Iterative Decision–Feedback Differential Demodulation of Bit–Interleaved Coded MDPSK for Flat Rayleigh Fading Channels

Lutz H.-J. Lampe and Robert Schober

Abstract — **A simple receiver structure recently proposed by the authors [1] for convolutional coded M–ary differential phase shift keying (MDPSK) transmission over flat Rayleigh fading channels without channel state information is analyzed in detail. We present a thorough discussion of the iterative decoding procedure, which is referred to as iterative decision–feedback differential demodulation (iterative DF–DM) [1]. The convergence behavior of iterative DF–DM is theoretically examined. The analysis supports the observation that the iterative decoding scheme works well for target bit error rates which are usually of interest. Furthermore, the associated cutoff rate for error–free decision feedback is studied. Judging from this performance parameter remarkable gains in power efficiency compared to conventional differential demodulation are indicated, while the computational complexity of the decoding remains low. The results from information theory are in good agreement with the given simulation results.**

Index terms: M–ary DPSK, Rayleigh fading channels, decision–feedback, noncoherent detection, bit–interleaved coded modulation, mobile communications

1 Introduction

Digital receivers performing noncoherent detection are very attractive because of their robustness against ambiguities and impairments of the phase of the received signal. In particular, reliable estimation of the carrier phase and, in the case of transmission over fading channels, the current channel state is often not practicable.

Usually, for noncoherent transmission, the underlying channel can be assumed to be slowly time–variant or even time–invariant over at least two consecutive symbols. Thus, there are dependences among the received symbols. As well known, this memory should be utilized in the receiver processing to improve performance [2].

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. Thus, the memory is partially taken into account. Generally, the larger N is chosen the more complex the receiver becomes. For the additive white Gaussian noise (AWGN) channel with constant (unknown) phase it has been shown that for sufficiently large N the performance of coherent transmission is approached [3, 4, 5, 6].

The classical solution to avoid the problem of phase ambiguity is M–ary differentially encoded phase shift keying (MDPSK) modulation at the transmitter. MDPSK and demodulation where decision variables on $N-1$ information symbols are provided based on the independent evaluation of blocks of $N \geq 2$ consecutive received symbols, which overlap by one symbol, are investigated in [3, 7, 8, 4]. In a more general framework, noncoherently detectable convolutional codes, which include differential encoding, are introduced in [5, 9]. Trellis–based noncoherent detection schemes are developed in [10, 11, 12]. There, when channel coding is applied, decoding is done in an augmented code trellis and interleaving is not employed. However, in a slow fading environment, it is desirable to combine error correction coding with interleaving to mitigate the effects of fading [13].

If differential encoding or some other modulation code, cf. e.g. [14, 15], succeeds the channel encoder, iterative decoding schemes regarding the modulation as inner component code of a serially concatenated code have been proposed in [14, 16, 17, 18, 19, 20]. Soft decisions of the outer (error correcting code) decoder are fed back to the inner decoder to improve the delivered decision values. Here, the decoding complexity of the inner code is determined by the considered channel memory length. In particular, computational complexity increases exponentially with N and linearly with the number of iterations.

In the case of differentially encoded transmission without error correction coding

decision–feedback differential detection (DF–DD), i.e., the feedback of previously decided transmitted symbols, has been proved to offer good performance at a very low computational complexity for both AWGN and fading channels [21, 22, 23, 24, 25].

A similar, but non–iterative reduced complexity decoding scheme using tentative decisions in the Viterbi algorithm [26] yields good results in terms of bit error rate for trellis coded noncoherent transmission over the AWGN channel, cf. e.g. [27, 12]. The number of states of the augmented trellis, which represents both the code memory and an implicit phase memory, is reduced by per–survivor processing [28]. Unfortunately, this scheme does not include interleaving, which is in fact not necessary for the AWGN channel, but very important for slow fading channels.

In this paper, we motivate and discuss extensively a simple–to–implement and low–complexity noncoherent receiver for convolutional coded MDPSK recently proposed by the authors in [1], which can also be employed in existing transmission systems without any modification of the transmitter. As a natural consequence of the considerations above, the DF–DD technique is adopted to the case of error correction coding with interleaving. In particular, bit–interleaved coded modulation (BICM) [29, 30] with convolutional codes is applied to increase the code diversity. Since as in DF–DD feedback of *hard* decisions is considered, the decoding is performed through the standard Viterbi algorithm. Thus, in contrast to the other iterative decoding schemes mentioned above, the complex soft output decoding using for example the decoding algorithm of [31], is not required. Moreover, the soft input for the Viterbi decoder can be computed by evaluating very simple expressions and without *a priori* knowledge on the fading statistics. Although the incorporation of hard decisions in the iterative decoding procedure is clearly suboptimum, performance gains of several dB compared to conventional differential demodulation are achieved, while the increase in receiver complexity is moderate. It should be mentioned that an iterative decoding technique with hard–decision feedback has also been proposed in [32, 33]. There, regarding the mapping as a modulation code enables an improvement of the performance of BICM with iterative coherent demodulation.

This paper is organized as follows. In Section 2, the discrete–time system model is introduced. The optimum metrics for the Viterbi decoder are derived in Section 3. In Section 4, the low complex iterative decoding algorithm is described, and its convergence is investigated by regarding the metric computation as an estimation problem. A theoretical performance assessment based on the cutoff rate [34] is provided in Section 5. The results of the cutoff rate analysis are in great accordance with the simulation

results presented in Section 6. Finally, conclusions are drawn in Section 7.

2 System Model

The block diagram of the system model is depicted in Figure 1. Here, we apply the discrete–time channel model. The channel and all signals are represented in the equivalent low–pass domain, i.e., all quantities are complex–valued, cf. e.g. [35, Appendix]. A sequence of information bits enters the convolutional encoder. The encoder output symbols are interleaved yielding the sequence of coded bits $c[i]$ ($i \in \mathbb{Z}$: bit discrete–time index). Then, $\ell \triangleq \log_2(M)$ interleaved coded bits are mapped to data–carrying symbols $a[k]$ ($k \in \mathbb{Z}$: symbol discrete–time index) from an M–ary PSK constellation $\mathcal{A} \triangleq \{\mathrm{e}^{\mathrm{j}2\pi m/M} | m = 0, 1, \ldots, M-1\}$. We use Gray labeling of PSK symbols $a[k]$ with respect to noncoherent distance [36], which equals the usual Gray labeling with respect to the Euclidean distance [37]. Note that the appropriate labeling is essential for BICM [30]. To enable noncoherent demodulation, $a[k]$ is differentially encoded. That is, the channel input symbols $x[k]$ are obtained from

$$x[k] = a[k] \cdot x[k-1] \; . \tag{1}$$

We assume sufficiently slow fading, i.e., the channel does not change significantly during one symbol interval, 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

$$y[k] = g[k] \cdot x[k] + n[k] \; , \tag{2}$$

where the fading process $g[\cdot]$ and the noise process $n[\cdot]$ are mutually independent correlated and uncorrelated zero–mean complex Gaussian random processes with variances σ_g^2 and σ_n^2, respectively. According to the widely used Clarke fading model [38] (also known as Jakes model [39]) with maximum Doppler frequency f_d, the autocorrelation function of the fading process is given by

$$R_g[\kappa] \triangleq \mathcal{E}_g\{g^*[k]g[k+\kappa]\} = \sigma_g^2 \cdot \mathrm{J}_0(2\pi f_d T \kappa) \; , \tag{3}$$

where $\mathcal{E}_\times\{\cdot\}$ denotes expectation with respect to $\times$, $\mathrm{J}_0(\cdot)$ is the zeroth order Bessel function of the first kind, and T is the symbol interval.

At the receiver, bit branch metrics $\lambda[i]$ are computed as described in Section 3. The deinterleaved metrics are the soft input for the standard Viterbi decoder [26]. The *hard*

decisions of the Viterbi decoder are interleaved and fed back to the metric calculation (cf. Figure 1) which now makes use of the decisions $\hat{c}[i]$. This procedure is repeated in a number of iterations.

3 Metric Calculation

In this section, the metric calculation based on decision–feedback symbols is presented. For this, we assume a sufficient interleaving depth such that the channel between encoder output and decoder input is essentially memoryless. Hence, the standard Viterbi algorithm with branch metrics $\lambda[i]$ can be applied for decoding.

As indicated in Section 1, for metric computation the dependences between consecutive channel output symbols should be incorporated as completely as possible. Thus, we base the metric on the observation of N received symbols. In particular, we start from the conditional probability density function (pdf) $p(\boldsymbol{y}[k]|\boldsymbol{a}[k])$ of $\boldsymbol{y}[k] \triangleq [y[k], y[k-1], \ldots, y[k-N+1]]^T$ under the assumption $\boldsymbol{a}[k] \triangleq [a[k], a[k-1], \ldots, a[k-N+2]]^T$, which is derived in e.g. [8].

Based on the pdf $p(\boldsymbol{y}[k]|\boldsymbol{a}[k])$ it is possible to compute the maximum–likelihood bit metrics for the $(N-1)\cdot\ell$ bits which correspond via mapping bijectively to each vector symbol $\boldsymbol{a}[k]$ as e.g. in [14]. Then, M^{N-1}/ℓ pdf calculations per $(N-1)\cdot\ell$ bit metrics are necessary, i.e., the computational effort grows exponentially with the observation length N. Here, we follow another approach, which clearly is suboptimum, but requires only a very low complexity for metric computations. Instead of considering blocks $\boldsymbol{a}[k]$ as trial symbols, we insert hard decision–feedback symbols $\hat{a}[k-\nu]$, $\nu = 1, 2, \ldots, N-2$, leaving only $a[k]$ as trial symbol. This procedure is essentially the decision rule of uncoded decision–feedback differential detection (DF–DD), e.g. [23, 24].

The μth bit, $\mu = 0, 1, \ldots, \ell-1$, of the label of $a[k]$ corresponds to the metric value $\lambda[i]$, where $i = k\cdot\ell + \mu$ holds. Subsequently, since the present symbol time index k is of no importance, we replace the index i by μ for clarity. Furthermore, the metric notation $\lambda[\mu]$ is completed by the subscript b, $b \in \{0, 1\}$, which gives the value of the considered bit, and the superscript "sym" indicating that symbols $\hat{a}[k-\nu]$ are fed back, i.e., $\lambda_b^{\mathrm{sym}}[\mu]$ is used. From the considerations above and assuming that the data symbols $a[k]$ are *a priori* equally likely, $\lambda_b^{\mathrm{sym}}[\mu]$ is given by

$$\lambda_b^{\mathrm{sym}}[\mu] = \log \sum_{a[k]\in\mathcal{A}_b^\mu} p(\boldsymbol{y}[k]|a[k], \hat{a}[k-1], \ldots, \hat{a}[k-N+2]) \;, \tag{4}$$

where $\mathcal{A}_b^\mu$ is the subset of symbols $a[k] \in \mathcal{A}$ whose label has the value b at position μ.

The metric computation is further simplified if not only decision–feedback symbols $\hat{a}[k-\nu]$, $\nu = 1, 2, \ldots, N-1$, but also $\ell - 1$ decision–feedback bits, which belong to the label $a[k]$, are inserted in (4). Then, denoting the μth estimated label bit of $a[k]$ by $\hat{c}[\mu]$ and the mapping function by $\mathcal{M}$, for calculation of $\lambda_b^{\mathrm{bit}}[\mu]$, solely $\tilde{a}[k] = \mathcal{M}(\hat{c}[0], \ldots, \hat{c}[\mu-1], b, \hat{c}[\mu+1], \ldots, \hat{c}[\ell-1])$ is considered and the resulting metric is given by:

$$\lambda_b^{\mathrm{bit}}[\mu] \;=\; \log(p(\boldsymbol{y}[k]|\tilde{a}[k], \hat{a}[k-1], \ldots, \hat{a}[k-N+2])) \; . \tag{5}$$

Instead of M, now only $\ell + 1 \leq M = 2^\ell$ pdf's have to be determined since only $\ell + 1$ different $\tilde{a}[k]$ are possible.

Noteworthy, the number of branch bit metrics $\lambda_b^{\mathrm{sym}}[\mu]$ and $\lambda_b^{\mathrm{bit}}[\mu]$, respectively, is independent of the observation length N, and identical to that for conventional differential demodulation. Because of the analogy to DF–DD and since the metrics constitute *soft* inputs for the decoder, both metric calculation strategies (4) and (5) are referred to as *decision–feedback differential demodulation (DF–DM)* [1].

To save complexity it is reasonable to neglect all multiplicative terms in the pdf's which do not depend on symbol $a[k] \in \mathcal{A}_b^\mu$ and $\tilde{a}[k]$, respectively. Using the notation $t[\nu] = t_{0\nu}$ with $t_{0\nu}$ defined in [24, Eq. (15)], $\lambda_b^{\mathrm{sym}}[\mu]$ and $\lambda_b^{\mathrm{bit}}[\mu]$ can be simplified to [24]

$$\check{\lambda}_b^{\mathrm{sym}}[\mu] = \log \sum_{a[k] \in \mathcal{A}_b^\mu} \exp\left(\mathrm{Re}\left\{ a[k] y^*[k] \cdot \sum_{\nu=1}^{N-1} t[\nu] y[k-\nu] \prod_{n=1}^{\nu-1} \hat{a}[k-n] \right\} \right) \; , \tag{6}$$

and [1]

$$\check{\lambda}_b^{\mathrm{bit}}[\mu] = \mathrm{Re}\left\{ \tilde{a}[k] y^*[k] \cdot \sum_{\nu=1}^{N-1} t[\nu] y[k-\nu] \prod_{n=1}^{\nu-1} \hat{a}[k-n] \right\} \; , \tag{7}$$

respectively, where $\mathrm{Re}\{\cdot\}$ denotes the real part of a complex number.

Since a positive multiplicative or additive constant is of no importance for the decoding decisions in the Viterbi algorithm, it is possible to replace $t[\nu]$ by $p[\nu] \triangleq c \cdot t[\nu]$, $c \in \mathbb{R}^+$, $\nu = 1, 2, \ldots, N-1$, in (6) and (7). If c is chosen properly (cf. [24]), $p[\nu]$ are the coefficients of a linear $(N-1)$st order minimum mean–squared error (MMSE) FIR predictor [40] for the random process $g[\cdot] + n[\cdot]x^*[\cdot]$. In this case, these coefficients $p[\nu]$ can be adaptively determined in a simple manner by employing, e.g., the recursive least–squares (RLS) algorithm [41, 42]. Note that the strict equivalence between maximum–likelihood based and prediction based approach holds for Rayleigh fading channels only cf. e.g. [10, 43, 17, 42].

4 Iterative Decoding Algorithm and Convergence

4.1 Iterative Decoding Algorithm

Now, the iterative decoding procedure as essentially proposed in [1] is formulated. For DF–DD of uncoded MDPSK the feedback symbols $\hat{a}[k-\nu]$ stem from immediate decisions on transmitted symbols. When error correction coding and bit–interleaving are applied it is reasonable to obtain the decision–feedback symbols from the bit decisions $\hat{c}[i]$ of the Viterbi decoder via remodulation. Of course, for the first demodulation of a received sequence (first decoding iteration), no previous decisions $\hat{c}[i]$ are available. Then, to keep the demodulation as simple as possible, we resort to conventional differential demodulation based on two consecutive received symbols, i.e., $N = 2$. For the further demodulations (decoding iterations) remodulated feedback symbols $\hat{a}[k-\nu]$ are used to calculate bit branch metrics based on an observation interval $N > 2$. If the metrics are determined from (7) address bits $\hat{c}[\mu]$ are additionally fed back.

It is worth mentioning that for BICM DF–DM can also lead to a performance improvement if $\lambda_b^{\mathrm{bit}}[\mu]$ with $N = 2$ is used in all but the first iterations. This is because the dependences between the address bits of the considered differential symbol are properly taken into account if the feedback bits are correct. Clearly, the improvement due to iterative decoding depends on the particular mapping. For coherent 8PSK and iterative decoding with hard decision–feedback the design of an optimal mapping is investigated in [32, 33]. There, another labeling than Gray labeling was found to perform best. For noncoherent iterative DF–DM with $N > 2$, our simulations have shown that the loss after the first decoding iteration could not be compensated in further iterations when using other labelings than Gray labeling. Furthermore, the gain of iterative DF–DM with $N = 2$ and non–Gray labeling was always lower than for $N > 2$ and Gray labeling. Therefore, we apply usual Gray labeling throughout this paper.

4.2 Convergence Analysis

Since the algorithm has an iterative structure the question of convergence arises. Regarding (6) and (7), respectively, inserting the predictor coefficients $p[\nu]$ for $t[\nu]$, and rewriting $y[k]$ as

$$y[k] = x[k] \cdot g[k] + n[k] = a[k] \cdot (g[k] + n[k]x^*[k]) \cdot x[k-1] \,, \tag{8}$$

we can describe the effect of DF–DM as to deliver an estimate for $\eta[k] \triangleq g[k]+n[k]x^*[k]$ by evaluating the term [42]

$$\hat{\eta}[k] \triangleq x^*[k-1] \cdot \sum_{\nu=1}^{N-1} p[\nu] y[k-\nu] \prod_{n=1}^{\nu-1} \hat{a}[k-n] \; . \tag{9}$$

From a closer examination of (9) we note that for $\hat{a}[k-\nu] = a[k-\nu]$, $\nu = 1, 2, \ldots, N-2$, $\hat{\eta}[k]$ is the MMSE estimate for $\eta[k]$ [10, 43, 17, 24, 42].

Adopting this point of view, an appropriate criterion for convergence is the variance of the estimation error

$$\sigma_e^2 \triangleq \mathcal{E}_{\eta,\hat{\eta}}\{|\eta[k] - \hat{\eta}[k]|^2\} \; , \tag{10}$$

where erroneous feedback symbols $\hat{a}[k-\nu] \neq a[k-\nu]$ occur in case of non–zero bit error rate (BER) in the previous iteration. If σ_e^2 decreases when the observation interval is increased to $N > 2$ using decision–feedback symbols, the corresponding BER will decrease as well. Thus, the algorithm converges. However, if due to erroneous decision–feedback symbols σ_e^2 increases for larger N, the corresponding BER will also increase. Then, the iterative decoding does not converge.

Therefore, we evaluate the expression for σ_e^2 in dependence of the BER in the previous iteration. Definition (10) may be rewritten to

$$\begin{aligned} \sigma_e^2 &= \mathcal{E}_{\eta}\{|\eta[k]|^2\} - 2 \cdot \mathrm{Re}\{\mathcal{E}_{\eta,\hat{\eta}}\{\eta[k]\hat{\eta}^*[k]\}\} + \mathcal{E}_{\hat{\eta}}\{|\hat{\eta}[k]|^2\} \\ &= \sigma_g^2 + \sigma_n^2 - 2 \cdot \mathrm{Re}\{\mathcal{E}_{\eta,\hat{\eta}}\{\eta[k]\hat{\eta}^*[k]\}\} + \mathcal{E}_{\hat{\eta}}\{|\hat{\eta}[k]|^2\} \; . \end{aligned} \tag{11}$$

For the further analysis, we assume that the decision–feedback binary symbols are mutually statistically independent and independent of the corresponding channel gain. This assumption might be violated in a practical system, because in case of low fading gain unreliable soft decoder inputs are likely to cause a detour in the Viterbi decoding. However, this simplification is necessary for mathematical tractability and the obtained results closely match with our simulations. After some straightforward manipulations we obtain

$$\mathcal{E}_{\eta,\hat{\eta}}\{\eta[k]\hat{\eta}^*[k]\} = \sum_{\nu=1}^{N-1} p^*[\nu] R_g[\nu] \prod_{n=1}^{\nu-1} \mathcal{E}_{a,\hat{a}}\{a[k-n]\hat{a}^*[k-n]\} \tag{12}$$

$$\begin{aligned} \mathcal{E}_{\hat{\eta}}\{|\hat{\eta}[k]|^2\} &= \sum_{\nu=1}^{N-1} \sum_{\mu=1}^{N-1} p[\nu] p^*[\mu] (R_g[\mu-\nu] + \sigma_n^2 \delta[\mu-\nu]) \\ &\quad \cdot \prod_{n=\min\{\nu,\mu\}}^{\max\{\nu,\mu\}-1} \mathcal{E}_{a,\hat{a}}\{a[k-n]\hat{a}^*[k-n]\} \; , \end{aligned} \tag{13}$$

where $\delta[\cdot]$ denotes the Kronecker delta, i.e., $\delta[0] = 1$, $\delta[\lambda] = 0$ for $\lambda \neq 0$. The evaluation of $\mathcal{E}_{a,\hat{a}}\{a[k-n]\hat{a}^*[k-n]\}$ as a function of BER requires the choice of the signal constellation and the labeling. As mentioned above, we apply Gray labeling. If, as in the following, 4PSK and 8PSK constellations are considered, and due to bit–interleaving statistically independent bit errors are assumed, the expected value reads

$$\begin{aligned} \alpha &\stackrel{\triangle}{=} \mathcal{E}_{a,\hat{a}}\{a[k-n]\hat{a}^*[k-n]\} \\ &= \begin{cases} 1 - 2 \cdot \mathrm{BER} & \text{for 4PSK}^1 \\ 1 - (3 - \sqrt{2}/2) \cdot \mathrm{BER} + (2 - \sqrt{2}) \cdot \mathrm{BER}^2 & \text{for 8PSK.} \end{cases} \end{aligned} \tag{14}$$

Finally, using (12)–(14) in (11) yields the error variance as a function of BER in the previous iteration:

$$\begin{aligned} \sigma_e^2 &= \sigma_g^2 + \sigma_n^2 - 2 \cdot \mathrm{Re}\left\{ \sum_{\nu=1}^{N-1} p^*[\nu] R_g[\nu] \cdot \alpha^{\nu-1} \right\} \\ &\quad + \sum_{\nu=1}^{N-1} \sum_{\mu=1}^{N-1} p[\nu] p^*[\mu] (R_g[\mu - \nu] + \sigma_n^2 \delta[\mu - \nu]) \cdot \alpha^{|\nu-\mu|} \,. \end{aligned} \tag{15}$$

As an example, we assume 4DPSK transmission over a Rayleigh fading channel with $f_d T = 0.01$, coding with the widely applied convolutional code of rate 1/2, memory 6, and generator polynomials $(133, 171)_8$, and Viterbi decoding.

Figure 2 shows σ_e^2 as a function of $\bar{E}_s/N_0$ ($\bar{E}_s$: average energy per received signal, N_0: one–sided noise power spectral density) for $N = 2$ in the first iteration (without feedback), and for $N = 3, 5, 10$ in the second iteration applying both DF–DM with BER $\neq 0$ (solid lines)[2] and genie–aided DF–DM, i.e., BER $= 0$ (dashed lines) for generating the feedback symbols. Clearly, the curves of genie–aided DF–DM constitute the performance limit of DF–DM. In these cases, increasing N results in a reduced variance of the estimation error for all SNR's. For comparison, also the curve for $N \to \infty$ obtained from [24, Eq. (51)] is included. If erroneous feedback symbols due to BER $\neq 0$ are taken into account the curves for the first ($N = 2$) and second iteration ($N > 2$) intersect. As expected, for relatively low SNR, where BER is high (cf. Figure 6), the error variance increases in the second iteration, whereas σ_e^2 decreases for higher SNR. Since an erroneous feedback symbol is more often used in the demodulation with larger N, the cross–over points of the curves depend on N. Overall, convergence of DF–DM is reached at relatively low values of $\bar{E}_s/N_0$. Hence, we conclude that the iterative decoding algorithm with DF–DM converges for usually desired BER's.

[1]Note that for 2PSK the same result as for 4PSK is obtained.

[2]Here, BER is taken from the simulated curve for $N = 2$ in Figure 6.

Noteworthy, for high SNR, the ultimate performance gain, which is illustrated by the genie–aided DF–DM curves, is already almost achieved after two iterations (one feedback of decision) for $10\log_{10}(\bar{E}_s/N_0) > 8$ dB.

A comparison of this convergence prediction with the simulation results presented in Figure 6 will be made in Section 6.

5 Cutoff Rate for Genie–Aided DF–DM

A good trade–off between powerful error correcting coding and fair complexity of decoding is offered by application of convolutional codes and decoding with the Viterbi algorithm. For this coding scheme, the cutoff rate R_0 of the associated memoryless channel is a common measure of performance [34]. Convolutional codes with soft input Viterbi decoding usually achieve a low BER when operating in the vicinity of the cutoff rate limit.

Hence, R_0 for genie–aided DF–DM, i.e., all decision–feedback symbols are assumed to be correct, is used to judge the achievable performance of DF–DM for different window lengths N. Following the derivations in [30], we define the so–called average Bhattacharyya factor

$$B \triangleq \frac{1}{\ell}\sum_{\mu=0}^{\ell-1}\mathcal{E}_{b,\boldsymbol{y}}\left\{\sqrt{\frac{\exp\left(\lambda_{\bar{b}}^{\times}[\mu]\right)}{\exp\left(\lambda_{b}^{\times}[\mu]\right)}}\right\}\ , \quad \times \in \{\mathrm{sym},\mathrm{bit}\}\ , \tag{16}$$

where $\lambda_b^{\mathrm{sym}}[\mu]$ and $\lambda_b^{\mathrm{bit}}[\mu]$ are taken from (4) and (5), respectively, and $\bar{b}$ denotes the complement of b. The cutoff rate of BICM in bits per channel use is given by

$$R_0 = \ell \cdot (1 - \log_2(B+1))\ . \tag{17}$$

Subsequently, this expression is numerically evaluated for DF–DM utilizing also decision–feedback bits labeling $a[k]$, i.e., $\lambda_b^{\mathrm{bit}}[\mu]$, which offers the largest computational savings. Since genie–aided DF–DM is assumed, there is no dependence on the number of iterations.

First, 4DPSK is considered. In Figure 3, the influence of the observation length N is examined for Jakes fading model with $f_dT = 0.01$. The curve for $N = 2$ represents the cutoff rate corresponding to conventional differential demodulation without decision–feedback. Noteworthy, in case of $N = 2$ DF–DM does not yield any performance improvement as for Gray labeling the difference metrics $(\lambda_0^{\mathrm{sym}}[\mu]-\lambda_1^{\mathrm{sym}}[\mu])$ and $(\lambda_0^{\mathrm{bit}}[\mu]-\lambda_1^{\mathrm{bit}}[\mu])$ are equivalent for 4DPSK (cf. also (6), (7)), i.e., the two bits per symbol

are transmitted independently of each other. As reference curve and upper bound, the cutoff rate of coherent 4PSK with perfect channel state information (CSI) at the receiver is also shown. As can be seen, there is a relatively large potential gain in power efficiency by expanding the demodulation window from $N = 2$ to $N = 3$ for DF–DM. A further increase of N leads to a higher R_0 for given $\bar{E}_s/N_0$, but the improvements are comparatively small. This observation will be confirmed by our simulation results in Section 6.

The interaction between the fading rate f_dT and the achievable performance of 4DPSK with DF–DM is illustrated in Figure 4, where the respective cutoff rates for $f_dT = 0.001$ and $f_dT = 0.05$ are plotted. For slow fading ($f_dT = 0.001$), there is a steady improvement in terms of R_0 by increasing N. On the other hand, for fast fading, ($f_dT = 0.05$) DF–DM is not expected to provide gains for all rates. In particular, for $R_0 \approx 1$ bit/(channel use) the required SNR is about the same for conventional demodulation with $N = 2$ and DF–DM with $N = 3, 4, 5$. This has been confirmed by simulations and it is also in accordance with simulation results presented in [17], where only small gains in power efficiency are achieved by iterative soft–feedback decoding with $N = 4$ for the same transmission rate and $f_dT = 0.05$ [3]. However, if high rate codes are employed, DF–DM is also advantageous in fast fading environments, since the flattening of the curve for $N = 2$ and $f_dT = 0.05$ at $R_0 \approx 1.65$ bit/(channel use) moves to values of about 2 bit/(channel use). This coincides with the results in [24] for uncoded 4DPSK transmission, where for differential detection with $N = 2$ an error floor is observed for fast fading, which is practically removed by DF–DD with $N = 4$. For comparison, the cutoff rate for coherent 4PSK, which is independent of f_dT, of course, is also included in Figure 4.

Next, the cutoff rate for 8DPSK is considered. Figure 5 shows the curves for conventional differential demodulation with $N = 2$ and DF–DM with $N = 3, 5, 10$ for fast fading with $f_dT = 0.05$ and for slow fading with $f_dT = 0.001$. As there is only a minor gain for DF–DM with $N = 2$ (see Section 6), the corresponding curves are omitted for clarity.

As for 4DPSK, in a slow fading environment relatively large performance gains are already achievable by DF–DM with $N = 3$. By increasing N, the cutoff rate for coherent transmission with perfect CSI is approached. For fast fading and DF–DM with $N = 3$, the potential gains depend strongly on the desired transmission rate. The

[3] Note that in [17] the cutoff rate of the applied iterative soft–feedback decoding scheme is not examined. Thus, there a satisfactory explanation for the absence of a gain for $N = 4$ and $f_dT = 0.05$ could not be given.

flattening of R_0 for conventional differential demodulation with $N = 2$ is significantly mitigated. However, due to the fast fading, a considerable gap remains between the curves for coherent transmission with perfect CSI and noncoherent transmission without CSI even for DF–DM with $N = 10$. These analytic results are in great accordance with the simulation results presented in Section 6 and in [1, Fig. 2].

Finally, we would like to mention that the cutoff rates for genie–aided DF–DM with branch metrics $\lambda_b^{\mathrm{sym}}[\mu]$ are only marginally inferior to the cutoff rates for metrics $\lambda_b^{\mathrm{bit}}[\mu]$, which are evaluated in Figures 3–5.

6 Simulation Results

To further assess the performance of iterative DF–DM, the system in Figure 1 with flat Rayleigh fading has been simulated. In this section, the measured BER's are presented as functions of $\bar{E}_b/N_0$ ($\bar{E}_b$: average energy per information bit). As for calculation of the cutoff rate, we concentrate on 4DPSK and 8DPSK transmission. For channel coding, BICM with Gray labeling and convolutional codes with standard Viterbi decoding are applied. The bit–interleavers are randomly generated for each transmitted BICM block to obtain results which are independent of a particular interleaver.

For 4DPSK, the target rate of 1 bit/(channel use) is regarded as a relevant example, which is generated through the standard rate 1/2 convolutional code with 64 states (see Section 4.2). The flat fading channel is specified by $f_dT = 0.01$. Bit–interleaving over 4000 bits, which correspond to 2000 channel symbols, is applied to provide transmission diversity.

In Figure 6, the convergence of the iterative decoding algorithm is illustrated. The curves correspond to the bit error rates for conventional differential demodulation with $N = 2$ and for iterative DF–DM with $N = 3$ and $N = 5$. As performance limits, the measured BER's for genie–aided DF–DM with $N = 3$ and $N = 5$ are plotted over the SNR. The curves show that iterative DF–DM provides considerable gains in power efficiency for a wide range of target BER's. In accordance with the convergence analysis in Section 4.2 the curves for iterative DF–DM with $N = 3$ and $N = 5$ intersect. Due to the (nonlinear) Viterbi decoding, the curves for σ_e^2 in Figure 2 do not translate one–to–one into the curves for BER in Figure 6, i.e., the intersection points are not located exactly at the same values of SNR. Comparing the results for genie–aided DF–DM and iterative DF–DM for $N = 3$ shows that almost the whole achievable performance gain is obtained after the second iteration, i.e., decision–feedback is performed only once.

For iterative DF–DM with $N = 5$ and regarding target bit error rates of about 10^{-4} and 10^{-5}, three decoding iterations are sufficient.

The performance of conventional differential demodulation, iterative DF–DM with several observation lengths N, and coherent 4PSK with perfect CSI at the receiver are compared in Figure 7. Here, DF–DM with four iterations is considered. Note that for $N = 3, 5$ almost identical BER's are obtained with fewer iterations (Figure 6). Again, the curves for genie–aided DF–DM give lower bounds on BER of iterative DF–DM. The large gap in power efficiency between coherent and conventional differential demodulation is significantly reduced by the simple iterative DF–DM scheme. About 1.8 dB can be gained for BER $\leq 10^{-4}$. The simulation results coincide with the cutoff rate analysis (see Figure 3), where very similar performance gains are predicted. In particular, iterative DF–DM with $N = 10$ is only marginally superior to iterative DF–DM with $N = 5$ for $f_dT = 0.01$ and code rate 1/2.

It can be seen from Figure 7 that the BER curves for genie–aided DF–DM and coherent detection have different slopes. This is in accordance with [24] where a similar behavior can be observed for uncoded transmission and DF–DD (cf. e.g. [24, Fig. 8] for $N = 3, 4$ and [24, Fig. 6] for $N \to \infty$).

For a comparison with iterative decoding schemes using soft decision–feedback simulation results with almost identical parameters (only the bit interleaver size is different) presented in [17, Fig. 6] can be considered. There, for 4DPSK and Rayleigh fading with $f_dT = 0.01$, a gain of about 2.8 dB over conventional differential demodulation is achieved after three iterations at BER $= 10^{-4}$, which is a further improvement of 1 dB compared to iterative DF–DM[4]. However, whereas for DF–DM 64 state standard Viterbi decoding and the evaluation of elementary metric expressions (cf. (7)) suffices, the scheme in [17] requires two 64 state soft–output algorithms. Furthermore, the number of states of the employed soft–output demodulator increases polynomially with the number of signal points M and exponentially with the length of the observation window N. Thus, the prize to be paid for further improvement in power efficiency is a much higher complexity, whereas for iterative DF–DM complexity is almost independent of M and N.

Now, more bandwidth efficient 8DPSK transmission with 2 bit/(channel use) is considered. The punctured rate 2/3 convolutional code with 64 states (generator polynomials $(135, 163)_8$) is taken from [44]. We assume fading with $f_dT = 0.001$ and ap-

[4]When comparing absolute quantities of SNR in Figure 7 and [17, Fig. 6] it should be noted that the larger interleaver size in [17] leads to a better performance in [17, Fig. 6] in general.

propriate bit–interleaving of 60000 bits, which corresponds to 20000 channel symbols[5]. Both bit branch metrics (6) and (7) are applied for comparison.

The simulation results are plotted in Figure 8. Note that in case of bit metric (6) using symbol feedback only, DF–DM with $N = 2$ is identical to conventional differential demodulation with $N = 2$, i.e., no feedback symbols are incorporated in the metric computation. Accordingly, the curves for $N = 2$ reflect the improvement exclusively due to bit feedback. Clearly, also for $N = 3$, 5 DF–DM with additional bit feedback is superior to DF–DM with symbol feedback only. However, consistent with the notion of Gray labeling the gains due to bit feedback are rather small. Hence, the main advantage of combined symbol and bit feedback is the simplicity of the corresponding bit branch metric (7). As the fading is relatively slow, the power efficiency of coherent 8PSK can be well approached by DF–DM as was expected from the cutoff rate analysis (cf. Figure 5). 8DPSK and DF–DM with $N = 5$ requires an SNR for BER $\leq 10^{-4}$ which is only about 1 dB higher than the SNR for coherent 8PSK with perfect CSI. The gain of using DF–DM with $N = 5$ compared to usual differential demodulation amounts to about 2.2 dB. Noteworthy, also for 8DPSK the iterative decoding algorithm converges for the target BER's which are usually of interest.

7 Conclusions

A simple iterative decoding scheme for bit–interleaved convolutional coded MDPSK transmission over flat Rayleigh fading channels without channel state information at the receiver recently proposed in [1] is discussed in detail. Due to the use of hard–decision feedback this decoding procedure has the distinct advantage of requiring only a very moderate increase in computational complexity compared to conventional differential demodulation with observation length $N = 2$.

Two different expressions for bit branch metrics are derived. Based on a convenient metric representation, the demodulation with observation length $N > 2$ is regarded as an estimation problem, which enables the investigation of the convergence of the iterative decoding algorithm. It turns out that the algorithm converges for usually desired BER's, which is exemplarily shown for 4DPSK transmission.

By a cutoff rate analysis the potential gains in power efficiency are quantified. It is shown that the achievable performance gains strongly depend on the transmission rate and the fading rate.

[5]For the fast fading case of $f_dT = 0.05$ we refer to [1, Fig. 2]

Both convergence and cutoff rate analysis well support and motivate the obtained simulation results, cf. also [1]. Compared to conventional differential demodulation remarkable performance improvements are achieved. Although larger gains can be obtained by iterative soft–feedback decoding, the considerably higher complexity of such schemes might be prohibitive. Thus, low–complexity iterative DF–DM is a promising solution for a number of applications, e.g. mobile communications, where computational complexity is limited.

References

[1] L.H.-J. Lampe and R. Schober. Decision–Feedback Differential Demodulation of Bit–Interleaved Coded MDPSK. *Electronics Letters*, 35(25):2170–2171, December 1999.

[2] J. Wolfowitz. Memory increases capacity. *Information and Control*, 11:423–428, 1967.

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

[4] 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. Com.*, 42(1):76–89, January 1994.

[5] Y. Kofman, E. Zehavi, and S. Shamai (Shitz). *nd*–Convolutional Codes—Part I: Performance Analysis. *IEEE Trans. Inf. Theory*, 43(2):558–575, March 1997.

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

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

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

[9] Y. Kofman, E. Zehavi, and S. Shamai (Shitz). *nd*–Convolutional Codes—Part II: Structural Analysis. *IEEE Trans. Inf. Theory*, 43(2):576–589, March 1997.

[10] 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. Com.*, 42(1):63–74, January 1994.

[11] D. Raphaeli. Noncoherent Coded Modulation. *IEEE Trans. Com.*, 44:172–183, February 1996.

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

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

[14] 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.

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

[16] 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.

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

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

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

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

[21] H. Leib and S. Pasupathy. The Phase of a Vector Pertubed by Gaussian Noise and Differentially Coherent Receivers. *IEEE Trans. Inf. Theory*, 34:1491–1501, November 1988.

[22] F. Edbauer. Bit Error Rate of Binary and Quarternary DPSK signals with Multiple Differential Feedback Detection. *IEEE Trans. Com.*, 40:457–460, March 1992.

[23] F. Adachi and M. Sawahashi. Decision Feedback Multiple–Symbol Differential Detection of M–ary DPSK. *Electronics Letters*, 29(15):1385–1387, July 1993.

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

[25] R. Schober, W.H. Gerstacker, and J.B. Huber. Improving Differential Detection of MDPSK by Nonlinear Noise Prediction and Sequence Estimation. *IEEE Trans. Com.*, 47(8):1161–1172, August 1999.

[26] A.J. Viterbi. Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. *IEEE Trans. Com.*, IT–13(2):260–269, April 1967.

[27] G. Colavolpe and R. Raheli. Noncoherent Sequence Detection of M-ary PSK. In *Proc. IEEE Int. Conf. Commun. (ICC)*, pages 21–25, Montreal, June 1997.

[28] R. Raheli, A. Polydoros, and C.-K. Tzou. Per–Survivor Processing: A General Approach to MLSE in Uncertain Environments. *IEEE Trans. Com.*, 43:354–364, February/April 1995.

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

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

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

[32] X. Li and J.A. Ritcey. Bit–Interleaved Coded Modulation with Iterative Decoding. *IEEE Com. Letters*, vol. 1:169–171, November 1997.

[33] X. Li and J.A. Ritcey. Trellis–Coded Modulation with Bit Interleaving and Iterative Decoding. *IEEE J. Select. Areas Commun.*, 17(4):715–724, April 1999.

[34] A.J. Viterbi and J.K. Omura. *Principles of Digital Communications and Coding.* McGraw–Hill, Tokyo, 1979.

[35] L.E. Franks. Carrier and Bit Synchronisation in Data Communication—A Tutorial Review. *IEEE Trans. Com.*, 28(8):1107–1121, August 1980.

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

[37] 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.

[38] R.H. Clarke. A Statistical Theory of Mobile–Radio Reception. *Bell System Technical Journal*, 47:957–1000, July/August 1968.

[39] W.C. Jakes, Jr. *Microwave Moble Communications.* John Wiley & Sons, Inc., New York, 1974.

[40] Simon Haykin. *Adaptive Filter Theory.* Prentice–Hall, Englewood Cliffs, NJ, third edition, 1996.

[41] R.J. Young and J.H. Lodge. Detection of CPM Signals in Fast Rayleigh Flat–Fading Using Adaptive Channel Estimation. *IEEE Trans. Vehicular Technology*, pages 338–347, May 1995.

[42] R. Schober and W.H. Gerstacker. Decision–Feedback Differential Detection Based on Linear Prediction for MDPSK Signals Transmitted over Ricean Fading Channels. *IEEE J. Select. Areas Commun.*, 18:391–402, March 2000.

[43] I. D. Marsland and P. T. Mathiopoulos. Multiple Differential Detection of Parallel Concatenated Convolutional (Turbo) Codes in Correlated Fast Rayleigh Fading. *IEEE J. Select. Areas Commun.*, 16(2):265–275, February 1998.

[44] G.C. Clark, Jr. and J.B. Cain. *Error–Correction Coding for Digital Communications.* Plenum Press, New York, 2nd edition, 1982.

List of Figures

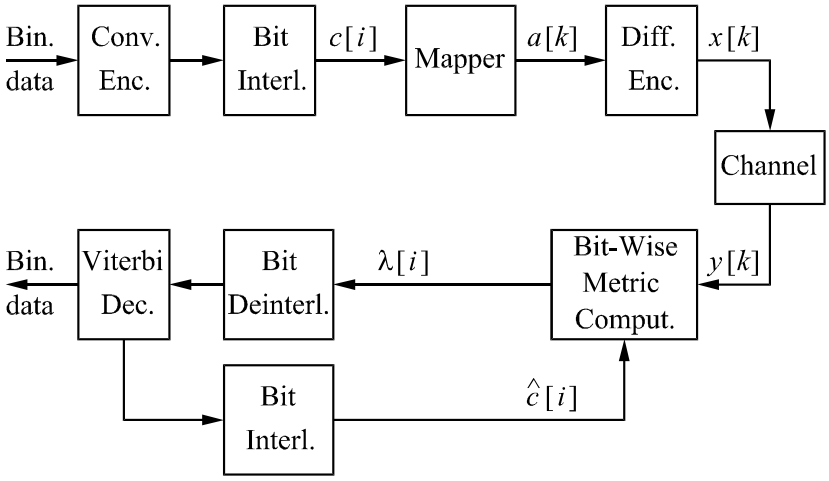

Figure 1: *Discrete–time system model.*

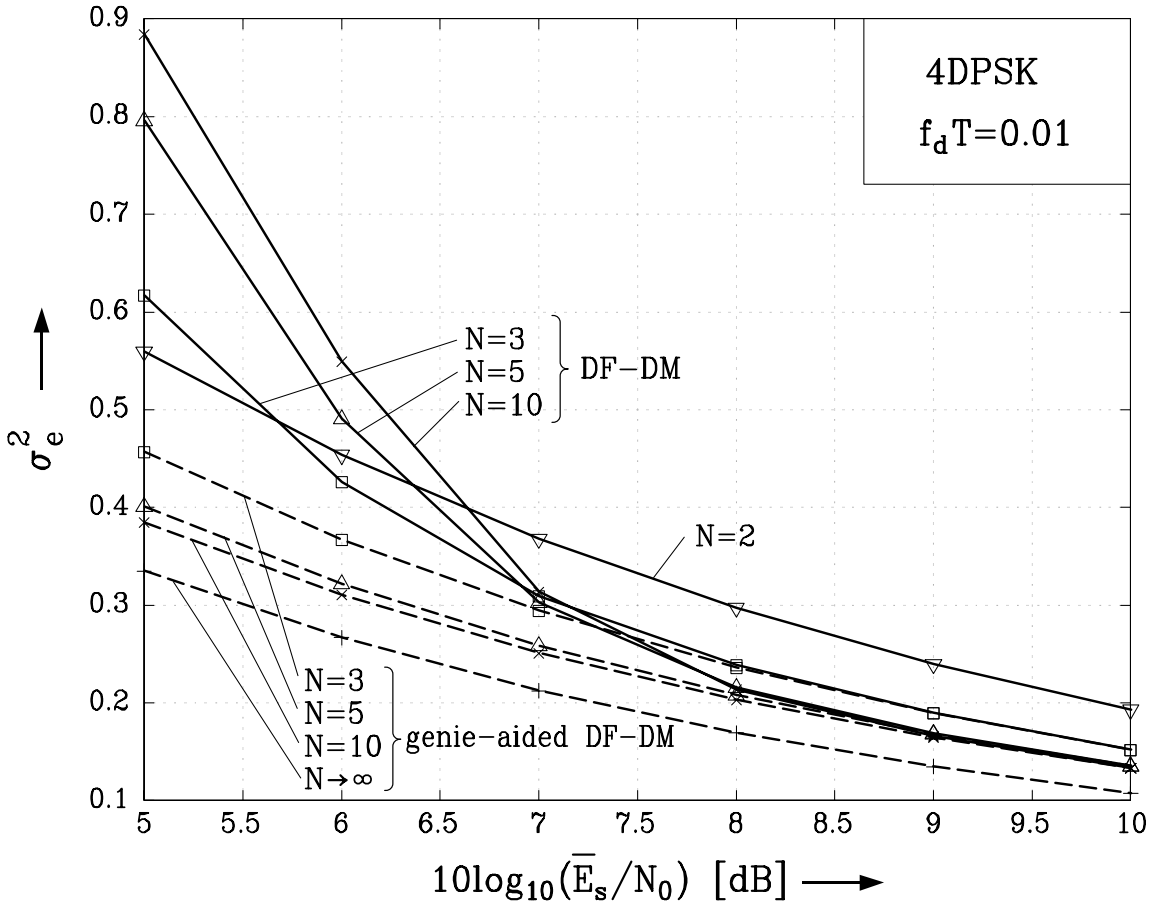

Figure 2: *Variance of the estimation error for first iteration (N=2) and second iteration with DF–DM (N=3,5,10). Solid lines: first iteration and DF–DM. Dashed lines: genie–aided DF–DM. 4DPSK and Rayleigh fading with* $f_dT = 0.01$ *is used.*

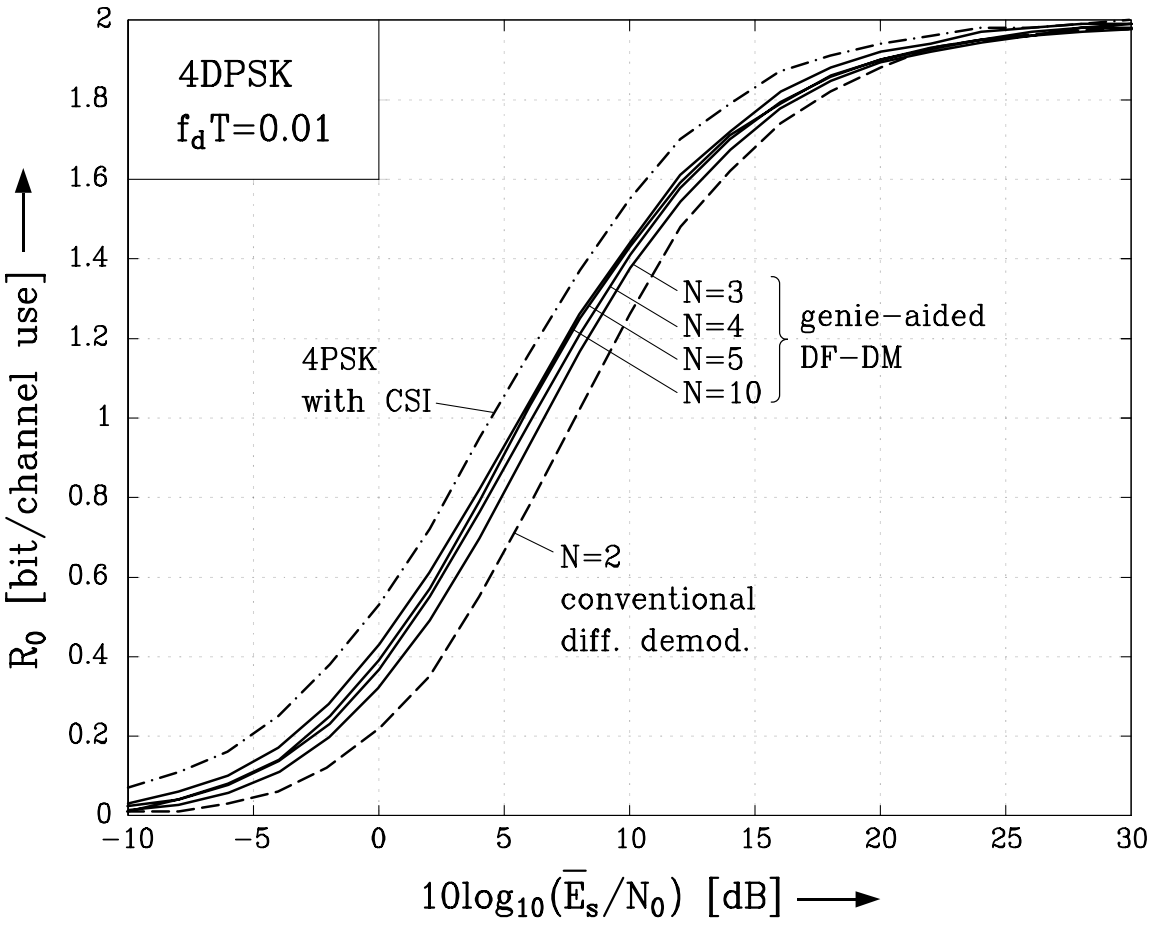

Figure 3: *Cutoff rate for 4DPSK and Rayleigh fading with $f_dT = 0.01$. Dashed line: conventional differential demodulation with $N = 2$. Solid lines: genie–aided DF–DM with $N = 3, 4, 5, 10$. Dash–dotted line: 4PSK with perfect CSI.*

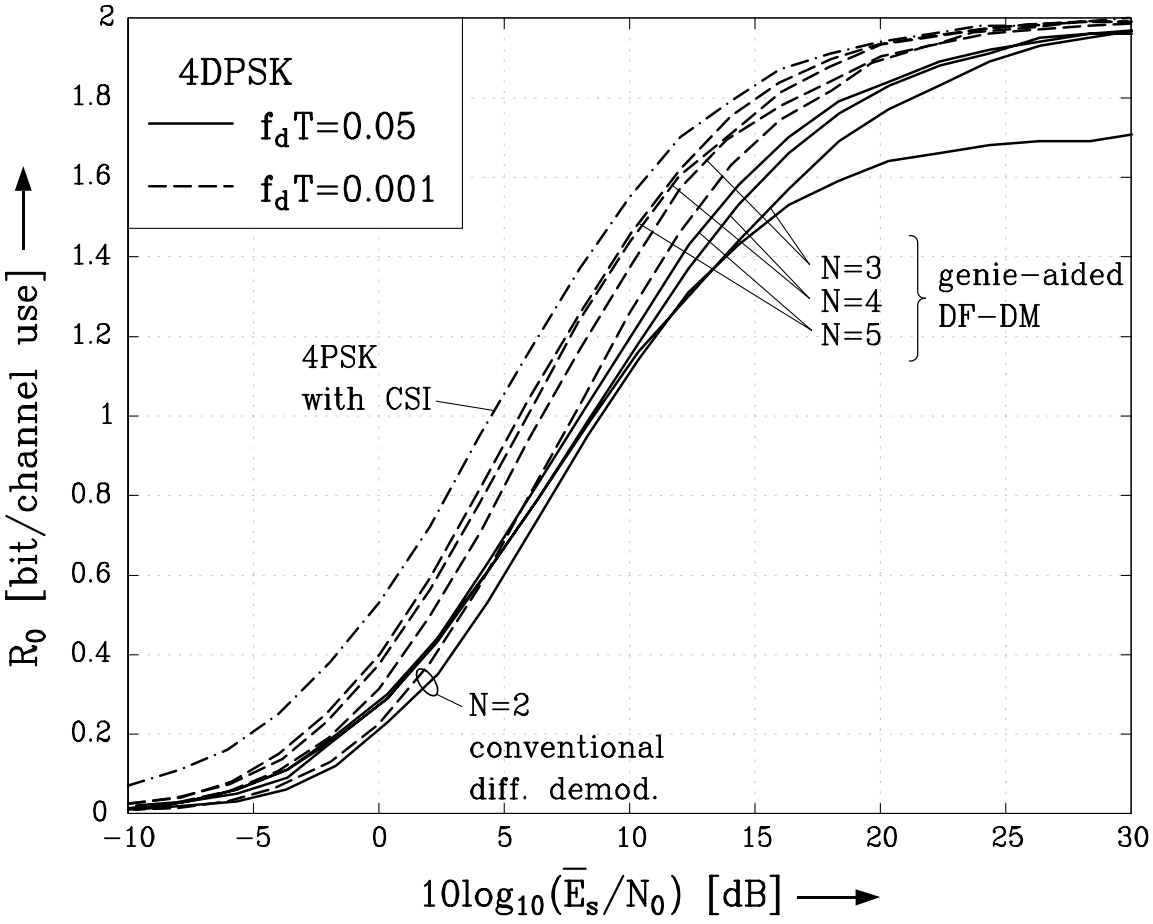

Figure 4: *Cutoff rate for 4DPSK and Rayleigh fading. Conventional differential demodulation with $N = 2$ and genie–aided DF–DM with $N = 3, 4, 5$. Dashed lines: $f_dT = 0.001$. Solid lines: $f_dT = 0.05$. Dash–dotted line: 4PSK with perfect CSI.*

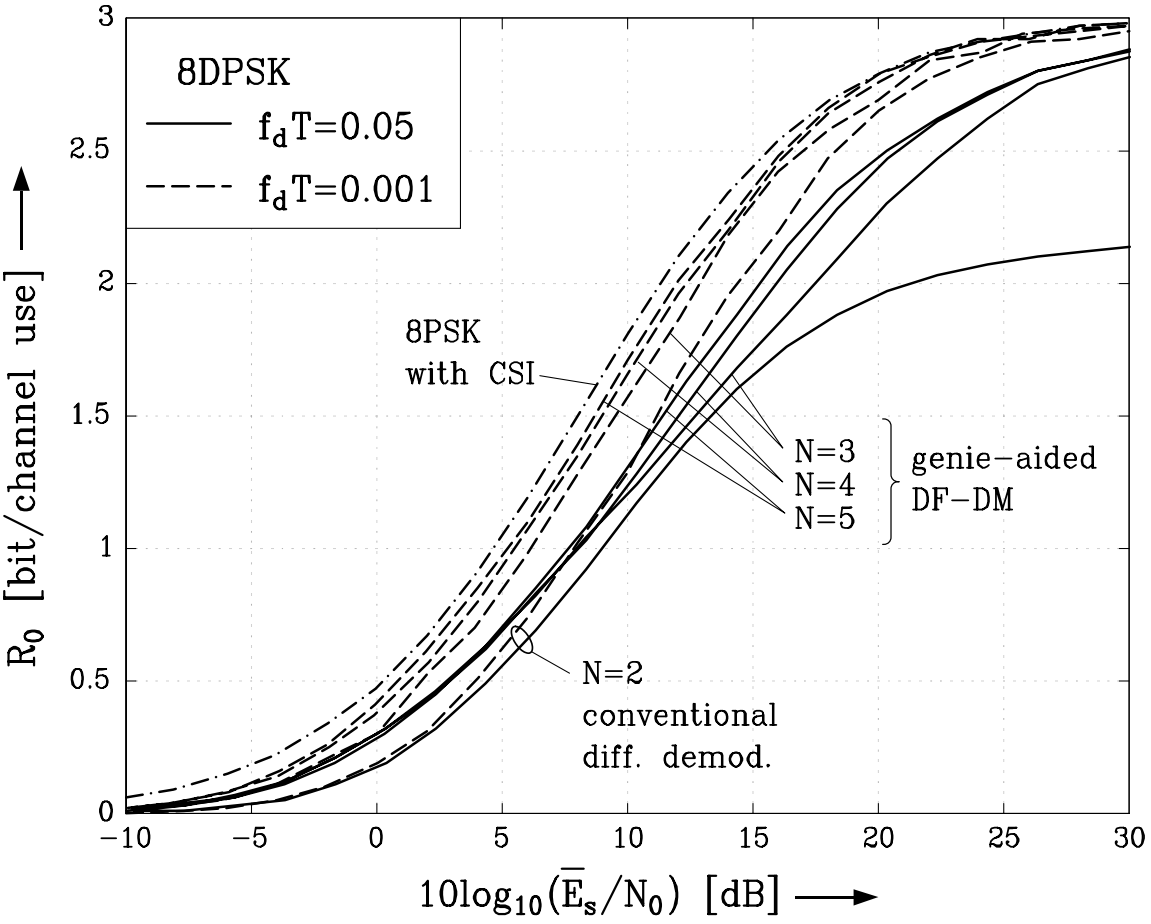

Figure 5: *Cutoff rate for 8DPSK and Rayleigh fading. Conventional differential demodulation with $N = 2$ and genie–aided DF–DM with $N = 3, 4, 5$. Dashed lines: $f_dT = 0.001$. Solid lines: $f_dT = 0.05$. Dash–dotted line: 8PSK with perfect CSI.*

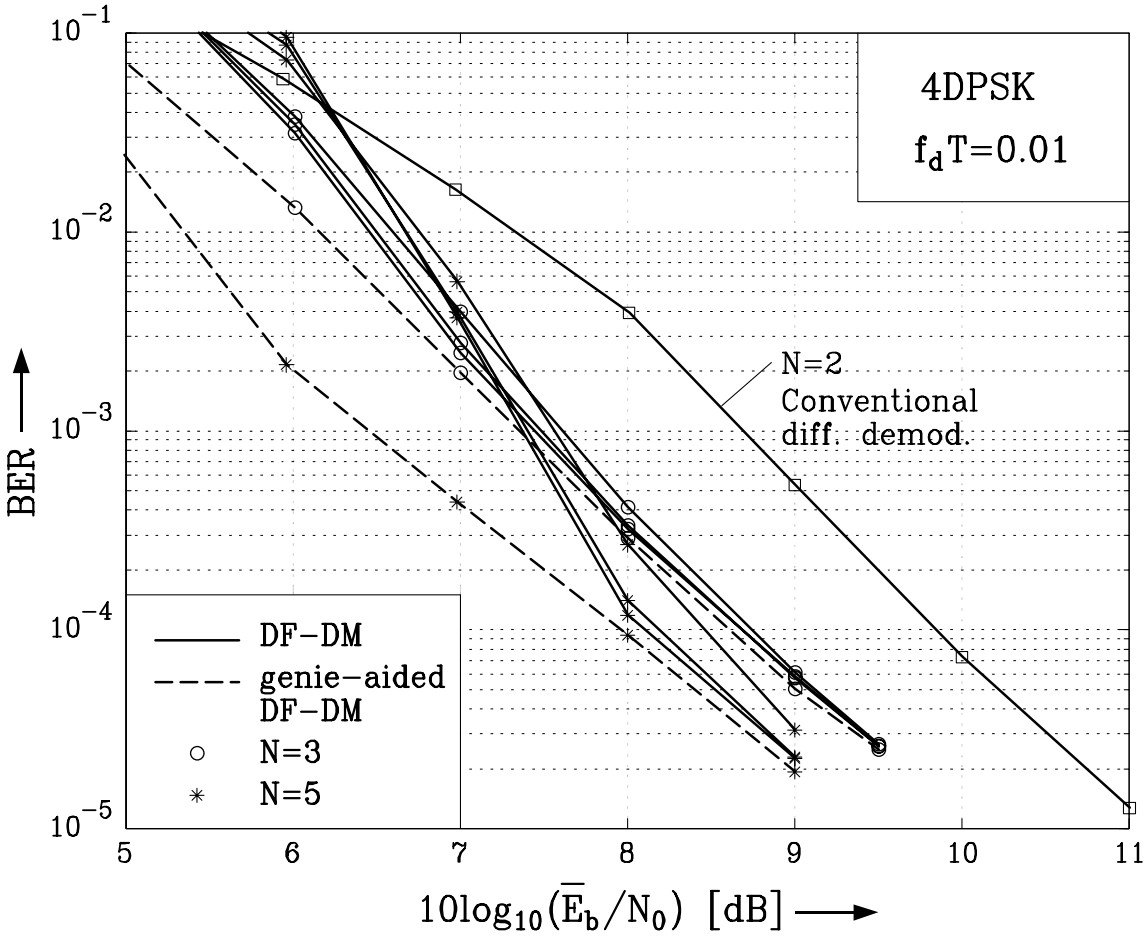

Figure 6: *BER versus* $10\log_{10}(\bar{E}_b/N_0)$ *for 4DPSK and Rayleigh fading with* $f_dT = 0.01$. *64 states convolutional code. Solid lines: conventional differential demodulation with* $N = 2$ *and DF–DM with* $N = 3, 5$. *First four iterations. Dashed lines: genie–aided DF–DM.*

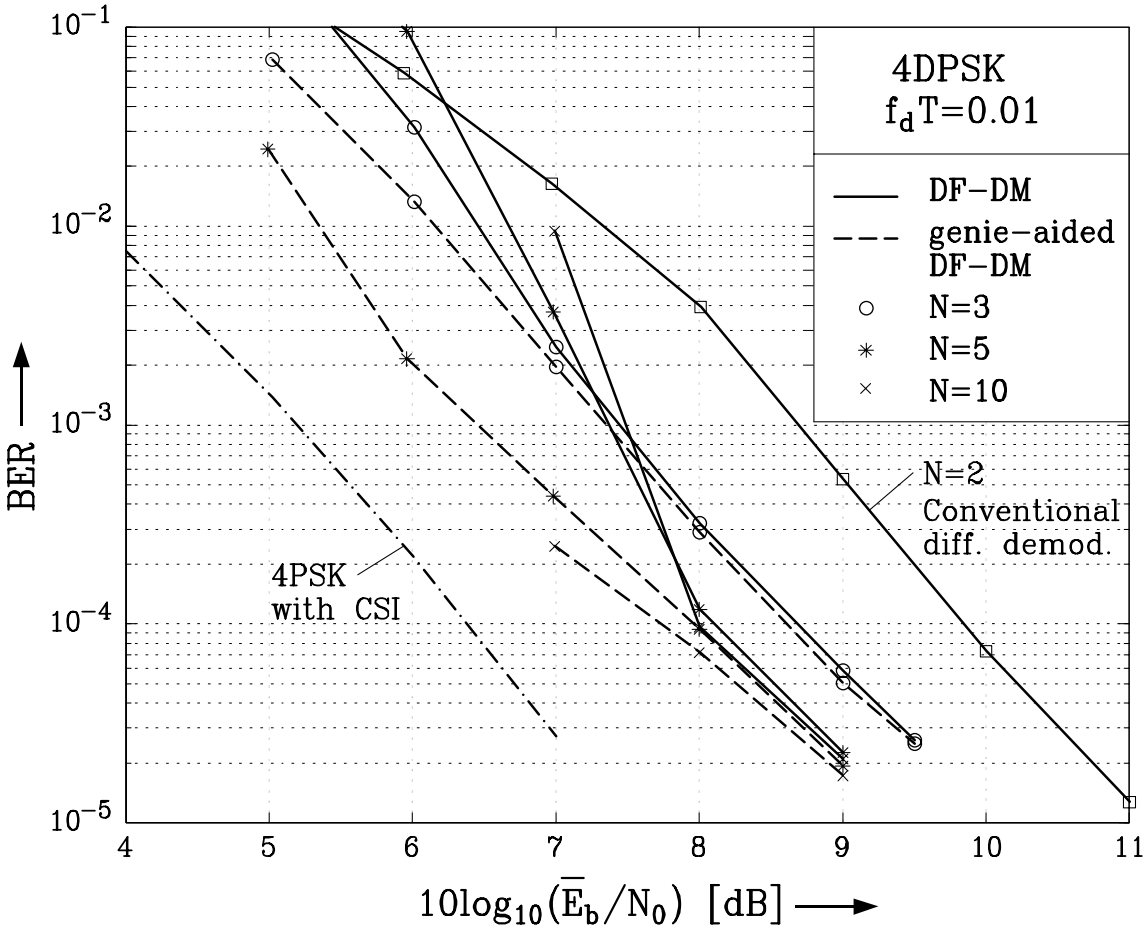

Figure 7: *BER versus* $10\log_{10}(\bar{E}_b/N_0)$ *for 4DPSK and Rayleigh fading with* $f_dT = 0.01$. *64 states convolutional code. DF–DM with four iterations. Solid lines: conventional differential demodulation with* $N = 2$ *and DF–DM with* $N = 3, 5, 10$. *Dashed lines: genie–aided DF–DM . Dash–dotted line: 4PSK with perfect CSI.*

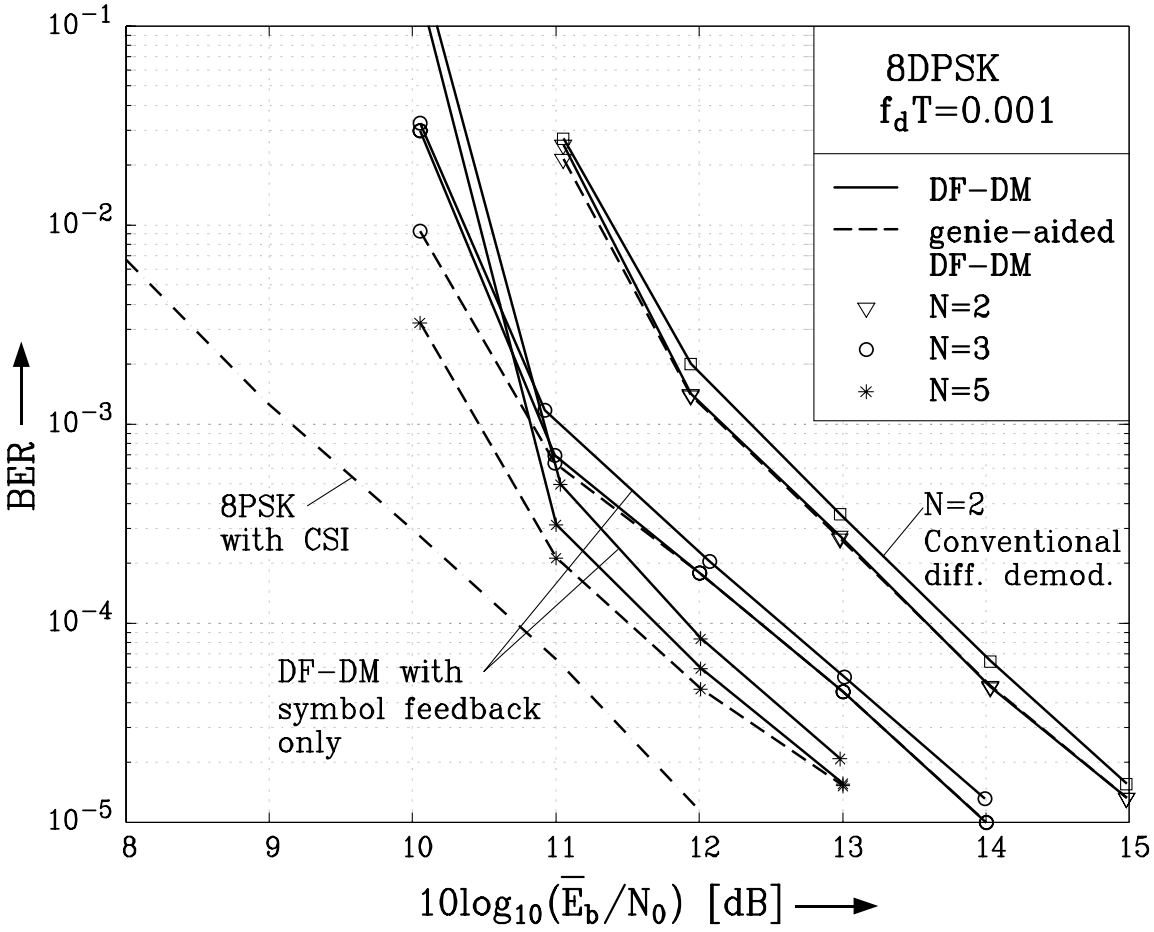

Figure 8: *BER versus* $10\log_{10}(\bar{E}_b/N_0)$ *for 8DPSK and Rayleigh fading with* $f_dT = 0.001$. *64 states convolutional code. DF–DM with four iterations. Solid lines: conventional differential demodulation with* $N = 2$ *and DF–DM with* $N = 2, 3, 5$. *Dashed lines: genie–aided DF–DM. Dash–dotted line: 8PSK with perfect CSI.*