# Coded Modulation for Noncoherent Reception with Application to OFDM

Robert F.H. Fischer, *Member, IEEE*, Lutz H.-J. Lampe, *Student Member, IEEE*,
Stefan H. Müller–Weinfurtner, *Member, IEEE*

*Abstract* — **Power and bandwidth efficient differentially encoded transmission over slowly time–varying fading channels with noncoherent reception and without channel state information is considered. For high bandwidth efficiencies, combined phase and amplitude modulation is used. For increased power efficiency, channel coding and multiple–symbol differential detection are applied, i.e., interleaving and detection are based on blocks of $N > 2$ consecutive symbols. The presented concepts are directly applicable to transmission over flat fading channels. However, concentrating on the situation of frequency–selective channels, we consider their application to multicarrier transmission using orthogonal frequency–division multiplexing (OFDM). When coding across subcarriers, OFDM transforms the actual frequency–selective channel into a slowly time–varying, frequency–non–selective fading channel. The paper presents a design for multilevel coding schemes to approach theoretical limits for power and bandwidth efficient noncoherent transmission over the equivalent fading channel. It is shown that bit–interleaved coded modulation, which relies on Gray labeling, is competitive only in case of conventional differential detection with $N = 2$. The theoretic considerations are well approved by simulation results, where turbo codes are applied as component codes.**

Index Terms: Multilevel coding, bit–interleaved coded modulation, OFDM, differential encoding, noncoherent reception

# 1 Introduction

In many transmission scenarios, e.g., for wireless communications, transmission systems have to cope with time–varying, frequency–selective channels due to multipath propagation. If, as usual, the channel is slowly time–varying compared to the transmitted symbol time, multicarrier modulation, which is often denoted as *orthogonal frequency–division multiplexing (OFDM)* [1, 2], is an attractive transmission technique. By use of OFDM, the actual dispersive channel is partitioned into frequency–*non*–selective narrow–band channels, the so–called tones or subcarriers. If channel state information is not available at the transmitter error correction coding is favorably performed across these tones to provide diversity in frequency direction. Then, the transmission channel is represented by a slowly time–varying, frequency–non–selective fading channel, where the current *channel state* is given by the respective value of the transfer function of the underlying channel.

Since reliable estimation of channel state information is often not practical, noncoherent detection at the receiver is advantageous. In order to avoid the problem of phase ambiguity in noncoherent receivers, classical differential phase–shift keying (DPSK) is performed at the transmitter. On the one hand, in state–of–the–art noncoherent receivers the decision variables are based on several received symbols within an observation interval of $N \geq 2$ symbols in order to take the dependences among the received symbols into account, cf. e.g. [3, 4, 5, 6, 7]. On the other hand, for increased reliability channel coding is required. Here, interleaving of decision variables is necessary in order to mitigate the effects of fading [8]. Both requirements can be met by performing *multiple–symbol differential detection (MSDD)*, where blocks of $N$ received symbols are independently evaluated to provide decision variables on $N-1$ information symbols [3, 9, 10, 11]. Thereby, the blocks overlap by one symbol. Moreover, if transmission with high bandwidth efficiency is desired, the straightforward extension of differential phase encoding is to transmit the information both in phase and in amplitude changes, which is known as *differential amplitude– and phase–shift keying (DAPSK)*, e.g. [12, 13].

In this paper, coded modulation schemes employing differential encoding of APSK constellations across the OFDM subchannels and MSDD reception without channel state information (CSI) for transmission over time–varying, frequency–selective channels are discussed. We focus on the application of *multilevel coding (MLC)* [14, 15] and *bit–interleaved coded modulation (BICM)* [16, 17] as competing coded modulation techniques. In particular, the design of MLC and the issue of Gray labeling, which BICM relies on, are treated in detail. Based on information theoretic arguments and by means of simulations we show that BICM is not well suited for differentially encoded transmission

and MSDD, whereas properly designed MLC schemes make *power and bandwidth efficient* digital transmission close to theoretical limits possible. As we employ turbo codes [18] as component codes, and we aim at approaching the ultimate limits as closely as possible, channel capacity is considered as reference for performance evaluation.

The paper is organized as follows: In Section 2 the system model is presented and the associated capacity is calculated. In Section 3 the coded modulation schemes are given. The design of MLC according to the capacity rule and the achievable transmission rates when applying MLC and BICM are shown in Section 4 for examples most relevant in practice. Simulation results also presented in Section 4 are in great accordance with the effects predicted from information theory. Finally, conclusions are drawn in Section 5.

# 2 System Model and Capacity

The discrete–time system model applied throughout the paper is sketched in Figure 1. The channel and all signals are given in the equivalent low–pass domain, i.e., all quantities are complex–valued, cf. e.g. [19].

## 2.1 Fading Channel Model of OFDM

We consider the situation where a number of distinctively delayed and attenuated versions of the transmitted signal are seen at the receiver side, which is typical for many applications, e.g. in wireless and powerline communications. Furthermore, we assume the communication channel to be time–varying, but (almost) static over a period of one OFDM symbol. Then, this multipath transmission can be represented by a discrete–time channel impulse response $h[\nu]$, $0 \leq \nu \leq p$, of finite length $p+1$. The coefficients $h[\nu]$ are mutually independent zero–mean complex Gaussian distributed random variables with variances $\sigma^2_{h[\nu]}$ according to a given channel power–delay profile. This model is often used for realistic scenarios, cf. e.g. [20].

Multicarrier transmission using OFDM has attracted a lot of attention because of its ability to combat effectively the problems with multipath propagation, cf. e.g. [21]. The transmitted symbols are denoted by $x'[k_\mathrm{c}]$ ($k_\mathrm{c} \in \mathbb{Z}$: discrete–time index of the actual channel). In OFDM, each block of $D$ symbols $x'[k_\mathrm{c}]$ is prefixed with the $D_0$ last samples, the so–called guard interval [2], of the same block at the transmitter, and $D$ out of $D+D_0$ received symbols $y'[k_\mathrm{c}]$ are further processed at the receiver. Thereby, if the length $D_0$ of the guard interval is chosen such that $D_0 \geq p$, the linear convolution of transmitted signal and channel impulse response is converted into a cyclic convolution. Subsequently,

we always assume that $D_0 \geq p$ holds. The IDFT/DFT–pair ((I)DFT: (inverse) discrete Fourier transform) of OFDM resolves this cyclic convolution such that the transmission between the IDFT input and the DFT output is performed in $D$ parallel, independent subchannels. For coding across these subchannels, a frequency–non–selective (flat) fading channel between $x[k_\mathrm{f}]$ and $y[k_\mathrm{f}]$ (cf. Figure 1, $k_\mathrm{f} \in \mathbb{Z}$: discrete–time index of the equivalent fading channel) with complex fading gains $g[k_\mathrm{f}]$ equal to samples of the underlying channel transfer function is obtained:

$$y[k_\mathrm{f}] = g[k_\mathrm{f}] \cdot x[k_\mathrm{f}] + n[k_\mathrm{f}] \,, \tag{1}$$

where $n[k_\mathrm{f}]$ denotes independent additive white Gaussian noise (AWGN) with variance $\sigma_n^2$.

By virtue of the DFT the fading gains $g[k_\mathrm{f}]$ are *correlated*, zero–mean complex Gaussian random variables. For differential encoding and noncoherent reception as described in Section 2.2, the correlation of fading samples corresponding to subchannels of one OFDM symbol has to be considered. The autocorrelation function is given by the DFT of the power–delay profile ($\mathrm{E}_\times\{\cdot\}$ denotes expectation with respect to $\times$, $G[k_\mathrm{f}]$ denotes the random variable corresponding to $g[k_\mathrm{f}]$)[1]

$$\mathrm{E}_G\{G[k_\mathrm{f} + \Delta k_\mathrm{f}] G^*[k_\mathrm{f}]\} = \sum_{\nu=0}^{p} \sigma_{h[\nu]}^2 \mathrm{e}^{-\mathrm{j}\frac{2\pi}{D}\nu \cdot \Delta k_\mathrm{f}} \,. \tag{2}$$

Given a particular communication environment, the values of $p$, $D$, and $\sigma_{h[\nu]}^2$ can be quantified and introduced in (2), cf. e.g. [22, 23]. Here, we focus on the discussion of coded modulation applied to noncoherent OFDM transmission in general. Consequently, the relation of the transmission parameters is specified rather than their actual quantities. In particular, we assume the channel length $p$ to be significantly shorter than the DFT size $D$, i.e., adjacent OFDM subchannels and thus, consecutive channel gains, are highly correlated (cf. Eq. (2)). More specifically, as very simple model of correlation, we expect the channel state $g[k_\mathrm{f}]$ to be constant over a block of $N$ consecutive symbols. This *block fading* model is convenient for the discussion of modulation and coding applied to noncoherent OFDM, since the number of adjustable parameters is limited. Of course, the incorporation of other correlation models is also possible.

[1] Throughout the paper we denote random variables corresponding to signals by the respective capital letter.

## 2.2 Differential Encoding and MSDD

In order not to require knowledge on the actual channel state, we employ differential encoding across the OFDM subchannels at the transmitter and MSDD at the receiver. Hence, information is represented in the transitions between consecutive fading channel input symbols $x[k_\mathrm{f}]$ rather than in the symbols themselves (cf. e.g. [1, 24, 3]).

Moreover, if high bandwidth efficiency ($> 2$ bit/s/Hz) is required it is advantageous not to solely modulate the phase but also represent information in the amplitude of the transmitted signal. Signal constellations $\mathcal{X}$ for the transmitted signal $x[k_\mathrm{f}]$, suited for differential encoding, are presented in [25]. These constellations consist of $M = \alpha \cdot \beta$ points arranged in $\alpha$ distinct concentric rings with radii taken from $\mathcal{R} = \{r_0, r_1, \ldots, r_{\alpha-1}\}$, $r_i < r_{i+1}$, and $\beta$ uniformly spaced phases. To avoid ambiguities, we denote these constellations by $\alpha\mathrm{A}\beta\mathrm{PSK}$. As usual for DPSK, we restrict the "differential symbols" $a[k_\mathrm{f}]$ to be taken from the same signal set as the transmit signal, i.e., $\mathcal{A} = \mathcal{X}$. Then, a simple formulation for the differential encoding results. Given $x[k_\mathrm{f}-1] = r_{i[k_\mathrm{f}-1]}\mathrm{e}^{\mathrm{j}\varphi_{n[k_\mathrm{f}-1]}}$ and $a[k_\mathrm{f}] = r_{j[k_\mathrm{f}]}\mathrm{e}^{\mathrm{j}\varphi_{m[k_\mathrm{f}]}}$, the encoder outputs

$$x[k_\mathrm{f}] = r_{(i[k_\mathrm{f}-1]+j[k_\mathrm{f}]) \bmod \alpha}\mathrm{e}^{\mathrm{j}\varphi_{(n[k_\mathrm{f}-1]+m[k_\mathrm{f}]) \bmod \beta}} \ . \tag{3}$$

Here, $i[k_\mathrm{f}]$ and $j[k_\mathrm{f}]$ are the index sequences of the radii, and $n[k_\mathrm{f}]$, $m[k_\mathrm{f}]$ are the respective index sequences of the phases.

Since we assume the fading channel to be (almost) constant for $N$ symbols, and additional information about the fading statistics is generally not available, it is reasonable to base the metric computation on the observation of $N$ consecutively received symbols. In MSDD, $N$ symbols $y[k_\mathrm{f}]$ are grouped and comprised into vectors ($k_\mathrm{v} \in \mathbb{Z}$: vector symbol time index) $\boldsymbol{y}[k_\mathrm{v}] \triangleq [y[k_\mathrm{v}(N-1)], \ldots, y[k_\mathrm{v}(N-1)+N-1]]$, which are overlapping by one symbol, cf. [11, Fig. 3]. In order to enable standard coding techniques and to combat the effects of fading, *sufficient* deinterleaving $\Pi^{-1}$ of $\boldsymbol{y}[k_\mathrm{v}]$ is performed. Of course, at the transmitter side interleaving $\Pi$ of vector symbols is necessary, where the vector $\boldsymbol{a}[k_\mathrm{v}] \triangleq [a[k_\mathrm{v}(N-1)], \ldots, a[k_\mathrm{v}(N-1)+N-2]]$ corresponds to $\boldsymbol{y}[k_\mathrm{v}]$. The respective non–interleaved versions of the vector symbols are denoted by $\boldsymbol{x}[\kappa_\mathrm{v}]$ and $\boldsymbol{y}[\kappa_\mathrm{v}]$ (cf. Fig. 1), where $k_\mathrm{v} = \Pi(\kappa_\mathrm{v})$.

With respect to the underlying OFDM transmission, interleaving is basically performed in the frequency–domain (across subchannels). For further increase of diversity, i.e., to accomplish *sufficient* decorrelation of the decoder input, time–domain interleaving, frequency hopping, and/or the use multiple transmitter/receiver antennas might be applied (cf. [21, 26] and references therein).

## 2.3 Vector Channel and Capacity

By the application of OFDM and differential encoding with MSDD a *vector channel* between $\boldsymbol{a}[\kappa_\mathrm{v}]$ and $\boldsymbol{y}[\kappa_\mathrm{v}]$ is obtained (cf. Fig. 1), which due to the interleaving can be regarded to be memoryless. For an optimum metric calculation in the decoder the complex $N$–dimensional probability density function (pdf) $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ of $\boldsymbol{y}[\kappa_\mathrm{v}]$ for given $\boldsymbol{a}[\kappa_\mathrm{v}]$ is required, which is independent of the index $\kappa_\mathrm{v}$. Defining the $N$–dimensional vector $\boldsymbol{x}[k_\mathrm{v}] \triangleq [x[k_\mathrm{v}(N-1)], \ldots, x[k_\mathrm{v}(N-1)+N-1]]$ of $N$ transmitted symbols corresponding to the received vector $\boldsymbol{y}[k_\mathrm{v}]$, $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ is obtained by averaging $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x})$ over all possible reference symbols $x[k_\mathrm{v}(N-1)] \triangleq x_\mathrm{ref} \in \mathcal{X}$:

$$p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a}) \triangleq \mathrm{E}_{X_\mathrm{ref}}\left\{p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x})\right\} \ , \tag{4}$$

where for flat Rayleigh fading with variance normalized to 1, the pdf $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x})$ reads [27]

$$p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\pi^N(\sigma_n^2)^{N-1}} \cdot \frac{1}{|\boldsymbol{x}|^2+\sigma_n^2} \cdot \exp\left(-\frac{1}{\sigma_n^2}\left[|\boldsymbol{y}|^2 - \frac{|\boldsymbol{y}\boldsymbol{x}^H|^2}{|\boldsymbol{x}|^2+\sigma_n^2}\right]\right) \tag{5}$$

Noteworthy, in case of PSK constellations $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{x})$ does not depend on $x_\mathrm{ref}$, and thus averaging in (4) can be omitted.

The metric computation based on $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ and the implementation of encoder and mapper are discussed in detail in Section 3. Here, we use $p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})$ to give an upper bound for the performance applying powerful channel coding in the system of Figure 1.

We consider the capacity of the memoryless vector channel with discrete input and continuous, complex–valued output. The average mutual information $I(\cdot\,;\cdot)$, measured in bits per vector symbol, is obtained by [28]

$$I(\boldsymbol{Y};\boldsymbol{A}) = \mathrm{E}_{\boldsymbol{Y},\boldsymbol{A}}\left\{\log_2\left(\frac{p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})}{p_{\boldsymbol{Y}}(\boldsymbol{y})}\right)\right\} \ , \tag{6}$$

where $p_{\boldsymbol{Y}}(\boldsymbol{y})$ is the average pdf of the channel output. It is well–known that for capacity calculation an optimization over all free parameters has to be performed. On the one hand, the ring ratios $r_{i+1}/r_i$, $i = 0, 1, \ldots, \alpha-2$, of the APSK constellation have to be chosen. On the other hand, the probabilities of the rings have to be regarded. Due to the rotational invariance of APSK constellations, uniformly distributed phases are optimal. Because of averaging for differential encoding of the amplitudes, uniformly distributed amplitudes result regardless the probabilities of the differential signal points $a[k]$. For these reasons, we fix the symbols $\boldsymbol{a}[\kappa_\mathrm{v}]$ to be uniformly, independently and identically distributed maximizing the throughput of the channel.

It should be noted that uniformly distributed APSK constellations already exhibit some kind of signal shaping. Since inner points are closer to each other than points at

the perimeter of the constellation, a non–uniform density is present. Hence, by selecting and optimizing the ring ratio, this density can be adjusted. Note that this procedure is comparable to the way of approaching AWGN channel capacity in [29], and has similarities with warping [30], where regularly spaced signal constellations are non–linearly distorted.

Consequently, we will use the term "capacity" for the mutual information given the uniformly distributed constellation and normalized with respect to the number $N-1$ of differential symbols:

$$C(N) \triangleq \frac{1}{N-1} \cdot I(\boldsymbol{Y};\boldsymbol{A}) = \frac{1}{N-1} \cdot \mathrm{E}_{\mathbf{Y},\mathbf{A}} \left\{ \log_2 \left( \frac{p_{\boldsymbol{Y}}(\boldsymbol{y}|\boldsymbol{a})}{p_{\boldsymbol{Y}}(\boldsymbol{y})} \right) \right\} . \tag{7}$$

Clearly, if the time–variance of the Rayleigh fading channel (coherence bandwidth of the actual channel) allows a longer observation interval $N$, $C(N)$ increases. If $N$ approaches infinity the channel is static (the actual channel is frequency–non–selective) and thus its state can be exactly estimated, e.g., by means of pilot symbols. Hence, $C(N)$ converges to the capacity $C_{\mathrm{CSI}}$ for the case of perfect channel state information (CSI) and coherent reception, which upper bounds $C(N)$, cf. e.g. [31, 6].

# 3 Coded Modulation

In this section, the application of *multilevel coding (MLC)* [14, 15] and *bit–interleaved coded modulation (BICM)* [16, 17] as powerful coded modulation schemes for transmission over the vector channel described in the previous section is shown. Here, we restrict the discussion to binary component codes. As $N-1$ symbols are comprised in the vector[2] $\boldsymbol{a}$ 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 amplitude and phase changes.

## 3.1 Multilevel Coding/Multistage Decoding

The capacity of a modulation scheme can be achieved by multilevel encoding and *multistage decoding (MSD)* if and only if the individual rates of the component codes are properly chosen, cf. e.g. [32, 15, 33]. Applying this procedure to the present vector fading channel, we arrive at a scheme, optimal in the sense of information theory.

[2] In this section, time indices are omitted for clarity.

Since the mapping $\mathcal{M}$ is bijective and using the mutual information chain rule [28] we can rewrite (7) as ($B^i$ denotes the random variable corresponding to $b^i$)

$$C(N) = \frac{1}{N-1} \cdot I(\boldsymbol{Y}; \boldsymbol{A}) = \frac{1}{N-1} \cdot I(\boldsymbol{Y}; B^0, B^1, \ldots, B^{\ell-1}) = \frac{1}{N-1} \cdot \sum_{i=0}^{\ell-1} C^i(N) \;, \quad (8)$$

with

$$C^i(N) \stackrel{\triangle}{=} I(\boldsymbol{Y}; B^i | [B^0, B^1, \ldots, B^{i-1}]) \;. \quad (9)$$

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 $\ell$ *equivalent channels* with capacity $C^i(N)$, provided that $b^0, \ldots, b^{i-1}$ are known (cf. [14, 15]). In MLC the digits $b^i$, $i = 0, \ldots, \ell-1$, result from independent encoding of the data symbols, see Fig. 2a).

According to [15, 32, 33] 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 [15]. 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 imaginable. Thereby, heuristically, also the error propagation from level to level can be taken into account. We emphasize that "classical" design parameters such as minimum squared Euclidean distance, minimum Hamming distance or product distance are not regarded.

From the chain rule of information theory it follows that the component codes $\mathcal{C}^i$ can be 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 $j < i$. Regarding (4), and taking into account that all differential symbols $\boldsymbol{a}$ are equally likely, the metric for maximum–likelihood (ML) decoding of code $\mathcal{C}^i$ at level $i$ is derived from the pdf

$$p_{\boldsymbol{Y}}(\boldsymbol{y} | b^i = c, [\hat{b}^0, \ldots, \hat{b}^{i-1}]) = \frac{1}{|\mathcal{L}_c^i(\hat{b}^0, \ldots, \hat{b}^{i-1})|} \sum_{\boldsymbol{a} \in \mathcal{L}_c^i(\hat{b}^0, \ldots, \hat{b}^{i-1})} p_{\boldsymbol{Y}}(\boldsymbol{y} | \boldsymbol{a}) \;. \quad (10)$$

Thereby, estimates $\hat{b}^j$, $j = 0, \ldots, i-1$, of the binary symbols at previous decoding stages are known and 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 $b^i = c$:

$$\mathcal{L}_c^i(b^0, \ldots, b^{i-1}) = \left\{ \boldsymbol{a} \;\middle|\; \begin{array}{c} \boldsymbol{a} = \mathcal{M}([b^0, \ldots, b^{i-1}, c, \delta^{i+1}, \ldots, \delta^{\ell-1}]), \\ \delta^j \in \{0, 1\},\; j = i+1, \ldots, \ell-1 \end{array} \right\} \;. \quad (11)$$

Additionally, regarding $|\mathcal{L}_0^i(b^0, \ldots, b^{i-1})| = |\mathcal{L}_1^i(b^0, \ldots, b^{i-1})|$, e.g., the log–likelihood ratio $\lambda^i$ for $b^i$ at level $i$ reads:

$$\lambda^i = \ln \frac{\sum_{\boldsymbol{a} \in \mathcal{L}_1^i(\hat{b}^0, \ldots, \hat{b}^{i-1})} p_{\boldsymbol{Y}}(\boldsymbol{y} | \boldsymbol{a})}{\sum_{\boldsymbol{a} \in \mathcal{L}_0^i(\hat{b}^0, \ldots, \hat{b}^{i-1})} p_{\boldsymbol{Y}}(\boldsymbol{y} | \boldsymbol{a})} \;. \quad (12)$$

In [15] it is shown that MLC can approach capacity with any labeling. For finite code length and coherent reception on the AWGN channel Ungerböck labeling [34] has some advantage. Thus, also here 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 optimal 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.

## 3.2 Bit–Interleaved Coded Modulation

In *bit–interleaved coded modulation (BICM)* only one binary code in combination with bit–interleaving is applied, and $\ell$ encoded bits are grouped to select the current symbol, see Fig. 2b). At the decoder, the dependences between address bits of one vector symbol are neglected. Hence,

$$C^i_{\mathrm{BICM}}(N) = I(\boldsymbol{Y};B^i) \le I(\boldsymbol{Y};B^i|[B^0,B^1,\ldots,B^{i-1}]) = C^i(N)\ , \quad i=0,1,\ldots,\ell-1 \tag{13}$$

is the usable capacity at level $i$.

The metric for the $i^{\mathrm{th}}$ address bit for ML decoding is calculated similar as above. Here, because no preselection is done, the sets $\mathcal{L}^i_c(\hat{b}^0,\ldots,\hat{b}^{i-1})$ have to be replaced by

$$\mathcal{L}^i_{c,\mathrm{BCIM}} = \left\{ \boldsymbol{a} \;\middle|\; \begin{array}{c} \boldsymbol{a} = \mathcal{M}([\delta^0,\ldots,\delta^{i-1},c,\delta^{i+1},\ldots,\delta^{\ell-1}]), \\ \delta^j \in \{0,1\},\ j \ne i \end{array} \right\}\ , \tag{14}$$

and the respective pdf for deriving the decoding metrics reads

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

Whereas MLC/MSD approaches capacity $C(N)$ and thus, its performance for $N \to \infty$ is bounded by $C_{\mathrm{CSI}}$, for BICM, based on the metric expression (15), another upper bound $C^u_{\mathrm{BICM}} < C_{\mathrm{CSI}}$ can be found. For symbol–interleaved $M$–ary DPSK over the noncoherent AWGN channel it is shown in [35] that MSDD with observation interval $N$ cannot outperform the ML detector using only two consecutive symbols, but with the initial phase given with ambiguity of $2\pi/M$ (differentially coherent DPSK with $N=2$). The derivation in [35] gives immediately the upper bound for our situation

$$C_{\mathrm{BICM}}(N) = \frac{1}{N-1}\cdot\sum_{i=0}^{\ell-1} C^i_{\mathrm{BICM}}(N) \le C^u_{\mathrm{BICM}} = \sum_{j=0}^{\log_2(M)-1} I(\boldsymbol{Y}'=[Y_1,Y_2];B^j|G)\ , \tag{16}$$

where $0 \leq j \leq \log_2(M) - 1$, since $\boldsymbol{y}'$ consists of two consecutively received symbols, only.

For several applications it has been shown [16, 17] that BICM suffers only a marginal capacity loss compared to MLC provided Gray labeling is applied. On the one hand, in practice, the advantage of BICM is its simple implementation, i.e., only one binary code is required compared to $\ell$, in general different, codes in MLC. On the other hand, BICM strongly relies on Gray labeling.

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 $\mathrm{PEP}(\boldsymbol{a} \to \hat{\boldsymbol{a}}) = \max_{\tilde{\boldsymbol{a}} \in \mathcal{A}^{N-1} \setminus \{\boldsymbol{a}\}} \mathrm{PEP}(\boldsymbol{a} \to \tilde{\boldsymbol{a}})$. For DPSK and Rayleigh fading, the pairwise error probabilities are evaluated in [11]. From the results therein it is apparent that the maximum of $\mathrm{PEP}(\boldsymbol{a} \to \tilde{\boldsymbol{a}})$ corresponds to the minimum of the *noncoherent distance* [36], cf. also [37, 38],

$$d^2_{\mathrm{NC}}(\tilde{\boldsymbol{a}}, \boldsymbol{a}) \stackrel{\triangle}{=} N r_0^2 - |\tilde{\boldsymbol{x}} \boldsymbol{x}^H| \,, \tag{17}$$

where $\tilde{\boldsymbol{x}}$ ($\boldsymbol{x}$) is related to $\tilde{\boldsymbol{a}}$ ($\boldsymbol{a}$) via differential encoding.

In Appendix A we prove that a DPSK constellation can be Gray labeled if and only if $N = 2$. The maximum number of pairs of nearest neighbors is Gray labeled when *quasi–Gray labeling* (see Appendix A) based on the Euclidean distance of the signal points is employed. Though quasi–Gray labeling is the best possible solution, the number of exceptions to Gray labeling with respect to noncoherent distance is given by

$$\begin{cases} \frac{1}{2} M^{N-1} \,, & \text{if} \quad M = 2 \\ M^{N-1} \,, & \text{if} \quad M > 2 \end{cases} \,, \tag{18}$$

i.e., increases exponentially with $N$.

For DAPSK numerical evaluations show that for the usual values of *ring ratios* $r_{i+1}/r_i$, $i = 0, \ldots, \alpha - 2$, both phase and amplitude transitions must be considered in order to identify the nearest neighbors. A similar and independent behavior for nearest neighbors in amplitude and phase can be observed. Since the subset of nearest neighbors corresponding only to phase transitions coincides with the set of nearest neighbors of a DPSK constellation, we conclude that no Gray labeling is possible when $N > 2$. We conjecture that the best practical solution is to separate the mapping and apply (quasi–) Gray labeling for the amplitude bits and quasi–Gray labeling for the phase bits.

# 4 Results

## 4.1 Capacity and Rate Design

In this section, the code design rules and the associated channel capacity for transmission over a frequency–selective channel when applying OFDM and differential encoding across the subchannels are presented. For it, (7), (9), and (13) using the pdfs from (4), (10), and (15), respectively, are numerically evaluated. In all examples for MLC with MSD Ungerböck labeling (UL) is employed, whereas BICM implies the use of (quasi–) Gray labeling (GL).

### 4.1.1 Differential Phase Modulation

First, we consider 8–ary differential PSK (D8PSK). In Figure 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 neglecting the cyclic prefix, $N_0$: one–sided noise power spectral density). For the solid lines, MLC and $N = 3$ are assumed. 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. Following the capacity rule, an example for the rate design of the component codes is given (dashed line). The target rate 1.5 bit/(channel use) divides optimally into the individual rates $R^i = C^i$, with $C^0/\cdots/C^5 = 0.15/0.18/0.51/0.58/0.76/0.82$. To compare MLC and BICM, the capacity of BICM ($N = 3$) is also displayed in Figure 3 (dash–dotted line). Apparently, there is a loss of about 1 dB in capacity compared to MLC.

A similar procedure for $N = 2$ (without figure) results in the rate design $R^i = C^i$ with $C^0/C^1/C^2 = 0.17/0.53/0.80$ for the three levels of MLC. In BICM, for the same transmission rate, always a single code with rate 1/2 has to be used.

The influence of the block size $N$ of MSDD, for which the equivalent flat Rayleigh fading channel is assumed to be constant, on the normalized capacity $C(N)$ is illustrated in Figure 4. Additionally, as references, the capacity $C_{\mathrm{CSI}}$ for coherent reception with perfect CSI and the bound for BICM capacity $C^u_{\mathrm{BICM}}$ are shown. As expected (cf. Section 2.3), with enlarging $N$ the capacity $C(N)$ increases. But the convergence of $C(N)$ to $C_{\mathrm{CSI}}$ as $N$ goes to infinity is rather slow.

Remarkably, the capacity of BICM for $N = 5$ is almost the same as for $N = 2$. This is caused by the fact, that for $N > 2$ no real Gray labeling exists. Furthermore, the values

of $C^u_{\mathrm{BICM}}$ indicate that the improvement possible with MSDD compared to conventional differential detection is rather limited, cf. [35]. Thus, for a given $N > 2$, we expect differentially encoded transmission using BICM with quasi–Gray labeling to be inferior to MLC.

### 4.1.2 Differential Amplitude/Phase Modulation

Second, we consider differential transmission schemes which modulate phase and amplitude. We restrict ourselves to D2A8PSK, consisting of two concentric rings with 8 signal points each and a ring ratio $r_1/r_0 = 2.0$, which was found to be advantageous for the Rayleigh fading channel in [39]. For $N = 3$, in Figure 5 the overall capacity $C$ and the capacities of the equivalent channels are sketched over $\bar{E}_{\mathrm{s}}/N_0$. An example for the optimal rate design of the component codes of MLC is given (dashed line). For the target rate 2.5 bit/(channel use), the rates of the eight equivalent channels $R^i = C^i$ are given by $C^0/\cdots/C^7 = 0.36/0.44/0.70/0.75/0.87/0.89/0.45/0.54$ (level 6 and 7: amplitude transitions). The capacity curve of BICM ($N = 3$), which is also plotted in Figure 5 (dash–dotted line), indicates a loss in power efficiency of about 1 dB compared to MLC.

For $N = 2$ (without figure) the rate design $R^i = C^i$ is obtained by $C^0/C^1/C^2/C^3 = 0.43/0.72/0.88/0.47$ (level 3: amplitude transition). In BICM, in each case, a single rate–5/8 code has to be applied.

The effect of an increased observation interval $N$ on the capacity is shown in Figure 6. Like for D8PSK, the convergence of the normalized capacity of MLC to the capacity $C_{\mathrm{CSI}}$ can be observed for increasing $N$. Again, for BICM the gains are negligible or at least very small, and the improvement over $N = 2$ is bounded to about 1 dB.

## 4.2 Simulation Results

As already noted in Section 2, we do not focus on a particular OFDM system, but aim at a general discussion on coded modulation for noncoherent OFDM. Therefore, system parameters are only specified qualitatively. We assume the coherence time of the frequency–selective channel to be of the order of at least one OFDM symbol. The channel coherence bandwidth and the spacing of OFDM subchannels are such that the channel state of the equivalent frequency non–selective fading channel can be expected to be constant over a block of $N$ consecutive symbols. E.g., regarding typical digital broadcasting systems for mobile receivers (cf. [21, 40, 41]), these conditions are satisfied. Furthermore, by means of interleaving and further diversity techniques [21, 26], the decoder input is sufficiently

decorrelated. For these reasons, in the simulations the overall OFDM transmission system is replaced by simply regarding the transmission over the equivalent vector channel.

In order to assess the performance of the proposed coded modulation schemes, numerical simulations using 8PSK and 2A8PSK constellations are performed. MLC in combination with Ungerböck's (natural) labeling and MSD is compared to BICM using (quasi–)Gray labeling. In each case, turbo codes [18] (parallel concatenated convolutional codes with 16 states each) are employed as component codes. Rate is adjusted by symmetric puncturing of parity symbols, cf. [42]. The systematical decoder concept according to [42] is used. For decoding of component codes, maximum a posteriori symbol–by–symbol estimation according to [43] is applied. The noncoherent decoding metrics are based on (10) and (15), respectively, and (12). 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 perform 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.

### 4.2.1 Differential Phase Modulation

First, we consider D8PSK over Rayleigh fading channel. We fix the target rate to 1.5 bit/symbol, which provides the required redundancy per complex dimension for Rayleigh fading, cf. [44].

In Fig. 7, D8PSK using MLC and BICM are compared for a fixed channel delay of 6000 transmitted symbols. For $N = 2$ the corresponding lengths of the three binary codes are 6000 for MLC and 18000 for the single code of BICM, respectively. With $N = 3$ the lengths of the six binary codes are 3000 for MLC and still 18000 for BICM, respectively. The rates of the components turbo codes are designed according to capacity of the individual levels (cf. Sec. 4.1.1).

As reference, using the numerical results from Section 4.1, the capacity limits taking the finite error rate into account ("rate–distortion capacities")[3] for MLC and BICM are shown. In the case of $N = 2$ and for low $\bar{E}_\mathrm{b}/N_0$ (bit error rates above BER $= 2 \cdot 10^{-4}$; $\bar{E}_\mathrm{b}$: average received energy per information bit neglecting the cyclic prefix) MLC outperforms

[3]Let $U_i$, $\hat{U}_i$, $1 \le i \le k$, be the binary source symbols and estimated symbols, respectively, with bit–error rate BER. We consider an $(n, k)$ binary block code of rate $R = k/n$ and a channel with capacity $C$. Then, regarding data processing theorem, the expansion of mutual information, and Fano's inequality, we have: $n \cdot C \ge k \cdot I(U; \hat{U}) = k \cdot (1 - H(U|\hat{U})) \ge k \cdot (1 - H_2(\mathrm{BER}))$, where $H$ denotes entropy and $H_2(p) \triangleq -p \log_2(p) - (1-p) \log_2(1-p)$ is the binary entropy function. Hence, $C \ge R \cdot (1 - H_2(\mathrm{BER}))$. Substituting inequality with equality gives the desired relationship between $C$ and BER.

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. 4. 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 in data delay is almost ineffective, because of the already very large code length.

### 4.2.2 Differential Amplitude/Phase Modulation

Next, D2A8PSK over Rayleigh fading channel is discussed. The ring ratio is chosen equal to 2.0 and a target rate of 2.5 bits/symbol is desired. For labeling, the most significant bit is assigned to the amplitude transition. The results in Fig. 8 refer to a channel delay equal to 6000 transmit symbols.

In the case of $N = 2$ the code length of the binary (component) codes are 6000 for MLC and 24000 for BICM, respectively. For $N = 3$ the lengths of the binary codes are 3000 for MLC and 24000 for BICM, respectively. Again, in MLC rate design according to capacity is applied (cf. Sec. 4.1.2).

Only for $N = 2$ BICM is superior compared to MLC. In the case $N = 3$, despite the much shorter block length, MLC clearly outperforms BICM. Again, this is due to BICM relying on Gray labeling, which does not exist for D(A)PSK and $N > 2$. In some applications the shorter block length in the MLC scheme are advantageous with respect to complexity.

To summarize, the simulation results appear to be in great accordance with the results predicted from information theory shown in Section 4.1. As already accepted for coherent communication, the capacity rule again turns out to be an appropriate criterion for the choice of the component codes in noncoherent MLC schemes.

In combination with OFDM using differential encoding, both MLC and BICM prove to be very effective methods to transfer the power efficiency of turbo codes into bandwidth–efficient noncoherent transmission schemes over frequency–selective channels. For moderate transmission delays and BER's around $10^{-4}$ an SNR gap between 1.4 and 1.8 dB to capacity limit results. For fixed, short–to–medium, transmission delays, when $N = 2$, BICM is slightly superior to MLC, because it employs binary codes of larger lengths. In

turn, MLC is indispensable to exploit the potential of multiple–symbol differential detection. For $N > 2$ BICM seems to be inappropriate because it is strongly tied to Gray labeling, which does not exist for D(A)PSK and $N > 2$.

# 5 Conclusions

In this paper, the application of coded modulation for OFDM transmission over slowly time–varying, frequency–selective channels is considered. Assuming OFDM combined with diversity and interleaving, the actual transmission channel is transformed into a stationary, slowly time–varying, flat Rayleigh fading channel. In order not to require channel information at the receiver side, differential encoding across OFDM subchannels and multiple–symbol differential detection are performed. Block interleaving is used as appropriate means to combat the effects of faded tones.

Multilevel coding and bit–interleaved coded modulation are applied as potential coded modulation techniques for the current situation. For the case of MLC and MSD the capacity rule, already well established for coherent transmission over the AWGN channel [15], is again verified. Results obtained with a proper design of the component code rates for 8PSK and 2A8PSK constellations are presented. By these MLC/MSD schemes the associated channel capacities can be approached.

Unfortunately, the simple BICM undergoes a significant shortcoming compared to multilevel coding when the time–variance of the fading channel allows multiple–symbol differential detection with $N > 2$, because no Gray labeling exists. For such situations, more complex iterative decoding of BICM with MSDD as proposed in e.g. [45, 46], cf. also [47], or other iterative decoding structures which allow for interleaving, e.g. [48, 49, 50, 51], might be advantageous. There, regarding error correction coding and modulation as a serially concatenated coding scheme, BICM does not require Gray labeling.

The theoretical results derived from information theory are confirmed by means of numerical simulations for D8PSK and D2A8PSK.

We believe that the use of MLC and MSDD in combination with OFDM and differential encoding across OFDM subchannels is an attractive method for bandwidth–efficient transmission over frequency–selective channels.

## Acknowledgment

The authors wish to thank S. Calabrò for many valuable discussions during his visit at the University of Erlangen, especially with respect to the proof of Appendix A. The authors also gratefully acknowledge the anonymous reviewers for their help in improving the paper.

# A Proof of Theorem 1

We show that for DPSK and MSDD Gray labeling exists if and only if $N = 2$. For it, it is convenient to introduce the equivalent data vector $\boldsymbol{a}' = [a'_1, \ldots, a'_{N-1}]$ with components obtained from the original data vector $\boldsymbol{a} = [a_1, \ldots, a_{N-1}]$ by

$$a'_i = \prod_{k=1}^{i} a_k \ . \tag{19}$$

Clearly, there is a bijective relation between $\boldsymbol{a}$ and $\boldsymbol{a}'$ and its components stem from the same PSK constellation[4]. Therefore, its sufficient to consider $\boldsymbol{a}'$ and the noncoherent distance

$$d^2_{\mathrm{NC}}(\tilde{\boldsymbol{a}}', \boldsymbol{a}') = N r_0^2 - \left| r_0^2 + \tilde{\boldsymbol{a}}' \boldsymbol{a}'^H \right| \ . \tag{20}$$

Furthermore, since $d^2_{\mathrm{NC}}(\tilde{\boldsymbol{a}}', \boldsymbol{a}')$ is invariant under componentwise simultaneous rotations of the vectors $\boldsymbol{a}'$ and $\tilde{\boldsymbol{a}}'$, we can restrict to the vector $\boldsymbol{a}' = r_0[\underbrace{1, 1, \ldots, 1}_{N-1}]$ as reference.

It is easy to verify that the list of the nearest neighbors is constituted by (i) all the permutations of the components of $r_0[\mathrm{e}^{\mathrm{j}\frac{2\pi}{M}}, 1, \ldots, 1]$ and (ii) the point $r_0[\mathrm{e}^{\mathrm{j}\frac{2\pi}{M}}, \mathrm{e}^{\mathrm{j}\frac{2\pi}{M}}, \ldots, \mathrm{e}^{\mathrm{j}\frac{2\pi}{M}}]$, together with their complex conjugates if $M > 2$. The points of the first kind are all the nearest neighbors with respect to the Euclidean distance, while the further points (ii) arise from differential encoding and are present only when $N > 2$. Hence, if $N = 2$ the usual Euclidean–distance Gray labeling is also a differential Gray labeling.

If $N > 2$ let us define the differential vectors $\boldsymbol{a}'_{j,i}$ with $j = 1, 2, \ldots, N-1$ and $i = 0, 1, \ldots, N-j-1$, such that all the components of $\boldsymbol{a}'_{j,i}$ are equal to $r_0$, with the exception of $j$ subsequent components, starting from index $i$ and ending at index $i+j-1$. All these components are set to $r_0 \mathrm{e}^{\mathrm{j}\frac{2\pi}{M}}$. Obviously, this definition implies $\boldsymbol{a}'_{0,0}$ be $r_0[1, 1, \ldots, 1]$. It is clear that $\boldsymbol{a}'_{j,i}$, $j = 1, \ldots, N-1$, $i = 0, \ldots, N-j-1$, is a nearest neighbor for $\boldsymbol{a}'_{j-1,i}$ and $\boldsymbol{a}'_{j-1,i+1}$ with respect both to the Euclidean and the noncoherent distance. Moreover, $\boldsymbol{a}'_{N-1,0}$ is also a nearest neighbor for $\boldsymbol{a}'_{0,0}$ according to the noncoherent distance, only.

These relations can be graphically displayed: Each differential vector $\boldsymbol{a}'$ will be represented as a point and each pair of points will be connected with a branch if and only if the corresponding vectors are nearest neighbors. Such a graph will be called adjacency diagram. In Fig. 9 we illustrate the corresponding adjacency diagram for the case of $N = 5$.

Let us suppose, without any restriction, that $\boldsymbol{a}'_{0,0}$ corresponds to the all–zero label and that the binary label of $\boldsymbol{a}'_{1,i}$ has a unique 1 in position $i$. In order to fulfill the requirements

[4] If binary information is directly mapped onto $\boldsymbol{a}'$, instead of usual "accumulated" differential encoding, "parallel" encoding is performed.

of Gray labeling the label of each point $\boldsymbol{a}'_{j,i}$, $j = 2, \ldots, N-1$, $i = 0, \ldots, N-j-1$, is constrained by the labels of $\boldsymbol{a}'_{j-1,i}$ and $\boldsymbol{a}'_{j-1,i+1}$ to have only $j$ bits 1 in the positions from $i$ to $i+j-1$. Specifically, the label of $\boldsymbol{a}'_{N-1,0}$ must have $N-1$ bits 1. From a comparison with the label of its nearest neighbor $\boldsymbol{a}'_{0,0}$ it is clear that a Gray labeling is not obtained.

As a consequence we have to renounce to Gray–label at least one pair of nearest neighbors in this fragment of the adjacency diagram; the most obvious (but not the only possible) choice is to exclude the pair $(\boldsymbol{a}'_{0,0}, \boldsymbol{a}'_{N-1,0})$. If we repeat such a consideration for any possible choice of the point $\boldsymbol{a}'_{0,0}$ and we neglect all the pairs of vectors which are nearest neighbors only with respect to the noncoherent distance, we can maximize the number of Gray–labeled pairs employing the usual Gray labeling based solely on the Euclidean distance. We denote this strategy as *quasi–Gray labeling*.

# References


[1] S.B. Weinstein and P.M. Ebert. Data Transmission by Frequency-Division Multiplexing Using the Discrete Fourier Transform. *IEEE Trans. on Communication Technology*, 19(5):628–634, October 1971.

[2] J.A.C. Bingham. Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come. *IEEE Communications Magazin*, pages 5–14, May 1990.

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

[4] 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 Communications*, 42(1):63–74, January 1994.

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

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

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

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

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

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

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

[12] A. Svensson. On Differentially Encoded Star 16QAM with Differential Detection and Diversity. *IEEE Trans. on Vehicular Technology*, 44:586–593, Aug. 1995.

[13] F. Adachi. Error Rate Analysis of Differentially Encoded and Detected 16APSK Under Rician Fading. *IEEE Trans. on Vehicular Technology*, 45(1):1–11, February 1996.

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

[15] U. Wachsmann, R.F.H. Fischer, and J.B. Huber. Multilevel Codes: Theoretical Concepts and Practical Design Rules. *IEEE Trans. on Information Theory*, 45(5):1361–1391, July 1999.

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

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

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

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

[20] B. Sklar. Rayleigh Fading Channels in Mobile Digital Communication Systems—Part I: Characterization. *IEEE Communications Magazin*, 35(7):90–101, July 1997.

[21] M. Alard and R. Lasalle. Principles of Modulation and Channel Coding for Digital Broadcasting for Mobile Receivers. *EBU Review*, pages 47–69, August 1987.

[22] O. Edfors, M. Sandell, J.-J. van de Beek, S.J. Wilson, and P.O. Börjesson. OFDM Channel Estimation by Singular Value Decomposition. *IEEE Trans. on Communications*, 46(7):931–939, July 1998.

[23] P.K. Frenger and N.A.B. Svensson. Decision–Directed Coherent Detection in Multicarrier Systems on Rayleigh Fading Channels. *IEEE Trans. on Vehicular Technology*, 48(2):490–498, March 1999.

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

[25] W.T. Webb, L. Hanzo, and R. Steele. Bandwidth Efficient QAM Schemes for Rayleigh Fading Channels. *IEE Proceedings–I*, 138(3):169–175, June 1991.

[26] J. Chuang and N. Sollenberger. Beyond 3G: Wideband Wireless Data Access Based on OFDM and Dynamic Packet Assignment. *IEEE Communications Magazin*, 38:78–87, July 2000.

[27] K. Miller. *Complex Stochastic Processes.* Addison–Wesley, New York, 1974.

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

[29] F.-W. Sun and H.C.A. van Tilborg. Approaching Capacity by Equiprobable Signaling on the Gaussian Channel. *IEEE Trans. on Information Theory*, 39:1714–1716, September 1993.

[30] W. Betts, A.R. Calderbank, and R. Laroia. Performance of Nonuniform Constellations on the Gaussian Channel. *IEEE Trans. on Information Theory*, 40:1633–1638, September 1994.

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

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

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

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

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

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

[37] Y. Kofman, E. Zehavi, and S. Shamai (Shitz). *nd*–Convolutional Codes—Part I: Performance Analysis, Part II: Structural Analysis. *IEEE Trans. on Information Theory*, 43(2):558–589, March 1997.

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

[39] L.H.-J. Lampe and R.F.H. Fischer. Comparison and Optimization of Differentially Encoded Transmission on Fading Channels. In *3rd Int. Symp. on Power-Line Communications (ISPLC'99)*, pages 107–113, Lancaster, UK, March 1999.

[40] European Telecommunications Standards Institute (ETSI). Radio broadcasting systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers. ETS 300 401, May 1997.

[41] European Telecommunications Standards Institute (ETSI). Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television. ETSI EN 300 744, July 1999.

[42] U. Wachsmann and J.B. Huber. Power and Bandwidth Efficient Digital Communication Using Turbo Codes in Multilevel Codes. *European Trans. Telecommunications (ETT)*, 6:557–567, Sept.-Oct. 1995.

[43] L.R. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate. *IEEE Trans. on Information Theory*, 20(3):284–287, March 1974.

[44] U. Hansson and T. Aulin. Channel Symbol Expansion Diversity — Improved Coded Modulation for the Rayleigh Fading Channel. In *Proc. IEEE Int. Conf. Communications (ICC)*, pages 891–895, Dallas, August 1996.

[45] 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.

[46] 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. Telecommunications (ICT)*, Porto Carras, Greece, June 1998.

[47] X. Li and J.A. Ritcey. Trellis–Coded Modulation with Bit Interleaving and Iterative Decoding. *IEEE Journal on Selected Areas in Communications*, 17(4):715–724, April 1999.

[48] I. D. Marsland and P. T. Mathiopoulos. Iterative Noncoherent Detection of Convolutionally Encoded Signals. In *Proc. IEEE Int. Conf. Telecommunications (ICT)*, Porto Carras, Greece, June 1998.

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

[50] G. Colavolpe, R. Raheli, and G. Ferrari. A Noncoherent Soft-Output Decoding Algorithm for Coded Linear Modulations. In *Proc. IEEE Int. Conf. Communications (ICC)*, pages 843–847, Vancouver, June 1999.

[51] K.R. Narayanan and G.L. Stüber. A Serial Concatenation Approach to Iterative Demodulation and Decoding. *IEEE Trans. on Communications*, 47(7):956–961, July 1999.

# List of Figures

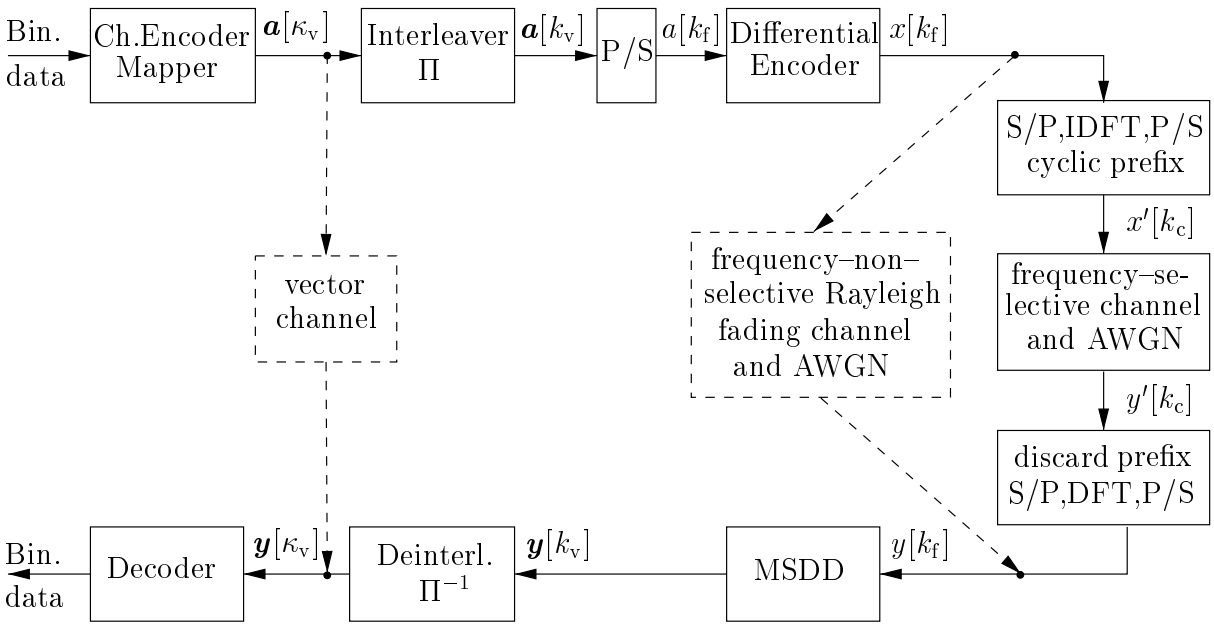


**Figure 1:** *Discrete–time system model for OFDM transmission with differential encoding and the resulting equivalent fading channel and vector channel.*

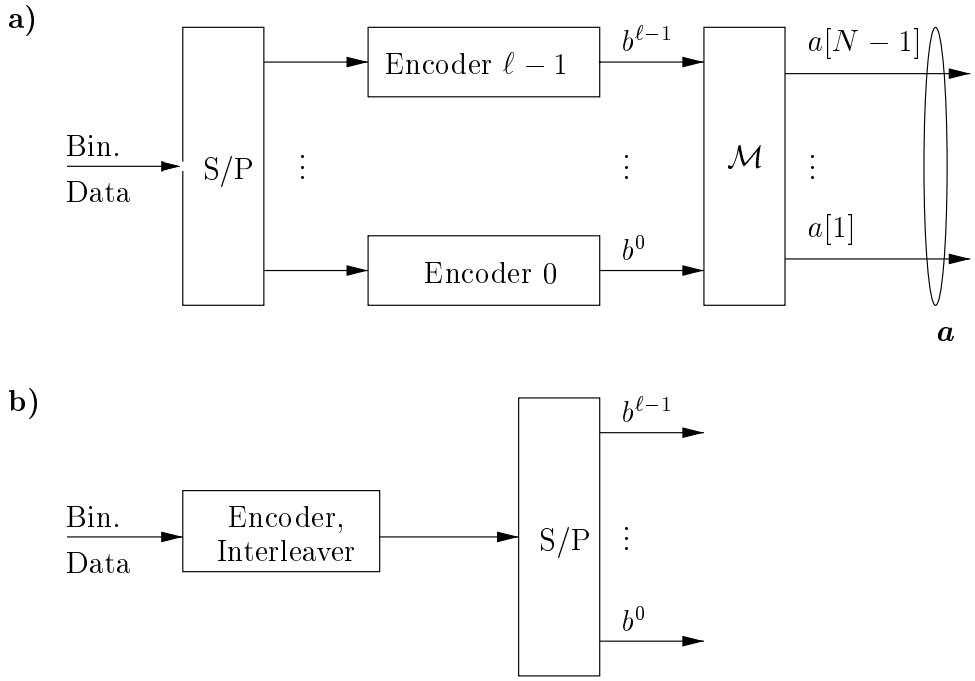


**Figure 2:** *Channel encoding and Mapping: a) MLC, b) Changes for BICM.*

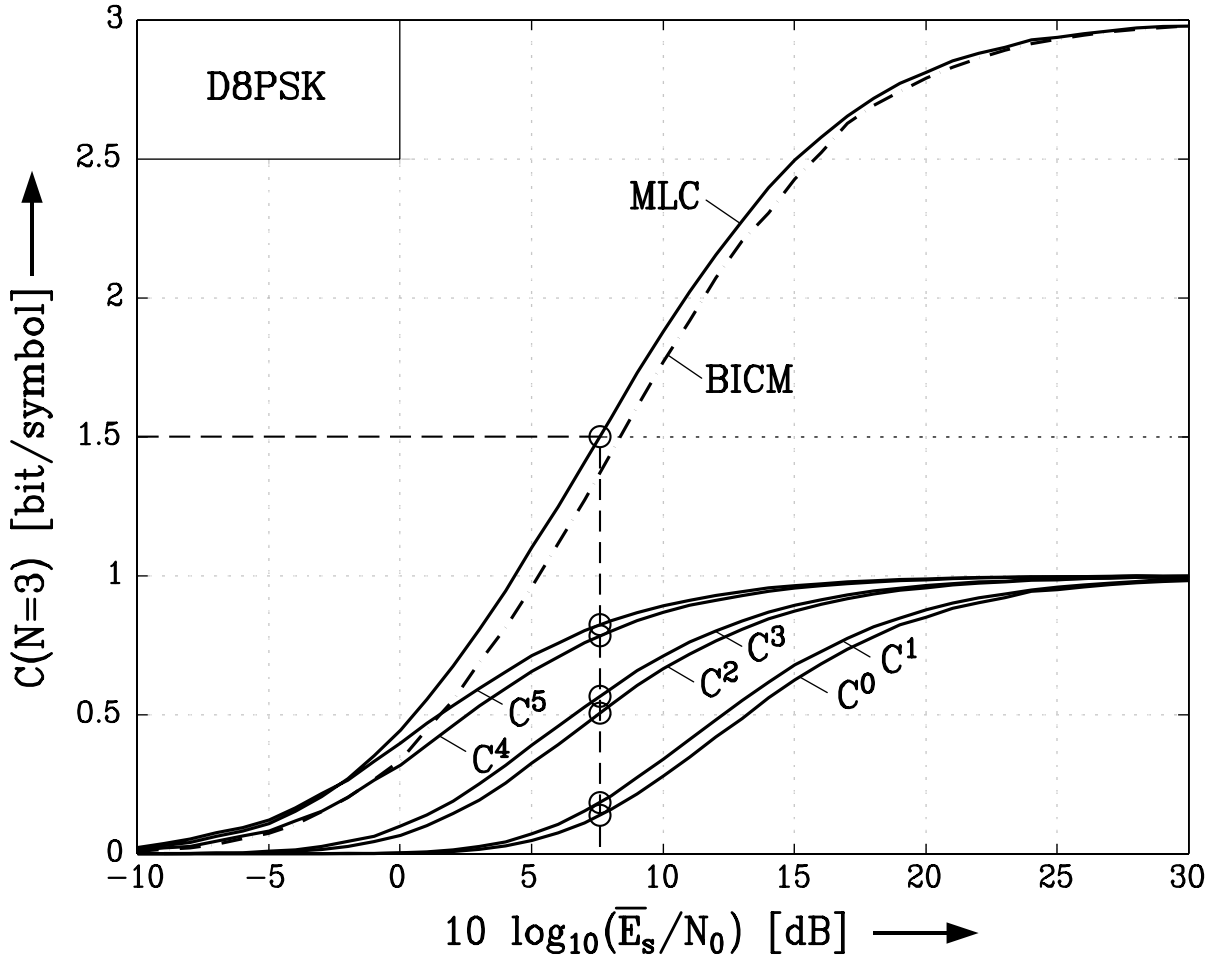


**Figure 3:** *Capacity $C(N = 3)$ and capacities $C^0, \ldots, C^5$ of the equivalent channels for D8PSK. Solid lines: MLC with UL. Dash–dotted lines: BICM with GL. Rayleigh fading channel. Dashed lines: Rate design for $C = 1.5$ bit/symbol.*

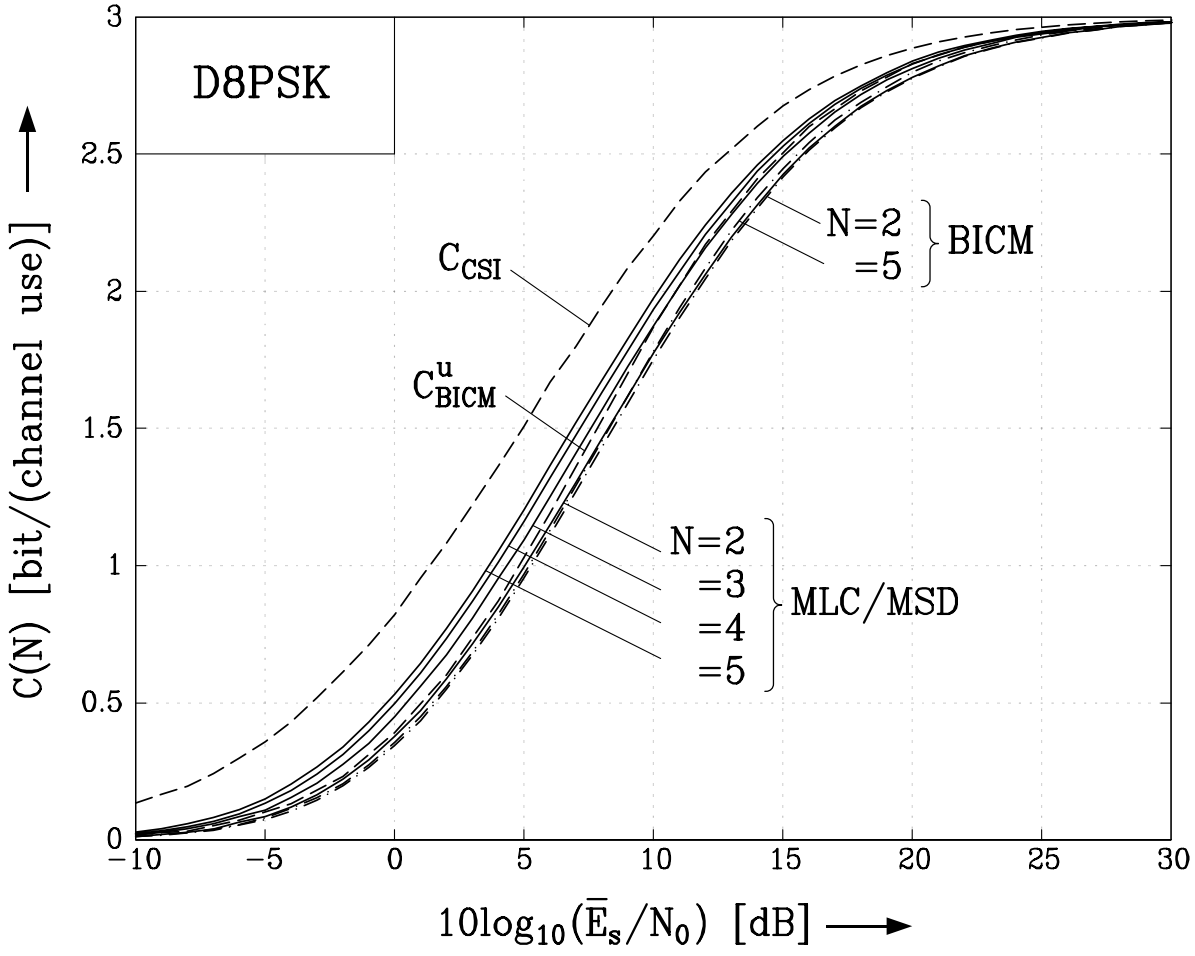


**Figure 4:** *Normalized capacity for D8PSK. Solid lines: MLC with UL, $N = 2, 3, 4, 5$ (from right to left). Dash–dotted lines: BICM with GL, $N = 2, 5$ (from right to left). Rayleigh fading channel. Dashed Lines: Capacity $C_{\mathrm{CSI}}$ and bound $C^{u}_{\mathrm{BICM}}$.*

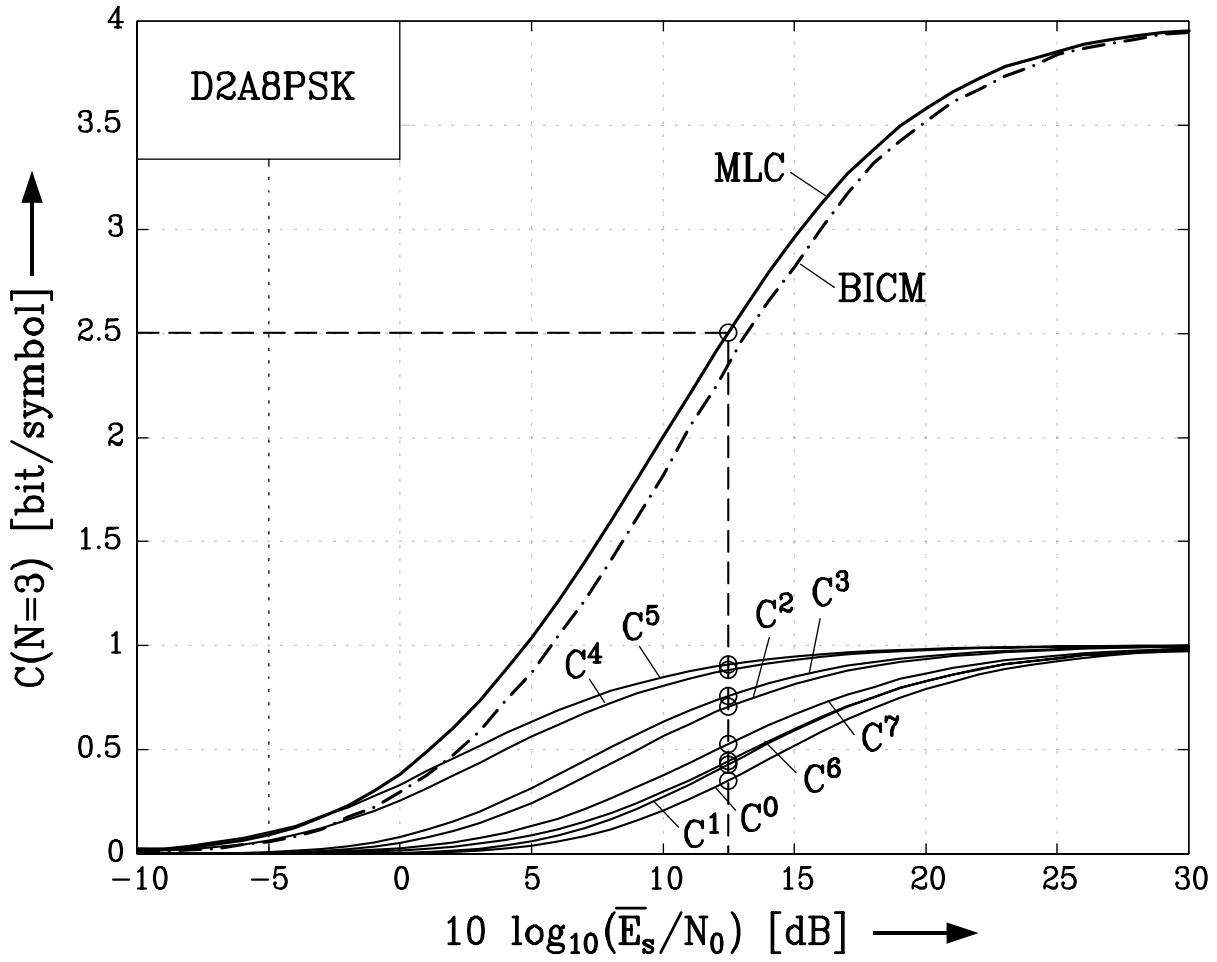


**Figure 5:** *Capacity $C(N = 3)$ and capacities $C^0, \ldots, C^7$ of the equivalent channels for D2A8PSK with ring ratio $r_1/r_0 = 2.0$. Solid lines: MLC with UL. Dash–dotted lines: BICM with GL. Rayleigh fading channel. Dashed lines: Rate design for $C = 2.5$ bit/symbol.*

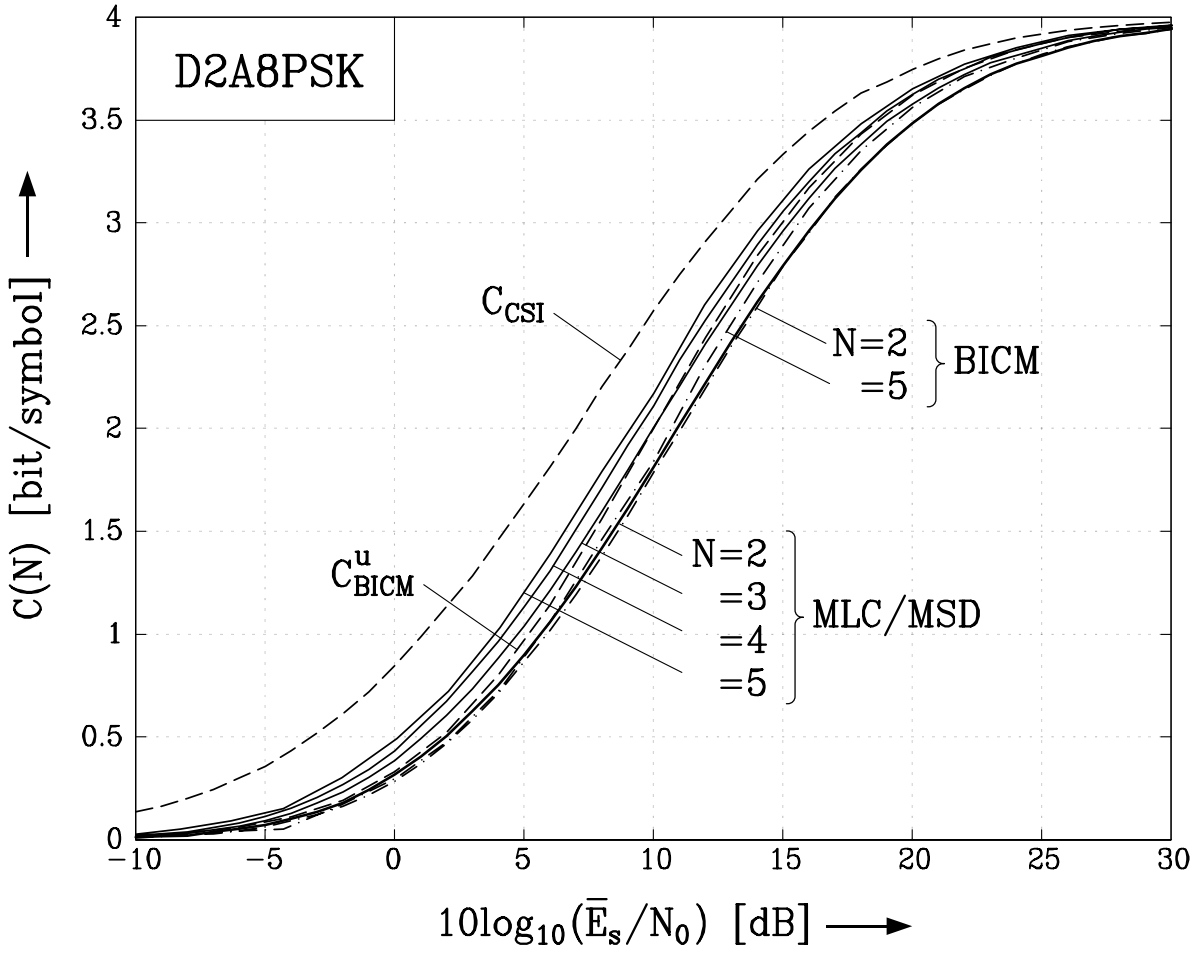


**Figure 6:** *Normalized capacity for D2A8PSK with ring ratio $r_1/r_0 = 2.0$. Solid lines: MLC with UL, $N = 2, 3, 4, 5$ (from right to left). Dash–dotted lines: BICM with GL, $N = 2, 5$ (from right to left). Rayleigh fading channel. Dashed Lines: Capacity $C_{\mathrm{CSI}}$ and bound $C^{u}_{\mathrm{BICM}}$.*

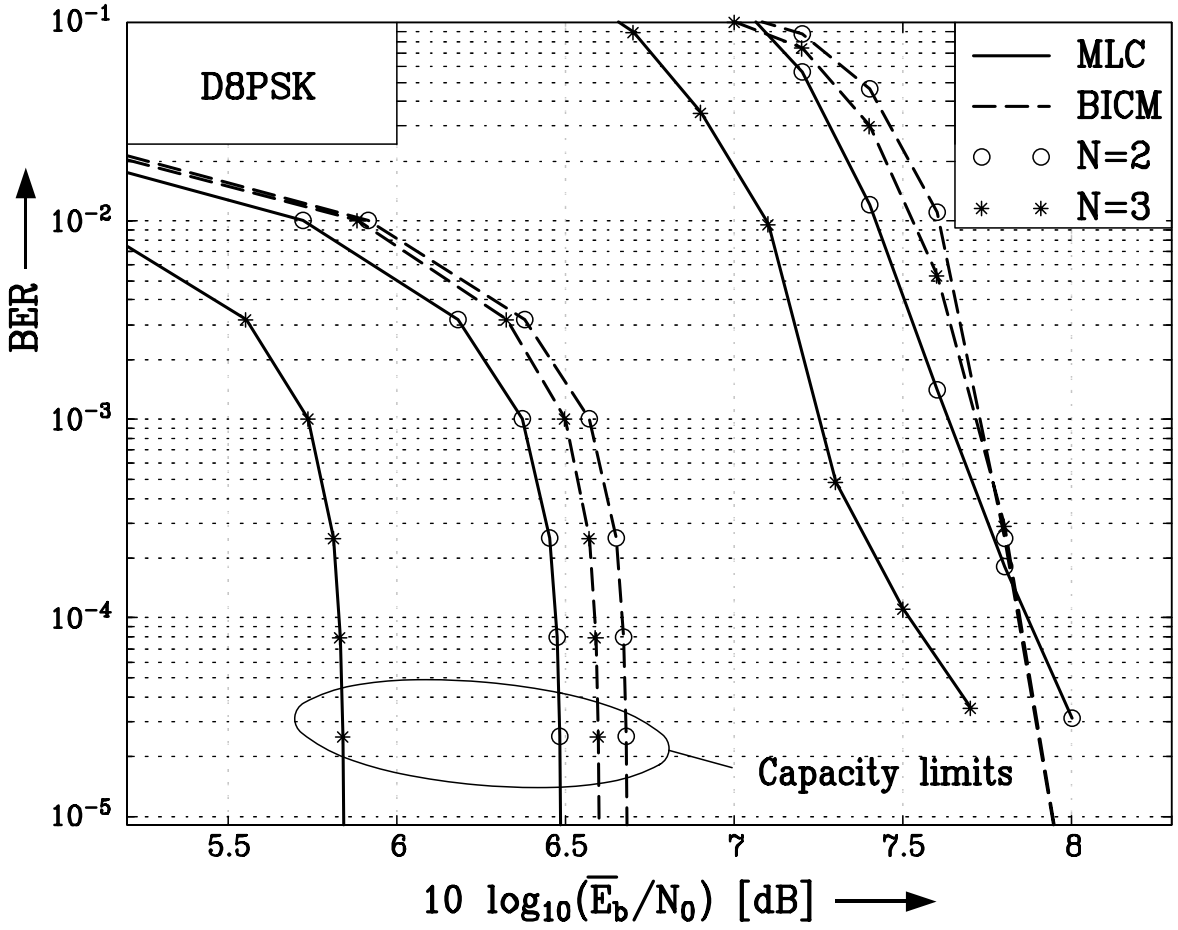


**Figure 7:** *BER versus $\bar{E}_b/N_0$ (in dB) for D8PSK with MLC (solid lines) and BICM (dashed lines). Rayleigh fading channel. Circles: $N = 2$, Stars: $N = 3$. Left hand side: respective rate–distortion capacity limits.*

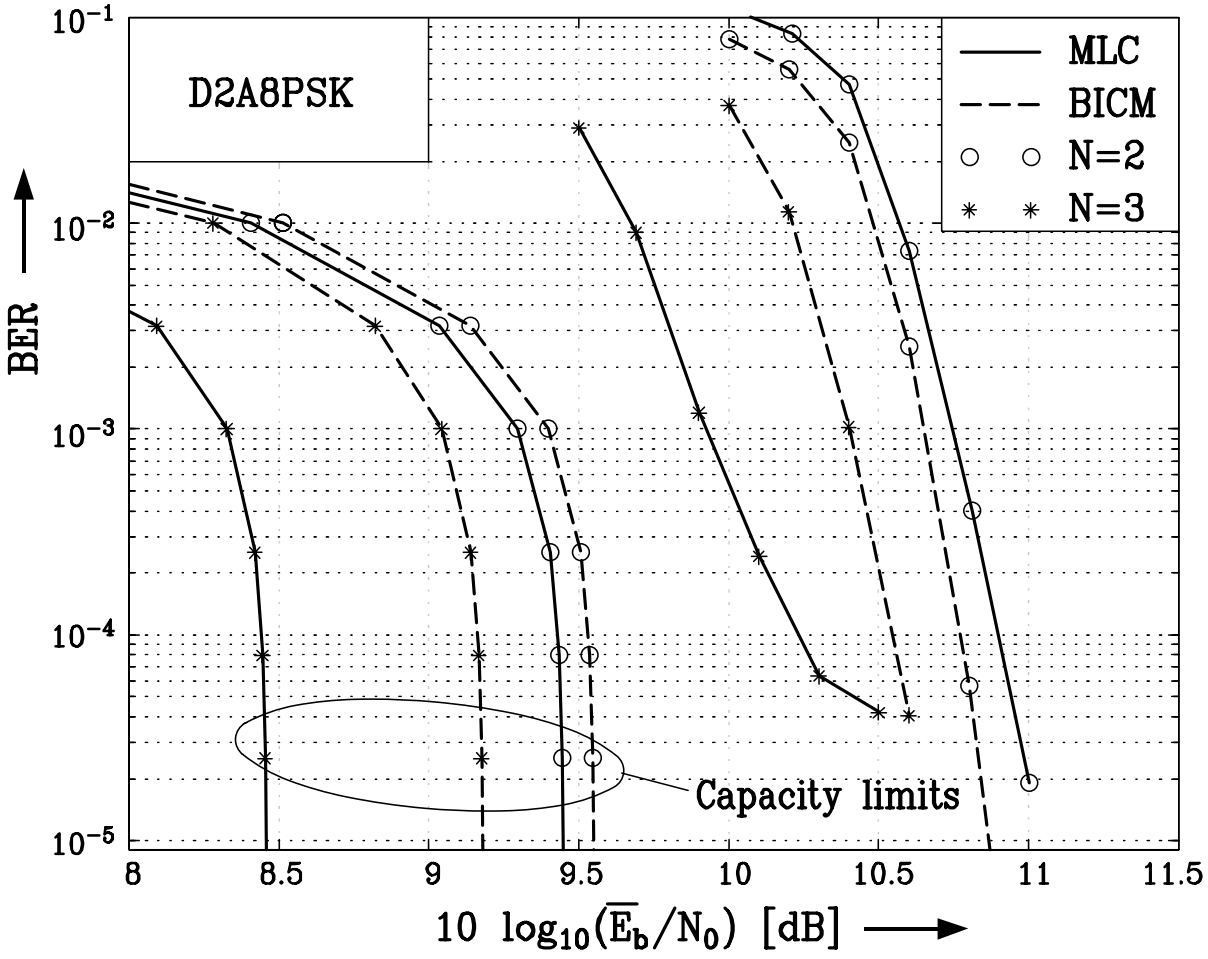


**Figure 8:** *BER versus $\bar{E}_\mathrm{b}/N_0$ (in dB) for D2A8PSK ($r_1/r_0 = 2.0$) with MLC (solid lines) and BICM (dashed lines). Rayleigh fading channel. Circles: $N = 2$, Stars: $N = 3$. Left hand side: respective rate–distortion capacity limits.*

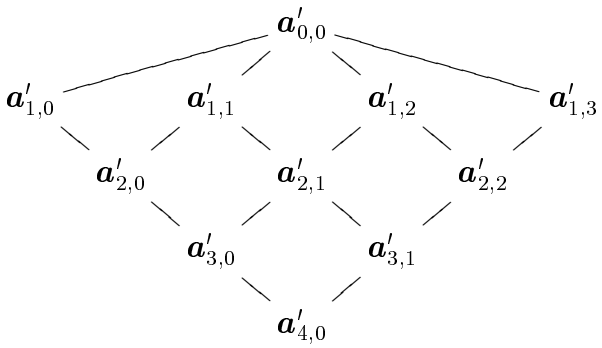


**Figure 9:** *Portion of an adjacency diagram.*