Foundations and Trends® in
sample
Vol. xx, No xx (xxxx) 1–144

DOI: xxxxxx

now
the essence of knowledge

Bit-Interleaved Coded Modulation

Albert Guillén i Fàbregas[1], Alfonso Martinez[2] and Giuseppe Caire[3]

[1] *Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom, guillen@ieee.org*
[2] *Centrum Wiskunde & Informatica (CWI), Kruislaan 413, Amsterdam, 1098 SJ, The Netherlands, alfonso.martinez@ieee.org*
[3] *Electrical Engineering Department, University of Southern California, 3740 McClintock Av., Los Angeles, 90080 CA, USA, caire@usc.edu*

Abstract

The principle of coding in the signal space follows directly from Shannon's analysis of waveform Gaussian channels subject to an input constraint. The early design of communication systems focused separately on *modulation*, namely signal design and detection, and *error correcting codes*, which deal with errors introduced at the demodulator of the underlying waveform channel. The correct perspective of signal-space coding, although never out of sight of information theorists, was brought back into the focus of coding theorists and system designers by Imai's and Ungerböck's pioneering work on coded modulation. More recently, powerful families of binary codes with a good tradeoff between performance and decoding complexity have been (re-) discovered. Bit-Interleaved Coded Modulation (BICM) is a pragmatic approach combining the best out of both worlds: it takes advantage of the signal-space coding perspective, whilst allowing for the use of powerful families of binary codes with virtually any modulation format.

BICM avoids the need for the complicated and somewhat less flexible design typical of coded modulation. As a matter of fact, most of today's systems that achieve high spectral efficiency such as DSL, Wireless LANs, WiMax and evolutions thereof, as well as systems based on low spectral efficiency orthogonal modulation, feature BICM, making BICM the *de-facto* general coding technique for waveform channels.
The theoretical characterization of BICM is at the basis of efficient coding design techniques and also of improved BICM decoders, e.g., those based on the belief propagation iterative algorithm and approximations thereof. In this monograph, we review the theoretical foundations of BICM under the unified framework of error exponents for mismatched decoding. This framework allows an accurate analysis without any particular assumptions on the length of the interleaver or independence between the multiple bits in a symbol. We further consider the sensitivity of the BICM capacity with respect to the signal-to-noise ratio (SNR), and obtain a wideband regime (or low-SNR regime) characterization. We review efficient tools for the error probability analysis of BICM that go beyond the standard approach of considering infinite interleaving and take into consideration the dependency of the coded bit observations introduced by the modulation. We also present bounds that improve upon the union bound in the region beyond the cutoff rate, and are essential to characterize the performance of modern randomlike codes used in concatenation with BICM. Finally, we turn our attention to BICM with iterative decoding, we review extrinsic information transfer charts, the area theorem and code design via curve fitting. We conclude with an overview of some applications of BICM beyond the classical coherent Gaussian channel.

Contents

List of Abbreviations, Acronyms and Symbols

APP	A posteriori probability
AWGN	Additive white Gaussian noise
BEC	Binary erasure channel
BICM	Bit-interleaved coded modulation
BICM-ID	Bit-interleaved coded modulation with iterative decoding
BIOS	Binary-input output-symmetric (channel)
BP	Belief propagation
CM	Coded modulation
EXIT	Extrinsic information transfer
FG	Factor graph
GMI	Generalized mutual information
ISI	Inter-symbol interference
LDPC	Low-density parity-check (code)
MAP	Maximum a posteriori
MIMO	Multiple-input multiple-output
MLC	Multi-level coding
MMSE	Minimum mean-squared error
MSD	Multi-stage decoding
PSK	Phase-shift keying

QAM	Quadrature-amplitude modulation
RA	Repeat-accumulate (code)
SNR	Signal-to-noise ratio
TCM	Trellis-coded modulation
TSB	Tangential sphere bound
A_d	Weight enumerator at Hamming distance d
A_{d,ρ_N}	Weight enumerator at Hamming distance d and pattern ρ_N
A'_d	Bit weight enumerator at Hamming distance d
b	Bit in codeword
$\bar{b}$	Binary complement of bit b
$b_j(x)$	Inverse labeling (mapping) function
C	Channel capacity
$\mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}}$	BICM capacity over set $\mathcal{X}$
$\mathsf{C}_{\mathrm{bpsk}}$	BPSK capacity
$\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}$	Coded modulation capacity over set $\mathcal{X}$
$\mathcal{C}$	Binary code
c_1	First-order Taylor capacity series coefficient
c_2	Second-order Taylor capacity series coefficient
d	Hamming distance
$\boldsymbol{d}$	Scrambling (randomization) sequence
$\bar{d}_{\mathrm{v}}$	Average variable degree (LDPC code)
$\bar{d}_{\mathrm{c}}$	Average check node degree (LDPC code)
ΔP	Power expansion ratio
ΔW	Bandwidth expansion ratio
$\mathbb{E}[U]$	Expectation (of a random variable U)
E_b	Average bit energy
$\frac{E_{\mathrm{b}}}{N_0}$	Ratio between average bit energy and noise spectral density
$\frac{E_{\mathrm{b}}}{N_0\,\mathrm{lim}}$	$\frac{E_{\mathrm{b}}}{N_0}$ at vanishing SNR
E_s	Average signal energy
$E(R)$	Reliability function at rate R
$E_0^{\mathrm{bicm}}(\rho, s)$	BICM random coding exponent
$E_0^{\mathrm{cm}}(\rho)$	CM random coding exponent
$E_0^q(\rho, s)$	Generalized Gallager function
$E_r^q(R)$	Random coding exponent with mismatched decoding
$\mathsf{exit}_{\mathrm{dec}}(\mathsf{y})$	Extrinsic information at decoder

$\mathsf{exit}_{\mathrm{dem}}(\mathsf{x})$	Extrinsic information at demapper
h_k	Fading realization at time k
$I(X;Y)$	Mutual information between variables X and Y
$I^{\mathrm{cm}}(X;Y)$	Coded modulation capacity ($\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}$)
$I^{\mathrm{gmi}}(X;Y)$	Generalized mutual information
$I_s^{\mathrm{gmi}}(X;Y)$	Generalized mutual information (function of s)
$I^{\mathrm{ind}}(X;Y)$	BICM capacity with independent-channel model
K	Number of bits per codeword $\log_2 \lvert\mathcal{M}\rvert$
$\kappa(s)$	Cumulant transform
$\kappa''(s)$	Second derivative of cumulant transform
$\kappa_1(s)$	Cumulant transform of bit score
$\kappa_v(s)$	Cumulant transform of symbol score with weight v
$\kappa_{\mathrm{pw}}(s)$	Cumulant transform of pairwise score
$\kappa_{\mathrm{pw}}(s,\rho_N)$	Cumulant transform of pairwise score for pattern ρ_N
M	Input set (constellation) $\mathcal{X}$ cardinality
m	Number of bits per modulation symbol
m_f	Nakagami fading parameter
μ	Labeling (mapping) rule
$\mathcal{M}$	Message set
m	Message
$\hat{\mathsf{m}}$	Message estimate
$\mathsf{mmse}(\mathsf{snr})$	MMSE of estimating input X (Gaussian channel)
N	Number of channel uses
N_0	Noise spectral density (one-sided)
$\mathcal{N}(\cdot)$	Neighborhood around a node (in factor graph)
$\nu^{f\to\vartheta}$	Function-to-variable message
$\nu^{\vartheta\to f}$	Variable-to-function message
$\mathrm{O}\big(f(x)\big)$	Term vanishing as least as fast as $af(x)$, for $a>0$
$\mathrm{o}\big(f(x)\big)$	Term vanishing faster than $af(x)$, for $a>0$
P	Signal power
P_b	Average probability of bit error
P_e	Average probability of message error
$P_j(y\vert b)$	Transition probability of output y for j-th bit
$P_{\boldsymbol{j}}(y\vert\boldsymbol{b})$	Output transition probability for bits $\boldsymbol{b}$ at $\boldsymbol{j}$ positions
$P_{B_j\vert Y}(b\vert y)$	j-th a posteriori marginal

$P.\cdot$	Probability distribution
$P_{Y\|X}(y\|x)$	Channel transition probability (symbol)
$P_{\boldsymbol{Y}\|\boldsymbol{X}}(\boldsymbol{y}\|\boldsymbol{x})$	Channel transition probability (sequence)
$\mathrm{PEP}(d)$	Pairwise error probability
$\mathrm{PEP}_1(d)$	Pairwise error probability (infinite interleaving)
$\mathrm{PEP}(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{x}_{\mathsf{m}})$	Pairwise error probability
π_n	Interleaver of size n
$\Pr^{\mathrm{dec}\to\mathrm{dem}}(b)$	Bit probability (from the decoder)
$\Pr^{\mathrm{dem}\to\mathrm{dec}}(b)$	Bit probability (from the demapper)
$Q(\cdot)$	Gaussian tail function
$q(x, y)$	Symbol decoding metric
$q(\boldsymbol{x}, \boldsymbol{y})$	Codeword decoding metric
$q_j(b, y)$	Bit decoding metric of j-th bit
R	Code rate, $R = \log_2 \lvert\mathcal{M}\rvert / N$
r	Binary code rate, $r = \log_2 \lvert\mathcal{C}\rvert / n = R/m$
R_0	Cutoff rate
R_0^{av}	Cutoff rate for average-channel model
R_0^{ind}	Cutoff rate for independent-channel model
R_0^q	Generalized cutoff rate (mismatched decoding)
ρ_N	Bit distribution pattern over codeword
$\sum_{\sim x}$	Summary operator —excluding x—
$\hat{s}$	Saddlepoint value
σ_X^2	Variance
$\hat{\sigma}_X^2$	Pseudo-variance, $\sigma_X^2 \triangleq \mathbb{E}\left[\lvert X\rvert^2\right] - \lvert\mathbb{E}[X]\rvert^2$
ζ_0	Wideband slope
snr	Signal-to-noise ratio
W	Signal bandwidth
$\mathcal{X}$	Input signal set (constellation)
$\mathcal{X}_b^j$	Set of symbols with bit b at j-th label
$\mathcal{X}_{b_{i_1},\ldots,b_{i_v}}^{j_{i_1},\ldots,j_{i_v}}$	Set of symbols with bits $b_{i_1}, \ldots, b_{i_v}$ at positions $j_{i_1}, \ldots, j_{i_v}$
x_k	Channel input at time k
$\boldsymbol{x}$	Vector of all channel inputs; input codeword
$\boldsymbol{x}_{\mathsf{m}}$	Codeword corresponding to message m
$\Xi_{m(k-1)+j}$	Bit log-likelihood of j-th bit in k-th symbol
y_k	Channel output at time k

$\boldsymbol{y}$	Vector of all channel outputs
$\mathcal{Y}$	Output signal set
z_k	Noise realization at time k
Ξ^{pw}	Pairwise score
Ξ_k^{s}	Symbol score for k-th symbol
$\Xi_{k,j}^{\mathrm{b}}$	Bit score at j-th label of k-th symbol
Ξ_1^{b}	Symbols score with weight 1 (bit score)
$\Xi_{m(k-1)+j}^{\mathrm{dec}\to\mathrm{dem}}$	Decoder LLR for j-th bit of k-th symbol
$\boldsymbol{\Xi}^{\mathrm{dec}\to\mathrm{dem}}$	Decoder LLR vector
$\boldsymbol{\Xi}_{\sim i}^{\mathrm{dec}\to\mathrm{dem}}$	Decoder LLR vector, excluding the i-th component
$\Xi_{m(k-1)+j}^{\mathrm{dem}\to\mathrm{dec}}$	Demodulator LLR for j-th bit of k-th symbol
$\boldsymbol{\Xi}^{\mathrm{dem}\to\mathrm{dec}}$	Demodulator LLR vector
$\boldsymbol{\Xi}_{\sim i}^{\mathrm{dem}\to\mathrm{dec}}$	Demodulator LLR vector, excluding the i-th component

1

Introduction

Since Shannon's landmark 1948 paper [105], approaching the capacity of the Additive White Gaussian Noise (AWGN) channel has been one of the more relevant topics in information theory and coding theory. Shannon's promise that rates up to the channel capacity can be reliably transmitted over the channel comes together with the design challenge of effectively constructing coding schemes achieving these rates with limited encoding and decoding complexity.

The complex baseband equivalent model of a bandlimited AWGN channel is given by

$$y_k = \sqrt{\mathsf{snr}}\, x_k + z_k, \tag{1.1}$$

where y_k, x_k, z_k are complex random variables and snr denotes the Signal-to-Noise Ratio (SNR), defined as the signal power over the noise power. The capacity C (in nats per channel use) of the AWGN channel with signal-to-noise ratio snr is given by the well-known

$$\mathsf{C} = \log(1 + \mathsf{snr}). \tag{1.2}$$

The coding theorem shows the existence of sufficiently long codes achieving error probability not larger than any $\epsilon > 0$, as long as the coding rate is not larger than C. The standard achievability proof of

(1.2) considers a random coding ensemble generated with i.i.d. components according to a Gaussian probability distribution.

Using a Gaussian code is impractical, as decoding would require an exhaustive search over the whole codebook for the most likely candidate. Instead, typical signaling constellations like Phase-Shift Keying (PSK) or Quadrature-Amplitude Modulation (QAM) are formed by a finite number of points in the complex plane. In order to keep the modulator simple, the set of elementary waveforms that the modulator can generate is a finite set, preferably with small cardinality. A practical way of constructing codes for the Gaussian channel consists of fixing the modulator signal set, and then considering codewords obtained as sequences over the fixed modulator signal set, or *alphabet*. These *coded modulation* schemes are designed for the equivalent channel resulting from the concatenation of the modulator with the underlying waveform channel. The design aims at endowing the coding scheme with just enough structure such that efficient encoding and decoding is possible while, at the same time, having a sufficiently large space of possible codes so that good codes can be found.

Driven by Massey's consideration on coding and modulation as a single entity [79], Ungerböck in 1982 proposed Trellis-Coded Modulation (TCM), based on the combination of trellis codes and discrete signal constellations through set partitioning [130] (see also [15]). TCM enables the use of the efficient Viterbi algorithm for optimal decoding [138] (see also [35]). An alternative scheme is multilevel coded modulation (MLC), proposed by Imai and Hirakawa in 1977 [56] (see also [140]). MLC uses several binary codes, each protecting a single bit of the binary label of modulation symbols. At the receiver, instead of optimal joint decoding of all the component binary codes, a suboptimal multi-stage decoding, alternatively termed successive interference cancellation, achieves good performance with limited complexity. Although not necessarily optimal in terms of minimizing the error probability, the multi-stage decoder achieves the channel capacity [140].

The discovery of turbo codes [11] and the re-discovery of low-density parity-check (LDPC) codes [38, 69] with their corresponding iterative decoding algorithms marked a new era in Coding Theory. These *modern*

codes [96] approach the capacity of binary-input channels with low complexity. The analysis of iterative decoding also led to new methods for their efficient design [96]. At this point, a natural development of coded modulation would have been the extension of these powerful codes to non-binary alphabets. However, iterative decoding of binary codes is by far simpler.

In contrast to Ungerböck's findings, Zehavi proposed bit-interleaved coded modulation (BICM) as a pragmatic approach to coded modulation. BICM separates the actual coding from the modulation through an interleaving permutation [142]. In order to limit the loss of information arising in this separated approach, soft information about the coded bits is propagated from the demodulator to the decoder in the form of bit-wise a posteriori probabilities or log-likelihood ratios. Zehavi illustrated the performance advantages of separating coding and modulation. Later, Caire *et al.* provided in [29] a comprehensive analysis of BICM in terms of information rates and error probability, showing that in fact the loss incurred by the BICM interface may be very small. Furthermore, this loss can essentially be recovered by using iterative decoding. Building upon this principle, Li and Ritcey [64] and ten Brink [122] proposed iterative demodulation for BICM, and illustrated significant performance gains with respect to classical non-iterative BICM decoding [29, 142] when certain binary mappings and convolutional codes are employed. However, BICM designs based on convolutional codes and iterative decoding cannot approach the coded modulation capacity, unless the number of states grows large [139]. Improved constructions based on iterative decoding and on the use of powerful families of modern codes can, however, approach the channel capacity for a particular signal constellation [120, 121, 127].

Since its introduction, BICM has been regarded as a pragmatic yet powerful scheme to achieve high data rates with general signal constellations. Nowadays, BICM is employed in a wide range of practical communications systems, such as DVB-S2, Wireless LANs, DSL, WiMax, the future generation of high data rate cellular systems (the so-called 4th generation). BICM has become the *de-facto* standard for coding over the Gaussian channel in modern systems.

In this monograph, we provide a comprehensive study of BICM. In particular, we review its information theoretic foundations, and review its capacity, cutoff rate and error exponents. Our treatment also covers the wideband regime. We further examine the error probability of BICM, and we focus on the union bound and improved bounds to the error probability. We then turn our attention to iterative decoding of BICM; we also review the underlying design techniques and introduce improved BICM schemes in a unified framework. Finally, we describe a number of applications of BICM not explicitly covered in our treatment. In particular, we consider the application of BICM to orthogonal modulation with non-coherent detection, to the block-fading channel, to the multiple-antenna channel as well as to less common channels such as the exponential-noise or discrete-time Poisson channels.

2

Channel Model and Code Ensembles

This chapter provides the reference background for the remainder of the monograph, as we review the basics of coded modulation schemes and their design options. We also introduce the notation and describe the Gaussian channel model used throughout this monograph. Chapter 6 briefly describes different channels and modulations not explicitly covered by the Gaussian channel model.

2.1 Channel Model: Encoding and Decoding

Consider a memoryless channel with input x_k and output y_k, respectively drawn from the alphabets $\mathcal{X}$ and $\mathcal{Y}$. Let N denote the number of channel uses, i. e. $k = 1, \ldots, N$. A block code $\mathcal{M} \subseteq \mathcal{X}^N$ of length N is a set of $|\mathcal{M}|$ vectors $\boldsymbol{x} = (x_1, \ldots, x_N) \in \mathcal{X}^N$, called codewords. The channel output is denoted by $\boldsymbol{y} \stackrel{\Delta}{=} (y_1, \ldots, y_N)$, with $y_k \in \mathcal{Y}$.

We consider memoryless channels, for which the channel transition probability $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ admits the decomposition

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{k=1}^{N} P_{Y|X}(y_k|x_k), \tag{2.1}$$

With no loss of generality, we limit our attention to continuous output and identify $P_{Y|X}(y|x)$ as a probability density function. We denote by X, Y the underlying random variables. Similarly, the corresponding random vectors are

$$\boldsymbol{X} \stackrel{\Delta}{=} (X_1, \ldots, X_N) \quad \text{and} \quad \boldsymbol{Y} \stackrel{\Delta}{=} (Y_1, \ldots, Y_N), \tag{2.2}$$

respectively drawn from the sets $\mathcal{X}^N$ and $\mathcal{Y}^N$.

At the transmitter, a message m drawn with uniform probability from a message set is mapped onto a codeword $\boldsymbol{x}_{\mathsf{m}}$, according to the encoding schemes described in Sections 2.2 and 2.3. We denote this encoding function by ϕ, i. e. $\phi(\mathsf{m}) = \boldsymbol{x}_{\mathsf{m}}$. Often, and unless strictly necessary, we drop the subindex m in the codeword $\boldsymbol{x}_{\mathsf{m}}$ and simply write $\boldsymbol{x}$. Whenever $|\mathcal{X}| < \infty$, we respectively denote the cardinality of $\mathcal{X}$ and the number of bits required to index a symbol by M and m,

$$M \stackrel{\Delta}{=} |\mathcal{X}|, \quad m \stackrel{\Delta}{=} \log_2 M. \tag{2.3}$$

The decoder outputs an estimate of the message $\widehat{\mathsf{m}}$ according to a given codeword *decoding metric*, denoted by $q(\boldsymbol{x}, \boldsymbol{y})$, so that

$$\widehat{\mathsf{m}} = \varphi(\boldsymbol{y}) = \underset{\mathsf{m} \in \{1, \ldots, |\mathcal{M}|\}}{\arg\max} \; q(\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y}). \tag{2.4}$$

The decoding metrics considered in this work are given as products of symbol decoding metrics $q(x, y)$, for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, namely

$$q(\boldsymbol{x}, \boldsymbol{y}) = \prod_{k=1}^{N} q(x_k, y_k). \tag{2.5}$$

For equally likely codewords, this decoder finds the most likely codeword as long as the metric $q(x, y)$ is a bijective (thus strictly increasing) function of the transition probability $P_{Y|X}(y|x)$ of the memoryless channel. Instead, if the decoding metric $q(x, y)$ is not a bijective function of the channel transition probability, we have a *mismatched decoder* [41, 59, 84].

2.1.1 Gaussian Channel Model

A particularly interesting, yet simple, case is that of complex-plane signal sets ($\mathcal{X} \subset \mathbb{C}$, $\mathcal{Y} = \mathbb{C}$) in AWGN with fully-interleaved fading,

$$y_k = h_k \sqrt{\mathsf{snr}}\, x_k + z_k, \quad k = 1, \ldots, N \tag{2.6}$$

where h_k are fading coefficients with unit variance, z_k are the zero-mean, unit-variance, circularly symmetric complex Gaussian samples, and snr is the signal-to-noise ratio (SNR). In Appendix 2.A we relate this discrete-time model to an underlying continuous-time model with additive white Gaussian noise. We denote the fading and noise random variables by H and Z, with respective probability density functions $P_H(h)$ and $P_Z(z)$. Examples of input set $\mathcal{X}$ are unit energy PSK or QAM signal sets.[1]

With perfect channel state information (coherent detection), the channel coefficient h_k is part of the output, i. e. it is given to the receiver. From the decoder viewpoint, the channel transition probability is decomposed as $P_{Y,H|X}(y,h|x) = P_{Y|X,H}(y|x,h)P_H(h)$, with

$$P_{Y|X,H}(y|x,h) = \frac{1}{\pi}\mathrm{e}^{-|y-h\sqrt{\mathsf{snr}}\ x|^2}. \tag{2.7}$$

Under this assumption, the phase of the fading coefficient becomes irrelevant and we can assume that the fading coefficients are real-valued. In our simulations, we will consider Nakagami-m_f fading, with density

$$P_H(h) = \frac{2m_f^{m_f}h^{2m_f-1}}{\Gamma(m_f)}e^{-m_f h^2}. \tag{2.8}$$

Here $\Gamma(x)$ is Euler's Gamma function, $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$, and $m_f > 0$. In this fading model, we recover the AWGN ($h = 1$) with $m_f \to +\infty$, the Rayleigh fading by letting $m_f = 1$ and the Rician fading with parameter $\mathcal{K}$ by setting $m_f = (\mathcal{K}+1)^2/(2\mathcal{K}+1)$.

Other cases are possible. For example, h_k may be unknown to the receiver (non-coherent detection), or only partially known, i. e. the receiver knows $\widehat{h}_k$ such that $(H, \widehat{H})$ are jointly distributed random variables. In this case, (2.7) generalizes to

$$P_{Y|X,\widehat{H}}(y|x,\widehat{h}) = \mathbb{E}\left[\frac{1}{\pi}\mathrm{e}^{-|y-H\sqrt{\mathsf{snr}}\ x|^2}\middle|\widehat{H} = \widehat{h}\right] \tag{2.9}$$

[1] We consider only one-dimensional complex signal constellations $\mathcal{X} \subset \mathbb{C}$, such as QAM or PSK signal sets (alternatively referred to as two-dimensional signal constellations in the real domain). The generalization to "multidimensional" signal constellations $\mathcal{X} \subset \mathbb{C}^{N'}$, for $N' > 1$, follows immediately, as briefly reviewed in Chapter 6.

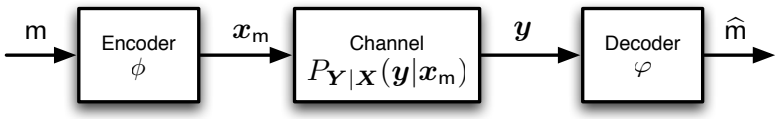


Fig. 2.1 Channel model with encoding and decoding functions.

The classical non-coherent channel where $h = e^{j\theta}$, with θ denoting a uniformly distributed random phase, is a special case of (2.9) [16, 95].

For simplicity of notation, we shall denote the channel transition probability simply as $P_{Y|X}(y|x)$, where the possible conditioning with respect to h or any other related channel state information $\widehat{h}$, is implicitly understood and will be clear from the context.

2.2 Coded Modulation

In a *coded modulation* (CM) scheme, the elements $x_k \in \mathcal{X}$ of the codeword $\boldsymbol{x}_{\mathsf{m}}$ are in general non-binary. At the receiver, a maximum metric decoder φ (as in Eq. (2.4)) generates an estimate of the transmitted message, $\varphi(\boldsymbol{y}) = \widehat{\mathsf{m}}$. The block diagram of a coded modulation scheme is illustrated in Figure 2.1.

The rate R of this scheme in bits per channel use is given by $R = \frac{K}{N}$, where $K \triangleq \log_2 |\mathcal{M}|$ denotes the number of bits per information message. We define the average probability of a *message* error as

$$P_e \triangleq \frac{1}{|\mathcal{M}|} \sum_{\mathsf{m}=1}^{|\mathcal{M}|} P_e(\mathsf{m}) \tag{2.10}$$

where $P_e(\mathsf{m})$ is the conditional error probability when message m was transmitted. We also define the probability of *bit* error as

$$P_b \triangleq \frac{1}{K|\mathcal{M}|} \sum_{k=1}^{K} \sum_{\mathsf{m}=1}^{|\mathcal{M}|} P_e(k, \mathsf{m}) \tag{2.11}$$

where $P_e(k, \mathsf{m}) \triangleq \Pr\{k\text{-th bit in error} \mid \text{message } \mathsf{m} \text{ was transmitted}\}$ is the conditional bit error probability when message m was transmitted.

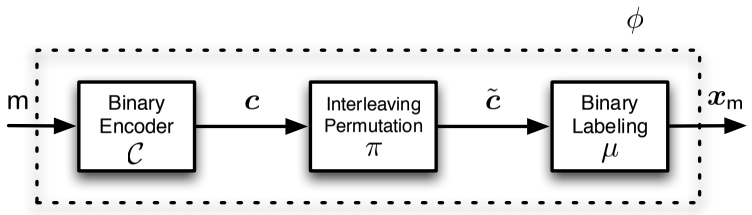


Fig. 2.2 BICM encoder model.

2.3 Bit-Interleaved Coded Modulation

2.3.1 BICM Encoder and Decoders

In a *bit-interleaved coded modulation* scheme, the encoder is restricted to be the serial concatenation of a binary code $\mathcal{C}$ of length $n \triangleq mN$ and rate $r = \frac{\log_2 |\mathcal{C}|}{n} = \frac{R}{m}$, a bit interleaver, and a binary labeling function $\mu : \{0,1\}^m \to \mathcal{X}$ which maps blocks of m bits to signal constellation symbols. The codewords of $\mathcal{C}$ are denoted by $\boldsymbol{c}$. The block diagram of the BICM encoding function is shown in Figure 2.2.

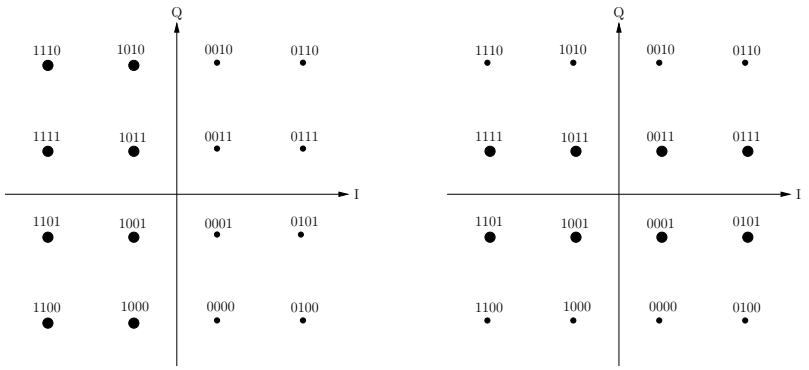
We denote the inverse mapping function for labeling position j as $b_j : \mathcal{X} \to \{0,1\}$, that is, $b_j(x)$ is the j-th bit of symbol x. Accordingly, we now define the sets

$$\mathcal{X}_b^j \triangleq \{x \in \mathcal{X} : b_j(x) = b\} \tag{2.12}$$

as the set of signal constellation points x whose binary label has value $b \in \{0,1\}$ in its j-th position. More generally, we define the sets $\mathcal{X}_{b_{i_1},\ldots,b_{i_v}}^{j_{i_1},\ldots,j_{i_v}}$ as the sets of constellation points having the v binary labels $b_{i_1},\ldots,b_{i_v}$ in positions $j_{i_1},\ldots,j_{i_v}$,

$$\mathcal{X}_{b_{i_1},\ldots,b_{i_v}}^{j_{i_1},\ldots,j_{i_v}} \triangleq \{x \in \mathcal{X} : b_{j_{i_1}}(x) = b_{i_1},\ldots,b_{j_{i_v}}(x) = b_{i_v}\}. \tag{2.13}$$

For future reference, we define the random variables B, X_b^j as a random variables taking values on $\{0,1\}$ or $\mathcal{X}_b^j$ with uniform probability, respectively. The bit $\bar{b} = b \oplus 1$ denotes the binary complement of b. The above sets prove key to analyze the BICM system performance. For reference, Figure 2.3 depicts the sets $\mathcal{X}_b^1$ and $\mathcal{X}_b^4$ for a 16-QAM signal constellation with the Gray labeling described in Section 2.3.3.



(a) $\mathcal{X}_0^1$ and $\mathcal{X}_1^1$. (b) $\mathcal{X}_0^4$ and $\mathcal{X}_1^4$.

Fig. 2.3 Binary labeling sets $\mathcal{X}_b^1$ and $\mathcal{X}_b^4$ for 16-QAM with Gray mapping. Thin dots correspond to points in $\mathcal{X}_0^i$ while thick dots correspond to points in $\mathcal{X}_1^i$.

The classical BICM decoder proposed by Zehavi [142] treats each of the m bits in a symbol as independent and uses a symbol decoding metric proportional to the product of the a posteriori marginals $P_{B_j|Y}(b|y)$. More specifically, we have the (mismatched) symbol metric

$$q(x,y) = \prod_{j=1}^{m} q_j\big(b_j(x), y\big), \tag{2.14}$$

where the j-th bit decoding metric $q_j(b,y)$ is given by

$$q_j\big(b_j(x)=b, y\big) = \sum_{x' \in \mathcal{X}_b^j} P_{Y|X}(y|x'). \tag{2.15}$$

We will refer to this metric as BICM Maximum A Posteriori (MAP) metric. This metric is proportional to the transition probability of the output y given the bit b at position j, which we denote for later use by $P_j(y|b)$,

$$P_j(y|b) \stackrel{\Delta}{=} P_{Y|B_j}(y|b) = \frac{1}{\big|\mathcal{X}_b^j\big|} \sum_{x' \in \mathcal{X}_b^j} P_{Y|X}(y|x'). \tag{2.16}$$

In practice, due to complexity limitations, one might be interested in the following lower-complexity version of (2.15),

$$q_j(b,y) = \max_{x \in \mathcal{X}_b^j} P_{Y|X}(y|x). \tag{2.17}$$

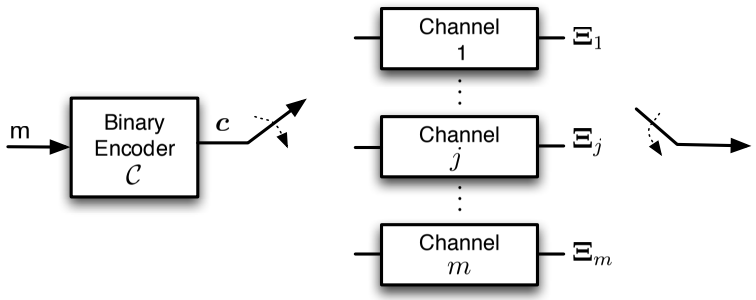


Fig. 2.4 Parallel channel model of BICM.

In the log-domain this is known as the max-log approximation. Either of the symbol metrics corresponding to Eq. (2.14) or Eq. (2.17) are mismatched and do not perform maximum likelihood decoding. Summarizing, the decoder of $\mathcal{C}$ uses a metric of the form given in Eq. (2.5) and outputs a binary codeword $\hat{\boldsymbol{c}}$ according to

$$\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}\in\mathcal{C}} \prod_{k=1}^{N} \prod_{j=1}^{m} q_j\big(b_j(x_k), y_n\big). \tag{2.18}$$

2.3.2 BICM Classical Model

The m probabilities $P_j(y|b)$ were used by Caire *et al.* [29] as starting point to define an equivalent BICM channel model. This equivalent BICM channel is the set of m parallel channels having bit $b_j(x_k)$ as input and the bit log-metric (log-likelihood) ratio for the k-th symbol

$$\Xi_{m(k-1)+j} = \log \frac{q_j\big(b_j(x_k)=1, y\big)}{q_j\big(b_j(x_k)=0, y\big)} \tag{2.19}$$

as output, for $j = 1, \ldots, m$ and $k = 1, \ldots, N$. We define the log-metric ratio vectors for each label bit as $\boldsymbol{\Xi}_j = (\Xi_j, \ldots, \Xi_{m(k-1)+j})$ for $j = 1, \ldots, m$. This channel model is schematically depicted in Figure 2.4.

With infinite-length interleaving, the m parallel channels were assumed to be independent in [29,140], or in other words, the correlations among the different subchannels are neglected. We will see later that

this "classical" representation of BICM as a set of parallel channels gives a good model, even though it can sometimes be optimistic. The alternative model which uses the symbol mismatched decoding metric achieves a higher accuracy at a comparable modeling complexity.

2.3.3 Labeling Rules

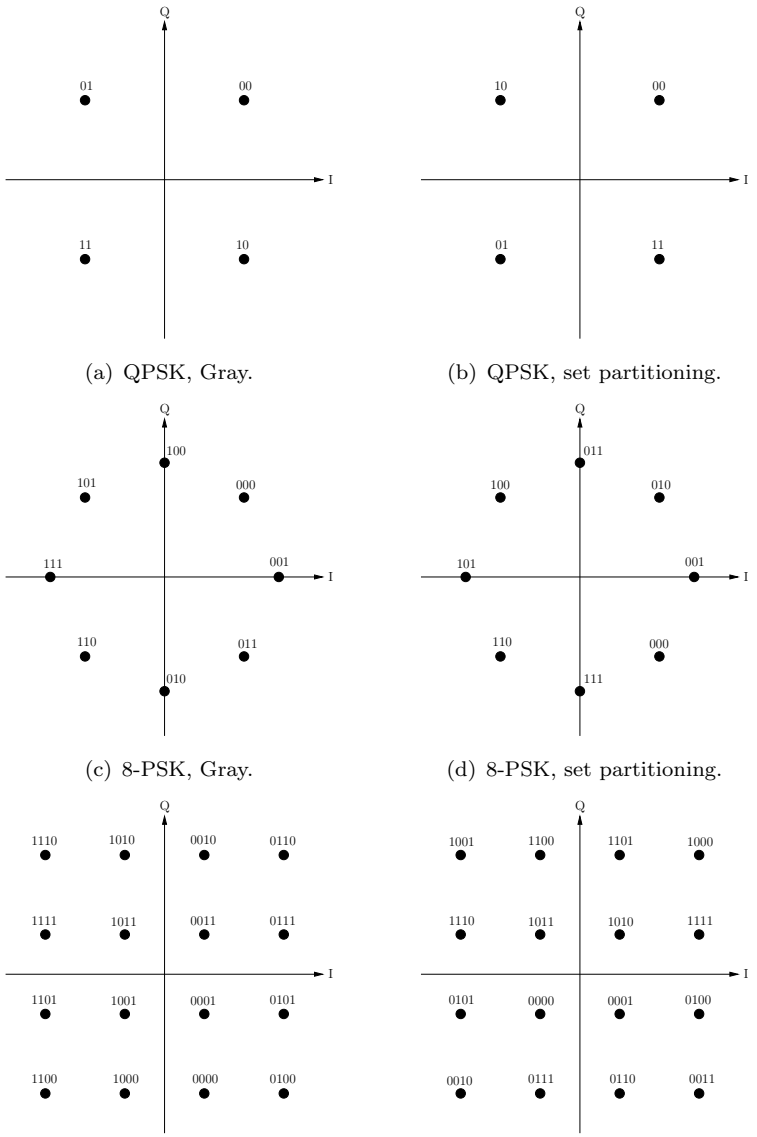
As evidenced in the results of [29], the choice of binary labeling is critical to the performance of BICM. For the decoder presented in previous sections, it was conjectured [29] that binary reflected Gray mapping was optimum, in the sense of having the largest BICM capacity. This conjecture was supported by some numerical evidence, and was further refined in [2, 109] to possibly hold only for moderate-to-large values of SNR. Indeed, Stierstorfer and Fischer [110] have shown that a different labeling —strictly regular set partitioning— is significantly better for small values of SNR. A detailed discussion on the different merits of the various forms of Gray labeling can be found in [2].

Throughout the monograph, we use for our simulations the labeling rules depicted in Figure 2.5, namely binary reflected Gray labeling [95] and set partitioning labeling [130]. Recall that the binary reflected Gray mapping for m bits may be generated recursively from the mapping for $m-1$ bits by prefixing a binary 0 to the mapping for $m-1$ bits, then prefixing a binary 1 to the reflected (i. e. listed in reverse order) mapping for $m-1$ bits. For QAM modulations, the symbol mapping is the Cartesian product of Gray mappings over the in-phase and quadrature components. For PSK modulations, the mapping table is wrapped so that the first and last symbols are contiguous.

2.A Continuous- and Discrete-Time Gaussian Channels

We follow closely the review paper by Forney and Ungerböck [36]. In the linear Gaussian channel, the input $x(t)$, additive Gaussian noise component $z(t)$, and output $y(t)$ are related as

$$y(t) = \int h(t;\tau)x(\tau - t)\,\mathrm{d}\tau + z(t), \tag{2.20}$$



(a) QPSK, Gray. (b) QPSK, set partitioning.

(c) 8-PSK, Gray. (d) 8-PSK, set partitioning.

(e) 16-QAM, Gray. (f) 16-QAM, set partitioning.

Fig. 2.5 Binary labeling rules (Gray, set partitioning) for QPSK, 8-PSK and 16-QAM.

where $h(t;\tau)$ is a (possibly time-varying) channel impulse response. Since all functions are real, their Fourier transforms are Hermitian and

we need consider only the positive-frequency components.

We constrain the signal $x(t)$ to have power P and a frequency content concentrated in an interval $(f_{\min}, f_{\max})$, with the bandwidth W given by $W = f_{\max} - f_{\min}$. Additive noise is assumed white in the frequency band of interest, i. e. noise has a flat power spectral density N_0 (one-sided). If the channel impulse response is constant with unit energy in $(f_{\min}, f_{\max})$, we define the signal-to-noise ratio snr as

$$\mathsf{snr} = \frac{P}{N_0 W}. \tag{2.21}$$

In this case, it is also possible to represent the received signal by its projections onto an orthonormal set,

$$y_k = x_k + z_k, \tag{2.22}$$

where there are only WT effective discrete-time components when the transmission time lasts T seconds ($T \gg 1$) [39]. The signal components x_k have average energy E_s, which is related to the power constraint as $P = E_s W$. The quantities z_k are circularly-symmetric complex Gaussian random variables of variance $\sigma_Z^2 = N_0$. We thus have $\mathsf{snr} = E_s/\sigma_Z^2$. Observe that we recover the model in Eq. (2.6) (with $h_k = 1$) by dividing all quantities in Eq. (2.22) by σ_Z and incorporating this coefficient in the definition of the channel variables.

Another channel of interest we use in the monograph is the frequency-nonselective, or flat fading channel. Its channel response $h(t;\tau)$ is such that Eq. (2.20) becomes

$$y(t) = h(t)x(t) + z(t). \tag{2.23}$$

The channel makes the signal $x(t)$ fade following a coefficient $h(t)$. Under the additional assumption that the coefficient varies quickly, we recover a channel model similar to Eq. (2.6),

$$y_k = h_k x_k + z_k. \tag{2.24}$$

With the appropriate normalization by σ_Z we obtain the model in Eq. (2.6). Throughout the monograph, we use the Nakagami-m_f fading model in Eq. (2.8), whereby coefficients are statistically independent

from one another for different values of k. The squared fading coefficient $g = |h|^2$ has density

$$P_G(g) = \frac{m_f^{m_f} g^{m_f - 1}}{\Gamma(m_f)} e^{-m_f g}. \tag{2.25}$$

3

Information-Theoretic Foundations

In this chapter, we review the information-theoretic foundations of BICM. As suggested in the previous chapter, BICM can be viewed as a coded modulation scheme with a mismatched decoding metric. We study the achievable information rates of coded modulation systems with a generic decoding metric [41, 59, 84] and determine the so-called generalized mutual information. We also provide a general coding theorem based on Gallager's analysis of the error probability by means of the random coding error exponent [39], thus giving an achievable rate and a lower bound to the random coding error exponent.

We compare these results (in particular, the mutual information, the cutoff rate and the overall error exponent) with those derived from the classical BICM channel model as a set of independent parallel channels [29,140]. Whereas the BICM mutual information coincides for both models, the error exponent of the mismatched-decoding model is always upper bounded by that of coded modulation, a condition which is not verified in the independent parallel-channel model mentioned in Section 2.3. We complement our analysis with a derivation of the error exponents of other variants of coded modulation, namely multi-level coding with successive decoding [140] and with independent decoding

of all the levels. As is well known, the mutual information attained by multi-level constructions can be made equal to that of coded modulation. However, this equality is attained at a non negligible cost in error exponent, as we will see later.

For Gaussian channels with binary reflected Gray labeling, the mutual information and the random coding error exponent of BICM are close to those of coded modulation for medium-to-large signal-to-noise ratios. For low signal-to-noise ratios —or low spectral efficiency— we give a simple analytic expression for the loss in mutual information or received power compared to coded modulation. We determine the minimum energy per bit necessary for reliable communication when BICM is used. For QAM constellations with binary reflected Gray labeling, this energy is at most 1.25 dB from optimum transmission methods. BICM is therefore a suboptimal, yet simple transmission method valid for a large range of signal-to-noise ratios. We also give a simple expression for the first derivative of the BICM mutual information with respect to the signal-to-noise ratio, in terms of Minimum Mean-Equare Error (MMSE) for estimating the input of the channel from its output, and we relate this to the findings of [51, 67].

3.1 Coded Modulation

3.1.1 Channel Capacity

A coding rate[1] R is said achievable if, for all $\epsilon > 0$ and all sufficiently large N there exists codes of length N with rate not smaller than R (i. e. with at least $\lceil e^{RN} \rceil$ messages) and error probability $P_e < \epsilon$ [31]. The capacity C is the supremum of all achievable rates. For memoryless channels, Shannon's theorem yields the capacity formula:

Theorem 3.1 (Shannon 1948). The channel capacity C is given by

$$\mathsf{C} = \sup_{P_X(\cdot)} I(X;Y), \tag{3.1}$$

where $I(X;Y)$ denotes the mutual information of between X and Y,

[1] Capacities and information rates will be expressed using a generic logarithm, typically the natural logarithm. However, all charts in this monograph are expressed in bits.

defined

$$I(X;Y) = \mathbb{E}\left[\log \frac{P_{Y|X}(Y|X)}{P_Y(Y)}\right]. \tag{3.2}$$

For the maximum-likelihood decoders considered in Section 2.1[2] Gallager studied the average error probability of randomly generated codes [39, Chapter 5]. Specifically, he proved that the error probability decreases exponentially with the block length according to a parameter called the *reliability function*. Denoting the error probability attained by a coded modulation scheme $\mathcal{M}$ of length N and rate R by $P_e(\mathcal{M})$, we define the reliability function $E(R)$ as

$$E(R) \stackrel{\Delta}{=} \lim_{N\to\infty} -\frac{1}{N} \log \inf_{\mathcal{M}} P_e(\mathcal{M}), \tag{3.3}$$

where the optimization is carried out over all possible coded modulation schemes $\mathcal{M}$. Since the reliability function is often not known exactly [39], upper and lower bounds to it are given instead. We are specially interested in a lower bound, known as the *random coding error exponent*, which gives an accurate characterization of the average error performance of the ensemble of random codes for sufficiently high rates. Furthermore, this lower bound is known to be tight for rates above a certain threshold, known as the critical rate [39, Chapter 5].

When the only constraint to the system is $\mathbb{E}[|X|^2] \leq 1$, then the channel capacity of the AWGN channel described in (2.6) (letting $H = 1$ with probability 1) is given by [31, 39, 105]

$$\mathsf{C} = \log\left(1 + \mathsf{snr}\right). \tag{3.4}$$

In this case, the capacity given by (3.4) is achieved by Gaussian codebooks [105], i. e. randomly generated codebooks with components independently drawn according to a Gaussian distribution, $X \sim \mathcal{N}_{\mathbb{C}}(0,1)$.

From a practical point of view, it is often more convenient to construct codewords as sequences of points from a signal constellation $\mathcal{X}$ of finite cardinality, such as PSK or QAM [95], with a uniform input

[2] Decoders with a symbol decoding metric $q(x,y)$ that is a bijective (increasing) function of the channel transition probability $P_{Y|X}(y|x)$.

distribution, $P_X(x) = \frac{1}{2^m}$ for all $x \in \mathcal{X}$. While a uniform distribution is only optimal for large snr, it is simpler to implement and usually leads to more manageable analytical expressions. In general, the probability distribution $P_X(x)$ that maximizes the mutual information for a given signal constellation depends on snr and on the specific constellation geometry. Optimization of this distribution has been termed in the literature as signal constellation *shaping* (see for example [34, 36] and references therein). Unless otherwise stated, we will always consider the uniform distribution throughout this monograph.

For a uniform input distribution, we refer to the corresponding mutual information between channel input X and output Y as the *coded modulation capacity*, and denote it by $\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}$ or $I^{\mathrm{cm}}(X;Y)$, that is

$$\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}} = I^{\mathrm{cm}}(X;Y) \triangleq \mathbb{E}\left[\log \frac{P_{Y|X}(Y|X)}{\frac{1}{2^m}\sum_{x' \in \mathcal{X}} P_{Y|X}(Y|x')}\right]. \tag{3.5}$$

Observe that the finite nature of these signal sets implies that they can only convey a finite number of bits per channel use, i. e. $\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}} \leq m$ bits.

Figure 3.1 shows the coded modulation capacity for multiple signal constellations in the AWGN channel, as a function of snr.

3.1.2 Error Probability with Random Codes

Following in the footsteps of Gallager [39, Chapter 5], this section provides an achievability theorem for a general decoding metric $q(x, y)$ using random coding arguments. The final result, concerning the error probability, can be found in Reference [59].

We consider an ensemble of randomly generated codebooks, for which the entries of the codewords $\boldsymbol{x}$ are i. i. d. realizations of a random variable X with probability distribution $P_X(x)$ over the set $\mathcal{X}$, i. e. $P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{k=1}^{N} P_X(x_k)$. We denote by $\bar{P}_e(\mathsf{m})$ the average error probability over the code ensemble when message m is transmitted and by $\bar{P}_e$ the error probability averaged over the message choices. Therefore,

$$\bar{P}_e = \frac{1}{|\mathcal{M}|} \sum_{\mathsf{m}=1}^{|\mathcal{M}|} \bar{P}_e(\mathsf{m}). \tag{3.6}$$

The symmetry of the code construction makes $\bar{P}_e(\mathsf{m})$ independent of

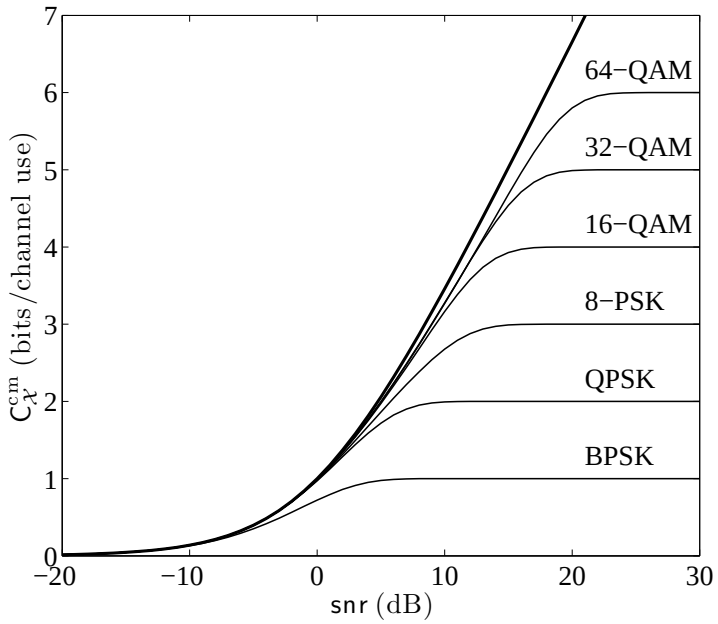


Fig. 3.1 Coded modulation capacity in bits per channel use for multiple signal constellations with uniform inputs in the AWGN channel. For reference, the channel capacity with Gaussian inputs (3.4) is shown in thick lines.

the index m, and hence $\bar{P}_e = \bar{P}_e(\mathsf{m})$ for any $\mathsf{m} \in \mathcal{M}$. Averaged over the random code ensemble, we have that

$$\bar{P}_e(\mathsf{m}) = \sum_{\boldsymbol{x}_\mathsf{m}} P_{\boldsymbol{X}}(\boldsymbol{x}_\mathsf{m}) \int_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_\mathsf{m}) \Pr\{\varphi(\boldsymbol{y}) \neq \mathsf{m}|\boldsymbol{x}_\mathsf{m}, \boldsymbol{y}\} \, \mathrm{d}\boldsymbol{y}, \quad (3.7)$$

where $\Pr\{\varphi(\boldsymbol{y}) \neq \mathsf{m}|\boldsymbol{x}_\mathsf{m}, \boldsymbol{y}\}$ is the probability that, for a channel output $\boldsymbol{y}$, the decoder φ selects a codeword other than the transmitted $\boldsymbol{x}_\mathsf{m}$.

The decoder φ, as defined in (2.4), chooses the codeword $\boldsymbol{x}_{\widehat{\mathsf{m}}}$ with largest metric $q(\boldsymbol{x}_{\widehat{\mathsf{m}}}, \boldsymbol{y})$. The pairwise error probability $\Pr\{\varphi(\boldsymbol{y}) = \mathsf{m}'|\boldsymbol{x}_\mathsf{m}, \boldsymbol{y}\}$ of wrongly selecting message m' when message m has been transmitted and sequence $\boldsymbol{y}$ has been received is given by

$$\Pr\{\varphi(\boldsymbol{y}) = \mathsf{m}'|\boldsymbol{x}_\mathsf{m}, \boldsymbol{y}\} = \sum_{\boldsymbol{x}_{\mathsf{m}'} \,:\, q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y}) \geq q(\boldsymbol{x}_\mathsf{m}, \boldsymbol{y})} P_{\boldsymbol{X}}(\boldsymbol{x}_{\mathsf{m}'}). \quad (3.8)$$

Using the union bound over all possible codewords, and for all $0 \leq \rho \leq 1$, the probability $\Pr\{\varphi(\boldsymbol{y}) \neq \mathsf{m}|\boldsymbol{x}_\mathsf{m}, \boldsymbol{y}\}$ can be bounded by [39, p.

136]

$$\Pr\{\varphi(\boldsymbol{y}) \neq \mathsf{m}|\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y}\} \leq \Pr\left\{\bigcup_{\mathsf{m}'\neq\mathsf{m}} \{\varphi(\boldsymbol{y}) = \mathsf{m}'|\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y}\}\right\} \tag{3.9}$$

$$\leq \left(\sum_{\mathsf{m}'\neq\mathsf{m}} \Pr\{\varphi(\boldsymbol{y}) = \mathsf{m}'|\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y}\}\right)^{\rho}. \tag{3.10}$$

Since $q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y}) \geq q(\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y})$ and the sum over all $\boldsymbol{x}_{\mathsf{m}'}$ upper bounds the sum over the set $\{\boldsymbol{x}_{\mathsf{m}'} : q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y}) \geq q(\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y})\}$, for any $s > 0$, the pairwise error probability in Eq. (3.8) can bounded by

$$\Pr\{\varphi(\boldsymbol{y}) = \mathsf{m}'|\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y}\} \leq \sum_{\boldsymbol{x}_{\mathsf{m}'}} P_{\boldsymbol{X}}(\boldsymbol{x}_{\mathsf{m}'}) \left(\frac{q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y})}{q(\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y})}\right)^{s}. \tag{3.11}$$

As m' is a dummy variable, for any $s > 0$ and $0 \leq \rho \leq 1$ it holds that

$$\Pr\{\varphi(\boldsymbol{y}) \neq \mathsf{m}|\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y}\} \leq \left((|\mathcal{M}| - 1) \sum_{\boldsymbol{x}_{\mathsf{m}'}} P_{\boldsymbol{X}}(\boldsymbol{x}_{\mathsf{m}'}) \left(\frac{q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y})}{q(\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y})}\right)^{s}\right)^{\rho}. \tag{3.12}$$

Therefore, Eq. (3.7) can be written as

$$\bar{P}_e \leq \left(|\mathcal{M}| - 1\right)^{\rho} \mathbb{E}\left[\left(\sum_{\boldsymbol{x}_{\mathsf{m}'}} P_{\boldsymbol{X}}(\boldsymbol{x}_{\mathsf{m}'}) \left(\frac{q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{Y})}{q(\boldsymbol{X}_{\mathsf{m}}, \boldsymbol{Y})}\right)^{s}\right)^{\rho}\right]. \tag{3.13}$$

For memoryless channels, we have a per-letter characterization [39]

$$\bar{P}_e \leq (|\mathcal{M}| - 1)^{\rho} \left(\mathbb{E}\left[\left(\sum_{x'} P_X(x') \left(\frac{q(x', Y)}{q(X, Y)}\right)^{s}\right)^{\rho}\right]\right)^{N}. \tag{3.14}$$

Hence, for any input distribution $P_X(x)$, $0 \leq \rho \leq 1$ and $s > 0$,

$$\bar{P}_e \leq \mathrm{e}^{-N(E_0^q(\rho,s) - \rho R)} \tag{3.15}$$

where

$$E_0^q(\rho, s) \triangleq -\log \mathbb{E}\left[\left(\sum_{x'} P_X(x') \left(\frac{q(x', Y)}{q(X, Y)}\right)^{s}\right)^{\rho}\right] \tag{3.16}$$

is the generalized Gallager function. The expectation is carried out according to the joint distribution $P_{X,Y}(x,y) = P_{Y|X}(y|x)P_X(x)$.

We define the mismatched *random coding exponent* as

$$E_{\mathrm{r}}^q(R) \stackrel{\Delta}{=} \max_{0\le\rho\le 1} \max_{s>0} \big(E_0^q(\rho,s) - \rho R\big). \tag{3.17}$$

This procedure also yields a lower bound on the reliability function, $E(R) \ge E_{\mathrm{r}}^q(R)$. Further improvements are possible by optimizing over the input distribution $P_X(x)$.

According to (3.15), the average error probability $\bar{P}_e$ goes to zero if $E_0^q(\rho,s) > \rho R$ for a given s. In particular, as ρ vanishes, rates below

$$\lim_{\rho\to 0} \frac{E_0^q(\rho,s)}{\rho} \tag{3.18}$$

are achievable. Using that $E_0^q(\rho,s) = 0$ for $\rho = 0$, and in analogy to the mutual information $I(X;Y)$, we define the quantity $I_s^{\mathrm{gmi}}(X;Y)$ as

$$I_s^{\mathrm{gmi}}(X;Y) \stackrel{\Delta}{=} \left.\frac{\partial E_0^q(\rho,s)}{\partial\rho}\right|_{\rho=0} = \lim_{\rho\to 0}\frac{E_0^q(\rho,s)}{\rho} \tag{3.19}$$

$$= -\mathbb{E}\left[\log \sum_{x'} P_X(x') \left(\frac{q(x',Y)}{q(X,Y)}\right)^s\right] \tag{3.20}$$

$$= \mathbb{E}\left[\log \frac{q(X,Y)^s}{\sum_{x'\in\mathcal{X}} P_X(x')q(x',Y)^s}\right]. \tag{3.21}$$

By maximizing over the parameter s we obtain the *generalized mutual information* for a mismatched decoder using metric $q(x,y)$ [41, 59, 84],

$$I^{\mathrm{gmi}}(X;Y) = \max_{s>0} I_s^{\mathrm{gmi}}(X;Y). \tag{3.22}$$

The preceding analysis shows that any rate $R < I^{\mathrm{gmi}}(X;Y)$ is achievable, i. e. we can transmit at rate $R < I^{\mathrm{gmi}}(X;Y)$ and have $\bar{P}_e \to 0$.

For completeness and symmetry with classical random coding analysis, we define the generalized cutoff rate as

$$R_0 \stackrel{\Delta}{=} E_{\mathrm{r}}^q(R=0) = \max_{s>0} E_0^q(1,s). \tag{3.23}$$

For a maximum likelihood decoder, $E_0^q(\rho,s)$ is maximized by letting

$s = \frac{1}{1+\rho}$ [39], and we have

$$E_0(\rho) \stackrel{\Delta}{=} -\log \mathbb{E}\left[\left(\sum_{x'} P_X(x') \left(\frac{P_{Y|X}(Y|x')}{P_{Y|X}(Y|X)}\right)^{\frac{1}{1+\rho}}\right)^{\rho}\right] \tag{3.24}$$

$$= -\log \int_y \left(\sum_x P_X(x) P_{Y|X}(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho} \mathrm{d}y, \tag{3.25}$$

namely the *coded modulation exponent*. For uniform inputs,

$$E_0^{\mathrm{cm}}(\rho) \stackrel{\Delta}{=} -\log \int_y \left(\frac{1}{2^m}\sum_x P_{Y|X}(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho} \mathrm{d}y. \tag{3.26}$$

Incidentally, the argument in this section proves the achievability of the rate $\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}} = I^{\mathrm{cm}}(X;Y)$ with random codes and uniform inputs, i. e. there exist coded modulation schemes with exponentially vanishing error probability for all rates $R < \mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}$.

Later, we will use the following data-processing inequality, which shows that the generalized Gallager function of any mismatched decoder is upperbounded by the Gallager function of a maximum likelihood decoder.

Proposition 3.1 (Data-Processing Inequality [59, 77]). For $s > 0$, $0 \le \rho \le 1$, and a given input distribution we have that

$$E_0^q(\rho, s) \le E_0(\rho) \tag{3.27}$$

The necessary condition for equality to hold is that the metric $q(x, y)$ is proportional to a power of the channel transition probability,

$$P_{Y|X}(y|x) = c' q(x, y)^{s'} \qquad \text{for all } x \in \mathcal{X} \tag{3.28}$$

for some constants c' and s'.

3.2 Bit-Interleaved Coded Modulation

In this section, we study the BICM decoder and determine the generalized mutual information and a lower bound to the reliability function. Special attention is given to the comparison with the classical analysis of BICM as a set of m independent parallel channels (see Section 2.3).

3.2.1 Achievable Rates

We start with a brief review of the classical results on the achievable rates for BICM. Under the assumption of an infinite-length interleaver, capacity and cutoff rate were studied in [29]. This assumption (see Section 2.3) yields a set of m independent parallel binary-input channels, for which the corresponding mutual information and cutoff rate are the sum of the corresponding rates of each subchannel, and are given by

$$I^{\mathrm{ind}}(X;Y) \triangleq \sum_{j=1}^{m} \mathbb{E}\left[\log \frac{\sum_{x'\in\mathcal{X}_B^j} P_{Y|X}(Y|x')}{\frac{1}{2}\sum_{x'\in\mathcal{X}} P_{Y|X}(Y|x')}\right], \tag{3.29}$$

and

$$R_0^{\mathrm{ind}} \triangleq m\log 2 - \sum_{j=1}^{m} \log\left(1+\mathbb{E}\left[\sqrt{\frac{\sum_{x'\in\mathcal{X}_{\bar{B}}^j} P_{Y|X}(Y|x')}{\sum_{x'\in\mathcal{X}_B^j} P_{Y|X}(Y|x')}}\right]\right), \tag{3.30}$$

respectively. An underlying assumption behind Eq. (3.30) is that the m independent channels are used the same number of times. Alternatively, the parallel channels may be used with probability $\frac{1}{m}$, and the cutoff rate is then m times the cutoff rate of an averaged channel [29],

$$R_0^{\mathrm{av}} \triangleq m\left[\log 2 - \log\left(1+\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left[\sqrt{\frac{\sum_{x'\in\mathcal{X}_{\bar{B}_j}^j} P_{Y|X}(Y|x')}{\sum_{x'\in\mathcal{X}_{B_j}^j} P_{Y|X}(Y|x')}}\right]\right)\right]. \tag{3.31}$$

The expectations are over the joint probability $P_{B_j,Y}(b,y) = \frac{1}{2}P_j(y|b)$. From Jensen's inequality one easily obtains that $R_0^{\mathrm{av}} \le R_0^{\mathrm{ind}}$.

We will use the following shorthand notation for the BICM capacity,

$$\mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}} \triangleq I^{\mathrm{ind}}(X;Y). \tag{3.32}$$

The following alternative expression [26, 76, 140] for the BICM mutual information turns out to be useful,

$$\mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}} = \sum_{j=1}^{m}\frac{1}{2}\sum_{b=0}^{1}\left(\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}} - \mathsf{C}_{\mathcal{X}_b^j}^{\mathrm{cm}}\right) \tag{3.33}$$

where $\mathsf{C}_{\mathcal{A}}^{\mathrm{cm}}$ is the mutual information for coded modulation over a general signal constellation $\mathcal{A}$.

We now relate this BICM capacity with the generalized mutual information introduced in the previous section.

Theorem 3.2 ([77]). The generalized mutual information of the BICM decoder is given by the sum of the generalized mutual informations of the independent binary-input parallel channel model of BICM,

$$I^{\mathrm{gmi}}(X;Y) = \sup_{s>0} \sum_{j=1}^{m} \mathbb{E}\left[\log \frac{q_j(b,Y)^s}{\frac{1}{2}\sum_{b'=0}^{1} q_j(b',Y)^s}\right]. \tag{3.34}$$

There are a number of interesting particular cases of the above theorem.

Corollary 3.1 ([77]). For the metric in Eq. (2.15),

$$I^{\mathrm{gmi}}(X;Y) = \mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}}. \tag{3.35}$$

Expression (3.35) coincides with the BICM capacity above, even though we have lifted the assumption of infinite interleaving. When the suboptimal metrics (2.17) are used, we have the following.

Corollary 3.2 ([77]). For the metric in Eq. (2.17),

$$I^{\mathrm{gmi}}(X;Y) = \sup_{s>0} \sum_{j=1}^{m} \mathbb{E}\left[\log \frac{\left(\max_{x\in\mathcal{X}_j^B} p(y|x)\right)^s}{\frac{1}{2}\sum_{b=0}^{1}\left(\max_{x'\in\mathcal{X}_b^j} p(y|x')\right)^s}\right]. \tag{3.36}$$

The information rates achievable with this suboptimal decoder have been studied by Szczecinski *et al.* [112]. The fundamental difference between their result and the generalized mutual information given in (3.36) is the optimization over s. Since both expressions are equal when $s = 1$, the optimization over s may induce a larger achievable rate.

Figure 3.2 shows the BICM mutual information for some signal constellations, different binary labeling rules and uniform inputs for

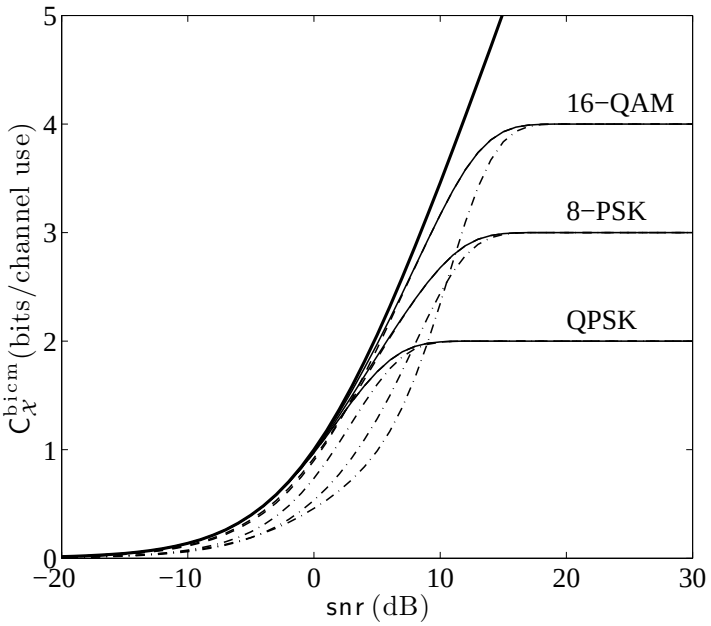


Fig. 3.2 Coded modulation and BICM capacities (in bits per channel use) for multiple signal constellations with uniform inputs in the AWGN channel. Gray and set partitioning labeling rules correspond to dashed and dashed-dotted lines respectively. In thick solid lines, the capacity with Gaussian inputs (3.4); with thin solid lines the CM channel capacity.

the AWGN channel, as a function of snr. For the sake of illustration simplicity, we have only plotted the information rate for the Gray and set partitioning binary labeling rules from Figure 2.5. Observe that binary reflected Gray labeling pays a negligible penalty in information rate, being close to the coded modulation capacity.

3.2.2 Error Exponents

Evaluation of the generalized Gallager function in Eq. (3.16) for BICM with a bit metric $q_j(b, y)$ yields a function $E_0^{\mathrm{bicm}}(\rho, s)$ of ρ and s,

$$E_0^{\mathrm{bicm}}(\rho, s) \stackrel{\Delta}{=} -\log \mathbb{E}\left[\left(\frac{1}{2^m}\sum_{x'\in\mathcal{X}}\prod_{j=1}^{m}\frac{q_j(b_j(x'), Y)^s}{q_j(b_j(X), Y)^s}\right)^{\rho}\right]. \qquad (3.37)$$

Moreover, the data processing inequality for error exponents in Proposition 3.1 shows that the error exponent (and in particular the cutoff rate) of the BICM decoder is upperbounded by the error exponent (and the cutoff rate) of the ML decoder, that is $E_0^{\mathrm{bicm}}(\rho, s) \leq E_0^{\mathrm{cm}}(\rho)$.

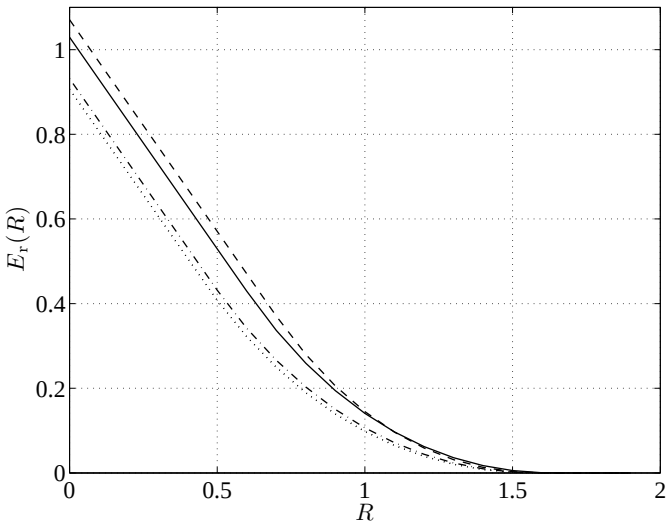
In their analysis of multilevel coding and successive decoding, Wachsmann *et al.* provided the error exponents of BICM modeled as a set of independent parallel channels [140]. The corresponding Gallager's function, which we denote by $E_0^{\mathrm{ind}}(\rho)$, is given by

$$E_0^{\mathrm{ind}}(\rho) \stackrel{\Delta}{=} -\sum_{j=1}^{m} \log \mathbb{E}\left[\left(\sum_{b'=0}^{1} P_{B_j}(b') \frac{P_j(Y|b')^{\frac{1}{1+\rho}}}{P_j(Y|B)^{\frac{1}{1+\rho}}}\right)^{\rho}\right]. \tag{3.38}$$

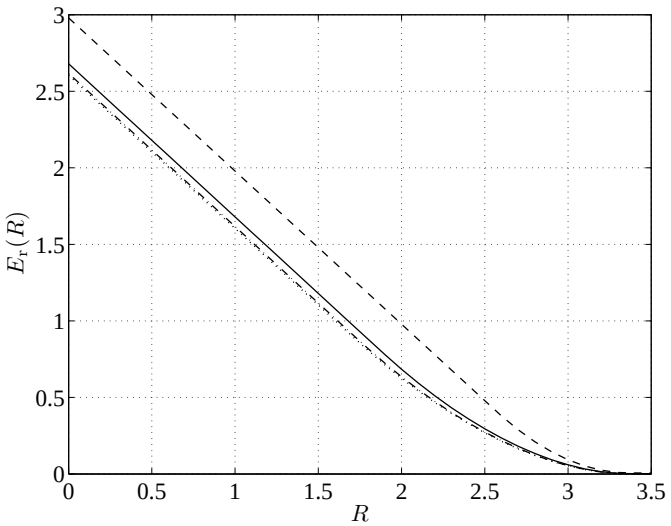
This quantity is the random coding exponent of the BICM decoder if the channel output $\boldsymbol{y}$ admits a decomposition into a set of parallel and independent subchannels. In general, this is not the case, since all subchannels are affected by the same noise —and possibly fading— realization, and the parallel-channel model fails to capture the statistics of the channel.

Figures 3.3(a), 3.3(b) and 3.4 show the error exponents for coded modulation (solid), BICM with independent parallel channels (dashed), BICM using metric (2.15) (dash-dotted), and BICM using metric (2.17) (dotted) for 16-QAM with the Gray labeling in Figure 2.5, Rayleigh fading and $\mathsf{snr} = 5, 15, -25$ dB, respectively. Dotted lines labeled with $s = \frac{1}{1+\rho}$ correspond to the error exponent of BICM using metric (2.17) letting $s = \frac{1}{1+\rho}$. The parallel-channel model gives a larger exponent than the coded modulation, in agreement with the cutoff rate results of [29]. In contrast, the mismatched-decoding analysis yields a lower exponent than coded modulation. As mentioned in the previous section, both BICM models yield the same capacity.

In most cases, BICM with a max-log metric (2.17) incurs a marginal loss in the exponent for mid-to-large SNR. In this SNR range, the optimized exponent and that with $s = \frac{1}{1+\rho}$ are almost equal. For low SNR, the parallel-channel model and the mismatched-metric model with (2.15) have the same exponent, while we observe a larger penalty when metrics (2.17) are used. As we observe, some penalty is incurred at low SNR for not optimizing over s. We denote with crosses the



(a) $\mathsf{snr} = 5$ dB.



(b) $\mathsf{snr} = 15$ dB.

Fig. 3.3 Error exponents for coded modulation (solid), BICM with independent parallel channels (dashed), BICM using metric (2.15) (dash-dotted), and BICM using metric (2.17) (dotted) for 16-QAM with Gray labeling, Rayleigh fading.

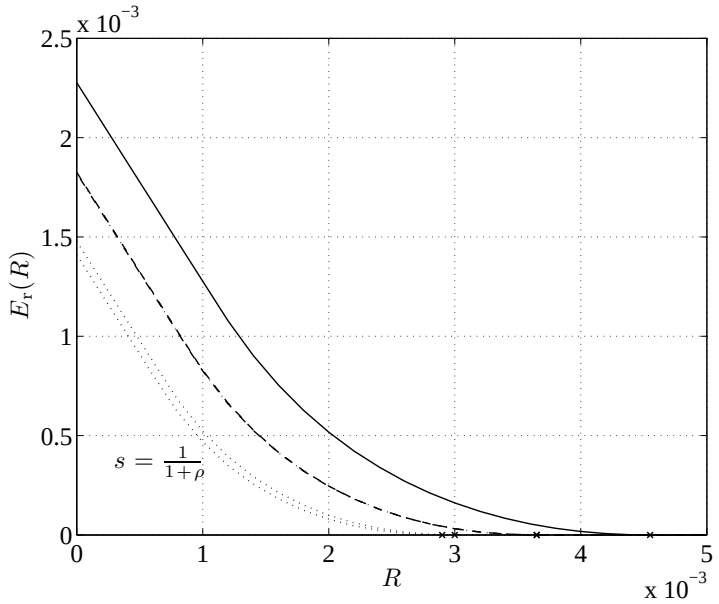


Fig. 3.4 Error exponents for coded modulation (solid), BICM with independent parallel channels (dashed), BICM using metric (2.15) (dash-dotted), and BICM using metric (2.17) (dotted) for 16-QAM with Gray labeling, Rayleigh fading and $\mathsf{snr} = -25$ dB. Crosses correspond to (from right to left) coded modulation, BICM with metric (2.15), BICM with metric (2.17) and BICM with metric (2.17) and $s = 1$.

corresponding achievable information rates.

An interesting question is whether the error exponent of the parallel-channel model is always larger than that of the mismatched-decoding model. The answer is negative, as illustrated in Figure 3.5, which shows the error exponents for coded modulation (solid), BICM with independent parallel channels (dashed), BICM using metric (2.15) (dash-dotted), and BICM using metric (2.17) (dotted) for 8-PSK with Gray labeling in the AWGN channel.

3.3 Comparison with Multilevel Coding

Multilevel codes (MLC) combined with multistage decoding (MSD) have been proposed [56,140] as an efficient method to attain the channel capacity by using binary codes. In this section, we compare BICM with

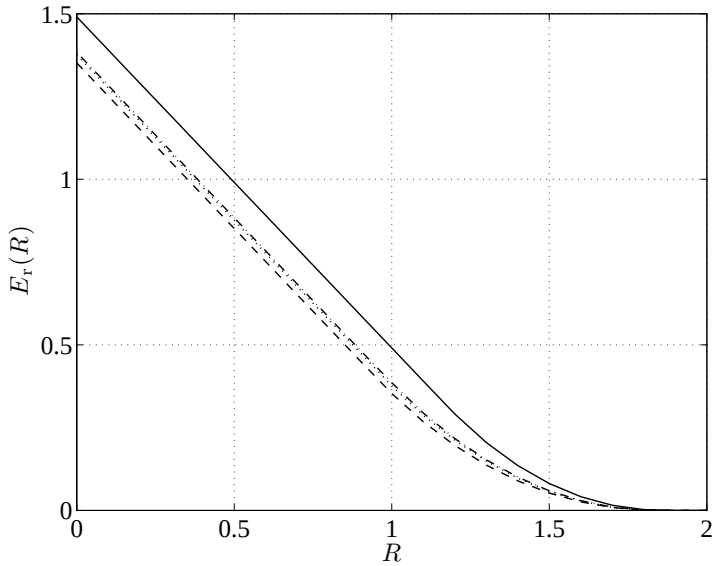


Fig. 3.5 Error exponents for coded modulation (solid), BICM with independent parallel channels (dashed), BICM using metric (2.15) (dash-dotted), and BICM using metric (2.17) (dotted) for 8-PSK with Gray labeling, AWGN and $\mathsf{snr} = 5$ dB.

MLC in terms of error exponents and achievable rates. In particular, we elaborate on the analogy between MLC and the multiple-access channel to present a general error exponent analysis of MLC with MSD. The error exponents of MLC have been studied in a somewhat different way in [12, 13, 140].

For BICM, a single binary code $\mathcal{C}$ is used to generate a binary codeword, which is used to select modulation symbols by a binary labeling function μ. A uniform distribution over the channel input set induces a uniform distribution over the input bits b_j, $j = 1, \ldots, m$. In MLC, the input binary code $\mathcal{C}$ is the Cartesian product of m binary codes of length N, one per modulation level, i. e. $\mathcal{C} = \mathcal{C}_1 \times \ldots \times \mathcal{C}_m$, and the input distribution for the symbol $x(b_1, \ldots, b_j)$ has the form

$$P_X(x) = P_{B_1,\ldots,B_M}(b_1, \ldots, b_m) = \prod_{j=1}^{m} P_{B_j}(b_j). \tag{3.39}$$

Denoting the rate of the j-th level code by R_j, the resulting total rate

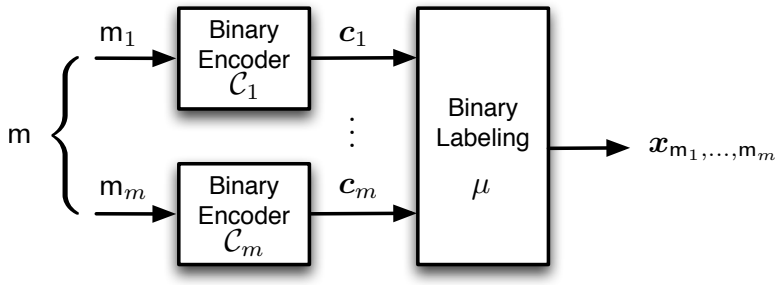


Fig. 3.6 Block diagram of a multi-level encoder.

of the MLC scheme is $R = \sum_{j=1}^{m} R_j$. We denote the codewords of $\mathcal{C}_j$ by $\boldsymbol{c}_j$. The block diagram of MLC is shown in Figure 3.6.

The multi-stage decoder operates by decoding the m levels separately. The symbol decoding metric is thus of the form

$$q(x, y) = \prod_{j=1}^{m} q_j(b_j(x), y). \tag{3.40}$$

A crucial difference with respect to BICM is that the decoders are allowed to pass information from one level to another. Decoding operates sequentially, starting with code $\mathcal{C}_1$, feeding the result to this decoder to $\mathcal{C}_2$, and proceeding across all levels. We have that the j-th decoding metric of MLC with MSD is given by

$$q_j\big(b_j(x) = b, y\big) = \frac{1}{\left|\mathcal{X}^{1,\ldots,j}_{b_1,\ldots,b_{j-1},b}\right|} \sum_{x' \in \mathcal{X}^{1,\ldots,j}_{b_1,\ldots,b_{j-1},b}} P_{Y|X}(y|x'). \tag{3.41}$$

Conditioning on the previously decoded levels reduces the number of symbols remaining at the j-th level, so that only 2^{m-j} symbols remain at the j-th level in Eq. (3.41). Figure 3.7 depicts the operation of an MLC/MSD decoder.

The MLC construction is very similar to that of a multiple-access channel, as noticed by [140]. An extension of Gallager's random coding analysis to the multiple-access channel was carried out by Slepian and Wolf in [108] and Gallager [40], and we now review it for a simple 2-level case. Generalization to a larger number of levels is straightforward.

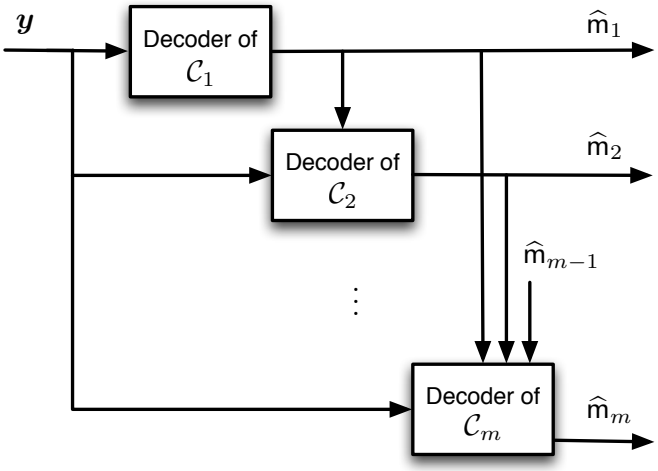


Fig. 3.7 Block diagram of a multi-stage decoder for MLC.

As in our analysis of Section 3.1.2, we denote by $\bar{P}_e$ the error probability averaged over an ensemble of randomly selected codebooks,

$$\bar{P}_e = \frac{1}{|\mathcal{C}_1||\mathcal{C}_2|} \sum_{\mathsf{m}_1=1}^{|\mathcal{C}_1|} \sum_{\mathsf{m}_2=1}^{|\mathcal{C}_2|} \bar{P}_e(\mathsf{m}_1, \mathsf{m}_2), \tag{3.42}$$

where $\bar{P}_e(\mathsf{m}_1, \mathsf{m}_2)$ denotes the average error probability over the code ensemble when messages m_1 and m_2 are chosen by codes $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. Again, the random code construction makes the error probability independent of the transmitted message and hence $\bar{P}_e = \bar{P}_e(\mathsf{m}_1, \mathsf{m}_2)$ for any messages $\mathsf{m}_1, \mathsf{m}_2$. If $\mathsf{m}_1, \mathsf{m}_2$ are the selected messages, then $\boldsymbol{x}_{\mathsf{m}_1,\mathsf{m}_2}$ denotes the sequence of modulation symbols corresponding to these messages.

For a given received sequence $\boldsymbol{y}$, the decoders φ_1 and φ_2 choose the messages $\widehat{\mathsf{m}}_1, \widehat{\mathsf{m}}_2$ with largest metric $q(\boldsymbol{x}_{\mathsf{m}_1,\mathsf{m}_2}, \boldsymbol{y})$. Let $(1, 1)$ be the selected message pair. In order to analyze the MSD decoder, it proves convenient to separately consider three possible error events,

(1) the decoder for $\mathcal{C}_1$ fails ($\varphi_1(\boldsymbol{y}) \neq 1$), but that for $\mathcal{C}_2$ is successful ($\varphi_2(\boldsymbol{y}) = 1$): the decoded codeword is $\boldsymbol{x}_{\mathsf{m}_1,1}$, $\mathsf{m}_1 \neq 1$;

(2) the decoder for $\mathcal{C}_2$ fails ($\varphi_2(\boldsymbol{y}) \neq 1$), but that for $\mathcal{C}_1$ is successful ($\varphi_1(\boldsymbol{y}) = 1$): the decoded codeword is $\boldsymbol{x}_{1,\mathsf{m}_2}$, $\mathsf{m}_2 \neq 1$;
(3) both decoders for $\mathcal{C}_1$ and $\mathcal{C}_2$ fail ($\varphi_1(\boldsymbol{y}) \neq 1, \varphi_2(\boldsymbol{y}) \neq 1$): the decoded codeword is $\boldsymbol{x}_{\mathsf{m}_1,\mathsf{m}_2}$, $\mathsf{m}_1 \neq 1$, $\mathsf{m}_2 \neq 1$.

We respectively denote the probabilities of these three alternative events by $P_{(1)}$, $P_{(2)}$, and $P_{(1,2)}$. Since the alternatives are not disjoint, application of the union bound to the error probability $\Pr\{\text{error}|\boldsymbol{x}_{1,1}, \boldsymbol{y}\}$ for a given choice of transmitted codeword $\boldsymbol{x}_{1,1}$ yields

$$\Pr\{\text{error}|\boldsymbol{x}_{1,1}, \boldsymbol{y}\} \leq P_{(1)} + P_{(2)} + P_{(1,2)}. \tag{3.43}$$

We next examine these summands separately. The probability in the first summand is identical to that of a coded modulation scheme with $|\mathcal{C}_1| - 1$ candidate codewords of the form $\boldsymbol{x}_{\mathsf{m}_1,1}$. Observe that the error probability is not decreased if the decoder has access to a genie giving the value of $\boldsymbol{c}_2$ [98]. As in the derivation of Eq. (3.12), we obtain that

$$P_{(1)} \leq (|\mathcal{C}_1| - 1)^{\rho} \left(\sum_{\boldsymbol{x}_{\mathsf{m}_1,1}} P_{\boldsymbol{X}}(\boldsymbol{x}_{\mathsf{m}_1,\mathsf{m}_2}) \frac{q(\boldsymbol{x}_{\mathsf{m}_1,1}, \boldsymbol{y})^s}{q(\boldsymbol{x}_{1,1}, \boldsymbol{y})^s} \right)^{\rho}. \tag{3.44}$$

Following exactly the same steps as in Section 3.1.2 we can express the average error probability in terms of a per-letter characterization, as

$$\bar{P}_{(1)} \leq \mathrm{e}^{-N\left(E^q_{0,(1)}(\rho,s) - \rho R_1\right)}, \tag{3.45}$$

where $E^q_{0,1}(\rho, s)$ is the corresponding generalized Gallager function,

$$E^q_{0,(1)}(\rho, s) \stackrel{\Delta}{=} -\log \mathbb{E}\left[\left(\sum_{b_1'} P_{B_1}(b_1') \frac{q(\mu(b_1', B_2), Y)^s}{q(\mu(B_1, B_2), Y)^s} \right)^{\rho} \right]. \tag{3.46}$$

Similarly, for the second summand, the probability satisfies

$$\bar{P}_{(2)} \leq \mathrm{e}^{-N\left(E^q_{0,(2)}(\rho,s) - \rho R_2\right)}, \tag{3.47}$$

with

$$E^q_{0,(2)}(\rho, s) \stackrel{\Delta}{=} -\log \mathbb{E}\left[\left(\sum_{b_2'} P_{B_2}(b_2') \frac{q(\mu(B_1, b_2'), Y)^s}{q(\mu(B_1, B_2), Y)^s} \right)^{\rho} \right]. \tag{3.48}$$

As for the third summand, there are $(|\mathcal{C}_1|-1)(|\mathcal{C}_2|-1)$ alternative candidate codewords of the form $\boldsymbol{x}_{\mathsf{m}_1,\mathsf{m}_2}$, which give the upper bound

$$\bar{P}_{(1,2)} \leq \mathrm{e}^{-N\left(E^q_{0,(1,2)}(\rho,s)-\rho(R_1+R_2)\right)}, \tag{3.49}$$

with the corresponding Gallager function,

$$E^q_{0,(1,2)}(\rho,s) \triangleq -\log \mathbb{E}\left[\left(\sum_{b_1',b_2'} P_{B_1}(b_1')P_{B_2}(b_2')\frac{q(\mu(b_1',b_2'),Y)^s}{q(\mu(B_1,B_2),Y)^s}\right)^{\rho}\right]. \tag{3.50}$$

Summarizing, the overall average error probability is bounded by

$$\begin{aligned}\bar{P}_e \leq\ & \mathrm{e}^{-N\left(E^q_{0,(1)}(\rho,s)-\rho R_1\right)} + \mathrm{e}^{-N\left(E^q_{0,(2)}(\rho,s)-\rho R_2\right)} \\ & + \mathrm{e}^{-N\left(E^q_{0,(1,2)}(\rho,s)-\rho(R_1+R_2)\right)}.\end{aligned} \tag{3.51}$$

For any choice of ρ, s, input distribution and $R_1, R_2 \geq 0$ such that $R_1+R_2=R$, we obtain a lower bound to the reliability function of MLC at rate R. For sufficiently large N, the error probability (3.51) is dominated by the minimum exponent. In other words,

$$E(R) \geq E^q_{\mathrm{r}}(R) \triangleq \max_{\substack{R_1,R_2 \\ R_1+R_2=R}} \min\left\{E^q_{\mathrm{r},(1)}(R_1)\,,\, E^q_{\mathrm{r},(2)}(R_2)\,,\, E^q_{\mathrm{r},(1,2)}(R)\right\} \tag{3.52}$$

where

$$E^q_{\mathrm{r},(1)}(R_1) \triangleq \max_{0\leq\rho\leq 1}\max_{s>0}\left(E^q_{0,(1)}(\rho,s)-\rho R_1\right) \tag{3.53}$$

$$E^q_{\mathrm{r},(2)}(R_2) \triangleq \max_{0\leq\rho\leq 1}\max_{s>0}\left(E^q_{0,(2)}(\rho,s)-\rho R_2\right) \tag{3.54}$$

$$E^q_{\mathrm{r},(1,2)}(R) \triangleq \max_{0\leq\rho\leq 1}\max_{s>0}\left(E^q_{0,(1,2)}(\rho,s)-\rho R\right). \tag{3.55}$$

Since $\mathcal{C}_1, \mathcal{C}_2$ are binary, the exponents $E^q_{\mathrm{r},(1)}(R)$ and $E^q_{\mathrm{r},(2)}(R)$ are always upper bounded by 1. Therefore, the overall exponent of MLC with MSD is smaller than one.

As we did in Section 3.1.2, analysis of Eq. (3.51) yields achievable rates. Starting the decoding at level 1, we obtain

$$R_1 < I(B_1;Y) \tag{3.56}$$

$$R_2 < I(B_2;Y|B_1). \tag{3.57}$$

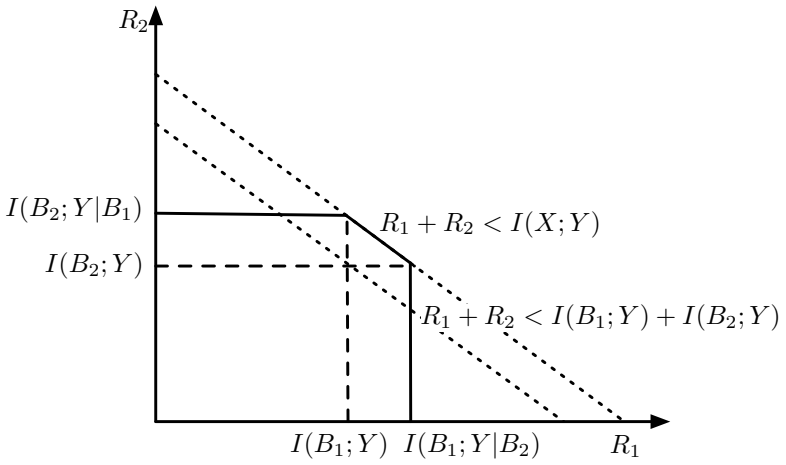


Fig. 3.8 Rate regions for MLC/MSD and BICM.

Generalization to a larger number of levels gives

$$R_j < I(B_j; Y|B_1, \ldots, B_{j-1}). \tag{3.58}$$

The chain rule of mutual information proves that MLC and MSD achieve the coded modulation capacity [56,140], $\sum_{j=1}^{m} R_j < I^{\mathrm{cm}}(X;Y)$. When MLC is decoded with the standard BICM decoder, i. e. without MSD, then the rates we obtain are

$$R_1 < I(B_1; Y) \tag{3.59}$$

$$R_2 < I(B_2; Y). \tag{3.60}$$

Figure 3.8 shows the resulting rate region for MLC. The Figure also compares with MLC with BICM decoding (i. e. without MSD), showing the corresponding achievable rates.

While MLC with MSD achieves the coded modulation capacity, it does not achieve the coded modulation error exponent. This is due to the MLC (with or without MSD) error exponent always being given by the minimum of the error exponents of the various levels, which results in an error exponent smaller than 1. While BICM suffers from a non-zero, yet small, capacity loss compared to CM and MLC/MSD,

BICM attains a larger error exponent, whose loss with respect to CM is small. This loss may be large for an MLC construction. In general, the decoding complexity of BICM is larger than that of MLC/MSD, since the codes of MLC are shorter. One such example is a code where only a few bits out of the m are coded while the rest are left uncoded. In practice, however, if the decoding complexity grows linearly with the number of bits in a a codeword, e. g. with LDPC or turbo codes, the overall complexity of BICM becomes comparable to that of MLC/MSD.

3.4 Mutual Information Analysis

In this section, we focus on AWGN channels with and without fading and study some properties of the mutual information as a function of snr. Building on work by Guo, Shamai and Verdú [51], we first provide a simple expression for the first derivative of the mutual information with respect to snr. This expression is of interest for the optimization of power allocation across parallel channels, as discussed by Lozano *et al.* [67] in the context of coded modulation systems.

Then, we study the BICM mutual information at low snr, that is in the wideband regime recently popularised by Verdú [134]. For a given rate, BICM with Gray labeling loses at most 1.25 dB in received power.

3.4.1 Derivative of Mutual Information

A fundamental relationship between the input-output mutual information and the minimum mean-squared error (MMSE) in estimating the input from the output in additive Gaussian channels was discovered by Guo, Shamai and Verdú in [51]. It is worth noting that, beyond its own intrinsic theoretical interest, this relationship has proved instrumental in optimizing the power allocation for parallel channels with arbitrary input distributions and in obtaining the minimum bit-energy-to-noise-spectral-density ratio for reliable communication [67].

For a scalar model $Y = \sqrt{\mathsf{snr}}X + Z$, it is shown in [51] that

$$\frac{d\,\mathsf{C}(\mathsf{snr})}{d\,\mathsf{snr}} = \mathsf{mmse}(\mathsf{snr}) \tag{3.61}$$

where $\mathsf{C}(\mathsf{snr}) = I(X;Y)$ is the mutual information expressed in nats,

$$\mathsf{mmse}(\mathsf{snr}) \triangleq \mathbb{E}\left[|X - \widehat{X}|^2\right] \tag{3.62}$$

is the MMSE of estimating the input X from the output Y of the given Gaussian channel model, and where

$$\widehat{X} \triangleq \mathbb{E}[X|Y] \tag{3.63}$$

is the MMSE estimate of the channel input X given the output Y. For Gaussian inputs we have that

$$\mathsf{mmse}(\mathsf{snr}) = \frac{1}{1+\mathsf{snr}}$$

while for general discrete signal constellations $\mathcal{X}$ we have that [67]

$$\mathsf{mmse}_{\mathcal{X}}(\mathsf{snr}) = \mathbb{E}\left[|X|^2\right] - \mathbb{E}\left[\left|\frac{\sum_{x'\in\mathcal{X}} x'\, \mathrm{e}^{-|\sqrt{\mathsf{snr}}(X-x')+Z|^2}}{\sum_{x'\in\mathcal{X}} \mathrm{e}^{-|\sqrt{\mathsf{snr}}(X-x')+Z|^2}}\right|^2\right]. \tag{3.64}$$

Figure 3.9 shows the function $\mathsf{mmse}(\mathsf{snr})$ for Gaussian inputs and various coded modulation schemes.

For BICM, obtaining a direct relationship between the BICM capacity and the MMSE in estimating the coded bits given the output is a challenging problem. However, the combination of Eqs. (3.33) and (3.61) yields a simple relationship between the first derivative of the BICM mutual information and the MMSE of coded modulation:

Theorem 3.3 ([49]). The derivative of the BICM mutual information is given by

$$\frac{d\,\mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}}(\mathsf{snr})}{d\,\mathsf{snr}} = \sum_{j=1}^{m} \frac{1}{2} \sum_{b=0}^{1} \left(\mathsf{mmse}_{\mathcal{X}}(\mathsf{snr}) - \mathsf{mmse}_{\mathcal{X}_b^j}(\mathsf{snr})\right) \tag{3.65}$$

where $\mathsf{mmse}_{\mathcal{A}}(\mathsf{snr})$ is the MMSE of an arbitrary input signal constellation $\mathcal{A}$ defined in (3.64).

Hence, the derivative of the BICM mutual information with respect to snr is a linear combination of MMSE functions for coded modulation.

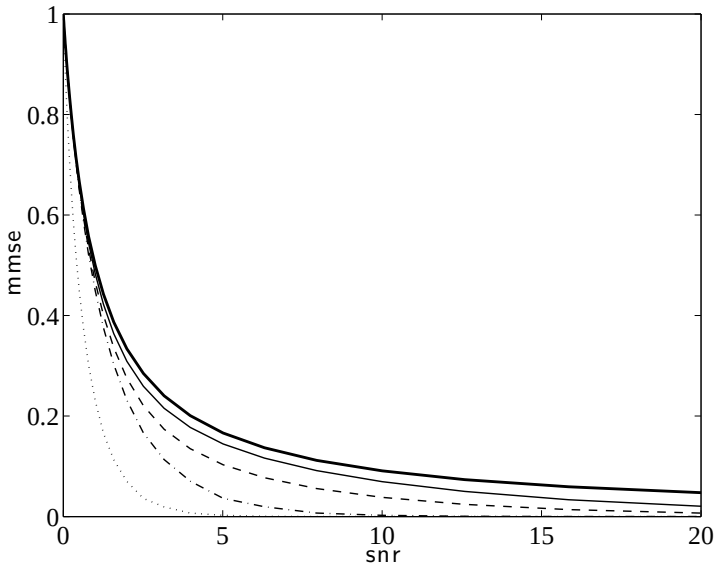


Fig. 3.9 MMSE for Gaussian inputs (thick solid line), BPSK (dotted line), QPSK (dash-dotted line), 8-PSK (dashed line) and 16-QAM (solid line).

Figure 3.10 shows an example (16-QAM modulation) of the computation of the derivative of the BICM mutual information. For comparison, the values of the MMSE for Gaussian inputs and for coded modulation and 16-QAM are also shown. At high snr we observe a very good match between coded modulation and BICM with binary reflected Gray labeling (dashed line). As for low snr, we notice a small loss, whose value is determined analytically from the analysis in the next section.

3.4.2 Wideband Regime

At very low signal-to-noise ratio snr, the energy of a single bit is spread over many channel degrees of freedom, leading to the wideband regime recently discussed at length by Verdú [134]. Rather than studying the exact expression of the channel capacity, one considers a second-order Taylor series in terms of snr,

$$\mathsf{C}(\mathsf{snr}) = c_1\mathsf{snr} + c_2\mathsf{snr}^2 + \mathrm{o}\big(\mathsf{snr}^2\big), \tag{3.66}$$

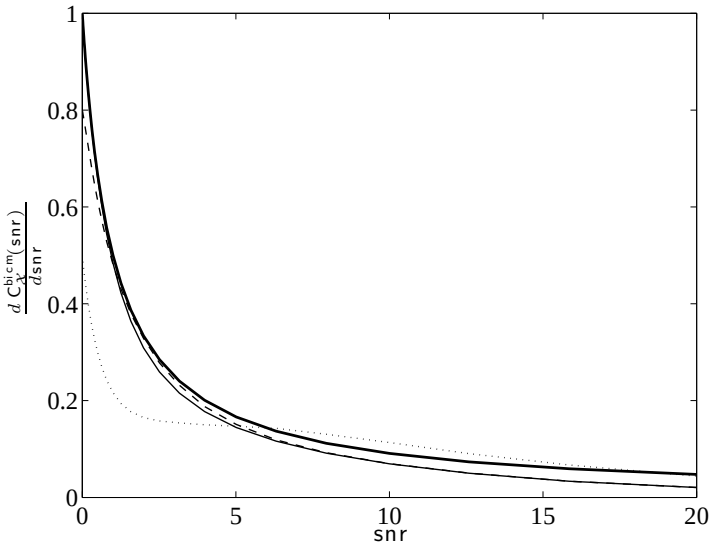


Fig. 3.10 Derivative of the mutual information for Gaussian inputs (thick solid line), 16-QAM coded modulation (solid line), 16-QAM BICM with Gray labeling (dashed line) and 16-QAM BICM with set partitioning labeling (dotted line).

where c_1 and c_2 depend on the modulation format, the receiver design, and the fading distribution. The notation $\mathrm{o}(\mathsf{snr}^2)$ indicates that the remaining terms vanish faster than a function $a\mathsf{snr}^2$, for $a > 0$ and small snr. Here the capacity may refer to the coded modulation capacity, or the BICM capacity.

In the following, we determine the coefficients c_1 and c_2 in the Taylor series (3.66) for generic constellations, and use them to derive the corresponding results for BICM. Before proceeding along this line, we note that [134, Theorem 12] covers the effect of fading. The coefficients c_1 and c_2 for a general fading distribution are given by

$$c_1 = \mathbb{E}\left[|H|^2\right] c_1^{\mathrm{awgn}}, \quad c_2 = \mathbb{E}\left[|H|^4\right] c_2^{\mathrm{awgn}}, \tag{3.67}$$

where the coefficients c_1^{awgn} and c_2^{awgn} are in absence of fading. Hence, even though we focus only on the AWGN channel, all results are valid for general fading distributions.

Next to the coefficients c_1 and c_2, Verdú also considered an equiv-

alent pair of coefficients, the energy per bit to noise power spectral density ratio at zero snr and the wideband slope [134]. These parameters are obtained by transforming Eq. (3.66) into a function of the $\frac{E_\mathrm{b}}{N_0} = \frac{\mathsf{snr}}{\mathsf{C}\log_2 e}$, so that one obtains

$$\mathsf{C}\left(\frac{E_\mathrm{b}}{N_0}\right) = \zeta_0 \left(\frac{E_\mathrm{b}}{N_0} - \frac{E_\mathrm{b}}{N_{0\,\mathrm{lim}}}\right) + \mathrm{O}\left(\left(\Delta\frac{E_\mathrm{b}}{N_0}\right)^2\right) \tag{3.68}$$

where $\Delta\frac{E_\mathrm{b}}{N_0} \triangleq \frac{E_\mathrm{b}}{N_0} - \frac{E_\mathrm{b}}{N_{0\,\mathrm{lim}}}$ and

$$\zeta_0 \triangleq -\frac{c_1^3}{c_2 \log^2 2}, \qquad \frac{E_\mathrm{b}}{N_{0\,\mathrm{lim}}} \triangleq \frac{\log 2}{c_1}. \tag{3.69}$$

The notation $\mathrm{O}(x^2)$ indicates that the remaining terms decay at least as fast as a function ax^2, for $a > 0$ and small x. The parameter ζ_0 is Verdú's wideband slope in linear scale [134]. We avoid using the word minimum for $\frac{E_\mathrm{b}}{N_{0\,\mathrm{lim}}}$, since there exist communication schemes with a negative slope ζ_0, for which the absolute minimum value of $\frac{E_\mathrm{b}}{N_0}$ is achieved at non-zero rates. In these cases, the expansion at low power is still given by Eq. (3.68).

For Gaussian inputs, we have $c_1 = 1$ and $c_2 = -\frac{1}{2}$. Prelov and Verdú determined c_1 and c_2 in [94] for proper-complex constellations, previously introduced by Neeser and Massey in [86]. These constellations satisfy $\mathbb{E}[X^2] = 0$, where $\mathbb{E}[X^2]$ is a second-order pseudo-moment [86]. We similarly define a pseudo-variance, denoted by $\hat{\sigma}_X^2$, as

$$\hat{\sigma}_X^2 \triangleq \mathbb{E}[X^2] - \mathbb{E}[X]^2. \tag{3.70}$$

Analogously, we define the constellation variance as $\sigma_X^2 \triangleq \mathbb{E}\left[|X|^2\right] - |\mathbb{E}[X]|^2$. The coefficients for coded modulation schemes with arbitrary first and second moments are given by the following result:

Theorem 3.4 ([76, 94]). Consider coded modulation schemes over a general signal set $\mathcal{X}$ used with probabilities $P_X(x)$ in the AWGN channel described by (2.6). Then, the first two coefficients of the Taylor expansion of the coded modulation capacity $\mathsf{C}(\mathsf{snr})$ around $\mathsf{snr} = 0$ are

$$c_1 = \sigma_X^2 \tag{3.71}$$

$$c_2 = -\frac{1}{2}\left(\left(\sigma_X^2\right)^2 + \left|\hat{\sigma}_X^2\right|^2\right). \tag{3.72}$$

For zero-mean unit-energy signal sets, we obtain the following

Corollary 3.3 ([76]). Coded modulation schemes over a signal set $\mathcal{X}$ with $\mathbb{E}[X] = 0$ (zero mean) and $\mathbb{E}[|X|^2] = 1$ (unit energy) have

$$c_1 = 1, \qquad c_2 = -\frac{1}{2}\Big(1 + \big|\mathbb{E}[X^2]\big|^2\Big). \tag{3.73}$$

Alternatively, the quantity c_1 can be simply obtained as $c_1 = \mathsf{mmse}(0)$. Observe also that $\frac{E_\mathrm{b}}{N_0}_{\lim} = \log 2$.

Plotting the mutual information curves as a function of $\frac{E_\mathrm{b}}{N_0}$ (shown in Figure 3.11) reveals the suboptimality of the BICM decoder. In particular, even binary reflected Gray labeling is shown to be information lossy at low rates. Based on the expression (3.33), one obtains

Theorem 3.5 ([76]). The coefficients c_1 and c_2 of $\mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}}$ for a constellation $\mathcal{X}$ with zero mean and unit average energy are given by

$$c_1 = \sum_{j=1}^{m} \frac{1}{2} \sum_{b=0}^{1} \Big|\mathbb{E}[X_b^j]\Big|^2, \tag{3.74}$$

$$c_2 = \sum_{j=1}^{m} \frac{1}{4} \sum_{b=0}^{1} \left(\Big(\sigma^2_{X_b^j}\Big)^2 + \Big|\hat{\sigma}^2_{X_b^j}\Big|^2 - 1 - \big|\mathbb{E}[X^2]\big|^2 \right). \tag{3.75}$$

Table 3.1 reports the numerical values for c_1, c_2, $\frac{E_\mathrm{b}}{N_0}_{\lim}$, and ζ_0 for various cases, namely QPSK, 8-PSK and 16-QAM with binary reflected Gray and Set Partitioning (anti-Gray for QPSK) mappings.

In Figure 3.12, the approximation in Eq. (3.68) is compared with the capacity curves. As expected, a good match for low rates is observed. We use labels to identify the specific cases: labels 1 and 2 are QPSK, 3 and 4 are 8-PSK and 5 and 6 are 16-QAM. Also shown is the linear approximation to the capacity around $\frac{E_\mathrm{b}}{N_0}_{\lim}$, given by Eq. (3.68). Two cases with Nakagami fading (with density in Eq. (2.8)) are also included in Figure 3.12, which also show good match with the estimate, taking

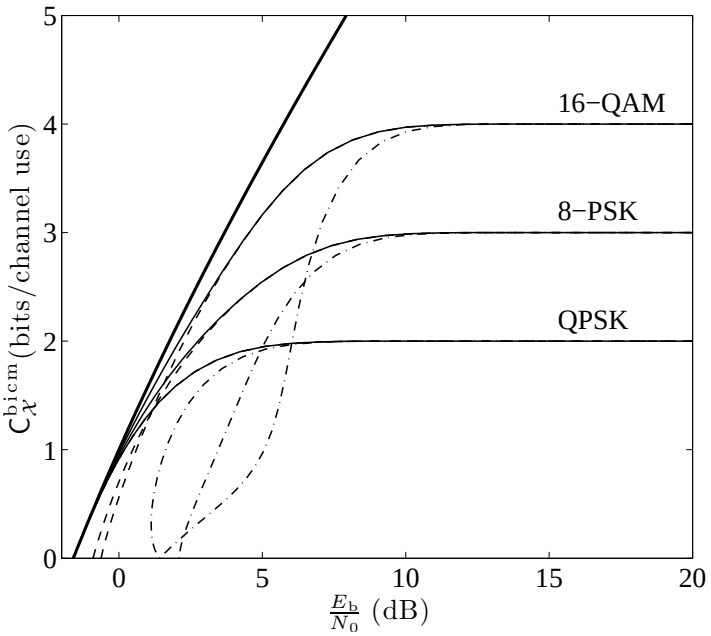


Fig. 3.11 Coded modulation and BICM capacities (in bits per channel use) for multiple signal constellations with uniform inputs in the AWGN channel. Gray and set partitioning labeling rules correspond to thin dashed and dashed-dotted lines respectively. For reference, the capacity with Gaussian inputs (3.4) is shown in thick solid lines and the CM channel capacity with thin solid lines.

into account that $\mathrm{E}[|H|^2] = 1$ and $\mathrm{E}[|H|^4] = 1 + 1/m_f$ for Nakagami-m_f fading [95]. An exception is 8-PSK with set-partitioning, where the large slope limits the validity of the approximation to very low rates.

It seems hard to make general statements for arbitrary labelings from Theorem 3.5. An important exception is the strictly regular set partitioning labeling defined by Stierstorfer and Fischer [110], which has $c_1 = 1$ for 16-QAM and 64-QAM. In constrast, for binary reflected Gray labeling (see Section 2.3) we have:

Theorem 3.6 ([76]). For M-PAM and M^2-QAM and binary-reflected Gray labeling, the coefficient c_1 is

$$c_1 = \frac{3 \cdot M^2}{4(M^2 - 1)}, \tag{3.76}$$

Table 3.1 $\frac{E_\mathrm{b}}{N_0}_{\mathrm{lim}}$ and wideband slope coefficients c_1, c_2 for BICM in AWGN.

	Modulation and Labeling					
	QPSK		8-PSK		16-QAM	
	GR	A-GR	GR	SP	GR	SP
c_1	1.000	0.500	0.854	0.427	0.800	0.500
$\frac{E_\mathrm{b}}{N_0}_{\mathrm{lim}}$ (dB)	-1.592	1.419	-0.904	2.106	-0.627	1.419
c_2	-0.500	0.250	-0.239	0.005	-0.160	-0.310
ζ_0	4.163	-1.041	5.410	-29.966	6.660	0.839

and the minimum $\frac{E_\mathrm{b}}{N_0}_{\mathrm{lim}}$ is

$$\frac{E_\mathrm{b}}{N_0}_{\mathrm{lim}} = \frac{4(M^2-1)}{3 \cdot M^2} \log 2. \tag{3.77}$$

As $M \to \infty$, $\frac{E_\mathrm{b}}{N_0}_{\mathrm{lim}}$ approaches $\frac{4}{3}\log 2 \simeq -0.3424$dB. from below.

The results for BPSK, QPSK (2-PAM×2-PAM), and 16-QAM (4-PAM×4-PAM), as presented in Table 3.1, match with the Theorem.

It is somewhat surprising that the loss incurred by binary reflected Gray labeling with respect to coded modulation is bounded at low snr. The loss for large M represents about 1.25 dB with respect to the classical CM limit, namely $\frac{E_\mathrm{b}}{N_0}_{\mathrm{lim}} = -1.59$ dB. Using a single modulation for all signal-to-noise ratio values, adjusting the transmission rate by changing the code rate using a suboptimal non-iterative demodulator, needs not result in a large loss with respect to optimal schemes, where both the rate and modulation can change. Another low-complexity solution is to change the mapping only (not the modulation) according to SNR, switching between Gray and the mappings of [110]. This has low-implementation complexity as it is implemented digitally.

3.4.2.1 Bandwidth and Power Trade-off

In the previous section we computed the first coefficients of the Taylor expansion of the CM and BICM capacities around $\mathsf{snr} = 0$. We now

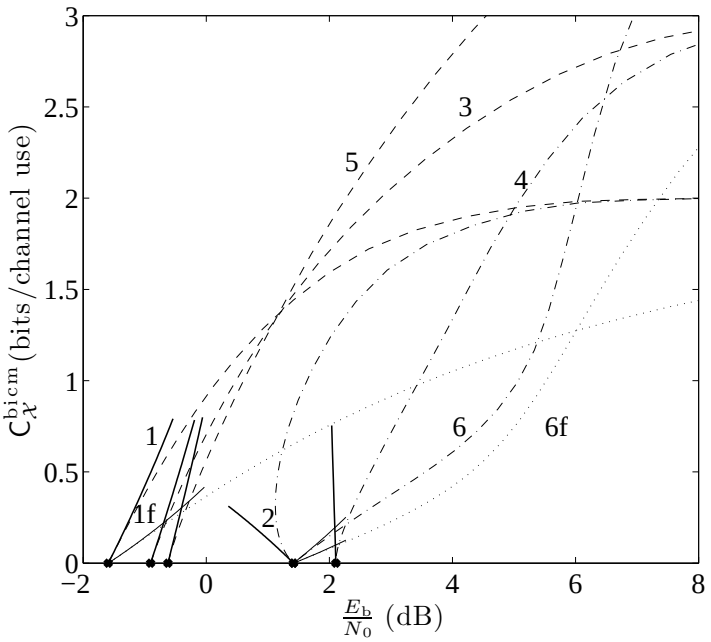


Fig. 3.12 BICM channel capacity (in bits per channel use). Labels 1 and 2 are QPSK, 3 and 4 are 8-PSK and 5 and 6 are 16-QAM. Gray and set partitioning labeling rules correspond to dashed (and odd labels) and dashed-dotted lines (and even labels) respectively. Dotted lines are cases 1 and 6 with Nakagami-0.3 and Nakagami-1 (Rayleigh) fading (an 'f' is appended to the label index). Solid lines are linear approximation around $\frac{E_b}{N_0}_{\lim}$.

use these coefficients to determine the trade-off between power and bandwidth in the low-power regime. We will see how to trade off part of the power loss incurred by BICM against a large bandwidth reduction.

As discussed in Section 2.A, the data rate transmitted across a waveform Gaussian channel is determined by two physical variables: the power P, or energy per unit time, and the bandwidth W, or number of channel uses per unit time. In this case, the signal-to-noise ratio snr is given by $\mathsf{snr} = P/(N_0 W)$, where N_0 is the noise power spectral density. Then, the capacity measured in bits per unit time is the natural figure of merit for a communications system. With only a constraint on snr, this capacity is given by $W \log(1+\mathsf{snr})$. For low snr, we have the Taylor

series expansion

$$W \log\left(1 + \frac{P}{N_0 W}\right) = \frac{P}{N_0} - \frac{P^2}{2N_0^2 W} + \mathrm{O}\left(\frac{P^3}{N_0^3 W^2}\right). \tag{3.78}$$

Similarly, for coded modulation systems with capacity $\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}$, we have

$$\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}} W = c_1 \frac{P}{N_0} + c_2 \frac{P^2}{N_0^2 W} + \mathrm{O}\left(\frac{P^{5/2}}{N_0^{5/2} W^{3/2}}\right). \tag{3.79}$$

Following Verdú [134], we consider the following scenario. Let two alternative transmission systems with respective powers P_i and bandwidths W_i, $i = 1, 2$, achieve respective capacities per channel use C_i. The corresponding first- and second-order Taylor series coefficients are denoted by c_{11}, c_{21} for the first system, and c_{12}, c_{22} for the second. A natural comparison is to fix a power ratio $\Delta P = P_2/P_1$ and then solve for the corresponding bandwidth ratio $\Delta W = W_2/W_1$ so that the data rate is the same, that is $\mathsf{C}_1 W_1 = \mathsf{C}_2 W_2$. For instance, option 1 can be QPSK and option 2 use of a high-order modulation with BICM.

When the capacities C_1 and C_2 can be evaluated, the exact trade-off curve $\Delta W(\Delta P)$ can be computed. For low power, a good approximation is obtained by keeping the first two terms in the Taylor series. Under this approximation, we have the following result.

Theorem 3.7 ([76]). Around $\mathsf{snr}_1 = 0$, and neglecting terms $\mathrm{o}(\mathsf{snr}_1)$, the capacities in bits per second, $\mathsf{C}_1 W_1$ and $\mathsf{C}_2 W_2$ are equal when the power and bandwidth expansion ratios ΔP and ΔW are related as

$$\Delta W \simeq \frac{c_{22} \mathsf{snr}_1 (\Delta P)^2}{c_{11} + c_{21} \mathsf{snr}_1 - c_{12} \Delta P}, \tag{3.80}$$

for ΔW as a function of ΔP and, if $c_{12} \neq 0$,

$$\Delta P \simeq \frac{c_{11}}{c_{12}} + \left(\frac{c_{21}}{c_{12}} - \frac{c_{22} c_{11}^2}{c_{12}^3 \Delta W}\right) \mathsf{snr}_1, \tag{3.81}$$

for ΔP as a function of ΔW.

The previous theorem leads to the following derived results.

Corollary 3.4. For $\Delta P = 1$, we obtain

$$\Delta W \simeq \frac{c_{22}\mathsf{snr}_1}{c_{11} + c_{21}\mathsf{snr}_1 - c_{12}}, \tag{3.82}$$

and for the specific case $c_{11} = c_{12}$, $\Delta W \simeq c_{22}/c_{21}$.

As noticed in [134], the loss in bandwidth may be significant when $\Delta P = 1$. But this point is just one of a curve relating ΔP and ΔW. For instance, with no bandwidth expansion we have

Corollary 3.5. For $c_{11} = c_{12} = 1$, choosing $\Delta W = 1$ gives $\Delta P \simeq 1 + \big(c_{21} - c_{22}\big)\mathsf{snr}_1$.

For signal-to-noise ratios below -10 dB, the approximation in Theorem 3.7 seems to be very accurate for "reasonable" power or bandwidth expansion ratios. A quantitative definition would lead to the problem of the extent to which the second order approximation to the capacity is correct, a question on which we do not dwell further.

Figure 3.13 depicts the trade-off for between QPSK and BICM over 16-QAM (with Gray labeling) for two values of signal-to-noise ratio. The exact result, obtained by using the exact formulas for $\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}$ and $\mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}}$, is plotted along the result by using Theorem 3.7. As expected from the values of c_1 and c_2, use of 16-QAM incurs in a non-negligible power loss. On the other hand, this loss may be accompanied by a significant reduction in bandwidth, which might be of interest in some applications. For signal-to-noise ratios larger than those reported in the figure, the assumption of low snr loses its validity and the results derived from the Taylor series are no longer accurate.

3.5 Concluding Remarks and Related Work

In this Chapter, we have reviewed the information-theoretic foundations of BICM and we have compared them with those of coded modulation. In particular, we have re-developed Gallager's analysis for the average error probability of the random coding ensemble for a generic mismatched decoding metric, of which BICM is a particular case. We

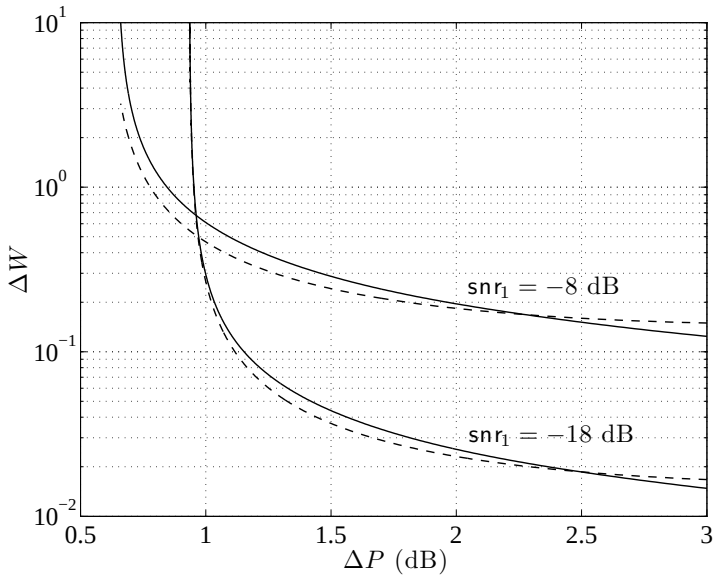


Fig. 3.13 Trade-off between ΔP and ΔW between QPSK and 16-QAM with Gray labeling. Exact tradeoff in solid lines, dashed lines for the low-SNR tradeoff.

have shown that the resulting error exponent cannot be larger than that of coded modulation and that the loss in error exponent with respect to coded modulation is small for binary reflected Gray labeling. We have shown that the largest rate achievable by the random coding construction, i. e. the generalized mutual information, coincides with the BICM capacity of [29], providing an achievability proof without resorting to the independent parallel channel model. We have compared the error exponents of BICM with those of multilevel coding with multi-stage decoding [56, 140]. We have shown that the error exponent of multilevel coding cannot be larger than one, while the error exponent of BICM does not show this restriction, and can hence be larger. Building upon these considerations, we have analyzed the BICM capacity in the wideband regime, or equivalently, when the SNR is very low. We have determined the minimum energy-per-bit-to-noise-power-spectral-density-ratio for BICM as well as the wideband slope with arbitrary labeling. We have also shown that, with binary reflected Gray labeling, the loss in minimum energy per bit to noise power spectral density

ratio with respect to coded modulation is at most 1.25 dB. We have also given a simple and general expression for the first derivative of the BICM capacity with respect to SNR.

A number of works have studied various aspects related to the BICM capacity and its application. An aspect of particular relevance is the impact of binary labeling on the BICM capacity. Based on the calculation of the BICM capacity, Caire et al. conjectured that Gray labeling maximizes the BICM capacity. As shown in [109], binary reflected Gray labeling for square QAM constellations maximizes the BICM capacity for medium-to-large signal-to-noise ratios, while different Gray labelings might show a smaller BICM capacity. A range of labeling rules have been proposed in the literature. However, there has been no attempt to systematically classify and enumerate all labeling rules for a particular signal constellation. Reference [26] provides an exhaustive labeling classification for 8-PSK based on bit-wise distance spectra for the BICM decoding metric.

Determination of the best labeling rule is thus an open problem. More generally, an analytic determination of the reason why the CM and BICM capacities (and error exponents) are very close for Gaussian channels, and whether this closeness extends to more general channels is also open. In Chapter 6 we review some applications of BICM to other channels. Also, in Chapter 5 we give an overview of the current information-theoretic characterization of iterative decoding of BICM via density evolution and EXIT charts.

4

Error Probability Analysis

In this Chapter, we present several bounds and approximations to the error probability of BICM. Our presentation incorporates fundamental traits from [29, 74, 104, 141]. As we mentioned in Chapter 2, special attention is paid to the union bound and the Gaussian-noise channel with and without fully-interleaved fading.

We first introduce a general method for estimating the pairwise error probability of a generic maximum-metric decoder, where the metric need not necessarily be the likelihood. As we saw in the previous chapter, BICM is a paramount example of such mismatched decoding, and our analysis is therefore directly applicable. The presentation is built around the concept of decoding score, a random variable whose positive tail probability yields the pairwise error probability. Our discussion of the computation of the pairwise error probability is similar to the analysis in Sections 5.3 and 5.4 of [39], or to the presentation of Chapters 2 and 3 of [139]. Exact expressions for the pairwise error probability are difficult to obtain, and we resort to bounds and approximations to estimate it. In particular, we will study the Chernoff and Bhattacharyya bounds, and the saddlepoint and Gaussian approximations, and show that these are simple to compute in practice. As we shall see, the sad-

dlepoint approximation often yields a good approximation.

Section 4.2 follows the path proposed by Caire *et al.* in [29], namely, modeling the BICM channel as a set of parallel binary-input output-symmetric channels. This analysis leads to a first approximation of the pairwise error probability, which we denote by $\mathrm{PEP}_1(d)$, d being the Hamming distance between the competing and reference codewords of the underlying binary code $\mathcal{C}$.

We then use the analysis of Yeh *et al.* [141] to derive general expressions for the error probability using a uniform interleaver, namely, the average over all possible interleavers, as done in [8, 9] for turbo-codes. In Section 4.3 we present this general expression, denoted by $\mathrm{PEP}(d)$, and discuss the extent to which it can be accurately approximated by $\mathrm{PEP}_1(d)$. We put forward the idea that the operation of BICM with uniform interleaving is close to that of Berrou's turbo codes in the following sense. In the context of fading channels, a deep fade affects all the m bits in the label. Consider now a pairwise error event. As noticed by Zehavi and Caire *et al.* [29, 142], thanks to the interleaver, BICM is able to achieve a larger diversity than that of standard coded modulation, since the bits corresponding to bad error events may be spread over different modulation symbols. We shall argue that these bad error events remain, but they are subtly weighted by a low error probability, remaining thus hidden for most practical purposes. We give a quantitative description of this behaviour by first focussing on the simple case of QPSK modulation with Gray labeling and fully-interleaved fading, and then extending the results to more general constellations.

Finally, we conclude this chapter with a brief section outlining possible extensions of the union bound for BICM to the region beyond the cutoff rate. We chiefly use the results reported in Sason and Shamai's monograph on improved bounds beyond the cutoff rate [101].

4.1 Error Probability and the Union Bound

In Chapter 2, we expressed the message error probability as

$$P_e = \frac{1}{|\mathcal{M}|} \sum_{\mathsf{m}=1}^{|\mathcal{M}|} P_e(\mathsf{m}) \tag{4.1}$$

where $P_e(\mathsf{m})$ is the probability of message error when message m was transmitted. As we saw in Chapter 3, obtaining exact expressions for P_e can be difficult. Instead, we commonly resort to bounds. From the standard union bound technique, we obtain that

$$P_e \leq \frac{1}{|\mathcal{M}|} \sum_{\mathsf{m}=1}^{|\mathcal{M}|} \sum_{\mathsf{m}' \neq \mathsf{m}} \mathrm{PEP}(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{x}_{\mathsf{m}}) \tag{4.2}$$

where

$$\mathrm{PEP}(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{x}_{\mathsf{m}}) \stackrel{\Delta}{=} \Pr\{q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y}) > q(\boldsymbol{x}_{\mathsf{m}}, \boldsymbol{y})\} \tag{4.3}$$

is the *pairwise error probability*, i. e. the probability that the maximum metric decoder decides in favor of $\boldsymbol{x}_{\mathsf{m}'}$ when $\boldsymbol{x}_{\mathsf{m}}$ was the transmitted codeword. In the analysis of BICM we quite often concentrate on the codewords of the underlying binary code $\boldsymbol{c}_{\mathsf{m}} \in \mathcal{C}$ instead of the modulated codewords $\boldsymbol{x}_{\mathsf{m}} \in \mathcal{M}$. Let $\boldsymbol{c}_{\mathsf{m}} = (c_1, \ldots, c_n)$ denote the reference codeword (corresponding to the transmitted message m) and $\boldsymbol{c}_{\mathsf{m}'} = (c'_1, \ldots, c'_n)$ denote the competing codeword (corresponding to the transmitted message m'). Since there is a one-to-one correspondence with the binary codewords $\boldsymbol{c}_{\mathsf{m}}$ and the modulated codewords $\boldsymbol{x}_{\mathsf{m}}$, with some abuse of notation we write that

$$P_e \leq \frac{1}{|\mathcal{M}|} \sum_{\mathsf{m}=1}^{|\mathcal{M}|} \sum_{\mathsf{m}' \neq \mathsf{m}} \mathrm{PEP}(\boldsymbol{c}_{\mathsf{m}'}, \boldsymbol{c}_{\mathsf{m}}) \tag{4.4}$$

where

$$\mathrm{PEP}(\boldsymbol{c}_{\mathsf{m}'}, \boldsymbol{c}_{\mathsf{m}}) \stackrel{\Delta}{=} \Pr\{q(\boldsymbol{c}_{\mathsf{m}'}, \boldsymbol{y}) > q(\boldsymbol{c}_{\mathsf{m}}, \boldsymbol{y})\}. \tag{4.5}$$

Since the symbol x_k is selected as $x_k = \mu(b_1(x_k), \ldots, b_m(x_k))$, we can write the decoder metric as

$$q(\boldsymbol{x}, \boldsymbol{y}) = \prod_{k=1}^{N} q(x_k, y_k) \tag{4.6}$$

$$= \prod_{k=1}^{N} \prod_{j=1}^{m} q_j\big(b_j(x_k), y_k\big), \tag{4.7}$$

and we can define the *pairwise score* as

$$\Xi^{\mathrm{pw}} \triangleq \sum_{k=1}^{N}\sum_{j=1}^{m} \log \frac{q_j(c'_{m(k-1)+j}, y_k)}{q_j(c_{m(k-1)+j}, y_k)}. \tag{4.8}$$

The pairwise score can be expressed as

$$\Xi^{\mathrm{pw}} = \sum_{k=1}^{N} \Xi_k^{\mathrm{s}} = \sum_{k=1}^{N}\sum_{j=1}^{m} \Xi_{k,j}^{\mathrm{b}} \tag{4.9}$$

where

$$\Xi_k^{\mathrm{s}} \triangleq \sum_{j=1}^{m} \log \frac{q_j(c'_{m(k-1)+j}, y_k)}{q_j(c_{m(k-1)+j}, y_k)} \tag{4.10}$$

is the k-th *symbol* score and

$$\Xi_{k,j}^{\mathrm{b}} \triangleq \log \frac{q_j(c'_{m(k-1)+j}, y_k)}{q_j(c_{m(k-1)+j}, y_k)} \tag{4.11}$$

is the *bit* score corresponding to the j-th bit of the k-th symbol. Clearly, only the bit indices in which the codewords $\boldsymbol{c}_{\mathsf{m}'}$ and $\boldsymbol{c}_{\mathsf{m}}$ differ have a non-zero bit score. Since some of these bit indices might be modulated in the same constellation symbol, we have m classes of symbol scores, each characterized by a different number of wrong bits (that is, the Hamming weight of the binary labels). Then, we have the following

---

Proposition 4.1. The pairwise error probability $\mathrm{PEP}(\boldsymbol{c}_{\mathsf{m}'}, \boldsymbol{c}_{\mathsf{m}})$ between a reference codeword $\boldsymbol{c}_{\mathsf{m}}$ and the competing codeword $\boldsymbol{c}_{\mathsf{m}'}$ is given by

$$\mathrm{PEP}(\boldsymbol{c}_{\mathsf{m}'}, \boldsymbol{c}_{\mathsf{m}}) = \Pr\{\Xi^{\mathrm{pw}} > 0\}, \tag{4.12}$$

where Ξ^{pw} is the pairwise decoding score.

These scores are random variables whose density function depends on all the random elements in the channel, as well as the transmitted bits, their position in the symbol and the bit pattern. In order to avoid this dependence, we will use the random coset code method used in [60] to analyze LDPC codes for the Inter-Symbol Interference (ISI) channel and in [10] to analyze nonbinary LDPC codes. This method consists of

adding to every transmitted codeword $\boldsymbol{c} \in \mathcal{C}$ a random binary word $\boldsymbol{d} \in \{0,1\}^n$, which is assumed to be known by the receiver. This is equivalent to scrambling the output of the encoder by a sequence known at the receiver. Scrambling guarantees that the symbols corresponding to two m-bit sequences $(c_1, \ldots, c_m)$ and $(c'_1, \ldots, c'_m)$ are mapped to all possible pairs of modulation symbols differing in a given Hamming weight, hence making the channel symmetric. Clearly, the error probability computed this way gives an average over all possible scrambling sequences.

Remark 4.1. In [29], the scrambler role was played by randomly choosing between a mapping rule μ and its complement $\bar{\mu}$ with probability $1/2$ at every channel use. As here, the choice between μ and $\bar{\mu}$ is known to the receiver. Scrambling is the natural extension of this random choice to symbols of Hamming weight larger than 1.

4.1.1 Linear Codes

If the underlying binary code $\mathcal{C}$ is linear and the channel is symmetric, the pairwise error probability depends on the transmitted codeword $\boldsymbol{c}_{\mathsf{m}}$ and the competing codeword $\boldsymbol{c}_{\mathsf{m}'}$ only through their respective Hamming distance d [139]. In this case, we may rewrite (4.4) as

$$P_e \leq \sum_d A_d \ \mathrm{PEP}(d), \tag{4.13}$$

where A_d is the weight enumerator, i. e. the number of codewords of $\mathcal{C}$ with Hamming weight d, and $\mathrm{PEP}(d)$ denotes the error probability of a pairwise error event of weight d. The quantity $\mathrm{PEP}(d)$ will be central in this chapter. In some cases it is possible to obtain a closed-form expression for $\mathrm{PEP}(d)$, whereas in other cases we will need to resort to further bounds or approximations. Despite its simplicity, the union bound accurately characterizes the error probability in the region above the cutoff rate [139].

The bit error probability can be simply bounded by

$$P_b \leq \sum_d A'_d \ \mathrm{PEP}(d), \tag{4.14}$$

where

$$A'_d \triangleq \sum_i \frac{i}{n} A_{i,d}, \tag{4.15}$$

where $A_{i,d}$ is the input-output weight enumerator, i. e. number of codewords of $\mathcal{C}$ of Hamming weight d generated with information messages of Hamming weight i.

4.1.2 Cumulant Transforms of Symbol Scores

In this section, we introduce the definition of the cumulant transform and apply it to the symbol scores. The cumulant transform of the symbol and bit scores, compared to other equivalent representations such as the characteristic function of the moment generating function, will show particularly convenient in accurately approximating the pairwise error probability with the saddlepoint approximation [28, 57].

Let U be a random variable. Then, for $s \in \mathbb{C}$ we define its cumulant transform as [28, 57]

$$\kappa(s) \triangleq \log \mathbb{E}\left[\mathrm{e}^{sU}\right]. \tag{4.16}$$

The cumulant transform is an equivalent representation of the probability density function. Whenever needed, the density can be recovered by an inverse Fourier transform.

The derivatives of the cumulant transform are respectively denoted by $\kappa'(s)$, $\kappa''(s)$, $\kappa'''(s)$, and $\kappa^{(\nu)}(s)$ for $\nu > 3$.

In Eq. (4.9) we expressed the pairwise decoding score as a sum of non-zero symbol scores, each of them corresponding to a symbol where the codewods differ in at least 1 bit. Since there are m bits in the binary label, there are m classes of symbol scores, one for each of the possible Hamming weigths. Consider a symbol score of weight v, $1 \leq v \leq m$.

Definition 4.1. The cumulant transform of a symbol score Ξ^{s} with Hamming weight v, denoted by $\kappa_v(s)$, is given by

$$\begin{aligned}\kappa_v(s) &= \log \mathbb{E}\left[\mathrm{e}^{s\,\Xi^{\mathrm{s}}}\right] \\ &= \log\left(\frac{1}{\binom{m}{v}} \sum_{\boldsymbol{j}=(j_1,\ldots,j_v)} \frac{1}{2^v} \sum_{\boldsymbol{b}\in\{0,1\}^v} \mathbb{E}\left[\frac{\prod_{i=1}^v q_{j_i}(\bar{b}_{j_i}, Y)^s}{\prod_{i=1}^v q_{j_i}(b_{j_i}, Y)^s}\right]\right),\end{aligned} \tag{4.17}$$

where $\boldsymbol{j} = (j_1, \ldots, j_v)$ is a sequence of v bit indices, drawn from all the possible such v-tuples, and Y are the channel outputs with bit v-tuple $\boldsymbol{b}$ transmitted at positions in $\boldsymbol{j}$. The remaining expectation in (4.17) is done according to the probability $P_{\boldsymbol{j}}(y|\boldsymbol{b})$,

$$P_{\boldsymbol{j}}(y|\boldsymbol{b}) = P_{j_1,\ldots,j_v}(y|b_1,\ldots,b_v) = \frac{1}{2^{m-v}} \sum_{x \in \mathcal{X}_{b_1,\ldots,b_v}^{j_1,\ldots,j_v}} P_{Y|X}(y|x) \quad (4.18)$$

where $\mathcal{X}_{b_1,\ldots,b_v}^{j_1,\ldots,j_v}$, defined in Eq. (2.13), is the set of symbols with bit labels in positions $j_1, \ldots, j_v$ equal to $b_1, \ldots, b_v$.

A particularly important case of the above definition is $v = 1$, for which the symbol score becomes the bit score. The binary labels of the reference and competing symbols in the symbol score differ only by a single bit, and all d different bits of the pairwise error between the reference and competing codewords are mapped onto different modulation symbols. This is the case for interleavers of practical length. As we will see in the next sections, this will significantly simplify the analysis.

Definition 4.2. The cumulant transform of the bit score corresponding to a symbol score with a single different bit, denoted by Ξ_1^{b}, is given by

$$\begin{aligned} \kappa_1(s) &= \log \mathbb{E}\big[\mathrm{e}^{s\,\Xi_1^{\mathrm{b}}}\big] \\ &= \log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{1}{2}\sum_{b\in\{0,1\}}\mathbb{E}\left[\frac{q_j(\bar{b},Y)^s}{q_j(b,Y)^s}\right]\right), \end{aligned} \quad (4.19)$$

where $\bar{b} = b \oplus 1$ is the binary complement of b. The expectation in (4.19) is done according to the transition probability $P_j(y|b)$ in (2.16).

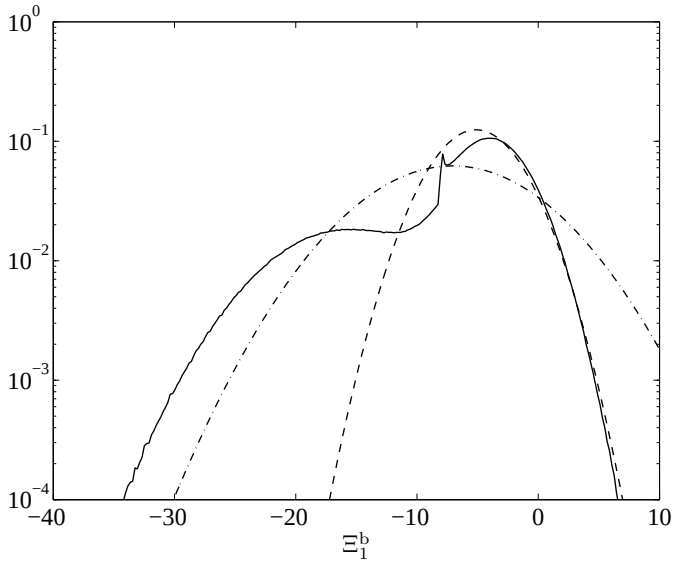
In general, obtaining closed-form expressions of the cumulant transforms can be difficult, and numerical methods are needed. A notable case for which closed-form expressions for $\kappa_1(s)$ exists is BPSK modulation with coherent detection over the AWGN channel with fully-interleaved Nakagami-m_f fading channel (density given in Eq. (2.8)). In that case we have that [75],

$$\kappa_1(s) = -m_f \log\left(1 + \frac{4\mathsf{snr}}{m_f}(s - s^2)\right). \quad (4.20)$$

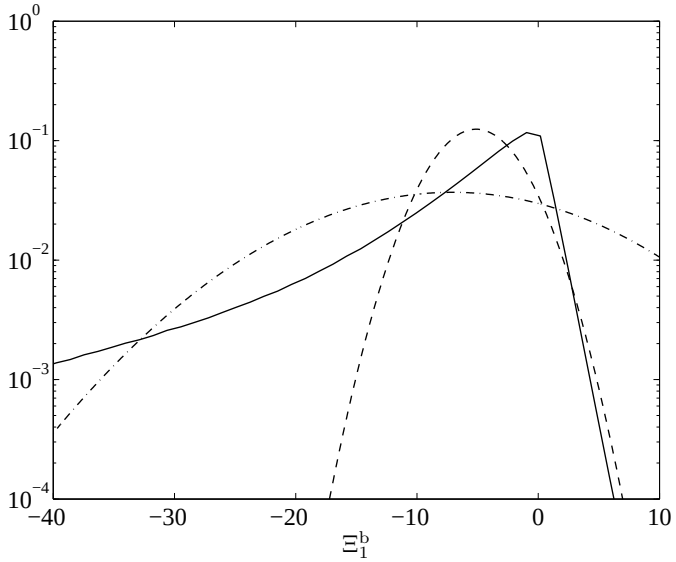
As $m_f \to \infty$, we have that $\kappa_1(s) \to -4\mathsf{snr}(s-s^2)$, which corresponds to the cumulant transform of a Gaussian random variable of mean $-4\mathsf{snr}$ and variance $8\mathsf{snr}$, the density of a log-likelihood ratio in this channel [119]. Another exception, although significantly more complicated than the above, is BICM over the Gaussian and fully-interleaved channels with max-log metric. In this case, as shown in [112] and references therein, the probability density function of Ξ_1^{b} can be computed in closed-form. As we will see, only the cumulant transform is needed to accurately approximate the error probability. Moreover, we can use efficient numerical methods based on Gaussian quadratures to compute it.

As we shall see in the coming sections, the error events that dominate the error probability for large SNR are those that assign multiple different bits to the same symbol, i. e. $v > 1$. For large SNR the curves change slope accordingly. Fortunately, this effect shows in error probability values of interest only for short interleaver lengths, and assuming symbols of weight 1 is sufficient for most practical purposes. Where short lengths are required, random interleaving can cause performance degradation due to these effects and an interleaving must be designed to allocate all different bits in the the smaller distance pairwise errors to different constellation symbols.

Figure 4.1 shows the computer-simulated density of Ξ_1^{b} for BICM using the MAP metric, 16-QAM with Gray mapping in the AWGN and fully-interleaved Rayleigh fading channels, respectively, both with $\mathsf{snr} = 10$ dB. For the sake of comparison, we also show with dashed lines the distribution of a Gaussian random variable with distribution $\mathcal{N}(-4\mathsf{snr}_{\mathrm{eq}}, 8\mathsf{snr}_{\mathrm{eq}})$, where $\mathsf{snr}_{\mathrm{eq}} = -\kappa_1(\hat{s})$ is the equivalent signal-to-noise ratio. This Gaussian approximation is valid in the tail of the distribution, rather than at the mean as would be the case for the standard Gaussian approximation (dash-dotted lines) with the same mean and variance as that of Ξ_1^{b}, i. e. $\mathcal{N}\big(\mathbb{E}[\Xi_1^{\mathrm{b}}], \mathbb{E}[(\Xi_1^{\mathrm{b}})^2] - \mathbb{E}[\Xi_1^{\mathrm{b}}]^2\big)$. For AWGN, the tail of Ξ_1^{b} is close to the Gaussian approximation, and remains similar for Rayleigh fading. Since the tail gives the pairwise error probability, this Gaussian approximation gives a good approximation to the latter. This Gaussian approximation, which will be properly intro-



(a) No fading.



(b) Fully-interleaved Rayleigh fading.

Fig. 4.1 Simulated density of Ξ_1^{b} for BICM with MAP metric, 16-QAM with Gray mapping with $\mathsf{snr} = 10$ dB. The solid lines correspond to the simulated density, dashed lines correspond to the Gaussian approximation $\mathcal{N}(-4\mathsf{snr}_{\mathrm{eq}}, 8\mathsf{snr}_{\mathrm{eq}})$, with $\mathsf{snr}_{\mathrm{eq}} = -\kappa_1(\hat{s})$, and dash-dotted lines correspond to the Gaussian approximation $\mathcal{N}\big(\mathbb{E}[\Xi_1^{\mathrm{b}}], \mathbb{E}[(\Xi_1^{\mathrm{b}})^2] - \mathbb{E}[\Xi_1^{\mathrm{b}}]^2\big)$.

duced in the next section, illustrates the benefits of using the cumulant transform to compute the pairwise error probability.

4.2 Pairwise Error Probability for Infinite Interleaving

In this section, we study the pairwise error probability assuming infinite-length interleaving [29]. As we saw in the previous chapter, this channel model does not fully characterize the fundamental limits of BICM. While the model yields the same capacity, the error exponent is in general different. In this and the following sections, we shall see that this model characterizes fairly accurately the error probability for medium-to-large signal-to-noise ratios when the union bound is employed. However, as shown in Section 4.3, for very large signal-to-noise ratio this model does not capture the overall error behavior, and more general techniques must be sought. Infinite-length interleaving implies that all d different bits in a pairwise error event are mapped onto d different symbols, i. e. there are no symbols with label-Hamming-weight larger than 1. We will first review expressions, bounds and approximations to the pairwise error probability, and we will then apply them to compute the union bound.

4.2.1 Exact Formulas, Bounds, and Approximations

We denote the pairwise error probability for infinite interleaving as $\mathrm{PEP}_1(d)$. We first relate it to the pairwise score Ξ^{pw}. We denote the cumulant transform of the pairwise score by

$$\kappa_{\mathrm{pw}}(s) \stackrel{\Delta}{=} \log \mathbb{E}[\mathrm{e}^{s\,\Xi^{\mathrm{pw}}}]. \tag{4.21}$$

Under the assumption that bit scores are independent and identically distributed, we have that

$$\kappa_{\mathrm{pw}}(s) = d\,\kappa_1(s). \tag{4.22}$$

Since the pairwise error probability is the tail probability of the pairwise score, we have that,

$$\mathrm{PEP}_1(d) = \Pr\left\{\sum_{i=1}^{d} \Xi_i^{\mathrm{b}} > 0\right\} \tag{4.23}$$

where the variables Ξ_i^{b} are bit scores corresponding to the d different bits in the pairwise error.

The above pairwise tail probability can be evaluated by complex-plane integration of the cumulant transform of $\kappa_{\mathrm{pw}}(s) = d\kappa_1(s)$, a method advocated by Biglieri *et al.* [18, 19]. We have that

---

Proposition 4.2. The pairwise error probability $\mathrm{PEP}_1(d)$ is given by

$$\mathrm{PEP}_1(d) = \frac{1}{2\pi j} \int_{s_0-j\infty}^{s_0+j\infty} \frac{1}{s} \, \mathrm{e}^{d\,\kappa_1(s)} \, \mathrm{d}s, \tag{4.24}$$

where $s_0 \in \mathbb{R}$ belongs to region where the cumulant transform $\kappa_1(s)$ is well defined.

In [18], the numerical evaluation of the above complex-plane integration using Gaussian quadrature rules was proposed.

As we next review, the tail probability of the pairwise or the symbol scores is to a large extent determined by the form of the cumulant transform around a special value of s, the saddlepoint $\hat{s}$.

Definition 4.3. The *saddlepoint* $\hat{s}$ is the value of s that makes the first derivative of the cumulant transform, $\kappa'(s)$ equal to zero.

An expression often used is the Chernoff bound, which gives the following upper bound to the pairwise error probability.

Proposition 4.3 (Chernoff Bound). The Chernoff bound to the pairwise error probability is

$$\mathrm{PEP}_1(d) \leq \mathrm{e}^{d\kappa_1(\hat{s})}, \tag{4.25}$$

where $\hat{s}$, the saddlepoint, is the root of the equation $\kappa_1'(s) = 0$.

In general, one needs to find the saddlepoint either analytically or numerically. As we shall see in Section 4.2.2, the choice $\hat{s} = \frac{1}{2}$ is optimum for the MAP BICM metric. We also note that the optimum value of s for the random coding error exponent at vanishing rate (or equivalently at the cutoff rate) is $\hat{s} = \frac{1}{2}$ (for $s = 1/(1+\rho)$ and $\rho = 1$). This

is not surprising since the error exponent analysis and our pairwise error probability are closely related. A somewhat looser version of the Chernoff bound for general metrics is the Bhatthacharyya bound, for which one replaces $\hat{s}$ in (4.25) by $\frac{1}{2}$, as if the saddlepoint was $\hat{s} = \frac{1}{2}$ (we remark that for general metrics and channels $\hat{s} \neq \frac{1}{2}$).

The Chernoff bound gives a true bound and is moreover easy to compute. It is further known to correctly give the asymptotic exponential decay of the error probability for large d and snr [74, 95]. In some cases, such as the AWGN or fully interleaved fading channels, the Chernoff bound is not tight. This looseness can be compensated for in practice by using the saddlepoint approximation.

Theorem 4.1. The pairwise error probability can be approximated to first-order by

$$\mathrm{PEP}_1(d) \simeq \frac{1}{\sqrt{2\pi d \kappa_1''(\hat{s})}\,\hat{s}}\, \mathrm{e}^{d\,\kappa_1(\hat{s})} \tag{4.26}$$

where $\kappa_1''(s)$ is the second derivative of $\kappa(s)$, and $\hat{s}$ is the saddlepoint.

The saddlepoint approximation may also be seen as an approximation of the complex-plane integration of Proposition 4.2. Essentially, the function $\kappa_1(s)$ is approximated by a second-order Taylor expansion at the saddlepoint, neglecting higher order terms, and the resulting integral is explicitly computed.

Alternatively, the saddlepoint approximation extends the Chernoff bound by including a multiplicative coefficient to obtain an expression of the form $\alpha \cdot \mathrm{e}^{d\kappa_1(\hat{s})}$. Higher-order expansions have a correction factor α, polynomial in inverse powers of $\left(d\kappa_1''(\hat{s})\right)^{-1}$. For instance, for the second-order approximation we have

$$\alpha = 1 + \frac{1}{d\kappa_1''(\hat{s})}\left(-\frac{1}{\hat{s}^2} - \frac{\kappa_1'''(\hat{s})}{2\hat{s}\kappa_1''(\hat{s})} + \frac{\kappa_1^{(4)}(\hat{s})}{8\kappa_1''(\hat{s})} - \frac{15}{72}\left(\frac{\kappa_1'''(\hat{s})}{\kappa_1''(\hat{s})}\right)\right)^2. \tag{4.27}$$

In general, and for the purpose of evaluating the error probability of BICM, the effect of the higher-order correction terms is found to be negligible and the formula in Eq. (4.26) gives a good approximation.

For the sake of completeness, we also mention two additional ap-

proximations. First, the Lugannani-Rice formula [68] (see also [78]),

$$\mathrm{PEP}_1(d) \simeq Q\big(\sqrt{-2d\kappa_1(\hat{s})}\big) \tag{4.28}$$

$$+ \frac{1}{\sqrt{2\pi}} \mathrm{e}^{d\kappa_1(\hat{s})} \left(\frac{1}{\sqrt{d\kappa_1''(\hat{s})}\hat{s}} - \frac{1}{\sqrt{-2d\kappa_1(\hat{s})}} \right), \tag{4.29}$$

Finally, we have a Gaussian approximation

$$\mathrm{PEP}_1(d) \simeq Q\left(\sqrt{-2d\kappa(\hat{s})}\right). \tag{4.30}$$

This approximation equals the zero-th order term in the Lugannani-Rice formula and the error probability of a BPSK AWGN channel with equivalent signal-to-noise ratio $\mathsf{snr}_{\mathrm{eq}} = -\kappa(\hat{s})$. This approximation was heuristically introduced in [50], and is precisely the Gaussian approximation depicted in Figure 4.1.

In the following, we particularize the above results to the MAP and max-log metrics in Eqs. (2.15) and (2.17) respectively.

4.2.2 MAP Demodulator

In this section, we study the MAP metric, i. e.

$$q_j\big(b_j(x) = b, y\big) = \sum_{x' \in \mathcal{X}_b^j} P_{Y|X}(y|x'). \tag{4.31}$$

First, we relate the BICM channel with MAP metric to the family of binary-input output-symmetric channels. Recall that the standard definition of BIOS channels starts with the posterior log-likelihood ratio

$$\Lambda = \log \frac{P_{C|Y}(c=1|y)}{P_{C|Y}(c=0|y)}. \tag{4.32}$$

Then, a channel with binary input is said output-symmetric [96] if the following relation between the densities of the posterior log-likelihood ratios, seen as function the channel input, holds:

$$P_{\Lambda|C}(\Lambda|c=1) = P_{\Lambda|C}(-\Lambda|c=0). \tag{4.33}$$

By construction, bit score and posterior log-likelihood ratio coincide when the transmitted bit is zero, i. e. $\Lambda = \Xi_{\mathrm{b}}$. Similarly, we have that $\Lambda = -\Xi_{\mathrm{b}}$ when the transmitted bit is one.

Proposition 4.4. The classical BICM model with infinite interleaving and MAP metric is a binary-input output-symmetric (BIOS) channel.

In addition, we have the following result.

Proposition 4.5. The saddlepoint is located at $\hat{s} = \frac{1}{2}$, i. e. $\kappa_1'\left(\frac{1}{2}\right) = 0$. The following higher-order derivatives of the cumulant transform verify

$$\kappa_1''(\hat{s}) = \frac{\mathbb{E}\left[(\Xi^{\mathrm{b}})^2 \mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}\right]}{\mathbb{E}\left[\mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}\right]} \geq 0, \quad \kappa_1'''(\hat{s}) = \frac{\mathbb{E}\left[(\Xi^{\mathrm{b}})^3 \mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}\right]}{\mathbb{E}\left[\mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}\right]} = 0. \tag{4.34}$$

Proof. See Appendix 4.A. □

Using Definition 4.2 and Theorem 4.1 we can write the saddlepoint approximation to $\mathrm{PEP}_1(d)$ for BICM with infinite interleaving in the fully-interleaved AWGN fading channel:

$$\mathrm{PEP}_1(d) \simeq \frac{2}{\sqrt{2\pi d\, \mathbb{E}\left[\mathrm{e}^{\frac{1}{2}\Xi^{\mathrm{b}}}(\Xi^{\mathrm{b}})^2\right]}} \left(\mathbb{E}[\mathrm{e}^{\frac{1}{2}\Xi^{\mathrm{b}}}]\right)^{d+\frac{1}{2}}, \tag{4.35}$$

where Ξ^{b} is the bit score. In the fully-interleaved AWGN fading channel, we have that

$$\Xi^{\mathrm{b}} = \log \frac{q_j\left(\bar{b}, \sqrt{\mathsf{snr}}hx + z\right)}{q_j\left(b, \sqrt{\mathsf{snr}}hx + z\right)} \tag{4.36}$$

$$= \log \frac{\sum_{x' \in \mathcal{X}_{\bar{b}}^j} \mathrm{e}^{-|\sqrt{\mathsf{snr}}h(x-x')+z|^2}}{\sum_{x' \in \mathcal{X}_b^j} \mathrm{e}^{-|\sqrt{\mathsf{snr}}h(x-x')+z|^2}}. \tag{4.37}$$

The expectation of a generic function $f(\cdot)$ of the bit score Ξ^{b} is done according to

$$\mathbb{E}[f(\Xi^{\mathrm{b}})] = \frac{1}{m2^m} \sum_{b=0}^{1} \sum_{j=1}^{m} \sum_{x \in \mathcal{X}_b^j} \iint f(\Xi^{\mathrm{b}})\, p d_H(h) P_Z(z)\, \mathrm{d}z \mathrm{d}h. \tag{4.38}$$

This expectation can be easily evaluated by numerical integration using the appropriate quadrature rules. Unfortunately, there seems to be no

tractable, simple expression for the final result. The only case in which (4.35) admits a closed form is the case of binary BPSK modulation with Nakagami-m_f fading for which using (4.20) we obtain [75],

$$\mathrm{PEP}_1(d) \simeq \frac{1}{2\sqrt{\pi d\,\mathsf{snr}}}\left(1+\frac{\mathsf{snr}}{m_f}\right)^{-m_f d+\frac{1}{2}}. \tag{4.39}$$

The exact expression of $\mathrm{PEP}_1(d)$ from Proposition 4.2 or the Chernoff bound in Proposition 4.3 admits a form similar to (4.35). In particular, since in this case $\hat{s}=\frac{1}{2}$, the Chernoff and Bhattacharyya bounds coincide. This Bhattacharyya union bound was first proposed in [29].

We next compare the estimate of $\mathrm{PEP}_1(d)$ with the BICM *expurgated bound* of [29]. The expurgated bound can be seen as the positive tail probability of the random variable with sample value

$$\log\frac{P_{Y|X}(y|\hat{x})}{P_{Y|X}(y|x)}, \tag{4.40}$$

where x is the transmitted symbol and $\hat{x}$ is its nearest neighbour in $\mathcal{X}_{\bar{b}}^{j}$, i. e., with complementary bit $\bar{b}$ in label index j. Compared with the MAP metric there is only one term in each summation, rather than the full set $\mathcal{X}_{b}^{j}$. For some metrics and/or labelings, Eq. (4.40) may not be accurate. For example, for the set-partitioning mapping considered in [29], the bound was not close to the simulation results. This inaccuracy was solved in [74] by using the saddlepoint approximation with the full MAP metric given in (4.35). The inaccuracy of the expurgated bound was also remarked by Sethuraman [104] and Yeh *al.* [141] who noticed that this "bound" is actually not a bound in general. A further point is discussed in Section 4.3.4, where we analyze the effect of considering only Hamming weight 1.

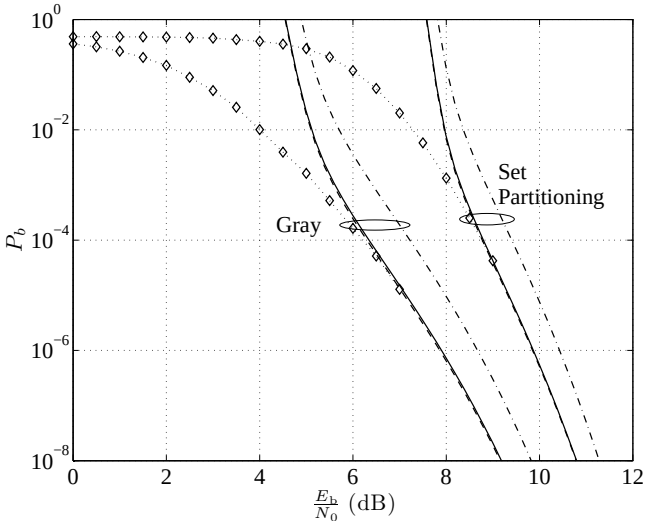
We next show some examples to illustrate the accuracy of these bounds and approximations for convolutional and repeat-accumulate (RA) codes [32]. In particular, we show the Chernoff/Bhattacharyya union bound (dash-dotted lines), the saddlepoint approximation (4.35) union bound (solid lines), the Gaussian approximation union bound (dashed lines) and the simulations. In the case of RA codes, we use the uniform interleaver [8,9] and an iterative decoder with 20 decoding iter-

ations. The assumption of uniform interleaver allows for the analytical calculation of the weight enumerator coefficients (see [32]).

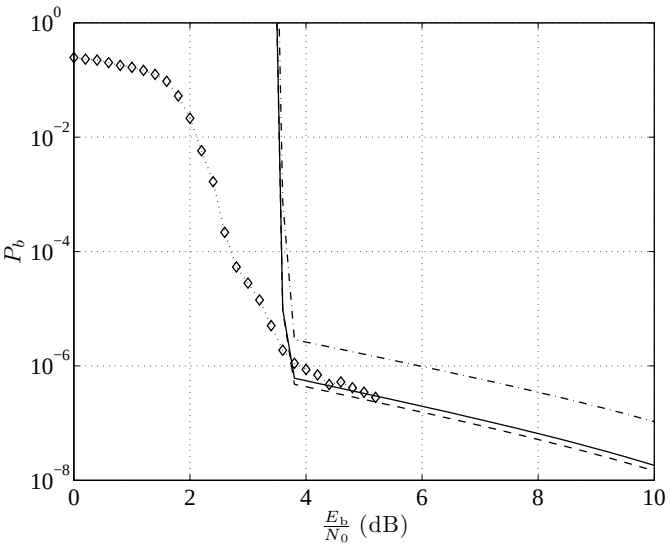
Figure 4.2 shows the bit error rate as a function of $\frac{E_b}{N_0}$ for the aforementioned methods with 16-QAM in the AWGN channel with no fading. In Figure 4.2(a) we use the optimum 64-state and rate-1/2 convolutional code with Gray and set partitioning mappings and in Figure 4.2(b) an RA code of rate 1/4 with Gray mapping. As we observe, the Chernoff/Bhattacharyya bound is loose when compared with the saddlepoint or Gaussian approximations, which are close to the actual simulations for large $\frac{E_b}{N_0}$. We notice that the saddlepoint and Gaussian approximations are close to each other for the convolutional code. However, for RA codes, the Gaussian approximation yields a slightly optimistic estimate of the error floor region. As we will see in the following examples, this effect is more visible in fading channels.

Figure 4.3 shows the estimates of the bit error rate for convolutional and RA codes respectively in a fully-interleaved AWGN channel with Rayleigh fading. Figure 4.3(a) shows two cases, a rate-2/3, 8-state code over 8-PSK, and the rate-1/2, 64-state code over 16-QAM both with Gray mapping (both codes have largest Hamming distance). Figure 4.3(b) shows the performance of a RA code of rate 1/4 with Gray mapping and 16-QAM modulation. All approximations are close to the simulated value, but now only the saddlepoint approximation gives an accurate estimate of the error probability. As we saw in Section 4.1.2, and in Figure 4.1(b) to be more precise, the tail of the bit score in the fully-interleaved Rayleigh fading channel is approximately exponential, rather than Gaussian, and this shape is not accurately tracked by the Gaussian approximation. As evidenced by the results of 16-QAM with the 64-state convolutional code, this effect becomes less apparent for codes with large minimum distance, since the pairwise score contains more terms and its tail is closer to a Gaussian. For RA codes, where the minimum distance is low, we appreciate the differences between saddlepoint and Gaussian approximations.

In the following section, we discuss in more detail the slope of decay of the error probability for large snr, by studying the asymptotic behavior of the cumulant transform and its second derivative evaluated at the saddlepoint. This analysis reveals that BICM mimics the behavior

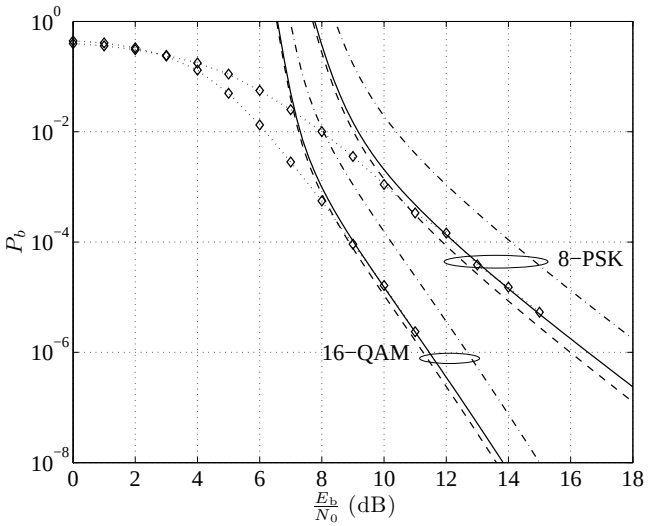


(a) 64-state, rate-1/2 convolutional code with generators $(133, 171)_8$

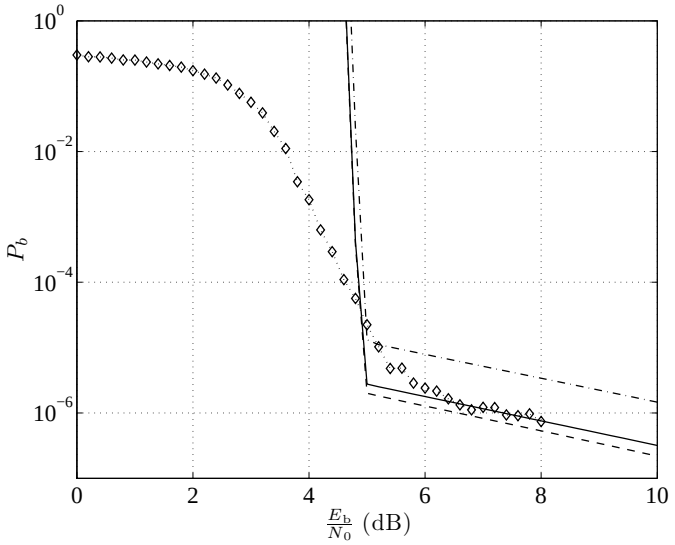


(b) RA code of rate 1/4 with $K = 1024$ information bits

Fig. 4.2 Comparison of simulation results (dotted lines and markers), saddlepoint (in solid lines) and Gaussian approximations (dashed lines), and Chernoff/Bhattacharyya union bound (in dash-dotted lines) to the bit error rate of BICM with 16-QAM modulation with Gray and Set Partitioning mapping, in the AWGN channel.



(a) 64-state, rate-1/2 convolutional code with generators $(133, 171)_8$ over 16-QAM modulation with Gray labeling and optimum 8-state rate-2/3 convolutional code with 8-PSK and Gray labeling



(b) RA code of rate 1/4 with $K = 512$ information bits over 16-QAM modulation with Gray mapping

Fig. 4.3 Comparison of simulation results (dotted lines and markers), saddlepoint (in solid lines) and Gaussian approximations (dashed lines), and Chernoff union bound (in dash-dotted lines) to the bit error rate of BICM in the fully-interleaved Rayleigh fading channel.

of binary modulation, preserving the properties of the binary code.

4.2.3 Cumulant transform asymptotic analysis

Inspection of Figures 4.2 and 4.3 suggests that the bounds and approximations considered in the previous section yield the same asymptotic behavior of the error probability for large snr. Indeed, BICM mimics the behavior of binary modulation, preserving the properties of the binary code:

Proposition 4.6 ([74]). The cumulant transform of the bit scores of BICM transmitted over the AWGN with no fading and its second derivative have the following limits for large snr

$$\lim_{\mathsf{snr}\to\infty} \frac{\kappa_1(\hat{s})}{\mathsf{snr}} = -\frac{d^2_{\mathcal{X},\min}}{4} \tag{4.41}$$

$$\lim_{\mathsf{snr}\to\infty} \frac{\kappa_1''(\hat{s})}{\mathsf{snr}} = 2d^2_{\mathcal{X},\min}, \tag{4.42}$$

where $d^2_{\mathcal{X},\min} \stackrel{\Delta}{=} \min_{x,x'} |x - x'|^2$ is the minimum squared Euclidean distance of the signal constellation $\mathcal{X}$.

For large snr, BICM behaves like a binary modulation with distance $d^2_{\mathcal{X},\min}$, regardless of the mapping. This result confirms that BICM preserves the properties of the underlying binary code $\mathcal{C}$, and that for large snr the error probability decays exponentially with snr as

$$\mathrm{PEP}_1(d) \simeq \mathcal{K}\, \mathrm{e}^{-\frac{1}{4}\mathsf{snr}\, d^2_{\mathcal{X},\min}} \tag{4.43}$$

Here $\mathcal{K}$ may depend on the mapping (see for example Figure 4.2(a)), but the exponent is not affected by it. To illustrate this point, Figure 4.4 shows $-\frac{\kappa(\hat{s})}{\mathsf{snr}}$ (thick lines) and $\frac{\kappa''(\hat{s})}{\mathsf{snr}}$ (thin lines) for 16-QAM with Gray (solid lines) and set partitioning (dashed lines) mappings in the AWGN channel. As predicted by Proposition 4.6, the asymptotic value of $-\frac{\kappa(\hat{s})}{\mathsf{snr}}$ is $-\frac{1}{4}d^2_{\mathcal{X},\min} = 0.1$ (recall that we consider signal constellations $\mathcal{X}$ normalized in energy, and that $d^2_{\mathcal{X},\min} = 0.4$ for 16-QAM). In the Gaussian approximation introduced in [50], the quantity $-\frac{\kappa(\hat{s})}{\mathsf{snr}}$ can be interpreted as a scaling in snr due to BICM; the asymptotic

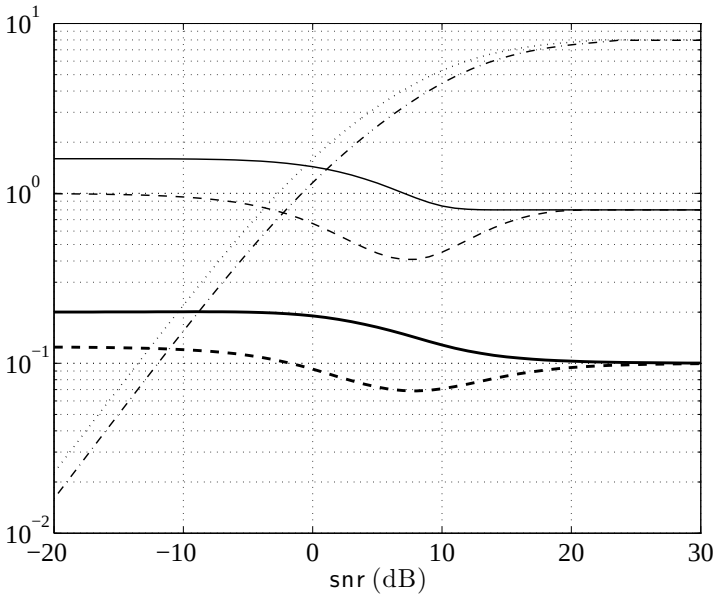


Fig. 4.4 Cumulant transform limits in the AWGN channel $-\frac{\kappa(\hat{s})}{\mathsf{snr}}$ (thick lines) and $\frac{\kappa''(\hat{s})}{\mathsf{snr}}$ (thin lines) for 16-QAM with Gray (solid lines) and set partitioning (dashed lines) mappings. For fully-interleaved Rayleigh fading channel, in dash-dotted lines $\kappa''(\hat{s})$ for 16-QAM and in dotted lines for 8-PSK.

scaling depends only on the signal constellation, but not on the binary labeling. Figure 4.4 also shows $\frac{\kappa''(\hat{s})}{\mathsf{snr}}$ for the same setup (thin lines). Again, the limit coincides with (4.42).

In the case of Nakagami-m_f fading (density in Eq. (2.8)), we have a similar result. All curves in Figures 4.3(a) and 4.3(b) show an asymptotic slope of decay, which is characterized by the following proposition.

Proposition 4.7 ([74]). The cumulant transform of BICM transmitted over the AWGN with fully-interleaved Nakagami-m_f fading and its second derivative have the following limits for large snr

$$\lim_{\mathsf{snr}\to\infty} \frac{\kappa_1(\hat{s})}{\log \mathsf{snr}} = -m_f \tag{4.44}$$

$$\lim_{\mathsf{snr}\to\infty} \kappa_1''(\hat{s}) = 8m_f. \tag{4.45}$$

The proof of this result is a straightforward extension of the proof in [74] for Rayleigh fading. The above result does not depend on the modulation nor the binary labeling, and confirms that BICM indeed behaves as a binary modulation and thus, the asymptotic performance depends on the Hamming distance of the the binary code $\mathcal{C}$ rather than on the Euclidean distance. Figure 4.4 also shows $\kappa''(\hat{s})$ for 16-QAM (dash-dotted line) and 8-PSK (dotted line), both with Gray labeling in the Rayleigh channel. As expected, the limit value is 8, which does not depend on the modulation.

A finer approximation to the exponent of the error probability is given by the the following result.

Proposition 4.8. The cumulant transform of the BICM bit score for the AWGN channel with fully-interleaved Nakagami-m_f fading behaves for large snr as

$$\lim_{\mathsf{snr}\to\infty} \frac{\mathrm{e}^{\kappa_1(\hat{s})}}{\mathsf{snr}^{-m_f}} = \mathbb{E}_{X,J}\left[\frac{4m_f}{d^2(x,\hat{x})}\right]^{m_f} \tag{4.46}$$

where $\hat{x}$ is the closest symbol in the constellation $\mathcal{X}_{\bar{b}}^{j}$ to x, $d^2(x,\hat{x}) = |x-\hat{x}|^2$ is the Euclidean distance between two symbols and the expectation $\mathbb{E}_{X,J}[\cdot]$ is with respect to X uniform over $\mathcal{X}$ and J uniform over the the label bit index $j = 1,\ldots,m$.

Proof. See Appendix 4.B. □

The above expectation over the symbols is proportional to a form of harmonic distance d_{h}^2,

$$\frac{1}{d_{\mathrm{h}}^2} = \left(\frac{1}{m\,2^m}\sum_{b=0}^{1}\sum_{j=1}^{m}\sum_{x\in\mathcal{X}_b^j}\left(\frac{1}{d^2(x,x')}\right)^{m_f}\right)^{\frac{1}{m_f}}, \tag{4.47}$$

where x' is the closest symbol in the constellation $\mathcal{X}_{\bar{b}}^{j}$ to x and $d^2(x,x')$ is the Euclidean distance between two symbols.

The asymptotic results given by Propositions 4.7 and 4.8 can be combined with the saddlepoint approximation (4.35) to produce a

heuristic approximation of $\mathrm{PEP}_1(d)$ for large snr as,

$$\mathrm{PEP}_1(d) \simeq \frac{1}{2\sqrt{\pi d\, m_f}} \left(\frac{4\, m_f}{d_{\mathrm{h}}^2 \mathsf{snr}} \right)^{d m_f}. \tag{4.48}$$

A similar harmonic distance was used by Caire *et al.* [29] to approximate the Chernoff bound for Rician fading.

4.2.4 Numerical Results with Max-Log Metric

In this section, we study the error probability of BICM with the max-log metric in Eq. (2.17),

$$q_j\big(b_j(x) = b, y\big) = \max_{x' \in \mathcal{X}_b^j} P_{Y|X}(y|x'). \tag{4.49}$$

As for the MAP metric, the saddlepoint approximation with infinite interleaving in the fully-interleaved AWGN fading channel is given by

$$\mathrm{PEP}_1(d) \simeq \frac{2}{\sqrt{2\pi d\, \mathbb{E}\left[\mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}(\Xi^{\mathrm{b}})^2\right]}} \left(\mathbb{E}[\mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}]\right)^{d+\frac{1}{2}}, \tag{4.50}$$

where Ξ^{b} is the bit score, in turn given by

$$\Xi^{\mathrm{b}} = \min_{x' \in \mathcal{X}_b^j} |\sqrt{\mathsf{snr}} h(x - x') + z|^2 - \min_{x' \in \mathcal{X}_{\bar{b}}^j} |\sqrt{\mathsf{snr}} h(x - x') + z|^2. \tag{4.51}$$

The saddlepoint $\hat{s}$ now needs to be found numerically, since in general, it will be different from that of the MAP metric, i. e. $\frac{1}{2}$. To any extent, as $\mathsf{snr} \to \infty$ the saddlepoint approaches $\frac{1}{2}$, since the MAP and the max-log metrics become indistinguishable. This result builds on the proofs of Propositions 4.6 and 4.7 [74], where only the dominant term in the MAP metric is kept for large snr.

For this metric, Sczeszinski *et al.* [112] have derived closed expressions for the density of the bit score Ξ^{b}. This expression allows for the calculation of the cumulant transform, however, it does not allow for the direct evaluation of $\mathrm{PEP}_1(d)$ and numerical methods are needed to either perform the numerical integration in (4.24) or compute the saddlepoint, and hence the saddlepoint approximation (4.50).

Figure 4.5 shows the comparison between MAP and max-log metrics for BICM with an RA code and 16-QAM. As we observe, the performance is with the max-log metric is marginally worse than that with

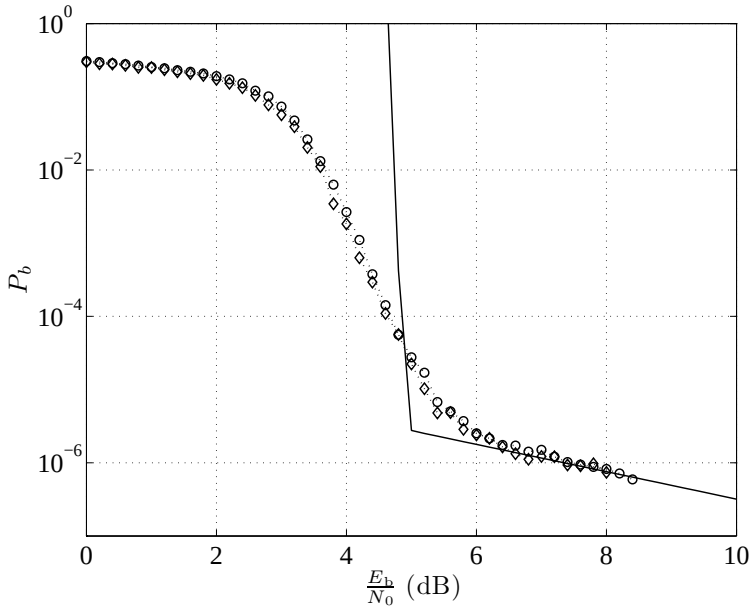


Fig. 4.5 Comparison of simulation results and saddlepoint approximations on the bit error rate of BICM with a rate-1/4 RA code with $K = 512$ information bits, 16-QAM modulation with Gray mapping, in the fully-interleaved Rayleigh fading channel using MAP (diamonds) and max-log (circles) metrics. In solid lines, saddlepoint approximation union bound.

the MAP metric. Note that we are using perfectly coherent detection. It should be remarked, though, that more significant penalties might be observed for larger constellations, in presence of channel estimation errors, or in non-coherent detection (see Chapter 6).

4.3 Pairwise Error Probability for Finite Interleaving

4.3.1 Motivation

So far we have assumed, as it was done in [29], that the bit interleaver has infinite length. For this case, all symbol scores have Hamming weight 1 and are thus bit scores. Moreover, since the channel is memoryless, the bit scores with infinite interleaving are independent. For finite interleaving, however, some of the d bits in which the two codewords in the pairwise error event differ belong to the same symbol with nonzero probability. Therefore, they are affected by the same realiza-

tion of the channel noise and possibly fading, and are thus statistically dependent. As we mentioned in Section 4.1, a general characterization of the error probability requires the whole set of possible symbol scores Ξ^{s}, for all the valid Hamming weights v, $1 \leq v \leq \min(m, d)$.

Since the task of determining the exact distribution of the d pairwise different bits onto the N symbols can be hard, we follow the results of [141] and compute an average pairwise error probability by averaging over all possible distributions of d bits onto N symbols, equivalent to uniform interleaving for turbo-codes [8, 9]. In Section 4.3.2 we present a general expression for the pairwise error probability, as well as its corresponding saddlepoint approximation.

In Section 4.3.3 we apply the theory to what is arguably the simplest case of BICM, QPSK under Nakagami fading. We discuss the extent to which QPSK can be modeled by two independent BPSK modulations each with half signal-to-noise ratio. We show that the slope of the pairwise error probability is reduced at sufficiently high snr due to the symbol scores of Hamming weight 2. An estimate of the snr at which this "floor" appears is given. Finally, we extend the main elements of the QPSK analysis to higher-order modulations in Section 4.3.4, where we discuss the presence of a floor in the error probabilities and give a method to estimate the snr at which it appears.

4.3.2 A General Formula for the Pairwise Error Probability

In this section, we closely follow Yeh *et al.* [141] to give a general characterization of the error probability. We distinguish the possible ways of allocating d bits on N modulation symbols, by counting the number of symbols with weight v, where $0 \leq v \leq w^\star$ and

$$w^\star \triangleq \min(m, d). \tag{4.52}$$

Here m is the number of bits in the binary label of a modulation symbol. Denoting the number of symbols of weight v by N_v, we have that

$$\sum_{v=0}^{w^\star} N_v = N \quad \text{and} \quad d = \sum_{v=1}^{w^\star} v N_v. \tag{4.53}$$

We further denote the symbol pattern by $\rho_N \triangleq (N_0, \ldots, N_{w^\star})$. For finite interleaving, every possible pattern corresponds to a (possibly) different conditional pairwise error probability, denoted by $\mathrm{PEP}(d, \rho_N)$. Then, we can write the following union bound

$$P_e \leq \sum_d \sum_{\rho_N} A_{d,\rho_N} \mathrm{PEP}(d, \rho_N) \tag{4.54}$$

where A_{d,ρ_N} is the number of codewords of $\mathcal{C}$ of Hamming weight d that are mapped onto the constellation symbols following patter ρ_N, i. e. mapped onto N_v symbols of Hamming weight v, for $0 \leq v \leq w^\star$. In some cases where the exact location of the d different bits is difficult to obtain, it might be interesting and convenient to study the average performance over all possible ways of choosing d locations in a codeword. In this case we have the following union bound

$$P_e \leq \sum_d A_d \sum_{\rho_N} P(\rho_N) \mathrm{PEP}(d, \rho_N) \tag{4.55}$$

$$= \sum_d A_d \mathrm{PEP}(d) \tag{4.56}$$

where $P(\rho_N)$ is the probability of a particular pattern ρ_N and $\mathrm{PEP}(d) \triangleq \sum_{\rho_N} P(\rho_N)\mathrm{PEP}(d, \rho_N)$ is the average pairwise error probability over all possible patterns. A counting argument [141] gives the probability of the pattern ρ_N as

$$P(\rho_N) \triangleq \Pr\big(\rho_N = (N_0, \ldots, N_{w^\star})\big) \tag{4.57}$$

$$= \frac{\binom{m}{1}^{N_1}\binom{m}{2}^{N_2}\cdots\binom{m}{w^\star}^{N_{w^\star}}}{\binom{mN}{d}} \frac{N!}{N_0!N_1!N_2!\ldots N_{w^\star}!}. \tag{4.58}$$

In general, the pairwise score depends on the pattern ρ_N; accordingly, we denote the cumulant transform of the symbol score by $\kappa_{\mathrm{pw}}(s, \rho_N)$, to make the dependence on ρ_N explicit. We have that

$$\kappa_{\mathrm{pw}}(s, \rho_N) = \sum_{v=1}^{w^\star} N_v \kappa_v(s), \tag{4.59}$$

since the symbol scores are independent. For infinite interleaving, we have $\rho_N^* = (N-d, d, 0, \ldots, 0)$, and we therefore recover the result shown

in the previous section, i. e. $\kappa_{\mathrm{pw}}(s, \rho_N^*) = d\kappa_1(s)$. When $N \to \infty$ and $m > 1$ the probability that the bit scores are dependent tends to zero, but does so relatively slowly, as proved by the following result.

Proposition 4.9 ([73]). The probability that all d bits are independent, i. e. the probability of $\rho_N = (N-d, d, 0, \dots)$ is

$$P_{\mathrm{ind}} \triangleq \Pr\big(\rho_N = (N-d, d, 0, \dots)\big) = \frac{m^d \binom{N/m}{d}}{\binom{N}{d}}. \tag{4.60}$$

Further, for large N, we have that

$$P_{\mathrm{ind}} \simeq \mathrm{e}^{-\frac{d(d-1)(m-1)}{2N}} \simeq 1 - \frac{d(d-1)(m-1)}{2N} \tag{4.61}$$

where the second approximation in (4.61) assumes that $N \gg d$.

For BPSK, or $m = 1$, there is no dependence, as it should be. However, for $m > 1$ the probability of having symbol scores independent is nonzero. As we shall see in the following, this effect will induce a change in the slope of the error probability. It is important to remark that this is an average result over the ensemble of all possible interleavers.

The conditional pairwise error probability is of fundamental importance for this analysis, and we devote the rest of the section to its study. In particular we have a result analogous to Proposition 4.2 [141].

Proposition 4.10. The conditional pairwise error probability for a particular pattern ρ_N can be computed as

$$\mathrm{PEP}(d, \rho_N) = \frac{1}{2\pi j} \int_{s_0 - j\infty}^{s_0 + j\infty} \frac{1}{s} \, \mathrm{e}^{\kappa_{\mathrm{pw}}(s, \rho_N)} \, \mathrm{d}s \tag{4.62}$$

where $s_0 \in \mathbb{R}$ belongs to region where the cumulant transforms $\kappa_v(s)$ are well defined.

Sethuraman *et al.* [104] pointed out that Caire's union bound for BICM (which assumes infinite interleaving, that is our $\mathrm{PEP}_1(d)$), was not a true bound. Taking into account all possible patterns ρ_N solves this inconsistency and yields a true bound to the average pairwise error probability with a uniform interleaver.

A looser bound given by the following result.

Proposition 4.11 (Chernoff Bound). The Chernoff bound to the conditional pairwise error probability is

$$\mathrm{PEP}(d, \rho_N) \leq \mathrm{e}^{d\kappa_{\mathrm{pw}}(\hat{s}, \rho_N)}. \tag{4.63}$$

where $\hat{s}$, the saddlepoint, is the root of the equation $\kappa'_{\mathrm{pw}}(s, \rho_N) = 0$.

Again, we can use the saddlepoint approximation to obtain a result similar to that shown in Theorem 4.1.

Theorem 4.2. The conditional pairwise error probability can be approximated to first-order by

$$\mathrm{PEP}(d, \rho_N) \simeq \frac{1}{\sqrt{2\pi\kappa''_{\mathrm{pw}}(\hat{s}, \rho_N)\hat{s}}}\, \mathrm{e}^{\kappa_{\mathrm{pw}}(\hat{s}, \rho_N)}. \tag{4.64}$$

The saddlepoint $\hat{s}$ depends in general on the specific pattern ρ_N.

In the following, we particularize the above analysis to QPSK with Gray labeling in the fully-interleaved Nakagami-m_f channel. This is possibly the simplest case of dependency between the bit sub-channels, with symbol scores of Hamming weight 2.

4.3.3 QPSK with Gray Labeling as BICM

In this section, we analyze QPSK in fully interleaved Nakagami-m_f fading (for the density, see Eq. (2.8)) for finite block lengths. The saddlepoint approximation was evaluated in [73].

Theorem 4.3 ([73]). The saddlepoint approximation corresponding to the average pairwise error probability $\mathrm{PEP}(d)$ of QPSK with codeword length N in Nakagami-m_f fading is

$$\mathrm{PEP}(d) \simeq \sum_{N_1, N_2} P(\rho_N) \frac{\left(1 + \frac{\mathsf{snr}}{2m_f}\right)^{-m_f N_1} \left(1 + \frac{\mathsf{snr}}{m_f}\right)^{-m_f N_2}}{\sqrt{2\pi\left(N_1 \frac{\mathsf{snr}}{1 + \frac{\mathsf{snr}}{2m_f}} + N_2 \frac{2\mathsf{snr}}{1 + \frac{\mathsf{snr}}{m_f}}\right)}} \tag{4.65}$$

where

$$P(\rho_N) = \frac{2^{N_1}}{\binom{2N}{d}} \frac{N!}{(N-d+N_2)!(d-2N_2)!N_2!} \tag{4.66}$$

and $\max(0, d-N) \le N_2 \le \lfloor d/2 \rfloor$, $N_1 + 2N_2 = d$.

For the AWGN channel, as $m_f \to \infty$, we have that

$$\mathrm{PEP}(d) \simeq \sum_{N_1,N_2} P(\rho_N) \frac{\mathrm{e}^{-\frac{\mathsf{snr}}{2}(N_1+2N_2)}}{\sqrt{2\pi \mathsf{snr}(N_1+2N_2)}} \tag{4.67}$$

$$= \frac{\mathrm{e}^{-d\frac{\mathsf{snr}}{2}}}{\sqrt{2\pi d \mathsf{snr}}}, \tag{4.68}$$

the usual exponential approximation to the error probability of BPSK at half signal-to-noise ratio. In the presence of fading, however, the symbols of Hamming weight 2 have an important effect. The slope of the pairwise error probability changes at sufficiently large signal-to-noise ratio. The approximate signal-to-noise ratio at which the error probability changes slope, denoted by $\mathsf{snr}_{\mathrm{th}}$, was estimated in [73],

$$\mathsf{snr}_{\mathrm{th}} = \begin{cases} 4m \left(\frac{2Ne}{8^{\frac{1}{d}} d} \right)^{\frac{1}{m}} & d \text{ even} \\ 4m(2Ne)^{\frac{1}{m}} \left(\frac{e}{2}\right)^{\frac{2}{m(d-1)}} \frac{(d-1)^{\frac{d}{m(d-1)}} (d+1)^{\frac{1}{m(d-1)}}}{d^{\frac{2(d+1)}{m(d-1)}}} & d \text{ odd.} \end{cases} \tag{4.69}$$

A coarse approximation to the bit error probability at the threshold is derived by keeping only the term at minimum distance d and the two summands with $N_2 = 0$ and $N_2 = \lfloor \frac{d}{2} \rfloor$. We have then

$$P_b \simeq 2A'_d \mathrm{PEP}(d) \simeq A'_d \frac{1}{\sqrt{\pi m_f d}} \left(\frac{\mathsf{snr}_{\mathrm{th}}}{2m} \right)^{-m_f d}. \tag{4.70}$$

Figure 4.6 depicts the bit error rate of QPSK with the $(5,7)_8$ convolutional code for $n = 40,200$ ($N = 20,100$) and $m_f = 0.5$. In all cases, the saddlepoint approximation to the union bound (solid lines) is accurate for large SNR. In particular, we observe the change in slope with respect to the union bound using $N_2 = \max\{0, d-N\}$ (the upper one corresponds to $n = 200$), i. e. assuming that all bits of the different codewords are mapped over different symbols (independent binary

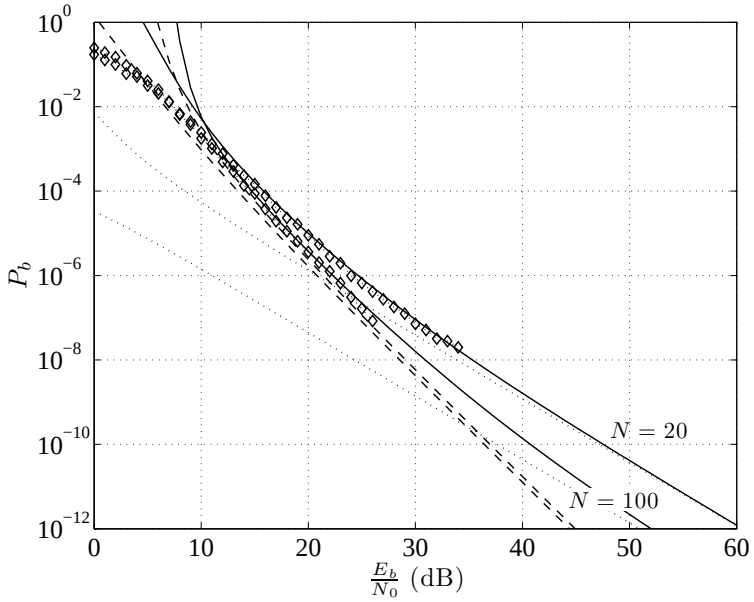


Fig. 4.6 Bit error rate of the $(5,7)_8$ convolutional code over QPSK in a fully interleaved fading channel with $m_f = 0.5$; uniform interleaver of length $n = 40,200$ ($N = 20,100$). Diamonds for the simulation, solid lines for the union bound, dashed lines for the union bound assuming $N_2 = N_2^{\min} = \max\{0, d-N\}$ (the upper one corresponds to $N = 100$) and dotted lines for the union bound for $N_2 = \lfloor \frac{d}{2} \rfloor$.

channels). The approximated thresholds (4.69) computed with the minimum distance, namely $\mathsf{snr}_{\text{th}} = 22$ dB for $N = 20$ and $\mathsf{snr}_{\text{th}} = 36$ dB for $N = 100$, are very close to the points where dashed and dotted lines cross. In the next section we generalize this result to higher order modulations.

Figure 4.7 depicts the approximate value of P_b (assuming $A'_d = 1$) at the threshold signal-to-noise ratio for several values of m_f and ℓ as a function of the minimum Hamming distance d. The error probability at the crossing rapidly becomes small, at values typically below the operating point of common communication systems.

4.3.4 High-order Modulations: Asymptotic Analysis

In this section, we closely follow the analysis in [73] for general constellations and mappings, and estimate the signal-to-noise ratio at which

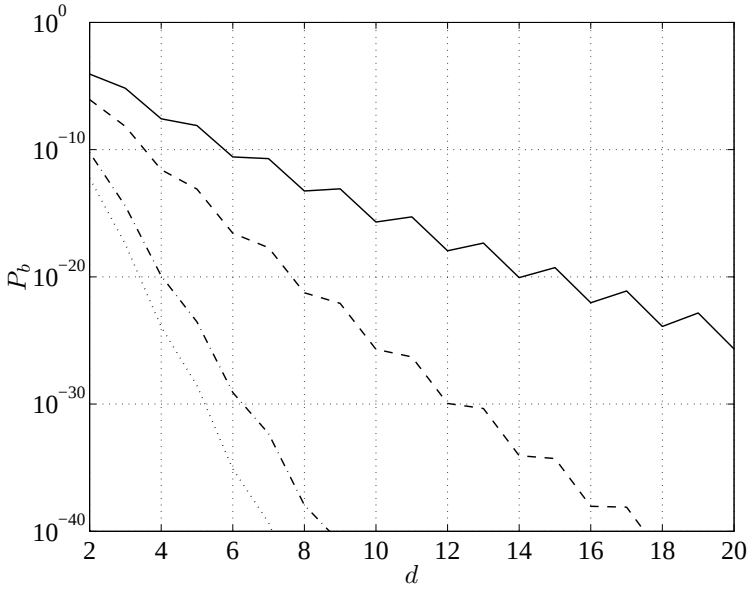


Fig. 4.7 Bit error probability of QPSK at the threshold signal-to-noise ratio $\mathsf{snr}_{\mathrm{th}}$ according to (4.70) as a function of the minimum Hamming distance d for several values of m_f and n. The solid line corresponds to $m_f = 0.5, n = 100$, the dashed line to $m_f = 0.5, n = 1000$, the dash-dotted to $m_f = 3, n = 100$ and the dotted to $m_f = 3, n = 1000$.

the slope of the error probability changes.

First, and in a similar vein to the asymptotic analysis presented in Section 4.2.3 for $v = 1$, for sufficiently large snr, we can approximate the Chernoff bound to the error probability in the AWGN channel by

$$\mathrm{e}^{\kappa_v(\hat{s})} \simeq \mathrm{e}^{-\frac{1}{4} d_{\mathrm{m}}^2(v) \mathsf{snr}}, \tag{4.71}$$

where $d_{\mathrm{m}}^2(v)$ is given by

$$d_{\mathrm{m}}^2(v) \triangleq \min_{x \in \mathcal{X}} \frac{\left(\sum_{i=1}^{v} |x - x_i'|^2\right)^2}{\left|\sum_{i=1}^{v} (x - x_i')\right|^2}, \quad x_i' \triangleq \underset{x' \in \mathcal{X}_{b_i(x)}^i}{\arg\min} |x - x'|^2. \tag{4.72}$$

For $v = 1$, we recover $d_{\mathrm{m}}^2(1) = d_{\mathcal{X},\min}^2$.

For the fully-interleaved Nakagami-m_f fading channel, we have

$$\mathrm{e}^{\kappa_v(\hat{s})} \simeq \left(d_{\mathrm{h}}^2(v) \frac{\mathsf{snr}}{4 m_f} \right)^{-m_f}, \tag{4.73}$$

where $d_{\mathrm{h}}^2(v)$ is a generalization of the harmonic distance given by

$$\frac{1}{d_{\mathrm{h}}^2(v)} = \left(\frac{1}{\binom{m}{v} 2^m} \sum_{\boldsymbol{b}} \sum_{\boldsymbol{j}} \sum_{x \in \mathcal{X}_{\boldsymbol{j}}^{\boldsymbol{b}}} \left(\frac{\left| \sum_{i=1}^{v} (x - x_i') \right|^2}{\left(\sum_{i=1}^{v} |x - x_i'|^2 \right)^2} \right)^{m_f} \right)^{\frac{1}{m_f}}. \quad (4.74)$$

For a given x, x_i' is the i-th symbol in the sequence of v symbols $(x_1', \ldots, x_v')$ which have binary label $\bar{c}_{j_i}$ at position j_i and for which the ratio $\frac{\left| \sum_{i=1}^{v} (x - x_i') \right|}{\sum_{i=1}^{v} |x - x_i'|^2}$ is minimum among all possible such sequences. For $m_f = 1$, $v = 1$ we recover the harmonic distance d_{h}^2 in Section 4.2.3.

As it happened with the bit score and $\mathrm{PEP}_1(d)$, Eq. (4.73) may be in the saddlepoint approximation to obtain a heuristic approximation to the pairwise error probability for large snr, namely

$$\mathrm{PEP}(d, \rho_n) \simeq \mathrm{PEP}_{\mathrm{H}}(d, \rho_n) \stackrel{\Delta}{=} \frac{1}{2\sqrt{\pi m_f \sum_{v \geq 1} n_v}} \prod_{v=1}^{m} \left(\frac{4 m_f}{d_{\mathrm{h}}^2(v)} \frac{1}{\mathsf{snr}} \right)^{n_v m_f}. \quad (4.75)$$

We use Eq. (4.75) to estimate the threshold SNR. For the sake of simplicity, consider $d \geq m$, that is $w^\star = m$ in our definitions at the beginning of this Section. Then, since $d = \sum_{v=1}^{m} N_v v$ by construction, we can view the pattern $\rho_N = (N_0, N_1, \ldots, N_m)$ as a (non-unique) representation of the integer d as a weighted sum of the integers $\{0, 1, 2, \ldots, m\}$. By construction, the sum $\sum_v N_v$ is the number of non-zero Hamming weight symbols in the candidate codeword. The lowest value of $\sum_v N_v$ gives the worst (flattest) pairwise error probability in the presence of fading. As found in [73], a good approximation to the threshold is given by

$$\mathsf{snr}_{\mathrm{th}} \simeq 4 m_f \left(\sum_{\rho_n : \min \sum_v n_v} \frac{P(\rho_n) \left(d_{\mathrm{h}}^2(1) \right)^d}{P(\rho_0) \prod_v \left(d_{\mathrm{h}}^2(v) \right)^{n_v}} \right)^{-\frac{1}{m_f (d - \sum_v n_v)}}. \quad (4.76)$$

For QPSK with Gray mapping, computation of this value of $\mathsf{snr}_{\mathrm{th}}$ ($d_{\mathrm{h}}^2(1) = 2$, $d_{\mathrm{h}}^2(2) = 4$) gives a result which is consistent with the result derived in Section 4.3.3, namely Eq. (4.69), with the minor difference that we use now the Chernoff bound whereas the saddlepoint approximation was used in the QPSK case.

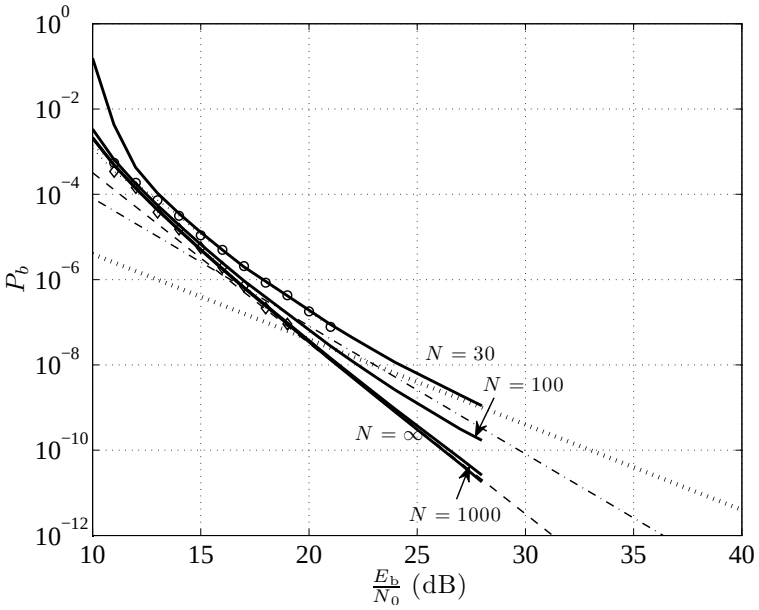


Fig. 4.8 Bit error probability union bounds and bit-error rate simulations of 8-PSK with the 8-state rate-2/3 convolutional code in a fully-interleaved Rayleigh fading channel. Interleaver length $N = 30$ (circles) and $N = 1000$ (diamonds). In solid lines, the saddlepoint approximation union bounds for $N = 30$, $N = 100$, $N = 1000$ and for infinite interleaving, with $\mathrm{PEP}_1(d)$. In dashed, dashed-dotted, and dotted lines, the heuristic approximations with weight $v = 1, 2, 3$ respectively.

Figure 4.8 shows the error performance of 8-PSK with Rayleigh fading and Gray labeling; the code is the optimum 8-state rate-2/3 convolutional code, with $d_{\min} = 4$. Again, an error floor appears due to the probability of having symbol scores of Hamming weight larger than 1. The figure depicts simulation results (for interleaver sizes $n = 90, 1000$) together with the saddlepoint approximations for finite N (for $N = 30, 100, 1000$), infinite interleaving (with $\mathrm{PEP}_1(d)$), and with the heuristic approximation $\mathrm{PEP}_{\mathrm{H}}(d, \rho_n)$ (only for $n = 90$ and $d = 4$).

For 8-PSK with Gray mapping, evaluation of Eq. (4.74) gives $d_{\mathrm{h}}^2(1) = 0.7664$, $d_{\mathrm{h}}^2(2) = 1.7175$, and $d_{\mathrm{h}}^2(3) = 2.4278$. Table 4.1 gives the values of $P(\rho_n)$ for the various patterns ρ_n. Table 4.1 also gives the threshold $\mathsf{snr}_{\mathrm{th}}$ given in Eq. (4.76) for all possible values of $\sum_{v \geq 1} n_v$,

Table 4.1 Asymptotic analysis for 8-PSK with varying interleaver length $n = 3N$ and minimum distance $d = 4$.

Pattern ρ_n	n	$P(\rho_n)$	Threshold $\frac{E_b}{N_0}$ (dB)
$(N-4,4,0,0)$	$n = 90$	0.8688	N/A
	$n = 300$	0.9602	N/A
	$n = 3000$	0.9960	N/A
$(N-3,2,1,0)$	$n = 90$	0.1287	16.0
	$n = 300$	0.0396	21.5
	$n = 3000$	0.0040	31.6
$(N-2,0,2,0)$,	$n = 90$	0.0015	20.5
$(N-1,1,0,1)$	$n = 300$	0.0002	26.0
	$n = 3000$	$2 \cdot 10^{-6}$	39.1

not only for the worst case. We observe that the main flattening of the error probability takes place at high snr. This effect essentially disappears for interleavers of practical length: for $n = 100$ (resp. $n = 1000$) the error probability at the first threshold is about 10^{-8} (resp. 10^{-12}). The saddlepoint approximation is remarkably precise; the heuristic approximation $\mathrm{PEP}_{\mathrm{H}}(d_{\min}, \rho_n)$ also gives very good results.

4.4 Bounds and Approximations Above the Cutoff Rate

Spurred by the appearance of turbo-codes [11] and the rediscovery of LDPC codes [69], there has been renewed interest in the past decade in the derivation of improved bounds for a region above the cutoff rate. A comprehensive review can be found in the monograph by Sason and Shamai [101]. In this section, we briefly discuss such bounds for BICM.

Of the available improved bounds, two are of special importance: the second Duman-Salehi (DS2) bound and the tangential sphere bound (TSB). The first bound, valid for general decoding metrics and channels, is directly applicable to BICM. The TSB is known to be the tightest bound in binary-input AWGN channels, and will be combined with the Gaussian approximation introduced in Section 4.2.

The error probability was analyzed in Section 3.1.2 for an ensemble

of random codes. For a fixed binary code $\mathcal{C}$, we consider the average over all possible interleavers π_n of length n and over all the scrambling sequences $\boldsymbol{d}$ described in Section 4.1. We denote the joint probability of the interleaver and the scrambling by $P(\pi_n, \boldsymbol{d})$. For a fixed message m (and binary codeword $\boldsymbol{c}_\mathsf{m}$), the transmitted codeword $\boldsymbol{x}_\mathsf{m}$ depends on the interleaver π_n and on the scrambling sequence $\boldsymbol{d}$. The conditional error probability $P_e(\mathsf{m})$, when message m and codeword $\boldsymbol{c}_\mathsf{m}$ are transmitted, can be derived from Eqs. (3.7) and (3.11), to obtain

$$P_e(\mathsf{m}) \leq \sum_{\pi_n, \boldsymbol{d}} P(\pi_n, \boldsymbol{d}) \int_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_\mathsf{m}) \left(\sum_{\mathsf{m}' \neq \mathsf{m}} \left(\frac{q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y})}{q(\boldsymbol{x}_\mathsf{m}, \boldsymbol{y})} \right)^s \right)^\rho d\boldsymbol{y} \tag{4.77}$$

for $0 \leq \rho \leq 1$ and $s > 0$.

The Duman-Salehi bound is derived by introducing a normalized measure $\psi(\boldsymbol{y})$ (possibly dependent on the codeword $\boldsymbol{c}_\mathsf{m}$), and transforming the integral in Eq. (4.77) as

$$\int_{\boldsymbol{y}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_\mathsf{m}) \frac{\psi(\boldsymbol{y})}{\psi(\boldsymbol{y})} \left(\sum_{\mathsf{m}' \neq \mathsf{m}} \left(\frac{q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y})}{q(\boldsymbol{x}_\mathsf{m}, \boldsymbol{y})} \right)^s \right)^\rho d\boldsymbol{y} \tag{4.78}$$

$$= \int_{\boldsymbol{y}} \psi(\boldsymbol{y}) \left(\sum_{\mathsf{m}' \neq \mathsf{m}} \frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_\mathsf{m})^{\frac{1}{\rho}}}{\psi(\boldsymbol{y})^{\frac{1}{\rho}}} \left(\frac{q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y})}{q(\boldsymbol{x}_\mathsf{m}, \boldsymbol{y})} \right)^s \right)^\rho d\boldsymbol{y}. \tag{4.79}$$

Now, applying Jensen's inequality we obtain

$$P_e(\mathsf{m}) \leq \left(\sum_{\mathsf{m}' \neq \mathsf{m}} \sum_{\pi_n, \boldsymbol{d}} P(\pi_n, \boldsymbol{d}) \int_{\boldsymbol{y}} \frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_\mathsf{m})^{\frac{1}{\rho}}}{\psi(\boldsymbol{y})^{\frac{1}{\rho}-1}} \left(\frac{q(\boldsymbol{x}_{\mathsf{m}'}, \boldsymbol{y})}{q(\boldsymbol{x}_\mathsf{m}, \boldsymbol{y})} \right)^s d\boldsymbol{y} \right)^\rho. \tag{4.80}$$

Next, we consider a decomposition $\psi(\boldsymbol{y}) = \prod_{n=1}^N \psi(y_n)$. We also assume that the all-zero binary codeword is transmitted, and group the summands m' having the same Hamming weight d together. Recall that A_d denotes the number of codewords with Hamming weight d and that $\rho_N = (N_0, \ldots, N_{w^\star})$, with $w^\star = \min(m, d)$, denotes the pattern

representing the allocation of d bits onto the N modulation symbols. The average over all possible patterns is equivalent to the average over all possible interleavers π_n. We thus have

$$\sum_{\pi_n,\boldsymbol{d}} P(\pi_n,\boldsymbol{d}) \int_{\boldsymbol{y}} \frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_{\mathsf{m}})^{\frac{1}{\rho}}}{\psi(\boldsymbol{y})^{\frac{1}{\rho}-1}} \left(\frac{q(\boldsymbol{x}_{\mathsf{m}'},\boldsymbol{y})}{q(\boldsymbol{x}_{\mathsf{m}},\boldsymbol{y})}\right)^s d\boldsymbol{y} = \sum_{\rho_N} P(\rho_N) \prod_{v=0}^{w^\star} P_{\mathrm{s}}(v)^{\rho_N(v)}, \tag{4.81}$$

where $P_{\mathrm{s}}(v)$ is defined as

$$P_{\mathrm{s}}(v) \triangleq \frac{1}{\binom{m}{v}2^m} \sum_{\boldsymbol{j}} \sum_{\boldsymbol{b}} \sum_{x\in\mathcal{X}_{\boldsymbol{j}}^{\boldsymbol{b}}} \int_y \frac{P_{Y|X}(y|x)^{\frac{1}{\rho}}}{\psi(y)^{\frac{1}{\rho}-1}} \left(\prod_{i=1}^{v} \frac{q_{j_i}(\bar{b}_{j_i},y)}{q_{j_i}(b_{j_i},y)}\right)^s dy. \tag{4.82}$$

In particular, for $v=0$ (for which the metrics $q(x_{\mathsf{m}},y)$ and $q(x_{\mathsf{m}'},y)$ coincide) and $v=1$ (for which the metrics differ in only 1 bit), we have

$$P_{\mathrm{s}}(0) = \frac{1}{2^m} \sum_{x\in\mathcal{X}} \int_y \frac{P_{Y|X}(y|x)^{\frac{1}{\rho}}}{\psi(y)^{\frac{1}{\rho}-1}} dy \tag{4.83}$$

$$P_{\mathrm{s}}(1) = \frac{1}{m2^m} \sum_{j=1}^{m} \sum_{b\in\{0,1\}} \sum_{x\in\mathcal{X}_j^b} \int_y \frac{P_{Y|X}(y|x)^{\frac{1}{\rho}}}{\psi(y)^{\frac{1}{\rho}-1}} \left(\frac{q_j(\bar{b},y)}{q_j(b,y)}\right)^s dy. \tag{4.84}$$

Finally, since the computation does not depend on the transmitted message, we have that the average error probability is bounded as

$$P_e \le \left(\sum_d A_d \sum_{\rho_N} P(\rho_N) \prod_{v=0}^{w^\star} P_{\mathrm{s}}(v)^{\rho_N(v)}\right)^\rho, \tag{4.85}$$

for all possible choices of ρ and ψ. For the specific choice $\rho=1$, the function ψ is inactive, and the functions $P_{\mathrm{s}}(v)$ are simply a Chernoff bound to the symbol score Ξ^{s} of weight v (see Definition 4.1).

As for the tangential sphere bound, the Gaussian approximation to the bit score discussed in Section 4.2 was used in [50] to give an approximation to the average error probability in the AWGN channel (see also Eq. (4.30)) . This analysis involves two approximations. First, all the symbol decoding scores of weight larger than 1 are neglected. Secondly,

the bit score is assumed to be Gaussian with equivalent signal-to-noise ratio $\mathsf{snr}_{\mathrm{eq}} = -\kappa_1(\hat{s})$ for all values of the score, not only the tail, so that distances between codewords follow a chi-square distribution. The resulting approximation is then given by

$$P_e \leq Q\left(\sqrt{-2n\kappa_1(\hat{s})}\right) + \int_{-\infty}^{\infty} \frac{\mathrm{d}\Xi^{\mathrm{b}}}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(\Xi^{\mathrm{b}})^2}{2}} \left(1 - \Gamma\left(\frac{n-1}{2}, \frac{\mathsf{r}^2}{2}\right)\right) + \sum_d \tilde{A}_d \int_{-\infty}^{\infty} \frac{\mathrm{d}\Xi^{\mathrm{b}}}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(\Xi^{\mathrm{b}})^2}{2}} \Gamma\left(\frac{n-2}{2}, \frac{\mathsf{r}^2 - \beta_d^2}{2}\right) \Big(Q\big(\beta_d\big) - Q(\mathsf{r})\Big), \tag{4.86}$$

where $n = mN$, and we have used that

$$\mathsf{r} \stackrel{\Delta}{=} \left(\sqrt{-2n\kappa_1(\hat{s})} - \Xi_{\mathrm{b}}\right) \tan\theta, \tag{4.87}$$

$$\beta_d \stackrel{\Delta}{=} \left(\sqrt{-2n\kappa_1(\hat{s})} - \Xi_{\mathrm{b}}\right) \tan\phi, \tag{4.88}$$

where $\tan\phi = \sqrt{\frac{d}{n-d}}$ and $\tan\theta$ is the solution of

$$\sum_d \tilde{A}_d \int_0^{\arccos\frac{\tan\phi}{\tan\theta}} \sin^{n-3}\theta \, \mathrm{d}\theta = \frac{\sqrt{\pi}\Gamma\big(\frac{n-2}{2}\big)}{\Gamma\big(\frac{n-1}{2}\big)}, \tag{4.89}$$

and the coefficients $\tilde{A}_d$ are $\tilde{A}_d = A_d$ if $\beta_d < \mathsf{r}$ and $\tilde{A}_d = 0$ otherwise.

Figure 4.9 shows the comparison between the simulation, saddlepoint approximation and the TSB with the Gaussian approximation for an RA code with 16-QAM with Gray mapping in the fully-interleaved Rayleigh fading channel. As we observe, the TSB with the Gaussian approximation yields an underestimation of the error floor, and gives an estimate of the error probability for low SNR. Unfortunately, validating the tightness of this approximation at low SNR is difficult, chiefly because of the use of approximations and of the suboptimality of iterative decoding compared to maximum-metric decoding.

4.5 Concluding Remarks and Related Work

In this Chapter, we have reviewed the error probability analysis of BICM. After considering the traditional error probability analysis of BICM based on the independent parallel channel model [29], we have

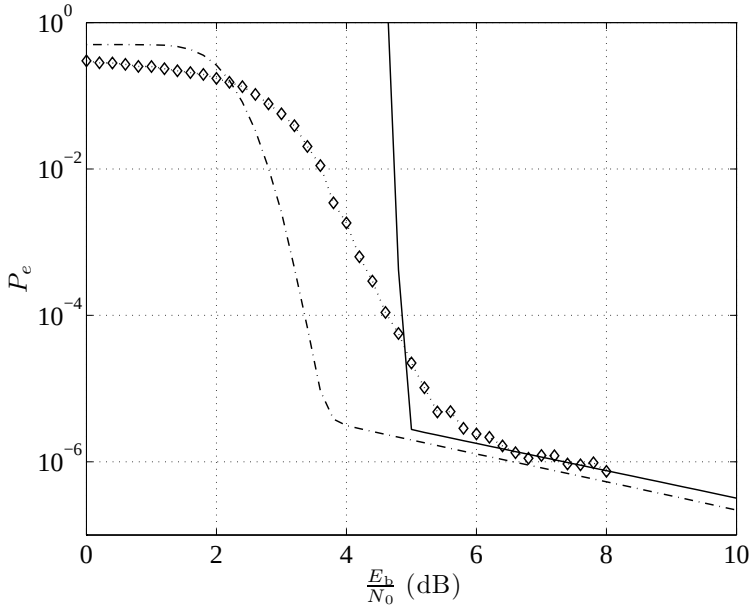


Fig. 4.9 Comparison of simulation results and saddlepoint (solid line) and TSB Gaussian approximations (dash-dotted line) on the bit error rate of BICM with an RA code of rate 1/4 with $K = 512$ information bits, 16-QAM modulation with Gray mapping, in the fully-interleaved AWGN Rayleigh fading channel.

presented a number of bounds and approximations to the error probability based on a union bound approach. In particular, we have shown that the saddlepoint approximation yields accurate results and is simple to compute. In passing, we have also seen why the expurgated bound proposed in [29] is not a true bound [104]. Next, and based on [141], we have considered the average over all possible interleavers (the uniform interleaver from turbo codes [8, 9]), and have shown that for interleavers of short length, the error probability is dominated by the symbols with Hamming weight larger than 1. This effect translates into a reduced rate of decay of the error probability with SNR. We have then reviewed improved bounds to the error probability like the Duman-Salehi and the tangential-sphere bound [101]. We have given the expression of the former for a general decoding metric, while the latter is an approximate bound based on the Gaussian approximation of the tail of the decoding score distribution.

Bounding techniques similar to those presented in this Chapter have been applied to the determination of the error floor of BICM with iterative decoding for convolutional codes or turbo-like codes [30, 63, 103,113]. In these analyses, the bit decoding score has access to perfect extrinsic side information on the values of the decoded bits.

4.A Saddlepoint Location

Since the metric $q_j(b,y)$ is proportional to $P_j(y|b)$, we have that

$$\mathbb{E}[\mathrm{e}^{s\Xi^{\mathrm{b}}}] = \int \left(\frac{1}{2} P_j^{1-s}(y|0) P_j^{s}(y|1) + \frac{1}{2} P_j^{1-s}(y|1) P_j^{s}(y|0) \right) \mathrm{d}y. \quad (4.90)$$

This quantity is symmetric around $s = \frac{1}{2}$, since it remains unchanged if we replace s by $1-s$. This also shows that the density of the bit score is independent of the value of the transmitted bit.

The first derivative of $\kappa_1(s)$, $\kappa_1'(s) = \frac{\mathbb{E}\left[\Xi^{\mathrm{b}} \mathrm{e}^{s\Xi^{\mathrm{b}}}\right]}{\mathbb{E}\left[\mathrm{e}^{s\Xi^{\mathrm{b}}}\right]}$, is zero when $\mathbb{E}\left[\Xi^{\mathrm{b}} \mathrm{e}^{s\Xi^{\mathrm{b}}}\right] = 0$. We concentrate on $\mathbb{E}\left[\Xi^{\mathrm{b}} \mathrm{e}^{s\Xi^{\mathrm{b}}}\right]$, readily computed as

$$\int \frac{1}{2} \left(P_j(y|0) \left(\frac{P_j(y|1)}{P_j(y|0)} \right)^s - P_j(y|1) \left(\frac{P_j(y|1)}{P_j(y|0)} \right)^s \right) \log \frac{P_j(y|1)}{P_j(y|0)} \, \mathrm{d}y, \quad (4.91)$$

which is zero at $\hat{s} = \frac{1}{2}$ thanks to the symmetry.

The other derivatives, evaluated at the saddlepoint, are given by

$$\kappa_1''(\hat{s}) = \frac{\mathbb{E}\left[(\Xi^{\mathrm{b}})^2 \mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}\right]}{\mathbb{E}\left[\mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}\right]}, \quad \kappa_1'''(\hat{s}) = \frac{\mathbb{E}\left[(\Xi^{\mathrm{b}})^3 \mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}\right]}{\mathbb{E}\left[\mathrm{e}^{\hat{s}\Xi^{\mathrm{b}}}\right]}. \quad (4.92)$$

We thus find that the second derivative is proportional to

$$\frac{1}{2} \int \left(P_j(y|0) \left(\frac{P_j(y|1)}{P_j(y|0)} \right)^s + P_j(y|1) \left(\frac{P_j(y|1)}{P_j(y|0)} \right)^s \right) \log^2 \frac{P_j(y|1)}{P_j(y|0)} \, \mathrm{d}y, \quad (4.93)$$

which is positive at $s = \frac{1}{2}$. Also, the third derivative is proportional to

$$\frac{1}{2} \int \left(P_j(y|0) \left(\frac{P_j(y|1)}{P_j(y|0)} \right)^s - P_j(y|1) \left(\frac{P_j(y|1)}{P_j(y|0)} \right)^s \right) \log^3 \frac{P_j(y|1)}{P_j(y|0)} \, \mathrm{d}y, \quad (4.94)$$

which is zero at $s = \frac{1}{2}$, thanks to the symmetry.

4.B Asymptotic Analysis with Nakagami Fading

We wish to compute the limit $\ell_v \stackrel{\Delta}{=} \lim_{\mathsf{snr}\to\infty} \frac{\mathrm{e}^{\kappa_v(s)}}{\mathsf{snr}^{-m_f}}$, given by

$$\ell_v = \lim_{\mathsf{snr}\to\infty} \frac{1}{\mathsf{snr}^{-m_f}} \left(\frac{1}{2^v \binom{m}{v}} \sum_{\boldsymbol{j},\boldsymbol{b}} \mathbb{E}\left[\frac{\prod_{i=1}^v q_{j_i}(\bar{b}_{j_i}, Y)^s}{\prod_{i=1}^v q_{j_i}(b_{j_i}, Y)^s} \right] \right). \tag{4.95}$$

As done in [74], only the dominant terms need to be kept in the bit scores $q_{j_i}(\cdot, y)$, which allows us to rewrite the expression inside the expectation as

$$\mathrm{e}^{-\sum_{i=1}^v s\left(\mathsf{snr}|H|^2|X-X_i'|^2 + 2\sqrt{\mathsf{snr}}\,\mathrm{Re}(H(X-X_i')Z^*)\right)}, \tag{4.96}$$

where x_i' is the closest symbol (in Euclidean distance) to the transmitted symbol x differing in the i-th bit label.

We now carry out the expectation over Z. Completing squares, and using that the formula for the density of Gaussian noise, we have that

$$\begin{aligned} \int \frac{1}{\pi} \mathrm{e}^{-|z|^2} \mathrm{e}^{-\sum_{i=1}^v s\left(\mathsf{snr}|H|^2|X-X_i'|^2 + 2\sqrt{\mathsf{snr}}\,\mathrm{Re}(H(X-X_i')z^*)\right)} \,\mathrm{d}z = \\ = \mathrm{e}^{-\mathsf{snr}|H|^2\left(s\sum_{i=1}^v |X-X_i'|^2 - s^2\left|\sum_{i=1}^v (X-X_i')\right|^2\right)}. \end{aligned} \tag{4.97}$$

In turn, the expectation over h of this quantity yields [75]

$$\left(1 + \left(s\sum_{i=1}^v |X-X_i'|^2 - s^2\left|\sum_{i=1}^v (X-X_i')\right|^2\right)\frac{\mathsf{snr}}{m_f}\right)^{-m_f}. \tag{4.98}$$

We next turn back to the limit ℓ_v. For large snr, we have that

$$\ell_v = \frac{m_f^{m_f}}{2^m \binom{m}{v}} \sum_{\boldsymbol{j},\boldsymbol{b}} \sum_{x\in\mathcal{X}_{\boldsymbol{b}}^{\boldsymbol{j}}} \left(s\sum_{i=1}^v |x-x_i'|^2 - s^2\left|\sum_{i=1}^v (x-x_i')\right|^2\right)^{-m_f}. \tag{4.99}$$

For each summand, the optimizing s is readily computed to be

$$\hat{s} = \frac{\sum_{i=1}^v |x-x_i'|^2}{2\left|\sum_{i=1}^v (x-x_i')\right|^2}, \tag{4.100}$$

which gives

$$\ell_v = \frac{1}{2^m \binom{m}{v}} \sum_{\boldsymbol{j},\boldsymbol{b}} \sum_{x \in \mathcal{X}_{\boldsymbol{b}}^{\boldsymbol{j}}} \left(\frac{4m_f \left| \sum_{i=1}^{v} (x - x_i') \right|^2}{\left(\sum_{i=1}^{v} |x - x_i'|^2 \right)^2} \right)^{m_f}. \tag{4.101}$$

In particular, for $v = 1$, we have

$$\ell_1 = \frac{1}{m2^m} \sum_{j,b} \sum_{x \in \mathcal{X}_b^j} \left(\frac{4m_f}{|x - x_1'|^2} \right)^{m_f}. \tag{4.102}$$

5

Iterative Decoding

In this Chapter, we study BICM with iterative decoding (BICM-ID). Inspired by the advent of turbo-codes and iterative decoding [11], BICM-ID was originally proposed by Li and Ritcey and ten Brink [64–66, 122] in order to improve the performance of BICM by exchanging messages between the demodulator and the decoder of the underlying binary code. As shown in [64–66, 122] and illustrated in Figure 5.1, BICM-ID can provide some performance advantages using convolutional codes combined with mappings different from Gray. With more powerful codes, the gains are even more pronounced [120, 121].

In BICM-ID, a message-passing decoding algorithm [96] is applied to the factor graph (FG) [62] representing the BICM scheme and the channel. In particular, we focus our attention on the Belief Propagation (BP) message-passing algorithm. The BP algorithm often yields excellent performance, even when it is not optimal, or if optimality cannot be shown. We describe a numerical analysis technique, known as density evolution, that is able to characterize the behavior of the iterative decoder for asymptotically large block length. Since density evolution is computationally demanding, we focus on a simpler, yet accurate, approximate technique based on extrinsic information transfer

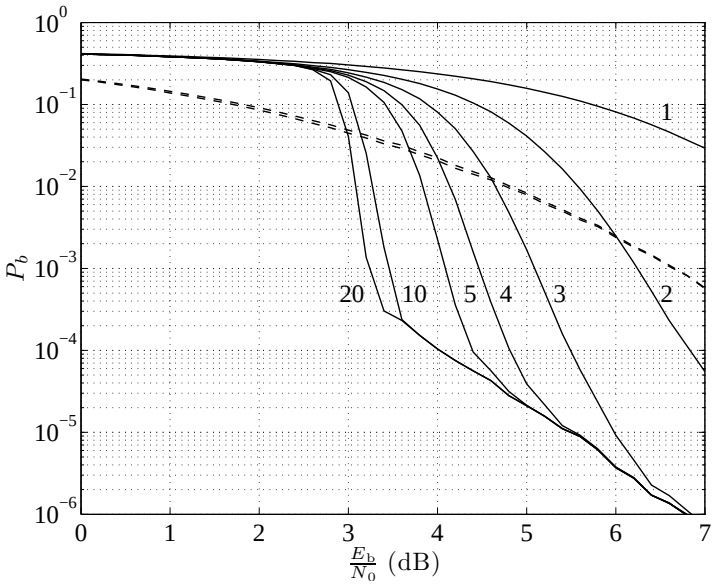


Fig. 5.1 Bit error rate for BICM-ID in the AWGN channel with the $(5,7)_8$ convolutional code, 16-QAM with set partitioning mapping, for $n = 20000$. Curve labels denote the iteration number. Upper curve corresponds to iteration 1, decreasing down to iteration 5. In dashed line, results for Gray mapping.

(EXIT) charts [116, 117, 119]. This technique is particularly useful for the optimization of BICM-ID schemes.

We also discuss the so-called area theorem that yields a condition (within the limits of the EXIT approximation) for which a BICM-ID scheme can approach the corresponding coded modulation capacity. Due to the shape of the EXIT curves and despite the good performance shown in Figure 5.1, standard convolutional codes with BICM-ID cannot achieve vanishing error probability as the block length increases. We particularize the area theorem to a number of improved constructions based on LDPC and RA codes and illustrate the impact of the design parameters (degree distributions). We will finally show how to design capacity-approaching BICM schemes using curve fitting. Simulations show that EXIT analysis is highly accurate.

5.1 Factor Graph Representation and Belief Propagation

In this section, we present the FG representation of BICM. In general, consider a coding problem where the information message m is represented in binary form, as a block of K information bits $\mathsf{m}_1, \ldots, \mathsf{m}_K$. The message is mapped by the encoding function ϕ onto a codeword $\boldsymbol{x}$. Let $\boldsymbol{y}$ denote the channel output. The optimal decoding rule that minimizes the bit-error probability P_b is the bit-wise MAP rule

$$\widehat{\mathsf{m}}_i = \underset{b\in\{0,1\}}{\arg\max} \ \Pr\{\mathsf{m}_i = b|\boldsymbol{y}\}. \tag{5.1}$$

We compute this marginal a posteriori probability (APP) $\Pr\{\mathsf{m}_i = b|\boldsymbol{y}\}$ starting with the joint APP of all information bits $\Pr\{\mathsf{m}|\boldsymbol{y}\}$, given by

$$\begin{aligned} \Pr\{\mathsf{m}|\boldsymbol{y}\} &= \sum_{\boldsymbol{x}\in\mathcal{X}^N} \frac{P_{\boldsymbol{Y},\boldsymbol{X},\mathsf{M}}(\boldsymbol{y},\boldsymbol{x},\mathsf{m})}{P_{\boldsymbol{Y}}(\boldsymbol{y})} \\ &= \sum_{\boldsymbol{x}\in\mathcal{X}^N} \frac{[\phi(\mathsf{m})=\boldsymbol{x}]\prod_{k=1}^{N} P_{Y|X}(y_k|x_k)}{|\mathcal{M}|\ P_{\boldsymbol{Y}}(\boldsymbol{y})} \end{aligned} \tag{5.2}$$

where

$$[\mathcal{A}] = \begin{cases} 1 & \text{if } \mathcal{A} \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

denotes the indicator function of the condition $\mathcal{A}$. Since the denominator is common to all information messages, the marginal APP for the i-th information bit is given by

$$\Pr\{\mathsf{m}_i = b|\boldsymbol{y}\} = \sum_{\sim\mathsf{m}_i} \Pr\{\mathsf{m}|\boldsymbol{y}\}\big|_{\mathsf{m}_i=b} \tag{5.3}$$

where $\sum_{\sim\mathsf{m}_i}$ indicates the *summary operator* [62] that sums the argument over the ranges of all variables, with the exception of the variable m_i, which is held to a fixed value. The marginalization (5.3) of the joint APP is a formidable task, since it requires summing over 2^{K-1} terms. For practical values of K this is typically infeasible, unless the problem has some special structure that can be exploited.

Factor graphs are a general tool to represent a complicated multivariate function when this splits into the product of simpler *local* factors, each one of which contains only a subset of the variables. Given a

factorization, the corresponding Factor Graph (FG) is a bipartite graph with two sets of nodes, the variable nodes $\mathcal{V}$ and the function nodes $\mathcal{F}$. The set of edges $\mathcal{E}$ in the FG contains all edges (ϑ, f) for which the variable $\vartheta \in \mathcal{V}$ is an argument of the factor (function) $f \in \mathcal{F}$ [62]. We use circles and squares in order to represent variable and function nodes in the FGs, respectively. As an example, Figure 5.2 shows the FG of a BICM scheme of rate $R = 1$, using a rate-1/3 binary convolutional code and an 8-ary modulation ($m = 3$). On the right side of the FG we see the convolutional structure of the binary code, with boxes representing the encoder state equation, black circles representing the encoder state variables, and white circles representing the encoder input and output variables. On the left, we notice the modulator part, where white boxes represent the modulation mapping μ (from binary labels to signal points), taking labels of three bits each and producing the corresponding signal point. In between, separating the code and the modulator, we have an interleaver, i. e. a permutation of the coded bits on the right-hand side to the label bits on the left-hand side. These connections are included into a big box for the sake of clarity.

A general method for computing or approximating the marginal APPs consists of applying the BP algorithm to a FG corresponding to the joint APP. If the FG is a tree, then BP computes the marginal APP exactly [62]. If the FG contains loops, then the BP algorithm becomes iterative in nature, and gives rise to a variety of iterative decoding algorithms. These algorithms are distinguished by the underlying FG used for decoding, by the scheduling of the local computation at the nodes, and by the approximations used in the computations, that sometimes are necessary in order to obtain low complexity algorithms. Generally speaking, the BP algorithm is a *message passing* algorithm that propagates messages between adjacent nodes in the FG. For any given node, the message output on a certain edge is a function of all the messages received on all other edges.

The BP general computation rules are given as follows [62]. Consider a variable node ϑ, connected to function nodes $\{f \in \mathcal{N}(\vartheta)\}$, where $\mathcal{N}(\vartheta)$ denotes the set of neighbors of node ϑ in the FG. The variable-

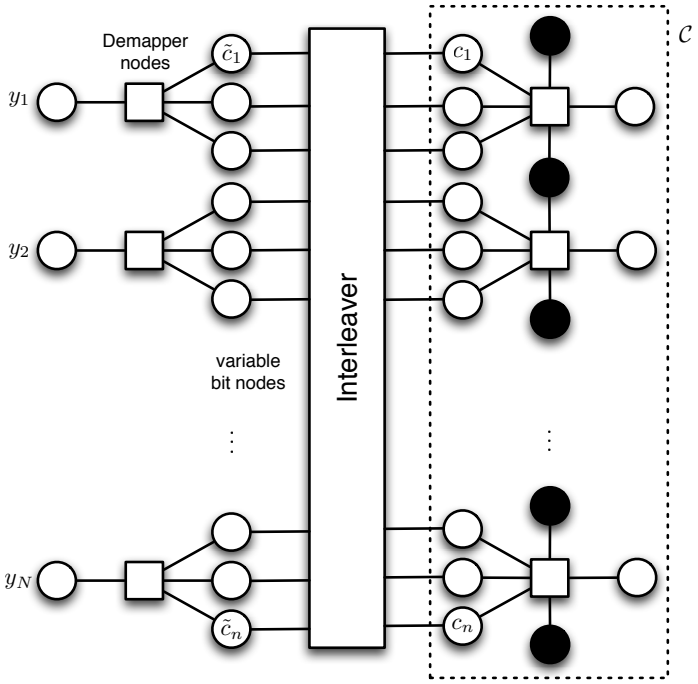


Fig. 5.2 Factor graph representation of a BICM scheme of rate $R = 1$. In this case, $\mathcal{C}$ is a binary convolutional code of rate $r = \frac{1}{3}$ and $m = 3$, i. e., 8-PSK or 8-QAM.

to-function message from ϑ to $f \in \mathcal{N}(\vartheta)$ is given by

$$\nu^{\vartheta \to f} = \prod_{\substack{g \in \mathcal{N}(\vartheta) \\ g \neq f}} \nu^{g \to v}. \tag{5.4}$$

Messages are functions of the variable node they come from or are sent to, i. e., both $\nu^{f \to \vartheta}$ and $\nu^{\vartheta \to f}$ are functions of the single variable ϑ. Consider a function node f, connected to variable nodes $\{\vartheta \in \mathcal{N}(f)\}$, where $\mathcal{N}(f)$ denotes the set of neighbors of node f in the FG. The function-to-variable message from f to $\vartheta \in \mathcal{N}(f)$ is given by

$$\nu^{f \to \vartheta} = \sum_{\sim \vartheta} f(\{\eta \in \mathcal{N}(f)\}) \prod_{\substack{\eta \in \mathcal{N}(f) \\ \eta \neq \vartheta}} \nu^{\eta \to f}. \tag{5.5}$$

where $\sum_{\sim \vartheta}$ indicates summation of the argument inside over the ranges of all variables, with the exception of ϑ.

We particularize the general BP computation rules at the nodes to the FG of a BICM scheme as shown in Figure 5.2. In this case, the code part of the FG is standard, and depending on the structure of the code, the BP computation can be performed with well-known methods. For example, if the code is convolutional, we can use the optimal forward backward (or BCJR) algorithm [7]. We treat in more detail the computation at the modulator mapping nodes, referred to in the following as *demapper* nodes[1].

For the FG of the joint APP given in (5.2), all messages are marginal probabilities, or proportional to marginal probabilities. With some abuse of notation, let $\Pr^{\text{dec}\to\text{dem}}\{b_j(x) = b\}$ denote the probability that the j-th label bit of a symbol x is equal to $b \in \{0, 1\}$, according to the message probability sent by the decoder to the demapper. Similarly, let $\Pr^{\text{dem}\to\text{dec}}\{b_j(x) = b\}$ denote the corresponding probability according to the message probability sent by the demapper to the decoder. A straightforward application of (5.5) in this context yields

$$\Pr^{\text{dem}\to\text{dec}}\{b_j(x_k) = b\} = \sum_{x' \in \mathcal{X}_b^j} P_{Y|X}(y_k \,|\, x') \prod_{\substack{j'=1 \\ j' \neq j}}^{m} \Pr^{\text{dec}\to\text{dem}}\{b_{j'}(x')\} \tag{5.6}$$

An equivalent formulation of the BP, which is sometimes more convenient, uses log-ratios instead of probabilities. In some cases, these can be log-likelihood ratios (LLRs). In this case, the LLRs propagated by the BP are the bit scores defined in the previous Chapter. For the j-th bit of the label of the k-th symbol, from (5.6), we define

$$\Xi_{m(k-1)+j}^{\text{dem}\to\text{dec}} = \log \frac{q_j\big(b_j(x_k) = 1, y_k\big)}{q_j\big(b_j(x_k) = 0, y_k\big)} \tag{5.7}$$

$$= \log \frac{\sum_{x' \in \mathcal{X}_1^j} P_{Y|X}(y_k|x') \prod_{\substack{j'=1 \\ j' \neq j}}^{m} \Pr^{\text{dec}\to\text{dem}}\big\{b_{j'}(x')\big\}}{\sum_{x' \in \mathcal{X}_0^j} P_{Y|X}(y_k|x') \prod_{\substack{j'=1 \\ j' \neq j}}^{m} \Pr^{\text{dec}\to\text{dem}}\big\{b_{j'}(x')\big\}} \tag{5.8}$$

This corresponds to the message from the demapper function node k to the code binary variable node $i = m(k-1) + j$. In this Chapter we

[1] Demapping indicates the fact that these nodes undo, in a probabilistic soft-output sense, the mapping μ from the label bits to the modulation symbol.

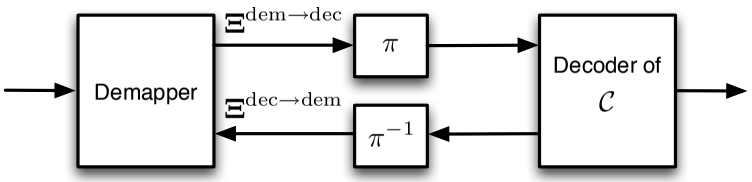


Fig. 5.3 Block diagram of BICM-ID.

will consider the analysis of BICM-ID using an infinite interleaver. We hence drop the superscript b in the bit scores to simplify the notation. Instead, we have added a superscript to indicate the direction of the message, i. e., from demapper to decoder or from decoder to demapper. We denote the vector of bit scores by

$$\boldsymbol{\Xi}^{\mathrm{dem}\to\mathrm{dec}} \triangleq (\Xi_1^{\mathrm{dem}\to\mathrm{dec}}, \ldots, \Xi_n^{\mathrm{dem}\to\mathrm{dec}}) \tag{5.9}$$

and the vector of all bit scores but the i-th by

$$\boldsymbol{\Xi}_{\sim i}^{\mathrm{dem}\to\mathrm{dec}} \triangleq (\Xi_1^{\mathrm{dem}\to\mathrm{dec}}, \ldots, \Xi_{i-1}^{\mathrm{dem}\to\mathrm{dec}}, \Xi_{i+1}^{\mathrm{dem}\to\mathrm{dec}}, \ldots, \Xi_n^{\mathrm{dem}\to\mathrm{dec}}). \tag{5.10}$$

We further denote the a priori log-messages as

$$\Xi_{m(k-1)+j}^{\mathrm{dec}\to\mathrm{dem}} = \log \frac{\Pr^{\mathrm{dec}\to\mathrm{dem}}\{b_j(x_k) = 1\}}{\Pr^{\mathrm{dec}\to\mathrm{dem}}\{b_j(x_k) = 0\}}. \tag{5.11}$$

Similarly, we define the vectors

$$\boldsymbol{\Xi}^{\mathrm{dec}\to\mathrm{dem}} \triangleq (\Xi_1^{\mathrm{dec}\to\mathrm{dem}}, \ldots, \Xi_n^{\mathrm{dec}\to\mathrm{dem}}) \tag{5.12}$$

and the vector of all log a priori messages but the i-th by

$$\boldsymbol{\Xi}_{\sim i}^{\mathrm{dec}\to\mathrm{dem}} \triangleq (\Xi_1^{\mathrm{dec}\to\mathrm{dem}}, \ldots, \Xi_{i-1}^{\mathrm{dec}\to\mathrm{dem}}, \Xi_{i+1}^{\mathrm{dec}\to\mathrm{dem}}, \ldots, \Xi_n^{\mathrm{dec}\to\mathrm{dem}}). \tag{5.13}$$

Figure 5.3 shows the block diagram of the basic BICM-ID decoder.

5.2 Density Evolution

In this section, we discuss a method to characterize the asymptotic performance of BICM in the limit of large interleaver size. This method,

named *density evolution*, describes how the densities of the messages passed along the graph evolve through the iterations.

In particular, we consider a *random ensemble* of BICM schemes defined by a random interleaver π_n, generated with uniform probability over the set of permutations of n elements, and a random binary scrambling sequence $\boldsymbol{d} \in \{0,1\}^n$ generated i.i.d. with uniform probability. The random interleaver and the random scrambling sequence were respectively introduced in Sections 4.3 and 4.4, and in Sections 4.1 and 4.4. As already mentioned in Section 4.1, these elements used to symmetrize the system with respect to the transmitted codewords. In order to transmit a binary codeword $\boldsymbol{c} \in \mathcal{C}$, the BICM encoder produces $\boldsymbol{c} \oplus \boldsymbol{d}$, interleaves the result and maps the interleaved scrambled binary codeword over a sequence of modulation symbols. The receiver knows $\boldsymbol{d}$, such that the modulo-2 scrambling can be undone at the receiver.

Let us denote the coded bit error probability at iteration ℓ when interleaver π_n and scrambling sequence $\boldsymbol{d}$ are used by $P_c^{(\ell)}(\pi_n, \boldsymbol{d})$. We wish to compute the averaged coded bit error probability $\bar{P}_c$ at iteration ℓ, given by

$$\bar{P}_c^{(\ell)} \triangleq \sum_{\pi_n, \boldsymbol{d}} P(\pi_n) P(\boldsymbol{d}) P_c^{(\ell)}(\pi_n, \boldsymbol{d}), \tag{5.14}$$

where $P(\pi_n)$ and $P(\boldsymbol{d})$ respectively denote the probabilities of the interleaver and the scrambling sequence.

For given interleaver and scrambling sequence, the message vectors $\boldsymbol{\Xi}^{\mathrm{dem}\to\mathrm{dec}}$ and $\boldsymbol{\Xi}^{\mathrm{dec}\to\mathrm{dem}}$ are random vectors, function of the channel noise and fading and of the transmitted information message. The BICM-ID decoder behavior can be characterized by studying the joint probability density of $\boldsymbol{\Xi}^{\mathrm{dem}\to\mathrm{dec}}$ (or, equivalently, of $\boldsymbol{\Xi}^{\mathrm{dec}\to\mathrm{dem}}$) as a function of the decoder iterations. This can be a highly complicated task due to n being large and to the statistical dependence between the messages passed along the FG. Moreover, the joint probability density of the message vector at a given iteration is a function of the (random) interleaver and of the (random) scrambling sequence. Fortunately, a powerful set of rather general results come to rescue [22, 60, 96, 97]. As we will state more precisely later, the marginal probability density of a message sent over a randomly chosen edge in the FG converges

almost surely to a deterministic probability density that depends on the BICM ensemble parameters for large block lengths, provided that the FG is sufficiently local (e.g., the maximum degree of each node is bounded). The proof of this concentration result for BICM-ID closely follows the footsteps of analogous proofs for LDPC codes over BIOS channels [96, 97], for LDPC codes over binary-input ISI channels [60], and for linear codes in multiple access Gaussian channels [22].

By way of illustration, consider a BICM ensemble based on the concatenation of a binary convolutional code with a modulation signal set through an interleaver, whose FG is shown in Figure 5.2. We focus on a randomly chosen coded symbol c_i. Also, we let $i = m(k-1)+j$ such that the coded symbol i corresponds to the k-th modulation symbol, in label position j. We wish to compute the limiting density of the message $\Xi_i^{\mathrm{dem}\to\mathrm{dec}}$, from the demapper node μ_k to the variable node c_i, at the ℓ-th iteration.

We define the oriented neighborhood $\mathcal{N}^\ell(\mu_k \to c_i)$ of depth 2ℓ as the subgraph formed by all paths of length 2ℓ starting at node μ_k and not containing the edge $(\mu_k \to c_i)$. As done in [22, 60], we decode the convolutional code using a windowed BCJR algorithm. This algorithm produces the extrinsic information output for symbol c_i by operating over a trellis finite window centered around the symbol position i. The trellis states at the window edges are assumed to be uniformly distributed. It is well-known that if the window size is much larger than the code constraint length, then the windowed BCJR approaches the exact (optimal) BCJR algorithm that operates over the whole trellis. Figure 5.4 shows an example of oriented neighborhood of depth 4.

In general, the neighbourhood $\mathcal{N}^\ell(\mu_k \to c_i)$ has cycles, or loops, by which we mean that some node of the FG appears more than once in $\mathcal{N}^\ell(\mu_k \to c_i)$. As $n \to \infty$, however, one can show that the probability that the neighborhood $\mathcal{N}^\ell(\mu_k \to c_i)$ is cycle-free for finite ℓ can be made arbitrarily close to 1. Averaged over all possible interleavers, the probability that the neighborhood has cycles is bounded as [22, 60]

$$\Pr\left\{\mathcal{N}^\ell(\mu_k \to c_i) \text{ has cycles}\right\} \leq \frac{\alpha}{n}, \tag{5.15}$$

as $n \to \infty$ and for some positive α independent of n. The key element in the proof is the locality of the FG, both for the code and for the

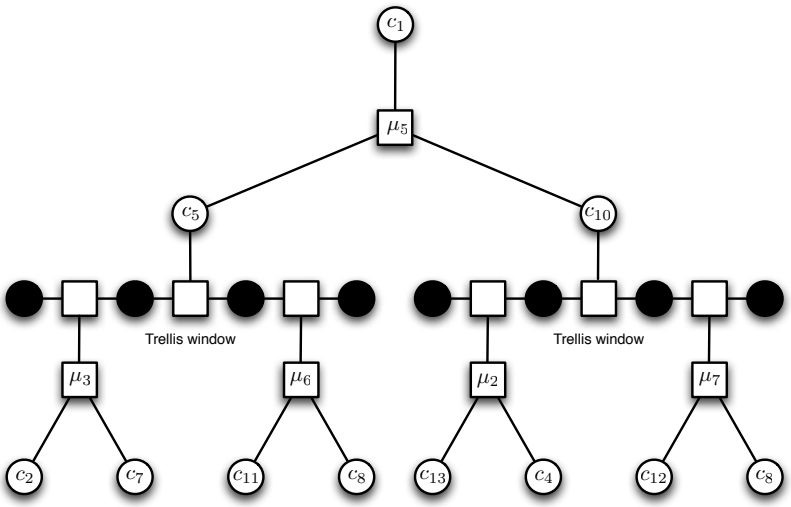


Fig. 5.4 An instance of the oriented neighborhood $\mathcal{N}^2(\mu_5 \to c_1)$ of depth 4 with a trellis window of size 1 for $m = 3$.

modulator. Notice that this result is reminiscent of Eq. (4.61) in Chapter 4, which concerned the probability that d bits randomly distributed over n bits belong to different modulation symbols.

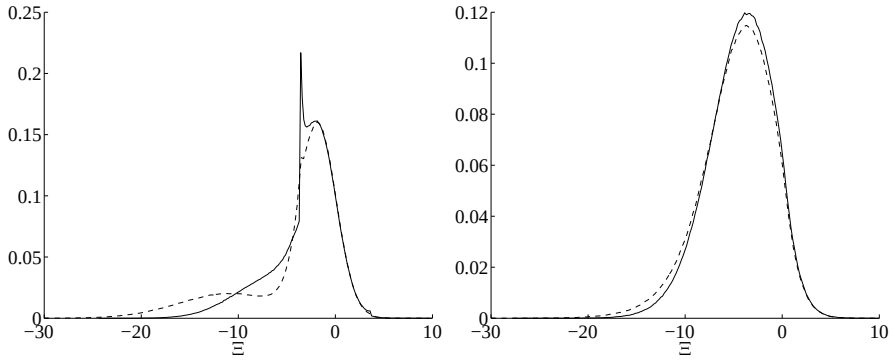
Furthermore, we have then the following concentration theorem [22, 60, 96, 97]:

Theorem 5.1. Let $Z^{(\ell)}(\boldsymbol{c}, \pi_n, \boldsymbol{d})$ denote the number of coded bits in error at the ℓ-th iteration, for an arbitrary choice of codeword $\boldsymbol{c}$, interleaver π_n, and scrambling sequence $\boldsymbol{d}$. For arbitrarily small $\epsilon > 0$, there exists a positive number β such that

$$\Pr\left\{\left|\frac{Z^{(\ell)}(\boldsymbol{c}, \pi_n, \boldsymbol{d})}{n} - \bar{P}_c^{(\ell)}\right| \geq \epsilon\right\} \leq e^{-\beta\epsilon^2 n}. \tag{5.16}$$

Therefore, for large n, the performance of a given instance of codeword, interleaver, and scrambling sequence, is close to the average error probability $\bar{P}_c^{(\ell)}$ of a randomly chosen bit.

Under the cycle-free assumption, all messages propagated along the edges of $\mathcal{N}^\ell(\mu_k \to c_i)$ are statistically independent. Finally, note that



(a) Density of $\Xi^{\mathrm{dem}\to\mathrm{dec}}$. (b) Density of $\Xi^{\mathrm{dec}\to\mathrm{dem}}$.

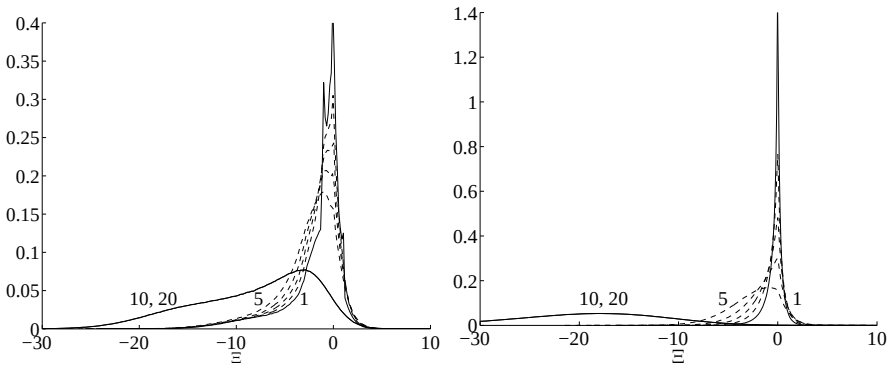
Fig. 5.5 Density evolution of messages $\Xi^{\mathrm{dem}\to\mathrm{dec}}$ and $\Xi^{\mathrm{dec}\to\mathrm{dem}}$ with the $(5,7)_8$ convolutional code with 16-QAM and Gray mapping at $\frac{E_\mathrm{b}}{N_0} = 3.5$ dB. Solid lines correspond to the first iteration; dashed lines correspond to iteration 5.

the coded bit error probability is simply the tail probability of the message $\Xi_i^{\mathrm{dec}\to\mathrm{dem}}$, from the variable node c_i to the demapper node μ_k[2]. Density evolution tracks precisely the density of this message. Note also that by construction, the message $\Xi_i^{\mathrm{dem}\to\mathrm{dec}}$ is the the *bit score* Ξ^b introduced in Chapter 4.

A detailed DE algorithm, for arbitrary modulation constellation and labeling, arbitrary (memoryless stationary) channel and arbitrary binary convolutional code, can be found in Appendix 5.A.

Figures 5.5 and 5.6 illustrate the density evolution method for the $(5,7)_8$ convolutional code, 16-QAM with Gray and set-partitioning mapping, respectively. Note that the $\frac{E_\mathrm{b}}{N_0} = 3.5$ dB is above the waterfall, which from Figure 5.1 occurs at approximately $\frac{E_\mathrm{b}}{N_0} = 3$ dB. Figure 5.5 shows that the positive tail of the message densities does not change much with the iterations, resulting in a nearly equal error probability. On the other hand, Figure 5.6, illustrates the improvement achieved with iterative decoding, which reduces the tail of the distribution. Notice, however, that there is an irreducible tail, which results in an error

[2] This probability corresponds to decisions based on extrinsic information. Usually, decisions are based on APP information for which the corresponding error probability is given by the tail of the APP message $\Xi_i^{\mathrm{dec}\to\mathrm{dem}} + \Xi_i^{\mathrm{dem}\to\mathrm{dec}}$.



(a) Density of $\Xi^{\mathrm{dem}\to\mathrm{dec}}$.

(b) Density of $\Xi^{\mathrm{dec}\to\mathrm{dem}}$.

Fig. 5.6 Density evolution of messages $\Xi^{\mathrm{dem}\to\mathrm{dec}}$ and $\Xi^{\mathrm{dec}\to\mathrm{dem}}$ with the $(5,7)_8$ convolutional code with 16-QAM and set-partitioning mapping at $\frac{E_\mathrm{b}}{N_0} = 3.5$ dB. Solid lines correspond to iterations $1, 10$ and 20; dashed lines correspond to iterations $2, 3, 4$ and 5.

floor. As opposed to turbo-like and LDPC codes, this error floor occurs even for infinite interleaving. As shown in [30], this floor matches the performance with perfect side information on the transmitted bits. The results match with the error probability simulations shown in Figure 5.1. Note that the densities would not evolve significantly through the iterations if $\frac{E_\mathrm{b}}{N_0} < 3$ dB.

5.3 EXIT Charts

In the previous section, we have described density evolution as a method that characterizes the message-passing process exactly in the limit for infinite interleavers. Density evolution is computationally demanding, since the whole density of the messages needs to be tracked. Unfortunately, density evolution does not yield simple criteria to optimize the BICM-ID scheme. However, there are a number of one-dimensional approximate methods that only track one parameter of the density. These methods are simpler to compute and yield good and accurate criteria to design efficient BICM-ID schemes. Among the possible one-dimensional parameters, the mutual information functional yields accurate approximations as well as analytical insight into the design of iteratively decodable systems [6, 129].

Extrinsic information transfer (EXIT) charts are a tool for predicting the behavior of iterative detection and decoding systems like BICM-ID, and where first proposed by ten Brink in [116, 117, 119]. EXIT charts are a convenient and accurate single-parameter approximation of density evolution that plots the mutual information between bits and their corresponding messages in the iterative decoder. In particular, let

$$\mathsf{x} \triangleq \frac{1}{n}\sum_{k=1}^{n} I(C_k; \Xi_k^{\mathrm{dec}\to\mathrm{dem}}), \tag{5.17}$$

$$\mathsf{y} \triangleq \frac{1}{n}\sum_{k=1}^{n} I(C_k; \Xi_k^{\mathrm{dem}\to\mathrm{dec}}) \tag{5.18}$$

respectively denote the average mutual information between the coded bits C_k, $k = 1, \ldots, n$, and the decoder-to-demapper messages $\Xi_k^{\mathrm{dec}\to\mathrm{dem}}$ and the demapper-to-decoder messages $\Xi_k^{\mathrm{dem}\to\mathrm{dec}}$. Note that x and y respectively correspond to the extrinsic informations at the decoder (demapper) output (input) and input (output). EXIT charts represent the extrinsic information as a function of a priori information,

$$\mathsf{y} = \mathsf{exit}_{\mathrm{dem}}(\mathsf{x}), \tag{5.19}$$

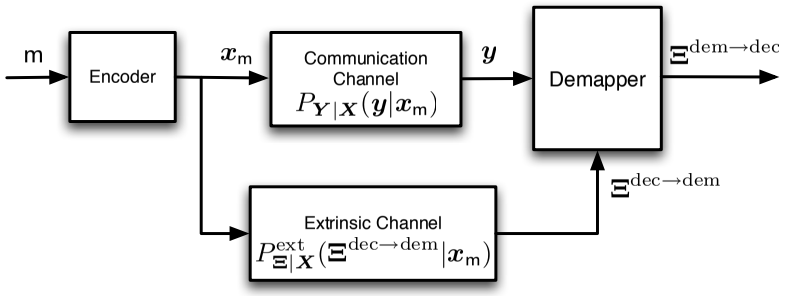
$$\mathsf{x} = \mathsf{exit}_{\mathrm{dec}}(\mathsf{y}), \tag{5.20}$$

and thus represent the transfer of extrinsic information in the demapper and decoder blocks. Since we deal with binary variables, the EXIT charts are represented in the axis $[0, 1]\times[0, 1]$. The fixed point of BICM-ID, where further decoding iterations do not improve the performance, is the leftmost intersection of the EXIT curves.

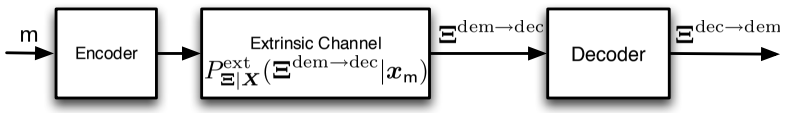
EXIT-chart analysis is based on the simplifying assumption that the a priori information, i. e., x for $\mathsf{y} = \mathsf{exit}_{\mathrm{dem}}(\mathsf{x})$ and y for $\mathsf{x} = \mathsf{exit}_{\mathrm{dec}}(\mathsf{y})$, comes from an independent extrinsic channel. Figures 5.7(a) and 5.7(b) show the EXIT decoding model and extrinsic channels for a priori information for the demapper and decoder blocks, respectively.

In the original EXIT charts proposed by ten Brink [116, 117, 119], the extrinsic channel was modeled as a binary-input AWGN channel with an equivalent SNR given by

$$\mathsf{snr}_{\mathrm{ap}}^{\mathrm{dem}} = \mathsf{C}_{\mathrm{bpsk}}^{-1}(\mathsf{x}) \tag{5.21}$$



(a) Demapper



(b) Decoder

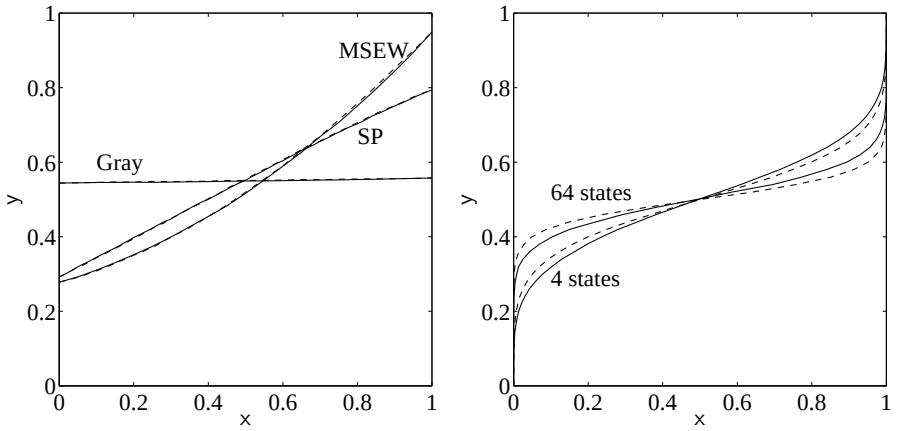
Fig. 5.7 Extrinsic channel models in BICM-ID.

for the demapper EXIT chart, and

$$\mathsf{snr}_{\mathrm{ap}}^{\mathrm{dec}} = \mathsf{C}_{\mathrm{bpsk}}^{-1}(\mathsf{y}) \tag{5.22}$$

for the decoder EXIT chart, where $\mathsf{C}_{\mathrm{bpsk}}(\mathsf{snr})$ is the capacity curve for BPSK modulation over the AWGN channel. In this model, the a priori messages are modeled as normal distributed. In practice, the AWGN model for the extrinsic channel predicts the convergence behavior of BICM-ID accurately, although not exactly, since the real messages exchanged in the iterative process are not Gaussian. Another particularly interesting type of extrinsic channel is the binary erasure channel (BEC), for which the EXIT chart analysis is exact.

Different mappings might behave differently in BICM-ID [117]. This is shown in Figure 5.8(a), where we show the EXIT curves for Gray, set-partitioning and the optimized maximum squared Euclidean weight (MSEW) [113] for 16-QAM. In particular, Gray mapping has an almost flat curve. This implies that not much will be gained through iterations, since as the input extrinsic information x increases, the output extrinsic information y virtually stays the same, i. e., when decoding a given label position, the knowledge of the other positions of the label does not

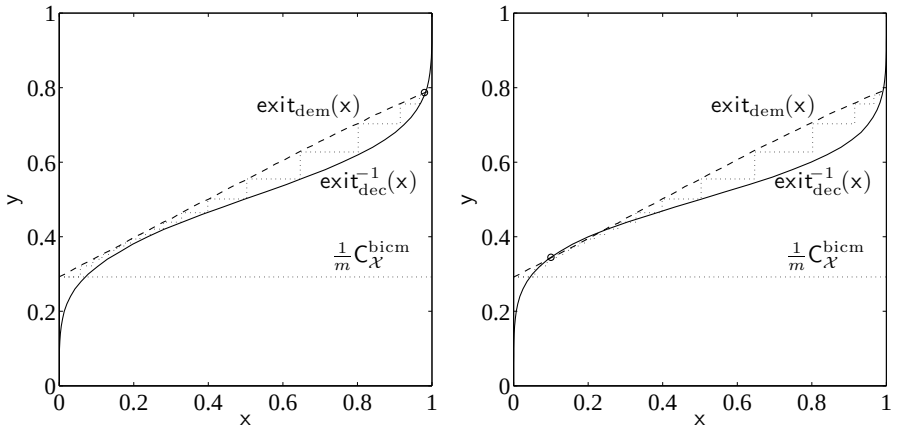


(a) Mapping comparison. (b) Convolutional code comparison.

Fig. 5.8 (a) EXIT curves for Gray, set partitioning and MSEW are shown at $\mathsf{snr} = 6$ dB with an AWGN (solid) and a BEC (dashed) extrinsic channel. (b) Exit curves for 4 and 64 states convolutional codes with an AWGN (solid) and a BEC (dashed) extrinsic channel.

help. Instead, set partitioning and MSEW show better EXIT curves. An important observation is that even with perfect extrinsic side information (see the point $\mathsf{x} = 1$ in the Figure), the mappings cannot perfectly recover the bit for which the message is being computed, i. e., these mappings cannot reach the $(\mathsf{x}, \mathsf{y}) = (1, 1)$ point in the EXIT chart. Notice that the AWGN and BEC EXIT curves are close to each other. Figure 5.8(b) shows the EXIT curves for 4- and 64-state convolutional codes; the more powerful the code becomes, the flatter its EXIT curve. The differences between BEC and AWGN extrinsic information are slightly more significant in this case.

Figures 5.9(a) and 5.9(b) show the EXIT chart of a BICM-ID scheme with a convolutional code, 16-QAM with set partitioning mapping, and with AWGN and BEC extrinsic channels respectively. We also show the simulated trajectory of the information exchange between demapper and decoder through the iterations as well as the fixed point (circle). The trajectory matches well the predictions made by the AWGN EXIT analysis, while it is not as accurate with the BEC EXIT, which predicts a fixed point not achieved in reality. Note that this matches with the results of Figure 5.1, where a waterfall is observed at



(a) AWGN EXIT vs Trajectory.

(b) BEC EXIT vs Trajectory.

Fig. 5.9 EXIT chart and simulated trajectory for BICM-ID with the $(5,7)_8$ convolutional code with 16-QAM with set partitioning mapping with $\mathsf{snr} = 6$ dB, and AWGN (a) and BEC (b) extrinsic information. The normalized BICM capacity $\frac{1}{m}\mathsf{C}^{\mathrm{bicm}}_{\mathcal{X}}$ is plotted for reference. Dashed lines correspond to the EXIT curve of the demapper, solid lines correspond to the EXIT curve of the decoder while circles denote the fixed point.

$\frac{E_\mathrm{b}}{N_0} \approx 3$ dB ($\mathsf{snr} \approx 6$ dB). This waterfall corresponds to the tunnel of the EXIT chart being open with a fixed point close to $(1, \mathsf{y})$ point. The iterative process starts at the normalized BICM capacity $\frac{1}{m}\mathsf{C}^{\mathrm{bicm}}_{\mathcal{X}}$ and stalls when approaching the fixed point. Since the mapping cannot reach the $(\mathsf{x}, \mathsf{y}) = (1, 1)$ point in the EXIT chart, the iterative decoder will not converge to vanishing error probability. Although it can significantly improve the performance, it will therefore not have a threshold [96]. The underlying reason is that the demodulator treats the bits as if they were uncoded. Figure 5.1 confirms the above discussion, namely a threshold behavior at low snr together with an irreducible error floor at large snr.

5.4 The Area Theorem

As discovered by Ashikhmin *et al.* EXIT charts exhibit a fundamental property when the extrinsic channel is a BEC [6]: the area under the EXIT curve with a BEC extrinsic channel is related to the rate of the iterative decoding element we are plotting the EXIT chart of. Consider

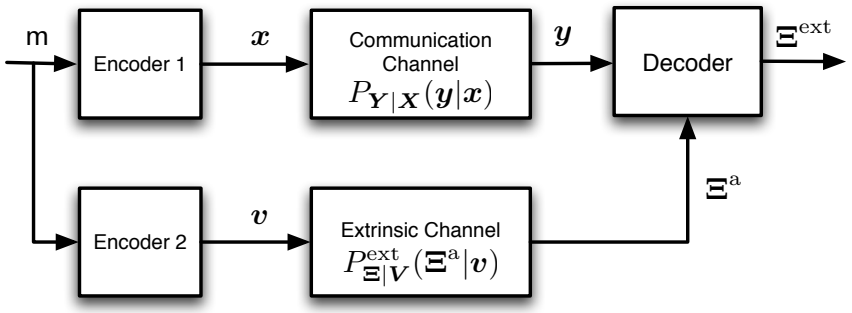


Fig. 5.10 General decoding model for a given component of the iterative decoding process.

the general decoding model for a given component shown in Figure 5.10 [6], where $\boldsymbol{\Xi}^{\text{a}}$ and $\boldsymbol{\Xi}^{\text{ext}}$ denote the a priori and extrinsic messages, respectively. The input of the extrinsic channel is denoted by $\boldsymbol{v} = (v_1, \ldots, v_n)$, and its associated random vector is denoted by $\boldsymbol{V}$. Then,

---

Theorem 5.2 ([6]). The area under the EXIT curve $\mathsf{y}(\mathsf{x})$ of the decoding model described in Figure 5.10 with a BEC extrinsic channel satisfies

$$A = \int_0^1 \mathsf{y}(\mathsf{x})\mathrm{d}\mathsf{x} = 1 - \frac{1}{n} H(\boldsymbol{V}|\boldsymbol{Y}) \tag{5.23}$$

where $H(\boldsymbol{V}|\boldsymbol{Y})$ denotes the conditional entropy of the random vector $\boldsymbol{V}$ given $\boldsymbol{Y}$.

Figure 5.10 encompasses the decoding models shown in Figures 5.7(a) and 5.7(b) for the demapper and decoder of BICM-ID, respectively. In particular, we recover the decoding model of the decoder shown in Figure 5.7(b) by having no communication channel and letting $\boldsymbol{v} = \boldsymbol{x}$, $\boldsymbol{\Xi}^{\text{a}} = \boldsymbol{\Xi}^{\text{dem}\to\text{dec}}$ and $\boldsymbol{\Xi}^{\text{ext}} = \boldsymbol{\Xi}^{\text{dec}\to\text{dem}}$. Applying Theorem 5.2 to the decoder extrinsic model yields the following result.

Corollary 5.1 ([6]). Consider the EXIT curve of the decoder of the binary code $\mathcal{C}$ of length n and rate r in a BICM-ID scheme, $\mathsf{x} = \mathsf{exit}_{\text{dec}}(\mathsf{y})$. Then,

$$\int_0^1 \mathsf{x}(\mathsf{y})\mathrm{d}\mathsf{y} = 1 - r \tag{5.24}$$

and

$$A_{\text{dec}} = \int_0^1 \mathsf{y}(\mathsf{x})\mathrm{d}\mathsf{x} = r. \tag{5.25}$$

An example of the above result with the $(5,7)_8$ convolutional code is shown in Figure 5.11 (left). From Figure 5.10, we also recover the decoding model of the demapper shown in Figure 5.7(a) by letting Encoder 2 be equal to Encoder 1, $\boldsymbol{v} = \boldsymbol{x}$, $\boldsymbol{\Xi}^{\text{a}} = \boldsymbol{\Xi}^{\text{dec}\to\text{dem}}$ and $\boldsymbol{\Xi}^{\text{ext}} = \boldsymbol{\Xi}^{\text{dem}\to\text{dec}}$. Applying Theorem 5.2 to the demapper case yields the following result.

Corollary 5.2. Consider the EXIT curve of the demapper in a BICM-ID scheme, $\mathsf{y} = \mathsf{exit}_{\text{dem}}(\mathsf{x})$. Then,

$$A_{\text{dem}} = \int_0^1 \mathsf{y}(\mathsf{x})\mathrm{d}\mathsf{x} = \frac{1}{m}\mathsf{C}_{\mathcal{X}}^{\text{cm}}. \tag{5.26}$$

Proof. From Figure 5.7(a), since the encoder is a one-to-one mapping, we can replace $\boldsymbol{V}$ by $\boldsymbol{X}$ to obtain that [6, p. 2662]

$$\begin{aligned} A_{\text{dem}} &= \int_0^1 \mathsf{y}(\mathsf{x})\mathrm{d}\mathsf{x} && (5.27)\\ &= 1 - \frac{1}{n}H(\boldsymbol{X}|\boldsymbol{Y}) && (5.28)\\ &= \frac{1}{m}\left(m - \frac{1}{N}H(\boldsymbol{X}|\boldsymbol{Y})\right) && (5.29)\\ &= \frac{1}{m}\mathsf{C}_{\mathcal{X}}^{\text{cm}}. && (5.30) \end{aligned}$$

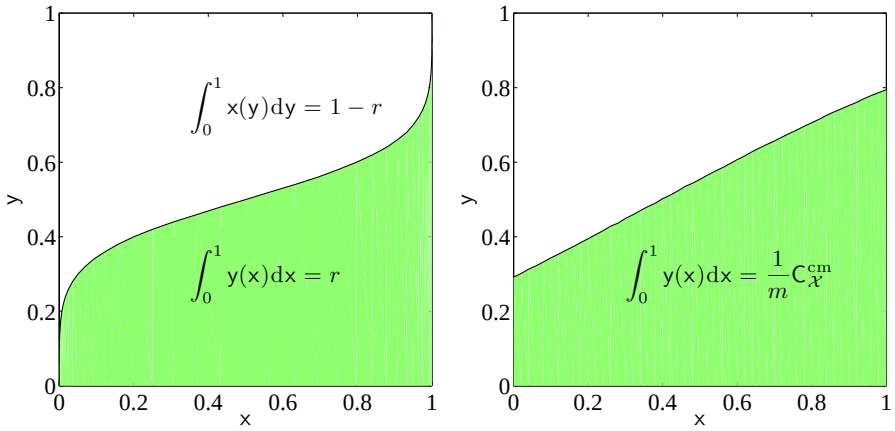
□

An example of the above result with a set partitioning demapper in the AWGN channel is shown in Figure 5.11 (right).

5.4.1 Code Design Considerations

The practical significance of the area theorem lies in the following observation. As mentioned earlier, as long as the demapper EXIT curve is above the code EXIT curve, i. e.,

$$\mathsf{exit}_{\text{dem}}(\mathsf{x}) > \mathsf{exit}_{\text{dec}}^{-1}(\mathsf{x}), \tag{5.31}$$



(a) Convolutional code. (b) Demapper.

Fig. 5.11 Area theorem for the convolutional code $(5,7)_8$ (left) and for a set partitioning demapper in an AWGN channel with $\mathsf{snr} = 6$ dB (right).

the decoder improves its performance with iterations eventually yielding vanishing error probability. This implies that as long as $A_{\mathrm{dem}} > A_{\mathrm{dec}}$ BICM-ID will improve its performance with iterations. Combining the results of Corollaries 5.1 and 5.2 we see that this happens if

$$\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}} > rm = R. \tag{5.32}$$

which is a satisfying result. Furthermore, this result suggests that any area gap between the two EXIT curves translates in a rate loss with respect to the capacity $\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}$ [6]. The implications for code design are clear: one should design codes and mappers so that the EXIT curve of the code matches (fits) with that of the demapper. This *matching condition* was illustrated in [82, 83, 96] for binary LDPC codes.

For the demapper EXIT curve, it is not difficult to show that

$$\mathsf{exit}_{\mathrm{dem}}(\mathsf{x} = 1) < 1 \tag{5.33}$$

for any binary labeling μ and finite snr. As discussed earlier, this implies that, even with perfect extrinsic side information the demapper is not able to correctly infer the value of the bit i. e., the demapper EXIT curve cannot reach the point $(1, 1)$. This is due to the fact that the demapper treats the bits as if they were uncoded. As $(1, 1)$ cannot be a

fixed point, the demapper and code curves will cross at some point and there is no threshold effect. Also, since the area is fixed, the higher the point $(1, \mathsf{y})$ is, the lower the point $(0, \mathsf{y})$, i. e., BICM capacity $\mathsf{C}_{\mathcal{X}}^{\text{bicm}}$.

5.5 Improved Schemes

In this section, we discuss some BICM-ID schemes whose demapper EXIT curve does not treat the bits as uncoded. In particular, we consider LDPC and RA-based constructions and show that significant gains can be achieved. These constructions use in one way or another coded mappers, i. e., mappers with memory. The underlying memory in the mapping will show to be the key to achieve the $(1, 1)$ point. The following result lies at the heart of improved constructions.

Proposition 5.1 (Mixing property, [6, 120, 128]). Suppose that encoder 1 in Figure 5.10 consists of the direct product of multiple codes, i. e., $\mathcal{C} = \mathcal{C}_1 \times \ldots \times \mathcal{C}_{d_{\max}}$ each of length n_i, i. e., $\sum_{i=1}^{d_{\max}} n_i = n$. Then,

$$\mathsf{y}(\mathsf{x}) = \sum_{i=1}^{d_{\max}} \frac{n_i}{n} \mathsf{y}_i(\mathsf{x}) \tag{5.34}$$

where $\mathsf{y}_i(\mathsf{x})$ is the EXIT curve of code $\mathcal{C}_i$.

The EXIT curve of a code mixture is the sum of the individual EXIT curves, appropriately weighted by the length fractions corresponding to each code. The above result is usually employed to design irregular code constructions by optimizing the code parameters [96,100,120,121,128].

5.5.1 Standard concatenated codes with BICM-ID

A possibility to improve over the above BICM-ID is to use a powerful binary code like a turbo code or an LDPC code. The corresponding block diagram for this scheme is shown in Figure 5.12. The decoding of this scheme involves message passing between the mapper and inner code, as well as between the inner and outer codes. This typically results in 3-dimensional EXIT charts and multiple scheduling algorithms, i. e., how decoding iterations of each type are scheduled [25,27,55]. Hence the design of such schemes based on curve fitting is potentially complicated.

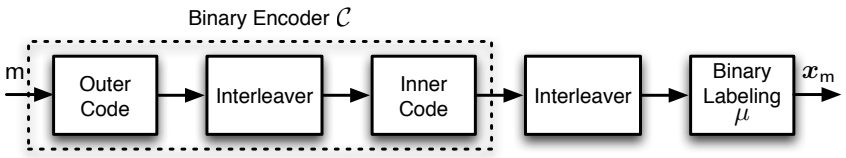


Fig. 5.12 Block diagram of concatenated codes with BICM-ID.

Although suboptimal, another way of looking at the problem considers a scheduling that allows the decoder of the powerful code to perform as many iterations as required, followed by a demapping iteration. In this case the EXIT curve of the code becomes a step function [111] and hence, due to the shape of the demapper EXIT curve (see e.g. Figure 5.9), the BICM-ID scheme would yield vanishing error probability as long as $R < \mathsf{exit}_{\mathrm{dem}}(\mathsf{x} = 0) = \mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}}$, i. e., the leftmost point of the demapper EXIT curve.

Since the largest $\mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}}$ is achieved for Gray mapping, it seems natural to design powerful binary codes for Gray mapping. As Gray mapping does not improve much its EXIT transfer characteristics through iterations, we could simplify the scheme —and reduce the decoding complexity— by not performing iterations at the demapper, e. g. the BICM decoding described in Chapter 3. This construction can be advantageous for high rates, where the BICM capacity is close to the coded modulation capacity. The FG is shown in Figure 5.13.

Our next example is a binary irregular RA codes [58, 100] with grouping factor a, for which a single-parity-check code of rate a is introduced before the accumulator. In Figure 5.14 we show the results for grouping factor $a = 2$, no demapping iterations, and 16-QAM with Gray mapping. A number of optimization methods are possible and we here use standard linear programming [100, 120, 121, 127]. Since no iterations are performed at the demapper, the observation messages correspond the classical BICM decoder, without extrinsic side information. Observe that the simulated error probability in Figure 5.15 ($N = 125000$ and 100 decoding iterations) matches well the EXIT analysis predictions, especially the threshold, which the EXIT analysis

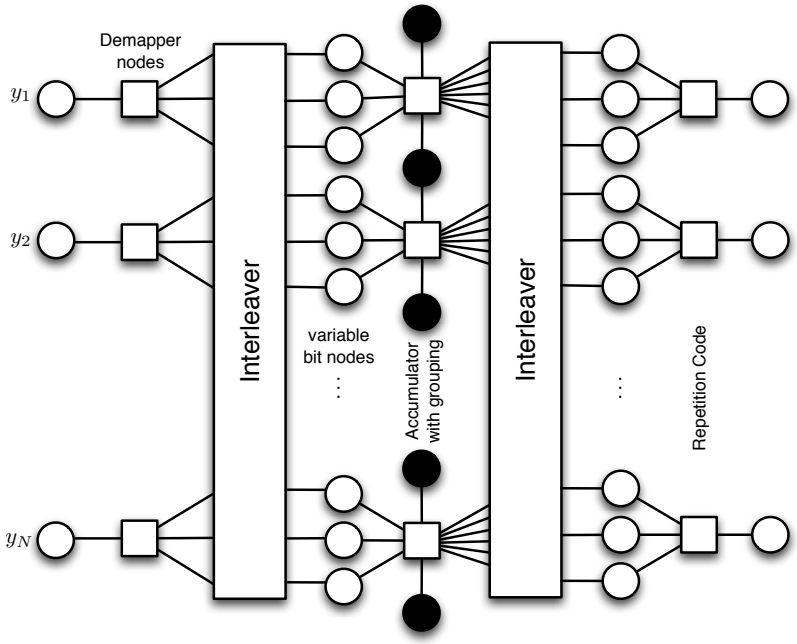


Fig. 5.13 Regular rate $r_{\text{rep}} = \frac{1}{3}$ RA BICM-ID factor graph with regular grouping with factor $a = 2$ and $m = 3$. The overall transmission rate is $R = 2$.

locates at 0.2157 dB from the BICM capacity.

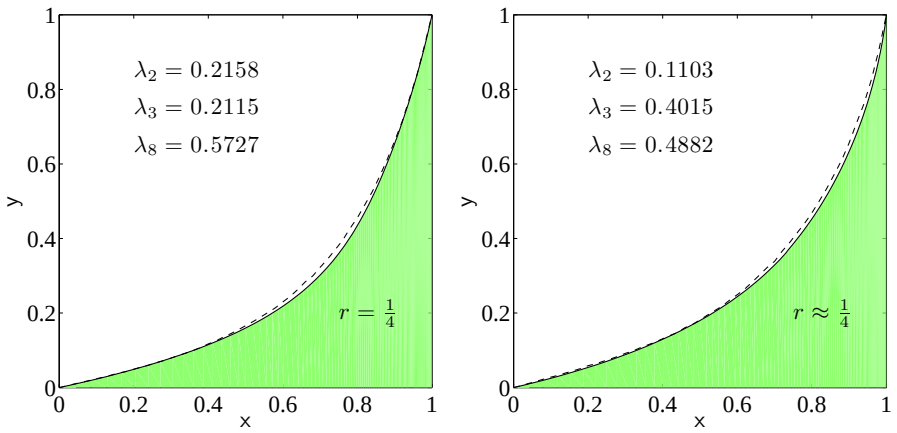
5.5.2 LDPC BICM-ID

Let us first consider the concatenation of an LDPC code with the binary labeling. The FG of this construction is shown in Figure 5.16 for a regular $(3, 6)$ LDPC code. In general, we consider irregular ensembles with left and right degree distributions respectively given by [96]

$$\lambda(z) = \sum_i \lambda_i z^{i-1} \quad \text{and} \quad \rho(z) = \sum_i \rho_i z^{i-1}, \tag{5.35}$$

where λ_i (resp. ρ_i) is the fraction of edges in the LDPC graph connected to variable (resp. check) nodes of degree i. With this notation, the average variable and check node degrees are given by

$$\bar{d}_{\text{v}} \triangleq \frac{1}{\int_0^1 \lambda(z)dz} \quad \text{and} \quad \bar{d}_{\text{c}} \triangleq \frac{1}{\int_0^1 \rho(z)dz}, \tag{5.36}$$



(a) BICM BEC extrinsic channel. (b) BICM AWGN extrinsic channel.

Fig. 5.14 EXIT chart fit with an irregular RA code a with BEC (upper left) and AWGN (upper right) extrinsic channel, for 16-QAM with Gray mapping in the AWGN channel with $\mathsf{snr} = 6$ dB using $a = 2$. Binary accumulator; no iterations at the demapper. In dashed lines the accumulator-demapper curve; in solid lines the irregular repetition code. Codes are at 0.2157 dB from the BICM capacity.

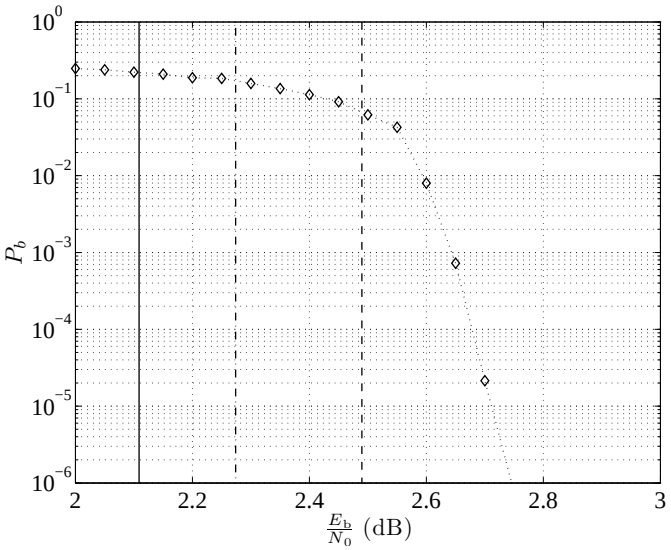


Fig. 5.15 Error probability for optimal AWGN code with $N = 125000$ and 100 decoding iterations. In solid, dash-dotted, and dashed vertical lines the coded modulation capacity, BICM capacity, and code threshold, respectively. Dotted line with diamonds is the simulation of the optimal AWGN code.

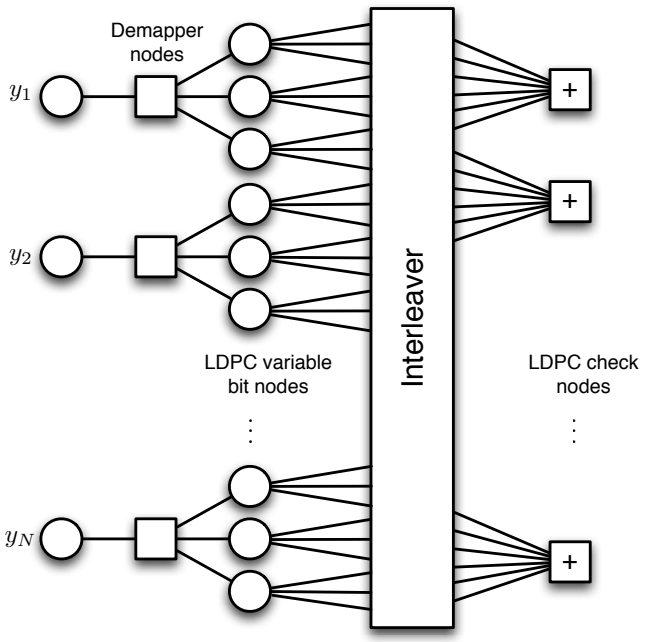


Fig. 5.16 Regular $(3, 6)$ LDPC BICM-ID factor graph with $m = 3$.

respectively. The design rate of this LDPC code ensemble is thus

$$r = 1 - \frac{\bar{d}_{\mathrm{v}}}{\bar{d}_{\mathrm{c}}}, \tag{5.37}$$

and the overall rate of the construction is $R = mr$. The interest of this improved BICM-ID construction is the lack of interleaver between the LDPC and the mapping. The decoding algorithm decodes the variable node bits from the LDPC directly and passes the messages along to the check node decoder [121]. The corresponding EXIT area result for this construction is given in the following.

Corollary 5.3 ([6]). Consider the joint EXIT curve of the demapper and LDPC variable nodes $\mathsf{y} = \mathsf{exit}^{\mathrm{ldpc}}_{\mathrm{dem,v}}(\mathsf{x})$ in a BICM-ID scheme with an LDPC code with left and right edge degree distributions given by

$\lambda(z)$ and $\rho(z)$, respectively. Then

$$A_{\mathrm{dem,v}}^{\mathrm{ldpc}} = 1 - \bar{d}_{\mathrm{v}} \left(1 - \frac{1}{m} \mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}\right). \tag{5.38}$$

Furthermore, the area under the check node decoder EXIT curve $\mathsf{x} = \mathsf{exit}_{\mathrm{dec,c}}^{\mathrm{ldpc}}(\mathsf{y})$ is given by

$$\int_0^1 \mathsf{x}(\mathsf{y})\mathrm{d}\mathsf{y} = \frac{1}{\bar{d}_{\mathrm{c}}}. \tag{5.39}$$

Then we have that

$$A_{\mathrm{dec,c}}^{\mathrm{ldpc}} = \int_0^1 \mathsf{y}(\mathsf{x})\mathrm{d}\mathsf{x} = 1 - \frac{1}{\bar{d}_{\mathrm{c}}}. \tag{5.40}$$

Since we need that the areas $A_{\mathrm{dem,v}}^{\mathrm{ldpc}} > A_{\mathrm{dec,c}}^{\mathrm{ldpc}}$ to ensure convergence to vanishing error probability, the above result implies that

$$\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}} > m \left(1 - \frac{\bar{d}_{\mathrm{v}}}{\bar{d}_{\mathrm{c}}}\right) = mr = R \tag{5.41}$$

where r denotes here the design rate of the LDPC code. That is, again we recover a matching condition, namely, that the overall design spectral efficiency should be less than the channel capacity.

Bennatan and Burshtein [10] proposed a similar construction based on non-binary LDPC codes, where the variable nodes belong to $\mathbb{F}_M$, followed by a mapping over a signal constellation of size M. The messages passed along the edges of the graph correspond to M-ary variables. The random coset technique described in Section 4.1 enabled them to consider i.i.d. messages and define density evolution and EXIT charts.

5.5.3 RA BICM-ID

Similarly to the LDPC BICM-ID construction, we can design an improved BICM-ID scheme based on RA codes [93,120,127]. In particular, we consider the accumulator shown in Figure 5.17 of rate $\frac{m}{m}$. In order to design a more powerful code, we incorporate *grouping*, i. e., introduce a check node of degree a before the accumulator. As an outer code, we consider irregular repetition code. We use the same notation to refer

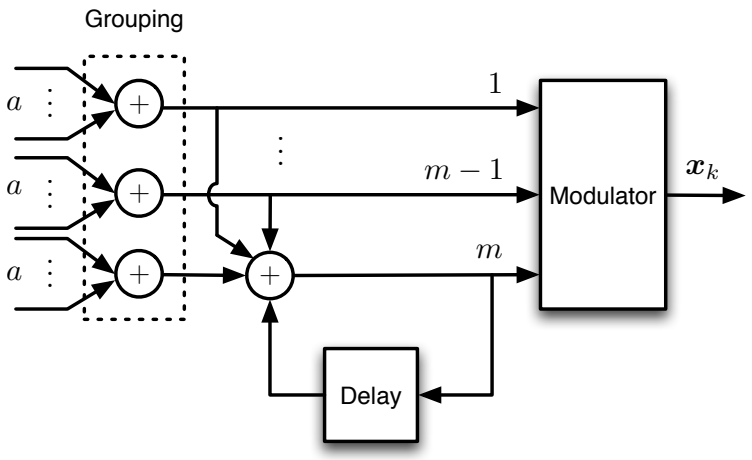


Fig. 5.17 Accumulator of rate $r_{\text{acc}} = a\frac{m}{m}$, with grouping factor a.

to the degree distributions of the code: $\lambda(z)$ denotes the edge degree distribution of the irregular repetition code while $\rho(z)$ denotes the edge degree distribution of the accumulators with grouping. Let also

$$\bar{d}_{\text{rep}} \triangleq \frac{1}{\int_0^1 \lambda(z)dz} \quad \text{and} \quad \bar{d}_{\text{acc}} \triangleq \frac{1}{\int_0^1 \rho(z)dz}, \tag{5.42}$$

denote the average degrees of the repetition code and accumulator, respectively. The rate of the outer repetition code is $r_{\text{rep}} = \bar{d}_{\text{rep}}$. The rate of the accumulator is $r_{\text{acc}} = m\,\bar{d}_{\text{acc}}$. The overall rate of the construction is hence $R = m\,\bar{d}_{\text{rep}}\,\bar{d}_{\text{acc}}$. The overall FG is shown in Figure 5.18 for a regular repetition code and regular grouping.

Figure 5.19 shows the EXIT curves for the accumulator in Figure 5.17 with 8-PSK and various grouping factors. The curves not only reach the $(1, 1)$ point, but also start at the point $(0, 0)$ when $a > 1$, and iterative decoding would not start. In order to overcome this problem, we can introduce a small fraction of bits of degree one [120] or use code doping [118], i. e., introduce a small fraction of pilot bits known to the receiver.

Following [6, Example 26], we have the following area property.

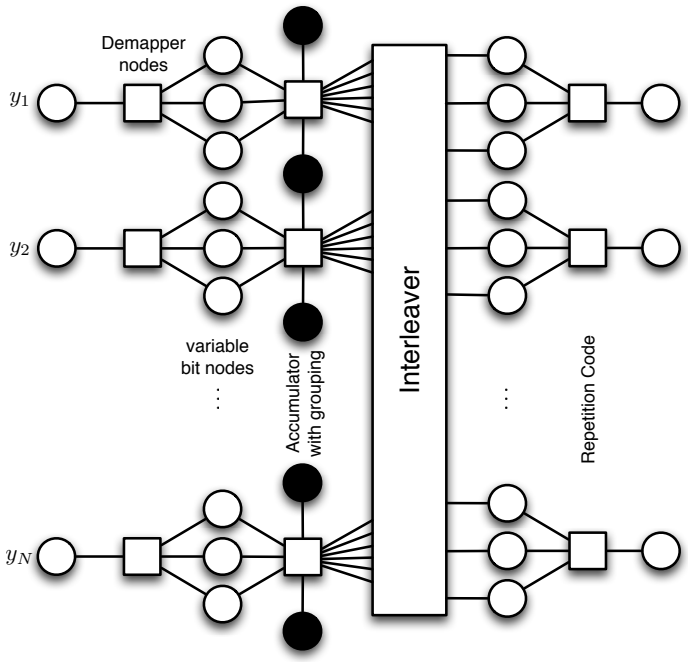


Fig. 5.18 Regular rate $r_{\text{rep}} = \frac{1}{3}$ RA BICM-ID factor graph with regular grouping with factor $a = 2$ and $m = 3$. The overall transmission rate is $R = 2$.

Corollary 5.4 ([6]). Consider the EXIT curve $\mathsf{y} = \mathsf{exit}^{\text{ra}}_{\text{dem,acc}}(\mathsf{x})$ of the accumulator shown in Figure 5.17 combined with irregular grouping with edge degree distribution $\rho(z)$. Then

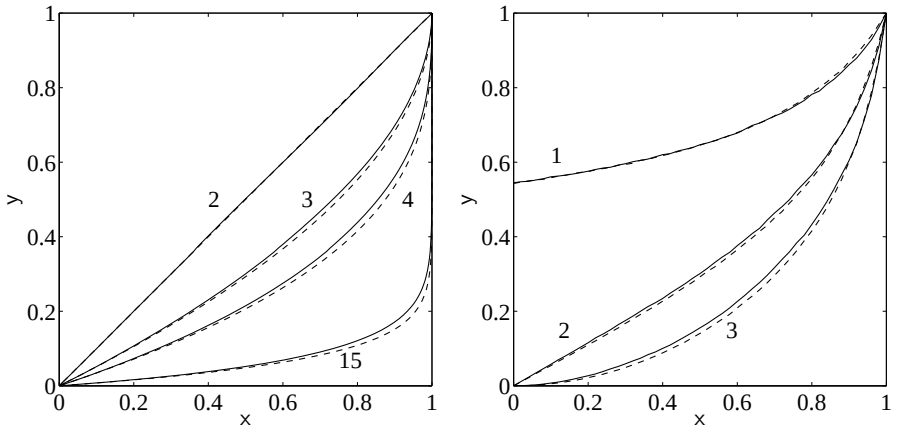
$$A^{\text{ra}}_{\text{dem,acc}} = \frac{1}{m\,\bar{d}_{\text{acc}}} \mathsf{C}^{\text{cm}}_{\mathcal{X}}. \tag{5.43}$$

Furthermore, the area under the EXIT curve of the outer irregular repetition code with edge degree distribution $\lambda(z)$ is given by

$$A^{\text{ra}}_{\text{dec,rep}} = \bar{d}_{\text{rep}}. \tag{5.44}$$

Since we again require the areas $A^{\text{ra}}_{\text{dem,acc}} > A^{\text{ra}}_{\text{dec,rep}}$ to ensure convergence to vanishing error probability, we have the matching condition

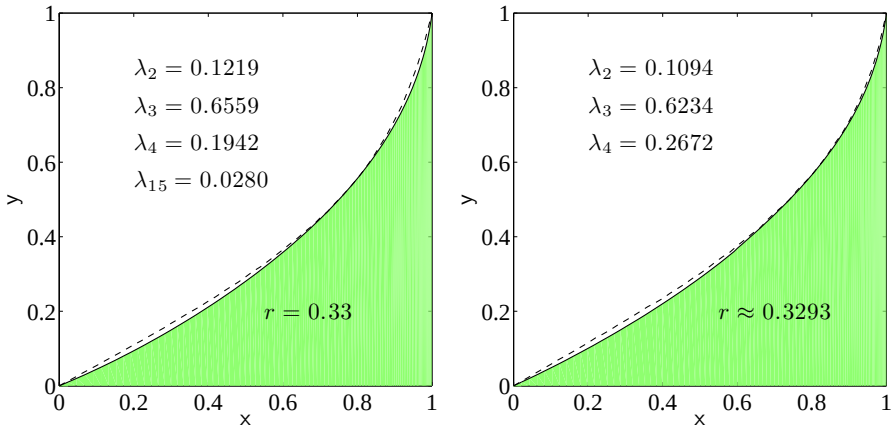
$$\mathsf{C}^{\text{cm}}_{\mathcal{X}} > m\,\bar{d}_{\text{acc}}\bar{d}_{\text{rep}} = R. \tag{5.45}$$



(a) Repetition codes. (b) Accumulators.

Fig. 5.19 EXIT curves for repetition codes with repetition factors $2, 3, 4, 15$ (left) and the accumulator shown in Figure 5.17 with 8-PSK and grouping factors $1, 2, 3$ (right). Solid and dashed lines correspond to AWGN and BEC extrinsic channels, respectively.

We now show a design example of RA-BICM-ID with 8-PSK. For simplicity, we consider regular grouping with $a = 2$, although irregular grouping would lead to improved codes with a better threshold. We use a curve-fitting approach to optimize the irregular code distribution and match it to that corresponding to the accumulator. As in Section 5.5.1, we optimize the design by using standard linear programming. Figure 5.20 shows the results of the curve fitting design at $\mathsf{snr} = 6$ dB, which is close to the capacity for a transmission rate $R = 2$. Results with AWGN and BEC extrinsic channels are shown. The rate of the overall schemes with a BEC extrinsic channel is $R = 1.98$ while for AWGN is $R = 1.9758$, which are respectively at 0.34 dB and 0.36 dB from capacity. The corresponding error probability simulation is shown in Figure 5.21. As we observe, the error probability matches well the EXIT predictions. Also, the code designed with a BEC extrinsic channel is 0.1 dB away from that designed for the AWGN channel. Hence, using the BEC as a design channel yields good codes in a similar way to [92] for LDPC codes.



(a) BICM-ID BEC extrinsic channel.

(b) BICM-ID AWGN extrinsic channel.

Fig. 5.20 EXIT chart fit with an irregular RA code a with BEC (left) and AWGN (right) extrinsic channel, for 8-PSK with Gray mapping in the AWGN channel with $\mathsf{snr} = 6$ dB. Grouping factor $a = 2$ and non-binary accumulator from Figure 5.17. In dashed lines the accumulator-demapper curve; in solid lines the irregular repetition code.

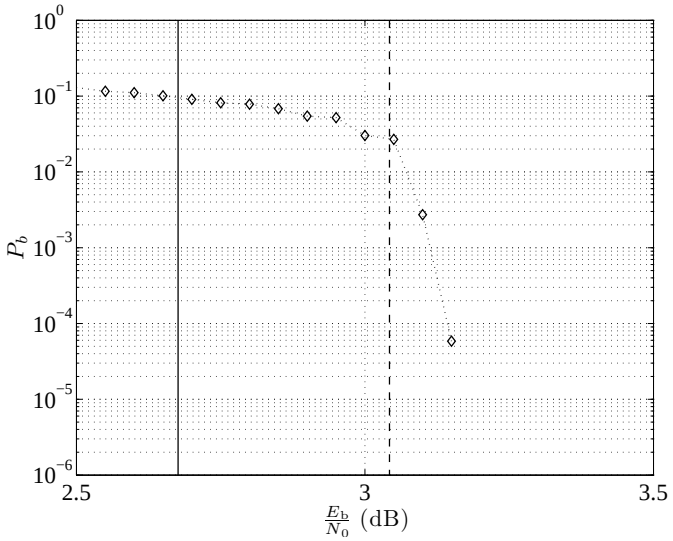


Fig. 5.21 Error probability for optimal AWGN code with $N = 100000$ and 100 decoding iterations. In solid and dashed vertical lines the coded modulation capacity and code threshold, respectively; actual simulation in dotted line with diamonds.

5.6 Concluding Remarks and Related Work

In this Chapter, we have reviewed iterative decoding of BICM. In particular, we have studied the properties of BICM-ID when the interleaver length is large. In this case, we have shown that the corresponding analysis tool to characterize the error probability of BICM-ID is density evolution, i. e., the characterization of how the density of the bit scores evolve through the iterations. A more compact and graphical, yet not exact, method for understanding the behavior of BICM-ID is EXIT chart analysis. A particularly interesting property of EXIT chart analysis is the area theorem for BEC extrinsic side information, which relates the area under the EXIT curve of a given component to the its rate. We have shown how to particularize this property to BICM-ID, and we have illustrated how the area property leads to the efficient design of capacity-approaching BICM-ID schemes. The optimized RA-BICM-ID codes could be further improved by introducing irregularity at the accumulator side, and by possibly optimizing the connections at the accumulator and the constellation mapping. A number of improved BICM-ID code constructions have been proposed in the literature [93,120,121,127]. Based on the concepts illustrated in this Chapter, turbo-coded modulation schemes [99] can also be put into the BICM-ID framework. Furthermore, a number of optimized MLC have also been proposed [54, 102]. We argue, however, that the optimization process and decoding is simpler for BICM-ID than for MLC, since there is only one code to be optimized.

Since the discovery of the improved performance of BICM-ID with mappings different from Gray, a number of works have focused on finding good mappings under a variety of criteria [26, 103, 113, 125, 131]. Further, multidimensional mappings have been shown to provide performance advantages with BICM-ID [106, 107, 124]. A popular mapping design criterion has been to optimize the perfect extrinsic side information point of the EXIT chart $(1, \mathsf{y})$, in order to make y as large as possible. Note that, according to the area theorem, larger y in the $(1, \mathsf{y})$ point, which corresponds to a lower error floor, implies lower y in the $(0, \mathsf{y})$ point, which implies a larger convergence SNR. A number of works, using bounding techniques similar to those pre-

sented in Chapter 4 using a metric with perfect extrinsic side information, have predicted the error floor of BICM-ID with convolutional codes [30, 63, 103, 113].

Recall that the area theorem yields exact results only when the extrinsic side information comes from a BEC channel, and only approximate results when the extrinsic channel is an AWGN. In [14], an MSE-based transfer chart was proposed. Building on the fundamental relationship between mutual information and MMSE [51], Bhattad and Narayanan showed that the area property holds exactly for the MSE-based chart when the extrinsic channel is an AWGN. Although the analysis that one might obtain in this case is still approximate, it seems reasonable to assume that it might be more accurate than the BEC analysis in terms of thresholds. Finally, generalized EXIT (GEXIT) charts were proposed in [82] as an EXIT chart where the area property is imposed by definition. GEXIT charts therefore exactly characterize the iterative decoding process. The slight drawback of GEXIT charts is that computing them is potentially demanding.

In this Chapter, we have only touched upon infinite-length analysis of BICM-ID. Finite-length analysis of BICM-ID is an open problem which is of significant interest due to the fact that practical systems, especially those meant to transmit real-time data, do not usually employ such large interleavers. Successful finite-length analysis techniques for LDPC codes, such as finite-length scaling [3,4,96], find BICM-ID as a natural application. The analysis of the finite-length case, could also be performed by employing improved bounds to the error probability, like the Duman-Salehi or the tangential-sphere bound. The tangential-sphere bound was used in [44] to characterize the error probability of finite-length binary RA codes with grouping.

5.A Density Evolution Algorithm for BICM-ID

Since the message $\Xi_i^{\mathrm{dem}\to\mathrm{dec}}$ is a function of the messages propagated upwards (from bottom to top) in the oriented neighborhood $\mathcal{N}^{\ell}(\mu_k \to c_i)$, it follows that the limiting probability density of $\Xi_i^{\mathrm{dem}\to\mathrm{dec}}$ can be recursively computed as follows. First, initialize the pdf of the messages $\Xi^{\mathrm{dec}\to\mathrm{dem}}$ (we drop the time index for simplicity of notation) to a single

mass-point at zero. This represents the fact that, at the beginning of the BICM-ID process, no information is available from the decoder to the demapper. Then, alternatively, compute the pdf of the messages $\Xi^{\mathrm{dem}\to\mathrm{dec}}$ given the pdf of the messages $\Xi^{\mathrm{dec}\to\mathrm{dem}}$ and then compute the pdf of the messages $\Xi^{\mathrm{dec}\to\mathrm{dem}}$ given the pdf of the messages $\Xi^{\mathrm{dem}\to\mathrm{dec}}$ obtained at the previous step until the exit condition is met.

Several criteria for the exit condition are possible and have been considered in the literature. A common criterion is to exit if the resulting error probability does not change significantly through the iterations.

Generally, it is much more convenient to work with cumulative distribution functions rather than with pdfs (we refer to pdfs here since the algorithm is called *density* evolution). In fact, the calculation of the empirical cdf of a data set can easily be accomplished, and also the sampling from a distribution with given cdf is well-known and simple to implement.

Algorithm 1: Density Evolution

Initialization *Let* $P_{\Xi}^{\mathrm{dec}\to\mathrm{dem},(0)}(\xi)=\delta(\xi)$ *(a single mass-point at zero) and let the iteration index* $\ell=1$.

Demapper Step *Generate a large sample of the random variable* $\Xi^{\mathrm{dem}\to\mathrm{dec},(\ell)}$ *as follows*

(1) *Generate random scrambling symbols* $\boldsymbol{d}=(d_1,\ldots,d_m)$.
(2) *Let* $x=\mu(\boldsymbol{d})$, *where* μ *is the modulator mapping.*
(3) *Generate* $y\sim P_{Y|X}(y|x)$.
(4) *Generate the samples* $\{\xi_j^{\mathrm{dec}\to\mathrm{dem},(\ell-1)}: j=1,\ldots,m\}$ *by sampling i.i.d. from the pdf* $P_{\Xi}^{\mathrm{dec}\to\mathrm{dem},(\ell-1)}(\xi)$.
(5) *Generate a label position* $j\in\{1,\ldots,m\}$ *with uniform probability.*
(6) *Compute the realization of* $\Xi^{\mathrm{dem}\to\mathrm{dec}}$ *as*

$$\xi^{\mathrm{dem}\to\mathrm{dec},(\ell)}=\log\frac{\displaystyle\sum_{x'\in\mathcal{X}_{d_j}^{j}}P_{Y|X}(y|x')\;\mathrm{e}^{\sum_{j'\neq j}b_{j'}(x')(-1)^{d_{j'}}\xi_{j'}^{\mathrm{dec}\to\mathrm{dem},(\ell-1)}}}{\displaystyle\sum_{x'\in\mathcal{X}_{d_j}^{j}}P_{Y|X}(y|x')\;\mathrm{e}^{\sum_{j'\neq j}b_{j'}(x')(-1)^{d_{j'}}\xi_{j'}^{\mathrm{dec}\to\mathrm{dem},(\ell-1)}}}$$

(7) *Compute the empirical pdf of the generated set of values from step (6) and call it* $P_{\Xi}^{\mathrm{dem}\to\mathrm{dec},(\ell)}(\xi)$.

Decoder Step *Generate a large sample of the random variable* $\Xi^{\mathrm{dec}\to\mathrm{dem},(\ell)}$ *as follows*

(1) *Generate the samples* $\{\xi_i^{\mathrm{dem}\to\mathrm{dec},(\ell)}: i=1,\ldots,L\}$ *by sampling i.i.d. from the pdf* $P_{\Xi}^{\mathrm{dem}\to\mathrm{dec},(\ell)}(\xi)$, *where* L *is the length of a sufficiently large trellis section of the code.*
(2) *Apply the BCJR algorithm to the code trellis with input* $\{\xi_i^{\mathrm{dem}\to\mathrm{dec},(\ell)}: i=1,\ldots,L\}$ *and compute the corresponding extrinsic messages* $\{\xi_i^{\mathrm{dec}\to\mathrm{dem},(\ell)}: i=1,\ldots,L\}$.
(3) *Compute the empirical pdf of the generated set of values and call it* $P_{\Xi}^{\mathrm{dec}\to\mathrm{dem},(\ell)}(\xi)$.

Exit Condition *If the exit condition is met, then exit. Otherwise, let* $\ell\leftarrow\ell+1$ *and go to the Demapper step.*

6

Applications

In this Chapter we briefly discuss some applications of BICM to cases of practical relevance not explicitly included in our presentation. In particular, we review current work and outline how to extend the results we presented throughout the monograph to non-coherent detection, block-fading, multiple-input multiple-output (MIMO) channels and non-standard channels such as the exponential-noise channel.

6.1 Non-Coherent Demodulation

Orthogonal modulation with non-coherent detection is a practical choice for situations where the received signal phase cannot be reliably estimated and/or tracked. Important examples include military communications using fast frequency hopping, airborne communications with high Doppler shifts due to significant relative motion of the transmitter and receiver, and high phase noise scenarios, due to the use of inexpensive or unreliable local oscillators. Common choices of implementation for the modulator are pulse-position modulation (PPM) or frequency-shift keying (FSK) [95].

BICM with orthogonal modulation was first studied by Caire *et al.*

in [29]. The study in [29] included BICM capacity, cutoff rate as well as error probability considerations. Valenti and Cheng applied BICM-ID techniques to orthogonal modulation with non-coherent detection with turbo-codes [132]. In [48], the design of capacity approaching codes is considered using the improved construction based on RA codes. Capacity-approaching codes within tenths of dB of capacity were found, also with suboptimal decoding metrics.

The application of our main results to orthogonal modulation is straightforward. Symbols $x \in \mathbb{R}^M$ belong now to an orthogonal modulation constellation, i. e., $x \in \mathcal{X} \stackrel{\Delta}{=} \{e_1, \ldots, e_M\}$, where $e_k = (\underbrace{0, \ldots, 0}_{k-1}, 1, \underbrace{0, \ldots, 0}_{M-k-1})$ is a vector has all zeros except in the k-th position, where there is a one. The received signal over a fading channel can still be expressed by (2.6), where now $y, z \in \mathbb{C}^M$, i. e., they are vectors of dimension M. The channel transition probability becomes

$$P_{Y|X,H}(y|x,h) = \frac{1}{\pi^M} e^{-\|y - h\sqrt{\mathsf{snr}}\ x\|^2}. \tag{6.1}$$

Depending on the knowledge of the channel coefficients at the receiver, the decoding metric might vary. In particular, for coherent detection the symbol decoding metric for hypothesis $x = e_k$ satisfies $q(x = e_k, y) \propto P_{Y|X,H}(y|x,h)$. When no knowledge of the carrier phase is available at the receiver, then the symbol decoding metric for hypothesis $x = e_k$ with coded modulation becomes [95]

$$q(x = e_k, y) \propto I_0\left(2\sqrt{\mathsf{snr}}|h|y_k\right) \tag{6.2}$$

where $I_0(.)$ is the zero-th order Bessel function of the first kind [1] and with some abuse of notation we let y_k denote the k-th entry of the received signal vector y. The bit metrics are

$$q_j\big(b_j(x) = b, y\big) = \sum_{k' \in \mathcal{X}_b^j} I_0\left(2\sqrt{\mathsf{snr}}|h|y_{k'}\right). \tag{6.3}$$

where now $\mathcal{X}_b^j$ is the set of indices from $\{0, \ldots, M-1\}$ with bit b in the j-th position. All general results from the previous chapters can be applied to orthogonal modulation by properly adapting the metric. Also, all integrals over y are now M-dimensional integrals. As an example,

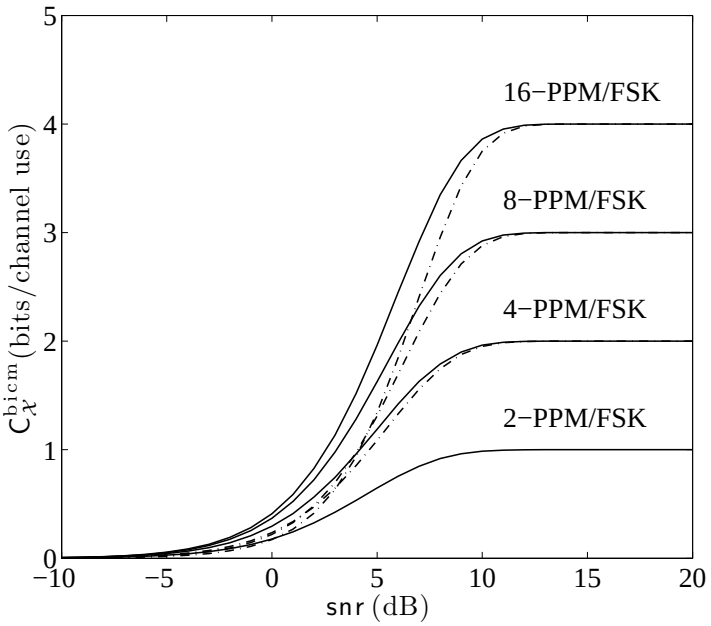


Fig. 6.1 Coded modulation capacity (solid lines), BICM capacity (dash-dotted lines) for orthogonal modulation (PPM/FSK) with non-coherent detection in the AWGN channel.

Figures 6.1 and 6.2 show the coded modulation and BICM capacities for the AWGN channel and the fully-interleaved Rayleigh fading channel with non-coherent detection, respectively. As we observe, the capacity loss of BICM with respect to coded modulation is somewhat larger than in the QAM/PSK modulation case with coherent-detection.

Another common choice for systems where the carrier phase cannot be reliably estimated and/or tracked is differential modulation where a reference symbol is included. Common choices for practical implementations in this case include differential PSK or block-differential PSK [53, 88–91].

6.2 Block-Fading

The block-fading channel [16, 87] is a useful channel model for a class of time- and/or frequency-varying fading channels where the duration of a block-fading period is determined by the product of the channel

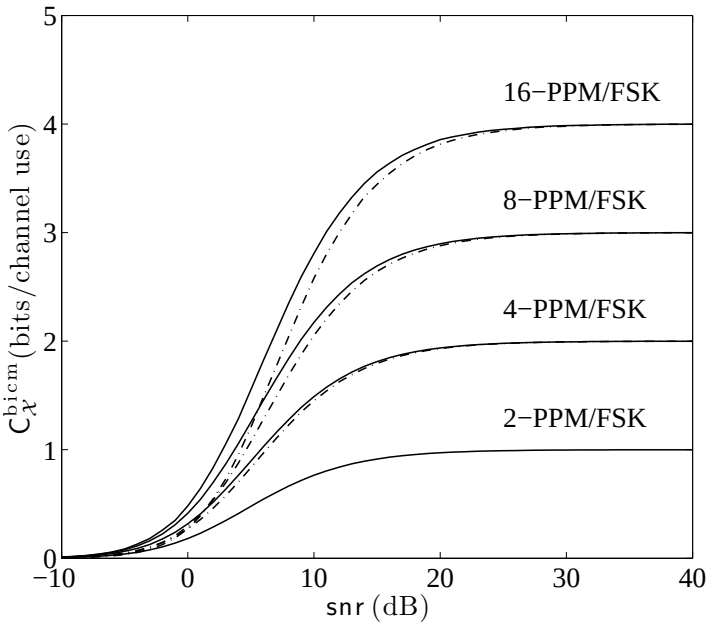


Fig. 6.2 Coded modulation capacity (solid lines), BICM capacity (dash-dotted lines) for orthogonal modulation (PPM/FSK) with non-coherent detection in the fully-interleaved Rayleigh fading channel.

coherence bandwidth and the channel coherence time [95]. Within a block-fading period, the channel fading gain remains constant. In this setting, transmission typically extends over multiple block-fading periods. Frequency-hopping schemes as encountered in the Global System for Mobile Communication (GSM) and the Enhanced Data GSM Environment (EDGE), as well as transmission schemes based on OFDM, can also conveniently be modeled as block-fading channels. The simplified model is mathematically tractable, while still capturing the essential features of the practical transmission schemes over fading channels.

Denoting the number of block per codeword by B, the codeword transition probability in a block-fading channel can be expressed as

$$P_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}) = \prod_{i=1}^{B}\prod_{k=1}^{N} P_{Y_i|X_i,H_i}(y_{i,k}|x_{i,k},h_i), \tag{6.4}$$

where $x_{i,k}, y_{i,k} \in \mathbb{C}$ are respectively the transmitted and received sym-

bols at block i and time k, and $h_i \in \mathbb{C}$ is the channel coefficient corresponding to block i. The symbol transition probability is given by

$$P_{Y_i|X_i,H_i}(y_{i,k}|x_{i,k},h_i) \propto e^{-|y_{i,k}-h_i\sqrt{\mathsf{snr}}x_{i,k}|} \tag{6.5}$$

The block-fading channel is equivalent to a set of parallel channels, each used a fraction $\frac{1}{B}$ of the time. The block-fading channel is not information stable [135] and its channel capacity is zero [16, 87]. The corresponding information-theoretic limit is the outage probability, and the design of efficient coded modulation schemes for the block-fading channel is based on approaching the outage probability. In [46] it was proved that the outage probability for sufficiently large snr behaves as

$$P_{\text{out}}(\mathsf{snr}) = \mathcal{K}\,\mathsf{snr}^{-d_{\text{sb}}}, \tag{6.6}$$

where d_{sb} is the slope of the outage probability in a log-log scale and is given by the Singleton bound [61, 70]

$$d_{\text{sb}} \triangleq 1 + \left\lfloor B\left(1-\frac{R}{m}\right)\right\rfloor. \tag{6.7}$$

Hence, the error probability of efficient coded modulation schemes in the block-fading channel must have slope equal to the Singleton bound. Furthermore, this diversity is achievable by coded modulation as well as BICM [46]. As observed in Figure 6.3, the loss in outage probability due to BICM is marginal. While the outage probability curves with Gaussian, coded modulation and BICM inputs have the same slope $d_{\text{sb}} = 2$ when $B = 2$, we note a change in the slope with coded modulation and BICM for $B = 4$. This is due to the fact that while Gaussian inputs yield slope 4, the Singleton bound gives $d_{\text{sb}} = 3$.

In [46] the family of blockwise concatenated codes based on BICM was introduced. Improved binary codes for the block-fading channel [23, 24] can be combined with blockwise bit interleaving to yield powerful BICM schemes.

In order to apply our results on error exponents and error probability, we need to follow Malkämaki' and Leib's approach [70] and derive the error exponent for a particular channel realization. Then, the error probability (and not the error exponent) is averaged over the channel realizations. Similarly, in order to characterize the error probability for

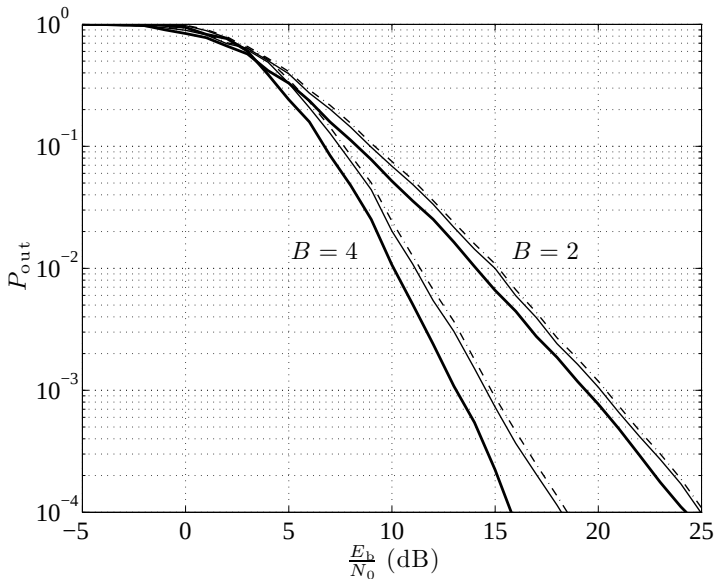


Fig. 6.3 Outage probability with 16-QAM for $R = 2$ in a block-fading channel with $B = 2, 4$. Solid lines correspond to coded modulation and dash-dotted lines correspond to BICM. The outage probability with Gaussian inputs is also shown for reference with thick solid lines.

particular code constructions, we need to calculate the error probability (or a bound) for each channel realization, and average over the realizations. Unfortunately, the union bound averaged over the channel realizations diverges, and improved bounds must be used. A simple, yet very powerful technique, was proposed by Malkämaki and Leib in [71], where the average over the fading is performed once the union bound has been truncated at 1.

6.3 MIMO

Multiple antenna or MIMO channels model transmission systems where either the transmitter, the receiver or both, have multiple antennas available for transmission/reception. Multiple antenna transmission has emerged as a key technology to achieve high spectral and power efficiency in wireless communications ever since the landmark works by Telatar [115] and Foschini and Gans [37] (see also [126]), which illustrate significant advantages from an information theoretic perspective.

Coded modulation schemes that are able to exploit the available degrees of freedom in both space and time are named space-time codes [114]. Design criteria based on the worst case PEP were proposed in [45,114] for quasi-static and fully-interleaved MIMO fading channels.

The design of efficient BICM space-time codes has been studied in a number of works [17,20,21,42,43,47,52,80,81,85,123]. The design of BICM for MIMO channels fundamentally depends on the nature of the channel (quasi-static, block-fading, fully-interleaved) and the number of antennas available at the transmitter/receiver. An important feature of the design of BICM for MIMO channels is decoding complexity. In the case of MIMO, we have a transition probability

$$P_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{H}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{H}) \propto e^{-\|\boldsymbol{y}-\sqrt{\mathsf{snr}}\boldsymbol{H}\boldsymbol{x}\|^2}, \tag{6.8}$$

where $\boldsymbol{y} \in \mathbb{C}^{n_r}$ is the received signal vector, $\boldsymbol{x} \in \mathbb{C}^{n_t}$ is the transmitted signal vector and $\boldsymbol{H} \in \mathbb{C}^{n_r \times n_t}$ is the MIMO channel matrix. In this case, the BICM decoding metric is

$$q_{j,t}(b_{j,t}(x) = b, \boldsymbol{y}) = \sum_{x' \in \mathcal{X}_b^{j,t}} e^{-\|\boldsymbol{y}-\sqrt{\mathsf{snr}}\boldsymbol{H}\boldsymbol{x}\|^2}, \tag{6.9}$$

where $\mathcal{X}_b^{j,t}$ denotes the set of vectors with bit b in the j-th position at the t-th antenna. In the fully-interleaved MIMO channel, this metric can be used to characterize the mismatched decoding error exponents, capacity and error probability. In the quasi-static channel, we would need to compute the error exponent using the metric (6.9) at every channel realization, and derive the corresponding error probability as done in [70] in block-fading channels. Note, however, that the size of $\mathcal{X}_b^{j,t}$ is exponential with the number of transmit antennas, which can make decoding very complex.

For BICM-ID, similarly to what we discussed in Chapter 5, the a priori probabilities need to be incorporated in the metric (6.9). Sphere decoding and linear filtering techniques have been successfully applied to limit the decoding complexity [20,47,52,136,137]. Figure 6.4 shows the ergodic capacities for coded modulation and BICM in a MIMO channel with $n_t = 2$, $n_r = 2$ and Rayleigh fading. A larger penalty, due to the impact of the BICM decoding metric that treats the bits

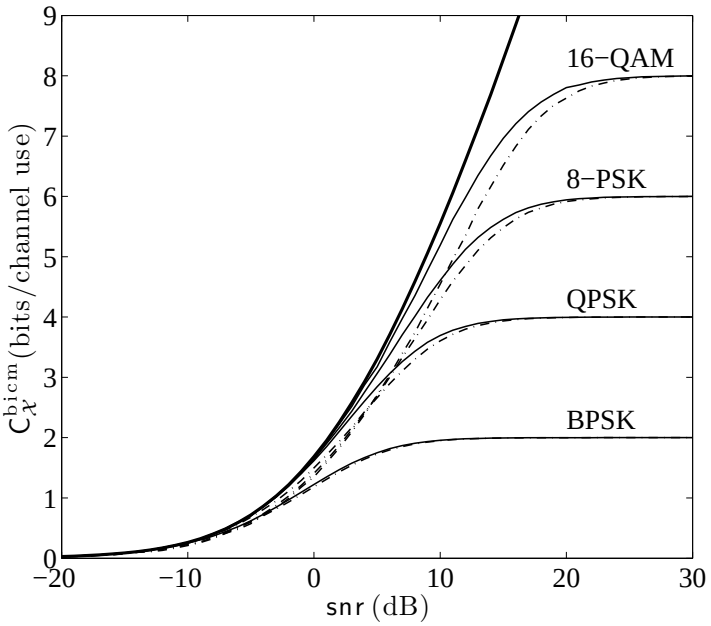


Fig. 6.4 Coded modulation capacity (solid lines), BICM capacity (dash-dotted lines) in MIMO channel with $n_t = n_r = 2$ and independent Rayleigh fading. The channel capacity (thick solid line) is shown for reference.

corresponding to a given vector symbol as independent, than in single-antenna channels is apparent. In particular, the BPSK capacities differ. When the number of antennas grows, the capacity computation becomes more complex. As illustrated in [20] sphere decoding techniques can also be employed to accurately estimate the coded modulation and BICM capacities.

6.4 Optical Communication: Discrete-Time Poisson Channel

The channel models we have mainly considered so far are variations of the additive Gaussian noise channel, which provide an accurate characterization for communication channels operating at radio and microwave frequencies. For optical frequencies, however, the family of Poisson channels is commonly considered a more accurate channel model. In particular, the so-called discrete-time Poisson (DTP) channel

with pulse-energy modulations (PEM) constitutes a natural counterpart to the PSK and QAM modulations considered throughout the monograph [72].

More formally, the input alphabet of the DTP channel is a set of M non-negative real numbers $\mathcal{X} = \{x_1, \ldots, x_M\}$, normalized to unit energy, $\frac{1}{M}\sum_{k=1}^{M} x_k = 1$. The output alphabet is the set of non-negative integers, $\mathcal{Y} = \{0, 1, 2, \ldots\}$. The output at time n is distributed according to a Poisson random variable with mean $\varepsilon_{\mathrm{s}} x_k + z_0$, where ε_{s} is the average energy constraint and z_0 a background noise component. The symbol transition probability $P_{Y|X}(y|x)$ is thus given by

$$P_{Y|X}(y|x) = e^{-(\varepsilon_{\mathrm{s}} x + z_0)} \frac{(\varepsilon_{\mathrm{s}} x + z_0)^y}{y!} \tag{6.10}$$

and therefore the symbol decoding metric can be written as

$$q(x, y) = e^{-\varepsilon_{\mathrm{s}} x} (\varepsilon_{\mathrm{s}} x + z_0)^y. \tag{6.11}$$

It is clear that BICM schemes can be used in the DTP channel. In particular, the bitwise BICM metric can be written as

$$q_j\big(b_j(x) = b, y\big) = \sum_{x' \in \mathcal{X}_b^j} e^{-\varepsilon_{\mathrm{s}} x'} (\varepsilon_{\mathrm{s}} x' + z_0)^y. \tag{6.12}$$

Figure 6.5 depicts the CM capacity attained by constellation points placed at points $\frac{6(k-1)^2}{(2M-1)(M-1)}$, for $k = 1, \ldots, M$, as well BICM capacity $\mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}}$ for BICM with binary reflected Gray mapping for several modulations. In all cases, $z_0 = 0$ is considered. This spacing between constellation points was proved in [33] to minimize the pairwise error probability in the DTP channel at high signal energy levels. As it happened in the Gaussian channel, BICM performs close to coded modulation in the DTP channel when Gray labeling is used [72].

6.5 Additive Exponential Noise Channel

Channels with additive exponential noise (AEN) have been considered in the context of queueing systems [5], and then in their own right [133] because their analytical characterization closely follows that of the Gaussian channel. For instance, the capacity of such channels with

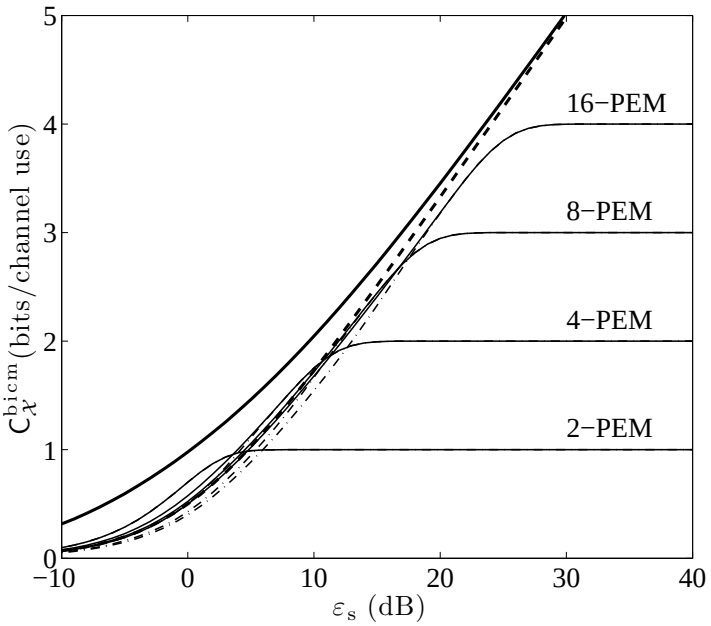


Fig. 6.5 Coded modulation capacity (solid lines), BICM capacity (dash-dotted lines) for PEM modulation in the DTP channel. Upper (thick solid line) and lower bounds (thick dashed line) to the capacity are shown for reference.

average signal-to-noise ratio snr is given by $\log(1+\mathsf{snr})$ [133]. In the AEN channel, inputs are non-negative real numbers $\mathcal{X}=\{x_1,\ldots,x_M\}$, $x_k\geq 0$, with average unit energy, $\frac{1}{M}\sum_{k=1}^{M}x_k=1$. The output alphabet is the set of non-negative real numbers, $\mathcal{Y}=[0,\infty)$. The symbol transition probability $P_{Y|X}(y|x)$ is given by

$$P_{Y|X}(y|x)=e^{-(y-\mathsf{snr}\,x)}u(y-\mathsf{snr}\,x), \tag{6.13}$$

where $u(t)$ is the unit step function. For the symbol decoding metric, a simple expression can be used, namely

$$q(x,y)=e^{\mathsf{snr}\,x}u(y-\mathsf{snr}\,x). \tag{6.14}$$

The bitwise BICM metric can be written as

$$q_j\big(b_j(x)=b,y\big)=\sum_{x'\in\mathcal{X}_b^j}e^{\mathsf{snr}\,x'}u(y-\mathsf{snr}\,x'). \tag{6.15}$$

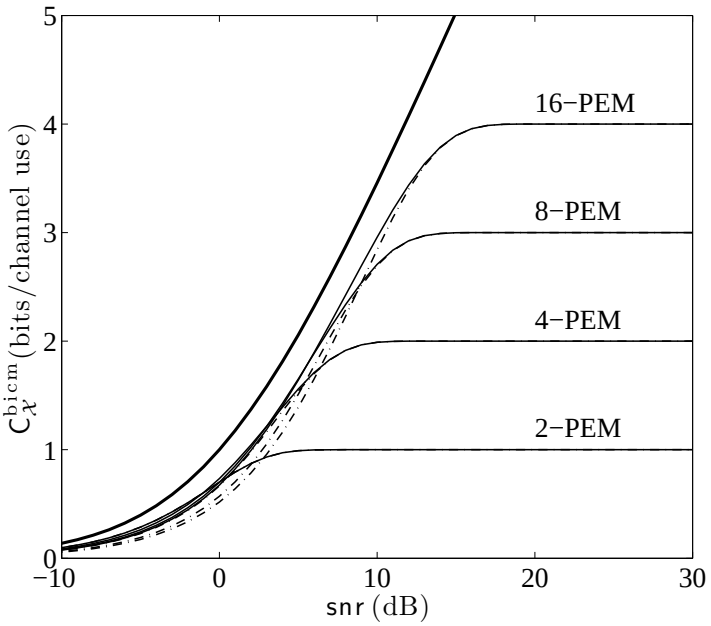


Fig. 6.6 Coded modulation capacity (solid lines), BICM capacity (dash-dotted lines) for PEM modulation in the AEN channel. The channel capacity (thick solid line) is shown for reference.

Figure 6.6 shows the CM capacity $\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}$ and the BICM capacity $\mathsf{C}_{\mathcal{X}}^{\mathrm{bicm}}$ of uniformly spaced, equiprobable constellation sets placed at locations $\frac{2(k-1)}{M-1}$. Gray mapping is used for BICM. As it happened in the Gaussian channel, BICM performs rather well in the AEN channel, although the loss in capacity with respect to $\mathsf{C}_{\mathcal{X}}^{\mathrm{cm}}$ is somewhat larger than the one in the Gaussian channel [72].

7

Conclusions

Coding in the signal space is dictated directly by Shannon capacity formula and suggested by the random coding achievability proof. In early days of digital communications, modulation and coding were kept separated because of complexity. Modulation and demodulation treated the physical channel, waveform generation, parameter estimation (e.g., frequency, phase and timing) and symbol-by-symbol detection. Error correcting codes were used to undo the errors introduced by the physical modulation/demodulation process. This paradigm changed radically with the advent of Coded Modulation. Trellis-coded modulation was a topic of significant research activities in the 80's, for approximately a decade. In the early 90's, new families of powerful random-like codes, such as turbo codes, LDPC codes and Repeat-Accumulate codes were discovered (or re-discovered), along with very efficient low-complexity Belief Propagation iterative decoding algorithms which allowed unprecedented performance close to capacity, at least for binary-input channels. Roughly at the same time, Bit-Interleaved Coded Modulation emerged as a very simple yet powerful tool to concatenate virtually any binary code to any modulation constellation, with only minor penalty with respect to the traditional joint modulation and decoding

paradigm. Therefore, BICM and modern powerful codes with iterative decoding were a natural marriage, and today virtually any modern telecommunication system that seeks high spectral efficiency and high performance, including DSL, Digital TV and Audio Broadcasting, Wireless LANs, WiMax, and next generation cellular systems use BICM as the central component of their respective physical layers.

In this monograph, we have presented a comprehensive review of the foundations of BICM in terms of information-theoretic, error probability and iterative decoding analysis. In particular, we have shown that BICM is a paramount example of mismatched decoding where the decoding metric is obtained as the product of bitwise metrics. Using this decoder, we have presented the derivation of the average error probability of the random coding ensemble and obtained the resulting error exponents, generalized mutual information and cutoff rate. We have shown that the error exponent —and hence the cutoff rate— of the BICM mismatched decoder is upper bounded by that of coded modulation and may thus be lower than in the infinite-interleaved model that was previously proposed and generally accepted for the analysis of BICM. Nevertheless, for binary reflected Gray mapping in Gaussian channels the loss in error exponent is small. We have also considered BICM in the wideband, or low SNR, regime and we have characterized the suboptimality of the BICM decoder in terms of minimum bit energy to noise ratio and wideband slope.

We have reviewed the error probability analysis of BICM with a generic decoding metric with finite and infinite interleavers and we have given exact formulas and tight approximations, such as the saddlepoint approximation. We have seen that a short uniform interleaver results in a performance loss that in a fading channel translates into a diversity loss. We have also reviewed the application of improved bounds to BICM, which are central to analyze the performance of modern codes such as turbo and LDPC codes with BICM, since these typically operates beyond the cutoff rate and the standard union bound fails to give meaningful results.

We have reviewed iterative decoding of BICM. In particular, we have shown that perfect extrinsic information closes the gap between the error exponents of BICM and that of coded modulation. We have

reviewed the density evolution analysis of BICM-ID and the application of the area theorem to BICM. We have finally shown how to design very powerful BICM-ID structures based on LDPC or RA codes. The analysis reviewed here is valid for infinitely large block lengths, which is anyway a meaningful assumption when modern random-like codes are considered, since the latter yield competitive performance at moderate to large block length. The analysis of BICM with iterative decoding and short block lengths remains an open problem.

Overall, this monograph casts BICM in a unified common framework, and we believe it provides analysis and code design tools valuable for both researchers and system designers.

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of mathematical funcions*. Dover Publications, 1964.

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

[3] A. Amraoui, *Asymptotic and finite-length optimization of LDPC codes*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2006.

[4] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-Length Scaling for Iteratively Decoded LDPC Ensembles," *submitted to IEEE Trans. Inf. Theory. Arxiv preprint cs/0406050*, 2004.

[5] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, 1996.

[6] A. Ashikhmin, G. Kramer, and S. Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2657–2673, 2004.

[7] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[8] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 909–926, 1998.

[9] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 409–428, 1996.

[10] A. Bennatan and D. Burshtein, "Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 549–583, 2006.

[11] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes.," in *IEEE Int. Conf. Commun., Geneva*, 1993.

[12] G. Beyer, K. Engdahl, and K. S. Zigangirov, "Asymptotical analysis and comparison of two coded modulation schemes using PSK signaling - Part I," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2782–2792, 2001.

[13] G. Beyer, K. Engdahl, and K. S. Zigangirov, "Asymptotical analysis and comparison of two coded modulation schemes using PSK signaling - Part II," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2793–2806, 2001.

[14] K. Bhattad and K. R. Narayanan, "An MSE-based transfer chart for analyzing iterative decoding schemes using a gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, p. 22, 2007.

[15] E. Biglieri, D. Divsalar, M. K. Simon, and P. J. McLane, *Introduction to Trellis-Coded Modulation with Applications*. Prentice-Hall, 1991.

[16] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[17] E. Biglieri, G. Taricco, and E. Viterbo, "Bit-interleaved space–time codes for fading channels," in *Proc. Conf. Inf. Sci. and Sys.*, 2000.

[18] E. Biglieri, G. Caire, and G. Taricco, "Error probability over fading channels: a unified approach," *Europ. Trans. Commun.*, January 1998.

[19] E. Biglieri, G. Caire, G. Taricco, and J. Ventura-Traveset, "Simple method for evaluating error probabilities," *Electron. Lett.*, vol. 32, pp. 191–192, February 1996.

[20] J. Boutros, N. Gresset, L. Brunel, and M. Fossorier, "Soft-input soft-output lattice sphere decoder for linear channels," in *Global Telecommun. Conf.*, 2003.

[21] J. J. Boutros, F. Boixadera, and C. Lamy, "Bit-interleaved coded modulations for multiple-input multiple-output channels," in *IEEE Int. Symp. Spread Spectrum Tech. App.*, 2000.

[22] J. J. Boutros and G. Caire, "Iterative multiuser joint decoding: unified framework and asymptotic analysis," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1772–1793, May 2002.

[23] J. J. Boutros, E. Calvanese Strinati, and A. Guillén i Fàbregas, "Turbo code design for block fading channels," *Proc. 42nd Annual Allerton Conference on Communication, Control and Computing, Allerton, IL*, Sept.-Oct. 2004.

[24] J. Boutros, A. Guillén i Fàbregas, E. Biglieri, and G. Zémor, "Low-density parity-check codes for nonergodic block-fading channels," *submitted to IEEE Trans. Inf. Theory. Available from http://arxiv.org/abs/0710.1182*, 2007.

[25] F. Brännström, *Convergence Analysis and Design of Multiple Concatenated Codes*. PhD thesis, Department of Computer Engineering, Chalmers University of Technology, 2004.

[26] F. Brännström and L. K. Rasmussen, "Classification of 8PSK mappings for BICM," in *2007 IEEE Int. Symp. Inf. Theory, Nice, France, June*, 2007.

[27] F. Brännström, L. K. Rasmussen, and A. J. Grant, "Convergence Analysis and Optimal Scheduling for Multiple Concatenated Codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3354–3364, 2005.

[28] R. W. Butler, *Saddlepoint Approximations with Applications*. Cambridge University Press, 2007.

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

[30] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 944–957, 2001.

[31] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley New York, 2nd ed., 2006.

[32] D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for turbo-like codes," *Proc. 1998 Allerton Conference on Communication, Control and Computing*, pp. 201–210, 1998.

[33] G. H. Einarsson, *Principles of lightwave communications*. Wiley New York, 1996.

[34] R. F. Fischer, *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, Inc. New York, NY, USA, 2002.

[35] G. D. Forney Jr, "The Viterbi algorithm," *Proc. of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

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

[37] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[38] R. G. Gallager, "Low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[39] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Inc. New York, NY, USA, 1968.

[40] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, pp. 124–142, Mar. 1985.

[41] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: general alphabets, channels with memory, and the wideband limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, 2000.

[42] N. Gresset, J. J. Boutros, and L. Brunel, "Multidimensional mappings for iteratively decoded bicm on multiple-antenna channels," *IEEE Trans. Inf. Theory.*, vol. 51, no. 9, pp. 3337–3346, 2005.

[43] N. Gresset, L. Brunel, and J. J. Boutros, "Space-time coding techniques with bit-interleaved coded modulations for MIMO block-fading channels," *IEEE Trans. Inf. Theory*, vol. 54, pp. 2156–2178, May 2008.

[44] S. Guemghar, *Advanced Coding Techniques and Applications to CDMA*. PhD thesis, Ecole Nationale des Télécommunications, Paris, 2004.

[45] J. C. Guey, M. P. Fitz, M. R. Bell, and W. Y. Kuo, "Signal design for transmitter diversity wireless communicationsystems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 527–537, 1999.

[46] A. Guillén i Fàbregas and G. Caire, "Coded modulation in the block-fading channel: coding theorems and code construction," *IEEE Trans on Inf. Theory*, vol. 52, pp. 91–114, Jan. 2006.

[47] A. Guillén i Fàbregas and G. Caire, "Impact of signal constellation expansion on the achievable diversity of pragmatic bit-interleaved space-time codes," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2032–2037, 2006.

[48] A. Guillén i Fàbregas and A. Grant, "Capacity approaching codes for noncoherent orthogonal modulation," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 4004–4013, Nov. 2007.

[49] A. Guillén i Fàbregas and A. Martinez, "Derivative of BICM mutual information," *IET Electronics Letters*, vol. 43, pp. 1219–1220, Oct. 2007.

[50] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Error probability of bit-interleaved coded modulation using the gaussian approximation," in *Proceedings of the Conf. on Inform. Sciences and Systems (CISS-04)*, (Princeton, USA), March 2004.

[51] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1261–1282, Apr. 2005.

[52] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, 2003.

[53] P. Hoeher and J. Lodge, ""Turbo DPSK": iterative differential PSK demodulationand channel decoding," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 837–843, 1999.

[54] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2141–2155, 2003.

[55] S. Huettinger and J. Huber, "Design of "multiple-turbo-codes" with transfer characteristics of component codes"," *Proc. Conf. Inf. Sciences and Systems (CISS)*, 2002.

[56] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, pp. 371–377, May 1977.

[57] J. L. Jensen, *Saddlepoint Approximations*. Oxford University Press, USA, 1995.

[58] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes," in *2nd Int. Symp. Turbo Codes and Related Topics*, pp. 1–8, 2000.

[59] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU. Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, 1993.

[60] A. Kavcic, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution, and code performance bounds," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1636–1652, 2003.

[61] R. Knopp and P. A. Humblet, "On coding for block fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 189–205, 2000.

[62] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[63] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8 PSK signaling," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1250–1257, 2002.

[64] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Letters*, vol. 1, pp. 169–171, Nov. 1997.

[65] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding using soft feedback," *Electronics Letters*, vol. 34, no. 10, pp. 942–943, 1998.

[66] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 715–724, 1999.

[67] A. Lozano, A. M. Tulino, and S. Verdú, "Opitmum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3033–3051, Jul. 2006.

[68] R. Lugannani and S. O. Rice, "Saddle point approximation for the distribution of the sum of independent random variables," *Adv. Appl. Prob.*, vol. 12, pp. 475–490, 1980.

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

[70] E. Malkämaki and H. Leib, "Coded diversity on block-fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 771–781, 1999.

[71] E. Malkämaki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1643–1646, 1999.

[72] A. Martinez, *Information-theoretic analysis of a family of additive energy channels*. PhD thesis, Technische Universiteit Eindhoven, 2008.

[73] A. Martinez and A. Guillén i Fàbregas, "Large-snr error probability analysis of bicm with uniform interleaving in fading channels," *submitted to IEEE Trans. Wireless Commun.*, Nov. 2008.

[74] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Error probability analysis of bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 52, pp. 262–271, Jan. 2006.

[75] 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, pp. 2051–2056, Jun. 2007.

[76] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation in the wideband regime," *accepted for publication in IEEE Trans. Inf. Theory*, 2008.

[77] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation revisited: a mismatched decoding perspective," *submitted to IEEE Trans. Inf. Theory. arXiv:0805.1327*, 2008.

[78] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "New simple evaluation of the error probability of bit-interleaved coded modulation using the saddle-point approximation," in *Proc. 2004 International Symposium on Information Theory and its Applications (ISITA 2004)*, (Parma (Italy)), October 2004.

[79] J. L. Massey, "Coding and modulation in digital communications," in *Int. Zurich Seminar Digital Commun., Zurich, Switzerland*, 1974.

[80] M. McKay and I. Collings, "Capacity and performance of MIMO-BICM with zero-forcing receivers," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 74–83, 2005.

[81] M. McKay and I. Collings, "Error performance of MIMO-BICM with zero-forcing receivers in spatially-correlated Rayleigh channels," *IEEE Trans. Wireless Commun*, vol. 6, no. 3, pp. 787–792, 2005.

[82] C. Measson, *Conservation laws for coding*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2006.

[83] C. Measson, A. Montanari, and R. Urbanke, "Why we can not surpass capacity: The matching condition," in *43rd Annual Allerton Conference on Communication, Control and Computing, Monticello, IL*, Oct. 2005.

[84] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai Shitz, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, 1994.

[85] S. H. Müller-Weinfurtner, "Coding approaches for multiple antenna transmission in fast fading and OFDM," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2442–2450, 2002.

[86] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.

[87] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, 1994.

[88] V. Pauli, L. Lampe, and R. Schober, ""Turbo DPSK" Using Soft Multiple-Symbol Differential Sphere Decoding," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1385–1398, Apr. 2006.

[89] M. Peleg, I. Sason, S. Shamai, and A. Elia, "On interleaved, differentially encoded convolutional codes," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2572–2582, 1999.

[90] M. Peleg and S. Shamai, "On the capacity of the blockwise incoherent MPSK channel," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 603–609, 1998.

[91] M. Peleg, S. Shamai, and S. Galan, "Iterative decoding for coded noncoherent MPSK communications overphase-noisy AWGN channel," *IEE Proc. Commun.*, vol. 147, no. 2, pp. 87–95, 2000.

[92] F. Peng, W. Ryan, and R. Wesel, "Surrogate-Channel Design of Universal LDPC Codes," *IEEE Commun. Letters*, vol. 10, no. 6, p. 480, 2006.

[93] 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, 2006.

[94] V. Prelov and S. Verdú, "Second order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1567–1580, Aug. 2004.

[95] J. G. Proakis, "Digital Communications," *McGraw-Hill USA*, 4th edition, 2001.

[96] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, Cambridge, UK, 2008.

[97] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, 2001.

[98] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, pp. 364–375, Mar. 1996.

[99] P. Robertson and T. Worz, "Bandwidth-efficient turbo trellis-coded modulation using puncturedcomponent codes," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 206–218, 1998.

[100] A. Roumy, S. Guemghar, G. Caire, and S. Verdu, "Design methods for irregular repeat-accumulate codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1711–1727, 2004.

[101] I. Sason and S. Shamai, *Performance Analysis of Linear Codes Under Maximum-likelihood Decoding: A Tutorial*. Vol. 3, Now Publishers Inc, 2006.

[102] H. E. Sawaya and J. J. Boutros, "Multilevel coded modulations based on asymmetric constellations," in *IEEE Int. Symp. Inf. Theory, Washington, DC, USA, June*, 2001.

[103] F. Schreckenbach, N. Gortz, J. Hagenauer, and G. Bauch, "Optimized symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Letters*, vol. 7, pp. 593 – 595, Dec. 2003.

[104] V. Sethuraman and B. Hajek, "Comments on Bit-Interleaved Coded Modulation," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1795–1797, 2006.

[105] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.

[106] F. Simoens, H. Wymeersch, H. Bruneel, and M. Moeneclaey, "Multidimensional mapping for bit-interleaved coded modulation with BPSK/QPSK signaling," *IEEE Commun. Letters*, vol. 9, no. 5, pp. 453–455, 2005.

[107] 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, pp. 87–99, Jan. 2008.

[108] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J*, vol. 52, no. 7, pp. 1037–1076, 1973.

[109] C. Stierstorfer and R. F. H. Fischer, "(Gray) mappings for bit-interleaved coded modulation," in *65th IEEE Vehic. Techn. Conf., Dublin, Ireland*, April 2007.

[110] C. Stierstorfer and R. F. H. Fischer, "Mappings for BICM in UWB scenarios," in *Proc. 7th Intl ITG Conf. Source and Channel Coding (SCC), Ulm, Germany*, January 2008.

[111] I. Sutskover, S. Shamai, and J. Ziv, "Extremes of information combining," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1313–1325, 2005.

[112] L. Szczecinski, A. Alvarado, R. Feick, and L. Ahumada, "On the distribution of L-values in gray-mapped M^2-QAM signals: Exact expressions and simple approximations," in *IEEE Global Commun. Conf, 26-30 November, Washington, DC, USA*, 2007.

[113] J. Tan and G. Stuber, "Analysis and Design of Symbol Mappers for Iteratively Decoded BICM," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 662–672, 2005.

[114] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[115] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[116] S. ten Brink, "Convergence of iterative decoding," *Electronics Letters*, vol. 35, p. 806, 1999.

[117] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," *AEU Int. J. Electron. Commun*, vol. 54, no. 6, pp. 389–398, 2000.

[118] S. ten Brink, "Code doping for triggering iterative decoding convergence," in *IEEE Int. Symp. Inf. Theory, Washington, DC, USA, June*, 2001.

[119] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct. 2001.

[120] S. ten Brink and G. Kramer, "Design of repeat-accumulate codes for iterative detection and decoding," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2764–2772, 2003.

[121] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, 2004.

[122] S. ten Brink, J. Speidel, and R. H. Yan, "Iterative demapping for QPSK modulation," *Electronics Letters*, vol. 34, no. 15, pp. 1459–1460, 1998.

[123] A. M. Tonello, "Space-time bit-interleaved coded modulation with an iterative decoding strategy," in *52nd Vehic. Tech. Conf.*, 2000.

[124] N. Tran and H. Nguyen, "Design and performance of BICM-ID systems with hypercube constellations," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1169–1179, 2006.

[125] N. Tran and H. Nguyen, "Signal mappings of 8-ary constellations for bit interleaved coded modulation with iterative decoding," *IEEE Trans. Broadcasting*, vol. 52, no. 1, pp. 92–99, 2006.

[126] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[127] M. Tüchler, "Design of serially concatenated systems depending on the block length," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 209–218, 2004.

[128] M. Tüchler and J. Hagenauer, "EXIT charts of irregular codes," in *Conf. Inf. Sciences and Systems*, pp. 748–753, 2002.

[129] M. Tüchler, J. Hagenauer, and S. ten Brink, "Measures for Tracing Convergence of Iterative Decoding Algorithms," in *Proc. 4th Intern. ITG Conf. on Source and Channel Coding, Berlin, Germany*, pp. 53–60, Jan. 2002.

[130] G. Ungerböck, "Channel Coding With Multilevel/Phase Signals.," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–66, 1982.

[131] M. Valenti, R. Doppalapudi, and D. Torrieri, "A genetic algorithm for designing constellations with low error floors," *Proc. Conf. Inf. Sciences and Systems (CISS)*, 2008.

[132] M. Valenti and S. Cheng, "Iterative demodulation and decoding of turbo-coded M-ary noncoherent orthogonal modulation," *IEEE J. Select. Areas Commun.*, vol. 23, no. 9, pp. 1739–1747, 2005.

[133] S. Verdú, "The exponential distribution in information theory," *Problemy peredachi informatsii*, vol. 32, no. 1, pp. 100–111, 1996.

[134] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1319–1343, Jun. 2002.

[135] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[136] H. Vikalo, *Sphere Decoding Algorithms for Digital Communications*. PhD thesis, Stanford University, 2003.

[137] H. Vikalo, B. Hassibi, and T. Kailath, "Iterative decoding for MIMO channels via modified sphere decoding," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2299–2311, 2004.

[138] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[139] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. McGraw-Hill, Inc. New York, NY, USA, 1979.

[140] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1361–1391, Jul. 1999.

[141] P. Yeh, S. Zummo, and W. Stark, "Error probability of bit-interleaved coded modulation in wireless environments," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 2, pp. 722–728, 2006.

[142] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, 1992.