Optimization of Delay Diversity for Fading ISI Channels

Robert Schober[1], Thorsten Hehn[2], and Wolfgang H. Gerstacker[3]

[1] The University of British Columbia, Vancouver, Canada, E-mail: rschober@ece.ubc.ca
[2] Chair for Information Transmission, University of Erlangen-Nürnberg
[3] Chair for Mobile Communications, University of Erlangen-Nürnberg, E-mail: gersta@LNT.de

Abstract

In this paper, we propose an optimized delay diversity (ODD) scheme for fading intersymbol interference (ISI) channels. We derive a cost function for optimization of the ODD transmit filters and we provide a steepest descent algorithm for iterative calculation of the filter coefficients. In addition, an upper bound on the cost function is derived and employed to prove the asymptotic optimality of the generalized DD (GDD) scheme in [1] for very high signal–to–noise (SNR) ratios and transmit filters of maximum length. However, for SNRs of practical interest and reasonable filter lengths the novel ODD scheme significantly outperforms GDD for both optimum and suboptimum equalization.

1 Introduction

Recently a number of space–time block coding schemes has been proposed to improve the performance of transmission over fading intersymbol interference (ISI) channels, cf. e.g. [2], [3], [4]. Unfortunately, if these schemes are used to upgrade existing systems from single antenna transmission to transmit diversity (TD), the burst structure has to be modified and the receiver processing (channel estimation, equalization, etc.) has to be changed significantly. A space–time trellis code (STTC) designed originally for flat fading with two transmit antennas was adopted in [5] for frequency–selective fading. Since this particular STTC can be interpreted as a data dependent delay diversity (DD) scheme, the overall channel impulse response (CIR) is similar to the single transmit antenna case, but data dependent. This data dependence of the overall CIR makes the design of low–complexity equalizers difficult. In addition, since this STTC was designed for the flat fading case, optimality in frequency–selective channels cannot be guaranteed.

In this paper, we optimize simple DD for frequency–selective fading channels. DD was first proposed in [6] for flat fading channels, but has been recently generalized to frequency–selective channels in [1]. DD has the advantage that the overall channel can be modeled as a single–input channel. Therefore, the same channel estimation, channel tracking, and equalization techniques as in the single transmit antenna case can be used. Also, existing mobile communication systems can be upgraded easily with DD, since the burst structure does not have to be modified. In the *generalized* DD (GDD) scheme in [1] transmit antenna n_T, $1 \leq n_T \leq N_T$, delays the transmitted data stream by $(n_T - 1)L$ symbols, where N_T and L are the number of transmit antennas and the CIR length, respectively. In that way, full TD is achieved, however, the resulting overall CIR may be excessively long, which implies a high equalizer complexity.

We further generalize DD for frequency–selective channels and employ a finite impulse response (FIR) filter of length N at each transmit antenna [6]. Based on a Chernoff bound on the pairwise error probability (PEP) for maximum–likelihood sequence estimation (MLSE), we derive a cost function that is suitable for optimization of the FIR filter coefficients. A closed–form optimization does not seem to be feasible and we devise a steepest descent (SD) type of algorithm for filter search. Since we cannot guarantee that the SD algorithm finds the global maximum of the cost function, we also derive an upper bound for the maximum of the cost function. The bound allows us to show that for very high signal–to–noise ratios (SNRs) the GDD scheme in [1] is optimum in the sense that it maximizes the proposed cost function. However, we also show that for the realistic case of low–to–moderate SNRs our *optimized* DD (ODD) scheme significantly outperforms the GDD scheme in [1]. Our simulations show that the performance advantage of ODD over GDD is preserved if suboptimum equalization strategies such as delayed decision–feedback sequence estimation (DFSE) [7] or decision–feedback equalization (DFE) [8] are used.

2 Preliminaries

2.1 Delay Diversity

A block diagram of the transmitter of a general delay diversity scheme is shown in Fig. 1. The independent,

This research was supported in part by the Canada Research Chair (CRC) Program.

identically distributed (i.i.d.) transmit symbols $b[k]$ belong to some scalar symbol alphabet $\mathcal{A}$ such as phase–shift keying (PSK) or quadrature amplitude modulation (QAM) [9]. We assume the normalization $\mathcal{E}\{|b[k]|^2\} = 1$ ($\mathcal{E}\{\cdot\}$: statistical expectation). The transmitted signal at antenna n_T, $1 \leq n_T \leq N_T$, in time interval k, $k \in \mathbb{Z}$, is given by

$$s_{n_T}[k] = \sqrt{\frac{E_S}{N_T}}\, g_{n_T}[k] * b[k], \qquad (1)$$

where $*$ refers to discrete–time convolution and E_S is the average transmit energy. The FIR transmit filters have length N and the energy of their coefficients $g_{n_T}[\nu]$, $0 \leq \nu \leq N-1$, is normalized to $\sum_{n_T=1}^{N_T}\sum_{\nu=0}^{N-1} |g_{n_T}[\nu]|^2 = N_T$. The optimization of these filters will be discussed in Section 3. The GDD scheme in [1] is obtained as a special case for $N = (N_T-1)L+1$, and $g_{n_T}[\nu] = 1$ for $\nu = (n_T-1)L$ and $g_{n_T}[\nu] = 0$ for $\nu \neq (n_T-1)L$, $1 \leq n_T \leq N_T$. However, in the following, we will use the term GDD for any DD scheme with $g_{n_T}[\nu] = 1$ for $\nu = (n_T - 1)L_0$ and $g_{n_T}[\nu] = 0$ for $\nu \neq (n_T-1)L_0$, where $0 \leq L_0 \leq L$, $1 \leq n_T \leq N_T$. Note that $L_0 < L$ may be desirable in a practical system for complexity reasons (cf. Section 2.2).

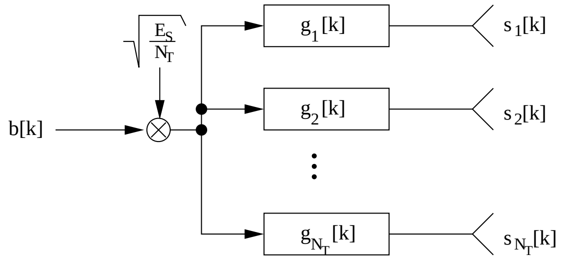

Fig. 1. Block diagram of transmitter for general delay diversity.

2.2 Channel Model

We assume N_R receive antennas and that the coefficients $h_{n_T n_R}[l]$, $0 \leq l \leq L-1$, of the discrete–time overall CIR between transmit antenna n_T and receive antenna n_R are mutually correlated zero–mean Gaussian random variables. Furthermore, in this paper, we assume that the CIRs $h_{n_T n_R}[l]$ are constant within one data burst but change independently from burst to burst (block fading model). Note that the coefficients of the discrete–time overall CIRs contain the combined effects of transmit pulse shaping, multipath Rayleigh fading, receive filtering, and sampling. In this paper, for all numerical and simulation results presented we adopt the GSM/EDGE system parameters, i.e., we assume a linearized Gaussian minimum–shift keying (GMSK) transmit pulse [10], the typical urban area (TU) and equalizer test (EQ) power delay profiles [11], and a square–root raised cosine receive filter with a roll–off factor of 0.3, which is near optimum [12]. The CIR coefficients are spatially and temporally correlated, and their $N_T N_R L \times N_T N_R L$ autocorrelation matrix (ACM) is defined as

$$\boldsymbol{\Phi} \triangleq \mathcal{E}\{\boldsymbol{h}\boldsymbol{h}^H\} \qquad (2)$$

with

$$\boldsymbol{h} \triangleq [h_{11}[0] \,\ldots\, h_{11}[L-1] \; h_{21}[0] \,\ldots\, h_{N_T N_R}[L-1]]^T, \qquad (3)$$

where $[\cdot]^T$ and $[\cdot]^H$ denote transposition and Hermitian transposition, respectively. The spatial correlation can be attributed to insufficient antenna spacing, whereas the temporal correlation is caused by transmit pulse shaping and receive filtering. Although all our derivations are valid for the general case $N_R \geq 1$, for all numerical and simulation results presented in this paper, we consider the practically most important case of one receive antenna. In addition, all CIRs corresponding to different antenna pairs have the same average energy and identical statistical properties. For simplicity, we assume that the spatial correlation between two CIRs is identical for all CIR coefficients and define the spatial correlation parameter $\rho_{\mu\nu} \triangleq \mathcal{E}\{h_{\mu 1}[l]h^*_{\nu 1}[l]\}/\sqrt{\sigma^2_{\mu 1}[l]\sigma^2_{\nu 1}[l]}$, $0 \leq l \leq L-1$, where $\sigma^2_{n_T n_R}[l] \triangleq \mathcal{E}\{|h_{n_T n_R}[l]|^2\}$ and $(\cdot)^*$ denotes complex conjugation.

The received signal at antenna n_R is given by

$$r_{n_R}[k] = \sqrt{\frac{E_S}{N_T}}\, h^{\text{eq}}_{n_R}[k] * b[k] + n_{n_R}[k], \qquad (4)$$

where the *equivalent* CIR corresponding to receive antenna n_R is defined as

$$h^{\text{eq}}_{n_R}[k] \triangleq \sum_{n_T=1}^{N_T} h_{n_T n_R}[k] * g_{n_T}[k]. \qquad (5)$$

$n_{n_R}[k]$ is the additive white Gaussian noise (AWGN) process at receive antenna n_R and has variance $\sigma^2_n = N_0$, where N_0 denotes the single–sided power spectral density of the underlying continuous–time passband noise process.

From Eq. (4) we observe that for delay TD an equivalent single–input multiple–output (SIMO) system with CIRs $h^{\text{eq}}_{n_R}[k]$, $1 \leq n_R \leq N_R$, exists. Therefore, the same channel estimation [13], channel tracking [14], and channel equalization [15], [12] techniques as in the single transmit antenna case can be used. However, the *equivalent* CIRs have length $L_{\text{eq}} = L + N - 1$ and in order to minimize receiver complexity, N should be chosen as small as possible. For example, the complexity of MLSE grows exponentially with L_{eq} [16]. In addition, for a given training sequence length the quality of the channel estimates decreases with increasing L_{eq} [13]. If N is chosen small enough so that L_{eq} does not exceed the maximum CIR length that can be tolerated by currently used receivers, existing single transmit antenna wireless systems can be upgraded to DD without changing the receiver at all.

3 Optimized Delay Diversity

Based on a Chernoff upper bound on the PEP for optimum MLSE, we first derive a cost function that is suitable for optimization of the FIR transmit filter coefficients. Then we pose the optimization problem to be solved and show that a closed–form optimization is not possible. Therefore, we derive an SD type of algorithm which can be used for filter search.

3.1 Cost Function

Let us first assume transmission of a block of K symbols $b[k]$, $0 \leq k \leq K-1$. For antenna n_R the vector $\boldsymbol{r}_{n_R} \triangleq [r_{n_R}[0]\; r_{n_R}[1]\; \ldots\; r_{n_R}[K+L_{\text{eq}}-1]]^T$ of received signal samples is given by

$$\boldsymbol{r}_{n_R} = \sqrt{\frac{E_S}{N_T}}\boldsymbol{B}\boldsymbol{G}\boldsymbol{h}_{n_R} + \boldsymbol{n}_{n_R}, \qquad (6)$$

where the definitions $\boldsymbol{h}_{n_R} \triangleq [h_{1n_R}[0]\; h_{1n_R}[1]\; \ldots\; h_{N_T n_R}[L-1]]^T$ and $\boldsymbol{n}_{n_R} \triangleq [n_{n_R}[0]\; n_{n_R}[1]\; \ldots\; n_{n_R}[K+L_{\text{eq}}-1]]^T$ are used. The $(K+L_{\text{eq}}-1)\times L_{\text{eq}}$ data matrix $\boldsymbol{B}$ contains the vector $[b[k]\; b[k-1]\; \ldots\; b[k-(L_{\text{eq}}-1)]]$ in row k, $0 \leq k \leq K+L_{\text{eq}}-2$ ($b[k]=0$ for $k<0$ and $k \geq K$). The $L_{\text{eq}} \times LN_T$ filter matrix $\boldsymbol{G}$ is defined as

$$\boldsymbol{G} \triangleq [\boldsymbol{G}_1\; \boldsymbol{G}_2\; \ldots\; \boldsymbol{G}_{N_T}], \qquad (7)$$

$$\boldsymbol{G}_{n_T} \triangleq \begin{pmatrix} \boldsymbol{g}_{n_T} & 0 & \cdots & 0 \\ 0 & \boldsymbol{g}_{n_T} & \ddots & 0 \\ \vdots & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \vdots & \boldsymbol{g}_{n_T} \end{pmatrix}, \qquad (8)$$

$$\boldsymbol{g}_{n_T} \triangleq [g_{n_T}[0]\; g_{n_T}[1]\; \ldots\; g_{n_T}[N-1]]^T. \qquad (9)$$

Now, we stack the received vectors of different antennas into one vector and obtain

$$\boldsymbol{r} = \sqrt{\frac{E_S}{N_T}}(\boldsymbol{I}_{N_R} \otimes \boldsymbol{B}\boldsymbol{G})\boldsymbol{h} + \boldsymbol{n}, \qquad (10)$$

where the definitions $\boldsymbol{r} \triangleq [\boldsymbol{r}_1^T\; \boldsymbol{r}_2^T\; \ldots\; \boldsymbol{r}_{N_R}^T]^T$ and $\boldsymbol{n} \triangleq [\boldsymbol{n}_1^T\; \boldsymbol{n}_2^T\; \ldots\; \boldsymbol{n}_{N_R}^T]^T$ are used. $\boldsymbol{I}_X$ denotes the $X \times X$ identity matrix and $\otimes$ is the Kronecker product [17].

In the following, we assume that a sequence $b_\alpha[\cdot] \in \mathcal{A}^K$ is transmitted but a different sequence $b_\beta[\cdot] \in \mathcal{A}^K$ is detected. The corresponding data matrices are denoted by $\boldsymbol{B}_\alpha$ and $\boldsymbol{B}_\beta$ ($\boldsymbol{B}_\beta \neq \boldsymbol{B}_\alpha$), respectively. Using Eq. (10) and similar techniques as in [18], it can be shown that the PEP for MLSE can be upper bounded as

$$\begin{aligned} P_e &= \Pr\Big\{||\boldsymbol{r} - \sqrt{\frac{E_S}{N_T}}(\boldsymbol{I}_{N_R} \otimes \boldsymbol{B}_\alpha\boldsymbol{G})\boldsymbol{h}||^2 > \\ &\qquad ||\boldsymbol{r} - \sqrt{\frac{E_S}{N_T}}(\boldsymbol{I}_{N_R} \otimes \boldsymbol{B}_\beta\boldsymbol{G})\boldsymbol{h}||^2\Big\} \\ &\leq \prod_{q=1}^{L_{\text{eq}}N_R} \frac{1}{1+\lambda_q(\boldsymbol{Z})E_S/(4N_TN_0)}, \end{aligned} \qquad (11)$$

where $\Pr\{\cdot\}$ and $||\cdot||^2$ denote the probability of the event in brackets and the L_2 norm, respectively. In this paper, $\lambda_q(\boldsymbol{X})$ denotes the eigenvalues of matrix $\boldsymbol{X}$, which are assumed to be real–valued and ordered as $\lambda_1(\boldsymbol{X}) \geq \lambda_2(\boldsymbol{X}) \geq \ldots$. In Eq. (11), the $(K+L_{\text{eq}}-1)N_R \times (K+L_{\text{eq}}-1)N_R$ matrix $\boldsymbol{Z}$ is given by

$$\boldsymbol{Z} \triangleq (\boldsymbol{I}_{N_R} \otimes \boldsymbol{B}_{\beta|\alpha}\boldsymbol{G})\boldsymbol{\Phi}(\boldsymbol{I}_{N_R} \otimes \boldsymbol{B}_{\beta|\alpha}\boldsymbol{G})^H, \qquad (12)$$

where the difference sequence matrix is defined as $\boldsymbol{B}_{\beta|\alpha} \triangleq \boldsymbol{B}_\beta - \boldsymbol{B}_\alpha$. Note that $\boldsymbol{Z}$ has at most $L_{\text{eq}}N_R$ non–zero eigenvalues. From Eqs. (11) and (12) it can be observed that the maximum possible *diversity order* LN_TN_R can be achieved only if the CIR ACM $\boldsymbol{\Phi}$ has full rank and if the FIR filter length is chosen as proposed in [1], i.e., $N=(N_T-1)L+1$. On the other hand, in practice $N<(N_T-1)L+1$ might be desirable for complexity reasons and it is not clear if the specific choice of $\boldsymbol{G}$ in [1] minimizes the upper bound on the PEP.

In order to minimize the upper bound in Eq. (11), we have to maximize the *cost function*

$$d(\boldsymbol{g}, \boldsymbol{B}_{\beta|\alpha}) \triangleq \prod_{q=1}^{L_{\text{eq}}N_R}\left(1+\frac{E_S}{4N_TN_0}\lambda_q(\boldsymbol{Z})\right), \qquad (13)$$

where the FIR coefficient vector of length NN_T is defined as $\boldsymbol{g} \triangleq [\boldsymbol{g}_1^T\; \boldsymbol{g}_1^T\; \ldots\; \boldsymbol{g}_{N_T}^T]^T$. In order to minimize the maximum PEP, we first find the optimum $\boldsymbol{g}_{\text{opt}}^{\beta|\alpha}$ for all possible difference matrices $\boldsymbol{B}_{\beta|\alpha}$. For the sake of mathematical tractability we assume that the *dominant* error event is independent of $\boldsymbol{g}$. In this case, the overall optimum vector $\boldsymbol{g}_{\text{opt}}$ is the one that is associated with that difference matrix $\boldsymbol{B}_{\beta_0|\alpha_0}$ which yields minimum $d(\boldsymbol{g}_{\text{opt}}^{\beta|\alpha}, \boldsymbol{B}_{\beta|\alpha})$, since the sequences $b_{\alpha_0}[\cdot]$ and $b_{\beta_0}[\cdot]$ correspond to the dominant error event.

We could reduce the complexity of the optimization problem significantly if we knew in advance, which error event is dominant. On the other hand, since it is well known that, in general, for frequency–selective fading channels the performance of MLSE can be tightly approximated by the matched–filter bound [19], it seems to be reasonable to assume that for most channels the dominant error event corresponds to sequences that differ only by the minimum Euclidean distance. For those sequences $b_\alpha[k]=b_\beta[k]$, $0 \leq k \leq K-1$, $k \neq k_0$, and $|b_\alpha[k_0]-b_\beta[k_0]|^2 = d_{\text{min}}^2$, where d_{min}^2 denotes the squared minimum Euclidean distance of the adopted signal constellation $\mathcal{A}$. With this assumption, Eq. (13) can be simplified to

$$d(\boldsymbol{g}) = \prod_{q=1}^{L_{\mathrm{eq}}N_R} \left(1 + \gamma\lambda_q\left((\boldsymbol{I}_{N_R} \otimes \boldsymbol{G})\boldsymbol{\Phi}(\boldsymbol{I}_{N_R} \otimes \boldsymbol{G})^H\right)\right)$$
$$= \det\left(\boldsymbol{I}_{L_{\mathrm{eq}}N_R} + \gamma(\boldsymbol{I}_{N_R} \otimes \boldsymbol{G})\boldsymbol{\Phi}(\boldsymbol{I}_{N_R} \otimes \boldsymbol{G})^H\right), \quad (14)$$

where the *effective* SNR γ is defined as $\gamma \triangleq d_{\min}^2 E_S/(4N_T N_0)$ and $\det(\cdot)$ denotes the determinant of a matrix. Although the optimization procedure proposed in the following sections can be easily extended to the more general cost function $d(\boldsymbol{g}, \boldsymbol{B}_{\beta|\alpha})$, for complexity reasons and practical importance we restrict our attention to the simpler cost function $d(\boldsymbol{g})$ in Eq. (14), which does not depend on the specific signal constellation $\mathcal{A}$ used, but only on the minimum Euclidean distance $d_{\min}$ of that constellation. Note that unlike most other papers on TD design, cf. e.g. [18], [20], [21], [22], we do not make the high SNR assumption ($\gamma \gg 1$) in Eq. (14). We will show that ODD filters designed for $\gamma \gg 1$ do not perform well at SNRs of practical interest.

3.2 Optimization Problem

The optimum filter coefficients are obtained by maximizing the cost function $d(\boldsymbol{g})$ subject to the energy constraint $\boldsymbol{g}^H\boldsymbol{g} = N_T$. In an attempt to arrive at a closed–form solution we may define the Lagrange cost function [17]

$$J(\boldsymbol{g}) \triangleq d(\boldsymbol{g}) + \nu_0 \cdot (\boldsymbol{g}^H\boldsymbol{g} - N_T), \quad (15)$$

where ν_0 denotes the Lagrange multiplier. In order to find the optimum coefficient vector $\boldsymbol{g}_{\mathrm{opt}}$, we have to differentiate $J(\boldsymbol{g})$ with respect to $\boldsymbol{g}^*$. The gradient vector of $d(\boldsymbol{g})$ is given by [23]

$$\frac{\partial d(\boldsymbol{g})}{\partial \boldsymbol{g}^*} = \left(\frac{\partial d(\boldsymbol{g})}{\partial g_1[0]}\ \frac{\partial d(\boldsymbol{g})}{\partial g_1[1]}\ \cdots\ \frac{\partial d(\boldsymbol{g})}{\partial g_{N_T}[N-1]}\right)^H \quad (16)$$

$$\frac{\partial d(\boldsymbol{g})}{\partial g_{n_T}[\nu]} = \gamma d(\boldsymbol{g}) \cdot \mathrm{tr}\Big((\boldsymbol{I}_{L_{\mathrm{eq}}N_R} + \gamma\bar{\boldsymbol{G}}\boldsymbol{\Phi}\bar{\boldsymbol{G}}^H)^{-1} \cdot \bar{\boldsymbol{E}}_{n_T\nu}\boldsymbol{\Phi}\bar{\boldsymbol{G}}^H\Big) \quad (17)$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix and the bar over a matrix $\boldsymbol{X}$ means $\bar{\boldsymbol{X}} \triangleq \boldsymbol{I}_{N_R} \otimes \boldsymbol{X}$. The elements of the $L_{\mathrm{eq}} \times LN_T$ matrix $\boldsymbol{E}_{n_T\nu}$ are 1 at those positions where matrix $\boldsymbol{G}$ has entries $g_{n_T}[\nu]$ (cf. Eq. (8)) and zero otherwise.

Eqs. (16) and (17) show that the system of equations that has to be solved for direct maximization of $J(\boldsymbol{g})$ is highly nonlinear in $\boldsymbol{g}$. Consequently, it seems to be not possible to find a closed–form solution for the optimum vector $\boldsymbol{g}_{\mathrm{opt}}$. Therefore, in the next section, we derive an SD algorithm [24] to maximize $J(\boldsymbol{g})$ recursively.

3.3 Steepest Descent (SD) Algorithm

Since we have a closed–form expression for the gradient vector $\partial d(\boldsymbol{g})/(\partial \boldsymbol{g}^*)$, an SD algorithm for maximization of $J(\boldsymbol{g})$ can be obtained. If we denote the number of iterations by i, $i \in \{0, 1, \ldots\}$, the SD algorithm is given by

$$\boldsymbol{g}_0[i+1] = \boldsymbol{g}[i] + \mu_0 \frac{\partial d(\boldsymbol{g}[i])}{\partial \boldsymbol{g}^*[i]} \quad (18)$$

$$\boldsymbol{g}[i+1] = \sqrt{\frac{N_T}{\boldsymbol{g}_0^H[i+1]\boldsymbol{g}_0[i+1]}}\, \boldsymbol{g}_0[i+1], \quad (19)$$

where μ_0 is a small positive adaptation constant. The ODD filter coefficients used in Sections 4 and 5 were obtained with $\mu_0 = 0.01$. Eq. (19) ensures that the constraint in Eq. (15) is fulfilled. Unfortunately, the cost function $J(\boldsymbol{g})$ has local maxima. Therefore, a careful initialization of the SD algorithm is important for finding the global maximum or at least a large local maximum of $J(\boldsymbol{g})$. Since we use the simple GDD scheme as a benchmark for the proposed ODD scheme, for the sake of comparison we will restrict our attention to the case $N = (N_T - 1)L_0 + 1$, $0 \leq L_0 \leq L$, and initialize $\boldsymbol{g}[i]$ with the corresponding GDD filter coefficients, cf. Section 2.1. Extensive simulations have shown that this is a good choice. Indeed, testing a variety of different channels, we could not find an example where another (randomly chosen) initialization would lead to a coefficient vector $\boldsymbol{g}$ that yielded a larger value of the respective cost function than the coefficient vector obtained with the GDD initialization.

We note that using a similar method as in [23] it is possible to derive a closed–form expression for the gradient of $d(\boldsymbol{g}, \boldsymbol{B}_{\beta|\alpha})$ defined in Eq. (13). Therefore, also in this case, filter optimization can be performed using an SD algorithm.

4 Performance Comparison

Since we cannot guarantee that the FIR filter coefficients found with the SD algorithm proposed in the last section correspond to the global (constrained) maximum of the cost function $d(\boldsymbol{g})$, in this section we derive an upper bound for $d(\boldsymbol{g}_{\mathrm{opt}})$. The maximum of the cost function for the filter coefficients found by our search can then be compared with this upper bound. We show that under certain conditions GDD is indeed optimum, i.e., the GDD filter coefficients achieve the upper bound. However, in most cases of practical interest, the ODD filter coefficients yield a significantly better performance.

4.1 Upper Bound on $d(\boldsymbol{g})$

The main difficulty in finding the maximum of $d(\boldsymbol{g})$ is the special structure of matrix $\bar{\boldsymbol{G}}$. Therefore, in the following, we consider unstructured matrices $\bar{\boldsymbol{G}}$ for maximization of the cost function. The resulting maximum is only an upper bound for the maximum achievable with structured matrices $\bar{\boldsymbol{G}}$ and the corresponding optimum matrix $\bar{\boldsymbol{G}}_{\text{opt}}$, in general, does not correspond to realizable FIR transmit filters. Using this approach, we show in the full paper [23] that $d(\boldsymbol{g})$ can be upper bounded as

$$d(\boldsymbol{g}_{\text{opt}}) \le \left(\frac{LN_TN_R}{Q}\gamma\right)^Q \prod_{q=1}^{Q}\lambda_q(\boldsymbol{\Phi}) \left(1+\frac{1}{\gamma LN_TN_R}\sum_{q=1}^{Q}\frac{1}{\lambda_q(\boldsymbol{\Phi})}\right)^Q, \quad (20)$$

where Q is the maximum integer $1 \le Q \le L_{\text{eq}}N_R$ that satisfies $\sigma_Q^2(\bar{\boldsymbol{G}}_{\text{opt}}) > 0$ with $\sigma_Q^2(\bar{\boldsymbol{G}}_{\text{opt}})$ given by

$$\sigma_Q^2(\bar{\boldsymbol{G}}_{\text{opt}}) = \frac{1}{Q}\left(LN_TN_R + \frac{1}{\gamma}\sum_{i=1}^{Q}\frac{1}{\lambda_i(\boldsymbol{\Phi})}\right) - \frac{1}{\gamma\lambda_Q(\boldsymbol{\Phi})}. \quad (21)$$

Regarding the upper bound in Eq. (20), we make the following interesting observations.

a) If all eigenvalues of the CIR ACM $\boldsymbol{\Phi}$ are larger than zero, for high SNRs ($\gamma \gg 1$) $Q = L_{\text{eq}}N_R$ is valid, cf. Eq. (21). Therefore, the right hand side of Eq. (20) is approximately proportional to $\gamma^{L_{\text{eq}}N_R}$, which suggests that exploiting the maximum available diversity order $L_{\text{eq}}N_R$ is the optimum strategy for high SNRs.

b) If not all eigenvalues of matrix $\boldsymbol{\Phi}$ are identical, which will be the case for most practical frequency–selective channels, for small–to–medium SNRs it can be inferred from Eq. (21) that $Q < L_{\text{eq}}N_R$ is optimum. This suggests that under these circumstances FIR transmit filters that do not exploit the maximum available diversity order may be preferable.

Considering the above discussion, the right hand side of Eq. (20) suggests that FIR transmit filters that are optimum for high SNRs may be suboptimum for SNRs of practical interest. This is supported by the numerical and simulation results presented in Sections 4.3 and 5, respectively.

4.2 Special Case: $N = (N_T-1)L+1$ and $\gamma\lambda_{LN_TN_R} \gg 1$

If we assume GDD with $N = (N_T-1)L+1$, we obtain from Eq. (7) $\boldsymbol{G} = \boldsymbol{I}_{LN_T}$. Consequently, Eq. (14) yields $d(\boldsymbol{g}) = \det(\boldsymbol{I}_{LN_TN_R} + \gamma\boldsymbol{\Phi})$. Thus, assuming $\lambda_q(\boldsymbol{\Phi}) > 0$, $1 \le q \le LN_TN_R$, and a large enough SNR to assure $\gamma\lambda_{LN_TN_R} \gg 1$, we obtain

$$d(\boldsymbol{g}) = \prod_{q=1}^{LN_TN_R}(1+\gamma\lambda_q(\boldsymbol{\Phi})) \approx \gamma^{LN_TN_R}\prod_{q=1}^{LN_TN_R}\lambda_q(\boldsymbol{\Phi}) = \gamma^{LN_TN_R}\det(\boldsymbol{\Phi}) \quad (22)$$

On the other hand, making the same assumptions, from Eq. (21) $Q = LN_TN_R$ results. Therefore, we obtain for the upper bound in Eq. (20)

$$d(\boldsymbol{g}_{\text{opt}}) \le \gamma^{LN_TN_R}\left(1+\frac{\text{tr}(\boldsymbol{\Phi}^{-1})}{\gamma LN_TN_R}\right)^{LN_TN_R} \cdot \det(\boldsymbol{\Phi}), \lesssim \gamma^{LN_TN_R}\det(\boldsymbol{\Phi}). \quad (23)$$

Obviously GDD as proposed in [1] achieves the upper bound for $d(\boldsymbol{g})$ for high SNRs. Therefore, in the considered special case GDD is indeed optimum with respect to the optimization criterion adopted in this paper. This result holds for arbitrary (non–singular) correlation matrices $\boldsymbol{\Phi}$, i.e., for arbitrary spatial and temporal correlation of the CIR coefficients. Note that the asymptotic optimality of GDD was not proved in [1].

On the other hand, as long as not all eigenvalues of $\boldsymbol{\Phi}$ are equal, for small–to–moderate SNRs $Q < LN_TN_R$ is valid and GDD is no longer optimum.

4.3 Example for $N < (N_T-1)L+1$

In this section, we compare GDD and ODD for $N_R = 1$ receive antenna, $N_T = 2$ transmit antennas with a spatial correlation of $\rho_{12} = 0.5$, and the EQ profile with $L = 7$. Our comparison is based on $P_d \triangleq 1/d(\boldsymbol{g})$ instead of $d(\boldsymbol{g})$ itself, since P_d is related to the PEP for MLSE, cf. Eqs. (11)–(14). The upper bound on $d(\boldsymbol{g})$ in Eq. (20) corresponds to a lower bound on P_d.

Fig. 2 shows P_d vs. $10\log_{10}(\gamma)$ for GDD and ODD with $N = 3$, respectively. For ODD the FIR transmit filters were optimized for each γ–value that is marked by a "$\circ$" using the SD algorithm described in Section 3.3. Also shown in Fig. 2 is the lower bound on P_d. Obviously, ODD significantly outperforms GDD, e.g. at $P_d = 10^{-3}$ the performance gain of ODD over GDD is 2.2 dB, whereas the remaining gap to the lower bound is 1.5 dB. In order to better understand why ODD achieves large gains for low–to–moderate SNRs, the optimized FIR filter coefficients for different values of γ are shown in Table I. Although the ODD FIR filters change with increasing γ, for -3 dB $\le 10\log_{10}(\gamma) \le 7$ dB, the filters for both transmit antennas are *identical*. It is easy to show that this means that unlike GDD the optimized TD scheme has only a diversity order of 7 and does not exploit the maximum possible diversity order of $L_{\text{eq}} = 9$. Interestingly, for calculation of the lower bound we also obtained $Q = 7$ in Eq. (20) for that range of γ. For $10\log_{10}(\gamma) > 7$ dB the two ODD transmit filters are different and ODD makes use of the full available diversity of the channel. From Table I we also observe that for $10\log_{10}(\gamma) = 47$ dB the ODD FIR filter coefficients are practically identical to the GDD coefficients. This example clearly shows that, in contrast to the frequency–nonselective case [18], [20], for frequency–selective channels and SNRs of practical interest diversity order should not be the primary TD design criterion. The reason for this difference is the inherent frequency diversity of frequency–selective channels.

In Fig. 3a) and b), P_d vs. N is depicted for $10\log_{10}(\gamma) = 7$ dB and $10\log_{10}(\gamma) = 47$ dB, respectively. For $10\log_{10}(\gamma) = 47$ dB the performance of GDD is identical to that of ODD for $N \geq 3$ and for $N \geq 5$ GDD achieves the lower bound. On the other hand, for the practically relevant case of $10\log_{10}(\gamma) = 7$ dB ODD achieves considerable gains over GDD. Note that increasing the FIR filter length N does not necessarily improve the performance of GDD. In fact, for GDD and $10\log_{10}(\gamma) = 7$ dB $N = 1$ is optimum and outperforms GDD with $N = L + 1 = 8$.[1] Fig. 3a) also shows that even if transmit filters of maximum length $N = L + 1 = 8$ are used, the GDD scheme in [1] is not optimum for finite SNRs.

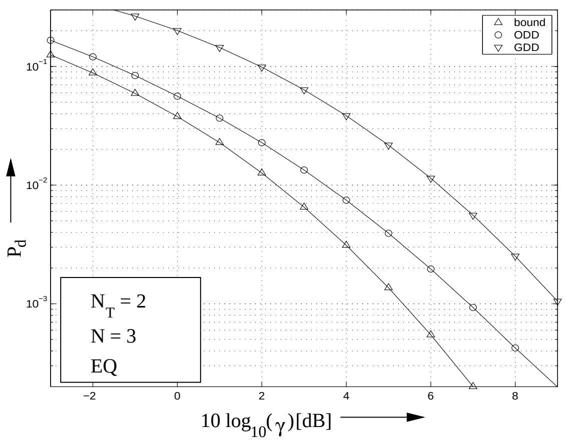

Fig. 2. P_d vs. $10\log_{10}(\gamma)$ for EQ profile with $L = 7$ and $N_T = 2$ transmit antennas with $\rho_{12} = 0.5$. ODD is compared with GDD and the derived bound.

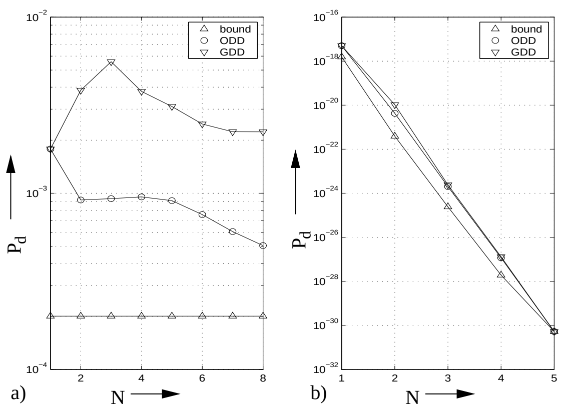

Fig. 3. P_d vs. N for a) $10\log_{10}(\gamma) = 7$ dB, and b) $10\log_{10}(\gamma) = 47$ dB. The EQ profile with $L = 7$ and $N_T = 2$ transmit antennas with $\rho_{12} = 0.5$ is adopted. ODD is compared with GDD and the derived bound.

[1]Note that $N = L + 1 = 8$ is the value proposed in [1].

TABLE I
ODD FIR TRANSMIT FILTER COEFFICIENTS FOR EQ PROFILE WITH $L = 7$ AND $N_T = 2$ TRANSMIT ANTENNAS WITH $\rho_{12} = 0.5$. $N = 3$ IS VALID.

$10\log_{10}(\gamma)$ [dB]	n_T	$g_{n_T}[0]$	$g_{n_T}[1]$	$g_{n_T}[2]$
-3	1	0.4501	0.7712	0.4501
	2	0.4501	0.7712	0.4501
-1	1	0.4256	0.7986	0.4256
	2	0.4256	0.7986	0.4256
1	1	0.1713	0.7770	0.6057
	2	0.1713	0.7770	0.6057
3	1	-0.0036	0.7053	0.7089
	2	-0.0036	0.7053	0.7089
5	1	-0.1031	0.6499	0.7530
	2	-0.1031	0.6499	0.7530
7	1	-0.1704	0.6063	0.7767
	2	-0.1704	0.6063	0.7767
9	1	0.0085	0.5525	0.8335
	2	-0.4999	0.3504	0.7921
17	1	0.8236	-0.1176	-0.5548
	2	0.8972	0.0409	0.4396
27	1	0.9503	-0.1245	-0.2854
	2	0.6238	-0.0513	0.7799
37	1	0.9470	-0.1566	-0.2041
	2	0.5960	-0.0535	0.8012
47	1	0.9904	-0.0914	-0.1037
	2	-0.1037	-0.0914	0.9904

5 Simulation Results

In this section, we present some simulation results for GDD and ODD. We only consider the practically most important case of $N_R = 1$ receive antenna. For equalization we adopt the same schemes as commonly used in the single transmit antenna case, i.e., we employ MLSE with Viterbi decoding [16], DFSE [7], [12], and DFE [8]. We assume perfect knowledge of the CIR at the receiver, and for ODD we also assume perfect knowledge of the CIR ACM at the transmitter. The ODD FIR filters are designed for a certain target E_b/N_0 ratio, i.e., they are not optimized for each SNR value. We use the channel model described in Section 2.2 and for each BER curve at least 10000 CIRs have been randomly generated in accordance with the respective power delay profile. In order to closely approximate the GSM and EDGE system parameters, respectively, we adopt binary PSK (BPSK) and 8–ary PSK (8PSK) modulation in the following.

5.1 BPSK Transmission

In Fig. 4, we consider the EQ channel ($L = 7$) and $N_T = 2$ transmit antennas with $\rho_{12} = 0.5$. We show the performance of ODD and GDD with $N = 3$, respectively. The ODD FIR filters were optimized for $10\log_{10}(E_b/N_0) = 10$ dB (corresponding to $10\log_{10}(\gamma) = 7$ dB). The same filters were used for all simulation points and their coefficients can be found in Table I. Obviously, ODD outperforms GDD for all considered equalization schemes. For $10\log_{10}(E_b/N_0) \approx 10$ dB, ODD gains approximately 1.4 dB, 1.4 dB, and 0.9 dB for MLSE, DFSE with $S = 4$ states, and DFE, respectively. We observe that the gain expected for

MLSE is also retained for suboptimum low–complexity equalization schemes. In addition, it is interesting to note that ODD also outperforms GDD for E_b/N_0 ratios different from the target value of 10 dB. From Fig. 3b) we know however that GDD is optimum for very high SNRs. This shows again that a TD design based on the assumption $\gamma \to \infty$ is not advisable for frequency–selective channels. For comparison in Fig. 4 also the performance of GDD with $N = L+1 = 8$ as proposed in [1] is depicted. For $10\log_{10}(E_b/N_0) > 8$ dB GDD with $N = 8$ outperforms GDD with $N = 3$, but has a worse performance than ODD with $N = 3$. Note that the Viterbi algorithm for $N = 8$ requires $S = 2^{13} = 8192$ states, whereas for $N = 3$ only $S = 2^8 = 256$ states are necessary.

In Fig. 5, the TU channel ($L = 5$) and $N_T = 3$ transmit antennas with spatial correlations $\rho_{12} = \rho_{23} = 0.5$ and $\rho_{13} = 0.2$ are assumed. Again, the ODD FIR filters optimized for $10\log_{10}(E_b/N_0) = 10$ dB are used for all considered E_b/N_0 ratios. Similar observations as in Fig. 4 can be made. In particular, for E_b/N_0 ratios around 10 dB the performance gain of ODD over GDD is approximately 0.95 dB, 0.8 dB, and 0.8 dB for MLSE, DFSE with $S = 4$ states, and DFE, respectively. For comparison also the performance of single antenna transmission ($N_T = 1$) with MLSE is shown. Both GDD and ODD yield large performance gains over the single antenna scheme, especially at high SNRs.

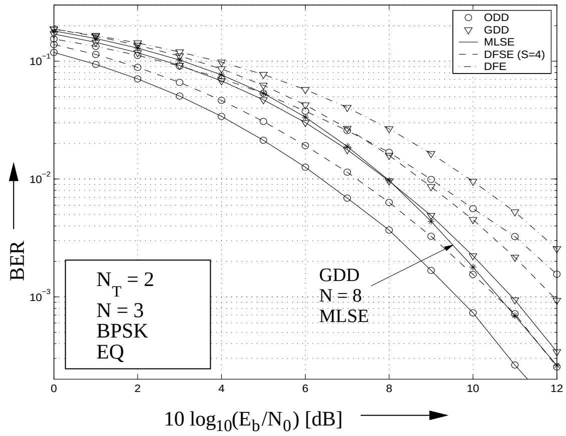

Fig. 4. BER of BPSK vs. $10\log_{10}(E_b/N_0)$ for EQ profile with $L = 7$ and $N_T = 2$ transmit antennas with $\rho_{12} = 0.5$. Simulation results for ODD (optimized for $10\log_{10}(E_b/N_0) = 10$ dB) and GDD (both with $N = 3$) are shown for MLSE, DFSE with $S = 4$ states, and DFE. For comparison also GDD with $N = 8$ and MLSE is included.

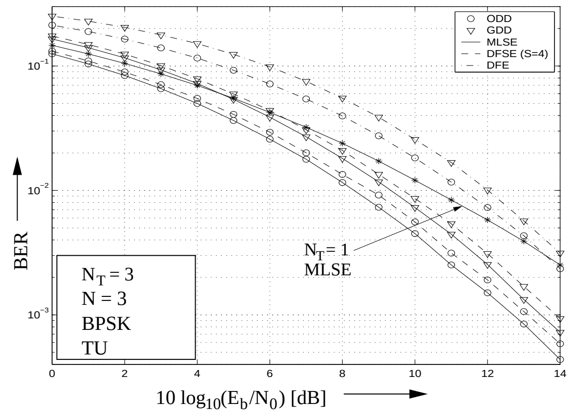

Fig. 5. BER of BPSK vs. $10\log_{10}(E_b/N_0)$ for TU profile with $L = 5$ and $N_T = 3$ transmit antennas with $\rho_{12} = \rho_{23} = 0.5$ and $\rho_{13} = 0.2$. Simulation results for ODD (optimized for $10\log_{10}(E_b/N_0) = 10$ dB) and GDD (both with $N = 3$) are shown for MLSE, DFSE with $S = 4$ states, and DFE. For comparison also the performance for $N_T = 1$ and MLSE is included.

5.2 8PSK Transmission

Finally, in Fig. 6, we consider the performance of 8PSK for the EQ channel ($L = 7$) and $N_T = 3$ transmit antennas with $\rho_{12} = \rho_{13} = \rho_{23} = 0.5$. Again $N = 3$ is adopted and the ODD FIR filters were optimized for $10\log_{10}(E_b/N_0) = 16$ dB. Since MLSE would require $S = 8^8 = 16777216$ states, we only show the results for suboptimum equalization schemes. For E_b/N_0 ratios around 16 dB the performance gain of ODD over GDD is approximately 2.5 dB and 1.9 dB for DFSE with $S = 64$ states and DFE, respectively. For comparison also the performance for a single transmit antenna and DFSE with $S = 64$ states is depicted in Fig. 6. Interestingly, in the considered SNR range the scheme with $N_T = 1$ has a better performance than GDD with $N_T = 3$ transmit antennas. In addition, for 10 dB $\leq 10\log_{10}(E_b/N_0) \leq 17$ dB the BER curves for $N_T = 1$ and GDD and ODD with $N_T = 3$, respectively, are practically parallel. This clearly shows that for this example and SNRs of practical interest diversity order is not important.

Further simulations not presented here have shown that the performance gain achieved by ODD is usually large for strong spatial correlations and large inherent frequency diversity of the channel.

6 Conclusions

In this paper, we have shown that DD is an attractive TD technique since the same receivers as for single antenna transmission can be adopted, and therefore, existing systems such as GSM and EDGE can be easily upgraded. In particular, based on the PEP of MLSE we have derived a cost function for optimization of the DD FIR transmit filters. Since a closed–form maximization of this cost function seems to be not tractable, we have provided an SD algorithm for filter optimization. Based on an upper bound on the cost function, we have proved that the GDD scheme in [1] is optimum for very high SNRs and FIR filters of maximum length. However, for SNRs of practical interest and reasonable FIR filter lengths, the proposed ODD scheme yields significant performance gains over GDD. Fortunately, these performance gains cannot only

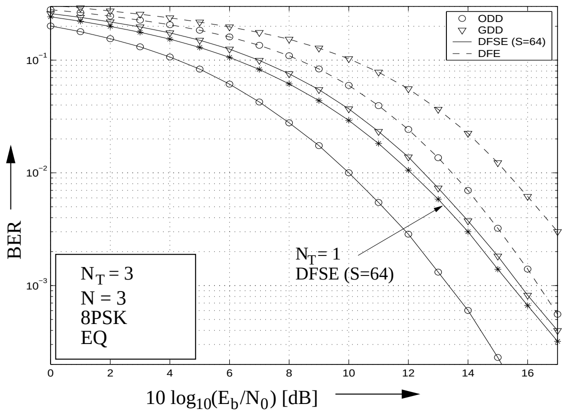

Fig. 6. BER of 8PSK vs. $10\log_{10}(E_b/N_0)$ for EQ profile with $L = 7$ and $N_T = 3$ transmit antennas with $\rho_{12} = \rho_{13} = \rho_{23} = 0.5$. Simulation results for ODD (optimized for $10\log_{10}(E_b/N_0) = 16$ dB) and GDD (both with $N = 3$) are shown for DFSE with $S = 64$ states and DFE. For comparison also the performance for $N_T = 1$ and DFSE with $S = 64$ states is included.

be observed for optimum MLSE but are also preserved for good suboptimum equalization schemes such as DFSE and DFE.

Finally, our investigations have shown that for TD design for frequency–selective channels and relevant SNR values the diversity order is not the most important parameter. Therefore, unlike the frequency–nonselective case [18], for frequency–selective channels TD schemes optimized for high SNRs may perform poorly for low to moderate SNRs.

References

[1] D. Gore, S. Sandhu, and A. Paulraj. Delay Diversity Codes for Frequency Selective Channels. In *Proceedings of the IEEE International Conference on Communications (ICC)*, pages 1949–1953, New York, May 2002.

[2] H. Bölcskei and A.J. Paulraj. Space–Frequency Coded Broadband OFDM Systems. In *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, Chicago (IL), September 2000.

[3] E. Lindskog and A. Paulraj. A Transmit Diversity Scheme for Channels with Intersymbol Interference. In *Proceedings of IEEE International Conference on Communications (ICC)*, pages 307–311, New Orleans, 2000.

[4] S. Zhou and G.B. Giannakis. Single–Carrier Space–Time Block–Coded Transmission Over Frequency–Selective Fading Channels. *IEEE Transactions on Information Theory*, IT-49:164–179, January 2003.

[5] N. Al-Dhahir. Overview and Comparison of Equalization Schemes for Space–Time–Coded Signals With Application to EDGE. *IEEE Transactions on Signal Processing*, SP:50:2477–2488, October 2002.

[6] A. Wittneben. A New Bandwidth Efficient Transmit Antenna Modulation Diversity Scheme for Linear Digital Modulation. In *Proceedings of the International Conference on Communications (ICC)*, pages 1630–1634, Geneva, May 1993.

[7] A. Duel-Hallen and A. C. Heegard. Delayed Decision–Feedback Sequence Estimation. *IEEE Trans. on Commun.*, COM-37:428–436, 1989.

[8] C. A. Belfiore and J. H. Park. Decision–Feedback Equalization. *Proceedings of the IEEE*, 67:1143–1156, August 1979.

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

[10] *"Tdoc SMG2 WPB 325/98", ETSI, Nov. 1998.*

[11] *GSM Recommendation 05.05: "Propagation Conditions", Vers. 5.3.0, Release 1996.*

[12] W. H. Gerstacker and R. Schober. Equalization Concepts for EDGE. *IEEE Transactions on Wireless Communications*, 1:190–199, January 2002.

[13] S.N. Crozier, D.D. Falconer, and S.A. Mahmoud. Least sum of squared errors (LSSE) channel estimation. *IEE Proceedings-F*, 138:371–378, August 1991.

[14] K.A. Hamied and G.L. Stüber. An Adaptive Truncated MLSE Receiver for Japanese Personal Digital Cellular. *IEEE Transactions on Vehicular Technology*, VT-45:41–50, February 1996.

[15] W. Gerstacker and J. Huber. Improved Equalization for GSM Mobile Communications. In *Proceedings of International Conference on Telecommunications (ICT)*, Istanbul, April 1996.

[16] G.D. Forney, Jr. Maximum–Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference. *IEEE Transactions on Information Theory*, IT-18:363–378, 1972.

[17] T.K. Moon and W.C. Stirling. *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, New York, 2000.

[18] V. Tarokh, N. Seshadri, and A.R. Calderbank. Space–Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction. *IEEE Transactions on Information Theory*, IT-44:744–765, March 1998.

[19] P. Balaban and J. Salz. Optimum Diversity Combining and Equalization in Digital Data Transmission with Applications to Cellular Mobile Radio – Part I and II. *IEEE Trans. on Commun.*, COM-40:885–907, May 1992.

[20] J.-C. Guey, M.P. Fitz, M.R. Bell, and W.-Y. Kuo. Signal Design for Transmitter Diversity Wireless Communication Systems Over Rayleigh Fading Channels. *IEEE Trans. on Commun.*, COM-47:527–537, April 1999.

[21] Y. Liu, M.P. Fitz, and O.Y. Takeshita. Space–Time Codes Performance Criteria and Design for Frequency Selective Fading Channels. In *Proceedings of IEEE International Conference on Communications (ICC)*, pages 2800–2804, Helsinki, Finland, June 2001.

[22] R. Schober, W.H. Gerstacker, and L.H.-J. Lampe. Performance Analysis and Design of STBCs for Fading ISI Channels. In *Proceedings of the International Conference on Communications (ICC)*, pages 1451–1455, May 2002.

[23] R. Schober, T. Hehn, and W. H. Gerstacker. Optimized Delay Diversity for Frequency–Selective Fading Channels. *Submitted to IEEE Trans. on Wireless Commun.*, April 2003.

[24] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Upper Saddle River, New Jersey, Third Edition, 1996.