SIGNAL MAPPING DESIGNS FOR BIT-INTERLEAVED CODED MODULATION WITH ITERATIVE DECODING (BICM-ID)

A Thesis Submitted
to the College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in the Department of Electrical Engineering
University of Saskatchewan

by

Nghi H. Tran

Saskatoon, Saskatchewan, Canada

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, it is agreed that the Libraries of this University may make it freely available for inspection. Permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised this thesis work or, in their absence, by the Head of the Department of Electrical Engineering or the Dean of the College of Graduate Studies and Research at the University of Saskatchewan. Any copying, publication, or use of this thesis, or parts thereof, for financial gain without the written permission of the author is strictly prohibited. Proper recognition shall be given to the author and to the University of Saskatchewan in any scholarly use which may be made of any material in this thesis.

Request for permission to copy or to make any other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Electrical Engineering
57 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
S7N 5A9

ACKNOWLEDGMENTS

I would like to express my deepest appreciation and gratitude to my supervisor, Professor Ha Nguyen for his support and guidance during my studies. He has been a great source of knowledge and has inspired me with many ideas which are very useful for my research. His patience, encouragement and availability made this thesis possible.

The financial support from Telecommunication Research Laboratories (TRLabs) are gratefully acknowledged. This financial source is important in order to help me keep working on my research and finish this thesis.

I would like to say thank you to my labmates as well as all of my friends here for their support and encouragement. Without them, life here would have been more difficult.

Finally, my deepest love and gratitude is devoted to my parents and my sister. They are always a source of love, supporting and encouragement. To them, I dedicate this thesis.

UNIVERSITY OF SASKATCHEWAN

Electrical Engineering Abstract

SIGNAL MAPPING DESIGNS FOR BIT-INTERLEAVED CODED MODULATION WITH ITERATIVE DECODING (BICM-ID)

Student: Nghi H. Tran

Supervisor: Prof. Ha H. Nguyen

M.Sc. Thesis Submitted to the

College of Graduate Studies and Research

December, 2004

ABSTRACT

Bit-interleaved coded modulation with iterative decoding (BICM-ID) is a spectral efficient coded modulation technique to improve the performance of digital communication systems. It has been widely known that for fixed signal constellation, interleaver and error control code, signal mapping plays an important role in determining the error performance of a BICM-ID system. This thesis concentrates on signal mapping designs for BICM-ID systems. To this end, the distance criteria to find the best mapping in terms of the asymptotic performance are first analytically derived for different channel models. Such criteria are then used to find good mappings for various two-dimensional 8-ary constellations. The usefulness of the proposed mappings of 8-ary constellations is verified by both the error floor bound and simulation results.

Moreover, new mappings are also proposed for BICM-ID systems employing the quadrature phase shift keying (QPSK) constellation. The new mappings are obtained by considering many QPSK symbols over a multiple symbol interval, which essentially creates hypercube constellations. Analytical and simulation results show that the use

of the proposed mappings together with very simple convolutional codes can offer significant coding gains over the conventional BICM-ID systems for all the channel models considered. Such coding gains are achieved without any bandwidth nor power expansion and with a very small increase in the system complexity.

Table of Contents

List of Tables

List of Figures

ABBREVIATIONS

ASK	Amplitude Shift Keying
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BICM	Bit-Interleaved Coded Modulation
BICM-EX	Bit-Interleaved Coded Modulation Expurgated
BICM-ID	Bit-Interleaved Coded Modulation with Iterative Decoding
BI-STCM-ID	Bit-Interleaved Space Time Coded-Modulation with Iterative Decoding
BPSK	Binary Phase Shift Keying
BSA	Binary Switching Algorithm
CC	Convolutional Code
CSI	Channel State Information
dB	Decibel
EF	Error-free Feedback
FED	Free Euclidean Distance
FEDC	Free Euclidean Distance Conditioned on the ideal feedback
MAP	Maximum A Posteriori Probability
MIMO	Multi-Input Multi-Output
ML	Maximum Likelihood
MLC	Multilevel Coding
PEP	Pairwise Error Probability
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
SISO	Soft-Input Soft-Output
SNR	Signal-to-Noise Ratio

SSP	Semi Set Partitioning
TCM	Trellis-Coded Modulation
TTCM	Turbo Trellis-Coded Modulation

1. Introduction

A general digital communication system includes three basic elements: a transmitter, a communication channel and a receiver. The transmitter translates the information bits to signals that can be effectively transmitted over the channel. The communication channel is the physical medium used to send the signal from the transmitter to the receiver. At the receiving end of a digital communication system, the receiver tries to recover the transmitted information as correctly as possible.

In almost any communication systems, the channel encoder is a very important part of the transmitter. The purpose of the channel encoder is to introduce, in a controlled manner, some redundancy in the binary information sequence. This redundant information can be used at the receiver to overcome the effects of noise and interference encountered during the transmission of the signal through the channel. The use of error control coding in digital communication results in coded systems, as illustrated in Fig. 1.1.

The transmitter of such systems consists of an encoder and a modulator. Most of the error control codes that have been devised for increasing the reliability of information transmission are effective when the channel is assumed to be memoryless (i.e., the channel imperfectness is independent from symbol to symbol). For correlated channels, the errors are no longer statistically independent and the memoryless assumption is invalid. The successive symbol-by-symbol detection becomes unreliable, hence, the performance of the codes designed for memoryless channel is degraded. One example of correlated channels is the class of channels characterized by multipaths and fading. One of the effective ways to deal with this class of channels is to

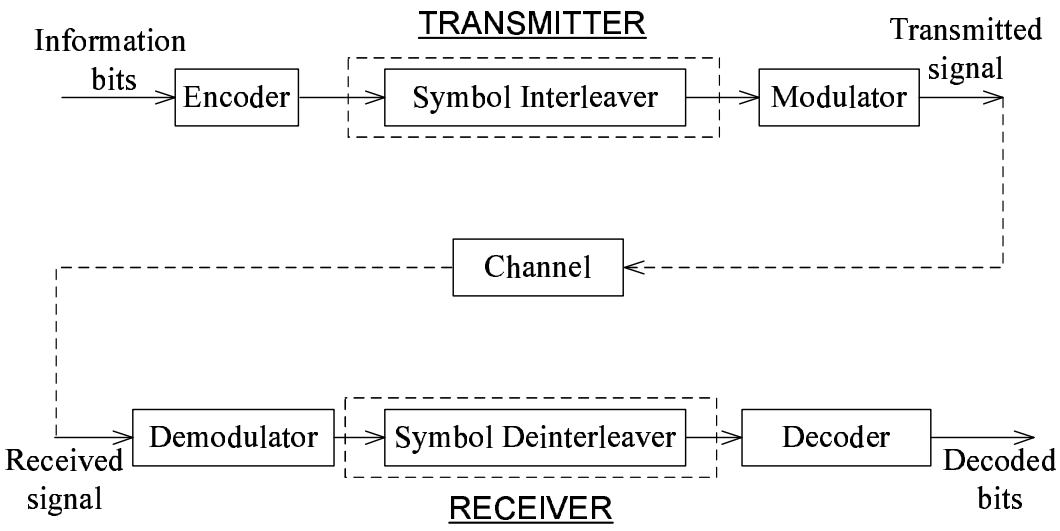

Figure 1.1 The block diagram of a coded communication system.

interleave coded data in such a way that the bursty channel is transformed into a channel having independent errors. In this case, the encoded data is reordered by the interleaver as shown in Fig. 1.1. The interleaver is used to break the correlation of the channel, and as a result, error bursts are spread out in time so that errors within a codeword appear to be independent. At the receiver, the opposite operations are performed accordingly. After demodulation, the deinterleaver puts the demodulated data in a proper sequence and passes it to the decoder. The decoder attempts to reconstruct the original information sequence from the knowledge of the code used by the channel encoder and the redundancy contained in the received data.

In general, there are two types of error control codes, namely linear block codes and trellis codes (or convolutional codes). This thesis only focuses on convolutional codes due to the existence of an efficient "soft decoding" method for these codes.

Traditionally, the channel encoder and the modulator are treated separately. The design of a good code strives to maximize its minimum Hamming distance d_H [3]. Using error control coding improves the performance of the system at the cost of spending more bandwidth for the transmitted signal, and an increase in the receiver implementation complexity. Generally, the use of a very powerful code is required to

trade the bandwidth and implementation complexity for the transmitted power.

On the other hand, if one wants to have the benefit of coding gain without increasing the transmission bandwidth, one must employ a multilevel/multiphase-modulation scheme. However, the use of such a modulation scheme results in a more crowded constellations and therefore it has a negative effect on the error performance of the overall system.

If the modulator and the encoder are designed as a single identity, the loss due to the expansion of the signal set can be overcome and a significant coding gain can be obtained by using relatively simple codes. The integration of these two components results in a digital coded modulation system and it was first studied by Massey [4]. Coded modulation that jointly optimizes coding and modulation is now a popular and powerful technique to improve the performance of digital communication systems whose bandwidths are limited. In 1982, Ungerboeck introduced a trellis-coded modulation (TCM) system as a bandwidth-efficient signaling over an additive white Gaussian noise (AWGN) channel [5]. The first important insight of Ungerboek is that for the non-binary signal constellations used in TCM (such as ASK, PSK and QAM), the minimum free Euclidean distance (FED), not the minimum free Hamming distance of the code, is the primary parameter that determines the code's performance. The technique of "mapping by set partitioning" suggested in [5], as opposed to the conventional Gray mapping used in uncoded systems, turned out to be the key to achieving performance gains with TCM compared to uncoded modulation. One of the important goals of mapping by set partitioning is to insure that parallel transitions in the code trellis are mapped into signals far apart in the signal constellation, thus minimizing the probability of short, one-branch error event due to the uncoded bits.

Independently, multilevel-coded modulation employing multiple binary codes was proposed by Imai and Hirakawa in 1977 [6]. The main idea is also to optimize the code in the Euclidean space rather than dealing with Hamming distance as in classical coding schemes. The multilevel coding (MLC) protects each bit of the signal point by an individual binary code. The individual codes are chosen in such a way that the

minimum distance of the Euclidean code space is maximized. In contrast to TCM, the MLC approach provides flexible transmission rates and it can employ block codes, convolutional codes or concatenated codes as component codes. The disadvantage of MLC compared to TCM is that it requires multiple encoders and decoders.

As mentioned earlier, Ungerboeck's TCM approach is based on mapping by set partitioning that maximizes the minimum intra-subset Euclidean distance [5]. Such an approach is only advantageous for AWGN channels. For fading channels, the primary design criterion is shifted to the diversity order of the coded modulation system [7]. More specifically, it is demonstrated in [7] that the Hamming distance between any two transmitted sequences is the primary factor that determines the system performance at high signal-to-noise ratio (SNR). Since conventional TCM is designed to maximize the Euclidean distance, it usually exhibits a low diversity order. Therefore, the performance of TCM over fading channels is significantly degraded. Adding a symbol interleaver and avoiding parallel transitions are the well known techniques to improve the performance of TCM over fading channels [7]. However, the diversity order for any symbol interleaved coded system is limited to the minimum number of distinct symbols along any error event. This implies that an increase in diversity order can only be done at the cost of increasing the complexity of the code. To improve the performance of TCM over fading channels, the principle of interleaving over coordinates was recently studied in [8], where it is shown that the number of different coordinates between two coded sequences is the diversity order in this case. Based on this technique, the signal set design for TCM that achieves a high level of diversity order while preserving the coding gain over AWGN channels was proposed in [9].

In [1], a different approach, called bit-interleaved coded modulation (BICM), was suggested by Zehavi to improve the performance of coded modulation systems over fading channels. It was shown that the diversity order can be increased to the minimum number of distinct bits rather than the number of distinct channel symbols by using bitwise interleaving. This scheme, however, increases the time diversity of

coded modulation at the expense of reducing the free squared Euclidean distance (FED), leading to a performance degradation over additive white Gaussian noise (AWGN) channels when compared to TCM systems [1, 10]. Reference [10] presents in a comprehensive fashion the theory of BICM, which provides tools for evaluating its performance and gives guidelines for the design. As BICM is the main topic of this thesis, the theory of BICM is reviewed in Chapter 2.

Since the invention of turbo codes [11], interleaving and iterative processing have also been applied to coded modulation systems. It was shown in [12–18] that with iterative decoding, BICM, a bandwidth-efficient approach primarily considered in the past for fading channels, can in fact be used to provide excellent error performance over both Gaussian and fading channels. More specifically, it was shown in [14] that with a careful design of signal mapping, iterative decoding can increase the minimum intersignal Euclidean distance of BICM while retaining the desirable Hamming distance. This makes BICM with iterative decoding (BICM-ID) greatly outperform TCM and compare favorably with the bandwidth-efficient turbo TCM (TTCM) [19]. The advantage of BICM-ID over TTCM is that it requires only one soft-input soft-output (SISO) decoder instead of two as normally used in turbo decoding [11, 19].

For given signal constellation, interleaver and convolutional code, signal mapping plays an important role in determining the error performance of a BICM-ID system. Generally, there are two approaches to study the effect of signal mapping on the performance of BICM-ID systems. The first technique is based on the error bound [12–14], which is only related to the asymptotic performance of the systems. Another technique based on mutual information was also introduced in [16, 18]. This thesis applies these above two techniques to study the signal mapping designs in order to improve the error performance of BICM-ID systems. The designs are considered for both AWGN and Rayleigh fading channels.

1.1 Thesis Contributions

The first part of this thesis considers BICM-ID systems employing two-dimensional constellations. The mapping problem is studied based on both the upper bound of the error performance and the bitwise mutual information. Good signal mappings of various 8-ary constellations are then proposed. These mappings are the best mappings in terms of asymptotic performance of BICM-ID systems. The bit error rate (BER) performance as well as the convergence behavior are discussed and verified by analytical and simulation results.

The second part of the thesis is concerned with BICM-ID systems employing a QPSK constellation. Using bitwise mutual information, reference [17] shows that while Gray mapping of QPSK is the best for BICM (i.e., when no iteration is implemented between the channel decoder and the demodulator), it is the anti-Gray mapping that is preferred for BICM-ID[1]. In contrast to [17], here, instead of mapping two coded bits to one QPSK symbol, a group of $m = 2n$ coded bits, $n > 1$, is mapped to n consecutive QPSK symbols. Since a QPSK constellation is built from two quadrature (orthogonal) carriers, there are 4^n distinct combinations of n QPSK symbols and they are the vertices of a $2n$-dimensional hypercube ($2n$-cube). Performing signal mapping on this $2n$-cube is exactly what is proposed in the second part of this thesis. It should be mentioned that using Gray or anti-Gray mapping for each QPSK constellation is just one special case of the above general mapping. The obvious question is "how to find the best mapping ?". Since the number of signal points in this hypercube constellation can be very large, a computer search to find the best mapping becomes impossible. As an example, there are $16! \geq 10^{13}$ possible mappings for a four-dimensional hypercube.

Nevertheless, by exploiting the symmetry of a hypercube constellation, a universal condition for the best mapping with respect to the asymptotic performance is established in this thesis. A simple and general algorithm to construct the best mapping

[1]Note that Gray and anti-Gray mappings are the only two mappings available for a QPSK constellation.

is also proposed. Extensive simulations have been carried out to demonstrate the advantage of the proposed systems compared to conventional ones. The tightness of the error bound derived in this thesis is also observed. It is shown that significant coding gains over the conventional systems can be obtained by using very simple convolutional codes in the proposed systems.

1.2 Thesis Organization

The remaining of the thesis is organized as follows.

Chapter 2 studies bit-interleaved coded modulation (BICM) systems [1]. The suboptimum decoding method is first introduced. The influence of the signal mapping to the error performance will be discussed based on the error bound and the technique of mutual information. It is then shown that for fixed convolutional code and signal constellation, signal mapping plays a crucial role in determining the error performance. The most suitable mappings are also mentioned with respect to the asymptotic performance.

Chapter 3 is devoted to the conventional bit-interleaved coded modulation with iterative decoding (BICM-ID) systems. As bandwidth efficiency is a primary concern in this thesis, the signal constellations employed are in two-dimensional signal space. The chapter starts with the introduction of the system model and the iterative processing between the demodulator and the channel decoder. Two types of feedback from the decoder to the demodulator will be discussed. They are the hard-decision feedback and the soft-decision feedback. As before, the performance evaluation of the system will be studied using the error bound and mutual information. Finally, the mapping problems of different 8-ary constellations are investigated and good signal mappings are proposed.

Chapter 4 is the main contribution of this thesis, which introduces the BICM-ID systems employing QPSK constellation, viewed as a hypercube. First, a hypercube and its distance properties is studied. Then the proposed system is described. The

influence of signal mapping on the hypercube to the error performance will be studied by evaluating the error floor of the system and analyzing the bitwise mutual information. By exploiting the symmetric properties of the hypercube, the distance criterion of the best mapping in terms of the asymptotic performance for different channel models is established. A universal condition for the best mapping is then introduced. A simple and general algorithm to construct the best mapping for the system is then proposed. Analytical and simulation results are also provided in this chapter to demonstrate the advantages of the proposed systems and the tightness of the error bound.

Finally, Chapter 5 draws the conclusions and gives suggestions for further studies.

2. Bit-Interleaved Coded Modulation (BICM)

For a communication system operating over a multi-path fading environment, the correlation between the sequential fading coefficients degrades the error performance of the system. It is well known that over a fading channel, the diversity order of any coded system is the key parameter in determining its error performance [7]. Traditionally, the symbol interleaver is used to break the fading correlation and the diversity order is measured at the symbol level. With a symbol interleaver, the diversity order is the minimum number of distinct symbols between any two codewords. Thus, the diversity order can only be increased by preventing parallel transitions and increasing the constraint length of the convolutional code.

The potential of BICM systems over a fading channel was first recognized by Zehavi in [1] where he suggested a coded system built from a convolutional encoder followed by random bit interleavers. More specifically, a system with an 8-state, rate–2/3 convolutional code and three random bit interleavers as shown in Fig. 2.1 was proposed. A group of 3 bits at the output of the interleavers is mapped to a symbol in 8-PSK constellation by Gray mapping[1]. By assuming random interleavers, the combined interleavers and mapping can be viewed as three statistically independent communication modulators and channels. It was shown in [1] that, due to the random modulation, higher diversity order can be achieved with the proposed system compared to the conventional coded system. Thus, BICM systems are quite attractive for transmission over fading channels. More specifically, it was pointed out in [1] that

[1]Gray mapping is the mapping in which the labels of two nearest signal points differ in only one bit.

the diversity order can be increased to the minimum number of distinct bits between two codewords rather than the number of distinct symbols. The comparison made in [1] showed that an 8-state, rate–2/3 convolutional code with 8-PSK Gray mapping and random bit interleaving outperforms the 8-state, 8-PSK Ungerboeck's TCM with symbol interleaving over a Rayleigh fading channel by more than 2dB at the bit error rate level of 10^{-5}.

Recently, an analytical framework for the evaluation of Zehavi's BICM systems was presented in [10]. The advantage of a BICM system is that it treats coding and modulation as separate components, hence, the code and the modulator can be flexibly selected. The diversity order can now be maximized by using the best convolutional code, which provides the largest d_H for given code rate and constraint length. The following discussion of BICM systems is mainly based on [10]. Special attention is paid to the signal mapping in order to improve the system performance.

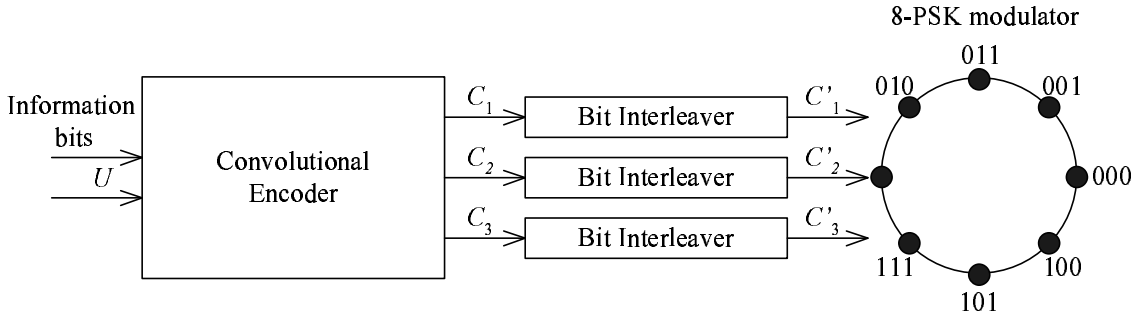

Figure 2.1 BICM transmitter with 8-PSK constellation and Gray mapping [1].

2.1 The BICM System Model

A conventional BICM system is built from a serial concatenation of a channel encoder, a bit interleaver π and an M-ary memoryless modulator (where $M = 2^m$) as shown in Fig. 2.2. The sequence of information bits $\underline{\mathbf{u}}$ is first encoded by a convolutional encoder to produce the output coded bit sequence $\underline{\mathbf{c}}$. The convolutional code should be chosen to be optimal in the sense that it gives the largest free Hamming distance d_H for given code rate and constraint length. The pseudoran-

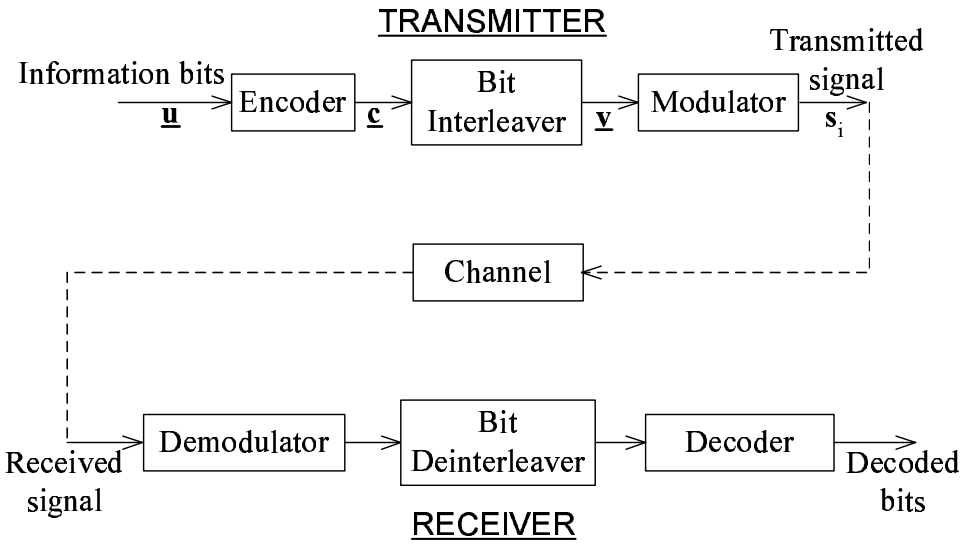

Figure 2.2 General block diagram of a BICM system.

dom interleaver permutes the encoded bits, as opposed to the channel symbols in the symbol-interleaved coded systems. The purpose of the interleaver is to break the fading correlation and increase the diversity order to the minimum Hamming distance d_H of the convolutional code. Note that there is only a single bit interleaver in Fig. 2.2 instead of m bit interleavers used in Zehavi's appoach. The reason for not using m bit interleavers is that it limits the flexibility of BICM and complicates the analysis [10]. Moreover, using m bit interleavers creates a fixed correspondence between the coded bits and the label positions. This leads to unequal error protection and suboptimal performance when the code is chosen at random [10].

After the interleaver, m consecutive bits of the interleaved coded sequence are grouped as a channel symbol $\mathbf{v} = (v_1, ..., v_m)$. The complex transmitted signal $\mathbf{s}_i = \xi(\mathbf{v})$, $1 \leq i \leq M$, is then chosen from the M-ary constellation Ψ to carry the m coded bits over each symbol duration. Here, ξ denotes the mapping scheme from the bit patterns to the constellation points. For a frequency non-selective Rayleigh fading channel and coherent detection, the received signal during each symbol interval can be written as follows:

$$\mathbf{r} = g\mathbf{s}_i + \mathbf{w} \tag{2.1}$$

In (2.1), $\mathbf{w}$ is complex white Gaussian noise with independent inphase and quadrature components having two-sided power spectral density $\sigma^2 = N_0/2$. The scalar g is a Rayleigh random variable representing the fading amplitude of the transmitted signal $\mathbf{s}_i$. It is assumed that the channel fades slowly so that the fading amplitude is constant over one symbol duration. It is also assumed that g is perfectly estimated at the receiver. Note that in the case of an AWGN channel, $g = 1$.

The receiver of a BICM system includes three elements: the demodulator, the de-interleaver and the convolutional decoder. Due to the presence of the random bit interleaver, the true maximum likelihood decoding of BICM requires joint demodulation and convolutional decoding and it is therefore too complicated to implement in practice. In [1], as a trade-off between the complexity and error performance, Zehavi suggested a suboptimal decoding method that includes two separate steps: the bit metric computation and the decoding of convolutional code by the Viterbi algorithm [3]. For each received signal $\mathbf{r}$, $2m$ log-likelihood bit metrics for m coded bits v_k, $1 \leq k \leq m$, are computed using the maximum likelihood (ML) rule as follows:

$$\lambda(v_k = b) = \log p(\mathbf{r}|v_k = b, g) \sim \log \sum_{\mathbf{s}_i \in \Psi_b^k} p(\mathbf{r}|\mathbf{s}_i, g) \tag{2.2}$$

where Ψ_b^k, $b \in \{0, 1\}$, denotes the subset of Ψ that contains all symbols whose labels have the value b in the kth position. The function $p(\mathbf{r}|\mathbf{s}_i, g)$ is the probability density function of the received signal $\mathbf{r}$ given the fading amplitude g and signal $\mathbf{s}_i$ was transmitted. The symbol $\sim$ in (2.2) indicates replacement by an equivalent statistic. With a two-dimensional M-ary signal constellation Ψ, $p(\mathbf{r}|\mathbf{s}_i, g)$ is given as:

$$p(\mathbf{r}|\mathbf{s}_i, g) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(r_1 - g s_{i1})^2 + (r_2 - g s_{i2})^2}{2\sigma^2}\right] \tag{2.3}$$

where (r_1, r_2) and (s_{i1}, s_{i2}) are the components of $\mathbf{r}$ and $\mathbf{s}_i$, respectively.

For practical implementation, the log-sum calculation is preferred and the suboptimum maximum log-likelihood bit metric is written as:

$$\lambda(v_k = b) \approx \max_{\mathbf{s}_i \in \Psi_b^k} \log p(\mathbf{r}|\mathbf{s}_i, g) = - \min_{\mathbf{s}_i \in \Psi_b^k} ||\mathbf{r} - g\mathbf{s}_i||^2 \tag{2.4}$$

Equation (2.4) basically implies that each bit metric is computed based on the minimum squared Euclidean distance between the received signal $\mathbf{r}$ and the signal point $\mathbf{s}_i$ over the subset Ψ_b^k. At the decoder, the trellis branch metrics are computed by summing the corresponding bit metrics after de-interleaving. The standard Viterbi algorithm for convolutional code is then applied in the final step.

An example of subset partitioning for each bit of 8-PSK constellation with Gray mapping is shown in Fig. 2.3. The bit metric for the first bit of a transmitted signal

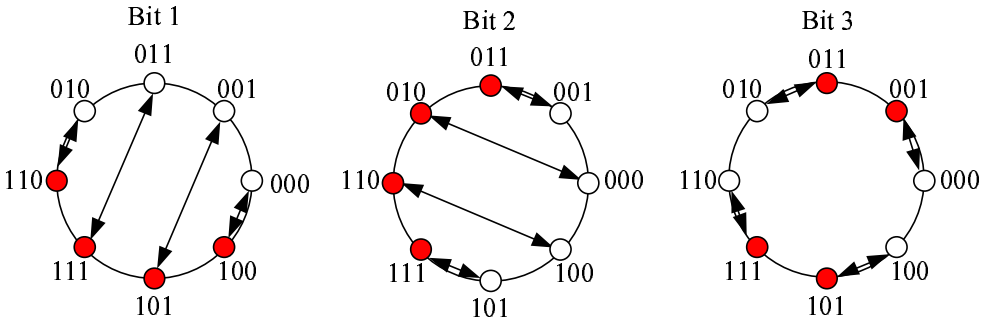

Figure 2.3 The subset partitioning for each bit of Gray mapping of 8-PSK constellation.

taking on value 1 is:

$$\lambda(v_1 = 1) = -\min\left(d_{100}^2, d_{101}^2, d_{110}^2, d_{111}^2\right) \tag{2.5}$$

Likewise, the bit metric for the first bit taking on value 0 is:

$$\lambda(v_1 = 0) = -\min\left(d_{000}^2, d_{001}^2, d_{010}^2, d_{011}^2\right) \tag{2.6}$$

In (2.5) and (2.6), $d_{\mathbf{v}}^2$, $\mathbf{v} = 000, \ldots, 111$, is the squared Euclidean distance between the received signal $\mathbf{r}$ and the transmitted signal associated with label $\mathbf{v}$. The bit metrics for the other two bit positions can be calculated similarly.

2.2 The Role of Signal Mapping

This section discusses the influence of signal mapping to the error performance of BICM systems. Two relevant techniques, namely the error bound and mutual information, will be reviewed first.

2.2.1 The Error Bound of BICM Systems

Using the same notations as in [10], the union bound of the bit-error-rate (BER) for a BICM system employing a rate-k_c/n_c convolutional code, a constellation Ψ and a mapping ξ is given by:

$$P_b \leq \frac{1}{k_c} \sum_{d=d_H}^{\infty} c_d f(d, \Psi, \xi) \tag{2.7}$$

In (2.7), c_d is the total information weight of all error events at Hamming distance d and d_H is the free Hamming distance of the code. The function $f(d, \Psi, \xi)$ is the average pairwise probability, which depends on the Hamming distance d, the constellation Ψ and the mapping ξ. Let $\underline{\mathbf{c}}$ and $\breve{\underline{\mathbf{c}}}$ denote the input sequence and the estimated sequence with the Hamming distance d between them. These binary sequences correspond to the signal sequences $\underline{\mathbf{x}}$ and $\breve{\underline{\mathbf{x}}}$, respectively. The function $f(d, \Psi, \xi)$ can be computed from the pairwise error probability (PEP) $P(\underline{\mathbf{x}} \rightarrow \breve{\underline{\mathbf{x}}})$ by taking the average over all the possible sequences $\underline{\mathbf{x}}$ and $\breve{\underline{\mathbf{x}}}$ with respect to the label positions and the label mappings [10].

It was pointed out in [10] that averaging over all sequences $\underline{\mathbf{x}}$ and $\breve{\underline{\mathbf{x}}}$ is impractical due to the huge number of coded sequences. By neglecting the irrelevant error events and by using the expurgation technique based on the principle of "decomposability" of error sequences, the authors in [10] derived a new tight bound called BICM expurgated bound (BICM-EX bound) for BICM systems. The key idea in obtaining BICM-EX bound is that for each coded sequence $\underline{\mathbf{x}}$, only one relevant error event $\underline{\mathbf{x}} \rightarrow \breve{\underline{\mathbf{x}}}$, where $\breve{\underline{\mathbf{x}}}$ is the unique nearest neighbor of $\underline{\mathbf{x}}$, is considered [10]. This technique results in a computationally efficient form of the bound. It was also shown in [10] that the BICM-EX bound provides a very tight approximation to the PEP for different signal sets employing Gray or near Gray mappings. For other mappings, however, this bound might yield too optimistic results since there are many more nearest neighbors of each transmitted signal sequence.

By studying the asymptotic behavior of PEP, reference [10] also provides some insights into the asymptotic performance of BICM at high signal-to-noise ratio (SNR).

Over an AWGN channel, the performance of BICM is asymptotically dominated by the following free squared Euclidean distance (FED):

$$\text{FED} = d_H d_{\min}^2 \tag{2.8}$$

where $d_{\min}$ is the minimum Euclidean distance of the signal constellation. In the case of a Rayleigh fading channel with perfect channel state information (CSI), the factor that determines the asymptotic performance of a BICM system is the harmonic mean squared Euclidean distance between the complementary subsets of Ψ, which is defined as [10]:

$$d_h^2 = \left(\frac{1}{m2^m} \sum_{k=1}^{m} \sum_{\mathbf{s}_i \in \Psi_{\bar{b}}^k} \frac{1}{|\mathbf{s}_i - \hat{\mathbf{s}}_i|^2} \right)^{-1} \tag{2.9}$$

where $\bar{b}$ is the complement of b and $\hat{\mathbf{s}}_i \in \Psi_{\bar{b}}^k$ denotes the nearest neighbor of $\mathbf{s}_i$. More specifically, the asymptotic performance of BICM over a Rayleigh fading channel is approximated by [10]:

$$\log_{10} P_b \simeq \frac{-d_H}{10} \left[(R d_h^2)_{\text{dB}} + \left(\frac{E_b}{N_0} \right)_{\text{dB}} \right] + \text{const} \tag{2.10}$$

where P_b is the probability of bit error and R is the information rate.

Another relevant parameter that affects to the performance of BICM systems over both AWGN and Rayleigh fading channels is the average number of signals at the minimum Euclidean distance, defined as [2, 10]

$$N_{\min} = \frac{1}{m2^m} \sum_{\mathbf{s}_i \in \Psi} \sum_{k=1}^{m} N_{\min}(\mathbf{s}_i, k) \tag{2.11}$$

where $N_{\min}(\mathbf{s}_i, k)$ is the number of signal points at the Euclidean distance $d_{\min}$ whose label differs at position k compared to that of $\mathbf{s}_i$. The above parameter obviously depends on a specific mapping and it should be kept as small as possible.

Based on the error bound of BICM systems, it can be observed that for a fixed error control code and a fixed signal constellation, signal mapping plays an important role in determining the error performance of the systems. It was also suggested in [10] that the Gray and quasi-Gray mappings are optimal for BICM systems.

2.2.2 The Technique of Mutual Information

A brief discussion on the concepts of mutual information and channel capacity is provided in Appendix A. Here, these concepts shall be applied to study the effect of signal mapping in BICM systems.

Consider a two-dimensional M-ary constellation Ψ and a Rayleigh fading channel. The channel output $\mathbf{r}$ is given in (2.1). Then the channel capacity of a coded modulation system employing constellation Ψ can be computed as follows:

$$\begin{aligned} C &= E_g\left[\frac{1}{M}\sum_{i=1}^{M} I_g(\mathbf{s}_i;\mathbf{r})\right] \\ &= E_g\left[\frac{1}{M}\sum_{i=1}^{M}\int_{-\infty}^{+\infty} p(\mathbf{r}|\mathbf{s}_i,g)\log_2\frac{p(\mathbf{r}|\mathbf{s}_i,g)}{p(\mathbf{r}|g)}\mathrm{d}\mathbf{r}\right] \end{aligned} \tag{2.12}$$

where $\mathbf{s}_i = (s_{i1}, s_{i2})$ and $\mathbf{r} = (r_1, r_2)$ represent the transmitted and received signals respectively, $I_g(\mathbf{s}_i;\mathbf{r})$ is the mutual information between the transmitted signal $\mathbf{s}_i$ and the received signal $\mathbf{r}$ given the fading coefficient g. The function $p(\mathbf{r}|\mathbf{s}_i,g)$ is given in (2.3) and the function $p(\mathbf{r}|g)$ is the probability density function of $\mathbf{r}$ given g. Assume that the transmitted signals are equally likely, the function $p(\mathbf{r}|g)$ can be expressed as:

$$p(\mathbf{r}|g) = \frac{1}{M}\sum_{i=1}^{M} p(\mathbf{r}|\mathbf{s}_i,g) \tag{2.13}$$

Note also that the integral in (2.12) is a two-fold integral. The channel capacity in (2.12) is simply the symbol-wise mutual information of any two-dimensional M-ary modulation scheme. It is clear that for a fixed constellation, the symbol-wise mutual information is independent of the applied mapping. The symbol-wise mutual information can also be expressed as the sum of the bitwise mutual information as [16, 20]:

$$C = \sum_{L=0}^{m-1} E_g\left[I_{L|g}\right] = \sum_{L=0}^{m-1} I_L \tag{2.14}$$

where $I_{L|g}$ is the bitwise mutual information given g and under the condition that "L other bits in one symbol are known". Let $\{v_1, v_2, \ldots, v_m\}$ denote the m-bit label (or mapping) of one symbol in the constellation Ψ where $m = \log_2 M$. Then the bitwise

mutual information can be computed as:

$$E_g[I_{L|g}] = I_L = \overline{I_L(v_k; \mathbf{r}|L \text{ bits in } \{v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_m\} \text{ are known})} \quad (2.15)$$

The bar in the above equation means that the bitwise mutual information is averaged over all possible cases of "other L bits known", all bits in mapping as well as the bit value 1 or 0 for v_k. Note that, since conditioning increases mutual information, one has $I_0 \leq I_1 \leq \ldots \leq I_{m-1}$. Under the condition of ideal interleaver and uniform-input distribution, the channel capacity of BICM system over a Rayleigh fading channel can be computed as [10]:

$$C_{\text{BICM}} = mI_0 \quad (2.16)$$

The roles of other bitwise mutual information will be discussed later when BICM systems with iterative decoding are considered.

Observe that Equation (2.16) indicates the suboptimality of a BICM system compared to a general coded modulation system. On the other hand, (2.16) shows that, since the bitwise mutual information I_0 depends on the signal mapping, the signal mapping is an important factor to achieve good performance of BICM systems.

The mapping problem for BICM was recently addressed in [2] for various two-dimensional M-ary constellations by using the concept of channel capacity limit. It was shown in [2] that the ideal mappings for BICM are Gray mapping and quasi-Gray mapping. Such an observation is also consistent with the observation obtained with the error bound technique. The most suitable mappings of different 8-ary constellations for BICM systems found in [2] are shown in Fig. 2.4.

Finally, it should be mentioned here that reference [2] only studies the signal mappings of BICM systems over AWGN channels. We have also verified that these mappings are also the most suitable mappings for BICM systems over Rayleigh fading channels [21].

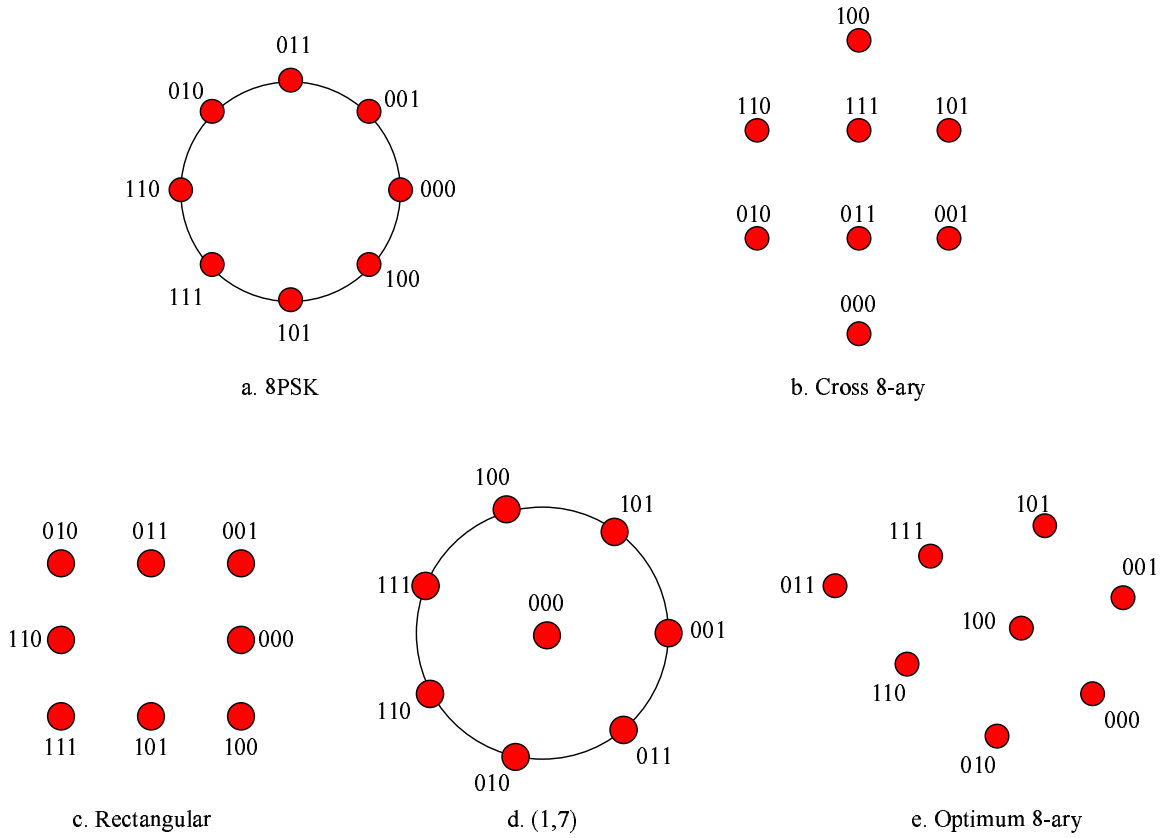

Figure 2.4 The best mappings in terms of channel capacity for different 8-ary constellations [2].

3. Bit-Interleaved Coded Modulation with Iterative Decoding (BICM-ID)

As demonstrated in Chapter 2, BICM performs well over fading channels by increasing the diversity order of the systems. However, due to "random modulation" caused by the bitwise interleaver, the free Euclidean distance of BICM is reduced. Recall that a large free Euclidean distance is the key for achieving good performance of a coded modulation system over an AWGN channel. This "random modulation" therefore results in the performance degradation of a BICM system compared to the conventional coded systems (i.e., TCM) over an AWGN channel [1, 10]. Even in the case of fading channels, when the fading is slow and interleaving is insufficient, the conventional TCM is still preferred.

In BICM, by using bitwise interleaving, the coded bits in one channel symbol are far apart to each other in the original coded sequence. With ideal interleaving (i.e., the interleaving depth is of infinite size and the interleaving pattern is completely random), the coded bits forming a channel symbol can be assumed to be independent. Obviously, the conventional decoder of a BICM system suggested in [1] does not fully exploit the advantages provided by bit interleaving and can therefore be improved.

Recently, the success of turbo codes has demonstrated the advantages of iterative processing in the decoding of concatenated coding schemes. Reference [12] appears to be the first paper that studies iterations between the demodulator and the decoder to overcome the drawbacks of conventional BICM systems. In that paper, a simple iterative decoding processing with a hard-decision feedback is proposed. It was pointed out in [12] that by carefully choosing the signal mapping, a large binary Hamming

distance between coded bits is indirectly translated into a large Euclidean distance. Using a simple decoder for convolutional code and the binary hard-decision feedback to the demodulator, it was shown in [12, 13] that BICM-ID with hard-decision feedback compares favorably with TCM over AWGN channels, while it significantly outperforms the latter over fading channels.

In iterative processing, the feedback from the section which is less affected by the channel noise removes the ambiguity in the high-order demodulation and enhances the decoding of the weak data sections. With perfect knowledge of the other $m-1$ bits, an M-ary constellation, $M = 2^m$, is translated to binary modulation selected from $M/2 = 2^{m-1}$ possible sets of binary constellations. It then follows that iterative decoding of BICM not only increases the intersubset Euclidean distance, but also reduces the number of nearest neighbors. This leads to a significant improvement over both AWGN and fading channels. Of course, if the feedback contains errors, wrong binary constellations are chosen and this leads to the degradation of the system performance. This also explains why it is important to control well the feedback and the error propagation.

By using the soft-decision feedback in the iterative processing, it has been shown in [14] that the performance of BICM-ID can be further improved. It was observed in [14] that the soft-decision feedback is the key to obtain coding gains in BICM while mitigating the error propagation. The results in [14] show that BICM-ID with soft-decision feedback significantly outperforms the conventional TCM and performs closely to the turbo-TCM [19] over both AWGN and Rayleigh fading channels.

This chapter is concerned with BICM-ID systems employing two-dimensional constellations. The system model with iterative processing at the receiver with two types of feedbacks mentioned before is first introduced. The effect of the interleaver will be briefly discussed. As in the previous chapter, here, the influence of the signal mapping to the error performance will also be discussed with two approaches. The first approach is based on the error bound of a BICM-ID system. The second one relies on the technique of bitwise mutual information. Finally, various good 8-ary

constellations and mappings for BICM-ID will be presented.

3.1 BICM-ID System Model

Figures 3.1 and 3.2 show the block diagram of BICM-ID systems using hard-decision and soft-decision feedback, respectively. Observe that the transmitters

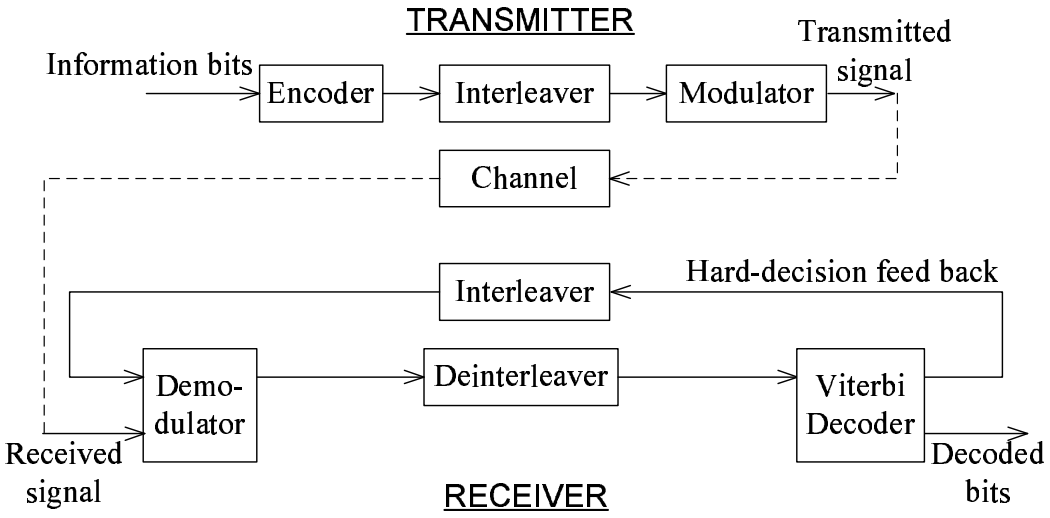

Figure 3.1 The block diagram of a BICM-ID system with hard-decision feedback.

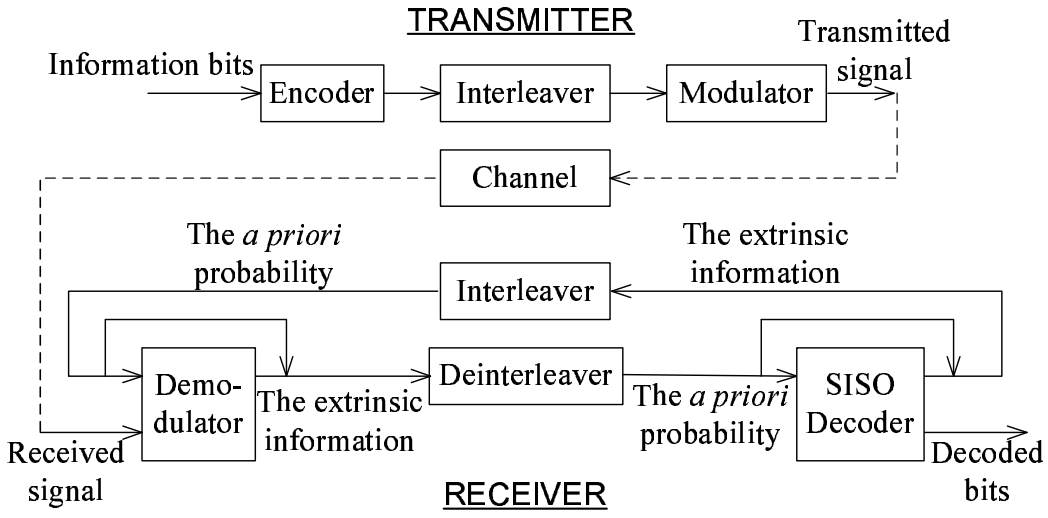

Figure 3.2 The block diagram of a BICM-ID system with soft-decision feedback.

of BICM-ID systems are similar to that of a BICM system, which are based on

the serial concatenation of a convolutional encoder, an interleaver and an M-ary modulator. The differences are in the receivers of these systems. In the case of systems with hard-decision feedback, a simple Virterbi decoder for convolutional code is used as in the conventional BICM systems. The hard-decision from the output of the Viterbi decoder is transferred back to the demodulator. On the other hand, the decoder of a BICM-ID system with soft-decision feedback is substituted by a more sophisticated decoder, namely the soft-input soft-output (SISO) decoder for convolutional code [22]. Here, the iterative processing between the demodulator and the SISO decoder transfers the *extrinsic information*, a term well explained in the literature of turbo codes [11, 22], of the coded bits to each other. The *extrinsic information* becomes the *a priori* information at the input of each component. The following discussion explains these two iterative receivers in more detail.

3.1.1 BICM-ID with Hard-Decision Feedback

Recall that in the conventional BICM systems, the bit metric for each coded bit is computed at the demodulator as in (2.2) based on the ML rule. This metric is essentially a simplification of the *a posteriori* probabilities in the *maximum a posteriori probability* (MAP) criterion [23], which can be computed as follows:

$$\lambda(v_k = b) = \log \sum_{\mathbf{s}_i \in \Psi_b^k} P\left(\mathbf{s}_i | \mathbf{r}, g\right) \tag{3.1}$$

In (3.1), $P\left(\mathbf{s}_i|\mathbf{r}, g\right)$ is the *a posteriori* probability of the transmitted signal $\mathbf{s}_i$ given the received signal $\mathbf{r}$ and the channel fading coefficient g. The set Ψ_b^k, $b \in \{0, 1\}$, denotes the subset of Ψ that contains all symbols whose labels have the value b at the kth position. Using Bayes rule [24], the *a posterior* probability $P\left(\mathbf{s}_i|\mathbf{r}, g\right)$ can be computed as:

$$P\left(\mathbf{s}_i|\mathbf{r}, g\right) = \frac{p(\mathbf{r}|\mathbf{s}_i, g)P(\mathbf{s}_i)}{p(\mathbf{r}|g)} \tag{3.2}$$

where $P(\mathbf{s}_i)$ is the *a priori* probability that the signal $\mathbf{s}_i$ is transmitted at the transmitter. The denominator of (3.2) can be expressed as:

$$p(\mathbf{r}|g) = \sum_{i=1}^{M} p(\mathbf{r}|\mathbf{s}_i, g)P(\mathbf{s}_i) \tag{3.3}$$

From (3.2) and (3.3), it can be observed that the computation of the *a posterior* probability $P(\mathbf{s}_i|\mathbf{r}, g)$ requires knowledge of the *a priori* probability $P(\mathbf{s}_i)$ and the conditional probability density function $p(\mathbf{r}|\mathbf{s}_i, g)$. Hence, the statistic for the bit metric $\lambda(v_k = b)$ can be written as follows:

$$\lambda(v_k = b) \sim \log \sum_{\mathbf{s}_i \in \Psi_b^k} p(\mathbf{r}|\mathbf{s}_i, g) P(\mathbf{s}_i) \tag{3.4}$$

Assume that the transmitted signals are equally likely, then the MAP rule becomes ML rule and the bit metrics in (3.4) and (2.2) are the same.

As mentioned in Chapter 1, the function of the channel encoder is to introduce some redundancy into the information sequence. This also creates the memory and redundancy in the signal sequence $\{\mathbf{s}_i\}$. The assumption of equally likely transmitted signals fails to use this information, mainly because it is too difficult to specify this information in advance for any decoding. However, the *a priori* information of the transmitted signals can be reflected from the output of the decoder and therefore can be included through the iterative processing.

The hard-decision feedback in [12] considers only binary-decision feedback for the calculation of the bit metrics. The following example illustrates the calculation of $\lambda(v_1 = 1)$ with an 8-ary signal constellation. The *a priori* information for any signal $\mathbf{s}_i$ with label $\mathbf{s}_i \sim \xi(v_1 = 1, v_2, v_3) \in \Psi_1^1$ is given by:

$$P(\mathbf{s}_i) = \begin{cases} 1, & \text{if } v_1 = 1, v_2 = \hat{v}_2, v_3 = \hat{v}_3 \\ 0, & \text{otherise} \end{cases} \tag{3.5}$$

where $\hat{v}_2$ and $\hat{v}_3$ are the results reflected from the previous decoding decisions. The bit metric is then computed as:

$$\lambda(v_1 = 1) = -||\mathbf{r} - g\mathbf{s}_i(1, \hat{v}_2, \hat{v}_3)||^2 \tag{3.6}$$

The bit metrics for different values of b and k can be computed similarly.

Given $(m-1)$ values of the feedback coded bits, an M-ary constellation with mapping ξ is translated into a binary modulation selected from a set of $M/2$ possible

binary constellations. It was shown in [12] that with an appropriate signal mapping, the minimum Euclidean distance between two coded sequences can be large for BICM-ID. It makes the BICM-ID suitable for both Rayleigh fading and AWGN channels.

3.1.2 BICM-ID with Soft-Decision Feedback

In BICM-ID systems with soft-decision feedback, instead of a standard Viterbi decoder, a more advanced soft feedback method is used to reduce the error propagation. As shown in Fig. 3.2, the receiver of the system uses a suboptimal, iterative method with soft-output demodulator and the SISO decoder. Though the two components of the receiver are individually optimal, the demodulation of the received signal and the decoding of the convolutional code act separately.

In the first iteration, it is reasonable to assume that the transmitted signals $\{\mathbf{s}_i\}$ are equally likely. The *a posteriori* probabilities of the coded bits can be calculated similarly to the case of BICM as well as BICM-ID systems with hard-decision feedback as in (2.2) and (3.4). This information is passed to the SISO decoder [22] which is used for decoding the convolutional code. Based on the information from the demodulator and the trellis structure of the convolutional code, the SISO module generates the *a posteriori* probabilities for the information and the coded bits.

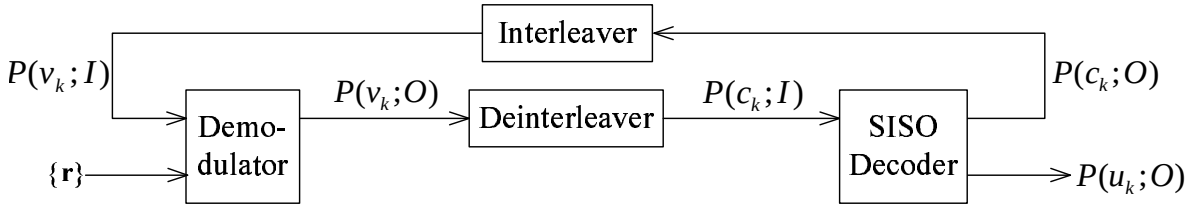

Figure 3.3 Block diagram of the receiver for a BICM-ID system with soft-decision feedback.

Figure 3.3 shows the block diagram of the receiver for a BICM-ID system with soft-decision feedback. Using the same notation as in [14, 22], the *a priori* probability and the *a posteriori* probability of a random variable q are denoted by $P(q;I)$ and $P(q;O)$, respectively. It should be mentioned here that $P(q;O)$ is often referred to

as the *extrinsic information* in turbo code literature [11]. From the second iteration, the *extrinsic information* $P(c_k; O)$ of the coded bits is transferred back to the demodulator. After being interleaved, it becomes the *a priori* information $P(v_k; I)$ at the input of the demodulator. Since the (ideal) interleaver makes m bits in one symbol independent, the *a priori* information $P(\mathbf{s}_i)$ of each signal $\mathbf{s}_i \in \Psi$ can be obtained as [14]:

$$P(\mathbf{s}_i) = P\left(v_1(\mathbf{s}_i), \ldots, v_m(\mathbf{s}_i)\right) = \prod_{k=1}^{m} P(v_k = v_k(\mathbf{s}_i); I) \tag{3.7}$$

where $v_k(\mathbf{s}_i) \in \{0, 1\}$, $1 \leq k \leq m$ is the value of the kth bit in the label of $\mathbf{s}_i$. Clearly, the calculation of $P(\mathbf{s}_i)$ in the case of hard-decision feedback in (3.5) is a special case of (3.7). Using (3.7) and (3.4), the *extrinsic information* from the second iteration can be obtained as [14]:

$$\begin{aligned} P(v_k = b; O) &= \frac{P(v_k = b|\mathbf{r})}{P(v_k = b; I)} = \frac{\left(\sum_{\mathbf{s}_i \in \Psi_b^k} P(\mathbf{r}|\mathbf{s}_i) P(\mathbf{s}_i)\right)}{P(v_k = b; I)} \\ &= \sum_{\mathbf{s}_i \in \Psi_b^k} \left(P(\mathbf{r}|\mathbf{s}_i) \prod_{i \neq k} P(v_j = v_j(\mathbf{s}_i); I) \right) \end{aligned} \tag{3.8}$$

It can be observed from (3.8) that the bit metric $P(v_k = b; O)$ is computed from the *a priori* probabilities $P(v_j; I)$ of the other bits ($i \neq k$) on the same channel symbol. The regenerated bit metrics are then delivered to the SISO decoder. The SISO decoder treats that information as the *a priori* information for the coded bits and the iterative demodulation and decoding keep running. The hard-decision for the information bits is made at the final iteration based on the *extrinsic information* $P(u_j; O)$. This information can also be understood as the *total a posteriori* probabilities of the information bits, since the *a priori* probabilities $P(u_j; I)$ are not available in the iterative processing [14].

3.2 The Effects of the Interleaver

The interleaver is critical to the high performance of BICM-ID systems. This section briefly discusses some aspects in the design of a good interelaver for BICM-ID systems.

The key idea in the design of a good interleaver for a BICM-ID system is to make the interleaved coded bits in the same channel symbol as far apart as possible. More specifically, as pointed out in [14], two design objectives of a good interleaver are as follows:

(i) To increase the minimum Euclidean distance between any two coded sequences

(ii) To mitigate the error propagation during the iterative decoding.

Some design rules are also suggested in [14] to obtain a good interleaver. These rules are quite similar to those for spread-random interleavers studied in [25] and given below:

- *Modularity*: The bit position before and after interleaving must have the same modulo-m value. This ensures that the coded bits with different protection, due to their different positions in the channel symbol labels, are distributed uniformly along the trellis.

- *Reverse Spread*: The m bits going to the same channel symbol should be spread as far apart as possible from each other. This ensures feedback independence in the bit metric recalculation and it mitigates the error propagation.

Another important parameter for the error performance of a BICM-ID system is the block length of the interleaver. It is well known for many iterative decoding schemes that the longer the interleaver, the better the error performance becomes. This is also the case for BICM-ID systems. A small block size might cause a substantial performance degradation to BICM-ID. It should be noted that increasing the block length leads to earlier convergence, in terms of both the SNR and the number of iterations. The reason for this is the reduction in the number of nearest neighbor codewords when the interleaver size is increased. However, increasing the interleaver size does not improve the performance of BICM-ID systems after a certain high SNR

level (as normally seen in standard turbo codes). This is due to the fact that the performance at a high SNR mainly depends on d_{free} of the whole code and this parameter is constant for most interleavers.

The question on whether there exists an ideal interleaver with a finite block length for BICM-ID still remains unanswered.

3.3 The Effect of Signal Mapping

Similar to the case of BICM systems, this section considers the effect of signal mapping on the error performance of BICM-ID systems. Since only soft-decision feedback is applied in this thesis, such a consideration concentrates on BICM-ID systems with soft-decision feedback. The effect of signal mapping to BICM-ID with hard-decision feedback is studied in [13]. Two different approaches, namely the error bound and the mutual information, will be taken into account.

3.3.1 The Upper Bound on the Error Performance of a BICM-ID System

Due to the large coding gain produced by iterative processing, one is most interested in the asymptotic performance to which the iterative processing converges. Such an asymptotic performance of the systems under consideration can be analyzed with the error-free feedback bound (EF bound) introduced in [14, 15]. The EF bound for the case of a Rayleigh fading channel is essentially the BICM expurgated bound [10] introduced in Chapter 2 modified for a BICM-ID system by assuming the error-free feedback. Over an AWGN channel, references [14] and [15] compute the error bound based on the Euclidean distances FED and FEDC. However, it was shown in [26] that these distances sometimes fail to discriminate the differences in error performance of different mappings. In this section, an upper bound on the BER performance of a BICM-ID system using the techniques in [27, 28] is derived. From the upper bound, a distance criterion is identified to find the best mappings in terms of asymptotic performance over AWGN and Rayleigh fading channels. For Rayleigh fading channels,

the result is similar to the one derived in [14,15]. More importantly, the derivations in this section will also be extended to BICM-ID systems with hypercube constellations proposed in Chapter 4.

The union bound of the BER , P_b, for a BICM-ID system can also be written similarly to that of a BICM system in (2.7). Using the same notation as in Chapter 2, let $\underline{\mathbf{c}}$ and $\breve{\underline{\mathbf{c}}}$ denote the input sequence and the estimated sequence with the Hamming distance d between them. These binary sequences correspond to the symbol sequences $\underline{\mathbf{x}}$ and $\breve{\underline{\mathbf{x}}}$, each consisting of d signal symbols. In what follows, the function $f(d, \Psi, \xi)$ is computed from the PEP $P(\underline{\mathbf{x}} \rightarrow \breve{\underline{\mathbf{x}}})$ under the assumption that the iterative processing works perfectly.

Without loss of generality, assume that $\underline{\mathbf{c}}$ and $\breve{\underline{\mathbf{c}}}$ differ in the first d consecutive bits. Hence, $\underline{\mathbf{x}}$ and $\breve{\underline{\mathbf{x}}}$ can be redefined as sequences of d M-ary signals as follows:

$$\underline{\mathbf{x}} = [\mathbf{x}_1, \ldots, \mathbf{x}_d] \tag{3.9}$$

$$\breve{\underline{\mathbf{x}}} = [\breve{\mathbf{x}}_1, \ldots, \breve{\mathbf{x}}_d] \tag{3.10}$$

Also let

$$\underline{\mathbf{g}} = [g_1, g_2, \ldots, g_d] \tag{3.11}$$

where g_e, $1 \leq e \leq d$, represents the Rayleigh fading coefficient corresponding to the eth M-ary symbol. Then the PEP conditioned on $\underline{\mathbf{g}}$ can be computed as[1]

$$P(\underline{\mathbf{x}} \rightarrow \breve{\underline{\mathbf{x}}} | \underline{\mathbf{g}}) = Q\left(\sqrt{\frac{1}{2N_0}\sum_{e=1}^{d} g_e^2 \|\mathbf{x}_e - \breve{\mathbf{x}}_e\|^2}\right) \tag{3.12}$$

It should be mentioned here that with the ideal interleaver, g_e, $1 \leq e \leq d$, can be assumed to be independent and identical distributed (i.i.d) random variables [10]. In the case of AWGN channels, all the fading coefficients $\{g_e\}$ equal 1. The PEP can then be obtained by removing the condition on $\underline{\mathbf{g}}$ in (3.12) as follows.

[1]The Q-function is defined as $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \mathrm{e}^{-t^2/2} \mathrm{d}t$.

First, for Rayleigh fading channels, by using the following Gaussian probability integral:

$$Q(\gamma) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\frac{\gamma^2}{2\sin^2\theta}\right) \mathrm{d}\theta \tag{3.13}$$

and averaging (3.12) over the sequence $\underline{\mathbf{g}}$ of d i.i.d Rayleigh random variables, it can be shown that [27, 28]:

$$P(\underline{\mathbf{x}} \to \breve{\underline{\mathbf{x}}}) \;=\; E_{\underline{\mathbf{g}}}\left\{P(\underline{\mathbf{x}} \to \breve{\underline{\mathbf{x}}}|\underline{\mathbf{g}})\right\} = \frac{1}{\pi}\int_0^{\pi}\left(\prod_{e=1}^{d}\Delta_e\right)\mathrm{d}\theta \tag{3.14}$$

where

$$\Delta_e = \left(1 + \frac{1}{4N_0}\frac{\|\mathbf{x}_e - \breve{\mathbf{x}}_e\|^2}{\sin^2\theta}\right)^{-1} \tag{3.15}$$

Note that the term inside the brackets of (3.15) is obtained by invoking the following equality for a Rayleigh random variable g:

$$E_g\{\exp(-\gamma g^2)\} = 1/(1+\gamma) \tag{3.16}$$

Owning to the success of decoding steps (i.e., the iterations between the SISO decoder and the demodulator are assumed to work perfectly), one needs to consider the two signal symbols $\mathbf{x}_e$ and $\breve{\mathbf{x}}_e$ whose labels differ in only one bit. Also observe that Δ_e are i.i.d. random variables. Hence, the function $f(d, \Psi, \xi)$ in (2.7) for BICM-ID systems can be obtained by averaging over all the signal symbols of the M-ary constellation Ψ [10, 27].

Consider a single signal symbol $\mathbf{s}_i \in \Psi$ and let $\mathbf{s}_{j(i,k)}$ denote the signal whose label differs at position k compared to the label of $\mathbf{s}_i$. Thus $j(i,k)$ is simply understood as an integer index of a signal symbol and it depends on i, k and the specific mapping ξ. Now taking the average over all the symbol $\mathbf{s}_i$ and the corresponding symbol $\mathbf{s}_{j(i,k)}$ in the M-ary constellation Ψ gives:

$$f(d, \Psi, \xi) \le \frac{1}{\pi}\int_0^{\pi/2}\left[E\left\{\left(1 + \frac{1}{4N_0}\cdot\frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{\sin^2\theta}\right)^{-1}\right\}\right]^d \mathrm{d}\theta \tag{3.17}$$

where

$$E\left\{\left(1 + \frac{1}{4N_0}\frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{\sin^2\theta}\right)^{-1}\right\} = \frac{1}{m2^m}\sum_{\mathbf{s}_i\in\Psi}\sum_{k=1}^{m}\left[\left(1 + \frac{1}{4N_0}\frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{\sin^2\theta}\right)^{-1}\right] \tag{3.18}$$

Although (3.17) can be efficiently computed with a single integral and it gives an accurate approximation for the asymptotic performance (i.e., the error floor) of the system, it does not give an insight on how to design good mappings for a given constellation. To understand the influence of the constellation mapping to the asymptotic behavior of the PEP, the following inequality is useful:

$$Q(\gamma) < \frac{1}{2}\exp(-\frac{\gamma^2}{2}) \tag{3.19}$$

By substituting (3.19) into (3.12) and taking the average over $\underline{\mathbf{g}}$, one obtains:

$$P(\underline{\mathbf{x}} \to \breve{\underline{\mathbf{x}}}) \;<\; \frac{1}{2}\prod_{e=1}^{d}\left(1+\frac{\|\mathbf{x}_e-\breve{\mathbf{x}}_e\|^2}{4N_0}\right)^{-1} \tag{3.20}$$

As before, taking the average over the all symbols $\mathbf{s}_i$ and the corresponding symbols $\mathbf{s}_{j(i,k)}$, the function $f(d,\Psi,\xi)$ can be approximated at high signal-to-noise ratio (SNR) as:

$$f(d,\Psi,\xi) \;\approx\; \frac{1}{2}\left[\delta_1(\Psi,\xi)\right]^d \tag{3.21}$$

where the parameter $\delta_1(\Psi,\xi)$ depends on the constellation and mapping, and is given by:

$$\delta_1(\Psi,\xi) = \frac{1}{m2^m}\sum_{\mathbf{s}_i\in\Psi}\sum_{k=1}^{m}\left(1+\frac{\|\mathbf{s}_i-\mathbf{s}_{j(i,k)}\|^2}{4N_0}\right)^{-1} \tag{3.22}$$

Since $\|\mathbf{s}_i-\mathbf{s}_{j(i,k)}\|^2 > 0$, we use the following approximation when $N_0 \to 0$:

$$\left(1+\frac{\|\mathbf{s}_i-\mathbf{s}_{j(i,k)}\|^2}{4N_0}\right)^{-1} \approx \frac{4N_0}{\|\mathbf{s}_i-\mathbf{s}_{j(i,k)}\|^2} \tag{3.23}$$

Thus, $f(d,\Psi,\xi)$ can be approximated as

$$f(d,\Psi,\xi) \approx \frac{1}{2}\left[4N_0\cdot\tilde{\delta}_1(\Psi,\xi)\right]^d \tag{3.24}$$

where

$$\tilde{\delta}_1(\Psi,\xi) = \left(\frac{1}{m2^m}\sum_{\mathbf{s}_i\in\Psi}\sum_{k=1}^{m}\frac{1}{\|\mathbf{s}_i-\mathbf{s}_{j(i,k)}\|^2}\right) \tag{3.25}$$

Note that $\tilde{\delta}_1(\Psi,\xi)^{-1}$ is exactly the harmonic mean Euclidean distance with ideal feedback defined in [14] for a BICM-ID system over a Rayleigh fading channel.

For the case of an AWGN channel, the probability $P(\underline{\mathbf{x}} \to \breve{\underline{\mathbf{x}}})$ is simply obtained as in (3.12) with all fading coefficients set to 1. It then follows that

$$f(d, \Psi, \xi) \leq \frac{1}{\pi} \int_0^{\pi/2} \left[E\left\{ \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{4N_0 \sin^2\theta} \right) \right\} \right]^d \mathrm{d}\theta \tag{3.26}$$

where

$$E\left\{ \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{4N_0 \sin^2\theta} \right) \right\} = \frac{1}{m2^m} \sum_{\mathbf{s}_i \in \Psi} \sum_{k=1}^{m} \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{4N_0 \sin^2\theta} \right) \tag{3.27}$$

Again, the function $f(d, \Psi, \xi)$ at high SNR can be approximated as

$$f(d, \Psi, \xi) \approx \frac{1}{2} \left[\delta_2(\Psi, \xi) \right]^d \tag{3.28}$$

where

$$\delta_2(\Psi, \xi) = \frac{1}{m2^m} \sum_{\mathbf{s}_i \in \Psi} \sum_{k=1}^{m} \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{4N_0} \right) \tag{3.29}$$

Obviously, the two parameters $\delta_1(\Psi, \xi)$, $\delta_2(\Psi, \xi)$ derived above characterize the influence of the signal mapping to the asymptotic performance of a BICM-ID system over the two channel models, respectively. Specifically, the smaller these parameters are, the lower the asymptotic BER performance of the systems becomes. The best mappings with respect to the asymptotic performance of a BICM-ID system over Rayleigh and AWGN channels are the ones that have the smallest $\delta_1(\Psi, \xi)$ and $\delta_2(\Psi, \xi)$, respectively.

3.3.2 The Technique of Mutual Information

A different technique based on mutual information was introduced in [16, 18] to explain the influence of signal mapping on the performance of BICM-ID. As mentioned in Chapter 2, the symbol wise mutual information C of a coded modulation scheme employing an M-ary constellation can be expressed as the sum of the bitwise mutual information I_L, $0 \leq L \leq (m-1)$ in (2.15). Furthermore, for a fixed constellation, C is independent of the applied mapping.

Using the concept of equivalent channels in [29], a channel accepting an M-ary signal symbol can be seen as consisting of m sub-channels, each carries one bit with a

different condition on "the other bits being known". For example, the sub-channel L carries one bit with the condition that the other $m-1-L$ bits are perfectly known. Thus the capacity of this channel is the bitwise mutual information I_{m-1-L}. Though, for a fixed constellation, $C = \sum_{L=0}^{m-1} I_L$ is always a constant regardless of the signal mapping, the bitwise mutual information I_L depends on the mapping.

Since the calculation of the bitwise mutual information I_0 assumes that there is no information from the other bits, I_0 is an important parameter for the performance of BICM-ID with no iteration, i.e., the performance of a BICM system studied in Chapter 2. If the iteration between the channel decoder and the demodulator works, then a more and more accurate information about the other bits in one symbol will be transferred to the next iteration [18]. Therefore, with iterations, I_0 is not very important any more. Instead, $I_1, I_2, \ldots, I_{m-1}$ become dominant parameters to the error performance of BICM-ID systems when the number of iterations becomes larger. Finally, if the perfect a priori knowledge of all other bits is available, only I_{m-1} dominates the performance of BICM-ID. Equivalently, it can be seen that I_{m-1} is the key parameter in determining the asymptotic performance of BICM-ID systems. The exact expression for the bitwise mutual information with the perfect knowledge of the other bits in one channel symbol is provided in Appendix B.

Since the sum of bitwise mutual information is always a constant for a fixed constellation, there is a trade-off in the values of bitwise mutual information among different mappings. To have a good performance with no iteration, one should apply the mapping that has large I_0. However, large value of I_0 means that the other mutual information with a priori knowledge will be smaller, implying that iteration may not helpful. A compromise solution is to design the mapping with I_0 big enough to make iteration work, while maximizing I_{m-1} to achieve good error performance after a large number of iterations.

It should be pointed out that the bitwise mutual information depends on the particular value of the signal-to-noise ratio (SNR), i.e., the channel condition. Thus the bitwise mutual information can be used to evaluate and compare different constella-

tions/mappings of a BICM-ID system operating at a specific range of SNR. This is the major advantage of mutual information compared to the error bound discussed in the previous section.

3.4 Proposed Mappings of 8-ary Constellations for BICM-ID

Based on the theories introduced in the previous sections, this section presents good mappings of various 8-ary constellations for BICM-ID systems operating over frequency non-selective block Rayleigh fading channels.

In [14], different mappings for an 8-PSK constellation were studied. It was shown that a BICM-ID system using semi-set partitioning (SSP) mapping shown in Fig. 3.4–(a) offers the best performance over other popular 8-PSK mappings provided that a large enough block length of the interleaver is used.

Here, three more popular constellations, including (1,7), cross 8-ary, and the optimum 8-ary constellations, are also considered. They are shown in Fig. 3.4–(b), (c), and (d), respectively. The exact locations of the signal points in the optimum 8-ary constellation are provided in [30]. This constellation is optimum in the sense that it performs optimally at high SNR for uncoded systems over an AWGN channel.

3.4.1 The Proposed Mappings

For each constellation, the mapping of interest to us is the one that minimizes $\tilde{\delta}_1(\Psi, \xi)$ computed as in (3.25). Such mappings are found for the above constellations by computer search and they are also presented in Figs. 3.4–(b), (c), and (d). Table 3.1 lists the parameter $\tilde{\delta}_1(\Psi, \xi)^{-1}$ of the proposed mappings where it is assumed that the average symbol energy is normalized to be one. Also shown in Table 3.1 is the harmonic mean distance d_h^2 defined in (2.9). Recall that this parameter affects the asymptotic performance of BICM systems (i.e., the performance of the first iteration of the BICM-ID systems).

It is interesting to observe that d_h^2 of SSP 8-PSK is smallest among four constel-

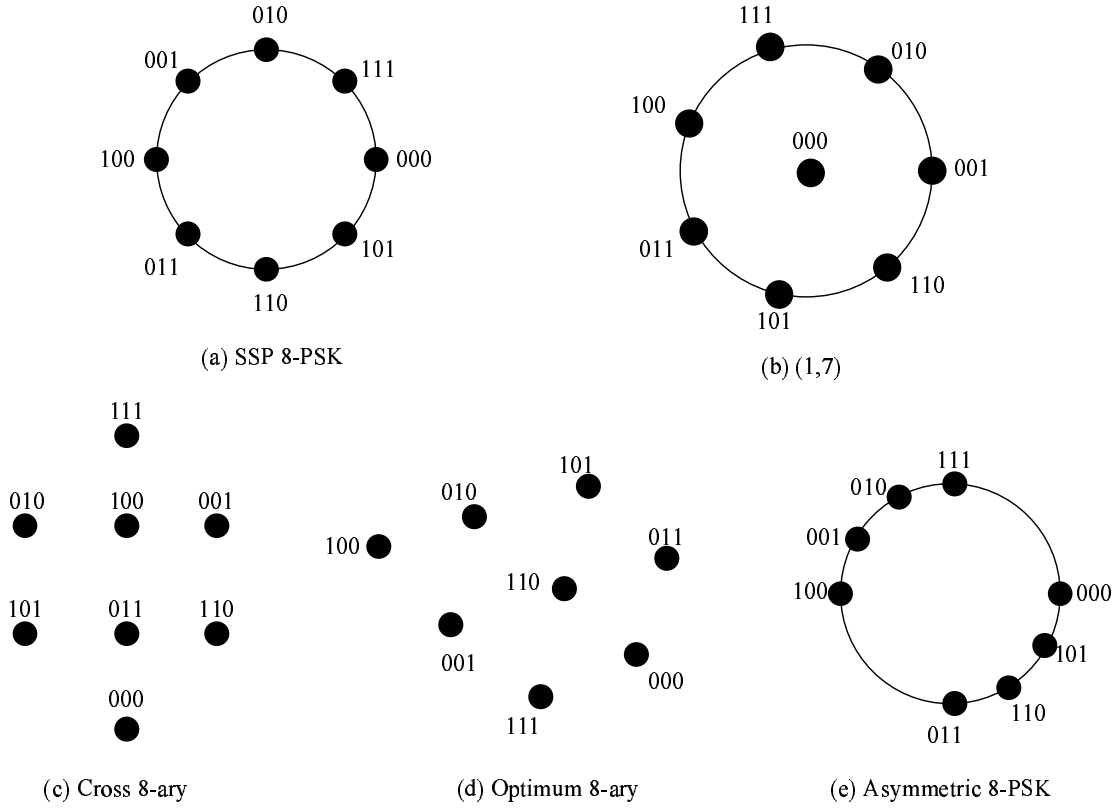

Figure 3.4 The best mappings in terms of asymptotic performance of various 8-ary constellations for BICM-ID systems over a Rayleigh fading channel.

Table 3.1 The parameters $\tilde{\delta}_1(\Psi,\xi)^{-1}$ and d_h^2 for the proposed mappings (minimized-$\tilde{\delta}_1(\Psi,\xi)$ mappings).

	Constellation	$\tilde{\delta}_1(\Psi,\xi)^{-1}$	d_h^2
(a)	SSP 8-PSK	2.8766	0.5858
(b)	(1,7)	2.3045	0.8880
(c)	Cross 8-ary	2.5946	0.8348
(d)	Optimum 8-ary	2.3420	0.9163
(e)	Asymmetric 8-PSK	3.5241	0.3052

lations/mappings in Figs. 3.4–(a), (b), (c), and (d). This suggests that if iterations work well with SSP 8-PSK, they also work well with other contellations/mappings.

Even with a short interleaver length which makes the performance of a BICM-ID system with SSP 8-PSK mapping degraded, the BICM-ID systems employing other constellations/mappings might still deliver good performance. In contrast to d_h^2, SSP 8-PSK has the biggest value of $\tilde{\delta}_1(\Psi,\xi)^{-1}$. This means that the asymptotic performance of a BICM-ID system using SSP 8-PSK mapping outperforms the systems using other constellations/mappings. The above discussion is confirmed by analytical and simulation results in Section 3.4.3.

As far as the asymptotic performance is concerned, it is also interesting to find a constellation/mapping that performs better than SSP 8-PSK, i.e., having a larger $\tilde{\delta}_1(\Psi,\xi)^{-1}$ than that of SSP 8−PSK. This task can be accomplished by a simple modification of SSP 8-PSK as follows. Observe that the signal points in SSP 8-PSK are divided into even and odd subsets. The even subset includes all the signal points whose labels have even Hamming weights. Similarly, the odd subset includes all the signal points whose labels have odd Hamming weights. Now if the signal points in each subset move closer to each other, $\tilde{\delta}_1(\Psi,\xi)^{-1}$ will increase. It is easy to verify that when all four signals in each subset collapse to one point, the parameter $\tilde{\delta}_1(\Psi,\xi)^{-1}$ is the biggest. However, the use of this ambiguity constellation/mapping will result in a very poor error performance at any practical SNR level due to the large value of $N_{\min}$ and the fact that the minimum Euclidean distance between the signal points is 0.

A compromise solution proposed here is a constellation/mapping shown in Fig. 3.4-(e), which shall be referred to as an asymmetric 8-PSK constellation. In each odd or even subset of this constellation/mapping, the angle between any two nearest signal points is $\pi/6$. The harmonic mean d_h^2 and $\tilde{\delta}_1(\Psi,\xi)^{-1}$ of this constellation/mapping are shown in Table 3.1. Observe that the parameter $\tilde{\delta}_1(\Psi,\xi)^{-1}$ of the asymmetric 8-PSK constellation/mapping is much higher than that of SSP 8-PSK and, therefore, it is expected to have a very good asymptotic performance. However, it should be mentioned that due to the small value of d_h^2, it might need a long interleaver and a large number of iterations to achieve the asymptotic performance of this asymmetric

8-PSK constellation/mapping. Again, this prediction is also confirmed by analytical and simulation results in Section 3.4.3.

3.4.2 A Study of the Proposed Mappings by Mutual Information

As discussed before, for a BICM-ID system with 8-ary constellation, I_0 is a dominant parameter for the performance after the first iteration. From the second iteration, if I_0 is big enough to make the iteration work, then I_1 and I_2 become dominant parameters. Finally, I_2 is the most important parameter for the asymptotic performance of a BICM-ID system.

With any 8-ary constellation and mapping, the bitwise mutual information can be calculated for different mappings. Especially, I_2 can be computed with a high accuracy using a series expansion as provided in Appendix B. Starting with the 8-PSK constellation, it can be seen that SSP mapping is the mapping with maximized I_2. It is also easy to verify that for given constellations, the mappings introduced in the previous section are the mappings that maximize I_2. With a fixed convolutional code, the length of the interleaver affects the convergence of the error performance of a BICM-ID system. In general, a longer interleaver leads to an earlier convergence. This also implies that, with a shorter interleaver and if only a few iterations can be used, the one with bigger I_2 might not be preferred for BICM-ID. The bitwise mutual information I_2 for the I_2-maximized mappings of the five different constellations are plotted in Fig. 3.5. Due to the arrangements of the signal points, I_2-maximized asymmetric mapping has the biggest I_2 among all the I_2-maximized mappings. This observation is consistent with the fact that this special mapping of 8-PSK has the largest value of $\tilde{\delta}_1(\Psi, \xi)^{-1}$. The $(1, 7)$ and optimum 8-ary constellations/mappings have almost the same value of I_2, which is also smallest. There is a small difference in I_2 between SSP 8-PSK and I_2-maximized cross 8-ary mapping and it is clear that SSP 8-PSK is preferred.

Finally, Fig. 3.6 shows the bitwise mutual information where it is assumed that

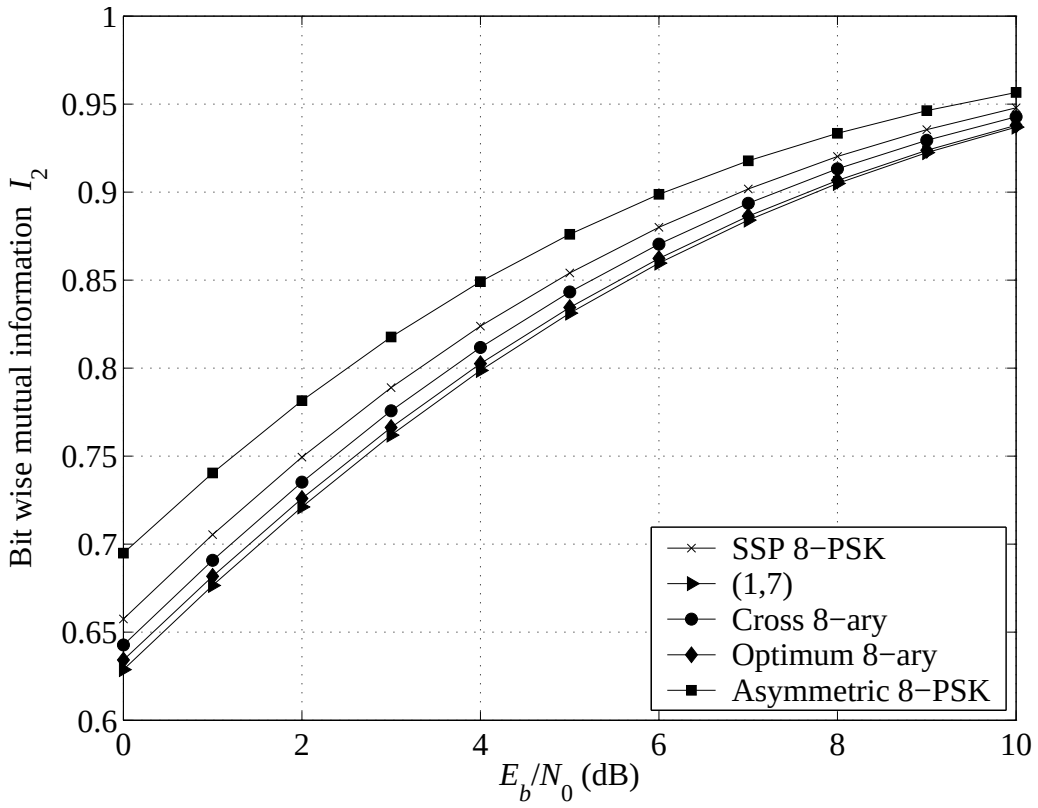

Figure 3.5 Bitwise mutual information with the perfect knowledge of the other two bits (I_2) of various 8-ary constellations/mappings: The channel model is Rayleigh fading and the code rate is 2/3.

there is no information from the other bits, I_0, for the above mappings. It can be seen that in contrast to I_2, the asymmetric 8-PSK mapping has the lowest I_0. The values of I_0 are almost the same for the cross 8-ary, optimum 8-ary and $(1, 7)$ mappings. This suggests that when using the same interleaver, especially a short interleaver, these three mappings are more suitable than SSP 8-PSK and the asymmetric 8-PSK. This expectation is again confirmed by the analytical and simulation results in the next section.

3.4.3 Analytical and Simulation Results

This subsection provides the analytical and simulation results to verify the performance of the mapping designs introduced in Section 3.4. As the investigation concentrates on signal mapping, a rate-2/3, 8-state convolutional code (CC) with the

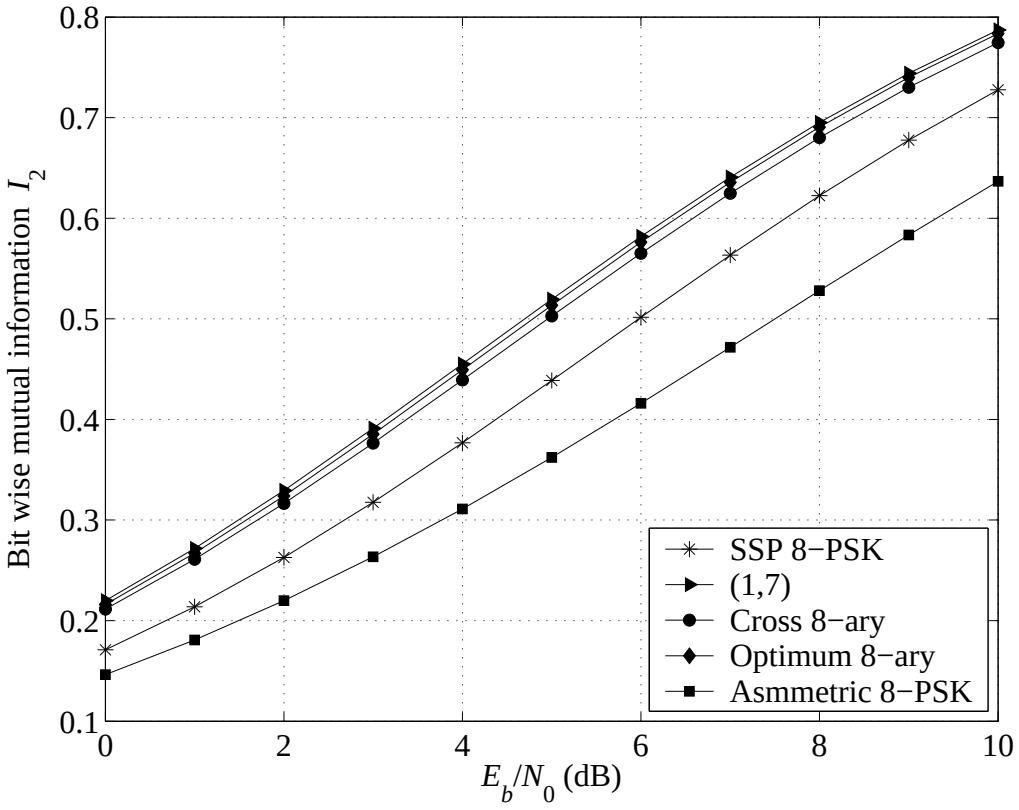

Figure 3.6 Bitwise mutual information I_0 for maximized-I_2 mappings.

generator sequences $\mathbf{g}_1 = (4, 2, 6)$ and $\mathbf{g}_2 = (1, 4, 7)$ is always assumed. The bitwise interleavers used in all simulations are designed according to the rules outlined in Section 3.2. Each point of the BER curves is obtained by simulating the systems with 10^7 to 10^8 coded bits. Also observe from Table 3.1 that the cross 8-ary, $(1, 7)$ and optimum 8-ary constellations/mappings have almost the same values of d_h^2 and $\tilde{\delta}_1(\Psi, \xi)^{-1}$. Thus it can be expected that these constellations/mappings perform very closely.

First, a long interleaver with a length of $12,000$ coded bits is used. Figures 3.7, 3.8 and 3.9 present the performances with 1 to 8 iterations of the BICM-ID systems employing the asymmetric 8-PSK, SSP 8-PSK, cross 8-ary and the proposed mappings (see Fig. 3.4), respectively. Also shown in each figure is the asymptotic performance calculated from (2.7) and (3.17) using the first 20 Hamming distances of the convolutional code. It is clear from these figures that iterations work well with all

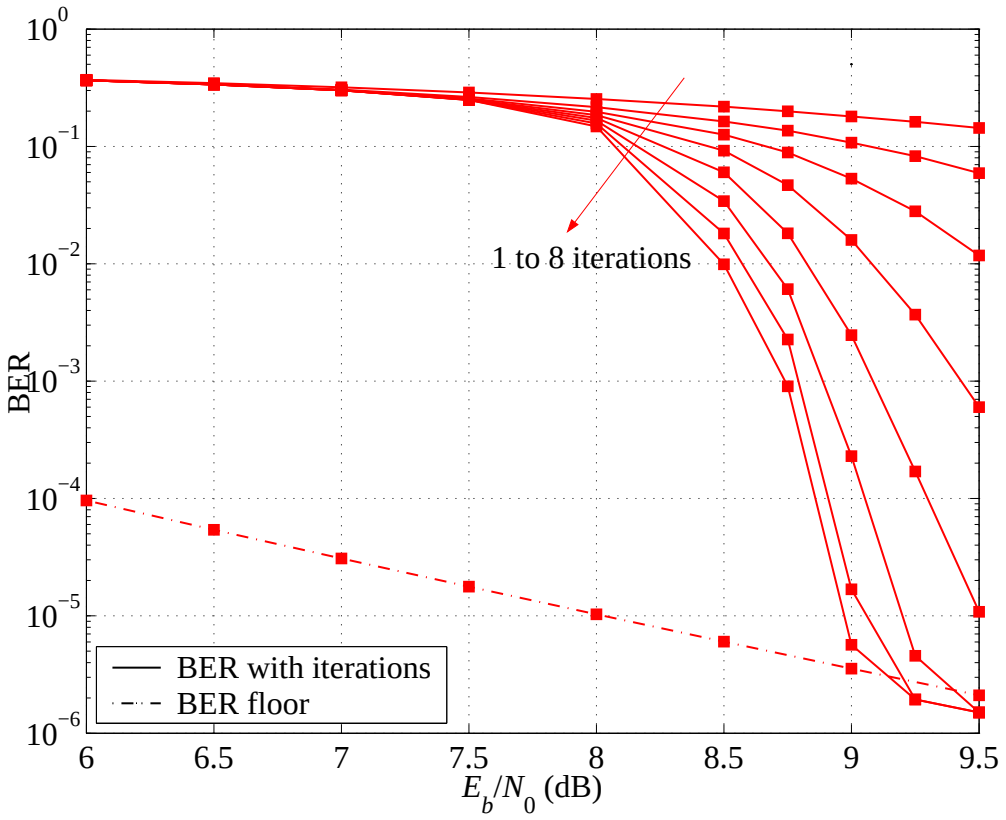

Figure 3.7 BER performance of a BICM-ID system: Asymmetric 8-PSK/proposed mapping, rate-2/3, 8-state CC, 12,000-bit interleaver, 8 iterations.

the constellations/mappings and they also converge to the asymptotic performances. Note, however, that the convergence behavior of each constellation/mapping is very different. This is of course due to the difference in the parameters d_h^2 and $\tilde{\delta}_1(\Psi,\xi)^{-1}$ and can be summarized as follows. For a constellation/mapping with a small d_h^2 and a large $\tilde{\delta}_1(\Psi,\xi)^{-1}$ (such as the asymmetric 8-PSK), it requires a higher value of SNR for the iteration to work, but it converges to a lower asymptotic performance. For a constellation/mapping with a large d_h^2 and a small $\tilde{\delta}_1(\Psi,\xi)^{-1}$ (such as the cross 8-ary), it is the other way around: the iterations start working at a lower SNR region but it converges to a higher asymptotic performance.

Figure 3.10 compares the BER performance of different constellations/mappings after 8 iterations. Due to having a bigger value of $\tilde{\delta}_1(\Psi,\xi)^{-1}$, the asymptotic performance of the asymmetric 8-PSK constellation is better than that of both the cross

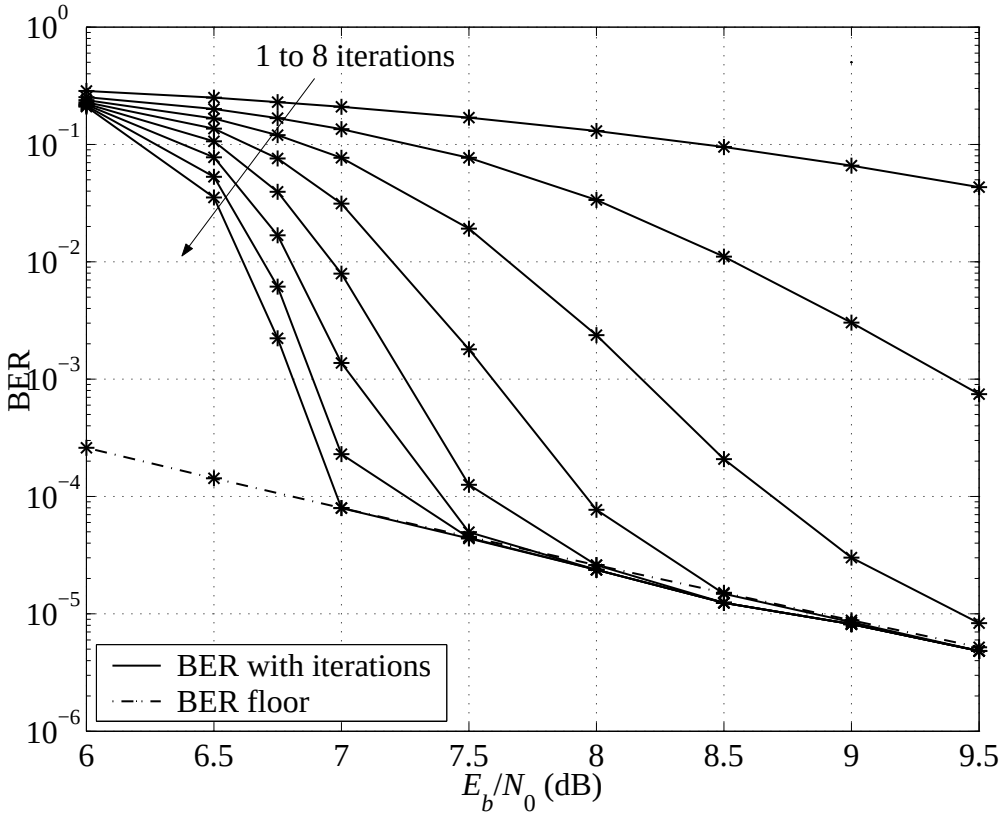

Figure 3.8 BER performance of a BICM-ID system: SSP 8-PSK mapping, rate-2/3, 8-state CC, 12,000-bit interleaver, 8 iterations.

8-ary and 8-PSK constellations. On the other hand the asymptotic performance of the 8-PSK is only slightly better than that of the cross 8-ary. As the convergence behaviors of these constellations/mapping are different, the appropriate choice of signal constellation/mapping depends on the desired BER level. For example, Fig. 3.10 suggests that the cross 8-ary, $(1, 7)$ and optimum 8-ary are the best choices if the targeted BER level is 10^{-3}. The 8-PSK performs the best at BER levels in the range from 10^{-3} to 10^{-4}. Finally, to achieve a BER of 10^{-6}, the asymmetric 8-PSK is the most efficient constellation (about 1dB is gained compared to other constellations).

It is also of interest to study the performance of BICM-ID systems employing different constellations/mappings, a short interleaver and when only a few iterations can be afforded. This is important when receiver complexity and/or short delay are of particular concern(s). Figure 3.11 presents the BER performance of BICM-ID systems

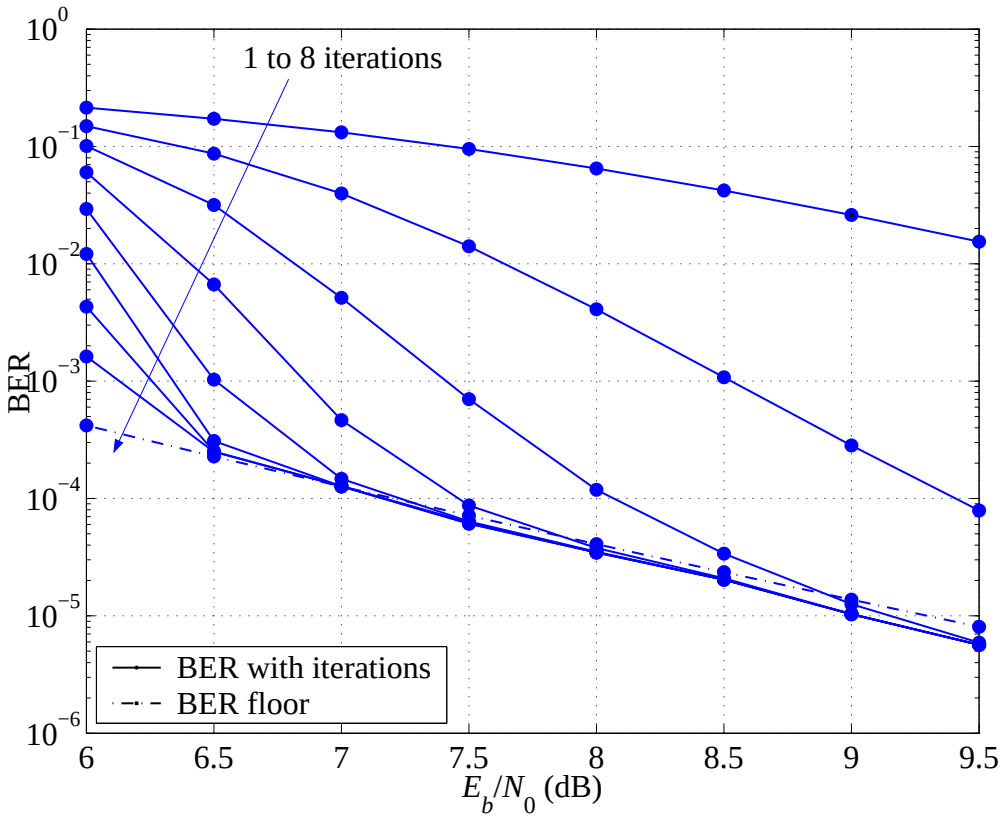

Figure 3.9 BER performance of a BICM-ID system: Cross 8-ary/proposed mapping, rate-2/3, 8-state CC, 12,000-bit interleaver, 8 iterations.

using an interleaver with a length of $1,200$ coded bits and after 5 iterations. As can be expected, the performance of the asymmetric 8-PSK constellation/mapping is very poor compared to that of other constellations/mappings. Obviously, the iterations do not work with this asymmetric 8-PSK constellation because of its small value of d_h^2 and the use of a short interleaver. With bigger values of d_h^2, the advantage of the cross 8-ary, $(1, 7)$ and the optimum 8-ary constellations is clear from Figure 3.11. They outperform both the asymmetric 8-PSK and 8-PSK constellations at practical range of BER level from 10^{-3} to 10^{-5}.

Finally, for AWGN channels, reference [26] also demonstrates that the mappings proposed in Fig. 3.4 are also the best mappings in terms of asymptotic performance. It can also be verified easily that these mapping achieve the minimum values of $\delta_2(\Psi, \xi)$. However, it should be pointed out that the iterative decoding gains offered

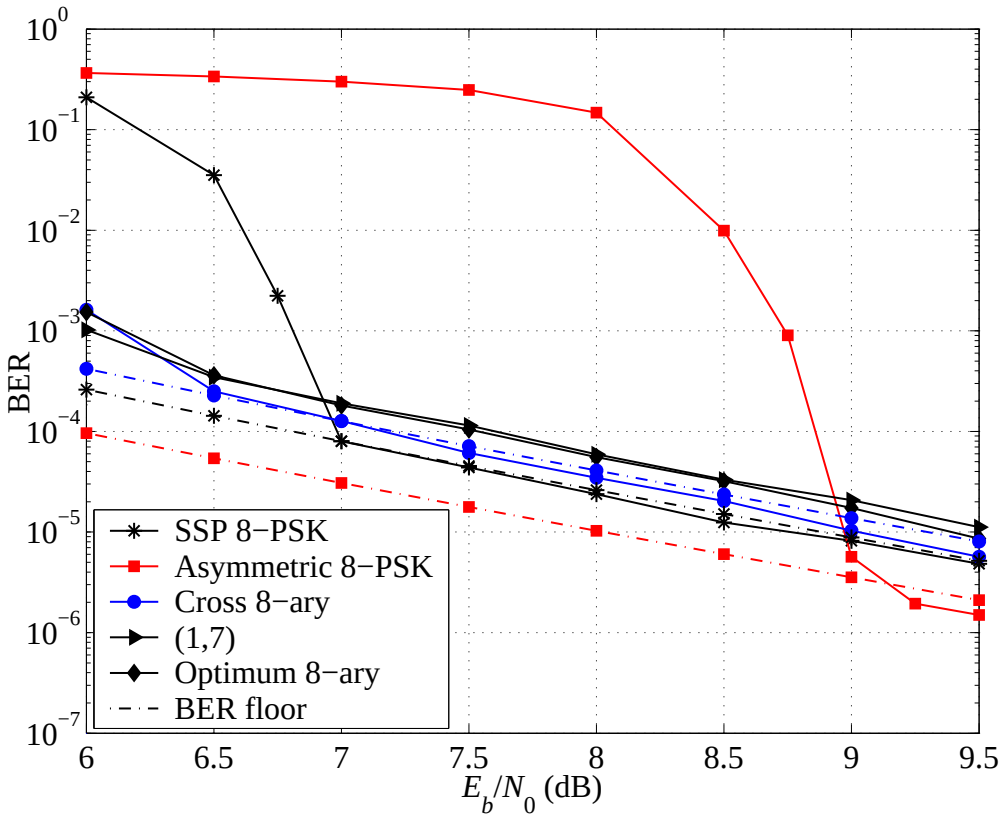

Figure 3.10 BER performance of BICM-ID systems: Different constellations/mappings, rate-2/3, 8-state CC, 12,000-bit interleaver, 8 iterations.

by these proposed mappings might not be evident at BER values of interest. This observation is also consistent with that in [14].

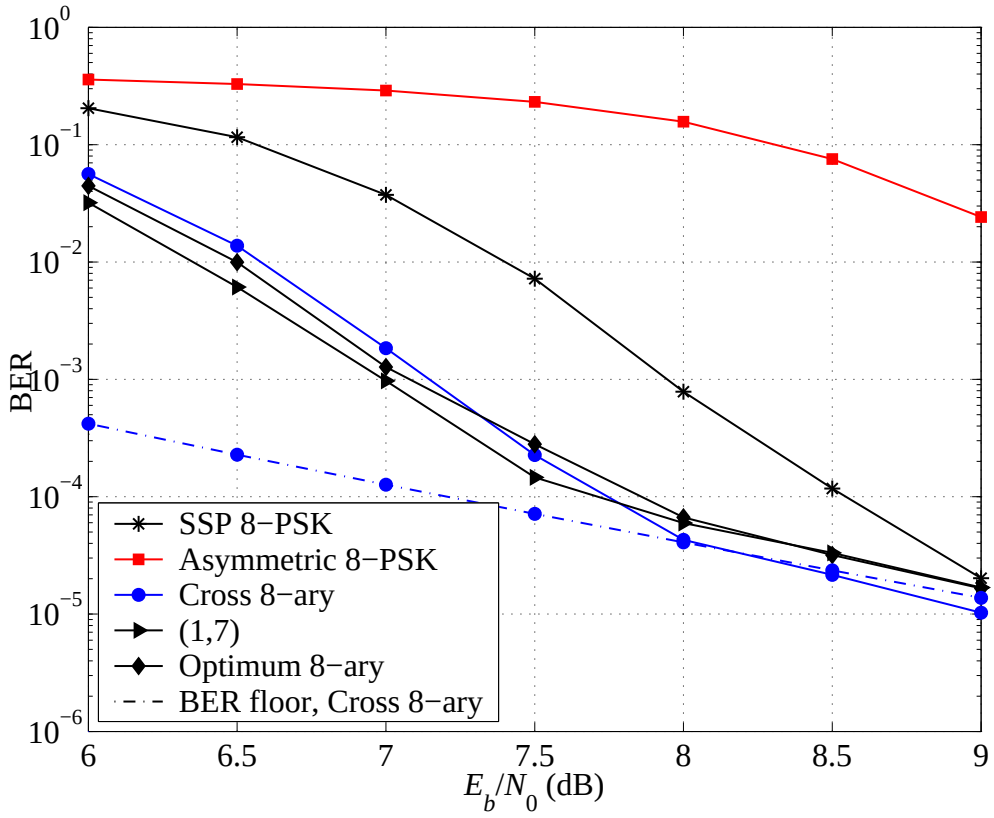

Figure 3.11 BER performance of BICM-ID systems: Different constellations/mappings, rate-2/3, 8-state CC, 1,200-bit iterleaver, 5 iterations.

4. BICM-ID with Hypercube Constellations

As pointed out in Chapter 3, BICM-ID is a spectral efficient coded modulation technique for both AWGN and Rayleigh fading channels. The main advantage of BICM-ID is that its implementation is quite simple compared to other complex coded modulation techniques while still providing large coding gains. This is because BICM-ID requires only one soft-input soft-output (SISO) decoder instead of two as normally seen in turbo decoding [19, 31]. In BICM-ID systems, the iterative processing is implemented between one SISO decoder and the soft-output demodulator. The complexity of the soft-output demodulator is relatively small compared to that of a SISO decoder [14] (which requires *forward* and *backward* recursions [22]), since the soft-output demodulator can be considered as a one-state (zero-memory) decoder. For example, it was shown in [14] that while the complexity of an 8-state BICM-ID system is about one half of that of an 8-state turbo TCM, its performance closely matches that of the turbo TCM system.

Using multi-dimensional constellations, it has been shown in [32,33] that the error performance of trellis coded modulation (TCM) systems can be improved significantly. The multi-dimensional constellation can be created by considering many symbols over a multiple symbol interval. Obviously, such a larger constellation offers a more flexibility for the mapping design compared to two-dimensional constellations.

Motivated by the above observation, this chapter shall study signal mapping in multi-dimensional constellations to improve the error performance of BICM-ID. More specifically, the chapter is concerned with BICM-ID systems employing QPSK. Since QPSK consists of orthogonal components, considering multiple QPSK symbols over

a multiple symbol interval is equivalent to employing a multi-dimensional hypercube as a signal constellation.

The chapter starts with a brief study of the hypercube and its distance properties. Then BICM-ID models employing hypercube constellations are described. As in the conventional BICM-ID systems, the effect of signal mapping to the system performance will be studied based on both the error bound and the technique of mutual information.

In the case of error bound, the guideline for mapping design derived from the error bound in [10, 14] cannot be applied directly to the proposed system. This is because the multiple QPSK symbols in our systems are treated as a single hypercube signal and need to be taken into account simultaneously. Nevertheless, by using the techniques in Chapter 3, the upper bounds for the bit error rate (BER) of the BICM-ID systems under consideration can also be evaluated. Distance criteria and related conditions are then obtained to find the best mapping of a hypercube with respect to the asymptotic performance. Moreover, by investigating the distance properties of a hypercube, a general and simple algorithm to construct the best mapping of a hypercube is proposed. It is then demonstrated by simulation that the use of the proposed mappings provide significant coding gains over the conventional systems for both AWGN and Rayleigh fading channels. The tightness of the error bound derived in this chapter is also verified.

4.1 A Hypercube and Its Distance Properties

A hypercube is a cube of more than three dimensions where the coordinate of any vertex in each dimension is either -1 or $+1$. A single point (or vertex) can be considered as a zero-dimensional cube. Two vertices (2^1) joined by a line form a one-dimensional cube. Four vertices (2^2) arranged in a square are a two-dimensional cube and eight (2^3) vertices are an ordinary three-dimensional cube. Continuing this geometric progression, the first hypercube is four-dimensional cube with $2^4 = 16$ vertices. Generally, m-dimensional cube (m-cube) has 2^m vertices and each vertex

can be identified by a set of m Cartesian coordinates, where each coordinate takes on the value $+1$ or -1. To make an $(m+1)$-dimensional cube, take two m-dimensional cubes and join each vertex on one cube to the corresponding vertex on the other. More simply, a $(m+1)$-cube can be created from m-cube and 1-cube by adding the single coordinate of 1-cube to the m Cartesian coordinates of m-cube. Examples of 1-cube, 2-cube, 3-cube and 4-cube are plotted in Fig. 4.1.

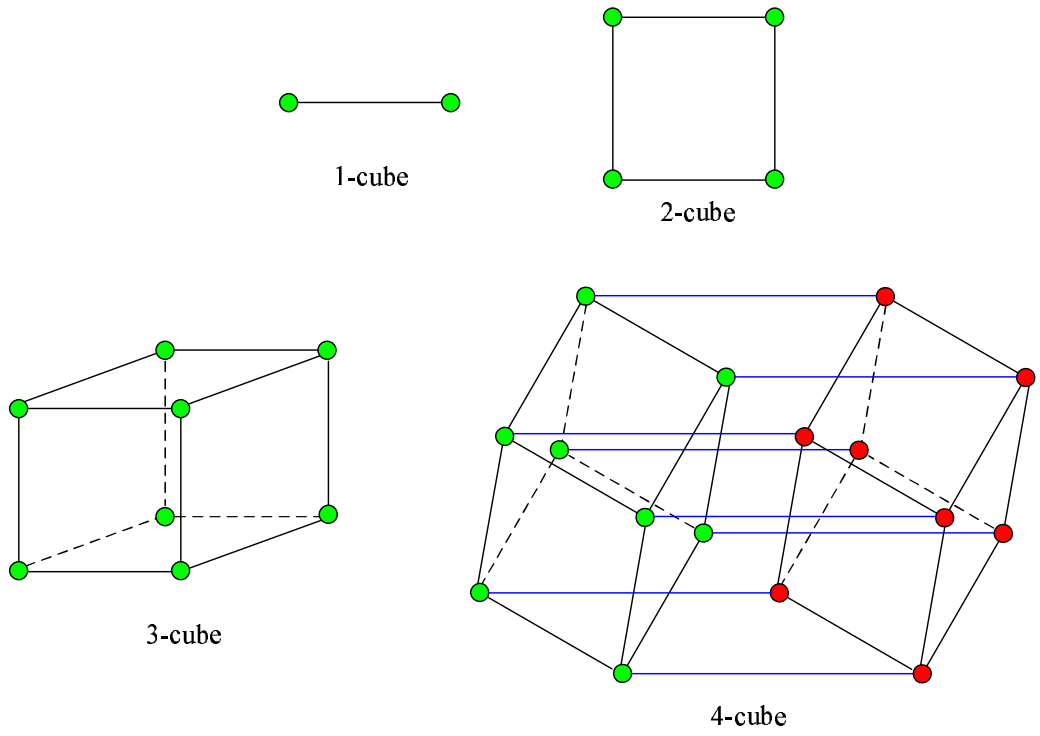

Figure 4.1 Cubes in 1, 2, 3 and 4-dimensional space.

Consider two vertices represented by two m Cartesian coordinates $\mathbf{s}_i = [s_{i,1}, \ldots, s_{i,m}]$ and $\mathbf{s}_j = [s_{j,1}, \ldots, s_{j,m}]$ where $s_{i,k}, s_{j,k}, 1 \leq k \leq m$, are either $+1$ or -1 and $1 \leq i, j \leq 2^m$. The squared Euclidean distance between the two vertices can be computed as follows:

$$d_{i,j}^2 = \sum_{k=1}^{m} (s_{i,k} - s_{j,k})^2 \tag{4.1}$$

Observe that since $(s_{i,k} - s_{j,k})^2$ is either 0 (when $s_{i,k} = s_{j,k}$) or 4 (when $s_{i,k} \neq s_{j,k}$), $d_{i,j}^2$ can be obtained from the coordinate Hamming distance $H_{i,j}$ between $\mathbf{s}_i$ and $\mathbf{s}_j$ (i.e.,

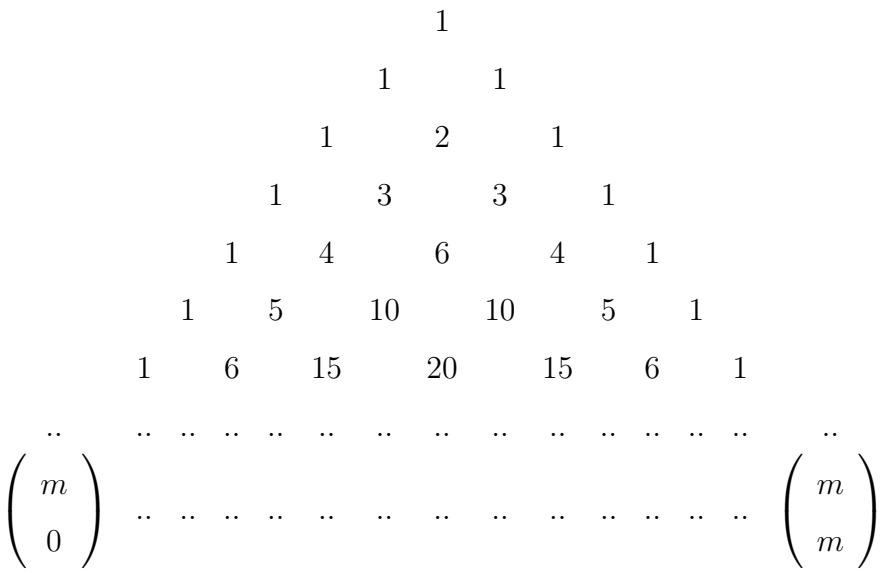

Figure 4.2 Coordinate Hamming distance profile of an m-cube.

the number of positions in which the two m-tuples of Cartesian coordinates disagree) as:

$$d_{i,j}^2 = 4H_{i,j} \tag{4.2}$$

Thus, to study the Euclidean distance properties of a cube, one may investigate the coordinate Hamming distance properties, instead.

There are a total of 2^m vertices in an m-dimensional cube, but the coordinate Hamming distance between any two vertices ranges from 0 to m. For any vertex $\mathbf{s}_i$, the coordinate Hamming distance is zero when compared to itself and it equals to m when compared to the vertex $\mathbf{s}_j$ whose coordinates are the complements of that of $\mathbf{s}_i$. It is also not difficult to show that for any vertex $\mathbf{s}_i$, the number of vertices $\{\mathbf{s}_j\}$ with coordinate Hamming distance $H_{i,j} = k, 0 \leq k \leq m$, to the vertices $\mathbf{s}_i$ is:

$$m_k = \binom{m}{k} \tag{4.3}$$

Hence, the coordinate Hamming distance profile of an m-cube can be represented by the famous Pascal triangle of Fig. 4.2.

Here, the mth horizontal row in this triangle represents the coordinate Hamming distance profile for the $(m-1)$-cube. As an example, the fourth row of the triangle

shows that, for any vertex in a three-dimensional cube, there is 1 vertex at coordinate Hamming distance 0, 3 vertices at coordinate Hamming distance 1, 3 vertices at coordinate Hamming distance 2 and 1 vertex at coordinate Hamming distance 3 away from this vertex.

Also observe from the above Pascal triangle that there are m vertices with coordinate Hamming distance 1, m vertices with coordinate Hamming distance $(m-1)$ and 1 vertex with Hamming distance m to any reference vertex in an m-cube. The vertex with coordinate Hamming distance 1 (i.e., the minimum Euclidean distance $d_{\min} = 4$) to a reference vertex is called the nearest neighbor of this vertex.

It should be emphasized that the above properties hold for any m-cube in general. As will be seen later, only the m-cube with m is even is considered in this thesis. This is because the hypercube considered in this thesis is constructed from QPSK components. For this reason, the m-cube mentioned hereafter always corresponds to $m = 2n$.

4.2 BICM-ID with Hypercube Constellations

Consider a BICM-ID system that employs QPSK modulation and a rate-1/2 convolutional code (CC). This combination yields a spectral efficiency of 1 bit/s/Hz. For every n symbol interval, a group of $m = 2n$ coded bits is mapped to n consecutive QPSK symbols. As a result, a $2n$-dimensional hypercube ($2n$-cube) is created. There are a total of 4^n vertices in this hypercube constellation. Clearly, the Gray or anti-Gray mapping for each QPSK constellation is just one special case of the above general mapping. Assume that the energy of each QPSK symbol is normalized to be 2 joules. Then the coordinates of the four QPSK symbols are $[+1, +1]$, $[+1, -1]$, $[-1, +1]$ and $[-1, -1]$. The vertex of the $2n$-cube can be represented by the following vector:

$$\mathbf{s}_i = [s_{i,1}, s_{i,2}, \ldots, s_{i,2n-1}, s_{i,2n}] \tag{4.4}$$

where $s_{i,j} \in \{\pm 1\}$, $1 \leq j \leq 2n$, and $i = 1, \ldots, 4^n$. Equivalently,

$$\mathbf{s}_i = [\mathbf{q}_{i,1}, \ldots, \mathbf{q}_{i,n}] \tag{4.5}$$

where $\mathbf{q}_{i,p} = [s_{i,2p-1}, s_{i,2p}], 1 \leq p \leq n$, represents the pth QPSK symbol. Clearly, the energy of each $2n$-cube signal is $2n$ and each signal carries $2n$ coded bits. Also there is no change in bandwidth efficiency due to the use of this general $2n$-cube constellation.

When a $2n$-cube is employed as a signal constellation in a BICM-ID system, each vertex $\mathbf{s}_i$ represents a transmitted signal and it is labelled by $2n$ binary digits (or bits) as follows:

$$\mathbf{a}_i = (a_{i,1}, a_{i,2}, \ldots, a_{i,2n}) = \xi^{-1}(\mathbf{s}_i) \tag{4.6}$$

where $a_{i,k}, 1 \leq k \leq 2n$, is either 0 or 1 and ξ denotes the symbol mapping from $2n$ bits to one $2n$-cube signal. It should be pointed out that while the Cartesian coordinates $\mathbf{s}_i$ define the location of the transmitted signal in the Euclidean space (i.e., the actual waveform of the transmitted signal), choosing the labels $\mathbf{a}_i$ of the signal points can be done freely and it is the main subject of this chapter. In the remaining of this chapter, the coordinate Hamming distance refers to the Hamming distance computed from the Cartesian coordinates of two vertices, while the label Hamming distance is determined from the labels of two vertices.

Let $\mathbf{r} = [\mathbf{r}_1, \ldots, \mathbf{r}_n]$ represent the received signal in a $2n$-dimensional signal space. For a frequency non-selective slowly Rayleigh fading channel, $\mathbf{r}_p$ is related to the transmitted symbol $\mathbf{q}_{i,p}$ as:

$$\mathbf{r}_p = g_p \mathbf{q}_{i,p} + \mathbf{w}_p \tag{4.7}$$

where $1 \leq p \leq n$ and i is the index of the transmitted signal. In (4.7), $\mathbf{w}_p$ is complex white Gaussian noise with independent components having two-sided power spectral density $N_0/2$ and the scalar g_p is a Rayleigh random variable representing the fading amplitude of the pth QPSK symbol. It is assumed that the channel fades slowly compared to one QPSK symbol interval and two cases of fading amplitude are considered:

(i) The coefficients $\{g_p\}$ are different for different QPSK symbols in one $2n$-cube signal.

(ii) The coefficients $\{g_p\}$ are constant over n consecutive QPSK symbols, i.e., over the duration of the $2n$-cube signal.

The channel model in the first case is referred to as the fast fading channel. The second channel model is usually referred to as a quasistatic fading channel [34]. The fading coefficients can be modeled as independent and identical distributed (i.i.d) Rayleigh random variables as mentioned in Chapter 3. It is also further assumed that perfect channel state information (CSI) is available at the receiver, i.e., g_p can be perfectly estimated. In the case of an AWGN channel, g_p equals 1 for all p.

The receiver in the proposed BICM-ID systems also includes the soft-output demodulator and SISO decoder as described in Chapter 3. The only difference here is in the soft-output demodulator, where it works with a hypercube constellation.

4.3 The Upper Bound of BER Performance

This section studies the asymptotic performance of the systems under consideration with the assumption of error-free feedback. Similar to Chapter 2 and Chapter 3, to obtain the upper bound of the bit error rate P_b, one needs to compute the average pairwise error probability $f(d, \Psi, \xi)$. The function $f(d, \Psi, \xi)$ depends on the Hamming distance d, the $2n$-cube constellation Ψ and the mapping ξ. Using the same notation as in Chapters 2 and 3, $\underline{\mathbf{c}}$ and $\underline{\breve{\mathbf{c}}}$ denote the input sequence and the estimated sequence with the Hamming distance d between them. These binary sequences correspond to the sequences $\underline{\mathbf{x}}$ and $\underline{\breve{\mathbf{x}}}$, each consisting of d $2n$-cube symbols.

In what follows, the technique in [27, 28] is also applied to compute the PEP for the BICM-ID systems under consideration. For the cases of Rayleigh quasistatic fading and AWGN channels, the analysis yields a similar result as in Chapter 3 with two-dimensional constellations. More significantly, the analysis obtains a distance criterion for mapping design when the channel is a Rayleigh fast fading channel men-

tioned earlier.

Without loss of generality, assume that $\underline{\mathbf{c}}$ and $\breve{\underline{\mathbf{c}}}$ differ in the first d consecutive bits. Hence, $\underline{\mathbf{x}}$ and $\breve{\underline{\mathbf{x}}}$ can be redefined as sequences of d $2n$-cube signals as follows:

$$\underline{\mathbf{x}} = [\mathbf{x}_1, \ldots, \mathbf{x}_d] = \left[\mathbf{y}_{1,1}, \ldots, \mathbf{y}_{1,n}, \ldots, \mathbf{y}_{d,1}, \ldots, \mathbf{y}_{d,n}\right] \tag{4.8}$$

$$\breve{\underline{\mathbf{x}}} = [\breve{\mathbf{x}}_1, \ldots, \breve{\mathbf{x}}_d] = \left[\breve{\mathbf{y}}_{1,1}, \ldots, \breve{\mathbf{y}}_{1,n}, \ldots, \breve{\mathbf{y}}_{d,1}, \ldots, \breve{\mathbf{y}}_{d,n}\right] \tag{4.9}$$

Also let

$$\underline{\mathbf{g}} = [g_{1,1}, g_{1,2}, \ldots g_{1,n}, \ldots, g_{d,1}, g_{d,2}, \ldots, g_{d,n}] \tag{4.10}$$

where $g_{e,p}$, $1 \leq e \leq d$, $1 \leq p \leq n$, represents the Rayleigh fading coefficient corresponding to the pth QPSK symbol ($\mathbf{y}_{e,p}$) in the eth $2n$-cube signal ($\mathbf{x}_e$). Then the PEP conditioned on $\underline{\mathbf{g}}$ can be computed as

$$P(\underline{\mathbf{x}} \to \breve{\underline{\mathbf{x}}}|\underline{\mathbf{g}}) = Q\left(\sqrt{\frac{1}{2N_0}\sum_{e=1}^{d}\sum_{p=1}^{n} g_{e,p}^2 \|\mathbf{y}_{e,p} - \breve{\mathbf{y}}_{e,p}\|^2}\right) \tag{4.11}$$

To further evaluate (4.11), three different cases of channel model are considered next.

4.3.1 The Fast Fading Channels

For this channel model, the QPSK symbols in each $2n$-cube signal fade independently. By using the Gaussian probability integral in (3.13) and averaging (4.11) over the sequence $\underline{\mathbf{g}}$ of $n \times d$ i.i.d. Rayleigh random variables, it can be shown that [27,28]:

$$P(\underline{\mathbf{x}} \to \breve{\underline{\mathbf{x}}}) \;=\; E_{\underline{\mathbf{g}}}\left\{P(\underline{\mathbf{x}} \to \breve{\underline{\mathbf{x}}}|\underline{\mathbf{g}})\right\} = \frac{1}{\pi}\int_0^{\pi}\left(\prod_{e=1}^{d}\Delta_e\right)\mathrm{d}\theta \tag{4.12}$$

where

$$\Delta_e = \prod_{p=1}^{n}\left(1 + \frac{1}{4N_0}\frac{\|\mathbf{y}_{e,p} - \breve{\mathbf{y}}_{e,p}\|^2}{\sin^2\theta}\right)^{-1} \tag{4.13}$$

The term inside the brackets of (4.13) is obtained similarly by invoking the equality for a Rayleigh random variable g in (3.16).

Similar to Chapter 3, assume that the iterative processing works perfectly. Thus, one needs to consider the two vertices $\mathbf{x}_e = [\mathbf{y}_{e,1}, \ldots, \mathbf{y}_{e,n}]$ and $\breve{\mathbf{x}}_e = [\breve{\mathbf{y}}_{e,1}, \ldots, \breve{\mathbf{y}}_{e,n}]$

whose labels differ in only one bit. Also observe that Δ_e are i.i.d. random variables. Therefore, the function $f(d, \Psi, \xi)$ can be obtained by averaging over all vertices of $2n$-cube Ψ [10, 27]. Consider a single vertex $\mathbf{s}_i \in \Psi$ and the vertex $\mathbf{s}_{j(i,k)}$ that is related to $\mathbf{s}_i$ as described in Chapter 3[1]. Taking the average over all the vertices $\mathbf{s}_i = [\mathbf{q}_{i,1}, \ldots, \mathbf{q}_{i,n}]$ and the corresponding vertices $\mathbf{s}_{j(i,k)} = [\mathbf{q}_{j(i,k),1}, \ldots, \mathbf{q}_{j(i,k),n}]$ in the $2n$-cube Ψ gives

$$f(d, \Psi, \xi) \leq \frac{1}{\pi} \int_0^{\pi/2} \left[E \left\{ \prod_{p=1}^{n} \left(1 + \frac{1}{4N_0} \frac{\|\mathbf{q}_{i,p} - \mathbf{q}_{j(i,k),p}\|^2}{\sin^2 \theta} \right)^{-1} \right\} \right]^d \mathrm{d}\theta \tag{4.14}$$

where

$$E \left\{ \prod_{p=1}^{n} \left(1 + \frac{1}{4N_0} \frac{\|\mathbf{q}_{i,p} - \mathbf{q}_{j(i,k),p}\|^2}{\sin^2 \theta} \right)^{-1} \right\} = \frac{1}{m2^m} \sum_{\mathbf{s}_i \in \Psi} \sum_{k=1}^{m} \left[\prod_{p=1}^{n} \left(1 + \frac{1}{4N_0} \frac{\|\mathbf{q}_{i,p} - \mathbf{q}_{j(i,k),p}\|^2}{\sin^2 \theta} \right)^{-1} \right] \tag{4.15}$$

and $m = 2n$ is the number of coded bits per vertex. To understand the effect of the constellation mapping on the asymptotic behavior of the PEP, the inequality in (3.19) is substituted into (4.11). Taking the average over $\underline{\mathbf{g}}$ gives

$$P(\underline{\mathbf{x}} \to \underline{\breve{\mathbf{x}}}) \;<\; \frac{1}{2} \prod_{e=1}^{d} \prod_{p=1}^{n} \left(1 + \frac{\|\mathbf{y}_{e,p} - \breve{\mathbf{y}}_{e,p}\|^2}{4N_0} \right)^{-1} \tag{4.16}$$

As before, by taking the average over all the vertices $\mathbf{s}_i = [\mathbf{q}_{i,1}, \ldots, \mathbf{q}_{i,n}]$ and the corresponding vertices $\mathbf{s}_{j(i,k)} = [\mathbf{q}_{j(i,k),1}, \ldots, \mathbf{q}_{j(i,k),n}]$, the function $f(d, \Psi, \xi)$ is approximated at high SNR as:

$$f(d, \Psi, \xi) \;\approx\; \frac{1}{2} \left[\delta_1^{(A)}(\Psi, \xi) \right]^d \tag{4.17}$$

where the parameter $\delta_1^{(A)}(\Psi, \xi)$ depends on the constellation and mapping. This parameter is given by

$$\delta_1^{(A)}(\Psi, \xi) = \frac{1}{m2^m} \sum_{\mathbf{s}_i \in \Psi} \sum_{k=1}^{m} \prod_{p=1}^{n} \left(1 + \frac{\|\mathbf{q}_{i,p} - \mathbf{q}_{j(i,k),p}\|^2}{4N_0} \right)^{-1} \tag{4.18}$$

[1]Recall that $\mathbf{s}_{j(i,k)}$ denotes the signal whose label differs at position k compared to the label of $\mathbf{s}_i$.

4.3.2 The Quasistatic Fading Channels

The assumption of quasistatic fading in [34] implies that the fading is sufficiently slow as to be constant over the duration of a number of QPSK symbols. It is therefore reasonable to assume that QPSK symbols in each $2n$-cube signal fade simultaneously. The fading coefficients $g_{e,p}$ just equal g_e for all $1 \leq p \leq n$ and $\underline{\mathbf{g}} = [g_1, \ldots, g_d]$. With an ideal interleaver, g_e, $1 \leq e \leq d$, can be assumed to be i.i.d random variables [10]. Hence, the upper bound derived in Chapter 3 for the conventional BICM-ID systems over Rayleigh fading channels can be applied in this case. It is straightforward to show that

$$f(d, \Psi, \xi) \leq \frac{1}{\pi} \int_0^{\pi/2} \left[E\left\{ \left(1 + \frac{1}{4N_0} \cdot \frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{\sin^2 \theta} \right)^{-1} \right\} \right]^d \mathrm{d}\theta \tag{4.19}$$

where $E\{\cdot\}$ denotes the average over all the vertices $\mathbf{s}_i$.

At high SNR, the function $f(d, \Psi, \xi)$ is approximated by:

$$f(d, \Psi, \xi) \approx \frac{1}{2} \left[\delta_1^{(B)}(\Psi, \xi) \right]^d \tag{4.20}$$

where

$$\delta_1^{(B)}(\Psi, \xi) = \frac{1}{m2^m} \sum_{\mathbf{s}_i \in \Psi} \sum_{k=1}^{m} \left(1 + \frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{4N_0} \right)^{-1} \tag{4.21}$$

4.3.3 The AWGN Channels

The result for this channel model is exactly what was derived in Chapter 3. The function $f(d, \Psi, \xi)$ at high SNR can be approximated as

$$f(d, \Psi, \xi) \approx \frac{1}{2} \left[\delta_2(\Psi, \xi) \right]^d \tag{4.22}$$

where

$$\delta_2(\Psi, \xi) = \frac{1}{m2^m} \sum_{\mathbf{s}_i \in \Psi} \sum_{k=1}^{m} \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{4N_0} \right) \tag{4.23}$$

As before, the three parameters $\delta_1^{(A)}(\Psi, \xi)$, $\delta_1^{(B)}(\Psi, \xi)$ and $\delta_2(\Psi, \xi)$ derived above characterize the influence of the signal mapping to the asymptotic performance of BICM-ID systems over the three channel models, respectively. Specifically, the smaller

these parameters are, the lower the asymptotic bit error rate performance of the proposed systems becomes. It is clear that for a high-dimensional cube, an exhaustive search to find the mappings that yield the smallest values of these parameters is impossible. In the next section, by obtaining the lower bounds on $\delta_1^{(A)}(\Psi,\xi)$, $\delta_1^{(B)}(\Psi,\xi)$ and $\delta_2(\Psi,\xi)$ for a hypercube constellation, a universal condition for the best mapping that minimizes all these parameters is established. An algorithm to construct the mapping that satisfies the condition is also provided.

As in the case of conventional BICM-ID system, it is also important to investigate the error performance of system after the first iteration (i.e., the performance of a BICM system). This is because such a performance influences the performance convergence behavior of a BICM-ID system. Similar to the previous derivations, performance evaluation of a BICM system can be obtained by substituting $\mathbf{s}_{j(i,k)}$ by $\mathbf{s}_{j(i,k)}^{\text{bicm}}$, where $\mathbf{s}_{j(i,k)}^{\text{bicm}}$ is the nearest vertex of $\mathbf{s}_i$ whose labels differ at position k [10][2]. In the case of an AWGN channel, the key parameter for the asymptotic performance of a BICM system is the minimum Euclidean distance between any two signal points in the constellation [2, 10]. The other relevant parameter is the average number of signals at the minimum Euclidean distance $N_{\min}$ defined in (2.11).

4.4 The Best Mapping of a Hypercube Constellation

4.4.1 The Lower Bounds on $\delta_1^{(A)}(\Psi,\xi)$, $\delta_1^{(B)}(\Psi,\xi)$ and $\delta_2(\Psi,\xi)$

First, the lower bound on $\delta_1^{(A)}(\Psi,\xi)$ is obtained as follows. Define the parameter $\delta_1^{(A)}(\Psi,\xi,\mathbf{s}_i)$ associated with each vertex $\mathbf{s}_i$ as:

$$\delta_1^{(A)}(\Psi,\xi,\mathbf{s}_i) = \frac{1}{m}\sum_{k=1}^{m}\prod_{p=1}^{n}\left(1+\frac{\|\mathbf{q}_{i,p}-\mathbf{q}_{j(i,k),p}\|^2}{4N_0}\right)^{-1} \tag{4.24}$$

where $m = 2n$. Observe that $\|\mathbf{q}_{i,p}-\mathbf{q}_{j(i,k),p}\|^2$ can only take a value in $\{0,4,8\}$. This observation together with the distance properties of a cube implies the following

[2]These two vertices are not necessarily at the minimum Euclidean distance.

inequality:

$$\delta_1^{(A)}(\Psi,\xi,\mathbf{s}_i) \geq \frac{1}{m}\left[\left(1+\frac{2}{N_0}\right)^{1-n}\left(\left[1+\frac{2}{N_0}\right]^{-1}+(m-1)\left[1+\frac{1}{N_0}\right]^{-1}\right)\right] \quad (4.25)$$

Hence

$$\delta_1^{(A)}(\Psi,\xi) = \frac{1}{2^m}\sum_{\mathbf{s}_i\in\Psi}\delta_1^{(A)}(\Psi,\xi,\mathbf{s}_i) \geq \frac{1}{m}\left[\left(1+\frac{2}{N_0}\right)^{1-n}\left(\left[1+\frac{2}{N_0}\right]^{-1}+(m-1)\left[1+\frac{1}{N_0}\right]^{-1}\right)\right] \quad (4.26)$$

The inequalities in (4.25) and (4.26) follow from the fact that for each vertex $\mathbf{s}_i$ of a hypercube, there is only one vertex with coordinate Hamming distance m and there are m vertices with coordinate Hamming distance $(m-1)$ to $\mathbf{s}_i$.

The lower bound in (4.26) is valid for any mapping. This means that the mapping that achieves the lower bound in (4.26), if it exists, is the best mapping as far as minimizing $\delta_1^{(A)}(\Psi,\xi)$ is concerned. Furthermore, it is simple to see that such a mapping must satisfy the following condition.

Condition 1: *For any vertex $\boldsymbol{s}_i$ of an m-cube, the vertices with coordinate Hamming distance m to $\boldsymbol{s}_i$ must have label Hamming distance 1 to $\boldsymbol{s}_i$. Furthermore, there must be $(m-1)$ vertices with coordinate Hamming distances $(m-1)$ and label Hamming distance 1 to $\boldsymbol{s}_i$.*

Next, to obtain the lower bound on $\delta_1^{(B)}(\Psi,\xi)$, define and bound the parameter $\delta_1^{(B)}(\Psi,\xi,\mathbf{s}_i)$ associated with each vertex $\mathbf{s}_i$ as follows:

$$\begin{aligned}\delta_1^{(B)}(\Psi,\xi,\mathbf{s}_i) &= \frac{1}{m}\sum_{k=1}^{m}\left(1+\frac{\|\mathbf{s}_i-\mathbf{s}_{j(i,k)}\|^2}{4N_0}\right)^{-1} = \frac{1}{m}\sum_{k=1}^{m}\left(1+\frac{H_{i,j(i,k)}}{N_0}\right)^{-1} \\ &\geq \frac{1}{m}\left[\left(1+\frac{m}{N_0}\right)^{-1}+(m-1)\cdot\left(1+\frac{m-1}{N_0}\right)^{-1}\right] \quad (4.27)\end{aligned}$$

The above implies that

$$\delta_1^{(B)}(\Psi,\xi) = \frac{1}{2^m}\sum_{\mathbf{s}_i\in\Psi}\delta_1^{(B)}(\Psi,\xi,\mathbf{s}_i) \geq \frac{1}{m}\left[\left(1+\frac{m}{N_0}\right)^{-1}+(m-1)\cdot\left(1+\frac{m-1}{N_0}\right)^{-1}\right] \quad (4.28)$$

Again, it is not hard to see that the best mapping of a hypercube that achieves the lower bound in (4.28) in this case is also the mapping that satisfies *Condition 1*.

Finally, for AWGN channels, define and bound the parameter $\delta_2(\Psi, \xi, \mathbf{s}_i)$ associated with each vertex $\mathbf{s}_i$ as follows:

$$\begin{aligned} \delta_2(\Psi, \xi, \mathbf{s}_i) &= \frac{1}{m}\sum_{k=1}^{m} \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_{j(i,k)}\|^2}{4N_0}\right) = \frac{1}{m}\sum_{k=1}^{m} \exp\left(-\frac{H_{i,j(i,k)}}{N_0}\right) \\ &\geq \frac{1}{m}\left[\exp\left(-\frac{m}{N_0}\right) + (m-1)\exp\left(-\frac{m-1}{N_0}\right)\right] \end{aligned} \tag{4.29}$$

The lower bound on $\delta_2(\Psi, \xi)$ is then given by

$$\delta_2(\Psi, \xi) = \frac{1}{2^m}\sum_{\mathbf{s}_i \in \Psi} \delta_2(\Psi, \xi, \mathbf{s}_i) \geq \frac{1}{m}\left[\exp\left(-\frac{m}{N_0}\right) + (m-1)\exp\left(-\frac{m-1}{N_0}\right)\right] \tag{4.30}$$

Similarly, the mapping, if it exists, that achieves the lower bound in (4.30) is the best mapping as far as minimizing $\delta_2(\Psi, \xi)$ is concerned. Such a mapping must also satisfy *Condition 1*.

Finally, it should be mentioned here that the technique of bitwise mutual information also lead to the same condition for the best mapping obtained above. The detailed justification can be found in Appendix C.

4.4.2 An Algorithm to Construct the Best Mapping

Instead of proving the existence of the mapping that satisfies *Condition 1*, it is convenient to introduce the equivalent condition, called *Condition 2*. The proof that the two conditions are equivalent is straightforward and follows from the distance properties of a hypercube discussed in Section 4.1.

Condition 2: For any vertex $\boldsymbol{s}_i$ in an m-cube, there must be one vertex $\boldsymbol{s}_j$ with coordinate Hamming distance m and label Hamming distance 1 to $\boldsymbol{s}_i$. Furthermore, there must exist $(m-1)$ vertices all of which have coordinate Hamming distance 1 to $\boldsymbol{s}_j$ (the nearest neighbors[3] of $\boldsymbol{s}_j$) and label Hamming distance 2 to $\boldsymbol{s}_j$. For any of these

[3]It should be noted that each vertex has a total of m nearest neighbors.

$(m-1)$ *vertices, one of the two positions in which its label differs from the label of* $\boldsymbol{s}_j$ *must be the same as the position where the labels of* $\boldsymbol{s}_i$ *and* $\boldsymbol{s}_j$ *differ.*

The proof that a mapping that satisfies *Condition 2* exists is by construction, where an explicit procedure to construct the desired mapping is provided below.

The Algorithm to Construct the Best Mapping ξ for m-Cube Ψ

- *Step 1*: Construct a $(m-1)$-cube with Gray labels as follows. For any vertex $\mathbf{s}_i$ of the $(m-1)$-cube with Cartesian coordinates $\mathbf{s}_i = [s_{i,1}, \ldots, s_{i,m-1}]$, label it with $\mathbf{a}_i = (a_{i,1}, \ldots, a_{i,m-1})$, where

$$a_{i,k} = (s_{i,k} + 1)/2 \tag{4.31}$$

  It is straightforward to show that with this labeling scheme, the label Hamming distance between any two vertices that are the nearest neighbors of each other (i.e., the coordinate Hamming distance between them is one) is one. Hence, the above labeling is Gray labeling.

- *Step 2*: For each vertex $\mathbf{s}_i$, modify its current label $\mathbf{a}_i = (a_{i,1}, \ldots, a_{i,m-1})$ to $\hat{\mathbf{a}}_i = (b, a_{i,1}, \ldots, a_{i,m-1}) = (b, \hat{a}_{i,1}, \ldots, \hat{a}_{i,m-1})$, where $b = 0$ if the weight of $\mathbf{a}_i$ is even and $b = 1$ if the weight of $\mathbf{a}_i$ is odd. Denote the $(m-1)$-cube with the modified m-bit labels by Ω. The above modification guarantees that:

  (i) the label Hamming distance between any two nearest neighbor vertices of Ω is two.

  (ii) the labels of the two nearest neighbor vertices always differ in the lowest significant bit position.

  (iii) the label Hamming weight of any vertex is always even.

- *Step 3*: From Ω, one-half of m-cube Ψ is constructed as follows. For each vertex $\mathbf{s}_i$ in Ω (with label $\hat{\mathbf{a}}_i$), the coordinates of $\mathbf{s}_i$ are changed by inserting a -1 in the lowest significant position to become $\hat{\mathbf{s}}_i = [-1, s_{i,1}, \ldots, s_{i,m-1}] =$

$[-1, \hat{s}_{i,1}, \ldots, \hat{s}_{i,m-1}]$. This procedure creates a subset $\hat{\Psi}$ of Ψ which contains 2^{m-1} vertices $\hat{\mathbf{s}}_i$ with label $\hat{\mathbf{a}}_i$. The following properties of $\hat{\Psi}$ can be obtained from the properties of Ω:

(i) The label Hamming weight of any vertex in $\hat{\Psi}$ is even (for this reason, denote the set of labels in $\hat{\Psi}$ by ξ^{e}).

(ii) For any vertex $\hat{\mathbf{s}}_i = [-1, \hat{s}_{i,1}, \ldots, \hat{s}_{i,m-1}]$ in $\hat{\Psi}$ with the label $\hat{\mathbf{a}}_i = (b, \hat{a}_{i,1}, \ldots, \hat{a}_{i,m-1})$, the relationship between the coordinate $\hat{s}_{i,k}$ and the labeling bit $\hat{a}_{i,k}$ can be obtained from (4.31) to be

$$\hat{a}_{i,k} = (\hat{s}_{i,k} + 1)/2 \tag{4.32}$$

(iii) Each vertex $\hat{\mathbf{s}}_i$ in $\hat{\Psi}$ has $(m-1)$ nearest neighbor vertices in $\hat{\Psi}$ and the label Hamming distance between any of these $(m-1)$ nearest neighbors to $\hat{\mathbf{s}}_i$ is two, including the difference in the lowest significant bit.

- *Step 4*: The remaining half of the vertices of Ψ (denoted by $\tilde{\Psi}$) is created as follows. For each $\hat{\mathbf{s}}_i = [-1, \hat{s}_{i,1}, \ldots, \hat{s}_{i,m-1}] \in \hat{\Psi}$ with label $\hat{\mathbf{a}}_i = (b, \hat{a}_{i,1}, \ldots, \hat{a}_{i,m-1}) \in \xi^{\mathrm{e}}$, obtain a new vertex $\tilde{\mathbf{s}}_i = [+1, -\hat{s}_{i,1}, \ldots, -\hat{s}_{i,m-1}] = [+1, \tilde{s}_{i,1}, \ldots, \tilde{s}_{i,m-1}] \in \tilde{\Psi}$ and label it with $\tilde{\mathbf{a}}_i = \left(\overline{b}, \hat{a}_{i,1}, \ldots, \hat{a}_{i,m-1}\right) = \left(\overline{b}, \tilde{a}_{i,1}, \ldots, \tilde{a}_{i,m-1}\right)$, where $\overline{b} = 1 - b$. By the above construction, there is a one to one correspondence between $\hat{\mathbf{s}}_i$ and $\tilde{\mathbf{s}}_i$ as well as their labels: If $\hat{\mathbf{s}}_j$ is the nearest neighbor of $\hat{\mathbf{s}}_i$ and their labels differ in two bits, including the lowest significant bit, then $\tilde{\mathbf{s}}_j$ is also the nearest neighbor of $\tilde{\mathbf{s}}_i$ and their labels also differ in two bits. Furthermore, the vertices in $\tilde{\Psi}$ possess the following properties similar to those in $\hat{\Psi}$:

(i) The label Hamming weight of any vertex in $\tilde{\Psi}$ is odd (hence, denote the set of labels in $\tilde{\Psi}$ by ξ^{o}).

(ii) For any vertex $\tilde{\mathbf{s}}_i = [+1, \tilde{s}_{i,1}, \ldots, \tilde{s}_{i,m-1}]$ in $\tilde{\Psi}$ with the label $\tilde{\mathbf{a}}_i = \left(\overline{b}, \tilde{a}_{i,1}, \ldots, \tilde{a}_{i,m-1}\right)$ ($\overline{b} = 0$ when $\sum_{k=1}^{m-1} \tilde{a}_{i,k}$ is odd and $\overline{b} = 1$ when $\sum_{k=1}^{m-1} \tilde{a}_{i,k}$ is even), the relationship between the coordinate $\tilde{s}_{i,k}$ and the bit $\tilde{a}_{i,k}$ can be obtained from (4.32) to be

$$\tilde{a}_{i,k} = \hat{a}_{i,k} = \frac{(\hat{s}_{i,k} + 1)}{2} = \frac{(-\tilde{s}_{i,k} + 1)}{2} \tag{4.33}$$

(iii) Each vertex $\tilde{\mathbf{s}}_i$ in $\tilde{\Psi}$ has $(m-1)$ nearest neighbor vertices in $\tilde{\Psi}$ and the label Hamming distance between any of these $(m-1)$ nearest neighbors to $\tilde{\mathbf{s}}_i$ is two, including the difference in the lowest significant bit.

- *Step 5*: The m-cube Ψ and the best labels ξ are finally obtained as $\Psi = \hat{\Psi} \cup \tilde{\Psi}$ and $\xi = \xi^{\mathrm{e}} \cup \xi^{\mathrm{o}}$.

Verifying that the m-cube Ψ with mapping ξ as constructed above satisfies *Condition 2* is quite simple. It can be seen that for any $\hat{\mathbf{s}}_i \in \hat{\Psi}$, the vertex $\tilde{\mathbf{s}}_i \in \tilde{\Psi}$ is the vertex whose label differs in only 1 bit at the lowest significant position and whose coordinate Hamming distance to $\mathbf{s}_i$ is m. There are $(m-1)$ nearest neighbor vertices $\tilde{\mathbf{s}}_j \in \tilde{\Psi}$ of $\tilde{\mathbf{s}}_i \in \tilde{\Psi}$ whose labels differ in 2 bits compared to that of $\tilde{\mathbf{s}}_i$, including the difference in the lowest significant position. Similarly, these properties are also true for any $\tilde{\mathbf{s}}_i \in \tilde{\Psi}$.

To illustrate the above procedure, an example is provided next.
Example 1: Consider constructing the best mapping for a four-dimensional hypercube (4-cube). The steps in the proposed algorithm are carried out as follows:

- *Step 1*: A 3-cube has $2^3 = 8$ vertices with the following Cartesian coordinates:

$$[-1,-1,-1],[-1,-1,+1],[-1,+1,-1],[-1,+1,+1]$$

$$[+1,-1,-1],[+1,-1,+1],[+1,+1,-1],[+1,+1,+1]$$

The Gray mapping for this 3-cube is shown in Fig. 4.3 and it is obtained as:

$$\mathbf{s}_i = [s_{i,1}, s_{i,2}, s_{i,3}] \longrightarrow \mathbf{a}_i = \left(\frac{s_{i,1}+1}{2}, \frac{s_{i,2}+1}{2}, \frac{s_{i,3}+1}{2} \right)$$

- *Step 2*: Modify the labels of 3-cube with Gray mapping as follows. The label with odd Hamming weight has binary bit 1 inserted. Otherwise, binary bit 0 is inserted. This step gives a 3-cube Ω with modified 4-bit labels as shown in

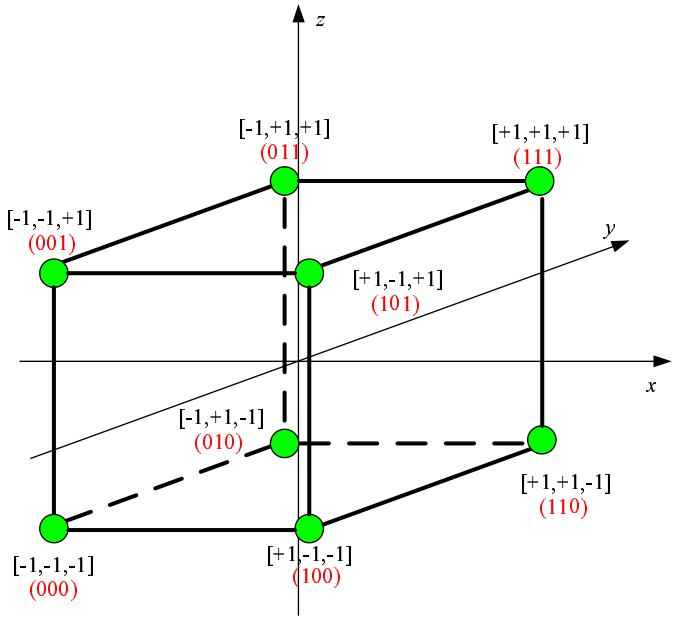

Figure 4.3 The Gray mapping for 3-cube.

Fig. 4.4. It is easy to verify that the label Hamming weight of any vertex in Ω is now even. The labels of any two nearest neighbors differ in 2 positions.

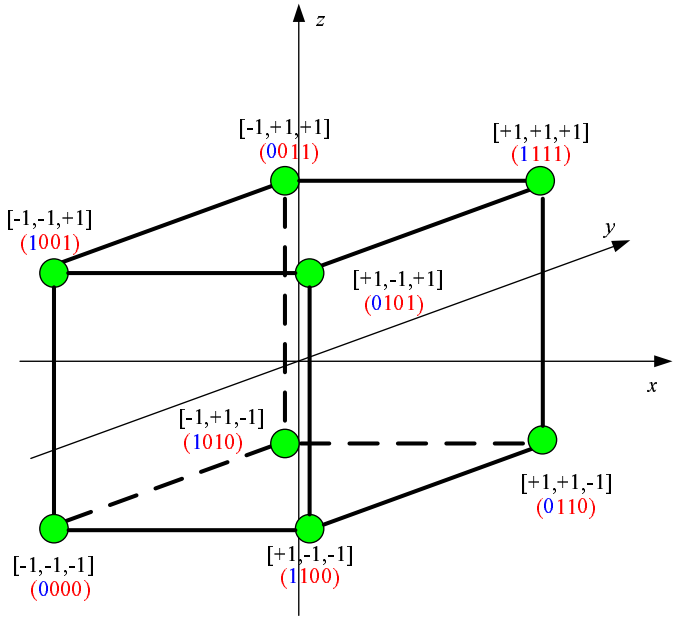

Figure 4.4 The 3-cube Ω with modified 4-bit labels.

- *Step 3*: Create subset $\hat{\Psi}$ that contains one-half of the vertices of Ψ by inserting -1 to the coordinates of each vertex of Ω. This procedure gives eight vertices $\hat{\mathbf{s}}_i$ with the corresponding labels $\hat{\mathbf{a}}_i$ as follows:

$$\begin{aligned}
\hat{\mathbf{s}}_1 &= [-1,-1,-1,-1] \longrightarrow \hat{\mathbf{a}}_1 = (0,0,0,0) \\
\hat{\mathbf{s}}_2 &= [-1,-1,-1,+1] \longrightarrow \hat{\mathbf{a}}_2 = (1,0,0,1) \\
\hat{\mathbf{s}}_3 &= [-1,-1,+1,-1] \longrightarrow \hat{\mathbf{a}}_3 = (1,0,1,0) \\
\hat{\mathbf{s}}_4 &= [-1,-1,+1,+1] \longrightarrow \hat{\mathbf{a}}_4 = (0,0,1,1) \\
\hat{\mathbf{s}}_5 &= [-1,+1,-1,-1] \longrightarrow \hat{\mathbf{a}}_5 = (1,1,0,0) \\
\hat{\mathbf{s}}_6 &= [-1,+1,-1,+1] \longrightarrow \hat{\mathbf{a}}_6 = (0,1,0,1) \\
\hat{\mathbf{s}}_7 &= [-1,+1,+1,-1] \longrightarrow \hat{\mathbf{a}}_7 = (0,1,1,0) \\
\hat{\mathbf{s}}_8 &= [-1,+1,+1,+1] \longrightarrow \hat{\mathbf{a}}_8 = (1,1,1,1)
\end{aligned}$$

- *Step 4*: For each $\hat{\mathbf{s}}_i = [-1, s_{i,1}, s_{i,2}] \in \hat{\Psi}$ with label $\hat{\mathbf{a}}_i = (b, a_{i,1}, a_{i,2}) \in \xi^{\mathrm{e}}$, obtain a new vertex $\tilde{\mathbf{s}}_i = [+1, -s_{i,1}, -s_{i,2}] \in \tilde{\Psi}$ and label it with $\tilde{\mathbf{a}}_i = (1-b, a_{i,1}, a_{i,2})$ (taking the complement $(1-b)$ of the lowest significant bit). This procedure gives the remaining half of the vertices of Ψ with the corresponding labels as follows:

$$\begin{aligned}
\tilde{\mathbf{s}}_1 &= [+1,+1,+1,+1] \longrightarrow \tilde{\mathbf{a}}_1 = (1,0,0,0) \\
\tilde{\mathbf{s}}_2 &= [+1,+1,+1,-1] \longrightarrow \tilde{\mathbf{a}}_2 = (0,0,0,1) \\
\tilde{\mathbf{s}}_3 &= [+1,+1,-1,+1] \longrightarrow \tilde{\mathbf{a}}_3 = (0,0,1,0) \\
\tilde{\mathbf{s}}_4 &= [+1,+1,-1,-1] \longrightarrow \tilde{\mathbf{a}}_4 = (1,0,1,1) \\
\tilde{\mathbf{s}}_5 &= [+1,-1,+1,+1] \longrightarrow \tilde{\mathbf{a}}_5 = (0,1,0,0) \\
\tilde{\mathbf{s}}_6 &= [+1,-1,+1,-1] \longrightarrow \tilde{\mathbf{a}}_6 = (1,1,0,1) \\
\tilde{\mathbf{s}}_7 &= [+1,-1,-1,+1] \longrightarrow \tilde{\mathbf{a}}_7 = (1,1,1,0) \\
\tilde{\mathbf{s}}_8 &= [+1,-1,-1,-1] \longrightarrow \tilde{\mathbf{a}}_8 = (0,1,1,1)
\end{aligned}$$

- *Step 5*: The 4-cube Ψ and the best mapping ξ are obtained by combining $\hat{\Psi}$ and $\tilde{\Psi}$ and their labels ξ^{e} and ξ^{o}. The final results are shown in Table 4.1.

In Table 4.1, the binary numbers under each signal (in the first row) denotes the proposed mapping. As an example, $\mathbf{s}_2 = [-1 - 1 - 1 + 1]$ is labeled with 1001, $\mathbf{s}_{15} = [+1 + 1 + 1 - 1]$ is labelled with 0001. These two labels differ in only the first bit and the Euclidean distance between these two 4-cube signals is 16. It can also be verified from Table 4.1 that the coordinate Hamming distance between any two vertices whose labels differ in the first bit, the second bit, the third bit and the fourth bit are 4, 3, 3 and 3, respectively. They correspond to 16, 12, 12 and 12 in terms of the squared Euclidean distances which are shown in bold numbers in Table 4.1.

Table 4.1 Distance parameters of the best mapping of 4-cube: The binary numbers under each signal (in the first row) denotes the proposed mapping.

	$\mathbf{s}_1$ 0000	$\mathbf{s}_2$ 1001	$\mathbf{s}_3$ 1010	$\mathbf{s}_4$ 0011	$\mathbf{s}_5$ 1100	$\mathbf{s}_6$ 0101	$\mathbf{s}_7$ 0110	$\mathbf{s}_8$ 1111	$\mathbf{s}_9$ 0111	$\mathbf{s}_{10}$ 1110	$\mathbf{s}_{11}$ 1101	$\mathbf{s}_{12}$ 0100	$\mathbf{s}_{13}$ 1011	$\mathbf{s}_{14}$ 0010	$\mathbf{s}_{15}$ 0001	$\mathbf{s}_{16}$ 1000
$\mathbf{s}_1 = [-1,-1,-1,-1]$	0	4	4	8	4	8	8	12	4	8	8	**12**	8	**12**	**12**	**16**
$\mathbf{s}_2 = [-1,-1,-1,+1]$	4	0	8	4	8	4	12	8	8	4	**12**	8	**12**	8	**16**	**12**
$\mathbf{s}_3 = [-1,-1,+1,-1]$	4	8	0	4	8	12	4	8	8	**12**	4	8	**12**	**16**	8	**12**
$\mathbf{s}_4 = [-1,-1,+1,+1]$	8	4	4	0	12	8	8	4	**12**	8	8	4	**16**	**12**	**12**	8
$\mathbf{s}_5 = [-1,+1,-1,-1]$	4	8	8	12	0	4	4	8	8	**12**	**12**	**16**	4	8	8	**12**
$\mathbf{s}_6 = [-1,+1,-1,+1]$	8	4	12	8	4	0	8	4	**12**	8	**16**	**12**	8	4	**12**	8
$\mathbf{s}_7 = [-1,+1,+1,-1]$	8	12	4	8	4	8	0	4	**12**	**16**	8	**12**	8	**12**	4	8
$\mathbf{s}_8 = [-1,+1,+1,+1]$	12	8	8	4	8	4	4	0	**16**	**12**	**12**	8	**12**	8	8	4
$\mathbf{s}_9 = [+1,-1,-1,-1]$	4	8	8	**12**	8	**12**	**12**	**16**	0	4	4	8	4	8	8	12
$\mathbf{s}_{10} = [+1,-1,-1,+1]$	8	4	**12**	8	**12**	8	**16**	**12**	4	0	8	4	8	4	12	8
$\mathbf{s}_{11} = [+1,-1,+1,-1]$	8	**12**	4	8	**12**	**16**	8	**12**	4	8	0	4	8	12	4	8
$\mathbf{s}_{12} = [+1,-1,+1,+1]$	**12**	8	8	4	**16**	**12**	**12**	8	8	4	4	0	12	8	8	4
$\mathbf{s}_{13} = [+1,+1,-1,-1]$	8	**12**	**12**	**16**	4	8	8	**12**	4	8	8	12	0	4	4	8
$\mathbf{s}_{14} = [+1,+1,-1,+1]$	**12**	8	**16**	**12**	8	4	**12**	8	8	4	12	8	4	0	8	4
$\mathbf{s}_{15} = [+1,+1,+1,-1]$	**12**	**16**	8	**12**	8	**12**	4	8	8	12	4	8	4	8	0	4
$\mathbf{s}_{16} = [+1,+1,+1,+1]$	**16**	**12**	**12**	8	**12**	8	8	4	12	8	8	4	8	4	4	0

4.4.3 Theoretical Performance Comparison Between the Proposed Mapping and the Anti-Gray Mapping

As mentioned earlier, the anti-Gray mapping is superior to Gray mapping for a QPSK constellation in the conventional BICM-ID systems. It is therefore of interest to compare the anti-Gray mapping and the proposed multi-dimensional mapping of an m-cube.

First, with the anti-Gray mapping of each QPSK symbol in an m-cube signal, it can be seen that the parameters $\delta_1^{(A)}(\Psi, \xi)$, $\delta_1^{(B)}(\Psi, \xi)$ and $\delta_2(\Psi, \xi)$ are independent of m. This property also holds for the parameters that affect the performance of the first iteration mentioned before. This implies that anti-Gray mapping results in the same performance for a BICM-ID system employing an m-cube constellation. Furthermore, it can be seen that $\delta_1^{(A)}(\Psi, \xi)$ and $\delta_1^{(B)}(\Psi, \xi)$ are identical for the anti-Gray mapping. Thus, there is no difference in the performance of a BICM system employing m-cube and anti-Gray mapping over the two different Rayleigh fading channels.

Next, observe that the anti-Gray mapping of the QPSK symbols is identical to the proposed mapping for the simplest case of $m = 2$. Furthermore, since all the achievable bounds in (4.26), (4.28) and (4.30) for the parameters $\delta_1^{(A)}(\Psi, \xi)$, $\delta_1^{(B)}(\Psi, \xi)$ and $\delta_2(\Psi, \xi)$ decrease as m increases, the asymptotic performance of a BICM-ID system with a hypercube constellation is improved by simply increasing m (i.e., the size of the hypercube).

Although the asymptotic performance of the proposed mapping can be significantly improved by increasing m, the performance after the first iteration becomes worse. This means that it might need a higher SNR to make the iterations work and a larger number of iterations to converge to the asymptotic performance. The worse performance of the first iteration when m increases can be explained by evaluating the parameter $N_{\min}$ described before. To compute this parameter for the proposed mapping, some properties of the nearest neighbors of a vertex in an m-cube are discussed next.

Consider any vertex $\hat{\mathbf{s}}_i = \left[-1, \hat{s}_{i_1}, \ldots, \hat{s}_{i_{m-1}}\right] \in \hat{\Psi}$ with label $\hat{\mathbf{a}}_i = \left(b, \hat{a}_{i_1}, \ldots, \hat{a}_{i_{m-1}}\right)$. It is clear that $\hat{\mathbf{s}}_i$ has $(m-1)$ vertices $\hat{\mathbf{s}}_j \in \hat{\Psi}$ as its nearest neighbors and the labels of $\hat{\mathbf{s}}_i$ and $\hat{\mathbf{s}}_j$ differ in 2 bits, including the difference in the lowest significant position. Denote the mth nearest neighbor of $\hat{\mathbf{s}}_i$ as $\tilde{\mathbf{n}}_i$ whose coordinate is $\tilde{\mathbf{n}}_i = \left[+1, \hat{s}_{i_1}, \ldots, s_{i_{m-1}}\right]$. Since the first coordinate of $\tilde{\mathbf{n}}_i$ equals $+1$, it can be seen that $\tilde{\mathbf{n}}_i \in \tilde{\Psi}$. Hence, its label $\tilde{\mathbf{l}}_i = \left(q, \tilde{l}_{i_1}, \ldots, \tilde{l}_{i_{m-1}}\right)$ has the following properties, which follows from (4.32) and (4.33):

$$\tilde{l}_{i_k} = \frac{-\hat{s}_{i_k} + 1}{2} = 1 - \frac{\hat{s}_{i_k} + 1}{2} = 1 - \hat{a}_{i_k} \tag{4.34}$$

Let $A = \sum_{k=1}^{m-1} \hat{a}_{i_k}$ and $L = \sum_{k=1}^{m-1} \tilde{l}_{i_k}$. It then follows that b equals 0 or 1 when A is even or odd, respectively. Similarly, q equals 1 or 0 when L is even or odd, respectively. From (4.34) one has

$$A + L = m - 1 \tag{4.35}$$

Thus, it is clear that $b = q$ when m is even, while $b \neq q$ when m is odd. Equivalently, it can be seen that when m is odd, the label Hamming distance between any vertex $\hat{\mathbf{s}}_i$ and its mth nearest neighbor is m. On the other hand, when m is even, this label Hamming distance is $(m-1)$ and their labels differ in all bits except the lowest significant bit. Similarly, this property can be obtained for $\tilde{\mathbf{s}}_i$. Thus the above property holds for any vertex $\mathbf{s}_i$ in the m-cube with the proposed mapping.

From the above properties, the parameter $N_{\min}$ can be computed as follows.

- m is even: For any vertex $\mathbf{s}_i$, the number of nearest vertices whose labels differ in the lowest significant bit is $(m-1)$. Moreover, for any k, $2 \leq k \leq m$, the number of nearest vertices whose labels differ in the kth bit is 2. Therefore,

$$N_{\min} = \frac{1}{m}\left[(m-1) + (m-1)\cdot 2\right] = \frac{3(m-1)}{m} \tag{4.36}$$

- m is odd: For any vertex $\mathbf{s}_i$, the number of nearest vertices whose labels differ in the lowest significant bit is m. For any k, $2 \leq k \leq m$, the number of nearest vertices whose labels differ in the kth bit is 2. Therefore,

$$N_{\min} = \frac{1}{m}\left[m + (m-1)\cdot 2\right] = \frac{3m-2}{m} \tag{4.37}$$

Notice that for the BICM-ID systems considered in this chapter, the value of m is always even, hence, $N_{\min}$ in (4.36) is of interest. For a conventional BICM-ID system employing QPSK constellation with anti-Gray mapping, then $m = 2$ and $N_{\min} = 1.5$. Observe that the value of $N_{\min}$ always increases as m increases, and for a large m, $N_{\min} \approx 3$.

Finally, the comparison for the BICM-ID systems operating over the two different Rayleigh fading channels can also be made by comparing $\delta_1^{(A)}(\Psi, \xi)$ and $\delta_1^{(B)}(\Psi, \xi)$. Using the approximation $(1 + x)^{-1} \approx x^{-1}$ when $x \to \infty$, (4.26) and (4.28) simplify to:

$$\delta_1^{(A)}(\Psi, \xi) \approx \frac{1}{m} \cdot \frac{N_0^n}{2^n}(2m - 1) \tag{4.38}$$

and

$$\delta_1^{(B)}(\Psi, \xi) \approx \frac{N_0(m + 1)}{m^2} \tag{4.39}$$

It then follows that:

$$\gamma = \frac{\delta_1^{(A)}(\Psi, \xi)}{\delta_1^{(B)}(\Psi, \xi)} \approx \frac{N_0^{n-1}(2m - 1)m}{2^n(m + 1)} \tag{4.40}$$

Note that for $m > 2$ (i.e., $n > 1$) and when N_0 is small enough, $\gamma << 1$. Therefore, with respect to the asymptotic performance, the system employing the proposed mapping performs better over a Rayleigh fast fading channel. This result is similar to the performance of a BICM-ID system with signal space diversity studied in [15]. It is also consistent with the fact that the diversity order dominates the performance of a communications system over a fading channel [7].

4.5 Simulation and Numerical Results

This section provides the simulation and numerical results to confirm the advantage of the BICM-ID systems that employ the hypercube constellations together with the proposed mappings over different channel models. For comparison, the performance of the conventional systems with anti-Gray mapping is also provided. The bitwise interleaver with a length of 12,000 coded bits is designed according to the rules outlined in Chapter 3. Each point in the BER curves is simulated with 10^7 to

10^8 coded bits. The error floors calculated according to (2.7) are also plotted to show how the iterations converge. In calculating (2.7), the first 20 Hamming distances of the convolutional codes are retained to guarantee the accuracy of the error bounds.

4.5.1 Performance over an AWGN Channel

For the AWGN channel model, we first use a very simple, rate-1/2, 4-state convolutional code with the generator sequences $\mathbf{g}_1 = 5$ and $\mathbf{g}_2 = 7$. Figure 4.5 presents the performance of a BICM-ID system employing a 4-cube constellation and the proposed mapping (see Table 4.1). Specifically, shown in the figure are the BER of the system after 1, 5, 9 and 12 iterations. For comparison, the performance after 12 iterations of the conventional system using anti-Gray mapping of QPSK is also plotted.

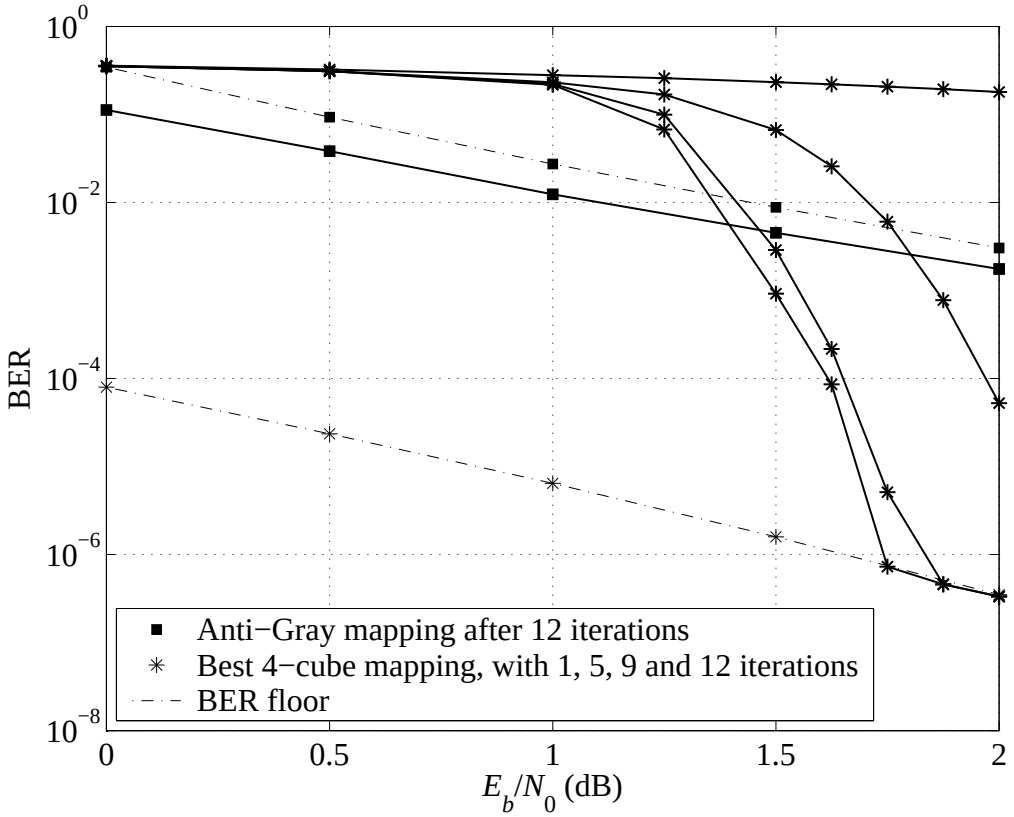

Figure 4.5 BER performance of BICM-ID systems over an AWGN channel: a rate-1/2, 4-state convolutional code and different mappings of a 4-cube.

It can be seen from Fig. 4.5 that, the simulation results converge to the error

floor bound for the proposed mapping in the range of the BER of interest (around 10^{-6}). For the anti-Gray mapping, the error floor bound underestimates the actual BER after 12 iterations. This is merely due to the poor performance of the anti-Gray mapping in the range of the signal-to-noise (SNR) shown in Fig. 4.5. The tightness of the error floor bound should be observed at a higher SNR region for the anti-Gray mapping. It is also obvious from Fig. 4.5 that a significant coding gain is achieved by the proposed mapping over the anti-Gray mapping.

To see what is the coding gain provided by the proposed mapping over the conventional anti-Gray mapping at a practical level of BER, Fig. 4.6 plots the error floor bounds for the systems employing the two mappings over the SNR range from 0 to 6 dB. Observe that with the selected 4-state convolutional code, the coding gain offered by the best 4-cube mapping at the BER level of 10^{-6} is about 3.8 dB. It should be emphasized that this coding gain is achieved with a very small increase in the receiver complexity which is only required to address the soft-output demodulation of a higher-order hypercube constellation.

It is also, of interest, to consider the trade-off between the complexity of the soft-output demodulator and that of the SISO decoder (i.e., using a higher-dimensional cube versus using a convolutional code with a larger constraint-length). Figure 4.7 compares the BER performance for two systems: the first one employs the proposed 4-cube mapping together with a 4-state convolutional code and the second one employs the proposed 6-cube mapping and a 2-state code (the simplest possible convolutional code). Observe that there is not much difference in the performance of the two systems after 12 iterations. Similar to Fig. 4.5, a significant coding gain over the conventional systems can also be achieved with the use of 6-cube mapping together with the 2-state code. In terms of complexity, the first system that uses 4-cube and 4-state code might be preferred. This is because, while the complexity of the SISO decoder of the first system is about twice as that of the second one, the complexity of its demodulator is only about one quarter of that of the demodulator in the second system.

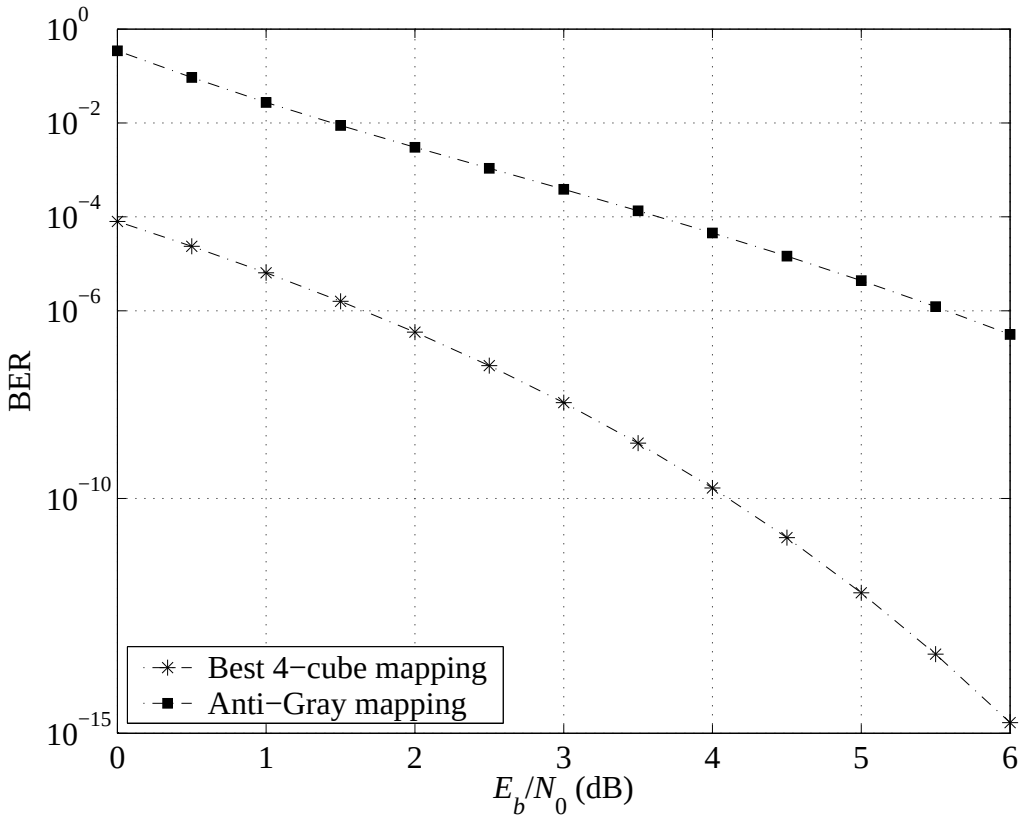

Figure 4.6 Error floors bounds for different mappings of a 4-cube and with a 4-state convolutional code over an AWGN channel.

It can also be observed from Fig. 4.5 that these two systems have different convergence behaviors. The use of a more complicated convolutional code (i.e., with a larger memory length) should result in a lower error floor due to the larger Hamming distances of the code. But because of the bigger error multiplicity, it needs a higher SNR and more iterations for the BER to converge to the very low error floor. However, the convergence behavior is also affected by using a higher-dimensional constellation as analyzed in the previous section. In the second system, a simple code leads to an earlier convergence compared to the first system. On the other hand, its higher-dimensional constellation results in the opposite influence. However, the effect of the code memory can compensate for that of the use of a higher-dimensional constellation. Hence, the system with 6-cube is preferred with respect to the convergence behavior as observed in Fig. 4.7. The above observation is based on the simulation results, but the interested reader is referred to the EXIT chart technique in [18] to

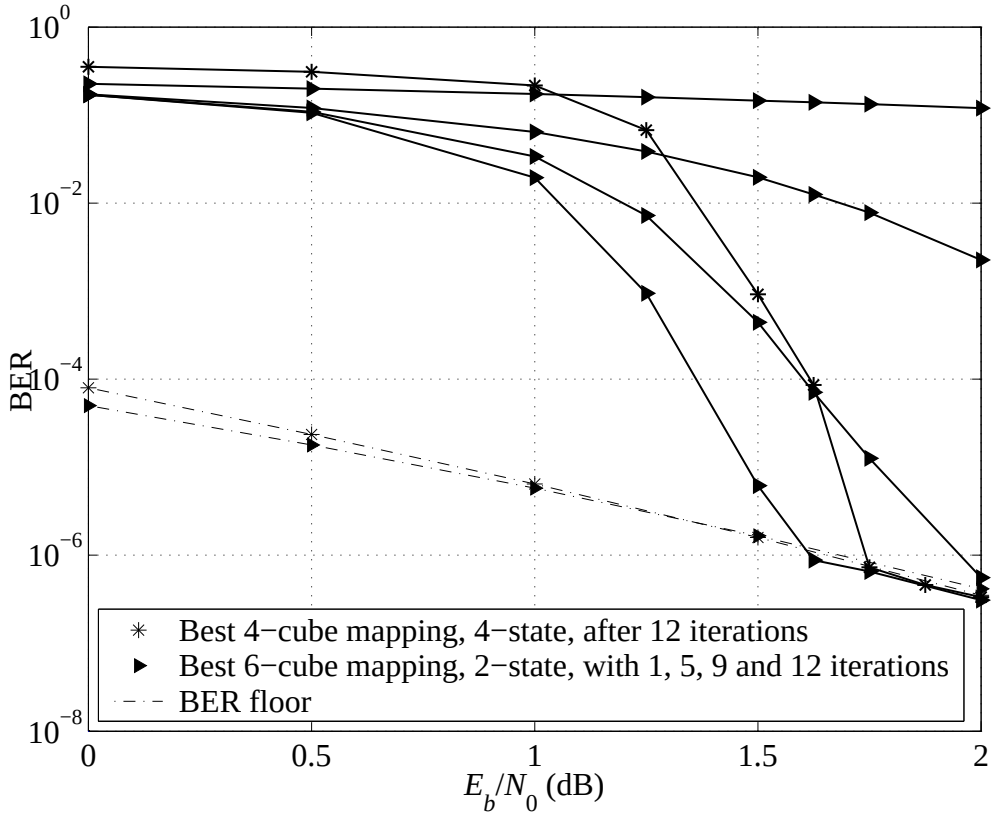

Figure 4.7 BER performance over an AWGN channel: Two systems with different code memories and different hypercube mappings.

investigate the convergence behavior of BICM-ID systems.

4.5.2 Performance over Rayleigh Fading Channels

First, the case in which QPSK symbols fade independently (i.e., fast fading channel) is considered. Figure 4.8 presents the performance of a BICM-ID system employing the proposed 4-cube mapping and a rate-1/2, 4-state convolutional code. More specifically, shown in the figure are the BER performances after 1, 5, 9 and 12 iterations. As before, the performance of the system using the anti-Gray mapping of QPSK after 12 iterations is provided and the error floor bounds calculated from (2.7) are also shown. It can be seen from Fig. 4.8 that, the simulation results converge to the error floor bound for the 4-cube mapping in the range of the BER of interest (around 10^{-5} to 10^{-6}). Similar to the case of an AWGN channel, the error floor bound

underestimates the actual BER performance for the anti-Gray mapping. The reason is also due to the poor performance of the anti-Gray mapping in the range of the SNR shown in Fig. 4.8. As can be seen clearly, a significant coding gain is obtained by the 4-cube mapping over the anti-Gray mapping. The approximated coding gain offered by the best 4-cube mapping can be determined from Fig. 4.9, which shows the error floor bounds for the two systems over the SNR range from 2 to 12 dB. With the selected 4-state convolutional code, the coding gain is about 7.4 dB at the BER level of 10^{-6}.

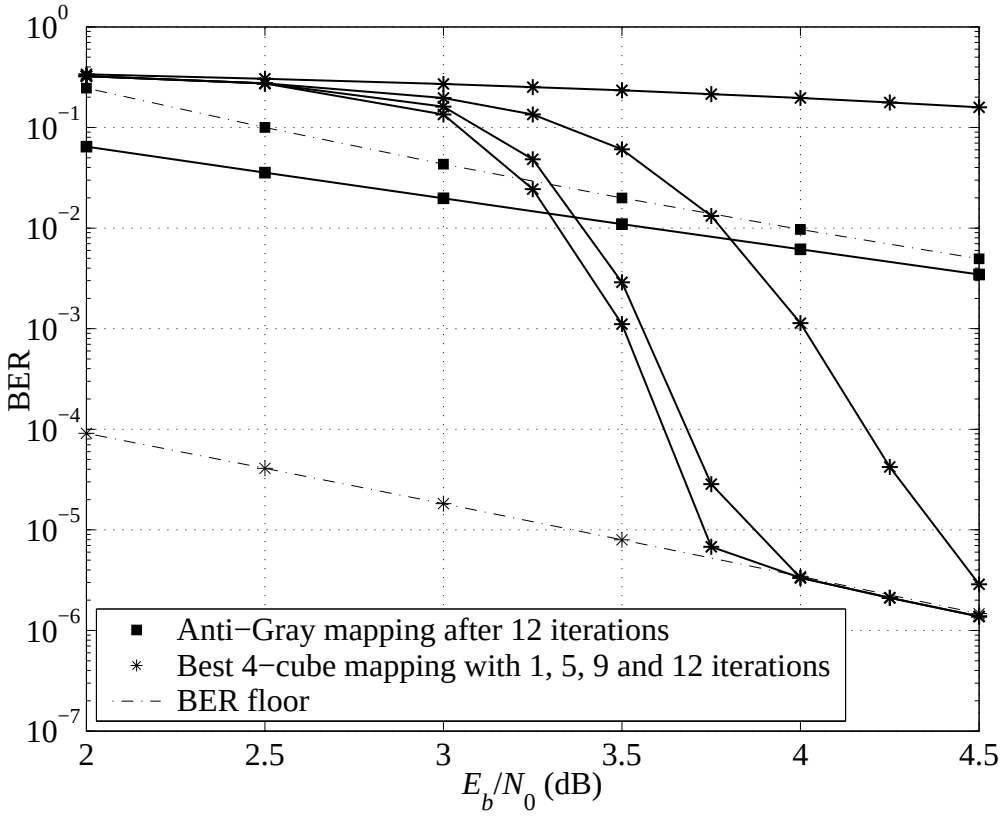

Figure 4.8 BER performance of a BICM-ID system employing a 4-cube constellation over a Rayleigh fast fading channel.

As in the case of an AWGN channel, Fig. 4.10 compares the performance of two systems that use 4-cube mapping/4-state CC and 6-cube mapping/2-state CC, respectively. Again, a significant coding gain compared to the conventional system can also be obtained by the use of 6-cube mapping together with a 2-state code. There is a slight difference for the asymptotic performance of these two systems, but the

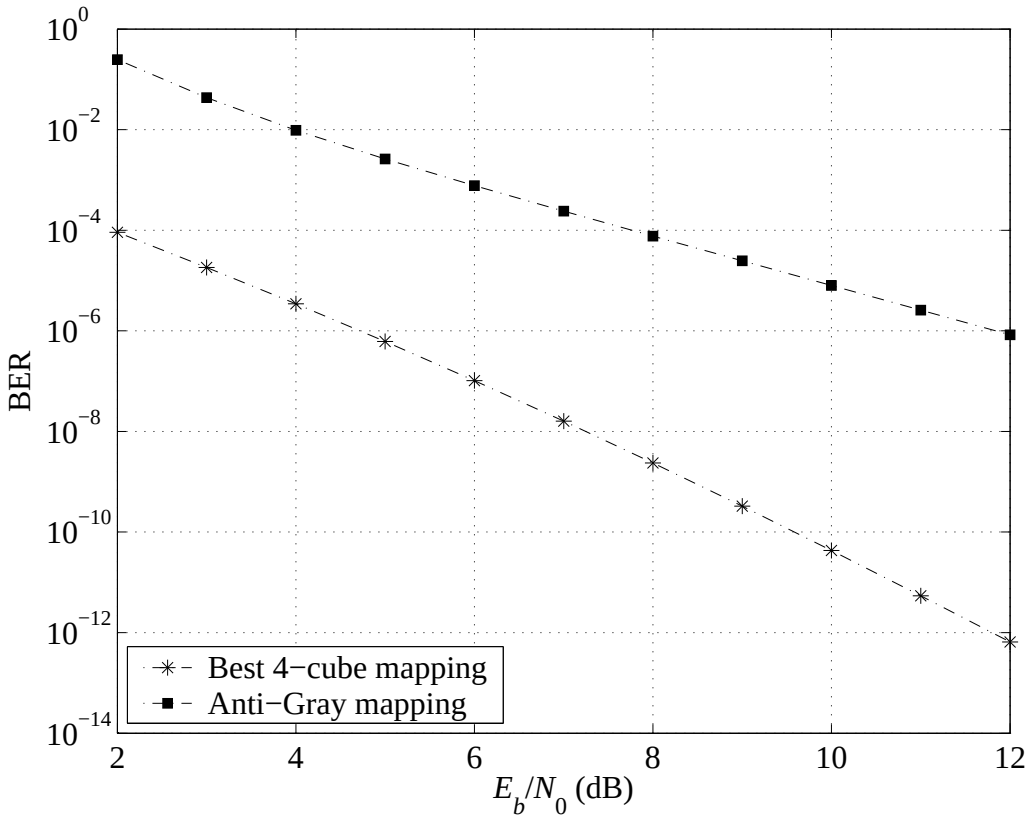

Figure 4.9 Error floor bounds for different mappings of a 4-cube and with a 4-state convolutional code over a Rayleigh fast fading channel.

first one is preferred in terms of the implementation complexity. However, the latter one is superior with respect to the convergence behavior as was seen in an AWGN channel.

For the case of Rayleigh fading channel in which QPSK components fades simultaneously (i.e., quasistatic fading channel), Fig. 4.11 shows the BER performance after 12 iterations of the system employing the best 4-cube mapping together with its error floor bound. Also for comparison, the performance after 12 iterations and the error bound for the system operating over a fast fading channel also shown. It is observed that for the anti-Gray mapping, there is no difference in BER performance between the two models of Rayleigh fading channels. This observation is consistent with the analytical evaluation made in the previous section. It can be seen from Fig. 4.11 that, the BER performance with iterations work better and it converges to the

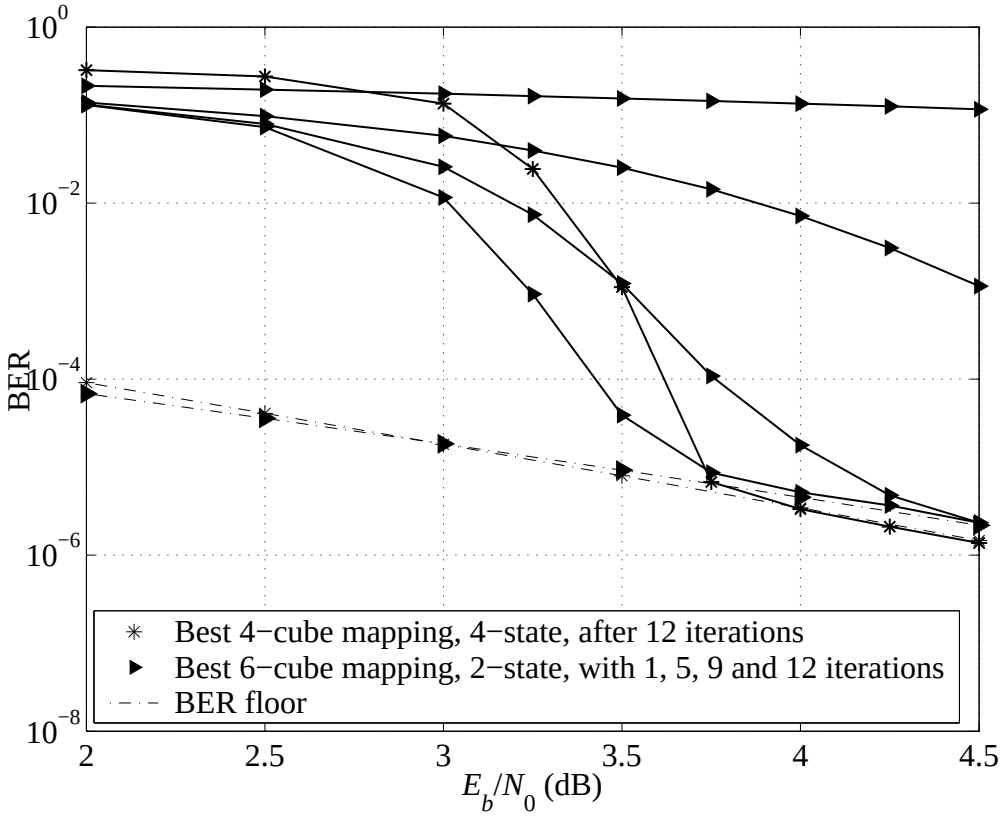

Figure 4.10 BER performance comparison over a Rayleigh fast fading channel: Two systems with different code memories and different hypercube mappings.

error floor faster for the case of a Rayleigh quasistatic fading channel than for the case of a Rayleigh fast fading channel. However, its error floor bound is higher as expected.

Finally, the coding gain over the conventional BICM-ID system employing the anti-Gray mapping calculated from error floor bounds for a Rayleigh quasistatic fading channel is about 4dB at the BER level of 10^{-6}. Although, this is still a significant gain, comparing to that of a Rayleigh fast fading channel, the coding gain drops about 3.4dB.

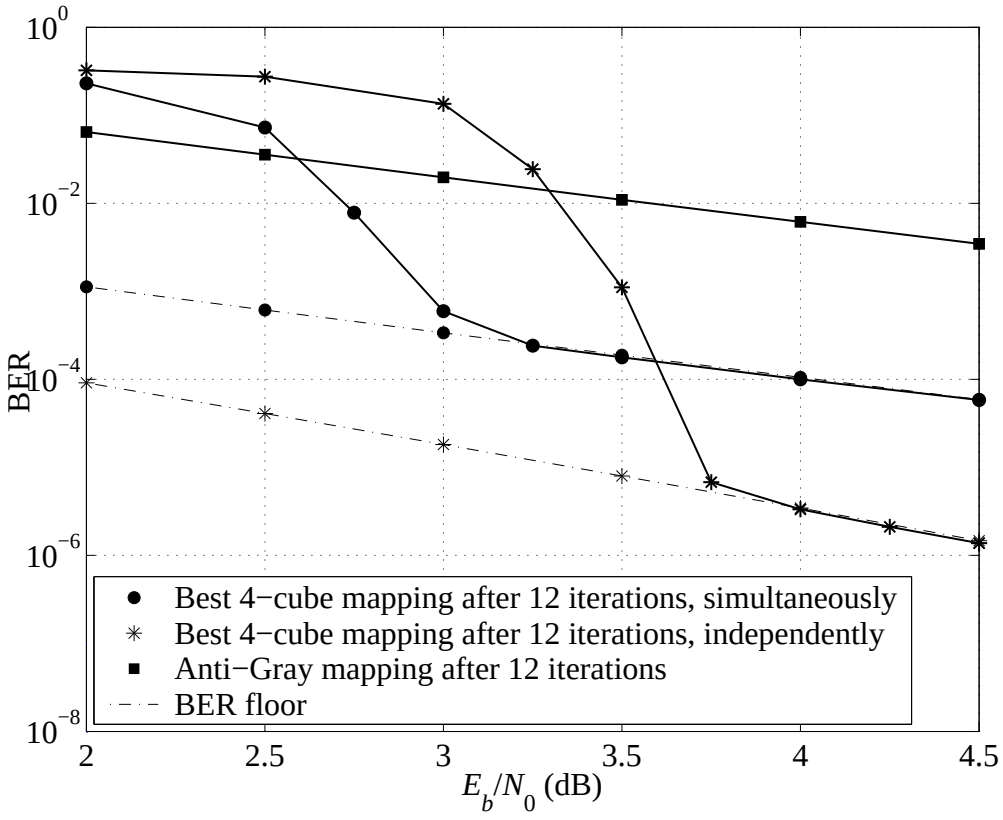

Figure 4.11 Performances of the proposed BICM-ID system over different fading channels.

5. Conclusions and Suggestions for Further Research

5.1 Conclusions

This thesis addressed the signal mapping problem to improve the error performance of bit-interleaved coded modulation (BICM) and BICM with iterative decoding (BICM-ID) over AWGN and Rayleigh fading channels. The influence of signal mapping was studied with two different approaches. The first approach relies on the upper bound of the system's error performance, whereas the second one is based on the technique of bitwise mutual information.

The effect of signal mapping on the error performance of BICM systems was first studied. It was observed that for any constellation, the Gray mapping and near Gray mapping are the optimal choices in terms of bit error rate performance. These mappings are suitable for both AWGN and Rayleigh fading channels. As an example, various Gray or near Gray mappings of many popular 8-ary constellations were introduced.

With BICM-ID systems, as demonstrated throughout the thesis, for given signal constellation, interleaver length, channel code and the channel itself, signal mapping plays an important role in determining the system performance. Starting with the conventional BICM-ID systems employing two-dimensional constellations, the techniques of error bound and mutual information were applied to provide an insight on how to obtain the best mapping of a given constellation with respect to the asymptotic performance. Good signal mappings of popular 8-ary constellations, including 8-PSK, cross 8-ary, (1,7) and optimum 8-ary were first proposed. A new 8-ary con-

stellation/mapping, referred to as the asymmetric 8-PSK, was then introduced to improve the error performance of BICM-ID. It was shown that the use of the new constellation/mapping results in about 1dB coding gain compared to other popular constellations/mappings with respect to the asymptotic performance. Nevertheless, it was concluded that for BICM-ID, the most suitable constellation/mapping depends on the length of the interleaver, the number of iterations one is willing to use and the desired BER level.

Signal mapping design was also carried out for BICM-ID systems with multi-dimensional constellations. More specifically, BICM-ID employing QPSK, viewed as hypercube constellations, was studied. To the best of our knowledge, it is the first time BICM-ID with multi-dimensional constellations is investigated. This work was motivated from the relatively small complexity of the soft-output demodulator compared to the SISO decoder and the flexibility of the signal mapping design in multi-dimensional constellations. In this thesis, the multi-dimensional constellation is a hypercube because it is constructed from QPSK. By analytically evaluating the error performance of the proposed systems over different channel models, the design criteria for the best mappings were established. It was pointed out that an exhaustive search to find the best mapping as previously done in conventional BICM-ID becomes intractable for a hypercube constellation. By studying the symmetry of a hypercube, universal conditions to find the best mapping of a hypercube constellation under different channel models were established. A general and simple algorithm to construct the best mapping was then introduced. It was demonstrated that, by employing the proposed mappings for four-dimensional and six-dimensional hypercubes and with very simple 2-state and 4-state convolutional codes, significant coding gains can be achieved over the conventional BICM-ID systems that use the anti-Gray mapping of QPSK symbols. More specifically, over an AWGN channel, the coding gain compared to the conventional systems is about 3.8dB. The coding gains over fast Rayleigh fading and quasistatic Rayleigh fading channels are 7.4dB and 4.0dB, respectively. Such coding gains are obtained with a very small increase in the receiver complexity

which is required to address the soft-output demodulation of a higher-order constellation. The tightness of the error bounds derived in this thesis for BICM-ID with multi-dimensional constellations was also verified.

5.2 Suggestions for Further Research

This thesis only considers the simplest case of multi-dimensional constellations constructed from QPSK. When bandwidth efficiency is the primary concern, it is of interest to study the signal mapping problems in multi-dimensional constellations built from other conventional higher order constellations (such as M-PSK, M-QAM). For these larger constellations, the binary switching algorithm (BSA) [35] might be useful. However, this algorithm only provides the locally optimized mappings. Therefore, the question on whether there exists an algorithm to construct the optimal mappings needs to be answered and it should be an interesting topic for further studies.

Recently, as further extensions of BICM and BICM-ID, bit-interleaved space time coded-modulation with iterative decoding (BI-STCM-ID) has been studied over Rayleigh block fading multi-input multi-output (MIMO) channels [36]. The signal constellations and mappings were still only studied in two-dimensional signal space. The performance of MIMO systems can certainly be further improved by extending the idea of multi-dimensional constellations and mappings proposed in this thesis.

Finally, for conventional BICM-ID systems and BICM-ID systems employing hypercube constellations in this thesis, the upper bound of the error performance was derived based on the assumption of infinite interleaver length and ideal feedback from the SISO decoder to the soft-output demodulator. This bound is only tight at high SNR, or equivalently, at the region of convergence. Therefore, obtaining a tighter bound for finite interleaver length of BICM-ID systems would be very attractive.

A. Appendix A: A Review of Mutual Information and Channel Capacity

In general, the mutual information between two random variables is a measure of information provided by one random variable about the other random variable. As an example, consider two discrete random variables $\mathbf{X}$ and $\mathbf{Y}$ with possible outcomes $\{\mathbf{x}_k\}_{k=1}^{M}$ and $\{\mathbf{y}_j\}_{j=1}^{N}$, respectively. If we observe some outcome $\mathbf{Y} = \mathbf{y}_j$, the mutual information $I(\mathbf{x}_k, \mathbf{y}_j)$ determines quantitatively the amount of information that the occurrence of the event $\mathbf{Y} = \mathbf{y}_j$ provides about the event $\mathbf{X} = \mathbf{x}_k$. This information is given as:

$$I(\mathbf{x}_k, \mathbf{y}_j) = \log \frac{P(\mathbf{x}_k|\mathbf{y}_j)}{P(\mathbf{x}_k)} \tag{A.1}$$

where $P(\mathbf{x}_k)$ is the probability mass function (PMF) that the discrete random variable $\mathbf{X}$ receives the value $\mathbf{x}_k$ and $P(\mathbf{x}_k|\mathbf{y}_j)$ is the conditional PMF of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}_j$. The unit of mutual information is bits or nats, depending on the base (2 or e) of the logarithm.

For the channel with discrete inputs $\mathbf{S} = \{\mathbf{x}_k\}, 1 \leq k \leq M$ and the continuous output $\mathbf{R} = \{\mathbf{r}\}$, the average mutual information provided by the output $\mathbf{R}$ about the input $\mathbf{S}$ is:

$$I(\mathbf{S}, \mathbf{R}) = \sum_{k=1}^{M} \int_{-\infty}^{+\infty} p(\mathbf{r}|\mathbf{s}_k) P(\mathbf{s}_k) \log \frac{p(\mathbf{r}|\mathbf{s}_k)}{p(\mathbf{r})} \mathrm{d}\mathbf{r} \tag{A.2}$$

where $P(\mathbf{s}_k)$, $p(\mathbf{r}|\mathbf{s}_k)$ and $p(\mathbf{r})$ are defined similarly as in Chapter 2.

The value of $I(\mathbf{S}, \mathbf{R})$ maximized over the set of input symbol probabilities $P(\mathbf{s}_k)$

is called the capacity of the channel. It is denoted by C and given below:

$$C = \max_{P(\mathbf{s}_k)} \sum_{k=1}^{M} \int_{-\infty}^{+\infty} p(\mathbf{r}|\mathbf{s}_k) P(\mathbf{s}_k) \log \frac{p(\mathbf{r}|\mathbf{s}_k)}{p(\mathbf{r})} \mathrm{d}\mathbf{r} \tag{A.3}$$

Over an AWGN channel, it has been shown that [23] the capacity of the channel is obtained when the input symbols are equally probable. However, nothing can be said in general about the set $\{P(\mathbf{s}_k)\}$ that maximizes the average mutual information. In this thesis, however, the channel capacity C refers to the average mutual information in (A.2).

B. Appendix B: The Bitwise Mutual Information I_0 and I_{m-1} for an M-ary Constellation

B.1 The Bitwise Mutual Information I_0

For an M-ary constellation, the explicit expressions for the bitwise mutual information I_0 can be obtained as follows. Let $\{v_1, \ldots, v_m\}$ denote the m-bit label of one symbol in an M-ary constellation Ψ. Then the unconditional bitwise mutual information I_0 given the fading amplitude g is:

$$\begin{aligned} I_{0|g} &= \frac{1}{m}\sum_{k=1}^{m} I(v_k;\mathbf{r}|g) = \frac{1}{m}\sum_{k=1}^{m}\left[\frac{1}{2}\sum_{b=0}^{1} I(v_k=b;\mathbf{r}|g)\right] \\ &= \frac{1}{2m}\sum_{k=1}^{m}\sum_{b=0}^{1} I(v_k=b;\mathbf{r}|g) \end{aligned} \tag{B.1}$$

where $I(c_i = b;\mathbf{r}|g)$ is calculated as:

$$I(v_k=b;\mathbf{r}|g) = \int_{-\infty}^{\infty}\left(\left[\frac{2}{M}\sum_{\mathbf{s}_i\in\Psi_b^k} p(\mathbf{r}|\mathbf{s}_i,g)\right]\cdot\log_2\frac{\left[\frac{2}{M}\sum_{\mathbf{s}_i\in\Psi_b^k} p(\mathbf{r}|\mathbf{s}_i,g)\right]}{p(\mathbf{r}|g)}\right)\mathrm{d}\mathbf{r} \tag{B.2}$$

Recall that Ψ_b^k in (B.2) denotes the subset of Ψ that contains all symbols whose labels have the value $b \in \{0,1\}$ at the position k. The functions $p(\mathbf{r}|\mathbf{s}_i,g)$ and $p(\mathbf{r}|g)$ are given in (2.3) and (2.13), respectively. The unconditional bitwise mutual information I_0 is then obtained by averaging $I_{0|g}$ over the fading amplitude g.

B.2 The Bitwise Mutual Information I_{m-1}

With the perfect knowledge of the other $(m-1)$ bits, an M-ary constellation is translated to BPSK constellations, each constellation contains 2 signal points whose

labels differ in only 1 bit. Therefore, the bitwise mutual information I_{m-1} can be computed by simply taking the average of the symbol mutual information of all BPSK constellations. This value for an M-ary constellation Ψ with the labelling scheme ξ can be computed as follows:

$$I_{m-1} = \frac{1}{m2^m} \sum_{\mathbf{s}_i \in \Psi} \sum_{k=1}^{m} I_k(\mathbf{s}_i) \tag{B.3}$$

where $I_k(\mathbf{s}_i)$ is interpreted as the average mutual information of the BPSK constellation consisting of two signal symbols $\mathbf{s}_i$ and $\mathbf{s}_{j(i,k)}$.

For the case of an AWGN channel, $I_k(\mathbf{s}_i)$ is exactly the channel capacity with two possible inputs $\mathbf{s}_i$ and $\mathbf{s}_{j(i,k)}$ [23], since all the signal symbols $\mathbf{s}_i \in \Psi$ are equally likely. This value can be numerically computed using a one-fold integral as demonstrated in [23].

For a Rayleigh fading channel, $I_k(\mathbf{s}_i)$ can be computed by numerical integration with a two-fold integral, but the accuracy of the numerical integration can be an issue. In this case, by applying the power series expansion techniques in [37], $I_k(\mathbf{s}_i)$ can be determined with any desired accuracy level as follows:

$$I_k(\mathbf{s}_i) = \frac{1}{2}\sqrt{\frac{\delta_{i,j(i,k)}}{1+\delta_{i,j(i,k)}}} \sum_{n=0}^{\infty} \frac{1}{2^n} \cdot \frac{n!}{g(\delta_{i,j(i,k)})[g(\delta_{i,j(i,k)})+1]\cdots[g(\delta_{i,j(i,k)})+n]} \tag{B.4}$$

where $g(x) = \frac{1}{2}\left(1+\sqrt{1+\frac{1}{x}}\right)$, $\delta_{i,j(i,k)} = \frac{d^2_{i,j(i,k)}}{4N_0}$ and N_0 is one-sided power spectral density of AWGN. Recall that $d_{i,j(i,k)}$ is the Euclidean distance between the two signal symbols $\mathbf{s}_i$ and $\mathbf{s}_{j(i,k)}$, respectively. The summation in (B.4) can be truncated to the first N terms with an error of at most $\sqrt{\frac{\delta_{i,j(i,k)}}{1+\delta_{i,j(i,k)}}}2^{-N}$ [37]. Thus, I_{m-1} can be efficiently computed with a high accuracy.

C. Appendix C: The Bitwise Mutual Information for a Hypercube Constellation

This Appendix studies the bitwise mutual information I_{m-1} of an m-hypercube to find the condition of the best mapping with respect to the asymptotic performance. With the perfect knowledge of the other $(m-1)$ bits, an m-hypercube constellation is also translated to BPSK constellations. The bitwise mutual information therefore can be computed as in (B.3). In the following, the bitwise mutual information in (B.3) is further evaluated for each channel model considered in this thesis.

In the case of an AWGN channel, $I_k(\mathbf{s}_i)$ is the channel capacity with two possible inputs $\mathbf{s}_i$ and $\mathbf{s}_{j(i,k)}$ [23]. For a fixed value of N_0, $I_k(\mathbf{s}_i)$ is a function of the Euclidean distance $d_{i,j(i,k)}$ between two vertices $\mathbf{s}_i$ and $\mathbf{s}_{j(i,k)}$. Thus, it depends on the coordinate Hamming distance $H_{i,j(i,k)}$. Next, denote $I_k(\mathbf{s}_i)$ by $f_a(H_{i,j(i,k)})$, where $f_a(H_{i,j(i,k)})$ is understood as a function of $H_{i,j(i,k)}$. Furthermore, this function increases monotonically as $H_{i,j(i,k)}$ increases [23]. Also define:

$$I(\mathbf{s}_i) = \frac{1}{m}\sum_{k=1}^{m} I_k(\mathbf{s}_i) \tag{C.1}$$

Then I_{m-1} can be rewritten as follows:

$$I_{m-1} \;=\; \frac{1}{2^m}\sum_{i=1}^{2^m} I(\mathbf{s}_i) \tag{C.2}$$

Since for each vertex $\mathbf{s}_i$, there is one and only one vertex with coordinate Hamming distance m and there are m vertices with coordinate Hamming distance $(m-1)$ to $\mathbf{s}_i$, then,

$$I(\mathbf{s}_i) = \frac{1}{m}\sum_{k=1}^{m} I_k(\mathbf{s}_i) \le \frac{1}{m}\left[f_a(m) + (m-1)f_a(m-1)\right] \tag{C.3}$$

It then follows that:

$$I_{m-1} = \frac{1}{2^m} \sum_{i=1}^{2^m} I(\mathbf{s}_i) \leq \frac{1}{m} \left[f_a(m) + (m-1) f_a(m-1) \right] \tag{C.4}$$

Therefore, the ideal mapping that achieves equality in (C.4), if it exists, should satisfy *Condition 1.*

For a Rayleigh quasistatic fading channel, $I_k(\mathbf{s}_i)$ can be determined with any desired accuracy level as in (B.4). It should be mentioned that the parameter $\delta_{i,j(i,k)}$ can also be expressed in terms of $H_{i,j(i,k)}$, the coordinate Hamming distance between the two vertices $\mathbf{s}_i$ and $\mathbf{s}_{j(i,k)}$, as follows:

$$\delta_{i,j(i,k)} = \frac{d^2_{i,j(i,k)}}{4N_0} = \frac{H_{i,j(i,k)}}{N_0} \tag{C.5}$$

For a fixed level of additive noise N_0, it can be seen that $I_k(\mathbf{s}_i)$ depends only on the coordinate Hamming distance $H_{i,j(i,k)}$ between two vertices $\mathbf{s}_i$ and $\mathbf{s}_{j(i,k)}$. As in the case of an AWGN channel, also denote $I_k(\mathbf{s}_i)$ by $f_{rs}(H_{i,j(i,k)})$. Since $g(x)$ is a decreasing function[1] of x and $\frac{\delta_{i,j(i,k)}}{1+\delta_{i,j(i,)}}$ is an increasing function of $\delta_{i,j(i,k)}$, it follows from (B.4) that $I_k(\mathbf{s}_i) = f_{rs}(H_{i,j(i,k)})$ is an increasing function of $H_{i,j(i,k)}$. Similarly, by defining $I(\mathbf{s}_i)$ for each vertex $\mathbf{s}_i$, one obtains:

$$I_{m-1} = \frac{1}{2^m} \sum_{i=1}^{2^m} I(\mathbf{s}_i) \leq \frac{1}{m} \left[f_{rs}(m) + (m-1) f_{rs}(m-1) \right] \tag{C.6}$$

Clearly, the ideal mapping that achieves equality in (C.6), if it exists, should also satisfy *Condition 1.*

Finally, for the case of a Rayleigh fast fading channel, unfortunately, there is no closed form expression for $I_k(\mathbf{s}_i)$ and the numerical integration is needed. However, observe that for a BPSK constellation, a bigger Euclidean distance between the two signal points results in a higher average mutual information, it is expected that the best mapping in this case should also satisfy *Condition 1.*

[1]See Appendix B.

References

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–883, May 1992.

[2] S. L. Goff, "Signal constellations for bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 49, pp. 307–313, Jan. 2003.

[3] S. Lin and D. J. Costello, *Error Control Coding*. Prentice-Hall, Inc. Englewood Cliffs, New Jersey 07632, 1983.

[4] J. L. Massey, "Coding and modulation in digital communications," in *Proc. of International Zurich Seminar on digital communications*, March, 1974.

[5] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[6] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 371–377, May 1977.

[7] D. Divsalar and M. K. Simon, "The design of trellis coded MPSK for fading channels: Performance criteria," *IEEE Trans. Commun.*, vol. 36, pp. 1004–1011, Sept. 1988.

[8] K. Boulle and J. C. Belfiore, "Modulation schemes designed for the Rayleigh channel," in *Proc. CISS'92*, pp. 288–293, March, 1992.

[9] B. D. Jelicic and S. Roy, "Design of trellis coded QAM for flat fading and AWGN channels," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 192–201, Feb. 1994.

[10] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.

[11] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correction coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Commun.*, pp. 1064–1070, 1993.

[12] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Letters*, vol. 1, pp. 169–171, Nov. 1997.

[13] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE J. Select. Areas in Commun.*, vol. 17, pp. 715–724, Apr. 1999.

[14] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8PSK signaling," *IEEE Trans. Commun.*, vol. 50, pp. 1250–1257, Aug. 2002.

[15] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Select. Areas in Commun.*, vol. 19, pp. 944–957, May 2001.

[16] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE Global Telecommun. Conf.*, pp. 579–584, 1998.

[17] S. ten Brink, J. Speidel, and R. H. Yan, "Iterative demapping for QPSK modulation," *Electronics Letters*, vol. 34, pp. 1459–1460, July 1998.

[18] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information chart," *AEU Int. J. Electron. Commun*, vol. 54, pp. 389–398, Sept. 2000.

[19] P. Robertson and T. Worz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component code," *IEEE J. Select. Areas in Commun.*, vol. 16, pp. 206–218, Feb. 1998.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[21] N. H. Tran and H. H. Nguyen, "On signal constellations and mappings for 8-ary BICM over a Rayleigh fading channel," in *Department of Electrical Engineering, University of Saskatchewan, Saskatoon, Canada. Unpublished.*

[22] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Letters*, vol. 1, pp. 22–24, Jan. 1997.

[23] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2001.

[24] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.

[25] D. Divsalar and F. Pollara, "Turbo codes for PCS applications," in *Proc. ICC'93*, pp. 54–59, June, 1998.

[26] N. H. Tran and H. H. Nguyen, "Signal constellations and mappings for 8-ary BICM-ID over an AWGN channel," in *Proc. Wireless 2004*, pp. 464–472, July, 2004.

[27] H. Schulze, "System design for bit interleaved coded QAM with iterative decoding in a Rician fading channel," *European Transactions on Telecommunications*, vol. 14, pp. 119–129, Apr. 2003.

[28] M. K. Simon and D. Divsalar, "Some new twists to problems involving the Gaussian probability integral," *IEEE Trans. Commun.*, vol. 46, pp. 200–210, Feb. 1998.

[29] R. H. F. Udo Wachsmann and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1361–1391, July 1999.

[30] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Trans. Commun.*, vol. 22, pp. 28–38, Jan. 1974.

[31] S. L. Goff, A. Glavieux, and C. Berrou, "Turbo-code and high spectral efficiency modulation," in *Proc. Int. Conf. Communication*, June, 1994.

[32] L.-F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. 33, pp. 483–501, July 1987.

[33] S. S. Pietrobon, R. H. Deng, A. Lafanechere, G. Ungerboeck, and D. J. Costello, "Trellis-coded multidimensional phase modulation," *IEEE Trans. Inform. Theory*, vol. 36, pp. 62–89, Jan. 1990.

[34] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[35] F. Schreckenbach, N. Gortz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Letters*, vol. 7, pp. 593–595, Dec. 2003.

[36] Y. Huang and J. A. Ritcey, "Tight BER bound for iteratively decoded bit-interleaved space-time coded modulation," *IEEE Commun. Letters*, vol. 8, pp. 153–155, Mar. 2004.

[37] T. F. Wong, "Numerical calculation of symmetric capacity of Rayleigh fading channel with BPSK/QPSK," *IEEE Commun. Letters*, vol. 5, pp. 328–330, Aug. 2001.