Coded Modulation for Differential Encoding and Non–Coherent Reception on Fading Channels

Robert F.H. Fischer, Lutz H.-J. Lampe, Stefan H. Müller–Weinfurtner, Johannes B. Huber

Lehrstuhl für Nachrichtentechnik II, Universität Erlangen–Nürnberg,
Cauerstraße 7/NT, 91058 Erlangen, Germany

***Abstract* — In this paper, multilevel encoded transmission over Rician fading channels is regarded. Especially, we focus on differentially encoded APSK constellations, and non–coherent reception without any channel state information at the receiver. Additional gains can be achieved if interleaving and detection are based on blocks of N consecutive symbols. This multiple–symbol differential detection can be applied to slowly time–varying channels. Regarding channel capacity as an appropriate performance measure when powerful component codes are applied in the multilevel coding scheme, we discuss several aspects, such as the differential encoding strategy, the mapping rule, and the optimization of the constellation. We show, that the competing coding scheme bit–interleaved coded modulation is not well suited for multiple–symbol differential detection, because no Gray labeling is possible for multiple–symbol differential detection and APSK constellations. In contrast to this, power and bandwidth efficient digital transmission close to capacity limit over fading channels using differential encoding is possible if properly designed MLC schemes are employed. The theoretic considerations are confirmed by means of simulations, where turbo codes are applied as component codes.**

1 Introduction

There is a great number of scenarios where transmission over fading channels has to be performed without having channel state information and/or reliable carrier phase estimation at the receiver. In these situations, differential encoding at the transmitter and non–coherent reception have proved to be advantageous. Moreover, in order to achieve high *bandwidth efficiencies*, mixed phase and amplitude modulation can be used. A straightforward extension of 'classical' differential phase–shift keying (DPSK) is to transmit information both in phase and in amplitude changes. This strategy is known as differential amplitude- and phase–shift keying (DAPSK), e.g. [1].

Power efficient transmission can be designed if channel coding and modulation are jointly optimized. Known *coded modulation* schemes are trellis–coded modulation (TCM) [2] and *multilevel coding (MLC)* [3]. Recently, a pragmatic approach called *bit–interleaved coded modulation (BICM)* [4, 5] has gained attention. If the underlying channel is slowly time–varying, power efficiency of differential schemes can further be improved by basing demodulation and decoding on blocks of N consecutive symbols, cf. e.g. [6], which is called *multiple–symbol differential detection*.

In this paper, multilevel coding for power and bandwidth efficient transmission using differentially encoded APSK constellations on flat fading channels is considered. We focus on the case where no channel state information is available at the receiver. Key aspects, such as the differential encoding strategy, the mapping rule, achievable capacity and decoding procedure are illuminated. We show, that in contrast to MLC, BICM is not well suited for multiple–symbol differential detection. It is pointed out, that power and bandwidth efficient digital transmission close to capacity limit over fading channels using differential encoding is possible if properly designed MLC schemes are employed. Simulations, using turbo codes [7] as component codes, affirm the theoretical statements.

In Section 2, the system model and encoding strategies are presented. Based on this, in Section 3, the capacity of the overall channel, including the actual channel and differential encoding/demodulation, is calculated. Multilevel coding and bit–interleaved coded modulation are reviewed in Section 4. Finally, Section 5 presents numerical and simulation results.

2 System Model and Differential Encoding Strategies

The system model regarded in this paper is plotted in Figure 1. The underlying channel is described by its complex

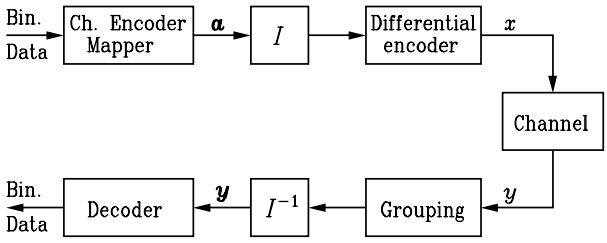

Figure 1: System model.

valued discrete–time model (i.e., in the equivalent low–pass domain) [8]. We assume a stationary, slowly time–varying, frequency non–selective (non–dispersive or flat) Rician fading channel. As usual, channel state and carrier phase offset are expected to be constant for at least N symbols.

We investigate the situation where neither channel state information nor reliable estimation of the carrier phase offset are available at the receiver side. For these scenarios, differential encoding at the transmitter and non–coherent demodulation at the receiver are convenient.

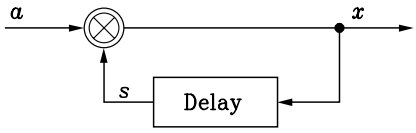

Figure 2: Differential encoder.

In *differential phase–shift keying (DPSK)* information is carried in *phase changes* rather than in the phase itself. For it, a phase accumulation is performed at the transmitter. Choosing the PSK symbols equally spaced on the unit circle, this so–called *differential encoding* is usually done by multiplying the data–carrying *differential symbol* (phase increment) by the previous transmit symbol, which also can be regarded as *state* of the encoder or phase reference.

The concept of differential encoding can easily be extended to constellations for mixed amplitude/phase modulation. Let s be the current state of the encoder (the reference symbol) and $a \in \mathcal{A}$ a differential symbol drawn from the signal set $\mathcal{A}$. From the tuple (a, s) the encoder generates the transmit symbol $x \in \mathcal{X}$, which also constitutes the next state of the differential encoder, and hence $s \in \mathcal{X}$. We formally specify the relation between a, s, and x by the operator "$\otimes$", with

$$x = s \otimes a \ , \tag{1}$$

where "$\otimes$" is a (possibly non–commutative) operation with

$$\otimes : \mathcal{X} \times \mathcal{A} \to \mathcal{X} \ , \tag{2}$$

see Figure 2. For usual DPSK, "$\otimes$" simply is the complex multiplication.

Especially for fading channels, constellations $\mathcal{C}$ consisting of $M = \alpha \cdot \beta$ points arranged in α concentric rings with radii r_i, $i = 0, \ldots, \alpha - 1$, $r_i < r_{i+1}$, and β uniformly spaced phases $\varphi_m = \frac{2\pi}{\beta} m$, $m = 0, \ldots, \beta - 1$, are preferred [9]:

$$\mathcal{C}_{\mathrm{APSK}}(\alpha, \beta) \triangleq \left\{ c = r_i \mathrm{e}^{\mathrm{j}\frac{2\pi}{\beta}m} \,\middle|\, \begin{smallmatrix} i \in \{0,1,\ldots,\alpha-1\} \\ m \in \{0,1,\ldots,\beta-1\} \end{smallmatrix} \right\} \ . \tag{3}$$

In order to avoid ambiguities, we denote these constellations by αAβPSK.

Here, for brevity, we restrict to $\mathcal{A} = \mathcal{X} = \mathcal{C}_{\mathrm{APSK}}$, i.e., the differential symbols a and the transmit symbols x are taken from the same APSK signal constellation. This is true, e.g., for classical DPSK, but there are also schemes for which both constellations differ (e.g., $\frac{\pi}{4}$–DQPSK [10]).

2.1 Differential Amplitude and Phase Encoding

A straightforward extension of classical DPSK is to perform differential encoding over phase and amplitude. In this case the differential symbols a give phase *and* amplitude increments. Like the phase, the amplitude is incremented cyclical, too. We denote this strategy by *differential amplitude and phase–shift keying (DAPSK)*.

Let the encoder state be $s = r_i \mathrm{e}^{\mathrm{j}\varphi_n}$ and the differential symbol $a = r_j \mathrm{e}^{\mathrm{j}\varphi_m}$. Now, the transmit symbol x (and, thus, implicitly the operator "$\otimes$") is given by

$$x = s \otimes a = r_i \mathrm{e}^{\mathrm{j}\varphi_n} \otimes r_j \mathrm{e}^{\mathrm{j}\varphi_m} \triangleq r_{(i+j) \bmod \alpha} \mathrm{e}^{\mathrm{j}\varphi_{(n+m) \bmod \beta}} \ . \tag{4}$$

2.2 Differential Phase and Absolute Amplitude Encoding

When designing a transmission scheme with non–coherent reception, the carrier phase is assumed to be completely unknown. In order to resolve ambiguities, differential encoding over the phase is strictly necessary. But, even if no channel state information is available at the receiver, the amplitude of the received signal can still provide information on the transmitted amplitude. This is because on the one hand the channel gain is non–uniformly distributed. On the other hand the observation of amplitude transitions enables a classification of the transmitted magnitude. Increasing N, the length of the observation interval at the receiver, this classification becomes more reliable since the observation of amplitude changes becomes more and more likely.

Thus, mapping the information onto the *phase change* and onto the *actual amplitude* might be advantageous in some situations. It should be noted that there are also scenarios for which this approach is not well suited, e.g., Rayleigh fading channels with high signal–to–noise ratios (see discussion below). Now, the definition (4) of the operator has to be modified to

$$x = s \otimes a = r_i \mathrm{e}^{\mathrm{j}\varphi_n} \otimes r_j \mathrm{e}^{\mathrm{j}\varphi_m} \triangleq r_j \mathrm{e}^{\mathrm{j}\varphi_{(n+m) \bmod \beta}} \ . \tag{5}$$

We call this differential encoding strategy *absolute amplitude and differential phase–shift keying (ADPSK)*.

2.3 Differential Encoding Matched to Multiple–Symbol Differential Detection

In multiple–symbol differential detection, demodulation is based on blocks of N consecutively received symbols, which are grouped into the vector $\boldsymbol{y}$. Successive vectors overlap at least in one symbol [11]. Applying this strategy, an obvious approach is to base channel coding and differential encoding on blocks on $N - 1$ (differential) symbols $a[\kappa]$, $\kappa = 1, 2, \ldots, N - 1$, combined into the vector

$$\boldsymbol{a} \triangleq [a[1], a[2], \ldots, a[N-1]] \ .$$

Together with the reference symbol s the block

$$\begin{aligned} \boldsymbol{x} &\triangleq [s, x[1], \ldots, x[N-1]] \\ &= [s,\ s \otimes a[1],\ (s \otimes a[1]) \otimes a[2],\ \ldots, \\ &\qquad (s \otimes a[1] \otimes \cdots) \otimes a[N-1]] \end{aligned}$$

of N transmit symbols at the transmitter corresponds to the respective block $\boldsymbol{y}$ at the receiver.

Instead of the usual accumulated differential encoding, decomposed ("parallel") encoding alternatively is possible,

which is the obvious counterpart to multiple–symbol differential detection at the receiver. Here, channel coding and mapping provide the vector

$$\boldsymbol{a}' \stackrel{\triangle}{=} [a'[1], a'[2], \ldots, a'[N-1]]$$

from which the differential encoder generates

$$\boldsymbol{x} = [s,\ s \otimes a'[1],\ s \otimes a'[2],\ \ldots,\ s \otimes a'[N-1]]$$

The value of $x[N-1]$ then gives the next state of the differential encoder. Both approaches are related by the simple accumulation $a'[\kappa] = a[1] \otimes \cdots \otimes a[\kappa]$. For $N = 2$ they coincide, for $N > 2$ they perform equivalent.

3 Capacity

Because of the differential encoding, the channel memory, and the overlapping of blocks $\boldsymbol{y}$, transmission is not memoryless. For the application of standard coding techniques and the ease of analysis (ideal) interleaving I (see Figure 1) at the transmitter and deinterleaving (I^{-1}) at the receiver are performed. Then, the channel between $\boldsymbol{a}$ and $\boldsymbol{y}$ is memoryless and can be entirely characterized by a single probability density function (pdf) $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ of $\boldsymbol{y}$ given $\boldsymbol{a}$.[1]

As we are intended to employ turbo codes [7] as channel coding scheme, which operates close to limits from information theory, capacity is considered as well suited measure for performance assessment. The normalized channel capacity[2] of the over–all transmission scheme (including differential encoding), measured in bits per symbol, is given by [12]

$$C(N) \stackrel{\triangle}{=} \frac{1}{N-1} \cdot \mathcal{E}_{\boldsymbol{Y},\boldsymbol{A}} \left\{ \log_2 \left(\frac{p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})}{p_{\boldsymbol{Y}}(\boldsymbol{y})} \right) \right\} , \quad (6)$$

where $p_{\boldsymbol{Y}}(\boldsymbol{y})$ is the average pdf of the channel output, and $\mathcal{E}_{\times}\{\cdot\}$ denotes expectation with respect to $\times$.

3.1 Conditional PDF

In order to calculate $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ we first consider the stationary, memoryless, non–dispersive, and multiplicative Rician fading channel between the N–dimensional vector symbols $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\boldsymbol{y}[k] = g[k] \cdot \mathrm{e}^{\mathrm{j}\phi[k]} \cdot \boldsymbol{x}[k] + \boldsymbol{n}[k] . \quad (7)$$

Here, $k \in \mathbb{Z}$ is the discrete–time index of the vector symbols, and $g[k]$ is the complex Gaussian distributed channel gain. Because of the assumption of slow fading all components of $\boldsymbol{x}[k]$ meet the same fading conditions. The average power of $g[k]$ is normalized to one and the Rician parameter, the ratio of directly to diffuse received power, is denoted by K. $\phi[k]$ is the carrier phase, which is uniformly distributed in $[0, 2\pi)$ and independent of $g[k]$. The vector $\boldsymbol{n}[k]$ denotes independent additive white Gaussian noise (AWGN) with variance $\sigma_n^2 = N_0/T$ per complex component.

In [11, Appendix] the pdf $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x})$ for such channels is calculated to be ($\mathrm{I}_0(\cdot)$ is the modified Bessel function of first kind and order zero and the superscript H denotes Hermitian transposition):

$$\begin{aligned} p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x}) \quad = \quad & \frac{1}{\pi^N (\sigma_n^2)^{N-1}} \cdot \frac{K+1}{|\boldsymbol{x}|^2 + (K+1)\sigma_n^2} \\ & \cdot \exp\left(-\frac{1}{\sigma_n^2} \left[|\boldsymbol{y}|^2 + \frac{K}{K+1}|\boldsymbol{x}|^2 \right.\right. \\ & \left.\left. - \frac{|\boldsymbol{y}\boldsymbol{x}^H|^2 + \frac{K}{K+1}|\boldsymbol{x}|^4}{|\boldsymbol{x}|^2 + (K+1)\sigma_n^2} \right] \right) \\ & \cdot \mathrm{I}_0\left(\frac{2\sqrt{K(K+1)}}{|\boldsymbol{x}|^2 + (K+1)\sigma_n^2} |\boldsymbol{y}\boldsymbol{x}^H| \right) . \quad (8) \end{aligned}$$

Since i) $\boldsymbol{x}$ is completely specified by s and $\boldsymbol{a}$, and ii) due to the uniformly distributed carrier phase the pdf only depends on $|s|$, $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x}) = p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}, |s|)$ holds. Finally, averaging $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}, |s|)$ over all possible amplitudes $|s| = r$ yields

$$p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}) = \mathcal{E}_r \{ p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}, r) \} . \quad (9)$$

3.2 Optimization of the Signal Constellation

Strictly speaking, in order to derive the channel capacity an optimization over all accessible parameters has to be performed. In our situation these are: i) the ring ratio of the APSK constellation, and ii) the probabilities of the rings. Due to the rotational invariance of APSK constellations, uniformly distributed phases are optimal.

If differential encoding of the amplitudes is used (encoding operator (4)), because of averaging, regardless the distribution of the differential symbols a, uniformly distributed amplitudes of the transmit symbols x result.

Contrary, in case of absolute amplitude encoding an optimization of the amplitude distribution $\Pr\{r_i\}$ is possible. But numerical evaluations turn out that the gains due to optimized non–equiprobable amplitudes are negligible when compared to equiprobable amplitudes. The reason is that uniformly distributed APSK constellations already exhibit some kind of shaping. Like in warping [13], a non–uniform density is achieved by optimizing the spacing of the signal points, rather than the probabilities. Hence, optimizing the ring ratio can (to some extent) replace an optimization of the amplitude distribution. The geometrical spacing of the rings, which is normally used, leads to a discrete second order hyperbolic distribution of the two–dimensional points a, i.e., $\Pr\{|a|\} \sim 1/|a|^2$.

In turn, the unavoidable uniform distribution of the amplitudes, which results in differential amplitude encoding, is no real disadvantage if the ring spacing is optimally adjusted. Thus, in all situations we fix the symbols $\boldsymbol{a}$ to be **uniformly**, independently and identically distributed, because this choice achieves the highest entropy and thus maximizes the throughput of the channel.

[1] We denote random variables corresponding to signals by the respective capital letter.

[2] Here, we always use the term "capacity" for the average mutual information given a **fixed** probability distribution of the constellation points.

4 Coded Modulation Schemes

In this section, the application of *multilevel coding (MLC)* [3] to differentially encoded transmission and non–coherent reception on fading channels is presented. For brevity, and without loss of generality, we restrict the discussion to binary component codes. Further, we match the channel coding scheme to multiple–symbol differential detection, i.e., coding is done with respect to vectors $\boldsymbol{a}$ of differential symbols.

As these vectors comprise $N-1$ symbols, each drawn from an M–ary set, $\ell = (N-1)\cdot\log_2(M)$ binary symbols are required to address $\boldsymbol{a}$. Thus, from vectors $\boldsymbol{b} = [b^0, b^1, \ldots, b^{\ell-1}]$ of binary symbols b^i, $i = 0, \ldots, \ell-1$, the mapping $\mathcal{M}$ generates vectors $\boldsymbol{a}$ of differential symbols. This in turn requires joint labeling and set partitioning for amplitude and phase changes.

As an alternate coded modulation scheme *bit–interleaved coded modulation (BICM)* [4, 5] is compared with MLC.

4.1 Multilevel Coding/Multistage Decoding

In [14, 15, 16, 17] it is proved that the capacity of the modulation scheme can be achieved by multilevel codes and multistage decoding iff the individual rates of the component codes are properly chosen. Key point is the well–known chain rule for mutual information [12]. Applying this procedure to the present channel including differential encoding, we arrive at a scheme, optimal in the sense of information theory.

Since the mapping $\mathcal{M}$ is bijective and regarding the chain rule to average mutual information we can rewrite Eq. (6) as

$$\begin{aligned} C(N) &= I(\boldsymbol{Y};\boldsymbol{A}) = I(\boldsymbol{Y};B^0,B^1,\ldots,B^{\ell-1}) \\ &= I(\boldsymbol{Y};B^0) + I(\boldsymbol{Y};B^1|B^0) + \ldots \\ &\quad + I(\boldsymbol{Y};B^{\ell-1}|[B^0,B^1,\ldots,B^{\ell-2}]) \ . \end{aligned} \tag{10}$$

Thus, transmission of vectors with binary digits b^i, $i = 0, \ldots, \ell-1$, over the physical channel can be virtually separated into the parallel transmission of individual digits b^i over ℓ *equivalent channels*, provided that $b^0, \ldots, b^{i-1}$ are known (cf. [3, 15]). In MLC the digits b^i, $i = 0, \ldots, \ell-1$, result from independent encoding of the the data symbols, see Fig. 3 a).

From (10) and the derivation in [15] the capacity of equivalent channel i is given by

$$\begin{aligned} C^i &\triangleq I(\boldsymbol{Y};B^i|[B^0,B^1,\ldots,B^{i-1}]) \\ &= \frac{1}{2^i} \sum_{[b^0\ldots b^{i-1}]\in\{0,1\}^i} I(\boldsymbol{Y};B^i|[b^0,b^1,\ldots,b^{i-1}]) \ . \end{aligned} \tag{11}$$

Since the subsets at one partitioning level may not be congruent, the average mutual information is calculated by averaging over all 2^i possible combinations of $b^0 \ldots b^{i-1}$.

The right hand side of (10) suggests the rule for a low complex staged decoding procedure, that is well–known as *multistage decoding (MSD)*. The component codes $\mathcal{C}^i$ are successively decoded by the corresponding decoders D^i starting at level 0. At stage i, decoder D^i processes not only the received signal, but also (hard) decisions of previous decoding stages $j = 0, 1, \ldots, i-1$. For details on encoder and decoder, see [3, 15].

The capacity $C(N)$ can be approached by MLC and MSD, iff the individual rates R^i are chosen to be equal to the capacities of the equivalent channels: $R^i = C^i$, which is applied subsequently (cf., e.g., [15]). Noteworthy, this procedure is optimum for capacity approaching codes [15] and we do not regard "traditional" design parameters such as minimum squared Euclidean distance, minimum Hamming distance or product distance. For short to medium code lengths, other design rules, e.g. based on error exponent or on the error rate of the individual levels are possible. For details on code design rules and a tutorial review on MLC see [15].

Starting from (9), and taking into account that all differential symbols $\boldsymbol{a}$ are equally likely, the metric for maximum–likelihood decoding of code $\mathcal{C}^i$ at level i is derived from the pdf

$$\begin{aligned} p_{\boldsymbol{Y}}(\boldsymbol{y}|b^i = c, [\hat{b}^0,\ldots,\hat{b}^{i-1}]) &= \\ \frac{1}{|\mathcal{L}_c^i(\hat{b}^0,\ldots,\hat{b}^{i-1})|} &\sum_{\boldsymbol{a}\in\mathcal{L}_c^i(\hat{b}^0,\ldots,\hat{b}^{i-1})} p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}) \ . \end{aligned} \tag{12}$$

The set $\mathcal{L}_c^i(b^0,\ldots,b^{i-1})$ comprises all differential symbols $\boldsymbol{a}$, which represent the given binary symbols b^0 through b^{i-1} and $b^i = c$:

$$\begin{aligned} \mathcal{L}_c^i(b^0,\ldots,b^{i-1}) &= \\ \left\{ \boldsymbol{a} \;\middle|\; \begin{array}{c} \boldsymbol{a} = \mathcal{M}([b^0,\ldots,b^{i-1},c,\delta^{i+1},\ldots,\delta^{\ell-1}]), \\ \delta^j \in \{0,1\},\ j = i+1,\ldots,\ell-1 \end{array} \right\} \ . \end{aligned} \tag{13}$$

In coherent transmission over the AWGN channel, often the decoding complexity is reduced by regarding only the nearest neighbours with respect to Euclidean distance of the received point. Using this sub–optimal metric usually results only in marginal losses. Here, in general, a similar approach can not be found. It seems to be impossible to obtain the dominant term in (12) by solely inspecting the vectors $\boldsymbol{a}$.

4.2 Bit–Interleaved Coded Modulation

In [5, 18, 15] modulation schemes using Gray labeling were investigated. Due to Gray labeling the (average) number of signal points at minimum Euclidean distance representing the inverse binary symbol is not greater than 1. Moreover, neither this number nor the minimum intra–subset Euclidean distance change when decisions of lower levels are not taken into account for decoding at higher levels. Hence, without significant loss, independent parallel processing of the levels is possible. Finally, only one binary code is applied and ℓ encoded bits are grouped to select the current symbol, see Fig. 3 b) and [5, 15].

Assuming ideal bit interleaving, the address symbols are independent of each other and, hence, this scheme can

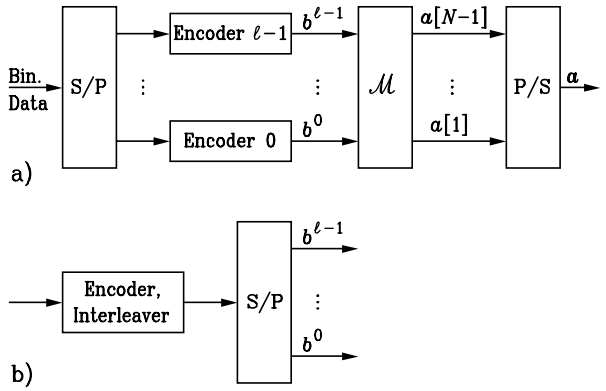

Figure 3: Channel encoding and Mapping: a) MLC, b) Changes for BICM.

be interpreted as binary transmission over a memoryless, but time–variant channel. This scheme is named *bit–interleaved coded modulation (BICM)* [5].

For several applications it has been shown [5] that BICM suffers only a marginal capacity loss compared to MLC. Here, only the reduced capacity

$$\begin{aligned} C^i_{\mathrm{BICM}} &= I(\boldsymbol{Y}; B^i) \\ &\leq I(\boldsymbol{Y}; B^i | [B^0, B^1, \ldots, B^{i-1}]) = C^i \ , \quad i = 0, 1, \ldots, \ell - 1 \end{aligned} \tag{14}$$

is usable.

On the one hand, in practice, the advantage of BICM is that only one binary code is required compared to ℓ, in general different, codes in MLC. On the other hand, BICM strongly relies on Gray labeling, which is now required for the differential symbols with an appropriate distance measure.

In BICM the metrics for maximum–likelihood decoding are derived from

$$p_{\boldsymbol{Y}}(\boldsymbol{y}|b^i = c) = \frac{1}{|\mathcal{L}^i_{c,\mathrm{BICM}}|} \sum_{\boldsymbol{a} \in \mathcal{L}^i_{c,\mathrm{BICM}}} p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}) \ , \tag{15}$$

with

$$\mathcal{L}^i_{c,\mathrm{BICM}} = \left\{ \boldsymbol{a} \;\middle|\; \begin{array}{c} \boldsymbol{a} = \mathcal{M}([\delta^0, \ldots, \delta^{i-1}, c, \delta^{i+1}, \ldots, \delta^{\ell-1}]), \\ \delta^j \in \{0, 1\}, \ j \neq i \end{array} \right\} . \tag{16}$$

4.3 Labeling for the Differential Symbols

Now, we discuss possible strategies for labeling the vector $\boldsymbol{a} = [a[1], \cdots, a[N-1]]$. MLC can approach capacity for any labeling. For finite code length however, Ungerböck labeling (UL) has a small advantage [15]. Thus, we use Ungerböck labeling in combination with MLC. Optimally, in combination with multiple–symbol differential detection a multidimensional set partitioning should be performed. Because up to now, no partitioning for the present situation is known, we partition the two–dimensional constituent constellation and simply take the Cartesian product (separation of the transitions). In terms of capacity this is still an optimal approach; for finite code length we conjecture that no significant differences are noticeable.

Contrary, BICM strongly relies on Gray labeling (GL). A mapping of binary address vectors to symbols in the signal space is *Gray labeled*, if the most likely error events result in the wrong decision of only a single binary digit. A suitable criterion is the pairwise error probability $\mathrm{PEP}(\boldsymbol{a} \to \hat{\boldsymbol{a}})$. The (differential) signal point $\hat{\boldsymbol{a}}$ will be denoted as *nearest neighbor* of $\boldsymbol{a}$ if and only if $\mathrm{PEP}(\boldsymbol{a} \to \hat{\boldsymbol{a}}) = \max_{\tilde{\boldsymbol{a}} \in \mathcal{A}^{N-1} \setminus \{\boldsymbol{a}\}} \mathrm{PEP}(\boldsymbol{a} \to \tilde{\boldsymbol{a}})$.

In [19] we have proved that a D(A)PSK constellation can only be Gray labeled if $N = 2$ and that the maximum number of pairs of nearest neighbors is Gray labeled when the usual Gray labeling, based on the Euclidean distance, is employed. Then, the number of exceptions to Gray labeling for differential symbols increases exponentially with N. We call this labeling subsequently *quasi–Gray labeling* because it is the best possible solution. For APSK constellations we perform labeling separately on phase and amplitude on the basis of the Euclidean distance.

Finally, we have to distinguish between labeling the symbols in conventional, accumulated differential encoding and parallel encoding. It can be proved that for both strategies the capacities of the individual coding levels are identical. Thus, from a practical point of view this choice is arbitrary and in each case the same MLC scheme, i.e., the same rates of the component codes, results.

5 Results

In this section numerical evaluations of the capacity and simulation results for coded modulation schemes using differential encoding are presented. We restrict to the 2A8PSK signal set with ring ratio $r_1/r_0 = 2.0$, which was found to be advantageous for fading channels [20]. In all MLC examples Ungerböck labeling is employed, whereas BICM implies the use of (quasi–)Gray labeling.

5.1 Numerical Results

First, we present capacity curves for differentially encoded transmission versus the (average) signal–to–noise ratio $\bar{E}_s/N_0$ ($\bar{E}_s$: average energy per symbol, N_0: one–sided noise power spectral density). Thereby, we mainly concentrate on transmission over the AWGN channel ($K \to \infty$) and the Rayleigh fading channel ($K = 0$), which are the most interesting special cases of the Rician fading model.

In Figure 4 the effect of increasing the block size N on the normalized capacity $C(N)$ is visualized for DAPSK transmission. As expected (e.g., [21]) enlarging N increases $C(N)$, because the statistical dependencies within the scalar received sequence are treated more completely. If N approaches infinity the channel is static and thus its state can be estimated (e.g., using pilot symbols). Hence, $C(N)$ converges to the capacity C_{CSI} for the case of perfect channel state information (CSI) and coherent reception. But the convergency is rather slow.

To compare MLC and BICM the capacity $C_{\mathrm{BICM}} \triangleq \sum_{i=0}^{\ell-1} C^i_{\mathrm{BICM}}$ is also displayed in Figure 4 (dashed dotted lines). Apparently, for $N = 2$ the capacity loss of BICM is

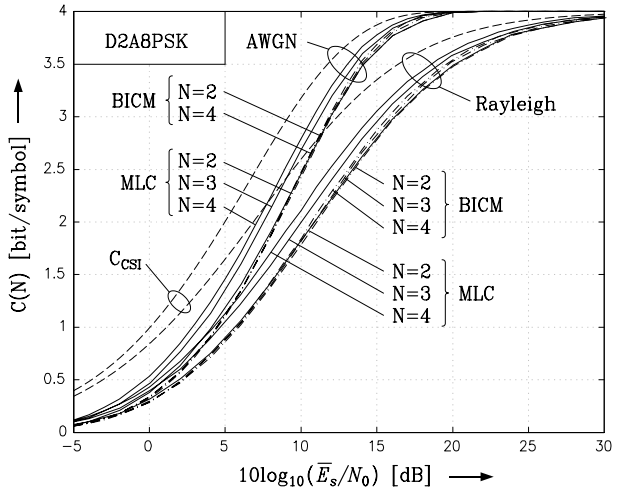

Figure 4: Normalized capacity for D2A8PSK, with ring ratio $r_1/r_0 = 2.0$. Solid lines: MLC, $N = 2, 3, 4$ (from right to left). Dash–dotted lines: BICM, $N = 2, 3, 4$ (from right to left). Dashed Line: Capacity C_{CSI}. AWGN and Rayleigh fading channel.

only marginal compared to MLC. But here, by increasing N almost no additional gain is provided. This is because for $N > 2$ no real Gray labeling exists. In particular, the number of pairs of nearest neighbors that are not Gray labeled increases exponentially [22].

Next, Figure 5 compares the two proposed differential encoding strategies DAPSK and ADPSK. Clearly, for the AWGN channel where the channel gain is constant, modulation of the actual transmit amplitude is superior to differential encoding over the amplitude. But, in the case of Rayleigh fading and capacities above 2. . .3 bit/symbol DAPSK outperforms ADPSK. Because a classification of the amplitudes becomes more reliable as N increases, the advantage of DAPSK over ADPSK becomes smaller for $N = 3$ compared to $N = 2$. Noteworthy, in both cases the channel amplitude is unknown at the receiver.

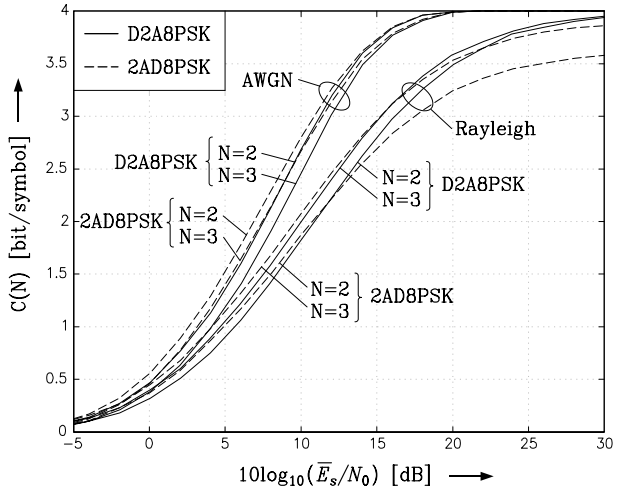

Figure 5: Normalized capacity for D2A8PSK (differential amplitude encoding, solid lines) and 2AD8PSK (absolute amplitude encoding, dashed lines), with ring ratio $r_1/r_0 = 2.0$; $N = 2, 3$. AWGN and Rayleigh fading channel.

This statement is confirmed by Figure 6, where the normalized capacity is plotted over the Rician fading factor K in the case $N = 2$. For low signal–to–noise ratios (low capacities) and due to the fixed ring ratio ADPSK is always preferable. For high $\bar{E}_s/N_0$ DAPSK clearly outperforms ADPSK for small K, and is almost identical with ADPSK as the channel tends to AWGN. In summary, it can be stated that differential amplitude encoding is preferable in the medium–to–high SNR region on channels with small K, whereas absolute amplitude encoding should be used as the channel goes to AWGN or the SNR is very low.

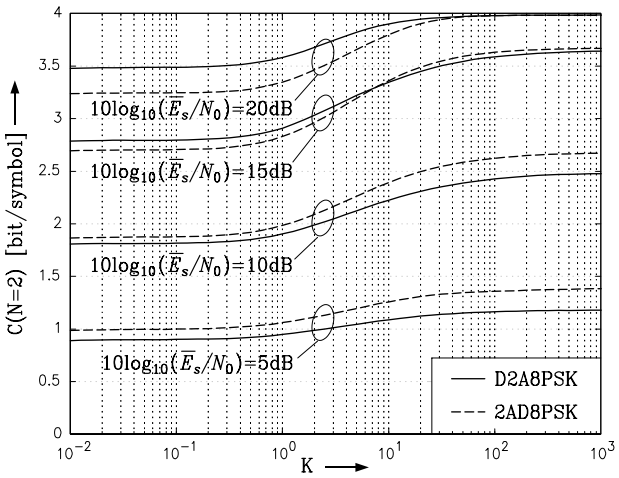

Figure 6: Normalized capacity for D2A8PSK (differential amplitude encoding, solid lines) and 2AD8PSK (absolute amplitude encoding, dashed lines) over Rician fading factor K. Ring ratio $r_1/r_0 = 2.0$; $N = 2$.

Finally, the MLC rate design is illustrated in Figure 7. For it, the overall capacity $C(N = 2)$ of D2A8PSK and the capacities of the equivalent channels are sketched over $\bar{E}_s/N_0$. Following the capacity rule [15], an example for the rate design of the component codes is given (dashed line). The target rate 2.5 bit/symbol divides optimally into the individual rates $R^i = C^i$, with $C^0/C^1/C^2/C^3 = 0.43/0.72/0.88/0.47$. Here, bit b^3 is assigned to the amplitude transition, whereas $b^0b^1b^2$ correspond to Ungerböck–type phase mapping.

5.2 Simulation Results

In order to further assess MLC for differentially encoded transmission numerical simulations of 2A8PSK over the Rayleigh fading channel were performed.

First, D2A8PSK with rate 2.5 bit/symbol over the Rayleigh fading channel is regarded. This rate is chosen from capacity arguments (see e.g. [23, Fig. 4]). The results in Figure 8 refer to a fixed channel delay equal to 6000 transmit symbols. As component codes Turbo codes [7] (two parallel concatenated systematic recursive convolutional codes with 16 states each and identical subgenerator $(1, \frac{27}{31})_8$ in octal notation) and the symmetrical decoder concept according to [24] is used. Rate is adjusted by symmetric puncturing of parity symbols and the interleavers of the

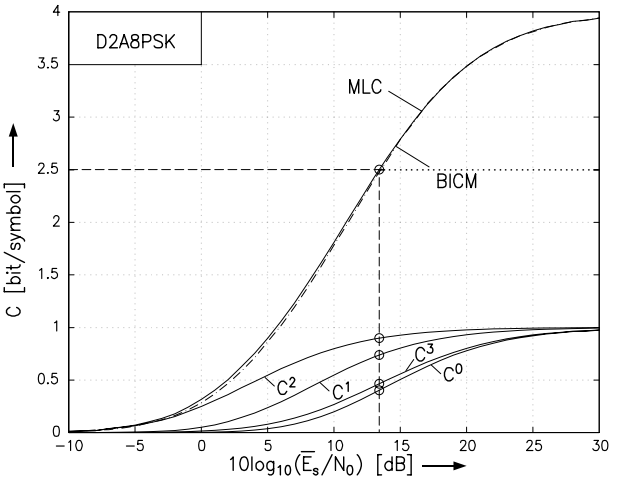

Figure 7: Capacity and capacities C^0, C^1, C^2, C^3 of the equivalent channels for D2A8PSK; $N = 2$. Rayleigh fading channel. Dashed lines: Rate design for $C = 2.5$ bit/symbol.

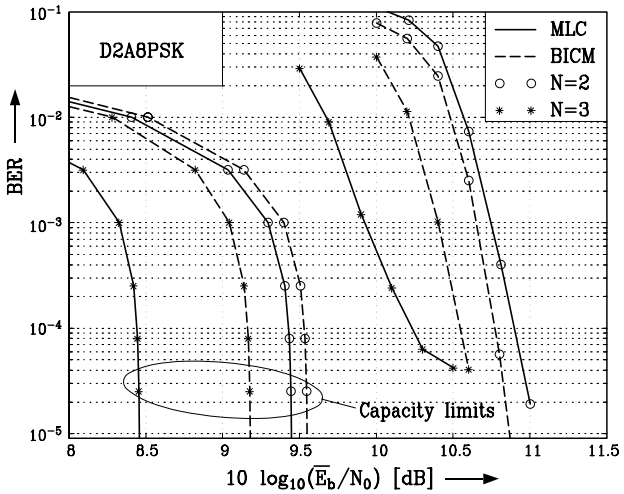

Figure 8: BER as a function of $\bar{E}_\mathrm{b}/N_0$ (in dB) for D2A8PSK with MLC (solid lines) and BICM (dashed lines), rate 2.5 bit/symbol. Rayleigh fading channel. Circles: $N = 2$, Stars: $N = 3$. Left hand side: respective rate–distortion capacity limits.

turbo codes are randomly generated. The decoders perform 6 iterations.

In the case of $N = 2$ the code length of the binary (component) codes in MLC are 6000. The above rate design according to channel capacity is applied. In order to arrive at the same over–all delay, for $N = 3$ the lengths of the binary codes are 3000. Here, the rates of the component codes are [3]

$$0.36/0.44/0.70/0.75/0.87/0.89/0.45/0.54\,.$$

As reference, the capacity limits taking the finite error rate into account ("rate–distortion capacities") [25] are shown.

The gain of multiple–symbol differential detection is clearly visible ($N = 2 \rightarrow N = 3$). Almost the whole gain of about 1 dB expected from capacity limits can be utilized with MLC, even though the code length at the individual levels becomes shorter. In each case for BERs around 10^{-4} an SNR–gap between 1.4 and 1.8 dB to capacity limit results. As the interleaver of the turbo code is randomly generated, some flattening of the error curve for MLC $N = 3$ is visible. Improvements are possible by designing the interleaver appropriately.

Again, for comparison BICM is applied. Here, independent of N, a single rate–$5/8$ code with code length 24000 is used to obtain a delay of 6000 channel symbols. As can be seen from Figure 8, only for $N = 2$ BICM is superior to MLC. But, using $N = 3$, despite the much shorter codes, MLC clearly outperforms BICM, as predicted from capacity. This is again due to the fact, that here no real Gray labeling is possible. But Gray labeling is the key point in BICM.

Finally, Figure 9 compares differential and absolute amplitude encoding of 2A8PSK and $N = 3$. The rates of the component codes for ADPSK are

$$0.34/0.41/0.69/0.74/0.88/0.91/0.49/0.54\,;$$

for DAPSK the design is already given above. In each case the transmission rate is 2.5 bit/symbol. As expected from capacity arguments for $N = 3$ and rate 2.5 bit/symbol differential amplitude encoding performs slightly worse on the Rayleigh channel (cf. Figure 5).

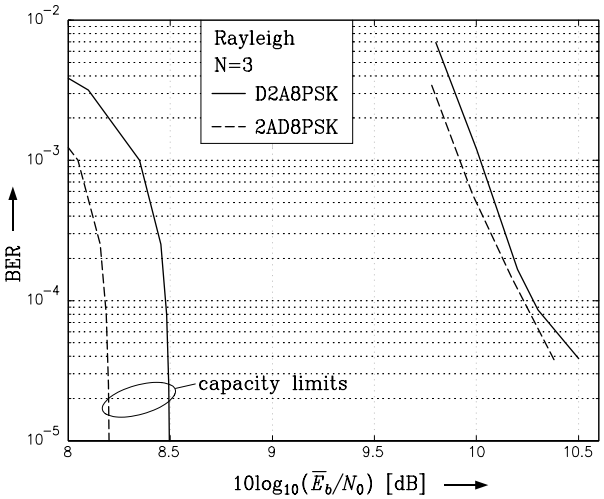

Figure 9: BER as a function of $\bar{E}_\mathrm{b}/N_0$ (in dB) for MLC encoded D2A8PSK (differential amplitude encoding, solid lines) and 2AD8PSK (absolute amplitude encoding, dashed lines), rate 2.5 bit/symbol. Rayleigh fading channel. $N = 3$. Left hand side: respective rate–distortion capacity limits.

In summary, the simulation results appear to be well in accordance with the results from information theory. As already accepted for the AWGN channel, the capacity rule again turns out to be the appropriate criterion for the choice of the component codes in MLC schemes. Both, MLC and BICM are effective methods to transfer the power

[3] The address bits of the differential symbols are first sorted according to the significance and then according to the position within the frame. Thus, the order of the bits, from the least significant to the most significant, is bit 0 of the label of the first transition, bit 0 of the label of the second transition, bit 1 of the label of the first transition, and so on. The last bit of the label corresponds to the amplitude transition.

efficiency of turbo codes into bandwidth–efficient transmission schemes using differential encoding and non–coherent reception over fading channels. MLC is indispensable to exploit the potential of multiple–symbol differential detection. For $N > 2$ BICM seems to be inappropriate because it strongly relies on Gray labeling, which does not exist for D(A)PSK and $N > 2$.

6 Conclusions

In this paper, multilevel coding for transmission over slowly fading channels with differential encoding and non–coherent reception is presented. For comparison, bit–interleaved coded modulation is regarded. At the receiver, no knowledge on the carrier phase and no information concerning the actual fading gain is assumed. Block interleaving and multiple–symbol differential detection of block length N is used to exploit the coherence time of the fading process.

The capacity rule, already well established for coherent transmission over the AWGN channel [15], is again the key point when designing the MLC scheme. For BICM, a suitable Gray labeling for the differential symbols is introduced. As no real Gray labeling exists for $N > 2$, shortcomings of the performance of BICM compared to MLC can be observed when multiple–symbol differential detection is applied. Exemplary, the theoretical results derived from information theory are confirmed by means of numerical simulations for the 2A8PSK constellation.

In summary, we believe that MLC with multiple–symbol differential detection is an attractive method to approach the capacity of fading channels with differential encoding and non–coherent detection very closely.

Acknowledgment

The authors greatly acknowledge Stefano Calabrò, visiting scientist at our lab in 1998, for valuable discussions, which stimulated some of the presented work.

References

[1] A. Svensson, "On Differentially Encoded Star 16QAM with Differential Detection and Diversity," *IEEE Tr. Vehic. Techn.*, 586–593, Aug. 1995.

[2] G. Ungerböck, "Channel Coding with Multilevel/ Phase Signals," *IEEE Tr. Inf. Theory*, 55–67, Jan. 1982.

[3] H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error Correcting Codes," *IEEE Tr. Inf. Theory*, 371–377, 1977.

[4] E. Zehavi, "8–PSK Trellis Codes for a Rayleigh Channel," *IEEE Tr. Comm.*, 873–884, May 1992.

[5] G. Caire, G. Taricco, and E. Biglieri, "Bit–Interleaved Coded Modulation," *IEEE Tr. Inf. Theory*, 927–946, May 1998.

[6] D. Divsalar and M.K. Simon, "Multiple–Symbol Differential Detection of MPSK," *IEEE Tr. Comm.*, 300–308, Mar. 1990.

[7] C. Berrou and A. Glavieux, "Near Optimum Limit Error Correcting Coding and Decoding: Turbo–Codes," *IEEE Tr. Comm.*, 1261–1271, Oct. 1996.

[8] J.G. Proakis, *Digital Communications*, McGraw–Hill, New York, third edition, 1995.

[9] W.T. Webb, L. Hanzo, and R. Steele, "Bandwidth Efficient QAM Schemes for Rayleigh Fading Channels," *IEE Proceedings–I*, 169–175, June 1991.

[10] T.S. Rappaport, *Wireless Communications — Principles & Practice*, Prentice Hall, Upper Saddle River, NJ, 1996.

[11] D. Divsalar and M.K. Simon, "Maximum-Likelihood Differential Detection of Uncoded and Trellis Coded Amplitude Phase Modulation over AWGN and Fading Channels — Metrics and Performance," *IEEE Tr. Comm.*, 76–89, Jan. 1994.

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

[13] W. Betts, A.R. Calderbank, and R. Laroia, "Performance of Nonuniform Constellations on the Gaussian Channel," *IEEE Tr. Inf. Theory*, 1633–1638, Sept. 1994.

[14] J. Huber, "Multilevel Codes: Distance Profiles and Channel Capacity," in *ITG–Fachtagung Codierung für Quelle und Kanal*, 305–319, München, Oct. 1994.

[15] U. Wachsmann, R.F.H. Fischer, and J.B. Huber, "Multilevel Codes: Theoretical Concepts and Practical Design Rules," *IEEE Tr. Inf. Theory*, 1361–1391, July 1999.

[16] Y. Kofman, E. Zehavi, and S. Shamai (Shitz), "Analysis of a Multilevel Coded Modulation System," in *1990 Bilkent International Conference on New Trends in Communications, Control, and Signal Processing*, 376–382, Ankara, Turkey, July 1990.

[17] G.D. Forney, Jr., M.D. Trott, S.-Y. Chung, "Sphere–Bound–Achieving Coset Codes and Multilevel Coset Codes," Accepted for publication at IEEE Tr. Inf. Theory, 1999.

[18] P. Schramm, "Multilevel Coding with Independent Decoding on Levels for Efficient Communication on Static and Interleaved Fading Channels," in *Proceedings of the PIMRC'97*, 1196–1200, Helsinki, Finland, Sept. 1997.

[19] L.H.-J. Lampe, S. Calabrò, R.F.H. Fischer, S. Müller–Weinfurtner, and J.B. Huber, "On the Difficulty of Bit–Interleaved Coded Modulation for Differentially Encoded Transmission over Fading Channels," in *Proc. 1999 IEEE Inf. Theory Workshop (ITW)*, Kruger National Park, South Africa, June 1999.

[20] L.H.-J. Lampe and R.F.H. Fischer, "Comparison and Optimization of Differentially Encoded Transmission on Fading Channels," in *In Proceedings 3rd International Symposium on Power–Line Communications and its Applications (ISPLC'99)*, 107–113, Lancaster, UK, Mar. 1999.

[21] R.J. McEliece and W.E. Stark, "Channels with Block Interference," *IEEE Tr. Inf. Theory*, 44–53, Jan. 1984.

[22] L.H.-J. Lampe, R.F.H. Fischer, S. Calabrò, and S. Müller-Weinfurtner, "Coded Modulation for DPSK on Fading Channels," in *Proceedings of Globecom'99*, Rio, Brasil, Nov. 1999.

[23] L.H-J. Lampe, R.F.H. Fischer, and S. Calabrò, "Channel Capacity of Fading Channels for differentially encoded transmission," *Electronics Letters*, 192–194, Jan. 1999.

[24] U. Wachsmann and J. Huber, "Power and bandwidth efficient digital communication using turbo codes in multilevel codes," *Europ. Trans. Telecommun. (ETT)*, 557–567, Sept.-Oct. 1995.

[25] S.A. Butman and R.J. McEliece, "The Ultimate Limits of Binary Coding for a Wideband Gaussian Channel," *JPL Deep Space Network Progress Report 42–22*, 78–80, Aug. 1974.