

# Equalization and Decoding for Multiple-Input Multiple-Output Wireless Channels

**Bjørn A. Bjerke**

Qualcomm, Inc., 9 Damonmill Square, Suite 2A, Concord, MA 01742, USA  
Email: [bbjerke@qualcomm.com](mailto:bbjerke@qualcomm.com)

**John G. Proakis**

Department of Electrical and Computer Engineering, Northeastern University,  
360 Huntington Avenue, Boston, MA 02115, USA  
Email: [proakis@neu.edu](mailto:proakis@neu.edu)

Received 31 May 2001 and in revised form 11 January 2002

We consider multiple-input multiple-output (MIMO) wireless communication systems that employ multiple transmit and receive antennas to increase the data rate and achieve diversity in fading multipath channels. We begin by focusing on an uncoded system and define optimal and suboptimal receiver structures for this system in Rayleigh fading with and without intersymbol interference. Next, we consider coded MIMO systems. We view the coded system as a serially concatenated convolutional code (SCCC) in which the code and the multipath channel take on the roles of constituent codes. This enables us to analyze the performance using the same performance analysis tools as developed previously for SCCCs. Finally, we present an iterative (“turbo”) MAP-based equalization and decoding scheme and evaluate its performance when applied to a system with  $N$  transmit antennas and  $M$  receive antennas. We show that by performing recursive precoding prior to transmission, significant interleaving gains can be realized compared to systems without precoding.

**Keywords and phrases:** MIMO systems, multiple antennas, fading channels, MLSE, linear equalization, DFE, MRC, union bound, turbo equalization.

## 1. INTRODUCTION

Recently, multiple-input multiple-output (MIMO) wireless systems have attracted considerable attention in the communications community. Such systems employ multiple antennas, or antenna arrays, at both the transmitter and the receiver to enable spatial multiplexing of data and, thus, increased data rates. Traditionally, multiple antennas have been used at the receiver to provide spatial diversity and mitigate the effects of signal fading due to multipath propagation in the channel. However, recent developments in information theory have shown that by using multiple transmit and receive antennas, signal fading can in fact be turned into an advantage. With multiple antennas at both the transmitter and the receiver, spatially distributed channels can be supported simultaneously in the same frequency band, and by transmitting data in parallel through these channels the data rate can be increased. When deployed in a rich scattering environment, such systems are capable of greatly increasing the spectral efficiency over traditional single channel systems. Foschini and Gans [1] showed that the capacity of the flat MIMO Rayleigh fading channel associated with a system

with  $N$  transmit antennas and  $M \geq N$  receive antennas is given as

$$C = \log_2 (\det [\mathbf{I}_M + \rho \mathbf{H}\mathbf{H}^H]) \quad \text{bit/s/Hz}, \quad (1)$$

where  $\mathbf{I}_M$  is the  $M \times M$  identity matrix,  $\rho$  is the signal-to-noise ratio (SNR), and  $\mathbf{H}$  is the  $M \times N$  matrix whose elements  $\{h_{nm}\}$  represent the channel gains between pairs of transmit and receive antennas. The achievable data rate depends on the rank of  $\mathbf{H}$ . For large SNR and large  $N$  and  $M$ , the capacity tends to the value  $r \log_2 \rho$ , where  $r = \text{rank}(\mathbf{H})$ . When the elements of  $\mathbf{H}$  are independent and identically distributed, the rank  $r = N$ . Hence, in this ideal scenario of independent fading, the data rate grows linearly with the number of transmit antennas, which is the key observation made by Foschini and Gans. Ideally, the  $M$  receive antennas can provide  $M$ th-order diversity reception for each of the  $N$  transmitted signals in addition to whatever implicit diversity the channel has to offer. Since there is no orthogonal structure imposed on the signals by the transmitter, the received signals contain interchannel interference. The receiver must therefore be able to separate

the  $N$  signals and at the same time take advantage of the inherent signal diversity. The rule of thumb is that in order to ensure independent fading, the antennas have to be separated by at least half a wavelength at the receiver and as much as several wavelengths at an elevated transmitting base station. However, for fairly typical indoor and outdoor nonline-of-sight scattering scenarios, experiments have shown that antenna spacing has only limited impact on the capacity [2, 3, 4].

There are many similarities between today's multiple-antenna systems and other MIMO systems studied in the past. Nichols et al. [5] described adaptive detection algorithms for dual-input, dual-output systems such as dually-polarized microwave radio. These systems are often degraded by interchannel interference due to channel distortion effects. However, rather than ignoring these mutual interferences, the receiver can be designed to take advantage of them, and thus, improve performance. In [6], Salz considered the  $N$ -input,  $N$ -output linear transmission channel and developed minimum mean squared error (MMSE) transmit and receive filters. Duel-Hallen considered MIMO equalizers in the case of completely known, fixed channels in [7]. The information theory advances reported by Foschini and Gans in [1] and Raleigh and Cioffi in [8] marked the beginning of a renewed interest in MIMO systems, with focus on the significantly increased data rate capabilities of multiple-antenna systems.

The theoretical capacity discussed above can be approached by using powerful signal processing and coding techniques designed to take advantage of the particular characteristics of MIMO channels. In [9], Foschini suggested a receiver architecture referred to as BLAST (Bell Labs Layered Space-Time), in which  $M$  receive antennas are used to decouple and detect  $N$  parallel data streams. The receiver architecture exploits the distinct spatial signatures of the different data streams that arise from rich multipath propagation to separate the  $N$  channels. Laboratory experiments carried out with a simplified version of BLAST, known as V-BLAST, have demonstrated spectral efficiencies as high as 40 bit/s/Hz in an indoor slow-fading environment with negligible delay spread [10]. Several authors have since elaborated on the BLAST approach, studying the impact of providing channel information to the transmitter, the receiver, or both. BLAST-type architectures with reduced computational complexity were studied in [10, 11, 12], and optimal BLAST processing for systems operating under the influence of spatially colored multiple-access interference was considered in [13].

Coded MIMO systems have also been studied, in particular by Tarokh et al., who proposed the so-called space-time codes [14, 15] to further improve the performance. MIMO systems that employ parallel concatenated codes, also known as turbo codes, were first proposed by Stefanov and Duman in [16]. In their paper, an arbitrary turbo code was chosen and shown to yield certain gains over the space-time codes developed by Tarokh et al. Other authors have proposed iterative detection and decoding techniques inspired by the success of the turbo equalization concept developed by Douillard et al. in [17]. In particular, Bauch

and Naguib proposed iterative equalization and decoding of space-time codes in multipath channels with intersymbol interference (ISI) in [18]. Followup papers on complexity reduction using channel-shortening filters were presented by Bauch and Al-Dahir in [19, 20]. In [21], Ariyavisitakul combined iterative detection and decoding with a BLAST-type detector in an effort to limit the effects of error propagation associated with the original BLAST structure. B ro et al. proposed a similar solution that employs soft interference cancellation in [22]. Tonello discussed iterative decoding of space-time bit-interleaved coded modulation in [23].

In this paper, we also consider coded MIMO systems and discuss several equalization and decoding schemes which are appropriate for such systems. In the first half of the paper, we focus on equalizers for uncoded MIMO systems and evaluate their performance analytically as well as through simulations. In the second half of the paper, we analyze coded MIMO systems and propose an iterative equalization and decoding scheme. The paper is organized as follows. We begin in Section 2 by presenting discrete-time system and channel models for MIMO systems in multipath and flat Rayleigh fading. The optimal MIMO detector is presented in Section 3, while linear and decision feedback detectors are considered in Section 4. These are MIMO versions of the conventional single-input, single-output maximum likelihood sequence estimator (MLSE) and linear and decision feedback equalizers, respectively. We also present a decision-directed detector for the special case of flat fading. In Section 5, we focus on the application of error-correcting codes in MIMO systems. We view the coded MIMO system as a serially concatenated code in which the code and the channel take on the roles of constituent "codes." This enables us to apply many of the same performance analysis tools that have been developed previously for serially concatenated codes. In Section 6, we consider an iterative equalization and decoding scheme based on the maximum a posteriori probability (MAP) criterion. Differential precoding is introduced to make the multipath channel appear recursive, thus enabling the receiver to benefit from interleaving gain in a similar way as is possible with serially concatenated codes. The performance of the iterative receiver is evaluated through simulations and the simulation results are compared to the analytical results developed in Section 5. Finally, in Section 7, a summary of the main results as well as some concluding remarks are given.

## 2. SYSTEM AND CHANNEL MODELS

The general configuration for a multiple-antenna communication system is shown in Figure 1. For simplicity, we assume that a binary modulation such as binary PSK or binary PAM is used. The data is encoded and interleaved by a pseudo-random interleaver of length  $K$ . A block of  $N$  coded data symbols is converted from serial to parallel and fed to  $N$  identical modulators, where each modulator is connected to a separate antenna. The transmitted signals are received by  $M$  receive antennas whose signals are fed to separate but

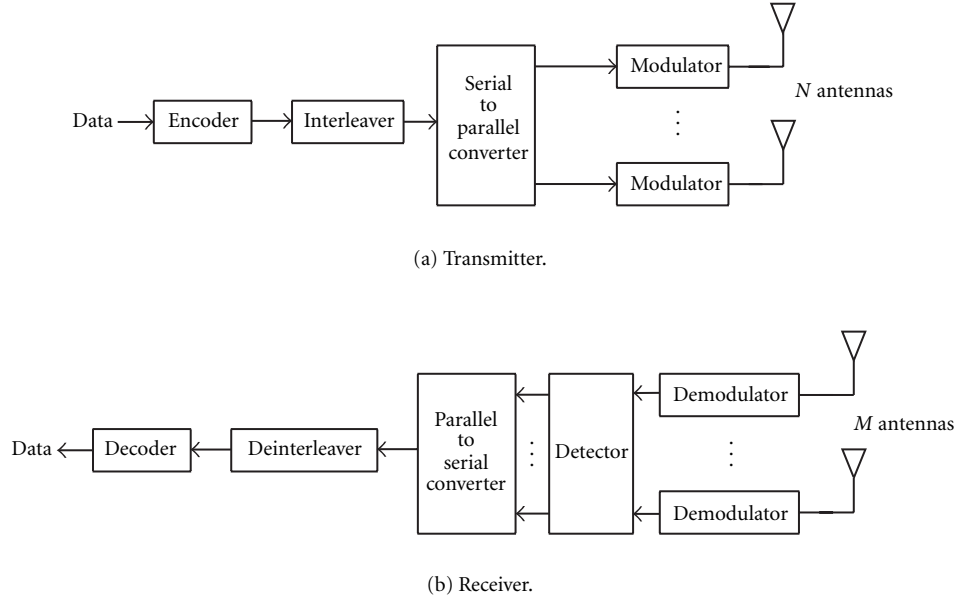


FIGURE 1: Model of MIMO digital communication system with multiple transmit and receive antennas.

identical demodulators. In the following, we will refer to such a system as an  $(N, M)$  system. The signal transmitted on the  $n$ th transmit antenna can be represented as

$$s_n(t) = \sum_k d_n(k)g(t - kT), \quad (2)$$

where  $g(t)$  is the pulse shape (impulse response) of the modulation filter,  $\{d_n(k)\}$  is the sequence of coded data symbols, and  $T$  is the symbol duration.

We assume that the channels between each transmit and receive antenna are completely known, independent time-varying multipath channels. We also assume that the differences in propagation times of the signals from the  $N$  transmit antennas to the  $M$  receive antennas are small relative to the symbol duration  $T$ , so that for practical purposes, the signals from the  $N$  transmit antennas to any receiving antenna are synchronous. The channel between each transmit and receive antenna, including transmitter and receiver filters, is modeled as a linear, discrete-time filter having a finite-duration impulse response. The tap coefficients of the equivalent discrete-time filter between the  $n$ th transmit antenna and the  $m$ th receive antenna are denoted as  $f_{nm}^{(l)}(k)$ ,  $l = 0, 1, \dots, L_{nm}$ ,  $n = 1, 2, \dots, N$ ,  $m = 1, 2, \dots, M$ . Hence, the received signals can be represented as

$$v_m(k) = \sum_{n=1}^N \sum_{l=0}^{L_{nm}} f_{nm}^{(l)}(k) d_n(k-l) + \eta_m(k), \quad m = 1, 2, \dots, M, \quad (3)$$

where  $d_n$  is the coded symbol transmitted on the  $n$ th antenna,  $L_{nm} + 1$  represents the span of the ISI on that particular discrete-time channel, and  $\eta_m(k)$  is a sample function of a zero-mean, temporally and spatially white Gaussian noise

process. For convenience, (3) may be written in matrix form as

$$\mathbf{v}(k) = \sum_{l=0}^L \mathbf{F}(k, l) \mathbf{d}(k-l) + \boldsymbol{\eta}(k), \quad (4)$$

where

$$\begin{aligned} \mathbf{v}(k) &= [v_1(k) v_2(k) \cdots v_M(k)]^T, \\ \mathbf{d}(k) &= [d_1(k) d_2(k) \cdots d_N(k)]^T, \\ \boldsymbol{\eta}(k) &= [\eta_1(k) \eta_2(k) \cdots \eta_M(k)]^T, \end{aligned} \quad (5)$$

and  $\{\mathbf{F}(k, l)\}$  is a set of channel matrices representing the equivalent discrete-time channels between the  $N$  transmit antennas and the  $M$  receive antennas. The maximal span of the ISI is given by  $L = \max\{L_{nm}\}$ . Figure 2 illustrates the model of the equivalent discrete-time system with white noise.

In the general case of multipath fading, the tap coefficients are assumed to be complex-valued, mutually statistically independent Gaussian random variables with zero mean and variance

$$\sigma_{n,m,l}^2 = E[|f_{nm}^{(l)}(k)|^2], \quad n = 1, 2, \dots, N, \quad m = 1, 2, \dots, M, \quad l = 0, 1, \dots, L_{nm}. \quad (6)$$

In the special case of frequency-nonselctive, or flat, fading, the parameter  $L = 0$  since the channel is memoryless, and (3) reduces to

$$v_m(k) = \sum_{n=1}^N f_{nm}(k) d_n(k) + \eta_m(k), \quad m = 1, 2, \dots, M, \quad (7)$$

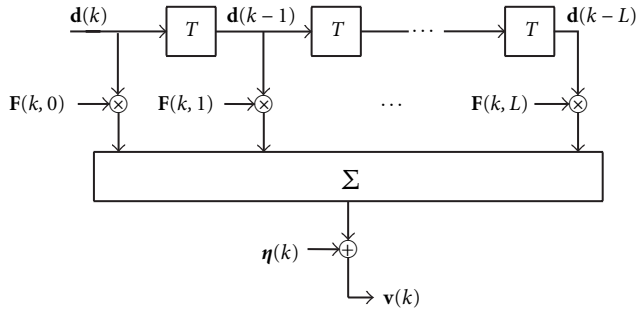


FIGURE 2: Discrete-time vector model for a time-invariant  $(N, M)$  system.

which may be conveniently represented in matrix form as

$$\mathbf{v}(k) = \mathbf{F}(k)\mathbf{d}(k) + \boldsymbol{\eta}(k). \quad (8)$$

In the general MIMO model depicted in Figure 1, we assume coded data symbols. However, in Sections 3 and 4, we will focus on equalization and detection schemes and will therefore consider an uncoded system. In Section 5, we return to coded MIMO systems and investigate receiver structures which combine the tasks of equalization, detection, and decoding.

### 3. THE MLSE FOR MIMO CHANNELS

The optimal receiver for an  $(N, M)$  MIMO multipath channel is based on joint maximum likelihood detection of the vector sequence  $\{\mathbf{d}(k)\}$ , where  $\mathbf{d}(k) = [d_1(k)d_2(k)\cdots d_N(k)]^T$  is the vector of data symbols transmitted simultaneously from antennas 1 through  $N$ . This receiver is known as the maximum likelihood sequence estimator (MLSE), and is a generalization of the well-established single-input, single-output MLSE [24]. In this section, we first describe the structure of the MIMO MLSE and then analyze its asymptotic performance on Rayleigh fading multipath channels.

#### 3.1. The MIMO MLSE

The MLSE criterion is equivalent to estimating the state sequence of a discrete-time, finite-state machine. In this case, the state machine is the equivalent discrete-time multipath channel with channel coefficients  $\{f_{nm}^{(l)}(k)\}$ . These coefficients are treated as known constants in the detection of the information sequence. The state at any time instant is given by the  $L$  most recent vector inputs, that is,

$$\mathbf{S}_k = (\mathbf{d}(k-1), \mathbf{d}(k-2), \dots, \mathbf{d}(k-L)), \quad (9)$$

where  $\mathbf{d}(k) = \mathbf{0}$  for  $k \leq 0$ . Assuming a binary modulation scheme, the state machine has  $2^{LN}$  states. Hence, the channel is described by a  $2^{LN}$ -state trellis and the vector Viterbi algorithm introduced by van Etten [25] may be used to determine the most probable path through the trellis.

With these signals as input, the MLSE decides in favor of

the vector sequence  $\{\mathbf{d}(k)\}$  that maximizes the joint conditional probability density function

$$p(\mathbf{v}_W|\mathbf{d}_W) = \prod_{k=1}^W p(\mathbf{v}(k)|\mathbf{d}(k), \mathbf{d}(k-1), \dots, \mathbf{d}(k-L)), \quad (10)$$

where  $W \gg L$  is the length of the transmitted sequence and the components  $\{v_m(k)\}$  of the vector  $\mathbf{v}(k)$  are complex-valued Gaussian random variables with mean

$$\bar{v}_m(k) = \sum_{n=1}^N \sum_{l=0}^{L_{nm}} f_{nm}^{(l)}(k)d_n(k-l) \quad (11)$$

and variance

$$\sigma^2 = E\left[|\eta_m(k)|^2\right] = N_0. \quad (12)$$

The variance is independent of  $m$ . Thus, the MLSE decides on the sequence that maximizes the joint probability density function

$$p(\mathbf{v}_W|\mathbf{d}_W) = \left(\frac{1}{2\pi\sigma^2}\right)^{MW} \times \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^W \sum_{m=1}^M |v_m(k) - \bar{v}_m(k)|^2\right) \quad (13)$$

over all  $2^{NW}$  possible sequences. Maximizing  $p(\mathbf{v}_W|\mathbf{d}_W)$  over  $\mathbf{d}_W$  is equivalent to minimizing the distance metric

$$J(\mathbf{d}_W) = \sum_{k=1}^W \sum_{m=1}^M |v_m(k) - \bar{v}_m(k)|^2. \quad (14)$$

Consequently, at each stage of the Viterbi algorithm the following incremental distance metric must be computed:

$$\mu_k(\mathbf{d}(k)) = \sum_{m=1}^M \left| v_m(k) - \sum_{n=1}^N \sum_{l=0}^{L_{nm}} f_{nm}^{(l)}(k)d_n(k-l) \right|^2. \quad (15)$$

In the case of flat fading, that is, where  $L = 0$ , the distance metric reduces to

$$\mu_k(\mathbf{d}(k)) = \sum_{m=1}^M \left| v_m(k) - \sum_{n=1}^N f_{nm}(k)d_n(k) \right|^2. \quad (16)$$

In this case, the channel is memoryless, and, assuming that the channel coefficients are statistically independent over successive symbol intervals, sequence detection is no longer necessary. Instead, the receiver performs maximum likelihood detection on a symbol-by-symbol basis, as suggested by Nichols et al. [5].

#### 3.2. Performance of the MLSE

An  $(N, M)$  system such as the one we consider here can be thought of as  $N$  parallel single transmitter systems with  $M$ th-order spatial diversity, each one experiencing multipath

fading. The error probability of one such component single transmitter system was derived by Proakis in [26]. In this section, we build upon those results as well as the results derived by van Etten for  $(N, N)$  systems in time-invariant multipath channels [25] to obtain the asymptotic bit error probability for our  $(N, M)$  system. For simplicity, we assume that BPSK modulation is employed. We also assume perfect knowledge of the channel coefficients at all times.

The expression for the upper bound on the probability of a bit error is identical to that derived by Forney [24] for the single-input, single-output MLSE on a time-invariant multipath channel, namely,

$$P_b(\gamma) \leq \sum_{\varepsilon \in E} K(\varepsilon) Q(\sqrt{2\gamma}). \quad (17)$$

The SNR parameter  $\gamma$  is given by  $\gamma = d^2(\varepsilon)/N_0$ , where  $d^2(\varepsilon)$  is the Euclidean weight of an error event  $\varepsilon$ ,  $E$  is the set of all error events starting at a given point in time, and  $K(\varepsilon)$  is a constant which is independent of  $d^2(\varepsilon)$ . The standard error function  $Q$  is given by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt, \quad x \geq 0. \quad (18)$$

An error event is defined as an estimated state sequence

$$\hat{S}_k, \hat{S}_{k+1}, \dots, \hat{S}_{k+v}, \quad (19)$$

where the first and last state estimates are correct, that is,  $\hat{S}_k = S_k$  and  $\hat{S}_{k+v} = S_{k+v}$ , but the intermediate state estimates are incorrect. Associated with the estimated state sequence is a normalized symbol error sequence  $\{\mathbf{e}(i)\}$  defined by

$$\mathbf{e}(i) = \frac{1}{2} [\mathbf{d}(i) - \hat{\mathbf{d}}(i)], \quad i = k, k+1, \dots, k+v-L-1, \quad (20)$$

where  $\mathbf{e}(i) = [e_1(i)e_2(i)\dots e_N(i)]^T$ , as well as a signal error sequence  $\{\boldsymbol{\varepsilon}(j)\}$  where the vector  $\boldsymbol{\varepsilon}(j) = [e_1(j)e_2(j)\dots e_M(j)]^T$  is given by

$$\boldsymbol{\varepsilon}(j) = \sum_{l=0}^L \mathbf{F}(j, l) \mathbf{e}(j-l), \quad j = k, k+1, \dots, k+v-1. \quad (21)$$

The Euclidean weight of the error event can be expressed as

$$\begin{aligned} d^2(\varepsilon) &= \sum_{j=k}^{k+v-1} \|\boldsymbol{\varepsilon}(j)\|^2 \\ &= \sum_{m=1}^M \left( \sum_{n=1}^N \sum_{n'=1}^N \mathbf{f}'(n, m) \mathbf{A}_{n, n'} \mathbf{f}(n', m) \right), \end{aligned} \quad (22)$$

where  $\mathbf{A}_{n, n'}$  is an  $(L+1) \times (L+1)$  positive definite matrix which depends on the symbol errors associated with the particular error event  $\varepsilon$ , and  $\mathbf{f}(n, m)$  is a vector which contains the coefficients of the channel between transmit antenna  $n$  and receive antenna  $m$ , and is given by

$$\mathbf{f}(n, m) = [f_{nm}^{(0)}(k) f_{nm}^{(1)}(k) \dots f_{nm}^{(L)}(k)]^T. \quad (23)$$

Due to the exponential behavior of the  $Q$  function, the bit error probability on a time-invariant channel as given by (17) is dominated by the term corresponding to the minimum value of  $d^2(\varepsilon)$ , which we denote by  $d_{\min}^2(\varepsilon)$ . This minimum weight is associated with the length-1 error event which consists of only one error vector  $\mathbf{e}(k)$  (i.e.,  $\mathbf{e}(j) = 0$ ,  $j \neq k$ ). Moreover, in this error vector, only one of the  $N$  components differs from zero. Thus, the minimum weight  $d_{\min}^2(\varepsilon)$  takes the form

$$d_{\min}^2(\varepsilon) = \sum_{m=1}^M \mathbf{f}'(n, m) \mathbf{A} \mathbf{f}(n, m) = \sum_{m=1}^M d_m^2(\varepsilon), \quad (24)$$

where  $\mathbf{A} = \mathbf{A}_{n, n}$  is given by (see [27])

$$\mathbf{A} = \begin{bmatrix} \beta_n(0) & \beta_n(1) & \dots & \beta_n(L) \\ \beta_n(1) & \beta_n(0) & \dots & \beta_n(L-1) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_n(L) & \beta_n(L-1) & \dots & \beta_n(0) \end{bmatrix}, \quad (25)$$

$$\beta_n(l) = \sum_{j=k}^{k+v-1-l} e_n(j) e_n(j+l), \quad l = 0, 1, \dots, L.$$

For the length-1 error event  $\mathbf{A} = \mathbf{I}$ , since  $\beta_n(l)$  takes on the following values:

$$\beta_n(l) = \begin{cases} 1, & l = 0, \\ 0, & l \neq 0. \end{cases} \quad (26)$$

For the Rayleigh fading channel, the vectors  $\{\mathbf{f}(n, m)\}$  are zero-mean, complex-valued, mutually statistically independent Gaussian vectors whose covariance matrices are diagonal matrices given by

$$\mathbf{R}_m = E[\mathbf{f}(n, m) \mathbf{f}'(n, m)], \quad m = 1, 2, \dots, M. \quad (27)$$

Consequently, the Euclidean weight  $d_{\min}^2(\varepsilon)$  is a random variable represented by the sum of quadratic forms  $d_m^2(\varepsilon)$  given in (24).

To obtain the bit error probability in Rayleigh fading, we must average the conditional error probability  $P_b(\gamma)$  over the statistics of the SNR parameter  $\gamma$ . The characteristic function of  $\gamma$  may generally be written as (see [28])

$$\psi_\gamma(j\omega) = \prod_{m=1}^M \prod_{l=0}^L \frac{1}{1 - j\omega \xi_{m,l}/N_0}, \quad (28)$$

where  $\{\xi_{m,l}\}$  are the  $(L+1)M$  eigenvalues of the matrices  $\mathbf{A} \mathbf{R}_m$ ,  $m = 1, 2, \dots, M$ . We are particularly interested in the special case where the channel coefficients  $\{f_{nm}^{(l)}(k)\}$  have equal variance  $\sigma_f^2$ , that is, where

$$\mathbf{R}_m = \sigma_f^2 \mathbf{I}, \quad m = 1, 2, \dots, M. \quad (29)$$

We define  $\bar{\gamma}_{m,l}$  as the average SNR per channel coefficient,

$$\bar{\gamma}_{m,l} = \frac{\xi_{m,l}}{N_0} = \frac{\sigma_f^2 \lambda_l}{N_0} = \bar{\gamma}_c \lambda_l, \quad (30)$$

where  $\{\lambda_l\}$  are the eigenvalues of  $\mathbf{A}$  and  $\bar{\gamma}_c = \sigma_f^2/N_0$ . The characteristic function then becomes

$$\psi_\gamma(j\omega) = \prod_{l=0}^L \left( \frac{1}{1 - j\omega\bar{\gamma}_c\lambda_l} \right)^M. \quad (31)$$

In the case where  $\mathbf{A} = \mathbf{I}$ , the eigenvalues are all equal to unity, that is,  $\lambda_l = 1$  for all  $l$ . In this case (31) becomes

$$\psi_\gamma(j\omega) = \left( \frac{1}{1 - j\omega\bar{\gamma}_c} \right)^{(L+1)M}. \quad (32)$$

The probability density function of  $\gamma$  is obtained by inverse-Fourier transforming the characteristic function, which yields

$$p(\gamma) = \frac{1}{(D-1)!\bar{\gamma}_c^D} \gamma^{D-1} e^{-\gamma/\bar{\gamma}_c}, \quad \gamma \geq 0, \quad (33)$$

where  $D = (L+1)M$ . We can compute a lower bound on the bit error probability by truncating the upper bound in (17) to include only the single error event which corresponds to the minimum Euclidean weight  $d_{\min}^2$ . By averaging this conditional lower bound over the probability density function in (33), we obtain

$$P_{b1} = \int_0^\infty Q(\sqrt{2\gamma}) p(\gamma) d\gamma \quad (34)$$

$$= \left[ \frac{1}{2}(1-\mu) \right]^D \sum_{k=0}^{D-1} \binom{D-1+k}{k} \left[ \frac{1}{2}(1+\mu) \right]^k,$$

where  $\mu = \sqrt{\bar{\gamma}_c/(1+\bar{\gamma}_c)}$ . For high SNR ( $\gamma \gg 1$ ),  $P_{b1}$  is approximated by

$$P_{b1} \approx \binom{2D-1}{D} \left( \frac{1}{4\bar{\gamma}_c} \right)^D. \quad (35)$$

We note that this result is identical to the average probability of error for a maximal ratio combiner with diversity of order  $D$  in flat fading. Hence, our system benefits from an implicit diversity of order  $L+1$  due to channel dispersion in addition to the explicit  $M$ th-order spatial diversity. Equation (34) also gives the lower bound on the error probability for an  $(N, M)$  system in flat fading (where  $L=0$ ). In this case, only spatial diversity gain is available.

Figure 3 shows the lower bound on the bit error probability as a function of the average received SNR per bit,  $\bar{\gamma}_b$ , for  $D=2, 4, 6$ . The average SNR per bit is related to the average SNR per channel coefficient  $\bar{\gamma}_c$  by the formula  $\bar{\gamma}_b = D\bar{\gamma}_c$ . Also shown is the simulated bit error rate (BER) for three different MIMO systems [29]. The first system is a (2, 2) system operating in a Rayleigh fading channel with no time dispersion ( $L=0$ ). The effective order of diversity is  $D=M=2$ . The second system is also a (2, 2) BPSK system, but operating in a two-path ( $L=1$ ) Rayleigh fading channel, where the variance of the channel coefficients is  $\sigma_f^2 = 1/D$ . In this

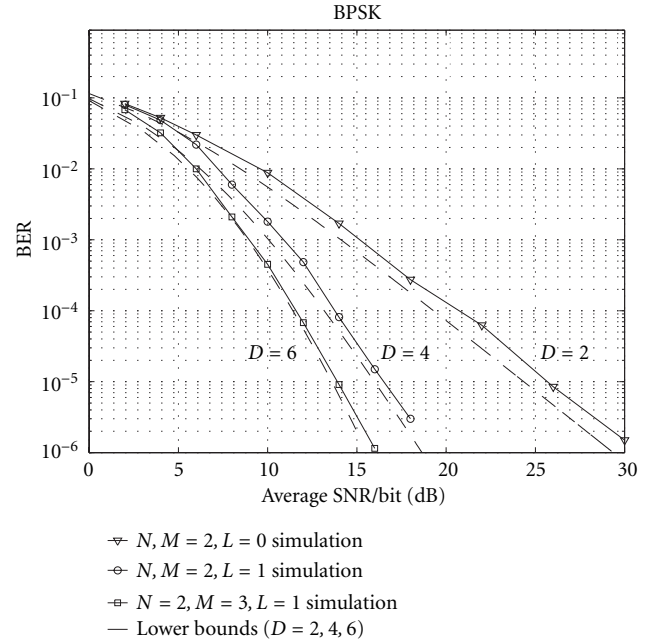


FIGURE 3: Performance of ML detectors in dispersive and nondispersive Rayleigh fading.

case, the effective order of diversity is  $D=(L+1)M=4$ , and the simulation results agree very well with the lower bound. Finally, the third system is a (2, 3) system operating in the same two-path Rayleigh fading channel as the previous system. Again, the simulation results are in agreement with the lower bound, which predicts an effective order of diversity  $D=(L+1)M=6$ .

Based on the analysis and simulation results presented above, we can conclude that the MIMO MLSE is capable of exploiting the full diversity potential offered by the channel, including explicit antenna diversity as well as implicit diversity due to channel dispersion.

#### 4. SUBOPTIMAL MIMO DETECTORS

In systems with many transmit antennas and significant ISI, that is, where the parameters  $N$  and  $L$  are large, the computational complexity of the MIMO maximum likelihood sequence estimator might be problematic, since the complexity grows exponentially with  $NL$ . We are therefore interested in evaluating receiver structures of lower complexity which can offer good, yet suboptimal performance. The MLSE may be replaced by a multiple-input multiple-output combiner-equalizer structure such as a MIMO linear equalizer or a MIMO decision feedback equalizer. These equalizers are completely analogous to single channel equalizers, which are well known from the general literature [27]. Optimal (in the minimum mean squared error sense) MIMO linear and decision feedback equalizers were presented in [7] for the case of completely known and fixed channels, while adaptive versions were presented in [30]. A related study comparing the performance of joint maximum likelihood detection

and MMSE combining of co-channel signals was presented in [31]. In this section, we consider similar MMSE linear and decision feedback equalizers for MIMO systems in multipath fading. We also consider alternative detectors for memoryless MIMO channels, based on both the MMSE and the zero-forcing criteria.

#### 4.1. MIMO linear equalizer

The MIMO linear equalizer performs cancellation of ISI and interchannel interference as well as diversity combining. It consists of a bank of transversal filters, each with  $2K + 1$  coefficients, operating on the  $M$  received signals. Hence, the estimate of the data symbol transmitted on the  $n$ th antenna at time  $k$  is represented by

$$\tilde{d}_n(k) = \sum_{m=1}^M \mathbf{w}_{mn}^T(k) \mathbf{v}_m(k), \quad n = 1, 2, \dots, N, \quad (36)$$

where  $\mathbf{w}_{mn}(k) = [w_{mn}^{(-K)}(k) \cdots w_{mn}^{(0)}(k) \cdots w_{mn}^{(K)}(k)]^T$  is the vector of filter coefficients corresponding to channel  $n$  and receive antenna  $m$  at time  $k$ , and  $\mathbf{v}_m(k)$  is the vector of symbol-spaced signal samples from receive antenna  $m$  stored in the transversal filter at time  $k$ , which is given by  $\mathbf{v}_m(k) = [v_m(k+K) \cdots v_m(k) \cdots v_m(k-K)]^T$ . Equation (36) may be represented in matrix form as

$$\tilde{\mathbf{d}}(k) = \mathbf{W}(k) \mathbf{v}_{1:M}(k), \quad (37)$$

where  $\tilde{\mathbf{d}}(k) = [\tilde{d}_1(k) \tilde{d}_2(k) \cdots \tilde{d}_N(k)]^T$ ,  $\mathbf{W}(k)$  is the  $N \times M(2K + 1)$  matrix given by

$$\mathbf{W}(k) = \begin{bmatrix} \mathbf{w}_{11}^T(k) & \mathbf{w}_{21}^T(k) & \cdots & \mathbf{w}_{M1}^T(k) \\ \mathbf{w}_{12}^T(k) & \mathbf{w}_{22}^T(k) & \cdots & \mathbf{w}_{M2}^T(k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{w}_{1N}^T(k) & \mathbf{w}_{2N}^T(k) & \cdots & \mathbf{w}_{MN}^T(k) \end{bmatrix}, \quad (38)$$

and  $\mathbf{v}_{1:M}(k)$  is the  $M(2K + 1) \times 1$  vector of received signals given by

$$\mathbf{v}_{1:M}(k) = [\mathbf{v}_1^T(k) \mathbf{v}_2^T(k) \cdots \mathbf{v}_M^T(k)]^T. \quad (39)$$

The optimal values of the filter coefficients at time  $k$  are determined using the MMSE criterion. The overall mean squared error is the sum of the mean squared errors for each of the  $N$  transmitted symbols,

$$\text{MSE} = \sum_{n=1}^N E \left[ |d_n(k) - \tilde{d}_n(k)|^2 \right]. \quad (40)$$

Minimization of the MSE with respect to the filter coefficient vectors  $\{\mathbf{w}_{mn}(k)\}$  yields the following solutions:

$$\begin{aligned} \mathbf{w}_{mn}^T(k) &= E[d_n(k) \mathbf{v}_m'(k)] (E[\mathbf{v}_m(k) \mathbf{v}_m'(k)])^{-1}, \\ m &= 1, 2, \dots, M, \quad n = 1, 2, \dots, N. \end{aligned} \quad (41)$$

In the case of flat fading, a somewhat simpler detector can be designed [32]. In this case, estimates of the transmitted symbols are formed by linearly combining the received signal samples  $v_1(k), v_2(k), \dots, v_M(k)$ . The linear combining may be represented in matrix form as

$$\tilde{\mathbf{d}}(k) = \mathbf{W}(k) \mathbf{v}(k), \quad (42)$$

where  $\mathbf{v}(k) = [v_1(k) v_2(k) \cdots v_M(k)]^T$ ,  $\mathbf{W}(k)$  is the matrix of weights given by  $\mathbf{W}(k) = [\mathbf{w}_1^T(k) \mathbf{w}_2^T(k) \cdots \mathbf{w}_N^T(k)]^T$  and  $\mathbf{w}_n(k) = [w_{1n}(k) w_{2n}(k) \cdots w_{Mn}(k)]^T$ . Minimization of the MSE with respect to the weights yields the solutions

$$\mathbf{w}_n^T(k) = E[d_n(k) \mathbf{v}'(k)] (E[\mathbf{v}(k) \mathbf{v}'(k)])^{-1}, \quad n = 1, 2, \dots, N. \quad (43)$$

If we use zero-forcing of the interchannel interference as optimality criterion instead of the MMSE criterion in this case, the combining weight matrix is simply given as the inverse of the channel matrix, that is,

$$\mathbf{W}(k) = \mathbf{F}^{-1}(k). \quad (44)$$

The price paid for the complete elimination of interference is noise enhancement in all channels. However, at high SNR levels, the zero-forcing solution approaches the MMSE solution.

#### 4.2. MIMO decision feedback equalizer

The MIMO decision feedback equalizer (DFE) consists of a feedforward section with transversal filters of length  $K_1 + 1$ , a feedback section with filters of length  $K_2 = L$  and  $N$  symbol decision devices. The structure of the feedforward section of the DFE is similar to the linear equalizer, and it performs cancellation of both interchannel interference and ISI, as well as diversity combining. Just as in the single channel case, the feedback section can further reduce ISI due to postcursors. In general, the estimate of the data symbol transmitted on the  $n$ th antenna at time  $k$  is represented by

$$\begin{aligned} \tilde{d}_n(k) &= \sum_{m=1}^M \mathbf{a}_{mn}^T(k) \mathbf{v}_m^f(k) \\ &\quad - \sum_{i=1}^N \mathbf{b}_{in}^T(k) \hat{d}_i^f(k), \quad n = 1, 2, \dots, N. \end{aligned} \quad (45)$$

The vector  $\mathbf{a}_{mn}(k) = [a_{mn}^{(-K_1)}(k) a_{mn}^{(-K_1+1)}(k) \cdots a_{mn}^{(0)}(k)]^T$  represents the feedforward filter coefficients corresponding to channel  $n$  and receive antenna  $m$  at time  $k$ , and  $\mathbf{v}_m^f(k)$  denotes a vector of signal samples received at antenna  $m$ , given by  $\mathbf{v}_m^f(k) = [v_m(k+K_1) v_m(k+K_1-1) \cdots v_m(k)]^T$ . The feedback filter  $\mathbf{b}_m(k)$  takes the symbol decisions for transmit antenna  $i$  as input and outputs an estimate of the interchannel interference and postcursor ISI seen by channel  $n$ . The feedback filter coefficients are given by  $\mathbf{b}_{in}(k) = [b_{in}^{(1)}(k) b_{in}^{(2)}(k) \cdots b_{in}^{(K_2)}(k)]^T$ . The  $K_2$  most recent decisions obtained for the symbols transmitted by antenna  $i$  are represented by the vector  $\hat{\mathbf{d}}_i^f(k) = [\hat{d}_i(k-1) \hat{d}_i(k-2) \cdots \hat{d}_i(k-K_2)]^T$ .

The feedback filters  $\mathbf{b}_{mn}(k)$  mainly perform ISI cancellation, while the filters  $\mathbf{b}_{ij}(k)$ , where  $i \neq j$ , mainly perform cancellation of interchannel interference. Equation (45) may be represented in matrix form as

$$\tilde{\mathbf{d}}(k) = \mathbf{A}(k)\mathbf{v}_{1:M}^f(k) - \mathbf{B}(k)\hat{\mathbf{d}}_{1:N}^f(k), \quad (46)$$

where  $\mathbf{A}(k)$  is the  $N \times M(K_1 + 1)$  matrix given by

$$\mathbf{A}(k) = \begin{bmatrix} \mathbf{a}_{11}^T(k) & \mathbf{a}_{21}^T(k) & \cdots & \mathbf{a}_{M1}^T(k) \\ \mathbf{a}_{12}^T(k) & \mathbf{a}_{22}^T(k) & \cdots & \mathbf{a}_{M2}^T(k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{1N}^T(k) & \mathbf{a}_{2N}^T(k) & \cdots & \mathbf{a}_{MN}^T(k) \end{bmatrix}, \quad (47)$$

$\mathbf{B}(k)$  is the  $N \times NK_2$  matrix given by

$$\mathbf{B}(k) = \begin{bmatrix} \mathbf{b}_{11}^T(k) & \mathbf{b}_{21}^T(k) & \cdots & \mathbf{b}_{N1}^T(k) \\ \mathbf{b}_{12}^T(k) & \mathbf{b}_{22}^T(k) & \cdots & \mathbf{b}_{N2}^T(k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_{1N}^T(k) & \mathbf{b}_{2N}^T(k) & \cdots & \mathbf{b}_{NN}^T(k) \end{bmatrix}, \quad (48)$$

$\mathbf{v}_{1:M}^f(k)$  is the  $M(K_1 + 1) \times 1$  vector of received signals given by

$$\mathbf{v}_{1:M}^f(k) = [\mathbf{v}_1^f(k) \mathbf{v}_2^f(k) \cdots \mathbf{v}_M^f(k)]^T, \quad (49)$$

and  $\hat{\mathbf{d}}_{1:N}^f(k)$  is the  $NK_2 \times 1$  vector of recent symbol decisions given by

$$\hat{\mathbf{d}}_{1:N}^f(k) = [\hat{\mathbf{d}}_1^f(k) \hat{\mathbf{d}}_2^f(k) \cdots \hat{\mathbf{d}}_N^f(k)]^T. \quad (50)$$

As in the case of the linear equalizer, the optimal DFE coefficients are determined using the MMSE criterion. In particular, the solution for the feedforward coefficient vector  $\mathbf{a}_{mn}(k)$  is given by

$$\mathbf{a}_{mn}^T(k) = E[d_n(k)\mathbf{v}_m^f(k)] (E[\mathbf{v}_m^f(k)\mathbf{v}_m^f(k)])^{-1}, \quad (51)$$

$$m = 1, 2, \dots, M, \quad n = 1, 2, \dots, N.$$

The feedback coefficient vectors are computed with the help of the feedforward vectors and are given by

$$\mathbf{b}_{in}^T(k) = \sum_{m=1}^M \mathbf{a}_{mi}^T(k) [\mathbf{f}_{nm}^{(1)}(k) \mathbf{f}_{nm}^{(2)}(k) \cdots \mathbf{f}_{nm}^{(L)}(k)], \quad (52)$$

$$i, n = 1, 2, \dots, N,$$

where  $\mathbf{f}_{nm}^{(l)}(k)$  is a column vector of length  $K_1 + 1$  which consists of the discrete-time coefficients of the channel between transmit antenna  $n$  and receive antenna  $m$  at subsequent time instants, and is given by

$$\mathbf{f}_{nm}^{(l)}(k) = [f_{nm}^{(l+K_1)}(k+K_1) f_{nm}^{(l+K_1-1)}(k+K_1-1) \cdots f_{nm}^{(l)}(k)]^T, \quad (53)$$

$$l = 1, 2, \dots, L.$$

### 4.3. Decision-directed MRC detector for flat fading channels

The idea behind the decision-directed maximal ratio combining (DD-MRC) detector is to exploit the knowledge of the channel coefficients to achieve  $M$ th-order diversity reception of each of the  $N$  transmitted signals in flat fading channels. The detector consists of two sections. The front section cancels interchannel interference due to signals transmitted on antennas other than the one of interest, and the back section performs maximal ratio combining [32]. In order to cancel interference, preliminary symbol decisions are made for the interfering signals. These decisions can, for example, be obtained with the zero-forcing or MMSE detectors discussed in Section 4.1. Using the preliminary decisions, estimates of the interchannel interference are calculated and subtracted from the received signals.  $M$ th-order diversity reception is achieved by performing maximal ratio combining of the (interference free) signals from each receiving antenna. The signals at the input of the  $n$ th combiner are given by

$$\begin{bmatrix} u_{1n} \\ u_{2n} \\ \vdots \\ u_{Mn} \end{bmatrix} = \begin{bmatrix} f_{n1}d_n + \sum_{i \neq n}^N f_{i1}(d_i - \hat{d}_i) + \eta_1 \\ f_{n2}d_n + \sum_{i \neq n}^N f_{i2}(d_i - \hat{d}_i) + \eta_2 \\ \vdots \\ f_{nM}d_n + \sum_{i \neq n}^N f_{iM}(d_i - \hat{d}_i) + \eta_M \end{bmatrix}, \quad (54)$$

where we have left out the time indices to simplify notation. The preliminary symbol estimates are given by  $\{\hat{d}_i\}$ ,  $i = 1, 2, \dots, N$ , and we observe that if the estimates are correct, the terms involving sums of interfering signals cancel. In the case of BPSK, the output of the combiners can be expressed as decision variables in the form

$$U_n = \text{Re} \left\{ \sum_{m=1}^M u_{mn} f_{nm}^* \right\}, \quad n = 1, 2, \dots, N. \quad (55)$$

### 4.4. Performance of suboptimal detectors

We do not provide analytical performance results for the suboptimal MIMO detectors discussed in this section. Instead, we give results for the bit error rate (BER) obtained through computer simulations. Figure 4 shows the BER as a function of average received SNR per bit for the linear equalizer and the DFE, respectively, in a two-path ( $L = 1$ ) fading channel. In both cases, the following systems were simulated:  $(N, M) = (2, 4)$ ,  $(2, 6)$ , and  $(2, 8)$ . In the case of the linear equalizer, the filter length was 5 ( $K = 2$ ). The feedforward filter lengths of the DFE were all  $K_1 = 2$ , while the feedback filters had  $K_2 = L = 1$  coefficients.

It is clear from the graphs that both suboptimal detectors perform significantly poorer than the MLSE. Both the linear equalizer and the DFE simulated here are unable to fully cancel the interference, as is evident from the error floors in the graphs. However, we observe that as the



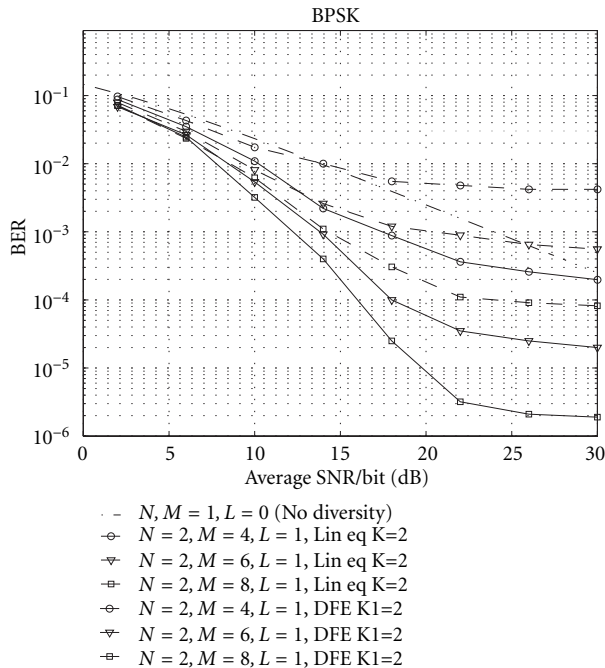


FIGURE 4: Performance of the MIMO linear and decision feedback equalizers in two-path fading channels.

number of receive antennas increases from 4 to 8, thus increasing the order of diversity in the systems, the error rates are improved. With a higher order of receiver diversity the error floors are lowered, and it appears that in the two-path channel, the improvements in BER are approximately one order of magnitude for every two receive antennas added. This observation is made for both detectors. We also note that the decision feedback equalizers provide improvements of 3–5 dB over the linear equalizers at a BER of  $10^{-3}$ .

Figure 5 shows the BER performance achieved with the simplified detectors designed specifically for the case of flat fading ( $L = 0$ ). The simulation results were obtained by simulating a (2, 2) system. In particular, we observe that the performance of the MMSE detector is within 0.5 dB of the “no diversity” curve, while the performance of the zero-forcing detector is about 2 dB worse, mainly due to noise enhancement caused by the matrix inversion. The performance curves for the DD-MRC detector were obtained with both correct (known) preliminary bit decisions and actual preliminary decisions provided by a zero-forcing detector. With correct decisions, the performance attains the lower bound on the performance of the optimal maximum likelihood detector, which we expect since interchannel interference is completely eliminated without compromising the order of diversity. In the other case, the DD-MRC’s performance is limited by the zero-forcing detector’s inability to make reliable decisions. We conclude that the linear detectors lose one order of diversity compared to the optimal detector on the flat fading channel.

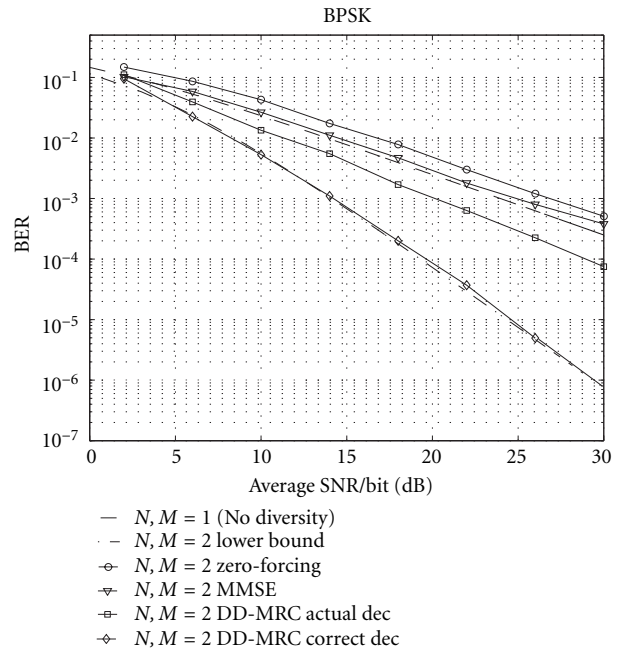


FIGURE 5: Performance of the simplified MIMO detectors in flat fading.

## 5. ANALYSIS OF CODED MIMO SYSTEMS

In this section, we turn our attention to coded MIMO systems. Error correcting codes may be applied to MIMO systems to improve the performance in fading and mitigate the degradation resulting from interchannel interference. Although there exist the so-called space-time codes which have been developed specifically for MIMO systems, any  $(n, k)$  convolutional code or block code may in principle be applied to improve the performance. In the MIMO systems that we consider here, the data are encoded with a convolutional code and interleaved before they are transmitted over spatially distributed fading multipath channels. This scheme resembles a serially concatenated convolutional code (SCCC), where the convolutional code is the outer code and the MIMO multipath channel takes on the role of an inner time-varying convolutional code. In realizing this resemblance, we may analyze the performance of MIMO systems using the same bounding techniques as developed previously for SCCCs [33, 34, 35].

### 5.1. The union bound for coded MIMO systems

First, we consider an SCCC which consists of an outer convolutional code  $C_o$  with rate  $R_o = k/p$  and an inner convolutional code  $C_i$  with rate  $R_i = p/n$ . The two codes are joined together by a uniform interleaver of length  $K$ . This interleaver maps an input word with Hamming weight  $l$  into any one of  $\binom{K}{l}$  distinct permutations of itself with uniform probability  $1/\binom{K}{l}$ . Assuming linear constituent codes, the resulting concatenated code  $C_s$  is a linear code with rate  $R_s = k/n$ . By terminating the trellis of the outer code at both the

beginning and the end of the codeword, we construct an equivalent block code. Thus, we may process the received signals on a codeword-by-codeword basis. The length of the input sequence is  $KR_o$ , while the length of the SCRC codeword is  $K/R_i$ . The optimal decoder for this code is a hypothetical maximum likelihood sequence decoder which operates on a hyper-trellis where the states are pairs of states of the outer and inner codes. Although prohibitively complex due to the interleaving, we will assume ML decoding for the purpose of calculating the union bound on the bit error probability.

In [33], Benedetto et al. demonstrated that high coding gains can be obtained at very low SNRs with serially concatenated codes, provided that the interleaver is large and the inner code is a recursive code. The outer code can be either recursive or nonrecursive. In particular, it was shown that every term of the union bound decreases asymptotically at least as rapidly as  $K^{-\lfloor (d_{\text{free}}^o + 1)/2 \rfloor}$ , where  $d_{\text{free}}^o$  is the free distance of the outer code and  $\lfloor \cdot \rfloor$  denotes the integer value. This is referred to as *interleaving gain*.

We view the MIMO system shown in Figure 6 as an SCRC. In order to benefit from interleaving gain, the inner "code," that is, the channel must be recursive. The multipath channel can be made to appear recursive to the outer code by performing recursive precoding prior to transmission. Narayanan [36] showed that the preferred precoder is of the form  $1/(1 + Z^R)$  due to the enhanced convergence properties of iterative decoders for such codes. For this reason and for simplicity, we will restrict our attention to a simple differential precoder with polynomial  $1/(1 + Z)$  and rate  $R_i = 1$ .

In the MIMO system of Figure 6, the output sequence of the outer code is passed through a uniform interleaver before being fed to  $N$  parallel and identical precoders and modulators, where each modulator is connected to a separate antenna. The  $N$  signals are then transmitted over an  $(N, M)$  channel and received by  $M$  receive antennas which are connected to an ML receiver capable of performing joint demodulation, detection and decoding. As before, the channels between each transmit and receive antenna are assumed to be independently fading multipath channels, each with  $L + 1$  paths. In Section 3, we saw that the uncoded performance in fading multipath channels is approximated by the performance of a  $(1, D)$  system with maximal ratio combining in flat fading. Both implicit diversity of order  $L + 1$  due to channel dispersion, and explicit spatial diversity of order  $M$  contribute to the effective diversity order  $D = (L + 1)M$  of the uncoded system. Analogously, the performance of a coded MIMO system is approximated by the performance of an SCRC in flat fading with diversity of order  $D$ . We use this knowledge in developing the union bound below.

Since the SCRC is a linear code, the bit error probability can be calculated under the assumption that the all-zero codeword was transmitted. We define a pairwise error event as the event in which the likelihood of a codeword with Hamming weight  $h$ , generated by an information word with Hamming weight  $w$ , is higher than the likelihood of the all-zero codeword. Assuming ML decoding and BPSK modulation, the union bound is given by (see [33])

$$P_b \leq \sum_{h=1}^{K/R_i} B_h P_h, \quad (56)$$

where  $P_h$  is the pairwise error probability in Rayleigh fading given by

$$P_h = \left[ \frac{1}{2}(1 - \mu) \right]^{Dh} \sum_{k=0}^{Dh-1} \binom{Dh-1+k}{k} \left[ \frac{1}{2}(1 + \mu) \right]^k, \quad (57)$$

and  $\mu = \sqrt{\bar{\gamma}_c R_s / (1 + \bar{\gamma}_c R_s)}$ ,  $R_s = R_o = k/n$ ,  $\bar{\gamma}_c = \bar{\gamma}_b/D$ , and  $D = (L + 1)M$ . The bit error multiplicity is expressed as

$$B_h = \sum_{w=1}^{KR_o} \sum_{l=0}^K \frac{w}{KR_o \binom{K}{l}} A_{w,l}^{C_o} \cdot A_{l,h}^{C_i}, \quad (58)$$

where  $A_{i,j}^{C_o}$  and  $A_{i,j}^{C_i}$  denote the number of codewords with Hamming weight  $j$  generated by input sequences with Hamming weight  $i$  of the outer and inner codes, respectively. These input-output weight spectra may be calculated using the recursive method described in [35].

For  $K$  much larger than the memory of the outer convolutional code, (56) can be approximated by (see [33])

$$P_b \leq \sum_{h=1}^{K/R_i} P_h \sum_{w=1}^{KR_o} \sum_{l=d_{\text{free}}^o}^K \sum_{n^o=1}^{n_M^o} \sum_{n^i=1}^{n_M^i} \frac{w}{KR_o} \times K^{n^o+n^i-l-1} \frac{l!}{p^{n^o+n^i} n^o! n^i!} T_{w,l,n^o}^{C_o} \cdot T_{l,h,n^i}^{C_i}, \quad (59)$$

where  $T_{i,j,n}^{C_o}$  and  $T_{i,j,n}^{C_i}$  denote the number of codewords of weight  $j$  generated by input sequences of weight  $i$  that are formed by the concatenation of  $n$  adjacent error events of the outer and inner codes, respectively. The free distance of the outer code is denoted by  $d_{\text{free}}^o$ , and  $n_M^o$  and  $n_M^i$  refer to the maximum number of adjacent error events of the outer and inner codes, respectively.

For large  $K$ , the dominant coefficient of  $P_h$  is the one for which the exponent of  $K$  is maximum. We define this maximum exponent as

$$\alpha(h) = \max_{w,l} \{n^o + n^i - l - 1\}. \quad (60)$$

In general,  $\alpha(h)$  cannot be evaluated without specifying the outer and inner codes, but general expressions can be found for two important cases, namely (i) the exponent corresponding to the minimum output weight, and (ii) the overall maximum exponent.

For large values of  $\bar{\gamma}_b$ , the union bound is dominated by the term corresponding to the minimum value of  $h$ , known as the free distance of the SCRC or  $d_{\text{free}}^s$ . For smaller values of  $\bar{\gamma}_b$ , the union bound is dominated by terms corresponding to other values of  $h$ . To determine what these values are, we must first find the maximum value of  $\alpha$ ,

$$\alpha_M = \max_h \{\alpha(h)\} = \max_{w,l,h} \{n^o + n^i - l - 1\}. \quad (61)$$

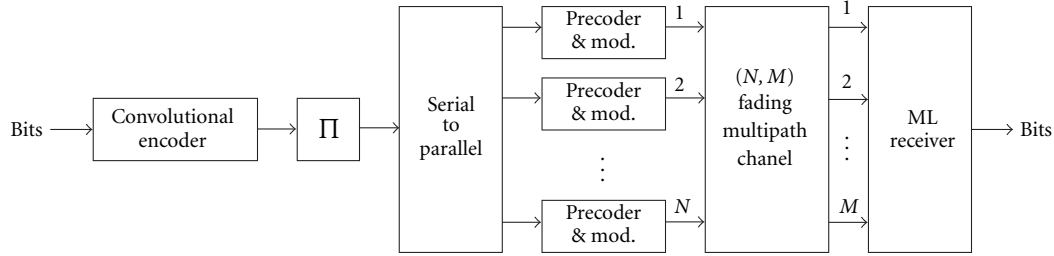


FIGURE 6: Coded MIMO system.

In the case of a nonrecursive inner code, input sequences with weight 1 exist, so that an input sequence with weight  $l$  will generate at most  $l$  error events in the inner code. Thus,  $n^i \leq l$  and

$$\alpha_M = n_M^o - 1 \geq 0, \quad (62)$$

which effectively eliminates interleaving gain.

For a recursive inner code, the minimum weight of input sequences that can generate error events is 2. Input sequences of weight  $l$  can therefore generate at most  $\lfloor l/2 \rfloor$  error events, and it is shown in [33] that

$$\alpha_M = - \left\lfloor \frac{d_{\text{free}}^o + 1}{2} \right\rfloor \leq 0, \quad (63)$$

when  $l = d_{\text{free}}^o$ . In this case, the exponents of  $K$  are always negative integers yielding interleaving gain. We can therefore conclude that the union bound is dominated by error events for which  $l = d_{\text{free}}^o$ , that is, error events from the inner code which generate error events in the outer code with weight equal to  $d_{\text{free}}^o$ .

## 5.2. Approximation of the union bound

As we have seen, calculating the union bound involves the calculation of input-output weight spectra for the constituent codes. The bound is dominated by a relatively small number of low-weight error events, and based on the observations made above, we can find an approximation of the probability of error by considering only these events. This approach allows us to circumvent the full calculation of the weight spectra.

Benedetto et al. [33] concluded that the union bound is dominated by the error event which is associated with the maximum exponent  $\alpha_M$ . The output weight of this event is denoted by  $h(\alpha_M)$ . Furthermore, they observed that the recursive inner code has minimum input weight 2, and the minimum output weight of such input weight-2 codewords is denoted by  $d_{\text{min},2}^i$ , also known as the effective free distance. The authors suggested that it is the concatenation of such minimum weight error events that result in the dominating error event with output weight  $h(\alpha_M)$ . In particular, error events with  $l = d_{\text{free}}^o$  consist of  $d_{\text{free}}^o/2$  concatenated error events in the inner code, each with input weight 2 and output weight  $d_{\text{min},2}^i$ . For even values of  $d_{\text{free}}^o$ , the SCCC output

weight is therefore given by

$$h(\alpha_M) = \frac{d_{\text{free}}^o d_{\text{min},2}^i}{2}. \quad (64)$$

For odd values of  $d_{\text{free}}^o$ , the SCCC error events are concatenations of several inner code error events with input weight 2 and one event with input weight 3, that is,

$$h(\alpha_M) = \frac{(d_{\text{free}}^o - 3)d_{\text{min},2}^i}{2} + d_{\text{min},3}^i, \quad (65)$$

where  $d_{\text{min},3}^i$  is the minimum output weight of codewords of the inner code with input weight 3. If  $d_{\text{min},3}^i = \infty$ , the output weight is given by

$$h(\alpha_M) = \frac{(d_{\text{free}}^o + 1)d_{\text{min},2}^i}{2}. \quad (66)$$

However, in [37], Gray claimed that the concatenation of  $d_{\text{free}}^o/2$  error events from the inner code, each with output weight  $d_{\text{min},2}^i$ , is not guaranteed to be the most likely error event with  $l = d_{\text{free}}^o$ . There may be other and more likely inner code error events which also have input weight  $d_{\text{free}}^o$ . Our approach here is to consider all concatenated error events of the inner code that have input weight  $l^* = d_{\text{free}}^o$ , regardless of the specific minimum-weight error events they consist of. If no such events exist in the inner code, we must consider error events with input weight  $l^* = d_{\text{free}}^o + 1$  or, if necessary, higher. The procedure can be simplified by considering only concatenated events composed of error events with minimum output weight  $d_{\text{min},j}^i$ , where  $j = 2, 3, \dots$ . With this simplification there is a risk that certain compound events with  $h > d_{\text{free}}^o$  are missed, but the risk is justified by the simplified calculations.

The union bound can now be approximated using only the multiplicity factors  $B_h$  that correspond to the particular values of  $h$  associated with the concatenated error events discussed above, that is,

$$B_h \approx \sum_{w=1}^k \frac{w}{KR_o} \cdot \frac{A_{w,l^*}^{C_o} \cdot A_{l^*,h}^{C_i}}{\binom{K}{l^*}}. \quad (67)$$

The value of  $A_{w,l^*}^{C_o}$  is found from the input-output weight spectrum of the outer code, while the value of  $A_{l^*,h}^{C_i}$  can be approximated using the fact that a concatenated error event

that consists of  $n^i$  adjacent events in a block of size  $K$ , can be obtained in

$$\binom{K}{n^i} = \frac{K!}{n^i!(K-n^i)!} < \frac{K^{n^i}}{n^i!} \quad (68)$$

ways. This quantity includes overlapping events, but these can be ignored if  $K$  is much greater than  $n^i$  and the length of the events. Also utilizing the inequality

$$\binom{K}{l^*} < \frac{K^{l^*}}{l^*!}, \quad (69)$$

we arrive at our final approximation of  $B_h$ ,

$$B_h \approx \sum_{w=1}^k \frac{w}{R_o} A_{w,l^*}^{C_o} \frac{l^*!}{n^i!} K^{n^i-l^*-1}, \quad (70)$$

where  $n^i \leq \lfloor l^*/2 \rfloor$ , as stated previously.

In order to calculate the bit error multiplicity given by (70), we must find the distances  $d_{\min,j}^i$  for various values of  $j$ . In most cases, this can be done by inspection of the trellis of the inner code. As an example, consider the (2, 2) system that uses a rate-1/2 nonrecursive terminated convolutional code with generating polynomial (5, 7)<sub>8</sub> as outer code and a rate-1 differential code as inner code. The outer code has a free distance of  $d_{\text{free}}^o = 5$ . Hence, the maximum exponent of  $K$  is  $\alpha_M = -3$ . For this particular system, we realize that the dominant error event has  $l^* = 6$ , and the minimum output weight for the input weight-2 error events in the inner code is  $d_{\min,2}^i = d_{\text{free}}^i = 1$ . Since  $d_{\min,3}^i = \infty$  in this case, the dominating output weight is  $h(\alpha_M) = 3$ . From the input-output weight spectrum of the outer code we find that

$$A_{2,6}^{C_o} = K - 7, \quad (71)$$

and that only a single input weight ( $w = 2$ ) is associated with the output weight  $l^* = 6$ , causing the summation over  $w$  in (70) to vanish. Hence, the bit error probability is approximated by just one term,

$$P_b \approx B_3 P_3, \quad (72)$$

where

$$B_3 \approx 480(K-7)K^{-4}, \quad (73)$$

and  $P_3$  is given by (57) with  $h = 3$ .

Figure 7 shows the bit error probability versus the average SNR per bit,  $\bar{\gamma}_b$ , for a (2, 2) system with an interleaver of length  $K = 512$ , calculated using the union bound in (56) as well as the approximate result in (72). The figure shows the error probability in a flat fading channel ( $L = 0$ , i.e.,  $D = 2$ ) and in a dispersive fading channel with two equal-strength paths ( $L = 1$ , i.e.,  $D = 4$ ). Also shown is the performance of a (1, 1) system in AWGN and in flat fading. As expected, the performance improves as the order of diversity increases. It is interesting to note that the performance

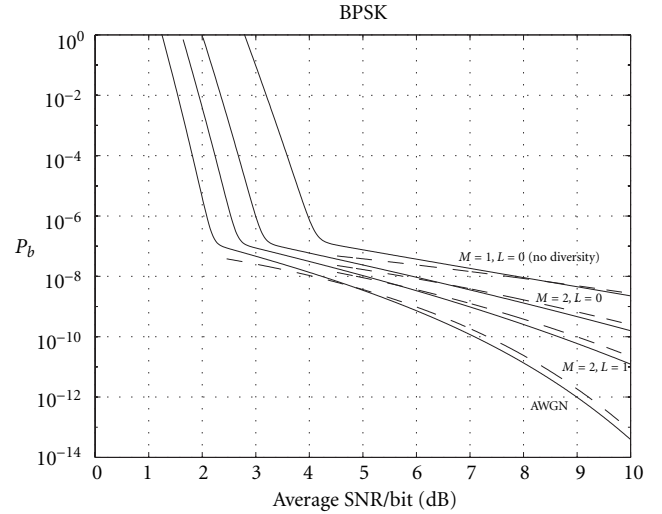
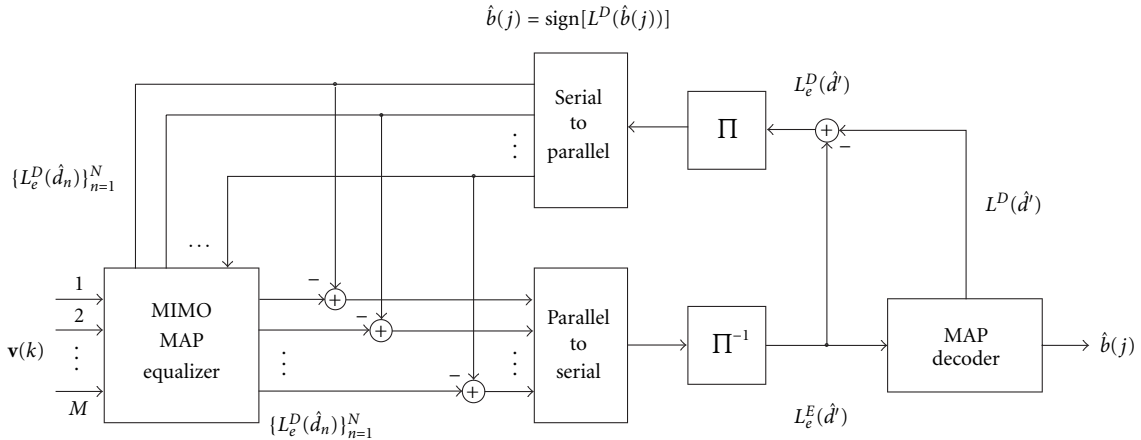


FIGURE 7: Union bounds for coded (2, 2) systems in fading.  $R_s = 1/2$ ,  $d_{\text{free}}^o = 5$ ,  $K = 512$ . Approximate bounds shown as dashed curves.

of the MIMO system with diversity of order  $D = 4$  is only 1–2 dB worse than the performance in AWGN for SNRs lower than 10 dB. Note that the union bound diverges for values of  $\bar{\gamma}_b$  less than 2.46 dB in AWGN and 4.52 dB in Rayleigh fading. These values correspond to the cutoff rate for a rate-1/2 code in AWGN and fading, respectively. The rather curious shape of the bound deserves some explanation. At medium to high SNR, the performance is dominated by only a few error events of low weight, resulting in the remarkable bit error rate displayed in this region. At low SNR, however, several error events of larger weight contribute to the bound and this causes the performance curve to have a different slope in the divergence region. Other techniques than the union bounding technique discussed here must be used to predict the performance in this region. A survey of such techniques can be found in the paper by Shamai and Sason [38]. We observe that the bit error probability of (72) is a good approximation of the union bound above the divergence region, even though only one value of  $h$  is used in the calculation. We conclude that we can obtain good approximations of the bit error probability without having to calculate the full input-output spectra of the constituent codes.

## 6. ITERATIVE EQUALIZATION AND DECODING

In Section 5, we viewed MIMO systems in fading multipath channels as serially concatenated codes in order to analyze their performance. We assumed that a maximum likelihood decoder was utilized to achieve optimal performance. While theoretically optimal in the ML sense, such a decoder would be prohibitively complex and therefore unrealizable in practice. In this section, we take a more practical approach and consider receivers which decode the constituent codes separately and in an iterative manner. This suboptimal approach has been shown to yield near-optimal performance in many cases. In our case, where the constituent “codes” are the


 FIGURE 8: Iterative MAP equalizer/decoder for  $(N, M)$  system.

outer convolutional code and the multipath channel, the receiver performs iterative equalization and decoding. This is also known as “turbo equalization,” a technique which was first proposed by Douillard et al. in [17]. The general idea is to exploit the error correcting capabilities of the interleaved channel code to enhance the performance of the equalizer. This is accomplished by iteratively passing soft a priori information between a soft-input, soft-output equalizer and a soft-input, soft-output channel decoder which are separated by an appropriate interleaver-deinterleaver pair. In this process, the reliability of the bit decisions is improved with each iteration. This is similar to the original “turbo decoding” principle which was first introduced by Berrou et al. in [39] for decoding of parallel concatenated codes.

As in the previous section, we consider a MIMO system which consists of an outer convolutional code followed by a pseudo-random interleaver and a differentially precoded MIMO channel. The differential precoder was chosen because of its simplicity and because it has been shown to enhance the convergence properties, and, hence, the performance, of the iterative equalizer/decoder [36]. The combination of differential precoding, modulation and multipath channel is modeled equivalently as a MIMO recursive transversal filter. We note that the number of memory elements in the precoded channel remains the same as in the nonprecoded channel. Hence, the number of states in the finite state machine that characterizes the channel remains the same, while the branch transitions are altered. Since the precoder is a rate-1 recursive convolutional code, the data rate is not decreased. The number of states in the channel state trellis is  $2^{NL}$ , assuming binary signaling. Codewords of length  $NW$  are formed from a rate- $1/p$  convolutional code which terminates in the all-zero state. These codewords are transmitted, received and processed as separate frames.

When employing iterative equalization and decoding in our MIMO system, the receiver structure must accommodate the passing of soft information between the equalizer and the decoder, and vice versa. We assume that the channel coefficients are known or can be perfectly estimated at all times, and use a symbol-by-symbol maximum a posteriori proba-

bility (MAP) algorithm for both equalization and decoding, as shown in Figure 8. The equalizer computes a posteriori probabilities for the coded bits, based on both the received signals and a priori probabilities derived from the outputs of the channel decoder. Since we assume binary signaling, it is convenient to compute this soft information in the form of log-likelihood ratios, or  $L$ -values, which are given by

$$L^E(\hat{d}_n(k)) = \log \left( \frac{P\{d_n(k) = +1|\mathbf{v}\}}{P\{d_n(k) = -1|\mathbf{v}\}} \right), \quad n = 1, 2, \dots, N. \quad (74)$$

Here,  $\mathbf{v}$  is the noisy received codeword

$$\mathbf{v} = [\mathbf{v}(1)\mathbf{v}(2)\dots\mathbf{v}(W)] = \mathbf{v}_1^W \quad (75)$$

of length  $W$ , where  $\{\mathbf{v}(k)\}$  are  $M$ -dimensional vectors, as before. The  $L$ -values contain channel information, extrinsic information and a priori information. The extrinsic information is incremental information about the coded bit in question that has been obtained from all the other coded bits in the equalization process. Only channel information and extrinsic information are passed from the equalizer to the decoder, where, after parallel-to-serial conversion and deinterleaving, they are used as a priori information in the decoding process. This is important since, ideally, the a priori information would be provided by an independent source. We do not have access to such a source, but we may mimic the independence by minimizing the correlation between the a priori information and the previous decisions made by the equalizer. This is done by subtracting the a priori  $L$ -values  $L_e^D(\hat{d}_n(k))$  from the  $L$ -values at the output of the equalizer, as shown in Figure 8

$$L_e^E(\hat{d}_n(k)) = L^E(\hat{d}_n(k)) - L_e^D(\hat{d}_n(k)), \quad n = 1, 2, \dots, N. \quad (76)$$

After parallel-to-serial conversion and deinterleaving, the MAP channel decoder computes the a posteriori probabilities for the information bits and the coded bits, denoted by  $L^D(\hat{b}(j))$  and  $L^D(\hat{d}(k))$ , respectively. The prime indicates

deinterleaved coded bits. The extrinsic information to be fed back to the equalizer is obtained by subtracting the a priori information provided by the equalizer from the  $L$ -values at the output of the decoder:

$$L_e^D(\hat{d}'(k)) = L^D(\hat{d}'(k)) - L_e^E(\hat{d}'(k)). \quad (77)$$

The process described above constitutes one iteration and is repeated until the bit error rate converges or reaches an acceptable level. The final bit decisions  $\{\hat{b}(j)\}$  are obtained after the last iteration and are given by

$$\hat{b}(j) = \text{sign}[L^D(\hat{b}(j))]. \quad (78)$$

### 6.1. The MIMO MAP equalizer

The MAP equalizer uses the well-known BCJR algorithm [40] to compute the  $L$ -values  $L^E(\hat{d}_n(k))$ ,  $n = 1, 2, \dots, N$ . This algorithm optimally computes the a posteriori probabilities  $p(d_n(k)|\mathbf{v})$  for the coded bits  $d_n(k)$ , taking into account the information gathered from all the  $NW$  bits of the codeword. The  $L$ -value for a coded bit is given by

$$\begin{aligned} L^E(\hat{d}_n(k)) &= \log \left( \frac{P\{d_n(k) = +1|\mathbf{v}\}}{P\{d_n(k) = -1|\mathbf{v}\}} \right) \\ &= \log \left( \frac{\sum_{d_n(k)=+1}^{(s',s)} p(s', s, \mathbf{v})}{\sum_{d_n(k)=-1}^{(s',s)} p(s', s, \mathbf{v})} \right), \quad n = 1, 2, \dots, N, \end{aligned} \quad (79)$$

where  $s'$  and  $s$  denote the states of the channel trellis at times  $k-1$  and  $k$ , respectively, and  $(s', s)$  denotes a transition from  $s'$  to  $s$ . The summations in the numerator and the denominator are performed over all transitions which correspond to coded bits  $d_n(k) = +1$  and  $d_n(k) = -1$ , respectively.

The BCJR algorithm specifies a method for computing the probability  $p(s', s, \mathbf{v})$ :

$$\begin{aligned} p(s', s, \mathbf{v}) &= p(s', \mathbf{v}_1^k) \cdot P\{s|s'\} \cdot p(\mathbf{v}(k)|s', s) \cdot p(\mathbf{v}_{k+1}^W|s) \\ &= \alpha_{k-1}(s') \gamma_k(s', s) \beta_k(s). \end{aligned} \quad (80)$$

Here,  $\mathbf{v}_a^b$  denotes the sequence  $[\mathbf{v}(a)\mathbf{v}(a+1)\dots\mathbf{v}(b)]$ . The probabilities  $\alpha_k(s) = p(s', \mathbf{v}_1^k)$  are computed recursively as

$$\alpha_k(s) = \sum_{s'} \gamma_k(s', s) \alpha_{k-1}(s'), \quad (81)$$

with initial conditions

$$\alpha_0(0) = 1, \quad \alpha_0(s) = 0 \quad \forall s \neq 0, \quad (82)$$

since we assume that the zero state is the starting state. The probabilities  $\beta_k(s) = p(\mathbf{v}_{k+1}^W|s)$  are computed using the backwards recursion

$$\beta_{k-1}(s') = \sum_s \beta_k(s) \gamma_k(s', s), \quad (83)$$

where the initial conditions are

$$\beta_W(s) = 1 \quad \forall s. \quad (84)$$

The branch transition probability  $\gamma_k(s', s)$  is given by

$$\gamma_k(s', s) = P\{s|s'\} \cdot p(\mathbf{v}(k)|s', s), \quad (85)$$

where  $P\{s|s'\}$  is the a priori probability defined by

$$\begin{aligned} P\{s|s'\} &= \prod_{n=1}^N P_e\{d_n(k)|\mathbf{v}\} \\ &= \prod_{n=1}^N \exp \left( \frac{1}{2} d_n(k) L(d_n(k)) \right), \end{aligned} \quad (86)$$

and the event  $\mathbf{d}(k) = [d_1(k)d_2(k)\dots d_N(k)]^T$  corresponds to the transition  $(s', s)$  [41]. As for the transition probability  $p(\mathbf{v}(k)|s', s)$ , we may write

$$\begin{aligned} p(\mathbf{v}(k)|s', s) &\propto \exp \left( -\frac{1}{2\sigma^2} \left| \mathbf{v}(k) - \sum_{l=0}^L \mathbf{F}(k, l) \mathbf{d}(k-l) \right|^2 \right) \\ &= \exp \left( -\frac{1}{2\sigma^2} \left| \sum_{m=1}^M v_m(k) - \sum_{n=1}^N \sum_{l=0}^{L_{nm}} f_{nm}^{(l)}(k) d_n(k-l) \right|^2 \right) \\ &\stackrel{\text{def}}{=} \gamma_k^*(s', s), \end{aligned} \quad (87)$$

where  $\sigma^2$  is the variance of the additive white Gaussian noise (AWGN) [42]. Finally, combining (79), (80), (81), (82), (83), (84), (85), (86), and (87), we obtain the  $L$ -values for the coded bits

$$\begin{aligned} L^E(\hat{d}_n(k)) &= \sum_{n=1}^N L(d_n(k)) \\ &+ \log \left( \frac{\sum_{d_n(k)=+1}^{(s',s)} \alpha_{k-1}(s') \gamma_k^*(s', s) \beta_k(s)}{\sum_{d_n(k)=-1}^{(s',s)} \alpha_{k-1}(s') \gamma_k^*(s', s) \beta_k(s)} \right), \end{aligned} \quad n = 1, 2, \dots, N, \quad (88)$$

where the a priori  $L$ -values  $L(d_n(k))$  are substituted with the  $L$ -values  $L_e^D(\hat{d}_n(k))$  provided by the MAP decoder.

### 6.2. The MAP decoder

The decoder also uses the BCJR MAP algorithm to compute the a posteriori probabilities  $p(b(j)|\mathbf{v})$  for the information bits in the form of  $L$ -values

$$\begin{aligned} L^D(\hat{b}(j)) &= \log \left( \frac{P\{b(j) = +1|\mathbf{v}\}}{P\{b(j) = -1|\mathbf{v}\}} \right) \\ &= \log \left( \frac{\sum_{b(j)=+1}^{(q',q)} p(q', q, \mathbf{v})}{\sum_{b(j)=-1}^{(q',q)} p(q', q, \mathbf{v})} \right), \end{aligned} \quad (89)$$

where  $q'$  and  $q$  denote the states of the code trellis at times  $j-1$  and  $j$ , respectively.

The only difference in the computation of  $L$ -values between the decoder and the equalizer occurs in the computation of the branch transition probability  $\gamma_j(q', q)$  (see [42]):

$$\gamma_j(q', q) = \exp \left( \sum_{v=1}^p \left( \frac{1}{2} d'(j; v) L(d'(j; v)) + \frac{1}{2} b(j) L(b(j)) \right) \right). \quad (90)$$

In this expression,  $p$  is the inverse of the code rate and  $[d'(j; 1)d'(j; 2) \cdots d'(j; p)]$  denotes the  $p$ -bit codeword produced by the convolutional encoder, associated with the  $j$ th information bit. The prime denotes deinterleaved bits. We use the equalizer outputs  $L_e^E(\hat{d}'(j; v))$ ,  $v = 1, 2, \dots, p$  as estimates for their  $L$ -values.  $L(b(j))$  is the a priori  $L$ -value for the information bit  $b(j)$ . The probabilities  $\alpha_j(q)$  and  $\beta_j(q)$  are computed in the same way as described for the MAP equalizer, with the following initial conditions

$$\begin{aligned} \alpha_0(0) &= 1, & \alpha_0(s) &= 0 \quad \forall s \neq 0, \\ \beta_{NW}(0) &= 1, & \beta_{NW}(s) &= 0 \quad \forall s \neq 0. \end{aligned} \quad (91)$$

From (89) and (90) we obtain the  $L$ -values for the information bits

$$\begin{aligned} L^D(\hat{b}(j)) &= L(b(j)) \\ &+ \log \left( \frac{\sum_{\substack{(q', q) \\ b(j)=+1}} \alpha_{j-1}(q') \gamma_j^e(q', q) \beta_j(q)}{\sum_{\substack{(q', q) \\ b(j)=-1}} \alpha_{j-1}(q') \gamma_j^e(q', q) \beta_j(q)} \right), \end{aligned} \quad (92)$$

where

$$\gamma_j^e(q', q) = \exp \left( \sum_{v=1}^p \frac{1}{2} d'(j; v) L(d'(j; v)) \right). \quad (93)$$

The decoder must also compute  $L$ -values for the coded bits to be fed back to the equalizer,

$$\begin{aligned} L^D(\hat{d}'(j; v)) &= \log \left( \frac{P\{d'(j; v) = +1 | \mathbf{v}\}}{P\{d'(j; v) = -1 | \mathbf{v}\}} \right) \\ &= \log \left( \frac{\sum_{\substack{(q', q) \\ d'(j; v)=+1}} p(q', q, \mathbf{v})}{\sum_{\substack{(q', q) \\ d'(j; v)=-1}} p(q', q, \mathbf{v})} \right). \end{aligned} \quad (94)$$

These are computed in a similar fashion as the  $L$ -values for the information bits, yielding as final result

$$\begin{aligned} L^D(\hat{d}'(j; v)) &= L(b(j)) + L(d'(j; v)) \\ &+ \log \left( \frac{\sum_{\substack{(q', q) \\ d'(j; v)=+1}} \alpha_{j-1}(q') \gamma_j^c(q', q) \beta_j(q)}{\sum_{\substack{(q', q) \\ d'(j; v)=-1}} \alpha_{j-1}(q') \gamma_j^c(q', q) \beta_j(q)} \right), \end{aligned} \quad (95)$$

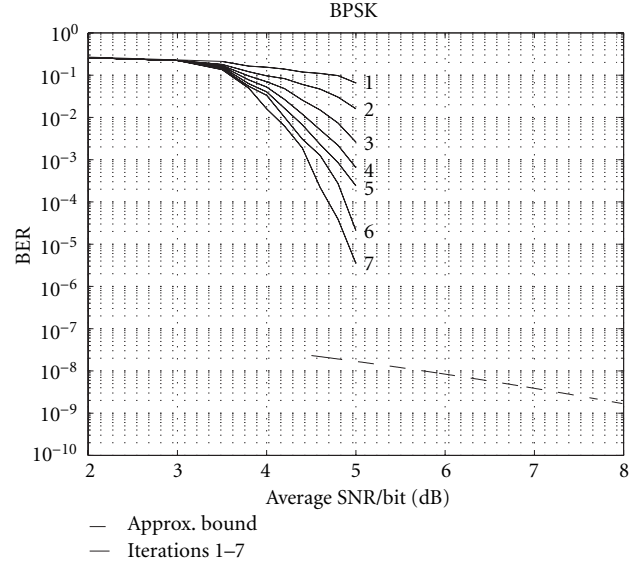


FIGURE 9: Performance of iterative MAP decoding for a (2, 2) system in flat fading ( $L = 0$ ). An approximate bound is shown for intermediate SNRs.  $R_s = 1/2$ ,  $K = 512$ .

where

$$\gamma_j^c(q', q) = \exp \left( \sum_{\substack{\mu=1 \\ \mu \neq v}}^p \frac{1}{2} d'(j; \mu) L(d'(j; \mu)) \right). \quad (96)$$

Again, we use the equalizer outputs  $L_e^E(\hat{d}'(j; v))$  as estimates for  $L(d'(j; v))$ ,  $v = 1, 2, \dots, p$ , and  $L(b(j))$  is the a priori  $L$ -value for the information bit  $b(j)$ . As we assume equiprobable information bits in our case,  $L(b(j)) = 0$  in both (92) and (95).

### 6.3. Performance of iterative receivers

To evaluate the performance of the iterative MIMO receiver, we performed computer simulations of a (2, 2) system which uses a rate-1/2, 4-state convolutional code with generator polynomial  $(5, 7)_8$  and  $d_{\text{free}}^o = 5$ , as outer code. The coded bits were passed through a pseudo-random interleaver of length  $K = 512$ , serial-to-parallel converted and differentially encoded prior to transmission over a (2, 2) fading channel. All simulations were carried out with BPSK modulation and with perfect knowledge of the channel coefficients. The channel fading was independent from symbol interval to symbol interval, and the variance of the channel coefficients was  $\sigma_f^2 = 1/D$ .

Figure 9 shows the BER performance in flat fading ( $L = 0$ ), as well as the approximate analytical bound developed in Section 5. Simulation results obtained with 1–7 iterations are shown. The performance is determined by interleaving gain as well as spatial diversity gain of order  $M = 2$ . The SNR corresponding to the cut-off rate for the rate-1/2 code in Rayleigh fading (without spatial diversity) is  $\bar{\gamma}_b = 4.52$  dB.

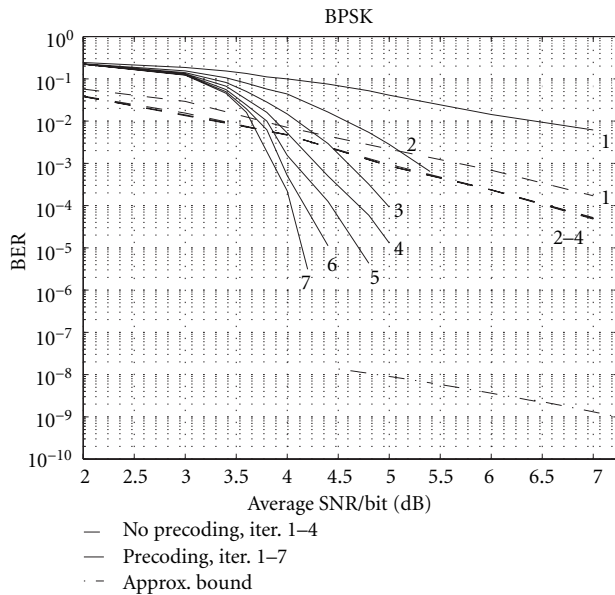


FIGURE 10: Performance of iterative MAP equalization and decoding for a (2, 2) system in a two-path fading channel with and without recursive precoding. An approximate bound is shown for intermediate SNRs.  $R_s = 1/2$ ,  $K = 512$ .

Figure 10 contrasts the performance in a two-path ( $L = 1$ ) fading channel with and without differential precoding. In particular, the dashed curves show the BER results obtained without precoding. In this case, 1–4 iterations were performed. The performance improves only by approximately 1 dB after the first iteration and converges thereafter. Evidently, there is no interleaving gain available since the inner “code” is nonrecursive. The solid curves show the BER performance *with* differential precoding. In this case, the performance continues to show improvement, even after seven iterations. We note that the BER exhibits a dramatic drop at  $\bar{\gamma}_b \approx 4$  dB, which indicates that interleaving gain due to the recursive inner “code” affects the performance. It is worth mentioning that recursive precoding results in a performance loss during the first iteration which amounts to approximately 3 dB, relative to the first iteration of the nonprecoded case. This is to be expected, as differential coding is known to incur a 3 dB performance loss. Also shown is the approximate analytical bound developed in Section 5.

## 7. CONCLUSIONS

In this paper, we have presented a general system and channel model for coded MIMO wireless systems that use multiple transmit and receive antennas. Multiple transmit antennas are used for the purpose of increasing the data rate, while coding and multiple receive antennas are employed to improve the performance in fading multipath channels by introducing signal diversity. For uncoded systems, we have examined the optimal MIMO MLSE detector and the suboptimal MIMO linear and decision feedback equalizers.

These detectors are fairly straightforward generalizations of their well-known single-input, single-output predecessors. We have also considered the case of flat fading and presented detectors targeted towards this special case. We have analyzed the performance of the maximum likelihood detector and found that it is capable of fully exploiting both the explicit diversity due to multiple receive antennas and implicit diversity due to multipath propagation. On the other hand, the suboptimal detectors are not capable of fully exploiting the diversity inherent in the channel and require many more receive antennas to achieve comparable performance.

For coded MIMO systems, we have presented a theoretical analysis of the bit error probability, assuming maximum likelihood decoding. We have also presented a more practical iterative equalization and decoding scheme based on the BCJR MAP algorithm, and evaluated its performance through computer simulations. We have seen that by introducing differential precoding, thus translating the channel into a recursive channel, significant interleaving gains can be realized compared to systems without precoding. The proposed receiver structure is applicable to any MIMO system which uses a convolutional code. However, as the complexity of trellis-based MIMO algorithms grows exponentially with the number of transmit antennas as well as the channel memory, it is of interest to design receivers of lower complexity. Methods for complexity reduction already developed for single-input, single-output systems may be appropriately modified for this purpose.

## ACKNOWLEDGMENT

This work was supported in part by the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564.

## REFERENCES

- [1] G. J. Foschini and M. J. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.
- [2] D. Gesbert, H. Bölcskei, D. Gore, and A. Paulraj, “MIMO wireless channels: capacity and performance prediction,” in *Proc. Globecom 2000, IEEE Global Commun. Conference*, pp. 1083–1088, San Francisco, Calif, USA, November 2000.
- [3] C. C. Martin, J. H. Winters, and N. R. Sollenberger, “Multiple-input multiple-output (MIMO) radio channel measurements,” in *Proc. IEEE Vehicular Technology Conference (VTC-2000)*, Boston, Mass, USA, 2000.
- [4] D. P. McNamara, M. A. Beach, P. Karlsson, and P. N. Fletcher, “Initial characterization of multiple-input multiple-output (MIMO) channels for space-time communication,” in *Proc. IEEE Vehicular Technology Conference (VTC-2000)*, Boston, Mass, USA, 2000.
- [5] H. E. Nichols, A. A. Giordano, and J. G. Proakis, “MLD and MSE algorithms for adaptive detection of digital signals in the presence of interchannel interference,” *IEEE Transactions on Information Theory*, vol. 23, no. 5, pp. 563–575, 1977.
- [6] J. Salz, “Digital transmission over cross-coupled linear channels,” *AT&T Technical Journal*, vol. 64, no. 6, pp. 1147–1159, 1985.



- [7] A. Duel-Hallen, "Equalizers for multiple input/output channels and PAM systems with cyclostationary input sequences," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 3, pp. 630–639, 1992.
- [8] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. on Communication*, vol. 46, no. 3, pp. 357–366, 1998.
- [9] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Lab. Tech. Journal*, vol. 1, no. 2, pp. 41–59, 1996.
- [10] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–15, 1999.
- [11] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI International Symposium on Signals, Systems, and Electronics (ISSSE-98)*, pp. 295–300, Pisa, Italy, 1998.
- [12] X. Li, H. C. Huang, A. Lozano, and G. J. Foschini, "Reduced-complexity detection algorithms for systems using multi-element arrays," in *Proc. Globecom 2000, IEEE Global Commun. Conference*, San Francisco, Calif, USA, 2000.
- [13] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal BLAST processing with multiple-access interference," in *Proc. IEEE Vehicular Technology Conference (VTC-2000)*, Boston, Mass, USA, 2000.
- [14] 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*, vol. 44, no. 2, pp. 744–765, 1998.
- [15] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 451–460, 1999.
- [16] A. Stefanov and T. M. Duman, "Turbo coded modulation for systems with transmit and receive antenna diversity," in *Proc. Globecom 99, IEEE Global Commun. Conference*, pp. 2336–2340, Rio De Janeiro, Brazil, 1999.
- [17] C. Douillard, M. Jézéquel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: turbo-equalization," *European Trans. on Telecommun.*, vol. 6, no. 5, pp. 507–511, 1995.
- [18] G. Bauch and A. F. Naguib, "MAP equalization of space-time coded signals over frequency selective channels," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC-99)*, pp. 261–265, New Orleans, La, USA, September 1999.
- [19] G. Bauch and N. Al-Dahir, "Reduced-complexity turbo-equalization with multiple transmit and receive antennas over multipath fading channels," in *Proc. Conference on Information Sciences and Systems (CISS-2000)*, pp. WP 3–13–WP 3–18, Princeton, NJ, USA, March 2000.
- [20] G. Bauch and N. Al-Dahir, "Iterative equalization and decoding with channel shortening filters for space-time coded modulation," in *Proc. IEEE Vehicular Technology Conference (VTC-2000)*, pp. 1575–1582, Boston, Mass, USA, September 2000.
- [21] S. L. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. Communications*, vol. 48, no. 8, pp. 1347–1359, 2000.
- [22] S. B ro, G. Bauch, A. Pavlic, and A. Semmler, "Improving BLAST performance using space-time block codes and turbo decoding," in *Proc. IEEE Globecom 2000*, pp. 1067–1071, San Francisco, Calif, USA, November 2000.
- [23] A. M. Tonello, "Space-time bit-interleaved coded modulation over frequency selective fading channels with iterative decoding," in *Proc. IEEE Globecom 2000*, San Francisco, Calif, USA, November 2000.
- [24] G. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, no. 3, pp. 363–378, 1972.
- [25] W. van Etten, "Maximum likelihood receiver for multiple channel transmission systems," *IEEE Trans. Communications*, vol. 24, no. 2, pp. 276–283, 1976.
- [26] J. G. Proakis, "Performance capabilities of the Viterbi algorithm for combatting intersymbol interference on fading multipath channels," in *Communication Systems and Random Process Theory*, pp. 403–424, Sijthoff and Noordhoff, 1978.
- [27] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 3rd edition, 1995.
- [28] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques*, McGraw-Hill, New York, NY, USA, 1966.
- [29] B. A. Bjerke and J. G. Proakis, "Multichannel antenna diversity combining techniques for fading multipath channels," in *Proc. Conference on Information Sciences and Systems (CISS-2000)*, pp. FA4–1–FA4–6, Princeton, NJ, USA, March 2000.
- [30] M. Stojanovic and Z. Zvonar, "Multichannel processing of broadband multiuser communication signals in shallow water acoustic channels," *IEEE Journal of Oceanic Engineering*, vol. 21, no. 2, pp. 156–166, 1996.
- [31] S. J. Grant and J. K. Cavers, "Performance enhancement through joint detection of cochannel signals using diversity arrays," *IEEE Trans. Communications*, vol. 46, no. 8, pp. 1038–1049, 1998.
- [32] B. A. Bjerke and J. G. Proakis, "Multiple-antenna diversity techniques for transmission over fading channels," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC-99)*, New Orleans, La, USA, September 1999.
- [33] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 909–926, 1998.
- [34] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," The telecommunications and mission operations progress report 42–139, Jet Propulsion Laboratory, Pasadena, Calif, USA, November 1999.
- [35] A. J. Viterbi, A. M. Viterbi, J. Nicolas, and N. T. Sindhushayana, "Perspectives on interleaved concatenated codes with iterative soft-output decoding," in *Proc. 1st International Symposium on Turbo Codes and Related Topics*, pp. 47–54, Brest, France, September 1997.
- [36] K. R. Narayanan, "Design considerations for coded systems with turbo equalization," in *Proc. Conference on Information Sciences and Systems (CISS-2000)*, pp. TP7–17–TP7–22, Princeton, NJ, USA, 2000.
- [37] P. K. Gray, *Serially Concatenated Trellis Coded Modulation*, Ph.D. dissertation, University of South Australia, Australia, August 1999.
- [38] S. Shamai and I. Sason, "Variations on Gallager's bounding techniques: performance bounds for turbo codes in Gaussian and fading channels," in *Proc. 2nd International Symposium on Turbo Codes and Related Topics*, pp. 27–34, Brest, France, September 2000.
- [39] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting and decoding: turbo codes," in *Proc. IEEE International Conference on Communication (ICC-93)*, pp. 1064–1070, Geneva, Switzerland, 1993.

- [40] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, 1974.
- [41] W. E. Ryan, "A turbo code tutorial," unpublished paper, <http://citeseer.nj.nec.com/ryan97turbo.html>.
- [42] G. Bauch, H. Khorram, and J. Hagenauer, "Iterative equalization and decoding in mobile communications systems," in *Proc. 2nd European Personal Mobile Communications Conference (EPMCC-97)*, pp. 307–312, Bonn, Germany, September 1997.

**Bjørn A. Bjerke** received the Siv. Ing. degree in electrical engineering from the Norwegian Institute of Technology (NTH), Trondheim, Norway, in 1995 and the M.S. and Ph.D. degrees from Northeastern University, Boston, Mass, USA in 1997 and 2001, respectively, both in electrical engineering. He is currently a Senior Engineer with Qualcomm, Inc., Concord, Mass, USA, where he focuses on physical layer algorithms and architectures for multi-antenna wireless communications.



**John G. Proakis** received the BSEE from the University of Cincinnati in 1959, the MSEE from MIT in 1961, and the Ph.D. from Harvard University in 1967. He is currently an Adjunct Professor at the University of California at San Diego and a Professor Emeritus at Northeastern University. From 1969 to 1998 he was a faculty member at the ECE department of Northeastern University and he held the positions of Department Chair (1984–1997), Associate Dean and Director of the Graduate School of Engineering (1982–1984), and Acting Dean (1992–1993). Prior to joining Northeastern University, he worked at GTE Laboratories and the MIT Lincoln Laboratory. His professional experience and interests are in the general areas of digital communications and digital signal processing. He is the author of the book *Digital Communications* (New York: McGraw-Hill, 1983, 1st ed.; 1989, 2nd ed.; 1995, 3rd ed.; 2001, 4th ed.), and co-author of the books, *Introduction to Digital Signal Processing* (New York: Macmillan, 1988, 1st ed.; 1992, 2nd ed.; 1996, 3rd ed.); *Digital Signal Processing Laboratory* (Englewood Cliffs: Prentice Hall, 1991); *Advanced Digital Signal Processing* (New York: Macmillan, 1992); *Algorithms for Statistical Signal Processing* (Englewood Cliffs: Prentice Hall, 2002); *Discrete-Time Processing of Speech Signals* (New York: Macmillan, 1992, IEEE Press, 2000); *Communication Systems Engineering*, (Englewood Cliffs: Prentice Hall, 1994, 1st ed.; 2002, 2nd ed.); *Digital Signal Processing Using MATLAB V.4* (Boston: Brooks/Cole-Thomson Learning, 1997, 2000); *Contemporary Communication Systems Using MATLAB* (Boston: Brooks/Cole-Thomson Learning, 1998, 2000). Dr. Proakis holds five patents and has published over 150 papers.

