Multilevel-Codes: Distance Profiles and Channel Capacity

Johannes Huber

Lehrstuhl für Nachrichtentechnik
Universität Erlangen–Nürnberg
Cauerstrasse 7, D-91058 Erlangen

Abstract

The calculation of the average profile of the squared Euclidean distance is derived for the individual coding levels of multilevel-coding schemes using the weight enumerators of the binary linear component codes. It is shown that the multiple representation of binary symbols at low coding levels causes a significant loss in power efficiency. This effect can also be observed for the capacities of the equivalent channels at the individual coding levels, which are derived. It is proven that the capacity of any digital modulation scheme can be achieved by a multilevel-coding scheme together with multistage-decoding in principle. The results allows an optimum tuning of the rates for the component codes.

Zusammenfassung

Für Multilevel-Codes mit linearen binären Komponentencodes wird das Distanzprofil bzgl. der quad. Euklidischen Distanz für die einzelne Codierebene aus der Gewichtsverteilung des Binärcodes abgeleitet. Es zeigt sich, daß infolge der mehrfachen Repräsentation der Binärsymbole durch Signalpunkte deutliche Störabstandsverluste zu verzeichnen sind. Dieser Effekt zeigt sich auch hinsichtlich der Kapazitäten der äquivalenten Kanäle für die einzelnen Codierebenen, deren Berechnung sehr einfach abgeleitet wird. Es wird nachgewiesen, daß durch Multilevel-Codierung und Multistage-Decodierung im Prinzip die Kapazität eines mehrstufigen digitalen Übertragungsverfahrens erreichbar ist. Die Ergebnisse erlauben eine optimale Aufteilung der Rate auf die einzelnen Komponentencodes.

Keywords: power and bandwidth efficient digital transmission schemes, multilevel-coding, multistage-decoding, distance profile, channel capacity

Schlagworte: Leistungs- und bandbreiteneffiziente digitale Übertragungsverfahren, Codierung auf mehreren Ebenen, Mehrstufendecodierung, Distanzprofil, Kanalkapazität

1 Introduction

For $M > 2$-ary digital transmission schemes like ASK, PSK, QAM or CPM (incl. FSK) an efficient combining of channel coding and modulation is possible using multilevel-coding (MLC). Transmission schemes with high power and bandwidth efficiency can be designed by this method in various ways. The concept of MLC has been developed almost at the same time as trellis-coded pulse-amplitude modulation (TCM), cf. [10, 18, 19] and various aspects of both approaches have carefully been investigated in detail, see e.g. [6, 15, 4, 2, 22, 20, 3, 17] and many other references. MLC and TCM are closely related because both methods are based on an iterative partitioning of the set of signal elements of the modulation scheme. TCM may be interpreted as a special MLC-scheme with only two levels, a lower one with coding and a higher one without coding, whereas MLC with short component block codes is applied to design multidimensional signal sets and partitions thereupon for TCM, see e.g. [13]. The distance structure of MLC-schemes is in principle known as methods of generalized concatenated codes can be applied, [14]. Often, a design of MLC-schemes according to the minimum Euclidean distance criterion for sequences of signal elements of eqs. (5) and (6) has been proposed, e.g. [2].

Decoding can efficiently be done individually for each component code beginning from the lowest level and later on using decisions of lower levels. This suboptimum process is called multistage-decoding (MSD), see e.g. [22, 20, 17]. Simulation results for simple examples have shown that surprisingly MSD with Maximum-Likelihood-Sequence-Estimation (MLSE) on each level is only very slightly inferior to an optimum joint MLSE over all levels, cf. [12, 16], while saving an enormous implementation effort. On the other hand, it can be observed that the performance of MLC-schemes with binary component codes which are designed according to the usual rule (5) and (6), is far away from desired asymptotic curves corresponding to the designed minimum Euclidean distance. Errors in the lowest level predominate, cf. e.g. [22, 21, 16]. For that reason an increase in the number of possible error events with low distances often has been supposed. In order to get rid of this effect interleaving between the levels, forwarding of reliability information from lower to higher levels using so-called soft-output decoding-algorithms (cf. e.g. [9]), and especially iterative MSD has been proposed [21]. But interleaving between coding levels causes a very high delay of data which often is not tolerable in practice. On the other hand, a comparison of schemes with and without interleaving is not fair because interleaving produces an enormous increase of the effective codeword length. Using such long component codes in a direct way would yield a quite better performance, especially since now there are powerful schemes available for which a near optimum decoding algorithm with constant effort per information symbol exists for all codeword lengths, cf. e.g. [1]. Although there is a rich bibliography on MLC, the properties of the *individual equivalent channels at the different coding levels* has yet not been in the focus of interest.

In this paper the influence of the multiple representation of symbols at low levels caused by the information at higher levels on the distance profile at the *individual level* is investigated in section 3. It turns out that a design of component codes at low levels according to the criterion of balanced minimum Euclidean distances is not sufficient as long as not very short codes are used. The capacities of the individual equivalent channels which are derived in section 4, give a quite better criterion. It is shown that the capacity of any digital modulation scheme can be achieved in principle by MLC together with MSD. Thus, the channel coding problem for *any* digital modulation scheme can be solved by the application of *binary* codes for components of a properly designed MLC-scheme. In order to specify the notation applied in this paper, a brief introduction to MLC is presented in section 2.

2 Multilevel-Coding

MLC is applied to a digital modulation scheme with $M = 2^n$; $n \geq 1$ signal elements, i.e. M signal points a_m ; $m \in \{1, 2, \ldots, M\}$ in a signal space with D dimensions per modulation interval defined by D well suited base-functions. The signal points are assumed to be equiprobable. (The restriction to 2^n-ary schemes with equiprobable signal points is only due to conciseness reasons. The general results hold for any number of signal points with arbitrary probability distributions, too.) This assumption is no serious restriction because possible gains by signal shaping are very small for small signal constellations $\mathbf{A} = \{a_m | m \in \{1, 2, \ldots, M\}\}$, whereas for large constellations channel coding and shaping can be treated separately, [5]. An iterative binary set partitioning with subsets of equal cardinality at each partitioning level is applied to $\mathbf{A}$ in order to define a bijective mapping of binary signal numbers $\vec{c} = (c^0, c^1, \ldots, c^{n-1})$; $c^q \in \{0, 1\}$; $q \in \{0, 1, \ldots, n-1\}$ to the signal points: $m \leftrightarrow \vec{c}$. The subsets on level q are characterized by the paths $c^0, c^1, \ldots, c^{q-1}$ from the root to these subsets in the partitioning tree of $\mathbf{A}$:

$$\mathbf{A}_{c^0 \ldots c^{q-1}} = \left\{ a_m | m \leftrightarrow (c^0, c^1, \ldots, c^{q-1},\ x^q, \ldots, x^{n-1}),\ x^i \in \{0, 1\} \right\} \tag{1}$$

We claim the set partitioning to be *regular*, i.e. all 2^q subsets at partitioning level q (regarded as constellations of individual modulation schemes) have equal capacities K_q on the given channel:

$$K(\mathbf{A}_{c^0 c^1 \ldots c^{q-1}}) = K_q; \quad c^i \in \{0; 1\};\ i \in \{0, 1, q-1\} \tag{2}$$

Here, $K(\mathbf{B})$ denotes the capacity of a modulation scheme with a signal set $\mathbf{B}$. The memoryless, discrete time channel without intersymbol interference is specified by the conditional probability density functions $f_{\mathbf{y}|a_m}(y)$ of the channel output variable y for transmission of signal point a_m. For equiprobable signal points

$$K(\mathbf{B}) = \frac{1}{|\mathbf{B}|} \int\limits_{\mathbf{Y}} \sum_{a_m \in \mathbf{B}} f_{\mathbf{y}|a_m}(y) \log_2 \left(\frac{f_{\mathbf{y}|a_m}(y)}{\frac{1}{|\mathbf{B}|} \sum\limits_{a_k \in \mathbf{B}} f_{\mathbf{y}|a_k}(y)} \right) \mathrm{d}y \tag{3}$$

is valid, cf. [11]. The integration has to be done over the set of channel output variables, e.g. over $\mathbb{R}^D$ (soft decision). If the channel output is quantized (e.g. hard decision) the integral degenerates to an ordinary sum. For the <u>a</u>dditive <u>w</u>hite <u>G</u>aussian <u>n</u>oise channel (AWGN) the partitioning is regular, if the profiles of the squared Euclidean distance are equal for all subsets on one partitioning level, i.e. if all these subsets differ only in translation and/or rotation in the signal space with D dimensions. Usually, for each sequence $\langle c_\mu^q \rangle := \langle \ldots, c_{-1}^q, c_0^q, c_1^q, c_2^q, \ldots \rangle$; $\mu \in \mathbb{Z}$ of the components of the binary signal numbers $\vec{c}$ an individual binary code $\mathbf{C}^q$ with rate R^q is applied, see **Fig. 1**. This special case is assumed in section 3. But it is also possible to form I classes of n_i consecutive components with

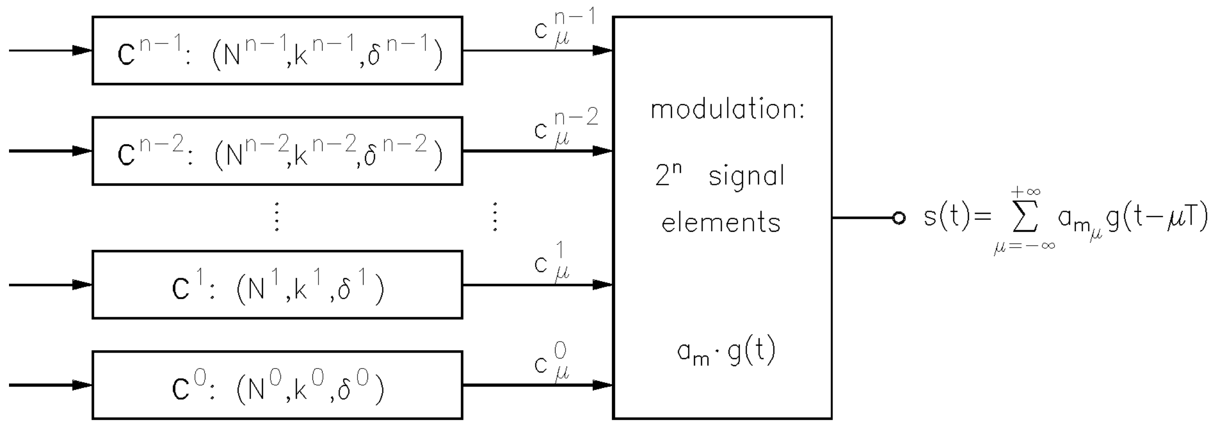

Figure 1 Block diagram of a multilevel-coding scheme (special case: binary component codes with wordlength N and min. Hamming distances δ^q, digital PAM)

$$\sum_{i=0}^{I-1} n_i = n\,, \tag{4}$$

and to use individual codes $\mathbf{C}^i$ over 2^{n_i}-ary symbols for these classes. Each class specifies a coding level. The equiprobability of the signal points results immediately from the assumed independence of the different coding levels. For the AWGN-channel the minimum squared Euclidean distance (MSED) $d^2_{\min}$ between any pair of sequences of signal points specified by the MLC-scheme is given by, see e.g. [6]:

$$d^2_{\min} = \min_i (d^2_{\min,i}) \ ; \quad i \in \{0, 1, \ldots, I-1\} \tag{5}$$

Here, $d^2_{\min,i}$ denotes the MSED at the i-th coding level provided that the components of the binary signal number below this level are specified, e.g. for the subset $\mathbf{A}_{\underbrace{00\ldots0}_{q}}$ with $q = \sum_{j=0}^{i-1} n_j$. Using a binary component code $\mathbf{C}^i$ with minimum Hamming distance δ^i on coding-level i, the MSED at this level is given by, see e.g. [2]

$$d^2_{\min,i} = \delta_i \cdot d^2_q \quad ; \qquad q = \sum_{j=0}^{i-1} n_j \tag{6}$$

Here, d_q denotes the minimum intra subset Euclidean distance at the q-th partitioning level.

3 Distance Profile for Linear, Binary Component Codes

For transmission schemes using MLC with binary component codes and a design of the individual rates R^q, or minimum Hamming distances δ^q, resp., using eqs. (5) and (6) according to the criterion of balanced $\delta^2_{\min,q}$ over all levels, errors at the lowest coding-level predominate the system performance. The coding gain is far away from the asymptotic curve promised by the designed MSED. The reason for this degradation is the multiple representation of binary symbols which causes an enormous multiplication of possible error events with low distances. In **Fig. 2**, the multiple representation of binary symbols is illustrated for a set partitioning according to the criterion of maximum intra subset MSED (Ungerboeck's set partitioning [19]) for large ASK- and QAM-constellations. For signal points which are far away from the boundary of the constellation, the inverse binary symbol is represented by two and four nearest neighbours, respectively. In general, each point in a D-dimensional constellation over $\mathbb{Z}^D$ has up to $2D$ neighbours representing the inverse binary symbol. A single binary codeword of $\mathbf{C}^0$ with minimum Hamming weight δ^0 causes up to $(2D)^{\delta^0}$ sequences of signal points with MSED $\delta^0 \cdot d_0^2$ from a sequence representing the all zero sequence. For the lowest coding-level the minimum intra set distance d_0 is the smallest one and, therefore, a code $\mathbf{C}^0$ with the highest minimum Hamming weight δ^0 has to be applied for this level. Thus, the greatest degradation due to multiple symbol representation occurs at level 0. For set partitioning according to Ungerboeck's criterion, the rates R^q rapidly goes up close to 1 at higher levels. Therefore, only a small portion of all possible multiple representations is cancelled by the code constraints at higher levels. The difference in the number of error events with relatively small Euclidean distances for independent symbols and for coded symbols at higher levels is not very great. But the only benefit of an overall MLSE over MSD with individual MLSE consists in the fact that the code constraints at higher levels are taken into account while decoding at lower levels. This observation gives an explanation of the small difference in performance of optimum overall MLSE and MSD for set-partitioning according to Ungerboeck's criterion.

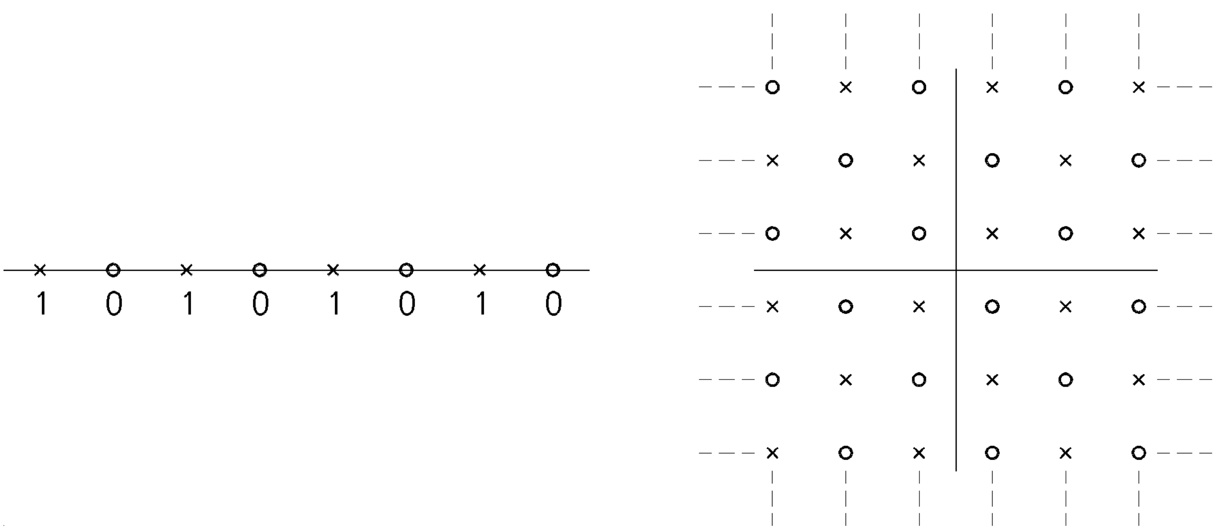

Figure 2 Multiple representation of the binary symbols c^0 at level 0 for signal constellation taken from lattices $(2\mathbb{Z}+1)$ and $(2\mathbb{Z}+1)^2$ ($\circ : c^q = 0$; $\times : c^q = 1$)

In this section we assume *binary, linear* component codes $\mathbf{C}^q$. Under this restriction the profiles of the squared Euclidean distance for each individual coding level can be derived from the weight distributions of the linear codes $\mathbf{C}^q$, see [16]. Assume that correct decisions have already been found for the binary symbol sequences $\langle c_\mu^j \rangle$; $j \in \{0, 1, \ldots, q-1\}$ before starting the q-th step of the MSD. Without loss of generality, we assume that the all zero sequence is transmitted at level q. Additionally we claim for the moment that *special* signal points $a_\lambda \in \mathbf{A}_{c^0 \ldots c^{q-1} 0}$; $\lambda = \{1, 2, \ldots, 2^q\}$ are used for representation of these symbols

$c^q_\mu = 0$. These special points – one per subset $\mathbf{A}_{c^0\ldots c^{q-1}0}$ – are characterized by equal distance distributions to all points of the corresponding opposite subset $\mathbf{A}_{c^0\ldots c^{q-1}1}$, whose elements represent the inverse symbol $c^q = 1$. The squared Euclidean distances from each a_λ to all $r = 2^{n-q-1}$ points of the corresponding opposite subset $\mathbf{A}_{c^0\ldots c^{q-1}1}$ are denoted by $\vartheta_1, \vartheta_2, \ldots, \vartheta_r$. Because of the regularity of the set partitioning a point a_λ with these squared distances exists for any of the subsets $\mathbf{A}_{c^0\ldots c^{q-1}0}$. Due to the information transmitted at higher levels one codeword of $\mathbf{C}^q$ with weight δ can be mapped to $2^{(n-q-1)\cdot\delta} = r^\delta$ different sequences of signal points. (Symbols $c^q = 0$ of this codeword are still represented by the special signal points a_λ at this point of our derivation.) There are

$$\frac{\delta!}{j_1!\, j_2!\, \ldots\, j_r!} \quad \text{with} \quad j_\ell \in \{0, 1, \ldots, \delta\} \quad \text{and} \quad \sum_{\ell=1}^{r} j_\ell = \delta \tag{7}$$

sequences for which the different squared Euclidean distances ϑ_ℓ occur j_ℓ times each; $\ell \in \{1, 2, \ldots, r\}$. Applying the polynomial theorem

$$\sum \cdots \sum_{j_1+j_2+\ldots+j_r=\delta} \frac{\delta!}{j_1!\, j_2!\, \ldots\, j_r!} X_1^{j_1} \cdot X_2^{j_2} \ldots X_r^{j_r} = (X_1 + X_2 + \ldots + X_r)^\delta, \tag{8}$$

the profile of the squared Euclidean distance can simply be expressed by substituting the variables X_ℓ in (8) by Z^{ϑ_ℓ}, where Z is a real dummy variable. The exponents are added to the entire squared Euclidean distance and the corresponding factors describe the multiplicities of sequences with these entire squared distances. If the distribution of the squared distances $\vartheta_1, \vartheta_2, \ldots, \vartheta_r$ is expressed by a polynomial (distance enumerator)

$$B^q_\lambda(Z) = \sum_{\ell=1}^{r} Z^{\vartheta_\ell} = \sum_s b^q_{\lambda,s} Z^{\vartheta_s} \tag{9}$$

as usual, the profile of the squared Euclidean distances for *one* binary codeword with weight δ to the uniquely represented all zero word is specified by the polynomial

$$\left(B^q_\lambda(Z)\right)^\delta \tag{10}$$

In (9), the coefficients $b^q_{\lambda,s}$ denote the multiplicities of squared distances ϑ_s from the points a_λ. Now we drop the restriction that all symbols $c^q = 0$ are represented by special signal points a_λ. The representators for $c^q = 0$ are chosen for each subset $\mathbf{A}_{c^0c^1\ldots c^{q-1}0}$ with equal probabilities $1/r$. The probability for a sequence of δ signal points for $c^q = 0$, for which points $a_i \in \mathbf{A}_{c^0c^1\ldots c^{q-1}0}$ with distance enumerators $B^q_i(Z)$ are used k_i times each, is given by

$$\frac{1}{r^\delta} \cdot \frac{\delta!}{k_1!\, k_2!\, \ldots\, k_r!} \quad \text{with} \quad k_i \in \{0, 1, \ldots, \delta\} \quad \text{and} \quad \sum_{i=1}^{r} k_i = \delta \tag{11}$$

The distance profile for a codeword with weight δ to the all zero sequence represented by k_i points a_i with distance profiles $B^q_i(Z)$ in the differing binary symbols is expressed by the polynomial:

$$\left(B^q_1(Z)\right)^{k_1} \cdot \left(B^q_2(Z)\right)^{k_2} \ldots \left(B^q_r(Z)\right)^{k_r} \tag{12}$$

Using (11), the average of the distance profile over all possible sequences of representators for the zero sequence reads:

$$\frac{1}{r^\delta} \sum \cdots \sum_{k_1+k_2+\ldots+k_r=\delta} \frac{\delta!}{k_1!\, k_2!\, \ldots\, k_r!} \left(B^q_1(Z)\right)^{k_1} \ldots \left(B^q_r(Z)\right)^{k_r} = \left(\frac{1}{r} \sum_{i=1}^{r} B^q_i(Z)\right)^\delta \tag{13}$$

Again the polynomial theorem (8) is applied. As a regular partitioning is assumed, the *average* profile

$$B^q(Z) = \frac{1}{r}\sum_{i=1}^{r} B_i^q(Z) \quad \text{with} \quad r = 2^{n-q-1} \tag{14}$$

of squared Euclidean distances between all representators of symbols $c^q = 0$ and $c^q = 1$ is equal for all subsets at the partitioning level q. The probabilities of errors from 0 to 1 and vice versa are equal and the result (13) is valid for any pair of the codewords with Hamming distance δ.

Let $N^q(D) = 1 \cdot D^0 + N_{\min} \cdot D^{\delta_{\min}} + N_2 \cdot D^{\delta_2} + \ldots$ be the weight enumerator of the linear binary code $\mathbf{C}^q$. For a convolutional component code $N^q(D)$ and $C^q(D)$ denote the enumerator of the free Hamming distance and the enumerator of the free distance weighted by the average number of information symbol errors per error event as usual, see e.g. [7]. By application of (13) the average distance profile of the squared Euclidean distance of the code at level q is expressed by the polynomials

$$N_E^q(Z) = N^q\left(D = B^q(Z)\right) \quad \text{and} \quad C_E^q(Z) = C^q\left(D = B^q(Z)\right) \tag{15}$$

Using this result, the error probability at level q (without taking into account a possible error propagation from lower levels) can be upper-bounded via the union-bound in the same way as usual for binary coding, antipodal signalling, and MLSE, see e.g. [7].

Example: 8-ary PSK, level 0, partitioning for maximum intra subset MSED (see **Fig. 3**)

$$\vartheta_1 = \vartheta_2 = 4\sin^2(\tfrac{\pi}{8});\ \vartheta_3 = \vartheta_4 = 4\sin^2(\tfrac{3\pi}{8})$$

$$B^0(Z) = 2 \cdot Z^{\vartheta_1} + 2 \cdot Z^{\vartheta_3}$$

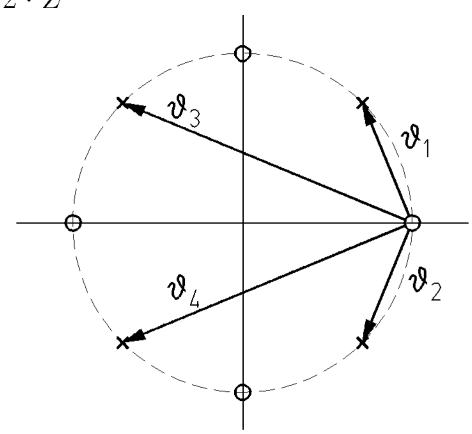

Figure 3 Squared Euclidean Distances ϑ_ℓ at level 0 for 8-ary PSK

Simulation results and the upper bound (15) (dashed-dotted line) taking multiple symbol representation into account are given for a convolutional code $\mathbf{C}^q$ with rate 1/3 and 8 states in **Fig. 4** ($\delta_{\min} = 10$; weighted weight enumerator $C^0(D)$ see [7]). At moderately high signal to noise ratios SNRs the simulation results are very close to the upper bound (15). Additionally, the union bound for a unique representation of the binary symbols by only two antipodal signal points with Euclidean distance $d_0 = \sqrt{\vartheta_1}$ is shown by a dashed line. Because of the enormous increase in the number of low distance error events a loss of more than 2 dB has to be accepted at a bit error rate BER $= 10^{-6}$ due to multiple symbol representation. The number of nearest neighbour error events is enlarged from 6 to 6144. Using the traditional design rule eqs. (5) and (6), this effect would not be taken into account. An error curve close to the dashed line would be expected. Therefore, this rough design rule does not lead to satisfying MLC-schemes.

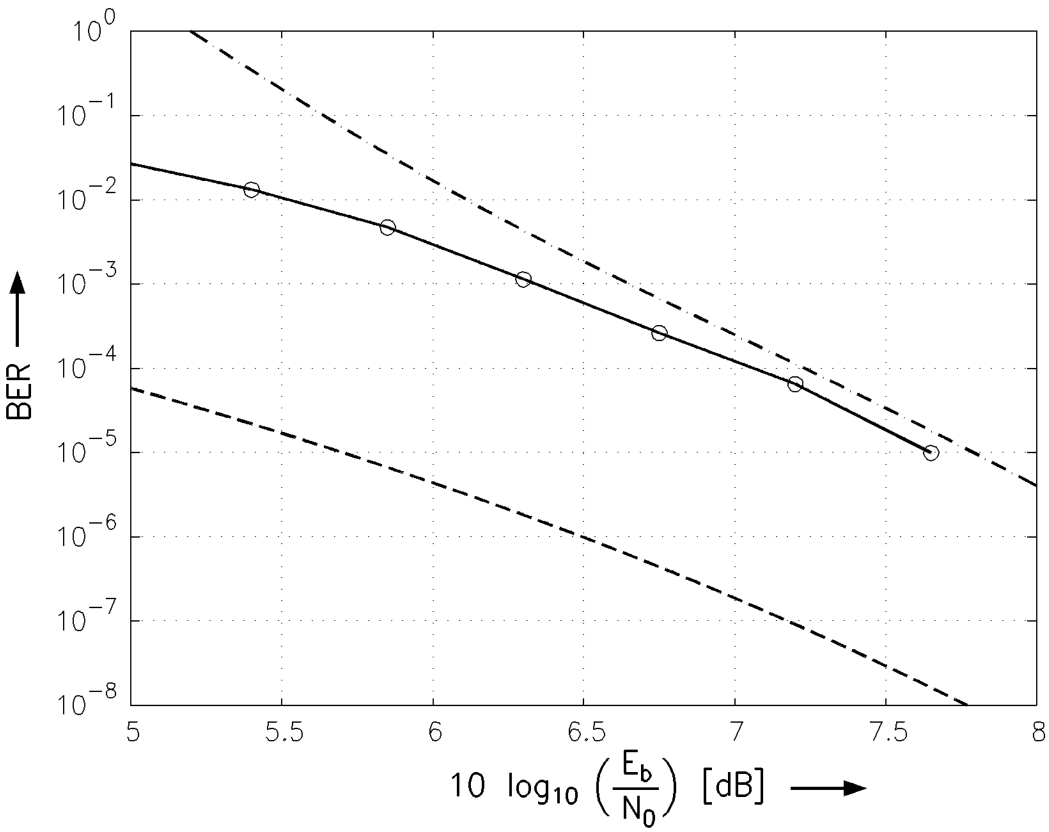

Figure 4 Simulation results and upper bound (union-bound) for the bit error rate (BER) at level 0 for 8-ary PSK; component code $\mathbf{C}^0$: rate 1/3 convolutional code with 8 states ($\delta_{\min} = 10$)

4 Capacity of the Equivalent Channels

As shown in the previous section a design of a MLC-scheme according to the criterion of balanced minimum Euclidean distances at all levels turned out to be problematical. Especially, for the application of component codes which cannot sufficiently be characterized by a minimum Hamming distance, e.g. product codes or TURBO-codes [1], this rough rule cannot be applied at all. Therefore, we propose to choose the rates R^q of the component codes from the capacities C^q of the equivalent channels at the individual coding levels.

We assume a regular set partitioning tree for the signal set satisfying eq. (2) on each level. At first capacity C^0 is calculated. The multiple representation of symbols by signal points due to the information at higher levels can be interpreted as an additional discrete noise in level 0, by which the signal points are shifted from unique to ambiguous representators, and which decreases capacity.

Let n_0 ; $1 \leq n_0 \leq n$ components of the binary signal number $\vec{c}$ be combined for coding level 0. The capacity C^0 of the equivalent channel of this coding level 0 is defined to be

$$C^0 = \frac{1}{2^{n_0}} \sum_{c^0=0}^{1} \cdots \sum_{c^{n_0-1}=0}^{1} \int_{\mathbf{Y}} f_{\mathbf{y}|c^0 \ldots c^{n_0-1}}(y) \log_2 \frac{f_{\mathbf{y}|c^0 \ldots c^{n_0-1}}(y)}{f_{\mathbf{y}}(y)} \, \mathrm{d}y \,, \tag{16}$$

if the signal points are equiprobable. The conditional probability density functions of the channel output variable y for given symbols $c^0, c^1, \ldots, c^{n_0-1}$ reads:

$$f_{\mathbf{y}|c^0 \ldots c^{n_0-1}}(y) = 2^{n_0-n} \cdot \sum_{a_m \in \mathbf{A}_{c^0 c^1 \ldots c^{n_0-1}}} f_{\mathbf{y}|a_m}(y) \tag{17}$$

The capacity $K(\mathbf{B})$ for an arbitrary (sub-)set $\mathbf{B}$ of equiprobable signal points (see eq. (3)) can be expanded to:

$$K(\mathbf{B}) = \frac{1}{|\mathbf{B}|} \int_{\mathbf{Y}} \sum_{a_m \in \mathbf{B}} f_{\mathbf{y}|a_m}(y) \left[\log_2 \left(\frac{f_{\mathbf{y}|a_m}(y)}{f_{\mathbf{y}}(y)} \right) - \log_2 \left(\frac{\frac{1}{|\mathbf{B}|} \sum_{a_k \in \mathbf{B}} f_{\mathbf{y}|a_k}(y)}{f_{\mathbf{y}}(y)} \right) \right] \mathrm{d}y \tag{18}$$

Eq. (18) is a simple application of the well known chain rule for mutual information, [11]. This expansion is now applied to all 2^{n_0} subsets $\mathbf{A}_{c^0 \ldots c^{n_0-1}}$ which contain all representators of the encoder output symbols $(c^0 \ldots c^{n_0-1})$ at level 0. Substitution of (17) yields:

$$\begin{aligned} &\sum_{c^0=0}^{1} \cdots \sum_{c^{n_0-1}=0}^{1} K\left(\mathbf{A}_{c^0 \ldots c^{n_0-1}}\right) = \\ &\sum_{c^0=0}^{1} \cdots \sum_{c^{n_0-1}=0}^{1} \left(2^{n_0-n} \int_{\mathbf{Y}} \sum_{a_m \in \mathbf{A}_{c^0 \ldots c^{n_0-1}}} f_{\mathbf{y}|a_m}(y) \log_2 \frac{f_{\mathbf{y}|a_m}(y)}{f_{\mathbf{y}}(y)} \, \mathrm{d}y - \right. \\ &\qquad \left. - \int_{\mathbf{Y}} f_{\mathbf{y}|c^0 \ldots c^{n_0-1}}(y) \log_2 \frac{f_{\mathbf{y}|c^0 \ldots c^{n_0-1}}(y)}{f_{\mathbf{y}}(y)} \, \mathrm{d}y \right) = \\ &= 2^{n_0} K(\mathbf{A}) - 2^{n_0} C^0 = 2^{n_0} K_{n_0} \end{aligned} \tag{19}$$

The last expression results from the regularity of the set partitioning. The simple result

$$C^0 = K_0 - K_{n_0} = K(\mathbf{A}) - K(\mathbf{A}_{\underbrace{0 \ldots 0}_{n_0}}) \tag{20}$$

says that the capacity of the equivalent channel at coding level 0 is the difference of the capacities of the entire constellation $\mathbf{A}$ and its subset specified by a certain output symbol of the level 0 encoder.

If rate $R^0 \leq C^0$ and the codeword length is chosen sufficiently high, an error free decoding in step 0 in the MSD-process is possible in principle. Thus, the sequence $\langle \mathbf{A}_{c^0 \ldots c^{n_0-1}} \rangle$ of subsets can correctly be specified. Because of this reason, eq. (20) can be applied to the capacity of the equivalent channel at coding level 1 in the same way as for level 0:

$$C^1 = K_{n_0} - K_{n_0+n_1} \tag{21}$$

This argument holds iteratively for all levels:

$$C^i = K_{\sum_{\ell=0}^{i-1} n_\ell} - K_{\sum_{\ell=0}^{i} n_\ell} \tag{22}$$

The subsets at the highest level of the set partitioning tree contain only a single signal point. Therefore, $K_n = 0$. Thus, we have proven the subsequent theorem:

Theorem

> The capacity C of a 2^n-ary digital modulation scheme is equal to the sum of the capacities C^q of the equivalent channels at the individual coding levels of a multilevel-coding scheme which is based on a regular binary partitioning tree of the signal set:
>
> $$C = \sum_{i=0}^{I-1} C^i \tag{23}$$
>
> The capacity C can be achieved via multilevel-encoding and multistage-decoding if and only if the individual rates R^i at the different coding levels are chosen to be the capacities of the equivalent channels, $R^i = C^i$.

This theorem can be generalized to arbitrary, even irregular partitioning trees and for nonequiprobable signal points in a straight forward way. Please notice that the *method* of partitioning has no influence on the fact that MLC together with MSD can lead to capacity. Only the distribution of the rates to the different levels is affected by the specific method of partitioning. Seen from this point of view, the usual set partitioning for maximum intra subset MSED has no special importance.

The theorem is not only an efficient way for an optimum design of a MLC-scheme, but also shows in general that out of the huge set of all possible codes with length N which map $N \cdot \sum_{i=0}^{I-1} R^i$ bits to N channel symbols of a transmission scheme, the (in comparison very small) subset of codes with an *independent* mapping of NR^i bits to the I levels is a selection whose performance is not below average. Additionally the theorem states that the channel coding problem for any digital transmission scheme can be solved in principle by the application of *binary codes* in an optimum way via MLC and MSD.

In **Fig. 5** the capacities of the equivalent channels and the total capacity for the application of binary component codes to 8-ary ASK are shown versus the variance σ^2 of the Gaussian noise per (real) dimension of the signal space. The signal points are taken from the lattice of odd integers ($\mathbf{A} = \{\pm 1, \pm 3, \pm 5, \pm 7\}$). The usual partitioning for maximum intra subset MSED is applied. Additionally, the capacity of 2ASK ($\{\pm 1\}$, level 0, but without multiple symbol representation) is illustrated by a dashed line in order to show the loss due to multiple symbol representation. In contrast to the usual abscissa parameter E_b/N_0, the variance σ^2 of noise is used for a better illustration of the relationships between the various capacity curves.

Analogous results for 8-ary PSK and 16-ary QAM with binary component codes are shown in **Figs. 6** and **7**. In **Fig. 8** the capacity curves C^0 for binary coding at level 0 ($n_0 = 1$) are given for large constellations taken from the lattices $2\mathbb{Z} + 1$; $D = 1$ and $(2\mathbb{Z} + 1)^2$; $D = 2$. The capacity C^0 decreases with increasing dimensionality of the signal constellation because the number of nearest neighbour error events increases with $(2D)^\delta$. For very large constellations, these curves can also be used for higher levels by rescaling of the abscissa.

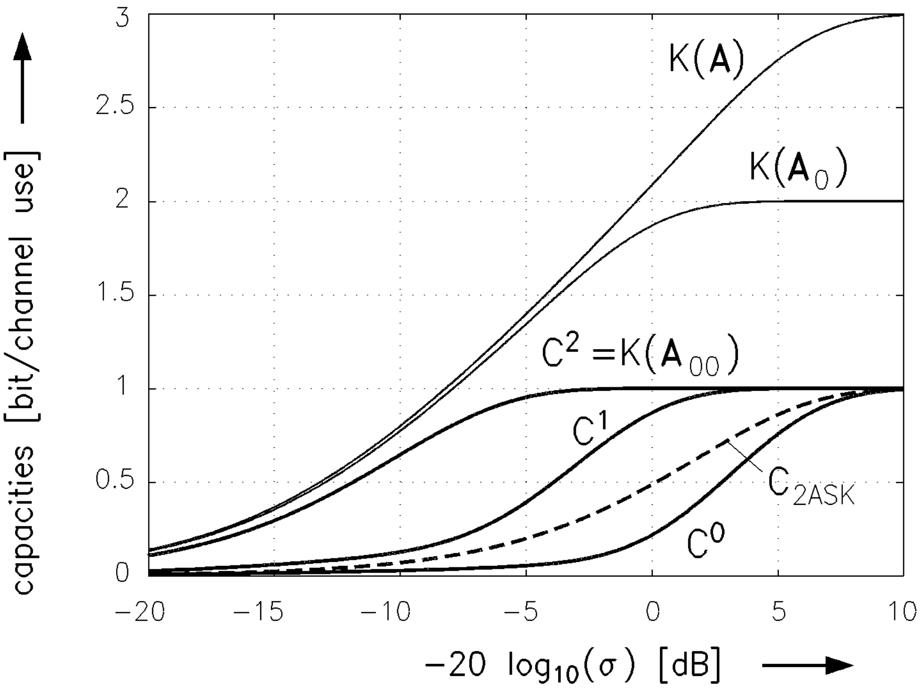

Figure 5 Capacities of the (sub-)sets and capacities of the equivalent channels at the individual coding levels for 8-ary ASK versus noise variance (AWGN-channel, set partitioning for maximum intra subset minimum Euclidean distance)

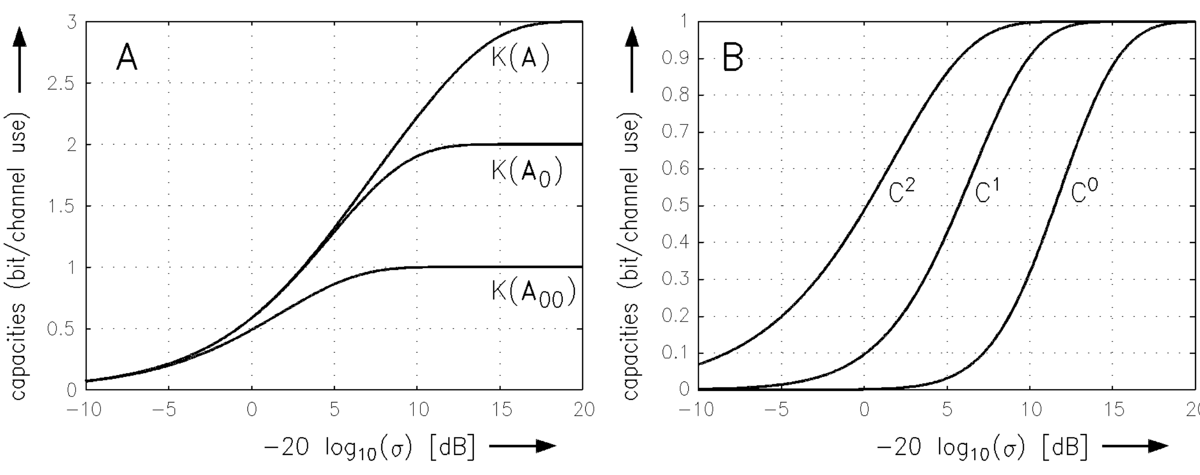

Figure 6 Capacities of the (sub-)sets (Fig. A) and capacities of the equivalent channels (Fig. B) at the individual coding levels for 8-ary PSK versus noise variance per dimension of the signal space ($|a_m| = 1$, AWGN-channel, set partitioning for maximum intra subset minimum Euclidean distance)

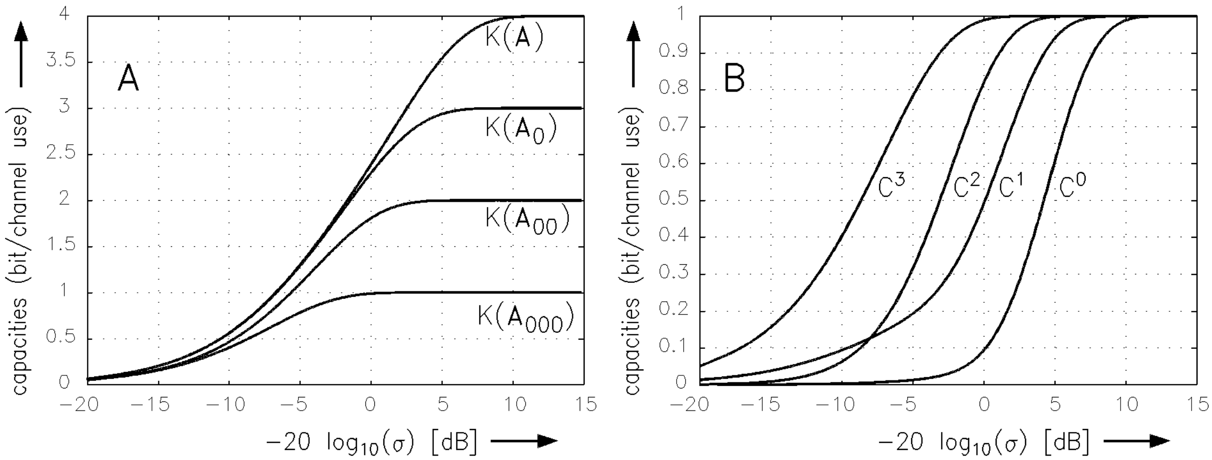

Figure 7 Capacities of the (sub-)sets (Fig. A) and capacities of the equivalent channels (Fig. B) at the individual coding levels for 16-ary QAM versus noise variance per dimension of the signal space (AWGN-channel, set partitioning for maximum intra subset minimum Euclidean distance)

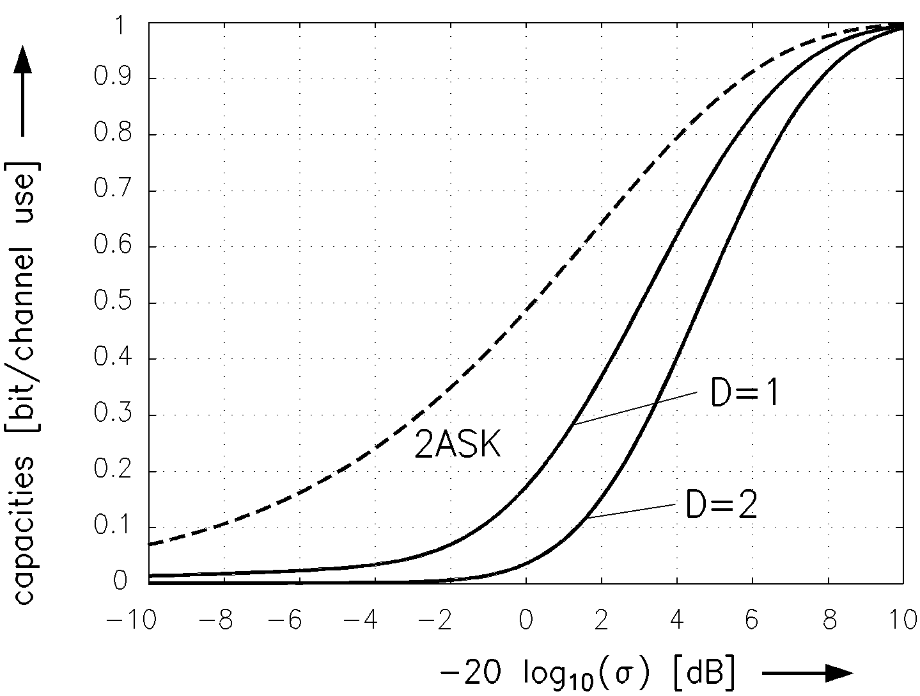

Figure 8 Capacities of the equivalent channels at level 0 for large constellations taken from lattices $(2\mathbb{Z}+1)$ and $(2\mathbb{Z}+1)^2$, cf. Fig. 2. (dashed line: capacity with equal minimum distance but without multiple representation of binary symbols ($\hat{=}$ 2-ary ASK))

Finally, **Fig. 9** shows the capacities C^q for 8-ary ASK and a partitioning of the signal set according to a criterion of minimum variance of intra subset signal points, see **Fig. 10**. Now, level 0 offers the highest capacity. The capacities C^q have a smaller divergency which may facilitate implementation in some cases. Additionally, schemes may be designed by this partitioning method with a natural adaptation of the total rate on the actual

SNR (gracefully degrading schemes). Choosing the rates R^q for all levels equally, errors appear for decreasing SNR at level 2, whereas the lower levels offer a reliable transmission furthermore. This behaviour may be suited better for MSD as the unequal distribution of capacities with increasing order as given in Fig. 5.

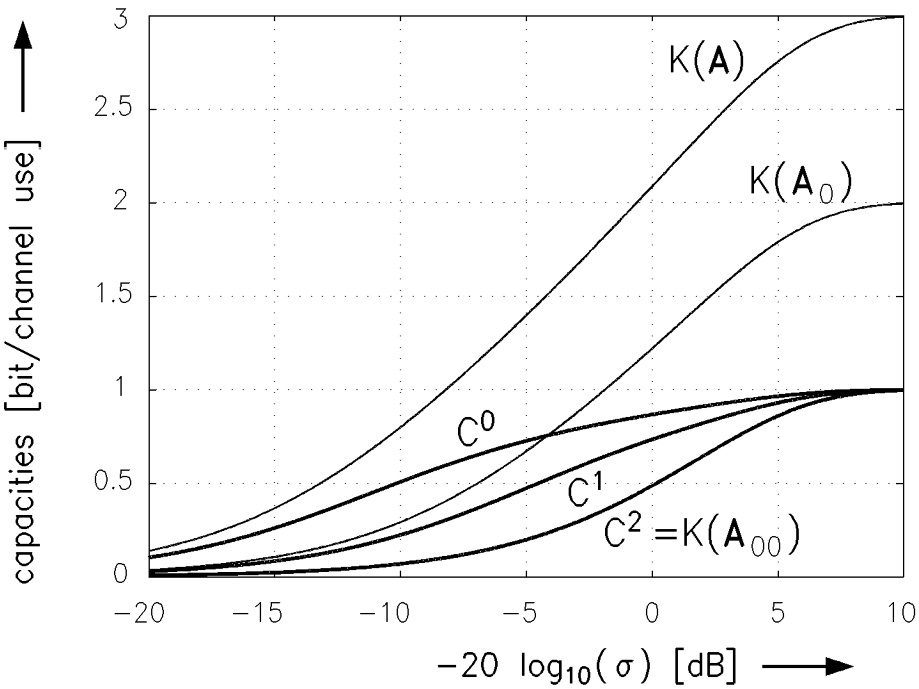

Figure 9 Capacities of the (sub-)sets and capacities of the equivalent channels at the individual coding levels for 8-ary ASK versus noise variance (AWGN-channel, set partitioning for minimum intra subset variance, cf. Fig. 10)

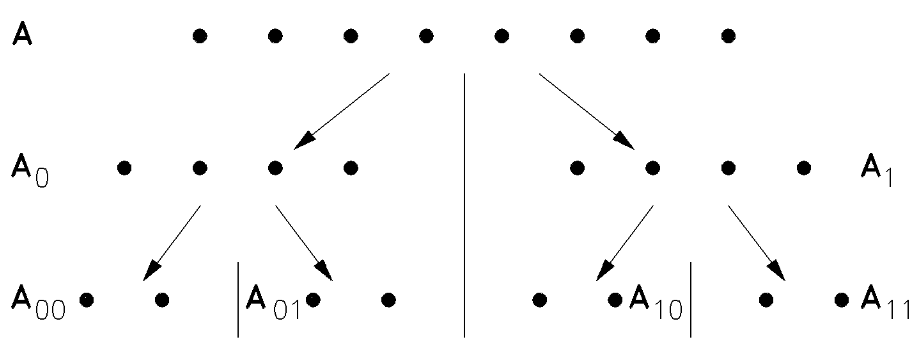

Figure 10 Set partitioning tree for 8-ary ASK for minimum intra subset variance

Besides capacities, we have calculated the individual cut-off rates R_0^i of the equivalent channels at the individual coding levels q, see [14], too. It turned out that for set partitioning for maximum intra subset MSED the sum $\sum_{i=0}^{I-1} R_0^i$ of individual cut-off-rates even exceeds the cut-off-rate R_0 of the entire digital modulation scheme. The author interprets this surprising result in that way that for a finite codeword length the MLC approach corresponds to a subset out of the huge set of all possible codes whose performance even is above average. On the contrary we observed $\sum_{i=0}^{I-1} R_0^i < R_0$ for set partitioning for minimum intra subset variance (e.g. Fig. 10). Thus, set partitioning according to Ungerboeck's criterion nevertheless seems to be well suited for MLC-schemes. The distribution of rates R^i to the levels according to the cut-off-rate criterion is very close to the distribution derived from the capacity criterion.

References

[1] C. Berrou, A. Glavieux, P. Thitimajshima: Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes. In *Proceedings of the International Conference on Communications (ICC'93)*, pp. 1064–1070, May 1993.

[2] E. Biglieri e.a.: *Introduction to Trellis–Coded Modulation with Applications*. Macmillan, New York, 1991.

[3] M. Bossert: *Kanalcodierung*. B.G. Teubner, Stuttgart, 1992.

[4] A.R. Calderbank: Multilevel Codes and Multistage Decoding: *IEEE Trans. on Commun.*, COM-37, pp. 222–229, March 1989.

[5] G.D. Forney Jr., L.-F. Wei: Multidimensional constellations – Part I: Introduction, figures of merit, and generalized cross constellations. *IEEE Journal on Selected Areas in Communications*, SAC-7, pp. 877–892, Aug. 1989.

[6] V.V. Ginzburg: Multidimensional signals for a continuous channel. *Probl. Inf. Transmission*, vol. 23, pp. 20–34, 1984. (Translation from Probl. Peredachi Inf., pp. 28-46, 1984).

[7] J. Huber: *Trelliscodierung*, Springer Verlag, Berlin, 1992.

[8] J. Huber and U. Wachsmann: Capacities of the Equivalent Channels in Multilevel Coding Schemes. *Electr. Letters*, vol. 30, pp. 557–558, March 1994.

[9] J. Huber, A. Rüppel: Zuverlässigkeitsabschätzung für die Ausgangssymbole von Trellis-Decodern. *AEÜ (Electronics and Communication)*, vol. 44, pp. 8–21, 1990.

[10] H. Imai, S. Hirakawa; A new multilevel coding method using error correcting codes. *IEEE Trans. on Information Theory*, vol. IT-23, pp. 371–377, May 1977.

[11] R. Johannesson: *Informationstheorie – Grundlage der (Tele-)Kommunikation.* Addison-Wesley, Lund, Schweden, 1988.

[12] M. Litzenburger: Gezielte Beeinflussung der Hüllkurve eines modulierten Signals durch Codierung und „Shaping". In *8. DFG-Kolloquium im Schwerpunktprogramm „Grundlagen digitaler Kommunikationssysteme"*, Saarbrücken, Juni 1993.

[13] S.S. Pietrobon, R.H. Deng, A. Lafanechére, G. Ungerboeck, D.J. Costello: Trellis–coded multidimensional phase modulation. *IEEE Trans. on Information Theory*, vol. IT-36, pp. 63–89, 1990.

[14] J. Portugheis: *Generalized Concatenated Codes for M-PSK Modulation.* Ph. D. Thesis, Technische Hochschule Darmstadt, Juli 1992.

[15] G.J. Pottie and D.P. Taylor: Multilevel Codes based on Partitioning. *IEEE Trans. on Information Theory*, vol. IT-35, pp. 87–98, Jan. 1989.

[16] K. Rebhan: *Schrittweise Decodierung von Mehrstufencodes.* Diploma Thesis, Lehrstuhl für Nachrichtentechnik, Universität Erlangen-Nürnberg, Okt. 1993.

[17] T. Takata, S. Ujita, T. Kasami, S. Lin: Multistage Decoding of Multilevel Block M-PSK Modulation Codes and Its Performance Analysis. *IEEE Trans. on Information Theory*, vol. IT-39, pp. 1204–1218, July 1993.

[18] G. Ungerböck, I. Csajka: On improving data-link performance by increasing channel alphabet and introducing sequence coding. In *Int. Sym. Inform. Theory*, Ronneby, Sweden, June 1976.

[19] G. Ungerböck: Channel coding with multilevel/phase signals. *IEEE Trans. on Information Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[20] T. Wörz, J. Hagenauer: Iterative Decoding for Multilevel Codes using Reliability Information. In *Proceedings of the IEEE Global Telecommunication Conference (GLOBECOM'92)*, Orlando, Dec. 1992.

[21] T. Wörz, J. Hagenauer: Decoding of M-PSK-multilevel codes. *European Transactions on Telecommunications*, vol. 4, pp. 65–74, May-June 1993.

[22] T. Wörz, J. Hagenauer: Multistage coding and decoding. In *Proceedings of the IEEE Global Telecommunication Conference (GLOBECOM'90)*, pp. 501.1.1–501.1.6, San Diego, Dec. 1990.