Coded Modulation by Multilevel–Codes: Overview and State of the Art

Johannes Huber, Udo Wachsmann, Robert Fischer

Lehrstuhl für Nachrichtentechnik II, Universität Erlangen–Nürnberg
Cauerstraße 7/NT, 91058 Erlangen, Germany
Phone: +49-9131-857113, Fax: +49-9131-858849, Email: lnt2@nt.e-technik.uni-erlangen.de

***Abstract* — The development of theory for multilevel coding during the last 20 years is reminded. Several design rules are compared and the capacity region of multilevel codes is outlined. We discuss the dimensionality of the constituent signal constellation and compare different labeling strategies by means of random coding exponent. Finally, by discussing bit interleaved coded modulation we show that the progress in theory now leads back to the origin of coded modulation.**

1 Introduction and Historical Overview

During the first 25 years after the foundation of information theory by C.E. Shannon 50 years ago research in channel coding was almost entirely restricted to coding for binary antipodal signalling like BPSK, QPSK with Gray mapping etc. It soon was recognized that a direct application of traditional forward error correcting codes (FEC) to bandwidth efficient modulation schemes with more than two signal points per dimension of a signal space, e.g. $M > 2$–ary ASK, $M > 4$–ary QAM, $M > 4$–ary PSK, see Fig. 1, offers no or only small gains over simple uncoded schemes at the same bandwidth efficiency. This effect becomes evident by a comparison of digital communication schemes in the power–bandwidth–plane for the additive white Gaussian noise channel (AWGN), see Fig. 2. The dashed lines mark the results of binary primitive BCH codes (cf. [30]). Power efficiency (i.e. min. equivalent energy per information bit E_b over one sided noise power density N_0 for desired reliability, here bit error rate BER $\leq 10^{-5}$) increases due to increasing error correctability, while bandwidth efficiency , (here depicted for bandwidth excess factor 0) together with code rate decrease. For fixed bandwidth efficiency slight gains over uncoded schemes (△) with smaller constellations are only possible using long FEC codes; for short codes FEC is useless at all.

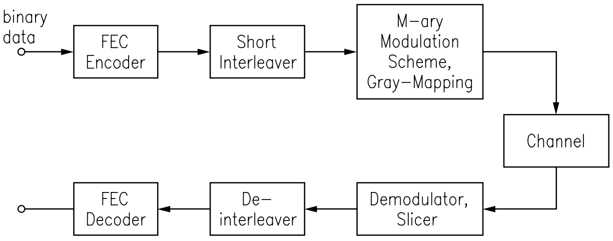

Figure 1: FEC scheme.

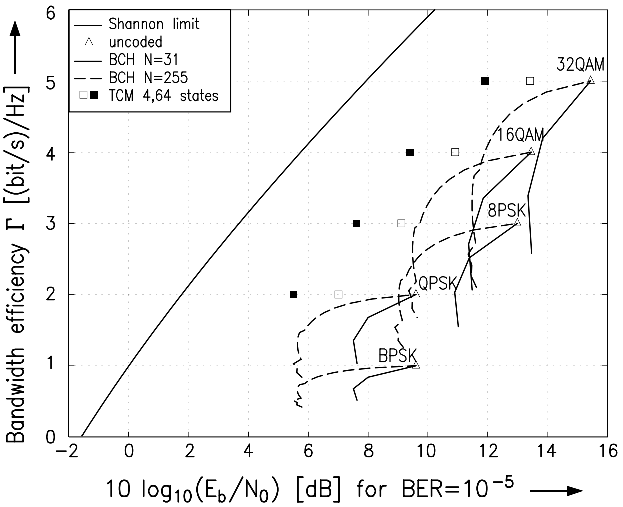

Figure 2: Power–bandwidth plane for the AWGN channel.

Having this effect in mind, the field of *coded modulation* was founded in 1974 by J.L. Massey [31] stating that *joint* coding and modulation can improve the performance of a digital transmission scheme. In 1976/77, independently Ungerböck [44, 41] and Imai [27] presented powerful and applicable coded modulation schemes. The common core is to optimize the code in Euclidean space rather than dealing with Hamming distance. For both approaches the mapping of binary address vectors $\mathbf{x}_m = (x_m^0, x_m^1, \ldots, x_m^{\ell-1})$ to the $M = 2^\ell$ signal points a_m, $m = 1, 2, \ldots, M$, of a PAM constellation $\mathbf{A}$ is based on iterative binary set partitioning. Ungerböck divides the binary components of the address vectors in two parts: the least significant binary symbols are convolutionally encoded whereas the high significant binary symbols 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 nature of sequences of signal points, Ungerböck's approach to coded modulation is named *trellis coded modulation* (TCM). TCM was originally proposed for one– and two–dimensional signal sets using one bit redundancy per signal point. Boxes in Fig. 2 indicate the efficiency of simple (□: 4 states) and fairly complex (■: 64 states) TCM schemes over two–dimensional constellations, i.e., for 0.5 bit redundancy per dimension. Thus, the deficiency of FEC was overcome by TCM. Lots of work were performed in order to provide more flexible transmission rates with TCM, using constituent signal constellations for coding in higher dimensions or signal constellations derived from lattice theory by combining several one– or two–dimensional PAM symbols, e.g. [7, 8, 42, 43, 17, 15, 34, 33]. This approach can be viewed as a special kind of code concatenation, cf. [22]. Although some more flexibility is offered no substantial improvement is possible by multidimensional TCM schemes at all, because less than 0.5 bit redundancy per dimension causes inevitable losses, cf. [41, 12].

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, see Fig. 3. Traditionally, it was proposed to choose the individual codes in such a way that the minimum distance of the Euclidean space code is maximized, [19, 6, 28, 35, 53] et al. In the following we refer to this concept of assigning codes to the individual levels as *balanced distances rule* (see Section 2). At the receiver side, each code $\mathcal{C}^i$ is decoded individually starting from the lowest level and taking decisions of prior decoding stages into account. This procedure is called *multistage decoding* (MSD), see Fig. 4. In contrast to Ungerböck's TCM, the MLC approach has the advantage of providing flexible transmission rates, because it decouples the dimensionality of the signal constellation from the code rate. Any code can be used as component code, e.g. block codes, convolutional codes or concatenated codes. Although MLC designed according to the balanced distances rule offers excellent asymptotic coding gains, it achieved only theoretical interest in the past. In practice, performance of such schemes is severely degraded due to high error rates at low levels. A lot of effort was done to overcome this effect, see e.g. [54].

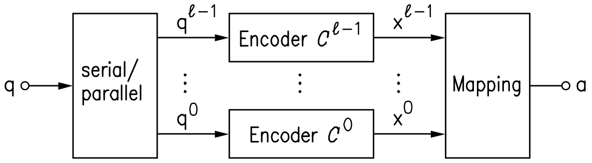

Figure 3: Multilevel encoder.

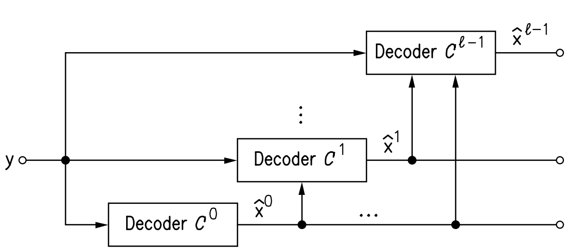

Figure 4: Multistage decoder.

A straightforward generalization of Imai's approach is to use q–ary component codes based on a non–binary partitioning of the signal set, cf. [23]. In this context, TCM is a special case of MLC using a single convolutional code with a non–binary output alphabet while higher levels remain uncoded. Therefore, TCM and MLC need not to be treated separately.

In his famous framework, G.D. Forney presented 1988 the concept of coset codes [13, 14]. By dealing only with infinite constellations (neglecting the boundary effects) and using the mathematics of lattice theory a general class of codes was established. Similar to TCM, not signal points but cosets are selected in the encoding process. Coset codes divide into two classes: *trellis codes* (a generalization of TCM) and *lattice 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 (see e.g. [16, 9]). De Buda [10] stated that lattice codes can approach the channel capacity of the AWGN channel. The proof was recently refined by Urbanke and Rimoldi [45] as well as by Forney [16].

For practical coded modulation schemes, where boundary effects have to be taken into account, Huber et al. [24, 23, 25, 51] proved that with multilevel codes and multistage decoding the capacity of the modulation scheme can be achieved if and only if the individual rates of the component codes are chosen properly. Key point is the well–known chain rule for mutual information, cf. also [29]. Main intention of this paper together with its companion papers [12] and [50] is to show that there are different ways to design power and bandwidth efficient digital communication schemes close to theoretical limits.

2 Design Rules and Capacity Regions

Bandwidth efficient communication in the area , $\geq 3\frac{\text{bit/s}}{\text{Hz}}$ preferably is based on digital pulse amplitude modulation (PAM) employing a constituent signal set $\mathbf{A} \subset \mathbb{R}^D$ in D dimensions with $M = 2^\ell$ points. A bijective mapping $\mathcal{M} : \{0\,;1\}^\ell \rightarrow \mathbf{A}$ of ℓ–dimensional binary vectors $\mathbf{x}_m = (x_m^0, x_m^1, \ldots, x_m^{\ell-1})$, $x_m^i \in \{0,1\}$, $m = 1,2,\ldots,M$, to all signal points $a_m \in \mathbf{A}$ is defined preferably by an iterative binary set partitioning procedure like proposed in [41]. We denote the minimum intra subset Euclidean distance over all subsets at partitioning level i by d_i. The most popular labelling strategy is to maximize d_i at every partitioning level $i = 0,\ldots,\ell-1$, which we call Ungerböck Labelling (UL).

For MLC each partitioning level correspond to a coding level, i.e., the sequences $x^i[k]$, $k \in \mathbb{Z}$, of components of the sequence $\mathbf{x}[k]$ of address vectors are encoded individually by binary $(N, K^i, \delta^i_{\text{min}})$ codes $\mathcal{C}^i$ with (not necessary, but here assumed) equal lengths N, rates $R^i = K^i/N$ and min. Hamming distance δ^i_{min}, see Fig. 3. The total rate of the MLC scheme reads:

$$R = \sum_{i=0}^{\ell-1} R^i = \frac{K}{N} \left[\frac{\text{bit}}{\text{symbol}}\right] \quad \text{with} \quad K = \sum_{i=0}^{\ell-1} K^i.$$

2.1 Balanced Distances Rules

Since the very beginning of channel coding, a traditional paradigm has been to assess a code by its minimum distance, either minimum Hamming distance δ_{min} in a vector space over a finite field or minimum Euclidean distance $d_{E\,\text{min}}$ between codewords in signal space. Based on the min. Euclidean distances d_i of the subsets and the min. Hamming distances δ_i of the component codes for MLC a lower bound on $d_{E\,\text{min}}$ can easily be derived (see e.g. [19]):

$$d^2_{E\,\text{min}} \geq \min_i (d_i^2 \cdot \delta_i), \qquad i = 0,1,\ldots,\ell-1.$$

In order to maximize the minimum Euclidean distance, the component codes $\mathcal{C}^i$ have to be selected in that way that the products $d_i^2 \cdot \delta_i$ are balanced over all ℓ levels and the total rate meets the desired value R. This *balanced distances rule* for the rate design of MLC schemes was initially proposed and usually used. Throughout this paper, for illustrations and comparisons we use the following simple

Example: 4–ary ASK or 16–ary QAM, resp.

$$D = 1, \quad M = 4, \quad \mathbf{A} = \{-3,-1,+1,+3\}$$
$$R = 1.5\,\frac{\text{bit}}{\text{dim.}}$$

Assuming natural mapping which here corresponds to UL, component codes with length $N = 2000$, and minimum Hamming distances according to the Gilbert–Varshamov bound (see [30]) lead to the following parameters:

$$\left.\begin{aligned} R^0 &= 0.63: \; d_0^2\delta_0 = 4\cdot 142 = 568 \\ R^1 &= 0.87: \; d_1^2\delta_1 = 16\cdot 36 = 576 \end{aligned}\right\} d^2_{E\,\text{min}} \geq 568$$

That means, the theoretical asymptotic gain in power efficiency over uncoded transmission ($d^2_{E\,\text{min}} = 4$, $R = 2$) is $(568\cdot 1.5)/(4\cdot 2) \widehat{=} 20.3$ dB. Of course, such an asymptotic gain has no relevance in practice, (i) because a corresponding performance (for desired error rate 10^{-5}) would be about 9 dB beyond the Shannon limit for the AWGN channel and (ii) because of the exponential growth of the number of nearest neighbour error events, see [23]. In fact, it turns out that a scheme designed in this way shows poor performance, when MSD is applied. Thus, it is quite obvious that minimum Euclidean distance should not be used for criterion to design MLC schemes. Moreover, it becomes more and more clear that minimum distance is not the most important performance parameter in channel coding, cf. e.g. turbo codes [3] or the result in [4]. Thus, a change of basic paradigms should take place in coding theory, cf. also [1].

2.2 Capacity Design Rule

A communication channel is characterized by the mutual information $I(A;Y)$ between transmitted signal point $a \in \mathbf{A}$ (channel input variable) and discrete time channel output variable $y \in \mathbf{Y}$ (after matched filter, sampling, equalization). (Random variables are denoted by corresponding capital letters as usual.) For fixed or optimized distribution of channel input variable, $I(A;\,Y)$ equals the *capacity* C of the channel (per channel use). The mapping $\mathcal{M}$ may formally be included into the channel and, since mapping is bijective, the equation

$$I(A;Y) = I((X^0, X^1, \ldots, X^{\ell-1})\,;\,Y) \qquad (1)$$

holds. Applying the well known chain rule to mutual information [18, p. 22] yields

$$\begin{aligned} I((X^0, X^1, \ldots, X^{\ell-1})\,;\,Y) = \qquad & (2) \\ I(X^0;Y) + I(X^1;Y \mid X^0) + \ldots & \\ +I(X^{\ell-1};Y \mid (X^0, X^1, \ldots, X^{\ell-2})). & \end{aligned}$$

This equation allows the following interpretation: The transmission of vectors $\mathbf{x}$ of binary digits x^i over the physical channel is separable into a parallel transmission of individual digits x^i over *equivalent channels*, provided $x^0, \ldots, x^{i-1}$ are already known. The equivalent channels correspond to the coding levels. *Conditional* mutual information represents the capacity C^i of the equivalent channel $i = 0, 1, \ldots, (\ell - 1)$:

$$C^i := I(X^i;Y \mid (X^0, X^1, \ldots, X^{i-1}))$$

MLC **together with** MSD for coded modulation is a direct consequence of Eq. (2), although this method originally was derived in the context of minimum Euclidean distance [27]. Eqs. (1) and (2) immediately lead to a fundamental theorem which includes a rate design rule for MLC:

Theorem 1

The capacity C 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 of a multilevel coding scheme:

$$C = \sum_{i=0}^{\ell-1} C^i. \qquad (3)$$

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, $R^i = C^i$.

Proofs for equiprobable signal points are given in [24, 23, 51] (see also [29]), and equations for calculation of the capacities C^i can also be found therein. The general case is treated in [47].

This fundamental theorem has important consequences for the design of 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 very small subset of codes generated by the MLC approach — where $NR^i = K^i$ binary symbols are mapped to N signal point label component x^i independently for each level i — is a selection with asymptotic optimum performance. Here, Shannon's coding theorem is proved for well structured codes opposite to the originally used completely random codes, see [16] and [45], too.

2. Although in multistage decoding (MSD) code constraints at higher levels are not taken into account while decoding lower levels, it is sufficient to approach capacity. 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.

3. The theorem states that for any digital transmission scheme, where 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 non–binary codes for bandwidth efficient transmission systems. Starting from the huge field of good binary codes, its 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*.

4. The theorem implies no restriction on the particular labeling of signal points. Thus, mapping by set partitioning according to a special criterion [41] is not essential to approach capacity, cf. Section 4.

Example (continued):

Capacity $C = 1.5$ bit/(channel use) of the signal set $\mathbf{A}$ for the AWGN channel, which is achieved for $E_s/N_0 \mathrel{\widehat{=}} 6.46$ dB, divides into $C^0 = R^0 = 0.52$ and $C^1 = R^1 = 0.98$.

For codes of length $N = 2000$ a minimum Hamming distance ≥ 4.4 at level 1, say $\delta_1 = 5$ (cf. binary primitive BCH–codes, $\delta = 7$), results from Gilbert–Varshamov bound. Thus, minimum Euclidean distance is reduced to $d_{E\,\min} = 80$ and asymptotic gain to $15 \mathrel{\widehat{=}} 11.76$ dB, which is fairly compatible with capacity limit (for desired bit error rate 10^{-5}). On the other hand, a rate reduction at level 0 compensates for the exponential increase in nearest neighbour error events.

Fig. 5 shows simulation results for this design rule employing turbo codes of length 2000 and 20000 as component codes for MLC/QAM schemes in the power–bandwidth–plane spending 0.5 bit of redundancy per dimension. Although minimum Euclidean distance is further reduced by the poor minimum Hamming distance of turbo codes [37, 2], the power efficiency of these schemes in the range of high bandwidth efficiency , > 2 turn out to be as close to capacity for equiprobable signal points (boxes ■) as the direct application of turbo codes to binary antipodal signalling [3]. Now the advantages of efficient binary codes are utilizable by MLC together with simple MSD over the total range of bandwidth efficiency. Here, information *theory* leads to very efficient communication schemes in *practice* in a straightforward way. An application of shaping methods in order to produce a near discrete Gaussian distribution of signal points (see companion paper [12]) provides MLC schemes closer than 1 dB to the Shannon limit in the power–bandwidth plane, even slightly beyond capacity for equiprobable signal points. The diamond in Fig. 5 marks an example based on 64–ary QAM and , $= 4$ bit/s/Hz, i.e., spending 1 bit/dimension of redundancy for coding and shaping. Design details are given in [49, 12].

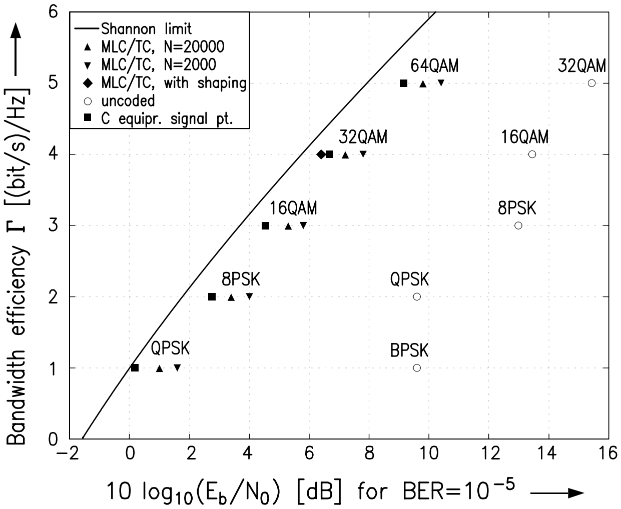

Figure 5: Power–bandwidth–plane for coded transmission over the AWGN channel with MLC/QAM schemes.

2.3 Design from Random Coding Exponent

An alternative approach to design MLC scheme is to balance word error probabilities p_w^i over all levels: $p_w^i = p_w$. In order not to be restricted to specific codes, bounds on random coding, may be applied. The random coding exponent $E_r^i(R^i)$, see [18], applied to the equivalent channel i additionally allows to consider finite codeword length [51, 47]. Individual rates R^i are obtained by numerical inversion of

$$2^{-NE_r^i(R^i)} \overset{!}{=} p_w. \qquad (4)$$

The results show that only for $N < 1000$ rates R^i slightly differ from that derived from capacity rule. For very short codewords and UL rate at lower levels increases whereas rate at higher levels decreases, i.e., rate design tends towards balanced distances rule.

Cut–off rates R_0^i may also be applied for the rate design: $R^i = R_0^i$. The distribution of individual rates is almost indistinguishable from a capacity design, too.

2.4 Bounds on Error Probabilities

As long as error propagation from lower to higher levels is neglected, all levels of a MSD should contribute equally to the error probability at minimum acceptable reliability. Such a design for balanced error rates requires sufficiently tight bounds on the individual error rates. The multiple representation of binary symbols at low levels according to information mapped to higher levels cause an exponential increase in the number of possible error events. Thus, key point is the calculation of the profiles of Euclidean distances at individual levels of MLC based on the profiles of Hamming distances of the corresponding component codes. In [36, 23] an analytical analysis of the distance profile is derived, but the effect of multiple representation of correct symbols was not taken into account. Biglieri et al. presented a complete analysis for component codes based on lattice constellations, thus neglecting boundary effects in real constellation. In [49] a refinement of these previous results is presented.

It turns out that the so–called minimum sufficient distance profile for an error rate estimation using the union bound (see [9]) is closely related to the result in [23]. Simulation results presented

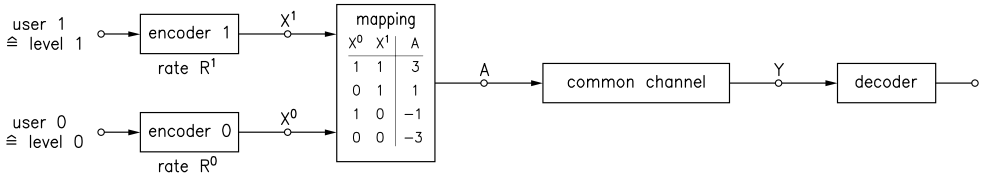

Figure 6: MLC as multiple access problem.

in [23, 49] indicate that the union bound technique based on these distance profiles promise tight estimates in the range of error probabilities $< 10^{-3}$. At low SNR, a tightening of the union bound according to [20, 21] helps to improve the estimation. Using correct distance profiles, i.e. including multiple representation of binary symbols by signal points, a rate design of MLC schemes for balanced error probabilities yields results almost not distinguishable from capacity rule.

2.5 Capacity Regions

A MLC design according to capacity rule (Section 2.2) leads to a poor minimum distance when compared to the balanced distances design rule. Codewords form a more or less irregular structure of 2^{NR} points in ND dimensions, being far away from dense lattices. Thus, Theorem 1 seems to be in contradiction to the well–known theorem of de Buda [10]. Therefore, we have to discuss conditions on the optimality of MLC in more detail.

Coded modulation with ℓ levels may be interpreted as a special multiple access problem, i.e., ℓ binary output sequences access a common channel via the mapping [29], cf. also Fig. 6. Using again our simple two–level example, rates R^0 and R^1 of both "users" are restricted by following facts:
(i) Total rate cannot exceed mutual information (capacity) provided by the common channel.
(ii) Rate for one user cannot exceed mutual information provided the message of the other one is known at the receiver side. Thus we have the conditions:

$$\begin{array}{c} R^0 + R^1 \le I((X^0, X^1); Y) \\ R^0 \le I(X^0; Y \mid X^1), \quad R^1 \le I(X^1; Y \mid X^0) \end{array} \qquad (5)$$

which form the typical polygon in the rate plane, known from multiuser information theory, see Fig. 7. Points $\boxed{1}$ and $\boxed{2}$ correspond to the chain rule Eq. (2) for both possible expansions and thus are achievable due to Theorem 1 even by suboptimum MSD. (Point $\boxed{2}$ simply corresponds to an exchange of components X^0 and X^1 when compared to point $\boxed{1}$, i.e., to a different mapping strategy, see Section 4. But mapping does not affect capacity, see Theorem 1.)

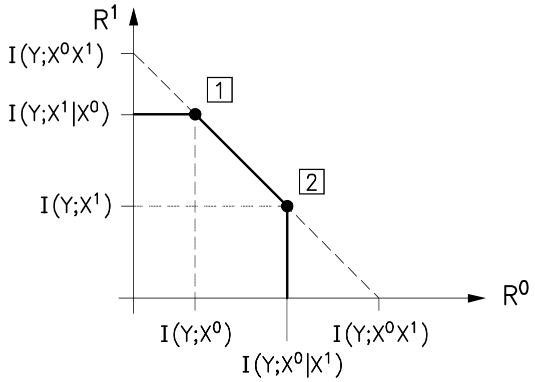

Figure 7: Capacity region for two–level coding.

In [47], we give the proof that all MLC schemes with rate design corresponding to the points on the straight line between $\boxed{1}$ and $\boxed{2}$ achieve capacity, too. Usually, such MLC schemes include those designed according to the balanced distances rule. Hence, even the balanced distances rule leads to optimum schemes. Thus, the ostensible contradiction to [10] is resolved and dense lattices lead towards capacity, too. But there is an important difference between points along the line and the vertices $\boxed{1}$ and $\boxed{2}$: As for points between $\boxed{1}$ and $\boxed{2}$ $R^0 > I(X^0; Y)$ as well as $R^1 > I(X^1; Y)$ holds, none of both sequences can successfully be decoded without taking the other one into account, i.e., MSD does not work at all. A maximum likelihood decoding over all levels with tremendous complexity is indispensable for all designs except the vertices of the polygon. All ideas to overcome the poor performance of MLC schemes with an unfortunate rate design, like forwarding of reliability information from lower to higher levels or/and iterative decoding [35, 54, 55, 46] may be interpreted as attempts to approximate maximum likelihood overall decoding in a similar way as it is done in turbo decoding. Unfortunately, besides increased complexity these methods suffer from the necessity of a quasi–perfect interleaving between the levels which introduces a tremendous delay of data. But in most applications data delay is restricted, and thereby performance loss due to short codewords and/or insufficient interleaving has to be taken into consideration. Please notice, all these problems are completely avoided, when a proper rate design according to Sections 2.2 to 2.4 is applied. Here, such extensions to MSD neither are necessary nor can substantially improve performance.

3 Dimensionality of the Constituent Signal Constellation

In contrast to TCM, dimensionality D of the constituent signal constellation and rate R of MLC schemes are completely decoupled. Therefore, the important question arises how to choose D? Concerning performance, this question cannot be answered by capacity arguments, because Eq. (2) leads to equal results for all possible values of D. But an answer for finite code word length can be given from random coding exponent and cut–off rate. A choice of D as small as possible yields minimum number of levels, i.e., of component codes and therefore may be preferred from complexity reasons.

For a fair comparison of MLC schemes employing block codes we fix the number $N_D D$ of dimensions of signal space, where N_D denotes the code word length of a MLC scheme based on a D–dimensional signal set. For our example (4–ary ASK or 16–ary QAM, resp.) sums of rates $\sum_{i=0}^{2^D-1} R^i$, calculated from random coding exponent (Eq. (4)), $p_w = 10^{-4}$ and $N_1 = 4000$ are shown in Fig. 8. The one–dimensional approach, $D = 1$, promises a performance gain of about 0.2 dB over $D = 2$. Thus, the simpler solution turns out to be the better one! Constraints over N_1 dimensions are more efficient embedded into binary component codes than into the constellation itself. This result has been verified by simulation using turbo codes for component codes, see [11] and Fig. 11. One–dimensional MLC only is possible for square QAM constellations over the lattice $(2\mathbb{Z}+1)^2$ in a direct way. But even for other constellations (e.g. cross–constellation) partitioning strategies exist which at least for the two lowest levels allow a multiplexing of inphase and quadrature component in one coding level. In [50] we show that only for the lowest level of a one–dimensional constellation, sophisticated coding, such as soft decision maximum likelihood decoding is necessary, whereas simple FEC schemes are sufficient at higher levels. Thus, the advantage using a one–dimensional scheme in performance and complexity can be exploited for every bandwidth efficient modulation schemes. In [48] examples are presented for the 32–ary cross constellation. Summarizing, multidimensional constituent constellations are neither necessary nor useful for coded modulation via MLC with block component codes.

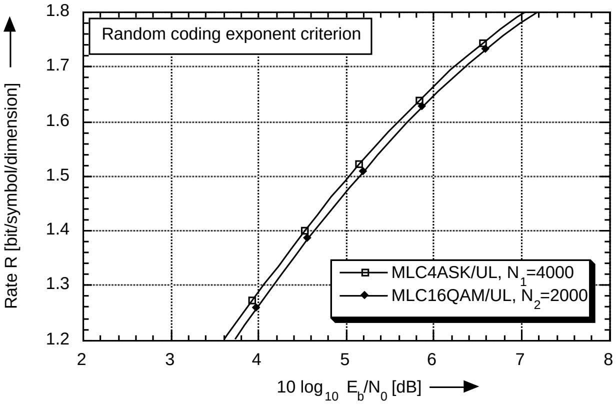

Figure 8: Sum R/D of individual rates R^i/D for MLC/4ASK/UL, $N_1 = 4000$ and for MLC/16QAM/UL with $N_2 = 2000$ derived via random coding exponents. AWGN channel.

If the comparison is done for cut–off rates the situation is quite different: The sum of cut–off rates increases with the dimensionality D of the constituent constellation, cf. [32]. Cut–off rate is an appropriate parameter for convolutional codes. For increasing D the number of codes increases and by this the effective constraint length, i.e., constraint length over channel symbols. (Data delay increases, too, but usually remains tolerably small.) Thus, it is not surprising that an effect opposite to block codes is observed.

4 Labelling Strategies

From Theorem 1, it follows that the strategy how to map address vectors $\mathbf{x}$ to M signal point does not affect total mutual information or capacity at all. Thus, from a capacity point of view labelling strategies like set partitioning for maximum intra subset minimum Euclidean distance (UL) seems to be not relevant in coded modulation. But this statement is only true for infinite codeword length. Using the random coding exponent, several labelling strategies can be compared for fixed length of the component codes. For our example Fig. 9 shows the required E_b/N_0 to achieve a reliability $p_w \leq 10^{-3}$ versus code word length N. Here, GL denotes Gray labelling; i.e., addres-

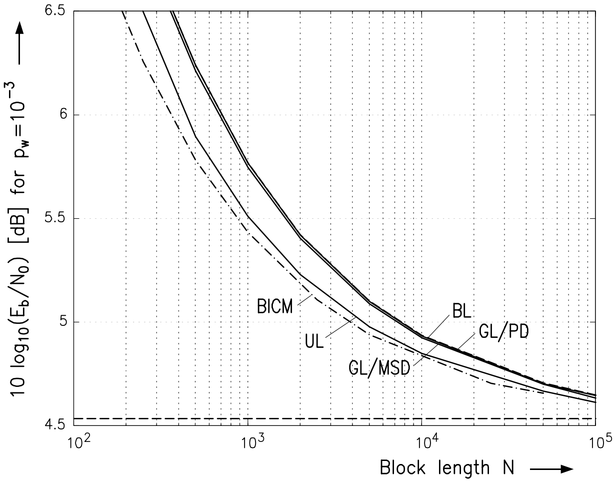

Figure 9: Required SNR of various coded modulation schemes derived from random coding exponent for 4–ary ASK, $R = 1.5$ bit/dim.

ses of neighboring signal points differ only in one binary digit, see Fig. 10. BL is an abbreviation for block labelling, i.e., keeping signal points in the subsets as close together as possible (set partitioning for minimum intra subset variance — the opposite to UL). Standard UL turns out to be superior over other labelling strategies and the gap increases for short codes, but the differences are surprisingly small. BL inverts the order of rates (in our example 4–ary ASK: $R^0 > R^1$) which can be applied to construct softly degrading MLC schemes, see [23, 26]. Here, decoding in the MSD procedure is done only up to that level which still delivers reliable data. Examples for softly degrading schemes via MLC together with optimization of constellations and codes can also be found e.g. in [40, 38]. Summarizing, labelling strategy has usually a quite smaller importance in coded modulation than stated in many papers, except the special case discussed in the next section. Of course, UL is an essential method to separate coded and uncoded levels, thus to save complexity. The approach presented in [50] to simplify MLC schemes by using only one sophisticated binary coding scheme at the lowest level can also only be applied for UL without noticeable performance loss.

5 Parallel Decoding and Bit Interleaved Coded Modulation

Gray labelling (GL), see the example in Fig. 10, usually leads to an irregular set partitioning tree, i.e., the subsets at one partitioning label are not congruent and provide unequal capacities on an additive noise channel. It is inherent to GL that for all signal points and components of the binary address vectors there is at most one point with minimum distance d_0 representing the inverse binary symbol. This involves that there is no exponential increase in the number of nearest neighbour error events due to multiple representation of binary symbols in MLC schemes. In particular, neither the error coefficient nor the minimum Euclidean distance changes when decisions at lower levels are taken into account for decoding higher levels. Hence, decoding for each level may be based on the **entire** signal constellation without any preselection of signal points at higher levels by decisions at lower levels. Therefore, a suboptimum *parallel decoding* (PD) of all levels is possible with only small performance loss. Mutual information I_p utilizable by PD corresponds to a lower bound on the expansion (2)

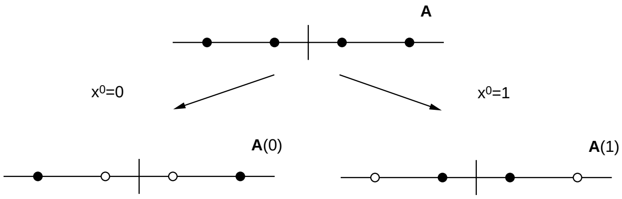

Figure 10: Set partitioning for 4–ary ASK with Gray labeling.

$$\begin{aligned} I((X^0, X^1, \ldots, X^{\ell-1}); Y) \geq & \qquad (6) \\ I(X^0; Y) + I(X^1; Y) + \ldots + I(X^{\ell-1}; Y) =: I_p. & \end{aligned}$$

Surprisingly, $I(\mathbf{X}; Y)$ and I_p differ only very slightly for GL. In the reasonable range of 0.5 bit/dimension of redundancy this difference typically corresponds to a loss < 0.1 dB and, hence, is irrelevant in practice. Furthermore, PD is not affected by error propagation. This statement also holds for finite codeword lengths as SNR curves derived from random coding exponent also show these small differences, see Fig. 9, comparing curves GL/MSD and GL/PD.

On the one hand PD offers the advantage of reducing delay of the data stream due to the decoding procedure by parallel processing of all levels. But usually this reduction usually will be small, since for UL and a one–dimensional constituent constellation ($D = 1$) only decoding time for level 1 is saved which may be very short, when simple FEC is applied at this level, cf. [50]. On the other hand, GL together with PD is well suited for communication over fading channels because code rates $R^i = I(X^i;Y)$ for all levels differ significantly from 1 and inherently high time diversity due to independent binary coding at all levels is achieved, see [39, 52].

Eq. (6) for I_p utilizable by PD exactly corresponds to the formula for the mutual information of a time–variant, memoryless (i.e. perfectly interleaved) channel with equiprobable discrete states $0, 1, \ldots, \ell - 1$, provided the actual state is known at the receiver. Interleaving together with a mapping of ℓ binary digits to the components of the address vector $\mathbf{x}$ in time–multiplex and the inverse process at the receiver side produce such a time–variant channel with mutual information I_p/ℓ per binary input symbol. Following the channel coding theorem a single binary code with rate $R_{\text{BICM}} \leq I_p/\ell$ is sufficient for reliable communication. This method called $\underline{\text{b}}$it $\underline{\text{i}}$nterleaved $\underline{\text{c}}$oded $\underline{\text{m}}$odulation (BICM) was proposed in [56] as an alternative approach to TCM over the Rayleigh fading channel using one convolutional code. Later it was recognized by [5] that BICM over the AWGN channel provides mutual information very close to capacity. The equivalence of BICM and MLC/PD is derived in [48, 39]. If the number N of channel symbols included in one codeword is high, no further increase of the blocklength needs to be introduced by interleaving, because in this case averaging of channel states (i.e. levels) within one codeword is sufficient. Interblock interleaving may be indicated here only for blockcodes being sensitive to special the error pattern, e.g. turbo codes. BICM offers the benefit that multiplexing levels in time results in an increased length of the binary code $N_{BICM} = \ell \cdot N$. Due to this effect the min. SNR calculated from the random coding exponent of BICM in Fig. 9 promises performance even better than for UL/MSD at a fixed number of channel symbols within one codeword. But, in simulating using turbo codes of length 4000 for component codes in UL/MSD (4–ary ASK, $R^0 = 0.52$, $R^1 = 0.98$) and 8000 for BICM ($R_{\text{BICM}} = 0.75$), UL/MSD is slightly superior by about 0.2 dB at bit error rate 10^{-5}, see Fig. 11. In this example, BICM achieves capacity for equiprobable signal points within 1.4 dB. On the other hand, BICM clearly outperforms BL/MSD and GL/PD as indicated in Fig. 9.

For simplicity, PD and BICM are discussed here to comprise all levels. For the AWGN channel and very large signal constellation it is more efficient to apply a mixed labelling strategy, i.e., to apply UL in order to separate coded from uncoded levels and by this to save complexity. Within coded levels a relabeling according to the Gray criterion allows to apply PD or BICM for subset coding/decoding. For fading channels GL/PD or BICM over all levels is recommended.

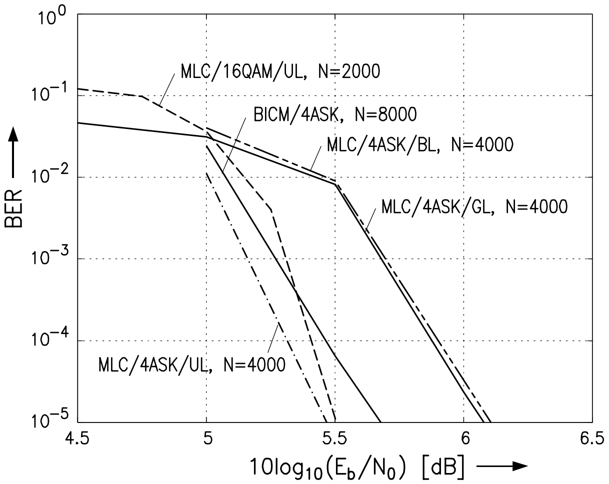

Figure 11: Simulation results: bit error rate over signal–to–noise ratio E_b/N_0.

Obviously, BICM is nothing else but the traditional approach shown in Fig. 1, i.e., the state of the art before the topic *coded modulation* was born in digital communications. Thus the progress in theory brought us back exactly to the status more than 24 years ago! Therefore, following questions arise: (i) What are the reasons for the poor performance of the traditional approach in the past? (ii) Did the topic *coded modulation* really exist at any time or was it simply a fata morgana, during the last 24 years? (iii) Have Caire, Taricco and Biglieri brought coded modulation to its end and will it now vanish at all?

Contradiction (i) is simply resolved by the observation, that binary FEC block coding with rate 0.5 for binary antipodal signalling per dimension together with $\underline{\text{b}}$ounded $\underline{\text{m}}$inimum $\underline{\text{d}}$istance decoding (BMD) exhibits a performance gap of about 5.5 – 7.5 dB to capacity limit for BER = 10^{-5}, see Fig. 2. In [50] we show that interleaving and hard decision decoding for bandwidth efficient schemes with 0.5 bit/dimension of redundancy causes the same capacity loss of about 2 dB as for binary signalling per dimension. Thus, it is not surprising that a similar distance of 5.5 – 7.5 dB to capacity limits can also be observed for schemes with higher bandwidth efficiency. But in this area there is a smaller gap between uncoded transmission (BER = 10^{-5}) and capacity which result in a poorer coding gain. Additionally, the shaping gap here arises. Therefore, the insufficient results for the traditional approaches fit quite well into the framework of quasi–optimality of BICM, as the gap completely is caused by the poor performance of the binary coding scheme itself and **not** by an improper combining of coding with modulation. The reason for the gap being greater than 2 dB is the lack of powerfull hard–decision near maximum likelihood decoding algorithms for long block FEC codes with minimum distance close to Gilbert–Varshamov bound. Additionally, in the range of high bandwidth efficiency competitive uncoded schemes with smaller constellations exist at the same bandwidth which amplifies the impression of poor performance of FEC schemes. Now, the gap to capacity limit for more than 2 signal points per dimension is more and more bridged by BICM based on long binary codes with quasi–random properties and near maximum likelihood soft decision decoding techniques.

Concerning question (ii) the topic coded modulation indeed did not really exist at all as long as only power efficiency is taken into account. But "classical" coded modulation is still a powerful method to save complexity as coding is restricted to that data really affected by noise. Additionally, BICM does not work at all if data delay is closely restricted. For example, BICM does not offer sufficient performance using a binary convolutional code, because of restricted constraint length. In this case, MLC/MSD or TCM clearly is the better choice. This statement includes the answer to question (iii): In practical applications under various restrictions coded modulation will still be a hot topic over a couple of years. But the authors think that indeed the field coded modulation may have lost some of its charm in theory by the paper [5].

6 Conclusions

In coded modulation a wide arch of research work now have us led back to the origin. Optimum or near optimum approaches to coding for bandwidth efficient communication would have been available since the early days, but were not recognized before 1994. Almost the entire research community was fixed to minimum Euclidean distances and asymptotic gains and by thus did not see the potential of MLC and BICM over a long time. Strict application of methods from information theory helped to overcome the traditional coding paradigms [29, 24], a further example of the inestimable benefits to be gained from theory for the technical progress. Nowadays we see many different ways to coded bandwidth efficient communication schemes with excellent power efficiency, i.e., MLC with almost arbitrary labeling strategies, independent parallel decoding and BICM, which are feasible in practice. This development was accompanied by the discovery and refinement of turbo codes which is characterized by the same change in the paradigm of optimality, i.e., not to pay too much attention on minimum distances. Employment of turbo codes for component codes in MLC or BICM together with a proper rate design leads to practicable schemes close to existence limits for all desired values of bandwidth efficiency.

For the AWGN channel MLC/MSD offers the advantage over BICM that only the lowest level needs to be encoded by sophisticated complex schemes, whereas BICM has at least to include two levels. On the other hand BICM over all levels is a very efficient solution for fading channels, thus, BICM offers more flexibility. If there are strong restrictions on data delay TCM or equivalently MLC/UL/MSD with convolutional codes as component codes and a rate design from the cut–off rate criterion still are proper choices. Here, MLC offers more flexibility in total rate.

Work on coded modulation is not finished by MLC or BICM. E.g., research has to be extended to more realistic channel models; there is a lot of work to refine equalization and synchronization techniques for such extreme low SNR that modern coding techniques allow to tolerate. Summarizing, there are still many further topics for paper to be presented in sessions on coded modulation at future ITG–conferences on source and channel coding.

Acknowledgment

The authors are greatly indebted to *Deutsche Forschungsgemeinschaft* for supporting this work under contract Hu 634/1-2.

References

[1] G. Battail. On random–like codes. In *Proc. 4-th Canad. Workshop Inf. Theory*, May 1995.

[2] S. Benedetto and G. Montorsi. Unveiling Turbo Codes: Some Results on Parallel Concatenated Coding Schemes. *IEEE Trans. Inf. Theory*, vol.42:pp.409–428, 1996.

[3] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes. In *Proc. IEEE Int. Conf. Commun. (ICC)*, pages 1064–1070, Geneva, Switzerland, May 1993.

[4] T. Beth and D.E. Lazic. If Binary Codes Existed That Exceed Gilbert–Varshamov Bound They Could Not Reach the Cutoff Rate Of BSC. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, page 132, Whistler,Canada, Sept. 1995.

[5] G. Caire, G. Taricco, and E. Biglieri. Capacity of bit–interleaved channels. *Electronics Letters*, vol.32:pp.1060–1061, June 1996.

[6] A.R. Calderbank, T.A. Lee, and J.E. Mazo. Baseband Trellis Codes with a Spectral Null at Zero. *IEEE Trans. Inf. Theory*, vol.34:pp.425–434, 1988.

[7] 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.

[8] A.R. Calderbank and N.J.A. Sloane. An eight–dimensional trellis code. *Proc. of the IEEE*, vol.74:pp.757–759, May 1986.

[9] J.H. Conway and N.J.A. Sloane. *Sphere Packings, Lattices and Groups.* Springer Verlag, New York, 1988.

[10] R. de Buda. Some optimal codes have structure. *IEEE J. Select. Areas Commun.*, vol.7:pp.893–899, Aug. 1989.

[11] R. Fischer, J. Huber, and U. Wachsmann. Multilevel coding: Aspects from information theory. In *Proc. CTMC at IEEE Global Telecommun. Conf. (GLOBECOM'96)*, pages 26–30, London, UK, Nov. 1996.

[12] R. Fischer, J. Huber, and U. Wachsmann. On the Combination of Multilevel Coding and Signal Shaping. In *Proc. 2nd ITG Conf. Source and Channel Coding*, Aachen, Germany, March 1998.

[13] G.D. Forney, Jr. Coset codes – part I: Introduction and geometrical classification. *IEEE Trans. Inf. Theory*, vol.34:pp.1123–1151, Sept. 1988.

[14] G.D. Forney, Jr. Coset codes – part II: Binary lattices and related codes. *IEEE Trans. Inf. Theory*, vol.34:pp.1152–1187, Sept. 1988.

[15] G.D. Forney, Jr. Multidimensional constellations—part II: Voronoi constellations. *IEEE J. Select. Areas Commun.*, vol.SAC-7:pp.941–958, Aug. 1989.

[16] G.D. Forney, Jr. Approaching the Capacity of the AWGN Channel with Coset Codes and Multilevel Coset Codes. Submitted to IEEE Trans. Inf. Theory, 1996.

[17] 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.SAC-7:pp.877–892, Aug. 1989.

[18] R. G. Gallager. *Information Theory and Reliable Communication.* John Wiley & Sons, Inc., New York, 1968.

[19] 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).

[20] H. Herzberg and G. Poltyrev. Techniques of Bounding the Decoding Error for Block Coded Modulation. *IEEE Trans. Inf. Theory*, vol.40:pp.903–911, May 1994.

[21] H. Herzberg and G. Poltyrev. On the Error Probability of M–ary PSK Block Coded Modulation. *IEEE Trans. Commun.*, vol.COM-44:pp.427–433, April 1996.

[22] J. Huber. *Trelliscodierung in der digitalen Übertragungstechnik — Grundlagen und Anwendungen.* Springer Verlag, Berlin, 1992.

[23] J. Huber. Multilevel Codes: Distance Profiles and Channel Capacity. In *ITG-Fachbericht 130*, pages 305–319, München, Oct. 1994.

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

[25] J. Huber and U. Wachsmann. Design of Multilevel Codes. In *Proc. IEEE Inf. Theory Workshop (ITW)*, Rydzyna, Poland, June 1995. paper 4.6.

[26] 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.-March 1996.

[27] H. Imai and S. Hirakawa. A new multilevel coding method using error correcting codes. *IEEE Trans. Inf. Theory*, vol.23:pp.371–377, May 1977.

[28] T. Kasami, T. Takata, T. Fujiwara, and S. Lin. On Multilevel Block Coded Modulation Codes. *IEEE Trans. Inf. Theory*, vol.37:pp.965–975, July 1991.

[29] Y. Kofman, E. Zehavi, and S. Shamai (Shitz). Performance Analysis of a Multilevel Coded Modulation System. *IEEE Trans. Commun.*, vol.42:pp.299–312, 1994.

[30] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error–Correcting Codes*. North–Holland, Amsterdam, 1978.

[31] J.L. Massey. Coding and modulation in digital communications. In *1974 Intern. Zürich Seminar on Digital Communications*, Zürich, Switzerland, March 1974.

[32] J. Persson. *Multilevel Coding Based on Convolutional Codes*. PhD thesis, Lund University, Sweden, June 1996.

[33] S.S. Pietrobon and D.J. Costello, Jr. Trellis coding with multidimensional QAM signal sets. *IEEE Trans. Inf. Theory*, vol.39:pp.325–336, Mar. 1993.

[34] S.S. Pietrobon, R.H. Deng, A. Lafanechére, G. Ungerböck, and D.J. Costello. Trellis–coded multidimensional phase modulation. *IEEE Trans. Inf. Theory*, vol.36:pp.63–89, 1990.

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

[36] K. Rebhan. *Schrittweise Decodierung von Mehrstufencodes*. Diplomarbeit, Lehrstuhl für Nachrichtentechnik der Universität Erlangen-Nürnberg, Okt. 1993.

[37] P. Robertson. Illuminating the Structure of Code and Decoder for Parallel Concatenated Recursive Systematic (Turbo) Codes. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pages 1298–1304, San Francisco, Nov./Dec. 1994.

[38] D.W. Schill and J.B. Huber. On Hierarchical Signal Constellations For the Gaussian Broadcast Channel. Subm. to Int. Conf. Telecommun. June 1998.

[39] P. Schramm. Multilevel Coding with Independent Decoding on Levels for Efficient Communication on Static and Interleaved Fading Channels. In *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, pages 1196–1200, Helsinki, Finland, Sept. 1997.

[40] A. Seeger. A new signal constellation for the hierarchical transmission of two equally sized data streams. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, page 169, Ulm, Germany, July 1997.

[41] G. Ungerböck. Channel coding with multilevel/phase signals. *IEEE Trans. Inf. Theory*, vol.28:pp.55–67, Jan. 1982.

[42] G. Ungerböck. Trellis coded modulation with redundant signal sets, part I. *IEEE Commun. Mag.*, 25:pp.5–11, Feb. 1987.

[43] G. Ungerböck. Trellis coded modulation with redundant signal sets, part II. *IEEE Commun. Mag.*, 25:pp.12–21, Feb. 1987.

[44] G. Ungerböck and I. Csajka. On improving data–link performance by increasing channel alphabet and introducing sequence coding. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Ronneby, Sweden, June 1976.

[45] R. Urbanke and B. Rimoldi. Lattice Codes can achieve Capacity on the AWGN channel. Submitted to IEEE Trans. Inf. Theory, 1996.

[46] B. Vucetic, E.J. Leonardo, and L. Zhang. Iterative decoding of block codes. In *Proc. IEEE Inf. Theory Workshop (ITW)*, Rydzyna, Poland, 1995.

[47] U. Wachsmann, R. Fischer, and J. Huber. Multilevel Coding: Basic Concepts, Capacity, and Rate Design. Submitted to IEEE Trans. Inf. Theory, 1997.

[48] U. Wachsmann, R. Fischer, and J. Huber. Multilevel Coding: Dimensionality of the Constituent Signal Set, Labeling, and Hard Decision Decoding. Submitted to IEEE Trans. Inf. Theory, 1997.

[49] U. Wachsmann, R. Fischer, and J. Huber. Multilevel Coding: Distance Profile and Signal Shaping. Submitted to IEEE Trans. Inf. Theory, 1997.

[50] U. Wachsmann, R. Fischer, and J. Huber. Multilevel Coding: Use of Hard Decisions in Multistage Decoding. In *Proc. 2nd ITG Conf. Source and Channel Coding*, Aachen, Germany, March 1998.

[51] 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.

[52] U. Wachsmann, P. Schramm, and J. Huber. Comparison of Coded Modulation Schemes for the AWGN and the Rayleigh Fading Channel. Subm. to Int. Symp. Inf. Theory, August 1998.

[53] T. Wörz and J. Hagenauer. Multistage coding and decoding. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pages 501.1.1–501.1.6, San Diego, Dec. 1990.

[54] T. Wörz and J. Hagenauer. Iterative Decoding for Multilevel Codes using Reliability Information. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Orlando, Dec. 1992.

[55] T. Wörz and J. Hagenauer. Decoding of M-PSK-multilevel codes. *Europ. Trans. Telecommun. (ETT)*, vol.4:pp.65–74, May-June 1993.

[56] E. Zehavi. 8–PSK Trellis Codes for a Rayleigh Channel. *IEEE Trans. Commun.*, vol.40:pp.873–884, May 1992.