Thesis for the Degree of Licentiate of Engineering

On Bit-Wise Decoders for Coded Modulation

Mikhail Ivanov

CHALMERS

Communication Systems Group
Department of Signals and Systems
Chalmers University of Technology

Gothenburg, Sweden 2013

On Bit-Wise Decoders for Coded Modulation
Mikhail Ivanov

Technical report number: R020/2013
ISSN 1403-266X

This thesis has been prepared using LaTeX and BibTeX.

Communication Systems Group
Department of Signals and Systems
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Telephone: + 46 (0)31-772 1000

Printed by Chalmers Reproservice,
Gothenburg, Sweden, November 2013.

To whom it may concern...

Abstract

Coded modulation is a technique that emerged as a response to the growing demand for high data rates. It was devised to achieve high spectral efficiency with high reliability, potentially approaching Shannon's capacity. The main idea of coded modulation consists in combining error-correcting coding with higher-order modulation. For fast fading channels, the use of binary codes with a bit-wise interleaver between the encoder and the modulator together with a bit-wise decoder was proposed in order to increase the code diversity. This coded modulation technique is called bit-interleaved coded modulation (BICM) and, due to its flexibility of design and good performance, it gained popularity in various wireless communication systems, e.g., WiFi, LTE, etc.

The key component of a BICM scheme is a demapper that, based on the channel observations, calculates L-values (also known as log-likelihood ratios) for the coded bits. In this thesis, we take a closer look at different properties of L-values and their implications when analyzing the performance of bit-wise decoders for coded modulation systems.

First, the demapper is studied in terms of uncoded bit-error rate (BER) over the additive white Gaussian noise (AWGN) channel, i.e., when hard decisions on the bits are made directly at the output of the demapper. A new expression for the BER is formulated for arbitrary one-dimensional constellations. Next, two demapping strategies are considered: when the demapper calculates exact L-values and when L-values are calculated using the so-called max-log approximation. Closed-form expressions for the BER for 4-ary and 8-ary pulse amplitude modulation constellations with some of the most popular binary labelings are found. The numerical results show that there is no difference between the two strategies in terms of BER for any signal-to-noise ratio of practical interest.

We then study the performance of coded systems when the demapper uses the max-log approximation to calculate L-values. We consider a 16-ary quadrature amplitude modulation constellation labeled with a Gray code over the AWGN channel, as well as flat fading channels. At the receiver, a bit-wise decoder is used for decoding, which finds the maximum correlation between L-values and coded bits. This decoder performs the maximum likelihood decoding for the binary-input AWGN channel, however, it is suboptimal when higher-order modulation is considered. We show that the asymptotic loss in terms of pairwise error probability of such a decoder compared to the maximum likelihood decoder is bounded by 1.25 dB for any flat fading channel (including the AWGN channel as a special case). The analysis also shows that for the AWGN channel, the asymptotic loss is zero for a wide range of linear binary codes.

Keywords: Additive white Gaussian noise, bit-error probability, coded modulation, flat fading channel, Gray code, interleaver, labeling, L-values, log-likelihood ratio, maximum likelihood decoder, pairwise error probability, quadrature amplitude modulation.

Acknowledgements

First of all, I would like to thank my main supervisor Fredrik Brännström, my co-supervisors Alex Alvarado and Alexandre Graell i Amat, and my examiner/co-supervisor Erik Agrell. Thanks to their expertise and support I am writing this thesis now. I am very happy to be a part of this coded modulation team.

Second of all, I want to thank all members of the Communication Systems Group. You all together create an amazing working (not only) atmosphere here, on the 6th floor of Hörsalsvägen 11. I really look forward to every working day.

Here comes the financial part. This work has been supported by the Swedish Research Council (VR) under grant #621-2011-5950. I wish to thank Swedish tax payers for that. I'm grateful to Ericsson's Research Foundation for grant #556016-0680 that facilitated my research visit to the University of Cambridge. Thanks to Chalmersska forskningsfonden I had a chance to attend GLOBECOM 2012 and present some of the results included in this thesis.

The calculations were performed in part on resources provided by the Swedish National Infrastructure for Computing (SNIC) at C3SE.

Misha

List of Publications

This thesis is based on the following publications:

Paper A
M. Ivanov, F. Brännström, A. Alvarado, and E. Agrell, "General BER expression for one-dimensional constellations," in *IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, Dec. 2012.

Paper B
M. Ivanov, F. Brännström, A. Alvarado, and E. Agrell, "On the exact BER of bit-wise demodulators for one-dimensional constellations," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1450–1459, Apr. 2013.

Paper C
M. Ivanov, A. Alvarado, F. Brännström, and E. Agrell, "On the asymptotic performance of bit-wise decoders for coded modulation," *Submitted to IEEE Trans. Inf. Theory*, June 2013, available at http://arxiv.org/abs/1306.4009.

Acronyms

ABD:	Approximated Bit-Wise Demodulator
AGC:	Anti-Gray Code
ARE:	Anti-Reflected
ASY:	Asymmetric
AWGN:	Additive White Gaussian Noise
BER:	Bit-Error Rate
BD:	Bit-Wise Demodulator
BICM:	Bit-Interleaved Coded Modulation
BICM-ID:	BICM with Iterative Decoding
bpcu:	Bit Per Channel Use
BPSK:	Binary Phase Shift Keying
BRGC:	Binary Reflected Gray Code
B-DEC:	Bit-Wise Decoder
CC:	Convolutional Code
CM:	Coded Modulation
DEM:	Demapper
ED:	Euclidean Distance
ENC:	Encoder
GL:	Gray Labeling
i.i.d.:	Independent and Identically Distributed
LDPC:	Low-Density Parity-Check
LLR:	Log-Likelihood Ratio
MFD:	Maximum Free Distance
ML:	Maximum Likelihood
MLCM:	Multilevel Coded Modulation
MOD:	Modulator

NBC:	Natural Binary Code
PAM:	Pulse Amplitude Modulation
PBER:	BER for a Pattern
PDF:	Probability Density Function
PEP:	Pairwise Error Probability
PSK:	Phase Shift Keying
QAM:	Quadrature Amplitude Modulation
QPSK:	Quadrature Phase Shift Keying
RE:	Reflected
SER:	Symbol-Error Rate
SD:	Symbol-Based Demodulator
SMD:	Symbol Metric Difference
SNR:	Signal-to-Noise Ratio
SP:	Set Partitioning
S-DEC:	Symbol-Wise Decoder
TCM:	Trellis-Coded Modulation
UB:	Union Bound
ZC:	Zero-Crossing

Contents

Part I

Introduction

Chapter 1

Preliminaries

1.1 Why Coded Modulation?

Coded modulation (CM) is referred to as a technique that combines error-correcting coding with higher-order modulation in order to achieve high spectral efficiency. Its story began in 1948, when Claude Shannon introduced a general communication system model for point-to-point communications and formulated the problem of reliable data transmission [1]. With slight changes, this system model is reproduced in Fig. 1.1. Throughout this thesis, different variations of this system model are considered. The main problem consists in communicating a message m, represented by an integer, through a noisy channel by encoding the message into a vector $\boldsymbol{x}$ (codeword) of symbols from a certain alphabet (constellation) with as small probability of error $\Pr\{\hat{\mathsf{M}} \neq \mathsf{m}\}$ as possible, where $\hat{\mathsf{M}}$ is the estimate of the message. Using his mathematical theory, Shannon showed that, for a wide range of channels, the fundamental limit at which such a system can operate is given by the mutual information (MI) between the channel input and the channel output optimized over the input distributions. This fundamental limit is called channel capacity and it is measured in bits per channel use (bpcu). It shows how many bits of information can be transmitted per symbol (or channel use) with an arbitrarily small probability of error.

In his work that marked the beginning of information theory, Shannon already introduced the technique we now call CM. He even suggested a hypothetical solution for CM to attain the channel capacity. Let us consider a small example of a CM scheme proposed by Shannon. Consider a complex Gaussian channel, where the output of the channel is given by the sum of the input and Gaussian-distributed noise. Under an average power constraint, the channel capacity is a function of the signal-to-noise ratio (SNR) and is given by the famous equation $C = \log_2(1 + \mathrm{SNR})$. How can we achieve this fundamental limit for a certain SNR? Shannon answers this question from a very theoretical point of view as follows. First, let the constellation be all complex numbers. Second, construct a codebook with 2^{nC} codewords of length n (numbered $1, \ldots, 2^{nC}$) by choosing elements of the codewords randomly and independently from the constellation according to a complex Gaussian distribution. Third, transmit the mth codeword in order to communicate message m through the channel. If n grows large and a maximum likelihood (ML) decoder is used at the receiver, the resulting error probability approaches zero and the coding scheme will achieve the channel capacity.

There are four main ingredients in the described coding scheme: the constellation, the code (or codebook), the assignment of the messages to the codewords (or encoding), and the decoding algorithm. The four ingredients described by Shannon lead to solutions which are highly complex and absolutely impractical from an implementation point of

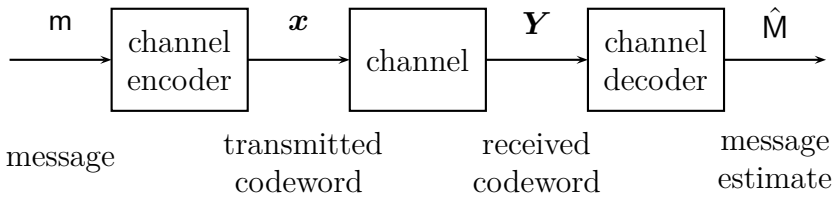
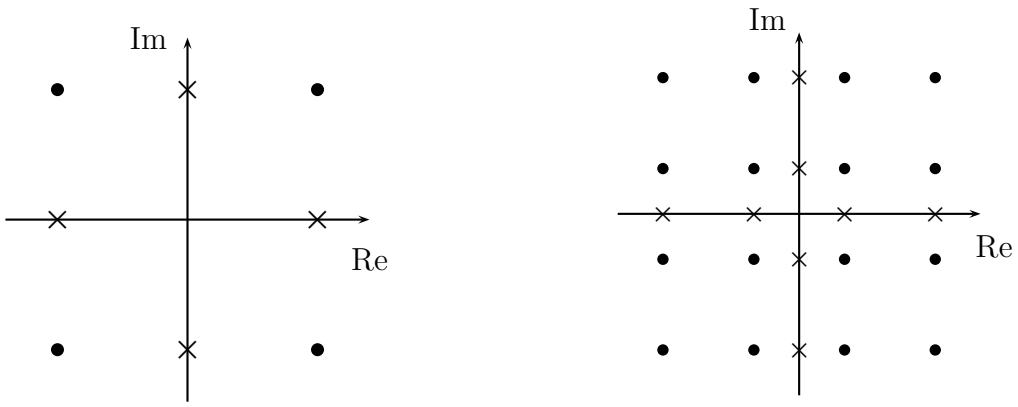


Figure 1.1: A block diagram of a general digital communication system.



(a) QPSK or 4-QAM constellation.

(b) 16-QAM constellation.

Figure 1.2: PAM and QAM constellations in 2-dimensions. Crosses in each dimension show PAM constellations and dots show QAM constellations.

view. Coding theorists and communication engineers spent a great deal of effort to propose alternative ways to approach capacity.

The first step in order to make the coding scheme more realistic was to use discrete constellations. Due to implementation issues, regularly spaced constellations, such as M-ary phase shift keying (PSK), pulse amplitude modulation (PAM), or quadrature amplitude modulation (QAM) formed as the Cartesian product of two PAM constellations, are usually used in practical systems. Examples of PAM constellations are shown in Fig. 1.2 with crosses in each dimension, and the corresponding QAM constellations obtained as the product are shown with dots.

Once a discrete constellation is chosen, labels (binary or nonbinary) can be assigned to the constellation points. By doing so, the codebook construction and encoding can be done in two steps, which are usually regarded as error-correcting coding and modulation. Due to the error-correcting code design, the simplest probability distribution to impose on constellation points is a uniform distribution. Such CM scheme construction generally leads to losses in potentially achievable performance which are clearly seen from Fig. 1.3. This figure shows the MI for M-PAM constellations and demonstrates that binary phase shift keying (BPSK) performs near the capacity in the so-called power-limited regime (low SNR values), whereas in the bandwidth-limited regime (high SNR values) the loss is

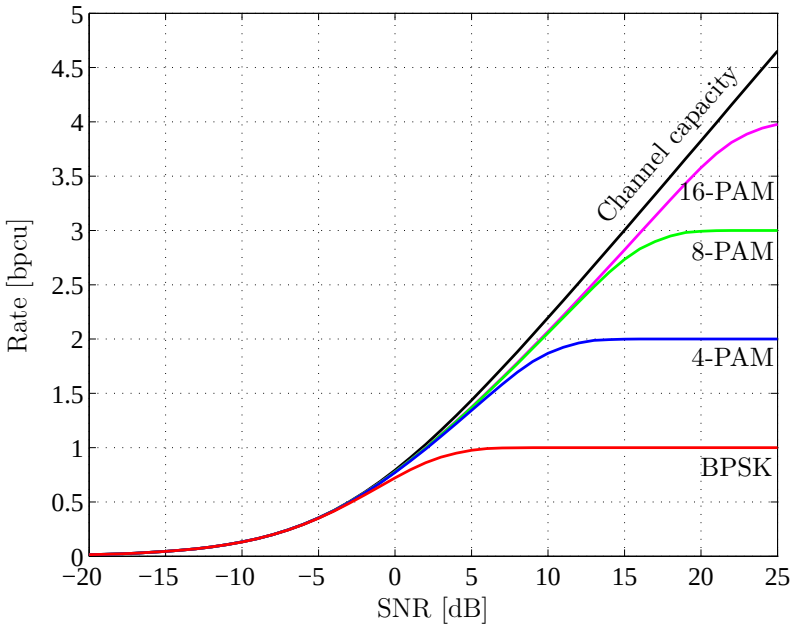


Figure 1.3: MI for M-PAM constellations together the channel capacity for the AWGN channel.

1.53 dB when $M \to \infty$ [2]. Note that this choice of constellations with regularly spaced and equally likely points is mainly due to implementation reasons. One could optimize the positions of the constellation points, for instance, in terms of bit error rate [3] or MI [4], which is referred to as geometrical shaping. Probabilistic shaping, i.e., changing the distribution of constellation points, could also be used to reduce this 1.53 dB gap [5].

For a long time engineers were mainly concerned with the power-limited regime (low SNR region) and the main focus was on achieving capacity for MI below one using BPSK. With the increase of the demand for high date rates, the bandwidth-limited regime came into play. The first approaches to combine higher-order modulation, where binary labels are assigned to the constellation points, with classical codes designed for binary transmission showed very disappointing results, which are well described in Ungerboeck's paper [6]. This even resulted in questioning whether coding is relevant for high spectral efficiencies [7]. The problem with those schemes was actually not a poor design[1] but rather bad decoding techniques adopted from the algebraic coding theory.

The second step towards a good practical CM scheme was to impose a very regular structure on the codebook. In [8] and [9], the technique called TCM was proposed, where convolutional codes (CCs), appropriately tailored to the constellation, are used to obtain codes with a trellis structure. This resulted in easy encoding and, most importantly, it enabled implementation of a "complicated" ML decoder with a reasonable complexity by means of the Viterbi algorithm [10]. In those works, the importance of the Euclidean distance (ED) between codewords for the code performance was highlighted and a set of rules to design TCM systems with a large minimum ED was formulated based on the so-called set partitioning (SP) technique [9]. However, very structured codes are in general

[1]The design of the scheme that Ungerboeck describes as a "bad scheme" follows the bit-interleaved coded modulation (BICM) approach and, in fact, corresponds to a good trellis-coded modulation (TCM) scheme.

unable to achieve the performance predicted by information theory.

The third and the most successful step to make a good practical CM scheme was to allow the code to have poor structure and to develop good suboptimal decoding algorithms. One of the first examples of such CM systems is multilevel coded modulation (MLCM) proposed by Imai and Hirakawa [11]. The main idea was to use different binary codes for different bit positions in the constellation, which made the code less structured compared to TCM, together with a sub-optimal multi-stage decoder, where the binary decoders for the corresponding binary codes are allowed to exchange information between one another.

Another important example of the development in this direction is BICM [12]. Zehavi in [12] suggested to place bit-wise interleavers between the encoder and the modulator and to use a suboptimal bit-wise decoder. The original motivation behind introducing interleavers was to improve the performance over fast fading channels. For such channels, the most important parameter of the code is the so-called code diversity whereas the ED is secondary. The code diversity can be seen as the minimum number of different symbols between any two codewords of the code and, in a sense, is equivalent to time diversity. Interleavers were shown to increase the code diversity making BICM very attractive for fading channels. At the same time, BICM appeared to be very powerful over the AWGN channel. It is currently used in various communications standards, e.g., [13–15] to name a few.

Further developments in coding theory led to iterative decoding algorithms, powerful binary and nonbinary turbo [16] and low-density parity-check (LDPC) [17,18] codes. This, in turn, resulted in powerful and efficient CM schemes, such as turbo TCM [19, 20] and BICM with iterative decoding (BICM-ID) [21, 22].

Even though BICM has been a research topic for more than 20 years, there are still many unsolved problems, e.g., those related to interleaver design, performance analysis for finite interleaver length, maximum achievable rates, etc. In early BICM systems, where convolutional codes were used, as well as in modern coded systems that use, for instance, LDPC codes, the calculation of log-likelihood ratios (LLRs or L-values) is one of the key operations. The knowledge about L-values and their properties can be used to obtain precise performance estimations, new design criteria, and may help to solve some of the open problems. In the following chapters, we take a closer look at different properties of the L-values.

1.2 Notation Convention

The following notation is used in Part I of the thesis. Lowercase letters x denote real or complex scalars and boldface letters $\boldsymbol{x}$ denote row vectors of scalars. The complex conjugate of x is denoted by x^*. Blackboard bold letters $\mathbb{X}$ denote matrices with elements $x_{i,j}$ in the ith row and the jth column and $(\cdot)^\intercal$ denotes transposition. Calligraphic capital letters $\mathcal{X}$ denote sets, where the set of real numbers is denoted by $\mathcal{R}$. Binary addition (exclusive-OR) of two bits a and b is denoted by $a \oplus b$. Random variables are denoted by capital letters X, probabilities by $\Pr\{\cdot\}$, the probability density function (PDF) of X by $f_X(x)$, and the conditional PDF of Y conditioned on $X = x$ by $f_{Y|X}(y|x)$. The Gaussian Q-function is defined as

$$\mathrm{Q}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-t^2/2)\,\mathrm{d}t.$$

1.3 Structure of the Thesis

In Sweden, the Licentiate degree is an intermediate degree between a Master's and a PhD. A doctoral student is given two options when writing a Licentiate thesis: a classical monograph or a collection of papers published by the student. This thesis is written as a collection of papers and it is divided into two parts. Part I gives a general overview of CM techniques with a particular focus on bit-wise decoding algorithms and prepares the reader for the papers that come in Part II. Part I is organized as follows. Chapter 2 deals with uncoded transmission and analyzes the bit-error rate (BER) for bit-wise demodulators for various constellations and labelings. In Chapter 3, TCM and BICM are discussed. Chapter 4 introduces a bit-wise decoder for a general CM scheme and tools for its analysis. Finally, Chapter 5 gives an overview of the contributions made by the author and presents possible future work directions in the area of CM.

Chapter 2

Uncoded Transmission

2.1 System Model

A system model for uncoded transmission can be seen as a special case of Shannon's model where only one channel use is allowed for the coding scheme (codewords have length one) and discrete constellations are used. Even though uncoded transmission is not of a big interest in coded systems, we describe and analyze the uncoded system model in detail since it has many similarities with coded models, which will be discussed in the next chapter.

The system model for uncoded transmission in Fig. 2.1 is a special case of the model in Fig. 1.1. The message m is represented by a binary vector $\boldsymbol{b} = [b_1, \ldots, b_m]$. The vector $\boldsymbol{b}$ is fed to a modulator $\Phi_{\mathcal{S}}$ that carries out a one-to-one mapping from $\boldsymbol{b}$ to one of the $M = 2^m$ constellation points $x \in \mathcal{S} \triangleq \{s_1, \ldots, s_M\}$ for transmission over the physical channel. The modulator is defined as the function $\Phi_{\mathcal{S}} : \{0,1\}^m \to \mathcal{S}$ with the corresponding inverse function $\Phi_{\mathcal{S}}^{-1} : \mathcal{S} \to \{0,1\}^m$. In this case, the constellation can be interpreted as the codebook.

The modulator is fully described by the constellation and its binary labeling. A binary labeling is specified by the matrix $\mathbb{C}$ of dimensions M by m, where the ith row is the binary label of the constellation point s_i. The labeling can be seen as an underlying binary code for this simple CM system. This explains the conventional name for Gray code labelings. An alternative and more compact way to represent a labeling is to use a vector $\boldsymbol{q} = [q_1, \ldots, q_M]$, where q_i is the integer representation of the ith row of $\mathbb{C}$ with the most significant bit to the left.

In this thesis, we are mostly interested in one-dimensional constellations. For PAM constellations, $s_i = -d(M - 2i + 1), i = 1, \ldots, M$, where d is half the distance between the constellation points. The scalar d determines the average symbol energy defined as $E_{\mathrm{s}} = \sum_{i=1}^{M} \Pr\{X = s_i\} s_i^2$. All M possible messages are assumed to be equiprobable, which implies equiprobable symbols, i.e., $\Pr\{X = s_i\} = 1/M$, $\forall i$. This also means that the bits transmitted in the jth position B_j are independent and identically distributed (i.i.d.) with $\Pr\{B_j = u\} = 0.5, \forall j$ and $u \in \{0,1\}$. This gives an average symbol energy of

$$E_{\mathrm{s}} = \frac{M^2 - 1}{3} d^2. \tag{2.1}$$

2.1.1 Channel Model

We consider a discrete time memoryless additive white Gaussian noise (AWGN) channel with output $Y = x + Z$, where $x \in \mathcal{S}$ and the noise sample Z is a zero-mean Gaussian

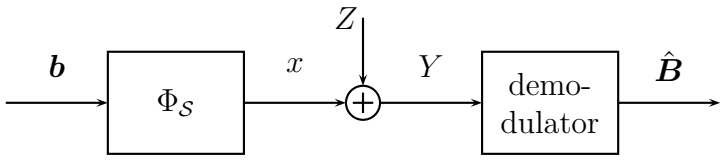


Figure 2.1: The system model for uncoded transmission.

random variable with variance $\sigma^2 = N_0/2$. The conditional PDF of the channel output, given the channel input, is

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(y-x)^2}{2\sigma^2}}, \tag{2.2}$$

and we define the SNR as $\gamma \triangleq E_\mathrm{s}/N_0$. In the rest of this chapter, we assume the constellation to be normalized to unit average energy, i.e., $\gamma = (2\sigma^2)^{-1}$.

The observation Y is used by the demodulator to decide on the received binary sequence, i.e., to produce $\hat{\boldsymbol{B}} = [\hat{B}_1, \ldots, \hat{B}_m]$. The function of the demodulator is equivalent to that of the channel decoder in Fig. 1.1. In the following section, we consider two demodulators that can be used for uncoded transmission.

2.2 Demodulators

2.2.1 Symbol-Based Demodulator

An ML symbol-based demodulator (SD) is the most natural way of decoding symbols transmitted through the channel. It is optimal in terms of minimizing the symbol-error probability and its performance is well documented in literature, e.g., [23, Ch. 5], [24, Ch. 10], [25–30], and references therein.

The decision of the SD can be formally defined as

$$\hat{\boldsymbol{B}}^{\mathrm{SD}} \triangleq {\Phi_{\mathcal{S}}}^{-1}\left(\hat{X}\right), \tag{2.3}$$

where

$$\hat{X} \triangleq \underset{s\in\mathcal{S}}{\operatorname{argmin}} (Y-s)^2. \tag{2.4}$$

The symbol-error rate (SER) is defined as

$$\mathrm{SER} \triangleq \Pr\{X \neq \hat{X}\} = \Pr\{\boldsymbol{B} \neq \hat{\boldsymbol{B}}\}$$

and it can be expressed as

$$\mathrm{SER} = \frac{1}{M} \sum_{s_i\in\mathcal{S}} \sum_{\substack{s_j\in\mathcal{S} \\ s_j\neq s_i}} \Pr\{\hat{X} = s_j | X = s_i\}. \tag{2.5}$$

For the sake of simplicity, a union bound (UB) on the SER is often considered instead, i.e.,

$$\mathrm{SER}_{\mathrm{UB}} \triangleq \frac{1}{M} \sum_{s_i\in\mathcal{S}} \sum_{\substack{s_j\in\mathcal{S} \\ s_j\neq s_i}} \mathrm{PEP}\{\hat{X} = s_j | X = s_i\}, \tag{2.6}$$

where $\mathrm{PEP}\{\hat{X} = s_j | X = s_i\}$ is the pairwise error probability (PEP), i.e., $\Pr\{\hat{X} = s_j | X = s_i\}$ when only two constellation points s_i and s_j are considered. Note that the UB is an upper bound for the exact SER. Taking into account that for the AWGN channel

$$\mathrm{PEP}\{\hat{X} = s_j | X = s_i\} = \mathrm{Q}\left(\sqrt{\frac{d^2(s_i, s_j)}{2N_0}}\right), \tag{2.7}$$

where $d(s_i, s_j) = |s_i - s_j|$ is the ED between the points s_i and s_j, the UB in (2.6) can be written as

$$\mathrm{SER}_{\mathrm{UB}} = \frac{1}{M} \sum_{s_i \in \mathcal{S}} \sum_{\substack{s_j \in \mathcal{S} \\ s_j \neq s_i}} \mathrm{Q}\left(\sqrt{\frac{d^2(s_i, s_j)}{2N_0}}\right). \tag{2.8}$$

The exact BER expression can be written as

$$\mathrm{BER} = \frac{1}{mM} \sum_{s_i \in \mathcal{S}} \sum_{\substack{s_j \in \mathcal{S} \\ s_j \neq s_i}} d_{\mathrm{H}}(s_i, s_j) \Pr\{\hat{X} = s_j | X = s_i\}, \tag{2.9}$$

where $d_{\mathrm{H}}(s_i, s_j)$ is the Hamming distance, i.e., the number of different bits between the labels of the constellation points. A UB for the BER can be obtained from (2.9) by replacing the probability with the PEP, i.e.,

$$\mathrm{BER}_{\mathrm{UB}} \triangleq \frac{1}{mM} \sum_{s_i \in \mathcal{S}} \sum_{\substack{s_j \in \mathcal{S} \\ s_j \neq s_i}} d_{\mathrm{H}}(s_i, s_j) \mathrm{Q}\left(\sqrt{\frac{d^2(s_i, s_j)}{2N_0}}\right). \tag{2.10}$$

2.2.2 Bit-Wise Demodulator

In a coded system such as BICM, soft information on the received (coded) bits is more relevant as this information can be used for further soft decoding. To obtain such soft information, L-values (also known as log-likelihood ratios (LLRs)) are usually calculated in practice. The exact L-value for the jth bit based on the observation Y can be expressed as

$$L_j(Y) \triangleq \log \frac{\Pr\{B_j = 1 | Y\}}{\Pr\{B_j = 0 | Y\}} = \log \frac{\sum_{s \in \mathcal{S}_{j,1}} \mathrm{e}^{-\gamma(Y-s)^2}}{\sum_{s \in \mathcal{S}_{j,0}} \mathrm{e}^{-\gamma(Y-s)^2}}, \tag{2.11}$$

where $j = 1, \ldots, m$ and $\mathcal{S}_{j,u} \subset \mathcal{S}$ is a subconstellation of points whose labels have bit u in the jth bit position. The second equality follows from Bayes' rule used together with the i.i.d. assumption of the bits and the conditional PDF in (2.2).

A bit-wise demodulator (BD) uses the L-values in (2.11) to make a decision on the received bit according to the rule

$$\hat{B}_j^{\mathrm{BD}} = \begin{cases} 1 & \text{if } \ L_j(Y) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

The BD implements an ML bit-wise demodulation and minimizes the BER. The uncoded performance of such a demodulator was studied in [31]. Among other results, closed-form expressions for the BER for 4-ary PAM constellation labeled with the binary reflected Gray code (BRGC) [28, 32, 33] were presented.

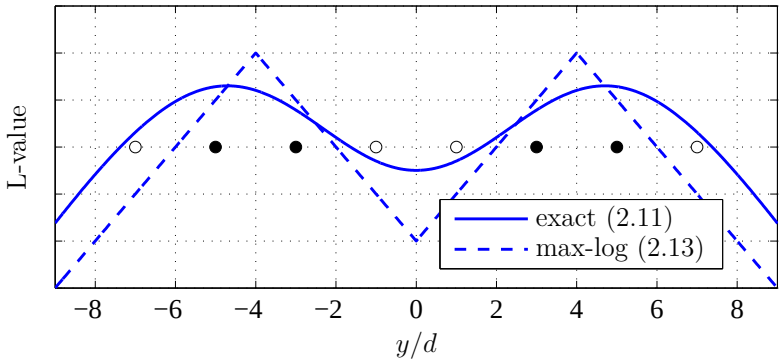


Figure 2.2: Exact and max-log L-values as functions of the (normalized) observation y for the third bit position of 8-PAM with the BRGC for $\gamma = 0$ dB. Empty and filled circles show constellation points labeled with 0 and 1, respectively.

The calculation of the L-values in their exact form (2.11) is complicated, especially for large constellations, as it requires calculation of the logarithm of a sum of exponentials. To overcome this problem, approximations are usually used in practice. The most common approximation is the so-called max-log approximation ($\log \sum_i \mathrm{e}^{\lambda_i} \approx \max_i \lambda_i$) [12, eq. (3.2)], [34, eq. (9)], [35, eq. (5)], [36, eq. (8)], which used in (2.11) gives

$$\tilde{L}_j(Y) \triangleq \gamma \left[\min_{x \in \mathcal{S}_{j,0}} (Y - x)^2 - \min_{x \in \mathcal{S}_{j,1}} (Y - x)^2 \right]. \tag{2.13}$$

The use of the max-log approximation transforms the nonlinear relationship (2.11) into a piece-wise linear relationship (2.13), as previously shown in, e.g., [37, Fig. 3], [38, eqs. (11)–(14)]. The exact and the max-log L-values are shown in Fig. 2.2. The exact L-value approaches the max-log L-value when the SNR increases. The demodulator that uses the decision rule (2.12) based on L-values calculated by (2.13) is called the approximated bit-wise demodulator (ABD).

2.3 Labelings and Patterns

Throughout Part I of this thesis, we will concentrate on the 8-PAM constellation with two particular labelings. The first one is the BRGC, which minimizes the uncoded BER at high SNR and is widely used in noniterative BICM schemes. The second one is the natural binary code (NBC), which follows Ungerboeck's SP principle, and, hence, is popular in TCM schemes. These labelings are given by the matrices

$$\mathbb{C}_{\mathrm{BRGC}} \triangleq \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}^{\mathsf{T}}, \quad \mathbb{C}_{\mathrm{NBC}} \triangleq \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}^{\mathsf{T}}. \tag{2.14}$$

In terms of their $\boldsymbol{q}$ vectors, the BRGC can be written as $\boldsymbol{q}_{\mathrm{BRGC}} = [0, 1, 3, 2, 6, 7, 5, 4]$ and the NBC as $\boldsymbol{q}_{\mathrm{NBC}} = [0, 1, 2, 3, 4, 5, 6, 7]$.

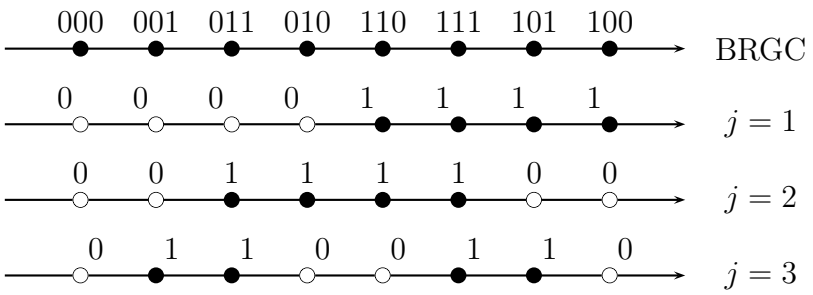


Figure 2.3: Subconstellations for 8-PAM with the BRGC. Subconstellations $\mathcal{S}_{j,0}$ are shown with empty circles and $\mathcal{S}_{j,1}$ are shown with filled circles.

It follows from (2.11) and (2.13) that the L-values depend on the subconstellations $\mathcal{S}_{j,u}$, which are determined by the bits in the jth column of the labeling matrix. An example of subconstellations for 8-PAM with the BRGC is shown in Fig. 2.3. When studying different properties of higher-order modulation, such as uncoded BER or bit-level MI, it can be convenient to consider the columns of the labeling matrix separately. We refer to these columns as *patterns*, which are formally defined below.

We define a bit pattern as a length-M binary vector $\boldsymbol{p} = [p_1, \ldots, p_M] \in \{0,1\}^M$ with $M/2$ ones and $M/2$ zeros. The labeling $\mathbb{C}$ can now be defined by m patterns, each corresponding to one column of $\mathbb{C}$. The patterns are indexed as $\boldsymbol{p}_w$ with w being the decimal representation of the vector $\boldsymbol{p}$, i.e., $w = \sum_{i=1}^{M} 2^{M-i} p_i$. The set of m indices of patterns in the labeling $\mathbb{C}$ is denoted by $\mathcal{W}$. The BRGC can be written in terms of patterns as $\mathbb{C}_{\text{BRGC}} = [\boldsymbol{p}_{15}^{\mathsf{T}}, \boldsymbol{p}_{60}^{\mathsf{T}}, \boldsymbol{p}_{102}^{\mathsf{T}}]$, where

$$\begin{aligned}\boldsymbol{p}_{15} &= [0,0,0,0,1,1,1,1],\\ \boldsymbol{p}_{60} &= [0,0,1,1,1,1,0,0],\\ \boldsymbol{p}_{102} &= [0,1,1,0,0,1,1,0],\end{aligned}$$

and $\mathcal{W}_{\text{BRGC}} = \{15, 60, 102\}$. The NBC can be represented as $\mathbb{C}_{\text{NBC}} = [\boldsymbol{p}_{15}^{\mathsf{T}}, \boldsymbol{p}_{51}^{\mathsf{T}}, \boldsymbol{p}_{85}^{\mathsf{T}}]$ with $\mathcal{W}_{\text{NBC}} = \{15, 51, 85\}$ and

$$\begin{aligned}\boldsymbol{p}_{15} &= [0,0,0,0,1,1,1,1],\\ \boldsymbol{p}_{51} &= [0,0,1,1,0,0,1,1],\\ \boldsymbol{p}_{85} &= [0,1,0,1,0,1,0,1].\end{aligned}$$

The number of different patterns for equally spaced one-dimensional constellations has been recently analyzed in [39] and [Paper A]. We note that not all combinations of patterns can form a valid labeling, as all rows of $\mathbb{C}$ must be different. In the following section, we discuss how to calculate the BER for a pattern (PBER). Once the PBERs are known for all patterns in $\mathcal{W}$, the BER for the labeling can be calculated as

$$\text{BER} = \frac{1}{m} \sum_{w \in \mathcal{W}} \text{PBER}_w. \tag{2.15}$$

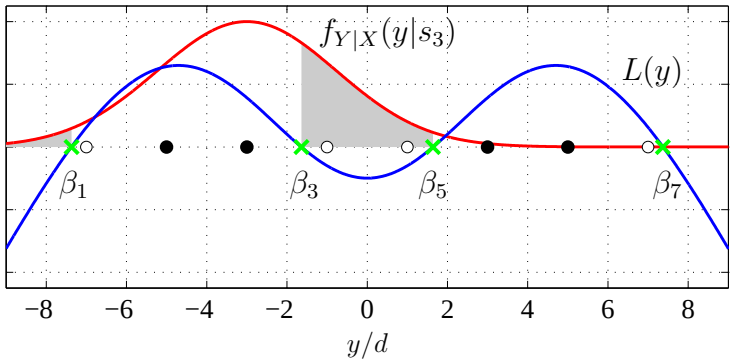


Figure 2.4: Exact L-value as functions of the (normalized) observation y for the pattern $\boldsymbol{p}_{102}$ for $\gamma = 0$ dB and the decision thresholds together with the conditional PDF $f_{Y|X}(y|s_3)$ (the PDF is normalized for illustration purposes). Empty and filled circles show constellation points labeled with 0 and 1, respectively. The gray area shows the probability of bit error if s_3 is transmitted.

2.4 BER Performance of Bit-Wise Demodulators

One way to analyze the performance of the BD is to use the PDFs of the L-values, as was proposed in [40,41]. While PDFs can be analytically calculated for max-log L-values, analytical expressions for the PDFs of exact L-values are unknown for $m > 1$. Another way to evaluate the performance of the BD was used in [31] where the error probability was formulated in terms of decision thresholds. Fig. 2.4 demonstrates how the PBER can be calculated using this approach for the pattern $\boldsymbol{p}_{102}$. The decision thresholds denoted by β_k, $k \in \mathcal{K}$ are the points where the L-value function crosses the zero-level, i.e., $L(Y) = 0$. The set $\mathcal{K}$ is defined as $\mathcal{K} = \{k = 1, \ldots, M-1 : p_k \neq p_{k+1}\}$. The thresholds are shown with crosses in Fig. 2.4. Given a particular transmitted symbol, the probability of error can be calculated as the integral of the conditional PDF over the observation region where the L-value has the wrong sign.

It was shown in [Paper A] that for an arbitrary one-dimensional constellation with a pattern $\boldsymbol{p}$, the PBER can be expressed as

$$\text{PBER} = \frac{1}{2} + \frac{1}{M} \sum_{i=1}^{M} \sum_{k \in \mathcal{K}} g_{i,k} \text{Q}\left((\beta_k - s_i)\sqrt{2\gamma} \right), \tag{2.16}$$

where $g_{i,k} \in \{\pm 1\}$ is

$$g_{i,k} \triangleq (p_{k+1} - p_k)(1 - 2p_i). \tag{2.17}$$

For max-log L-values, the decision thresholds are the midpoints between the constellation points labeled with different bits, i.e., they are independent of the SNR and coincide with the thresholds for the SD. The PBER in (2.16) is then equivalent to (2.9), where the Hamming distance is calculated for only one bit position. The ABD is therefore equivalent to the SD in terms of uncoded BER.

The problem of finding thresholds for exact L-values and equally spaced one-dimensional constellations was addressed in [Paper B], where closed form expressions for these thresholds were found for all labelings for 4-PAM and the most popular labelings for 8-PAM. As

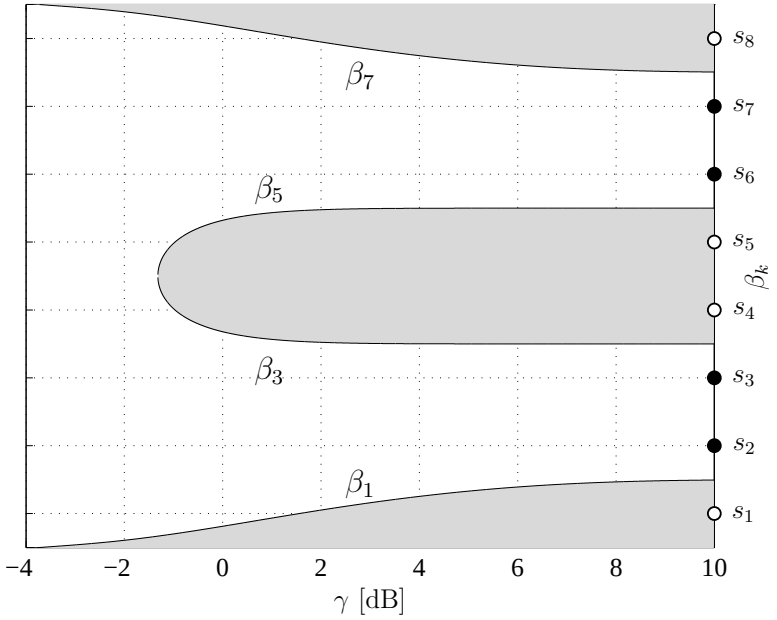


Figure 2.5: Threshold for the pattern $\boldsymbol{p}_{102}$. Empty and filled circles show constellation points labeled with 0 and 1, respectively. Gray and white areas indicate the decision regions for 0 and 1, respectively.

an example, the thresholds for the pattern $\boldsymbol{p}_{102}$ are shown in Fig. 2.5, where it is clear that the thresholds depend on the SNR. Moreover, at low SNR the thresholds β_3 and β_5 disappear and become *virtual* [Paper C]. In [Paper C, Theorem 2] we show how to choose the values of virtual thresholds in order to use them in (2.16). For high SNR, the thresholds approach the midpoints, i.e., coincide with the thresholds for max-log L-values. Numerical results confirm that the BD outperforms the SD. However, for any BER of practical interest (below 0.1), the difference between the demodulators is negligible (see [Paper A, Fig. 2] and [Paper B, Figs. 4 and 5]).

Chapter 3

Practical Approaches to Coded Modulation

3.1 Trellis-Coded Modulation

TCM introduced in [8, 9] was one of the first practical CM schemes with reasonably good performance. It was rather quickly implemented in modem standards [42] in the beginning of the 90s. Since then, coding theory has made a huge progress and TCM is not widely used in modern (especially wireless) communication systems nowadays. However, it gives insights into how a binary code and modulation interact. It is also a good example of a coded scheme where the optimal ML decoder can be implemented with affordable complexity. These are the reasons why TCM is considered in this thesis.

The system model of a TCM scheme is shown in Fig. 3.1. A length-K vector of information bits $\boldsymbol{c} = [c[1], \ldots, c[K]] \in \{0,1\}^K$ is fed to a convolutional encoder[2], which produces a vector of coded bits $\boldsymbol{b} = [b[1], \ldots, b[N_{\mathcal{B}}]] \in \{0,1\}^{N_{\mathcal{B}}}$ of length $N_{\mathcal{B}}$. All information vectors are assumed to be equiprobable. All possible codewords form a binary code $\mathcal{B}$. The rate of the CC is $R_{\mathcal{B}} = \frac{K}{N_{\mathcal{B}}}$. The encoder is defined by a generator matrix $\mathbb{G}$ as in [43, Ch. 4], where the elements of the generator matrix are polynomials over the binary field given in octal form.

The coded bits are fed to the modulator, defined in the previous chapter, which outputs a vector of symbols $\boldsymbol{x} = [x[1], \ldots, x[N]] \in \mathcal{S}^N$ of length N. The vector $\boldsymbol{x}$ is called a TCM codeword. All possible TCM codewords form a TCM code $\mathcal{X}$. The overall rate (or spectral efficiency) is therefore $R = \frac{K}{N}$ and defines how many information bits are transmitted per channel use. The equality $N_{\mathcal{B}} = mN$ should hold to match the lengths of the vectors. The rate of the CC can be then expressed as $R_{\mathcal{B}} = \frac{K}{mN}$.

Examining the capacity curves in Fig. 1.3, Ungerboeck suggested to use CCs with a rate $R_{\mathcal{B}} = \frac{m-1}{m}$. For such a rate, a spectral efficiency of $m-1$ bpcu can be achieved with a 2^m-point constellation at SNRs very close to the channel capacity. For instance, a spectral efficiency of 2 bpcu can be achieved with 8-PAM and a rate-2/3 code at around 10 dB. Increasing the constellation cardinality further does not give any significantly noticeable gain for the desired spectral efficiency.

The vector $\boldsymbol{x}$ is transmitted over the memoryless AWGN channel, where the output of the channel is given by $\boldsymbol{Y} = \boldsymbol{x} + \boldsymbol{Z}$ and $\boldsymbol{Z} = [Z[1], \ldots, Z[N]]$ is a vector of i.i.d. zero-mean Gaussian random variables. CCs allow a reasonably simple implementation of the ML decoder by means of the Viterbi algorithm [10]. The choice of the rate $R_{\mathcal{B}} = \frac{m-1}{m}$ facilitates the decoding because m coded bits on each trellis section of the CC are replaced by one

[2] We will consider only feedforward encoders.

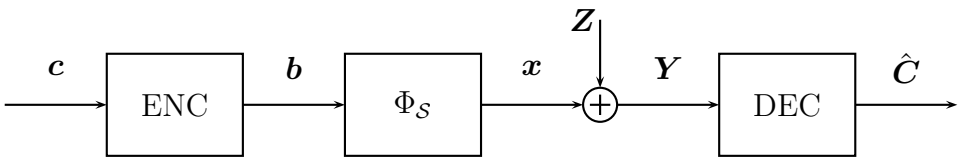


Figure 3.1: TCM system model

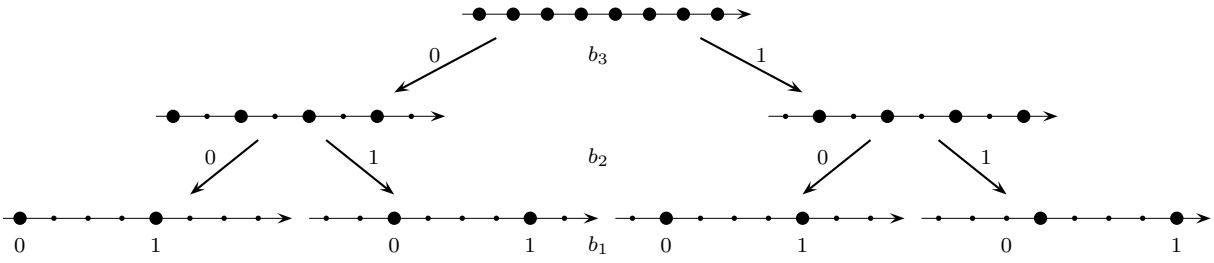


Figure 3.2: Set partitioning for 8-PAM that results in the NBC labeling.

symbol. Other rates could be implemented by using multidimensional TCM schemes [44], where one trellis section of the code comprises several symbols. The decoder will be formally described in the next chapter, where symbol-wise and bit-wise decoders are compared.

As pointed out by Ungerboeck, the figure of merit for a TCM system is the EDs between the codewords of the TCM code rather than the Hamming distance between the codewords of the underlying CC. For example, the free Hamming distance d_{free}, i.e., the minimum Hamming weight of a nonzero codeword, is an important property of a CC. Its TCM counterpart is the minimum ED between two codewords of a TCM code. Even though these two quantities are a measure of distance between codewords, there is no explicit relation between them.

Ungerboeck also suggested a set of rules to design good TCM schemes. The key element in his design is the SP technique for constructing labelings for constellations. The main idea is to partition the constellation into subconstellations with nondecreasing minimum ED between the subconstellation points. The SP procedure for the 8-PAM constellation is shown in Fig. 3.2. It is easy to see that for this constellation, each step of SP doubles the minimum ED. After appropriately labeling all the points, we can verify that the NBC defined in (2.14) follows the SP principle.

Consider a TCM code $\mathcal{X}$ obtained by concatenating a labeling $\mathbb{C}$ with a CC $\mathcal{B}$. Recent results in [45] showed that the same TCM encoder, i.e., the same TCM code $\mathcal{X}$ and the same mapping between the information bits and the codewords, can be implemented by using other labelings with properly modified binary encoders. This is not surprising since the labeling is not important once the trellis structure of the code is fixed and the symbols are assigned to the trellis branches. As [45] reveals, different labelings can be grouped into equivalence classes. Any labeling $\mathbb{C}'$ within the same equivalence class as $\mathbb{C}$, together with a properly modified CC of the same memory, can be used to obtain exactly the same TCM

encoder. We denote the new binary code by $\mathcal{B}'$. It turns out that for one-dimensional constellations, the SP (or NBC) and the BRGC belong to the same class.

As mentioned earlier, the labeling is not important for TCM and any other labeling outside the equivalence class can be used as well. However, the binary code $\mathcal{B}'$ in this case may no longer be a CC (for instance, it may not contain the all-zero codeword and may therefore be nonlinear).

As an example of equivalent TCM schemes, we consider 8-PAM labeled with the NBC and a rate-2/3 CC defined by the generator matrix [43, Ch. 4]

$$\mathbb{G} = \begin{bmatrix} g_{1,1} & g_{1,2} & g_{1,3} \\ g_{2,1} & g_{2,2} & g_{2,3} \end{bmatrix}.$$

By looking at the patterns for the two labelings in (2.14), it is easy to establish the relations

$$\boldsymbol{p}_{60} = \boldsymbol{p}_{51} \oplus \boldsymbol{p}_{15}, \tag{3.1}$$

$$\boldsymbol{p}_{102} = \boldsymbol{p}_{85} \oplus \boldsymbol{p}_{51}, \tag{3.2}$$

i.e., all bit patterns in the BRGC can be obtained from the patterns of the NBC. To obtain the same TCM scheme when switching labelings, the new generator matrix $\mathbb{G}'$ is given by

$$\mathbb{G}' = \begin{bmatrix} g_{1,1} & g_{1,2} + g_{1,1} & g_{1,3} + g_{1,2} \\ g_{2,1} & g_{2,2} + g_{2,1} & g_{2,3} + g_{2,2} \end{bmatrix}, \tag{3.3}$$

where addition is performed for the corresponding polynomials over the binary field.

Let us consider an example of a TCM scheme with the CC defined by $\mathbb{G}_{\text{NBC}} = [7, 6, 2; 7, 3, 0]$ and the NBC labeling $\mathbb{C}_{\text{NBC}}$. We note that this TCM scheme is not optimal in terms of minimum ED and also does not posses the structure devised by Ungerboeck, where the most protected bit position is uncoded. However, this example makes the performance analysis for TCM and BICM intuitive and illustrative. After applying the linear operations on the generator polynomials, we conclude that the same TCM scheme can be implemented by using the BRGC and the CC with the generator matrix

$$\mathbb{G}_{\text{BRGC}} = \begin{bmatrix} 7 & 1 & 4 \\ 7 & 4 & 3 \end{bmatrix}. \tag{3.4}$$

We denote this configuration by the tuple $(\mathbb{C}_{\text{BRGC}}, \mathbb{G}_{\text{BRGC}})$ and it will be used later on for illustration purposes. The steps of this "transformation" are schematically shown in [45, Fig. 1].

We repeated the same procedure for 4-PAM with the NBC and the rate-1/2 codes specified in [9, Table I] up to memory $\nu = 9$. The results are summarized in Table 3.1, where the generator matrices together with the corresponding free Hamming distances d_{free} are shown. The second column shows the maximum free distance (MFD) for CCs of a given memory [46]. An interesting observation is that codes that provide good TCM schemes when used with the BRGC have in general a large free Hamming distance d_{free}.

Table 3.1: Convolutional encoders for equivalent TCM schemes. $\mathbb{G} = [g_1, g_2]$ corresponds to the encoders for the NBC, whereas $\mathbb{G}' = [g_1, g_1 + g_2]$ is the encoder for the BRGC. The polynomials are given in octal form. The free Hamming distances for the codes $\mathbb{G}$ and $\mathbb{G}'$ are denoted by d_{free} and d'_{free}, respectively. The second column shows the MFD for a given code memory ν.

ν	MFD	$\mathbb{G}$	d_{free}	$\mathbb{G}'$	d'_{free}
2	5	[5, 2]	3	[5, 7]	5
3	6	[13, 4]	4	[13, 17]	6
4	7	[23, 4]	4	[23, 27]	7
5	8	[45, 10]	4	[45, 55]	7
6	10	[103, 24]	5	[103, 127]	8
7	10	[235, 126]	8	[235, 313]	10
8	12	[515, 362]	8	[515, 677]	12
9	12	[1017, 342]	8	[1017, 1355]	12

In fact, most of the codes have the maximum possible free Hamming distance d_{free}. When the BRGC labeling is used, this result is intuitively expected as a code with a large free Hamming distance d_{free} is likely to result in a TCM code with a large minimum ED. However, the number of different bits (Hamming distance) between the codewords is not the only parameter to define the ED. The way different bits between the codewords are grouped also plays an important role and sometimes it could be beneficial to use a code with a slightly smaller free Hamming distance d_{free} but better "bit grouping". This is the case for the codes with $\nu = 5, 6$ in Table 3.1. Good and bad groupings will be clear when we discuss the performance of TCM and BICM in the next sections.

3.1.1 Performance Analysis

The performance analysis of uncoded transmission in Sec. 2.2.1 can be easily extended to the coded case by replacing the constellation symbols by TCM codewords in (2.5)–(2.9). For instance, the PEP for the coded case is given by

$$\text{PEP}\{\hat{\boldsymbol{X}} = \boldsymbol{x}_j | \boldsymbol{X} = \boldsymbol{x}_i\} = \text{Q}\left(\sqrt{\frac{d^2(\boldsymbol{x}_i, \boldsymbol{x}_j)}{2N_0}}\right), \tag{3.5}$$

where $d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$. The counterpart of the SER in the coded case is called frame error rate (FER). Exact FER and BER are usually too complex to calculate and union bounds are considered in most cases. The union bound for the BER is given by

$$\text{BER}_{\text{UB}} = \frac{1}{K2^K} \sum_{\boldsymbol{x}_i \in \mathcal{X}} \sum_{\substack{\boldsymbol{x}_j \in \mathcal{X}, \\ \boldsymbol{x}_j \neq \boldsymbol{x}_i}} d_{\text{H}}(\boldsymbol{c}_i, \boldsymbol{c}_j) \text{Q}\left(\sqrt{\frac{d^2(\boldsymbol{x}_i, \boldsymbol{x}_j)}{2N_0}}\right), \tag{3.6}$$

where $\boldsymbol{c}_i$ is the information vector assigned to the codewords $\boldsymbol{x}_i$. This expression can be seen as a "coded" version of (2.10) if the constellation symbols are replaced by the codewords of $\mathcal{X}$. By combining Q-functions with the same arguments, (3.6) can be rewritten as in [45, eq. (22)] in terms of the distances $\mathcal{D} = \{d(\boldsymbol{x}_i, \boldsymbol{x}_j) : \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{X}, \boldsymbol{x}_i \neq \boldsymbol{x}_j\}$, i.e.,

$$\mathrm{BER}_{\mathrm{UB}} = \sum_{u \in \mathcal{D}} B_u \mathrm{Q}\left(\sqrt{\frac{u^2}{2N_0}}\right), \tag{3.7}$$

where B_u is the bit multiplicity for the distance $u \in \mathcal{D}$. The distances in $\mathcal{D}$ with the corresponding multiplicities form the so-called distance spectrum.

CM codes in general and TCM codes in particular are nonlinear codes, i.e, the sum of two codewords does not have to be another codeword. This makes the performance analysis of TCM more complicated compared to linear codes, where a performance analysis can be based on the assumption that the all-zero codeword is transmitted. In the case of TCM, the distance spectrum cannot be obtained based on a state machine as for binary CC [23, Ch. 8].

For high SNR, the performance will be dominated by pairs of codewords at minimum ED. We will illustrate the high-SNR analysis for the setup $(\mathbb{C}_{\mathrm{BRGC}}, \mathbb{G}_{\mathrm{BRGC}})$. The free Hamming distance of the code $\mathbb{G}_{\mathrm{BRGC}}$ is $d_{\mathrm{free}} = 5$. We can identify binary codewords that give two TCM codewords at minimum ED. The binary codewords are

$$\begin{aligned} \boldsymbol{b} &= [0,0,0,0,0,0,0,0,0\ldots], \\ \hat{\boldsymbol{b}} &= [0,1,1,0,0,1,0,1,1\ldots], \end{aligned} \tag{3.8}$$

where dots represent some arbitrary common bits of the two codewords. For simplicity, we assume these bits are zero. Using Fig. 2.3, we conclude that the corresponding TCM codewords are

$$\begin{aligned} \boldsymbol{x} &= [s_1, s_1, s_1, \ldots], \\ \hat{\boldsymbol{x}} &= [s_3, s_2, s_3, \ldots], \end{aligned} \tag{3.9}$$

where dots represent some common symbols. The ED between the codewords $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ can be calculated as

$$d^2(\boldsymbol{x}, \hat{\boldsymbol{x}}) = 2d^2(s_1, s_3) + d^2(s_1, s_2) = 36d^2.$$

The Hamming distance between the codewords happens to be 5, i.e., it is equal to the free Hamming distance of the code. We note that this is not always the case and binary codewords at free Hamming distance may result in TCM codewords with the ED larger than the minimum ED of the code.

For the code $\mathbb{G}_{\mathrm{BRGC}}$, the output distribution of the symbols is uniform[3] and this allows to relate the average symbol energy to the constellation parameter d using (2.1)

$$E_{\mathrm{s}} = 21d^2.$$

The high-SNR approximation of the BER can thus be written as

$$\mathrm{BER} \approx \mathrm{Q}\left(\sqrt{\frac{18d^2}{N_0}}\right) = \mathrm{Q}\left(\sqrt{\frac{18}{21}\gamma}\right). \tag{3.10}$$

[3]This is not generally true for an arbitrary CC.

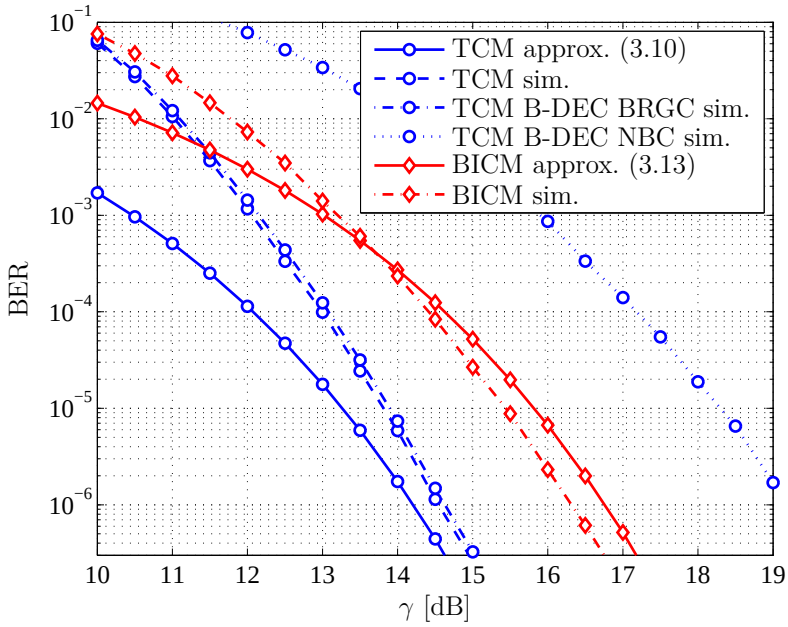


Figure 3.3: The BER performance for a coding scheme $(\mathbb{C}_{\text{BRGC}}, \mathbb{G}_{\text{BRGC}})$ without and with interleaver (see Sec. 3.2).

The high-SNR approximation in (3.10) is shown in Fig. 3.3 with the solid line marked with circles. The dashed line marked with circles shows the simulation results for the TCM scheme $(\mathbb{C}_{\text{BRGC}}, \mathbb{G}_{\text{BRGC}})$ and demonstrates a good agreement with the theoretical prediction at high SNR. A better agreement could be achieved if more terms of the distance spectrum with their multiplicities are taken into account according to (3.7). The figure also shows results for a BICM system, described in Sec. 3.2.1, and the performance curves for bit-wise decoders (B-DECs), which will be discussed in Sec. 4.2.

3.2 Bit-Interleaved Coded Modulation

BICM was proposed in [12] as a CM scheme for fast fading channels. Two new elements compared to TCM were introduced in a BICM system by Zehavi in [12], i.e., the bit-wise interleaver at the transmitter and the bit-wise decoder at the receiver. It is not always clear which one is an intrinsic part of a BICM system. A system without an interleaver but with the bit-wise decoder may still be called BICM, for instance, when a CM scheme uses LDPC codes [47]. We begin with describing a classical scheme where both elements are present, highlighting the differences between BICM and TCM schemes.

The system model of a BICM scheme is shown in Fig. 3.4. After the information vector $\boldsymbol{c}$ is encoded into a vector of coded bits $\boldsymbol{b}$, a bit-wise interleaver Π produces a permuted version of the coded bits $\boldsymbol{b}^{\pi}$. The modulator maps the interleaved bits to symbols and sends the vector of the symbols to the channel. At the receiver, the demapper (DEM)[4] calculates the L-values $\boldsymbol{L}^{\pi}$ for the coded bits using (2.11) or (2.13). After deinterleaving,

[4]A bit-wise demodulator defined in Sec. 2.2.2 calculates the L-values and makes hard decisions on them, whereas the demapper only calculates the L-values.

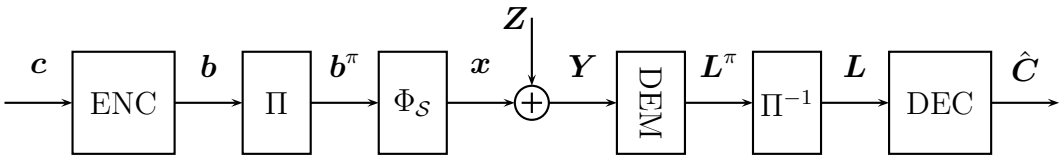


Figure 3.4: BICM system model.

the L-values $\boldsymbol{L}$ are passed to a bit-wise decoder that maximizes the correlation between the coded bits and the L-values [34, eq. (9)]. When the binary code is a CC, such a decoder can be implemented using a soft-input Viterbi decoder for the binary code. The decoder will be formally defined and analyzed in the next chapter. Here, we will try to gain some intuition on how the interleaver affects the performance.

The key assumption when analyzing BICM is that the L-values are independent conditioned on a given codeword [48, Sec. 4.3.2]. This is assumed to be achieved by using an ideal bit interleaver, that is, an interleaver that has an infinite length and is "completely" random [34]. If two coded bits of a codeword are transmitted within one symbol, the independence assumption is clearly not correct, since the L-values are calculated from an observation affected by the same noise realization. To make the L-values conditionally independent, all bits of a codeword need to be transmitted over different symbols. This essentially implies that ideal interleaving has to be performed over several codewords. The following interleaving strategy guarantees the conditional independence of the L-values.

For the sake of illustration, we set $m = 3$ and assume that three codewords $\boldsymbol{b}_i$, $i = 1, 2, 3$ are jointly interleaved. In accordance with Fig. 3.4, the interleaved codewords $\boldsymbol{b}_i^\pi$ are obtained by interleaving each of the codewords $\boldsymbol{b}_i$ using the interleaver Π. We note that at this point interleaving is performed within one codeword and the interleaver may be "completely" random, may have a certain structure (e.g., an M-interleaver in [49]), or it may be deterministic (for example, no interleaver). We divide each codeword into m subcodewords $\boldsymbol{b}_i^\pi = [\boldsymbol{b}_{i,1}^\pi, \boldsymbol{b}_{i,2}^\pi, \boldsymbol{b}_{i,3}^\pi]$ where $\boldsymbol{b}_{i,j}^\pi$ are the bits of the codeword $\boldsymbol{b}_i^\pi$ that are supposed to be transmitted on the jth bit position. The rest of the transmission is schematically shown in Fig. 3.5 and explained below.

According to the model in Fig. 3.4, each of the interleaved codewords $\boldsymbol{b}_i^\pi$, represented by the ith column of the top-left matrix, is mapped to $\boldsymbol{x}_i$. At the receiver, the L-values $\boldsymbol{L}_i^\pi$, represented by the ith column of the top-right matrix, are calculated. As explained earlier, for a chosen i, the elements of $\boldsymbol{L}_i^\pi$, and hence, the elements of $\boldsymbol{L}_i$ are not conditionally independent.

To make the L-values conditionally independent without changing their individual properties, we can perform an inter-codeword interleaving as shown in Fig. 3.5 in the bottom-left matrix, where interleaving is done across the rows to produce $\tilde{\boldsymbol{b}}_i^\pi$. At the receiver, deinterleaving is done to produce vectors of L-values $\boldsymbol{L}_i^\pi$. Since all bits of one codeword are sent over different symbols, the L-values $\boldsymbol{L}_i^\pi$, as well as $\boldsymbol{L}_i$, within one codeword will be conditionally independent. Such an interleaver was studied in [50]. Of course, several interleaved codewords can be viewed as a codeword of a new longer code. When decoding the new code, the L-values are no longer conditionally independent. Even though the ideal interleaver is an abstraction for the scheme in Fig. 3.4, the ideal interleaving assumption gives very precise performance estimations for CCs. This is due

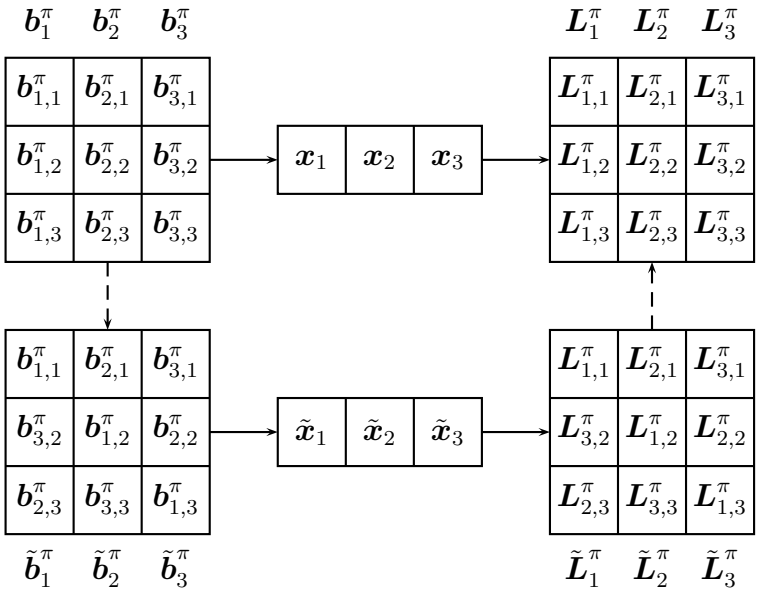


Figure 3.5: An implementation of the interleaver that guarantees the independence of the L-values.

to the fact that a long CC to a large degree can be seen as a concatenation of multiple shorter codes.

The infinite interleaver assumption does not comply with the system model in Fig. 3.4 since it accounts for only one codeword. A more suitable channel model in the case of ideal interleaving was proposed in [34, Fig. 3], where the channel is represented by a set of m parallel independent channels corresponding to m bit positions in the constellation. Many authors studied the MI of such a channel, e.g., [34, 51, 52]. In Fig. 3.6, the MI for 8-PAM labeled with the BRGC and the NBC is shown under the assumption of uniformly distributed input bits. We will refer to this quantity as BICM MI[5]. The figure demonstrates that the BICM MI strongly depends on the labeling. It is commonly assumed that the BRGC labeling maximizes the BICM MI for moderate and high SNR. The asymptotic high SNR optimality of the BRGC was shown in [53]. This is the main theoretical justification of the use of the BRGC in virtually all BICM systems.

More recent studies [54, 55] based on the framework of mismatched decoding and generalized MI showed that the ideal interleaving assumption is not necessary for achieving the rates predicted by the BICM MI. The maximum rate that can be achieved with a bit-wise decoder is still an open question.

The performance of a BICM system is highly dependent on the distributions of the L-values. It is easy to verify that these distributions are different for different bit positions. Moreover, the distribution of the L-values in a certain bit position may depend on whether zero or one was transmitted. This complicates the analysis and some simplifications are usually made. In [34], a time-varying labeling was proposed in order to symmetrize the channel. The same effect can be achieved by using a random scrambler as in [56] or [57]. Even though a BICM system in Fig. 3.4 does not contain any scrambler, the symmetrization technique allows us to obtain very precise performance estimations. We

[5]It is also commonly called BICM capacity or BICM generalized MI.

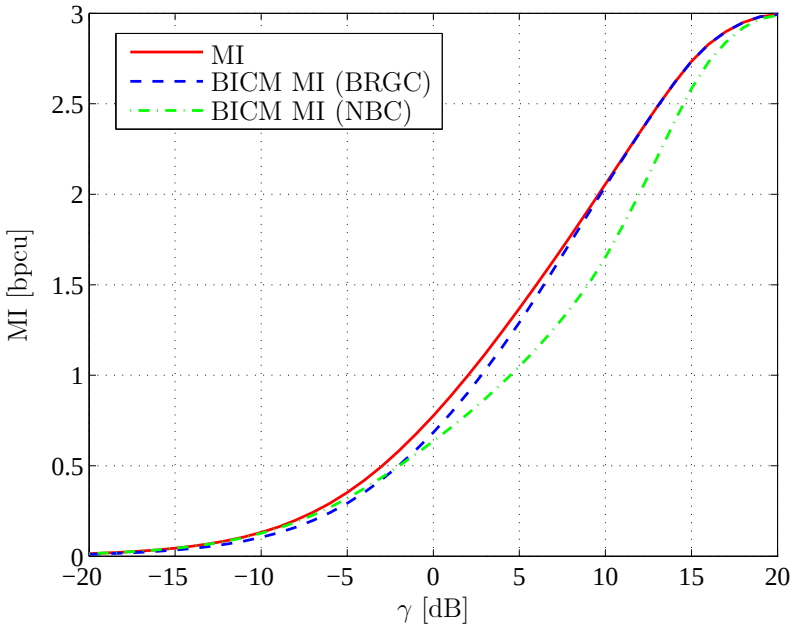


Figure 3.6: BICM MI for uniform input bits for the BRGC and the NBC labelings. The solid line shows the 8-PAM MI for uniform input symbols.

will comment on this later on in Sec. 4.2.

3.2.1 BICM with Convolutional Codes

In this section, we try to understand the effects that the interleaver causes under a hypothetical ML decoder by considering a simple example. Note that this "analysis" is far from being rigorous and complete but it explains many phenomena of BICM on an intuitive level. Assume the setup $(\mathbb{C}_{\mathrm{BRGC}}, \mathbb{G}_{\mathrm{BRGC}})$ and the pair of binary codewords in (3.8). The binary code is assumed to be sufficiently long and the interleaver matches the length of the code, which corresponds to Fig. 3.4. The number of ones in $\hat{\boldsymbol{b}}$ is 5, which corresponds to the free Hamming distance of $\mathbb{G}_{\mathrm{BRGC}}$. Since the number of zeros is much larger than the number of ones, a vector of interleaved bits $\hat{\boldsymbol{b}}^{\pi}$ is very likely to have a structure, where each triplet of bits contains at most one nonzero bit. For instance, after interleaving, the two codewords may look like

$$\begin{aligned}\boldsymbol{b}^{\pi} &= [\ldots, 0, 0, 0, \ldots, 0, 0, 0, \ldots, 0, 0, 0, \ldots, 0, 0, 0, \ldots, 0, 0, 0, \ldots],\\ \hat{\boldsymbol{b}}^{\pi} &= [\ldots, 0, 0, 1, \ldots, 0, 0, 1, \ldots, 0, 0, 1, \ldots, 0, 0, 1, \ldots, 0, 0, 1, \ldots].\end{aligned} \tag{3.11}$$

After the modulator, the corresponding CM codewords will have the structure

$$\begin{aligned}\boldsymbol{x} &= [\ldots, s_1, \ldots, s_1, \ldots, s_1, \ldots, s_1, \ldots, s_1, \ldots],\\ \hat{\boldsymbol{x}} &= [\ldots, s_2, \ldots, s_2, \ldots, s_2, \ldots, s_2, \ldots, s_2, \ldots].\end{aligned} \tag{3.12}$$

As we can see, the number of different symbols is equal to the Hamming distance between the binary codewords. For fast fading channels, the minimum number of symbols between all pairs of codewords is the so-called code diversity. Comparing the codewords in (3.9)

with (3.12), we conclude that the interleaver increases the code diversity and makes it equal to the free Hamming distance of the code. This was the original motivation behind the BICM scheme proposed in [12] for fading channels. We will however concentrate on the analysis for the AWGN channel.

Obviously, there are many other possible interleaved versions of the codewords that have the same structure (i.e., each triplet has at most one nonzero bit), but the pair shown in (3.11) is special. The squared distance between the codewords in (3.12) is $d^2(\boldsymbol{x}, \hat{\boldsymbol{x}}) = 5d^2(s_1, s_2) = 5(2d)^2 = 20d^2$. It is easy to verify that, if the BRGC is used, this is the minimum ED distance among all possible interleaved versions of all codewords, which is in agreement with [34, eq. (64)]. Analogously to TCM, the high-SNR approximation of the BER can be written as

$$\text{BER} \approx \text{Q}\left(\sqrt{\frac{10d^2}{N_0}}\right) = \text{Q}\left(\sqrt{\frac{10}{21}\gamma}\right) \tag{3.13}$$

and it is shown in Fig. 3.3 with the solid line marked with diamonds. The dashed-dotted line marked with diamonds shows the simulation results for the BICM scheme $(\mathbb{C}_{\text{BRGC}}, \mathbb{G}_{\text{BRGC}})$ with an interleaver. We note that this rough analysis predicts the performance of the BICM system and should agree well at asymptotically high SNR. A better agreement for medium SNR could be achieved by considering more terms of the distance spectrum with their multiplicities in an expurgated bound type of analysis [34].

Comparing the codewords in (3.9) with (3.12), it can be seen that distributing the bits over different symbols decreases the minimum ED when the BRGC is used. This explains the improved performance of BICM systems with CCs when the interleaver is removed, first observed in [58] and then analyzed in [59], as bit differences between codewords of a CC are always closely situated.

We note that the results for this simple example are based on the assumption of the ML decoder and do not account for the bit-wise decoder which is usually used in practice. In [Paper C], we show that, when the BRGC is used, the performance of the bit-wise decoder in many cases is very similar to that of the optimal ML decoder. This will also be illustrated in Sec. 4.2. This makes the intuition for this simple example hold even when the suboptimal bit-wise decoder is used.

3.2.2 BICM-ID

BICM-ID was originally presented in [21,22] and it was shown to be a powerful technique for both AWGN and fading channels. It was also shown that non-Gray labelings may give better performance when BICM-ID is considered. In BICM-ID, the decoder is allowed to exchange information about the coded bits with the demapper. The resulting iterative decoder is outside the scope of this thesis and it is not discussed at all here. In this short section, we only try to understand what happens when a non-Gray labeling is used in BICM under an ML decoding assumption. For this purpose, we introduce the labeling

$$\mathbb{C}_{\text{AGC}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}^{\intercal}. \tag{3.14}$$

For 8-PAM, this labeling has the largest possible Hamming distance between the closest constellation points. This labeling is called the anti-Gray code (AGC) [22], see also [53]. We further consider the 8-PAM constellation labeled with the AGC. The most important observation is that the closest points differ by either 2 or 3 bits.

We consider the CC given by $\mathbb{G}_{\text{BRGC}}$ in (3.4) and the two codewords in (3.8). As pointed out in the previous section, the most likely vectors of interleaved bits $\hat{\boldsymbol{b}}^{\pi}$ will have a structure where each triplet of bits contains at most one nonzero bit, e.g., as in (3.11). If the AGC is used, the modulated codewords are

$$\begin{aligned}\boldsymbol{x} &= [\ldots, s_2, \ldots, s_2, \ldots, s_2, \ldots, s_2, \ldots, s_2, \ldots],\\ \hat{\boldsymbol{x}} &= [\ldots, s_4, \ldots, s_4, \ldots, s_4, \ldots, s_4, \ldots, s_4, \ldots].\end{aligned} \tag{3.15}$$

The ED distance between these codewords $d^2(\boldsymbol{x}, \hat{\boldsymbol{x}}) = 5d^2(s_2, s_4) = 5(4d)^2 = 80d^2$ is much larger compared to the distance between the codewords in (3.12). Since for the AGC, the closest points cannot differ in one bit, we can conclude that, with very high probability, interleaved binary codewords will result in CM codewords, which on average have larger ED distance than the same codewords would have if the BRGC was used. Note that such codewords will differ in 5 symbols, i.e., the code diversity is not affected when the labeling is changed. We have to note, however, that the interleaver may result in an "unfortunate" pair of codewords[6], e.g.,

$$\begin{aligned}\boldsymbol{b}^{\pi} &= [\ldots, 0, 0, 0, \ldots, 0, 0, 0, \ldots],\\ \hat{\boldsymbol{b}}^{\pi} &= [\ldots, 1, 1, 1, \ldots, 1, 1, 0, \ldots],\end{aligned} \tag{3.16}$$

with the CM codewords of the form

$$\begin{aligned}\boldsymbol{x} &= [\ldots, s_2, \ldots, s_2, \ldots],\\ \hat{\boldsymbol{x}} &= [\ldots, s_1, \ldots, s_3, \ldots].\end{aligned} \tag{3.17}$$

The distance between these codewords $d^2(\boldsymbol{x}, \hat{\boldsymbol{x}}) = d^2(s_2, s_1) + d^2(s_2, s_3) = 8d^2$ is even smaller than the minimum ED if the BRGC would have been used. This formally reduces both the minimum ED and the code diversity. Luckily, the probability to observe such pairs of codewords is very small. So, on average, most of the codewords are separated by a large ED but there are also codewords very close to one another. This effect is very similar to the so-called "spectral thinning" [60] observed for turbo codes. Codewords far apart explain the good performance of BICM-ID at moderate SNR and codewords close to one another cause flattening of the BER at high SNR, which is usually referred to as the error floor [61].

The performance, predicted based on the ML decoder, cannot be obtained if a standard bit-wise decoder is used in combination with a non-Gray labeling, as will be illustrated in Sec. 4.2. This motivates the use of iterative decoders for non-Gray labelings.

[6] If the BRGC was used, this pair of codewords would be very "fortunate", as the squared distance between the codewords would be $d^2(\boldsymbol{x}, \hat{\boldsymbol{x}}) = d^2(s_6, s_1) + d^2(s_5, s_1) = 164d^2$.

Chapter 4

Symbol-Wise and Bit-Wise Decoders

In this section, we generalize the system models of TCM and BICM and study a general CM scheme shown in Fig. 4.1. A CM encoder (CM-ENC) maps information vectors $\boldsymbol{c} \in \{0,1\}^K$ to codewords $\boldsymbol{x} \in \mathcal{X} \subset \mathcal{S}^N$, i.e., $\Phi_{\mathcal{X}} : \{0,1\}^K \to \mathcal{X}$, where no particular structure of the CM code is assumed. The inverse mapping is denoted by $\Phi_{\mathcal{X}}^{-1} : \mathcal{X} \to \{0,1\}^K$. By assigning binary labels to the constellation points, the CM encoder can be seen as a concatenation of a modulator (MOD) and a binary encoder (ENC) that performs a mapping of information vectors $\boldsymbol{c}$ to binary codewords $\boldsymbol{b} \in \mathcal{B}$. The binary encoder is defined as the function $\Phi_{\mathcal{B}} : \{0,1\}^K \to \mathcal{B}$ with the corresponding inverse function $\Phi_{\mathcal{B}}^{-1} : \mathcal{B} \to \{0,1\}^K$. The described coding scheme is equivalent to TCM in Fig. 3.1 if $\mathcal{B}$ is a CC, and it is equivalent to BICM in Fig. 3.4 if $\mathcal{B}$ is an interleaved version of another binary code.

We consider a more general discrete-time channel model given by $Y[k] = H[k]x[k] + Z[k]$, where the conditional PDF of the channel output $Y[k]$ is given by (2.2) and $H[k]$ for $k = 1, \ldots, N$ are real channel coefficients, which are assumed to be known at the receiver. No particular assumption on the distribution or correlation properties of the channel coefficients is made. This channel model is usually referred to as a flat fading channel. The fading may be fast if the coefficients change significantly between adjacent time instants, or slow if the coefficients are highly correlated. As the symbols in the CM code are not required to be equiprobable, we use the ratio d/σ as a measure of SNR. In the following, we describe two decoding algorithms adopted from TCM and BICM and show how they can be compared.

4.1 Symbol-Wise Decoder

The symbol-wise decoder (S-DEC) in Fig. 4.1 is an ML decoder. The ML decoding rule for CM over the AWGN channel is equivalent to the one for uncoded transmission in (2.3) if the symbol and the observation are replaced with the corresponding vectors. For the flat fading channel, the ML decision is

$$\hat{\boldsymbol{C}}^{\mathcal{X}} \triangleq \Phi_{\mathcal{X}}^{-1}\left(\underset{\boldsymbol{u}\in\mathcal{X}}{\operatorname{argmin}}\left\{\|\boldsymbol{Y} - \boldsymbol{H} \odot \boldsymbol{u}\|^2\right\}\right), \tag{4.1}$$

where $\odot$ denotes an element-wise multiplication. The argument of the minimization function is the distance between the observation and a weighted codeword. For a given observation $\boldsymbol{Y} = \boldsymbol{y}$ and a given channel realization $\boldsymbol{H} = \boldsymbol{h}$, the distance is a function of

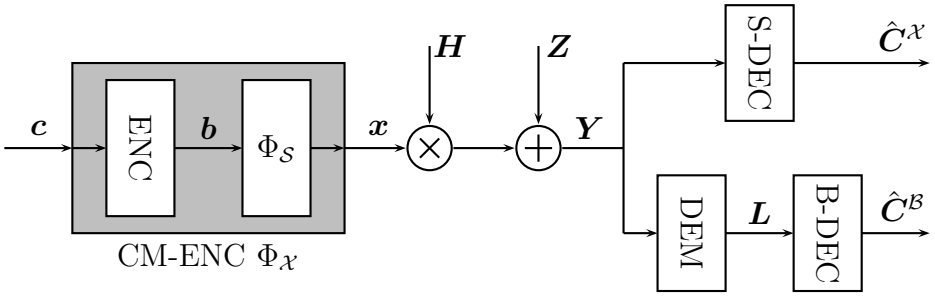


Figure 4.1: Block diagram of a general CM system. The CM encoder $\Phi_{\mathcal{X}}$ is used at the transmitter. At the receiver, two decoders are considered: the ML symbol-wise decoder S-DEC or a suboptimal bit-wise decoder B-DEC.

a codeword $\boldsymbol{u} \in \mathcal{X}$ and it can be written as

$$D^{\mathcal{X}}(\boldsymbol{u}) = \sum_{k=1}^{N} (y[k] - h[k]u[k])^2.$$

For the transmitted codeword $\boldsymbol{x}$, the distance is

$$D^{\mathcal{X}}(\boldsymbol{x}) = \sum_{k=1}^{N} (y[k] - h[k]x[k])^2 = \sum_{k=1}^{N} (h[k]x[k] + z[k] - h[k]x[k])^2 = \sum_{k=1}^{N} z^2[k].$$

For a competing codeword $\hat{\boldsymbol{x}} = [\hat{x}[1], \ldots, \hat{x}[N]] \in \mathcal{X}$, the distance can be expressed as

$$\begin{aligned} D^{\mathcal{X}}(\hat{\boldsymbol{x}}) &= \sum_{k=1}^{N} (y[k] - h[k]\hat{x}[k])^2 = \sum_{k=1}^{N} (h[k]x[k] + z[k] - h[k]\hat{x}[k])^2 \\ &= \sum_{k=1}^{N} h^2[k](x[k] - \hat{x}[k])^2 + 2h[k](x[k] - \hat{x}[k])z[k] + z^2[k]. \end{aligned}$$

The PEP for these two codewords can then be rewritten as

$$\mathrm{PEP}^{\mathcal{X}}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \Pr\{D^{\mathcal{X}}(\hat{\boldsymbol{x}}) - D^{\mathcal{X}}(\boldsymbol{x}) < 0\}. \tag{4.2}$$

We therefore take a closer look at the difference $\Delta^{\mathcal{X}}(\boldsymbol{x}, \hat{\boldsymbol{x}}) \triangleq D^{\mathcal{X}}(\hat{\boldsymbol{x}}) - D^{\mathcal{X}}(\boldsymbol{x})$ between the distances for the codewords $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, i.e.,

$$\Delta^{\mathcal{X}}(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \sum_{k=1}^{N} h^2[k](x[k] - \hat{x}[k])^2 + 2h[k](x[k] - \hat{x}[k])z[k] = 4d \sum_{k=1}^{N} \Lambda^{\mathcal{X}}(x[k], \hat{x}[k]), \tag{4.3}$$

where

$$\Lambda^{\mathcal{X}}(x[k], \hat{x}[k]) \triangleq h^2[k] \left(\frac{x[k] - \hat{x}[k]}{2d} \right)^2 d + h[k] \frac{x[k] - \hat{x}[k]}{2d} z[k] \tag{4.4}$$

is called the symbol metric difference (SMD). Since $Z[k]$ is a Gaussian random variable with zero mean and variance σ^2, the SMD in (4.4) is also a Gaussian random variable with mean $h^2[k]\mu_{\mathcal{X}}(x[k], \hat{x}[k])d$ and variance $h^2[k]\sigma^2_{\mathcal{X}}(x[k], \hat{x}[k])\sigma^2$, where

$$\mu_{\mathcal{X}}(x[k], \hat{x}[k]) = \sigma^2_{\mathcal{X}}(x[k], \hat{x}[k]) \tag{4.5}$$

$$\triangleq \left(\frac{x[k] - \hat{x}[k]}{2d} \right)^2. \tag{4.6}$$

Using (4.2)–(4.3) and the fact that SMDs in (4.4) are independent Gaussian random variables for different $k = 1, \ldots, N$, the PEP for the codewords $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ can be calculated as

$$\begin{aligned}\mathrm{PEP}^{\mathcal{X}}(\boldsymbol{x}, \hat{\boldsymbol{x}}) &= \Pr\left\{\sum_{k=1}^{N} \Lambda^{\mathcal{X}}(x[k], \hat{x}[k]) < 0\right\} \\ &= \mathrm{Q}\left(\frac{\sum_{k=1}^{N} h^2[k]\mu_{\mathcal{X}}(x[k], \hat{x}[k])}{\sqrt{\sum_{k=1}^{N} h^2[k]\sigma^2_{\mathcal{X}}(x[k], \hat{x}[k])}}\frac{d}{\sigma}\right) \\ &= \mathrm{Q}\left(\sqrt{\sum_{k=1}^{N} h^2[k]\mu_{\mathcal{X}}(x[k], \hat{x}[k])}\frac{d}{\sigma}\right), \end{aligned} \tag{4.7}$$

where the last equality was obtain using (4.5). For the AWGN channel, i.e., when $h[k] = 1$ for all $k = 1, \ldots, N$, it can be easily verified that the PEP expression in (4.7) is equivalent to the one in (3.5). The distribution parameters in (4.5)–(4.6) are the key result in this section. In the following, we will show how the distribution parameters of SMDs can be obtained for the B-DEC.

4.2 Bit-Wise Decoder

The B-DEC in Fig. 4.1 operates on the L-values provided by the DEM. The DEM acts independently of the B-DEC and calculates a vector of L-values $\boldsymbol{L}$ using the max-log approximation. For future use, we represent the vector of L-values as a concatenation of length-m vectors, i.e., $\boldsymbol{L} = [\boldsymbol{L}[1], \ldots, \boldsymbol{L}[N]]$, where $\boldsymbol{L}[k] = [L_1[k], \ldots, L_m[k]]$ are the L-values obtained from the kth observation. For the fading channel we consider here, the L-values are calculated similarly to (2.13) accounting for the channel coefficients as[7]

$$L_j[k] \triangleq \frac{1}{4d}\left[\min_{s \in \mathcal{S}_{j,0}} (Y[k] - H[k]s)^2 - \min_{s \in \mathcal{S}_{j,1}} (Y[k] - H[k]s)^2\right]. \tag{4.8}$$

The calculated L-values are passed to the B-DEC, which uses the decoding rule [56, Sec. 2.2], [54, eq. (13)]

$$\hat{C}^{\mathcal{B}} \triangleq \Phi_{\mathcal{B}}^{-1}\left(\operatorname*{argmin}_{\boldsymbol{u} \in \mathcal{B}} \{(2\boldsymbol{u} - 1)\boldsymbol{L}^{\intercal}\}\right). \tag{4.9}$$

For the transmitted codeword $\boldsymbol{b}$, the PEP of detecting another codeword $\hat{\boldsymbol{b}}$ is

$$\mathrm{PEP}^{\mathcal{B}}(\boldsymbol{b}, \hat{\boldsymbol{b}}) = \Pr\{(2\boldsymbol{b} - 1)\boldsymbol{L}^{\intercal} - (2\hat{\boldsymbol{b}} - 1)\boldsymbol{L}^{\intercal} < 0\} = \Pr\{\Delta^{\mathcal{B}}(\boldsymbol{b}, \hat{\boldsymbol{b}}) < 0\}, \tag{4.10}$$

where

$$\Delta^{\mathcal{B}}(\boldsymbol{b}, \hat{\boldsymbol{b}}) \triangleq (\boldsymbol{b} - \hat{\boldsymbol{b}})\boldsymbol{L}^{\intercal} \tag{4.11}$$

can be interpret as a distance between the codewords when the B-DEC is used. In [56, Ch. 4], $-\Delta^{\mathcal{B}}(\boldsymbol{b}, \hat{\boldsymbol{b}})$ was called a pairwise score and the distribution of the pairwise score

[7] The L-value in (4.8) is scaled by $\sigma^2(2d)^{-1}$ compared to (2.13) for convenience. The scaling is irrelevant for the B-DEC though it may be important if other decoders are used, e.g., a decoder based on the sum product algorithm [43, Ch. 5].

		$\hat{x}_i$							
		$s_1\,(000)$	$s_2\,(001)$	$s_3\,(011)$	$s_4\,(010)$	$s_5\,(110)$	$s_6\,(111)$	$s_7\,(101)$	$s_8\,(100)$
x_i	$s_1\,(000)$	0	$-L_3$	$-L_2-L_3$	$-L_2$	$-L_1-L_2$	$-L_1-L_2-L_3$	$-L_1-L_3$	$-L_1$
	$s_2\,(001)$	L_3	0	$-L_2$	$-L_2+L_3$	$-L_1-L_2+L_3$	$-L_1-L_2$	$-L_1$	$-L_1+L_3$
	$s_3\,(011)$	L_2+L_3	L_2	0	L_3	$-L_1+L_3$	$-L_1$	$-L_1+L_2$	$-L_1+L_2+L_3$
	$s_4\,(010)$	L_2	L_2-L_3	$-L_3$	0	$-L_1$	$-L_1-L_3$	$-L_1+L_2-L_3$	$-L_1+L_2$
	$s_5\,(110)$	L_1+L_2	$L_1+L_2-L_3$	L_1-L_3	L_1	0	$-L_3$	L_2-L_3	L_2
	$s_6\,(111)$	$L_1+L_2+L_3$	L_1+L_2	L_1	L_1+L_3	L_3	0	L_2	L_2+L_3
	$s_7\,(101)$	L_1+L_3	L_1	L_1-L_2	$L_1-L_2+L_3$	$-L_2+L_3$	$-L_2$	0	L_3
	$s_8\,(100)$	L_1	L_1-L_3	$L_1-L_2-L_3$	L_1-L_2	L_2	$-L_2-L_3$	$-L_3$	0

Table 4.1: SMDs as different combinations of L-values conditioned on different transmitted symbols for the BRGC.

was analyzed under different assumptions. In this chapter, we are also interested in the distribution of $\Delta^{\mathcal{B}}(\boldsymbol{b},\hat{\boldsymbol{b}})$. However, our main goal is to compare this distribution with that of $\Delta^{\mathcal{X}}(\boldsymbol{x},\hat{\boldsymbol{x}})$ in (4.3). Since the mapping between $\boldsymbol{b}$ and $\boldsymbol{x}$ is one-to-one, with a slight abuse of notation, we use $\Delta^{\mathcal{B}}(\boldsymbol{x},\hat{\boldsymbol{x}})$ instead to highlight the similarity with (4.3).

We can rewrite (4.11) as

$$\Delta^{\mathcal{B}}(\boldsymbol{x},\hat{\boldsymbol{x}}) = \sum_{k=1}^{N} \Lambda^{\mathcal{B}}(x[k],\hat{x}[k]), \tag{4.12}$$

where

$$\Lambda^{\mathcal{B}}(x[k],\hat{x}[k]) \triangleq (\Phi_{\mathcal{S}}^{-1}(x[k]) - \Phi_{\mathcal{S}}^{-1}(\hat{x}[k]))\boldsymbol{L}[k]^{\intercal} \tag{4.13}$$

is the SMD[8] for the B-DEC. Since the channel is memoryless, the SMDs for different $k = 1,\ldots,N$ are independent. Once the distributions of all SMDs are known, the PDF of $\Delta^{\mathcal{B}}(\boldsymbol{x},\hat{\boldsymbol{x}})$ in (4.12) can be found as a convolution of the PDFs of the summands. The PEP in (4.10) will then be given by the integral of the negative tail of this PDF. In the following, we discuss distributions of SMDs for a given time instant k and we omit the index k for clarity.

In order to proceed further, we need to specify the labeling used by the DEM since it greatly affects the SMD in (4.13). We will concentrate on the 8-PAM constellation labeled with the BRGC. Different SMDs $\Lambda^{\mathcal{B}}(x,\hat{x})$ can then be expressed in terms of the L-values L_1, L_2, L_3 as shown in Table 4.1. The binary labels for the constellation symbols are also shown so that the entries of the table become evident.

As mentioned in Sec. 3.2, a random scrambler is usually assumed to be used to symmetrize the channel when analyzing the performance of the B-DEC. This makes the SMDs for the pairs of symbols whose labels differ in the same bits have the same distribution. For example, the SMDs on the main anti-diagonal $\Lambda^{\mathcal{B}}(s_1,s_8)$, $\Lambda^{\mathcal{B}}(s_2,s_7),\ldots,\Lambda^{\mathcal{B}}(s_8,s_1)$ in Table 4.1 should all have the same distribution under such an assumption. This, however, is not generally true. The distributions of the aforementioned SMDs are given by the distribution of L_1 conditioned on different transmitted symbols, and hence, are different. The assumption of a random scrambler gives good predictions of the PEP but cannot be used when the S-DEC and the B-DEC are compared. Here, we do not use this assumption and treat every SMD in Table 4.1 individually.

[8] $-\Lambda^{\mathcal{B}}(x[k],\hat{x}[k])$ is called a symbol score in [56, Ch. 4].

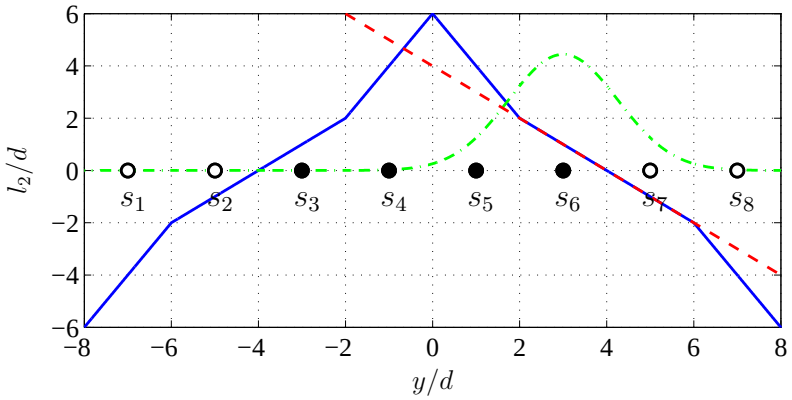


Figure 4.2: Approximation of L-values. The solid line shows (normalized) l_2 as a function of the (normalized) observation y. The dashed line shows the approximation (consistent and ZC) of the L-value. The dash-dotted line shows the distribution $f_{Y|X}(y|s_6)$ for $d/\sigma = -5$ dB; the distribution is scaled for illustration purposes. Empty and filled circles show constellation points labeled with 0 and 1, respectively.

4.3 Distribution of L-values

We first assume that the channel coefficient $h = 1$. As mentioned in Sec. 2.2.2, the max-log L-values are linear functions of the observation. This implies that the SMDs in Table 4.1 are linear functions of the observation as well. In this section, we will show how to obtain the PDFs of two SMDs in Table 4.1 and how these PDFs can be approximated. The distributions of all other SMDs in Table 4.1 can be obtained and approximated in a similar fashion.

It follows from Table 4.1 that the SMD $\Lambda^{\mathcal{B}}(s_6, s_7) = L_2$ and hence, its distribution is the distribution of L_2 given that symbol s_6 was transmitted. Fig. 4.2 shows the piecewise linear function $l_2(y)$ together with the distribution of the observation conditioned on the transmitted symbol s_6 for $d/\sigma = -5$ dB, which corresponds to

$$\gamma = \frac{21d^2}{2\sigma^2} = 5.2 \text{ dB}$$

if the symbols are equiprobable. The PDF $f_{L_2|X}(l|s_6)$ is thus a sum of piecewise Gaussian functions with the mean and the variance determined by the parameters of the corresponding linear pieces of $l_2(y)$. The PDF $f_{L_2|X}(l|s_6)$, or equivalently, the PDF of $\Lambda^{\mathcal{B}}(s_6, s_7)$, is shown in Fig. 4.3 with the solid line.

The PDFs of the SMDs are easy to obtain but they become intractable analytically when several PDFs need to be convolved, and hence, approximations of the distributions are usually used. We consider two ways to approximate such PDFs. The first one is the so-called consistent approximation [59] for which the L-value is approximated by a linear function corresponding to the linear piece over the Voronoi region of the transmitted symbol. This approximation is shown with the dashed line in Fig. 4.2. The approximated PDF is thus a single Gaussian function that approximates the exact PDF "at the mean" of the L-value and it is shown with the dashed line in Fig. 4.3. The second approach approximates the negative tail of the PDF, which is more important when analyzing the PEP. For that, the L-value is approximated using the so-called zero-crossing (ZC)

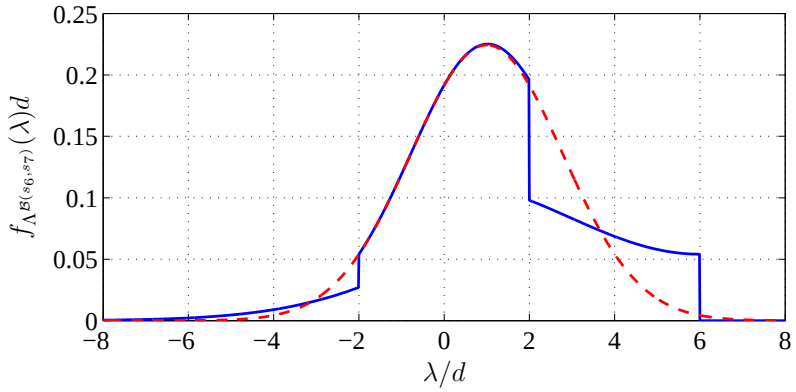


Figure 4.3: Distribution of the SMD $\lambda^{\mathcal{B}}(s_6, s_7)$ for $d/\sigma = -5$ dB. The solid line represents the exact PDF and the dashed line shows the approximated PDF. The consistent and the ZC approximations give the same results.

approximation [40], i.e., the L-value is approximated by a linear function corresponding to the linear piece of the L-value function intersecting the zero-level at the closest point to the transmitted symbol. For the example in Fig. 4.2, the ZC approximation coincides with the consistent approximation (dashed line) and gives the same approximated PDF, which is shown with the dashed line in Fig. 4.3. However, the two approximations are not always the same. In the following, we give an example showing that the ZC approximation should be preferred when analyzing the probability of error.

As shown in Table 4.1, the SMD $\Lambda^{\mathcal{B}}(s_6, s_8) = L_2 + L_3$ and it is shown in Fig. 4.4 as a piece-wise linear function of the observation y together with the conditional distribution of the observation $f_{Y|X}(y|s_6)$. The exact PDF of $\Lambda^{\mathcal{B}}(s_6, s_8)$ is a sum of piece-wise Gaussian functions shown in Fig. 4.5. Due to the horizontal parts of the SMD in Fig. 4.4, the exact PDF contains a Dirac delta function with amplitude $A = 1-2\mathrm{Q}(d/\sigma)+\mathrm{Q}(5d/\sigma)-\mathrm{Q}(7d/\sigma)$ at $\lambda = 2d$. The consistent approximation approximates the SMD by a linear function $\lambda = 2d$ and hence, the approximated PDF is a Dirac delta function of unit amplitude at $\lambda = 2d$. Clearly, such an approximation will give poor results when the PEP is analyzed.

The ZC approximation of the SMD $\Lambda^{\mathcal{B}}(s_6, s_8)$ is shown with the dashed line in Fig. 4.4 and it results in a Gaussian distribution shown with the dashed line in Fig. 4.5. It is possible to show that the ZC crossing approximation is asymptotically tight (when $\gamma \to \infty$) in terms of PEP, if $\Delta^{\mathcal{B}}(\boldsymbol{x}, \hat{\boldsymbol{x}})$ in (4.12) consists of only one SMD. We also considered different $\Delta^{\mathcal{B}}(\boldsymbol{x}, \hat{\boldsymbol{x}})$ consisting of several SMDs, for instance $\Delta^{\mathcal{B}}(\boldsymbol{x}, \hat{\boldsymbol{x}})$ for the codewords in [Paper C, Fig. 2], and the asymptotic tightness of the ZC approximation was successfully verified. For a general case, where $\Delta^{\mathcal{B}}(\boldsymbol{x}, \hat{\boldsymbol{x}})$ is a summation of an arbitrary number of different SMDs, the tightness of the ZC approximation remains an open question.

The above examples were presented under the assumption that the channel coefficient $h = 1$. It can be shown that, regardless of the chosen approximation, the mean and the variance of the SMD scale with h^2, similarly to the distribution parameters of the SMD for the S-DEC in (4.4). Once the parameters of the approximated distributions (the mean and the variance) for all entries of Table 4.1 are known, they can be compared to those of the S-DEC. We performed this analysis for 4-PAM labeled with a Gray code in [Paper C]. The results showed that the parameters are very similar for the two decoders, which allowed us to bound the asymptotic performance loss of the B-DEC compared to

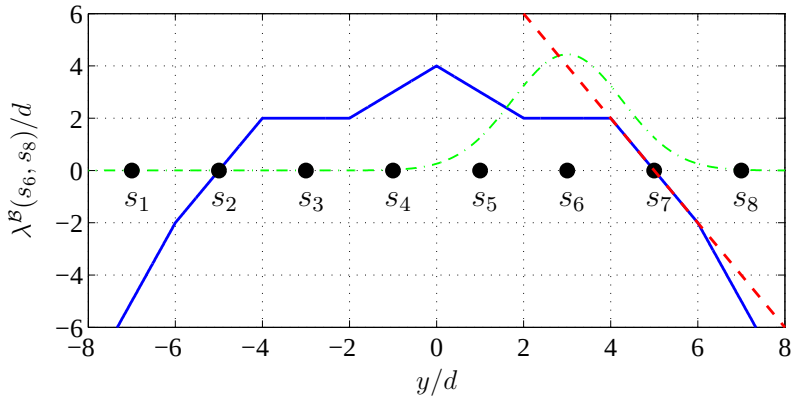


Figure 4.4: Approximation of the SMD. The solid line shows the (normalized) SMD $\lambda^{\mathcal{B}}(s_6, s_8)$ as a function of the (normalized) observation y. The dashed line shows the ZC approximation of the SMD. The dash-dotted line shows the distribution $f_{Y|X}(y|s_6)$ for $d/\sigma = -5$ dB; the distribution is scaled for illustration purposes.

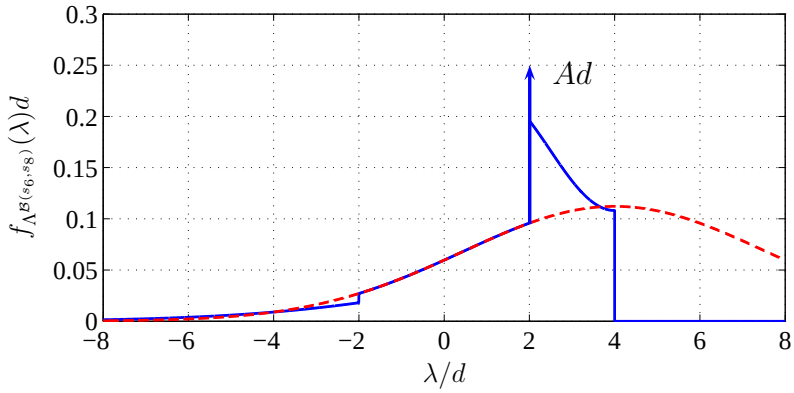


Figure 4.5: Distribution of the SMD $\lambda^{\mathcal{B}}(s_6, s_8)$ for $d/\sigma = -5$ dB. The solid line represents the exact PDF and the dashed line shows the approximated PDF using the ZC approximation. The PDF obtained based on the consistent approximation is a Dirac delta function with amplitude 1 at $\lambda = 2d$ and is not shown in the figure.

the S-DEC. The analysis also showed that for a wide range of codes, the performance loss is equal to zero. This phenomenon is illustrated in Fig. 3.3 for 8-PAM. The dash-dotted line marked circles shows the performance of the B-DEC with the BRGC for the TCM scheme $(\mathbb{C}_{\text{BRGC}}, \mathbb{G}_{\text{BRGC}})$. The B-DEC causes a loss of fractions of a dB compared to the S-DEC shown with the dashed line marked with circles.

A similar analysis could be performed for other than Gray labelings. For example, we derived the distribution parameters of the SMDs for 4-PAM labeled with the NBC. Unfortunately, the results do not allow us to bound the performance loss and draw quantitative conclusions. However, they clearly indicate that the loss can be large when the B-DEC is used with the NBC. We illustrate this with Fig. 3.3 for 8-PAM. The dotted line marked with circles shows the performance of the B-DEC with the NBC for the TCM scheme $(\mathbb{C}_{\text{BRGC}}, \mathbb{G}_{\text{BRGC}})$. For that, the demapper for the NBC together with a decoder for the binary code $\mathbb{C}_{\text{NBC}}$ is used at the receiver. As we can see, the loss is approximately

4.5 dB compared to the S-DEC. This gives an intuition why the B-DEC performs bad when used with a non-Gray labeling and motivates the use of iterative decoders to obtain a near-optimal performance.

Chapter 5

Contributions and Future Work

5.1 Contributions

The main objective of this thesis is to compare bit-wise and symbol-wise decoders for different scenarios. This work is an attempt to understand and explain how different labelings affect the performance of coded systems for the two decoders. The contributions made by the author in this area are presented in Part II of the thesis and summarised in the following.

5.1.1 Paper A: "General BER Expression for One-Dimensional Constellations"

In Paper A, we present a novel general ready-to-use BER expression for arbitrary one-dimensional constellations over the AWGN channel. The BER expression is formulated in terms of decision thresholds. The BER analysis is performed for bit patterns that form a labeling, which allows us to use the derived expression for the BD if the decision thresholds are known. This paper serves as the basis for the results in Paper B, where the thresholds are found analytically. We also analyze the number of bit patterns for equally spaced M-PAM constellations with different BER.

5.1.2 Paper B: "On the Exact BER of Bit-Wise Demodulators for One-Dimensional Constellations"

In Paper B, we study the decision thresholds for the BD for equally spaced one-dimensional constellations. Closed form expressions for the decision thresholds are found for all patterns for 4-PAM and for 11 out of 23 patterns for 8-PAM. This enables us to obtain closed form BER expressions for 4-PAM with any labeling and for 8-PAM with some of the most popular labelings, including the BRGC, the NBC, and the AGC. Numerical results show that, regardless of the labeling, there is no difference between the optimal demodulator and the symbol-wise demodulator for any BER of practical interest (below 0.1).

5.1.3 Paper C: "On the Asymptotic Performance of Bit-Wise Decoders for Coded Modulation"

In Paper C, we analyze the system model in Fig. 4.1 and compare the S-DEC and the B-DEC for a 16-QAM constellation formed as a direct product of two 4-PAM constellations. The labeling of the 16-QAM constellation is also obtained as a direct product of two

Gray-labeled 4-PAM constellations. We therefore concentrate the analysis only on the constituent 4-PAM constellation. We find the SMD distributions for the B-DEC and compare them with the SMD distributions for the S-DEC. This allows us to bound the asymptotic performance loss in terms of the PEP for any pair of codewords and any channel realization by 1.25 dB. For the AWGN channel, we show that the asymptotic loss for codes is zero for two out of the four nonequivalent Gray labelings if the underlying binary code $\mathcal{B}$ is linear. For other Gray labelings, including the BRGC, the analysis shows that the asymptotic loss is zero for a wide range of linear codes, including all rate-1/2 convolutional codes.

5.2 Future Work

The results of Paper C rely on the ZC approximation for max-log L-values. Even though it has been shown numerically in [40], [41], [59] that the ZC approximation is good in terms of predicting the coded BER, the proof for its asymptotic tightness is an open question. Moreover, the ZC approximation has never been thoroughly studied for other than Gray labelings, which could be a future work direction.

Information-theoretic studies of BICM usually address the question of achievable rates for the B-DEC with exact L-values. Closed form expressions for the PDFs of max-log L-values enable us to study the MI and the generalized MI for the B-DEC with max-log L-values, which has not received too much attention in the literature.

Finally, one could also try to extend the results of equivalent TCM schemes to other than convolutional codes, for instance, LDPC codes. This may allow us to compare different labelings under the sum-product decoding algorithm. This could also enable us to perform a fair comparison of noniterative and iterative BICM schemes that use different labelings. As for now, it is still unclear if BICM with a Gray code is better than BICM-ID with some other labeling or vice versa.

References

[1] C. E. Shannon, "A mathematical theory of communications," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

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

[3] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Transactions on Communications*, vol. 22, no. 1, pp. 28–38, Jan. 1974.

[4] M. Barsoum, C. Jones, and M. Fitz, "Constellation design via capacity maximization," in *IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007.

[5] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 726–740, July 1990.

[6] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets, Part I: Introduction," *IEEE Communications Magazine*, vol. 25, no. 2, pp. 5–11, Feb. 1987.

[7] J. L. Massey, "Coding and modulation in digital communications," in *International Zurich Seminar on Digital Communications*, Zurich, Switzerland, Mar. 1974.

[8] G. Ungerboeck and I. Csajka, "On improving data-link performance by increasing channel alphabet and introducing sequence decoding," in *International Symposium on Information Theory (ISIT)*, Ronneby, Sweden, June 1976, (Book of abstracts).

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

[10] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

[11] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Transactions on Information Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.

[12] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 927–946, May 1992.

[13] ETSI, "Digital video broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2)," ETSI, Tech. Rep. ETSI EN 302 755 V1.3.1 (2012-04), Apr. 2012.

[14] IEEE 802.11, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE Std 802.11-2012, Tech. Rep., Mar. 2012.

[15] ETSI, "LTE; Evolved universal terrestrial radio access (E-UTRA); Physical channels and modulation," ETSI, Tech. Rep. ETSI TS 136 211 V11.2.0 (2013-04), Apr. 2013.

[16] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *IEEE International Conference on Communications (ICC)*, Geneva, Switzerland, May 1993.

[17] R. Gallager, "Low-density parity-check codes," *IRE Trans. Information Theory*, pp. 21–28, Jan. 1962.

[18] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electronics Letters*, vol. 33, no. 6, pp. 457–458, Mar. 1997.

[19] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Bandwidth efficient parallel concatenated coding schemes," *Electronics Letters*, vol. 31, no. 24, pp. 2067–2069, Nov. 1995.

[20] ——, "Parallel concatenated trellis coded modulation," in *IEEE International Conference on Communications (ICC)*, Dallas, TX, June 1996.

[21] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Communications Letters*, vol. 1, no. 6, pp. 169–171, Nov. 1997.

[22] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multi-level modulation," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Sydney, Australia, Nov. 1998.

[23] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2000.

[24] M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection.* Prentice Hall, 1995.

[25] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, July 2002.

[26] P. J. Lee, "Computation of the bit error rate of coherent M-ary PSK with Gray code bit mapping," *IEEE Trans. Commun.*, vol. COM-34, no. 5, pp. 488–491, May 1986.

[27] J. Lassing, E. G. Ström, E. Agrell, and T. Ottosson, "Computation of the exact bit-error rate of coherent M-ary PSK with Gray code bit mapping," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1758–1760, Nov. 2003.

[28] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.

[29] J. Lassing, E. G. Ström, E. Agrell, and T. Ottosson, "Unequal bit-error protection in coherent M-ary PSK," in *IEEE Vehicular Technology Conference (VTC-Fall)*, Orlando, FL, Oct. 2003.

[30] L. Szczecinski, C. Gonzalez, and S. Aissa, "Exact expression for the BER of rectangular QAM with arbitrary constellation mapping," *IEEE Trans. Commun.*, vol. 54, no. 3, pp. 389–392, Mar. 2006.

[31] M. K. Simon and R. Annavajjala, "On the optimality of bit detection of certain digital modulations," *IEEE Trans. Commun.*, vol. 53, no. 2, pp. 299–307, Feb. 1988.

[32] F. Gray, "Pulse code communications," U. S. Patent 2 632 058, Mar. 1953.

[33] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "Gray coding for multilevel constellations in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 224–235, Jan. 2007.

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

[35] A. J. Viterbi, "An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 260–264, Feb. 1998.

[36] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *IEEE International Conference on Communications (ICC)*, Seattle, WA, June 1995.

[37] K. Hyun and D. Yoon, "Bit metric generation for Gray coded QAM signals," *IEEE Proc.-Commun.*, vol. 152, no. 6, pp. 1134–1138, Dec. 2005.

[38] M. S. Raju, R. Annavajjala, and A. Chockalingam, "BER analysis of QAM on fading channels with transmit diversity," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 481–486, Mar. 2006.

[39] C. Stierstorfer, "A bit-level-based approach to coded multicarrier transmission," Ph.D. dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, 2009, available at http://www.opus.ub.uni-erlangen.de/opus/volltexte/2009/1395/.

[40] M. Benjillali, L. Szczecinski, S. Aissa, and C. Gonzalez, "Evaluation of bit error rate for packet combining with constellation rearrangement," *Wiley Journal Wireless Comm. and Mob. Comput.*, vol. 8, no. 7, pp. 831–844, Sep. 2008.

[41] A. Alvarado, L. Szczecinski, R. Feick, and L. Ahumada, "Distribution of L-values in Gray-mapped M^2-QAM: Closed-form approximations and applications," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2071–2079, July 2009.

[42] ITU, "A duplex modem operating at data signalling rates of up to 14 400 bit/s for use on the general switched telephone network and on leased point-to-point 2-wire telephone-type circuits," ITU-T Recommendation V.32 bis, Tech. Rep., Feb. 1991.

[43] W. E. Ryan and S. Lin, *Channel codes: Classical and Modern.* Cambridge University Press, 2009.

[44] L.-F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Transactions on Information Theory*, vol. IT-33, no. 4, pp. 483–501, July 1987.

[45] A. Alvarado, A. Graell i Amat, F. Brännström, and E. Agrell, "On optimal TCM encoders," *IEEE Trans. Commun.*, 2013 (to appear), available at http://arxiv.org/abs/1210.2107.

[46] K. Larsen, "Short convolutional codes with maximal free distance for rates 1/2, 1/3, and 1/4," *IEEE Transactions on Information Theory*, vol. 19, no. 3, pp. 371–372, May 1973.

[47] Y. Li and W. E. Ryan, "Bit-reliability mapping in LDPC-coded modulation systems," *IEEE Communications Letters*, vol. 9, no. 1, pp. 1–3, Jan. 2005.

[48] A. Alvarado, "Towards fully optimized BICM transmissions," Ph.D. dissertation, Chalmers University of Technology, Göteborg, Sweden, 2010, available at http://publications.lib.chalmers.se/records/fulltext/130824.pdf.

[49] A. Alvarado, L. Szczecinski, E. Agrell, and A. Svensson, "On BICM-ID with multiple interleavers," *IEEE Commun. Lett.*, vol. 14, no. 9, pp. 785–787, Sep. 2010.

[50] A. Ingber and M. Feder, "Parallel bit interleaved coded modulation," in *48th Annual Allerton Conference on Communication, Control, and Computing*, Allerton, IL, 2010.

[51] A. Guillén i Fàbregas and A. Martinez, "Derivative of BICM mutual information," *IET Electronics Letters*, vol. 43, no. 22, pp. 1219–1220, Oct. 2007.

[52] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Bit-interleaved coded modulation in the wideband regime," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5447–5455, Dec. 2008.

[53] A. Alvarado, F. Brännström, E. Agrell, and T. Koch, "High-SNR asymptotics of the mutual information with applications to BICM," Mar. 2013, available at http://arxiv.org/1212.6526.

[54] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.

[55] J. Jaldén, P. Fertl, G., and Matz, "On the generalized mutual information of BICM systems with approximate demodulation," in *IEEE Information Theory Workshop (ITW)*, Jan. 2010.

[56] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.

[57] A. Alvarado, L. Szczecinski, and E. Agrell, "On the performance of BICM with trivial interleavers in nonfading channels," in *IEEE International Conference on Communications (ICC)*, Kyoto, Japan, June 2011.

[58] C. Stierstorfer, R. F. H. Fischer, and J. B. Huber, "Optimizing BICM with convolutional codes for transmission over the AWGN channel," in *International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2010.

[59] A. Alvarado, L. Szczecinski, and E. Agrell, "On BICM receivers for TCM transmission," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2692–2707, Oct. 2011.

[60] L. C. Perez, J. Seghers, and D. J. Costello, Jr., "A distance spectrum interpretation of turbo codes," *IEEE Transactions on Information Theory*, vol. 42, no. 16, pp. 1698–1709, Nov. 1996.

[61] S. Pfletschinger and F. Sanzi, "Error floor removal for bit-interleaved coded modulation with iterative detection," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3174–3181, Nov. 2006.