CODED MODULATION FOR DPSK ON FADING CHANNELS

L. H.-J. Lampe, R. F.H. Fischer, S. Calabrò*, and S. H. Müller–Weinfurtner

Laboratorium für Nachrichtentechnik, Universität Erlangen–Nürnberg
Cauerstraße 7/NT, D–91058 Erlangen, Germany

Abstract — Coded modulation for differentially encoded PSK with non–coherent reception on flat fading channels is considered. We derive theoretical limits on the performance and show how to design multilevel coding schemes and bit–interleaved coded modulation. Since Gray labeling is not possible for multiple symbol differential detection and PSK constellations, bit–interleaved coded modulation exhibits severe degradation compared to the multilevel coding approach. By means of simulations, where turbo codes are used as component codes, the theoretical considerations are confirmed and we show how close theoretical limits of differentially encoded PSK transmission over flat fading channels can be approached by applying properly designed coded modulation schemes.

1 Introduction

Coded modulation schemes for power and bandwidth efficient transmission over the AWGN channel were proposed by Ungerböck [1] (*trellis–coded modulation*) and by Imai/Hirakawa [2] (*multilevel coding (MLC)*). For coherent transmission over Rayleigh fading channels, a pragmatic approach to coded modulation, called *bit–interleaved coded modulation (BICM)*, was introduced by Zehavi [3].

In many situations, where neither channel state information is available, nor coherent reception is desired or possible, it is convenient to use differential encoding at the transmitter, cf. e.g. [4]. If the underlying channel is slowly time–varying *multiple symbol differential detection*, i.e., detection is based on blocks of N consecutive symbols, is proved (e.g. [4]) to provide further gains.

In this paper, the application of MLC and BICM for differential encoding of PSK on flat fading channels is developed and compared. We show analytically and by simulations that BICM, in contrast to MLC, is not well suited for multiple symbol differential detection. Aim of this contribution is to show that differentially encoded transmission over fading channels close to theoretical limits is possible via optimal designed coded modulation schemes.

*now with Siemens AG, München, Germany.

2 System Model and Capacity

The transmission scheme applied throughout the paper is sketched in Fig. 1. Core is the actual, discrete–

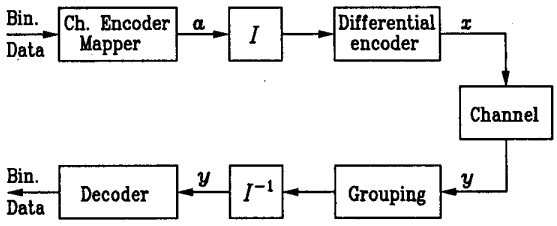

Figure 1: System model.

time channel model, given in the equivalent low–pass domain, i.e., all quantities are complex [5]. The channel is assumed to be a stationary, slowly time–varying, frequency non–selective (non–dispersive) Rician fading channel with finite memory. As usual, channel state and carrier phase offset are expected to be constant over a block of N consecutive symbols.

In order not to require knowledge on the carrier phase and actual channel state we employ differential encoding at the transmitter and differential detection at the receiver. Due to the coherence time of the channel of (at least) N symbols, the encoder operates on blocks of N consecutive channel input symbols x.

As usual for differentially encoded PSK (DPSK) we restrict the differential symbols a to be taken from the same signal set as the transmit signal, i.e., M uniformly spaced phases $\varphi_m = \frac{2\pi}{\beta}m$, $m = 0, \ldots, \beta - 1$. Then, a simple formulation of the operation of the differential encoder results.[1] Given a reference symbol (state of the differential encoder) $s = e^{j\varphi_n}$ and $N-1$ differential symbols (phase increments) $a[\nu] = e^{j\varphi_{m[\nu]}}$, $\nu = 1, \ldots, N-1$, the encoder outputs $x[\nu] = e^{j(\varphi_n + \varphi_{m[\nu]})}$. The value of $x[N-1]$ then gives the next state of the differential encoder. This definition of the differential encoder, which performs "parallel" encoding, is the obvious counterpart to multiple symbol differential detec-

[1] For the moment, interleaving and deinterleaving are ignored.

0-7803-5796-5/99/$10.00 © 1999 IEEE

tion at the receiver. If the usual "accumulated" encoding (with phase increments $\varphi_{m[\nu]}$) is desired, instead of $a[\nu]$, $a'[\nu] = e^{j\sum_{\mu=1}^{\nu}\varphi_{m[\mu]}}$ has to be passed to this encoder. For $N = 2$ both strategies coincide, for $N > 2$ they are equivalent; but for analysis parallel encoding is advantageous.

At the receiver, N consecutive channel output symbols y are grouped and comprised into the vector $\boldsymbol{y}$. Thereby, the blocks are overlapping by one symbol [6]. After performing ideal interleaving (I) of vector symbols $\boldsymbol{a} := [a[1], \ldots, a[N-1]]$ at the transmitter and the inverse operation (I^{-1}) operating on vectors $\boldsymbol{y}$ at the receiver side the transmission between $\boldsymbol{a}$ and $\boldsymbol{y}$ is regarded to be memoryless, and hence, can completely be characterized by a single probabilty density function (pdf)[2] $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ of $\boldsymbol{y}$ for given $\boldsymbol{a}$. In order to obtain $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ we consider the transmission of the N–dimensional symbols $\bar{\boldsymbol{x}} := [s, x[1], \ldots, x[N-1]]$ through the memoryless channel, modeled to be a stationary, non–dispersive and multiplicative Rician fading channel. The corresponding pdf $p_{\boldsymbol{Y}}(\boldsymbol{y}|\bar{\boldsymbol{x}})$, and hence $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}, s)$, is derived in the Appendix of [6]. Averaging $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}, s)$ with respect to s yields $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$.

Having established the channel model the capacity of the memoryless channel can be calculated. The normalized capacity $C(N)$, measured in bit per symbol, is obtained by [7]

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

where $\mathcal{E}\{\cdot\}$ denotes expectation and $p_{\boldsymbol{Y}}(\boldsymbol{y})$ is the average pdf of the channel output. Regardless the choice of the distribution of $\boldsymbol{a}$ the differentially encoded symbols x are uniformly distributed over $\mathcal{X}$. If, furthermore, x is uniformly, independently, and identically distributed (u.i.i.d) the entropy of the channel input is maximized. Because this implies an u.i.i.d. input to the differential encoder we choose the vector symbols $\boldsymbol{a}$ to be u.i.i.d., and thus no optimization on the channel input distribution can be performed.

As already observed in [8] the normalized capacity increases with the block size N. If N approaches infinity $C(N)$ converges to the capacity C_{CSI} for the case of perfect channel state information (CSI) and coherent reception, which upper bounds $C(N)$.

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

3 Coded Modulation

In this section, the application of *multilevel coding (MLC)* [2] to fading channels and differential encoding is considered. We restrict the discussion to binary component codes. As $N-1$ symbols are comprised in the vector $\boldsymbol{a}$ and each component is taken 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, i.e., labeling is done with respect to phase changes. As an alternative, *bit–interleaved coded modulation (BICM)* [3, 9] is applied to the differentially encoded transmission scheme.

3.1 Multilevel Coding/Multistage Decoding

It is well–known, e.g. [10], that multilevel encoding in combination with *multistage decoding (MSD)* can achieve capacity. Applying this procedure to the present channel including differential encoding, we arrive at an optimal scheme.

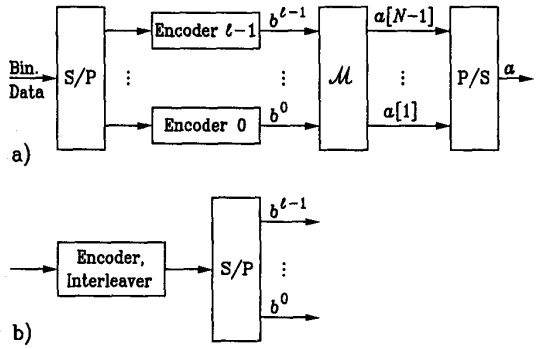

Figure 2: Channel encoding and Mapping: a) MLC, b) Changes for BICM.

In the MLC scheme transmission of vectors of binary digits b^i, $i = 0, \ldots, \ell-1$, over the physical channel is separated into the parallel transmission of individual digits b^i over ℓ *equivalent channels* with capacity C^i, provided that $b^0, \ldots, b^{i-1}$ are known (cf. [2, 10]). In MLC the digits b^i, $i = 0, \ldots, \ell-1$, result from independent encoding of the data symbols, see Fig. 2.

The component codes C^i with rate R^i are successively decoded by the corresponding decoders starting at level 0. At stage i, the decoder processes not only the received signal, but also decisions of previous decoding stages j. For details on encoder and decoder, see [2, 10].

According to [10] we chose the individual rates R^i to be equal to the capacities of the equivalent channels: $R^i = C^i$ (capacity rule). Noteworthy, this procedure is optimum for capacity approaching codes and we do not regard parameters such as minimum squared Euclidean distance, minimum Hamming distance or product distance.

3.2 Bit–Interleaved Coded Modulation

BICM is a suboptimum but simple scheme, applying only one binary code and (ideal) bit interleaving. Here, ℓ encoded and interleaved bits are grouped to select the current symbol, see Fig. 2.
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. When increasing the observation interval N the complexity remains constant in BICM whereas in MLC the number of levels ℓ increases, resulting in a considerably higher complexity. On the other hand, BICM strongly relies on Gray labeling.

3.3 Labeling for Differential Symbols

In [10] it is shown that MLC can approach capacity with any labeling. For finite code length however, Ungerböck labeling has a small advantage. 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.
For BICM, Gray labeling has to be applied. A mapping of binary address vectors to symbols in the signal space is *Gray labeled*, if the most likely error event results in the wrong decision of only a single binary digit. In contrast to usual coherent transmission, Euclidean distance is not a significant parameter in the case of differential encoding. By defining a signal point $\tilde{\boldsymbol{a}}$ as *nearest neighbor* of $\boldsymbol{a}$ iff the pairwise error probability $\mathrm{PEP}(\boldsymbol{a} \to \tilde{\boldsymbol{a}})$ is maximum, using the results of [6], it is easily shown that a DPSK constellation can be Gray labeled iff $N = 2$ [11]. The maximum number of pairs of nearest neighbors is Gray labeled when the usual Gray labeling, based on the Euclidean distance of the signal points $\boldsymbol{a}$, $\tilde{\boldsymbol{a}}$ is employed.

If $N > 2$, using Gray labeling with respect to Euclidean distance, the labels of

$$\begin{cases} \frac{1}{2}M^{N-1}, & \text{if } M = 2 \\ M^{N-1}, & \text{if } M > 2 \end{cases} \tag{2}$$

pairs of nearest neighbor points differ by more than one binary symbol. This labeling is subsequently called *quasi–Gray labeling* because it is the best possible solution.

3.4 Numerical Results

In this section numerical results for coded modulation using differential encoding and 8–ary differential PSK (8DPSK) are presented. Thereby, we restrict to capacity curves for the Rayleigh fading channel. In all examples for MLC with MSD Ungerböck labeling (UL) [1] is employed, whereas BICM implies the use of (quasi–) Gray labeling (GL).
For $N = 2$, in Fig. 3 the overall capacity and the capacities of the equivalent channels are sketched over $\bar{E}_s/N_0$ ($\bar{E}_s$: average energy per symbol, N_0: one–sided noise power spectral density). For the solid lines, MLC is assumed. Following the capacity rule, an example for the rate design of the component codes is given (dashed line). The target rate 1.5 bit/ch. use divides optimally into the individual rates $R^i = C^i$, with $C^0/C^1/C^2 = 0.17/0.53/0.80$. To compare MLC and BICM the capacity of BICM ($N = 2$) is also displayed in Fig. 3 (dash–dotted line). Apparently, the loss in capacity is very small compared to MLC.

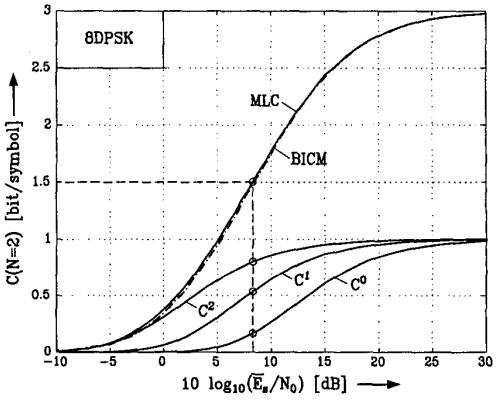

Figure 3: Capacity $C(N = 2)$ and capacities C^0, C^1, C^2 of the equivalent channels for 8DPSK. Solid lines: MLC with UL. Dash–dotted lines: BICM with GL. Rayleigh fading channel. Dashed lines: Rate design for $C = 1.5$ bit/symbol.

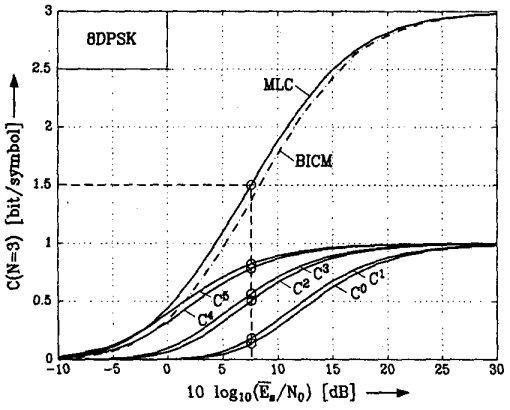

Figure 4: Capacity $C(N=3)$ and capacities $C^0,\ldots,C^5$ of the equivalent channels for 8DPSK. Solid lines: MLC with UL. Dash-dotted lines: BICM with GL. Rayleigh fading channel. Dashed lines: Rate design for $C = 1.5$ bit/symbol.

The capacity curves for $N = 3$ are shown in Fig. 4. The address bits of the differential symbols are first sorted according to the significance and then according to the position of the component within $\boldsymbol{a}$. 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. Aiming again at a target rate of 1.5 bit/ch. use, the capacities of the six equivalent channels of MLC are $C^0/\cdots/C^5 = 0.15/0.18/0.51/0.58/0.76/0.82$. With $N = 3$, a significant loss in capacity for BICM can be recognized.

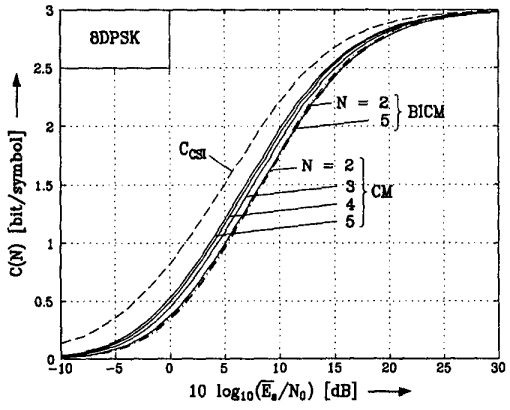

Figure 5: Normalized capacity for 8DPSK. Solid lines: MLC with UL, $N = 2,3,4,5$ (from right to left). Dash-dotted lines: BICM with GL, $N = 2,5$ (from right to left). Rayleigh fading channel. Dashed Line: Capacity C_{CSI}.

The influence of the block size N on the normalized capacity $C(N)$ is illustrated in Fig. 5. Additionally, as reference, the capacity C_{CSI} for coherent reception is shown. As expected (cf. Section 2), by enlarging N the capacity $C(N)$ increases. But the convergency of $C(N)$ to C_{CSI} as N goes to infinity is rather slow. Similar results for AWGN channel are shown in [12].
Remarkably, the capacity of BICM and $N = 5$ is almost the same as for $N = 2$. This is caused by the fact, that for $N > 2$ no real Gray labeling exists; in particular the exponential increase of the number of pairs of nearest neighbors that are not Gray labeled, cf. (2). Thus, for a given $N > 2$, we expect differentially encoded transmission using BICM with quasi-Gray labeling to be inferior to MLC.

4 Simulation Results

In order to assess the performance of the proposed coded modulation schemes, numerical simulations of differentially encoded 8PSK with target rate 1.5 bit/symbol over the Rayleigh fading channel are performed. Thereby, MLC in combination with Ungerböck's (natural) labeling and MSD is compared to BICM using (quasi-)Gray labeling. In each case, turbo codes [13] (parallel concatenated convolutional codes with 16 states) are employed as component codes. The systematical decoder concept according to [14] is used. Rate is adjusted by symmetric puncturing of parity symbols, cf. [14]. The interleavers of the turbo codes are randomly generated, i.e., no optimization has been performed. This leads to some flattening of the error curve at moderate to low error rates (error floor). The decoders are allowed to perform a maximum number of 6 iterations. For a fair comparison, the code length of the binary (component) codes are chosen such that the overall delay is equal for all strategies.
In Fig. 6, 8DPSK using MLC and BICM are compared for a fixed channel delay of 6000 transmitted symbols. For $N = 2$ the corresponding lengths of the binary codes are 6000 for MLC and 18000 for BICM, respectively. The rates of the component turbo-codes are designed according to capacity of the individual levels (see above and Fig. 3). In BICM, a single code with rate 1/2 is present.
With $N = 3$ the lengths of the six binary codes are 3000 for MLC and 18000 for BICM, respectively. The rates of the six component codes for MLC are given above (Fig. 4); in BICM again a single code with rate 1/2 is used.
As reference, the capacity limits taking the finite error

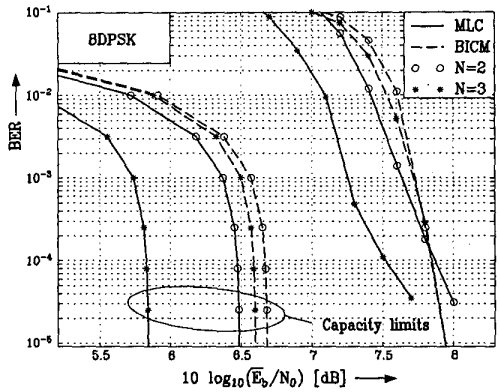

Figure 6: BER as a function of $\bar{E}_b/N_0$ (in dB) for 8DPSK with MLC (solid lines) and BICM (dashed lines). Rayleigh fading channel. Circles: $N = 2$, Stars: $N = 3$. Left hand side: respective rate-distortion capacity limits.

rate into account ("rate-distortion capacities") [15] for MLC and BICM are shown. In the case of $N = 2$ and for low $\bar{E}_b/N_0$ (bit error rates above BER $= 2 \cdot 10^{-4}$; $\bar{E}_b$: average energy per information bit) MLC outperforms BICM. Only for large $\bar{E}_b/N_0$ (very low error rates) BICM turns out to be superior, because it takes advantage of the greater blocklength. Here, the error floor of the turbo codes is much lower than for the shorter codes used in MLC.

For $N = 3$, despite the shorter codes, a clear superiority of MLC over BICM is visible as predicted from the capacity gap in Fig. 5. This is due to the fact, that for $N > 2$ and DPSK no exact Gray labeling is possible. But Gray mapping is the key point in BICM.

Noteworthy, an increase of the admissible data delay is extremely rewarding in the case of MLC, where only rather short component codes can be applied. Moreover, the error floor is lowered considerably. For BICM an increase is almost ineffective, because of the already very large code length.

5 Conclusions

The application of MLC and BICM to differentially PSK over fading channels is studied. Both the capacity of the channel and the maximum mutual information which can be utilized by means of BICM are calculated. The severe degradation of the performance of BICM compared to MLC and MSD due to the non-existence of real Gray labeling in the case of multiple symbol differential detection is shown. Examples of the rate design for MLC are given. Applying the properly chosen individual rates for the MLC component codes numerical simulations for 8DPSK confirm the theoretical results.

We believe that the use of MLC and multiple symbol differential detection is an attractive method to approach the capacity of fading channels with differential encoding and bandwidth efficient constellations very closely.

References

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

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

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

[4] D. Divsalar and M.K. Simon. Multiple–Symbol Differential Detection of MPSK. *IEEE Tr. Commun.*, 300–308, March 1990.

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

[6] 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. Commun.*, 76–89, Jan. 1994.

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

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

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

[10] 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.

[11] 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. In *Proc. ITW'99*, page 111, Kruger National Park, SA, June 1999.

[12] M. Peleg and S. Shamai (Shitz). On the Capacity of the Blockwise Incoherent MPSK Channel. *IEEE Tr. Commun.*, 603–609, May 1998.

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

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

[15] 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.