

Multilevel Codes: Theoretical Concepts and Practical Design Rules

Udo Wachsmann, *Member, IEEE*, Robert F. H. Fischer, *Member, IEEE*, and Johannes B. Huber, *Member, IEEE*

Abstract— This paper deals with 2^ℓ -ary transmission using multilevel coding (MLC) and multistage decoding (MSD). The known result that MLC and MSD suffice to approach capacity if the rates at each level are appropriately chosen is reviewed. Using multiuser information theory, it is shown that there is a large space of rate combinations such that MLC and full maximum-likelihood decoding (MLD) can approach capacity. It is noted that multilevel codes designed according to the traditional balanced distance rule tend to fall in the latter category and, therefore, require the huge complexity of MLD. The capacity rule, the balanced distances rules, and two other rules based on the random coding exponent and cutoff rate are compared and contrasted for practical design. Simulation results using multilevel binary turbo codes show that capacity can in fact be closely approached at high bandwidth efficiencies. Moreover, topics relevant in practical applications such as signal set labeling, dimensionality of the constituent constellation, and hard-decision decoding are emphasized. Bit interleaved coded modulation, recently proposed by Caire *et al.* is reviewed in the context of MLC. Finally, the combination of signal shaping and coding is discussed. Significant shaping gains are achievable in practice only if these design rules are taken into account.

Index Terms— Bit-interleaved coded modulation, channel capacity, dimensionality, hard decision, multilevel coding, multistage decoding, set partitioning, signal shaping, trellis-coded modulation.

I. INTRODUCTION

THE idea of *coded modulation* is to jointly optimize coding and modulation in order to improve the performance of digital transmission schemes, see, e.g., [2]. Independently of each other, the most powerful applicable coded modulation schemes were presented in 1976/1977 by Ungerboeck [3], [4] and Imai and Hirakawa [5]. The common core is to optimize

the code in Euclidean space rather than dealing with Hamming distance as in classical coding schemes.

Ungerboeck's approach to coded modulation is based on mapping by set partitioning. Thereby, the signal set (constellation) $\mathcal{A} = \{a_0, a_1, \dots, a_{M-1}\}$ of an $M = 2^\ell$ -ary modulation scheme is successively binary partitioned in ℓ steps defining a mapping of binary addresses $\mathbf{x} = (x^0, x^1, \dots, x^{\ell-1})$ to signal points a_m . In almost all work dealing with coded modulation the set partitioning strategy introduced by Ungerboeck is chosen: maximize the minimum intra-subset Euclidean distance. In the encoder, the binary addresses are usually divided into two parts: the least significant binary symbols are convolutionally encoded and the most significant binary symbols (if present) remain uncoded. The code parameters are chosen by means of an exhaustive computer search in order to maximize the minimum distance of the coded sequences in Euclidean space. Because of the trellis constraints on sequences of signal points, Ungerboeck's approach to coded modulation is named *trellis-coded modulation* (TCM). Originally, TCM was proposed for one- and two-dimensional signal sets using one bit redundancy per signal point. Significant work was performed in order to provide more flexible transmission rates with TCM, using signal constellations in higher dimensions [6], [7] or signal constellations derived from lattice theory, e.g., [8]–[13]. Alternatively, a pragmatic approach of punctured TCM was proposed in [14]–[16]. Further work was done to achieve desired properties such as rotational invariance [17]–[21] or spectral zeros [22], [23].

Imai's idea of *multilevel coding* (MLC) is to protect each address bit x^i of the signal point by an individual binary code \mathcal{C}^i at level i . (In retrospect, the constructions B through E for dense lattices reviewed in [24] can be regarded as the first MLC approach.) Originally, MLC was proposed for one-dimensional signaling combined with labeling by binary counting off the signal levels. The individual codes were chosen in such a way that the minimum distance of the Euclidean space code was maximized. In the following we refer to this concept of assigning codes to the individual levels as *balanced distances rule* (see Section II). At the receiver side, each code \mathcal{C}^i is decoded individually starting from the lowest level and taking into account decisions of prior decoding stages. This procedure is called *multistage decoding* (MSD). In contrast to Ungerboeck's TCM, the MLC approach provides flexible transmission rates, because it decouples the dimensionality of the signal constellation from the code rate. Additionally, any code, e.g., block codes, convolutional codes, or concatenated codes, can be used as component code.

Manuscript received August 30, 1997; revised July 4, 1998. This work was supported in part by Deutsche Forschungsgemeinschaft under Contract Hu 634/1-1. The material in this paper was presented in part at the IEEE Information Theory Workshop (ITW'95), Rydzyna, Poland, June 1995; at the IEEE International Symposium on Information Theory (ISIT'95), Whistler, BC, Canada, September 1995; at the Mediterranean Workshop on Coding and Information Integrity, Palm de Mallorca, Spain, February–March 1996; the IEEE Information Theory Workshop (ITW'96), Haifa, Israel, June 1996; the CTMC at the IEEE Global Telecommunications Conference (GLOBECOM'96), London, U.K., November 1996; the IEEE International Symposium on Information Theory (ISIT'97), Ulm, Germany, July 1997; and the ITG-Fachtagung: Codierung für Quelle, Kanal und Übertragung (Source and Channel Coding), Aachen, Germany, March 1998.

U. Wachsmann is with Ericsson Eurolab GmbH, D-90411 Nürnberg, Germany (e-mail: Udo.Wachsmann@eedn.ericsson.se).

R. F. H. Fischer and J. B. Huber are with Lehrstuhl für Nachrichtentechnik II, Universität Erlangen-Nürnberg, D-91058 Erlangen, Germany (e-mail: {fischer}{huber}@nt.e-technik.uni-erlangen.de).

Communicated by A. Vardy, Associate Editor for Coding Theory.

Publisher Item Identifier S 0018-9448(99)04373-4.

Although MLC offers excellent asymptotic coding gains, it achieved only theoretical interest in the past. In practice, system performance was severely degraded due to high error rates at low levels. A lot of effort was devoted to overcome this effect, see, e.g., [25].

A straightforward generalization of Imai's approach is to use q -ary, $q > 2$, component codes based on a nonbinary partitioning of the signal set. In this context, TCM is a special case of MLC using a single convolutional code with a nonbinary output alphabet while higher levels remain uncoded, see, e.g., [26], [27]. In this paper we restrict ourselves to binary codes, because a) the large class of binary codes has been well established in coding theory for a long time, b) mainly binary codes are of practical interest, and c) binary codes in conjunction with multilevel codes turn out to be asymptotically optimum, see Section III.

In 1988, Forney presented the concept of coset codes [28], [29], a class of codes originally considered in [30]. By dealing only with infinite constellations (neglecting boundary effects) and using the mathematics of lattice theory, a general class of codes was established. Similar to TCM, cosets instead of signal points are selected in the encoding process. Coset codes divide into two classes: *trellis codes* (a generalization of TCM) which employ a convolutional encoder and *lattice codes* (based on block codes) where the signal points in N dimensions exhibit group structure. Lattice codes can also be generated by the MLC approach, if the individual codes are subcodes of each other. Already de Buda [31] stated that lattice codes can approach the channel capacity of the additive white Gaussian noise (AWGN) channel. The proof was recently refined by Urbanke and Rimoldi [32] as well as by Forney [33].

For practical coded modulation schemes where boundary effects have to be taken into account, Huber *et al.* [34]–[37] and Kofman *et al.* [38], [39] independently proved that the capacity of the modulation scheme can be achieved by multilevel codes together with multistage decoding if and only if (iff) the individual rates of the component codes are properly chosen. In these papers, equiprobable signal points and regular partitioning were assumed. In this paper, the results are generalized for arbitrary signaling and labeling of signal points. We present theoretical concepts for the design and analysis of practical coded-modulation schemes. The key point is the well-known chain rule for mutual information [40], [41]. As shown in Section II, the chain rule provides a model with virtually independent parallel channels for each address bit at the different partitioning levels, called equivalent channels. Considering the information-theoretic parameters of these channels leads to

- a) *theoretical statements* about coded modulation and
- b) *practical rules* for designing and constructing coded modulation schemes.

The main intention of this paper is to show that power- and bandwidth-efficient digital communication close to theoretical limits is possible with properly designed MLC schemes not only in theory but also in practice. For that purpose a variety of aspects of coded modulation are discussed.

The organization of this paper is as follows. In Section II the system model is given and the concept of equivalent channels is introduced. It is shown in Section III that coded modulation via the multilevel coding approach is optimum in the sense of capacity. Additionally, the capacity region for coded modulation schemes when using overall maximum-likelihood decoding is derived, leading to conditions on optimality. In Section IV, random coding exponents as well as cutoff rates of the equivalent channels are investigated. Several rules to assign rates to the individual codes are discussed and compared. In particular, the optimality conditions of the rate distribution according to the traditional balanced distances rule are given and compared to the design according to the capacities of the equivalent channels. Simulation results show that the application of information-theoretical design rules to multilevel codes leads to power- and bandwidth-efficient transmission systems close to the Shannon limit.

In the remaining sections we deal with further aspects of coded modulation schemes which are essential in practice. In Section V, the optimal dimensionality of the signal set, on which MLC should be based, is derived. Although Ungerboeck's strategy is often used as "natural" method of partitioning a signal set, the proof of optimality of MLC does not depend on the actual set partitioning strategy. In Section VI we investigate set partitioning strategies for finite codeword length. Caire *et al.* [42] recently presented the derivation of capacity for a pragmatic approach to coded modulation. In this scheme, only one individual binary code with subsequent bit interleaving is necessary to address the signal points. In terms of capacity the loss with regard to Imai's MLC scheme is surprisingly small if and only if Gray labeling of signal points is employed. A derivation of this scheme starting from the standard approach of MLC and its multistage decoding together with a discussion for finite block length is also addressed in Section VI. In Section VII the use of hard-decision instead of soft-decision decoding at the receiver is investigated.

Since signal shaping is well known to provide an additional gain by replacing a uniformly distributed signal by a Gaussian distributed one in order to reduce average transmit power, we address the optimum combination of shaping and channel coding using MLC in Section VIII. The optimum assignment of code rates to the individual levels and optimum sharing of redundancy between coding and shaping is given.

Section IX summarizes the main results. The Appendix sketches the derivation of the distance profile of multilevel codes and an efficient technique to bound the error rate.

II. SYSTEM MODEL

Consider a modulation scheme with $M = 2^\ell$, $\ell > 1$, signal points in a D -dimensional signal space. The signal points are taken from the signal set $\mathbf{A} = \{a_0, a_1, \dots, a_{M-1}\}$ with $\mathbf{A} \subset \mathbb{R}^D$. Since we mainly focus on the AWGN channel, the channel output signal points \mathbf{y} come from the alphabet $\mathbf{Y} = \mathbb{R}^D$ of real numbers in D dimensions. In order to create powerful Euclidean-space codes for such an M -ary signal alphabet, labels have to be assigned to each signal point,

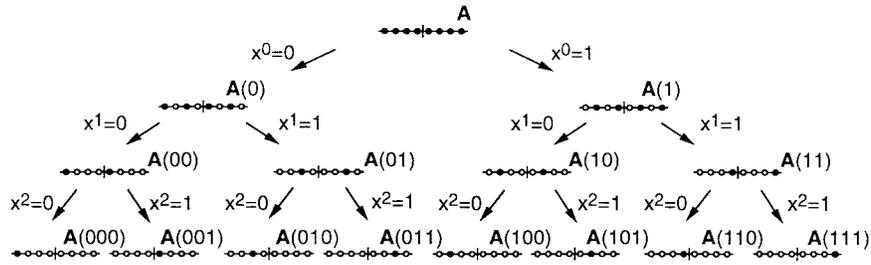


Fig. 1. Binary partitioning of the 8-ASK signal set.

see, e.g., [26]. Therefore, a (bijective) mapping $a = \mathcal{M}(\mathbf{x})$ of binary address vectors $\mathbf{x} = (x^0, x^1, \dots, x^{\ell-1})$, $x^i \in \{0, 1\}$, to signal points $a \in \mathbf{A}$ is defined. Usually, the mapping is derived by successively partitioning the signal set \mathbf{A} into subsets.

As an example, binary partitioning of the 8-ASK (8-ary amplitude shift keying) signal ($D = 1$) set is illustrated in Fig. 1. In contrast to most of the literature, most examples in this paper are based on one-dimensional constellations. This is not only for simplicity of presentation, but also due to the results given in Section V. Of course, M -ary ASK will represent one quadrature component in a QAM scheme in practice.

In the first step, at partitioning level 1, the signal set \mathbf{A} is divided into two parts, namely, the subsets $\mathbf{A}(x^0=0)$ and $\mathbf{A}(x^0=1)$. Then, all subsets $\mathbf{A}(x^0 \dots x^{i-1})$ at partitioning level i are iteratively divided into two further subsets $\mathbf{A}(x^0 \dots x^{i-1}0)$ and $\mathbf{A}(x^0 \dots x^{i-1}1)$ at partitioning level $i+1$. Each subset at partitioning level i is uniquely labeled by the path $(x^0 \dots x^{i-1})$ in the partitioning tree from the root to the subset

$$\mathbf{A}(x^0 \dots x^{i-1}) = \{a = \mathcal{M}(\mathbf{x}) | \mathbf{x} = (x^0 \dots x^{i-1} b^i \dots b^{\ell-1}), b^j \in \{0, 1\}, j = i, \dots, \ell - 1\}. \quad (1)$$

The iteration stops when each subset at level ℓ only contains one signal point. Then, the subset label $(x^0 \dots x^{\ell-1})$ equals the address vector of the signal point. As we will see in Section VI, the particular strategy for this mapping by set partitioning influences the properties of the coded modulation scheme.

Since the mapping \mathcal{M} is bijective independently of the actual partitioning strategy, the mutual information¹ $I(Y; \mathbf{A})$ between the transmitted signal point $a \in \mathbf{A}$ and the received signal point $y \in \mathbf{Y}$ equals the mutual information $I(Y; X^0, X^1, \dots, X^{\ell-1})$ between the address vector $\mathbf{x} \in \{0, 1\}^\ell$ and the received signal point $y \in \mathbf{Y}$. The discrete-time physical channel is characterized by the set $\{f_Y(y|a) | a \in \mathbf{A}\}$ of conditional probability density functions (pdf's) of the received point y given the signal point a .

Applying the chain rule of mutual information [40, p. 22] we get

$$\begin{aligned} I(Y; \mathbf{A}) &= I(Y; X^0, X^1, \dots, X^{\ell-1}) \\ &= I(Y; X^0) + I(Y; X^1 | X^0) + \dots \\ &\quad + I(Y; X^{\ell-1} | X^0, X^1, \dots, X^{\ell-2}). \end{aligned} \quad (2)$$

¹We denote the random variables corresponding to the transmitted and received signal points, the binary address vector, and its components by capital letters.

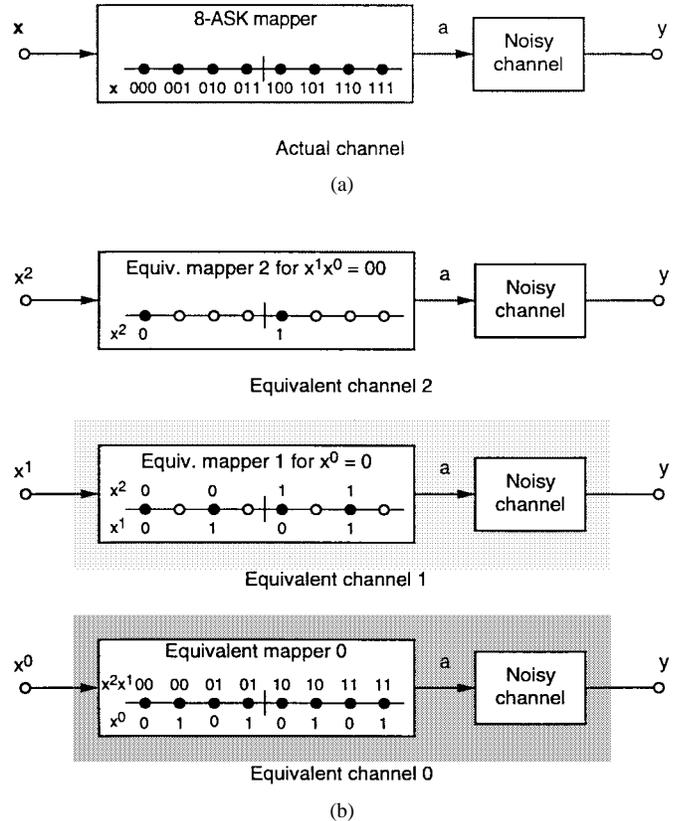


Fig. 2. Equivalent channels for coded 8-ASK modulation.

The equation may be interpreted in the following way. Transmission of vectors with binary digits $x^i, i = 0, \dots, \ell - 1$, over the physical channel can be separated into the parallel transmission of individual digits x^i over ℓ equivalent channels, provided that x^0, \dots, x^{i-1} are known (cf. [5]). This fundamental principle is illustrated in Fig. 2 for 8-ASK modulation and the natural labeling of the signal points. Fig. 2(a) shows the usual transmission scheme with the binary address vector entering the 8-ASK modulator. The addressed signal point is transmitted over the noisy channel. According to the chain rule of mutual information, the scenario illustrated in Fig. 2 (b) provides the same mutual information as Fig. 2 (a). Fig. 2 (b) shows parallel transmission schemes for the individual binary digits x^i . Binary digits x^i of “low” levels $i, i < \ell - 1$, are multiply represented in the modulator signal sets. The underlying signal set of the equivalent modulator i for digit x^i is time-variant depending on the digits of lower levels $j < i$.

For example, in Fig. 2 the equivalent modulators are shown for one time instant where $(x^0 x^1) = (00)$ holds. The actual signal point to be transmitted for digit x^i is selected by digits x^j of higher levels j ; $i < j \leq \ell - 1$. This example illustrates that the equivalent channel i for transmission of the binary digit x^i comprises the time-variant equivalent modulator i and the actual noisy (physical) channel.

From the chain rule the mutual information

$$I(Y; X^i | X^0 \dots X^{i-1})$$

of the equivalent channel i can be easily calculated by (cf. [37], [33])

$$I(Y; X^i | X^0 \dots X^{i-1}) = I(Y; X^i \dots X^{\ell-1} | X^0 \dots X^{i-1}) - I(Y; X^{i+1} \dots X^{\ell-1} | X^0 \dots X^i). \quad (3)$$

Since the subsets at one partitioning level may not be congruent, the mutual information $I(Y; X^i \dots X^{\ell-1} | X^0 \dots X^{i-1})$ is calculated by averaging over all possible combinations of $x^0 \dots x^{i-1}$

$$I(Y; X^i \dots X^{\ell-1} | X^0 \dots X^{i-1}) = E_{x^0 \dots x^{i-1} \in \{0,1\}^i} \{I(Y; X^i \dots X^{\ell-1} | x^0 \dots x^{i-1})\}. \quad (4)$$

This concept of equivalent channels of a coded modulation scheme is the basic tool for the analysis and design of such schemes. For the moment, let the digits x^0, \dots, x^{i-1} of the lower levels $j, j < i$, be fixed. Then, the pdf $f_Y(y | x^i, x^0 \dots x^{i-1})$ characterizes the equivalent channel i . The underlying signal subset for the equivalent modulator i is given by $\mathbf{A}(x^0 \dots x^{i-1})$. For all but the highest level $i, i < \ell - 1$, the binary symbol x^i is multiply represented in the subset $\mathbf{A}(x^0 \dots x^{i-1})$. Therefore, the signal point a is taken from the subset $\mathbf{A}(x^0 \dots x^i)$. Thus the pdf $f_Y(y | x^i, x^0 \dots x^{i-1})$ is given by the expected value of the pdf $f_Y(y | a)$ over all signal points a out of the subset $\mathbf{A}(x^0 \dots x^i)$

$$\begin{aligned} f_Y(y | x^i, x^0 \dots x^{i-1}) &= E_{a \in \mathbf{A}(x^0 \dots x^i)} \{f_Y(y | a)\} \\ &= \frac{1}{\Pr\{\mathbf{A}(x^0 \dots x^i)\}} \sum_{a \in \mathbf{A}(x^0 \dots x^i)} \Pr\{a\} \cdot f_Y(y | a). \end{aligned} \quad (5)$$

In general, the equivalent channel i is completely characterized by a set of probability density functions $f_Y(y | x^i)$ of the received signal point $y \in \mathbf{Y}$ if the binary symbol x^i is transmitted. Due to the time-variant equivalent modulator i , i.e., the underlying subset for transmission of symbol x^i depends on the symbols at lower levels $j, 0 \leq j < i$, this set of pdf's comprises the pdf's $f_Y(y | x^i, x^0 \dots x^{i-1})$ for all combinations of $(x^0 \dots x^{i-1})$

$$f_Y(y | x^i) = \{f_Y(y | x^i, x^0 \dots x^{i-1}) | (x^0 \dots x^{i-1}) \in \{0,1\}^i\}. \quad (6)$$

Note that the characterization of the equivalent channel i by its set of pdf's $f_Y(y | x^i)$ given in (5) and (6) is valid for every memoryless channel that can be characterized by pdf's $f_Y(y | a)$. This fact is quite notable since almost all work

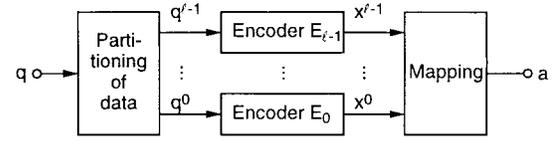


Fig. 3. Multilevel encoder.

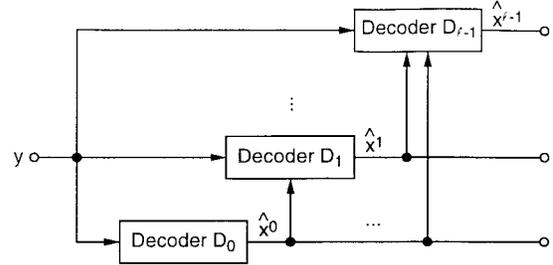


Fig. 4. Multistage decoding.

concerning coded modulation deals with minimum distance as the essential design parameter which is relevant for at most the Gaussian channel. Indeed, we will see later that even for the Gaussian channel optimizing minimum distance does not necessarily lead to asymptotically optimum schemes. With the set of pdf's $f_Y(y | x^i)$ the theoretical treatment of coded modulation schemes is possible for arbitrary probabilities of signal points as well as for arbitrary labeling of signal points. This is the basis for analyzing schemes employing signal shaping (approximately Gaussian-distributed signal points are generated, see Section VIII), as well as for the assessment of schemes with different labelings (Section VI).

Having the above derivation of the equivalent channels in mind, the multilevel coding approach together with its multistage decoding (MSD) procedure, originally presented by Imai and Hirakawa [5], is a straightforward consequence of the chain rule (2). The digits $x^i, i = 0, \dots, \ell - 1$, result from independent encoding of the data symbols. The encoder is sketched in Fig. 3. A block of K binary source data symbols $\mathbf{q} = (q_1, \dots, q_K), q_i \in \{0,1\}$, is partitioned into ℓ blocks

$$\mathbf{q}^i = (q_1^i, \dots, q_{K_i}^i), i = 0, \dots, \ell - 1$$

of length K_i with $\sum_{i=0}^{\ell-1} K_i = K$. Each data block \mathbf{q}^i is fed into an individual binary encoder E_i generating words $\mathbf{x}^i = (x_1^i, \dots, x_{N_i}^i), x_\gamma^i \in \{0,1\}$, of the component code \mathcal{C}^i . For simplicity, we here assume equal code lengths N at all levels, but in principle the choice of the component codes is arbitrary. For example, binary block and convolutional codes or concatenated codes like turbo codes [43], [44] can be employed. The codeword symbols $x_\gamma^i, \gamma = 1, \dots, N$, of the codewords $\mathbf{x}^i, i = 0, \dots, \ell - 1$, at one time instant γ , form the binary label $\mathbf{x}_\gamma = (x_\gamma^0, \dots, x_\gamma^{\ell-1})$, which is mapped to the signal point a_γ . The code rate R of the scheme is equal to the sum of the individual code rates $R^i = K_i/N$, namely,

$$R = \sum_{i=0}^{\ell-1} R^i = \frac{1}{N} \sum_{i=0}^{\ell-1} K_i = \frac{K}{N}.$$

The right-hand side of the chain rule (2) suggests a rule for a low-complexity staged decoding procedure that is well

known as *multistage decoding* (MSD), see Fig. 4. The component codes C^i are successively decoded by the corresponding decoders D^i . At stage i , decoder D^i processes not only the block $\mathbf{y} = (y_1, \dots, y_N)$, $y_\gamma \in \mathbf{Y}$, of received signal points, but also decisions $\hat{\mathbf{x}}^j$, $j = 0, \dots, i-1$, of previous decoding stages j . (Notice that decoding delays are not shown in Fig. 4.) Actually, the staged decoding according to the chain rule in (2) would require the transmitted symbol \mathbf{x}^j instead of the estimate $\hat{\mathbf{x}}^j$. But as long as error-free decisions $\hat{\mathbf{x}}^j = \mathbf{x}^j$ are generated by the decoder D^j , MSD can be interpreted as an implementation of the chain rule.

III. CAPACITY OF MULTILEVEL CODING

A. Capacity

In order to approach channel capacity, a maximization of the mutual information over all selectable parameters has to be performed. Usually, these are the *a priori* probabilities of the signal points. Thus a specific probability distribution $\Pr\{a\}$ over the channel inputs is required to achieve the capacity C . These probabilities cannot be optimized independently for each individual level, but only for the entire signal set. Thus the capacity C^i of the equivalent channel i is given by the respective mutual information $I(Y; X^i | X^0 \dots X^{i-1})$ for these specific channel input probabilities. In view of (3) and (7), C^i is given by

$$\begin{aligned} C^i &= I(Y; X^i | X^0 \dots X^{i-1}) \\ &= \mathbb{E}_{x^0 \dots x^{i-1}} \{C(\mathbf{A}(x^0 \dots x^{i-1}))\} \\ &\quad - \mathbb{E}_{x^0 \dots x^i} \{C(\mathbf{A}(x^0 \dots x^i))\} \end{aligned} \quad (7)$$

where $C(\mathbf{A}(x^0 \dots x^i))$ denotes the capacity when using (only) the (sub)set $\mathbf{A}(x^0 \dots x^i)$ with *a priori* probabilities $\Pr\{a\} / \Pr\{\mathbf{A}(x^0 \dots x^i)\}$. Hence, in order to avoid confusion, we use the term “capacity” throughout this paper for given and fixed *a priori* probabilities of signal points.

Because the MLC approach directly results from the chain rule for mutual information (Section II), the derivation of the capacity of multilevel coded schemes is obvious. We review the theorem given in [34]–[37], cf. also [38] and [39].

Theorem 1: The capacity $C = C(\mathbf{A})$ of a 2^ℓ -ary digital modulation scheme under the constraint of given *a priori* probabilities $\Pr\{a\}$ of the signal points $a \in \mathbf{A}$ is equal to the sum of the capacities C^i of the equivalent channels i of a multilevel coding scheme

$$C = \sum_{i=0}^{\ell-1} C^i. \quad (8)$$

The capacity C can be approached via multilevel encoding and multistage decoding, if and only if the individual rates R^i are chosen to be equal to the capacities of the equivalent channels, $R^i = C^i$.

The proof is obvious from the chain rule and given in the papers cited above.

Theorem 1 has the following consequences for digital transmission schemes:

- 1) Out of the huge set of all possible codes with length N , where $NR = K$ binary symbols are mapped to N signal points, the (comparatively very small) subset of codes generated by the MLC approach—where $NR^i = K^i$ binary symbols are mapped to N signal point label elements x^i independently for each level i —is a selection with asymptotically optimum performance. As already mentioned by Forney [33] and Urbanke and Rimoldi [32] for the case of lattice codes, here Shannon’s coding theorem is proved with well-structured in contrast to random codes.
- 2) Although in multistage decoding (MSD) the code constraints at higher levels are not taken into account while decoding lower levels, suboptimum MSD suffices to achieve capacity. Optimum overall *maximum-likelihood decoding* (MLD) of the Euclidean-space code cannot improve the asymptotic performance of the scheme as long as the rates R^i are chosen equal to the capacities C^i . Even in practice with finite code lengths, the gain of MLD over MSD is expected to be relatively small as long as the rate design is appropriate.
- 3) The theorem states that for any digital transmission scheme (provided that the number of points is a power of two), the problem of channel coding can be solved in principle in an optimum way via MLC and MSD by employing *binary codes*. That means there is no need to search for good nonbinary codes to be used in bandwidth-efficient transmission systems. Starting from the huge field of good binary codes, their properties can be directly translated to any bandwidth efficiency via the MLC approach. Therefore, similar to the theoretical separability of source and channel coding, channel coding and modulation can be treated and optimized separately. (In practice, i.e., for finite data delay, nonbinary codes may, of course, have some advantages in performance or complexity, cf. TCM.)
- 4) The theorem implies no restriction on the particular labeling of signal points. Thus mapping by set partitioning according to Ungerboeck’s criterion [4] is not essential to approach capacity. Nevertheless, for finite code length, Ungerboeck’s partitioning strategy turns out to lead to the highest performance among MLC schemes with different partitioning strategies, see Section VI. However, alternative partitioning strategies may be favorable for some other practical purposes as discussed also in Section VI.

B. Capacity Region

In this section we regard MLC as a multiuser communication scheme as in [39], and show that a much larger variety of rate combinations can approach capacity, provided that overall MLD is used rather than MSD. Therefore, this theoretical result may not be of much practical importance, although it does help to elucidate the strengths and weaknesses

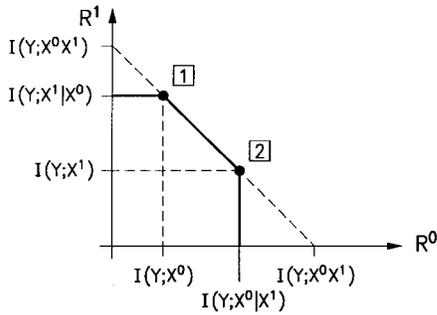


Fig. 5. Bounding region for the individual rates of a two-level code.

of a broader class of MLC designs than those suggested by Theorem 1 (cf. also [45]).

Applying known results from multiuser information theory we can state the following theorem.

Theorem 2: The capacity $C = C(\mathbf{A})$ of a 2^ℓ -ary digital modulation scheme with given *a priori* probabilities $\Pr\{a\}$ of signal points $a \in \mathbf{A}$ can be achieved by multilevel encoding and overall maximum-likelihood decoding if and only if the rates R^i satisfy the following conditions:

- 1) $\sum_{i=0}^{\ell-1} R^i = I(Y; X^0 \dots X^{\ell-1}) = C;$
- 2) $\sum_{i \in \mathcal{S}} R^i \leq I(Y; \{X^i | i \in \mathcal{S}\} | \{X^j | j \in \bar{\mathcal{S}}\})$, for all possible sets $\mathcal{S} \subset \{0, \dots, \ell-1\}$ of indices, where $\bar{\mathcal{S}}$ is the complementary set of \mathcal{S} such that $\mathcal{S} \cup \bar{\mathcal{S}} = \{0, \dots, \ell-1\}$ and $\mathcal{S} \cap \bar{\mathcal{S}} = \{\}$.

Proof: For simplicity, let us start with a two-level code, i.e., 4-ary signaling. The symbols x^0 and x^1 of two single independent users are combined via the mapping of symbols to signal points and transmitted over the channel. Therefore, we actually have to deal with a multiple-access channel where the maximum feasible sum of rates R^i is bounded by the mutual information of the total scheme, see, e.g., [46]

$$R^0 + R^1 \leq I(Y; X^1 X^0). \quad (9)$$

Since overall MLD takes into account the constraints of code \mathcal{C}^1 at level 1 for decoding of symbol x^0 at level 0, the maximum rate R^0 at level 0 is given by the mutual information $I(Y; X^0 | X^1)$ [46], [39]. This argument is also valid for decoding symbol x^1 at level 1, thus

$$R^0 \leq I(Y; X^0 | X^1) \quad \text{and} \quad R^1 \leq I(Y; X^1 | X^0). \quad (10)$$

The bounding regions for R^0 and R^1 according to (10) and (9) are shown in Fig. 5. Regarding point $\square 1$, $R^0 = I(Y; X^0)$ and $R^1 = I(Y; X^1 | X^0)$ hold, which corresponds to the chain rule: $I(Y; X^0 X^1) = I(Y; X^0) + I(Y; X^1 | X^0)$. Hence, symbol x^0 can be decoded as in the single-user case without any knowledge of the actually transmitted symbol x^1 , and point $\square 1$ marks the special result described above that MSD suffices to achieve capacity. The analogous situation is given for point $\square 2$ by interchanging x^0 and x^1 . Here, MSD starts with the decoding of symbol x^1 . To be more specific, points $\square 1$ and $\square 2$ differ only in the labeling of the signal points.

For example, if point $\square 1$ corresponds to the labeling of a 4-ASK constellation according to Ungerboeck, then point $\square 2$ represents the situation for the labeling defined by a *block partitioning*, see Section VI. Notice that the individual rates depend strongly on the particular labeling; i.e., $I(Y; X^0)$ is not equal to $I(Y; X^1)$ in the general case (also illustrated in Fig. 5).

Following Gallager [40], total randomness is not indispensable to prove the channel coding theorem. A careful analysis shows that pairwise independence of codewords is sufficient. It is easy to see that the pairwise independence of codewords in the ensemble of multilevel codes is valid. Hence, the channel coding theorem applies to this ensemble if overall maximum-likelihood decoding is used. This is still valid even if the rates R^i are not chosen to be equal to the capacities C^i of the equivalent channels i but the total rate is less than the capacity of the channel. Thus capacity in the range $I(Y; X^0) < R^0 < I(Y; X^0 | X^1)$ can still be achieved with MLC iff $R = R^0 + R^1 = I(Y; X^1 X^0)$. However, in this case it is *unavoidable*² to replace the low-complexity MSD by an extremely complex overall maximum-likelihood decoding in order to come close to capacity. For rate R^0 exceeding $I(Y; X^0 | X^1)$ or for rate R^1 exceeding $I(Y; X^1 | X^0)$, respectively, the pair of rates (R^0, R^1) is outside the capacity region.

In summary, if (R^0, R^1) lies on the straight line connecting points $\square 1$ and $\square 2$ in Fig. 5, the capacity of the modulation scheme can be achieved by MLC and MLD. In the special case where (R^0, R^1) lies on one of the vertices of the capacity region (point $\square 1$ or $\square 2$), MSD is sufficient to achieve capacity.

Using the same arguments, the results for the two-level code can be extended to the case of an ℓ -level code. Here, multiuser theory gives not only upper bounds for the individual rates R^i and the sum of all rates $\sum_{i=0}^{\ell-1} R^i$, but also for the sums of two or more rates [41]. This proves the theorem. \square

The consequences of Theorem 2 are discussed in the next section.

IV. COMPARISON OF RATE DESIGN RULES

The essential point for the design of a coded modulation scheme is the assignment of code rates to the individual coding levels. In this section, five rules for rate design are reviewed, and their similarities and differences are discussed.

A. Capacity Design Rule

Following Theorem 1, the first design rule is quite obvious:

Rate Design Rule 1 (Capacity Rule)

For a 2^ℓ -ary digital modulation scheme the rate R^i at the individual coding level i of a multilevel coding scheme should be chosen equal to the capacity C^i of the equivalent channel i , $i = 0, \dots, \ell - 1$

$$R^i = C^i.$$

²It is true that in the entire region MSD in combination with time sharing is still sufficient to achieve capacity. Since this implies switching periodically between code rates given by $\square 1$ and $\square 2$, this strategy is not of interest in practice for MLC.

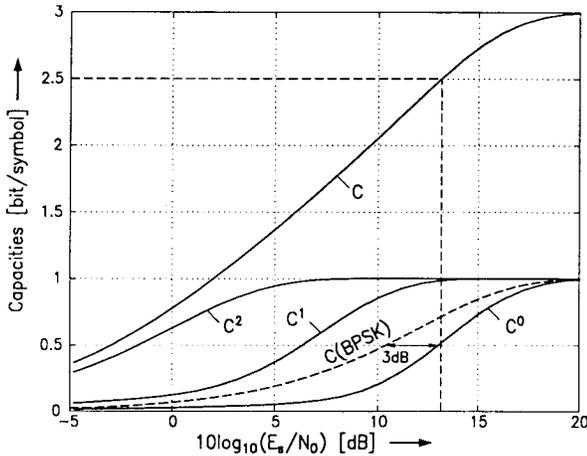


Fig. 6. Capacity $C(\mathbf{A})$ and capacities C^0 , C^1 , and C^2 of the equivalent channels for 8-ASK with natural labeling. AWGN channel. Dashed vertical line: $C(\mathbf{A}) = 2.5$ bits/symbol. Dashed curve: Capacity $C(\text{BPSK})$ for a BPSK scheme (same spacing of signal points relative to noise variance).

This capacity rule was proposed in [34], [35], [33] and independently in [38], and [39].

For example, we assume an 8-ASK constellation $\mathbf{A} = \{\pm 1, \pm 3, \pm 5, \pm 7\}$ with natural labeling, cf. Fig. 1. In Fig. 6, the capacity $C(\mathbf{A})$ of the scheme as well as the capacities C^i , $i = 0, \dots, \ell - 1$, of the equivalent channel i for transmission over the AWGN channel are shown. In order to get a unified representation, the abscissa is labeled by E_s/N_0 , where E_s denotes the average symbol energy. For the design of a scheme with rate 2.5, see dashed line in Fig. 6, the capacities C^i coinciding with the optimum individual rates, are given by $C^0/C^1/C^2 = 0.52/0.98/1$. Additionally, for comparison, the capacity $C(\text{BPSK})$ for binary signaling corresponding to the situation at level 0 without multiple representation of symbol x^0 is sketched in Fig. 6. Here the same spacing of signal points relative to noise variance as in 8-ASK is assumed (i.e., $E_{s,\text{BPSK}} = E_{s,8\text{-ASK}}/21$). The gap between $C(\text{BPSK})$ and C^0 (e.g., 3 dB at $C = 0.5$) is due to the multiple representation of binary symbols by signal points at level 0 (cf. also [35]).

B. Balanced Distances Rule

The second design rule, which was used traditionally, is solely based on the minimum Euclidean distance in signal space. Let us denote the minimum Hamming distance of code C^i at coding level i by δ_i and the minimum Euclidean distance of signal points in subsets $\mathbf{A}(x^0 \dots x^{i-1})$ at partitioning level i by d_i . Then, the squared minimum Euclidean distance of multilevel codewords is lower-bounded by

$$d^2 \geq \min_{i=0, \dots, \ell-1} \{d_i^2 \delta_i\} \quad (11)$$

see, e.g., [47], [26], and [48]. In order to maximize the minimum Euclidean distance at lowest cost in total rate, it is reasonable to choose the product $d_i^2 \delta_i$ equal for all levels i , i.e., the product of distances should be balanced. This leads to

Rate Design Rule 2 (Balanced Distances Rule)

For a 2^ℓ -ary digital modulation scheme, the rate R^i at the individual coding level i of a multilevel coding scheme should

be chosen such that the following conditions are satisfied:

$$d_i^2 \delta_i = \text{const.}, \quad i = 0, \dots, \ell - 1. \quad (12)$$

The intra-subset distances d_i are given by the signal constellation and the particular labeling strategy. Because the choice of the minimum Hamming distance δ_i of the component codes C^i balances the levels, we refer to Rate Design Rule 2 as the balanced distances rule (BDR).

The balanced distances rule has been used by most previous MLC work starting with the lattice constructions B through E presented in [24].

Consider again the 8-ASK constellation sketched in Fig. 1. The intra-subset minimum Euclidean distances d_i for this example are $d_0/d_1/d_2 = 2/4/8$. According to the balanced distances rule, the normalized minimum Hamming distances of the component codes with length N must satisfy

$$2^2 \delta_0/N = 4^2 \delta_1/N = 8^2 \delta_2/N. \quad (13)$$

As component codes, we use linear binary block codes with minimum Hamming distances which meet the well-known Gilbert–Varshamov bound with equality, see, e.g., [49]. Then, for long component codes C^i , the rate R^i and minimum Hamming distance δ_i are related by

$$R^i = 1 - H_2(\delta_i/N) \quad (14)$$

where

$$H_2(x) = -x \log_2 x - (1-x) \log_2 (1-x)$$

denotes the binary entropy function. Using again the example of a design for total rate $R = 2.5$ bits/symbol, the rate distribution according to the balanced distances rule yields

$$R^0/R^1/R^2 = 0.66/0.88/0.96, \quad \text{with } d_i^2 \delta_i/N = 0.26. \quad (15)$$

Since the multiple representation of symbols by signal points is ignored, the rate $R^0 = 0.66$ according to the BDR is substantially higher than the capacity $C^0 = 0.52$ of the equivalent channel at level 0. According to the converse of Shannon’s channel coding theorem, transmission over this channel is not possible with arbitrarily high reliability. Nevertheless, assuming MSD and comparing the required SNR for $C^0 = 0.52$ and 0.66 , respectively, we can observe a loss of about 1.2 dB for an MLC scheme designed by the BDR compared to an MLC scheme designed according to the capacity rule. This degradation due to the increased rate R^0 results mainly from a tremendous increase in the number of nearest neighbor error events because of the multiple representation of symbols by signal points, cf. [35] and the examples in the Appendix. To illustrate this crucial effect, let us consider the following example assuming one-dimensional constellations: Regard two codewords at level 0, say \mathbf{x}_0^0 and \mathbf{x}_1^0 , with Hamming distance δ_0 . If each symbol of \mathbf{x}_0^0 is only represented by “inner” signal points, then two nearest neighbor points represent the complementary binary symbol. Hence, there are 2^{δ_0} words in the Euclidean space representing \mathbf{x}_1^0 with minimum squared Euclidean distance $d_0^2 \delta_0$ to codeword \mathbf{x}_0^0 . For Ungerboeck’s labeling, the minimum Hamming distance

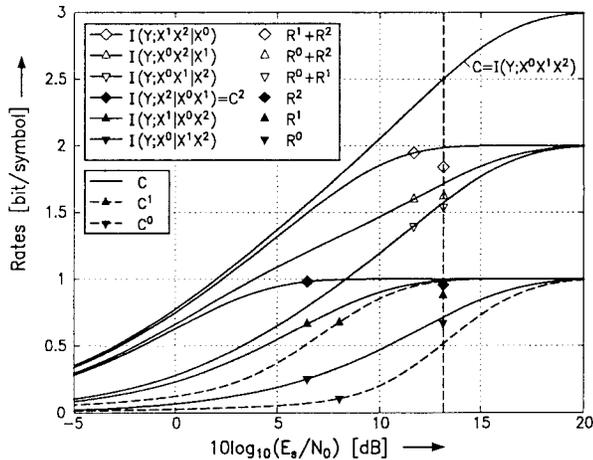


Fig. 7. Mutual informations according to (16) and capacities C^0 , C^1 , and C^2 of the equivalent channels for a multilevel coded 8-ASK scheme. Ungerboeck's set partitioning. AWGN channel. Rates R^0 , R^1 , and R^2 are chosen by the balanced distances rule for total rate $R = 2.5$ bits/symbol.

at level 0 is the greatest one and hence, the greatest degradation due to multiple symbol representation occurs at level 0.

Let us check whether the rate distribution (11) is optimum in sense of capacity. As in Theorem 2, the individual code rates R^i must satisfy the following conditions:

$$\begin{aligned}
 R^0 &\leq I(Y; X^0 | X^1 X^2) \\
 R^1 &\leq I(Y; X^1 | X^0 X^2) \\
 R^2 &\leq I(Y; X^2 | X^0 X^1) \\
 R^0 + R^1 &\leq I(Y; X^0 X^1 | X^2) \\
 R^0 + R^2 &\leq I(Y; X^0 X^2 | X^1) \\
 R^1 + R^2 &\leq I(Y; X^1 X^2 | X^0) \\
 R = R^0 + R^1 + R^2 &\leq I(Y; X^0 X^1 X^2). \quad (16)
 \end{aligned}$$

In Fig. 7 the curves of the mutual informations appearing in (16) are plotted together with the capacities C^i of the equivalent channels. The dashed vertical line marks the situation for $C(\mathbf{A}) = 2.5$ bits/symbol. We observe that the rate distribution according to the balanced distances rule, see (15), satisfies Theorem 2, because all marked points on the vertical dashed line lie below the corresponding curves, marked by the same symbol. But error-free multistage decoding of the MLC scheme designed by the BDR is impossible in principle, since the rate R^0 exceeds the capacity C^0 . Only by employing overall MLD can the capacity of the 8-ASK constellation be approached.

These differences between MLC schemes designed by the capacity rule (Cap-MLC) and MLC schemes designed by the balanced distances rule (BDR-MLC) can be similarly observed for other schemes [50].

Nonetheless, in the majority of the work concerning coded modulation, see (e.g., [48], [26], [27], [51], and [52]), the balanced distances rule was used for rate design although multistage decoding was applied. These papers gave schemes with excellent asymptotic coding gains, because d^2 was maximized. In practice, however, the real coding gains are far below the asymptotic gains. At least at the lowest level, the rate R^0

significantly exceeds the capacity of the equivalent channel 0, and, therefore, many errors occur at decoder D^0 . These errors propagate through higher levels in the MSD procedure, so the performance of the scheme is quite poor.

In order to overcome this effect, several strategies have been proposed [53], [54]. First, reliability information about decoding decisions may be passed to higher decoding stages in order to mitigate the effect of error propagation, cf. also [26] and [25]. But this method requires the use of sophisticated decoders generating reliability information for each output symbol. Moreover, especially for concatenated codes which are iteratively decoded such as turbo codes [43], [44], it is not yet possible to generate reliability information of good quality.³

Second, interleaving between binary codewords at the individual coding levels has been proposed. Then, error bursts at decoding stage i can be broken up into almost independent single errors, facilitating the task for the decoders D^j , $j > i$, at higher levels. But in fact, interleaving increases the effective code length. Hence, using codes whose performance increases with the block length, e.g., turbo codes, it seems more efficient to increase the code length in a direct way instead of introducing interleaving. Nevertheless, interleaving within one codeword can still improve performance without introducing additional delay when employing codes which are sensitive to the structure of the error event. For example, this holds for turbo codes or terminated convolutional codes.

Third, instead of performing multistage decoding once per multilevel codeword, it has been proposed to iterate the decoding procedure, now using decoding results of higher levels at lower levels [25]. Consider, e.g., a second decoding of symbol x^0 by the decoder D^0 , where estimates \hat{x}^i of all symbols x^i , $i \neq 0$, are already known. Now, symbol x^0 is no longer multiply represented in the signal set \mathbf{A} —only the signal points

$$a^{(0)} = \mathcal{M}(x^0 = 0, \hat{x}^1 \dots \hat{x}^{\ell-1})$$

and

$$a^{(1)} = \mathcal{M}(x^0 = 1, \hat{x}^1 \dots \hat{x}^{\ell-1})$$

remain and thus the effective error coefficient is decreased. If the decisions at other levels are error-free, then simple binary signaling remains. (Fortunately, this is true for a practical decoder, since the number of errors usually decreases during the iteration proceeds [37].) In this way, the enormous increase in the number of error events by multiple representation of binary symbols is avoided and performance is greatly improved. Obviously, such an iterative decoding strategy over several levels only works if there is sufficient interleaving between levels. Thus this method not only causes a multiple decoding effort, but also an increased data delay, which is not usefully exploited in the sense of information theory.

These methods proposed to improve the suboptimum MSD procedure are indeed more or less good attempts to approximate an overall maximum-likelihood decoder. The complexity

³Due to the remaining statistical dependencies in the feedback loop, the reliability information of iteratively decoded symbols tends to a hard-decision value with an increasing number of iterations.

and the decoding delay due to this approximate MLD are substantially larger than for usual MSD. If the rates at the individual coding levels of an MLC scheme are designed according to the BDR, the use of MLD is indispensable to avoid a significant performance loss. However, as shown in Section III, there is no need to use an extremely complex overall MLD if the individual code rates are chosen according to the capacity rule. Therefore, we conclude that the reason for performance deficiencies of BDR-MLC schemes is actually not the suboptimality of MSD, but the rate design itself.

Strictly speaking, the optimality of MSD holds only in the asymptotic case of capacity-approaching component codes. But in practice the gain of approximate MLD over MSD is expected to be very small if the rate design is appropriate.

Let us now consider the special case of trellis-coded modulation (TCM) in this context. TCM is actually a two-level MLC scheme in which the second level is “uncoded”—or, more precisely, coded only per symbol. TCM schemes are always designed for maximum minimum Euclidean distance in signal space. Hence, TCM schemes are actually a special case of BDR-MLC schemes. The traditional decoding method for TCM involves first doing ML decoding at the **higher** levels for all possible lower level symbol values (finding the closest point in each subset), and then using these decisions to do ML decoding at the lower levels. The combination is an overall ML decoder. In contrast, to apply MSD principles, one should first decode the **lower** levels, e.g., using lattice decoding for coset codes, and then, given the lower level decoder outputs, decode once per symbol at the higher level. The results in this paper and in [33] show that the latter approach will also work provided that the lower level rate is below capacity, although it will clearly not be as good as full MLD.

C. Coding Exponent Rule

Now, we sketch a rate design rule suited for practical applications in which the data delay and the codeword length N are restricted, see also [37]. Additionally, a certain error rate can often be tolerated depending on the particular application. Therefore, we subsequently employ the well-known random coding bound [40], which provides a relation between codeword length N and word error rate p_w , for design and discussion of MLC schemes.

Although some of the results are illustrated only for the particular example of 8-ASK, they are in principle valid for all pulse amplitude modulation (PAM) transmission schemes based on sets of real or complex signal points, i.e., ASK, PSK, QAM.

From [40], the random coding exponent $E_r(R)$ is⁴

$$E_r(R) := \max_{0 \leq \rho \leq 1} \{E_0(\rho) - \rho R\} \quad (17)$$

where

$$E_0(\rho) := -\log_2 \left\{ \int_{\mathbf{Y}} \left[\sum_{x \in \mathbf{X}} \Pr\{x\} (f_Y(y|x))^{\frac{1}{1+\rho}} \right]^{1+\rho} dy \right\}.$$

⁴Note that the symbol E followed by curly brackets denotes the expected value operation, whereas the symbol E followed by parentheses denotes a coding exponent.

The channel input symbol and input alphabet are denoted by x and \mathbf{X} , respectively. Using the random coding exponent, the word error rate is bounded by $p_w \leq 2^{-N E_r(R)}$.

Now, let us consider the i th level of an MLC scheme. As shown in Section II, the underlying subset $\mathbf{A}(x^0 \dots x^{i-1})$ for transmission over this equivalent channel i varies. Hence, the parameter $E_0^i(\rho)$ of the equivalent channel i is given by averaging the parameter $E_0^i(\rho, x^0 \dots x^{i-1})$ over all combinations of $x^0 \dots x^{i-1}$, cf. [40]

$$E_0^i(\rho) = E_{x^0 \dots x^{i-1}} \{E_0^i(\rho, x^0 \dots x^{i-1})\} \quad (18)$$

where

$$E_0^i(\rho, x^0 \dots x^{i-1}) = -\log_2 \left\{ \int_{\mathbf{Y}} \left[\sum_{x^i=0}^1 \Pr\{x^i\} (f_Y(y|x^i, x^0 \dots x^{i-1}))^{\frac{1}{1+\rho}} \right]^{1+\rho} dy \right\}. \quad (19)$$

Usually (19) has to be evaluated numerically. Clearly, the random coding exponent $E_r^i(R^i)$ of the equivalent channel i is

$$E_r^i(R^i) := \max_{0 \leq \rho \leq 1} \{E_0^i(\rho) - \rho R^i\}. \quad (20)$$

Fixing the random coding exponent and hence the error rate to a constant value, a tradeoff between rate and SNR is possible. We obtain the so-called isoquants of the random coding exponent

$$E_r^i(R^i) = -\frac{1}{N} \log_2 p_w = \text{const.} \quad (21)$$

For high SNR's the bound can be improved employing the expurgated coding exponent $E_{\text{ex}}^i(R^i)$, see [40]. The coding exponent $E^i(R^i)$ for the subsequent analysis is simply the maximum of the random and the expurgated coding exponent.

For illustration, the isoquants $R^i(E_s/N_0)$ of the coding exponent $E^i(R^i)$ for all levels i of 8-ASK with natural labeling (cf. Fig. 1) and the AWGN channel are plotted versus E_s/N_0 in Fig. 8. Additionally, the sum of individual rates $R = \sum_{i=0}^{\ell-1} R^i$ is shown. The presented isoquants are used to compare the power efficiency of different MLC schemes as well as to assign rates to the individual codes for a given length N . Similar to Design Rule 1 (capacity rule), we propose the following design rule for MLC schemes with given block length N :

Rate Design Rule 3 (Coding Exponent Rule)

Consider a 2^ℓ -ary digital modulation scheme combined with a multilevel coding scheme applying binary component codes of length N . For a maximum tolerable word error rate p_w , the rates R^i , $i = 0, \dots, \ell - 1$, at the individual coding levels i should be chosen according to the corresponding isoquants of the coding exponents $E^i(R^i)$.

Continuing the example of 8-ASK with a total rate $R = 2.5$ (see the dashed line), we obtain from Fig. 8 the rate design

$$R^0/R^1/R^2 = 0.531/0.97/0.999 \quad (22)$$

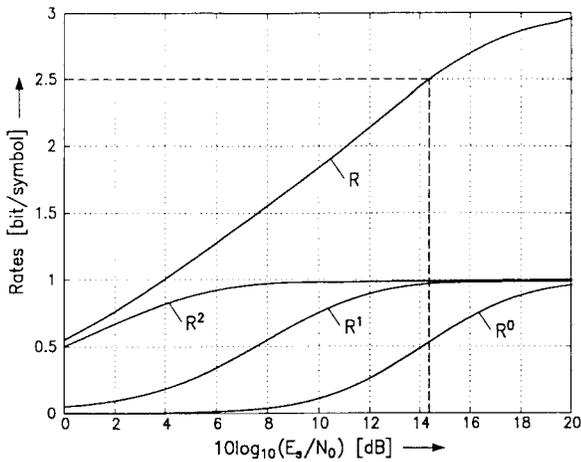


Fig. 8. 8-ASK, Ungerboeck's set partitioning. Rates R^i , $i = 0, 1, 2$, and total rate $R = \sum_{i=0}^2 R^i$ from isoquants of $E^i(R^i)$ for $N = 500$ and $p_w = 10^{-3}$ versus E_s/N_0 . Dashed line: $R = 2.5$ bits/symbol. AWGN channel.

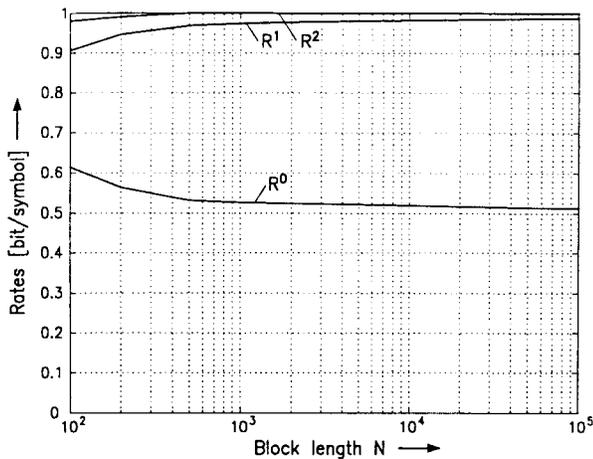


Fig. 9. 8-ASK with natural labeling. AWGN channel. Distribution of rates R^i , $i = 0, 1, 2$, derived from isoquants of $E^i(R^i)$ versus block length N .

for codeword length $N = 500$ and word error rate $p_w = 10^{-3}$. (The particular values for N and p_w are chosen because experience shows that with $N = 500$ and $p_w = 10^{-3}$ a bit-error rate $< 10^{-5}$ is achievable.)

For a given total rate R and desired reliability p_w , the isoquant curves serve to determine the distribution of individual rates R^i dependent on the block length N . It is worth noting that if the error probability of the equivalent channels is not low enough to neglect error propagation, then one may choose different error rates for different levels to compensate for this effect; i.e., lower levels should operate at lower (designed) error rates than higher levels.

For the particular case of $p_w = 10^{-3}$ and total rate $R = 2.5$, the rate distribution is plotted versus the block length N for the 8-ASK example in Fig. 9. This representation gives a nice illustration of the variation of the rate distribution versus the codeword length N . The rates R^i of the component codes of an MLC scheme derived from the capacity or the coding exponent rule are quite different from those derived from the balanced distances rule, as long as the block length N is

not very small. In particular, the rates at lower levels are substantially decreased due to the multiple representation of binary symbols by signal points.

For moderate to high block lengths ($N \gtrsim 1000$) and word error rate $p_w = 10^{-3}$, the individual rates according to the coding exponent rule differ only slightly from the asymptotic values for $N \rightarrow \infty$ given by the capacity rule. Therefore, the capacity rule is a good choice for practical codes. However, for short codes, the rates derived from the coding exponent rule tend to those derived by the balanced distances rule, because in this region error probabilities are mainly determined by the minimum Euclidean distance rather than by the effective error coefficient. Thus rate design according to the coding exponent rule connects the capacity rule and the balanced distances rule in a certain way.

D. Cutoff Rate Rule

In order to complete the discussion of design rules based on information theoretical parameters, a rate design rule employing cutoff rates of the equivalent channels is stated here, cf. [55]. The cutoff rate is an appropriate parameter for convolutional codes using sequential decoding [2]. More generally, the cutoff rate criterion may be useful for those classes of codes that are not capacity approaching but rather R_0 approaching.

The cutoff rate R_0^i for equivalent channel i is simply [40]

$$R_0^i = E_0^i(\rho = 1). \quad (23)$$

Thus we arrive at the following design rule for MLC schemes when R_0 is the significant parameter.

Rate Design Rule 4 (Cutoff Rate Rule)

For a 2^ℓ -ary digital modulation scheme, the rates R^i , $i = 0, \dots, \ell - 1$, at the individual coding levels i of a multilevel coding scheme using R_0 -approaching codes and decoding schemes should be chosen equal to the cutoff rates R_0^i of the equivalent channels:

$$R^i = R_0^i.$$

The rates derived from the cutoff rate rule are very similar to those derived from the capacity rule. Therefore, we do not present an example for this rate design rule.

It is noteworthy that there is no simple relation between the sum of the individual cutoff rates R_0^i and the cutoff rate of the underlying modulation scheme as there is for the respective capacities (Theorem 1). Interestingly, the results of [55] show that the sum of the R_0^i in an MLC scheme using Ungerboeck's partitioning can exceed the cutoff rate of the modulation scheme.

E. Equal Error Probability Rule

The design rules given so far are mainly based on random coding principles rather than on specific component codes. An alternative way is to choose codes with known properties in such a way that the word or bit-error probabilities of the equivalent channels or their bounds are equal. This leads to an *equal error probability rule*. For this purpose,

TABLE I
RATE DISTRIBUTION FOR DIFFERENT RATE DESIGNS. 8-ASK
CONSTELLATION, UNGERBOECK LABELING, $R = 2.5$ BITS/SYMBOL

Rate design according to	R^0	R^1	R^2
Capacity rule	0.52	0.98	1.0
Balanced distances rule	0.66	0.88	0.96
Coding exponent rule ($N = 500$)	0.53	0.97	1.0
Cut-off rate rule	0.51	0.99	1.0
Equal error probability rule	0.51	0.99	1.0

an analytic expression for the error probability is required. In the Appendix, calculations of the distance enumerators of the individual levels are presented which allow estimating the error probability using a union bound approach.

To summarize, Table I displays the different rate designs proposed in this section. Again the 8-ASK constellation with Ungerboeck labeling and a total rate $R = 2.5$ bits/symbol are assumed. We see that all rules give similar results except the balanced distances rule.

F. Examples

In order to confirm the relevance of the presented design rules in practice, simulations for several digital PAM transmission schemes with MLC and MSD over the AWGN channel were performed. In particular, 8-PSK with $R = 2$ bits/symbol, 16-QAM with $R = 3$ bits/symbol, 32-QAM with $R = 4$ bits/symbol, and 64-QAM with $R = 5$ bits/symbol (all with equiprobable signal points) were investigated. Turbo codes (TC) using 16-state convolutional codes [44], [56] with rates derived from the coding exponent rule are used as component codes. Flexible rates of turbo codes are achieved via puncturing as in [57], [37]. Block lengths $N = 2000$ and $N = 20000$ of turbo codes with interleaver lengths $K_i = R^i \cdot N$ (number of information symbols fed to level i) are used.

The results are presented in the power–bandwidth plane for the band-limited AWGN channel, see Fig. 10. The bandwidth efficiency of the digital communication schemes, measured in bits/s/Hz, is plotted versus the required E_b/N_0 to achieve a desired reliability. As usual, E_b denotes the energy per bit at the receiver input and N_0 the one-sided spectral noise power density. The solid line marks the Shannon limit for reliable digital transmission. The squares mark the capacity limits for these PAM schemes with equiprobable signal points. Note that in order to overcome the gap to the Shannon limit, shaping methods as described in Section VIII are indispensable.

Each sketched triangle marks the required E_b/N_0 to achieve a bit-error rate (BER) of 10^{-5} by the corresponding transmission scheme. Of course, the results for these QAM schemes can be extended to $M > 64$ -QAM schemes by imposing further uncoded levels. Additionally, results for the “original” turbo codes with the above given block lengths transmitted with QPSK and $R = 1$ as well as results for several uncoded schemes are plotted for reference. The results show that the gap between the Shannon limit and the power efficiency of the investigated transmission schemes remains quite constant, nearly independent of the bandwidth efficiency of the scheme.

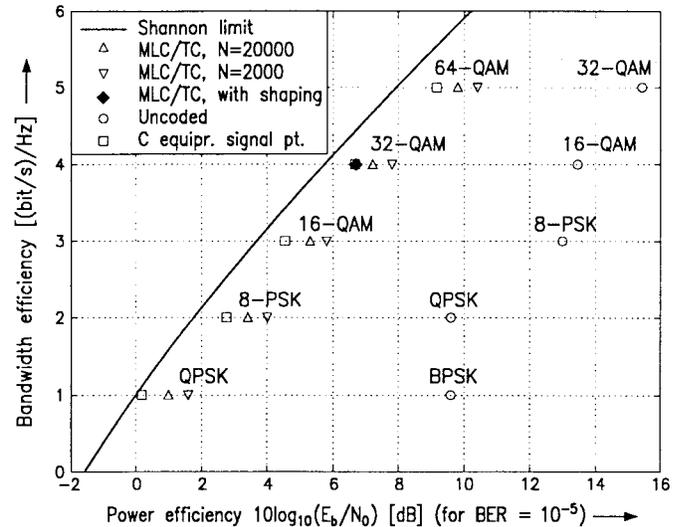


Fig. 10. Power–bandwidth plane for the band-limited AWGN channel. (MLC/TC: Multilevel coded PAM scheme employing turbo codes with block length N as components, MSD.)

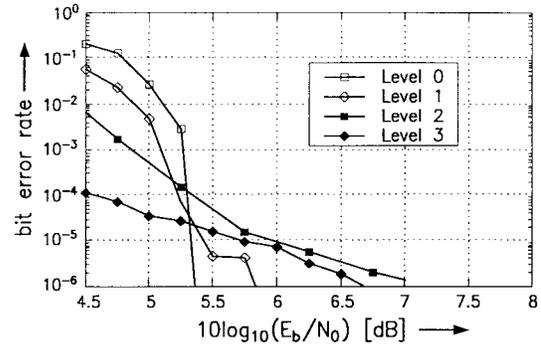


Fig. 11. Bit-error rate for the individual levels of a multilevel code for 16-QAM using turbo codes as component codes, block length $N = 2000$, total rate $R = 3$, individual rates: $R^0/R^1/R^2/R^3 = 0.29/0.75/0.96/1.0$, Ungerboeck’s set partitioning, no error propagation in multistage decoding.

For brevity, we have included an additional simulation result which relates to Section VIII. The diamond in Fig. 10 shows a result for a multilevel coded 64-QAM constellation using signal shaping.

Finally, in order to check the proposed design of the MLC scheme, measurements of the BER for the individual levels i , $i = 0, \dots, \ell - 1$, without any error propagation in MSD were performed. Instead of using the symbols \hat{x}^j of lower levels j , $j = 0, \dots, i - 1$, that were estimated by previous decoding stages, the correct transmitted symbols x^j were fed to higher stages so that the correct subset is always addressed. The results in Fig. 11 show the desired error behavior of the component codes with a “crossover area” of the different curves in the interesting range (BER = 10^{-5}). Note that the abscissa parameter E_b/N_0 refers to the entire scheme. The slopes of the curves stem from the very different code rates and increase as the code rate decreases. Additionally, as one can see from Fig. 11 the SNR required to achieve a bit-error rate of 10^{-5} increases from the lowest to the highest level. Hence, in the case of this particular example, error propagation would only marginally degrade the total performance of the scheme.

The reason is that the code rate increases significantly from the lowest to the highest coded levels, as is always the case when using a labeling according to Ungerboeck's set partitioning. Thus Ungerboeck's set partitioning is essential to reduce the effect of error propagation in practical schemes.

Our main result is that by using binary codes with the capacity design rule for the rates at each level, the unshaped Shannon limit can be approached within tenths of a decibel independently of spectral efficiency in the bandwidth-limited regime, with no more decoding complexity than is needed for comparable capacity-approaching codes on ordinary binary channels. Moreover, the complexity per information bit is decreased because of further uncoded levels.

In order to conclude this section the similarities and differences between MLC and TCM are briefly discussed:

- Turbo TCM schemes perform similarly over the AWGN channel as the presented MLC schemes with turbo component codes at comparable complexity and coding delay [58], [59]. The main difference is that in MLC binary turbo codes are the building blocks of the entire scheme, whereas in turbo TCM the component TCM schemes are building blocks of the turbo scheme.
- The usual TCM schemes and MLC with convolutional component codes perform similarly, but with a clear advantage to MLC at the same entire number of trellis branches per information bit, especially for high numbers of states.
- The main difference between MLC and TCM is not performance, but, first, how to design the coding scheme, and second, the achievable code rates. While for TCM the best codes for a particular modulation scheme are found by an exhaustive computer search, the best binary codes may be applied as component codes in an MLC scheme in conjunction with the proposed rate design. Because for TCM the code rate is strongly related to the dimensionality of the signal set, only integer code rates with respect to the considered dimensionality can be achieved. In contrast, the choice of the total rate in an MLC scheme is arbitrary; however, component codes with "strange" rates usually have to be implemented, e.g., by puncturing. In summary, there is much more flexibility in MLC design.

V. DIMENSIONALITY OF THE CONSTITUENT SIGNAL CONSTELLATION

A TCM scheme usually generates one bit of redundancy per D -dimensional constituent signal set. Thus the redundancy per dimension is $1/D$ bits/dimension. In contrast, for MLC schemes, the dimensionality D and the rate per dimension R/D can be chosen independently. Thus an important question when employing MLC schemes is the optimum dimensionality of the constituent signal constellation, cf. also [60]. In this section, we discuss the differences between multilevel codes based on one- or two-dimensional constellations. As we aim at approaching capacity, we neglect the necessity for higher dimensional constellations which may be present in some practical applications. Moreover, we have to distinguish

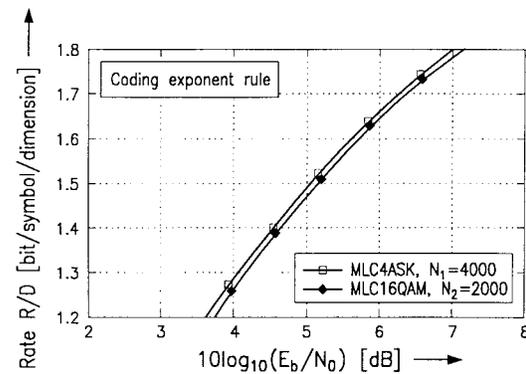


Fig. 12. Sum R/D of individual rates R^i/D for multilevel coded 4-ASK (MLC4ASK) with code length $N_1 = 4000$ and for multilevel coded 16-QAM (MLC16QAM) with $N_2 = 2000$ derived via isoquants of corresponding coding exponents. AWGN channel.

between MLC employing block codes and convolutional codes as component codes.

A. Block Component Codes

Consider an MLC scheme based on a D -dimensional signal set \mathcal{A}_D , which is the D -fold Cartesian product of a one-dimensional constellation \mathcal{A}_1 . The set partitioning is performed in ℓ_D steps in the D -dimensional signal space. To minimize complexity, it is preferable to base the MLC scheme on a one-dimensional signal set, because the least number (ℓ_1) of individual encoders and decoders are necessary compared to a D -dimensional approach with $\ell_D = D \cdot \ell_1$. For a fair comparison of MLC schemes using block codes as component codes, we fix the dimensionality of the entire multilevel codeword to, say, $N_D D$, where N_D denotes the length of the binary component codes of an MLC scheme based on a D -dimensional signal constellation. Thus all schemes can be compared based on the same delay per codeword.

As an example, we look at the power efficiency of MLC for 16-QAM, first based on the one-dimensional 4-ASK signal set per dimension, and second based on the two-dimensional 16-ary signal set. Because capacities are equal for both approaches, the coding exponents of the equivalent channels are applied (cf. Section IV-C) to assess the power efficiency of an MLC scheme with block codes of fixed length N_D . The sum of rates $\sum_{i=0}^1 R^i = R$ for MLC with 4-ASK derived via the isoquants of the coding exponent $E^i(R^i)$ is compared to the sum of rates $\sum_{i=0}^3 R^i = R$ for MLC with 16-QAM. We fix the code length to 2000 QAM symbols, resulting in $N_1 = 4000$ and $N_2 = 2000$. Fig. 12 shows the code rate R/D per dimension versus SNR for both approaches. An MLC scheme based on the one-dimensional ASK constellation with component codes of length $N_1 = 4000$ promises better performance than MLC based on the two-dimensional QAM constellation with codes of length $N_2 = 2000$.

In order to verify this result, simulations were performed for 4-ASK and 16-QAM schemes with MLC of code lengths $N_1 = 4000$ and $N_2 = 2000$, respectively, and transmission over the AWGN channel. Again, turbo codes using 16-state convolutional codes are employed as component codes. The normalized total code rate per dimension is fixed to $R/D =$

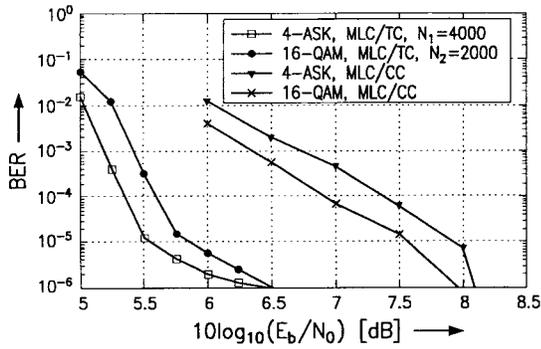


Fig. 13. Simulation results for 4-ASK and 16-QAM schemes with MLC of total rate $R/D = 1.5$ bits/dimension, transmission over the AWGN channel and MSD. Individual rates for 4-ASK: $R^0/R^1 = 0.52/0.98$. Individual rates for 16-QAM: $R^0/R^1/R^2/R^3 = 0.29/0.75/0.96/1$. MLC/TC denotes an MLC scheme using turbo codes (using 16-state convolutional codes) as components, MLC/CC denotes an MLC scheme using (64-state) convolutional codes as components.

1.5 bits/dimension. The individual code rates are assigned according to the coding exponent rule. The two leftmost curves, depicted in Fig. 13, represent the results for both schemes. Indeed, it can be observed that multilevel coded transmission based on the one-dimensional constellation exhibits a power efficiency which is about 0.25 dB higher than that based on the two-dimensional constellation. This result is quite close to the predictions derived from the coding exponent calculations.

Concluding, we state that, for a fixed data delay, it is more *power and complexity efficient* to apply block codes as long as possible to MLC schemes based on a one-dimensional constellation instead of increasing the dimensionality of the signal constellation. For MLC schemes based on a one-dimensional signal set, the signal points in N_1 dimensions, i.e., the multilevel codeword, are exclusively given by the code constraints of the component codes, whereas for a D -dimensional signal set we have concatenated coding consisting of two constituent parts. The first part is covered by the component codes applying only to $N_D = N_1/D$ dimensions. The second part is given by the constraints between the components $a_k, k = 1, \dots, D$, of a signal point $a = (a_1, \dots, a_D)$, which are introduced by the set partitioning strategy. Our previous result indicates that the direct approach is more efficient than the concatenated one, as long as the overall dimensionality of the multilevel codeword is fixed.

This result is quite obvious since the set of codes that can be realized by MLC of two binary codes of length $N_2 = N_1/2$ on a four-way two-dimensional partition is a subset of the set of codes that can be realized by a single binary code of length N_1 on a two-way one-dimensional partition. Moreover, even if the codes were the same, one-dimensional MLD will be superior to two-dimensional two-level multistage decoding.

This disadvantage of two-dimensional constituent constellations could be overcome if quaternary component codes were admitted. Then the same set of codes could be generated as in the one-dimensional case. Therefore, in the case of nonbinary component codes the question of the optimal dimensionality of the constituent signal set translates to the question of the optimal partitioning depth of the lowest level, cf. [33].

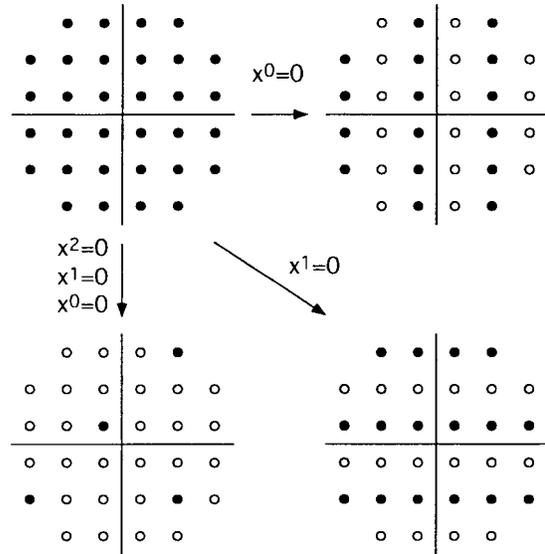


Fig. 14. Selection of subsets of a set partitioning for a QAM cross-constellation with 32 points implementing a particular labeling.

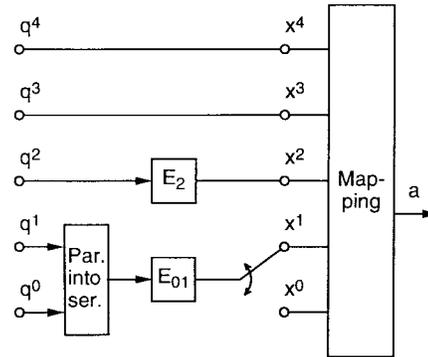


Fig. 15. Multilevel encoding scheme for a 32-QAM cross-constellation with set partitioning according to Fig. 14.

B. Nonsquare QAM Constellations

In QAM schemes based on nonsquare constellations, like the popular cross constellations with 32 or 128 points [61], dimensions are not completely separable. But interdependencies between the dimensions are only relevant at the perimeter of the constellation. Hence, if we neglect these interdependencies, it is possible to encode the lowest partitioning levels of cross constellations exactly as is done for ASK signal sets. Because of the high intra-subset Euclidean distance, the highest levels (where the interdependency of dimensions has to be taken into account) remain uncoded. For example, Fig. 14 shows some subsets of a set partitioning for the 32-ary cross-constellation implementing a specific labeling. Obviously, since the subsets $x^0 = 0$ and $x^1 = 0$, respectively, differ only by a rotation they provide equal capacities. Additionally, if we again neglect the boundary effects, only the in-phase (quadrature) component of the noise is effective at level 0 (1). Thus the transmission of x^0 (x^1) is essentially based on a 6-ASK constellation, multiplexed to in-phase and quadrature components, and the coding of the symbols x^0 and x^1 can be done by a single encoder E_{01} . The block diagram of the resulting MLC scheme of rate $R = 4$ bits/symbol = 2 bits/dimension is given in Fig. 15.

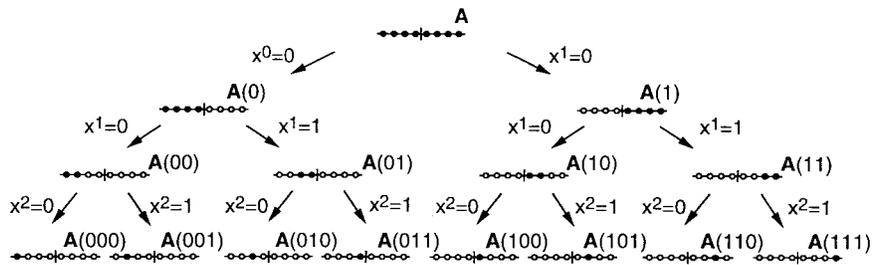


Fig. 16. Partitioning tree for the 8-ASK signal set when block partitioning is used.

When signal shaping for minimum average signal power is applied, a square QAM constellation can always be used, because a proper constellation boundary is determined in a high-dimensional space by the shaping procedure. Therefore, simple MLC based on a one-dimensional constellation is always possible.

C. Convolutional Component Codes

Using convolutional codes as component codes, a comparison is much more difficult, because in this case no clear concept of dimensionality exists. Simulation results using 64-state convolutional codes (CC) as component codes promise better performance for two-dimensional signaling, see Fig. 13. Again, the total code rate per dimension is fixed to $R/D = 1.5$ bits/dimension. (The individual code rates, which are implemented by nonoptimized, regular puncturing, are assigned by the cutoff rate rule, cf. Section IV.) 16-QAM with MLC using convolutional codes outperforms the MLC scheme based on 4-ASK. The reason is that increasing the dimensionality of the constituent signal constellation and applying MLC with Ungerboeck's set partitioning leads to an increased effective constraint length, as we have chosen equal constraint lengths for each code.

But this improvement has to be paid for twice: the receiver complexity is increased, because the two-dimensional approach needs three individual decoders instead of two for the one-dimensional case, and the signal delay is doubled when the lengths of the survivor memories are fixed. However, when employing convolutional codes, data delay generally is not a problem, and only half the processing speed for each individual decoder is necessary under a two-dimensional approach.

Strictly speaking, the result of Section V-A applies only if the block length of the component code is a relevant parameter, i.e., for codes where performance is related to the code length. In the case of convolutional codes, the higher dimensional constituent constellation leads to a more power-efficient concatenated coding scheme since the effective constraint length is enlarged.

VI. LABELING STRATEGIES

In this section labeling strategies that offer some practical advantages, are presented. MLC/MSD with finite code length is considered, cf. also [35], [37], [62], and [63]. Moreover, an analysis of bit-interleaved coded modulation, cf. [42], for finite code length is presented.

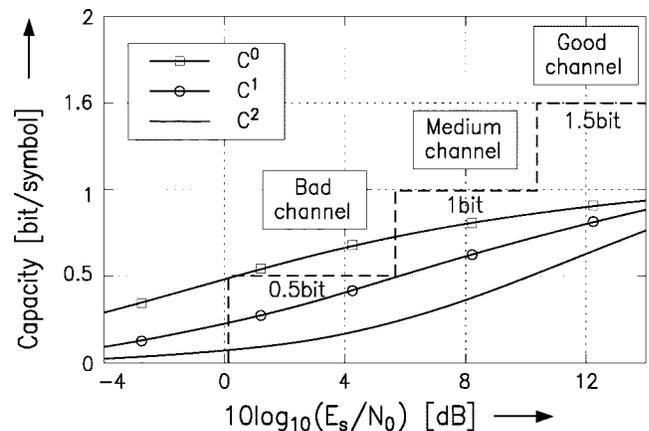


Fig. 17. Capacities C^i of the equivalent channels versus E_s/N_0 for the 8-ASK signal set when block partitioning is used. AWGN channel. Dashed line: aggregate transmittable rate for $R^i = 0.5$, $i = 0, 1, 2$. Bad channel marks the range with low SNR, medium channel with medium SNR, and good channel with high SNR.

A. Labeling Strategies for MLC/MSD

1) *Block Partitioning*: Usually, the strategy for partitioning a signal set is to maximize the minimum intra-subset Euclidean distance, as introduced by Ungerboeck [4] and Imai [5]. Here, an alternative labeling with the opposite strategy, called block partitioning (BP), is investigated. The minimum intra-subset Euclidean distance remains constant for all partitioning levels and the partitioned subsets form dense clusters of signal points. The strategy for block partitioning is to minimize the intra-subset variance. As an example, the block partitioning tree for the 8-ASK signal set is depicted in Fig. 16.

Fig. 17 sketches the corresponding capacity curves for 8-ASK. Since the minimum intra-subset Euclidean distance is equal for all partitioning levels, the capacities C^i of the equivalent channels i of an MLC scheme with block partitioning decrease from the lowest level to the highest level

$$C^i \geq C^{i+1}, \quad i = 0, \dots, \ell - 2. \quad (24)$$

This property is well suited for softly degrading schemes for transmission over channels with time-varying SNR. Consider the following example: The individual code rates R^i , $i = 0, 1, 2$, of an MLC scheme with BP and 8-ASK signal set can be chosen to be equal for all levels, say $R^i = 0.5$. According to Fig. 17, three states of the channel—"good," "medium," "bad"—are defined, depending on the current E_s/N_0 at the receiver side. First, for transmission over the "good" channel, the capacities C^i exceed the code rates R^i at each level i and,

hence, in principle, error-free transmission of $R^0 + R^1 + R^2 = 1.5$ bits/symbol is possible. Second, for transmission over the “medium” channel, the code rate R^2 at the highest level exceeds the capacity C^2 while R^0 and R^1 are still smaller than the corresponding capacities C^0 and C^1 . Hence, starting at level 0, error-free multistage decoding at levels 0 and 1 is still possible resulting in a total transmission rate of $R^0 + R^1 = 1.0$ bit/symbol. Third, for transmission over the “bad” channel, only code rate R^0 is smaller than capacity C^0 , while code rates R^1 and R^2 exceed the corresponding capacities C^1 and C^2 . Hence, over the “bad” channel error-free transmission of 0.5 bit/symbol is possible. To summarize, using this construction a soft degradation of the scheme over a wide range of SNR is enabled. It adapts automatically to the channel state without the need of a feedback channel, simply by estimation of the SNR at the receiver side or a cyclic redundancy check (CRC), and by delivering only the reliable part of the information to the sink. In the above example, the information increases in steps of 0.5 bit/symbol with increasing SNR. Such softly degrading schemes may be employed in “scalable” source coding schemes (video, audio), cf. [64]–[66], in broadcasting systems, cf. [67], or in mobile communications.

Of course, this way of constructing softly degrading schemes by designing rates for upper levels according to transmission at high SNR can be based on any labeling strategy. But with natural mapping, the rate steps are quite different and very little rate remains at low SNR. Block partitioning is a simple way to achieve equal rate steps. Moreover, it is interesting to note that, in the case of equal code rates R^i at the individual levels i , MLC schemes with BP do not suffer from error propagation in MSD.

The concept of capacities of the equivalent channels provides the framework for optimization of constellations for softly degrading schemes, sometimes called *multiresolutional constellations* [68]. Results presented in [69] show that an optimization via information-theoretic parameters leads to constellations completely different from those constructed from Euclidean distance arguments, like those used in digital video broadcasting in Europe [70]. If powerful coding is applied, the Euclidean distance becomes a less important parameter and the latter approach becomes very inefficient. The differences are comparable to those observed in our comparison of the capacity rule and the balanced distances rule.

2) *Mixed Partitioning*: Now, we will focus on a labeling strategy for 8-PSK such that for MLC/MSD the capacities C^0 and C^1 of the equivalent channels 0 and 1 are approximately equal. The important feature of such a scheme is that if the code rates R^i at the individual levels i are chosen to be approximately equal to the capacities C^i , then $R^0 = R^1$ is possible. Hence, it is sufficient to implement one encoder and one decoder for both levels. The proposed scheme is depicted in Fig. 18. Encoders E_0 and E_1 are both implemented by the encoder E_{01} , and decoders D_0 and D_1 are both implemented by the decoder D_{01} .

In order to provide similar capacities C^0 and C^1 , the (sub)sets \mathbf{A} and $\mathbf{A}(x^0)$ must exhibit similar distance properties. This can be achieved by mixing Ungerboeck and block

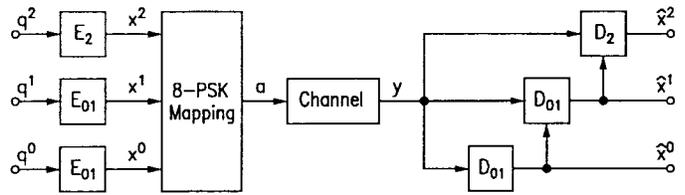


Fig. 18. MLC/MSD transmission system for mixed partitioning of 8-PSK set.

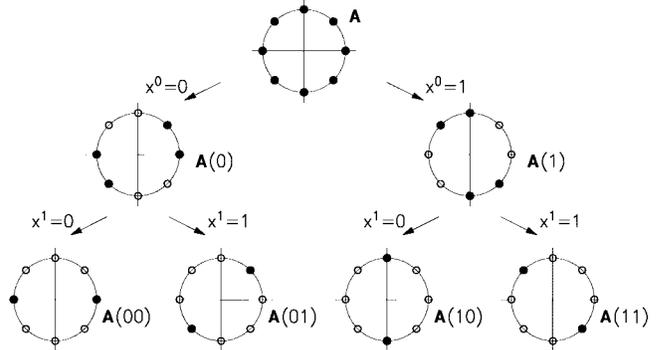


Fig. 19. Tree for mixed partitioning of 8-PSK set.

partitioning, called mixed partitioning (MP). At partitioning level 2 the subsets $\mathbf{A}(x^0x^1)$ should maximize minimum intra-subset Euclidean distance as in Ungerboeck’s approach. The goal of the first partitioning step is to minimize the intra-subset variance under the constraint that subsets $\mathbf{A}(x^0x^1)$ with the properties described above are still possible.

Fig. 19 shows an example of mixed partitioning for the 8-PSK signal set. Let us compare distance properties for transmission of symbol x^0 by the set \mathbf{A} and symbol x^1 by the set $\mathbf{A}(0)$ or set $\mathbf{A}(1)$. In each case, even for small SNR’s, the performance is dominated by the respective minimum distance d_i , $i = 0, 1$. From Fig. 19 it is obvious that

$$d_0 = d_1 = 2\sqrt{E_s} \sin \frac{\pi}{8}.$$

Additionally, the number of nearest neighbors representing the complementary binary symbol coincides in both cases. Hence, coded transmission using the (sub)set \mathbf{A} and $\mathbf{A}(x^0)$ exhibits similar distance properties.

The capacities C^i , $i = 0, 1, 2$, of the 8-PSK scheme with MP operating on the AWGN channel are plotted in Fig. 20 versus $10 \cdot \log_{10}(E_s/N_0)$. Indeed, one can see that C^0 and C^1 are hardly distinguishable. If we design the corresponding MLC scheme for total rate $R = 2$ bits/symbol according to the capacity rule, the rates are

$$R^0/R^1/R^2 = C^0/C^1/C^2 = 0.51/0.51/0.98. \quad (25)$$

In comparison, the corresponding rates for an MLC scheme and 8-PSK signal set based on Ungerboeck’s partitioning are

$$R^0/R^1/R^2 = C^0/C^1/C^2 = 0.2/0.81/0.99. \quad (26)$$

In this case, three different encoders and decoders are required. For the MLC scheme based on mixed partitioning, $R^0 = R^1$ holds, and hence two different encoders (E_{01} and E_2) and decoders (D_{01} and D_2) are sufficient, cf. Fig. 18.

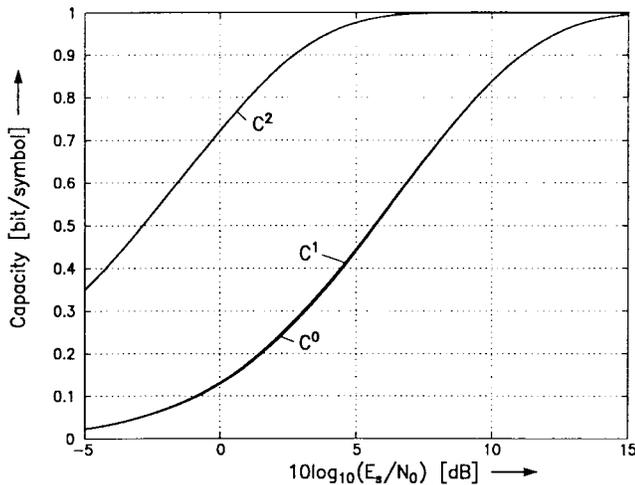


Fig. 20. Capacities C^i of the 8-PSK scheme with MP versus E_s/N_0 for the AWGN channel.

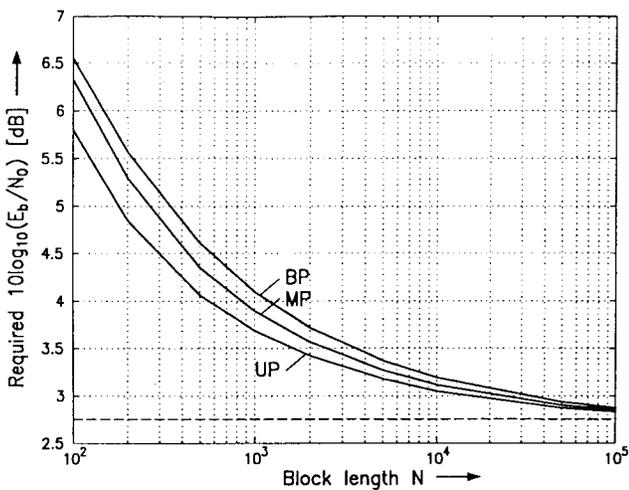


Fig. 21. Required E_b/N_0 versus codeword length N for transmission over the AWGN channel using MLC and MSD with 8-PSK for a tolerable word-error rate $p_w = 10^{-3}$. $R = 2$ bits/symbol. UP: Labeling by Ungerboeck's partitioning. MP: Mixed partitioning. BP: Block partitioning.

It is worth mentioning that with this mixed partitioning approach, coding is still based on MLC/MSD. Levels 0 and 1 are not combined into a single level as in bit-interleaved coded modulation, cf. Section VI-B. Instead, with the proposed scheme, the hardware of one encoder and one decoder can be saved. Clearly, since usually the individual encoders E_0 and E_1 of an MLC scheme work in parallel, encoder E_{01} has to work at double speed.

3) *Coding Exponent Analysis*: Using block codes of length N at all levels, the power efficiency of these labeling strategies may be evaluated by the coding exponents of the equivalent channels, cf. Section IV. As an example, we consider transmission of $R = 2$ bits/symbol over the AWGN channel using an 8-PSK constellation. The tolerable word-error rate is assumed to be $p_w = 10^{-3}$. In each case, the required E_b/N_0 for an MLC scheme with block length N is calculated via isoquants of coding exponents, see Fig. 21. As expected, as block length tends to infinity, the required SNR for all labeling

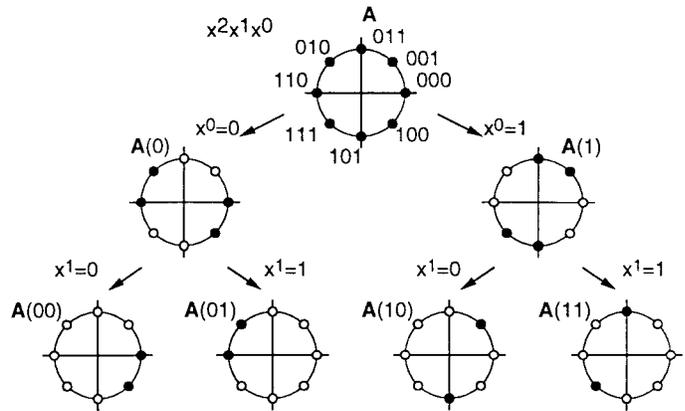


Fig. 22. Set partitioning tree of a 8-PSK constellation with Gray labeling of the signal points.

strategies merges into the capacity limit of $10 \log_{10}(E_b/N_0) = 2.7$ dB for $C = 2$ bits/symbol and 8-PSK. For finite N , the labeling strategy introduced by Ungerboeck and Imai in an intuitive manner shows the best performance, but the gain compared to the other labelings is relatively small. Simulation results confirm this statement, cf. [60] and [63]. Thus the alternative labeling strategies, which offer some interesting practical advantages, do not suffer a significant degradation in power efficiency.

B. Gray Labeling

Beside the set partitioning strategies discussed up to now, *Gray labeling* of the signal points is an interesting alternative. In [71], a pragmatic approach to coded 8-PSK modulation is presented using a single rate-2/3 convolutional encoder. In this approach, the three encoded bits are (randomly) interleaved independent of each other and mapped to the signal points using Gray labeling. It was shown for the Rayleigh fading channel that because of increased diversity due to bit interleaving, this scheme outperforms the best known trellis codes. Stimulated by this work, Caire *et al.* [42], [1] recently investigated the capacity of *bit-interleaved coded modulation* (BICM) schemes over the AWGN channel. In BICM schemes one binary encoder is used with subsequent (random) bit interleaving, followed by demultiplexing the bits to select the signal point. The results show that for 8-PSK and 16-QAM schemes in the range of practical interest, the capacity loss of BICM versus the optimum approach is negligible if (and only if) Gray labeling is used.

In what follows, we present a strict derivation of BICM starting from MLC and MSD using Gray labeling. For this discussion 8-PSK is assumed as an example.

1) *MLC and MSD Using Gray Labeling*: First, we study the properties of an 8-PSK MLC scheme with MSD and Gray labeling. The corresponding set partitioning tree is sketched in Fig. 22. Notice that, in contrast to Ungerboeck's set partitioning, the minimum intra-subset Euclidean distance remains constant for all partitioning levels. Moreover, in contrast to the other labeling strategies discussed here, the set partitioning for Gray labeling is irregular, e.g., subset $A(01)$ is not a rotated version of $A(10)$. Hence, according to (7) the

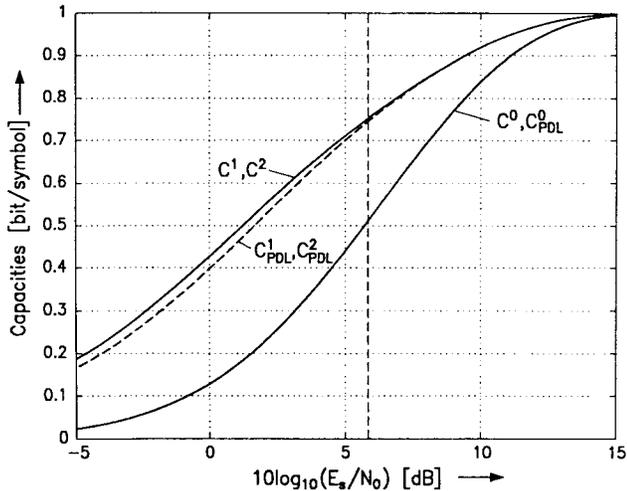


Fig. 23. Capacities C^i of the equivalent channels of an MLC scheme based on 8-PSK constellation with Gray labeling for transmission over the AWGN channel. Solid lines: MSD, dashed line: PDL. Vertical line: total rate $R = 2$ bit/symbol.

capacities C^1 and C^2 of the equivalent channels have to be averaged over the different subsets. Using the fact that some subsets have equal capacities, the individual capacities

$$\begin{aligned} C^0 &= C(\mathbf{A}) - C(\mathbf{A}(0)) \\ C^1 &= C(\mathbf{A}(0)) - \frac{1}{2}[C(\mathbf{A}(01)) + C(\mathbf{A}(10))] \end{aligned}$$

and

$$C^2 = \frac{1}{2}[C(\mathbf{A}(01)) + C(\mathbf{A}(10))]$$

for transmission over the AWGN channel have been plotted in Fig. 23 (solid lines). The equivalent channels at level 1 and 2 provide nearly the same capacity; the difference is not visible in Fig. 23. The reason is that the transmission of symbol x^1 by $\mathbf{A}(0)$ and $\mathbf{A}(1)$ exhibits very similar distance properties to the transmission of symbol x^2 by $\mathbf{A}(01)$ and $\mathbf{A}(10)$, respectively. The rate distribution according to the capacity rule for MLC and MSD based on 8-PSK with Gray labeling and rate $R = 2$ bit/symbol yields (see solid vertical line in Fig. 23)

$$R^0/R^1/R^2 = C^0/C^1/C^2 = 0.51/0.745/0.745. \quad (27)$$

Note that, compared to Ungerboeck labeling with individual capacities $C^0/C^1/C^2 = 0.2/0.81/0.99$, the range of individual code rates for Gray labeling is substantially smaller.

Regarding the subsets of the set partitioning tree in Fig. 22, a careful analysis shows that in each case the number of nearest neighbor signal points representing the complementary binary symbol (error coefficient) is equal to 1. In particular, neither the error coefficient nor the minimum intra-subset distance changes when decisions of lower levels are not taken into account for decoding at higher levels. Hence, we conjecture that without significant loss, the transmission of the address symbol x^i , $i = 0, 1, 2$, can be based on the **entire** signal constellation, i.e., the individual levels may be decoded *in parallel* without any preselection of signal points at higher levels. Subsequently, independent *parallel decoding* of levels is investigated in detail.

2) *MLC and Parallel Independent Decoding of the Individual Levels Using Gray Labeling*: In MLC with parallel, independent decoding of the individual levels (PDL), the decoder D^i makes no use of decisions of other levels $i \neq j$. In order to investigate this scheme, the definition of the equivalent channel and its characterizing pdf has to be adapted appropriately. In the case of MLC/MSD, the equivalent channel i for transmission of digit x^i comprises the time-varying equivalent modulator i , depending on the binary digits x^j of lower levels j , $j = 0, \dots, i-1$, and the actual physical channel. Here, in the case of MLC/PDL, the equivalent modulator i is no longer time-varying. Since decoding at level i is done independently of other levels, the equivalent modulator i for MLC/PDL is based on the entire signal set, whereby all signal points with address digit $x^i = b$, $b \in \{0, 1\}$, represent the binary symbol b . Therefore, the equivalent channel i for MLC/PDL is characterized by the pdf's

$$\begin{aligned} f_{Y, \text{PDL}}(y|x^i, i) &= E_{a \in \mathbf{A}_i(x^i)} \{f_Y(y|a)\}, \\ x^i &\in \{0, 1\}, \quad i = 0, \dots, \ell-1. \end{aligned} \quad (28)$$

Thereby, subset $\mathbf{A}_i(x^i)$ is defined by

$$\begin{aligned} \mathbf{A}_i(x^i) &:= \{a = \mathcal{M}(x^0 \dots x^{i-1} x^i x^{i+1} \dots x^{\ell-1}) \\ &\quad (x^0 \dots x^{i-1}, x^{i+1} \dots x^{\ell-1}) \in \{0, 1\}^{\ell-1}\}. \end{aligned} \quad (29)$$

Since decoder D^i makes no use of decisions of other levels j , $j \neq i$, in MLC/PDL the maximum individual rate at level i to be transmitted at arbitrary low error rate is bounded by

$$R^i \leq I(Y; X^i), \quad i = 0, \dots, \ell-1. \quad (30)$$

Consequently, the total rate is restricted to

$$R = \sum_{i=0}^{\ell-1} R^i \leq \sum_{i=0}^{\ell-1} I(Y; X^i). \quad (31)$$

The bound

$$I(Y; X^i) \leq I(Y; X^i | X^0 \dots X^{i-1}) \quad (32)$$

is valid in general, with equality iff for a given channel output variable y the input symbols x^i and x^j , $0 \leq j < i$, are independent. This is true as the signal-to-noise ratio goes to infinity. Therefore, combining (6), (31), and (32) yields that the sum of maximum rates R^i in an MLC/PDL scheme must be less than or equal to the mutual information for the total scheme

$$\begin{aligned} R &= \sum_{i=0}^{\ell-1} R^i \leq \sum_{i=0}^{\ell-1} I(Y; X^i) \\ &\leq \sum_{i=0}^{\ell-1} I(Y; X^i | X^0 \dots X^{i-1}) = I(Y; X^0 \dots X^{\ell-1}). \end{aligned} \quad (33)$$

Thus asymptotically, the mutual information of the modulation scheme can be approached with MLC/PDL iff $R^i = I(Y; X^i)$.

Equation (33) shows that the MLC/PDL approach is simply a suboptimum approximation of an optimum coded modulation scheme. In contrast to the optimum scheme, the capacity of the MLC/PDL scheme strongly depends on the particular labeling

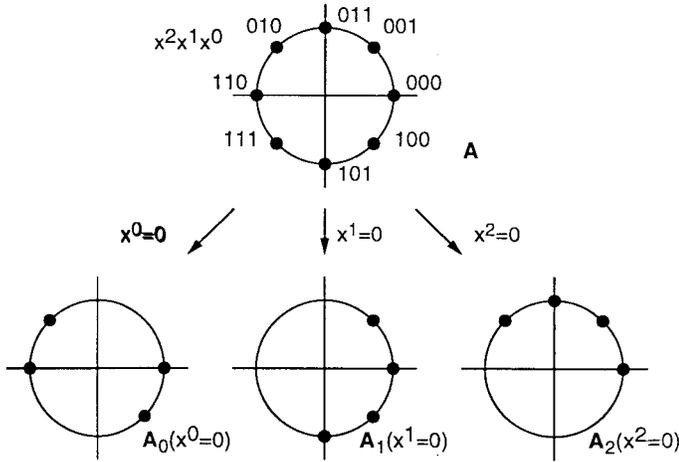


Fig. 24. Subsets $\mathbf{A}_i(x^i = 0)$ according to (29) with 8-PSK with Gray labeling.

of signal points. Caire *et al.* showed in [42] and [1] that the gap to an optimum scheme is surprisingly small with Gray labeling of the signal points. However, we will show now that the gap increases significantly if the finite length of the code is taken into account.

For example, the 8-PSK subsets \mathbf{A}_i , $i = 0, 1, 2$, for $x^i = 0$ and Gray labeling are sketched in Fig. 24. Since the subsets $\mathbf{A}_i(x^i = 0)$ and $\mathbf{A}_i(x^i = 1)$, $i = 0, 1, 2$, differ only by rotation, they provide equal capacities. Moreover, the subsets $\mathbf{A}_1(x^1 = 0)$ and $\mathbf{A}_2(x^2 = 0)$ differ only by rotation, too. Hence, according to (7), the individual capacities for the equivalent channels using PDL are given by

$$\begin{aligned} C_{\text{PDL}}^0 &= C(\mathbf{A}) - C(\mathbf{A}_0(0)) \\ C_{\text{PDL}}^1 &= C_{\text{PDL}}^2 = C(\mathbf{A}) - C(\mathbf{A}_1(0)). \end{aligned} \quad (34)$$

(See dashed line in Fig. 23.) Since at level 0 the equivalent channels for MLC using MSD and PDL are equal $C_{\text{PDL}}^0 = C^0$ holds. Clearly, the capacities C_{PDL}^1 and C_{PDL}^2 using PDL are lower than those using MSD. But from Fig. 23 it is apparent that in the range of interest the loss of parallel compared to multistage decoding is very small. Additionally, the rate distribution for MLC using PDL according to the capacity design rule for a total rate $R = 2$ bits/symbol is nearly the same as for MSD (cf. dashed vertical line in Fig. 23).

Finally, following Section IV and considering (28), the coding exponent $E_{0,\text{PDL}}^i(\rho)$ of the equivalent channel i for an MLC scheme using PDL is

$$\begin{aligned} E_{0,\text{PDL}}^i(\rho) &= -\log_2 \left\{ \int_{\mathbf{Y}} \left[\sum_{x^i=0}^1 \Pr\{x^i\} (f_{Y,\text{PDL}}(y|x^i, i))^{\frac{1}{1+\rho}} \right]^{1+\rho} dy \right\}. \end{aligned} \quad (35)$$

3) *Bit Interleaved Coded Modulation*: As described above, in MLC schemes using PDL the output symbol x^i of encoder i is transmitted over the time-invariant equivalent channel i . Thus ℓ independent binary coding and decoding levels are present in parallel. An obvious approach is to apply only one

binary code and to group ℓ encoded bits to address the current symbol. Assuming ideal bit interleaving,⁵ the address symbols are independent of each other and, hence, this scheme can be interpreted—at least for infinite code length—as multilevel encoding together with parallel, independent decoding of the individual levels.

BICM transmission of binary symbol x^i can again be viewed as transmission over the equivalent channel i for MLC/PDL. But here, the equivalent channels i , $i = 0, \dots, \ell - 1$, for MLC/PDL are not used in parallel; rather, they are time-multiplexed. Hence, the equivalent channel using BICM is characterized by a set $f_{Y,\text{BICM}}(y|x)$ of pdf's for the binary encoder output symbol x

$$f_{Y,\text{BICM}}(y|x) = \{f_{Y,\text{PDL}}(y|x^i = x, i) | i = 0, \dots, \ell - 1\}. \quad (36)$$

Since the equivalent channel using BICM is time-variant with ideal channel state information at the receiver (side information $\hat{=}$ actual level), the corresponding coding parameter $E_{0,\text{BICM}}(\rho)$ is the average of $E_{0,\text{PDL}}^i(\rho)$ for $i = 0, \dots, \ell - 1$. Assuming equal probability for all channels i , we have

$$E_{0,\text{BICM}}(\rho) = \frac{1}{\ell} \sum_{i=0}^{\ell-1} E_{0,\text{PDL}}^i(\rho). \quad (37)$$

The random coding exponent for the equivalent channel using BICM is then

$$E_{r,\text{BICM}}(R_B) := \max_{0 \leq \rho \leq 1} \{E_{0,\text{BICM}}(\rho) - \rho R_B\}. \quad (38)$$

Here, R_B denotes the rate of the binary code used in BICM.

For simplicity, MLC/PDL and BICM are assumed here at all levels. For the AWGN channel and very large signal constellations, it is more efficient to apply a mixed labeling strategy, i.e., to use Ungerboeck labeling in order to separate coded from uncoded levels and thus to save complexity. A relabeling within the coded levels according to the Gray criterion allows use of MLC/PDL or BICM for subset coding/decoding. For fading channels, Gray labeling and PDL or BICM over all levels is recommended.

The same principle applies when hard-decision decoding at higher levels is possible without significant loss, see Section VII. In this case also, only the levels with soft-decision decoding should be combined.

4) *Coding Exponent Analysis*: To compare the power efficiency of the previously presented schemes, the coding exponents given above are evaluated. Coded 8-PSK transmission at $R = 2$ bits/symbol over the AWGN channel is chosen with a tolerable word-error rate $p_w = 10^{-3}$. Additionally, the result of the previous section for the usual MLC scheme with Ungerboeck partitioning and MSD is included for reference. In all cases, the required SNR for the different schemes and fixed total rate R was calculated via isoquants of coding exponents. For MLC/MSD and MLC/PDL the code length N is used for all component codes, whereas for BICM only a single code of this length is used. In the case of BICM, the calculation

⁵Ideal interleaving increases data delay to infinity, whereas code length is not affected.

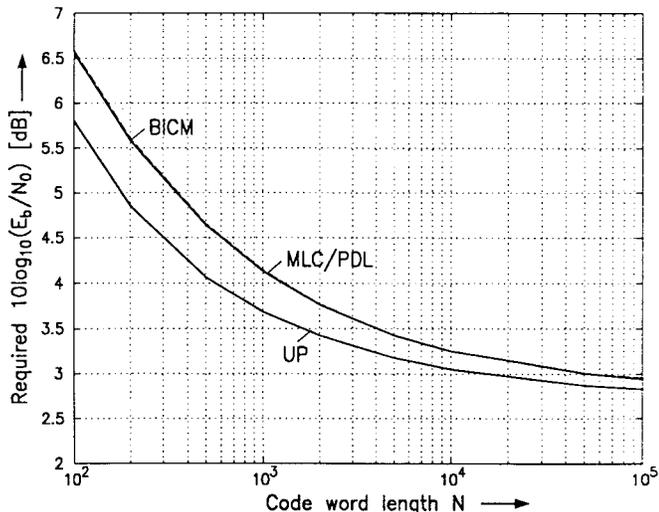


Fig. 25. Required E_b/N_0 versus code length N for coded transmission over the AWGN channel with 8-PSK for a tolerable word-error rate $p_w = 10^{-3}$. Coding schemes: 1. UP: MLC and MSD with Ungerboeck labeling. 2. MLC + PDL: MLC and PDL with Gray labeling (solid line). 3. BICM with Gray labeling (dashed line).

of isoquants provides the rate R_B of the single code, which has to be multiplied by ℓ to obtain the total rate $R = \ell R_B$ for BICM. The results for the competing schemes are shown in Fig. 25 as a function of the code length N .

From Fig. 25 it is evident that the performance of BICM schemes and that of MLC schemes using PDL is almost identical and inferior to that of MLC/MSD. For decreasing N , the loss of BICM and MLC/PDL compared to the optimum MLC/MSD approach using Ungerboeck labeling (UP) increases. The reason for this is as follows: Assume that the signal point $\mathcal{M}(000)$ in the subset $\mathbf{A}_1(x^1 = 0)$ is chosen for transmission of binary symbol $x^1 = 0$, cf. Fig. 24. Then, both nearest neighbor signal points represent the same binary symbol $x^1 = 0$. Hence, the minimum distance of $\mathcal{M}(000)$ to signal points representing the complementary binary symbol is relatively large, and the transmission of $x^1 = 0$ by $\mathcal{M}(000)$ is relatively robust against the channel noise. The situation changes if, e.g., the signal point $\mathcal{M}(001)$ in the subset $\mathbf{A}_1(x^1 = 0)$ is chosen for transmission of binary symbol $x^1 = 0$. Then, one nearest neighbor point represents the complementary binary symbol $x^1 = 1$ resulting in a substantially lower minimum distance than in the former case. Hence, the transmission of $x^1 = 0$ by $\mathcal{M}(001)$ is relatively sensitive to the channel noise. The same is true for the transmission of the binary symbol x^2 .

This example shows that the distance properties of a BICM scheme with Gray labeling are extremely time-varying. Therefore, in order to compensate for these varying distances a large block code is required. In particular, since convolutional codes can be viewed as short block codes with a sliding encoding window, they are not well suited to BICM schemes over the AWGN channel.

VII. HARD-DECISION DECODING AT INDIVIDUAL LEVELS

Up to now, we have considered only coded modulation schemes in which the individual decoding stages optimally de-

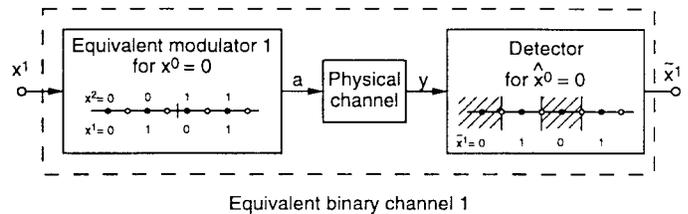


Fig. 26. Equivalent binary channel 1 for the transmission of symbol x^1 for an 8-ASK scheme when hard-decision decoding is used.

code the channel output. In this section, we investigate coded modulation schemes in which error-correcting binary codes are employed at selected coding levels; i.e., the corresponding decoders use only binary decisions without additional reliability information.

In the previous MLC literature, usually all component codes are either exclusively error-correcting algebraic block codes, e.g., [26], [72] or all codes are soft-decodable codes like convolutional or turbo codes [39], [25], [37]. With error-correcting codes, decoding complexity is usually much lower than with soft-decodable codes. It is well known, however, that hard decisions incur a performance loss compared to soft decisions; for binary antipodal signaling the loss is about 2 dB.

Results concerning the performance loss due to error-correcting codes in MLC schemes are thus desirable. In this section, we compare hard- and soft-decision decoding in MLC schemes. We show that the performance gap between hard- and soft-decision decoding decreases with increasing bandwidth efficiency. It turns out that soft-decision decoding at only the lowest level is sufficient to achieve near-optimum performance.

A. Capacity

For hard decisions, the concept of equivalent channels introduced in Section II has to be modified appropriately. Only binary variables should be fed to the decoder D^i at decoding stage i ; i.e., the channel output y is quantized to binary symbols \tilde{x}^i . To distinguish between equivalent channel i with hard and soft decisions, we refer to the former one as equivalent *binary* channel i .

For example, let us consider the equivalent binary channel for transmission of symbol x^1 for an 8-ASK scheme (level 1), see Fig. 26. Assuming equiprobable signal points, the detection thresholds are half-way between two adjacent points. In general, for a nonuniform distribution of signal points, the decision region $\mathcal{R}_{\mathbf{A}}(a)$ of signal point a out of the set \mathbf{A} has to be optimized according to the maximum *a posteriori* (MAP) criterion

$$\mathcal{R}_{\mathbf{A}}(a) = \{y | \Pr\{a\} \cdot f_Y(y|a) \geq \Pr\{a'\} \cdot f_Y(y|a'), a' \in \mathbf{A}, a \neq a'\}, \quad a \in \mathbf{A}. \quad (39)$$

Due to the multiple representation of binary symbols by signal points, the decision region of binary symbol \tilde{x}^i is the union of the corresponding regions $\mathcal{R}_{\mathbf{A}(x^0 \dots x^{i-1})}(a)$ (see Fig. 26)

$$\mathcal{R}(\tilde{x}^i | x^0 \dots x^{i-1}) = \bigcup_{a \in \mathbf{A}(x^0 \dots x^{i-1} \tilde{x}^i)} \mathcal{R}_{\mathbf{A}(x^0 \dots x^{i-1})}(a). \quad (40)$$

For the calculation of the capacity of the equivalent binary channel i the transition probabilities $\Pr\{\tilde{x}^i|x^i, x^0 \dots x^{i-1}\}$ are needed. Let us first assume that $x^i = 0$ is transmitted using the fixed signal point $a \in \mathbf{A}(x^0 \dots x^{i-1})$. Then, the probability of detection error is⁶

$$\Pr\{\tilde{x}^i = 1|a, x^0 \dots x^{i-1}\} = \sum_{\tilde{a} \in \mathbf{A}(x^0 \dots x^{i-1})} \Pr\{y \in \mathcal{R}_{\mathbf{A}(x^0 \dots x^{i-1})}(\tilde{a})|a\}. \quad (41)$$

The desired detection error probability is given by expectation over all possible signal points a representing symbol $x^i = 0$

$$\Pr\{\tilde{x}^i = 1|x^i = 0, x^0 \dots x^{i-1}\} = E_{a \in \mathbf{A}(x^0 \dots x^{i-1})} \{\Pr\{\tilde{x}^i = 1|a, x^0 \dots x^{i-1}\}\}. \quad (42)$$

If the *a priori* probabilities of the signal points are fixed, the capacity C_H^i of the equivalent binary channel i (hard decision) equals the corresponding mutual information:

$$\begin{aligned} C_H^i &= I(\tilde{X}^i; X^i | X^0 \dots X^{i-1}) \\ &= E_{(x^0 \dots x^{i-1})} \{I(\tilde{X}^i; X^i | x^0 \dots x^{i-1})\} \\ &= E_{(x^0 \dots x^{i-1}, \tilde{x}^i, x^i)} \left\{ \log_2 \left(\frac{\Pr\{\tilde{x}^i|x^i, x^0 \dots x^{i-1}\}}{\Pr\{\tilde{x}^i|x^0 \dots x^{i-1}\}} \right) \right\}. \end{aligned} \quad (43)$$

The results derived above are valid for an arbitrary distribution of signal points and for arbitrary labeling. When equiprobable signal points and regular partitioning (subsets at one partitioning level differ only by translation and/or rotation) are considered, the transition probabilities $\Pr\{\tilde{x}^i|x^i, x^0 \dots x^{i-1}\}$ are independent from the particular values of $\tilde{x}^i, x^i, x^0 \dots x^{i-1}$. In this particular case, the resulting equivalent binary channel i is symmetric; i.e., it is a *binary symmetric channel* (BSC). The transition probability $\Pr\{\tilde{x}^i \neq x^i|x^i, x^0 \dots x^{i-1}\}$ is simply the bit-error probability ε_i of the equivalent BSC i . Thus its capacity C_H^i is given by

$$C_H^i = 1 - H_2(\varepsilon_i) \quad (44)$$

where $H_2(\cdot)$ again denotes the binary entropy function.

Now, the capacity of a coded modulation scheme with a mix of hard- and soft-decision decoding at the individual levels is simply the sum of the capacities of the corresponding equivalent channels $i, i = 0, \dots, \ell - 1$. For example, consider a three-level scheme with soft-decision decoding at level 0 and hard-decision decoding at levels 1 and 2

$$C_{\text{SHH}} = C^0 + C_H^1 + C_H^2. \quad (45)$$

Here, the index of the capacity C_{SHH} denotes the decoding manner (soft or hard) at the individual levels. If the index is omitted, soft-decision decoding is assumed at all levels.

⁶The probability of detection error may be upper-bounded by the respective symbol error probability for transmitting $a \in \mathbf{A}(x^0 \dots x^{i-1})$. Thereby, the detection of signal points $\tilde{a} \in \mathbf{A}(x^0 \dots x^{i-1}), \tilde{a} \neq a$, are counted as errors although they represent the transmitted symbol $x^i = 0$. Especially when the detection error probability is mainly determined by the nearest neighbor signal points as it is the case with Ungerboeck set partitioning, this approximation becomes quite tight.

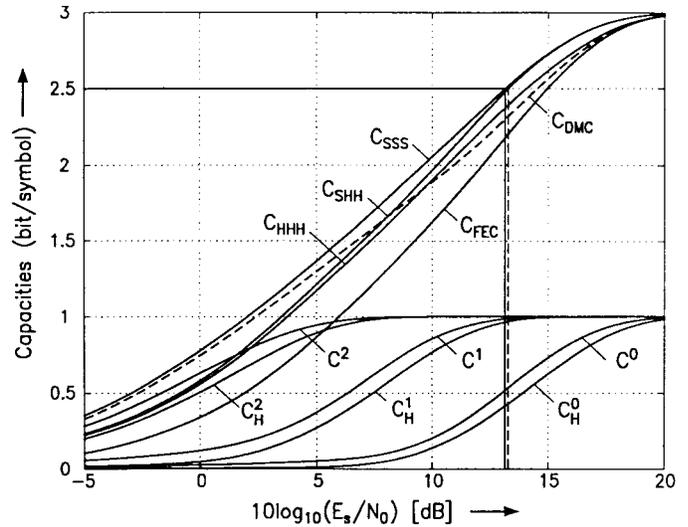


Fig. 27. 8-ASK with natural labeling. AWGN channel. Capacities for the curves SSS, SHH, and HHH (see text) and corresponding capacities C^i and C_H^i for the equivalent and binary equivalent channel $i, i = 0, 1, 2$. Dashed line: Capacity for the equivalent DMC. Solid vertical line: $C_{\text{SSS}} = 2.5$ bits/symbol. Dashed vertical line: $C_{\text{SHH}} = 2.5$ bits/symbol. C_{FEC} : binary FEC, Gray mapping and interleaving.

Using the chain rule of mutual information (2), it was shown in Section II that with soft-decision decoding at all levels, MLC together with MSD is an asymptotically optimum coding approach. However, when hard-decision decoding is used at several levels, the chain rule no longer holds, and the MLC approach is not necessarily optimum. Thus the MLC/MSD transmission scheme with hard-decision decoding at all levels operating on the AWGN channel is compared to a coded modulation scheme operating on an M -ary discrete memoryless channel (DMC). For example, we look at 8-ASK. The channel input and output alphabet is equal to the 8-ASK signal set $\mathbf{A} = \{\pm 1, \pm 3, \pm 5, \pm 7\}$. The transition probability $\Pr\{a_k|a_j\}$ for receiving symbol a_k is given by the probability that the output y of the underlying AWGN channel falls into the decision region $\mathcal{R}_{\mathbf{A}(a_k)}$ if symbol a_j is transmitted. The capacity of this scheme is given by

$$C_{\text{DMC}} = E_{(a_k \in \mathbf{A}, a_j \in \mathbf{A})} \left\{ \log_2 \frac{\Pr\{a_k|a_j\}}{\Pr\{a_k\}} \right\}. \quad (46)$$

B. Examples and Discussion

Again the coded modulation scheme with 8-ASK and natural labeling operating on the AWGN channel will be investigated. Fig. 27 depicts:

- capacity C , soft-decision decoding at all levels (case SSS, reference);
- capacity C_{SHH} , soft-decision decoding at level 0 and hard-decision decoding at levels 1 and 2 (case SHH);
- capacity C_{HHH} , hard-decision decoding at all levels (case HHH);

together with the corresponding capacities of the equivalent and the equivalent binary channels.

First, we compare the capacities C^i and $C_H^i, i = 0, 1, 2$. At level 2, the underlying signal set for the transmission of symbol x^2 is the ordinary BPSK signal set, i.e., symbol x^2

is not multiply represented. In this case, the well-known gap of about 1.7 dB between the capacities $C^2 = C_H^2 = 0.5$ is observed. At level 1, symbol x^1 is represented twice in the underlying 4-ASK signal set. Here, the gap between the capacities $C^1 = C_H^1 = 0.5$ is only 1.1 dB. This gap is further reduced to 0.9 dB at level 0. Thus when symbols are multiply represented by signal points, the gap between soft- and hard-decision channels becomes smaller.

We interpret this different behavior by the following observation: For binary antipodal signaling, the gain of soft-decision decoding mainly results from those highly reliable symbols that are received far away from the decision boundary. But for multi-amplitude/multi-phase modulation, the existence of boundaries of the decision regions on all sides of inner points reduces the soft-decision advantage. Also, as usual, the gap between the capacities C^i , $i = 0, 1, 2$, and C_H^i , respectively, decreases for increasing SNR.

Second, we compare the cases SSS and SHH. It is apparent from Fig. 27 that the gap between C and C_{SHH} is negligible for rates above 2.5 bits/symbol. In particular, for $R = 2.5$ bits/symbol, the loss for the case SHH versus the optimum case SSS is 0.15 dB. Even for rates down to 2.0 bits/symbol a loss of only 0.6 dB is visible. Hence, the performance loss due to hard-decision decoding at higher levels is dramatically reduced compared to BPSK. The explanation is as follows. For set partitioning according to Ungerboeck's criterion, the individual rates increase from lowest level to highest level. Thus the performance loss due to hard-decision decoding decreases. Hence, if hard-decision decoding is used at higher levels, where high-rate codes are employed, the loss remains small.

An additional loss occurs if hard-decision maximum-likelihood decoding is replaced by bounded-distance decoding. This loss cannot be assessed exactly, but for high-rate codes bounded-distance decoding is close to hard-decision maximum-likelihood decoding. In the case SHH, where only high-rate codes are used for hard-decision decoding, there is a little additional loss for practical bounded-distance decoding algorithms.

If we design the 8-ASK system for a total rate $R = 2.5$ bits/symbol, the rate distribution according to capacities in the case SSS (solid vertical line) is

$$R^0/R^1/R^2 = 0.516/0.984/1 \quad (47)$$

whereas in the case SHH (dashed vertical line) it is

$$R^0/R^1/R^2 = 0.532/0.968/1. \quad (48)$$

For SHH, the rate at level 0 with soft-decision decoding is slightly increased while the rate at level 1 with hard-decision decoding is decreased by the same amount when compared to SSS.

Next, we assess hard-decision decoding at all levels (HHH). For a total rate $R = 2.5$ bits/symbol, the gap between C and C_{HHH} is about 0.9 dB (see Fig. 27). The loss due to full hard-decision decoding in a coded modulation scheme is thus substantially less than with BPSK. The reason is that with Ungerboeck's set partitioning, the lowest rate is transmitted at

level 0 and, hence, the capacity loss at this level dominates. But, as shown above, at level 0 the loss due to hard-decision decoding is moderate because of the multiple representation of the symbol x^0 in the underlying signal set.

In conclusion, we see that for 8-ASK and $R = 2.5$ bits/symbol, it is sufficient to employ soft-decision decoding at level 0. Hard-decision decoding at higher levels can be done without any significant performance loss while offering a reasonable reduction in complexity, cf. also [33]. Even if hard-decision decoding is used at all levels, the loss compared to soft-decision decoding is less than 1 dB.

Finally, the capacities C_{DMC} of the equivalent DMC and C_{HHH} for the MLC/MSD scheme with hard-decision decoding at each level are compared. It can be observed from Fig. 27 that, in the region of interest (rates above 2.0 bits/symbol), the MLC approach with hard-decision decoding at each level outperforms the scheme operating on the equivalent DMC. The difference between the schemes lies in the way the soft output of the underlying AWGN channel is exploited. For the 8-ary DMC, the soft channel output is quantized once. In multistage decoding, the soft value is used at each level for binary detection. This leads to the important observation that, in the case of hard-decision decoding, an efficient way to exploit the channel information is to split up the coding and decoding procedure in multiple, ideally binary, levels so as to use the soft channel output multiply. Therefore, when designing coded modulation schemes, the two cases of soft- and hard-decision decoding have to be carefully distinguished. As shown in the previous section, in the case of soft-decision decoding and very large block length it is possible to link several coding levels into a single one without significant performance loss when Gray mapping is used. In the case of hard-decision decoding, this is not true since combining multiple levels into a single one results in an unavoidable performance loss due to a suboptimum exploitation of the soft information. It is well known that adapting binary error-correcting codes (forward error correction (FEC)) to an M -ary transmission scheme by Gray mapping does not lead to satisfactory results. This effect is confirmed by the curve C_{FEC} in Fig. 27, where for $C = 2.5$ bits/symbol a loss of approximately 2 dB compared to C_{SSS} results, which is in accordance to binary transmission. In contrast, the MLC approach with individual error-correcting codes at each level promises much better performance.

The question of the optimum dimensionality of the constituent signal set (cf. Section V) has to be discussed again for hard-decision decoding. Especially, hard-decision decoding at all but the lowest level is addressed, because of its interesting tradeoff between performance and complexity. As in Section V, an MLC scheme based on a 4-ASK signal set is compared to an MLC scheme based on a 16-QAM signal set. Again, the overall block length is assumed to be equal. In Fig. 28 the corresponding isoquants for a block length of 2000 QAM symbols are shown. A total rate of $R = 1.5$ bits/dimension is chosen. Here, the loss of the two-dimensional QAM scheme when compared to the one-dimensional ASK scheme is higher than with soft-decision decoding at all levels, cf. Fig. 16. The reason is that for the 16-QAM scheme there are three levels with hard-decision decoding with a rate

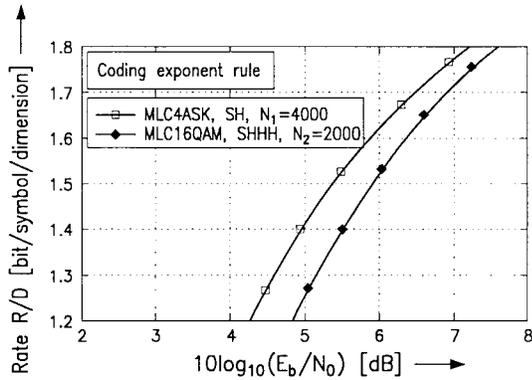


Fig. 28. Total rates $R/D = \sum_i R^i/D$ per dimension from isoquants $E^i(R^i)$ of 4-ASK in the case SH (soft-decision decoding at level 0, hard-decision decoding at level 1) and of 16-QAM in the case SHHH. For 4-ASK block length $N = 4000$ and for 16-QAM $N = 2000$ is assumed. $p_w = 10^{-3}$. Ungerboeck set partitioning. AWGN channel.

$R^1 = 0.7$ at level 1 rather far away from 1. Contrary, there is only one level with hard-decision decoding in the one-dimensional approach with a rate $R^1 = 0.96$ quite close to 1. This result further supports the statements concerning dimensionality in Section V. If hard-decision decoding at all but the lowest level is used, the less complex MLC scheme based on a one-dimensional signal set is more power-efficient than an MLC scheme based on a two-dimensional set.

As stated in Section V, if nonbinary codes are considered these differences are conjectured to vanish. Moreover, we assume that a two-dimensional SSHH approach will perform close to the one-dimensional SH scheme. But since two soft-decision decoders are required, complexity is almost doubled.

C. Simulation Results

In order to verify these capacity results, simulations for 8-ASK transmission with an MLC/MSD scheme over the AWGN channel have been performed. In particular, we are interested in the loss of an MLC scheme using hard-decision decoding at the two highest levels when compared to a scheme using entirely soft-decision decoding. For reference, the MLC scheme with the individual rates $R^0/R^1/R^2 = 0.52/0.98/1.0$ derived from the coding exponent rule is used, where turbo codes of 16-state convolutional codes [44], [56] and code length $N = 4095$ are employed as component codes. Again, flexible rates of turbo codes are achieved via puncturing, see [57] and [37]. For the competing scheme, a turbo code with $R^0 = 0.52$ at level 0 and a primitive Bose-Chaudhuri-Hocquenghem (BCH) code of length $N = 4095$ at level 1 are employed. Level 2 remains uncoded. The error-correcting capability of the BCH code is adjusted such that the individual performance at level 0 and level 1 is similar (equal error probability rule). As a result, the required error-correcting capability of the BCH code is $t = 14$ errors and, hence, $R^1 = 0.96$.

The simulation results are depicted in Fig. 29. For the reference scheme $10\log_{10}(E_b/N_0) = 10.1$ dB is required to achieve $BER = 10^{-5}$ with a total rate $R = 2.5$ bits/symbol. Since the capacity $C = 2.5$ bits/symbol is reached for

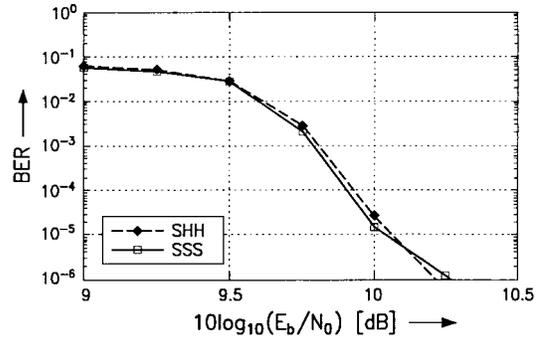


Fig. 29. BER of 8-ASK transmission with MLC/MSD over the AWGN channel. Codeword length $N = 4095$. Solid line (case SSS): Rate distribution $R^0/R^1/R^2 = 0.52/0.98/1.0$, total rate $R = 2.5$ bits/symbol. Turbo codes are employed as component codes. Dashed line (case SHH): Rate distribution $R^0/R^1/R^2 = 0.52/0.96/1.0$, total rate $R = 2.48$ bits/symbol. Component codes: Turbo code at level 0, 14-error-correcting BCH code at level 1. Simulation results.

$10\log_{10}(E_b/N_0) = 9.1$ dB, this scheme works about 1.0 dB above capacity. From Fig. 29, it can be seen that the competing scheme with the BCH code at level 1 achieves $BER = 10^{-5}$ at $10\log_{10}(E_b/N_0) = 10.15$ dB with a total rate $R = 2.48$ bits/symbol. Here, the capacity $C = 2.48$ bits/symbol is reached for $10\log_{10}(E_b/N_0) = 8.9$ dB. Hence, the MLC scheme using a BCH code at level 1 works about 1.25 dB above capacity, resulting in a loss of about 0.25 dB versus the MLC scheme using turbo codes at the coded levels. This loss of 0.25 dB observed in simulations corresponds well to the loss of 0.15 dB predicted from capacity arguments.

In conclusion, in practice we recommend multilevel coding schemes based on Ungerboeck set partitioning, where soft-decision decoding is employed only at level 0 and hard-decision decoding is applied at higher levels. With this approach, MLC transmission based on a one-dimensional M -ary signal set requires only modest additional complexity (since low-complexity hard-decision decoding is used at higher levels) compared to binary transmission. Thus to approach capacity with bandwidth-efficient digital transmission requires much less decoding complexity per bit than to approach capacity with binary antipodal signaling.

VIII. APPLICATION OF SIGNAL SHAPING

It is well known that signal shaping provides further gain by replacing a uniform signal distribution by a Gaussian-like distribution in order to reduce average transmit power. In many situations, it is easier to obtain shaping gain than to obtain a similar gain by more powerful coding. In order to approach the Shannon limit, shaping is indispensable.

In this section, the combination of MLC and signal shaping is discussed. We find that the achievable shaping gain does not correspond directly to a gain in capacity. The optimum assignment of code rates to the individual levels and optimum sharing of redundancy between coding and shaping is given. The section closes with implementation issues and simulation results.

In view of the results of Sections V and VII, we restrict ourselves to one-dimensional constellations throughout this

section. As mentioned earlier, because shaping algorithms automatically impose proper spherical boundaries in many dimensions on a given signal set, it is sufficient to restrict the discussion to uniformly spaced M -ary one-dimensional constituent constellations.

A. Aspects from Theory

As is well known, optimization of the *a priori* probabilities of signal points is necessary in order to approach channel capacity. In the literature, the gain due to *shaping* is mostly derived only for very large constellations using the continuous approximation. This leads to the following two statements [10].

I: The maximum shaping gain, i.e., the maximum reduction in average transmit power, is given by

$$G_{s,\max} = \frac{\pi c}{6} \hat{=} 1.53 \text{ dB.}$$

But for situations most relevant in practice, using “small” signal sets, the limit of 1.53 dB can never be achieved. In [73] the shaping gain of finite constellations is calculated to be approximately

$$G_s \approx \frac{\pi c}{6}(1 - 2^{-2R_m}) \quad (49)$$

where R_m is the transmission rate per dimension. The term $(1 - 2^{-2R_m})$ can be regarded as a quantization loss due to approximating a continuous distribution by a discrete distribution with entropy R_m bits/dimension. In terms of ASK signaling, it is the ratio of the power of a 2^{R_m} -ary equiprobable constellation to that of a continuous, uniform distribution.

II: Coding gain and shaping gain are separable.

This statement is true only asymptotically for very large constellations. In contrast to other authors (e.g., [10], [33]) we are interested in the analysis of finite constellations. Here, coding and shaping gains interact and cannot simply be added (in decibels). The reason is that, on the one hand, signal power is decreased, leading to a gain. But, on the other hand, a loss in channel output differential entropy $h(Y)$ and, hence, in mutual information $I(A; Y) = h(Y) - h(N)$ is observable, where $h(N)$ denotes the differential entropy of the additive Gaussian noise: $N = Y - A$. Shaping fixes the entropy $H(A)$ of the transmit symbols A instead of the differential entropy of the channel output symbols Y . Thus we have to distinguish between the pure *shaping* (power) *gain* (fixing $H(A)$) and the gain for fixed mutual information $I(A; Y)$, which we denote by *capacity gain*.

Subsequently, the maximum capacity gain in using an optimized channel input distribution is derived (cf. [74]). Consider the capacity $C = \frac{1}{2} \cdot \log_2(1 + 2\frac{E_s}{N_0})$ (bits/dimension) of the AWGN channel with a continuous Gaussian distributed input. To transmit a certain rate $R = C$, the minimum signal-to-noise ratio is thus

$$E_s/N_0 = (2^{2R} - 1)/2. \quad (50)$$

Without loss of generality, we force a uniformly distributed signal to be $f_A(a) = 1$ for $|a| \leq 1/2$ and $f_A(a) = 0$ otherwise.

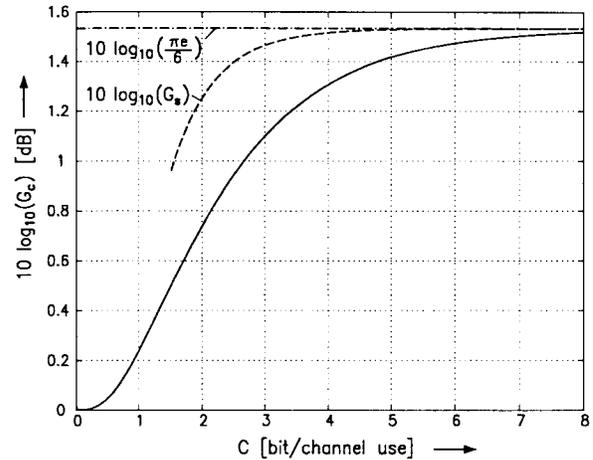


Fig. 30. Capacity gain G_c according to (53) (solid line) and shaping gain G_s for discrete constellation according to (49) (dashed line), respectively, versus capacity.

Since here $\sigma_a^2 = 1/12$, the noise variance is related to the signal-to-noise ratio by $\sigma_n^2 = (24E_s/N_0)^{-1}$. Thus the capacity reads

$$C = C_u\left(\frac{E_s}{N_0}\right) = h(Y) + \frac{1}{2} \cdot \log_2\left(\frac{6}{\pi c} \cdot 2\frac{E_s}{N_0}\right) \quad (51)$$

where $h(Y)$ again denotes the differential entropy of the channel output with density

$$f_Y(y) = \sqrt{\frac{24E_s/N_0}{2\pi}} \cdot \int_{-1/2}^{1/2} \exp\left\{-\frac{(y-x)^2}{2} 24E_s/N_0\right\} dx. \quad (52)$$

Hence, assuming $R \rightarrow C$ and replacing a uniformly distributed channel input signal by a continuous Gaussian distributed one, the maximum capacity gain is given by

$$G_c(C) := \frac{C_u^{-1}(C)}{(2^{2C} - 1)/2} \quad (53)$$

where $C_u^{-1}(C)$ denotes the inverse function of $C_u(E_s/N_0)$.

In Fig. 30, G_c is plotted versus the desired capacity C . Additionally, the maximum shaping gain of discrete constellations (49) is shown.⁷ As one can see, in a wide range the shaping gain is much greater than the gain in capacity. Strictly speaking, the true shaping gain is even greater, because some constellation expansion is necessary to realize coding gain. Hence, the shaping gain of a constellation supporting a certain rate R plus the coding redundancy can be exhausted. Thus the shaping gain curve has to be moved left by the redundancy. Only for asymptotically high rates does the whole shaping gain translate directly to a gain in capacity, approaching the *ultimate shaping gain* of $\frac{\pi c}{6}$. This is because

$$C_u \approx \frac{1}{2} \cdot \log_2\left(\frac{6}{\pi c}(1 + 2E_s/N_0)\right)$$

for high signal-to-noise ratios (cf. [74], [75]). In contrast, for $C \rightarrow 0$ the capacity gain completely vanishes.

⁷The approximation is tight only for $C > 1.5$ bits/dimension.

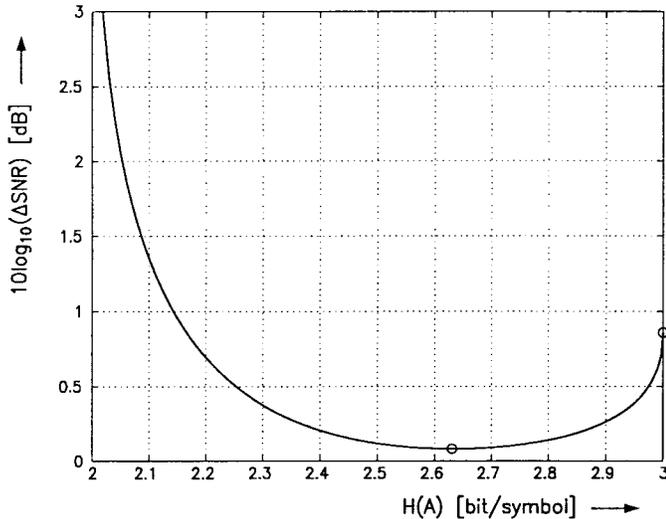


Fig. 31. SNR gap (capacity limit to Shannon limit, see (55)) for rate $R = 2.0$ bits/dimension and 8-ASK constellation as a function of the entropy $H(A)$ of the constellation.

Note that an additional loss appears for discrete constellations compared to the corresponding continuous ones. Hence, for discrete constellations (53) is actually a lower bound.

In order to come close to the optimum, an optimization over the probabilities of the signal points of finite constellations has to be performed, cf., e.g., [76]. This procedure is quite difficult. Therefore, we force the channel input to be (discrete) Gaussian distributed, i.e.,

$$\Pr\{a_m\} = K(\lambda) \cdot e^{-\lambda|a_m|^2}, \quad \lambda \geq 0 \quad (54)$$

where

$$K(\lambda) = \left(\sum_{a_m} e^{-\lambda|a_m|^2} \right)^{-1}$$

normalizes the distribution. This distribution, which maximizes the entropy under an average power constraint, is sometimes called a *Maxwell-Boltzmann distribution* [73]. The parameter λ governs the tradeoff between average power σ_a^2 of signal points and entropy $H(A)$. For $\lambda = 0$, a uniform distribution results, whereas for $\lambda \rightarrow \infty$, only the two signal points closest to the origin remain (M even). From (54), higher dimensional constellations may be simply obtained by taking the Cartesian product. As we will see later, by selecting λ (and hence $H(A)$) properly, the performance of the optimum (not necessarily discrete Gaussian) distribution is approached very closely.

For a given M -ary constellation and target transmission rate $R < \log_2(M)$, this variation of entropy $H(A)$ moreover leads directly to the optimum partitioning of redundancy between coding and shaping. For example, consider the transmission of $C = 2.0$ bits/dimension using an 8-ASK signal constellation. In Fig. 31 the SNR gap to the Shannon limit (normalized SNR)

$$\Delta \text{SNR} = \frac{\text{required } E_b/N_0, \text{ discrete Gaussian distribution}}{2^{2C} - 1} \quad (55)$$

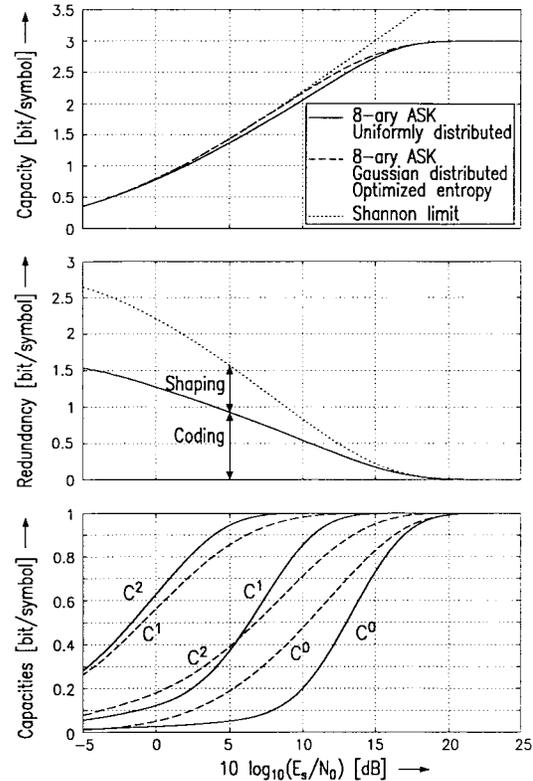


Fig. 32. Top: Shannon limit and the capacity C of 8-ASK (AWGN channel, Ungerboeck set partitioning) with uniform signaling (solid) and with an optimized discrete Gaussian constellation (dashed). Middle: Sharing of coding and shaping redundancy. Bottom: Capacities C^i of the equivalent channels of the corresponding MLC scheme.

is plotted over the entropy $H(A)$. There are three important points: First, for $H(A) = 3$ a uniformly distributed 8-ASK constellation results where only coding is active. Second, as $H(A)$ approaches 2 only signal shaping is used. Because for $C = H(A)$ error-free transmission is in principle only possible for a noiseless channel, the gap here goes to infinity. Third, the minimum is obtained for $H(A) = 2.63$. Thus in the optimum point redundancy of $\log_2(M) - C = 1$ bit has to be divided into 0.63 bit coding redundancy and 0.37 bit shaping redundancy for this specific example. Additionally, it should be noted that for the entropy $H(A)$ in the range of 2.5 to 2.8 bits/symbol ΔSNR differs only slightly. Thus the selection of the optimum entropy is not very sensitive. In the optimum, an additional capacity gain G_c of about 0.78 dB over channel coding only results. Since shaping is done without extra constellation expansion the shaping gain is somewhat smaller than would be possible in principle. But this residual loss in shaping gain is very small; about 0.06 dB at the optimum point, cf. Fig. 31.

Since we have now derived the quasi-optimal distribution of the signal points, the individual rates of the MLC scheme can be calculated and plotted in Fig. 32. Again, an 8-ASK constellation and Ungerboeck set partitioning is assumed. At the top, the capacity C is plotted versus the signal-to-noise ratio E_s/N_0 in decibels. The solid line is valid for uniform signaling, whereas the dashed one assumes an optimized discrete Gaussian constellation. It is important to notice that

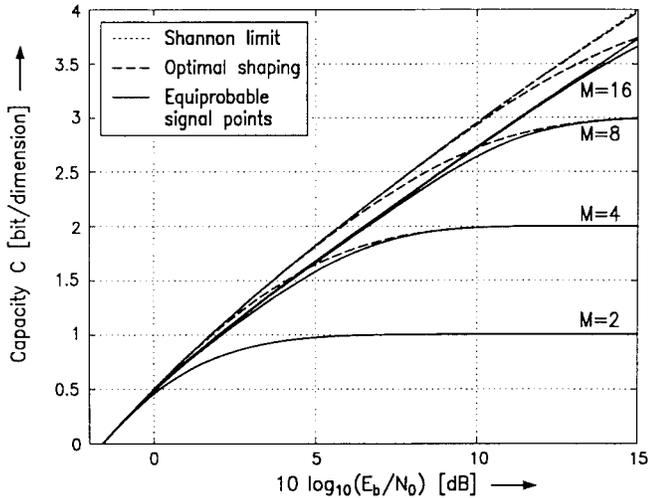


Fig. 33. Capacity of M -ary ASK schemes with and without shaping.

at each point the optimization is performed individually and, hence, different distributions result. Additionally, the Shannon limit is shown (dotted line). As one can see, the Shannon limit can be approached very closely over a wide range.

In addition, an optimization for an arbitrary discrete distribution was done using a modified version of the Blahut–Arimoto algorithm [77]. The resulting optimum is not exactly Gaussian, but the difference from the curve shown in Fig. 32 is invisible (below 0.001 dB).

The plot in the middle displays the optimal sharing of total redundancy $3 - C$ between coding and shaping. As a rule of a thumb, we propose that one bit total redundancy should be divided into $2/3$ bit coding redundancy and $1/3$ bit shaping redundancy.

On the bottom, the capacities C^i of the equivalent channels i of the MLC scheme according to (7) are shown. Again, the solid lines correspond to uniform signaling and the dashed lines hold for the optimized Gaussian distributions. It is important to observe that rate design completely changes when shaping is active. In particular, the rate of the highest level decreases strongly. The reason is that this level has to carry the entire shaping redundancy. This observation leads directly to a simple construction of MLC schemes with signal shaping, see Section VIII-B.

Finally, in Fig. 33 this optimization is performed for $M = 2, 4, 8, 16$, and 32-ary ASK constellations. (Obviously no shaping gain can be achieved for 2-ASK.) The gain increases as the size of the constellation increases (cf. Fig. 30).

Ungerboeck stated [4] that by doubling the size of a QAM signal constellation, i.e., by introducing 0.5 bit redundancy per dimension, almost the entire coding gain can be realized; going to larger constellations is not rewarding. Less total redundancy than 0.5 bit/dimension causes an inevitable loss from the maximum coding gain for equiprobable signal points. This statement has now to be formulated more precisely: as long as no signal shaping is applied. For combined coding and shaping no improvement can be gained beyond doubling the number of signal points per dimension, i.e., by introducing 1 bit redundancy per dimension. Here, the Shannon limit curve is

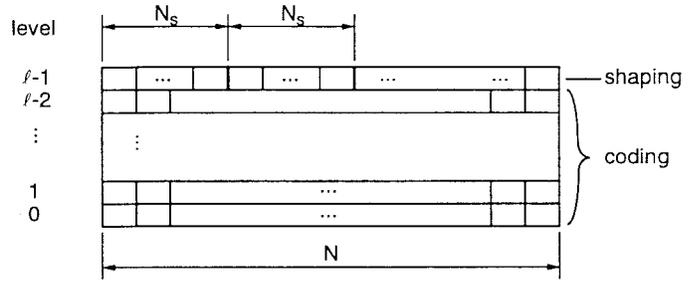


Fig. 34. Application of trellis shaping to MLC.

approached very closely by applying the usual ASK or QAM constellations. Reducing total redundancy to 0.5 bit/dimension already causes an inevitable loss of approximately 0.5 dB to the Shannon limit curve, although even here, the SNR gain for shaping and coding is already much greater than for coding solely. For $E_b/N_0 \rightarrow \infty$ shaping becomes inactive and the curves merge.

As our aim is to approach capacity, we are not concerned with the constellation expansion ratio. In these examples we have chosen 1 bit total redundancy, i.e., approximately $1/3$ bits/dimension shaping redundancy. But in contrast to coded uniform constellations, shaping is here done without further constellation expansion.

B. Implementation and Simulation Results

We now apply the theoretical considerations of the previous section to practical schemes. A shaping algorithm has to generate a distribution of signal points approximating the theoretical Maxwell–Boltzmann distribution while preserving the optimum entropy. In principle all shaping techniques, e.g., *shell mapping* [78], [79] or *trellis shaping* [80] can be combined with an MLC scheme. Here, we prefer trellis shaping, because shell mapping is only suited for very large constellations, partitioned in a large enough number of two-dimensional shells. By contrast, since trellis shaping takes the lower levels into account (without modifying them), only a small portion of data has to be scrambled with shaping redundancy. Here, we will not describe trellis shaping in detail (see [80]), but we give a possible approach to combining MLC with shaping. The idea is sketched in Fig. 34. The lower $\ell - 1$ levels are assumed to be coded using block codes of length N with appropriate rates. Only the highest level to which no coding is applied is involved in shaping. This approach preserves the MLC philosophy, and coding and shaping are decoupled. In spite of this separation, in terms of rate design and calculation of gains, coding and shaping still interacts. N_s consecutive symbols form one shaping step (N does not need to be a multiple of N_s). Shaping redundancy can be adjusted properly by combining an appropriate number N_s of modulation intervals into a shaping step. Because a rate- $1/N_s$ shaping convolutional code is used, $N_s - 1$ data bits are scrambled with one shaping bit, resulting in an N_s -dimensional shaping scheme. This construction can be interpreted as shaping based on an N_s -dimensional lattice, where the dimension N_s is equal to the number of bits selecting one of 2^{N_s} regions (cf. [80]). A generalization would

be the application of rate- N_r/N_s , $N_r < N_s$, shaping codes in order to get a finer granularity of rate.

Simulations for a multilevel coded 64-QAM constellation, where MLC is based on a 8-ASK constellation with turbo codes of large block length $N = 40000$ and total rate $R = 2.0$ bits/dimension were performed. Three-dimensional trellis shaping is applied at the highest level, i.e., $1/3$ bit shaping plus $2/3$ bit coding redundancy per dimension is spent as discussed above. Using this scheme we achieve $\text{BER} = 10^{-5}$ within 1 dB of the Shannon limit (see Fig. 10). Notice that, while using shaping, the nonuniform distribution of signal points has to be taken into account in the decoding procedure, i.e., maximum-likelihood decoding has to be replaced by maximum *a posteriori* decoding. Further simulation results can be found in [81].

In conclusion, these rules lead to very powerful transmission systems bridging the gap between signaling with uniformly distributed signal points and the Shannon limit.

Similar results can be derived if hard-decision decoding is used at certain levels. Here also an optimal tradeoff between coding and shaping redundancy can be found. But because the poorer performance of channel coding, it is advantageous to spend more redundancy for shaping and achieve a higher shaping gain. For 8-ary signaling with strategy SHH (see Section VII) and a target rate of 2.0 bits/dimension, a two-dimensional shaping scheme with $H(A) = 2.5$ is close to the optimum, cf. [82].

IX. CONCLUSIONS

The concept of equivalent channels for the individual coding levels of an MLC scheme establishes a basis to derive tools for the analysis and design of coded modulation schemes. The key point for a power-efficient design of MLC schemes is the proper assignment of individual rates to the component codes. If the individual rates are chosen to be equal to the capacities of the equivalent channels, the capacity of the underlying 2^ℓ -ary modulation scheme is achieved by an ℓ -level MLC scheme together with the suboptimum MSD for arbitrary *a priori* probabilities and for an arbitrary labeling of signal points. Thus the problem of channel coding can be solved in principle in an optimum way by employing binary codes in an MLC approach.

There exists a wide region for the individual code rates within which capacity is achievable. But, except for the vertices of this region, which correspond to the capacities of the equivalent channels, one must replace relatively low-complexity MSD by extremely complex MLD. In particular, the individual rates of coded modulation schemes which are designed by the BDR differ substantially from the capacities of the equivalent channels, leading to an unavoidable performance loss when using MSD. In this case capacity is only achievable by overall MLD.

Various information-theoretic parameters (capacity, coding exponent, and cutoff rate) of the equivalent channels of an MLC scheme lead to various design rules for coded modulation schemes based, e.g., on block codes with given length or R_0 -approaching codes as component codes. Simulation results for the AWGN channel show that transmission schemes

designed by these rules exhibit power and bandwidth efficiency close to the Shannon limit. In practice these rules do not lead to very different rate designs.

From the coding exponent and the cutoff rate of the equivalent channel of an MLC scheme, the optimum dimensionality of the constituent signal set operating on the AWGN channel was derived, assuming binary codes at each level. For a fixed data delay it is more efficient in power and complexity to base MLC schemes on a one-dimensional constellation combined with as long block codes as possible, instead of increasing the dimensionality of the constellation. In contrast, using binary convolutional codes with equal constraint lengths, it is more efficient to use MLC in combination with a multidimensional constellation.

As shown in Section III, the capacity of MLC schemes is independent of the particular labeling. For finite code length, the labeling introduced by Ungerboeck and Imai, leads to the most power-efficient schemes. However, two MLC schemes based on block and mixed partitioning were also presented which are suited for softly degrading transmission schemes and for a reduction of hardware complexity, respectively. With hard-decision decoding, low individual code rates lead to a significant performance loss and thus should be avoided. With an appropriate labeling strategy, rates can be assigned much more uniformly. Additionally, combining several levels into a single one and applying a sufficiently large block code with subsequent bit interleaving was discussed. This BICM approach using Gray labeling of signal points seems to be a relatively low-complexity attractive alternative approach to coded modulation. On the other hand, convolutional coding is not suited to BICM.

Employment of hard-decision decoding at several coding levels is an efficient method to save complexity in coded modulation schemes. With Ungerboeck labeling of signal points, the performance loss compared to soft-decision decoding (in terms of capacity as well as in simulations) is only about 0.2 dB for 8-ASK transmission over the AWGN channel, when hard-decision decoding is employed at all but the lowest level. Since in general the complexity of hard-decision decoding is substantially lower than that of soft-decision decoding, we observe that power-efficient coded 2^ℓ -ary, $\ell > 1$, modulation requires only slightly more complexity than coded binary transmission. If hard-decision decoding is employed at all levels, then an MLC approach with ℓ individual binary error-correcting codes is recommended. Since multistage decoding exploits the soft channel output to some extent, it promises better performance than a single binary code adapted to the 2^ℓ -ary modulation scheme by Gray mapping.

In combination with channel coding, signal shaping provides further gain by reducing average transmit power. Since for finite constellations coding and shaping are not separable, their interaction has to be taken into account when designing an MLC scheme. Assuming discrete Gaussian constellations, the key design point is the optimum sharing of redundancy between coding and shaping. It turns out that a redundancy of 1 bit/dimension is sufficient to approach to the Shannon limit very closely. Less redundancy results in an unavoidable performance loss. In order to achieve significant shaping gains

in practice, one must take into account these rate design principles. Using coding *and* shaping results in completely different rate designs than without shaping. Moreover, maximum *a posteriori* decoding should be used instead of maximum-likelihood decoding.

In the Appendix, a tight upper bound on the error probability for binary transmission over the equivalent channels of an MLC scheme is derived. Here finite constellations and boundary effects are taken into account leading to results that are more useful in practice.

APPENDIX

REVIEW OF UPPER BOUNDS ON THE ERROR PROBABILITY OF MULTILEVEL CODED TRANSMISSION

Selected previous work on the performance evaluation of coded modulation schemes over the AWGN channel includes the derivation of the following.

- Upper bounds on the word-error probability of modulation schemes based on binary lattices and M -ary PSK constellations, respectively, with MLC where overall maximum-likelihood decoding and MSD are considered [83] and [52].
- The probability of decoding error when overall maximum-likelihood decoding is used [51].
- The Euclidean distance enumerator of MLC [35].
- The bit-error probability for 8-PSK with MLC and MSD when error propagation is taken into account [39].
- The probability of decoding error at level i as a function of the Chernoff bounding parameter for M -ary PSK, 4-QAM and QAM with an infinite number of signal points [84], [85].
- A minimum-distance bound on the word error probability for block-coded modulation schemes using MSD when error propagation is taken into account [86].
- A lower bound on the symbol error probability for lattice codes [87].
- An upper bound on the word error probability for lattice codes [88].

In this appendix, we sketch the derivation of a tight upper bound on the word error probability for $M = 2^\ell$ -ary transmission using an MLC scheme with linear binary component codes. In [83] a derivation is given for component codes based on lattice constellations that neglects boundary effects. Here, the results are generalized to constellations with a finite number of signal points and almost arbitrary labelings. Thus boundary effects are included. Note that the main results are in [35].

Let p_w denote the probability that the word of estimated source symbols $\hat{\mathbf{q}} = [\hat{\mathbf{q}}^0, \hat{\mathbf{q}}^1, \dots, \hat{\mathbf{q}}^{\ell-1}]$ is in error and let $p_{w,i}$ denote the probability that the estimated component data word $\hat{\mathbf{q}}^i$ contains errors, provided the correct subset is known at each decoding stage, i.e., error-free decisions at lower levels, or parallel decoding of the individual levels (PDL) are assumed. Using the union bound principle, the error probability p_w is strictly upperbounded by $\sum_{i=0}^{\ell-1} p_{w,i}$. Consequently, p_w can be upper-bounded by bounding the individual error rates $p_{w,i}$.

If the mean Euclidean distance enumerator

$$N_i(Z) = \sum_{d \geq d_{C^i}} n_i(d) \cdot Z^{d^2}$$

of component code C^i in Euclidean space is known (d_{C^i} : minimum Euclidean distance for coded transmission at level i ; $n_i(d)$: average number of codewords at Euclidean distance d ; Z : indeterminate) the usual union bound for the error rate $p_{w,i}$ can be applied, cf., e.g., [89]. Notice that the upper bound for the error probability given by the union bound is rather tight for $p_{w,i} < 10^{-3}$ only for codes with rates smaller than or close to the cutoff rate.

For low-to-moderate SNR, where this bound may not be sufficiently tight, an improved bounding technique was proposed by Hughes [90] and refined by Herzberg and Poltyrev [91], [92]. As the Hughes–Herzberg–Poltyrev (HHP) bound is also based on the union bound (cf. [83]), the distance enumerator $N_i(Z)$ of codewords of the component code C^i in Euclidean space is a key parameter, which we now derive.

A. Euclidean Distance Enumerator of Multilevel Component Codes

For brevity, we restrict ourselves to equiprobable signal points and a regularly partitioned constellation; i.e., all subsets at one partitioning level are congruent. The Euclidean distance enumerator $N_i(Z)$ of the Euclidean space representation of the multilevel component code C^i of length N is given by

$$N_i(Z) = [B_i^{(0)}(Z)]^N \cdot W_i\left(D = \frac{B_i^{(1)}(Z)}{B_i^{(0)}(Z)}\right) \quad (56)$$

where

$$W_i(D) = \sum_{\delta \geq \delta_i} w_i(\delta) \cdot D^\delta \quad (57)$$

denotes the weight enumerator of the linear binary component code C^i (minimum Hamming distance δ_i).

$$B_i^{(1)}(Z) = E_{a_k \in \mathcal{A}(x^0 \dots x^{i-1})} \left\{ \sum_{a_j \in \mathcal{A}(x^0 \dots x^{i-1})} Z^{|a_k - a_j|^2} \right\} \quad (58)$$

is the averaged different-subset constellation enumerator at level i and

$$B_i^{(0)}(Z) = E_{a_k \in \mathcal{A}(x^0 \dots x^{i-1})} \left\{ \sum_{a_j \in \mathcal{A}(x^0 \dots x^{i-1})} Z^{|a_k - a_j|^2} \right\} \quad (59)$$

is the averaged same-subset constellation enumerator at level i .

The bound on the error probability can be tightened by using a *relevant* Euclidean distance enumerator

$$N_{i,\text{rel}}(Z) = W_i(D = B_{i,\text{rel}}(Z)) \quad (60)$$

with

$$B_{i,\text{rel}}(Z) = E_{a_k \in \mathbf{A}(x^0 \dots x^{i-1})} \left\{ \sum_{a_j \in \mathbf{V}(a_k, \mathbf{A}(x^0 \dots x^{i-1}))} Z^{|a_k - a_j|^2} \right\} \quad (61)$$

where $\mathbf{V}(a_k, \mathbf{A}(x^0 \dots x^{i-1}))$ denotes the set of signal points which determine the Voronoi region of a_k with respect to $\mathbf{A}(x^0 \dots x^{i-1})$.

Sketch of Proof: For simplicity of notation and without loss of generality subsequently the considerations are restricted to the level $i = 0$. We start with a simple two word linear code $\mathcal{C}^0 = \{\mathbf{c}_0^0 = \mathbf{0}, \mathbf{c}_1^0\}$ of length N , where the weight of \mathbf{c}_1^0 is δ .

If we first assume the all-zero codeword \mathbf{c}_0^0 to be transmitted, the signal points representing the complementary binary symbol $x^0 = 1$ have to be considered to upper-bound the error probability by the union bound. In this case the averaged different-subset constellation enumerator $B_{0,0 \rightarrow 1}^{(1)}(Z)$ at level 0 is given by the mean Euclidean distance enumerator counting each signal point $a_k \in \mathbf{A}(0)$ representing $x^0 = 0$ with respect to each signal point $a_j \in \mathbf{A}(1)$ representing $x^0 = 1$

$$B_{0,0 \rightarrow 1}^{(1)}(Z) = E_{a_k \in \mathbf{A}(0)} \left\{ \sum_{a_j \in \mathbf{A}(1)} Z^{|a_k - a_j|^2} \right\}. \quad (62)$$

By the same argument, the constellation enumerator $B_0^{(1)}(Z)$ in the case of the transmission of \mathbf{c}_1^0 is

$$B_{0,1 \rightarrow 0}^{(1)}(Z) = E_{a_j \in \mathbf{A}(1)} \left\{ \sum_{a_k \in \mathbf{A}(0)} Z^{|a_k - a_j|^2} \right\}. \quad (63)$$

For regular partitions, $B_{0,0 \rightarrow 1}^{(1)}(Z) = B_{0,1 \rightarrow 0}^{(1)}(Z) := B_0^{(1)}(Z)$ holds; the average constellation enumerator $B_0^{(1)}(Z)$ is independent of the actual transmitted codeword. The same derivation applies to the averaged same-subset constellation enumerator $B_0^{(0)}(Z)$, which is required to consider all possible Euclidean space representations of those positions, which are identical in both codewords.

Due to the orthogonality of the signal space, the Euclidean distance enumerator of the component code \mathcal{C}^0 with respect to all possible Euclidean-space codeword representations is given by the product of the different-subset constellation enumerators for all differing positions times the product of the same-subset constellation enumerators for all equal positions (see, e.g., [35]), hence

$$\begin{aligned} N_0(Z) &= [B_0^{(0)}(Z)]^{N-\delta} \cdot [B_0^{(1)}(Z)]^\delta \\ &= [B_0^{(0)}(Z)]^N \cdot \left[\frac{B_0^{(1)}(Z)}{B_0^{(0)}(Z)} \right]^\delta. \end{aligned} \quad (64)$$

Applying this result to a general linear binary code \mathcal{C}^0 with weight enumerator $W_{\mathcal{C}^0}(D)$ yields (56).

In order to calculate an upper bound on the error probability, the entire Euclidean distance enumerator $N_i(Z)$ of the code is not required. Here, due to the union bound, many error events are counted more than once. The number of terms comprising

the constellation enumerator $B_0^{(1)}(Z)$ can be reduced in order to tighten the error probability bound if the following fact (given in [24] and also used in [83]) is taken into account: The error probability $p_{w,i}$ is still upper-bounded by the union bound when only adjacent signal points are counted, i.e., those points determining the walls of the Voronoi region of the considered signal point. Thus the resulting *relevant* constellation enumerator in (61) is valid if $x^0 = 0$ is assumed to be transmitted. Additionally, since signal points representing the same symbol are irrelevant for determining the Voronoi cells, the same-subset enumerator equals one. In the case of regular partitions, (61) also holds for transmission of $x^0 = 1$, namely

$$\begin{aligned} B_{0,\text{rel}}(Z) &= E_{a_j \in \mathbf{A}(1)} \left\{ \sum_{a_k \in \mathbf{V}(a_j, \mathbf{A}(0))} Z^{|a_k - a_j|^2} \right\} \\ &= E_{a_k \in \mathbf{A}(0)} \left\{ \sum_{a_j \in \mathbf{V}(a_k, \mathbf{A}(1))} Z^{|a_k - a_j|^2} \right\}. \end{aligned} \quad (65)$$

□

The derivation of the upper bound on the error probability of multilevel codes on the AWGN channel is sketched for regular partitions. This symmetry property of a Euclidean space code was already exploited by Zehavi and Wolf [93] and in a weaker form by Ungerboeck in his code search criteria [4]; Forney called it ‘‘Ungerboeck–Zehavi–Wolf symmetry’’ [94]. However, the calculation can still be applied to nonregular partitions if the formulas are adapted to the time variance of the equivalent channel in a similar way as is done for the Rayleigh fading channel, see, e.g., [95].

Additionally, in almost all cases relevant in practice the results also hold for nonuniform, i.e., shaped, transmission.

Analogous to the derivation of an upper bound on the word error probability $p_{w,i}$, an upper bound on the bit-error probability $p_{b,i}$ can be obtained by the modified weight enumerator as usual [89].

If error propagation in MSD is neglected, the bit-error probability p_b for multilevel coded transmission is given by

$$p_b = \sum_{i=0}^{\ell-1} \frac{R^i}{R} p_{b,i}$$

cf. [39], where $p_{b,i}$ denotes the bit-error probability for decoding at level i when error-free decisions are assumed at decoding stages of lower levels.

B. Examples

Two special cases of the calculation of the Euclidean distance enumerator that are of practical interest are discussed in the following examples.

Example 1: The use of a linear (n, k, δ) code \mathcal{C}^0 with minimum Hamming distance δ at level 0 of an equiprobable 8-ASK constellation $\mathbf{A} = \{\pm 1, \pm 3, \pm 5, \pm 7\}$ with natural labeling is investigated (regular partitions). The relevant constellation

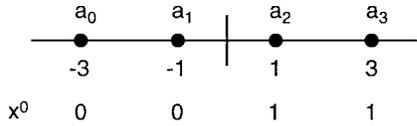


Fig. 35. Signal constellation at level 0 of a 4-ASK constellation with block partitioning.

enumerator $B_0(Z)$ is calculated according to (61)

$$B_0(Z) = \frac{1}{4}2Z^4 + \frac{1}{4}2Z^4 + \frac{1}{4}2Z^4 + \frac{1}{4}Z^4 = \frac{7}{4}Z^4. \quad (66)$$

Hence, the relevant Euclidean distance enumerator of the code \mathcal{C}^0 in this example is given by

$$N_{0,\text{rel}}(Z) = w_0(\delta) \cdot \left(\frac{7}{4}\right)^\delta Z^{4\delta} + \dots \quad (67)$$

From (67) one can see that the minimum-distance error coefficient of the code \mathcal{C}^0 is dramatically increased by a factor of $(\frac{7}{4})^\delta$ due to the multiple representation of binary symbols in the signal constellation at level 0. Even for relatively small Hamming distances δ , e.g., $\delta = 15$, an increase of the effective error coefficient by $(\frac{7}{4})^{15} \approx 4422$ results. In general, the factor is given by the δ th power of the average number of nearest neighbors in the constituent signal set.

Example 2: The 4-ASK constellation $\mathbf{A} = \{\pm 1, \pm 3\}$ with block partitioning for transmission at level 0 of an MLC scheme as shown in Fig. 35 is investigated. Equiprobable signal points are assumed. The relevant constellation enumerator $B_0(Z)$ reads:

$$B_0(Z) = \frac{1}{2}(Z^4 + Z^{16}). \quad (68)$$

Here not only the minimum squared Euclidean distance of 4 contributes to the relevant constellation enumerator but also the additional distance 16, because walls of the Voronoi region are at different distances for signal points a_1 and a_2 .

Remarkably, as δ increases the minimum-distance error coefficient $(\frac{1}{2})^\delta$ decreases. Thus in case of block labeling asymptotically the minimum distance is of almost no interest. This again emphasizes our conclusion that minimum Euclidean distance is not the most appropriate design criterion.

ACKNOWLEDGMENT

The authors are greatly indebted to the anonymous reviewers for their very careful reading and detailed critique of an earlier version of the manuscript. Their insights and suggestions greatly improved the paper. The authors would also like to thank Rolf Weber for performing some of the calculations in Section VII.

REFERENCES

- [1] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.
- [2] J. L. Massey, "Coding and modulation in digital communications," in *1974 Int. Zürich Seminar on Digital Communications* (Zürich, Switzerland, Mar. 1974).
- [3] G. Ungerboeck and I. Csajka, "On improving data-link performance by increasing channel alphabet and introducing sequence coding," in *Proc. IEEE Int. Symp. Information Theory (ISIT)* (Ronneby, Sweden, June 1976).
- [4] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.
- [5] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 371–377, May 1977.
- [6] G. Ungerboeck, "Trellis coded modulation with redundant signal sets, part I," *IEEE Commun. Mag.*, vol. 25, pp. 5–11, Feb. 1987.
- [7] ———, "Trellis coded modulation with redundant signal sets, part II," *IEEE Commun. Mag.*, vol. 25, pp. 12–21, Feb. 1987.
- [8] A. R. Calderbank and N. J. A. Sloane, "Four-dimensional modulation with an eight-state trellis code," *AT&T Tech. J.*, vol. 64, pp. 1005–1018, May–June 1985.
- [9] ———, "An eight-dimensional trellis code," *Proc. IEEE*, vol. 74, pp. 757–759, May 1986.
- [10] G. D. Forney Jr. and L.-F. Wei, "Multidimensional constellations—Part I: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 877–892, Aug. 1989.
- [11] G. D. Forney Jr., "Multidimensional constellations—Part II: Voronoi constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 941–958, Aug. 1989.
- [12] S. S. Pietrobon, R. H. Deng, A. Lafaneché, G. Ungerboeck, and D. J. Costello, "Trellis-coded multidimensional phase modulation," *IEEE Trans. Inform. Theory*, vol. 36, pp. 63–89, 1990.
- [13] S. S. Pietrobon and D. J. Costello Jr., "Trellis coding with multidimensional QAM signal sets," *IEEE Trans. Inform. Theory*, vol. 39, pp. 325–336, Mar. 1993.
- [14] A. Viterbi, J. K. Wolf, E. Zehavi, and R. Padovani, "A pragmatic approach to trellis coded modulation," *IEEE Commun. Mag.*, vol. 27, pp. 11–19, July 1989.
- [15] J. K. Wolf and E. Zehavi, " P^2 codes: Pragmatic trellis coding utilizing punctured convolutional codes," *IEEE Commun. Mag.*, vol. 33, pp. 94–99, Feb. 1995.
- [16] J. Kim and G. J. Pottie, "On punctured trellis-coded modulation," *IEEE Trans. Inform. Theory*, vol. 42, pp. 627–636, Mar. 1996.
- [17] L.-F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 483–501, July 1987.
- [18] ———, "Rotationally invariant trellis-coded modulation with multidimensional M-PSK," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 1281–1295, Dec. 1989.
- [19] W. Henkel, "Conditions for 90° phase-invariant block-coded QAM," *IEEE Trans. Commun.*, vol. 41, pp. 524–527, Apr. 1993.
- [20] M. D. Trott, S. Benedetto, R. Garello, and M. Mondin, "Rotational invariance of trellis codes—Part I: Encoders and precoders," *IEEE Trans. Inform. Theory*, vol. 42, pp. 751–765, May 1996.
- [21] S. Benedetto, R. Garello, M. Mondin, and M. D. Trott, "Rotational invariance of trellis codes—Part II: Group codes and decoders," *IEEE Trans. Inform. Theory*, vol. 42, pp. 766–778, May 1996.
- [22] A. R. Calderbank, T. A. Lee, and J. E. Mazo, "Baseband trellis codes with a spectral null at zero," *IEEE Trans. Inform. Theory*, vol. 34, pp. 425–434, 1988.
- [23] G. D. Forney Jr. and A. R. Calderbank, "Coset codes for partial response channels; or coset codes with spectral nulls," *IEEE Trans. Inform. Theory*, vol. 35, pp. 925–943, 1989.
- [24] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer Verlag, 1988.
- [25] T. Wörz and J. Hagenauer, "Iterative decoding for multilevel codes using reliability information," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)* (Orlando, FL, Dec. 1992).
- [26] G. J. Pottie and D. P. Taylor, "Multilevel codes based on partitioning," *IEEE Trans. Inform. Theory*, vol. 35, pp. 87–98, Jan. 1989.
- [27] A. R. Calderbank, "Multilevel codes and multistage decoding," *IEEE Trans. Commun.*, vol. 37, pp. 222–229, Mar. 1989.
- [28] G. D. Forney Jr., "Coset codes—Part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, Sept. 1988.
- [29] ———, "Coset codes—Part II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1152–1187, Sept. 1988.
- [30] A. R. Calderbank and N. J. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 177–195, 1987.
- [31] R. de Buda, "Some optimal codes have structure," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 893–899, Aug. 1989.
- [32] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 273–278, Jan. 1998.
- [33] G. D. Forney Jr., "Approaching AWGN channel capacity with coset codes and multilevel coset codes," *IEEE Trans. Inform. Theory*, 1997,

- submitted for publication.
- [34] J. Huber and U. Wachsmann, "Capacities of equivalent channels in multilevel coding schemes," *Electron. Lett.*, vol. 30, pp. 557–558, Mar. 1994.
- [35] J. Huber, "Multilevel codes: Distance profiles and channel capacity," in *ITG-Fachbericht 130, Conf. Rec.* (München, Germany, Oct. 1994), pp. 305–319.
- [36] J. Huber and U. Wachsmann, "Design of multilevel codes," in *Proc. IEEE Inform. Theory Workshop (ITW)* (Rydzyzna, Poland, June 1995), paper 4.6.
- [37] U. Wachsmann and J. Huber, "Power and bandwidth efficient digital communication using Turbo codes in multilevel codes," *Europ. Trans. Telecommun. (ETT)*, vol. 6, pp. 557–567, Sept.–Oct. 1995.
- [38] Y. Kofman, E. Zehavi, and S. Shamai (Shitz), "Analysis of a multilevel coded modulation system," in *1990 Bilkent Int. Conf. New Trends in Communications, Control, and Signal Processing, Conf. Rec.* (Ankara, Turkey, July 1990), pp. 376–382.
- [39] ———, "Performance analysis of a multilevel coded modulation system," *IEEE Trans. Commun.*, vol. 42, pp. 299–312, 1994.
- [40] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [41] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [42] G. Caire, G. Taricco, and E. Biglieri, "Capacity of bit-interleaved channels," *Electron. Lett.*, vol. 32, pp. 1060–1061, June 1996.
- [43] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. Commun. (ICC)* (Geneva, Switzerland, May 1993), pp. 1064–1070.
- [44] C. Berrou and A. Glavieux, "Near optimum limit error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, pp. 1261–1271, Oct. 1996.
- [45] U. Wachsmann and J. Huber, "Conditions on the optimality of multilevel codes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (Ulm, Germany, July 1997).
- [46] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124–142, Mar. 1985.
- [47] V. V. Ginzburg, "Multidimensional signals for a continuous channel," *Probl. Inform. Transm.*, vol. 23, pp. 20–34, 1984; Translation from *Probl. Pered. Inform.*, pp. 28–46, 1984.
- [48] S. I. Sayegh, "A class of optimum block codes in signal space," *IEEE Trans. Commun.*, vol. COM-34, pp. 1043–1045, Oct. 1986.
- [49] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1978.
- [50] U. Wachsmann, "Coded modulation: Theoretical concepts and practical design rules," Ph.D. dissertation, Universität Erlangen-Nürnberg, Erlangen, Germany, May 1998.
- [51] T. Kasami, T. Takata, T. Fujiwara, and S. Lin, "On multilevel block coded modulation codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 965–975, July 1991.
- [52] 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, pp. 1204–1218, July 1993.
- [53] T. Wörz and J. Hagenauer, "Multistage coding and decoding," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)* (San Diego, CA, Dec. 1990), pp. 501.1.1–501.1.6.
- [54] ———, "Decoding of M-PSK-multilevel codes," *Europ. Trans. Telecommun. (ETT)*, vol. 4, pp. 65–74, May–June 1993.
- [55] J. Persson, "Multilevel coding based on convolutional codes," Ph.D. dissertation, Lund Univ., Lund, Sweden, June 1996.
- [56] P. Robertson, "Illuminating the structure of code and decoder for parallel concatenated recursive systematic (turbo) codes," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM), Conf. Rec.* (San Francisco, CA, Nov./Dec. 1994), pp. 1298–1304.
- [57] J. Huber and U. Wachsmann, "Power efficient rate design for multilevel codes with finite blocklength," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)* (Whistler, BC, Canada, Sept. 1995).
- [58] P. Robertson and T. Wörz, "Bandwidth-efficient Turbo trellis-coded modulation using punctured component codes," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 206–218, Feb. 1998.
- [59] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Parallel concatenated trellis coded modulation," in *Proc. IEEE Int. Conf. Commun. (ICC)* (Dallas, June 1996), pp. 974–978.
- [60] R. Fischer, J. Huber, and U. Wachsmann, "Multilevel coding: Aspects from information theory," in *Proc. CTMC at IEEE Global Telecommun. Conf. (GLOBECOM'96)* (London, U.K., Nov. 1996), pp. 26–30.
- [61] W. Webb and L. Hanzo, *Quadrature Amplitude Modulation*. New York: IEEE Press and Pentech Press, 1994.
- [62] J. Huber and U. Wachsmann, "On set partitioning strategies for multilevel coded modulation schemes," in *Proc. Mediterranean Workshop on Coding and Information Integrity* (Palma de Mallorca, Spain, Feb.–Mar. 1996).
- [63] R. Fischer, J. Huber, and U. Wachsmann, "Multilevel coding: Aspects from information theory," in *Proc. IEEE Inform. Theory Workshop (ITW'96)* (Haifa, Israel, June 1996), p. 9.
- [64] K. Ramchandran, A. Ortega, K. M. Uz, and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source/channel coding," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 6–21, Jan. 1993.
- [65] U. Horn, B. Girod, and B. Belzer, "Scalable video coding for multimedia applications and robust transmission over wireless channels," in *Proc. 7th Int. Workshop on Packet Video* (Brisbane, Australia, Mar. 1996), pp. 43–48.
- [66] B. Girod, U. Horn, and B. Belzer, "Scalable video coding with multiscale motion compensation and unequal error protection," in *Multimedia Communications and Video Coding*, Y. Wang, S. Panwar, S.-P. Kim, and H. L. Bertoni, Eds. New York: Plenum, Oct. 1996, pp. 475–482.
- [67] R. H. Morelos-Zaragoza, O. Y. Takeshita, H. Imai, M. P. C. Fossorier, and S. Lin, "Coded modulation for satellite broadcasting," in *Proc. CTMC at IEEE Global Telecommun. Conf. (GLOBECOM'96)* (London, U.K., Nov. 1996), pp. 31–35.
- [68] K. Fazel and M. J. Ruf, "Combined multilevel coding and multiresolution modulation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 1993, pp. 1081–1085.
- [69] D. Schill and J. Huber, "On hierarchical signal constellations for the Gaussian broadcast channel," in *Proc. ICT'98* (Chalkidiki, Greece, 1998).
- [70] ETS 400 744 ETSI, "Digital video broadcasting (DVB); framing structure, channel coding, and modulation for digital terrestrial television," Draft, 1997.
- [71] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–884, May 1992.
- [72] E. L. Cusack, "Error control codes for QAM signaling," *Electron. Lett.*, vol. 20, pp. 62–63, Jan. 1984.
- [73] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 913–929, May 1993.
- [74] L. H. Ozarow and A. D. Wyner, "On the capacity of the Gaussian channel with a finite number of input levels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1426–1428, Nov. 1990.
- [75] F.-W. Sun and H. C. A. van Tilborg, "Approaching capacity by equiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1714–1716, Sept. 1993.
- [76] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inform. Theory*, vol. 36, pp. 726–740, July 1990.
- [77] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, 1972.
- [78] R. Laroia, N. Farvardin, and S. A. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1044–1056, July 1994.
- [79] M. V. Eyuboğlu, G. D. Forney Jr., P. Dong, and G. Long, "Advanced modulation techniques for V.fast," *Europ. Trans. Telecommun. (ETT)*, vol. 4, pp. 243–256, May–June 1993.
- [80] G. D. Forney Jr., "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38, pp. 281–300, Mar. 1992.
- [81] R. Fischer, U. Wachsmann, and J. Huber, "On the combination of multilevel coding and signal shaping," in *ITG-Fachbericht: Codierung für Quelle, Kanal und Übertragung* (Aachen, Germany, Mar. 1998), pp. 273–278.
- [82] R. Weber, "Einsatz von fehlerkorrigierenden block-codes in der codierten modulation," *Diploma Thesis*, Lehrstuhl für Nachrichtentechnik, Univ. Erlangen-Nürnberg, Erlangen, Germany, 1997, in German.
- [83] E. Biglieri, A. Sandri, and A. Spalvieri, "Computing upper bounds to error probability of coded modulation schemes," *IEEE Trans. Commun.*, vol. 44, pp. 786–790, July 1996.
- [84] K. Engdahl and K. Sh. Zigangirov, "On the calculation of the error probability for a multilevel modulation scheme using QAM-signalling," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1612–1620, 1998.
- [85] ———, "On the calculation of the error probability for a multilevel modulation scheme using PSK-signalling," submitted to *Europ. Trans. Telecommun. (ETT)*, 1997.
- [86] 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*, vol. 43, pp. 373–385, Jan. 1997.
- [87] V. Tarokh, A. Vardy, and K. Zeger, "Universal bound on the per-

- formance of lattice codes," *IEEE Trans. Inform. Theory*, vol. 45, pp. 670–681, Mar. 1999.
- [88] H. Magalhães de Oliveira and G. Battail, "Performance of lattice codes over the Gaussian channel," *Ann. Telecommun.*, vol. 47, pp. 293–305, 1992.
- [89] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [90] B. Hughes, "On the error probability of signals in additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 37, pp. 151–155, Jan. 1991.
- [91] H. Herzberg and G. Poltyrev, "Techniques of bounding the decoding error for block coded modulation," *IEEE Trans. Inform. Theory*, vol. 40, pp. 903–911, May 1994.
- [92] ———, "On the error probability of M -ary PSK block coded modulation," *IEEE Trans. Commun.*, vol. 44, pp. 427–433, Apr. 1996.
- [93] E. Zehavi and J. K. Wolf, "On the performance evaluation of trellis codes," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 196–202, Mar. 1987.
- [94] G. D. Forney Jr., "Geometrically uniform codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, Sept. 1991.
- [95] E. Biglieri *et al.*, *Introduction to Trellis-Coded Modulation with Applications*. New York: Macmillan, 1991.