

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

---

Analysis and Design of Spatially-Coupled  
Codes with Application to Fiber-Optical  
Communications

CHRISTIAN HÄGER



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Communication Systems Group  
Department of Signals and Systems  
Chalmers University of Technology  
Göteborg, Sweden, 2016

# **Analysis and Design of Spatially-Coupled Codes with Application to Fiber-Optical Communications**

CHRISTIAN HÄGER

Copyright © 2016 CHRISTIAN HÄGER, except where  
otherwise stated. All rights reserved.

ISBN 978-91-7597-351-7

Doktorsavhandlingar vid Chalmers tekniska högskola

Series No. 4032

ISSN 0346-718X

This thesis has been prepared using  $\text{\LaTeX}$  and PSTricks.

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

Printed by Chalmers Reproservice  
Göteborg, Sweden, May 2016

## Abstract

The theme of this thesis is the analysis and design of error-correcting codes that are suitable for high-speed fiber-optical communication systems. In particular, we consider two code classes. The codes in the first class are protograph-based low-density parity-check (LDPC) codes which are decoded using iterative soft-decision decoding. The codes in the second class are generalized LDPC codes with degree-2 variable nodes—henceforth referred to as generalized product codes (GPCs)—which are decoded using iterative bounded-distance decoding (BDD). Within each class, our focus is primarily on spatially-coupled codes. Spatially-coupled codes possess a convolutional structure and are characterized by a wave-like decoding behavior caused by a termination boundary effect. The contributions of this thesis can then be categorized into two topics, as outlined below.

First, we consider the design of systems 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 system that is based on the bit-interleaved coded modulation paradigm. As an example, for the (protograph-based) AR4JA code family, the transmission reach can be extended by roughly up to 8% by using an optimized bit mapper, without significantly increasing the system complexity. For terminated spatially-coupled codes with long spatial length, the bit mapper optimization only results in marginal performance improvements, suggesting that a sequential allocation is close to optimal. On the other hand, an optimized allocation can significantly improve the performance of tail-biting spatially-coupled codes which do not possess an inherent termination boundary. In this case, the unequal error protection offered by the modulation bits of a nonbinary signal constellation can be exploited to create an artificial termination boundary that induces a wave-like decoding for tail-biting spatially-coupled codes.

As a second topic, we study deterministically constructed GPCs. GPCs are particularly suited for high-speed applications such as optical communications due to the significantly reduced decoding complexity of iterative BDD compared to iterative soft-decision decoding of LDPC codes. We propose a code construction for GPCs which is sufficiently general to recover several well-known classes of GPCs as special cases, e.g., irregular product codes (PCs), block-wise braided codes, and staircase codes. Assuming transmission over the binary erasure channel, it is shown that the asymptotic performance of the resulting codes can be analyzed by means of a recursive density evolution (DE) equation. The DE analysis is then applied to study three different classes of GPCs: spatially-coupled PCs, symmetric GPCs, and GPCs based on component code mixtures.

**Keywords:** Bit-interleaved coded modulation, bit mapper, bounded-distance decoding, braided codes, density evolution, generalized low-density parity-check codes, generalized product codes, spatial coupling, staircase codes.



# List of Publications

This thesis is based on the following publications:

## Paper A

C. Häger, A. Graell i Amat, F. Brännström, A. Alvarado, and E. Agrell, “Improving Soft FEC Performance for Higher-Order Modulations via Optimized Bit Channel Mappings,” *Optics Express*, vol. 22, no. 12, pp. 14544–14558, June 2014.

## Paper B

C. Häger, A. Graell i Amat, A. Alvarado, F. Brännström, and E. Agrell, “Terminated and Tailbiting Spatially Coupled LDPC Codes with Optimized Bit Mapping for Spectrally Efficient Fiber-Optical Systems,” (invited paper) *J. Lightw. Technol.*, vol. 33, no. 7, pp. 1275–1285, April 2015.

## Paper C

C. Häger, A. Graell i Amat, H. D. Pfister, A. Alvarado, F. Brännström, and E. Agrell, “On Parameter Optimization for Staircase Codes,” in *Proc. Optical Fiber Communication Conf. and Exposition (OFC)*, Los Angeles, CA, 2015.

## Paper D

C. Häger, H. D. Pfister, A. Graell i Amat, and F. Brännström, “Density Evolution for Deterministic Generalized Product Codes on the Binary Erasure Channel,” submitted to *IEEE Trans. Inf. Theory*.

## Paper E

C. Häger, H. D. Pfister, A. Graell i Amat, and F. Brännström, “Density Evolution and Error Floor Analysis for Staircase and Braided Codes,” in *Proc. Optical Fiber Communication Conf. and Exposition (OFC)*, Los Angeles, CA, 2016.

## Paper F

C. Häger, H. D. Pfister, A. Graell i Amat, and F. Brännström, “Deterministic and Ensemble-Based Spatially-Coupled Product Codes,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Barcelona, Spain, 2016.

Other publications by the author not included in this thesis:

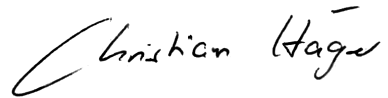
- C. Häger, A. Graell i Amat, A. Alvarado, and E. Agrell, “Constellation Optimization for Coherent Optical Channels Distorted by Nonlinear Phase Noise,” in *Proc. IEEE Global Communications Conf. (GLOBECOM)*, Anaheim, CA, 2012.
- C. Häger, A. Graell i Amat, A. Alvarado, and E. Agrell, “Design of APSK Constellations for Coherent Optical Channels with Nonlinear Phase Noise,” *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3362–3373, Aug. 2013.
- C. Häger, L. Beygi, E. Agrell, P. Johanisson, M. Karlsson, and A. Graell i Amat, “A Low-Complexity Detector for Memoryless Polarization-Multiplexed Fiber-Optical Channels,” *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 368–371, Jan. 2014.
- C. Häger, A. Graell i Amat, A. Alvarado, F. Brännström, and E. Agrell, “Optimized Bit Mappings for Spatially Coupled LDPC Codes over Parallel Binary Erasure Channels,” in *Proc. IEEE Int. Conf. Communications (ICC)*, Sydney, Australia, 2014.
- C. Häger, A. Graell i Amat, F. Brännström, A. Alvarado, and E. Agrell, “Comparison of Terminated and Tailbiting Spatially Coupled LDPC Codes With Optimized Bit Mapping for PM-64-QAM,” in *Proc. European Conf. Optical Communication (ECOC)*, Cannes, France, 2014.
- A. Graell i Amat, C. Häger, F. Brännström, and E. Agrell, “Spatially-Coupled Codes for Optical Communications: State-of-the-Art and Open Problems,” (invited paper) in *Proc. Optoelectronics and Communications Conf. (OECC)*, Shanghai, China, 2015.
- M. Ivanov, C. Häger, F. Brännström, A. Graell i Amat, A. Alvarado, and E. Agrell, “On the Information Loss of the Max-Log Approximation in BICM Systems,” *IEEE Trans. Inf. Theory* (to appear), 2016.
- C. Häger, H. D. Pfister, A. Graell i Amat, F. Brännström, and E. Agrell, “A Deterministic Construction and Density Evolution Analysis for Generalized Product Codes,” (invited paper) in *Proc. Int. Zurich Seminar on Communications*, Zürich, Switzerland, 2016.
- C. Häger, A. Graell i Amat, H. D. Pfister, and F. Brännström, “Density Evolution for Deterministic Generalized Product Codes with Higher-Order Modulation,” (invited paper) in *Proc. Int. Symp. on Turbo Codes and Iterative Information Processing (ISTC)*, Brest, France, 2016.

## Acknowledgements

First of all, I would like to express my deepest gratitude to my doctoral advisor (*Doktorvater* in German) Professor Alexandre Graell i Amat. Àlex has been a constant source of support and guidance during my studies and I am very grateful for the invaluable feedback that I have received from him throughout the past four and a half years. I would also like to express my gratitude to my co-advisor Professor Fredrik Brännström for all the fruitful discussions and for spotting even the slightest inconsistencies in my paper drafts. I would further like to thank Dr. Alex Alvarado and Professor Erik Agrell who have been my co-advisors until April 2014. Alex has also been an excellent host during my research visit to the University of Cambridge in 2013 for which I am very grateful.

I am especially indebted to Professor Henry Pfister who has had a great influence on the shape and direction of my doctoral studies. I had the pleasure to visit Henry for the first time in autumn 2014, where he suggested to work on the asymptotic analysis of deterministic codes. At least half of the work contained in this thesis directly originates from this collaboration which is still ongoing.

I would like to extend my thanks to Professor 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 my “early” PhD years is Dr. Lotfollah Beygi. I am immensely grateful for all the time that he spent with me discussing various aspects of fiber-optical communications. I would also like to thank Mikhail Ivanov for being an outstanding office mate, research collaborator, and friend. Many thanks also to the administrative staff of the department, in particular Agneta Kinnander, Madeleine Persson, and Natasha Adler. Special thanks goes to the late Lars Börjesson, who will be sorely missed as a computer administrator and colleague. 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.



Christian Häger  
Göteborg, April 2016

*Regarding the financial support, I would like to thank and acknowledge the Ericsson Research Foundation for their generous financial support of my research visits to Cambridge and Duke University. Furthermore, this work was partially funded by the Swedish Research Council under grant #2011-5961.*





# Acronyms

The following acronyms are used in the introductory part of this thesis.

Abbreviation	Meaning
<b>ASE</b>	amplified spontaneous emission
<b>AWGN</b>	additive white Gaussian noise
<b>BCH</b>	Bose–Chaudhuri–Hocquenghem
<b>BDD</b>	bounded-distance decoding
<b>BEC</b>	binary erasure channel
<b>BICM</b>	bit-interleaved coded modulation
<b>BP</b>	belief propagation
<b>BSC</b>	binary symmetric channel
<b>CM</b>	coded modulation
<b>CN</b>	constraint node
<b>DE</b>	density evolution
<b>EDFA</b>	erbium-doped fiber amplifier
<b>EXIT</b>	extrinsic information transfer
<b>GLDPC</b>	generalized low-density parity-check
<b>GPC</b>	generalized product code
<b>HPC</b>	half-product code
<b>LDPC</b>	low-density parity-check
<b>LLR</b>	log-likelihood ratio
<b>MI</b>	mutual information
<b>NLSE</b>	nonlinear Schrödinger equation
<b>OOK</b>	on-off keying
<b>PC</b>	product code
<b>PDF</b>	probability density function
<b>PM</b>	polarization-multiplexed
<b>PSD</b>	power spectral density
<b>QPSK</b>	quadrature phase-shift keying
<b>SMF</b>	single-mode fiber
<b>sNLSE</b>	stochastic nonlinear Schrödinger equation
<b>SNR</b>	signal-to-noise ratio
<b>SSFM</b>	split-step Fourier method
<b>VN</b>	variable node



---

## Contents

---

<b>Abstract</b>	<b>i</b>
<b>List of Papers</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Acronyms</b>	<b>vii</b>
<b>I Introduction</b>	<b>1</b>
<b>1 Background</b>	<b>3</b>
1.1 Thesis Organization . . . . .	6
1.2 Notation . . . . .	6
<b>2 Fiber-Optical Channel Modeling</b>	<b>9</b>
2.1 The Additive White Gaussian Noise Channel . . . . .	10
2.2 The Nonlinear Schrödinger Equation . . . . .	10
2.3 Optical Amplification and Noise . . . . .	12
2.4 Polarization Multiplexing . . . . .	14
2.5 Linear Pulse Modulation and Linear Receiver . . . . .	15
<b>3 Bit-Interleaved Coded Modulation</b>	<b>17</b>
3.1 Introduction to Coded Modulation . . . . .	18
3.2 BICM System Model . . . . .	19

<b>4</b>	<b>Low-Density Parity-Check Codes</b>	<b>23</b>
4.1	Introduction . . . . .	24
4.2	Iterative Belief Propagation Decoding . . . . .	24
4.3	Code Construction via Protographs . . . . .	26
4.4	Density Evolution . . . . .	27
4.5	Spatially-Coupled LDPC Codes . . . . .	29
<b>5</b>	<b>Generalized Product Codes</b>	<b>33</b>
5.1	Generalized Low-Density Parity-Check Codes . . . . .	34
5.2	Product Codes . . . . .	34
5.3	Generalized Product Codes . . . . .	36
5.3.1	Braided Codes . . . . .	36
5.3.2	Staircase Codes . . . . .	37
5.3.3	Half-Product Codes . . . . .	38
5.4	Iterative Bounded-Distance Decoding . . . . .	39
5.5	Density Evolution . . . . .	39
5.5.1	Code Ensembles . . . . .	40
5.5.2	Deterministic Codes . . . . .	41
<b>6</b>	<b>Conclusions and Future Work</b>	<b>43</b>
	<b>Bibliography</b>	<b>47</b>

## **II Papers 53**

<b>A</b>	<b>Improving Soft FEC Performance for Higher-Order Modulations via Optimized Bit Channel Mappings</b>	<b>A1</b>
1	Introduction . . . . .	A3
1.1	Notation . . . . .	A5
2	System Model . . . . .	A6
2.1	Continuous-Time Channel . . . . .	A6
2.2	Discrete-Time Channel . . . . .	A6
2.3	Bit-Interleaved Coded Modulation . . . . .	A7
3	Protograph-Based LDPC Codes . . . . .	A9
3.1	AR4JA Codes . . . . .	A10
3.2	Spatially Coupled LDPC Codes . . . . .	A10
3.3	Decoding and Asymptotic EXIT Analysis . . . . .	A10
4	Bit Mapper Optimization . . . . .	A12
4.1	Asymptotic Bit Mapper Model . . . . .	A12
4.2	Optimization . . . . .	A13
5	Results and Discussion . . . . .	A15

5.1	Linear Transmission . . . . .	A16
5.2	Nonlinear Transmission . . . . .	A17
6	Conclusion . . . . .	A19
	References . . . . .	A19

**B Terminated and Tailbiting Spatially-Coupled Codes with Optimized Bit Mappings for Spectrally Efficient Fiber-Optical Systems B1**

1	Introduction . . . . .	B3
2	System Model . . . . .	B5
3	Protograph-Based SC-LDPC Codes . . . . .	B7
3.1	Code Construction . . . . .	B7
3.2	Soft-Decision Decoding and Density Evolution . . . . .	B8
4	SC-GLDPC Codes with BCH Component Codes . . . . .	B10
4.1	Code Construction . . . . .	B10
4.2	Hard-Decision Decoding and Density Evolution . . . . .	B12
5	Bit Mapper Optimization . . . . .	B14
6	Results and Discussion . . . . .	B14
6.1	Structure of the Optimized Bit Mapper for Tailbiting Codes . . . .	B15
6.2	Optimization Gain . . . . .	B16
6.3	Gap to Capacity . . . . .	B18
6.4	Simulation Results . . . . .	B19
7	Conclusions . . . . .	B20
8	Acknowledgements . . . . .	B21
	References . . . . .	B21

**C On Parameter Optimization for Staircase Codes C1**

1	Introduction . . . . .	C3
2	Staircase Codes . . . . .	C4
2.1	Iterative Hard-Decision Decoding . . . . .	C4
2.2	Density Evolution . . . . .	C4
3	Extended Staircase Codes . . . . .	C6
4	Results and Discussion . . . . .	C6
5	Conclusions . . . . .	C8
	References . . . . .	C8

**D Density Evolution for Deterministic Generalized Product Codes on the Binary Erasure Channel D1**

1	Introduction . . . . .	D3
1.1	Notation . . . . .	D6
2	Half-Product Codes . . . . .	D6
2.1	Code Construction . . . . .	D6
2.2	Binary Erasure Channel . . . . .	D8

2.3	Iterative Decoding . . . . .	D8
2.4	Asymptotic Performance . . . . .	D9
3	Random Graphs and Branching Processes . . . . .	D12
3.1	Random Graphs . . . . .	D12
3.2	Neighborhood Exploration Process . . . . .	D13
3.3	Branching Processes . . . . .	D15
4	Proof of Theorem 1 . . . . .	D15
4.1	Tree-like Neighborhood . . . . .	D16
4.2	Convergence to the Poisson Branching Process . . . . .	D17
4.3	Density Evolution . . . . .	D18
4.4	Concentration . . . . .	D19
5	Generalized Product Codes . . . . .	D21
5.1	Code Construction . . . . .	D21
5.2	Inhomogeneous Random Graphs . . . . .	D25
5.3	Iterative Decoding . . . . .	D27
5.4	Asymptotic Performance . . . . .	D27
5.5	Density Evolution . . . . .	D29
6	Discussion . . . . .	D30
6.1	Thresholds and Code Comparisons . . . . .	D30
6.2	Upper Bound on the Decoding Threshold . . . . .	D31
6.3	Modified Decoding Schedules . . . . .	D31
6.4	Performance on the Binary Symmetric Channel . . . . .	D32
7	Irregular Half-Product Codes . . . . .	D33
7.1	Preliminaries . . . . .	D33
7.2	Lower Bounds on the Threshold . . . . .	D34
7.3	Optimization via Linear Programming . . . . .	D35
7.4	Initial Component Code Loss . . . . .	D36
7.5	Simulation Results . . . . .	D40
8	Conclusions and Future Work . . . . .	D42
A	Proof of Lemma 1 . . . . .	D42
B	Bound on the Second Moment of $T_\ell$ . . . . .	D43
	References . . . . .	D44

## **E Density Evolution and Error Floor Analysis for Staircase and Braided Codes E1**

1	Introduction . . . . .	E3
2	Density Evolution for Deterministic GPCs . . . . .	E3
2.1	Code Construction . . . . .	E4
2.2	Iterative Decoding . . . . .	E4
2.3	Density Evolution . . . . .	E4
3	Case Study: Comparison of Staircase, Braided, and Half-Braided Codes . . . . .	E5
3.1	Parameters and Error Floor . . . . .	E6

3.2	Results and Discussion . . . . .	E7
4	Conclusion . . . . .	E8
	References . . . . .	E8

<b>F</b>	<b>Deterministic and Ensemble-Based Spatially-Coupled Product Codes</b>	<b>F1</b>
1	Introduction . . . . .	F3
2	Code Construction and Density Evolution for Deterministic Generalized Product Codes . . . . .	F4
2.1	Code Construction . . . . .	F4
2.2	Iterative Decoding . . . . .	F5
2.3	Density Evolution . . . . .	F5
3	Spatially-Coupled Product Codes . . . . .	F6
3.1	Deterministic Spatially-Coupled Product Codes . . . . .	F6
3.2	Spatially-Coupled Product Code Ensembles . . . . .	F7
4	Comparison of Deterministic and Ensemble-Based Codes . . . . .	F8
4.1	Ensemble Performance via Deterministic Codes . . . . .	F9
4.2	Simpler Deterministic Codes . . . . .	F10
5	Potential Threshold Optimization . . . . .	F11
6	Conclusion . . . . .	F11
A	Proof of Theorems 3 and 4 . . . . .	F12
	References . . . . .	F13





## **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].

There is currently a significant interest in determining the ultimate capacity limits of fiber-optical systems [2–4] and developing practical schemes that can achieve these limits [5–7]. Error-correcting codes are an integral part of communication systems that operate close to capacity. In theory, the proper use of such codes allows the system to achieve an arbitrarily low error rate if the data rate is chosen below the capacity [8]. In practice, however, operating at lower error rates and closer to capacity comes at the expense of an increased system complexity and communication delay. Code design thus requires assessment of nontrivial trade-offs between performance, complexity, and delay.

Fiber-optical communication systems operate at very high data rates that can exceed several hundreds of Gbit/s. At such high speeds, one of the main challenges is to keep the decoding complexity at an acceptable level. For example, the amount of processing power that can be spent on decoding each bit is severely limited [6]. Fiber-optical communication systems also require extremely low error rates below  $10^{-15}$  [6]. On the other hand, communication delay caused by coding is typically not an issue. This implies that the use of codes with long block lengths is relatively unproblematic. As an example, the code proposals in [9, 10] have effective block lengths in the order of  $10^6$  bits. Assuming

a transmission rate of 100 Gbit/s, the corresponding delay is then only  $20\,\mu\text{s}$ . This is negligible compared to the propagation delay caused by the light traversing hundreds or even thousands of kilometers of fiber.<sup>1</sup> In a nutshell, the challenge is therefore to devise coding schemes that offer very low error rates with affordable decoding complexity, while potentially using relatively long block lengths.

Current state-of-the-art codes are defined on graphs and decoded iteratively by passing messages along the edges of the graph [11, 12]. In this thesis, our focus is primarily on spatially-coupled codes. The first instance of spatially-coupled codes are the low-density parity-check (LDPC) convolutional codes introduced in [13], which are now also referred to as spatially-coupled LDPC codes. It was later realized in [14] that these codes have much better performance than conventional regular LDPC codes [15]. The reason for this phenomenon is a wave-like decoding behavior caused by a termination boundary effect. Several different proofs now exist for the fact that spatially-coupled LDPC codes can operate arbitrarily close to the capacity of a variety of different communication channels [16, 17]. Other instances of spatially-coupled codes include braided codes [18] and staircase codes [9], which have been shown to offer outstanding performance using very low-complexity decoding algorithms. In this thesis, we address several challenges that arise in the analysis and design of spatially-coupled codes when used in fiber-optical communication systems.

We start by studying the design of spectrally-efficient systems. 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. To improve the spectral efficiency over OOK, the data can be encoded into multiple amplitude and/or phase levels of the optical carrier. A further increase in spectral efficiency can be achieved by utilizing both polarizations of the optical light, which is referred to as polarization-multiplexed (PM) transmission. PM signals can be represented as points in a four-dimensional signal space, also referred to as a signal constellation [19]. In Papers A and B, we study how to combine such signal constellations with error-correcting codes. In Paper A, we consider LDPC codes which are defined via protographs [20]. Protograph-based LDPC codes allow for an efficient hardware implementation, which makes them attractive candidates for high-speed application such as fiber-optical communications [6].

Spatially-coupled LDPC codes can also be constructed using protographs. An in-depth analysis focusing exclusively on spatially-coupled LDPC codes for spectrally-efficient systems is presented in Paper B. In order to highlight one of the contributions of this thesis, we first note that the termination boundary that induces the wave-like decoding behav-

---

<sup>1</sup>As a rule of thumb, light takes approximately 5 ms to propagate through 1000 kilometer of fiber.

---

ior for spatially-coupled LDPC codes comes at the price of a so-called rate loss, i.e., a larger redundancy overhead, compared to “uncoupled” regular LDPC codes. This rate loss can be avoided by considering a tail-biting termination scheme. The resulting codes are referred to as tail-biting spatially-coupled LDPC codes. Unfortunately, by default, these codes behave essentially the same as regular LDPC codes because the absence of an explicit termination boundary prevents a wave-like decoding behavior. One of the main findings of this work is that the different modulation bits of a nonbinary signal constellation can be exploited to create an artificial termination boundary. This artificial boundary is sufficient to initiate a decoding wave and can significantly improve the performance of “rate-loss-free” tail-biting spatially-coupled LDPC codes.

As a second topic, we investigate so-called generalized product codes (GPCs). GPCs are extensions of classical product codes (PCs) [21]. PCs are one of the first examples of the idea to build long and powerful codes from shorter component codes. In particular, each coded bit in a PC is protected by two component codes, where the coded bits are assumed to be arranged in a rectangular array. This assumption is relaxed for GPCs which allows the array shape to be arbitrary. GPCs are very appealing for high-speed applications such as fiber-optical communication systems. The reason is rooted in their low-complexity decoding algorithm, which is based on iteratively decoding the component codes. When compared to message-passing decoding of LDPC codes, this approach can result in significant complexity advantages [9]. Indeed, PCs are already implemented in certain communication standards for fiber-optical systems [22]. Moreover, several constructions of GPCs, e.g., braided [10] and staircase codes [9], have been recently proposed and investigated for such systems.

GPCs are the main focus in Papers C–F (although we already consider the application of GPCs for the bit mapper optimization in Paper B as a side application). In Paper C, we study parameter optimization for staircase codes. This work is inspired by the work in [23], where staircase code parameters are found using a simulation-based approach. The parameter optimization in Paper C on the other hand is based on density evolution (DE). DE is an analytical tool to analyze the behavior of codes under iterative decoding in the limit of infinite block lengths [24]. However, the DE analysis used in Paper C does not directly apply to staircase codes. Rather, we observed that staircase codes share some structural similarities with the spatially-coupled PC ensemble defined in [25]. A code ensemble is a collection or set of codes, typically defined via suitable randomized connections in the underlying graphical representation. The approach used in Paper C is therefore only heuristically motivated. While it appears to work well, it raises the question whether an asymptotic DE analysis is possible by directly targeting specific deterministic GPCs such as staircase codes.

This question is answered positively in Paper D, which represents the main theoretical contribution of this thesis. In Paper D, we propose a parametrized family of deterministic GPCs that includes staircase codes (and also many other code classes) as special cases. Based on the theory of inhomogeneous random graphs [26], we provide a DE analysis that

characterizes the asymptotic code performance. It is important to stress that the analysis does not rely on the definition of a code ensemble but directly applies to sequences of deterministically constructed codes. The resulting DE analysis can be useful for a variety of different applications. For example, it can be used to predict and compare the waterfall performance of different GPCs, optimize code parameters for particular classes of GPCs, find suitable windowed-decoding schedules for spatially-coupled PCs, or design new classes of GPCs. Papers E and F are based on the theoretical tools derived in Paper D, where we apply the theory to design and study deterministically constructed GPCs in more detail.

## 1.1 Thesis Organization

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 introductory 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. In Chapter 3, we give a brief introduction to bit-interleaved coded modulation (BICM), which is a pragmatic way to combine signal constellations with error-correcting codes to operate at high spectral efficiencies. In Chapter 4, we review some basic background about LDPC codes. In particular, we discuss iterative belief propagation (BP) decoding, DE, and protograph-based constructions including spatially-coupled LDPC codes. The content of this chapter is mainly relevant for Papers A and B, where the reader is assumed to be somewhat familiar with LDPC codes and iterative decoding. In Chapter 5, we then discuss GPCs. Starting from the concept of generalized low-density parity-check (GLDPC) codes, we give several examples of well-known classes of GPCs and discuss the decoding via iterative bounded-distance decoding (BDD). We also briefly review two approaches to perform an asymptotic analysis for GPCs. Finally, the main conclusions from the appended papers are summarized in Chapter 6, where we also discuss future work.

## 1.2 Notation

The following notation is used in the introductory part of this thesis.

- Vectors and matrices are typeset in bold font by lowercase letters  $\mathbf{a}$  and capital letters  $\mathbf{A}$ , respectively.
- The transpose of a matrix is denoted by  $(\cdot)^\top$ .
- $\mathbb{Z}$ ,  $\mathbb{N}_0$ ,  $\mathbb{N}$ ,  $\mathbb{R}$ , and  $\mathbb{C}$  denote the set of integers, nonnegative integers including zero, nonnegative integers excluding zero, real numbers, and complex numbers,

respectively.

- The cardinality of a set  $\mathcal{A}$  is denoted by  $|\mathcal{A}|$ .
- Random variables are denoted by capital letters  $X$  and their realizations by lower-case letters  $x$ .
- The probability density function (PDF) of a random variable  $X$  is denoted by  $f_X(x)$ . The PDF of a random variable  $Y$  conditioned on the realization of another random variable  $X$  is denoted by  $f_{Y|X}(y|x)$ .
- Expectation is denoted by  $\mathbb{E}[\cdot]$ .
- $\delta(t)$  denotes Dirac's delta function while  $\delta[k]$  denotes the Kronecker delta.
- Convolution is denoted by  $\otimes$ .
- The imaginary unit is denoted by  $j \triangleq \sqrt{-1}$ .
- Complex conjugation is denoted by  $(\cdot)^*$ .

### Notational Inconsistencies

The reader should be aware of the following inconsistencies in the notation across the appended papers and the thesis introduction.

- The variable  $n$  is used to denote the block length of an LDPC code and GLDPC code in Papers A, B, and the thesis introduction. It is used for different purposes in Papers C–F, e.g., it denotes the length of a Bose–Chaudhuri–Hocquenghem (BCH) code in Paper C.
- The block size of staircase codes and braided codes is denoted differently in different papers and the thesis introduction. For example, for staircase codes, the block size is denoted by  $a$  in Paper C, E and the thesis introduction, while it is denoted by  $d$  in Paper F. In Paper D, the block size is denoted by  $n_i$ , where the index  $i$  indicates the position in the Tanner graph.
- The error-correcting capability of a component code is denoted by  $t$  in Papers B, C, F, and the thesis introduction, while it is denoted by  $\mathbf{t}$  in Papers D and E. The variable  $t$  is also used as a time index for the signals in Papers A, B, and Chapter 2 of the thesis introduction.
- The matrix  $\boldsymbol{\eta}$  that defines the Tanner graph connectivity of the deterministic GPC construction is typeset in normal weight font as  $\eta$  in Paper E.
- The two light polarizations are denoted by  $\mathbf{a}$  and  $\mathbf{b}$  in the thesis introduction, while we use  $\mathbf{x}$  and  $\mathbf{y}$  in Papers A and B.

- The acronym “SC” is used for “spatially-coupled” in Papers A, B, and F, while it is used for “staircase code” in Paper E.
- The acronym “CN” is used for “constraint node” in Papers D, E, F, and the thesis introduction, while it is used for “check node” in Papers A and B.
- The spatially-coupled code ensemble in [27] is referred to as a “SC-GLDPC ensemble” in Papers B and C, while it is referred to as a “spatially-coupled PC ensemble” in Paper F and the thesis introduction.



---

## 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, we assume “traditional” channel models, in particular the (discrete and memoryless) additive white Gaussian noise (AWGN) channel. This chapter is intended to put this channel model into the context of fiber-optical communications. It should also give the reader a broader picture about optical channel modeling in general. Other channel models that are mainly relevant for iterative hard-decision decoding are discussed in Chapter 5.

We are concerned with coherent, long-haul (i.e., distances exceeding 2000 km) data transmission over single-mode fibers (SMFs). The main challenge from a channel modeling perspective 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 [28, p. 18].

The chapter is structured as follows. In Section 2.1, we review the AWGN channel model. In Section 2.2, we discuss the nonlinear Schrödinger equation (NLSE) which is a deterministic channel model for an SMF. Multi-span links consisting of several SMFs including optical amplification elements are covered in Section 2.3. PM systems are discussed in 2.4. Lastly, linear modulation and receivers are described in Section 2.5.

## 2.1 The Additive White Gaussian Noise Channel

Consider the discrete and memoryless (complex-valued) AWGN channel model

$$y_k = x_k + n_k \quad (2.1)$$

for  $k \in \mathbb{Z}$ , where  $x_k \in \mathbb{C}$  denotes an information symbol,  $n_k$  is a realization of a zero-mean complex Gaussian random variable  $N_k$  with  $\mathbb{E}[N_k N_{k'}^*] = N_0 \delta[k - k']$  and power spectral density (PSD)  $N_0$ , and  $y_k \in \mathbb{C}$  is the channel output or simply the observation. The channel from  $x_k$  to  $y_k$  is characterized by the conditional PDF

$$f_{Y_k|X_k}(y_k|x_k) = \frac{1}{\pi N_0} \exp\left(-\frac{|y_k - x_k|^2}{N_0}\right). \quad (2.2)$$

The primary goal of this section is to motivate (2.1) in the context of fiber-optical communication systems. In particular, our target application is coherent, long-haul data transmission over SMFs. The validity of (2.1) in this case depends heavily on the assumed system parameters, e.g., the type of dispersion-compensation scheme that is being used.

## 2.2 The Nonlinear Schrödinger Equation

The starting point for fiber-optical channel modeling is the NLSE, which can be derived from Maxwell's equations under some assumptions that are appropriate for SMFs [29]. The NLSE is a partial differential equation that defines the input–output relationship for optical baseband signals<sup>1</sup> propagating through SMFs.

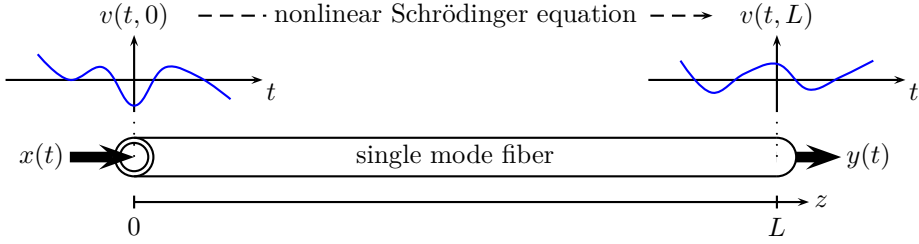
Let us first introduce a continuous-time parameter  $t \in \mathbb{R}$  and a distance parameter  $0 \leq z \leq L$  that denotes the propagation distance of the signal from the beginning of the fiber, where  $L$  is the total length of the fiber. The baseband signal of interest is a function of two parameters, denoted by  $v(t, z)$ . We define the input and output signals as  $x(t) \triangleq v(t, 0)$  and  $y(t) \triangleq v(t, L)$ , 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. 2.1. Before we continue, we also define the instantaneous signal power  $P(t, z) \triangleq |v(t, z)|^2$  and the power profile  $P(z) \triangleq \lim_{T \rightarrow \infty} (\int_{-T}^T P(t, z) dt) / (2T)$ , where  $P = P(0)$  is the power of the input signal.

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

$$\frac{\partial v(t, z)}{\partial z} = -\frac{\alpha}{2} v(t, z) - j \frac{\beta_2}{2} \frac{\partial^2 v(t, z)}{\partial t^2} + j \gamma v(t, z) |v(t, z)|^2, \quad (2.3)$$

where  $\alpha$  is the attenuation coefficient,  $\beta_2$  is the chromatic dispersion coefficient, and  $\gamma$  is the nonlinear Kerr parameter. If we take into account only the first term on the

<sup>1</sup>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.



**Figure 2.1:** Conceptual representation of the signal evolution through an SMF. The NLSE describes the relationship between the input signal  $x(t) = v(t, 0)$  and the output signal  $y(t) = v(t, L)$ .

right-hand side of (2.3), one obtains  $v(t, z) = \exp(-\alpha z/2)v(t, 0)$  as a solution<sup>2</sup>, 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 (2.3), one obtains an alternative and somewhat simpler version of the NLSE as [29, eq. (4)]

$$\frac{\partial u(t, z)}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 u(t, z)}{\partial t^2} + j\gamma e^{-\alpha z}u(t, z)|u(t, z)|^2. \quad (2.4)$$

In general, there are no closed-form solutions to the NLSE and one has to resort to numerical methods in order to obtain a solution. In the following, we briefly describe one of the most widely used numerical methods to solve (2.4), namely the split-step Fourier method (SSFM). Conceptually, we start by discretizing the spatial dimension and subdividing the entire fiber of length  $L$  into small segments of length  $\Delta$ , where  $M = L/\Delta \in \mathbb{N}$  is the total number of segments. For the  $i$ -th 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 then assumed that an approximate solution to obtain  $u(t, i\Delta)$  based on  $u(t, (i-1)\Delta)$  is given by

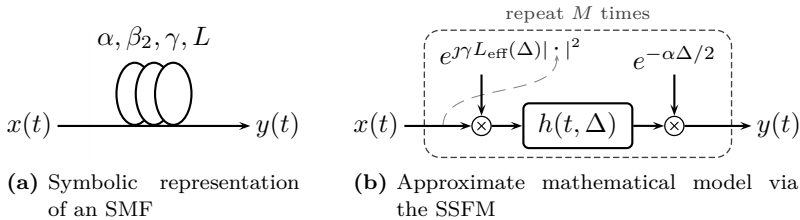
$$u(t, i\Delta) \approx h(t, \Delta) \otimes \left( u(t, (i-1)\Delta) e^{j\gamma L_{\text{eff}}(\Delta)|u(t, (i-1)\Delta)|^2} \right), \quad (2.5)$$

where  $h(t, z) = \exp(jt^2/(2\beta_2 z)) / \sqrt{j2\pi\beta_2 z}$  is the impulse response of a linear filter representing dispersive effects and

$$L_{\text{eff}}(z) \triangleq \int_0^z e^{-\alpha z'} dz' = \frac{1 - \exp(-\alpha z)}{\alpha} \quad (2.6)$$

is referred to as the effective nonlinear length with  $L_{\text{eff}}(z) \leq z$  and  $L_{\text{eff}}(z) \rightarrow z$  as  $\alpha \rightarrow 0$ . The reasoning behind (2.5) is that over a short segment of length  $\Delta \ll L$ , the linear (i.e., dispersive) and nonlinear effects act almost independently of one another.

<sup>2</sup>Recall that the solution of  $\partial f(z)/\partial z = cf(z)$  is given by  $f(z) = \exp(cz)f(0)$ .



**Figure 2.2:** Illustrations for an SMF. The notation  $|\cdot|^2$  in (b) stands for the instantaneous power of the signal that arrives at the corresponding multiplication block as indicated by the dashed, gray line.

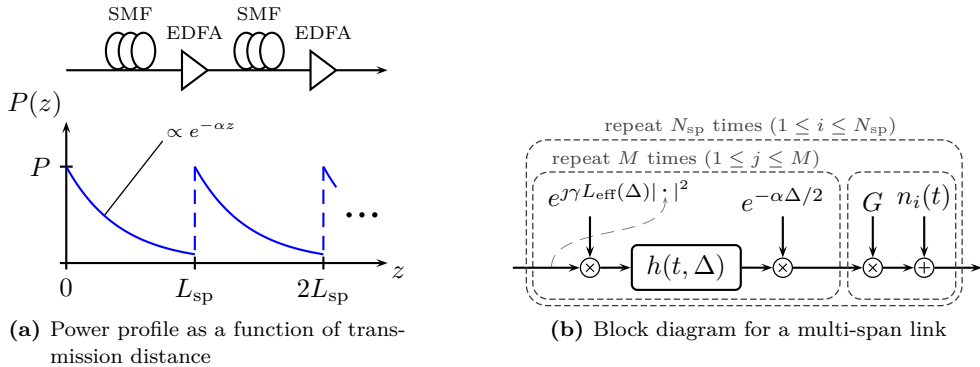
Using this assumption, an approximate solution for an entire SMF of length  $L$  is given by repeatedly applying (2.5), starting with the first segment where  $i = 1$ , i.e., with the input signal  $u(t, 0) = x(t)$ . The SSFM step in (2.5) is given in terms of the normalized signal  $u(t, z)$ . In order 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. Fig. 2.2 shows the resulting numerical method in terms of a block diagram. 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  $\Delta \rightarrow 0$  [28, p. 42]. Practical guidelines on the choice of the segment size are developed in [30].

The name of the method originates from the fact that the nonlinear phase-shift operation and the linear filtering in Fig. 2.2(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 the baseband signal  $u(t, z)$  (or  $v(t, z)$ ) is considered which facilitates the application of the computationally efficient fast Fourier transform. Such an implementation is for example provided in [28, App. B].

## 2.3 Optical Amplification and Noise

The numerical value of the attenuation coefficient  $\alpha$  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.

We briefly discuss one common type of amplification, referred to as lumped amplification, in terms of its effect on the power profile of the signal and the type of noise that is introduced. Modeling the power profile is important due to the dependency of the nonlinear effect on the instantaneous signal power. Thus, one cannot simply ignore atten-



**Figure 2.3:** Illustrations for a fiber-optical communication link including a lumped amplification scheme and noise.

uation 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., [31, 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. This means that noise from other sources, e.g., 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 (2.3) can be extended by inserting a gain profile  $g(z)$  and a complex-valued stochastic process  $w(t, z)$ , resulting in

$$\frac{\partial v(t, z)}{\partial z} = -\frac{\alpha - g(z)}{2}v(t, z) - j\frac{\beta_2}{2}\frac{\partial^2 v(t, z)}{\partial t^2} + j\gamma v(t, z)|v(t, z)|^2 + w(t, z). \quad (2.7)$$

Equation (2.7) is referred to as the stochastic nonlinear Schrödinger equation (sNLSE) [32]. We start by discussing the gain profile  $g(z)$  and its effect on the power profile of the signal  $v(t, z)$ , ignoring all other effects (including  $w(t, z)$ ). Signal amplification is applied periodically, in the sense that the entire transmission distance  $0 \leq z \leq L$  is split up into spans of length  $L_{\text{sp}}$  varying between 60 and 120 km, where  $N_{\text{sp}} = L/L_{\text{sp}} \in \mathbb{N}$  denotes the total number of spans. For 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 (2.7), 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 schematically illustrated in Fig. 2.3(a). It can be seen that the signal power decreases exponentially according to the loss coefficient  $\alpha$  and is periodically restored to the input power  $P$  after each span.

Next, we discuss the noise that is generated by optical amplification through a process called amplified spontaneous emission (ASE). Noise can be thought of as being added to the signal at discrete locations  $z_i \triangleq iL_{\text{sp}}$ ,  $1 \leq i \leq N_{\text{sp}}$ . Thus, if we think about  $z_i^-$  and  $z_i^+$  as the locations right before and after the amplifiers, we have  $v(t, z_i^+) = Gv(t, z_i^-) + n_i(t)$ ,

where  $n_i(t)$  is the additive noise originating from the  $i$ -th amplifier [33, 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.,  $\mathbb{E}[N_i(t)N_j^*(t')] = N_\ell \delta(t - t')\delta[i - j]$ , where  $N_\ell$  denotes the noise PSD per amplifier. (The index  $\ell$  refers to the lumped amplification type.) We further set  $N_0 = N_{\text{sp}}N_\ell$ , which one might think of as the cumulative PSD at the end of the transmission link for  $N_{\text{sp}}$  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  $\delta_{W_n}(x) = W_n \text{sinc}(W_n x)$  [33]. Further limitations of the optical bandwidth can occur due to the insertion of optical bandpass filters and/or reconfigurable optical add-drop multiplexer along the transmission line [2].

Based on the previous description, a block diagram of a continuous-time model for a multi-span transmission link with lumped amplification is depicted in Fig. 2.3(b). The model consists of the concatenation of the deterministic model for an SMF based on the SSFM (cf. Fig. 2.2(b)) with a multiplicative gain factor and additive noise representing the optical amplifier. For completeness, we also indicate how the additive noise terms  $n_i(t)$  can be related to  $w(t, z)$  in (2.7). Note that if we neglect all terms on the right-hand side of (2.7) except  $w(t, z)$ , we have  $\partial v(t, z)/\partial z = w(t, z)$  and integrating this equation leads to

$$v(t, z) = v(t, 0) + \int_0^z w(t, \xi) d\xi = v(t, 0) + n(t, z). \quad (2.8)$$

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  $w(t, z) = \sum_{i=1}^{N_{\text{sp}}} n_i(t)\delta(z - iL_{\text{sp}})$  [29, p. 84], so that  $n(t, z) = \sum_{i=1}^{\lfloor z/L_{\text{sp}} \rfloor} n_i(t)$  corresponds the addition of all  $n_i(t)$  up to distance  $z$  (the upper integral limit in (2.8) is interpreted as  $z^+$ ).

## 2.4 Polarization Multiplexing

In addition to the amplitude and phase (or, alternatively, the in-phase and quadrature component), data may also be encoded using the polarization of the light source. Systems where both polarizations of the light are used to transmit data are referred to as PM. For PM transmission, the sNLSE equation can be further extended by considering the vector signal  $\mathbf{v}(t, z) = (v_a(t, z), v_b(t, z))^T$ , where the indices indicate the two polarizations

**a** and **b**.<sup>3</sup> The resulting equation is referred to as the Manakov equation which includes amplifier noise, gain, and loss terms. It is given by [34, p. 8]

$$\frac{\partial \mathbf{v}(t, z)}{\partial z} = -\frac{\alpha - g(z)}{2} \mathbf{v}(t, z) - j \frac{\beta_2}{2} \frac{\partial^2 \mathbf{v}(t, z)}{\partial t^2} + j \gamma \mathbf{v}(t, z) \|\mathbf{v}(t, z)\|^2 + \mathbf{w}(t, z), \quad (2.9)$$

where  $\mathbf{w}(t, z) = (w_a(t, z), w_b(t, z))^\top$  are two (independent) stochastic processes describing the ASE noise generated in both polarizations. The major difference between (2.9) and (2.7) is that (2.9) models the nonlinearity that is due to the sum of the instantaneous power in both polarizations  $\|\mathbf{v}(t, z)\|^2 = P_a(t, z)^2 + P_b(t, z)^2$ . We should mention that (2.7) ignores the fact that amplifier noise is always generated “in two polarizations”, i.e., even if we assume one of the two signals in  $\mathbf{v}(t, z)$  to be zero, technically the amplifier noise in that polarization still contributes via (2.9) through the fiber nonlinearity.

For simplicity, we ignore polarization-specific impairments. This includes for example polarization mode dispersion, which would cause different group velocities of the signals in polarization **a** and **b** caused by natural imperfections and asymmetries of the fiber cross-section area.

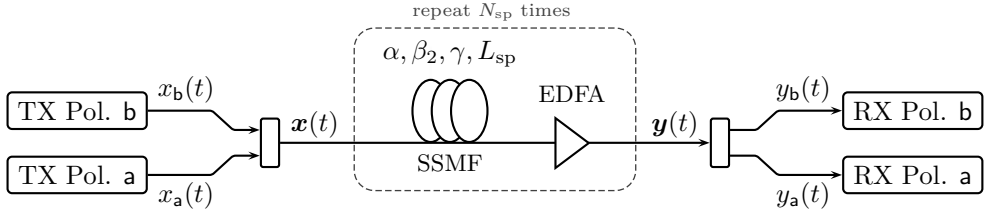
## 2.5 Linear Pulse Modulation and Linear Receiver

So far, we have discussed models for waveform channels. In order to arrive at a discrete-time channel model, we have to make some assumptions about the type of modulation that is used in the transmitter and the type of receiver structure. The statistics of the resulting discrete-time channel may depend heavily on these assumptions.

In Fig. 2.4, a generic block diagram for a PM transmission scheme is shown. We assume that the transmitters (TX) employ a linear pulse modulation according to  $x_a(t) = \sum_k x_{a,k} p(t - kT_s)$  for polarization **a** and similarly for polarization **b**, where  $T_s$  is the symbol period. The evolution of the PM signal is then described by the Manakov equation (2.9), where  $v_a(t, 0) = x_a(t)$  and  $v_b(t, 0) = x_b(t)$ . The received signal in each polarization is assumed to be processed according to a linear receiver. In particular, for polarization **a**, it is assumed that  $y_a(t) = v_a(t, L)$  is passed through an equalizer, a pulse-matched filter, and a sampler, in order to obtain  $y_{a,k'} = y_a(t) \otimes h(t, -L) \otimes p(-t)|_{t=k'T_s}$  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. Here, we focus on optical transmission links without any inline dispersion compensation, which are referred to as non-dispersion-managed or uncompensated transmission links. Recently, there has been a substantial amount of work on this type of transmission link with the goal to find such

<sup>3</sup>This nonstandard notation for the polarizations is an attempt to avoid confusion with the transmitted and received signals.



**Figure 2.4:** Block diagram of the PM transmission scheme considered in Papers A and B.

a statistical relationship [35–38].<sup>4</sup>

In [35], it is shown that the discrete-time channel for non-dispersion-managed 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  $1/T_s$  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 this nonlinear noise is additive, Gaussian, and uncorrelated (both in time and across polarizations). A discrete-time channel model in polarization *a* is then given according to

$$y_a = \zeta x_{a,k} + n_{a,k} + \tilde{n}_{a,k}, \quad (2.10)$$

where  $\zeta \in \mathbb{C}$  is a complex scaling factor,  $n_{a,k}$  corresponds to the linear ASE noise with  $\mathbb{E}[N_{a,k}N_{a,k}^*] = N_0/T_s = P_{\text{ASE}}$ ,  $\tilde{n}_{a,k}$  accounts for nonlinear noise with  $\mathbb{E}[\tilde{N}_{a,k}\tilde{N}_{a,k}^*] = \eta P^3$ , and the same transmit power  $P$  is assumed for the signals in both polarizations. The parameter  $\eta$  (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 the “conventional” discrete-time additive Gaussian channel in (2.1) 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  $(P, P_{\text{ASE}})$  or, more practically relevant, the pair  $(P, L)$ . This parameter pair leads in turn to both a linear and a nonlinear noise variance based on which an effective SNR can be computed.

<sup>4</sup>This case is 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.”



---

## Bit-Interleaved Coded Modulation

---

In this chapter, we provide a brief introduction on how to design systems that reliably transmit data at high spectral efficiencies. Spectrally-efficient communication can be achieved in practice by combining error-correcting codes with nonbinary signal constellations, which is commonly referred to as coded modulation (CM). We focus on BICM, which is a pragmatic approach to CM and often implemented in practice, due to its inherent simplicity and flexibility.

We start by outlining the main principles behind CM in Section 3.1. In Section 3.2, we explain the building blocks of a BICM system. We also cover the parallel independent channel model for BICM which is used for the bit mapper optimization problem studied in Papers A and B.

### 3.1 Introduction to Coded Modulation

Consider again the discrete memoryless AWGN channel in (2.1). The goal is to reliably transmit data at high spectral efficiencies over this channel. To do so, one can formally define an encoder  $\mathcal{E} : \{0, 1\}^d \rightarrow \mathcal{C}_c$ , which maps a vector of  $d$  information bits to a codeword in the code  $\mathcal{C}_c \subset \mathbb{C}^N$ . 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  $\mathcal{D} : \mathbb{C}^N \rightarrow \{0, 1\}^d$ , which maps a vector of  $N$  channel observations 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  $\kappa = \log_2(|\mathcal{C}_c|)/N = d/N$ . Notice that the communication rate of the discrete-time channel 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 time  $T_s$ . Shannon proved that all rates up to the channel capacity

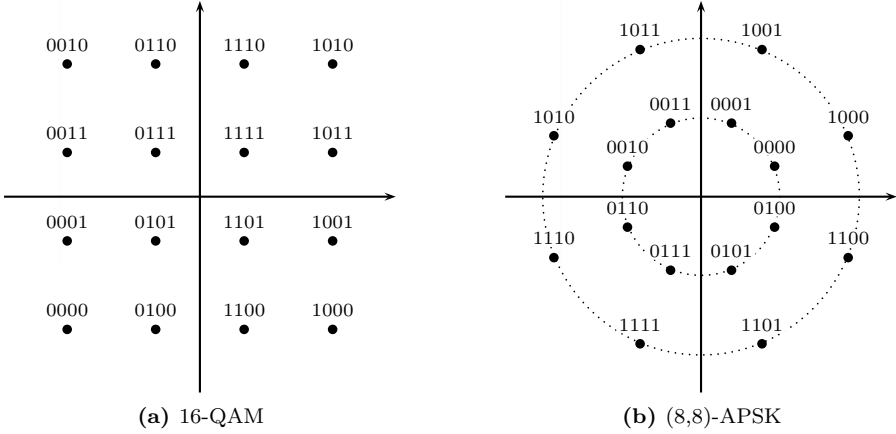
$$C = \log_2(1 + \text{SNR}) \quad (3.1)$$

are achievable, in the sense that there exists an encoder/decoder pair that can provide an arbitrarily small error probability as long as  $N \rightarrow \infty$  [39].

While Shannon's proof provides communication engineers with an invaluable benchmark, the problem of designing practical encoders and decoders that operate close to 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  $\kappa \leq 2$  is referred to as the power-limited regime and  $\kappa > 2$  as the bandwidth-limited regime [40]. 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  $\mathcal{X} = \{1 + j, 1 - j, -1 + j, -1 - j\}$  and scaled by  $\sqrt{P/2}$ ), in combination with binary error-correcting codes in order to operate close to the capacity. On the other hand, spectrally-efficient communication requires the use of signal constellations with cardinality larger than 4, which are referred to as nonbinary or higher-order<sup>1</sup> constellations. By invoking the capacity formula, it follows directly that operating at high spectral efficiencies (where  $\kappa > 2$ ) requires the signal power to be at least three times the noise power. In other words, spectrally-efficient communication requires a reasonably high SNR.

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 [41], CM with nonbinary codes [42],

<sup>1</sup>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.



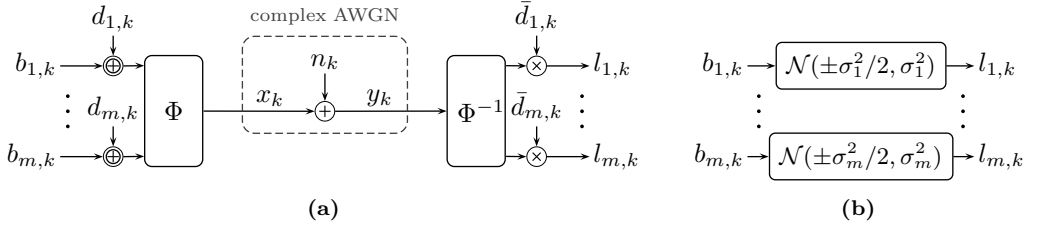
**Figure 3.1:** Two examples of higher-order signal constellations with 16 points.

multilevel coded modulation [43], or BICM [44]. Our focus here 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 [45]. 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., [46] or [47] and references therein.

## 3.2 BICM System Model

In the following, 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  $b_i(x)$ ,  $1 \leq i \leq m$ , denotes the  $i$ -th bit in the binary string assigned to  $x \in \mathcal{X}$  (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-QAM and (8,8)-APSK.

We now 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  $d_{i,k} \in \{0, 1\}$  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  $\Phi$  takes  $m$  bits  $b_{i,k}$ ,  $1 \leq i \leq m$ , and maps them to one of the constellation points according to the binary labeling of the signal constellation. At the receiver, the demodulator  $\Phi^{-1}$  computes soft reliability information about the



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

transmitted bits in the form of the log-likelihood ratios (LLRs)

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

where  $\mathcal{X}_{i,u} \triangleq \{x \in \mathcal{X} : b_i(a) = u\}$  is the subconstellation where all points have the bit  $u$  at the  $i$ -th 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  $\Phi$ , the AWGN channel, and demodulator  $\Phi^{-1}$  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  $f_{L_{i,k}|B_{i,k}}(\cdot|b)$ ,  $1 \leq i \leq m$ , 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 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 [48, Ch. 5]. While  $f_{L_{i,k}|B_{i,k}}(\cdot|b)$  is an LLR channel, the channel is not necessarily symmetric in general.<sup>2</sup> Symmetry can be enforced by adding modulo 2 independent and identically distributed bits  $d_{i,k}$  to the bits  $b_{i,k}$  [49]. 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  $f_{L_{i,k}|B_{i,k}}(\cdot|b)$  can be shown to be symmetric [49].

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

<sup>2</sup>The symmetry condition will become important when discussing DE 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

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

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

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

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

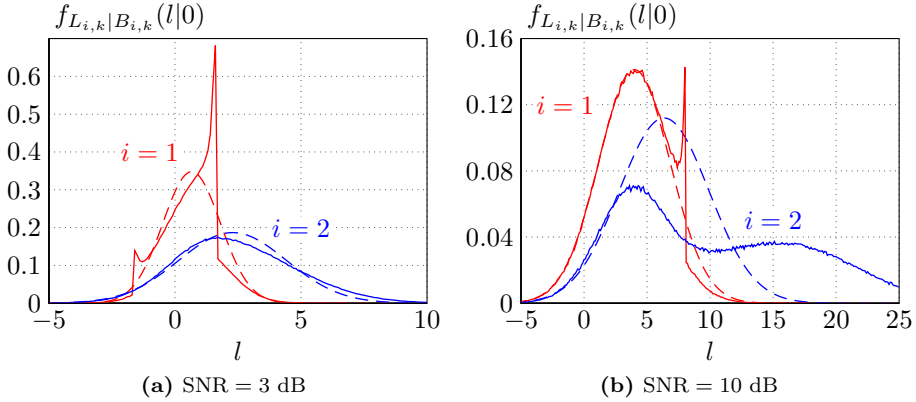
$$= 1 - \int_{-\infty}^{+\infty} f_{L|B}(l|0) \log_2(1 + \exp(-l)) dl. \quad (3.7)$$

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

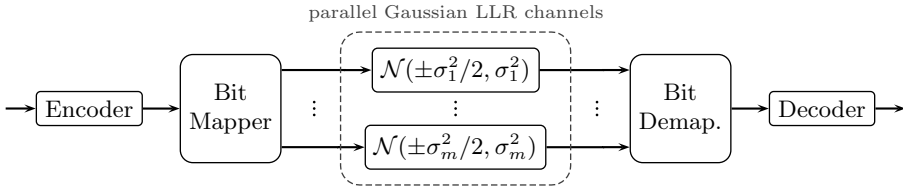
It turns out that, while the channel quality of the bit channels can be determined quite efficiently, it is very difficult to find exact analytical expressions for the actual densities  $f_{L_{i,k}|B_{i,k}}(\cdot|\cdot)$ . A common approach in the analysis of BICM is to make the simplifying assumption that the densities  $f_{L_{i,k}|B_{i,k}}(\cdot|\cdot)$  are 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  $\sigma^2$  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  $B = 1$ , where  $\mathcal{N}(\mu, \sigma^2)$  denotes the Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . 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  $\sigma_i^2$ . In order to find a correspondence between the LLR channels  $f_{L_{i,k}|B_{i,k}}(\cdot|\cdot)$  and the parameters  $\sigma_i^2$ , one may match the MI according to  $J(\sigma_i) = I_i(\text{SNR}) \Leftrightarrow \sigma_i^2 = J^{-1}(I_i(\text{SNR}))^2$ , where  $I_i(\text{SNR}) = I(B_{i,k}; L_{i,k})$  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).<sup>3</sup> The densities  $f_{L_{i,k}|B_{i,k}}(l|0)$  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 Papers A and B (i.e., predicting the iterative

<sup>3</sup>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 constellations.



**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.

performance behavior of LDPC codes), the approximation turns out to be 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 Papers A and B, our goal is to find good bit mappers for a given code and signal constellation.

As a side note, we remark that the term “bit interleaver” is also commonly used instead of “bit mapper”. In fact, the modulator  $\Phi$  is sometimes referred to as the (symbol) mapper (and the demodulator  $\Phi^{-1}$  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  $\Phi$  is explicitly studied or optimized, see, e.g., [50, 51]. 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 [52, 53].

---

## Low-Density Parity-Check Codes

---

LDPC codes were proposed by Gallager in his Ph.D. thesis [15]. 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” [54].

In this chapter, we review some basic concepts behind LDPC codes and iterative decoding, focusing on protograph-based codes. In Section 4.1, we give a formal definition of an LDPC code. In Section 4.2, we review BP decoding which is based on message passing. The protograph-based construction of LDPC codes is explained in 4.3. The basic idea behind the asymptotic analysis of LDPC codes via DE is outlined in Section 4.4. Finally, in Section 4.5 we briefly cover spatially-coupled LDPC codes, which are one of the code classes considered for the problem statement addressed in Papers A and B.

## 4.1 Introduction

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.,

$$\mathcal{C} = \{\mathbf{c} \in \{0, 1\}^n : \mathbf{H}\mathbf{c}^\top = \mathbf{0}\}, \quad (4.1)$$

where  $n > c$  and operations (i.e., additions and multiplications) 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  $R = d/n = 1 - c/n$ .

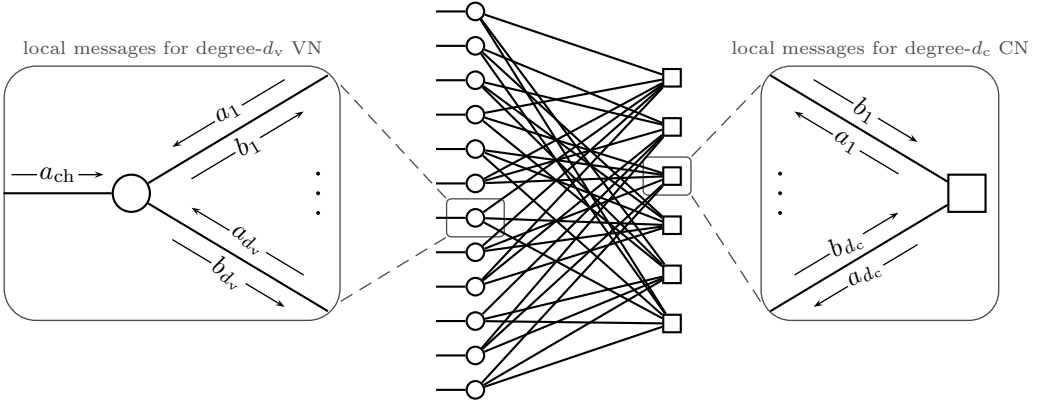
The definition in (4.1) does indeed apply to an arbitrary binary linear code with a given parity-check matrix  $\mathbf{H}$ . It is of course up to interpretation when exactly the matrix  $\mathbf{H}$  should be classified as sparse (and, hence, the resulting code should be classified as an LDPC code). As an example, consider the case where  $\mathbf{H}$  is such that each row contains exactly  $d_c$  ones and each column contains exactly  $d_v$  ones. Choosing  $n$  and  $c$  large compared to  $d_c$  and  $d_v$  then leads to a sparse matrix  $\mathbf{H}$ . The code defined by such a matrix  $\mathbf{H}$  is referred to as a *regular* LDPC code.

## 4.2 Iterative Belief Propagation Decoding

Consider the scenario where each bit in the codeword of an LDPC code is transmitted over an LLR channel  $f_{L|B}(\cdot|\cdot)$  (recall the definition of an LLR channel in Section 3.2). 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. The decoding is based on a graphical representation of the code. In particular, the parity-check matrix of an LDPC code can be represented by using a bipartite Tanner graph consisting of  $n$  variable nodes (VNs) and  $c$  constraint nodes (CNs), where the  $i$ -th CN is connected to the  $j$ -th VN if and only if  $h_{i,j} = 1$ . During the decoding process, the decoder tries to iteratively improve the accuracy of the initial 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 BP decoding, we refer the reader to [12, Ch. 5.3]. Here, we only briefly review the basic steps of the decoding algorithm. We use the following convention. Messages arriving at VNs are denoted by  $a$  and 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. 4.1. 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





**Figure 4.1:** Illustration of the messages involved in the iterative BP decoding algorithm.

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  $a_1, \dots, a_{d_v}$  are messages from CNs and  $a_{ch}$  corresponds to the channel LLR. The  $d_v$  outgoing messages  $b_1, \dots, b_{d_v}$  are computed according to

$$b_i = \sum_{\sim i} a_j + a_{ch}, \quad (4.2)$$

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

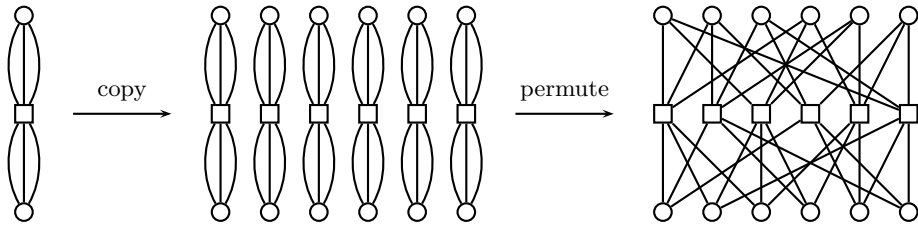
$$a_i = 2 \tanh^{-1} \left( \prod_{\sim i} \tanh(b_j/2) \right), \quad (4.3)$$

where the product is over the index set  $j \in \{1, \dots, d_c\}$  *excluding* the index  $i$ . Since the CN operation (4.3) is central to the analysis of LDPC codes under iterative decoding, it is very common to rewrite it in terms of the binary boxplus operator defined by

$$b_1 \boxplus b_2 = 2 \tanh^{-1} (\tanh(b_1/2) \tanh(b_2/2)). \quad (4.4)$$

The box-addition of an arbitrary number of terms is evaluated by recursively applying (4.4), 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

$$a_i = \bigboxplus_{\sim i} b_j. \quad (4.5)$$



**Figure 4.2:** Illustration of the protograph lifting procedure for  $\mathbf{P} = (3, 3)$  and  $M = 6$ .

The decoding process can now be described as follows. Set  $a_{\text{ch}}$  for all VNs to the corresponding channel LLR and set all other messages to 0. (As an example, in a BICM system, the channel LLRs are computed according to (3.2).) Then, repeat the following two steps. First, compute outgoing messages for all VNs according to (4.2). After that, compute the outgoing messages for all CNs according to (4.3). Stop if either a maximum number of iterations has been reached or the proper combination of the hard decisions on the messages

$$\sum_{j=1}^{d_v} a_j + a_{\text{ch}} \quad (4.6)$$

for all VNs forms a valid codeword.

### 4.3 Code 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 [20]. A protograph is 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 so-called 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. 4.2.

Designing codes via protographs has several practical advantages, e.g., a quasi-cyclic code is easily obtained by constraining the  $M$ -by- $M$  matrices to have a circulant structure. This in turn allows for hardware-efficient implementation [12, p. 263] suitable for high-speed optical communications [6]. Moreover, codes of different lengths can be obtained simply by adjusting the lifting factor.

## 4.4 Density Evolution

DE is a powerful tool to analyze the iterative decoding behavior and performance of LDPC codes in the limit of infinite block length [24]. 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  $\sigma^2$  of a symmetric Gaussian LLR channel) into a region where reliable decoding is possible and where it is not.

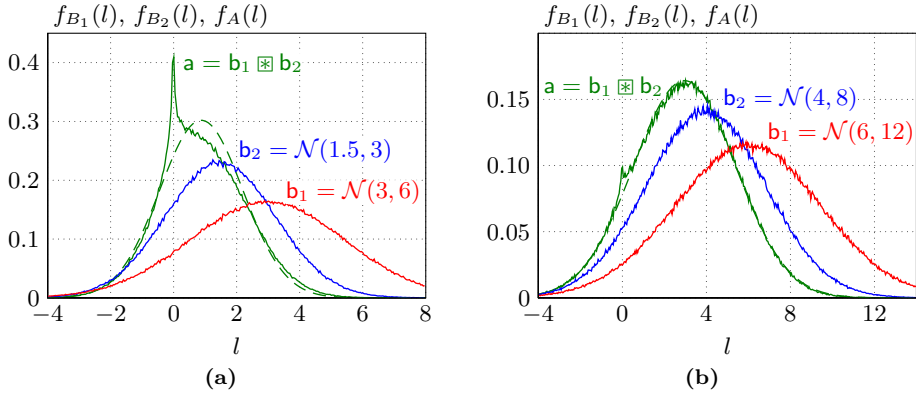
The main steps in the DE algorithm can be understood by considering the update equations (4.2) and (4.3) for the VNs and CNs, respectively. If we assume that the involved incoming messages are random variables, then they have a certain probability distribution or density. For example,  $a_{\text{ch}}$  is distributed according to the 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  $f_C(c) = f_A(a) \otimes f_B(b)$ , where  $\otimes$  denotes convolution. It is convenient to introduce the short notation  $\mathbf{a} \otimes \mathbf{b}$ , where  $\mathbf{a}$  and  $\mathbf{b}$  are placeholders for the densities of the random variables  $A$  and  $B$  [11]. Given the densities of the incoming messages, the densities of the outgoing messages can then be computed according to

$$\mathbf{b}_i = \bigotimes_{\sim i} \mathbf{a}_j \otimes \mathbf{a}_{\text{ch}}. \quad (4.7)$$

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  $\mathbf{b}_1$  and  $\mathbf{b}_2$  are processed according to the boxplus operation  $a = b_1 \boxplus b_2$ . In order to obtain the density  $\mathbf{a}$ , 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  $\mathbf{a}$ . This method is illustrated in Fig. 4.3, where it is shown how two consistent Gaussian densities “evolve” under the boxplus operation. A density  $\mathbf{a}$  is called a consistent Gaussian density<sup>1</sup> with parameter  $\sigma^2$  if  $A \sim \mathcal{N}(\sigma^2/2, \sigma^2)$ . As a short notation, one may introduce the operator  $\mathbf{a} = \mathbf{b}_1 \boxtimes \mathbf{b}_2$ , which is referred to as box-convolution [11]. In practice, the box-convolution of two densities can be implemented by using a look-up table approach [55]. Similar to (4.7), the densities of the outgoing CN messages can then be computed according to

$$\mathbf{a}_i = \bigboxtimes_{\sim i} \mathbf{b}_j. \quad (4.8)$$

<sup>1</sup>Note that the conditional distribution  $f_{L|B}(l|0)$  of a symmetric Gaussian LLR channel corresponds to a consistent Gaussian density.



**Figure 4.3:** Illustration of the box-convolution of two consistent Gaussian densities. The green dashed line corresponds to the consistent Gaussian approximation obtained via EXIT functions.

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. Due to the channel symmetry, one may assume the transmission of the all-zero codeword [12, p. 389]. The iterative decoding behavior can be predicted via DE as follows. Set  $\mathbf{a}_{\text{ch}}$  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 (4.7). After that, calculate the outgoing message densities for all CNs in the protograph according to (4.8). Stop if the error probability associated with the density

$$\bigotimes_{j=1}^{d_v} \mathbf{a}_j \boxtimes \mathbf{a}_{\text{ch}} \quad (4.9)$$

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

$$p_e(\mathbf{a}) = \int_{-\infty}^0 f_A(a) da, \quad (4.10)$$

or a maximum number of iterations is reached (decoding failure). In order to find the decoding threshold, we start from a channel quality where the decoding is predicted to be successful. The above procedure is then repeatedly applied for decreasing channel quality until the decoding fails.

### 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 [56]. Let us assume that the density  $\mathbf{a}$  fulfills the condition  $f_A(a)e^a = f_A(-a)$ . Then, the density can be associated with the MI measure

$$I(\mathbf{a}) = 1 - \int_{-\infty}^{\infty} f_A(a) \log_2(1 + e^{-a}) da. \quad (4.11)$$

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

$$I(\mathbf{b}_i) \approx \tilde{J} \left( \sum_{\sim i} \tilde{J}^{-1}(I(\mathbf{a}_j)) + \tilde{J}^{-1}(I(\mathbf{a}_{\text{ch}})) \right), \quad (4.12)$$

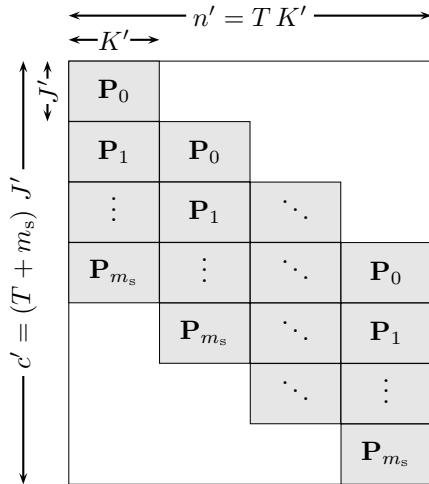
whereas, under the CN operation it evolves approximately according to

$$I(\mathbf{a}_i) \approx 1 - \tilde{J} \left( \sum_{\sim i} \tilde{J}^{-1}(1 - I(\mathbf{b}_j)) \right), \quad (4.13)$$

where  $\tilde{J}(x) = J(\sqrt{x})$ . These two equations can be motivated as follows. Eq. (4.12) is exact under the assumption that all incoming densities  $\mathbf{a}_1, \dots, \mathbf{a}_{d_v}$ , and  $\mathbf{a}_{\text{ch}}$  are consistent Gaussian densities. To see this, note that the convolution of two consistent Gaussian densities with parameters  $\sigma_1^2$  and  $\sigma_2^2$  is another consistent Gaussian density with parameter  $(\sigma_1^2 + \sigma_2^2)/2$ . Furthermore, if  $\mathbf{a}$  is a consistent Gaussian density with parameter  $\sigma^2$ , the operation  $\tilde{J}^{-1}(I(\mathbf{a}))$  simply returns  $\sigma^2$ . Without going into the details, (4.13) can be heuristically motivated by a duality property that holds for the binary erasure channel (BEC) [12, p. 415]. It is important to point out that (4.13) 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. 4.3 have been obtained using (4.13), where the resulting MI measure is plotted in the form of the associated consistent Gaussian density.

## 4.5 Spatially-Coupled LDPC Codes

Spatial coupling of regular LDPC codes has emerged as a powerful technique to construct capacity-achieving codes for a large class of channels using iterative BP decoding [14, 57]. The main idea is to make several copies of the Tanner graph that defines the regular 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



**Figure 4.4:** Illustration of the base matrix  $\mathbf{P}_{[T]}$  of a  $(J, K)$  regular, protograph-based SC-LDPC code.

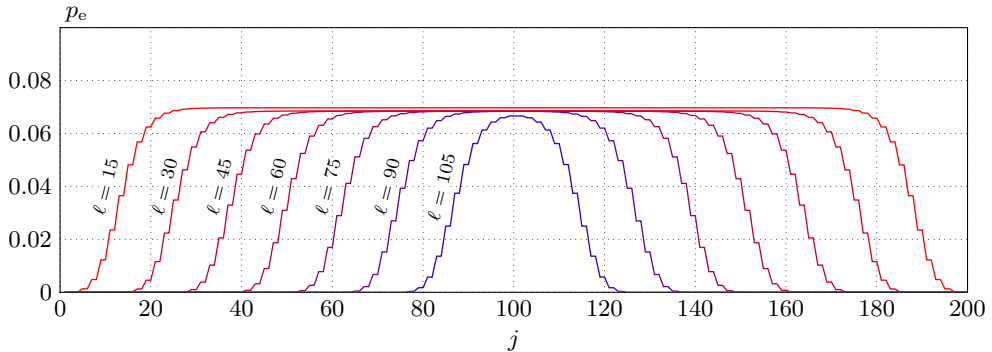
a way is a boundary effect due to slight irregularities at the two ends of the resulting Tanner graph.

In general, spatially-coupled LDPC codes have parity-check matrices with a band-diagonal structure, see, e.g., [57] for a formal definition. Here, we briefly introduce their construction via protographs [58], [59, Sec. II-B]. The base matrix  $\mathbf{P}_{[T]}$  of a  $(J, K)$  regular, protograph-based spatially-coupled LDPC code with termination length  $T$  can be constructed by specifying matrices  $\mathbf{P}_i$ ,  $0 \leq i \leq m_s$  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  $\mathbf{P}_{[T]}$  is constructed as shown in Fig. 4.4. From the dimensions of  $\mathbf{P}_{[T]}$  one can infer a design rate of  $R(T) = 1 - (T + m_s)J'/(TK')$ . As  $T$  grows large, the rate approaches  $R(\infty) = 1 - J'/K'$ .

Before continuing, it is insightful to recall the following statement from [60], where the design of irregular LDPC codes is studied. (VNs are referred to as message nodes and CNs are referred to as check 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 spatially-coupled LDPC codes, one can give a similar heuristic explanation for their



**Figure 4.5:** Illustration of the wave-like decoding behavior of spatially-coupled LDPC codes.

outstanding performance as follows (see [57] for a detailed explanation). By inspecting the structure of the base matrix in Fig. 4.4, one may verify that the CN degrees corresponding to the first and last couple of rows are 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. 4.5, we show the predicted bit error rates  $p_e$  via (approximate) DE for the coded bits corresponding to the  $j$ -th column of the spatially-coupled 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 Gaussian LLR channel with parameter  $\sigma^2 = 4$ . In the figure,  $\ell$  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 LDPC 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 [61]. In contrast, spatially-coupled 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 spatially-coupled LDPC, see [62] for a recent overview. For example, the price to pay for the wave-like decoding behavior is a rate loss with respect to regular codes that are defined by the protograph  $\mathbf{P} = \sum_i \mathbf{P}_i$ .





---

## Generalized Product Codes

---

The practical implementation of BP decoding for LDPC codes at very high data rates poses a significant challenge. This motivates the use of coding schemes that are less complex (potentially sacrificing some performance). One particular example of such a coding scheme is discussed in this chapter, namely the use of GPCs in combination with iterative BDD.

We start in Sections 5.1 and 5.2 by reviewing GLDPC codes and PCs, respectively. GPCs can be regarded as a subclass of GLDPC codes and a formal definition is given in Section 5.3 together with several examples of GPCs. In Section 5.4, we discuss the assumed channel model and describe the decoding of GPCs via iterative BDD. Finally, in Section 5.5 we briefly outline and compare two approaches to perform an asymptotic DE analysis for GPCs.

## 5.1 Generalized Low-Density Parity-Check Codes

In the previous chapter, we have seen that the parity-check matrix  $\mathbf{H}$  of an LDPC code can be represented in terms of a bipartite Tanner graph where coded bits and parity-check equations are represented by VNs and CNs, respectively. An edge in the graph indicates if a certain bit participates in a certain parity-check equation (i.e., a row in  $\mathbf{H}$ ). The code can then be defined as the set of all VN bit assignments such that the parity-check equations corresponding to the CNs are satisfied.

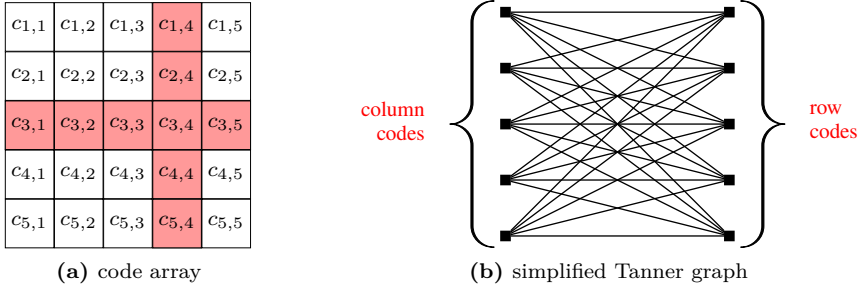
This concept can be generalized by interpreting the CNs not just as simple parity-check equations but as component code constraints corresponding to smaller block codes (e.g., Hamming or BCH codes). In order to specify the code, one uses a binary matrix  $\mathbf{\Gamma} \in \{0,1\}^{m \times n}$ . This matrix is interpreted as the adjacency matrix for an associated bipartite graph. The graph consists again of  $n$  VNs representing coded bits (one for each column in  $\mathbf{\Gamma}$ ) and  $m$  CNs representing component code constraints (one for each row in  $\mathbf{\Gamma}$ ). An edge between a VN and CN indicates if a certain bit participates in a certain constraint enforced by a component code. In addition to the matrix  $\mathbf{\Gamma}$  (or the corresponding graph), one also needs to specify  $m$  component codes  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_m$  that are associated with the  $m$  CNs in the graph. The overall code is defined as the set of all VN bit assignments that satisfy all component code constraints. The code thus defined is referred to as a GLDPC code.

Assuming that all component codes  $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_m$  are linear codes, the resulting overall GLDPC code is also a linear code [12]. This implies that the code can alternatively be represented by using a parity-check matrix  $\mathbf{H}$  and a corresponding Tanner graph where CNs correspond again to simple parity-check equations. It should thus be stressed that the term “generalized” refers to the extended graphical representation by means of “generalized” CNs. The main reason for introducing these generalized CNs is that they add a layer of abstraction into the code representation. This may be helpful when constructing new codes or devising decoding algorithms. For example, assume that we have at our disposal an efficient decoding algorithm for some linear block code. The GLDPC code framework then allows us to build longer and potentially more powerful codes by using this block code as a building block. When decoding the overall code, we may take advantage of the available component code decoder, thereby allowing for an efficient overall decoding scheme.

## 5.2 Product Codes

PCs are one of the first examples that use the idea of building longer codes from shorter component codes [21]. In the following, we review the code construction and the representation as a GLDPC code.

Let  $\mathcal{B}$  be some binary linear block code of length  $n_{\mathcal{B}}$ . A PC is defined as the set of  $n_{\mathcal{B}} \times n_{\mathcal{B}}$  arrays such that each row and each column in the array is a codeword in  $\mathcal{B}$ .



**Figure 5.1:** Illustrations for a PC with  $n_B = 5$ .

The code array for  $n_B = 5$  is visualized in Fig. 5.1(a), where we use a two-dimensional indexing to refer to the coded bits  $c_{i,j}$  for  $i, j \in \{1, 2, \dots, n_B\}$ . In the figure, one particular row/column constraint is highlighted in red.

A PC can be interpreted as a GLDPC code with a very structured Tanner graph representation [63]. In particular, for  $n_B = 5$ , the adjacency matrix  $\mathbf{\Gamma}$  of the Tanner graph is given by

$$\mathbf{\Gamma} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (5.1)$$

where the first and last five rows of  $\mathbf{\Gamma}$  correspond to the row and column component codes, respectively. We further have  $\mathcal{B}_1 = \mathcal{B}_2 = \dots = \mathcal{B}_{10} = \mathcal{B}$  in the GLDPC code representation since all component codes are identical.

From the structure of  $\mathbf{\Gamma}$ , it can be seen that all VNs in the corresponding Tanner graph have degree 2. This is due to the fact that each coded bit (i.e., each entry in the array) is protected by precisely two component codes. In this case, it turns out to be convenient to represent these degree-2 VNs by simple edges. Fig. 5.1(b) shows the simplified Tanner graph corresponding to the  $\mathbf{\Gamma}$ -matrix in (5.1). With this simplified representation, the Tanner graph corresponds to a complete bipartite graph: There exist two types of CNs representing the row and column component codes, respectively, and each CN of one type is connected to *all* CNs of the other type. This gives rise to exactly  $n_B^2$  edges, each of which corresponds to one coded bit in the array.

As a side note, one subtlety that arises in the graphical representation of a GLDPC code is that, strictly speaking, the edges emanating from each CN should also be labeled with the corresponding component code bit positions [63]. If the component code has length  $n_B$  (i.e., the corresponding CN has degree  $n_B$ ), then in principle any of the  $n_B!$  possible permutations of the bit positions can be assigned to the edges. Choosing different assignments may result in an overall code with different properties, e.g., rate and minimum distance [63]. The reason this is not an issue for LDPC codes is that a parity-check equation is invariant under the order in which the bits appear in the equation. Moreover, for PCs and related code structures, the code is typically unambiguously defined by an accompanying array description.

## 5.3 Generalized Product Codes

With the background about GLDPC codes and PCs described in the previous two sections, we are now in the position to give a formal definition of a GPC. In particular, we adopt the convention in [64] and define a GPC as any GLDPC code whose Tanner graph representation consists exclusively of degree-2 VNs.<sup>1</sup> This implies that, similar to PCs, each coded bit is protected by exactly two component codes. However, the bits may not necessarily be arranged in the form of a rectangular array.

We remark that this terminology is nonstandard and GPCs are sometimes also referred to simply as “product-like” codes [9]. In the following, we review several examples of GPCs that are relevant for the appended papers.

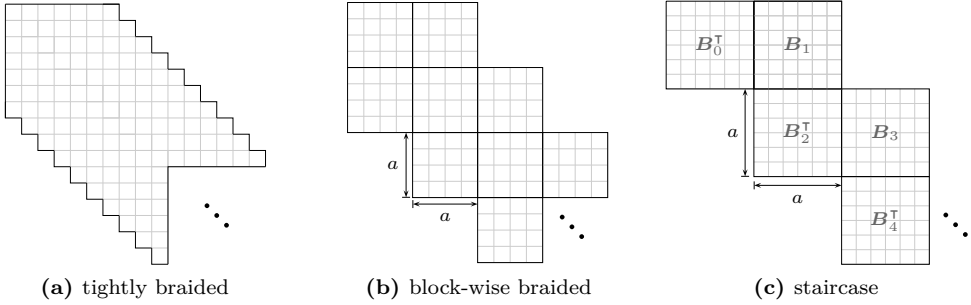
### 5.3.1 Braided Codes

Braided codes are proposed in [18] as “convolutional (or sliding) versions” of PCs. Similar to PCs, the code construction is based on a two-dimensional array where bits are placed in the array under the constraint that rows and columns form codewords in some component code. Depending on the type of component code, the resulting braided code is referred to either a braided block code or a braided convolutional code. In this thesis, we focus exclusively on the case where the component codes are block codes, and for simplicity we refer to the resulting codes simply as braided codes.

Braided codes come in different flavors, depending on the precise structure of the code array. Figs. 5.2(a) and (b) show two examples which are referred to as tightly braided codes and block-wise braided codes, respectively. In both cases, the array is conceptually infinite, i.e., one assumes the transmission of a continuous data stream. For tightly braided codes, the array structure consists of rows and columns that are shifted by one array element at each step. The simplified Tanner graph for a tightly braided code is shown for example in [18, Fig. 2(b)]. For the block-wise braided codes the array consists

---

<sup>1</sup>In [10], a slightly different definition of a GPC is given.



**Figure 5.2:** Differently shaped code arrays for the various GPCs discussed in Section 5.3.

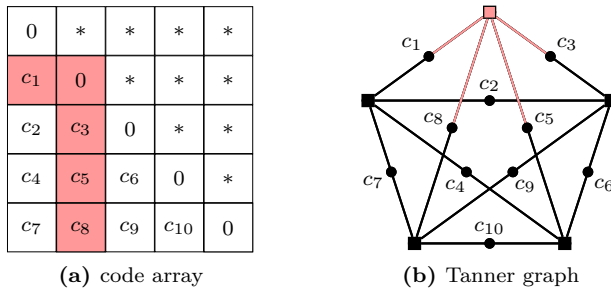
of three block ribbons with block size  $a$ . For example, the block size in Fig. 5.2(b) is given by  $a = 4$ .

Braided codes can be classified as spatially-coupled PCs or alternatively as convolutional PCs. (In [18], the term “GLDPC convolutional codes” is used instead.) Braided codes have been explicitly considered for the use in fiber-optical communication systems for example in [10]. The code construction we propose in Paper D includes block-wise braided codes as special cases, thereby enabling an asymptotic analysis for these codes.

### 5.3.2 Staircase Codes

Staircase codes are proposed in [9] as a new class of error-correcting codes for optical transport networks by “combining ideas from convolutional and block coding”. Given a component code  $\mathcal{B}$  of length  $n_{\mathcal{B}}$ , a staircase code is defined as the set of all matrix sequences  $\mathbf{B}_i \in \{0,1\}^{a \times a}$ , where  $a = n_{\mathcal{B}}/2$  and  $i = 0, 1, 2, \dots$ , such that the rows in  $[\mathbf{B}_{i-1}^T \mathbf{B}_i]$  for all  $i \geq 1$  form valid codewords in  $\mathcal{B}$ . The matrix  $\mathbf{B}_0$  is assumed to be initialized to the all-zero matrix. The code array that corresponds to this definition has a characteristic staircase structure and is shown in Fig. 5.2(c), where  $n_{\mathcal{B}} = 12$  and  $a = 6$ . Similar to braided codes, staircase codes can also be classified as instances of spatially-coupled PCs.

When comparing staircase and braided codes, it should be mentioned that for braided codes, only soft-decision decoding of the component codes is studied in [18]. The authors in [9] consider this to be “unsuitable for high-speed fiber-optic communications”. Motivated by the excellent performance of staircase codes under iterative hard-decision decoding, the design of braided codes intended for fiber-optical communications is considered for example in [10]. The resulting braided code is found to be “competitive” to the staircase code designed in [65] suggesting that the performance of staircase and braided codes can be quite similar. This conclusion is also confirmed in Paper E, where we compare staircase codes and braided codes. For the considered parameters, both code classes perform almost identically in terms of waterfall performance and error floor.



**Figure 5.3:** Illustrations for an HPC with  $n_B = 5$ . In the array, “\*” means “equal to the transposed element”. The highlighted array elements illustrate one particular code constraint, which is also highlighted in the Tanner graph.

### 5.3.3 Half-Product Codes

Consider again the simplified Tanner graph representation of a PC shown in Fig. 5.1(b) which corresponds to a complete bipartite graph. The graph structure is a consequence of the array description of a PC which appears to be quite natural. On the other hand, in [66], Justesen points out that if one focuses instead on the graph, “it is not clear why a bipartite graph is preferable” and that in this case “the more natural concept [...] is a complete graph”. Such a complete Tanner graph indeed appears as one of the first examples in [63]. The resulting codes, however, have received very little attention in the literature thus far and, to the best of our knowledge, Justesen was the first to provide an interpretation of the graph structure in terms of a code array [66, Sec. III-B]. He also found a direct connection to conventional PCs which we briefly review in the following based on the descriptions in [67, Sec. IV-B].

Consider a conventional  $n_B \times n_B$  PC based on a component code  $\mathcal{B}$  of length  $n_B$ . Based on this PC, a new code is formed by keeping only codeword arrays that have zeros on the diagonal and are symmetric, in the sense that the array is equal to its transpose. Since the diagonal and upper (or lower) triangular part of the array do not contain “useful” bits, they can be ignored so that the effective length of the resulting code is given by  $m = \binom{n_B}{2}$ . Such codes are referred to as half-product codes (HPCs) in [66]. The Tanner graph of an HPC corresponds exactly to a “complete Tanner graph” with  $n_B$  CNs where each CN is connected to all other CNs through a VN. As an example, Figs. 5.3(a) and (b) show the code array and Tanner graph for an HPC assuming that  $n_B = 5$ , where one particular code constraint is highlighted in red.

HPCs and PCs are compared for example in [64], where it is shown that HPCs can have larger normalized minimum distance than PCs. It is also possible to extend the above definition to other classes of GPCs, i.e., other array shapes. For example, for the arrays shown in Figs. 5.2(a) and (b), it poses no conceptual problem to enforce the additional constraint (additional to the usual row and column component code constraints) that

the array should be symmetric with a zero diagonal. The resulting codes belong to the class of symmetric GPCs [64]. In general, symmetric GPCs use symmetry to reduce the block length of a GPC while employing the same component code [64].

## 5.4 Iterative Bounded-Distance Decoding

We assume that the intended “target” channel for GPCs is the binary symmetric channel (BSC) where each bit is flipped independently of all other bits with a certain crossover probability  $p$ . In the context of fiber-optical communications, this channel can be motivated by considering PM-QPSK transmission (i.e., independent binary modulation in both quadratures and polarizations) in combination with a minimum-distance detector that provides a hard decision about the transmitted signal. The BSC can be shown to be exact in the case where nonlinear transmission effects are ignored.

For the BSC, there exist very efficient algebraic BDD for several linear block codes, e.g., BCH codes. BDD corrects all error patterns with Hamming weight up to the error-correcting capability  $t$  of the code. The idea is then to use such codes as component codes for a GPC and decode the overall code by iteratively performing BDD of all component codes. While this decoding scheme is suboptimal, it has been shown to offer excellent performance provided that the code rate of the GPC is relatively high. For example, the staircase code designed in [9] has rate  $R = 239/255 \approx 0.937$  and performs only about 0.56 dB away from the channel capacity of the BSC under iterative BDD. Moreover, the decoder data flow requirements in this case are estimated to be more than two orders of magnitude smaller compared to the requirements for a comparable LDPC code with message-passing decoding [9].

The main conceptual problem that arises in the theoretical analysis of iterative BDD for GPCs over the BSC is that the component decoders may miscorrect in which case a successful (component) decoding is declared but the found codeword is not the correct one. Such miscorrections introduce additional bit errors into the iterative decoding process which makes a rigorous analysis challenging. One approach to avoid this issue is to ignore miscorrections entirely and assume the use of so-called idealized BDD. Such a decoder either outputs the correct codeword or declares a decoding failure. The assumption of idealized BDD over the BSC is conceptually equivalent to transmission over the BEC. For the BEC, each bit is erased independently of all other bits with a certain erasure probability  $p$ . In that case, the error-correcting capability  $t$  of the component code is interpreted as the erasure-correcting capability. The BEC is used in Papers D–F in order to allow for a rigorous theoretical analysis.

## 5.5 Density Evolution

The purpose of this section is to discuss two different approaches to perform an asymptotic performance analysis for GPCs under iterative BDD assuming transmission over

the BEC. The first approach is based on an ensemble argument and uses the ideas and techniques discussed in [24, 68]. The second approach directly targets a sequence of deterministically constructed GPCs and is based on the work in [69, 70]. Here, we only give a high-level overview of the basic approach idea in both cases. The main goal of this section is to contrast the two approaches and discuss potential advantages and disadvantages.

As a side note, the reader should be aware that the term “density evolution” for the asymptotic analysis may be somewhat misleading. This is because the parameter of interest in this case turns out to correspond to a simple probability and not a density.

### 5.5.1 Code Ensembles

The first and most widely used approach to perform an asymptotic analysis for GPCs is to define a “suitable” code ensemble. In the following, we review one such ensemble which is taken from [71] as an illustrative example. Let  $\mathcal{B}$  be a component code of length  $n_{\mathcal{B}}$ . Assume that there are  $m$  CNs of degree  $n_{\mathcal{B}}$  (each corresponding to the component code  $\mathcal{B}$ ) and  $mn_{\mathcal{B}}/2$  VNs of degree 2. In order to construct the Tanner graph, it is helpful to imagine that there are  $mn_{\mathcal{B}}$  half-edges emanating from all the CNs and VNs, respectively. One particular code in the ensemble is defined by the Tanner graph obtained by connecting these half-edges using a uniform random permutation. The ensemble is defined as the set of all codes defined by all possible choices of permutations.

The above ensemble definition is conceptually not much different from the definition of the regular LDPC code ensemble [24], except that the graph consists of generalized CNs. The asymptotic scenario considers the limit  $m \rightarrow \infty$ , i.e., one increases the number of VNs and CNs in the graph, while using a fixed component code. The principal conclusions from [24] (see also [68]) can be paraphrased as follows.

1. Asymptotically, the fixed-depth neighborhood of a randomly chosen VN or CN in the Tanner graph becomes a tree with high probability.
2. Assuming that the neighborhood is tree-like, the analysis of the expected iterative decoding performance is drastically simplified to the extent that it can usually be accomplished “in closed form” using a recursive expression.
3. There exists a concentration phenomenon that ensures that with high probability, a particular code taken uniformly at random from the ensemble will have actual performance close to the expected decoding performance computed in the previous step.

By combining these three conclusions, one can then study and analyze the asymptotic performance of the GPC ensemble defined above.

This ensemble approach appears to be appealing at first. For instance, one may be interested in using a fixed component code  $\mathcal{B}$  in practice (e.g., a triple-error correcting BCH code of length  $n_{\mathcal{B}} = 1023$ ) and the ensemble analysis enables an asymptotic analysis



for precisely this component code. Moreover, the approach is not limited to the analysis of iterative BDD but can be used to analyze the ensemble performance for a variety of different channels and iterative decoding algorithms [24].

On the other hand, for certain applications, the ensemble approach may not be appropriate. In particular, assume that we are interested in implementing a GPC with a fixed deterministic structure, e.g., a PC. In that case, it is not clear to what extent the ensemble approach is useful. In order to illustrate this, note that a PC is contained in the ensemble defined above for  $m = 2n_{\mathcal{B}}$  and a particular interleaver permutation. However, the ensemble approach only makes a statement about codes that are sampled uniformly at random from the ensemble and not particular ones. It is therefore not clear if the expected ensemble performance is somehow indicative for the performance of a PC.

### 5.5.2 Deterministic Codes

For sequences of *deterministic* GPCs, choosing a meaningful asymptotic scenario is not straightforward. For simplicity, let us restrict ourselves to “square” PCs, i.e., the case where both the row and column component codes have length  $n_{\mathcal{B}}$ . For the asymptotic scenario, we consider sequences of PCs with increasing array size. Increasing the array size, however, has *two* consequences. First, it leads to an increase in the number of component codes and thereby an increase in the number of CNs in the underlying Tanner graph. This is similar to the ensemble approach described in the previous subsection. However, increasing the array size also changes the component codes. In particular, the length  $n_{\mathcal{B}}$  of each component code does not remain fixed but it increases. This is different from the ensemble approach where the component code properties (including the length) are assumed to remain fixed.

When dealing with sequences of component codes with increasing lengths, one should also specify what happens to the other component code parameters, in particular the erasure-correcting capability  $t$ . There are essentially two possibilities that have been studied in the literature before. In the following, we briefly review both cases.

In the first case, one assumes that the erasure-correcting capability increases linearly with  $n_{\mathcal{B}}$ , i.e.,  $t$  is assumed to be a function of  $n_{\mathcal{B}}$ . In particular, one may assume that each component code can correct a fixed fraction  $\alpha \in (0, 1)$  of erasures in terms of its block length, i.e., we have  $t = \alpha n_{\mathcal{B}}$ . This case has been studied in [69]. The analysis is based on Chernoff bounds and the conclusion is quite simple: If one has access to component codes with  $t = \alpha n_{\mathcal{B}}$ , then, asymptotically, it is pointless to construct a PC out of these component codes since both the PC and each component code can operate reliably if (and only if) the erasure probability satisfies  $p < \alpha$ . In other words, the product construction is useless in this asymptotic scenario.

In the second case, one assumes that the erasure-correcting capability remains fixed. This is reasonable from a practical perspective because the complexity of algebraic BDD for BCH codes increases drastically with  $t$ . With this assumption, however, one finds that for any fixed erasure probability  $p$ , the decoding will fail with high probability in

the limit  $n_B \rightarrow \infty$ . The reason for this is simple. Even if we choose  $p$  very small, the expected number of erasures per row and column in the PC array grows linearly with  $n_B$  and will eventually far exceed the assumed finite erasure-correcting capability  $t$  of each component code.

In summary, both cases lead to somewhat unsatisfactory answers. The important thing to realize in the second case, however, is that we are essentially considering sequences of PCs with code rate approaching 1 as  $n_B \rightarrow \infty$ . Due to this vanishing redundancy, it should not come as a surprise that operating at any fixed erasure probability is futile asymptotically. With this in mind, a meaningful asymptotic analysis can be performed by allowing the erasure probability to decay slowly as  $c/n_B$  for some fixed constant  $c > 0$ . In other words, we are not considering a fixed channel anymore but the channel is changing in accordance with the strength of the PC. This approach is pioneered for PCs in [69, 70] and it heavily relies on properties from random graph theory. In Paper D, this approach is extended to a large class of deterministic GPCs based on properties of so-called inhomogeneous random graphs [26].

---

## 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. The conclusions are structured in terms of the two topics that are investigated in this thesis.

### **Bit Mapper Optimization for Spectrally-Efficient Systems (Papers A and B)**

In Papers A and B, we study coded transmission systems that operate at high spectral efficiency over fiber-optical links without inline dispersion compensation. Assuming a linear coherent receiver, the classical AWGN channel with a modified SNR expression is used as a design channel.

In Paper A, we propose a method to optimize the bit mapper that determines the allocation of the coded bits from the channel encoder to the modulation or labeling bits of the signal constellation. The proposed method applies to an arbitrary protograph-based LDPC code. Compared to previous approaches for protograph-based codes, we use a fractional allocation between the modulation bits and the VN classes in the protograph. This allows for an unrestricted matching between different protographs and modulation formats. We also discuss the bit mapper optimization for spatially-coupled LDPC codes that are based on protographs and decoded using a windowed decoder. Our 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. We also provide a simulative verification for a nonlinear transmission scenario based on the SSFM which confirms the accuracy of the assumed channel model.

In Paper B, we consider the bit mapper optimization for spatially-coupled codes in

more detail. In particular, we consider both spatially-coupled LDPC codes based on protographs and the spatially-coupled PC ensemble in [27]. In the first case, standard BP decoding is assumed, while in the second case we assume an iterative hard-decision decoding algorithm which is significantly less complex. For both cases, we consider terminated and tail-biting spatially-coupled codes. In general, for terminated spatially-coupled codes with long spatial length, the bit mapper optimization only results in marginal performance improvements, suggesting that a sequential or random allocation is close to optimal. On the other hand, an optimized allocation can significantly improve the performance of tail-biting spatially-coupled codes. Such codes do not possess an inherent termination boundary. In this case, the unequal error protection offered by the modulation bits can be used to create an artificial termination boundary that induces a wave-like decoding behavior. Unlike for terminated spatially-coupled codes, the wave effect in this case does not come at the price of a rate loss: tail-biting spatially-coupled codes have the same design rate as the underlying uncoupled codes.

An interesting direction for future work could be to study the bit mapper optimization for irregular spatially-coupled LDPC codes. In this thesis, we have assumed the use of spatially-coupled LDPC codes that are based on regular LDPC codes. While such codes are capacity-achieving, it has been shown in [72] that irregular spatially-coupled LDPC codes can offer some advantages when taking into consideration practical restrictions for parameters, e.g., a limited number of decoding iterations. For the irregular case, there is an opportunity to optimize the bit allocation by only considering the different VN degrees at one spatial position instead of considering the entire code chain. An optimized bit mapper allocation could for example lead to a decreased number of decoding iterations that are necessary to reach a certain target error rate.

### **Analysis and Design of Deterministic Generalized Product Codes (Papers C–F)**

In Paper C, we study parameter optimization for staircase codes assuming an extrinsic iterative hard-decision decoding algorithm. The optimization is based on a DE analysis for a spatially-coupled PC ensemble that shares structural similarities with staircase codes. Using this approach, staircase code parameters can be found at a significantly reduced optimization time compared to a simulation-based approach. We also propose an extension of staircase codes by allowing for multiple code constraints per row and column in the corresponding array description. This construction leads to larger staircase block sizes and steeper waterfall performance that better matches the predicted DE performance.

The optimization approach in Paper C is, however, only heuristically motivated. In particular, the DE analysis does not directly apply to staircase codes. This issue is addressed in Paper D, where we consider an asymptotic DE analysis for deterministic GPCs with a fixed Tanner graph structure. The main conclusion from Paper D can be summarized as follows. There exists a large class of deterministic GPCs for which a rigorous asymptotic DE analysis assuming iterative BDD over the BEC is possible.

---

For example, the proposed code construction and analysis can be used to study the asymptotic performance of

- conventional PCs [21], staircase codes [65], and block-wise braided codes [18],
- GPCs employing a mixture of different component codes with varying erasure-correcting capabilities such as irregular PCs [73, 74],
- symmetric GPCs [64] such as HPCs [66] which can be seen as symmetric subcodes of certain GPCs.

The DE analysis in Paper D can also be used to study different decoding schedules that are practically relevant such as row/column decoding for PCs or windowed decoding of staircase and braided codes.

In Papers E and F, we use the proposed code construction and DE analysis to study some relevant classes of GPCs in more detail. In Paper E, we provide a comparison between staircase codes, block-wise braided codes, and the symmetric subcode of a block-wise braided code which is referred to as a half-braided code. Our results indicate that half-braided codes can outperform both staircase codes and braided codes in the waterfall performance, at a lower error floor and decoding delay.

In Paper F, we consider spatially-coupled PCs in more detail. In particular, we revisit the spatially-coupled PC ensemble that is used in Paper C for the parameter optimization of staircase codes. It is shown in [27] that this ensemble can approach the capacity of the BSC at high rates. Motivated by this result, our main goal is to compare the ensemble performance to the performance of deterministic GPCs with a spatially-coupled structure via their respective DE equations. For the BEC, it is shown that there exists a family of deterministic braided codes that performs asymptotically equivalent to the ensemble. It is also shown that there exists a related but structurally simpler family of braided codes with essentially the same performance, even though the DE equations are not exactly equivalent. Lastly, we consider spatially-coupled PCs with a mixture of different component codes. In that case, the conclusion is that employing such component code mixtures for spatially-coupled PCs is not beneficial from an asymptotic point of view.

In the following, we suggest two potentially interesting topics for future work. The first topic concerns the asymptotic performance of deterministic GPCs over the BSC. While the obtained results for the BEC can be used to approximate the code performance over the BSC, it would be desirable to rigorously characterize the BSC performance including the effect of decoder miscorrections. For example, the equivalence between the spatially-coupled PC ensemble in [27] and the deterministic spatially-coupled PCs in Paper F only holds for the BEC. This is not sufficient to show that the same deterministic codes are also capacity-achieving over the BSC. On the other hand, the proof in [27] relies partially on the fact that the impact of miscorrections becomes small if the error-correction capability increases. Similar conclusions should also hold for deterministic GPCs, at least qualitatively.

Another potentially interesting topic is the investigation of the finite-length scaling behavior of deterministic GPCs, similar to the scaling analysis for certain LDPC code ensembles presented in [75]. A generalization to deterministic GPCs may give a relatively complete picture of the performance of these codes under iterative decoding.

---

## 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, 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] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, Jul. 1948.
- [9] B. P. Smith, A. Farhood, A. Hunt, F. R. Kschischang, and J. Lodge, “Staircase codes: FEC for 100 Gb/s OTN,” *J. Lightw. Technol.*, vol. 30, no. 1, pp. 110–117, Jan. 2012.

- [10] Y.-Y. Jian, H. D. Pfister, K. R. Narayanan, R. Rao, and R. Mazahreh, “Iterative hard-decision decoding of braided BCH codes for high-speed optical communication,” in *Proc. IEEE Glob. Communication Conf. (GLOBECOM)*, Atlanta, GA, 2014.
- [11] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.
- [12] W. Ryan and S. Lin, *Channel Codes Classical and Modern*. Cambridge University Press, 2009.
- [13] 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.
- [14] 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, 2005.
- [15] R. Gallager, “Low-density parity-check codes,” Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, 1963.
- [16] 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, 2012.
- [17] A. Yedla, Y.-Y. Jian, P. S. Nguyen, and H. D. Pfister, “A simple proof of threshold saturation for coupled scalar recursions,” in *Proc. 7th Int. Symp. on Turbo Codes and Iterative Information Processing (ISTC)*, Gothenburg, Sweden, 2012.
- [18] A. J. Felström, D. Truhachev, M. Lentmaier, and K. S. Zigangirov, “Braided block codes,” *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2640–2658, Jul. 2009.
- [19] E. Agrell and M. Karlsson, “Power-efficient modulation formats in coherent transmission systems,” *J. Lightw. Technol.*, vol. 27, no. 22, pp. 5115–5126, Nov. 2009.
- [20] J. Thorpe, “Low-density parity-check (LDPC) codes constructed from protographs,” *IPN Progress Report 42-154*, JPL, 2005.
- [21] P. Elias, “Error-free coding,” *IRE Trans. Inf. Theory*, vol. 4, no. 4, pp. 29–37, Apr. 1954.
- [22] “Forward error correction for high bit-rate DWDM submarine systems,” ITU-T Recommendation G.975.1, 2004.
- [23] L. M. Zhang and F. R. Kschischang, “Staircase codes with 6% to 33% overhead,” *J. Lightw. Technol.*, vol. 32, no. 10, pp. 1999–2002, May 2014.



- 
- [24] T. J. Richardson and R. L. 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.
- [25] Y.-Y. Jian, H. D. Pfister, and K. R. Narayanan, "Approaching capacity at high-rates with iterative hard-decision decoding," *arxiv:1202.6095v3 [cs.IT]*, May 2015. [Online]. Available: <http://arxiv.org/abs/1202.6095.pdf>
- [26] B. Bollobás, S. Janson, and O. Riordan, "The phase transition in inhomogeneous random graphs," *Random Structures and Algorithms*, vol. 31, no. 1, pp. 3–122, Aug. 2007.
- [27] Y.-Y. Jian, H. D. Pfister, and K. R. Narayanan, "Approaching capacity at high rates with iterative hard-decision decoding," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Cambridge, MA, 2012.
- [28] G. P. Agrawal, *Nonlinear Fiber Optics*, 4th ed. Academic Press, 2006.
- [29] M. Secondini and E. Forestieri, "The nonlinear Schrödinger equation in fiber-optic systems," *Riv. Mat. Univ. Parma*, vol. 8, pp. 69–98, Sep. 2008.
- [30] 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.
- [31] G. P. Agrawal, *Fiber-optic Communication Systems*, 4th ed. Wiley-Interscience, 2010.
- [32] 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.
- [33] M. I. Yousefi, "Information transmission using the nonlinear Fourier transform," Ph.D. dissertation, University of Toronto, 2013.
- [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] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, 1949.
- [40] 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.
- [41] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.
- [42] 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.
- [43] 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.
- [44] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] L. Szczecinski and A. Alvarado, *Bit-Interleaved Coded Modulation: Fundamentals, Analysis, and Design*, 1st ed. Wiley-Interscience, 2014.
- [49] 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.

- 
- [50] 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, 2007.
  - [51] 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.
  - [52] 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.
  - [53] 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.
  - [54] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.
  - [55] 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.
  - [56] 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.
  - [57] 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.
  - [58] D. G. M. Mitchell, M. Lentmaier, and D. J. Costello, Jr., "AWGN channel analysis of terminated LDPC convolutional codes," in *Proc. Information Theory and Applications Workshop (ITA)*, La Jolla, CA, 2011.
  - [59] 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.
  - [60] M. G. Luby, M. Mitzenmacher, M. A. Shokrollah, and D. A. Spielman, "Analysis of low density codes and improved designs using irregular graphs," in *Proc. AMC Symp. on Theory of Computing (STOC)*, New York, USA, 1998.
  - [61] 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.

- [62] D. J. Costello, L. Dolecek, T. E. Fuja, J. Kliever, D. G. M. Mitchell, and R. Smarandache, “Spatially coupled sparse codes on graphs – theory and practice,” *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 168–176, Jul. 2014.
- [63] R. Tanner, “A recursive approach to low complexity codes,” *IEEE Trans. Inf. Theory*, vol. 27, no. 5, pp. 533–547, Sep. 1981.
- [64] H. D. Pfister, S. K. Emmadi, and K. Narayanan, “Symmetric product codes,” in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, 2015.
- [65] B. P. Smith, “Error-correcting codes for fibre-optic communication systems,” Ph.D. dissertation, University of Toronto, 2011.
- [66] J. Justesen, “Performance of product codes and related structures with iterated decoding,” *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 407–415, Feb. 2011.
- [67] C. Häger, H. D. Pfister, A. Graell i Amat, F. Brännström, and E. Agrell, “A deterministic construction and density evolution analysis for generalized product codes,” in *Proc. Int. Zurich Seminar on Communications (IZS)*, Zürich, Switzerland, 2016.
- [68] M. G. Luby, M. Mitzenmacher, and M. A. Shokrollahi, “Analysis of random processes via and-or tree evaluation,” in *Proc. 9th Annual ACM-SIAM Symp. Discrete Algorithms*, San Francisco, CA, 1998.
- [69] M. Schwartz, P. Siegel, and A. Vardy, “On the asymptotic performance of iterative decoders for product codes,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, SA, 2005.
- [70] J. Justesen and T. Høholdt, “Analysis of iterated hard decision decoding of product codes with Reed-Solomon component codes,” in *Proc. IEEE Information Theory Workshop (ITW)*, Tahoe City, CA, 2007.
- [71] Y.-Y. Jian, “On the analysis of spatially-coupled GLDPC codes and the weighted min-sum algorithm,” Ph.D. dissertation, Texas A&M University, 2013.
- [72] A. Leven and L. Schmalen, “Status and recent advances on forward error correction technologies for lightwave systems,” *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2735–2750, Aug. 2014.
- [73] S. Hirasawa, M. Kasahara, Y. Sugiyama, and T. Namekawa, “Modified product codes,” *IEEE Trans. Inf. Theory*, vol. 30, no. 2, pp. 299–306, Mar. 1984.
- [74] M. Alipour, O. Etesami, G. Maatouk, and A. Shokrollahi, “Irregular product codes,” in *Proc. IEEE Information Theory Workshop (ITW)*, Lausanne, Switzerland, 2012.
- [75] A. Amraoui, A. Montanari, T. J. Richardson, and R. L. Urbanke, “Finite-length scaling for iteratively decoded LDPC ensembles,” *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 473–498, Feb. 2009.