power to lower-frequency channels due to Raman scattering. Finding the optimum input power spectral density requires an iterative water-filling approach. As Raman amplification is infeasible in such a broadband system, the capacity would have to be calculated assuming lumped amplification.
- ▷ The FWM noise spectrum is not flat. Optimizing the input PSD using an iterative water-filling algorithm would increase (or rather achieve) the capacity. Performing this procedure separately for each channel in a WDM system is expected to produce soft guard bands between WDM channel.
- ▷ The proposed frequency-domain model ignores any information that may be contained in the correlation of the FWM noise between the sub-channels. Decorrelation techniques could therefore be applied to obtain more accurate (higher) capacity estimates.
- ▷ In the context of signal-ASE interaction, the capacity calculation can be refined using an effective ASE power level instead of assuming that the full ASE noise power is added at the transmitter.
- ▷ Polarization multiplexing can double the capacity of fiber-optic channels almost effortlessly. Presumably, the main additional limitation in WDM PolMux systems is set by cross-polarization modulation. This question requires a more detailed (possibly numerical) investigation.
- ▷ Many practical optical communication systems still use direct-detection receivers. The noise statistics encountered at this receiver are very complicated due to the electrical filtering that follows the photodiode. Detailed numerical calculations are required to determine whether the capacity of DD systems is accurately predicted by noncoherent channel models.

A

Calculation of the number of four-wave mixing products

The generation of four-wave mixing products obeys the condition $f_m = f_n + f_p - f_q$ with $f_q \neq f_n$, $f_q \neq f_p$. The number of FWM products at a specific frequency f_m as well as the total number of FWM products are derived in the following as a function of N_{sc}.

This problem was adressed in the context of intermodulation interference in radio systems long before communication over fiber was possible [Bab53]. Our calculations assume equally spaced frequencies $f_1, f_2, \ldots, f_{N_{\mathrm{sc}}}$. Methods for finding the number of FWM products in the more general scenario of unequally spaced carriers were reported in [HT98, CYK00].

In Section A.1, the calculation is explained in detail and separately for degenerate and non-degenerate products. Section A.2 gives an example that may be helpful to consult when reproducing the calculations.

A.1 The number of FWM products

We consider a system with N_{sc} channels or subcarriers, located at equally spaced frequencies $f_1, f_2, \ldots, f_{N_{\mathrm{sc}}}$. What we are interested in is the number of FWM products arising at a given frequency f_m.

We differentiate between *degenerate* (DG) ($f_n = f_p$) and *non-degenerate* (NDG) ($f_n \neq f_p$) FWM products. No distinction is made between $f_m = f_n + f_p - f_q$ and $f_m = f_p + f_n - f_q$, so that every non-degenerate FWM product is actually two-fold. In cases where $f_q = f_n$ or $f_q = f_p$, no new frequency is generated. These cases give rise to SPM ($f_m = f_n = f_p = f_q$)

and XPM ($f_m = f_n$, $f_p = f_q$).

To formalize the problem, we introduce integer indices m, n, p, q and ask

- ▷ $m = n + p - q$,
- ▷ $\{n, p, q\} \in \mathbb{N}$ with $1 \leq n, p, q \leq N_{\mathrm{sc}}$,
- ▷ $m \in \mathbb{Z}$ with $2 - N_{\mathrm{sc}} \leq m \leq 2N_{\mathrm{sc}} - 1$,
- ▷ $q \neq n$ and $q \neq p$,

i. e. the original channels are located at $n = 1 \ldots N_{\mathrm{sc}}$, and new products can arise anywhere between $m = 2 - N_{\mathrm{sc}}$ and $m = 2N_{\mathrm{sc}} - 1$.

The total number of FWM products is obtained by counting all combinatorial combinations of $\{n, p, q\}$ satisfying the above-mentioned constraints:

- ▷ There are N_{sc} different possibilities for n.
- ▷ Either $p = n$ (degenerate product, one possibility) or $p \neq n$ (non-degenerate product, $N_{\mathrm{sc}} - 1$ possibilities).
- ▷ For q, $N_{\mathrm{sc}} - 1$ (degenerate) or $N_{\mathrm{sc}} - 2$ (non-degenerate) options are left.

Dividing the number of non-degenerate products by two, this sums up to

$$N_{\mathrm{FWM,total}} = N_{\mathrm{sc}} \cdot 1 \cdot (N_{\mathrm{sc}} - 1) + \frac{N_{\mathrm{sc}} \cdot (N_{\mathrm{sc}} - 1) \cdot (N_{\mathrm{sc}} - 2)}{2} = \frac{(N_{\mathrm{sc}}^3 - N_{\mathrm{sc}}^2)}{2}. \tag{A.1}$$

A.1.1 Degenerate FWM products

Degenerate FWM products satisfy $f_m = f_n + f_n - f_q$ or $m = 2n - q$. The total number of degenerate FWM products is

$$N_{\mathrm{FWM,total,DG}} = N_{\mathrm{sc}} \cdot (N_{\mathrm{sc}} - 1) = N_{\mathrm{sc}}^2 - N_{\mathrm{sc}}. \tag{A.2}$$

Now we determine the number of combinations at a given index m (i. e. the number of degenerate FWM products falling on a given frequency f_m). There are N_{sc} possibilities for n. For each of these, and for any given m, there exists exactly one or no q that satisfies $m = 2n - q$. No q exists if

(i) $q = n$, or if

(ii) $q < 1$, or if

(iii) $q > N_{\mathrm{sc}}$.

Ad (i): $q = n$ is equivalent to $m = n$. If $1 \leq m \leq N_{\mathrm{sc}}$, this is one single case, otherwise none.

Ad (ii): $q < 1$ is equivalent to $2n \leq m$ or $n \leq m/2$. If $m > 0$, there are $\lfloor m/2 \rfloor$ possibilities. If $m \leq 0$, there is no case in which $q < 1$.

Ad (iii): $q > N_{\mathrm{sc}}$ is equivalent to $2n > N_{\mathrm{sc}}+m$ or $n > (N_{\mathrm{sc}}+m)/2$. If $(N_{\mathrm{sc}}+m)/2 \leq N_{\mathrm{sc}}$ or $m \leq N_{\mathrm{sc}}$, there are $\lceil N_{\mathrm{sc}}/2 - m/2 \rceil$ possible q's. If $m > N_{\mathrm{sc}}$, there is no case in which $q > 1$.

In summary, the number of degenerate FWM products for a given m is

$$N_{\mathrm{FWM,DG}}(m) = \begin{cases} 0, & m < -N_{\mathrm{sc}}, \\ N_{\mathrm{sc}} - \lceil (N_{\mathrm{sc}} - m)/2 \rceil, & -N_{\mathrm{sc}} \leq m \leq 0, \\ N_{\mathrm{sc}} - 1 - \lfloor m/2 \rfloor - \lceil (N_{\mathrm{sc}} - m)/2 \rceil, & 1 \leq m \leq N_{\mathrm{sc}}, \\ N_{\mathrm{sc}} - \lfloor m/2 \rfloor, & N_{\mathrm{sc}} < m \leq 2N_{\mathrm{sc}}, \\ 0, & m > 2N_{\mathrm{sc}}. \end{cases} \tag{A.3}$$

The number of degenerate FWM products falling into the original signal band ($1 \leq m \leq N_{\mathrm{sc}}$) can be written as

$$N_{\mathrm{FWM,DG}}(m) = \begin{cases} N_{\mathrm{sc}}/2 - 1, & 1 \leq m \leq N_{\mathrm{sc}}, N_{\mathrm{sc}} \text{ even}, \\ N_{\mathrm{sc}}/2 - 1 - (-1)^m/2, & 1 \leq m \leq N_{\mathrm{sc}}, N_{\mathrm{sc}} \text{ odd}. \end{cases} \tag{A.4}$$

A.1.2 Non-degenerate FWM products

Non-degenerate FWM products can arise anywhere between $3 - N_{\mathrm{sc}} \leq m \leq 2N_{\mathrm{sc}} - 2$. The total number is given by (A.1)−(A.2):

$$N_{\mathrm{FWM,total,NDG}} = \frac{N_{\mathrm{sc}}^3}{2} - \frac{3N_{\mathrm{sc}}^2}{2} + N_{\mathrm{sc}}. \tag{A.5}$$

To determine the number of non-degenerate FWM products at a given index m, we follow the procedure of Section A.1.1, however with the additional complication that n and p are now independent. We have the following conditions:

- ▷ $m = n + p - q$,
- ▷ all $\{n, p, q\}$ are different, i. e. $n \neq p$, $n \neq q$, and $p \neq q$,
- ▷ $1 \leq n, p, q \leq N_{\mathrm{sc}}$,
- ▷ The order of n and p is not considered, i. e. we make no distinction between $\{n, p\}$ and $\{p, n\}$. This simply introduces a factor $1/2$ in every calculation.

For every combination $\{n, p\}$ and given m there exists exactly one or no q that satisfies all conditions. No q exists if

(i) $q = n$ or $q = p$, or if

(ii) $q < 1$, or if

(iii) $q > N_{\mathrm{sc}}$.

There are $N_{\mathrm{sc}} \cdot (N_{\mathrm{sc}} - 1)/2$ possibilities for $\{n, p\}$. Hence, the number of non-degenerate FWM products is given by $(N_{\mathrm{sc}}^2 - N_{\mathrm{sc}})/2$ less the number of cases (i), (ii), (iii).

Ad (i): $q = n$ or $q = p$ is equivalent to $n = m$ or $p = m$. If $1 \leq m \leq N_{\mathrm{sc}}$, there are $((N_{\mathrm{sc}} - 1) + (N_{\mathrm{sc}} - 1))/2 = N_{\mathrm{sc}} - 1$ possibilities for this case. Outside this range, there are no possibilities to count.

Ad (ii): Because $m = n + p - q$, $q < 1$ is equivalent to $n + p \leq m$. Counting possibilities for the case of equality, we have $m - 1$ possible n's, and exactly one p for every n. Subtracting the case $n = p$ if m is even, we get

- ▷ 0 possibilities if $m \leq 2$ or $m > 2N$,
- ▷ $(m - 2)/2$ possibilities if $2 < m \leq N_{\mathrm{sc}}$, m even,
- ▷ $(m - 1)/2$ possibilities if $2 < m \leq N_{\mathrm{sc}}$, m odd,
- ▷ $(2N_{\mathrm{sc}} - m)/2$ possibilities if $N_{\mathrm{sc}} < m \leq 2N_{\mathrm{sc}}$, m even,
- ▷ $(2N_{\mathrm{sc}} - m + 1)/2$ possibilities if $N_{\mathrm{sc}} < m < 2N_{\mathrm{sc}}$, m odd.

The total number of possibilities for $n + p \leq m$ is obtained by summing up the number of cases of each equality contained in the inequality. For instance, the calculation for even $m \in [3, N_{\mathrm{sc}}]$ is

$$\begin{aligned} N_{(\mathrm{ii})}(m) &= \sum_{a=1}^{m} (\text{possibilities for } n + p = a) \\ &= 0 + 0 + (3 - 1)/2 + (4 - 2)/2 + (5 - 1)/2 + \ldots \\ &= \frac{1}{2} \cdot \left(\sum_{a=3}^{m} q - \frac{m-2}{2} \cdot 2 - \frac{m-2}{2} \cdot 1 - 1 - 2 \right) \\ &= m^2/4 - m/2. \end{aligned}$$

In summary, the number of possibilities for (ii) is

$$N_{(\mathrm{ii})}(m) = \begin{cases} 0, & m < 1, \\ m^2/4 - m/2, & 1 \leq m \leq N_{\mathrm{sc}}, m \text{ even}, \\ m^2/4 - m/2 + 1/4, & 1 \leq m \leq N_{\mathrm{sc}}, m \text{ odd}, \\ -N_{\mathrm{sc}}^2/2 - m^2/4 - N_{\mathrm{sc}}/2 + N_{\mathrm{sc}} m, & N_{\mathrm{sc}} < m \leq 2N_{\mathrm{sc}}, m \text{ even}, \\ -N_{\mathrm{sc}}^2/2 - m^2/4 - N_{\mathrm{sc}}/2 + N_{\mathrm{sc}} m + 1/4, & N_{\mathrm{sc}} < m \leq 2N_{\mathrm{sc}}, m \text{ odd}, \\ N_{\mathrm{sc}}^2/2 - N_{\mathrm{sc}}/2, & m > 2N_{\mathrm{sc}}. \end{cases} \tag{A.6}$$

Ad (iii): Because $m = n + p - q$, $q > N_{\mathrm{sc}}$ is equivalent to $n + p > N_{\mathrm{sc}} + m$. Again, we count possibilities for the case of equality, $n + p = N_{\mathrm{sc}} + m$, and have

- ▷ 0 possibilities if $m \leq -N_{\mathrm{sc}} + 1$ or $m > N_{\mathrm{sc}}$,
- ▷ $(N_{\mathrm{sc}} + m - 2)/2$ possibilities if $-N_{\mathrm{sc}} + 1 < m \leq 0$, $N_{\mathrm{sc}} + m$ even,
- ▷ $(N_{\mathrm{sc}} + m - 1)/2$ possibilities if $-N_{\mathrm{sc}} + 1 < m \leq 0$, $N_{\mathrm{sc}} + m$ odd,
- ▷ $(N_{\mathrm{sc}} - m)/2$ possibilities if $0 < m \leq N_{\mathrm{sc}}$, $N_{\mathrm{sc}} + m$ even,
- ▷ $(N_{\mathrm{sc}} - m + 1)/2$ possibilities if $0 < m \leq N_{\mathrm{sc}}$, $N_{\mathrm{sc}} + m$ odd.

As before, the total number of possibilities for $n + p > N_{\mathrm{sc}} + m$ is obtained by summing up the number of cases of each equality contained in the inequality. Then, the number of possibilities for (iii) is

$$N_{(\mathrm{iii})}(m) = \begin{cases} \left(N_{\mathrm{sc}}^2 - N_{\mathrm{sc}}\right)/2, & m < -N_{\mathrm{sc}}, \\ \left(N_{\mathrm{sc}}^2 - m^2 + 2m - 2N_{\mathrm{sc}}m\right)/4, & -N_{\mathrm{sc}} \leq m < 1, N_{\mathrm{sc}} + m \text{ even}, \\ \left(N_{\mathrm{sc}}^2 - m^2 + 2m - 2N_{\mathrm{sc}}m - 1\right)/4, & -N_{\mathrm{sc}} \leq m < 1, N_{\mathrm{sc}} + m \text{ odd}, \\ \left(N_{\mathrm{sc}}^2 + m^2 - 2N_{\mathrm{sc}}m\right)/4, & 1 \leq m \leq N_{\mathrm{sc}}, N_{\mathrm{sc}} + m \text{ even}, \\ \left(N_{\mathrm{sc}}^2 + m^2 - 2N_{\mathrm{sc}}m - 1\right)/4, & 1 \leq m \leq N_{\mathrm{sc}}, N_{\mathrm{sc}} + m \text{ odd}, \\ 0, & m > N_{\mathrm{sc}}. \end{cases} \tag{A.7}$$

The last step is to evaluate $N_{\mathrm{FWM,NDG}}(m) = (N_{\mathrm{sc}}^2 - N_{\mathrm{sc}})/2 - N_{(\mathrm{i})}(m) - N_{(\mathrm{ii})}(m) - N_{(\mathrm{iii})}(m)$. We obtain

$$N_{\mathrm{FWM,NDG}}(m) = \begin{cases} 0, & m < -N_{\mathrm{sc}}, \\ \left\lceil (N_{\mathrm{sc}}^2 + m^2 - 2N_{\mathrm{sc}} - 2m + 2N_{\mathrm{sc}}m)/4 \right\rceil, & -N_{\mathrm{sc}} \leq m < 1, \\ \left\lceil (N_{\mathrm{sc}}^2 - 6N_{\mathrm{sc}} - 2m^2 + 2m + 4)/4 + \lfloor N_{\mathrm{sc}}m/2 \rfloor \right\rceil, & 1 \leq m \leq N_{\mathrm{sc}}, \\ \left\lfloor N_{\mathrm{sc}}^2 + m^2/4 - N_{\mathrm{sc}}m \right\rfloor, & N_{\mathrm{sc}} < m \leq 2N_{\mathrm{sc}}, \\ 0, & m > 2N_{\mathrm{sc}}. \end{cases} \tag{A.8}$$

The number of non-degenerate FWM products falling into the original signal band ($1 \leq m \leq N_{\mathrm{sc}}$) can be written as

$$N_{\mathrm{FWM,NDG}}(m) = \begin{cases} (N_{\mathrm{sc}}^2 - 6N_{\mathrm{sc}} + 2N_{\mathrm{sc}}m - 2m^2 + 2m + 4)/4, & 1 \leq m \leq N_{\mathrm{sc}}, N_{\mathrm{sc}} \text{ even}, \\ (N_{\mathrm{sc}}^2 - 6N_{\mathrm{sc}} + 2N_{\mathrm{sc}}m - 2m^2 + 2m + 4 + (-1)^m)/4, & 1 \leq m \leq N_{\mathrm{sc}}, N_{\mathrm{sc}} \text{ odd}. \end{cases} \tag{A.9}$$

A.2 An example

For the reader who would like to reproduce the combinatorics of Section A.1 by means of an example, Table A.1 lists all possible combinations of $n + p - q$ for $N_{\mathrm{sc}} = 4$.

Table A.1: All possible FWM cases for $N_{\mathrm{sc}} = 4$.

n	p	q	m	DG	NDG	other	n	p	q	m	DG	NDG	other
1	1	1	1			SPM	3	1	1	3			XPM
1	1	2	0	x			3	1	2	2		x	
1	1	3	-1	x			3	1	3	1			XPM
1	1	4	-2	x			3	1	4	0		x	
1	2	1	2			XPM	3	2	1	4		x	
1	2	2	1			XPM	3	2	2	3			XPM
1	2	3	0		x		3	2	3	2			XPM
1	2	4	-1		x		3	2	4	1		x	
1	3	1	3			XPM	3	3	1	5	x		
1	3	2	2		x		3	3	2	4	x		
1	3	3	1			XPM	3	3	3	3			SPM
1	3	4	0		x		3	3	4	2	x		
1	4	1	4			XPM	3	4	1	6		x	
1	4	2	3		x		3	4	2	5		x	
1	4	3	2		x		3	4	3	4			XPM
1	4	4	1			XPM	3	4	4	3			XPM
2	1	1	2			XPM	4	1	1	4			XPM
2	1	2	1			XPM	4	1	2	3		x	
2	1	3	0		x		4	1	3	2		x	
2	1	4	-1		x		4	1	4	1			XPM
2	2	1	3	x			4	2	1	5		x	
2	2	2	2			SPM	4	2	2	4			XPM
2	2	3	1	x			4	2	3	3		x	
2	2	4	0	x			4	2	4	2			XPM
2	3	1	4		x		4	3	1	6		x	
2	3	2	3			XPM	4	3	2	5		x	
2	3	3	2			XPM	4	3	3	4			XPM
2	3	4	1		x		4	3	4	3			XPM
2	4	1	5		x		4	4	1	7	x		
2	4	2	4			XPM	4	4	2	6	x		
2	4	3	3		x		4	4	3	5	x		
2	4	4	2			XPM	4	4	4	4			SPM

The number of degenerate and non-degenerate FWM products is shown in Figure A.1 for $N_{\mathrm{sc}} = 4$ and $N_{\mathrm{sc}} = 64$. It can be seen that FWM products arise symmetrically around the original band ($m = 1, \ldots, N_{\mathrm{sc}}$). With increasing N_{sc}, the number of non-degenerate FWM products becomes dominant compared to the number of degenerate products.

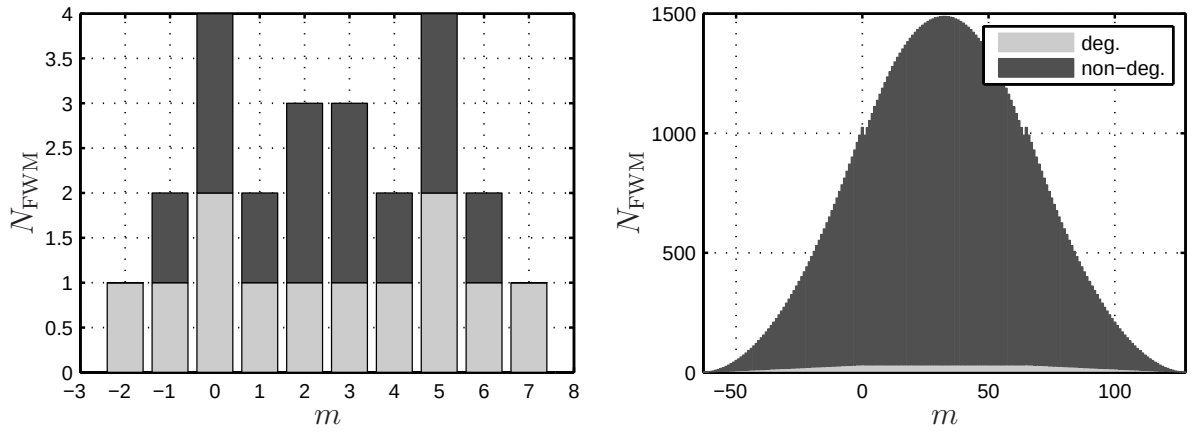

Figure A.1: Number of degenerate and non-degenerate FWM products at frequency $f_m = f_n + f_p - f_q$ for $N_{\mathrm{sc}} = 4$ (left) and $N_{\mathrm{sc}} = 64$ (right).

B

Notation, symbols and abbreviations

Mathematical notation

∇	Nabla operator
Δ	Laplace operator
$\overrightarrow{\boldsymbol{X}}$	geometric vector
$\overleftrightarrow{\boldsymbol{X}}$	tensor
$\times$	vector cross product
$\cdot$	vector dot product
$\star$	convolution
x^*	conjugate of argument
$\boldsymbol{H}^{-1}$	inverse of a matrix
$\boldsymbol{H}^{\mathrm{T}}$	transpose of a matrix
$\boldsymbol{H}^{\mathrm{H}}$	complex conjugate and transpose of a matrix
x^+	positive part of argument
$\overline{x}$	expectation of a random variable
$\lfloor x \rfloor$	floor function
$\lceil x \rceil$	ceiling function
$\lvert x \rvert$	absolute value or cardinality of argument
$x_{\shortparallel}$	short for $\lvert x \rvert$
$x_{\sphericalangle}$	phase angle of argument
$X \sim p(x)$	denotes distribution of random variable

$\det \boldsymbol{H}$	determinant of matrix $\boldsymbol{H}$
$\operatorname{erfc}(x)$	complementary error function
$\mathcal{E}_a\{X\}$	expectation of a random variable X with respect to a
$\mathcal{F}(x(t))$	Fourier transform
$\mathcal{F}^{-1}(X(f))$	inverse Fourier transform
$\mathrm{I}_\alpha(x)$	modified Bessel function of the first kind with order α
$\Im\{x\}$	imaginary part of argument
$\ln(x)$	natural logarithm
$\log(x)$	logarithm to unspecified base
$\log_b(x)$	logarithm to base b
$\max_a(\ldots)$	maximum value of arguments with respect to a
$\min_a(\ldots)$	minimum value of arguments with respect to a
$\operatorname{Prob}(x)$	probability of argument
$\Re\{x\}$	real part
$\operatorname{rect}(x)$	rectangular function
$\operatorname{sgn}(x)$	sign of x
$\operatorname{sinc}(x)$	sinus cardinalis, $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$

Physical constants

γ	Euler's constant	0.577215665
c_0	speed of light in vacuum	$2.99792458 \cdot 10^8\,\mathrm{m\,s^{-1}}$
e	base of the natural logarithm	2.718281828
ε_0	vacuum permittivity	$8.8541878 \cdot 10^{-12}\,\mathrm{A\,s\,V^{-1}\,m^{-1}}$
h	Planck constant	$6.62606896 \cdot 10^{-34}\,\mathrm{W\,s^2}$
$\jmath$	imaginary unit	$\sqrt{-1}$
k_B	Boltzmann constant	$1.3806504 \cdot 10^{-23}\,\mathrm{W\,s\,K^{-1}}$
μ_0	vacuum permeability	$4\pi \cdot 10^{-7}\,\mathrm{V\,s\,A^{-1}\,m^{-1}}$
π		3.14159265
Z_0	impedance of free space	$\sqrt{\mu_0/\varepsilon_0} \approx 376.73\ \Omega$

List of Symbols

α	attenuation	m^{-1} or $\mathrm{dB}\,/\,\mathrm{km}$
β	propagation constant	m^{-1}
β_m	m^{th}-order coefficient in Taylor series expansion of $\beta(\omega)$	$\mathrm{s}^m\,\mathrm{m}^{-1}$
$\Delta\beta$	phase mismatch	m^{-1}
γ	nonlinear fiber parameter	$\mathrm{W}^{-1}\,\mathrm{m}^{-1}$
γ_{eff}	effective nonlinear fiber parameter	$\mathrm{W}^{-1}\,\mathrm{m}^{-1}$
ε	ellipticity (polarization)	
ε	material permittivity	
$\overleftrightarrow{\varepsilon_r}$	relative permittivity tensor	
η	FWM efficiency	
θ	realization of circular random variable Θ	
θ	MZM asymmetry factor (2.166)	
θ	random angle of fiber segment, cf. (4.4)	
κ	parameter of von Mises distribution (3.70)	
λ_i	parameter of maximum entropy distribution (3.21)	
λ	wavelength	m
μ	mean of random variable	
μ°	mean direction of circular random variable	
ξ	random PMD phase angle	
ρ	water-filling constant (3.28)	
ρ°	resultant length of circular random variable	
ρ_f	free charge density	$\mathrm{A\,s\,m}^{-3}$
σ	Electronic contribution to cubic material polarization	$\mathrm{A\,m\,s\,V}^{-3}$
σ°	circular standard deviation of circular random variable	
$\sigma^2_{(n)}$	variance of real-valued Gaussian (noise) process	W
σ^2_{P}	variance of signal power in a WDM channel	W^2
τ	time variable	s
τ	normalized time variable	
τ_I, τ_1, τ_2	time constants	s
ϕ	inclination (polarization)	
ϕ_{NL}	nonlinear phase shift	
ϕ_{SPM}	nonlinear phase shift induced by SPM	
ϕ_{XPM}	nonlinear phase shift induced by XPM	
$\varphi(\tau)$	auto-correlation function	W
$\Phi(f)$	power spectral density	$\mathrm{W\,Hz}^{-1}$
χ	material susceptibility (general, time domain)	
$\overleftrightarrow{\boldsymbol{\chi}}^{(1)}$	linear material susceptibility tensor (time domain)	s^{-1}
$\chi^{(1)}_{ij}(t)$	component of linear susceptibility tensor (time domain)	s^{-1}

$\overleftrightarrow{\boldsymbol{X}}^{(1)}(f)$	linear material susceptibility tensor (frequency domain)	
$X_{ij}^{(1)}(f)$	component of linear susceptibility tensor (frequency domain)	
$\overleftrightarrow{\boldsymbol{\chi}}^{(3)}$	cubic material susceptibility tensor (time domain)	$\mathrm{m^2\,V^{-2}\,s^{-3}}$
$\chi_{ijkl}^{(3)}(t)$	component of cubic susceptibility tensor (time domain)	$\mathrm{m^2\,V^{-2}\,s^{-3}}$
$\overleftrightarrow{\boldsymbol{X}}^{(3)}(f)$	cubic material susceptibility tensor (frequency domain)	$\mathrm{m^2\,V^{-2}}$
$X_{ijkl}^{(3)}(f)$	component of cubic susceptibility tensor (frequency domain)	$\mathrm{m^2\,V^{-2}}$
$X_{\mathrm{eff}}^{(3)}$	effective cubic susceptibility	$\mathrm{m^2\,V^{-2}}$
ω	angular frequency, $\omega = 2\pi f$	Hz
$a(t)$	Nuclear contribution to cubic material polarization (time domain)	$\mathrm{A\,m\,V^{-3}}$
$A(\Delta f)$	Nuclear contribution to cubic material polarization (frequency domain)	$\mathrm{A\,m\,s\,V^{-3}}$
$a(z,t)$	normalized signal envelope	$\sqrt{\mathrm{W}}$
$A(z,f)$	Fourier transform of $a(z,t)$	$\sqrt{\mathrm{W}}\,\mathrm{s}$
A_{eff}	effective fiber core area	$\mathrm{m^2}$
B	bandwidth (of a signal)	Hz
B_{PSP}	bandwidth of PSPs	Hz
$b(t)$	Nuclear contribution to cubic material polarization (time domain)	$\mathrm{A\,m\,V^{-3}}$
$B(\Delta f)$	Nuclear contribution to cubic material polarization (frequency domain)	$\mathrm{A\,m\,s\,V^{-3}}$
$\overrightarrow{\boldsymbol{B}}(\overrightarrow{\boldsymbol{r}},t)$	Magnetic flux density	$\mathrm{V\,s\,m^{-2}}$
c_χ, c_ψ	eigenpolarization coefficients	
C	channel capacity	bit/s bit/channel use bit/symbol bit/s/Hz
$\overline{C}$	ergodic capacity	(as above)
C_x	x-% outage capacity	(as above)
$c_{\mathrm{R2}}, c_{\mathrm{R3}}$	parameters of phenomenological channel model	$\mathrm{W^{-2}}$
d_F	degeneracy factor (FWM)	
D	dispersion parameter	$\mathrm{ps\,nm^{-1}\,km^{-1}}$
$\overrightarrow{\boldsymbol{D}}$	Electric flux density	$\mathrm{A\,s\,m^{-2}}$
$D_\parallel, D_\perp$	polarization coupling parameters	$\mathrm{m^2\,V^{-2}}$
$D(P\|\|Q)$	Kullback-Leibler distance between PMFs P, Q (or PDFs p, q)	

$\overrightarrow{\boldsymbol{e}}_\chi, \overrightarrow{\boldsymbol{e}}_\psi$	eigenpolarization vectors	
$\overrightarrow{\boldsymbol{e}}, \overrightarrow{\boldsymbol{e}}^{(n)}$	polarization vectors	
$\hat{E}$	Normalized field amplitude $\hat{E}$ (2.52))	$\mathrm{V\,m^{-1}}$
$\overrightarrow{\boldsymbol{E}}(\overrightarrow{\boldsymbol{r}}, t)$	Electric field (time domain)	$\mathrm{V\,m^{-1}}$
$\overrightarrow{\boldsymbol{E}}(\overrightarrow{\boldsymbol{r}}, f)$	Electric field (frequency)	$\mathrm{V\,s\,m^{-1}}$
f, f'	frequency variable	Hz
Δf_{np}	$\lvert f_n - f_p \rvert$	Hz
$\overrightarrow{\boldsymbol{F}}$	modal distribution	
F_n	noise figure	∅ or dB
$F_{n,\mathrm{eff}}$	effective noise figure	∅ or dB
G	gain	∅ or dB
$G(z)$	(differential) gain	$\mathrm{m^{-1}}$ or dB / m
$G_{\mathrm{on-off}}$	on-off gain	∅ or dB
G_1	linear dispersion kernel (Volterra series)	$\mathrm{m^{-1}}$
G_3	fiber nonlinearity kernel (Volterra series)	$\mathrm{W^{-1}\,m^{-1}}$
$G_B(f_P, f_S)$	Brillouin gain spectrum	$\mathrm{W\,m^{-1}}$
$G_{\mathrm{B,max}}$	Maximum value of Brillouin gain spectrum	$\mathrm{W\,m^{-1}}$
$G_R(f_P, f_S)$	Raman gain spectrum	$\mathrm{W\,m^{-1}}$
$G_{\mathrm{R,max}}$	Maximum value of Raman gain spectrum	$\mathrm{W\,m^{-1}}$
$h(X)$	differential entropy	
$H(X)$	entropy	
$h(t)$	impulse response of LTI system	$\mathrm{s^{-1}}$
$H(f)$	LTI channel transfer function	
h_n	n^{th}-order Volterra kernel (general, time domain)	$\mathrm{s^{-}}^n$
H_1	first-order Volterra kernel (frequency domain)	
H_3	third-order Volterra kernel (frequency domain)	$\mathrm{W^{-1}}$
H_5	fifth-order Volterra kernel (frequency domain)	$\mathrm{W^{-2}}$
$h_R(t)$	Raman impulse response	$\mathrm{W\,m^{-1}\,s^{-1}}$
$\boldsymbol{H}$	transfer matrix	
$\overrightarrow{\boldsymbol{H}}(\overrightarrow{\boldsymbol{r}}, t)$	Magnetic field	$\mathrm{A\,m^{-1}}$
$\boldsymbol{I}_n$	$n \times n$ identity matrix	
i_P	photocurrent	A
I	intensity	$\mathrm{W\,m^{-2}}$
$I(X;Y)$	mutual information between random variables	
$\overrightarrow{\boldsymbol{J}}$	current density	$\mathrm{A\,m^{-2}}$
k	angular wavenumber, $k_0 = \omega/c_0 = 2\pi/\lambda$	$\mathrm{m^{-1}}$
L	fiber link length	km
L_{B}	beat length	m
L_c	constant length of fiber segment	m
L_{coh}	coherence length	m

L_{corr}	correlation length	m
L_{D}	dispersion length	m
L_{eff}	effective length	m
L_{NL}	nonlinear length	m
L_{span}	fiber length per span	km
L_{W}	walk-off length	m
$\overrightarrow{\boldsymbol{M}}$	magnetic polarization	V s m^{-2}
m_i	i^{th} moment of random variable	
m_i°	i^{th} trigonometric moment of circular random variable	
n	refractive index	
n_2	nonlinear index coefficient	$\text{m}^2\,\text{W}^{-1}$
$n_{2,\text{eff}}$	effective nonlinear index coefficient	$\text{m}^2\,\text{W}^{-1}$
n_{sp}	spontaneous emission factor (EDFA) or spontaneous scattering factor (Raman)	
$n(t)$	continuous-time noise process	$\sqrt{\text{W}}$
n	realization of random variable N (noise)	$\sqrt{W}$
$N_0/2$	power spectral density of real-valued white noise process	W Hz^{-1}
N_{ASE}	ASE noise power spectral density	W Hz^{-1}
$N_{\text{ASE}}(z)$	differential ASE noise PSD	$\text{W Hz}^{-1}\,\text{m}^{-1}$
N_{FWM}	number of FWM products	
N_{sc}	number of spectral components, sub-carriers or sub-channels	
N_{spans}	number of spans in multi-span system	
$\mathcal{N}_{\mathbb{C}}\left(\mu, 2\sigma_n^2\right)$	complex-valued Gaussian distribution	
$\mathcal{N}_{\mathbb{R}}\left(\mu, \sigma_n^2\right)$	real-valued Gaussian distribution	
P	(optical) power	W
$\overline{P}$	mean power	W
$P_{\text{critical,R}}$	critical pump power for SRS	W
$P_{\text{critical,B}}$	critical pump power for SBS	W
P_n	average noise power	W
P_s	average signal power	W
$P_{s,\text{rx}}$	average received signal power	W
$P^{(n)}$	average signal power of spectral component at f_n	W
P_{FWM}	total FWM power	W
$P_{\text{FWM}}^{(m)}$	total FWM power at f_m	W
$P_{\text{FWM}}^{(npq)}$	power of a single FWM product	W
$\overrightarrow{\boldsymbol{P}}(\overrightarrow{\boldsymbol{r}}, t)$	material polarization (time domain)	A s m^{-2}
$\overrightarrow{\boldsymbol{P}}(\overrightarrow{\boldsymbol{r}}, f)$	material polarization (frequency domain)	$\text{A s}^2\,\text{m}^{-2}$
$p(x)$	probability density function (PDF) of continuous random variable	

$P(x)$	probability mass function (PMF) of discrete random variable	
$\overrightarrow{\boldsymbol{r}}$	position vector	m
r	radius, $r = \sqrt{x^2 + y^2}$	m
r_{core}	fiber core radius	m
r_{eff}	effective fiber core radius	m
R	responsivity of photodiode	$\text{A}\,\text{W}^{-1}$
$\boldsymbol{R}$	rotation matrix (4.4)	
$s(z,t)$	dimensionless signal (time domain)	
$S(z,f)$	Fourier transform of $s(z,t)$	s
S	dispersion slope	$\text{ps}\,\text{nm}^{-2}\,\text{km}^{-1}$
T	temperature	K
t, t'	time variable	s
t_g	group delay	$\text{s}\,\text{m}^{-1}$
Δt_g	differential group delay	s
$\overline{\Delta t_g}$	mean differential group delay	s
$u(z,t)$	normalized signal envelope	
$\boldsymbol{U}$	transfer matrix of fiber segment (4.5)	
v	normalized frequency	
v_g	group velocity	$\text{m}\,\text{s}^{-1}$
V	variance of random variable	
V_{SPM}	variance of random SPM-induced phase shift	
V_{XPM}	variance of random XPM-induced phase shift	
V°	circular variance of circular random variable	
V_{xt}	variance of XPolM crosstalk term (4.29)	W
W	bandwidth (of a channel)	Hz
x, y, z	cartesian coordinates	m
x, y	realizations of random variables X, Y	$\sqrt{\text{W}}$
$\mathcal{X}, \mathcal{Y}$	support of random variables X, Y	
$x(t)$	(input) signal	$\sqrt{\text{W}}$
$X(f)$	Fourier transform of $x(t)$	$\sqrt{\text{W}}\,\text{s}$
$X^{(n)}(z)$	spectral component at f_n	$\sqrt{\text{W}}$
$y(t)$	(output) signal	$\sqrt{\text{W}}$
$Y(f)$	Fourier transform of $y(t)$	$\sqrt{\text{W}}\,\text{s}$
Z	wave impedance, $Z = Z_0/n$	Ω

List of abbreviations

ACF	auto-correlation function
AM	amplitude modulation
AMI	alternate mark inversion
APD	avalanche photodiode
ASE	amplified spontaneous emission
ASK	amplitude-shift keying
AWGN	additive white Gaussian noise
BCJR	algorithm by Bahl, Cocke, Jelinek, Raviv
BER	bit error rate
BPSK	binary phase-shift keying
cc	conjugate complex
CDF	cumulative distribution function
CSI	channel state information
CSRZ	carrier-suppressed return-to-zero
DC	direct current
DCF	dispersion-compensating fiber
DD	direct detection
DFB	distributed feedback (laser)
DFT	discrete Fourier transform
DGD	differential group delay
DOP	degree of polarization
DSF	dispersion-shifted fiber
DSP	digital signal processing
DWDM	dense wavelength division multiplexing
ECL	external cavity laser
EDFA	erbium-doped fiber amplifier
FEC	forward error correction
FNM	full numerical model
FWHM	full width at half maximum
FWM	four-wave mixing
$\text{FWM}^{\mathcal{S}}$	four-wave mixing (of WDM channels)
IDFT	inverse discrete Fourier transform
iFWM	intra-channel four-wave mixing
i. i. d.	independent and identically distributed
IM/DD	intensity modulation / direct detection
iXPM	intra-channel cross-phase modulation
LO	local oscillator
LTI	linear time-invariant
MI	mutual information

MIMO	multiple-input/multiple-output
MZM	Mach-Zehnder modulator
NLSE	Nonlinear Schroedinger equation
NRZ	non-return-to-zero
NZDSF	non-zero dispersion-shifted fiber
OFDM	orthogonal frequency division multiplexing
OOK	on-off keying
OSNR	optical signal-to-noise ratio
PBS	polarization beam splitter
PCM	phenomenological channel model
PDF	probability density function
PDG	polarization-dependent gain
PDL	polarization-dependent loss
PDM	polarization division multiplexing
PolMux	polarization multiplexing
PM	phase modulation
PMD	polarization mode dispersion
PMF	probability mass function
PSD	power spectral density
PSK	phase-shift keying
PSP	principal state of polarization
PTF	power transfer function
QAM	quadrature amplitude modulation
RDPS	residual dispersion per span
RZ	return-to-zero
SBS	stimulated Brillouin scattering
SL	spectral loss
SMF	single mode fiber
SNR	signal-to-noise ratio
SOP	state of polarization
SPM	self-phase modulation
$\mathrm{SPM}^{\mathcal{S}}$	self-phase modulation (of WDM channel)
SRS	stimulated Raman scattering
SSF	split-step Fourier method
SSMF	standard single mode fiber
WDM	wavelength division multiplexing
XPM	cross-phase modulation
$\mathrm{XPM}^{\mathcal{S}}$	cross-phase modulation (of WDM channels)
XPolM	cross-polarization modulation

Bibliography

[AB93] J. P. Aldis and A. G. Burr. The channel capacity of discrete time phase modulation in AWGN. *IEEE Transactions on Information Theory*, 39(1):184–185, January 1993.

[Agr01] G. P. Agrawal. *Nonlinear fiber optics*. Academic Press, San Diego, 3rd edition, 2001.

[Agr02] G. P. Agrawal. *Fiber-optic communication systems*. John Wiley & Sons, Inc., New York, 3rd edition, 2002.

[Agr04] G. P. Agrawal. *Lightwave technology: Components and devices*. John Wiley & Sons, Inc., Hoboken, 2004.

[Agr05] G. P. Agrawal. *Lightwave technology: Telecommunication systems*. John Wiley & Sons, Inc., Hoboken, 2005.

[AL01] D. Arnold and H.-A. Loeliger. On the information rate of binary-input channels with memory. In *Proceedings of the IEEE International Conference on Communications (ICC)*, volume 9, pages 2692–2695, Helsinki, Finland, June 2001.

[ALV+06] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and Wei Zeng. Simulation-based computation of information rates for channels with memory. *IEEE Transactions on Information Theory*, 52(8):3498–3508, August 2006.

[Arn03] D. Arnold. *Computing Information Rates of Finite-State Models with Application to Magnetic Recording*. PhD thesis, Eidgenössische Technische Hochschule Zürich, 2003.

[AS72] M. Abramowitz and I. A. Stegun, editors. *Handbook of mathematical functions*. United States Government Printing Office, Washington, 10th edition, 1972.

[AY78] J. AuYeung and A. Yariv. Spontaneous and stimulated Raman scattering in long low loss fibers. *IEEE Journal of Quantum Electronics*, 14(5):347–352, May 1978.

[Bab53] W. Babcock. Intermodulation interference in radio systems. *Bell System Technical Journal*, 32(1):63–73, January 1953.

[Bah06] C. Bahlmann. Directional features in online handwriting recognition. *Pattern Recognition*, 39(1):115–125, January 2006.

[Bar87] R. Barakat. Weak-scatterer generalization of the K-density function. II. Probability density of total phase. *Journal of the Optical Society of America A*, 4(7):1213–1219, July 1987.

[BBC87] S. Benedetto, E. Biglieri, and V. Castellani. *Digital transmission theory*. Prentice Hall, Englewood Cliffs, 1987.

[BC02] N. Benvenuto and G. Cherubini. *Algorithms for communications systems and their applications*. John Wiley & Sons, Inc., Chichester, 2002.

[BCJR74] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Transactions on Information Theory*, 20(2):284–287, March 1974.

[BDV07] H. G. Batshon, I. B. Djordjevic, and B. V. Vasic. An improved technique for suppression of intrachannel four-wave mixing in 40-Gb/s optical transmission systems. *IEEE Photonics Technology Letters*, 19(2):67–69, January 2007.

[BFBT06] M. Brodsky, N. J. Frigo, M. Boroditsky, and M. Tur. Polarization mode dispersion of installed fibers. *IEEE/OSA Journal of Lightwave Technology*, 24(12):4584–4599, December 2006.

[Bla53] N. M. Blachman. A comparison of the informational capacities of amplitude- and phase-modulation communication systems. *Proceedings of the IRE*, 41(6):748–759, June 1953.

[Bla87] R. E. Blahut. *Principles and practice of information theory*. Addison-Wesley, Reading, MA, 1987.

[BPRC+09] O. Bertran-Pardo, J. Renaudier, G. Charlet, P. Tran, H. Mardoyan, M. Salsi, and S. Bigo. Experimental assessment of interactions between nonlinear impairments and polarization-mode dispersion in 100-Gb/s coherent systems versus receiver complexity. *IEEE Photonics Technology Letters*, 21(1):51–53, January 2009.

[BPS98] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: information-theoretic and communications aspects. *IEEE Transactions on Information Theory*, 44(6):2619–2692, October 1998.

[BRHF02] B. J. Belzer, A. D. Risley, P. Hou, and T. R. Fischer. Capacity of AM-PSK on partially coherent fading channels. *IEEE Transactions on Communications*, 50(3):369–373, March 2002.

[Bri56] L. Brillouin. *Science and information theory*. Academic Press, New York, 1956.

[Bro04] J. Bromage. Raman amplification for fiber communications systems. *IEEE/OSA Journal of Lightwave Technology*, 22(1):79–93, January 2004.

[BW89] K. J. Blow and D. Wood. Theoretical description of transient stimulated Raman scattering in optical fibers. *IEEE Journal of Quantum Electronics*, 25(12):2665–2673, December 1989.

[BW99] M. Born and E. Wolf. *Principles of optics*. Cambridge University Press, Cambridge, 7th edition, 1999.

[Cah60] C. Cahn. Combined digital phase and amplitude modulation communication systems. *IRE Transactions on Communications Systems*, 8(3):150–155, September 1960.

[CFS86] D. Cygan, J. Franz, and G. Söder. Einfluß eines Filters auf nicht-gaußverteilte Zufallsprozesse. *Archiv für Elektronik und Übertragungstechnik (AEÜ)*, 40(6):377–384, June 1986.

[CMG+09] L. D. Coelho, L. Molle, D. Gross, N. Hanik, R. Freund, C. Caspar, E.-D. Schmidt, and B. Spinnler. Modeling nonlinear phase noise in differentially phase-modulated optical communication systems. *Optics Express*, 17(5):3226–3241, March 2009.

[CMLB05] A. Castaño-Martínez and F. López-Blázquez. Distribution of a sum of weighted noncentral chi-square variables. *Test*, 14(2):397–415, December 2005.

[CR01] G. Colavolpe and R. Raheli. The capacity of the noncoherent channel. *European Transactions on Telecommunications*, 12(4):289–296, July-August 2001.

[CT91] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, Inc., New York, 1991.

[CYK00] K.-D. Chang, G.-C. Yang, and W. C. Kwong. Determination of FWM products in unequal-spaced-channel WDM lightwave systems. *IEEE/OSA Journal of Lightwave Technology*, 18(12):2113–2122, December 2000.

[dA07] G. T. F. de Abreu. On the simulation of Tikhonov random processes. In *Proceedings of the IEEE International Conference on Communications (ICC)*, pages 5021–5027, Glasgow, UK, June 2007.

[DAPR07] I. B. Djordjevic, N. Alic, G. C. Papen, and S. Radic. Determination of achievable information rates (AIRs) of IM/DD systems and AIR loss due to chromatic dispersion and quantization. *IEEE Photonics Technology Letters*, 19(1):12–14, January 2007.

[DBDB02] E. Desurvire, D. Bayart, B. Desthieux, and S. Bigo. *Erbium-doped fiber amplifiers: Device and system developments*. John Wiley & Sons, Inc., New York, 2002.

[Des02a] E. Desurvire. *Erbium-doped fiber amplifiers: Principles and applications*. John Wiley & Sons, Inc., New York, 2002.

[Des02b] E. Desurvire. A quantum model for optically amplified nonlinear transmission systems. *Optical Fiber Technology*, 8(3):210–230, July 2002.

[Des02c] E. Desurvire. Quantum noise model for ultimate information-capacity limits in long-haul wdm transmission. *Electronics Letters*, 38(17):983–984, August 2002.

[Des06] E. B. Desurvire. Capacity demand and technology challenges for lightwave systems in the next two decades. *IEEE/OSA Journal of Lightwave Technology*, 24(12):4697–4710, December 2006.

[DSS91] Y. E. Dallal and S. Shamai (Shitz). An upper bound on the error probability of quadratic-detection in noisy phase channels. *IEEE Transactions on Communications*, 39(11):1635–1650, November 1991.

[DV06] I. B. Djordjevic and B. Vasic. Constrained coding techniques for the suppression of intrachannel nonlinear effects in high-speed optical transmission. *IEEE/OSA Journal of Lightwave Technology*, 24(1):411–419, January 2006.

[DVIG05] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov. Achievable information rates for high-speed long-haul optical transmission. *IEEE/OSA Journal of Lightwave Technology*, 23(11):3755–3763, November 2005.

[EFKW08] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer. Capacity limits of information transport in fiber-optic networks. *Physical Review Letters*, 101(16):163901–1–4, October 2008.

[EFKW10] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer. Capacity limits of information transmission in optically-routed fiber networks. *Bell Labs Technical Journal*, 14(4):149–162, 2010.

[EFW+08] R.-J. Essiambre, G. J. Foschini, P. J. Winzer, G. Kramer, and E. C. Burrows. The capacity of fiber-optic communication systems. In *Proceedings of the Optical Fiber Communication Conference (OFC)*, page OTuE1, San Diego, CA, USA, February 2008.

[EFWK08] R.-J. Essiambre, G. J. Foschini, P. J. Winzer, and G. Kramer. Exploring capacity limits of fibre-optic communication systems. In *Proceedings of the European Conference on Optical Communications (ECOC)*, page We.1.E.1, Brussels, Belgium, September 2008.

[EFWK09] R.-J. Essiambre, G. Foschini, P. Winzer, and G. Kramer. Capacity limits of fiber-optic communication systems. In *Proceedings of the Optical Fiber Communication Conference (OFC)*, page OThL1, San Diego, CA, USA, March 2009.

[EKFW08] R.-J. Essiambre, G. Kramer, G. J. Foschini, and P. J. Winzer. High spectral efficiency modulation for high capacity transmission. In *Digest of the IEEE/LEOS Summer Topical Meetings*, pages 113–114, Acapulco, Mexico, July 2008.

[EKW+10] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel. Capacity limits of optical fiber networks. *IEEE/OSA Journal of Lightwave Technology*, 28(4):662–701, February 2010.

[EMR99] R. J. Essiambre, B. Mikkelsen, and G. Raybon. Intra-channel cross-phase modulation and four-wave mixing in high-speed TDM systems. *Electronics Letters*, 35(18):1576–1578, September 1999.

[EWW+06] R.-J. Essiambre, P. J. Winzer, X. Q. Wang, W. Lee, C. A. White, and E. C. Burrows. Electronic predistortion and fiber nonlinearity. *IEEE Photonics Technology Letters*, 18(17):1804–1806, September 2006.

[EZC10] A. D. Ellis, J. Zhao, and D. Cotter. Approaching the non-linear Shannon limit. *IEEE/OSA Journal of Lightwave Technology*, 28(4):423–433, February 2010.

[FBF+07] M. Franceschini, G. Bongiorni, G. Ferrari, R. Raheli, F. Meli, and A. Castoldi. Fundamental limits of electronic signal processing in direct-detection optical communications. *IEEE/OSA Journal of Lightwave Technology*, 25(7):1742–1753, July 2007.

[Fel03] G. Feldhaus. *Volterrafilter zur Modellierung und Entzerrung optischer Übertragungssysteme mit Direktempfänger*. PhD thesis, Technische Universität Darmstadt, 2003.

[FEW+09] T. Freckmann, R.-J. Essiambre, P. J. Winzer, G. J. Foschini, and G. Kramer. Fiber capacity limits with optimized ring constellations. *IEEE Photonics Technology Letters*, 21(20):1496–1498, October 2009.

[FFRB06] M. Franceschini, G. Ferrari, R. Raheli, and G. Bongiorni. Fundamental limits of electronic dispersion compensation in optical communications with direct photodetection. *Electronics Letters*, 42(15):874–875, July 2006.

[FGV88] G. J. Foschini, L. J. Greenstein, and G. Vannucci. Noncoherent detection of coherent lightwave signals corrupted by phase noise. *IEEE Transactions on Communications*, 36(3):306–314, March 1988.

[FGW73] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein. On the selection of a two-dimensional signal constellation in the presence of phase jitter and Gaussian noise. *Bell System Technical Journal*, 52(6):927–965, July-August 1973.

[Fis96] N. I. Fisher. *Statistical analysis of circular data*. Cambridge University Press, Cambridge, 1996.

[FU98] G. D. Forney and G. Ungerboeck. Modulation and coding for linear Gaussian channels. *IEEE Transactions on Information Theory*, 44(6):2384–2415, October 1998.

[FVG89] G. J. Foschini, G. Vannucci, and L. J. Greenstein. Envelope statistics for filtered optical signals corrupted by phase noise. *IEEE Transactions on Communications*, 37(12):1293–1302, December 1989.

[Gei90] J. M. Geist. Capacity and cutoff rate for dense M-ary PSK constellations. In *Proceedings of the Military Communications Conference (MILCOM)*, volume 2, pages 768–770, Monterey, CA, USA, September/October 1990.

[GFCH08] B. Goebel, B. Fesl, L.D. Coelho, and N. Hanik. On the effect of FWM in coherent optical OFDM systems. In *Proceedings of the Optical Fiber Communication Conference (OFC)*, page JWA58, San Diego, CA, USA, February 2008.

[GG95] R. M. Gray and J. W. Goodman. *Fourier transforms*. Kluwer Academic Publishers, Boston, 1995.

[GHH10] B. Goebel, S. Hellerbrand, and N. Hanik. Link-aware precoding for nonlinear optical OFDM transmission. In *Proceedings of the Optical Fiber Communication Conference (OFC)*, page OTuE4, San Diego, CA, USA, March 2010.

[GHHH09] B. Goebel, S. Hellerbrand, N. Haufe, and N. Hanik. PAPR reduction techniques for coherent optical OFDM transmission. In *Proceedings of the International Conference on Transparent Optical Networks (ICTON)*, page Mo.B2.4, Sao Miguel, Portugal, June/July 2009.

[GJJV03] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE Journal on Selected Areas in Communications*, 21(5):684–702, June 2003.

[GK00] J. P. Gordon and H. Kogelnik. PMD fundamentals: Polarization mode dispersion in optical fibers. *Proceedings of the National Academy of Sciences*, 97(9):4541–4550, April 2000.

[GKH07] B. Goebel, M. Kuschnerov, and N. Hanik. On the effect of polarisation-mode dispersion on the channel capacity of coherent fibre-optic communication systems. In *Proceedings of the International Conference on Trans-*

parent Optical Networks (ICTON), volume 1, pages 155–159, Rome, Italy, July 2007.

[GLW63] J. P. Gordon, W. H. Louisell, and L. R. Walker. Quantum fluctuations and noise in parametric processes. II. *Physical Review*, 129(1):481–485, January 1963.

[GM05] A. Galtarossa and C. R. Menyuk, editors. *Polarization mode dispersion*. Springer, New York, 2005.

[GWL63] J. P. Gordon, L. R. Walker, and W. H. Louisell. Quantum statistics of masers and attenuators. *Physical Review*, 130(2):806–812, April 1963.

[HA91] P. A. Humblet and M. Azizoglu. On the bit error rate of lightwave systems with optical amplifiers. *IEEE/OSA Journal of Lightwave Technology*, 9(11):1576–1582, November 1991.

[Han95] N. Hanik. *Auswirkung nichtlinearer Fasereffekte auf die digitale Signalübertragung in Monomode-Glasfasern*. PhD thesis, Technische Universität München, Munich, Germany, May 1995.

[Han02] N. Hanik. Modelling of nonlinear optical wave propagation including linear mode-coupling and birefringence. *Optics Communications*, 214(1-6):207–230, December 2002.

[Har28] R. V. L. Hartley. The transmission of information. *Bell System Technical Journal*, 3:535–564, July 1928.

[HBF02] P. Hou, B. J. Belzer, and T. R. Fischer. Shaping gain of the partially coherent additive white Gaussian noise channel. *IEEE Communications Letters*, 6(5):175–177, May 2002.

[HBF03] P. Hou, B. J. Belzer, and T. R. Fischer. On the capacity of the partially coherent additive white Gaussian noise channel. In *Proceedings of the IEEE International Symposium on Information Theory*, page 372, Yokohama, Japan, June/July 2003.

[HC02] D. Hollenbeck and C. D. Cantrell. Multiple-vibrational-mode model for fiber-optic Raman gain spectrum and response function. *Journal of the Optical Socienty of America B*, 19(12):2886–2892, December 2002.

[Hel77] R. W. Hellwarth. Third-order optical susceptibilities of liquids and solids. *Progress in Quantum Electronics*, 5(1):1–68, 1977.

[HGH09] S. Hellerbrand, B. Goebel, and N. Hanik. Trellis shaping for reduction of the peak-to-average power ratio in coherent optical OFDM systems. In *Proceedings of the Optical Fiber Communication Conference (OFC)*, page JThA48, San Diego, CA, USA, March 2009.

[HJKM78] K. O. Hill, D. C. Johnson, B. S. Kawasaki, and R. I. MacDonald. CW three-wave mixing in single mode optical fibers. *Journal of Applied Physics*, 49(10):5098–5106, October 1978.

[HK02] K.-P. Ho and J. M. Kahn. Channel capacity of WDM systems using constant-intensity modulation formats. In *Proceedings of the Optical Fiber Communication Conference (OFC)*, pages 731–733, Anaheim, CA, USA, March 2002.

[HK04] S. Hranilovic and F. R. Kschischang. Capacity bounds for power- and band-limited optical intensity channels corrupted by Gaussian noise. *IEEE Transactions on Information Theory*, 50(5):784–795, May 2004.

[HL05] Y. Han and G. Li. Coherent optical communication using polarization multiple-input-multiple-output. *Optics Express*, 13(19):7527–7534, September 2005.

[HM10] H. Haunstein and M. Mayrock. OFDM spectral efficiency limits from fiber and system non-linearities. In *Proceedings of the Optical Fiber Communication Conference (OFC)*, page OThM7, San Diego, CA, USA, March 2010.

[Ho05a] K.-P. Ho. *Phase-modulated optical communication systems*. Springer, New York, 2005.

[Ho05b] Keang-Po Ho. Exact evaluation of the capacity for intensity-modulated direct-detection channels with optical amplifier noises. *IEEE Photonics Technology Letters*, 17(4):858–860, April 2005.

[HT98] B. Hwang and O. K. Tonguz. A generalized suboptimum unequally spaced channel allocation technique—Part I: In IM/DD WDM systems. *IEEE/OSA Journal of Lightwave Technology*, 46(8):1027–1037, August 1998.

[HWHP07] S. Hellerbrand, C. Weber, N. Hanik, and K. Petermann. Advanced modulation formats for electronic predistortion of intra-channel nonlinearities at 40 Gbit/s. In *Proceedings of the European Conference on Optical Communications (ECOC)*, page 6.2.4, Berlin, Germany, September 2007.

[HWK00] L. Hanzo, W. Webb, and T. Keller. *Single- and multi-carrier quadrature amplitude modulation*. John Wiley & Sons, Inc., Chichester, 2000.

[IDV07] M. Ivkovic, I. Djordjevic, and B. Vasic. Calculation of achievable information rates of long-haul optical transmission systems using instanton approach. *IEEE/OSA Journal of Lightwave Technology*, 25(5):1163–1168, May 2007.

[IK08] E. Ip and J. M. Kahn. Compensation of dispersion and nonlinear impairments using digital backpropagation. *IEEE/OSA Journal of Lightwave Technology*, 26(20):3416–3425, October 2008.

[ILBK08] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn. Coherent detection in optical fiber systems. *Optics Express*, 16(2):753–791, January 2008.

[Ip10] E. Ip. Nonlinear compensation using backpropagation for polarization-multiplexed transmission. *IEEE/OSA Journal of Lightwave Technology*, 28(6):939–951, March 2010.

[IT95] K. Inoue and H. Toba. Fiber four-wave mixing in multi-amplifier systems with nonuniform chromatic dispersion. *IEEE/OSA Journal of Lightwave Technology*, 13(1):88–93, January 1995.

[ITU02] ITU-T. *G.694.1: Spectral grids for WDM applications: DWDM frequency grid*, June 2002.

[JBS00] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan. *Simulation of communication systems*. Kluwer Academic / Plenum Publishers, New York, 2nd edition, 2000.

[JMST08] S. L. Jansen, I. Morita, T. C. W. Schenk, and H. Tanaka. Long-haul transmission of 16 $\times$ 52.5 Gbits/s polarization-division multiplexed OFDM enabled by MIMO processing. *Journal of Optical Networking*, 7(2):173–182, February 2008.

[JMST09] S. L. Jansen, I. Morita, T. C. W. Schenk, and H. Tanaka. 121.9-Gb/s PDM-OFDM transmission with 2-b/s/Hz spectral efficiency over 1000 km of SSMF. *IEEE/OSA Journal of Lightwave Technology*, 27(3):177–188, February 2009.

[JNK99] R. M. Jopson, L. E. Nelson, and H. Kogelnik. Measurement of second-order polarization-mode dispersion vectors in optical fibers. *IEEE Photonics Technology Letters*, 11(9):1153–1155, September 1999.

[KAvWW03] G. Kramer, A. Ashikhmin, A. J. van Wijngaarden, and X. Wei. Spectral efficiency of coded phase-shift keying for fiber-optic communication. *IEEE/OSA Journal of Lightwave Technology*, 21(10):2438–2445, October 2003.

[KB99] M. Karlsson and J. Brentel. Autocorrelation function of the polarization-mode dispersion vector. *Optics Letters*, 24(14):939–941, July 1999.

[KBA00] M. Karlsson, J. Brentel, and P. A. Andrekson. Long-term measurement of PMD and polarization drift in installed fibers. *IEEE/OSA Journal of Lightwave Technology*, 18(7):941–951, July 2000.

[KERR09] H. Kee, R. Epworth, A. Robinson, and R. Rickard. Spectral shaping for optical OFDM transmission. United States patent 7580630, August, 25, 2009.

[KH04] J. M. Kahn and K.-P. Ho. Spectral efficiency limits and modulation/detection techniques for DWDM systems. *IEEE Journal of Selected Topics in Quantum Electronics*, 10(2):259–272, March/April 2004.

[KJ82] S. Kotz and N. L. Johnson, editors. *Encyclopedia of statistical sciences*, volume 2. John Wiley & Sons, Inc., New York, 1982.

[KK97a] I. P. Kaminov and T. L. Koch, editors. *Optical fiber telecommunications III A*. Academic Press, San Diego, 1997.

[KK97b] I. P. Kaminov and T. L. Koch, editors. *Optical fiber telecommunications III B*. Academic Press, San Diego, 1997.

[KL02a] I. P. Kaminov and T. Li, editors. *Optical fiber telecommunications IV A: Components*. Academic Press, San Diego, 2002.

[KL02b] I. P. Kaminov and T. Li, editors. *Optical fiber telecommunications IV B: Systems and impairments*. Academic Press, San Diego, 2002.

[KLW08a] I. P. Kaminov, T. Li, and A. E. Willner, editors. *Optical fiber telecommunications V A: Components and subsystems*. Academic Press, San Diego, 2008.

[KLW08b] I. P. Kaminov, T. Li, and A. E. Willner, editors. *Optical fiber telecommunications V B: Systems and networks*. Academic Press, San Diego, 2008.

[KP93] F. R. Kschischang and S. Pasupathy. Optimal nonuniform signaling for Gaussian channels. *IEEE Transactions on Information Theory*, 39(3):913–929, May 1993.

[Kra09] G. Kramer. Fundamentals of coding and modulation. In *Proceedings of the European Conference on Optical Communications (ECOC)*, page 2.6.1, Vienna, Austria, September 2009.

[KS06] M. Karlsson and H. Sunnerud. Effects of nonlinearities on PMD-induced system impairments. *IEEE/OSA Journal of Lightwave Technology*, 24(11):4127–4137, November 2006.

[KSS04] M. Katz and S. Shamai (Shitz). On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels. *IEEE Transactions on Information Theory*, 50(10):2257–2270, October 2004.

[KSV06] N. Kashyap, P. H. Siegel, and A. Vardy. Coding for the optical channel: The ghost-pulse constraint. *IEEE Transactions on Information Theory*, 52(1):64–77, January 2006.

[Kus06] M. Kuschnerov. The effect of polarization-mode dispersion on the channel capacity of single-mode optical fibers. Master's thesis, Technische Universität München, Munich, Germany, November 2006.

[KWM+05] R.I. Killey, P.M. Watts, V. Mikhailov, M. Glick, and P. Bayvel. Electronic dispersion compensation by signal predistortion using digital processing and a dual-drive Mach-Zehnder modulatorelectronic precompensation of intrachannel. *IEEE Photonics Technology Letters*, 17(3):714–716, March 2005.

[LA06] Q. Lin and G. P. Agrawal. Raman response function for silica fibers. *Optics Letters*, 31(21):3086–3088, November 2006.

[LAR07] J. Leibrich, A. Ali, and W. Rosenkranz. Optical OFDM as a promising technique for bandwidth-efficient high-speed data transmission over optical fiber. In *Proceedings of the 12th International OFDM Workshop*, pages 31–35, Hamburg, Germany, August 2007.

[Lei07] J. Leibrich. *Modeling and simulation of limiting impairments on next generation's transparent optical WDM transmission systems with advanced modulation schemes*. PhD thesis, Christian-Albrechts-Universität zu Kiel, January 2007.

[Lou06] H. Louchet. *Top-down analysis of high-capacity fiber-optic transmission*. PhD thesis, Technische Universität Berlin, May 2006.

[Low07] A. Lowery. Fiber nonlinearity mitigation in optical links that use OFDM for dispersion compensation. *IEEE Photonics Technology Letters*, 19(19):1556–1558, October 2007.

[LS94] J.-S. Lee and C.-S. Shim. Bit-error-rate analysis of optically preamplified receivers using an eigenfunction expansion method in optical frequency domain. *IEEE/OSA Journal of Lightwave Technology*, 12(7):1224–1229, July 1994.

[LWP07] A. J. Lowery, S. Wang, and M. Premaratne. Calculation of power limit due to fiber nonlinearity in optical OFDM systems. *Optics Express*, 15(20):13282–13287, October 2007.

[LWR02] J. Leibrich, C. Wree, and W. Rosenkranz. CF-RZ-DPSK for suppression of XPM on dispersion-managed long-haul optical WDM transmission on standard single-mode fiber. *IEEE Photonics Technology Letters*, 14(2):155–157, February 2002.

[Mar72] K. V. Mardia. *Statistics of directional data*. Academic Press, New York, 1972.

[Mar78] D. Marcuse. Gaussian approximation of the fundamental modes of graded-index fibers. *Journal of the Optical Society of America*, 68(1):103–109, January 1978.

[MBG+96] D. Mahgerefteh, D. L. Butler, J. Goldhar, B. Rosenberg, and G. L. Burdge. Technique for measurement of the Raman gain coefficient in optical fibers. *Optics Letters*, 21(24):2026–2028, December 1996.

[MDB+07] L. L. Minkov, I. B. Djordjevic, H. G. Batshon, L. Xu, T. Wang, M. Cvijetic, and F. Kueppers. Demonstration of PMD compensation by LDPC-coded turbo equalization and channel capacity loss characterization due to PMD and quantization. *IEEE Photonics Technology Letters*, 19(22):1852–1854, November 2007.

[Men99] C. R. Menyuk. Application of multiple-length-scale methods to the study of optical fiber transmission. *Journal of Engineering Mathematics*, 36(1-2):113–136, August 1999.

[Men06] C. R. Menyuk. Statistical errors in biasing Monte Carlo simulations with applications to polarization-mode dispersion compensators. *IEEE/OSA Journal of Lightwave Technology*, 24:4184–4196, November 2006.

[MG09] E. F. Mateo and G.Li. Compensation of interchannel nonlinearities using enhanced coupled equations for digital backward propagation. *Applied Optics*, 48(25):F6–F10, September 2009.

[Mid60] D. Middleton. *An introduction to statistical communication theory.* McGraw-Hill, New York, 1960.

[MK08] Wei Mao and J. M. Kahn. Lattice codes for amplified direct-detection optical systems. *IEEE Transactions on Communications*, 56(7):1137–1145, July 2008.

[MM06] C. R. Menyuk and B. S. Marks. Interaction of polarization mode dispersion and nonlinearity in optical fiber transmission systems. *IEEE/OSA Journal of Lightwave Technology*, 24(7):2806–2826, July 2006.

[MMW97] D. Marcuse, C. R. Menyuk, and P. K. A. Wai. Application of the manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence. *IEEE/OSA Journal of Lightwave Technology*, 15(9):1735–1746, September 1997.

[MP92] A. M. Mathai and S. B. Provost. *Quadratic forms in random variables.* Marcel Dekker, New York, 1992.

[MS01a] A. Mecozzi and M. Shtaif. On the capacity of intensity modulated systems using optical amplifiers. *IEEE Photonics Technology Letters*, 13(9):1029–1031, September 2001.

[MS01b] P. P. Mitra and J. B. Stark. Nonlinear limits to the information capacity of optical fibre communications. *Nature*, 411:1027–1030, June 2001.

[NM02] E. E. Narimanov and P. Mitra. The channel capacity of a fiber optics communication system: Perturbation theory. *IEEE/OSA Journal of Lightwave Technology*, 20(3):530–537, March 2002.

[NMS09] A. Nafta, E. Meron, and M. Shtaif. Capacity limitations in fiber-optic communication systems as a result of polarization-dependent loss. *Optics Letters*, 34(23):3613–3615, December 2009.

[NTR97] M. Nikles, L. Thevenaz, and P.A. Robert. Brillouin gain spectrum characterization in single-mode optical fibers. *IEEE/OSA Journal of Lightwave Technology*, 15(10):1842–1851, October 1997.

[OL02] J.-R. Ohm and H. D. Lüke. *Signalübertragung*. Springer, Berlin, 8th edition, 2002.

[PBP97] K.V. Peddanarappagari and M. Brandt-Pearce. Volterra series transfer function of single-mode fibers. *IEEE/OSA Journal of Lightwave Technology*, 15(12):2232–2241, December 1997.

[Pin63] M. S. Pinsker. *Information und Informationsstabilität zufälliger Größen und Prozesse*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1963.

[PK06] V. Pechenkin and F. R. Kschischang. Constrained coding for quasi-linear optical data transmission systems. *IEEE/OSA Journal of Lightwave Technology*, 24(12):4895–4902, December 2006.

[Pro95] J. G. Proakis. *Digital communications*. McGraw-Hill, New York, 3rd edition, 1995.

[PSS01] H. D. Pfister, J. B. Soriaga, and P. H. Siegel. On the achievable information rates of finite state isi channels. In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, volume 5, pages 2992–2996, San Antonio, TX, USA, November 2001.

[PSSG00] M. Peleg, S. Shamai (Shitz), and S. Galán. Iterative decoding for coded noncoherent MPSK communications over phase-noisy AWGN channel. *IEE Proceedings-Communications*, 147(2):87–95, April 2000.

[PW86] C. D. Poole and R. E. Wagner. Phenomenological approach to polarisation dispersion in long single-mode fibres. *Electronics Letters*, 22(19):1029–1030, September 1986.

[PW02] V. E. Perlin and H. G. Winful. Optimizing the noise performance of broadband WDM systems with distributed Raman amplification. *IEEE Photonics Technology Letters*, 14(8):1199–1201, August 2002.

[PWS06] M. R. Phillips, S. L. Woodward, and R. L. Smith. Cross-polarization modulation: Theory and measurement in subcarrier-modulated WDM systems.

IEEE/OSA Journal of Lightwave Technology, 24(11):4089–4099, November 2006.

[RLS+06] K. Roberts, C. Li, L. Strawczynski, M. O'Sullivan, and I. Hardcastle. Electronic precompensation of optical nonlinearity. *IEEE Photonics Technology Letters*, 18(2):403–405, January 2006.

[Sal85] J. Salz. Coherent lightwave communication. *AT&T Technical Journal*, 64(10):2153–2209, December 1985.

[SBW87] N. Shibata, R. P. Braun, and R. G. Waarts. Phase-mismatch dependence of efficiency of wave generation through four-wave mixing in a single-mode optical fiber. *IEEE Journal of Quantum Electronics*, 23(7):1205–1210, July 1987.

[Sch91] D. G. Schadt. Effect of amplifier spacing on four-wave mixing in multichannel coherent communications. *Electronics Letters*, 27(20):1805–1807, September 1991.

[Sch04] T. Schneider. *Nonlinear optics in telecommunications*. Springer, Berlin, 2004.

[Sei08] M. Seimetz. Laser linewidth limitations for optical systems with high-order modulation employing feed forward digital carrier phase estimation. In *Proceedings of the Optical Fiber Communication Conference (OFC)*, page OTuM2, San Diego, CA, USA, February 2008.

[SGTH89] R. H. Stolen, J. P. Gordon, W. J. Tomlinson, and H. A. Haus. Raman response function of silica-core fibers. *Journal of the Optical Society of America B*, 6(6):1159–1166, June 1989.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948.

[Sha49] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, January 1949.

[Sha56] C. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, 2(3):8–19, September 1956.

[SHZM03] O. V. Sinkin, R. Holzlöhner, J. Zweck, and C. R. Menyuk. Optimization of the split-step Fourier method in modeling optical-fiber communications systems. *IEEE/OSA Journal of Lightwave Technology*, 21(1):61–68, January 2003.

[SI73] R. H. Stolen and E. P. Ippen. Raman gain in glass optical waveguides. *Applied Physics Letters*, 22(6):276–278, March 1973.

[SLJ84] R. H. Stolen, C. Lee, and R. K. Jain. Development of the stimulated Raman spectrum in single-mode silica fibers. *Journal of the Optical Society of America B*, 1(4):652–657, August 1984.

[Smi72] R. G. Smith. Optical power handling capacity of low loss optical fibers as determined by stimulated Raman and Brillouin scattering. *Applied Optics*, 11(11):2489–2494, November 1972.

[SMN00] M. Shtaif, A. Mecozzi, and J. A. Nagel. Mean-square magnitude of all orders of polarization mode dispersion and the relation with the bandwidth of the principal states. *IEEE Photonics Technology Letters*, 12(1):53–55, January 2000.

[SMS01] J. B. Stark, P. Mitra, and A. Sengupta. Information capacity of nonlinear wavelength division multiplexing fiber optic transmission line. *Optical Fiber Technology*, 7(4):275–288, October 2001.

[SS89] S. Shamai (Shitz). On information rates of bandwidth restricted noisy phase channels. *Archiv für Elektronik und Übertragungstechnik (Electronics and Communication)*, 43(6):350–360, November–December 1989.

[STT07] A. Shafarenko, K. S. Turitsyn, and S. K. Turitsyn. Information-theory analysis of skewed coding for suppression of pattern-dependent errors in digital communications. *IEEE Transactions on Communications*, 55(2):237–241, February 2007.

[Tan01a] J. Tang. The multispan effects of Kerr nonlinearity and amplifier noises on Shannon channel capacity of a dispersion-free nonlinear optical fiber. *IEEE/OSA Journal of Lightwave Technology*, 19(8):1110–1115, August 2001.

[Tan01b] J. Tang. The Shannon channel capacity of dispersion-free nonlinear optical fiber transmission. *IEEE/OSA Journal of Lightwave Technology*, 19(8):1104–1109, August 2001.

[Tan02] J. Tang. The channel capacity of a multispan DWDM system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver. *IEEE/OSA Journal of Lightwave Technology*, 20(7):1095–1101, July 2002.

[Tan06] J. Tang. A comparison study of the Shannon channel capacity of various nonlinear optical fibers. *IEEE/OSA Journal of Lightwave Technology*, 24(5):2070–2075, May 2006.

[TAP09] Z. Taghavi, N. Alic, and G. Papen. Maximum-likelihood detection and constrained coding on optical channels. *IEEE/OSA Journal of Lightwave Technology*, 27(11):1469–1479, June 2009.

[TDYT03] K. S. Turitsyn, S. A. Derevyanko, I.V. Yurkevich, and S. K. Turitsyn. Information capacity of optical fiber channels with zero average dispersion. *Physical Review Letters*, 91(20):203901–1–4, November 2003.

[Tel10] Telegeography Research. *Global Bandwidth Research Service*, 2010.

[TPS06] M. H. Taghavi, G. C. Papen, and P. H. Siegel. On the multiuser capacity of WDM in a nonlinear optical fiber: Coherent communication. *IEEE Transactions on Information Theory*, 52(11):5008–5022, November 2006.

[Unb98] R. Unbehauen. *Systemtheorie 2*. Oldenbourg, Munich, 7th edition, 1998.

[vBv+08] J. Řeháček, Z. Bouchal, R. Čelechovský, Z. Hradil, and L. L. Sánchez-Soto. Experimental test of uncertainty relations for quantum mechanics on a circle. *Physical Review A*, 77(3):032110, March 2008.

[Vit63] A. J. Viterbi. Phase-locked loop dynamics in the presence of noise by Fokker-Planck techniques. *Proceedings of the IEEE*, 51(12):1737–1753, December 1963.

[Vit65] A. Viterbi. Optimum detection and signal selection for partially coherent binary communication. *IEEE Transactions on Information Theory*, 11(2):239–246, April 1965.

[VP02] E. Voges and K. Petermann, editors. *Optische Kommunikationstechnik*. Springer, Berlin, 2002.

[VRD+04] B. Vasic, V. S. Rao, I. B. Djordjevic, R. K. Kostuk, and I. Gabitov. Ghost-pulse reduction in 40-Gb/s systems using line coding. *IEEE Photonics Technology Letters*, 16(7):1784–1786, July 2004.

[WBSP09] M. Winter, C.-A. Bunge, D. Setti, and K. Petermann. A statistical treatment of cross-polarization modulation in DWDM systems. *IEEE/OSA Journal of Lightwave Technology*, 27(17):3739–3751, September 2009.

[WE06] P. J. Winzer and R.-J. Essiambre. Advanced optical modulation formats. *Proceedings of the IEEE*, 94(5):952–985, May 2006.

[WFBP06] C. Weber, J.K. Fischer, C.-A. Bunge, and K. Petermann. Electronic precompensation of intrachannel nonlinearities at 40 Gb/s. *IEEE Photonics Technology Letters*, 18(16):1759–1761, August 2006.

[Wit05] M. Witzke. *Iterative Detektion von MIMO-Signalen mit linearen und erweitert-linearen Filtern*. PhD thesis, Technische Universität München, 2005.

[WK82] G. E. Walrafen and P. N. Krishnan. Model analysis of the Raman spectrum from fused silica optical fibers. *Applied Optics*, 21(3):359–360, February 1982.

[WMC91] P. K. A. Wai, C. R. Menyuk, and H. H. Chen. Stability of solitons in randomly varying birefringent fibers. *Optics Letters*, 16(16):1231–1233, August 1991.

[WN05] B. Wu and E. E. Narimanov. Information capacity of nonlinear fiber-optical systems. In *Proceedings of the Optical Fiber Communication Conference (OFC)*, page OME58, Anaheim, CA, USA, March 2005.

[Wol67] J. Wolfowitz. Memory increases capacity. *Information and Control*, 11:423–428, 1967.

[WPG+04] L. G. L. Wegener, M. L. Povinelli, A. G. Green, P. P. Mitra, J. B. Stark, and P. B. Littlewood. The effect of propagation nonlinearities on the information capacity of WDM optical fiber systems: cross-phase modulation and four-wave mixing. *Physica D*, 189(1–2):81–89, February 2004.

[WSP10] M. Winter, D. Setti, and K. Petermann. Cross-polarization modulation in polarization-division multiplex transmission. *IEEE Photonics Technology Letters*, 22(8):538–540, April 2010.

[XBP03] B. Xu and M. Brandt-Pearce. Comparison of FWM- and XPM-induced crosstalk using the Volterra series transfer function method. *IEEE/OSA Journal of Lightwave Technology*, 21(1):40–53, January 2003.

[XP03] W. Xiang and S. S. Pietrobon. On the capacity and normalization of ISI channels. *IEEE Transactions on Information Theory*, 49(9):2263–2268, September 2003.

[XWP+08] T. J. Xia, G. Wellbrock, D. Peterson, W. Lee, M. Pollock, B. Basch, D. Chen, M. Freiberger, M. Alfiad, H. de Waardt, M. Kuschnerov, B. Lankl, T. Wuth, E. D. Schmidt, B. Spinnler, C. J. Weiske, E. de Man, C. Xie, D. van den Borne, M. Finkenzeller, S. Spaelter, R. Derksen, M. Rehman, J. Behel, J. Stachowiak, and M. Chbat. Multi-rate transmission at 50-GHz channel spacing over 1040-km field-deployed fiber. In *Proceedings of the European Coference on Optical Communications (ECOC)*, page Th.2.E.2, Brussels, Belgium, September 2008.

[YDT02] A. Yeniay, J.-M. Delavaux, and J. Toulouse. Spontaneous and stimulated brillouin scattering gain spectra in optical fibers. *IEEE/OSA Journal of Lightwave Technology*, 20(8):1425–1432, August 2002.

[YL09] F. Yaman and G. Li. Nonlinear impairment compensation for polarization-division multiplexed WDM transmission using digital backward propagation. *IEEE Photonics Journal*, 1(2):144–152, August 2009.