THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

On Signal Constellations and Coding for Long-Haul Fiber-Optical Systems

Christian Häger

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Signals and Systems
Chalmers University of Technology
Göteborg, Sweden, 2014

On Signal Constellations and Coding for Long-Haul Fiber-Optical Systems

Christian Häger

Technical Report No. R008/2014
ISSN 1403-266X

This thesis has been prepared using LaTeX and PSTricks.

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

Printed by Chalmers Reproservice
Göteborg, Sweden, April 2014

Abstract

Motivated by the realization that even the enormous bandwidth available in an optical fiber is finite and valuable, the design of *spectrally efficient* long-haul fiber-optical communication systems has become an important research topic. Compared to other wireline technologies, e.g., transmission over coaxial cables, the main challenge comes from the inherent nonlinearity of the underlying communication channel caused by the relatively high signal intensities. In this thesis, we study the design of spectrally efficient fiber-optical systems for both uncoded and coded transmission scenarios.

We consider the problem of designing higher-order signal constellations for a system that is severly impaired by nonlinear phase noise. By optimizing amplitude phase-shift keying constellations, which can be seen as the union of phase-shift keying constellation with different amplitude levels, gains of up to 3.2 dB at a symbol error probability of 10^{-2} are shown to be achievable compared to conventional constellations. We also illustrate a somewhat counterintuitive behavior of optimized constellations for very high input powers and nonlinear distortions. In particular, sacrificing a constellation point or ring may be beneficial in terms of the overall performance of the constellation.

Furthermore, we study polarization-multiplexed transmission, where spectral efficiency is increased by encoding data onto both polarizations of the light. For a memoryless fiber-optical channel, we introduce a low-complexity detector which is based on an amplitude-dependent phase rotation and subsequent threshold detection. The complexity compared to the four-dimensional maximum likelihood detector is considerably reduced, albeit at the expense of some performance loss.

Lastly, we consider the design of a coded fiber-optical system operating at high spectral efficiency. In particular, we study the optimization of the mapping of the coded bits to the modulation bits for a polarization-multiplexed fiber-optical system that is based on the bit-interleaved coded modulation paradigm. This technique, which we refer to as bit mapper optimization, is extended to the class of spatially coupled low-density parity-check codes, which have shown outstanding performance over memoryless binary-input channels. For a transmission scenario without optical inline dispersion compensation, the results show that the transmission reach can be extended by roughly up to 8%, without significantly increasing the system complexity.

Keywords: Bit-interleaved coded modulation, bit mapper, constellation optimization, detector, fiber-optical communication, low-density parity-check codes, spatial coupling.

List of Publications

This thesis is based on the following publications:

Paper A
Christian Häger, Alexandre Graell i Amat, Alex Alvarado, Erik Agrell, "Design of APSK Constellations for Coherent Optical Channels with Nonlinear Phase Noise", *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3362–3373, Aug. 2013

Paper B
Christian Häger, Lotfollah Beygi, Erik Agrell, Pontus Johannisson, Magnus Karlsson, Alexandre Graell i Amat, "A Low-Complexity Detector for Memoryless Polarization-Multiplexed Fiber-Optical Channels", *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 368–371, Jan. 2014

Paper C
Christian Häger, Alexandre Graell i Amat, Fredrik Brännström, Alex Alvarado, Erik Agrell "Improving Soft FEC Performance for Higher-Order Modulations by Bit Mapper Optimization", submitted to *Optics Express*, Apr. 2014

Other publications by the author not included in this thesis:

- Christian Häger, Alexandre Graell i Amat, Alex Alvarado, Erik Agrell, "Constellation Optimization for Coherent Optical Channels Distorted by Nonlinear Phase Noise", in Proc. *IEEE Global Communications Conf. (GLOBECOM)*, Anaheim, CA, Dec. 2012
- Christian Häger, Alexandre Graell i Amat, Alex Alvarado, Fredrik Brännström, Erik Agrell, "Optimized Bit Mappings for Spatially Coupled LDPC Codes over Parallel Binary Erasure Channels", in Proc. *IEEE Int. Conf. Communications (ICC)*, Sydney, Australia, June 2014

Acknowledgements

First of all, I would like to thank my supervisor Alexandre Graell i Amat and my co-supervisors Alex Alvarado, Fredrik Brännström, and Erik Agrell for all the fruitful discussions and the invaluable feedback that I have received throughout the past two and a half years. In particular, I would like to thank Alexandre for his constant support and guidance. I would also like to thank Alex again for being an outstanding host during my research visit in Cambridge.

I would like extend my thanks to Iryna Andriyanova for sharing her insight on spatially coupled LDPC codes during my stay in Paris. Also, one of the persons who taught me a lot during these "early" PhD years is Lotfollah Beygi. I am immensely grateful for all the time that he spent with me discussing various aspects of fiber-optical communication. Paper B in this thesis is a collaborative effort which was initiated by Lotfollah. I would also like to thank Mikhail Ivanov for proofreading parts of the thesis.

Many thanks also to the administrative staff of the department, in particular Lars Börjesson, Agneta Kinnander, Madeleine Persson, and Natasha Adler. Last but not least, I would like to thank everyone in our group, and also everyone involved in FORCE, the fiber-optic research center at Chalmers, for creating a fantastic and stimulating work environment.

Regarding the financial support, I would like to thank and acknowledge the Ericsson Research Foundation for their generous financial support of my research visit to Cambridge.

Christian Häger, Göteborg, March 2014

Acronyms

Abbreviation	Meaning
APSK	amplitude phase-shift keying
ASE	amplified spontaneous emission
AWGN	additive white Gaussian noise
BEC	binary erasure channel
BEP	bit error probability
BICM	bit-interleaved coded modulation
BP	belief propagation
BRGC	binary reflected Gray code
CM	coded modulation
CN	check node
DBP	digital backpropagation
DCF	dispersion compensating fiber
DE	density evolution
DM	dispersion-managed
EDFA	erbium-doped fiber amplifier
EXIT	extrinsic information transfer
FEC	forward error correction
FFT	fast Fourier transform
GMI	generalized mutual information
IID	independent and identically distributed
LB	lower bound
LDPC	low-density parity-check
LLR	log-likelihood ratio
MAP	maximum a posteriori
MI	mutual information
ML	maximum likelihood
MLCM	multilevel coded modulation
MLSE	maximum likelihood sequence estimation
NLPN	nonlinear phase noise
NLSE	nonlinear Schrödinger equation
OOK	on-off keying
PDF	probability density function
PM	polarization-multiplexed
PMD	polarization mode dispersion
PSD	power spectral density
PSK	phase-shift keying

Abbreviation	Meaning
QAM	quadrature amplitude modulation
QPSK	quadrature phase-shift keying
ROADM	reconfigurable optical add-drop multiplexer
RPS	Raman pump station
SC-LDPC	spatially coupled low-density parity-check
SEP	symbol error probability
SMF	single-mode fiber
sNLSE	stochastic nonlinear Schrödinger equation
SNR	signal-to-noise ratio
SP	single-polarization
SSFM	split-step Fourier method
TCM	trellis coded modulation
TS	two-stage
VN	variable node
WD	windowed decoder
WDM	wavelength-division multiplexing

Contents

Part I

Introduction

CHAPTER 1

Background

When requesting a website, most internet users are probably unaware that the digital data is modulated onto a light source and transmitted over thousands of kilometers in an optical waveguide, a so-called optical fiber, at some point on the way from the remote server to their home computer or mobile device. In fact, more than 99% of the global intercontinental traffic is carried over optical fiber and such "long-haul" fiber-optical communication systems are the key enabler of high-speed internet data transfer connecting cities, countries, and continents [1].

Optical fiber as a transmission medium is very well suited for sending large amounts of data over long distances. Engineers and physicists have spent a great effort to refine the fiber material to be very transparent for optical light over a large frequency range and to improve the components in an optical transmission system, e.g., lasers and amplifiers. Meanwhile, however, fiber-optical systems traditionally employ digital modulation techniques that are rather wasteful with the available frequency spectrum. As an example, switching the light source on and off according to the digital data stream, referred to as on-off keying (OOK), is highly inefficient from a spectral viewpoint.

To keep up with the increasing data rate demands of current applications, and to enable innovative broadband technologies in the future, it becomes more and more apparent that next generation fiber-optical systems need to use the available spectrum more efficiently. The bandwidth of optical fibers is now considered a limited resource and due to this realization, there is currently a great deal of interest in determining the ultimate limits of optical systems in terms of spectral efficiency [2–4] and developing practical schemes that can achieve these limits [5–7]. The distinctive feature, and at the same time the main challenge in fiber-optical communication, is that the underlying communication

channel is fundamentally nonlinear. The fiber nonlinearity is considered as one of the limiting factors for increasing the data rates in long-haul systems [8–10]. Some important problems that arise in the design of spectrally efficient fiber-optical systems are addressed in this thesis.

To improve the spectral efficiency over OOK, the data can be encoded into multiple amplitude and/or phase levels of the optical carrier. This leads to the problem of designing good signal constellations that are robust to the severe nonlinear distortions. Signal constellation design can be thought of as optimizing the placement of a given number of points in two or more dimensions under some constraints. This problem can be considered as a classical problem in communication theory [11, Ch. 1] and it is revisited in Paper A, paying special attention to fiber nonlinearities.

A further increase in spectral efficiency can be made by properly utilizing both polarizations of the optical light, which we refer to as polarization-multiplexed (PM) transmission. Compared to single-polarization (SP) transmission, signals are now represented as points in a four- rather than a two-dimensional space. Optimal maximum likelihood (ML) detection in four dimensions can be computationally very complex, particularly for constellations with many points. Therefore, in Paper B, we study a detector design for a recently developed four-dimensional, memoryless fiber-optical channel model with the intention to significantly reduce the detection complexity.

The problems described above are related to uncoded transmission schemes and do not consider forward error correction (FEC). However, FEC needs to be considered in order to operate close to the ultimate transmission limits of optical fibers. We therefore also consider the problem of designing coded transmission systems at high spectral efficiencies in Paper C. Here, we restrict ourselves to systems that are based on the bit-interleaved coded modulation (BICM) paradigm. BICM can be seen as a pragmatic way to combine signal constellations consisting of many points with powerful binary error correction codes. In particular, we study how the coded bits should be allocated to the modulation bits. We refer to this problem as bit mapper optimization. As one particular example, we consider spatially coupled low-density parity-check (SC-LDPC) codes which have recently been shown to achieve outstanding performance over a variety of communication channels [12–14].

1.1 Thesis Organization

The licentiate degree is an intermediate step for a doctoral student towards the final PhD degree and the licentiate thesis documents the progress that has been made over a period of roughly two to two and a half years. The format of this thesis is a so-called collection of papers. It is divided into two parts, where the first part serves as an introduction to the appended papers in the second part.

The remainder of the first part of this thesis is structured as follows. In Chapter 2, we provide an introduction to fiber-optical channel modeling and describe the origin of

the channel models that are used in the appended papers. Our main goal is to show that all utilized models have the same mathematical foundation, namely the nonlinear Schrödinger equation (NLSE), albeit assuming somewhat different material parameters and system configurations. In Chapter 3, we explain the necessary theoretical background for the optimization problem that we consider in Paper C. For Paper C, the reader is assumed to be somewhat familiar with coding theory and in particular low-density parity-check (LDPC) codes, which we discuss in Chapter 3. Finally, some conclusions are summarized in Chapter 4, where we also briefly discuss future work.

1.2 Notation

Throughout this thesis, vectors are denoted by boldface letters $\mathbf{A}$, matrices by blackboard letters $\mathbb{A}$, sets by calligraphic letters $\mathcal{A}$ (except the sets containing the real numbers $\mathbb{R}$, complex numbers $\mathbb{C}$, integers $\mathbb{Z}$, and natural numbers $\mathbb{N}$), and random variables by capital letters A. The probability density function (PDF) of a continuous random variable Y is denoted by $f_Y(\,\cdot\,)$ and $\mathbb{E}[\,\cdot\,]$ denotes expectation. A matrix transpose is denoted by $(\,\cdot\,)^\intercal$. $\delta(t)$ denotes Dirac's delta function while $\delta[k]$ denotes the Kronecker delta. Convolution is denoted by $\circledast$. For a continuous-time signal $x(t)$, we write $x(t) \circ\!\!-\!\!\bullet X(f)$ to indicate its Fourier transform. The imaginary unit is denoted by $\jmath \triangleq \sqrt{-1}$.

We also acknowledge the following notational inconsistencies. In Paper A, x and y are used for the channel input and output, respectively, whereas in Paper B and C, x and y are used to differentiate between polarizations. In Paper A, conditioning on particular realization of a random variable is denoted by $f_{Y|X=x}(y)$, whereas in the introductory part of the thesis we use $f_{Y|X}(y|x)$ for readability purposes. In Papers A and B, the spontaneous emission factor n_{sp} appeared in the context of distributed amplification, which should be replaced with the photon occupancy factor K_T for Raman amplification [2]. Further, the additive noise power spectral density N_0 in Paper A is defined per unit length and this definition is inconsistent with the definition used in Part I of the thesis where it has units of [W/Hz]. We use a different font N_0 in Part I to indicate the difference.

CHAPTER 2

Fiber-Optical Channel Modeling

A channel model is a mathematical description of the propagation medium and possibly also includes certain elements of the transmitter and receiver (e.g., filters). In the appended papers, the starting point for the analysis is a discrete memoryless channel, which can be specified in the form of a conditional PDF. This chapter is intended to describe the origin of the assumed PDFs and also to give the reader a broader picture about optical channel modeling in general.

We are concerned with coherent, long-haul (i.e., distances exceeding 2000 km) data transmission over single-mode fibers (SMFs) and the main challenge is a nonlinear effect caused by the relatively high signal power in relation to the small cross-section area of the fiber. Without going further into the physical details, a useful way to think about this effect is to imagine that the presence of an optical signal can compress the fiber material (in most cases silica) to such a degree that its propagation properties, in particular the refractive index, are changed in a nonlinear way [9, p. 18].

If nonlinear effects are ignored, an optical fiber can be regarded as a linear dispersive channel. Motivated by this, we start by reviewing some important results for this channel in Section 2.1. In Section 2.2, we discuss the NLSE, which is a deterministic channel model for an SMF. Multi-span links consisting of several SMFs and different amplification types are described in Section 2.3. In Section 2.4, the channel models assumed in the papers are introduced and compared.

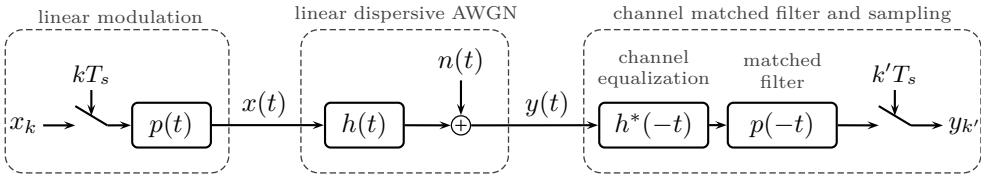


Figure 2.1: Block diagram of a system with linear modulation, a linear dispersive AWGN channel, and a receiver that obtains a sufficient statistic for optimal detection.

2.1 Linear Dispersive Channels

Consider the complex-valued linear dispersive additive white Gaussian noise (AWGN) channel

$$y(t) = h(t) \circledast x(t) + n(t), \tag{2.1}$$

where $x(t)$ is the baseband representation of the input signal, $y(t)$ is the output signal, $h(t)$ is the channel impulse response, and $n(t)$ is a circularly symmetric complex Gaussian stochastic process with zero mean with power spectral density (PSD) N_0, i.e., $\mathbb{E}[N(t)N^*(t')] = \mathsf{N}_0\delta(t-t')$. Given an appropriate impulse response $h(t)$, equation (2.1) is widely used as a model for, e.g., wireline transmission over coaxial cables. A vast amount of literature exists on the linear dispersive AWGN channel and many important problems such as the optimal receiver/detector structure, the ultimate achievable transmission rates, or practical schemes that achieve these rates can be considered well studied and understood by now, see [15] and references therein.

In the following, we review some important concepts and results under the assumption that $h(t)$ is a unit gain all-pass filter, i.e., $|H(f)| = 1$, where $h(t) \circ\!\!-\!\!\bullet H(f)$. This assumption will turn out to be accurate for the fiber-optical channel later if nonlinear effects are ignored. Henceforth, the receiver is always assumed to have perfect knowledge about $h(t)$, obtained through an appropriate channel estimation technique. Furthermore, it is assumed that perfect carrier and timing synchronization between transmitter and receiver has been achieved.

In Fig. 2.1, a block diagram of the considered system is shown. We start with a linearly pulse-modulated[1] input signal

$$x(t) = \sum_k x_k p(t - kT_s), \tag{2.2}$$

where $x_k \in \mathbb{C}$ are the information symbols for $k \in \mathbb{Z}$, $p(t)$ is the real-valued pulse shape, and T_s is the symbol period. The symbol rate is defined as $R_s = 1/T_s$. The input signal

[1] A different approach to communicate over linear dispersive channels is through multicarrier transmission, which we do not consider in this thesis.

$x(t)$ is assumed to be bandlimited to a bandwidth $W \geq 1/T_s$ and the power of $x(t)$ is

$$P \triangleq \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2 \, \mathrm{d}t. \tag{2.3}$$

The receiver observes the filtered and noisy version of $x(t)$ according to (2.1) as

$$y(t) = h(t) \circledast x(t) + n(t) = \sum_k x_k \tilde{p}(t - kT_s) + n(t), \tag{2.4}$$

where $\tilde{p}(t) = h(t) \circledast p(t)$ is the convolution of the pulse shape with the channel impulse response. A sufficient statistic[2] for detecting the symbols x_k based on $y(t)$ can be obtained by filtering $y(t)$ with the channel matched filter $\tilde{p}^*(-t)$ and sampling at time instances $t = k'T_s$, $k' \in \mathbb{Z}$ [16, Prop. 28.5.2]. Since $\tilde{p}^*(-t) = h^*(-t) \circledast p(-t)$, one may interpret the application of the channel matched filter as a two-step process. The first step (filtering with $h^*(-t)$) is referred to as "channel equalization" or simply "equalization". Due to the assumption that $h(t)$ is an all-pass filter, one may also think of applying $h^*(-t)$ as a compensation technique[3], in the sense that $h(t) \circledast h^*(-t) = \delta(t)$, which is readily seen by applying the Fourier transform and invoking the all-pass filter assumption, i.e., $H(f)H^*(f) = |H(f)|^2 = 1$. After applying the (pulse) matched filter $p(-t)$ and sampling, one finally obtains the discrete-time channel model [16, Prop. 28.5.2 (ii)]

$$y_{k'} = \sum_k x_k \mathsf{R}_p((k - k')T_s) + n_{k'}, \tag{2.5}$$

where $\mathsf{R}_p(t) = p(t) \circledast p(-t)$ is the self-similarity (or autocorrelation) function of the pulse shape [16, Def. 11.2.1] and $n_{k'}$ is a zero mean Gaussian random variable with $\mathbb{E}[N_k N_{k'}^*] = \mathsf{N}_0 \mathsf{R}_p((k - k')T_s)$. A discrete memoryless channel $y_k = x_k + n_k$ is obtained by choosing the pulse shape such that its self-similarity function $\mathsf{R}_p(t)$ satisfies Nyquist's criterion $\mathsf{R}_p((k-k')T_s) = \delta[k-k']$ [16, Def. 11.3.1], e.g., choosing $p(t)$ to be a root-raised cosine pulse with an arbitrary roll-off factor. In this case, the channel from x_k to y_k is completely characterized by the conditional PDF

$$f_{Y_k|X_k}(y_k|x_k) = \frac{1}{\pi \mathsf{N}_0} \exp\left(-\frac{|y_k - x_k|^2}{\mathsf{N}_0}\right). \tag{2.6}$$

On the other hand, if the Nyquist criterion is not fulfilled (or $h(t)$ is not an all-pass filter), the discrete-time channel model is not memoryless (one may write the first term on the right-hand side of (2.5) as a discrete convolution) and the noise samples are correlated. The optimal detection approach in that case is maximum likelihood sequence estimation (MLSE) either with [17] or without [18] the insertion of a noise whitening filter.

[2] See [16, Ch. 26] for a formal definition.

[3] The term "dispersion compensation" is often used instead of "equalization" in the context of fiber-optical communication systems.

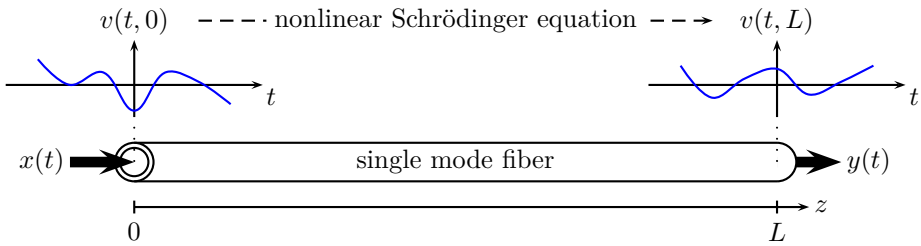


Figure 2.2: Conceptual representation of the signal evolution through a SMF. The NLSE describes the relationship between the input signal $x(t) = u(t, 0)$ and the output signal $y(t) = u(t, L)$.

The previous description is relevant to fiber-optical systems for several important reasons. The first one is that, while the actual fiber-optical channel model is nonlinear, the linear (pulse) modulation and receiver structure in Fig. 2.1 are nonetheless ubiquitously used in practical fiber-optical systems. Obviously, they are not necessarily optimal anymore[4], but their performance can still be analyzed and seen as a baseline. Secondly, the previous discussion illustrates how an originally continuous-time channel model (eq. (2.1)) can be simplified to a discrete-time model (eq. (2.5)) which in turn can then be used to study, e.g., detection algorithms or channel coding schemes. This is the approach taken in all the appended papers, where the analysis is based on discrete-time channel models.

2.2 The Nonlinear Schrödinger Equation

The starting point for coherent, long-haul fiber-optical channel modeling is the NLSE, which can be derived from the Maxwell equations under some assumptions that are appropriate for SMFs [20]. The NLSE is a partial differential equation that defines the input–output relationship for optical baseband signals[5] propagating through SMFs.

Let us first, in addition to the time parameter t, introduce a distance parameter $0 \leq z \leq L$ that denotes the propagation distance of the signal from the beginning of the fiber, where L is the total length of the fiber. The baseband signal of interest is then a function of two parameters, denoted by $v(t, z)$. To be consistent with the previous notation, we define the input and output signals as $x(t) = v(t, 0)$ and $y(t) = v(t, L)$, i.e., $x(t)$ is the signal launched into the fiber at $z = 0$, and $y(t)$ is the signal received after propagating through an SMF of length L. This is conceptually illustrated in Fig. 2.2. Before we continue, we also define the instantaneous signal power $P(t, z) \triangleq |v(t, z)|^2$ and the power profile $P(z) \triangleq \lim_{T\to\infty}(\int_{-T}^{T} P(t, z)\,\mathrm{d}t)/(2T)$, where $P(0) = P$ is the power of the input

[4] The author in [19, p. 42] goes so far to say that these methods "borrowed from linear system theories, are inappropriate for communication over optical fiber networks".

[5] Often called "slowly varying envelope" in the literature. The carrier frequency is assumed to be the equivalent of a 1550 nm light wave, corresponding to roughly 193.4 THz.

signal.

The NLSE accounts for signal attenuation, chromatic dispersion, and nonlinear effects in an SMF and can be written as

$$\frac{\partial v(t,z)}{\partial z} = -\frac{\alpha}{2}v(t,z) - \jmath\frac{\beta_2}{2}\frac{\partial^2 v(t,z)}{\partial t^2} + \jmath\gamma v(t,z)|v(t,z)|^2, \tag{2.7}$$

where α is the attenuation coefficient, β_2 is the chromatic dispersion coefficient, and γ is the nonlinear Kerr parameter. If we take into account only the first term on the right-hand side of (2.7), one obtains $v(t,z) = \exp(-\alpha z/2)v(t,0)$ as a solution[6], i.e., we immediately see that the signal amplitude in an SMF decays exponentially with the propagation distance. By defining a renormalized version of $v(t,z)$ as $u(t,z) \triangleq \exp(\alpha z/2)v(t,z)$ and substituting it into (2.7), one obtains an alternative and somewhat simpler version of the NLSE as [20, eq. (4)]

$$\frac{\partial u(t,z)}{\partial z} = -\jmath\frac{\beta_2}{2}\frac{\partial^2 u(t,z)}{\partial t^2} + \jmath\gamma e^{-\alpha z}u(t,z)|u(t,z)|^2. \tag{2.8}$$

Unfortunately, there are no closed-form solutions to the NLSE and one has to use numerical methods in order to solve (2.7) or (2.8).

We proceed by discussing two special cases of (2.8) in Sections 2.2.1 and 2.2.2, for $\gamma = 0$ and $\beta_2 = 0$, respectively. In both cases, a closed-form solution can be obtained. These solutions are also the key ingredients for one of the most widely used numerical methods to solve (2.8), namely, the split-step Fourier method (SSFM), which is described in Section 2.2.3.

2.2.1 Absence of Nonlinear Effects

As we will see, when nonlinear effects are ignored, the solution of (2.8) can be more conveniently written as the convolution of the input signal with a dispersive filter. For $\gamma = 0$, (2.8) becomes

$$\frac{\partial u(t,z)}{\partial z} = -\jmath\frac{\beta_2}{2}\frac{\partial^2 u(t,z)}{\partial t^2}, \tag{2.9}$$

which can be solved analytically by first transforming (2.9) into the Fourier domain[7] using the correspondence $\frac{\partial^n}{\partial t^n}x(t) \circ\!\!-\!\!\bullet (\jmath 2\pi f)^n X(f)$ to obtain [20]

$$\frac{\partial U(f,z)}{\partial z} = -\jmath\frac{\beta_2}{2}(\jmath 2\pi f)^2 U(f,z), \tag{2.10}$$

[6] Recall that the solution of $\partial f(z)/\partial z = cf(z)$ is given by $f(z) = \exp(cz)f(0)$.

[7] The solution of (2.9) is sometimes immediately written as $u(t,z) = \exp(z\hat{D})u(t,0)$ with $\hat{D} = -\jmath\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2}$. The operator $\exp(z\hat{D})$ may be interpreted with the help of the Taylor expansion of the exponential function around 0, i.e., using $e^x = 1 + x + x^2/2 + \cdots$ with $x = z\hat{D}$.

where $U(f,z) \bullet\!\!-\!\!\circ\, u(t,z)$. The solution of (2.10) can be verified to be

$$U(f,z) = \exp\left(\jmath 2\beta_2\pi^2 f^2 z\right) U(f,0) \tag{2.11}$$

$$= H(f,z)U(f,0), \tag{2.12}$$

where $H(f,z) = \exp(\jmath 2\beta_2\pi^2 f^2 z)$ can be seen as the frequency response of a (z-dependent) dispersive filter. Applying the inverse Fourier transform to (2.12) finally leads to the expression

$$u(t,z) = h(t,z) \circledast u(t,0), \tag{2.13}$$

where $h(t,z) = \exp\left(\jmath t^2/(2\beta_2 z)\right)/\sqrt{\jmath 2\pi\beta_2 z} \circ\!\!-\!\!\bullet\, H(f,z)$ is the impulse response of the filter. We further have $|H(f,z)| = 1$. In summary, in the absence of nonlinear effects, chromatic dispersion manifests itself as a unit-gain all-pass filter.

2.2.2 Absence of Dispersion

Another special case that allows for an exact and explicit solution of the NLSE is when chromatic dispersion is completely ignored. In this case, i.e., for $\beta_2 = 0$, (2.8) becomes

$$\frac{\partial u(t,z)}{\partial z} = \jmath\gamma e^{-\alpha z}|u(t,z)|^2 u(t,z). \tag{2.14}$$

One may then verify that the solution to (2.14) is given by [20, eq. (17)]

$$u(t,z) = u(t,0)e^{\jmath\gamma L_{\text{eff}}(z)|u(t,0)|^2}, \tag{2.15}$$

where

$$L_{\text{eff}}(z) \triangleq \int_0^z e^{-\alpha z'}\,\mathrm{d}z' = \frac{1-\exp(-\alpha z)}{\alpha} \tag{2.16}$$

is the effective nonlinear length, where $L_{\text{eff}}(z) \le z$ with $L_{\text{eff}}(z) \to z$ as $\alpha \to 0$.

From (2.15), we see that, for a given z, the nonlinear effect by itself causes a phase-shift of the signal that is proportional to the instantaneous power $|u(t,0)|^2$, whereas the amplitude $|u(t,z)| = |u(t,0)|$ remains unchanged. This effect is called self-phase modulation and an important consequence of its nonlinear nature is that the bandwidth of $u(t,z)$ may grow during propagation through the fiber.

2.2.3 Split-Step Fourier Method

As was mentioned earlier, the NLSE with $\gamma \neq 0$ and $\beta_2 \neq 0$ cannot be solved analytically and one has to resort to numerical methods in order to obtain the relationship between the input and output signals. A popular and computationally efficient numerical method is the SSFM which we describe in the following.[8]

[8] There exist several different versions of the SSFM and the one presented here is referred to as asymmetric and non-iterative. For more details we refer the reader to [9, Sec. 2.4.1].

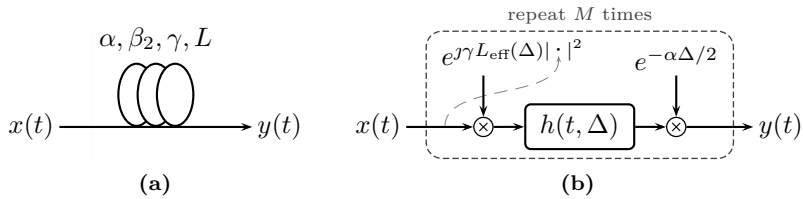


Figure 2.3: Symbolic representation of an SMF in (a) and an (approximate) mathematical model via the SSFM in (b). The notation $|\cdot|^2$ stands for the instantaneous power of the signal arriving at the corresponding multiplication block as indicated by the dashed, gray line.

Conceptually, we start by discretizing the spatial dimension and subdividing the entire fiber of length L into small segments of length Δ, where $M = L/\Delta \in \mathbb{N}$ is the total number of segments. For the ith segment, $1 \leq i \leq M$, the input signal is denoted by $u(t,(i-1)\Delta)$ and the corresponding output signal by $u(t,i\Delta)$. It is assumed that an approximate solution to obtain $u(t,i\Delta)$ based on $u(t,(i-1)\Delta)$ is given by first applying (2.15) and then (2.13), i.e., for small Δ, we assert that

$$u(t,i\Delta) \approx h(t,\Delta) \circledast \left(u(t,(i-1)\Delta) e^{\jmath\gamma L_{\text{eff}}(\Delta)|u(t,(i-1)\Delta)|^2} \right). \tag{2.17}$$

Then, an approximate solution for an entire SMF of length L is given by repeatedly applying (2.17), starting with the first segment $i = 1$, i.e., with the input signal $u(t,0) = x(t)$. The SSFM step in (2.17) is given in terms of the normalized signal $u(t,z)$ and to incorporate the signal attenuation, the output signal $u(t,i\Delta)$ is multiplied by $\exp(-\alpha\Delta/2)$ to obtain $v(t,i\Delta)$ after each step. The resulting numerical method is shown in terms of a block diagram in Fig. 2.3. In the figure, the notation $|\cdot|^2$ stands for the instantaneous power of the signal that arrives at the corresponding multiplication block (e.g., $|x(t)|^2$ in the first segment, $|u(t,\Delta)\exp(-\alpha\Delta/2)|^2$ in the second, and so on). It has been shown that the above method converges to the true solution for $\Delta \to 0$ [9, p. 42]. Practical guidelines on the choice of the segment size are developed in [21].

The name of the method originates from the fact that the nonlinear phase-shift operation and the linear filtering in Fig. 2.3(b) are commonly carried out in the time and frequency domain, respectively. Therefore, one forward and one inverse Fourier transform have to be performed per segment. In computer implementations, a sampled version of $x(t)$ is considered which facilitates the application of the computationally efficient fast Fourier transform (FFT). Such an implementation is for example provided in [9, App. B].

2.3 Optical Amplification and Noise

The numerical value of the attenuation coefficient α is typically between 0.2 and 0.4 dB/km. Assuming $\alpha = 0.2$ dB/km and a transmission distance of $L = 2000$ km, the input signal would be attenuated by 400 dB implying that $y(t)$ is practically zero [2, Sec. IX-B]. It is therefore necessary to amplify the signal along the transmission path, which invariably introduces noise into the system.

In this section, we briefly discuss two types of amplification, lumped and distributed, in terms of their effect on the power profile of the signal and the type of noise that they introduce. Modeling the power profile is important due to the dependency of the nonlinear effect on the instantaneous signal power. Thus, one cannot simply ignore attenuation effects and make a link budget analysis as is common for linear channels. Details about the underlying physical aspects of optical amplification can be found in standard textbooks on optical data transmission, e.g., [22, Ch. 6]. It should, however, be pointed out that the optical amplifier noise is in fact the dominant source of noise in long-haul systems meaning that noise from other sources, for example thermal noise from electrical components, is negligible in comparison and can therefore be ignored [2, Sec. IX-A].

To account for amplification and noise, the NLSE (2.7) can be extended by inserting a gain profile $g(z)$ and a complex-valued stochastic process $w(t, z)$, resulting in

$$\frac{\partial v(t,z)}{\partial z} = -\frac{\alpha - g(z)}{2} v(t,z) - \jmath \frac{\beta_2}{2} \frac{\partial^2 v(t,z)}{\partial t^2} + \jmath \gamma v(t,z)|v(t,z)|^2 + w(t,z). \tag{2.18}$$

Equation (2.18) is referred to as the stochastic nonlinear Schrödinger equation (sNLSE) [23]. We first discuss the gain profile $g(z)$ and its effect on the power profile of the signal $v(t, z)$, ignoring all other effects (including $w(t, z)$). Both amplification types are applied periodically, in the sense that the entire transmission distance $0 \leq z \leq L$ is split up into spans of length L_{sp}, varying between 60 and 120 km, where $N_{\text{sp}} = L/L_{\text{sp}} \in \mathbb{N}$ is the total number of spans. In the case of lumped amplification, an optical amplifier, most often an erbium-doped fiber amplifier (EDFA) [2, Sec. IX-B], is inserted after each span, where the amplifier gain G matches the power loss of the signal in that span, i.e., $G = e^{\alpha L_{\text{sp}}}$. In (2.18), this is accounted for by setting $g(z) = \alpha L_{\text{sp}} \sum_{i=1}^{N_{\text{sp}}} \delta(z - iL_{\text{sp}})$. The corresponding power profile is illustrated in Fig. 2.4(a). The signal power decreases exponentially according to the loss coefficient α and is periodically restored to the input power P after each span. In the case of distributed amplification, it is assumed that the signal power can be held at an approximately constant level as shown in Fig. 2.4(b). In order to achieve this, pump waves are launched into the fiber at Raman pump stations (RPSs) which are located at the beginning and after each span [2, Sec. IX-B]. The pump waves co-propagate together with the signal $v(t, z)$ and the nonlinear nature of the fiber is exploited to continuously transfer energy from the pump wave to the signal. In Fig. 2.4(b), the "realistic" power profile (dashed line) is schematically reproduced from [24, Fig. 3] and assumes two pump waves per span, one propagating co-directionally

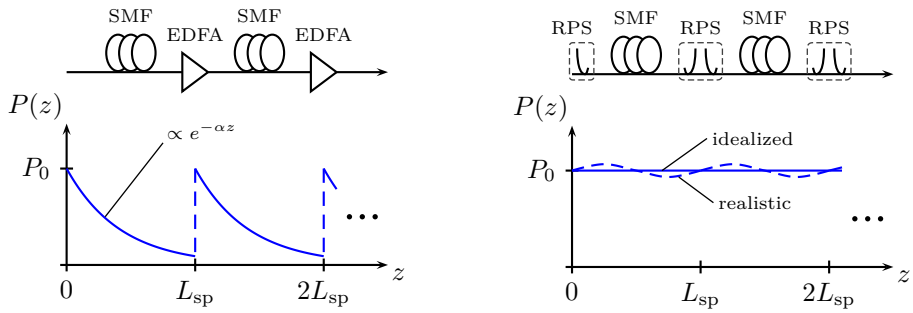
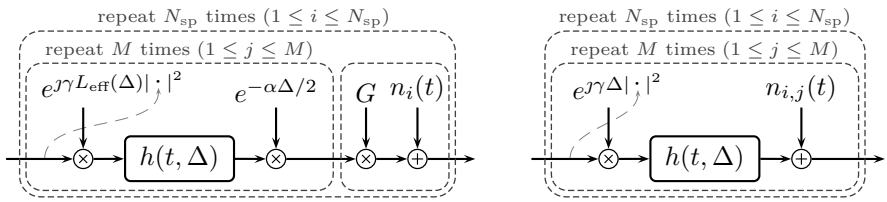


Figure 2.4: Schematic comparison of the power profile as a function of the transmission distance z for the two considered amplification types.

and one contra-directionally to the signal. From a modeling perspective, the "idealized" constant profile (solid line) is assumed for simplicity, where $g(z) = \alpha$ and hence the first term on the right-hand side of (2.18) simply disappears.

Next, we discuss the noise that is generated by the optical amplification schemes through a process called amplified spontaneous emission (ASE). For lumped amplification, noise can be thought of as being added to the signal at discrete locations $z_i \triangleq iL_{\mathrm{sp}}$, $1 \leq i \leq N_{\mathrm{sp}}$. Thus, if $v(t, z_i^-)$ is the output signal after the ith fiber span, the input signal to the next span is given by $v(t, z_i^+) = Gv(t, z_i^-) + n_i(t)$, where $n_i(t)$ is the additive noise originating from the ith amplifier [19, p. 36]. It has been experimentally verified that ASE noise can be accurately modeled as circularly symmetric complex Gaussian [2, p. 667] and therefore it remains to specify the autocorrelation function of $n_i(t)$, where processes from different amplifiers are uncorrelated. The most common assumption is white Gaussian noise, i.e., $\mathbb{E}[N_i(t)N_j^*(t')] = \mathsf{N}_\ell \delta(t-t')\delta[i-j]$, where the noise PSD per amplifier for EDFAs is computed as $\mathsf{N}_\ell = (G-1)h\nu_s n_{\mathrm{sp}}$ [2, eq. (54)]. The meaning and values of the quantities appearing in this expression are summarized in Table 2.1 at the end of this section. We further set $\mathsf{N}_0 = N_{\mathrm{sp}}\mathsf{N}_\ell$ in the case of lumped amplification, which one might think of as the cumulative PSD at the end of the transmission link for N_{sp} amplifiers. Since temporally white noise has infinite instantaneous power, this assumption would, however, lead to infinite phase rotations due to the nonlinear effect. In reality, the noise power is of course finite, and the PSD of ASE noise is comparable to the gain spectrum of the amplifier. For an idealized EDFA that provides flat gain over a certain frequency range W_N, one would then replace $\delta(t-t')$ with $\delta_{W_N}(t-t')$ where $\delta_{W_N}(x) = W_N \mathrm{sinc}(W_N x)$ [19]. Further limitations of the optical bandwidth can occur due to the insertion of optical bandpass filters and/or reconfigurable optical add-drop multiplexers (ROADMs) along the transmission line [2].

Based on the previous description, a block diagram of a continuous-time model describing a multi-span transmission link with lumped amplification is depicted in Fig. 2.5(a).



(a) lumped **(b)** distributed

Figure 2.5: Block diagram for a multi-span link including amplification and noise for (a) lumped and (b) distributed amplification.

The model consists of the concatenation of the deterministic model for an SMF based on the SSFM (cf. Fig. 2.3(b)) with a multiplicative gain and additive noise for the optical amplifier. For completeness, we also indicate how the additive noise terms $n_i(t)$ can be related to $w(t,z)$ in (2.18) for lumped amplification. Note that if we neglect all terms on the right-hand side of (2.18) except $w(t,z)$, we have $\partial v(t,z)/\partial z = w(t,z)$ and integrating this equation leads to

$$v(t,z) = v(t,0) + \int_0^z w(t,\xi)\,\mathrm{d}\xi = v(t,0) + n(t,z). \tag{2.19}$$

Here, $n(t,z)$ represents the noise that is added to the signal up to a certain distance z. For lumped amplification, one may set $w(t,z) = \sum_{i=1}^{N_{\mathrm{sp}}} n_i(t)\delta(z - iL_{\mathrm{sp}})$ [20, p. 84], so that $n(t,z) = \sum_{i=1}^{\lfloor z/L_{\mathrm{sp}} \rfloor} n_i(t)$ corresponds the addition of all $n_i(t)$ up to distance z (the upper integral limit in (2.19) is interpreted as z^+).

Next, we discuss distributed amplification, where ASE noise is continuously added throughout the entire transmission link. A common assumption is that $w(t,z)$ is a white Gaussian stochastic process in both time and space, and hence [2, eq. (53)] [19, p. 37]

$$\mathbb{E}\left[W(t,z)W^*(t',z')\right] = \bar{\mathsf{N}}_{\mathrm{d}}\delta(t-t')\delta(z-z'), \tag{2.20}$$

where $\bar{\mathsf{N}}_{\mathrm{d}}$ is the distributed PSD per unit length (in [W/km/Hz]) computed as $\bar{\mathsf{N}}_{\mathrm{d}} = \alpha h \nu_s K_T$, where K_T is the photon occupancy factor. Similarly as for lumped amplification, we set $\mathsf{N}_0 = L\bar{\mathsf{N}}_{\mathrm{d}}$ as the PSD at the end of the entire transmission line of length L. The expressions for N_0 for the two amplification types are related via $N_{\mathrm{sp}} = L/L_{\mathrm{sp}}$ and letting $L_{\mathrm{sp}} \to 0$, and replacing n_{sp} with K_T. Regarding the temporal correlation of $w(t,z)$, one can make similar arguments as for the lumped case and replace $\delta(t-t')$ with $\delta_{W_N}(t-t')$ for some W_N to account for the bandwidth limitation of physically realistic noise. Under the assumption that $w(t,z)$ is uncorrelated in space, $n(t,z)$ in (2.19) is a Wiener process, i.e., the integral of a white Gaussian process. As pointed out in [25, Sec. III], the sNLSE then has to be interpreted via an equivalent integral representation (similar to (2.19) but including all terms of (2.18)), since a Wiener process is

not differentiable for any z to satisfy $\partial n(t,z)/\partial z = w(t,z)$. Here, we will ignore such issues and rely on some intuition to describe how the Wiener process can be included in the SSFM to numerically solve the sNLSE with distributed noise. In particular, we assert that the distributed nature of the noise can be accurately captured by adding a Gaussian stochastic process *in each segment* of the SSFM. We denote the process that is added in the jth segment of the ith fiber span by $n_{i,j}(t)$, where processes from different segments and spans are uncorrelated. In order for all MN_{sp} processes $n_{i,j}(t)$ to produce the same PSD as the Wiener process at the end of the link, i.e., $\mathsf{N}_0 = L\bar{\mathsf{N}}_{\mathrm{d}}$, we require $\mathbb{E}[N_{j,i}(t)N_{j,i}^*(t')] = L\bar{\mathsf{N}}_{\mathrm{d}}/(MN_{\mathrm{sp}})\delta(t-t')$. The corresponding block diagram is shown in Fig. 2.5(b). Compared to the lumped amplification case in Fig. 2.5(a), it can be seen that the attenuation step is removed (also $L_{\mathrm{eff}}(\Delta) = \Delta$, since $\alpha = 0$) and an additive noise term is included in each segment.

All relevant quantities that have been described in this section for the two amplification types are summarized in Table 2.1.

2.3.1 Linear Regime

We briefly discuss the case where $\gamma = 0$ but now including ASE noise from the optical amplification schemes. To that end, consider again the two block diagrams in Fig. 2.5(a) and (b). If the nonlinear phase rotations are removed, the model indeed reverts to the linear dispersive AWGN channel (2.1) (see also Fig. 2.1) for both amplification types. To see this, first note that the attenuation and gain factors in Fig. 2.5(a) cancel out due to the linearity of the model. Further, due to the all-pass nature of the dispersive filters, one may freely rearrange the additive noise terms because filtered noise remains Gaussian with the same PSD. Thus, we may assume that all noise processes are added together at the end of the transmission link and the model in fact corresponds to the linear dispersive AWGN channel in both cases [20, Sec. 5.1]. We have $h(t) = h(t,L)$ (convolving $h(t,\Delta)$ MN_{sp} times with itself) and $n(t) = \sum_{i=1}^{N_{\mathrm{sp}}} n_i(t)$ and $n(t) = \sum_{i=1}^{N_{\mathrm{sp}}}\sum_{j=1}^{M} n_{i,j}(t)$ for lumped and distributed amplification, respectively.

2.4 Channel Models in the Appended Papers

2.4.1 Paper A: Zero-Dispersion Fiber, Single Polarization

In Paper A, we consider the special case where dispersive effects are absent, i.e., $\beta_2 = 0$, in combination with a distributed amplification scheme. In Section 2.2.2, it has already been shown that it is possible to find an analytical solution for the (deterministic) NLSE if $\beta_2 = 0$. When ASE noise from optical amplifiers is also considered, the signal and noise interact through the fiber nonlinearity and give rise to the phenomenon of nonlinear phase noise (NLPN). Fortunately, it turns out that an exact analytical characterization of a discrete-time channel model can still be found. We proceed by first describing how the discrete-time channel is obtained from the continuous-time channel. Based on this

Table 2.1: Comparison of amplification types

quantity	meaning	lumped	distributed
$g(z)$	gain profile	$\alpha L_{\mathrm{sp}} \sum_{i=1}^{N_{\mathrm{sp}}} \delta(z - iL_{\mathrm{sp}})$	α
$w(t,z)$	added to $\partial v(t,z)/\partial z$	$\sum_{i=1}^{N_{\mathrm{sp}}} n_i(t)\delta(z - iL_{\mathrm{sp}})$	white Gaussian
$n(t,z)$	$\int_0^z w(t,\xi)\,\mathrm{d}\xi$	$\sum_{i=1}^{\lfloor z/L_{\mathrm{sp}} \rfloor} n_i(t)$	Wiener process
$\mathbb{E}\left[W(t,z)W^*(t',z')\right]$	-	$\mathsf{N}_\ell \delta(t-t') \sum_{i=1}^{N_{\mathrm{sp}}} \delta(z - iL_{\mathrm{sp}})\delta(z' - iL_{\mathrm{sp}})$	$\bar{\mathsf{N}}_{\mathrm{d}}\delta(t-t')\delta(z-z')$
$\mathbb{E}[N(t,z)N^*(t',z')]$	-	$\mathsf{N}_\ell \delta(t-t') \min(\lfloor z/L_{\mathrm{sp}} \rfloor, \lfloor z'/L_{\mathrm{sp}} \rfloor)$	$\bar{\mathsf{N}}_{\mathrm{d}}\delta(t-t') \min(z,z')$
N_0	PSD after distance L	$N_{\mathrm{sp}}\mathsf{N}_\ell = N_{\mathrm{sp}}(e^{\alpha L_{\mathrm{sp}}} - 1)h\nu n_{\mathrm{sp}}$	$L\bar{\mathsf{N}}_{\mathrm{d}} = N_{\mathrm{sp}} L_{\mathrm{sp}} \alpha h\nu K_T$
F_n	amplifier noise figure	typically 4–7 dB	-
n_{sp}	spontaneous emission factor	$F_n(1 - G^{-1})^{-1}/2$	-
K_T	photon occupancy factor	-	≈ 1.13 (at room temp.)
h	Planck's constant	$6.626 \cdot 10^{-34}$ [Js]	
ν_s	optical carrier frequency	$1.936 \cdot 10^{14}$ [Hz]	

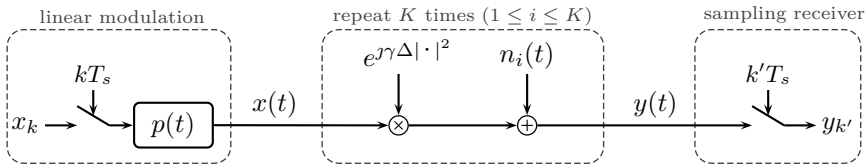


Figure 2.6: Block diagram of the transmission system considered in Paper A. Note that the absence of a filter front-end at the receiver means that the noise bandwidth is implicitly assumed to be limited due to inline optical filters.

discrete-time channel, an analytical expression for the conditional PDF is presented, which is essentially the starting point for the discussions in Paper A.

For the case of distributed amplification and in the absence of dispersion, the sNLSE (2.18) reduces to

$$\frac{\partial v(t,z)}{\partial z} = \jmath\gamma v(t,z)|v(t,z)|^2 + w(t,z). \tag{2.21}$$

A block diagram of the corresponding continuous-time channel model (via the SSFM) is then derived from Fig. 2.5(b) by simply removing the dispersive filters. A discrete-time channel model can be obtained by assuming a conventional linear pulse modulation (cf. Fig. 2.1) in combination with a sampling receiver. The resulting block diagram depicting this scenario is shown in Fig. 2.6. For simplicity, and to make the notation consistent with Paper A, we have replaced the double enumeration over segments and spans with a single enumeration over $1 \leq i \leq K = MN_{\mathrm{sp}}$.

Due to the absence of a filter prior to sampling the received signal $y(t)$ (see Fig. 2.6), it is implicitly assumed that the noise bandwidth W_N is somehow limited during propagation, e.g., through inline optical filters (otherwise the samples $y_{k'}$ would have infinite variance) [26]. Furthermore, due to the absence of a matched filter (matched to the pulse shape), the pulse $p(t)$ itself rather than the self-similarity function of the pulse $\mathsf{R}_p(t)$ needs to fulfill the Nyquist criterion $p(kT_s) = \delta[k]$, e.g., one may assume sinc pulses.

It is important to point out that the step from continuous-time to discrete-time is not necessarily optimal, i.e., the samples $y_{k'}$ do not necessarily form a sufficient statistic for detecting x_k based on $y(t)$. Therefore, statements about optimality (e.g., "ML detection") are implicitly understood with respect to the discrete-time channel only, not with respect to the actual waveform channel.

The channel from the transmitted symbols to the received samples is memoryless and the indices can hence be dropped, i.e., x and y denote the (complex-valued) channel input and output, respectively. The system model presented thus far has been extensively studied in the literature and there exist several different derivations of the conditional PDF $f_{Y|X}(y|x)$, some of which we mention in the following in chronological order.

- Gordon and Mollenauer were the first to recognize that the interaction between

the signal and inline amplifier noise leads to a phase noise effect [27] and NLPN is sometimes referred to as the Gordon–Mollenauer effect.

- Their model was made rigorous by Mecozzi in [23] and later in [28].
- In [29], Turitsyn *et al.* proposed a technique to derive the PDF based on the Martin–Siggia–Rose formalism from statistical mechanics. It was also recognized that the capacity of the discrete channel grows unbounded with input power.
- Ho provided an in-depth treatment on the subject of NLPN including a derivation of the PDF based on characteristic functions [10, Ch. 5], see also [30].
- Most recently, Yousefi and Kschischang have considered the zero-dispersion case in a series of papers [25,26,31,32]. They provide several additional derivations of the PDF based on a sum-product and a Fokker–Planck differential equation approach.

The PDF is given as follows. Adopting a polar notation for the channel input and output including a magnitude normalization with $\sigma_{\mathrm{d}}^2 = L\bar{\mathsf{N}}_{\mathrm{d}}W_N = \mathsf{N}_0 W_N$ according to $x/\sigma_{\mathrm{d}} = r_0 e^{\jmath\theta_0}$ and $y/\sigma_{\mathrm{d}} = re^{\jmath\theta}$, one may write the PDF in the form of a Fourier series as[9]

$$f_{\Theta,R|\Theta_0,R_0}(\theta, r|\theta_0, r_0) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} C_k(r, r_0)e^{\jmath k(\theta-\theta_0)}, \tag{2.22}$$

where the Fourier coefficients are given by

$$C_k(r, r_0) = \frac{2z_k}{\sin z_k} r \exp\left(-\frac{r^2 + r_0^2}{(\tan z_k)/z_k}\right) \mathrm{I}_k\left(\frac{2z_k}{\sin z_k} r r_0\right), \tag{2.23}$$

where $\mathrm{I}_k(\,\cdot\,)$ is the modified Bessel function of the first kind and $z_k = \sqrt{\jmath\gamma k\sigma_{\mathrm{d}}^2}$. The singularities for $z_0 = 0$ are understood to be resolved as $(\sin 0)/0 = 1$ and $(\tan 0)/0 = 1$. Since the PDF is a real function, the symmetry condition $C_k(r, r_0) = C_{-k}(r, r_0)$ holds.

The zero-dispersion assumption can be motivated by the fact that the dispersion coefficient β_2 can be physically engineered to take on values over a certain range including (approximately) zero. However, this scenario is generally assumed to be unrealistic due to the severe spectral broadening that may occur during propagation. For example, in [2] the authors state that "the zero dispersion region is generally to be avoided as the effects of fiber nonlinearity are enhanced dramatically". Similarly, in [20] it is noted that "the zero-dispersion regime is not practical for communications". A discussion about this topic can also be found in [26, Sec. VIII].

The model can nonetheless be useful since the assumption of zero dispersion is sometimes fulfilled, at least approximately, in dispersion-managed (DM) transmission links.

[9] The joint density of the magnitude and phase of a complex random variable with density $f_Y(y)$ can be obtained via $f_{\Theta,R}(\theta, r) = r f_Y(re^{\jmath\theta})$ [16, Lem. 17.3.5].

These links consist of the concatenation of a standard SMF with a dispersion compensating fiber (DCF) whose dispersion coefficient β_2' has been engineered to be of opposite sign, i.e., $\mathrm{sgn}(\beta_2') = -\mathrm{sgn}(\beta_2)$. If the length of the DCF L_{sp}' is matched to the length of the standard fiber L_{sp} according to the simple linear relation $|\beta_2|L_{\mathrm{sp}} = |\beta_2'|L_{\mathrm{sp}}'$, then the net effect of the dispersion after distance $L_{\mathrm{sp}}+L_{\mathrm{sp}}'$ is zero (neglecting nonlinear effects). To see this, consider the concatenation of two dispersive filters $H(f, L_{\mathrm{sp}}) = \exp(\jmath 2\beta_2\pi^2 f^2 L_{\mathrm{sp}})$ (cf. (2.10)) and $H'(f, L_{\mathrm{sp}}') = \exp(\jmath 2\beta_2'\pi^2 f^2 L_{\mathrm{sp}}')$ which exactly cancel in this case. If nonlinear effects are taken into consideration for such DM links, the accuracy of the zero-dispersion model depends essentially on the symbol rate R_s and lengths L_{sp} and L_{sp}'. In general, the lower the symbol rate and the shorter the fibers, the more accurate is the model.

2.4.2 Paper B: Zero-Dispersion Fiber, Polarization Multiplexing

In Paper B, we study an extension of the model used in Section 2.4.1 which was recently derived in [33]. The model takes into account PM transmission, where both polarizations of the light are used to transmit data. For PM transmission, the sNLSE equation can be further extended by considering the vector signal $\boldsymbol{v}(t,z) = (v_{\mathsf{a}}(t,z), v_{\mathsf{b}}(t,z))^{\intercal}$, where the indices indicate the signals in the two polarizations a and b.[10] The resulting equation is referred to as the Manakov equation including loss and gain terms and amplifier noise and is given by [34, p. 8]

$$\frac{\partial \boldsymbol{v}(t,z)}{\partial z} = -\frac{\alpha - g(z)}{2}\boldsymbol{v}(t,z) - \jmath\frac{\beta_2}{2}\frac{\partial^2 \boldsymbol{v}(t,z)}{\partial t^2} + \jmath\gamma \boldsymbol{v}(t,z)\|\boldsymbol{v}(t,z)\|^2 + \boldsymbol{w}(t,z), \tag{2.24}$$

where $\boldsymbol{w}(t,z) = (w_{\mathsf{a}}(t,z), w_{\mathsf{b}}(t,z))^{\intercal}$ are two (independent) stochastic processes describing the ASE noise generated in both polarizations. The major difference between (2.24) and (2.18) is that (2.24) models the nonlinearity that is due to the sum of the instantaneous power in both polarizations $\|\boldsymbol{v}(t,z)\|^2 = P_{\mathsf{a}}(t,z)^2 + P_{\mathsf{b}}(t,z)^2$. We should mention that (2.18) ignores the fact that amplifier noise is always generated "in two polarizations", i.e., even if we assume one of the two signals in $\boldsymbol{v}(t,z)$ to be zero (as was done in Section 2.3), technically the amplifier noise in that polarization still contributes via (2.24) through the fiber nonlinearity.

The derivation presented in [33] makes similar assumptions as in, e.g., [10, 23, 26, 29] for the continuous-to-discrete time conversion and the subsequent analysis. In particular, dispersive effects are ignored. This also includes polarization mode dispersion (PMD), which would cause different group velocities of the signals in polarization a and b because of natural imperfections and asymmetries of the fiber cross-section area. The simplified Manakov equation for distributed amplification and neglecting all dispersive effects is

[10] This nonstandard notation for the polarizations is an attempt to avoid confusion with the transmit and received signals. However, we acknowledge inconsistent notation with respect to Paper B, where the polarizations are denoted by x and y.

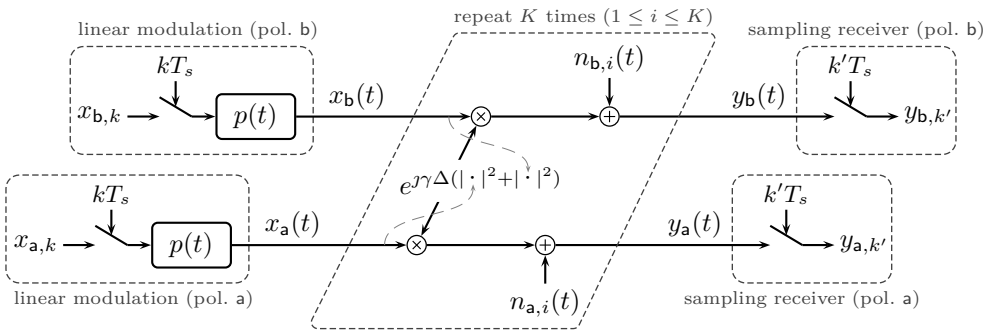


Figure 2.7: Block diagram of the transmission system considered in Paper B.

obtained from (2.24) as

$$\frac{\partial \boldsymbol{v}(t,z)}{\partial z} = \jmath\gamma \boldsymbol{v}(t,z)\|\boldsymbol{v}(t,z)\|^2 + \boldsymbol{w}(t,z). \tag{2.25}$$

Again, noise is assumed to be bandlimited and a discrete channel is obtained based on samples that are taken at the receiver at a rate $1/T_s$ in both polarizations. A block diagram of the assumed transmission system including a graphical representation of (2.25) via the SSFM is shown in Fig. 2.7. It can be seen that there are essentially two SP transmission systems which are now coupled due to the fiber nonlinearity (and hence uncoupled when $\gamma = 0$).

As for the SP case, the discrete-time model is memoryless. Dropping the discrete-time indices, the joint input and output in both polarizations are denoted by $\mathbf{y} = (y_{\mathsf{a}}, y_{\mathsf{b}})^{\mathsf{T}}$ and $\mathbf{x} = (x_{\mathsf{a}}, x_{\mathsf{b}})^{\mathsf{T}}$, respectively. A notation based on polar coordinates with the same normalization as before is preferable and we therefore set $x_{\mathsf{a}}/\sigma_{\mathrm{d}} = r_{0\mathsf{a}}e^{\jmath\theta_{0\mathsf{a}}}$, $x_{\mathsf{b}}/\sigma_{\mathrm{d}} = r_{0\mathsf{b}}e^{\jmath\theta_{0\mathsf{b}}}$, $y_{\mathsf{a}}/\sigma_{\mathrm{d}} = r_{\mathsf{a}}e^{\jmath\theta_{\mathsf{a}}}$, $y_{\mathsf{b}}/\sigma_{\mathrm{d}} = r_{\mathsf{b}}e^{\jmath\theta_{\mathsf{b}}}$, and collect the corresponding transmitted and received magnitudes and phases in the vectors $\boldsymbol{r}_0 = (r_{0\mathsf{a}}, r_{0\mathsf{b}})^{\mathsf{T}}$, $\boldsymbol{r} = (r_{\mathsf{a}}, r_{\mathsf{b}})^{\mathsf{T}}$, $\boldsymbol{\theta}_0 = (\theta_{0\mathsf{a}}, \theta_{0\mathsf{b}})^{\mathsf{T}}$, $\boldsymbol{\theta} = (\theta_{\mathsf{a}}, \theta_{\mathsf{b}})^{\mathsf{T}}$. Using this notation, the conditional PDF is given in the form of a two-dimensional Fourier series as

$$f_{\boldsymbol{\Theta},\boldsymbol{R}|\boldsymbol{\Theta}_0,\boldsymbol{R}_0}(\boldsymbol{\theta},\boldsymbol{r}|\boldsymbol{\theta}_0,\boldsymbol{r}_0) = \frac{1}{4\pi^2}\sum_{k_{\mathsf{a}}\in\mathbb{Z}}\sum_{k_{\mathsf{b}}\in\mathbb{Z}} C_{\boldsymbol{k}}(\boldsymbol{r},\boldsymbol{r}_0)e^{\jmath\boldsymbol{k}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)}, \tag{2.26}$$

where $\boldsymbol{k} = (k_{\mathsf{a}}, k_{\mathsf{b}})^{\mathsf{T}}$ and the Fourier coefficients are

$$C_{\boldsymbol{k}}(\boldsymbol{r},\boldsymbol{r}_0) = \left(\frac{2z_{\boldsymbol{k}}}{\sin z_{\boldsymbol{k}}}\right)^2 r_{\mathsf{a}}r_{\mathsf{b}} \exp\left(-\frac{\|\boldsymbol{r}\|^2+\|\boldsymbol{r}_0\|^2}{(\tan z_{\boldsymbol{k}})/z_{\boldsymbol{k}}}\right) \mathrm{I}_{k_{\mathsf{a}}}\left(\frac{2z_{\boldsymbol{k}}}{\sin z_{\boldsymbol{k}}}r_{\mathsf{a}}r_{0\mathsf{a}}\right) \mathrm{I}_{k_{\mathsf{b}}}\left(\frac{2z_{\boldsymbol{k}}}{\sin z_{\boldsymbol{k}}}r_{\mathsf{b}}r_{0\mathsf{b}}\right), \tag{2.27}$$

with $z_{\boldsymbol{k}} = \sqrt{\jmath\gamma(k_{\mathsf{a}}+k_{\mathsf{b}})\sigma_{\mathrm{d}}^2}$. Since the PDF is a real function, the Fourier coefficients satisfy the symmetry condition $C_{\boldsymbol{k}}(\boldsymbol{r},\boldsymbol{r}_0) = C_{-\boldsymbol{k}}(\boldsymbol{r},\boldsymbol{r}_0)$.

Visualizing such a four-dimensional PDF can be difficult and we make an attempt by showing various scatter plots further below and in Paper B. When showing a scatter plot of the received points in only polarization a, one is effectively showing a particle representation of the marginal distribution

$$f_{\Theta_\mathsf{a},R_\mathsf{a}|\boldsymbol{\Theta}_0,\boldsymbol{R}_0}(\theta_\mathsf{a},r_\mathsf{a}|\boldsymbol{\theta}_0,\boldsymbol{r}_0) = \int_0^{2\pi}\int_0^{\infty} f_{\boldsymbol{\Theta},\boldsymbol{R}|\boldsymbol{\Theta}_0,\boldsymbol{R}_0}(\boldsymbol{\theta},\boldsymbol{r}|\boldsymbol{\theta}_0,\boldsymbol{r}_0)\,\mathrm{d}r_\mathsf{b}\,\mathrm{d}\theta_\mathsf{b} \tag{2.28}$$

$$= \frac{1}{4\pi^2}\sum_{k_\mathsf{a}\in\mathbb{Z}}\sum_{k_\mathsf{b}\in\mathbb{Z}} e^{\jmath k_\mathsf{a}\theta_\mathsf{a}}\int_0^{\infty} C_{\boldsymbol{k}}(\boldsymbol{r},\boldsymbol{r}_0)\,\mathrm{d}r_\mathsf{b}\int_0^{2\pi} e^{\jmath k_\mathsf{b}\theta_\mathsf{b}}\,\mathrm{d}\theta_\mathsf{b} \tag{2.29}$$

$$= \frac{1}{2\pi}\sum_{k_\mathsf{a}}\tilde{C}_{k_\mathsf{a}}(r_\mathsf{a},\boldsymbol{r}_0)e^{\jmath k_\mathsf{a}\theta_\mathsf{a}}, \tag{2.30}$$

where the modified Fourier coefficient for the above marginal PDF is [33]

$$\tilde{C}_{k_\mathsf{a}}(r_\mathsf{a},\boldsymbol{r}_0) = \int_0^{\infty} C_{(k_\mathsf{a},0)}(\boldsymbol{r},\boldsymbol{r}_0)\,\mathrm{d}r_\mathsf{b} \tag{2.31}$$

$$= \frac{2z_{\boldsymbol{k}}}{\sin z_{\boldsymbol{k}}} r_\mathsf{a}\exp\left(-\frac{r_\mathsf{a}^2+r_{0\mathsf{a}}^2}{(\tan z_{\boldsymbol{k}})/z_{\boldsymbol{k}}}\right) I_{k_\mathsf{a}}\left(\frac{2z_{\boldsymbol{k}}}{\sin z_{\boldsymbol{k}}} r_\mathsf{a} r_{0\mathsf{a}}\right)\cdot\frac{\exp\left(r_{0\mathsf{b}}^2\cdot z_{\boldsymbol{k}}\tan z_{\boldsymbol{k}}\right)}{\cos z_{\boldsymbol{k}}}. \tag{2.32}$$

Due to symmetry, the polarizations a and b have the same marginal distributions. Also, the above distributions are conditioned on a given symbol in *both* polarizations.

In Paper B, we restrict ourselves to M-PSK with the same power in both polarizations in order to simplify the detection problem.

2.4.3 Paper C: Non-Dispersion-Managed Links with a Linear Receiver

For Paper C, we consider PM transmission without neglecting dispersive effects that are due to β_2 (we do, however, neglect PMD). The optical transmission link consists of the periodic concatenation of a standard SMF and an EDFA (i.e., a lumped amplification scheme) and there is no optical inline dispersion compensation through DCFs. A block diagram of this setup is shown in Fig. 2.8. The transmitters (TX) employ a linear pulse modulation according to $x_\mathsf{a}(t) = \sum_k x_{\mathsf{a},k}p(t-kT_s)$ for polarization a and similarly for polarization b. The evolution of the PM signal is described by the Manakov equation (2.24). The received signal in each polarization is processed according to the linear matched filter receiver shown in Fig. 2.1. For polarization a, this amounts to passing $y_\mathsf{a}(t)$ through an equalizer, a pulse-matched filter, and a sampler, to obtain $y_{\mathsf{a},k'} = y_\mathsf{a}(t)\circledast h(t,-L)\circledast p(-t)|_{t=k'T_s}$ and similarly for polarization b.

Characterizing the statistical relationship between the transmitted symbols and received samples is a challenging task due to the complicated interaction of the signal with itself, the noise, and the signal in the orthogonal polarization. The crucial difference with respect to the setup in Paper A and Paper B is the presence of dispersive filtering

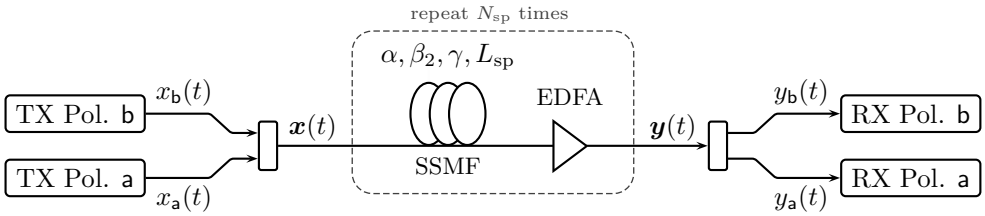


Figure 2.8: Block diagram of the PM transmission scheme considered in Paper C.

effects throughout the signal propagation. Optical transmission links without any inline dispersion compensation are referred to as non-DM or uncompensated transmission links. Recently, there has been a substantial amount of work on these types of transmission links with the goal to find such a statistical relationship [35–38].[11]

In [35], it is shown that the discrete-time channel for non-DM links is well modeled by a circularly symmetric complex additive Gaussian channel including a complex scaling factor. In the derivation of the model, the assumption is that dispersive effects are dominant (i.e., the symbol rate is high enough) and that the nonlinear effects are not too strong. The complex scaling accounts for a constant phase offset as well as the fact that part of the signal is converted into noise-like interference through the interaction between the dispersive and nonlinear effects. For simplicity, it is then assumed that the nonlinear noise is additive, Gaussian, and uncorrelated (both in time and across polarizations). Thus, the discrete-time channel model in polarization a is given according to

$$y_{\mathsf{a}} = \zeta x_{\mathsf{a}} + n_{\mathsf{a}} + \tilde{n}_{\mathsf{a}}, \tag{2.33}$$

where $\zeta \in \mathbb{C}$ is a complex scaling factor, n_{a} corresponds to the linear ASE noise with $\mathbb{E}[N_{\mathsf{a}} N_{\mathsf{a}}^*] = \mathsf{N}_0/T_s$, and $\tilde{n}_{\mathsf{a}}$ accounts for nonlinear noise with $\mathbb{E}[\tilde{N}_{\mathsf{a}} \tilde{N}_{\mathsf{a}}^*] = \eta P^3$, where the same transmit power P is assumed for both signals in the two polarizations. η (and hence the nonlinear noise variance) is a function of the link parameters and the symbol time, i.e., $\eta = f(\alpha, \beta_2, \gamma, L_{\text{sp}}, N_{\text{sp}}, T_s)$ [35, eq. (15)], and $|\zeta|^2 = 1 - |\eta| P^2$.

The main difference with respect to a "conventional" discrete-time additive Gaussian channel is that the signal-to-noise ratio (SNR) (defined as the ratio of the input power to the *additive* noise power) is not sufficient to characterize the operating point of the channel but rather one needs to consider the pair (P, P_{ASE}) or, more practically relevant, the pair (P, L). This parameter pair in turn leads to both a linear and a nonlinear noise variance based on which an effective SNR can be computed.

[11] These links are also of high practical relevance and according to [36], "the current consensus is that green-field installations, as well as major overhauling and refurbishing of existing links, should adopt uncompensated transmission."

2.4.4 Comparison

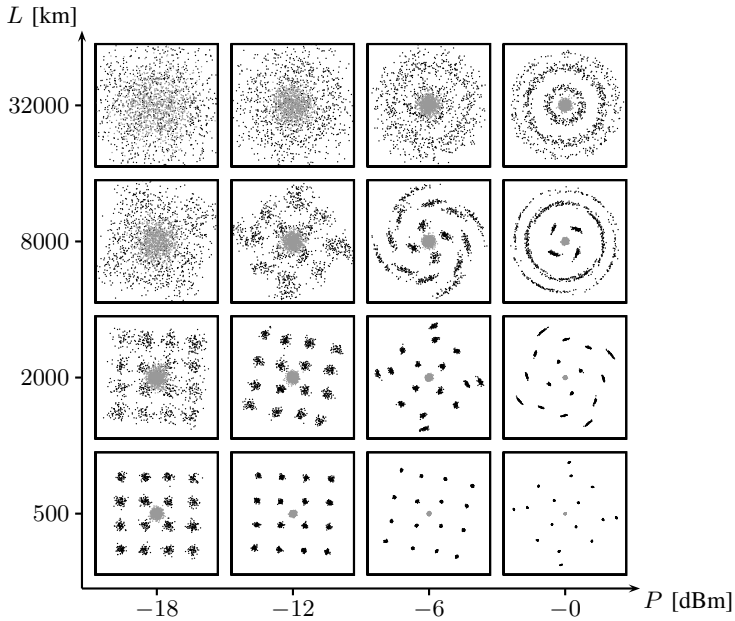
In order to illustrate the previously described models, we show scatter plots of the received symbols for different combinations of input power P and transmission distance L in Fig. 2.9. The transmitted symbols are taken uniformly at random from the 16-QAM signal constellation. For Fig. 2.9(a) and (b), we use the same link parameters as in Paper A. However, when generating the scatter plots shown in Fig. 2.9(a), we take into account the noise in *both* polarizations. In fact, the gray points in Fig. 2.9(a) correspond to the received symbols in the unused polarization and hence they represent only the noise. The scatter plots along diagonals in Fig. 2.9 correspond to the same signal-to-additive-noise ratio. For Fig. 2.9(c), we use the same parameters as in Paper C. We caution the reader that a quantitative comparison between the scatter plots shown for the models in Paper A/B and C is not fair due to the different link parameters and amplification schemes. Furthermore, spectral considerations are not reflected in the scatter plots. For example, operating at high input power and transmission length seems feasible for the memoryless models, since the phase is predominantly distorted. However, in this transmission regime severe spectral broadening of the signal is to be expected. Also, for the PM transmission shown in Fig. 2.9(b), the nonlinear phase rotation in one polarization depends on the selected point in the other polarization. Therefore, the noise clouds at high input power begin to separate into three smaller clouds, where each cloud corresponds to a different magnitude of the transmitted symbol in the orthogonal polarization.

From Fig. 2.9(c), it can be observed that the Gaussian noise assumption for the nonlinear interference appears to be valid for a wide range of transmission parameters. For the scatter plot in the lower right corner, is seems that the outer symbols are more affected by phase noise rather than circularly symmetric Gaussian noise, which can be explained by the fact that the transmission distance (two fiber spans) is relatively short.

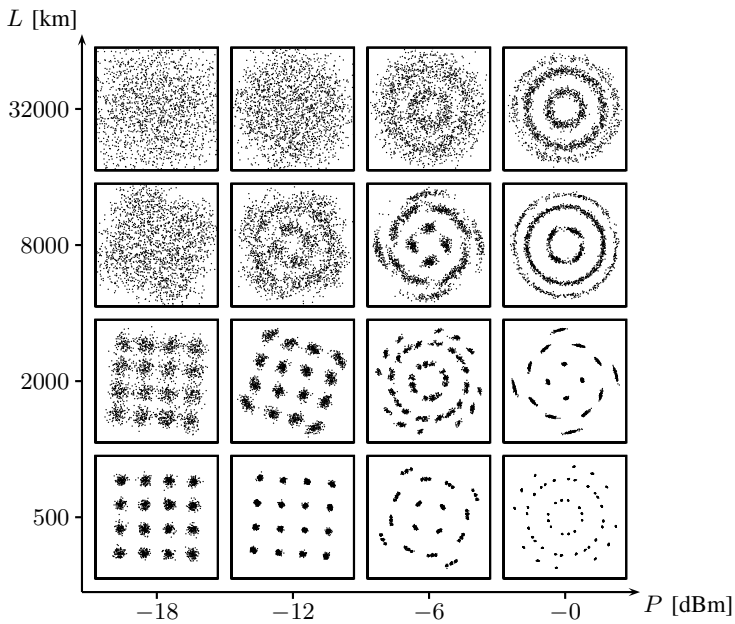
2.5 Further Reading

In this chapter, we have presented an overview of the basic concepts regarding fiber-optical channel modeling for long-haul data transmission. However, there are many important issues that have not been addressed, some of which we briefly mention in this section.

One important topic is the study of wavelength-division multiplexing (WDM) systems, where many input signals are multiplexed in the frequency domain at different carrier frequencies and simultaneously transmitted over the fiber. In this case, the underlying channel model, i.e., the sNLSE together with the corresponding block diagrams via the SSFM in Fig. 2.5, are still valid. However, nonlinear effects may be more pronounced due to the additional input power. Furthermore, it was mentioned in the beginning of this chapter that the matched filter receiver can be seen as a baseline. The development of improved receiver structures is an active area of research. For example, digital

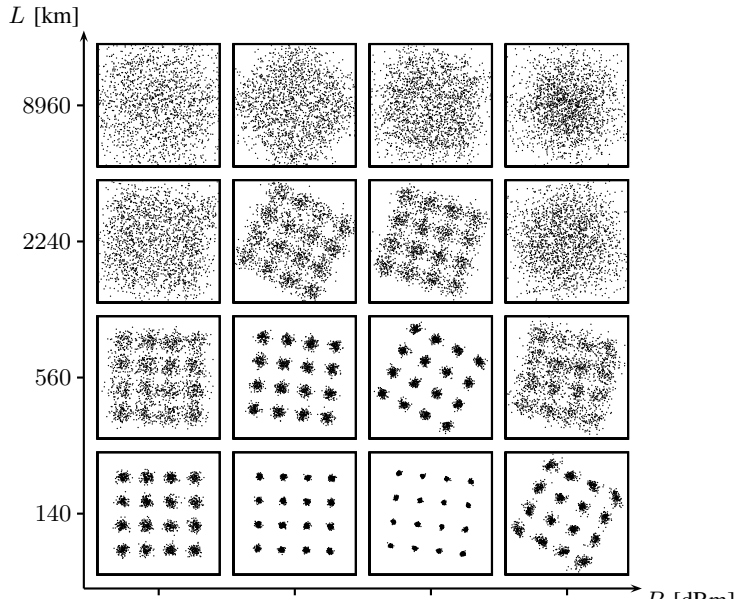


(a) Paper A



(b) Paper B

Figure 2.9: Comparison of scatter plots for different combinations of input power P and transmission length L.



(c) Paper C

Figure 2.9: (cont.)

backpropagation (DBP) can be employed to improve performance, which is based on the invertibility of the Schrödinger equation in the absence of noise. In [39, 40], the authors extend DBP via a factor-graph approach showing significant improvements over DBP for DM links in the nonlinear operating regime. In [41, 42], different detection approaches are compared, with a focus on MLSE. In [19], a new transmission scheme based on the nonlinear Fourier transform is proposed.

We have also ignored hardware imperfections, e.g., laser phase noise, which might potentially have a significant impact on the discrete-time channel model that should be used to design practical modulation and coding schemes.

CHAPTER 3

Bit-Interleaved Coded Modulation with Low-Density Parity-Check Codes

In this chapter, we provide a brief introduction on how to reliably transmit data at high spectral efficiencies. Spectrally efficient communication can be achieved in practice by combining forward error correction with higher-order signal constellations, which is commonly referred to as coded modulation (CM). We focus on bit-interleaved coded modulation (BICM), which is a pragmatic approach to CM and often implemented in practice, due to its inherent simplicity and flexibility. This chapter should be seen mainly as supplementary material for the problem statement that is addressed in Paper C, where it is assumed that the reader is somewhat familiar with coding theory and iterative decoding techniques.

We start by outlining the main principles behind coded modulation in Section 3.1. In Section 3.2, we explain the building blocks of a BICM system. In Section 3.3, we review some basic concepts behind LDPC codes and iterative decoding, focusing on protograph-based codes. We also briefly cover SC-LDPC codes, which are one of the code classes considered for the problem statement addressed in Paper C.

3.1 Introduction

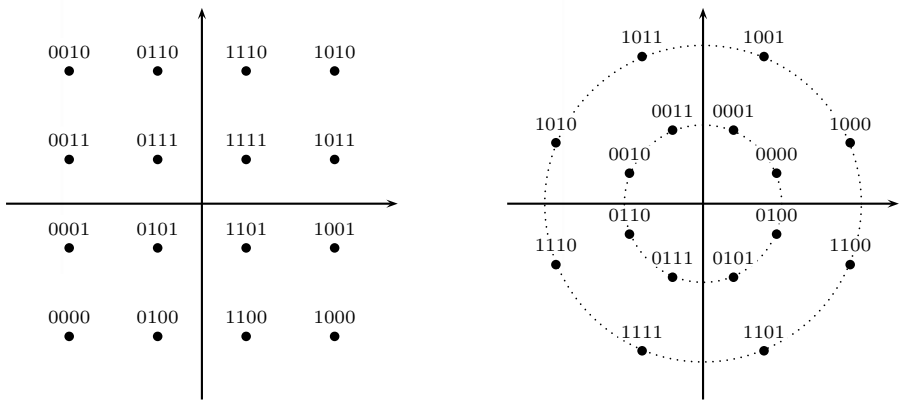
Consider again the discrete memoryless AWGN channel $y_k = x_k + n_k$ (see Section 2.1), where $x_k \in \mathbb{C}$ is the channel input at time instant k, y_k is the corresponding channel output, and n_k is the realization of a zero-mean, circularly symmetric complex Gaussian random variable N_k with $\mathbb{E}[|N_k|^2] = P_{\mathrm{N}}$. The channel input is assumed to satisfy the average power constraint $\mathbb{E}[|X_k|^2] = P$ and the SNR is given by $\mathrm{SNR} = P/P_{\mathrm{N}}$. The goal is to reliably transmit data at high spectral efficiencies over this channel. To do so, one can formally define an encoder $\mathcal{E} : \{0,1\}^d \to \mathcal{C}_{\mathrm{c}}$, which maps a vector of d information bits to a codeword in the code $\mathcal{C}_{\mathrm{c}} \subset \mathbb{C}^N$. Each codeword is a complex vector of length N and its components serve as the input for N consecutive uses of the AWGN channel. Similarly, one can define a decoder $\mathcal{D} : \mathbb{C}^N \to \{0,1\}^d$, which maps a vector of N channel outputs back to a sequence of d estimated bits. Assuming equally likely information bits, the communication rate (measured in [bits/complex symbol]) of such a system is given by $\kappa = \log_2(|\mathcal{C}_{\mathrm{c}}|)/N = d/N$. Notice that the communication rate of the discrete channel (in [bits/complex symbol]) is intimately related to the spectral efficiency of the continuous-time channel (in [bits/s/Hz]) via the bandwidth of the pulse shape $p(t)$ and the symbol rate. Shannon proved that all rates up to the channel capacity

$$C = \log_2(1 + \mathrm{SNR}) \tag{3.1}$$

are achievable, in the sense that there exists an encoder/decoder pair that can provide an arbitrarily small error probablitity as long as $N \to \infty$ [43].

While Shannon's proof provides communication engineers with an invaluable benchmark, the problem of designing practical encoders and decoders that operate close to the capacity and are implementable with reasonable complexity was not directly addressed by Shannon. In practical systems, the channel input x_k commonly does not take on arbitrary complex values, but is constrained to a discrete signal constellation $\mathcal{X} \subset \mathbb{C}$. Given this premise, it is useful to introduce a soft dividing line between two different operating regimes for this channel. This dividing line is at $\kappa = 2$, where $\kappa < 2$ is referred to as the power-limited regime and $\kappa > 2$ as the bandwidth-limited regime [15]. Roughly speaking, in the power-limited regime, it is sufficient to consider a binary modulation, independently in the real and imaginary part (e.g., Gray-labeled quadrature phase-shift keying (QPSK) according to $\mathcal{X} = \{1 + \jmath, 1 - \jmath, -1 + \jmath, -1 - \jmath\}$ and scaled by $\sqrt{P/2}$), in combination with binary error correction codes in order to operate close to the capacity. On the other hand, spectrally efficient communication, i.e., $\kappa > 2$, requires the use of signal constellations with cardinality larger than 4, which are referred to as higher-order[1] constellations. By invoking the capacity formula, it follows directly that operating at high spectral efficiencies $\kappa > 2$ requires the signal power to be at least three times the noise power. In other words, spectrally efficient communication requires a reasonably high SNR.

[1] One may also classify complex constellations with 4 points as "higher-order", as long as they cannot be viewed as two independent binary modulations per real and complex dimension.



(a) 16-QAM **(b)** (8,8)-APSK

Figure 3.1: Two examples of higher-order signal constellations with 16 points.

Devising practical encoder/decoder pairs where x_k is constrained to a higher-order signal constellation is commonly referred to as CM design. There exist several different approaches, for example trellis coded modulation (TCM) [44], CM with nonbinary codes [45], multilevel coded modulation (MLCM) [46], or BICM [47]. Our focus is on BICM in combination with (binary) LDPC codes, which is one of the most popular capacity-approaching coding schemes for achieving high spectral efficiency, due to its simplicity and flexibility [48]. BICM is employed as the de-facto standard in many wireless communication standards and has also been studied by many authors for fiber-optical communication systems, see, e.g., [49] or [50] and references therein.

3.2 BICM System Model

The transmitted symbols x_k in each time instant k are assumed to take on values from a discrete signal constellation $\mathcal{X} \subset \mathbb{C}$ with $|\mathcal{X}|$ points, where $|\mathcal{X}|$ is a power of two. Furthermore, each point in the constellation is assumed to be labeled with a unique binary string of length $m = \log_2 |\mathcal{X}|$, where $b_i(a)$, $1 \leq i \leq m$, denotes the ith bit in the binary string assigned to $a \in \mathcal{X}$ (counting from left to right). Two examples of signal constellations with $|\mathcal{X}| = 16$ points are shown in Fig. 3.1 and referred to as 16-quadrature amplitude modulation (QAM) and $(8, 8)$-amplitude phase-shift keying (APSK). For a detailed definition of APSK constellations, we refer the reader to Paper A.

In the following, we describe the main components of a BICM system. First, consider the block diagram shown in Fig. 3.2(a), where the modulo 2 addition of $d_{i,k}$ and multiplication by $\bar{d}_{i,k} = (-1)^{d_{i_k}}$ are explained further below and can be ignored for now. At each time instant, the modulator Φ takes m bits $b_{i,k}$, $1 \leq i \leq m$, and maps them to one

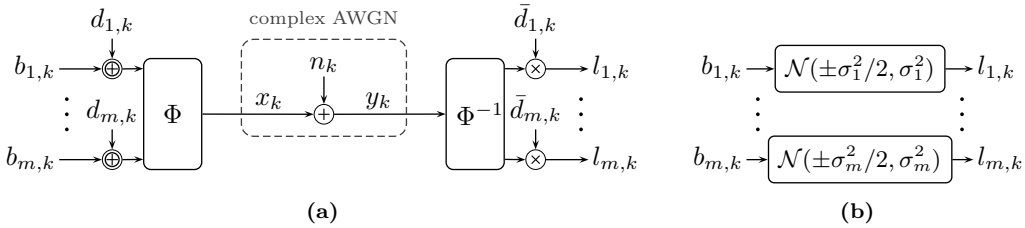


Figure 3.2: The modulator Φ, demodulator Φ^{-1}, and channel symmetrization technique in (a). A helpful approximate channel model via parallel symmetric Gaussian LLR channels in (b).

of the constellation points according to the binary labeling of the signal constellation. At the receiver, the demodulator Φ^{-1} computes soft reliability information about the transmitted bits in the form of the log-likelihood ratios (LLRs)

$$l_{i,k} \triangleq \log \frac{f_{Y_k|B_{i,k}}(y_k|0)}{f_{Y_k|B_{i,k}}(y_k|1)} = \log \frac{\sum_{x \in \mathcal{X}_{i,0}} f_{Y_k|X_k}(y_k|x)}{\sum_{x \in \mathcal{X}_{i,1}} f_{Y_k|X_k}(y_k|x)}, \tag{3.2}$$

where $\mathcal{X}_{i,u} \triangleq \{a \in \mathcal{X} : b_i(a) = u\}$ is the subconstellation where all points have the bit u at the ith position of their binary label. The LLR is a function of the observation, and, since the observation is a random variable, the LLR is also a random variable.

One way to interpret the setup depicted in Fig. 3.2(a) is as follows. The concatenation of the modulator Φ, the AWGN channel, and demodulator Φ^{-1} establishes a binary interface for the complex-valued AWGN channel. It is useful to imagine transmitting data over a set of m parallel binary-input continuous-output channels, or simply bit channels, where one may view the conditional distribution of the LLR $f_{L_{i,k}|B_{i,k}}(\cdot|\cdot)$, $1 \leq i \leq m$, as a bit channel. In the following, a bit channel $f_{L|B}(l|b)$ is called symmetric if $f_{L|B}(l|0) = f_{L|B}(-l|1)$ and referred to as an LLR channel if $f_{L|B}(l|0)e^l = f_{L|B}(l|1)$. The terminology here is used to emphasize that, if the second condition is fulfilled, the output of the channel corresponds to a "true" LLR. This is important because, in practice, low-complexity approximations of (3.2) are often considered, and the resulting bit channel in that case is not necessarily an LLR channel. One can show that $f_{L_{i,k}|B_{i,k}}(\cdot|\cdot)$ is an LLR channel. However, the channel is not necessarily symmetric in general.[2] Symmetry can be enforced by adding modulo 2 independent and identically distributed bits $d_{i,k}$ to the bits $b_{i,k}$ [51]. After the demodulator, the corresponding LLR is multiplied by $\bar{d}_{i,k} = (-1)^{d_{i,k}}$, which implies that the bits $d_{i,k}$ are known to both the transmitter and receiver. The resulting bit channel $f_{L_{i,k}|B_{i,k}}(\cdot|\cdot)$ can be shown to be symmetric.

We proceed by quantifying the quality of the m bit channels, where we rely on the mutual information (MI) as a measure of quality. The MI between the output of a

[2] The symmetry condition will become important when discussing density evolution and LDPC codes, where one relies on the all-zero codeword assumption.

symmetric LLR channel $f_{L|B}(l|b)$ and uniform input bits is given by

$$I(L;B) = \mathbb{E}\left[\log_2\left(\frac{f_{L|B}(L|B)}{f_L(L)}\right)\right] \tag{3.3}$$

$$= 1 - \mathbb{E}\left[\log_2\left(\frac{f_{L|B}(L|B) + f_{L|B}(L|1-B)}{f_{L|B}(L|B)}\right)\right] \tag{3.4}$$

$$= 1 - \mathbb{E}\left[\log_2\left(1 + \frac{f_{L|B}(L|1-B)}{f_{L|B}(L|B)}\right)\right] \tag{3.5}$$

$$= 1 - \mathbb{E}\left[\log_2\left(1 + \exp((-1)^{1-B}L)\right)\right] \tag{3.6}$$

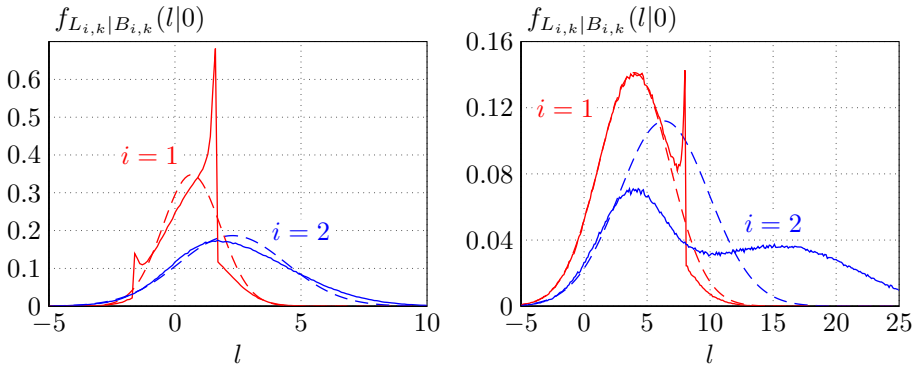
$$= 1 - \int_{-\infty}^{+\infty} f_{L|B}(l|0)\log_2(1+\exp(-l))\,\mathrm{d}l. \tag{3.7}$$

Writing the MI in the form (3.7) can be useful in order to compute the MI with the help of Monte Carlo integration.

It turns out that, while the channel quality of the bit channels can be determined quite efficiently, it is very difficult to find exact analytical expressions for the actual densities $f_{L_{i,k}|B_{i,k}}(\cdot\,|\,\cdot)$. A common approach in the analysis of BICM is to make the simplifying assumption that the densities $f_{L_{i,k}|B_{i,k}}(\cdot\,|\,\cdot)$ are approximately Gaussian. An LLR channel with a Gaussian density is particularly simple, because it can be parametrized by a single parameter. More precisely, we refer to a bit channel $f_{L|B}(l|b)$ as a symmetric Gaussian LLR channel with parameter σ^2 if $L \sim \mathcal{N}(\sigma^2/2, \sigma^2)$ conditioned on $B = 0$ and $L \sim \mathcal{N}(-\sigma^2/2, \sigma^2)$ conditioned on $B = 1$. The MI between the output of a symmetric Gaussian LLR channel and uniform input bits is denoted by $J(\sigma)$. Under the Gaussian assumption, a helpful approximation of the setup in Fig. 3.2(a) is shown in Fig. 3.2(b), where transmission takes place over m parallel symmetric Gaussian LLR channels with different parameters σ_i^2. In order to find a correspondence between the LLR channels $f_{L_{i,k}|B_{i,k}}(\cdot\,|\,\cdot)$ and the parameters σ_i^2, one may match the MI according to $J(\sigma_i) = I_i(\mathrm{SNR}) \Leftrightarrow \sigma_i^2 = J^{-1}(I_i(\mathrm{SNR}))^2$, where $I_i(\mathrm{SNR}) = I(B_{i,k}; L_{i,k})$ is independent of k.

While the parallel Gaussian model can be quite useful, one should, however, be aware of the inaccuracies of this simplified model. In particular, the bit channels are not independent as suggested in Fig. 3.2(b) and the true distribution of the LLRs is not Gaussian. To illustrate the latter inaccuracy, in Fig. 3.3, we compare the actual densities with the approximated Gaussian densities for two different SNRs for the first two bit positions of the 16-QAM constellation shown in Fig. 3.1(a).[3] The densities $f_{L_{i,k}|B_{i,k}}(\cdot\,|0)$ are estimated via histograms and shown by the solid lines, whereas the Gaussian densities are shown by the dashed lines. It can be seen that the actual densities are clearly non-Gaussian and the accuracy of the Gaussian approximation therefore depends on the application scenario. For the application in Paper C (predicting the iterative performance

[3]The third and fourth bit positions lead to identical distributions, due to the fact that 16-QAM with the shown labeling can be seen as a product constellation of two one-dimensional constellation.



(a) SNR = 3 dB **(b)** SNR = 10 dB

Figure 3.3: Comparison of the true LLR channels (including channel symmetrization) with the symmetric Gaussian LLR channels that have the same MI.

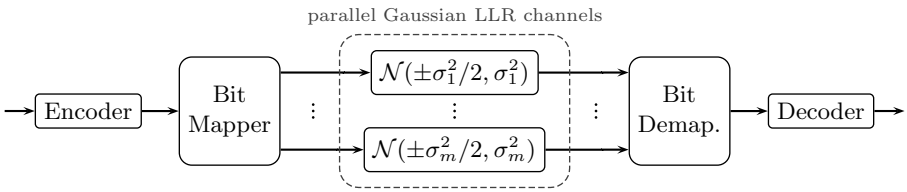


Figure 3.4: A useful approximate system model for BICM systems.

behavior of LDPC codes), the approximation is quite accurate and at the same time allows for a major simplification of the analysis, thereby justifying its use.

Consider now the case where we employ a single binary code $\mathcal{C} \subset \{0,1\}^n$ of length n, and each codeword is transmitted using $N = n/m$ symbols x_k. The allocation of the coded bits to the modulator (i.e., the different bit channels in Fig. 3.2(b)) is determined by a bit mapper as shown in Fig. 3.4. In Paper C, our goal is to find good bit mappers for a given code and signal constellation, where we focus on protograph-based LDPC codes.

As a side note, we remark that the term "bit interleaver" is also commonly used instead of "bit mapper". In fact, the modulator Φ is sometimes referred to as the (symbol) mapper (and the demodulator Φ^{-1} as the demapper), which the reader should be aware of in order to avoid confusion. However, the terms "bit mapper", "bit mapping", or "mapping" seem to be preferred in the literature when the allocation of the coded bits to the constellation symbols Φ is explicitly studied or optimized, see, e.g., [52,53]. Moreover, outside the context of BICM, the terms "mapping device" or "channel mapper" are used when studying parallel channels in combination with binary codes, e.g., in [54, 55].

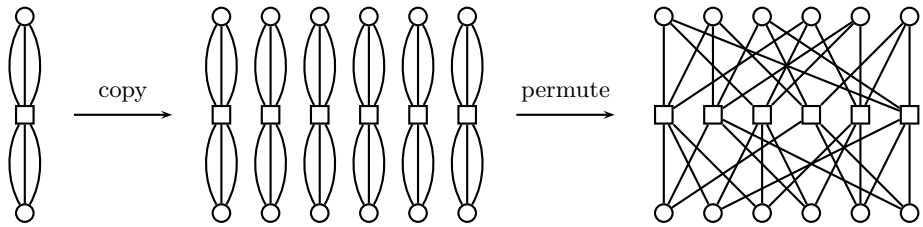


Figure 3.5: Illustration of the protograph lifting procedure for $\mathbf{P} = (3, 3)$ and $M = 6$.

3.3 Low-Density Parity-Check Codes

LDPC codes were proposed by Gallager in his PhD thesis [56]. They were conceived as practically decodable codes, able to "utilize the long block lengths necessary for low error probability without requiring excessive equipment or computation" [57]. Formally, a binary LDPC code $\mathcal{C}$ of length n is defined as the null space of a sparse parity-check matrix $\mathbf{H} = [h_{i,j}] \in \{0,1\}^{c \times n}$, i.e., $\mathcal{C} = \{\boldsymbol{c} \in \{0,1\}^n : \mathbf{H}\boldsymbol{c}^\intercal = \mathbf{0}\}$, where $n > c$ and operations are over the binary field. Assuming that $\mathbf{H}$ has full rank c, one can invoke the fundamental theorem of linear algebra to infer that the code has $|\mathcal{C}| = 2^d$ codewords, where $d = n - c$ is the dimension of the code. The code rate is defined as $R = d/n = 1 - c/n$.

3.3.1 Construction via Protographs

There exist different methods to construct "good" LDPC codes, i.e., good matrices $\mathbf{H}$, and one popular method is by using protographs [58]. The parity-check matrix of an LDPC code can be represented by using a bipartite Tanner graph consisting of n variable nodes (VNs) and c check nodes (CNs), where the ith CN is connected to the jth VN if $h_{i,j} = 1$. A protograph is also a "small" bipartite graph defined by an adjacency matrix $\mathbf{P} = [p_{i,j}] \in \mathbb{N}_0^{c' \times n'}$, called the base matrix. Given $\mathbf{P}$, a parity-check matrix $\mathbf{H}$ is obtained by replacing each entry $p_{i,j}$ in $\mathbf{P}$ with a random binary M-by-M matrix which contains $p_{i,j}$ ones in each row and column. This procedure is called lifting and $M \geq \max_{i,j} p_{i,j}$ is the lifting factor. Graphically, it amounts to copying the protograph M times and subsequently permuting edges, in order to obtain the Tanner graph. Parallel edges, i.e., for $p_{i,j} > 1$, are permitted in the protograph and are resolved in the lifting procedure. The design rate of the code is given by $R = 1 - c/n = 1 - c'/n'$, where $c = c'M$ and $n = n'M$. As an example, the lifting procedure for $\mathbf{P} = (3, 3)$ and $M = 6$ is illustrated in Fig. 3.5.

Designing codes via protographs has several practical advantages, e.g., a quasi-cyclic code construction is easily applied by constraining the M-by-M matrices to have a circulant structure which in turn allows for hardware-efficient implementation [59, p. 263], suitable for high-speed optical communications [6]. Moreover, codes of different lengths can be obtained simply by adjusting the lifting factor.

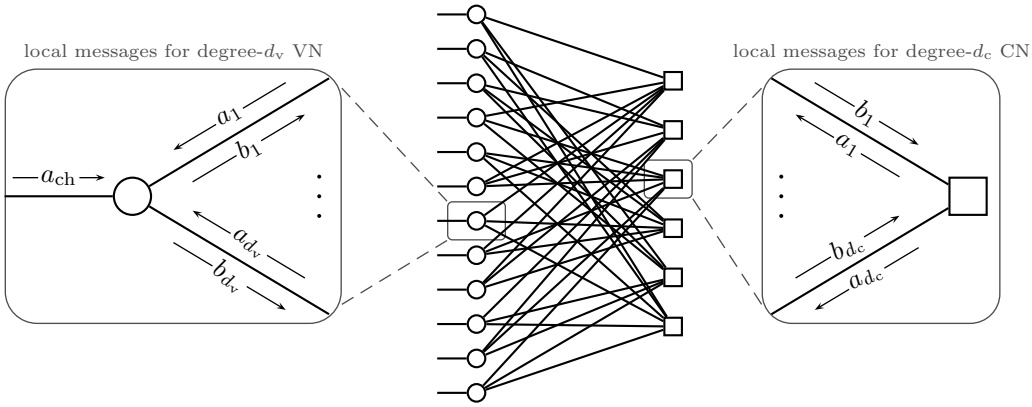


Figure 3.6: Illustration of the messages involved in the iterative BP decoding algorithm.

3.3.2 Iterative Belief Propagation Decoding

Consider the scenario where each bit in the codeword is transmitted over an LLR channel $f_{L|B}(\cdot\,|\,\cdot)$. The goal of the decoder is to recover the transmitted codeword based on the observation from the channel, which consists of n LLRs. These LLRs can be interpreted as the initial belief about each coded bit. During the decoding process, the decoder tries to iteratively improve the accuracy of the belief, by exchanging messages in the form of extrinsic LLRs between VNs and CNs along the edges of the Tanner graph.

For an excellent and comprehensive description of belief propagation (BP) decoding, we refer the reader to [59, Ch. 5.3]. Here, we will only briefly review the basic steps of the decoding algorithm. We use the following convention. Messages arriving at VNs are denoted by a, while messages emanating from VNs are denoted by b. For CNs, it is the other way around, i.e., arriving messages are denoted by b, while emanating messages by a. In an attempt to avoid cluttered notation, only one index is appended to a or b in order to *locally* distinguish between messages along different edges for the same node. The corresponding picture we have in mind is illustrated in Fig. 3.6. By locally we mean that, for example, the message b_1 emanating from the magnified VN does not correspond to the message b_1 arriving at the magnified CN. (In fact, from the way the figure is drawn, the message b_1 arriving at the magnified CN would emanate from the fourth VN, counting from the top.)

Consider now an arbitrary VN of degree d_v, where the degree of a VN corresponds to the number of CNs that are connected to it. There are $d_v + 1$ messages arriving at this VN, where $a_1, \ldots, a_{d_v}$ are messages from CNs and a_{ch} corresponds to a channel LLR. The d_v outgoing messages $b_1, \ldots, b_{d_v}$ are computed according to

$$b_i = \sum_{\sim i} a_j + a_{ch}, \tag{3.8}$$

where the summation is over the index set $j \in \{1, \ldots, d_\mathrm{v}\}$ *excluding* the index i. Similarly, if we consider an arbitrary CN of degree d_c, there are d_c messages $b_1, \ldots, b_{d_\mathrm{c}}$ arriving and the outgoing messages are computed according to

$$a_i = 2 \tanh^{-1} \left(\prod_{\sim i} \tanh(b_j/2) \right), \tag{3.9}$$

where the product is over the index set $j \in \{1, \ldots, d_\mathrm{c}\}$ excluding the index i. Since the CN operation (3.9) is central in the analysis of LDPC codes under iterative decoding, it is very common to rewrite it in terms of the binary boxplus operator defined as

$$b_1 \boxplus b_2 = 2 \tanh^{-1} \left(\tanh(b_1/2) \tanh(b_2/2) \right), \tag{3.10}$$

where the box-addition of an arbitrary number of terms is evaluated by recursively applying (3.10), e.g., $b_1 \boxplus b_2 \boxplus b_3 = (b_1 \boxplus b_2) \boxplus b_3$. With this convention, one can write the CN operation more concisely as

$$a_i = \boxplus_{\sim i} b_j. \tag{3.11}$$

The decoding process can now be described as follows. Initialize a_ch for all VNs to the corresponding channel LLR, and set all other messages to 0. Then, repeat the following two steps. First, compute outgoing messages for all VNs according to (3.8). After that, compute the outgoing messages for all CNs according to (3.9). Stop if either a maximum number of iterations has been reached, or the proper combination of the hard decisions on the messages

$$\sum_{j=1}^{d_\mathrm{v}} a_j + a_\mathrm{ch} \tag{3.12}$$

for all VNs forms a valid codeword.

3.3.3 Density Evolution

Density evolution (DE) is a powerful tool to analyze the iterative decoding behavior and performance of LDPC codes [60]. DE mimics the decoding process under a cycle-free graph assumption by tracking how the densities of the messages evolve with iterations. DE is commonly used to find so-called decoding thresholds, which can be interpreted as the capacity for LDPC codes under BP decoding. Similar to the channel capacity, the threshold divides the channel quality parameter range (for example the parameter σ^2 of a symmetric Gaussian LLR channel) into a region where reliable decoding is possible and where it is not.

The main steps in the DE algorithm can be understood by considering the update equations for the VNs (3.8) and CNs (3.9). If we assume that the involved incoming messages

are random variables, then they have a certain probability distribution or density (for example, a_{ch} is distributed according to LLR channel). The main question is, how can we obtain the densities of the outgoing messages? For the VN update, the answer turns out to be a simple convolution. In particular, for two independent random variables A and B with distributions $f_A(a)$ and $f_B(b)$, their sum $C = A + B$ is distributed according to $f_C(c) = f_A(a) \circledast f_B(b)$. It is convenient to introduce the short notation $\mathsf{a} \circledast \mathsf{b}$, where a and b are placeholders for the densities of the random variables A and B [61]. With this notation, the densities of the outgoing messages, given the densities of the incoming messages, can be computed according to

$$\mathsf{b}_i = \mathop{\circledast}_{\sim i} \mathsf{a}_j \circledast \mathsf{a}_{\mathrm{ch}}. \tag{3.13}$$

For the CN update, it is somewhat more challenging to obtain the densities of the outgoing messages. The most straightforward approach is by using Monte Carlo techniques and histograms. Consider the case where two messages b_1 and b_2 with densities b_1 and b_2 are processed according to the boxplus operation $a = b_1 \boxplus b_2$. In order to obtain the density a, one can simply generate many independent realizations of the random variables B_1 and B_2, perform the boxplus operation, and collect the resulting samples. These samples can be seen as a particle representation of the density a. This method is illustrated in Fig. 3.7, where it is shown how two consistent Gaussian densities "evolve" under the boxplus operation. A density a is called a consistent Gaussian density[4] with parameter σ^2, if $A \sim \mathcal{N}(\sigma^2/2, \sigma^2)$. As a short notation, one may introduce the operator $\mathsf{a} = \mathsf{b}_1 \boxtimes \mathsf{b}_2$, referred to as box-convolution [61]. In practice, the box-convolution of two densities can be implemented by using a look-up table approach [62]. Similar to (3.14), the densities of the outgoing CN messages can then be computed according to

$$\mathsf{a}_i = \boxed{\mathop{\circledast}_{\sim i}} \mathsf{b}_j. \tag{3.14}$$

For protograph-based codes, DE can be used to analyze the iterative decoding behavior by tracking one density for each edge in the protograph. This asserts that the messages exchanged during the decoding process over edges belonging to the same edge-type (defined by one protograph edge) have the same density. Assume that the transmission takes place over a symmetric LLR channel with a fixed channel quality and we wish to predict the iterative decoding behavior. Due to the channel symmetry, one may assume the transmission of the all-zero codeword [59, p. 389]. Start by initializing a_{ch} for all VNs in the protograph to $f_{L|B}(l|0)$ and set all other densities to $\delta(l)$. Then, repeat the following two steps. First, calculate the outgoing message densities for all VNs in the protograph according to (3.13). After that, calculate the outgoing message densities for all CNs in the protograph according to (3.14). Stop if the error probability associated

[4] Note that the conditional distribution $f_{L|B}(l|0)$ of a symmetric Gaussian LLR channel corresponds to a consistent Gaussian density.

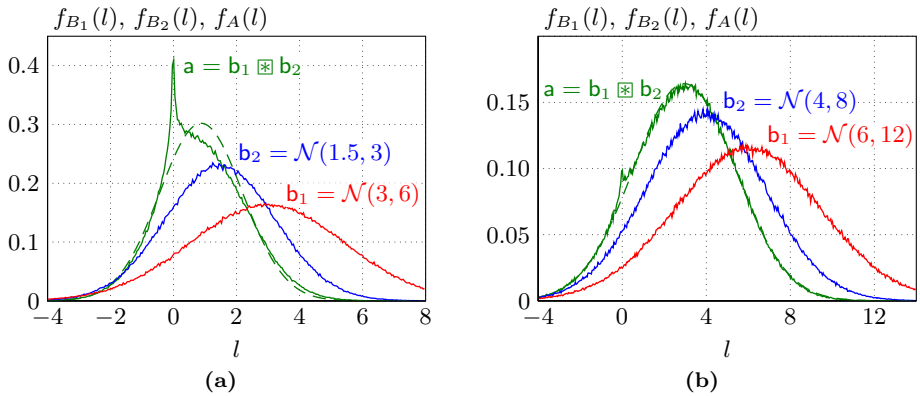


Figure 3.7: Illustration of the box-convolution of two consistent Gaussian densities. The green dashed line corresponds to the consistent Gaussian approximation obtained via EXIT functions.

with the density

$$\overset{d_v}{\underset{j=1}{\circledast}} \mathsf{a}_j \circledast \mathsf{a}_{\text{ch}} \tag{3.15}$$

for each VN is below a certain target bit error probability (successful decoding), where the error probability associated with a density a is given by

$$p_e(\mathsf{a}) = \int_{-\infty}^{0} f_A(a)\,\mathrm{d}a, \tag{3.16}$$

or a maximum number of iterations is reached (decoding failure). In order to find the decoding threshold, the above procedure is repeated many times for decreasing channel quality until the decoding fails, starting from a channel quality where the decoding is successful.

Approximate Density Evolution via EXIT Functions

Tracking the full densities (or quantized densities in practice) is computationally demanding and extrinsic information transfer (EXIT) functions are usually considered to be a good compromise between computational efficiency and accuracy [63]. Let us assume that the density a fulfills the condition $f_A(a)e^a = f_A(-a)$. Then, the density can be associated with the MI measure

$$I(\mathsf{a}) = 1 - \int_{-\infty}^{\infty} f_A(a)\log_2(1 + e^{-a})\,\mathrm{d}a. \tag{3.17}$$

Now, instead of tracking the evolution of densities, one may track the evolution of the MI measure associated with the densities (which is just a scalar value for each density). Let us assert that, under the VN operation, this measure evolves approximately according to

$$I(\mathsf{b}_i) \approx \tilde{J}\left(\sum_{\sim i} \tilde{J}^{-1}(I(\mathsf{a}_j)) + \tilde{J}^{-1}(I(\mathsf{a}_{\mathrm{ch}}))\right), \tag{3.18}$$

whereas, under the CN operation it evolves approximately according to

$$I(\mathsf{a}_i) \approx 1 - \tilde{J}\left(\sum_{\sim i} \tilde{J}^{-1}(1 - I(\mathsf{b}_j))\right), \tag{3.19}$$

where $\tilde{J}(x) = J(\sqrt{x})$. These two equations can be motivated as follows. Eq. (3.18) is exact under the assumption that all incoming densities $\mathsf{a}_1, \ldots, \mathsf{a}_{d_\mathrm{v}}$, and a_{ch} are consistent Gaussian densities. To see this, note that the convolution of two consistent Gaussian densities with parameters σ_1^2 and σ_2^2 is another consistent Gaussian density with parameter $(\sigma_1^2 + \sigma_2^2)/2$. Furthermore, if a is a consistent Gaussian density with parameter σ^2, the operation $\tilde{J}^{-1}(I(\mathsf{a}))$ simply returns σ^2. Without going into details, (3.19) can be heuristically motivated by a duality property that holds for the binary erasure channel (BEC) [59, p. 415]. It is important to point out that (3.19) it is not exact, even if all incoming densities are consistent Gaussians, but it turns out to be surprisingly accurate nonetheless. For example, the green dashed lines in Fig. 3.7 have been obtained using (3.19), where the resulting MI measure is plotted in the form of the associated consistent Gaussian density.

3.3.4 Spatially Coupled LDPC Codes

Spatial coupling of regular[5] LDPC codes has emerged as a powerful technique to construct capacity-achieving codes for a large class of channels using iterative BP decoding [13,64]. The main idea is to make several copies of the Tanner graph that defines the regular base code, arrange the copies next to each other, and then interconnect neighboring graphs in a particular way. The key to the outstanding performance of codes constructed in such a way is a boundary effect due to slight irregularities at the two ends of the resulting Tanner graph.

In general, SC-LDPC codes have parity-check matrices with a band-diagonal structure, see, e.g., [64] for a formal definition. Here, we briefly introduce their construction via protographs [65], [66, Sec. II-B]. The base matrix $\mathbf{P}_{[T]}$ of a (J, K) regular, protograph-based SC-LDPC code with termination length T can be constructed by specifying matrices $\mathbf{P}_i$, $0 \leq i \leq m_\mathrm{s} + 1$ of dimension J' by K', where m_s is referred to as the memory. The matrices are such that $\mathbf{P} = \sum_i \mathbf{P}_i$ has column weight J and row weight K for all columns and rows, respectively. Given T and the matrices $\mathbf{P}_i$, the base matrix $\mathbf{P}_{[T]}$

[5] An LDPC codes is called regular if all VNs have degree d_v and all CNs have degree d_c.

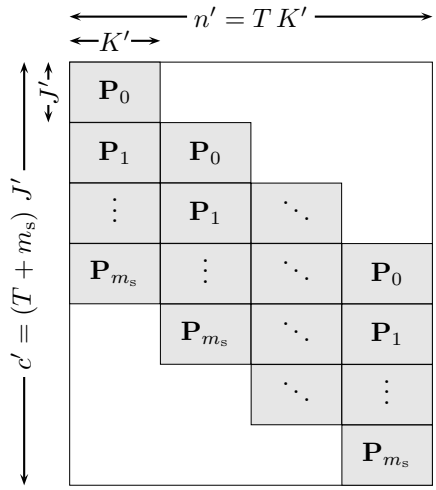


Figure 3.8: Illustration of the base matrix $\mathbf{P}_{[T]}$ of a (J, K) regular, protograph-based SC-LDPC code.

is constructed as shown in Fig. 3.8. From the dimensions of $\mathbf{P}_{[T]}$ one can infer a design rate of $R(T) = 1 - (T + m_\mathrm{s})J'/(TK')$. As T grows large, the rate approaches $R(\infty) = 1 - J'/K'$.

Before continuing, it is insightful to recall the following statement from [67], where the design of irregular LDPC codes is studied. (VNs are referred to as message nodes.)

> "[...] we offer some intuition as to why irregular graphs prove useful. [...] Message nodes with high degree tend to their correct value quickly. These nodes then provide good information to the check nodes, which subsequently provide better information to lower degree message nodes. Irregular graph constructions thus *lead to a wave effect*, where high degree message nodes tend to get corrected first, and then message nodes with slightly smaller degree, and so on down the line." [emphasis added]

For SC-LDPC codes, one can give a similar heuristic explanation for their outstanding performance as follows (see [64] for a detailed explanation). By inspecting the structure of the base matrix in Fig. 3.8, one may verify that the CN degrees corresponding to the first and last couple of rows is lower than the CN degrees corresponding to the rows in between. The lower degree CNs lead to a locally better decoding capability which helps decoding neighboring VNs. This local boundary effect turns out to initiate a wave-like behavior and can have a global effect on the decoding capability of the entire code with increasing number of decoding iterations. To illustrate this behavior, in Fig. 3.9, we show the predicted bit error rates p_e via (approximate) DE for the coded bits corresponding to the jth column of the SC-LDPC protograph $\mathbf{P}_{[T]}$ with component matrices $\mathbf{P}_1 = \mathbf{P}_2 = \mathbf{P}_3 = (1, 1)$ and $T = 100$. We assume transmission over a symmetric

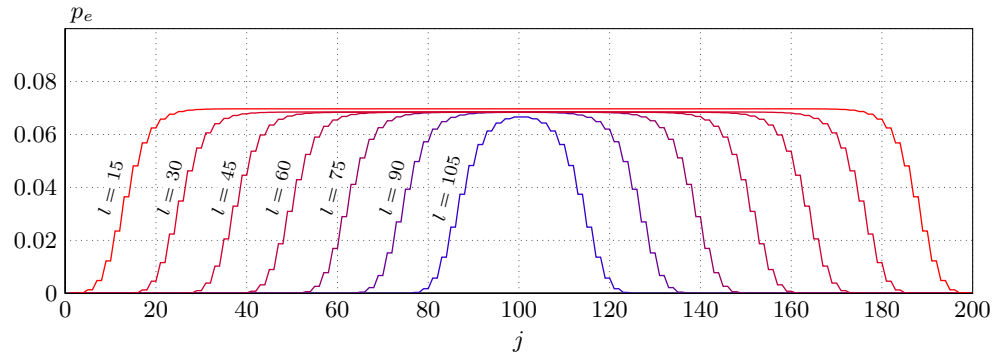


Figure 3.9: Illustration of the wave-like decoding behavior of SC-LDPC codes.

Gaussian LLR channel with parameter $\sigma^2 = 4$. In the figure, l denotes the iteration number. It can be observed that the error probability of the VNs at the two ends of the graph converges to zero after 15 iterations. Due to the spatial coupling, this boundary effect propagates inwards all the way to the center of the protograph in a wave-like fashion.

An important reason for the tremendous interest in spatially coupled codes is their universality. While irregular LDPC codes have been optimized for various communication channels, the degree distribution pairs that achieve the best performance usually vary from channel to channel [68]. In contrast, SC-LDPC codes derived from simple regular codes have been shown to universally achieve capacity for a variety of channels. However, there are also many open research problems concerning the practical implementation of SC-LDPC, see [69] for a recent overview. For example, the price to pay for the wave-like decoding behavior is a rate loss with respect to regular, uncoupled codes that are defined by the protograph $\boldsymbol{B} = \sum_i \boldsymbol{B}_i$.

CHAPTER 4

Conclusions and Future Work

In this chapter, we summarize the main conclusions from the appended papers and outline some potentially interesting ideas for future work.

Paper A

In Paper A, we study the design of APSK signal constellations under the assumption of a memoryless fiber-optical channel model with NLPN as the main system impairment. Optimized APSK constellations can offer significant performance advantages over conventional QAM constellations for the assumed channel model and detection scheme. It is also shown that the optimization of the signal constellation in the presence of severe nonlinear distortions can lead to somewhat counterintuitive results in the form of sacrificial points or sacrificial rings. As outlined in [70], these effects become particularly important when studying the channel capacity for such channels. Furthermore, when the bit error probability (BEP) is taken as a performance measure, it is important to consider the joint design of both the constellation and the labeling. In particular, an optimized constellation for symbol error probability (SEP) with an optimal labeling does not necessarily provide the best BEP. In fact, more structured constellations such as the considered rectangular APSK constellations may give better performance due to the possibility of Gray-like labeling methods.

An interesting direction for future work would be to replace the relevant measures for uncoded transmission (SEP and BEP) with the relevant measures for coded transmission (MI and generalized mutual information (GMI)). It may also be rewarding to find an accurate approximate characterizations of the channel PDF for the considered channel

model which may simplify the optimization procedure in this case.

Paper B

In Paper B, we study the design of a low-complexity detector for a memoryless fiber-optical channel assuming PM transmission. The detector uses a phase compensation scheme based on the received signal amplitudes in both polarizations, followed by a subsequent threshold detection. The complexity can be significantly reduced compared to a four-dimensional ML detector, albeit at some performance loss.

An important conclusion for the corresponding detector for SP transmission is that a nonlinear phase compensation scheme and subsequent threshold detection is equivalent to the ML detector. Unfortunately, a similar conclusion does not hold for the proposed detector for PM signals and it would be interesting to study the design of an improved detector that has this property (or show that this is fundamentally not possible). Furthermore, the detector in Paper B is limited to PM-M-PSK constellations and similar detection schemes for constellations with multiple amplitude levels could be investigated.

Paper C

In Paper C, we study a coded transmission system that operates over a fiber-optical link without inline dispersion compensation. Assuming a linear coherent receiver, the classical AWGN channel with a modified SNR expression is used as a design channel. We propose a method to optimize the bit mapper that determines the allocation of the coded bits from the FEC encoder to the labeling bits of the signal constellation that is applicable to any protograph-based LDPC code. Protograph-based codes are particularly interesting for fiber-optical systems because they allow for an efficient hardware implementation. We also extend the technique to SC-LDPC codes using a windowed decoder. The results show that by using an optimized bit mapper, the transmission reach can be extended by up to 8%, with almost no added system complexity.

An interesting direction for future work could be to study the bit mapper optimization assuming hard-decision decoding. For example, Smith and Kschischang have proposed such a setup in [49] using staircase codes for BICM with an additional shaping unit. An appropriate model in this case would be to study parallel binary symmetric channels with different crossover probabilities. To the best of our knowledge, bit mapper optimization for such a scenario has not yet been considered in the literature.

Bibliography

[1] K. F. Rauscher, "ROGUCCI study final report," IEEE Communications Society, Tech. Rep., 2010.

[2] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[3] E. Agrell and M. Karlsson, "Satellite constellations: Towards the nonlinear channel capacity," in *Proc. IEEE Photon. Conf.*, Burlingame, CA, Sep. 2012.

[4] A. D. Ellis, J. Zhao, and D. Cotter, "Approaching the non-linear Shannon limit," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 423–433, Feb. 2010.

[5] B. P. Smith and F. R. Kschischang, "Future prospects for FEC in fiber-optic communications," *IEEE J. Sel. Topics. Quantum Electron.*, vol. 16, no. 5, pp. 1245–1257, Oct. 2010.

[6] L. Schmalen, A. J. de Lind van Wijngaarden, and S. ten Brink, "Forward error correction in optical core and optical access networks," *Bell Labs Tech. J*, vol. 18, no. 3, pp. 39–66, Mar. 2013.

[7] L. Beygi, E. Agrell, J. M. Kahn, and M. Karlsson, "Coded modulation for fiber-optic networks," *IEEE Signal Processing Mag.*, vol. 31, no. 2, pp. 93–103, Mar. 2014.

[8] I. Djordjevic, W. Ryan, and B. Vasic, *Coding for Optical Channels*. Springer, 2010.

[9] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed. Academic Press, 2006.

[10] K.-P. Ho, *Phase-modulated Optical Communication Systems*. Springer, 2005.

[11] L. Hanzo, W. Webb, and T. Keller, *Single- and Multi-carrier Quadrature Amplitude Modulation: Principles and Applications for Personal Communications, WLANs and Broadcasting.* Wiley, 2000.

[12] A. J. Felström and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2181–2191, Sep. 1999.

[13] M. Lentmaier, A. Sridharan, K. S. Zigangirov, and D. J. Costello, Jr., "Terminated LDPC convolutional codes with thresholds close to capacity," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005.

[14] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Cambridge, MA, Jul. 2012.

[15] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.

[16] A. Lapidoth, *A Foundation in Digital Communication.* Cambridge University Press, 2009.

[17] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 18, no. 3, pp. 363–378, May 1972.

[18] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. 22, no. 5, pp. 624–636, May 1974.

[19] M. I. Yousefi, "Information transmission using the nonlinear Fourier transform," Ph.D. dissertation, University of Toronto, 2013.

[20] M. Secondini and E. Forestieri, "The nonlinear Schrödinger equation in fiber-optic systems," *Riv. Mat. Univ. Parma*, vol. 8, pp. 69–98, Sep. 2008.

[21] O. V. Sinkin, R. Holzlohner, J. Zweck, and C. R. Menyuk, "Optimization of the split-step Fourier method in modeling optical-fiber communications systems," *J. Lightw. Technol.*, vol. 21, no. 1, pp. 61–68, Jan. 2003.

[22] G. P. Agrawal, *Fiber-optic Communication Systems*, 4th ed. Wiley-Interscience, 2010.

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

[24] T. J. Ellingham, J. D. Ania-Castanón, R. Ibbotson, X. Chen, L. Zhang, and S. K. Turitsyn, "Quasi-lossless optical links for broad-band transmission and data processing," *IEEE Photon. Technol. Lett.*, vol. 18, pp. 268–270, Jan. 2006.

[25] M. I. Yousefi and F. R. Kschischang, "A Fokker-Planck differential equation approach for the zero-dispersion optical fiber channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Austin, TX, Jun. 2010.

[26] ——, "On the per-sample capacity of nondispersive optical fibers," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7522–7541, Nov. 2011.

[27] J. P. Gordon and L. F. Mollenauer, "Phase noise in photonic communications systems using linear amplifiers," *Opt. Lett.*, vol. 15, no. 23, pp. 1351–1353, Dec. 1990.

[28] A. Mecozzi, "Probability density functions of the nonlinear phase noise," *Opt. Lett.*, vol. 29, no. 7, pp. 673–675, Apr. 2004.

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

[30] K.-P. Ho, "Probability density of nonlinear phase noise," *J. Opt. Soc. Am. B*, vol. 20, no. 9, pp. 1875–1879, Sep. 2003.

[31] M. I. Yousefi and F. R. Kschischang, "A probabilistic model for optical fiber channels with zero dispersion," in *Proc. 25th Biennial Symp. on Communications*, Kingston, ON, May 2010.

[32] ——, "The per-sample capacity of zero-dispersion optical fibers," in *Proc. 12th Canadian Workshop on Information Theory*, Kelowna, BC, May 2011.

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

[34] L. Beygi, "Channel-aware multilevel coded modulation for coherent fiber-optic communications," Ph.D. dissertation, Chalmers University of Technology, 2013.

[35] L. Beygi, E. Agrell, P. Johannisson, M. Karlsson, and H. Wymeersch, "A discrete-time model for uncompensated single-channel fiber-optical links," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3440–3450, Nov. 2012.

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

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

[38] P. Johannisson, "Analytical modeling of nonlinear propagation in a strongly dispersive optical communication system," *arXiv:1205.2193v2 [physics.optics]*, May 2012. [Online]. Available: http://arxiv.org/abs/1205.2193

[39] N. V. Irukulapati, H. Wymeersch, P. Johannisson, and E. Agrell, "Extending backpropagation to account for noise," in *Proc. European Conf. Optical Communication (ECOC)*, London, UK, 2013, p. We.3.C.4.

[40] N. V. Irukulapati, "On nonlinear compensation techniques for coherent fiber-optical channel," Licentiate Thesis, Chalmers University of Technology, 2014.

[41] D. Marsella, M. Secondini, E. Forestieri, R. Magri, G. Moruzzi, and I. Pisa, "Detection strategies in the presence of fiber nonlinear effects," in *Proc. European Conf. Optical Communication (ECOC)*, Amsterdam, NL, 2012, p. P4.06.

[42] D. Marsella, M. Secondini, and E. Forestieri, "Maximum likelihood sequence detection for mitigating nonlinear effects," *J. Lightw. Technol.*, vol. 32, no. 5, pp. 908–916, Mar. 2014.

[43] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[44] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[45] A. Bennatan and D. Burshtein, "Design and analysis of nonbinary ldpc codes for arbitrary discrete-memoryless channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 549–583, Feb. 2006.

[46] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[47] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[48] D. J. Costello and G. D. Forney, Jr., "Channel coding: The road to channel capacity," *Proc. IEEE*, vol. 95, no. 6, pp. 1150–1177, Jun. 2007.

[49] B. Smith and F. R. Kschischang, "A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications," *J. Lightw. Technol.*, vol. 30, no. 13, pp. 2047–2053, Jul. 2012.

[50] I. B. Djordjevic, M. Arabaci, and L. L. Minkov, "Next generation FEC for high-capacity communication in optical transport networks," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3518–3530, Aug. 2009.

[51] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2141–2155, Sep. 2003.

[52] G. Richter, A. Hof, and M. Bossert, "On the mapping of low-density parity-check codes for bit-interleaved coded modulation," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Nice, Italy, Jun. 2007.

[53] T. Cheng, K. Peng, J. Song, and K. Yan, "EXIT-aided bit mapping design for LDPC coded modulation with APSK constellations," *IEEE Commun. Lett.*, vol. 16, no. 6, pp. 777–780, Jun. 2012.

[54] R. Liu, P. Spasojevic, and E. Soljanin, "Reliable channel regions for good binary codes transmitted over parallel channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1405–1424, Apr. 2006.

[55] I. Sason and I. Goldenberg, "Coding for parallel channels: Gallager bounds and applications to turbo-like codes," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2394–2428, Jul. 2007.

[56] R. Gallager, "Low-density parity-check codes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, 1963.

[57] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.

[58] J. Thorpe, "Low-density parity-check (LDPC) codes constructed from protographs," *IPN Progress Report 42-154, JPL*, 2005.

[59] W. Ryan and S. Lin, *Channel Codes Classical and Modern.* Cambridge University Press, 2009.

[60] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[61] T. T. Richardson and R. Urbanke, *Modern Coding Theory.* Cambridge University Press, 2008.

[62] Sae-Young Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 db of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, no. 2, pp. 58–60, Feb. 2001.

[63] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[64] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.

[65] D. G. M. Mitchell, M. Lentmaier, and D. J. Costello, Jr., "AWGN channel analysis of terminated LDPC convolutional codes," *Proc. Information Theory and Applications Workshop (ITA)*, 2011.

[66] A. R. Iyengar, M. Papaleo, P. H. Siegel, J. K. Wolf, A. Vanelli-coralli, and G. E. Corazza, "Windowed decoding of protograph-based LDPC convolutional codes over erasure channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2303–2320, Apr. 2012.

[67] M. Luby, M. Mitzenmacher, A. Shokrollah, and D. Spielman, "Analysis of low density codes and improved designs using irregular graphs," in *Proc. 30th Annual ACM Symp. on Theory of Computing (STOC)*, New York, USA, 1998.

[68] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[69] D. J. Costello, L. Dolecek, T. E. Fuja, J. Kliewer, D. G. M. Mitchell, and R. Smarandache, "Spatially coupled sparse codes on graphs - theory and practice," *arXiv:1310.3724v1 [cs.IT]*, Oct. 2013. [Online]. Available: http://arxiv.org/abs/1310.3724

[70] E. Agrell, "On monotonic capacity–cost functions," *arXiv:1209.2820v1 [cs.IT]*, Sep. 2012. [Online]. Available: http://arxiv.org/abs/1209.2820