Lehrstuhl für Nachrichtentechnik

Iterative Decoding of Bit-Interleaved Coded Modulation

Frank Schreckenbach

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

Doktor–Ingenieurs

genehmigten Dissertation.

Vorsitzender:	Univ. Prof. Dr.–Ing. Alexander W. Koch
Prüfer der Dissertation:	1. Univ. Prof. Dr.–Ing. Dr.–Ing. E.h. Joachim Hagenauer
	2. Univ. Prof. Dr.–Ing. Joachim Speidel, Universität Stuttgart

Die Dissertation wurde am 23.01.2007 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 20.04.2007 angenommen.

Preface

This thesis has been written during my time as research assistant at the Institute for Communications Engineering (LNT) at the Munich University of Technology (TUM). During this time, many persons have contributed in various ways to the successful outcome of this work.

First of all, I am deeply grateful to my supervisor, Prof. Dr.–Ing. Dr.–Ing. E.h. Joachim Hagenauer for giving me the chance to work at his institute and for his support in all aspects. A special thanks goes to him for giving me the opportunity to teach the lecture and tutorial of a graduate-level class in channel coding. I highly appreciated this rewarding experience.

I am also very grateful to Prof. Dr.–Ing. Joachim Speidel for acting as co-supervisor and to Dr. Mark Reed for the excellent collaboration and his hospitality during my research visit at National ICT Australia (NICTA) in Canberra.

For the fruitful co-operation with DoCoMo Eurolabs in Munich I would like to thank, representatively, Dr. Gerhard Bauch.

I am very thankful to all my colleagues and diploma students for the enjoyable atmosphere and the inspiring working environment at the Institute for Communications Engineering. I would like to mention in particular Dr. Norbert Görtz for his valuable contributions at the beginning of my thesis, Christoph Hausl for the collaboration in the DoCoMo Eurolabs project and my office mate Dr. Hrvoje Jenkac for the great time.

Finally, I would like to thank my family for their unique support and encouragement they provided me throughout the years. Last but not least, thanks to Jana for her overwhelming love and for making our life a wonderful journey.

München, September 2007 Frank Schreckenbach

Contents

Abstract

This thesis addresses techniques for wireless communication systems to transmit high data rates with high reliability. The investigations are focused on the bit-interleaved coded modulation (BICM) system that is the serial concatenation of a channel encoder, an interleaver and a symbol mapper for high order modulation. The performance of BICM can be greatly improved through iterative information exchange between the demapper and the decoder at the receiver. This system is usually referred to as BICM with iterative decoding (BICM-ID). The major contributions of this work are the study of the iterative BICM receiver and the investigation of modulation techniques to improve the adaptability and performance of the system: First, properties of decoder and demapper EXIT functions are derived to gain additional insights on the convergence of the iterative BICM receiver. Then, the focus is on mapping techniques, i.e. on the assignment of bit sequences to signal points of a high order signal constellation. Methods to characterize and optimize mappings for a variety of applications are investigated. In particular, mappings optimized for iterative receivers are proposed. Along the same line of thought, the combination of different signal constellations and mappings within one code word, non-unique mappings and superimposed mappings are investigated to improve the adaptability and performance of the system using modulation techniques.

Zusammenfassung

In der vorliegenden Dissertation werden Methoden für drahtlose Kommunikationssysteme untersucht, die es ermöglichen Information sowohl mit hohen Datenraten als auch mit einer hohen Zuverlässigkeit zu übertragen. Es wird insbesondere das *Bit-Interleaved Coded Modulation* (BICM)-System betrachtet, das aus einer seriellen Verkettung eines Kanalencoders, eines Interleavers sowie einer Abbildung der Bits auf Symbole einer höherstufigen Signalraumkonstellation (sog. *Mapping*) besteht. Die Leistungsfähigkeit von BICM kann durch iterativen Informationsaustausch zwischen dem Demapper und dem Decoder stark verbessert werden. Dieses System wird üblicherweise als BICM mit iterativer Decodierung bezeichnet (BICM-ID). Die Hauptbeiträge dieser Arbeit sind die Untersuchung des iterativen BICM Empfängers und die Entwicklung von Modulationstechniken, um die Anpassungs- und Leistungsfähigkeit des Systems zu verbessern: Zuerst werden Eigenschaften der Decoder- und Demapper EXIT-Funktionen abgeleitet, wodurch zusätzliche Einblicke in das Konvergenzverhalten des iterativen BICM Empfängers erlangt werden. Anschließend liegt der Schwerpunkt auf Mapping-Techniken. Methoden um Mappings für eine Vielzahl von Anwendungen zu Beschreiben und zu Optimieren werden untersucht. Insbesondere werden optimierte Mappings für iterative Empfänger hergeleitet. Weiterhin werden die Kombination von verschiedenen Signalraumkonstellationen und Mappings innerhalb eines Codewortes, mehrdeutige Mappings sowie überlagerte Mappings untersucht, die eine verbesserte Anpassungs- und Leistungsfähigkeit des Systems mit Hilfe von Modulationstechniken ermöglichen.

1

Introduction

The vision of being able to communicate and access data anywhere and anytime has driven the research and development of powerful communication systems forward at a remarkable speed. Today, digital communication systems have become an integral part of our everyday life. The wireless communication standards GSM[1], UMTS[2] and WLAN[3], as well as DAB[4], DVB[4] and Internet applications are only few examples.

The next generation wireless communication systems (namely, 3GPP long term evolution (LTE) and fourth generation (4G) communications systems) will be used for a large variety of advanced applications and should be able to support widely varying user needs, service requirements and radio environments (home, office, vehicular, cellular/mobile, satellite, etc). Nevertheless, the unambiguous trend is the need of very high data transmission rates to cope with the increasing demand of multimedia communication services such as video teleconferencing, network gaming, and high quality audio/video streaming.

The technical challenge is to enable high data transmission rates with a high power and bandwidth efficiency. Power efficiency describes the minimum ratio of signal to noise power required to achieve a desired quality of service. Bandwidth efficiency describes the minimum bandwidth required to transmit at the desired data rate. Further aims in the design of communication systems include manageable computational complexity and low end-to-end delay.

The theoretical limits of data transmission have been derived by Claude Shannon in his landmark paper in 1948 [Sha48]. Shannon showed that information can be transmitted with an arbitrary small probability of error as long as the data transmission rate is below the so-called channel capacity. To approach this capacity, high order modulation and powerful channel coding schemes are required. High order modulation improves the bandwidth efficiency by transmitting several bits per channel use through a signal point of a given signal constellation. Channel coding techniques introduce redundancy to improve the power efficiency. The redundancy

[1] GSM: Global System for Mobile Communications
[2] UMTS: Universal Mobile Telecommunications System
[3] WLAN: Wireless Local Area Network (802.11 standards)
[4] DAB: digital audio broadcast; DVB: digital video broadcast

is used to protect the transmitted data against channel impairments and enable the receiver to correct transmission errors. The breakthrough towards capacity approaching channel codes was the idea of iterative decoding of concatenated codes, the so-called turbo codes by Berrou, Glavieux, and Thitimajshima in 1993 [BGT93]. The turbo decoder consists of the component decoders that exchange soft extrinsic information in an iterative fashion. The idea of iterative decoding has been shown to be valid in a more general sense and the *turbo principle* [Hag97] can be applied at the receiver of a communication system with serial and parallel concatenated components.

Massey proposed in 1974 to jointly design coding and modulation to optimize the performance of digital transmission schemes and introduced the field of coded modulation [Mas74]. Powerful coded modulation schemes include trellis-coded modulation (TCM) [Ung82], multilevel codes (MLC) [IH97] [WFH99] and bit-interleaved coded modulation (BICM) [Zeh92] [CTB98]. BICM is the serial concatenation of a channel code, interleaver and mapper and is used in most recent wireless standards due to its simplicity, flexibility and performance. At the receiver, the signal is consecutively demapped, deinterleaved and decoded. The performance of this standard BICM receiver can be greatly improved through iterative information exchange between the demapper and the decoder according to the turbo principle. This system, introduced in [LR97] [tBSY98a], is usually referred to as BICM with iterative decoding (BICM-ID).

It was soon recognized that the choice of the mapping, i.e. the assignment of bit sequences to signal points of a high order signal constellation, is a crucial design parameter in coded modulation schemes. Probably the most famous mappings are Gray mapping [Gra53] for BICM and Ungerböcks set partitioning mapping [Ung82] for TCM.

The present thesis describes the characterization and optimization of mappings for a wide range of advanced applications. We design mappings for BICM systems with iterative demapping and decoding, for automatic repeat request (ARQ) and multi-antenna (MIMO) systems, irregular (hybrid) mappings for increased system flexibility, non-unique mappings to shape the probability distribution of the transmitted signal and mappings constructed by signal superposition.

The motivation for several approaches within this thesis stem from advances in channel coding. Channel codes are described by their Hamming distance spectrum, we describe mappings by an Euclidean distance spectrum; Channel codes with an optimum Hamming distance spectrum have been proposed, we derive mappings with an optimum Euclidean distance spectrum; Irregular codes have been proposed to optimize the iterative receiver, we propose irregular (hybrid) mappings and signal constellations to optimize the iterative receiver, just to mention three examples. Actually, a fundamental approach within this thesis is to *consider the mapping as a code* that introduces dependencies between the bits mapped to one signal point.

This thesis is organized as follows:

Chapter 2 introduces the general framework and tools that will be used throughout the thesis. This chapter defines the discrete time channel model, derives information theoretical limits on data transmission, reviews channel coding schemes and describes three major coded modulation schemes, namely trellis-coded modulation (TCM), multi-level codes (MLC) and bit-interleaved coded modulation (BICM) without and with iterative demapping and decoding.

Chapter 3 investigates the EXIT chart proposed by ten Brink [tB01b] as a tool to analyze and optimize iterative receivers. In particular EXIT functions of the decoder and the demapper for the BICM system with iterative demapping and decoding are considered. A bit-level and symbol-level analysis is introduced. Furthermore, an analytic and numeric computation of the EXIT functions of the decoder and demapper is proposed that circumvent the need of Monte-Carlo simulations.

In Chapter 4, an Euclidean distance spectrum (EDS) for BICM based systems is proposed to characterize mappings for arbitrary signal constellations. The Euclidean distance spectrum provides an universal framework to analyze mappings for a variety of applications. Then, the Euclidean distance spectrum and a binary switching algorithm are used to find new optimized mappings. Finally, we discuss the use of an additional inner encoder to reduce the error rate at high SNR.

Chapter 5 introduces bit-interleaved coded *irregular* modulation (BICIM), where different signal constellations and mappings may be used within one code word for channel adaptation and to optimize iterative receivers.

Chapter 6 investigates a simple method to shape the probability distribution of the transmitted signal to approach Shannon's channel capacity. The idea is to use non-unique mappings, where multiple bit sequences may be assigned to the same signal point. More bit sequences are assigned to signal points with low energy, less bit sequences to signal points with high energy.

In Chapter 7, signal constellations and bit-to-symbol mappings that arise from the linear superposition of several signal layers are considered. We define a *mapping by superposition* for multi-level codes and bit-interleaved coded modulation. Advantages of this approach are discussed and EXIT charts are used to optimize the power allocation to the superimposed signal layers.

Chapter 8 finally summarizes the results and states possible directions for future research.

Parts of the work presented in this thesis has been published in the following conference proceedings [SGHB03b], [SB04a], [SB04b], [SHGB05], [SB05], [BSS05a], [BSS05b], [SH05], [Sch06], [SSRS06], [SSRS07] and journal papers [SGHB03a], [SB06], [BSHA07].

2

Fundamentals

This Chapter introduces the general framework and specific techniques that will be used throughout this work. We start by defining the considered discrete time channel model in Section 2.1 and the resulting fundamental limitations on the performance in Section 2.2. Optimal decision rules and channel coding techniques are reviewed in Section 2.3 and 2.4, respectively. Finally, we investigate different approaches to combine channel coding and high order modulation in Section 2.5.

2.1 Channel Model

2.1.1 Mobile Radio Channel

The physical mechanisms of electromagnetic wave propagation are diverse. Three basic propagation mechanisms are usually considered for wireless communication systems: reflection from large objects, diffraction at sharp edges, and scattering from rough surfaces [Rap02]. Therefore, the electromagnetic waves may travel from the transmitter to the receiver on the direct line-of-sight path or may reach the receiver after being reflected, diffracted and/or scattered. A strong signal from the line-of-sight path is usually received with satellite or microwave point-to-point links. In mobile communication systems, the line-of-sight path is rarely available. The various propagation paths have different length and cause the corresponding signals to arrive at the receiver with different delays, amplitudes and phases. The different phases yield to situations of constructive and destructive interference of the electromagnetic waves superimposed at the receiver antenna.

Since besides the free space propagation loss, a detailed deterministic computation of the channel characteristics is in general not possible, a variety of statistical propagation models based on empiric observations have been proposed. We distinguish between propagation models that predict the average received signal power (large scale propagation models) and propagation models that predict the variations of the received signal power over small distances or time periods (small scale propagation models).

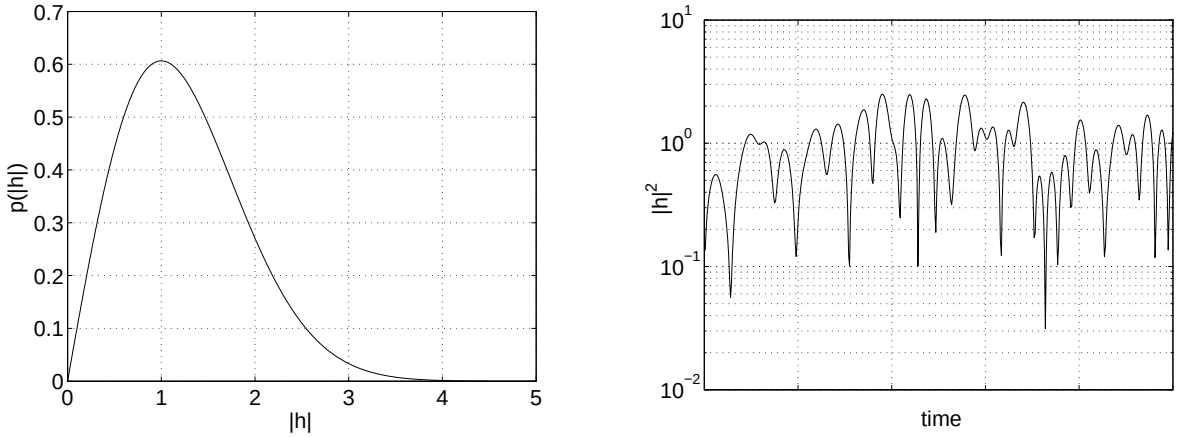
The large scale *path loss* is determined by the distance between transmitter and receiver, the transmitted power, the transmitter and receiver antenna gains, the wavelength, and the propagation environment (line-of-sight path, rural or urban environment).

The small scale *fading* results from multipath propagation. Since the phases of the waves superimposed at the receiver change rapidly with the traveled distance or with moving reflection and scattering objects, the situation may change between constructive and destructive signal superposition within a short time period or traveled distance (in the order of a wavelength).

The multipath channel is assumed to be a bandlimited bandpass channel and can be modeled as a linear filter with time varying impulse response $h_{BP}(t,\tau)$. The variable t represents the time variations and τ the channel multipath delay for a fixed value of t. $h_{BP}(t,\tau)$ may be equivalently described by a complex lowpass impulse response $h_{LP}(t,\tau)$ in the baseband with the relation [Pro01]:

$$h_{BP}(t,\tau) = 2\mathrm{Re}\{h_{LP}(t,\tau)e^{j2\pi f_c t}\}, \tag{2.1}$$

where f_c is the considered carrier frequency. The factor 2 stems from the convenient definition of the Fourier transform of $h_{BP}(t,\tau)$ in [Pro01]. If we assume a large number of propagation paths and that the waves arrive at the receiver with random amplitude and phase distribution, $h_{LP}(t,\tau)$ may be modeled as a complex random variable with uncorrelated and Gaussian distributed real and imaginary part according to the wide-sense stationary uncorrelated scattering (WSSUS) channel model [Bel63]. Without line-of-sight path, $h_{LP}(t,\tau)$ has zero mean and the magnitude $|h_{LP}(t,\tau)|$ follows a Rayleigh distribution shown in Fig. 2.1a). With a line-of-sight path, a real valued constant is added and $|h_{LP}(t,\tau)|$ is Rician distributed with the so-called Rice factor K [Rap02].



a) Rayleigh distribution b) Small scale fading over the time.

Figure 2.1: Qualitative characteristics of the equivalent lowpass channel impulse response $h_{LP}(t,\tau)$.

A common approach to model the time and spectral characteristics of $h_{LP}(t,\tau)$ is to assume that the phases and angles of arrival of the received signals are uniformly distributed, the average amplitudes of the received signals are equal (no line-of-sight path), and the Doppler shift of the received signals is set according to the motion of the receiver [Rap02]. Then, a typical time characteristic of the Rayleigh distributed $|h_{LP}(t,\tau)|$ is illustrated in Fig. 2.1b) and its power density spectrum is the so-called Jakes spectrum.

A useful parameter in this context is the coherence time T_{coh} defined as the time for a mobile receiver required to travel a wavelength λ at velocity v:

$$T_{\text{coh}} = \frac{\lambda}{v}. \tag{2.2}$$

T_{coh} is a measure of the time duration over which the channel impulse response is essentially invariant. The Doppler shift is the change in frequency due to the motion of the mobile. The maximum Doppler shift $f_{d,\text{max}}$ and the coherence time are inversely proportional to each other:

$$f_{d,\text{max}} = \frac{1}{T_{\text{coh}}} = \frac{v}{\lambda}. \tag{2.3}$$

A further important parameter of a multipath channel is the time delay at the receiver between the waves traveling along the different paths. The inverse of the time delays is proportional to the bandwidth of the channel. The coherence bandwidth B_{coh} is a measure of the bandwidth over which the channel has a constant gain and phase shift. If the bandwidth B of the transmitted signal is larger than the coherence bandwidth B_{coh}, the received signal is distorted and we have a frequency selective channel. Otherwise, the channel is frequency non-selective or flat.

Independent of the impairments of the mobile radio channel, the received signal contains additive noise due to the thermal noise of the receiver front end and interference from other users. We will not consider the interference in this work.

2.1.2 Discrete Time Channel Model

The standard discrete time channel model is introduced. This model comprises pulse shaping, modulation to carrier frequency, physical transmission, and at the receiver perfect time and frequency synchronization, demodulation, matched filtering, and sampling, as shown in Fig. 2.2. A few basics and the main assumptions are highlighted in the following.

The complex data symbols x_n from the symbol alphabet $\mathcal{X}$ are processed at the transmitter as shown in Fig. 2.2. The pulse shaping filter with impulse response $g_T(t)$ together with the modulation interval T determine the spectral characteristics of the transmitted signal that have to match the restrictions imposed by the communication system and the given communication channel.

The pulse shaped signals are modulated by a carrier with frequency f_c. Quadrature amplitude modulation (QAM) uses the in-phase and quadrature carriers $\cos(2\pi f_c t)$ and $-\sin(2\pi f_c t)$, respectively. The transmitted bandpass signal is given by:

$$\begin{aligned} x_{BP}(t) &= \cos(2\pi f_c t) \cdot \sum_{n=1}^{N_s} \text{Re}\{x_n\} \cdot g_T(t - nT) - \sin(2\pi f_c t) \cdot \sum_{n=1}^{N_s} \text{Im}\{x_n\} \cdot g_T(t - nT) \\ &= \text{Re}\{x_{LP}(t) e^{j2\pi f_c t}\}, \end{aligned} \tag{2.4}$$

with the equivalent complex lowpass signal

$$x_{LP}(t) = \sum_{n=1}^{N_s} x_n \cdot g_T(t - nT), \tag{2.5}$$

and the complex symbol x_n at time nT, $n = 1, \ldots, N_s$. The signal $x_{BP}(t)$ is transmitted over a channel modeled by a linear system with impulse response $h_{BP}(t, \tau)$ and additive white Gaussian noise (AWGN) $n_{BP}(t)$. The received signal is the convolution of $x_{BP}(t)$ and $h_{BP}(t, \tau)$:

$$y_{BP}(t) = \int_{\tau=t-\tau_{\max}}^{t} x_{BP}(\tau) \cdot h_{BP}(t, t-\tau)\, \mathrm{d}\tau + n_{BP}(t) = x_{BP}(t) \star h_{BP}(t, \tau) + n_{BP}(t), \quad (2.6)$$

where $\tau_{\max}$ is the maximum delay experienced by any of the arriving signal components.

The AWGN $n_{BP}(t)$ is defined to have a constant power spectral density $N_0/2$ over the entire frequency range. It is mathematically convenient to postulate that the signals and noise at the receiver have passed through an ideal bandpass filter with a bandwidth B large enough to not distort the signals. The equivalent lowpass noise $n_{LP}(t)$ has then a power spectral density N_0 for $|f| \leq \frac{1}{2}B$ and 0 otherwise.

At the receiver, the demodulation is done by multiplying the received signal with the two quadrature carriers $\cos(2\pi f_c t)$ and $-\sin(2\pi f_c t)$. Signal parts at twice the carrier frequency are removed by a low pass filter and we obtain the lowpass in-phase and quadrature signal components. The relation in equation (2.6) can be directly applied to the equivalent lowpass signals [Pro01]:

$$y_{LP}(t) = x_{LP}(t) \star h_{LP}(t, \tau) + n_{LP}(t) \quad \text{with} \quad y_{BP}(t) = \mathrm{Re}\{y_{LP}(t) e^{j2\pi f_c t}\}. \quad (2.7)$$

We omit in the following the subscript LP since only equivalent lowpass signals will be considered in this work.

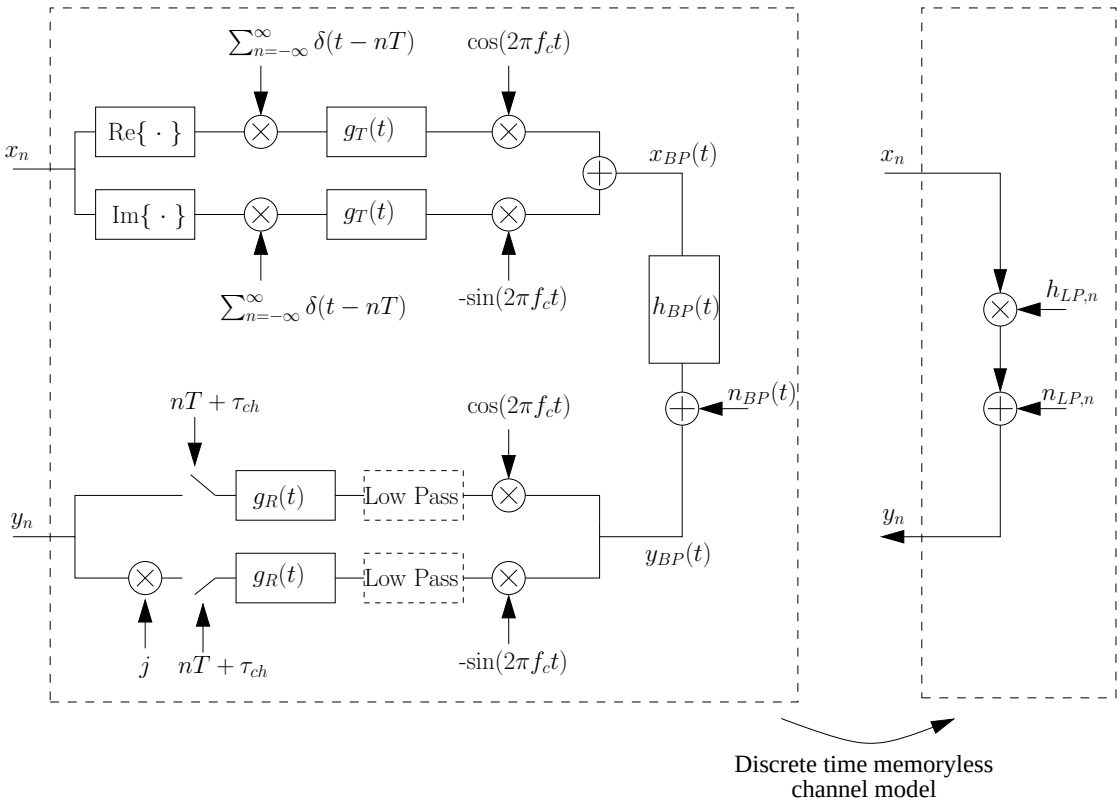


Figure 2.2: Bandpass and discrete time baseband channel model.

The receiver filter $g_R(t) = c \cdot g_T^*(T - t)$ with a scaling factor c is matched to the pulse shaping filter at the transmitter to maximize the signal-to-noise ratio. The output of the matched filter is sampled at time $nT + \tau_{ch}$, $n = 1, \ldots, N_s$, to obtain the signal y_n from the alphabet $\mathcal{Y}$ of infinite size. The delay τ_{ch} from the channel is neglected in the following.

We arrive at the time discrete channel model used in this work that describes the relation between the transmitted time discrete samples x_n and the received samples y_n. We consider only frequency non-selective channels where the channel is memoryless: The received signal y_n depends only on the values of the transmitted signal x_n, the channel coefficient h_n and the noise n_n at the same time nT. Then we have:

$$y_n = h_n \cdot x_n + n_n. \tag{2.8}$$

The channel model is adapted such that $E_s = \mathrm{E}\{|x_n|^2\}$ is the average energy per symbol, $P_s = E_s/T$ is the average power of the symbols and $\mathrm{E}\{|n_n|^2\} = N_0$ is the variance σ_n^2 of the complex noise after matched filtering and sampling. Then, both the real and the imaginary part of n_n have zero mean and variance $N_0/2$.

The channel coefficients are normalized to $\mathrm{E}\{|h_n|^2\} = 1$ such that $10 \cdot \log_{10}(E_s/N_0)$ is the signal-to-noise ratio (SNR) at the receiver in dB. By setting the fading coefficient $h_n = 1, \forall n$, in (2.8), we obtain the AWGN channel model. To simulate a mobile environment, the values h_n should be samples of a curve as given in Fig. 2.1b). For the design and simulation of communication systems, the two marginal cases of block fading and symbol fading are often considered for convenience, where h_n changes independently after each transmitted data block or symbol, respectively. The fading coefficients are complex Gaussian distributed and both the real and the imaginary part of h_n have zero mean and variance $1/2$; $|h_n|$ is then Rayleigh distributed.

We focus on the scenario where the samples h_n and σ_n^2 are ideally known for all n at the receiver, i.e. that the receiver has ideal channel state information.

Fig. 2.3 summarizes the main parameters of the discrete time memoryless channel. The channel input is from the alphabet $\mathcal{X}$, the channel output from the alphabet $\mathcal{Y}$. The probability distribution of x and y is given by $p(x)$ and $p(y)$, respectively. The channel is further specified by the transition probabilities $p(y|x)$. If the alphabets are of finite size, we have the probabilities $P(x)$, $P(y)$ and $P(y|x)$.

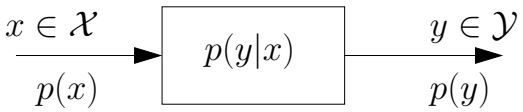
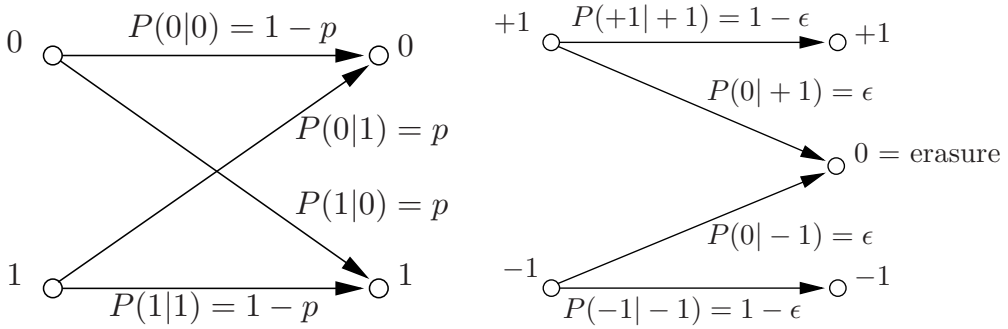


Figure 2.3: Discrete time memoryless channel.

With an AWGN or fading channel model as described in equation (2.8), $\mathcal{X}$ is often assumed continuous for theoretical investigations. In practice, the set $\mathcal{X}$ comprises discrete symbols of a signal constellation. The channel output alphabet $\mathcal{Y}$ has a continuous distribution. For real valued signals, the channel transition probability is given by the Gaussian distribution:

$$p(y|x,h) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \, \mathrm{e}^{-|y-h\cdot x|^2/2\sigma_n^2}. \tag{2.9}$$

Two other channel models will be of importance in this work: The binary symmetric channel (BSC) has a binary channel input and output alphabet $\mathcal{X} = \mathcal{Y} = \{0; 1\}$. The channel transition probabilities $P(y|x)$ for an error probability p are given in Fig. 2.4a). The binary erasure channel (BEC) has two possible channel input values $\mathcal{X} = \{+1; -1\}$ and three output values $\mathcal{Y} = \{+1; 0; -1\}$. With the erasure probability ϵ, the channel transition probabilities are given in Fig. 2.4b).



a) Binary symmetric channel (BSC) b) Binary erasure channel (BEC)

Figure 2.4: BSC and BEC channel models.

2.2 Information Theory

Information theory uses the entropy as a statistical measure to determine the amount of information contained in a given amount of data. In this work, information theoretic results are applied to analyze and derive performance limits for the investigated communication systems. We refer to [CT91] [Gal68] for a detailed overview about this topic.

2.2.1 Entropy and Mutual Information

The notions of entropy, differential entropy, conditional entropy and mutual information are briefly reviewed. Let X be a random variable from the alphabet $\mathcal{X}$ with realizations x and let Y be a random variable from the alphabet $\mathcal{Y}$ with realizations y. The entropy of X for a finite size alphabet $\mathcal{X}$ is defined as follows:

$$H(X) = \sum_{x \in \mathcal{X}} P(x) \log_2 \frac{1}{P(x)} = -\mathrm{E}\{\log_2 P(x)\}. \tag{2.10}$$

With the logarithm to the base of 2, the entropy can be thought of as the number of bits which are on average necessary to represent a realization x of X, or as the amount of information contained in x. The entropy $H(X)$ is maximized if all values of X are equally likely:

$$\max_{P(X)} \{H(X)\} = \log_2(|\mathcal{X}|). \tag{2.11}$$

For an alphabet $\mathcal{X}$ of infinite size, the so-called differential entropy is given by:

$$h(X) = \int_{\mathcal{X}} p(x) \log_2 \frac{1}{p(x)} \, \mathrm{d}x = -\mathrm{E}\{\log_2 p(x)\}. \tag{2.12}$$

For a real valued X, the differential entropy is bounded by

$$h(X) \leq \frac{1}{2} \log_2(2\pi e \mathrm{Var}\{X\}), \tag{2.13}$$

and maximized when X is a zero mean and Gaussian distributed variable with variance σ_x^2: $X \sim \mathcal{N}_{\mathbb{R}}(0, \sigma_x^2)$. Then, the maximum differential entropy is

$$\max_{p(X)}\{h(X)\} = \frac{1}{2}\log_2(2\pi e\sigma_x^2). \tag{2.14}$$

For a circularly symmetric complex Gaussian distributed variable X with $X \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_x^2)$, the variance of the real and imaginary part is $\sigma_x^2/2$. Then, the differential entropy is equal to the sum of the differential entropies of the real and imaginary part. With equation (2.14) we have:

$$\max_{p(X)}\{h(X)\} = \log_2(\pi e\sigma_x^2). \tag{2.15}$$

The conditional entropies of X given Y are defined as follows:

$$\begin{aligned} H(Y|X) &= \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P(y,x)\log_2\frac{1}{P(y|x)} = -\mathrm{E}_{XY}\{\log_2 P(y|x)\}, \\ h(Y|X) &= \int_{\mathcal{Y}}\int_{\mathcal{X}} p(y,x)\log_2\frac{1}{p(y|x)}\,\mathrm{d}x\,\mathrm{d}y = -\mathrm{E}_{XY}\{\log_2 p(y|x)\}. \end{aligned} \tag{2.16}$$

If X and Y are independent, the amount of information $H(Y|X)$ contained in the realization y is equal to $H(Y)$, since x does not have any information about y. On the other hand, the conditional entropy $H(Y|X)$ is zero if $X = Y$ holds.

The mutual information (MI) between X and Y is defined as:

$$\begin{aligned} I(X;Y) &= \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P(y,x)\log_2\frac{P(y,x)}{P(y)P(x)} = \mathrm{E}_{XY}\left\{\log_2\frac{P(y,x)}{P(y)P(x)}\right\}, \\ I(X;Y) &= \int_{\mathcal{Y}}\int_{\mathcal{X}} p(y,x)\log_2\frac{p(y,x)}{p(y)p(x)}\,\mathrm{d}x\,\mathrm{d}y = \mathrm{E}_{XY}\left\{\log_2\frac{p(y,x)}{p(y)p(x)}\right\}. \end{aligned} \tag{2.17}$$

The mutual information $I(X;Y)$ is the amount of information y provides about x and vice versa. If X and Y are independent, the mutual information is zero. $I(X;Y)$ is upper bounded by the minimum of $H(X)$ and $H(Y)$. The mutual information is related to the entropy and conditional entropy (and in a similar way to the differential entropy) as follows:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{2.18}$$

Note that in this work, we will often compute the conditional entropy and mutual information for an alphabet $\mathcal{X}$ of finite size and an alphabet $\mathcal{Y}$ of infinite size in equations (2.16) and (2.17).

2.2.2 Limits on Data Transmission

Shannon's famous channel coding theorem states that there is a fundamental limit on the data transmission rate R_t for a given channel [Sha48]. R_t has the unit of information bit per symbol and is the amount of information measured in bit contained in a transmitted channel input symbol $x \in \mathcal{X}$. The channel coding theorem states that R_t is upper bounded by the channel capacity: $R_t < \mathcal{C}$ [Sha48]. For a memoryless channel, the channel capacity $\mathcal{C}$ is equal to the maximum mutual information over all input distributions $p(X)$ between the channel input X and channel output Y:

$$\mathcal{C} = \max_{p(x)} I(X;Y). \tag{2.19}$$

We use the term capacity in a wide sense and investigate the maximum achievable data transmission rate for different constraints. We consider the constraints of a maximum transmitted power (*power constrained capacity*), a finite size channel input alphabet (*constellation constrained capacity*), different probability distributions of the channel input alphabet, and different receiver architectures and components (*receiver constrained capacity*).

Power constrained capacity

If no constraint on the power of the channel input signal X is set, there is no capacity limit and $\mathcal{C} \to \infty$. Therefore, we always consider the capacity for a limited power. The power constrained capacity $\mathcal{C}_{\mathbb{R}}$ of a real-valued AWGN channel $Y = X + N$ with $N \sim \mathcal{N}_{\mathbb{R}}(0, \sigma_n^2)$ is obtained from equation (2.19) by determining the distribution $p(x)$ that maximizes $I(X;Y)$. Since N and X are independent, with equation (2.14) and a zero mean X with variance σ_x^2 we write:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) = H(Y) - H(X+N|X) = H(Y) - H(N) \\ &\leq \frac{1}{2}\log_2(2\pi e \mathrm{Var}\{Y\}) - \frac{1}{2}\log_2(2\pi e \sigma_n^2) \\ &= \frac{1}{2}\log_2(2\pi e(\sigma_x^2 + \sigma_n^2)) - \frac{1}{2}\log_2(2\pi e \sigma_n^2) \\ &= \frac{1}{2}\log_2\left(1 + \frac{\sigma_x^2}{\sigma_n^2}\right) = \frac{1}{2}\log_2\left(1 + \frac{2E_s}{N_0}\right) = \mathcal{C}_{\mathbb{R}}. \end{aligned} \tag{2.20}$$

We have equality if the received signal Y is Gaussian distributed with $Y \sim \mathcal{N}_{\mathbb{R}}(0, \sigma_x^2 + \sigma_n^2)$. With a zero mean Gaussian noise N, this implies that the pdf of the transmitted signal should be Gaussian too: $X \sim \mathcal{N}_{\mathbb{R}}(0, \sigma_x^2)$. The units of the capacity given in equation (2.20) are bits per channel use. Since the discrete time channel can be used any $T = 1/(2B)$ seconds with an appropriate pulse shaping filter, B is the bandwidth, the capacity $\mathcal{C}'_{\mathbb{R}}$ in bits per second is

$$\mathcal{C}'_{\mathbb{R}} = B\log_2\left(1 + \frac{2E_s}{N_0}\right). \tag{2.21}$$

The complex AWGN channel with $X \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_x^2)$ and $N \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_n^2)$ can be considered as two independent, parallel, real AWGN channels due to the circular symmetry of the complex channel noise. Then, the capacity in bits per channel use is computed with equation (2.15) to:

$$\mathcal{C}_{\mathbb{C}} = \log_2\left(1 + \frac{\sigma_x^2}{\sigma_n^2}\right) = \log_2\left(1 + \frac{E_s}{N_0}\right). \tag{2.22}$$

The power constrained capacities of a real and complex AWGN channel are shown in Fig. 2.6. The figure illustrates that we aim at operating close to the channel capacity to either reduce the transmitted power at a given data rate or increase the data rate at a given transmitted power.

A symbol wise flat fading channel can be interpreted as an AWGN channel with a transmitted power attenuated or amplified by the fading value h. If h is known at the receiver, the capacity is obtained by averaging over the fading states:

$$\mathcal{C}_{\mathbb{R},f} = \mathrm{E}_h\left\{\frac{1}{2}\log_2\left(1 + \frac{h \cdot \sigma_x^2}{\sigma_n^2}\right)\right\} \quad \text{and} \quad \mathcal{C}_{\mathbb{C},f} = \mathrm{E}_h\left\{\log_2\left(1 + \frac{h \cdot \sigma_x^2}{\sigma_n^2}\right)\right\}, \tag{2.23}$$

for real and complex channel inputs, respectively.

The capacity of a BSC channel with error probability p is [CT91]

$$\mathcal{C} = \max_{p(x)} H(Y) - H(p) = 1 - H(p), \tag{2.24}$$

with uniformly distributed binary random variables X and Y. The capacity of a BEC channel with erasure probability ϵ is likewise achieved with a uniformly distributed X [CT91]:

$$\mathcal{C} = \max_{p(x)} H(Y) - H(\epsilon) = 1 - \epsilon. \tag{2.25}$$

Constellation constrained capacity

An infinite size channel input alphabet $\mathcal{X}$ is not of practical relevance for digital communications. Therefore, we constrain the input alphabet to signal points x taken from a signal alphabet $\mathcal{X}$ of finite size $|\mathcal{X}|$; M bits are associated to each signal point from the signal constellation of size $|\mathcal{X}| = 2^M$. The placement of the signal points in the real or complex plane strongly affects the capacity. For the moment, we apply the additional constraint of a uniform probability distribution of the signal points. The optimization of the signal point probabilities is investigated in Chapter 6.

A combined phase and amplitude modulation with equally probable signal points on concentric rings was proposed in [Cah60]. Based on this work, a quadrature amplitude modulation (QAM) signal constellation was first introduced in [CG62]. Foschini *et al.* investigated in [FGW74] signal constellations optimized for Gaussian channels. Recent wireless standards use standard signal constellations including amplitude shift keying (ASK), phase shift keying (PSK) and QAM schemes. Fig. 2.5 depicts some common signal alphabets. As example, the GSM extension EDGE (enhanced data for GSM evolution [3GPP01]) has the option to use 8PSK modulation, the UMTS extension HSDPA (high speed downlink packet access [3GPP06a]) allows 16QAM for high data rates, IEEE 802.11 WLAN [IEEE99] (wireless local area network) and IEEE 802.16 WiMAX [IEEE04] (worldwide interoperability for microwave access) standards allow up to 64QAM signal constellations.

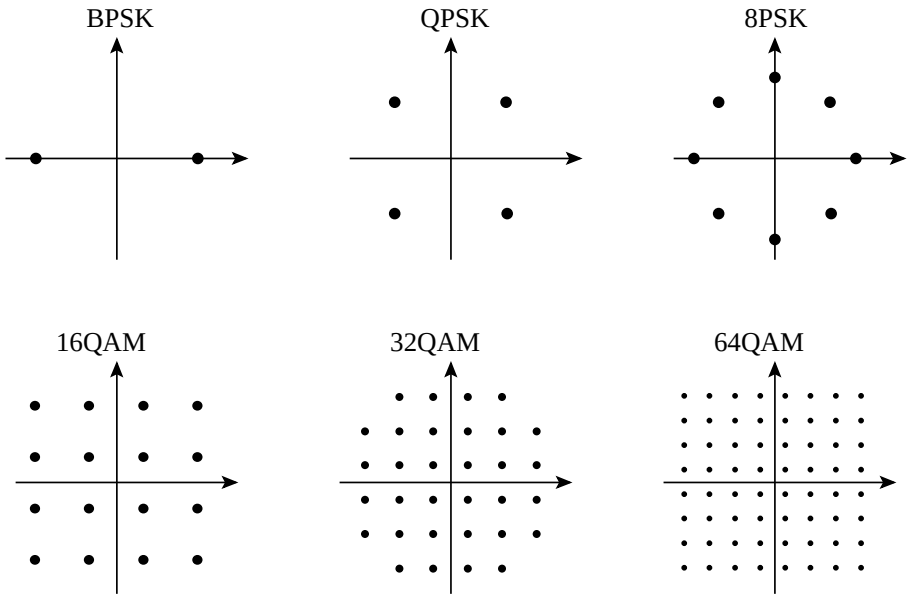


Figure 2.5: Signal constellations.

The constellation constrained capacity with equally probable signal points is obtained directly from equation (2.17). For the signal constellation $\mathcal{X}$ we write using Bayes' rule:

$$\mathcal{C} = I(X;Y) = E_{X,Y}\left\{\log_2 \frac{p(y|x)}{p(y)}\right\}. \tag{2.26}$$

$p(y|x)$ is the channel transition probability and

$$p(y) = \sum_{x\in\mathcal{X}} P(x)\cdot p(y|x). \tag{2.27}$$

Fig. 2.6 depicts the numerically obtained constellation constrained capacity with an AWGN channel and different signal constellations $\mathcal{X}$. The important fact to note is that if we want to achieve a high bandwidth efficiency at large SNR, it is essential to apply high order modulation schemes with a large number of signal points.

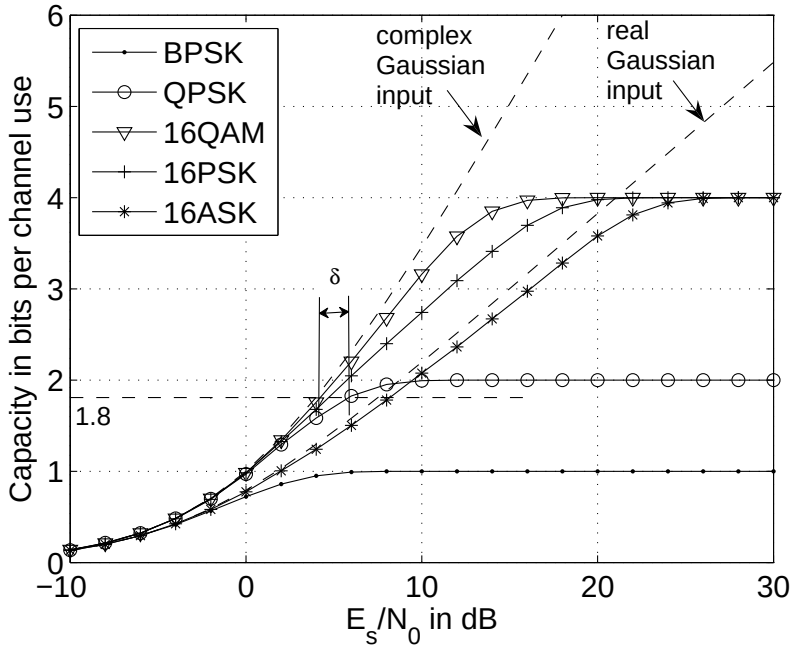


Figure 2.6: Capacity with AWGN channel and different channel input alphabets $\mathcal{X}$; δ is the gain at $R_t = 1.8$ bits per channel use for using 16QAM instead of QPSK modulation.

Receiver constrained capacity

The considered channel can be defined to include receiver elements, i.e. Y is not the physical channel output but the output of e.g. a detector at the receiver. If the detector does not use the optimal decision rules described in the next Section 2.3, information is lost and the receiver constrained capacity is reduced. An example is the receiver constrained capacity of bit-interleaved coded modulation (BICM) without iterative demapping and decoding shown in Fig. 2.21.

2.3 Optimal Decision Rules

Optimal decision rules for the receiver in a communication system are discussed in this Section. In general, the receiver has the observation vector $\mathbf{y}$ and would like to find the corresponding most probable bit sequence $\hat{\mathbf{v}}$ or specific bit value $\hat{v}_i$ at time index i. Then, the sequence error rate or the bit error rate is minimized, respectively.

The receiver has to exploit the dependencies between the elements in $\mathbf{y}$ and $\mathbf{v}$ to output the optimal solution. Table 2.1 summarizes for some wide spread applications the values of $\mathbf{y}$ and $\mathbf{v}$ that should be considered for optimal decoding and detection.

Application	Observation vector $\mathbf{y}$	Estimated bit vector $\hat{\mathbf{v}}$
Channel decoder	N coded bits	K information bits
Demapper, 2^M-ary signal constellation	1 complex symbol	M bits
Equalizer, channel with L taps and terminating sequence	$N + L$ complex symbols	$M \cdot N$ bits
MIMO detection, n_R receive antennas, rate-1 space time mapper	n_R complex symbols	$M \cdot n_R$ bits
CDMA multiuser detection, K users, spreading code of length R	R complex symbols	$M \cdot K$ bits

Table 2.1: Values of the observation vector $\mathbf{y}$ and estimated bit vector $\hat{\mathbf{v}}$ for optimal detection and for different applications.

The most probable sequence $\hat{\mathbf{v}}$ is in general obtained by

$$\hat{\mathbf{v}} = \underset{\forall \mathbf{v}}{\text{argmax}}\ P(\mathbf{v}|\mathbf{y}) = \underset{\forall \mathbf{v}}{\text{argmax}}\ P(\mathbf{y}|\mathbf{v})P(\mathbf{v}), \tag{2.28}$$

assuming that the probability $p(\mathbf{y})$ is independent of the investigated $\mathbf{v}$. The bit error probability is minimized if the bit estimate $\hat{v}_i$ is computed with symbol-by-symbol decoding:

$$\hat{v}_i = \underset{v \in \{0,1\}}{\text{argmax}}\ P(v_i = v|\mathbf{y}) = \underset{v \in \{0,1\}}{\text{argmax}} \sum_{\forall \mathbf{v}: v_i = v} P(\mathbf{y}|\mathbf{v})P(\mathbf{v}). \tag{2.29}$$

In this work, we focus on symbol-by-symbol decision rules.

The probabilities $P(\mathbf{v}|\mathbf{y})$ and $P(v_i|\mathbf{y})$ are called *a posteriori probabilities* (APP) and $P(\mathbf{v})$ *a priori probabilities* in this context. Therefore, the algorithms implementing the equations above perform *maximum a posteriori probability* (MAP) decoding. If the a priori probabilities are not considered, we have *maximum likelihood* (ML) decoding.

We distinguish between hard and soft decision decoding. With hard decision decoding, the size of the considered channel output alphabet $\mathcal{Y}$ is equal to the size of the input alphabet $\mathcal{X}$. With soft decision decoding, the channel state information is incorporated in the decoding process. Then, the channel output alphabet is larger than the channel input alphabet and the reliability of the received values is taken into account to improve the performance. In this work, we focus on soft decision decoding.

Log-Likelihood Algebra

Instead of probabilities, it is often convenient to use log-likelihood ratios (LLRs) defined as [HOP93]

$$L(v_i) = \log \frac{P(v_i = 0)}{P(v_i = 1)}. \tag{2.30}$$

The sign of $L(v_i)$ corresponds to its hard decided binary value and the magnitude $|L(v_i)|$ represents a measure for the reliability of the hard decision. The bit probabilities are obtained from the LLRs by evaluating

$$P(v_i) = \frac{1}{1 + \mathrm{e}^{-L(v_i)}} \cdot \mathrm{e}^{-v_i \cdot L(v_i)} \tag{2.31}$$

or

$$P(v_i) = \frac{\mathrm{e}^{-L(v_i)/2}}{1 + \mathrm{e}^{-L(v_i)}} \cdot \mathrm{e}^{\tilde{v}_i \cdot L(v_i)/2} \quad \text{with} \quad \tilde{v}_i = \begin{cases} +1, & v_i = 0, \\ -1, & v_i = 1. \end{cases} \tag{2.32}$$

The so-called soft-bit is the expected value of the bit v_i and is given by

$$\mathrm{E}\{v_i\} = \frac{\mathrm{e}^{L(v_i)} - 1}{\mathrm{e}^{L(v_i)} + 1} = \tanh\left(\frac{L(v_i)}{2}\right). \tag{2.33}$$

The error probability of the hard decisions can be evaluated from the magnitude $|L(\hat{v}_i)|$ as follows:

$$P(\hat{v}_i \neq v_i) = \frac{1}{1 + \mathrm{e}^{|L(\hat{v}_i)|}}. \tag{2.34}$$

Optimal Decision Rules using LLRs

For symbol-by-symbol decoding, the optimum decision principle using LLRs can be written as

$$\hat{v}_i = \begin{cases} 0, & L(\hat{v}_i) \geq 0, \\ 1, & L(\hat{v}_i) < 0. \end{cases} \tag{2.35}$$

The LLR $L(\hat{v}_i)$ for a given channel observation $\mathbf{y}$ is computed as follows:

$$L(\hat{v}_i) = L(v_i|\mathbf{y}) = \log \frac{P(v_i = 0|\mathbf{y})}{P(v_i = 1|\mathbf{y})} = \log \frac{\sum_{\forall \mathbf{v}: v_i = 0} P(\mathbf{y}|\mathbf{v}) P(\mathbf{v})}{\sum_{\forall \mathbf{v}: v_i = 1} P(\mathbf{y}|\mathbf{v}) P(\mathbf{v})}. \tag{2.36}$$

It is convenient to define a metric $\Lambda(\mathbf{v})$ to write

$$L(\hat{v}_i) = \log \frac{\sum_{\forall \mathbf{v}: v_i = 0} \mathrm{e}^{\Lambda(\mathbf{v})}}{\sum_{\forall \mathbf{v}: v_i = 1} \mathrm{e}^{\Lambda(\mathbf{v})}} \quad \text{with} \quad \Lambda(\mathbf{v}) = \log P(\mathbf{y}|\mathbf{v}) P(\mathbf{v}). \tag{2.37}$$

A low complexity approximation of the MAP estimation is the max-log MAP approach. Using the approximation $\log(\sum_j \mathrm{e}^{x_j}) \approx \max_j(x_j)$, we have:

$$L(\hat{v}_i) \approx \max_{\forall \mathbf{v}: v_i = 0} (\Lambda(\mathbf{v})) - \max_{\forall \mathbf{v}: v_i = 1} (\Lambda(\mathbf{v})). \tag{2.38}$$

We obtain the optimum log-MAP by adding a correction term as described in [RVH95]. This correction term can be efficiently pre-computed and stored in a look-up table.

Channel Decoder

As described in detail in the next Section 2.4, the channel encoder adds redundancy to protect the information bits against channel impairments by mapping a length K information bit sequence to a length N code word, $N \geq K$. Consider a memoryless channel and binary BPSK modulation where the code bits $c_n \in \{0;1\}$ are mapped to the transmitted symbols $x_n \in \{\pm 1\}$ according to the mapping rule $0 \rightarrow +1$ and $1 \rightarrow -1$. The channel decoder uses the channel observations $\mathbf{y} = (y_1, \ldots, y_{N_s})$ corresponding to the code bits $\mathbf{c} = (c_1, \ldots, c_N)$, $N_s = N$ for BPSK, to obtain estimates about the information bits $\mathbf{u} = (u_1, \ldots, u_K)$. The LLR for the information bit estimate $\hat{u}_i$ is:

$$L(\hat{u}_i) = \log \frac{\sum_{\forall \mathbf{u}: u_i = 0} \mathrm{e}^{\Lambda(\mathbf{u})}}{\sum_{\forall \mathbf{u}: u_i = 1} \mathrm{e}^{\Lambda(\mathbf{u})}}, \tag{2.39}$$

with the metric

$$\Lambda(\mathbf{u}) = \sum_{n=1}^{N} \log\ p(y_n|x_n) + \sum_{j=1}^{K} \log P(u_j). \tag{2.40}$$

Using equation (2.31) and omitting the terms that are independent of the investigated sequence $\mathbf{u}$, we rewrite the metric using LLRs:

$$\Lambda(\mathbf{u}) = -\sum_{n=1}^{N} c_n L(y_n|x_n) - \sum_{j=1}^{K} u_j L(u_j). \tag{2.41}$$

For iterative decoding, it is of interest to separate the information into a priori, channel and extrinsic information. For a systematic code where $c_i = u_i$ for $i = 1, \ldots, K$, we have:

$$\begin{aligned} L(\hat{u}_i) &= \log \frac{\sum_{\forall \mathbf{u}: u_i = 0} (\prod_{n=1}^{N} p(y_n|x_n))(\prod_{j=1}^{K} P(u_j))}{\sum_{\forall \mathbf{u}: u_i = 1} (\prod_{n=1}^{N} p(y_n|x_n))(\prod_{j=1}^{K} P(u_j))} \\ &= \log \frac{\sum_{\forall \mathbf{u}: u_i = 0} (\prod_{n=1:n\neq i}^{N} p(y_n|x_n))(\prod_{j=1:j\neq i}^{K} P(u_j))}{\sum_{\forall \mathbf{u}: u_i = 1} (\prod_{n=1:n\neq i}^{N} p(y_n|x_n))(\prod_{j=1:j\neq i}^{K} P(u_j))} + \\ &\qquad \log \frac{p(y_i|x_i = +1)}{p(y_i|x_i = -1)} + \log \frac{P(u_i = 0)}{P(u_i = 1)} \\ &= \underbrace{L_e(u_i)}_{\text{extrinsic inform.}} + \underbrace{L_c(u_i)}_{\text{channel inform.}} + \underbrace{L_a(u_i)}_{\text{a priori inform.}} \end{aligned} \tag{2.42}$$

Demapper

The demapper uses the received value y_n of the transmitted symbol x_n from a 2^M-ary signal constellation to obtain estimates about the corresponding bits $c_{n,m}$, $m = 1, \ldots, M$. The LLR for the bit estimate $\hat{c}_{n,m}$ is

$$L(\hat{c}_{n,m}) = \log \frac{\sum_{\forall \mathbf{c}_n : c_{n,m}=0} \mathrm{e}^{\Lambda(\mathbf{c}_n)}}{\sum_{\forall \mathbf{c}_n : c_{n,m}=1} \mathrm{e}^{\Lambda(\mathbf{c}_n)}}, \tag{2.43}$$

with the metric

$$\Lambda(\mathbf{c}_n) = \log p(y_n|x_n) + \sum_{m=1}^{M} \log P(c_{n,m}), \tag{2.44}$$

where the bit sequence $\mathbf{c}_n$ is mapped to the symbol $x_n = \mu(\mathbf{c}_n)$ according to the bit-to-symbol mapping μ. Using equation (2.9) for the channel transition probability $p(y_n|x_n)$, equation (2.31) and omitting the terms that are independent of the investigated sequence $\mathbf{c}_n$, we rewrite the metric using LLRs:

$$\Lambda(\mathbf{c}_n) = -\frac{|y_n - h_n x_n|^2}{2\sigma_n^2} - \sum_{m=1}^{M} c_{n,m} \cdot L(c_{n,m}). \tag{2.45}$$

To separate the information into extrinsic and a priori information, we rewrite:

$$L(\hat{c}_{n,m}) = \log \frac{\sum_{\forall \mathbf{c}_n : c_{n,m}=0} p(y_n|x_n) \prod_{j=1}^{M} P(c_{n,j}))}{\sum_{\forall \mathbf{c}_n : c_{n,m}=1} p(y_n|x_n) \prod_{j=1}^{M} P(c_{n,j}))} \tag{2.46}$$

$$= \log \frac{\sum_{\forall \mathbf{c}_n : c_{n,m}=0} p(y_n|x_n) \prod_{j=1:j\neq m}^{M} P(c_{n,j}))}{\sum_{\forall \mathbf{c}_n : c_{n,m}=1} p(y_n|x_n) \prod_{j=1:j\neq m}^{M} P(c_{n,j}))} + \log \frac{P(c_{n,m}=0)}{P(c_{n,m}=1)}. \tag{2.47}$$

$$= \underbrace{L_e(c_{n,m})}_{\text{extrinsic inform.}} + \underbrace{L_a(c_{n,m})}_{\text{a priori inform.}} \tag{2.48}$$

2.4 Error Control Coding

When data is transmitted over unreliable channels, error-free reception of the transmitted symbols is not guaranteed. To detect and correct symbols falsified by the channel, we should introduce redundancy in the transmitted data using forward error correction (FEC) techniques. We proceed as follows: The data stream of information bits is partitioned into information words of fixed length K. Each information word is mapped to a code word of length N. Since redundancy is added, we have $N \geq K$ and out of the 2^N possible bit sequences of length N, there are only 2^K valid code words. The set $\mathcal{C}$ of valid code words is called channel code. The code rate is defined by

$$R = \frac{K}{N}. \tag{2.49}$$

The coding redundancy is $1 - R$. In combination with a signal constellation with M bits associated to each symbol, the data transmission rate is $R_t = R \cdot M$. Then, the energy per information bit is given by

$$E_b = E_s/(R \cdot M). \tag{2.50}$$

We consider first BPSK modulation with $M = 1$ and $R_t = R$. The combination of channel coding and high order modulation is investigated in Section 2.5.

We reviewed the channel coding theorem in Section 2.2 that states, that arbitrary small transmission error rates are possible for a large code word length N as long as the data transmission rate R_t is smaller than the capacity $\mathcal{C}$. However, this theorem does not say anything about how to design a set of code words $\mathcal{C}$ that achieves this performance with feasible encoding and decoding complexity. The code $\mathcal{C}$ can be constructed by randomly selecting 2^K binary codewords out of the 2^N possible length N bit sequences. This approach yields a so-called random code. However, the complexity and storage requirements of a random code grow exponentially with the information word length and are impractical for real communication systems.

The aim of forward error coding (FEC) techniques is to introduce structures in the code that enable a good trade-off between feasible complexity and strong error protection. Other aims include good performance for small and medium block length, efficient implementation possibilities, and high adaptability through a variable code rate support.

We refer to [LC04] [JZ99] [Fri95] [Bos98] for an extensive coverage of FEC techniques. In this thesis, we focus on convolutional codes and concatenations of these codes.

2.4.1 Convolutional Codes

Convolutional codes were introduced by Elias [Eli55] and are now widely used in wireless communications. These codes are highly structured, allowing a simple implementation and a good performance with short block length, but they are still far away from reaching the capacity limit predicted by Shannon.

Convolutional codes are a specific class of binary linear codes. A code is linear if the sum $\mathbf{c}+\mathbf{c}'$ of any two length N code words $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$ is again a code word in $\mathcal{C}$. It follows that the code $\mathcal{C}$ is a K dimensional subspace of the vector space of all 2^N binary length N vectors. K linearly independent code words in $\mathcal{C}$ form the basis of the subspace $\mathcal{C}$, i.e. any code word $\mathbf{c} \in \mathcal{C}$ can be uniquely expressed as a linear combination of these K linearly independent vectors. These K base vectors entirely define the code and are commonly arranged as the rows of a $K \times N$ generator matrix $\mathbf{G}$. This offers a convenient linear encoding rule from the set of information words to the set of code words:

$$\mathbf{c} = \mathbf{u} \cdot \mathbf{G}. \tag{2.51}$$

The columns of $\mathbf{G}$ correspond to the code word positions, the rows to the information word positions. The encoding mapping is systematic if the K information bits are contained in the code word $\mathbf{c}$.

Alternatively, the code $\mathcal{C}$ may be defined as the null space of a $(N-K) \times N$ parity check matrix $\mathbf{H}$:

$$\mathbf{c} \cdot \mathbf{H}^T = \mathbf{0}_{1\times(N-K)}, \tag{2.52}$$

where $\mathbf{0}_{1\times(N-K)}$ is the all-zero vector of length $N-K$. The columns in $\mathbf{H}$ correspond to the code word positions, the rows to the parity check equations fulfilled by a valid code word.

The code words of a convolutional code are the output sequence of a linear encoder circuit fed by the information bits. This code construction sets additional constraints on the characteristics of the corresponding $\mathbf{G}$ and $\mathbf{H}$ matrices.

Consider as example the convolutional encoder depicted in Fig. 2.7. At any given time i, an information bit u_i is shifted into the circuit and two coded bits $\mathbf{c}_i = (c_{i,1}, c_{i,2})$ are generated from u_i and the two previous input bits u_{i-1} and u_{i-2}.

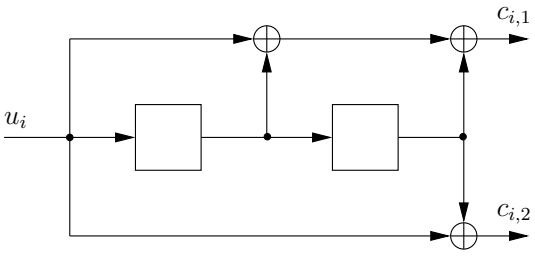


Figure 2.7: Rate $1/2$ convolutional encoder.

We only consider convolutional codes with one encoder input and N_c decoder outputs, resulting in a code rate of $R = 1/N_c$. Convolutional codes of rate k_c/N_c are conveniently constructed by periodically puncturing a low rate mother code. The puncturing pattern is usually defined through a $N_c \times p_c$ puncturing matrix $\mathbf{P}$ for a the puncturing period p_c.

***Example** 1 For the encoder given in Fig. 2.7, we apply as example the puncturing matrix*

$$\mathbf{P} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}. \tag{2.53}$$

Then, the 4th bit of the first decoder output and the 2nd bit of the second output are deleted. This puncturing pattern of length $p_c = 4$ is periodically repeated. The code rate of the punctured code is $R = 2/3$.

The semi-infinite sequence $\mathbf{u} = (u_1, u_2, \dots)$ is most conveniently described by the series

$$u(D) = \sum_{i=1}^{\infty} u_i \cdot D^{i-1}. \tag{2.54}$$

Then, the jth output of a rate $R = 1/N_c$ convolutional encoder is given by the series

$$c_j(D) = u(D) \cdot g_j(D), \quad j = 1, \dots, N_c, \tag{2.55}$$

where $g_j(D)$ are the transfer functions characterizing the convolutional code. The encoding equation can be compactly described by

$$\mathbf{c}(D) = u(D) \cdot \mathbf{G}(D), \tag{2.56}$$

with $\mathbf{c}(D) = (c_1(D), \dots, c_{N_c}(D))$ and the generator matrix $\mathbf{G}(D) = (g_1(D), \dots, g_{N_c}(D))$.

***Example** 2 For the encoder given in Fig. 2.7, we have $g_1(D) = 1+D+D^2$, $g_2(D) = 1+D^2$, and $\mathbf{G}(D) = (1 + D + D^2, 1 + D^2)$.*

In general, the transfer function $g_j(D)$ may be rational:

$$g_j(D) = \frac{f_j(D)}{q_j(D)} = \frac{f_{j,0} + f_{j,1}D + \dots f_{j,m}D^m}{1 + q_{j,1}D + \dots + q_{j,m}D^m}. \tag{2.57}$$

The code memory M_c is the largest degree of $f_j(D)$ and $q_j(D)$:

$$M_c = \max_{j=1,\ldots,N_c} \{\deg\, f_j(D), \deg\, q_j(D)\}. \tag{2.58}$$

The encoding circuit of the single-input, single-output linear system with transfer function $g_j(D)$ in its controller canonical form is depicted in Fig. 2.8 [JZ99].

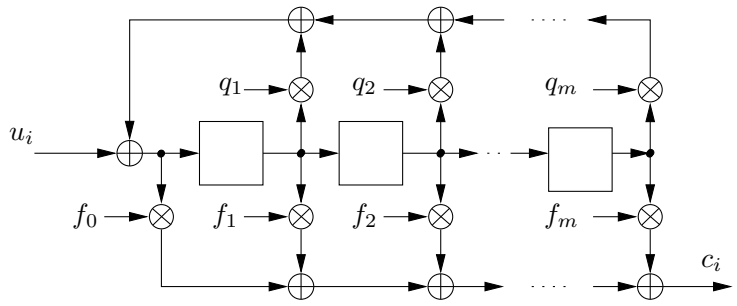


Figure 2.8: Controller canonical form of a rational transfer function.

If the encoder does not contain a feedback path, we have a *feedforward* encoder, otherwise we have a *feedback* or *recursive* encoder. A non-systematic feedforward encoder with generator matrix $\mathbf{G}(D) = (g_1(D),\ g_2(D))$ has an equivalent systematic recursive encoder. The rational transfer functions

$$\begin{aligned} \tilde{g}_1(D) &= 1, \\ \tilde{g}_2(D) &= g_2(D)/g_1(D), \end{aligned} \tag{2.59}$$

yield the same set of codewords and thus, the same convolutional code. However, the mapping of information bits to code words differs. The recursive systematic encoder corresponding to the code generated by the encoder given in Fig. 2.7 is depicted in Fig. 2.9.

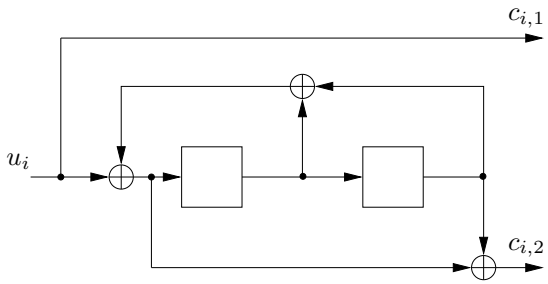


Figure 2.9: Rate $1/2$ convolutional recursive systematic encoder.

***Example** 3 For the encoder given in Fig. 2.9, we have $\tilde{g}_1(D) = 1$, $\tilde{g}_2(D) = 1 + D^2/(1 + D + D^2)$, and $\mathbf{G}(D) = (1, 1 + D^2/(1 + D + D^2))$. The code memory is $M_c = 2$.*

Convolutional codes are compactly described in the widely used octal notation. Then, the coefficients of the polynomials $f_j(D)$ and $q_j(D)$ from equation (2.57) are expressed in octal form, where $f_{j,0}$ is the most significant bit (MSB).

***Example** 4 In the example of Fig. 2.7, the coefficients of the generator matrix $\mathbf{G}(D) = (1 + D + D^2, 1 + D^2)$ are given in octal notation by $(7, 5)$. The recursive code in Fig. 2.9 is denoted by $(1, 5/7)$. A code rate $R = 1/2$, memory $M_c = 3$ code with $g_1(D) = 1$ and $g_2(D) = 1 + D + D^3/(1 + D^2 + D^3)$ is compactly described by $(1, 15/13)$ in octal notation.*

Since convolutional codes are linear codes, the encoding can be described with the binary generator matrix $\mathbf{G}$:

$$\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \dots) = (u_1, u_2, \dots) \cdot \mathbf{G} = \mathbf{u} \cdot \mathbf{G}, \tag{2.60}$$

with, for a feedforward encoder, a banded generator matrix of the type:

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \cdots & \mathbf{G}_{M_c} & & & \\ & \mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \cdots & \mathbf{G}_{M_c} & & \\ & & \mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \cdots & \mathbf{G}_{M_c} & \\ & & & \ddots & \ddots & & & \ddots \end{pmatrix}. \tag{2.61}$$

The $1 \times N_c$ submatrix $\mathbf{G}_l$ governs how the input u_{i-l} affects the output $\mathbf{c}_i = (c_{i,1}, c_{i,2}, \dots, c_{i,N_c})$:

$$\mathbf{c}_i = \sum_{l=0}^{M_c} u_{i-l} \cdot \mathbf{G}_l. \tag{2.62}$$

One column in the generator matrix corresponds to the impulse response of the encoder shift register. Note that the encoding operation in equation (2.62) is essentially a convolution of the impulse response $\mathbf{G}_l$, $l = 0, \dots, M_c$ of the encoder with the information sequence, which illuminates the name convolutional codes.

With a recursive encoder, the impulse response is infinite and the generator matrix has the following structure:

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \mathbf{G}_3 & \mathbf{G}_4 & \cdots \\ & \mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \mathbf{G}_3 & \cdots \\ & & \mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \cdots \\ & & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{2.63}$$

***Example** 5 For the encoder given in Fig. 2.7, we have $\mathbf{G}_0 = (1, 1)$, $\mathbf{G}_1 = (1, 0)$, and $\mathbf{G}_2 = (1, 1)$ resulting in:*

$$\mathbf{G} = \begin{pmatrix} 11 & 10 & 11 & & & \\ & 11 & 10 & 11 & & \\ & & 11 & 10 & 11 & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{2.64}$$

With the recursive encoder of Fig. 2.9, the impulse response of the first output (systematic part) is $1, 0, 0, \dots$ and the impulse response of the second output (parity part) is the infinite repetition of the sequence $1, 1, 1, 0, 1$, resulting in the generator matrix:

$$\mathbf{G} = \begin{pmatrix} 11 & 01 & 01 & 00 & 01 & \cdots \\ & 11 & 01 & 01 & 00 & \cdots \\ & & 11 & 01 & 01 & \cdots \\ & & & 11 & 01 & \cdots \\ & & & & 11 & \cdots \\ & & & & & \ddots \end{pmatrix}. \tag{2.65}$$

We described the encoder output as a function of the present and previous encoder input bits. The encoder output can also be given as a function of the present input and the content of the memory elements defined as the encoder state at time index i. A widely used graphical representation of the sequence of states traversed by the encoder for a given input sequence is the trellis. The trellis depicts the 2^{M_c} possible states for time i and $i+1$ and the state transitions through the 2^K branches leaving and entering each state. A trellis section for the encoder of Fig. 2.9 is shown in Fig. 2.10.

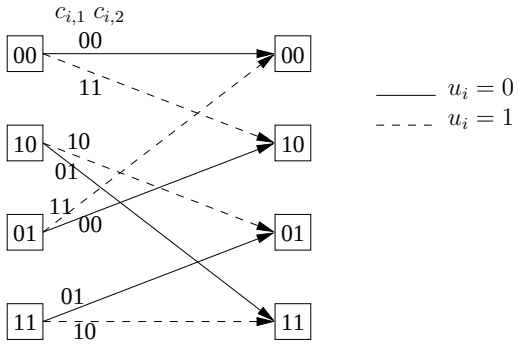


Figure 2.10: Trellis section of the rate $R = 1/2$, memory $M_c = 2$, $(1, 5/7)$ convolutional code.

The trellis need to be terminated to the zero state at the end of a code word by appending M_c deterministic bits to the information bits, resulting in a slightly decreased code rate. For a large code word length, this rate decrease is negligible. With an input sequence

$$\mathbf{u} = (u_1, \ldots, u_K), \tag{2.66}$$

the coded bits are multiplexed to the sequence

$$\mathbf{c} = (c_1, \ldots, c_N). \tag{2.67}$$

The code rate of the terminated code is then

$$R = \frac{K}{N} = \frac{K}{N_c(K + M_c)} \approx \frac{1}{N_c} \tag{2.68}$$

for a large block length $K \gg M_c$. For a short block length, tail-biting convolutional codes may be used to avoid the termination, see [Wei01] and references therein for an extensive coverage.

The Viterbi algorithm [Vit67] is an efficient implementation of the optimum ML word based decoding for convolutional codes. The basic concept is the sequential computation of the metric and the tracking of survivor pathes in the trellis. The algorithm was extended in [HH89b] for soft-outputs (SOVA algorithm).

We use in this work the BCJR algorithm, an abbreviation due to the initials of the inventors Bahl, Cocke, Jelinek, and Raviv [BCJR74]. The BCJR algorithm performs MAP symbol based decoding using a recursive forward and backward calculation of state probabilities over the trellis.

A low complexity approximation of the BCJR algorithm is the max-log BCJR algorithm [RVH95] [KB90] that uses the principle given in equation (2.38). Only one path is tracked in the forward and backward recursion of the BCJR algorithm. We can interpret the max-log BCJR as two Viterbi algorithms, one running in the forward direction, one in the backward direction. The

hard decisions are identical to the standard Viterbi algorithm with a maximum decision delay. The addition of a correction term in the max-log BCJR algorithm [RVH95] yields the performance of the optimal BCJR and is convenient to implement in a real communication system. The correction term can be efficiently implemented with a look-up table.

The error probability performance of a convolutional code is directly related to its distance spectrum. The distance spectrum $\{a_d\}$ specifies the number of code bit sequences of Hamming weight d that leave the all zero state of the trellis at time $i = 1$ and end in the all zero state for some $i > 1$. The first non-zero a_d, $d = 2, 3, \ldots$ specifies the so-called free distance d_f of the code. The cumulated Hamming weight of the a_d information bit sequences generating weight d code bit sequences is denoted by c_d. Note that for a terminated convolutional code, the free distance d_f does not depend on the block length and that the bit error probability is nearly independent of the block length.

2.4.2 Parallel and Serial Concatenated Convolutional Codes

Instead of using a single code, several codes may be concatenated and sequentially decoded [For66]. The main advantage is that the computational demanding task of decoding the strong overall code is broken into simpler computationally feasible sequential decoding steps. An attractive solution first implemented for space communications was the concatenation of a Reed-Solomon code as *outer* code with a convolutional code as *inner* code. The Viterbi decoder will occasionally output bursts of errors and therefore, the Reed-Solomon code is well suited as outer code to reduce the error rate. A further improvement is to insert an interleaver between the two codes and to introduce a feedback from the Reed-Solomon code to the convolutional code (”reiterated-decoding” [HOP93]).

The breakthrough towards capacity approaching codes was the invention of the so-called turbo codes that comprise iterative decoding of parallel concatenated convolutional codes (PCCC) [BGT93]. The astonishing performance of 0.5dB from to the Shannon limit was presented for a code of rate $R = 1/2$ transmitted over an AWGN channel. The main concept of the proposed system is to iteratively exchange extrinsic information between the constituent soft-in/soft-out decoders of concatenated codes separated by interleavers. Iterative decoding of serial concatenated convolutional codes (SCCC) was investigated in [BDMP98]. The successive exchange of extrinsic information is now commonly understood as an instance of the sum-product algorithm [KFL01].

Parallel Concatenated Codes

Fig. 2.11 depicts the encoder of a rate $R = 1/3$ parallel concatenated turbo code. The information bits are encoded by a first systematic $R = 1/2$ encoder. An interleaved version of the information sequence is encoded by a second systematic $R = 1/2$ encoder. Finally, the code word is formed by the information bits and the two parity sequences of both encoders. The interleaver aims at generating independent parity sequences.

The originally proposed turbo code employed two identical systematic convolutional encoders [BGT93]. Several other parallel code constructions have been proposed, including multiple concatenated turbo codes [DP95], asymmetric turbo codes [TCMC99] or non-systematic turbo codes [CTC00] [Brä04].

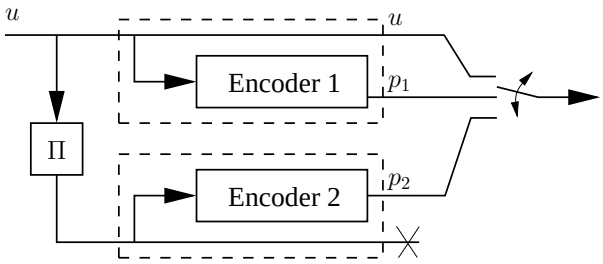


Figure 2.11: Encoding circuit of a parallel concatenated code.

A decoding circuit for a parallel concatenated code is shown in Fig. 2.12. The two constituent codes are decoded alternatively using soft-in/soft-out symbol based MAP decoding, e.g. the BCJR algorithm for convolutional codes. Each decoder uses the extrinsic information on the information bits of the other decoder as a priori information. After a decent number of iterations, usually less than 15, no improvement by further iterations is achieved, the decoder converges and outputs the computed a posteriori estimate.

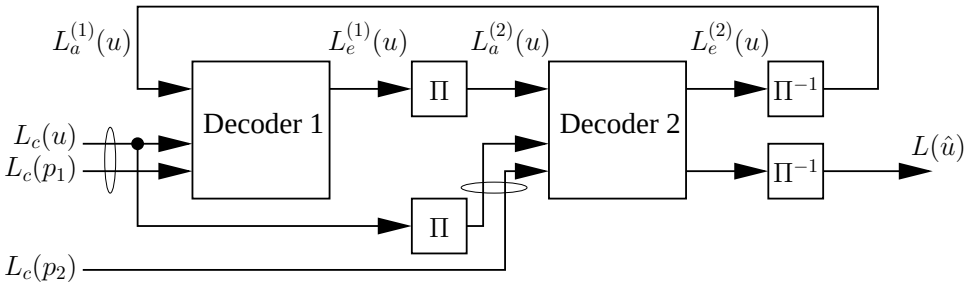


Figure 2.12: Decoding circuit for a parallel concatenated code; $L(\hat{u})$, $L_c(u)$, $L_a(u)$, and $L_e(u)$ denote the a posteriori, channel, a priori and extrinsic information, respectively; $L_e(u) = L(\hat{u}) - L_c(u) - L_a(u)$, see equation (2.42).

The turbo code proposed in [BGM98] and used in the UMTS physical layer standard [3GPP06b] has two systematic constituent convolutional codes of memory 3 with generator matrix $\mathbf{G}(D) = (g_1(D),\ g_2(D))$ and

$$g_1(D) = 1; \qquad g_2(D) = \frac{1 + D + D^3}{1 + D^2 + D^3}. \tag{2.69}$$

As mentioned in Example 4, this corresponds to $(1, 15/13)$ in octal notation.

To highlight the relation to general linear codes, the generator matrix $\mathbf{G}$ of the overall PCCC is derived. We start by constructing the following equivalent PCCC matrix:

$$\mathbf{G}' = (\mathbf{I}_K,\ \ \mathbf{G}_1,\ \ \mathbf{\Pi} \cdot \mathbf{G}_2)\,, \tag{2.70}$$

where $\mathbf{I}_K$ is the $K \times K$ identity matrix, $\mathbf{G}_1$ and $\mathbf{G}_2$ are the $K \times K$ generator matrices for the rate $R = 1$ parity part with the upper triangular structure given in equation (2.63), and $\mathbf{\Pi}$ is a $K \times K$ matrix defining the interleaving. The code termination is not considered here. By rearranging the columns of $\mathbf{G}'$ according to the multiplexing of the systematic and parity bits of the two encoders, we obtain the generator matrix $\mathbf{G}$ of the overall PCCC.

Serial Concatenated Codes

The encoder of a serially concatenated code is depicted in Fig. 2.13. The coded bits of the outer encoder are interleaved and fed as information bits to the inner encoder.

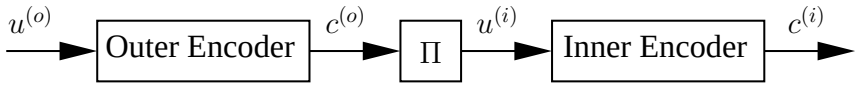


Figure 2.13: Encoding circuit of a serial concatenated code.

The decoding algorithm shown in Fig. 2.14 has several differences to the one for parallel concatenated codes: the outer decoder receives no direct channel information, the inner decoder therefore forwards the extrinsic information $L_e(u^{(i)}) = L(\hat{u}^{(i)}) - L_a(u^{(i)})$ that includes the channel information, see equation (2.42). The extrinsic information on the inner information bits is the a priori information of the outer coded bits. The outer decoder feed back the extrinsic information on the outer coded bits to the inner decoder.

The inner code needs to be recursive in order to achieve the so-called *interleaver gain* [BDMP98]. Then, the decrease of the error rate with the code word length is unbounded. The outer code should have a large free distance and a good distance spectrum.

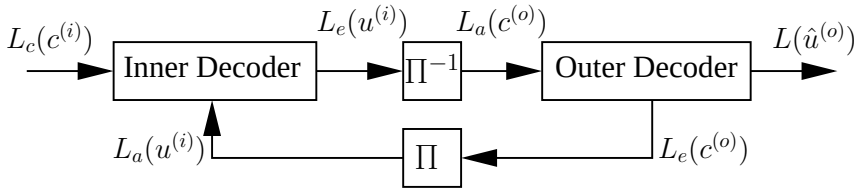


Figure 2.14: Decoding circuit for a serial concatenated code. $L(\hat{u})$, $L_c(c)$, $L_a(u)$, and $L_e(u)$ denote the a posteriori, channel, a priori and extrinsic information, respectively; $L_e(u) = L(\hat{u}) - L_a(u)$.

As for the PCCC, we can easily give the generator matrix of the overall code:

$$\mathbf{G} = \left(\mathbf{G}^{(\mathrm{o})} \cdot \mathbf{\Pi} \cdot \mathbf{G}^{(\mathrm{i})}\right), \tag{2.71}$$

where $\mathbf{\Pi}$ is a square matrix defining the interleaving and $\mathbf{G}^{(\mathrm{o})}$ and $\mathbf{G}^{(\mathrm{i})}$ are the generator matrices of the outer and inner encoder, respectively.

The overall serial or parallel concatenated code has less structure and more randomness than the simple convolutional code. The convolutional code has a satisfactory performance for short block length, the concatenated code approaches the capacity for large block length.

There is a trade-off between a code with a random-like structure for capacity approaching performance with large block length and high complexity decoding vs. a well structured code for good performance with short block length and low complexity decoding. Random codes are the marginal example for codes without any structure that approach capacity for very large block length, but with impractical complexity. Low-density parity check (LDPC) codes [Gal62] have attracted great attention in recent research and standardization work and can be classified as more random than concatenated convolutional codes with in general a better performance for very large block length.

2.4.3 Iterative Decoding in Communication Systems

The concept of iterative decoding has proven to be valid for a large variety of decoding and detection problems in communication systems [Hag97]. Table 2.2 gives an overview of some applications of the so-called turbo principle for serial concatenated systems.

application	inner code	outer code
serial code concatenation	FEC en-/decoder	FEC en-/decoder
turbo equalization	FEC en-/decoder	Multipath channel / equalizer
Turbo BICM	FEC en-/decoder	Mapper/demapper
Turbo MIMO	FEC en-/decoder	Mapper/MIMO detector
Turbo multiuser detection	FEC en-/decoder	Multiuser channel / detector
Turbo source channel coding	source en-/decoder	FEC en-/decoder
LDPC code	check nodes	variable nodes
Joint network channel coding	network code	FEC en-/decoder

Table 2.2: Some applications of the turbo-principle [Hag04].

2.5 Coded Modulation

Recall that the data transmission rate is $R_t = R \cdot M$, where R is the code rate and M is the number of bits associated to one transmitted symbol. To achieve high data rates at good channel conditions, we have to use large signal constellations since the code rate is reasonably limited to $R < 1$.

Furthermore, from an information theoretical point of view, it is preferable to use large signal constellations with low rate channel codes than small signal constellations with a high rate code to transmit at a specific data rate R_t. In the example of Fig. 2.6 and for a data rate of $R_t = M \cdot R = 1.8$ bits per channel use, it would be better to use 16QAM ($M = 4$) with a rate $R = 9/20$ code instead of QPSK ($M = 2$) with a rate $R = 9/10$ code. In practice, smaller signal constellations with code rates larger than $R = 1/4$ are often used due to complexity issues, synchronization problems or sensitivity to non-linear distortions.

For a good overall performance, we have to efficiently combine high order modulation for high data rates with strong channel coding schemes for high reliability. The success of the first attempts to directly combine binary codes with high order modulation has been decent, and for increasing spectral efficiency, only slight coding gains over uncoded schemes with smaller signal constellations were possible to achieve.

To overcome this unsatisfactory performance, Massey proposed in 1974 to jointly design coding and modulation [Mas74] and thus founded the field of coded modulation. In 1976/77, Ungerböck presented trellis coded modulation (TCM) [Ung76] [Ung82] and Imai proposed multilevel coding (MLC) [IH97] [WFH99] as powerful and applicable coded modulation schemes. In both approaches, an Euclidean distance measure was optimized instead of the commonly used Hamming distance in channel coding. In 1992, Zehavi introduced bit-interleaved coded modulation (BICM) [Zeh92], that is nothing else than a serial concatenation of a code, an interleaver and a mapper. A thorough analysis performed by Caire, Taricco and Biglieri in [CTB98] revealed that very close to capacity performance is possible with BICM and Gray mapping. Iterative decoding of BICM, so-called BICM-ID was introduced in [LR97] and [tBSY98a]

to achieve a good performance with simple codes in combination with carefully chosen mappings different from Gray.

Obviously, BICM is nothing else but the traditional approach used before Massey introduced the field of coded modulation. The question arises what were the reasons for the poor performance of the traditional approach in the past? As remarked in [HWF98], the poor overall performance is due to the weak binary channel coding schemes and not to an improper combining of coding and modulation. The distance to capacity limits with the traditional schemes is approximately similar for different bandwidth efficiencies. However, the gap between uncoded transmission and capacity is smaller for high bandwidth efficiency, resulting in a smaller coding gain.

Nevertheless, for the innovative iterative schemes like BICM-ID, it is once again crucial to jointly design coding and modulation.

The design guidelines for all coded modulation schemes result from the following Chernoff upper bounds on the pairwise error probability resulting from the equations (B.16), (B.5) and (B.6) derived in Appendix B. The pairwise error probability is the probability that a symbol sequence $\mathbf{x}$ is chosen instead of the sequence $\hat{\mathbf{x}}$ that differs in d positions. Assume $\mathbf{x}$ and $\hat{\mathbf{x}}$ differ in the first d positions. Using equation (B.7) we have for an AWGN channel

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \leq \exp\left(-\frac{E_s}{4N_0}\sum_{n=1}^{d}|x_n - \hat{x}_n|^2\right). \tag{2.72}$$

The cumulated squared Euclidean distance $\sum_{n=1}^{d}|x_n - \hat{x}_n|^2$ between $\mathbf{x}$ and $\hat{\mathbf{x}}$ is decisive and not the number d of symbols that differ. For a fading channel at high SNR, the pairwise error probability is with equation (B.8)

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \leq \prod_{n=1}^{d}\frac{1}{1+\frac{E_s}{4N_0}|x_n - \hat{x}_n|^2}. \tag{2.73}$$

We observe that at high SNR, the error probability behaves as:

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \sim \left(\frac{E_s}{N_0}\right)^{-d}. \tag{2.74}$$

This relation is similar to what is normally achieved with time diversity techniques [Pro01]. Because of this similarity, the parameter d is usually called time or code diversity in this context. The time diversity and the squared product distance $\prod_{n=1}^{d}|x_n - \hat{x}_n|^2$ are the design parameters for fading channels. Therefore the value of d has a major influence on the error probability with fading channels in contrast to AWGN channels. This corresponds to the intuitive guess, stating that valid symbol sequences should differ in several positions with fading channels since transmitted symbols may be erased through deep fades.

In this Section, we review the main aspects of the three coded modulation schemes TCM, MLC and in particular BICM without and with iterative demapping and decoding.

2.5.1 Trellis Coded Modulation (TCM)

Trellis coded modulation (TCM) was proposed by Ungerböck [Ung76] [Ung82] as a coding scheme that improves the error rate performance without sacrificing data rate or requiring more bandwidth. This was achieved by constellation expansion and carefully designed combination of coding and modulation.

TCM is the serial concatenation of a rate $R = (M-1)/M$ convolutional code and a mapper for a 2^M-ary signal constellation. Each state transition in the convolutional code then corresponds to one transmitted symbol. If 2^{M-1} exceeds the number of states, it is unavoidable that the trellis has parallel state transitions and the encoder leaves some information bits uncoded. Fig. 2.15 depicts the encoder, where $M - \tilde{M}$ information bits remain uncoded.

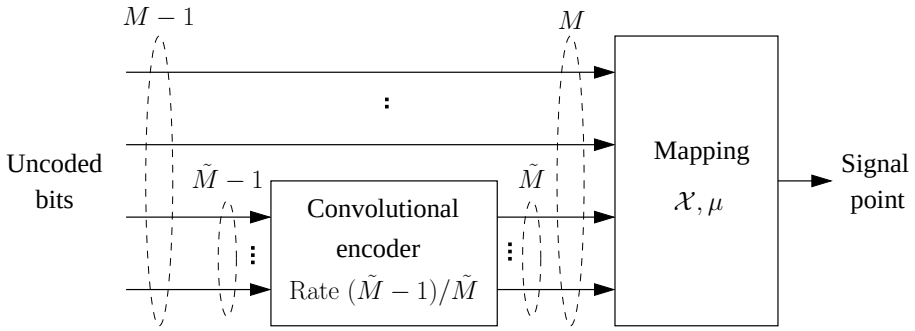


Figure 2.15: General structure of a TCM encoder.

For each trellis time step, M bits are mapped to a complex symbol chosen from the 2^M-ary signal constellation $\mathcal{X}$ (e.g. 16QAM for $M = 4$) according to a mapping $\mu : \{0,1\}^M \rightarrow \mathcal{X}$ that defines the assignment of the binary bit sequences to the signal points. The decoding is done on the joint trellis using e.g. the Viterbi algorithm.

TCM was first designed for AWGN channels and according to equation (2.72), the aim was to maximize the free Euclidean distance. The free Euclidean distance is defined as the minimum cumulated Euclidean distance between symbol sequences corresponding to different pathes in the trellis. Ungerböck proposed a mapping by set partitioning well suited for this application. Set partitioning is described in detail in Section 4.2. The basic idea is to divide the signal constellation recursively into subsets of signal points such that the minimum Euclidean distance between two signal points within a subset is maximized at each stage. Ungerböck formulated the following heuristic design rules for AWGN channels [Ung82]: First, the parallel trellis transitions should be associated to signal points from the subset with the maximum inter-symbol Euclidean distance; Second, trellis transitions that originate or join the same state should be associated to signal points from a subset with large inter-symbol Euclidean distance.

TCM codes designed for AWGN channels have usually a poor performance in fading channels. The minimum time diversity d in equation (2.73), i.e. the minimum number of symbols that differ along different trellis pathes may be reduced to one due to the uncoded bits and the parallel trellis transitions. TCM codes designed for fading channels have been proposed in [DS88] [SC89].

The parallel concatenation of two TCM encoder with iterative decoding at the receiver was proposed in [RW98] under the name of Turbo Trellis Coded Modulation (TTCM). The symbols from the two TCM encoders separated by symbol-wise interleavers are punctured so that all information bits are send exactly once and the partity bits are provided alternatively by the two encoders.

A drawback of TCM is that it is not straightforward to implement adaptive modulation and coding to cope with time varying channels since the channel code and the modulation are tightly linked together.

2.5.2 Multilevel Codes (MLC)

Imai's multilevel encoder [IH97] is depicted in Fig. 2.16. A sequence of information bits is demultiplexed to M layers or levels that correspond to the M bit positions in the binary label of the signal points. Each layer is separately encoded by a binary code and interleaved by a different interleaver Π_m, $m = 1, \ldots, M$. The interleavers are included for a better performance in fading channels and for iterative decoding. At time index n, one bit from each coded and interleaved level is grouped to the sequence

$$\mathbf{c}_n = (c_{n,1}, \ldots, c_{n,m}, \ldots, c_{n,M}) \tag{2.75}$$

and mapped to the complex symbol $x_n = \mu(\mathbf{c}_n)$ chosen from the 2^M-ary signal constellation $\mathcal{X}$ according to the bit-to-symbol mapping $\mu : \{0,1\}^M \rightarrow \mathcal{X}$ (e.g. Gray mapping).

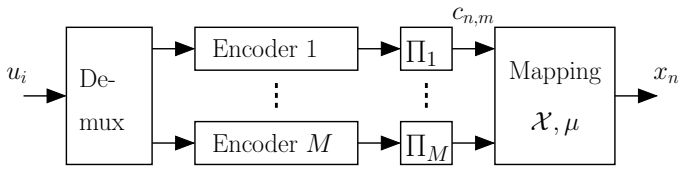


Figure 2.16: Multi-level coding (MLC): encoder.

The optimum receiver would be the joint overall ML decoder. The ML algorithm implements equation (2.29) and the most probable information bit value u_i is selected given the received complex sequence $\mathbf{y}$. However, a joint ML demapper and decoder of all layers has an impractical complexity. Therefore, we separate the demapping and decoding task.

Commonly used approaches to decode MLC include multistage decoding (MSD), parallel decoding of the individual levels (PDL), and iterative decoding [WFH99][Wör96]. For symbol-by-symbol MAP soft demapping and decoding, we use the demapper described in equation (2.48) and the BCJR algorithm for the decoder.

The multistage decoding (MSD) receiver is a direct interpretation of the chain rule for mutual information. Let Y and X be the random variables for the channel output and input, respectively, and $\mathbf{C} = (C_1, \ldots, C_m, \ldots, C_M)$ the vector of random variables for the coded bits associated to one channel input symbol. Applying the chain rule for mutual information leads to:

$$\begin{aligned} I(Y;X) &= I(Y;\mathbf{C}) = I(Y;C_1,\ldots,C_M) \\ &= \sum_{m=1}^{M} I(Y;C_m|C_1,\ldots,C_{m-1}). \end{aligned} \tag{2.76}$$

Equation (2.76) reveals that we should decode each layer $m = 1, \ldots, M$ individually, taking decisions of prior decoding stages into account. The MSD circuit is depicted in Fig. 2.17. MSD is closely related to multiple-access or multiple-antenna receivers with serial interference cancelation (also called stripping or onion peeling receivers). Instead of subtracting the interference from other users or antennas, we reduce the number of possible signal points with each decoding stage to improve the reliability of the decisions. After e.g. a first ideal or genie decoding stage, the signal constellation is reduced from 2^M to 2^{M-1} signal points since the first position in the binary label of length M is a priori known.

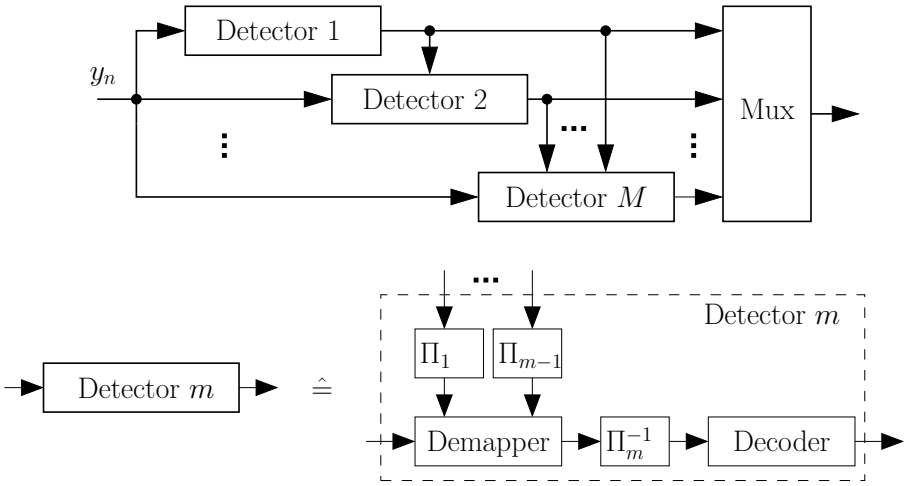


Figure 2.17: Multi-level coding (MLC): multistage decoding (MSD).

From equation (2.76) we can define M *equivalent channels* [WFH99] with capacity

$$\mathcal{C}^{(m)} = I(Y; C_m | C_1, \ldots, C_{m-1}), \quad m = 1, \ldots, M. \tag{2.77}$$

Then, the transmission over the physical channel can be virtually separated into the parallel transmission of the individual bits over these M equivalent channels. If the code rates $R^{(m)}$ of the individual codes are chosen according to the capacity of the equivalent channels of the corresponding bit position ($R^{(m)} = \mathcal{C}^{(m)}$), MSD is sufficient to achieve the constellation constrained capacity and in theory no additional gain is obtained with overall joint ML demapping and decoding of all levels.

Beside this capacity rule, we should mention the balanced design rule and the coding exponent rule described in [WFH99] for assigning the individual channel codes to the M levels with MSD.

Parallel decoding of the individual levels (PDL) has been proposed as a further decoding scheme for MLC. With PDL, all decoders are operating in parallel and do not use the decisions of other decoders [WFH99] [Sch97]. From equation (2.76), the receiver constrained capacity $\mathcal{C}^{(m)}$ of the equivalent channel m with PDL is

$$\mathcal{C}^{(m)} = I(Y; C_m). \tag{2.78}$$

The overall receiver constrained PDL capacity $\mathcal{C}$ is then

$$\mathcal{C} = \sum_{m=1}^{M} \mathcal{C}^{(m)} = \sum_{m=1}^{M} I(Y; C_m) \leq \sum_{m=1}^{M} I(Y; C_m | C_1, \ldots, C_{m-1}) = I(Y; \mathbf{C}). \tag{2.79}$$

The PDL capacity strongly depends on the particular mapping. We have equality if the bit positions are independent to each other. Gray mapping closely approaches this scenario.

Finally, Fig. 2.18 shows the receiver circuit for MLC with iterative decoding [WH92] [II01]. All decoders are working in parallel as with PDL, but the decisions of the decoder are fed back as a priori information to the demapper. The extrinsic LLRs of the demapper and decoder are iteratively exchanged, similar to the decoding of a serial concatenated code in Fig. 2.14.

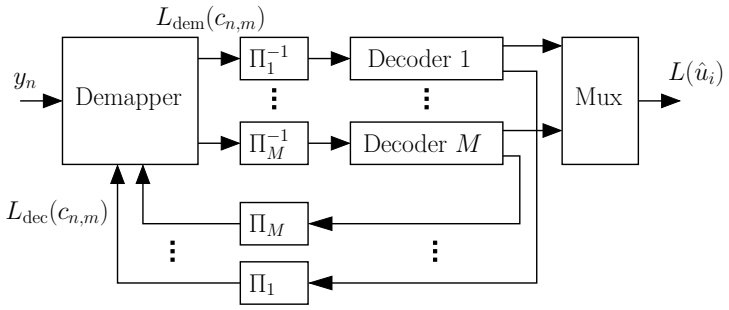


Figure 2.18: Multi-level coding (MLC): Iterative decoding.

2.5.3 Bit-Interleaved Coded Modulation (BICM and BICM-ID)

Bit-interleaved coded modulation (BICM) is the serial concatenation of a code, interleaver and mapper, as depicted in Fig. 2.19. Zehavi proposed this scheme in [Zeh92] and a thorough investigation was done by Caire, Taricco and Biglieri in [CTB98].

The information bits are processed by a single encoder and random interleaver Π. The coded and interleaved bit sequence $\mathbf{c}$ is partitioned in N_s subsequences $\mathbf{c}_n$ of length M:

$$\mathbf{c} = (\mathbf{c}_1, \ldots, \mathbf{c}_n, \ldots, \mathbf{c}_{N_s}), \quad \text{with} \quad \mathbf{c}_n = (c_{n,1}, \ldots, c_{n,m}, \ldots, c_{n,M}). \tag{2.80}$$

The bits $\mathbf{c}_n$ are mapped at time index n to a symbol x_n chosen from the 2^M-ary signal constellation $\mathcal{X}$ according to the binary labeling map $\mu : \{0,1\}^M \to \mathcal{X}$.

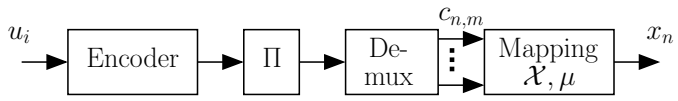


Figure 2.19: Bit-interleaved coded modulation (BICM) encoder.

Similar to MLC, the optimum BICM receiver would be the overall ML decoder. However, the complexity of a joint ML demapper and decoder is not manageable. Therefore, we separate the demapping and decoding task and consider a BICM receiver without and with iterative demapping and decoding. For symbol-by-symbol MAP soft demapping and decoding, we use the demapper described in equation (2.48) and the BCJR algorithm for the decoder.

With the standard BICM receiver without iterative demapping and decoding shown in Fig. 2.20, the demapper uses the received complex signals y_n and output the LLRs $L_{\text{dem}}(c_{n,m})$, $m = 1, \ldots, M$ of the corresponding coded bits. These LLRs are deinterleaved and decoded to obtain estimates about the information bits. Both the decoder and the demapper may use ML algorithms, but the overall BICM detector is no more ML.

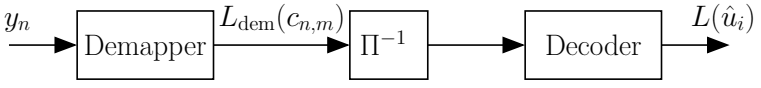


Figure 2.20: Bit-interleaved coded modulation (BICM) decoder.

The BICM receiver in Fig. 2.20 is closely related to MLC with PDL. We can also define equivalent channels for each bit position $m = 1, \ldots, M$ in the binary label with the bit-wise capacities $\mathcal{C}^{(m)}$ given in equation (2.78). Similar to MLC with PDL, the demapper in the standard BICM scheme has no a priori information and makes no use of previous decoder decisions. However, in contrast to MLC, the equivalent channels are not used in parallel, they are rather time multiplexed. If we assume ideal interleaving, the equivalent channels are selected randomly and the capacity of the BICM channel is the average over the equivalent channels of the bit positions. The overall capacity constrained by the BICM receiver is then identical to the PDL capacity given in equation (2.79).

Similar to PDL, the BICM capacity strongly depends on the applied mapping. We can approach the performance of the overall ML receiver if the bit positions in the symbol labels are independent. Gray mapping closely approaches this scenario [CTB98]. Fig. 2.21 depicts the receiver constrained capacity with a standard BICM receiver using different mappings. The values can be computed with equation (3.4) without a priori information at the demapper. A significant loss is obtained for mappings different from Gray.

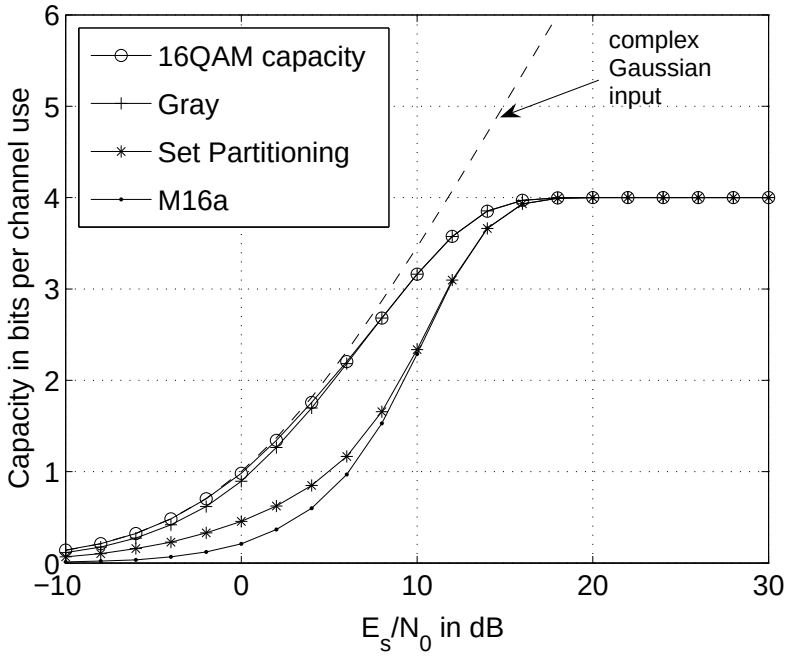


Figure 2.21: Receiver constrained capacity for bit-interleaved coded modulation (BICM) without iterative demapping and decoding. Mappings defined in Appendix A.

Li and Ritcey [LR97] [LR98] and ten Brink, Speidel and Yan [tBSY98a][tBSY98b] proposed the receiver with iterative demapping and decoding depicted in Fig. 2.22. The BICM system is considered as a serial concatenated code with the channel code as outer code and the mapper as inner code. The mapper can be truly considered as code that may introduce dependencies between the M bits associated to one symbol. The iterative receiver works as described in Fig. 2.14. The extrinsic LLRs $L_{\text{dec}}(c_{n,m})$ fed back from the decoder to the demapper may improve the reliability of the demapper decisions. During the initial demapping step, the a priori LLRs are equal to zero.

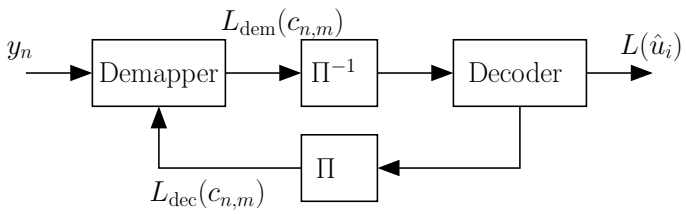


Figure 2.22: Bit-interleaved coded modulation decoder with iterative demapping and decoding (BICM-ID).

With this so-called BICM-ID scheme, large gains over the iterations are achieved in practical systems with carefully chosen mappings different from Gray. Which mappings are well suited for BICM-ID and how to optimize mappings for arbitrary signal constellations will be investigated in detail in Chapter 4.

With BICM-ID, the overall ML decoding performance and the constellation constrained capacity $I(Y;X)$ can be approached for all mappings, as shown in the following: Let I_L denote the average mutual information conditioned on the number $0 \leq L \leq M-1$ of a priori known bits. We obtain I_L by averaging over the M bit positions and the $\binom{M-1}{L}$ possibilities to choose L a priori known bit positions:

$$I_L = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{\binom{M-1}{L}} \sum_{j=1}^{\binom{M-1}{L}} I(Y; C_m | \mathbf{C}^L), \tag{2.81}$$

where $\mathbf{C}^L$ denotes the vector of L a priori known bits C_j, $j \neq m$ and $1 \leq j \leq M$. For the example of a 2^3-ary signal constellation (e.g. 8PSK) with a label length of $M = 3$, we have:

$$\begin{aligned} I_L = \frac{1}{3} \cdot (\quad & I(Y;C_1) + \tfrac{1}{2}\left(I(Y;C_1|C_2) + I(Y;C_1|C_3)\right) + I(Y;C_1|C_2C_3) + \\ & I(Y;C_2) + \tfrac{1}{2}\left(I(Y;C_2|C_1) + I(Y;C_2|C_3)\right) + I(Y;C_2|C_1C_3) + \\ & \underbrace{I(Y;C_3)}_{L=0} + \underbrace{\frac{1}{2}\left(I(Y;C_3|C_1) + I(Y;C_3|C_2)\right)}_{L=1} + \underbrace{I(Y;C_3|C_1C_2)}_{L=2}) \end{aligned} \tag{2.82}$$

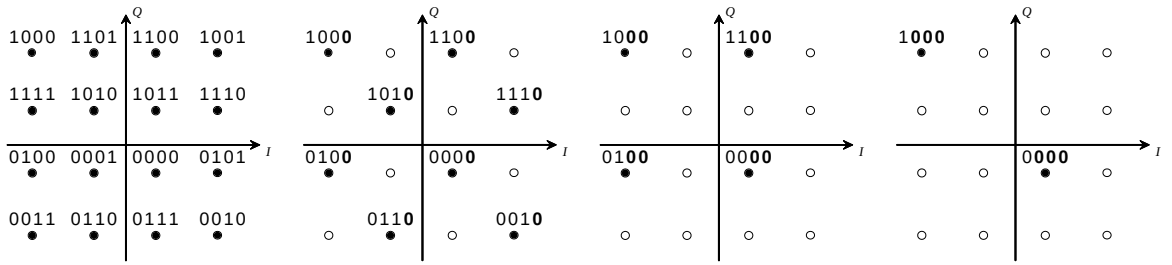
By reordering the terms from the example in equation (2.82) according to the chain rule of mutual information given in equation (2.76) and since the ordering of the arguments of the mutual information is irrelevant, i.e. $I(Y;C_1,C_2,C_3) = I(Y;C_2,C_1,C_3) = \ldots$, we deduce that

$$I(Y;X) = \sum_{L=0}^{M-1} I_L. \tag{2.83}$$

For the EXIT chart analysis in Section 3.3 and for the design of mappings in Chapter 4, we will often consider the two marginal scenarios with *no a priori information* and with *ideal a priori information* at the demapper. No a priori information is available during the initial demapping step or if the feedback from the channel decoder to the demapper is not implemented. The case of ideal a priori information (so-called genie or error free feedback case) is a lower bound on the performance at high SNR after several demapper and decoder iterations. With ideal a priori

information, the demapper knows exactly $M-1$ bits of the label but not the bit $c_{n,m}$ to be detected, since only extrinsic information is exchanged. Thus, the demapper has only to decide between the two symbols with bit labels differing solely in the a priori unknown mth bit and the 2^M-ary signal constellation is reduced to a symbol pair.

Fig. 2.23 illustrates for the example of 16QAM signal constellation and Ungerböck's set partitioning the reduction of the number of possible signal points for the very specific case where the symbol with label (0000) is transmitted and the last $0, 1, 2, 3$ bits are a priori known. This corresponds to a BEC a priori information where we have either ideal or no a priori knowledge on a bit position.



a) no a priori b) last bit c) last two bits d) last three bits

Figure 2.23: 16QAM signal constellation with set partitioning mapping. Symbol with label (0000) transmitted, remaining signal constellation if the last $0, 1, 2, 3$ bits are a priori known.

Similar to the definition of a soft-bit in equation (2.33), we define the soft-symbol $\bar{x}_n$ as the expectation of the symbol $x_n = \mu(\mathbf{c}_n)$. Assuming that the bit estimates corresponding to one symbol are independent through ideal interleaving, we have

$$\bar{x}_n = \mathrm{E}\{x_n | L(\mathbf{c}_n)\} = \sum_{x_n \in \mathcal{X}} x_n \cdot P(x_n | L(\mathbf{c}_n)) = \sum_{x_n \in \mathcal{X}} x_n \cdot \prod_{m=1}^{M} P(c_{n,m} | L(c_{n,m})). \tag{2.84}$$

$P(x_n|L(\mathbf{c}_n))$ is the a priori probability of the signal point x_n computed using the a priori LLRs $L(\mathbf{c}_n) = (L(c_{n,1}), \ldots, L(c_{n,M}))$ with equation (2.31).

Fig. 2.24 depicts the received signals y_n and the soft symbols from the a posteriori LLRs at the output of the demapper after $2, 4, 6$ demapping operations. The symbol with label (0000) is transmitted and Ungerböck's set partitioning mapping from Fig. 2.23 is applied. We observe that after 6 iterations, no more errors are made after the demapper. Furthermore, it is interesting to note that with a decent a priori information after 4 iterations, most soft symbols lie between the signal points with labels differing in only one bit.

A further concept to improve the performance of BICM not treated in this work but worth mentioning is the modulation or signal space diversity for fading channels [BV98]. There, the in-phase and quadrature components of the transmitted symbol are independently interleaved. The aim is to have in-phase and quadrature components that fade independently to improve the diversity. To maximize the diversity, the constellation should be properly rotated such that all distinct signal points are separable on every coordinate. A simple signal space diversity scheme for standard non-rotated signal constellations is to introduce a time delay larger than the coherence time of the fading process between in-phase and quadrature component [CR01]. Optimized mappings for a signal space diversity scheme with iterative demapping and decoding and independent in-phase and quadrature component interleaving are given in [CGV06].

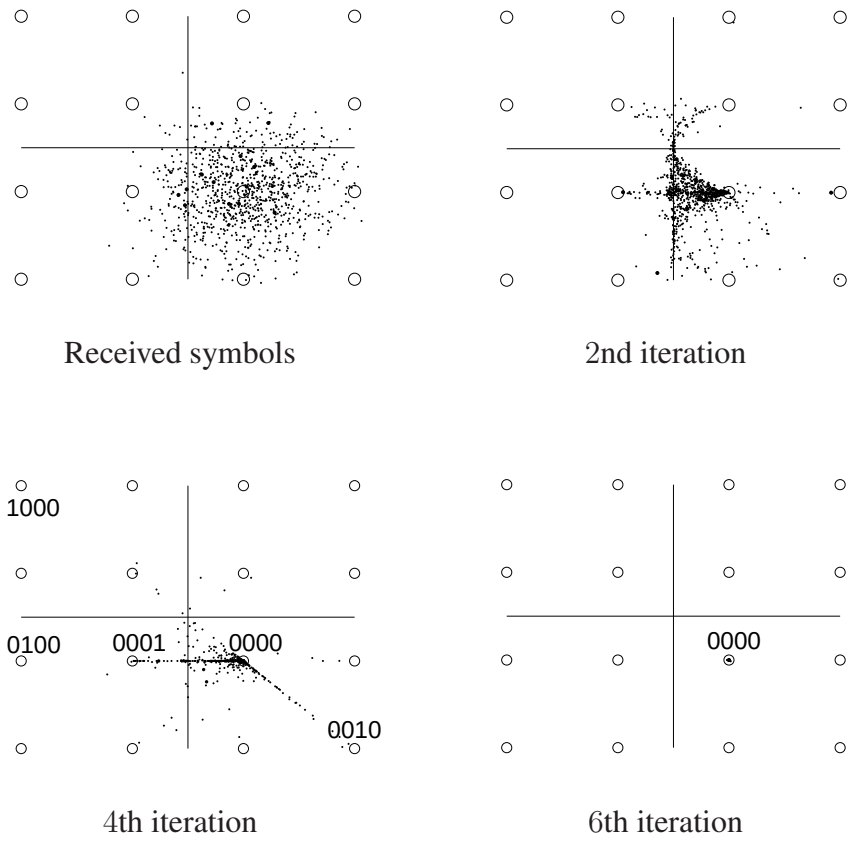


Figure 2.24: Received symbols and soft symbols after the demapper, 16QAM signal constellation with set partitioning mapping from Fig. 2.23. Symbol with label (0000) transmitted, AWGN channel, rate $1/2$ $(1, 5/7)$ convolutional code, $10\log_{10}(E_s/N_0)\text{dB} = 6\text{dB}$.

3

EXIT Chart Analysis

Extrinsic information transfer (EXIT) charts were proposed by ten Brink [tB01b] to analyze, design and optimize concatenated schemes with iterative decoding described in Section 2.4.2 and 2.4.3.

The error rate performance of these schemes has three characteristic SNR regions: First, at low SNR, no gain over the iterations is attained and the error rate is high. The amount of information exchanged between the soft-in/soft-out receiver components is negligible. Then, in the so-called turbo cliff or waterfall region, the iterative system starts to converge and the error rate curve drops within a small SNR range. Finally, at high SNR, the error rate curve flattens out, resulting in a sometimes negligible error floor. Each receiver component has close to ideal a priori information from the other receiver components.

To analyze and optimize the convergence behavior of iterative receivers in the SNR range of the waterfall region, density evolution techniques have been proposed in [RU01] [RSU01]. The main idea is to track the pdfs of the information messages exchanged in the iterative decoding algorithm. To simplify the analysis, we can assume Gaussian pdfs [CRU01] or use a single parameter to describe the pdfs [DDP00a] [GH01] [AGR98] [RS98] [Nar01]. EXIT charts [tB01b] visualize the density evolution of the extrinsic LLRs over the iterations using as single parameter the average mutual information between the coded bits at the transmitter and the LLRs at the receiver.

To predict the evolution of the investigated parameter, the soft-in/soft-out components of the iterative receiver are interpreted as non-linear filters and characterized by their input/output function. For the EXIT chart analysis, we construct the characteristic *EXIT functions*.

We focus on the bit-interleaved coded modulation scheme with iterative demapping and decoding (BICM-ID) introduced in Section 2.5.3 and depicted in Fig. 3.1. In this Chapter, EXIT functions of the two receiver components, namely the demapper and the decoder, are thoroughly analyzed for a variety of parameters. A bit-level and symbol-level analysis is introduced. Furthermore, an analytic and numeric computation of the EXIT functions of the decoder and demapper is investigated that circumvent the need of Monte-Carlo simulations.

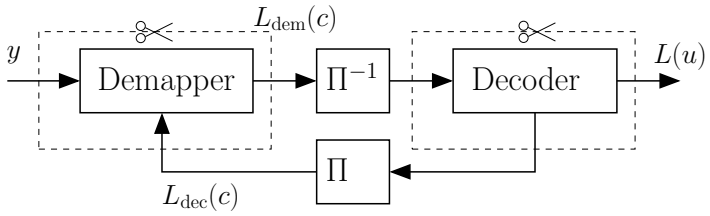


Figure 3.1: BICM-ID system model for EXIT chart analysis.

3.1 EXIT Chart Construction and Properties

Fig. 3.2 depicts the general model introduced in [AKtB04] to construct characteristic input/output functions for any receiver component of Table 2.2 in Section 2.4.3. The channel and a priori inputs of the considered decoder are the outputs of the communication channel and the virtual extrinsic or a priori channel, respectively. For a given communication channel, we vary the a priori channel quality and observe the extrinsic output of the decoder. Different channel and a priori information should be considered for the decoder, depending on the system architecture and on the position of the decoder in the receiver. We have to differentiate between serial and parallel concatenated schemes and between the inner and outer code in a serial concatenated scheme as described in Section 2.4.2.

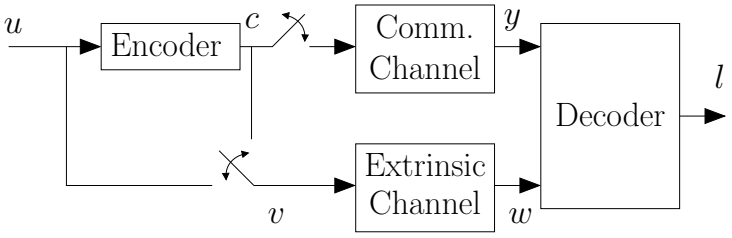


Figure 3.2: System model for the measurement of a decoder input/ouput function.

The crucial step to obtain a characteristic decoder function is to determine a single parameter that describes the extrinsic a priori input and the output of the decoder.

The extrinsic channel output and the decoder output are usually modeled as the output of AWGN, BEC or BSC channels. The BEC and BSC channels are directly characterized by a single parameter, namely the erasure probability ϵ and error probability p, respectively. For the AWGN channel, the channel output soft values – we consider the LLRs values – can be precisely described through their pdfs. To derive the characteristics of these pdfs, consider first the conditional LLRs $L(y|x)$ for BPSK modulation with $x \in \{\pm 1\}$ and an AWGN channel $y = x + n$ with $n \sim \mathcal{N}(0, \sigma_n^2)$. Then, the LLRs are given by:

$$L(y|x) = \frac{2}{\sigma_n^2} \cdot y. \tag{3.1}$$

Alternatively, we can write

$$L(y|x) = \mu_L \cdot x + n_L, \tag{3.2}$$

with a mean $\mu_L = 2/\sigma_n^2$ and a Gaussian distributed n_L with variance $\sigma_L^2 = 4/\sigma_n^2$ for $x = 1$ fixed. The mean and the variance are related by:

$$\mu_L = \frac{\sigma_L^2}{2}. \tag{3.3}$$

Therefore, we have Gaussian distributed LLRs $L(y|x) \sim \mathcal{N}(\sigma_L^2/2, \sigma_L^2)$ that are characterized by the single statistic σ_L^2 or μ_L. The assumption on an AWGN extrinsic channel and uncorrelated information turns out to be appropriate for large interleavers, for all optimal and most suboptimal receiver elements.

The mean μ_L and variance σ_L^2 have been used in [DDP00a] and [tB01b] to describe the LLR pdfs in iterative receivers, respectively. An early attempt to visualize the behavior of a soft-in/soft-out decoder used a similar measure based on the SNR [HH89a]. Other choices include the error rate [GH01], the mutual information [tB01b], and the fidelity or closely related measures based on "soft bits" [AGR98] [RS98] [Nar01] [Hag04]. An extensive comparison of these parameters for the analysis of iterative receivers is given in [TtBH02].

The two receiver components of a concatenated scheme iteratively exchange extrinsic information to improve the performance. The output of one component is the input of the other. Therefore, we can plot both characteristic functions in a single figure. We expect that the effective measured evolution of the considered parameter follows a staircase trajectory between the characteristic functions of the two components, i.e. that the possible deviations from the assumptions made while constructing the characteristic function – namely the Gaussian distribution and the uncorrelated a priori information – do not have a significant impact.

Out of the above mentioned parameters, the mutual information and the fidelity provide the most accurate prediction [TtBH02], mainly because the assumption of a Gaussian distribution is only made on the input pdf and not on the output as for the other measures. Using the mutual information gives the EXIT chart, using the fidelity gives the soft bit transfer (SOBIT) [Hag04] or fidelity chart. The advantage of the EXIT chart is that the mutual information has an information theoretical meaning, resulting in several interesting properties. A circuit to generate EXIT functions with AWGN channels is presented in [Hag04]. An example of an EXIT chart for a serially concatenated system is shown in Fig. 3.6.

Mutual Information Measure for EXIT Chart

For the EXIT chart, the mutual information (MI) between the coded bits C_i at the transmitter and the LLRs L_i at the receiver is used as the single parameter to characterize the pdf of the LLRs. We consider the average mutual information defined for a codeword of length N as follows [AKtB04]:

$$I = \frac{1}{N} \sum_{i=1}^{N} I(C_i; L_i). \tag{3.4}$$

Three promising approaches have been proposed to compute the mutual information for EXIT charts.

First, we can use the general equation (2.17) of the mutual information and simplify it using the following assumptions:

$$p(l, c) = p(l|c) \cdot p(c); \quad p(y) = \frac{1}{2}(p(l|c = 0) + p(l|c = 1)); \quad p(c) = \frac{1}{2}, \tag{3.5}$$

where c and l are realizations of the random variables C and L of the coded bits and the LLRs, respectively. Then, for a continuous distribution of the LLRs, we have

$$I = \frac{1}{2} \cdot \sum_{c=0,1} \int_{-\infty}^{\infty} p(l|c) \cdot \log_2 \frac{2 \cdot p(l|c)}{p(l|c = 0) + p(l|c = 1)} \mathrm{d}l. \tag{3.6}$$

This expression is most conveniently computed using Monte Carlo simulation and histogram measurements. With a BSC or BEC channel, the LLRs adopt only a finite number of possible values and the integral in equation (3.6) is replaced by the sum over these values; Or we determine the error or erasure probability and use directly equation (2.24) and (2.25) for BSC and BEC channels, respectively.

Second, if we apply additional assumptions on the distributions $p(l|c)$, we can compute the mutual information as time average and avoid the numerical integration. We assume symmetric distributions $p(l|c=0) = p(-l|c=1)$, consistent distributions $p(l|c=0) = p(-l|c=1) \cdot \mathrm{e}^{l}$, and therefore

$$p(l|c=0) = p(l|c=1) \cdot \mathrm{e}^{l}. \tag{3.7}$$

The distributions are consistent if the LLRs are "correct", i.e. if they reflect the true reliability [HSL00][RSU01][GH01]. Often, distributions are not exactly consistent, e.g. due to suboptimal detectors. However, it turns out that we can mostly assume consistency and still obtain accurate results. Combining the new constraints with equation (3.6) leads to:

$$I = 1 - \int_{-\infty}^{\infty} p(l|c=0) \cdot \log_2(1+\mathrm{e}^{-l})\mathrm{d}l \tag{3.8}$$

$$= 1 - \mathrm{E}\{\log_2(1+\mathrm{e}^{-l})|c=0\}. \tag{3.9}$$

By invoking the ergodicity theorem, namely that the expectation can be replaced by the time average, we can approximate the mutual information for a large number N of samples as follows [TH02]:

$$I \approx 1 - \frac{1}{N}\sum_{i=1}^{N} \log_2(1+\mathrm{e}^{-\tilde{c}_i \cdot l_i}), \quad \text{where} \quad \tilde{c}_i = \begin{cases} +1, & c_i = 0, \\ -1, & c_i = 1. \end{cases} \tag{3.10}$$

This equation is conveniently used for LLRs with AWGN, BSC and BEC channels.

Finally, it was shown in [Lan05] that the knowledge of the values c_i is not required to compute the mutual information for the EXIT charts. The bit error probability p_i of c_i is obtained from the magnitude of the LLRs l_i by evaluating equation (2.34). Then, the average mutual information is computed for a BSC channel using equation (2.24):

$$I = 1 - \frac{1}{N}\sum_{i=1}^{N} H(p_i). \tag{3.11}$$

Fidelity Measure

Beside of the mutual information, the fidelity or measures based on the soft bits as defined in equation (2.33) provide an accurate prediction of the convergence behavior.

The fidelity is the expectation of the soft bits and it can be approximated by the time average without knowledge of the corresponding code bits [Nar01] [SBR06]:

$$M = \mathrm{E}\left\{\tilde{c} \cdot \tanh\left(\frac{l}{2}\right)\right\} = \mathrm{E}\left\{\tanh^2\left(\frac{l}{2}\right)\right\} \approx \frac{1}{N}\sum_{i=1}^{N} \tanh^2\left(\frac{l_i}{2}\right). \tag{3.12}$$

The fidelity or soft bit transfer (SOBIT) chart and the EXIT chart are closely related. This fact is clarified if we compare equations (3.10) and (3.12) for $\tilde{c} = +1$ using the following relation [Hag04]:

$$1 - \log_2(1+\mathrm{e}^{-l}) = \log_2(1+\tanh(l/2)) \approx \tanh(l/2). \tag{3.13}$$

Some Useful Functions

The relationship between mutual information, variance of LLRs and fidelity can be formally described by the so-called J, S and T functions using the Gaussian approximation, as illustrated in Fig. 3.3. These functions can either be pre-computed and stored in a look-up table, or they can be closely approximated by analytical expressions obtained using curve fitting techniques. We will use these functions extensively in Chapter 7.

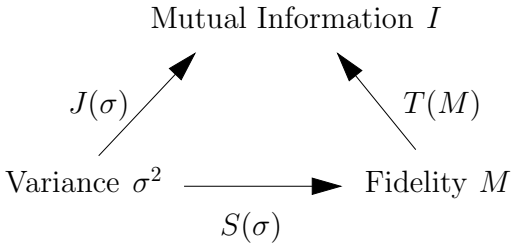


Figure 3.3: Relationship between mutual information, variance of LLRs and fidelity using the J, S, and T functions.

The J function introduced in [tB01b] and depicted in Fig. 3.4 relates the mutual information to the variance σ_L^2 of the LLRs:

$$J(\sigma) = I(\sigma_L = \sigma). \tag{3.14}$$

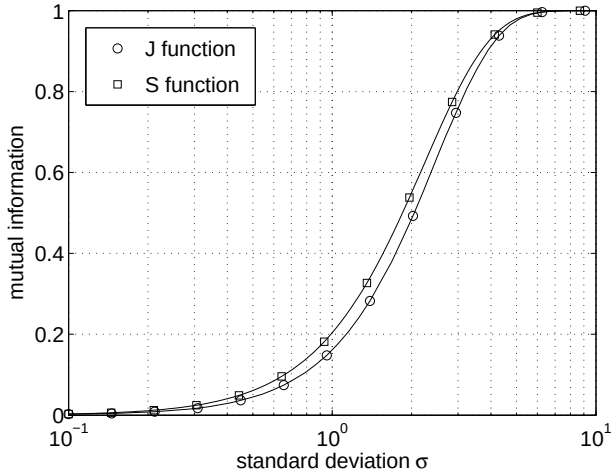


Figure 3.4: J and S functions over the standard deviation σ of the LLRs.

A good analytical approximation of the J function is derived in [BRG05]:

$$J(\sigma) = \left(1 - 2^{-H_1 \sigma^{2H_2}}\right)^{H_3} \quad \text{and} \quad J^{-1}(I) = \left(-\frac{1}{H_1} \log_2 \left(1 - I^{\frac{1}{H_3}}\right)\right)^{\frac{1}{2H_2}}, \tag{3.15}$$

with $H_1 = 0.3073$, $H_2 = 0.8935$ and $H_3 = 1.1064$. An application of the J function is to estimate the error rate from the EXIT chart: From the mutual information at the output of the receiver component, we compute the variance of the soft LLRs through the inverse J function as described in [tB01b] for a parallel concatenation and in [tB01a] for a serial concatenation. Assuming that the distribution $p(l|c)$ of the LLRs is Gaussian, the bit error rate is easily derived from the variance using the $\mathrm{Q}(\cdot)$ or $\mathrm{erfc}(\cdot)$ function. We obtain accurate results down to medium BER, i.e. down to approximately 10^{-4}.

The S function defines the relationship between the fidelity M from equation (3.12) and the variance σ_L^2 of the LLRs [SBR06]:

$$S(\sigma) = M(\sigma_L = \sigma). \tag{3.16}$$

According to equation (3.13), the fidelity is closely related to the mutual information. Therefore, it is not surprising that the J and S functions are quite similar, as shown in Fig. 3.4. The S function can be approximated with the expression from equation (3.15) and the values $H_1 = 0.4282$, $H_2 = 0.8130$ and $H_3 = 1.1699$ [SBR06].

The T function combines the J and S function and relates the mutual information I to the fidelity M [SBR06]:

$$I = T(M) = J(S^{-1}(M)). \tag{3.17}$$

EXIT Chart Properties

We highlight one important property of the EXIT chart, namely the area theorem that relates the area under EXIT functions to the code rate and capacity. We assume linear codes and independent and identically distributed (i.i.d.) sources. General expressions for non-i.i.d. sources have been derived in [AKtB04] and [Düt05]. We consider the area $\mathcal{A}$ under the EXIT function $f(I)$:

$$\mathcal{A} = \int_0^1 f(I)\mathrm{d}I. \tag{3.18}$$

With the notation from the model of Fig. 3.2, the following relationship was proven in [AKtB04] for BEC a priori information:

$$\mathcal{A} = 1 - \frac{1}{N_v} H(\mathbf{V}|\mathbf{Y}), \tag{3.19}$$

where N_v is the length of $\mathbf{V}$. If we assume an encoder with a one-to-one (invertible) mapping $H(\mathbf{V}) = H(\mathbf{C}) = H(\mathbf{U})$ and $H(\mathbf{V}|\mathbf{Y}) = H(\mathbf{C}|\mathbf{Y}) = H(\mathbf{U}|\mathbf{Y})$. Furthermore, we assume an optimal MAP detector and therefore $H(\mathbf{C}|\mathbf{Y}) = H(\mathbf{C}|\mathbf{L})$. The analysis of the outer and inner decoder in a serial concatenated system is considered.

The outer decoder receives a priori information on the coded bits but no information from the communication channel. Therefore, $\mathbf{V} = \mathbf{C}$ and $H(\mathbf{V}|\mathbf{Y}) = H(\mathbf{V})$. The length of $\mathbf{C}$ is N and the length of the i.i.d. information bit sequence $\mathbf{U}$ is K. Then, equation (3.19) becomes

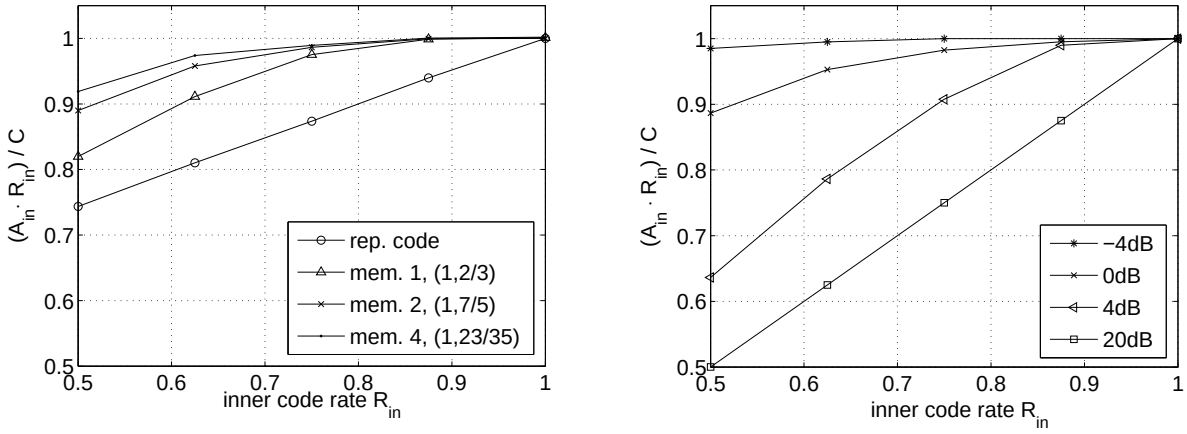
$$\mathcal{A}_{\text{out}} = 1 - \frac{1}{N} H(\mathbf{V}) = 1 - \frac{K}{N} = 1 - R_{\text{out}}. \tag{3.20}$$

The inner decoder in a serial concatenated system receives channel information and a priori information on the information bits, which correspond to the code bits of outer code. We have $\mathbf{V} = \mathbf{U}$ and equation (3.19) becomes

$$\mathcal{A}_{\text{in}} = 1 - \frac{1}{K} H(\mathbf{U}|\mathbf{Y}) = 1 - \frac{1}{K}\left(H(\mathbf{U}) - I(\mathbf{U};\mathbf{Y})\right) = \frac{I(\mathbf{U};\mathbf{Y})}{K} = \frac{I(\mathbf{U};\mathbf{Y})/N}{R_{\text{in}}}. \tag{3.21}$$

The value of $I(\mathbf{U};\mathbf{Y})/N$ is upper bounded by the capacity $\mathcal{C}$. We have equality if $N \leq K$ and $R_{\text{in}} \geq 1$. With $R_{\text{in}} = 1$, we obtain the remarkable property that the area $\mathcal{A}_{\text{in}}$ is equal to the capacity $\mathcal{C}$. With $R_{\text{in}} > 1$, the encoder mapping is no more invertible but we have no capacity loss.

Fig. 3.5 depicts the fraction $(\mathcal{A}_{\text{in}} \cdot R_{\text{in}})/\mathcal{C} = I(\mathbf{U};\mathbf{Y})/(N \cdot \mathcal{C})$ obtained from EXIT chart measurements for different code rates $R_{\text{in}} \leq 1$. We observe that $I(\mathbf{U};\mathbf{Y})/N$ approaches the capacity only at low SNR and with strong codes. However, low-complexity codes are usually applied in iterative decoding schemes and these codes will inevitably leave significant gaps between $I(\mathbf{U};\mathbf{Y})/N$ and $\mathcal{C}$. For large SNR, $I(\mathbf{U};\mathbf{Y})/N$ is bounded by R_{in} since the area $\mathcal{A}_{\text{in}}$ approaches one.



a) b)

Figure 3.5: Fraction $(\mathcal{A}_{\text{in}} \cdot R_{\text{in}})/\mathcal{C}$ as a function of the inner code rate R_{in}, BPSK modulation; **a)** repetition code and convolutional codes of different memory, $10\log_{10}(E_s/N_0)\text{dB} = 0\text{dB}$; **b)** $(1, 7/5)$ memory 2 convolutional code and different E_s/N_0 in dB.

For successful decoding, the inner code EXIT function must lie above the outer code EXIT function and we must have

$$1 - \mathcal{A}_{\text{out}} < \mathcal{A}_{\text{in}}, \tag{3.22}$$

or, using (3.20) and (3.21):

$$R_{\text{out}} R_{\text{in}} < I(\mathbf{U};\mathbf{Y})/N \leq \mathcal{C}. \tag{3.23}$$

The overall rate should be smaller than the capacity, as stated by Shannon's famous theorem. The inner code causes an inherent capacity loss if $R_{\text{in}} < 1$, since then $I(\mathbf{U};\mathbf{Y})/N < \mathcal{C}$. For $R_{\text{in}} = 1$, the area between the two curves $\mathcal{A}_{\text{in}} - (1 - \mathcal{A}_{\text{out}})$ corresponds to the rate loss $\mathcal{C} - R_{\text{out}}$. These relations are illustrated in Fig. 3.6.

If we do not apply optimal MAP decoding, we have $I(\mathbf{U};\mathbf{L}) \leq I(\mathbf{U};\mathbf{Y})$. Then, for $R_{\text{in}} = 1$, the area $\mathcal{A}_{\text{in}}$ corresponds to the receiver-constrained capacity, that is the maximum achievable rate with the not necessarily optimum detector.

These facts have important consequences for the system design, as further discussed in Section 3.4.

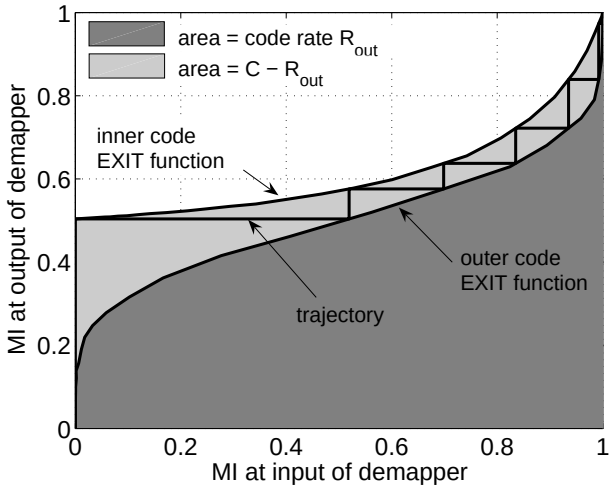


Figure 3.6: EXIT chart example of a serially concatenated system with $R_{\text{in}} = 1$ and $R_{\text{out}} = 1/2$. The area properties of the inner and outer code are illustrated.

3.2 EXIT Functions of Channel Codes

3.2.1 Examples for Various Scenarios

EXIT functions $f_{\text{dec}}(I)$ of different channel codes and for different communication and a priori channel models are investigated. Interesting observations are made and interpreted.

Different channel models

We first compare the EXIT functions for AWGN, BEC, and BSC a priori channel models for the example of an outer channel code in a serial concatenated system.

In [Lan05], upper and lower bounds on the information combined by the decoder (keyword *information combining*) are derived for repetition codes, single parity check codes and accumulators. For the single parity check code, the performance with a BEC a priori channel provides a lower bound and with BSC an upper bound. With a repetition code, the relations are reversed: the lower bound is achieved with BSC, the upper bound with BEC. The bounds for the accumulator are given in [Lan05].

Fig. 3.7 depicts EXIT functions for the three investigated a priori channels models AWGN, BEC, and BSC, and for a repetition code and a convolutional code. We observe the following: As expected from the bounds derived in [Lan05], the EXIT function for the repetition code with BEC information upperbounds the EXIT function with information from a BSC. The EXIT function with an AWGN channel lies in between. The EXIT function of the convolutional code with BEC information corresponds to a "stronger" convolutional code than with BSC information, i.e. with a higher memory and a slightly lower rate. The EXIT function for AWGN a priori information lies in between.

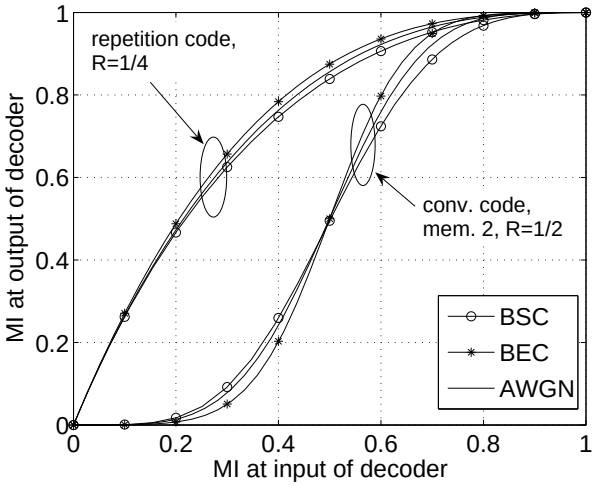


Figure 3.7: EXIT functions of outer codes in a serially concatenated system; BSC, BEC and AWGN decoder input information.

Convolutional codes with different rate and memory

EXIT functions of convolutional codes as outer codes in a serial concatenated scheme with different code parameters are investigated in Fig. 3.8. Convolutional codes with a higher memory have a lower extrinsic information output with low a priori information, but a steeper EXIT function and perform better with more a priori information. The crossover point between the EXIT functions of codes with different memory is around the mutual information corresponding to the code rate. The shape of the EXIT functions is restricted through the code rate since the area under the curves should be equal to $\mathcal{A}_{\text{out}} = 1 - R_{\text{out}}$ according to equation (3.20).

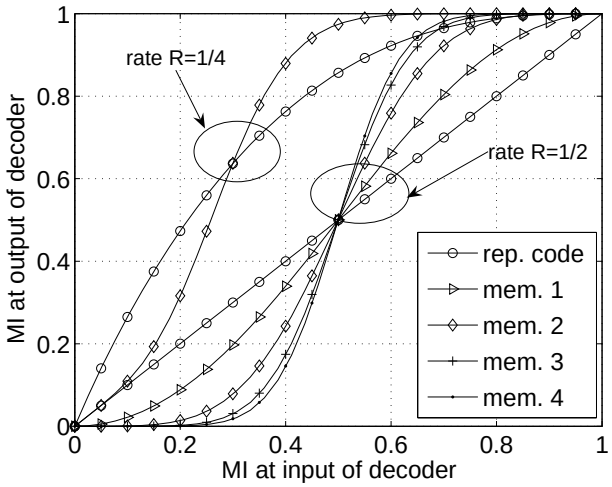
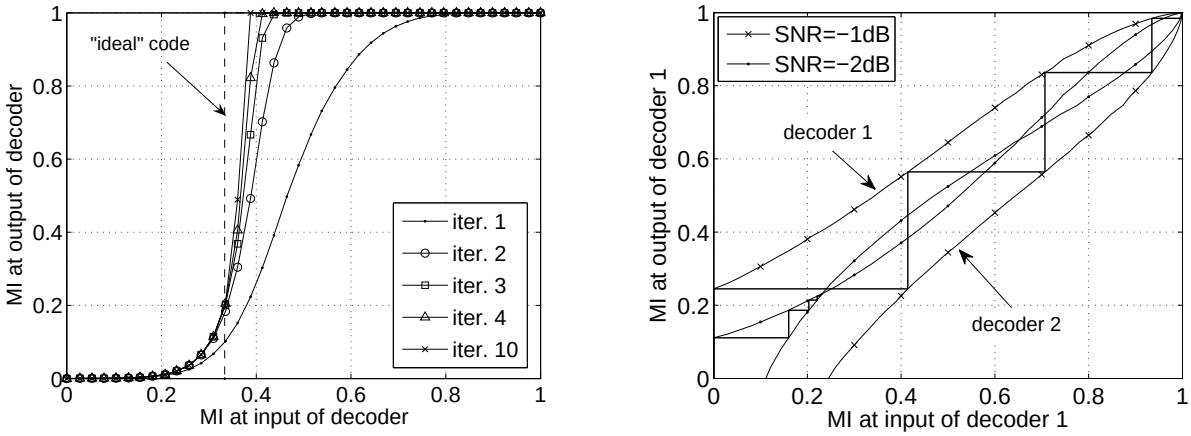


Figure 3.8: EXIT functions of rate $R = 1/2$ and $R = 1/4$ outer codes: repetition code and convolutional codes with different memory.

Parallel concatenated convolutional codes

The EXIT function of the rate $1/3$ parallel concatenated UMTS turbo code defined by the generators in equation (2.69) is investigated in Fig. 3.9. The performance after a different number of iterations is depicted in Fig. 3.9a). The area under the EXIT function is smaller than $1 - R_{\text{out}}$ since we use iterative decoding and not the optimum ML decoding. An ideal capacity achieving code would have an EXIT function that is 0 for $I < R$ and 1 for $I > R$, as depicted in Fig. 3.9a) for a rate $R = 1/3$.

Fig. 3.9b) shows the EXIT functions of the constituent codes of the turbo code and the internal iterations. Since both decoders receive channel information, both curves do not start from 0 and both curves dependent on the channel quality.



a) Turbo code b) Constituent codes

Figure 3.9: EXIT functions of the rate $R = 1/3$ parallel concatenated UMTS turbo code, information word length of 50000 bits.

Different rate-1 inner codes

Fig. 3.10 depicts rate $R = 1$ inner codes of different memory with recursive and non-recursive structures. The two questions of main relevance that arise are: When does the EXIT function of these codes start from 0 without a priori information and prevent the iterative process to start? When does the EXIT function reach the point $[1; 1]$ in the EXIT chart for low error rates after convergence of the iterative process?

Without a priori information, some rate $R = 1$ codes suffer from *catastrophic error propagation*: a finite number of channel errors may result in an unbounded number of errors on the estimated information bits. A code is catastrophic if the generator matrix does not have a feedforward inverse $\mathbf{G}^{-1}(D)$ with $\mathbf{G}(D)\mathbf{G}^{-1}(D) = D^i \cdot \mathbf{I}_K$, for some $i \geq 0$, where $\mathbf{I}_K$ is the $K \times K$ identity matrix [LC04]. For a rate-1 code, we have $\mathbf{G}(D) = f(D)/q(D)$. Then, for a non-catastrophic code, $D^i \cdot q(D)$ must divide $f(D)$ for at least one value of i to obtain a feedforward inverse $\mathbf{G}^{-1}(D) = D^i \cdot q(D)/f(D)$.

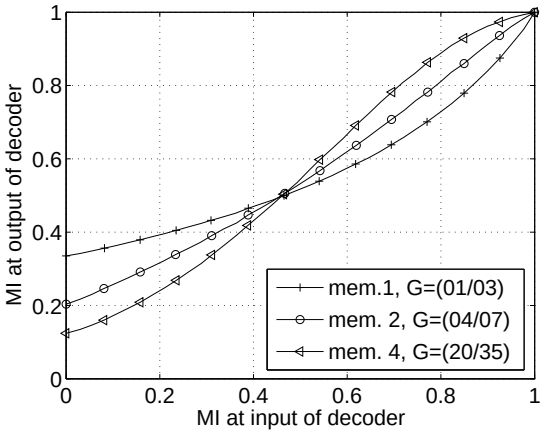
Examples of non-catastrophic rate-1 codes include the $(4/7)$ code

$$\mathbf{G}(D) = \frac{D^2}{1 + D + D^2}; \quad \mathbf{G}^{-1}(D) = 1 + D + D^2; \quad \mathbf{G}(D)\mathbf{G}^{-1}(D) = D^2, \tag{3.24}$$

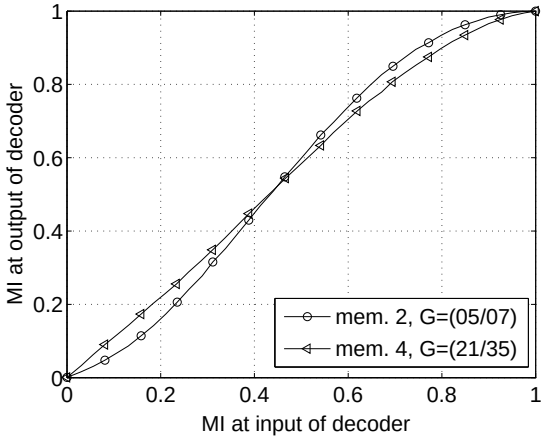
and the $(20/35)$ code

$$\mathbf{G}(D) = \frac{D^4}{1 + D^2 + D^3 + D^4}; \;\; \mathbf{G}^{-1}(D) = 1 + D^2 + D^3 + D^4; \;\; \mathbf{G}(D)\mathbf{G}^{-1}(D) = D^4. \quad (3.25)$$

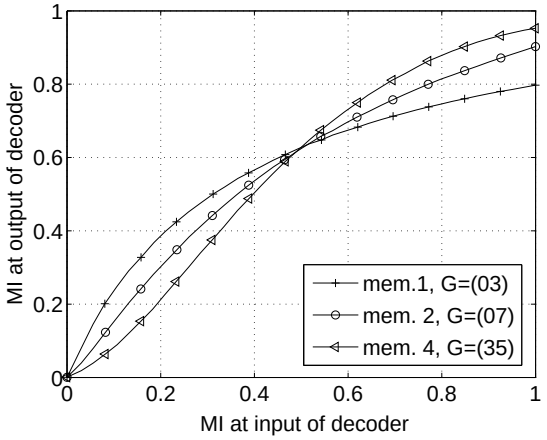
A systematic code is always non-catastrophic. The rate-1 codes can be interpreted as rate $1/2$ systematic codes where the systematic bit is punctured. With the a priori information on the information bits, the catastrophic rate-1 codes become non-catastrophic, yielding in an extrinsic output information different from zero and an ascending EXIT function. Fig. 3.10 illustrates these facts. Note that non-recursive or feedforward rate-1 codes are non-catastrophic only if $\mathbf{G}(D) = D^i$ for some $i \geq 0$. Then, the encoder output is simply a delayed version of the encoder input and a priori information has no influence on the extrinsic decoder output. A method to let the EXIT function of a catastrophic rate-1 code start at a value above zero is the insertion ("doping") of information bits at a certain ratio to the coded bits.



a) Recursive, non-catastrophic



b) Recursive, catastrophic



c) Non-recursive, catastrophic

Figure 3.10: EXIT functions of different rate $R = 1$ inner codes; $10\log_{10}(E_s/N_0)\text{dB} = 1\text{dB}$, BPSK modulation.

With ideal a priori information, it was shown in [BDMP98] that only recursive inner codes achieve a so called *interleaver gain*, where the error rate at high SNR tends towards zero for an interleaver length going towards infinity. Then, the LLRs are unbounded and the EXIT function will approach the point $[1; 1]$ in the EXIT chart [tB01a].

This fact can be intuitively explained as follows: Assume without loss of generality that the all zero information word $\mathbf{u} = (0, \ldots, 0)$ is transmitted. With ideal extrinsic a priori information, all information bits but the bit to be detected at discrete time i are a priori known. Therefore, the decoder has to decide only between the two remaining code words corresponding to $u_i = 0$ and $u_i = 1$. With a non-recursive rate-1 code, the information bit u_i influences only $M_c + 1$ code bits, where M_c is the code memory. For a recursive code, the value of u_i influences all remaining code bits of the codeword. Therefore, the Hamming distance between the two remaining code words with ideal a priori information is limited by the code memory for non-recursive codes and by the code word length for recursive codes. For sufficiently large block length, the Hamming distance with a recursive code becomes large enough to output a mutual information close to one. Comparing Fig. 3.10a), b) and c), we observe as expected that the EXIT functions of the recursive codes reach the point $[1; 1]$ and that the value $f(1)$ with non-recursive codes depends on the code memory.

3.2.2 Bit-Level EXIT Functions

The idea of bit-level EXIT functions of channel codes is to generate a separate EXIT function for each characteristic bit position. Three examples are shown in Fig. 3.11.

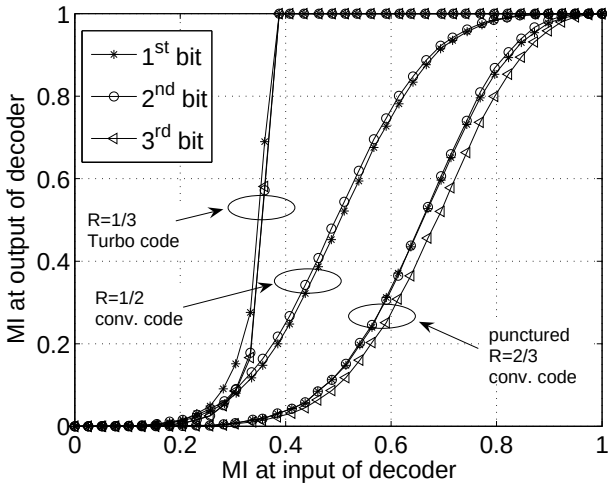


Figure 3.11: Bit-level decoder EXIT functions: rate $R = 1/3$ UMTS turbo code, 10 iterations; $R = 1/2$ convolutional code $(1, 5/7)$; $R = 2/3$ convolutional code punctured from $(1, 5/7)$ code with periodic puncturing pattern (1110), i.e. every second parity bit is punctured.

First, with the rate $R = 1/3$ parallel concatenated turbo code, we observe that the systematic bit has a higher reliability than the two parity bits, especially with low input mutual information. We explain this by the fact that in contrast to the parity bits, both constituent decoders receive the channel information on the systematic bits.

Then, with the systematic rate $R = 1/2$, $(1, 5/7)$ convolutional code, the systematic bit is less reliable than the parity bit. By considering the trellis of the code in Fig. 2.10, we find out that for the two most probable error paths of Hamming weight 5, 3 bit errors are made on the systematic bits and 2 on the parity bits, resulting in the respective reliability. For the two error paths of weight 6, the systematic and parity bit encounter on average the same number of bit errors.

Finally, we consider the $(1, 5/7)$ convolutional code punctured to rate $R = 2/3$ and observe the lower reliability of the systematic bit of which the corresponding parity bit is punctured.

This additional information on the reliability of the characteristic bit positions can be exploited when the code is combined with other components that expect a certain reliability on specific bit positions.

3.2.3 Analytical Computation

We are highly interested in closed form expressions of decoder EXIT functions to avoid simulations, obtain exact decoding thresholds, and derive new design guidelines. Analytical expressions of decoder EXIT functions for simple block codes have been derived for BEC and BSC channels in [AKtB04] [LHH04].

We derived closed form expressions of decoder EXIT functions of convolutional codes with BEC channels in [SB04a] using a similar approach than proposed in [MU02][Méa06]. The results are based on work done to determine polynomial expressions for the exact erasure and error probability of convolutional codes for asymptotically large code word length on BEC and BSC channels. Results for BSC channels and Viterbi decoding are derived in [BBL$^+$95][LTZ04]. We consider BEC channels and symbol-by-symbol MAP decoding as investigated in [KSW03]. The key observation in these papers is that the number of possible state probability vectors in the decoding algorithm is finite and reasonably small.

Example 6 *Consider the rate $R = 1/2$, memory 2, $(1, 5/7)$ systematic convolutional code with the trellis depicted in Fig. 2.10 as constituent code of a PCCC. Assuming that the all zero word is transmitted over a BEC, the five possible state probability vectors are*

$$\mathcal{S} = \left\{ (1,0,0,0), (\frac{1}{2},\frac{1}{2},0,0), (\frac{1}{2},0,\frac{1}{2},0), (\frac{1}{2},0,0,\frac{1}{2}), (\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}) \right\}.$$

Here, the state probability vector $(1/2, 1/2, 0, 0)$ means that the encoder has the probability $1/2$ of being in the trellis state 00 or 01.

We construct a Markov chain with the distinct state probability vectors as Markov states. The erasure probabilities given in the system model of Fig. 3.12 for a systematic rate $R = 1/2$ convolutional code are used to determine in the following the Markov state transition probabilities.

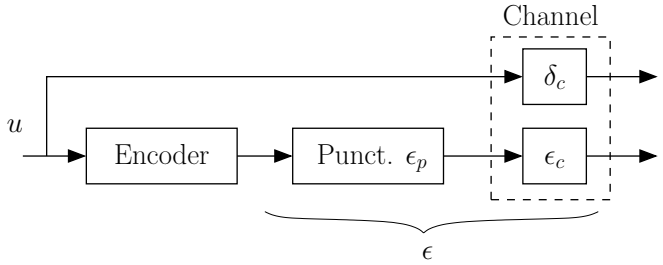


Figure 3.12: Rate $R = 1/2$ convolutional code model for EXIT chart computation with BEC channels.

- δ_c: channel erasure probability of systematic bits
- δ_a: a priori erasure probability of systematic bits
- ϵ_c: channel erasure probability of parity bits
- ϵ_p: erasure probability of parity bits from random puncturing

The erasure probability of the parity bits after puncturing and the channel is $\epsilon = \epsilon_p + (1-\epsilon_p)\epsilon_c$. The probabilities of the four possible erasure events are:

- systematic and parity bit erased: $p^{(00)} = \delta_a\delta_c\epsilon$
- systematic bit erased, parity bit received $p^{(01)} = \delta_a\delta_c(1-\epsilon)$
- systematic bit received, parity bit erased $p^{(10)} = (1-\delta_a\delta_c)\epsilon$
- systematic and parity bit received $p^{(11)} = (1-\delta_a\delta_c)(1-\epsilon)$

To obtain the EXIT function of a constituent code of a PCCC, we fix δ_c and ϵ and compute the decoder output for different δ_a. For an outer code of a SCCC, we set $\delta_a = 1$ and compute the decoder output for different $\delta_c = \epsilon_c$. Code rates between $1/2 \leq R \leq 1$ are obtained by setting ϵ_p. For the special case of a rate $R_{in} = 1$ inner code of a SCCC, we set $\delta_c = 1$, $\epsilon_p = 0$, fix ϵ_c and compute the decoder output for different δ_a.

For the forward and backward recursions of the BCJR symbol-by-symbol MAP algorithm, we determine the Markov transition probability matrices $\mathbf{P}^{(\alpha)}$ and $\mathbf{P}^{(\beta)}$, respectively. The i,j matrix element represents the probability of making the transition from the ith to the jth Markov state.

***Example** 7 For the investigated example with a constituent code of a PCCC, the forward transition probability matrix is with the probabilities given above:*

$$\mathbf{P}^{(\alpha)} = \begin{bmatrix} 1-p^{(00)} & p^{(00)} & 0 & 0 & 0 \\ p^{(11)} & 0 & p^{(01)} & p^{(10)} & p^{(00)} \\ 0 & 1 & 0 & 0 & 0 \\ p^{(11)} & 0 & p^{(10)} & p^{(01)} & p^{(00)} \\ 0 & p^{(11)} & 0 & 0 & 1-p^{(11)} \end{bmatrix}. \tag{3.26}$$

The steady-state distribution $\pi^{(\alpha)}$ of the Markov chain for the forward recursion is the solution to

$$\pi^{(\alpha)} = \mathbf{P}^{(\alpha)^T} \cdot \pi^{(\alpha)} \quad \text{with} \quad \sum_i^{|\mathcal{S}|} \pi_i^{(\alpha)} = 1, \tag{3.27}$$

where $\pi_i^{(\alpha)}$ represents the ith element of the row vector $\pi^{(\alpha)}$; $\pi^{(\beta)}$ for the backward recursion is computed similarly.

The transition matrix $\mathbf{P}^{(\gamma)}$ with elements $P_{i,j}^{(\gamma)}$ determines whether an information bit is erased depending on the knowledge of the code bit and for the transition from the ith to the jth Markov state. We distinguish between three scenarios: the information bit is always erased ($P_{i,j}^{(\gamma)} = 1$), never erased ($P_{i,j}^{(\gamma)} = 0$), or erased only if the code bit is erased ($P_{i,j}^{(\gamma)} = \epsilon$).

***Example** 8 For the investigated example, the transition matrix is*

$$\mathbf{P}^{(\gamma)} = \begin{bmatrix} 0 & \epsilon & 0 & 0 & \epsilon \\ 0 & \epsilon & 1 & 0 & 1 \\ \epsilon & \epsilon & \epsilon & \epsilon & \epsilon \\ 0 & \epsilon & 0 & 1 & 1 \\ \epsilon & \epsilon & 1 & 1 & 1 \end{bmatrix}.$$

Finally the erasure probability after the decoder is

$$p_e = \pi^{(\alpha)} \cdot \mathbf{P}^{(\gamma)} \cdot \pi^{(\beta)}.$$

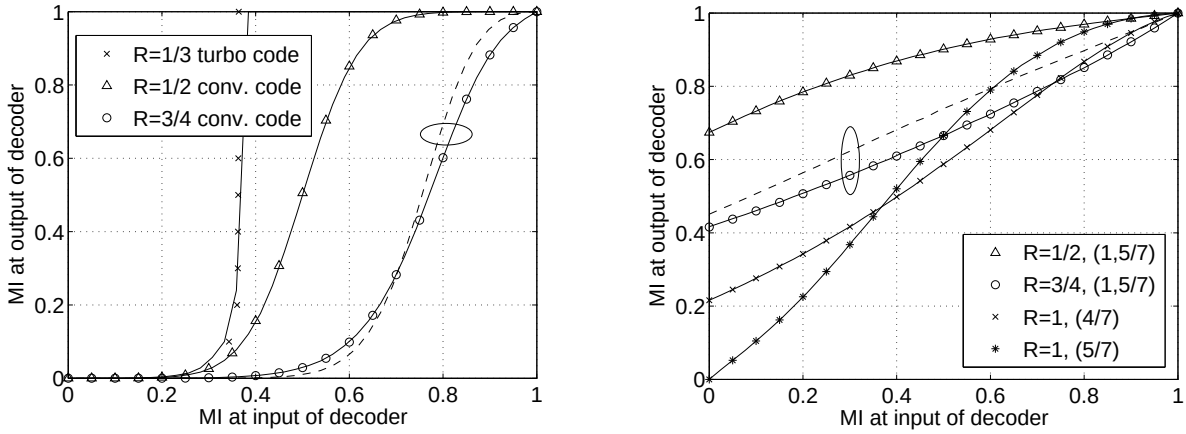
and the mutual information $I = 1 - p_e$.

Example 9 *For the investigated example with $\delta_a = 1$ and $\epsilon = \epsilon_c = \delta_c$ (outer code of a SCCC), the extrinsic erasure probability is:*

$$p_e = \frac{3\epsilon^4 - 18\epsilon^5 + 56\epsilon^6 - 107\epsilon^7 + 132\epsilon^8 - 101\epsilon^9 + 42\epsilon^{10} - 3\epsilon^{11} - 4\epsilon^{12} + \epsilon^{13}}{(1 - 4\epsilon + 7\epsilon^2 - 5\epsilon^3 + 3\epsilon^5 - \epsilon^6)^2}.$$

The polynomial degree increases rapidly with the code memory but the results are manageable up to memory 4 convolutional codes for BEC channels.

Fig. 3.13 depicts computed and simulated decoder EXIT functions for different scenarios. The EXIT function of the overall turbo code is obtained by determining the crossing point of the analytical constituent codes. The curves for $R = 3/4$ codes obtained by random puncturing the rate $R = 1/2$ mother code ($\epsilon_p = 2/3$) match exactly the simulated curves, but the performance is definitely lower than with regular puncturing. The other analytical curves correspond precisely to the simulated curves.



a) b)

Figure 3.13: Analytical (markers) and simulated (lines) EXIT functions for BEC channels; dashed lines correspond to regular puncturing; **a)** outer codes in a SCCC: $R = 1/2$ and $R = 3/4$, $(1, 15/17)$ convolutional code and $R = 1/3$ UMTS turbo code defined in equation (2.69); **b)** $R = 1/2$ ($\delta_c = \epsilon_c = 0.45$) and $R = 3/4$ ($\delta_c = \epsilon_c = 0.25$) memory 2 constituent code of a PCCC; $R = 1$, $(4/7)$ and $(5/7)$ inner code of a SCCC ($\epsilon_c = 0.4$), cp. to Fig. 3.10.

The analytical computation of EXIT charts with similar methods is also possible for BSC channels. However, the number of Markov states is much larger than with BEC channels. For a memory 4 encoder, the number of Markov states becomes as large as 25641 with BSC channels [LTZ04], while this number is 67 for an example given in [KSW03].

3.3 EXIT Functions of Mappings

We consider the EXIT function $f_{\text{dem}}(I)$ of the demapper integrated in a bit-interleaved coded modulation scheme with iterative demapping and decoding (BICM-ID) as described in Fig. 2.22. The major parameters that affect $f_{\text{dem}}(I)$ are the channel characteristics, the signal constellation and the bit-to-symbol mapping. The characterization with an Euclidean distance spectrum and the optimization with a binary switching algorithm of mappings for different signal constellations will be investigated in detail in Chapter 4. In this Section, we focus on the characteristics and the computation of average, bit-level and symbol-level demapper EXIT functions for different scenarios.

The demapper can be truly considered as a rate-1 inner code in the serial concatenation of a channel code, interleaver and mapper in the BICM scheme since it may introduce dependencies between the M bits associated to one symbol.

Three characteristic values of the demapper EXIT function $f_{\text{dem}}(I)$ are of interest: First, a large value of $f_{\text{dem}}(0)$ without a priori information is desired to avoid an early crossing with the decoder EXIT function which would cause the iterative process to stop. Second, the value of $f_{\text{dem}}(1)$ with ideal a priori information (genie or error free feedback case) determines the gain over the iterations, i.e. the performance at high SNR after several iterations. Finally, the area $\mathcal{A}_{\text{dem}}$ under the demapper EXIT function is of interest since it is related to the capacity as follows: In equation (3.19), $\mathbf{V}$ is the length $N_v = M$ sequence of bits associated to the complex symbol X and Y is the received complex symbol. The invertible one-to-one mapping does not introduce a loss of information and we consider an optimal MAP demapper. Following the derivation of equation (3.21), the area under the rate $R_{\text{in}} = 1$ demapper EXIT function is:

$$\mathcal{A}_{\text{dem}} = I(\mathbf{V}, Y)/M = I(X, Y)/M \quad \in [0, 1]. \tag{3.28}$$

$I(X, Y)$ is the constellation constrained capacity. The important fact to note is that the capacity and the area under the EXIT function depend only on the signal constellation, but are independent on the applied mapping. This restricts the shape of the EXIT function for a given signal constellation and we have a trade-off between the value $f_{\text{dem}}(0)$ without a priori information and $f_{\text{dem}}(1)$ with ideal a priori information. For different mappings, we expect to have either a good performance at low SNR with a high $f_{\text{dem}}(0)$ or a low error bound with a high $f_{\text{dem}}(1)$ at high SNR and after several demapping and decoding iterations. Note that the mapping as inner code does not have a recursive structure. Therefore, no interleaver gain is possible and the EXIT function will not reach the point $[1, 1]$ in the EXIT chart.

These facts are illustrated in the following examples. All investigated mappings are defined in the tables of Appendix A.

3.3.1 Examples for Various Scenarios

We analyze the demapper EXIT function $f_{\text{dem}}(I)$ for different mappings, signal constellations and channel models.

Fig. 3.14 shows demapper EXIT functions with 16QAM and Gray mapping, Ungerböcks set partitioning (SP) mapping [Ung82] and the turbo optimized mapping $M16a$ that maximizes the value $f_{\text{dem}}(1)$, see Chapter 4. With Gray mapping, the value $f_{\text{dem}}(0)$ is maximized, but the a priori information has only a negligible influence on the demapper performance and the slope of the EXIT function is very low. The best performance with large a priori information is expected with the turbo optimized $M16a$ mapping. Set partitioning mapping is not well suited for iterative demapping and decoding because of the convex shape of the EXIT function that results in both a low $f_{\text{dem}}(0)$ and a low $f_{\text{dem}}(1)$. Note again that the area $\mathcal{A}_{\text{dem}}$ is independent of the applied mapping.

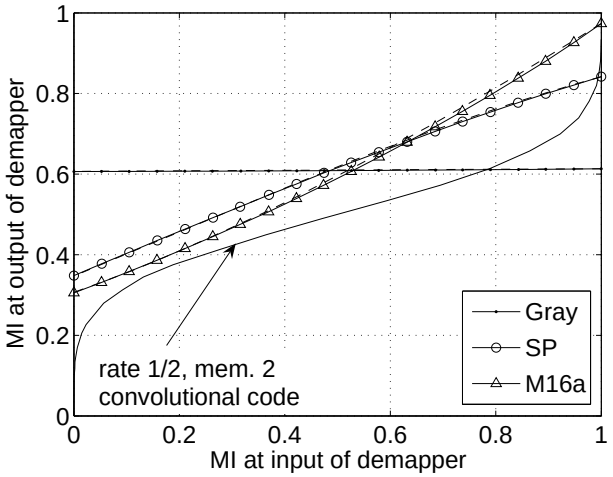


Figure 3.14: EXIT functions of 16QAM with different mappings, AWGN channel, $10\log_{10}(E_s/N_0)\text{dB} = 7\text{dB}$. Solid lines: simulation, dashed lines: numerical computation with BEC a priori information.

For different SNR values, the capacity and therefore the area under the demapper EXIT function varies, as illustrated for the Gray and turbo-optimized $M16a$ mapping in Fig. 3.15.

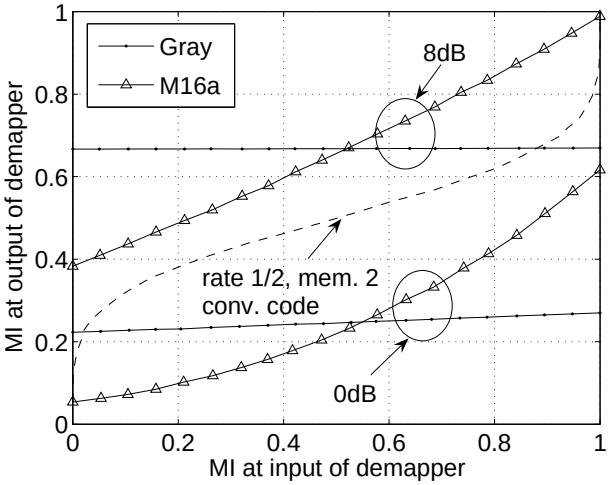


Figure 3.15: EXIT functions with 16QAM, AWGN channel, different values of E_s/N_0 in dB.

Fig. 3.16 depicts demapper EXIT functions with Gray mapping and different signal constellations. The crossing point between the curves of the demapper and the decoder are shifted towards the point $[1, 1]$ for smaller signal constellations, resulting in an increased reliability of the transmission with the cost of a reduced throughput.

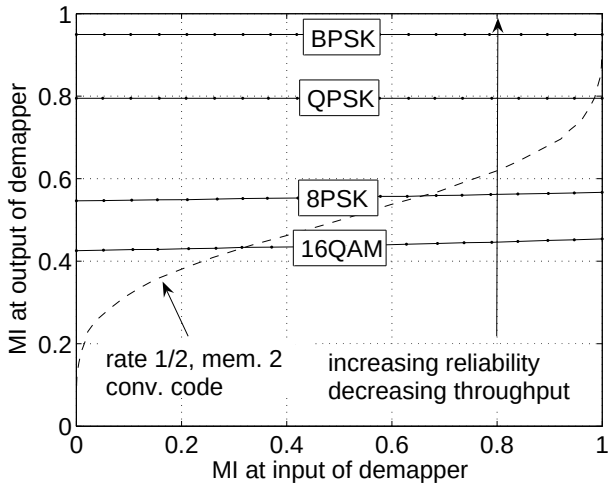


Figure 3.16: EXIT functions of different signal constellations with Gray mapping, AWGN channel, $10\log_{10}(E_s/N_0)\text{dB} = 4\text{dB}$.

In Fig. 3.17, we compare demapper EXIT functions for an AWGN channel and a Rayleigh fading channel with independent fading coefficients for every symbol. 16QAM with Gray and the turbo-optimized $M16a$ mapping are considered. The different SNRs are set to have a similar capacity with the Rayleigh fading and AWGN channel in order to obtain the same area under the EXIT function for a better comparison. We observe that with the $M16a$ mapping, the EXIT function is not very robust against channel model variations: the slope of the function is lower with the Rayleigh than with the AWGN channel, requiring a different channel code for optimized iterative decoding.

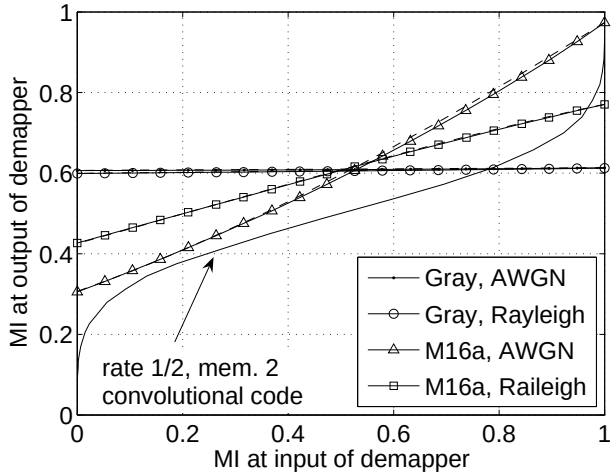
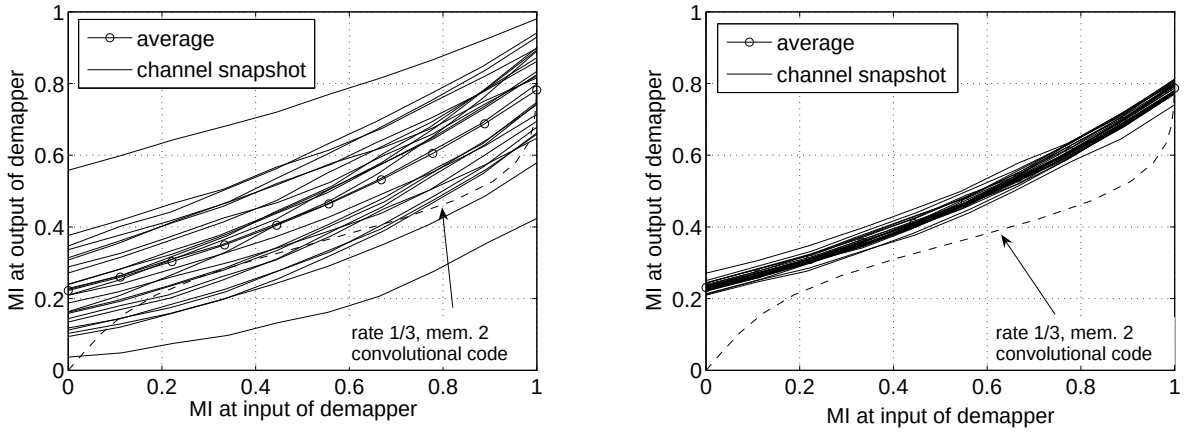


Figure 3.17: EXIT functions of Gray and $M16a$ 16QAM mappings for a Rayleigh channel ($10\log_{10}(E_s/N_0)\text{dB} = 9.1\text{dB}$) and AWGN channel ($10\log_{10}(E_s/N_0)\text{dB} = 7\text{dB}$) with rate $1/2$ code. Solid lines: simulation, dashed lines: numerical computation with BEC a priori information.

A symbol-level fading channel model is given only if we assume a large interleaver, a channel changing with the symbol rate, or with ideal frequency hopping. Otherwise, we have to consider the coherence time T_{coh} that describes the time duration over which the fading coefficient is essentially invariant, as defined in equation (2.2). For a large coherence time, we approach a block fading channel model, where we assume that the fading coefficient is constant over the duration of one interleaver frame. Fig. 3.18 shows sets of EXIT functions for qualitative investigations with different values of the coherence time. The variance of the EXIT functions is larger for a large coherence time. The EXIT function of the symbol-level fading model is approached for a short coherence time. When the coherence time is large, we can either set the code rate for a worst case scenario, which is really a waste of bandwidth, or we accept a certain error rate and use ARQ techniques to enable reliable transmission.



a) Large T_{coh} b) Small T_{coh}

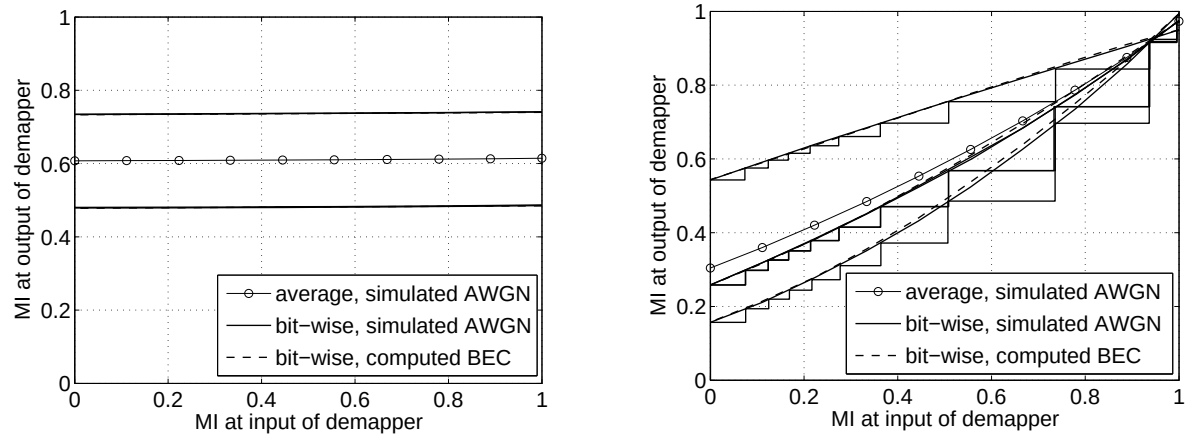
Figure 3.18: Demapper EXIT functions of 16QAM, $M16a$ mapping, fading channel with different coherence time T_{coh}; $10\log_{10}(E_s/N_0)\text{dB} = 6\text{dB}$.

3.3.2 Bit- and Symbol-Level EXIT Functions

The *bit-level* demapper EXIT function reflects the average EXIT characteristics of a single bit position $m = 1, \ldots, M$ in the binary label. Fig. 3.19a) and 3.19b) depict the EXIT functions and the decoding trajectory split up in the four bit positions of a Gray and $M16a$ 16QAM mapping, respectively. These plots illustrate the different reliability of the bit positions. The trajectories reveal that the output of the decoder corresponding to the a priori input of the demapper is similar for all bit positions. This fact results from the averaging over the different bit position reliabilities through the interleaver and decoder.

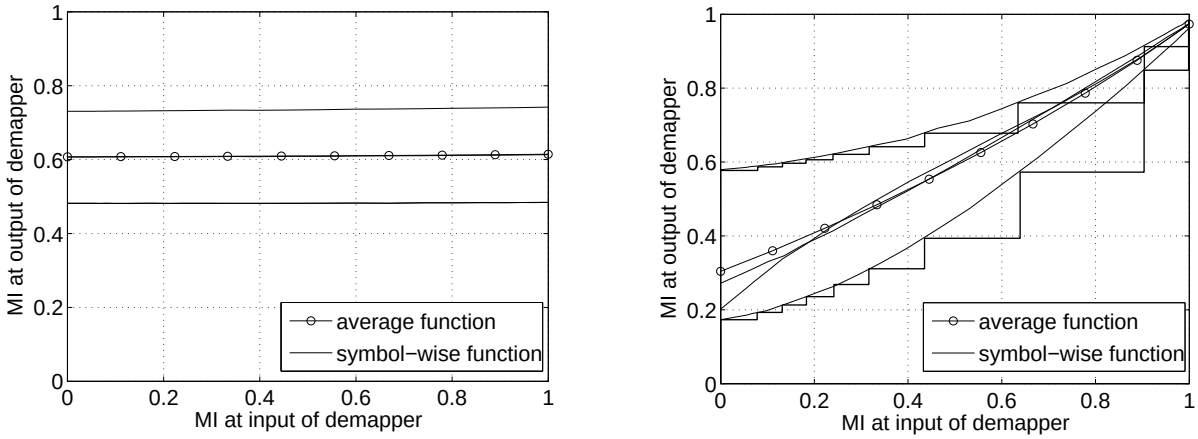
A *symbol-level* description of mappings is obtained if we fix the transmitted signal point and observe the LLRs at the output of the demapper as a function of the a priori information. To compute the symbol-level mutual information, we should use equation (3.10) to get rid of the problems that arise with a fixed transmitted bit sequence. In Fig. 3.20a) and 3.20b), we observe that the 16 symbols of the 16QAM Gray and $M16a$ mappings are grouped in three and four groups with similar characteristics, respectively. These plots illustrate the different reliability of the symbols.

Note that the full characterization of a signal constellation would include the reliabilities of all $M \cdot 2^M$ bit positions: With e.g. 16QAM, we would have to consider separately $4 \cdot 16 = 64$ bits positions.



a) Gray mapping b) $M16a$ mapping

Figure 3.19: Bit-level demapper EXIT functions with trajectory; $10\log_{10}(E_s/N_0)\text{dB} = 7\text{dB}$. Solid lines: simulation, dashed lines: numerical computation with BEC a priori information.



a) Gray mapping b) $M16a$ mapping

Figure 3.20: Symbol-level demapper EXIT functions with trajectory; $10\log_{10}(E_s/N_0)\text{dB} = 7\text{dB}$.

3.3.3 Numerical Computation

We derive a numerical solution for demapper EXIT functions without Monte Carlo simulations. The derivation is useful for a better understanding and to verify the results. We consider a AWGN and symbol-level fading channel with BEC a priori information at the demapper. The approach is based on the capacity considerations of Section 2.5 and the ideas presented in [tB01c].

We start from equation (2.81) and (2.83) and average over the 2^L possible a priori known bit values $\mathbf{c}^L$ to obtain the average mutual information I_L conditioned on the number $0 \leq L \leq M-1$ of a priori known bits:

$$I(X,Y) \;=\; \sum_{L=0}^{M-1} I_L \;=\; \sum_{L=0}^{M-1} \frac{1}{M} \frac{1}{\binom{M-1}{L}} \frac{1}{2^L} \sum_{m=1}^{M} \sum_{j=1}^{\binom{M-1}{L}} \sum_{\forall \mathbf{c}^L} I(Y, C_m|\mathbf{c}^L). \tag{3.29}$$

If no a priori information is available ($L = 0$), we only cumulate over the M bit positions:

$$I(X;Y) \geq \sum_{m=1}^{M} I(Y;C_m). \tag{3.30}$$

$I(Y;C_m|\mathbf{c}^L)$ describes the information that can be gained by the demapper on the transmitted bit C_m by observing the received symbol Y and using the a priori known bits $\mathbf{c}^L$. $I(Y;C_m|\mathbf{c}^L)$ is evaluated by numerical integration over the complex signal space and by averaging over the two possible values $\{0;1\}$ of C_m. We start from the general definition of the mutual information given in equation (2.17), use Bayes' rule and the following facts for simplification, similar as done in equation (3.6):

$$p(y,c_m|\mathbf{c}^L) = p(y|c_m,\mathbf{c}^L)\cdot p(c_m); \quad p(c_m) = \frac{1}{2}; \quad p(y|\mathbf{c}^L) = \frac{1}{2}\left(p(y|0,\mathbf{c}^L) + p(y|1,\mathbf{c}^L)\right). \tag{3.31}$$

Then we have

$$\begin{aligned} I(Y;C_m|\mathbf{c}^L) &= \mathrm{E}_{c_m,h,y}\left(\log_2\frac{p(y|c_m,\mathbf{c}^L)}{p(y|\mathbf{c}^L)}\right) \qquad (3.32)\\ &= \frac{1}{2}\cdot\sum_{c_m=\{0,1\}}\int_{\mathbb{C}} p(h)\int_{\mathbb{C}} p(y|c_m,\mathbf{c}^L)\cdot\log_2\frac{2\cdot p(y|c_m,\mathbf{c}^L)}{p(y|0,\mathbf{c}^L)+p(y|1,\mathbf{c}^L)}\,\mathrm{d}y\,\mathrm{d}h, \end{aligned}$$

with

$$p(y|c_m,\mathbf{c}^L)) = \frac{1}{2^{M-(L+1)}}\cdot\sum_{x\in\mathcal{X}_m^{c_m,\mathbf{c}^L}} p(y|x), \tag{3.33}$$

where $\mathcal{X}_m^{c_m,\mathbf{c}^L}$ denotes the subset of symbols x of the signal constellation $\mathcal{X}$ whose bit labels have the a priori known values $\mathbf{c}^L$ in L positions and the value c_m in position $m \in \{1,\ldots,M\}$. $|\mathcal{X}_m^{c_m,\mathbf{c}^L}| = 2^{M-(L+1)}$ since there are $M-(L+1)$ undetermined bit values. $p(y|x)$ is given by the two-dimensional Gaussian distribution:

$$p(y|x,h) = \frac{1}{2\pi\sigma_n^2}\exp\left(-\frac{|y-h\cdot x|^2}{\sigma_n^2}\right), \tag{3.34}$$

where σ_n^2 is the noise variance per dimension. For an AWGN channel, the integration in equation (3.32) over the fading coefficient h can be omitted.

In the special case of ideal a priori information, where $L = M-1$ bits are a priori known, the set $\mathcal{X}_m^{c_m,\mathbf{c}^L}$ is reduced to one symbol. Since the signal constellation is then reduced to a one-dimensional BPSK constellation, the corresponding LLRs are easy to obtain and the mutual information can be obtained through numerical integration over the LLR distributions as described in [Muh04] [MS05].

Recall that I_L from equation (3.29) is the average conditional mutual information given that L bits out of a maximum of $M-1$ bits are a priori known at the demapper, $0 \leq L \leq M-1$. This corresponds to BEC a priori information with erasure probability $\epsilon = 1 - L/(M-1)$. With the mutual information of the BEC channel given in equation (2.25), we obtain the following values of the demapper EXIT function:

$$f_{\text{dem}}(L/(M-1)) = I_L. \tag{3.35}$$

For a transition to continuous a priori knowledge, we consider a sequence of k symbols and associated bits to obtain a smaller granularity $1/(kM-1)$ of a priori known bits [tB01c]. Then, we have the EXIT function values $f_{\text{dem}}(L'/(kM-1))$ for the a priori mutual information $1-\epsilon = L'/(kM-1)$ and obtain a continuous function for $k \to \infty$.

Numerically computed EXIT functions with AWGN and Rayleigh channel are compared to the simulation results in Fig. 3.14 and 3.17. The results coincide well, even though a BEC a priori information was assumed for the numerical computation.

The numerically computed bit-level EXIT functions are obtained if we omit the averaging over the M bit positions in equation (3.29). The simulated and numerically computed EXIT functions are compared in Fig. 3.19 and we observe again a good match.

An other approach to compute the EXIT functions of mappings has been presented in [QZZW05]. To avoid the numerical integration over the signal space, a hard decision virtual communication channel is defined that has the same capacity than the real AWGN or Rayleigh fading channel. This approach leads to a good approximation of the demapper EXIT functions.

3.4 Design of Communication Systems Using EXIT Charts

According to the area properties given in equations (3.20) and (3.21) in Section 3.1, we would approach capacity with iterative decoding of a serially concatenated system if the EXIT functions of the receiver components precisely match without crossing since the area between the curves is the distance to capacity and an early crossing would result in a high error rate. The design of capacity approaching iterative systems is therefore reduced to a *curve-fitting problem* [AKtB04][TH02] where we have to carefully shape the EXIT functions. Furthermore, we have observed in Section 3.1 that an inner code with a rate R_{in} less than one implies an inherent capacity loss. Therefore, R_{in} should be larger or equal to one.

For the considered BICM-ID scheme with iterative demapping and decoding, we have to match the curve of the demapper to the curve of the decoder to optimize the performance. The main parameters of the demapper EXIT function are the signal constellation and the mapping. With a convolutional code, the code rate and the code memory are most relevant. With a given signal constellation and code rate, the shape of the EXIT functions is restricted since the area under the curves does not change with e.g. the mapping or the code memory.

The decoder curve can be further shaped if we apply an irregular puncturing [Tüc04]. Then, different puncturing rates are applied within a code word and the overall decoder EXIT function is the linear combination of the EXIT functions of the punctured codes of different rate. Similarly, we can use irregular or hybrid signal constellations and mappings to shape the EXIT function of the demapper [SB06], as investigated in detail in Chapter 5. A further possibility to adapt EXIT functions would be a non-uniform power distribution of the transmitted symbols within one code word.

From the observations made in Section 3.2 and 3.3, we state that the steeper the EXIT function of the demapper is, the less powerful the channel code has to be and the more iterations between the decoder and demapper are required. With Gray mapping as example, iterations between decoder and demapper are not necessary but a powerful capacity approaching code is required. The pragmatic approach described in [GGB94] combines a powerful turbo code with Gray mapping.

We can generalize these observations to the following trade-off: if we use low complexity components in iterative schemes, a large number of iterations is required. With more complex and powerful components, a reduced number of iterations is required to achieve a similar performance.

For low-density parity check (LDPC) codes [Gal62] as example, the constituent codes can be interpreted as repetition and single parity check codes with very low decoding complexity. However, the number of required iterations is higher than with concatenated convolutional codes. Another example is the use of irregular codes to optimize the iterative receiver. With irregular codes, we deliberately lower the performance of the code itself but make it better suited for the iterative receiver. However, optimized irregular codes require more iterations than regular ones.

Another issue is the trade-off between low complexity receivers and low achievable error rate, illustrated in Fig. 3.21. In practical systems, we may set this trade-off according to the available resources and the varying quality of service (QoS) requirements. The complexity is affected by the number of required decoding iterations and the complexity of the components. As already mentioned, an early crossing of the EXIT functions would result in a higher error rate. In the considered BICM system, we can use only Gray mapping and implement a ”strong” or ”weak” code for low error rates or low complexity, respectively. Or we only use a ”weak” channel code and vary the mapping. The advantage of the second approach is that the change of the mapping characteristic is easier to implement than different channel codes. The disadvantage is that we need the feedback from the decoder to the demapper and that the performance might be lower than with ”strong” channel codes.

Beside the analysis and optimization of iterative receivers, the EXIT chart provides the possibility to easily measure e.g. the receiver-constrained capacity. We generate the EXIT function of the considered receiver component that may use a sub-optimal estimation algorithm. The area under the EXIT function then approximates the receiver-constrained capacity as long as the LLR pdf does not strongly differs from the assumed distributions given by equation (3.2) and (3.3).

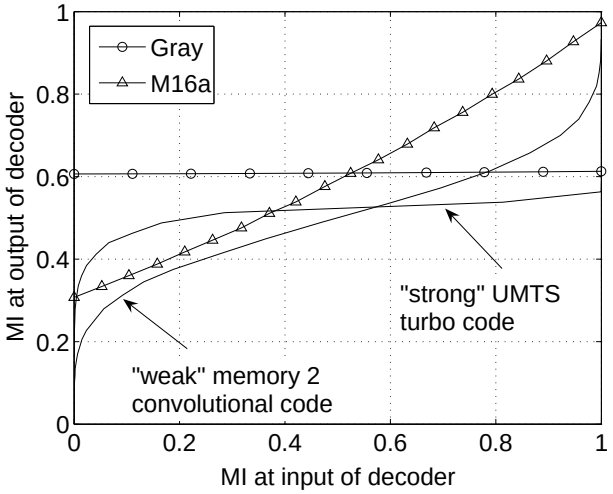


Figure 3.21: EXIT chart: combinations of Gray and turbo optimized (M16a) mapping with ”weak” and ”strong” rate $R = 1/2$ codes; $10\log_{10}(E_s/N_0)\text{dB} = 7\text{dB}$, AWGN channel.

3.5 Summary

The EXIT chart was investigated as a tool to analyze the convergence properties of iterative receivers and in particular of bit-interleaved coded modulation with iterative demapping and decoding (BICM-ID).

We reviewed the EXIT chart construction and area properties. The capacity loss of rate $R_{\text{in}} < 1$ inner codes in serially concatenated schemes, properties of catastrophic and non-catastrophic rate $R_{\text{in}} = 1$ inner codes and the impact of different channel models (BSC, BEC, and AWGN) on the EXIT functions were studied. We derived analytic decoder EXIT functions of convolutional codes and turbo codes for BEC channels. A random puncturing approach was proposed to obtain a closed form solution for codes of different rates. Demapper EXIT functions for different system parameters and in particular for different channel models were investigated. We considered a bit-level and symbol-level analysis and the numerical computation of demapper EXIT functions. Using these findings, design guidelines were elucidated for BICM with iterative demapping and decoding.

4

Mappings for Coded Modulation

High order modulation schemes are required to transmit high data rates over a limited bandwidth, as emphasized in Section 2.2. For a given number of bits that should be transmitted within one complex symbol, the signal constellation and the bit-to-symbol mapping need to be specified.

We use standard ASK, PSK and QAM signal constellations described in Section 2.2 and focus on the mapping, i.e. the assignment of bit sequences of length M to the complex symbols of a 2^M-ary signal constellation. Different mappings have been proposed for a variety of applications. Probably the most famous ones are Gray mapping [Gra53] for bit-interleaved coded modulation (BICM) and set partitioning mapping [Ung82] for trellis coded modulation (TCM). The degree of freedom in the design of mappings grows rapidly with the size of the signal constellation. With a 16QAM signal constellation, we can already select one out of $16! \approx 2 \cdot 10^{13}$ possible mappings. Note that several mappings may have similar properties and are just mirrored or rotated versions of other mappings, but the number of mappings with different characteristics is still very large.

In the EXIT chart analysis in Section 3.3 we already considered the mapping as a true coding entity. The memoryless mapper has a single state and generates encoded complex symbols corresponding to its input binary sequence; M consecutive bits determine which complex symbol is transmitted. Therefore, each bit influences the transmission of $M-1$ other bits. The single state trellis of the mapper with 2^M parallel transitions is depicted in Fig. 4.1. A state transition is characterized by the Hamming weight of the binary label and the Euclidean distance to the symbols corresponding to the other state transitions.

Similar to a channel code, we are interested in characterizing and optimizing the mapping. We characterize the mappings by introducing a suitable Euclidean distance spectrum (EDS) for BICM based systems in Section 4.1. In Section 4.2 we discuss the design of mappings for different applications using the EDS, including the standard Gray and set partitioning mappings, mappings for equal and unequal error protection, mappings for iterative receivers, multi-

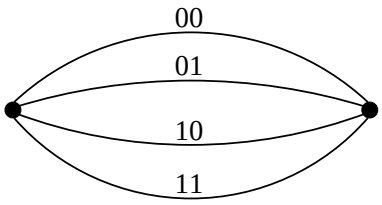


Figure 4.1: Trellis of a memoryless, single state 16QAM mapping [Ung82].

dimensional mappings and mappings for ARQ. In Section 4.3 we optimize mappings for different applications by casting the optimization to a quadratic assignment problem (QAP) and by solving it using a binary switching algorithm (BSA). The use of an additional inner encoder is finally discussed in Section 4.4.

All mappings considered in this Chapter are defined in the tables of Appendix A.

4.1 Euclidean Distance Spectrum

The term Euclidean distance spectrum or Euclidean distance profile is often used in the literature in the context of coded modulation and especially trellis coded modulation (TCM), where coding and modulation are considered as a single entity on a joint trellis. The Euclidean distance spectrum is then defined as the number of error events with a certain cumulated Euclidean distance along the wrong path in the trellis. The minimum Euclidean distance is denoted as the free Euclidean distance.

However, we consider the memoryless, single-state mapper as a stand-alone coding entity and focus on applications to BICM without and with iterative decoding. We define an Euclidean distance spectrum (EDS) [SGHB03a] to characterize and analyze mappings and to easily derive precise error bounds. The properties of the signal constellation are inherently included in the EDS. Roughly speaking, the EDS is the average or cumulated Hamming distance between bit labels at a specific Euclidean distance. The Hamming distance is equivalent to the number of bit errors made while choosing the wrong signal point.

We derive the EDS from a bit-wise and a symbol-wise point of view. These descriptions correspond to the bit-wise and symbol-wise EXIT charts introduced in Section 3.3, where the bit reliabilities are averaged over one bit position or one bit label, respectively. Note that the full characterization of a signal constellation would include the reliabilities of all $M \cdot 2^M$ bit positions.

Furthermore, the EDS depends on the amount of available a priori information. Similar to the EXIT chart analysis in Section 3.3, we assume BEC a priori information and characterize the mappings with the EDS depending on the number of a priori known bits. The EDS provides an exact characterization of the mapping with ideal a priori information. For the other scenarios, we introduce an expurgated EDS to obtain a satisfactory approximation of the mapping characteristics.

Other approaches to characterize signal constellation and mapping properties have been proposed in the literature. Brännström considered in [Brä04] a distance spectrum for PSK modulation for no and ideal a priori information. In [TS05a], a description of the mappings by a transfer function is proposed. A so-called average distance spectrum is defined in [LSAO03a] [LSAO03b] [ALSO04] for PSK and ASK signal constellations without a priori information.

Furthermore, Wachsmann presented in [Wac98] a distance profile in the context of multilevel coding (MLC) with multistage decoding (MSD). The EDS proposed in this section can be used for arbitrary signal constellations, any number of a priori known bits, and for a bit-wise and symbol-wise mapping description.

4.1.1 Bit-Wise Definition

We separately investigate the M bit positions in the binary label to obtain a bit-wise EDS. For each bit position, we define decision regions in the complex plane and list the Euclidean distances between signal points belonging to different decision regions. The results are in direct relation to the bit-wise EXIT charts investigated in Section 3.3.

Consider the example of 16QAM with set partitioning mapping shown in Fig. 4.2. The shaded regions correspond to the regions for the received signal where the optimal MAP demapper would decide for the investigated bit to be equal to 0. Large decision regions around the symbols provide a high protection for the corresponding bits. Note that the decision regions for each bit position may have different size and therefore the bit estimates may have different reliability. As example, we observe in Fig. 4.2 that the decision regions for the first bit are larger than for the last bit, resulting in a different reliability of the demapper soft output values.

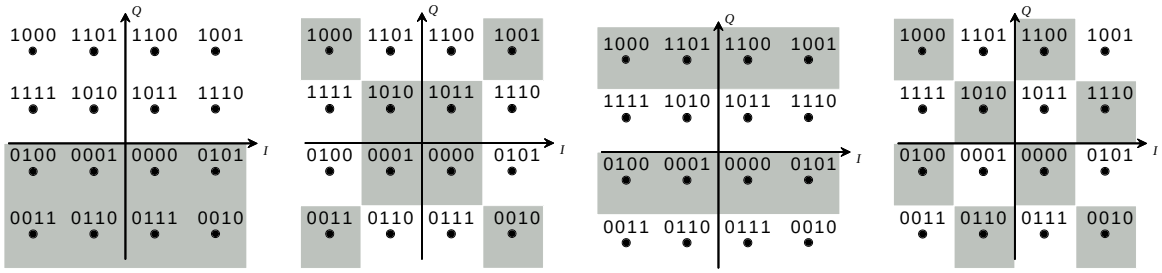
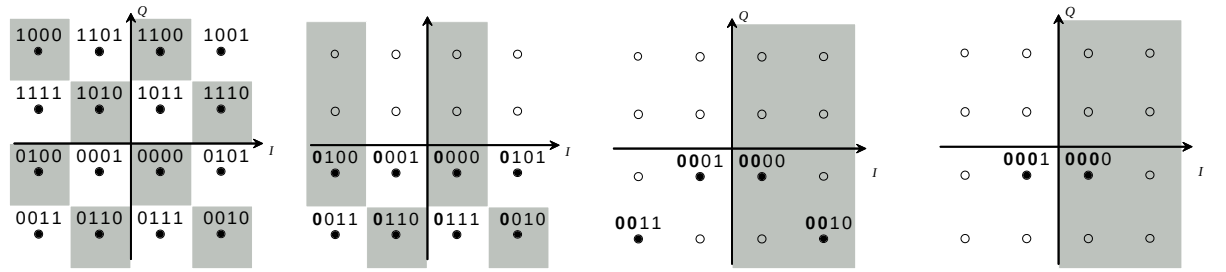


Figure 4.2: 16QAM with set partitioning mapping. Decision regions for bit $m = 1, 2, 3, 4$ with no a priori information. Shaded region: bit m is equal to 0.

With BEC a priori information, the number of possible signal points is reduced as described in Section 2.5. This a priori information may result in an enlargement of the decision regions and therefore in an improvement of the reliability of the soft demapper outputs.

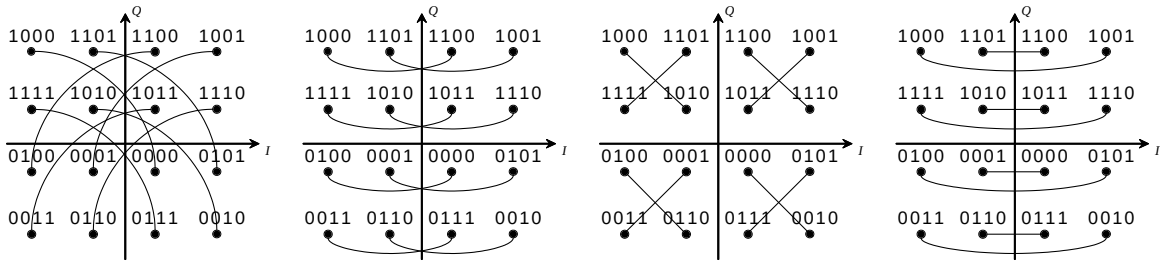
The effect of a priori information on the decision regions is illustrated in Fig. 4.3 for the example of 16QAM with set partitioning mapping and if the signal point with label (0000) is transmitted. The decision on the last bit is considered and we observe the remaining signal points and decision regions if the *first* 0, 1, 2 and 3 bits are a priori known. For that specific example, the knowledge of the third bit does not increase the decision region for the last bit. To determine the reliability of a specific bit position m in a BICM based scheme after the demapper, we have to average over all possible a priori known bit positions and a priori known bit values, as done in the EXIT chart computation in equation (3.29). Note that this reasoning is valid for the perspective of BICM with iterative demapping and decoding and does not correspond to the basic idea of set partitioning mapping, where the Euclidean distances within subsets of signal points are considered. These subsets correspond to the remaining signal points if the *last* 1, 2, and 3 bit are known in this example. A more comprehensive study of set partitioning mapping is given in Section 4.2.



a) zero bits known b) first bit known c) first two bits known d) first three bits known

Figure 4.3: 16QAM with set partitioning mapping. Symbol with label (0000) transmitted, remaining signal points and decision regions for the last bit $m = 4$ if $0, 1, 2, 3$ bits are a priori known.

With ideal a priori information (genie or error free feedback case), all bits except the bit to be detected are a priori known and the signal constellation is reduced to a binary constellation. Depending on the actual a priori information, different symbol pairs remain. The labels of the two remaining signal points differ only in the bit to be detected. The decision regions with ideal a priori information then correspond to the decision regions with BPSK modulation. Fig. 4.4 depicts the possible binary signal constellations for a 16QAM signal constellation with set partitioning mapping for the bit positions $m = 1, \ldots, 4$ to be detected. We achieve a good performance if the Euclidean distance between the corresponding signal points is maximized. Therefore, the first bit has the highest and the fourth bit the lowest reliability in this example.



a) 1^{st} bit b) 2^{nd} bit c) 3^{rd} bit d) 4^{th} bit

Figure 4.4: 16QAM with set partitioning mapping. Detection of bit $m = 1, \ldots, 4$ with ideal a priori information.

From the investigations done so far we have seen that signal constellations and mappings are precisely characterized by their decision regions for each bit position. The aim of the Euclidean distance spectrum (EDS) is to retrieve and list the relevant Euclidean distances between the decision regions. With ideal a priori information, we obtain a precise characterization of the mapping if we enumerate the distances between all possible remaining symbol pairs. Without a priori information, the EDS cannot reflect in general the details of the two dimensional decision regions, at least for signal constellations that are not based on ASK and PSK modulation schemes. However, we can obtain a good and helpful approximation.

To formally derive the EDS from a bit-wise perspective, we define the three sets D_{ex}, D and Λ:

The set $D_{\text{ex}} = \{d_{\text{ex},1}, \ldots, d_{\text{ex},v}, \ldots, d_{\text{ex},V}\}$ is defined as the set of all possible distinct (*ex*purgated) Euclidean distances between any two distinct signal points of the signal set $\mathcal{X}$. For example, the possible Euclidean distances for a 16QAM signal constellation are shown in Fig. 4.5, where Δ denotes the minimum squared Euclidean distance between any two signal

points:

$$\Delta = \min_{\forall x_i, x_j \in \mathcal{X}} |x_i - x_j|^2. \tag{4.1}$$

In this example, we have $V = 9$ distinct Euclidean distances and

$$D_{\text{ex}}^2 = \{\Delta, 2\Delta, 4\Delta, 5\Delta, 8\Delta, 9\Delta, 10\Delta, 13\Delta, 18\Delta\}. \tag{4.2}$$

In general, an ASK signal constellation has $V = 2^M - 1$ distinct Euclidean distances, square QAM $V = 2^{M-1} + 2^{M/2-1} - 1$, and PSK $V = 2^{M/2}$.

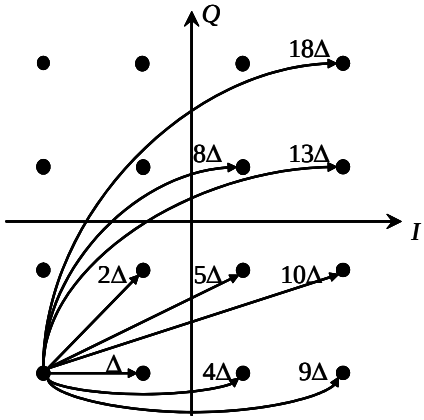


Figure 4.5: Possible Euclidean distances in the set D_{ex}^2 with 16QAM.

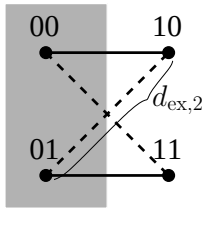
Furthermore, we enumerate for each bit position $m = 1, \ldots, M$ all not necessarily distinct Euclidean distances between symbols that belong to different decision regions, where two symbols x_i and x_j belong to different decision regions for bit position m if the bit label of x_i differs in the mth bit position from the bit-label of x_j. These Euclidean distances are grouped in the set $D = \{d_1, \ldots, d_{V'}\} \supseteq D_{\text{ex}}$.

With L a priori known bits, the reduced signal constellation has 2^{M-L} signal points and we cumulate the distances over the M bit positions and $2^L \cdot \binom{M-1}{L}$ possible reduced signal constellations. The cardinality of the set D is then:

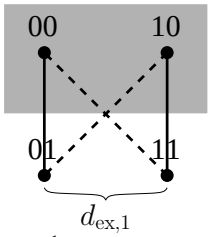
$$V' = |D| = M \cdot 2^L \cdot \binom{M-1}{L} \cdot \frac{2^{M-L} \cdot 2^{M-L-1}}{2} = M \cdot 2^{2M-L-2} \cdot \binom{M-1}{L}. \tag{4.3}$$

Finally, let λ_v denote the frequency of the distance $d_{\text{ex},v}$ in the set D and $\Lambda = \{\lambda_1, \ldots, \lambda_V\}$. D_{ex} depends only on the signal constellation, whereas D and Λ characterize the bit mapping. The Euclidean distance spectrum lists the frequencies λ_v of the Euclidean distances $d_{\text{ex},v}$.

To illustrate these definitions, we consider the basic example of QPSK modulation. The set D_{ex} of possible distinct Euclidean distances includes the two distances $D_{\text{ex}} = \{d_{\text{ex},1}, d_{\text{ex},2}\} = \{\sqrt{\Delta}, \sqrt{2\Delta}\}$. Fig. 4.6(a) depicts the distances from the set D between the decision regions for Gray mapping and for the two bit positions. With ideal a priori information, the distances marked with dashed lines are eliminated. The corresponding EDS lists all possible distances sketched in Fig. 4.6(a) and is given in the Tables of Fig. 4.6(b) for no and ideal a priori information. The distances in brackets are omitted in the expurgated EDS described later on. Fig. 4.7 shows the results for QPSK Anti-Gray mapping.



a) 1st bit ($m = 1$),



2nd bit ($m = 2$).

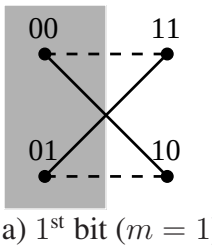
D_{ex}^2	Δ	2Δ
Λ	λ_1	λ_2
1st bit	2	(2)
2nd bit	2	(2)
Overall	4	(4)

b) no a priori,

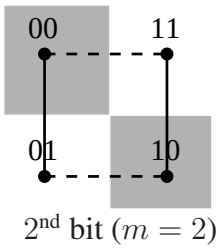
D_{ex}^2	Δ	2Δ
Λ	λ_1	λ_2
1st bit	2	0
2nd bit	2	0
Overall	4	0

ideal a priori

Figure 4.6: **a)** QPSK Gray mapping with distances from the set D. Dashed lines: relevant only without a priori information. Shaded regions: bit m has value 0. **b)** Distance spectrum: frequencies λ_1 and λ_2 of distances $d_{\text{ex},1} = \sqrt{\Delta}$ and $d_{\text{ex},2} = \sqrt{2\Delta}$, respectively.



a) 1st bit ($m = 1$),



2nd bit ($m = 2$).

D_{ex}^2	Δ	2Δ
Λ	λ_1	λ_2
1st bit	2	(2)
2nd bit	4	(0)
Overall	6	(2)

b) no a priori,

D_{ex}^2	Δ	2Δ
Λ	λ_1	λ_2
1st bit	0	2
2nd bit	2	0
Overall	2	2

ideal a priori

Figure 4.7: **a)** QPSK Anti-Gray mapping with distances from the set D. Dashed lines: relevant only without a priori information. Shaded regions: bit m has value 0. **b)** Distance spectrum: frequencies λ_1 and λ_2 of distances $d_{\text{ex},1} = \sqrt{\Delta}$ and $d_{\text{ex},2} = \sqrt{2\Delta}$, respectively.

For a good performance, the minimum Euclidean distance should be maximized and the number of minimum Euclidean distances minimized. In other words, the minimum index v of $d_{\text{ex},v}$ where $\lambda_v \neq 0$ should be maximized and the corresponding value of λ_v minimized. As a consequence of the area theorem of the demapper EXIT chart, we stated in Section 3.4 that there is a strong trade-off in the design of mappings between the performance with no a priori information and with ideal a priori information. A Gray mapping optimized for no a priori information will take little advantage of this a priori information, as illustrated in Fig. 4.6 where the overall value $\lambda_1 = 4$ is independent of the a priori information. With Anti-Gray mapping, the number of distances at minimum Euclidean distance is reduced from $\lambda_1 = 6$ to $\lambda_1 = 2$ by a priori information, see Fig. 4.7. Therefore, we expect with Anti-Gray mapping a performance that is lower than the one with Gray mapping with no a priori information, but higher for the assumption of ideal a priori information.

Furthermore, we consider in Fig. 4.6 and 4.7 the distances for each bit separately. With QPSK Gray mapping, the 2 bit positions are equally reliable, whereas with QPSK Anti-Gray mapping, the first bit is more reliable than the second one, both for no and ideal a priori information.

The relations in the EDS are directly reflected in the corresponding EXIT chart shown in Fig. 4.8. With Gray mapping, both bit positions are equally reliable and the a priori information has no influence on the performance. With Anti-Gray mapping, the two bit positions have different reliability and the a priori information improves the performance. The EXIT chart results correspond directly to the EDS in Fig. 4.6 and 4.7, where with Anti-Gray mapping the value of $\lambda_1 = 2$ of the first bit without a priori information and of the second bit with ideal a priori information is equal to the value of $\lambda_1 = 2$ with Gray mapping.

In addition to the marginal cases of no a priori information and ideal a priori information, an EDS can be given for signal constellations with $M \geq 3$ for any number $1, \ldots, M-2$ of a priori known bits.

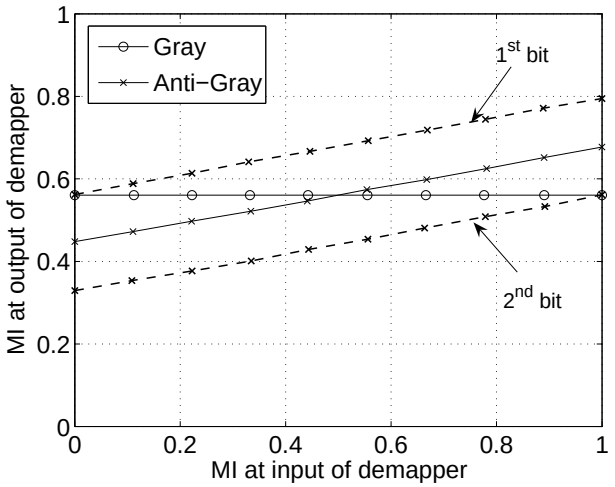


Figure 4.8: Average and bit-wise EXIT functions with QPSK Gray and Anti-Gray mapping. AWGN channel, $10\log_{10}(E_s/N_0)\text{dB} = 1\text{dB}$.

A comprehensive example is given in Table 4.1. The EDS for 16QAM with set partitioning mapping as depicted in Fig. 4.2 and 4.4 is given for a different number of a priori known bits and for the different bit positions. We observe that the EDS is improved with a priori information and the different reliabilities of the bit positions.

4.1.2 Symbol-Wise Definition

Instead of processing each bit position $1, \ldots, M$ for the EDS, the symbol-wise approach investigates the relevant Euclidean and Hamming distances for each symbol $1, \ldots, 2^M$ separately. The cumulated overall distances of the bit-wise and symbol-wise approach are identical.

The symbol-wise investigation gives the reliability of a specific symbol and is useful for the optimization of mappings, where symbols with a large contribution to a bad performance need to be determined, see Section 4.3. The results are consistent with the symbol-wise EXIT charts introduced in Section 3.3.

To obtain the symbol-wise EDS, we proceed as follows: We cumulate and sort the Hamming distances (or number of bit errors) for each of the 2^M symbols to the other $2^M - 1$ symbols according to the Euclidean distance between the respective symbols. A mapping has a uniform symbol-wise EDS if the EDS is identical for each symbol.

Consider again the example of QPSK modulation with Gray and Anti-Gray mapping shown in Fig. 4.6 and 4.7, respectively. With Gray mapping and no a priori information, the cumulated Hamming distance for symbols at squared Euclidean distance $d^2_{\text{ex},1} = \Delta$ and $d^2_{\text{ex},2} = 2\Delta$ is 2 for all symbols. With Anti-Gray mapping, we have 3 cumulated bit errors at $d^2_{\text{ex},1} = \Delta$ and 1 bit error at $d^2_{\text{ex},2} = 2\Delta$ for all symbols. Therefore, these mappings are uniform and we obtain the overall EDS as given in Fig. 4.6(b) and 4.7(b) with the factor of two. With a priori information, the procedure is similar but we consider the reduced signal constellations.

A further example for 16QAM with set partitioning mapping and for the symbol labeled with (0000) is given in Fig. 4.9. No a priori information is assumed.

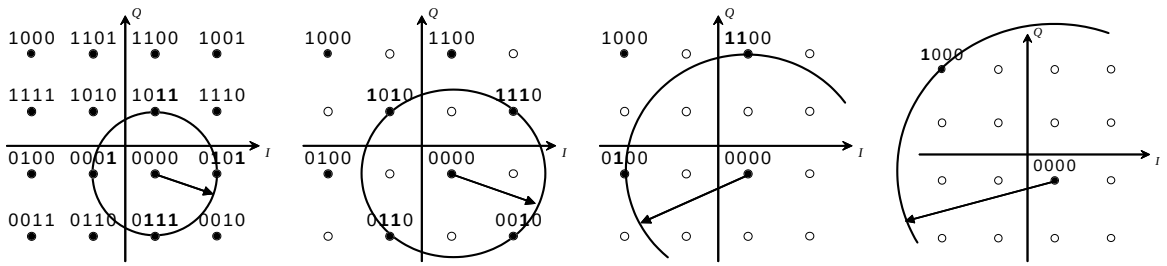


Figure 4.9: 16QAM with set partitioning mapping. Symbol with label (0000) transmitted, enumerate cumulated Hamming distances to symbols at different Euclidean distances.

4.1.3 Expurgated Definition

So far, we have considered all possible Euclidean distances between the decision regions. However, not all of these distances are relevant. Therefore, we define an expurgated EDS that contains only the most significant distances. The idea is similar to the approach presented in [CTB98], where relevant distances for BICM with Gray mapping have been determined to obtain tighter error bounds.

With ideal a priori information, the EDS as described above accurately characterizes the mapping. Otherwise, we use the bit-wise definition of the EDS to approximate the properties of the decision regions as follows: For each symbol, we include in the set D only the Euclidean distances to the nearest symbols from the other decision region. With Gray mapping, each symbol has only one so called *nearest neighbor* from the other decision region. In general, several nearest neighbors are possible.

For the example of QPSK modulation, the expurgated EDS does not consider the values of λ_2 in brackets in Fig. 4.6(b) and 4.7(b). Consider in particular QPSK Gray mapping without a priori information as shown in Fig. 4.6(a). Then, if we investigate e.g. the decision on the first bit of the signal point labeled with (00), only the Euclidean distance to its nearest neighbor (10) should be taken into account. If we reduce the problem of identifying regions contributing to the probability of error to a single dimensional problem, the Euclidean distance to the point labeled with (11) has no influence on the performance and we should ignore it in that case. If non-Gray mappings are used, a signal point may have several nearest neighbors, as it is the case for the second bit of the anti-Gray mapping in Fig. 4.7(a).

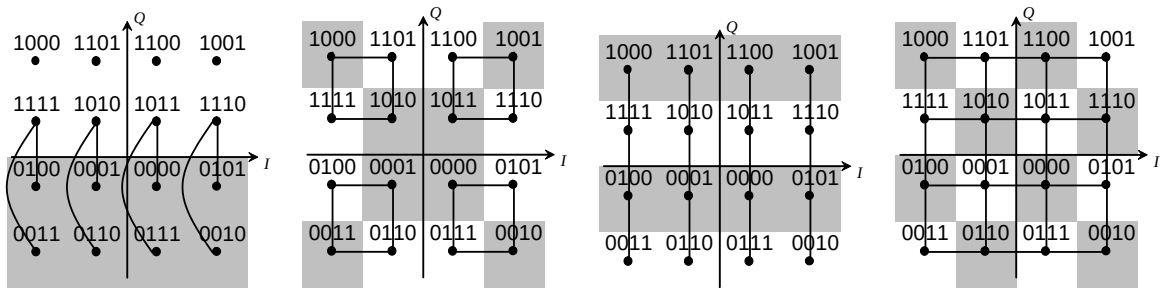
As a further example, the Euclidean distances for the expurgated EDS without a priori information with 16QAM and set partitioning mapping are depicted in Fig. 4.10.

Table 4.1 summarizes the EDS for different setups for the example of 16QAM and set partitioning mapping. The given values should be considered in relation with the Figures mentioned in the table.

Fig. 4.11 depicts the bit-wise EXIT chart corresponding to the EDS in Table 4.1. We readily observe that the 3rd bit is more reliable than the 2nd and 4th bit without a priori information, and that its performance is just slightly better than the 4th bit with ideal a priori information. This corresponds to the relations in the EDS.

D^2_{ex} (Fig. 4.5)	$d^2_{\text{ex},1}=\Delta$	$d^2_{\text{ex},2}=2\Delta$	$d^2_{\text{ex},3}=4\Delta$	5Δ	8Δ	9Δ	10Δ	13Δ	18Δ
Λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
Number of a priori known bits: $L=0$ (No a priori, Fig. 4.2)									
1^{st} bit	4	6	8	16	8	4	8	8	2
2^{nd} bit	16	8	16	8	0	0	8	8	0
3^{rd} bit	12	18	0	12	0	4	12	4	2
4^{th} bit	24	0	0	24	0	8	0	8	0
Overall	56	32	24	60	8	16	28	28	4
Number of a priori known bits: $L=0$ (No a priori), expurgated (Fig. 4.10)									
1^{st} bit	4	0	4	0	0	0	0	0	0
2^{nd} bit	16	0	0	0	0	0	0	0	0
3^{rd} bit	12	0	0	0	0	0	0	0	0
4^{th} bit	24	0	0	0	0	0	0	0	0
Expurgated	56	0	4	0	0	0	0	0	0
Number of a priori known bits: $L=1$									
Expurgated	26.7	12	5.4	2.7	0	0	0	0	0
Number of a priori known bits: $L=2$									
Expurgated	9.4	12	9.4	5.4	2	0	0	0	0
Number of a priori known bits: $L=3$ (Ideal a priori, Fig. 4.4)									
1^{st} bit	0	0	0	0	8	0	0	0	0
2^{nd} bit	0	0	8	0	0	0	0	0	0
3^{rd} bit	0	8	0	0	0	0	0	0	0
4^{th} bit	4	0	0	0	0	4	0	0	0
Overall	4	8	8	0	8	4	0	0	0

Table 4.1: Euclidean distance spectrum (EDS) for 16QAM with set partitioning mapping. The values are normalized by $\binom{M-1}{L}$.



a) 1^{st} bit b) 2^{nd} bit c) 3^{rd} bit d) 4^{th} bit

Figure 4.10: 16QAM with set partitioning mapping. Euclidean distances for expurgated EDS for bit $m = 1, \ldots, 4$ and no a priori information.

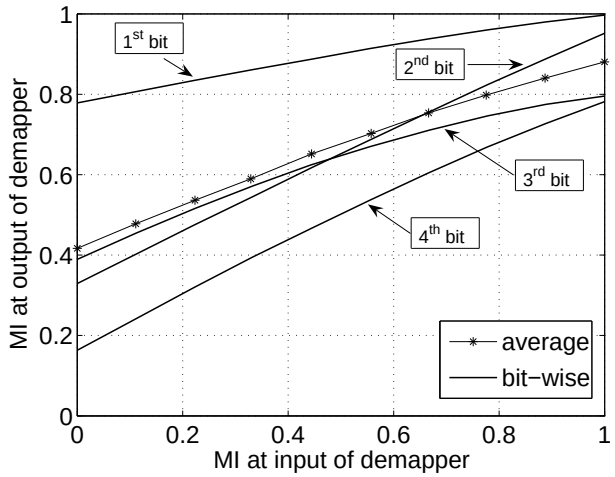


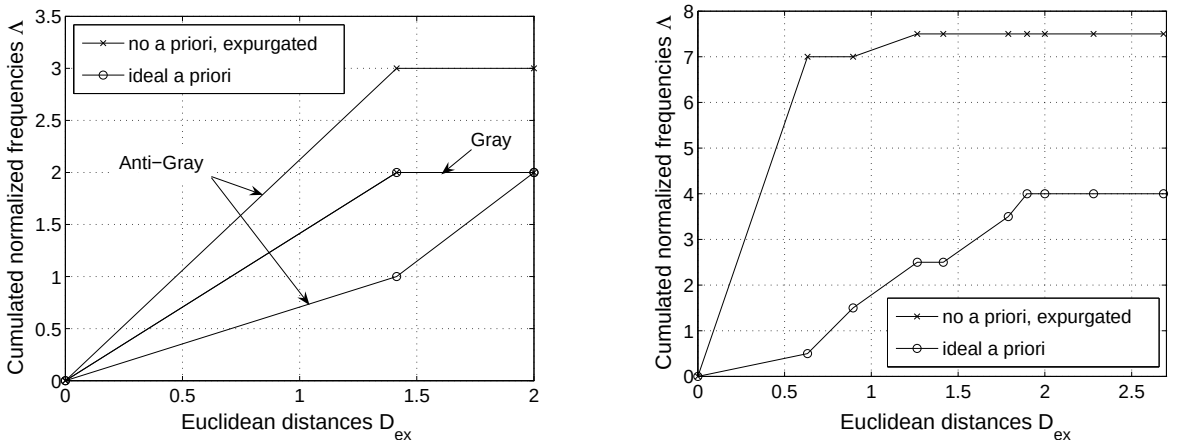
Figure 4.11: Average and bit-wise EXIT function of 16QAM with set partitioning mapping, AWGN channel, $10\log_{10}(E_s/N_0)\text{dB} = 8\text{dB}$.

4.1.4 Graphical Representation

The description of the EDS like in Table 4.1 can become cumbersome. As an alternative, we can describe the EDS as coefficients of a polynomial or using a graphical representation. A possible graphical representation is shown in Fig. 4.12 for QPSK and 16QAM with set partitioning mapping, corresponding to the tables in Fig. 4.6, 4.7 and to Table 4.1, respectively. We plot the cumulated values of the frequencies λ_v normalized by the number of signal points 2^M over the Euclidean distances $d_{ex,v}$, $v = 1, \ldots, V$.

4.2 Mapping Strategies

Mappings can be designed and optimized for a variety of applications. Before discussing optimization algorithms for mappings, it is worth examining the most important design aims and some corresponding mappings. Furthermore, we show how to apply the EXIT chart and the Euclidean distance spectrum (EDS) as powerful tools to describe the specific properties of the different mappings.



a) QPSK b) 16QAM, set partitioning

Figure 4.12: Graphical representation of expurgated EDS. Normalized frequencies: $\lambda_v/2^M$, Euclidean distances from the set D_{ex} with normalized signal constellations and minimum Euclidean distances $\sqrt{\Delta} = \sqrt{2}$ and $\sqrt{\Delta} = \sqrt{0.4}$ for QPSK and 16QAM, respectively.

4.2.1 Gray Mapping

Gray mapping was proposed by Frank Gray in a patent from 1953 as a means of reducing the error rate in a pulse code communication system [Gra53]. With Gray mapping, the binary labels of signal points at minimum Euclidean distance differ in only one bit, thus minimizing the error probability for the most probable symbol error event at minimum Euclidean distance. This commonly used definition of a Gray mapping does not say anything about the Hamming distance between binary labels of signal points at distances larger than the minimum Euclidean distance. The error events at minimum Euclidean distance are the most relevant, but error events at other small Euclidean distances may also be taken into account.

Gray proposed a recursive construction method that generates the mapping by binary reflexion as follows: If we start from a Gray labeling of order $m-1$, we append a sequence of 2^{m-1} binary vectors formed by repeating the labels in reverse order. To this new sequence of binary labels, an extra bit is added from the left, where this extra bit is 0 for the first half of the 2^m labels and 1 for the second half. The example in Fig. 4.13 illustrates this procedure.

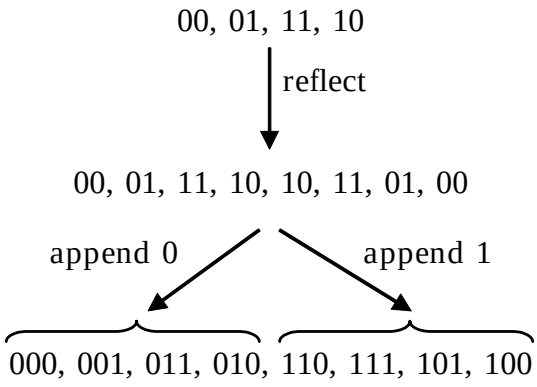


Figure 4.13: Construction of a binary reflected Gray code.

A binary reflected Gray code of order M is obtained by $M-1$ recursive reflexions of the trivial Gray labeling (01). Through the direct product of two one-dimensional Gray mappings constructed by reflection, we obtain the corresponding Gray mapping for two-dimensional QAM signal constellations.

It is shown in [ALSO04] that this construction method leads to the best Gray mappings with optimal performance without a priori information at the demapper.

Consider as example the EDS in Table 4.2 and the EXIT functions in Fig. 4.14 of two Gray mappings for 16PSK modulation obtained by different construction methods. We observe that the Gray mapping constructed by reflection has a slightly better performance without a priori information. The value $\lambda_1 = 16$ is identical for all Gray mappings. In general, $\lambda_1 = 2^M$ for all PSK signal constellations with Gray mapping. However, λ_2 and then $\lambda_3, \lambda_4, \ldots$ are minimized with Gray mapping constructed by binary reflexion. Since the area under the demapper EXIT function is identical for all mappings, the performance of the Gray mapping constructed by reflection is slightly lower with ideal a priori information than with other Gray mappings. However, this is not the scenario Gray mapping was designed for.

Λ	λ_1	λ_2	λ_3	λ_4	λ_5	$\ldots$
No a priori information (Expurgated)						
Binary reflected	16	8	4	4	0	$\ldots$
Balanced	16	11	7	0	0	$\ldots$
Ideal a priori information						
Binary reflected	16	0	8	0	4	$\ldots$
Balanced	16	0	5	0	0	$\ldots$

Table 4.2: EDS, 16PSK with different Gray mappings.

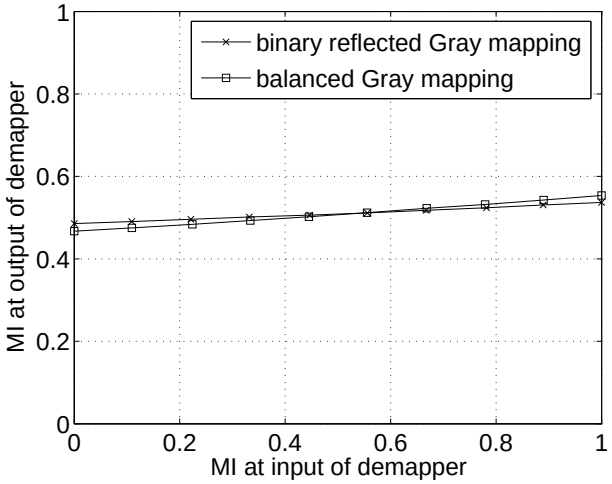


Figure 4.14: EXIT function, 16PSK with different Gray mappings, $10\log_{10}(E_s/N_0)\text{dB} = 6\text{dB}$.

4.2.2 Mapping by Set Partitioning

Mapping by set partitioning was introduced by Ungerböck in [Ung82] [Ung87a] [Ung87b] for trellis coded modulation (TCM) described in Section 2.5. The mapping is designed in conjunction with the code to increase the minimum Euclidean distance between pairs of coded signal sequences.

The mapping by set partitioning is constructed by successive partitioning of the signal constellation into disjoint subsets with increasing minimum Euclidean distance between the signal points of these subsets, as illustrated for the example of 8ASK modulation in Fig. 4.15.

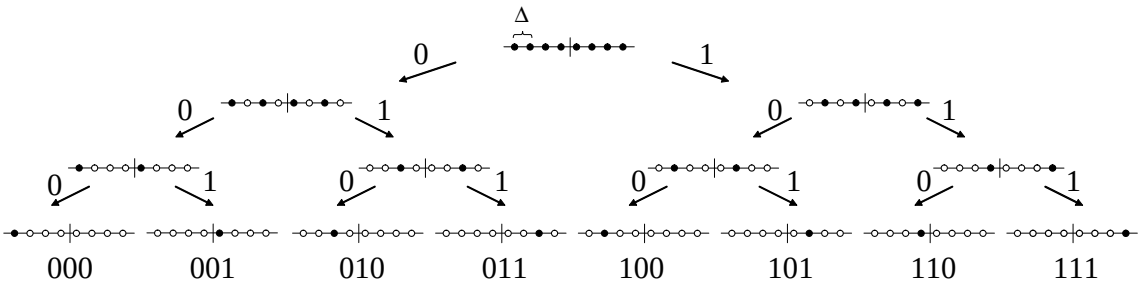


Figure 4.15: Construction of a set partitioning mapping for 8ASK.

At each partitioning level, the intra-set Euclidean distance is maximized. For one-dimensional ASK signal constellations, the minimum Euclidean distance at a certain partition level is twice the minimum Euclidean distance of the previous partition level. For two dimensional QAM signal constellations, the minimum *squared* Euclidean distance at a certain partition level is twice the minimum squared Euclidean distance of the previous partition level. Note that this approach does not lead to optimized Euclidean distances for the last partitioning level, as required for iterative demapping and decoding. It is more a greedy approach, where the next partition is chosen to achieve the best improvement at that level without considering the next levels. Each bit position in the binary label specifies a subset at a certain partition level. If e.g. the bit corresponding to the first partition level is known, the mapping is optimized for the second partition level and a certain minimum Euclidean distance between the signal points is guaranteed. This important property is used when the mapping is combined with a code in a TCM scheme.

This explains in part the convex shape of the average EXIT function $f_{\text{dem}}(I)$ in Fig. 4.11: Starting from a low value $f_{\text{dem}}(0)$ without a priori information, the improvement is large with low a priori information but there is no more potential to reach high values $f_{\text{dem}}(1)$ with ideal a priori information. Set partitioning is therefore not well suited for iterative demapping and decoding schemes, where both the values $f_{\text{dem}}(0)$ and $f_{\text{dem}}(1)$ should be large for an early start of the iterative process and low error rates at high SNR, respectively.

The properties of set partitioning are not directly reflected by the EDS since we average over all combinations of a priori known bit positions. Nevertheless, with ideal a priori information, symbol pairs that differ in a bit position corresponding to the first partition level have a smaller minimum Euclidean distance than symbol pairs that differ in a bit position corresponding to last partition level: In the example of Fig. 4.15, the Euclidean distance between the signal points labeled with (000) and (100) is Δ, between (000) and (010) 2Δ, and between (000) and (001) 4Δ, where Δ is the minimum Euclidean distance between any two signal points. A similar observation can be made for 16QAM with set partitioning in the EDS of Table 4.1 with ideal a priori information.

In contrast to set partitioning, block partitioning mapping proposed in [WFH99] for multilevel coding (MLC) has the aim to keep the minimum intra-set Euclidean distance as constant as possible for all partitioning levels. For iterative demapping and decoding, the characteristics are however similar to set partitioning and we will not further address this mapping.

4.2.3 Mappings for Equal and Unequal Error Protection

The individual bit positions in the binary label may have a different reliability, as observed in bit-wise EXIT charts and bit-wise EDS. It is often desirable to design mappings with specific bit-wise performance characteristics. The aim is to achieve either a uniform or non-uniform reliability distribution of the bit positions, resulting in an equal or unequal error protection, respectively.

Consider the example of a multiplexing system in which each bit position carries the data of an independent source. The mapping can be optimized to provide equal or unequal error protection for the different sources. Equal error protection is of interest if all sources provide information of same relevance. Unequal error protection is of interest for e.g. hierarchical transmission in a broadcast scenario. The more reliable bit positions carry crucial information that should be received by all users whereas the less reliable bit positions carry enhancement information. By selecting a specific mapping, we do not change the capacity. To improve the overall capacity for hierarchical broadcast transmission, we should shift the signal points and choose a signal constellation where the signal points are grouped to clusters, as proposed and investigated in [Cov72] [ROUV93] [See99].

The BICM system is another example where the reliability of the individual bit positions is of relevance. A convolutional code has the best performance if all channel values have a similar reliability.

Mappings with similar properties for each bit position are usually denoted as *balanced mappings*. The class of balanced Gray mappings is investigated in [Sav97] and references therein.

The bit-wise expurgated EDS without a priori information is given in Table 4.3 for the example of the two 16PSK Gray mappings already investigated in Table 4.2 and Fig. 4.14. The cumulated EDS is slightly better with the binary reflected Gray mapping, with smaller values of λ_2 and λ_3. However, the single bit positions in the binary reflected Gray mapping have quite different values of λ_1, resulting in different protection levels. The balanced mapping has identical values of $\lambda_1 = 8$ and similar values of λ_2 for all bit positions.

These facts are reflected in the bit-wise EXIT chart in Fig. 4.16, where we clearly observe the different reliability of the bit positions with the binary reflected Gray mapping.

4.2.4 Mappings for Iterative Receivers

The binary reflected Gray mapping is optimal without a priori information at the demapper. With a receiver performing iterative demapping and decoding, close to ideal a priori information may be available at the demapper at high SNR after several iterations. But what is the "optimal" mapping with ideal a priori information? With QPSK modulation, only two distinct mappings are possible. We denoted them as Gray and Anti-Gray mapping in Fig. 4.6 and 4.7, respectively. But what is in general the definition of an "Anti-Gray" mapping? What are the mappings that provide the best performance with ideal a priori information for high order signal constellations? How can we find them, how can we construct them?

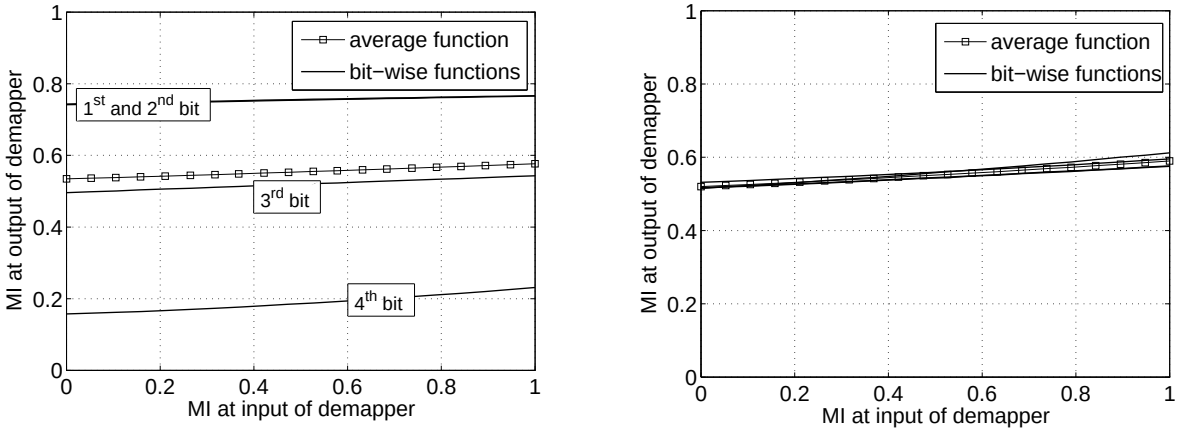
Λ	λ_1	λ_2	λ_3	λ_4	...
1^{st} bit	4	4	4	4	...
2^{nd} bit	4	4	4	4	...
3^{rd} bit	8	8	0	0	...
4^{th} bit	16	0	0	0	...
Overall	16	8	4	4	...

a) Binary reflected

Λ	λ_1	λ_2	λ_3	λ_4	...
1^{st} bit	8	6	4	0	...
2^{nd} bit	8	6	4	0	...
3^{rd} bit	8	6	2	0	...
4^{th} bit	8	4	4	0	...
Overall	16	11	7	0	...

b) Balanced

Table 4.3: Bit-wise expurgated EDS without a priori information, 16PSK with different Gray mappings.



a) Binary reflected b) Balanced

Figure 4.16: Average and bit-wise EXIT functions, 16PSK with different Gray mappings; $10\log_{10}(E_s/N_0)\text{dB} = 7\text{dB}$.

With ideal a priori information, the signal constellation is reduced to two signal points whose labels differ only in the bit to be detected. To maximize the gain over the iterations and to optimize the performance at high SNR, the aim is to *maximize* the minimum Euclidean distance between symbols whose bit label differ in one bit. This is in contrast to the design rules for Gray mapping, where the aim is to *minimize* the Euclidean distance between symbols whose bit label have Hamming distance one.

Several mappings for iterative receivers have been investigated in [tB00] [CR01] [CR02] [TS02] [ZLH03] [SWK03] [CGV04] [TN04] [TS05a]. We propose in Section 4.3 the general approach published in [SGHB03b] [SGHB03a] to optimize mappings.

As example, we consider in Fig. 4.17 the modified set partitioning (MSP) mapping from [CR01] and the maximum squared Euclidean weight (MSEW) mapping from [TS02]. The corresponding EDS are given in Table A.3 in Appendix A. We observe the strong performance improvement with ideal a priori information compared to Gray mapping.

It is interesting that, if the binary bit labels are converted to decimal numbers, the MSEW mapping is a perfect magic square, since the sum of all rows, columns, diagonals, 2×2 sub-squares and 2×2 cyclic sub-squares is equal to 30. We also observed that every perfect magic square has the distance spectrum of the MSEW mapping.

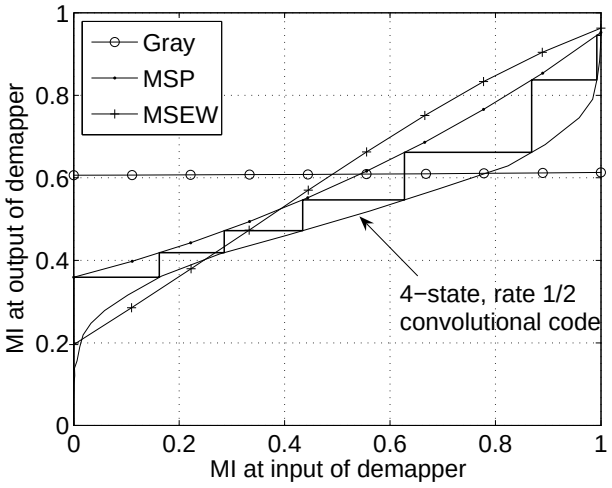


Figure 4.17: EXIT functions, 16QAM with different mappings, AWGN, $10\log_{10}(E_s/N_0)\text{dB} = 7\text{dB}$.

4.2.5 Multi-Dimensional Mappings

Instead of mapping a bit sequence of length M to one symbol, we can jointly map a bit sequence $\mathbf{c}$ of length MN_d to a vector $\mathbf{x}$ of N_d symbols. These multi-dimensional mappings provide further design and optimization possibilities. The N_d jointly mapped symbols can be transmitted sequentially in time or in parallel through N_d transmit antennas. Fig. 4.18 depicts an example for two-dimensional QPSK with $N_d = 2$ and $M = 2$. The joint mapping implies that there may be no one-to-one relationship between the first M bits of the label and the first symbol.

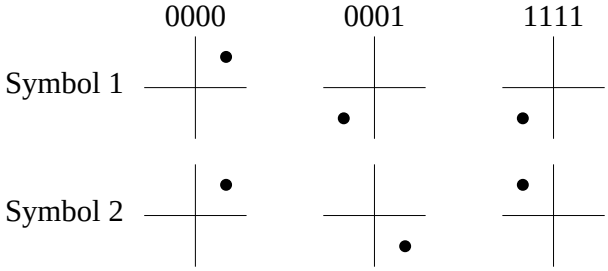


Figure 4.18: Example of two-dimensional QPSK mapping.

According to equation (2.43), the APP LLR for the bit estimate $\hat{c}_m$ with multi-dimensional mapping is

$$L(\hat{c}_m) = \log \frac{\sum_{\forall \mathbf{c}: c_m = 0} \mathrm{e}^{\Lambda(\mathbf{c})}}{\sum_{\forall \mathbf{c}: c_m = 1} \mathrm{e}^{\Lambda(\mathbf{c})}}. \tag{4.4}$$

Similar to equation (2.45), the metric $\Lambda(\mathbf{c})$ is computed for the AWGN channel as follows:

$$\Lambda(\mathbf{c}) = -\sum_{n=1}^{N_d} \frac{|y_n - x_n|^2}{2\sigma_n^2} - \sum_{m=1}^{MN_d} c_m \cdot L(c_m), \tag{4.5}$$

where the bit sequence $\mathbf{c}$ with elements c_m is jointly mapped to the symbol sequence $\mathbf{x}$ with elements x_n.

Multi-dimensional mappings have been extensively studied in the 80's for trellis coded modulation (TCM) [CS87] [Ung87b] [Wei87]. The application of multi-dimensional mappings to multi-antenna systems has been investigated in [Hon03] [Bär04] [GBB05] [MS06] and to BICM in [SWM05] [TN06].

With BICM, standard Gray mapping is the best choice for all symbols if no a priori information is available at the demapper. Then, multi-dimensional mappings provide no benefits. The advantages of multi-dimensional mappings arise when mappings are designed for iterative receivers with a priori information at the demapper.

Consider the example of two-dimensional QPSK mappings. Then, the possible squared Euclidean distances are $D^2_{\text{ex}} = \{\Delta, 2\Delta, 3\Delta, 4\Delta\}$, where Δ is the minimum squared Euclidean distance between two signal points in a QPSK signal constellation. Table 4.4 gives the EDS of different two-dimensional QPSK mappings. As expected, Gray mapping is optimal without a priori information. With ideal a priori information however, the turbo optimized two-dimensional mapping yields a large gain over the one-dimensional QPSK Gray and Anti-Gray mappings.

D^2_{ex}	$d^2_{\text{ex},1} = \Delta$	$d^2_{\text{ex},2} = 2\Delta$	$d^2_{\text{ex},3} = 3\Delta$	$d^2_{\text{ex},4} = 4\Delta$
Λ	λ_1	λ_2	λ_3	λ_4
No a priori information (Expurgated)				
Gray	32	0	0	0
1-D Anti-Gray	48	0	0	0
2-D Turbo opt.	72	0	0	0
Ideal a priori information				
Gray	32	0	0	0
1-D Anti-Gray	16	16	0	0
2-D Turbo opt.	0	0	24	8

Table 4.4: EDS, two-dimensional QPSK mappings ($M = 2$, $N_d = 2$), mappings defined in Table A.8.

4.2.6 Mapping Rearrangement and Mapping Diversity for ARQ

Balanced Gray mappings with equal reliabilities of all bits do not exist for all signal constellations, in particular not for 16QAM. Then, the receiver performance is lower than with a homogeneous bit reliability distribution. To overcome this fact in ARQ schemes - and in general transmit diversity schemes - with maximum ratio combining (or Chase combining), each diversity branch may use a different Gray mapping [WES04]. With 16QAM Gray mapping, the $M \cdot 2^M = 4\dot{1}6$ possible bit positions have four different reliabilities. With carefully chosen Gray mappings for each retransmission, the bit reliabilities are homogeneous after four retransmissions. This approach was adopted for hybrid ARQ operation in high speed downlink packet access (HSDPA) [3GPP06b].

The idea of symbol mapping diversity for ARQ is more general: A different mapping is selected for each retransmission and mappings different from Gray may be used for the retransmissions. In [SDH05], optimized mappings for each retransmission have been proposed for BICM without iterative demapping and decoding. The mapping for the ith retransmission is chosen to maximize the performance after the ith retransmission. It is interesting to note that even though no feedback to the demapper is implemented, a Gray mapping is selected only for the first initial

transmission. We derived optimized sequences of mappings for BICM with iterative demapping and decoding in [Vu05].

The MAP demapper with mapping diversity and N_t retransmissions is given by equation (2.43) with the new metric implementing maximum ratio combining:

$$\Lambda(\mathbf{c}_n) = -\sum_{i=1}^{N_t} \frac{|y_n^{(i)} - x_n^{(i)}|^2}{2\sigma_n^2} - \sum_{m=1}^{M} c_{n,m} \cdot L(c_{n,m}), \tag{4.6}$$

for the AWGN channel and where $x_n^{(i)}$ and $y_n^{(i)}$ are the transmitted and received signals of the ith retransmission, respectively.

Assume as example QPSK modulation with the distinct squared Euclidean distances $D_{\text{ex}}^2 = \{\Delta, 2\Delta\}$ and up to three retransmissions in an ARQ scheme. Then, the possible cumulated distances over three transmissions are $D_{\text{ex}}^2 = \{\Delta, 2\Delta, 3\Delta, 4\Delta, 5\Delta, 6\Delta\}$. The EDS for three retransmissions of QPSK Gray mapping is given in Table 4.5a). A significant improvement is achieved if different QPSK mappings are applied for the retransmissions, as observed in the EDS in Fig. 4.5b). Anti-Gray QPSK mappings are used in the 2nd and 3rd retransmission and the performance is better than with only Gray mapping, even without a priori information.

D_{ex}^2	Δ	2Δ	3Δ	4Δ	5Δ
Λ	λ_1	λ_2	λ_3	λ_4	λ_5
No a priori information (Expurgated)					
1^{st} tx	4	0	0	0	0
2^{nd} tx	0	4	0	0	0
3^{rd} tx	0	0	4	0	0
Ideal a priori information					
1^{st} tx	4	0	0	0	0
2^{nd} tx	0	4	0	0	0
3^{rd} tx	0	0	4	0	0

a) QPSK Gray.

D_{ex}^2	Δ	2Δ	3Δ	4Δ	5Δ
Λ	λ_1	λ_2	λ_3	λ_4	λ_5
No a priori information (Expurgated)					
1^{st} tx	4	0	0	0	0
2^{nd} tx	0	1	4	0	0
3^{rd} tx	0	0	0	7	0
Ideal a priori information					
1^{st} tx	4	0	0	0	0
2^{nd} tx	0	1	2	0	0
3^{rd} tx	0	0	0	3	1

b) QPSK mixed.

Table 4.5: EDS, symbol mapping diversity, QPSK mappings defined in Table A.9.

4.3 Optimization of Mappings

Mappings for specific applications can be either constructed according to specific rules, e.g. Gray mapping for no a priori information or set partitioning for TCM, or we have to use appropriate optimization algorithms. An exhaustive search becomes impracticable for high order signal constellations ($2^M \geq 8$) since a maximum of $2^M!$ different mappings have to be checked.

No construction methods are known for mappings optimized for ideal a priori information or mappings with an optimized trade-off between the performance without and with a priori information. The optimization of mappings in general can be formulated as a quadratic assignment problem (QAP) [SHGB05] [HR05] which is NP-hard (non-deterministic polynomial-time hard) and therefore not solvable in polynomial time. However, several efficient algorithms have been proposed to solve quadratic assignment problems.

A *binary switching algorithm* as introduced in [ZG90] to optimize the index assignment for vector quantization and applied in [SGHB03a] to optimize mappings of high order signal constellations is a promising method to overcome the complexity problems of the brute-force approach. This algorithm finds through systematic switches of label positions a local optimum on a given cost function. If the algorithm is executed several times with different random initializations, the global optimum may be found with high probability. With an appropriate choice of the cost function, optimized mappings for no a priori information, ideal a priori information or any trade-off may be found. Further heuristic approaches to solve the optimization problem include greedy algorithms and a tabu search algorithm.

Note that these mapping optimization approaches can be extended to optimize signal constellations: we define a large grid where only a few positions are occupied by actual signal points. In the binary switching algorithm, either two labels are switched or a label is placed on an unoccupied grid position.

4.3.1 Optimization Criteria and Cost Functions

The first step in the optimization procedure is to define a cost function for the mapping to be optimized. Values from the Euclidean distance spectrum (EDS), error bounds or mutual information measures are promising candidates to serve as cost function.

First, we consider Chernoff error bounds obtained from the EDS as derived in Appendix B as cost function for the optimization algorithms. Our main goal is not to use very tight error bounds but to have a reliable qualitative measure to define costs of different mappings. We use only the relevant term ω_v that characterizes the influence of the Euclidean distance $d_{\text{ex},v}$ from the EDS in the Chernoff error bound:

$$\omega_v = \exp\left(-\frac{E_s}{4N_0} d_{\text{ex},v}^2\right) \tag{4.7}$$

for the AWGN channel and

$$\omega_v = \frac{1}{d_{\text{ex},v}^2} \tag{4.8}$$

for the fully interleaved Rayleigh fading channel from equation (B.7) and (B.8), respectively. Using the frequencies λ_v from the EDS, the optimization of the mapping μ is done by minimizing the total cost Ω as follows:

$$\min_{\mu}(\Omega) = \min_{\mu}\left(\sum_{v=1}^{V} \lambda_v \cdot \omega_v\right). \tag{4.9}$$

Remember that the frequencies λ_v depend on the applied mapping μ and the number of a priori known bits. Therefore, the mapping μ can be optimized for any number of a priori known bit positions by using the corresponding λ_v values. We can also optimize a trade-off between e.g. the performance without and with ideal a priori information if we define the cost function as

$$\Omega = w_0 \cdot \Omega_0 + w_1 \cdot \Omega_1, \tag{4.10}$$

where Ω_0 and Ω_1 are the overall cost functions without and with ideal a priori information, respectively; w_0 and w_1 are weighting factors.

Instead of error bounds, we can use as cost function Ω mutual information values from the EXIT function $f_{\text{dem}}(I)$. The mutual information is a robust performance measure but expensive Monte-Carlo simulations or numerical integration are required. An approximation of the

mutual information can be obtained from the error bounds: Either by simulation or, according to [tB01b] or [tB01a], using the inverse $\mathrm{erfc}(\cdot)$ or $Q(\cdot)$ function together with the J function given in equation (3.15), we can construct a look-up table that gives the relationship between the error rate and the mutual information. This look-up table may be used to map the results from the error bounds to the mutual information measure. The advantage of this approach is the straightforward weighting of the EXIT chart values to obtain the desired shape of the EXIT function:

$$\Omega = w_0 \cdot f_{\text{dem}}(0) + w_1 \cdot f_{\text{dem}}(1), \tag{4.11}$$

with the negative weighting factors w_0 and w_1. The negative sign is necessary to use the cost function for the minimization problem in equation (4.9).

In addition to the overall cost Ω, we need the cost of the single symbols for the optimization to know the contribution of a specific symbol to a good or bad overall performance. We use the values λ_v from the symbol-wise EDS for the error bounds or the symbol-wise EXIT functions.

Furthermore, we can optimize single bit positions to obtain mappings for equal or unequal error protection. As overall cost Ω, we use the bit-wise EDS for the error bounds or the bit-wise EXIT functions.

4.3.2 Quadratic Assignment Problem and Algorithms

The optimization of mappings can be formulated as a quadratic assignment problem (QAP) [Cel98]. The QAP was introduced by Koopmans and Beckmann in 1957 [KB57]. The QAP can best be described as the problem of assigning a set of facilities to a set of locations with given distances between the locations and given flows between the facilities. The aim is to place the facilities on locations in such a way that the sum of the product between flows and distances is minimal. Beside the facility location problems, the QAP has applications in scheduling, wiring problems in electronics, statistical data analysis, design of control panels and typewriter keyboards, archeology and even sports [Cel98].

The formal description of the QAP is as follows. Given are N facilities and N locations. Let us define the two $N \times N$ matrices $\mathbf{A}$ and $\mathbf{B}$ with elements $a_{k,l}$ and $b_{i,j}$, respectively. $a_{k,l}$ is the distance between locations k and l and $b_{i,j}$ is the flow between facilities i and j. The quadratic assignment problem is then stated as:

$$\min_{\pi \in \mathcal{S}_N} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{\pi(i),\pi(j)} \cdot b_{i,j}, \tag{4.12}$$

where $\mathcal{S}_N$ is the set of permutations of the set of integers $\{1, 2, \ldots, N\}$ and $a_{\pi(i),\pi(j)}$ are the elements of a matrix obtained by permuting the rows and columns of $\mathbf{A}$ according to the permutation π. Here, $a_{\pi(i),\pi(j)} \cdot b_{i,j}$ describes the cost contribution of simultaneously assigning facility i to location $\pi(i)$ and facility j to location $\pi(j)$. For the optimization of mappings, the facilities are the binary bit labels that have to be assigned to the signal points as locations.

The QAP is of interest in the research in various fields, both because of its wide applicability and its difficulty to provide a reliable computer solution. The QAP belongs to the class of NP-complete problems, i.e. the run time to optimally solve this problem cannot be bounded by a polynomial that is a function of the problem size. Exact algorithms and heuristic approaches have been proposed to solve quadratic assignment problems.

Among the exact algorithms used to solve QAPs, branch-and-bound has been the most successful. As the name already reveals, the two steps of this approach is branching and bounding.

Branching is used to split the overall problem in less complex subproblems. Through recursive execution of branching, we obtain a tree structure. The task of bounding is to eliminate some branches of the tree. This reduces the complexity since the eliminated subproblems are no more considered in the subsequent computations. A subregion from the search tree may be safely discarded from the search if, for a minimization task, its lower bound is greater than the upper bound for any other previously examined subregion. Any node whose lower bound is greater than the global minimum upper bound seen among all subregions so far can be discarded. The algorithm stops if ideally all nodes are pruned or solved, i.e. when the upper bound matches the lower bound of a node. The lack of a sharp lower bound is one of the major difficulties. Indeed, either the bound is too loose, or the time needed to compute the bound is prohibitive. Subproblems are often solved with the Simplex method not further discussed here. A software solving quadratic assignment problems using branch-and-bound techniques is available at [BCKR].

Heuristic approaches often provide less reliable optimization results but are easier to implement and fast to compute. We focus on three interesting heuristic approaches, a Greedy algorithm, a tabu search, and the binary switching algorithm.

The greedy algorithm makes at each stage a locally optimum decision, without regard on future consequences. Applying the greedy strategy to the mapping optimization problem yields the following algorithm: "Assign the 2^M binary labels one by one to the signal points in an optimal way at each stage". The total costs after the placement of the first few labels would be very low. However, this strategy does not guarantee to find the global optimal solution when the algorithm is terminated, i.e. when all labels are assigned to signal points. Intuitively, the lack of options when most labels are assigned and most points are occupied inflate the costs. The greedy algorithm is simple to implement and require minimal amount of resources, but often it is not powerful enough to provide satisfactory results.

Another heuristic approach is the tabu search algorithm [BT94][GL02]. The basic idea is to start with a random mapping of the 2^M binary labels and to switch the labels between all possible pairs of signal points. The switch that results in the best overall cost function is finally done. $M(M-1)/2$ switches are tested in the first stage. Then, the actually performed switch is set on a tabu list. The tabu list contains switches performed in the past and that are not allowed to be repeated in subsequent switches to avoid cycles. By selecting the best switch that is not in the tabu list, it may happen that a switch increases the costs, i.e. we may accept new inferior solutions in order to avoid paths already investigated. This allows to investigate new regions in the problem solution space with the goal of avoiding local minima and ultimately finding the desired solution. The tabu list has usually a limited size. Old entries are erased to continue the search. A software for the tabu search algorithm is available at [BB].

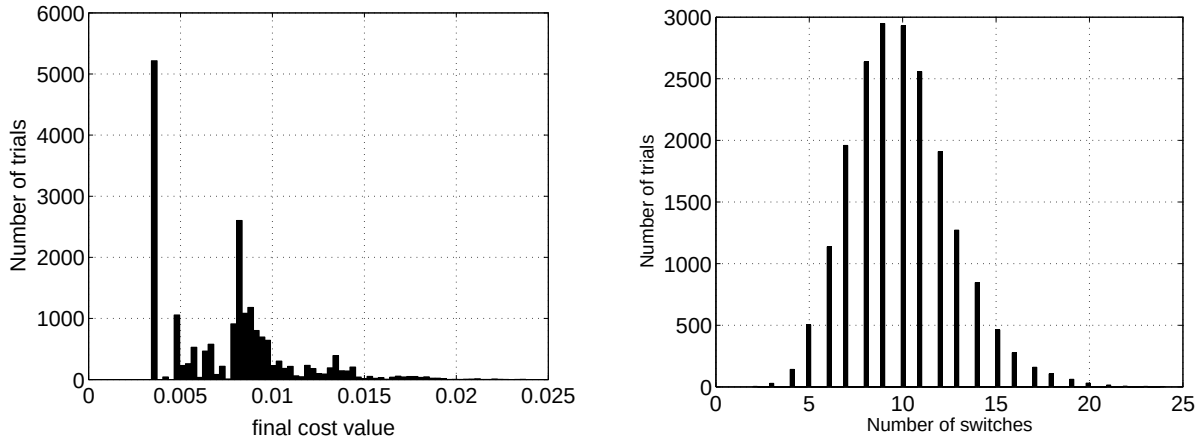
The binary switching algorithm (BSA) is based on a similar idea than the tabu search and was proposed in [ZG90] and [GL03] to optimize the index assignment in the context of vector quantization. We used this simple and yet efficient method in [SGHB03a] [SGHB03b] [SB04a] [SHGB05] to optimize mappings according to a given cost function.

The binary switching algorithm is started with a random initial mapping. Using one of the cost functions, the cost of each symbol and the total cost are calculated. An ordered list of symbols, sorted by decreasing costs, is generated. The idea is to pick the symbol with the highest cost in the list (which has the strongest contribution to a "bad" performance), and to try to switch the label of this symbol with the label of another symbol. The latter is selected such that the decrease of the total cost due to the switch is as large as possible. However, it is not verified if the switch of two other labels with lower costs would result in a larger decrease of the overall cost. If no switch partner can be found for the symbol with the highest cost, the label with

the second-highest cost will be tried to switch next. This process continues for symbols in the list with decreasing costs until a symbol is found that allows a switch that lowers the total cost. After an accepted switch, a new ordered list of symbols is generated, and the algorithm continues as described above until no further reduction of the total cost is possible. The BSA finds a local optimum. Several algorithm executions with random initial mappings may yield to the presumed global optimum.

Fig. 4.19a) depicts the histogram of the costs obtained by the BSA after different random initializations. The example of 16QAM modulation, AWGN channel and the optimization for ideal a priori information at the demapper is investigated. The presumed optimal solution is obtained for approximately $1/4$ of the trials, which is a satisfactory result. For larger signal constellations, where the optimization procedure is more complex, the presumed optimal solution is obtained less frequently, but the runtime for a reliable result is still manageable. The number of switches required to reach a local optimum for this example is depicted in Fig. 4.19b). The average number of switches is approximately 10.

Optimized mappings for different signal constellations are listed in Appendix A. With 16QAM, the mappings $M16a$ and $M16r$ are optimized for ideal a priori information and for the AWGN and Rayleigh fading channel, respectively. The $I16$ mapping has an optimized trade-off between the performance with ideal a priori information and without a priori information to enable low error rates at high SNR and an early convergence of the iterative process, respectively.



a) Cost distribution of BSA trials

b) Number of performed switches

Figure 4.19: Analysis of the binary switching algorithm (BSA) for 16QAM, ideal a priori information, AWGN, $10\log_{10}(E_s/N_0)\text{dB} = 10\text{dB}$; Cost function from equation (4.7) and (4.9), 20000 trials with random initialization: **a)** Histogram of cost values after one execution of the BSA algorithm; minimum cost: 0.003417. **b)** Histogram of number of switches until the BSA converges to a local optimum.

4.3.3 Simulation Results

Fig. 4.20-4.24 depict the bit error rate (BER) performance of BICM with iterative demapping and decoding (BICM-ID) for different scenarios. A different random interleaver is applied for each transmitted block (so-called uniform interleaver). The information block length is $K = 5000$. The other parameters are given in the figure labels.

From Fig. 4.20 and 4.21 we observe that even though the error rate with the Gray and set partitioning mapping is lower at low SNR, significant gains can be achieved at high SNR with other, well chosen mappings. The error bound at high SNR is minimized with the optimized mapping $M16a$. The mapping $I16$ with an optimized trade-off between the performance with and without a priori information is particularly interesting: the difference to the error bound of the $M16a$ mapping is negligible but we achieve a slightly earlier convergence. The error rate performance over the iterations with the $M16a$ mapping is shown in Fig. 4.22.

According to the EXIT chart in Fig. 3.17, the error bound at high SNR is higher with a Rayleigh fading channel than with an AWGN channel. We clearly observe this fact if we compare the error bounds in Fig. 4.21 and 4.23 for the same memory 2 code. Furthermore, Fig. 4.23 illustrates the impact of the choice of the convolutional code: we have a trade-off between an early convergence with a low memory code and a low error bound at high SNR with a high memory code. This fact results from the different shapes of the decoder EXIT functions depicted in Fig. 3.8.

Finally, the impact of the interleaver length is shown in Fig. 4.24. The performance degrades severely with small block length.

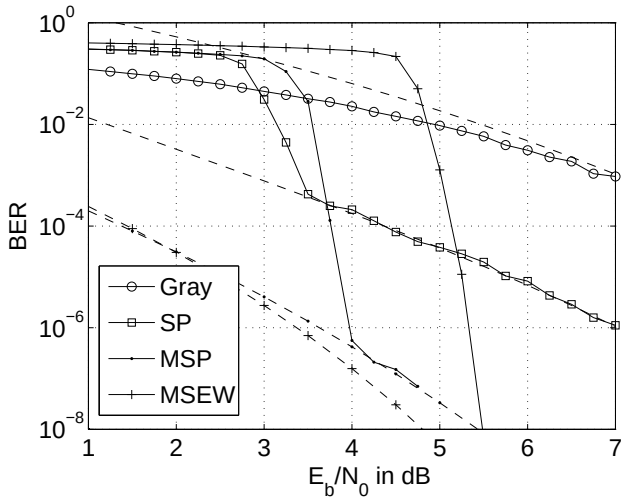


Figure 4.20: BICM-ID, AWGN channel, 16QAM, $(1, 7/5)$ rate $R = 1/2$ conv. code, 25 demapping and decoding iterations; dashed lines: analytical bounds for ideal a priori information from Appendix B; comparison of different previously proposed mappings: Gray, set partitioning (SP) [Ung82], modified set partitioning (MSP) [CR01], maximum squared Euclidean weight (MSEW) [TS02]; compare to the EXIT chart of Fig. 4.17.

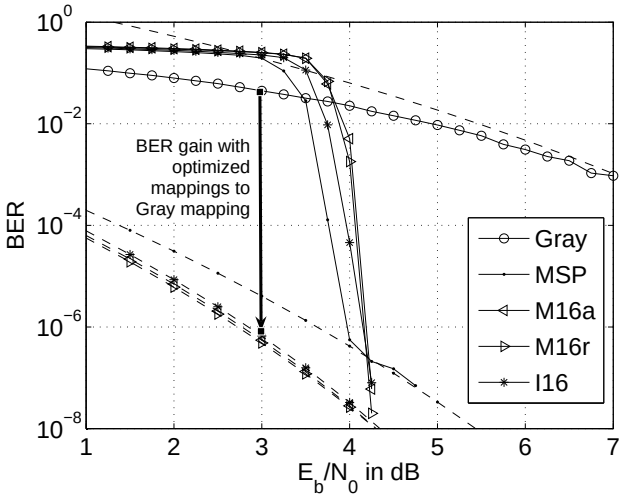


Figure 4.21: BICM-ID, AWGN channel, 16QAM, $(1, 7/5)$ rate $R = 1/2$ conv. code, 25 demapping and decoding iterations, dashed lines: analytical bounds for ideal a priori information from Appendix B, optimized mappings compared to Gray and modified set partitioning (MSP) [CR01].

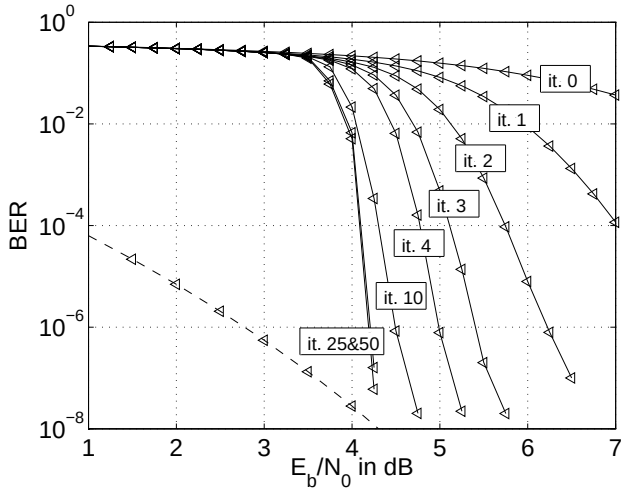


Figure 4.22: BICM-ID, AWGN channel, 16QAM, $M16a$ mapping, $(1, 7/5)$ rate $R = 1/2$ conv. code, performance over the demapping and decoding iterations.

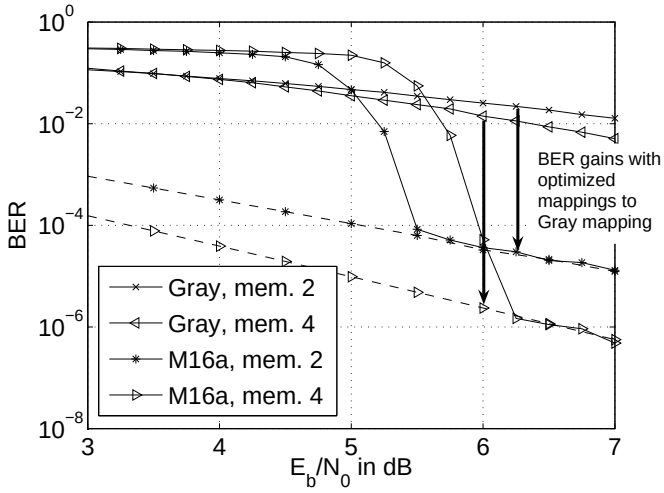


Figure 4.23: BICM-ID, Rayleigh fading channel (symbol-wise fading or fully interleaved fading), 16QAM, Gray and $M16a$ mappings, $(1, 7/5)$ and $(1, 15/17)$, rate $R = 1/2$ conv. codes, 25 demapping and decoding iterations.

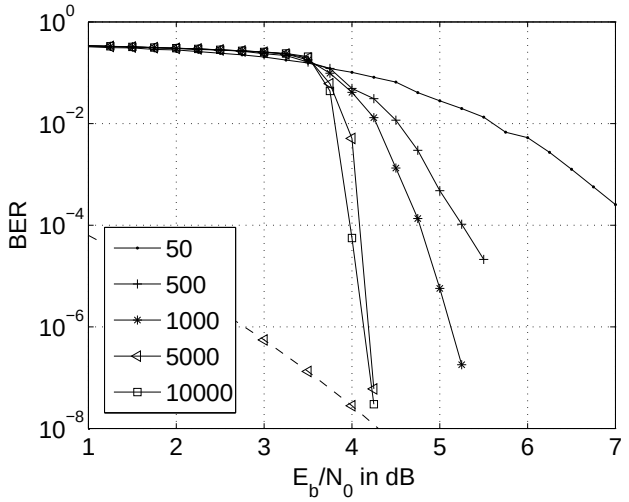


Figure 4.24: BICM-ID, AWGN channel, 16QAM, $M16a$ mapping, $(1, 7/5)$ rate $R = 1/2$ conv. code, 25 demapping and decoding iterations, different information word length K.

4.4 Additional Inner Encoder

To further decrease the error probability with iterative demapping and decoding at high SNR, an additional recursive inner encoder should be used. The encoder should have rate $R \geq 1$ in order not to waste capacity, see Section 3.1. This code adds no redundancy and thus has no error correcting capabilities at all. Nevertheless, due to its recursive structure, it introduces dependencies between all bits in a code sequence, whereas a mapping introduces dependencies only between the M bits associated to one symbol. These dependencies allow arbitrary low error rates with an interleaver length going to infinity (so-called *interleaver gain*) [BDMP98]. Additional inner encoder were used in [TS05b] [DDP00b] [SSS01] [Tüc04] [BMS05] to reduce the error floor in iterative schemes based on BICM. Note that some modulation schemes, e.g. continuous phase modulation scheme (CPM) or differential PSK (DPSK), have an inherent recursive structure and are therefore well suited to serve as inner code for iterative decoding [MA01] [NS99].

We focus on a rate one and memory one inner encoder for complexity reasons. Thus, the straightforward solution would be to use a standard accumulator as inner encoder, since it has a recursive structure and memory one. The recursive rate-1 inner encoder inevitably leads to a performance loss if no a priori information is available. We focus on two promising ideas that have been proposed to increase the performance of the standard accumulator without a priori information:

In [Tüc04], an accumulator operating on all the M bits corresponding to one complex symbol is proposed. If all bits contribute to the accumulator but only one bit-label per symbol is affected by the output of the accumulator, as shown in Fig. 4.25, the performance without a priori information is increased without degrading the performance with ideal a priori information.

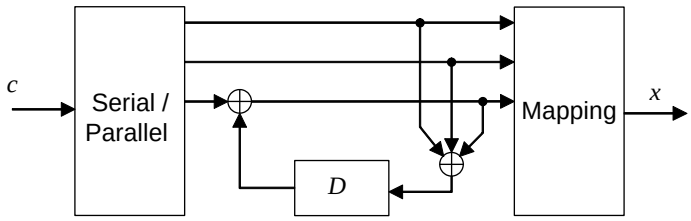


Figure 4.25: Accumulator operating on $M = 3$ bits corresponding to e.g. a 8PSK symbol.

The effect of the structure proposed in [SSS01] and depicted in Fig. 4.26 is similar. Most of the bits at the output of the encoder are systematic bits. Only every Dth systematic bit is replaced ("doped") with a parity bit. Thus, since all bits influence the accumulator state but only every Dth bit is affected by the output of the accumulator, the performance without a priori information is increased for a large doping period D without loosing in performance for ideal a priori information. Note that the system in Fig. 4.25 is only a special case of the system depicted in Fig. 4.26 with $D = 3$ in this example.

The outer channel code, the inner recursive encoder and the mapper form a double serially concatenated system. This structure requires an additional inner detection loop, i.e. iterations between the demapper and the inner decoder. An alternative would be to perform the demapping and the decoding of the inner code on the same trellis. The latter has two states and 2^M incoming/outgoing edges per state. It was shown in [tBK03] that the joint trellis detection is more robust against code parameter mismatches. A receiver without inner detection loop proposed in [SSS01] for the system of Fig. 4.26 is depicted in Fig. 4.27. The a priori information

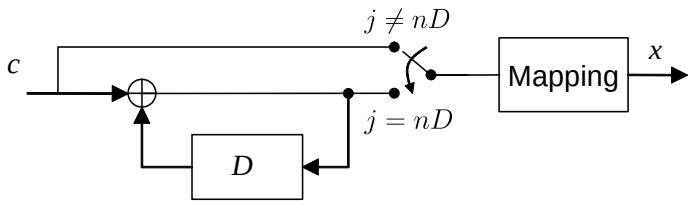


Figure 4.26: Doped accumulator with doping period D.

on the doped parity bits of the inner code is directed to the inner decoder, while the a priori information on the un-doped information bits of the inner code is fed to the demapper. The demapper has no a priori information on every Dth bit, while the inner decoder combines all the available a priori information. This receiver performs well if the doping period is quite high ($D > 10$) or if a mapping is chosen that is less sensitive to a priori information.

With Gray mapping, the feedback to the demapper can be omitted. If e.g. iterative equalization or iterative MIMO detection is done and the demapper is extended to an equalizer or MIMO detector, the receiver depicted in Fig. 4.27 with feedback to the detector should be used even with Gray mapping.

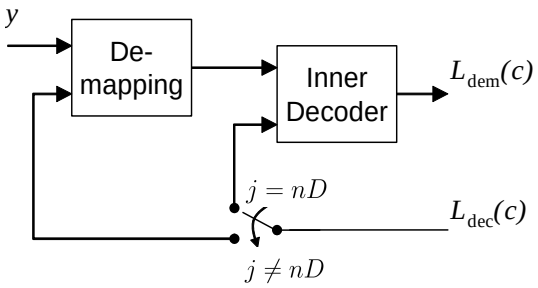


Figure 4.27: Decoder for doped accumulator [SSS01].

The doped inner encoder of Fig. 4.26 can be generalized to the encoder depicted in Fig. 4.28. The doped inner code can be interpreted as punctured rate $1/2$ recursive systematic code. The puncturing ratio of the systematic and coded bits correspond to the doping ratio. The inner code should have rate $R_{\text{in}} \geq 1$ to avoid an inherent capacity loss as explained in Section 3.1. With an outer code of fixed low code rate R_{out}, the rate adaptation can be done by puncturing the inner code to rates larger than one. The characteristics of the inner code are adjusted by carefully choosing the ratio of inner systematic and coded bits. The puncturing matrix

$$\mathbf{P}_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.13}$$

with puncturing period 3 results in the rate-1 inner code with doping period $D = 3$ as depicted in Fig. 4.26. With the puncturing matrix

$$\mathbf{P}_2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{4.14}$$

we obtain a rate $R_{\text{in}} = 4/3$ inner code. A similar approach for serially concatenated codes using systematic doping of the inner code has been proposed in [tB01a], but only for rate $R_{\text{in}} = 1$.

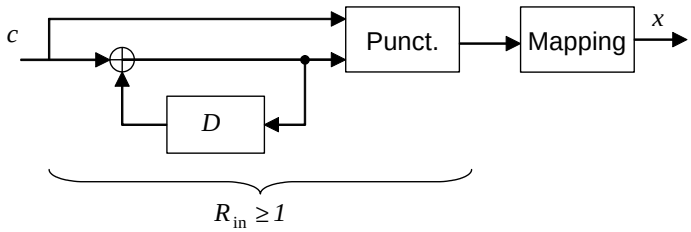
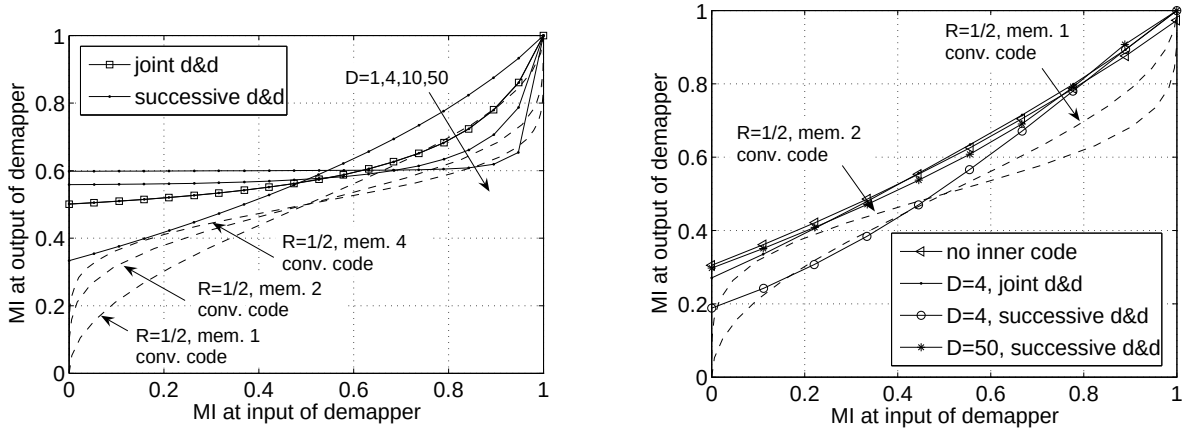


Figure 4.28: Serial concatenated convolutional code with doped inner code of rate $R_{\text{in}} \geq 1$ for rate adaptation.

The elaborated code designs proposed in [BGM+05] and [BMV04] are also based on similar ideas.

Fig. 4.29 depicts EXIT functions $f_{\text{dem}}(I)$ with an inner encoder and decoder according to Fig. 4.26 and 4.27. All curves reach the point $(1, 1)$ in the EXIT chart. The values of $f_{\text{dem}}(0)$ without a priori information are higher with a large doping period D and Gray mapping. Mappings different from Gray may be of interest if further shapes the EXIT function $f_{\text{dem}}(I)$ are required to optimize the receiver. Then, we should take care to use a large doping period D to ensure that the demapper receives most a priori information. In Fig. 4.29b) we observe that the area under the EXIT function with the $M16a$ mapping is lower with $D = 4$ than with $D = 20$. This corresponds to the reduction of the receiver constrained capacity of the considered suboptimal receiver for the combination of low doping periods and mappings sensitive to a priori information.



a) Gray mapping b) $M16a$ mapping

Figure 4.29: EXIT functions with 16QAM, AWGN channel, $10\log_{10}(E_s/N_0)\text{dB} = 7\text{dB}$, recursive inner encoder with different doping periods D, joint demapping and decoding (d&d) on a single trellis, successive d&d with the receiver of Fig. 4.27, rate $R = 1/2$ convolutional codes.

Fig. 4.30 depicts the bit error rate for different convolutional codes and doping periods D with an inner encoder and decoder according to Fig. 4.26 and 4.27, respectively. With the inner encoder, the mapping is set to Gray. The performance without inner encoder with Gray and $I16$ mapping is given for comparison. The channel is AWGN, the information block length is $K = 5000$.

The results correspond to the expectations from the EXIT chart in Fig. 4.29: An early convergence is achieved for the combination of a memory-1 convolutional code and an inner code with doping period $D = 1$ and for a memory-4 code and a doping period $D = 4$. The performance of a memory 2 code with $D = 1$ is similar to the performance without inner code and the optimized $I16$ mapping.

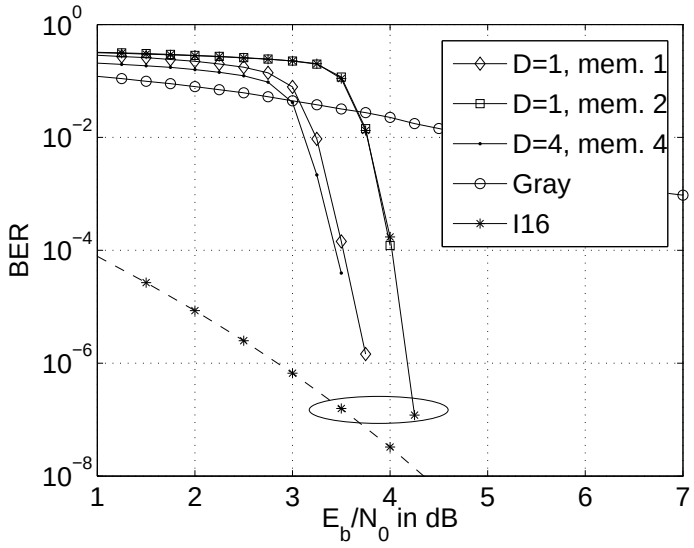


Figure 4.30: BICM-ID, AWGN channel, 16QAM, with and without recursive inner code, $(1, 2/3)$, $(1, 7/5)$ and $(1, 23/35)$ rate $R = 1/2$ conv. codes, 20 iterations, different doping periods D.

4.5 Summary

An Euclidean distance spectrum (EDS) for BICM schemes was derived to characterize mappings for arbitrary signal constellations and any number of a priori known bits. A bit-wise, symbol-wise, expurgated and graphical representation of the Euclidean distance spectrum is introduced. The Euclidean distance spectrum is shown to provide an universal framework to characterize mappings for a wide range of applications.

Then, we showed how the optimization of mappings can be cast to a quadratic assignment problem and a simple binary switching algorithm is used to find new optimized mappings. Cost functions based on the Euclidean distance spectrum, error bounds and mutual information measures were proposed. The main result are mappings optimized for iterative receivers in contrast to Gray mapping optimized for systems without iterative demapping and decoding. The optimization approach can be used to derive optimized mappings for further applications, e.g. equal or unequal error protection, multi-dimensional mappings or mappings for ARQ.

Finally, the use of an additional inner encoder to reduce the error rate at high SNR is discussed.

5

Adaptive Bit-Interleaved Coded Irregular Modulation

A simple method to improve the adaptiveness and flexibility of bit-interleaved coded modulation (BICM) is proposed in this Chapter. The basic idea is to apply different signal constellations and mappings within one code word. We call this approach bit-interleaved coded *irregular* modulation (BICIM) [SB06].

We focus on two advantages of BICIM: First, the combination of different signal constellations allows a fine adaptation of the data rate to the channel characteristics with the modulation. Given knowledge of the average channel quality at the transmitter, we can determine for which fraction of the code word we should use a certain signal constellation. Then, a fractional average number of bits per complex symbol may be obtained on average and we can adapt the data rate very accurately and with fine grid to the channel state. Second, for a receiver performing iterative demapping and decoding, the mixture of different mappings enables an optimization of the iterative decoding process according to the system requirements. Instead of a cumbersome design of new mappings for different applications, a large variety of mapping characteristics is obtained by the combination of a few mappings.

5.1 Adaptive Modulation

The two main applications for bit-interleaved coded irregular modulation (BICIM) are further discussed in this Section, namely the improved adaptation of the communication system to the channel characteristics and to the iterative receiver.

5.1.1 Adaptation to Channel Characteristics

In most wireless mobile communication systems, the propagation environment and thus the channel characteristics are changing over the time. To maximize the achievable data rate over a time varying channel, two basic approaches are usually considered:

First, we can adapt the transmission scheme using channel estimates available at the transmitter. The adaptation may be done by varying the transmitted power level, the size of the signal constellation or the code rate, as discussed in [GC97] and references therein. However, if the channel estimate at the transmitter is unreliable or not available, the system should tolerate a significant outage rate, or the transmission parameters should be set for a worst case scenario, resulting in an inefficient use of the channel capacity due to an unnecessary overhead for good channel conditions.

To avoid these drawbacks, automatic repeat request (ARQ) schemes have been proposed. There, we start by assuming good channel characteristics. If the decoding at the receiver fails, the data is retransmitted. The performance is further improved if we do not simply retransmit the data, but if additional redundancy for the same information data is transmitted. This *incremental redundancy* approach is usually realized with rate-compatible punctured codes (RCPC) [Hag88]. With a fine granularity, the channel capacity is well exploited, but the signaling overhead and the delay may be significant when multiple retransmissions are required.

A reasonable approach is to combine channel adaptation to reduce the delay and ARQ schemes to maximize the throughput.

We focus on the task of channel adaptation with the signal constellation and with the knowledge of only the average channel quality at the transmitter. With state-of-the-art adaptive modulation and coding (AMC), as e.g. included in the recent EDGE [3GPP01] (enhanced data for GSM evolution), HSDPA [3GPP06a] (high speed downlink packet access), IEEE 802.11 WLAN [IEEE99] (wireless local area network) and IEEE 802.16 WiMAX [IEEE04] (worldwide interoperability for microwave access) standards, the channel code rate can be varied in small steps using appropriate puncturing patterns. However, the choice of the signal constellation allows only a very raw adaptation to the channel quality since the granularity is at least one bit per symbol.

If we extend these systems to bit-interleaved coded irregular modulation (BICIM), where different signal constellations may be used *within* one code word, a fractional average number of bits per complex symbol may be obtained. Thus, a highly flexible adaptation to the channel quality is now possible with the modulation. Only a few system extensions are required to apply BICIM. The system should be able to switch the signal constellation within one code word and to support the increased switch rate. However, already one or two switches within a code word are often enough to exploit the advantages of BICIM, as discussed later on. Furthermore, the signalling should be adapted to be able to specify several signal constellations within one code word and to specify the positions within the code word when the switches take place.

Different signal constellations within one code word have already been widely investigated for fading channels [GC97][ÖLGW01] and OFDM systems [CCB95][FH96], where according to the channel quality of a sub-carrier, a corresponding signal constellation is chosen. Then, the instantaneous channel state should be available at the transmitter. However, we focus on the scenario where only the average channel quality of one code word is known at the transmitter. We may assign different signal constellations within one code word to maximize the achievable rate even if the channel quality is constant over that code word.

Note that the combination of different signal constellations may also help in carrier recovery [HS89].

5.1.2 Adaptation to Iterative Receiver

In addition of using different signal constellations, we may assign different bit-to-symbol mappings within one code word. This approach is useful in a system where the demapper is part of the iterative process at the receiver. As mentioned in Section 3.4, we would stick to Gray mapping if a capacity approaching "strong" channel code is used. However, if we apply a "weak" convolutional code, a carefully chosen mapping different from Gray should be used. Then, we can select the mapping to optimize the convergence of the iterative demapping and decoding procedure, to set the number of required iterations, or to minimize the error bound at high SNR.

The combination of different mappings within one code word offers advanced optimization possibilities since a large number of mapping characteristics can be derived out of a small set of underlying mappings, as also investigated recently in [SCH05].

Note that the basic idea is similar to irregular channel codes, where the combination of different code rates within one code word offers advanced optimization possibilities of the iterative decoding of concatenated codes. Irregular channel codes have been proposed in [FM00] and [TH02] for parallel and serial concatenated convolutional codes, respectively, and in [LMS+97] [RU01] for LDPC codes.

5.2 System Structure with Irregular Modulation

The bit-interleaved coded modulation (BICM) system with optional iterative demapping and decoding (BICM-ID) was introduced in Section 2.5. This system is extended to bit-interleaved coded irregular modulation (BICIM) as depicted in Fig. 5.1, where different signal constellations $\mathcal{X}_p$ and mappings μ_p may be used within one code word.

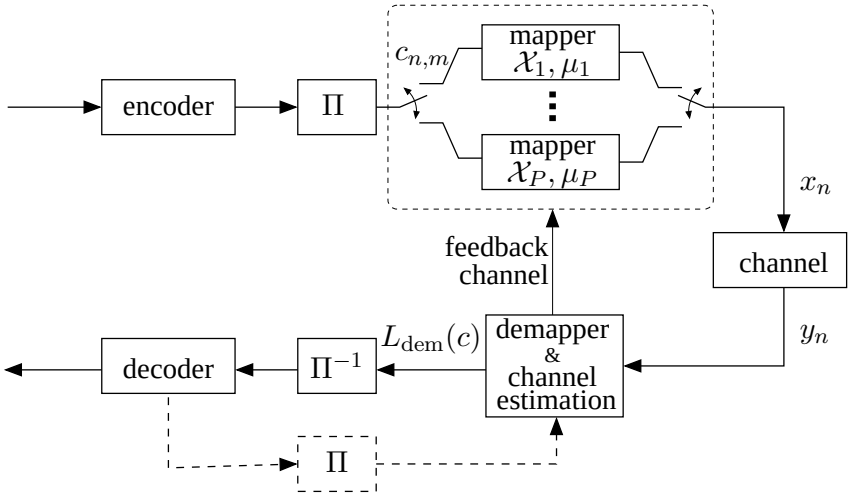


Figure 5.1: BICIM system model with optional iterative demapping and decoding.

The transmitter operates as follows: A sequence of information bits is encoded and bit-interleaved by a random interleaver Π. The interleaved sequence of N code bits is divided in P subblocks of length $\alpha_p N$, as illustrated in Fig. 5.2, where α_p is the ratio of the pth subblock, $p = 1, \ldots, P$.

M_p consecutive bits of the subblock p are grouped to form the subsequence $\mathbf{c}_n = (c_{n,1}, \ldots, c_{n,M_p})$. Each subsequence $\mathbf{c}_n$ is mapped to a complex symbol $x_n = \mu_p(\mathbf{c}_n)$

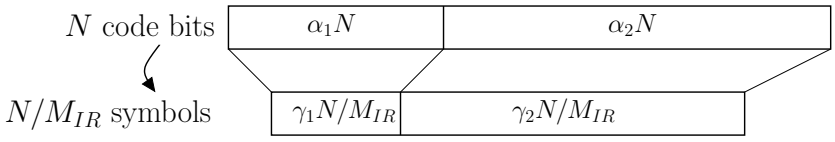


Figure 5.2: Construction of irregular modulation schemes.

chosen from the 2^{M_p}-ary signal constellation $\mathcal{X}_p$ according to the binary labeling map $\mu_p : \{0,1\}^{M_p} \to \mathcal{X}_p$. An inner recursive rate-1 encoder may be used in conjunction with the mapping as described in Section 4.4. A different signal constellation $\mathcal{X}_p$ and/or mapping μ_p is used in each subblock. Then, an average of $M_{IR} \in \mathbb{R}^+$ coded bits per symbol is obtained.

The ratios α_p are chosen according to the information available at the transmitter. This information may include the knowledge of the average channel quality obtained through a feedback channel from the receiver or quality of service (QoS) requirements of the transmitted data. The ratios $\alpha_p \in [0,1]$ must satisfy the conditions

$$\sum_{p=1}^{P} \alpha_p = 1 \quad \text{and} \quad \sum_{p=1}^{P} \alpha_p / M_p = 1/M_{IR}. \tag{5.1}$$

The second condition follows from the computation of the symbol block length $N/M_{IR} = \sum_{p=1}^{P} N \cdot \alpha_p / M_p$. The ratios $\gamma_p = \alpha_p \cdot M_{IR}/M_p$ may be used instead of the ratios α_p to determine the segmentation of the code block on a symbol level instead of on a bit level, as illustrated in Fig. 5.2. Then, with $\gamma_p \in [0,1]$, the following conditions should be satisfied:

$$\sum_{p=1}^{P} \gamma_p = 1 \quad \text{and} \quad \sum_{p=1}^{P} \gamma_p M_p = M_{IR}. \tag{5.2}$$

The following discussion includes the AWGN and Rayleigh fading channel models as introduced in equation (2.8). We focus on the scenario where the instantaneous channel state is available at the receiver but only the average channel state is known at the transmitter.

The receiver is similar to the BICM receiver described in Section 2.5 with optional iterative demapping and decoding. The only modification is that the demapper must be able to switch the signal constellation and/or mapping within one code word.

5.3 EXIT Chart Analysis and Optimization

EXIT charts have been investigated in Chapter 3 as a tool to analyze and optimize the convergence of iterative systems. We use EXIT charts and linear programming methods to optimize and efficiently use irregular modulation schemes.

With equation (3.10) or by using the results of [AKtB04], the average mutual information I_{dem} at the output of the demapper with irregular modulation is the linear combination of the average mutual information $I_{\text{dem},p}$ of the P subblocks of length $\alpha_p N$ bits:

$$I_{\text{dem}} = \sum_{p=1}^{P} \alpha_p \cdot \left(\frac{1}{\alpha_p N} \sum_{j=1}^{\alpha_p N} I(C_{p,j}; L_{p,j}) \right) = \sum_{p=1}^{P} \alpha_p \cdot I_{\text{dem},p}, \tag{5.3}$$

where $C_{p,j}$ and $L_{p,j}$ are the jth bit after the encoder and LLR value after the demapper of the pth subblock, respectively. Therefore, the demapper EXIT function $f_{\text{dem}}(I)$ is the linear combination of the EXIT functions $f_{\text{dem},p}(I)$ of the P subblocks:

$$f_{\text{dem}}(I) = \sum_{p=1}^{P} \alpha_p \cdot f_{\text{dem},p}(I). \tag{5.4}$$

As stated in Section 3.4, we have to match the EXIT function of the demapper to the function of the decoder to achieve close to capacity performance. With irregular modulation, we can obtain a high number of possible demapper characteristics out of a small number of underlying signal constellations and mappings. Both the data rate and the desired shape of the EXIT function are adjusted by setting the ratios α_p in equation (5.4) in an appropriate way. By combining different signal constellations, we can shift the demapper curve upwards or downwards in the EXIT chart. By combining different mappings, we can change the slope and in general the shape of the demapper curve. A fine adjustment of the data rate and a precise design of the shape of the demapper EXIT function is now easily possible.

Let $f_{\text{dec}}^{-1}(I)$ denote the inverse EXIT function of the channel decoder as plotted in Fig. 5.3, 5.4 and 5.5. Similar to the optimization of irregular convolutional codes [Tüc04], the optimization of the ratios α_p for a fixed channel code to maximize the data rate can be formulated as a linear programming problem:

$$\begin{aligned} \text{minimize} \quad & 1/M_{IR} = \textstyle\sum_{p=1}^{P} \alpha_p / M_p, \\ \text{subject to} \quad & \textstyle\sum_{p=1}^{P} \alpha_p \cdot f_{\text{dem},p}(I) > f_{\text{dec}}^{-1}(I), \\ & \textstyle\sum_{p=1}^{P} \alpha_p = 1, \quad \alpha_p \in [0,1], \forall p. \end{aligned} \tag{5.5}$$

The first constraint is to ensure that the EXIT function of the demapper is always above the one of the decoder and that they do not intersect. Since the EXIT function of the demapper is not reaching the point $[1;1]$ in the EXIT chart, a crossing with the function of the decoder is unavoidable at a priori values I from the decoder close to 1. Thus, the first condition in equation (5.5) should be relaxed to be valid up to a value of I close to 1.

Fig. 5.3 shows the result of the optimization for the combination of a turbo decoder with a mixture of QPSK and 16QAM, both with Gray labeling. Note that if a more complex scenario than in Fig. 5.3 is chosen, i.e. if more than two signal constellations and mappings are combined, the proposed linear programming method will be more useful. Further examples of EXIT functions of irregular modulation are shown in Fig. 5.4 and 5.5, where the ratios α_p of the underlying signal constellations and mappings are not optimized and only chosen according to illustrative purposes: In Fig. 5.4, the combination of Gray and $M16a$ 16QAM mappings results in an earlier convergence and a lower number of iterations than if only $M16a$ mapping would be used. However, a higher error bound with ideal a priori information is expected. This will be confirmed in Section 5.5. EXIT functions for irregular modulation with an additional inner encoder are depicted in Fig. 5.5. Gray and $M16a$ 16QAM mappings are combined to shape the EXIT function of the demapper to allow the trajectory to sneak through the open tunnel.

5.4 Capacity Analysis and Optimization

In the previous Section, we have optimized irregular modulation schemes with the EXIT chart to achieve close to capacity performance. In the following, we investigate the information theoretical limits and the actual capacity of irregular modulation.

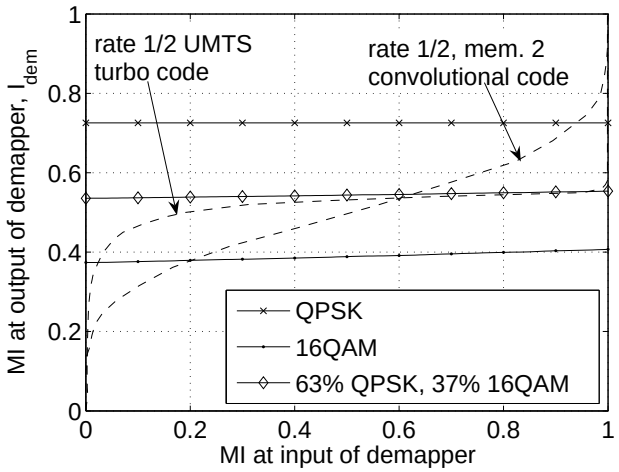


Figure 5.3: EXIT functions with Gray mapping, AWGN channel, $10\log_{10}(E_s/N_0)\text{dB} = 3\text{dB}$. Optimized combination of QPSK and 16QAM signal constellations with $\gamma_1 = 0.63$ and $\gamma_2 = 0.37$, respectively. $M_{IR} = 2.74$.

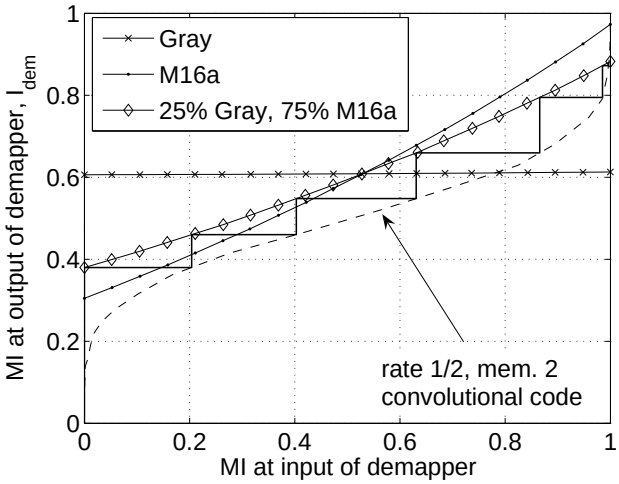


Figure 5.4: EXIT functions with 16QAM, AWGN channel, $10\log_{10}(E_s/N_0)\text{dB} = 7\text{dB}$. Combination of Gray and $M16a$ mapping with $\gamma_1 = 0.25$ and $\gamma_2 = 0.75$, respectively.

The constellation constrained capacity $\mathcal{C}$ of the signal constellation $\mathcal{X}$ is given by the average mutual information [CTB98]

$$\mathcal{C} = \frac{M}{N} \sum_{n=1}^{N/M} I(X_n, Y_n), \tag{5.6}$$

where X_n and Y_n denote the random variables for the nth transmitted and received symbol, respectively. Recall that the mapping has no influence on the capacity. The capacity $\mathcal{C}_{IR}$ with irregular modulation is the linear combination of the capacities $\mathcal{C}_p$ of the P subblocks of length $\gamma_p N/M_{IR}$ symbols:

$$\mathcal{C}_{IR} = \sum_{p=1}^{P} \gamma_p \cdot \left(\frac{M_{IR}}{\gamma_p N} \sum_{n=1}^{\gamma_p N/M_{IR}} I(X_{p,n}, Y_{p,n}) \right) = \sum_{p=1}^{P} \gamma_p \cdot \mathcal{C}_p, \tag{5.7}$$

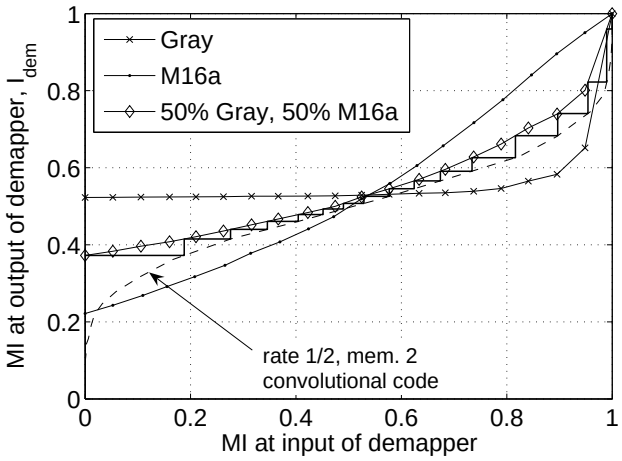


Figure 5.5: EXIT functions with 16QAM, AWGN channel, $10\log_{10}(E_s/N_0)\text{dB} = 6\text{dB}$, recursive inner encoder with doping period $D = 20$ according to Fig. 4.26. Combination of Gray and $M16a$ mapping with $\gamma_1 = \gamma_2 = 0.5$.

where $X_{p,n}$ and $Y_{p,n}$ are the nth values of the pth subblock.

As mentioned in Section 5.1, the power level, the size of the signal constellation and/or the code rate should be adapted to the channel conditions to maximize the achievable data rate. If the instantaneous channel quality is available at the transmitter, the power and the size of the signal constellation may be adapted within one code word frame according to optimized power loading and bit loading strategies. With an AWGN, block fading or symbol fading channel with only average channel state information at the transmitter and equal power distribution, equal bit loading within one frame would be optimal. For these scenarios irregular modulation schemes may lead to a suboptimum solution since the bit loading within one frame may not be constant. However, we will see that the loss is negligible for some system setups and that the advantages of irregular modulation may predominate.

In the following we investigate with simple examples the performance of irregular modulation schemes in terms of achievable data rates, i.e. in terms of achievable number of information bits per channel use.

Arbitrary Code Rate, Constrained Modulation

First, we vary the channel code rate with the channel quality in arbitrarily small steps and fix the modulation scheme. If irregular modulation schemes are used, the ratios γ_p, or equivalently, the ratios α_p are fixed and independent from the channel quality.

Fig. 5.6 and 5.7 show for this scenario the achievable rates of regular and irregular modulation for an AWGN and fully interleaved Rayleigh fading channel, respectively.

From an information theoretical point of view, the best would be here to always use 16QAM and just vary the code rate. However, this may not be desired in real world systems, because of e.g. sensitivity to nonlinear distortions or complexity. Irregular modulation schemes offer the possibility to achieve any maximum transmission rate between 2 (QPSK) and 4 (16QAM) bits per channel use in this example with only small performance degradation, especially in combination with low to moderate channel code rates.

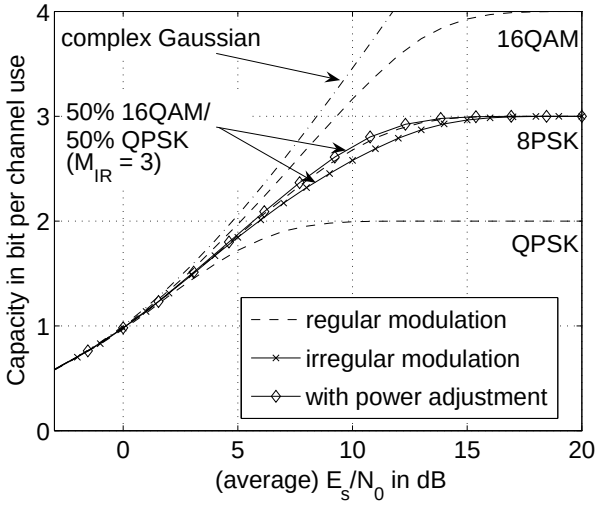


Figure 5.6: Capacity of regular and irregular modulation for an AWGN channel; irregular modulation without and with optimized power distribution.

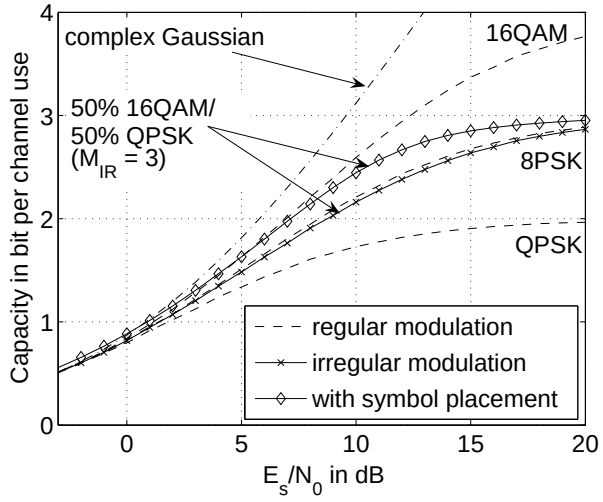


Figure 5.7: Capacity of regular and irregular modulation over a fully interleaved fading channel, equal power distribution; irregular modulation without and with optimized placement of symbols belonging to different signal constellations.

In Fig. 5.6 the effect of an optimized power assignment to the different signal constellations is investigated for an AWGN channel with irregular modulation and $\gamma_{QPSK} = \gamma_{16QAM} = 0.5$. In general more power should be assigned to large signal constellations and less power to small signal constellations. The gain is significant only for data rates of 2 bits per channel use and more, i.e. for a channel code with rate $R > 2/3$ in this example.

With an AWGN or block fading channel, the placement of the symbols belonging to different signal constellations is not relevant due to the subsequent interleaver. However, with a symbol fading channel, symbols from high order signal constellations should be transmitted when the channel quality is better and symbols from low order signal constellations when the channel quality is worse, according to optimized bit loading strategies [GC97]. Fig. 5.7 depicts for irregular modulation with $\gamma_{QPSK} = \gamma_{16QAM} = 0.5$ and equal power distribution the achievable

rates if the instantaneous instead of only the average channel state information is available at the transmitter and if the symbols are placed accordingly. This scenario may be realistic in an OFDM system, where the estimated quality of the subcarriers may be available at the transmitter. The gain is quite substantial for data rates above 1.5 bits per channel use, i.e. for a channel code with rate $R > 1/2$ in this example.

Constrained Code Rate, Arbitrary Modulation

Now we allow only a small number of code rates and use irregular modulation to adapt the data rate to the channel characteristics. To maximize the average number of coded bits per symbol M_{IR} and thus the achievable data rate, the ratios γ_p should be optimized for every channel state. If more than two signal constellations are combined, it is useful to describe the optimization as a linear programming problem. The problem can be formulated as

$$\begin{aligned} \text{maximize} \qquad & M_{IR} = \textstyle\sum_{p=1}^{P} \gamma_p M_p, \\ \text{subject to} \quad & \textstyle\sum_{p=1}^{P} (\mathcal{C}_p - R M_p) \cdot \gamma_p > 0, \\ & \textstyle\sum_{p=1}^{P} \gamma_p = 1, \quad \gamma_p \in [0,1], \forall p. \end{aligned} \tag{5.8}$$

The first condition follows from the constraint of a fixed channel code rate R, stating that

$$\mathcal{C}_{IR} = \sum_{p=1}^{P} \mathcal{C}_p \gamma_p > R \cdot \sum_{p=1}^{P} M_p \gamma_p = R \cdot M_{IR}. \tag{5.9}$$

Fig. 5.8 depicts the achievable rates with regular and irregular modulation if the code rates $R = 1/4$, $1/2$ and $3/4$ and the modulation schemes QPSK, 16QAM and 64QAM are used. The code rates and signal constellations are chosen in an optimum way with respect to capacity. The staircase trajectory indicates the maximum achievable rates with regular modulation. With irregular modulation, we observe that close to optimum data rates may be achieved. The shaded area is the rate gain of irregular over regular modulation.

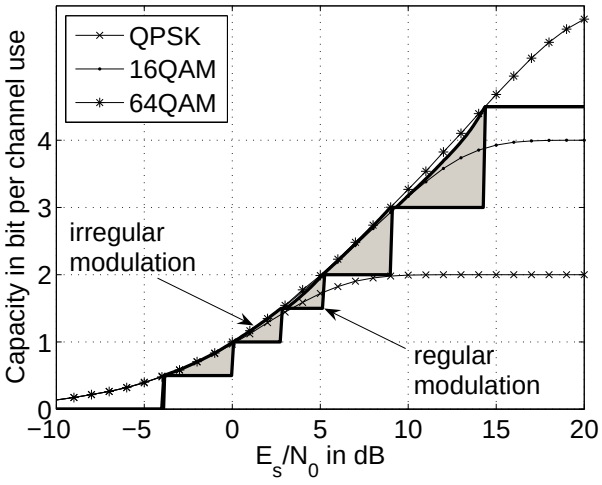


Figure 5.8: Achievable data rates of irregular modulation with AWGN channel, three channel code rates $R = 1/4$, $1/2$ and $3/4$ and optimized ratios γ_p for the combination of three signal sets (QPSK, 16QAM, 64QAM) with equal power loading. The shaded area is the rate gain of irregular over regular modulation.

5.5 Simulation Results

Fig. 5.9 and 5.10 depict the bit error rate (BER) performance with different irregular modulation schemes. The channel is AWGN, the information word length is 5000 bits, and we use an uniform (random) interleaver between the channel code and the mapping.

First, the BER performance for the combination of different signal constellations is investigated in Fig. 5.9. We apply 16QAM and QPSK modulation with Gray mapping in combination with the UMTS turbo code defined in equation (2.69), similar to the EXIT chart in Fig. 5.3. The UMTS interleaver is used to reduce the error floor. 20 iterations between the constituent codes of the turbo code but no iterations between the demapper and the decoder are performed. With a fixed code rate of $R = 1/2$, we can adjust the data transmission rate to the SNR in an arbitrary way by setting the ratios of 16QAM and QPSK.

Second, the BER performance with the combination of different mappings is shown in Fig. 5.10. We use the same system parameters as in the EXIT chart example of Fig. 5.4, i.e. 16QAM modulation with Gray and $M16a$ mapping and the $(1, 7/5)$ convolutional code. The performance after 20 demapping and decoding iterations is depicted. We can set the trade-off early convergence and low number of required iterations vs. low error bound with ideal a priori information by varying the ratios of the underlying mappings. As depicted in Fig. 5.10, Gray mapping has the best performance at low SNR, $M16a$ the best at high SNR. By mixing 25% of Gray mapping with 75% of $M16a$ mapping, we achieve an earlier convergence than with only $M16a$ mapping at the expense of a higher error bound at high SNR.

To obtain the analytical error bounds with irregular modulation, we have to extend equation (B.9) given in Appendix B and average over the P subblocks to compute the average Laplace transform $\Phi_\Delta(s)$ of the probability density function $f_\Delta(\Delta)$ of the overall decision metric Δ:

$$\Phi_{\Delta}(s) = \left(\sum_{p=1}^{P} \gamma_p \cdot \frac{1}{\sum_{v=1}^{V} \lambda_{p,v}} \sum_{v=1}^{V} \lambda_{p,v} \cdot \Phi_{\Delta(d_{\mathrm{ex},v})}(s) \right)^{d} . \tag{5.10}$$

where $\lambda_{p,v}$ is the frequency in the Euclidean distance spectrum (EDS) of the distance $d_{\mathrm{ex},v}$ in the pth subblock.

From the equations (B.7) and (B.8) related to the error rate, we expect that small Euclidean distances will be more harmful in an AWGN channel than in an fading channel. If e.g. a mapping optimized for ideal a priori information is combined with a Gray mapping, the performance with ideal a priori information at the demapper (i.e. with genie or error free feedback) is degraded quite severely in an AWGN channel, as observed in Fig. 5.10.

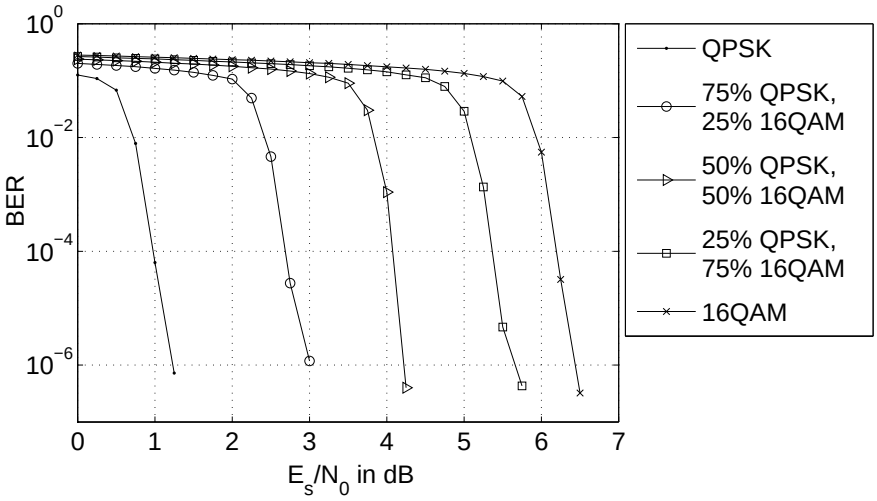


Figure 5.9: BICIM, AWGN channel, QPSK and 16QAM with Gray mapping, rate $R = 1/2$ UMTS turbo code, 20th iteration; Example of different data transmission rates $R_t = \{1, 1.25, 1.5, 1.75, 2\}$ information bits per channel use.

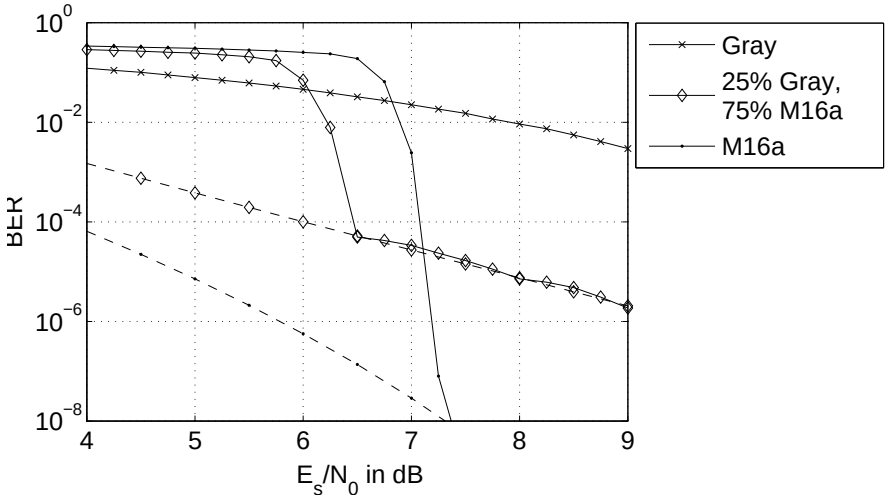


Figure 5.10: BICIM, AWGN channel, 16QAM with Gray and $M16a$ mapping; $(1, 7/5)$, rate $R = 1/2$ conv. code, data rate $R_t = 2$; 20 demapping and decoding iterations.

5.6 Summary

A bit-interleaved coded irregular modulation (BICIM) scheme was introduced, where different signal constellations and mappings may be used within one code word, even if only the average channel quality is available at the transmitter. The proposed system offers a promising possibility to adapt the transmission system to the channel quality, the complexity and error rate requirements. The analysis included EXIT charts to analyze and optimize the convergence of iterative demapping and decoding, capacity to determine the fundamental limits, error bounds to predict the performance at high SNR and error rate simulations to validate the results. In particular, we can state that:

- BICIM is promising if only a limited number of channel code rates are available or desired. The fine adaptation of the data transmission rate to the channel quality could be performed by combining different signal constellations.
- By combining different mappings, we optimize the convergence of the iterative decoding and demapping procedure, similar to the effect of irregular codes. Instead of using a large number of different mappings, the desired mapping characteristics can be obtained by combining only two mappings, namely Gray mapping and a mapping optimized for iterative demapping and decoding. The iterative decoding threshold, the required number of iterations and the remaining error rate can be adjusted in a very flexible way.
- The additional complexity of BICIM is very low. The transmitter should support only a small number of signal constellations and mappings. Furthermore, the rate at which the transmitter has to change the signal constellation and mapping is not excessive if only the average channel characteristics are considered.
- BICIM is well suited for the combination with bit loading and power loading schemes if additional channel knowledge is available at the transmitter.

6

Combination of Signal Shaping and Bit-Interleaved Coded Modulation

In most state-of-the-art communication systems using high order modulation schemes, each signal point of the selected signal constellation is equally likely to be transmitted. However, for a continuous channel input alphabet, the pdf of the received signal should be Gaussian to approach Shannon's channel capacity, as derived in Section 2.2. Similarly, with a finite size channel input alphabet, a so-called shaping gain is achieved if the standard uniform probability distribution of the signal points is replaced by a Gaussian-like distribution.

Gains of more than 1dB may be achieved in practice and it is often easier to obtain a shaping gain than to obtain a similar gain through more powerful coding. Furthermore, the adaptation of the probability distribution of the signal points to the channel characteristics is an inherent way of "soft" adaptive modulation: using optimized probability distributions, the high energy signal points become more probable with increasing SNR.

To obtain non-equiprobable signal points, we assign a non-uniform number of distinct bit labels to the signal points. The proposed method is integrated in the bit-interleaved coded modulation scheme with iterative demapping and decoding (BICM-ID). The channel code and the iterative process resolve the ambiguities that arise when multiple bit labels are assigned to one signal point. A greedy algorithm is derived to obtain a close to optimum probability distribution of the signal points and we use the binary switching algorithm introduced in Section 4.3 to optimize the bit-to-symbol mapping. At a rate of 3 bits per channel use, we obtain capacity gains over 1dB for a 16ASK signal constellation. In combination with irregular convolutional codes, we achieve an iterative decoding threshold at 0.7dB below the capacity of the uniform distribution.

6.1 Introduction to Signal Shaping

The aim of signal shaping is to "shape" the probability distribution of the transmitted signal to achieve a capacity gain. Fig. 6.1 shows as example Shannon's AWGN channel capacity given in

equation (2.20) valid for a Gaussian distributed continuous channel input and the constellation-constrained capacity for uniformly distributed ASK signal constellations [FU98]. We observe that at high SNR, a gap to the AWGN capacity remains for all signal constellations. This gap can be reduced through signal shaping methods.

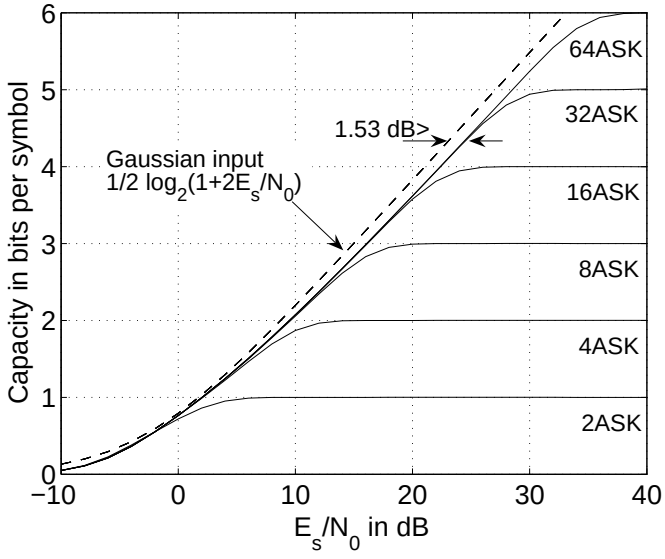


Figure 6.1: Channel capacity with ASK modulation schemes.

Two basic approaches have been proposed for signal shaping: Either the signal constellation is non-uniformly spaced [SvT93] or the signal points have a non-uniform probability distribution [CO90]. We use standard uniformly spaced signal constellations and adapt the probabilities of the signal points.

The following quantities are of special interest throughout this Chapter: average energy, entropy $H(X)$ of transmitted symbols, and peak-to-average power ratio (PAPR). For a signal constellation $\mathcal{X}$ with bit labels of length M, the average symbol energy is

$$\mathrm{E}\{|x_i|^2\} = \sum_{i=1}^{|\mathcal{X}|} P(x_i) \cdot |x_i|^2, \tag{6.1}$$

where $P(x_i)$ is the probability of the signal point x_i. The entropy $H(X)$ of a signal constellation is the average number of bits per signal point and is given from equation (2.10) by

$$H(X) = -\sum_{i=1}^{|\mathcal{X}|} P(x_i) \cdot \log_2(P(x_i)). \tag{6.2}$$

For equiprobable signal points with $P(x_i) = 1/|\mathcal{X}|$, the entropy is maximized to $H(X) = \log_2(|\mathcal{X}|)$. If $|\mathcal{X}| = 2^M$ and $P(x_i) = 1/2^M$, $\forall i$, we have a standard uniformly distributed signal constellation with a unique bit-to-symbol mapping and the maximum entropy is $H(X) = M$. With signal shaping, the signal points with less energy are used more frequently than points with high energy to reduce the average symbol energy. Such non-uniform signaling reduces the entropy $H(X)$ and thus the data rate. However, the energy savings may more than compensate for this loss of data rate.

Note the duality to source coding: With signal shaping, an equiprobable, redundancy free bit sequence is converted to a non-equiprobable symbol sequence and a *shaping redundancy* $r_s = M - H(X)$ is introduced. It is well known that source coding performs the reverse operation.

A drawback of signal shaping techniques is that the peak-to-average power ratio (PAPR) is increased. For the signals points x_i from the constellation $\mathcal{X}$, the PAPR is defined as

$$\text{PAPR} = \frac{\max_{\forall x_i \in \mathcal{X}}(|x_i|^2)}{\text{E}\{|x_i|^2\}}. \tag{6.3}$$

A large PAPR results in high energy peaks in the modulated signal and high demands on the linear amplifier.

6.1.1 Optimized Probability Distribution

The probability distribution that minimizes the average energy for a given entropy $H(X)$ or that maximizes $H(X)$ for a given average energy is the so-called *Maxwell Boltzmann* distribution obtained through a Lagrange optimization [KP93] [Fis02]:

$$P(x_i) = K(\lambda) \cdot \text{e}^{-\lambda |x_i|^2}, \quad \lambda \geq 0. \tag{6.4}$$

The probability distribution is therefore discrete Gaussian. The factor

$$K(\lambda) = \left(\sum_{m=1}^{2^M} e^{-\lambda |x_i|^2} \right)^{-1} \tag{6.5}$$

normalizes the distribution and the parameter λ governs the trade-off between average energy and entropy of signal points. With $\lambda = 0$, we obtain a uniform distribution, whereas with $\lambda \to \infty$, only the signal points closest to the origin remain.

However, when transmitting over an AWGN channel with a given average energy, the quantity to optimize is not the entropy $H(X)$ of the channel input but the mutual information $I(X;Y)$. Instead of the mutual information, we can optimize the differential entropy $h(Y)$ of the channel output since $I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(N)$ and $h(N)$ is independent of the distribution of X. We have to distinguish between the energy gain achieved for a fixed $H(X)$ and the gain for a fixed $I(X;Y)$ or $h(Y)$. At low SNR, the energy gain with a fixed $I(X;Y)$ or $h(Y)$ is smaller [WFH99]. For large SNR, both gains are similar since then, $h(Y) \approx H(X)$. We focus on the energy gain with a fixed $I(X;Y)$ or $h(Y)$ and denote it as shaping gain.

A Gaussian distributed continuous channel input signal maximizes the mutual information $I(X;Y)$ in an AWGN channel. However, a continuous signal is not of practical relevance in a digital communication system and we transmit discrete signal points. Then, the task of finding the probabilities $P(x_i)$ of the signal points that maximize the mutual information $I(X;Y)$ is not trivial. Blahut and Arimoto independently introduced an iterative method to determine the optimized probabilities $P(x_i)$ [Bla72][Ari72].

As shown in [WFH99], the performance of the optimum and not necessarily Gaussian distribution from the Blahut-Arimoto algorithm is closely approached with the discrete Gaussian distribution of equation (6.4). Therefore, we use the discrete Gaussian distribution with the parameter λ adapted to the channel characteristics.

6.1.2 Ultimate Shaping Gain

The question that arises is, what is the maximum shaping gain? How large is asymptotically the gap between the capacity of a uniform ASK signal constellation and the AWGN capacity in the example of Fig. 6.1? The maximum shaping gain is achieved for signal constellations with a large number of signal points and high SNR where $h(Y) \approx H(X)$. Then, we can approximate the distribution of the signal points by a continuous pdf and compare the differential entropies of the uniform pdf and the optimal Gaussian pdf.

For a uniform pdf with $p(x) = 1/a$ for $-a/2 \leq x \leq a/2$ and 0 otherwise, the average energy is $E_u = a^2/12$. The differential entropy is therefore [CT91]:

$$h(X) = \frac{1}{2} \cdot \log_2(12E_u). \tag{6.6}$$

For a Gaussian pdf with average energy E_g, we have [CT91]

$$h(X) = \frac{1}{2} \cdot \log_2(2\pi e E_g). \tag{6.7}$$

We consider the relation between the average energy for the same differential entropy. Therefore, we set

$$\frac{1}{2} \cdot \log_2(12E_u) = \frac{1}{2} \cdot \log_2(2\pi e E_g) \tag{6.8}$$

and obtain

$$G_\infty = \frac{E_u}{E_g} = \frac{\pi e}{6}, \quad \text{and} \quad 10 \log_{10}\left(\frac{\pi e}{6}\right) \text{dB} = 1.53 \text{ dB}, \tag{6.9}$$

which is the ultimate shaping gain [FGL$^+$84] [FW89]. With signal constellations of practical size, the ultimate shaping gain will never be achieved. The shaping gain is not overwhelming. However, we can achieve these energy savings by simple means and, as already mentioned, it is often easier to obtain a shaping gain than to obtain a similar gain through powerful coding.

6.1.3 Combined Coding and Shaping Techniques

Several techniques have been proposed to obtain a non-uniform probability distribution of the signal points. A promising approach is to partition the signal constellation into several equal-sized sub-constellations corresponding to different average energy levels [CO90]. A shaping algorithm is then used to specify the sequence of sub-constellations so that low-energy signals are transmitted more frequently than high-energy signals. *Shell mapping* ([Fis02] and references therein) is part of the ITU recommendation V.34 [FBEM96] and uses a multi-dimensional mapping to select sequences of low energy signal points. Forney introduced *trellis shaping* in [For92], where the mapping of the bits to the transmitted symbol sequence is done via a search through the trellis (e.g. with a Viterbi algorithm) of a shaping code. For BICM, a system using a shaping block code to select the transmitted sequence has been proposed in [GSJ05].

Another approach to obtain non-equiprobable signal points is to design specific bit-to-symbol mappings. A simple example of such a scheme is given in [FGL$^+$84] and further analyzed in [KP93]. The idea is to assign less bits to more frequent signal points and more bits to less frequent ones. A promising solution is the use of a Huffman prefix code. The main drawback of this method is that it requires a variable rate input, which results in buffering and synchronization problems.

To circumvent this problem, we proposed in [SH05] a practical method for signal shaping based on the ideas of [RG04] and [CV04]: In [RG04], a bit-to-symbol mapping is proposed to obtain binary prefix codes for signal shaping without buffer and synchronization problems. In [CV04], the idea of non-unique mappings for iterative demapping and decoding has been proposed for PSK modulation to reduce the error bound at high SNR.

We assign an arbitrary non-uniform number of bit labels to the signal points and the resulting ambiguities are resolved by the channel code and through iterative demapping and decoding. More bit-labels are assigned to signal points with low energy and less bit-labels to those with high energy.

Note that shaping gains may also be achieved from superimposed signals, as further investigated in Chapter 7.

6.2 Signal Shaping Using Non-Unique Mappings

6.2.1 Bit-Interleaved Coded Modulation with Non-Unique Mappings

We consider the bit-interleaved coded modulation (BICM) system depicted in Fig. 6.2 with a feedback from the channel decoder to the demapper for iterative demapping and decoding (BICM-ID). Extensions and modifications to the standard BICM-ID scheme as introduced in Section 2.5 are mentioned in the following.

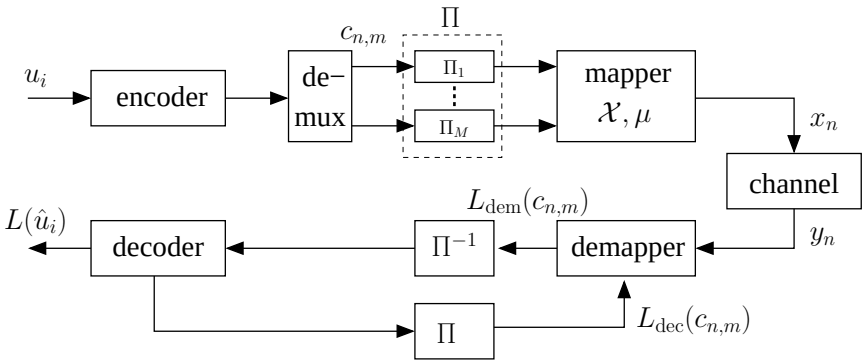


Figure 6.2: BICM-ID system model for signal shaping.

A sequence of information bits is encoded by a binary code and multiplexed to M subsequences. The interleaving is done separately over each subsequence $m = 1, \ldots, M$ with distinct random interleavers Π_m. This interleaver design was proposed in [CR02] to ensure that the coded bits with different protection due to their different positions in the symbol labels are distributed uniformly along the trellis. In the considered non-uniform mapping scheme, this interleaver should be used since the difference in the reliability of the bit positions may be significant.

At transmission time interval n, the bits $\mathbf{c}_n = (c_{n,1}, \ldots, c_{n,M})$ from the M interleaved subsequences are mapped to a complex symbol $x_n = \mu(\mathbf{c}_n)$ chosen from the signal constellation $\mathcal{X}$ according to the binary labeling map $\mu : \{0,1\}^M \rightarrow \mathcal{X}$. The number of possible bit labels 2^M may exceed the number of signal points $|\mathcal{X}|$ and multiple distinct bit sequences may be mapped to the same signal point.

An example for a non-unique mapping of bit sequences of length $M = 6$ to a 16ASK signal constellation is given in Fig. 6.3. Without a priori information, a transmitted bit at position m in the binary label of the ith signal point is deleted or punctured by the non-unique mapping only if $n_i = 2^1, 2^2, 2^3, \ldots$ labels are mapped to the signal point and if the number of ones and zeros at position m in the labels is equal. In the example of Fig. 6.3, the last two bits of the signal point with $n_i = 2$ are punctured by the mapping. With ideal a priori information, the ambiguities can be resolved provided that the sequences that are mapped to the same signal point differ in at least two positions. In Fig. 6.3, all the depicted bit labels differ in two or more positions.

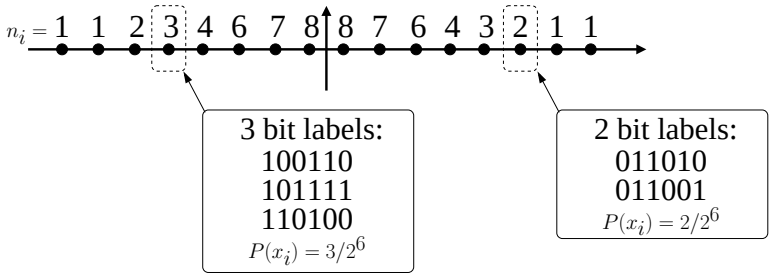


Figure 6.3: Non-unique mapping with 16ASK, 2^6 bit labels of length $M = 6$, given number n_i of bit labels per signal point. Entropy of the channel input signal: $H(X) = 3.69$ bits per symbol.

We consider the transmission over an additive white Gaussian noise (AWGN) channel described in equation (2.8) and an optimum MAP demapper according to equation (2.48). The demapper considers the non-unique mapping as a 2^M-ary signal constellation with superimposed signal points. Note that we use the *full* available a priori information in contrast to the punctured one in [RG04].

In all the following investigations, we consider a 16ASK signal constellation. The investigated methods are applicable to any arbitrary one- or two-dimensional signal constellation. Note that we are no more restricted to signal constellations with a number of signal points equal to a power of two. Fig. 6.4 depicts an example of a 24QAM signal constellation with a non-uniform number of bit labels of length $M = 6$ per signal point. The entropy $H(X) = 4$ corresponds to the entropy of the uniform 16QAM signal constellation.

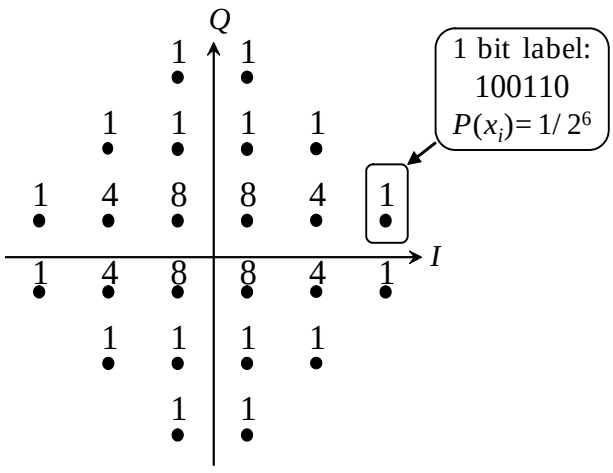
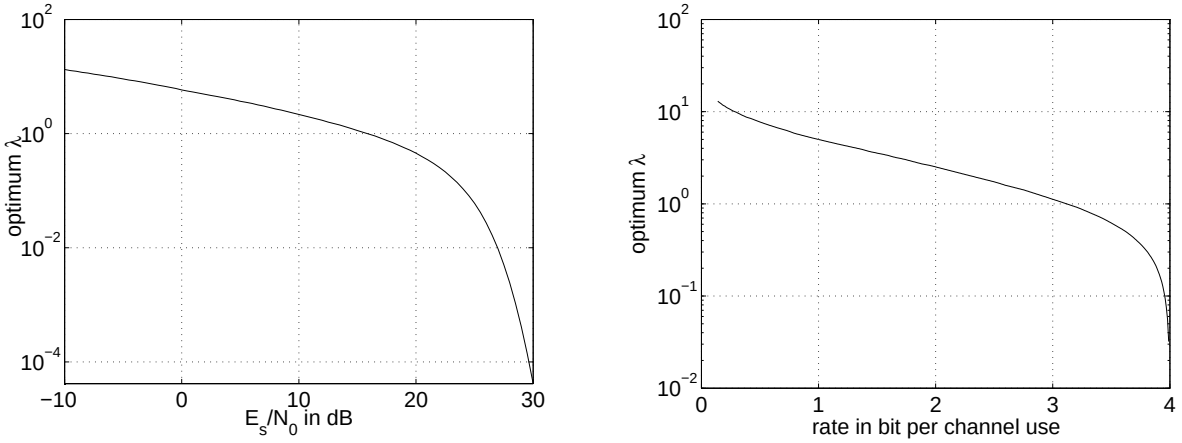


Figure 6.4: Non-unique mapping with 24QAM, 2^6 bit labels of length $M = 6$, given number n_i of bit labels per signal point. Entropy of the channel input signal: $H(X) = 4$ bits per symbol.

6.2.2 Optimization of Symbol Probabilities

To achieve a maximum possible shaping gain, the probabilities of the signal points should closely reflect the probability distribution given in equation (6.4). We first determine the value of λ in equation (6.4) that maximizes the resulting signal set capacity for a given channel quality E_s/N_0 or data rate. The result for a 16ASK signal constellation is depicted in Fig. 6.5.



a) b)

Figure 6.5: Values of λ (equation (6.4)) that maximize the capacity of a 16ASK signal constellation as a function of **a)** E_s/N_0 and of **b)** the data rate in bits per channel use.

To achieve a probability distribution close to the one given for an optimized value of λ, we assign a non-uniform number of distinct bit labels to the signal points and optimize the number of labels per signal point. For this purpose, the Huffman algorithm may be used [KP93]. Note that we have to apply a modified suboptimal Huffman algorithm since we limit the length of the binary label to a maximum value M. To construct a binary Huffman tree, we have to stick to the restriction that the resulting probabilities of the signal points are given by $P(x_i) = 1/2^{l_i}$, with $1 \leq l_i \leq M$, and that at least one label is assigned to each signal point. Then, the number of labels n_i per signal point is restricted to powers of two: $n_i = 2^{M-l_i}$.

To overcome these restrictions, we propose a *label filling* algorithm, where the 2^M labels are allocated to the signal points according to a greedy approach: The algorithm assigns two labels at a time to signal points in a symmetric way to obtain a zero mean constellation. In each assignment, the signal point that results in the smallest Kullback-Leibler distance to the optimum probabilities $P_{\text{opt}}(x_i)$ is chosen, where the Kullback-Leibler distance is given by [CT91] :

$$D(P_{\text{opt}}(x_i)||P(x_i)) = \sum_{i=1}^{|\mathcal{X}|} P_{\text{opt}}(x_i) \cdot \log_2 \frac{P_{\text{opt}}(x_i)}{P(x_i)}. \tag{6.10}$$

A number n_i of labels is assigned to a signal point x_i and $\sum_{i=1}^{|\mathcal{X}|} n_i = 2^M$. Then, $P(x_i) = n_i/2^M$ as illustrated in the examples in Fig. 6.3 and 6.4. Note that the problem of assigning a certain number of bit labels to signal points can be in general formulated as a quadratic assignment problem, similar to the mapping optimization problem discussed in Section 4.3.

The probability distribution of the signal points resulting from the label filling algorithm is shown in Fig. 6.6 for different SNR values. We clearly observe that by reducing the SNR, the signal points close to the origin become more probable and therefore, the entropy $H(X)$ of the signal constellation is reduced. We can interpret this as "soft" adaptive modulation with a soft transition between 16ASK and 2ASK.

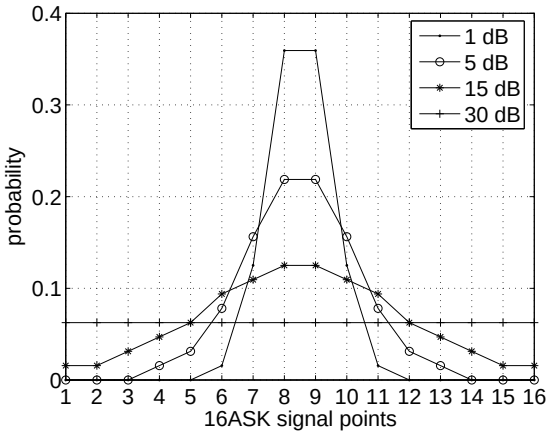


Figure 6.6: Probability distributions of signal points from the label filling algorithm for 16ASK, different values of E_s/N_0 in dB.

Fig. 6.7 shows the Kullback-Leibler distance of some signal point probability distributions to the optimum distribution as a function of the parameter λ and E_s/N_0. 16ASK and a label length $M = 6$ are used. We observe that, as expected, the distributions obtained with the label filling algorithm better approach the optimum distribution than the ones obtained with the Huffman algorithm, especially for high values of λ and low values of E_s/N_0. Asymptotically, the uniform distribution is optimum for very low λ, whereas for a high λ, only the two inner signal points should be used.

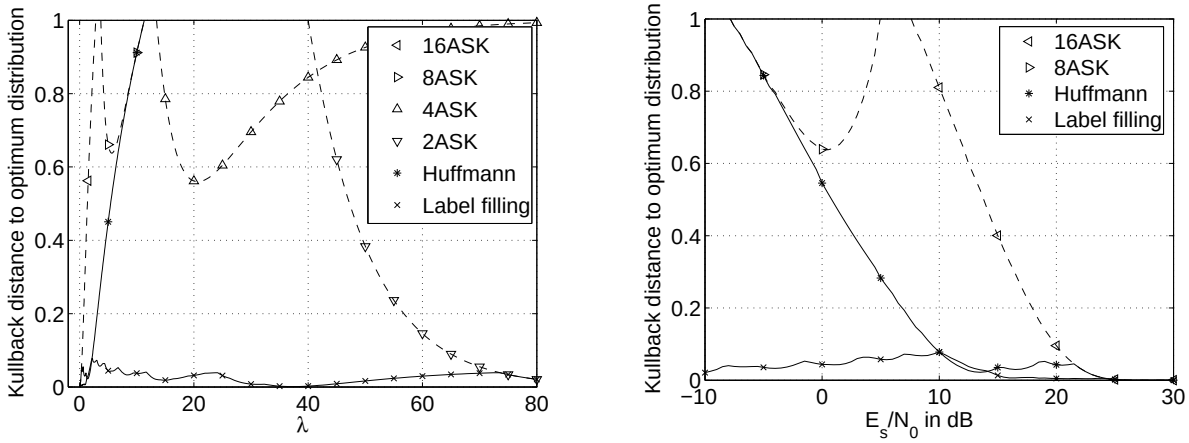
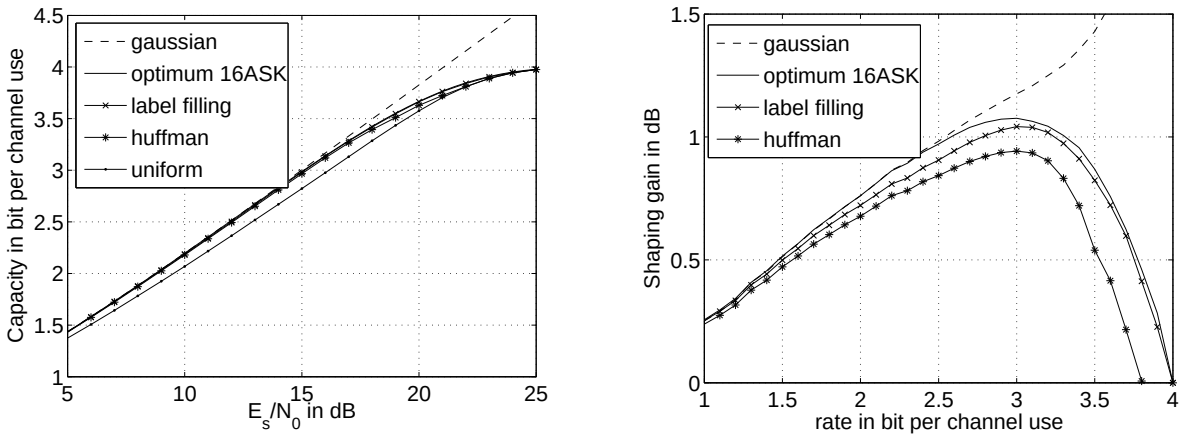


Figure 6.7: Kullback-Leibler distance as defined in equation (6.10) as a function of **a)** λ (equation (6.4)) and **b)** E_s/N_0. Label length $M = 6$, 16ASK signal constellation normalized for a uniform probability distribution. 8ASK, 4ASK and 2ASK use only the inner points of the given 16ASK constellation.

We expect a larger shaping gain with a probability distribution with a small Kullback distance to the optimum distribution. This is confirmed in Fig. 6.8a), where the capacities with 16ASK are depicted for the optimum probability distribution and the probability distributions obtained using the label filling and the Huffman algorithm. We achieve a shaping gain up to 1.08dB, 1.04dB and 0.94dB with respect to the uniform probability distribution if we apply the optimum distribution, the distribution obtained with label filling and the distribution obtained with the Huffman algorithm, respectively (see Fig. 6.8b)).



a) b)

Figure 6.8: For a 16ASK constellation: **a)** Capacity in bits per channel use for different signal point probability distributions. **b)** Corresponding shaping gain.

Fig. 6.9 illustrates the distribution of the overall redundancy $M - \mathcal{C}$ into coding redundancy $r_c = H(X) - \mathcal{C}$ and shaping redundancy $r_s = M - H(X)$, where $\mathcal{C}$ denotes the capacity as given in Fig. 6.8a). The results show how much loss r_s of data rate we should accept because of the non-uniform probability distribution to optimize the capacity while reducing the amount of redundancy r_c from channel coding [WFH99]. For the considered 16ASK constellation with the optimized probability distribution, it is quite interesting to observe that the coding redundancy should be only slightly higher than the shaping redundancy.

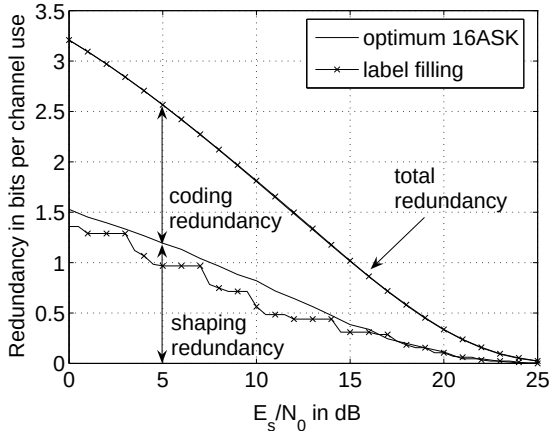


Figure 6.9: Distribution of shaping and coding redundancy.

6.2.3 Optimization of Mapping

For capacity-optimized probabilities of the signal points, we have to determine the bit-to-symbol mapping. In contrast to the signal point probabilities, the mapping does not change the capacity if iterative demapping and decoding is applied at the receiver. However, the mapping has a considerable influence on the system performance.

We use the methods introduced in Chapter 4 to characterize and optimize non-unique mappings: We generate an Euclidean distance spectrum (EDS, see Section 4.1) for no and ideal a priori information to characterize the mapping and use the cost functions of equation (4.9) and the binary switching algorithm to optimize the mapping. With non-unique mappings, Euclidean distances $d_{\text{ex},1} = 0$ between decision regions appear in the EDS due to superimposed labels.

Consider as example a 16ASK signal constellation and a label length of $M = 6$. For a rate of 3 bits per channel use, the capacity limit is at $10\log_{10}(E_s/N_0)\text{dB} = 15\text{dB}$. The optimum value of λ is then $\lambda = 1.134$. With the label filling algorithm, we assign $\{n_i\} = \{1, 1, 2, 3, 4, 6, 7, 8, 8, 7, 6, 4, 3, 2, 1, 1\}$ labels to the 16 signal points to optimize the constellation-constrained capacity. The resulting entropy is $H(X) = 3.69$ bits per symbol and the shaping redundancy $M - H(X) = 0.31$ bits.

For this setup, we optimize the mapping with the binary switching algorithm for no and ideal a priori information at the demapper. The resulting Gray-like and turbo optimized mapping are defined in Table A.5 in Appendix A and the Euclidean distance spectra are given in Table 6.1.

D_{ex}	$d_{\text{ex},1} = 0$	$d_{\text{ex},2} = \Delta$	$d_{\text{ex},3} = 2\Delta$	3Δ	4Δ	5Δ	6Δ	7Δ	...
Λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	...
No a priori information (Expurgated)									
Gray	488	716	350	124	56	32	16	16	...
Turbo opt.	832	139	11	0	0	0	0	0	...
Ideal a priori information									
Gray	70	56	8	16	6	6	14	8	...
Turbo opt.	0	0	0	43	58	44	26	11	...

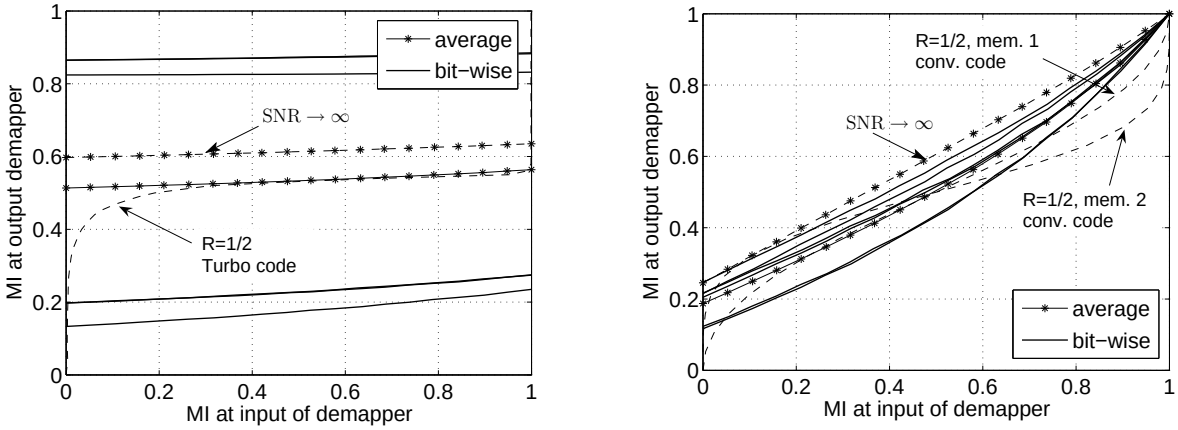
Table 6.1: Euclidean distance spectrum (EDS) for 16ASK.

The corresponding demapper EXIT functions are shown in Fig. 6.10. The area under the demapper EXIT function is upper bounded at high SNR by $\mathcal{A}_{\text{dem}} \leq H(X)/M = 3.69/6 = 0.615$ in the considered example. The EXIT functions for an asymptotically large SNR are included in Fig. 6.10. Recall that the area under the EXIT function of the outer decoder is equal to $\mathcal{A}_{\text{dec}} = R_{\text{out}}$ and that $\mathcal{A}_{\text{dec}} < \mathcal{A}_{\text{dem}}$ to have a chance of successful decoding. Therefore, a code rate smaller than 0.615 is always required in this example to resolve the ambiguities that arise when multiple bit labels are assigned to one signal point.

6.2.4 Optimization of Iterative Receiver

The EXIT functions of the demapper and the channel decoder should be well matched to achieve close to capacity performance, as discussed in Section 3.4.

For the example investigated in Section 6.2.3, we consider first the combination of a parallel concatenated turbo code with the Gray-like mapping. We observe in Fig. 6.10a) that this setup results in a good match of the EXIT functions.



a) Gray mapping b) Turbo opt. mapping

Figure 6.10: Average and bit-wise EXIT chart of 16ASK with different mappings, $10\log_{10}(E_s/N_0)\text{dB} = 16.5\text{dB}$ and $E_s/N_0 \to \infty$.

Then, we use the turbo optimized mapping with a convolutional code and shape the decoder EXIT function using the concept of irregular convolutional codes [TH02]. A code word of an irregular code of length N and overall rate R_{IR} consists of P subblocks of length $\alpha_p N$, $p = 1, \ldots, P$. These subblocks may have different code rates R_p obtained by applying different puncturing patters in each subblock. We deliberately reduce the performance of the convolutional code by irregular puncturing to optimize it for the iterative scheme. The EXIT function $f_{\text{dec}}(I)$ of the overall irregular code is the linear combination of the EXIT functions $f_{\text{dec},p}(I)$ of the P subcodes, similar to the EXIT functions of irregular modulation introduced in Chapter 5:

$$f_{\text{dec}}(I) = \sum_{p=1}^{P} \alpha_p \cdot f_{\text{dec},p}(I). \tag{6.11}$$

The optimization of the ratios α_p can be formulated as a linear programming problem. Similar to equation (5.5) for irregular modulation, the optimization problem can be written as:

$$\begin{aligned} \text{maximize} \quad & R_{IR} = \textstyle\sum_{p=1}^{P} \alpha_p \cdot R_p, \\ \text{subject to} \quad & \textstyle\sum_{p=1}^{P} \alpha_p \cdot f_{\text{dec},p}^{-1}(I) < f_{\text{dem}}(I), \\ & \textstyle\sum_{p=1}^{P} \alpha_p = 1, \quad \alpha_p \in [0,1], \forall p. \end{aligned} \tag{6.12}$$

For the considered example, we use a memory 3, rate $R = 1/4$ mother code with generator polynomial $(1, 15/13, 15/13, 17/13)$ in octal notation. The possible rates of the $P = 14$ subblocks obtained using the puncturing pattern given in [Tüc04] are $R_p = \{0.25, 0.3, 0.35, \ldots, 0.85, 0.9\}$. Fig. 6.11a) and 6.11b) show the EXIT functions of the subcodes and of the two irregular codes optimized for an early convergence threshold for the reference system without signal shaping and for the proposed system with signal shaping, respectively. In both systems, we set the data rate to 3 bits per channel use.

In the reference system, we use a 16ASK signal constellation with label length $M = 4$ and the optimized mapping given in Table A.5. The irregular code is punctured to rate $R = 3/4$ with the optimized ratios $\{\alpha_p\} = \{0, 0, 0, 0, 0, 0.005, 0.004, 0, 0.43, 0.14, 0.03, 0.033, 0.041, 0.317\}$. The decoder EXIT function is given in Fig. 6.11a).

For the proposed system with signal shaping, we use a 16ASK constellation with a label length $M = 6$, the optimized probability distribution and the non-unique turbo optimized mapping defined in Table A.5. The irregular code is punctured to rate $R = 1/2$ with the optimized ratios $\{\alpha_p\} = \{0.2, 0.18, 0.01, 0.03, 0.034, 0.118, 0.11, 0.022, 0.013, 0.116, 0.025, 0.026, 0.028, 0.088\}$. The resulting decoder EXIT function is depicted in Fig. 6.11b).

The iterative decoding threshold predicted by the EXIT chart with the optimized irregular codes is at $10\log_{10}(E_s/N_0)\text{dB} = 16.6\text{dB}$ and $10\log_{10}(E_s/N_0)\text{dB} = 15.5\text{dB}$ for the considered systems without and with signal shaping, respectively. Using the values of Table 6.2, the EXIT threshold with signal shaping is therefore 0.5dB above the capacity of the continuous Gaussian input and 0.7dB below the capacity of the uniform distribution.

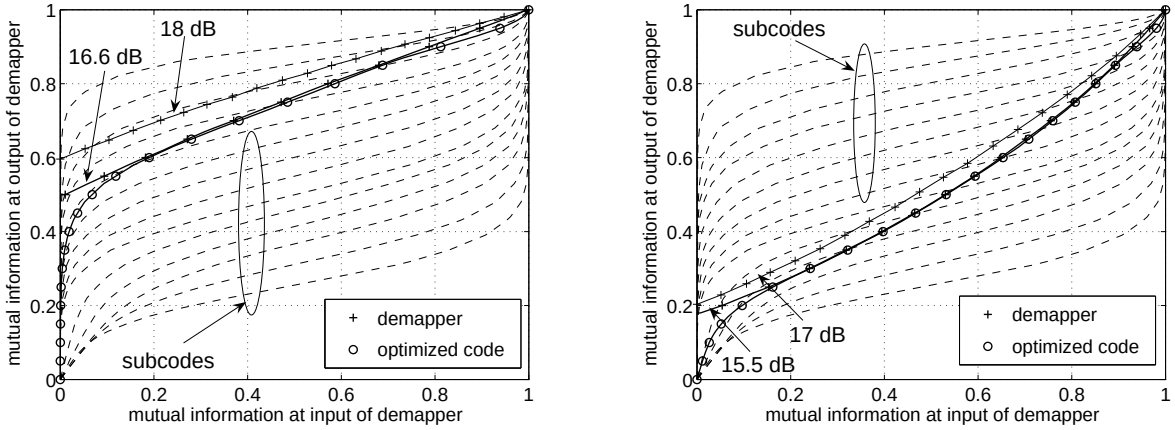


Figure 6.11: EXIT functions of optimized mappings with different E_s/N_0 values in dB and of optimized irregular codes for **a)** the reference system without signal shaping; **b)** the proposed system with signal shaping. The data transmission rate is $R_t = R \cdot M = 3$ bits per channel use.

6.3 Simulation Results

Fig. 6.12 depicts the bit error rate (BER) performance of the optimized BICM-ID systems without and with signal shaping. The channel is AWGN and the interleaver length is 100000 bits. Table 6.2 summarizes the results.

	Capacity	EXIT threshold	BER at 10^{-5}
Gaussian input	15dB	–	–
Non-uniform distribution	15.1dB	15.5dB	16.2dB
Uniform distribution	16.2dB	16.6dB	16.9dB

Table 6.2: Characteristic E_s/N_0 values in dB (rounded) for 16ASK, data rate of $R_t = 3$ bits per channel use.

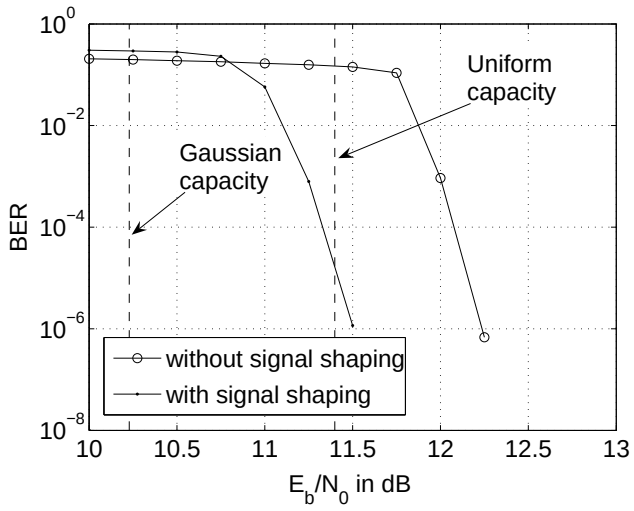


Figure 6.12: BER with AWGN channel, data rate of $R \cdot M = 3$ bits per channel use, 16ASK, memory 3 irregular convolutional code. Optimized BICM-ID systems as derived in Section 6.2.4; $E_b/N_0 = E_s/N_0 \ / \ (R \cdot M)$; 50 iterations.

6.4 Summary

A shaping gain is obtained by assigning a non-uniform number of distinct bit labels to the signal points in a BICM scheme with iterative demapping and decoding (BICM-ID). We apply a greedy label filling method to optimize the probability distribution of the signal points and a binary switching algorithm to optimize the bit-to-symbol mapping. EXIT charts are used to optimize an irregular convolutional code and to achieve a performance close to Shannon's AWGN capacity. The results for the example of 16ASK and a data transmission rate of 3 bits per channel use are summarized in Table 6.2. We achieve approximately a BER of 10^{-5} at the capacity of the uniform distribution and an iterative decoding threshold 0.7dB below the capacity of the uniform distribution.

7

Coded Modulation with Mapping by Superposition

We consider signal constellations and bit-to-symbol mappings that arise from the linear superposition of several signal layers. A *mapping by superposition* is defined that is applicable to both multi-level coding (MLC) and bit-interleaved coded modulation (BICM). The approach is based on the idea of superposition coding [CT91] and in particular on the system proposed in [DRU97] and further developed in [Cro05] [MP04a] [TP06]. There, the superposition of independent signals yields a Gaussian-like channel input, as required to approach Shannon's capacity.

Furthermore, mapping by superposition is closely related to general multiple-access (MA) schemes and is in particular very similar to trellis coded multiple access (TCMA) [BAR02] and interleave-division multiple-access (IDMA) [PLWL06]. However, we focus on the single user system where the data of the superimposed layers all belong to one user. Then, the cooperation among the layers at the transmitter is perfect. By adjusting the power and phase allocation to the layers, we are able to design a large variety of overall signal constellations.

Advantages of mapping by superposition include that data from different sources can be easily combined for bandwidth efficient transmission, that for some power allocations, the transmitted signal is approximately Gaussian, as required to approach the AWGN capacity, and that low complexity detectors like the investigated soft interference canceler may be used. As main drawback we should mention the design constraints on the overall bit-to-symbol mapping.

In Section 7.2, we construct different overall signal constellations by the superposition of signals from smaller signal constellations. We investigate their constellation-constrained capacity and their peak-to-average power ratios (PAPR), similar to the investigations performed in [BAR01] for TCMA. By setting the powers and phases of the layers, a large variety of overall signal constellations is obtained, including standard QAM constellations and constellations with non-unique signal points.

To analyze the iterative receiver, we derive in Section 7.3 analytic EXIT functions of a low complexity interference canceler (IC) for the MLC and BICM systems. For the MLC scheme,

we break down the multi-dimensional EXIT analysis to a one-dimensional problem by scaling the output of the detector and decoder in an appropriate way. The approach used in [SS03] and [SSRS06] for a large number of users or layers is extended to be applicable for the considered system with a low number of superimposed layers. Different power allocations and bit-to-symbol mappings of the layers are considered. The analytic IC EXIT functions are compared to the simulated IC and MAP detector characteristics.

In Section 7.4, the mapping by superposition is optimized by adjusting the power allocation of the layers. For successive interference cancelation, the power allocation for multi-user systems has been optimized in [Mec03] using mutual information measures and assuming a capacity approaching code. For iterative detection, the power allocation has been optimized in [CMT04] using multiuser efficiency as performance measure and in [MP04b] using error rate simulations. We use the results of the EXIT chart analysis in Section 7.3 to solve the optimization of the power allocation to the layers for iterative detection using constrained nonlinear optimization techniques [SSRS06]. The approach is similar to the optimization of irregular modulation in Chapter 5 and irregular codes in Chapter 6: Instead of assigning different signal constellations, mappings or different code rates within one code word, we assign here different powers to shape the EXIT functions to obtain a good match of the detector and decoder EXIT functions.

Section 7.5 finally shows simulation results and analytical error bounds of the investigated systems.

7.1 System Structure with Mapping by Superposition

The MLC scheme with mapping by superposition is depicted in Fig. 7.1. A sequence of information bits is de-multiplexed to K data layers. Each layer is separately encoded by a binary code and interleaved by a different random interleaver Π_k; M consecutive coded and interleaved bits of each layer $k = 1, \ldots, K$ are grouped to $\mathbf{c}_k = (c_{k,1}, \ldots, c_{k,m}, \ldots, c_{k,M})$.

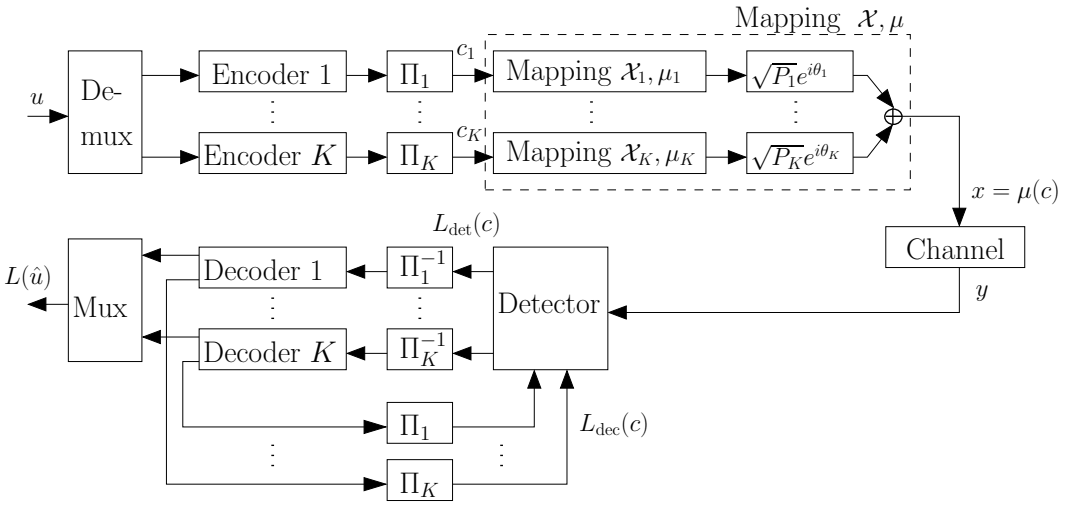


Figure 7.1: Multi-level coding (MLC) with mapping by superposition.

The BICM approach is shown in Fig. 7.2. A single encoder and random interleaver Π is used; M consecutive coded and interleaved bits are grouped to $\mathbf{c}_k = (c_{k,1}, \ldots, c_{k,m}, \ldots, c_{k,M})$; K consecutive subsequences $\mathbf{c}_k$, $k = 1, \ldots, K$, are de-multiplexed to K layers.

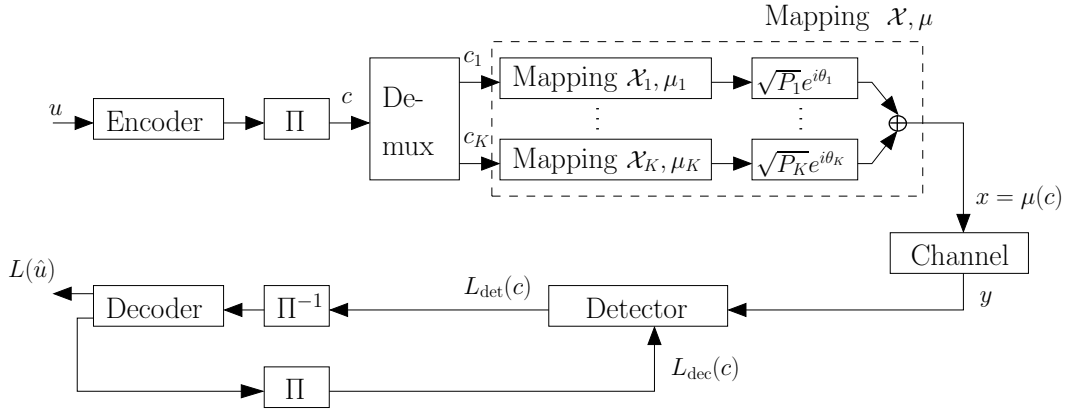


Figure 7.2: Bit-interleaved coded modulation (BICM) with mapping by superposition.

The mapping by superposition for *both* MLC and BICM is done as follows: The subsequences $\mathbf{c}_k$ of length M from the layers $k = 1, \ldots, K$ are mapped to the complex symbols $x_k = \mu_k(\mathbf{c}_k)$ chosen from the 2^M-ary signal constellation $\mathcal{X}_k$ (e.g. QPSK for $M = 2$) according to the bit-to-symbol mapping $\mu_k : \{0, 1\}^M \to \mathcal{X}_k$ that defines the assignment of the binary bit sequences $\mathbf{c}_k$ to the signal points (e.g. Gray mapping). The transmitted signal x is the linear superposition of the signals x_k of the K layers:

$$x = \sum_{k=1}^{K} \sqrt{P_k}\, \mathrm{e}^{i\theta_k} x_k, \tag{7.1}$$

with a power P_k and phase θ_k associated to each layer. We omit the time index for x for ease of presentation.

In other words, MK coded and interleaved bits from the K subsequences $\mathbf{c}_k$ are grouped to the sequence $\mathbf{c} = (\mathbf{c}_1, \ldots, \mathbf{c}_k, \ldots, \mathbf{c}_K)$ and mapped to a complex symbol $x = \mu(\mathbf{c})$ chosen from the overall 2^{MK}-ary signal constellation $\mathcal{X}$ according to the overall mapping $\mu : \{0, 1\}^{MK} \to \mathcal{X}$. The signal constellation $\mathcal{X}$ and mapping μ depend on the signal constellation $\mathcal{X}_k$ and mapping μ_k of the underlying layers and on the power P_k and phase θ_k used for the linear superposition.

In the following investigations, we set the signal constellation $\mathcal{X}_k$ to QPSK with $M = 2$, and the mapping μ_k to Gray or Anti-Gray. The best performance in the first receiver iteration is achieved with Gray mapping but the performance gain with the iterations is maximized with Anti-Gray mapping, as described in Chapter 4. We use the same signal constellation and mapping for each layer. The combination of different signal constellations and mappings would further enhance the design and optimization possibilities, similar as with irregular or hybrid modulation investigated in Chapter 5.

The superimposed signal x is transmitted over the basic AWGN channel $y = x + n$ given in equation (2.8).

At the receiver, the detector uses the received symbols y and the extrinsic a priori LLRs $L_{\text{dec}}(c_{k,m})$ fed back from the decoder to compute the extrinsic LLRs $L_{\text{det}}(c_{k,m})$, for $k = 1, \ldots, K$ and $m = 1, \ldots, M$. The a priori knowledge reduces the number of interfering layers as well as the number of possible received signal points in the underlying signal constellation $\mathcal{X}_k$ of the layer k to be detected. The extrinsic LLRs of the detector and decoder are iteratively

exchanged, similar to the decoding of a serial concatenated code shown in Fig. 2.14. We assume that the receiver has perfect channel knowledge and knows through signalling the parameters used at the transmitter, namely the applied signal constellation, mapping, number of layers, power and phase allocation to the layers.

We consider two detectors: the optimum maximum a-posteriori (MAP) detector and a low complexity interference canceler (IC). Note that other multi-user detectors not considered here may also be a good choice, e.g. the list-sequential detector described in [HK06].

The optimum detector corresponds to the MAP demapper of the overall signal constellation $\mathcal{X}$. The MAP demapper is described in Section (2.3), where a metric $\Lambda(\mathbf{c})$ is defined to compute the LLRs according to equation (2.43). The metric for the extrinsic LLRs $L_{\text{det}}(c_{k,m})$ is

$$\Lambda(\mathbf{c}) = -\frac{|y-\mu(\mathbf{c})|^2}{2\sigma_n^2} - \sum_{l=1}^{K}\sum_{j=1}^{M} c_{l,j} \cdot L_{\text{dec}}(c_{l,j}) + c_{k,m} \cdot L_{\text{dec}}(c_{k,m}). \tag{7.2}$$

The considered low complexity IC detector is shown in Fig. 7.3. It is composed of two functional units which are soft interference cancelation and QPSK soft demapping to compute the extrinsic LLRs of the coded bits.

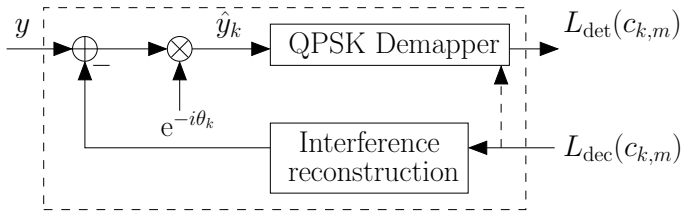


Figure 7.3: Interference canceler (IC).

In the first stage of the IC, a soft estimate of the interference is subtracted from the received signal y, that is often referred to as parallel interference cancelation (PIC).

The general definition of a soft-symbol $\bar{x}_k$ with a priori knowledge is given in equation (2.84). In the special case of QPSK with Gray mapping, the signal constellation is an orthogonal superposition of two independent BPSK signals and we can compute $\bar{x}_k$ using the definition of soft-bits in equation (2.33):

$$\bar{x}_k = \tanh(L_{\text{dec}}(c_{k,1})/2) + i \cdot \tanh(L_{\text{dec}}(c_{k,2})/2). \tag{7.3}$$

The interference-reduced and rotated signal estimate $\hat{y}_k$ is

$$\hat{y}_k = \mathrm{e}^{-i\theta_k} \cdot (y - \sum_{l \neq k} \sqrt{P_l}\mathrm{e}^{i\theta_l}\bar{x}_l). \tag{7.4}$$

The QPSK demapper uses the log-MAP algorithm described in Section 2.3 to compute the extrinsic LLRs out of $\hat{y}_k$ and the a priori information fed back from the decoder; $\hat{y}_k$ is assumed to be Gaussian distributed with mean $\sqrt{P_k}$ and variance $\text{Var}(\hat{y}_k)$. For a QPSK signal with $|x_k| = 1, \forall k$, the variance of $\bar{x}_k$ is

$$\sigma^2_{\text{dec},k} = \mathrm{E}\{|x_k - \bar{x}_k|^2 | L_{\text{dec}}(\mathbf{c}_k)\} = 1 - |\bar{x}_k|^2. \tag{7.5}$$

Then, if the variance of the real and imaginary part of $\bar{x}_k$ are identical, or if $\theta_k = 0$, $\forall k$, the variance of $\hat{y}_k$ is

$$\mathrm{Var}(\hat{y}_k) = \sum_{l \neq k} P_l \sigma^2_{\mathrm{dec},l} + \sigma_n^2. \tag{7.6}$$

Otherwise, we proceed as described in [TP06].

In the special case of QPSK and Gray mapping, no a priori information needs to be passed to the demapper after the interference cancelation stage. Then, if the bit $c_{k,1}$ corresponds to the decision on the real part of the QPSK signal constellation and with the Gaussian assumption, the QPSK demapper output is similar to equation (3.1) and given by:

$$L_{\mathrm{det}}(c_{k,1}) = \frac{2\sqrt{P_k}}{\mathrm{Var}(\mathrm{Re}(\hat{y}_k))} \cdot \mathrm{Re}(\hat{y}). \tag{7.7}$$

The channel decoder uses the BCJR algorithm to compute extrinsic estimates about the coded bits that are fed back and regarded as a priori information at the detector.

To ensure the stability of the iterative receiver, we apply a mask on the coded and interleaved bits and flip every second bit at the transmitter and reverse this operation at the receiver by flipping the sign of every second LLR [PLWL06].

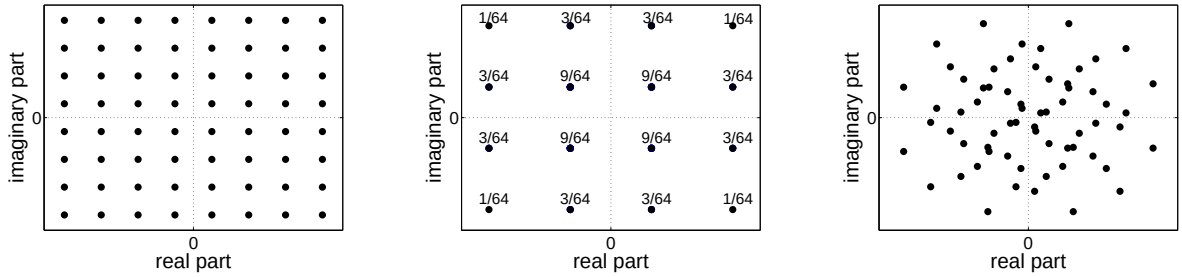
7.2 Constellation-Constrained Capacity

We investigate the properties of the overall signal constellation $\mathcal{X}$ resulting from the superposition of smaller signal constellations $\mathcal{X}_k$ of the layers $k = 1, \ldots, K$, as described in the previous Section 7.1.

Fig. 7.4 depicts for three different power and phase allocations the overall signal constellations $\mathcal{X}$ for $K = 3$ superimposed layers. With a uniform phase distribution and the power allocation $\mathbf{P}^{(1)}$ given in Fig. 7.4, we obtain a standard 64QAM signal constellation. With a uniform phase and uniform power distribution $\mathbf{P}^{(2)}$, the superposition of signals leads to non-unique signal points. The distinct signal points with low energy are more probable than those with high energy as required to achieve a shaping gain. Finally, we expect a shaping gain with the signal constellation obtained with the power allocation $\mathbf{P}^{(3)}$ due to the non-uniformly spaced signal points.

The three signal constellations of Fig. 7.4 are examples of the large variety of signal constellations we can design by setting the power and phase of the layers. The standard QAM signal constellations, the signal shaping approach with non-uniform probability distribution of the signal points [RG04][SH05] and the signal shaping approach with non-uniformly spaced signal points [SvT93] are only special cases of mapping by superposition. Note however that for a given signal constellation constructed by superposition, only a small number of different overall mappings are possible. As example, it is not possible to design an overall Gray mapping if we use the power allocation $\mathbf{P}^{(1)}$ resulting in a 64QAM signal constellation.

Different power allocations imply a different peak-to-average power ratio (PAPR) defined in equation (6.3). The PAPR of the investigated examples is PAPR$= 2.33$, 3.00, and 2.46 for the signal constellations constructed with the power allocations $\mathbf{P}^{(1)}$, $\mathbf{P}^{(2)}$, and $\mathbf{P}^{(3)}$, respectively. For more layers, the PAPR will increase. The issue of reducing the PAPR by clipping for a similar system has been investigated in [TP06].



a) $\mathbf{P}^{(1)} = [1/21, 4/21, 16/21]$
$\theta_k = 0, \forall k$

b) $\mathbf{P}^{(2)} = [1/3, 1/3, 1/3]$
$\theta_k = 0, \forall k$

c) $\mathbf{P}^{(3)} = [0.2, 0.3, 0.5]$
$\theta_k = k\pi/(2K)$

Figure 7.4: Signal constellations for 3 layers, QPSK modulation, different normalized power allocations $\mathbf{P} = [P_1, \ldots, P_K]$ and phase distributions θ_k. **a)** Non-uniform power allocation resulting in a 64QAM signal constellation with $\sqrt{P_k} = 2 \cdot \sqrt{P_{k-1}}$ for $k = 2, \ldots, K$; every signal point is transmitted with equal probability. **b)** Uniform power allocation; each signal point is labeled with the probability of being transmitted. **c)** Non-uniform power allocation; every signal point is transmitted with equal probability.

Fig. 7.5 depicts the constellation-constrained capacity $\mathcal{C}_{cc}$ in bits per channel use of the signal constellations investigated in Fig. 7.4. A gap to the AWGN capacity remains with the standard 64QAM signal constellation at medium and high SNR. The other two curves approach the AWGN channel capacity at medium SNR. The capacity of the signal constellation resulting from the uniform power allocation $\mathbf{P}^{(2)}$ is limited by the number of distinct signal points. The maximum achievable data rate is equal to the entropy $H(X)$ of the transmitted signal:

$$H(X) = -\sum_i P(x_i) \log_2(P(x_i)), \tag{7.8}$$

where $P(x_i)$ is the probability of the ith *distinct* signal point. For the superposition of 3 layers with the uniform power allocation $\mathbf{P}^{(2)}$, we have $H(X) = 3.66$. With the power allocations $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(3)}$, the distinct signal points are equiprobable and $H(X) = 4$ (cf. Fig. 7.5).

7.3 EXIT Chart Analysis

We use the EXIT charts investigated in Chapter 3 to analyze and optimize the considered BICM and MLC systems. We investigate the power allocation to the layers, compare QPSK Gray and Anti-Gray mapping, and the MAP and IC detector.

Consider the average mutual information (MI) defined in equation (3.4) between the coded bits at the transmitter and the extrinsic LLRs at the receiver. Let I_{det} denote the average extrinsic MI at the output of the detector and the a priori MI at the input of the decoder; I_{dec} is the average extrinsic MI at the output of the decoder and the a priori MI at the input of the detector.

The detector is characterized by the EXIT function $I_{\text{det}} = f_{\text{det}}(I_{\text{dec}})$ for a given SNR, power P_k and phase θ_k of the layers $k = 1, \ldots, K$. Monte-Carlo simulations or numerical integration are used to obtain the EXIT function of the MAP detector, see Section 3.3. For the low complexity IC detector, we derive in this Section the analytic EXIT functions for the BICM and MLC schemes. The channel decoder is characterized by the EXIT function $I_{\text{dec}} = f_{\text{dec}}(I_{\text{det}})$, see Section 3.2.

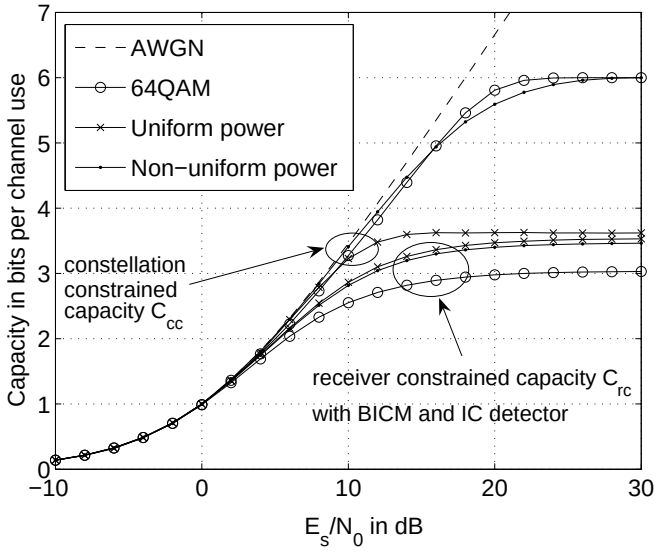


Figure 7.5: Constellation-constrained capacity $\mathcal{C}_{cc}$ and receiver-constrained capacity $\mathcal{C}_{rc}$ with BICM and IC detector obtained from the EXIT function as described in Section 7.3; 3 QPSK layers; power allocations $\mathbf{P}^{(1)}$ (64QAM), $\mathbf{P}^{(2)}$ (uniform power), and $\mathbf{P}^{(3)}$ (non-uniform power) given in Fig. 7.4.

For the case of non-uniform power allocation, we have to differentiate between the EXIT chart analysis of the BICM and MLC schemes. With BICM, we consider the MI averaged over one code word, i.e. we average over all layers since a single interleaver and channel code are used. The analysis is a one-dimensional problem. With MLC, the analysis becomes a K-dimensional problem since the output of the detector depends on the a priori MI of every single layer and not just on the average MI.

To translate between the variance of the LLRs and the MI after the detector, we use the J function defined in equation (3.14). To translate between the variance of the soft-bits or soft-symbols and the MI after the decoder, we use the T function of equation (3.17). Both functions assume Gaussian distributed values.

Bit-interleaved coded modulation scheme

We investigate the EXIT function of the IC detector in the BICM scheme. From the a priori MI I_{dec}, we compute the variance σ^2_{dec} of the soft-bits used for interference cancelation:

$$\sigma^2_{\text{dec}} = 1 - T^{-1}(I_{\text{dec}}). \tag{7.9}$$

For QPSK with Gray mapping and with equation (7.6) and (7.7), the variance $\sigma^2_{\text{det},k}$ of the LLRs of the kth layer after the QPSK demapper is given by:

$$\sigma^2_{\text{det},k} = \frac{4P_k}{\sum_{l\neq k} P_l \sigma^2_{\text{dec}} + \sigma^2_n}. \tag{7.10}$$

The MI $I_{\text{det},k}$ of the kth layer is then

$$I_{\text{det},k} = J\left(\sigma_{\text{det},k}\right). \tag{7.11}$$

With QPSK and Anti-Gray mapping, we use EXIT chart characteristics to derive the MI $I_{\text{det},k}$. We know that the QPSK signal constellation is reduced to two signal points with ideal a priori information. If $|x_k| = 1, \forall k$, the Euclidean distance between the two remaining signal points is either $\sqrt{2}$ or 2 with equal probability, as illustrated in Fig. 4.6. Therefore, the detector EXIT function with QPSK and Anti-Gray mapping at $I_{\text{dec}} = 1$ is

$$I_{\text{det},k}\big|_{I_{\text{dec}}=1} = \frac{1}{2} J\left(\sigma_{\text{det},k}\right) + \frac{1}{2} J\left(\sqrt{2}\sigma_{\text{det},k}\right). \tag{7.12}$$

Furthermore, from the derivations performed in Section 3.3 and in particular from Fig. 4.8, we know that the EXIT functions of QPSK *without superposition* are straight lines and that the functions with Gray and Anti-Gray mapping intersect at an a priori information of $I_{\text{dec}} = 0.5$. Using basic geometry, equation (7.11) for Gray mapping and (7.12) for Anti-Gray mapping, the detector EXIT function for the kth layer for *superimposed* QPSK and Anti-Gray mapping is given by:

$$I_{\text{det},k} = \left[J\left(\sqrt{2}\sigma_{\text{det},k}\right) - J\left(\sigma_{\text{det},k}\right)\right] I_{\text{dec}} + 1.5 J\left(\sigma_{\text{det},k}\right) - 0.5 J\left(\sqrt{2}\sigma_{\text{det},k}\right). \tag{7.13}$$

For both Gray and Anti-Gray mapping, we average over the K layers to obtain:

$$I_{\text{det}} = \frac{1}{K} \sum_{k=1}^{K} I_{\text{det},k}. \tag{7.14}$$

Three characteristic values of the EXIT function of the detector are of interest (see also Chapter 3): First, a large value of I_{det} without a priori information ($I_{\text{dec}} = 0$) is desired to open the tunnel between the EXIT functions and to ensure the start of the iterative process at low SNR. Second, the value of I_{det} with ideal a priori information (genie or error free feedback case, $I_{\text{dec}} = 1$) determines the gain over the iterations, i.e. the performance at high SNR after several iterations. Finally, the area $\mathcal{A}_{\text{det}}$ under the detector EXIT function $f_{\text{det}}(I_{\text{dec}})$ is of interest.

We have defined in Section 2.2 the receiver-constrained capacity $\mathcal{C}_{rc}$ as the maximum achievable rate with the (not necessarily optimum) detector. Compared to the constellation-constrained capacity $\mathcal{C}_{cc}$, we have $\mathcal{C}_{rc} \leq \mathcal{C}_{cc}$ with equality for the MAP detector. Then, the following relation results from (3.28):

$$\mathcal{A}_{\text{det}} = \mathcal{C}_{rc}/(MK), \tag{7.15}$$

where MK is the number of bits associated to one symbol of the overall signal constellation.

The values I_{det} for no and ideal a priori information are given directly from the analytic EXIT function of the IC detector. By integrating over this EXIT function, we obtain the area $\mathcal{A}_{\text{det}}$. We are interested in optimizing the overall signal constellation $\mathcal{X}$ and mapping μ with respect to the three considered values by setting the powers P_k. This optimization problem can be cast to a constrained nonlinear optimization problem:

$$\begin{aligned} &\text{maximize} && I_{\text{det}}\big|_{I_{\text{dec}}=0} \text{ or } I_{\text{det}}\big|_{I_{\text{dec}}=1} \text{ or } \mathcal{A}_{\text{det}}, \\ &\text{subject to} && \textstyle\sum_{k=1}^{K} P_k = 1;\ P_{\min} \leq P_k \leq P_{\max}, \forall k. \end{aligned} \tag{7.16}$$

For the considered IC detector, the best performance without a priori information is achieved with a single dominant layer with high power, whereas the best performance with ideal a priori

information is reached with an equal power distribution. Note that without a priori information ($I_{\text{dec}} = 0$ and $\sigma^2_{\text{dec}} = 1$) and for high SNR, the value $\sigma^2_{\text{det},k}$ in equation (7.10) is bounded by:

$$\sigma^2_{\text{det},k}\big|_{I_{\text{dec}}=0,\sigma_n^2=0} = \frac{4P_k}{\sum_{l\neq k} P_l}. \tag{7.17}$$

Therefore, the value I_{det} at $I_{\text{dec}} = 0$ is interference limited and will not reach the value of 1 at high SNR, both with Gray mapping in equation (7.11) and Anti-Gray mapping in equation (7.13). With a MAP detector however, I_{det} will reach the value of 1 at $I_{\text{dec}} = 0$ and high SNR if the signal points in the signal constellation $\mathcal{X}$ are distinct, i.e. with a non-uniform power or phase distribution.

If we consider the IC detector again, the area $\mathcal{A}_{\text{det}}$ related to the receiver-constrained capacity $\mathcal{C}_{rc}$ is maximized with a uniform power distribution and minimized with a single dominant layer with high power. $\mathcal{C}_{rc}$ is compared to the constellation-constrained capacity $\mathcal{C}_{cc}$ with 3 layers in Fig. 7.5. Recall that $\mathcal{C}_{cc}$ corresponds to the achievable rates with the MAP detector. The difference between $\mathcal{C}_{cc}$ and $\mathcal{C}_{rc}$ is large especially for a non-uniform power distribution at high SNR, where $\mathcal{C}_{cc}$ reaches the maximum data rate of MK bits per channel use. Therefore, the use of the IC detector in the BICM system is reasonable only for a uniform power distribution or in combination with a low rate code. Note that this reasoning does not apply to the MLC system where the constellation-constrained capacity can be theoretically achieved with the IC detector, see e.g. [Mec03].

Fig. 7.6 depicts EXIT functions for 3 superimposed layers with BICM. As expected, the EXIT function of the MAP detector upperbounds the function of the IC detector and both curves merge at $I_{\text{dec}} = 1$ where the interference from the other layers is eliminated. These observations suggest a hybrid detection approach not further investigated here, where the MAP detector is used for the first few iterations to start the iterative process and the IC detector is used for the last iterations. Then, the performance of the MAP detector could be approached with lower complexity.

We furthermore observe in Fig. 7.6 that the analytic IC function is an acceptable approximation of the simulated curve, considering that the assumption of Gaussian signals is not really valid with 3 layers. As expected from the optimization results, I_{det} is higher with the non-uniform power distribution for low a priori information while the uniform power distribution performs better for high a priori information. Similarly, the values I_{det} with Anti-Gray QPSK mapping are lower at $I_{\text{dec}} = 0$ and higher at $I_{\text{dec}} = 1$ than with Gray QPSK mapping.

Multi-level coding scheme

In the MLC based scheme, an EXIT function of the detector cannot be defined in a straightforward way for the general case of unequal power allocation. For a given SNR, power P_k and phase θ_k of the layers $k = 1, \ldots, K$, the MI $I_{\text{det},k} = f_{\text{det}}(I_{\text{dec},j}, \forall j)$ depends on the a priori MI of every single layer, and not just on the average MI as it is the case with the BICM scheme. The EXIT chart analysis becomes a K-dimensional problem, which would be intractable even for a small number of layers. In the equal power case, i.e. $P_k = P, \forall k$, both the a priori MI at the detector input and the extrinsic MI at the detector output are identical for all layers. For that special case, we obtain a single EXIT function and the analysis is similar to the one with BICM.

An approach to break down the multi-dimensional EXIT chart analysis of a multi-user detector to a one-dimensional analysis has been proposed in [SS03]. The main idea is to scale the

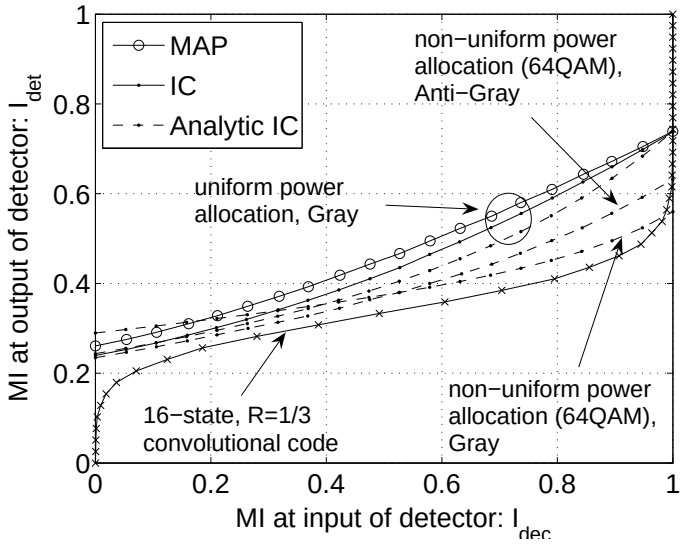


Figure 7.6: EXIT chart with BICM, 3 layers, QPSK with Gray and Anti-Gray mapping, $10\log_{10}(E_s/N_0) = 8$dB; simulated MAP, simulated IC, analytic IC; power allocations $\mathbf{P}^{(1)}$ (64QAM) and $\mathbf{P}^{(2)}$ (uniform power) given in Fig. 7.4.

variance of the LLRs after the detector and the variance of the a priori soft symbols before the detector in an appropriate way. The results in [SS03] are derived for a large number of users. Therefore, it is assumed that the variance of the residual interference embedded in any user's signal is the same for all users (Lemma 1 in [SS03]). This assumption is not valid in the considered system described in Fig. 7.1, where usually a small number of users or layers are superimposed. That has two consequences: First, we have to generate a layer-specific a priori information, and second, we will not obtain a single detector EXIT function for all layers. To handle that, we generate the a priori information for the detector according to the channel code characteristics and obtain a single detector EXIT function by averaging over all layers.

We consider first the EXIT function of the IC detector. With the knowledge of the *layer specific* a priori MI $I_{\text{dec},k}$ fed back from the decoder, we calculate the variance $\sigma^2_{\text{dec},k}$ of the a priori soft-bits (cf. equation (7.9)):

$$\sigma^2_{\text{dec},k} = 1 - T^{-1}(I_{\text{dec},k}). \tag{7.18}$$

To obtain a *layer independent* a priori MI value I_{dec}, we combine the values $\sigma^2_{\text{dec},k}$ as follows:

$$\sigma^2_{\text{dec}} = \frac{1}{\sum_{k=1}^{K} P_k} \sum_{k=1}^{K} P_k \sigma^2_{\text{dec},k}, \tag{7.19}$$

and transfer the variance of the soft-bits to the MI I_{dec} with

$$I_{\text{dec}} = T(1 - \sigma^2_{\text{dec}}). \tag{7.20}$$

For QPSK with Gray mapping, the variance $\sigma^2_{\text{det},k}$ of the LLRs of the kth layer after the IC detector is given by (cf. equation (7.10)):

$$\sigma^2_{\text{det},k} = \frac{4P_k}{\sum_{l \neq k} P_l \sigma^2_{\text{dec},l} + \sigma^2_n}. \tag{7.21}$$

Note that with a large number of layers, the sum in the denominator in equation (7.21) could be performed over all layers and with equation (7.19), we would obtain a simple relation that is independent of the specific distribution of the input variances $\sigma^2_{\text{dec},k}$:

$$\sigma^2_{\text{det},k} = \frac{4P_k}{\sigma^2_{\text{dec}} \cdot \sum_{l=1}^{K} P_l + \sigma_n^2}. \tag{7.22}$$

To obtain a layer independent MI I_{det} for the extrinsic output of the detector for a low number of layers, we scale the LLR variance to normalize the numerator in equation (7.21) and average over all layers since the residual interference in the denominator of equation (7.21) is different for each layer:

$$I_{\text{det}} = \frac{1}{K} \sum_{k=1}^{K} J\left(\sigma_{\text{det},k} \cdot \sqrt{\frac{P_{\text{ref}}}{P_k}} \right), \tag{7.23}$$

where P_{ref} is a reference power and is set to the average power of the layers for convenience. The prediction of the actual trajectory using the proposed approach is more accurate when the differences between the powers P_k are small or with high a priori information when the values $\sigma^2_{\text{dec},k}$ are low, since then, the denominator in equation (7.21) is similar for each layer k.

For the EXIT function of the decoder, the LLR scaling done in equation (7.23) is reversed to obtain a layer specific a priori MI $I_{\text{det},k}$ for the decoder out of the layer independent MI I_{det}:

$$I_{\text{det},k} = J\left(J^{-1}(I_{\text{det}}) \cdot \sqrt{\frac{P_k}{P_{\text{ref}}}} \right). \tag{7.24}$$

Then, $I_{\text{dec},k} = f_{\text{dec}}(I_{\text{det},k})$ is the standard EXIT function of the channel code. To obtain a layer independent MI value I_{dec}, we scale the variance of the corresponding soft symbols according to equations (7.18), (7.19) and (7.20). Fig. 7.7 depicts for $K = 3$ superimposed layers an example of the layer specific decoder EXIT functions $I_{\text{dec},k} = f^k_{\text{dec}}(I_{\text{det}})$, $k = \{1, 2, 3\}$, and the overall decoder EXIT function $I_{\text{dec}} = f^*_{\text{dec}}(I_{\text{det}})$.

Finally, we generate the layer specific a priori information $I_{\text{dec},k}$ for the IC detector in equation (7.18) according to the channel code characteristics: For different values of I_{det}, we use the set of values $I_{\text{dec},k} = f^k_{\text{dec}}(I_{\text{det}})$, $k = 1, \ldots, K$, as shown in Fig. 7.7 for the example of $I_{\text{det}} = 0.3$. The resulting IC detector EXIT function $I_{\text{det}} = f_{\text{det}}(I_{\text{dec}})$ is included in Fig. 7.7. We observe that the analytic IC detector curve matches quite well the simulated curve.

For the EXIT trajectory, the soft-bit output to the detector is processed according to equation (7.23) and the MI output of the decoder is processed according to equations (7.19) and (7.20). We observe in Fig. 7.7 that despite the weakness in the Gaussian assumption when only 3 layers are superimposed, the snapshot trajectory matches fairly well the prediction from the EXIT functions. The EXIT functions with a uniform power allocation would be identical to the BICM counterpart in Fig. 7.6.

Fig. 7.8 illustrates and summarizes the relations between the considered variables.

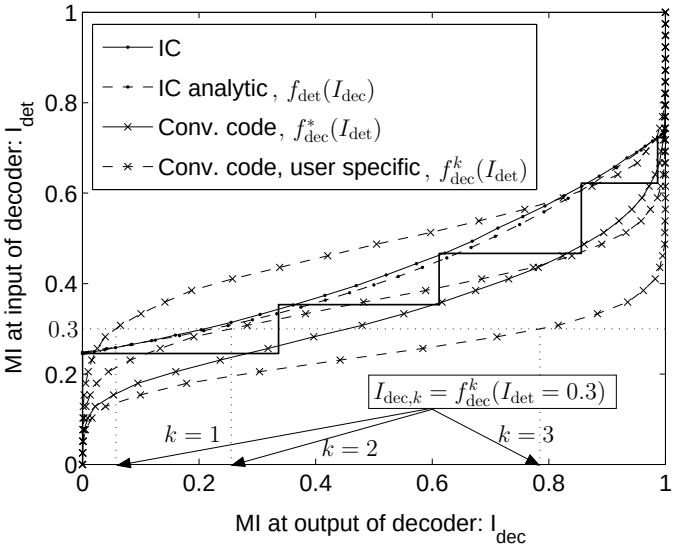


Figure 7.7: EXIT chart with MLC, IC detector, 3 layers, QPSK with Gray mapping, $10\log_{10}(E_s/N_0) = 8$dB; power allocation $\mathbf{P}^{(3)}$ and phase distribution as given in Fig. 7.4c); memory 4, rate $R = 1/3$ convolutional code.

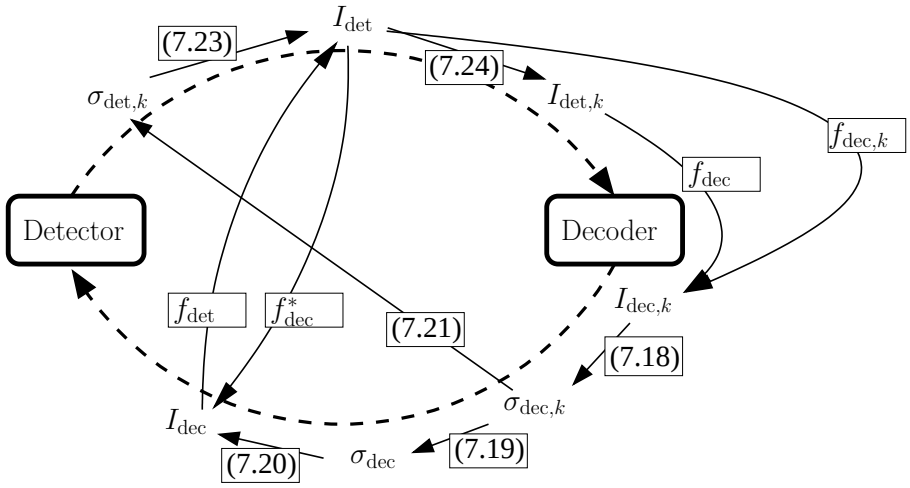


Figure 7.8: Relations between the considered variables with the corresponding equation numbers and functions; σ^2_{dec} is the variance of the soft-bits after the decoder, σ^2_{det} the variance of the LLRs after the detector.

7.4 Optimization of Power Allocation

To achieve close to optimum performance, the detector and decoder EXIT functions should be well matched to avoid an early crossing and to approach capacity. For a fixed set of parameters such as number of layers, signal constellation and mapping of the layers, code rate and code polynomial, we shape these EXIT functions by adjusting the power allocation to the layers. Note that in contrast to BICM, both the detector and the decoder EXIT function depend on the power distribution in the MLC scheme.

The optimization of the power allocation for both BICM and MLC can be cast to a constrained nonlinear optimization problem [SSRS06]:

$$\begin{aligned} &\text{minimize} && \textstyle\sum_{k=1}^{K} P_k, \\ &\text{subject to} && f_{\text{dec}}^{-1}(I_{\text{dec}}) - f_{\text{det}}(I_{\text{dec}}) + \delta < 0, \\ & && P_{min} \leq P_k \leq P_{max}, \forall k, \end{aligned} \tag{7.25}$$

where δ is an arbitrary scalar that determines the minimum width of the tunnel between the EXIT functions. The first constraint is to ensure that the EXIT function of the demapper is always above the one of the decoder and that they do not intersect. Since the EXIT function of the detector does not reach the point $[1; 1]$ in the EXIT chart, a crossing with the function of the decoder is unavoidable at values of I_{dec} close to 1. Thus, the first condition in (7.25) should be relaxed to be valid up to a value of I_{dec} close to 1, depending on the target error rate.

With BICM, the power allocation mainly adjusts the slope of the detector EXIT function and governs the trade-off convergence at low SNR vs. error rate at high SNR, as explained in Section 7.3 and shown in Fig. 7.6. We already optimized I_{det} with no a priori information for an early convergence and with ideal a priori information for a low error rate at high SNR using equation (7.16). With equation (7.25), we can now optimize the trade-off and find the power allocation that optimizes the convergence threshold at low SNR for a desired target error rate at high SNR.

More design possibilities are given with MLC, where both the EXIT function of the detector and decoder depend on the power allocation and the shape of these functions can be precisely adjusted.

An optimized power allocation for MLC with 3 and 6 layers and IC detector is shown in Fig. 7.9. We observe a good match of the curves.

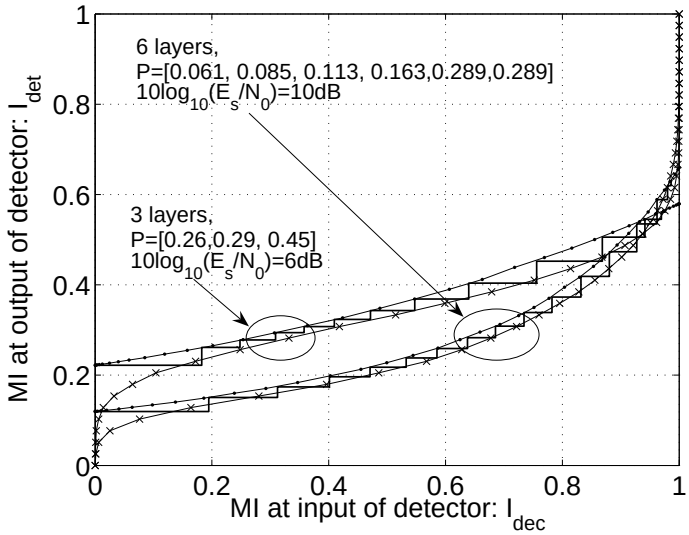


Figure 7.9: EXIT chart with MLC, IC detector, 3 and 6 layers, QPSK with Gray mapping; optimized power distribution; phase $\theta_k = k\pi/(2K)$; memory 4 convolutional code with rate $R = 1/3$ for 3 layers and rate $1/4$ for 6 layers.

D_{ex}	$d_{\text{ex},1}$	$d_{\text{ex},2}$	$d_{\text{ex},3}$	$d_{\text{ex},4}$	...
Λ	λ_1	λ_2	λ_3	λ_4	...
No a priori information (Expurgated)					
Gray	0	176	0	32	...
Anti-Gray	0	264	0	32	...
Ideal a priori information					
Gray	0	64	0	64	...
Anti-Gray	0	32	32	32	...

a) $\mathbf{P}^{(1)}$ (64QAM)

D_{ex}	$d_{\text{ex},1}$	$d_{\text{ex},2}$	$d_{\text{ex},3}$	$d_{\text{ex},4}$	...
Λ	λ_1	λ_2	λ_3	λ_4	...
No a priori information (Expurgated)					
Gray	480	120	0	0	...
Anti-Gray	538	72	0	0	...
Ideal a priori information					
Gray	0	192	0	0	...
Anti-Gray	0	96	96	0	...

b) $\mathbf{P}^{(2)}$ (uniform power)

Table 7.1: Euclidean distance spectrum (EDS) of the overall signal constellations with Gray and Anti-Gray QPSK mapping and the power allocations $\mathbf{P}^{(1)}$ (64QAM) and $\mathbf{P}^{(2)}$ (uniform power) given in Fig. 7.4. $D_{\text{ex}}^2 = \{0, \Delta, 2\Delta, 4\Delta, 5\Delta, \dots\}$ with the minimum squared Euclidean distances $\Delta = 2/21$ and $\Delta = 2/3$ for the normalized signal constellations corresponding to $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$, respectively.

7.5 Simulation Results

In the following bit error rate (BER) simulations, we assume an AWGN channel and uniform (random) interleavers.

Fig. 7.10 depicts the BER for BICM with Gray and Anti-Gray QPSK mapping, corresponding to the EXIT chart in Fig. 7.6. The 3 layers are encoded with a $R = 1/3$ convolutional code with generator polynomial $(1, 33/25, 37/25)$ in octal notation. The information word size is 10000 bits. To compute the analytical error bounds with ideal a priori information at the detector (error free feedback or genie case) shown in Fig. 7.10, we use the Euclidean distance spectrum (EDS) defined in Section 4.1 of the overall signal constellation $\mathcal{X}$ and the error bound derived in Appendix B. The EDS for the power allocations $\mathbf{P}^{(1)}$ (64QAM) and $\mathbf{P}^{(2)}$ (uniform power) as well as for Gray and Anti-Gray QPSK mapping are given in Table 7.1. With a uniform power allocation and no a priori information, we have distances at $d_{\text{ex},1} = 0$ due to the non-unique overall mapping.

The EDS of Table 7.1 and the results in Fig. 7.10 confirm the expectations from the EXIT chart analysis: we achieve a lower BER at high SNR but a later convergence to the genie error bound with a uniform power allocation or Anti-Gray QPSK mapping than with a non-uniform power allocation (64QAM) or Gray QPSK mapping. Any power allocation resulting from the optimization in equation (7.25) will have a higher error bound at high SNR than with the uniform power allocation but an earlier convergence. As already mentioned, a large number of overall signal constellations but only a limited number of overall mappings can be constructed with mapping by superposition. Unfortunately, the error bound at large SNR is quite high with these mappings, especially for non-uniform power allocations. To reduce this error bound, either a stronger channel code or an additional recursive inner code as described in Section 4.4 should be used. If the MAP detector is used instead of the IC detector, a gain of 0.4dB is achieved at low SNR for the example with uniform power allocation and Gray mapping in Fig. 7.10.

Fig. 7.11 shows the BER for MLC averaged over all layers for 3 layers with a convolutional code of rate $R = 1/3$ and generator polynomial $(1, 33/25, 37/25)$, and for 6 layers with a convolutional code of rate $1/4$ and generator polynomial $(1, 27/25, 33/25, 37/25)$. The information word size per layer is 10000 bits. The performance with a uniform power allocation is

good with 3 layers, but no convergence is achieved with 6 layers. With an optimized power allocation according to Fig. 7.9, the convergence threshold is optimized. However, the genie error bounds in Fig. 7.11 are quite high and the earlier convergence of the power optimized receiver may not be of practical relevance. To reduce this error bound, either a stronger channel code or an additional recursive inner code should be used, similar to the approach proposed for BICM with mapping by superposition.

The analytic genie bounds are computed in a slightly different way than with BICM: instead of considering the overall signal constellation, we determine for each layer the EDS of the mappings and signal constellations scaled by the respective allocated power and average over the analytic error bounds of the single layers.

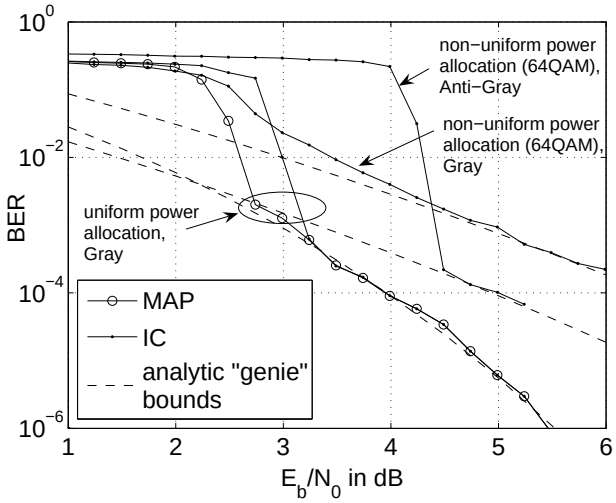


Figure 7.10: BER with BICM, 3 layers, QPSK with Gray and Anti-Gray mapping; memory 4, rate $R = 1/3$ convolutional code; 2 bits per channel use, AWGN channel, 50 iterations; power allocations $\mathbf{P}^{(1)}$ (64QAM) and $\mathbf{P}^{(2)}$ (uniform power) as given in Fig. 7.4; cf. EXIT chart in Fig. 7.6.

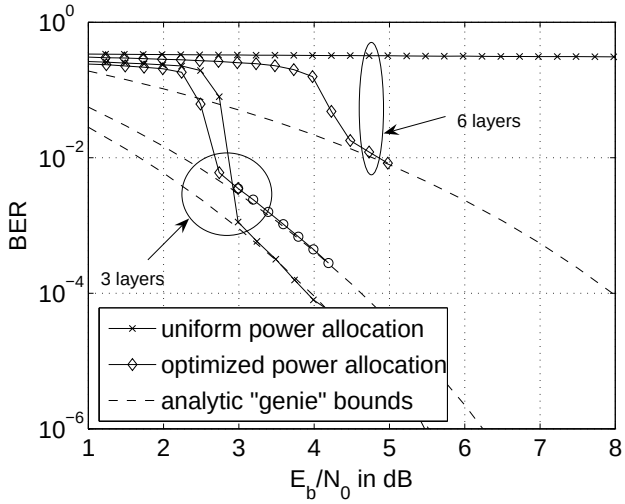


Figure 7.11: Average BER with MLC, IC detector, QPSK Gray mapping; 3 layers with memory 4, rate $1/3$ convolutional code (2 bits per channel use); 6 layers with memory 4, rate $1/4$ convolutional code (3 bits per channel use); AWGN channel, 50 iterations; optimized power allocations according to Fig. 7.9.

7.6 Summary

Signal constellations and bit-to-symbol mappings that arise from the linear superposition of several signal layers are investigated and a mapping by superposition is defined for multilevel coding (MLC) and bit-interleaved coded modulation (BICM). By adjusting the power and phase allocation to the layers, we can construct a large variety of signal constellations, including signal constellations with approximately Gaussian distributed signal points as required to approach the AWGN capacity, standard QAM constellations and QAM constellations with non-unique mappings investigated in Chapter 6. However, only a limited number of overall mappings is obtained. Therefore, a high error bound remains at large SNR and either a strong channel code or an additional recursive inner code should be used.

Analytic EXIT charts are derived for a low complexity interference canceler (IC) and the differences between the EXIT chart analysis of the BICM and MLC schemes with a non-uniform power allocation are highlighted. The BICM approach requires only a single channel code and a single interleaver and is easier to integrate in most recent wireless standards. However, the performance of BICM with the interference canceler is limited by the low receiver-constrained capacity at high SNR and the MLC approach offers better design possibilities.

Using the analytic EXIT charts, the power allocation to the superimposed layers is optimized using constrained nonlinear optimization techniques. For BICM, the performance of the interference canceler without a priori information, the performance with ideal a priori information (genie or error free feedback case) and the receiver-constrained capacity are optimized. For MLC, power allocations are given that optimize the shape of the EXIT functions for capacity approaching performance.

8

Conclusions and Outlook

In this work methods to characterize and optimize bit-to-symbol mappings for a wide range of applications have been developed. The mapping was truly considered as a channel code and integrated in bit-interleaved coded modulation schemes with iterative demapping and decoding (BICM-ID). Using the approaches described in this thesis, additional possibilities arise to design efficient communication systems through advanced modulation techniques. The main contributions that have been achieved in the course of this work are summarized as follows:

- **Properties and computation methods of EXIT charts** were derived for improved analysis and optimization possibilities of iterative decoding. We focused on the demapper and decoder EXIT functions for the BICM system with iterative demapping and decoding. A bit-level and symbol-level EXIT analysis was introduced which reveals the different reliability levels of the decoder and demapper outputs. The capacity loss of rate $R_{\text{in}} < 1$ inner codes in serially concatenated schemes, properties of catastrophic and non-catastrophic rate $R_{\text{in}} = 1$ inner codes and the impact of different channel models (BSC, BEC, and AWGN) on the EXIT functions were studied. Analytic decoder EXIT functions are derived for convolutional codes and turbo codes for BEC channels and different code rates. Demapper EXIT functions are numerically computed. Design guidelines for BICM with iterative demapping and decoding are elucidated to set the trade-off complexity vs. error rate performance and low complexity components vs. number of required iterations.

- An **Euclidean distance spectrum** (EDS) was proposed for BICM based applications. The Euclidean distance spectrum is shown to provide an universal framework to characterize mappings for arbitrary signal constellations and any number of a priori known bits at the demapper. A bit-wise, symbol-wise, expurgated and graphical representation of the Euclidean distance spectrum was investigated.

- The **optimization of mappings** can be cast to a quadratic assignment problem and solved using a simple binary switching algorithm. Cost functions based on the Euclidean distance spectrum, error bounds and mutual information measures were proposed. The main result are mappings optimized for iterative receivers in contrast to Gray mapping optimized for systems without iterative demapping and decoding. The optimization approach

can be used to derive optimized mappings for further applications, e.g. equal or unequal error protection, multi-dimensional mappings or mappings for ARQ.

- The bit-interleaved coded **irregular modulation** (BICIM) scheme was introduced, where different signal constellations and mappings may be used within one code word, even if only the average channel quality is available at the transmitter. The proposed system offers a promising possibility to adapt the transmission system to the channel quality, the complexity and error rate requirements: By combining different signal constellations, the data transmission rate can be precisely adapted to the channel quality. By combining different mappings, the convergence behavior of the iterative decoding and demapping procedure can be optimized. Instead of a cumbersome design of new mappings for different applications, a large variety of mapping characteristics is obtained by the combination of only two mappings, namely Gray mapping and a mapping optimized for iterative demapping and decoding.

- **Non-unique mappings**, where multiple bit labels may be mapped to the same signal point, were used to shape the probability distribution of the signal points. The proposed method is integrated in the BICM scheme with iterative demapping and decoding. The channel code and the iterative process may resolve the ambiguities that arise when multiple bit labels are assigned to one signal point. A greedy label filling algorithm is introduced to optimize the number of labels per signal point and the binary switching algorithm is used to optimize the mapping. Shannon's AWGN channel capacity is approached at high data transmission rates with an optimized Gaussian-like probability distribution of the transmitted signal. The adaptation of the probability distribution of the signal points to the channel characteristics is an inherent way of "soft" adaptive modulation: using optimized probability distributions, the high energy signal points become more probable with increasing SNR. Furthermore, it might be often easier in practical systems to obtain a performance gain by signal shaping than to obtain a similar gain through more powerful coding.

- Signal constellations and bit-to-symbol mappings that arise from the linear superposition of several signal layers were investigated and a **mapping by superposition** was defined for multilevel coding and bit-interleaved coded modulation. With different power and phase allocations to the layers, a large variety of signal constellations can be constructed. Advantages of mapping by superposition include that data from different sources can be easily combined for bandwidth efficient transmission, that for some power allocations, the transmitted signal is approximately Gaussian, as required to approach the AWGN capacity, and that low complexity detectors like the investigated soft interference canceler may be used. As main drawback we should mention the design constraints on the overall bit-to-symbol mapping.

  Analytic EXIT charts were derived for the low complexity interference canceler and for a small number of superimposed layers. For the multilevel coding scheme, the multi-dimensional EXIT analysis is reduced to an one-dimensional problem by scaling the output of the detector and decoder in an appropriate way. Using the analytic EXIT charts, the power allocation to the superimposed layers is optimized for an early convergence with constrained nonlinear optimization techniques.

In the course of this work, several interesting problems have appeared among which the following are of particular interest:

- The EXIT chart analysis could be improved with further derivations of analytic EXIT functions for convolutional codes. In particular, closed form EXIT functions for BSC channels would be useful since from the observations in Section 3.2 and in combination with the BEC results, we would obtain upper and lower bounds on the EXIT functions with an AWGN channel.

- Characterization and optimization methods for mapping diversity techniques for ARQ have been described in Section 4.2. Further interesting results could be derived, in particular a detailed capacity and EXIT chart analysis for mapping diversity techniques is currently missing.

- In this work, the proposed mappings have been used in a BICM scheme with iterative demapping and decoding. An interesting application of the new mappings would be to integrate them in e.g. iterative equalization, iterative multi-user, iterative multi-antenna or joint source/channel coding and modulation systems. Additional performance gains are expected with mappings different from Gray. The mapping characteristics can be easily adapted to the system requirements with irregular (hybrid) modulation.

- The optimization of mappings and the optimization of the number of non-unique bit labels for signal shaping was cast to a quadratic assignment problem (QAP). Furthermore, linear and non-linear programming methods have been used to optimize irregular modulation and the power allocation, respectively. Further optimization and detection problems in communications might be solved by casting them to these standard problems.

- For large signal constellations like 64QAM and with iterative demapping and decoding, the optimal demapping rules are too complex to be implemented in low cost practical systems. We only considered the optimal demapper and for mapping by superposition a simple interference canceler. Further work is required to investigate well suited low-complexity receivers.

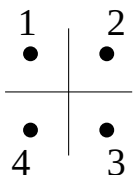
A

Optimized Mappings for Different Signal Constellations

This appendix defines the mappings investigated in this thesis for different signal constellations. The expurgated Euclidean distance spectrum (EDS) for no a priori information and the EDS with ideal a priori information are given as reference.

The mappings are defined as follows: the 2^M binary indices of length M are enumerated in ascending order and mapped to the respective signal points given in the tables. With e.g. 8PSK Gray mapping defined in the table of Fig. A.2, the label (000) is mapped to signal point number 8, (001) is mapped to signal point number 7, (010) is mapped to signal point number 1, a.s.o.

QPSK



Numeration of signal points

Gray	1,2,4,3
Anti-Gray	1,2,3,4

Definition of mappings

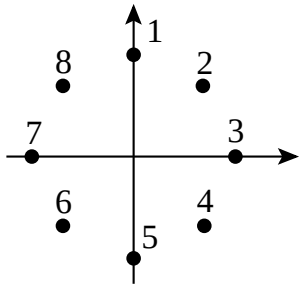
D_{ex}	$d_{\text{ex},1}$	$d_{\text{ex},2}$
Λ	λ_1	λ_2
No a priori information (Expurgated)		
Gray	4	0
Anti-Gray	6	0
Ideal a priori information		
Gray	4	0
Anti-Gray	2	2

Euclidean distance spectrum (EDS)

Table A.1: QPSK mappings

8PSK

SP: set partitioning mapping [Ung82]
SSP: semi set-partitioning mapping [CR02]



Numeration of signal points

Gray	8,7,1,2,5,6,4,3
SP	3,2,1,8,7,6,5,4
SSP	3,6,1,4,7,2,5,8

Definition of mappings

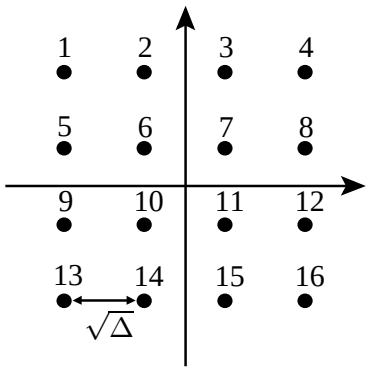
D_{ex}	$d_{\text{ex},1}$	$d_{\text{ex},2}$	$d_{\text{ex},3}$	$d_{\text{ex},4}$
Λ	λ_1	λ_2	λ_3	λ_4
No a priori information (Expurgated)				
Gray	8	4	0	0
SP	14	2	0	0
SSP	18	0	0	0
Ideal a priori information				
Gray	8	0	4	0
SP	4	4	0	4
SSP	0	4	4	4

Euclidean distance spectrum (EDS)

Table A.2: 8PSK mappings

16QAM

SP: set partitioning mapping [Ung82]
MSP: modified set partitioning mapping [CR01]
Rand: [CR01]
MSEW: maximum squared Euclidean weight [TS02]
M16a, M16r: [SGHB03b] [SGHB03a]
I16: [SB04a]



Numeration of signal points

Gray, HSPDA	7,3,8,4,11,15,12,16,6,2,5,1,10,14,9,13
SP	11,10,16,13,9,12,14,15,1,4,6,7,3,2,8,5
MSP	11,5,6,12,9,7,8,10,1,15,16,2,3,13,14,4
Rand	5,15,16,7,11,4,3,9,12,6,1,14,2,13,10,8
MSEW	11,2,5,16,13,8,3,10,4,9,14,7,6,15,12,1
M16a	13,6,7,16,3,12,14,5,8,15,9,2,10,1,4,11
M16r	5,12,15,2,7,13,1,11,4,10,6,16,14,3,8,9
I16	15,9,1,7,6,4,12,14,5,3,11,13,16,10,2,8

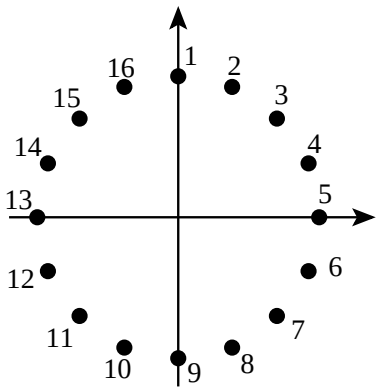
Definition of mappings

D^2_{ex}	$d^2_{\text{ex},1}=\Delta$	$d^2_{\text{ex},2}=2\Delta$	$d^2_{\text{ex},3}=4\Delta$	5Δ	8Δ	9Δ	10Δ	13Δ	18Δ
Λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
No a priori information (Expurgated)									
Gray	24	0	8	0	0	0	0	0	0
SP	56	0	4	0	0	0	0	0	0
MSP	52	0	4	0	0	0	0	0	0
Rand	52	2	0	0	0	0	0	0	0
MSEW	72	0	0	0	0	0	0	0	0
M16a	56	0	4	0	0	0	0	0	0
M16r	56	2	0	0	0	0	0	0	0
I16	52	0	4	0	0	0	0	0	0
Ideal a priori information									
Gray	24	0	0	0	0	8	0	0	0
SP	4	8	8	0	8	4	0	0	0
MSP	0	2	8	4	8	0	4	4	2
Rand	0	0	4	12	4	0	6	4	2
MSEW	0	0	0	24	0	0	0	8	0
M16a	0	0	0	16	4	0	4	8	0
M16r	0	0	4	8	8	0	8	4	0
I16	0	0	0	16	8	0	0	8	0

Euclidean distance spectrum (EDS)

Table A.3: 16QAM mappings

16PSK



Numeration of signal points

Binary Reflected Gray	1,2,4,3,8,7,5,6,16,15,13,14,9,10,12,11
Balanced Gray	1,2,10,3,16,15,11,4,8,7,9,6,13,14,12,5
Turbo AWGN	14,8,9,2,3,12,13,7,6,16,1,10,11,4,5,15

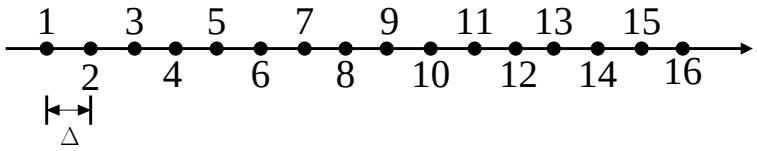
Definition of mappings

D_{ex}	$d_{\text{ex},1}$	$d_{\text{ex},2}$	$d_{\text{ex},3}$	$d_{\text{ex},4}$	$d_{\text{ex},5}$	$d_{\text{ex},6}$	$d_{\text{ex},7}$	$d_{\text{ex},8}$
Λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
No a priori information (Expurgated)								
Binary Reflected Gray	16	8	4	4	0	0	0	0
Balanced Gray	16	11	7	0	0	0	0	0
Turbo AWGN	38	36	42	26	30	14	22	4
Ideal a priori information								
Binary Reflected Gray	16	0	8	0	4	0	4	0
Balanced Gray	16	0	5	0	4	0	7	0
Turbo AWGN	0	0	0	2	8	4	2	4

Euclidean distance spectrum (EDS)

Table A.4: 16PSK mappings

16ASK



Numeration of signal points

Gray	1,2,4,3,8,7,5,6,10,11,13,12,9,16,14,15
Turbo AWGN	5,16,10,6,13,8,2,14,11,7,1,12,3,15,9,4
Non-unique Gray	11,8,11,8,12,7,13,7,11,8,14,8,12,1,14,7,10,9,10,9,4,6,4,6, 10,9,10,9,4,2,3,3,11,8,11,8,12,7,13,7,11,8,15,8,12,7,13,7, 10,9,10,9,5,6,5,6,10,9,16,9,5,6,5,6
Non-unique Turbo AWGN	10,6,5,10,6,11,11,6,15,9,8,13,10,6,4,9,16,9,8,4,10,6,5,10, 7,12,11,7,3,9,8,14,1,9,8,3,10,6,5,10,7,12,11,7,2,9,8,14,7, 13,12,8,4,9,9,5,11,7,7,11,8,13,12,8

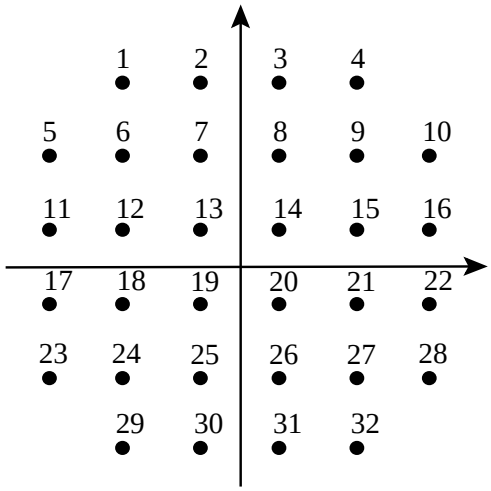
Definition of mappings

D_{ex}	$d_{\text{ex},1}=0$	$d_{\text{ex},2}=\Delta$	2Δ	3Δ	4Δ	5Δ	6Δ	7Δ	...
Λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	...
No a priori information (Expurgated)									
Gray	0	15	9	4	1.5	1	1	1	...
Turbo AWGN	0	35	8	1	0	0	0	0	...
Non-unique Gray	488	716	350	124	56	32	16	16	...
Non-unique Turbo AWGN	832	139	11	0	0	0	0	0	...
Ideal a priori information									
Gray	0	15	0	5	0	3	0	2	...
Turbo AWGN	0	0	0	0	2	4	4	2	...
Non-unique Gray	70	56	8	16	6	6	14	8	...
Non-unique Turbo AWGN	0	0	0	43	58	44	26	11	...

Euclidean distance spectrum (EDS)

Table A.5: 16ASK mappings

32QAM



Numeration of signal points

Gray-like	18,12,24,6,17,11,23,5,21,15,27,9,22,16,32,4,19,13,25,7,29, 1,30,2,20,14,26,8,28,10,31,3
Turbo AWGN	27,2,23,20,12,31,9,11,1,26,14,5,16,6,30,15,18,32,8,17,10, 13,25,4,22,7,29,21,19,28,3,24

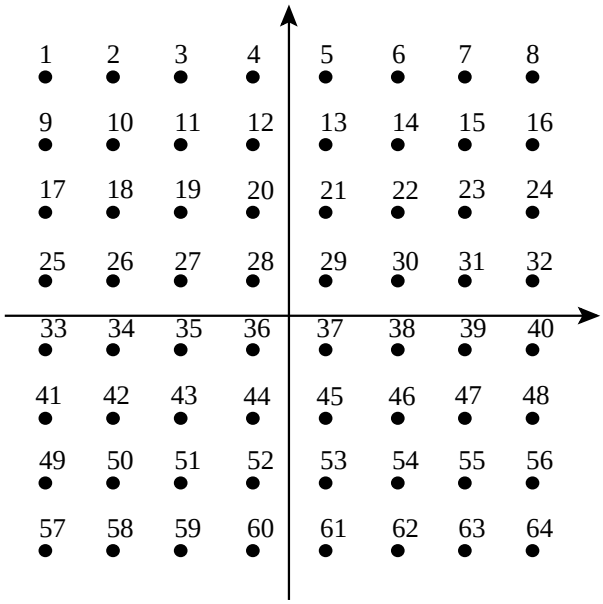
Definition of mappings

D_{ex}	$d_{\text{ex},1}$	$d_{\text{ex},2}$	$d_{\text{ex},3}$	$d_{\text{ex},4}$	$d_{\text{ex},5}$	$d_{\text{ex},6}$	$d_{\text{ex},7}$	$d_{\text{ex},8}$	...
Λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	...
No a priori information (Expurgated)									
Gray-like	60	4	19	0	0	8	0	0	...
Turbo AWGN	126	8	4	0	0	0	0	0	...
Ideal a priori information									
Gray-like	48	0	0	12	0	10	0	0	...
Turbo AWGN	0	0	0	0	4	2	16	16	...

Euclidean distance spectrum (EDS)

Table A.6: 32QAM mappings

64QAM



Numeration of signal points

Gray	1,2,4,3,8,7,5,6,9,10,12,11,16,15,13,14,25,26,28,27,32,31, 29,30,17,18,20,19,24,23,21,22,57,58,60,59,64,63,61,62,49, 50,52,51,56,55,53,54,33,34,36,35,40,39,37,38,41,42,44,43, 48,47,45,46,
Turbo AWGN	31,43,41,5,34,15,30,60,50,4,21,48,22,62,57,27,42,3,13,55, 23,53,58,20,12,47,56,26,61,18,28,40,35,8,24,44,32,36,33, 16,6,45,59,10,49,2,29,63,7,37,52,9,51,1,14,54,46,17,19,39, 11,38,64,25

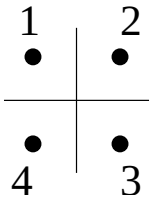
Definition of mappings

D_{ex}	$d_{\text{ex},1}$	$d_{\text{ex},2}$	$d_{\text{ex},3}$	$d_{\text{ex},4}$	$d_{\text{ex},5}$	$d_{\text{ex},6}$	$d_{\text{ex},7}$	$d_{\text{ex},8}$	$d_{\text{ex},9}$	$d_{\text{ex},10}$	...
Λ	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	...
No a priori information (Expurgated)											
Gray	112	0	48	0	0	16	0	0	16	0	...
Turbo AWGN	304	24	5.5	2	0.5	0	0	0	0	0	...
Ideal a priori information											
Gray	112	0	0	0	0	48	0	0	0	0	...
Turbo AWGN	0	0	0	0	0	0	0	0	0	9	...

Euclidean distance spectrum (EDS)

Table A.7: 64QAM mappings

Two-Dimensional QPSK



Numeration of signal points

2-D Turbo opt	(2,2), (3,4), (3,3), (2,1), (4,3), (1,1), (1,2), (4,4), (1,3), (4,1), (4,2), (1,4), (3,2), (2,4), (2,3), (3,1)

Definition of mappings

Table A.8: Definition of two-dimensional QPSK mappings: the $2^{MN_s} = 16$ bit labels of length $MN_s = 4$ are jointly mapped to the $N_s = 2$ QPSK symbols in brackets, respectively. The Euclidean distance spectrum (EDS) given in Table 4.4.

Symbol Mapping Diversity with QPSK

Gray	1^{st} tx:(1, 2, 4, 3), 2^{nd} tx:(1, 2, 4, 3), 3^{rd} tx:(1, 2, 4, 3)
Mixed	1^{st} tx:(1, 2, 4, 3), 2^{nd} tx:(1, 3, 4, 2), 3^{rd} tx:(1, 4, 3, 2)

Table A.9: Definition of QPSK mappings with symbol mapping diversity for three retransmissions: the $2^M = 4$ bit labels of length $M = 2$ are mapped to the QPSK symbols with eventually different mappings for every retransmission or diversity branch. The Euclidean distance spectrum (EDS) given in Table 4.5.

B

Error Probability Analysis

The union bound for the probability of bit error for convolutional codes of rate k_c/N_c is given by

$$P_b \leq \frac{1}{k_c} \sum_{d=d_f}^{\infty} c_d f(d, \mu, \mathcal{X}), \tag{B.1}$$

where d_f is the free distance of the code and c_d is the cumulated Hamming weight of all information words generating weight d code words as defined in Section 2.4. Values of c_d for various non-systematic feedforward convolutional codes are tabulated in [FOO99]. The tables in Appendix C give values of c_d for systematic recursive convolutional codes used in this work. $f(d, \mu, \mathcal{X})$ is the pairwise error probability (PEP) that depends on the Hamming distance d, the labeling map μ and the signal constellation $\mathcal{X}$.

We derive in the following bounds on the pairwise error probability for bit-interleaved coded modulation (BICM) without and with iterative demapping and decoding based on the approaches proposed in [BB99] [CTB98]. A new feature is to use the Euclidean distance spectrum (EDS) as defined in Section 4.1 to determine these bounds.

The pairwise error probability is in general given by the tail probability of an *average* metric Δ:

$$f(d, \mu, \mathcal{X}) = P(\Delta \leq 0). \tag{B.2}$$

Consider the two encoded and interleaved bit sequences $\mathbf{c}$ and $\hat{\mathbf{c}}$ that differ in d bits. M consecutive bits are grouped to form the subsequences $\mathbf{c}_n$ and $\hat{\mathbf{c}}_n$ that are mapped to the symbols x_n and $\hat{x}_n$, $n = 1, \ldots, N_s$, respectively. We assume ideal interleaving and that each subsequence $\mathbf{c}_n$ contains at most one bit that differs from the subsequence $\hat{\mathbf{c}}_n$. Therefore, the symbol sequences $\mathbf{x}$ and $\hat{\mathbf{x}}$ of length N_s differ in d symbols.

The sequence $\mathbf{x}$ is transmitted and $\mathbf{y}$ is received. From equation (2.28), the ML metric difference of the two sequences is

$$\Delta(\mathbf{x}, \hat{\mathbf{x}}) = \log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{y}|\hat{\mathbf{x}}). \tag{B.3}$$

If $\Delta(\mathbf{x}, \hat{\mathbf{x}}) \leq 0$, the receiver would decide for the wrong sequence $\hat{\mathbf{x}}$. For a memoryless channel, we write

$$\Delta(\mathbf{x}, \hat{\mathbf{x}}) \quad = \quad \sum_{n=1}^{N_s} \Delta(x_n, \hat{x}_n) = \sum_{n=1}^{N_s} \left(\log p(y_n|x_n) - \log p(y_n|\hat{x}_n) \right).$$

For the fading channel introduced in equation (2.8), the metric difference can be simplified to

$$\Delta(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{n=1}^{N_s} \left(-|y_n - h_n \cdot x_n|^2 + |x_n - h_n \cdot \hat{x}_n|^2 \right). \tag{B.4}$$

To ease the calculation of the average pairwise error probability $f(d, \mu, \mathcal{X}) = P(\Delta \leq 0)$, we use the Laplace transform. The aim is to determine an expression for the Laplace transform $\Phi_\Delta(s)$ of the probability density function $f_\Delta(\Delta)$ of the average metric Δ.

We start by considering the metric $\Delta(\mathbf{x}, \hat{\mathbf{x}})$ and the Laplace transform $\Phi_{\Delta(\mathbf{x},\hat{\mathbf{x}})}(s)$ of $f_\Delta(\Delta(\mathbf{x}, \hat{\mathbf{x}}))$ for two specific sequences $\mathbf{x}$ and $\hat{\mathbf{x}}$:

$$\Phi_{\Delta(\mathbf{x},\hat{\mathbf{x}})}(s) = \mathrm{E}_\Delta\{\mathrm{e}^{-s\Delta(\mathbf{x},\hat{\mathbf{x}})}\} = \prod_{n=1}^{N_s} \mathrm{E}_\Delta\{\mathrm{e}^{-s\Delta(x_n,\hat{x}_n)}\} = \prod_{n=1}^{N_s} \Phi_{\Delta(x_n,\hat{x}_n)}(s). \tag{B.5}$$

For the symbol combination $(x_n, \hat{x}_n)$, $\Phi_{\Delta(x_n,\hat{x}_n)}(s)$ is given for the general case of a Rician fading channel with Rice factor K [BB99] and $E_s = 1$:

$$\Phi_{\Delta(x_n,\hat{x}_n)}(s) = \frac{1+K}{(1+K) - s(N_0 s - 1)|x_n - \hat{x}_n|^2} \cdot \exp\left(\frac{Ks(N_0 s - 1)|x_n - \hat{x}_n|^2}{(1+K) - s(N_0 s - 1)|x_n - \hat{x}_n|^2} \right). \tag{B.6}$$

We set $K \to \infty$ for an AWGN channel to obtain

$$\Phi_{\Delta(x_n,\hat{x}_n)}(s) = \exp\left(s(N_0 s - 1)|x_n - \hat{x}_n|^2 \right), \tag{B.7}$$

and $K = 0$ for a fully interleaved fading channel to obtain:

$$\Phi_{\Delta(x_n,\hat{x}_n)}(s) = \frac{1}{1 - s(N_0 s - 1)|x_n - \hat{x}_n|^2}. \tag{B.8}$$

Then, we use the Euclidean distance spectrum (EDS) defined in Section 4.1 to obtain the Laplace transform $\Phi_\Delta(s)$ of the average metric Δ. Instead of averaging over all possible sequences $\mathbf{x}$ and $\hat{\mathbf{x}}$, we use the information from the EDS on the number λ_v of relevant distinct Euclidean distances $d_{\mathrm{ex},v} = |x_n - \hat{x}_n|$. With $\Phi_{\Delta(d_{\mathrm{ex},v})}(s)$ from equation (B.6) and using equation (B.5) with sequences that differ in d symbols, we have

$$\Phi_\Delta(s) = \left(\frac{1}{\sum_{v=1}^{V} \lambda_v} \sum_{v=1}^{V} \lambda_v \cdot \Phi_{\Delta(d_{\mathrm{ex},v})}(s) \right)^d. \tag{B.9}$$

According to the selected Euclidean distance spectrum, the case of no a priori information and ideal a priori information at the demapper is considered.

Using equations (B.6) and (B.9), we obtain the exact pairwise error probability through numerical integration or an estimate of the pairwise error probability using the Chernoff bound as described in the following.

Numerical Integration

Recall that the Laplace transform $\Phi_\Delta(s)$ derived in equation (B.9) of the pdf $f_\Delta(\Delta)$ of Δ is

$$\Phi_\Delta(s) = E(\mathrm{e}^{-s\Delta}) = \int_{-\infty}^{\infty} \mathrm{e}^{-s\Delta} f_\Delta(\Delta)\mathrm{d}\Delta. \tag{B.10}$$

The inverse Laplace transform is

$$f_\Delta(\Delta) = \frac{1}{2\pi j} \int_{\alpha-j\infty}^{\alpha+j\infty} \mathrm{e}^{-s\Delta} \Phi_\Delta(s)\mathrm{d}\Delta. \tag{B.11}$$

Using the inverse Laplace formula and $\int_{-\infty}^{0} \mathrm{e}^{-s\Delta}\mathrm{d}\Delta = 1/s$, the exact value of $P(\Delta \leq 0)$ is calculated as follows:

$$P(\Delta \leq 0) = \int_{-\infty}^{0} f_\Delta(\Delta)\mathrm{d}\Delta = \frac{1}{2\pi j} \int_{\alpha-j\infty}^{\alpha+j\infty} \Phi_\Delta(s)\frac{ds}{s}. \tag{B.12}$$

To evaluate this integral, we use the Gauss-Chebyshev quadrature [BB99]:

$$P(\Delta \leq 0) = \frac{1}{2\nu} \sum_{l=1}^{\nu} \mathrm{Re}(\Phi_\Delta(s)) + \tau_l \cdot \mathrm{Im}(\Phi_\Delta(s)) + \epsilon \ , \tag{B.13}$$

where $s = c + jc\tau_l$ and $\tau_l = \tan((2l-1)\pi/\nu)$. The error term ϵ vanishes as $\nu \to \infty$. Note that c affects only the value of ν which is necessary to achieve a prescribed accuracy. A reasonable choice in this respect is the Chernoff bound parameter: $c = 1/(2N_0)$.

Chernoff bound

The Chernoff bound provides a simple, although frequently loose upper bound on the tail probability in the interval (δ, ∞) of a pdf $p(y)$. To derive this bound, we define the function $f(y)$ [Gal68][Pro01]:

$$f(y) = \begin{cases} 1 & y \geq \delta, \\ 0 & y < \delta. \end{cases} \quad \text{and} \quad \mathrm{E}_y\{f(y)\} = P(y \geq \delta). \tag{B.14}$$

This function is overbounded by

$$f(y) \leq \mathrm{e}^{\lambda(y-\delta)} \quad \text{and} \quad \mathrm{E}_y\{f(y)\} = P(y \geq \delta) \leq \mathrm{E}_y\{e^{\lambda(y-\delta)}\}, \tag{B.15}$$

where $\lambda \geq 0$ is the parameter to be optimized. We set $\Delta = -y + \delta$. With equation (B.9) and the definition of the Laplace transform, we obtain the Chernoff bound on the pairwise error probability:

$$P(\Delta \leq 0) \leq \min_{\lambda \geq 0} \mathrm{E}_\Delta\{\mathrm{e}^{-\lambda\Delta}\} = \min_{\lambda \geq 0} \Phi_\Delta(\lambda). \tag{B.16}$$

The optimum λ is $1/(2N_0)$ [BB99].

C

Optimum Systematic Recursive Convolutional Codes

The generators are given in octal notation as described in Section 2.4. As example, $15 \equiv (1101) \equiv 1 + D + D^3$; M_c is the code memory, d_f is the free distance, a_d is the number of error paths in the trellis of Hamming weight d and c_d is the cumulated Hamming weight of all information words that are encoded to weight d code words. The tables extend the results of [FOO99] to systematic recursive encoders.

M_c	Generators	d_f	a_d, c_d with $d = d_f, d_f + 1, \ldots,$
2	5,7	5	(1,2,4,8,16,32,64,128,256,512) (1,4,12,32,80,192,448,1024,2304,5120)
3	15,17	6	(1,3,5,11,25,55,121,267,589,1299) (2,7,18,49,130,333,836,2069,5060,12255)
4	23,35	7	(2,3,4,16,37,68,176,432,925,2156) (4,12,20,72,225,500,1324,3680,8967,22270)

Table C.1: Rate $R = 1/2$ optimum distance spectrum, feedforward non-systematic encoders.

M_c	Generators	d_f	a_d, c_d with $d = d_f, d_f + 1, \ldots,$
2	1,7/5	5	(1,2,4,8,16,32,64,128,256,512) (2,6,14,32,72,160,352,768,1664,3584)
3	1,15/17	6	(1,3,5,11,25,55,121,267,589,1299) (2,12,20,48,126,302,724,1732,4112,9714)
4	1,23/35	7	(2,3,4,16,37,68,176,432,925,2156) (6,12,20,76,194,410,1132,2944,6854,17124)

Table C.2: Rate $R = 1/2$ optimum distance spectrum, recursive systematic encoders.

M_c	Generators	d_f	a_d, c_d with $d = d_f, d_f + 1, \ldots,$
2	1,5/7, 5/7	8	(2,0,5,0,13,0,34,0,89,0) (5,0,15,0,46,0,139,0,413,0)
3	1,15/17,13/17	10	(3,0,2,0,15,0,24,0,87,0) (10,0,8,0,70,0,128,0,523,0)
4	1,33/25, 37/25	12	(5,0,3,0,13,0,62,0,108,0) (20,0,14,0,72,0,368,0,711,0)

Table C.3: Rate $R = 1/3$ optimum distance spectrum, recursive systematic encoders.

M_c	Generators	d_f	a_d, c_d with $d = d_f, d_f + 1, \ldots,$
2	1,7/5,7/5,5/5	10	(1,0,2,0,4,0,8,0,16,0) (2,0,6,0,14,0,32,0,72,0)
3	1,15/13,15/13,17/13	13	(2,1,0,3,1,4,8,4,15,16) (6,4,0,10,3,16,36)
4	1,27/25,33/25,37/25	16	(4,0,2,0,4,0,15,0,30,0) (16,0,8,0,22,0,74,0)

Table C.4: Rate $R = 1/4$ optimum distance spectrum, recursive systematic encoders.

D

Abbreviations and Notation

List of Abbreviations

AMC	adaptive modulation and coding
APP	a posteriori probability
ARQ	automatic repeat request
ASK	amplitude shift keying
AWGN	additive white Gaussian noise
BCJR	MAP algorithm by Bahl, Cocke, Jelinek, Raviv [BCJR74]
BEC	binary erasure channel
BER	bit error rate
BICM	bit-interleaved coded modulation
BICM-ID	bit-interleaved coded modulation with iterative demapping and decoding
BPSK	binary phase shift keying
BSA	binary switching algorithm
BSC	binary symmetric channel
CDMA	code division multiple access
EDGE	enhanced data rates for GSM evolution
EDS	Euclidean distance spectrum
FEC	forward error correction
GSM	global system for mobile communications
HSDPA	high speed downlink packet access
i.i.d.	independent and identically distributed
LLR	log likelihood ratio
MAP	maximum a posteriori probability
MI	mutual information
MIMO	multiple-input/multiple output
ML	maximum likelihood
MLC	multilevel coding
MSD	multistage decoding of MLC
OFDM	orthogonal frequency division multiplex

PAPR	peak-to-average power ratio
PCCC	parallel concatenated convolutional code
pdf	probability density function
PDL	parallel decoding of MLC
PSK	phase shift keying
QAM	quadrature amplitude modulation
QAP	quadratic assignment problem
RCPC	rate compatible punctured codes
SCCC	serial concatenated convolutional code
SISO	soft-in/soft-out
SNR	signal-to-noise ratio
SOVA	soft-output Viterbi algorithm
TCM	trellis coded modulation
UMTS	universal mobile telecommunication system
WiMAX	worldwide interoperability for microwave access
WLAN	wireless local area network

Mathematical Notation

$(.)^*$	conjugate of the argument
$(.)^{-1}$	inverse of the argument
$(.)^T$	transpose of a vector or matrix
$\|.\|$	absolute value or cardinality of the argument
$\overline{(.)}$	expectation of a random variable
$\mathbf{0}_K$	$K \times K$ matrix containing all zeros
arg	operator that delivers the argument
$\mathbb{C}$	field of the complex numbers
E$\{.\}$	expectation of a random variable
$\mathrm{erfc}(.)$	complementary error function
$h(x)$	differential entropy of the random variable X
$H(X)$	entropy of the random variable X
$I(X;Y)$	mutual information between the random variables X and Y
Im$(.)$	imaginary part of argument
$\mathbf{I}_K$	$K \times K$ identity matrix
$\log(.)$	natural logarithm (to base e)
$\log_2(.)$	binary logarithm (to base 2)
$\log_{10}(.)$	decimal logarithm (to base 10)
$\max(.)$	maximum of arguments
$\min(.)$	minimum of arguments
$\mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$	real Gaussian pdf with mean μ and variance σ^2
$\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$	circularly symmetric complex Gaussian pdf where the real and imaginary part are independent Gaussians of mean μ and variance $\sigma^2/2$
$p(x)$	probability density function of the random variable X, same as $p_X(x)$
$P(x)$	probability mass function of the random variable X, same as $P_X(x)$
$\mathbb{R}$	field of the real numbers
Re$(.)$	real part of argument
$\tanh(.)$	hyperbolic tangent
Var$(.)$	variance of random variable

List of Symbols

α_p	ratio for irregular modulation and irregular codes
a_d	Hamming weight distribution of a convolutional code
B_{coh}	coherence bandwidth
c_d	cumulated information weight distribution of a convolutional code
$c_{n,m}$	coded bits
$\mathbf{c}$	code bit sequence with elements $c_{n,m}$
$\mathcal{C}$	channel capacity in bits per channel use
d_f	free distance of a convolutional code
d_v^{ex}	vth distinct Euclidean distance between signal points, for EDS
D^{ex}	set of distinct Euclidean distances, for EDS
D	set of Euclidean distances between signal points that belong to different decision regions, for EDS
ϵ	erasure probability in a BEC channel
E_s	average symbol energy
E_b	average energy per information bit
f_c	carrier frequency
$f_{d,\max}$	maximum doppler shift
$f_{\text{dem}}(.)$	demapper EXIT function
$f_{\text{dec}}(.)$	decoder EXIT function
$f_{\text{det}}(.)$	detector EXIT function
γ_p	ratio for irregular modulation
$g_T(.)$	pulse shaping filter
$g_R(.)$	receiver matched filter
$\mathbf{G}$	binary generator matrix
$\mathbf{G}(D)$	generator matrix
h_n	fading coefficient
$\mathbf{H}$	binary parity check matrix
I	average mutual information
I_L	average mutual information if L bits are a priori known
I_{dem}	average mutual information after the demapper
I_{dec}	average mutual information after the decoder
I_{det}	average mutual information after the detector
λ_v	frequency of the vth distinct Euclidean distance, for EDS
Λ	set of frequencies of distinct Euclidean distances, for EDS
$\Lambda(.)$	metric for MAP decoding
L	number of a priori known bits of a symbol label
$L(.)$	log-likelihood ratio (LLR) of binary random variable
$L_a(.)$	a priori LLR
$L_e(.)$	extrinsic LLR
$L_c(.)$	channel LLR
$L(\hat{.})$	a posteriori LLR
$L_{\text{dem}}(.)$	extrinsic LLR after demapper
$L_{\text{dec}}(.)$	extrinsic LLR after decoder
$L_{\text{det}}(.)$	extrinsic LLR after detector
μ	bit-to-symbol mapping
M	bit label length
M_c	memory of convolutional code

n_n	noise sample
$N_0/2$	power spectral density of white noise process per real dimension
N	code word length
N_c	number of outputs of the encoding circuit of a convolutional code
N_d	number of dimensions in a multidimensional mapping
N_s	symbol word length
N_t	number of retransmissions in an ARQ scheme
Π	interleaver
p	error probability in a BSC channel
p_c	puncturing period
$\mathbf{P}$	puncturing matrix
P_k	power of layer k
R	code rate
R_t	data transmission rate ($R \cdot M$)
R_{out}	code rate of outer code in a serial concatenated system
R_{in}	code rate of inner code in a serial concatenated system
σ_n^2	noise variance (real or complex)
σ_L^2	variance of LLRs
σ_{dec}^2	variance of the soft bits after the decoder
σ_{det}^2	variance of the LLRs after the detector
θ_k	phase of layer k
T	symbol duration
T_{coh}	coherence time
μ_L	mean of LLRs
u_i	information bits
$\mathbf{u}$	information bit sequence with elements u_i
x_n	transmitted complex symbols
$\mathbf{x}$	sequence of transmitted symbols with elements x_n
$\mathcal{X}$	channel input alphabet, signal constellation
y_n	received symbols
$\mathbf{y}$	sequence of received symbols with elements y_n
$\mathcal{Y}$	channel output alphabet

Bibliography

[3GPP06a] 3GPP TS 25.213 version 7.0.0. *Universal Mobile Telecommunications System (UMTS), spreading and modulation (FDD)*. European Telecomunications Standards Institute (ETSI), 2006.

[3GPP06b] 3GPP TS 25.212 version 7.1.0. *Universal Mobile Telecommunications System (UMTS), multiplexing and channel coding (FDD)*. European Telecomunications Standards Institute (ETSI), 2006.

[3GPP01] 3GPP TS 05.04 version 8.4.0. *Digital cellular telecommunications system (Phase 2+), modulation*. European Telecomunications Standards Institute (ETSI), 2001.

[AGR98] P. Alexander, A. Grant, and M. Reed. Iterative detection and code-division multiple-access with error control coding. *European Transactions on Telecommunications (ETT)*, 9:419–425, September-October 1998.

[AKtB04] A. Ashikhmin, G. Kramer, and S. ten Brink. Extrinsic information transfer functions: Model and erasure channel properties. *IEEE Transactions on Information Theory*, 50:2657–2673, November 2004.

[ALSO04] E. Agrelland, J. Lassing, E. Stromand, and T. Ottosson. On the optimality of the binary reflected Gray code. *IEEE Transactions on Information Theory*, 50:3170–3182, December 2004.

[Ari72] S. Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18:14–20, January 1972.

[BAR01] F. Brännström, T. Aulin, and L. Rasmussen. Constellation-constrained capacity for trellis code multiple access systems. In *Proc. IEEE Globecom Conference*, pages 791–795, San Antonio, USA, November 2001.

[BAR02] F. Brännström, T. Aulin, and L. Rasmussen. Iterative detectors for trellis-code multiple-access. *IEEE Transactions on Communications*, 50:1478–1485, September 2002.

[Bär04] S. Bäro. Turbo detection for MIMO systems: Bit labeling and pre-coding. *European Transactions on Telecommunications (ETT)*, 15:1–8, July 2004.

[BB] R. Battiti and M. Brunato. *The reactive search website*. URL: http://www.reactive-search.org.

[BB99] S. Benedetto and E. Biglieri. *Principles of Digital Transmission*. Kluwer Academic / Plenum publishers, 1999.

[BBL+95] M. Best, M. Burnashev, Y. Levy, A. Rabinovich, P. Fishburn, A. Calderbank, and D. Costello. On a technique to calculate the exact performance of a convolutional code. *IEEE Transactions on Information Theory*, 41:441–447, March 1995.

[BCJR74] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Transactions on Information Theory*, 20:284–287, March 1974.

[BCKR] R. Burkard, E. Cela, S. Karisch, and F. Rendl. *QAPLIB - A quadradic assignment problem library*. URL: http://www.seas.upenn.edu/qaplib.

[BDMP98] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara. Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding. *IEEE Transactions on Information Theory*, 44:909–926, May 1998.

[Bel63] P. Bello. Characterization of randomly time-variant linear channels. *IEEE Transactions on Communications*, 11:360–393, December 1963.

[BGM98] S. Benedetto, R. Garello, and G. Montorsi. A search for good convolutional codes to be used in the construction of turbo codes. *IEEE Transactions on Communications*, 46:1101–1105, September 1998.

[BGM+05] S. Benedetto, R. Garello, G. Montorsi, C. Berrou, C. Douillard, D. Giancristofaro, A. Ginesi, L. Giugno, and M. Luise. MHOMS: High speed ACM modem for satellite applications. *IEEE Transactions on Wireless Communications*, 12:66–77, April 2005.

[BGT93] C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error–correcting coding and decoding: Turbo-codes. In *IEEE International Conference on Communications (ICC)*, pages 1064–1070, Geneva, Switzerland, May 1993.

[Bla72] R. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18:460–473, July 1972.

[BMS05] A. Boronka, N. S. Muhammad, and J. Speidel. Removing error floor for bit-interleaved coded modulation MIMO transmission with iterative detection. In *IEEE International Conference on Communications (ICC)*, Seoul, Korea, May 2005.

[BMV04] F. Babich, G. Montorsi, and F. Vatta. Partially systematic rate-compatible punctured SCCCs. *IEEE Communications Letters*, 8:241–243, March 2004.

[Bos98] M. Bossert. *Kanalcodierung*. Teubner, Stuttgart, Germany, 2nd edition, 1998.

[Brä04] F. Brännström. *Convergence Analysis and Design of Multiple Concatenated Codes*. PhD thesis, Department of Computer Engineering, Chalmers University of Technology, Göteborg, Sweden, March 2004.

[BRG05] F. Brännström, L. Rasmussen, and A. Grant. Convergence analysis and optimal scheduling for multiple concatenated codes. *IEEE Transactions on Information Theory*, 51:3354–3364, September 2005.

[BSHA07] G. Bauch, F. Schreckenbach, C. Hausl, and T. Abe. Turbo modulation and coding: Design and evaluation of iterative bit-interleaved coded modulation methods for wireless systems beyond 3G. *European Transactions on Telecommunications (ETT)*, accepted for publication, early view available online, 2007.

[BSS05a] G. Bauch, P. Sethuraman, and F. Schreckenbach. Non-unique differential turbo matrix modulation. In *16th IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Berlin, Germany, September 2005.

[BSS05b] G. Bauch, P. Sethuraman, and F. Schreckenbach. Partially unique mappings for bit-interleaved coded modulation with iterative detection. In *Proc. IEEE Globecom Conference*, St. Louis, USA, December 2005.

[BT94] R. Battiti and G. Tecchiolli. The reactive tabu search. *ORSA Journal on Computing*, 6:126–140, 1994.

[BV98] J. Boutros and E. Viterbo. Signal space diversity: A power- and bandwidth-efficient diversity technique for the rayleigh fading channel. *IEEE Transactions on Information Theory*, 44:1453–1467, July 1998.

[Cah60] C. Cahn. Combined digital phase and amplitude modulation communication system. *IRE Transactions on Communications*, 8:150–155, September 1960.

[CCB95] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham. A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels. *IEEE Transactions on Communications*, 43:773–775, February/March/April 1995.

[Cel98] E. Cela. *The Quadratic Assignment Problem, Theory and Algorithms*. Kluwer Academic, 1998.

[CG62] C. Campopiano and B. Glazer. A coherent amplitude and phase modulation scheme. *IRE Transactions on Communications*, 10:90–95, March 1962.

[CGV04] T. Clevorn, S. Godtmann, and P. Vary. PSK versus QAM for iterative decoding of bit-interleaved coded modulation. In *Proc. IEEE Globecom Conference*, pages 341–345, Dallas, USA, December 2004.

[CGV06] T. Clevorn, S. Godtmann, and P. Vary. Optimized mappings for iteratively decoded BICM on rayleigh channels with IQ interleaving. In *IEEE Vehicular Technology Conference (VTC) Spring*, pages 2083–2087, Melbourne, Australia, May 2006.

[CMT04] G. Caire, R. Müller, and T. Tanaka. Iterative multiuser jont decoding: Optimal power allocation and low-complexity implementation. *IEEE Transactions on Information Theory*, 50:1950–1973, September 2004.

[CO90] A. Calderbank and L. Ozarow. Nonequiprobable signaling on the Gaussian channel. *IEEE Transactions on Information Theory*, 36:726–740, July 1990.

[Cov72] T. Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18:2–14, January 1972.

[CR01] A. Chindapol and J. Ritcey. Design, analysis and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels. *IEEE Journal on Selected Areas in Communications*, 19:944–957, May 2001.

[CR02] A. Chindapol and J. Ritcey. Bit-interleaved coded modulation with iterative decoding and 8PSK signaling. *IEEE Transactions on Communications*, 50:1250–1257, August 2002.

[Cro05] H. S. Cronie. Sparse graph codes for multilevel modulation with signal shaping. In *IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, September 2005.

[CRU01] S. Chung, T. Richardson, and R. Urbanke. Analysis of sum-product decoding of low-density parity-check codes using the Gaussian approximation. *IEEE Transactions on Information Theory*, 47:657–670, February 2001.

[CS87] A. Calderbank and N. Sloane. New trellis codes based on lattice and cosets. *IEEE Transactions on Information Theory*, 33:177–195, March 1987.

[CT91] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[CTB98] G. Caire, G. Taricco, and E. Biglieri. Bit-interleaved coded modulation. *IEEE Transactions on Information Theory*, 44:927–946, May 1998.

[CTC00] O. Collins, O. Takeshita, and D. Costello. Iterative decoding of non-systematic turbo-codes. In *IEEE International Symposium on Information Theory (ISIT)*, page 172, Sorrento, Italy, June 2000.

[CV04] T. Clevorn and P. Vary. Iterative decoding of BICM with non-regular signal constellation sets. In *5th International ITG-Conference on Source and Channel Coding (SCC)*, Erlangen, Germany, January 2004.

[DDP00a] D. Divsalar, S. Dolinar, and F. Pollara. Low complexity turbo-like codes. In *Proc. 2nd international symposium on Turbo codes*, pages 73–80, Brest, France, September 2000.

[DDP00b] D. Divsalar, S. Dolinar, and F. Pollara. Serial concatenated trellis coded modulation with rate-1 inner code. In *Proc. IEEE Globecom Conference*, pages 777–782, San Francisco, November 2000.

[DP95] D. Divsalar and F. Pollara. Multiple turbo codes for deep-space communications. In *The Telecommunications and Data Acquisition Progress Report 42-121*, pages 66–77, Jet Propulsion Laboratory, Pasadena, USA, January–March 1995.

[DRU97] L. Duan, B. Rimoldi, and R. Urbanke. Approaching the AWGN channel capacity without active shaping. In *IEEE International Symposium on Information Theory (ISIT)*, page 374, Ulm, Germany, June/July 1997.

[DS88] D. Divsalar and M. Simon. The design of trellis coded MPSK for fading channels: performance criteria. *IEEE Transactions on Communications*, 36:1004–1012, September 1988.

[Düt05] N. Dütsch. Code optimization for lossless turbo source coding. In *Proc. of 14th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June 2005.

[Eli55] P. Elias. Coding for noisy channels. *IRE Convention Record*, pages 37–47, March 1955.

[FBEM96] D. Forney, L. Brown, V. Eyuboglu, and L. Moran. The V.34 high-speed modem standard. *IEEE Communications Magazine*, 34:28–33, December 1996.

[FGL+84] D. Forney, R. Gallager, G. Lang, F. Longstaff, and S. Qureshi. Efficient modulation for band-limited channels. *IEEE Journal on Selected Areas in Communications*, 2:632–647, September 1984.

[FGW74] G. Foschini, R. Gitlin, and S. Weinstein. Optimization of two-dimensional signal constellations in the presence of Gaussian noise. *IEEE Transactions on Communications*, 22:28–38, January 1974.

[FH96] R. Fischer and J. Huber. A new loading algorithm for discete multitone transmission. In *Proc. IEEE Globecom Conference*, pages 724–728, London, November 1996.

[Fis02] R. Fischer. *Precoding and Signal Shaping for Digital Transmission*. John Wiley, 2002.

[FM00] B. Frey and D. MacKay. Irregular turbo-like codes. In *2nd International Symposium on Turbo codes*, pages 67–72, Brest, France, September 2000.

[FOO99] P. Frenger, P. Orten, and T. Ottosson. Convolutional codes with optimum distance spectrum. *IEEE Communications Letters*, 3:317–319, November 1999.

[For66] G. Forney. *Concatened Codes*. MIT Press, Cambridge, 1966.

[For92] G. D. Forney. Trellis shaping. *IEEE Transactions on Information Theory*, 38:281–300, March 1992.

[Fri95] B. Friedrichs. *Kanalcodierung – Grundlagen und Anwendungen in modernen Kommunikationssystemen*. Springer, Berlin, Germany, 1995.

[FU98] D. Forney and G. Ungerböck. Modulation and coding for linear Gaussian channels. *IEEE Transactions on Information Theory*, 44:2384–2409, October 1998.

[FW89] G. D. Forney and L.-F. Wei. Multidimensional constellations - part I: Introduction, figures of merit, and generalized cross constellations. *IEEE Journal on Selected Areas in Communications*, 7:877–892, August 1989.

[Gal62] R. G. Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*, 8:21–28, January 1962.

[Gal68] R. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, USA, 1968.

[GBB05] N. Gresset, J. Boutros, and L. Brunel. Multi-dimensional mappings for iteratively decoded BICM on multiple antenna channels. *IEEE Transactions on Information Theory*, 51:3337–3346, September 2005.

[GC97] A. Goldsmith and S.-G. Chua. Variable-rate variable-power MQAM for fading channels. *IEEE Transactions on Communications*, 45:1218–1230, October 1997.

[GGB94] S. Le Goff, A. Glavieux, and C. Berrou. Turbo codes and high spectral efficiency modulation. In *IEEE International Conference on Communications (ICC)*, pages 645–649, New Orleans, USA, May 1994.

[GH01] H. El Gamal and A. Hammons. Analyzing the Turbo decoder using the Gaussian approximation. *IEEE Transactions on Information Theory*, 47:671–686, February 2001.

[GL02] F. Glover and M. Laguna. *Tabu Search*. Kluwer Academic, Dordrecht, Netherlands, 5th edition, November 2002.

[GL03] N. Görtz and P. Leelapornchai. Optimization of the index assignments for multiple description vector quantizers. *IEEE Transactions on Communications*, 51:336–340, March 2003.

[Gra53] F. Gray. Pulse code communications. *U.S.Patent No. 2632058*, March 1953.

[GSJ05] S. Le Goff, B. Sharif, and S. Jimaa. Bit-interleaved turbo-coded modulation using shaping coding. *IEEE Communications Letters*, 9:246–248, March 2005.

[Hag88] J. Hagenauer. Rate-compatible punctured convolutional codes (RCPC codes) and their applications. *IEEE Transactions on Communications*, 36:389–400, April 1988.

[Hag97] J. Hagenauer. The turbo principle: Tutorial introduction and state of the art. In *International Symposium on Turbo-Codes*, Brest, France, September 1997.

[Hag04] J. Hagenauer. The EXIT chart - introduction to the extrinsic information transfer in iterative processing. In *12th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, September 2004.

[HH89a] J. Hagenauer and P. Höher. Concatenated viterbi-decoding. In *Proc. of the 4th Joint Swedisch-Soviet International Workshop on Information Theory*, pages 29–33, Gotland, Sweden, September 1989.

[HH89b] J. Hagenauer and P. Höher. A viterbi algorithm with soft-decision outputs and its applications. In *Proc. IEEE Globecom Conference*, pages 1680–1686, Dallas, USA, January 1989.

[HK06] J. Hagenauer and C. Kuhn. The list-sequential (LISS) algorithm and its application. Accepted for Publication in the *IEEE Transactions on Communications*, 2006.

[Hon03] A. Hong. Entwicklung von Signalraumkonstellationen für bit-interleaved coded modulation mit iterativer Detektion. Master's thesis, Intitute for Communications Engineering, Munich University of Technology, Munich, Germany, August 2003.

[HOP93] J. Hagenauer, E. Offer, and L. Papke. Improving the standard coding system for deep space missions. In *IEEE International Conference on Communications (ICC)*, pages 1092–1097, Geneva, Switzerland, May 1993.

[HR05] Y. Huang and J. Ritcey. Optimal constellation labeling for iteratively decoded bit-interleaved space-time coded modulation. *IEEE Transactions on Information Theory*, 51:1865–1871, May 2005.

[HS89] J. Hagenauer and C. Sundberg. On hybrid trellis-coded 8/4-PSK modulation schemes. In *IEEE International Conference on Communications (ICC)*, pages 568–572, Boston, USA, June 1989.

[HSL00] P. Höher, U. Sorger, and I. Land. Log-likelihood values and monte-carlo simulation - some fundamental results. In *Proc. International Symposium on Turbo Codes and Related Topics*, pages 43–46, Brest, France, September 2000.

[HWF98] J. Huber, U. Wachsmann, and R. Fischer. Coded modulation by multilevel-codes: Overview and state of the art. In *Proc. ITG Conference on Source and Channel Coding*, pages 255–266, Aachen, Germany, March 1998.

[IEEE99] IEEE 802.11. *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications*, 1999.

[IEEE04] IEEE 802.16. *IEEE Standard for local and metropolitan area networks, air interface for fixed and mobile broadband wireless access systems*, 2004.

[IH97] H. Imai and S. Hirakawa. A new multilevel coding method using error-correcting codes. *IEEE Transactions on Information Theory*, 23:371–377, May 1997.

[II01] M. Isaka and H. Imai. On the iterative decoding of multilevel codes. *IEEE Journal on Selected Areas in Communications*, 19:935–943, May 2001.

[JZ99] R. Johannesson and K. Zigangirov. *Fundamentals of Convolutional Coding*. IEEE Press, Piscataway, USA, 1999.

[KB57] T. C. Koopmans and J. Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25:53–76, 1957.

[KB90] W. Koch and A. Baier. Optimum and sub-optimum detection of coded data distributed by time-varying intersymbol interference. In *Proc. IEEE Globecom Conference*, pages 1679–1685, San Diego, USA, December 1990.

[KFL01] F. Kschischang, B. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47:498–519, February 2001.

[KP93] F. R. Kschischang and S. Pasupathy. Optimal nonuniform signaling for Gaussian channels. *IEEE Transactions on Information Theory*, 39:913–929, March 1993.

[KSW03] B. Kurkoski, P. Siegel, and J. Wolf. Exact probability of erasure and a decoding algorithm for convolutional codes on the binary erasure channel. In *Proc. IEEE Globecom Conference*, San Francisco, December 2003.

[Lan05] I. Land. *Reliability Information in Channel Decoding*. PhD thesis, Christian-Albrechts-University Kiel, Information and Coding Lab, Kiel, Germany, September 2005.

[LC04] S. Lin and D. Costello. *Error Control Coding*. Prentice-Hall, 2nd edition, 2004.

[LHH04] I. Land, P. Höher, and J. Huber. Analytical derivation of EXIT charts for simple block codes and for LDPC codes using information combining. In *12th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, September 2004.

[LMS+97] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, and V. Stemann. Practical loss-resilient codes. In *Proceeding 29th Annual ACM Symposium Theory of Computing*, pages 150–159, 1997.

[LR97] X. Li and J. Ritcey. Bit-interleaved coded modulation with iterative decoding. *Electronic Letters*, 1:169–171, November 1997.

[LR98] X. Li and J. Ritcey. Bit-interleaved coded modulation with iterative decoding using soft feedback. *Electronic Letters*, 34:942–943, May 1998.

[LSAO03a] J. Lassing, E. Ström, E. Agrell, and T. Ottosson. Computation of the exact bit error rate of coherent M-ary PSK with Gray code bit mapping. *IEEE Transactions on Communications*, 51:1758–1760, November 2003.

[LSAO03b] J. Lassing, E. Ström, E. Agrell, and T. Ottosson. Unequal bit-error protection in coherent M-ary PSK. In *IEEE Vehicular Technology Conference (VTC) Fall*, pages 251–256, Orlando, USA, October 2003.

[LTZ04] M. Lentmaier, D. Truhachev, and K. Zigangirov. Analytic expressions for the bit error probability of rate-$1/2$ memory 2 convolutional encoders. *IEEE Transactions on Information Theory*, 50:1303–1311, June 2004.

[MA01] P. Moqist and T. Aulin. Serially concatenated continuous phase modulation with iterative decoding. *IEEE Transactions on Communications*, 49:1901–1915, November 2001.

[Mas74] J. Massey. Coding and modulation in digital communications. In *International Zürich Seminar on Digital Communications*, Zürich, Switzerland, March 1974.

[Méa06] C. Méasson. *Conservation Laws for Coding*. PhD thesis, School of Computer and Communication Sciences, Ecole Polytechnique Federale de Lausanne - EPFL, Lausanne, January 2006.

[Mec03] M. Mecking. *Fading Multiple-Access with Channel State Information*. PhD thesis, Munich University of Technology, Institute for Communications Engineering, Munich, Germany, July 2003.

[MP04a] X. Ma and L. Ping. Coded modulation using superimposed binary codes. *IEEE Transactions on Information Theory*, 50:3331–3343, December 2004.

[MP04b] X. Ma and L. Ping. Power allocations for multilevel coding with sigma mapping. *Electronic Letters*, 40:609–611, May 2004.

[MS05] N. Muhammad and J. Speidel. Joint optimization of signal constellation and bit-labeling for bit-interleaved coded modulation with iterative decoding. *IEEE Communications Letters*, 9:775–777, September 2005.

[MS06] N. Muhammad and J. Speidel. Design of multidimensional mappings for iterative MIMO detection with minimized bit error floor. In *6th International Symposium on Turbo Codes*, Munich, Germany, April 2006.

[MU02] C. Méasson and R. Urbanke. Asymptotic analysis of turbo codes over the binary erasure channel. In *Proc. of the 12th Joint Conference on Communications and Coding (JCCC)*, Saas Fee, Switzerland, March 2002.

[Muh04] N. Muhammad. Analytische Berechnung von Kenngrößen des EXIT-Charts und Optimierung von QAM-Signalraumpunkten. In *3. Diskussionssitzung der ITG Fachgruppe Angewandte Informationstheorie*, Ulm, Germany, April 2004.

[Nar01] K. Narayanan. Effect of precoding on the convergence of turbo equalization for partial response channels. *IEEE Journal on Selected Areas in Communications*, 19:686–698, April 2001.

[NS99] K. Narayanan and G. Stüber. A serial concatenation approach to iterative demodulation and decoding. *IEEE Transactions on Communications*, 47:956–961, July 1999.

[ÖLGW01] P. Örmeci, X. Liu, D. Goeckel, and R. Wesel. Adaptive bit-interleaved coded modulation. *IEEE Transactions on Communications*, 49:1572–1581, September 2001.

[PLWL06] L. Ping, L. Liu, K. Wu, and W. Leung. Interleave-division multiple-access. to appear in *IEEE Transactions on Wireless Communications*, 2006.

[Pro01] J. Proakis. *Digital Communications*. McGraw-Hill, 4th edition, 2001.

[QZZW05] X. Qi, S. Zhou, M. Zhao, and J. Wang. Design of constellation labelling maps for iteratively demapped modulation schemes based on the assumption of hard-decision virtual channels. *IEE Proc. Communications*, 152:1139–1148, December 2005.

[Rap02] T. Rappaport. *Wireless Communications - Principles and Practice*. Prentice Hall, 2nd edition, 2002.

[RG04] D. Raphaeli and A. Gurevitz. Constellation shaping for pragmatic turbo-coded modulation with high spectral efficiency. *IEEE Transactions on Communications*, 52:341–345, March 2004.

[ROUV93] K. Ramchandran, A. Ortega, M. Uz, and M. Vetterli. Multiresolution broadcast for digial HDTV using joint source/channel coding. *IEEE Journal on Selected Areas in Communications*, 11:6–23, January 1993.

[RS98] M. Reed and C. Schlegel. An iterative receiver for the partial response channel. In *IEEE International Symposium on Information Theory (ISIT)*, page 63, Cambridge, USA, August 1998.

[RSU01] T. Richardson, M. Shokrollahi, and R. Urbanke. Design of capacity-approaching irregular low density parity-check codes. *IEEE Transactions on Information Theory*, 47:619–637, February 2001.

[RU01] T. Richardson and R. Urbanke. The capacity of low density parity-check codes under message passing decoding. *IEEE Transactions on Information Theory*, 47:599–618, February 2001.

[RVH95] P. Robertson, E. Villebrun, and P. Höher. A comparison of optimal and sub-optimal map decoding algorithms in the log domain. In *IEEE International Conference on Communications (ICC)*, pages 1009–1013, Seattle, USA, June 1995.

[RW98] P. Robertson and T. Wörz. Bandwidth-efficient turbo trellis-coded modulation using punctured component codes. *IEEE Journal on Selected Areas in Communications*, 16:206–218, February 1998.

[Sav97] C. Savage. A survey of combinatorial Gray codes. *SIAM Rev.*, 39:605–629, December 1997.

[SB04a] F. Schreckenbach and G. Bauch. EXIT charts for iteratively decoded multilevel modulation. In *12th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, September 2004.

[SB04b] F. Schreckenbach and G. Bauch. Irregular signal constellations, mappings and precoder. In *International Symposium on Information Theory and its Applications (ISITA)*, Parma, Italy, October 2004.

[SB05] F. Schreckenbach and G. Bauch. Bit-interleaved coded irregular modulation. In *Proc. of 14th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June 2005.

[SB06] F. Schreckenbach and G. Bauch. Bit-interleaved coded irregular modulation. *European Transactions on Telecommunications (ETT)*, 17:269–282, March/April 2006.

[SBR06] D. Shepherd, F. Brännström, and M. Reed. Fidelity charts and stopping/termination criteria for iterative multiuser detection. In *4th International Symposium on Turbo Codes and Related Topics*, Munich, Germany, April 2006.

[SC89] C. Schlegel and D. J. Costello. Bandwidth efficient coding for fading channels: code construction and performance analysis. *IEEE Journal on Selected Areas in Communications*, 7:1356–1368, December 1989.

[Sch97] P. Schramm. Multilevel coding with independent decoding on levels for efficient communication on static and interleaved fading channels. In *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1196–1220, Helsinki, Finnland, September 1997.

[SCH05] L. Szczecinski, H. Chafnaji, and C. Hermosilla. Modulation doping for iterative demapping of bit-interleaved coded modulation. *IEEE Communications Letters*, 9:1031–1033, December 2005.

[Sch06] F. Schreckenbach. Approaching AWGN channel capacity using non-unique symbol mappings. In *7th Australian Communications Theory Workshop (AusCTW)*, Perth, Australia, February 2006.

[SDH05] H. Samra, Z. Ding, and P. Hahn. Symbol mapping diversity design for multiple packet transmissions. *IEEE Transactions on Communications*, 53:810–817, May 2005.

[See99] A. Seeger. *Hierarchical Channel Coding for Digital Video Broadcasting*. PhD thesis, Munich University of Technology, Institute for Communications Engineering, Munich, Germany, January 1999.

[SGHB03a] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch. Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding. *IEEE Communications Letters*, 7:593–595, December 2003.

[SGHB03b] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch. Optimized symbol mappings for bit-interleaved coded modulation with iterative decoding. In *Proc. IEEE Globecom Conference*, San Francisco, December 2003.

[SH05] F. Schreckenbach and P. Henkel. Signal shaping using non-unique symbol mappings. In *43rd Annual Allerton Conference on Communication, Control, and Computing*, Monticello, USA, September 2005.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948.

[SHGB05] F. Schreckenbach, P. Henkel, N. Görtz, and G. Bauch. Analysis and design of mappings for iterative decoding of BICM. In *11th National Symposium of Radio Sciences URSI*, Poznan, Poland, April 2005.

[SS03] Z. Shi and C. Schlegel. Performance analysis of iterative detection for unequal power coded CDMA systems. In *Proc. IEEE Globecom Conference*, pages 1537–1542, San Francisco, USA, December 2003.

[SSRS06] D. Shepherd, Z. Shi, M. Reed, and F. Schreckenbach. Optimization of unequal power coded multiuser DS-CDMA using extrinsic information transfer charts. In *40th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, USA, March 2006.

[SSRS07] F. Schreckenbach, D. Shepherd, M. Reed, and Z. Shi. Optimization of coded modulation with mapping by superposition using EXIT charts. In *8th Australian Communications Theory Workshop (AusCTW)*, Adelaide, Australia, February 2007.

[SSS01] F. Sanzi, A. Slama, and J. Speidel. Multi code division multiplex with iterative map symbol-by-symbol estimation. In *Proc. IEEE Globecom Conference*, San Antonio, USA, November 2001.

[SvT93] F. Sun and H. van Tilborg. Approaching capacity by equiprobable signaling on the Gaussian channel. *IEEE Transactions on Information Theory*, 39:1714–1716, September 1993.

[SWK03] A. Sezgin, D. Wübben, and V. Kühn. Analysis of mapping strategies for turbo-coded space-time block codes. In *IEEE Infromation Theory Workshop (ITW)*, pages 103–106, Paris, France, March 2003.

[SWM05] F. Simoens, H. Wymeersch, and M. Moeneclaey. Multi-dimensional mapping for bit-interleaved coded modulation with BPSK/QPSK signaling. *IEEE Communications Letters*, 9:453–455, May 2005.

[tB00] S. ten Brink. Designing iterative decoding schemes with the extrinsic information transfer chart. *AEÜ International Journal of Electronics and Communications*, 54:389–398, November 2000.

[tB01a] S. ten Brink. Code characteristic matching for iterative decoding of serially concatenated codes. *Annals of Telecommunications*, 56:394–408, April 2001.

[tB01b] S. ten Brink. Convergence behavior of iteratively decoded parallel concatenated codes. *IEEE Transactions on Communications*, 49:1727–1737, October 2001.

[tB01c] S. ten Brink. Exploiting the chain rule of mutual information for the design of iterative decoding schemes. In *Proc. 39th Allerton Conf. on Communications, Control and Computing*, Monticello, USA, October 2001.

[tBK03] S. ten Brink and Gerhard Kramer. Turbo processing for scalar and vector channels. In *3rd International Symposium on Turbo Codes*, pages 23–30, Brest, France, September 2003.

[tBSY98a] S. ten Brink, J. Speidel, and R. Yan. Iterative demapping and decoding for multi-level modulation. In *Proc. IEEE Globecom Conference*, pages 579–584, Sydney, Australia, November 1998.

[tBSY98b] S. ten Brink, J. Speidel, and R. Yan. Iterative demapping for QPSK modulation. *Electronic Letters*, 34:1459–1460, July 1998.

[TCMC99] O. Takeshita, O. Collins, P. Massey, and D. Costello. A note on asymmetric turbo codes. *IEEE Communications Letters*, 3:69–71, March 1999.

[TH02] M. Tüchler and J. Hagenauer. EXIT charts of irregular codes. In *Conference on Information Sciences and Systems (CISS)*, Princeton University, March 2002.

[TN04] N. Tran and H. Nguyen. Signal mappings of 8-ARY constellations for BICM-ID systems over a rayleigh fading channel. In *International Conference on Wireless Communications*, pages 464–472, Calgary, Canada, July 2004.

[TN06] N. Tran and H. Nguyen. Design and performance of BICM-ID systems with hypercube constellation. *IEEE Transactions on Wireless Communications*, 5:1169–1179, May 2006.

[TP06] J. Tong and L. Ping. Iterative decoding of superposition coding. In *4th International Symposium on Turbo Codes and Related Topics*, Munich, Germany, April 2006.

[TS02] J. Tan and G. Stüber. Analysis and design of interleaver mappings for iteratively decoded BICM. In *IEEE International Conference on Communications (ICC)*, pages 1403–1407, New York, USA, May 2002.

[TS05a] J. Tan and G. Stüber. Analysis and design of symbol mappers for iteratively decoded BICM. *IEEE Transactions on Wireless Communications*, 4:662–672, March 2005.

[TS05b] H. Tullberg and P. Siegel. Serial concatenated TCM with an inner accumulate code - part I: Maximum-likelihood analysis. *IEEE Transactions on Communications*, 53:64–72, January 2005.

[TtBH02] M. Tüchler, S. ten Brink, and J. Hagenauer. Measures for tracing convergence of iterative decoding algorithms. In *Proc. ITG Conference on Source and Channel Coding*, pages 53–60, Berlin, Germany, January 2002.

[Tüc04] M. Tüchler. Design of serially concatenated systems depending on the block length. *IEEE Transactions on Communications*, 52:209–218, February 2004.

[Ung76] G. Ungerböck. On improving data-link performance by increasing channel alphabet and introducing sequence decoding. In *IEEE International Symposium on Information Theory (ISIT)*, Ronneby, Sweden, June 1976.

[Ung82] G. Ungerböck. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, 28:55–67, January 1982.

[Ung87a] G. Ungerböck. Trellis-coded modulation with redundant signal sets, partI: Introduction. *IEEE Communications Magazine*, 25:5–11, February 1987.

[Ung87b] G. Ungerböck. Trellis-coded modulation with redundant signal sets, partII: State of the art. *IEEE Communications Magazine*, 25:12–21, February 1987.

[Vit67] A. Viterbi. Error bounds for convolutional codes and asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13:260–269, April 1967.

[Vu05] V. T. Vu. Optimization of multi-dimensional mappings for advanced applications. Master's thesis, Intitute for Communications Engineering, Munich University of Technology, Munich, Germany, October 2005.

[Wac98] U. Wachsmann. *Coded Modulation: Theoretical Concepts and PRactical Design Rules*. PhD thesis, University of Erlangen-Nürnberg, Erlangen, February 1998.

[Wei87] L. Wei. Trellis-coded modulation with multidimensional constellations. *IEEE Transactions on Information Theory*, 33:483–501, July 1987.

[Wei01] C. Weiß. *Error Correction with Tail-Biting Convolutional Codes*. PhD thesis, Munich University of Technology, Institute for Communications Engineering, Munich, Germany, September 2001.

[WES04] C. Wengerter, A. Elbwart, and E. Seidel. Constellation rearrangement: enhancement for multilevel modulation formats and transmit diversity. *IEEE Wireless Personal Communications*, 29:35–45, 2004.

[WFH99] U. Wachsmann, R. Fischer, and J. Huber. Multilevel codes: Theoretical concepts and practical design rules. *IEEE Transactions on Information Theory*, 45:1361–1391, July 1999.

[WH92] T. Wörz and J. Hagenauer. Iterative decoding for multilevel codes using reliability information. In *Proc. IEEE Globecom Conference*, pages 1779–1784, Orlando, USA, December 1992.

[Wör96] T. Wörz. *Decodierung von mehrstufiger codierter Modulation*. PhD thesis, Munich University of Technology, Munich, Germany, 1996.

[Zeh92] E. Zehavi. 8-PSK trellis codes for a Rayleigh channel. *IEEE Transactions on Communications*, 40:873–884, May 1992.

[ZG90] K. Zeger and A. Gersho. Pseudo-Gray coding. *IEEE Transactions on Communications*, 38:2147–2158, December 1990.

[ZLH03] L. Zhao, L. Lampe, and J. Huber. Study of bit-interleaved coded space-time modulation with different labeling. In *IEEE Infromation Theory Workshop (ITW)*, pages 199–202, Paris, France, March 2003.