Space–Time Decoding With Imperfect Channel Estimation

Giorgio Taricco, *Senior Member, IEEE*, and Ezio Biglieri, *Fellow, IEEE*

***Abstract*—Under the assumption of a frequency-flat slow Rayleigh fading channel with multiple transmit and receive antennas, we examine the effects of imperfect estimation of the channel parameters on error probability when known pilot symbols are transmitted among information data. Three different receivers are considered. The first one derives an estimate of the channel [by using either a maximum-likelihood (ML) or a minimum mean square error (MMSE) criterion], and then uses this estimate in the same metric that would be applied if the channel were perfectly known. The second receiver derives again an estimate of the channel, but uses the ML metric conditioned on the channel estimate. Our last receiver simultaneously processes the pilot and data symbols received. Simulation results are exhibited, showing that only a relatively small percentage of the transmitted frame need be allocated to pilot symbols in order to experience an acceptable degradation of error probability due to imperfect channel knowledge. Algorithms for the recursive calculation of the decision metric of the last two receivers are also developed for application to sequential decoding of trellis space–time codes.**

***Index Terms*—Channel estimation, fading channels, MIMO systems, space–time codes, wireless communications.**

I. Introduction and Motivation of the Work

A typical scenario for multiple-input multiple-output (MIMO) radio communication systems assumes a channel changing so slowly that an entire frame can be transmitted without any appreciable variation in channel characteristics [7]. If this assumption (also known as block or quasistatic fading channel) is valid, the system performance can be enhanced if the receiver is made aware of the so-called channel state information (CSI), i.e., the realization of the random fading gains affecting the transmission paths from each transmit to each receive antenna. To this purpose, a portion of the transmitted frame may consist of pilot symbols, whose composition is known to the receiver and is used by the latter to estimate the channel parameters. Due to noise and to the finite number of pilot symbols in a frame, the channel estimate is not perfect: The main purpose of this paper is to investigate the effects that this imperfect estimation has on system performance.

A similar investigation was carried out in [4] for scalar channels. An information-theoretic analysis of the capacity of the MIMO channel with imperfect side information generated by the estimation procedure can be found in [11] and [26]. There, the effect of capacity reduction due to the transmission of training symbols carrying no information is accounted for. MIMO channels with a BLAST receiver interface and orthogonal training signals were examined in [18]; it was shown that the training sequence length required to control the estimation-error-induced outage probability is approximately proportional to the number of transmit antennas, and is independent of the number of receive antennas. Tarokh *et al.* [22] examine space–time codes and their design criteria in the presence of channel estimation errors, but their analysis is partially invalidated by a flaw in the derivation of their (2) (see below and [23]). If the channel fades so rapidly that reliable channel estimation is impossible, one may choose a reception scheme that assumes a known probabilistic model for the channel, but no knowledge of the actual channel state [12].

Recent work in this area can be classified in two main streams: OFDM and narrowband techniques. A rich literature exists in the former area, basically addressing the problem of estimating the channel matrices with different goodness criteria. As an example, [17] derives the condition required for optimal training sequences, while [24] addresses the specific case of OFDM and carrier frequency offset. Reference [20] presents a training-based MIMO channel estimation scheme operating in wideband frequency-selective fading with layered space–time coding and using a pilot matrix to resolve both intersymbol and intercarrier interference. Narrowband studies include [25], where an uncoded V-BLAST system is considered, and [14], where a semiblind channel estimation technique aims at minimizing the mean square error. Reference [8] exhibits results [in terms of error rate versus signal-to-noise ratio (SNR)] with the Alamouti code and a mismatched receiver with minimum mean square error (MMSE) CSI estimation. Recently, Larsson [16] (see also [15, Sec. 4.2.3]) showed independently that the diversity gain attained by pilot-aided channel estimation based on a MMSE procedure and minimum distance detection is the same as that of a maximum-likelihood (ML) receiver with ideal channel estimation.

Our work addresses the case of narrowband MIMO systems, with the aim of finding the error probability degradation of three receiver architectures using standard space–time codes.

This paper is organized as follows. Section II describes the MIMO channel model. Section III illustrates three types of detection schemes for MIMO channels. They are the mismatched receiver, the ML receiver, and the optimum receiver. The first two are based on a channel matrix estimate obtained by ML or MMSE criteria. Section IV studies the ML receiver in detail: We exhibit the receiver metric, an algorithm for its sequential computation, and show that the receiver output does not depend on the type of channel matrix estimation considered.

Manuscript received September 1, 2003; revised January 26, 2004; accepted June 4, 2004. The editor coordinating the review of this paper and approving it for publication is C. Xiao. This work was supported by CERCOM.

The authors are with Politecnico di Torino, Torino 10129, Italy (e-mail: giorgio.taricco@polito.it; ezio.biglieri@polito.it).

Digital Object Identifier 10.1109/TWC.2005.850324

Limiting cases are considered and compared with known results from the literature. Section V studies the optimum receiver in detail, providing the receiver metric and an algorithm for its sequential computation. This receiver is not based on an explicit estimation of the channel matrix, but bases its decisions directly on the pilot signals and the receiver data signal. Section VI studies the asymptotic (in the SNR) error performance and the diversity gain of the receivers considered with orthogonal pilot matrices. Section VII includes numerical results: Specifically, Section VII-A considers a simple space–time block code; Section VII-B considers more elaborate trellis space–time codes; and Section VII-C compares the results obtained in this paper with information-theoretical results adapted from [10], [11]. Finally, Section VIII summarizes our findings.

II. MIMO Channel Model

We consider a radio system with t transmit and r receive antennas, $r \geq t$. Assume that a space–time code $\mathcal{X}$ is used with block length N. Specifically, the transmitted signal is represented by a $t \times N$ matrix $\mathbf{X} \triangleq (\mathbf{x}_1, \ldots, \mathbf{x}_N)$. The received signal is modeled by an $r \times N$ matrix

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}. \tag{1}$$

Here, the channel is described by the $r \times t$ complex random matrix $\mathbf{H}$ (referred to in the following as channel matrix or CSI) whose entries are independent identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian (ZMCSCG) random variables (i.e., each random variable has independent real and imaginary parts with zero mean and same variance) with unit variance (i.e., twice the variance of the real or imaginary parts). The noise matrix $\mathbf{Z}$ consists of ZMCSCG random variables with variance N_0 so that the noise affecting the received signal is spatially and temporally white, with $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\dagger] = NN_0\mathbf{I}_r$, where $\mathbf{I}_r$ denotes the $r \times r$ identity matrix and $(\cdot)^\dagger$ denotes Hermitian transposition. The average energy of the entries of the transmitted signal is

$$E_\mathrm{s} \triangleq \frac{1}{tN}\mathbb{E}\left[\|\mathbf{X}\|^2\right] = \frac{1}{tN}\mathbb{E}\left[\mathrm{tr}(\mathbf{X}\mathbf{X}^\dagger)\right] \tag{2}$$

where the expectation is taken over the *a priori* distribution of the transmitted signal code words, and where $\mathrm{tr}(\mathbf{A})$ denotes the matrix trace, i.e., $\mathrm{tr}(\mathbf{A}) \triangleq \sum_i \mathbf{A}_{ii}$, and $\|\mathbf{A}\|$ the Frobenius norm, i.e., $\|\mathbf{A}\| \triangleq (\sum_{i,j} |\mathbf{A}_{ij}|^2)^{1/2}$. The channel matrix $\mathbf{H}$ is independent of both $\mathbf{X}$ and $\mathbf{Z}$, and remains constant during the transmission of an entire code word.

III. Pilot-Based Detection Schemes

In a large number of studies, it is assumed that the receiver has perfect knowledge of the channel matrix $\mathbf{H}$ and code design criteria are derived using this assumption (see, e.g., [22]). However, in a real communication environment, the receiver has no *a priori* knowledge of the realization of $\mathbf{H}$ and has to estimate it using the available received data samples. A standard technique to allow the receiver to estimate $\mathbf{H}$ consists of transmitting pilot symbols among the data, i.e., a set of symbols whose location and values are known to the receiver. Without loss of generality (after possibly rearranging the order of transmitted symbols), we assume that the transmitter sends a pilot matrix $\mathbf{X}_\mathrm{p}$ and a data matrix $\mathbf{X}$, both affected by the same channel matrix $\mathbf{H}$, so that the receiver separately observes the matrices

$$\mathbf{Y}_\mathrm{p} = \mathbf{H}\mathbf{X}_\mathrm{p} + \mathbf{Z}_\mathrm{p} \tag{3}$$

and

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z} \tag{4}$$

where $\mathbf{Z}_\mathrm{p}$ and $\mathbf{Z}$ are the noise matrices affecting the transmission of data and pilot symbols, respectively. $\mathbf{Z}$ and $\mathbf{Z}_\mathrm{p}$ have the same distribution. We also assume that $\mathbf{X}_\mathrm{p}$ and $\mathbf{X}$ are $t \times P$ and $t \times N$ matrices, respectively, and the average pilot symbol energy is

$$E_\mathrm{p} \triangleq \frac{1}{tP}\|\mathbf{X}_\mathrm{p}\|^2 = \frac{1}{tP}\mathrm{tr}\left(\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger\right). \tag{5}$$

Since to estimate the $r \times t$ matrix $\mathbf{H}$, we need at least rt independent measurements, and each symbol time yields r measurements at the receiver, we must have $P \geq t$. Moreover, the matrix $\mathbf{X}_\mathrm{p}$ must have full rank t, since otherwise t linearly independent columns would not be available to yield rt independent measurements. As a consequence of the above, the matrix $\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger$ must be nonsingular.

Among the receiver structures that can be envisaged we focus on the following:

1) The receiver estimates the channel matrix $\mathbf{H}$ from $\mathbf{Y}_\mathrm{p}$ and $\mathbf{X}_\mathrm{p}$ and uses the result $\widehat{\mathbf{H}}$ in the same metric that would be applied if the channel were perfectly known. The output is

$$\widehat{\mathbf{X}} \triangleq \arg\min_{\mathbf{X}} \widetilde{\mu}(\mathbf{X}) \tag{6}$$

where

$$\widetilde{\mu}(\mathbf{X}) \triangleq \|\mathbf{Y} - \widehat{\mathbf{H}}\mathbf{X}\|^2. \tag{7}$$

Since (7) is sometimes referred to as mismatched metric, we call this a mismatched receiver. There are several possible mismatched receivers, depending on the type of CSI estimation criterion.

2) The receiver estimates the channel matrix $\mathbf{H}$ from $\mathbf{Y}_\mathrm{p}$ and $\mathbf{X}_\mathrm{p}$ and uses the result $\widehat{\mathbf{H}}$ to detect the transmitted signal $\mathbf{X}$ according to an ML criterion. The output is

$$\widehat{\mathbf{X}} \triangleq \arg\max_{\mathbf{X}} p(\mathbf{Y}|\mathbf{X}, \widehat{\mathbf{H}}). \tag{8}$$

With a slight abuse of terminology, we call this an ML receiver. There are several possible ML receivers, depending on the type of CSI estimation criterion. However, we will show in Section IV that ML and MMSE CSI estimation yield the same ML receiver.

3) The receiver detects the transmitted signal $\mathbf{X}$ by jointly processing $\mathbf{Y}$, $\mathbf{Y}_{\rm p}$, and $\mathbf{X}_{\rm p}$ (as suggested, e.g., in [11]) without explicit estimation of $\mathbf{H}$. In this case, the detection problem can be written as

$$\widehat{\mathbf{X}} \triangleq \arg\max_{\mathbf{X}} p(\mathbf{Y}, \mathbf{Y}_{\rm p}|\mathbf{X}, \mathbf{X}_{\rm p}). \tag{9}$$

We refer to this detection scheme as the optimum receiver.

The mismatched and ML receivers are more common in the literature [see [4] for single-input single-output (SISO) ML receiver]. They share the need of recovering CSI in order to detect the transmitted signal. The optimum receiver skips the CSI recovery step, and focuses instead on the detection of the transmitted signal. The performances of these receivers are studied and compared in the following.

A. CSI Estimation

Here, we focus on ML and MMSE estimation of $\mathbf{H}$ from the pilot signals $\mathbf{X}_{\rm p}$ and $\mathbf{Y}_{\rm p}$.

1) The ML estimate of $\mathbf{H}$ is obtained by maximizing $p(\mathbf{Y}_{\rm p}|\mathbf{H}, \mathbf{X}_{\rm p})$ or, equivalently, by minimizing $\|\mathbf{Y}_{\rm p} - \mathbf{H}\mathbf{X}_{\rm p}\|^2$ with respect to $\mathbf{H}$. This yields

$$\widehat{\mathbf{H}}_{\rm ML} = \mathbf{Y}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\left(\mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\right)^{-1} = \mathbf{H} + \mathbf{E} \tag{10}$$

where

$$\mathbf{E} \triangleq \mathbf{Z}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\left(\mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\right)^{-1} \tag{11}$$

denotes the CSI estimation error matrix. Here, we assume that $(\mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger})^{-1}$ exists, consistently with remarks above.

2) An MMSE estimate of $\mathbf{H}$ can be obtained by the linear transformation $\mathbf{Y}_{\rm p}\mathbf{F}$, with $\mathbf{F}$ the $P \times t$ matrix that minimizes the mean square error $\mathbb{E}\|\mathbf{Y}_{\rm p}\mathbf{F} - \mathbf{H}\|^2$. This yields

$$\begin{aligned}\widehat{\mathbf{H}}_{\rm MMSE} &= \mathbf{Y}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\left(N_0\mathbf{I}_t + \mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\right)^{-1}\\ &= \mathbf{H}\mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\left(N_0\mathbf{I}_t + \mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\right)^{-1}\\ &\quad + \mathbf{Z}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\left(N_0\mathbf{I}_t + \mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\right)^{-1}\end{aligned} \tag{12}$$

where $\mathbf{I}_t$ denotes the $t \times t$ identity matrix.

We notice that

$$\widehat{\mathbf{H}}_{\rm MMSE} = \widehat{\mathbf{H}}_{\rm ML}\mathbf{B}_{\rm MMSE} \tag{13}$$

where $\mathbf{B}_{\rm MMSE} \triangleq \mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}(N_0\mathbf{I}_t + \mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger})^{-1}$. Since $\widehat{\mathbf{H}}_{\rm ML}$ is plainly an unbiased estimator of $\mathbf{H}$, we will call $\mathbf{B}_{\rm MMSE}$ biasing matrix. This leads us to consider the general class of CSI estimators

$$\widehat{\mathbf{H}} = \widehat{\mathbf{H}}_{\rm ML}\mathbf{B} \tag{14}$$

characterized by an invertible biasing matrix $\mathbf{B}$. In Section IV, we show that the ML receiver output does not depend on $\mathbf{B}$ so that ML and MMSE estimation of the CSI lead to the same result in that case.

Analysis of the ML receiver (see Section IV) requires the statistics of the error matrix $\mathbf{E}$ defined in (11). First, it is plain to see that $\mathbf{H}$ and $\mathbf{E}$ are independent. Next, denoting by $(\cdot)_i$ the ith row of a matrix $(\cdot)$, we can write

$$\mathbf{E}_i = (\mathbf{Z}_{\rm p})_i\mathbf{X}_{\rm p}^{\dagger}\left(\mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\right)^{-1}. \tag{15}$$

Then, the rows of $\mathbf{E}$ are independent vectors of ZMCSCG random variables with covariance matrix

$$\boldsymbol{\Sigma}_{\rm e} \triangleq \mathbb{E}\left[\mathbf{E}_i^{\dagger}\mathbf{E}_i\right] = N_0\left(\mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger}\right)^{-1}. \tag{16}$$

This suggests the use of an orthogonal pilot matrix[1] yielding a white error matrix (the entries of $\mathbf{E}$ are i.i.d. ZMCSCG random variables with variance $N_0/(PE_{\rm p})$ and $\boldsymbol{\Sigma}_{\rm e} = N_0\mathbf{I}_t/(PE_{\rm p})$). For example, in [26], orthogonality is achieved by a diagonal $\mathbf{X}_{\rm p}$, which corresponds to having only one transmit antenna active at a time.[2]

IV. ML Receiver

The solution of the maximization problem (8) requires an explicit expression of the conditional probability density function (pdf) $p(\mathbf{Y}|\mathbf{X}, \widehat{\mathbf{H}})$, where we assume $\widehat{\mathbf{H}}$ to be given in its general form (14). This can be obtained since both $\mathbf{Y}$ and $\widehat{\mathbf{H}}$ are ZMCSCG random matrices and we can apply known results such as Theorem 1 of the Appendix I.

In this section, we derive the receiver metric for the detection of the transmitted signal, we show that this metric is independent of the type of CSI estimation assumed (ML or MMSE), and we provide a sequential algorithm for its computation.

A. Receiver Metric

The pdf required to calculate the receiver metric is derived in Appendix II for general pilot matrices. In particular, with orthogonal pilot matrices, we obtain[3]

$$p(\mathbf{Y}|\mathbf{X}, \widehat{\mathbf{H}}) = \frac{\operatorname{etr}\left\{-(\mathbf{Y}-\rho\widehat{\mathbf{H}}\mathbf{X})\left[N_0\mathbf{I}_N + (1-\rho)\mathbf{X}^{\dagger}\mathbf{X}\right]^{-1}(\mathbf{Y}-\rho\widehat{\mathbf{H}}\mathbf{X})^{\dagger}\right\}}{\det\{\pi\left[N_0\mathbf{I}_N + (1-\rho)\mathbf{X}^{\dagger}\mathbf{X}\right]\}^r}, \tag{17}$$

[1]Hereafter, we refer to a pilot matrix $\mathbf{X}_{\rm p}$ with orthogonal rows (i.e., such that $\mathbf{X}_{\rm p}\mathbf{X}_{\rm p}^{\dagger} = PE_{\rm p}\mathbf{I}_t$) by the name orthogonal pilot matrix.

[2]Orthogonality, albeit attractive, may be incompatible with standard signal constellations such as PSK. In fact, it is desirable that the rows of $\mathbf{X}_{\rm p}$ have good autocorrelation and cross correlation properties. This is achieved by perfect root of unity sequences only in some special cases (see [6] and references therein).

[3]We define $\operatorname{etr}(\mathbf{A}) \triangleq \exp[\operatorname{tr}(\mathbf{A})]$.

where ρ is defined as

$$\rho \triangleq \frac{PE_{\mathrm{p}}}{N_0 + PE_{\mathrm{p}}}. \tag{18}$$

Taking logarithms and dropping constant terms yields the ML estimate

$$\widehat{\mathbf{X}} = \arg\min_{\mathbf{X}} \mu_{\mathrm{ML}}(\mathbf{X}) \tag{19}$$

where

$$\mu_{\mathrm{ML}}(\mathbf{X}) \triangleq rN_0 \ln\det\left(\mathbf{I}_N + \frac{(1-\rho)\mathbf{X}^\dagger\mathbf{X}}{N_0}\right) + \mathrm{tr}\left[(\mathbf{Y}-\rho\widehat{\mathbf{H}}\mathbf{X})\left(\mathbf{I}_N + \frac{(1-\rho)\mathbf{X}^\dagger\mathbf{X}}{N_0}\right)^{-1}(\mathbf{Y}-\rho\widehat{\mathbf{H}}\mathbf{X})^\dagger\right] \tag{20}$$

is the metric of $\mathbf{X}$. Specializing this result to extremal cases we get:

1) with perfect CSI (i.e., $PE_{\mathrm{p}}/N_0 \to \infty$ and consequently $\rho \to 1$)

$$\mu_{\mathrm{ML}}(\mathbf{X}) = \|\mathbf{Y} - \widehat{\mathbf{H}}\mathbf{X}\|^2 \tag{21}$$

2) without CSI (i.e., $PE_{\mathrm{p}}/N_0 \to 0$ and consequently $\rho \to 0$)

$$\mu_{\mathrm{ML}}(\mathbf{X}) = rN_0 \ln\det\left(\mathbf{I}_N + \frac{\mathbf{X}^\dagger\mathbf{X}}{N_0}\right) + \mathrm{tr}\left[\mathbf{Y}\left(\mathbf{I}_N + \frac{\mathbf{X}^\dagger\mathbf{X}}{N_0}\right)^{-1}\mathbf{Y}^\dagger\right] \tag{22}$$

consistently with [12].[4] It is worth noting that the metric expression does not contain the CSI estimate $\widehat{\mathbf{H}}$ in this case, consistently with the fact that there is no available CSI.

B. Iterative Metric Computation

Here, we consider the problem of calculating the metric (20) for implementation of a sequential decoding algorithm such as the Viterbi decoder. For simplicity, we restrict ourselves to consider the case of orthogonal pilot matrices. Assuming the transmitted and received signal matrices to be split as $\mathbf{X} = (\mathbf{X}^-, \mathbf{x})$ and $\mathbf{Y} = (\mathbf{Y}^-, \mathbf{y})$ (where $\mathbf{x}$ and $\mathbf{y}$ are column vectors), we obtain the following result (see Appendix III):

$$\begin{aligned}\Delta\mu_{\mathrm{ML}}(\mathbf{x};\mathbf{X}^-) &\triangleq \mu_{\mathrm{ML}}(\mathbf{X}) - \mu_{\mathrm{ML}}(\mathbf{X}^-)\\ &= \|\mathbf{y} - \rho\widehat{\mathbf{H}}\mathbf{x}\|^2 - rN_0\ln\left(1 + \frac{(1-\rho)\mathbf{x}^\dagger\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{x}}{N_0}\right)\\ &\quad - \frac{1-\rho}{N_0}\mathrm{tr}\left[\mathbf{\Xi}(\mathbf{X})^\dagger\mathbf{\Lambda}(\mathbf{X})\mathbf{\Xi}(\mathbf{X}) - \mathbf{\Xi}(\mathbf{X}^-)^\dagger\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{\Xi}(\mathbf{X}^-)\right]\end{aligned} \tag{23}$$

where

$$\begin{cases}\widehat{\mu}(\mathbf{X}) \triangleq \|\mathbf{Y} - \rho\widehat{\mathbf{H}}\mathbf{X}\|^2\\ \qquad = \widehat{\mu}(\mathbf{X}^-) + \|\mathbf{y} - \rho\widehat{\mathbf{H}}\mathbf{x}\|^2\\ \mathbf{\Xi}(\mathbf{X}) \triangleq \mathbf{X}(\mathbf{Y} - \rho\widehat{\mathbf{H}}\mathbf{X})^\dagger\\ \qquad = \mathbf{\Xi}(\mathbf{X}^-) + \mathbf{x}(\mathbf{y} - \rho\widehat{\mathbf{H}}\mathbf{x})^\dagger\\ \mathbf{\Lambda}(\mathbf{X}) \triangleq \left(\mathbf{I}_t + \frac{(1-\rho)\mathbf{X}\mathbf{X}^\dagger}{N_0}\right)^{-1}\\ \qquad = \mathbf{\Lambda}(\mathbf{X}^-) - \frac{1-\rho}{N_0 + (1-\rho)\mathbf{x}^\dagger\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{x}}\\ \qquad\quad \cdot \mathbf{\Lambda}(\mathbf{X}^-)\mathbf{x}\mathbf{x}^\dagger\mathbf{\Lambda}(\mathbf{X}^-).\end{cases} \tag{24}$$

This equation can be applied to a sequential decoding algorithm and, in particular, used to calculate the branch metrics for a Viterbi decoder. Notice that the computational load involved in evaluating the branch metric is independent of the code word length N, and $O(t^2)$, i.e., it grows asymptotically with the square of t as $t \to \infty$. This is fundamental to the sequential implementation of the decoding algorithm for space–time trellis codes.

V. Optimum Receiver

Here, we consider the problem of detecting the transmitted word $\mathbf{X}$ maximizing the pdf $p(\mathbf{Y}, \mathbf{Y}_{\mathrm{p}}|\mathbf{X}, \mathbf{X}_{\mathrm{p}})$ without any prior estimate of the channel matrix $\mathbf{H}$ (though we assume that the joint pdf of the channel matrix $\mathbf{H}$ and the noise $\mathbf{Z}$ is known).

Applying Theorem 3 of Appendix I, we obtain (25), (shown at the bottom of the next page).

The logarithm of this expression yields the corresponding metric to be minimized by the optimum receiver

$$\mu_{\mathrm{opt}}(\mathbf{X}) = r\ln\det\left[\mathbf{I}_t + \frac{\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger}{N_0}\right] - \mathrm{tr}\left\{\left(\mathbf{Y}\mathbf{X}^\dagger + \mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\right)\left[\mathbf{I}_t + \frac{\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger}{N_0}\right]^{-1} \cdot \frac{\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger}{N_0^2}\right\}. \tag{26}$$

[4] An erroneous result is obtained if we set $\mu = 1$ (corresponding to $\rho = 0$ in our notation) in [22, eq. (3)]. This is caused by the invalid conditional independence assumption on the received vector components implicitly made in (2) therein (see also [23]).

If the pilot matrix $\mathbf{X}_\mathrm{p}$ is orthogonal, we have the surprising result that the metrics (20) and (26) are equivalent (see Appendix IV).

Caveat: The assumptions used in the derivation of metric (26) are that $\mathbf{H}$ and $\mathbf{Z}$ have i.i.d. entries distributed as $\mathcal{N}_\mathrm{c}(0,1)$ and $\mathcal{N}_\mathrm{c}(0,N_0)$, respectively. Whenever the channel fails to meet these assumptions, the metric is not optimum.

A. Iterative Metric Computation

Under the same assumptions of Section IV-B, we set $\mathbf{X}=(\mathbf{X}^-,\mathbf{x})$ and $\mathbf{Y}=(\mathbf{Y}^-,\mathbf{y})$. Then, we define

$$\begin{cases}\mathbf{\Xi}(\mathbf{X}) \triangleq \dfrac{\mathbf{X}\mathbf{Y}^\dagger+\mathbf{X}_\mathrm{p}\mathbf{Y}_\mathrm{p}^\dagger}{N_0}\\ \mathbf{\Lambda}(\mathbf{X}) \triangleq \left[\mathbf{I}_t+\dfrac{\mathbf{X}\mathbf{X}^\dagger+\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger}{N_0}\right]^{-1}\end{cases} \tag{27}$$

so that the metric can be written as

$$\mu_\mathrm{opt}(\mathbf{X}) = -r\ln\det\mathbf{\Lambda}(\mathbf{X}) - \mathrm{tr}\left[\mathbf{\Xi}(\mathbf{X})^\dagger\mathbf{\Lambda}(\mathbf{X})\mathbf{\Xi}(\mathbf{X})\right]. \tag{28}$$

After some algebra (similar to Section IV-B), we obtain

$$\begin{cases}\mathbf{\Xi}(\mathbf{X}) = \mathbf{\Xi}(\mathbf{X}^-)+\dfrac{\mathbf{x}\mathbf{y}^\dagger}{N_0}\\ \mathbf{\Lambda}(\mathbf{X}) = \mathbf{\Lambda}(\mathbf{X}^-)-\dfrac{\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{x}\mathbf{x}^\dagger\mathbf{\Lambda}(\mathbf{X}^-)}{N_0+\mathbf{x}^\dagger\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{x}}\\ \ln\det\mathbf{\Lambda}(\mathbf{X}) = \ln\det\mathbf{\Lambda}(\mathbf{X}^-) - \ln\left(1+\dfrac{\mathbf{x}^\dagger\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{x}}{N_0}\right)\end{cases} \tag{29}$$

and the metric increment is

$$\begin{aligned}\Delta\mu_\mathrm{opt}(\mathbf{x};\mathbf{X}^-) &\triangleq \mu_\mathrm{opt}(\mathbf{X})-\mu_\mathrm{opt}(\mathbf{X}^-)\\ &= r\ln\left(1+\frac{\mathbf{x}^\dagger\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{x}}{N_0}\right)\\ &\quad+\mathrm{tr}\left[\mathbf{\Xi}(\mathbf{X}^-)^\dagger\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{\Xi}(\mathbf{X}^-)\right.\\ &\qquad\left.-\mathbf{\Xi}(\mathbf{X})^\dagger\mathbf{\Lambda}(\mathbf{X})\mathbf{\Xi}(\mathbf{X})\right].\end{aligned} \tag{30}$$

VI. Diversity Gain

In this section, we investigate the asymptotic performance of the receivers considered as the SNR grows to infinity, and prove that the diversity gain achievable with perfect CSI is still attainable by using a pilot-aided scheme. A similar result was asserted in [22] for the ML receiver, but their proof is affected by an erroneous assumption of conditional independence of the received vector components [22, eq. (2)].

In our setting, the diversity gain can be defined as the limit

$$\gamma \triangleq \lim_{N_0\to 0}\frac{\log\mathbb{P}(e)}{\log N_0}$$

where $\mathbb{P}(e)$ is the symbol error probability, which can be approximated by the asymptotically tight "union bound"

$$\mathbb{P}(e) \le \frac{1}{|\mathcal{X}|}\sum_{\mathbf{X}\in\mathcal{X}}\sum_{\widehat{\mathbf{X}}\neq\mathbf{X}}\mathbb{P}(\mathbf{X}\to\widehat{\mathbf{X}}). \tag{31}$$

It is known [21] that if the pairwise error probability (PEP) [1, p. 190] is given by $\mathbb{P}(\mathbf{X}\to\widehat{\mathbf{X}})=\mathbb{E}[Q(\|\mathbf{H}\mathbf{\Delta}\|/(2N_0)^{1/2})]$ for a code word difference $\mathbf{\Delta}\triangleq\mathbf{X}-\widehat{\mathbf{X}}$, and consequently

$$\gamma = r\cdot\min_{\mathbf{\Delta}}\mathrm{rank}(\mathbf{\Delta}). \tag{32}$$

To simplify the calculations that follow, we assume orthogonal pilot matrices and define the $r\times N$ and $r\times t$ matrices $\widetilde{\mathbf{Z}}$ and $\widetilde{\mathbf{E}}$ with i.i.d. ZMCSCG random variables with unit variance.

A. Mismatched Receiver

The received signal corresponding to the transmission of code word $\mathbf{X}$ can be written as

$$\mathbf{Y}=\mathbf{H}\mathbf{X}+\sqrt{N_0}\widetilde{\mathbf{Z}}.$$

The estimated CSI is $\widehat{\mathbf{H}}=\mathbf{H}+(N_0/(PE_\mathrm{p}))^{1/2}\widetilde{\mathbf{E}}$ for the ML estimator and $\widehat{\mathbf{H}}=\rho(\mathbf{H}+(N_0/(PE_\mathrm{p}))^{1/2}\widetilde{\mathbf{E}})$ for the MMSE estimator. However, since $\rho=1+O(N_0)$ as $N_0\to 0$, we can

$$\begin{aligned}p(\mathbf{Y},\mathbf{Y}_\mathrm{p}|\mathbf{X},\mathbf{X}_\mathrm{p}) &= \mathbb{E}_\mathbf{H}\left[\frac{\exp\left(-\left(\|\mathbf{Y}-\mathbf{H}\mathbf{X}\|^2+\|\mathbf{Y}_\mathrm{p}-\mathbf{H}\mathbf{X}_\mathrm{p}\|^2\right)/N_0\right)}{(\pi N_0)^{(P+N)r}}\right]\\ &= \mathbb{E}_\mathbf{H}\left[\frac{\mathrm{etr}\left(-\left(\mathbf{H}^\dagger\mathbf{H}\left(\mathbf{X}\mathbf{X}^\dagger+\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger\right)-\mathbf{H}\left(\mathbf{X}\mathbf{Y}^\dagger+\mathbf{X}_\mathrm{p}\mathbf{Y}_\mathrm{p}^\dagger\right)-\left(\mathbf{Y}\mathbf{X}^\dagger+\mathbf{Y}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger\right)\mathbf{H}^\dagger+\left(\mathbf{Y}\mathbf{Y}^\dagger+\mathbf{Y}_\mathrm{p}\mathbf{Y}_\mathrm{p}^\dagger\right)\right)/N_0\right)}{(\pi N_0)^{(P+N)r}}\right]\\ &= \frac{1}{(\pi N_0)^{(P+N)r}}\det\left[\mathbf{I}_t+\frac{\mathbf{X}\mathbf{X}^\dagger+\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger}{N_0}\right]^{-r}\\ &\quad\cdot\mathrm{etr}\left\{\left(\mathbf{Y}\mathbf{X}^\dagger+\mathbf{Y}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger\right)\left[\mathbf{I}_t+\frac{\mathbf{X}\mathbf{X}^\dagger+\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger}{N_0}\right]^{-1}\frac{\mathbf{X}\mathbf{Y}^\dagger+\mathbf{X}_\mathrm{p}\mathbf{Y}_\mathrm{p}^\dagger}{N_0^2}-\frac{\mathbf{Y}\mathbf{Y}^\dagger+\mathbf{Y}_\mathrm{p}\mathbf{Y}_\mathrm{p}^\dagger}{N_0}\right\}\end{aligned} \tag{25}$$

write in both cases

$$\widehat{\mathbf{H}} = \mathbf{H} + \sqrt{\frac{N_0}{PE_{\rm p}}}\widetilde{\mathbf{E}} + O(N_0).$$

Then, the mismatched metric is given by

$$\begin{aligned}\widetilde{\mu}(\widehat{\mathbf{X}}) &= \|\mathbf{Y} - \widehat{\mathbf{H}}\widehat{\mathbf{X}}\|^2\\ &= \left\|\mathbf{H}\mathbf{X} + \sqrt{N_0}\widetilde{\mathbf{Z}} - \left(\mathbf{H} + \sqrt{\frac{N_0}{PE_{\rm p}}}\widetilde{\mathbf{E}}\right)\widehat{\mathbf{X}} + O(N_0)\right\|^2\\ &= \left\|\mathbf{H}\boldsymbol{\Delta} + \sqrt{N_0}\widetilde{\mathbf{Z}} - \sqrt{\frac{N_0}{PE_{\rm p}}}\widetilde{\mathbf{E}}\widehat{\mathbf{X}} + O(N_0)\right\|^2\\ &= \|\mathbf{H}\boldsymbol{\Delta}\|^2 + 2\sqrt{N_0}\mathrm{Re}\left[\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}\widetilde{\mathbf{Z}}^\dagger)\right]\\ &\quad - 2\sqrt{\frac{N_0}{PE_{\rm p}}}\mathrm{Re}\left[\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}\widehat{\mathbf{X}}^\dagger\widetilde{\mathbf{E}}^\dagger)\right] + O(N_0).\end{aligned} \tag{33}$$

After some algebra, we get the following asymptotic result:

$$\mathbb{P}(\mathbf{X}\to\widehat{\mathbf{X}}) = \mathbb{E}\left[Q\left(\sqrt{\frac{\|\mathbf{H}\boldsymbol{\Delta}\|^2}{2N_0}\left(1 + \frac{\|\mathbf{H}\boldsymbol{\Delta}\widehat{\mathbf{X}}^\dagger\|^2}{PE_{\rm p}\|\mathbf{H}\boldsymbol{\Delta}\|^2}\right)^{-1} + O(1)}\right)\right]. \tag{34}$$

From the Cauchy–Schwarz inequality applied to Frobenius norms [13, p. 291], we have the inequalities

$$1 \le 1 + \frac{\|\mathbf{H}\boldsymbol{\Delta}\widehat{\mathbf{X}}^\dagger\|^2}{PE_{\rm p}\|\mathbf{H}\boldsymbol{\Delta}\|^2} \le 1 + \frac{\|\widehat{\mathbf{X}}\|^2}{PE_{\rm p}}$$

and, hence, we obtain the following upper and lower bounds to the PEP:

$$\begin{aligned}\mathbb{E}\left[Q\left(\sqrt{\frac{\|\mathbf{H}\boldsymbol{\Delta}\|^2}{2N_0} + O(1)}\right)\right] &\le \mathbb{P}(\mathbf{X}\to\widehat{\mathbf{X}})\\ &\le \mathbb{E}\left[Q\left(\sqrt{\frac{\|\mathbf{H}\boldsymbol{\Delta}\|^2}{2N_0}\left(1+\frac{\|\widehat{\mathbf{X}}\|^2}{PE_{\rm p}}\right)^{-1} + O(1)}\right)\right].\end{aligned}$$

Then, using (32), we see that the diversity gain achieved by the mismatched receiver is the same as for the ML receiver with perfect CSI, consistently with a claim in [22].

B. ML/Optimum Receiver

The same result holds for the ML and optimum receivers (which are equivalent in the case considered of orthogonal pilot matrices, as shown in Appendix IV, and do not depend on the type of CSI estimation as shown in Section IV). In this case, we can write the ML metric as in (35) (shown at the bottom of the page).

After some algebra, we obtain

$$\mathbb{P}(\mathbf{X}\to\widehat{\mathbf{X}}) = \mathbb{E}\left[Q\left(\sqrt{\frac{\mathrm{tr}\{\mathbf{H}\boldsymbol{\Delta}[\mathbf{I}_N + \widehat{\mathbf{X}}^\dagger\widehat{\mathbf{X}}/(PE_{\rm p})]^{-1}\boldsymbol{\Delta}^\dagger\mathbf{H}^\dagger\}}{2N_0} + O(1)}\right)\right].$$

From positive semidefinite ordering theory (see [13, Section 7.7]), we can easily see that

$$\mathbf{I}_N \succeq \left(\mathbf{I}_N + \frac{\widehat{\mathbf{X}}^\dagger\widehat{\mathbf{X}}}{PE_{\rm p}}\right)^{-1} \succeq \left(1 + \frac{\|\widehat{\mathbf{X}}\|^2}{PE_{\rm p}}\right)^{-1}\mathbf{I}_N$$

and, hence

$$\begin{aligned}\mathbb{E}\left[Q\left(\sqrt{\frac{\|\mathbf{H}\boldsymbol{\Delta}\|^2}{2N_0} + O(1)}\right)\right] &\le \mathbb{P}(\mathbf{X}\to\widehat{\mathbf{X}})\\ &\le \mathbb{E}\left[Q\left(\sqrt{\frac{\|\mathbf{H}\boldsymbol{\Delta}\|^2}{2N_0}\left(1+\frac{\|\widehat{\mathbf{X}}\|^2}{PE_{\rm p}}\right)^{-1} + O(1)}\right)\right].\end{aligned}$$

Again, we have from (32) that the diversity gain achieved by the ML receiver with imperfect CSI estimation is the same as for the ML receiver with perfect CSI.

VII. Numerical Results

We now provide numerical results based on space–time codes obtained from a simple short block code [the (8,4,4) binary Hamming code] and certain binary convolutional codes [5]. Throughout this section, we assume that the pilot matrix

$$\begin{aligned}\mu_{\rm ML}(\widehat{\mathbf{X}}) &= \mathrm{tr}\left\{\left[\mathbf{H}\mathbf{X} + \sqrt{N_0}\widetilde{\mathbf{Z}} - \left(\mathbf{H} + \sqrt{\frac{N_0}{PE_{\rm p}}}\widetilde{\mathbf{E}}\right)\widehat{\mathbf{X}}\right]\left(\mathbf{I}_N + \frac{\widehat{\mathbf{X}}^\dagger\widehat{\mathbf{X}}}{N_0 + PE_{\rm p}}\right)^{-1}\right.\\ &\qquad \left.\cdot\left[\mathbf{H}\mathbf{X} + \sqrt{N_0}\widetilde{\mathbf{Z}} - \left(\mathbf{H} + \sqrt{\frac{N_0}{PE_{\rm p}}}\widetilde{\mathbf{E}}\right)\widehat{\mathbf{X}}\right]^\dagger\right\} + rN_0\ln\det\left(\mathbf{I}_N + \frac{\widehat{\mathbf{X}}^\dagger\widehat{\mathbf{X}}}{N_0 + PE_{\rm p}}\right)\\ &= \mathrm{tr}\left\{\mathbf{H}\boldsymbol{\Delta}\left(\mathbf{I}_N + \frac{\widehat{\mathbf{X}}^\dagger\widehat{\mathbf{X}}}{PE_{\rm p}}\right)^{-1}\boldsymbol{\Delta}^\dagger\mathbf{H}^\dagger\right\} + 2\sqrt{N_0}\mathrm{Re}\,\mathrm{tr}\left\{\mathbf{H}\boldsymbol{\Delta}\left(\mathbf{I}_N + \frac{\widehat{\mathbf{X}}^\dagger\widehat{\mathbf{X}}}{PE_{\rm p}}\right)^{-1}\left(\widetilde{\mathbf{Z}} - \frac{\widetilde{\mathbf{E}}\widehat{\mathbf{X}}}{\sqrt{PE_{\rm p}}}\right)^\dagger\right\} + O(N_0)\end{aligned} \tag{35}$$

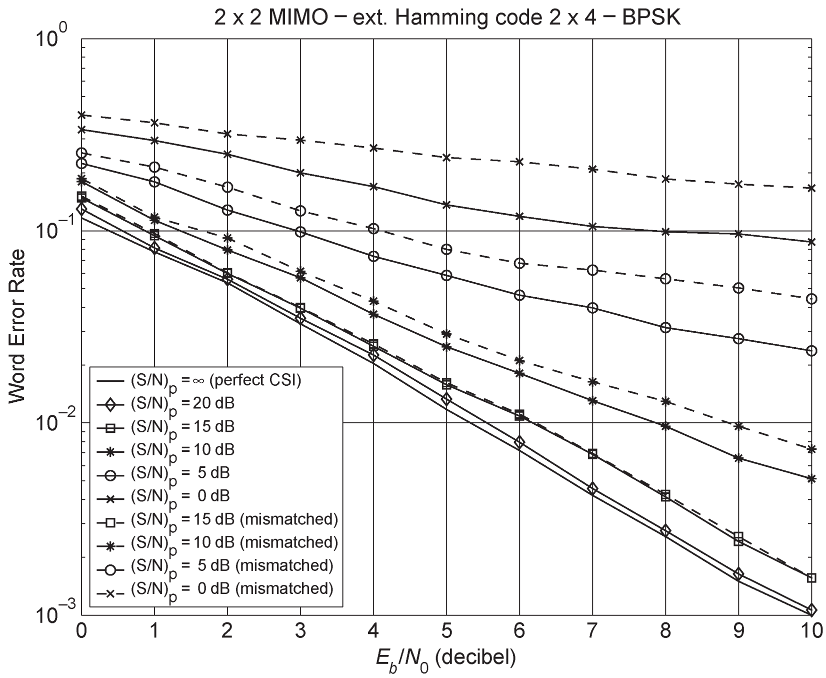


Fig. 1. Word error rate of a 2×2 ($t = 2$, $r = 2$) independent Rayleigh fading MIMO channel with a space–time code obtained by mapping the (8,4,4) binary Hamming code to 2×4 BPSK code words. Solid curves show the performance with the ML metric (20). Dashed curves show the performance with the mismatched metric (7).

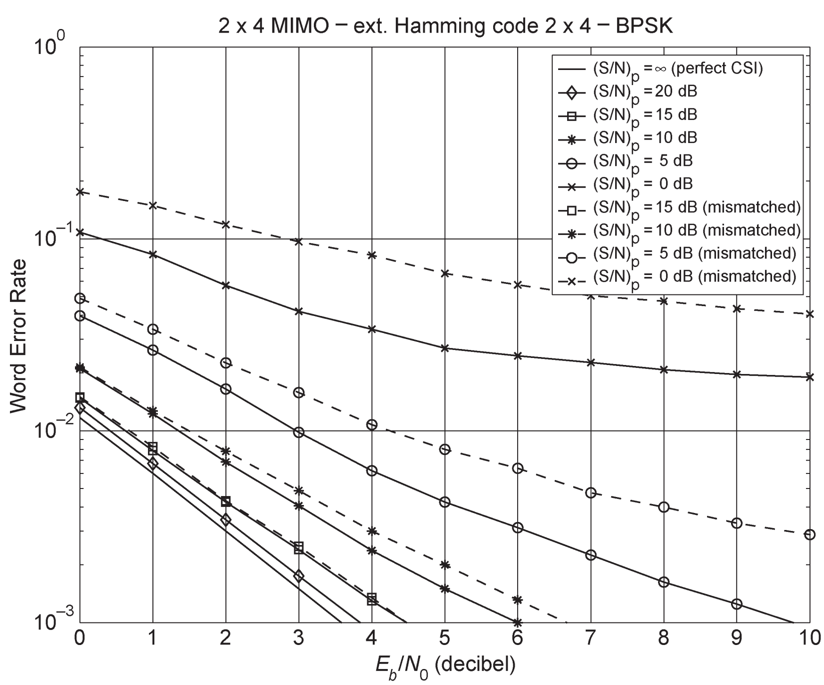


Fig. 2. Same as Fig. 1 for a 2×4 MIMO channel.

$\mathbf{X}_\mathrm{p}$ is orthogonal and that the same power is allocated to pilot and data, i.e., $E_\mathrm{p} = E_\mathrm{s}$. An upside of this choice is that it prevents fluctuations in the interference level in a multiuser setting.

A. Simple Block Space–Time Codes

We consider first a simple space–time block code obtained by mapping the (8,4,4) binary Hamming code to 2×4 binary phase-shift keying (BPSK) code words.[5] The 8-bit c_i $(i = 1, \dots, 8)$ are mapped to the BPSK symbols $x_i = (-1)^{c_i}$, which are used to produce the matrix code words filling the 2×4 matrices by rows. Figs. 1 and 2 illustrate the results obtained. The channel is assumed to be independent block Rayleigh fading. The figures show the word error rate versus E_b/N_0 for

[5]The rationale behind this choice is discussed in [2].

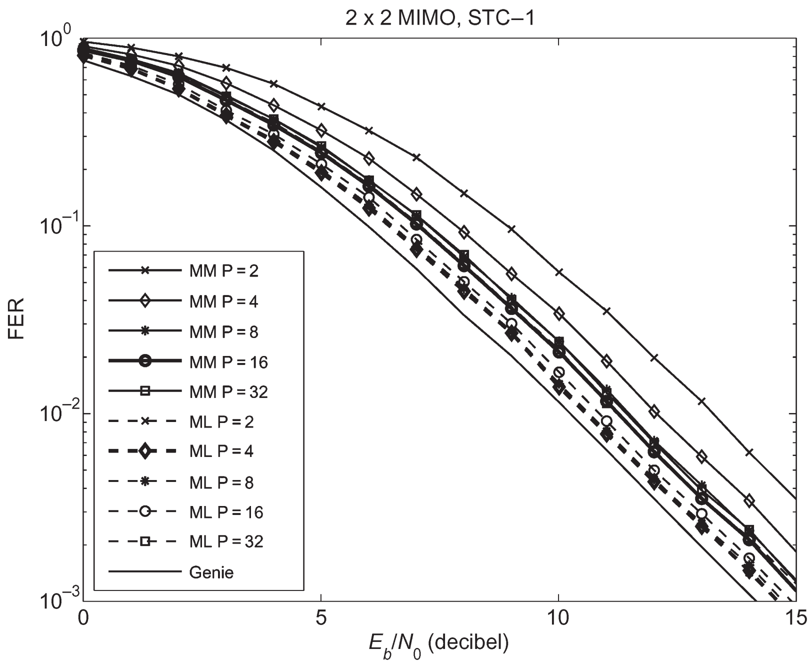


Fig. 3. Word error rate of a 2×2 ($t = 2, r = 2$) independent Rayleigh fading MIMO channel with the trellis space–time code STC-1 versus E_b/N_0 for several values of pilot intervals $P = 2$, 4, 8, and 16 and frame length $N = 130$. Solid curves with markers (labeled MM) show the performance with the mismatched metric (7). Dashed curves with markers (labeled ML) show the performance with the ML metric (20). The lowest solid curve shows the performance with perfect CSI.

different values of $(\mathrm{S/N})_\mathrm{p} \triangleq PE_\mathrm{p}/N_0$ with the mismatched metric (7) and the ML metric (20).

It is worth saying that E_b denotes the average received energy per information bit accounting for the rate loss due to pilot symbols, i.e.,

$$E_b = \frac{N+P}{N}\,\frac{tE_\mathrm{s}}{\mu_b}$$

where μ_b denotes the number of information bits per symbol interval conveyed by the space–time code. This definition reflects the fact that an increase of the number of pilot symbols P increases the quality of the channel estimate, but lowers the transmission rate.

It can be noticed that: 1) the performance loss with respect to ideal coherent detection is negligible with $(\mathrm{S/N})_\mathrm{p} \geq 20$ dB (recall that $(\mathrm{S/N})_\mathrm{p} \triangleq PE_\mathrm{p}/N_0$); and 2) the performance loss of the mismatched receiver with respect to the ML receiver is significant when $(\mathrm{S/N})_\mathrm{p} \leq 10$ dB. These results depend on the code used, so it makes sense to investigate the influence of the code selection on the accuracy of the channel estimate required for a marginal performance loss.

B. Trellis Space–Time Codes

The simulation examples below show the performance degradation of space–time codes due to imperfect channel estimation with the mismatched metric (7) and with the ML metric (20). We show in Appendix IV that the optimum receiver is equivalent to the ML receiver in this case. Simulations have been performed by implementing the sequential algorithm derived in Section IV-B leading to the branch metrics (23). Performance results relevant to an ML receiver with perfect CSI (provided by a genie without rate loss) are also reported for comparison.

Two trellis space–time codes are considered. They are obtained by mapping the rate-2/4 binary convolutional codes with 4 and 16 states, whose generator matrices are [5]

$$\mathbf{G}_1 = \begin{pmatrix} 0 & 3 & 1 & 2 \\ 3 & 1 & 2 & 1 \end{pmatrix}$$

and

$$\mathbf{G}_2 = \begin{pmatrix} 3 & 7 & 1 & 6 \\ 4 & 7 & 6 & 3 \end{pmatrix}$$

to QPSK so that the four encoded bits are mapped to QPSK symbol pairs sent to the $t = 2$ transmit antennas. Hereafter, we refer to these codes by the names STC-1 and STC-2.

Fig. 3 shows the frame error rate (FER) versus E_b/N_0 of a 2×2 MIMO block Rayleigh fading channel with the trellis space–time code STC-1 with the mismatched and ML metrics (7) and (20) labeled MM and ML, respectively. The frame length is $N = 130$ including trellis termination. Moreover, it is assumed that the average pilot symbol energy is equal to the average data symbol energy (i.e., $E_\mathrm{p} = E_\mathrm{s}$) and P (number of pilot intervals per frame) takes on the values 2, 4, 8, 16, and 32.

Similarly, Fig. 4 shows the performance of a 2×4 MIMO channel with the trellis space–time code STC-1. Fig. 5 shows the performance of a 2×2 MIMO channel with the trellis space–time code STC-2.

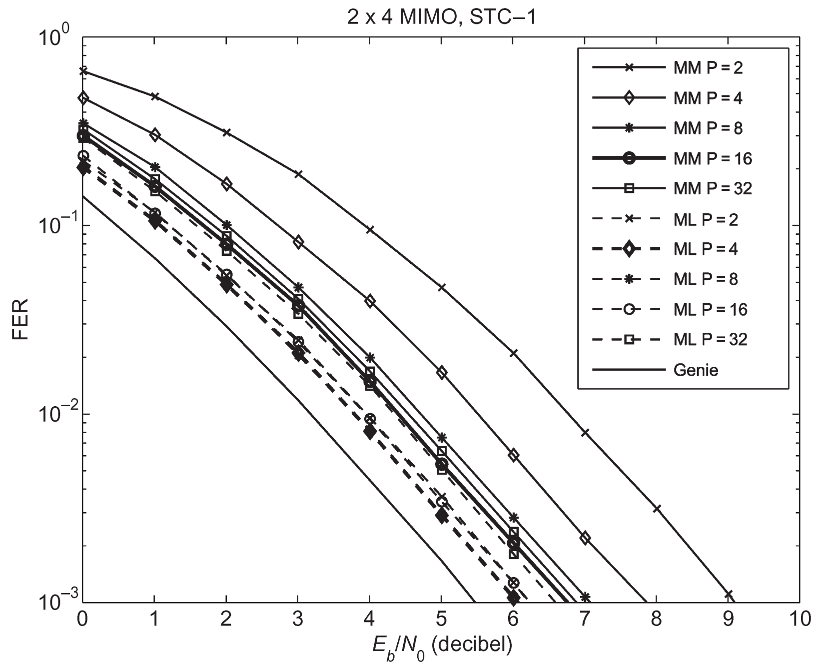


Fig. 4. Same as Fig. 3 for a 2 × 4 MIMO channel.

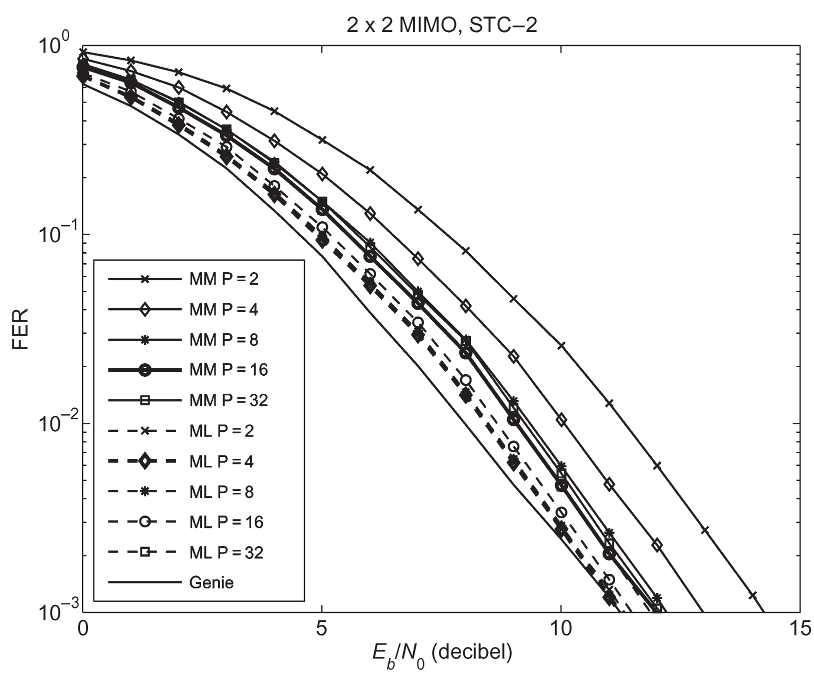


Fig. 5. Same as Fig. 3 for the code STC-2.

It can be noticed that the improvement achieved by increasing the number of pilot symbols used with the ML metric is always limited, and there is a gap to the lower bound performance of the genie ML receiver, which has perfect CSI available at no expense. The gap depends on t and r—the code and the receiver considered. When $t = r = 2$, it is about 0.3 dB (STC-1 at $\mathrm{FER} = 10^{-2}$ with $P = 4$, ML receiver) or 1.0 dB (STC-1 at $\mathrm{FER} = 10^{-2}$ with $P = 16$, mismatched receiver). When $t = 2$ and $r = 4$, it is about 0.6 dB (STC-1 at $\mathrm{FER} = 10^{-2}$ with $P = 4$, ML receiver) or 1.3 dB (STC-1 at $\mathrm{FER} = 10^{-2}$ with $P = 16$, mismatched receiver). Finally, when $t = r = 2$, it is about 0.45 dB (STC-2 at $\mathrm{FER} = 10^{-2}$ with $P = 4$, ML receiver) or 1.1 dB (STC-2 at $\mathrm{FER} = 10^{-2}$ with $P = 16$, mismatched receiver).

Notice that the mismatched receiver attains its optimum performance when the number of pilot symbols per frame P is equal to 16 (about 11% of the overall frame of pilot and data symbols), while the ML receiver performance attains its

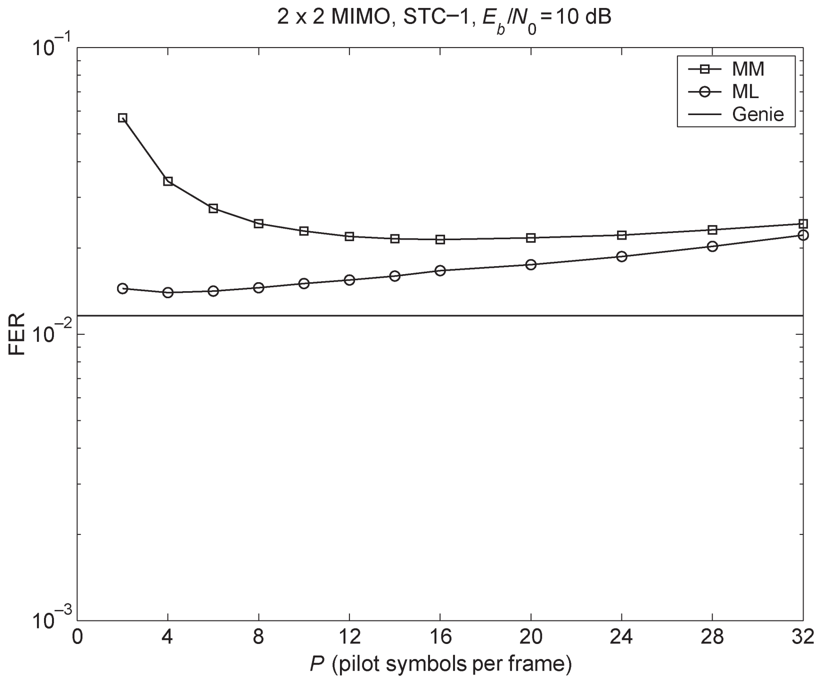


Fig. 6. Word error rate of a 2×2 ($t = 2, r = 2$) independent Rayleigh fading MIMO channel with the trellis space–time code STC-1 versus the number of pilot symbols at $E_b/N_0 = 10$ dB. Solid curves with $\square$ (labeled MM) show the performance with the mismatched metric (7). Solid curves with $\circ$ (labeled ML) show the performance with the ML metric (20). The lowest straight line shows the performance with perfect CSI.

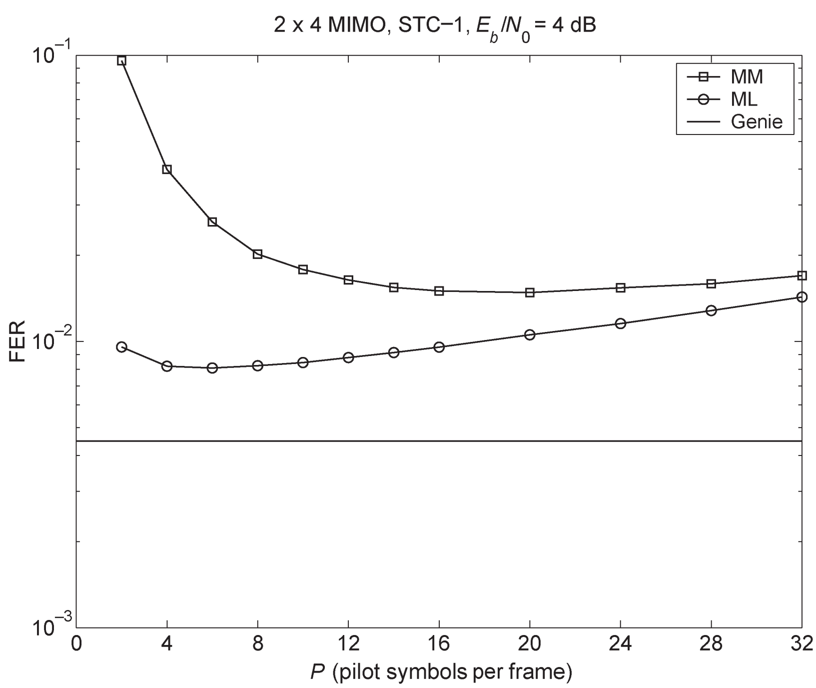


Fig. 7. Same as Fig. 6 for a 2×4 MIMO channel at $E_b/N_0 = 4$ dB.

optimum performance when the number of pilot symbols per frame P is equal to 4 (about 3% of the overall frame of pilot and data symbols).

Figs. 6–8 show the FER performance versus the number of pilot symbols P at fixed E_b/N_0. They refer to 2×2 and 2×4 MIMO systems with trellis space–time codes STC-1 and STC-2. The ML receiver performs very close to the genie receiver and the optimum number of pilot symbols is about 4 for the ML receiver and 16 for the mismatched receiver.

To summarize, we can say that the complexity increase entailed by the ML receiver buys in the cases considered about 0.7 dB of E_b/N_0.

C. Comparison With [11]

It was shown in [11], under our assumption of equal signal and pilot energies, i.e., $E_{\rm s} = E_{\rm p}$, that the achievable rate of a MIMO system based on orthogonal pilot matrices is lower

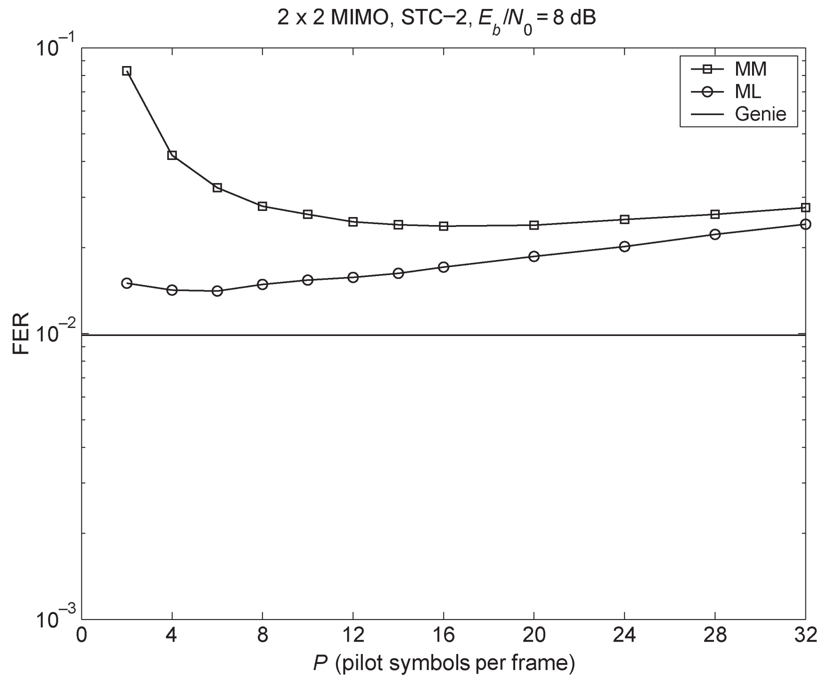


Fig. 8. Same as Fig. 6 for the code STC-2 at $E_b/N_0 = 8$ dB.

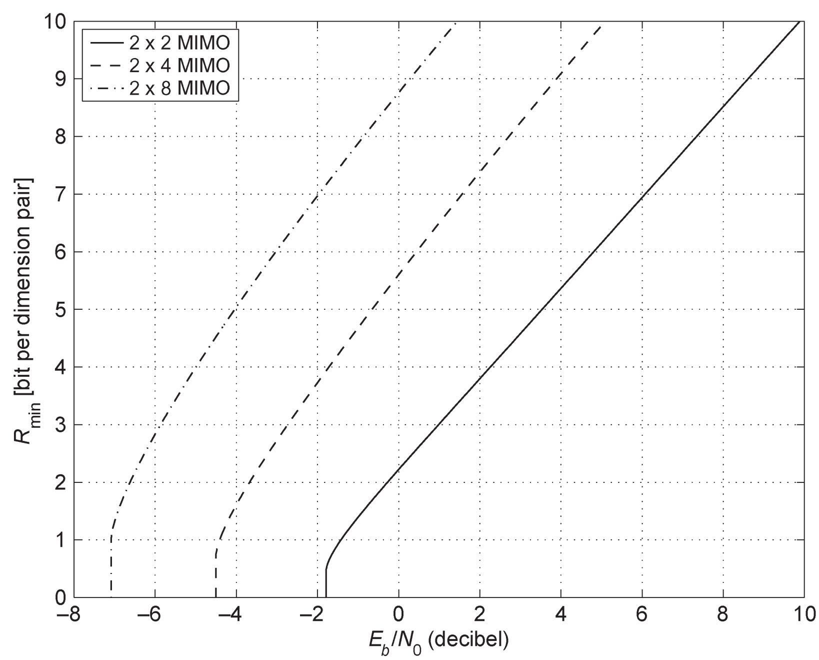


Fig. 9. Plot of the minimum achievable rate for 2×2, 2×4, and 2×8 independent Rayleigh MIMO channels versus E_b/N_0.

bounded by $R_{\min}$, a quantity obtained by solving the implicit equation

$$R_{\min} = \frac{N}{N+P}\mathbb{E}\log\det\left(\mathbf{I}_r + \frac{\rho/t}{1+(t/P)(1+1/\rho)}\mathbf{H}\mathbf{H}^\dagger\right) \tag{36}$$

bit/dimension pair, where

$$\rho \triangleq \frac{E_b}{N_0}R_{\min}. \tag{37}$$

Fig. 9 plots the maximum (over the number of pilot symbols per frame, P) $R_{\min}$ versus E_b/N_0 assuming a frame length $N = 130$ for a 2×2, 2×4, and 2×8 independent Rayleigh MIMO channel. Fig. 10 plots the corresponding optimum number of pilot symbols per frame P versus E_b/N_0. As expected, when E_b/N_0 is low, a large number of pilot symbols per frame is required, reflecting the fact that the channel is more difficult to estimate. On the contrary, when E_b/N_0 is large, fewer pilot symbols are needed. The optimum P depends also on the number of receive antennas: If more antennas are available at the receiver at a fixed E_b/N_0, fewer pilot symbols are sufficient

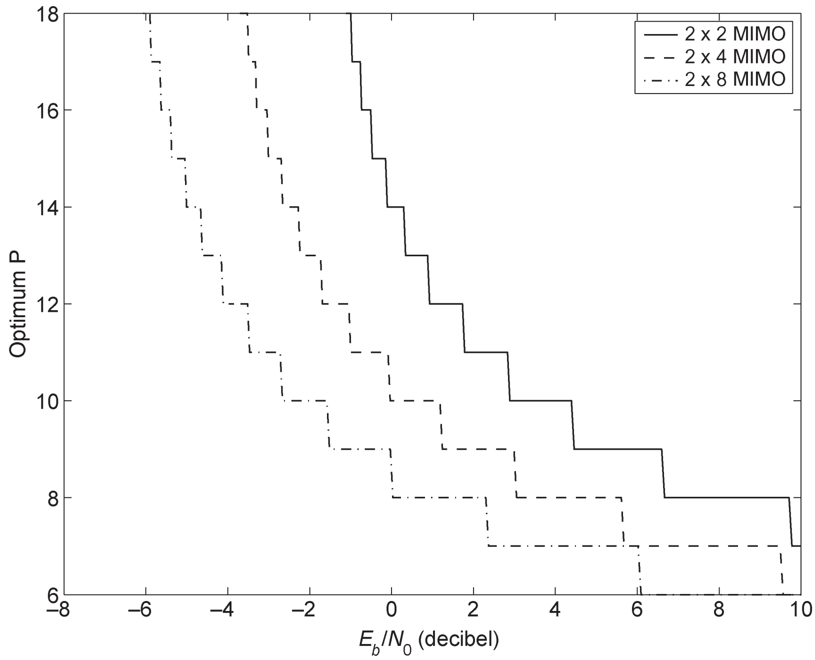


Fig. 10. Plot of the optimum number of pilot symbols corresponding to the minimum achievable rate plotted in Fig. 9 versus E_b/N_0.

to obtain the optimum performance. Nevertheless, the variation of the optimum fraction of pilot symbols required, $P/(P+N)$, is quite limited.

Moreover, using the fact that $\log\det(\mathbf{I}+\alpha\mathbf{X}) \approx \alpha\mathrm{tr}(\mathbf{X})$ as $\alpha \to 0$ and [9, Lemma A.2], the asymptotic behavior of $R_{\min}$ versus ρ is given as

$$R_{\min} \approx \begin{cases} \dfrac{NP}{t(N+P)}\rho^2 \log e, & (\rho \to 0) \\ \dfrac{N}{N+P}\left\{t \log \dfrac{P\rho}{(t+P)t} + \displaystyle\sum_{i=0}^{t-1} \Psi(r-i)\right\}, & (\rho \to \infty) \end{cases} \tag{38}$$

where $\Gamma(x) \triangleq \int_0^\infty u^{x-1}\mathrm{e}^{-u}\mathrm{d}u$ and $\Psi(x) \triangleq \Gamma'(x)/\Gamma(x)$ are, respectively, Euler's Gamma and Digamma functions. Hence, it is straightforward to see that the optimum P approaches N as $\rho \to 0$, and t (the minimum value allowed according to the inequality $P \geq t$ justified in Section III) as $\rho \to \infty$. The former case is also addressed by [18, Section 6] and by [11, Section II-C]. These results are in fair agreement with Figs. 6 and 8, which show the optimum value of P to be 4 in both cases. The difference is due to the fact that those results refer to finite space–time codes with finite signal constellations.

VIII. Conclusion

In this work, we studied the optimum receiver for a coded MIMO channel with CSI obtained by transmitting pilot symbols among the data. We derived some basic conditions on the pilot matrix $\mathbf{X}_\mathrm{p}$ and the statistics of the channel estimation error matrix. We obtained the optimum metric for a receiver that estimates the channel matrix with ML and for a receiver that does not form the channel matrix explicitly, but jointly processes all the available signals. We showed that when the pilot matrix is orthogonal, the receivers are equivalent. We provided numerical results showing that an optimum allocation of pilot space exists, and this amounts to an overhead of about 3% for the cases considered. This is in contrast with information-theoretic studies showing that for a BLAST receiver, an overhead of 100% is optimal to maximize the theoretical throughput [18]. However, we showed that this is in fair agreement with the lower bound to the information rate computed in [11].

Appendix I
Results on Circularly Symmetric Complex Gaussian Random Vectors

The following result is derived in [3].

Theorem 1: Let $\mathbf{z}_1$ and $\mathbf{z}_2$ be circularly symmetric complex Gaussian random vectors with zero means and full-rank covariance matrices $\mathbf{\Sigma}_{ij} \triangleq \mathbb{E}[\mathbf{z}_i\mathbf{z}_j^\dagger]$. Then, conditionally on $\mathbf{z}_2$, the random vector $\mathbf{z}_1$ is circularly symmetric complex Gaussian with mean $\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{z}_2$ and covariance matrix $\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$.

The following result is derived in [19].

Theorem 2: For a circularly symmetric complex Gaussian random vector $\mathbf{z} \sim \mathcal{N}_\mathrm{c}(\boldsymbol{\mu}, \mathbf{\Sigma})$ with mean $\boldsymbol{\mu} = \mathbb{E}[\mathbf{z}]$ and covariance matrix $\mathbf{\Sigma} = \mathbb{E}[\mathbf{z}\mathbf{z}^\dagger] - \boldsymbol{\mu}\boldsymbol{\mu}^\dagger$, and a Hermitian matrix $\mathbf{A}$ such that $\mathbf{I} + \mathbf{\Sigma}\mathbf{A} > \mathbf{0}$, we have

$$\mathbb{E}\big[\exp(-\mathbf{z}^\dagger\mathbf{A}\mathbf{z})\big] = \det(\mathbf{I}+\mathbf{\Sigma}\mathbf{A})^{-1}\exp\big[-\boldsymbol{\mu}^\dagger\mathbf{A}(\mathbf{I}+\mathbf{\Sigma}\mathbf{A})^{-1}\boldsymbol{\mu}\big]. \tag{39}$$

Using Theorem 2 and some algebra, we have the following.

Theorem 3: Given a Hermitian $n \times n$ matrix $\mathbf{A}$ such that $\mathbf{I}_n + \mathbf{A} > 0$, an $m \times n$ complex matrix $\mathbf{B}$, and an $m \times n$

matrix $\mathbf{Z}$ of i.i.d. ZMCSCG random variables with unit variance, the following identity holds:

$$\mathbb{E}\left[\operatorname{etr}(-\mathbf{Z}\mathbf{A}\mathbf{Z}^\dagger - \mathbf{Z}\mathbf{B}^\dagger - \mathbf{B}\mathbf{Z}^\dagger)\right] = \det(\mathbf{I}_n + \mathbf{A})^{-m} \operatorname{etr}\left[\mathbf{B}(\mathbf{I}_n + \mathbf{A})^{-1}\mathbf{B}^\dagger\right].$$

Appendix II
Details on the Derivation of (17)

To compute the pdf $p(\mathbf{Y}|\mathbf{X}, \widehat{\mathbf{H}})$, we notice that it can be written as

$$p(\mathbf{Y}|\mathbf{X}, \widehat{\mathbf{H}}) = \prod_{i=1}^{r} p(\mathbf{Y}_i|\mathbf{X}, \widehat{\mathbf{H}}_i) \tag{40}$$

since conditionally on $\mathbf{X}$, $\mathbf{Y}_i$ depends only on $\mathbf{H}_i$ and $\mathbf{Z}_i$. Then, we apply Theorem 1 of Appendix I to each factor $p(\mathbf{Y}_i|\mathbf{X}, \widehat{\mathbf{H}}_i)$. Letting

$$\mathbf{z}_1 = \mathbf{Y}_i^\dagger = \mathbf{X}^\dagger\mathbf{H}_i^\dagger + \mathbf{Z}_i^\dagger \quad \text{and} \quad \mathbf{z}_2 = \widehat{\mathbf{H}}_i^\dagger = \mathbf{B}^\dagger\left(\mathbf{H}_i^\dagger + \mathbf{E}_i^\dagger\right)$$

in Theorem 1, we have

$$\begin{aligned}
\mathbf{\Sigma}_{11} &= N_0\mathbf{I}_N + \mathbf{X}^\dagger\mathbf{X}\\
\mathbf{\Sigma}_{12} &= \mathbf{X}^\dagger\mathbf{B}\\
\mathbf{\Sigma}_{22} &= \mathbf{B}^\dagger\left[\mathbf{I}_t + N_0\left(\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger\right)^{-1}\right]\mathbf{B}.
\end{aligned}$$

Then, the conditional pdf of $\mathbf{Y}_i^\dagger$, given $\mathbf{X}$ and $\widehat{\mathbf{H}}_i$, is a circularly symmetric complex Gaussian distribution, with

$$\begin{aligned}
\text{mean} &= \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{z}_2\\
&= \mathbf{X}^\dagger\mathbf{B}\mathbf{B}^{-1}\left[\mathbf{I}_t + N_0\left(\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger\right)^{-1}\right]^{-1}(\mathbf{B}^\dagger)^{-1}\mathbf{B}^\dagger\left(\mathbf{H}_i^\dagger + \mathbf{E}_i^\dagger\right)\\
&= \mathbf{X}^\dagger\left[\mathbf{I}_t + N_0\left(\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger\right)^{-1}\right]^{-1}\left(\mathbf{H}_i^\dagger + \mathbf{E}_i^\dagger\right)
\end{aligned} \tag{41}$$

and

$$\begin{aligned}
\text{covariance matrix} &= \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^\dagger\\
&= N_0\mathbf{I}_N + \mathbf{X}^\dagger\mathbf{X} - \mathbf{X}^\dagger\mathbf{B}\mathbf{B}^{-1}\\
&\quad\times\left[\mathbf{I}_t + N_0\left(\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger\right)^{-1}\right]^{-1}(\mathbf{B}^\dagger)^{-1}\mathbf{B}^\dagger\mathbf{X}\\
&= N_0\mathbf{I}_N + \mathbf{X}^\dagger\mathbf{X} - \mathbf{X}^\dagger\\
&\quad\times\left[\mathbf{I}_t + N_0\left(\mathbf{X}_\mathrm{p}\mathbf{X}_\mathrm{p}^\dagger\right)^{-1}\right]^{-1}\mathbf{X}.
\end{aligned} \tag{42}$$

Therefore, the biasing matrix $\mathbf{B}$ does not affect the detection of the transmitted signal $\mathbf{X}$ since it does not affect $p(\mathbf{Y}|\mathbf{X}, \widehat{\mathbf{H}})$.

For simplicity, we restrict to the case of orthogonal pilot matrices. The previous expressions simplify to

$$\text{mean} = \rho\mathbf{X}^\dagger\widehat{\mathbf{H}}_i^\dagger \tag{43}$$

$$\text{covariance matrix} = N_0\mathbf{I}_N + (1-\rho)\mathbf{X}^\dagger\mathbf{X} \tag{44}$$

where we define $\rho \triangleq PE_\mathrm{p}/(N_0 + PE_\mathrm{p})$. As a result, with orthogonal pilot matrices, we have (17).

Appendix III
Derivation of the Branch Metric (23)

We want to obtain $\mu_\mathrm{ML}(\mathbf{X})$ from $\mu_\mathrm{ML}(\mathbf{X}^-)$ by a smaller number of calculations than required by using the method of direct evaluation of (20). Thus, we resort to two well-known results from linear algebra (available, e.g., from [13]) recalled here for easy reference.

1) For any $m \times n$ matrix $\mathbf{A}$ and $n \times m$ matrix $\mathbf{B}$, we have the following determinant identity:

$$\det(\mathbf{I}_m + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_n + \mathbf{B}\mathbf{A}). \tag{45}$$

2) For a given invertible $m \times m$ matrix $\mathbf{A}$, an $m \times n$ matrix $\mathbf{X}$, and an $n \times m$ matrix $\mathbf{Y}$, we have the following matrix inversion lemma:

$$(\mathbf{A} + \mathbf{X}\mathbf{Y})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}(\mathbf{I} + \mathbf{Y}\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{Y}\mathbf{A}^{-1}. \tag{46}$$

From (46), we have

$$\left(\mathbf{I}_N + \frac{(1-\rho)\mathbf{X}^\dagger\mathbf{X}}{N_0}\right)^{-1} = \mathbf{I}_N - \frac{1-\rho}{N_0}\mathbf{X}^\dagger\left(\mathbf{I}_t + \frac{(1-\rho)\mathbf{X}\mathbf{X}^\dagger}{N_0}\right)^{-1}\mathbf{X}.$$

Applying this result and identity (45), we can write the ML metric (20) as

$$\begin{aligned}
\mu_\mathrm{ML}(\mathbf{X}) = &\|\mathbf{Y} - \rho\widehat{\mathbf{H}}\mathbf{X}\|^2 - \frac{1-\rho}{N_0}\\
&\times \operatorname{tr}\left[(\mathbf{Y} - \rho\widehat{\mathbf{H}}\mathbf{X})\cdot\mathbf{X}^\dagger\left(\mathbf{I}_t + \frac{(1-\rho)\mathbf{X}\mathbf{X}^\dagger}{N_0}\right)^{-1}\right.\\
&\left.\times\mathbf{X}(\mathbf{Y} - \rho\widehat{\mathbf{H}}\mathbf{X})^\dagger\right]\\
&+ rN_0\ln\det\left(\mathbf{I}_t + \frac{(1-\rho)\mathbf{X}\mathbf{X}^\dagger}{N_0}\right).
\end{aligned} \tag{47}$$

Now, since $\mathbf{X} = (\mathbf{X}^-, \mathbf{x})$ and $\mathbf{Y} = (\mathbf{Y}^-, \mathbf{y})$, we have the following recursions:

$$\begin{aligned}
\widehat{\mu}(\mathbf{X}) &\triangleq \|\mathbf{Y} - \rho\widehat{\mathbf{H}}\mathbf{X}\|^2\\
&= \widehat{\mu}(\mathbf{X}^-) + \|\mathbf{y} - \rho\widehat{\mathbf{H}}\mathbf{x}\|^2
\end{aligned} \tag{48}$$

$$\begin{aligned}
\mathbf{\Xi}(\mathbf{X}) &\triangleq \mathbf{X}(\mathbf{Y} - \rho\widehat{\mathbf{H}}\mathbf{X})^\dagger\\
&= \mathbf{\Xi}(\mathbf{X}^-) + \mathbf{x}(\mathbf{y} - \rho\widehat{\mathbf{H}}\mathbf{x})^\dagger
\end{aligned} \tag{49}$$

$$\begin{aligned}
\mathbf{\Lambda}(\mathbf{X}) &\triangleq \left(\mathbf{I}_t + \frac{(1-\rho)\mathbf{X}\mathbf{X}^\dagger}{N_0}\right)^{-1}\\
&= \left[\mathbf{\Lambda}(\mathbf{X}^-)^{-1} + \frac{(1-\rho)\mathbf{x}\mathbf{x}^\dagger}{N_0}\right]^{-1}\\
&= \mathbf{\Lambda}(\mathbf{X}^-) - \frac{1-\rho}{N_0 + (1-\rho)\mathbf{x}^\dagger\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{x}}\\
&\quad\cdot\mathbf{\Lambda}(\mathbf{X}^-)\mathbf{x}\mathbf{x}^\dagger\mathbf{\Lambda}(\mathbf{X}^-)
\end{aligned} \tag{50}$$

and the ML metric can be written as

$$\mu_{\mathrm{ML}}(\mathbf{X}) = \widehat{\mu}(\mathbf{X}) - \frac{1-\rho}{N_0}\mathrm{tr}\left[\mathbf{\Xi}(\mathbf{X})^\dagger \mathbf{\Lambda}(\mathbf{X})\mathbf{\Xi}(\mathbf{X})\right] + rN_0 \ln\det \mathbf{\Lambda}(\mathbf{X}).$$

Finally, the corresponding metric increment is given by (23).

Appendix IV
Equivalence of the Metrics (20) and (26)

We prove the equivalence of the metrics (20) and (26) under the assumption $\mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger = PE_{\mathrm{p}}\mathbf{I}_t$. First, recall that from (18)

$$\mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger = PE_{\mathrm{p}}\mathbf{I}_t = \frac{\rho}{1-\rho}N_0\mathbf{I}_t$$

$$\mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger = PE_{\mathrm{p}}\widehat{\mathbf{H}} = \frac{\rho}{1-\rho}N_0\widehat{\mathbf{H}}.$$

Then, we obtain

$$\begin{aligned}\mu_{\mathrm{ML}}(\mathbf{X}) &= rN_0 \ln\det\left(\mathbf{I}_N + \frac{(1-\rho)\mathbf{X}^\dagger\mathbf{X}}{N_0}\right)\\ &\quad + \mathrm{tr}\left[(\mathbf{Y}-\rho\widehat{\mathbf{H}}\mathbf{X})\left(\mathbf{I}_N + \frac{(1-\rho)\mathbf{X}^\dagger\mathbf{X}}{N_0}\right)^{-1} \times(\mathbf{Y}-\rho\widehat{\mathbf{H}}\mathbf{X})^\dagger\right]\\ &= rN_0 \ln\det\left(\mathbf{I}_N + \frac{(1-\rho)\mathbf{X}^\dagger\mathbf{X}}{N_0}\right)\\ &\quad + \mathrm{tr}\left\{\left(\mathbf{Y}-\frac{1-\rho}{N_0}\mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\mathbf{X}\right)\left(\mathbf{I}_N+\frac{1-\rho}{N_0}\mathbf{X}^\dagger\mathbf{X}\right)^{-1} \times\left(\mathbf{Y}-\frac{1-\rho}{N_0}\mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\mathbf{X}\right)^\dagger\right\}\end{aligned} \tag{51}$$

and $\mu_{\mathrm{opt}}(\mathbf{X})$ is defined in (52) at the bottom of the page. Comparing (51) and (52), we obtain

$$\mu_{\mathrm{ML}}(\mathbf{X}) = N_0\mu_{\mathrm{opt}}(\mathbf{X}) + rtN_0\ln(1-\rho) + \mathrm{tr}(\mathbf{Y}\mathbf{Y}^\dagger) + \frac{1-\rho}{N_0}\mathrm{tr}\left(\mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\mathbf{X}_{\mathrm{p}}\mathbf{Y}_{\mathrm{p}}^\dagger\right)$$

which shows that the two metrics are equivalent.

$$\begin{aligned}\mu_{\mathrm{opt}}(\mathbf{X}) &= r\ln\det\left[\mathbf{I}_t + \frac{\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger}{N_0}\right] - \mathrm{tr}\left\{\left(\mathbf{Y}\mathbf{X}^\dagger + \mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\right)\left[\mathbf{I}_t + \frac{\mathbf{X}\mathbf{X}^\dagger + \mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger}{N_0}\right]^{-1}\frac{\mathbf{X}\mathbf{Y}^\dagger + \mathbf{X}_{\mathrm{p}}\mathbf{Y}_{\mathrm{p}}^\dagger}{N_0^2}\right\}\\ &= r\ln\det\left(\frac{1}{1-\rho}\mathbf{I}_t + \frac{\mathbf{X}\mathbf{X}^\dagger}{N_0}\right) - \frac{1}{N_0^2}\mathrm{tr}\left\{\left(\mathbf{Y}\mathbf{X}^\dagger + \mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\right)\left(\frac{1}{1-\rho}\mathbf{I}_t + \frac{\mathbf{X}\mathbf{X}^\dagger}{N_0}\right)^{-1}\left(\mathbf{X}\mathbf{Y}^\dagger + \mathbf{X}_{\mathrm{p}}\mathbf{Y}_{\mathrm{p}}^\dagger\right)\right\}\\ &= -rt\ln(1-\rho) + r\ln\det\left[\mathbf{I}_N + \frac{(1-\rho)\mathbf{X}^\dagger\mathbf{X}}{N_0}\right]\\ &\quad - \frac{1-\rho}{N_0^2}\mathrm{tr}\left\{\left(\mathbf{Y}\mathbf{X}^\dagger + \mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\right)\left[\mathbf{I}_t + \frac{(1-\rho)\mathbf{X}\mathbf{X}^\dagger}{N_0}\right]^{-1}\left(\mathbf{X}\mathbf{Y}^\dagger + \mathbf{X}_{\mathrm{p}}\mathbf{Y}_{\mathrm{p}}^\dagger\right)\right\}\\ &= -rt\ln(1-\rho) + r\ln\det\left[\mathbf{I}_N + \frac{(1-\rho)\mathbf{X}^\dagger\mathbf{X}}{N_0}\right]\\ &\quad - \frac{1-\rho}{N_0^2}\mathrm{tr}\left\{\left(\mathbf{Y}\mathbf{X}^\dagger + \mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\right)\left[\mathbf{I}_t - \frac{1-\rho}{N_0}\mathbf{X}\left(\mathbf{I}_N + \frac{1-\rho}{N_0}\mathbf{X}^\dagger\mathbf{X}\right)^{-1}\mathbf{X}^\dagger\right]\left(\mathbf{X}\mathbf{Y}^\dagger + \mathbf{X}_{\mathrm{p}}\mathbf{Y}_{\mathrm{p}}^\dagger\right)\right\}\\ &= -rt\ln(1-\rho) + r\ln\det\left[\mathbf{I}_N + \frac{(1-\rho)\mathbf{X}^\dagger\mathbf{X}}{N_0}\right]\\ &\quad - \frac{1-\rho}{N_0^2}\mathrm{tr}\left\{\mathbf{Y}\mathbf{X}^\dagger\mathbf{X}\left(\mathbf{I}_N + \frac{1-\rho}{N_0}\mathbf{X}^\dagger\mathbf{X}\right)^{-1}\mathbf{Y}^\dagger + \mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\mathbf{X}\left(\mathbf{I}_N + \frac{1-\rho}{N_0}\mathbf{X}^\dagger\mathbf{X}\right)^{-1}\mathbf{Y}^\dagger\right.\\ &\qquad + \mathbf{Y}\left(\mathbf{I}_N + \frac{1-\rho}{N_0}\mathbf{X}^\dagger\mathbf{X}\right)^{-1}\mathbf{X}^\dagger\mathbf{X}_{\mathrm{p}}\mathbf{Y}_{\mathrm{p}}^\dagger + \mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^\dagger\\ &\qquad \left.\times\left[\mathbf{I}_t - \frac{1-\rho}{N_0}\mathbf{X}\left(\mathbf{I}_N + \frac{1-\rho}{N_0}\mathbf{X}^\dagger\mathbf{X}\right)^{-1}\mathbf{X}^\dagger\right]\mathbf{X}_{\mathrm{p}}\mathbf{Y}_{\mathrm{p}}^\dagger\right\}\end{aligned} \tag{52}$$

ACKNOWLEDGMENT

The authors would like to thank E. G. Larsson for useful discussions and the anonymous reviewers for their helpful comments.

REFERENCES

[1] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*. New York: Kluwer/Plenum, 1999.

[2] E. Biglieri, G. Taricco, and A. Tulino, "Performance of space–time codes for a large number of antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1794–1803, Jul. 2002.

[3] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*. New York: Springer, 1999.

[4] J. K. Cavers and P. Ho, "Analysis of the error performance of trellis-coded modulations in Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 40, no. 1, pp. 74–83, Jan. 1992.

[5] J.-J. Chang, D.-J. Hwang, and M.-C. Lin, "Some extended results on the search for good convolutional codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1682–1697, Sep. 1997.

[6] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, "Great expectations: The value of spatial diversity in wireless networks," *Proc. IEEE*, vol. 92, no. 2, pp. 219–270, Feb. 2004.

[7] G. J. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, Aut. 1996.

[8] P. Garg, R. K. Mallik, and H. A. Gupta, "Performance analysis of space–time coding with imperfect channel estimation," in *Proc. Int. Conf. Personal Wireless Communications (ICPWC) 2002*, New Delhi, India, pp. 71–75.

[9] A. Grant, "Rayleigh fading multiple antenna channels," *EURASIP J. Appl. Signal Process., Special Issue on Space–Time Coding and its Applications—Part I*, vol. 2002, no. 3, pp. 316–329, Mar. 2002.

[10] B. Hassibi and B. M. Hochwald, "Optimal training in space–time systems," in *Proc. 34th Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 2000, vol. 1, pp. 743–747.

[11] ——, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[12] B. M. Hochwald and T. L. Marzetta, "Unitary space–time modulation for multiple-antenna communications in Rayleigh fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[13] R. Horn and C. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.

[14] A. Jagannatham and B. D. Rao, "A semi-blind technique for MIMO channel matrix estimation," in *Signal Processing Advances Wireless Communications (SPAWC) 2003*, Rome, Italy, pp. 304–308.

[15] E. G. Larsson and P. Stoica, *Space–Time Block Coding for Wireless Communications*. New York: Cambridge Univ. Press, 2003.

[16] E. G. Larsson, "Diversity and channel estimation errors," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 205–208, Feb. 2004.

[17] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 67–75, Jan. 2002.

[18] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space–time wireless," in *Proc. 37th Annual Allerton Conf. Communication, Control, and Computing*, Monticello, IL, 1999, pp. 958–966.

[19] M. Schwartz, W. R. Bennett, and S. Stein, *Communications Systems and Techniques*. New York: McGraw-Hill, 1966.

[20] M. F. Siyau, P. Nobles, and R. F. Ormondroyd, "Channel estimation for layered space–time systems," in *Signal Processing Advances Wireless Communications (SPAWC) 2003*, Rome, Italy, pp. 482–486.

[21] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

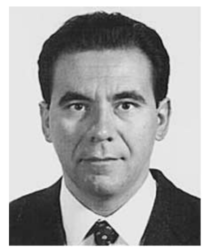
[22] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 199–207, Feb. 1999.

[23] ——, "Errata to 'Space–time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths'," *IEEE Trans. Commun.*, vol. 51, no. 12, p. 2141, Dec. 2003.

[24] E. Yoon, M. Vu, and A. Paulraj, "Exploiting channel knowledge at the Tx in MISO and MIMO wireless," in *1st Int. Workshop Signal Processing Wireless Communications*, London, U.K., 2003.

[25] X. Zhang and B. Ottersten, "Performance analysis of V-BLAST structure with channel estimation errors," in *Signal Processing Advances Wireless Communications (SPAWC) 2003*, Rome, Italy, pp. 487–491.

[26] L. Zheng and D. N. C. Tse, "Communication on the Grassman manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.



Giorgio Taricco (M'91–SM'03) was born in Torino, Italy. He received the Ingegnere Elettronico degree (*cum laude*) from Politecnico di Torino, Torino, Italy, in 1985.

In 1985, he joined the Telecom Italian Laboratories (CSELT) where he was involved in the design of the channel coding subsystem of GSM. Since 1991, he has been with the Dipartimento di Elettronica of Politecnico di Torino, now as an Associate Professor. In 1996, he was a Research Fellow at ESTEC. Among his research interests are the following: error-control coding, multiuser detection, space–time coding, and MIMO communications. He is the author or coauthor of about 50 journal papers and 100 conference contributions, and holds two international patents in applied error-control coding.

Mr. Taricco took part in the committees of several IEEE conferences and he is currently an Associate Editor of the Journal on Communications and Networks and of the IEEE COMMUNICATIONS LETTERS.

Ezio Biglieri (M'73–SM'82–F'89) was born in Aosta, Italy, in 1944. He received the Dr. Engr. degree in electrical engineering from Politecnico di Torino, Torino, Italy, in 1967.

From 1968 to 1975, he was with the Institute of Electronics and Telecommunications, Politecnico di Torino, first as a Research Engineer, then as an Associate Professor (jointly with the Institute of Mathematics). In 1975, he was made a Professor of Electrical Engineering at the University of Napoli, Naples, Italy. In 1977, he returned to Politecnico di Torino as a Professor in the Department of Electrical Engineering. From 1987 to 1989, he was a Professor of Electrical Engineering at the University of California, Los Angeles. He has been, again, a Professor at Politecnico di Torino since 1990.

He has held visiting positions with the Department of System Science, UCLA, the Mathematical Research Center, Bell Laboratories, Murray Hill, NJ, the Bell Laboratories, Holmdel, NJ, the Department of Electrical Engineering, UCLA, the Telecommunication Department of the Ecole Nationale Supérieure des Télécommunications, Paris, France, the University of Sydney, Australia, the Yokohama National University, Japan, the Electrical Engineering Department of Princeton University, the University of South Australia, Adelaide, the University of Melbourne, and the Institute for Communications Engineering, Munich Institute of Technology, Germany.

Dr. Biglieri was elected three times to the Board of Governors of the IEEE Information Theory Society, and he served as its President in 1999. He was an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE COMMUNICATIONS LETTERS, the Journal on Communications and Networks, and the Editor-in-Chief of the European Transactions on Telecommunications. Since 2004, he has been the Editor-in-Chief of the IEEE COMMUNICATIONS LETTERS. He has edited three books and coauthored five. Among other honors, in 2000, he received the IEEE Third-Millennium Medal and the IEEE Donald G. Fink Prize Paper Award, and in 2001, the IEEE Communications Society Edwin Howard Armstrong Achievement Award.