

Coded Modulation

Motohiko Isaka

School of Science and Technology, Kwansei Gakuin Univ.,

isakam@kwansei.ac.jp

1 Introduction

As its name suggests, coded modulation is a subject treating the joint design of (error correcting) coding and modulation. The goal of this technique may depend on the scenario in which it is used, and we are particularly interested in bandwidth-efficient transmission from a transmitter to a receiver over the Gaussian and flat Rayleigh fading channels in this chapter.

A straightforward transmission scheme for such channels is to use a binary signaling together with error correcting codes over the binary field $\mathbb{F}_2 = \{0, 1\}$. In this case, designing good binary codes directly results in an improvement of error performance as the mapping from the finite field to the signal space is immediate. However, the use of multilevel signaling is essential in enhancing bandwidth efficiency. Also, for practical channels, the received signals inherently take analog values, and the codes that can efficiently utilize such information are mostly limited to binary codes (at least historically). This fact raises the problem of designing a coding scheme for non-binary alphabet by employing binary codes, preferably with a computationally efficient soft-decision decoding algorithm. This problem of combined coding and modulation was first suggested by Massey [1].

Coded modulation was explored in the mid-1970's by Ungerboeck who invented trellis coded modulation (TCM) [2][3] as well as Imai and Hirakawa [4] who proposed multilevel coding and multistage decoding. Another approach called bit-interleaved coded modulation (BICM) [5] was presented in 1992 by Zehavi [6] (originally intended for use on Rayleigh fading channels). These coding schemes were introduced primarily with the use of binary convolutional codes or relatively short linear block codes for which an efficient soft-decision maximum likelihood decoding algorithm is known. In this case, a distance-based design criterion has been pursued from various aspects.

Later in the mid-1990's, the invention of turbo codes [7][8] and the re-discovery of low-density parity-check (LDPC) codes [9][10] attracted interests in capacity-approaching performance with high bandwidth efficiency in the coding community. Since then, the use of turbo and LDPC codes, or the related iterative decoding schemes in the context of coded modulation has been intensively studied. The design and performance of such coding schemes may be substantially different from the traditional coded modulation schemes.

In the rest of this chapter, after providing preliminaries in Sec. 2, we illustrate the idea of the three coded modulation techniques: TCM in Sec. 3, multilevel coding and multistage decoding in Sec. 4, and BICM in Sec. 5.

Readers who are further interested in this topic are referred to the following publications among others: tutorial documents on TCM by the inventor are found in [12][13]. Survey papers published at the occasion of the golden jubilee of information theory in 1998 [14][15][16] review the first 50 years of coding techniques. Books that are entirely devoted to TCM include [17][18]. A monograph on BICM was recently published [19], and various bandwidth-efficient coding schemes are presented in [20][21][22][23]. Books on general coding theory [24] and

digital communications with emphasis on wireless channels [25][26] also have chapters on coded modulation techniques.

2 Preliminaries

2.1 Gaussian and Fading Channels

In this chapter, we focus on a communication system consisting of a sender and a receiver. Two channel models, Gaussian and flat Rayleigh fading channels, are considered and summarized in the following. The sender is supposed to transmit a signal X through the channel whose output Y is received. Both of these channel input and output are regarded as random variables, and the alphabet in which X and Y take values are denoted by $\mathcal{X}$ and $\mathcal{Y}$, respectively. In the following, the alphabet $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be one-dimensional or two-dimensional Euclidean space. For practical communication systems, the alphabet $\mathcal{X}$ of the input signal is a discrete set whose size is assumed to be $|\mathcal{X}| = 2^M$ for an integer M unless stated otherwise.

We impose a constraint on the average transmit power such that $\mathbb{E}_{P_X}[\|X\|^2] \leq S$, where $\|X\|^2$ represents the squared norm of X and $\mathbb{E}_{P_X}$ is the expectation with respect to the probability distribution P_X of X .

The output of the additive white Gaussian noise (AWGN) channel is given as $Y_i = X_i + Z_i$ for $i = 1, \dots, n$, where Z_i 's are independent Gaussian random variables of mean 0 and variance σ^2 .

In this chapter, we also consider the flat Rayleigh fading channel with coherent detection and perfect channel state information in which the amplitude of the transmitted signal X_i is multiplied by a factor H_i before the Gaussian noise Z_i is added. If H_i at each transmission is independent of the other factors, it is called fast (fully-interleaved) fading channel. In contrast, another channel model in which the values H_i 's remain the same in a block of symbols but are block-wise independently distributed is called block fading.

For a one-dimensional signal space, the signal set called Pulse Amplitude Modulation (PAM) is represented as $\mathcal{X} = \{\pm A, \pm 3A, \dots, \pm(2M-1)A\}$, where $A = \sqrt{S/(2 \sum_{i=1}^M (2i-1)^2)}$ is set to satisfy the constraint on the average transmit power. For a two-dimensional constellation, we focus only on PSK (Phase Shift Keying) which has the signal set $\mathcal{X} = \{(S \cos(\frac{2i}{M}\pi), S \sin(\frac{2i}{M}\pi)) : i = 0, \dots, M-1\}$.

In the following, for simplicity of description, we may let the average transmit power to be unity ($S = 1$).

2.2 Capacity of Signal Sets over the Gaussian Channel

The capacity of the AWGN channel under average transmit power constraint are reviewed from basic information theory [27] in this subsection. For simplicity, we only deal with transmission of one-dimensional signals, but a similar development is made for the two-dimensional case.

Suppose that we have no restriction on the channel input X in one-dimensional space $\mathbb{R}$ except that the average transmit power is limited to $\mathbb{E}[\|X\|^2] \leq S$. The capacity of this channel, denoted by C_{AWGN} , is defined as the maximum of the mutual information $I(X; Y)$ with respect to the probability density function f_X of the channel input X , and computed as

$$C_{\text{AWGN}} = \max_{f_X} I(X; Y) = \frac{1}{2} \log_2(1 + \text{SNR}), \quad (1)$$

in bits per channel use and the signal-to-noise ratio (SNR) is defined as S/σ^2 for this channel. The maximum Eq. (1) is achieved when f_X follows the Gaussian distribution with mean zero and variance S . From the channel coding theorem for memoryless channels, there exists a

sequence of encoder and decoder pairs such that the decoder correctly estimates the transmitted codeword except for arbitrarily small probability if the rate is smaller than C_{AWGN} . However, computationally efficient encoding and decoding schemes with such performance have not been discovered in the literature unfortunately largely because of the difficulty in designing codes over continuous alphabet (indeed, the transmitted signals should be Gaussian distributed as described above).

As a result, discrete channel input alphabets $\mathcal{X}$ are considered in essentially all the practical digital communications systems. For a fixed channel input alphabet $\mathcal{X}$, the mutual information between X and Y is computed as

$$I(X;Y) = \sum_{x \in \mathcal{X}} \int_{-\infty}^{\infty} f_{Y|X}(y|x) P_X(x) \log_2 \frac{f_{Y|X}(y|x)}{f_Y(y)} dy, \quad (2)$$

where $f_{Y|X}$ is the conditional probability density function of the channel. The most important special case in practice is the uniform input distribution $P_X(x) = 1/M$ for an M -ary signal set, on which we basically focus in this chapter except for Sec. 5.5. We denote the mutual information $I(X;Y)$ for $\mathcal{X}$ under this distribution by $I_{\mathcal{X}}$, which is sometimes called the (constrained) capacity of $\mathcal{X}$ for simplicity.

The constrained capacity $I_{\mathcal{X}}$ for various M -PAM with $M = 2^m$ under uniform input distribution over the alphabet $\mathcal{X}$ as well as the capacity of the AWGN channel C_{AWGN} are plotted in Fig. 1, where the horizontal axis represents the SNR in decibel. This is also called the Shannon-limit for each alphabet especially when the SNR which allows vanishingly small error probability is discussed at a fixed rate.

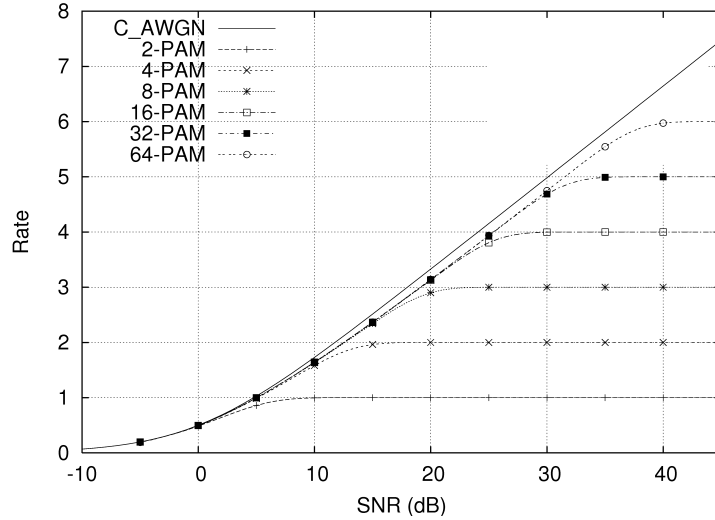


Figure 1: Constrained capacity of PAM over the AWGN channel

The constrained capacity $I_{\mathcal{X}}$ associated with the alphabet $\mathcal{X}$ is upper bounded by the entropy of the channel input, $H(X) = \log_2 |\mathcal{X}|$. It is indeed closely approached at high SNR provided that each signal point is transmitted by the uniform distribution on the alphabet. Accordingly, we need to employ alphabets of larger size in achieving high rate (bits/channel use) or high bandwidth efficiency.

Note also that at high rates the PAM constellations do not approach the capacity of the AWGN channel whose ideal channel input alphabet is the Gaussian distribution. At asymptotically high rates, the gap between C_{AWGN} and PAM constellations amounts to

$\pi e/6$ or 1.53dB, which corresponds to the ratio of the normalized second moment between the N -dimensional sphere to the cube for $N \rightarrow \infty$. Techniques to make up for this gap are generally called shaping [28] whose problem is defining signal sets in a near-spherical region in the multi-dimensional spaces rather than employing signal sets as the Cartesian product of PAM. Efficient two-dimensional constellations together with coded modulation techniques were reviewed in [11].

2.3 Overview of Coded Modulation Schemes

Before introducing the details of various coded modulation schemes, it would be instructive to describe them in a unified framework and see their structural similarity and difference. In the following of this chapter, for descriptive convenience, convolutional codes may be described as an equivalent block codes to which termination or truncation is applied.

Suppose that the message is a binary k -dimensional vector $\mathbf{u} = (u_1, \dots, u_k) \in \mathbb{F}_2^k$. Assuming n channel uses and transmission with rate $R = k/n$ [bits/channel use], the problem is simply to determine a mapping of these 2^k messages into distinct sequences of length n over the channel input alphabet $\mathcal{X}$. Rather than designing coding schemes directly over the alphabet $\mathcal{X}^n$, the encoding for the overall coded modulation system is performed through the use of binary linear codes by the following two steps.

1. the message vector $\mathbf{u} = (u_1, \dots, u_k) \in \mathbb{F}_2^k$ is mapped to n -dimensional vectors $\mathbf{c}_1, \dots, \mathbf{c}_M$, where $\mathbf{c}_i = (c_{i,1}, \dots, c_{i,n}) \in \mathbb{F}_2^n$ for $i = 1, \dots, M$. The encoder f is regarded as a one-to-one mapping from $\mathbb{F}_2^k$ to $\mathbb{F}_2^{Mn}$. Set of these vectors $\mathbf{c}_1, \dots, \mathbf{c}_M$ constitutes codeword(s) of certain linear codes over $\mathbb{F}_2$.
2. The binary vectors $\mathbf{c}_1, \dots, \mathbf{c}_M \in \mathbb{F}_2^n$ are mapped to an n -dimensional vector $\mathbf{x} = (x_1, \dots, x_n)$ over $\mathcal{X}$. Let $\phi : \mathbb{F}_2^M \rightarrow \mathcal{X}$ be a bijective map that brings M -bit strings to a signal point on $\mathcal{X}$ as $x_j = \phi(c_{1,j}, c_{2,j}, \dots, c_{M,j})$.

We denote the overall code over the alphabet $\mathcal{X}^n$ by $\mathcal{C} = \{\phi(f(\mathbf{u})) : \mathbf{u} \in \mathbb{F}_2^k\}$, and call each x_j a symbol over $\mathcal{X}$.

Upon receiving $\mathbf{y} = (y_1, \dots, y_n)$ as output of the noisy channel, the decoder determines the estimate $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_n)$ of the transmitted signal sequence $\mathbf{x}$. Let this decoding map be $g : \mathcal{Y}^n \rightarrow \mathcal{X}^n$. Note that because of the one-to-one nature of the encoding process, this is equivalent to computing the estimate $\hat{\mathbf{c}}_i = (\hat{c}_{i,1}, \dots, \hat{c}_{i,n})$ for $\mathbf{c}_i$, $i = 1, \dots, M$ as well as the one $\hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_k)$ for the original message $\mathbf{u}$.

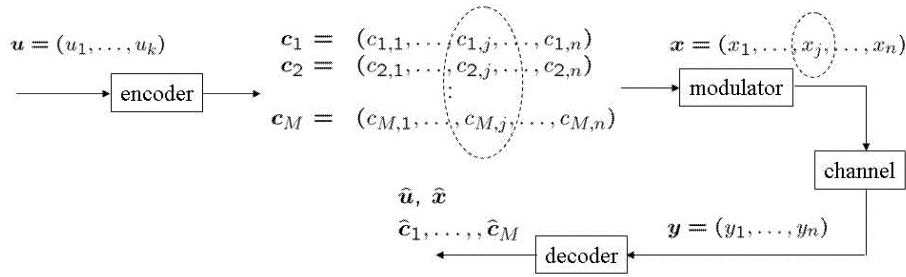


Figure 2: Coding, Modulation and Decoding

Loosely speaking, the study of coded modulation schemes is reduced to designing the encoder f as well as the mapping ϕ in this framework.

The encoder f and decoder g for each coded modulation scheme work as follows:

x	$-7A$	$-5A$	$-3A$	$-A$	$+A$	$+3A$	$+5A$	$+7A$
Gray mapping	000	001	011	010	110	111	101	100
natural mapping	000	001	010	011	100	101	110	111

Table 1: Mappings for 8-PAM ($A = \sqrt{S/21}$)

- for TCM, as shown in Fig. 5 of Sec. 3, some $N (\leq M)$, $\mathbf{c}_1, \dots, \mathbf{c}_N$ are the output of a binary convolutional encoder typically of rate $(N-1)/N$ while $\mathbf{c}_{N+1}, \dots, \mathbf{c}_M$ are uncoded (portion of the message vector $\mathbf{u}$ appears as they are).

The decoding for TCM is usually performed with the Viterbi algorithm [24] on the trellis diagram associated with the underlying convolutional encoder.

- for multilevel codes, as in Fig. 11 of Sec. 4, $\mathbf{c}_1, \dots, \mathbf{c}_M$ are the codewords of independent M binary (block or convolutional) codes.

The decoding for multilevel codes is usually performed in a staged fashion. Decoding associated with the M binary codes are performed successively based on the assumption that the decisions at the former stages are correct as in Fig. 12.

- for BICM, as in Fig. 19 of Sec. 5, a single binary linear code is used to encode $\mathbf{u}$. The codeword of length Mn is bit-wise interleaved, and is decomposed into a set of M binary vectors $\mathbf{c}_1, \dots, \mathbf{c}_M$ by a serial-to-parallel transform.

The decoding for BICM is based on the (soft-decision) decoder of the binary code employed. The metric is computed for each codeword bit rather than for the symbol over $\mathcal{X}$.

Suppose that each of the 2^M signal points in $\mathcal{X}$ are given a label of length M , and the mapping ϕ brings an M -bit string to the corresponding signal point on $\mathcal{X}$. The following two mappings are frequently considered among others,

- natural mapping: the 2^M signal points are labeled by an M -tuple as a binary representation of the increasing integers from 0 to $M-1$ (which are given counterclockwise in the left of Fig. 3).
- Gray mapping: the signal points are labeled so that the neighboring points differ at most in one of the labeling bits.

An example of natural and Gray mapping for 8-PAM and 8-PSK constellations are shown in Table 2.3 and Fig. 3, respectively. For natural mapping, the least and the most significant bits in each label are arranged from right to left.

It is easily observed that, for uncoded transmission, Gray mapping minimizes the bit error probability as decision errors most frequently occurs with respect to the neighboring signal point(s). See [30][31] for the characterization of mapping for non-binary alphabets.

As we will see in the upcoming sections, the selection of mapping ϕ has major effects on the design of the encoder f as well as the decoder g . For example, the use of natural mapping and the associated partitioning of the signal set plays an important role in TCM. Proper selection of component codes strongly depends on the mapping ϕ in multilevel codes and multistage decoding. Gray mapping should be used in BICM in terms of the error probability, but use of natural mapping may be more preferable when iterative decoding is performed as in Sec. 5.4.

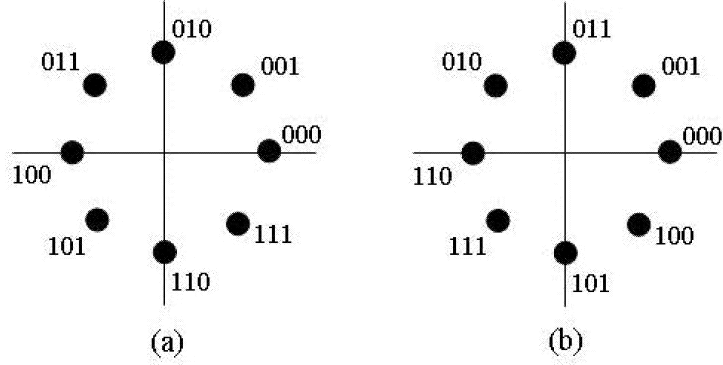


Figure 3: (a) natural mapping and (b) Gray mapping for an 8-PSK constellation

2.4 Performance Measure

Suppose that a codeword $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$ of $\mathcal{C}$ is transmitted and $\mathbf{y} = (y_1, \dots, y_n)$ is received. The maximum likelihood decoding rule, which achieves the smallest codeword-wise error probability (under the assumption that every codeword is transmitted with the same probability), is to determine a codeword $\mathbf{x}' = (x'_1, \dots, x'_n)$ that maximizes the likelihood function $p(\mathbf{y}|\mathbf{x}')$ as the estimate of transmitted codeword $\mathbf{x}$. Formally, over memoryless channels, the decision rule is given by

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}' \in \mathcal{C}} p(\mathbf{y}|\mathbf{x}') = \arg \max_{\mathbf{x}' \in \mathcal{C}} \prod_{j=1}^n p(y_j|x'_j). \quad (3)$$

For the AWGN channel and binary channel input alphabet, it is equivalent to determine

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}' \in \mathcal{C}} \prod_{j=1}^n x'_j m(x'_j), \quad (4)$$

where

$$m(x'_j) = \ln \frac{p(y_j|x'_j = +1)}{p(y_j|x'_j = -1)}. \quad (5)$$

This decoder for binary signaling is directly used for decoding some of the coded modulation schemes.

We define a pairwise error probability $P(\mathbf{x} \rightarrow \mathbf{x}')$ that the decoder incorrectly determines a code sequence $\mathbf{x}'$ under the condition that another codeword $\mathbf{x}$ ($\neq \mathbf{x}'$) is actually transmitted. When maximum likelihood decoding is performed, this event occurs when $p(\mathbf{y}|\mathbf{x}) < p(\mathbf{y}|\mathbf{x}')$ or equivalently $d_E^2(\mathbf{x}, \mathbf{y}) > d_E^2(\mathbf{x}', \mathbf{y})$ over the AWGN channel, where the squared Euclidean distance between $\mathbf{x}$ and $\mathbf{x}'$ is defined as $d_E^2(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^n \|x_i - x'_i\|^2$. The pairwise error probability is computed as

$$P(\mathbf{x} \rightarrow \mathbf{x}') = Q \left(\sqrt{\frac{\text{SNR}}{2}} d_E^2(\mathbf{x}, \mathbf{x}') \right), \quad (6)$$

where $Q(x) = 1/\sqrt{2\pi} \int_x^\infty e^{-z^2/2} dz$. By the union bound, the error probability p_e , which is the probability that the transmitted codeword is not correctly estimated by the decoder, is upper bounded by

$$p_e \leq \sum_{\mathbf{x} \in \mathcal{C}} P(\mathbf{x}) \sum_{\mathbf{x}' \in \mathcal{C}: \mathbf{x}' \neq \mathbf{x}} P(\mathbf{x} \rightarrow \mathbf{x}'), \quad (7)$$

where $P(\mathbf{x})$ is usually assumed to be $1/2^k$. This bound is usually tight at high SNR. At high SNR, the upper bound on the error probability in Eq. (7) is well approximated only by the terms for pairs of codewords, $\mathbf{x}$ and $\mathbf{x}'$, whose distance $d_E^2(\mathbf{x}, \mathbf{x}')$ gives the minimum distance of the code $\mathcal{C}$,

$$d_{\min, \mathcal{C}} = \arg \min_{\mathbf{x}, \mathbf{x}' \in \mathcal{C}: \mathbf{x} \neq \mathbf{x}'} d_E^2(\mathbf{x}, \mathbf{x}'). \quad (8)$$

Consequently, selecting coded modulation schemes with large $d_{\min, \mathcal{C}}$ for a given pair of n and k is one of the important design criterion for the AWGN channel. It is indeed optimal in the sense of the asymptotic coding gain with respect to the uncoded transmission over an alphabet $\mathcal{X}_u$ and is defined as

$$G_{\mathcal{C}} = 10 \log_{10} \frac{d_{\min, \mathcal{C}}}{d_{\min}(\mathcal{X}_u)}, \quad (9)$$

where $d_{\min}(\mathcal{X}_u) = \min_{\mathbf{x}, \mathbf{x}' \in \mathcal{X}_u: \mathbf{x} \neq \mathbf{x}'} \|\mathbf{x} - \mathbf{x}'\|^2$ is the minimum squared Euclidean distance between two signal points in $\mathcal{X}_u$. Note that the coding gain is defined as the difference in SNR required to achieve a given error probability between coded and uncoded systems, and the asymptotic coding gain indicates the one at asymptotically high SNR.

However, it should be noted that a code with the largest asymptotic coding gain does not necessarily outperform the others at low to moderate SNR. Indeed, the number of nearest neighbor codewords at distance $d_{\min, \mathcal{C}}$ has more effect at lower SNR. Also, the pairwise error probability $P(\mathbf{x} \rightarrow \mathbf{x}')$ with $d_E^2(\mathbf{x}, \mathbf{x}') > d_{\min, \mathcal{C}}$ may contribute to the overall error probability more significantly.

On the other hand, the design criterion for the Rayleigh fading channels should be remarkably different. In [29], upper bound on the error probability over the fully-interleaved Rayleigh fading channel is derived as

$$P(\mathbf{x} \rightarrow \mathbf{x}') \leq (\text{SNR})^{-d_H(\mathbf{x}, \mathbf{x}')} \frac{1}{\prod_{i=1, \dots, n: x_i \neq x'_i} \|x_i - x'_i\|^2}, \quad (10)$$

where $d_H(\mathbf{x}, \mathbf{x}')$ denotes the Hamming distance between $\mathbf{x}$ and $\mathbf{x}'$ over $\mathcal{X}$, or the number of positions in which the symbols of $\mathbf{x}$ and $\mathbf{x}'$ differ. The bound in Eq. (10) implies that the symbol-wise Hamming distance $d_H(\mathbf{x}, \mathbf{x}')$ is the most important parameter that dominates the pairwise error probability as it determines the power to the SNR. Also, the product of the squared Euclidean norm between the signal points in $\mathbf{x}$ and $\mathbf{x}'$ (restricted to the positions in which symbols differ) is the secondly important parameter that affects the coding gain. Consideration on fully interleaved channel with perfect channel state information, though not very realistic, gives insights on the code design over more practical channels. For instance, an analysis of coding for block fading channels in [?] reveals that the error rate behaves as a negative power of the SNR, having the block-wise Hamming distance between two signal sequences as the exponent.

3 Trellis Coded Modulation

Trellis coded modulation [3], which are also called trellis codes, specify a set of signal sequences on the expanded signal set by a binary convolutional code so that it can accommodate the redundancy due to the encoding, but without sacrificing bandwidth efficiency. Note that at the time these codes were invented, binary convolutional codes were the most useful class of codes in practice due to the availability of efficient maximum likelihood decoding by Viterbi algorithm.

3.1 Set Partitioning

The problem of joint design of the encoder f and mapping ϕ is the central issue in coded modulation as outlined in Sec. 2.3. An approach called set partitioning facilitates this problem. The key idea is to partition the signal set into subsets so that the squared Euclidean distance between signal points in each subset is increasing.

For illustration, we take the 8-PSK constellation shown in Fig. 4 with labeling due to the natural mapping. At the first level, the eight signal points in two-dimensional space are partitioned into two subsets which are defined by the rightmost bit of the labels. Each of the two subsets is further partitioned into two subsets by the middle labeling bit at the second partitioning level. The same procedure is applied to each subset repeatedly until only one signal point remains.

Let $\mathcal{X}_{x^M \dots x^1}$ denote the set of signal points that have a label $x^M \dots x^1$, while x^i is supposed to take arbitrary values when denoted as $x^i = *$. Under this notation, the 8-PSK signal set $\mathcal{X}$ is alternatively expressed as $\mathcal{X}_{***}$, which is partitioned into two subsets $\mathcal{X}_{**0}$ and $\mathcal{X}_{**1}$. The former subset is further partitioned into $\mathcal{X}_{*00}$ and $\mathcal{X}_{*10}$, and the latter yields $\mathcal{X}_{*01}$ and $\mathcal{X}_{*11}$. These four sets, each consisting of two points, are arranged from left to right at the bottom of Fig. 4 in which the third partitioning level is not explicitly depicted. Define the intra-set distance at level- i as

$$\delta_i^2 = \min_{x^1, \dots, x^{i-1} \in \{0,1\}} d_{\min}(\mathcal{X}_{* \dots * x^{i-1} \dots x^1}). \quad (11)$$

The intra-set distance at each level for the partitioning in Fig. 4 is $\delta_1^2 = d_{\min}(8\text{PSK}) = 0.586$, $\delta_2^2 = 2.0$, and $\delta_3^2 = 4.0$ as indicated by the arrows. We see that the intra-set distance due to the natural mapping is larger at higher levels.

3.2 Encoder and Decoder for Trellis Codes

The distance structure due to set partitioning introduced above can be effectively used in the design of trellis coded modulation. Suppose that the target asymptotic coding gain is 3.0dB over the uncoded QPSK with $d_{\min}(\text{QPSK}) = 2.0$. Since $\delta_3^2 = 4.0$ at the third level, the vector $\mathbf{c}_3$ associated with this level need not be encoded because of $10 \log_{10} \delta_3^2 / d_{\min}(\text{QPSK}) = 10 \log_{10} 4.0 / 2.0 \approx 3(\text{dB})$. Accordingly, we only need to specify the sequence of subsets $\mathcal{X}_{*00}$, $\mathcal{X}_{*01}$, $\mathcal{X}_{*10}$, and $\mathcal{X}_{*11}$ at the lower two levels by the output of a convolutional code so that $d_{\min,C}$ is at least 4.0. One simple example of encoder for trellis coded modulation is shown in Fig. 5. The output of the rate-1/2 convolutional encoder determines the signal subsets through the rightmost and middle labeling bit of each signal point. The uncoded bit x^3 corresponding to the leftmost bit in the label determines the signal points in the subset $\mathcal{X}_{*x^2x^1}$ for x^1 and x^2 of the partition chain in Fig. 4. The trellis diagram with four states for this trellis code is illustrated in Fig. 6. The state associated with the value a and b for the left and right delay element, respectively, in the convolutional encoder in Fig. 5 is denoted by S_{ab} . The trellis structure is defined by the rate 1/2 binary convolutional code in the encoder in Fig. 5. Each branch is labeled by one of the signal points on $\mathcal{X}$, expressed in octal form as $4x^2 + 2x^1 + x^0$. Also, there are parallel branches per transition between states, and each branch is associated with one of the two signal points in $\mathcal{X}_{*x^2x^1}$.

The minimum distance for distinct two paths that do not share parallel branches is 4.586: consider two paths in the trellis with the signal points 0 (000), 0 (000), 0 (000) and 2 (010), 1 (001), 2 (010) that diverge at the initial state and merge again at the state S_{00} . The squared Euclidean distance between these signal sequences are computed as $2.0 + 0.586 + 2.0 = 4.586$.

Consequently, the minimum squared Euclidean distance is given as δ_3^2 by $d_{\min,C} = \min\{4.586, 4.0\} = 4.0$ or equivalently the asymptotic coding gain is $G_C = 3\text{dB}$.

The encoder of this trellis code allows two input bits per transmission, one is to be encoded and the other is left uncoded. The overall rate of this trellis code is accordingly

$R = 2$ (bis/channel use) which is exactly the same as the uncoded QPSK. As a result, the asymptotic coding gain of 3dB is obtained by introducing expanded signal set (8-PSK, rather than uncoded QPSK) without sacrificing the bandwidth efficiency. In essence, this is possible through careful choice of the convolutional codes even though the minimum squared Euclidean distance of the 8-PSK constellation is smaller than the uncoded QPSK. Note that if the target asymptotic coding gain is larger than 3.0 dB for transmission with the 8-PSK signal set, all the input bits from $\mathbf{c}_1, \dots, \mathbf{c}_3$ to the mapper ϕ should be encoded as otherwise $d_{\min, \mathcal{C}}$ is upper bounded by $\delta_3^2 = 4.0$.

Trellis codes for 8-PSK can be also realized by recursive systematic convolutional (RSC) encoders as in Fig. 7 besides the feedback-free encoder of Fig. 5. This general encoder has ν delay elements and 2^ν states in the associated trellis. Note that the trellis code due to the encoder in Fig. 5 corresponds to the one in Fig. 7 by letting $\nu = 2$, $h_0^{(0)} = h_2^{(0)} = 1$ and $h_1^{(1)} = 1$ and $h_i^{(j)} = 0$ for the other i and j .

As illustrated in Sec. 2.4, designing trellis codes of a given rate with large $d_{\min, \mathcal{C}}$ is desirable for the AWGN channel. Trellis codes with large $d_{\min, \mathcal{C}}$ are usually designed by computer search as is usually done for constructing binary convolutional codes. Some good codes found by Ungeboeck [3] for the 8-PSK constellation are shown in Table 2. It is seen that larger asymptotic coding gains $G_{\mathcal{C}}$ are obtained as the number of delay elements in the encoder increases.

TCM is usually decoded by using the Viterbi algorithm on the trellis diagram for the binary convolutional code employed. Over the AWGN channel, the squared Euclidean distance between the code sequence and the received sequence can be used as path metric. The complexity of this maximum likelihood decoder grows exponentially to the number of delay elements ν .

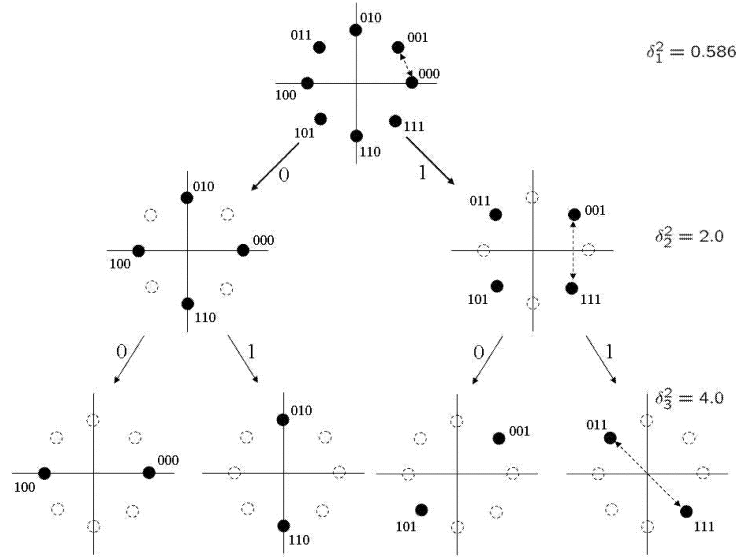


Figure 4: Set Partitioning for 8-PSK constellation

A general encoding structure of trellis coded modulation is depicted in Fig. 8. There are K input and N output bits for the convolutional encoder while $M - N$ bits of the output are uncoded. A rate $(N - 1)/N$ convolutional code is most frequently employed, and it doubles the size of the constellation compared with the uncoded transmission of the same rate. However, employing convolutional codes of lower rate is possible. The value N should

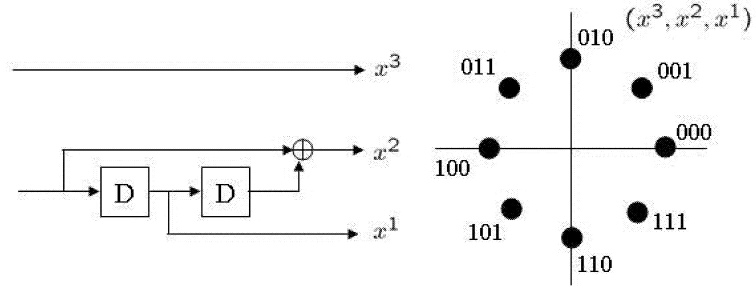


Figure 5: An encoder for trellis coded modulation

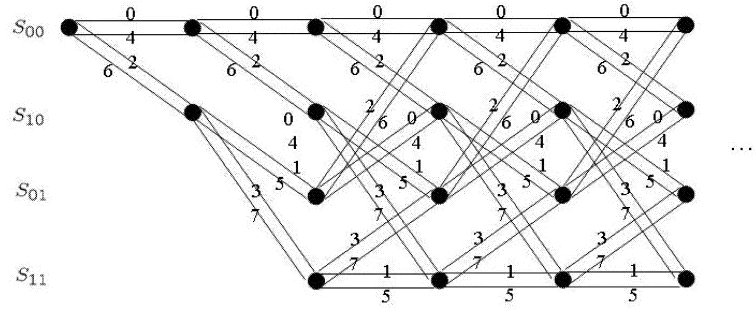


Figure 6: Trellis diagram for the encoder in Fig. 5

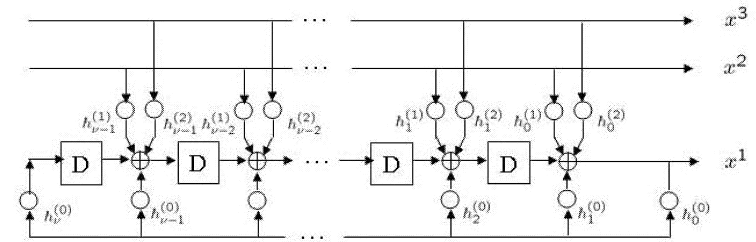


Figure 7: Recursive Systematic Encoder for 8-PSK Trellis Codes

Table 2: Trellis Coded Modulation for 8-PSK Constellation

ν	$H^{(0)}(D)$	$H^{(1)}(D)$	$H^{(2)}(D)$	$G_c(\text{dB})$
2	5	2	–	3.0
3	11	02	04	3.6
4	23	04	16	4.1
5	45	16	34	4.6
6	105	036	074	4.8
7	203	014	016	5.0

be determined based on the target coding gain, or by the intra-set distance δ_{N+1}^2 of the lowest uncoded level when the intra-set distance δ_i^2 is non-decreasing. Note that there are 2^{M-N} parallel branches between a pair of states when uncoded sequence is considered for $M - N$ higher levels.

If we employ a rate $(N - 1)/N$ binary convolutional code, a state is connected to 2^{N-1} states by 2^{M-N} parallel branches. In this case, each branch is associated with one of the signal points on $\mathcal{X}$, the squared Euclidean distance between the signal points and the received signal y_j is used as the branch metric in the trellis for the Viterbi algorithm.

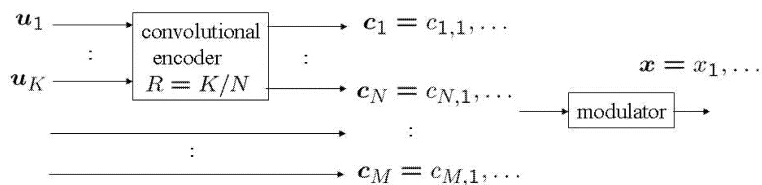


Figure 8: Encoder of trellis coded modulation

3.2.1 Design of Trellis Coded Modulation

Design and construction of trellis coded modulation is briefly reviewed in the following.

In designing TCM, the asymptotic coding gain, or equivalently $d_{\min, c}$, is often employed as the design criterion for transmission over the AWGN channel. This is in part due to the fact that analysis of error probability is not easy for general trellis codes in that the pairwise error probability $P(\mathbf{x} \rightarrow \mathbf{x}')$ needs to be evaluated for all pairs of code sequences $\mathbf{x}$ and $\mathbf{x}'$. On the other hand, restricting the class of trellis codes to those with certain symmetric structure may facilitate the analysis on error probability or designing good codes. For example, generating function approach commonly used for computation of the distance spectrum of convolutional codes is extended to trellis codes in [33]. A class of codes with geometrical uniformity is defined and formulated in [34], with the property that the Voronoi region of each codeword is congruent. An overview of performance analysis for trellis codes is given in [35]. For recent development in searching for good TCM encoders, see [36] and the references therein.

A good trellis code for the AWGN channel may not perform well over the Rayleigh fading channel. Recall that the most important design criterion for the fully interleaved channel is the symbol-wise Hamming distance. However, codes with parallel branches in the trellis diagram perform poorly because the symbol-wise Hamming distance is $d_H(\mathbf{x}, \mathbf{x}') = 1$. Based on this observation, analysis and design of trellis codes for fading channels were studied under various settings, see for example [18][37][38][39]. Note that BICM appears to be recognized

as a more useful coding scheme for fading channels than TCM-type codes [5].

Trellis codes based on multi-dimensional (more than two-dim.) constellations were proposed in [41][42]. For illustration, consider the simplest four-dimensional (4-D) signal set as the Cartesian product $\mathcal{X} \times \mathcal{X}$ of the 8-PSK constellation $\mathcal{X}$ [42]. There are $8^2 = 64$ signal points with intra-set distance $\delta_1^2 = 0.586$ in the 4-D space, and this is partitioned in six levels. At the first level, $\mathcal{X} \times \mathcal{X}$ is partitioned into $\mathcal{X}_{**0} \times \mathcal{X}_{**0} \cup \mathcal{X}_{**1} \times \mathcal{X}_{**1}$ and $\mathcal{X}_{**0} \times \mathcal{X}_{**1} \cup \mathcal{X}_{**1} \times \mathcal{X}_{**0}$, which results in $\delta_2^2 = 0.586 * 2 = 1.172$. At the second level, the former set is partitioned into $\mathcal{X}_{**0} \times \mathcal{X}_{**0}$ and $\mathcal{X}_{**1} \times \mathcal{X}_{**1}$ with $\delta_3^2 = 2.0$. Repeating the same procedure for the second labeling bit, $\delta_4^2 = 4.0$ and $\delta_5^2 = 4.0$, and for the leftmost bit, and finally $\delta_6^2 = 8.0$ results. As in Fig. 8, a binary convolutional code is applied to this partition chain. Advantages in employing multi-dimensional signal set includes the flexibility in the design of the overall rate R of the trellis code, and the availability of signal sets with lower average transmit power. The desirable features of multi-dimensional QAM-type constellations are reviewed and developed in [44]. It is also shown that rotationally-invariant trellis codes, which resolves ambiguity in carrier phase in conjunction with differential encoding, are more easily available with multi-dimensional constellations [42][43] than with two-dimensional constellations [40] for which the use of non-linear encoders are necessary.

Trellis codes based on lattices are proposed in [45], and further extended and characterized in [46] as coset codes. As in Fig 9(a), a lattice Λ and its partition Λ/Λ' induced by the sublattice Λ' is considered. Part of the message bits are encoded by a linear code and identifies one of the coset of Λ' in Λ . While lattices define infinite signal sets and cannot be used on (average) power limited channels, it decouples the functionality of coding and shaping (determining the constellation boundary) and make the fundamental properties of coding clearer. For example, let Λ be an n -dimensional integer lattice $\mathbb{Z}^n$ and its sublattice $\Lambda' = 2^N \mathbb{Z}^n$. If the output of a rate $(N-1)/N$ binary convolutional encoder selects the coset of Λ' , it can be regarded as a template for Ungerboeck-type TCM based on PAM constellations.

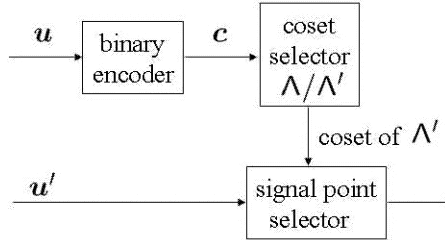


Figure 9: Encoder of coset codes

3.3 Concatenated Trellis Coded Modulation

Turbo codes [7][8] were first presented in 1993 and shown to achieve near-Shannon limit performance over the binary-input AWGN, and later the effectiveness of the codes are also verified. The encoder structure is the parallel concatenation of RSC encoders: more precisely, a message sequence u is encoded by the first RSC encoder, while the interleaved message enters the second RSC encoder. The output of those two encoders constitutes codeword symbols of the overall turbo code. Provided that rate 1/2 convolutional encoders are employed, the overall rate of the turbo code is 1/3 because the systematic part needs not be included in the codeword twice, and higher rate codes are available by puncturing some portion of the parity bits. For decoding turbo codes, soft-input/soft-output (SISO) decoders [47] associated with the two constituent encoders are used so that they iteratively exchange the so-called

extrinsic information [47]. See another chapter of this book for the details of turbo codes and turbo decoding.

We can make use of this parallel concatenation structure and iterative decoding to coded modulation by noticing that trellis codes have been realized by RSC encoders. The encoder of turbo trellis coded modulation proposed in [48] for the 8-PSK constellations is presented in Fig. 10. The parity bits are taken alternately from the first and the second encoders, and the other parity bits are punctured. To allow this structure, two message bits are grouped and pairwise interleaved. The rate of this coding scheme is thus $R = 2$ [bits/channel use]. Decoding for each of the trellis codes is performed by the BCJR algorithm [49], and the extrinsic information for the systematic part are exchanged between the two decoders as for binary turbo codes. Note that since both systematic bits and parity bits are accommodated in a single 8-PSK symbol in contrast to the BPSK transmission, symbol-wise extrinsic information for the systematic part should be extracted. It was shown in [48] that the code in Fig. 10 successfully achieves BER of 10^{-5} within 1 dB from the Shannon limit for the block length $n = 5000$.

Parallel concatenated coded modulation scheme using bit-interleaving was presented in [50] among others.

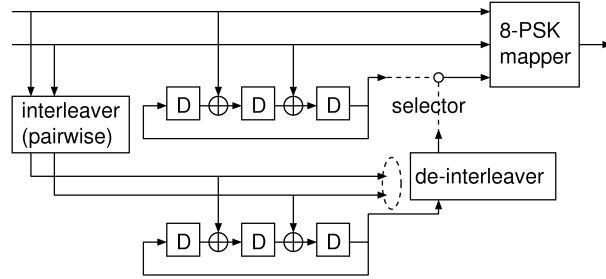


Figure 10: Encoder of parallel concatenation of trellis codes

An alternative coding scheme for iterative decoding is the serial concatenation of two convolutional encoders with an interleaver in between [51]. In serially concatenated trellis coded modulation in [52], messages are encoded by an outer convolutional encoder followed by interleaving, and the output is encoded by the inner convolutional encoder which should be of recursive form. In particular, the output of the outer encoder in [52] is mapped to a four-dimensional signal set given as the Cartesian product of two-dimensional one. Use of rate-1 inner encoder for serially concatenated TCM is proposed and analyzed in [53][54]. As emphasized in these publications, one advantage of serially concatenated TCM over the parallel counterpart is that they exhibit an error floor only at lower error probabilities.

4 Multilevel Codes and Multistage Decoding

At almost the same time as the first presentation of TCM was made, Imai and Hirakawa [4] published another coding scheme using M independent binary codes for 2^M -ary signal sets. This approach is now called multilevel coding scheme which we illustrate by the 8-PSK constellation in the following.

4.1 Multilevel Codes

The encoder for multilevel codes works as follows: a message vector $\mathbf{u} = (u_1, \dots, u_k)$ is decomposed into M vectors $\mathbf{u}_1, \dots, \mathbf{u}_M$ in which $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,k_i})$ is a k_i -dimensional

Table 3: An example of multilevel code of length 8.

message	codewords
(0)	$\mathbf{c}_1 = (0, 0, 0, 0, 0, 0, 0, 0)$
(1, 0, 1, 1, 0, 0, 0, 0)	$\mathbf{c}_2 = (1, 0, 1, 1, 0, 0, 0, 1)$
(0, 1, 1, 1, 0, 0, 1, 0)	$\mathbf{c}_3 = (0, 1, 1, 1, 0, 0, 1, 0)$
	$\mathbf{x} = (2, 4, 6, 6, 0, 0, 4, 2)$

vector and $k = \sum_{i=1}^M k_i$. Let $\mathcal{C}_i$ be an (n, k_i, d_i) binary linear block code, where n , k_i , and d_i are the code length, dimension, and the minimum Hamming distance of $\mathcal{C}_i$, respectively. We call each code $\mathcal{C}_i$ as the component code for level i . The vector $\mathbf{u}_i$ is encoded into a codeword $\mathbf{c}_i = (c_{i,1}, \dots, c_{i,n}) \in \mathcal{C}_i$. A vector $(c_{1,j}, c_{2,j}, \dots, c_{M,j})$ consisting of j -th elements from $\mathbf{c}_1, \dots, \mathbf{c}_M$ is mapped to the corresponding signal point $x_j \in \mathcal{X}$ by ϕ to yield the signal sequence $\mathbf{x} = (x_1, \dots, x_n)$. Note that convolutional codes are also available as component codes.

Since there are $k = k_1 + \dots + k_M$ message bits, the rate of the overall multilevel code is given as $R = k/n = \sum_{i=1}^M k_i/n$. The minimum distance of the overall multilevel code is derived as

$$d_{\min, \mathcal{C}} = \min_{i \in \{1, \dots, M\}} d_i \delta_i^2. \quad (12)$$

As an example, consider a multilevel code of block length $n = 8$: the first level component code is the repetition code with $k_1 = 8, d_1 = 8$, the second level code is the single parity-check with $k_2 = 7, d_2 = 2$, and the third level is uncoded or conveniently regarded as a trivial code with $k_3 = 8, d_3 = 1$. For three codewords $\mathbf{c}_1 = (0, 0, 0, 0, 0, 0, 0, 0)$, $\mathbf{c}_2 = (1, 0, 1, 1, 0, 0, 0, 1)$, and $\mathbf{c}_3 = (0, 1, 1, 1, 0, 0, 1, 0)$, the transmitted signal sequence is determined as $x_j = c_{1,j} + 2c_{2,j} + 4c_{3,j}$ for the integer representation of signal points in the natural mapping. Consequently, the eight symbols to be transmitted are those with labels x (x^3, x^2, x^1) = 2 (010), 4 (100), 6 (110), 6 (110), 0 (000), 0 (000), 4 (100), and 2 (010) in octal (binary) form. The number of information symbols is $k = k_1 + k_2 + k_3 = 16$ and the rate of the overall rate is $R = k/n = 2$ (bits/channel use). The minimum distance of this multilevel code is computed as $d_{\min, \mathcal{C}} = \min\{0.586 \times 8, 2.0 \times 2, 4.0 \times 1\} = 4.0$.

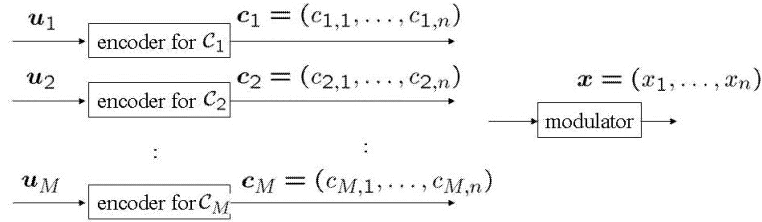


Figure 11: Encoder of multilevel codes

4.2 Multistage Decoding

Maximum likelihood decoding of the overall multilevel code is in general computationally demanding as independent encoding is performed at each level. One of the outstanding contributions of [4] is the introduction of multistage decoding, in which a (soft-decision) decoder for each binary component code successively works. Even though this decoding is not optimal in general with respect to the error rate, the decoding complexity is just on

the order of maximum computational cost of the component decoders, and is significantly reduced compared to the optimum decoding.

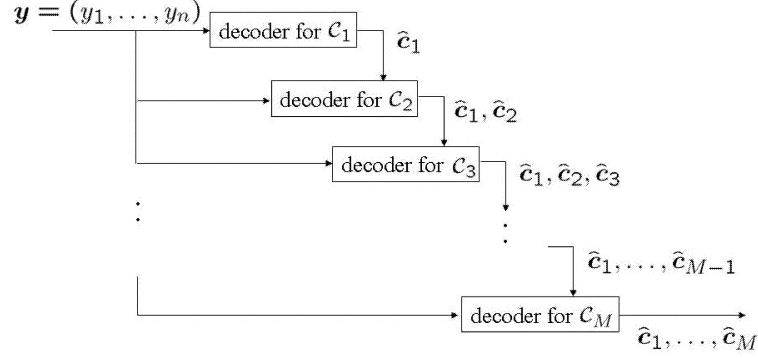


Figure 12: Multistage Decoder

Multistage decoding for multilevel codes is closely related to the notion of set partitioning and works as follows: i -th stage decoder for the binary component code $\mathcal{C}_i$ works under the assumption that the decisions made at lower stages $\hat{\mathbf{c}}_1, \dots, \hat{\mathbf{c}}_{i-1}$ are correct and binary vectors of the higher levels, $\mathbf{c}_{i+1}, \dots, \mathbf{c}_M$ are uncoded. This procedure is sketched in Fig. 12 and described in more detail for the 8-PSK constellation in the following.

This effective channel encountered by each stage decoder is shown in Fig. 13 with lower to higher stages from the left to the right. They are binary channels in the sense that the decision should be made between 0 and 1 for the symbols in the codeword $\mathbf{c}_i$. The multistage decoding proceeds as follows:

1. At the first stage, the decoder for the binary code $\mathcal{C}_1$ at the first level computes the estimate of the codeword $\mathbf{c}_1 = (c_{1,1}, \dots, c_{1,n})$ under the assumption that the binary vectors at higher levels $\mathbf{c}_2, \dots, \mathbf{c}_M$ are uncoded. In the leftmost constellation in Fig. 13, the log-likelihood ratio for the codeword symbol $c_{1,j}$ of the first level component code is given as

$$m_1(y_j) = \ln \frac{\sum_{x \in \mathcal{X}_{**0}} f(y_j|x)}{\sum_{x \in \mathcal{X}_{**1}} f(y_j|x)} \approx \ln \frac{\max_{x \in \mathcal{X}_{**0}} f(y_j|x)}{\max_{x \in \mathcal{X}_{**1}} f(y_j|x)}. \quad (13)$$

The approximation in Eq. (13) essentially considers only the signal point which are the closest to the received signal $\mathbf{y}_j$ in $\mathcal{X}_{**0}$ and $\mathcal{X}_{**1}$ if the transmission is over the Gaussian channel. These values are used as the input to the decoder for $\mathcal{C}$ as in Eq. (5). We denote the decoder output by $\hat{\mathbf{c}}_1$.

2. The second stage decoder for $\mathcal{C}_2$ works under the assumption that the output of the previous stage $\hat{\mathbf{c}}_1$ correctly estimated the transmitted codeword $\mathbf{c}_1$. Also, the vector $\mathbf{c}_3$ of the higher level is regarded as uncoded. Assuming $\hat{c}_{1,j} = 0$, the decision is made between $\mathcal{X}_{*00}$ and $\mathcal{X}_{*10}$ as in the middle of Fig. 13. This decision to be made at this stage is again binary, and is for the symbols in the codeword $\mathbf{c}_2$. The metric $m_2(y_j)$ to be used in Eq. (5) follows the same line as Eq. (13) but with respect to $\mathcal{X}_{*00}$ and $\mathcal{X}_{*10}$. The decision $\hat{\mathbf{c}}_2$ together with $\hat{\mathbf{c}}_1$ is fed to the third stage decoder.
3. In the same way, based on the assumption of correct estimates $\hat{c}_{2,j}$ and $\hat{c}_{1,j}$, the decision at the third stage is made between the subsets $\mathcal{X}_{0\hat{c}_{2,j}\hat{c}_{1,j}}$ and $\mathcal{X}_{1\hat{c}_{2,j}\hat{c}_{1,j}}$.

Multistage decoding is suboptimal with respect to the error probability even when maximum likelihood decoder for $\mathcal{C}_i$ is employed at every stage because of the following two issues,

- Suppose that a decoding error occurs at the i -th stage. Then, an erroneous decision $\hat{c}_i (\neq c_i)$ is passed to the subsequent stages, while this decision is assumed to be correct. Accordingly, it is likely that the subsequent stage decoders also make decoding error. This behavior is called error propagation and can deteriorate the error performance of the overall multistage decoding.
- The i -th stage decoder regards the vectors $c_{i+1}, \dots, c_M$ of the higher levels as uncoded. It results in an increased number of effective nearest neighbor signal sequences for the codeword c_i . This number is called multiplicity for an erroneous codeword $c' (\neq c_i)$. For example, on the binary channel associated with the first stage of Fig. 13, there are two neighboring signal points with the opposite labeling bit for every signal point: namely, the signal points with xx0 has two neighboring point with xx1. Consider the transmitted binary codeword $c_1 \in \mathcal{C}_1$ and another one $c'_1 \in \mathcal{C}_1$ where the Hamming distance between them is w . Then, there could be effectively 2^w signal sequences associated with c'_1 with squared Euclidean distance $w d_1^2$ from the transmitted signal sequence. This multiplicity affects the decoding error probability especially when the minimum distance of the component code d_1^2 is large.

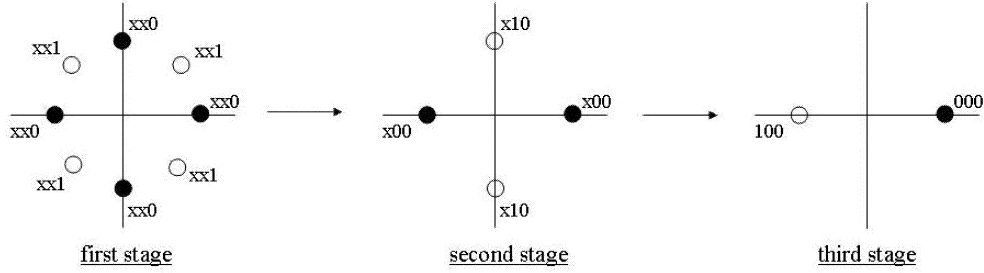


Figure 13: Binary channels encountered at each stage of multistage decoding

4.3 Code Design for Multilevel Codes and Multistage Decoding

4.3.1 Component Codes for Low Error Probability

The minimum squared Euclidean distance $d_{\min, \mathcal{C}}$ of the overall multilevel code $\mathcal{C}$ is the key parameter on the error probability under maximum likelihood decoding as in Sec. 2.4. Since $d_{\min, \mathcal{C}}$ is determined as Eq. (12), one may come up with a code design such that $d_i \delta_i^2$ at each level is (almost) balanced. However, we should be careful about the choice of the component codes when multistage decoding is assumed. Recall that for an error event of Hamming weight w for the decoder, the multiplicity amounts to 2^w at the first and second stage decoding for 8-PSK with natural mapping. Consequently, over the AWGN channel, upper bound on the error probability at the stage- i ($i = 1, 2$) is given from the union bound as

$$p_{e,i} \leq \sum_{w=d_i}^n A_{i,w} 2^w Q \left(\sqrt{\frac{\text{SNR}}{2}} w \delta_i^2 \right), \quad (14)$$

where $A_{i,w}$ is the number of codewords with weight w in the component code $\mathcal{C}_i$ of level- i (note that error propagation from the first to the second stage is ignored in Eq. (14)). At

high SNR, where the bound is tight, the term corresponding to $w = d_i$ dominates the error probability, and the effect of multiplicity results in the increase of error rate by 2^{d_i} times. Degradation of error performance is also observed at low to moderate SNR as well especially for large d_i .

Based on these observations, component codes should be selected so as to satisfy

$$d_1\delta_1^2 > d_2\delta_2^2 > \dots > d_M\delta_M^2 \quad (15)$$

holds in order to avoid severe error propagation.

As an example, simulation results on bit error rate (BER) are shown in Fig. 14 for a multilevel code and multistage decoding of block length $n = 32$ and uncoded QPSK. The component codes are the extended BCH (eBCH) code with $k_1 = 6, d_1 = 16$ for the first level, eBCH code with $k_2 = 26, d_2 = 4$ for the second level, and single parity-check code with $k_3 = 31, d_3 = 2$ for the third level, and where practically optimum decoding is performed at each stage. The rate of the overall multilevel code is $R = (6 + 26 + 31)/32 \approx 1.97$, which is almost the same as the uncoded QPSK. The upper bound on BER due to Eq. (14) (more precisely, w/n is multiplied to each term in the summation to incorporate the bit error rate) for the first stage is also depicted. In this case, we observe a coding gain of about 3 dB at BER of 10^{-5} . Also, it appears that the error performance of the second and third stage decoding follows that of the first stage. This observation implies that the error rate of the overall multistage decoding is dominated by that of the first stage decoder due to error propagation although the squared Euclidean distance at the first level is larger than others, or $d_1\delta_1^2 \approx 9.376 > d_2\delta_2^2 = d_3\delta_3^2 = 8.0$.

Theoretical analysis on the error probability is useful in designing multilevel codes with high coding gain. Performance analysis of multistage decoding is given in [57] for various decoders at each stage (soft-/hard-decision decoding, minimum-distance/bounded-distance decoding). Analysis under the effect of error propagation is given in [58] for multilevel codes in which each component codeword is interleaved before being mapped by ϕ . Tight bounds on the error probability is discussed in [59] by Chernoff bound arguments, and in [60][61] by (tangential) sphere bound. Some good selections of short and simple component codes are discussed in [62] with a relatively high target error rate.

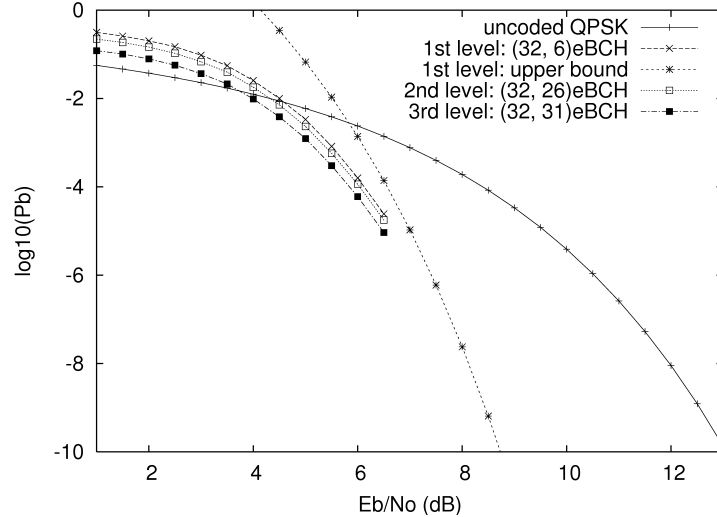


Figure 14: Error performance of multilevel codes with multistage decoding with the extended BCH codes of length 32

4.3.2 Design for Capacity-Approaching Component Codes

Use of capacity-approaching component codes in multilevel coding was studied in [63][64][58], with an emphasis on turbo codes soon after its invention in 1993 [7]. An important observation made in [63][64] is that information theoretic quantities can be used as a powerful design guideline. In particular, with a proper selection of component codes and the associated decoders, constrained capacity of a signal set can be achieved with multilevel codes and multistage decoding. To illustrate this fact, let random variables representing the symbols from M binary codewords be $C_1, \dots, C_M$, which are mapped to $\mathcal{X}$ by ϕ . From the assumption that ϕ is a one-to-one mapping and the chain rule of mutual information [27], the constrained capacity of $\mathcal{X}$ is expressed as

$$\begin{aligned} I_{\mathcal{X}} &= I(X; Y) \\ &= I(C_1, C_2, \dots, C_M; Y) \\ &= I(C_1; Y) + I(C_2, \dots, C_M; Y|C_1) \\ &= \vdots \\ &= \sum_{i=1}^M I(C_i; Y|C_1, \dots, C_{i-1}), \end{aligned} \tag{16}$$

where $I(\cdot|\cdot)$ is the conditional mutual information. Recall that the i -th stage decoder in multistage decoding is to estimate $\mathbf{c}_i$ from the received signal $\mathbf{y}$ based on the knowledge of the decisions at lower stages $\hat{\mathbf{c}}_1, \dots, \hat{\mathbf{c}}_{i-1}$. A key observation is that $I(C_i; Y|C_1, \dots, C_{i-1})$ represents the information rate for the binary channel in which the i -th stage decoding for C_i is performed. Accordingly, the optimum rate $I_{\mathcal{X}}$ of the overall signal set $\mathcal{X}$ can be attained by multilevel codes and multistage decoding if ideal binary component codes of rate $R_i = I(C_i; Y|C_1, \dots, C_{i-1}) - \epsilon$ for sufficiently small positive number ϵ is selected at each level.

An alternative proof for the optimality of multilevel coding with multistage decoding was given in [55] in the context of infinite signal sets (with unbounded cardinality) to show that optimum (achieving the sphere-bound [56]) multilevel coset codes exist.

In Fig. 15, for 8-PSK with the set partitioning rule, $I(C_1; Y)$, $I(C_2; Y|C_1)$, and $I(C_3; Y|C_1, C_2)$ are plotted for the AWGN channel with SNR represented at the horizontal axis. Since we see that

$$I(C_1; Y) < I(C_2; Y|C_1) < I(C_3; Y|C_1, C_2) \tag{17}$$

holds, the first level code should be of lower rate at a fixed SNR. From Fig. 15, the information rate $I_{\mathcal{X}} = I(X; Y) = 2.0$ is achieved at about SNR ≈ 5.75 (dB) in which $I(C_1; Y) \approx 0.20$, $I(C_2; Y|C_1) \approx 0.81$, and $I(C_3; Y|C_1, C_2) \approx 0.99$. These are the optimum rate allocation for each level.

Turning our attention to the component code selection, turbo codes and LDPC codes achieve performance close to the Shannon limit in their waterfall region at wide range of coding rates for BPSK modulations. It is therefore reasonable that the rate selection rule $R_i \approx I(C_i; Y|C_1, \dots, C_{i-1})$ based on the mutual information of the equivalent binary channel at each stage can be a suitable design criterion for waterfall region of the iteratively decoded coding schemes. In [63][64], it was indeed shown that multilevel codes with turbo component codes according to this rate selection results in approaching the Shannon limit of the signal set $\mathcal{X}$. Later, irregular LDPC codes for multilevel codes and multistage decoding are developed in [65]. Note that turbo codes may exhibit error floor at relatively high error probabilities due to the possible small minimum distance. In such a regime, traditional design criterion based on the distance spectrum (or error probability) should be taken into account as discussed in Sec. 4.2.

Note that the above rate selection rule in Eq. (16) is also available to other partitioning rule.

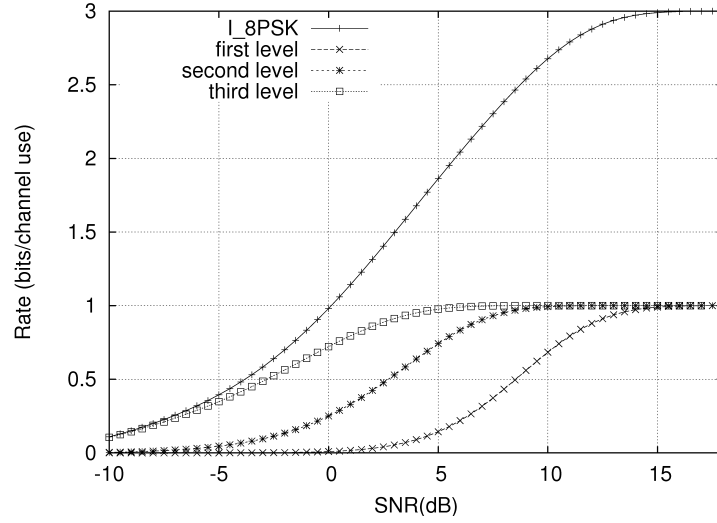


Figure 15: Mutual information of each binary channels in the multistage decoding based on natural mapping for 8-PSK multilevel codes

4.4 Iterative Decoding of Multilevel Codes

As was illustrated in Sec. 4.3.2, multistage decoding does not perform as well as the maximum likelihood decoding for the overall multilevel code unless component codes are appropriately selected, especially when lower level component codes are not “powerful” enough. To overcome this problem, iterative decoding for multilevel codes was proposed in [66] and later studied in [67][68]. The underlying idea is that the decoding performance at stage- i can be improved when side information on the symbols in the higher level codewords $c_{i+1}, \dots, c_M$ are available, rather than regarding them as uncoded. Operationally, SISO decoders for component codes are utilized and extrinsic information [47] is exchanged between them. The underlying notion of iterative multistage decoding is similar to that of BICM-ID to be illustrated in Sec. 5.4, except for the fact that the former approach uses M decoders associated with each component code while the latter one uses only a single SISO decoder.

4.5 Multilevel Codes for Unequal Error Protection

One advantage of multilevel coding and multistage decoding is that unequal error protection (UEP) capability is conveniently provided as was pointed out by [69][70]. This capability is achieved through an appropriate design of signal set, partitioning, and the component codes.

Due to the inherent nature of multistage decoding with possible error propagation, the more important class of message bits should be assigned to lower levels. However, the set partitioning due to the natural mapping may not be very efficient in providing UEP capability because of the large multiplicity as expressed in Eq. (14) or the small capacity of the equivalent channel $I(C_1; Y)$ as illustrated in Fig. 15.

An alternative approach called block partitioning shown in Fig. 16 was proposed and analyzed in [71] and [64]. In this scheme, the labeling is the same as the Gray mapping (but the partition is done from the rightmost labeling bit at the first level). In contrast to the natural mapping, the intra-set distance is not increasing at higher levels and given as $\delta_1^2 = \delta_2^2 = \delta_3^2 = 0.586$. It was shown in [71] that, for an error event of Hamming weight w at the first stage decoding, with probability $2^{-w} \binom{w}{i}$, the squared Euclidean distance to the

decision hyperplane is

$$(i \sin(\pi/8) + (w - i) \cos(\pi/8))^2 / w \quad (18)$$

assuming that the symbols in the component codewords at higher levels are randomly chosen. This is in contrast to the the multiplicity 2^w associated with the squared Euclidean distance of $0.586w = 4 \sin^2(\pi/8)$ which corresponds to $i = w$ in Eq. (18). It implies that applying lower rate codes at results in UEP capability. The advantage of block partitioning at the first stage is also seen by the capacity of equivalent binary channel as shown in Fig. 17, indicating $I(C_1; Y) = I(C_2; Y|C_1) > I(C_3; Y|C_1, C_2)$.

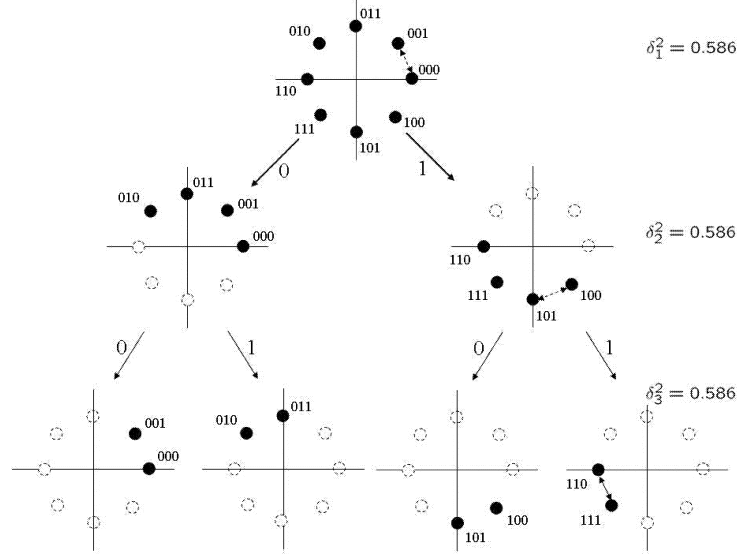


Figure 16: Block Partitioning

Fig. 18 shows the BER of multilevel codes with UEP capabilities together with the upper bound on the error probability derived in [71] based on the distance profile from Eq. (18). The component codes employed are the extended BCH codes of length $n = 64$, and $k_1 = 18$, $k_2 = 45$, and $k_3 = 63$ with $d_1 = 22$, $d_2 = 8$, and $d_3 = 2$, resulting in $d_1 \delta_1^2 \approx 12.9$, $d_2 \delta_2^2 \approx 4.7$, and $d_3 \delta_3^2 \approx 1.17$.

5 Bit-Interleaved Coded Modulation

Bit-Interleaved coded modulation was introduced by Zehavi [6] with a principal goal in improving the performance of trellis codes over the flat Rayleigh fading channels. It is now quite universally considered both in theory and practice under various kind of channels. For a comprehensive treatment and recent developments of BICM, the readers are referred to [19].

5.1 Encoding of BICM

As in Fig. 19, the message vector $\mathbf{u}$ is first encoded into a codeword of the linear code of length $n' = Mn$, dimension k and minimum Hamming distance $d_{\min}$. This codeword $\mathbf{c} = (c_1, \dots, c_{Mn})$ goes through a bit-wise interleaver. Finally, $\mathbf{c}_1, \dots, \mathbf{c}_M$ is obtained by the serial-to-parallel conversion of the interleaved codeword and mapped by ϕ to the constellation

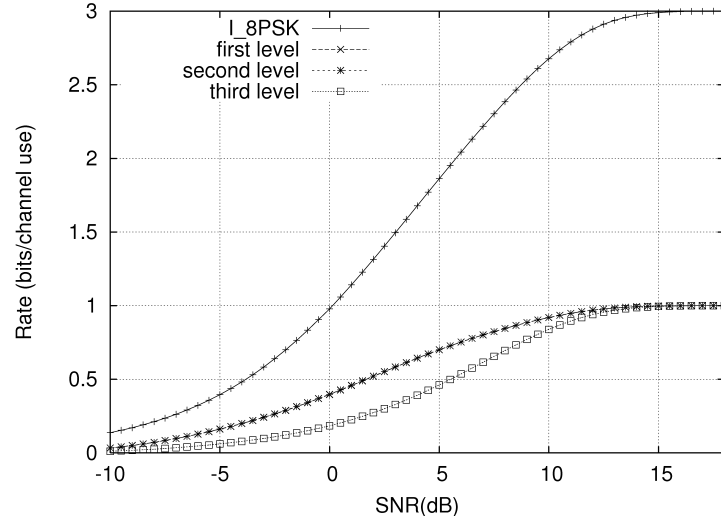


Figure 17: Mutual information of each binary channels in the multistage decoding based on block partitioning for 8-PSK multilevel codes: results for first and second levels overlap.

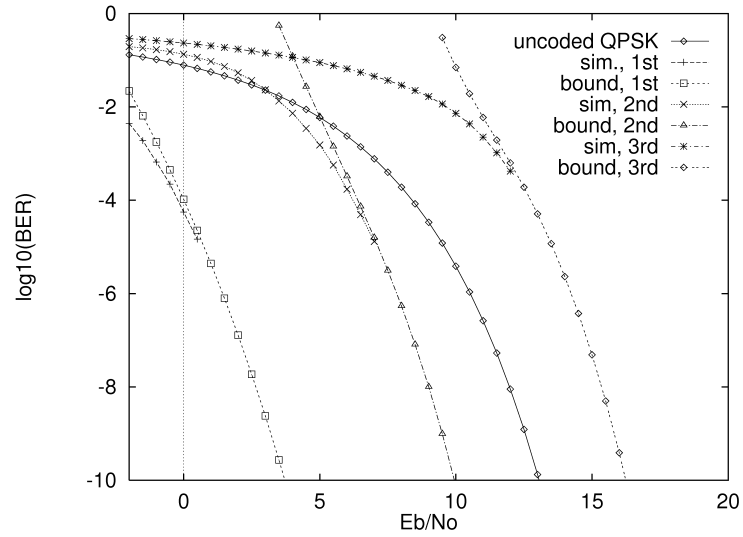


Figure 18: UEP capability with multilevel codes based on component codes of length 64.

$\mathcal{X}$. This structure essentially treats the binary encoder and modulator to be separate two entities.

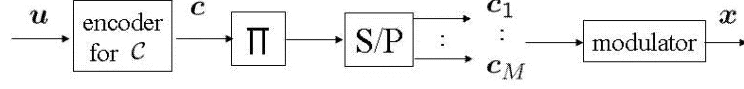


Figure 19: Encoder of BICM

Recall from Sec. 2.4 that the most important parameter for the fading channels is the symbol-wise Hamming distance between two different code sequences $d_H(\mathbf{x}, \mathbf{x}')$ such that $\mathbf{x} \neq \mathbf{x}'$. An intuitive interpretation on the advantage of BICM is that $d_H(\mathbf{x}, \mathbf{x}')$ is expected to be equal or close to the minimum Hamming distance $d_{\min}$ of the underlying code due to the bit-wise interleaving. On the other hand, BICM does not perform as well as the best TCM code of similar complexity over the AWGN channel because of the smaller $d_{\min, \mathcal{C}}$.

5.2 Decoding BICM

Due to the bit-interleaver in the encoding process, the optimum decoding of BICM can be computationally prohibitive. Instead, the decoder for BICM proposed in [6] and assumed in [5] uses bit-wise metric as input to the binary decoder rather than symbol-wise metric. This is in contrast to TCM where symbol-wise metric is explicitly computed for decoding the overall code in the Viterbi algorithm, and to multilevel codes in which symbol-wise metric is implicitly used based on the assumption of correct decisions at lower decoding stages in multistage decoding. As in the block diagram in Fig. 20, the demodulator (demapper) computes bit-wise metric for each codeword symbol, and feed it to the decoder of the binary code $\mathcal{C}$.

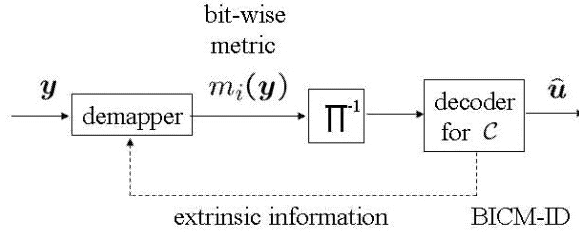


Figure 20: Decoder for BICM (feedback from the decoder to demapper is used for BICM-ID in Sec. 5.4)

Let us first see how it works for Gray mapping. The effective binary channels for metric computation are determined by the three labeling bits as shown in Fig. 21. The signal points associated with value 0 are colored black while those with 1 are colored white. For the rightmost labeling bit of the Gray mapping, the log-likelihood ratio for the bit $c_{1,j}$ in the codeword $\mathbf{c}_1$ is given as Eq. (13) for the first stage decoding of multilevel codes with natural mapping. In the same way, the bit-wise metric $m_2(y_j)$ for the middle labeling bit $c_{2,j}$ is computed with respect to the binary channel defined by $\mathcal{X}_{*0*}$ and $\mathcal{X}_{*1*}$, while $m_3(y_j)$ for $c_{3,j}$ is given with respect to $\mathcal{X}_{0**}$ and $\mathcal{X}_{1**}$. These bit-wise metrics are used in the decoder for the binary code as in Eq. (5). Based on this computation, the decoder for the binary code effectively encounters three parallel binary channels defined by each labeling bit as illustrated

for Gray mapping in Fig. 21. This treatment essentially deals with these parallel channels as if they were independent as in Fig. 22. However, this decoding is not optimum in terms of error probability in general even with a maximum likelihood decoder is used for the binary code $\mathcal{C}$.

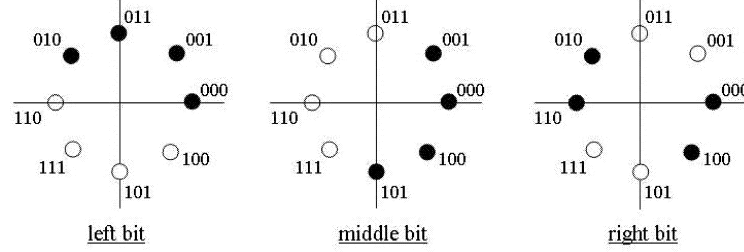


Figure 21: Parallel binary channels in BICM decoding with Gray mapping

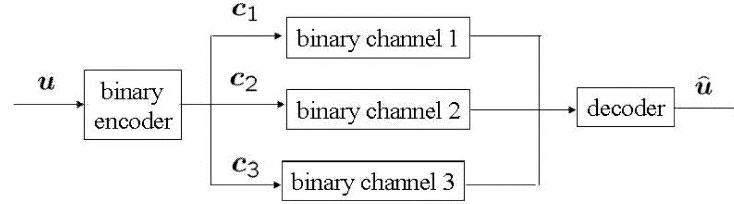


Figure 22: Equivalent channel for BICM with decoding by bit-wise metric.

5.3 BICM Capacity and Capacity-Approaching Codes

In [5], a quantity called BICM capacity was defined as the sum of constrained capacity of the M parallel binary channels. The BICM capacity of the 8-PSK constellation over the AWGN channel under Gray and natural mappings as well as $I_{8\text{PSK}}$ are shown in Fig. 24. It is observed that the BICM capacity for the Gray mapping is almost identical to $I_{8\text{PSK}}$ at rates larger than 2 bits per channel use despite the suboptimality of the bit-wise metric. This fact indicates that the three parallel channels in Fig. 21 are “almost independent” for this mapping when SNR is high. It also advocates the use of turbo codes and LDPC codes in BICM together with Gray mapping to approach the Shannon limit of the underlying constellation. Numerical results based on turbo codes [72] and irregular LDPC codes [65] actually exhibit near-Shannon limit performance as in the case of binary signaling, and verifies this view.

On the other hand, BICM capacity of natural mapping is much smaller at wide range of SNR. This observation would be understood by noticing the difference in the leftmost binary channel in Fig. 21 and the rightmost binary channel in Fig. 23 depicted for natural mapping. Indeed, the capacity of the former and the latter channels correspond to the first stage decoding of the multilevel codes in Fig. 15 for natural mapping and Fig. 17 for Gray mapping, respectively.

On the other hand, the study of the desired mappings at (asymptotically) low SNR regimes is more involved [73][74][75]. One of the findings is that the Gray mapping is not optimal in contrast to the high SNR scenario.

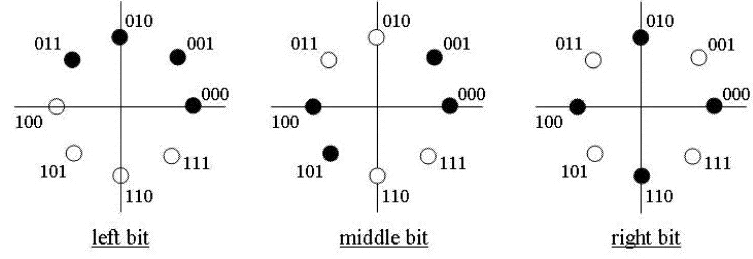


Figure 23: Parallel binary channels in BICM decoding with natural labeling

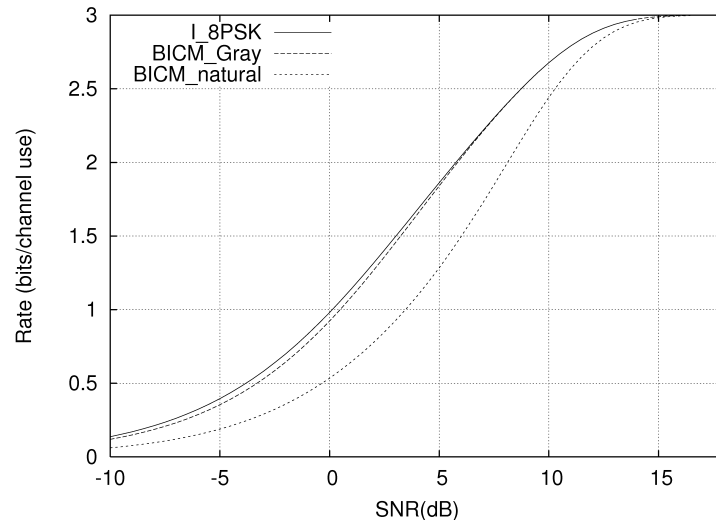


Figure 24: BICM capacity of 8-PSK under Gray and natural mapping compared with $I_{8\text{-PSK}}$.

5.4 BICM with Iterative Decoding

The decoding for BICM in the previous subsection is “one-shot” in the sense that the decoder for the binary code is activated only once. The suboptimality of this approach in Sec. 5.2, especially for natural mapping, comes from the fact that only bit-wise metric is computed without any information on the other labeling bits. BICM with iterative decoding (BICM-ID) in [76][77][78] at least partially overcomes this problem. As in turbo decoding, the extrinsic information for each symbol in a codeword is computed in the SISO decoder, and fed back to the demapper as in Fig. 20. A series of demapping and decoding is iteratively performed in contrast to the original BICM decoding. Suppose that the *a priori* probability that a signal point x is transmitted is available and given as $P(x)$. Then the demapper computes

$$m_1(y_j) = \ln \frac{\sum_{x \in \mathcal{X}_{*0}} f(y_j|x)P(x)}{\sum_{x \in \mathcal{X}_{*1}} f(y_j|x)P(x)} \quad (19)$$

for the updated log-likelihood ratio for $c_{1,j}$. The values for all the symbols are passed to the SISO decoder for the binary code as the extrinsic information for this binary symbol. Again, the signal sets to be used for $c_{2,j}$ are $\mathcal{X}_{*0*}$ and $\mathcal{X}_{*1*}$ for the nominator and denominator of Eq. (19), respectively. Similarly, $\mathcal{X}_{0**}$ and $\mathcal{X}_{1**}$ are used for $c_{3,j}$. Note that in the metric computation by Eq. (13) in the decoding scheme of Sec. 5.2 uniform distribution over $\mathcal{X}$, or $P(x) = 1/2^M$ is implicitly assumed. Let us see how the *a priori* probability on the transmitted signal points for the modified natural mapping in Fig. 25 in which the labels 011 and 111 of natural mapping are switched.

- Suppose that the extrinsic information from the SISO decoder suggests that the signal points with $x^1 = x^2 = 0$ would have been transmitted with high probability. As indicated by an arrow in the leftmost picture in Fig. 25, the effective binary channel for the leftmost labeling bit then only consists of the signal points with labels 100 and 000 whose squared Euclidean distance is 4.0. In contrast, for the Gray mapping the smallest distance between those signal points in different subsets (like those with “010” and “111”) is 0.586 as found in Fig. 21. This observation indicates that the minimum distance of the overall code $d_{\min,C}$ with the mapping in Fig. 25 can be much larger than that of the Gray mapping if iterative decoding is successful.
- Provided that a signal point with $x^2 = 0$ and $x^3 = 0$ is thought to be transmitted owing to the extrinsic information from the decoder. For the rightmost labeling bit x^1 , the updated log-likelihood ratio is to be computed only with respect to the two signal points (000 and 001) as bridged by an arrow in the rightmost figure. As a result, the number of nearest neighbors for the signal point with the opposite subset can be reduced: indeed, the signal point with 000 originally had two neighbors with 001 and 011).

Through these two effects, error performance of BICM can be significantly improved by introducing iterative decoding especially for (modified) natural mapping. In contrast, the improvement is much less in the case of the Gray mapping, due to the fact mentioned in Sec. 5.2 that the three parallel channels are nearly independent at high SNR. Accordingly, BICM with decoding in Sec. 5.2 and Gray mapping is a simple and low-complexity approach for use with capacity-approaching binary codes (assuming that the error floor appears at low error probability). On the other hand, BICM-ID with the mapping like Fig. 25 may be preferred for convolutionally coded system because of the expected relatively small $d_{\min,C}$ with the Gray mapping.

The *a priori* probability used in Eq. (19) for each signal point is computed in the following way. Suppose that the SISO decoder for the binary code provides the extrinsic information

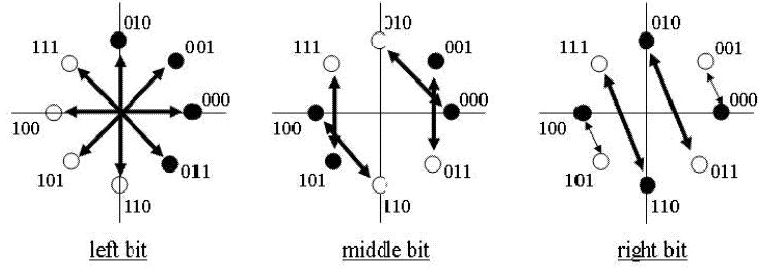


Figure 25: Effective binary channel for modified natural mapping

for the code symbol $c_{i,j}$ in the vector $\mathbf{c}_i$ in the form of $P(c_{i,j} = 0)$ and $P(c_{i,j} = 1)$. Also, the M -bit string $(c_{1,j}, c_{2,j}, \dots, c_{M,j})$ is supposed to be mapped to a signal point x with a label $(x^1, \dots, x^M)$. The *a priori* probability for the signal point x is given by

$$P(x) = \prod_{i' \in \{1, \dots, M\} \setminus \{i\}} P(c_{i',j} = x^j). \quad (20)$$

Note that in computing the updated log-likelihood ratio for the i -th level symbol $c_{i,j}$, the *a priori* probability $P(c_{i,j} = x^j)$ for that symbol is excluded as in turbo decoding. Note that in practice the extrinsic information is given in the form of L -value [47], and the computations in Eq. (19) are indeed in the log-domain as in [47][79].

Performance of BICM-ID in the waterfall (turbo cliff) region depends on the behavior of the SISO decoder and the demapper and often evaluated by the EXIT chart [80] especially for large block lengths. On the other hand, error performance in the error floor region depends on the distance spectrum of the overall code assuming that almost optimum decoding is achieved. Various mappings for BICM-ID are investigated in [81][82][83][84] based on these two criteria. The underlying principle of BICM-ID decoding has been widely used in various communication scenarios including iterative detection of differentially encoded signals [85][86] and detection on the multiple-input/multiple-output (MIMO) channels [87].

5.5 Shaping and BICM

As illustrated in Sec. 2.2, there is a gap between the capacity C_{AWGN} of the AWGN channel and the constrained capacity $I_{\mathcal{X}}$ of the signal set $\mathcal{X}$. Shaping techniques to fill up this gap is of increasing interest as the Shannon limit of non-binary alphabet are approached within 1 dB for large block lengths through the use of iteratively decodable codes in coded modulations, as we saw in Secs. 3.3, 4.3.2 and 5.4.

A practical implementation of shaping in the context of coded modulation is obtained for trellis codes and multilevel coding which accompany set partitioning by applying the shaping techniques [28][32] at the highest level being decoupled from coding structure at lower levels. As an instance, combination of multilevel codes and trellis shaping [88] is investigated in [64].

On the other hand, combined coding and shaping scheme have been recently studied in the context of BICM and BICM-ID. One approach for BICM is the so-called geometrical shaping, in which signal points are non-equally spaced so that the discrete probability distribution approximates the Gaussian distribution. The asymptotic optimality of this approach was proved in [89]. Accordingly, the signal points are densely populated near the center and sparsely at the edge of the constellation. In [90][91][92], combined coding and shaping gain due to the use of turbo codes were examined and it is shown that large shaping gains are achieved over the conventional PAM. An instance of 16-PAM shown in [90] is $\mathcal{X} = \{\pm 0.08, \pm 0.25, \pm 0.42, \pm 0.60, \pm 0.81, \pm 1.05, \pm 1.37, \pm 1.94\}$ which achieves about 0.6dB

shaping gain with respect to $I_{16\text{PAM}}$. Capacity-approaching performance by non-binary LDPC codes in [93] also follows this line.

The other approach is the probabilistic shaping that allows the mapping ϕ to be many-to-one to yield non-uniform probability distribution [94][95]. In [94], a mapping $\phi : \{0, 1\}^6 \rightarrow \mathcal{X}$, where $\mathcal{X}$ is the equally-spaced 16-PAM, is employed. The probability of each signal point to be transmitted is set to an integral power of two so that it approximates Maxwell-Boltzmann distribution: namely, $1/64$ for four points $(\pm 15A, \pm 13A)$, $1/32$ for two points $(\pm 11A)$, $1/16$ for six points $(\pm 9A, \pm 7A, \pm 5A)$, and $1/8$ for four points $(\pm 3A, \pm A)$, where A is the normalization factor. The mapping for this signal set follows from the Huffman prefix code [96], but part of the turbo codeword symbols are punctured in order to keep the rate constant. Since the bit-wise metric for the punctured bits can be poor, use of BICM-ID is indispensable in ensuring coding and shaping gain with this scheme. It was shown in [97] that signal set with non-uniform distribution provides shaping gain to BICM (without iterative decoding) and the information rate approaches that of the optimum coded modulation techniques. The use of a tailored shaping encoder which yields a non-uniform distribution over subsets of the constellation was proposed in [98].

Another approach for achieving shaping gain is the superposition modulation [99][100] where the mapping ϕ is expressed as

$$x = \sum_{i=1}^M \mu_i (-1)^{c_{i,j}}, \quad (21)$$

where the set of constants $\mu_1, \dots, \mu_M$ (> 0) is carefully determined to let the mutual information $I(X; Y)$ be close to C_{AWGN} . Such signal sets for BICM-ID were studied in [101][102].

References

- [1] J.L. Massey, "Coding and modulation in digital communications," in *Proc. of the 1974 Int. Zurich Seminar on Digital Commun.*, Zurich, Switzerland, pp. E2(1)-E2(4), Mar. 1974.
- [2] G. Ungerboeck and I. Csajka, "On improving data-link performance by increasing channel alphabet and introducing sequence coding," in *Proc. of Int. Symp. Inform. Theory*, Ronneby, Sweden, June 1976.
- [3] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol.IT-28, pp.55-67, Jan. 1982.
- [4] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inform. Theory*, vol. IT-23, no.3, pp.371-377, May 1977.
- [5] G. Caire, G. Tarrico and E. Biglieri, "Bit interleaved coded modulation," *IEEE Trans. on Inform. Theory*, vol. 44, no. 3, pp. 927-946, May 1998.
- [6] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. on Commun.*, vol. 40, pp. 873-884, May 1992.
- [7] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding: Turbo codes," in *Proc. of ICC93*, Geneva, Switzerland, pp. 1064-1070, May 1993.
- [8] C. Berrou, A. Glavieux, "Near optimum error-correcting coding: Turbo codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261-1271, Oct. 1996.
- [9] R.G. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory*, vol. IT-8, pp.21-28, Jan. 1962.

- [10] R.G. Gallager, *Low-density parity-check codes*, MIT Press, 1963.
- [11] G.D. Forney, Jr., R.G. Gallager, G.R. Lang, F.M. Longstaff, and S.U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas in Commun.*, vol. 2, no. 5, pp. 632-647, Sep. 1984.
- [12] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets Part I: Introduction," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 5-11, Feb. 1987.
- [13] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets Part II: State of the art," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 12-25, Feb. 1987.
- [14] G.D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2384-2415, 1998.
- [15] D.J. Costello, Jr., J. Hagenauer, H. Imai, and S.B. Wicker, "Applications of error-control coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2531-2560, Oct. 1998.
- [16] A.R. Calderbank, "The art of signaling: Fifty years of coding theory," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2561-2595, 1998.
- [17] E. Biglieri, D. Divsalar, M.K. Simon, and P.J. McLane, *Introduction to Trellis-Coded Modulation with Applications*, Prentice Hall, 1991.
- [18] S.H. Jamali and T. Le-Ngoc, *Coded-modulation techniques for fading channels*, Springer, 1994.
- [19] A. Guillén i Fabregas, A. Martinez, and G. Caire, *Bit-interleaved coded modulation*, Foundations and Trends in Communications and Information Theory 5 (1-2), now Publishers, 2008.
- [20] J.B. Anderson and A. Svensson, *Coded modulation systems*, Kluwer Academic/Plenum Publishers, 2003.
- [21] C.B. Schlegel and L.C. Pérez, *Trellis and turbo coding*, Wiley Interscience, 2004.
- [22] M. Franceschini, G. Ferrari, and R. Raheli, *LDPC Coded Modulations*, Springer, 2009.
- [23] E. Biglieri, *Coding for wireless channels*, Springer, 2005.
- [24] S. Lin and D.J. Costello, Jr., *Error control coding*, 2nd Ed., Pearson, 2004.
- [25] S.G. Wilson, *Digital Modulation and Coding*, Prentice Hall, 1986.
- [26] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*, Kluwer Academic/Plenum Publishers, 1999.
- [27] R.G. Gallager, *Information theory and reliable communication*, John Wiley, 1968.
- [28] R.F.H. Fischer, *Precoding and signal shaping for digital transmission*, Wiley, 2002.
- [29] D. Divsalar and M. K. Simon, "The design of trellis codes for fading channels: Performance criteria," *IEEE Trans. on Commun.*, vol. 36, No. 9, Sep. 1988.
- [30] R.D. Wesel, X. Liu, J.M. Cioffi, and C. Komninakis, "Constellation labeling for linear encoders," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2417-2431, Sep. 2001.
- [31] E. Agrell, J. Lassing, E.G. Ström, and T. Ottoson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Infor. Theory*, vol. 50, no. 12, pp. 3170-3182, Dec. 2004.

- [32] S.A. Tretter, *Constellation shaping, nonlinear precoding, and trellis coding for voiceband telephone channel modems*, Kluwer Academic Publishers, 2002.
- [33] E. Zehavi and J.K. Wolf, "On the performance evaluation of trellis codes," *IEEE Trans. Inform. Theory*, vol. 33, no. 2, pp. 196-202, Mar. 1987.
- [34] G.D. Forney, Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, no. 5, pp. 1241-1260, Sep. 1991.
- [35] S. Benedetto, M. Mondin, and G. Montorsi, "Performance evaluation of trellis-coded modulation schemes," *Proc. of IEEE*, vol. 82, no. 6, pp. 833-855, June 1994.
- [36] A. Alvarado, A. Graell i Amat, F. Brännström, and E. Agrell, "On optimal TCM encoders," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2178-2189, Oct. 2012.
- [37] J.K. Cavers and P. Ho, "Analysis of the error performance of trellis-coded modulations in Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 40, no. 1, pp. 74-83, Jan. 1992.
- [38] D. Divsalar and M.K. Simon, "Trellis coded modulation for 4800-9600 bits/s transmission over a fading mobile satellite channel," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 162-175, Feb. 1987.
- [39] C. Schlegel and D.J. Costello, Jr., "Bandwidth efficient coding for fading channels: Code construction and performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 12, pp. 1356-1368, Dec. 1987.
- [40] L.-F. Wei, "Rotationally invariant convolutional channel encoding with expanded signal space, Part II: Nonlinear codes," *IEEE J. Sel. Areas in Commun.*, vol. 2, pp. 672-686, Sep. 1984.
- [41] L.-F. Wei, "Trellis-coded modulation using multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. 33, no. 4, pp. 483-501, July 1987.
- [42] S.S. Pietrobon, G. Ungerboeck, L.C. Perez, D.J. Costello, Jr., "Trellis-coded multidimensional phase modulation," *IEEE Trans. Inform. Theory*, vol. 36, pp. 63-89, Jan. 1990.
- [43] L.-F. Wei, "Rotationally invariant trellis-coded modulations with multidimensional MPSK," *IEEE J. Sel. Areas in Commun.*, vol. 7, pp. 1285-1295, Dec. 1989.
- [44] G.D. Forney, Jr., and L.-F. Wei, "Multidimensional constellations - Part I: Introduction, figures of merit, and generalized constellations," *IEEE J. Sel. Areas in Commun.*, vol. 7, no. 6, pp. 877-892, Aug. 1989.
- [45] A.R. Calderbank and N.J.A. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inform. Theory*, vol. 33, pp. 177-195, 1987.
- [46] G.D. Forney, Jr., "Coset codes - Part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1123-1151, Sep. 1988.
- [47] J. Hagenauer, E. Offer and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory*, vol. 42, no.2, pp.429-445, March 1996.
- [48] P. Robertson and T. Woertz, "Bandwidth-efficient Turbo trellis-coded modulation using punctured component codes," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 206-218, Feb. 1998.
- [49] L.R. Bahl, J. Cocke, F. Jelinek and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Inform. Theory*, vol.20, pp.284-287, Mar. 1974.

- [50] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara, "Parallel concatenated trellis coded modulation," in *Proc. of ICC96*, pp. 962-967, Dallas, TX, June 1996.
- [51] S. Benedetto, S. Divsalar, G. Montorsi and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design and iterative decoding," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 909-926, May 1998.
- [52] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara, "Serial concatenated trellis coded modulation with iterative decoding: design and performance," in *Proc. of CTMC97, GLOBECOM97*, pp. 38-43, Phoenix, AZ, Nov. 1997.
- [53] D. Divsalar, S. Dolinar, and F. Pollara, "Serial concatenated trellis coded modulation with rate-1 inner code," in *Proc. of IEEE Global Telecommun. Conf.*, pp. 777-782, San Francisco, CA, Nov.-Dec. 2000.
- [54] H.M. Tullberg and P.H. Siegel, "Serial concatenated TCM with an inner accumulate code - Part I: Maximum-likelihood analysis," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 64-73, Jan. 2005.
- [55] G.D. Forney, Jr., M.D. Trott, and S.-Y. Chung, "Sphere bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 820-850, May 2000.
- [56] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 40, pp. 409-417, Mar. 1994.
- [57] T. Takata, S. Ujita, T. Kasami, and S. Lin, "Multistage decoding of multilevel block M -PSK modulation codes and its performance analysis," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1204-1218, July 1993.
- [58] Y. Kofman, E. Zehavi, and S. Shamai, "Performance analysis of a multilevel coded modulation system," *IEEE Trans. Commun.*, vol. 42, pp. 299-312, Feb./Mar./Apr. 1994.
- [59] K. Engdahl and K. Zigangirov, "On the calculation of the error probability for a multilevel modulation scheme using QAM-signaling," *IEEE Trans. Inform. Theory*, vol. 44, no. 4, pp. 1612-1620, July 1998.
- [60] H. Herzberg and G. Poltyrev, "Techniques of bounding the probability of decoding error for block coded modulation structures," *IEEE Trans. Inform. Theory*, vol. 40, no. 3, pp. 903-911, May 1994.
- [61] H. Herzberg and G. Poltyrev, "The error probability of M -ary PSK block coded modulation schemes," *IEEE Trans. Commun.*, vol. 44, no. 4, pp. 427-433, Apr. 1996.
- [62] A.G. Burr and T.J. Lunn, "Block-coded modulation optimized for finite error rate on the white Gaussian noise channel," *IEEE Trans. Inform. Theory*, pp. 373-385, vol. 42, no. 1, pp. 373-385, Jan. 1997.
- [63] U. Wachsmann and J. Huber, "Power and bandwidth efficient digital communication using Turbo codes in multilevel codes," *European Trans. Telecommun.*, vol. 6, no. 5, 1995.
- [64] U. Wachsmann, R.F.H. Fischer and J.B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1361-1391, July 1999.
- [65] J. Hou, P. H. Siegel, L.B. Milstein, and H.D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 49, no. 9, pp. 2141-2155, Sep. 2003.

- [66] T. Woertz and J. Hagenauer, "Decoding of M-PSK-Multilevel codes," *European Trans. Telecommun.*, vol. 4, no. 3, pp. 299-308, May-June 1993.
- [67] M. Isaka and H. Imai, "On the iterative decoding of multilevel codes," *IEEE J. Sel. Areas in Commun.*, vol. 19, no. 5, pp. 935-943, May, 2001.
- [68] P.A. Martin and D.P. Taylor, "On multilevel codes and iterative multistage decoding," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1916-1925, Nov. 2001.
- [69] A.R. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1234-1248, July 1993.
- [70] L.-F. Wei, "Coded modulation with unequal error protection," *IEEE Trans. Commun.*, vol. 41, pp. 1439-1449, Oct. 1993.
- [71] R.H. Morelos-Zaragoza, M.P.C. Fossorier, S. Lin and H. Imai "Multilevel coded modulation for unequal error protection and multistage decoding — part I: symmetric constellations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 204-213, Feb. 2000.
- [72] S. Le Goff, A. Glavieux and C. Berrou, "Turbo-code and high spectral efficiency modulation," in *Proc. of International Conference on Communications (ICC94)*, pp. 645-649, New Orleans, USA, May 1994.
- [73] A. Martinez, A. Guillén i Fabregas, G. Caire and F. Willems, "Bit-interleaved coded modulation in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5447-5455, Dec. 2008.
- [74] C. Stiestorfer and R.F.H. Fischer, "Mappings for BICM in UWB scenarios," in *Proc. of 7th Int. Conf. Source and Channel Coding*, Ulm, Germany, Jan. 2008.
- [75] C. Stiestorfer and R.F.H. Fischer, "Asymptotically optimal mappings for BICM with M-PAM and M^2 -QAM," *IET Electronics Lett.*, vol. 45, no. 3, Jan. 2009.
- [76] X. Li and J.A. Ritcey, "Bit interleaved coded modulation with iterative decoding," in *Proc. of IEEE ICC99*, Vancouver, Canada, June, 1999.
- [77] X. Li, A. Chindapol, and J.A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8PSK signaling," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 1250-1257, Aug. 2002.
- [78] S. ten Brink, J. Spiedel and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. of Global Communications Conference (GLOBECOM98)*, pp. 579-584, Sydney, Australia, Nov. 1998.
- [79] P. Robertson, P. Hoeher, and E. Villebrun, "Optimal and suboptimal maximum a posteriori algorithms suitable for turbo decoding," *European Trans. Telecommun.*, vol. 8, pp. 119-125, Mar.-Apr. 1997.
- [80] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727-1737, Oct. 2001.
- [81] F. Schreckenbach, N. Görtz, J. Hagenauer, and G. Bauch, "Optimization of symbol mappings for bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 7, no. 12, pp. 593-595, Dec. 2003.
- [82] J. Tan and G.L. Stüber, "Analysis and design of symbol mappers for iteratively decoded BICM," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 662-672, Mar. 2005.

- [83] N.H. Tran and H.H. Nguyen, "Signal mapping of 8-ary constellations for bit interleaved coded modulation with iterative decoding," *IEEE Trans. Broadcast.*, vol. 52, no. 1, pp. 92-99, Mar. 2006.
- [84] F. Brännström and L.K. Rasmussen, "Classification of 8PSK mappings for BICM," in *Proc. of 2007 IEEE International Symposium on Inform. Theory*, Nice, France, June 2007.
- [85] P. Hoeher and J. Lodge, "'Turbo DPSK': iterative differential PSK demodulation and channel decodings," *IEEE Trans. on Commun.*, vol. 47, no. 6, pp. 837-843, June 1999.
- [86] K.R. Narayanan and G.L. Stuber, "A serial concatenation approach to iterative demodulation and decoding," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 956-961, July 1999.
- [87] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670-678, Apr. 2004.
- [88] G.D. Forney, Jr., "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 281-300, Mar. 1992.
- [89] F.W. Sun and H.C.A. van Tilborg, "Approaching capacity by equiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 39, no. 5, pp. 1714-1716, Sep. 1993.
- [90] D. Sommer and G.P. Fettweis, "Signal shaping by non-uniform QAM for AWGN channel and applications to turbo coding,"—in *Proc. of ITG Conf. Source and Channel Coding*, pp. 81-86, Jan. 2000.
- [91] C. Fragouli, R.D. Wesel, D. Sommer, and G.P. Fettweis, "Turbo codes with nonuniform constellations," in *Proc. of IEEE Int. Conf. Commun. (ICC2001)*, pp. 70-73, June 2001.
- [92] Md.J. Hossain, A. Alvarado, and L. Szczecinski, "BICM transmission using non-uniform QAM constellations: Performance analysis and design," in *Proc. of IEEE Int. Conf. Commun.*, 2010.
- [93] A. Bennatan and D. Burstein, "Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 549-583, 2006.
- [94] D. Raphaeli and A. Gurevits, "Constellation shaping for pragmatic turbo-coded modulation for high spectral efficiency," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 341-345, Mar. 2004.
- [95] F. Schreckenbach and P. Henkel, "Signal shaping using non-unique symbol mappings," in *Proc. of 43rd Annual Allerton Conf. on Commun., Cont., and Comp.*, Monticello, IL, Sep. 2005.
- [96] F.R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 913-928, May 1993.
- [97] A.G. i Fabregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *Proc. of IEEE Inform. Theory Workshop*, Dublin, Ireland, 2010.
- [98] S. Le Goff, B. K. Khoo, and C. C. Tsimenidis, "Constellation shaping for bandwidth-efficient turbo-coded modulation with iterative receiver," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2223-2233, June 2007.

- [99] L. Duan, B. Rimoldi, and R. Urbanke, "Approaching the AWGN channel capacity without active shaping," in *Proc. of IEEE Int. Symp. Inform. Theory*, p. 34, Ulm, Germany, June/July 1997.
- [100] X. Ma and L. Ping, "Coded modulation using superimposed binary codes," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3331-3343, Dec. 2004.
- [101] H.S. Cronie, "Signal shaping for bit-interleaved coded modulation on the AWGN channel," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3428-3435, Dec. 2010.
- [102] P.A. Hoeher and T. Wo, "Superposition modulation: Myths and facts," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 110-116, Dec. 2011.