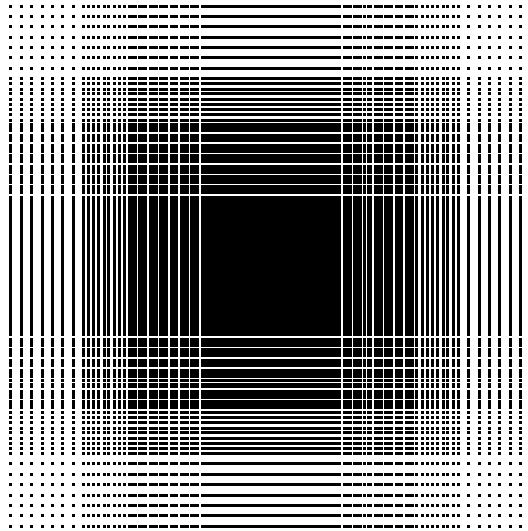


Coding and Modulation for Power and Bandwidth Efficient Communication



Harm S. Cronie

Composition of the graduation committee:

Chairman and Secretary:

Prof. Dr. Ir. A.J. Mouthaan

Promotor:

Prof. Dr. Ir. C.H. Slump

Internal members:

Prof. Dr. Ir. W. van Etten

Prof. Dr. Ir. B. Nauta

External members:

Dr. Ir. J.H. Weber (Delft University of Technology)

Dr. Ir. F.M.J. Willems (Eindhoven University of Technology)

Prof. Dr. R.L. Urbanke (Ecole Polytechnique Federale de Lausanne)

The research in this thesis was carried out at the Signals & Systems group of the University of Twente, Enschede, The Netherlands.

Copyright © 2008 by Harm S. Cronie

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written consent of the copyright owner.

ISBN: XX-XXX-XXXX-X

Printed by Ipskamp Printpartners, Enschede, The Netherlands. Typeset in L^AT_EX.

CODING AND MODULATION FOR POWER AND
BANDWIDTH EFFICIENT COMMUNICATION

DISSERTATION

to obtain

the doctor's degree at the University of Twente,

on the authority of the rector magnificus,

prof. dr. W.H.M. Zijm,

on account of the decision of the graduation committee,

to be publicly defended

on Thursday, September 11, 2008 at 16:45

by

Harm Stefan Cronie

born on 3 December 1978

in Utrecht, The Netherlands

Abstract

We investigate methods for power and bandwidth efficient communication. The approach we consider is based on powerful binary error correcting codes and we construct coded modulation schemes which are able to perform close to the capacity of the channel.

We focus on the additive white Gaussian noise channel. For this channel a Gaussian distribution maximizes mutual information and signal shaping has to be used to get close to capacity. We investigate a simple method of signal shaping based on the superposition of binary random variables. With multi-stage decoding at the receiver, the original coding problem is transformed into a coding problem for a set of equivalent binary-input output-symmetric channels. It is shown that with the method signal constellations can be designed for high spectral efficiencies which have their capacity limit within 0.1 dB of the capacity of the AWGN channel. Furthermore, low-density parity-check codes are designed for the equivalent binary channels resulting from this modulation method. We show how to approach the constrained capacity limit of the signal constellations we design very closely.

A downside of multistage decoding is that multiple binary error-correcting codes are used. We show how one can limit the number of error-correcting codes used by merging bit-interleaved coded modulation and signal shaping. This results in a coded modulation scheme which is able to approach the capacity of the AWGN channel closely for any spectral efficiency.

These coded modulation methods transform the coding problem for the original channel into a coding problem for a set of binary channels. Depending on the design of the modulation scheme these channels are symmetric or not. We show how to characterize channel symmetry in general and how these results can be used to design coded modulation schemes resulting in a set of symmetric binary channels.

Contents

Abstract	v
Contents	ix
1 Introduction	1
1.1 Information Theory	2
1.2 Coded Modulation	3
1.3 Channels with Additive Gaussian Noise	4
1.3.1 Discrete-time AWGN Channel	5
1.4 State of the Art and Summary of the Results	7
1.4.1 Binary Channel Inputs	7
1.4.2 Multilevel Codes and Bit-Interleaved Coded Modulation	7
1.4.3 Non-binary LDPC Codes	8
1.4.4 Overview of results	8
1.5 Outline	10
2 Superposition Modulation on the Gaussian Channel	11
2.1 Introduction	11
2.2 Modulation and Coding	13
2.2.1 Modulation by Superposition	15
2.2.2 Multilevel Encoding with Multistage Decoding	16
2.2.3 Equivalent Binary Channels	18
2.3 Signal Constellations	20
2.3.1 Signal Constellation Properties	21
2.3.2 Properties of Constellations generated by Superposition	21
2.3.3 Families of Signal Constellations	22
2.4 Error-control Coding with Binary LDPC Codes	31

2.4.1	Equivalent Binary Channels	33
2.4.2	Computation of Log-likelihood Ratios	35
2.4.3	LDPC Codes	38
2.4.4	Analysis and Design of LDPC Codes	38
2.5	Design Examples and Simulation Results	43
2.5.1	Decoding Order and Equivalent Binary Channels	44
2.5.2	Illustration of EXIT Chart Design	44
2.5.3	LDPC Codes for the Equivalent Binary Channels	48
2.6	Conclusions	55
2.7	Acknowledgments	55
3	Signal Shaping for Bit-Interleaved Coded-Modulation	57
3.1	Introduction	57
3.2	Coded Modulation	60
3.2.1	Introduction	60
3.2.2	Signal Constellations and Modulation	61
3.2.3	Coding Schemes and Decoding	62
3.3	Signal Shaping for Bit-interleaved Coded Modulation	65
3.3.1	Signal Constellations for BICM	66
3.3.2	Signal Constellations for Shaping	69
3.3.3	Shaping of PAM Constellations for BICM	70
3.3.4	Numerical Optimization of Constellations for BICM	73
3.4	Error-Control Coding with Binary Codes	76
3.4.1	Log-likelihood Ratios and Channel Symmetry	77
3.4.2	Equivalent Binary Channels for Modulation Maps	82
3.4.3	Binary LDPC Codes	85
3.5	Design Examples and Numerical Results	86
3.5.1	PAM-LDPC Codes	86
3.5.2	Shaped PAM-LDPC Codes	90
3.6	Conclusions and Final Remarks	94
4	Symmetric Channels and Coded Modulation	95
4.1	Introduction	95
4.2	Preliminaries	97
4.2.1	Information Theory	97
4.2.2	Geometry and Algebra	100
4.3	Memoryless Discrete-Input Symmetric Channels	102
4.3.1	Group Characterization of Channel Symmetry	103
4.3.2	Representation of Cyclic Symmetry Groups in \mathbb{R}^n	104

4.3.3	Channel Symmetry for Channels with a Binary Input . . .	105
4.4	Applications to Coded Modulation	108
4.4.1	Modulation for Channels with Additive Noise	108
4.4.2	Symmetric Channels and Symmetric Constellations . . .	110
4.4.3	Design Example for the AWGN Channel	113
4.5	Open Questions and Future Research	116
	Bibliography	116

Chapter 1

Introduction

The subject of this thesis is reliable communication over general channels close to the theoretical limits. The theoretical limit is given by the Shannon capacity of the channel and for many practical channel models the Shannon limit is a function of transmission power and signal bandwidth. Once these two are fixed we wish to achieve reliable communication while transmitting at a rate close to the Shannon limit. In this thesis we investigate coding and modulation methods for power and bandwidth efficient communication.

Our work is inspired by the success of binary sparse graph codes on binary channels. Low-density parity-check (LDPC) codes [1] can be constructed for which it can be proven that they are capable of achieving capacity on the binary erasure channel [2]. Furthermore, for the binary-input additive white Gaussian noise (BIAWGN) channel, LDPC codes have been designed which perform very close to the theoretical limit¹. Similar results can be obtained for other families of sparse graph codes such as repeat accumulate (RA) codes.

We investigate methods for achieving a near-capacity performance on non-binary channels with binary error-correcting codes. We focus on high spectral efficiencies where the use of binary signaling suffers from a large loss in capacity. In the end the goal is to construct schemes which perform within tenths of a decibel from capacity at high spectral efficiencies.

¹In [3] LDPC codes are designed which have a threshold within 0.0045 dB of the BIAWGN channel. Moreover, a low bit-error rate is achieved within 0.04 dB of the capacity of the channel.

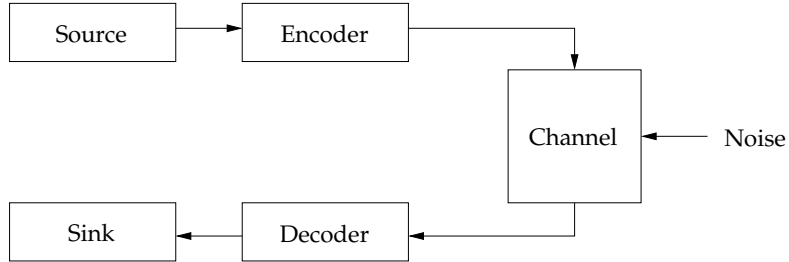


Figure 1.1: Block diagram of communication system.

1.1 Information Theory

One of the major contributions of Shannon's *A Mathematical Theory of Communication* [4] is the stochastic model of the communication system. A physical communication system is divided in several parts as shown in Figure 1.1. The *source* provides us with information to be transmitted across the *channel*. The *encoder* and *decoder* have to be designed in such a way that information can be transmitted across the channel efficiently and reliably. In information theory mathematical models are derived for the source and the channel and these models are usually stochastic in nature.

We assume that the source can be modeled as follows. The source provides a sequence $\{S_i\}_{i=1}^n$ of independent and identically distributed (i.i.d.) random bit variables. Moreover, we assume that the distribution of each of the S_i is uniform. This stochastic process provides us with a maximum entropy and there is no need for source encoding and decoding. Hence the encoder and decoder in Figure 1.1 are a *channel encoder* and a *channel decoder*.

A fundamental channel model is the discrete memoryless channel (DMC). Consider a DMC with input alphabet \mathbb{X} and output alphabet \mathbb{Y} . The channel is defined by a probability mass function $f_{Y|X}(y|x)$ where $f_{Y|X}(y|x)$ denotes the probability of observing y as a channel output when x is transmitted. For a DMC the mutual information between the channel input X and channel output Y is given by

$$I(Y; X) = \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} f_{Y|X}(y|x) f_X(x) \log_2 \frac{f_{Y|X}(y|x)}{\sum_{x' \in \mathbb{X}} f_{Y|X}(y|x') f_X(x')}, \quad (1.1)$$

where f_X defines the distribution over the input alphabet \mathbb{X} . The *capacity* of the

channel is defined as the maximum value of $I(Y; X)$ where the maximization is performed over all distributions on the channel input

$$C = \max_{f_X} I(Y; X). \quad (1.2)$$

The operational characterization of the channel capacity is given by a coding theorem. The capacity of the channel is the maximum amount of information we can transmit across the channel with arbitrary reliability.

Although, the DMC is a very simple channel model, it shares the important features with the channels we are interested in. Given a channel we associate with the channel input a stochastic process. This process is disturbed by noise and the output of the channel is a stochastic process also. Next, we associate a quantity $I(Y; X)$ to the channel whose operational meaning is related to the amount of information we can transmit reliably on the channel. Furthermore, the capacity of the channel is denoted by C and it is related to the maximum rate at which we can transmit information reliably. Note that not all channels fit this picture and Figure 1.1 is a simplified model.

Usually we only have limited options to change the characteristics of the channel. However, there are often degrees of freedom in designing the input process such that a performance close to the theoretical limit C becomes possible at acceptable computational complexity. We investigate low-complexity schemes for *coded modulation* which have the potential to approach the capacity of several channels very closely.

1.2 Coded Modulation

Consider a channel on which we wish to communicate reliably. We assume that the channel has capacity C which is achieved for some optimal input stochastic process. A capacity achieving coding scheme should essentially lead to this optimal input process. However, from a practical point of view this process is often difficult to realize with error-control coding. We cannot simply use a random codebook by sampling from the optimal input process. The reason for this is that description complexity and decoding complexity would be too high.

For certain codeword alphabets error-correcting codes can be defined which allow for low-complexity storage, encoding and decoding. We investigate methods to generate a channel input process based on a binary process. The characteristics of the resulting stochastic process at the output of the channel

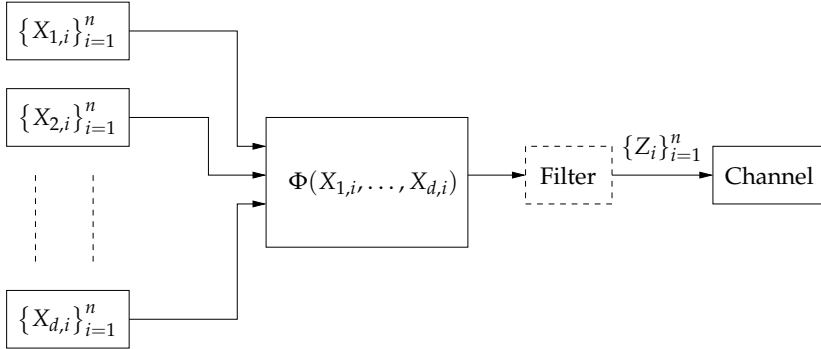


Figure 1.2: Illustration of Coded Modulation.

should be such that capacity is approached. We use *modulation* to transform the source process into a suitable channel input process. An overview of the method we is illustrated in Figure 1.2. We start with a set of d independent binary stochastic processes. These processes can be obtained from a common binary i.i.d. source. Next, a map Φ is applied to the realizations of the random variables. We refer to Φ as the modulation map and it transforms a tuple of bits to a channel input symbol. Furthermore, the resulting sequence of Z_i can be passed to a linear filter to further modify the properties of the channel input process.

The use of multiple binary processes is related to the encoding and decoding scheme used. In the end we can view the system as a collection of d binary channels for which we can employ binary codes. The choice of the number of processes and the decoding scheme employed has several consequences. First, some schemes are easier to analyze and design. Second, the performance, encoding complexity and decoding complexity depend on the number of processes and decoding method applied.

1.3 Channels with Additive Gaussian Noise

Our main example is the additive white Gaussian noise (AWGN) channel. Our results can be extended to other channels and an initial result in this direction is presented in [5] where we investigate coding for the continuous-time AWGN channel with intersymbol interference.

1.3.1 Discrete-time AWGN Channel

The discrete-time memoryless AWGN channel with input X and output Y is defined by

$$Y = X + N, \quad (1.3)$$

where N is zero-mean Gaussian noise with variance σ^2 . The density of N is given by

$$f_N(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{n^2}{2\sigma^2}}. \quad (1.4)$$

The channel is defined by its transition probability density function $f_{Y|X}$

$$f_{Y|X}(y|x) = f_N(y-x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-x)^2}{2\sigma^2}}. \quad (1.5)$$

We denote the amount of energy expended per channel use by E_s and it is given by

$$E_s = E[X^2], \quad (1.6)$$

where $E[\cdot]$ denotes mathematical expectation. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{E_s}{\sigma^2}. \quad (1.7)$$

The mutual information between X and Y is given by

$$I(Y; X) = H(Y) - H(Y|X) = H(Y) - H(N), \quad (1.8)$$

and its maximum value is achieved for a Gaussian distribution on X which leads to the following capacity formula

$$C = \frac{1}{2} \log_2(1 + \text{SNR}). \quad (1.9)$$

To achieve capacity on the AWGN channel the distribution of the channel input X should be Gaussian. The use of another input distributions leads to a loss in capacity. This is illustrated in Figure 1.3 which shows a plot of the capacity of the AWGN channel and the achievable rate when we restrict the input to a discrete pulse-amplitude modulation (PAM) constellation with 64

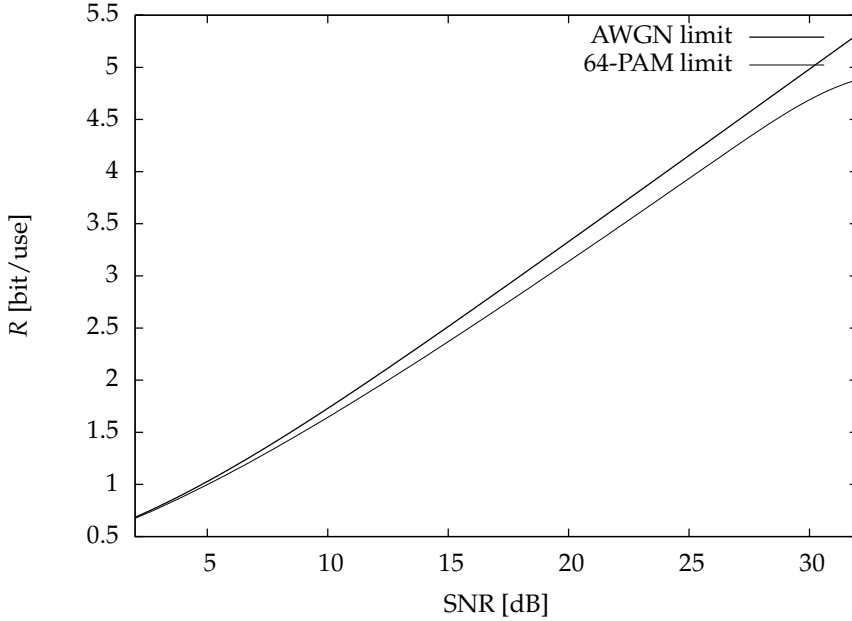


Figure 1.3: Capacity of the AWGN channel.

symbols. The achievable rate when the input is constrained to a signal constellation is called the *constrained constellation capacity*. A PAM constellation with 64 symbols is defined by

$$\mathcal{S} = \left\{ -2^6 + 2i - 1 \mid i = 1, 2, 3, \dots, 2^6 \right\}, \quad (1.10)$$

and the elements of the constellation symbols are selected with equal probability. The figure shows that for low SNRs there is hardly a loss compared to a Gaussian channel input. However, for higher SNRs there is a substantial loss. Techniques to bridge this gap are called *signal shaping* techniques and the main theme of this thesis is how to bridge this gap with the coded modulation scheme of Figure 1.2.

1.4 State of the Art and Summary of the Results

In this section we give a short overview of the state of the art in modulation and coding for the AWGN channel. We do not intend to give an exhaustive overview, but present a summary of prior work and give a comparison with our work.

1.4.1 Binary Channel Inputs

For low SNRs where the capacity of the AWGN is low, the loss resulting from using binary channel inputs is small. At a transmission rate of 0.5 bit/use, the loss with respect to capacity is only 0.18 dB and we can resort to binary signaling schemes.

Turbo codes are introduced in [6] and they perform within 0.5 dB of the constrained capacity limit while transmitting at a rate of 0.5 bit/use. In [3] LDPC codes are designed which perform extremely close to capacity. At a transmission rate of 0.5 bit/use, the distance to the constrained capacity limit is only 0.04 dB.

1.4.2 Multilevel Codes and Bit-Interleaved Coded Modulation

In [7] capacity approaching schemes based on LDPC codes are investigated for transmission over the AWGN channel. The authors use multilevel coding (MLC) [8] and bit-interleaved coded-modulation (BICM) [9] together with binary LDPC codes. The focus is on conventional signal constellations and signal shaping is not employed. At a transmission rate of 1 bit/use with a 4-PAM constellation and a channel block length of 10^6 , a low bit-error rate is achieved within 0.14 dB of the constrained constellation capacity.

In [10] trellis shaping is combined with the use of binary LDPC codes. At a transmission rate of 2 bit/use and a channel block length of 10^5 a low BER is achieved within 0.81 dB of the capacity of the AWGN channel.

In [11] a method for signal shaping is proposed and combined with turbo codes. For spectral efficiencies of 1 bit/use, 1.5 bit/use and 2 bit/use a low BER is achieved at a distance of 1.0 dB, 1.2 dB and 1.4 dB of the capacity of the AWGN channel. In Chapter 2 we show that with the method of signal shaping presented in [11], we can achieve a good performance very close to the capacity of the AWGN channel.

1.4.3 Non-binary LDPC Codes

In [12] non-binary LDPC codes are designed for coded modulation on the AWGN channel. One of the motivations of this paper is that for power and bandwidth efficient communications binary LDPC codes are not that suitable. For transmission on the AWGN channel spectral efficiencies of 3 bit/use and 4 bit/use are considered. Shaped signal constellations are designed by a method proposed in [13]. The code designed for 3 bit/use has a channel block length of $1.8 \cdot 10^5$ and a low bit-error rate is achieved at a distance of 0.56 dB from the capacity of the AWGN channel. The distance to the constrained constellation limit is 0.3 dB. The code designed for 4 bit/use has a channel block length of 10^5 and a low bit-error rate is achieved at a distance of 1 dB from the capacity of the AWGN channel. The distance to the constrained constellation limit is 0.72 dB.

1.4.4 Overview of results

To illustrate the performance of these results and compare with our results, we have plotted the capacity of the AWGN channel in Figure 1.4. The figure also shows the constrained constellation capacity of a 256-PAM constellation. Furthermore, we have indicated the SNR and rate points which are achieved by state-of-the-art schemes presented in literature and the schemes we present. The block length is denoted by N and it is equal to the number of channel input symbols. Furthermore, the SNR and rate points are defined as the SNR where the scheme achieves a bit-error rate $< 10^{-5}$.

The figure shows the performance of the non-binary LDPC codes from [12] and a trellis shaped code from [10]. Furthermore, in Chapter 2 we investigate modulation by superposition combined with multilevel coding. The figure shows the performance of two schemes which are designed in Chapter 2. In Chapter 3 we introduce shaped PAM-LDPC codes and the figure shows the performance of these codes.

At a rate around 5 bit/use, we present two schemes which operate very close to the capacity of the AWGN channel. We have not found any schemes in literature transmitting at such a high spectral efficiency. At a rate around 3 bit/use and 4 bit/use, the performance of the shaped PAM-LDPC codes is comparable to the performance of the non-binary LDPC codes. However, PAM-LDPC codes are based on binary LDPC codes and in general decoding complexity for these codes will be less. The schemes we present for a transmission at a rate around 2 bit/use perform slightly better than the trellis shaped

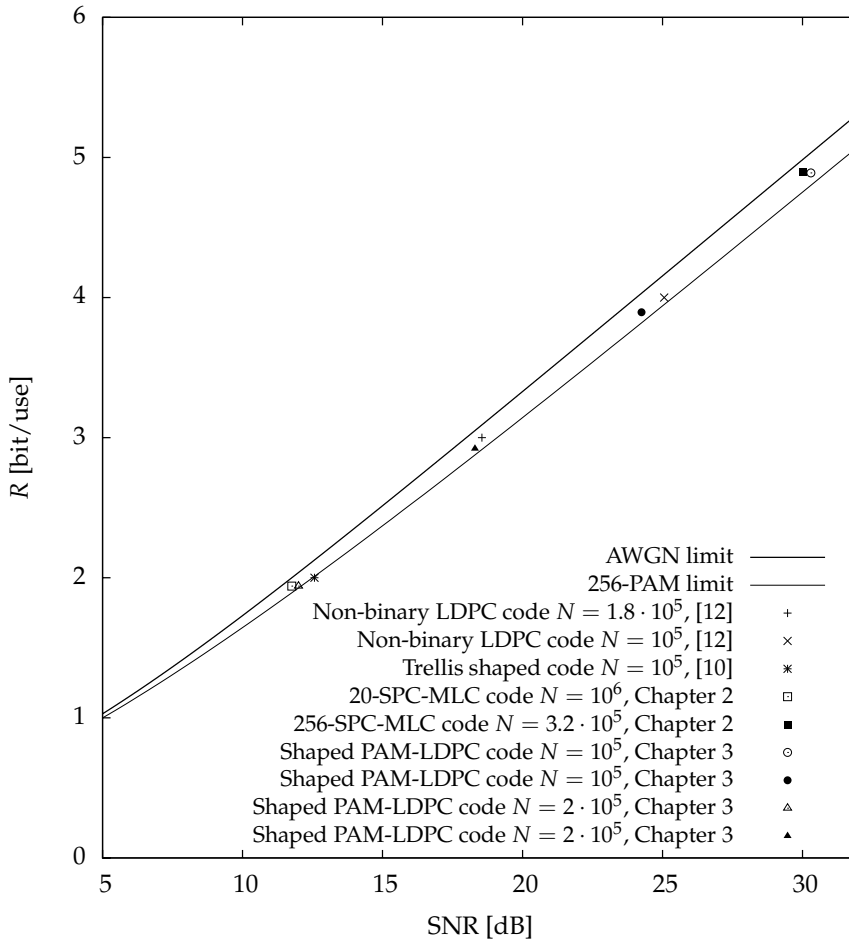


Figure 1.4: Coded modulation schemes for the AWGN channel.

code which is presented in [10]. We conclude that the schemes we present perform very close to the capacity of the AWGN channel.

1.5 Outline

The outline of this thesis is as follows. In Chapter 2 we investigate the use of *superposition modulation* for the design of signal constellations. In this case the modulation map is simply a scaled addition over the real numbers. We show that signal constellations can be designed which have a constrained capacity within 0.1 dB of the capacity of the AWGN channel for target rates between 2 bit/use to 5 bit/use. Furthermore, we show that the use of superposition modulation transforms the coding problem for the AWGN channel into a coding problem for a set of binary memoryless *symmetric* channels for which powerful binary codes can be designed.

The disadvantage of the approach followed in Chapter 2 is that in the context of Figure 1.2 the required value of d becomes high for higher spectral efficiencies. In Chapter 3 we show how to prevent this by merging bit-interleaved coded-modulation and multilevel coding. With this method we are able to achieve a good performance for a relatively small value of d (3 or 4) for any spectral efficiency.

The use of superposition modulation results in a set of equivalent symmetric binary channels. In Chapter 4 we investigate the concept of channel symmetry in more detail. We show how channel symmetry is related to the properties of the output space of the channel. As an application we show how the modulation map Φ can be chosen such that the equivalent binary channels are symmetric. This leads to a rich family of modulation maps suitable for coded modulation on the AWGN channel. The work presented in Chapter 4 is not to be seen as a completed piece of research. However, we feel that it is sufficiently mature to be included. An argument in favor for this is that the partial results we provide lead to an interesting application.

Chapter 2

Superposition Modulation on the Gaussian Channel

2.1 Introduction

In this chapter, we consider power- and bandwidth efficient communication over the discrete-time memoryless additive white Gaussian noise (AWGN) channel. The goal is to achieve reliable communication at a rate close to the capacity of the channel for high spectral efficiencies where the use of binary signaling incurs a large loss in rate. In this case one has to resort to so-called signal shaping methods to get close to capacity. A restriction to signal constellations with a uniform spacing and an equiprobable selection of the constellation symbols leads to a maximum loss of 1.53 dB compared to a Gaussian channel input [14]. A so-called *shaping gain* is available.

Power- and bandwidth efficient communication with signal shaping has been studied by several authors. A comprehensive overview of modulation and coding for general Gaussian channels can be found in [14]. Most methods are either based on non-equiprobable signaling and non-uniform signaling or multi-dimensional signal constellations. The former approach considers the problem at hand from a modulation point of view and the latter approach from a coding point of view.

The use of multi-dimensional signal constellations is closely related to the concept of lattice codes [15], [16], [17]. An essential observation is that coding

and shaping gain can be separated when the dimensionality of the constellation tends to infinity. Recent research on lattice codes shows that the capacity of the AWGN channel can be achieved with lattice codes under suboptimal lattice decoding [18], [19]. However, from a complexity point of view suboptimal lattice decoding is only feasible for relatively small lattices.

In non-equiprobable signaling, methods are devised to generate channel inputs with a non-uniform probability distribution [20], [21]. The main issue here is how to choose the distribution in the first place and how to generate channel inputs from this distribution keeping in mind that the source usually provides uniformly distributed bits. In non-uniform signaling the channel inputs have a non-uniform spacing [13] and design issues here are how to choose the actual spacing. Methods to design these signal constellations are proposed in [13]. These methods can be combined with binary error-correcting codes. Two well-known schemes are bit-interleaved coded-modulation [9] and multilevel coding [8]. These schemes have the potential to provide reliable communication with feasible encoding and decoding complexity.

Some recent research has focused on the combination of powerful binary error-correcting codes and signaling methods. In [7], [10] low-density parity-check (LDPC) codes are combined with conventional pulse-amplitude modulation (PAM) constellations in a multilevel coding (MLC) context. The analysis and design of LDPC codes is simplified for binary-input output-symmetric (BIOS) channels. However, the use of MLC does not necessarily lead to symmetric channels at the bit level. The analysis and design of LDPC codes is more involved in this case. Moreover, in [12] the main motivation for using non-binary LDPC codes is that for power- and bandwidth efficient modulation the channels at the bit level are not symmetric. However, analysis and design of non-binary LDPC codes is more complex and decoding complexity is increased.

In this chapter we investigate the use of a conceptually very simple modulation method which allows one to generate signal constellations with a non-uniform spacing and a non-equiprobable distribution on the constellation symbols. The method has its roots in the work of Imai et al. on multilevel coding [8]. The method is easily combined with binary error-correcting codes to provide reliable communication. We show that if one uses a MLC approach with multistage decoding, the original problem of achieving capacity on the AWGN channel reduces to achieving capacity on a set of *binary-input output-symmetric channels*. Hence it is more or less straightforward to analyze and design binary LDPC codes to get close to the capacity of the AWGN channel once a proper signal constellation is designed. We show that one can get very close to the ca-

capacity of the AWGN channel for high signal-to-noise ratios with *binary* LDPC codes.

The outline of this chapter is as follows. In Section 2.2 we introduce the modulation method and show how to combine it with binary block codes. In Section 2.3 we consider the design of signal constellations and present a few design examples of signal constellations for a high spectral efficiency. In Section 2.4 we consider the use of binary LDPC codes on the binary channels defined by the signal constellations. Moreover, in this section we derive some properties of these binary channels which are relevant for the analysis and design of LDPC codes. In Section 2.5, we present design examples and simulation results. We end with conclusions in Section 2.6.

2.2 Modulation and Coding

We consider power and bandwidth efficient communication over the AWGN channel which is defined by

$$Y = X + N, \quad (2.1)$$

where the channel input X is disturbed by the random variable N which has a zero-mean Gaussian distribution with variance σ^2

$$f_N(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{n^2}{2\sigma^2}}. \quad (2.2)$$

The energy expended per channel use E_s is equal to the mathematical expectation of X^2

$$E_s = E \left[X^2 \right], \quad (2.3)$$

where the mathematical expectation is denoted by $E[\cdot]$. The capacity of the AWGN channel is achieved for a Gaussian distribution on X and it is given by the well-known formula

$$C = \frac{1}{2} \log_2 (1 + \text{SNR}) \quad \text{bit/use}, \quad (2.4)$$

where SNR is the signal-to-noise ratio and is defined as

$$\text{SNR} = \frac{E_s}{\sigma^2}. \quad (2.5)$$

In practical communication systems we transmit a symbol Z from a discrete alphabet \mathcal{S} . The set \mathcal{S} is called the *signal constellation* and its elements

are *constellation symbols*. Moreover, we define a probability measure P_S on the elements of \mathcal{S} , where $P_S(z)$ denotes the probability that Z is equal to z

$$P_S(z) = \Pr[Z = z] \text{ for } z \in \mathcal{S}. \quad (2.6)$$

Now, the channel output Y is given by

$$Y = Z + N. \quad (2.7)$$

The achievable rate is upper-bounded by the so-called *constrained constellation capacity* $I(Z; Y)$, which is the mutual information between Z and Y . The goal is to design \mathcal{S} and P_S in such a way that $I(Z; Y)$ is as close to C as possible. However, once we have designed \mathcal{S} and P_S , it is not straightforward to come up with a method of error-control coding which results in this signal constellation with the corresponding probability distribution and has feasible encoding and decoding algorithms.

On the other hand, it is not difficult to generate a near-Gaussian distribution which comes close to the optimal input distribution for the AWGN channel. One way to generate a Gaussian distribution is by adding independent and identically-distributed (i.i.d.) random variables. Let X_1, \dots, X_d denote a sequence of uniform i.i.d. random bit variables taking values in $\{-1, 1\}$ ¹. Next, we define a random variable Z as

$$Z = \frac{1}{\sqrt{d}} \sum_{i=1}^d X_i. \quad (2.8)$$

The distribution of Z is binomial and when we let $d \rightarrow \infty$ the distribution of Z converges to the Gaussian distribution by the central limit theorem. We investigate the use of this method to generate signal constellations for power- and bandwidth efficient communication over the AWGN channel.

The idea of superimposing bits is not new and is sometimes referred to as *superposition coding*. In [8] multilevel coding is introduced where the output of d independent binary encoders is summed. Moreover, in [11] and [22] the authors show that for $d = 2$ and $d = 3$ and low spectral efficiencies, the method can be combined with turbo codes leading to a low bit-error rate within 1.4 dB of the capacity of the AWGN channel. We elaborate on this idea and show that for a whole range of spectral efficiencies we can design signal constellations

¹Throughout this chapter binary random variables will take values in $\{-1, 1\}$. Algebraic operations on these variables are the algebraic operations defined on the real numbers.

with a constrained capacity close to the capacity of the AWGN channel. Furthermore, we show that superposition coding reduces the problem of achieving the capacity on the AWGN channel to achieving the capacity on a set of equivalent binary-input output-symmetric channels. For these binary channels LDPC codes can be designed such that an overall near-capacity performance is achieved.

2.2.1 Modulation by Superposition

Let X_1, \dots, X_d be a tuple of independent random bit variables where each bit takes values in $\{-1, 1\}$. The distribution of X_i for $i = 1, \dots, d$ is defined by $P_{X_i}(x_i)$

$$P_{X_i}(x_i) = \Pr[X_i = x_i]. \quad (2.9)$$

A channel input Z is generated by a scaled addition of these random bit variables

$$Z = \sum_{i=1}^d \alpha_i X_i, \quad (2.10)$$

where the α_i are constants taken from \mathbb{R} . The α_i define the signal constellation \mathcal{S}

$$\mathcal{S} = \left\{ z \mid z = \sum_{i=1}^d \alpha_i x_i, x_1 \in \{-1, 1\}, \dots, x_d \in \{-1, 1\} \right\}. \quad (2.11)$$

The probability that a constellation symbol $z \in \mathcal{S}$ is selected is given by

$$P_{\mathcal{S}}(z) = \sum_{x_1} \dots \sum_{x_d} \left(\prod_{i=1}^d P_{X_i}(x_i) \right) \mathbb{1}_{\{z\}} \left(\sum_{i=1}^d \alpha_i x_i \right), \quad (2.12)$$

where $\mathbb{1}_{\{z\}}$ is the set indicator function which for a set A is defined as

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases} \quad (2.13)$$

The distribution of X_1, \dots, X_d induces a distribution on the elements of \mathcal{S} . In what follows we will choose the distribution of X_1, \dots, X_d as the uniform distribution. The reason for this is that in the end we are interested in using binary linear codes for which the ensemble is defined by a uniform distribution

on the codeword symbols. We generate a channel input Z by a scaled addition of d uniform i.i.d. random bit variables

$$Z = \sum_{i=1}^d \alpha_i X_i. \quad (2.14)$$

The signal constellation is defined by (2.11) and the distribution of the constellation symbols which is defined by (2.12) reduces to

$$P_S(z) = \frac{1}{2^d} \sum_{x_1} \cdots \sum_{x_d} \mathbb{1}_{\{z\}} \left(\sum_{i=1}^d \alpha_i x_i \right). \quad (2.15)$$

The α_i determine the constellation geometry, the distribution of the constellation symbols and the mapping from bits to constellation symbols. The mapping from bits to constellation symbols can be injective or not. In case the map is not injective $P_S(z)$ can be a non-uniform distribution. In Section 2.3 we discuss the properties of the signal constellations generated by (2.14) in more detail. Next, we turn to error-control coding.

2.2.2 Multilevel Encoding with Multistage Decoding

To combine modulation by superposition with error-control coding, we consider the mutual information between Y and X_1, \dots, X_d which we can express as

$$I(Y; (X_1, \dots, X_d)) = I(Y; X_1) + I(Y; X_2 | X_1) + \dots + I(Y; X_d | X_1, \dots, X_{d-1}). \quad (2.16)$$

This is the chain rule of mutual information. This identity suggests a multilevel encoding procedure with multistage decoding at the receiver [8], [23]. Consider a set of d binary error-correcting codes, where we denote the code at level i by \mathcal{C}_i . We assume that the codeword bits are represented on the real numbers by 1 and -1 . The rate of \mathcal{C}_i is denoted by r_i and the length of each code is n . Now, let $\mathbf{x}_i \in \mathcal{C}_i$ and denote its k th coordinate by $x_{i,k}$. A channel input at time k is generated by a scaled addition of the k th coordinate of the codewords

$$z_k = \sum_{i=1}^d \alpha_i x_{i,k}. \quad (2.17)$$

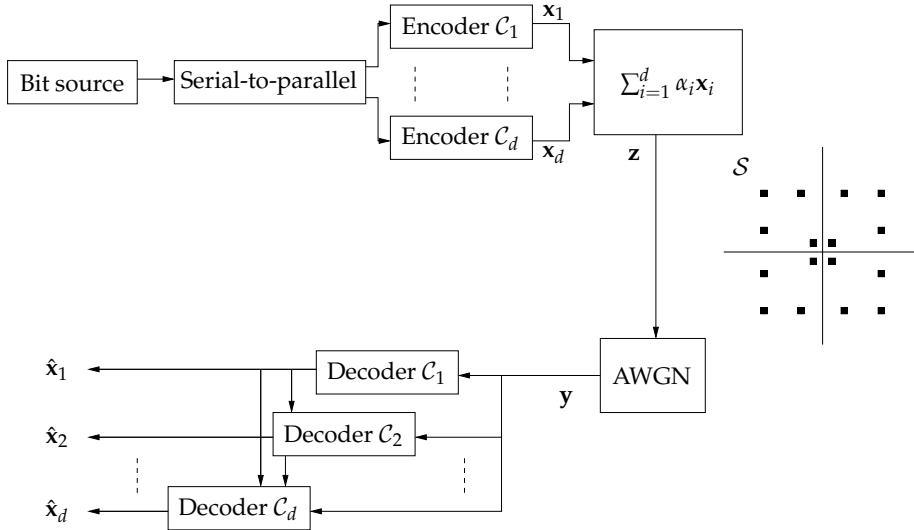


Figure 2.1: Block diagram of the modulation method with multilevel coding and multistage decoding.

Hence, the channel input word \mathbf{z} of length n is generated by a scaled component-wise addition of the codewords

$$\mathbf{z} = \sum_{i=1}^d \alpha_i \mathbf{x}_i. \quad (2.18)$$

At the receiver we employ a multistage decoding procedure which is inspired by (2.16). We decode each of the codes in a sequential order and without loss of generality we assume that the decoding sequence is C_1, C_2, \dots, C_d . C_1 is decoded first and the decision is passed on the next decoder which decodes C_2 . This procedure continues up to the last level where C_d is decoded². An overview of this system is shown in Figure 2.1.

We assume that codewords from C_1 to C_d are independently selected with equal probability and each code is such that the marginal distribution of the codeword bits is uniform. The latter will be the case if we use codes from a suitable ensemble of binary random codes or binary linear block codes. In this

²An alternative approach is to consider joint decoding of C_1 to C_d . However, we do not consider this approach in this chapter .

case the signal constellation is generated by the superposition of i.i.d. uniform random bit variables as in (2.14).

The use of multilevel coding with multistage decoding reduces the problem of achieving the left-handside of (2.16) to achieving each of the terms of the right-handside of (2.16) in a sequential fashion with binary codes. In [23] and [24] it is shown that multilevel coding with multistage decoding is optimal in the sense that $I((X_1, \dots, X_d); Y)$ can be achieved if the code rates are chosen properly.

2.2.3 Equivalent Binary Channels

When we use multilevel coding with binary codes and multistage decoding at the receiver, the coding problem for the original channel is transformed into a coding problem for a set of equivalent binary channels. Consider the case that we are decoding at level l . We assume that all previous levels are decoded correctly which implies that the values of X_1, \dots, X_{l-1} are known and we denote their realizations by x_1, \dots, x_{l-1} . The channel for X_l takes the form

$$Y = c'_l + \alpha_l X_l + \sum_{i=l+1}^d \alpha_i X_i + N, \quad (2.19)$$

where c'_l is given by

$$c'_l = \sum_{i=1}^{l-1} \alpha_i x_i. \quad (2.20)$$

Furthermore, X_{l+1}, \dots, X_d are unknown and considered to be noise. The additive noise for X_l is defined by

$$N'_l = \sum_{i=l+1}^d \alpha_i X_i + N, \quad (2.21)$$

and the density of N'_l is given by

$$f_{N'_l}(n) = \frac{1}{2^{d-l} \sqrt{2\pi\sigma^2}} \sum_{x_{l+1}} \dots \sum_{x_d} \exp\left(-\frac{(n - \alpha_{l+1}x_{l+1} - \dots - \alpha_d x_d)^2}{2\sigma^2}\right). \quad (2.22)$$

For future reference we note that this density has the following symmetry

$$f_{N'_l}(n) = f_{N'_l}(-n). \quad (2.23)$$

Now, we can write the equivalent channel for X_l as

$$Y = \alpha_l X_l + c'_l + N'_l. \quad (2.24)$$

This equivalent binary channel is defined by the channel transition density $f_{Y|X_l, \dots, X_1}$

$$f_{Y|X_l, \dots, X_1}(y|x_l, \dots, x_1) = f_{N'_l}(y - \alpha_l x_l - c'_l). \quad (2.25)$$

For the purpose of error-control coding we are interested in the achievable rate $I(Y; X_l | X_{l-1}, \dots, X_1)$ on this equivalent channel. $I(Y; X_l | X_{l-1}, \dots, X_1)$ is the average mutual information between Y and X_l given X_1, \dots, X_{l-1} . However, when decoding at level l the values of X_1, \dots, X_{l-1} are assumed to be known and the achievable rate is equal to

$$I(Y; X_l | X_{l-1} = x_{l-1}, \dots, X_1 = x_1).$$

A convenient consequence of the use of (2.14) is that this quantity is independent of the realization of X_1, \dots, X_{l-1} as the following theorem shows

Theorem 1 *Let channel inputs be generated by (2.14) where X_1, \dots, X_d are uniform i.i.d. random bit variables. $I(Y; X_l | X_{l-1} = x_{l-1}, \dots, X_1 = x_1)$ is independent of the realization of X_1, \dots, X_{l-1} . Hence*

$$I(Y; X_l | X_{l-1}, \dots, X_1) = I(Y; X_l | X_{l-1} = x_{l-1}, \dots, X_1 = x_1), \quad (2.26)$$

and the capacity C_l of the equivalent binary channel at level l is given by

$$C_l = \int_{-\infty}^{\infty} f_{N'_l}(y + \alpha_l) \log_2 \frac{2f_{N'_l}(y + \alpha_l)}{f_{N'_l}(y + \alpha_l) + f_{N'_l}(y - \alpha_l)} dy. \quad (2.27)$$

Proof 1 *First note that we can write*

$$I(Y; X_l | X_{l-1} = x_{l-1}, \dots, X_1 = x_1) = \sum_{x_l} \frac{1}{2} I(Y; X_l = x_l | X_{l-1} = x_{l-1}, \dots, X_1 = x_1), \quad (2.28)$$

where

$$\begin{aligned} & I(Y; X_l = x_l | X_{l-1} = x_{l-1}, \dots, X_1 = x_1) \\ &= \int_{-\infty}^{\infty} f_{Y|X_l, \dots, X_1}(y|x_l, \dots, x_1) \cdot \log_2 \frac{2f_{Y|X_l, \dots, X_1}(y|x_l, \dots, x_1)}{\sum_{x' \in \{-1, 1\}} f_{Y|X_l, \dots, X_1}(y|x', x_{l-1}, \dots, x_1)} dy \\ &= \int_{-\infty}^{\infty} f_{N'_l}(y - \alpha_l x_l - c'_l) \cdot \log_2 \frac{2f_{N'_l}(y - \alpha_l x_l - c'_l)}{f_{N'_l}(y - \alpha_l - c'_l) + f_{N'_l}(y + \alpha_l - c'_l)} dy. \end{aligned} \quad (2.29)$$

The right-handside of this equation does not depend on c'_1 since we integrate over \mathbb{R} . From this we conclude that the left-handside of (2.28) does not depend on the realization of X_1, \dots, X_{l-1} . Moreover, we can make use of the symmetry of $f_{N'_1}$ to show that the value of (2.28) does not depend on x_l . Equation (2.27) follows when we take $x_l = -1$ and $c'_1 = 0$.

By the chain rule of mutual information the constrained constellation capacity $I(Y; Z)$ is given by

$$I(Y; Z) = I(Y; (X_1, \dots, X_d)) = \sum_{i=1}^d C_i. \quad (2.30)$$

As mentioned before, multilevel coding with multistage decoding allows us to achieve $I((X_1, \dots, X_d); Y)$. Now it is clear that we require that the code rates satisfy $r_i \leq C_i$.

The use of superposition coding with multilevel encoding at the transmitter and multistage decoding at the receiver allows one to treat modulation and coding separately. First, a signal constellation can be designed for which the constrained constellation capacity is close to the capacity of the AWGN channel. Second, binary error-correcting codes can be designed for the set of equivalent binary channels defined by the constellation. We continue along this path in this chapter. First, we describe several families of signal constellations in Section 2.3 and show that for the AWGN channel constellations can be designed which have a constrained capacity close to the capacity of the AWGN channel. Second, we consider the design of binary LDPC codes for the equivalent binary channels in Section 2.4.

2.3 Signal Constellations

In this section we consider the properties of signal constellations generated by the superposition of uniform i.i.d. random bit variables and identify several families of constellations. We consider conventional pulse-amplitude modulation (PAM) signal constellations, binomial signal constellations and numerically optimized signal constellations. Furthermore, we compare the performance of different signal constellations.

2.3.1 Signal Constellation Properties

We use a signal constellation \mathcal{S} to communicate over the AWGN channel

$$Y = Z + N, \quad (2.31)$$

where Z takes a value $z \in \mathcal{S}$ with probability $P_{\mathcal{S}}(z)$. There are several performance measures on which signal constellations can be compared. These include uncoded symbol error rate, Euclidian distance profile and peak-to-average power ratio. We are interested in achieving capacity on the AWGN channel and we will only be concerned with the information theoretical limits. Thus we compare signal constellations on their constrained constellation capacity. For this purpose recall that the capacity of the AWGN channel is given by

$$C = \frac{1}{2} \log_2(1 + \text{SNR}). \quad (2.32)$$

Let R denote the constrained constellation capacity which is achieved at some SNR. Next, denote the SNR at which the capacity of the AWGN channel is equal to R by SNR_{AWGN}

$$\text{SNR}_{\text{AWGN}} = 2^{2R} - 1. \quad (2.33)$$

This motivates the definition of the normalized SNR [14] as

$$\text{SNR}_{\text{norm}} = \frac{\text{SNR}}{\text{SNR}_{\text{AWGN}}} = \frac{\text{SNR}}{2^{2R} - 1}. \quad (2.34)$$

The value of SNR_{norm} for which a constrained constellation capacity R is achieved signifies how far the constellation is operating from the capacity of the AWGN channel. The baseline performance is $\text{SNR}_{\text{norm}} = 0$ [dB], which is the required SNR_{norm} for a Gaussian channel input to achieve any rate on the AWGN channel. We use this benchmark to compare different signal constellations.

2.3.2 Properties of Constellations generated by Superposition

Recall from Section 2.2 that a channel input Z is generated by a scaled addition of uniform i.i.d. random bit variables

$$Z = \sum_{i=1}^d \alpha_i X_i. \quad (2.35)$$

The average energy expended per channel use E_s can be expressed as

$$E_s = E[Z^2] = E \left[\left(\sum_{i=1}^d \alpha_i X_i \right)^2 \right] = \sum_{i=1}^d \alpha_i^2. \quad (2.36)$$

The signal constellation \mathcal{S} is given by (2.11) and the probability with which the constellation symbols are selected by (2.15). To compute the elements of the signal constellation and the probability with which the constellation symbols are generated in an efficient way, we consider the generating function of Z . For this, note that the generating function of $\alpha_i X_i$ is given by

$$g_i(x) = \frac{1}{2}x^{\alpha_i} + \frac{1}{2}x^{-\alpha_i}, \quad (2.37)$$

which allows us to express the generating function of Z as

$$g_Z(x) = \prod_{i=1}^d \left(\frac{1}{2}x^{\alpha_i} + \frac{1}{2}x^{-\alpha_i} \right). \quad (2.38)$$

The righthand side of this equation can be expanded as

$$g_Z(x) = \sum_{i=1}^{2^d} p_i x^{t_i}. \quad (2.39)$$

Now, the signal constellation \mathcal{S} is given by

$$\mathcal{S} = \left\{ t_i \mid i = 1, \dots, 2^d \right\}. \quad (2.40)$$

The probability assignment on the constellation symbols can be obtained by collecting terms in (2.39). $P_S(z)$ is equal to the coefficient of the term of power z .

2.3.3 Families of Signal Constellations

PAM Signal Constellations

A signal constellation with a uniform spacing and a uniform distribution on the constellation symbols is generated by taking the α_i as consecutive powers

of two. A constellation symbol Z from an M -PAM constellation with $M = 2^d$ constellation symbols is generated by

$$Z = \sum_{i=1}^d 2^{i-1} X_i. \quad (2.41)$$

The signal constellation is given by

$$\mathcal{S} = \left\{ -2^d + 2i - 1 \mid i = 1, 2, 3, \dots, 2^d \right\}, \quad (2.42)$$

and the probability distribution is uniform

$$P_{\mathcal{S}}(z) = \frac{1}{2^d} \text{ for } z \in \mathcal{S}. \quad (2.43)$$

The average energy expended per channel use for this constellation is

$$E_s = \frac{2^{2d} - 1}{3}. \quad (2.44)$$

The constrained constellation capacity of the M -PAM constellations is plotted in Figure 2.2 for $d = 2$ to $d = 8$. For low rates there is only a small loss with respect to the capacity of the AWGN channel. However, for higher rates the loss is substantial. At a rate of 3 bit/use a shaping gain of over 1 dB is available. Note that the capacity curves all converge to a limit since the constellations have a finite number of constellation symbols.

Binomial Signal Constellations

Signal constellations with a uniform spacing and a binomial distribution are generated by

$$Z = \sum_{i=1}^d X_i. \quad (2.45)$$

The signal constellation is given by

$$\mathcal{S} = \{-d + 2(i - 1) \mid i = 1, 2, 3, \dots, d + 1\}. \quad (2.46)$$

The map from bits to constellation symbols is not injective and the distribution of the constellation symbols is binomial

$$P_{\mathcal{S}}(z) = \binom{d}{\frac{1}{2}(z + d)} 2^{-d} \text{ for } z \in \mathcal{S}. \quad (2.47)$$

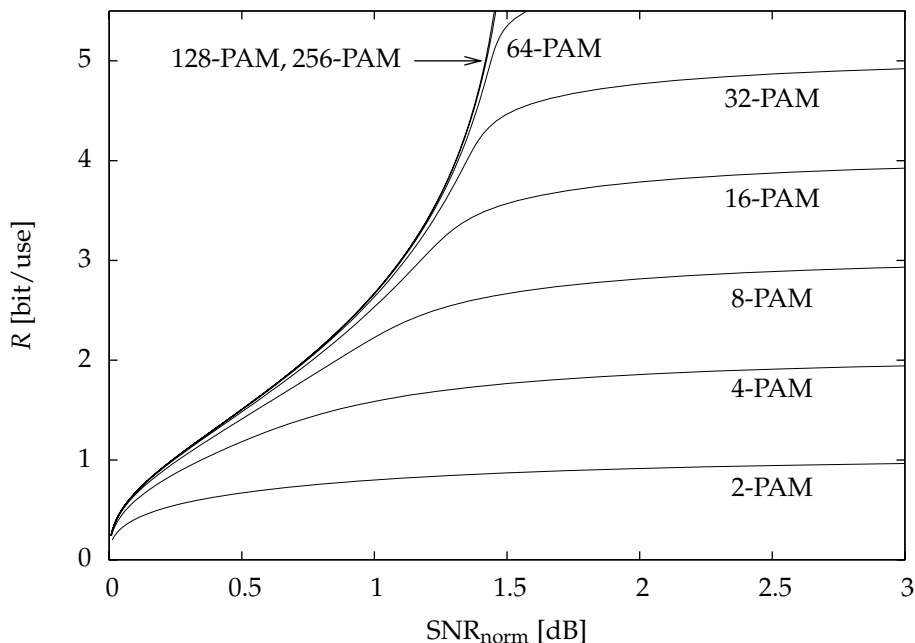


Figure 2.2: The constrained capacity limits of the M -PAM constellations.

The size of the signal constellation is $|\mathcal{S}| = d + 1$ and the average energy per channel use is given by

$$E_s = d. \quad (2.48)$$

The constrained capacity curves of the binomial signal constellations for $d = 2$ to $d = 10$ are shown in Figure 2.3. The figure also shows the 16-PAM constrained capacity limit. The binomial signal constellations have their constrained capacity limit very close to the AWGN limit. At least where the constrained capacity is not too close to the finite constellation entropy. For the signal constellation with $d = 10$ a rate of 2 bit/use is achieved at $\text{SNR}_{\text{norm}} = 0.027$ dB. This constellation has 11 constellations symbols and compared to the 16-PAM constellation, we achieve a shaping gain of 0.74 dB.

A drawback of the binomial signal constellations is that they are only useful for low to moderate rates. The reason for this is that the supported rate grows

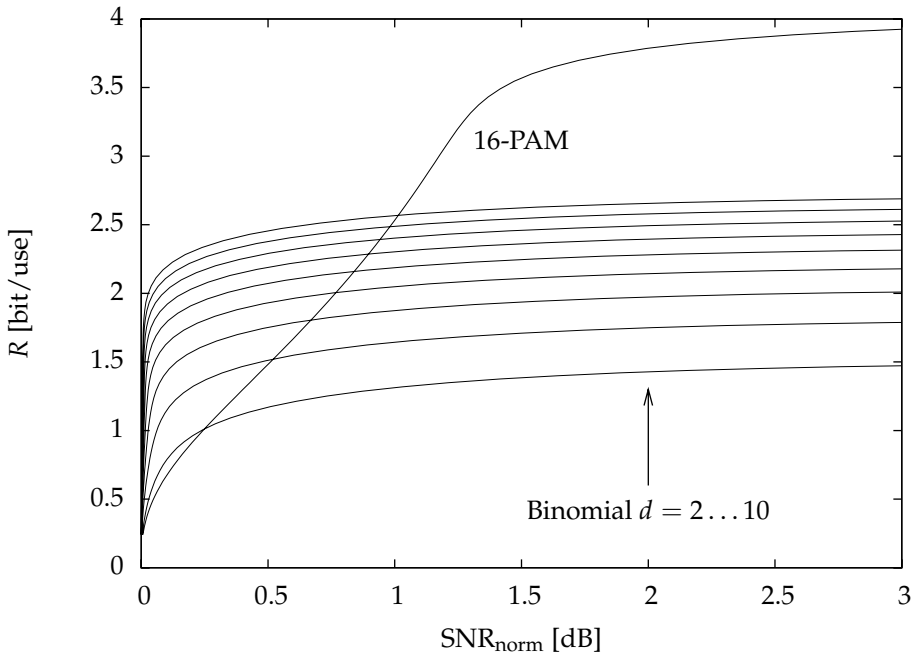


Figure 2.3: The constrained capacity limits of the binomial constellations.

only logarithmically with d and a high number of levels is required for multi-level coding with multistage decoding. To see this note that for the binomial signal constellations, the size of the signal constellation is equal to $d + 1$. Hence the entropy of the signal constellation is upper bounded by $\log_2(d + 1)$. Thus to transmit at a rate of R bit/use, we should at least have $d \geq 2^R - 1$.

Finally, note that these signal constellations are useless for uncoded transmission, because the map from bits to constellation symbols is not injective. Regardless of the SNR, the bit-error rate will always be lower bounded by a fixed constant. However, when we combine modulation with error-control coding the binomial distribution provides a shaping gain which saves transmission power.

Numerically Optimized Signal Constellations

A major advantage of the use of superposition to generate a signal constellation is that relatively few degrees of freedom determine the constellation geometry and mapping from bits to constellation symbols. This makes a numerical optimization feasible. The objective is to find a set of α_i defining a signal constellation with a constrained constellation capacity close to the capacity of the AWGN channel. We can formulate this optimization problem as follows

$$\begin{aligned} \max I(Y; Z), \text{ where } Z = \sum_{i=1}^d \alpha_i X_i \\ \text{subject to } \sum_{i=1}^d \alpha_i^2 = E_s, \end{aligned} \quad (2.49)$$

where we have only incorporated a power constraint, but other constraints, such a maximum peak-to-average power ratio, can be included as well.

To illustrate the potential of numerical optimization we design several signal constellations for target rates in the range from 2 bit/use to 5 bit/use for several values of d . Note that to transmit at a rate of R bit/use, we require at least $d \geq R$. The optimization is carried out as follows. First, we determine the SNR for which the capacity of the AWGN channel is equal to the target rate. Second, the power constraint is set accordingly and (2.49) is solved.

For the actual optimization, we have experimented with several optimization strategies. One strategy giving good results in acceptable optimization time is the use of differential evolution [25] and we limit ourselves to the results obtained by this optimizer. Differential evolution is a global optimization strategy based on hill-climbing and a genetic algorithm and is sometimes used in the design of error-correcting codes [26], [27].

The optimization results are shown in Table 2.1. The table gives for each rate R the SNR for which the capacity of the AWGN channel is equal to R bit/use. Furthermore, for several values of d the optimized α_i are given with the resulting constrained constellation capacity $I(Y; Z)$. Note that $I(Y; Z)$ is independent of the order in which the α_i are given in the table. However, the capacities of the binary equivalent channels depend on the order in which the levels are decoded. Changing the order of the α_i changes the capacities of the equivalent binary channels. In Table 2.1 the α_i are given in ascending order. Finally, the table gives the value of SNR_{norm} where a rate of R bit/use is achieved. This value signifies the gap to capacity of the signal constellation.

R SNR	2.0 bit/use 11.76				3.0 bit/use 17.99			
	d	3	4	5	6	4	5	6
α_1	0.3488	0.2968	0.2632	0.2373	0.2090	0.2118	0.1978	0.1918
α_2	0.5671	0.4165	0.3353	0.2901	0.3738	0.3316	0.2767	0.2418
α_3	0.7461	0.5597	0.4827	0.4636	0.4928	0.4280	0.3752	0.3386
α_4		0.6520	0.5410	0.4636	0.7575	0.5052	0.4212	0.3697
α_5			0.5410	0.4636		0.6378	0.5003	0.4179
α_6				0.4636			0.5620	0.4706
α_7								0.5074
$ \mathcal{S} $	8	16	24	20	16	32	64	128
$I(Y;Z)$	1.930	1.972	1.987	1.993	2.905	2.958	2.980	2.990
SNR_{norm}	0.50	0.19	0.09	0.05	0.64	0.27	0.13	0.06

R SNR	4.0 bit/use 24.07				5.0 bit/use 30.10			
	d	5	6	7	8	6	7	8
α_1	0.1571	0.1313	0.1175	0.1160	0.0910	0.0700	0.0634	0.0581
α_2	0.2688	0.2345	0.2123	0.1995	0.1716	0.1356	0.1498	0.1099
α_3	0.3473	0.3046	0.2737	0.2490	0.2223	0.2558	0.2591	0.2075
α_4	0.5675	0.4624	0.3999	0.3080	0.3513	0.2923	0.2989	0.2380
α_5	0.6785	0.5146	0.4424	0.3801	0.5491	0.4761	0.3475	0.3423
α_6		0.5970	0.4726	0.4189	0.6986	0.5208	0.3986	0.3757
α_7			0.5360	0.4639		0.5727	0.5107	0.4187
α_8				0.5046			0.5260	0.4609
α_9								0.4887
$ \mathcal{S} $	32	64	128	256	64	128	256	512
$I(Y;Z)$	3.895	3.956	3.978	3.989	4.887	4.950	4.970	4.988
SNR_{norm}	0.73	0.28	0.13	0.07	0.77	0.32	0.18	0.07

Table 2.1: Parameters of the designed signal constellations.

We observe that for the target rates given in the table, the designed signal constellations achieve a considerable shaping gain. All constellations given in the table outperform conventional PAM constellations. At the lowest R in the table, a 256-PAM constellation requires an SNR_{norm} of 0.74 dB to achieve a rate of 2 bit/use. The constellation for $R = 2$ with $d = 3$ achieves a rate of 2 bit/use at an SNR_{norm} of 0.50 dB. However, this constellation has only 8 constellation symbols instead of 256. For higher rates and higher values of d the achievable shaping gain is more profound.

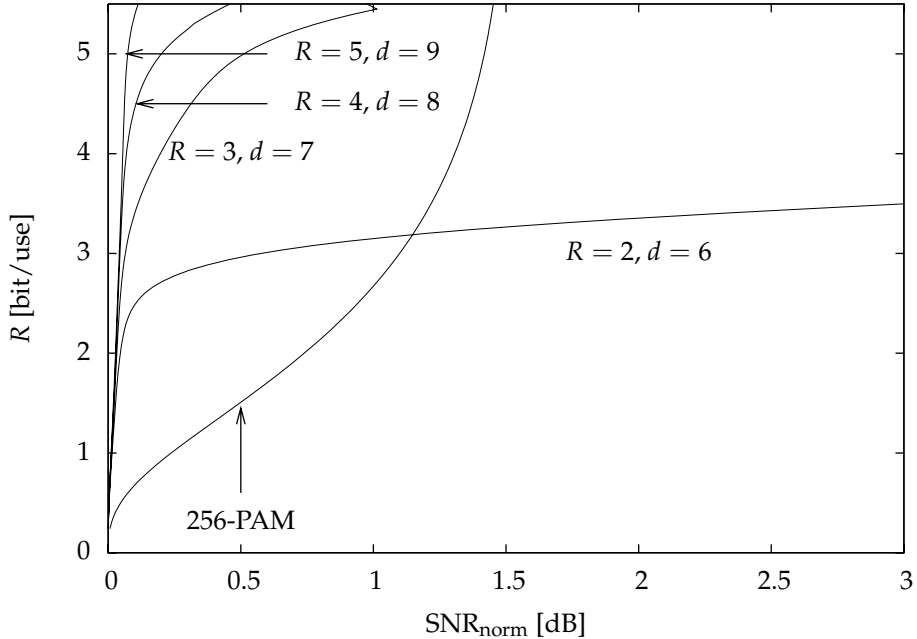


Figure 2.4: The constrained capacity limits of the numerically optimized constellations.

A plot of SNR_{norm} versus the rate of the signal constellations is given in Figure 2.4. The plot shows for each target rate the constrained capacity curve for the signal constellation with the highest value of d . For each of the target rates we have designed a signal constellation which achieves the target rate within 0.1 dB of the capacity of the AWGN channel. By increasing the value of d one can even get closer to the capacity of the AWGN channel.

Two of the constellations defined by Table 2.1 we discuss in greater detail. The parameters of these constellations are printed in bold in the table and these constellations serve as an example in the next section when we consider error-control coding. We refer to the constellation for 2 bit/use and 5 bit/use as constellation A and constellation B, respectively. Constellation A has 20 constellation symbols and a non-uniform spacing of the constellation symbols. Moreover, the distribution of the constellation symbols is non-uniform. It is in-

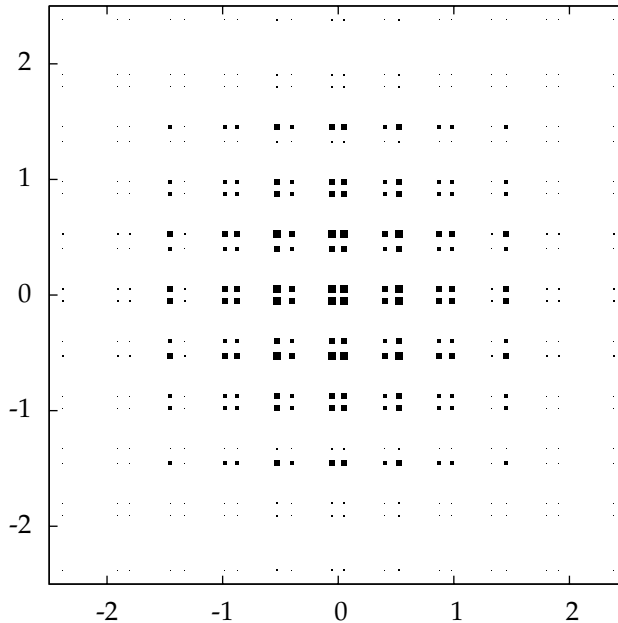


Figure 2.5: Signal constellation A.

interesting to see that the last four coefficients converge to the same value. This implies that X_3 to X_6 generate a binomial distribution. To give an impression of the geometry of the constellation, the resulting quadrature constellation is shown in Figure 2.5. This quadrature constellation is generated by using each dimension independently. The size of each square is proportional to the probability with which the constellation symbols are selected. The figure clearly shows the non-uniform spacing and non-uniform distribution of the constellation symbols. Figure 2.6 shows the constrained capacity limit of the constellation. The constrained capacity curve is close to the AWGN capacity curve for a wide range of SNRs. At $\text{SNR} = 11.76$ dB the constrained capacity is 1.993 which is very close to the capacity of the AWGN channel. In terms of dB the distance to the capacity of the AWGN channel is only 0.05 dB. Furthermore, a 32-PAM constellation requires $\text{SNR} = 12.51$ dB to achieve a constrained capacity of 2 bit/use while constellation A requires $\text{SNR} = 11.81$ dB to achieve the

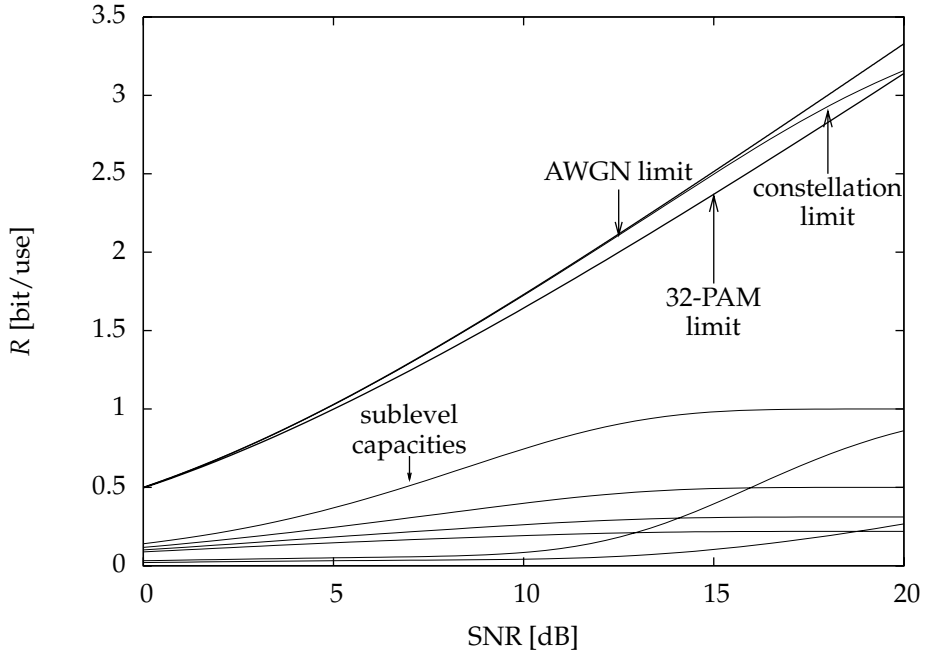


Figure 2.6: The capacity limit of constellation A.

same rate. Compared to a 32-PAM constellation, we achieve a shaping gain of 0.7 dB. The figure also shows the capacities of the equivalent binary channels whose sum is equal to the total capacity.

Constellation B has 256 constellation symbols and the spacing of the symbols is non-uniform. Figure 2.7 shows the quadrature constellation and unlike constellation A, the mapping from bits to constellation symbols is one-to-one which results in a uniform distribution over the constellation symbols. Figure 2.8 shows the constrained capacity of the signal constellation together with the constrained capacity of a 256-PAM signal constellation. We observe that at $\text{SNR} = 30.10$ dB the constrained capacity of the constellation is 4.97 bit/use. In terms of SNR the distance to the capacity of the AWGN channel is 0.18 dB. Compared to 256-PAM constellation we achieve a shaping gain of 1.22 dB.

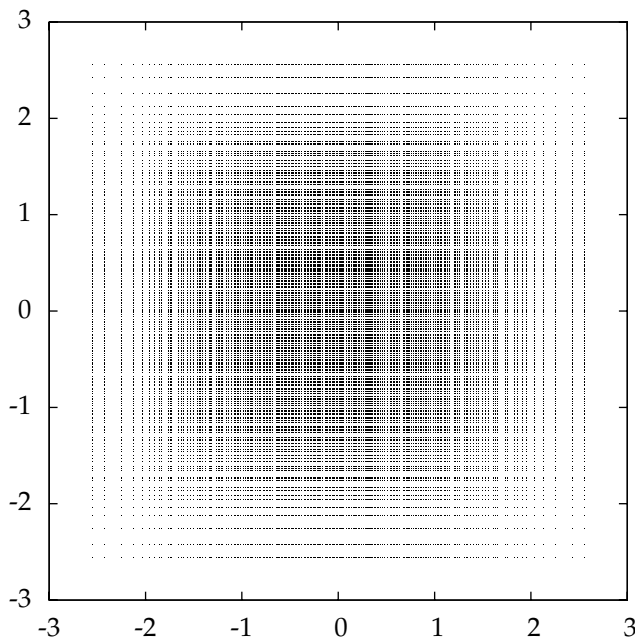


Figure 2.7: Signal constellation B.

2.4 Error-control Coding with Binary LDPC Codes

In this section we consider the use of binary error-correcting codes on the set of equivalent binary channels defined by the signal constellations. Constellation A and constellation B defined in the previous section will serve as a running example in this section and the next section. From the chain rule of mutual information it follows that the constrained constellation capacity can be achieved if we achieve capacity on each of the equivalent binary channels. When we generate channel inputs by (2.14) each of the equivalent binary channels is defined by (2.24). In the previous sections the capacity of this equivalent binary channel is denoted by C_l and achieved for a uniform distribution on X_l . Ensembles of binary linear block codes have a uniform distribution on the codeword bits and if the rate of the code satisfies $r_l \leq C_l$, they are capable of achieving C_l under

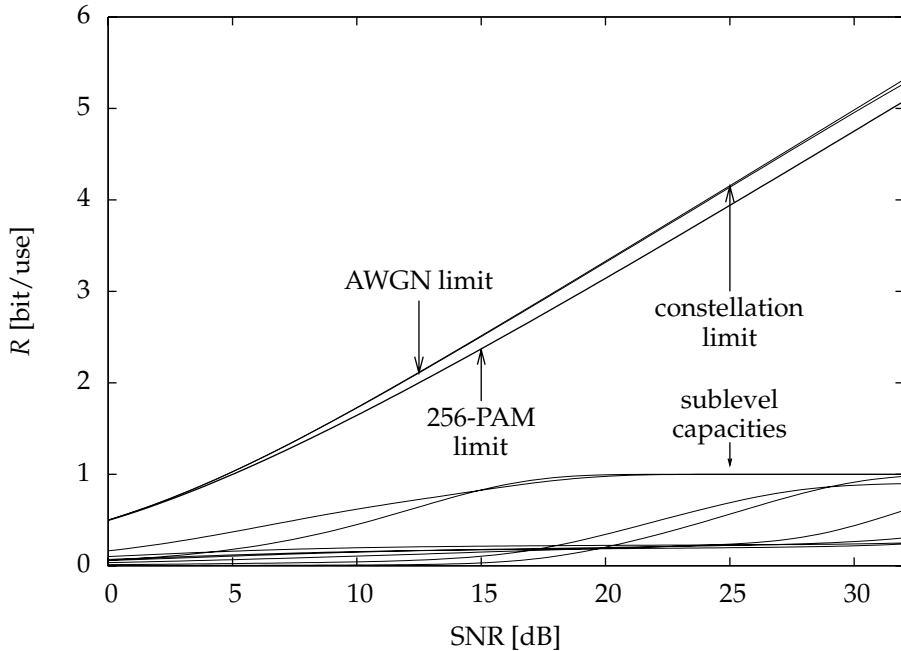


Figure 2.8: The capacity limit of constellation B.

maximum likelihood decoding [28]. However, maximum likelihood decoding is not feasible from a practical point of view.

Binary sparse-graph codes such as turbo codes [6] and LDPC codes [1] admit low-complexity decoding algorithms. In [26] it is shown that for several memoryless binary-input output-symmetric channels, LDPC codes can be designed which perform very close to channel capacity. We show that this also holds for the equivalent binary channels defined by the signal constellations.

In this section we start with the derivation of some additional properties of the equivalent binary channels which are relevant for the analysis and design of LDPC codes. We show that the equivalent binary channels are in fact output-symmetric channels. Furthermore, LDPC codes are usually decoded by message-passing algorithms where the messages represent log-likelihood ratios (LLRs). From a practical point of view the computation of LLRs is important and we show how to accomplish this in an efficient manner for signal

constellations generated by superposition. Finally, we discuss the design of LDPC codes for the equivalent binary channels.

2.4.1 Equivalent Binary Channels

Recall from Section 2.2 that with superposition coding and multistage decoding at the receiver, the equivalent binary channel at level l is given by

$$Y = \alpha_l X_l + c'_l + N'_l, \quad (2.50)$$

where c'_l is defined as

$$c'_l = \sum_{i=1}^{l-1} \alpha_i X_i, \quad (2.51)$$

and N'_l as

$$N'_l = \sum_{i=l+1}^d \alpha_i X_i + N. \quad (2.52)$$

Furthermore, the density of N'_l is given by

$$f_{N'_l}(n) = \frac{1}{2^{d-l} \sqrt{2\pi\sigma^2}} \sum_{x_{l+1}} \dots \sum_{x_d} \exp\left(-\frac{(n - \alpha_{l+1}x_{l+1} - \dots - \alpha_d x_d)^2}{2\sigma^2}\right). \quad (2.53)$$

A sufficient statistic to make a decision on X_l is the log-likelihood ratio. Let y denote a realization of Y . The LLR for X_l is defined as

$$L_l(y) = \log \frac{f_{Y|X_l, \dots, X_1}(y|1, x_{l-1}, \dots, x_1)}{f_{Y|X_l, \dots, X_1}(y|-1, x_{l-1}, \dots, x_1)} = \log \frac{f_{N'_l}(y - \alpha_l - c'_l)}{f_{N'_l}(y + \alpha_l - c'_l)}. \quad (2.54)$$

We can view $L_l(y)$ as a random variable by noting that it is a function of the channel output Y which is a function of the random variables X_1, \dots, X_l and N'_l . As a random variable we denote $L_l(y)$ by $L_l(Y)$

Lemma 2 $L_l(Y)$ is independent of the realization of X_1, \dots, X_{l-1} .

Proof 2 First, note that the realization of X_1, \dots, X_{l-1} is summarized in the value of c'_l . We can write $L_l(Y)$ as

$$\begin{aligned} L_l(Y) &= \log \frac{f_{Y|X_l, \dots, X_1}(Y|1, x_{l-1}, \dots, x_1)}{f_{Y|X_l, \dots, X_1}(Y|-1, x_{l-1}, \dots, x_1)} \\ &= \log \frac{f_{N'_l}(Y - \alpha_l - c'_l)}{f_{N'_l}(Y + \alpha_l - c'_l)} = \log \frac{f_{N'_l}(\alpha_l X_l - \alpha_l + N'_l)}{f_{N'_l}(\alpha_l X_l + \alpha_l + N'_l)}, \end{aligned} \quad (2.55)$$

which is only a function of X_l and N_l' .

In the analysis and design of binary LDPC codes the density of $L_l(Y)$ conditioned on the transmission of a 1 ($X_l = 1$) plays a crucial role. We assume that this density exists and refer to such a density as an ℓ -density. An ℓ -density $a(y)$ is said to be symmetric if it satisfies [29]

$$a(y) = e^y a(-y). \quad (2.56)$$

For a channel with a symmetric ℓ -density the analysis and design of LDPC codes is greatly simplified. The analysis of a message passing decoder satisfying some symmetry properties can be restricted to the all-ones codeword. In case the channel is a BIOS channel, i.e.

$$f_{Y|X}(y|1) = f_{Y|X}(-y|-1), \quad (2.57)$$

the corresponding ℓ -density is easily shown to be symmetric [29]. However, the channel of (2.50) does not satisfy (2.57). Nevertheless, the ℓ -density of the channel defined by (2.50) is symmetric as the following theorem shows.

Theorem 3 *The ℓ -density of the binary channel defined by (2.50) is symmetric.*

Proof 3 *First, define*

$$Y' = Y - c_l' = \alpha_l X_l + N_l', \quad (2.58)$$

which effectively cancels the contribution of c_l' . The LLR of X_l for this channel is defined as

$$\begin{aligned} L_l'(Y') &= \log \frac{f_{Y'|X_l, \dots, X_1}(Y'|1, x_{l-1}, \dots, x_1)}{f_{Y'|X_l, \dots, X_1}(Y'|-1, x_{l-1}, \dots, x_1)} = \log \frac{f_{N_l'}(Y' - \alpha_l)}{f_{N_l'}(Y' + \alpha_l)} \\ &= \log \frac{f_{N_l'}(\alpha_l X_l - \alpha_l + N_l')}{f_{N_l'}(\alpha_l X_l + \alpha_l + N_l')} = L_l(Y), \end{aligned} \quad (2.59)$$

which shows that $L_l(Y)$ and $L_l'(Y)$ are equal and will have the same ℓ -density. Next note that the channel defined by (2.58) has a channel transition probability density function which satisfies

$$f_{Y|X}(y|1) = f_{Y|X}(-y|-1). \quad (2.60)$$

The ℓ -density corresponding to this channel is symmetric from which we conclude that the ℓ -density of the binary channel defined by (2.50) is symmetric.

Several parameters of binary channels with a symmetric ℓ -density are easily expressed in terms of this ℓ -density. For an overview we refer to [29]. The capacity of the equivalent binary channel at level l in terms of its ℓ -density $a_l(y)$ is given by

$$C_l = 1 - \int_{-\infty}^{\infty} a_l(y) \log_2(1 + e^{-y}) dy. \quad (2.61)$$

2.4.2 Computation of Log-likelihood Ratios

From a practical point of view an important issue is the actual computation of LLRs. To derive a method to compute the LLRs for all levels efficiently, we define a random variable Z_l

$$Z_l = \sum_{i=1}^l \alpha_i X_i, \quad (2.62)$$

and we define Z_0 as a constant random variable equal to 0 with probability 1. Hence for $l \geq 1$ we can write

$$Z_l = Z_{l-1} + \alpha_l X_l. \quad (2.63)$$

The sequence of random variables Z_0, Z_1, \dots, Z_d forms a Markov chain where the state space can be identified with the signal constellation \mathcal{S} . However, the support of Z_l is \mathcal{S}_l

$$\mathcal{S}_l = \left\{ \sum_{i=1}^l \alpha_i x_i \mid x_1 \in \{-1, 1\}, \dots, x_l \in \{-1, 1\} \right\} \quad l \geq 1, \quad (2.64)$$

and by definition $\mathcal{S}_0 = \{0\}$. The possible transitions in state space are conveniently depicted by a trellis. Figure 2.9 shows the trellis for constellation A. The trellis consists of $d + 1$ rows of nodes where we start counting rows from 0. The i th row consists of nodes corresponding to the elements of \mathcal{S}_i . Hence the root node corresponds to \mathcal{S}_0 and the leave nodes to \mathcal{S}_d . Each node at a particular row i can be identified by an element of \mathcal{S}_i and in Figure 2.9 we have labeled the nodes accordingly. We refer to a node corresponding to $z \in \mathcal{S}_i$ as node z at row i . The edges between the nodes depict the possible state transitions. A node z_i at row i is connected to a node z_{i+1} at row $i + 1$ if and only if $z_{i+1} = z_i \pm \alpha_{i+1}$.

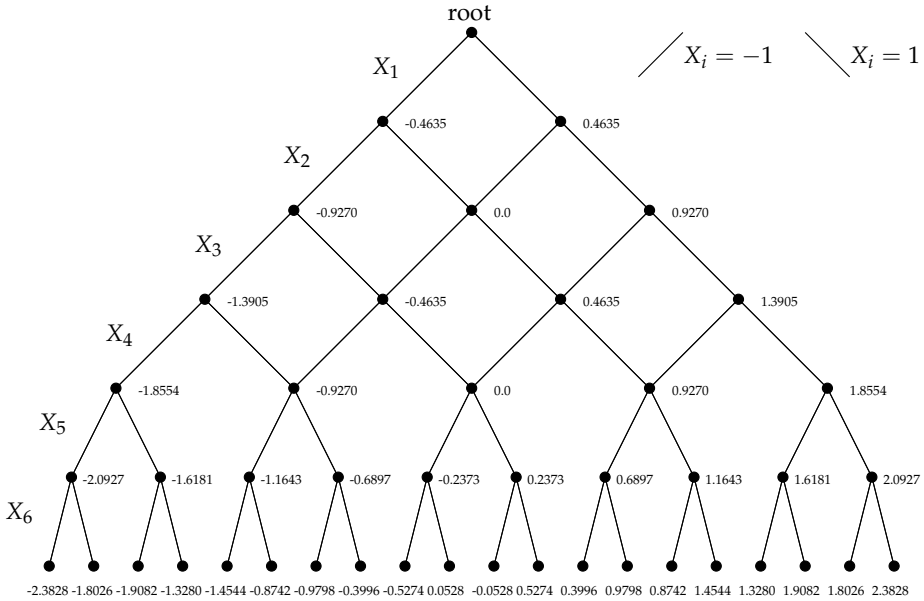


Figure 2.9: The trellis of constellation A.

We can use the trellis to compute the LLRs for each of the levels in multi-stage decoding. For this purpose we carry out a backward pass of messages on the trellis. Let $\beta_z^{(d)}$ denote the initial message at the leave node corresponding to constellation symbol $z \in \mathcal{S}_d$. We initialize $\beta_z^{(d)}$ as

$$\beta_z^{(d)} = f_N(y - z), \quad (2.65)$$

where y denotes the channel output and f_N the Gaussian noise density with variance σ^2 . At each node at row d of the trellis the corresponding $\beta_z^{(d)}$ is sent to its parent node at row $d - 1$. For a node z at row i we compute a message $\beta_z^{(i)}$ as

$$\beta_z^{(i)} = \beta_{z+\alpha_{i+1}}^{(i+1)} + \beta_{z-\alpha_{i+1}'}^{(i+1)}, \quad (2.66)$$

where $\beta_{z+\alpha_{i+1}}^{(i+1)}$ and $\beta_{z-\alpha_{i+1}'}^{(i+1)}$ are the messages sent by the descendants of node z . In multistage decoding we assume decoding proceeds from X_1 to X_d and we

can compute the LLR for X_1 as follows

$$L_1 = \log \frac{\sum_{x_2} \cdots \sum_{x_d} f_N(y - \alpha_1 - \sum_{i=2}^d \alpha_i x_i)}{\sum_{x_2} \cdots \sum_{x_d} f_N(y + \alpha_1 - \sum_{i=2}^d \alpha_i x_i)} = \log \frac{\beta_{+\alpha_1}^{(1)}}{\beta_{-\alpha_1}^{(1)}}. \quad (2.67)$$

Once a decision on X_1 has been made the LLR for X_2 can be computed. In case $X_1 = 1$ the LLR for X_2 is computed as

$$L_2 = \log \frac{\beta_{+\alpha_1+\alpha_2}^{(2)}}{\beta_{+\alpha_1-\alpha_2}^{(2)}}, \quad (2.68)$$

and in case $X_1 = -1$ the LLR for X_2 is computed as

$$L_2 = \log \frac{\beta_{-\alpha_1+\alpha_2}^{(2)}}{\beta_{-\alpha_1-\alpha_2}^{(2)}}. \quad (2.69)$$

In general at level l we compute a LLR for X_l as

$$L_l = \log \frac{\beta_{\alpha_l+c'_l}^{(l)}}{\beta_{-\alpha_l+c'_l}^{(l)}}. \quad (2.70)$$

We only need to compute the values of the $\beta_z^{(l)}$ messages at the beginning of the multistage decoding process. When decoding a level we compute the LLRs for the bits at that level by taking the logarithm of the two $\beta^{(l)}$ messages depending on the decision at the previous levels. The complexity of computing LLRs depends on the actual trellis and hence on the α_i defining the signal constellation. One can show that for a constellation of size $|\mathcal{S}|$, one requires at least $|\mathcal{S}|$ evaluations of f_N , $|\mathcal{S}| - 2$ additions, $\log_2 |\mathcal{S}|$ divisions and $\log_2 |\mathcal{S}|$ evaluations of the natural logarithm. For constellations where (2.14) is one-to-one this bound is tight.

Example 1 Consider the use of constellation A where transmission takes place at a rate of 2 bit/use. To compute LLRs for X_1 to X_6 during multistage decoding, we require at least 20 evaluations of f_N , 24 additions, 6 divisions and 6 logarithms. In total we require 56 operations and per actual data bit we require 28 operations. Compared to the decoding of e.g. LDPC codes this is negligible.

2.4.3 LDPC Codes

An LDPC code is a linear block code which is defined by a low-density parity-check matrix $\mathbf{H} \in \text{GF}(2)^{(n-k) \times n}$, where n is the length of a codeword and k the number of source bits which is mapped to a codeword. Each codeword \mathbf{x} satisfies $\mathbf{H}\mathbf{x} = \mathbf{0}$. An LDPC code is conveniently represented by a Tanner graph [30]. The Tanner graph is a bipartite graph consisting of variable nodes and check nodes. Each variable nodes represents a bit of a codeword and each check node a row of the parity-check matrix. An ensemble of LDPC codes can be defined by a degree distribution pair (λ, ρ) where $\lambda(x)$ and $\rho(x)$ define the distribution of variable nodes and check nodes, respectively

$$\lambda(x) = \sum_i \lambda_i x^{i-1} \quad \rho(x) = \sum_i \rho_i x^{i-1}, \quad (2.71)$$

where λ_i denotes the fraction of edges connected to variable nodes of degree i and ρ_i denotes the fraction of edges connected to check nodes of degree i . In terms of the degree distribution pair (λ, ρ) the design rate r of the code is given by

$$r = 1 - \frac{\int_0^1 \rho(x) dx}{\int_0^1 \lambda(x) dx}. \quad (2.72)$$

The analysis and design of LDPC codes is greatly simplified for BIOS channels. In case the channel is a BIOS channel and the decoder satisfies certain symmetry conditions, the performance of the decoder is independent of the transmitted codeword. This implies that we can restrict the analysis of the decoding process to the case where the all-ones codeword is transmitted. In the previous section we have shown that the equivalent binary channels resulting from superposition coding with multistage decoding are BIOS channels.

2.4.4 Analysis and Design of LDPC Codes

The average performance of an ensemble of LDPC codes used on a channel depends on the degree distribution pair (λ, ρ) . Given a channel the goal of LDPC design is to find a degree distribution pair which defines an LDPC ensemble of rate r with an iterative decoding threshold close to the capacity of the channel. Given a degree distribution pair the asymptotic performance of the corresponding ensemble can be evaluated by density evolution [26]. To design good degree distributions density evolution can be combined with a global optimizer.

Another approach to the design of LDPC codes is the use of EXIT charts analysis combined with curve fitting techniques. EXIT chart analysis is basically a one-dimensional approximation to density evolution based on a Gaussian approximation. The design problem can be cast into a curve fitting or linear programming problem which is easy to solve. The actual performance of the degree distribution pair designed with the approximation depends on the channel. In many cases the approximation is good enough in the sense that one can find degree distributions with a threshold close to the capacity of the channel. The use of superposition coding with multistage decoding gives rise to a whole family of BIOS channels. For many of these channels to use of EXIT chart design alone does not give satisfying results.

As pointed out in [31] and [29], density evolution and EXIT chart design can be combined as well. Initially EXIT chart design is used to find an initial degree distribution pair. Next, the performance of this pair is evaluated by density evolution. The results of density evolution are used to improve the EXIT chart approximation and a new degree distribution pair is obtained with hopefully a better performance. This approach allows to design good degree distributions for the channels resulting from superposition coding with multistage decoding as we will see later. First, we describe the design procedure in more detail.

Asymptotic Performance Evaluation with Density Evolution

In order to analyze the performance of an ensemble of LDPC codes on one of the equivalent binary channels, we consider the transmission of the all-ones codeword on this channel. In this case a negative realization of the LLR signifies a bit error. Let a_0 denote the ℓ -density of the LLRs corresponding to the channel output. Density evolution is a recursive process where the evolution of the densities of LLRs during the decoding process is tracked on a cycle-free graph corresponding to a degree distribution pair (λ, ρ) . Let f_m denote the density of messages sent by the variable to check nodes at iteration m . We can describe density evolution as

$$f_m = a_0 \otimes \lambda \left(\Gamma^{-1} \left(\rho \left(\Gamma \left(f_{m-1} \right) \right) \right) \right), \quad (2.73)$$

where

$$\begin{aligned} \lambda(f) &= \sum_j \lambda_j f^{\otimes(j-1)} \\ \rho(f) &= \sum_j \rho_j f^{\otimes(j-1)}, \end{aligned} \quad (2.74)$$

and \otimes denotes convolution and the operators Γ^{-1} and Γ are defined in [29]. Density evolution can be used to find the threshold of LDPC codes, and, moreover, we can combine density evolution with an optimization strategy to design degree distributions with a threshold close to the capacity of the channel. We will combine density evolution with EXIT chart design.

EXIT Chart Design with a Gaussian Approximation

EXIT charts are introduced in [32] and in [33] it is shown that EXIT chart techniques are very useful for the design of LDPC codes. In EXIT chart analysis we parameterize the ℓ -densities which are observed during decoding by a single parameter. Several choices are possible and this parameter can be taken as e.g. mutual information, entropy or mean of the density under consideration. We follow [29] and use the entropy corresponding to the ℓ -density as the main parameter. Furthermore, we assume that the ℓ -densities are from some convenient family of densities \mathcal{F} . In its original form, EXIT chart analysis uses a Gaussian approximation where \mathcal{F} is taken as the family of symmetric Gaussian ℓ -densities which we denote by $\mathcal{F}_{\mathcal{N}}$. Other choices are possible and useful in obtaining accurate one-dimensional approximations to density evolution.

The next step is to derive the so-called EXIT function for the variable nodes and check nodes. The EXIT function defines the input-output relation of the single parameter characterizing the density. For several families of codes there is some parameter for which the input-output relation is easy to derive. For the variable nodes in LDPC codes this parameter is the mean of the ℓ -density. To derive the entropy input-output relation for a variable node it is convenient to define a function $\psi_{\mathcal{F}}$ which given a family of ℓ -densities \mathcal{F} converts the mean of the ℓ -density to the entropy of the ℓ -density

$$\psi_{\mathcal{F}} : \mathbb{R} \mapsto [0, 1]. \quad (2.75)$$

We require that \mathcal{F} is chosen in such a way that this map is injective. Given a family of ℓ -densities \mathcal{F} , the input-output entropy relation of a variable node of degree i is denoted by $v_{\mathcal{F},i}$ and can be expressed as

$$v_{\mathcal{F},i}(h, h_0) = \psi_{\mathcal{F}} \left((i-1)\psi_{\mathcal{F}}^{-1}(h) + \psi_{\mathcal{F}}^{-1}(h_0) \right), \quad (2.76)$$

where h_0 denotes the entropy corresponding to the channel ℓ -density and h the input entropy. Using the duality of variable nodes and check nodes [29] a similar relation can be derived for the input-output entropy relation of a check

node. The input-output entropy relation for a check node of degree j is denoted by $c_{\mathcal{F},j}$ and can be expressed as

$$c_{\mathcal{F},j}(h) = 1 - \psi_{\mathcal{F}} \left((j-1)\psi_{\mathcal{F}}^{-1}(1-h) \right). \quad (2.77)$$

Now, let (λ, ρ) be a degree distribution pair defining an ensemble of LDPC codes. The average input-output entropy relation of the variable nodes is denoted by $v_{\mathcal{F},\lambda}$ and given by

$$v_{\mathcal{F},\lambda}(h, h_0) = \sum_i \lambda_i v_{\mathcal{F},i}(h, h_0). \quad (2.78)$$

In a similar way we define $c_{\mathcal{F},\rho}$ as the average input-output relation of the check nodes

$$c_{\mathcal{F},\rho}(h) = \sum_j \rho_j c_{\mathcal{F},j}(h). \quad (2.79)$$

Under the approximation that during decoding the messages of LLRs are distributed according to densities from \mathcal{F} , we can approximate the entropies of the densities during decoding for a certain degree distribution pair. Let h_m denote the average output entropy of the variable nodes during decoding at iteration m . We assume decoding is initialized by sending the channel LLRs to the check nodes at iteration 0. Given a degree distribution pair (λ, ρ) the output entropy of the variable nodes at iteration m is given by

$$h_m = \sum_i \lambda_i v_{\mathcal{F},i}(c_{\mathcal{F},\rho}(h_{m-1}), h_0), \quad m \geq 1. \quad (2.80)$$

The key to the design with EXIT charts is the observation that for successful decoding we should have $h_m \leq h_{m-1}$ for $m \geq 1$. In case we fix ρ and we want to design λ we can formulate the following optimization problem. To design λ , we ask for the highest rate code such that $h_m \leq h_{m-1}$ for $h_{m-1} \in [0, 1]$. Since ρ is fixed the rate of an LDPC code is proportional to $\sum_i \frac{\lambda_i}{i}$ and the optimization problem is easily formulated as a linear program which can be solved efficiently:

$$\begin{aligned} & \text{maximize } \sum_i \frac{\lambda_i}{i} \\ & \text{subject to } \sum_i \lambda_i = 1, \\ & \sum_i \lambda_i v_{\mathcal{F},i}(c_{\mathcal{F},\rho}(h), h_0) \leq h \text{ for } h \in [0, 1] \end{aligned} \quad (2.81)$$

Improving EXIT Chart Design with Density Evolution

The EXIT chart design can be improved by a simple method which is proposed in [31]. The idea is to start with an initial EXIT chart design by solving (2.81) for $\mathcal{F} = \mathcal{F}_N$. Next, density evolution is used to find the threshold of the designed degree distribution pair (λ, ρ) . At the SNR where density evolution just converges to negligible bit-error rate, we compute the densities originating from the variable nodes and check nodes. Moreover, we can compute the “real” entropy input-output relation of the variable and check nodes. Now, to design an improved λ we proceed as follows. Let $h_{\mathcal{D},m}$ denote the sequence of entropies corresponding to the ℓ -densities of the variable node output messages which we compute with density evolution for $m \geq 0$. For $m = 0$ we have $h_{\mathcal{D},0} = h_0$. For each $h_{\mathcal{D},m}$, we denote the average output entropy of the check node by $c_{\mathcal{D},\rho}(h_{\mathcal{D},m})$. The average output entropy h_m of the variable nodes at iteration $m \geq 1$ is given by

$$h_m = \sum_i \lambda_i v_{\mathcal{D},i}(c_{\mathcal{D},\rho}(h_{\mathcal{D},m-1}), h_0), \quad (2.82)$$

where $v_{\mathcal{D},i}(c_{\mathcal{D},\rho}(h_{\mathcal{D},m-1}), h_0)$ denotes the output entropy of a degree i variable node when the input entropy is $c_{\mathcal{D},\rho}(h_{\mathcal{D},m-1})$ and the entropy of the channel ℓ -density is h_0 . We compute $v_{\mathcal{D},i}$ during density evolution and to improve the EXIT chart design method we define $v'_{\mathcal{F},i}(c_{\mathcal{D},\rho}(h_{\mathcal{D},m}), h_0)$ as

$$v'_{\mathcal{F},i}(c_{\mathcal{D},\rho}(h_{\mathcal{D},m}), h_0) = \gamma_i(h_{\mathcal{D},m}) v_{\mathcal{F},i}(c_{\mathcal{D},\rho}(h_{\mathcal{D},m}), h_0), \quad (2.83)$$

where $\gamma_i(h_{\mathcal{D},m})$ is given by

$$\gamma_i(h_{\mathcal{D},m}) = \frac{v_{\mathcal{D},i}(c_{\mathcal{D},\rho}(h_{\mathcal{D},m}), h_0)}{v_{\mathcal{F},i}(c_{\mathcal{D},\rho}(h_{\mathcal{D},m}), h_0)}. \quad (2.84)$$

The result is that for the initial degree distribution we have computed a correction $\gamma_i(h_{\mathcal{D},m})$ such that $v'_{\mathcal{F},i}$ agrees with the entropies computed during density evolution. We can replace $v_{\mathcal{F},i}$ in (2.81) by $v'_{\mathcal{F},i}$ and solve the linear program for the sequence of $h_{\mathcal{D},m}$. Hopefully by using $v'_{\mathcal{F},i}$ the approximation will improve. Furthermore, the role of λ and ρ can be exchanged to design ρ . Finally, one can perform several iterations between linear programming and density evolution. Our observation is that 2 or 3 iterations usually suffice to get good results.

Rate Design

In a system employing multistage decoding we opt for a target bit-error rate (BER) of p_b at a certain SNR. We can achieve this overall p_b if we achieve a

BER of p_b on each of the equivalent binary channels. However, each of these binary channels behaves differently for finite block lengths. The reason for this is that the noise densities of these channels are different. The first level which is decoded will have Gaussian noise and noise from the other bits. However, the last level will only have Gaussian noise. As a consequence we cannot just design codes with rates equal to the capacity of the equivalent binary channels at a target SNR. For finite block lengths each channel requires a different back-off from the target SNR to achieve a BER of p_b . Hence to achieve a BER of p_b for the target SNR on each of the equivalent binary channels, we need to choose the code rates low enough such that this BER is achieved.

Rate design techniques for multilevel coded systems are described and compared in [23], but none of these techniques gives satisfactory results for LDPC codes under iterative decoding. We expect that an extension of the finite-length scaling techniques for LDPC codes as introduced in [34] can be very useful for the purpose of rate design. However, these techniques are not yet fully developed for general BIOS channels and general decoders. In this chapter we use a rudimentary approach to rate design which can be summarized as follows

1. Consider the design of a set of LDPC codes for the equivalent binary channels at an SNR given by SNR_t . Let C_l denote the capacity of the binary channel at SNR_t and we design an LDPC code of rate $r = C_l$ for the equivalent binary channel.
2. For a codeword length n and target BER p_b , we determine the SNR for which p_b is achieved by means of a Monte Carlo simulation. Now, denote this SNR by SNR' and define SNR_Δ as $\text{SNR}_\Delta = \text{SNR}' [\text{dB}] - \text{SNR}_t [\text{dB}]$. Next, we design an LDPC code for the equivalent binary channel at a rate equal to the capacity of the equivalent binary channel at an SNR of $\text{SNR}_t - \text{SNR}_\Delta$.

An additional Monte Carlo simulation can be performed and step 2 can be repeated. One or two Monte Carlo simulations usually suffice to give good results. In the next section we illustrate the design process for constellation A and B.

2.5 Design Examples and Simulation Results

In this section we present the overall performance of constellation A and constellation B with LDPC codes. First, we choose the decoding order of the equivalent binary channels and we illustrate the EXIT chart design process combined

i	Constellation A		Constellation B	
	α_i	C_i	α_i	C_i
1	0.4636	0.166	0.5107	0.239
2	0.4636	0.216	0.3986	0.219
3	0.4636	0.305	0.3475	0.266
4	0.4636	0.501	0.2591	0.446
5	0.2901	0.373	0.1498	0.881
6	0.2373	0.433	0.0634	0.918
7			0.5260	1.000
8			0.2989	1.000

Table 2.2: The order of α_i for constellation A and B.

with density evolution. Second, we present the design of LDPC codes for the equivalent channels corresponding to constellation A and B and evaluate their performance on these channels by simulations.

2.5.1 Decoding Order and Equivalent Binary Channels

Constellation A and B are defined in Table 2.1. As pointed out before, the constrained constellation capacity does not depend on the order in which the α_i are given. However, we assume that we decode the levels in the order X_1, X_2, \dots, X_d and the capacities of the equivalent binary channels do depend on the order of the α_i . We wish to choose the order in such a way that the capacities of the equivalent binary channels are not too low which would require low-rate codes. Furthermore, if the capacity of one of the equivalent binary channels is nearly equal to 1 that channel can be used uncoded. Table 2.2 gives the decoding order we use for constellation A and B.

The resulting capacities of the equivalent binary channels at the SNR for which the constellations have been designed are given as well. All equivalent binary channels corresponding to constellation A have to be used coded. However, for constellation B the channels corresponding to X_7 and X_8 can be used uncoded. Note that there are $d!$ permutations of the α_i and we just have chosen an order which is acceptable in terms of required code rates.

2.5.2 Illustration of EXIT Chart Design

Figure 2.10 shows the ℓ -densities of the equivalent binary channels of constellation A at SNR = 11.76 dB. The ℓ -densities defining the channels for X_1 to X_5

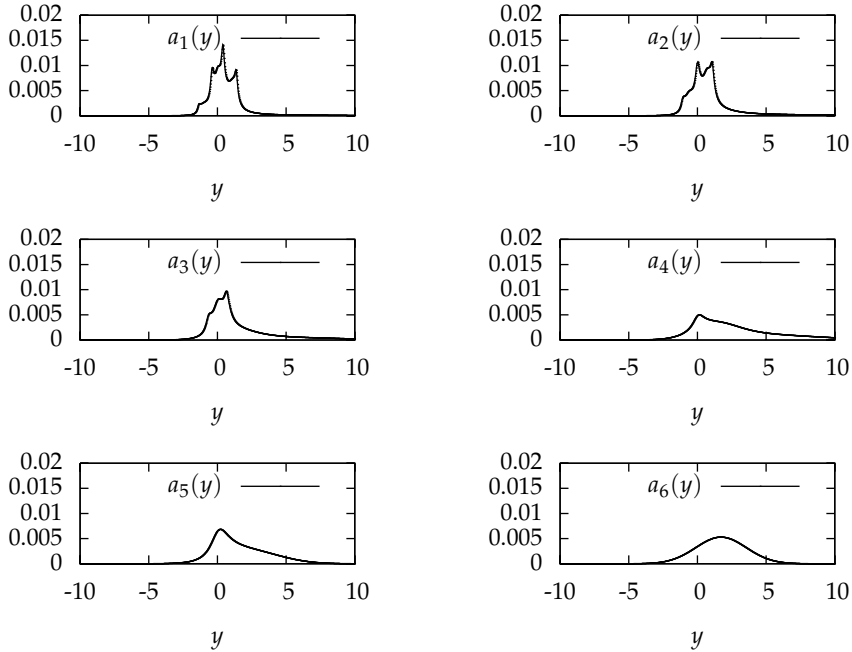


Figure 2.10: The ℓ -densities of the equivalent binary channels for constellation A at SNR = 11.76 dB.

are clearly non-Gaussian. The ℓ -density of X_6 is Gaussian since all the previous levels are assumed to be decoded correctly and only Gaussian noise is left. As an example consider the design of LDPC codes for the equivalent binary channel of X_1 corresponding to constellation A. We assume that the SNR is equal to 11.76 dB and at this SNR the capacity of the equivalent binary channel is 0.166 bit/use. We design a degree distribution pair with a maximum variable node degree of 50 by solving the linear program of (2.81) where \mathcal{F} is taken to be the family of symmetric Gaussian densities $\mathcal{F}_{\mathcal{N}}$. This gives the following degree distribution pair

$$\begin{aligned}
 \lambda(x) &= 0.3240x + 0.2526x^2 + 0.0046x^5 + 0.2078x^6 + 0.0404x^{14} \\
 &\quad + 0.0426x^{15} + 0.0894x^{32} + 0.0117x^{35} + 0.0269x^{36} \\
 \rho(x) &= 0.7279x^3 + 0.2721x^4 \quad (2.85)
 \end{aligned}$$

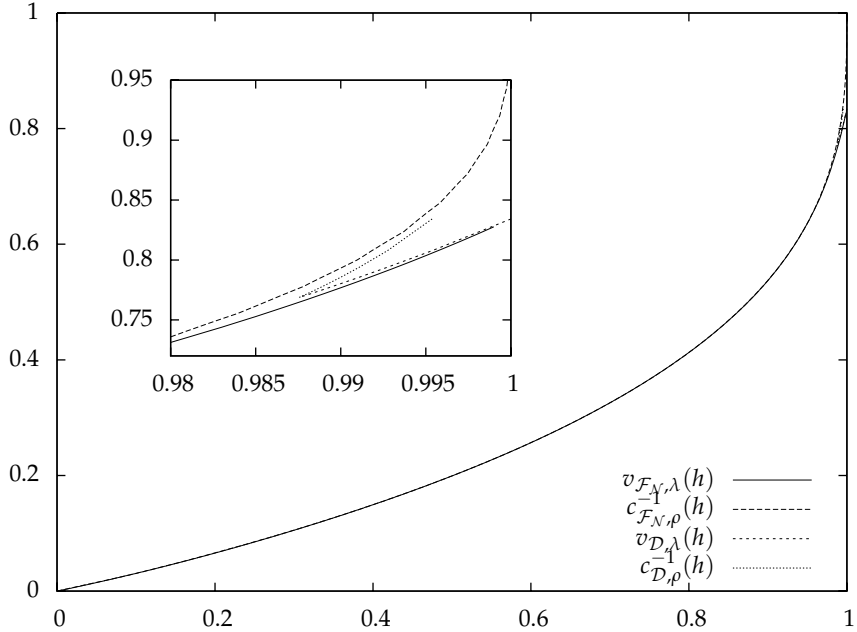


Figure 2.11: The EXIT chart entropy curves compared with density evolution entropy curves.

This pair corresponds to an LDPC code of rate 0.173 and obviously the threshold of this pair can not be at $\text{SNR} = 11.76$ dB since the capacity of the channel is only 0.166. With density evolution we find a threshold at $\text{SNR} = 18.92$ dB. At this SNR the capacity of the channel for X_1 is equal to 0.189 bit/use. Hence this degree distribution pair is able to transmit at about 90% of the capacity of the equivalent binary channel for codeword lengths tending to infinity. Both in terms of rate and SNR the distance to capacity is substantial. Figure 2.11 shows the EXIT curves for the degree distribution pair of (2.85) at $\text{SNR} = 11.76$ dB for $\mathcal{F} = \mathcal{F}_N$. The figure also shows the entropies corresponding to the densities computed with density evolution. The EXIT curves obtained from the Gaussian approximation match well. However, the entropy curves computed with density evolution hit a fixed point after about 30 iterations. Hence iterative decoding will not succeed and if we wish to achieve near-capacity performance,

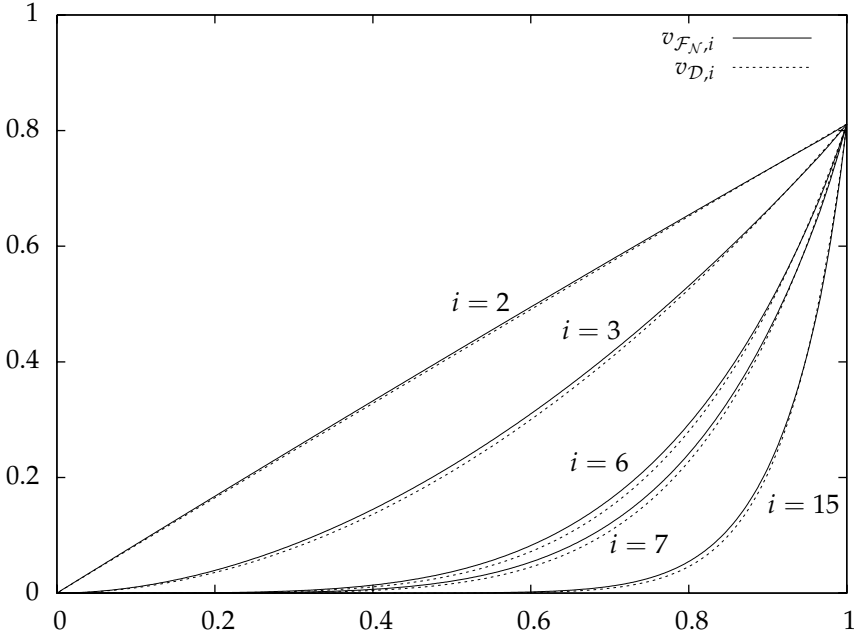


Figure 2.12: EXIT functions of several variable node degrees.

EXIT chart design based on a Gaussian assumption is not sufficient. Figure 2.12 shows the EXIT curves for several variable node degrees at an SNR slightly higher than the threshold.

We clearly observe a mismatch between the curves, especially for higher degree nodes. We can improve on the EXIT chart design method as described in the previous section. We use the results obtained by density evolution to compute the correction from (2.83). We solve the linear program of (2.81) with the computed correction and this gives the following degree distribution pair

$$\begin{aligned}
 \lambda(x) &= 0.3283x + 0.2273x^2 + 0.1315x^6 + 0.08393x^7 + 0.1052x^{18} \\
 &\quad + 0.003x^{19} + 0.1205x^{49} \\
 \rho(x) &= 0.6253x^3 + 0.3747x^4 \quad (2.86)
 \end{aligned}$$

The rate of the corresponding code is 0.166 and the threshold of the code is at

SNR = 13.72 dB. At this SNR the capacity of the equivalent binary channel is 0.172 bit/use and the code is able to transmit at 97% of the capacity. This is a substantial improvement compared to the Gaussian approximation and when we perform an additional iteration between density evolution and linear programming this results in a further improvement. Finally, note that there is a gap of almost 2 dB between the threshold of the code and the capacity of the channel. This seems substantial, but for this equivalent binary channel the capacity hardly depends on the SNR. Hence the loss in rate is only small and this is what matters in the end.

2.5.3 LDPC Codes for the Equivalent Binary Channels

LDPC Codes for Constellation A

For constellation A we design LDPC codes for a codeword length of $n = 10^6$ where we use a maximum variable node degree of 100. The degree distributions of the LDPC codes for constellation A are given in Table 2.3. The table gives the rates r_i of the codes and their thresholds which are denoted by SNR_* .

We transmit at a total rate R of 1.94 bit/use and simulation results are given in Figure 2.13. The LDPC codes are decoded by the sum-product algorithm and the figure shows the performance of the individual codes and the total average performance. In the first case, the simulation is performed with perfect decisions at the higher levels, i.e. there is no error propagation between different levels. In the second case, error propagation is taken into account. We observe that once the bit-error rate (BER) has reached a sufficiently low value, error propagation does not play a role. If we compare the performance of the levels, we observe that the behavior of the binary channels of the levels is different. This is a consequence of the difference in the noise structure of these binary channels. A BER of 10^{-5} is achieved at an SNR of 11.76 dB, which corresponds to a distance of 0.34 dB to the constellation capacity limit. The distance to the capacity of the Gaussian channel is 0.39 dB. Moreover, we achieve this BER 0.34 dB before the constrained capacity limit of a 32-PAM constellation.

LDPC Codes for Constellation B

For constellation B we design LDPC codes for codeword lengths of $n = 32000$ and $n = 320000$ for X_1 to X_6 where we use a maximum variable node degree of 20 and 50, respectively. The equivalent binary channels for X_7 and X_8 can

Level	X_1	X_2	X_3	X_4	X_5	X_6
λ_2	0.278549	0.248909	0.221353	0.170690	0.197001	0.173440
λ_3	0.202881	0.195496	0.189844	0.180700	0.188196	0.182454
λ_7	0.130993	0.140359	0.127853	0.092980	0.102967	0.099392
λ_8	0.076606	0.068553	0.082873	0.121127	0.111663	0.117011
λ_{15}						
λ_{17}						
λ_{20}						
λ_{21}		0.009717				
λ_{22}	0.075288	0.145815	0.130456	0.019257	0.031783	0.015465
λ_{23}	0.071948		0.035774	0.164600	0.142729	0.169356
λ_{100}	0.163735	0.191151	0.211847	0.250648	0.225662	0.242882
ρ_4	0.10196					
ρ_5	0.89804	0.403370				
ρ_6		0.596630	0.27505			
ρ_7			0.724950		0.13070	
ρ_8					0.86930	
ρ_9						0.67462
ρ_{10}				0.32433		0.32538
ρ_{11}				0.67567		
r_i	0.157	0.207	0.296	0.491	0.363	0.425
SNR* [dB]	10.36	10.98	11.30	11.52	11.59	11.70

Table 2.3: Degree distributions designed for constellation A for $n = 10^6$, total rate $R = 1.939$.

be used uncoded. The degree distributions of the LDPC codes for $n = 32000$ are given in Table 2.4. There is a large spread in the thresholds of the codes for the different levels. For instance the code for X_1 has a threshold of 13.61 dB. In other words, we require a back-off of $30.10 - 13.61 = 16.49$ dB to achieve a low BER at the target SNR. However, the capacity of the equivalent binary channel of X_1 hardly varies with the SNR. Between SNR= 13.69 dB and SNR= 30.10 dB the difference in capacity of the channel for X_1 is only 0.03 bit/use. Transmission takes place at a rate of 4.77 bit/use and Figure 2.14 shows the simulation results. The figure shows the performance of the individual codes and the total average performance. For a codeword length of 32000 error propagation hardly plays a role once the BER is sufficiently low. An average BER of 10^{-4} is achieved at an SNR of 30 dB, which corresponds to a distance of 1.14 dB

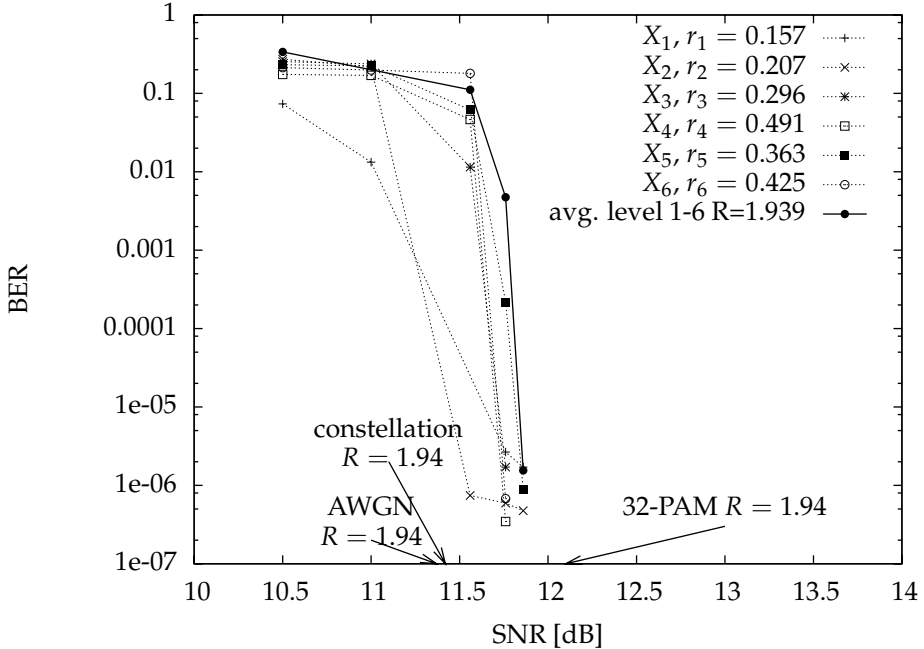


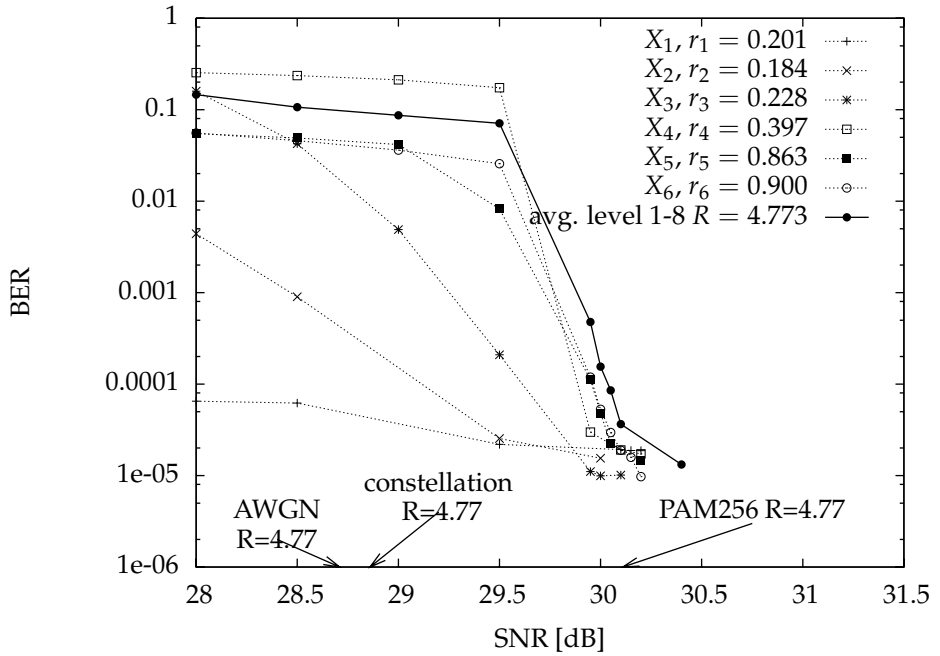
Figure 2.13: Simulation results for constellation A, $n = 10^6$

to the constellation capacity limit. The distance to the capacity of the AWGN channel is 1.27 dB. Furthermore, this BER is achieved just before the 256-PAM constrained capacity limit.

The degree distributions for a codeword length of 320000 for X_1 to X_6 are given in Table 2.5. Compared with the codeword length of 32000 the rates of the codes can be chosen higher and transmission takes place at a rate R of 4.90 bit/use. The simulation results are shown in Figure 2.15. A BER $< 10^{-5}$ is achieved at an SNR slightly higher than 30 dB. Hence a low BER is achieved within 0.6 dB of the capacity of the AWGN channel. This BER is achieved far before the 256-PAM capacity limit which shows that with a 256-PAM constellation it is unlikely that one can achieve the same performance. Compared with constellation A it is easier to get close to the constrained capacity limit of the constellation. Constellation A is generated by a map which is not injective and in general it seems harder to get close in that case.

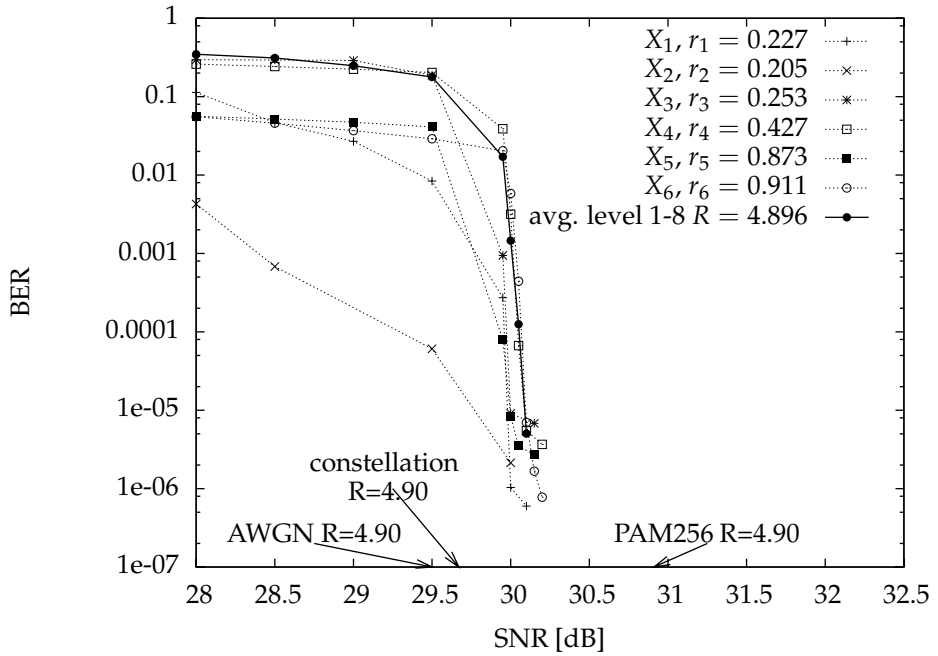
Level	X_1	X_2	X_3	X_4	X_5	X_6
λ_2	0.362979	0.361939	0.335337	0.277036	0.196059	0.147179
λ_3	0.299368	0.279302	0.270413	0.233689	0.239578	0.261360
λ_6				0.060805		
λ_7		0.108425	0.154350	0.163249	0.235428	0.215362
λ_8	0.124747	0.054619	0.034019		0.002427	0.080529
λ_9	0.086918					
λ_{19}						0.220558
λ_{20}	0.125988	0.195714	0.205882	0.265221	0.326509	0.075012
ρ_4	1.000000	1.000000	0.545450			
ρ_5			0.454550			
ρ_6				0.666667		
ρ_7				0.333333		
ρ_{32}					1.000000	
ρ_{46}						1.000000
r_i	0.201	0.184	0.228	0.397	0.863	0.900
SNR* [dB]	13.61	22.51	27.45	29.54	29.19	29.78

Table 2.4: Degree distributions designed for constellation A for $n = 32000$, total rate $R = 4.773$.

Figure 2.14: Simulation results for constellation B, $n = 32000$.

Level	X_1	X_2	X_3	X_4	X_5	X_6
λ_2	0.546364		0.281285	0.218469	0.165561	0.103892
λ_3	0.322008		0.176337	0.200442	0.180557	0.191428
λ_4						0.019178
λ_6			0.165870			
λ_7				0.187371	0.173931	
λ_8	0.075676			0.009430	0.038578	
λ_9	0.0297550					0.172671
λ_{10}						0.135522
λ_{11}			0.017232			
λ_{12}			0.116111			
λ_{14}				0.0588134		
λ_{15}				0.0764956		
λ_{16}					0.130306	
λ_{18}					0.019119	
λ_{24}	0.015300					
λ_{26}	0.002224					
λ_{50}	0.008673		0.243164	0.248979	0.291948	0.377308
ρ_4	0.081633					
ρ_5	0.918367		0.45455			
ρ_6			0.54545			
ρ_7						
ρ_8				1.000000		
ρ_9						
ρ_{42}					1.000000	
ρ_{70}						1.000000
r_i	0.226		0.252	0.427	0.873	0.911
SNR* [dB]	28.00		29.47	29.91	29.62	29.88

Table 2.5: Degree distributions designed for constellation B for $n = 320000$, total rate $R = 4.896$

Figure 2.15: Simulation results for constellation B, $n = 320000$

2.6 Conclusions

We have shown that with the superposition of binary random variables, signal constellations can be generated with a constrained capacity limit very close to the capacity of the AWGN channel. We have designed signal constellations for spectral efficiencies from 4 bit/s/Hz to 10 bit/s/Hz which have their constrained capacity limit within 0.1 dB of the capacity of the AWGN channel. For the constellations we have designed we achieve shaping gains ranging from 0.5 to 1.4 dB compared with conventional PAM constellations.

We have shown that combined with multistage decoding at the receiver the problem of achieving capacity on the AWGN channel reduces to achieving capacity on a set of binary-input output-symmetric channels. We have considered the design of binary LDPC codes for these channels and show that one can get very close to the capacity of the AWGN channel. With one of the constellations which is designed for 10 bit/s/Hz we achieve a low BER with binary LDPC codes within 0.6 dB of the capacity of the AWGN channel.

2.7 Acknowledgments

The author wishes to thank R.L. Urbanke and A. Amraoui for useful discussions while visiting EPFL, Lausanne and J.H. Weber, C.H. Slump and N.J. Bouman for reviewing a draft version of this chapter .

Chapter 3

Signal Shaping for Bit-Interleaved Coded-Modulation

3.1 Introduction

In this chapter we consider communication over discrete-time memoryless channels where the channel inputs and outputs take values in \mathbb{R} . Furthermore, we assume that the capacity of the channel is significantly larger than 1 bit/use. In this case binary signaling incurs a large loss in rate. The channels we consider often have a continuous-time counterpart and for these channels we are interested in spectral-efficient communication.

We are interested in reliable communication and for this purpose we consider *coded modulation* schemes which are based on binary error-correcting codes. The reason that we opt for binary codes is that these are well understood and the information source is often binary. Furthermore, binary codes are often easier to decode than their counterparts defined over larger alphabets. Finally, some families of binary codes such as low-density parity-check codes [35], [1] can be designed such that a near capacity performance becomes possible on many memoryless binary channels.

Coded modulation based on binary error-correcting codes has gained considerable attention in literature. A well-known scheme is multilevel coding

(MLC) as introduced in [8]. In MLC a set of binary codes is used and the bits of the codewords are mapped to constellation symbols. At the receiver a multistage decoding (MSD) procedure can be used where the codes are decoded in a sequential fashion. Multilevel coding has the potential to achieve a near-capacity performance on several channels. However, a major disadvantage is that for high spectral efficiencies a considerable number of levels is required.

Another major breakthrough was the invention of trellis-coded modulation (TCM) by Ungerboeck [36]. Since [36] it has been accepted that modulation and coding should be combined for improved performance. With TCM a binary code is used to select constellation symbols in such a way that the distance profile of the overall Euclidian space code is improved. On several channels this leads to a large improvement compared to conventional systems.

A scheme that breaks with Ungerboeck's paradigm in the sense that the focus is not on the Euclidian distance profile is bit-interleaved coded-modulation (BICM). The bits of a single codeword are interleaved and mapped to constellation symbols. The overall performance depends on the constellation, the map from bits to constellation symbols, the code used and the decoding strategy. If one uses Gray mapped signal constellations the achievable rate is quite close to the capacity limit of the signal constellation. Another popular decoding strategy is BICM with iterative decoding (BICM-ID) where a relatively simple code is combined with a signal constellation [37], [38]. At the receiver iterations are performed between the constellation demapper and the decoder of the code. A major advantage of BICM over MLC-MSD is that for the same number of channel input symbols a long binary codeword is used instead of several shorter binary codewords.

So far we did not consider the actual channel. For several interesting channels mutual information is maximized for a Gaussian distribution on the channel input. This is for instance the case for the additive white Gaussian noise (AWGN) channel and the Rayleigh fading channel with perfect channel state information at the receiver. When conventional pulse-amplitude modulation (PAM) constellations with a uniform spacing and uniform selection of the constellation symbols are used, capacity cannot be achieved and a so-called shaping gain is available [14].

Several signal shaping techniques are proposed in literature [20], [21], [13], [11]. Following [11], we show in [39] that with a very simple technique based on the superposition of binary random variables one can design signal constellations which are able to perform very close to the capacity of the AGWN channel with LDPC codes in a MLC-MSD setting. The major disadvantage is that for high spectral efficiencies one requires a large number of decoding

stages.

These coded modulation schemes based on binary codes lead to an equivalent binary channel or set of equivalent binary channels at the bit level. This is regardless of the channel and the degrees of freedom such as the signal constellation, the map from bits to constellation symbols and decoding strategy. From an information theoretical point of view we wish to choose these parameters in such a way that the overall achievable rate is close to the capacity of the channel. Moreover, from a practical point of view we want to use a decoding scheme which is favorable in terms of complexity.

Once a choice has been made for the modulation method and decoding scheme one requires a set of codes for the equivalent channels such that a target performance is reached on each of these channels. In this chapter the focus is not on the error-correcting codes, but we assume that a family of powerful binary error-correcting codes is available. This family of codes should be practical in the sense that feasible encoding and decoding algorithms with a good performance exist. Moreover, this family of codes should be amendable to analysis and should have sufficiently many degrees of freedom such that given a channel we can find a code which performs reasonably close to the capacity of this channel.

One such family of codes are binary low-density parity-check (LDPC) codes [35], [1]. The performance of an ensemble of LDPC codes is amendable to analysis as is shown in [40], [41], [26]. Moreover, given a binary-input output-symmetric channel, we can often design LDPC codes which have a performance close to the capacity of the channel under suboptimal decoding [26], [39].

We show how to construct a coded modulation scheme which is able to perform very close to the capacity of the AWGN channel. The method merges the merits of BICM with uniformly spaced pulse-amplitude modulation (PAM) constellations and signal shaping. These results are based on the work presented in [39]. However, unlike in [39], we only require 3 or 4 decoding stages to get close to the capacity of the AWGN channel for several spectral efficiencies. The advantage of this is that there is less error-propagation for short codeword lengths and for most of the source bits a single long codeword can be used. The method we use is based on shaping a conventional uniformly spaced PAM constellation.

The outline of this chapter is as follows. In Section 3.2 we give an overview of coded modulation techniques and decoding strategies. In Section 3.3 we consider signal shaping and introduce a method to generate signal constellations for the AWGN channel which have a constrained capacity limit close to

the capacity of the AWGN channel. Furthermore, these constellations allows one to use a decoding strategy which is favorable in terms of complexity. In Section 3.4 we consider the use of binary codes on the equivalent binary channels defined by the signal constellation and decoding strategy. First, we design M -PAM-LDPC codes which are LDPC codes optimized for uniformly spaced PAM constellations. These extend the results of [7] to higher spectral efficiencies. Second, we design shaped PAM-LDPC codes which provide a shaping gain over M -PAM-LDPC codes. In Section 3.5 we provide some design examples where we design binary LDPC codes for the decoding strategy we propose. We end with conclusions in Section 3.6.

3.2 Coded Modulation

3.2.1 Introduction

Consider a discrete-time memoryless channel with input $X \in \mathbb{R}$ and output $Y \in \mathbb{R}$. Furthermore, the channel is defined by a channel transition density $f_{Y|X}(y|x)$. The capacity of the channel is defined as

$$C = \max_{f_X} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) \log_2 \frac{f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x') f_X(x') dx'} dx dy, \quad (3.1)$$

where f_X defines the distribution of X and capacity is achieved for an optimal f_X . Our main example is the AWGN channel which is defined by

$$Y = X + N, \quad (3.2)$$

where the channel input X is disturbed by zero-mean Gaussian noise N with variance σ^2 . The channel transition density is given by

$$f_{Y|X}(y|x) = f_N(y - x), \quad (3.3)$$

where f_N is the density of N

$$f_N(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{n^2}{2\sigma^2}}. \quad (3.4)$$

The capacity of the AWGN channel is given by the well-known capacity formula

$$C = \frac{1}{2} \log_2 (1 + \text{SNR}), \quad (3.5)$$

where SNR is defined as

$$\text{SNR} = \frac{E_s}{\sigma^2}, \quad (3.6)$$

and where E_s denotes the mathematical expectation of X^2 . Furthermore, capacity is achieved for a Gaussian distribution on X . Thus to achieve capacity we require a coding and modulation scheme which results effectively in a Gaussian distribution on the channel input X .

3.2.2 Signal Constellations and Modulation

To achieve capacity we require a coding scheme which results in the optimal f_X . However, as for the AWGN channel defined in the previous section the optimal distribution can be a continuous distribution. The information source in a digital communication system is usually binary and we use a finite *signal constellation* \mathcal{S} to generate a channel input Z . The elements of \mathcal{S} are the *constellation symbols* and the probability with which they are selected is defined by a probability measure $P_{\mathcal{S}}$

$$\Pr [Z = z] = P_{\mathcal{S}}(z) \text{ for } z \in \mathcal{S}. \quad (3.7)$$

Now, the achievable rate is given by $I(Y; Z)$

$$I(Y; Z) = \int_{-\infty}^{\infty} \sum_{z \in \mathcal{S}} f_{Y|X}(y|z) P_{\mathcal{S}}(z) \log_2 \frac{f_{Y|X}(y|z)}{\sum_{z' \in \mathcal{S}} f_{Y|X}(y|z') P_{\mathcal{S}}(z')} dy. \quad (3.8)$$

We refer to $I(Y; Z)$ as the *constrained constellation capacity* limit. The loss in rate with respect to the optimal f_X depends on the choice of \mathcal{S} and $P_{\mathcal{S}}$. Once a suitable \mathcal{S} and $P_{\mathcal{S}}$ are chosen it is not straightforward to come up with a coding scheme which has practical encoding and decoding algorithms and which generates the desired \mathcal{S} with the probability measure $P_{\mathcal{S}}$. We cannot simply take a code from a random ensemble of codes since description and decoding complexity would be too high.

To solve this problem we assume that we are able to construct practical codes with a codeword alphabet much smaller than the cardinality of \mathcal{S} . Furthermore, we assume that the distribution of the codeword symbols is uniform. We restrict ourselves to binary codes.

Let X_1, \dots, X_d denote a tuple of i.i.d. random bit variables taking values in $\{-1, 1\}$. We can think of X_1, \dots, X_d as a realization of a set of codeword symbols of one or more binary codes. Next, we define a *modulation map* Φ

$$\Phi : \text{GF}(2)^d \rightarrow \mathbb{R}, \quad (3.9)$$

where $x_i \in \{-1, 1\}$ for $i = 1 \dots d$. The modulation map is used to generate a channel input Z from X_1, \dots, X_d

$$Z = \Phi(X_1, \dots, X_d), \quad (3.10)$$

and it defines the signal constellation with its probability measure P_S

$$\mathcal{S} = \{\Phi(x_1, \dots, x_d) | x_1 \in \{-1, 1\}, \dots, x_d \in \{-1, 1\}\} \quad (3.11)$$

$$P_S(z) = \frac{1}{2^d} \sum_{x_1} \dots \sum_{x_d} \mathbb{1}_{\{z\}}(\Phi(x_1, \dots, x_d)), \quad (3.12)$$

where $\mathbb{1}_{\{z\}}$ is the set indicator function which for a set A is defined as

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases} \quad (3.13)$$

The map Φ can be either injective or not injective. In case Φ is not injective the distribution of the constellation symbols is possibly non-uniform. We postpone a discussion on the actual choice of Φ and first consider the combination of modulation with error-control coding.

3.2.3 Coding Schemes and Decoding

Let a channel input Z be generated by a modulation map Φ

$$Z = \Phi(X_1, \dots, X_d), \quad (3.14)$$

where X_1, \dots, X_d are uniform i.i.d. random bit variables and let Y denote the channel output. To compare different coding schemes and decoding strategies our benchmark is achievable rate. For this we consider the mutual information between the channel output Y and X_1, \dots, X_d

$$I(Y; (X_1, \dots, X_d)) = \frac{1}{2^d} \int_{-\infty}^{\infty} \sum_{x_1} \dots \sum_{x_d} f_{Y|X}(y | \Phi(x_1, \dots, x_d)) \cdot \log_2 \frac{f_{Y|X}(y | \Phi(x_1, \dots, x_d))}{\frac{1}{2^d} \sum_{x'_1} \dots \sum_{x'_d} f_{Y|X}(y | \Phi(x'_1, \dots, x'_d))} dy. \quad (3.15)$$

Multilevel Coding with Multistage Decoding

By the chain rule of mutual information we can express $I(Y; (X_1, \dots, X_d))$ as

$$I(Y; (X_1, \dots, X_d)) = I(Y; X_1) + I(Y; X_2|X_1) + \dots + I(Y; X_d|X_1, \dots, X_{d-1}). \quad (3.16)$$

This identity suggests a multilevel encoding procedure with multistage decoding at the receiver [8], [23]. Let \mathcal{C}_i for $i = 1 \dots d$ define a binary error-correcting code of rate r_i and length n . A codeword of \mathcal{C}_i is denoted by \mathbf{x}_i and the j th coordinate of \mathbf{x}_i is denoted by $x_{i,j}$. To generate a channel input word \mathbf{z} we apply Φ in a componentwise fashion to a set of d codewords. The j th component of a channel input word \mathbf{z} is given by

$$z_j = \Phi(x_{1,j}, \dots, x_{d,j}). \quad (3.17)$$

With a slight abuse of notation we write

$$\mathbf{z} = \Phi(\mathbf{x}_1, \dots, \mathbf{x}_d), \quad (3.18)$$

where $\mathbf{z} \in \mathbb{R}^n$ is a vector with channel input symbols.

At the receiver we use a multistage decoding procedure where decoding starts with \mathcal{C}_1 and afterwards we decode $\mathcal{C}_2, \dots, \mathcal{C}_d$ one after the other. While decoding \mathcal{C}_i at level i we make use of the decisions made in the previous levels and we assume we have decoded the codes at these levels correctly. An overview of this system is shown in Figure 3.1. We can achieve the left-handside of (3.16) if the rates of the codes are chosen properly. See e.g. [23] for more details. However, multilevel encoding with several codes and multistage decoding has some disadvantages. First, we cannot start decoding all codes at once when all channel symbols are received. We first require the decision of the lower levels which results in a decoding delay. Moreover, we assume that the decisions on the decoded levels are correct and when this is not the case there can be error propagation. Second, source bits are encoded into d binary codewords of length n . If we were able to use a single codeword of length nd , we would improve the performance for finite codeword lengths.

Multilevel Coding with Parallel and Independent Decoding

To overcome the disadvantage of multistage decoding we reconsider the mutual information between Y and X_1, \dots, X_d and write

$$\begin{aligned} I(Y; (X_1, \dots, X_d)) &= I(Y; X_1) + I(Y; X_2|X_1) + \dots + I(Y; X_d|X_1, \dots, X_{d-1}) \\ &\geq I(Y; X_1) + I(Y; X_2) + \dots + I(Y; X_d), \end{aligned} \quad (3.19)$$

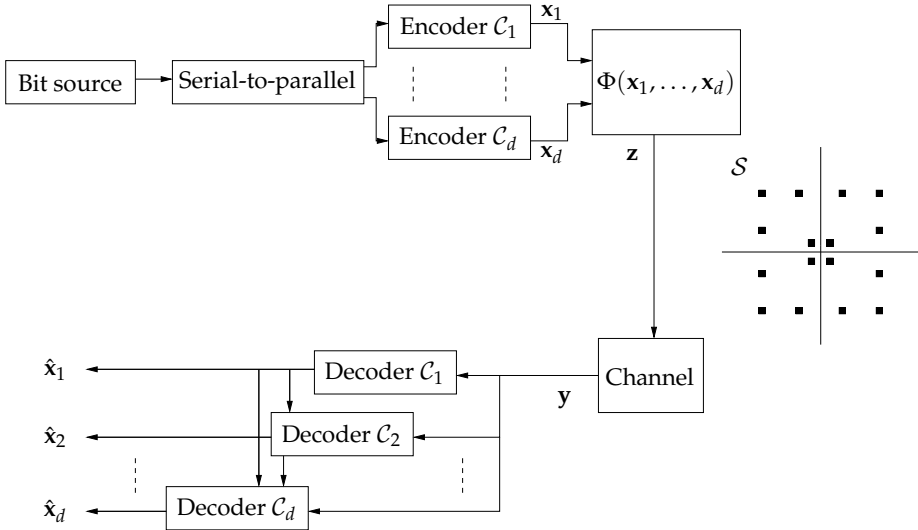


Figure 3.1: Block diagram of multilevel coding with multistage decoding.

where on the right-hand side we have dropped the conditioning. We define

$$I_{\text{PID}}(Y; (X_1, \dots, X_d)) = I(Y; X_1) + I(Y; X_2) + \dots + I(Y; X_d), \quad (3.20)$$

and (3.19) suggests a parallel-and-independent decoding (PID) scheme. We use multilevel encoding at the transmitter and at the receiver we decode \mathcal{C}_1 to \mathcal{C}_d independently. In this case the achievable rate is $I_{\text{PID}}(Y; (X_1, \dots, X_d))$ and the loss with respect to $I(Y; (X_1, \dots, X_d))$ depends on the channel and the modulation map Φ . We will see later that for the AWGN channel and a properly chosen Φ the loss is negligible. The receiver structure of the PID scheme is illustrated in Figure 3.2.

Bit-interleaved Coded Modulation

Parallel-and-independent decoding still uses d codes and to overcome the second disadvantage of multistage decoding, we can use bit-interleaved coded modulation [9] which is shown in Figure 3.3.

In bit-interleaved coded modulation we use a single binary code \mathcal{C} of length nd and rate r . Let $\mathbf{x}' = \pi(\mathbf{x})$ denote a random permutation of the bits of a

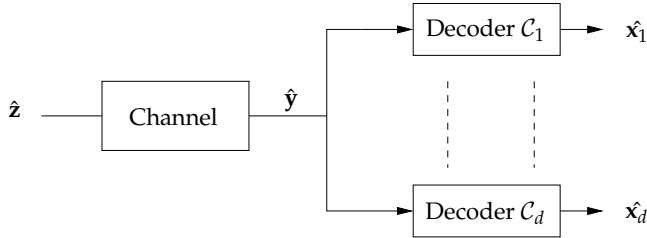


Figure 3.2: Block diagram of parallel-and-independent decoding.

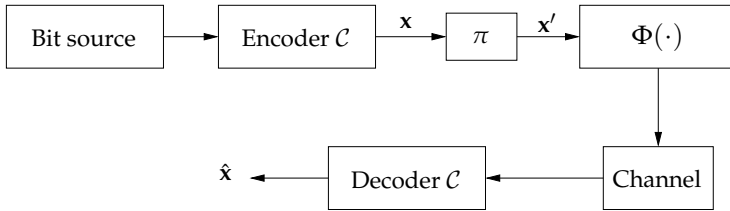


Figure 3.3: Block diagram of bit-interleaved coded modulation.

codeword $\mathbf{x} \in \mathcal{C}$. The j th coordinate of the channel input word \mathbf{z} is generated as

$$z_j = \Phi(x'_{d(j-1)+1}, x'_{d(j-1)+2}, \dots, x'_{d(j-1)+d}). \quad (3.21)$$

The achievable rate with BICM is equal to the achievable rate with PID [9].

In the next section we consider how to choose the modulation map Φ such that $I(Y; (X_1, \dots, X_d))$ and $I_{\text{PID}}(Y; (X_1, \dots, X_d))$ are close to the capacity of the AWGN channel.

3.3 Signal Shaping for Bit-interleaved Coded Modulation

In this section we present a hybrid scheme which merges the merits of BICM with signal shaping. This scheme is able to bridge the shaping gap on the AWGN channel with a small set of binary error-correcting codes.

First, we describe how to choose Φ such that I_{PID} is close to the constrained constellation capacity. The resulting signal constellations are uniformly spaced

PAM constellations with a Gray map from bits to constellation symbols. It is well-known that these constellations perform well in a BICM setting [9]. However, these constellations do not provide any shaping gain. Next, we show how to construct a Φ together with a decoding scheme such that shaping is possible with an acceptable number of multistage decoding levels. Furthermore, we give some numerically optimized Φ which are well-suited for communication on the AWGN channel. These Φ are used in the next section when we consider error-control coding.

3.3.1 Signal Constellations for BICM

A M -PAM constellation with a uniform spacing of the $M = 2^d$ constellation symbols is defined by

$$\mathcal{S} = \left\{ \frac{1}{\sqrt{E_{\text{PAM},d}}} (-M + 2i - 1) \mid i = 1, 2, 3, \dots, 2^d \right\}, \quad (3.22)$$

where $E_{\text{PAM},d}$ is a normalization constant given by

$$E_{\text{PAM},d} = \frac{2^{2d} - 1}{3}, \quad (3.23)$$

such that when the constellation symbols are selected with equal probability, the average energy expended per channel use is unity. The constrained constellation capacity is plotted for several M in Figure 3.4. The figure shows the rate versus the normalized SNR. The normalized SNR is defined as [14]

$$\text{SNR}_{\text{norm}} = \frac{\text{SNR}}{2^{2R} - 1}, \quad (3.24)$$

where R denotes the rate at which a scheme transmits. The value of SNR_{norm} signifies how far a signal constellation is operating from the capacity of the AWGN channel. The baseline performance is $\text{SNR}_{\text{norm}} = 0$ dB which is the required SNR to achieve any rate on the AWGN channel for a Gaussian channel input. We observe that depending on the transmission rate there is a substantial gap to the capacity of the AWGN channel. For instance to transmit at a rate of 3 bit/use with a 16-PAM constellation there is a shaping gain available of 1.19 dB.

Let x_1, \dots, x_d denote the realization of a tuple of binary random variables where each x_i takes values in $\{-1, 1\}$. A natural labeling of the symbols of

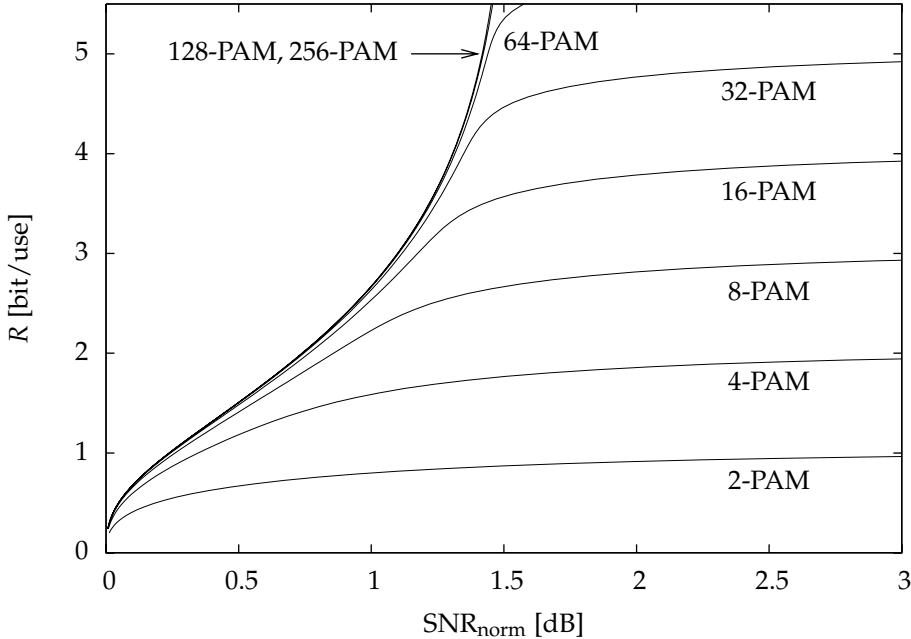


Figure 3.4: The constrained capacity limits of the PAM constellations.

an M -PAM constellation is generated by the superposition of d binary random variables where $M = 2^d$

$$\Phi_{\text{nat}}(x_1, \dots, x_d) = \frac{1}{\sqrt{E_{\text{PAM},d}}} \cdot \sum_{i=1}^d 2^{i-1} x_i. \quad (3.25)$$

As an alternative a Gray labeling of the constellation symbols can be generated by a Φ of the form

$$\Phi_{\text{gray}}(x_1, \dots, x_d) = \frac{1}{\sqrt{E_{\text{PAM},d}}} \cdot \sum_{i=1}^d \left(\prod_{j=i}^d x_j \right) 2^{i-1}. \quad (3.26)$$

The achievable rate does not depend on the labeling when a multistage decoding procedure is used. However, the achievable rate with PID and BICM

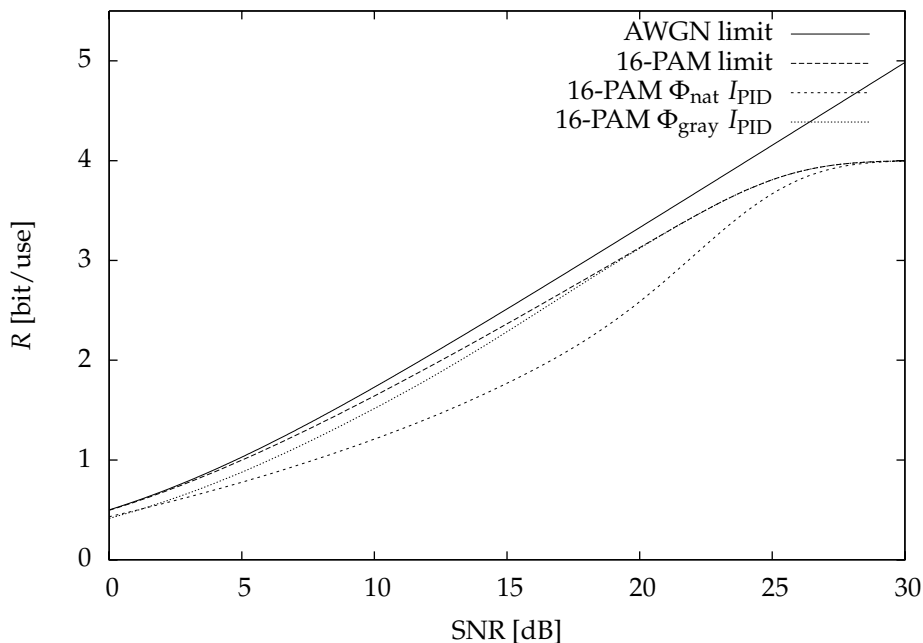


Figure 3.5: Capacity curves for a 16-PAM constellation with several decoding strategies.

does depend on the labeling of the constellation symbols. Figure 3.5 shows a plot of the constrained constellation capacity of a 16-PAM constellation together with I_{PID} for a 16-PAM constellation generated by (3.25) and (3.26). We observe that the use of (3.25) results in a substantial loss with respect to the constrained constellation capacity. The use of (3.26) results in a small loss especially in the regime where the rate is above 75 percent of the constrained constellation capacity. For different values of M results are similar and we conclude that modulation maps given by (3.26) are well-suited for BICM. However, there is still a gap to the capacity of the AWGN channel and our next topic is how to bridge this gap.

3.3.2 Signal Constellations for Shaping

In [39] we show that with superposition modulation we can generate signal constellations with a constrained constellation capacity very close to the capacity of the AWGN channel. In this case Φ is given by

$$\Phi_{\text{spm}}(x_1, \dots, x_d) = \sum_{i=1}^d \alpha_i x_i \quad \alpha_i \in \mathbb{R}. \quad (3.27)$$

The potential of (3.27) to generate good signal constellations can be seen as follows. Suppose we choose $\alpha_i = 1$ and let $d \rightarrow \infty$. In this case the distribution of

$$Z = \frac{1}{\sqrt{d}} \sum_{i=1}^d X_i, \quad (3.28)$$

converges to the Gaussian distribution by the central limit theorem. Furthermore, for finite values of d we can numerically maximize the constrained constellation capacity. In [39] several design examples are presented. For instance to transmit at a rate of 5 bit/use we can use the following Φ

$$\begin{aligned} \Phi(x_1, \dots, x_d) = & 0.0634x_1 + 0.1498x_2 + 0.2591x_3 + 0.2989x_4 + 0.3475x_5 \\ & + 0.3986x_6 + 0.5107x_7 + 0.5260x_8. \end{aligned} \quad (3.29)$$

Figure 3.6 shows the constrained capacity limit of the resulting signal constellation. We observe that for a wide range of SNRs the constrained constellation capacity limit is very close to the capacity of the AWGN channel. For a Gaussian channel input a rate of 5 bit/use is achieved at SNR = 30.10 dB. The signal constellation defined by (3.29) requires 30.28 dB to achieve 5 bit/use. However, a conventional 256-PAM constellation with a uniform spacing and uniform distribution on the constellation symbols requires 31.55 dB to achieve the same rate. Hence the constellation achieves a shaping gain of 1.27 dB.

For modulation maps of the form given by (3.27) the constrained constellation capacity is achievable with multilevel encoding and multistage decoding. However, the disadvantage is that a large number of levels is required with multistage decoding. For instance for the map of (3.29), we require a set of 8 binary codes.

The figure also shows the parallel-and-independent decoding capacity limit which can be achieved with a code in the BICM setting. Unfortunately, the loss with respect to the constrained constellation capacity limit is substantial and bit-interleaved coded modulation will only be able to achieve a small fraction of

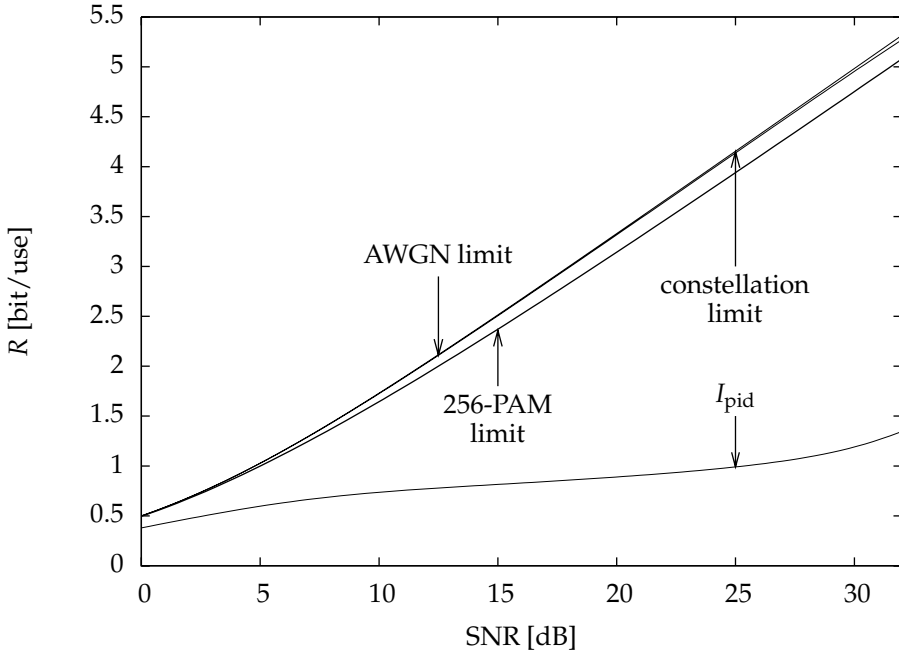


Figure 3.6: The capacity limit of the signal constellation for 5 bit/use.

the capacity of the AWGN channel. One can think of changing the natural labeling of the constellation symbols into a Gray labeling, but this hardly improves the PID capacity. In [9] it is actually conjectured that uniformly spaced PAM constellations with a Gray map from bits to constellation symbols maximize the PID capacity. Next, we show how to choose Φ such that the capacity of the AWGN channel can be approached closely with a decoding scheme that requires only a few levels of multistage decoding.

3.3.3 Shaping of PAM Constellations for BICM

To generate a signal constellation with a constrained constellation capacity close to the capacity of the AWGN channel, we can use a modulation map

of the form

$$\Phi_{\text{spm}}(x_1, \dots, x_d) = \sum_{i=1}^d \alpha_i x_i, \quad (3.30)$$

where the α_i are chosen such that $I(Y; Z)$ is maximized as is shown in [39]. With multilevel encoding and multistage decoding we can in principle achieve $I(Y; Z)$. However, the number of codes required is equal to the number of levels. In general the difference between the constrained constellation capacity and the PID capacity of the constellations generated by (3.30) is large. Signal constellations with a PID capacity close to the constrained constellation capacity are generated by modulation maps of the form

$$\Phi_{\text{gray}}(x_1, \dots, x_d) = \frac{1}{\sqrt{E_{\text{PAM},d}}} \cdot \sum_{i=1}^d \left(\prod_{j=i}^d x_j \right) 2^{i-1}. \quad (3.31)$$

However the resulting signal constellations do not provide any shaping. To have a shaping gain and the possibility to use a BICM code, we combine (3.30) and (3.31) to construct a modulation map of the form

$$\Phi_{\text{hybrid}}(x_1, \dots, x_d) = \sum_{i=1}^{d_1} \alpha_i x_i + \beta \cdot \frac{1}{\sqrt{E_{\text{PAM},d_2}}} \cdot \sum_{i=d_1+1}^{d_1+d_2} \left(\prod_{j=i}^{d_1+d_2} x_j \right) 2^{i-1}, \quad (3.32)$$

where $\alpha_i \in \mathbb{R}$ and $\beta \in \mathbb{R}$. Furthermore d_1 and d_2 are positive integers with $d = d_1 + d_2$. Essentially, a base signal constellation is generated by Φ_{gray} and this constellation is scrambled by d_1 random bit variables in such a way that a shaping gain is achieved.

Now, we use Φ_{hybrid} as follows. Let \mathcal{C}_i for $i = 1 \dots d_1$ denote a binary code of length n and rate r_i . A codeword of \mathcal{C}_i is denoted by \mathbf{x}_i and the j th coordinate of \mathbf{x}_i by $x_{i,j}$. Next, let \mathcal{C} denote a binary code of length nd_2 and rate r . We denote a codeword of \mathcal{C} by \mathbf{x} and a random permutation of the bits of a codeword by \mathbf{x}' . Furthermore, the j th coordinate of \mathbf{x}' is denoted by x'_j . The j th coordinate of the channel input word \mathbf{z} is generated as

$$z_j = \Phi_{\text{hybrid}}(x_{1,j}, \dots, x_{d_1,j}, x'_{d_2(j-1)+1}, x'_{d_2(j-1)+2}, \dots, x'_{d_2(j-1)+d_2}). \quad (3.33)$$

The total transmission rate R of this scheme is

$$R = \sum_{i=1}^{d_1} r_i + d_2 r. \quad (3.34)$$

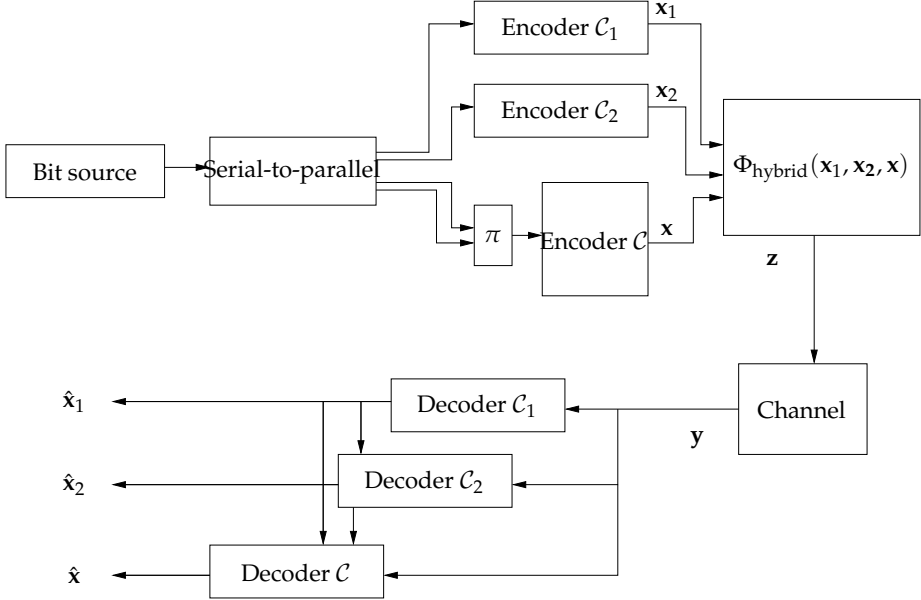


Figure 3.7: Hybrid scheme combining multistage decoding with bit-interleaved coded modulation.

At the receiver we first decode C_1 to C_{d_1} with a multistage decoding approach. Next, we decode C where we make use of the decisions made on C_1 to C_{d_1} . An overview of this scheme is given in Figure 3.7. The potential of this scheme is that with only a few (3 or 4) multistage decoding levels we are able to get close to the capacity of the AWGN channel for any spectral efficiency. Furthermore, most of the source bits will be encoded by C which takes full advantage of the larger codeword length.

The constrained constellation capacity of the signal constellation generated by (3.32) is given by

$$I(Y; Z) = I(Y; (X_1, \dots, X_d)), \quad (3.35)$$

and with the chain rule of mutual information we can write

$$I(Y; (X_1, \dots, X_d)) = I_{\text{MSD}_1} + I_{\text{MSD}_2}, \quad (3.36)$$

where

$$I_{\text{MSD}_1} = I(Y; X_1) + I(Y; X_2 | X_1) + \dots + I(Y; X_{d_1} | X_1, \dots, X_{d_1-1}), \quad (3.37)$$

and

$$I_{\text{MSD}_2} = I(Y; X_{d_1+1} | X_1, \dots, X_{d_1}) + I(Y; X_{d_1+2} | X_1, \dots, X_{d_1+1}) \\ + \dots + I(Y; X_{d_1+d_2} | X_1, \dots, X_{d_1+d_2-1}) \quad (3.38)$$

With the proposed decoding scheme I_{MSD_1} is achievable since we use multi-stage decoding for X_1 to X_{d_1} . However, I_{MSD_2} is not achievable, but we are able to achieve I_{PID_2} which is obtained from (3.38) by dropping the conditioning on $X_{d_1+1}, \dots, X_{d_1+d_2}$

$$I_{\text{PID}_2} = I(Y; X_{d_1+1} | X_1, \dots, X_{d_1}) + I(Y; X_{d_1+2} | X_1, \dots, X_{d_1}) \\ + \dots + I(Y; X_{d_1+d_2} | X_1, \dots, X_{d_1}) \leq I_{\text{MSD}_2} \quad (3.39)$$

Since we use Φ_{gray} for $X_{d_1+1}, \dots, X_{d_1+d_2}$ the loss in (3.39) is only small. We refer to the total rate which is achievable given the proposed decoding scheme as $I'(Y; Z)$

$$I'(Y; Z) = I_{\text{MSD}_1} + I_{\text{PID}_2}. \quad (3.40)$$

3.3.4 Numerical Optimization of Constellations for BICM

Once we have chosen d_1 and d_2 in (3.32) the degrees of freedom are the α_i and β . As in [39] we can numerically maximize $I'(Y; Z)$ for a given target SNR. For this note that the transmission power E_s is given by

$$E_s = \sum_{i=1}^d \alpha_i^2 + \beta^2. \quad (3.41)$$

We can formulate the optimization problem as follows

$$\max I'(Y; Z), \text{ where } Z = \sum_{i=1}^{d_1} \alpha_i x_i + \beta \cdot \Phi_{\text{gray}}(x_{d_1+1}, \dots, x_{d_1+d_2}) \\ \text{subject to } E_s = \sum_{i=1}^d \alpha_i^2 + \beta^2. \quad (3.42)$$

We have experimented with several optimization strategies and the use of differential evolution [25] gives good results. Differential evolution is a global optimizer based on a genetic algorithm and hill-climbing. We have solved (3.42) for target rates between 2 bit/use and 5 bit/use for $d_1 = 2$ and $d_1 = 3$.

R SNR [dB]	2.0 bit/use 11.76		3.0 bit/use 17.99	
d_1	2	3	2	3
d_2	2	2	3	3
α_1	0.5601	0.4722	0.5735	0.5105
α_2	0.3646	0.4722	0.3515	0.4024
α_3		0.2802		0.2632
β	0.7439	0.6896	0.7400	0.7129
$ \mathcal{S} $	16	24	32	64
$I(Y;Z)$	1.970	1.986	2.953	2.978
$I'(Y;Z)$	1.970	1.984	2.952	2.976
Δ_{AWGN} [dB]	0.19	0.10	0.29	0.14
Δ_{PAM} [dB]	0.56	0.64	0.81	0.95
R SNR [dB]	4.0 bit/use 24.07		5.0 bit/use 30.11	
d_1	2	3	2	3
d_2	4	4	5	5
α_1	0.5132	0.5118	0.5069	0.4956
α_2	0.3054	0.3950	0.3225	0.3894
α_3		0.2659		0.2555
β	0.8021	0.7151	0.7994	0.7331
$ \mathcal{S} $	64	128	128	256
$I(Y;Z)$	3.945	3.973	4.939	4.969
$I'(Y;Z)$	3.944	3.971	4.938	4.967
Δ_{AWGN} [dB]	0.34	0.18	0.37	0.20
Δ_{PAM} [dB]	0.97	1.12	1.04	1.22

Table 3.1: Parameters of the designed signal constellations.

The latter implies that with the proposed scheme only 3 or 4 multistage decoding stages are required.

The optimization results are shown in Table 3.1. The table gives the target rate R for which we design a signal constellation and the SNR for which the capacity of the AWGN channel is equal to this target rate. The table gives the values of the α_i and β . The constrained constellation capacity $I(Y;Z)$ of the resulting signal constellation is given and the table shows the value of $I'(Y;Z)$ which is achievable with the proposed decoding scheme. The distance in dB to the capacity of the AWGN channel is given by Δ_{AWGN} . Finally, the shaping gain which the signal constellation achieves with respect to a conventional PAM constellation with a size at least as large as the designed constellation is

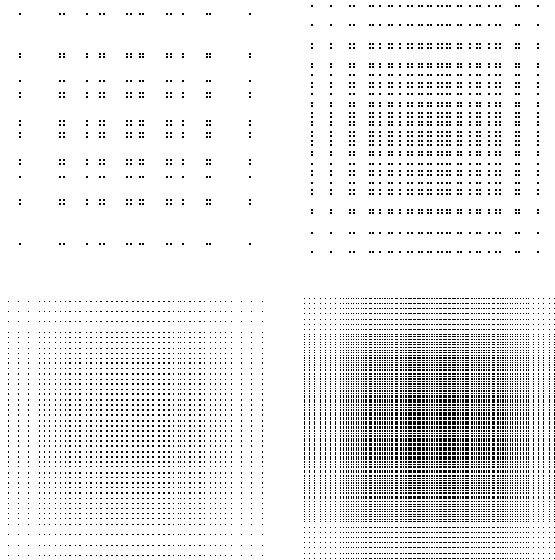


Figure 3.8: The numerically optimized signal constellations for $d_1 = 2$. From left to right and top to bottom are shown the constellations for $d_2 = 2$, $d_2 = 3$, $d_2 = 4$ and $d_2 = 5$.

given by Δ_{PAM} .

The optimization results show that for $d_1 = 2$ and the range of rates we consider, shaping gains from 0.56 dB to 1.04 dB are achieved compared to uniformly spaced PAM constellations. This under the constraint that we use the decoding scheme we proposed in this section. Furthermore, the distance from the capacity of the AWGN channel is less than 0.4 dB for all designs. For $d_1 = 3$ shaping gains are higher and the distance to the capacity of the AWGN channel is less than 0.2 dB.

In the next section we consider the use of binary error-correcting codes for the Φ corresponding to $d_1 = 2$ and we discuss these maps in somewhat more detail. Figure 3.8 gives an impression of the geometrical structure of the signal constellations for $d_1 = 2$. The figure shows the quadrature constellations which are generated by using both dimensions independently. We clearly ob-

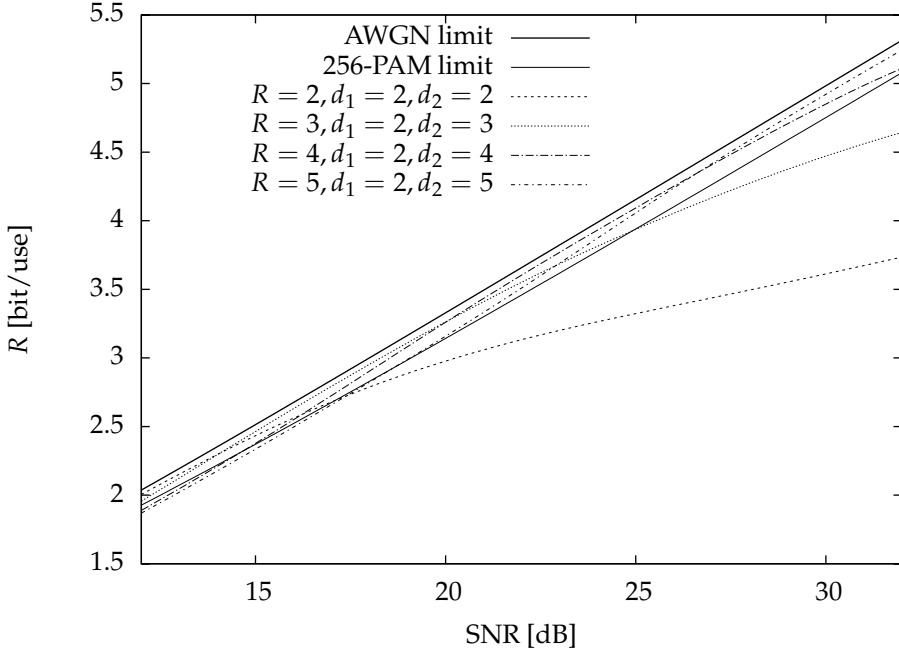


Figure 3.9: Capacity limits of the designed signal constellations with the hybrid decoding strategy.

serve a non-uniform spacing of the constellation symbols, but also the underlying uniformly spaced PAM constellation is visible. Figure 3.9 shows a plot of the achievable rate with the proposed decoding schemes for a larger range of SNRs. From this figure we conclude that the designed Φ are useful in a range of rates around the target rate for which it has been designed.

3.4 Error-Control Coding with Binary Codes

In this section we study the properties of the binary channels resulting from the modulation maps we have defined in the previous section. For the purpose of error-control coding it is convenient when memoryless binary channels are symmetric. For the Φ we consider the equivalent binary channels are

not necessarily symmetric. However, by the use of a technique which is called *channel adapters* in [7], one can make a binary channel symmetric. We derive some general properties of binary channels and densities of log-likelihood ratios and summarize the idea behind channel adapters. Once these properties are established, error-correcting codes can be chosen to get to capacity on the binary channels. We use binary LDPC codes and we briefly review these codes and their analysis and design techniques.

3.4.1 Log-likelihood Ratios and Channel Symmetry

Let $f_{Y|X}$ be a transition density of a memoryless binary channel. This channel is called symmetric if

$$f_{Y|X}(y|1) = f_{Y|X}(-y|-1). \quad (3.43)$$

In this case the performance of a message-passing decoder satisfying some symmetry conditions does not depend on the transmitted codeword and the analysis and design of the code can be restricted to the all-zeros codeword of the LDPC code.

Given a channel output Y , the log-likelihood ratio (LLR) is defined as

$$L = \log \frac{f_{Y|X}(Y|1)}{f_{Y|X}(Y|-1)}. \quad (3.44)$$

In this notation we view L as a random variable which is a function of the random variable Y . The LLR is a sufficient statistic and from an information theoretical point of view this implies that L contains all information about X and no information is lost as the following lemma shows.

Lemma 4 *Consider a memoryless binary channel with input X and output Y . The channel is defined by a transition density $f_{Y|X}$. Next, define a random variable L by (3.44). All information about X is contained in L , i.e.,*

$$I(L; X) = I(Y; X). \quad (3.45)$$

Proof 4 *A proof can be found e.g. in [29] and for convenience we provide a proof here also. First, note that X, Y and L form a Markov chain, i.e.,*

$$X \rightarrow Y \rightarrow L. \quad (3.46)$$

By the chain rule of information we have

$$I((Y, L); X) = I(L; X) + I(Y; X|L) = I(Y; X) + I(L; X|Y). \quad (3.47)$$

Since $X \rightarrow Y \rightarrow L$, it follows that $I(L; X|Y) = 0$ and we have

$$I(L; X) = I(Y; X) - I(Y; X|L). \quad (3.48)$$

Hence it suffices to show that Y and X are conditionally independent given L which would imply that $I(Y; X|L) = 0$. To show that Y and X are conditionally independent given L it is sufficient to show that we can write $f_{Y|X}$ as

$$f_{Y|X}(y|x) = a(x, l)b(y), \quad (3.49)$$

for suitable $a(x, l)$ and $b(y)$ where l denotes a realization of L . By definition of the LLR we have

$$f_{Y|X}(y|1) = f_{Y|X}(y|-1)e^l, \quad (3.50)$$

and we can write

$$f_{Y|X}(y|x) = f_{Y|X}(y|-1) \cdot e^{\frac{x+1}{2}l} = b(y) \cdot a(x, l). \quad (3.51)$$

Hence $I(Y; X|L) = 0$ and the claim follows.

We can view the computation of the LLR as part of the channel and this new channel is defined by a transition density $f_{L|X}$. We always assume that this density exists. Furthermore, note that in general it is not easy to obtain an analytical expression for $f_{L|X}$. However, one can resort to numerical techniques to obtain $f_{L|X}$. Now, we can compute the LLR of this new channel as

$$L' = \log \frac{f_{L|X}(L|1)}{f_{L|X}(L|-1)}. \quad (3.52)$$

Because of the nature of the LLR, we expect that $L' = L$. To show this we require the following lemma.

Lemma 5 *Let a binary memoryless channel be defined by a transition density $f_{Y|X}$. If the density of LLRs $f_{L|X}$ exists it will satisfy (for l not equal to $\pm\infty$)*

$$f_{L|X}(l|1) = e^l f_{L|X}(l|-1). \quad (3.53)$$

Proof 5 *We give a sketch of the proof based on measure theory. First note that $f_{Y|X}$ defines a measure on the measurable space $(\mathbb{R}, \mathcal{F})$ where \mathcal{F} denotes a σ -field of subsets of \mathbb{R} . Let $A \in \mathcal{F}$ and define $P_{Y|X}(A)$ as*

$$P_{Y|X}(A) = \int_A f_{Y|X}(y|x) dy. \quad (3.54)$$

Since the LLR is a random variable its distribution is defined by the measure $\mu_{L|X}$

$$\mu_{L|X}(A) = P_{Y|X}(L^{-1}(A)). \quad (3.55)$$

Next, we show that for an integrable function $h(x)$ we have

$$\int h(x) d\mu_{L|1}(x) = \int h(x) e^x d\mu_{L|-1}(x). \quad (3.56)$$

We write the left-handside of this equation as

$$\begin{aligned} \int h(x) d\mu_{L|1}(x) &= \int h(L(y)) dP_{Y|1}(y) = \int h(L(y)) f_{Y|1}(y) dy \\ &= \int h(L(y)) e^{L(y)} f_{Y|-1}(y) dy = \int h(L(y)) e^{L(y)} dP_{Y|-1}(y) \\ &= \int h(x) e^x d\mu_{L|-1}(x). \end{aligned} \quad (3.57)$$

Since we assume that $f_{L|X}$ exists it follows that

$$f_{L|X}(l|1) = e^l f_{L|X}(l|-1). \quad (3.58)$$

This lemma shows that a memoryless binary channel is fully defined by specifying $f_{L|X}(l|1)$ since $f_{L|X}(l|-1)$ can be obtained from $f_{L|X}(l|1)$. The next theorem is a generalization of the channel equivalence lemma as defined in [29].

Theorem 6 Let a binary memoryless channel defined by a transition density $f_{Y|X}$ and denote the LLR and corresponding density of LLRs by L and $f_{L|X}$, respectively. Next, consider a binary memoryless channel with transition density $f_{L|X}$ and denote its LLR by L' . The following holds

$$L' = L, \quad (3.59)$$

from which it follows that

$$f_{L'|X}(l|x) = f_{L|X}(l|x). \quad (3.60)$$

Proof 6 By definition of the LLR we can write

$$L' = \log \frac{f_{L|X}(L|1)}{f_{L|X}(L|-1)}, \quad (3.61)$$

and by Lemma 5 we can write

$$L' = \log \frac{f_{L|X}(L|1)}{e^{-L} f_{L|X}(L|1)} = \log e^L = L, \quad (3.62)$$

and it follows that $f_{L|X}(l|x) = f_{L'|X}(l|x)$.

This theorem shows that when we consider the set of all $f_{Y|X}$ defining a binary memoryless channel, we can call two channels equivalent if they have the same density of LLRs. This relation is an equivalence relation and Theorem 6 proves reflexivity of the relation. From a practical point of view the relevance of Theorem 6 is that it shows that when we consider the computation of LLRs as part of the channel, we do not have to recompute LLRs and can just take the channel output as the LLR.

For a binary memoryless symmetric channel we have the following lemma which can be found in [29].

Lemma 7 *Let a binary memoryless channel be defined by a transition density $f_{Y|X}$ which satisfies*

$$f_{Y|X}(y|1) = f_{Y|X}(-y|-1). \quad (3.63)$$

In this case the density of LLRs satisfies

$$f_{L|X}(l|1) = e^l f_{L|X}(-l|1), \quad (3.64)$$

Proof 7 *First, note that for a symmetric channel the following holds*

$$f_{L|X}(l|1) = f_{L|X}(-l|-1). \quad (3.65)$$

To see this note that we can write the LLR as

$$L = \log \frac{f_{Y|X}(Y|1)}{f_{Y|X}(Y|-1)} = \log \frac{f_{Y|X}(-Y|-1)}{f_{Y|X}(-Y|1)} = -\log \frac{f_{Y|X}(-Y|1)}{f_{Y|X}(-Y|-1)}. \quad (3.66)$$

This shows that the density of L as a function of the random variable Y is equal to the density of $-L$ as a function of $-Y$. Now, $f_{L|X}(l|1)$ defines the distribution of L when $X = 1$ and the density of Y is by definition $f_{Y|X}(Y|1)$. Since the distribution of Y given $X = -1$ is defined by $f_{Y|X}(Y|-1) = f_{Y|X}(-Y|1)$, we conclude that $f_{L|X}(l|1) = f_{L|X}(-l|-1)$. Now, the claim follows from Lemma 5.

A density of LLRs satisfying (3.64) is called symmetric.

In case the original binary channel is not symmetric we can construct an equivalent binary channel which is symmetric by using a common source of randomness at the transmitter and receiver.

Theorem 8 *Let a binary memoryless channel be defined by a $f_{Y|X}$ and denote the density of LLRs by $f_{L|X}$. Next, let B be a uniform distributed binary random variable taking values in $\{-1, 1\}$ of which the realization is known at the transmitter and the receiver. Instead of transmitting X we transmit BX and at the receiver we take the value of B into account and compute an LLR L' . We can consider L' as the output of an equivalent binary symmetric channel defined by an $f_{L'|X}$ which satisfies*

$$f_{L'|X}(l|1) = f_{L'|X}(-l - 1), \quad (3.67)$$

and

$$f_{L'|X}(l|1) = e^l f_{L'|X}(l - 1). \quad (3.68)$$

Proof 8 *In case we transmit BX instead of X the LLR is given by*

$$L' = \log \left(\frac{f_{Y|X}(y|1)}{f_{Y|X}(y|-1)} \right)^B = B \log \frac{f_{Y|X}(y|1)}{f_{Y|X}(y|-1)}. \quad (3.69)$$

Let $f_{L'|X, B=b}$ denote the density of L' given X and $B = b$. We can write $f_{L'|X}$ as

$$f_{L'|X}(l|x) = \frac{1}{2} f_{L'|X, B=1}(l|x) + \frac{1}{2} f_{L'|X, B=-1}(l|x). \quad (3.70)$$

Now, note that when $B = 1$, we just transmit X on the channel and

$$f_{L'|X, B=1}(l|x) = f_{L|X}(l|x). \quad (3.71)$$

In case $B = -1$ the value of X is flipped before transmission. Moreover, the value of the LLR is negated at the receiver. Hence we have

$$f_{L'|X, B=-1}(l|x) = f_{L|X}(-l - x), \quad (3.72)$$

from which it follows that

$$f_{L'|X}(l|x) = \frac{1}{2} f_{L|X}(l|x) + \frac{1}{2} f_{L|X}(-l - x). \quad (3.73)$$

From this it is clear that $f_{L'|X}(l|1) = f_{L|X}(-l - 1)$ and $f_{L'|X}(l|1) = e^l f_{L|X}(l - 1)$ follows from the fact that $f_{L|X}(l|1) = e^l f_{L|X}(l - 1)$ by Lemma 5.

This idea is introduced in [7] under the name of *channel adapters* and it is related to ensembles of binary coset codes.

3.4.2 Equivalent Binary Channels for Modulation Maps

Before we consider error-control coding, we derive the properties of the equivalent binary channels resulting from Φ_{hybrid} and the hybrid decoding scheme for transmission on the AWGN channel. For this purpose we first derive the properties of the binary channel resulting from Φ_{spm} with multistage decoding. Next, we consider the equivalent binary channels resulting from Φ_{gray} with parallel and independent decoding. Finally, we combine these results and consider Φ_{hybrid} .

Equivalent Channels for Φ_{spm}

Let X_1, \dots, X_d denote a tuple of uniform i.i.d. bits and let a channel input Z for the AWGN channel be generated by Φ_{spm} . The channel output Y is given by

$$Y = Z + N = \Phi_{\text{spm}}(X_1, \dots, X_d) + N = \sum_{i=1}^d \alpha_i X_i + N, \quad (3.74)$$

where N is Gaussian noise with variance σ^2 . Without loss of generality we assume that the decoding sequence for multistage decoding is X_1, X_2, \dots, X_d . While decoding X_l at level l the values of X_1, \dots, X_{l-1} are assumed to be known. We can write the channel for X_l as

$$Y = \alpha_l X_l + c'_l + N'_l, \quad (3.75)$$

where c'_l summarizes the values of the known bits X_1, \dots, X_{l-1} and is defined as

$$c'_l = \sum_{i=1}^{l-1} \alpha_i X_i. \quad (3.76)$$

The channel for X_l is disturbed by additive noise N'_l

$$N'_l = \sum_{i=l+1}^d \alpha_i X_i, \quad (3.77)$$

and the density of N'_l is given by

$$f_{N'_l}(n) = \frac{1}{2^{d-l}} \sum_{x_{l+1}} \dots \sum_{x_d} f_N(n - \alpha_{l+1} x_{l+1} - \dots - \alpha_d x_d). \quad (3.78)$$

Now, we can cancel the contribution of c'_l from the channel output which gives us a new channel output Y'

$$Y' = Y - c'_l = \alpha_l X_l + N'_l, \quad (3.79)$$

and this equivalent channel is defined by a transition density $f_{Y'|X_l}$

$$f_{Y'|X_l}(y|x_l) = f_{N'_l}(y - \alpha_l x_l). \quad (3.80)$$

We compute an LLR L_l for X_l based on Y'

$$L_l = \log \frac{f_{Y'|X_l}(Y'|1)}{f_{Y'|X_l}(Y'|-1)}. \quad (3.81)$$

Since $f_{N'_l}(n) = f_{N'_l}(-n)$ the transition density $f_{Y'|X_l}(y|x_l)$ defines a symmetric channel and the density of LLRs $f_{L_l|X_l}$ defining the channel for X_l is symmetric.

Equivalent Channels for Φ_{gray}

Let a channel input for the AWGN channel be generated by Φ_{gray}

$$Y = Z + N = \Phi_{\text{gray}}(X_1, \dots, X_d) + N. \quad (3.82)$$

When we use parallel and independent decoding the channel for a bit X_l is defined by the following transition density

$$f_{Y|X_l}(y|x_l) = \frac{1}{2^{d-1}} \sum_{x_1} \cdots \sum_{x_{l-1}} \sum_{x_{l+1}} \cdots \sum_{x_d} f_N(y - \Phi_{\text{gray}}(x_1, \dots, x_d)). \quad (3.83)$$

The use of Φ_{gray} provides a set of d equivalent binary channels and these channels are not necessarily symmetric. Nevertheless an LLR for X_l is computed as

$$L_l = \log \frac{f_{Y|X_l}(Y|1)}{f_{Y|X_l}(Y|-1)}, \quad (3.84)$$

and let $f_{L_l|X_l}$ define the density of LLRs for X_l . We apply Theorem 8 to make the channel for each of the bits symmetric. For this purpose let B_1, \dots, B_d denote a set of uniform i.i.d. random bit variables of which the realization is known at the transmitter and the receiver. A channel input is generated as

$$Z = \Phi_{\text{gray}}(B_1 X_1, \dots, B_d X_d). \quad (3.85)$$

Now, an LLR L'_l is computed for X_l as

$$L'_l = B_l \log \frac{f_{Y|X_l}(Y|1)}{f_{Y|X_l}(Y|-1)}, \quad (3.86)$$

and the density of L'_l which we denote by $f_{L'_l|X_l}$ defines a symmetric channel.

In a BICM setting the assignment of the bits as the input of the modulation map is randomized as well. This renders the channel for each of the bits the same and the channel for a bit X_l is defined by a density of LLRs $f_{L'_l|X_l}$

$$f_{L'_l|X_l}(l|x_l) = \frac{1}{d} \sum_{i=1}^d f_{L'_i|X_i}(l|x_l), \quad (3.87)$$

which in turn defines a symmetric channel.

Equivalent Binary Channels for Φ_{hybrid}

Consider the generation of channel inputs for the AWGN channel by Φ_{hybrid}

$$\begin{aligned} Y &= \Phi_{\text{hybrid}}(X_1, \dots, X_d) + N \\ &= \sum_{i=1}^{d_1} \alpha_i X_i + \beta \cdot \frac{1}{\sqrt{E_{\text{PAM}, d_2}}} \cdot \sum_{i=d_1+1}^{d_1+d_2} \left(\prod_{j=i}^{d_1+d_2} X_j \right) 2^{i-1} + N. \end{aligned} \quad (3.88)$$

To derive the properties of the equivalent binary channel for X_1, \dots, X_{d_1} note that for these bits $X_{d_1+1}, \dots, X_{d_1+d_2}$ generate a uniformly spaced PAM constellation which acts as noise. However, as noise the mapping from bits to constellation symbols is irrelevant. Hence, for a bit X_l with $l = 1, \dots, d_1$ the channel takes the form

$$Y = \alpha_l X_l + c'_l + N'_l, \quad (3.89)$$

where c'_l summarizes the values of the decoded bits X_1, \dots, X_{l-1} . Furthermore, the channel for X_l is disturbed by additive noise N'_l

$$N'_l = \sum_{i=l+1}^{d_1} \alpha_i X_i + \beta \cdot \frac{1}{\sqrt{E_{\text{PAM}, d_2}}} \cdot \sum_{i=d_1+1}^{d_1+d_2} 2^{i-1}. \quad (3.90)$$

From which we conclude that the equivalent binary channel is defined by a transition density given by (3.80) and LLRs are computed by (3.81). Furthermore, the equivalent binary channels for X_1, \dots, X_{d_1} are symmetric.

Once X_1, \dots, X_{d_1} are decoded they can be cancelled from the channel output and we are in the same setting as if channel inputs had been generated by a scaled version of Φ_{gray} .

3.4.3 Binary LDPC Codes

We do not give an extensive overview of LDPC codes and their design. We refer to [29] for more details. Furthermore, in [39] we discuss the design of LDPC codes for modulation maps of the form Φ_{spm} in more detail.

An LDPC code is a linear block code which is defined by a low-density parity-check matrix $\mathbf{H} \in \text{GF}(2)^{(n-k) \times n}$, where n is the length of a codeword and k the number of source bits which is mapped to a codeword. Each codeword \mathbf{x} satisfies $\mathbf{H}\mathbf{x} = \mathbf{0}$. An LDPC code is conveniently represented by a Tanner graph [30]. The Tanner graph is a bipartite graph consisting of variable nodes and check nodes. Each variable nodes represents a bit of a codeword and each check node a row of the parity-check matrix. An ensemble of LDPC codes can be defined by a degree distribution pair (λ, ρ) where $\lambda(x)$ and $\rho(x)$ define the distribution of variable nodes and check nodes, respectively

$$\lambda(x) = \sum_i \lambda_i x^{i-1} \quad \rho(x) = \sum_i \rho_i x^{i-1}, \quad (3.91)$$

where λ_i denotes the fraction of edges connected to variable nodes of degree i and ρ_i denotes the fraction of edges connected to check nodes of degree i . In terms of the degree distribution pair (λ, ρ) the design rate r of the code is given by

$$r = 1 - \frac{\int_0^1 \rho(x) dx}{\int_0^1 \lambda(x) dx}. \quad (3.92)$$

The analysis and design of LDPC codes is greatly simplified for symmetric channels. In case the channel is a symmetric channel and the decoder satisfies certain symmetry conditions, the performance of the decoder is independent of the transmitted codeword. We can restrict the analysis of the decoding process to the case where the all-zeros codeword is transmitted. In the previous section we have shown that for the modulation maps we consider, the equivalent binary channels are either symmetric or we can make them symmetric.

SNR [dB]	11.76	17.99	24.07	30.10
AWGN capacity	2.0	3.0	4.0	5.0
d	3	4	5	6
$I(Z;Y)$	1.879	2.819	3.784	4.765
I_{PID}	1.853	2.792	3.755	4.736

Table 3.2: Parameters of the PAM constellations for the PAM-LDPC codes.

3.5 Design Examples and Numerical Results

In this section we provide several numerical examples. First, we design LDPC codes for the equivalent binary channel resulting from Φ_{gray} for several constellation sizes and target rates. We refer to these codes as M -PAM-LDPC codes and we show that these codes can perform close to the constrained constellation capacity limit. Second, we design LDPC codes for the equivalent channels resulting from several Φ_{hybrid} and the proposed multistage decoding scheme. We refer to these codes as shaped M -PAM-LDPC codes. With these codes the gap to the capacity of the AWGN channel due to shaping can be bridged.

3.5.1 PAM-LDPC Codes

We consider transmission on the AWGN channel at rates in the range of 2 bit/use to 5 bit/use and the parameters of the PAM constellation generated by Φ_{gray} we use are given in Table 3.2. The table gives the SNRs at which a target rate of 2, 3, 4 and 5 bit/use is achieved. Furthermore, the table specifies the value of d used to generate the signal constellation and the constrained constellation capacity. Finally, the table also gives I_{PID} and we observe that the loss with respect to the constrained constellation capacity is only small.

We have designed LDPC codes for the equivalent binary channels and the parameters of the codes are given in Table 3.3. The table defines the degree distributions and the design rate r of the LDPC codes. The threshold is given by SNR_* and it is close to the target SNR for which we have designed the codes. The overall transmission rate $R = rd$ is given also and all codes achieve a low asymptotic BER while the rate is higher than 99.7% of the achievable rate I_{PID} .

We perform a Monte Carlo simulation of the codes where we choose a block length of $N = 20000$ and $N = 100000$ channel input symbols, respectively. The length n of the LDPC codeword is given by $n = Nd$. The simulation results

Code	$d = 3$	$d = 4$	$d = 5$	$d = 6$
λ_2	0.154145	0.145337	0.1407388	0.137750
λ_3	0.156311	0.151913	0.1492723	0.147196
λ_6	0.003218	0.004692	0.0065441	0.010026
λ_7	0.200050	0.194657	0.1909628	0.185561
λ_{20}		0.0627174	0.0837469	0.110721
λ_{21}	0.180790	0.136561	0.1167630	0.090856
λ_{22}	0.022166			
λ_{100}	0.283320	0.304124	0.3119721	0.317890
ρ_{15}	0.748250			
ρ_{16}	0.251750			
ρ_{19}		0.006943		
ρ_{20}		0.993057		
ρ_{24}			0.25675	
ρ_{25}			0.74325	
ρ_{29}				0.37185
ρ_{30}				0.62815
r	0.616	0.696	0.749	0.787
SNR* [dB]	11.77	18.00	24.07	30.11
$R = dr$	1.848	2.784	3.745	4.722

Table 3.3: Degree distributions for the M -PAM-LDPC codes.

of the codes are shown in Figure 3.10 and Figure 3.11. The figures show the bit-error rate performance of the codes for the two block lengths. Furthermore, the figures indicate the value of the SNR where the AWGN capacity is equal to the transmission rate. The SNR limit required to achieve the transmission rate with BICM (I_{PID}) is shown also.

All codes perform reasonably close to the I_{PID} limit. A $\text{BER} < 10^{-5}$ is achieved within 0.7 dB of I_{PID} for $N = 20000$ and within 0.4 dB for $N = 100000$. However, the gap to the capacity of the AWGN channel is substantial. To bridge this gap, we require shaping of the signal constellation.

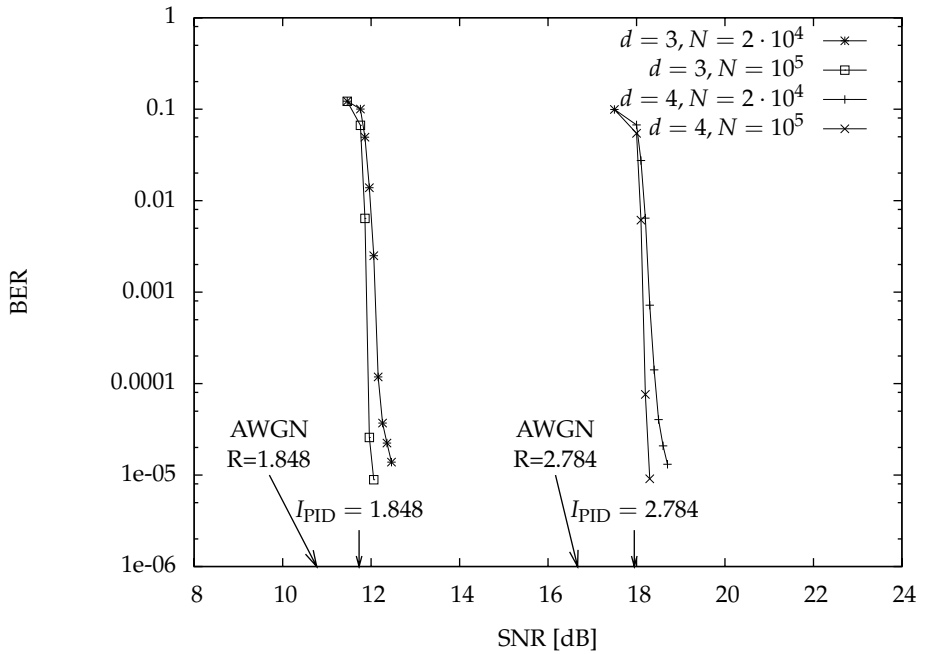


Figure 3.10: Simulation results for M-PAM-LDPC codes on the AWGN channel for target rates 2 bit/use and 3 bit/use.

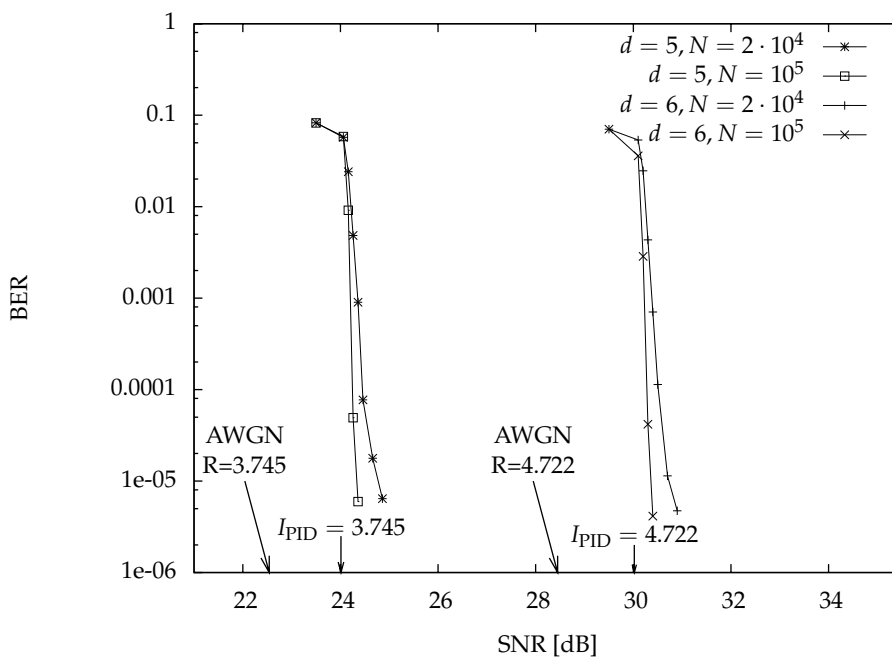


Figure 3.11: Simulation results for M -PAM-LDPC codes on the AWGN channel for target rates 4 bit/use and 5 bit/use.

Code	$d_1 = 2, d_2 = 2$			$d_1 = 2, d_2 = 3$		
	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}
λ_2	0.235819	0.264719	0.131359	0.228782	0.259637	0.128759
λ_3	0.189501	0.196177	0.144553	0.181581	0.186984	0.141871
λ_5			0.038020			
λ_6			0.031272			0.0711654
λ_7	0.157022	0.164893		0.176631	0.199624	0.1014825
λ_8	0.045821	0.033304	0.157238	0.025225		
λ_9						0.0382066
λ_{13}						0.0132644
λ_{14}			0.000707			
λ_{19}		0.017823				
λ_{20}	0.044399	0.122066		0.066201	0.123820	
λ_{21}	0.109676			0.094364	0.029331	0.0838960
λ_{22}			0.082780			
λ_{23}			0.102322			0.1084474
λ_{24}			0.018804			0.0098793
λ_{100}	0.217762	0.201018	0.292947	0.227215	0.200604	0.3030290
ρ_5		0.696302			0.366039	
ρ_6	0.771165	0.303698		0.214057	0.633961	
ρ_7	0.228835			0.785943		
ρ_{24}			0.558954			
ρ_{25}			0.441046			
ρ_{31}						0.37587
ρ_{32}						0.62413
r	0.263	0.190	0.741	0.307	0.223	0.797
SNR* [dB]	10.79	11.02	11.77	16.20	16.65	18.01
R	1.935			2.921		

Table 3.4: Parameters of the shaped M -PAM-LDPC codes for $d_2 = 2$ and $d_2 = 3$.

3.5.2 Shaped PAM-LDPC Codes

We design LDPC codes for the numerically optimized Φ_{hybrid} of Section 3.3 corresponding to $d_1 = 2$. The parameters of the codes are given in Table 3.4 and Table 3.5.

For each of the target rates we have first designed \mathcal{C} which is the code used in the BICM setting. For a target rate of 2 bit/use and 3 bit/use we choose a channel input block length of $2 \cdot 10^5$. For 4 bit/use and 5 bit/use we choose

Code	$d_1 = 2, d_2 = 4$			$d_1 = 2, d_2 = 5$		
	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}	\mathcal{C}_1	\mathcal{C}_2	\mathcal{C}
λ_2	0.251874	0.278142	0.118440	0.252996	0.269500	0.119845
λ_3	0.188163	0.187461	0.148027	0.186016	0.182617	0.142770
λ_6		0.025427			0.033295	0.022921
λ_7	0.171387	0.166436	0.199406	0.186855	0.163103	0.180233
λ_8	0.033614			0.015486		
λ_{18}		0.080523	0.084243			
λ_{19}		0.055880			0.145201	0.066506
λ_{20}				0.030448		
λ_{21}	0.109757			0.126781		0.098216
λ_{22}	0.047746					
λ_{23}			0.129713			
λ_{25}						0.059848
λ_{100}	0.197459	0.206131	0.320171	0.201418	0.206284	0.309660
ρ_5	0.067603	0.82170		0.074365	0.58550	
ρ_6	0.932397	0.17830		0.925635	0.41450	
ρ_{48}			0.67369			
ρ_{49}			0.32631			
ρ_{57}						0.034591
ρ_{58}						0.965409
r	0.255	0.188	0.863	0.254	0.205	0.886
SNR* [dB]	17.78	20.98	24.07	20.68	23.06	30.10
R	3.895			4.889		

Table 3.5: Parameters of the shaped M -PAM-LDPC codes $d_2 = 4$ and $d_2 = 5$.

a channel input block length of 10^5 . The code rates for \mathcal{C}_1 and \mathcal{C}_2 are chosen slightly lower than the capacity of the equivalent binary channel such that all codes achieve a low BER at approximately the same SNR.

The simulation results are shown in Figure 3.12 and Figure 3.13. All schemes achieve a low BER within 0.5 dB of $I'(Y; Z)$. Furthermore, the distance to the capacity of the AWGN channel is within 0.7 dB for all schemes. Moreover, the schemes all beat the uniform signaling limit for a signal constellation with a uniform spacing and uniform distribution on the constellation symbols. The schemes achieve a low BER at around the same SNR where the PAM-LDPC codes achieve a low BER. However, the transmission rate of the shaped PAM-LDPC codes is higher.

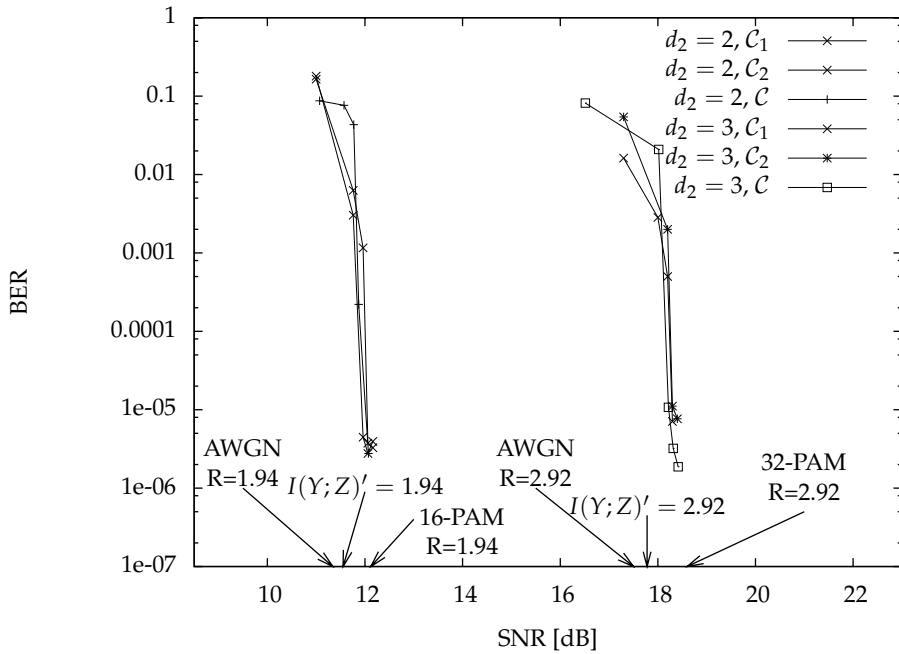


Figure 3.12: Simulation results for shaped M -PAM-LDPC codes on the AWGN channel for target rates 2 bit/use and 3 bit/use.

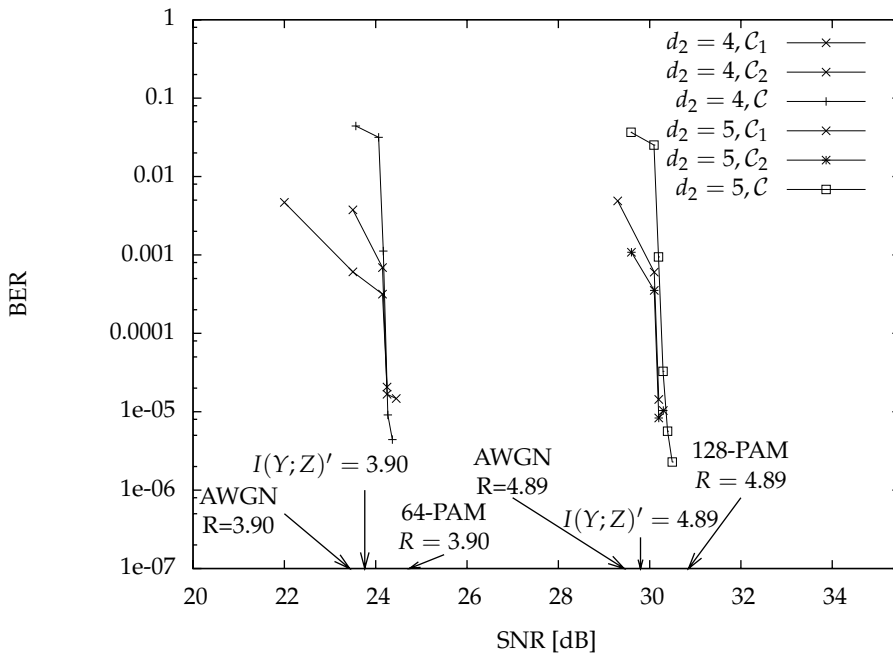


Figure 3.13: Simulation results for shaped M-PAM-LDPC codes on the AWGN channel for target rates 4 bit/use and 5 bit/use.

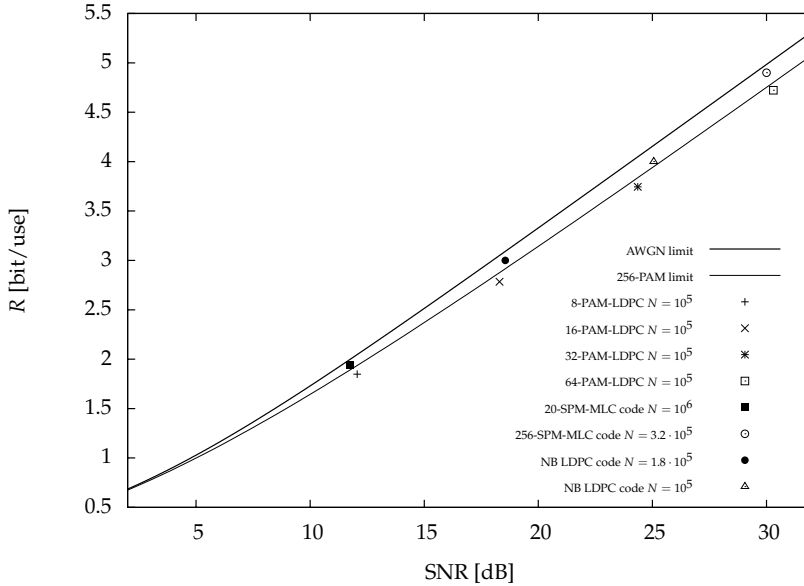


Figure 3.14: Summary of the power and bandwidth efficiency of the coding schemes for $N = 10^5$ and $\text{BER} = 10^{-5}$.

3.6 Conclusions and Final Remarks

A summary of the results is given in Figure 3.14 which shows the power-rate plane of the AWGN channel. The figure shows the capacity of the AWGN channel together with the constrained constellation capacity of a 256-PAM constellation. Furthermore, the rate achieved by each of the schemes presented in this chapter are shown also. We have plotted the SNR where each scheme achieves a $\text{BER} \leq 10^{-5}$. Each of the M -PAM-LDPC codes operates close to the capacity of the 256-PAM limit. Furthermore, the shaped M -PAM-LDPC all beat this limit and operate close to the capacity of the AWGN channel.

In this chapter we considered coded modulation on the AWGN channel and introduced a scheme which is able to perform close to the capacity of the AWGN channel. The scheme combined with binary LDPC codes is able to approach the capacity of the AWGN channel for several spectral efficiencies very closely.

Chapter 4

Symmetric Channels and Coded Modulation

4.1 Introduction

In this chapter we characterize channel symmetry for a broad class of memoryless channels. This characterization unifies several notions of channel symmetry. Furthermore, we give an application of the developed theory to coded modulation. We show how to devise coded modulation schemes for channels with additive noise. This in such a way that the original coding problem is transformed into a coding problem for a set of symmetric channels.

Channel symmetry is e.g. defined in [28] for discrete memoryless channels (DMCs). For symmetric DMCs mutual information is maximized for a uniform distribution on the channel input alphabet. A related concept of channel symmetry for DMCs is a regular channel which is defined in [42]. A regular channel is an algebraic characterization of the symmetry of DMCs and it can be shown that a regular channel is symmetric in the Gallager sense [43]. A major contribution of [42] is that it shows that linear codes over finite fields can be used to achieve capacity on regular channels. This generalizes the fact that binary linear codes can be used to achieve capacity on binary memoryless symmetric channels.

Binary sparse graph codes such as low-density parity-check (LDPC) codes are capable of achieving a near-capacity performance on several binary-input

output-symmetric (BIOS) channels [2], [3], [26]. The analysis of message passing decoders satisfying certain symmetry conditions can be restricted to the all-zeros codeword for symmetric channels. This greatly simplifies the analysis and design of LDPC codes.

For channels with a capacity significantly larger than one the use of binary inputs can lead to a loss in rate, because the channel capacity might allow transmission at higher rates. To achieve higher rates for power and bandwidth efficient communication, one can opt for non-binary codes as in [12]. An alternative is to use codes for which the codeword symbols are taken from a relatively small alphabet and map groups of codeword symbols to channel inputs. Several coding techniques apply this approach and we refer to these techniques as coded modulation.

A well-known coded modulation strategy is multi-level coding (MLC) with multi-stage decoding (MSD) or parallel-and-independent decoding (PID) [8], [23]. With MLC-MSD a set of codes is used at the transmitter and at the receiver successive decoding of the codes is used. One can approach the capacity of the channel with maximum-likelihood decoding of the individual codes. The use of MLC-PID is suboptimal in the sense that the achievable information rate is less than the capacity of the channel. However, the loss might only be small and the codes can be decoded in parallel. Moreover, one can use a single code at the transmitter which naturally leads to bit-interleaved coded-modulation (BICM) [9] when binary codes are used.

From a coding point of view MLC and BICM reduce the problem of achieving capacity on a channel with a large input alphabet to achieving capacity on a set of equivalent channels with a smaller input alphabet. However, these equivalent channels are not necessarily symmetric.

In this chapter we characterize channel symmetry for memoryless channels with a discrete input alphabet \mathcal{X} and a general output space \mathcal{Y} . The channels we consider are defined by a set of functions $\{f_i\}$ for $i = 1, \dots, |\mathcal{X}|$

$$f_i : \mathcal{Y} \mapsto \mathbb{R}. \quad (4.1)$$

For DMCs we can identify these functions with the conditional transition probabilities of the channel and for channels with outputs in \mathbb{R} these functions are the conditional transition density functions of the channel. Our starting point to define channel symmetry is information theoretic. We establish a set of conditions such that mutual information is maximized for a uniform distribution and call such a channel symmetric. We will see that we can express these conditions in terms of the existence of algebraic groups acting on the set $\{f_i\}$.

The motivation of the work we present is its practical application. We show how to construct coded modulation schemes for high spectral efficiencies leading to symmetric equivalent channels. The outline of this chapter is as follows. In Section 4.2 we introduce preliminary concepts and the notation used throughout the chapter. In Section 4.3, we revisit the concept of channel symmetry and characterize it for general channels. We give an application to coded modulation in Section 4.4 and end with conclusions in Section 4.5.

4.2 Preliminaries

In this section we review some basic concepts from information theory and give an overview of several notions of channel symmetry. Furthermore, we review some concepts from algebra and geometry which are used in later sections.

4.2.1 Information Theory

Channel Symmetry for DMCs

Consider a channel where the channel input X takes values in the set \mathcal{X} and the channel output Y takes values in the set \mathcal{Y} . A DMC is defined as a memoryless channel where both \mathcal{X} and \mathcal{Y} are discrete sets. In this case the channel is specified by a transition probability assignment $f_{Y|X}(y|x)$. By definition $f_{Y|X}(y|x)$ is the probability of receiving y when x is transmitted.

For a DMC the average mutual information between the channel input and channel output is defined as

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} f_{Y|X}(y|x) f_X(x) \log_2 \frac{f_{Y|X}(y|x)}{f_Y(y)}, \quad (4.2)$$

where $f_X(x)$ denotes the probability that the channel input is chosen as x . Furthermore, $f_Y(y)$ is given by

$$f_Y(y) = \sum_{x' \in \mathcal{X}} f_{Y|X}(y|x') f_X(x'). \quad (4.3)$$

The capacity of the channel is defined as the maximum mutual information between X and Y where the maximization is over all input distributions.

$$C = \max_{f_X} I(X;Y) \quad (4.4)$$

A logical way to define channel symmetry is to relate it to the properties of the f_X achieving the capacity of the channel. One property we wish symmetric channels to share is that capacity is achieved when f_X is the uniform distribution. In Gallager [28] the path to define channel symmetry is as follows. First, conditions for a f_X to maximize $I(X; Y)$ can be derived as in the following theorem.

Theorem 9 (Theorem 4.5.1 in [28]) *A set of necessary and sufficient conditions on a probability assignment $f_X(x)$ to achieve capacity of the DMC with transition probabilities $f_{Y|X}$ is that for some number C ,*

$$\begin{aligned} I(X = k; Y) &= C; & \text{all } x \text{ with } f_X(x) > 0 \\ I(X = k; Y) &\leq C; & \text{all } x \text{ with } f_X(x) = 0 \end{aligned} \quad (4.5)$$

Proof 9 *The proof follows from the convexity of mutual information with respect to the input distribution and can be found in [28].*

This theorem holds also for channels with a discrete input alphabet and outputs from \mathbb{R}^n or \mathbb{C}^n . Next, a symmetric DMC is defined as follows.

Definition 1 ([28], pp. 94) *A DMC is defined to be symmetric if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities (using inputs as rows and outputs of the subset as columns) has the property that each row is a permutation of each other row and each column (if more than 1) is a permutation of each other column.*

Now, we can show that if a DMC is symmetric a uniform $f_X(x)$ achieves capacity.

Theorem 10 (Theorem 4.5.2 in [28]) *For a symmetric DMC, capacity is achieved by using the inputs with equal probability.*

Proof 10 *The proof can be found in [28].*

Channel Symmetry for channels with outputs in \mathbb{R}^n or \mathbb{C}^n

In case $\mathcal{Y} = \mathbb{R}^n$ or $\mathcal{Y} = \mathbb{C}^n$, we assume that the channel is defined by a transition probability density function $f_{Y|X}(y|x)$. In this case $f_{Y|X}(y|x)$ is the density of the channel output y conditioned on the channel input x . For channels with outputs in \mathbb{R}^n the mutual information between X and Y is defined as

$$I(X; Y) = \sum_{x \in \mathcal{X}} f_X(x) \int_{\mathbb{R}^n} f_{Y|X}(y|x) \log_2 \frac{f_{Y|X}(y|x)}{f_Y(y)} dy. \quad (4.6)$$

A channel with binary inputs and outputs in \mathbb{R} is called symmetric if the channel transition probability density function satisfies [41]

$$f_{Y|X}(y|0) = f_{Y|X}(-y|1). \quad (4.7)$$

For a channel defined by a $f_{Y|X}(y|x)$ satisfying this relation (4.5) holds and capacity is achieved for a uniform distribution on the channel input. BIOS channels play a crucial role in the analysis and design of iterative decoding methods for binary sparse graph codes [26]. This has been one of the main motivations for the work presented in this chapter and some implications are given in Section 4.4 when we discuss applications.

Channel and Mutual Information in an Abstract Setting

For the purpose of defining channel symmetry irrespective of the channel output space \mathcal{Y} , we can cast mutual information in an abstract measure theoretical setting. We will derive most of our results without resorting to measure theory. However, we borrow some notational aspects which are common in measure theory.

We consider channels with inputs from a finite discrete set $\{0, 1, \dots, N-1\}$ and outputs from an output space \mathcal{Y} . The output space will either be a finite discrete set, \mathbb{R} or \mathbb{C} . As before let f_X define the distribution on the channel inputs and let the channel defined by a set of functions

$$f_i : \mathcal{Y} \rightarrow \mathbb{R}, \quad (4.8)$$

for $i = 1$ to N and f_i satisfies

$$\int_{\mathcal{Y}} f_i(y) dy = 1. \quad (4.9)$$

In case the channel output space is a finite discrete set the f_i define the channel transition probabilities and we interpret the integral as a sum. In case the channel output space is \mathbb{R} or \mathbb{C} the f_i are the channel transition densities. We write mutual information for the output spaces we consider as

$$I(X; Y) = \sum_{i \in \mathcal{X}} P_X(x) \int_{\mathcal{Y}} f_i(y) \log_2 \frac{f_i(y)}{\sum_{j \in \mathcal{X}} P_X(j) f_j(y)} dy, \quad (4.10)$$

and again, depending on the context we interpret the integral as a sum.

4.2.2 Geometry and Algebra

In this section we review some concepts from geometry and algebra which we use to define channel symmetry. These concepts can be found in textbooks on algebra and geometry [44], [45].

Metric space and isometry

Let $\mathcal{F}(\mathcal{Y})$ denote the set of all bijective functions from \mathcal{Y} to \mathcal{Y} . We equip the channel output space \mathcal{Y} with a metric d . The resulting *metric space* is denoted by (\mathcal{Y}, d) . In the case of a DMC, \mathcal{Y} is a discrete set and a natural metric to use is the discrete metric $\delta(i, j)$ which is defined as

$$\delta(i, j) = \begin{cases} 0 & i \neq j \\ 1 & i = j. \end{cases} \quad (4.11)$$

In later sections we consider the case that

$$\mathcal{Y} = \mathbb{R}^n,$$

and the metric we consider is the Euclidian metric

$$d_e(\mathbf{v}, \mathbf{w}) = \sqrt{(\mathbf{v} - \mathbf{w})^T (\mathbf{v} - \mathbf{w})} \quad \mathbf{v}, \mathbf{w} \in \mathbb{R}^n. \quad (4.12)$$

We are interested in maps from \mathcal{Y} to \mathcal{Y} which preserve distance.

Definition 2 An isometry ϕ of a metric space (\mathcal{Y}, d) is a bijective map

$$\phi : \mathcal{Y} \rightarrow \mathcal{Y}, \quad (4.13)$$

which preserves distance

$$d(\phi(v), \phi(w)) = d(v, w) \quad \forall v, w \in \mathcal{Y}. \quad (4.14)$$

We denote the set of all isometries of (\mathcal{Y}, d) by $\mathcal{I}(\mathcal{Y})$. Let $A \subseteq \mathcal{Y}$ and denote the image of A under the isometry ϕ by

$$\phi A = \{y \in \mathcal{Y} : y = \phi(x), x \in A\}. \quad (4.15)$$

In case $\phi A = A$ we call ϕ an *symmetry* of A . Next, we denote the set of all symmetries of A by \mathcal{S}_A

$$\mathcal{S}_A = \{\phi \in \mathcal{I}(\mathcal{Y}) : \phi A = A\}. \quad (4.16)$$

Groups and Group Actions

The set \mathcal{S}_A with composition as an operation constitutes a group. To see this note that the identity map is a member of \mathcal{S}_A . Furthermore, if $\phi A = A$ then $\phi^{-1}A = A$. Associativity follows from the fact that composition of functions is associative. The group \mathcal{S}_A is called the symmetry group of A . The set $\mathcal{F}(\mathcal{Y})$ is a group also and \mathcal{S}_A is a subgroup of $\mathcal{F}(\mathcal{Y})$.

Example 2 Let (\mathcal{Y}, δ) be a discrete metric space where $\mathcal{Y} \subseteq \mathbb{N}$. In this case the isometry group is given by the group of permutations on \mathcal{Y} .

It is important to distinguish between the abstract group and a particular form it takes on a set of invertible functions. For this the concept of homomorphism and group action are very useful.

Definition 3 Let G_1 and G_2 be groups. A homomorphism is a map $\psi : G_1 \mapsto G_2$ for which $\psi(ab) = \psi(a)\psi(b)$ for all a and b in G_1 .

The map ψ has to respect the group law but is not assumed to be injective or surjective.

Definition 4 A group action of a group G on \mathcal{Y} is a homomorphism from G into $\mathcal{F}(\mathcal{Y})$.

These definitions formalize the notion of a symmetry group acting on \mathcal{Y} or one of its subsets. Finally, a group action by a group G acting on a set A is called transitive if for any $a, b \in A$ there exists a $g \in G$ such that $g \cdot a = b$.

To define channel symmetry we use a group action on the space of probability distributions defined on \mathcal{Y} . Now, let f be a function satisfying (4.8) and (4.9) and consider the action of an isometry ϕ on f . We define ϕf as

$$\phi f : x \mapsto f(\phi^{-1}(x)) \quad (4.17)$$

With this definition a group acting on \mathcal{Y} induces a group action on the set of probability distributions. To show this let g_1 and g_2 be members of the group acting on \mathcal{Y} and consider $g_1 g_2$ acting on f

$$\begin{aligned} (g_1 g_2)f(x) &= f((g_1 g_2)^{-1}x) = f((g_2^{-1} g_1^{-1})x) \\ &= f(g_2^{-1}(g_1^{-1}x)) = (g_2 f)(g_1^{-1}x) = (g_1(g_2 f))(x). \end{aligned} \quad (4.18)$$

From this we see that we have the homomorphism required for the group action. Next, consider a set of functions $\{f_0, \dots, f_{N-1}\}$ where

$$f_i : \mathcal{Y} \mapsto \mathbb{R} \quad \text{for } i = 0, \dots, N-1. \quad (4.19)$$

We define the symmetry group of $\{f_0, \dots, f_{N-1}\}$ as the set of all $\phi \in \mathcal{I}(\mathcal{Y})$ which fix $\{f_0, \dots, f_{N-1}\}$ as a set. In the next sections we will see that we can characterize channel symmetry in terms of the existence of these symmetry groups.

Isometry and Integration

Isometries respect the geometrical properties of objects they are acting on and this has consequences when we consider integration.

Proposition 11 *Let $\phi \in \mathcal{I}(\mathcal{Y})$ and let $B \subset \mathcal{Y}$. For a function $f : \mathcal{Y} \mapsto \mathbb{R}$ we have*

$$\int_B f(y) dy = \int_{\phi B} (\phi f)(y) dy. \quad (4.20)$$

Proof 11 *For DMCs the isometries of \mathcal{Y} are the permutations and the proof is straightforward. For channels with outputs in \mathbb{R} or \mathbb{C} a proof can be constructed by using an explicit form of the isometries of \mathbb{R} or \mathbb{C} . In a general measure theoretical setting a proof can be given for measures constructed from the metric of the underlying metric space.*

Corollary 12 *In case $B = \mathcal{Y}$ we have*

$$\int_{\mathcal{Y}} (\phi f)(y) dy = \int_{\mathcal{Y}} f(y) dy. \quad (4.21)$$

4.3 Memoryless Discrete-Input Symmetric Channels

Our starting point in defining channel symmetry is mutual information. For discrete memoryless channels which are symmetric in the Gallager sense, mutual information is maximized for a uniform distribution on the channel inputs. This is the property we wish a symmetric channel to have and we use it to define channel symmetry. One of the features that this extension shares with the conventional definition is that channel symmetry will be a sufficient condition for a uniform distribution to maximize the mutual information between the channel input and output. The following proposition is useful for this purpose.

Proposition 13 Consider a memoryless channel with input alphabet $\mathcal{X} = \{0, 1, \dots, N - 1\}$. Furthermore, the channel is defined by a sequence of f_i for $i = 0, 1, \dots, N - 1$. The maximum value of $I(X; Y)$ is achieved for a uniform distribution on the channel input if and only if for each $i, j \in \mathcal{X}$ we have

$$\int_{\mathcal{Y}} f_i(y) \log_2 \frac{f_i(y)}{\sum_k f_k(y)} dy = \int_{\mathcal{Y}} f_j(y) \log_2 \frac{f_j(y)}{\sum_k f_k(y)} dy \quad (4.22)$$

Proof 12 The proposition is a more general version of Theorem 9 and the proof follows from the Karush-Kuhn-Tucker (KKT) conditions for convex functions.

4.3.1 Group Characterization of Channel Symmetry

Next, we investigate when equality in (4.22) occurs. For this it is instructive to first consider a channel with binary inputs. In this case (4.22) reduces to

$$\int_{\mathcal{Y}} f_0(y) \log_2 \frac{f_0(y)}{f_0(y) + f_1(y)} dy = \int_{\mathcal{Y}} f_1(y) \log_2 \frac{f_1(y)}{f_0(y) + f_1(y)} dy. \quad (4.23)$$

Suppose we can find an isometry ϕ of \mathcal{Y} acting on $f_0 \log_2 \frac{f_0}{f_0 + f_1}$ such that

$$\begin{aligned} & \phi \left(f_0 \log_2 \frac{f_0}{f_0 + f_1} \right) (y) \\ &= (\phi f_0)(y) \log_2 \frac{(\phi f_0)(y)}{(\phi f_0)(y) + (\phi f_1)(y)} = f_1(y) \log_2 \frac{f_1(y)}{f_0(y) + f_1(y)}. \end{aligned} \quad (4.24)$$

Now, from Proposition 11 it follows that we will have equality in (4.23). For this isometry we would have that

$$\phi f_0 = f_1, \quad (4.25)$$

and,

$$\phi \{f_0, f_1\} = \{f_0, f_1\}, \quad (4.26)$$

which implies that

$$\phi f_1 = f_0. \quad (4.27)$$

Combining (4.25) and (4.27), we observe that

$$\phi^2 = I, \quad (4.28)$$

where I denotes the identity map. Hence if there exists an isometry of the metric space such that $\phi f_0 = f_1$ and $\phi^2 = I$, we have equality and mutual information is maximized for a uniform distribution on the channel input. Such an isometry generates a cyclic subgroup of order 2 of $\mathcal{I}(\mathcal{Y})$. We can extend this to channels with a larger input alphabet as follows.

Theorem 14 *Consider a memoryless channel with input alphabet $\mathcal{X} = \{0, 1, \dots, N-1\}$ and output metric space (\mathcal{Y}, d) . Furthermore, the channel is defined by a set of functions $\{f_i\}$ for $i = 0, \dots, N-1$. If there exists a transitive symmetry group \mathcal{S} of $\{f_i\}$, the maximum value of $I(X; Y)$ is achieved for a uniform distribution on \mathcal{X} .*

Proof 13 *Since \mathcal{S} is a symmetry group of $\{f_i\}$ it follows that any $\phi \in \mathcal{S}$ fixes $\{f_i\}$ as a set. Furthermore, since \mathcal{S} is transitive, $\{f_i\}$ is in the orbit of any $f \in \{f_i\}$ under \mathcal{S} . Hence we can find a $\phi_{k,l} \in \mathcal{S}$ such that*

$$\phi_{k,l} f_k = f_l, \quad (4.29)$$

for any $k, l \in \mathcal{X}$. Now, we can write for any $l \in \mathcal{X}$

$$\begin{aligned} \int_{\mathcal{Y}} f_l(y) \log_2 \frac{f_l(y)}{\sum_j f_j(y)} dy &= \int_{\mathcal{Y}} (\phi_{k,l} f_k)(y) \log_2 \frac{(\phi_{k,l} f_k)(y)}{\sum_j f_j(y)} dy \\ &= \int_{\mathcal{Y}} (\phi_{k,l} f_k)(y) \log_2 \frac{(\phi_{k,l} f_k)(y)}{\sum_j (\phi_{k,l} f_j)(y)} dy = \int_{\phi_{k,l}^{-1} \mathcal{Y}} f_k(y) \log_2 \frac{f_k(y)}{\sum_j f_j(y)} dy \\ &= \int_{\mathcal{Y}} f_k(y) \log_2 \frac{f_k(y)}{\sum_j f_j(y)} dy, \end{aligned} \quad (4.30)$$

where we have made use of Proposition 11. Since this holds for any $k, l \in \mathcal{X}$ we will have equality in (4.22) for any $k, l \in \mathcal{X}$.

4.3.2 Representation of Cyclic Symmetry Groups in \mathbb{R}^n

We are mainly concerned with channels with a binary input and outputs from \mathbb{R}^n . We derive some properties in this section which are applied in the next section when we consider applications to coded modulation.

The group of isometries of the metric space (\mathbb{R}^n, d_e) is called the *Euclidian* group and we denote this group by \mathbb{E} . Any element of the Euclidian group can be represented by [45]

$$\begin{aligned} \phi : \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ \mathbf{x} &\mapsto \mathbf{A}\mathbf{x} + \mathbf{b}, \end{aligned} \quad (4.31)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix and $\mathbf{b} \in \mathbb{R}^n$. Now, let N be an integer. A subset of transitive group actions on \mathbb{R}^n are defined by transformation of the form

$$\phi : \mathbf{x} \rightarrow \mathbf{A}\mathbf{x} + \left(\mathbf{I}_n - \mathbf{A}^{N-1}\right) \mathbf{b}, \quad (4.32)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix for which satisfies $\mathbf{A}^N = \mathbf{I}_n$, $\mathbf{b} \in \mathbb{R}^n$ and \mathbf{I}_n is the identity matrix of dimension n . These transformations generate a cyclic group of order N .

4.3.3 Channel Symmetry for Channels with a Binary Input

In this section we specialize some of the results for channels with binary inputs and outputs from \mathbb{R}^n and show that the notion of channel symmetry defined in the previous section coincides with the notion of BIOS channels.

BIOS Channels

We define a BIOS channel as follows

Definition 5 A BIOS channel is a memoryless channel with input alphabet $\mathcal{X} = GF(2)$ and output alphabet $\mathcal{Y} = \mathbb{R}^n$ defined by two functions f_0 and f_1 related by an isometry of \mathbb{R}^n

$$\phi f_0 = f_1, \quad (4.33)$$

where $\phi^2 = I$.

Transformations given by (4.32) with $N = 2$ satisfy $\phi^2 = I$ and we can write

$$\phi : \mathbf{x} \rightarrow \mathbf{A}\mathbf{x} + (\mathbf{I}_n - \mathbf{A}) \mathbf{b}, \quad (4.34)$$

where \mathbf{A} is a matrix for which $\mathbf{A}^2 = \mathbf{I}_n$ and $\mathbf{b} \in \mathbb{R}^n$. Since $\mathbf{A}^2 = \mathbf{I}_n$, \mathbf{A} is an orthogonal symmetric matrix. Moreover any real symmetric matrix is diagonalizable by an orthogonal matrix. Hence \mathbf{A} has the form

$$\mathbf{A} = \mathbf{Q}\mathbf{S}\mathbf{Q}^T, \quad (4.35)$$

where \mathbf{Q} is an orthogonal matrix and \mathbf{S} is a diagonal matrix with entries equal to 1 or -1 . From this we conclude that we can write ϕ as

$$\phi : \mathbf{x} \mapsto \mathbf{Q}\mathbf{S}\mathbf{Q}^T (\mathbf{x} - \mathbf{b}) + \mathbf{b} \quad (4.36)$$

Example 3 Consider a memoryless channel with a binary input and outputs from \mathbb{R} . We have defined a binary-input output-symmetric channel as a channel for which the transition density satisfies

$$\phi f_{Y|X}(y|0) = f_{Y|X}(y|1), \quad (4.37)$$

where ϕ is an isometry for which $\phi^2 = I$. In this case the nontrivial isometries are of the form

$$\phi : x \mapsto -x + b, \quad (4.38)$$

where $b \in \mathbb{R}$. For $b = 0$ we have the conventional definition of a binary-input output-symmetric channel.

$$f_{Y|X}(y|0) = f_{Y|X}(-y|1), \quad (4.39)$$

Log-likelihood Ratio and Densities

In case the channel input is binary the log-likelihood ratio (LLR) is defined as

$$L(\mathbf{y}) = \log \frac{f_0(\mathbf{y})}{f_1(\mathbf{y})}. \quad (4.40)$$

The LLR is a sufficient statistic to make the decision on the transmitted bit. Many decoders of binary error-correcting codes use log-likelihood ratios during the decoding process. A good example is the message-passing decoder which is for instance used to decode LDPC codes. To analyze the performance of the decoders, we are interested in the distribution of the LLRs. For symmetric channels the distribution of LLRs satisfies some convenient properties which we will derive in this section for symmetric channels. As a random variable the LLR is a measurable function

$$L : \mathcal{Y} \mapsto \mathbb{R}. \quad (4.41)$$

We are interested in the distribution of this random variable conditioned on the channel input. Let $\mu_{L|0}$ and $\mu_{L|1}$ define a probability measure on \mathbb{R} defining the distribution of the LLR conditioned on the transmission of a 0 and 1, respectively. This measure is induced by L and defined as

$$\mu_{L|0}(A) = P_{\mathcal{Y}|0}(L^{-1}(A)), \quad (4.42)$$

where $P_{\mathcal{Y}|0}$ defines the channel and denotes a measure on the channel output space given that a zero is transmitted. This measure is induced by the channel transition density.

In [26] a distribution F is defined to be symmetric if it satisfies:

$$\int_{\mathbb{R}} h(x) dF(x) = \int_{\mathbb{R}} e^{-x} h(-x) dF(x), \quad (4.43)$$

for any $h(x)$ for which the integral exists. In a similar fashion we define $\mu_{L|0}$ to be symmetric if it satisfies

$$\int_{\mathbb{R}} h(x) d\mu_{L|0}(x) = \int_{\mathbb{R}} e^{-x} h(-x) d\mu_{L|0}(x), \quad (4.44)$$

Theorem 15 [Symmetry of distributions of output-symmetric channels]: *Consider a binary-input output-symmetric channel with output space \mathcal{Y} defined by an f_0 and f_1 . Conditioned on the transmission of a 0 the measure defining the distribution of log-likelihood ratios is symmetric.*

Proof 14 *We give a sketch of the proof. First, note that we have the following basic properties of the log-likelihood ratio*

$$f_0(y) = e^{L(y)} f_1(y), \quad (4.45)$$

and,

$$L(y) = \log \frac{f_0(y)}{f_1(y)} = \log \frac{\phi f_1(y)}{\phi f_0(y)} = -\log \frac{\phi f_0(y)}{\phi f_1(y)} = -\phi L(y). \quad (4.46)$$

Now, consider a measurable h for which $\int h d\mu_{L|0}$ exists. Furthermore, we assume that $P_{\mathcal{Y}|0}$ and $P_{\mathcal{Y}|1}$ are absolutely continuous with respect to some measure μ_d . For many channels this measure is related to the underlying metric space. In the case that $\mathcal{Y} = \mathbb{R}^n$, μ_d is the Lebesgue measure. In this case we can write,

$$\begin{aligned} \int_{\mathbb{R}} h(x) d\mu_{L|0}(x) &= \int_{\mathcal{Y}} h(L(y)) dP_{\mathcal{Y}|0}(y) \\ &= \int_{\mathcal{Y}} h(L(y)) f_0(y) d\mu_d(y) = \int_{\mathcal{Y}} h(L) f_0 d\mu_d, \end{aligned} \quad (4.47)$$

where the before last step follows from the absolute continuity of $P_{\mathcal{Y}|0}$ with respect to μ_d . In the last step we have dropped the use of the dummy y to simplify notation. Next by using the symmetry property of the channel and write

$$\begin{aligned} \int_{\mathcal{Y}} h(L) f_0 d\mu_d &= \int_{\mathcal{Y}} h(L) e^L f_1 d\mu_d = \int_{\phi^{-1}\mathcal{Y}} h(\phi L) e^{\phi L} \phi f_1 d\mu_d \\ &= \int_{\mathcal{Y}} h(-L) e^{-L} f_0 d\mu_d \int_{\mathcal{Y}} h(-L) e^{-L} dP_{\mathcal{Y}|0} = \int_{\mathbb{R}} h(-x) e^{-x} d\mu_{L|0}(x). \end{aligned} \quad (4.48)$$

Corollary 16 Consider a binary-input output-symmetric channel with output space \mathbb{R}^n defined by an f_0 and f_1 . Consider a binary-input output-symmetric channel with output space \mathcal{Y} defined by an f_0 and f_1 . In case $\mu_{L|0}$ admits a density f_L it satisfies

$$f_L(x) = e^x f_L(-x). \quad (4.49)$$

4.4 Applications to Coded Modulation

4.4.1 Modulation for Channels with Additive Noise

In this section we give an application of the results developed in the previous section. Consider communication on discrete-time memoryless channels with inputs and outputs from \mathbb{R}^n . Moreover, we assume that the channel is an additive noise channel such that a channel output Y is given by

$$Y = X + N, \quad (4.50)$$

where $X \in \mathbb{R}^n$ denotes the channel input and $N \in \mathbb{R}^n$ the additive noise. We assume that N is a random variable which admits a density f_N . Hence the channel is defined by a channel transition probability density function $f_{Y|X}$

$$f_{Y|X}(y|x) = f_N(y - x). \quad (4.51)$$

Let f_X denote a probability density function defining the distribution of X . The mutual information between Y and X is given by:

$$I(Y; X) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f_{Y|X}(y|x) f_X(x) \log_2 \frac{f_{Y|X}(y|x)}{\int_{\mathbb{R}^n} f_{Y|X}(y|x') f_X(x')} dx dy, \quad (4.52)$$

and the maximum value of $I(Y; X)$ is achieved for some f_X . For many channels it is not straightforward to generate channel inputs from f_X , because it is possible that the optimal f_X is the density of a continuous random variable while the source provides bits. We consider generating channel inputs from a discrete set $\mathcal{S} \subset \mathbb{R}^n$ to which we refer to as the *signal constellation*. If a channel input Z is generated from \mathcal{S} according to a probability assignment $P_{\mathcal{S}}$ on the elements of \mathcal{S} , we can achieve an average mutual information of

$$I(Y; Z) = \sum_{z \in \mathcal{S}} \int_{\mathbb{R}^n} f_{Y|X}(y|z) P_{\mathcal{S}}(z) \log_2 \frac{f_{Y|X}(y|z)}{\sum_{z' \in \mathcal{S}} f_{Y|X}(y|z') P_{\mathcal{S}}(z')} dy. \quad (4.53)$$

We want to design \mathcal{S} and $P_{\mathcal{S}}$ in such a way that $I(Y; Z)$ is as close to the capacity of the channel as possible. To achieve reliable communication one can opt for using an error-correcting code with a codeword symbol alphabet equal to \mathcal{S} where the marginal distribution on the codeword symbols is given by $P_{\mathcal{S}}$. However, with this approach it is not easy to come up with practical encoding and decoding methods. Thus we take a different approach.

At the transmitter we use a set of codes with codeword symbol alphabet \mathcal{A} considerably smaller than the cardinality of \mathcal{S} . Here we implicitly assume that we can find codes over this smaller alphabet with feasible encoding and decoding algorithms. In this chapter we opt for binary codes and we can use for instance LDPC codes. Let X_1, \dots, X_d denote a tuple of uniform i.i.d. random variables taking values in \mathcal{A} . The realization of the tuple is used to generate an element from the signal constellation. For this purpose we define a *modulation map* Φ

$$\Phi : (X_1, \dots, X_d) \rightarrow \mathbb{R}^n. \quad (4.54)$$

With the modulation map we generate a channel input Z as

$$Z = \Phi(X_1, \dots, X_d), \quad (4.55)$$

and the additive noise channel takes the form

$$Y = Z + N = \Phi(X_1, \dots, X_d) + N. \quad (4.56)$$

The mutual information between X_1, \dots, X_d and Y is given by

$$\begin{aligned} I(Y; (X_1, \dots, X_d)) &= I(Y; Z) \\ &= \sum_{z \in \mathcal{S}} \int_{\mathbb{R}^n} f_{Y|X}(y|z) P_{\mathcal{S}}(z) \log_2 \frac{f_{Y|X}(y|z)}{\sum_{z' \in \mathcal{S}} f_{Y|X}(y|z') P_{\mathcal{S}}(z')} dy, \end{aligned} \quad (4.57)$$

where the first equality follows from the fact that Z is a deterministic function of X_1, \dots, X_d . Furthermore, we can also write $I(Y; (X_1, \dots, X_d))$ as

$$\begin{aligned} I(Y; (X_1, \dots, X_d)) &= I(X_1; Y) + I(X_2; Y|X_1) + \\ &\quad \dots + I(X_d; Y|X_1, \dots, X_{d-1}), \end{aligned} \quad (4.58)$$

by the chain rule of mutual information. If each of the X_i is encoded separately, we can use a sequential decoding procedure at the receiver. This is the multilevel coding (MLC) strategy with multi-stage decoding (MSD) [8]. An alternative is to drop the conditioning on the X_i in (4.58) which gives us:

$$I(Y; (X_1, \dots, X_d)) \geq I(X_1; Y) + I(X_2; Y) + \dots + I(X_d; Y). \quad (4.59)$$

In the latter case we can also opt for a single code at the transmitter which naturally leads to bit-interleaved coded modulation (BICM) [9]. Both approaches reduce the problem of approaching capacity on the channel to achieving capacity on a set of equivalent channels. In MLC-MSD we can achieve a rate given by the left-handside of (4.58) by multi-stage decoding. The rates achievable on each of the equivalent channels is given by the right-handside of (4.58). With BICM we can achieve a rate given by the right-handside of (4.59) by parallel-and-independent (PID) decoding. However, in both cases the achievable rate and the nature of the equivalent channels depends on the choice of Φ . Next, we use the results of the previous sections to identify Φ for which the equivalent binary channels are symmetric. We will use MLC-MSD.

4.4.2 Symmetric Channels and Symmetric Constellations

Consider a channel with additive noise defined by (4.50). Moreover, we assume that N admits a density f_N . We consider the AWGN channel for which f_N is given by

$$f_N(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\mathbf{x}^T \mathbf{x}}{2\sigma^2}\right), \quad (4.60)$$

where $\mathbf{x} \in \mathbb{R}^n$. We assume that f_N exhibits symmetry and we consider f_N which are invariant with respect to the elements of the orthogonal group of \mathbb{R}^n which we denote by \mathcal{O}_n .

$$\phi f_N = f_N \quad \text{for } \phi \in \mathcal{O}_n. \quad (4.61)$$

The multivariate Gaussian density of (4.60) satisfies this condition.

To communicate over this channel we use a signal constellation \mathcal{S} generated by a modulation map Φ

$$\Phi: \text{GF}(2) \mapsto \mathbb{R}^n. \quad (4.62)$$

Next, consider the case where we encode each of the X_i by a binary code and use a multistage decoding procedure at the receiver. For the purpose of error-control coding we are interested in the properties of the equivalent binary channels. These properties are determined by the choice of Φ and we wish to choose Φ such that the equivalent binary channels are output-symmetric.

Now define $\mathcal{S}_{0|x_1, \dots, x_{l-1}}$ and $\mathcal{S}_{1|x_1, \dots, x_{l-1}}$ as

$$\begin{aligned} \mathcal{S}_{0|x_1, \dots, x_{l-1}} &= \{\Phi(x_1, \dots, x_{l-1}, 0, x_{l+1}, \dots, x_d) | x_j \in \{0, 1\}, j = l+1 \dots d\} \\ \mathcal{S}_{1|x_1, \dots, x_{l-1}} &= \{\Phi(x_1, \dots, x_{l-1}, 1, x_{l+1}, \dots, x_d) | x_j \in \{0, 1\}, j = l+1 \dots d\}. \end{aligned}$$

The sets $\mathcal{S}_{0|x_1, \dots, x_{l-1}}$ and $\mathcal{S}_{1|x_1, \dots, x_{l-1}}$ partition the signal constellation in two parts. $\mathcal{S}_{0|x_1, \dots, x_{l-1}}$ contains all constellation elements for which X_1 to X_{l-1} are given by x_1 to x_{l-1} and $X_l = 0$. $\mathcal{S}_{1|x_1, \dots, x_{l-1}}$ contains all constellation elements for which X_1 to X_{l-1} are given by x_1 to x_{l-1} and $X_l = 1$. With these definitions we observe that the binary channel for X_l is defined by a pair of transition densities $f_{l,0}$ and $f_{l,1}$

$$\begin{aligned} f_{l,0}(\mathbf{y}) &= \frac{1}{2^{d-l}} \sum_{\mathbf{x} \in \mathcal{S}_{0|x_1, \dots, x_{l-1}}} f_N(\mathbf{y} - \mathbf{x}) \\ f_{l,1}(\mathbf{y}) &= \frac{1}{2^{d-l}} \sum_{\mathbf{x} \in \mathcal{S}_{1|x_1, \dots, x_{l-1}}} f_N(\mathbf{y} - \mathbf{x}) \end{aligned} \quad (4.63)$$

Now, let \mathcal{T}_n define the isometry group of translations of \mathbb{R}^n and write the elements of \mathcal{T}_n as

$$\phi_T(\mathbf{x})\mathbf{y} = \mathbf{y} + \mathbf{x}, \quad (4.64)$$

where $\mathbf{y} \in \mathbb{R}^n$. In terms of the elements of \mathcal{T}_n we can write $f_{l,0}$ and $f_{l,1}$ as

$$\begin{aligned} f_{l,0} &= \frac{1}{2^{d-l}} \sum_{\mathbf{x} \in \mathcal{S}_{0|x_1, \dots, x_{l-1}}} \phi_T(\mathbf{x})f_N \\ f_{l,1} &= \frac{1}{2^{d-l}} \sum_{\mathbf{x} \in \mathcal{S}_{1|x_1, \dots, x_{l-1}}} \phi_T(\mathbf{x})f_N \end{aligned} \quad (4.65)$$

The following theorem provides a sufficient condition on the constellation partition $\{\mathcal{S}_{0|x_1, \dots, x_{l-1}}, \mathcal{S}_{1|x_1, \dots, x_{l-1}}\}$ such that the equivalent binary channel for X_l is symmetric.

Theorem 17 *Consider a channel with inputs and output from \mathbb{R}^n as defined by (4.50). Moreover, the noise density f_N is invariant with respect to the elements of \mathcal{O}_n . We use a modulation map $\Phi(X_1, \dots, X_d)$ to generate a signal constellation and at the receiver we use a multistage decoding procedure where the order in which we decode the bits is X_1, X_2, \dots, X_d . The equivalent binary channel at level l is defined by $f_{l,0}$ and $f_{l,1}$. This binary channel is symmetric if there exists an isometry ϕ of \mathbb{R}^n with $\phi^2 = I$ such that*

$$\phi\mathcal{S}_{0|x_1, \dots, x_{l-1}} = \mathcal{S}_{1|x_1, \dots, x_{l-1}}. \quad (4.66)$$

Proof 15 In case (4.66) holds there exists for each $x_0 \in \mathcal{S}_{0|x_1, \dots, x_{l-1}}$ a $x_1 \in \mathcal{S}_{0|x_1, \dots, x_{l-1}}$ with $x_1 = \phi x_0$. Now, if we show that

$$\phi \phi_{T(x_0)} f_N = \phi_{T(x_1)} f_N, \quad (4.67)$$

we have

$$\begin{aligned} \phi f_{l,0} &= \phi \sum_{x \in \mathcal{S}_{0|x_1, \dots, x_{l-1}}} \phi_{T(x)} f_N = \sum_{x \in \mathcal{S}_{0|x_1, \dots, x_{l-1}}} \phi \phi_{T(x)} f_N \\ &= \sum_{x \in \mathcal{S}_{1|x_1, \dots, x_{l-1}}} \phi_{T(x)} f_N = f_{l,1}, \end{aligned} \quad (4.68)$$

which would proof the theorem. To show this, first note that we can write any isometry ϕ of \mathbb{R}^n for which $\phi^2 = I$ as

$$\phi = \phi_{T(\mathbf{b})} \phi_{O(\mathbf{A})} \phi_{T(-\mathbf{b})}, \quad (4.69)$$

for $\mathbf{b} \in \mathbb{R}^n$ and where $\phi_{O(\mathbf{A})}$ is an element of \mathcal{O}_n and is defined by an orthogonal matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. Next, note that for the elements of the translational group we have

$$\phi_{T(\phi_T(\mathbf{b})\mathbf{c})} = \phi_{T(-\mathbf{b})} \phi_{T(\mathbf{c})}, \quad (4.70)$$

for $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$. This allows us to write the left-handside of (4.67) as

$$\begin{aligned} \phi \phi_{T(x_0)} f_N &= \phi \phi_{T(\phi^{-1}x_1)} f_N = \phi \phi_{T(\phi_{T(-\mathbf{b})} \phi_{O(\mathbf{A})} \phi_{T(\mathbf{b})} x_1)} f_N \\ &= \phi \phi_{T(\mathbf{b})} \phi_{T(\phi_{O(\mathbf{A})} \phi_{T(\mathbf{b})} x_1)} f_N. \end{aligned} \quad (4.71)$$

Next, note that we have the following property for composition of elements from the translational isometry group and the orthogonal isometry group of \mathbb{R}^n

$$\phi_{T(\phi_{O(\mathbf{A})}\mathbf{b})} = \phi_{O(\mathbf{A})} \phi_{T(\mathbf{b})} \phi_{O(\mathbf{A})}, \quad (4.72)$$

for $\mathbf{b} \in \mathbb{R}^n$. This allows us to write

$$\begin{aligned} \phi \phi_{T(x_0)} f_N &= \phi \phi_{T(\mathbf{b})} \phi_{O(\mathbf{A})} \phi_{T(\phi_{T(\mathbf{b})} x_1)} \phi_{O(\mathbf{A})} f_N \\ &= \phi \phi_{T(\mathbf{b})} \phi_{O(\mathbf{A})} \phi_{T(-\mathbf{b})} \phi_{T(x_1)} \phi_{O(\mathbf{A})} f_N \\ &= \phi^2 \phi_{T(x_1)} \phi_{O(\mathbf{A})} f_N = \phi_{T(x_1)} \phi_{O(\mathbf{A})} f_N. \end{aligned} \quad (4.73)$$

Now we make use of the symmetry of f_N with respect to the elements of \mathcal{O}_n and write (4.73) as

$$\phi \phi_{t(x_0)} f_N = \phi_{T(x_1)} \phi_{O(\mathbf{A})} f_N = \phi_{T(x_1)} f_N. \quad (4.74)$$

R	SNR [dB]	d	$ S $	$I(Y; Z)$	Δ_{AWGN} [dB]	Δ_{QAM} [dB]
3	8.45	5	32	2.931	0.23	0.30
3	8.45	6	64	2.968	0.11	0.44
4	11.76	6	64	3.885	0.37	0.47

Table 4.1: Parameters of the designed isometry signal constellations for the two-dimensional AWGN channel

We observe in the proof of the theorem that the symmetry group of the noise density gives us the essential degrees of freedom in choosing the signal constellation.

4.4.3 Design Example for the AWGN Channel

Now, consider a multilevel coded system for which we want to design a modulation map which is able to perform close to the capacity of the AWGN channel. We can generate channel inputs for the AWGN channel by choosing the modulation map as

$$Z = \phi(X_1, \dots, X_d) = \phi_1^{X_1} \phi_2^{X_2} \dots \phi_d^{X_d} \mathbf{c}_0, \quad (4.75)$$

where \mathbf{c}_0 is a constant from \mathbb{R}^n and ϕ_1 to ϕ_d are representations of the cyclic isometry group of order 2 on \mathbb{R}^n . Furthermore, X_1, \dots, X_d are uniform i.i.d. random bit variables taking values in $\{0, 1\}$. When we use a multistage decoding procedure at the receiver with decoding order X_1, \dots, X_d , the equivalent binary channel are output symmetric.

As in [39] we can consider a numerical maximization of the constrained constellation capacity. For this purpose we consider an AWGN channel in two dimensions. As an illustration we have performed a numerical maximization of the constrained constellation capacity for modulation maps given by (4.75) for target rates of 3 bit/use and 4 bit/use and the optimization results are given in Table 4.1. The table gives the target rate R for which we the design the signal constellation and the SNR for which the two-dimensional AWGN channel achieves the rate. The rate which is achieved at the SNR is given by $I(Y; Z)$. The distance from the capacity of the channel is given by Δ_{AWGN} and the gain with respect to a conventional quadrature amplitude modulation (QAM) constellation with the same number of constellation symbols is given by Δ_{QAM} . The geometrical structure of these constellations is illustrated in Figure 4.1, Figure 4.2 and Figure 4.3.

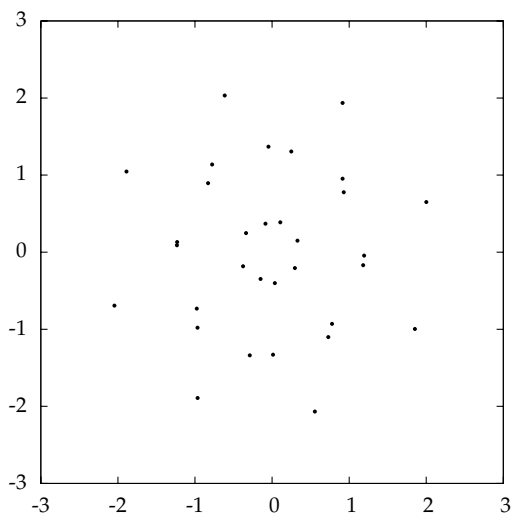


Figure 4.1: Isometry constellation with $d = 5$ designed for 3 bit/s/Hz.

The constrained constellation capacity of these constellation is close to the capacity of the AWGN and all constellations achieve a gain with respect to QAM constellations. However, the target rates are relatively low and for higher rates gains will be larger. As in [39] we can use binary linear codes (e.g. LDPC) to achieve the constrained constellation capacity since the equivalent binary channels are symmetric.

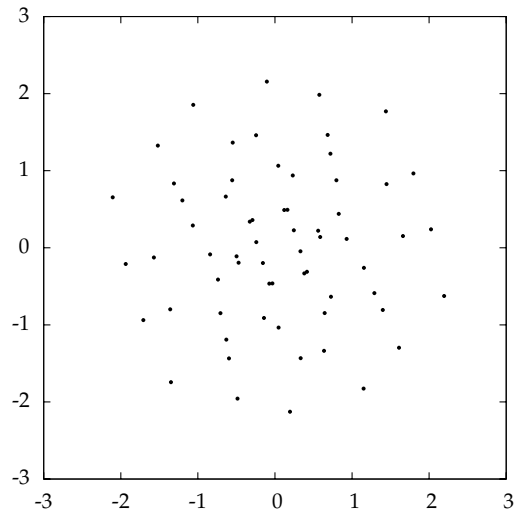


Figure 4.2: Isometry constellation with $d = 6$ designed for 3 bit/s/Hz.

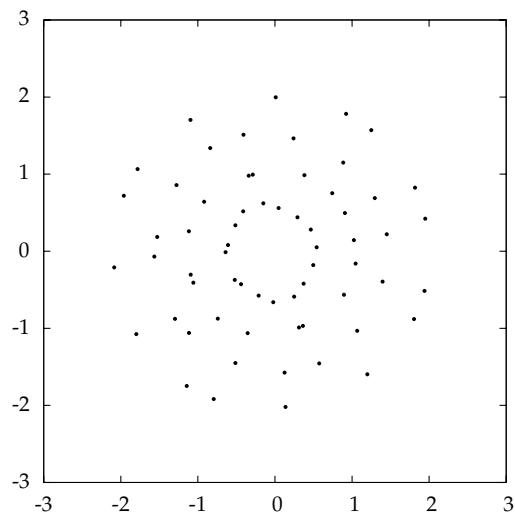


Figure 4.3: Isometry constellation with $d = 6$ designed for 4 bit/s/Hz.

4.5 Open Questions and Future Research

We conclude this chapter with some open questions and suggestions for future research. The notion of channel symmetry we have defined has relations with Gallager symmetry and the concept of a regular channel. We expect that these two concepts are a subset of our definition of channel symmetry. For instance one can restrict oneself to cyclic isometry groups which would provide a stronger notion of channel symmetry. In the case of a channel with binary channel inputs the only group is the cyclic group with two elements.

In the end we provided an application by letting isometries of \mathbb{R}^n act on some initial vector. These results are related to group codes for the AWGN channel which is a widely studied subject. Furthermore, the results we have presented in this chapter have a relation with the so-called geometrically uniform constellations of [46]. In fact, the modulation maps we have constructed for \mathbb{R}^2 lead to a geometrically uniform constellation partition.

We only considered multilevel coding with multistage decoding as an application. However, an interesting question is if one can construct constellations for channels with Gaussian noise in e.g. \mathbb{R}^3 and \mathbb{R}^4 which have a good parallel-and-independent decoding capacity and provide signal shaping. For these constellations one can use a single binary error-correcting code in a bit-interleaved coded modulation setting.

For some modulation maps an analytical maximization of the constrained constellation capacity might be feasible. This is worth investigating since for higher values of d the number of degrees of freedom increases and numerical optimization becomes difficult.

Bibliography

- [1] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.
- [2] P. Oswald and A. Shokrollahi, "Capacity-achieving sequences for the erasure channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3017 – 3028, Dec. 2002.
- [3] S.-Y. Chung, G. D. Forney, Jr., T. J. Richardson, and R. L. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, pp. 58–60, Feb. 2001.
- [4] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, pp. 379–423, 623–656, July 1948.
- [5] N. Bouman and H. S. Cronie, "Power- and bandwidth-efficient coded modulation for linear Gaussian channels," submitted to ISIT2008.
- [6] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Communications '93*, May 1993, pp. 1064–1070.
- [7] 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. Inform. Theory*, vol. 49, pp. 2141–2155, Sept. 2003.
- [8] H. Imai and S. Hiraakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inform. Theory*, vol. 23, pp. 371–377, May 1977.

-
- [9] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.
- [10] P. Limpaphayom and K. A. Winick, "Power- and bandwidth-efficient communications using LDPC codes," *IEEE Trans. Commun.*, vol. 52, pp. 350 – 354, Mar. 2004.
- [11] L. Duan, B. Rimoldi, and R. L. Urbanke, "Approaching the AWGN channel capacity without active shaping," in *Proc. IEEE Int. Symp. on Information Theory*, June/July 1997, p. 374.
- [12] A. Bennatan and D. Burshtein, "Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels," *IEEE Trans. Inform. Theory*, vol. 52, pp. 549 – 583, Feb. 2006.
- [13] F.-W. Sun and H. C. A. van Tilborg, "Approaching capacity by equiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1714–1716, Sept. 1993.
- [14] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 23, pp. 2384–2415, Oct. 1998.
- [15] G. D. Forney, Jr. and L.-F. Wei, "Multidimensional constellations. I. introduction, figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 877–892, Aug. 1989.
- [16] G. D. Forney, Jr., "Coset codes. I. introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, Sept. 1988.
- [17] —, "Coset codes. II. binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1152–1187, Sept. 1988.
- [18] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 273–278, Jan. 1998.
- [19] U. Erez and R. Zamir, "Achieving $1/2 \log(1+\text{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2293–2314, Oct. 2004.

- [20] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 36, pp. 726–740, July 1990.
- [21] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 913–929, May 1993.
- [22] L. Duan, "Approaching the capacity of the AWGN channel via multilevel coding," Ph.D. dissertation, Washington University, 1997.
- [23] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1361–1391, July 1999.
- [24] R. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. 31, pp. 124–142, Mar. 1985.
- [25] K. V. Price, "Differential evolution: a fast and simple numerical optimizer," in *Biennial Conference of the North American Fuzzy Information Processing Society, NAFIPS*, June 1996.
- [26] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.
- [27] K. V. Price, R. Storn, and J. A. Lampinen, *Differential Evolution*. Springer, 2005.
- [28] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley and Sons, Inc., 1968.
- [29] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008. [Online]. Available: <http://lthcwwww.epfl.ch/mct/index.php>
- [30] R. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, pp. 533–547, Sept. 1981.
- [31] A. Amraoui, "Asymptotic and finite-length optimization of LDPC codes," Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, 2006.

- [32] S. ten Brink, "Convergence of iterative decoding," *Electronics Letters*, vol. 35, pp. 1117–1119, June 1999.
- [33] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, pp. 670–678, Apr. 2004.
- [34] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-length scaling for iteratively decoded LDPC ensembles," *submitted to IEEE Trans. Inform. Theory*.
- [35] R. G. Gallager, "Low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. IT-8, pp. 21–28, Jan. 1962.
- [36] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. 28, pp. 55–67, Jan. 1982.
- [37] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 1, pp. 169–171, Nov. 1997.
- [38] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8 PSK signaling," *IEEE Trans. Commun.*, vol. 50, pp. 1250–1257, Aug. 2002.
- [39] H. S. Cronie, "Power and bandwidth efficient communication on the AWGN channel by the superposition of binary error-correcting codes," *submitted to IEEE Transactions on Information Theory*, Oct. 2007.
- [40] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inform. Theory*, vol. 47, pp. 585–598, Feb. 2001.
- [41] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.
- [42] P. Delsarte and P. Piret, "Algebraic constructions of Shannon codes for regular channels," *IEEE Trans. Inform. Theory*, vol. 28, pp. 593–599, July 1982.
- [43] G. D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, Sept. 1991.

-
- [44] A. W. Knap, *Basic Algebra*. Birkhäuser, 2006.
- [45] D. L. Johnson, *Symmetries*. Springer-Verlag, 2001.
- [46] G. D. Forney, Jr., M. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inform. Theory*, vol. 46, pp. 820–850, May 2000.