

---

# Bit-Interleaved Coded Modulation

---



# Bit-Interleaved Coded Modulation

---

**Albert Guillén i Fàbregas**

*University of Cambridge  
Cambridge, CB2 1PZ  
UK*

*guillen@ieee.org*

**Alfonso Martinez**

*Centrum Wiskunde & Informatica (CWI)  
Amsterdam, 1098 SJ  
The Netherlands*

*alfonso.martinez@ieee.org*

**Giuseppe Caire**

*University of Southern California  
Los Angeles, 90080 CA  
USA*

*caire@usc.edu*

**now**

the essence of knowledge

Boston – Delft

## Foundations and Trends<sup>®</sup> in Communications and Information Theory

*Published, sold and distributed by:*

now Publishers Inc.  
PO Box 1024  
Hanover, MA 02339  
USA  
Tel. +1-781-985-4510  
[www.nowpublishers.com](http://www.nowpublishers.com)  
[sales@nowpublishers.com](mailto:sales@nowpublishers.com)

*Outside North America:*

now Publishers Inc.  
PO Box 179  
2600 AD Delft  
The Netherlands  
Tel. +31-6-51115274

The preferred citation for this publication is A. Guillén i Fàbregas, A. Martinez and G. Caire, Bit-Interleaved Coded Modulation, Foundations and Trends<sup>®</sup> in Communications and Information Theory, vol 5, nos 1-2, pp 1-153, 2008

ISBN: 978-1-60198-190-5

© 2008 A. Guillén i Fàbregas, A. Martinez and G. Caire

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

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: [www.copyright.com](http://www.copyright.com)

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; [www.nowpublishers.com](http://www.nowpublishers.com); [sales@nowpublishers.com](mailto:sales@nowpublishers.com)

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, [www.nowpublishers.com](http://www.nowpublishers.com); e-mail: [sales@nowpublishers.com](mailto:sales@nowpublishers.com)

**Foundations and Trends® in  
Communications and Information Theory**  
Volume 5 Issue 1–2, 2008  
**Editorial Board**

**Editor-in-Chief:**

**Sergio Verdú**

*Depart of Electrical Engineering  
Princeton University  
Princeton, New Jersey 08544*

**Editors**

|                                            |                                      |
|--------------------------------------------|--------------------------------------|
| Venkat Anantharam (UC. Berkeley)           | Amos Lapidoth (ETH Zurich)           |
| Ezio Biglieri (U. Torino)                  | Bob McEliece (Caltech)               |
| Giuseppe Caire (U. Southern<br>California) | Neri Merhav (Technion)               |
| Roger Cheng (U. Hong Kong)                 | David Neuhoff (U. Michigan)          |
| K.C. Chen (Taipei)                         | Alon Orlitsky (UC. San Diego)        |
| Daniel Costello (U. Notre Dame)            | Vincent Poor (Princeton)             |
| Thomas Cover (Stanford)                    | Kannan Ramchandran (UC.<br>Berkeley) |
| Anthony Ephremides (U. Maryland)           | Bixio Rimoldi (EPFL)                 |
| Andrea Goldsmith (Stanford)                | Shlomo Shamai (Technion)             |
| Dave Forney (MIT)                          | Amin Shokrollahi (EPFL)              |
| Georgios Giannakis (U. Minnesota)          | Gadiel Seroussi (MSRI)               |
| Joachim Hagenauer (TU Munich)              | Wojciech Szpankowski (Purdue)        |
| Te Sun Han (Tokyo)                         | Vahid Tarokh (Harvard)               |
| Babak Hassibi (Caltech)                    | David Tse (UC. Berkeley)             |
| Michael Honig (Northwestern)               | Ruediger Urbanke (EPFL)              |
| Johannes Huber (Erlangen)                  | Steve Wicker (Cornell)               |
| Hideki Imai (Tokyo)                        | Raymond Yeung (Hong Kong)            |
| Rodney Kennedy (Canberra)                  | Bin Yu (UC. Berkeley)                |
| Sanjeev Kulkarni (Princeton)               |                                      |

## Editorial Scope

**Foundations and Trends<sup>®</sup> in Communications and Information Theory** will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

### Information for Librarians

Foundations and Trends<sup>®</sup> in Communications and Information Theory, 2008, Volume 5, 6 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.

## Bit-Interleaved Coded Modulation

Albert Guillén i Fàbregas<sup>1</sup>,  
Alfonso Martinez<sup>2</sup> and Giuseppe Caire<sup>3</sup>

<sup>1</sup> *Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, UK, guillen@ieee.org*

<sup>2</sup> *Centrum Wiskunde & Informatica (CWI), Kruislaan 413, Amsterdam, 1098 SJ, The Netherlands, alfonso.martinez@ieee.org*

<sup>3</sup> *Electrical Engineering Department, University of Southern California, 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 works 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 text, 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

---

|                                                          |           |
|----------------------------------------------------------|-----------|
| <b>List of Abbreviations, Acronyms and Symbols</b>       | <b>1</b>  |
| <b>1 Introduction</b>                                    | <b>7</b>  |
| <b>2 Channel Model and Code Ensembles</b>                | <b>11</b> |
| 2.1 Channel Model: Encoding and Decoding                 | 11        |
| 2.2 Coded Modulation                                     | 14        |
| 2.3 Bit-Interleaved Coded Modulation                     | 15        |
| <b>3 Information-Theoretic Foundations</b>               | <b>23</b> |
| 3.1 Coded Modulation                                     | 24        |
| 3.2 Bit-Interleaved Coded Modulation                     | 30        |
| 3.3 Comparison with Multilevel Coding                    | 36        |
| 3.4 Mutual Information Analysis                          | 43        |
| 3.5 Concluding Remarks and Related Works                 | 53        |
| <b>4 Error Probability Analysis</b>                      | <b>57</b> |
| 4.1 Error Probability and the Union Bound                | 58        |
| 4.2 Pairwise Error Probability for Infinite Interleaving | 66        |
| 4.3 Pairwise Error Probability for Finite Interleaving   | 79        |
| 4.4 Bounds and Approximations Above the Cutoff Rate      | 89        |
| 4.5 Concluding Remarks and Related Works                 | 92        |

|          |                                                      |            |
|----------|------------------------------------------------------|------------|
| <b>5</b> | <b>Iterative Decoding</b>                            | <b>97</b>  |
| 5.1      | Factor Graph Representation and Belief Propagation   | 99         |
| 5.2      | Density Evolution                                    | 103        |
| 5.3      | EXIT Charts                                          | 108        |
| 5.4      | The Area Theorem                                     | 112        |
| 5.5      | Improved Schemes                                     | 116        |
| 5.6      | Concluding Remarks and Related Works                 | 126        |
| <b>6</b> | <b>Applications</b>                                  | <b>131</b> |
| 6.1      | Noncoherent Demodulation                             | 131        |
| 6.2      | Block-Fading                                         | 133        |
| 6.3      | MIMO                                                 | 136        |
| 6.4      | Optical Communication: Discrete-Time Poisson Channel | 138        |
| 6.5      | Additive Exponential Noise Channel                   | 139        |
| <b>7</b> | <b>Conclusions</b>                                   | <b>143</b> |
|          | <b>Acknowledgments</b>                               | <b>147</b> |
|          | <b>References</b>                                    | <b>149</b> |

## 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 $d$                      |
| $A_{d,\rho_N}$                  | Weight enumerator at Hamming distance $d$ and pattern $\rho_N$ |
| $A'_d$                          | 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                                               |
| $C_{\mathcal{X}}^{\text{bicm}}$ | BICM capacity over set $\mathcal{X}$                           |
| $C_{\text{bpsk}}$               | BPSK capacity                                                  |
| $C_{\mathcal{X}}^{\text{cm}}$   | 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                                               |
| $\mathbf{d}$                    | Scrambling (randomization) sequence                            |
| $\bar{d}_v$                     | Average variable degree (LDPC code)                            |
| $\bar{d}_c$                     | Average check node degree (LDPC code)                          |
| $\Delta P$                      | Power expansion ratio                                          |
| $\Delta W$                      | Bandwidth expansion ratio                                      |
| $\mathbb{E}[U]$                 | Expectation (of a random variable $U$ )                        |
| $E_b$                           | Average bit energy                                             |
| $\frac{E_b}{N_0}$               | Ratio between average bit energy and noise spectral density    |
| $\frac{E_b}{N_0} \lim$          | $\frac{E_b}{N_0}$ at vanishing SNR                             |
| $E_s$                           | Average signal energy                                          |
| $E(R)$                          | Reliability function at rate $R$                               |
| $E_0^{\text{bicm}}(\rho, s)$    | BICM random coding exponent                                    |
| $E_0^{\text{cm}}(\rho)$         | CM random coding exponent                                      |

|                                 |                                                             |
|---------------------------------|-------------------------------------------------------------|
| $E_0^q(\rho, s)$                | Generalized Gallager function                               |
| $E_r^q(R)$                      | Random coding exponent with mismatched decoding             |
| $\text{exit}_{\text{dec}}(y)$   | Extrinsic information at decoder                            |
| $\text{exit}_{\text{dem}}(x)$   | Extrinsic information at demapper                           |
| $h_k$                           | Fading realization at time $k$                              |
| $I(X; Y)$                       | Mutual information between variables $X$ and $Y$            |
| $I^{\text{cm}}(X; Y)$           | Coded modulation capacity ( $C_{\mathcal{X}}^{\text{cm}}$ ) |
| $I^{\text{gmi}}(X; Y)$          | Generalized mutual information                              |
| $I_s^{\text{gmi}}(X; Y)$        | Generalized mutual information (function of $s$ )           |
| $I^{\text{ind}}(X; Y)$          | BICM capacity with independent-channel model                |
| $K$                             | Number of bits per codeword $\log_2  \mathcal{M} $          |
| $\kappa(s)$                     | 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 $v$          |
| $\kappa_{\text{pw}}(s)$         | Cumulant transform of pairwise score                        |
| $\kappa_{\text{pw}}(s, \rho_N)$ | Cumulant transform of pairwise score for pattern $\rho_N$   |
| $M$                             | Input set (constellation) $\mathcal{X}$ cardinality         |
| $m$                             | Number of bits per modulation symbol                        |
| $m_f$                           | Nakagami fading parameter                                   |
| $\mu$                           | Labeling (mapping) rule                                     |
| $\mathcal{M}$                   | Message set                                                 |
| $\mathbf{m}$                    | Message                                                     |
| $\hat{\mathbf{m}}$              | Message estimate                                            |
| $\text{mmse}(\text{snr})$       | 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)                |
| $\nu^{f \rightarrow \vartheta}$ | Function-to-variable message                                |
| $\nu^{\vartheta \rightarrow f}$ | Variable-to-function message                                |
| $\mathcal{O}(f(x))$             | Term vanishing as least as fast as $af(x)$ , for $a > 0$    |
| $\mathcal{o}(f(x))$             | 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                        |

|                                                                   |                                                                                           |
|-------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| $P_j(y b)$                                                        | Transition probability of output $y$ for $j$ th bit                                       |
| $P_{\mathbf{j}}(y \mathbf{b})$                                    | Output transition probability for bits $\mathbf{b}$ at $\mathbf{j}$ positions             |
| $P_{B_j Y}(b y)$                                                  | $j$ th a posteriori marginal                                                              |
| $P_{\cdot}$                                                       | Probability distribution                                                                  |
| $P_{Y X}(y x)$                                                    | Channel transition probability (symbol)                                                   |
| $P_{\mathbf{Y} \mathbf{X}}(\mathbf{y} \mathbf{x})$                | Channel transition probability (sequence)                                                 |
| $\text{PEP}(d)$                                                   | Pairwise error probability                                                                |
| $\text{PEP}_1(d)$                                                 | Pairwise error probability (infinite interleaving)                                        |
| $\text{PEP}(\mathbf{x}_{\text{m}'}, \mathbf{x}_{\text{m}})$       | Pairwise error probability                                                                |
| $\pi_n$                                                           | Interleaver of size $n$                                                                   |
| $\text{Pr}^{\text{dec} \rightarrow \text{dem}}(b)$                | Bit probability (from the decoder)                                                        |
| $\text{Pr}^{\text{dem} \rightarrow \text{dec}}(b)$                | Bit probability (from the demapper)                                                       |
| $Q(\cdot)$                                                        | Gaussian tail function                                                                    |
| $q(x, y)$                                                         | Symbol decoding metric                                                                    |
| $q(\mathbf{x}, \mathbf{y})$                                       | Codeword decoding metric                                                                  |
| $q_j(b, y)$                                                       | Bit decoding metric of $j$ th bit                                                         |
| $R$                                                               | Code rate, $R = \log_2  \mathcal{M} /N$                                                   |
| $r$                                                               | Binary code rate, $r = \log_2  \mathcal{C} /n = R/m$                                      |
| $R_0$                                                             | Cutoff rate                                                                               |
| $R_0^{\text{av}}$                                                 | Cutoff rate for average-channel model                                                     |
| $R_0^{\text{ind}}$                                                | Cutoff rate for independent-channel model                                                 |
| $R_0^q$                                                           | Generalized cutoff rate (mismatched decoding)                                             |
| $\rho_N$                                                          | Bit distribution pattern over codeword                                                    |
| $\sum_{\sim x}$                                                   | Summary operator — excluding $x$ —                                                        |
| $\hat{s}$                                                         | Saddlepoint value                                                                         |
| $\sigma_X^2$                                                      | Variance                                                                                  |
| $\hat{\sigma}_X^2$                                                | Pseudo-variance, $\sigma_X^2 \triangleq \mathbb{E}[ X ^2] -  \mathbb{E}[X] ^2$            |
| $\zeta_0$                                                         | Wideband slope                                                                            |
| $\text{snr}$                                                      | 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                                               |
| $\mathcal{X}_{b_{i_1}, \dots, b_{i_v}}^{j_{i_1}, \dots, j_{i_v}}$ | Set of symbols with bits $b_{i_1}, \dots, b_{i_v}$ at positions $j_{i_1}, \dots, j_{i_v}$ |

|                                                      |                                                        |
|------------------------------------------------------|--------------------------------------------------------|
| $x_k$                                                | Channel input at time $k$                              |
| $\mathbf{x}$                                         | Vector of all channel inputs; input codeword           |
| $\mathbf{x}_m$                                       | 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 $k$                             |
| $\mathbf{y}$                                         | Vector of all channel outputs                          |
| $\mathcal{Y}$                                        | Output signal set                                      |
| $z_k$                                                | Noise realization at time $k$                          |
| $\Xi^{\text{pw}}$                                    | Pairwise score                                         |
| $\Xi_k^{\text{s}}$                                   | Symbol score for $k$ th symbol                         |
| $\Xi_{k,j}^{\text{b}}$                               | Bit score at $j$ th label of $k$ th symbol             |
| $\Xi_1^{\text{b}}$                                   | Symbols score with weight 1 (bit score)                |
| $\Xi_{m(k-1)+j}^{\text{dec} \rightarrow \text{dem}}$ | Decoder LLR for $j$ th bit of $k$ th symbol            |
| $\Xi^{\text{dec} \rightarrow \text{dem}}$            | Decoder LLR vector                                     |
| $\Xi_{\sim i}^{\text{dec} \rightarrow \text{dem}}$   | Decoder LLR vector, excluding the $i$ th component     |
| $\Xi_{m(k-1)+j}^{\text{dem} \rightarrow \text{dec}}$ | Demodulator LLR for $j$ th bit of $k$ th symbol        |
| $\Xi^{\text{dem} \rightarrow \text{dec}}$            | Demodulator LLR vector                                 |
| $\Xi_{\sim i}^{\text{dem} \rightarrow \text{dec}}$   | 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

$$y_k = \sqrt{\text{snr}} x_k + z_k, \quad (1.1)$$

where  $y_k, x_k, z_k$  are complex random variables and  $\text{snr}$  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  $\text{snr}$  is given by the well-known

$$C = \log(1 + \text{snr}). \quad (1.2)$$

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 [17]). 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* [97] approach the capacity of binary-input channels with low complexity. The analysis of iterative decoding also led to new methods for their efficient design [97]. At this point, a natural development of coded modulation would have been the extension of these powerful codes to nonbinary 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 [65] and ten Brink [122] proposed iterative demodulation for BICM, and illustrated significant performance gains with respect to classical noniterative 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 communication 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 text, 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 noncoherent 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 section provides the reference background for the remainder of the text, 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 text. Section 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.,  $k = 1, \dots, N$ . A block code  $\mathcal{M} \subseteq \mathcal{X}^N$  of length  $N$  is a set of  $|\mathcal{M}|$  vectors  $\mathbf{x} = (x_1, \dots, x_N) \in \mathcal{X}^N$ , called codewords. The channel output is denoted by  $\mathbf{y} \triangleq (y_1, \dots, y_N)$ , with  $y_k \in \mathcal{Y}$ .

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

$$P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^N P_{Y|X}(y_k|x_k). \quad (2.1)$$

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

$$\mathbf{X} \triangleq (X_1, \dots, X_N) \quad \text{and} \quad \mathbf{Y} \triangleq (Y_1, \dots, Y_N), \quad (2.2)$$

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

At the transmitter, a message  $\mathbf{m}$  drawn with uniform probability from a message set is mapped onto a codeword  $\mathbf{x}_{\mathbf{m}}$ , according to the encoding schemes described in Sections 2.2 and 2.3. We denote this encoding function by  $\phi$ , i.e.,  $\phi(\mathbf{m}) = \mathbf{x}_{\mathbf{m}}$ . Often, and unless strictly necessary, we drop the subindex  $\mathbf{m}$  in the codeword  $\mathbf{x}_{\mathbf{m}}$  and simply write  $\mathbf{x}$ . 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  $m$ ,

$$M \triangleq |\mathcal{X}|, \quad m \triangleq \log_2 M. \quad (2.3)$$

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

$$\hat{\mathbf{m}} = \varphi(\mathbf{y}) = \arg \max_{\mathbf{m} \in \{1, \dots, |\mathcal{M}|\}} q(\mathbf{x}_{\mathbf{m}}, \mathbf{y}). \quad (2.4)$$

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

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

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 Additive White Gaussian Noise (AWGN)

with fully interleaved fading,

$$y_k = h_k \sqrt{\text{snr}} x_k + z_k, \quad k = 1, \dots, N \quad (2.6)$$

where  $h_k$  are fading coefficients with unit variance,  $z_k$  are the zero-mean, unit-variance, circularly symmetric complex Gaussian samples, and  $\text{snr}$  is the signal-to-noise ratio (SNR). In Appendix 2.A we relate this discrete-time model to an underlying continuous-time model with AWGN. 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.<sup>1</sup>

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

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

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

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

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  $m_f \rightarrow +\infty$ , 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 (noncoherent detection), or only partially known, i.e., the receiver knows  $\hat{h}_k$  such that  $(H, \hat{H})$  are jointly distributed random variables. In this case, (2.7) generalizes to

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

<sup>1</sup> 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 Section 6.

The classical noncoherent channel where  $h = e^{j\theta}$ , with  $\theta$  denoting a uniformly distributed random phase, is a special case of (2.9) [18, 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  $\hat{h}$ , 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  $\mathbf{x}_m$  are in general nonbinary. At the receiver, a maximum metric decoder  $\varphi$  (as in Equation (2.4)) generates an estimate of the transmitted message,  $\varphi(\mathbf{y}) = \hat{\mathbf{m}}$ . 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

$$P_e \triangleq \frac{1}{|\mathcal{M}|} \sum_{m=1}^{|\mathcal{M}|} P_e(\mathbf{m}), \quad (2.10)$$

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

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

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

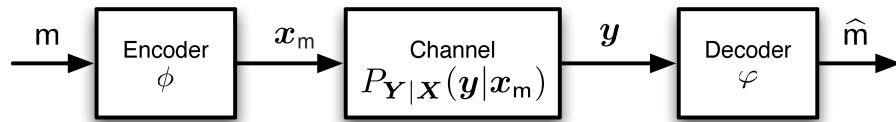


Fig. 2.1 Channel model with encoding and decoding functions.

## 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  $\mu: \{0,1\}^m \rightarrow \mathcal{X}$  which maps blocks of  $m$  bits to signal constellation symbols. The codewords of  $\mathcal{C}$  are denoted by  $\mathbf{c}$ . 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  $b_j: \mathcal{X} \rightarrow \{0,1\}$ , that is,  $b_j(x)$  is the  $j$ th bit of symbol  $x$ . Accordingly, we now define the sets

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

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  $\mathcal{X}_{b_{i_1}, \dots, b_{i_v}}^{j_{i_1}, \dots, j_{i_v}}$  as the sets of constellation points having the  $v$  binary labels  $b_{i_1}, \dots, b_{i_v}$  in positions  $j_{i_1}, \dots, j_{i_v}$ ,

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

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.

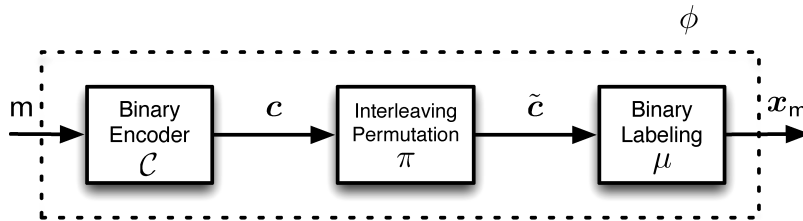


Fig. 2.2 BICM encoder model.

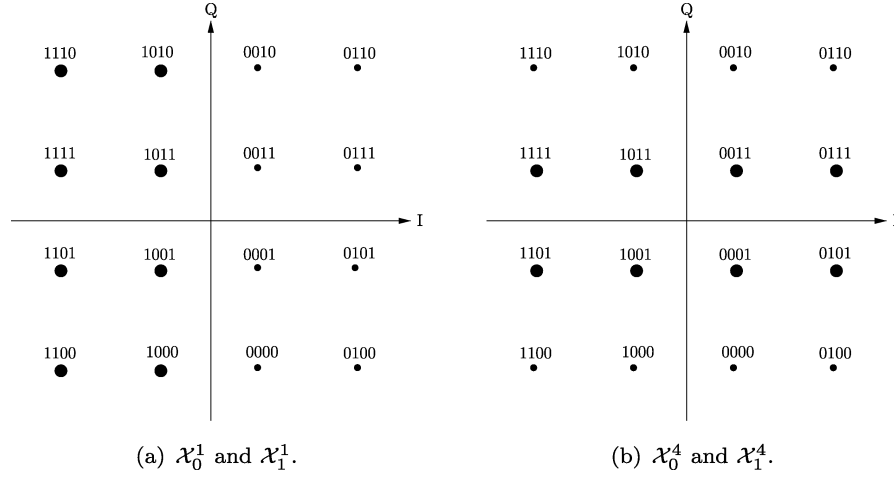


Fig. 2.3 Binary labeling sets  $\mathcal{X}_b^1$  and  $\mathcal{X}_b^4$  for 16-QAM with Gray mapping. Thin dots correspond to points in  $\mathcal{X}_0^i$  while thick dots correspond to points in  $\mathcal{X}_1^i$ .

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

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

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

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

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  $P_j(y|b)$ ,

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

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

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

In the log-domain this is known as the max-log approximation. Either of the symbol metrics corresponding to Equations (2.14) or (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 Equation (2.5) and outputs a binary codeword  $\hat{c}$  according to

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

### 2.3.2 BICM Classical Model

The  $m$  probabilities  $P_j(y|b)$  were used by Caire *et al.* in [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

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

as output, for  $j = 1, \dots, m$  and  $k = 1, \dots, N$ . We define the log-metric ratio vectors for each label bit as  $\Xi_j = (\Xi_j, \dots, \Xi_{m(k-1)+j})$  for  $j = 1, \dots, m$ . 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

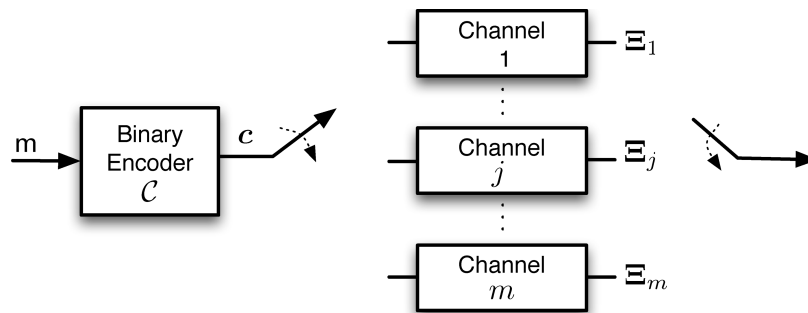


Fig. 2.4 Parallel channel model of BICM.

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 text, 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.

## Appendix 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

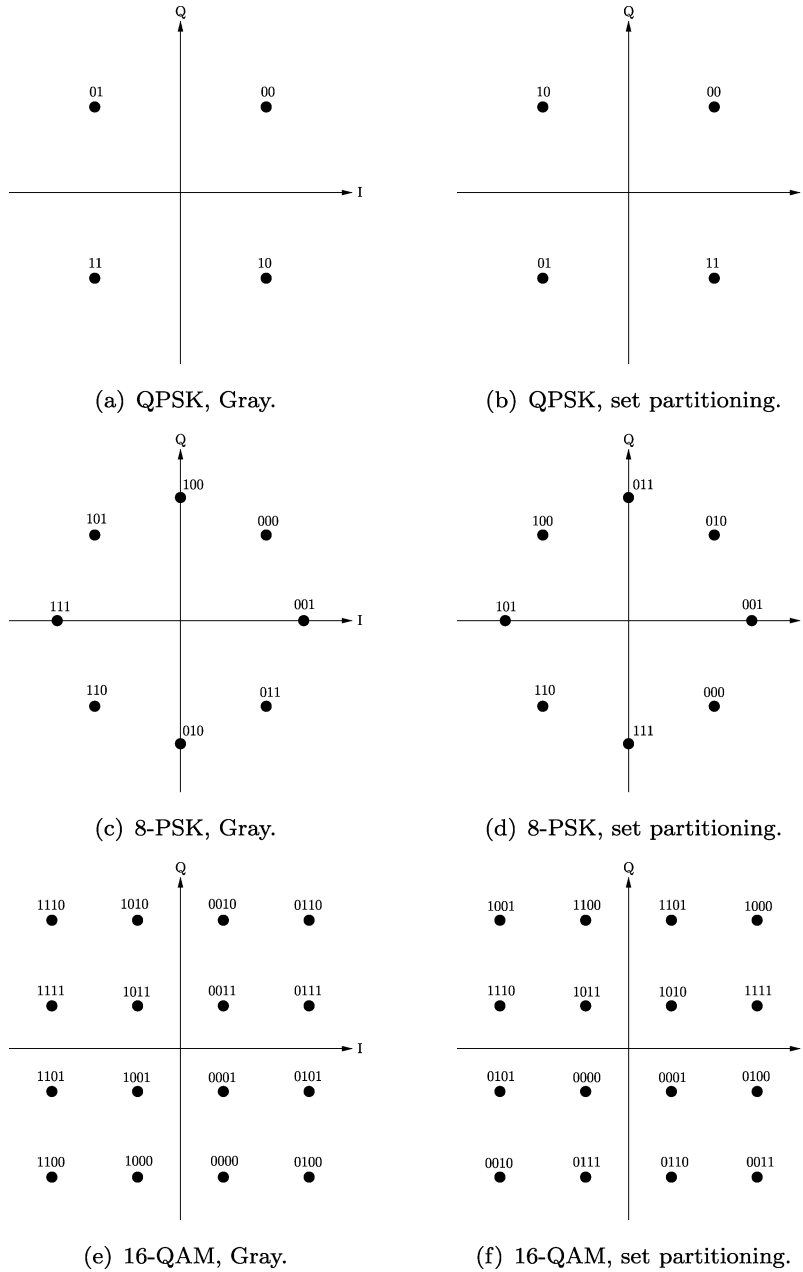


Fig. 2.5 Binary labeling rules (Gray, set partitioning) for QPSK, 8-PSK, and 16-QAM.

component  $z(t)$ , and output  $y(t)$  are related as

$$y(t) = \int h(t; \tau) x(\tau - t) d\tau + z(t), \quad (\text{A.1})$$

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  $\text{snr}$  as

$$\text{snr} = \frac{P}{N_0 W}. \quad (\text{A.2})$$

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

$$y_k = x_k + z_k, \quad (\text{A.3})$$

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  $\text{snr} = E_s / \sigma_Z^2$ . Observe that we recover the model in Equation (2.6) (with  $h_k = 1$ ) by dividing all quantities in Equation (A.3) by  $\sigma_Z$  and incorporating this coefficient in the definition of the channel variables.

Another channel of interest we use in the text is the frequency-nonselective, or flat fading channel. Its channel response  $h(t; \tau)$  is such that Equation (A.1) becomes

$$y(t) = h(t)x(t) + z(t). \quad (\text{A.4})$$

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 Equation (2.6),

$$y_k = h_k x_k + z_k. \quad (\text{A.5})$$

With the appropriate normalization by  $\sigma_Z$  we obtain the model in Equation (2.6). Throughout the text, we use the Nakagami- $m_f$  fading model in Equation (2.8), whereby coefficients are statistically independent from one another for different values of  $k$ . The squared fading coefficient  $g = |h|^2$  has density

$$P_G(g) = \frac{m_f^{m_f} g^{m_f-1}}{\Gamma(m_f)} e^{-m_f g}. \quad (\text{A.6})$$



# 3

---

## Information-Theoretic Foundations

---

In this section, we review the information-theoretic foundations of BICM. As suggested in the previous section, 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 (SNRs). For low SNRs — 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 SNRs. We also give a simple expression for the first derivative of the BICM mutual information with respect to the SNR, in terms of Minimum Mean-Squared Error (MMSE) for estimating the input of the channel from its output, and we relate this to the findings of Guo *et al.* [51] and Lozano *et al.* [67].

## 3.1 Coded Modulation

### 3.1.1 Channel Capacity

A coding rate<sup>1</sup>  $R$  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

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

---

<sup>1</sup> Capacities and information rates will be expressed using a generic logarithm, typically the natural logarithm. However, all charts in this text are expressed in bits.

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

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

For the maximum-likelihood decoders considered in Section 2.1,<sup>2</sup> 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

$$E(R) \triangleq \lim_{N \rightarrow \infty} -\frac{1}{N} \log \inf_{\mathcal{M}} P_e(\mathcal{M}), \quad (3.3)$$

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. In particular, 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]

$$C = \log(1 + \text{snr}). \quad (3.4)$$

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  $\mathcal{X}$

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

of finite cardinality, such as PSK or QAM [95], with a uniform input distribution,  $P_X(x) = \frac{1}{2^m}$  for all  $x \in \mathcal{X}$ . While a uniform distribution is only optimal for large  $\text{snr}$ , 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  $\text{snr}$  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 text.

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  $C_{\mathcal{X}}^{\text{cm}}$  or  $I^{\text{cm}}(X; Y)$ , that is

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

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.,  $C_{\mathcal{X}}^{\text{cm}} \leq m$  bits.

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

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

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

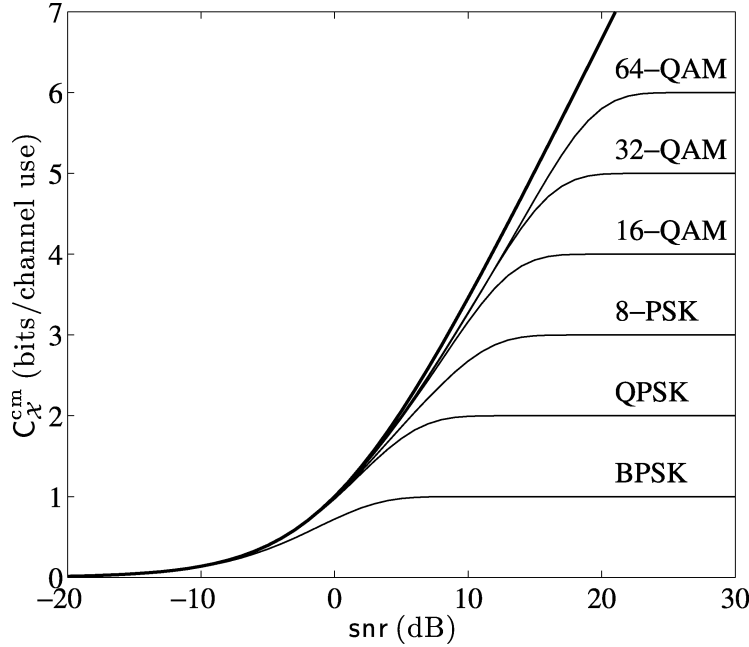


Fig. 3.1 Coded modulation capacity in bits per channel use for multiple signal constellations with uniform inputs in the AWGN channel. For reference, the channel capacity with Gaussian inputs (3.4) is shown in thick lines.

The symmetry of the code construction makes  $\bar{P}_e(\mathbf{m})$  independent of the index  $\mathbf{m}$ , and hence  $\bar{P}_e = \bar{P}_e(\mathbf{m})$  for any  $\mathbf{m} \in \mathcal{M}$ . Averaged over the random code ensemble, we have that

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

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

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

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

Using the union bound over all possible codewords, and for all  $0 \leq \rho \leq 1$ , the probability  $\Pr\{\varphi(\mathbf{y}) \neq \mathbf{m} | \mathbf{x}_m, \mathbf{y}\}$  can be bounded by [39, p. 136]

$$\begin{aligned} \Pr\{\varphi(\mathbf{y}) \neq \mathbf{m} | \mathbf{x}_m, \mathbf{y}\} &\leq \Pr\left\{\bigcup_{\mathbf{m}' \neq \mathbf{m}} \{\varphi(\mathbf{y}) = \mathbf{m}' | \mathbf{x}_m, \mathbf{y}\}\right\} \\ &\leq \left(\sum_{\mathbf{m}' \neq \mathbf{m}} \Pr\{\varphi(\mathbf{y}) = \mathbf{m}' | \mathbf{x}_m, \mathbf{y}\}\right)^\rho. \end{aligned} \quad (3.9)$$

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

$$\Pr\{\varphi(\mathbf{y}) = \mathbf{m}' | \mathbf{x}_m, \mathbf{y}\} \leq \sum_{\mathbf{x}_{m'}} P_{\mathbf{X}}(\mathbf{x}_{m'}) \left(\frac{q(\mathbf{x}_{m'}, \mathbf{y})}{q(\mathbf{x}_m, \mathbf{y})}\right)^s. \quad (3.10)$$

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

$$\Pr\{\varphi(\mathbf{y}) \neq \mathbf{m} | \mathbf{x}_m, \mathbf{y}\} \leq \left((|\mathcal{M}| - 1) \sum_{\mathbf{x}_{m'}} P_{\mathbf{X}}(\mathbf{x}_{m'}) \left(\frac{q(\mathbf{x}_{m'}, \mathbf{y})}{q(\mathbf{x}_m, \mathbf{y})}\right)^s\right)^\rho. \quad (3.11)$$

Therefore, Equation (3.7) can be written as

$$\bar{P}_e \leq (|\mathcal{M}| - 1)^\rho \mathbb{E} \left[ \left( \sum_{\mathbf{x}_{m'}} P_{\mathbf{X}}(\mathbf{x}_{m'}) \left(\frac{q(\mathbf{x}_{m'}, \mathbf{Y})}{q(\mathbf{x}_m, \mathbf{Y})}\right)^s \right)^\rho \right]. \quad (3.12)$$

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

$$\bar{P}_e \leq (|\mathcal{M}| - 1)^\rho \left( \mathbb{E} \left[ \left( \sum_{x'} P_X(x') \left(\frac{q(x', Y)}{q(X, Y)}\right)^s \right)^\rho \right] \right)^N. \quad (3.13)$$

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

$$\bar{P}_e \leq e^{-N(E_0^q(\rho, s) - \rho R)}, \quad (3.14)$$

where

$$E_0^q(\rho, s) \triangleq -\log \mathbb{E} \left[ \left( \sum_{x'} P_X(x') \left(\frac{q(x', Y)}{q(X, Y)}\right)^s \right)^\rho \right] \quad (3.15)$$

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

$$E_r^q(R) \triangleq \max_{0 \leq \rho \leq 1} \max_{s > 0} (E_0^q(\rho, s) - \rho R). \quad (3.16)$$

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

According to (3.14), 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  $\rho$  vanishes, rates below

$$\lim_{\rho \rightarrow 0} \frac{E_0^q(\rho, s)}{\rho} \quad (3.17)$$

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  $I_s^{\text{gmi}}(X;Y)$  as

$$\begin{aligned} I_s^{\text{gmi}}(X;Y) &\triangleq \left. \frac{\partial E_0^q(\rho, s)}{\partial \rho} \right|_{\rho=0} = \lim_{\rho \rightarrow 0} \frac{E_0^q(\rho, s)}{\rho} \\ &= -\mathbb{E} \left[ \log \sum_{x'} P_X(x') \left( \frac{q(x', Y)}{q(X, Y)} \right)^s \right] \\ &= \mathbb{E} \left[ \log \frac{q(X, Y)^s}{\sum_{x' \in \mathcal{X}} P_X(x') q(x', Y)^s} \right]. \end{aligned} \quad (3.18)$$

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

$$I^{\text{gmi}}(X;Y) = \max_{s > 0} I_s^{\text{gmi}}(X;Y). \quad (3.19)$$

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

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

$$R_0 \triangleq E_r^q(R=0) = \max_{s > 0} E_0^q(1, s). \quad (3.20)$$

For a maximum likelihood decoder,  $E_0^q(\rho, s)$  is maximized by letting  $s = \frac{1}{1+\rho}$  [39], and we have

$$\begin{aligned} E_0(\rho) &\triangleq -\log \mathbb{E} \left[ \left( \sum_{x'} P_X(x') \left( \frac{P_{Y|X}(Y|x')}{P_{Y|X}(Y|X)} \right)^{\frac{1}{1+\rho}} \right)^\rho \right] \\ &= -\log \int_y \left( \sum_x P_X(x) P_{Y|X}(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} dy, \end{aligned} \quad (3.21)$$

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

$$E_0^{\text{cm}}(\rho) \triangleq -\log \int_y \left( \frac{1}{2^m} \sum_x P_{Y|X}(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} dy. \quad (3.22)$$

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

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 3.2 (Data-Processing Inequality [59, 78]).** For  $s > 0$ ,  $0 \leq \rho \leq 1$ , and a given input distribution we have that

$$E_0^q(\rho, s) \leq E_0(\rho). \quad (3.23)$$

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

$$P_{Y|X}(y|x) = c' q(x, y)^{s'} \quad \text{for all } x \in \mathcal{X} \quad (3.24)$$

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

$$I^{\text{ind}}(X; Y) \triangleq \sum_{j=1}^m \mathbb{E} \left[ \log \frac{\sum_{x' \in \mathcal{X}_B^j} P_{Y|X}(Y|x')}{\frac{1}{2} \sum_{x' \in \mathcal{X}} P_{Y|X}(Y|x')} \right], \quad (3.25)$$

and

$$R_0^{\text{ind}} \triangleq m \log 2 - \sum_{j=1}^m \log \left( 1 + \mathbb{E} \left[ \sqrt{\frac{\sum_{x' \in \mathcal{X}_B^j} P_{Y|X}(Y|x')}{\sum_{x' \in \mathcal{X}_B^j} P_{Y|X}(Y|x')}} \right] \right), \quad (3.26)$$

respectively. An underlying assumption behind Equation (3.26) 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],

$$R_0^{\text{av}} \triangleq m \left[ \log 2 - \log \left( 1 + \frac{1}{m} \sum_{j=1}^m \mathbb{E} \left[ \sqrt{\frac{\sum_{x' \in \mathcal{X}_B^j} P_{Y|X}(Y|x')}{\sum_{x' \in \mathcal{X}_B^j} P_{Y|X}(Y|x')}} \right] \right) \right]. \quad (3.27)$$

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  $R_0^{\text{av}} \leq R_0^{\text{ind}}$ .

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

$$C_{\mathcal{X}}^{\text{bicm}} \triangleq I^{\text{ind}}(X; Y). \quad (3.28)$$

The following alternative expression [26, 77, 140] for the BICM mutual information turns out to be useful,

$$C_{\mathcal{X}}^{\text{bicm}} = \sum_{j=1}^m \frac{1}{2} \sum_{b=0}^1 \left( C_{\mathcal{X}}^{\text{cm}} - C_{\mathcal{X}_b^j}^{\text{cm}} \right), \quad (3.29)$$

where  $C_{\mathcal{A}}^{\text{cm}}$  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 3.3 ([78]).** 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,

$$I^{\text{gmi}}(X; Y) = \sup_{s>0} \sum_{j=1}^m \mathbb{E} \left[ \log \frac{q_j(b, Y)^s}{\frac{1}{2} \sum_{b'=0}^1 q_j(b', Y)^s} \right]. \quad (3.30)$$


---

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

---

**Corollary 3.1 ([78]).** For the metric in Equation (2.15),

$$I^{\text{gmi}}(X; Y) = C_{\mathcal{X}}^{\text{bicm}}. \quad (3.31)$$


---

Expression (3.31) 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 ([78]).** For the metric in Equation (2.17),

$$I^{\text{gmi}}(X; Y) = \sup_{s>0} \sum_{j=1}^m \mathbb{E} \left[ \log \frac{(\max_{x \in \mathcal{X}_j^B} p(y|x))^s}{\frac{1}{2} \sum_{b=0}^1 (\max_{x' \in \mathcal{X}_b^j} p(y|x'))^s} \right]. \quad (3.32)$$


---

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 (3.32) 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

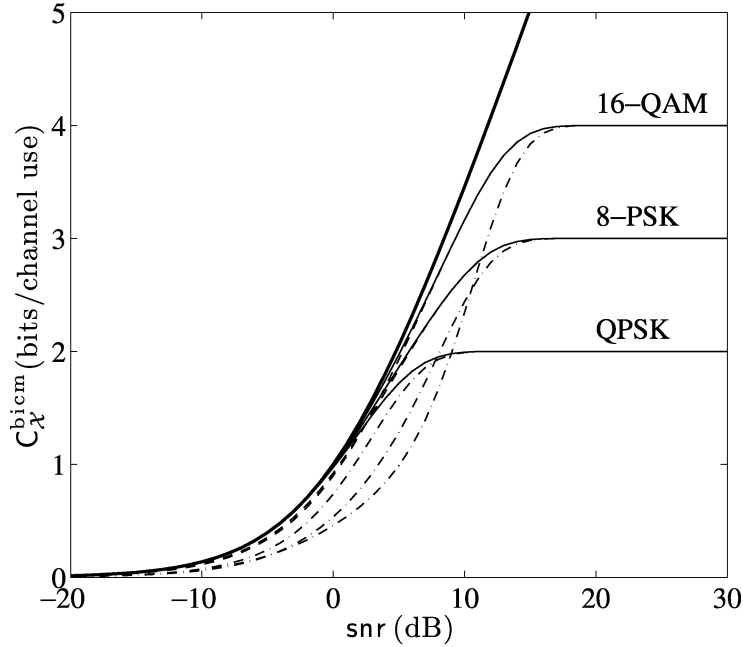


Fig. 3.2 Coded modulation and BICM capacities (in bits per channel use) for multiple signal constellations with uniform inputs in the AWGN channel. Gray and set partitioning labeling rules correspond to dashed and dash-dotted lines, respectively. In thick solid lines, the capacity with Gaussian inputs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

the AWGN channel, as a function of  $\text{snr}$ . 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 Equation (3.15) for BICM with a bit metric  $q_j(b, y)$  yields a function  $E_0^{\text{bicm}}(\rho, s)$  of  $\rho$  and  $s$ ,

$$E_0^{\text{bicm}}(\rho, s) \triangleq -\log \mathbb{E} \left[ \left( \frac{1}{2^m} \sum_{x' \in \mathcal{X}} \prod_{j=1}^m \frac{q_j(b_j(x'), Y)^s}{q_j(b_j(X), Y)^s} \right)^\rho \right]. \quad (3.33)$$

Moreover, the data processing inequality for error exponents in Proposition 3.2 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  $E_0^{\text{bicm}}(\rho, s) \leq E_0^{\text{cm}}(\rho)$ .

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

$$E_0^{\text{ind}}(\rho) \triangleq - \sum_{j=1}^m \log \mathbb{E} \left[ \left( \sum_{b'=0}^1 P_{B_j}(b') \frac{P_j(Y|b')^{\frac{1}{1+\rho}}}{P_j(Y|B)^{\frac{1}{1+\rho}}} \right)^\rho \right]. \quad (3.34)$$

This quantity is the random coding exponent of the BICM decoder if the channel output  $\mathbf{y}$  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  $\text{snr} = 5, 15, -25$  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

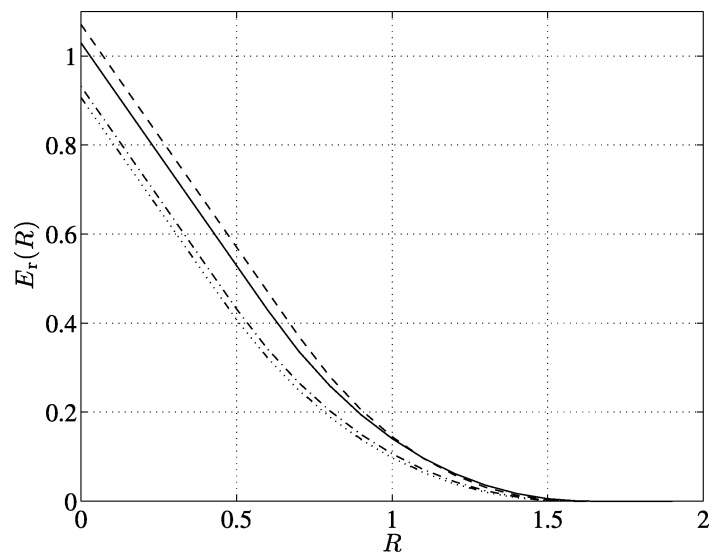
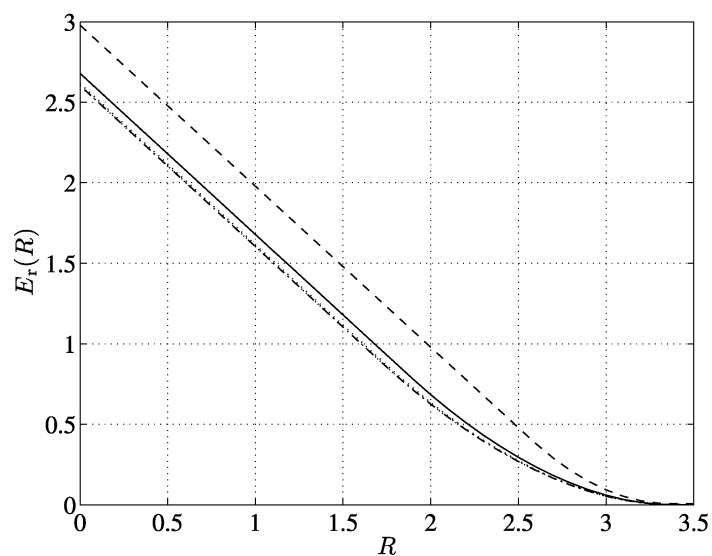
(a)  $\text{snr} = 5$  dB.(b)  $\text{snr} = 15$  dB.

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.

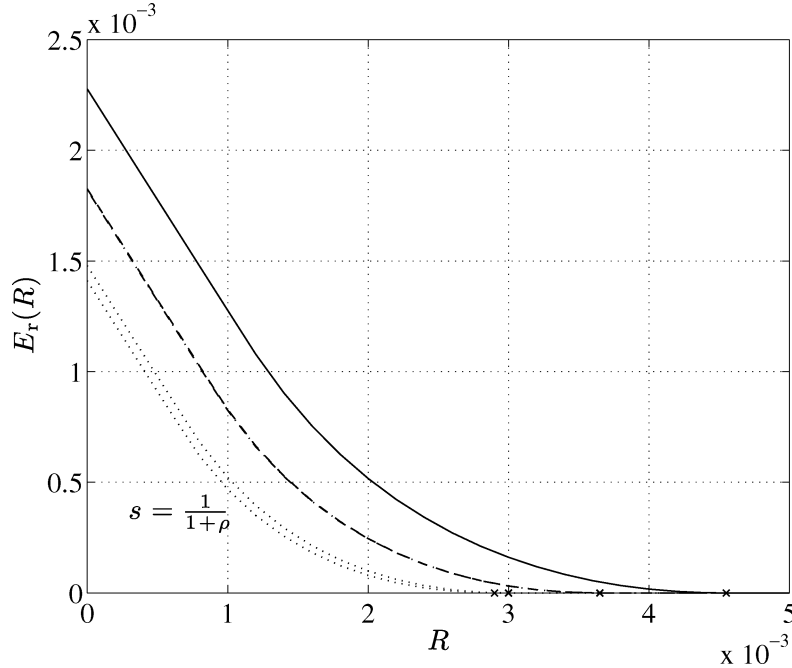


Fig. 3.4 Error exponents for coded modulation (solid), BICM with independent parallel channels (dashed), BICM using metric (2.15) (dash-dotted), and BICM using metric (2.17) (dotted) for 16-QAM with Gray labeling, Rayleigh fading and  $\text{snr} = -25$  dB. Crosses correspond to (from right to left)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

at low SNR for not optimizing over  $s$ . We denote with crosses the 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

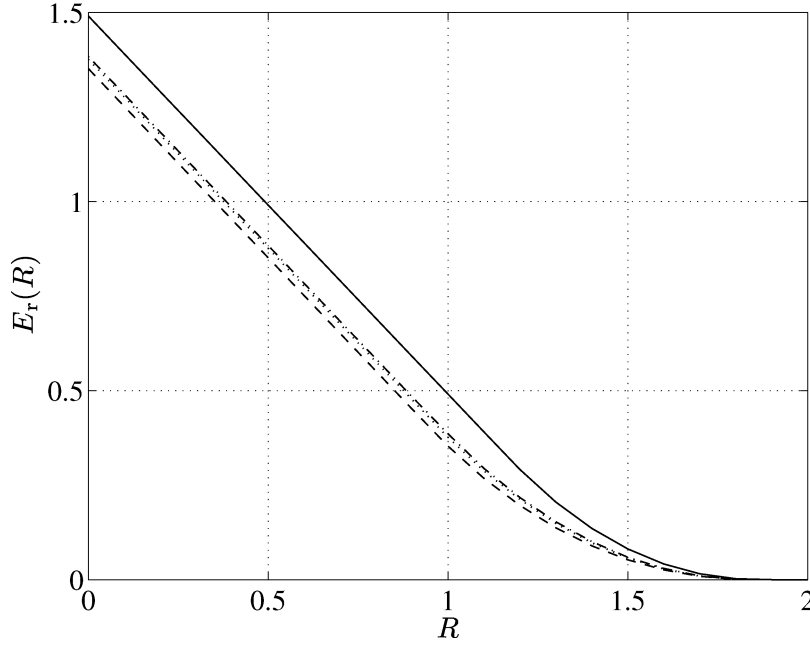


Fig. 3.5 Error exponents for coded modulation (solid), BICM with independent parallel channels (dashed), BICM using metric (2.15) (dash-dotted), and BICM using metric (2.17) (dotted) for 8-PSK with Gray labeling, AWGN and  $\text{snr} = 5$  dB.

channel capacity by using binary codes. In this section, we compare BICM with 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  $\mu$ . A uniform distribution over the channel input set induces a uniform distribution over the input bits  $b_j$ ,  $j = 1, \dots, m$ . 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.,  $\mathcal{C} = \mathcal{C}_1 \times \dots \times \mathcal{C}_m$ , and the input distribution for the symbol  $x(b_1, \dots, b_j)$  has the form:

$$P_X(x) = P_{B_1, \dots, B_M}(b_1, \dots, b_m) = \prod_{j=1}^m P_{B_j}(b_j). \quad (3.35)$$

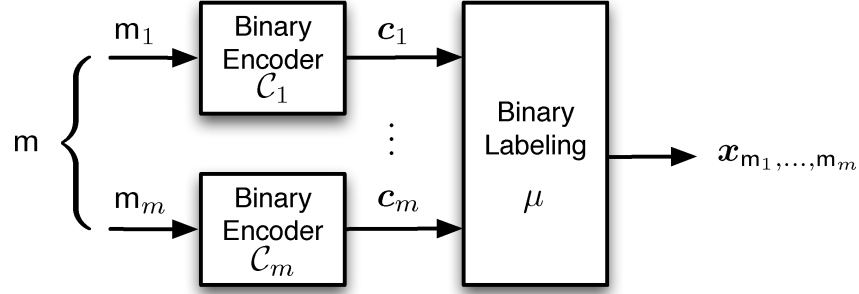


Fig. 3.6 Block diagram of a multi-level encoder.

Denoting the rate of the  $j$ th level code by  $R_j$ , the resulting total rate of the MLC scheme is  $R = \sum_{j=1}^m R_j$ . We denote the codewords of  $\mathcal{C}_j$  by  $c_j$ . 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:

$$q(x, y) = \prod_{j=1}^m q_j(b_j(x), y). \quad (3.36)$$

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

$$q_j(b_j(x) = b, y) = \frac{1}{|\mathcal{X}_{b_1, \dots, b_{j-1}, b}^{1, \dots, j}|} \sum_{x' \in \mathcal{X}_{b_1, \dots, b_{j-1}, b}^{1, \dots, j}} P_{Y|X}(y|x'). \quad (3.37)$$

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 Equation (3.37). 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 Wachsmann *et al.* [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 in [40], and

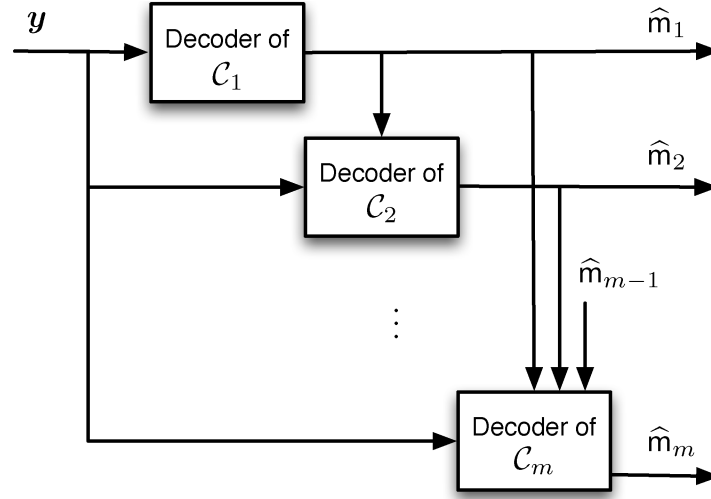


Fig. 3.7 Block diagram of a multi-stage decoder for MLC.

we now review it for a simple 2-level case. Generalization to a larger number of levels is straightforward.

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,

$$\bar{P}_e = \frac{1}{|\mathcal{C}_1||\mathcal{C}_2|} \sum_{\mathbf{m}_1=1}^{|\mathcal{C}_1|} \sum_{\mathbf{m}_2=1}^{|\mathcal{C}_2|} \bar{P}_e(\mathbf{m}_1, \mathbf{m}_2), \quad (3.38)$$

where  $\bar{P}_e(\mathbf{m}_1, \mathbf{m}_2)$  denotes the average error probability over the code ensemble when messages  $\mathbf{m}_1$  and  $\mathbf{m}_2$  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  $\bar{P}_e = \bar{P}_e(\mathbf{m}_1, \mathbf{m}_2)$  for any messages  $\mathbf{m}_1, \mathbf{m}_2$ . If  $\mathbf{m}_1, \mathbf{m}_2$  are the selected messages, then  $\mathbf{x}_{\mathbf{m}_1, \mathbf{m}_2}$  denotes the sequence of modulation symbols corresponding to these messages.

For a given received sequence  $\mathbf{y}$ , the decoders  $\varphi_1$  and  $\varphi_2$  choose the messages  $\hat{\mathbf{m}}_1, \hat{\mathbf{m}}_2$  with largest metric  $q(\mathbf{x}_{\mathbf{m}_1, \mathbf{m}_2}, \mathbf{y})$ . 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 ( $\varphi_1(\mathbf{y}) \neq 1$ ), but that for  $\mathcal{C}_2$  is successful ( $\varphi_2(\mathbf{y}) = 1$ ): the decoded codeword is  $\mathbf{x}_{\mathbf{m}_1,1}$ ,  $\mathbf{m}_1 \neq 1$ .
- (2) The decoder for  $\mathcal{C}_2$  fails ( $\varphi_2(\mathbf{y}) \neq 1$ ), but that for  $\mathcal{C}_1$  is successful ( $\varphi_1(\mathbf{y}) = 1$ ): the decoded codeword is  $\mathbf{x}_{1,\mathbf{m}_2}$ ,  $\mathbf{m}_2 \neq 1$ .
- (3) Both decoders for  $\mathcal{C}_1$  and  $\mathcal{C}_2$  fail ( $\varphi_1(\mathbf{y}) \neq 1, \varphi_2(\mathbf{y}) \neq 1$ ): the decoded codeword is  $\mathbf{x}_{\mathbf{m}_1,\mathbf{m}_2}$ ,  $\mathbf{m}_1 \neq 1, \mathbf{m}_2 \neq 1$ .

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  $\Pr\{\text{error}|\mathbf{x}_{1,1}, \mathbf{y}\}$  for a given choice of transmitted codeword  $\mathbf{x}_{1,1}$  yields

$$\Pr\{\text{error}|\mathbf{x}_{1,1}, \mathbf{y}\} \leq P_{(1)} + P_{(2)} + P_{(1,2)}. \quad (3.39)$$

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  $\mathbf{x}_{\mathbf{m}_1,1}$ . Observe that the error probability is not decreased if the decoder has access to a genie giving the value of  $\mathbf{c}_2$  [98]. As in the derivation of Equation (3.11), we obtain that

$$P_{(1)} \leq (|\mathcal{C}_1| - 1)^\rho \left( \sum_{\mathbf{x}_{\mathbf{m}_1,1}} P_{\mathbf{X}}(\mathbf{x}_{\mathbf{m}_1,\mathbf{m}_2}) \frac{q(\mathbf{x}_{\mathbf{m}_1,1}, \mathbf{y})^s}{q(\mathbf{x}_{1,1}, \mathbf{y})^s} \right)^\rho. \quad (3.40)$$

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

$$\bar{P}_{(1)} \leq e^{-N(E_{0,(1)}^q(\rho,s) - \rho R_1)}, \quad (3.41)$$

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

$$E_{0,(1)}^q(\rho, s) \triangleq -\log \mathbb{E} \left[ \left( \sum_{b'_1} P_{B_1}(b'_1) \frac{q(\mu(b'_1, B_2), Y)^s}{q(\mu(B_1, B_2), Y)^s} \right)^\rho \right]. \quad (3.42)$$

Similarly, for the second summand, the probability satisfies

$$\bar{P}_{(2)} \leq e^{-N(E_{0,(2)}^q(\rho,s) - \rho R_2)}, \quad (3.43)$$