

CHALMERS



Towards Fully Optimized BICM Transmissions

ALEX ALVARADO

Department of Signals and Systems
Communication Systems Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2010

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Towards Fully Optimized BICM Transmissions

ALEX ALVARADO



CHALMERS

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

Göteborg, Sweden 2010

Towards Fully Optimized BICM Transmissions
ALEX ALVARADO
ISBN 978-91-7385-465-8

Copyright ©2010 Alex Alvarado except where
otherwise stated. All rights reserved.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 3146
ISSN 0346-718X

This thesis has been prepared using L^AT_EX and B_IB_TE_X.

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

Printed by Chalmers Reproservice,
Göteborg, Sweden, Dec. 2010.

To Mirta and Luis (bis)

Abstract

Bit-interleaved coded modulation (BICM) was introduced in 1992 as a new coded modulation (CM) scheme particularly well suited for fading channels and is nowadays considered the *de facto* CM scheme in wireless standards (HSPA, IEEE 802.11a/g/n, DVB-T2/S2/C2, etc.). A typical BICM configuration is based on a convolutional code, a single bit-level interleaver (S-interleavers), and an equally spaced quadrature amplitude modulation (QAM) input alphabet. In this thesis, we investigate the design of these three building blocks and show that traditional configurations are suboptimal. The main contribution of this thesis is to formally study the design of BICM systems in fading and nonfading channels and to propose new BICM designs.

First, the use of multiple interleavers (M-interleavers) in BICM and BICM with iterative decoding (BICM-ID) is formally studied. For BICM, equally spaced QAM input alphabets labeled by the binary reflected Gray code are analyzed. Analytical bounds on the bit error rate (BER) are developed and based on these bounds, the optimum interleaver design is presented. It is shown that in both BICM and BICM-ID, M-interleavers outperform S-interleavers. Furthermore, the asymptotical optimality of BICM-ID with M-interleavers over BICM-ID with S-interleavers is proven. It is also shown that the use of M-interleaver redefines the optimality of the codes used in BICM and BICM-ID. The selection of these optimal codes is also studied in this thesis.

A new BICM scheme based on nonequally spaced (hierarchical) QAM input alphabets, a bit-level multiplexer, and M-interleavers is also proposed in this thesis. It is shown that this new scheme outperforms previous BICM designs in fading and nonfading channels. Analytical bounds on the BER are developed and used to optimize the design of the system. The results show that, compared to traditional BICM designs, gains of a few decibels can be obtained.

Finally, we study the use of BICM when the interleaver takes a trivial form (BICM-T), i.e., when it does not interleave the bits at all. An analytical model based on a new type of distance spectrum for convolutional codes is developed. This model is used to explain why BICM-T offers considerable gains compared to previous BICM configurations. It is also shown that properly designed BICM-T systems can be asymptotically as good as TCM.

Keywords: Bit-interleaved coded modulation, binary reflected Gray code, coded modulation, convolutional codes, hierarchical constellations, interleaver design, iterative decoding, pulse amplitude modulation, quadrature amplitude modulation, trellis-coded modulation.

Towards Fully Optimized BICM Transmissions
ALEX ALVARADO
Department of Signals and Systems
Chalmers University of Technology

List of Publications

This thesis is based on the following publications:

Paper A

Alex Alvarado, Erik Agrell, Leszek Szczecinski, and Arne Svensson, “Exploiting UEP in QAM-based BICM: Interleaver and Code Design,” *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 500–510, Feb. 2010.

Paper B

Alex Alvarado, Leszek Szczecinski, Erik Agrell, and Arne Svensson, “On BICM-ID with Multiple Interleavers,” *IEEE Commun. Lett.*, vol. 14, no. 9, pp. 785–787, Sep. 2010.

Paper C

Md. Jahangir Hossain, **Alex Alvarado**, and Leszek Szczecinski “*Towards Fully Optimized BICM Transceivers*,” submitted to *IEEE Trans. Commun.*, Dec. 2010.

Paper D

Alex Alvarado, Leszek Szczecinski, and Erik Agrell “On BICM receivers for TCM transmission,” submitted to *IEEE Trans. Commun.*, Aug. 2010 (revised Dec. 2010).

Acknowledgements

First of all, I would like to express my gratitude to Professor Arne Svensson for being my supervisor and for giving me the opportunity of becoming a PhD student at the Communication Systems Group. This gratitude also goes to Professor Erik Ström for giving me the opportunity to teach in the MPCOM program. Special thanks to Associate Professor Thomas Eriksson for all the informal discussions.

If luck exists, I must admit that over the years, I have been extremely lucky: I always happened to be at the right place at the right time. It all started at home with my parents being my teachers, both of them very strict. I also remember me and my brother playing a game that consisted of writing consecutive numbers in a very old notebook. The rule was simple: the one that writes more numbers wins. If I only knew back then what I was going to be doing 20 years after... High school came, and I met my math teacher Jorge Azócar who would do anything to make sure everyone in the class understood. I remember participating in a math reading club with him, on Saturdays, at 9 AM. Now you know who to blame...

University came, and at UTFSM in Valparaíso, my MSc thesis was guided by Professor Rodolfo Feick who introduced me to research. I then moved to INRS-EMT in Montreal Canada, where I worked with Professor Leszek Szczeciński. Finally, at Chalmers, I ended up doing research with Professor Erik Agrell. I must admit that on countless opportunities I was shocked by Leszek's and Erik's comments, ideas, and suggestions. These comments were always very different in style, but looking at them in retrospect, I now cannot do anything else than plainly admit that these two people are geniuses. Thank you both for all the help and guidance.

I thank the current and former members of the Communication Systems Group at Chalmers, and in particular Dr. A. Wolfgang, Dr. D. Persson, Dr. A. Tahmasebi, and Dr. F. Munier for their friendship. Special thanks go to Assistant Professor F. Brännström for carefully proofreading this thesis. From "the other corridor", I thank all the cool people at the Signal Processing Group for all the fun we had over the years. Particular thanks go to Livia, Panagiota, Daniel, Lars, and Fredrik, for being there throughout our PhD studies, and also to Lennart and Eija for all the music and friendship. Also, many thanks to the "Spanish crew" formed by Dr. Guillermo, Carmen, Elena, Astrid and Miriam, for all the coffees and chocolates at 6 PM, as well as for the Spanish lessons.

For helping me with all the administrative tasks, I would like to send my gratitude to Agneta and Madeleine. I cannot imagine how the department would look

like without them. My gratitude also goes to Lars for arranging all the computer-related stuff and for having such an interesting sense of humor.

Of course, I cannot miss this opportunity to thank friends from outside Chalmers. Many thanks to my “old friends” in Göteborg: Petra, Alice and Fredrik, Renata, Jessica, Kieran, and Cecile, and many thanks to the “best roommates/neighbors ever”: Adriana, Bastian, Berni and Tess, and Katarina. I also send my regards to my “Canadian” friends Sami and Annemarie who gave me shelter (and food!) in Montreal many times as well as to my *new family* in the old continent: Corrie, Wim, Robert, and Eric. Finally, many thanks to Tomas and Emilie, and yes Tomas, I *will* take you to the Nobel prize ceremony, just sit and wait...

Last not but least, I thank the love of my life Kirsten. Your support over the years is something I would never ever forget. *Ik hou heel veel van jou.*

Alex Alvarado
Göteborg, December of 2010

Different parts of this research were partially supported by the Swedish Research Council, Sweden (under research grant #2006-5599), by the European Commission (under project NEWCOM++ #216715), by FQRNT, Canada (Merit Scholarship Program for Foreign Students), and by the Department of Signals and Systems (Solveig and Karl G Eliasson Memorial Fund).

Acronyms

AMI	Average Mutual Information
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BICM	Bit-Interleaved Coded Modulation
BICM-ID	BICM with Iterative Decoding
BICM-M	BICM with Multiple interleavers
BICM-S	BICM with Single interleavers
BICM-T	BICM with Trivial interleavers
BICO	Binary Input Continuous Output
BPSK	Binary Phase-Shift Keying
BRGC	Binary Reflected Gray Code
BSGC	Binary Semi-Gray Code
CM	Coded Modulation
DICO	Discrete Input Continuous Output
DVB	Digital Video Broadcasting
DVB-T/S/C	DVB Terrestrial/Satellite/Cable
EDS	Euclidean Distance Spectrum
EFF	Error-Free Feedback
GB	Giga Byte
GEDS	Generalized Euclidean Distance Spectrum
GWDS	Generalized Weight Distribution Spectrum
HPAM	Hierarchical Pulse Amplitude Modulation
HQAM	Hierarchical Quadrature Amplitude Modulation
HSPA	High Speed Packet Access

i.i.d.	Independent and Identically Distributed
ISI	Intersymbol Interference
kbps	Kilobit Per Second
MAP	Maximum A Posteriori
ML	Maximum Likelihood
MLC	Multilevel Coding
MLC-MSD	MLC with Multistage Decoding
MLC-PDL	MLC with Parallel Decoding of the individual Levels
MMSE	Minimum Mean Square Error
NBC	Natural Binary Code
ODS	Optimum Distance Spectrum
OFDM	Orthogonal Frequency-Division Multiplexing
PAM	Pulse Amplitude Modulation
PDF	Probability Density Function
PEP	Pairwise Error Probability
PF	Perfect Feedback
PMF	Probability Mass Function
PSK	Phase-Shift Keying
PSD	Power Spectral Density
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
RSCC	Recursive and Systematic Convolutional Code
SEK	Swedish Kronor
SNR	Signal to Noise Ratio
SP	Set Partitioning
3GPP	Third Generation Partnership Project
TCM	Trellis-Coded Modulation
UB	Union Bound
UEP	Unequal Error Protection
USB	Universal Serial Bus
WDS	Weight Distribution Spectrum

Contents

Abstract	i
List of Publications	iii
Acknowledgements	v
Acronyms	vii
I Introduction	1
1 Introduction	3
1.1 Preliminaries	4
1.1.1 Notation Convention	4
1.1.2 Thesis Scope	5
1.2 Channel Models	6
1.2.1 The AWGN Channel	6
1.2.2 Fading Channels	8
1.2.3 A General Discrete-Time Model	9
1.3 Binary Labelings and Signal Sets	10
1.3.1 Binary Labelings	10
1.3.2 Input Alphabets	11
2 Coded Modulation	15
2.1 Introduction to Coded Modulation	15
2.1.1 The Need for Coded Modulation	15
2.1.2 Historical Overview	16
2.1.3 A General Model	17
2.2 TCM, MLC, and BICM	18
2.2.1 Trellis-Coded Modulation	18
2.2.2 Multilevel Coding	19
2.2.3 BICM and BICM-ID	20

3	Capacity Aspects of Coded Modulation	25
3.1	AMI and Channel Capacity	25
3.2	Coded Modulation Capacity	28
3.3	Capacity of BICM Systems	29
3.3.1	BICM with Arbitrary Input Distributions	29
3.3.2	BICM Capacity	30
3.3.3	Maximum BICM Capacity	32
4	Modeling BICM and BICM-ID	39
4.1	The BICM Channel	39
4.1.1	Perfect Feedback Assumption in BICM-ID	40
4.1.2	Probability Density Function of L-values in BICM	44
4.2	Unequal Error Protection in BICM and BICM-ID	51
4.2.1	Unequal Error Protection in BICM	52
4.2.2	Unequal Error Protection in BICM-ID	53
4.2.3	Taking Advantage of the UEP	55
4.3	Performance Evaluation	56
4.3.1	Union Bound and Pairwise Error Probability	56
4.3.2	The standard UB	58
4.3.3	Generalized UB	60
4.3.4	A New Generalized UB	62
4.3.5	Additional Remarks	63
5	Conclusions	65
5.1	Optimization Opportunities	65
5.1.1	Interleaver Alternatives for BICM and BICM-ID	65
5.1.2	Code Selection	66
5.1.3	Signal Shaping	66
5.2	Included Papers	67
5.3	Future Work	68
5.4	Related Contributions	68
	Appendix: Numerical Implementation of the CM and BICM Capacities	70
	Bibliography	71
II	Included Papers	85
A	Exploiting UEP in QAM-based BICM: Interleaver and Code Design	A1
1	Introduction	A2
2	System Model	A4
2.1	The interleavers and the multiplexing unit	A4

2.2	Equivalent Channel Model	A6
3	Interleaver and Code Design	A9
3.1	Generalized weight distribution spectrum	A10
3.2	Union bounds for BICM-QAM	A11
4	Numerical Results	A15
4.1	UB for BICM-QAM	A15
4.2	Optimum Interleaver and Code Design for BICM with Convolutional Codes	A19
5	Conclusions	A22
	References	A26
B On BICM-ID with Multiple Interleavers		B1
1	Introduction	B2
2	System Model and Preliminaries	B2
2.1	System Model	B2
2.2	Perfect Feedback and the BICM-ID Channel	B4
2.3	Union Bound	B4
3	Main Results and Conclusions	B6
3.1	BICM-ID-M with 4-PSK	B6
3.2	Optimality of BICM-ID-M	B6
3.3	Optimal Convolutional Codes	B8
	References	B9
C Towards Fully Optimized BICM Transceivers		C1
1	Introduction	C2
2	Proposed BICM Transceiver	C4
2.1	Encoder, Multiplexing, and Interleaving	C4
2.2	HPAM Constellations	C6
3	Equivalent Channel Model	C9
4	Performance Analysis	C11
4.1	Union Bound	C11
4.2	Equivalent Weight Distribution Spectrum	C12
4.3	Computation of $PEP(\mathbf{w}; \bar{\gamma})$	C13
5	Numerical Results	C15
5.1	Spectral efficiency 1 bit/dimension	C15
5.2	Spectral efficiency 1.5 bit/dimension	C18
6	Conclusions	C22
	References	C27
D On BICM receivers for TCM transmission		D1
1	Introduction	D2
2	System Model and Preliminaries	D3
2.1	System Model	D3
2.2	The Interleaver	D5

2.3	The Decoder	D7
3	Performance Evaluation	D8
3.1	BER Performance	D8
3.2	PDF of the L-values	D11
4	Discussion and Applications	D14
4.1	Performance of BICM-T	D14
4.2	Asymptotic Performance	D16
4.3	Asymptotically Optimum Convolutional Codes	D18
4.4	BICM-T vs. TCM	D19
5	Conclusions	D21
	References	D22

Part I

Introduction

Chapter 1

Introduction

Nowadays, for many people around the world, being for more than one week without access to a reliable internet connection sounds like a nightmare. The “always online syndrome” is commonplace in modern developed societies, where some time ago we became—because of work or for pleasure—email-dependent (and sometimes also Skype-dependent, Facebook-dependent, Twitter-dependent, etc.). During 2010, in Sweden, it was possible to buy a 3G USB modem for \$43 (299 SEK) and surf the internet from practically everywhere in the country with speeds of a few megabits per second for approximately \$10 (69 SEK) per month (up to 2GB of traffic), i.e., the price of a hamburger with French fries in a fast food restaurant. The fact that one could sit in a sailing boat and use Spotify to stream music for \$10 per month would have been hard to imagine 15 years ago (when we used dial-up internet connections of 28.8 kbps). “What will the internet access speeds (and technologies, applications, etc.) be in the future?” is nowadays a somehow common question. However, an even more interesting question exists: How did we go from a fixed internet connection of 28.8 kbps to a wireless link with a few tens of Mbps?

Behind these plug-and-play devices like the 3G USB modem, there is a whole complex communication system which is the result of years of efforts. Such devices usually follow a set of recommendations—the so-called *communication standards*—that assure interoperability, backward compatibility, etc. These standards define how particular communication systems are implemented.

Bit-interleaved coded modulation (BICM) is nowadays the *de facto* alternative for wireless communications standards such as HSPA [1] [2, Ch. 12], IEEE 802.11a/g [3] IEEE 802.11n [4, Sec. 20.3.3], and the DVB standards (DVB-T2 [5], DVB-S2 [6], and DVB-C2 [7]). BICM is also likely to become the *de facto* choice for most future standards, and therefore, its understanding is a key element for improving the efficiency of current and future

telecommunications systems.

In this thesis, we investigate different ways of improving the design of BICM systems. By doing this, the reliability of the transmission is improved and/or the power consumption is lowered (i.e., longer battery life). The general methodology followed throughout this thesis is the development of mathematical models that can be used to predict the performance of BICM systems. These analytical models are then used to improve current BICM designs and to propose new ones.

This thesis is divided into two parts. In Part I, we discuss general concepts about digital transmission and coded modulation (CM) techniques, with particular focus on BICM. We also present some information-theoretic aspects of CM systems and give a general background that serves as an introduction to the second part of the thesis. In Part II, four contributions made by the author to the field of BICM are included as appended papers.

This first chapter of Part I is intended to give a quick introduction to the main topic of the thesis and also to serve as a basis for the following chapters. This chapter starts by briefly introducing BICM and setting the notation convention to be used throughout this thesis. Later, the scope of the thesis is discussed, and a brief review of signal space concepts is presented. We conclude this chapter by introducing the equivalent baseband discrete-time channel models that will be used throughout this thesis.

The remaining of Part I of this thesis is organized as follows. Chapter 2 describes general aspects of coded modulation and it introduces BICM and BICM with iterative decoding (BICM-ID). Chapter 3 discusses capacity aspects of coded modulation systems, paying special attention to the so-called BICM capacity. In Chapter 4, analytical models that can be used to analyze BICM and BICM-ID are presented. Finally, in Chapter 5, the conclusions and some open research problems in the field are presented.

1.1 Preliminaries

1.1.1 Notation Convention

Hereafter we use lowercase letters x to denote a scalar, boldface letters \mathbf{x} to denote a row vector of scalars, and underlined letters $\underline{\mathbf{x}}$ to denote a sequence. The all-zeros and the all-ones vectors are denoted by $\mathbf{0}$ and $\mathbf{1}$, respectively. The norm of a vector is denoted by $\|\mathbf{x}\|$ and defined as $\|\mathbf{x}\|^2 = \sum_{i=1}^N x_i^2$, and the element-wise (Schur) product between the vectors \mathbf{x} and \mathbf{y} by $\mathbf{x} \circ \mathbf{y}$. Blackboard bold letters \mathbb{X} represent matrices and $x_{i,j}$ represents the entry of \mathbb{X} at row i , column j , where all the indices start at one. The transpose of \mathbb{X} is denoted by \mathbb{X}^T . The set of real numbers is denoted by \mathbb{R} , the set of natural numbers by \mathbb{N} , and the set of nonnegative integers by \mathbb{Z}^+ .

We denote random variables by capital letters Y , probabilities by $\Pr\{\cdot\}$,

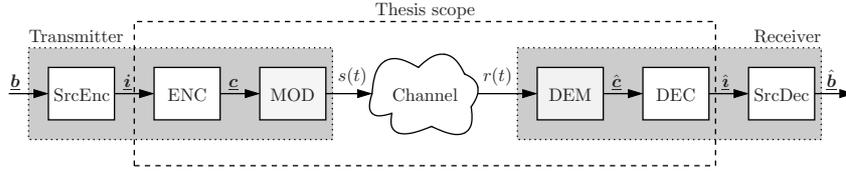


Fig. 1.1: A simplified block diagram of a digital transmitter/receiver pair.

the probability mass function (PMF) of the random vector \mathbf{Y} by $P_{\mathbf{Y}}(\mathbf{y})$, and the probability density function (PDF) of the random vector \mathbf{Y} by $p_{\mathbf{Y}}(\mathbf{y})$. The joint PDF of the random vectors \mathbf{X} and \mathbf{Y} is denoted by $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$ and the conditional PDF of \mathbf{Y} conditioned on $\mathbf{X} = \mathbf{x}$ is denoted by $p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})$. The same notation applies to joint and conditional PMFs, i.e., $P_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$ and $P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})$. The expectation of an arbitrary function $f(\mathbf{X}, \mathbf{Y})$ over the joint PDF of \mathbf{X} and \mathbf{Y} is denoted by $\mathbb{E}_{\mathbf{X},\mathbf{Y}}[f(\mathbf{X}, \mathbf{Y})]$, and the expectation over the conditional PDF $p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})$ is denoted by $\mathbb{E}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}[f(\mathbf{X}, \mathbf{Y})]$. A Gaussian distribution with mean value μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$, the Gaussian PDF with the same parameters by $\Psi(\lambda; \mu, \sigma^2)$, and the Q-function by $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\frac{u^2}{2}) du$.

1.1.2 Thesis Scope

In this thesis, we deal with transmission of binary information over a (wired or wireless) transmission channel. A simplified block diagram of a typical digital transmitter/receiver pair is shown in Fig. 1.1. At the transmitter side, the source encoder (SrcEnc) transforms the sequence of bits $\underline{\mathbf{b}}$ into a new sequence of bits $\underline{\mathbf{i}}$. The main purpose of the source encoder is to reduce redundancies in the sequence $\underline{\mathbf{b}}$. The next task at the transmitter's side—performed by the channel encoder (ENC)—is to convert the bits $\underline{\mathbf{i}}$ into a sequence of coded bits $\underline{\mathbf{c}} = [c[1]^T, \dots, c[N_s]^T]$, where $c[k]$ represents the bits to be transmitted in the k th time interval and N_s is the number of transmitted symbols. Unlike the source encoder, the channel encoder adds redundancy to the bits $\underline{\mathbf{i}}$. This redundancy is added to protect the information sequence $\underline{\mathbf{i}}$ against errors that can occur during transmission.

The last step at the transmitter is to map the bits $c[k]$ into a waveform selected as one of the possible M finite-energy signal alternatives $s_i(t)$ with $i = 1, 2, \dots, M$. Every T_s seconds, the modulator (MOD) selects one of these signal alternatives and transmits it over the channel, where the transmitted waveform $s(t)$ is then a sequence of (possibly overlapping) signal alternatives. The mapping between the bits $c[k]$ and the M signal al-

ternatives is a one-to-one mapping, where the transmitted bits in the k th transmitted signal alternative are grouped in a length- m binary codeword, i.e., $\mathbf{c}[k] \in \{0, 1\}^m$, where m is the number of bits required to index the M signal alternatives, $2^m = M$.

The channel will introduce some random perturbation such that at the receiver's side, the received waveform $r(t)$ will in general be different to the transmitted waveform $s(t)$, i.e., $r(t) \neq s(t)$. The general problem in communications is to *guess*, based only on the channel observation $r(t)$, which signal alternatives $s_i(t)$ were transmitted. This task is performed by the demodulator (DEM), which passes the corresponding decisions $\hat{\mathbf{c}}$ to the channel decoder (DEC). The channel decoder generates an estimate of the information bits $\hat{\mathbf{z}}$. Finally, these bits are passed to the source decoder (SrcDec) which generates an estimate of the transmitted bits $\hat{\mathbf{b}}$, cf. Fig. 1.1. Usually, the transmitter/receiver pair is designed to maximize the transmission rate and minimize the probability of error, under an energy constraint at the transmitter.

This thesis focuses on the channel encoder, modulator, transmission channel, demodulator, and decoder in Fig. 1.1. In very general terms, the thesis deals with the design of these blocks in order to maximize the transmission rate while minimizing the coded bit error probability, under an average energy constraint on the signal alternatives

1.2 Channel Models

1.2.1 The AWGN Channel

The continuous-time additive white Gaussian noise (AWGN) channel is defined as

$$r(t) = s(t) + z(t), \quad (1.1)$$

where $z(t)$ is a white Gaussian random process with power spectral density (PSD) equal to $N_0/2$. This model is justified by the thermal noise (Johnson–Nyquist noise) introduced by electronic components in the receiver [8, Sec. 5.2.2], [9, Sec. 1.3].

The transmission of the sequence of signal alternatives over the channel in (1.1) can be abstracted to the transmission of a sequence of N -dimensional real symbols.¹ More particularly, for each $k = 1, \dots, N_s$, the codeword $\mathbf{C}[k]$ is mapped onto a vector of N amplitudes $\mathbf{X}[k]$ by the mapper Φ .² This mapper is defined as a one-to-one mapping rule $\Phi : \{0, 1\}^m \rightarrow \mathcal{X}$,

¹A detailed explanation of the equivalence between the continuous-time and discrete time models is beyond the scope of this thesis. For more details about this, we refer the reader to standard textbooks like [9, Ch. 3], [10, Ch. 2], [11, Ch. 5].

²From now on, we use \mathbf{C} and \mathbf{X} to explicitly show that the transmitted codeword and symbol are random variables.

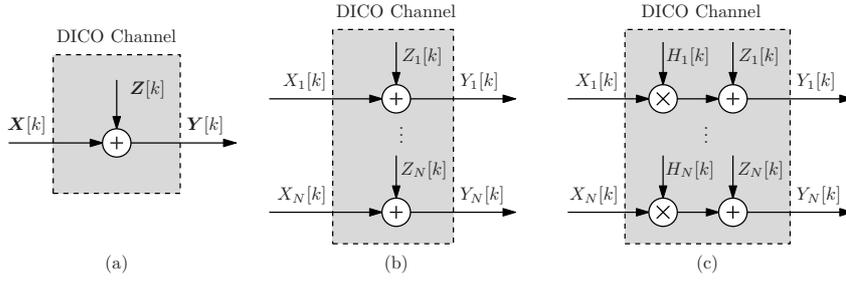


Fig. 1.2: Equivalent discrete-time model for the AWGN channel: without fading in (a) and (b), and with fading in (c).

where each element of $\mathcal{X} \subset \mathbb{R}^N$ is an N -dimensional real vector. The set \mathcal{X} contains all the values that $\mathbf{X}[k]$ can take, it has a cardinality $|\mathcal{X}| = M$, and it is called the input alphabet.

The equivalent discrete-time model for the channel in (1.1) is

$$\mathbf{Y}[k] = \mathbf{X}[k] + \mathbf{Z}[k], \quad (1.2)$$

where $k = 1, \dots, N_s$, $\mathbf{Y}[k] \in \mathbb{R}^N$, $\mathbf{X}[k] \in \mathcal{X}$, and $\mathbf{Z}[k]$ is a vector of i.i.d. Gaussian random variables, each of them with zero mean and variance $N_0/2$. The conditional transition PDF of the channel in (1.2) is given by

$$p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = \frac{1}{(N_0\pi)^{N/2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{x}\|^2}{N_0}\right). \quad (1.3)$$

The model in (1.2) is shown in Fig. 1.2 (a) and Fig. 1.2 (b) as a discrete-input continuous-output (DICO) channel. The model in Fig. 1.2 (b) is obtained because of the independence of the noise samples.

One practically relevant case of (1.2) is obtained when $N = 2$, i.e.,

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad (1.4)$$

where $Z_1, Z_2 \sim \mathcal{N}(0, N_0/2)$.

The model in (1.4) corresponds to so-called quadrature transmitter/receiver, which is the transmitter/receiver used in all the papers in Part II of this thesis. Since in this model all the vectors involved have two dimensions, the input alphabet the noise samples and the received signals are usually represented using complex numbers. This is in fact the model used in the papers in Part II of this thesis. Nevertheless, in the first part

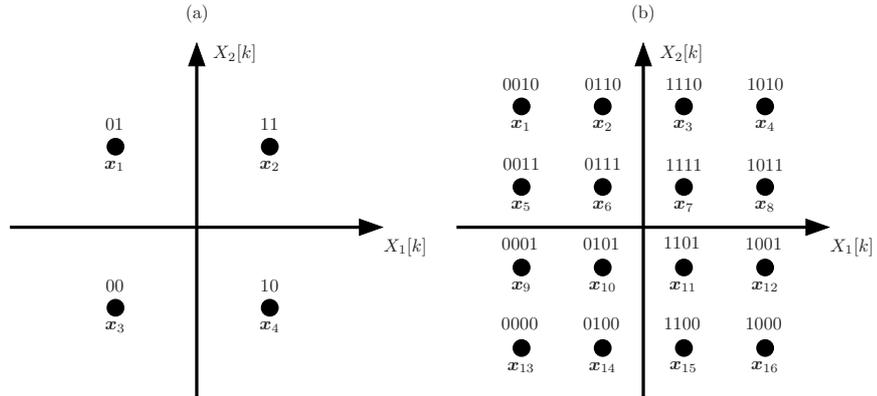


Fig. 1.3: Input alphabets with $N = 2$: (a) $M = 4$ and (b) $M = 16$. The mapper is based on the BRGC.

of this thesis, we refrain from using complex numbers and we use, without loss of generality, the real N -dimensional channel model.

Fig. 1.3 shows two commonly used two dimensional ($N = 2$) input alphabets (with $M = 4$ and $M = 16$). The mapper shown in Fig. 1.3 is based on the binary reflected Gray code (BRGC). More details about input alphabets are given in Sec. 1.3.2. The BRGC is formally defined in Sec. 1.3.1.

1.2.2 Fading Channels

In the presence of fading, the DICO channel for $N = 2$ is modeled as

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = R \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad (1.5)$$

where R is a random channel gain, which is assumed to be constant during the transmission of one symbol and independent between symbols. The PDF of R can be described using the Nakagami- m distribution [11, Sec. 3.2.2], [12] which is parametrized by the fading coefficient $1/2 \leq f < \infty$.³ This model allows us to consider a wide range of channels. For example, the Rayleigh fading channel is obtained when $f = 1$, and the AWGN channel when $f \rightarrow \infty$. The Nakagami- m fading model is used in Paper C.

³We refrain from using m as the parameter of the Nakagami- m PDF because, throughout this thesis, m represents the number of bits per symbols.

The average signal to noise ratio (SNR) is given by

$$\text{SNR} \triangleq \frac{\mathbb{E}_{R, \mathbf{X}}[\|R\mathbf{X}\|^2]}{N_0} = \frac{E_s}{N_0} \mathbb{E}_R[R^2], \quad (1.6)$$

where $E_s \triangleq \mathbb{E}_{\mathbf{X}}[\|\mathbf{X}\|^2]$ is the average symbol energy. For the Nakagami- m fading, the PDF of R is given by

$$p_R(r; f, \text{SNR}) = \frac{2f^f r^{2f-2}}{G(f)\text{SNR}^f} \exp\left(-\frac{fr^2}{\text{SNR}}\right), \quad (1.7)$$

where $G(v+1) \triangleq \int_0^\infty \lambda^v e^{-\lambda} d\lambda$ is the Gamma function. The instantaneous SNR is defined as $\gamma \triangleq R^2(E_s/N_0)$, which has a PDF given by

$$p_\Gamma(\gamma; \text{SNR}, f) = \frac{\gamma^{f-1}}{G(f)} \left(\frac{f}{\text{SNR}}\right)^f \exp\left(-\frac{f\gamma}{\text{SNR}}\right). \quad (1.8)$$

From (1.8), we observe that the instantaneous SNR follows a Gamma distribution. This becomes an exponential distribution when $f = 1$ (or equivalently, a chi-square distribution with one degree of freedom), which gives the Rayleigh fading.

1.2.3 A General Discrete-Time Model

In Sec. 1.2.2, we consider $N = 2$ and a channel gain that affects each dimension equally (the channel gain R), where the fading is Nakagami- m distributed. Inspired by (1.2) and (1.5), we conclude this section by presenting a more general model, which will be used throughout this thesis. In this model, we have

$$\mathbf{Y}[k] = \mathbf{H}[k] \circ \mathbf{X}[k] + \mathbf{Z}[k], \quad (1.9)$$

where $\mathbf{H} = [H_1, \dots, H_N]$ contains real fading coefficients and $H_i \in \mathbb{R}$ are random variables, possibly dependent, with the same PDF $p_H(h)$. This model could represent a fading channel where each component of the transmitted symbol fades independently and with the same distribution and could be used in transmission using multiple antenna elements, OFDM transmission, etc.

The model in (1.9) is presented in Fig. 1.2 (c). Clearly, if $N = 2$ and $R = H_1 = H_2$, the model in (1.9) leads to Rayleigh and Nakagami- m fading channels presented in Sec. 1.2.2. The AWGN channel in (1.2) can be also obtained from this model using $\mathbf{H} = \mathbf{1}$.

As a generalization of (1.6), the average SNR for the model in (1.9) is defined as

$$\text{SNR} \triangleq \frac{\mathbb{E}_{\mathbf{H}, \mathbf{X}}[\|\mathbf{H} \circ \mathbf{X}\|^2]}{N_0} = \frac{E_s}{N_0} \mathbb{E}_H[H^2], \quad (1.10)$$

where H is a random variable with the PDF $p_H(h)$ defined above.

The model in (1.9) represents a fast fading channel where each dimension of the transmitted symbol could fade independently and with the same distribution. Moreover, the conditional transition PDF of the channel in (1.9) is given by

$$p_{Y|X=\mathbf{x}, H=\mathbf{h}}(\mathbf{y}) = \frac{1}{(N_0\pi)^{N/2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{h} \circ \mathbf{x}\|^2}{N_0}\right). \quad (1.11)$$

1.3 Binary Labelings and Signal Sets

1.3.1 Binary Labelings

A binary labeling \mathbb{L} of order $m \in \mathbb{N}$ is defined using a matrix of dimensions $M = 2^m$ by m , where each row corresponds to one of the M length- m distinct binary codewords, $\mathbb{L} \triangleq [\mathbf{c}_1^T, \dots, \mathbf{c}_M^T]^T$, where $\mathbf{c}_i = [c_{i,1}, c_{i,2}, \dots, c_{i,m}] \in \{0, 1\}^m$. In order to recursively define some particular binary labelings, we define *expansions* and *reflections* of binary labelings as follows [13].

- To expand a labeling $\mathbb{L}_m = [\mathbf{c}_1^T, \dots, \mathbf{c}_M^T]^T$ into a labeling \mathbb{L}_{m+1} , we repeat each binary codeword once to obtain a new matrix given by $[\mathbf{c}_1^T, \mathbf{c}_1^T, \dots, \mathbf{c}_M^T, \mathbf{c}_M^T]^T$, and then we obtain \mathbb{L}_{m+1} by appending the extra column $[0, 1, 1, 0, 0, 1, 1, 0, \dots, 0, 1, 1, 0]^T$ of length $2M$.
- To generate a labeling \mathbb{L}_{m+1} from a labeling $\mathbb{L}_m = [\mathbf{c}_1^T, \dots, \mathbf{c}_M^T]^T$ by reflection, we join \mathbb{L}_m and a reversed version of \mathbb{L}_m to obtain a new matrix given by $[\mathbf{c}_1^T, \dots, \mathbf{c}_M^T, \mathbf{c}_M^T, \dots, \mathbf{c}_1^T]^T$, and we add an extra column from the left, consisting of M zeros followed by M ones.

In this thesis, we are particularly interested in the BRGC [13–15], the natural binary code (NBC), and the folded binary code (FBC) [16]. We also introduce a new binary labeling denoted binary semi-Gray code (BSGC) which has some interesting properties to be analyzed in Sec. 3.3. These binary labelings are generated as follows:

- The BRGC of order $m \geq 1$, denoted by \mathbb{G}_m , is generated by $m - 1$ recursive expansions of the trivial labeling $\mathbb{L}_1 = [0, 1]^T$, for any $m \geq 1$.
- The NBC of order m , denoted by \mathbb{N}_m , is defined as

$$\mathbb{N}_m = [\mathbf{b}(0)^T, \dots, \mathbf{b}(M-1)^T]^T, \quad (1.12)$$

where $\mathbf{b}(i) = [b_m(i), b_{m-1}(i), \dots, b_1(i)]$ is the base-2 representation of the integer $0 \leq i \leq M - 1$ and $b_m(i)$ is the most significant bit of i .

- The FBC of order $m \geq 2$, denoted by \mathbb{F}_m , is generated by one reflection of \mathbb{N}_{m-1} .

- The BSGC of order $m \geq 3$, denoted by \mathbb{S}_m , is generated by replacing the first column of \mathbb{G}_m by the modulo-2 sum of the first and last columns.

Example 1 (Binary labelings of order $m = 3$) *The four binary labelings defined above for $m = 3$ are given by the following matrices*

$$\mathbb{G}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \mathbb{N}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \mathbb{F}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \mathbb{S}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

The BRGC (\mathbb{G}_m) will be used in Papers A, C, and D, and the NBC (\mathbb{N}_m) in Paper B. The four labelings above will also be analyzed in Ch. 3.

1.3.2 Input Alphabets

Each element of the input alphabet \mathcal{X} is an N -dimensional symbol \mathbf{x}_i , $i = 1, \dots, M$, where $|\mathcal{X}| = M = 2^m$ and $\mathcal{X} \subset \mathbb{R}^N$. With slight abuse of terminology, we also use the name input alphabet to denote all the elements of \mathcal{X} ordered in a matrix $\mathbb{X} \triangleq [\mathbf{x}_1^T, \dots, \mathbf{x}_M^T]^T$ of dimensions M by N . Throughout this thesis, we use the names input alphabet and signal set interchangeably.

For practical reasons, we are interested in well structured discrete input alphabets. We start by defining three of the most popular M -ary one- and two-dimensional input alphabets, i.e., pulse amplitude modulation (M -PAM), phase-shift keying (M -PSK), and quadrature amplitude modulation (M -QAM). M -PAM and M -QAM signal sets are used in all the papers in Part II.

- An M -PAM signal set is defined by the matrix \mathbb{X}_{PAM} where $x_i = -(M - 2i + 1)$ with $i = 1, \dots, M$.
- An M -PSK signal set is defined by the matrix \mathbb{X}_{PSK} where $\mathbf{x}_i = [\cos(2\pi i/M), \sin(2\pi i/M)]$ with $i = 1, \dots, M$.
- An M -QAM signal set is defined by the direct product of two \sqrt{M} -PAM signal sets.

The last input alphabets we define are the so-called hierarchical PAM (HPAM) and hierarchical QAM (HQAM) signal sets [17–21]. HPAM and

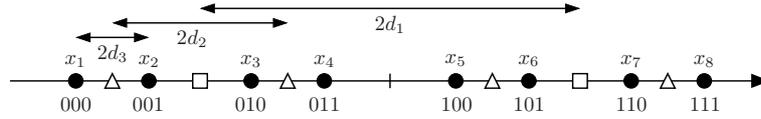


Fig. 1.4: HPAM labeled constellation with $M = 8$.

HQAM signal sets are nonequally spaced input alphabets that preserve certain symmetries and have been standardized for DVB [22]. HPAM signal sets are defined by the matrix \mathbb{X}_{HPAM} , where

$$x_i = \sum_{k=1}^m (-1)^{b_k(i-1)-1} d_k, \quad (1.13)$$

$b_k(i)$ is the base-2 representation of the integer i (see (1.12)), and where $d_k > 0$ for $k = 1, \dots, m$ are the distances defining the input alphabet⁴. The additional condition $x_i < x_{i+1}$ for $i = 1, \dots, M - 1$ is usually imposed so that overlapping points in the input alphabet are avoided. To clarify the definition in (1.13), an example is presented below (Example 2). An HQAM signal set is defined by the direct product of two HPAM signal sets. HPAM and HQAM signal sets are used in Paper C.

For a given input alphabet \mathbb{X} , we define the input distribution of the symbols using the PMF $P_{\mathbf{X}}(\mathbf{x})$, which represents the probabilities of transmitting the symbols \mathbf{x} , i.e., $\Pr\{\mathbf{X} = \mathbf{x}\}$. We define the matrix \mathbb{P} as an ordered list containing the probabilities of the symbols, i.e., $\mathbb{P} \triangleq [P_{\mathbf{X}}(\mathbf{x}_1), \dots, P_{\mathbf{X}}(\mathbf{x}_M)]^T$. The matrix associated with the discrete uniform input distribution is denoted by $\mathbb{U}_M \triangleq [1/M, \dots, 1/M]^T$.

We define a *constellation* as an input alphabet with a given input distribution, i.e., a constellation is defined by the pair $[\mathbb{X}, \mathbb{P}]$. Moreover, for a given constellation $[\mathbb{X}, \mathbb{P}]$, we define a *labeled constellation* as the triplet of matrices $\Omega \triangleq [\mathbb{X}, \mathbb{P}, \mathbb{L}]$. For a given labeled constellation Ω , where $\mathbb{X} = [\mathbf{x}_1^T, \dots, \mathbf{x}_M^T]^T$, $\mathbb{P} = [P_{\mathbf{X}}(\mathbf{x}_1), \dots, P_{\mathbf{X}}(\mathbf{x}_M)]^T$, and $\mathbb{L} = [\mathbf{c}_1^T, \dots, \mathbf{c}_M^T]^T$, we define $\mathcal{I}_{k,u} \subset \{1, \dots, M\}$ as the set of indexes of the symbols with a binary label $u \in \{0, 1\}$ at bit position $k \in \{1, \dots, m\}$, i.e., $\mathcal{I}_{k,u} \triangleq \{i \in \{1, \dots, M\} : c_{i,k} = u\}$.

Example 2 (HPAM Labeled Constellation with $M = 8$) In Fig. 1.4, we show the labeled constellation $\Omega = [\mathbb{X}_{\text{HPAM}}, \mathbb{P}, \mathbb{N}_3]$, i.e., an HPAM signal set with $M = 8$ labeled by the NBC (see Example 1). In this figure, the elements of the input alphabet are shown with black circles, while the

⁴An equivalent definition for (1.13) is $x_i = \sum_{k=1}^m (2b_k(i-1) - 1)d_k$. This expression differs to the one given in [18, eq. (3)] only by a minus sign which is caused by an inverse (right-to-left) enumeration of the constellation symbols.

white squares/triangles represent the HPAM with $M = 2$ and HPAM with $M = 4$ signal sets from which the HPAM with $M = 8$ can be recursively constructed. The selection of the NBC as the binary labeling in this example is motivated by the fact that HPAM input alphabets labeled by the NBC were found to be first order optimal (for asymptotically low SNR) in [23].

Chapter 2

Coded Modulation

Generally speaking, coded modulation refers to the design of the channel encoder and modulator blocks in Fig. 1.1 and its counterparts at the receiver (demodulator and decoder), when the channel conditions are good enough (high SNR) so that more than one bit can be transmitted in each symbol period. In the following sections, we present a brief introduction to CM systems, and we review three of the most popular ways of implementing a CM in practice: trellis-coded modulation, multilevel coding, and bit-interleaved coded modulation (with and without iterative decoding).

2.1 Introduction to Coded Modulation

2.1.1 The Need for Coded Modulation

The problem of reliable transmission of digital information through a noisy channel dates back to the works of H. Nyquist [24, 25] and R. V. L. Hartley [26] almost 90 years ago. Their efforts were capitalized by C. E. Shannon who formulated a unified mathematical theory of communication in 1948 [27, 28].¹ After he introduced the famous capacity formula for the AWGN channel, the problem of designing a system that operates close to that limit has been one of the most important and challenging problems in information/communication theory. Probably, one of the simplest ways of approaching capacity in the so-called power-limited regime (the low SNR region) is by using two signal alternatives and a binary channel encoder of rate $R < 1$ that corrects errors caused by the channel. In this case, by changing the rate of the encoder, the bit rate can be modified at expense of a higher/lower error correction capability. In such a system, the maximum

¹An excellent summary of the contributions that influenced Shannon's work can be found in [29, Sec. 1].

number of bits per symbol is bounded by one ($0 \leq R_c \leq 1$), and therefore, it is not spectrally efficient for good channel conditions, i.e., in the so-called bandwidth-limited regime (the high SNR region). One straightforward answer to the question of how to efficiently transmit more than 1 bit per symbol is a CM scheme, where the channel encoder is connected to a modulator, in which several bits are associated to one channel symbol (like in the system in Fig. 1.1 with $m > 1$). What is not straightforward is how to construct such a system, which operates close to the capacity limit, with a reasonably low complexity.

2.1.2 Historical Overview

After Shannon's work was published in 1948, many efforts were made to design binary channel codes that approach Shannon's limit in the power-limited regime (spectral efficiencies below 1 bit/symbol). More than 40 years had to pass until Berrou *et al.* in 1993 introduced the so-called turbo codes [30], which turned out to approach the channel capacity as no other code did before. A few years after, another family of powerful codes—the low-density parity check (LDPC) codes invented in 1962 by Gallager [31]—was rediscovered by MacKay and Neal [32, 33]. Both turbo codes and LDPC codes entered the commercial world very fast and based on titles of papers like [34], one could say that the problem of how to approach the capacity has actually been solved. This is true as far as binary modulation is concerned, i.e., when $M = 2$.

The early works on CM include those by de Buda [35, 36], Massey [37], Miyakawa *et al.* [38], Anderson and Taylor [39], and Aulin [40]. The first breakthroughs for coding in the bandwidth-limited regime (spectral efficiencies above 1 bit/symbol) came with Ungerboeck's trellis-coded modulation (TCM) [41, 42] and Imai and Hirakawa's multilevel coding (MLC) [43, 44].² They both aim to maximize a Euclidean distance measure, and thus, they perform very well over the AWGN channel. Some years later, with the expansion of wireless communications, channels affected by fading became more and more important. In this type of channels, the so-called "code diversity" is the key performance measure, and not the Euclidean measure used in the design of TCM and MLC.

The next breakthrough for coding in the bandwidth-limited regime came in 1992, when Zehavi introduced the so-called bit-interleaved coded modulation (BICM) [48]. BICM is usually referred to as a pragmatic approach for CM design and increases the code diversity when compared to TCM [49, Table III]. BICM is considered the dominant technique for CM in fading chan-

²For a detailed historical overview of the early works on CM, we refer the reader to [8, Sec. 1.2] and [45, pp. 952–953]. Also, good summaries of the efforts made over the years to approach Shannon's limit in both power- and bandwidth-limited regimes can be found in [46, 47].

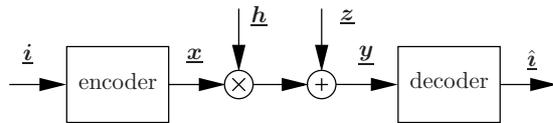


Fig. 2.1: A CM scheme: An encoder, the channel, and the decoder.

nels [11, Sec. 8.8.3] and is also very attractive from an implementation point of view because of its flexibility, i.e., the channel encoder and the modulator can be selected independently, somehow breaking Massey’s joint design paradigm. The flexibility of BICM also allows straightforward implementation of adaptive modulation and coding, as well as hybrid automatic repeat request. When compared with TCM, BICM decreases the minimum Euclidean distance, and consequently, it is suboptimal for the AWGN channel. Nevertheless, BICM’s flexibility is arguably the main advantage of BICM over other CM schemes, and also the reason of why it is used in almost all of the current wireless communications standards, cf. Ch. 1.

2.1.3 A General Model

In this thesis, we analyze CM schemes as the one shown in Fig. 2.1, where the encoder (decoder) in this figure corresponds to the blocks channel encoder and modulator (demodulator and channel decoder) in Fig. 1.1. Each of the 2^{k_c} possible messages is represented by the binary vector $\mathbf{i} \in \{0, 1\}^{k_c}$. The encoder maps each message to a sequence $\mathbf{x} = [\mathbf{x}[1]^T, \dots, \mathbf{x}[N_s]^T] \in \mathcal{X}^{N_s}$, which corresponds to N_s N -dimensional symbols (N_s channel uses³). The code \mathbb{C} is a subset of \mathcal{X}^{N_s} such that $|\mathbb{C}| = 2^{k_c}$, which is used for transmission.⁴ The encoder is then defined as a one-to-one function that assigns each information message \mathbf{i} to one of the 2^{k_c} possible sequences $\mathbf{x} \in \mathbb{C}$. The code rate in information bits per coded bits is then given by $R = k_c/mN_s$ or, equivalently, $R_c = k_c/N_s$ information bits per channel use (information bits per symbol, or information bits per N real dimensions). At the receiver side, based on the channel observations, the decoder generates an estimate of the information bits $\hat{\mathbf{i}}$ selecting the most likely transmitted message.

Each transmitted symbol conveys R_c information bits and thus, the relation between the average symbol energy E_s and the average information

³A “channel use” corresponds to the transmission of one N -dimensional symbol, i.e., it can be considered as a “vectorial channel use”.

⁴We note that the selection of $\mathbf{x} \in \mathcal{X}_s^N$ is in fact not completely general since it assumes the use of the same input alphabet at each time instant. Nevertheless, this model is what it is commonly done in practice and is general enough for the purposes of this thesis.

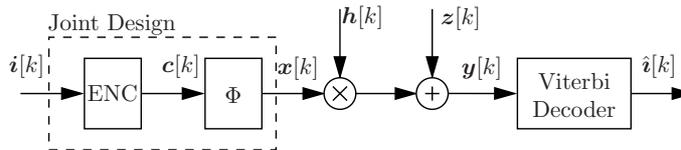


Fig. 2.2: A TCM scheme: A channel encoder followed by a mapper. The decoder is based on the Viterbi algorithm.

bit energy E_b is given by $E_s = R_c E_b$. By using this and (1.10), the average SNR can be expressed as

$$\text{SNR} = \frac{E_b}{N_0} R_c \mathbb{E}_H[H^2] \quad (2.1)$$

In the following sections, we review three common ways of constructing the pair Encoder/Decoder in Fig. 2.1.

2.2 TCM, MLC, and BICM

2.2.1 Trellis-Coded Modulation

TCM was proposed by Ungerboeck in 1982 [42] (see also [50, 51]) as a way of increasing the spectral efficiency of a communication system, and it is based on the concept of mapping by set partitioning.⁵ This is obtained by concatenating a convolutional code and a modulator with an increased number of signal alternatives ($M > 2$), and at the same time, a careful design of both encoder and mapper. A general TCM structure is shown in Fig. 2.2, where we explicitly show that the encoder (usually a binary convolutional code of rate $R = k_c/(mN_s)$) and the mapper Φ should be jointly designed. At the receiver, the decoder uses a Viterbi decoder which finds the most likely coded sequence.

Ungerboeck's famous example in [42] considered the comparison of two schemes giving a spectral efficiency of 2 bit/symbol. The first one is an uncoded system with a 4-PSK input alphabet, and the second one is a rate 2/3 convolutional code combined with an 8-PSK input alphabet. Based on capacity arguments (i.e., assuming unbounded coding/decoding complexity), it was shown that 7 dB of gain (for an uncoded bit error probability of 10^{-5}) could be achieved. In [42], Ungerboeck also proposed a design strategy (based on the so-called set-partitioning principle), and showed that for his famous example with a spectral efficiency of 2 bit/symbol, by using a

⁵In fact, TCM was originally proposed in ISIT 1976 [41].

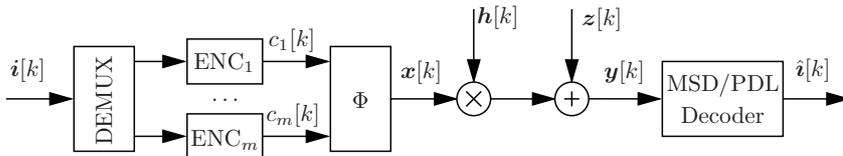


Fig. 2.3: A MLC scheme: a demultiplexer, m parallel encoders, and a mapper. The decoder can be based on MSD or PDL.

convolutional code with constraint length $K = 3$, asymptotic gains of 3 dB are obtained [42, Table II]. Moreover, if a convolutional code with a larger constraint length is used, this gain increases, and for example reaches 5 dB for $K = 8$ (at the expense of a more complex Viterbi decoder).

After Ungerboeck introduced TCM, it became a very popular research topic. Moreover, TCM went rather quickly from research to practice; it was introduced in the modem standards in the early 90s (V.32 [52] and V.32bis [53]) increasing the transmission rates up to 14.4 kbps. TCM is a well studied topic and extensive information about it can be found for example in [54], [9, Sec. 8.12], [55, Ch. 4], [56, Sec. 8.2], [57, Ch. 14], [45, Ch. 18]. TCM for fading channels is studied in [55, Ch. 5].

2.2.2 Multilevel Coding

MLC was proposed by Imai and Hirakawa in 1977 [43, 44]. MLC is another way of constructing a CM system, and it is shown in Fig. 2.3. The MLC encoder is based on the use of a demultiplexer (DEMUX) and m parallel encoders, each of them connected to one of the m bit positions in the mapper.

The decoder for MLC can be based on multistage decoding (MSD) or on parallel decoding of the individual levels (PDL). MLC with MSD was analyzed in detail in [58], and it corresponds to decoding the first code and passing the decoded sequence to the second decoder, which passes information to the third decoder, and so on. The PDL strategy is the same as MSD with the difference that no information is passed between the decoders. PDL can be implemented using decoders running in parallel, and therefore, it introduces less delay than MSD. PDL works particularly well with Gray-labeled signal sets.

In MLC, the selection of the m rates of the codes is crucial. One way of doing this is based on information-theoretic aspects (which we will clarify in Sec. 3.2). Other rules include the “balanced design rule” or the “coding exponent rule”, cf. [58], and the recently proposed “Lagrange multiplier approach” [59]. Finally, MLC for fading channels—which includes

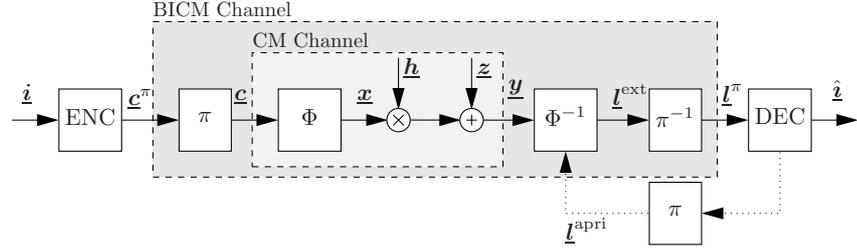


Fig. 2.4: A BICM(-ID) scheme: A binary channel encoder, a bit-level interleaver, a memoryless mapper, the channel, and the inverse processes at the receiver's side. The CM and the BICM channels are also shown.

bit-interleavers in each level—has been proposed in [60] and MLC using capacity-approaching (turbo) codes was proposed in [61].

2.2.3 BICM and BICM-ID

BICM was introduced in 1992 in [48] and later analyzed in detail in [49] and [62]. A general BICM system model is shown in Fig. 2.4. In this system, the encoder in Fig. 2.1 is realized using a serial concatenation of a binary encoder (ENC) of rate $R = k_c/(mN_s)$, a bit level interleaver (π), and a memoryless mapper (Φ). The mapper Φ is defined as a one-to-one mapping rule that maps the length- m binary random vector $\mathbf{C} = [C_1, \dots, C_m]$ to one symbol \mathbf{X} , i.e., $\Phi : \{0, 1\}^m \rightarrow \mathcal{X}$.

At the receiver side, the demapper (Φ^{-1}) computes extrinsic soft information (L-values) on the coded bits. These extrinsic L-values are then deinterleaved (π^{-1}) and passed to the channel decoder (DEC). The a posteriori logarithmic likelihood ratios (L-values) for the k th bit in the symbol are given by

$$l_k(\mathbf{y}) \triangleq \log \frac{\Pr\{C_k = 1 | \mathbf{Y} = \mathbf{y}\}}{\Pr\{C_k = 0 | \mathbf{Y} = \mathbf{y}\}} \quad (2.2)$$

$$= l_k^{\text{ext}}(\mathbf{y}) + l_k^{\text{apri}}, \quad (2.3)$$

where the *extrinsic* and *a priori* L-values are, respectively, defined as

$$l_k^{\text{ext}}(\mathbf{y}) \triangleq \log \frac{\Pr\{\mathbf{Y} = \mathbf{y} | C_k = 1\}}{\Pr\{\mathbf{Y} = \mathbf{y} | C_k = 0\}} \quad (2.4)$$

$$l_k^{\text{apri}} \triangleq \log \frac{\Pr\{C_k = 1\}}{\Pr\{C_k = 0\}}. \quad (2.5)$$

The a priori L-values l_k^{apri} are obtained from the decoder. This is shown with dotted lines in Fig. 2.4. When this feedback loop is present, the BICM system is referred to as BICM with iterative decoding (BICM-ID) and the decoding is performed in an iterative fashion by exchanging information between the decoder and the demapper. BICM-ID was introduced in [63–65] after BICM was recognized as a serial concatenation of codes (the encoder and the mapper) and further studied in [66–70]. Soon after BICM-ID was introduced, the key role of the binary mapping was recognized. Abundant literature exists discussing the design of bit mappings for improving the system performance, for example [67, 68, 71–76]. More details about this will be given in Sec. 4.1.1. BICM-ID is studied in Paper B.

When the feedback loop is not present, i.e., $l_k^{\text{apri}} = 0$, the system is called noniterative BICM, or simply BICM.⁶ In this case, the demapper computes the extrinsic (or equivalently, the a posteriori) L-values, which are passed to the deinterleaver, and then to the decoder, which produces an estimate of the information sequence $\hat{\mathbf{i}}$. BICM is studied in Papers A, C, and D.

By using the relation between the probabilities of a bit U being $u \in \{0, 1\}$ and its corresponding L-value l

$$P_U(u) = \Pr\{U = u\} = \frac{\exp(ul)}{1 + \exp(l)}, \quad (2.6)$$

and assuming independent bits C_k , it can be shown that the extrinsic L-values for the channel in (1.9) (with conditional PDF given by (1.11)) can be expressed as

$$l_k^{\text{ext}}(\mathbf{y}) = \sum_{u \in \{0,1\}} (-1)^{u+1} \log \sum_{i \in \mathcal{I}_{k,u}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{h} \circ \mathbf{x}_i\|^2}{N_0}\right) \prod_{\substack{j=1 \\ j \neq k}}^m P_{C_j}(c_{i,j}) \quad (2.7)$$

$$= \sum_{u \in \{0,1\}} (-1)^{u+1} \log \sum_{i \in \mathcal{I}_{k,u}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{h} \circ \mathbf{x}_i\|^2}{N_0} + \sum_{\substack{j=1 \\ j \neq k}}^m c_{i,j} l_j^{\text{apri}}\right), \quad (2.8)$$

where $c_{i,k}$ is the k th bit of the codeword labeling the symbol \mathbf{x}_i and l_j^{apri} is the a priori information on the coded bits given by (2.5), and to pass from (2.7) to (2.8), we used the fact that the denominator of (2.6) does not depend on u . Equation (2.8) tells us how the extrinsic L-values are calculated in BICM-ID for the k th bit in the symbol, based on both the received symbol \mathbf{y} and the available a priori information l_j^{apri} for $j \neq k$.

When a BICM receiver is used, the demapper Φ^{-1} has no a priori information on the coded bits, i.e., $l_j^{\text{apri}} = 0$ in (2.8). In this case, the extrinsic

⁶Throughout this thesis, we use the name BICM to refer to the noniterative BICM.

L-values are calculated as

$$l_k^{\text{ext}}(\mathbf{y}) = \sum_{u \in \{0,1\}} (-1)^{u+1} \log \sum_{i \in \mathcal{I}_{k,u}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{h} \circ \mathbf{x}_i\|^2}{N_0}\right). \quad (2.9)$$

This expression shows how to compute the L-values in their exact form, which depends on the received signal \mathbf{y} , the input alphabet and its binary labeling, and the noise variance N_0 . Since in BICM $l_k^{\text{ext}}(\mathbf{y}) = l_k(\mathbf{y})$, for notation simplicity, from now on, we will only use $l_k(\mathbf{y})$ to refer to the L-values computed by the demapper in BICM.

The computation of $l_k(\mathbf{y})$ in (2.9) involves computing the logarithm of a sum of exponential functions, where the number of exponential functions grows exponentially with m . When the constellation size increases, the computation of $l_k(\mathbf{y})$ in (2.9) becomes more complex because of the non-linear functions involved, and therefore, approximations are sought. One of the most common approximations used in practice is the so-called max-log approximation [77]

$$\log \sum_i \exp \lambda_i \approx \max_i \lambda_i. \quad (2.10)$$

Using (2.10) in (2.9), the approximated L-values can be expressed as

$$\tilde{l}_k(\mathbf{y}) = \frac{1}{N_0} \sum_{u \in \{0,1\}} (-1)^u \min_{i \in \mathcal{I}_{k,u}} \|\mathbf{y} - \mathbf{h} \circ \mathbf{x}_i\|^2. \quad (2.11)$$

Although the max-log metric in (2.11) is suboptimal, it is very popular in practical implementations of the demapper because of its low complexity. This simplification, proposed already in [48, 49], is frequently adopted for ease of the resulting implementation, e.g., by the 3GPP working groups [78]. It is also known that when Gray-labeled input alphabets are used, the use of this simplification results in a negligible impact on the receiver's performance [79, Fig. 9], [80]. Also, we note from (2.10) that the max-log approximation is tighter when there is one dominant term among the exponential functions. As a consequence of this, that the approximation $\tilde{l}_k(\mathbf{y}) \approx l_k(\mathbf{y})$ will be tight when one of the squared Eclidean distances $\|\mathbf{y} - \mathbf{h} \circ \mathbf{x}_i\|^2$ is very small compared to the others, i.e., for high SNR values.

A desired feature of the max-log approximation is that the dependence of the L-values on N_0 essentially disappears, i.e., since all the L-values in (2.11) are scaled by $1/N_0$, this scaling factor can be removed. An additional desired feature of the max-log approximation is that it transforms the nonlinear relation between $l_k(\mathbf{y})$ and \mathbf{y} in (2.9) into a piecewise linear relation, cf. [81, 82]. This has been used to develop expressions for the PDF of the L-values in (2.11). More details about this are given in Sec. 4.1.2.

It is important to mention that the max-log approximation given by (2.10) can also be used in the decoding algorithm. When the maximum a posteriori probability (MAP) algorithm is implemented in the logarithmic domain (Log-MAP) using the Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm, computations involving logarithms of sum of terms appear very often. Consequently, the max-log approximation can be used to reduce the decoding complexity, which yields the so-called Max-Log-MAP algorithm [45, 83]. Different approaches for simplified metric calculation—most of them based on the Jacobi logarithm [77, 84]—have been investigated in the literature for both the decoding algorithm and the metric calculation in the demapper. The design of decoding algorithms is however beyond the scope of this thesis. For more details about this topic, the reader is referred to [79, 85, 86].

We conclude this chapter by noting that BICM-ID is a particular case of a more general concept called iterative decoding (also referred as a turbo decoding). Turbo decoding is a process when different elements in the receiver exchange information in an iterative fashion in order to improve some parameter estimation. Although the concept of iterative decoding dates back to 1954 with the work of Elias [83], a major breakthrough came when Berrou, Glavieux, and Thitimajshima introduced the so-called turbo codes in [30]. The original turbo codes were formed by a parallel concatenation of two recursive systematic convolutional codes (RSCC) separated by an interleaver. At the receiver, both decoders exchange information in an iterative fashion, continuously improving the decisions about the transmitted bits. After turbo codes were introduced in 1993, the concept of iterative processing has been extended to different parts of the receiver. Nowadays it is common to design iterative receivers with algorithms for turbo synchronization, turbo equalization, turbo channel estimation, turbo multi-user detection, etc.

An important property of BICM-ID, is that its performance in terms of bit-error rate (BER) can be divided into two distinct regions: an *early convergence* region (also called waterfall region) and an *error-floor* region.⁷ The early convergence region is the SNR range where a small increase in the SNR value will produce a substantial decrease in the BER (steep BER curve). The error-floor region, instead, is the SNR range where no substantial improvements are obtained by increasing the SNR (flat BER curve).

⁷We recognize that the error floor in BICM-ID can be decreased or even completely removed if particular techniques are used, for example, linear precoders [87, 88], code/modulation doping [67, 69], or interleaver optimization [89, 90].

Chapter 3

Capacity Aspects of Coded Modulation

The so-called *channel capacity* is a numerical value that represents a fundamental limit for the maximum transmission rate supported by a given channel. In the following subsections, we analyze this fundamental limit for the coded modulation systems presented in Chapter 2. We pay particular attention to the so-called BICM capacity, which is the channel capacity when the transmitter and receiver are constrained to follow a structure shown in Fig. 2.4.

3.1 AMI and Channel Capacity

In this section, we assume a continuous input alphabet, i.e., $\mathcal{X} = \mathbb{R}^N$ and the input symbols are selected with PDF $p_{\mathbf{X}}(\mathbf{x})$, which upperbounds the performance of finite input alphabets.

The average mutual information (AMI) in bits¹ per channel use between the random vectors \mathbf{X} and \mathbf{Y} when the channel \mathbf{H} is perfectly known at the receiver is defined as

$$I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}) \triangleq \mathbb{E}_{\mathbf{X}, \mathbf{Y}} \left[\log_2 \frac{p_{\mathbf{X}, \mathbf{Y}}(\mathbf{X}, \mathbf{Y})}{p_{\mathbf{Y}}(\mathbf{Y})p_{\mathbf{X}}(\mathbf{X})} \right], \quad (3.1)$$

where we use \mathbf{X} as the index of $I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y})$ to emphasize that the AMI depends on the input PDF $p_{\mathbf{X}}(\mathbf{x})$.

For an arbitrary channel parameter \mathbf{H} , the AMI in (3.1) can be expressed

¹Throughout this thesis all the AMIs are given in bits.

as

$$I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}) = \mathbb{E}_{\mathbf{H}} [I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}|\mathbf{H})] \quad (3.2)$$

$$= \mathbb{E}_{\mathbf{X}, \mathbf{Y}, \mathbf{H}} \left[\log_2 \frac{p_{\mathbf{Y}|\mathbf{X}, \mathbf{H}}(\mathbf{Y})}{p_{\mathbf{Y}|\mathbf{H}}(\mathbf{Y})} \right] \quad (3.3)$$

$$= \int_{\mathbb{R}^N} p_{\mathbf{H}}(\mathbf{h}) \int_{\mathbb{R}^N} p_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^N} p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}, \mathbf{H}=\mathbf{h}}(\mathbf{y}) \cdot \log_2 \frac{p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}, \mathbf{H}=\mathbf{h}}(\mathbf{y})}{p_{\mathbf{Y}|\mathbf{H}=\mathbf{h}}(\mathbf{y})} d\mathbf{y} d\mathbf{x} d\mathbf{h}, \quad (3.4)$$

where $p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}, \mathbf{H}=\mathbf{h}}(\mathbf{y})$ is given by (1.11) and $I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}|\mathbf{H})$ in (3.2) represents the AMI between \mathbf{X} and \mathbf{Y} for a given value of \mathbf{H} .

The expressions in (3.2)–(3.4) state that, when the channel is perfectly known at the receiver, the AMI $I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y})$ in (3.1) is an average of the AMI of an AWGN channel (with gain $\mathbf{H} = \mathbf{h}$) over the distribution of \mathbf{H} . We also recognize that the AMI with perfect channel state information $\mathbb{E}_{\mathbf{H}} [I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}|\mathbf{H})]$ is usually denoted by $I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}|\mathbf{H})$ (see for example [91, eq. (3)]); however, to be consistent with our previously defined notation, we use $I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y})$.

The *channel capacity* of a continuous-input continuous-output memoryless channel, when the channel is perfectly known at the receiver, is defined as the maximum AMI between its input and output [91, eq. (3)], [92, Ch. 4]

$$C(\text{SNR}) \triangleq \max_{p_{\mathbf{X}}(\mathbf{x})} I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}), \quad (3.5)$$

where the maximization is over all possible input distributions. The capacity in (3.5) has units of [bit/channel use] (or equivalently [bit/symbol]), and it is an upper bound on the number of bits per symbol that can be reliably transmitted through the channel, where a symbol consists of N real dimensions.

Shannon's channel coding theorem states that it is not possible to transmit information reliably above this fundamental limit, i.e.,

$$R_c \leq C(\text{SNR}) = C \left(R_c \mathbb{E}_{\mathbf{H}} [H^2] \frac{E_b}{N_0} \right), \quad (3.6)$$

where we used (2.1).

In order to analyze the behavior of the capacity curves in the low SNR regime, it is common to plot them versus E_b/N_0 instead of versus SNR. From (3.6), we can see that E_b/N_0 is bounded from below by a function $f(R_c)$, i.e.,

$$\frac{E_b}{N_0} \geq f(R_c) \triangleq \frac{1}{\mathbb{E}_{\mathbf{H}} [H^2]} \frac{C^{-1}(R_c)}{R_c}. \quad (3.7)$$

The function $f(R_c)$ always exists because the capacity $C(\text{SNR})$ is a strictly increasing function², and thus, $C(\text{SNR})$ is invertible.

Example 3 (Capacity of the AWGN Channel) For the AWGN channel in (1.2) ($\mathbf{H} = \mathbf{1}$), since the noise is independent in each dimension, the transmission of \mathbf{X} can be considered as a transmission through N parallel independent Gaussian channels (see Fig. 1.2 (b)), i.e., $Y_n = X_n + Z_n$, where Z_n are i.i.d. zero-mean Gaussian random variables with variance $N_0/2$, with $n = 1, \dots, N$.

The AWGN capacity, denoted by $C^{\text{AW}}(\text{SNR})$, is defined as the channel capacity of the AWGN channel in (1.2), and it is given by [92, Sec. 9.4]

$$C^{\text{AW}}(\text{SNR}) = \frac{N}{2} \log_2 \left(1 + \frac{2}{N} \text{SNR} \right), \quad (3.8)$$

which is obtained when \mathbf{X} are i.i.d. zero-mean Gaussian random variables with variance E_s/N in each dimension.

The expressions in (3.8) and (3.7) allow us to find a lower bound on the E_b/N_0 for zero-rate transmission over the AWGN channel, i.e.,

$$\lim_{R_c \rightarrow 0^+} f^{\text{AW}}(R_c) = -1.59 \text{ dB}, \quad (3.9)$$

which we refer to as the Shannon limit (SL).

We conclude this section by defining the AMI for discrete input alphabets. The AMI in this case is (3.4) where the middle integral is replaced by a sum, and the PDF $p_{\mathbf{X}}(\mathbf{x})$ by the PMF $P_{\mathbf{X}}(\mathbf{x})$, i.e.,

$$I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}) = \int_{\mathbb{R}^N} p_{\mathbf{H}}(\mathbf{h}) \sum_{i=1}^M P_{\mathbf{X}}(\mathbf{x}_i) \int_{\mathbb{R}^N} p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}_i, \mathbf{H}=\mathbf{h}}(\mathbf{y}) \cdot \log_2 \frac{p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}_i, \mathbf{H}=\mathbf{h}}(\mathbf{y})}{p_{\mathbf{Y}|\mathbf{H}=\mathbf{h}}(\mathbf{y})} d\mathbf{y} d\mathbf{h}, \quad (3.10)$$

where again the notation $I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y})$ emphasizes the dependence of the AMI on the input distribution $P_{\mathbf{X}}(\mathbf{x})$.

In this following two sections, we analyze the capacity of CM schemes, i.e., the so-called CM and BICM capacities. We review their relation and we analyze how the selection of the input alphabet influences them. We pay special attention to the selection of the binary labeling and the use of probabilistic shaping in BICM.

²This can be proved using the relation between the AMI and the minimum mean square error (MMSE) presented in [93], i.e., that the derivative of the AMI with respect to SNR is proportional to the MMSE for any SNR. Since the MMSE is a strictly decreasing function of SNR, the AMI is a strictly increasing function of SNR.

3.2 Coded Modulation Capacity

The CM capacity is defined as the AMI between \mathbf{X} and \mathbf{Y} for a given labeled constellation Ω , i.e.,

$$I_{\Omega}^{\text{CM}}(\text{SNR}) \triangleq I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}) \quad (3.11)$$

$$= I_{\mathbf{X}}(\mathbf{C}; \mathbf{Y}) \quad (3.12)$$

$$= \sum_{k=1}^m I_{\mathbf{X}}(C_k; \mathbf{Y} | C_1, \dots, C_{k-1}), \quad (3.13)$$

where to pass from (3.11) to (3.12), we used the fact that the mapping rule between \mathbf{C} and \mathbf{X} is one-to-one. To pass from (3.12) to (3.13) we have used the chain rule of mutual information [92, Sec. 2.5], where we use $I_{\mathbf{X}}(C_k; \mathbf{Y} | C_1, \dots, C_{k-1})$ to represent a *bit level* AMI. This bit level AMI represents the maximum rate that can be used at the k th bit position, given a perfect knowledge of the previous k bits.

The CM capacity in (3.11) corresponds to the capacity of the memoryless “CM channel” in Fig. 2.4 for a given labeled constellation Ω . We note that different binary labelings will produce different values of $I_{\mathbf{X}}(C_k; \mathbf{Y} | C_1, \dots, C_{k-1})$ in (3.13), however, the overall sum will remain constant, i.e., the CM capacity does not depend on the binary labeling. We use the name “CM capacity” for $I_{\Omega}^{\text{CM}}(\text{SNR})$ in (3.11) following the standard terminology³ used in the literature (cf. [49, 96–99]); however, we recognize the misuse of the word capacity since no optimization over the input distribution is performed (cf. (3.5)). Moreover, it is also possible to optimize the input alphabet in order to obtain an increase in the AMI (so-called signal shaping, see [100–103] and references therein). Nevertheless, throughout this thesis we will refer to the AMI for a given Ω in (3.11) as the CM capacity.

The CM capacity in (3.11), for a given labeled constellation Ω , is an upper bound on the number of bits per symbol that can be reliably transmitted using for example TCM (Fig. 2.2) or MLC with multistage decoding (MLC-MSD) (Fig. 2.3). MLC-MSD is in fact a direct application of the summation in (3.13), i.e., m parallel encoders are used, each of them having a rate $R_k \leq I_{\mathbf{X}}(C_k; \mathbf{Y} | C_1, \dots, C_{k-1})$. At the receiver side, the first bit level is decoded and the decisions are passed to the second decoder, which then passes the decisions to the third decoder, and so on.

In case we are interested in optimal labeled constellations, in analogy with (3.5), it is possible to define the maximum coded modulation capacity

³Sometimes, this is also called joint capacity [94], or (constellation) constrained capacity [76, 95].

as

$$C^{\text{CM}}(\text{SNR}) \triangleq \max_{\Omega} I_{\Omega}^{\text{CM}}(\text{SNR}) \quad (3.14)$$

$$= \max_{[\mathbb{X}, \mathbb{P}]} \sum_{k=1}^m I_{\mathbf{X}}(C_k; \mathbf{Y} | C_1, \dots, C_{k-1}). \quad (3.15)$$

As mentioned before, the CM capacity *does not* depend on the binary labeling, i.e., it does not depend on how the mapping rule Φ is implemented, and therefore, in (3.15) we only show two optimization parameters: the input alphabet and the input distribution.

The maximum CM capacity $C^{\text{CM}}(\text{SNR})$ in (3.15) represents an upper bound on the number of bits per symbol that can be reliably transmitted using a fully optimized system, i.e., a system where for each SNR value, the input alphabet and the input distribution are selected in order to maximize the CM capacity $I_{\Omega}^{\text{CM}}(\text{SNR})$.

3.3 Capacity of BICM Systems

In this section, we are interested in the BICM capacity, i.e., the maximum number of bits per symbol that can be reliably transmitted when a BICM scheme is used.

3.3.1 BICM with Arbitrary Input Distributions

It is commonly assumed that the sequence \mathbf{c}^{π} generated by the binary encoder in Fig. 2.4 is infinitely long and symmetric⁴, and also that the interleaver (π) operates over this infinite sequence, simply permuting it in a pseudo-random fashion.⁵ Under these assumptions, the input symbol distribution in BICM will be always $\mathbb{P} = \mathbb{U}_M$. We are interested in analyzing a more general setup where the input symbol distribution can be modified, and thus, we develop a more general model in which we relax the equiprobable input distribution assumption.

Let $C_k \in \{0, 1\}$ be the binary random variable representing the bits at the k th modulator's input, where the PMF $P_{C_k}(u)$ represents the probability of transmitting a bit u at bit position k . We assume that in general $\sum_{k=1}^m P_{C_k}(0) \neq \sum_{k=1}^m P_{C_k}(1)$, i.e., the coded and interleaved sequence could have more zeros than ones (or vice versa). Note that since $P_{C_k}(u)$ is a PMF, $P_{C_k}(0) + P_{C_k}(1) = 1$.

⁴A symmetric sequence is defined as a sequence of bits that contains equal number of zeros and ones.

⁵The pseudo-random sequence is assumed to be known at the receiver so that the de-interleaving can be performed.

Let $\mathbf{c}_i = [c_{i,1}, \dots, c_{i,m}]$ be the binary label of the symbol \mathbf{x}_i . We assume that the bits at the input of the modulator are independent, and therefore, the input symbol probabilities are

$$P_{\mathbf{X}}(\mathbf{x}_i) = \prod_{k=1}^m P_{C_k}(c_{i,k}). \quad (3.16)$$

The independence condition on the coded bits that results in (3.16) can be obtained if the interleaver block in Fig. 2.4 completely breaks the temporal correlation of the coded bits. The condition that the coded and interleaved sequence could be nonsymmetric (i.e., a sequence with more zeros than ones, or vice-versa) can be obtained for example by using an encoder with nonuniform outputs, or by a particular puncturing scheme applied to the coded bits. This can be combined with the use of a deterministic multiplexer and multiple interleavers (more details about this can be found in Paper A, and C), which would allow $P_{C_k}(u) \neq 1/2$. Examples of how to construct a BICM scheme where nonuniform input symbol distributions are obtained include the “shaping encoder” of [104, 105], and the nonuniform signaling scheme based on a Huffman code of [106].

For future use, we define the conditional input symbol probabilities, conditioned on the k th bit being u , as

$$P_{\mathbf{X}|C_k=u}(\mathbf{x}_i) = \begin{cases} \prod_{\substack{k'=1 \\ k' \neq k}}^m P_{C_{k'}}(c_{i,k'}), & \text{if } c_{i,k} = u \\ 0, & \text{if } c_{i,k} \neq u \end{cases} \\ = \begin{cases} \frac{P_{\mathbf{X}}(\mathbf{x}_i)}{P_{C_k}(u)}, & \text{if } i \in \mathcal{I}_{k,u} \\ 0, & \text{if } i \notin \mathcal{I}_{k,u} \end{cases}, \quad (3.17)$$

where $\mathcal{I}_{i,k}$ are the sets of indexes defining a labeled constellation defined in Sec. 1.3.2.

3.3.2 BICM Capacity

In order to define the BICM capacity, which represents the capacity of the “BICM channel” in Fig. 2.4, we use the equivalent channel model proposed in [49]. This model replaces the BICM channel by m parallel memoryless binary-input continuous-output (BICO) channels [49, Fig. 3]. Using this model, the BICM capacity for a given labeled constellation Ω is defined as

$$I_{\Omega}^{\text{BI}}(\text{SNR}) \triangleq \sum_{k=1}^m I_{C_k}(C_k; \mathbf{Y}). \quad (3.18)$$

The AMI of each BICO channel $I_{C_k}(C_k; \mathbf{Y})$ in (3.18), in contrast to (3.15), it is not conditioned on the previous bit values C_1, \dots, C_{k-1} . As a consequence of this, unlike the CM capacity, the binary labeling strongly affects the BICM capacity $I_{\Omega}^{\text{BI}}(\text{SNR})$ in (3.18).

The BICM capacity in (3.18) can be expressed as

$$I_{\Omega}^{\text{BI}}(\text{SNR}) = \sum_{k=1}^m \mathbb{E}_{C_k, \mathbf{H}, \mathbf{Y}} \left[\log_2 \frac{p_{\mathbf{Y}|C_k, \mathbf{H}}(\mathbf{Y})}{p_{\mathbf{Y}|\mathbf{H}}(\mathbf{Y})} \right] \quad (3.19)$$

$$= \sum_{k=1}^m \sum_{u \in \{0,1\}} P_{C_k}(u) \mathbb{E}_{\mathbf{H}, \mathbf{Y}|C_k=u} \left[\log_2 \frac{p_{\mathbf{Y}|\mathbf{H}, C_k=u}(\mathbf{Y})}{p_{\mathbf{Y}|\mathbf{H}}(\mathbf{Y})} \right] \quad (3.20)$$

$$= \sum_{k=1}^m \sum_{u \in \{0,1\}} P_{C_k}(u) \sum_{i=1}^M P_{\mathbf{X}|C_k=u}(\mathbf{x}_i) \cdot \mathbb{E}_{\mathbf{H}, \mathbf{Y}|\mathbf{X}=\mathbf{x}_i, C_k=u} \left[\log_2 \frac{p_{\mathbf{Y}|\mathbf{H}, C_k=u}(\mathbf{Y})}{p_{\mathbf{Y}|\mathbf{H}}(\mathbf{Y})} \right] d\mathbf{h} \quad (3.21)$$

$$= \sum_{k=1}^m \sum_{u \in \{0,1\}} \sum_{i \in \mathcal{I}_{k,u}} P_{\mathbf{X}}(\mathbf{x}_i) \int_{\mathbb{R}^N} p_{\mathbf{H}}(\mathbf{h}) \cdot \mathbb{E}_{\mathbf{Y}|\mathbf{H}=\mathbf{h}, \mathbf{X}=\mathbf{x}_i, C_k=u} \left[\log_2 \frac{p_{\mathbf{Y}|\mathbf{H}=\mathbf{h}, C_k=u}(\mathbf{Y})}{p_{\mathbf{Y}|\mathbf{H}=\mathbf{h}}(\mathbf{Y})} \right] d\mathbf{h} \quad (3.22)$$

$$= \sum_{k=1}^m \sum_{u \in \{0,1\}} \sum_{i \in \mathcal{I}_{k,u}} P_{\mathbf{X}}(\mathbf{x}_i) \int_{\mathbb{R}^N} p_{\mathbf{H}}(\mathbf{h}) \int_{\mathbb{R}^N} p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}_i, \mathbf{H}=\mathbf{h}}(\mathbf{y}) \cdot \log_2 \frac{\frac{1}{P_{C_k}(u)} \sum_{j \in \mathcal{I}_{k,u}} P_{\mathbf{X}}(\mathbf{x}_j) p_{\mathbf{Y}|\mathbf{H}=\mathbf{h}, \mathbf{X}=\mathbf{x}_j}(\mathbf{y})}{\sum_{\mathbf{x} \in \mathcal{X}} P_{\mathbf{X}}(\mathbf{x}) p_{\mathbf{Y}|\mathbf{H}=\mathbf{h}, \mathbf{X}=\mathbf{x}}(\mathbf{y})} d\mathbf{y} d\mathbf{h}, \quad (3.23)$$

where to pass from (3.20) to (3.21) we used the law of total probability applied to expectations, and to pass from (3.21) to (3.22) we used (3.17) and we also expanded the expectation $\mathbb{E}_{\mathbf{H}, \mathbf{Y}|\mathbf{X}=\mathbf{x}, C_k=u}[\cdot]$ over \mathbf{H} . Finally, to pass from (3.22) to (3.23), we used (3.17) again, and also the fact that the value of C_k does not affect the conditional channel transition probability, i.e., $p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}, \mathbf{H}=\mathbf{h}, C_k=u}(\mathbf{y}) = p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}, \mathbf{H}=\mathbf{h}}(\mathbf{y})$.

The BICM capacity is equivalent to the capacity achieved by MLC based on PDL (Fig. 2.3), i.e., when no information is passed between the m decoders, cf. Sec. 2.2.2. In BICM, the bits are treated as independent, and therefore, BICM is somehow analogous to MLC-PDL. The differences are that BICM uses only one encoder, and that in BICM the equivalent channels are not used in parallel, but time multiplexed [58].

Again, following the standard terminology⁶ used in the literature (cf. [49,

⁶It is also called parallel decoding capacity in [94], or receiver constrained capacity in [76].

96–99]), we use the name “BICM capacity” even though no optimization over the input distribution is performed (nor over the input alphabet or the binary labeling).

The BICM capacity in (3.23) is a general expression that depends on all the labeled constellation parameters Ω . In the Appendix we show how to efficiently evaluate the CM and BICM capacities numerically using Gauss–Hermite quadratures for the AWGN channel. An alternative method for this is to use a one-dimensional integration based on the PDF of the L-values developed in [107–110].

Example 4 (BICM Capacity for $\mathbb{P} = \mathbb{U}_M$ and $\mathbf{H} = \mathbf{1}$) *If all the bits at the input of the BICO channels are equally likely, i.e., $P_{C_k}(u) = 1/2$ for $k = 1, \dots, m$ and $u \in \{0, 1\}$, we obtain from (3.16) $P_{\mathbf{X}}(\mathbf{x}) = 1/M$ for all $\mathbf{x} \in \mathcal{X}$. Under these constraints, the BICM capacity for the AWGN channel in (3.23) for the labeled constellation $\Omega = [\mathbb{X}, \mathbb{U}_M, \mathbb{L}]$ is given by*

$$I_{\Omega}^{\text{BI}}(\text{SNR}) = \frac{1}{M} \sum_{k=1}^m \sum_{u \in \{0,1\}} \sum_{i \in \mathcal{I}_{k,u}} \int_{\mathbb{R}^N} p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}_i}(\mathbf{y}) \cdot \log_2 \frac{2 \sum_{j \in \mathcal{I}_{k,u}} p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}_j}(\mathbf{y})}{\sum_{\mathbf{x} \in \mathcal{X}} p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})} d\mathbf{y}, \quad (3.24)$$

This expression coincides with the most common BICM capacity formula found in the literature, cf. [62, Sec. 3.2.1], [49, eq. (14)], [96, eq. (5)], [97, eq. (11)].

3.3.3 Maximum BICM Capacity

Example 4 raises the question about the optimum binary labeling from a capacity maximization point of view. Once this question is answered, approaching the fundamental limit will depend only on a good design of the channel encoder/decoder. Caire *et al.* conjectured the optimality of the BRGC [49, Sec. III-C], which was first disproved in [111], where an exhaustive search for M -PAM labelings up to $M = 8$ was carried out. The NBC was recently shown to be the optimal labeling for M -PAM and asymptotically low SNR [23, 98].

The results about optimal binary labelings suggest an even more general question: What are the optimal labeled constellations for BICM at a given SNR? To formalize this question, and in analogy to the maximum CM capacity in (3.14), we define the maximum BICM capacity as

$$C^{\text{BI}}(\text{SNR}) \triangleq \max_{\Omega} I_{\Omega}^{\text{BI}}(\text{SNR}), \quad (3.25)$$

where the optimization is in this case over the three parameters defining Ω .

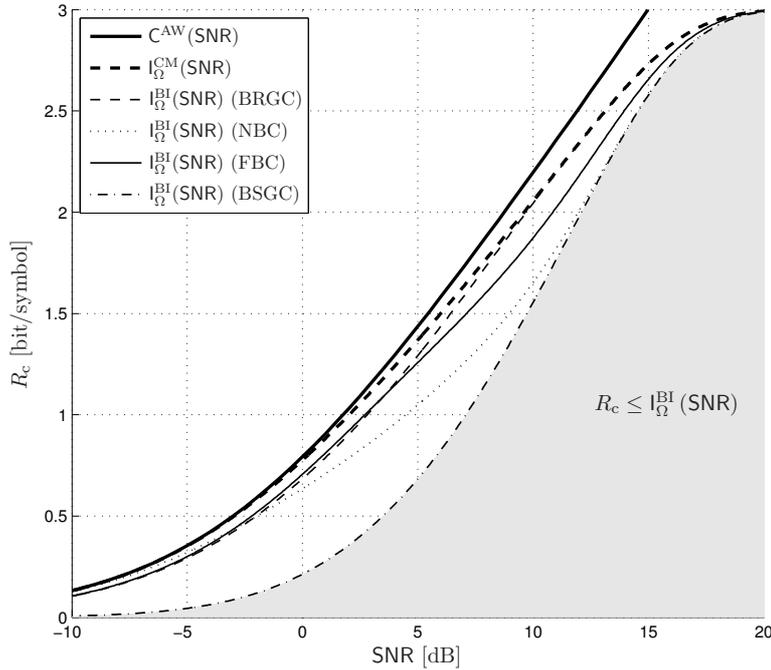


Fig. 3.1: CM capacity and BICM capacity for 8-PAM with \mathbb{U}_8 using the four different labelings defined in Example 1 for the AWGN channel. The shadowed region represents the achievable rates for $I_{\Omega}^{BI}(SNR)$ using the BSGC, cf. (3.6).

In analogy to the maximum CM capacity, the maximum BICM capacity in (3.25) represents an upper bound on the number of bits per symbol that can be reliably transmitted using a fully optimized BICM system, i.e., a system where, for each SNR, the labeled constellation is selected in order to maximize the BICM capacity $I_{\Omega}^{BI}(SNR)$. In general, the problem in (3.25) does not have a closed-form solution, even for the simplest cases. The particular case of $SNR \rightarrow 0$ has been recently analyzed in detail in [23]. In the following examples, we discuss the maximum BICM capacity $C^{BI}(SNR)$ under different constraints.

Example 5 (Binary labelings versus SNR) Fig. 3.1 shows the BICM capacity in (3.24) and the CM capacity in (3.11) for 8-PAM ($M = 8$, $m = 3$, $N = 1$), $\mathbb{P} = \mathbb{U}_8$, and for the four binary labelings in Example 1. From these curves we can see that the difference between the CM capacity and the

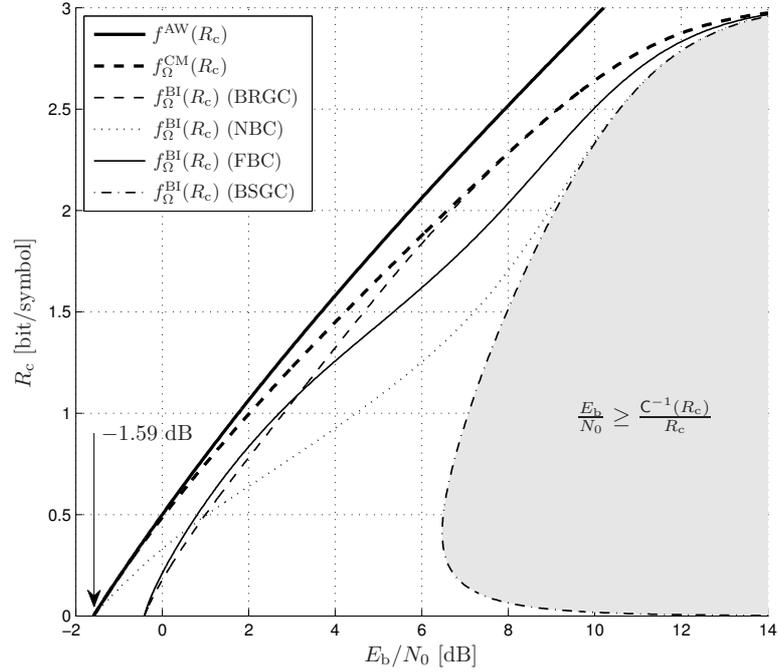


Fig. 3.2: CM capacity and BICM capacity versus E_b/N_0 (cf. (3.7)) for the same cases as in Fig. 3.1. The shadowed region represents the achievable rates for $I_0^{\text{BI}}(\text{SNR})$ using the BSGC given by (3.6).

BICM capacity is small if the binary labeling is properly selected. The best of the four binary labelings is the NBC for low SNR ($R_c \leq 0.43$ bit/symbol), the FBC for medium SNR ($0.43 \leq R_c \leq 1.09$ bit/symbol), and the BRGC for high SNR ($R_c \geq 1.09$ bit/symbol). On the other hand, the gap between the CM capacity and the BICM capacity for the BSGC is quite large. Fig. 3.1 shows that the BRGC is suboptimal for at least approximately 36% of the range of R_c . The FBC was analyzed in [16] for uncoded transmission.

Example 6 (Binary labelings versus E_b/N_0) In Fig. 3.2, we show the CM capacity and BICM capacity versus E_b/N_0 (based on (3.7)), for the same cases as in Fig. 3.1. The SL given by (3.9) is also shown. This figure clearly shows how the NBC and the FBC outperform the BRGC in the low SNR region and it also shows a particular property of the BSGC. When

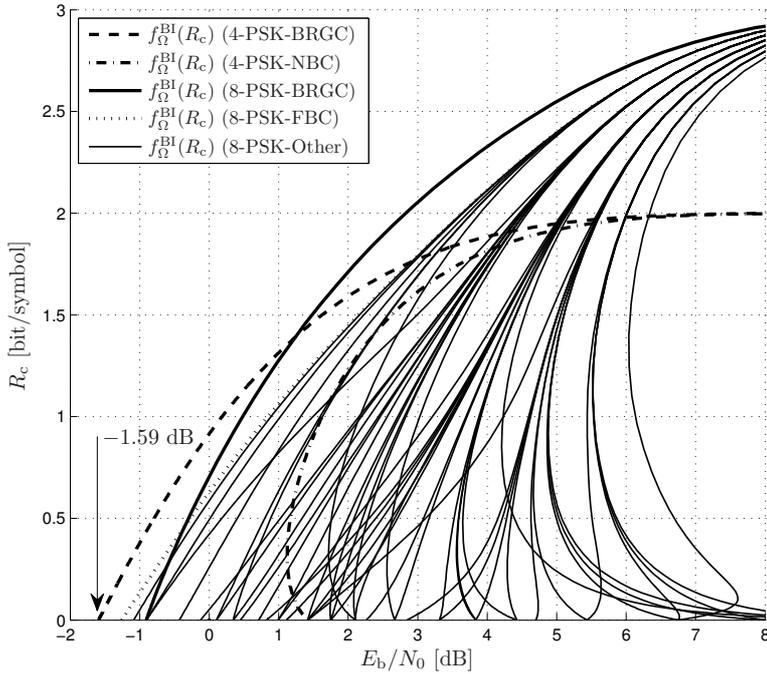


Fig. 3.3: BICM capacity for 4-PSK and 8-PSK for all binary labelings that give a different BICM capacity. The input distribution is $\mathbb{P} = \mathbb{U}_8$ and the channel is AWGN.

$\text{SNR} \rightarrow 0$ (or equivalently when $R_c \rightarrow 0$), the E_b/N_0 needed for reliable transmission grows to infinity. Moreover, this figure also shows that NBC achieves the SL, which none of the other labelings do. More details about this can be found in [23], where the analysis of BICM in the wideband regime presented in [96, 112] was generalized.

Example 7 (Binary labelings for PSK) In Fig. 3.3, we show the BICM capacity for $\mathbb{P} = \mathbb{U}_8$ and 4-PSK and 8-PSK. This figure shows the BICM capacity for all binary labelings that give a different BICM capacity. For 4-PSK, there are only two (the BRGC and NBC), and for 8-PSK there are 49 [113, Tables A.1 and A.2]. The results in Fig. 3.3 show that the optimum binary labeling for 4-PSK is the BRGC, but for 8-PSK this depends on the SNR. For low SNR the optimum binary labeling for 8-PSK is the FBC and for high SNR the optimum binary labeling is the BRGC. Moreover, this

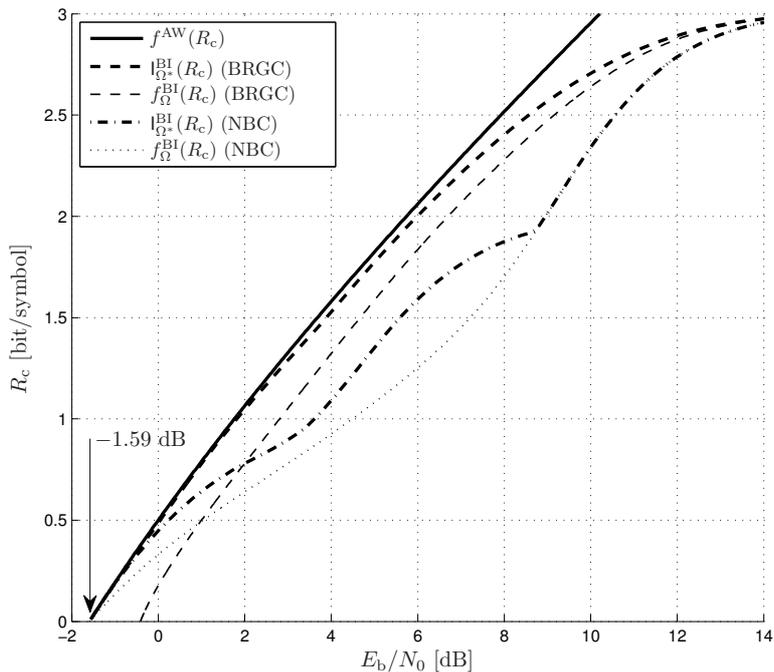


Fig. 3.4: BICM capacity for 8-PAM with the BRGC and the NBC, using \mathbb{U}_8 (Ω) and the optimized input distribution for each SNR (Ω^*). The AWGN capacity is shown for reference.

figure shows that no binary labeling for 8-PSK achieves the SL. This result is in fact a particular case of [23, Corollary 19] where it is proven that no binary labeling achieves the SL if M -PSK input alphabets with $\mathbb{P} = \mathbb{U}_M$ and $M > 4$ are used. More importantly, if we allow the system to switch the input alphabet (from $M = 4$ to $M = 8$) and the binary labeling, the BRGC is the optimum binary labeling for this particular case and any SNR. Also, we note that the capacity curves for 8-PSK and high SNR merge into 7 different groups, which are shown in [113, Table A.1].

Example 8 (BICM with nonuniform \mathbb{P} for 8-PAM) In Fig. 3.4, we show the BICM capacity for an 8-PAM signal set labeled by the BRGC and the NBC when nonuniform input distributions are used. In this figure, we use the notation $I_{\Omega^*}^{\text{BI}}(R_c)$ to show that the input distribution was optimized. We performed a grid search (with steps of 0.01) using the numer-

ical implementation of the BICM capacity shown in the Appendix over the three variables defining the input distribution: $P_{C_1}(0)$, $P_{C_2}(0)$, and $P_{C_3}(0)$. For each SNR value, we selected the input distribution that maximizes the BICM capacity.

The results in this figure show how, by properly selecting the input distribution, the BICM capacity increases and gets very close to the AWGN capacity for $R_c \leq 2$ bit/symbol. Interestingly, the results in Fig. 3.4 show that if the input distribution is optimized, the NBC is not the optimal binary labeling for low SNR anymore. The figure also shows that the BRGC with an optimized input distribution achieves the SL. Similar results have been recently presented in [114] for 4-PAM.

Based on the results in Example 5–8, some conclusions about the maximum BICM capacity can be drawn. First of all, these results show the suboptimality of the BRGC in terms of maximizing the BICM capacity for a given equally spaced input alphabet and uniform input distributions. For 8-PAM and for low and medium SNR values, the NBC and the FBC give higher BICM capacities than the BRGC. Moreover, for asymptotically low rates, the optimum binary labeling for M -PAM input alphabets is the NBC. The results in these examples also show that a BICM system can be optimum for asymptotically low rates (i.e., it can achieve the SL). This can be obtained by using probabilistic shaping or by properly selecting the binary labeling. Furthermore, some binary labelings were found to be extremely bad for BICM since they make the required E_b/N_0 for reliable transmission and asymptotically low rates grow to infinity. For more details and results about the maximum BICM capacity, we refer the reader to [23] (see also [98, 111, 114]).

Chapter 4

Modeling BICM and BICM-ID

When analyzing communication systems, it is common to study their performance by comparing them with the performance of other systems that can be used to perform the same tasks. The performance analysis uses models that predict the behavior of the system, assumptions that simplify the analysis, and it is usually based on mathematical tools. In particular, we are usually interested in a numerical quantity that represents how likely it is that the bits we are sending through the channel do not arrive at the destination correctly (the probability of error). In this chapter, we review models and tools that can be used to analyze the performance of BICM and BICM-ID systems.

4.1 The BICM Channel

In coded binary transmission, each coded bit is transmitted through the channel as one of the two possible signal alternatives and the channel is usually symmetric¹ and memoryless. However, in CM systems, since multi-level modulation is used, each signal alternative carries more than one coded bit, and thus, the bits are transmitted through a channel that is in general not symmetric and not memoryless. Specifically, in BICM, all the elements between the encoder and the decoder can be grouped in what is called the BICM channel, as shown in Fig. 2.4. The BICM channel was introduced in [115] and corresponds to the channel that separates the binary encoder and decoder in BICM transmission. Although this concept was introduced for BICM, it can also be used for BICM-ID.

¹A channel is said to be symmetric if the channel transition probability fulfills $p_{\mathbf{Y}|U=1}(\mathbf{y}) = p_{\mathbf{Y}|U=0}(-\mathbf{y})$.

Assuming infinitely long coded sequences, and an interleaver that randomly permutes the bits, the BICM channel is completely determined by the PDF of the random variables L^π passed to the decoder (see Fig. 2.4). In the following subsections, we will review how approximations for these PDFs can be obtained for the AWGN channel when an adequate model for the demapper is available. The importance of having such approximations for the PDF of the L-values is that they allow us to approximate the performance of BICM systems.

In this chapter we are interested in a probabilistic characterization of the L-values, and therefore, to emphasize their random nature, we will use capital letters to denote them. The same notation will be used for the transmitted and received symbols, which are also random variables.

4.1.1 Perfect Feedback Assumption in BICM-ID

In BICM-ID, after a certain number of iterations, and for a sufficiently high SNR, it can be assumed that the demapper has perfect knowledge about all the bits except the one it is calculating the L-value for. This situation is called perfect feedback (PF).² The PF assumption can be used to analyze the BER of BICM-ID in the error-floor region and it transforms the detection process of a multilevel signal set into the detection of binary signals. This assumption has proved to be simple yet very effective in predicting the BER performance of BICM-ID in the error-floor region.

To characterize the demapper under the PF assumption, we analyze the so-called generalized Euclidean distance spectrum (GEDS) of a labeled constellation with full a priori information. An identical concept called bit-wise distance spectra was presented in [95, 113], where all the binary labelings for 8-PSK having a different GEDS were classified. The same concept was also presented in [76, Ch. 4], where it is called bit-wise EDS. Throughout this thesis we use the name GEDS to be consistent with the name used in Paper B.

The GEDS with full a priori information is defined by the matrix $\mathbb{G}^{\text{full}} \triangleq [\mathbf{g}_1^T, \dots, \mathbf{g}_m^T]^T$ whose entries $g_{k,l}$ with $k = 1, \dots, m$ and $l = 1, \dots, D$ are twice the number of pairs of constellation points in \mathbb{X} at Euclidean distance d_l^2 (with $d_1^2 < d_2^2 < \dots < d_D^2$) whose binary labelings differ in the k th bit position only. The number of different Euclidean distances in the signal set is denoted by D . The GEDS is a generalization of the Euclidean distance spectrum of [116] in the sense that it considers the m bit positions separately.

Example 9 Consider a 4-PSK signal set labeled by the NBC and the BRGC as shown in Fig. 4.1. In this figure, we show the relevant Euclidean

²Also known as error-free feedback (EFF).

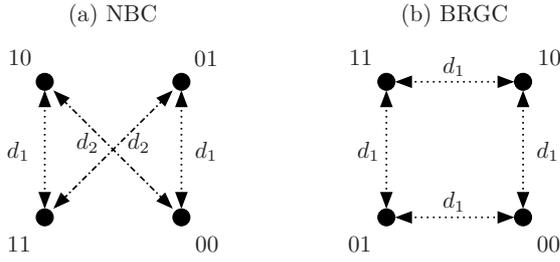


Fig. 4.1: 4-PSK signal sets labeled by the BRGC and NBC. The relevant Euclidean distances under the PF assumption are shown.

distances under the PF assumption. The GEDS of these labeled constellations are

$$\mathbb{G}_{\text{NBC}}^{\text{full}} = \begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix}, \quad \mathbb{G}_{\text{BRGC}}^{\text{full}} = \begin{bmatrix} 4 & 0 \\ 4 & 0 \end{bmatrix}. \quad (4.1)$$

Example 10 Consider an 8-PSK signal set labeled by the NBC. This labeled constellation, together with its relevant Euclidean distances, is shown in Fig. 4.2. In this case, the GEDS is given by

$$\mathbb{G}_{\text{NBC}}^{\text{full}} = \begin{bmatrix} 0 & 0 & 0 & 8 \\ 0 & 8 & 0 & 0 \\ 8 & 0 & 0 & 0 \end{bmatrix}. \quad (4.2)$$

Under the PF assumption, since only binary modulations are considered (with different distances d_l), the distribution of an L-value at the demodulator's output, conditioned on a transmitted symbol and bit position, follows a Gaussian distribution. To understand this, consider an input alphabet with only two N -dimensional symbols \mathbf{x}_1 and \mathbf{x}_2 (labeled by 0 and 1, respectively) and the AWGN channel. The extrinsic L-values³ computed by the modulator in this case are given by (2.4), i.e.,

$$L(\mathbf{Y}(\mathbf{X})) = \log \frac{p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}_2}(\mathbf{Y})}{p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}_1}(\mathbf{Y})} \quad (4.3)$$

$$= -\frac{\|\mathbf{Y}(\mathbf{X}) - \mathbf{x}_2\|^2}{N_0} + \frac{\|\mathbf{Y}(\mathbf{X}) - \mathbf{x}_1\|^2}{N_0}, \quad (4.4)$$

³Since the received signal \mathbf{Y} is a random variable that depends on the transmitted symbol \mathbf{X} , from now on, we use the notation $L(\mathbf{Y}(\mathbf{X}))$ to emphasize that the L-values are also random variables that depend on the transmitted symbol \mathbf{X} (through \mathbf{Y}).

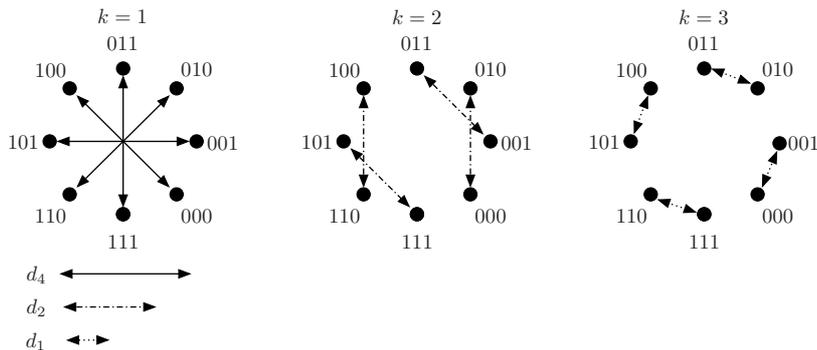


Fig. 4.2: 8-PSK signal set labeled by the NBC. The relevant Euclidean distances under the PF assumption for the different bit positions are shown.

where to pass from (4.3) to (4.4) we used the conditional transition PDF in (1.3). Alternatively, the expression in (4.4) can also be directly obtained from (2.9) by using $\mathbf{H} = \mathbf{1}$ and $m = 1$.

It is possible to show that since $\mathbf{Y}(\mathbf{X}) \sim \mathcal{N}(\mathbf{X}, N_0/2)$, the distribution of $L(\mathbf{Y}(\mathbf{X}))$ in (4.4) for a given transmitted symbol \mathbf{x}_i is

$$L(\mathbf{Y}(\mathbf{X} = \mathbf{x}_i)) \sim \mathcal{N}\left((-1)^i \frac{1}{N_0} \|\mathbf{x}_2 - \mathbf{x}_1\|^2, \frac{2}{N_0} \|\mathbf{x}_2 - \mathbf{x}_1\|^2\right), \quad (4.5)$$

with $i = 1, 2$ and where the sign of the mean depends on the transmitted symbol. Moreover, the mean of the Gaussian distribution depends on the squared Euclidean distance between the constellation points, and the variance is twice the absolute value of the mean. This last condition corresponds to the so-called consistency condition.⁴

The result in (4.5) simply states that the L-values computed by the modulator under the PF assumption are Gaussian random variables fully defined by the Euclidean distance of the two points involved in the detection and the noise variance $N_0/2$. One important consequence of this is that the L-values passed to the decoder (after de-interleaving) can be modeled as a Gaussian mixture, where the way of how this mixture is done depends on the interleaver design. More details about this will be given in Sec. 4.2.3.

Assuming equally likely symbols, the uncoded BER of a labeled constel-

⁴L-values are said to be consistent if their conditional PDF fulfill the condition $p_{L|U=1}(\lambda) = e^\lambda p_{L|U=0}(\lambda)$ [67].

lation under the PF assumption can be expressed as

$$P_b = \frac{1}{m} \sum_{k=1}^m P_b^{(k)} \quad (4.6)$$

$$= \frac{1}{mM} \sum_{k=1}^m \sum_{i=1}^M Q \left(\sqrt{\frac{[\hat{d}_i(k)]^2}{2N_0}} \right), \quad (4.7)$$

where $P_b^{(k)}$ in (4.6) is the uncoded BER for the k th bit position, $\hat{d}_i(k) \triangleq \|\mathbf{x}_i - \hat{\mathbf{x}}_{i,k}\|$ where $\hat{\mathbf{x}}_{i,k}$ is the constellation symbol with a binary labeling equal to the binary labeling of \mathbf{x}_i except in the k th bit position, and the Q-function in (4.7) come from (4.5). Moreover, the uncoded BER in (4.7) can be expressed in terms of the entries of the GEDS matrix \mathbb{G}^{full} as

$$P_b = \frac{1}{mM} \sum_{k=1}^m \sum_{l=1}^D g_{k,l} Q \left(\sqrt{\frac{d_l^2}{2N_0}} \right). \quad (4.8)$$

The expressions (4.7)–(4.8) show the dependency of the uncoded BER on the binary labeling (as well as on the input alphabet). Similar expressions to those in (4.7)–(4.8) have been presented in [95, eqs. (48)–(50)] and [113, App. A, eq. (A.17)]. Also, approximations of (4.7) have been used in the literature to find good binary labelings for BICM-ID. For example, binary labelings for multidimensional input alphabets were designed in [117], and the binary switching algorithm [118] was used in [68] to find good labelings for 8-PSK input alphabets. The search in [68] resulted in a the so-called M8 labeling shown in the following example. The same labeling was also found in [70], where it is called semi set partitioning.

Example 11 *The 8-PSK signal set labeled by the M8 labeling is shown in Fig. 4.3. In this case, the GEDS is given by*

$$\mathbb{G}_{\text{M8}}^{\text{full}} = \begin{bmatrix} 0 & 0 & 0 & 8 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \end{bmatrix}. \quad (4.9)$$

The key difference between the labelings shown in Examples 10 and 11 is that the M8 labeling gives a smaller P_b in (4.8). This can be observed directly from $\mathbb{G}_{\text{NBC}}^{\text{full}}$ in (4.2) and $\mathbb{G}_{\text{M8}}^{\text{full}}$ in (4.9). Both matrices are identical for $k = 1, 2$ and they differ for $k = 3$: the NBC has $\mathbf{g}_3 = [8, 0, 0, 0]$ (4 pairs of constellation points at distance d_1) and the M8 has $\mathbf{g}_2 = [0, 0, 8, 0]$ (4 pairs of constellation points at distance d_3).

For the purposes of this thesis, the most important property of the binary labelings proposed for BICM-ID (such as the NBC in Examples 9 and

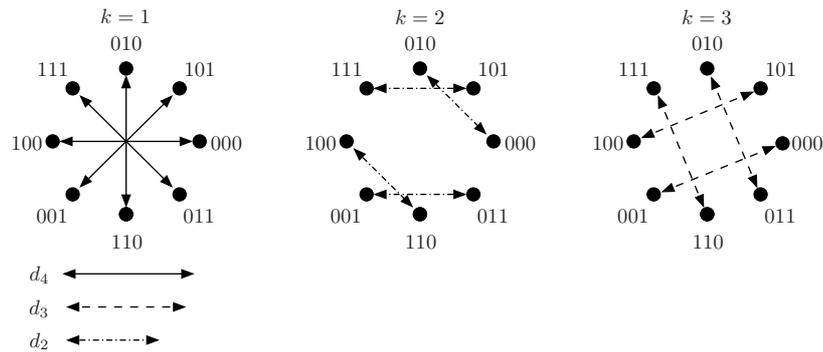


Fig. 4.3: 8-PSK signal set labeled by the M8 labeling [68, 70]. The relevant Euclidean distances under the PF assumption for the different bit positions are shown.

10, and the M8 in Example 11) is that they offer unequal error protection (UEP) in the different bit positions. This means that the minimum Euclidean distances of different rows of the matrix \mathbb{G}^{full} are in general different, cf. (4.2) and (4.9). The consequence of this is that coded bits that are transmitted in a given bit position will receive a higher (or lower) protection than other bits. This is a property of the binary labeling, and it can be exploited in the design of the BICM-ID system. More details about this will be given in Sec. 4.2 and in Paper B.

4.1.2 Probability Density Function of L-values in BICM

In BICM, the statistical properties of the L-values computed by the demapper in (2.9) (see Fig. 2.4) completely determine (after deinterleaving) the behavior of the BICM channel. The knowledge of the PDF of these L-values can then be used to predict the performance of the encoder/decoder pair, as explained in Sec. 4.3. The knowledge of the PDF of the L-values can also be used to compute the BICM capacity of the BICM channel. Closed-form expressions for the PDF of the L-values in (2.9) are in general not known; however, good analytical approximations (based on (2.11)) can be developed. In this subsection, we review these approximations.

Example 12 Consider a BPSK signal set over the AWGN channel where the symbols $x_1 = -\sqrt{E_s}$ and $x_2 = +\sqrt{E_s}$ are labeled by 0 and 1, respectively. In this case, the L-values in (2.9) (or directly from (4.4)) are given

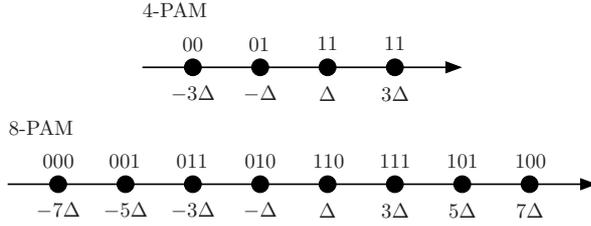


Fig. 4.4: 4-PAM and 8-PAM input alphabets labeled by the BRGC.

by

$$l(y) = -\frac{(y - \sqrt{E_s})^2}{N_0} + \frac{(y + \sqrt{E_s})^2}{N_0} = \frac{4\sqrt{E_s}}{N_0}y. \quad (4.10)$$

Since for a transmitted symbol x_i , $Y(X = x_i) \sim \mathcal{N}((-1)^i \sqrt{E_s}, N_0/2)$, the PDF of $L(Y(X))$ in (4.10) conditioned on the transmitted symbol x_i is

$$L(Y(X = x_i)) \sim \mathcal{N}((-1)^i 4\text{SNR}, 8\text{SNR}), \quad (4.11)$$

where $\text{SNR} = E_s/N_0$ and where the sign of the mean value depends on the transmitted symbol.

The expression in (4.11) shows the exact expression for the conditional PDF of the L-values. This comes from the linear relation between $l(y)$ and y , cf. (4.10). Moreover, this result can be extended if the input alphabet is 4-PSK labeled by the BRGC (cf. Fig. 4.1 (b)). This is simply because this particular labeled constellation can be separated into two independent BPSK signal sets in each dimension. This in general not true for other labeled constellations, or even for 4-PSK with the NBC.

Unfortunately, for any labeled constellation with $M > 2$ (except 4-PSK with the BRGC), the relation between the L-values and the received signal is not linear. The consequence of this is that in general closed-form expressions for the PDF of the conditional L-values are not available. To clarify this, consider the following example, which analyzes the 4-PAM and 8-PAM signal sets labeled by the BRGC shown in Fig. 4.4. From now on, the input alphabet are assumed to be normalized to have unit energy, i.e., $E_s = 1$. This gives $\Delta^2 = 1/5$ and $\Delta^2 = 1/21$ for 4-PAM and 8PAM, respectively.

Example 13 Consider a 4-PAM signal set labeled by the BRGC. The relation between the L-values and the received signal in (2.9) is shown in Fig. 4.5. This figure clearly shows the nonlinear relation between the L-values and the received signal, which is particularly clear for $k = 2$. Analogous results are shown in Fig. 4.6 for an 8-PAM signal set labeled by the BRGC.

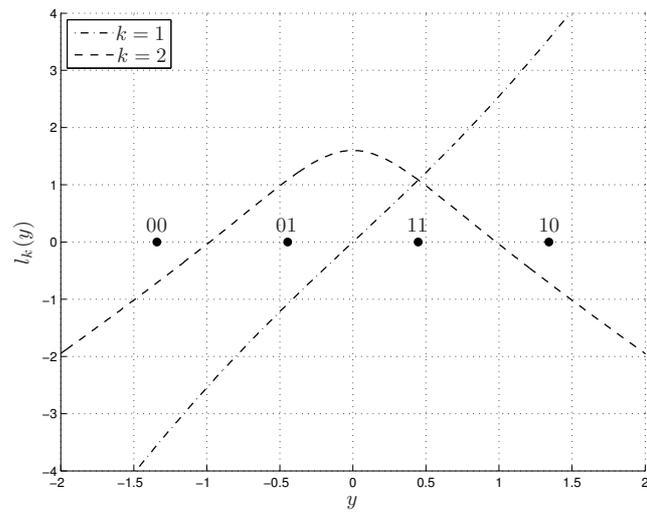


Fig. 4.5: Relation between the L-values and the received signal given by (2.9) for a 4-PAM signal set labeled by the BRGC and for SNR = 0 dB. The constellation points are also shown.

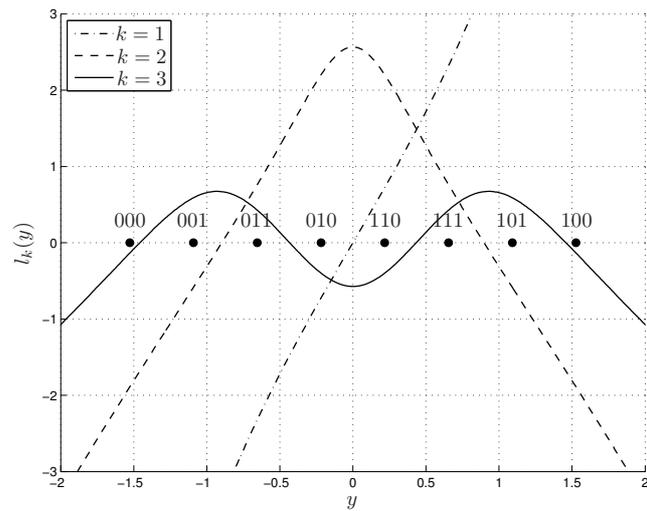


Fig. 4.6: Relation between the L-values and the received signal given by (2.9) for an 8-PAM signal set labeled by the BRGC and for SNR = 3 dB. The constellation points are also shown.

The previous example shows that the relation in (2.9) is nonlinear. However, if the max-log metrics in (2.11) are used instead, the relation becomes piece-wise linear. For example, for a one-dimensional input alphabet (e.g., an M -PAM labeled constellation), the L-values computed by the demapper are given by

$$\tilde{l}_k(y) = \frac{1}{N_0} \left[\min_{x \in \mathcal{X}_{k,0}} (y-x)^2 - \min_{x \in \mathcal{X}_{k,1}} (y-x)^2 \right]. \quad (4.12)$$

For the 4-PAM and 8-PAM labeled constellations in Example 13, this piece-wise linear relationship is shown in Fig. 4.7 and in Fig. 4.8, respectively. Similar results have been shown in [81, Fig. 3] and [82, eqs. (11)–(14), Tab. 1].

The use of the max-log approximation, together with the piece-wise linearization that it introduces, has been used to develop expressions for the PDF of the L-values in (2.11). An algorithmic approach for arbitrary labeled constellations was introduced in [107] and closed-form expressions for QAM input alphabets labeled with the BRGC for the AWGN channel were presented in [108, 119] and for PSK input alphabets in [120]. Results for fading channels have been reported in [109], and recently, closed-form approximations for arbitrary labeled constellations for fading channels have been developed in [110].

The expression in (4.12) can be expressed as a sum of piecewise linear functions

$$\tilde{l}_k(y) = \sum_{j=1}^M I_{[Y \in \mathcal{Y}_j]}(a_{k,j}y + b_{k,j}), \quad (4.13)$$

where $I_{[K]}$ is an indicator function, i.e., $I_{[K]} = 1$ if K is true, and $I_{[K]} = 0$ if K is false, \mathcal{Y}_j with $j = 1, \dots, M$ are the Voronoi region of the M -PAM signal set, and $a_{k,j}$ and $b_{k,j}$ are the slope and the intercept of the linear pieces, respectively. In Fig. 4.8, we show the regions \mathcal{Y}_j for an 8-PAM input alphabet.

For a given transmitted symbol $X = x_i$, $Y(X = x_i) \sim \mathcal{N}(x_i, N_0/2)$, and thus, the expression (4.13) shows that the PDF of the conditional L-values is a sum of piece-wise Gaussian functions. The closed-form expressions for the PDFs of the max-log L-values for QAM input alphabets and the BRGC presented in [107, 108] were found based on this observation. To clarify this idea, in Fig. 4.9, we show the PDF of the conditional max-log L-values for $M = 4$, $k = 1$, and conditioned to $X = x_4$. This figure shows the piece-wise Gaussian characteristic of the conditional PDF of the max-log L-values. In Fig. 4.10, we show similar results (in logarithmic scale), where the three Gaussian pieces defining $p_{L_1|X=x_4}(\lambda)$ are clearly visible.

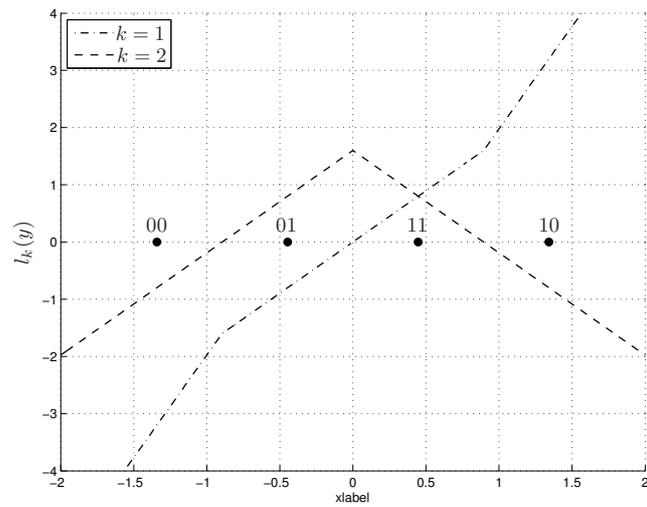


Fig. 4.7: Relation between the max-log L-values and the received signal given by (4.12) for a 4-PAM signal set labeled by the BRGC and for SNR = 0 dB. The constellation points are also shown.

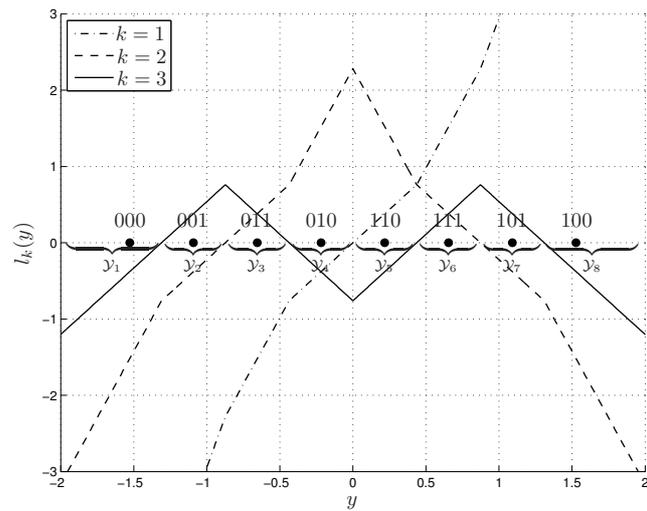


Fig. 4.8: Relation between the max-log L-values and the received signal given by (4.12) for an 8-PAM signal set labeled by the BRGC and for SNR = 3 dB. The constellation points and the partitioning of the received signal space for $k = 2$ are also shown.

In [108], a further approximation for (4.13) was introduced. In this approximation, for a given transmitted symbol $X = x_i$, only one interval in y is considered, i.e.,

$$\tilde{L}_k(Y(X = x_i)) \approx \hat{a}_k(x_i)Y(x_i) + \hat{b}_k(x_i), \quad (4.14)$$

where the coefficients $\hat{a}_k(x_i)$ and $\hat{b}_k(x_i)$ depend on the transmitted symbol x_i , the bit position k , and the noise variance.

The expression (4.14) simply states that the conditional max-log L-values can be approximated by an affine function, and therefore, the PDF of the conditional max-log L-values follows a Gaussian distribution. In [108], two different ways of selecting the parameters $\hat{a}_k(x_i)$ and $\hat{b}_k(x_i)$ defining this affine function were proposed:

- The *zero-crossing model* (ZCMod), initially proposed in [121], states that the affine function will be determined by the values of $a_{k,j}$ and $b_{k,j}$ in the closest partition to the transmitted symbol where $l_k(y) = 0$.⁵
- The *consistent model* (CoMod), initially proposed in [122, 123], which considers the Voronoi region of the transmitted symbol, i.e., $\hat{a}_k(x_i) = a_{k,i}$ and $\hat{b}_k(x_i) = b_{k,i}$.

To clarify these models, in Fig. 4.10, we show the ZCMod and CoMod approximations, and we compare them with the exact PDF of the conditional max-log L-values $p_{L_1|X=x_4}(\lambda)$.

It can be shown that for M -PAM signal sets labeled by the BRGC (see more details in Paper C), the distribution of the conditional max-log L-values in (4.14) using the CoMod approximation is given by

$$\tilde{L}_k(Y(X = x_i)) \sim \mathcal{N}\left((-1)^{1-c_{i,k}} \frac{1}{N_0} [\bar{d}_i(k)]^2, \frac{2}{N_0} [\bar{d}_i(k)]^2\right), \quad (4.15)$$

where $\bar{d}_i(k) \triangleq \min_{x \in \mathcal{X}_{k,1-c_{i,k}}} |x_i - x|$, i.e., $\bar{d}_i(k)$ corresponds to the Euclidean distance between the transmitted symbol x_i and the closest symbol in the labeled constellation with the opposite value at bit position k .

Motivated by the results in (4.15), and in analogy to the GEDS with full a priori information \mathbb{G}^{full} , we define here the GEDS with zero a priori information. The GEDS with zero a priori information is defined by the matrix $\mathbb{G}^{\text{zero}} \triangleq [\mathbf{g}_1^T, \dots, \mathbf{g}_m^T]^T$ whose entries $g_{k,l}$ with $k = 1, \dots, m$ and $l = 1, \dots, D$ are the number of elements in the set $\{\bar{d}_1(k), \dots, \bar{d}_D(k)\}$ equal to the distance d_l , and where $D = M/2$ for M -PAM and M -PSK labeled constellations.

⁵The two Voronoi regions around the zero-crossing $l_k(y) = 0$ have the same parameters $a_{k,j}$ and $b_{k,j}$ in (4.13), see for example Fig. 4.8. This is a property of the BRGC and leads to parameters $\hat{a}_k(x_i)$ and $\hat{b}_k(x_i)$ in (4.14) which are not affected by the selection of the interval on the left or on right of the zero-crossing.

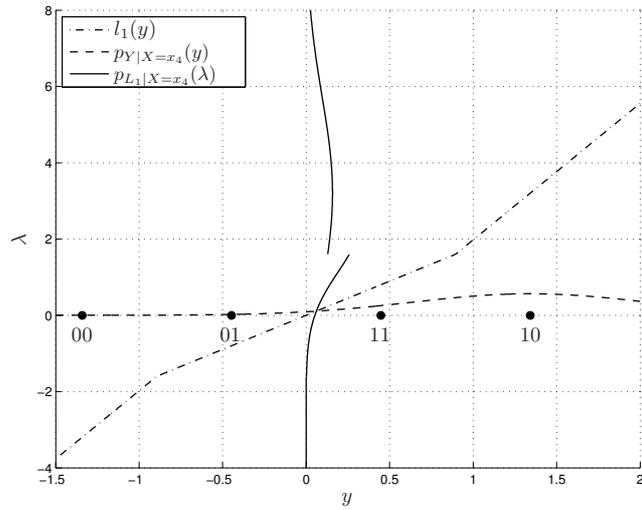


Fig. 4.9: Relation between the max-log L-values and the received signal given by (4.12) and the conditional PDFs ($X = x_4$) of the received signal Y and the max-log L-values. 4-PAM signal set labeled by the BRGC, $k = 1$, and SNR = 0 dB.

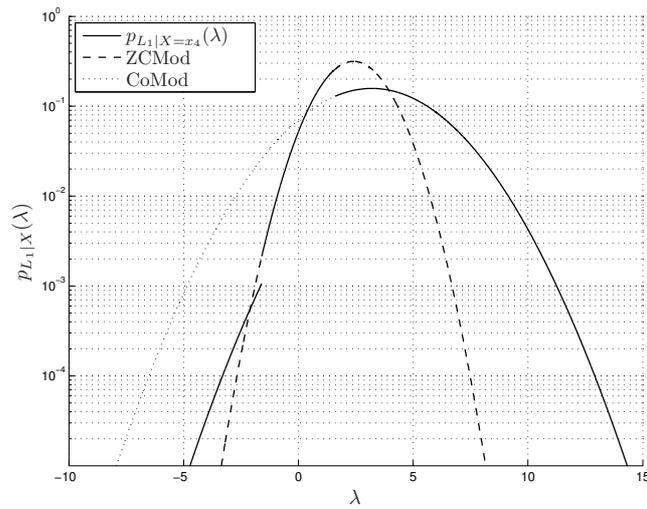


Fig. 4.10: Conditional PDF ($X = x_4$) of the max-log L-values for a 4-PAM signal set labeled by the BRGC, $k = 1$, and SNR = 0 dB. The CoMod and ZCMod approximations are also shown.

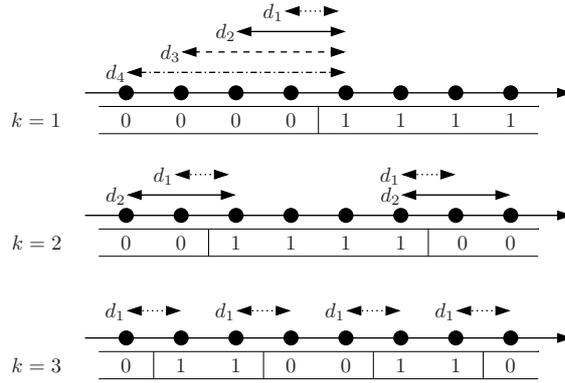


Fig. 4.11: 8-PAM signal set labeled by the BRGC. The relevant Euclidean distances for the GEDS with zero a priori information are shown.

Example 14 ($\mathbb{G}_{\text{BRGC}}^{\text{zero}}$ for 8-PAM and the BRGC) *The GEDS with zero a priori information for an 8-PAM signal set labeled by the BRGC shown in Fig. 4.11 is given by*

$$\mathbb{G}_{\text{BRGC}}^{\text{zero}} = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 4 & 4 & 0 & 0 \\ 8 & 0 & 0 & 0 \end{bmatrix}. \quad (4.16)$$

In analogy to the analysis of \mathbb{G}^{full} in Sec. 4.1.1, we can analyze $\mathbb{G}_{\text{BRGC}}^{\text{zero}}$ in (4.16). A quick examination of (4.16) together with the PDF of the L-values in (4.15) reveals that the BRGC introduces UEP, i.e., it offers different (average) protection levels to the different bit positions. For example, for the labeled constellation in Example 14, we see that for $k = 3$ all the bits have a closest neighbor at distance d_1 , but for $k = 1$, this happens only 25% of the time, i.e., it depends on the transmitted symbol.

We conclude this subsection by noting that the PDF of the L-values passed to the decoder will be a Gaussian mixture, where the mixture will depend on how the interleaver is designed. More details about this will be given in Sec. 4.2.3.

4.2 Unequal Error Protection in BICM and BICM-ID

The binary labelings that are commonly used in BICM and BICM-ID introduce UEP. This was suggested in the previous section by discussing the

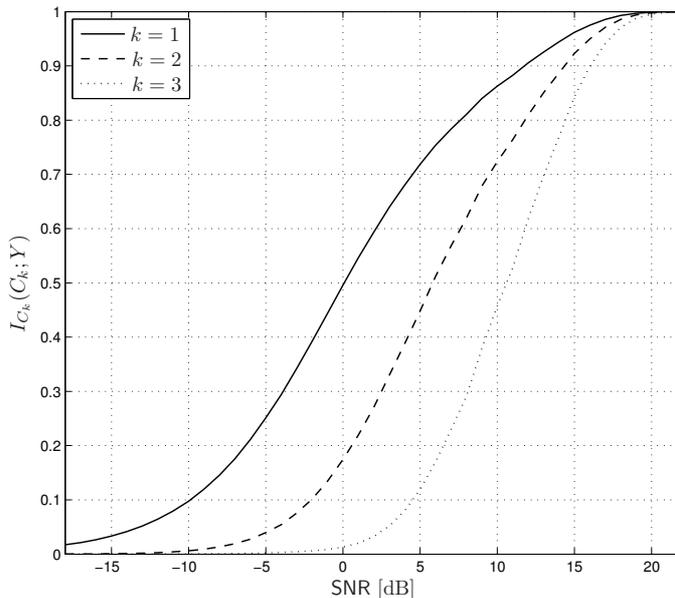


Fig. 4.12: Bit level AMI $I_{C_k}(C_k; Y)$ for an 8-PAM signal set labeled by the BRGC over the AWGN channel.

GEDS with zero and full a priori information. In this section, we discuss the concept of UEP in more detail. We will show that there are other ways to support the idea of UEP in BICM, which may be simpler to understand. We will also show how this can be exploited by using a properly designed BICM system.

4.2.1 Unequal Error Protection in BICM

The UEP introduced by the binary labeling can be explained based on the analysis of the bit level AMIs $I_{C_k}(C_k; Y)$ defining the BICM capacity in (3.18). In Fig. 4.12, we show these AMIs for an 8-PAM signal set labeled by the BRGC over the AWGN channel. This figure shows that $I_{C_1}(C_1; Y) \geq I_{C_2}(C_2; Y) \geq I_{C_3}(C_3; Y)$, which should be understood as the first bit position giving a higher protection than the second one, which in turn gives a higher protection than the third bit position.

Another way of supporting the claim that the binary labeling introduces UEP is by analyzing the uncoded BER. In Fig. 4.13, we show the uncoded BER per bit position $P_b^{(k)}$, for an 8-PAM signal set labeled by the BRGC

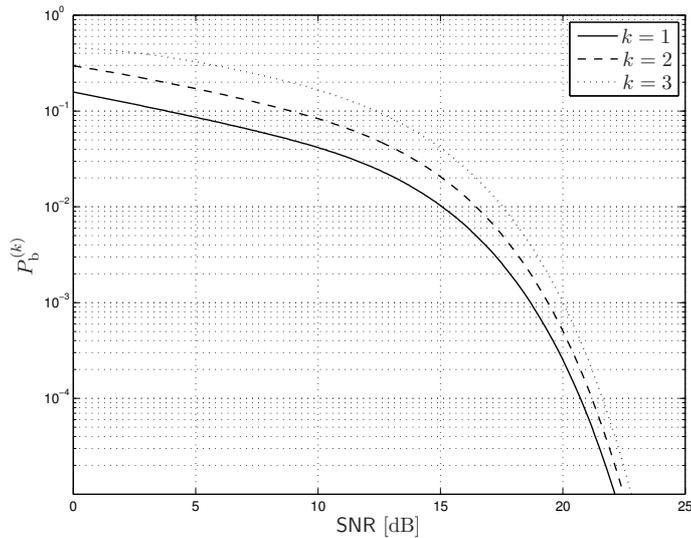


Fig. 4.13: Uncoded BER per bit position (BICM) for an 8-PAM signal set labeled by the BRGC over the AWGN channel.

over the AWGN channel. This figure shows how the bit position $k = 1$ offers a higher protection than $k = 2$, which in turn offers a higher protection than $k = 3$. Similar results have been presented for M -PSK labeled constellations in [16], and in [124, Figs. 5.2 and 5.4] for 16-QAM and 64-QAM.

4.2.2 Unequal Error Protection in BICM-ID

In BICM-ID, under the PF assumption, the binary labeling also introduces UEP. For example, if we analyze the 4-PSK signal set labeled by the NBC in Example 9, we see that for $k = 1$, since we assume perfect knowledge about $k = 2$, we obtain two BPSK signal sets, each of them separated by a distance d_2 (see Fig. 4.1 (a)). If we analyze $k = 2$, we again obtain two BPSK signal sets, but in this case, the Euclidean distance of those input alphabets is d_1 . Since $d_2 > d_1$, bits transmitted in the first bit position receive higher protection than the bits transmitted in the second bit position. We note that the BRGC does not introduce UEP; however, the BRGC would never be used in BICM-ID because it does not offer considerable gains over the iterations, i.e., the BER performance of a BICM-ID system using a signal set labeled by the BRGC does not improve over the iterations [76, Ch. 4].

A similar analysis can be done for the labelings for 8-PSK in Examples 10

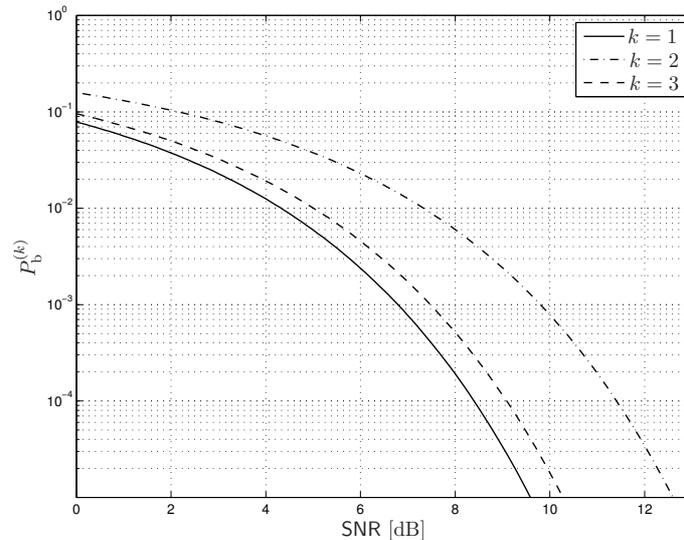


Fig. 4.14: Uncoded BER per bit position (BICM-ID) for an 8-PSK signal set labeled by the M8 labeling (Fig. 4.3) over the AWGN channel under the PF assumption, cf. (4.6)–(4.8).

and 11. The main conclusion obtained in this case is the same as for 4-PSK labeled by the NBC: the binary labeling introduces UEP. We note that this effect is in fact also caused by other binary labelings such as the maximum squared Euclidean weight [75], the modified set partitioning [125], or other labelings optimized for BICM-ID, cf. [68], [76, Ch. 4].

In analogy to the uncoded BER analysis for BICM with 8-PAM and the BRGC made in Sec. 4.2.1, we can analyze the uncoded BER of a labeled constellation under the PF assumption, cf. (4.6)–(4.8). In Fig. 4.14, we present the uncoded BER for the labeled constellation in Fig. 4.3 (cf. Example 11), i.e., an 8-PSK signal set labeled by the M8 labeling. This figure shows the UEP introduced by the binary labeling, i.e., the first bit position offers a high protection level compared to the third bit position, which in turn offers a higher protection when compared to the second bit position. Similar results have been presented for 8-PSK and the NBC in [95, Fig. 7].

We conclude this section by mentioning that the UEP introduced by the labeling can be used to design the interleaver and the code for BICM and BICM-ID. The interleaver and code design for BICM and M -QAM signal sets labeled by the BRGC is analyzed in Paper A. The interleaver and code design for BICM-ID for 4-PSK labeled by the NBC is analyzed in Paper B.

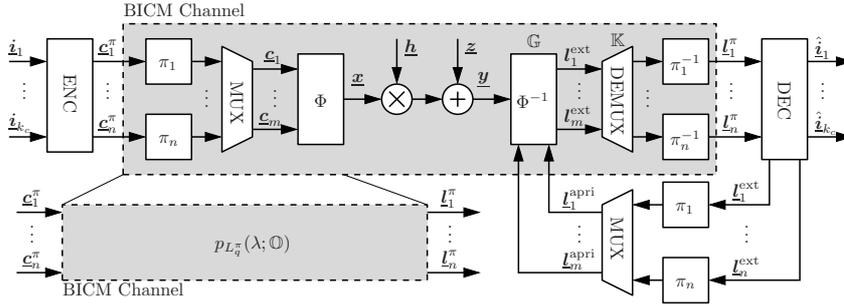


Fig. 4.15: General BICM(-ID) model. The BICM channel is also shown.

The design of unequally spaced input alphabets to take advantage of the UEP in BICM is analyzed in Paper C. A generalization of the CoMod for the PDF of the conditional max-log L-values presented Sec. 4.1.2 in BICM is used in Paper D.

4.2.3 Taking Advantage of the UEP

In the previous subsections, we showed that the UEP is caused by the binary labeling in both BICM and BICM-ID. However, if the standard BICM configuration in Fig. 2.4 is used, and due to the bit interleaving, the UEP is eliminated. In this subsection, we briefly introduce a general BICM model that can be used to exploit the UEP caused by the binary labeling. This model is shown in Fig. 4.15 and will also be used in Sec. 4.3.

As mentioned above, when a BICM system in Fig. 2.4 is used, the UEP caused by the binary labeling is eliminated. This can be explained by the fact that the interleaver spreads the bits over time and over the bit positions, so in average, the coded bits experience the same BICM channel. However, if instead of using one single interleaver, multiple interleavers (M-interleavers) are used, the UEP introduced by the binary labeling can be conserved. These M-interleavers are shown in Fig. 4.15 and will be discussed in more detail in Sec. 5.1.1.

The system model in Fig. 4.15 also includes a multiplexing (MUX) unit that simply distributes bits from the interleaver's outputs to the mapper's input. This multiplexing unit is defined using a matrix \mathbb{K} of dimensions $n \times m$, whose elements, $0 \leq \kappa_{q,k} \leq 1$, denote the fraction of bits \underline{c}_q^π that will be assigned to the k th mapper's input.⁶ Moreover, for a given \mathbb{G} (with full or zero a priori information), and because of the M-interleavers, the

⁶Some additional constraints on $\kappa_{q,l}$ must be added; they are detailed in Paper A.

probability of a bit in \mathbf{c}_q^π being transmitted using the l th distribution of the L-values (defined by the l th column of \mathbb{G}) is denoted by $\omega_{p,l}$, where $\omega_{p,l}$ is the (p,l) th entry of the matrix

$$\mathbb{O} = \frac{1}{M} \mathbb{K} \mathbb{G} \quad (4.17)$$

of dimensions $n \times D$, and where the coefficient $1/M$ is added to give \mathbb{G} a meaning of probability. More details about this multiplexing are given in Paper A, B, and C.

If the CoMod is used, the PDF of the conditional max-log L-values at the n th input of the decoder (cf. Fig. 4.15) is given by the following Gaussian mixture

$$p_{L_q^\pi}(\lambda; \mathbb{O}) = \sum_{l=1}^D \omega_{p,l} \Psi(\lambda; \pm\mu_l, \sigma^2), \quad (4.18)$$

where $\mu_l = d_l^2$ and $\sigma^2 = 2\mu_l$.

UEP was considered an “undesired feature” in [49, Sec. II], and thus, removed from the BICM configurations. When UEP is preserved by using M-interleavers, the BICM channel in Fig. 2.4 changes. As shown in Fig. 4.15, the BICM channel has n inputs and n outputs, and it is characterized by the PDFs $p_{L_q^\pi}(\lambda; \mathbb{O})$ given by (4.18). The performance gains offered by such a system are studied in Paper A, B, and C, and selection of the code for a BICM system like the one in Fig. 4.15 is discussed in Sec. 5.1.2.

4.3 Performance Evaluation

In this section we briefly describe the different methodologies that can be used to approximate the coded performance of BICM systems. We pay special attention to the techniques based on the knowledge of the PDF of the L-values.

4.3.1 Union Bound and Pairwise Error Probability

A maximum likelihood sequence decoder (e.g., the Viterbi decoder) chooses the most likely coded sequence assuming that the sequence of L-values representing the reliabilities of a given codeword are independent.⁷ The decoder assigns a metric $\Lambda(\mathbb{C}^\pi)$ to each codeword $\mathbb{C}^\pi = [(\mathbf{c}_1^\pi)^\top, \dots, (\mathbf{c}_n^\pi)^\top]^\top$ based on the L-values available for those bits $\mathbb{L}^\pi = [(\mathbf{l}_1^\pi)^\top, \dots, (\mathbf{l}_n^\pi)^\top]^\top$, where

⁷Throughout this section, all the L-values are conditioned on a given transmitted codeword, however, for notation simplicity, we do not make this conditionals explicit.

$\mathbf{c}_q^\pi = [c_q^\pi[1], \dots, c_q^\pi[N_c]]$ and $\mathbf{l}_q = [l_q^\pi[1], \dots, l_q^\pi[N_c]]$ with $q = 1, \dots, n$ and $N_c n = N_s m$. This metric is given by

$$\Lambda(\mathbb{C}^\pi) = \log \left(\prod_{q=1}^n \prod_{i=1}^{N_c} \Pr\{C_q^\pi[i] = c_q^\pi[i] | L_q^\pi[i] = l_q^\pi[i]\} \right) \quad (4.19)$$

$$= \sum_{q=1}^n \sum_{i=1}^{N_c} \log(\Pr\{C_q^\pi[i] = c_q^\pi[i] | L_q^\pi[i] = l_q^\pi[i]\}) \quad (4.20)$$

$$= \sum_{q=1}^n \sum_{i=1}^{N_c} c_q^\pi[i] l_q^\pi[i] - \sum_{q=1}^n \sum_{i=1}^{N_c} \log(1 + \exp(l_q^\pi[i])), \quad (4.21)$$

where (4.20) follows from the independence assumption of the L-values and to pass from (4.20) to (4.21) we used (2.6).

From (4.21), we can see that the decoder will select a codeword \mathbb{C}^π instead of the transmitted all-zero codeword if $\Lambda(\mathbb{C}) > \Lambda(\mathbf{0})$, or equivalently, if

$$\sum_{q=1}^n \sum_{i=1}^{N_c} c_q[i] l_q^\pi[i] > 0. \quad (4.22)$$

Let $\mathbb{E} = [\mathbf{e}_1^T, \dots, \mathbf{e}_n^T]^T$ be a $n \times I$ matrix, where $\mathbf{e}_q = [e_q[1], \dots, e_q[I]]$ with $q = 1, \dots, n$ denote a codeword that diverges from the zero state at an arbitrary time instant and remerges with it after I trellis stages, and let \mathcal{E}_I denote the set of all such divergent codewords with $I = 2, 3, \dots$. The UB on the coded BER can be defined as

$$\text{BER} \leq \text{UB} \triangleq \frac{1}{k_c} \sum_{i=2}^{\infty} \sum_{\mathbb{E} \in \mathcal{E}_i} d_H(\mathbf{i}_{\mathbb{E}}) \text{PEP}(\mathbb{E}), \quad (4.23)$$

where $d_H(\mathbf{i}_{\mathbb{E}})$ is the Hamming weight of the input sequence $\mathbf{i}_{\mathbb{E}}$ corresponding to the codeword \mathbb{E} and $\text{PEP}(\mathbb{E})$ is the probability that the decoder selects the codeword \mathbb{E} instead of the transmitted all-zero codeword.

The decoder in Fig. 4.15 will select the codeword \mathbb{E} based on nI consecutive L-values (outputs of the BICM channel). We denote the L-value associated to each bit $e_q[i]$ in the codeword \mathbb{E} by $l_q^\pi[i]$, with $q = 1, \dots, n$ and $i = 1, \dots, I$. The PEP in (4.23) associated to a particular divergent codeword \mathbb{E} can be expressed as

$$\text{PEP}(\mathbb{E}) = \Pr\{D(\mathbb{E}) > 0\} = \int_0^\infty p_{D(\mathbb{E})}(\lambda) d\lambda, \quad (4.24)$$

where $p_{D(\mathbb{E})}(\lambda)$ is the PDF of the decision variable associated to \mathbb{E} , which

is defined using (4.22) as

$$D(\mathbb{E}) = \sum_{q=1}^n \sum_{i=1}^I e_q[i] l_q^\pi[i]. \quad (4.25)$$

In what follows, we will show that (4.23)–(4.25) are general expressions that simplifies to simpler and well-known expressions for the UB.

4.3.2 The standard UB

If the conditional L-values passed to the decoder are i.i.d. with PDF $p_{L^\pi}(\lambda)$, the PDF of the decision variable in (4.25) is given by

$$p_{D(\mathbb{E})}(\lambda) = \underbrace{p_{L^\pi}(\lambda) * \dots * p_{L^\pi}(\lambda)}_{w \text{ times}}, \quad (4.26)$$

where w is the total Hamming weight of \mathbb{E} . For this particular case, since the PEP only depends on w , the UB in (4.23) becomes

$$\text{UB} = \frac{1}{k_c} \sum_{w=w^{\text{free}}}^{\infty} \alpha(w) \text{PEP}(w), \quad (4.27)$$

where w^{free} is the free distance of the code and $\alpha(w)$ is the weight distribution spectrum (WDS) of the code which counts the total input weight of error events with output Hamming weight w .

The UB in (4.27) can be found in almost any existing book on digital communications and/or coding, see for example [56, eq. (7.9)], [9, eq. (8.2–19)], [45, eq. (12.28)], and it was originally defined for channels in which the metrics for the coded bits passed to the decoder are i.i.d. (e.g., BPSK modulation over the AWGN channel, in which the conditional L-values follow a Gaussian distribution given by (4.11)). In BICM, this i.i.d. characteristic of the metrics passed to the decoder can be obtained if one infinitely long random interleaver is used. This case is illustrated in Fig. 4.16 (a) and it corresponds to the BICM configuration analyzed and popularized by Caire *et al.* in [49].

The WDS of the code $\alpha(w)$ in (4.27) is the total Hamming weight of all input sequences that generate error events with Hamming weight w , where an error event is defined as a path in the trellis representation of the code that leaves the zero state and remerges with it after an arbitrary number of trellis stages. Although this concept was first used for convolutional codes, it is also possible to calculate the weight distribution spectrum of a turbo code using the uniform interleaver concept introduced in [126, 127]. In the following example, we show how to calculate this WDS for a convolutional code analytically.

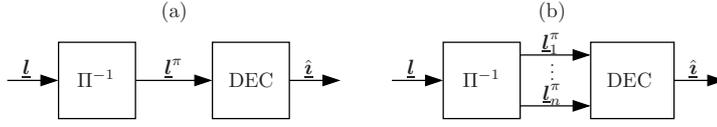


Fig. 4.16: Two possible configurations for the deinterleaver and decoder: The L-values at the input of the decoder are modeled using (a) one PDF $p_{L^\pi}(\lambda)$ or (b) n different PDFs $\{p_{L_1^\pi}(\lambda), \dots, p_{L_n^\pi}(\lambda)\}$.

Example 15 (WDS of the $(5, 7)_8$ code) Consider a convolutional code with constraint length $K = 3$ and polynomial generators $(5, 7)_8$.⁸ The encoder for this code is shown in Fig. 4.17 (a). Fig. 4.17 (b) shows the state machine associated to the encoder, where the four states represent the states of the memories $[M_1, M_2]$. We label each state transition with dummy variables I and W , where the exponents of I and W represent the Hamming weight of the input and output of the encoder, respectively. The presence of an I means that the input bit was one (marked as a solid line) and the absence of an I means that the input bit was a zero (marked as a dashed line). Also, each state is labeled by a dummy variable $\{\xi_a, \xi_b, \xi_c, \xi_d, \xi_e\}$, which can be used to write the state equations:

$$\xi_b = IW^2\xi_a + I\xi_c \quad (4.28)$$

$$\xi_c = W\xi_d + W\xi_b \quad (4.29)$$

$$\xi_d = IW\xi_b + IW\xi_d \quad (4.30)$$

$$\xi_e = W^2\xi_c. \quad (4.31)$$

The equations (4.28)–(4.31) can be used to find a transfer function of the code $T(I, W)$ as

$$T(I, W) = \frac{\xi_e}{\xi_a} = \frac{IW^5}{1 - 2IW} = W^5I + 2W^6I^2 + 4W^7I^3 + 8W^8I^4 + \dots$$

The WDS of the code can be finally obtained as

$$\alpha(w) = \frac{1}{w!} \frac{\partial^w}{\partial W^w} \frac{\partial}{\partial I} T(I, W) \Big|_{W=0, I=1}, \quad (4.32)$$

for $w = w^{\text{free}}, w^{\text{free}} + 1, \dots$, which for the $(5, 7)_8$ code gives

$$\alpha(w) = \begin{cases} (w - 4)2^{w-5}, & \text{if } w \geq 5 \\ 0, & \text{otherwise} \end{cases}. \quad (4.33)$$

⁸This code and its WDS is a standard textbook example, see for example [56, Sec 7.4.3], [9, Sec. 8.1–2], and [10, Sec. 4.3].

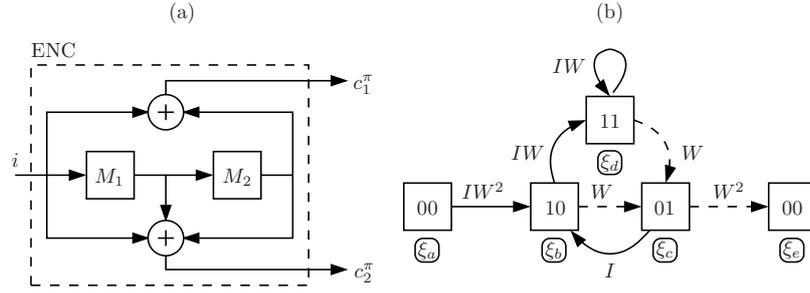


Fig. 4.17: Convolutional code $(5,7)_g$: encoder (a) and state machine (b).

4.3.3 Generalized UB

If instead of one single interleaver, M -interleavers are used (see Fig. 4.15), at the receiver's side, the conditional L -values at each input of the decoder are i.i.d., they are independent of each other, but the distributions are not necessarily the same across the decoder's inputs. This case is illustrated in Fig. 4.16 (b) and more details are given in Sec. 5.1.1. In this case, L -values at the q th decoder's input can be modeled using a different PDF $p_{L_q^\pi}(\lambda)$ with $q = 1, \dots, n$, and therefore, the PDF of the decision variable in (4.25) can be expressed as

$$p_{D(\mathbb{E})}(\lambda) = \underbrace{p_{L_1^\pi}(\lambda) * \dots * p_{L_1^\pi}(\lambda)}_{w_1 \text{ times}} * \dots * \underbrace{p_{L_n^\pi}(\lambda) * \dots * p_{L_n^\pi}(\lambda)}_{w_n \text{ times}}, \quad (4.34)$$

where w_q is the Hamming weight of the q th row of \mathbb{E} .

Since the PEP in (4.34) depends on $\mathbf{w} = [w_1, \dots, w_n]$, the UB in (4.23) becomes

$$\text{UB} = \frac{1}{k_c} \sum_{l=w^{\text{free}}}^{\infty} \sum_{\mathbf{w} \in \mathcal{W}_n(l)} \beta(\mathbf{w}) \text{PEP}(\mathbf{w}), \quad (4.35)$$

where $\beta(\mathbf{w})$ is a generalized WDS (GWDS) of the code which considers not only consider the total Hamming weight of \mathbb{E} , but the Hamming weight of each row of \mathbb{E} . In (4.35), $\mathcal{W}_n(l)$ is the set of all combinations of n nonnegative integers such that their sum is l : $\mathcal{W}_n(l) \triangleq \{[w_1, \dots, w_n] \in (\mathbb{Z}^+)^n : \sum_{q=1}^n w_q = l\}$, and $\text{PEP}(\mathbf{w})$ represents the probability of detecting a codeword with generalized weight \mathbf{w} instead of the transmitted all-zero codeword. Moreover, the WDS and in the GWDS are related via

$$\alpha(w) = \sum_{\mathbf{w} \in \mathcal{W}_n(w)} \beta(\mathbf{w}). \quad (4.36)$$

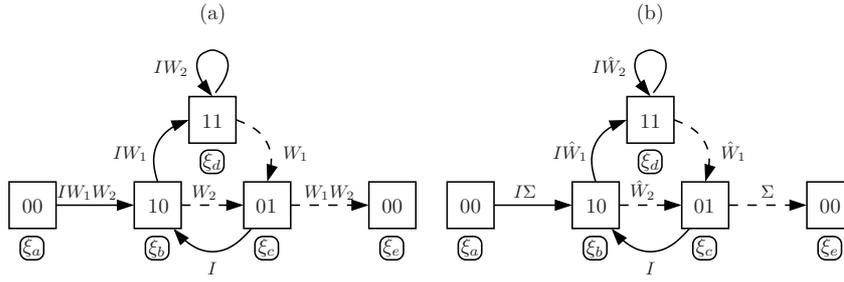


Fig. 4.18: Generalized state machine for the code $(5, 7)_8$: $\beta(w_1, w_2)$ in (a) and $\beta(\hat{w}_1, \hat{w}_2, w_\Sigma)$ in (b).

In the following example, we show how to compute this GWDS of the code as a generalization of the procedure in Example 4.33.

Example 16 (GWDS $\beta(w_1, w_2)$ of the $(5, 7)_8$ code) Consider the convolutional code analyzed in Example 15. To calculate the GWDS, a generalized transfer function can be used. In this case, however, instead of using only one dummy variable for the output weight, two dummy variables are used: W_1 and W_2 . The generalized state machine in this case is shown in Fig. 4.18 (a). The state equations in this case are

$$\xi_b = IW_1W_2\xi_a + I\xi_c \quad (4.37)$$

$$\xi_c = W_1\xi_d + W_2\xi_b \quad (4.38)$$

$$\xi_d = IW_1\xi_b + IW_2\xi_d \quad (4.39)$$

$$\xi_e = W_1W_2\xi_c. \quad (4.40)$$

We use again the equations (4.37)–(4.40) to obtain the generalized transfer function of the code $T(I, W_1, W_2)$ as

$$T(I, W_1, W_2) = \frac{\xi_e}{\xi_a} = \frac{IW_1^2W_2^2(I(W_1^2 - W_2^2) + W_2)}{1 - I(I(W_1^2 - W_2^2) + 2W_2)}.$$

The WDS of the code can be finally calculated using a generalization of (4.32):

$$\beta(w_1, w_2) = \frac{1}{w_1!w_2!} \frac{\partial^{w_1}}{\partial W_1^{w_1}} \frac{\partial^{w_2}}{\partial W_2^{w_2}} \frac{\partial}{\partial I} T(I, W_1, W_2) \Big|_{W_1=0, W_2=0, I=1}. \quad (4.41)$$

The generalized transfer function of the code was briefly introduced in the original paper of Zehavi [48, eq. (4.8)] and the GWDS in [70, Sec. IV-A].

We also note that the procedure presented in Example 16 ($n = 2$) can be straightforwardly extended to any value of n , and in general, the GWDS can be represented by an n -dimensional array. For the code $(5, 7)_8$, the first terms of the WDS and the GWDS are given, respectively, by

$$[\alpha(1) \quad \dots \quad \alpha(8)] = [0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 4 \quad \mathbf{12} \quad 32] \quad (4.42)$$

and

$$\begin{bmatrix} \beta(1, 1) & \dots & \beta(1, 7) \\ \vdots & \ddots & \vdots \\ \beta(7, 1) & \ddots & \beta(7, 7) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \mathbf{0} & 0 \\ 0 & 0 & 1 & 2 & \mathbf{3} & 4 & 5 \\ 0 & 0 & 0 & \mathbf{0} & 0 & 0 & 0 \\ 0 & 2 & \mathbf{9} & 24 & 50 & 90 & 147 \\ 0 & \mathbf{0} & 0 & 0 & 0 & 0 & 0 \\ \mathbf{0} & 4 & 25 & 90 & 245 & 560 & 1134 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4.43)$$

In (4.42) and (4.43), we highlighted the element $\alpha(7)$ and the elements $\beta(\mathbf{w})$ for all $\mathbf{w} \in \mathcal{W}_2(7)$; this can be useful to visualize the relation (4.36).

More details about the GWDS of convolutional codes can be found in Papers A, B, and C.

4.3.4 A New Generalized UB

Finally, if we consider a rate $R = 1/2$ ($k_c = 1$ and $n = 2$) code, a 4-PAM input alphabet, and a BICM configuration where the interleaver is not present (more details about this are given in Paper D), yet a different WDS of the code should be considered. In this case, the relevant WDS of the code should consider not the total Hamming weight w nor the generalized Hamming weight $[w_1, w_2]$, but a generalized Hamming weight $[\hat{w}_1, \hat{w}_2, w_\Sigma]$ where \hat{w}_1 , \hat{w}_2 , and w_Σ count, respectively, the number of columns of \mathbb{E} where

- $e_1[i] = 1$ and $e_2[i] = 0$,
- $e_1[i] = 0$ and $e_2[i] = 1$, and
- $e_1[i] = e_2[i] = 1$.

In this case, the GWDS of the code is of the form $\beta(\hat{w}_1, \hat{w}_2, w_\Sigma)$, and clearly

$$w = \hat{w}_1 + \hat{w}_2 + 2w_\Sigma. \quad (4.44)$$

In this case, the UB is a special case of (4.35), i.e.,

$$\text{UB} = \sum_{l=w^{\text{free}}}^{\infty} \sum_{\hat{\mathbf{w}} \in \mathcal{W}_2(l)} \beta([\hat{w}_1, \hat{w}_2, w_\Sigma]) \text{PEP}([\hat{w}_1, \hat{w}_2, w_\Sigma]), \quad (4.45)$$

where $\hat{\mathbf{w}} = [\hat{w}_1, \hat{w}_2, w_\Sigma]$ and $\beta([\hat{w}_1, \hat{w}_2, w_\Sigma])$ can be calculated using the methodology presented in Example 16. The only difference is that another dummy variable must be introduced, namely the variable Σ (for w_Σ). More details about this, as well as the PEP computation in (4.45) are given in Paper D.

For the code $(5, 7)_8$, and following a procedure analogous to the one in Example 16, we obtain the generalized transfer function

$$T(I, \hat{W}_1, \hat{W}_2, \Sigma) = \frac{I\Sigma^2(I(\hat{W}_1^2 - \hat{W}_2^2) + \hat{W}_2)}{1 - I(I(\hat{W}_1^2 - \hat{W}_2^2) + 2\hat{W}_2)}.$$

In Fig. 4.18 (b) we show the modified state machine, where the occurrences of W_1W_2 in Fig. 4.18 (a) are replaced by Σ .

4.3.5 Additional Remarks

The transfer function method used in the examples above is a well known method to find the WDS of a code analytically, cf. [56, Sec. 7.4.3], [9, Sec. 8.1–2], and as shown above, it can also be generalized to find in the GWDS of the code. An alternative analytical approach is the so-called signal flow graph approach [45, Sec. 11.2]. Alternative numerical methods for calculating the WDS of the code include the recursive algorithm in [128] and the breadth-first search algorithm in [129].

PEP calculations like the ones in (4.24) for BPSK or QPSK with the BRGC are straightforward since the PDF of the conditional L-values is known. In this case, the conditional L-values are i.i.d. with density given by (4.11), and thus, the PEP is given by (4.26). In this case, for a given Hamming weight w , the PEP obtained is the well known expression

$$\text{PEP}(w) = Q\left(\sqrt{2w\text{SNR}}\right).$$

However, PEP calculations become complicated when multilevel signal sets are considered. One alternative to exact PDF calculations is the Gaussian approximations presented in Sec. 4.1.1 and 4.1.2. In this case, since the PDFs are Gaussian mixtures, the PEP in (4.34) can be evaluated in closed form. Alternative methods for the PEP calculation include the expurgated bounds of [49, 130], the approximation of the BICM channel by a BPSK channel with scaled SNR of [131]. The main drawback of these approaches is that they rely on numerical integration. Another alternative is the use of the so-called saddlepoint approximation [115, 132]. This approximation is used in Paper C.

Chapter 5

Conclusions

The purpose of this thesis is to develop analytical tools for the analysis and design of BICM transmission. In the following sections, we describe the optimization opportunities discovered when analyzing BICM systems. We then present a short summary of the four papers appended in this thesis and present some open research problems in the field. Finally, a list of related publications by the author is also included.

5.1 Optimization Opportunities

5.1.1 Interleaver Alternatives for BICM and BICM-ID

Both the original BICM paper [48] and the paper introducing BICM-ID [63] postulated the application of M-interleavers (see Sec. 4.2.3) connecting each of the encoder's output to one modulator's input. For example, Zehavi's configuration [48] uses a rate-2/3 convolutional encoder and an 8-PSK input alphabet where three parallel interleavers connect the first/second/third encoder's output to the first/second/third modulator's input. Following the framework set in [49], since it simplifies the analysis, most of the existing literature on the design of BICM and BICM-ID assumes the use of one single interleaver (S-interleaver).

M-interleavers¹ have been used for BICM in [136] and for BICM-ID [66, 70, 137], for serially concatenated systems [133], and for BICM-OFDM [134]. M-interleavers have also been proposed in the 3GPP [138, Fig. 18], [2, Fig. 9.17] and DVB standards [22, Fig. 7a, Fig. 7b]. M-interleavers are relevant from an implementation point of view since for example, one of the reasons for using two parallel interleavers in HSPA with 16-QAM (or

¹Different names have been given to these interleavers: "in-line" [133], "intralevel" [134], "M" [135], "dual" [2], or "modular" [70].

three parallel interleavers for 64-QAM) is that the already implemented interleaver for 4-QAM can be “re-used”.

BICM with M-interleavers (BICM-M) and BICM with S-interleaver (BICM-S) were first compared in [135], where gains of 0.2–0.3 dB for BICM-M were reported; however, no explanation was given for these improvements. BICM-ID with M-interleavers (BICM-ID-M) was introduced in [66, 70, 135, 137], but also without theoretical justification. BICM-M and BICM-ID-M are formally analyzed in Papers A, B, and C, where we exploit the UEP introduced by the binary labeling, previously considered an “undesired feature” [49, Sec. II].

Another interleaver alternative recently proposed in [139] for BICM with convolutional codes is the use of a trivial interleaver (T-interleaver), i.e., no interleaving at all. Somehow surprisingly, it was shown in [139] that when BICM with T-interleavers (BICM-T) is used in nonfading channels, the performance of BICM can be significantly improved. BICM-T is analyzed in Paper D.

5.1.2 Code Selection

The best known convolutional codes are the so-called optimum distance spectrum (ODS) codes tabulated for example in [140]. ODS convolutional codes are found by exhaustive search and are characterized by their WDS $\alpha(w)$, cf. (4.27). Their optimality is based on a UB on the BER for asymptotically high SNR and on the inputs to the decoder being i.i.d., for example, when BPSK modulation is used over the AWGN channel.

The performance of BICM-S can be analyzed using (4.27), and therefore, the optimum codes are the ODS codes. However, the performance of BICM-M and BICM-ID-M should be analyzed using (4.35) instead. In this case, the UB is based on the GWDS of the code $\beta(\mathbf{w})$ (and not on $\alpha(w)$), and therefore, the optimality of the ODS codes is questionable. The optimality of the ODS codes is analyzed in Paper A, B, and D.

When BICM-T (for a rate $R = 1/2$ code and a 4-PAM signal set) is analyzed, the optimal codes are designed based on $\beta(\hat{w}_1, \hat{w}_2, w_\Sigma)$, cf. Sec. 4.3.4. Optimal convolutional codes for BICM-T are analyzed in Paper D.

5.1.3 Signal Shaping

Signal shaping refers to the use of nonequally spaced input alphabets and/or nonequally likely symbols, which, following the terminology in [94], are called geometrical shaping and probabilistic shaping, respectively.

Signal shaping is in fact a subject that has been studied during many years, cf. [100–103] and references therein. In the context of BICM, however, shaping techniques have not received a lot of attention. Only recently, geometrical shaping for BICM was proposed in [94, 141, 142], and probabilistic

shaping in [114]. All these approaches aim to optimize the BICM capacity for particular scenarios and under particular constraints.

In this thesis, we are interested in geometrical shaping for BICM. In principle, when designing a nonequally spaced input alphabet, all the points in the input alphabet can move freely in signal space. However, for simplicity, we will focus on input alphabets with certain symmetries, i.e., the HPAM and HQAM signal sets defined in Sec. 1.3.2. The design of BICM-M systems using HQAM signal sets in combination with M-interleavers and a deterministic bit-level multiplexer is studied in Paper C.

5.2 Included Papers

Paper A: “Exploiting UEP in QAM-based BICM: Interleaver and Code Design”

In Paper A, we formally analyze BICM-M using QAM input alphabets labeled by the BRGC. We show that the BRGC introduces UEP and also show how this can be exploited by using M-interleavers. A GWDS for the code is introduced. Optimum convolutional codes for BICM-M are tabulated. Numerical results show that gains can be obtained for both turbo codes and convolutional codes without increasing the complexity of the receiver.

Paper B: “On BICM-ID with Multiple Interleavers”

In Paper B, we analyze BICM-ID-M. We show that under the PF assumption, good binary labelings for BICM-ID introduce UEP, and that this can be exploited by using M-interleavers. We show that gains obtained by using M-interleavers instead of S-interleavers appear even when a very simple configuration is used (4-QAM and a constraint length $K = 3$ ODS convolutional code). The asymptotic optimality of BICM-ID-M over BICM-ID-S is proved. Optimal convolutional codes for BICM-ID-M are also analyzed.

Paper C: “Towards Fully Optimized BICM Transceivers”

In Paper C, we study geometrical shaping for BICM in fading channels based on HQAM signal sets, and we combine them with M-interleavers and a bit-wise multiplexing. Closed-form expressions for the PDF of the L-values are obtained as a function of the distances defining the HPAM signal set. A new WDS of the code is also introduced. Optimal signal sets and multiplexers for different channels and SNRs are tabulated. The gains obtained by using this optimized BICM configuration instead of the traditional one (equally spaced signal sets and S-interleavers) are quantified.

Paper D: “On BICM receivers for TCM transmission”

In Paper D, we formally explain why the use of BICM-T in nonfading channels results in performance gains. It is shown that BICM-T is simply a BRGC-based TCM transmitter and a BICM receiver. It is also shown that a properly designed BICM-T system performs asymptotically as well as TCM. Optimum convolutional codes are tabulated, and a new WDS for the code is introduced.

5.3 Future Work

In this thesis, different aspects of BICM are studied. In particular, UEP in BICM and BICM-ID and different weight distribution spectra of the code are analyzed, and different interleaver alternatives as well as geometrical shaping for BICM are formally studied.

The main contribution of this thesis is to show that standard BICM designs are suboptimal, and also to propose new (improved) designs. We have advanced one step forward towards fully optimized BICM designs, however, there are still many open research questions to be answered. Some of these are detailed below.

The full potential of the use of M-interleavers in a more general scenario where the channel introduces UEP is still unknown. For example, M-interleavers could be designed for MIMO channels where parallel channels are available after the singular value decomposition of the channel. M-interleavers could also be designed for OFDM systems where groups of subcarriers have different SNR.

As mentioned before, signal shaping for BICM has not been studied in detail. In particular, the question about optimal unequally spaced signal sets with full freedom is still unanswered. The same applies to probabilistic shaping, as well as to the relation between geometrical and probabilistic shaping in BICM.

Apart from the results at asymptotically low SNR or other very particular cases, there are no general results on capacity-maximizing labeled constellations, i.e., it is still unknown what the optimal binary labelings, signal sets, and input distributions for BICM are.

5.4 Related Contributions

Other related publications by the author, which are not included in this thesis, are listed below.

- [J4] **A. Alvarado**, E. Agrell, A. Guillén i Fàbregas, and A. Martinez, “Corrections to ‘Bit-interleaved coded modulation in the wideband regime’,” *IEEE Trans. Inf. Theory*, vol. 56, no. 12, p. 6513, Dec. 2010.

- [J3] E. Agrell and **A. Alvarado**, “On the BICM capacity,” *submitted to IEEE Trans. Inf. Theory*, Jan. 2010 (revised in Dec. 2010), available at <http://arxiv.org/abs/1001.4548>.
- [J2] L. Szczecinski, **A. Alvarado**, and R. Feick, “Distribution of max-log metrics for QAM-based BICM in fading channels,” *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2558–2563, Sep. 2009.
- [J1] **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.
- [C6] Md. J. Hossain, **A. Alvarado**, and L. Szczecinski, “BICM transmission using non-uniform QAM constellations: Performance analysis and design,” in *IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010.
- [C5] E. Agrell and **A. Alvarado**, “On optimal constellations for BICM at low SNR,” in *IEEE Information Theory Workshop (ITW)*, Taormina, Italy, Oct. 2009, (Best Poster Award).
- [C4] **A. Alvarado**, V. Núñez, L. Szczecinski, and E. Agrell, “Correcting suboptimal metrics in iterative decoders,” in *IEEE International Conference on Communications (ICC)*, Dresden, Germany, June 2009.
- [C3] **A. Alvarado**, E. Agrell, and A. Svensson, “On the capacity of BICM with QAM constellations,” in *International Wireless Communications and Mobile Computing Conference 2009 (IWCMC)*, Leipzig, Germany, June 2009, (Invited Paper).
- [C2] **A. Alvarado**, L. Szczecinski, and R. Feick, “On the distribution of extrinsic L-values in gray-mapped 16-QAM,” in *AMC International Wireless Communications and Mobile Computing Conference (IWCMC)*, Honolulu, HI, USA, Aug. 2007.
- [C1] **A. Alvarado**, H. Carrasco, and R. Feick, “On adaptive BICM with finite block-length and simplified metrics calculation,” in *IEEE Vehicular Technology Conference (VTC-Fall)*, Montreal, QC, Canada, Sep. 2006.

Appendix: Numerical Implementation of the CM and BICM Capacities

The Gauss–Hermite quadrature approximates integrals as [143, Sec. 7.3.4]

$$\int_{-\infty}^{\infty} g(t)e^{-t^2} dt \approx \sum_{k=1}^J \alpha_k g(\xi_k), \quad (5.1)$$

where ξ_k is the k th zero of the J th Hermite polynomial $H_J(x)$ and $\alpha_k = 2^{J+1} J! \sqrt{\pi} / [H_{J+1}(\xi_k)]^2$. Tables with α_k and ξ_k for different values of J can be found for example in [143, Appendix 7.3(b)]. The expression in (5.1) can be generalized to an N -dimensional vector $\mathbf{t} = [t_1, \dots, t_N]$ as²

$$\int_{\mathbb{R}^N} g(\mathbf{t})e^{-\|\mathbf{t}\|^2} d\mathbf{t} \approx \sum_{k_1=1}^J \dots \sum_{k_N=1}^J g(\xi_{k_1}, \dots, \xi_{k_N}) \prod_{i=1}^N \alpha_{k_i}. \quad (5.2)$$

Using (1.3) and the change of variables $\sqrt{N_0}\mathbf{t} = \mathbf{y} - \mathbf{x}_i$, the AMI in (3.10) for the AWGN channel can be expressed as

$$I_{\mathbf{X}}(\mathbf{X}; \mathbf{Y}) = \frac{1}{\pi^{N/2}} \sum_{i=1}^M P_{\mathbf{X}}(\mathbf{x}_i) \int_{\mathbb{R}^N} e^{-\|\mathbf{t}\|^2} \log_2 \frac{e^{-\|\mathbf{t}\|^2}}{\sum_{\mathbf{x} \in \mathcal{X}} P_{\mathbf{X}}(\mathbf{x}) e^{-\frac{1}{N_0} \|\sqrt{N_0}\mathbf{t} + \mathbf{x}_i - \mathbf{x}\|^2}} d\mathbf{t}.$$

Using (5.2), the CM capacity in (3.11) can be approximated as

$$I_{\Omega}^{\text{CM}}(\text{SNR}) \approx \frac{1}{\pi^{N/2}} \sum_{i=1}^M P_{\mathbf{X}}(\mathbf{x}_i) \sum_{k_1=1}^J \dots \sum_{k_N=1}^J g_i^{\text{CM}}(\xi_{k_1}, \dots, \xi_{k_N}) \prod_{n=1}^N \alpha_{k_n},$$

where

$$g_i^{\text{CM}}(\mathbf{t}) \triangleq -\|\mathbf{t}\|^2 \log_2 e - \log_2 \sum_{\mathbf{x} \in \mathcal{X}} P_{\mathbf{X}}(\mathbf{x}) e^{-\frac{1}{N_0} \|\sqrt{N_0}\mathbf{t} + \mathbf{x}_i - \mathbf{x}\|^2},$$

with $i = 1, \dots, M$. Using an analogous procedure, the BICM capacity in (3.23) can be approximated as

$$I_{\Omega}^{\text{BI}}(\text{SNR}) \approx \frac{1}{\pi^{N/2}} \sum_{k=1}^m \sum_{u \in \{0,1\}} \sum_{i \in \mathcal{I}_{k,u}} P_{\mathbf{X}}(\mathbf{x}_i) \sum_{k_1=1}^J \dots \sum_{k_N=1}^J g_i^{\text{BI}}(\xi_{k_1}, \dots, \xi_{k_N}) \prod_{n=1}^N \alpha_{k_n},$$

where

$$g_i^{\text{BI}}(\mathbf{t}) \triangleq \log_2 \frac{\frac{1}{P_{C_k}(u)} \sum_{j \in \mathcal{I}_{k,u}} P_{\mathbf{X}}(\mathbf{x}_j) e^{-\frac{1}{N_0} \|\sqrt{N_0}\mathbf{t} + \mathbf{x}_i - \mathbf{x}_j\|^2}}{\sum_{\mathbf{x} \in \mathcal{X}} P_{\mathbf{X}}(\mathbf{x}) e^{-\frac{1}{N_0} \|\sqrt{N_0}\mathbf{t} + \mathbf{x}_i - \mathbf{x}\|^2}}.$$

²We note that the selection of an equally spaced grid for the multidimensional integral in (5.2) is suboptimal; however, we use it since it is the simplest alternative.

Bibliography

- [1] 3GPP, “Universal mobile telecommunications system (UMTS); multiplexing and channel coding (FDD),” 3GPP, Tech. Rep. TS 125.212, V7.11.0 Release 7, Sep. 2009.
- [2] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, 2nd ed. Academic Press, 2008.
- [3] IEEE 802.11, “Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5GHz band,” IEEE Std 802.11a-1999(R2003), Tech. Rep., July 1999.
- [4] IEEE 802.11n, “Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. Amendment 5: Enhancements for higher throughput,” IEEE Std 802.11n-2009, Tech. Rep., Oct. 2009.
- [5] 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 V1.1.1 (2009-09), Sep. 2009.
- [6] —, “Digital video broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications (DVB-S2),” ETSI, Tech. Rep. ETSI EN 302 307 V1.2.1 (2009-08), Aug. 2009.
- [7] —, “Digital video broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital transmission system for cable system (DVB-C2),” ETSI, Tech. Rep. ETSI EN 302 769 V1.1.1 (2010-04), Apr. 2010.
- [8] J. B. Anderson and A. Svensson, *Coded Modulation Systems*. Springer, 2003.

-
- [9] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-Hill, 2008.
- [10] A. J. Viterbi and J. K. Omura, *Principles of Digital Communications and Coding*. McGraw-Hill, 1979.
- [11] A. Goldsmith, *Wireless Communications*. New York, NY: Cambridge University Press, 2005.
- [12] M. Nakagami, “The m -distribution—A general formula of intensity distribution of rapid fading,” in *Statistical Methods in Radio Wave Propagation*, W. G. Hoffman, Ed. Oxford, U.K.: Pergamon, 1960.
- [13] 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.
- [14] F. Gray, “Pulse code communications,” U. S. Patent 2 632 058, Mar. 1953.
- [15] 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.
- [16] 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, USA, Oct. 2003.
- [17] M. Morimoto, M. Okada, and S. Komaki, “A hierarchical image transmission system in fading channel,” in *IEEE International Conference on Universal Personal Communications (ICUPC)*, Tokyo, Japan, Oct. 1995, pp. 769–772.
- [18] P. K. Vitthaladevuni and M.-S. Alouini, “A recursive algorithm for the exact BER computation of generalized hierarchical QAM constellations,” *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 297–307, Jan. 2003.
- [19] Md. J. Hossain, M.-S. Alouini, and V. K. Bhargava, “Hierarchical constellation for multi-resolution data transmission over block fading channels,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 849–857, Apr. 2006.
- [20] Md. J. Hossain, P. K. Vitthaladevuni, M.-S. Alouini, V. K. Bhargava, and A. J. Goldsmith, “Adaptive hierarchical constellations for simultaneous voice and multi-class data transmission over fading channels,” *IEEE Trans. Veh. Technol.*, vol. 55, no. 4, pp. 1181–1194, July 2006.

- [21] M. R. Chari, F. Ling, A. Mantravadi, R. Krishnamoorthi, R. Vijayan, G. K. Walker, and R. Chandhok, "FLO physical layer: an overview," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 107–145, Mar. 2007.
- [22] ETSI, "Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television," ETSI, Tech. Rep. ETSI EN 300 744 V1.6.1 (2009-01), Jan. 2009.
- [23] E. Agrell and A. Alvarado, "On the BICM capacity," *submitted to IEEE Trans. Inf. Theory*, Jan. 2010 (revised in Dec. 2010), available at <http://arxiv.org/abs/1001.4548>.
- [24] H. Nyquist, "Certain factors affecting telegraph speed," *Bell System Technical Journal*, pp. 324–346, Apr. 1924.
- [25] —, "Certain topics in telegraph transmission theory," *Trans. of the A.I.E.E.*, vol. 47, no. 2, pp. 617–644, Apr. 1928.
- [26] R. V. L. Hartley, "Transmission of information," *Bell System Technical Journal*, pp. 535–563, July 1928.
- [27] C. E. Shannon, "A mathematical theory of communications," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.
- [28] —, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.
- [29] J. R. Pierce, "The early days of information theory," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 1, pp. 3–8, Jan. 1973, (Invited Paper).
- [30] 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.
- [31] R. Gallager, "Low-density parity-check codes," *IRE Trans. Information Theory*, pp. 21–28, Jan. 1962.
- [32] 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.
- [33] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.

- [34] S. Chung, G. D. Forney, Jr., T. J. Richardson, and R. L. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, no. 2, pp. 58–60, Feb. 2001.
- [35] R. de Buda, "Fast FSK signals and their demodulation," *Can. Elec. Eng. J.*, vol. 1, pp. 28–34, 1972.
- [36] —, "Coherent demodulation of frequency-shift keying with low deviation ratio," *IEEE Trans. Commun.*, vol. COM-20, pp. 429–435, June 1972.
- [37] J. L. Massey, "Coding and modulation in digital communications," in *International Zurich Seminar on Digital Communications*, Zurich, Switzerland, Mar. 1974.
- [38] H. Miyakawa, H. Harashima, and Y. Tanaka, "A new digital modulation scheme—Multi-mode binary cfsk," in *3rd Int. Conf. Digital Satellite Communications*, Kyoto, Japan, Nov. 1975.
- [39] J. B. Anderson and D. P. Taylor, "A bandwidth-efficient class of signal-space codes," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 6, pp. 703–712, 1978.
- [40] T. Aulin, "CPM—A power and bandwidth efficient digital constant envelope modulation scheme," Ph.D. dissertation, Lund University, Lund, Sweden, Nov. 1979.
- [41] 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).
- [42] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.
- [43] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.
- [44] —, "Correction to 'A new multilevel coding method using error-correcting codes'," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 6, p. 784, Nov. 1977.
- [45] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2004.

- [46] D. J. Costello, Jr. and G. D. Forney, Jr., "Channel coding: The road to channel capacity," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1150–1177, June 2007.
- [47] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998, (Invited Paper).
- [48] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.
- [49] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [50] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets Part I: Introduction," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 5–11, Feb. 1987.
- [51] —, "Trellis-coded modulation with redundant signal sets Part II: State of the art," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 12–21, Feb. 1987.
- [52] ITU, "A family of 2-wire, duplex modems operating at data signalling rates of up to 9600 bit/s for use on the general switched telephone network and on leased telephone-type circuits," ITU-T Recommendation V.32, Tech. Rep., Mar. 1993.
- [53] —, "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.
- [54] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. Prentice Hall, 1992.
- [55] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*. Kluwer Academic Publishers, 1994.
- [56] A. Burr, *Modulation and coding for wireless communications*. Prentice Hall, 2001.
- [57] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Prentice Hall, 1995.
- [58] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.

- [59] L. Beygi, E. Agrell, M. Karlsson, and B. Makki, "A novel rate allocation method for multilevel coded modulation," in *IEEE International Symposium on Information Theory (ISIT)*, Austin, TX, USA, June 2010.
- [60] Y. Kofman, E. Zehavi, and S. Shamai, "Performance analysis of a multilevel coded modulation system," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 299–312, Feb./Mar./Apr. 1994.
- [61] U. Wachsmann and J. B. Huber, "Power and bandwidth efficient digital communication using turbo codes in multilevel codes," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 557–567, Sep. 1995.
- [62] 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.
- [63] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 169–171, Nov. 1997.
- [64] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping for QPSK modulation," *IEE Electronics Letters*, vol. 34, no. 15, pp. 1459–1460, July 1998.
- [65] S. Benedetto, G. Montorsi, D. Divsalar, and F. Pollara, "Soft-input soft-output modules for the construction and distributed iterative decoding of code networks," *Eur. Trans. on Telecommun.*, vol. 9, no. 2, pp. 155–172, Mar.–Apr. 1998.
- [66] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 944–957, May 2001.
- [67] M. Tüchler, "Design of serially concatenated systems depending on the block length," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 209–218, Feb. 2004.
- [68] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 7, no. 12, pp. 593–595, Dec. 2003.
- [69] L. Szczecinski, H. Chafnaji, and C. Hermosilla, "Modulation doping for iterative demapping of bit-interleaved coded modulation," *IEEE Commun. Lett.*, vol. 9, no. 12, pp. 1031–1033, Dec. 2005.

- [70] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interlaved coded modulation with iterative decoding and 8PSK signaling," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 1250–1257, Aug. 2002.
- [71] J. Tan and G. L. Stüber, "Analysis and design of interleaver mappings for iteratively decoded BICM," in *IEEE International Conference on Communications (ICC)*, New York, USA, May 2002.
- [72] S. ten Brink, "Convergence behaviour of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
- [73] L. Zhao, L. Lampe, and J. Huber, "Study of bit-interleaved coded space-time modulation with different labeling," in *IEEE Information Theory Workshop (ITW)*, Paris, France, Mar. 2003.
- [74] T. Clevorn, S. Godtmann, and P. Vary, "Optimized mappings for iteratively decoded BICM on rayleigh channels with IQ interleaving," in *IEEE Vehicular Technology Conference (VTC-Spring)*, Melbourne, Australia, May 2006.
- [75] J. Tan and G. L. Stüber, "Analysis and design of symbol mappers for iteratively decoded BICM," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 662–672, Mar. 2005.
- [76] F. Schreckenbach, "Iterative decoding of bit-interleaved coded modulation," Ph.D. dissertation, Technische Universität München, Munich, Germany, 2007, available at <http://mediatum2.ub.tum.de/doc/644160/644160.pdf>.
- [77] 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.
- [78] Ericsson, Motorola, and Nokia, "Link evaluation methods for high speed downlink packet access (HSDPA)," TSG-RAN Working Group 1 Meeting #15, TSGR1#15(00)1093, Tech. Rep., Aug. 2000.
- [79] B. Classon, K. Blankenship, and V. Desai, "Channel coding for 4G systems with adaptive modulation and coding," *IEEE Wireless Commun. Mag.*, vol. 9, no. 2, pp. 8–13, Apr. 2002.
- [80] A. Alvarado, H. Carrasco, and R. Feick, "On adaptive BICM with finite block-length and simplified metrics calculation," in *IEEE Vehicular Technology Conference (VTC-Fall)*, Montreal, QC, Canada, Sep. 2006.

-
- [81] K. Hyun and D. Yoon, "Bit metric generation for Gray coded QAM signals," *IEE Proc.-Commun.*, vol. 152, no. 6, pp. 1134–1138, Dec. 2005.
- [82] 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.
- [83] R. H. Morelos-Zaragoza, *The Art of Error Correcting Codes*, 1st ed. New York, NY: John Wiley & Sons, 2002.
- [84] 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)*, June 1995.
- [85] M. Valenti and J. Sun, "The UMTS turbo code and an efficient decoder implementation suitable for software defined radios," *International Journal on Wireless Information Networks*, vol. 8, no. 4, pp. 203–216, Oct. 2001.
- [86] W. Gross and P. Gulak, "Simplified MAP algorithm suitable for implementation of turbo decoders," *Electronics Letters*, vol. 34, no. 16, pp. 1577–1578, Aug. 1998.
- [87] F. Simoens, H. Wymeersch, and M. Moeneclaey, "Linear precoders for bit-interleaved coded modulation on AWGN channels: Analysis and design criteria," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 87–99, Jan. 2008.
- [88] S. Pfletschinger and F. Sanzi, "Error floor removal for bit-interleaved coded modulation with iterative detection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3174–3181, Nov. 2006.
- [89] A. A. Mohammed, Y. Gong, and M. Bossert, "On multidimensional BICM-ID with 8-PSK constellation labeling," in *IEEE International Symposium on Personal, Indoor and Mobile Communications (PIMRC)*, Athens, Greece, Sep. 2007.
- [90] X. Zou, M. Wang, and G. Feng, "A new interleaver design for iteratively decoded bit-interleaved coded modulation," *International Journal of Soft Computing*, vol. 3, no. 5, pp. 338–343, 2008.
- [91] T. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1629–1638, June 2002.

- [92] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, USA: John Wiley & Sons, 2006.
- [93] D. Guo, S. Shamai, and S. Verdú, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
- [94] M. Barsoum, C. Jones, and M. Fitz, “Constellation design via capacity maximization,” in *IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007.
- [95] F. Brännström and L. K. Rasmussen, “Classification of unique mappings for 8PSK based on bit-wise distance spectra,” *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1131–1145, Mar. 2009.
- [96] A. Martinez, A. Guillén i Fàbregas, and G. Caire, “Bit-interleaved coded modulation in the wideband regime,” *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5447–5455, Dec. 2008.
- [97] —, “Bit-interleaved coded modulation revisited: A mismatched decoding perspective,” *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.
- [98] C. Stierstorfer and R. F. H. Fischer, “Asymptotically optimal mappings for BICM with M -PAM and M^2 -QAM,” *IET Electronics Letters*, vol. 45, no. 3, pp. 173–174, Jan. 2009.
- [99] 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/>.
- [100] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, 2002.
- [101] G. D. Forney, Jr. and L.-F. Wei, “Multidimensional constellations—Part I: Introduction, figures of merit, and generalized cross constellations,” *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.
- [102] G. D. Forney, Jr., “Multidimensional constellations—Part II: Voronoi constellations,” *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 941–957, Aug. 1989.
- [103] A. R. Calderbank and L. H. Ozarow, “Nonequiprobable signaling on the Gaussian channel,” *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 726–740, July 1990.

- [104] S. Le Goff, B. S. Sharif, and S. A. Jimaa, "Bit-interleaved turbo-coded modulation using shaping coding," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 246–248, Mar. 2005.
- [105] S. Le Goff, B. K. Khoo, C. C. Tsimenidis, and B. S. Sharif, "Constellation shaping for bandwidth-efficient turbo-coded modulation with iterative receiver," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2223–2233, June 2007.
- [106] D. Raphaeli and A. Gurevitz, "Constellation shaping for pragmatic turbo-coded modulation with high spectral efficiency," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 341–345, Mar. 2004.
- [107] L. Szczecinski, R. Bettancourt, and R. Feick, "Probability density function of reliability metrics in BICM with arbitrary modulation: Closed-form through algorithmic approach," *IEEE Trans. Commun.*, vol. 56, no. 5, pp. 736–742, May 2008.
- [108] 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.
- [109] L. Szczecinski, A. Alvarado, and R. Feick, "Distribution of max-log metrics for QAM-based BICM in fading channels," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2558–2563, Sep. 2009.
- [110] A. Kenarsari-Anhari and L. Lampe, "An analytical approach for performance evaluation of BICM over Nakagami- m fading channels," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1090–1101, Apr. 2010.
- [111] C. Stierstorfer and R. F. H. Fischer, "(Gray) Mappings for bit-interleaved coded modulation," in *IEEE Vehicular Technology Conference (VTC-Spring)*, Dublin, Ireland, Apr. 2007.
- [112] A. Alvarado, E. Agrell, A. Guillén i Fàbregas, and A. Martinez, "Corrections to 'Bit-interleaved coded modulation in the wideband regime'," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, p. 6513, Dec. 2010.
- [113] F. Brännström, "Convergence analysis and design of multiple concatenated codes," Ph.D. dissertation, Chalmers University of Technology, Göteborg, Sweden, Mar. 2004.
- [114] A. Guillén i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *IEEE Information Theory Workshop (ITW)*, Dublin, Ireland, Aug.–Sep. 2010.

- [115] A. Martinez, A. Guillén i Fàbregas, and G. Caire, “Error probability analysis of bit-interleaved coded modulation,” *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 262–271, Jan. 2006.
- [116] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch, “Optimized symbol mappings for bit-interleaved coded modulation with iterative decoding,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, San Francisco, CA, USA, Dec. 2003.
- [117] F. Simoens, H. Wymeersch, H. Bruneel, and M. Moeneclaey, “Multidimensional mapping for bit-interleaved coded modulation with BPSK/QPSK signaling,” *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 453–455, May 2005.
- [118] K. Zeger and A. Gersho, “Pseudo-Gray coding,” *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2147–2158, Dec. 1990.
- [119] M. Benjillali, L. Szczecinski, and S. Aissa, “Probability density functions of logarithmic likelihood ratios in rectangular QAM,” in *Twenty-Third Biennial Symposium on Communications*, Kingston, ON, Canada, May 2006.
- [120] L. Szczecinski and M. Benjillali, “Probability density functions of logarithmic likelihood ratios in phase shift keying BICM,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, San Francisco, USA, Nov. 2006.
- [121] 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.*, pp. 831–844, Sep. 2008.
- [122] C. Hermosilla and L. Szczecinski, “Performance evaluation of linear turbo-receivers using analytical extrinsic information transfer functions,” *EURASIP Journal on Applied Signal Processing*, pp. 892–905, May 2005.
- [123] E. Rosnes and Ø. Ytrehus, “On the design of bit-interleaved turbo-coded modulation with low error floors,” *IEEE Trans. Commun.*, vol. 54, no. 9, pp. 1563–1573, Sep. 2006.
- [124] L. Hanzo, W. Webb, and T. Keller, *Single- and Multi-carrier Quadrature Amplitude Modulation*. John Wiley & Sons, 2000.
- [125] N. H. Tran and H. H. Nguyen, “Signal mappings of 8-ary constellations for bit interleaved coded modulation with iterative decoding,” *IEEE Trans. Broadcast.*, vol. 52, no. 1, pp. 92–99, Mar. 2006.

- [126] S. Benedetto and G. Montorsi, "Average performance of parallel concatenated block codes," *Electronics Letters*, vol. 31, no. 3, pp. 156–158, Feb. 1995.
- [127] —, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 409–428, Mar. 1996.
- [128] D. Divsalar, S. Dolinar, R. J. McEliece, and F. Pollara, "Transfer function bounds on the performance of turbo codes," Jet Propulsion Laboratory, Pasadena, CA, TDA Progr. Rep. 42-121, Aug. 1995.
- [129] J. Belzile and D. Haccoun, "Bidirectional breadth-first algorithms for the decoding of convolutional codes," *IEEE Trans. Commun.*, vol. 41, no. 2, pp. 370–380, Feb. 1993.
- [130] V. Sethuraman and B. Hajek, "Comments on 'Bit-interleaved coded modulation'," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1795–1797, Apr. 2006.
- [131] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Error probability of bit-interleaved coded modulation using the Gaussian approximation," in *Conference on Information Sciences and Systems (CISS)*, Princeton University, NJ, USA, Mar. 2004.
- [132] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "A closed-form approximation for the error probability of BPSK fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2051–2054, June 2007.
- [133] A. Nilsson and T. M. Aulin, "On in-line bit interleaving for serially concatenated systems," in *IEEE International Conference on Communications (ICC)*, May 2005.
- [134] C. Stierstorfer and R. F. H. Fischer, "Intralevel interleaving for BICM in OFDM scenarios," in *12th International OFDM Workshop*, Hamburg, Germany, Aug. 2007.
- [135] I. Abramovici and S. Shamai, "On turbo encoded BICM," *Annales des Telecommunications*, vol. 54, no. 3–4, pp. 225–234, 1999.
- [136] U. Hansson and T. Aulin, "Channel symbol expansion diversity—improved coded modulation for the Rayleigh fading channel," in *IEEE International Conference on Communications (ICC)*, Dallas, USA, June 1996.
- [137] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 715–724, Apr. 1999.

-
- [138] 3GPP, “Universal mobile telecommunications system (UMTS); multiplexing and channel coding (FDD),” 3GPP, Tech. Rep. TS 25.212, V8.5.0 Release 8, Mar. 2009.
- [139] 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.
- [140] P. Frenger, P. Orten, and T. Ottosson, “Convolutional codes with optimum distance spectrum,” *IEEE Trans. Commun.*, vol. 3, no. 11, pp. 317–319, Nov. 1999.
- [141] D. Sommer and G. P. Fettweis, “Signal shaping by non-uniform QAM for AWGN channels and applications using turbo coding,” in *International ITG Conference on Source and Channel Coding (SCC)*, Munich, Germany, Jan. 2000.
- [142] S. Y. L. Goff, “Signal constellations for bit-interleaved coded modulation,” *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 307–313, Jan. 2003.
- [143] F. B. Hildebrand, Ed., *Handbook of applicable mathematics. Vol 3: Numerical Methods*. John Wiley & Sons, 1981.

Part II

Included Papers

