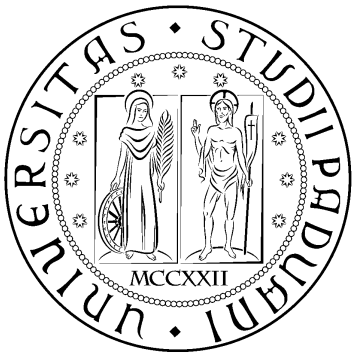

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
CORSO DI LAUREA TRIENNALE IN
INGEGNERIA DELLE TELECOMUNICAZIONI
TESI DI LAUREA

BIT-INTERLEAVED CODED MODULATION

RELATORE: Prof. Nevio Benvenuto

LAUREANDO: *Alberto Desiderà*

Padova, 30 settembre 2010

Opera in modo che la massima della tua volontà

possa sempre valere in ogni tempo come principio

di una legislazione universale.

(Immanuel Kant)

Contents

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*, and *error correcting codes*. More recently, powerful families of binary codes with a good trade-off between performance and decoding complexity have been 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 matter of fact, has established itself as a quasi-standard (*de-facto*) for bandwidth - and power - efficient communication, like DSL, wireless LANs, WiMax. The aim of this thesis is to describe the main aspects of the system, focusing the attention on model characteristics and on the error analysis (based on bit-error rate approximations). Finally I also consider the BICM with iterative decoding and I conclude with an overview of some applications of BICM.

Chapter 1

Introduction

Since Shannon's landmark 1948 paper [1], 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. A simply 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.

Starting from the trellis-coded modulation analysed in [2], it has been generally accepted that modulation and coding should be combined in a single entity for improved performance, based on the combination of trellis codes and discrete signal constellations through set partitioning. The code performance in this situation depends strongly, rather than on the minimum Euclidean distance of the code, on its minimum Hamming distance (the 'code diversity').

The discovery of turbo-codes and the re-discovery of low-density parity-check (LDPC) codes, with their corresponding iterative decoding algorithms marked a new era in coding theory. These modern codes approach the capacity of binary-input channels with low-complexity. The analysis of iterative decoding also led to new methods for their efficient design.

In contrast with findings in [2], in [3] is proposed the bit-interleaved coded modulation (BICM) as a pragmatic approach to coded modulation. Here is recognized that the code diversity, and hence the reliability of coded modulation over the

channel, could be further improved. The idea is to make the code diversity equal to the smallest number of distinct bits (rather than channel symbols) along any error event. This is achieved by bit-wise interleaving at the encoder output, and by using an appropriate soft-decision bit metric as an input to the decoder. In [4] there is a comprehensive analysis of BICM in terms of information rates and error probability, and is showed 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. Since its introduction, BICM has been regarded as a pragmatic yet a 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 system (the so-called 4th generation). BICM has become the *de-facto* standard for coding over Gaussian channel in modern systems.

In this work, I introduce in a comprehensive fashion the theory underlying BICM and also tools for evaluating the performance of BICM. First I define the system model, analysing the channel model, the scheme of coded modulation and finally the bit-interleaved coded modulation. After that I focus my attention on BICM, considering a more specific characterization; starting from this I introduce some approximations for the study of the error-rate, achieving a much important result: the closed-form expressions for some kind of well-chosen subsets of signal points, useful for computing the numerical integration of approximations. These are valid for general quadrature amplitude modulation and phase shift keying signal constellations and arbitrary bit-to-symbol mapping rules. I then turn my attention to iterative decoding of BICM (BICM-ID) that provide some performance advantages using convolutional codes combined with mappings different from the standards (as Gray labelling). After all I describe a number of application of BICM; in particular I consider the application of BICM to orthogonal modulation with non-coherent detection, to block-fading channel and to the multiple-antenna channel (MIMO).

Chapter 2

System model

In this chapter I recall the baseline model of coded modulation(CM) and introduce the model of bit-interleaved coded modulation(BICM). Here I consider a generic stationary finite-memory vector channel, that in the next analysis will be more specified.

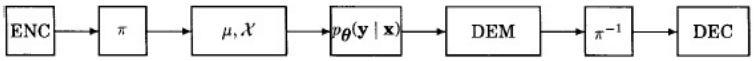

Figure 2.1: Block diagram of transmission with coded modulation(CM) and bit-interleaved coded modulation(BICM). For CM, π denotes interleaving at the symbol level; for BICM, π denotes interleaving at the bit level.

The CM and BICM models are illustrated in the block diagram of Figure 2.1. The main blocks of both schemes are 1) an encoder (ENC); 2) an interleaver π ; 3) a modulator, modelled by a labelling map μ and a signal set $\mathcal{X}$, as finite set of points in the complex N-dimensional Euclidean space $\mathbb{C}^N$; 4) a stationary finite-memory vector channel whose transition probability density function $p_{\boldsymbol{\theta}}(\boldsymbol{y} \mid \boldsymbol{x}), \boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$; 5) a demodulator (DEM); 6) a branch metric deinterleaver π^{-1}; 7) a decoder (DEC). In this first part of discussion I make a detailed description of these blocks for CM and BICM.

2.1 Channel model

Now, assuming that $\boldsymbol{x}$ is the input and $\boldsymbol{y}$ is the output of system ($\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N$), I consider a vector channel characterized by a family of transition probability

density function (pdf)

$$\left\{p_{\boldsymbol{\theta}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right) : \boldsymbol{\theta} \in \mathbb{C}^K; \boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^N\right\} \tag{2.1}$$

parametrized by a complex-valued vector $\boldsymbol{\theta} \in \mathbb{C}^K$ (where $\mathbb{C}^K$ is a complex K-dimensional Euclidean space, different from space of input and output vectors) representing the channel state. I assume that $\boldsymbol{\theta}$ is independent of the channel input $\boldsymbol{x}$, and that, conditionally on the sequence $\underline{\boldsymbol{\theta}}$, the channel is memoryless,

$$p_{\underline{\boldsymbol{\theta}}}\left(\underline{\boldsymbol{y}} \mid \underline{\boldsymbol{x}}\right) = \prod_k p_{\boldsymbol{\theta}_k}\left(\boldsymbol{y}_k \mid \boldsymbol{x}_k\right), \tag{2.2}$$

where $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ are the input and output sequences interested in the transmission. Moreover, $\underline{\boldsymbol{\theta}}$ is assumed to be a stationary, finite memory random process. By this model I can represent a large number of typical communication channels, for example the additive white Gaussian noise (AWGN) channel ($\boldsymbol{\theta}$ = constant). In the following sections I will take into account also the fading contribution, and then the channel will be consider as a pdf according to Nakagami-m distribution.

2.2 Coded modulation

A CM scheme can be viewed as the concatenation of an encoder (ENC) for a code $\mathcal{C}$ defined over a label alphabet $\mathcal{A}$ with an N-dimensional memoryless modulator over a signal set $\mathcal{X} \subseteq \mathbb{C}^N$. I assume that $\mathcal{A}$ and $\mathcal{X}$ are finite and $|\mathcal{A}| = |\mathcal{X}| = M$, where then I define $r = \log_2 M$. The transmitted code sequence $\underline{\boldsymbol{c}} \in \mathcal{C}$ is mapped onto the signal sequence $\underline{\boldsymbol{x}}$ by the labelling map $\mu : \mathcal{A} \longrightarrow \mathcal{X}$, acting componentwise. If the code $\mathcal{C}$ is designed to correct random errors, it may be convenient to introduce a *symbol interleaver* π between the encoder (ENC) and the mapper μ. For all this work I assume that π is an ideal interleaver. Let $\underline{\boldsymbol{y}} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_k, \cdots)$ denote the channel output *sequence* resulting from the transmission of the *sequence* $\underline{\boldsymbol{x}}$. Optimum decoding depends on the channel state information (CSI) available at the receiver. Here I will consider only the case of *perfect CSI.*

Detection: assuming that the receiver has perfect knowledge of the values taken on by each $\boldsymbol{\theta}_k$, the set of maximum-likelihood (ML) symbol metrics is given by

$$\left\{\log p_{\boldsymbol{\theta}_k}\left(\boldsymbol{y}_k \mid \boldsymbol{z}\right)\right\}_{\boldsymbol{z} \in \mathcal{X}} \tag{2.3}$$

After metric deinterleaving (denoted by $\pi^{-}1$ in Figure 2.1), ML decision about the transmitted code sequence is made according to the rule,

$$\hat{\underline{\boldsymbol{c}}} = \arg\max_{\underline{\boldsymbol{c}} \in \mathcal{C}} \sum_{k} \log p_{\boldsymbol{\theta}_k} \left(\boldsymbol{y}_k \mid \mu(c_k) \right) \tag{2.4}$$

where the summation runs over the whole sequence. This equation represents the ML decision rule also with finite-depth or with no interleaving, since conditionally on $\underline{\boldsymbol{\theta}}$, the channel is memoryless.

2.3 Bit-interleaved coded modulation

BICM can be obtained by concatenating an encoder ENC for binary code $\mathcal{C}$, with an N-dimensional memoryless modulator over a signal set $\mathcal{X} \subseteq \mathbb{C}^N$ of size $|\mathcal{X}| = M = 2^r$, through a bit interleaver π and a one-to-one binary labelling map $\mu : \{0,1\}^r \to \mathcal{X}$. The code sequence $\underline{\boldsymbol{c}}$ is first interleaved by π. Next, the interleaved sequence $\pi\left(\underline{\boldsymbol{c}}\right)$ is broken into subsequences of r bits each, which are mapped onto signals in $\mathcal{X}$. Finally, the resulting signal sequence $\underline{\boldsymbol{x}}$ is transmitted over the vector channel. The bit interleaver can be seen as a one-to-one correspondence $\pi :\to (k', i)$, where k denotes the time ordering of the coded bits c_k, k' denotes the time ordering of the signals $\boldsymbol{x}'_k$ transmitted over the vector channel, and i indicates the position of the bit c_k in the label of $\boldsymbol{x}'_k$. From here I assume ideal bit interleaving to make simplified analysis. Let $l^i\left(\boldsymbol{x}\right)$ denote the ith bit of the label of $\boldsymbol{x}$, and $\mathcal{X}_b^i$ the subset of all the signals $\boldsymbol{x} \in \mathcal{X}$ whose label has the value $b \in \{0,1\}$ in position i. Further, let $\bar{b}$ denote the complement of b and $\boldsymbol{y}$ the channel output resulting from the transmission of $\boldsymbol{x}$. The conditional pdf of $\boldsymbol{y}$, given $l^i\left(\boldsymbol{x}\right) = b \in \{0,1\}$ and $\boldsymbol{\theta}$, is

$$\begin{aligned} p_{\boldsymbol{\theta}}\left(\boldsymbol{y} \mid l^i\left(\boldsymbol{x}\right) = b\right) &= \sum_{\boldsymbol{x} \in \mathcal{X}} p_{\boldsymbol{\theta}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right) P\left(\boldsymbol{x} \mid l^i\left(\boldsymbol{x}\right) = b\right) \\ &= 2^{-(r-1)} \sum_{\boldsymbol{x} \in \mathcal{X}_b^i} p_{\boldsymbol{\theta}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right) \end{aligned} \tag{2.5}$$

The last expression is obtained by noting that

$$p_{\boldsymbol{\theta}}\left(\boldsymbol{y} \mid \boldsymbol{x}, l^i\left(\boldsymbol{x}\right) = b\right) = p_{\boldsymbol{\theta}}\left(\boldsymbol{y} \mid \boldsymbol{x}\right)$$

and that

$$P\left(\boldsymbol{x} \mid l^i\left(\boldsymbol{x}\right) = b\right) = \begin{cases} 0, & if \quad l^i\left(\boldsymbol{x}\right) = \bar{b} \\ 2^{-(r-1)}, & if \quad l^i\left(\boldsymbol{x}\right) = b \end{cases}$$

where I assume uniform input distribution

$$P(b=0) = P(b=1) = \frac{1}{2}$$

The optimum receiver depends on the available CSI that in this case is ideal; the bit interleaving is also ideal and I can assume that at each time k' the demodulator (DEM) produces the set of ML bit metrics

$$\lambda^i(\boldsymbol{y}_{k'}, b) = \log \sum_{\boldsymbol{x} \in \mathcal{X}_b^i} p_{\boldsymbol{\theta}_{k'}}(\boldsymbol{y}_{k'} \mid \boldsymbol{x}) \tag{2.6}$$

for $b \in \{0,1\}$ and $i = 1, \cdots, r$. Finally, the ML decoder (DEC) makes decisions according to the rule

$$\hat{\underline{\boldsymbol{c}}} = \arg \max_{c_k \to \underline{\boldsymbol{c}} \in \mathcal{C}} \sum_k \lambda^i(\boldsymbol{y}_{k'}, c_k), \tag{2.7}$$

where $c_k \to \underline{\boldsymbol{c}} \in \mathcal{C}$ indicates that c_k is a coded bit of the code sequence $\underline{\boldsymbol{c}}$. The decision rule (2.7) is completely analogous to (2.4), and can be implemented in a simply way.

Chapter 3

BICM analysis and approximations

After the introduction to the system model of previous chapter, I want to get a more specific characterization for BICM systems. As in my analysis I will consider the transmission performances over fading channels, I use a Nakagami-m distribution to model the channel aspects. Assuming this, I evaluate some approximations for the error-rate, comparing the results with union bound found in [4]. This work is very useful for the system design and in next sections I will derive other simplified expressions for this purpose.

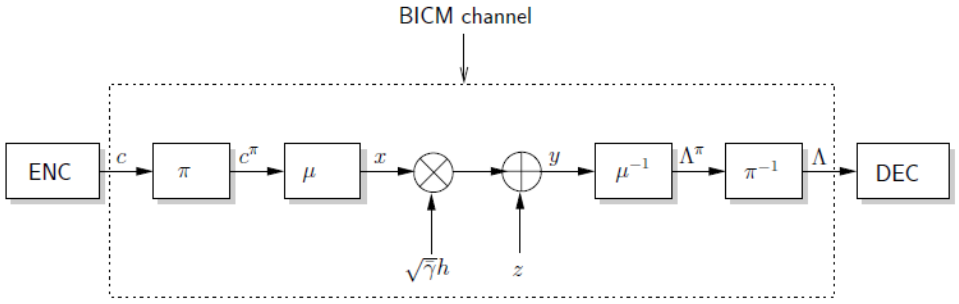

Figure 3.1: Block diagram of BICM transmission over a fading AWGN channel.

3.1 Preliminaries

Figure 3.1 shows the block diagram of equivalent BICM transmission system:
Transmitter: The BICM codeword $\boldsymbol{x} = [x_1, x_2, \cdots, x_L] \in \mathcal{C}$ comprises L complex

valued symbols and is obtained by first interleaving (π) the output of a binary encoder $\boldsymbol{c} = [c_1, c_2, \cdots, c_N]$ into $\boldsymbol{c^{\pi}} = [c_1^{\pi}, c_2^{\pi}, \cdots, c_N^{\pi}]$ and a subsequent mapping $\mu : \{0,1\}^r \to \mathcal{X}$ of each $r = \log_2(M)$ bits. While the pdf approximation that I will present next is applicable to arbitrary signal constellations, for practical relevance I assume that $\mathcal{X}$ is an M-ary QAM or PSK constellation with average unit symbol energy. As usual, the coding and mapping results in a uniform distribution signal points. When presenting numerical results next, I consider set

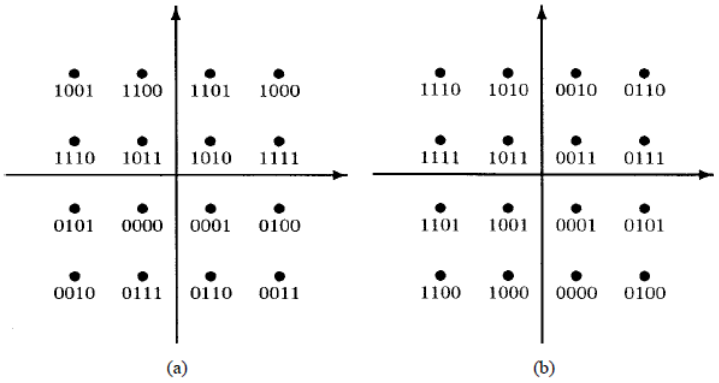

Figure 3.2: 16QAM signal set with (a) a suitable set partitioning labelling and (b) Gray Labelling.

partitioning labelling (SPL), semi set partitioning labelling (SSPL), modified set partitioning labelling (MSPL), and mixed labelling (ML), in addition to popular Gray labelling (GL) for QAM and PSK signal constellations. As Gray labelling plays a key role in BICM theory, I restate here its *definition* in term of BICM notation: let $\mathcal{X}$ denote a signal set of size $M = 2^r$, with minimum Euclidean distance d_{min}. A binary map $\mu : \{0,1\}^r \to \mathcal{X}$ is a Gray labelling for $\mathcal{X}$ if, for all $i = 1, \cdots, r$ and $b \in \{0,1\}$, each $\boldsymbol{x} \in \mathcal{X}_b^i$ has at most one $\boldsymbol{z} \in \mathcal{X}_{\bar{b}}^i$ at distance d_{min}. Figure 3.2 shows a labelled 16QAM signal set.

Channel: I consider BICM transmission over AWGN channels. The equivalent baseband discrete-time transmission model can be written as

$$y_i = \sqrt{\bar{\gamma}} h_i x_i + z_i, \tag{3.1}$$

where $y_i \in \mathbb{C}$ is the received sample, $h_i \in \mathbb{R}$ denotes the fading gain, $z_i \in \mathbb{C}$ is the additive noise sample at discrete-time i. The noise samples are independent

and identically distributed (i.i.d.) according to a zero-mean complex Gaussian distribution. I further assume that interleaving effectively renders the fading coefficients h_i i.i.d. random variables. Here $\bar{\gamma}$ represents the average SNR, so the instantaneous SNR is given by

$$\gamma_i = \bar{\gamma} h_i^2. \tag{3.2}$$

To make matters concrete I consider the Nakagami-m distribution to model multipath fading. Adjustment of the fading parameter renders this distribution very flexible. It includes Rayleigh fading ($m = 1$), nonfading AWGN ($m \to \infty$) and Rician fading $\left(m = \frac{(k+1)^2}{2k+1}\right)$ channels as special cases. The corresponding distribution of the SNR (3.2) is [5] ($\Gamma(\cdot)$ denotes the Gamma function)

$$f_{\gamma|\bar{\gamma},m}(\gamma) = \frac{m^m \gamma^{m-1}}{\bar{\gamma}^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\bar{\gamma}}\right) \tag{3.3}$$

Receiver: at the receiver, the demapper (μ^-1 in Figure 3.1) produces r bitwise reliability metrics Λ^π per symbol. The Λ^π are deinterleaved into Λ, which are then input to the decoder for binary code. The bit metric for the jth bit of the ith symbol has the form

$$\Lambda = -\min_{a \in \mathcal{X}_{j,1}} \left(\left\| y_i - \sqrt{\bar{\gamma}} h_i a \right\|^2 \right) + \min_{a \in \mathcal{X}_{j,0}} \left(\left\| y_i - \sqrt{\bar{\gamma}} h_i a \right\|^2 \right) \tag{3.4}$$

where $\mathcal{X}_{j,b}$ is the set of symbols with the jth bit in the binary label fixed to b. The (3.4) is the so-called max-log simplification of log-likelihood ratio (LLR), which is known to provide practically maximum-likelihood decoding performance [4].

3.2 Error-rate approximation using Union Bounding and Saddlepoint Approximation

The transmission channel between encoder output c and the decoder input Λ can be considered as an equivalent binary input output-symmetric (BIOS) channel (Figure 3.3), which is known as equivalent BICM channel. Assuming maximum likelihood decoding, the error-rate of linear codes transmitted over this kind of channels is well approximated by the union bound in the region above the cutoff

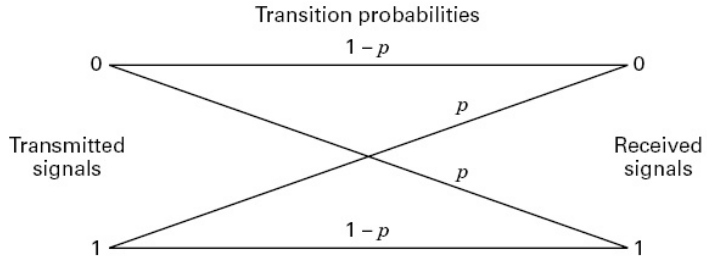

Figure 3.3: BIOS channel.

rate. As can be seen in [4], the BER union bound for a convolutional code of rate $R_c = k_c/n_c$ is given by

$$P_b \leq \frac{1}{k_c} \sum_{d_H = d_{H,min}}^{\infty} w_{d_H} \mathrm{PEP}\left(d_H \mid \bar{\gamma}, m\right) \tag{3.5}$$

where w_{d_H} denotes total input weight of error events at Hamming distance d_H, $d_{H,min}$ denotes the free distance of the convolutional code, and $\mathrm{PEP}\left(d_H \mid \bar{\gamma}, m\right)$ is the pairwise error probability corresponding to an error event with Hamming weight d_H. For BIOS channels, the PEP can be considered as the tail probability of a random variable generated by summing d_H i.i.d LLRs $\Lambda_1, \cdots, \Lambda_{d_H}$. Choosing all-one codeword as reference codeword, I get

$$\mathrm{PEP}\left(d_H \mid \bar{\gamma}, m\right) = \mathrm{P}\left(\Delta_{d_H} = \sum_{i=1}^{d_H} \Lambda_i < 0 \mid \bar{\gamma}, m\right). \tag{3.6}$$

For computing such a probability, a common way is through the use of the Laplace transform $\Phi_{\Lambda|\bar{\gamma},m}\left(s\right)$ of the pdf of Λ. That is

$$\mathrm{P}\left(\Delta_{d_H} < 0 \mid \bar{\gamma}, m\right) = \frac{1}{2\pi j} \int_{\alpha-\infty}^{\alpha+\infty} \left[\Phi_{\Lambda|\bar{\gamma},m}\left(s\right)\right]^{d_H} \frac{\mathrm{d}s}{s}, \tag{3.7}$$

where j is the imaginary unit and $\alpha \in \mathbb{R}$, $0 < \alpha < \alpha_{max}$, is chosen in the region of convergence of integral. The computation of this is often not straightforward and invokes the use of numerical methods. For this reason in [6] are proposed a few bounds and estimations; one of the most important for its simply form and accuracy is the saddlepoint approximations, that get an approximation for the previous result:

$$\mathrm{P}\left(\Delta_{d_H} < 0 \mid \bar{\gamma}, m\right) \approx \frac{\left(\Phi_{\Lambda|\bar{\gamma},m}\left(\hat{s}\right)\right)^{d_H+0.5}}{\hat{s}\sqrt{2\pi d_H \Phi''_{\Lambda|\bar{\gamma},m}\left(\hat{s}\right)}}, \tag{3.8}$$

where $\Phi''_{\Lambda|\bar{\gamma},m}(s)$ denotes the second-order derivative of $\Phi_{\Lambda|\bar{\gamma},m}(s)$ and $\hat{s} \in \mathbb{R}^+$, $0 < \hat{s} < \alpha_{max}$, is the saddlepoint defined through the first-order derivative as

$$\Phi'_{\Lambda|\bar{\gamma},m}(\hat{s}) = 0. \tag{3.9}$$

Now I first derive an approximation for pdf of the log-likelihood ratios defined in (3.4), that will be useful to derive the closed-form expressions for $\Phi_{\Lambda|\bar{\gamma},m}(\hat{s})$ for $s \in \mathbb{R}^+$ and arbitrary m. Since the channel is BIOS, the symmetry property

$$f_{\Lambda|c=b,\gamma}(\lambda) = f_{\Lambda|c=\bar{b},\gamma}(-\lambda) \tag{3.10}$$

holds. Here $f_{\Lambda|c=b,\gamma}(\lambda)$ is the pdf when $c = b$ is transmitted and $\bar{b}$ is the complement of b. I consider the transmission of $c = 1$, and I value the pdf as weight sum of pdf conditioned on the bit position $1 \leq j \leq r$, and I find this expression

$$f_{\Lambda|c=1,\gamma}(\lambda) = \frac{2}{rM} \sum_{j=1}^{r} \sum_{x \in \mathcal{X}_{j,1}} f_{\Lambda|j,x,\gamma}(\lambda). \tag{3.11}$$

where $f_{\Lambda|j,x,\gamma}(\lambda)$ is a function depending to three parameters, because the condition '$c = 1$' can be represented also by two condition on 'j, x'. In fact c is the output binary to transmit and j, x are the parameters of the codeword. Now the task is to find expressions for $f_{\Lambda|j,x,\gamma}(\lambda)$. For this I have to consider all signal points in the constellation. As in [7] a simply way is to consider the set of all nearest signal points in $\mathcal{X}_{j,\bar{b}}$ for a given $x \in \mathcal{X}_{j,b}$. So I define a set of nearest competitive signal point of x to approximate $f_{\Lambda|j,x,\gamma}(\lambda)$. This corresponds to the approximation

$$\Lambda \approx -\left(\left\|y_i - \sqrt{\bar{\gamma}} h_i x_i\right\|^2\right) + \min_{a \in \mathcal{X}_{j,x_i}} \left(\left\|y_i - \sqrt{\bar{\gamma}} h_i a\right\|^2\right), \tag{3.12}$$

for the log-likelihood ratio in (3.4), that is accurate in the SNR range in which the minimum Euclidean distance events dominate and the BER union bound converges to the true error rate. To this goal I quote here the six non-equivalent formations for the set of nearest competitive signal points $\mathcal{A}_{j,x}$ for QAM and PSK constellations (Figure 3.4). Starting from here I will determine the pdf $f_{\Lambda|j,x,\gamma}(\lambda)$ and in next sections I will consider all-cases closed-form. The pdf is expressed by

$$f_{\Lambda|j,x,\gamma}(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda} \int_{\mathbb{D}(\lambda|j,x,\gamma)} \frac{1}{\pi} \exp\left(-\|z\|^2\right) \mathrm{d}z \tag{3.13}$$

where $\mathbb{D}(\lambda \mid j, x, \gamma)$ is the part of complex plane in which the log-likelihood ratio is less than λ (in Figure 3.4 $\lambda = 0$)

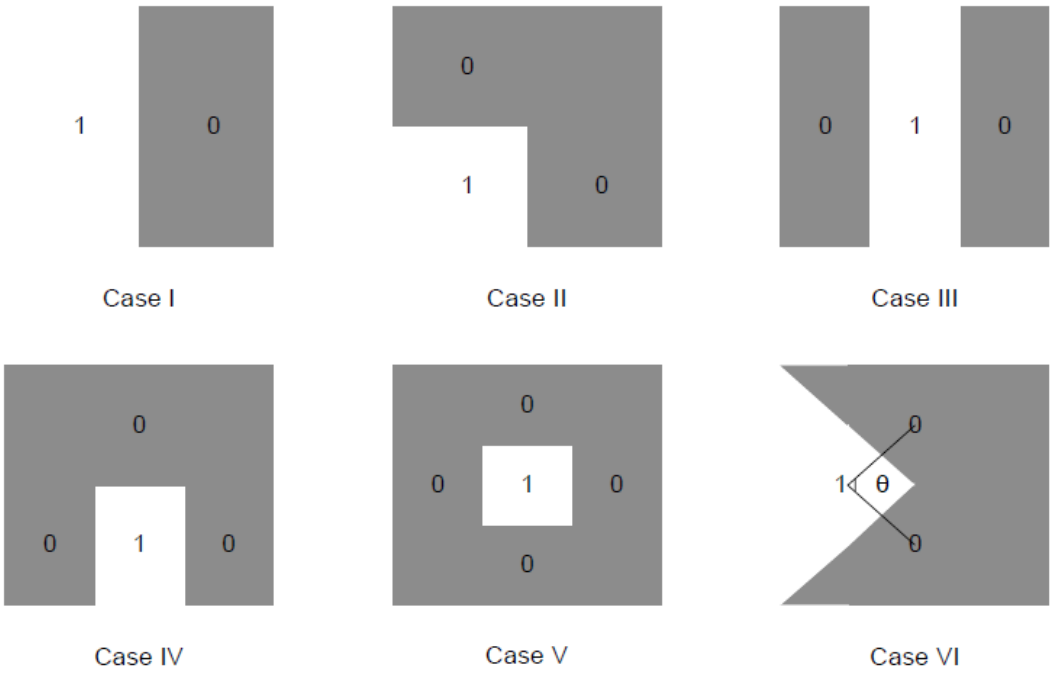

Figure 3.4: Illustration for possible sets of nearest competitive signal points $\mathcal{A}_{j,x}$ for general QAM and PSK constellations. The shaded areas indicates $\mathbb{D}(\lambda \mid j, x, \gamma)$ for $\lambda = 0$. For $\lambda > 0$ the boundaries move toward x, for $\lambda < 0$ the boundaries move towards the competitive signal points.

3.3 Closed-form expressions for results

From the expressions founded in the previous section, I consider here the most important result for the analysis: the closed-form solution for the kth configuration from Figure 3.4. In Table 3.1 are specified the $f_{\lambda,k|\cdot,\gamma}(\lambda)$,that are the probability density function of log-likelihood ratios, and the Laplace transforms over nonfading AWGN channel and over Nakagami-m fading channels. The expressions in Table 3.1 are obtained in [7] and the notations used are here reported: $\mathbb{N}_{\mu,\sigma^2}$ is the Gaussian normal distribution with mean μ and variance σ^2; $\operatorname{erf}(x)$ is the Gauss error function; $u(x)$ is the unit step function. Then substituting $f_{\Lambda|j,x,\gamma}(\lambda)$ in (3.11) with the corresponding functions from Table 3.1, I can obtain the desired closed-form expression for $f_{\Lambda|c=1,\gamma}(\lambda)$. In [7] there are all parameters of the most important constellations and for different labellings (as GL, SPL, MSPL, ML, SSPL). In Figure 3.5 are reported some results.

After this, other important expressions can be obtained substituting the approximations in Laplace transform, which become in closed-form. Using what I find in the previous calculation I can analyse the expressions of Table 3.1 like this:

1) Nonfading Channel: in this case $m \to \infty$ and $\gamma = \bar{\gamma}$. Using the expressions for $f_{\Lambda,k|\cdot,\gamma}(\lambda)$ from Table 3.1, the $\Phi_{\Lambda,k|\cdot,\gamma}(s)$ can be written as a weighted sum of the the integrals [7] $I_{1|\mu}(s)$ and $I_{2|\mu,\nu}(s)$ that in closed-form are:

$$I_{1|\mu}(s) = -\frac{1}{4} e^{\mu\left(s^2-s\right)} \left(1 + \operatorname{erf}\left(\sqrt{\frac{\mu}{2}} s\right)\right)^2 \tag{3.14}$$

$$I_{2|\mu,\nu}(s) = -\frac{1}{4} e^{\mu\left(s^2-s\right)} \operatorname{erf}\left(\frac{2\nu}{\sqrt{1+(2\nu)^2}} \sqrt{\mu} s\right) \tag{3.15}$$

where $\mu = d^2\gamma$ and $\nu = \tan(\theta/2)$.

2) Nakagami-m Fading Channel: in this case $\Phi_{\Lambda,k|\cdot,\bar{\gamma},m}(s) = \int_0^\infty f_{\gamma|\bar{\gamma},m}(s)\, \Phi_{\Lambda,k|\cdot,\gamma}(s)\, \mathrm{d}\gamma$ for which expressions are given in Table 3.1 in terms of integrals [7] $I_{3|\mu,\nu}(s)$ and $I_{4|\mu,\nu}(s)$ that in closed-form are:

$$I_{3|\mu,\nu}(s) \approx \sum_{i=1}^{3} a_i \left(\frac{m}{m + \left[\mu\left(s - s^2\right) + b_i\left(\nu s\right)^2\right]\bar{\gamma}}\right)^m \tag{3.16}$$

where

$$[a_1, a_2, a_3] = \left[1, -\frac{1}{6}, -\frac{1}{2}\right]$$

Case 1	$f_{\Lambda,1\|d,\gamma}(\lambda)$	$\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)$
	$\Phi_{\Lambda,1\|d,\gamma}(s)$	$\exp\left(d^2\gamma\left(s^2-s\right)\right)$
	$\Phi_{\Lambda,1\|d,\bar{\gamma},m}(s)$	$\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m$
Case 2	$f_{\Lambda,2\|d,\gamma}(\lambda)$	$\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)\left(1-\mathrm{erf}\left(\frac{\lambda-d^2\gamma}{2d\sqrt{\gamma}}\right)\right)$
	$\Phi_{\Lambda,2\|d,\gamma}(s)$	$\exp\left(d^2\gamma\left(s^2-s\right)\right)$ $\times\left(1+\mathrm{erf}\left(d\sqrt{\frac{\gamma}{2}}\,s\right)\right)$
	$\Phi_{\Lambda,2\|d,\bar{\gamma},m}(s)$	$\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m+I_{3\|d^2,d/\sqrt{2}}(s)$
Case 3	$f_{\Lambda,3\|d,\gamma}(\lambda)$	$2\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)u\left(d^2\gamma-\lambda\right)$
	$\Phi_{\Lambda,3\|d,\gamma}(s)$	$\exp\left(d^2\gamma\left(s^2-s\right)\right)\left(1+\mathrm{erf}\left(d\sqrt{\gamma}\,s\right)\right)$
	$\Phi_{\Lambda,3\|d,\bar{\gamma},m}(s)$	$\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m+I_{3\|d^2,d}(s)$
Case 4	$f_{\Lambda,4\|d,\gamma}(\lambda)$	$\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)\left(1-2\mathrm{erf}\left(\frac{\lambda-d^2\gamma}{2d\sqrt{\gamma}}\right)\right)$ $\times u\left(d^2\gamma-\lambda\right)$
	$\Phi_{\Lambda,4\|d,\gamma}(s)$	$\frac{1}{2}\exp\left(d^2\gamma\left(s^2-s\right)\right)$ $\times\Big[1+\mathrm{erf}\left(d\sqrt{\gamma}\,s\right)$ $+\left(1+\mathrm{erf}\left(d\sqrt{\frac{\gamma}{2}}\,s\right)\right)^2\Big]$
	$\Phi_{\Lambda,4\|d,\bar{\gamma},m}(s)$	$\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m+\frac{1}{2}I_{3\|d^2,d}(s)$ $+I_{3\|d^2,d/\sqrt{2}}(s)+\frac{1}{2}I_{4\|d^2,d/\sqrt{2}}(s)$
Case 5	$f_{\Lambda,5\|d,\gamma}(\lambda)$	$-4\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)$ $\times\,\mathrm{erf}\left(\frac{\lambda-d^2\gamma}{2d\sqrt{\gamma}}\right)u\left(d^2\gamma-\lambda\right)$
	$\Phi_{\Lambda,5\|d,\gamma}(s)$	$\exp\left(d^2\gamma\left(s^2-s\right)\right)$ $\times\left(1+\mathrm{erf}\left(d\sqrt{\frac{\gamma}{2}}\,s\right)\right)^2$
	$\Phi_{\Lambda,5\|d,\bar{\gamma},m}(s)$	$\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m+2I_{3\|d^2,d/\sqrt{2}}(s)$ $+I_{4\|d^2,d/\sqrt{2}}(s)$
Case 6	$f_{\Lambda,6\|d,\theta,\gamma}(\lambda)$	$\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)$ $\times\left(1-\mathrm{erf}\left(\tan\left(\frac{\theta}{2}\right)\frac{\lambda-d^2\gamma}{2d\sqrt{\gamma}}\right)\right)$
	$\Phi_{\Lambda,6\|d,\theta,\gamma}(s)$	$\exp\left(d^2\gamma\left(s^2-s\right)\right)$ $\times\left(1+\mathrm{erf}\left(\sin\left(\frac{\theta}{2}\right)d\sqrt{\gamma}\,s\right)\right)$
	$\Phi_{\Lambda,6\|d,\theta,\bar{\gamma},m}(s)$	$\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m$ $+I_{3\|d^2,\sin(\theta/2)d}(s)$

Table 3.1: pdf of log-likelihood ratios for transmission over nonfading AWGN channel, and its Laplace transform over nonfading AWGN channel and over Nakagami-m channel, for the six different set of Figure 3.4.

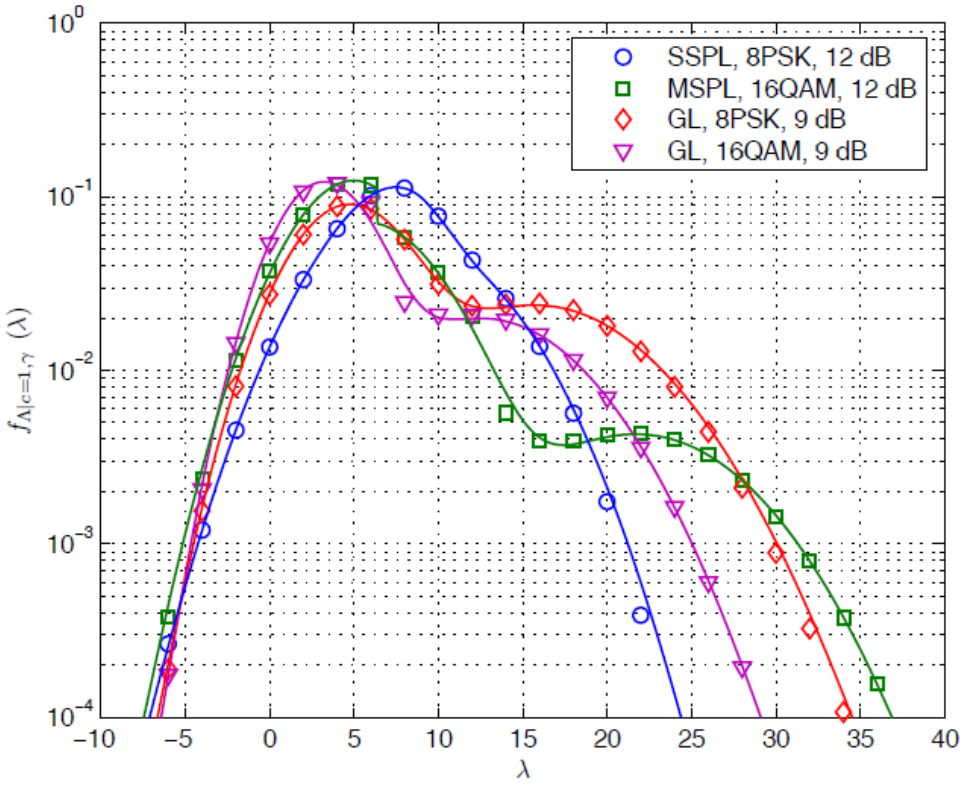

Figure 3.5: pdf of reliability metrics for BICM transmission over the nonfading AWGN channel for different constellations and labelling. Lines are pdf approximations founded while markers represent the estimated histograms through simulative measurement.

and

$$[b_1, b_2, b_3] = \left[0, 1, \frac{4}{3}\right].$$

$$I_{4|\mu,\nu}(s) \approx \sum_{i=1}^{6} a_i \left(\frac{m}{m + \left[\mu\left(s - s^2\right) + b_i\left(\nu s\right)^2\right]\bar{\gamma}}\right)^m \tag{3.17}$$

where

$$[a_1, a_2, a_3, a_4, a_5, a_6] = \left[1, -\frac{1}{3}, -1, \frac{1}{36}, \frac{1}{6}, \frac{1}{4}\right]$$

and

$$[b_1, b_2, b_3, b_4, b_5, b_6] = \left[0, 1, \frac{4}{3}, 2, \frac{7}{3}, \frac{8}{3}\right].$$

Here, $\mu = d^2$ and $\nu \in \left/\sqrt{2}, d, d\sin\left(\theta/2\right)\right\}$. Using what I have find, I can evaluate the BER that is an important performance indicator for BICM transmission. To do this I consider the case of asymptotically high SNR to further simplify the expressions for the pdf of log-likelihood ratios and its Laplace transform. These results are reported in Table 3.2. From the simplified expressions, I consider the case of asymptotic high SNR to further simplify the expressions for the pdf of log-likelihood ratios and its Laplace transform, and then I can make a direct evaluation of the PEP for the 2 cases. By these results I can obtain the BER in the common way.

1) Nonfading Channel: for transmission over nonfading AWGN channel the approximation is

$$\mathrm{PEP}\left(d_H \mid \gamma\right) = \frac{N_1^{d_H}}{d_1\sqrt{\pi d_H \gamma}} \exp\left(-\frac{1}{4} d_H d_1^2 \gamma\right). \tag{3.18}$$

2) Nakagami-m Fading Channel: for this case the approximation gives

$$\mathrm{PEP}\left(d_H \mid \bar{\gamma}, m\right) = \frac{\Gamma\left(md_H + 1/2\right)}{2\sqrt{\pi}\Gamma\left(md_H + 1\right)} \left(\sum_{l=1}^{l_{max}} \frac{N_l}{d_l^{2m}}\right)^{d_H} \left(\frac{4m}{\bar{\gamma}}\right)^{md_H}. \tag{3.19}$$

Here I make the next assumptions [7]:

$$l_{max} = \begin{cases} q_{max}, & for \quad QAM \\ M/2, & for \quad PSK \end{cases}$$

$$N_l = \begin{cases} \frac{2\left(n_{1,l} + 2n_{2,l} + 2n_{3,l} + 3n4,l + 4n_{5,l}\right)}{rM}, & for \quad QAM \\ \frac{2}{rM}\left(n_{1,l} + 2n_{6,l}\right), & for \quad PSK \end{cases}$$

Case 1	$f^{\mathrm{a}}_{\Lambda,1\vert d,\gamma}(\lambda)$	$\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)$
	$\Phi^{\mathrm{a}}_{\Lambda,1\vert d,\gamma}(s)$	$\exp\left(d^2\gamma\left(s^2-s\right)\right)$
	$\Phi^{\mathrm{a}}_{\Lambda,1\vert d,\bar{\gamma},m}(s)$	$\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m$
Case 2	$f^{\mathrm{a}}_{\Lambda,2\vert d,\gamma}(\lambda)$	$2\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)$
	$\Phi^{\mathrm{a}}_{\Lambda,2\vert d,\gamma}(s)$	$2\exp\left(d^2\gamma\left(s^2-s\right)\right)$
	$\Phi^{\mathrm{a}}_{\Lambda,2\vert d,\bar{\gamma},m}(s)$	$2\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m$
Case 3	$f^{\mathrm{a}}_{\Lambda,3\vert d,\gamma}(\lambda)$	$2\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)$
	$\Phi^{\mathrm{a}}_{\Lambda,3\vert d,\gamma}(s)$	$2\exp\left(d^2\gamma\left(s^2-s\right)\right)$
	$\Phi^{\mathrm{a}}_{\Lambda,3\vert d,\bar{\gamma},m}(s)$	$2\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m$
Case 4	$f^{\mathrm{a}}_{\Lambda,4\vert d,\gamma}(\lambda)$	$3\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)$
	$\Phi^{\mathrm{a}}_{\Lambda,4\vert d,\gamma}(s)$	$3\exp\left(d^2\gamma\left(s^2-s\right)\right)$
	$\Phi^{\mathrm{a}}_{\Lambda,4\vert d,\bar{\gamma},m}(s)$	$3\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m$
Case 5	$f^{\mathrm{a}}_{\Lambda,5\vert d,\gamma}(\lambda;d,\gamma)$	$4\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)$
	$\Phi^{\mathrm{a}}_{\Lambda,5\vert d,\gamma}(s)$	$4\exp\left(d^2\gamma\left(s^2-s\right)\right)$
	$\Phi^{\mathrm{a}}_{\Lambda,5\vert d,\bar{\gamma},m}(s)$	$4\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m$
Case 6	$f^{\mathrm{a}}_{\Lambda,6\vert d,\gamma}(\lambda)$	$2\mathcal{N}_{d^2\gamma,2d^2\gamma}(\lambda)$
	$\Phi^{\mathrm{a}}_{\Lambda,6\vert d,\gamma}(s)$	$2\exp\left(d^2\gamma\left(s^2-s\right)\right)$
	$\Phi^{\mathrm{a}}_{\Lambda,6\vert d,\bar{\gamma},m}(s)$	$2\left(\frac{m}{m-d^2\bar{\gamma}\left(s^2-s\right)}\right)^m$

Table 3.2: asymptotic values for pdf of log-likelihood ratios for transmission over nonfading AWGN channel, and its Laplace transform over nonfading AWGN channel and over Nakagami-m channel, for the six different set of Figure 3.4.

3.4 Numerical results and discussions

Now I compare simulated and analytical BER results as function of the bitwise SNR indicates with $\bar{\gamma}_b$ as in [7]. Figure 3.6 shows analytical (lines) and simulated (markers) BER results for different constellations and labelling for transmission over the nonfading AWGN channel. Solid lines represent the BER union bound, while dashed lines represent the asymptotic approximation for $d_H = d_{H,min}$. The BER union bound is fairly tight for all modulation schemes and BERs below about 10^{-4}. Likewise, the proposed simple expression (3.18) accurately predicts the asymptotic error-rate performance at high SNR.

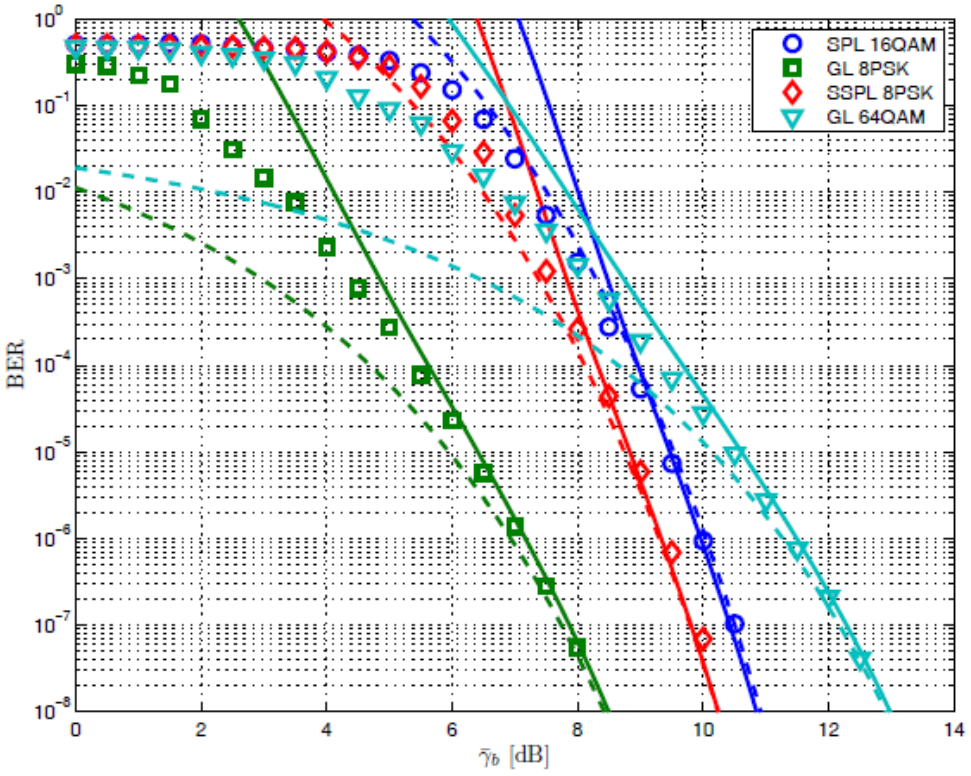

Figure 3.6: BER of BICM transmission over nonfading channel for a 64-state convolutional code of rate 1/2. Solid lines: BER union bound. Dashed lines: asymptotic analysis. Markers: simulation results.

I then compare analytical and simulated BER results for BICM transmission over fading channels with different constellations and labelling rules. Figure 3.7 shows BER curves obtained from the BER union bound and exact closed-form solutions for integrals $I_{3|\mu,\nu}(s)$ and $I_{4|\mu,\nu}(s)$ (solid lines) and their approximations

(3.16) and (3.17) (dashed lines). Again there is an excellent match between results from analysis and simulations, which confirm the validity of closed-form BER approximations. Finally in Figure 3.8 the asymptotic BER results from

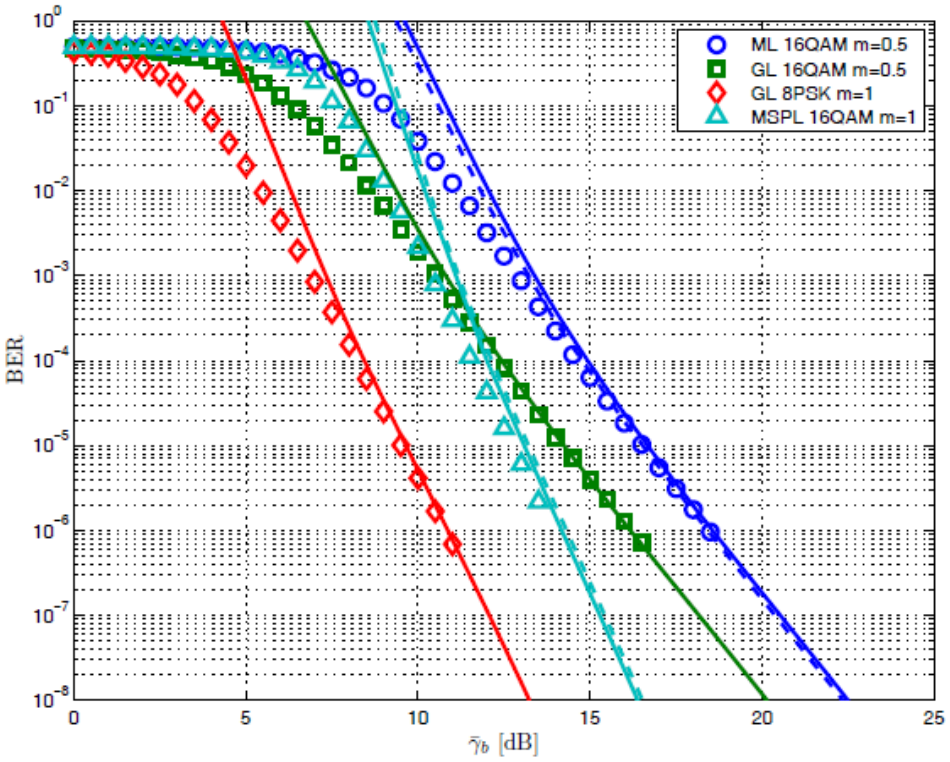

Figure 3.7: BER of BICM transmission over Nakagami-m fading channel for 64-state convolutional code of rate 1/2. Solid lines: BER union bound using exact closed-form solution for integrals. Dashed lines: BER union bound using the approximations. Markers: simulation results.

(3.19) and $d_H = d_{H,min}$ (solid lines) are plotted together with the BER union bound (markers) for the same transmission scenarios as Figure 3.7. It can be seen that asymptotic results correctly predict coding and fading gain of BICM scheme. Hence I conclude that the simple expressions founded are very valuable to quickly determine the asymptotic performance of BICM transmission over fading channels.

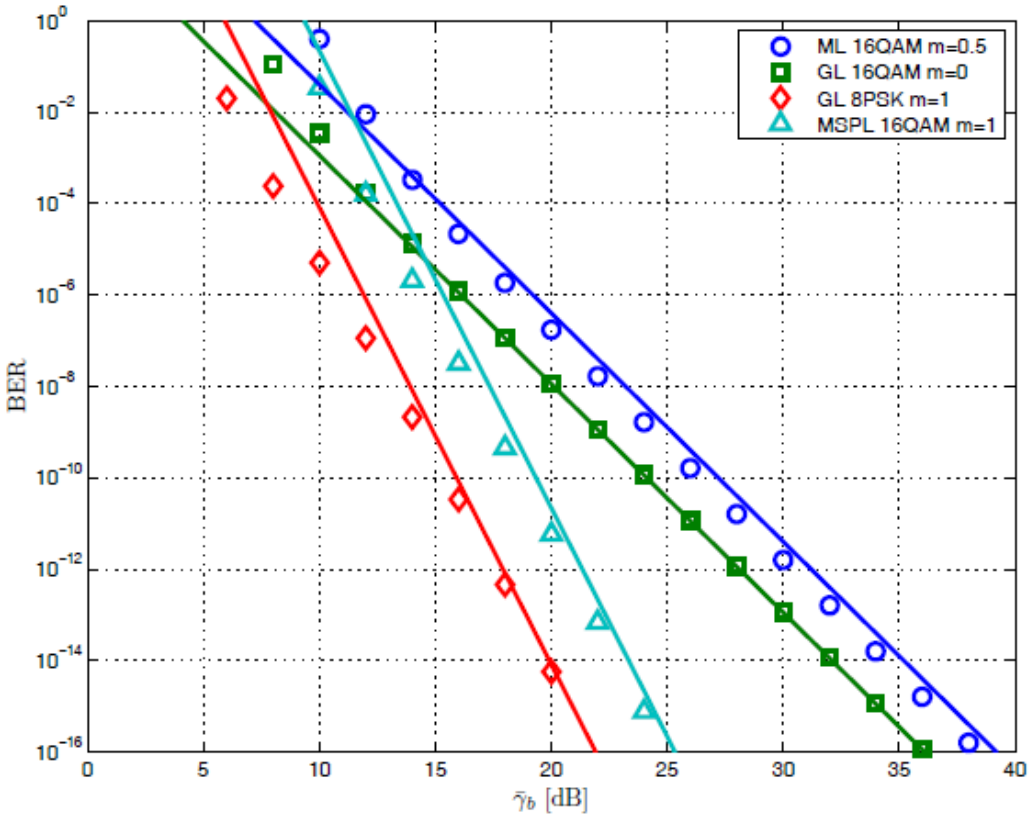

Figure 3.8: BER of BICM transmission over Nakagami-m fading channel for 64-state convolutional code of rate 1/2. Solid lines: asymptotic analysis with (3.19). Markers: BER union bound.

Chapter 4

Introduction of iterative decoding

In this chapter, I study BICM with iterative decoding (BICM-ID). Inspired by the advent of turbo-codes and iterative decoding, BICM-ID was proposed in order to improve the performance of BICM by exchanging messages between the demodulator and the decoder of the underlying binary code. As illustrated in Figure 4.1 and shown in [8], [9], BICM-ID can provide some performance advantages using convolutional codes combined with mapping different from Gray. With this kind of system with 8-state convolutional code of rate-2/3, and 8-PSK modulation, using hard-decision feedback, achieves an improvement over the conventional BICM scheme that exceeds 1 dB for a fully-interleaved Rayleigh flat-fading channel and that exceeds 1.5 dB for a channel with additive white Gaussian noise. This robust performance makes BICM with iterative decoding suitable for both types of channels [8]. In this analysis I will use only a system with this characteristics (8-state convolutional code of rate-2/3, and 8-PSK modulation); the extension to other coding rates and modulation types is straightforward.

4.1 System description

The system block is shown in Figure 4.2. Note the addition of a feedback loop compared with the conventional decoder. I represent the output of the encoder by

$$\boldsymbol{C} = \left[c_0^1, c_0^2, c_0^3, \cdots, c_t^1, c_t^2, c_t^3, \cdots\right] = [C_0, \cdots, C_t, \cdots] \tag{4.1}$$

where c_t^i is the ith output bit at time position t and $C_t = [c_t^1, c_t^2, c_t^3]$. Three independent bit interleavers permute bits to break the correlation of fading channels

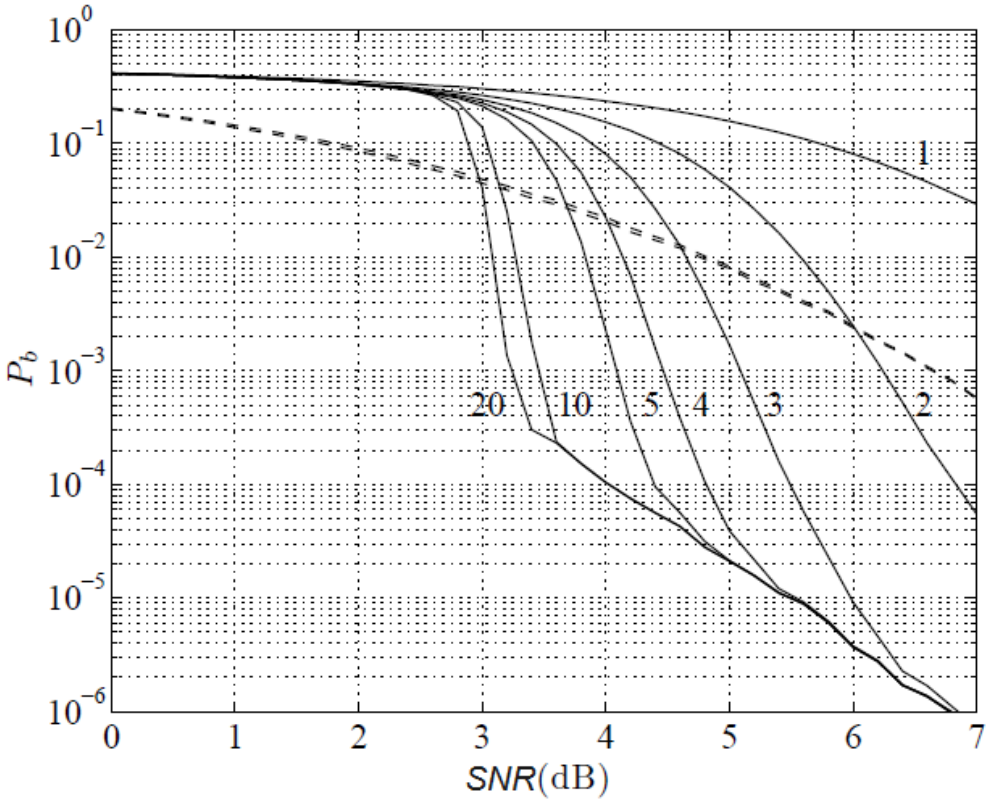

Figure 4.1: BER for BICM-ID in the AWGN channel with the $(5, 7)_8$ convolutional code, 16-QAM with set partitioning mapping. Curve labels denote the iteration number. Upper curve corresponds to iteration 1, decreasing down to iteration 20. In dashed lines, results for Gray mapping.

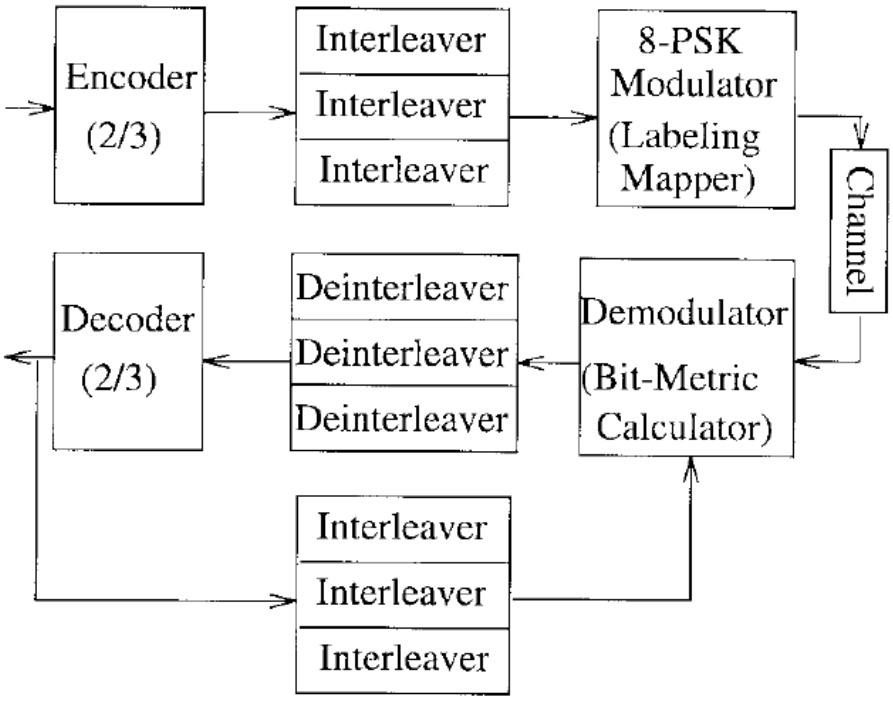

Figure 4.2: Block diagram of the BICM-ID scheme.

as well as the correlation between the bits in the same symbol. At the deinterleavers, the permutation is inverted. The output of the interleavers is represented by

$$\boldsymbol{V} = \left[v_0^1, v_0^2, v_0^3, \cdots, v_t^1, v_t^2, v_t^3, \cdots\right] = \left[V_0, \cdots, V_t, \cdots\right] \tag{4.2}$$

where v_t^i is the ith output bit at time position t and $V_t = [v_t^1, v_t^2, v_t^3]$. The interleaving is followed by a signal labelling map μ, an isomorphism between a 3-tuple $V_t = [v_t^1, v_t^2, v_t^3]$ and a signal constellation point x_t

$$x_t = \mu\left(V_t\right), x_t \in \mathcal{X} \tag{4.3}$$

where $\mathcal{X}$ is the signal set of 8-PSK. The Rayleigh-fading channel with coherent detection give this output

$$y_t = \rho_t x_t + z_t \tag{4.4}$$

as seen in the previous chapter.

4.2 Conventional decoding

For each received signal y_t, a log-likelihood function is calculated for the two possible binary values of each coded bit

$$\begin{aligned} \lambda\left(v_t^i = b\right) &= \log \sum_{x \in \mathcal{X}(i,b)} P\left(x \mid y_t, \rho_t\right) \\ &= \log \left[\sum_{x \in \mathcal{X}(i,b)} P\left(y_t \mid x, \rho_t\right) P\left(x\right) \right], \\ & i = 1,2,3; b = 0,1 \end{aligned} \tag{4.5}$$

where the signal subset $\mathcal{X}\left(i,b\right) = \left\{\mu\left(v^1, v^2, v^3\right) \mid v^i = b\right\}$ and the terms common to all i and b are disregarded. In conventional decoding, the *a priori* probability $P\left(x\right)$ is assumed equal for any $x \in \mathcal{X}\left(i,b\right)$. Then, the bit metric becomes

$$\begin{aligned} \lambda\left(v_t^i = b\right) &= \log \sum_{x \in \mathcal{X}(i,b)} P\left(y_t \mid x, \rho_t\right) \\ &\approx \max_{x \in \mathcal{X}(i,b)} \log P\left(y_t \mid x, \rho_t\right) \\ &= - \min_{x \in \mathcal{X}(i,b)} \left\| y_t - \rho_t x \right\|^2 \end{aligned} \tag{4.6}$$

where a constant scalar is disregarded and the approximation is good at high SNR. At the decoder (as Viterbi decoder), the branch metric corresponding to each of the eight possible 3-tuple is the sum of the corresponding bit metrics after deinterleaving.

4.3 Iterative decoding with hard-decision feedback

Convolutional encoding introduces redundancy and memory into the signal sequence. Yet, the equally likely assumption made before fails to use this information, primarily because it is difficult to specify in advance of any decoding. The *a priori* information is reflected in the decoding results and therefore can be included through iterative decoding. One approach is to use the Viterbi algorithm (see Appendix A) with soft outputs, although this is computationally complex. Instead, I consider only binary-decision feedback for the calculation of the bit

metrics in the second round of decoding. For example, to calculate $\lambda\left(v_t^1=0\right)$ I assume, for any $x=\mu\left(v^1=0,v^2,v^3\right)\in\mathcal{X}\left(1,0\right)$,

$$P\left(x\right)=\begin{cases}1, & if \quad v^1=0,v^2=\hat{v}_t^2,v^3=\hat{v}_t^3\\ 0, & otherwise\end{cases} \tag{4.7}$$

where $\hat{v}_t^2$ and $\hat{v}_t^3$ are the first-round decoding decisions. Then the bit metric with the decision feedback becomes

$$\lambda\left(v_t^1=0\right)=-\left\|y_t-\rho_t\mu\left(0,\hat{v}_t^2,\hat{v}_t^3\right)\right\|^2. \tag{4.8}$$

The bit metrics for other bit positions and bit values follow similarly. Given the feedback $\hat{v}_t^2$ and $\hat{v}_t^3$ the Euclidean distance between the signals $\mu\left(0,\hat{v}_t^2,\hat{v}_t^3\right)$ and $\mu\left(1,\hat{v}_t^2,\hat{v}_t^3\right)$ can be significantly larger than the minimum Euclidean distance between the signals in the subset $\mathcal{X}\left(1,0\right)$ and those $\mathcal{X}\left(1,1\right)$. With an appropriate signal labelling, the minimum Euclidean distance between coded sequences can be made large for BICM-ID. This is the key that BICM-ID outperforms conventional BICM. To avoid severe error propagation, the bits feedback should be independent of the bit for which the bit metric is calculated. This is made possible by independent bit interleavers - the three bits making up a channel symbol are typically far apart in the coded sequence. This is clearly a feature not available in a symbol-interleaved system.

4.4 Signal labelling

The performance of BICM-ID strongly depends on the signal labelling methods. It can be seen that a mixed labelling method outperforms both Gray labelling and set-partitioning labelling. With mixed labelling, the eight sequential labels for 8-PSK signals are $\{000,001,010,011,110,111,100,101\}$. The details of the labelling design are discussed in [9]. Here I report only the results.

4.5 Simulation results

Now I show the BER performance of 8-PSK BICM-ID for both Rayleigh fading and AWGN channels. Also other type of scheme are included for comparison. The performance for Rayleigh fading channels is shown in Figure 4.3.

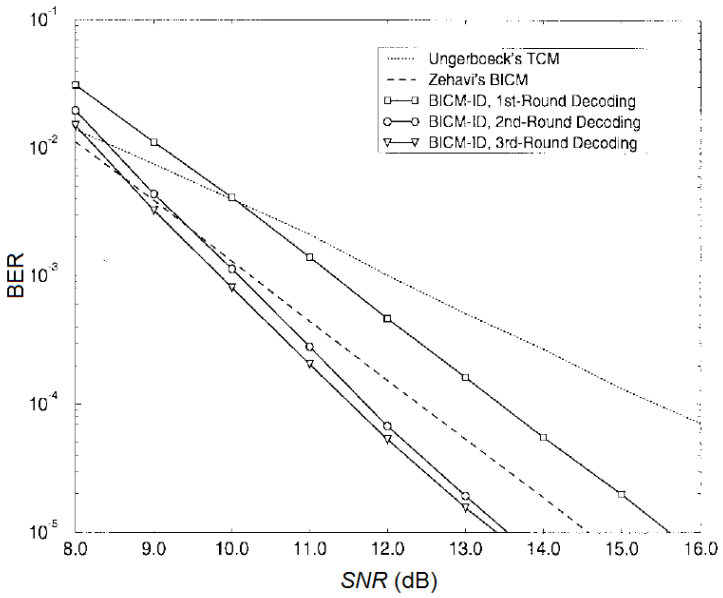

Figure 4.3: The performance of BICM-ID with the 8-state convolutional code of rate-2/3 and 8-PSK modulation for Rayleigh fading channels.

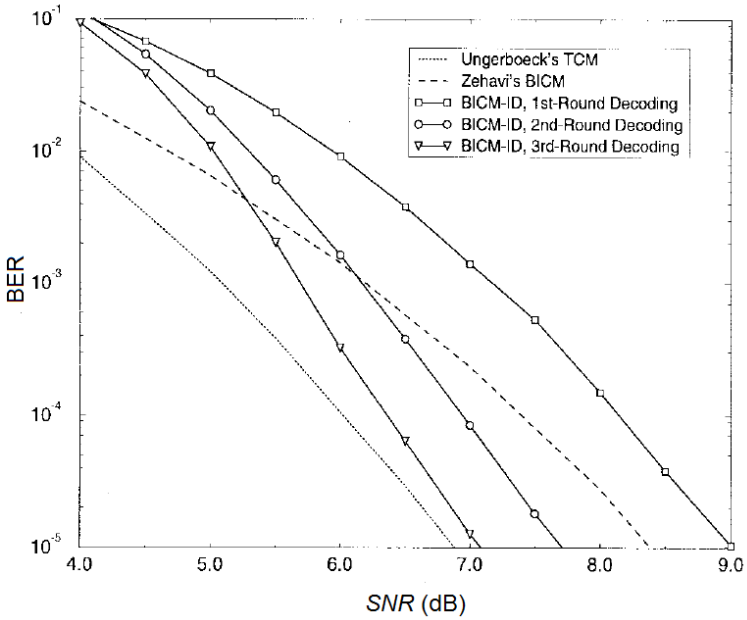

Figure 4.4: The performance of BICM-ID with the 8-state convolutional code of rate-2/3 and 8-PSK modulation for AWGN channels.

Compared with BICM scheme proposed in Chapter 2, there is a 1 dB performance degradation after the first round of decoding (without decision feedback) of BICM-ID due to mixed labelling. However, with a second round of decoding, BICM-ID quickly catches up and outperforms BICM by 1 dB. A third round of decoding adds a slight improvement. The performance for AWGN channels is shown in Figure 4.4. The SNR gap between BICM-ID and TCM scheme is only 0.2 dB. The gain of BICM-ID over BICM is more than 1.5 dB.

Chapter 5

Applications

In this chapter I briefly discuss some applications of BICM to cases of practical relevance not explicitly included in my presentation. In particular, I review current work and outline how to extend the results I presented throughout the thesis to non-coherent detection, block-fading and multiple-input multiple-output (MIMO) channels.

5.1 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) [11]. BICM with orthogonal modulation was first studied in [4]: here the study included BICM capacity, cutoff rate as well as error-probability considerations. Then in [8] are applied BICM-ID techniques to orthogonal modulation with non-coherent detection with turbo-codes. In [12], the design of capacity approaching codes is considered using the improved construction based on repeat-accumulate (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, $x \in \mathcal{X} = \{e_1, \cdots, e_M\}$, where $e_k = (\underbrace{0, \cdots, 0}_{k-1}, 1, \underbrace{0, \cdots, 0}_{M-k-1})$ 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 (3.1), where now $y, z \in \mathbb{C}^M$, they are vectors of dimension M. The channel transition probability becomes

$$P_\theta(y \mid x, h) = \frac{1}{\pi^M} e^{-\|y - \sqrt{\bar{\gamma}} x h\|^2}. \tag{5.1}$$

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 $q(x = e_k, y) \propto P_\theta(y \mid x, h)$. Here $q(x, y)$ is a function that represents $\sum_k \lambda^i(\boldsymbol{y}_{k'}, c_k)$ of (2.7). 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 [11]

$$q(x = e_k, y) \propto I_0\left(2\sqrt{\bar{\gamma}}|h|y_k\right) \tag{5.2}$$

where $I_0(\cdot)$ is the zero-th order Bessel function of the first kind and with some abuse of notation let y_k denote the k-th entry of the received signal vector $\boldsymbol{y}$. The bit metrics are

$$q_j(b_j(x) = b, \boldsymbol{y}) = \sum_{k' \in \mathcal{X}_b^j} I_0\left(2\sqrt{\bar{\gamma}}|h|y'_k\right) \tag{5.3}$$

where now the set $\mathcal{X}_b^j$ is the set of indices from $\{0, \cdots, M-1\}$ with bit b in the j-th position. As an example, Figure 5.1 and Figure 5.2 show the coded modulation and BICM capacities for the AWGN channel and the fully-interleaved Rayleigh fading channel with non-coherent detection, respectively. Observing, the capacity loss of BICM with respect to coded modulation is somewhat larger than in the QAM/PSK modulation case with coherent-detection [4].

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 implementation in this case include differential PSK or block-differential PSK.

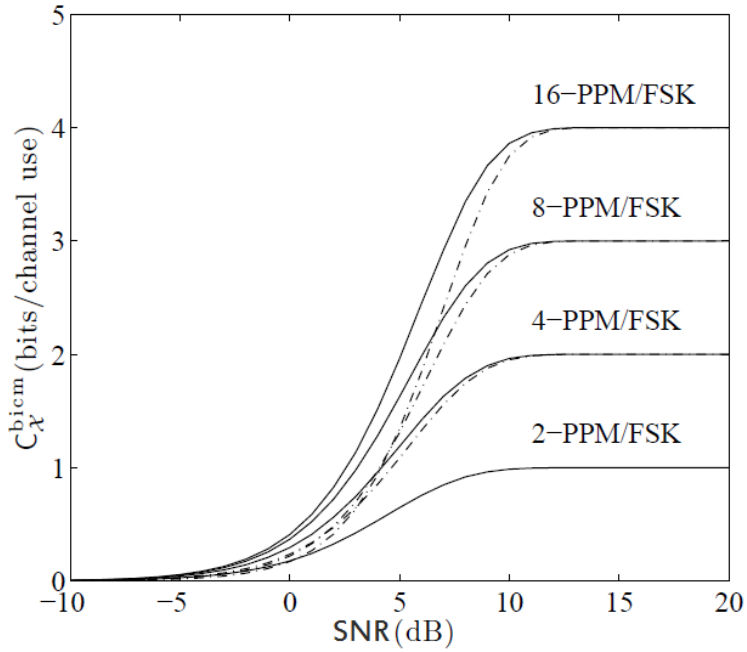

Figure 5.1: Coded modulation capacity (solid lines), BICM capacity (dash-dotted lines) for orthogonal modulation (PPM/FSK) with non-coherent detection in the AWGN channel.

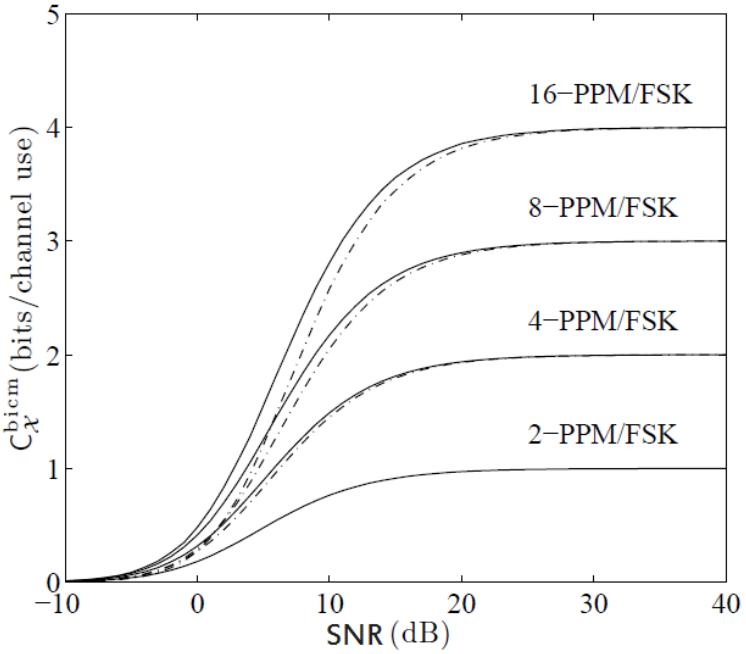

Figure 5.2: Coded modulation capacity (solid lines), BICM capacity (dash-dotted lines) for orthogonal modulation (PPM/FSK) with non-coherent detection in the fully-interleaved Rayleigh fading channel.

5.2 Block-fading

The block-fading channel [13] 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 coherence bandwidth and the channel coherence time [11]. 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 (EGDE), as well as transmission schemes based on OFDM, can also conveniently be modelled 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

$$P_{\boldsymbol{\theta}}\left(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{h}\right) = \prod_{i=1}^{B} \prod_{k=1}^{N} P_{\theta_i}\left(y_{i,k} \mid x_{i,k}, h_i\right) \tag{5.4}$$

where $x_{i,k}, h_i \in \mathbb{C}$ are respectively the transmitted and received symbols 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

$$P_{\theta_i}\left(y_{i,k} \mid x_{i,k}, h_i\right) \propto e^{-\left|y_{i,k} - \sqrt{\bar{\gamma}} h_i x_{i,k}\right|}. \tag{5.5}$$

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 and its channel capacity is zero. 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 [14] it was proved that the outage probability for sufficiently large SNR behaves as

$$P_{out}\left(\mathrm{SNR}\right) = \mathrm{SNR}^{-d_{sb}} \mathcal{K} \tag{5.6}$$

where d_{sb} is the slope of the outage probability in a log-log scale and is given by this bound

$$d_{sb} = 1 + \left\lfloor B\left(1 - \frac{R}{r}\right)\right\rfloor. \tag{5.7}$$

where the definition of r and R is the same that in Chapter 1. Hence, the error probability of efficient coded modulation schemes in the block-fading channel must have slope equal to this bound. Furthermore, this diversity is achievable by coded modulation as well as BICM. As in Figure 5.3, 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 $d_{sb} = 2$ when $B = 2$, I 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 previous bound gives $d_{sb} = 3$. In [14] the family of blockwise concatenated codes based on BICM was introduced. Improved binary codes for the block-fading channel can be combined with blockwise bit interleaving to yield powerful BICM schemes. Then, the error probability is averaged over the channel realizations. Similarly, in order to characterize the error probability for the particular code constructions, I 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 in [15], where the average over the fading is performed once the union bound has been truncated at 1.

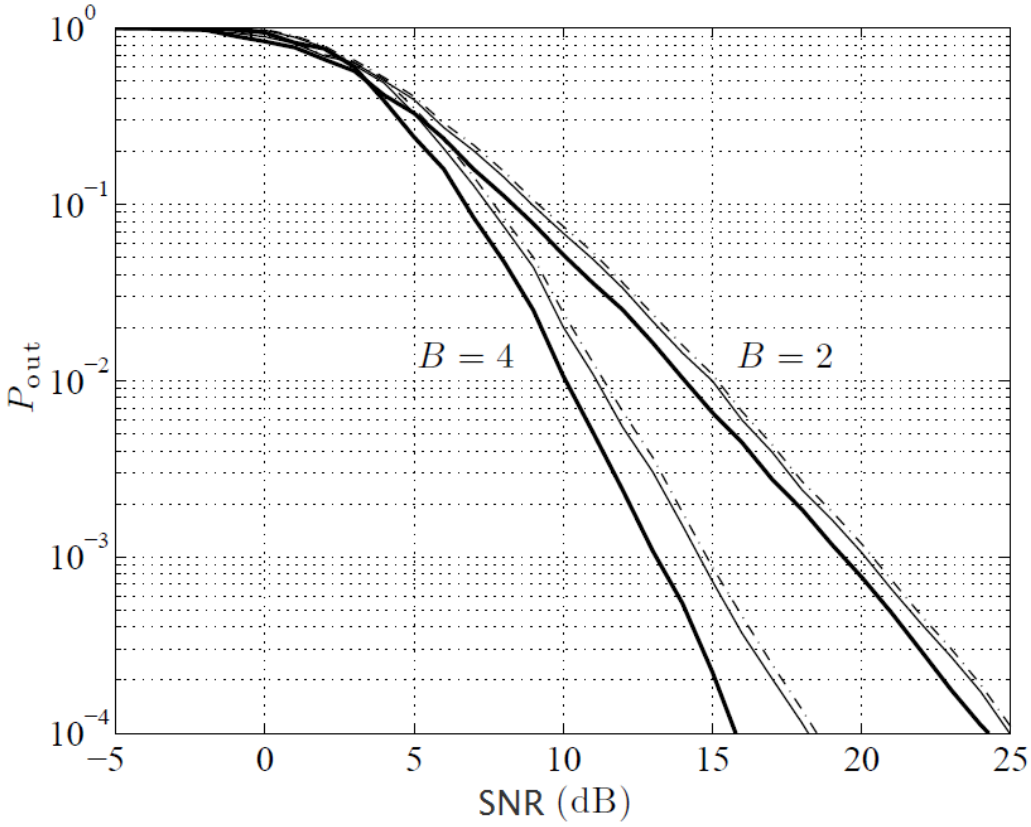

Figure 5.3: Outage probability with 16-QAM for $R = 2$ in a block-fading channel with $B = 2, 4$. Solid lines corresponds to coded modulation and dash-dotted lines correspond to BICM. The outage probability with Gaussian inputs is also shown for reference with thick solid lines.

5.3 MIMO - Multiple Input Multiple Output

Channel fading often limits the reliability of wireless communications but diversity can dramatically counteract the fading effects. Using several antennas at the transmitter and at the receiver brings space diversity. Multiple-input multiple-output (MIMO) systems can therefore offer an increased capacity and/or a better transmission reliability. The target of most MIMO schemes is either to increase the reliability by using the full diversity or to increase the bit rate by spatial multiplexing. Here I focus on a hybrid scheme. It is not a true spatial-multiplexing scheme because redundant information is added in the sent signal by means of bit-interleaved coded modulation [4]. This hybrid scheme can partially exploit the transmit diversity but it cannot claim that it benefits from the full diversity. It offers high flexibility because the encoder, the modulation and the number of transmit antennas can be chosen independently.

Another recent powerful technique is iterative processing at the receiver. By analogy with a turbo decoder, the iterations enable to jointly mitigate the interference and decode the bits (see Chapter 4). The turbo principle is well suited to the MIMO BICM scheme because there is an encoder at the transmitter and there is interference to deal with. Indeed, there is at least one kind of interference with a MIMO BICM structure: it is the co-antenna interference (CAI), which is due to the different signals that are simultaneously sent from the transmit antennas. In the sequel, two low-complexity turbo receivers will be described. One of them is well suited to a single-carrier (SC) modulation and the other one is designed for an orthogonal frequency-division multiplexing (OFDM) modulation. Then I show the ability of scheme to deal with the interference mitigation and with the decoding, and to exploit the space diversity and the frequency diversity.

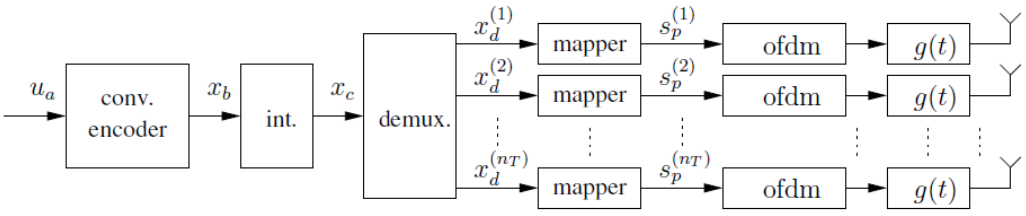

Figure 5.4: MIMO BICM OFDM transmitter.

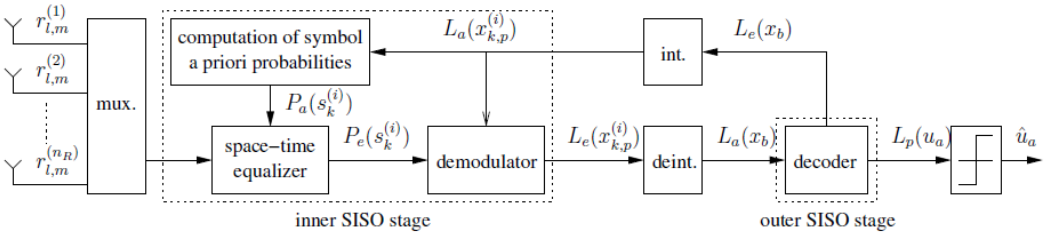

Figure 5.5: MIMO BICM SC receiver.

Transmitter: The MIMO BICM transmitter is depicted in Figure 5.4. The information bits are organized in frames. A frame of information bits is first encoded by a rate-r convolutional encoder. The coded bits are then interleaved by a random permutation. The frame of interleaved coded bits is then split into n_T sub-blocks, corresponding to the n_T transmit antennas. Within each of these sub-blocks, the interleaved coded bits are grouped and mapped to one of the complex symbols in the considered multilevel/phase constellation. Figure 5.4 illustrates the OFDM transmitter. After the mapping, the symbols enter the OFDM modulator and then the pulse-shaping filter. As far as the single-carrier (SC) transmitter is concerned, the OFDM boxes have to be removed from the figure and the pulse-shaping filter $g(t)$ is applied before the SC modulation. During one symbol period T, n_T symbols are transmitted: one per antenna. Since I assume that the transmitter does not have any knowledge of the channel impulse response, the antennas send signals with identical powers. They also use the same shaping filter and the same modulation and mapping rule.

This transmitter enables to benefit from the transmit diversity because the coding is performed before the frame is split between the antennas. Moreover, the presence of the bit-interleaver greatly reduces the correlation between successive coded bits. The combination of the interleaver and the encoder enables to use the turbo principle at the receiver. In the SC case, I can consider a bit-interleaved space-time code because the redundant information added by the encoder is spread over the space and time dimensions. In the OFDM case, I can consider a bit-interleaved space-frequency code because this redundant information is spread over the space and frequency dimensions. Note that both of these schemes offer high flexibility because the encoder, the modulation and the number of transmit antennas can be chosen independently.

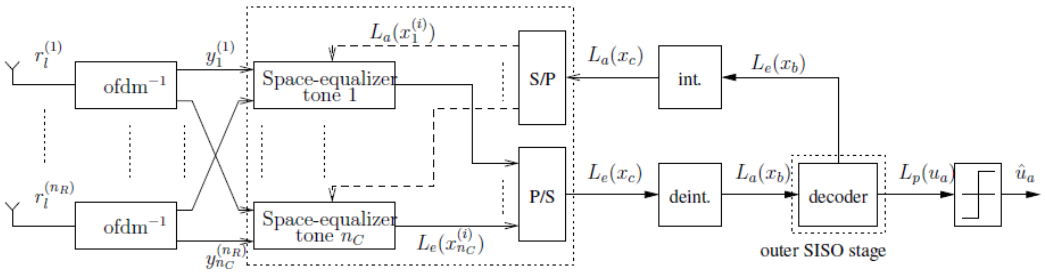

Figure 5.6: MIMO BICM OFDM receiver.

Receiver: Both receivers follow the same outline. They are iterative and make use of the turbo principle. They are made of the association of two soft-in soft-out (SISO) stages, separated by bit (de)interleavers, exchanging soft extrinsic information under the form of bit log-likelihood ratio's. I assume, as before, perfect channel knowledge and perfect synchronization.
The outer SISO stage performs optimal decoding with the algorithm of [16]. But the inner SISO stage is not the same in both schemes. The SC inner SISO stage has to deal jointly with the CAI and inter-symbol interference (ISI) mitigation and the demodulation as shown in Figure 5.5. OFDM transforms the frequency-selective MIMO channel into a set of frequency-flat MIMO channels. Therefore, the OFDM inner SISO stage has only to mitigate the CAI on each sub-carrier and to demodulate the symbols, as shown in Figure 5.6. The optimal inner stages need a trellis with a huge number of state to be analysed. Here I use low-complexity equalizers with linear filters based on the minimum mean square error (MMSE) criterion [17].

Channel model and simulation parameters: The considered channels are frequency-selective MIMO fading channels. The multipath channels are described by tap delays and tap gains, which are chosen according to either the GSM typical urban (TU) channel model or the Hiperlan one. The tap gains and delays are given in Figure 5.7. One tap does not represent one single path but the iteration of many paths with about the same delays. Assuming that the tap amplitudes are i.i.d. and that the number of paths is large enough, the taps are independent zero-mean complex Gaussian random variables. Moreover, the channel impulse responses between the various pairs of transmit and receive antennas are assumed to be independent from each other. That is realistic if the antenna spacing and

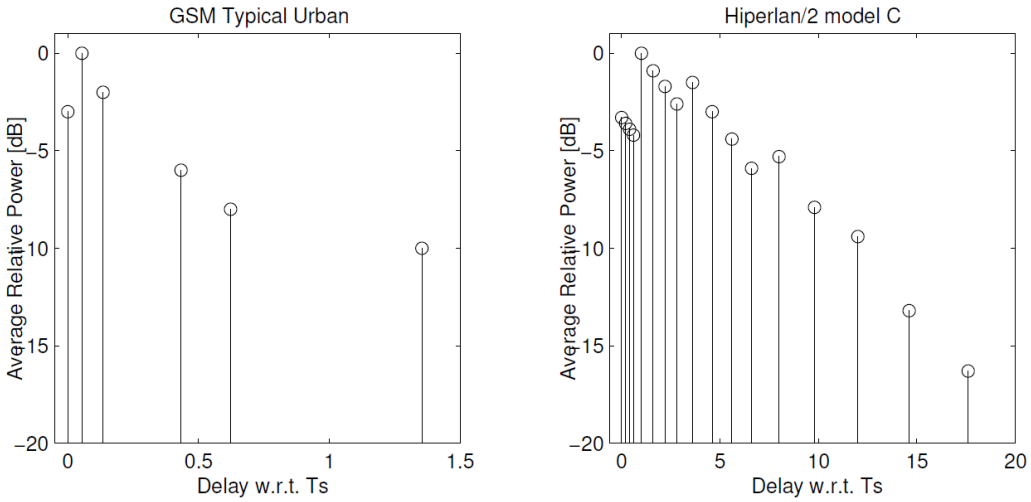

Figure 5.7: Tap gains and delays for the GSM TU and Hiperlan channel models.

the number of scatterers are sufficient. I assume quasi-static fading, which means that the channel taps remain constant over a frame length. It is reasonable assumption for relatively small frames if the channel is not varying too fast.
The simulations are stopped after at least 100 frame error for each SNR. SNR stands for the average bit energy to noise power spectral density ratio at each receive antenna. As in [18], for the SC modulation, the frame size is 400 information bits. The modulation is 8-PSK with Gray mapping. Here I use a convolutional encoder of rate 1/2 with constraint length 5 and generator polynomials $[23_8, 35_8]$. As in [18], for the OFDM modulation, the frame size is 574 bits. The modulation is 8-PSK with set partitioning mapping. I use a convolutional encoder of rate 1/2 with constraint length 3 and generator polynomials $[5_8, 7_8]$. (Here I reported the assumptions made in [18] for the analysis, in order to report the results and make them comprehensible by the reader).

Space diversity: At first sight, it may seem silly to use several antennas that transmit different information in the same frequency band and towards the same direction: it creates CAI and therefore degrades the performance. However, that is true in the non fading case but not over fading channels.
Fading is due to the multipath propagation. Depending on their relative phase, the different paths interact constructively or destructively. Since the relative phases of the paths may vary with the position of antenna, the channel gain be-

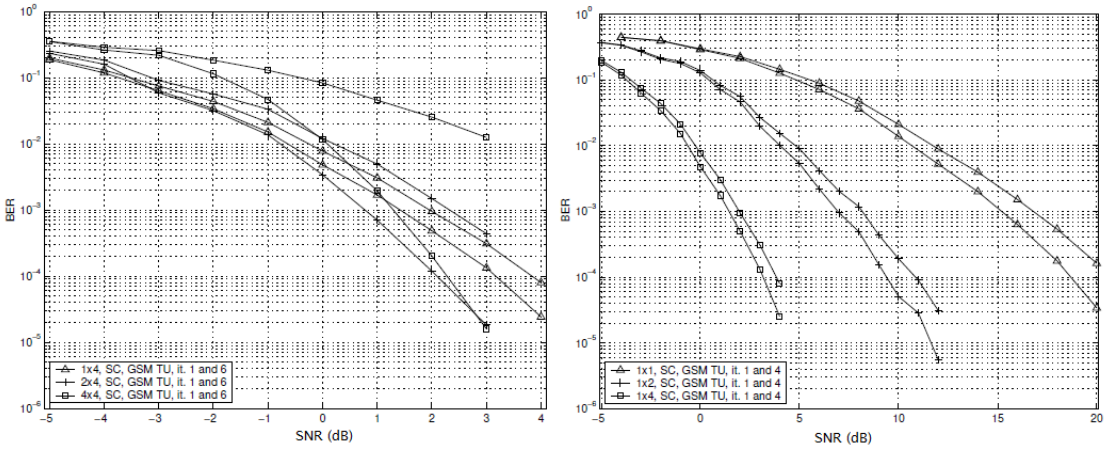

Figure 5.8: Transmit (left) and receive (right) diversity with the SC scheme over the GSM TU channel.

tween a transmit antenna and a receive antenna is different for every transmit antenna. The correlation between those gains is very low if the spacing between the transmit antennas is large and if there are many scatterers and thus many paths. It is possible to exploit this decorrelation to increase the capacity or the communications-link reliability. Two symbols carrying information about the same bit are sent from two different antennas. The reliability is increased since the probability that both channels fade simultaneously is lower than the probability that one channel fades. This is called the transmit diversity. The considered MIMO BICM transmitter benefits from transmit diversity. The convolutional encoder adds redundant information so that several coded bits carry information about the same bit. The coded bits are then interleaved and the stream is split into n_T sub-blocks. The information about one bit is very likely to be spread over several sub-blocks and thus to be sent by different transmit antennas.

The *left part* of Figure 5.8 shows the performance improvement when the number of transmit antennas n_T increases with the SC scheme over GSM TU channel. The number of receive antennas, n_R, is fixed to 4 while n_T takes the values 1,2 and 4. Without iteration the receiver cannot suppress the CAI: the more transmit antennas, the poorer the performance. However, at the sixth iteration, a large part of CAI can be cancelled. We see for instance that with two transmit antennas, the BER is better than a single transmit antenna. Regarding the $n_T = 4$ case, the performance is bad for low SNRs because the CAI cannot be cancelled. On the contrary, the CAI can be mitigated for higher SNRs and the scheme benefits

from transmit diversity: the $n_T = 2$ and $n_T = 4$ curves at iteration 6 cross each other at an SNR of abut 3 dB. Note that for a fixed SNR, the total transmitted power and the bit rate are proportional to n_T. Therefore, adding one more transmit antenna increases the spectral efficiency with a linear growth of the total transmitted power instead of the exponential growth that is usually needed.
The *right part* of Figure 5.8 shows the performance improvement when the number of receive antennas n_R increases. There is only one transmit antenna ($n_T = 1$) while n_R takes the values 1,2 and 4. The performance is dramatically improved when n_R increases. Two reasons can be identified. The first one is the array gain. If we are able to constructively combine the signals coming from the receive antennas and if the noises are independent, then each time n_R is doubled, the noise power is doubled too but the useful power of the signal is multiplied by four. So each time n_R is multiplied by two, the SNR increase is 3 dB what is called the array gain. The second improvement is due to the receive diversity because not all the receive antennas are likely to fade simultaneously.

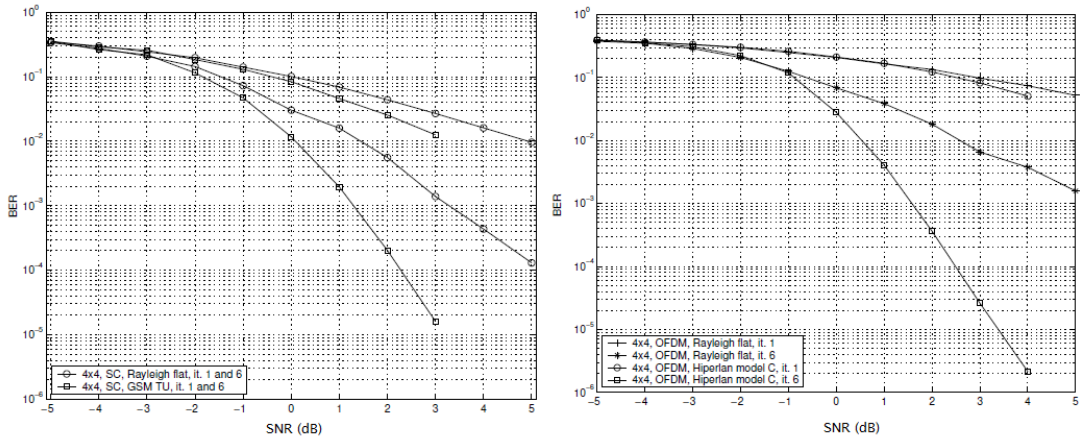

Figure 5.9: Frequency diversity with the SC scheme (left) and the OFDM scheme (right).

Frequency diversity: When the path delays are small with respect to the symbol period, around the inverse of the bandwidth, the different paths can not be distinguished by the receiver. In the case, the channel is frequency flat and there is no ISI. However, when the symbol period decreases, several groups of paths can be distinguished by the receiver. There is then ISI and the channel is frequency selective. Despite the ISI, this case offers an advantage. Indeed, if a distinguish-

able group of paths is referred to as a tap, the different taps will not suffer from the same fading. This kind of diversity is called the multipath diversity in the SC scheme. In the frequency domain, the relative phase differences of the paths become significant over the bandwidth, so that the fading becomes tone specific. This kind of diversity is called the frequency diversity in the OFDM scheme.
The multipath diversity and the frequency diversity are due to the same physical effect but they are not exploited the same way by the SC scheme and the OFDM one. In the SC scheme, every transmitted symbol is spread over several received symbols. The different replicas of the transmitted symbol are not identical because the fading is different for each tap. However, the receiver has to mitigate the ISI otherwise the performance are poorer than in the flat-fading case. The turbo receiver considered here is able to mitigate the ISI and thus to take advantage of the multipath diversity. That is shown in the *left part* of Figure 5.9. With a $(n_T, n_R) = (4, 4)$ system, we can see that the BER is better in the GSM TU case than in the flat-fading case.
Contrary to the SC scheme, the OFDM scheme needs an encoder in the transmitter to exploit the frequency diversity. The information about a bit needs to be spread over several tones by the encoder to exploit the diversity. In the considered scheme, this is implemented by the convolutional encoder, the interleaver and the OFDM modulator. The advantage brought by the frequency diversity of the Hiperlan channel is shown in the *right part* of Figure 5.9.

Chapter 6

Conclusion

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 and symbol-by-symbol detection. Error correcting codes were used to undo errors introduced by the physical modulation/demodulation process. This paradigm changed radically with the advent of Coded Modulation. Trellis-coded modulation was a topic 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 and LDPC codes were discovered, along with very efficient low-complexity Belief Propagation iterative decoding algorithms [10] 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 the first part of the thesis the main theme is that on some channels the separation of demodulation and decoding might be beneficial, provided that the

encoder output is interleaved bitwise and a suitable soft-decision metric is used in the Viterbi decoder. A comprehensive analysis of BICM, based on BER, shows this. Optimum and simpler, sub-optimum bit metrics are derived for channels with state information at the receiver. The central role of labelling map is also pinpointed.

After that, I have presented a generalized method for analysing the performance of BICM transmission. Its key element is a new approximation of the pdf of the bitwise reliability metrics, which is a valuable contribution in its own right. This approximation has led to closed-form expressions for Laplace transform of pdf, in term of which BER and cutoff rate of BICM can be expressed. Notably, the results are applicable to BICM with arbitrary QAM and PSK constellations and labelling rules, and transmission over Nakagami-m fading channels for arbitrary m. Furthermore, it was developed an asymptotic analysis which provides valuable insights into the performance of BICM over fading channels namely expressions for diversity order and asymptotic coding gain. Selected numerical results have confirmed the accuracy of the proposed analytical results for SNR regions of interest for convolutional coded BICM.

Then I have reviewed iterative decoding of BICM. In particular I have proved that BICM-ID significantly outperforms conventional BICM and this makes it suitable for both AWGN and Rayleigh fading channels.

In conclusion, reporting some applications for BICM, I have shown that in the real systems this pragmatic approach outperforms a lot of situations, overall where the *old* kind of systems are difficult to implement, and where is more constructively the use of new codes, as turbo-codes.

Appendix A

Viterbi Algorithm

The Viterbi algorithm (VA) is a dynamic programming algorithm for finding the most likely sequence of hidden states - called the *Viterbi path* - that results in a sequence of observed events. This algorithm is widely used for estimation and detection problems in digital communications and signal processing. It is used to detect signals in communication channels with memory, and to decode sequential error-control codes that are used to enhance the performance of digital communication systems [19]. In this appendix I explain the basics of the Viterbi algorithm as applied to digital communication systems.

Given a received sequence of symbols corrupted by AWGN, the VA finds the sequence of symbols in the given trellis that is closest in distance to the received sequence of noisy symbols. This sequence computed is the global *most likely sequence*. When Euclidean distance is used as a distance measure, the VA is the optimal maximum-likelihood detection method in AWGN. The use of this algorithm in this case is focused on the convolutional codes (Figure A.1), for which exists two way of decision decoding at the Viterbi decoder; one using Hamming distance (*hard decision decoding*) and the other using Euclidean distance (*soft decision decoding*). I consider here the second one [20].

To keep the information contained in the received noisy symbols, soft decision decoding computes the branch metrics by using the noisy symbols directly instead of quantizing them into bits. For each noisy symbol in the sequence, it calculates the squared distance between each noisy symbol and the corresponding coded bit in the coded bit sequence of the given transition, and then adds these distances together to form the branch metric of the transition. That is, if the given noisy

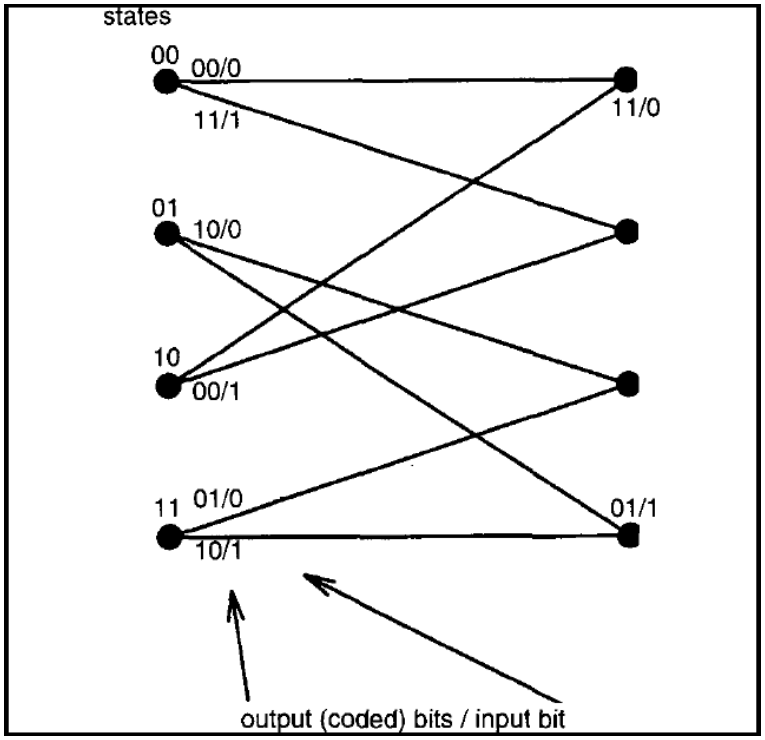

Figure A.1: An example 4-state convolutional code of a rate 1/2.

symbol is x and the coded bits are represented by a and b, then the distance to coded bits are $(x-a)^2$ and $(x-b)^2$, respectively. For a commonly used rate k/n convolutional code with binary output symbols represented by, say $-a$ and a, one can use negation operations to compute the metric for each bit [20]. Since there are n bits in the output (or coded) bit sequence of each transition, n noisy symbols are received at the input of the Viterbi decoder at each recursion. Thus, n negation operations are required to find the metric of all the received symbols and $(n-1)$ additions are required to compute the branch metric for a given coded bit sequence. Therefore, if m is the number of distinct coded bit sequences in the given trellis, at most $M(n-1)$ additions and n negations are required to compute the branch metrics for all the transitions and m memory locations are required to store these branch metrics.

Appendix B

Turbo-code

Forward-error-correcting (FEC) channel codes are commonly used to improve the energy efficiency of wireless communication systems. On the transmitter side, an FEC encoder adds redundancy to the data in the form of parity information. Then at the receiver, a FEC decoder is able to exploit the redundancy in such a way that a reasonable number of channel errors can be corrected. Because more channel errors can be tolerated with than without an FEC code, coded systems can afford to operate with a lower transmit power, transmit over longer distances, tolerate more interference, use smaller antennas, and transmit at a higher data rate. A binary FEC encoder takes in k bits at a time and produces an output (or code word) of n bits, where $n > k$. While there are 2^n possible sequences of n bits, only a small subset of them, 2^k to be exact, will be valid code words. The ratio k/n is called the *code rate* and is denoted by r. Lower rate codes, characterized by small values of r, can generally correct more channel errors than higher rate codes and are thus more energy efficient. However, higher rate codes are more bandwidth efficient than lower rate codes because the amount of overhead (in the form of parity bits) is lower. Thus the selection of the code rate involves a trade-off between energy efficiency and bandwidth efficiency. For every combination of code rate r, code word length n, modulation format, channel type, and received noise, there is a theoretic lower limit on the amount of energy that must be expended to convey one bit of information. This limit is called channel capacity or Shannon capacity. Since the dawn of information theory, engineers and mathematicians have tried to construct codes that achieve performance close to Shannon capacity.

The major advancement in coding theory occurred when a group of researchers developed *turbo codes*. The initial results showed that turbo codes could achieve energy efficiencies within only a half decibel of the Shannon capacity. After the research focused on improving the practicality of turbo codes, which have some peculiarities that make implementation less than straightforward.

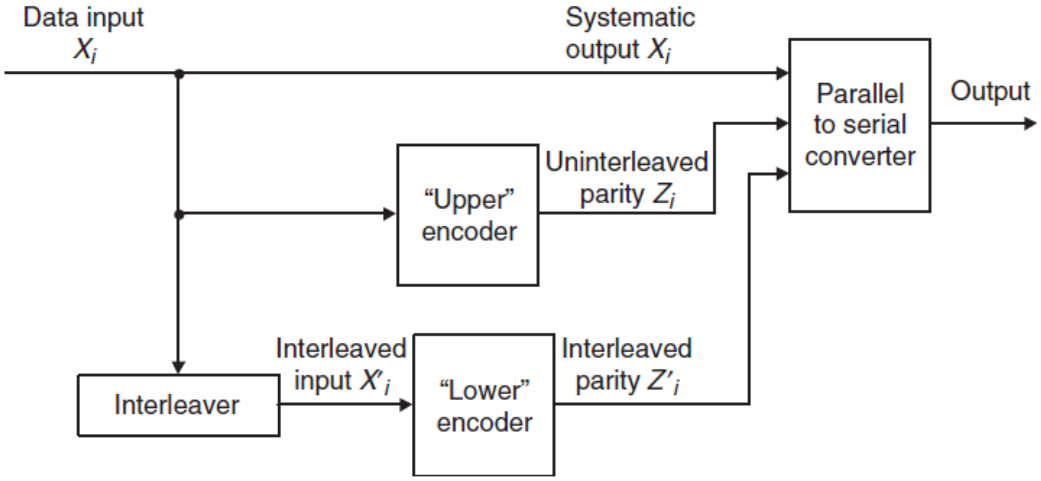

Figure B.1: A generic turbo encoder.

One of the most interesting characteristics of a turbo code is that is not just a single code. It is, in fact, a combination of two codes that work together to achieve a synergy that would not be possible by merely using one code by itself. In particular, a turbo code is formed from the parallel concatenation of two constituent codes separated by an interleaver. Each constituent code may be any type of FEC code used for conventional data communications. Although the two constituent encoders may be different, in practice they are normally identical. A generic structure for generating turbo codes is shown in Figure B.1. As can be seen, the turbo code consists of two identical constituent encoders, denoted as *ENC 1* and *ENC 2*. The input data stream and the parity outputs of the two parallel encoders are then serialized into a single turbo code word.

The interleaver is a critical part of the turbo code. It is a simple device that rearranges the order of the data bits in a prescribed, but irregular, manner. Although the same set of data bits is present at the output of the interleaver, the order of these bits has been changed, much like a shuffled deck of cards (although each input word is shuffled in exactly the same way). Without the interleaver, the

two constituent encoders would receive the data in the exact same order and thus - assuming identical constituent encoders - their outputs would be the same. This would not make for a very interesting (or powerful) code. However, by using an interleaver, the data X_i is rearranged so that the second encoder receives it in a different order, denoted X_i'. Thus the output of the second encoder will almost surely be different than the output of the first encoder - except in the rare case that the data looks exactly the same after it passes through the interleaver. Note that the interleaver used by a turbo code is quite different than the rectangular interleavers that are commonly used in wireless systems to help break up deep fades. While a rectangular channel interleaver tries to space the data out according to a regular patter, a turbo code interleaver tries to randomize the ordering of the data in an irregular manner.

Bibliography

[1] C.E. Shannon, *A Mathematical Theory on Communication*, Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, 1948.

[2] G. Ungerboeck, *Channel coding with multilevel/phase signal*, IEEE Transactions on Information Theory, vol. 28, pp. 55-67, 1982.

[3] E. Zehavi, *8-PSK trellis codes for Rayleigh channel*, IEEE Transactions on Communications, vol. 40, no. 5, pp. 873-884, 1992.

[4] G. Caire, G. Taricco, and E. Biglieri, *Bit-interleaved coded modulation*, IEEE Transactions on Information Theory, vol. 44, no. 3, pp. 927-946, 1998.

[5] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*, Kluwer Academic Publishers, Boston, Mass, USA, 1994.

[6] E. Biglieri, G. Caire, G. Taricco,J. Ventura-Traveset, *Computing Error Probabilities over Fading Channels: a Unified Approach*,European Transactions on Telecommunications, vol. 9, no. 1, pp. 15-25, 1998.

[7] A. Kenarsari Anhari and L. Lampe, *An Analytical Approach for Performance Evaluation of BICM Transmission over Nakagami-m Fading Channels*, IEEE Transactions on Communications, vol. 58, no. 4, pp. 1090-1101, 2010.

[8] X. Li and J. A. Ritcey, *Bit-interleaved coded modulation with iterative decoding*, IEEE Communication Letters, vol. 1, pp. 169-171, 1997.

[9] X. Li and J. A. Ritcey, *Trellis-code modulation with bit interleaving and iterative decoding*, IEEE Journal on Selected Areas in Communications, vol. 17, no. 4, pp. 715-724, 1999.

[10] A. G. i Fàbregas, A. Martinez and G. Caire, *Bit-Interleaved Coded Modulation*, Foundations and Trends in Communications and Information Theory: Vol. 5: No 1-2, pp. 1-153, 2008.

[11] J. G. Proakis, *Digital Communications*, McGraw-Hill USA, 4th edition, 2001.

[12] A. Guillén i Fàbregas and G. Caire, *Capacity approaching codes for non-coherent orthogonal modulation*, IEEE Transaction Wireless Communications, vol. 6, pp. 4004-4013, 2007.

[13] E. Biglieri, G. Caire and G. Taricco, *Error probability over fading channels. a unified approach*, European Transactions on Communications, 1998.

[14] A. Guillén i Fàbregas and G. Caire, *Coded modulation in the block-fading channel: Coding theorems and code construction*, IEEE Transactions on Information Theory, vol. 52, pp. 91-114, 2006.

[15] E. Malkämaki and H. Leib, *Evaluating the performance of convolutional codes over block fading channels*, IEEE Transactions on Information Theory, vol. 45, no.5, pp. 1643-1646, 1999.

[16] L.R. Bahl, J. Cocke, F. Jelinek and J. Raviv, *Optimal decoding of linear codes for minimizing symbol error rate*, IEEE Transactions on Information Theory, vol. 20, pp. 284-287, 1974.

[17] X. Wang, H.V. Poor, *Iterative (Turbo) soft interference cancellation and decoding for coded CDMA*, IEEE Transactions on Communications, vol. 47, pp. 1046-1061, 1999.

[18] X. Wautelet, D. Zuyderhoff and L. Vandendorpe, *Bit-interleaved coded modulation in MIMO systems*, communications and Remote Sensing Laboratory, Université catholique de Louvain, Belgium, paper.

[19] G. D. Forney Jr., *The Viterbi Algorithm*, IEEE Proceedings, vol. 61, no. 3, pp. 268-278, 1973.

[20] H. Lou, *Implementing the Viterbi Algorithm*, IEEE Signal Processing Magazine, pp. 42-52, 1995.

[21] M. Abramowitz and I.A. Stegun, *Handbook of mathematical functions*, Dover Publications, 1964.

[22] E. Biglieri, J. Proakis and S. Shamani, *Fading channels: information-theoretic and communications aspects*, IEEE Transactions on Information Theory, vol. 44, no. 6, pp. 2619-2691, 1998.

[23] R.W. Butler, *Saddlepoint Approximations with Applications*, Cambridge University Press, 2007.

[24] A. Guillén i Fàbregas and A. Grant, *Capacity approaching codes for non-coherent orthogonal modulation*, IEEE Transactions on Wireless Communications, vol. 6, pp. 4004-4013, 2007.

Acknowledgements

I miei più sentiti ringraziamenti vanno alla mia fidanzata Eleonora e a tutti i componenti della mia la mia famiglia, papà Piergiorgio, mamma Giovanna, fratello Enrico, che mi hanno accompagnato lungo questo percorso di studi. Senza di loro tutto ciò sarebbe stato impossibile.

Ringrazio inoltre tutte le persone con cui in questi anni ho condiviso la mia esperienza universitaria, partendo dai compagni di corso Daniele, Andrea, Davide, Paolo, Alessio, Fabio, Enrico, Damiano, Matteo, Giulio, Viviana e Francesca, i compagni di studio della biblioteca di Roncade, Giulio, Giovanni, Gaia, Eleonora, Chiara, Elisa, Marta, Francesca, Mattia e Marcello.

Il più affettuoso dei ringraziamneti va a tutta la mia compagnia del Parkejo nella quale ho sempre trovato tutta la serenità, la felicità e il supporto sia morale che fisico di cui avevo bisogno: Pè, Gero, Sturo, Alessia, Kalta, Bene, Davide, Erica, Gandi, John B, Cavei, Ivano, Lu, Jim, Nico, Teo, Monia, Monica, Filtro, Dorothy, Arp, Vale e Valeri.

Come ultimo ma non meno importante ('last but not least') un ringraziamneto speciale al mio amico Emanuele con il quale ho trascorso dei fantastici periodi.

GRAZIE MILLE A TUTTI!

Alberto Desiderà