Noncoherent Coded Modulation for Fading Channels

Lutz H.-J. Lampe[1] and Robert F.H. Fischer
Lehrstuhl für Nachrichtentechnik II, Universität Erlangen–Nürnberg
Cauerstraße 7/NT, D–91058 Erlangen, Germany
Phone: +49-9131-85-28718, Fax: +49-9131-85-28919
Email: {LLampe,fischer}@LNT.de

Abstract — **Coded modulation for noncoherent transmission over slowly time–variant flat fading channels without channel state information is considered. We focus on differentially encoded M–ary PSK with multiple–symbol differential detection (MSDD). Interleaving of blocks of symbols is used as a compromise between the conflicting requirements of exploiting the channel coherence time and providing diversity for decoding.**

We study multilevel coding (MLC) which is perfectly matched to the overall vector channel. As already well–known for coherent transmission, properly designed MLC is proved to be asymptotically optimum in the noncoherent case, too. A favorable strategy for labeling of signal points is given. As a promising alternative to MLC, bit–interleaved coded modulation (BICM) is addressed. But since no Gray labeling is possible, BICM can only marginally benefit from MSDD. Conversely, in case of strict decoding delay constraints, MLC suffers from component codes with very short code length. To overcome this drawback, we propose hybrid coded modulation schemes, which are able to combine the advantages of MLC and BICM, respectively. For performance assessment, we evaluate both the achievable channel capacity and the random coding exponent associated with noncoherent coded modulation.

Moreover, we show that in multilevel coding and multistage decoding (MSD) the complexity of MSDD can be reduced significantly. Remarkably, the performance gain of MSDD can be exploited almost completely with practically no increase in computational complexity compared to conventional differential detection.

Index Terms: Multilevel coding, Bit–interleaved coded modulation, flat fading channel, Differential encoding, Multiple–symbol differential detection, Random coding exponent

[1]Corresponding author

1 Introduction

The information theoretic analysis and the design of digital communications systems for transmission over fading channels are widely addressed fields of research for many years, cf. e.g. [1, 2, 3] and references therein. In many scenarios, e.g., regarding wireless communication with mobile terminals or even powerline communication where network characteristics fluctuate, the coherence time of the channel is of the order of only a few modulation intervals. Consequently, noncoherent transmission over fading channels requiring no (reliable) channel state information (CSI) at the receiver has attracted considerable attention, e.g. [4, 5, 6, 7, 8, 9, 10]. In order to bridge the gap to coherent transmission with perfect CSI, power–efficient noncoherent receivers utilize the dependences among consecutive received symbols, i.e., the memory of the fading process is (partially) taken into account.

In this paper, multiple–symbol differential detection (MSDD) [11, 12, 4, 5] is considered for noncoherent reception. That is, decision variables are calculated from blocks of $N \geq 2$ consecutive received symbols. At the transmitter classical M–ary differentially encoded phase shift keying (MDPSK) modulation [13] is performed to avoid phase ambiguity. Furthermore, we apply block interleaving matched to MSDD. This decorrelation of successive fading states is necessary if standard coding techniques should be applied [14]. Throughout the paper, we regard transmission over *flat* Rayleigh fading channels. Specifically, we focus on *block Rayleigh fading channels* [15, 16, 17], where the channel is virtually time–invariant during N modulation intervals. Though the given analysis is easily extended to more general flat fading channels, for clarity we restrict the discussion to this specific model. The obtained results provide general statements on coded noncoherent transmission.

We present multilevel coding (MLC) and multistage decoding (MSD) [18, 19] matched to noncoherent transmission with MSDD. It is shown that MLC/MSD asymptotically approach channel capacity iff the rates of the component codes are properly designed. Because of its efficiency for coherent transmission over the flat Rayleigh fading channel, suboptimal, but simple bit–interleaved coded modulation (BICM) [20, 21] is also considered as a promising alternative. A further class of *hybrid* coding schemes is introduced to combine the distinct advantages of MLC and BICM. For MLC/MSD a suboptimal

MSDD is proposed, which offers an almost arbitrary exchange between complexity and performance. By this approach, remarkable gains over conventional differential detection ($N = 2$) can be achieved with practically no increase in computational complexity. The various coding techniques are assessed by evaluating both the achievable capacity assuming infinite code length and the random coding exponent, taking finite decoding delay into account. Moreover, favorite labeling strategies for the signal points are discussed.

Besides MSDD, noncoherent sequence detection (NSD) techniques have proved to be very efficient [22, 23]. There, when channel coding is applied, decoding is done in an augmented code trellis. Since interleaving cannot be employed, such NSD schemes are not suited to coded noncoherent transmission over fading channels. NSD with soft outputs as described in [24] is an interesting alternative, which allows for interleaving and iterative decoding, cf. also [25, 26, 8, 27]. However, these schemes based on soft–output algorithms require high computational effort and give no insight on how to design powerful coded modulation schemes based on a information theoretic analysis. Note that MLC in combination with MSDD has also been discussed in [28]. Contrary to our approach, code design is based on "traditional" distance properties and error probability discussions.

This paper is organized as follows. In Section 2, we describe the channel model and give its (constrained) channel capacity and its random coding exponent. The design of the various coded modulation schemes and the derivation of the respective information theoretic parameters are presented in Section 3. In Section 4, the performance of these schemes is assessed and discussed. Finally, conclusions are drawn in Section 5.

2 System Model and Performance Bounds

The block diagram of the discrete–time system model is depicted in Fig. 1. All signals and systems are given in the equivalent low–pass domain, i.e., all quantities are complex–valued, cf. e.g. [29].

A sequence of information bits enters the encoder. After encoding, which is explained in detail in Section 3, $\ell \triangleq (N-1) \cdot \log_2(M)$ coded bits are mapped to data–carrying *vector symbols* $\boldsymbol{a}[k] \triangleq [a_1[k], \ldots, a_{N-1}[k]]^T$ (k: vector symbol discrete–time index, $.^T$ denotes

transposition). The $N-1$ components $a_i[\cdot]$ stem from an M–ary PSK constellation $\mathcal{A} \triangleq \{\mathrm{e}^{\mathrm{j}2\pi m/M} | m = 0, 1, \ldots, M-1\}$. Since $a_i[\cdot]$ establish phase differences, we denote $a_i[\cdot]$ as *differential signal points*. Then, vector symbols $\boldsymbol{a}[\cdot]$ are interleaved. In order to enable noncoherent demodulation, the scalar symbols $a_i[\cdot]$ are differentially encoded. That is, the channel input symbols $x_i[\cdot]$ are obtained from the recursion

$$x_i[k] = \begin{cases} x_{N-1}[k-1]a_1[k] \ , & i = 1 \\ x_{i-1}[k]a_i[k] \ , & 2 \leq i \leq N-1 \end{cases} \ . \tag{1}$$

To each vector symbol $\boldsymbol{a}[k]$ a block $\boldsymbol{x}[k] \triangleq [x_{N-1}[k-1], x_1[k], \ldots, x_{N-1}[k]]^T$ of N consecutive symbols $x_i[\cdot]$ is corresponding. Thereby, the first (reference) symbol $x_{N-1}[k-1]$ of a block is the last symbol of the previously transmitted block. This overlapping by one symbol is advantageous regarding spectral and power efficiency [30, 31].

We assume sufficiently slow fading and transmitter and receiver filters with square–root Nyquist characteristics. Hence, the discrete–time Rayleigh fading channel is frequency–nonselective (flat) and the input–output–relation reads

$$y_i[k] = g_i[k] \cdot x_i[k] + n_i[k] \ , \quad 1 \leq i \leq N-1 \ , \tag{2}$$

where the fading gain $g_i[\cdot]$ is a correlated zero–mean complex Gaussian process. The noise process $n_i[\cdot]$ is white with variance σ_n^2 per complex sample and also zero–mean complex Gaussian ($\mathcal{E}\{n_\nu[\kappa] n_\mu^*[l]\} = \sigma_n^2$ for $\nu = \mu, \kappa = l$ and 0 else, $\mathcal{E}\{\cdot\}$ denotes expectation). Both processes are mutually independent. Defining $\boldsymbol{g}[k] \triangleq [g_{N-1}[k-1], g_1[k], \ldots, g_{N-1}[k]]^T$ and assuming appropriate normalization, according to the block fading model [16] the autocorrelation function of the fading process is required to satisfy

$$\mathcal{E}\{\boldsymbol{g}[k]\boldsymbol{g}^H[k]\} = \mathbf{1} \ , \tag{3}$$

where $\mathbf{1}$ is the $N \times N$ all–one matrix.

At the receiver, N consecutive channel output symbols are grouped and comprised into the vector $\boldsymbol{y}[k] \triangleq [y_{N-1}[k-1], y_1[k], \ldots, y_{N-1}[k]]^T$. Thereby, the blocks are overlapping by one symbol corresponding to $\boldsymbol{x}[\cdot]$, cf. [5, Fig. 3]. The vector symbols $\boldsymbol{y}[\cdot]$ are deinterleaved[2]

[2]Noteworthy, interleaving is done on vectors of length $N-1$, whereas deinterleaving is done for vectors of length N.

and the decoder provides an estimate of the binary data sequence to the sink. Since the decision variables for decoding are based on joint observations of N received symbols, the inherent memory of the transmission is partially taken into account.

2.1 Vector Channel Model

Regarding differential encoding and MSDD as integrate part of the channel, after interleaving a memoryless *vector channel* between $\boldsymbol{a}$ and $\boldsymbol{y}$ is obtained, cf. Fig. 1. Without loss of clarity, we can omit the time index k and adopt the notations $\boldsymbol{a} = [a_1, \ldots, a_{N-1}]^T$ and $\boldsymbol{y} = [y_1, \ldots, y_N]^T$. Though for finite interleaver depth in case of some delay constraint the "memoryless channel" assumption might be violated, the resulting decoding strategy is of practical interest due to its simplicity and compatibility with standard coding techniques. Moreover, if space or frequency diversity is exploited, e.g., applying multicarrier transmission [32] with differential encoding in time direction and coding in frequency direction, the independent block fading model retains its validity. Thus, the vector channel can completely be characterized by a single complex N–dimensional probability density function (pdf)[3] $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ of $\boldsymbol{y}$ for given $\boldsymbol{a}$. Since the multiplicative channel gain is zero–mean complex Gaussian distributed, conditioned on $\boldsymbol{a}$, $\boldsymbol{y}$ is also complex Gaussian distributed. Defining the autocorrelation matrix $\boldsymbol{R}_{\boldsymbol{y}|\boldsymbol{a}}$ of $\boldsymbol{y}$ for given data–carrying symbol vector $\boldsymbol{a}$ as

$$\boldsymbol{R}_{\boldsymbol{y}|\boldsymbol{a}} \stackrel{\triangle}{=} \mathcal{E}\{\boldsymbol{y}\boldsymbol{y}^H|\boldsymbol{a}\} \tag{4}$$

the pdf reads [4, 5] ($[\cdot]^H$ and $\det(\cdot)$ denote Hermitian transpose and determinant of a matrix, respectively)

$$p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}) = \frac{1}{\pi^N \det(\boldsymbol{R}_{\boldsymbol{y}|\boldsymbol{a}})} \cdot \mathrm{e}^{-\boldsymbol{y}^H \boldsymbol{R}_{\boldsymbol{y}|\boldsymbol{a}}^{-1} \boldsymbol{y}} \; . \tag{5}$$

For MDPSK and block Rayleigh fading considered here, we obtain

$$\boldsymbol{R}_{\boldsymbol{y}|\boldsymbol{a}}^{-1} = \frac{1}{\sigma_n^2}\left(\boldsymbol{I} - \frac{[1, \boldsymbol{a}'^T]^T [1, \boldsymbol{a}'^H]}{N + \sigma_n^2}\right) \tag{6}$$

$$\det(\boldsymbol{R}_{\boldsymbol{y}|\boldsymbol{a}}) = (\sigma_n^2)^{(N-1)} \cdot (N + \sigma_n^2) \; , \tag{7}$$

where we introduced $\boldsymbol{a}' \stackrel{\triangle}{=} [a'_1, \ldots, a'_{N-1}]^T$ with

$$a'_i = \prod_{k=1}^{i} a_k \tag{8}$$

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

and $\boldsymbol{I}$ is the $N \times N$ identity matrix. Finally, using (6), (7), the vector channel pdf (5) follows to

$$p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}) = \frac{1}{\pi^N (\sigma_n^2)^{N-1}} \cdot \frac{1}{N+\sigma_n^2} \cdot \exp\left(-\frac{1}{\sigma_n^2}\left[\boldsymbol{y}^H\boldsymbol{y} - \frac{|[1, \boldsymbol{a}'^H]\boldsymbol{y}|^2}{N+\sigma_n^2}\right]\right) . \tag{9}$$

Clearly, there is a bijective relation between $\boldsymbol{a}$ and $\boldsymbol{a}'$. If binary information is directly mapped onto $\boldsymbol{a}'$, instead of usual "accumulated" differential encoding, "parallel" encoding is performed [33].

2.2 Capacity and Random Coding Exponent

As information theoretic parameters to quantify potential data rates supported by the vector channel the associated channel capacity C and the random coding exponent $E_r(R)$ [34] are considered.

The average mutual information $I(\cdot;\cdot)$, measured in bits per vector symbol, is given by [34]

$$I(\boldsymbol{Y};\boldsymbol{A}) = \mathcal{E}_{\boldsymbol{y},\boldsymbol{a}}\left\{\log_2\left(\frac{p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})}{p_{\boldsymbol{Y}}(\boldsymbol{y})}\right)\right\} , \tag{10}$$

where $p_{\boldsymbol{Y}}(\boldsymbol{y}) = \mathcal{E}_{\boldsymbol{a}}\{p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})\}$ is the average pdf of the channel output. In order to calculate the channel capacity an optimization over all accessible parameters has to be performed. In our situation, only the distribution of $\boldsymbol{a}$ is accessible. Regardless of the choice of the distribution of $\boldsymbol{a}$ the differentially encoded symbols x are uniformly distributed over $\mathcal{A}$. Since uniformly, independently, and identically distributed (u.i.i.d.) x are shown to achieve capacity [35], cf. also [36], we choose the vector symbols $\boldsymbol{a}$ to be u.i.i.d. Then, the constrained capacity normalized with respect to the number $N-1$ of differential symbols reads

$$C(N) = \frac{I(\boldsymbol{Y};\boldsymbol{A})}{N-1} . \tag{11}$$

Whereas C constitutes the absolute performance limit assuming infinite code length and zero error probability, the random coding exponent $E_r(R)$ relates decoding delay and reliability of decoded data. The random coding exponent associated with the memoryless vector channel is given by [34]

$$E_r(R,N) = \max_{0\le\rho\le 1}\{E_0(\rho,N) - \rho R\} , \tag{12}$$

with Gallager's function

$$E_0(\rho, N) = -\log_2 \left(\mathcal{E}_{\boldsymbol{y}} \left\{ \mathcal{E}_{\boldsymbol{a}} \left\{ \left(\frac{p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})}{p_{\boldsymbol{Y}}(\boldsymbol{y})} \right)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \right\} \right) , \tag{13}$$

where R is the rate in bit per vector symbol. Using the random coding exponent and given the number n of vector symbols $\boldsymbol{a}$ per code word, the word error rate p_w is bounded by

$$p_w \leq 2^{-n \cdot E_r(R,N)} . \tag{14}$$

Clearly, performance is expected to improve with increasing observation window N, as the dependencies among received symbols are more completely comprised. In terms of capacity, when N goes to infinity the performance of coherent transmission with perfect CSI is approached [16, 35]. In terms of cutoff rate, which is obtained as the special case $E_0(\rho = 1, N)$, a reverse effect has been observed in [16, 7]. This is due to the loss of diversity in case of large N and limited delay.

3 Design of Coded Modulation for MSDD

In this section, coded modulation schemes matched to multiple–symbol differential detection are presented. In particular, we study well–known *multilevel coding (MLC)* [18, 19] and *bit–interleaved coded modulation (BICM)* [20, 21]. Furthermore, hybrid strategies combining power–efficiency of MLC and low decoding delay of BICM are introduced. To enable a performance assessment of the different coding schemes, the corresponding capacity and the random coding exponent are given.

For brevity, and without loss of generality, we restrict the discussion to binary component codes. As coding is done with respect to vectors $\boldsymbol{a}$ comprising $N-1$ components 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. This in turn requires joint labeling for blocks of phase changes.

3.1 Multilevel Coding/Multistage Decoding

Huber et al. [37, 19], Kofman et al. [38, 39], and Forney [40] proved that the capacity of the modulation scheme can be achieved by multilevel encoding and *multistage decoding (MSD)* iff the individual rates of the component codes are properly chosen. Applying this procedure to the present channel including differential encoding, we arrive at a scheme, optimal in the sense of information theory.

3.1.1 Capacity

In MLC the digits b^i, $i = 0, \ldots, \ell - 1$, result from independent encoding of the data symbols, see Fig. 2 a). At the receiver, the component codes 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 MLC encoder and MSD decoder, see [18, 19]). 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. [18, 19]). The capacity of equivalent channel i is given by

$$\begin{aligned} C^i(N) &\stackrel{\triangle}{=} I(\boldsymbol{Y}; B^i | [B^0, B^1, \ldots, B^{i-1}]) \\ &= \mathcal{E}_{b^0,\ldots,b^{i-1}} \{ I(\boldsymbol{Y}; B^i | [b^0, b^1, \ldots, b^{i-1}]) \} \\ &= \mathcal{E}_{\boldsymbol{y},b^i,b^0,\ldots,b^{i-1}} \left\{ \log_2 \left(\frac{p_{\boldsymbol{Y}}(\boldsymbol{y} | b^i, [b^0, \ldots, b^{i-1}])}{p_{\boldsymbol{Y}}(\boldsymbol{y})} \right) \right\} \end{aligned} \tag{15}$$

where averaging over all 2^i possible combinations of $b^0 \ldots b^{i-1}$ is required, because the subsets at one partitioning level may not be congruent[4].

Since the mapping $\mathcal{M}$ is bijective and due to the chain rule to average mutual information we can rewrite Eq. (10) as

$$\begin{aligned} C(N) &= \frac{1}{N-1} I(\boldsymbol{Y}; \boldsymbol{A}) = \frac{1}{N-1} I(\boldsymbol{Y}; B^0, B^1, \ldots, B^{\ell-1}) \\ &= \frac{1}{N-1} \left(I(\boldsymbol{Y}; B^0) + I(\boldsymbol{Y}; B^1 | B^0) + \ldots + I(\boldsymbol{Y}; B^{\ell-1} | [B^0, B^1, \ldots, B^{\ell-2}]) \right) \\ &= \frac{1}{N-1} \sum_{i=0}^{\ell-1} C^i(N) \, . \end{aligned} \tag{16}$$

[4]Subsets are congruent if they can be transformed into each other by applying only translation and rotation.

We conclude that for noncoherent transmission with MSDD the channel capacity 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: $R^i = C^i$. Thus, for MLC operating close to capacity, "traditional" design parameters such as minimum squared Euclidean distance, minimum Hamming distance or product distance are not appropriate.

3.1.2 Random Coding Exponent

For the random coding exponent corresponding to MLC with MSD the equivalent channels have to be considered. Gallager's function $E_0^i(\rho, N)$ of the i^{th} equivalent channel is given by averaging $E_0^i(\rho, N|b^0, \ldots, b^{i-1})$ with respect to all combinations $b^0, \ldots, b^{i-1}$:

$$E_0^i(\rho, N) = \mathcal{E}_{b^0,\ldots,b^{i-1}}\{E_0^i(\rho, N|b^0, \ldots, b^{i-1})\} \tag{17}$$

with

$$E_0^i(\rho, N|b^0, \ldots, b^{i-1}) = -\log_2\left(\mathcal{E}_{\boldsymbol{y}}\left\{\mathcal{E}_{b^i}\left\{\left(\frac{p_{\boldsymbol{Y}}(\boldsymbol{y}|b^i, [b^0, \ldots, b^{i-1}])}{p_{\boldsymbol{Y}}(\boldsymbol{y})}\right)^{\frac{1}{1+\rho}}\right\}^{1+\rho}\right\}\right) . \tag{18}$$

Averaging again accounts for irregular, i.e., non–congruent, partitions.

Applying the random coding exponent $E_r^i(R^i, N)$ of equivalent channel i

$$E_r^i(R^i, N) = \max_{0\le\rho\le 1}\{E_0^i(\rho, N) - \rho R^i\} \tag{19}$$

the word error rate of level i is bounded by

$$p_w^i \le 2^{n\cdot E_r^i(R^i,N)} . \tag{20}$$

Noteworthy, for usually desired (low) error rates and well–designed MLC, the effect of error propagation over levels can be neglected. Moreover, word error rates of levels are almost equal, and the overall word error rate p_w is approximated by

$$p_w \approx p_w^i , \quad 0 \le i \le \ell - 1 . \tag{21}$$

3.1.3 Decoding and Metrics Calculation

At level i metrics for decoding are based on the two pdf's $p_{\boldsymbol{Y}}(\boldsymbol{y}|b^i, [b^0, \ldots, b^{i-1}])$ of $\boldsymbol{y}$ given the decisions $b^0, \ldots, b^{i-1}$ from lower levels and a trial symbol $b^i = c \in \{0, 1\}$. Optimally,

these pdf's are obtained by averaging the vector channel pdf $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ over all $\boldsymbol{a}$ which represent the given binary symbols $b^0, \ldots, b^i$ ($|\cdot|$ denotes cardinality of a set):

$$p_{\boldsymbol{Y}}(\boldsymbol{y}|b^i = c, [b^0, \ldots, b^{i-1}]) = \frac{1}{|\mathcal{L}_c^i(b^0, \ldots, b^{i-1})|} \sum_{\boldsymbol{a} \in \mathcal{L}_c^i(b^0, \ldots, b^{i-1})} p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}) \tag{22}$$

with the set of vector symbols to be averaged over ($\mathcal{M}(\cdot)$ is the mapping)

$$\mathcal{L}_c^i(b^0, \ldots, b^{i-1}) = \left\{ \boldsymbol{a} \;\middle|\; \begin{array}{c} \boldsymbol{a} = \mathcal{M}([b^0, \ldots, b^{i-1}, c, \delta^{i+1}, \ldots, \delta^{\ell-1}]), \\ \delta^j \in \{0, 1\},\ j = i+1, \ldots, \ell-1 \end{array} \right\} . \tag{23}$$

3.2 Bit–Interleaved Coded Modulation

Stimulated by Zehavi's work [20], a pragmatic approach to coded 8PSK, modulation schemes using Gray labeling were investigated in [21] and [19]. Here, only one binary code is applied and ℓ encoded bits are grouped to select the current symbol, see Fig. 2b) For several applications it has been shown [21] that this bit–interleaved coded modulation suffers only a marginal capacity loss compared to MLC. 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 [41].

3.2.1 Capacity

In BICM, the dependencies of binary symbols transmitted over different equivalent channels are not taken into account. Thus, only the reduced capacities

$$C_{\mathrm{BICM}}^i(N) = I(\boldsymbol{Y}; B^i) \leq I(\boldsymbol{Y}; B^i | [B^0, B^1, \ldots, B^{i-1}]) = C^i \ , \quad i = 0, 1, \ldots, \ell-1 \tag{24}$$

summing up to

$$C_{\mathrm{BICM}}(N) = \frac{1}{N-1} \sum_{i=0}^{\ell-1} C_{\mathrm{BICM}}^i(N) \tag{25}$$

are usable.

3.2.2 Random Coding Exponent

To calculate the random coding exponent associated with BICM we consider Gallager's function of equivalent channel i:

$$E_{0,\mathrm{BICM}}^{i}(\rho, N) = -\log_2 \left(\mathcal{E}_{\boldsymbol{y}} \left\{ \mathcal{E}_{b^i} \left\{ \left(\frac{p_{\boldsymbol{Y}}(\boldsymbol{y}|b^i, i)}{p_{\boldsymbol{Y}}(\boldsymbol{y})} \right)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \right\} \right) . \tag{26}$$

Since BICM uses all equivalent channels equiprobably with perfect knowledge of i, Gallager's function of BICM is given by

$$E_{0,\mathrm{BICM}}(\rho, N) = \frac{1}{\ell} \sum_{i=0}^{\ell-1} E_{0,\mathrm{BICM}}^{i}(\rho, N) . \tag{27}$$

From that, the coding exponent of BICM reads

$$E_{r,\mathrm{BICM}}(R, N) = \max_{0 \le \rho \le 1} \{ E_{0,\mathrm{BICM}}(\rho, N) - \rho R/\ell \} . \tag{28}$$

3.2.3 Decoding and Metrics Calculation

Analogous to MLC, in BICM the metrics for maximum–likelihood decoding are derived from

$$p_{\boldsymbol{Y}}(\boldsymbol{y}|b^i = c) = \frac{1}{|\mathcal{L}_{\mathrm{BICM},c}^{i}|} \sum_{\boldsymbol{a} \in \mathcal{L}_{\mathrm{BICM},c}^{i}} p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}) , \tag{29}$$

with

$$\mathcal{L}_{\mathrm{BICM},c}^{i} = \left\{ \boldsymbol{a} \;\middle|\; \begin{array}{c} \boldsymbol{a} = \mathcal{M}([\delta^0, \ldots, \delta^{i-1}, c, \delta^{i+1}, \ldots, \delta^{\ell-1}]), \\ \delta^j \in \{0, 1\},\ j \neq i \end{array} \right\} . \tag{30}$$

3.3 Labeling for the Differential Symbols

The different encoding schemes require different mapping strategies. Now, we discuss possible approaches. Note, for $N > 2$ a "multi–dimensional mapping" is present.

3.3.1 Labeling for MLC

The considerations in Section 3.1 show that MLC can approach capacity with any labeling (cf. [19]). For finite code length however, the choice of the labeling influences performance of MLC/MSD, i.e., random coding exponent (19) and hence word error rate (21) depend

on the labeling strategy. Optimally, in combination with MSDD a multi–dimensional partitioning should be performed. Because up to now, no partitioning designed for the present situation is known, we (i) label the two–dimensional constituent constellation and simply take the Cartesian product (separation of the transitions), and (ii) apply multi–dimensional labelings designed for coherent PSK [42, 43].

If the differential symbols a_ν, $1 \leq \nu \leq N-1$, are labeled separately, the question of how to assign the ℓ levels to address bits arises. Denoting the $\log_2(M)$ address bits of a_ν, by $b^{0,\nu}, \ldots, b^{\log_2(M)-1,\nu}$ (from the least to the most significant bit), two approaches seem to be natural. Either (a) levels in MLC are first sorted according to their significance and then according to their position ν within $\boldsymbol{a}$, or (b) levels are first sorted according to ν and then according to significance. These two schemes are illustrated in Fig. 3 for 8PSK and $N = 4$. A performance evaluation of MLC shows very similar results for either strategy; with respect to capacity they are of course equivalent. For this reason, in Section 4 we concentrate on labeling according to Fig. 3 b).

3.3.2 Labeling for BICM

BICM strongly relies on Gray labeling. A mapping of binary address vectors to symbols in signal space is called to be *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 $\mathrm{PEP}(\boldsymbol{a} \to \hat{\boldsymbol{a}}) = \max_{\tilde{\boldsymbol{a}} \in \mathcal{A}^{N-1} \setminus \{\boldsymbol{a}\}} \mathrm{PEP}(\boldsymbol{a} \to \tilde{\boldsymbol{a}})$. From (9) it is apparent that the maximum of $\mathrm{PEP}(\boldsymbol{a} \to \tilde{\boldsymbol{a}})$ corresponds to the minimum of

$$d^2_{\mathrm{NC}}(\tilde{\boldsymbol{a}}, \boldsymbol{a}) \stackrel{\triangle}{=} N - \left| 1 + \tilde{\boldsymbol{a}}'^H \boldsymbol{a}' \right| \tag{31}$$

with $\boldsymbol{a}'$ uniquely determined from $\boldsymbol{a}$, cf. (8). The measure $d^2_{\mathrm{NC}}(\tilde{\boldsymbol{a}}, \boldsymbol{a})$ can be regarded as a counterpart to Euclidean distance for coherent transmission and thus is often named *noncoherent distance*, e.g. [31].

In [41] we have proved that a DPSK constellation can only be Gray labelled if $N = 2$. For $N > 2$, the maximum number of pairs of nearest neighbors is Gray labelled when the usual Gray labeling, based on the Euclidean distance of the differential signal points is employed.

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

$$\begin{cases} \frac{1}{2}M^{N-1}\,, & \text{if} \quad M = 2 \\ M^{N-1}\,, & \text{if} \quad M > 2 \end{cases} \tag{32}$$

pairs of nearest neighbor points differ by more than one binary symbol. Although the number of exceptions to Gray labeling with respect to noncoherent distance increases exponentially with N, this labeling is subsequently called *quasi–Gray labeling* because it is the best possible solution.

3.4 Hybrid Coded Modulation Schemes

MLC and BICM as discussed above are the two extreme cases with respect to encoding. An obvious modification is to combine levels in the MLC approach and to perform BICM within these "hyper levels".

If, for a fair comparison, the code length of the binary (component) codes are chosen such that the over–all delay is equal for all strategies, BICM uses codes with ℓ times the length of the component codes in MLC. Especially for multiple–symbol differential detection ($N > 2$) very short code lengths result for MLC. This effect can be mitigated by a hybrid strategy; if a number of levels are combined the code length can be increased by the same factor.

The rational behind designing hybrid coding schemes is, starting from MLC with MSD, to find levels with small dependencies among each other. Then, these levels can be merged without significant loss in performance. This will be demonstrated in Section 4, where level dependencies are judged from capacities C^i of equivalent channels. If separated labeling is applied (see Section 3.3.1), a special situation results when $N-1$ levels corresponding to address bits of equal significance (i.e., $b^{m,1}, \ldots, b^{m,N-1}$ for fixed m) are combined. In this case, multilevel encoding is done as for $N = 2$ and at the receiver MSDD with adaptively adjustable N can be performed. I.e., the transmitter is not required to have knowledge of the observation window size.

Again, the metrics for decoding are obtained by averaging the vector channel pdf over a set of vector symbols $\boldsymbol{a}$. The specific set is straightforwardly found as a mix of (23)

used for MLC and (30) used for BICM. For decoding, decision from lower levels are taken into account, but an averaging over bits within the same level and higher levels has to be performed. For calculation of capacity and random coding exponent the same modifications apply.

3.5 Simplified MSDD for MLC and Separated Labeling

For MLC and separated labeling depicted in Fig. 3 b), we can significantly reduce MSDD complexity in terms of number of pdf's to be computed, whereas the resulting performance loss remains small.

According to (22), for optimal metrics all M^{N-1} N–dimensional pdf's $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ have to be evaluated. That is, computational effort increases *exponentially in* N and polynomially in M. However, the computational load is not equally distributed over the levels. At lower levels, much more averaging, and therefore calculations of pdf's, is required than at higher levels, where decisions from lower levels are taken into account. The idea to simplify metric computation is to *adapt* the observation interval size within multistage decoding of levels. I.e., instead of fixing N, an adapted observation interval N'_ν, $1 \leq \nu \leq N-1$, — only depending on the number of the current transition — is applied for metrics corresponding to the address bits $b^{m,\nu}$, $0 \leq m \leq \log_2(M)-1$, of a_ν.

A particularly interesting situation occurs if N'_ν increases linearly with ν, i.e., $N'_\nu = \nu+1$, $1 \leq \nu \leq N-1$. This procedure is illustrated in Fig. 4 for 8DPSK and $N=4$. Now, overall $M/\log_2(M)$ channel pdf's per bit have to be calculated for MSDD, which is exactly the number of pdf's required for performing conventional differential detection with $N=2$. Consequently, this simplified MSDD for MLC with MSD does *not* lead to an increased complexity compared to conventional differential detection.

More generally, any other strategy of choosing N'_ν over ν is feasible. Through adaptation of the observation window simplified MSDD offers a sliding exchange between complexity and performance of noncoherent detection.

Finally we note that disregarding higher levels in MSD by applying $N'_\nu < N$ is in the same spirit as using only "nearest–neighbor" signal points for decoding of lower levels,

which is often done in case of coherent MLC.

4 Performance of Coded Noncoherent Transmission

In this section the proposed coded modulation schemes for differentially encoded transmission with MSDD over flat Rayleigh fading channels are assessed based on the associated capacity and random coding exponent. Subsequently, we exemplary present numerical results for 8DPSK.

4.1 Channel Capacity

The limit for reliable communication over flat Rayleigh fading channels without CSI using MSDD and optimal coded modulation is given by the channel capacity $C(N)$ (11). With MLC and MSD this capacity is achievable, however, under the constraint of BICM data rates are bounded by the decreased capacity $C_{\mathrm{BICM}}(N)$ (25). Whereas $C(N)$ is independent of the labeling of vector symbols $\boldsymbol{a}$, BICM implies the use of (quasi–) Gray labeling (GL) to maximize $C_{\mathrm{BICM}}(N)$.

Figure 5 compares $C(N)$ and $C_{\mathrm{BICM}}(N)$ as functions of the received signal–to–noise ratio (SNR) $\bar{E}_s/N_0$ ($\bar{E}_s$: average received energy per symbol, N_0: one–sided noise power spectral density) for different observation lengths N. Additionally, as reference, the capacity C_{CSI} in case of no differential encoding and coherent reception with perfect CSI is shown, which upper bounds $C(N)$ [16]. Clearly, utilizing the channel memory by enlarging the observation window the capacity $C(N)$ increases. But, convergence of $C(N)$ to C_{CSI} as N grows is rather slow, cf. e.g. [16].

Apparently, applying BICM with $N = 2$ the loss in capacity is very small compared to MLC. However, with $N = 5$ the capacity of BICM is almost the same as for $N = 2$. This is caused by the fact that for $N > 2$ no true Gray labeling exists; in particular the exponential increase of the number of nearest–neighbor pairs that are not Gray labeled, cf. (32). Since the potential performance gain offered by an enlarged demodulation interval cannot be exploited, we conclude that BICM is *not suited* to MSDD. This observation

coincides with the results in [44]. There, based on the cutoff rate of the AWGN channel with unknown phase, the improvement of differentially encoded transmission with MSDD is shown to be strictly limited in case of BICM.

A design example, how to exploit MLC/MSD gains in a practical system with a target rate of 2 bit/(channel use) and MSDD with $N = 3$, is depicted in Fig. 6. There, the capacities of the equivalent channels (15) are plotted for separately Ungerboeck labeling (UL) [45] of the two constituent 8PSK constellations. The six address bits of the data–carrying symbols are sorted according to the strategy of Fig. 3 b). Following the capacity rule [19], the target rate 2 bit/(channel use) divides optimally into the individual rates $R^i = C^i$ of the component codes. In this particular example, the capacities of the six equivalent channels of MLC are $C^0/\cdots/C^5 = 0.35/0.70/0.87/0.43/0.77/0.88$. Noteworthy, simulations with powerful Turbo codes [46] as component codes and rates according to capacity rule are in great accordance with the information theoretic results presented here, cf. e.g. [33, 47]. Conversely, if code design is based on traditionally used parameters such as minimum squared Euclidean distance, minimum Hamming distance or product distance, channel capacity can not be approached in principle.

4.2 Random Coding Exponent

Now, the coded modulation schemes are assessed under a decoding delay constraint. As measure of average performance the minimum required signal–to–noise ratio yielding a random coding exponent, which in turn leads to a word error rate p_w lower than a given threshold, will be discussed. In all presented examples, 8DPSK with target rate of 2 bit/(channel use) is considered.

4.2.1 Labeling for MLC/MSD

In contrast to the case of infinite decoding delay, for finite code length random coding exponent of (19) and word error rate (21) of MLC/MSD depend on the labeling strategy. We will compare labeling, (i) where scalar symbols a are independently and identically labeled using UL and GL of the two–dimensional constituent constellation, and (ii) where

joint labeling according to the designs from Pietrobon et al. (PL) [42] and Leonardo et al. (LL) [43] is performed. In case of separated labeling, the levels are assigned as in Fig. 3 b). Furthermore, random labeling (RL), i.e., the average over randomly labeled multi–dimensional mappings, is studied for reference.

For MSDD with $N = 3$ Fig. 7 shows $\bar{E}_b/N_0$ ($\bar{E}_b$: average received energy per information bit) required for fixed $p_w = p_w^i = 10^{-3}$, $0 \leq i \leq 5$, over the length n of the component codes with different labelings. As performance limit, the curve for overall maximum–likelihood decoding (MLD), which does not depend on the particular labeling, is also plotted (see Eq. (12)).

As can be seen, for large code length MLC with MSD approaches the performance of MLD regardless the chosen labeling. In case of relatively short codes, i.e., $10^{-2} \leq n \leq 10^{-3}$, there is a gap of up to 0.4 dB in power–efficiency between different labelings. More specifically, UL is clearly superior to GL and shows also better performance than the average over randomly generated labelings (RL). The curve corresponding to partitioning taken from [43] (LL) is almost indistinguishable from the curve of separated UL. PL compares unfavorably to UL. Consequently, partitioning of multi–dimensional signal points $\boldsymbol{a}$ based on UL of the constituent PSK constellation proves to be an advantageous as well as simple strategy.

4.2.2 MLC/MSD vs. BICM

According to the results given above, for a comparison of MLC/MSD with BICM in case of limited decoding delay we apply UL in combination with MLC and (quasi) GL for BICM. Again, for MLC the levels are assigned as depicted in Fig. 3 b). Even though in terms of capacity MLC/MSD is always superior to BICM, for relatively short decoding delays, we expect BICM to have an advantage over MLC/MSD also in case of MSDD with $N > 2$. This is because for fixed delay, the code length of BICM is ℓ times larger than the length of the MLC component codes.

In Fig. 8, the curves of $\bar{E}_b/N_0$ required for $p_w = 10^{-3}$ and MSDD with $N = 2, 3, 4$ are plotted over the decoding delay measured in scalar channel symbols. For long delays, the situation is very similar to Fig. 5 considering capacities. MLC with MSD benefits

from increased observation interval, but the power–efficiency of BICM does practically not improve with N. When the decoding delay is reduced, the performance gain of MLC/MSD over BICM decreases until the curves intersect. The larger N, the lower the value of the delay at the point of intersection. This means, to a certain extent, the loss of MLC due to shorter codes when increasing N (and hence the number of levels ℓ) is more than compensated by the gain due to MSDD. E.g. with $N = 4$ MLC/MSD outperforms simple BICM for delays larger than 2000 channel symbols. Consequently, when decoding delay is limited to values in this range, there is a need for *hybrid* coded modulation schemes, which favorably combine the advantages of MLC and BICM.

4.2.3 Hybrid Coded Modulation Schemes

As explained in Section 3.4, to design power–efficient hybrid coded modulation schemes we consider MLC and merge deliberately selected levels in order to increase the code length for given delay. We will discuss two promising approaches based on UL and GL of the constituent PSK constellation, respectively.

Since the investigation of partitioning for MLC showed separated UL to be a good choice, first a hybrid scheme (Hybrid I) based on MLC with UL is proposed. The capacity curves in Fig. 6 for this case and $N = 3$ indicate that decoding at level 3 (4,5) can only little benefit from the decisions already made at levels 0 to 2. A similar observation can be made regarding the random coding exponent. Consequently, levels (0,3), (1,4), (2,5) can be merged to each one level of a hybrid scheme. For general N, this strategy is continued, ending up with $\ell = \log_2(M)$ levels (combining the levels in horizontal direction in Fig. 3 a)). Noteworthy, the number of component encoders is independent of N, and thus, the receiver can adjust a desired observation window without informing and modifying the transmitter.

The resulting performance curves for MSDD with $N = 3, 4, 5$ are displayed in Fig. 9 (the curves of MLC and BICM are repeated from Fig. 8). Unfortunately, the Hybrid I scheme shows a similar behavior as BICM. For large delays a considerably gap in power–efficiency compared to MLC remains. For short delays BICM is still superior. If $N = 5$, in a relatively small range of delays from about 2000 to 3000 symbols Hybrid I can

outperform MLC and BICM. Clearly, this effect becomes more manifest and Hybrid I becomes more appealing if N further increases.

Generally, Gray labeling implies that although bits are jointly mapped onto one signal point they are almost independently transmitted. Adopting this property to our situation straightforwardly leads to a second hybrid scheme (Hybrid II) by merging all levels, which correspond to the same component symbol a_ν, $1 \leq \nu \leq N-1$, and perform GL of PSK symbols. That is, address bits $b^{m,\nu}$, $0 \leq m \leq \log_2(M)-1$, for fixed ν are assigned to one level (combining the levels in vertical direction in Fig. 3 b)). Hence, instead of $\ell = (N-1) \cdot \log_2(M)$, now, $\ell = N-1$ levels remain. This scheme can be interpreted as MLC with efficient BICM at different levels.

In Fig. 10 the curves for the required $\bar{E}_b/N_0$ over decoding delay for $p_w = 10^{-3}$ are depicted. We compare MLC and BICM (the curves are repeated from Fig. 8) with the hybrid scheme (Hybrid II) for MSDD with $N = 3, 4$. As can be seen, for very short delay ≤ 500 symbols BICM still performs best, and for long delay of more than $4 \cdot 10^4$ symbols MLC/MSD is of course superior. But in a wide range of medium decoding delay, the proposed hybrid coded modulation significantly reduces the gap to optimal MLD. Here, the advantages of the original coding schemes are united.

4.3 Simplified MSDD for MLC/MSD

Finally, we assess simplified MSDD for MLC/MSD introduced in Section 3.5 by numerical evaluation of the corresponding random error exponent. We focus on linearly increasing the observation window from $N' = 2$ for the lowest $\log_2(M)$ levels up to $N' = N$ for the highest $\log_2(M)$ levels as described in Section 3.5. In this case, the computational complexity (with respect to the number of pdf's to be evaluated) of MSDD with $N > 2$ and conventional differential detection with $N = 2$ are identical.

The required $\bar{E}_b/N_0$ for $p_w = 10^{-3}$ when using MSDD with optimal MSD and with reduced complex MSD, respectively, are compared in Fig. 11 for $N = 2, 3, 4$. As can be seen, the proposed metric simplification causes almost no loss in power–efficiency. Of course, for $N = 2$ both schemes are identical. When N increases, more terms are neglected during

metric calculation, leading to a somewhat stronger degradation. Therefore, a slightly more distinct difference between the curves of optimal and simplified decoding can be observed for $N = 4$ compared to $N = 3$. Remarkably, using this low complex MSDD approach, with respect to the number of pdf's to be computed the gain of more than 1 dB in required $\bar{E}_b/N_0$ by enlarging N from 2 to 4 is for free.

5 Conclusions

In this paper coded modulation schemes for noncoherent transmission over slow fading channels without CSI are proposed and compared. We focus on differentially encoded M–ary PSK with multiple–symbol differential detection (MSDD). Block channel interleaving is used to comply with the conflicting requirements of exploiting the channel coherence time in MSDD and providing diversity for decoding.

A straightforward approach of coded modulation matched to MSDD is to apply MLC to blocks of differential symbols. As well–known for coherent transmission, also for noncoherent reception MLC/MSD is asymptotically optimum in terms of capacity. We give the appropriate design of MLC, which is based on capacities of equivalent channels. As an promising alternative to MLC we consider BICM. However, for MSDD with observation size $N > 2$ no Gray labeling exists. This causes a severe performance degradation of BICM compared to MLC/MSD when transmitting close to capacity limits.

To further assess the coded modulation schemes, the associated random coding exponents are considered as performance measures. Regarding MLC with limited decoding delay, separate Ungerboeck labeling of PSK symbols turns out to have advantages and is subsequently applied. Furthermore, if relatively short delay constraints are imposed, MLC is shown to suffer from reduced length of its component codes as N increases. As a solution to this dilemma we design hybrid coding schemes, which are able to combine the advantages of both MLC and BICM. Doing so, performance is considerably improved for short to medium delays.

Finally, we discuss a simplified MSDD with adaptive observation interval for bit metric calculation in multistage decoding. This technique enables an almost arbitrary trade–off

between complexity and performance of MSDD. In particular, we can achieve nearly the whole gain in power–efficiency owing to MSDD with practically no increase in computational complexity compared to conventional differential detection. Thus, in summary MLC in combination with MSDD is an attractive method for power and bandwidth–efficient noncoherent transmission.

Acknowledgment

The authors gratefully acknowledge J.B. Huber and S. Calabrò for many valuable discussions relating to this work. Thanks to P. Amon and B. Helal for contributing numerical results.

References

[1] E. Biglieri and M. Luise, editors. *Coded Modulation and Bandwidth–Efficient Transmission.* Elsevier, Amsterdam, 1992.

[2] C.-E. Sundberg and N. Seshadri. Coded Modulation for Fading Channels: An Overview. *Europ. Trans. Telecommun. (ETT)*, 4:309–324, May/June 1993.

[3] E. Biglieri, J. Proakis, and S. Shamai (Shitz). Fading Channels: Information–Theoretic and Communications Aspects. *IEEE Trans. on Inf. Theory*, 44(6):2619–2692, October 1998.

[4] P. Ho and D. Fung. Error Performance of Multiple–Symbol Differential Detection of PSK Signals Transmitted over Correlated Rayleigh Fading Channels. *IEEE Trans. on Commun.*, 40:25–29, October 1992.

[5] 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 Trans. on Commun.*, 42(1):76–89, January 1994.

[6] D. Makrakis, P.T. Mathiopoulos, and D.P. Bouras. Optimal Decoding of Coded PSK and QAM Signals in Correlated Fast Fading Channels and AWGN: A Combined Envelope, Multiple Differential and Coherent Detection Approach. *IEEE Trans. on Commun.*, 42(1):63–74, January 1994.

[7] J. Ventura-Traveset, G. Caire E. Biglieri, and G. Taricco. Impact of Diversity Reception on Fading Channels with Coded Modulation — Part II: Differential Block Detection. *IEEE Trans. on Commun.*, 45(6):576–586, June 1997.

[8] P. Hoeher and J. Lodge. "Turbo DPSK": Iterative Differential PSK Demodulation and Channel Decoding. *IEEE Trans. on Commun.*, 47(6):837–843, June 1999.

[9] R. Schober, W.H. Gerstacker, and J.B. Huber. Decision–Feedback Differential Detection of MDPSK for Flat Rayleigh Fading Channels. *IEEE Trans. on Commun.*, 47(7):1025–1035, July 1999.

[10] D. Warrier and U. Madhow. Noncoherent communication in space and time. *submitted for publication. http://www.ece.ucsb.edu/Faculty/Madhow/publications.html.*

[11] D. Divsalar and M.K. Simon. Multiple–Symbol Differential Detection of MPSK. *IEEE Trans. on Commun.*, 38(3):300–308, March 1990.

[12] D. Divsalar, M.K. Simon, and M. Shahshahani. The Performance of Trellis-Coded MDPSK with Multiple Symbol Detection. *IEEE Trans. on Commun.*, 38(9):1391–1403, September 1990.

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

[14] B. Sklar. Rayleigh Fading Channels in Mobile Digital Communication Systems—Part II: Mitigation. *IEEE Commun. Mag.*, 35(7):102–109, July 1997.

[15] T. Ericson. A Gaussian Channel with Slow Fading. *IEEE Trans. on Inf. Theory*, 16:353–355, May 1970.

[16] R.J. McEliece and W.E. Stark. Channels with Block Interference. *IEEE Trans. on Inf. Theory*, 30(1):44–53, January 1984.

[17] G. Kaplan and S. Shamai (Shitz). Error Probabilities for the Block–Fading Gaussian Channel. *Archiv für Elektronik und Übertragungstechnik (International Journal of Electronics)*, 49(4):192–205, 1995.

[18] H. Imai and S. Hirakawa. A New Multilevel Coding Method Using Error Correcting Codes. *IEEE Trans. on Inf. Theory*, 23:371–377, 1977.

[19] U. Wachsmann, R.F.H. Fischer, and J.B. Huber. Multilevel Codes: Theoretical Concepts and Practical Design Rules. *IEEE Trans. on Inf. Theory*, 44(3):927–946, May 1999.

[20] E. Zehavi. 8–PSK Trellis Codes for a Rayleigh Channel. *IEEE Trans. on Commun.*, 40(5):873–884, May 1992.

[21] G. Caire, G. Taricco, and E. Biglieri. Bit–Interleaved Coded Modulation. *IEEE Trans. on Inf. Theory*, 44(3):927–946, May 1998.

[22] D. Raphaeli. Noncoherent Coded Modulation. *IEEE Trans. on Commun.*, 44:172–183, February 1996.

[23] G. Colavolpe and R. Raheli. Nonoherent Sequence Detection. *IEEE Trans. on Commun.*, 47(9):1376–1385, September 1999.

[24] G. Colavolpe, G. Ferrari, and R. Raheli. Nonoherent Iterative (Turbo) Decoding. *IEEE Trans. on Commun.*, 48(9):1488–1498, September 2000.

[25] M. Peleg and S. Shamai (Shitz). Iterative Decoding of Coded and Interleaved Noncoherent Multiple Symbol Detected DPSK. *Electronics Letters*, 33(12):1018–1020, June 1997.

[26] M. Peleg, S. Shamai (Shitz), and S. Galán. On Iterative Decoding for Coded Noncoherent MPSK Communications over Block–Noncoherent AWGN Channel. In *Proc. IEEE Int. Conf. Telecom. (ICT)*, Porto Carras, Greece, June 1998.

[27] I. D. Marsland and P. T. Mathiopoulos. On the Performance of Iterative Noncoherent Detection of Coded M–PSK Signals. *IEEE Trans. on Commun.*, 48(4):588–596, April 2000.

[28] R. van Nobelen and D.P. Taylor. Multiple Symbol Differential Detected Multilevel Codes for the Rayleigh–Fading Channel. *IEEE Trans. on Commun.*, 45(12):1529–1537, December 1997.

[29] L.E. Franks. *Signal Theory.* Prentice–Hall, Englewood Cliffs, NJ, 1969.

[30] D. Raphaeli. Improvement of Noncoherent Orthogonal Coding by Time Overlapping. *IEE Proc.-Comm.*, 141(5):309–311, October 1994.

[31] R. Knopp and H. Leib. M–ary Phase Coding for the Noncoherent AWGN Channel. *IEEE Trans. on Inf. Theory*, 40(6):1968–1984, November 1994.

[32] J.A.C. Bingham. *ADSL, VDSL, and Multicarrier Modulation.* John Wiley & Sons, Inc., New York, 2000.

[33] L. Lampe, R. Fischer, S. Calabrò, and S. Müller-Weinfurtner. Coded Modulation for DPSK on Fading Channels. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pages 2259–2263, Rio, Brazil, December 1999.

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

[35] M. Peleg and S. Shamai (Shitz). On the Capacity of the Blockwise Incoherent MPSK Channel. *IEEE Trans. on Commun.*, 46(5):603–609, May 1998.

[36] G. Colavolpe and R. Raheli. The Capacity of Noncoherent Channels. In *Proc. IEEE Int. Conf. Commun. (ICC)*, pages 1182–1186, Vancouver, June 1999.

[37] J.B. Huber. Multilevel Codes: Distance Profiles and Channel Capacity. In *ITG–Fachtagung Codierung für Quelle und Kanal (ITG–Fachbericht 130)*, pages 305–319, München, October 1994.

[38] 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*, pages 376–382, Ankara, Turkey, July 1990. Conf. Record.

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

[40] G.D. Forney, M.D. Trott, and S.-Y. Chung. Sphere–Bound–Achieving Coset Codes and Multilevel Coset Codes. *IEEE Trans. on Inf. Theory*, 46(3):820–850, May 2000.

[41] L. Lampe, S. Calabrò, R. Fischer, S. Müller-Weinfurtner, and J. Huber. On the Difficulty of Bit-Interleaved Coded Modulation for Differentially Encoded Transmission. In *Proc. IEEE Inf. Theory Workshop (ITW)*, page 111, Kruger National Park, SA, June 1999.

[42] S.S. Pietrobon, R.H. Deng, A. Lafanechere, G. Ungerboeck, and D.J. Costello Jr. Trellis–Coded Multidimensional Phase Modulation. *IEEE Trans. on Inf. Theory*, 36:63–89, January 1990.

[43] É.L. Leonardo, L. Zhang, and B.S. Vucetic. Multidimensional M–PSK Trellis Codes for Fading Channels. *IEEE Trans. on Inf. Theory*, 42:1093–1108, July 1996.

[44] M. Peleg and S. Shamai (Shitz). On Coded and Interleaved Noncoherent Multiple Symbol Detected MPSK. *Europ. Trans. Telecommun. (ETT)*, 10(1):65–73, January/February 1999.

[45] G. Ungerböck. Channel Coding with Multilevel/Phase Signals. *IEEE Trans. on Inf. Theory*, 28(1):55–67, January 1982.

[46] C. Berrou and A. Glavieux. Near Optimum Limit Error Correcting Coding and Decoding: Turbo–Codes. *IEEE Trans. on Commun.*, COM–44:1261–1271, Oct. 1996.

[47] R. Fischer, L. Lampe, S. Müller-Weinfurtner, and J. Huber. Coded Modulation for Differential Encoding and Non-Coherent Reception on Fading Channels. In *3. ITG Conference Source and Channel Coding*, pages 129–146, München, January 2000.

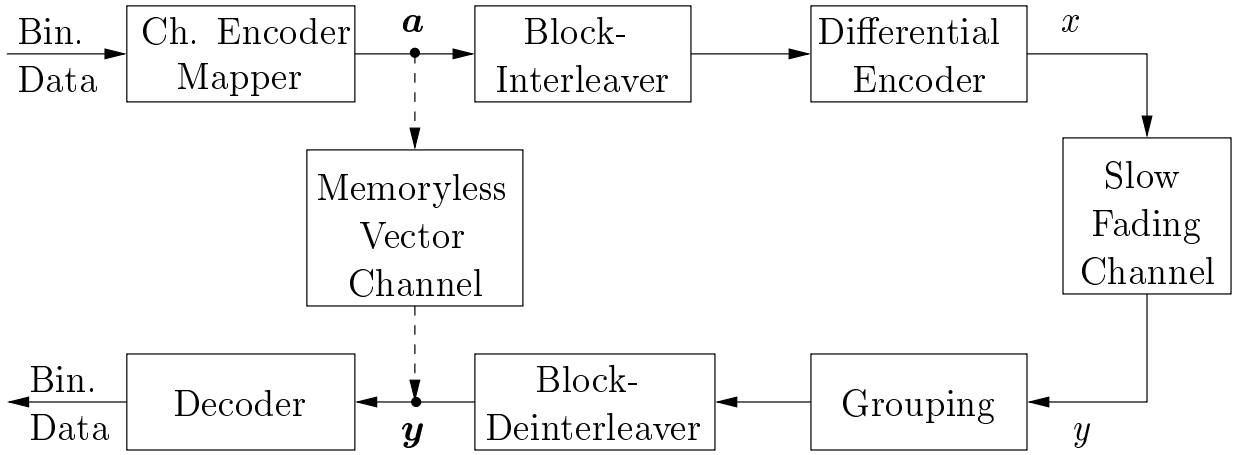

Figure 1: *System model.*

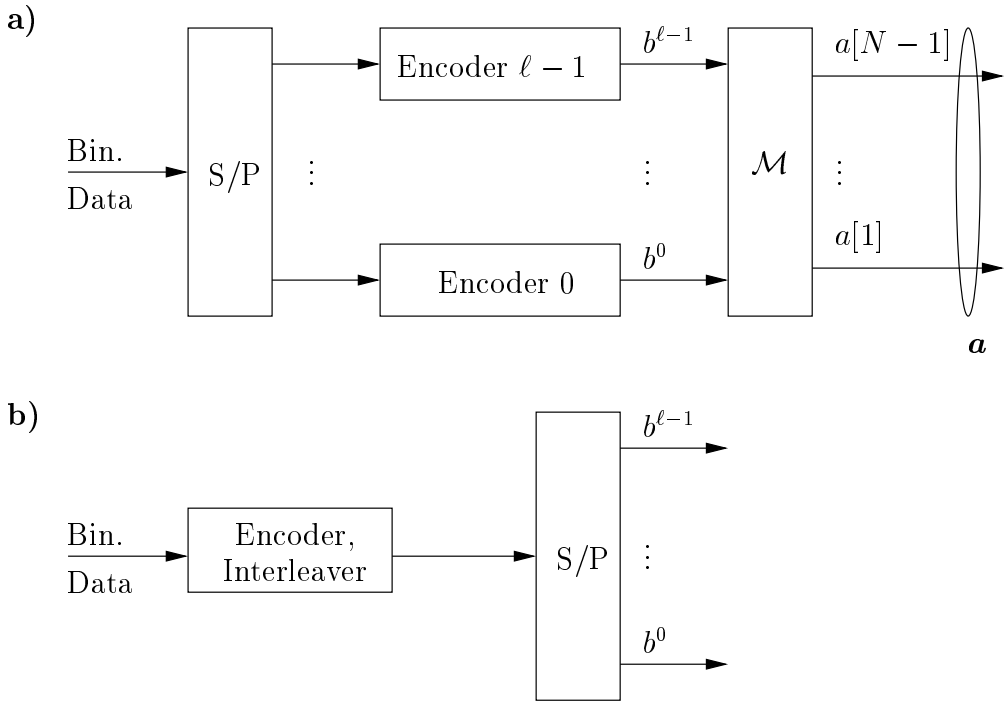

Figure 2: *Channel encoding and Mapping: a) MLC, b) Changes for BICM.*

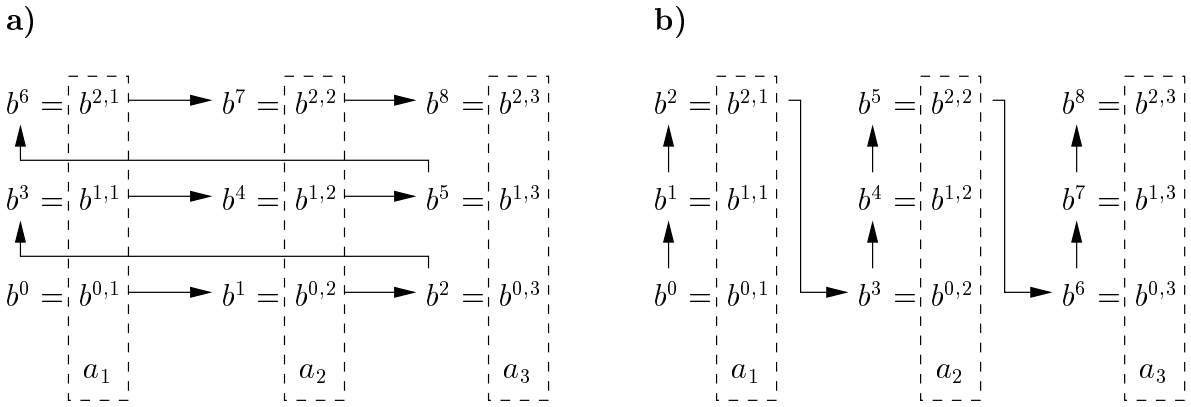

Figure 3: *Example of MLC with separated labeling for 8DPSK and $N = 4$. a): first sorting according to significance, then according to position. b): first sorting according to position, then according to significance.*

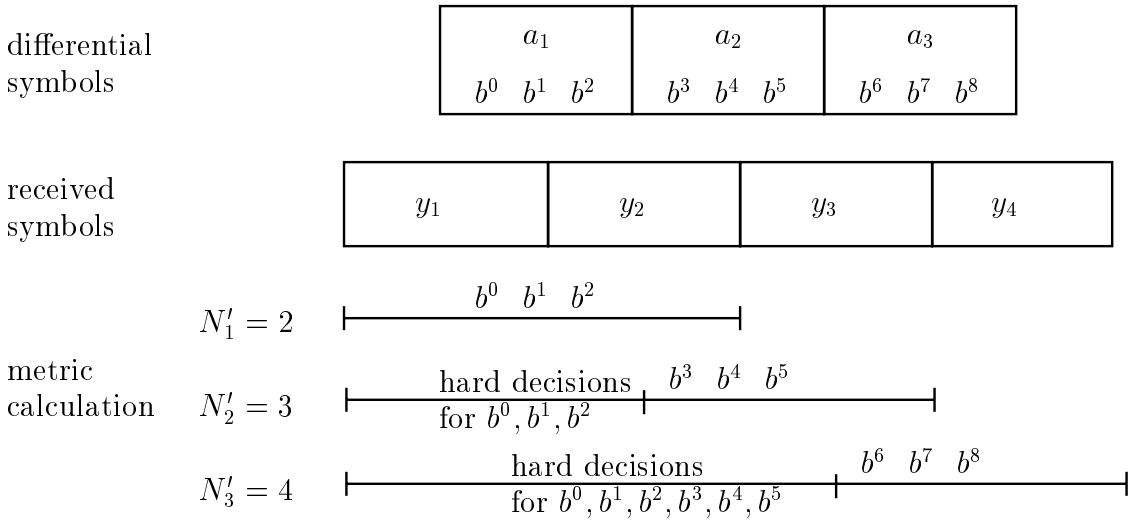

Figure 4: *MLC/MSD with MSDD and increasing observation window $N_\nu' = \nu+1$. $N = 4$ and 8DPSK. Differential symbols a_ν, received symbols y_ν, $1 \le \nu \le N(-1)$, address bits b^i, $0 \le i \le 8$.*

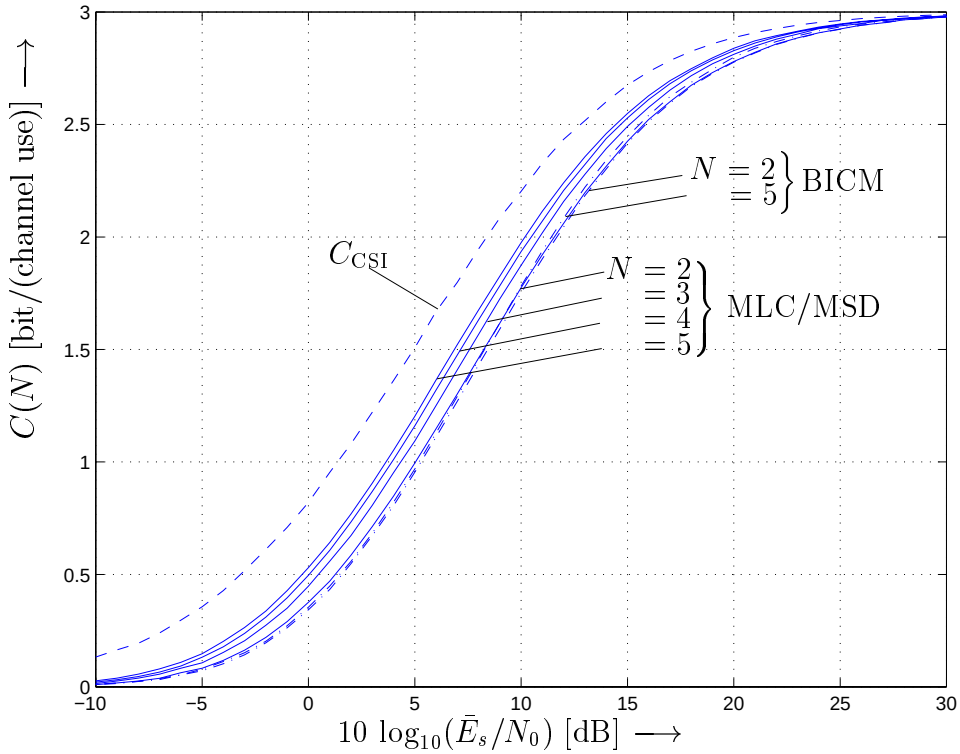

Figure 5: *Normalized capacity for 8DPSK. Solid lines: MLC/MSD, $N = 2, 3, 4, 5$ (from right to left). Dash–dotted lines: BICM with GL, $N = 2, 5$ (from right to left). Dashed Line: Capacity C_{CSI}.*

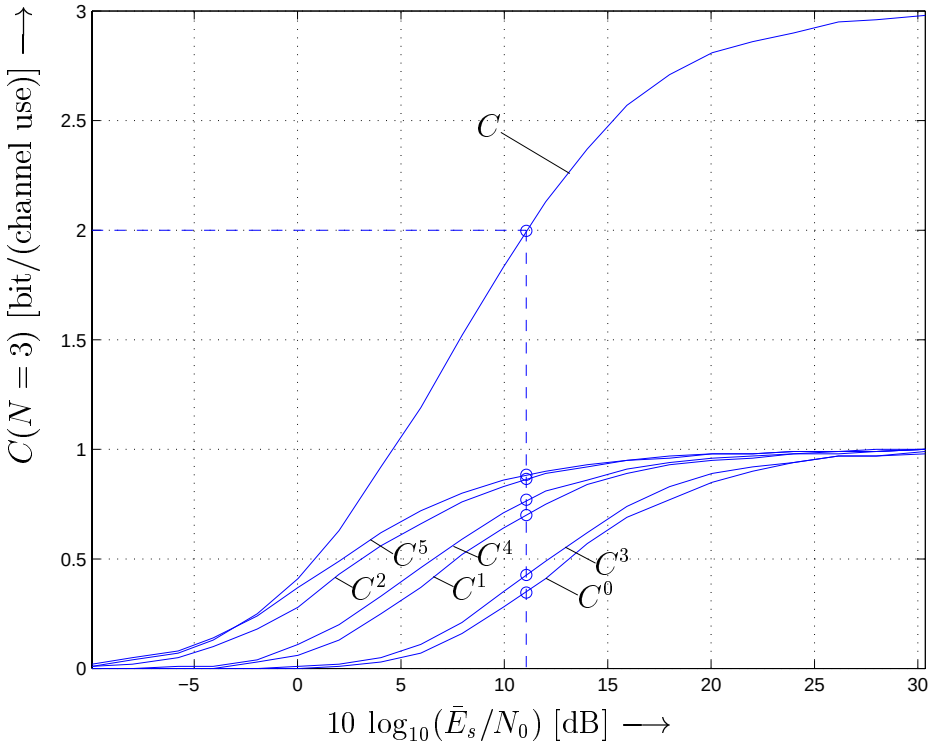

Figure 6: *Capacity $C(N = 3)$ and capacities $C^0, \ldots, C^5$ of the equivalent channels for 8DPSK. Solid lines: MLC with UL. Dashed lines: Rate design for $C = 2$ bit/(channel use).*

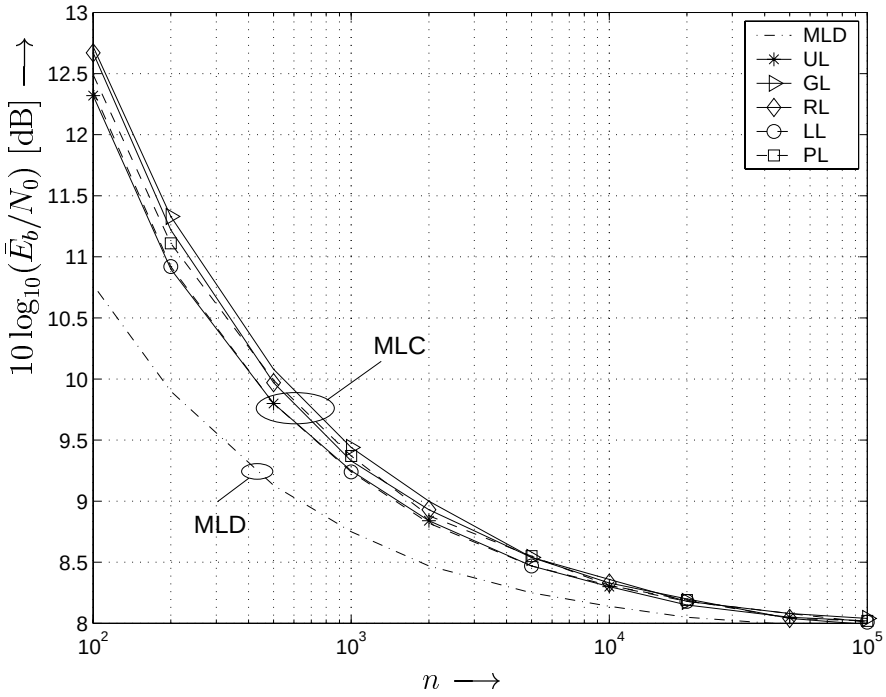

Figure 7: *Required $\bar{E}_b/N_0$ over code length n for $p_w = 10^{-3}$. 8DPSK with 2 bits/(channel use). MLC/MSD with various labelings for MSDD with $N = 3$. Reference curve: MLD.*

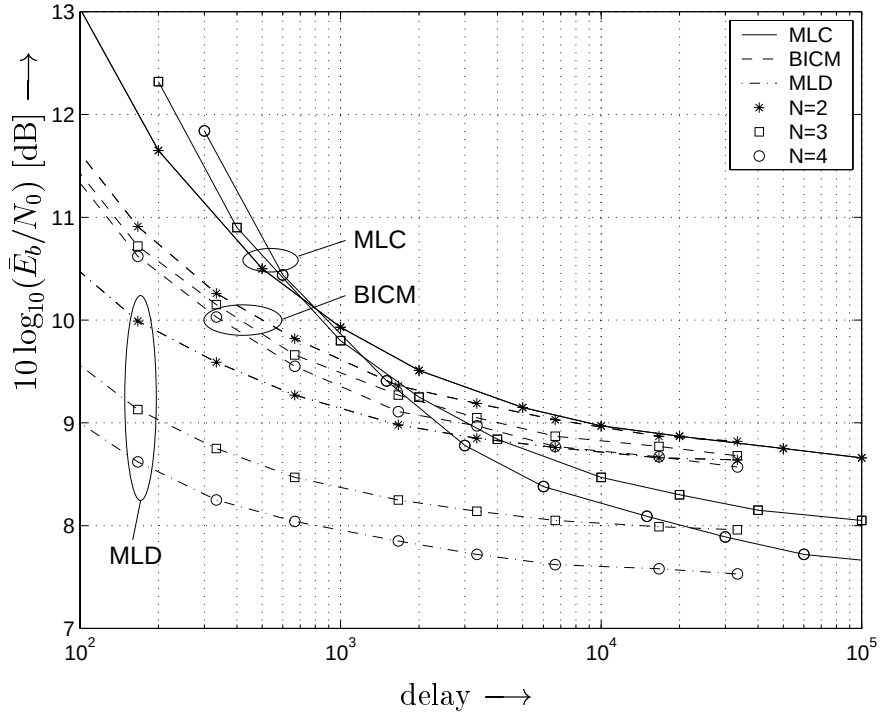

Figure 8: *Required $\bar{E}_b/N_0$ over decoding delay for $p_w = 10^{-3}$. 8DPSK with 2 bits/(channel use). MLC/MSD with UL vs. BICM with GL for MSDD with $N = 2, 3, 4$. Reference curves: MLD.*

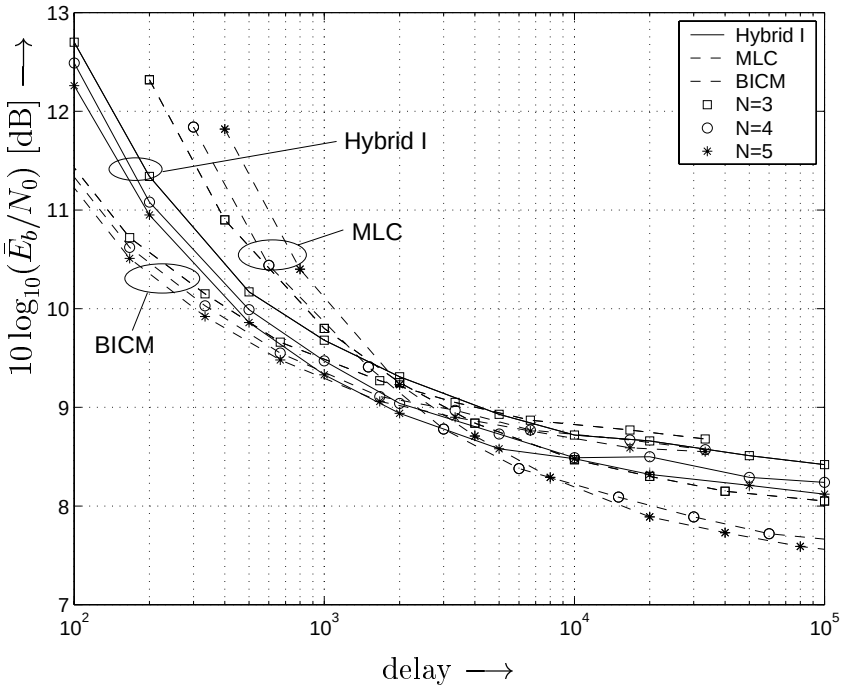

Figure 9: *Required $\bar{E}_b/N_0$ over decoding delay for $p_w = 10^{-3}$. 8DPSK with 2 bits/(channel use). MSDD with $N = 3, 4, 5$. Solid Lines: Hybrid I. Dashed Lines: MLC and BICM.*

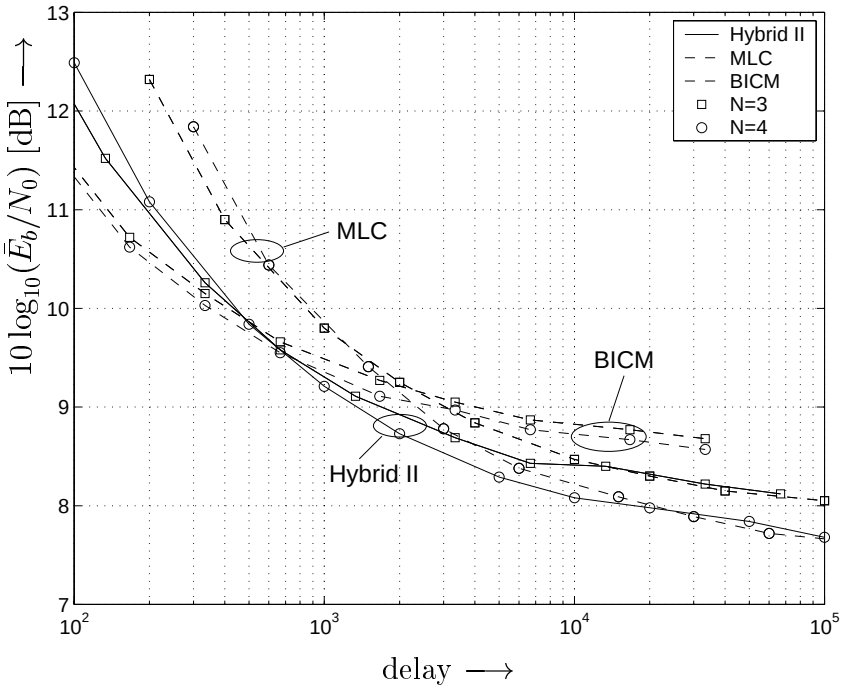

Figure 10: *Required $\bar{E}_b/N_0$ over decoding delay for $p_w = 10^{-3}$. 8DPSK with 2 bits/(channel use). MSDD with $N = 3, 4$. Solid Lines: Hybrid II. Dashed Lines: MLC and BICM.*

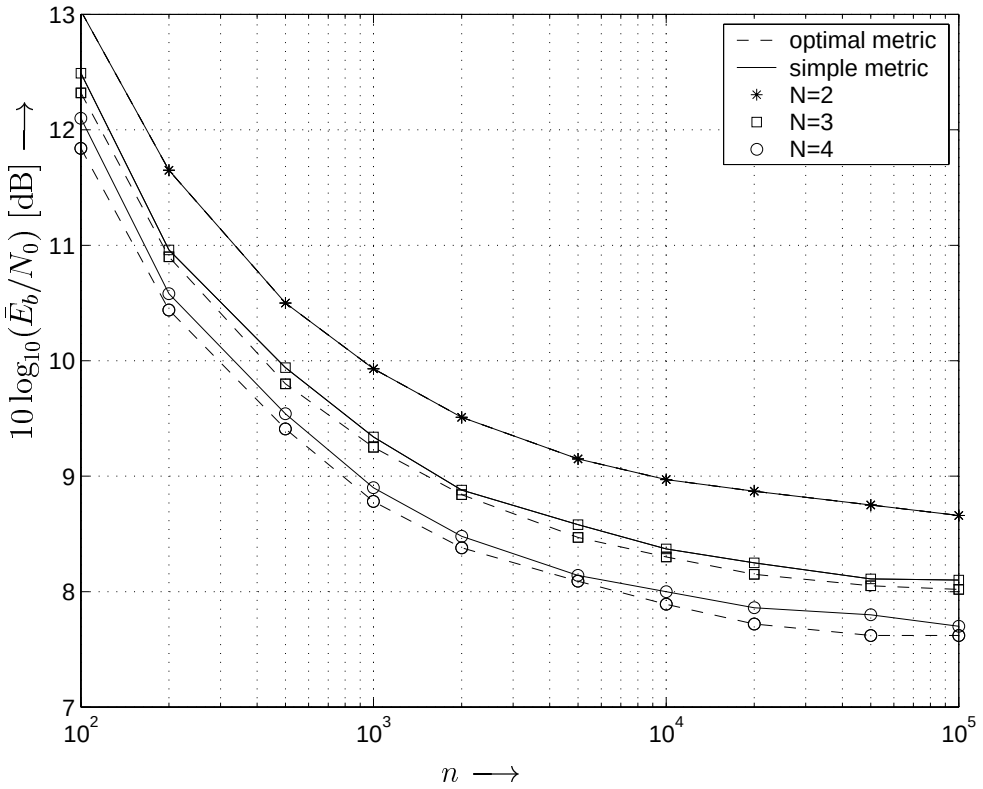

Figure 11: *Required $\bar{E}_b/N_0$ over code length n for $p_w = 10^{-3}$. 8DPSK with 2 bits/(channel use). MLC/MSD with UL and optimal (solid lines) vs. simplified metric computation (dashed lines) for MSDD with $N = 2, 3, 4$.*