Combining Beamforming and Orthogonal Space–Time Block Coding

George Jöngren, *Student Member, IEEE*, Mikael Skoglund, *Member, IEEE*, and Björn Ottersten, *Senior Member, IEEE*

***Abstract*—Multiple transmit and receive antennas can be used in wireless systems to achieve high data rate communication. Recently, efficient space–time codes have been developed that utilize a large portion of the available capacity. These codes are designed under the assumption that the transmitter has no knowledge about the channel. In this work, on the other hand, we consider the case when the transmitter has partial, but not perfect, knowledge about the channel and how to improve a predetermined code so that this fact is taken into account. A performance criterion is derived for a frequency-nonselective fading channel and then utilized to optimize a linear transformation of the predetermined code. The resulting optimization problem turns out to be convex and can thus be efficiently solved using standard methods. In addition, a particularly efficient solution method is developed for the special case of independently fading channel coefficients. The proposed transmission scheme combines the benefits of conventional beamforming with those given by orthogonal space–time block coding. Simulation results for a narrow-band system with multiple transmit antennas and one or more receive antennas demonstrate significant gains over conventional methods in a scenario with nonperfect channel knowledge.**

***Index Terms*—Array processing, beamforming, diversity, fading channels, space–time codes, wireless communication.**

I. Introduction

THE use of transmit diversity in wireless communication systems has recently received considerable attention [1]–[4]. By utilizing antenna arrays at both the transmitter as well as the receiver the limitations of the radio channel may be overcome and the data rates increased. The high data rates that these multi-input–multi-output (MIMO) systems may offer were demonstrated in [5], [6]. There, calculations of the information-theoretic capacity assuming a flat Rayleigh-fading environment were presented. This triggered the development of space–time codes that utilize both the spatial and temporal dimension to achieve a significant portion of the aforementioned capacity [2]–[4].

The present work considers how side information in the form of channel estimates at the transmitter can be used in conjunction with certain space–time codes. A transmission scheme is developed which adapts a predetermined space–time code to the available channel knowledge by means of a linear transformation. Accurate estimates are in practice not always possible, particularly in environments where the parameters of the channel are rapidly time-varying. This fact is taken into account in the proposed transmission scheme by assuming that the channel knowledge is nonperfect. In fact, the scheme continues to work well also when the quality of the side information is low.

Early attempts at designing transmission schemes for exploiting the potential offered by antenna arrays at the transmit side are generally concerned with increasing the diversity order of the system. Examples of such work include techniques for introducing artificial frequency/phase offsets [7]–[9] and time offsets [10] between the transmitted signals. The latter technique is commonly referred to as delay diversity and is closely related to the layered space–time architecture in [11]. A more systematic approach to find appropriate codes was pioneered in [2], [12]. A major contribution in [2], [12] was the development of a design criterion involving the rank and eigenvalues of certain matrices. The design criterion was later generalized to multiple receive antennas and to other channel models in [3], where the now popular notion of space–time coding was introduced. Examples of trellis codes based on the design criterion were also provided. A simple block code for two transmit antennas that leads to a low-complexity receiver was developed in [13]. In [4], this concept was extended to up to eight transmit antennas. Furthermore, it was shown that these so-called orthogonal space–time block codes satisfy the rank constraint of the previously mentioned design criterion. Consequently, these codes also provide the system with its maximum diversity order.

Common to the space–time coding schemes mentioned above is that they do not exploit channel knowledge at the transmitter. Information about the channel realization, if it is available, should of course be utilized in order to maximize the performance. As a simple example, consider a scenario with perfect channel state information at both sides of the communication link. It is well known that by appropriate linear processing at the transmitter and the receiver, the system can be transformed into a set of parallel scalar channels. The available transmit power may then be optimally allocated to the individual channels. An example of combining such an approach with coding is found in [14].

In this paper, we present a transmission scheme that combines the two extremes regarding the degree of channel knowledge. Codes belonging to the class of orthogonal space–time block codes [4] are linearly processed in order to take the side information into account. The side information is modeled using a purely statistical approach. Previous and related work includes

Manuscript received November 26, 2000; revised October 1, 2001. The material in this paper was presented in part at the Asilomar Conference on Signals, Systems and Computers, October 1999 and at the IEEE Sensor Array and Multichannel Signal Processing Workshop, March 2000.

The authors are with the Department of Signals, Sensors and Systems, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (e-mail: gjongren@s3.kth.se; skoglund@s3.kth.se; otterste@s3.kth.se).

Communicated by G. Caire, Associate Editor for Communications.
Publisher Item Identifier S 0018-9448(02)01056-8.

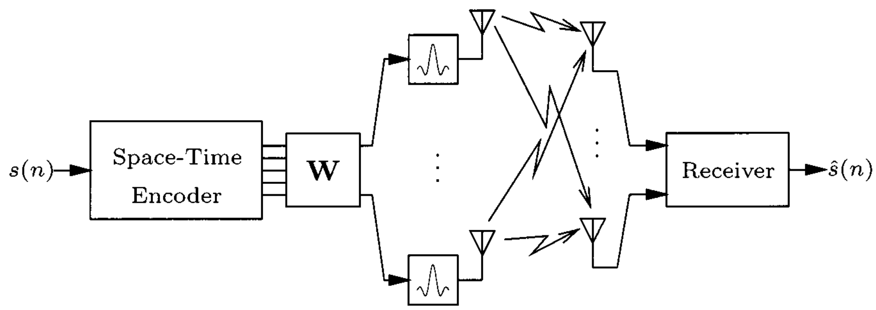

Fig. 1. An overview of the system.

[15], [16], and [17]–[19]. While a similar model for the side information was utilized in [17], [18] for determining suitable transmit beamformers in the presence of channel estimation errors, those papers neither considered space–time codes nor multiple antennas at the receiver. Another possibility for modeling partial channel knowledge is to take on a physical perspective. This is the approach used in [19] for adapting a predetermined space–time code to the particular channel. The assumption in that work is that the signals propagate along a finite number of directions, known at the transmitter, before entering a rich local scattering environment near the receiver.

The main contribution of the present work is the development and study of a low-complexity transmission scheme which combines the benefits of transmit beamforming and orthogonal space–time block coding. To accomplish this, a performance criterion is derived which takes the available side information into account. The performance criterion could, in principle, be used for designing codes from scratch. Although the derivation is to some extent similar in spirit to the derivation of the design criterion in [3], that paper did not treat side information and, therefore, utilized certain approximations which prohibit the resulting design criterion from exploiting any channel knowledge. The setup considered here may also be motivated by current standardization proposals for the WCDMA system, where an orthogonal space–time block code is used in one of the proposed transmission modes whereas one of the other proposed modes uses transmit beamforming [20], [21].

The paper is organized as follows. In Section II, the data model and the model for the side information are introduced. A derivation of the performance criterion as well as some general interpretations are given in Section III. The proposed transmission scheme is described in Section IV, leading to what first seems like a nontrivial optimization problem. However, a simple reformulation of the parameters to be determined turns out to give a convex optimization problem which permits a reasonably efficient implementation. Also, closed-form expressions for asymptotically optimum linear transformations are derived for a number of cases. These also reveal the highly intuitive behavior of the transmission scheme. Furthermore, a particularly efficient optimization algorithm is presented for a class of simplified fading scenarios. This algorithm is then applied in Section V to an actual example of a simplified fading scenario. Finally, simulation examples are presented in Section VI showing significant gains compared to conventional beamforming as well as conventional orthogonal space–time block coding.

II. SYSTEM MODEL

As illustrated in Fig. 1, a wireless communication system employing multiple antennas at both the transmitter and the receiver is considered. The transmitter is assumed to have some, but not necessarily perfect, knowledge about the current channel realization. In order to utilize this channel knowledge, while at the same time not sacrificing the benefits offered by conventional space–time codes, we propose the use of a transmitter consisting of a space–time encoder followed by a linear transformation $\boldsymbol{W}$. The space–time encoder maps the data to be transmitted $s(n)$, where n is a discrete-time index, into codewords that are split into a set of parallel and generally different symbol sequences. These codewords are linearly transformed in order to adapt the code to the available channel knowledge. As a result, a new set of parallel symbol sequences is formed. Each symbol sequence is first pulse-shaped and then transmitted over the corresponding antenna. Finally, the transmitted data is recovered at the receiver by means of maximum-likelihood (ML) decoding.

The information-carrying signals are transmitted over a wireless fading channel. The time dispersion introduced by the channel is assumed to be short compared with the symbol period. Therefore, the individual channel between each transmit and receive antenna is frequency-nonselective. Let M and N denote the number of transmit and receive antennas, respectively. The signal output from each receive antenna is then a weighted superposition of the M transmitted signals, corrupted by additive noise.

By collecting the filtered and symbol sampled complex baseband equivalent outputs from the receiving antenna array in an $N \times 1$ vector $\boldsymbol{x}(n)$, the received signal at time n can be written as

$$\boldsymbol{x}(n) = \boldsymbol{H}^* \boldsymbol{c}(n) + \boldsymbol{e}(n)$$

where $(\cdot)^*$ denotes the complex conjugate transpose operator and where the linearly transformed symbols, transmitted from the M antennas at time instant n are represented by

$$\boldsymbol{c}(n) = [c_1(n) \quad c_2(n) \quad \cdots \quad c_M(n)]^{\mathrm{T}} = \boldsymbol{W}\bar{\boldsymbol{c}}(n). \qquad (1)$$

Here, $\bar{\boldsymbol{c}}(n)$ is the corresponding output from the space–time encoder and $\boldsymbol{W}$ is the previously mentioned linear transformation matrix. As will be seen in the following sections, $\boldsymbol{W}$ is determined so as to minimize a certain upper bound on the probability of a codeword error. The noise term $\boldsymbol{e}(n)$ is assumed to be a zero-mean, temporally and spatially white, complex Gaussian random process with covariance matrix $\sigma^2 \boldsymbol{I}_N$. Furthermore, the MIMO channel is described by the $M \times N$ matrix

$$\boldsymbol{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{MN} \end{bmatrix}$$

where h_{ij}^* is a complex scalar denoting the channel between the ith transmit antenna and the jth receive antenna.[1] The MIMO channel is also represented by the $MN \times 1$ vector

$$\boldsymbol{h} = \mathrm{vec}(\boldsymbol{H})$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization operator which stacks the columns of its argument into a vector. The fading of the channel is assumed to obey a complex Gaussian distribution with mean vector $\boldsymbol{m_h}$ and covariance matrix $\boldsymbol{R_{hh}}$. Note that this assumption includes both independent Rayleigh fading and independent Ricean fading as special cases. In addition, more realistic fading environments with correlated channel coefficients may be modeled.

A quasi-static scenario is considered where the channel is assumed to be constant during the transmission of a burst of codewords but may vary from one burst to another in a statistically stationary fashion. For simplicity, the channel realizations are in this work assumed to be independent. However, it is straightforward to generalize the proposed methods to a correlated fading scenario.

A. Side Information at the Transmitter

An estimate of the channel realization is assumed to be available at the transmitter. There are several examples of how such an estimate may be obtained. An explicit channel can be used to feed back the channel estimates from the receiver or, if a time-division duplex system is used, it is possible for the transmitter to estimate the channel directly. In order to use the latter method in frequency-division duplex systems, estimates of the channel in the receive mode must first be transformed to the desired carrier frequency before they can be utilized in the transmission mode. In general, the channel estimates are not only noisy but may also be outdated due to feedback delay or duplex time. In addition, the frequency shift transformation in frequency-division duplex systems is another possible source of error.

For simplicity, the receiver is from now on assumed to have perfect channel knowledge. We make this assumption in order to simplify the exposition. However, it is straightforward to extend the following development to also take channel estimation errors at the receiver into account. Moreover, keep in mind that the development in this section is exemplified by, but not limited to, systems where the channel estimates are obtained through a dedicated feedback channel.

The channel estimates at the transmitter are assumed to be correlated (to an arbitrary degree) with the true channel. This assumption is motivated, for example, by the well-known Jakes model [22, p. 26], which describes the variations of the channel, due to movement of the mobile receiver, as a function of time. In this model, the channel coefficients are samples of a stationary Gaussian process with an autocorrelation function proportional to $J_0(2\pi f_m \tau)$, where $J_0(x)$ is the zero-order Bessel function of the first kind, τ denotes the time lag, and f_m is the maximum Doppler frequency. Hence, the outdated channel estimates available at the transmitter are correlated with the current channel and the amount of such correlation is determined by the time it takes to feed back the estimates.

For the purpose of describing the side information, let the matrix $\hat{\boldsymbol{H}}$, with the corresponding channel coefficients $\hat{h}_{ij}$, denote the estimate of $\boldsymbol{H}$ available at the transmitter. Let $\hat{\boldsymbol{h}}$ denote the vectorized counterpart. We assume that $\hat{\boldsymbol{h}}$ and $\boldsymbol{h}$ are jointly complex Gaussian. With obvious notation, the statistics of the side information and its relation to the true channel are now completely described by the mean vector $\boldsymbol{m}_{\hat{\boldsymbol{h}}}$, the covariance matrix $\boldsymbol{R}_{\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}}$, and the cross-covariance matrix $\boldsymbol{R}_{\boldsymbol{h}\hat{\boldsymbol{h}}}$. In view of the Jakes model, the joint Gaussian assumption is reasonable since the side information and the true channel are samples of the same Gaussian random process.

Clearly, the quality of the side information is closely related to the degree of correlation with the true channel, as represented by the cross-covariance matrix. A more general measure of the quality of the side information that will be used extensively in this paper is the covariance of the true channel, conditioned on the side information. Let $\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}$ denote this quantity. Since $\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}$ describes the remaining uncertainty when the side information is known, it should be appearant that, loosely speaking, high-quality side information corresponds to a small $\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}$ (measured in a suitable norm) whereas a large $\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}$ corresponds to side information of low quality. Let us now formally define the two notions of "perfect side information" (or "perfect channel knowledge") and "no side information" (or "no channel knowledge") as follows.

- "Perfect side information" $\Leftrightarrow \|\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\| \to 0$.
- "No side information" $\Leftrightarrow \|\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}^{-1}\| \to 0$.

Here, $\|\cdot\|$ denotes the spectral norm [23, p. 295]. A salient consequence of this choice of measure is that the distribution of the true channel is considered as part of the channel knowledge as well.

III. Performance Criterion

In this section, we derive a performance criterion for space–time codes which takes the quality of the side information into account. Rather than at this point limiting the application of the performance criterion to the proposed transmission scheme, the development in this section is structured in such a manner that the derived performance criterion is applicable to the design of a wide class of space–time codes.

[1]Since the focus of this work is on the transmitting side, it is convenient to define the MIMO channel using $\boldsymbol{H}^*$, as opposed to $\boldsymbol{H}$, so that each column of $\boldsymbol{H}$ represents the vector channel between the transmitter's antenna array and the corresponding receive antenna.

In the following sections, the performance criterion will be applied to the proposed transmission scheme.

As is shown next, a suitable performance criterion is based on the pairwise codeword error probability, conditioned on the side information. We start by motivating this fact. Let $\mathcal{C} = \{\boldsymbol{C}_1, \ldots, \boldsymbol{C}_K\}$ denote the set of codewords, where K is the number of codewords. Assume the codewords are of length L. Each codeword is described by an $M \times L$ matrix

$$\boldsymbol{C}_k = [\boldsymbol{c}_k(0) \quad \boldsymbol{c}_k(1) \quad \cdots \quad \boldsymbol{c}_k(L-1)], \qquad k = 1, \ldots, K$$

where $\boldsymbol{c}_k(n)$ is the nth transmitted vector of the kth codeword. Assume that a codeword $\boldsymbol{C} \in \mathcal{C}$ is transmitted. The received signal vectors corresponding to one codeword may then be arranged in an $N \times L$ matrix $\boldsymbol{X}$, given by

$$\boldsymbol{X} = \boldsymbol{H}^* \boldsymbol{C} + \boldsymbol{E}$$

where $\boldsymbol{E}$ is a matrix of noise vectors. The receiver is assumed to employ ML decoding of the codewords based on ideal channel state information. For the problem at hand this amounts to decoding the codewords according to

$$\hat{\boldsymbol{C}} = \arg\min_{\boldsymbol{C} \in \mathcal{C}} ||\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{C}||_{\mathrm{F}}^2$$

where $\hat{\boldsymbol{C}}$ denotes the codeword chosen by the receiver and $||\cdot||_{\mathrm{F}}$ is the Frobenius norm. The previously mentioned pairwise codeword error probability, conditioned on the side information,[2] can now be given a clear meaning. It is denoted by $P(\boldsymbol{C}_k \to \boldsymbol{C}_l | \hat{\boldsymbol{h}})$ and defined, for $k \neq l$, as

$$\Pr\left[||\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{C}_k||_{\mathrm{F}}^2 > ||\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{C}_l||_{\mathrm{F}}^2 \,\middle|\, \boldsymbol{C} = \boldsymbol{C}_k, \hat{\boldsymbol{h}}\right]$$

which is the probability that, given a transmitted codeword $\boldsymbol{C}_k$, the metric corresponding to the codeword $\boldsymbol{C}_l$ is smaller. Obvious variations of this notation will also be used. Let $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}]$ denote the codeword error probability, i.e., the probability that $\hat{\boldsymbol{C}}$ is different from $\boldsymbol{C}$. The overall design goal is to minimize this quantity with respect to the set of codewords. Since side information is available, the set of possible codewords $\mathcal{C}$ is a function of the channel estimate $\hat{\boldsymbol{h}}$. Conditioning on the side information gives the following relation:

$$\Pr\left[\hat{\boldsymbol{C}} \neq \boldsymbol{C}\right] = \int \Pr\left[\hat{\boldsymbol{C}} \neq \boldsymbol{C} \,\middle|\, \hat{\boldsymbol{h}}\right] p_{\hat{\boldsymbol{h}}}\left(\hat{\boldsymbol{h}}\right) d\hat{\boldsymbol{h}}$$

where $p_{\hat{\boldsymbol{h}}}(\hat{\boldsymbol{h}})$ is the probability density function (pdf) of the side information. It is now clear that minimizing $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C} | \hat{\boldsymbol{h}}]$ for each $\hat{\boldsymbol{h}}$ also minimizes $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}]$, since $p_{\hat{\boldsymbol{h}}}(\hat{\boldsymbol{h}}) > 0$. In order to obtain a closed-form expression for $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C} | \hat{\boldsymbol{h}}]$, we make the common assumption that the signal-to-noise ratio (SNR) is sufficiently high for the union bound technique to be applicable. More specifically, it is assumed that the largest pairwise codeword error probability, conditioned on the side information, is the dominating term in the union bound. Thus, $P(\boldsymbol{C}_k \to \boldsymbol{C}_l | \hat{\boldsymbol{h}})$ is a reasonable performance criterion. Note that, although based on a high-SNR assumption, we will demonstrate that the performance criterion can be successfully utilized also in situations when the SNR cannot be considered high.

We now turn our attention to deriving a closed-form expression for the performance criterion. Similarly to [3], we start by conditioning on the true channel realization and utilize a well-known upper bound on the Gaussian tail function to arrive at

$$P\left(\boldsymbol{C}_k \to \boldsymbol{C}_l \,\middle|\, \boldsymbol{h}, \hat{\boldsymbol{h}}\right) \leq \tfrac{1}{2}\, e^{-d^2(\boldsymbol{C}_k, \boldsymbol{C}_l)/4\sigma^2} \tag{2}$$

where

$$d^2(\boldsymbol{C}_k, \boldsymbol{C}_l) = ||\boldsymbol{H}^*(\boldsymbol{C}_k - \boldsymbol{C}_l)||_{\mathrm{F}}^2 \tag{3}$$

is the Euclidean distance between the codewords. At this point, the following standard relations, found in, e.g., [24, pp. 121–122], turn out to be useful:

$$\mathrm{tr}(\boldsymbol{AB}) = (\mathrm{vec}(\boldsymbol{A}^*))^* \mathrm{vec}(\boldsymbol{B}) \tag{4}$$

$$\mathrm{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A}) \mathrm{vec}(\boldsymbol{B}). \tag{5}$$

Here, $\otimes$ denotes the Kronecker product and $\mathrm{tr}(\cdot)$ is the trace operator. By utilizing (4) and (5) it is possible to rewrite (3) as

$$\begin{aligned} d^2(\boldsymbol{C}_k, \boldsymbol{C}_l) &= \mathrm{tr}(\boldsymbol{H}^* \boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l) \boldsymbol{H}) \\ &= (\mathrm{vec}(\boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l)\boldsymbol{H}))^* \mathrm{vec}(\boldsymbol{H}) \\ &= \boldsymbol{h}^* (\boldsymbol{I}_N \otimes \boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l)) \boldsymbol{h} \end{aligned} \tag{6}$$

where, similarly to [3]

$$\boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l) = (\boldsymbol{C}_k - \boldsymbol{C}_l)(\boldsymbol{C}_k - \boldsymbol{C}_l)^*$$

contains the codeword pair. Since the true channel and the side information are jointly complex Gaussian, the pdf of the true channel, conditioned on the side information, is also a complex Gaussian pdf, given by

$$p_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\left(\boldsymbol{h} \,\middle|\, \hat{\boldsymbol{h}}\right) = \frac{e^{-(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}})^* \boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}^{-1} (\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}})}}{\pi^{MN} \det\left(\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\right)} \tag{7}$$

where $\det(\cdot)$ denotes the determinant operator and where $\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}$ and $\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}$ represent the conditional mean and covariance, respectively. By averaging both sides of (2) over the distribution in (7), an upper bound to the pairwise error probability is formed as

$$P\left(\boldsymbol{C}_k \to \boldsymbol{C}_l \,\middle|\, \hat{\boldsymbol{h}}\right) \leq \int \tfrac{1}{2}\, e^{-d^2(\boldsymbol{C}_k, \boldsymbol{C}_l)/4\sigma^2} p_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\left(\boldsymbol{h} \,\middle|\, \hat{\boldsymbol{h}}\right) d\boldsymbol{h}. \tag{8}$$

This upper bound is denoted by $V(\boldsymbol{C}_k, \boldsymbol{C}_l)$. Now, introduce the following expression:

$$\boldsymbol{\Psi}(\boldsymbol{C}_k, \boldsymbol{C}_l) = (\boldsymbol{I}_N \otimes \boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l))/4\sigma^2 + \boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}^{-1}.$$

After expanding the exponent of (7) and combining it with (6), it is straightforward to verify that the sum of the exponents in the integrand of (8) can be written as

$$\begin{aligned} &\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}^* \boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}^{-1} \left(\boldsymbol{\Psi}^{-1} - \boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\right) \boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}^{-1} \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}} \\ &\qquad - \left(\boldsymbol{h} - \boldsymbol{\Psi}^{-1} \boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}^{-1} \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\right)^* \boldsymbol{\Psi} \left(\boldsymbol{h} - \boldsymbol{\Psi}^{-1} \boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}^{-1} \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\right) \end{aligned}$$

[2]For notational convenience, we will sometimes refer to this as simply the "pairwise error probability."

where the dependence on the codeword pair has been temporarily omitted in order to simplify the notation. The integral in (8) is now easily solved by making use of the fact that

$$\int \frac{e^{-(\boldsymbol{h}-\boldsymbol{\Psi}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}})^{*}\boldsymbol{\Psi}(\boldsymbol{h}-\boldsymbol{\Psi}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}})}}{\pi^{MN}\det(\boldsymbol{\Psi}^{-1})}\,d\boldsymbol{h}$$

is the integral of a complex Gaussian pdf and thus equals one. Consequently, the upper bound in (8) can be expressed as

$$V(\boldsymbol{C}_k, \boldsymbol{C}_l) = \frac{e^{\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}^{*}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}^{-1}(\boldsymbol{\Psi}(\boldsymbol{C}_k,\boldsymbol{C}_l)^{-1}-\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}})\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}}}{2\det\left(\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}\right)\det(\boldsymbol{\Psi}(\boldsymbol{C}_k, \boldsymbol{C}_l))}.$$

Taking the logarithm and neglecting parameter-independent terms yields the desired form of the performance criterion as

$$\ell(\boldsymbol{C}_k, \boldsymbol{C}_l) = \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}^{*}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{\Psi}(\boldsymbol{C}_k,\boldsymbol{C}_l)^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}} - \log\det(\boldsymbol{\Psi}(\boldsymbol{C}_k, \boldsymbol{C}_l)). \quad (9)$$

We stress that this result constitutes a new performance criterion for channel estimate dependent space–time codes. One possible approach to designing the corresponding codewords is to minimize the maximum, taken over all codeword pairs, of $\ell(\boldsymbol{C}_k, \boldsymbol{C}_l)$. For a similar solution based on the classic design criterion see, e.g., [25]. However, this procedure is deemed too computationally demanding for the side information model under consideration, since the optimum codewords depend on the actual channel estimate and the number of possible channel estimate realizations is infinite. Hence, a complicated optimization problem would have to be solved in real time for each new channel estimate that arrives at the transmitter.

On the other hand, for the case of quantized channel estimates, such an approach may be perfectly viable due to the finite number of channel estimate realizations. The entire channel estimate dependent space–time code can therefore be precalculated and stored in a lookup table, suitable for real-time use. Variations of this approach are further explored in [26]. However, such a study is beyond the scope of the present work, since we focus on unquantized channel information. Instead, certain constraints on the code will be introduced in the sections to follow in order to arrive at a tractable scheme.

A. Interpretations of the Performance Criterion

The two terms in (9), i.e., in the performance criterion, can be given some interesting interpretations. The first term mainly deals with the channel knowledge obtained from the actual realization of the channel estimate, as contained in $\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}$. The second term, on the other hand, does not depend on the realization of the channel estimate and therefore strives for a code design suitable for an open-loop system, which has no side information except prior knowledge of the distribution of the true channel. This interpretation is further supported by considering the two special cases of perfect side information ($\|\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}\| \to 0$) and of no side information ($\|\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}^{-1}\| \to 0$), respectively. In the first case, the first term is seen to dominate and in the second case, the second term is dominating. The performance criterion is in the second case equivalent to

$$\ell(\boldsymbol{C}_k, \boldsymbol{C}_l) = \frac{1}{\det \boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l)} \quad (10)$$

which is basically the same as the criterion used in [2], [3] for designing conventional space–time codes.

As a final remark, it should be emphasized that the proposed performance criterion can also be used for designing conventional space–time codes in various scenarios involving open-loop systems. To see this, bear in mind that if the side information is statistically independent of the true channel, the performance criterion reduces to

$$\ell(\boldsymbol{C}_k, \boldsymbol{C}_l) = \boldsymbol{m}_{\boldsymbol{h}}^{*}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{-1}\boldsymbol{\Psi}(\boldsymbol{C}_k, \boldsymbol{C}_l)^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{-1}\boldsymbol{m}_{\boldsymbol{h}} - \log\det(\boldsymbol{\Psi}(\boldsymbol{C}_k, \boldsymbol{C}_l)) \quad (11)$$

where now $\boldsymbol{\Psi}(\boldsymbol{C}_k, \boldsymbol{C}_l) = (\boldsymbol{I}_N \otimes \boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l))/4\sigma^2 + \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{-1}$. Thus, the development in the present work also applies to situations where the transmitter only knows the distribution of the true channel and nothing about the current realization. This version of the performance criterion is therefore closely related to the design criterion in [2], [3]. In fact, after some simple manipulations, it can be shown that (11) includes several of the results from the various fading scenarios in [2], [3] as special cases.

IV. The Transmission Scheme

This section deals with the construction of a tractable transmission scheme based on the performance criterion given by (9). As discussed in the previous section, one obvious alternative is to design a space–time code using the proposed performance criterion and an exhaustive search over all possible codewords. However, a more viable approach, for the scenario under consideration, is taken in this work.

As outlined in Section II, we assume that a space–time code is already determined and try to improve the code by a linear transformation. Hence, from (1) it follows that the kth transformed codeword may be written as

$$\boldsymbol{C}_k = \boldsymbol{W}\overline{\boldsymbol{C}}_k$$

where $\boldsymbol{W}$ is an $M \times M$ matrix, shared by all codewords, and $\overline{\boldsymbol{C}}_k \in \overline{\mathcal{C}}$ is the kth predetermined codeword. Here, the predetermined set of codewords $\overline{\mathcal{C}}$ is defined in a similar manner as $\mathcal{C}$. In order to limit the average output power, the constraint $\|\boldsymbol{W}\|_{\mathrm{F}}^2 = 1$ is imposed. Furthermore, orthogonal space–time block codes as found in [4] are considered. These codes are designed for open-loop type of systems and have the appealing property that

$$\boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l) = \mu_{kl}\boldsymbol{I}_M, \qquad \forall\, k \neq l \quad (12)$$

where μ_{kl} is a scaling factor which depends on the codeword pair. Substituting $\boldsymbol{C}_k = \boldsymbol{W}\overline{\boldsymbol{C}}_k$, $\boldsymbol{C}_l = \boldsymbol{W}\overline{\boldsymbol{C}}_l$, and (12) into (9) leads to the performance criterion

$$\ell(\boldsymbol{W}\boldsymbol{W}^{*}, \mu_{kl}) = \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}^{*}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{\Psi}(\boldsymbol{W}\boldsymbol{W}^{*}, \mu_{kl})^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\hat{\boldsymbol{h}}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}} - \log\det(\boldsymbol{\Psi}(\boldsymbol{W}\boldsymbol{W}^{*}, \mu_{kl})) \quad (13)$$

where

$$\boldsymbol{\Psi}(\boldsymbol{W}\boldsymbol{W}^*, \mu_{kl}) = (\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\mu_{kl}/4\sigma^2 + \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}.$$

As seen, with a slight abuse of notation, we have retained $\ell(\cdot)$ and $\boldsymbol{\Psi}(\cdot)$ as function names, even though the arguments have changed. Obvious variations of this notation will be used in the following.

The dependence on the codeword pair is now only through the scaling factor μ_{kl}. Since (13) is a decreasing function of μ_{kl}, the error probability is dominated by the codeword pairs corresponding to the minimum μ_{kl}. Thus, only one such pair is considered in the optimization procedure. An optimal $\boldsymbol{W}$ (optimal in the sense that it minimizes the criterion function under consideration) could now, at least in principle, be determined by minimizing $\ell(\boldsymbol{W}\boldsymbol{W}^*, \mu_{\min})$ with respect to $\boldsymbol{W}$, while satisfying the power constraint. Here, $\mu_{\min}$ denotes the minimum μ_{kl}. However, a highly challenging optimization problem, with a criterion function possessing multiple minimas, would need to be solved.

In order to obtain a tractable solution, we take on an alternative approach involving a reparameterization. An inspection of both (13) and the power constraint suggests the parameterization $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{W}^*$. A two-step procedure is now used for finding an optimal solution to the problem outlined in the previous paragraph. Rewriting the criterion function and the constraints in terms of the new parameters gives the following optimization problem:

$$\boldsymbol{Z}_{\rm opt} = \arg \min_{\substack{\boldsymbol{Z} \\ \boldsymbol{Z}=\boldsymbol{Z}^* \succeq 0 \\ {\rm tr}(\boldsymbol{Z})=1}} \ell(\boldsymbol{Z}) \tag{14}$$

where $\boldsymbol{Z} \succeq 0$ means that $\boldsymbol{Z}$ is positive semidefinite.[3] With a slight abuse of notation, the performance criterion $\ell(\boldsymbol{W}\boldsymbol{W}^*, \mu_{\min})$ is written here as

$$\ell(\boldsymbol{Z}) = \boldsymbol{m}^*_{\boldsymbol{h}|\hat{\boldsymbol{h}}} \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}} \left((\boldsymbol{I}_N \otimes \boldsymbol{Z})\eta + \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\right)^{-1} \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}} \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}} - \log\det\left((\boldsymbol{I}_N \otimes \boldsymbol{Z})\eta + \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\right) \tag{15}$$

where $\eta = \mu_{\min}/4\sigma^2$. An optimal linear transformation is finally obtained as $\boldsymbol{W}_{\rm opt} = \boldsymbol{Z}^{1/2}_{\rm opt}$, where $(\cdot)^{1/2}$ is a matrix square root such that $\boldsymbol{Z}_{\rm opt} = \boldsymbol{W}_{\rm opt}\boldsymbol{W}^*_{\rm opt}$. Note that a square root always exists since $\boldsymbol{Z}_{\rm opt}$ is a nonnegative definite matrix. Clearly, the solution is not unique.

The described reformulation is attractive since (14) is now a convex optimization problem. To see that the criterion function is convex, first note that $\boldsymbol{\Psi}(\boldsymbol{Z}, \mu_{\min})$ defines an affine transformation of $\boldsymbol{Z}$ and that $\boldsymbol{\Psi}(\boldsymbol{Z}, \mu_{\min})$ is positive definite over the set of all positive semidefinite $\boldsymbol{Z}$. Since affine transformations preserve convexity [27], [28], we can now establish the convexity of the criterion function by showing that

$$\boldsymbol{m}^*_{\boldsymbol{h}|\hat{\boldsymbol{h}}} \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}} \boldsymbol{\Psi}^{-1} \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}} \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}} + \log\det(\boldsymbol{\Psi}^{-1}) \tag{16}$$

is convex over the set of positive definite matrices $\boldsymbol{\Psi}$.

To see that the first term is convex, we utilize a theorem saying that a function is convex over a set $\mathcal{X}$ if it is convex when restricted to any line that intersects $\mathcal{X}$ [29, p. 94]. For this purpose, let $\boldsymbol{\Psi} = \lambda\boldsymbol{\Psi}_1 + (1-\lambda)\boldsymbol{\Psi}_2$, where $\boldsymbol{\Psi}_1 = \boldsymbol{\Psi}_1^*$, $\boldsymbol{\Psi}_2 = \boldsymbol{\Psi}_2^*$ are arbitrary positive definite matrices and $0 \le \lambda \le 1$ represent any line in the set of positive definite matrices. By making use of the identity

$$\frac{\partial \boldsymbol{X}(\theta)^{-1}}{\partial\theta} = -\boldsymbol{X}(\theta)^{-1}\frac{\partial \boldsymbol{X}(\theta)}{\partial\theta}\boldsymbol{X}(\theta)^{-1}$$

the second-order derivative of the first term with respect to λ is easily obtained as

$$2\boldsymbol{m}^*_{\boldsymbol{h}|\hat{\boldsymbol{h}}} \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}} \boldsymbol{\Psi}^{-1}(\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2)\boldsymbol{\Psi}^{-1}(\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2)\boldsymbol{\Psi}^{-1} \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}} \boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}.$$

This quadratic form is nonnegative since the matrix

$$\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}} \boldsymbol{\Psi}^{-1}(\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2)\boldsymbol{\Psi}^{-1}(\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2)\boldsymbol{\Psi}^{-1} \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}$$

is positive semidefinite. Because the second-order derivative is nonnegative it follows that the first term is convex with respect to λ (see e.g., [29, p. 91]) and thus, according to the theorem in [29, p. 94], also over the set of positive definite matrices $\boldsymbol{\Psi}$.

The convexity of the second term is established in, e.g., [23, p. 466]. Consequently, (16) is convex over all positive definite $\boldsymbol{\Psi}$. Due to the affine relation between $\boldsymbol{\Psi}$ and $\boldsymbol{Z}$, the criterion function is convex also with respect to $\boldsymbol{Z}$. It is easily verified that the constraints are convex [27], [28]. The entire optimization problem is therefore convex, which implies that all local minima are also global minima.

We will not go into great detail describing an algorithm that solves this particular optimization problem, since there are a number of standard techniques that are applicable. For example, interior point methods can be used for efficiently solving this kind of problem [30].

A. Asymptotic Properties of the Solution

Although the optimization problem given by (14) must, in general, be solved numerically, there are a few special cases that permit a closed-form solution. These special cases concern the asymptotic properties of the solution. In particular, here our attention is turned to the behavior of the solution when the channel quality is perfect and when there is no channel information, respectively. In addition, the influence of the SNR level is investigated. The solutions turn out to agree well with intuition and allow for some interesting interpretations. Detailed derivations can be found in Appendix I.

In the first case, no channel knowledge is assumed, i.e., $\|\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\| \to 0$. The criterion function used in (14) is then minimized for $\boldsymbol{Z}_{\rm as} = \boldsymbol{I}_M/M$. As a result, the optimal linear transformation is a scaled unitary matrix, an obvious choice given by $\boldsymbol{W}_{\rm as} = \boldsymbol{I}_M/\sqrt{M}$. Thus, the codewords are transmitted without modification. This makes sense considering the assumptions under which the predetermined space–time code was designed. It also makes sense in view of the fact that the transmitter does not know the channel and therefore has to choose a "neutral" solution. We refer to the resulting transmission technique as conventional orthogonal space–time block coding (OSTBC).

The second case concerns infinite SNR, in the sense that $\eta = \mu_{\min}/4\sigma^2 \to \infty$. Similarly to the case of no channel knowledge, the optimal linear transformation can be chosen to be a scaled unitary matrix. This indicates that the usefulness of

[3] In general, we take $\boldsymbol{A} \succ \boldsymbol{B}$ and $\boldsymbol{A} \succeq \boldsymbol{B}$ to mean that the matrix $\boldsymbol{A} - \boldsymbol{B}$ is positive definite and positive semidefinite, respectively.

channel knowledge diminishes as the SNR increases. Simulation results described in Section VI further support this claim.

In the third case, the channel knowledge is assumed to be perfect. Let $\boldsymbol{\Omega}_k^{(m)}$ denote the kth block of size $M \times M$ on the diagonal of $\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}^*$ and define $\boldsymbol{\Theta}$ as

$$\boldsymbol{\Theta} = \sum_{k=1}^{N} \boldsymbol{\Omega}_k^{(m)}. \tag{17}$$

To simplify the analysis, it is assumed that one of the eigenvalues of $\boldsymbol{\Theta}$ is strictly larger than all the other. This assumption is further commented on in the following. Studying the behavior of the solution as $\|\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\| \to 0$ gives the following asymptotically optimal linear transformation:

$$\boldsymbol{W}_{\mathrm{as}} = [\,\boldsymbol{v}_M \quad \boldsymbol{0} \quad \cdots \quad \boldsymbol{0}\,] \tag{18}$$

where $\boldsymbol{v}_M$ is the eigenvector of $\boldsymbol{\Theta}$ corresponding to the largest eigenvalue.

Due to the special structure of orthogonal space–time block codes, and since only one column of $\boldsymbol{W}_{\mathrm{as}}$ is nonzero, (18) may be interpretated as beamforming in the direction of $\boldsymbol{v}_M$. To see this, consider, for example, the two transmit antenna case and assume that the codewords of the predetermined space–time code are given by

$$\overline{\boldsymbol{C}} = \begin{bmatrix} s(n) & s(n+1) \\ -s(n+1)^* & s(n)^* \end{bmatrix} \tag{19}$$

where $s(n)$ is a sequence representing the data symbols to be transmitted. The code in (19) is the well-known Alamouti space–time code [13]. By utilizing the asymptotic result in (18) and the expression for the space–time code it is seen that the signal transmitted over the two antennas during time instant n and $n+1$ can be written as

$$\boldsymbol{C} = \boldsymbol{W}_{\mathrm{as}}\overline{\boldsymbol{C}} = [\,\boldsymbol{c}(n) \;\; \boldsymbol{c}(n+1)\,] = [\,\boldsymbol{v}_M s(n) \;\; \boldsymbol{v}_M s(n+1)\,].$$

Clearly, beamforming in the direction of $\boldsymbol{v}_M$ is performed. The present development can be generalized to all the orthogonal space–time block codes found in [4].

Note that because of the perfect channel knowledge, $\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}} \to \boldsymbol{h}$ in the mean square sense, i.e., $\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}$ is essentially the same as $\boldsymbol{h}$. Consequently, $\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}^* \to \boldsymbol{hh}^*$ which, in turn, means that $\boldsymbol{\Theta} \to \boldsymbol{HH}^*$, both in mean square. Hence, for all practical purposes, $\boldsymbol{v}_M$ can be considered equal to the left singular vector of $\boldsymbol{H}$ corresponding to the largest singular value [23, p. 414]. Thus, the transmission is now conducted in much the same way as in a scheme which utilizes the singular value decomposition of the channel matrix to convert the MIMO system into a set of parallel subchannels. Such a method was examined, for example, in [14] where a water-filling procedure was used for allocating transmit power (and thus distribution of data rates) among all the subchannels. However, our transmission scheme differs, among other things, in that only the strongest subchannel is used. This is due to the structure of the underlying orthogonal space–time block code. Because of the orthogonality, the decoding of the constituent data symbols decouples, allowing the transmission scheme to be studied by considering the symbols separately from each other. For example, consider again the Alamouti code in (19) and observe that the contribution to the transmitted signal from $s(n)$ is

$$\begin{aligned}\boldsymbol{C} &= \boldsymbol{W}_{\mathrm{as}} \begin{bmatrix} s(n) & 0 \\ 0 & s(n)^* \end{bmatrix} \\ &= [\,\boldsymbol{c}(n) \quad \boldsymbol{c}(n+1)\,] = [\,\boldsymbol{w}_1 s(n) \quad \boldsymbol{w}_2 s(n)^*\,]\end{aligned}$$

where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are the columns of $\boldsymbol{W}_{\mathrm{as}}$. It is now clear that to maximize the SNR for both $\boldsymbol{c}(n)$ and $\boldsymbol{c}(n+1)$, the two columns of $\boldsymbol{W}_{\mathrm{as}}$ should be matched to the channel, i.e., both should be parallel to the strongest left singular vector of the channel. Similar arguments apply to $s(n+1)$ and also to other orthogonal space–time block codes.

Also, note that the assumption that $\boldsymbol{\Theta}$ has a strictly largest eigenvalue is weak. The reason why is because of the often random nature of $\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}$ (or $\boldsymbol{h}$, since $\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}} \to \boldsymbol{h}$), and thus also of $\boldsymbol{\Theta}$, in practical fading scenarios. In particular, the probability that the assumption is violated is, except for some degenerate choices of the first- and second-order moments of the channel, in general vanishingly small for the Gaussian fading model used here. This is, for example, the case in the simplified fading scenario described in Section V.

Finally, in the fourth case, we consider an SNR value tending to zero, i.e., $\eta \to 0$. It turns out that the result is similar to the one derived in the previous case. Hence, the asymptotically optimal linear transformation is again given by (18). However, $\boldsymbol{\Theta}$ is now defined as

$$\boldsymbol{\Theta} = \sum_{k=1}^{N} \boldsymbol{\Omega}_k^{(m)} + \boldsymbol{\Omega}_k^{(R)}$$

where $\boldsymbol{\Omega}_k^{(R)}$ is the kth block of size $M \times M$ on the diagonal of $\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}$. Again, the existence of an eigenvalue of $\boldsymbol{\Theta}$ that is strictly larger than all the other is assumed.

Note that in the case of one receive antenna, the beamforming strategy proposed in [17], although derived using a different performance criterion, is seen to also give a beamformer proportional to $\boldsymbol{v}_M$. The approach taken on in [17] was to maximize the average SNR. As also pointed out there, such a performance criterion makes sense for small SNR values. Hence, the result for the fourth case in this section provides a generalization of the corresponding result in [17] to multiple receive antennas when a predetermined space–time code is used.

B. *An Algorithm for a Simplified Scenario*

In this subsection, we consider a simplified fading scenario in order to obtain a semi-closed-form solution of the optimization problem given in (14). In spite of the existence of a fairly efficient numerical optimization technique for the general case, the complexity of the algorithm described in this section is substantially lower.

Let us first make the simplifying assumption that the conditional covariance matrix is diagonal, also expressed as $\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}} = \alpha \boldsymbol{I}_{MN}$. Here, α represents the conditional variance of the channel coefficients. A scenario where this assumption is reasonable is considered in Section V. By introducing

$$\hat{\boldsymbol{\Upsilon}} = \frac{1}{\alpha}\sum_{k=1}^{N} \boldsymbol{\Omega}_k^{(m)}$$

and neglecting parameter independent terms, it is now possible to rewrite the performance criterion (15) in the following way:

$$\begin{aligned}\ell(\boldsymbol{Z}) &= \frac{1}{\alpha}\,\boldsymbol{m}^*_{\boldsymbol{h}|\hat{\boldsymbol{h}}}((\boldsymbol{I}_N \otimes \boldsymbol{Z}\alpha\eta) + \boldsymbol{I}_{MN})^{-1}\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\\ &\quad - \log\det((\boldsymbol{I}_N \otimes \boldsymbol{Z}\alpha\eta) + \boldsymbol{I}_{MN})\\ &= \frac{1}{\alpha}\,\mathrm{tr}\left((\boldsymbol{I}_N \otimes (\boldsymbol{Z}\alpha\eta + \boldsymbol{I}_M))^{-1}\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\boldsymbol{m}^*_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\right)\\ &\quad - \log\det(\boldsymbol{I}_N \otimes (\boldsymbol{Z}\alpha\eta + \boldsymbol{I}_M))\\ &= \mathrm{tr}\left((\boldsymbol{Z}\alpha\eta + \boldsymbol{I}_M)^{-1}\hat{\boldsymbol{\Upsilon}}\right) - N\log\det(\boldsymbol{Z}\alpha\eta + \boldsymbol{I}_M)\end{aligned}$$

where the second equality is due to the well-known relation $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$. In order to minimize this criterion we let $\boldsymbol{Z} = \boldsymbol{V\Lambda V}^*$ and $\hat{\boldsymbol{\Upsilon}} = \hat{\boldsymbol{V}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}}^*$ represent the eigenvalue decomposition (EVD) of $\boldsymbol{Z}$ and $\hat{\boldsymbol{\Upsilon}}$, respectively. The diagonal elements of $\boldsymbol{\Lambda}$ and $\hat{\boldsymbol{\Lambda}}$, representing the eigenvalues, are here denoted by $\{\lambda_i\}_{i=1}^M$ and $\{\hat{\lambda}_i\}_{i=1}^M$, respectively. In each set, the eigenvalues are assumed to be sorted in ascending order. It is also assumed that $\boldsymbol{V}$ and $\hat{\boldsymbol{V}}$ are unitary. Substituting for this new parameterization into $\ell(\boldsymbol{Z})$ and into the constraints results in an equivalent optimization problem given by the criterion function

$$\ell(\{\lambda_i\}_{i=1}^M, \boldsymbol{V}) = \mathrm{tr}\left((\boldsymbol{\Lambda}\alpha\eta + \boldsymbol{I}_M)^{-1}\boldsymbol{V}^*\hat{\boldsymbol{V}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}}^*\boldsymbol{V}\right) - N\log\det(\boldsymbol{\Lambda}\alpha\eta + \boldsymbol{I}_M) \quad (20)$$

subject to the constraints

$$\sum_{k=1}^M \lambda_k = 1 \quad (21)$$

$$\lambda_i \geq 0, \qquad i = 1, \ldots, M \quad (22)$$

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M. \quad (23)$$

It is seen that $\boldsymbol{V}$ is independent of the constraints and that it only affects the first term in (20). Keeping $\boldsymbol{\Lambda}$ constant and following the development in [29, p. 131], the optimum $\boldsymbol{V}$ can then be chosen as $\boldsymbol{V} = \hat{\boldsymbol{V}}$. For this to hold, (23) is needed.

The remaining optimization problem is clearly convex. The solution may therefore be obtained by means of the Karush–Kuhn–Tucker (KKT) conditions [29, p. 164]. Temporarily relaxing the problem by omitting the last constraint, and then finding a set of eigenvalues which satisfy the KKT conditions for the relaxed problem, is the approach used for deriving the solution. A detailed derivation is provided in Appendix II. The optimal eigenvalues for the relaxed problem turn out to be given by

$$\lambda_i = \max\left\{0,\ \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta}\right\} \quad (24)$$

where μ is the Lagrange multiplier corresponding to the power constraint. Note that this is also the optimum for the original problem since the above solution automatically satisfies (23).

The value of μ is obtained by inserting (24) into the power constraint (21) and solving the resulting equation. One possible procedure for accomplishing this is now described. To start with, assume that the number of eigenvalues equal to zero in the optimum solution is known. Let $l-1$ denote this quantity. Inserting (24) into the power constraint (21) then gives the equation

$$1 + \frac{M-l+1}{\alpha\eta} - \sum_{k=l}^M \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2N^2 + 4\alpha\eta\hat{\lambda}_k\mu}}{2\alpha\eta\mu} = 0 \quad (25)$$

from which μ can be determined. Let $f(\mu, l)$ represent the left-hand side of the equation. Since $f(\mu, l)$ is strictly increasing as a function of μ, the solution is unique and may be found numerically. For example, applying Newton's method gives rapid convergence. In this case, a suitable starting value is

$$\mu = \frac{\alpha\eta(M-l+1)^2}{(M-l+1+\alpha\eta)^2}\left(\hat{\lambda}_1 + \frac{N(M-l+1+\alpha\eta)}{M-l+1}\right)$$

obtained by using equal power on all eigenvectors whose eigenvalues are assumed to be nonzero. In order to arrive at the correct value of l, an iterative approach is used where, starting at $l = 1$, successive values of l are tried. An algorithm similar to the one utilized when computing the well-known water-filling power profile can be used for this purpose [32, p. 253]. The optimum linear transformation is finally obtained by an appropriate matrix square root of $\boldsymbol{Z}$. Thus, the whole procedure can be summarized as follows.

1) Set $l = 1$.
2) Solve $f(\mu, l) = 0$ with respect to μ.
3) Compute
$$\lambda_i = \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta}, \quad i = l, \ldots, M.$$
4) If $\lambda_l < 0$ then set $\lambda_l = 0$, $l = l + 1$ and repeat from 2).
5) Compute $\boldsymbol{W}_{\mathrm{opt}} = \hat{\boldsymbol{V}}\boldsymbol{\Lambda}^{1/2}$.

Note that we have tacitly assumed that the predetermined space–time code is designed for M transmit antennas, since $\boldsymbol{W}$ is a square matrix. However, the algorithm for the simplified scenario can easily be adapted to also handle the important case of an $M \times M'$ linear transformation, where M' denotes the number of rows in the predetermined codeword matrix and where $M' \leq M$. Toward this end, the starting value of l should be modified to $M - M' + 1$ and only the columns of $\boldsymbol{W}_{\mathrm{opt}}$ that correspond to $\{\lambda_i\}_{i=M-M'+1}^M$ should be retained from execution step 5). In this way, a simple predetermined code, designed for a small number of transmit antennas, can be used in conjunction with a much larger antenna array. Such a transmission scheme is also interesting in view of the fact that orthogonal space–time codes exist for only a limited number of transmit antennas [4].

V. A Simplified Scenario

Let us now detour from the general complex Gaussian fading assumption and instead consider a simplified fading scenario. The transmission scheme from Section IV-B will in this section be tailored specifically to this scenario.

In the simplified scenario, it is assumed that the antennas at both the transmitter and the receiver are spaced sufficiently far apart so that the fading is independent. A rich scattering environment with non-line-of-sight conditions is also assumed.

It is then reasonable to model the true channel coefficients h_{ij} as independent and identically distributed (i.i.d.) zero-mean complex Gaussian. Let σ_h^2 denote the variance of each individual channel coefficient. The coefficients of the channel estimates $\hat{h}_{ij}$ are modeled in the same way. Similarly to [17], each estimated channel coefficient $\hat{h}_{ij}$ is assumed to be correlated with the corresponding true channel coefficient h_{ij}, and uncorrelated with all others. In order to describe the degree of correlation, introduce the normalized correlation coefficient $\rho = E[h_{ij}\hat{h}_{ij}^*]/\sigma_h^2$. Thus, assuming h_{ij} and $\hat{h}_{ij}$ are jointly complex Gaussian, the distribution of the true channel and the side information is completely characterized by the covariance matrices $\boldsymbol{R}_{\boldsymbol{hh}} = \sigma_h^2 \boldsymbol{I}_{MN}$, $\boldsymbol{R}_{\boldsymbol{h}\hat{\boldsymbol{h}}} = \sigma_h^2 \rho \boldsymbol{I}_{MN}$, $\boldsymbol{R}_{\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}} = \sigma_h^2 \boldsymbol{I}_{MN}$, and the mean vectors $\boldsymbol{m}_{\boldsymbol{h}} = \boldsymbol{m}_{\hat{\boldsymbol{h}}} = \boldsymbol{0}$. Straightforward calculations show that this model leads to a conditional channel distribution described by

$$\boldsymbol{m}_{\boldsymbol{h}|\hat{\boldsymbol{h}}} = \rho \hat{\boldsymbol{h}} \qquad \boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}} = \sigma_h^2(1-|\rho|^2)\boldsymbol{I}_{MN}. \tag{26}$$

Although the previous measure of channel quality, as represented by $\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}$, can be retained in this scenario, we opt for ρ as the quantity describing the channel quality. Such a quantity was also used in [17]. Perfect channel knowledge now corresponds to $\rho \to 1$. As seen from (26), this in turn implies that $||\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}|| \to 0$. Hence, we also have perfect channel quality as defined by our original channel quality measure. On the other hand, no channel knowledge corresponds to $\rho \to 0$. For this case, $||\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}^{-1}||$ does not tend to zero. The two measures thus disagree. However, what seems like an inconsistency is really not, since the asymptotically optimum linear transformation can, for this case, be shown to be $\boldsymbol{W}_{\mathrm{as}} = \boldsymbol{I}_M/\sqrt{M}$, regardless of which of the two quality measures is used. The similarity in the asymptotic results is explained by the inherent symmetry in the distribution implied by (26) as $\rho \to 0$. Due to the symmetry, the distribution can be considered noninformative from the perspective of a transmitter, resulting in an open-loop type of system. This is clearly not true in general, since even if $\boldsymbol{h}$ and $\hat{\boldsymbol{h}}$ are uncorrelated, the distribution of the true channel represents a form of channel knowledge on its own.

A. Applying the Transmission Scheme

This subsection deals with how the transmission scheme that was described in Section IV-B can be customized for the simplified scenario. In addition, the behavior of the optimal linear transformation is studied.

In order to use the transmission scheme for the simplified scenario, α and $\hat{\boldsymbol{\Upsilon}}$ need to be computed. Based on (26), it is seen that

$$\hat{\boldsymbol{\Upsilon}} = \frac{1}{\alpha}\sum_{k=1}^{N} \boldsymbol{\Omega}_k^{(m)} = \frac{|\rho|^2}{\alpha}\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^* \tag{27}$$

where $\alpha = \sigma_h^2(1-|\rho|^2)$. It is now straightforward to apply the algorithm described in Section IV-B.

Let us investigate how the transmission scheme distributes the available power. Assuming perfect side information, i.e., $\rho \to 1$, the asymptotic result for the perfect channel knowledge case, discussed in Section IV-A, is applicable. Hence, all the power is allocated to the eigenvector $\boldsymbol{v}_M$, corresponding to the largest eigenvalue of $\hat{\boldsymbol{\Upsilon}}$, i.e., only λ_M is nonzero. In view of (27), it is clear that $\boldsymbol{v}_M$ is also the strongest left singular vector of $\hat{\boldsymbol{H}}$.

The second case considers no channel knowledge, i.e., $\rho \to 0$. From (27) it is obvious that $\hat{\boldsymbol{\Upsilon}}$ then tends to zero, which means that the corresponding eigenvalues $\{\hat{\lambda}_i\}_{i=1}^M$ also tend to zero. Hence, from (24), it follows that $\{\lambda_i\}_{i=1}^M$ will all be equal. As a result, $\boldsymbol{W}_{\mathrm{as}} = \hat{\boldsymbol{V}}/\sqrt{M}$. Such a linear transformation implies that $\boldsymbol{Z}_{\mathrm{as}} = \boldsymbol{I}_M/M$, which constitutes a transmission scheme equivalent[4] to conventional OSTBC.

For the special case when the number of antennas at either the transmitter or the receiver is two or lower, i.e., $\min\{M, N\} \le 2$, the transmission scheme can be further simplified. Only two of the eigenvalues $\hat{\lambda}_i$ are then nonzero, since $\hat{\boldsymbol{\Upsilon}}$ is the sum of N rank-one matrices of size $M \times M$. This allows the transmission scheme to be simplified by reorganizing the terms in (25) and then squaring repeatedly so that a polynomial equation is obtained. For example, consider a system with one receive antenna and assume the simplified scenario. It follows that

$$\hat{\boldsymbol{\Upsilon}} = \frac{|\rho|^2}{\alpha}\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^*.$$

Analytical expressions for the eigenvalues, as well as for the eigenvector corresponding to the largest eigenvalue, are easily found to be given by

$$\hat{\lambda}_1 = \cdots = \hat{\lambda}_{M-1} = 0, \qquad \hat{\lambda}_M = \frac{|\rho|^2}{\alpha}||\hat{\boldsymbol{h}}||^2$$

and

$$\boldsymbol{v}_M = \frac{\hat{\boldsymbol{h}}}{||\hat{\boldsymbol{h}}||} \tag{28}$$

respectively. Tedious but straightforward calculations now show that theprocedure for determining the optimum eigenvalues reduces to the following.

1) Let $\kappa = \alpha(M + \alpha\eta)$ and compute the equation at the bottom of the following page.
2) Compute $\lambda = \frac{1}{\mu} - \frac{1}{\alpha\eta}$.
3) If $\lambda > 0$ then set $\lambda_1 = \cdots = \lambda_{M-1} = \lambda$ and compute $\lambda_M = 1 - (M-1)\lambda$
4) If $\lambda \le 0$ then set $\lambda_1 = \cdots = \lambda_{M-1} = 0$, $\lambda_M = 1$

Although we have assumed the simplified fading scenario, the development generalizes easily to all scenarios where $\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}$ is diagonal. One important example of such a scenario is line-of-sight conditions in which the mean value of the true channel is nonzero, e.g., an environment with Ricean fading.

By analyzing the above procedure it is possible to make some interesting observations regarding the distribution of power among the eigenmodes. The expression for λ in the second step of the procedure is clearly decreasing as a function of $||\hat{\boldsymbol{h}}||$. Hence, when $||\hat{\boldsymbol{h}}||$ is above some threshold, the expression after the comparison in step four will be executed and all the power is allocated to the direction of the channel estimate $\hat{\boldsymbol{h}}$. On the other hand, falling below the threshold means that a part of the total power is allocated to $\hat{\boldsymbol{h}}$ and the remaining power is

[4]Equivalent in the sense that the corresponding values of the criterion function are the same.

divided equally between the $M-1$ directions orthogonal to the channel estimate. Recall that σ^2 is inversely proportional to η. A slightly more involved analysis then shows that the power distribution behaves similarly with respect to the noise variance σ^2 as well. The opposite behavior is observed for the fading variance σ_h^2. Thus, when σ_h^2 is below a certain threshold all power is allocated to $\hat{\boldsymbol{h}}$, whereas exceeding the same threshold leads to a portion of the total power being allocated to $\hat{\boldsymbol{h}}$ and the remaining power equally divided among the orthogonal directions. Simulation results presented in Section VI illustrate how the allocation of power affects the performance.

In the case of only two transmit antennas, a closed-form expression for $\boldsymbol{W}_{\text{opt}}$ may be formulated. For this purpose, let $\hat{h}_1$ and $\hat{h}_2$ denote the two elements of $\hat{\boldsymbol{h}}$. The eigenvector $\boldsymbol{v}_2$ is obtained from (28), whereas the other remaining eigenvector is obtained by forming a vector orthogonal to $\hat{\boldsymbol{h}}$. It follows that the optimal linear transformation can be written as

$$\boldsymbol{W}_{\text{opt}} = \frac{1}{\sqrt{|\hat{h}_1|^2 + |\hat{h}_2|^2}} \begin{bmatrix} -\hat{h}_2^* & \hat{h}_1 \\ \hat{h}_1^* & \hat{h}_2 \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix}.$$

Thus, our transmission scheme basically consists of a threshold test and some simple computations which are easily implemented using a lookup table. The complexity of the algorithm must therefore be considered very low. Note that this agrees well with the information-theoretic results outlined in [33], where a similar threshold effect was observed.

Based on the assumptions in the present section and using the corresponding transmission scheme, $\boldsymbol{W}_{\text{opt}}$ may now be efficiently determined. However, in order for the optimization to be carried out, the variances σ^2 and σ_h^2 and the correlation coefficient ρ must be known. In practice, these may be estimated at the receiver and fed back to the transmitter. Another approach is to treat them as design parameters chosen such that they roughly match the conditions the system is operating in. Nevertheless, in the simulations to follow, we assume these parameters are perfectly known at the transmitter.

VI. Simulation Results

In order to examine the performance of the proposed transmission scheme, and to investigate how it compares with conventional methods, simulations were conducted for several different cases. The performance was compared with three other methods—conventional OSTBC, conventional beamforming, and, what is here referred to as, ideal beamforming. Ideal beamforming is similar to conventional beamforming except that the beamformer is based on perfect channel knowledge.

For all examined cases, the simplified scenario with perfect knowledge of σ^2, σ_h^2, and ρ was assumed. The variance of the channel coefficients was arbitrarily set at $\sigma_h^2 = 1$. The channel was constant during the transmission of a codeword and independently fading from one codeword to another. Furthermore, the predetermined orthogonal space–time block codes were taken from the rate one codes found in [4]. The particular code used in each case is therefore directly determined by the number of transmit antennas. All elements of the codewords were taken from a binary phase-shift keying (BPSK) constellation. The input to the space–time encoder was assumed to form an i.i.d. sequence of equally probable symbol alternatives. Throughout the simulations, the bit error rate (BER) was used as the performance measure. Finally, the SNR was defined as

$$\text{SNR} = \frac{E\left[\|\boldsymbol{H}^*\boldsymbol{C}\|_{\text{F}}^2\right]}{LMN\sigma^2}.$$

For a conventional OSTBC system, the expression for the SNR is equal to the total received average signal power divided by the total noise power. Since the codes under consideration span as many time instants as the number of transmit antennas, L is here equal to M.

A. Varying the SNR

In the first case, a system with two transmit antennas and one receive antenna was considered. The channel quality was set to $\rho = 0.9$. The BER as a function of the SNR for the various transmission methods is depicted in Fig. 2. As seen, the performance of the proposed transmission scheme is for all SNR values better than conventional OSTBC but, as expected, worse than ideal beamforming. As the SNR decreases, the curve for the proposed scheme approaches the one for ideal beamforming whereas for increasing SNR it approaches the performance of conventional OSTBC. Thus, the proposed scheme combines the advantages of both beamforming and OSTBC. This is also in good agreement with both the asymptotic results of Section IV-A as well as the observations in Section V-A regarding the allocation of power among the eigenmodes. Note that the two curves for conventional OSTBC and ideal beamforming also show the performance of our transmission scheme in the case of $\rho \to 0$ and $\rho \to 1$, respectively. Conventional beamforming is seen to give good performance at low SNR values, but as the SNR increases, the lack of correct channel knowledge leads to a serious performance degradation.

In the second case, the number of transmit antennas was increased to eight. This was done in order to illustrate how the number of transmit antennas influences the performance. The channel quality was now set to $\rho = 0.7$. The BER versus the SNR for the four methods are presented in Fig. 3. As seen, the potential gains due to channel knowledge are now considerably higher. These gains remain to a large extent even when the number of receive antennas is increased, as illustrated in

$$\mu = \frac{\eta\left(\kappa(2M-1) + |\rho|^2\left\|\hat{\boldsymbol{h}}\right\|^2 + \sqrt{2\kappa(2M-1)|\rho|^2\left\|\hat{\boldsymbol{h}}\right\|^2 + |\rho|^4\left\|\hat{\boldsymbol{h}}\right\|^4 + \kappa^2}\right)}{2(M+\alpha\eta)^2}.$$

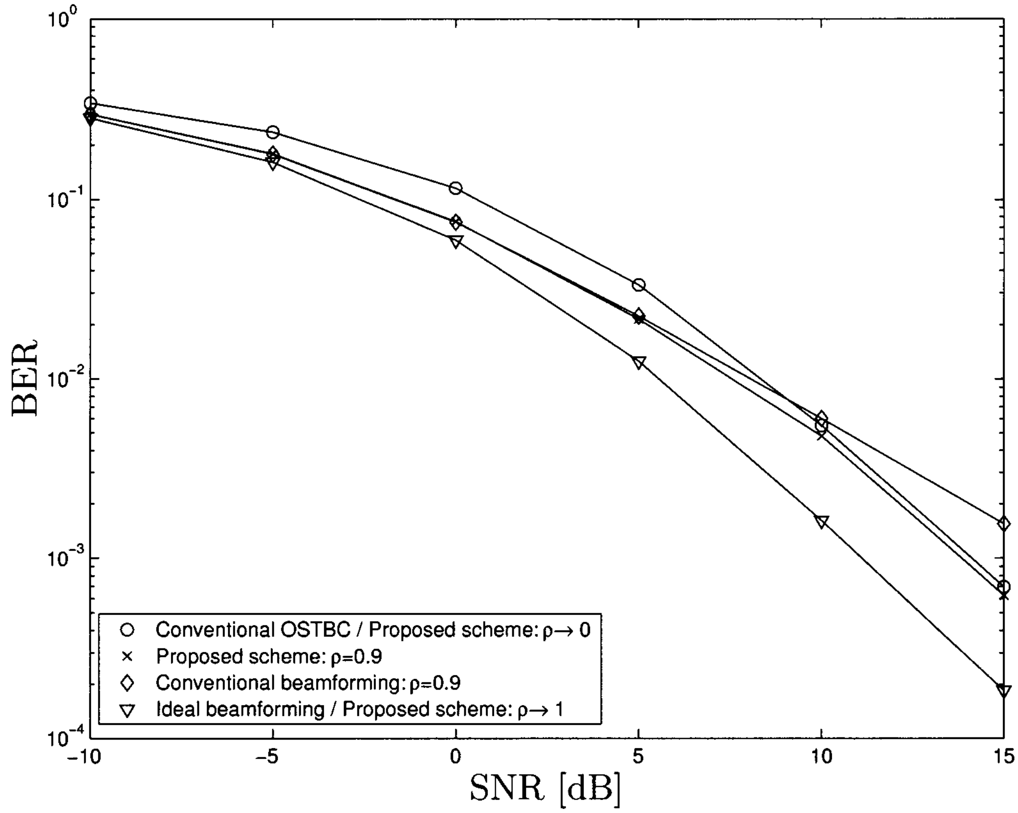

Fig. 2. Two transmit antennas, one receive antenna, and BPSK modulation.

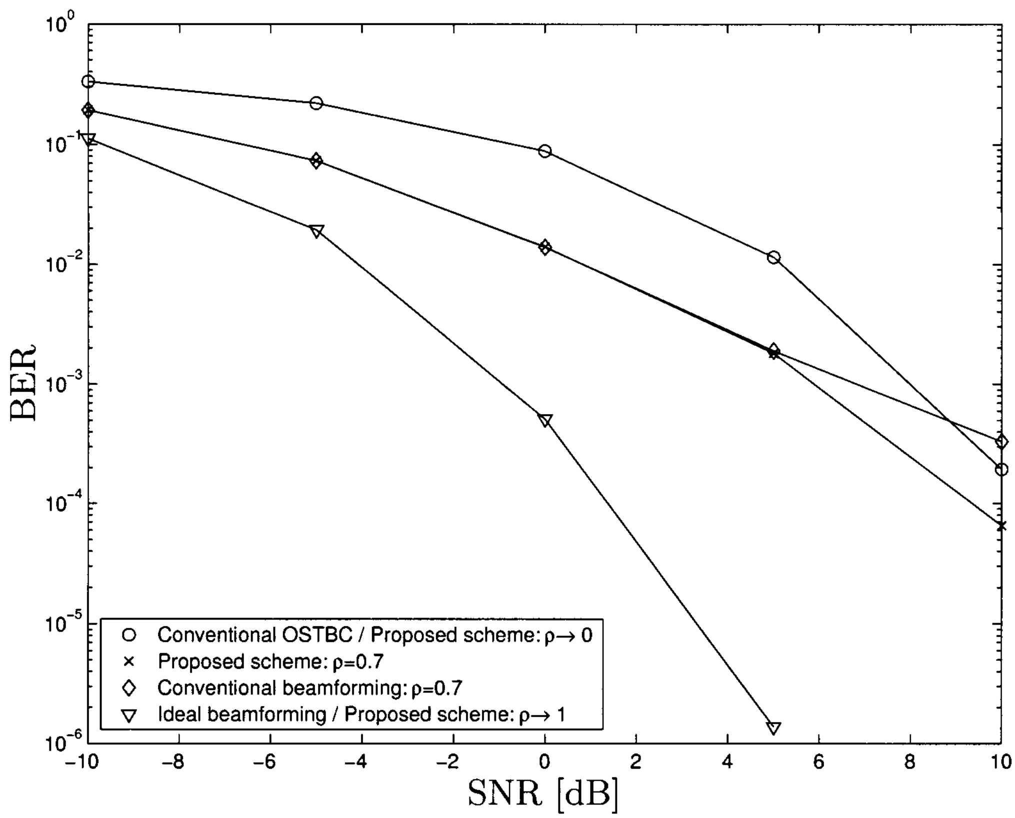

Fig. 3. Eight transmit antennas, one receive antenna, and BPSK modulation.

a comparison between the proposed method and conventional OSTBC in Fig. 4. Although not presented, simulation results demonstrating significant gains were also obtained for scenarios with fewer transmit antennas.

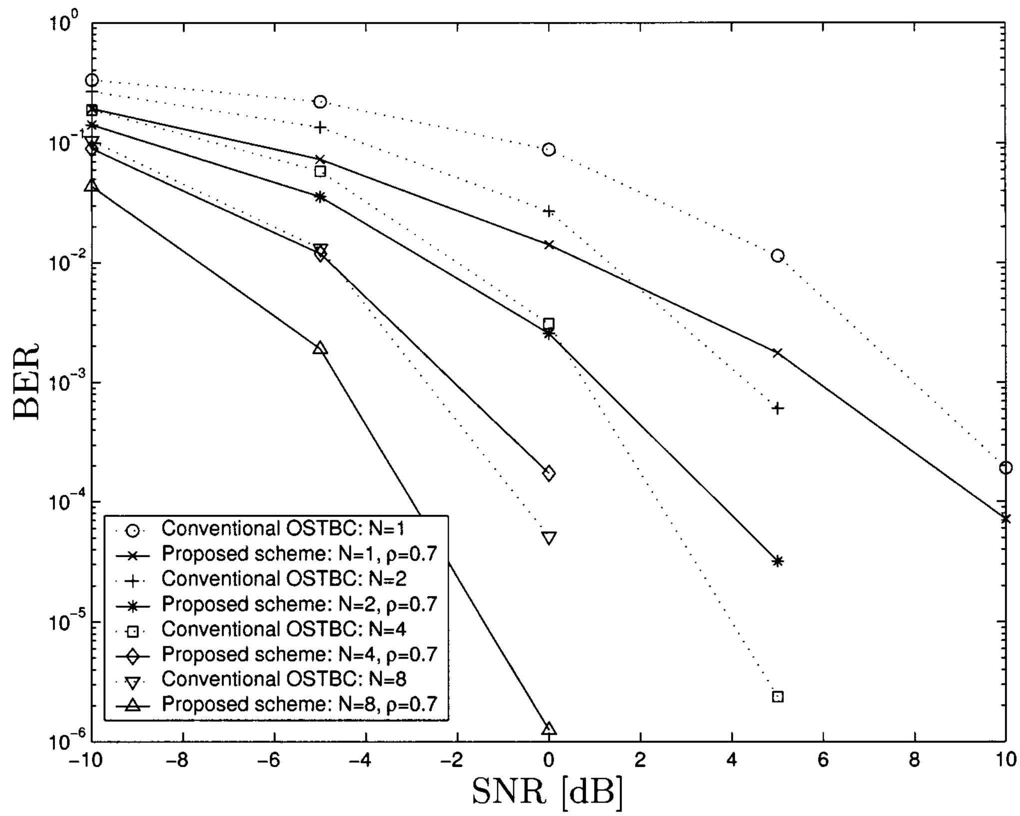

Fig. 4. Eight transmit antennas and BPSK modulation.

B. Varying the Channel Quality

The third and last case concerns how the channel quality affects the performance. Again, a system with two transmit antennas and one receive antenna is considered. The SNR was set at 10 dB and the BER versus the channel quality was plotted. The result is shown in Fig. 5, which thus provides an illustration of how the proposed scheme adapts to the variations in the channel quality. Hence, when the channel quality is low it is similar to conventional OSTBC and when it is high it is essentially the same as ideal beamforming.

VII. Conclusion

In this work, side information was utilized for improving a predetermined orthogonal space–time block code by means of a linear transformation. A transmission scheme that effectively combines conventional transmit beamforming with orthogonal space–time block coding was proposed. The resulting optimization problem was shown to be convex and could therefore be solved efficiently. Closed-form solutions were derived under certain asymptotic assumptions. Furthermore, the assumption of a simplified fading scenario resulted in a particularly efficient optimization algorithm. Numerical results demonstrated significant gains over both an open-loop system and a system using conventional beamforming.

Appendix I
Asymptotic Results

The strategy for deriving the asymptotic results presented in Section IV-A is to make use of the fact that, under certain conditions, it is possible to interchange the order of the limit and minimization operator, i.e.,

$$\begin{aligned}\boldsymbol{x}_{\rm as} &= \lim_{\rho\to a} \arg\min_{\boldsymbol{x}\in\mathcal{X}} V_\rho(\boldsymbol{x})\\ &= \arg\min_{\boldsymbol{x}\in\mathcal{X}} \lim_{\rho\to a} V_\rho(\boldsymbol{x})\\ &= \arg\min_{\boldsymbol{x}\in\mathcal{X}} \overline{V}(\boldsymbol{x})\end{aligned}$$

where $\mathcal{X}$ denotes the feasibility set and $\overline{V}(\boldsymbol{x}) \triangleq \lim_{\rho\to a} V_\rho(\boldsymbol{x})$. From [34, p. 221] it follows that this holds if $V_\rho(\boldsymbol{x})$ converges uniformly in $\boldsymbol{x}$ over $\mathcal{X}$ to the limit function $\overline{V}(\boldsymbol{x})$, $\mathcal{X}$ is a compact set (i.e., closed and bounded), and $\overline{V}(\boldsymbol{x})$ is continuous and has a unique global minimum.

To apply this theorem to the problem at hand, introduce a criterion function $\ell'(\boldsymbol{Z})$ that is equal to the original criterion function $\ell(\boldsymbol{Z})$, except for parameter-independent terms and factors. Let $\overline{\ell}_i(\boldsymbol{Z}) = \lim \ell(\boldsymbol{Z})$, where the limit is taken as either $\|\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\| \to 0$, $\eta\to\infty$, $\|\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\| \to 0$, or $\eta\to 0$, depending on the asymptotic case under consideration. Moreover, define the set of allowable parameters as

$$\mathcal{Z}(\varepsilon) = \{\boldsymbol{Z} | \boldsymbol{Z} = \boldsymbol{Z}^* \succeq \varepsilon \boldsymbol{I}_M,\ \mathrm{tr}(\boldsymbol{Z}) = 1\}. \tag{29}$$

Clearly, the requirement that $\mathcal{Z}(\varepsilon)$ is compact is satisfied. Normally, ε is taken to be zero. The set in (29) then corresponds to the feasibility set of the original optimization problem, as described in (14). However, in order to satisfy the requirement of a continuous limit function $\overline{\ell}_i(\boldsymbol{Z})$, we will for some of the cases first restrict $\mathcal{Z}(\varepsilon)$ by assuming that ε is small and positive and then argue why we can let $\varepsilon = 0$ without affecting the re-

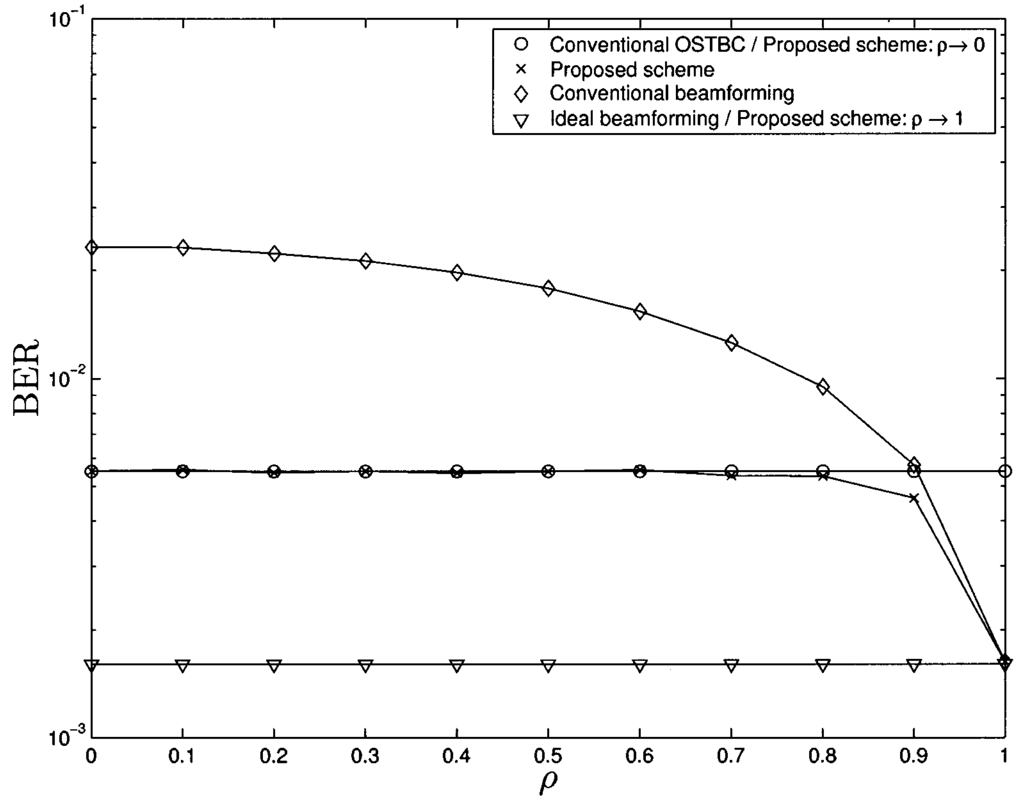

Fig. 5. Two transmit antennas, one receive antenna, SNR $= 10$ dB, and BPSK modulation.

sult. Proving that $\ell(\boldsymbol{Z})$ converges uniformly to $\bar{\ell}_i(\boldsymbol{Z})$ over $\mathcal{Z}(\varepsilon)$ amounts to showing that

$$\lim \sup_{\boldsymbol{Z}\in\mathcal{Z}(\varepsilon)} \left|\ell'(\boldsymbol{Z}) - \bar{\ell}_i(\boldsymbol{Z})\right| = 0.$$

In order to simplify the notation, the criterion function is, for the remaining part of this section, written as

$$\ell(\boldsymbol{Z}) = \boldsymbol{m}^*\boldsymbol{R}^{-1}((\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \boldsymbol{R}^{-1})^{-1}\boldsymbol{R}^{-1}\boldsymbol{m} - \log\det((\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \boldsymbol{R}^{-1})$$

where we note that $\eta = \mu_{\min}/4\sigma^2$ is a quantity proportional to the SNR. Furthermore, recall that $\|\boldsymbol{X}\| = \sigma_{\max}$, where $\sigma_{\max}$ is the maximum singular value of $\boldsymbol{X}$. Keep also in mind that if the argument is a vector, the result is the usual vector norm.

A. Case 1: No Channel Knowledge

The first case that is considered is no channel knowledge, i.e., $\|\boldsymbol{R}^{-1}\| \rightarrow 0$. To remove parameter-independent terms in the limit function, the equivalent criterion function $\ell'(\boldsymbol{Z}) = \ell(\boldsymbol{Z}) + MN\log(\eta)$ is considered. For now, assume that $\varepsilon > 0$. Since $\boldsymbol{Z}$ is then nonsingular, $\ell'(\boldsymbol{Z})$ can be written as

$$\begin{aligned}\ell'(\boldsymbol{Z}) = &\boldsymbol{m}^*\boldsymbol{R}^{-1}((\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \boldsymbol{R}^{-1})^{-1}\boldsymbol{R}^{-1}\boldsymbol{m}\\ &- \log\det(\boldsymbol{I}_{MN} + (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2}\\ &\cdot \boldsymbol{R}^{-1}(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2})\\ &- \log\det(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + MN\log(\eta)\end{aligned} \tag{30}$$

where $(\cdot)^{1/2}$ is now a matrix square root with Hermitian symmetry. As shown below, this function converges uniformly in $\boldsymbol{Z}$ to the limit function

$$\begin{aligned}\bar{\ell}_1(\boldsymbol{Z}) &= \lim_{\|\boldsymbol{R}^{-1}\|\rightarrow 0} \ell'(\boldsymbol{Z}) = -\log\det(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + MN\log\eta\\ &= -N\log\det(\boldsymbol{Z}).\end{aligned} \tag{31}$$

The limit function is obviously continuous. By utilizing Lagrangian multipliers and an EVD of $\boldsymbol{Z}$, it is straightforward to show that $\bar{\ell}_1(\boldsymbol{Z})$, subject to $\boldsymbol{Z} \in \mathcal{Z}(\varepsilon)$, has a unique global minimum $\boldsymbol{Z}_{\text{as}} = \boldsymbol{I}_M/M$. This fact, and the uniform convergence in $\boldsymbol{Z}$ over $\mathcal{Z}(\varepsilon)$, implies that, for a fixed positive $\varepsilon \leq 1/M$

$$\boldsymbol{Z}_{\text{as}} = \lim_{\|\boldsymbol{R}^{-1}\|\rightarrow 0} \arg\min_{\boldsymbol{Z}\in\mathcal{Z}(\varepsilon)} \ell'(\boldsymbol{Z}) = \boldsymbol{I}_M/M. \tag{32}$$

However, the solution is valid even if $\varepsilon = 0$. The reason is that $\ell'(\boldsymbol{Z})$ is a convex function and the solution to the above problem does not render the first constraint tight. Hence, relaxing the first constraint to $\varepsilon = 0$ does not change the optimum. Since $\boldsymbol{W}\boldsymbol{W}^* = \boldsymbol{Z}$ and $\boldsymbol{Z}_{\text{as}} = \boldsymbol{I}_M/M$, the optimum linear transformation in the case of no channel knowledge may, therefore, be chosen as $\boldsymbol{W}_{\text{as}} = \boldsymbol{I}_M/\sqrt{M}$.

To see the uniform convergence in $\boldsymbol{Z}$ over $\mathcal{Z}(\varepsilon)$ as $\|\boldsymbol{R}^{-1}\| \rightarrow 0$, consider the difference

$$\begin{aligned}&\ell'(\boldsymbol{Z}) - \bar{\ell}_1(\boldsymbol{Z}) = \boldsymbol{m}^*\boldsymbol{R}^{-1}((\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \boldsymbol{R}^{-1})^{-1}\boldsymbol{R}^{-1}\boldsymbol{m}\\ &- \log\det(\boldsymbol{I}_{MN} + (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2}\boldsymbol{R}^{-1}(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2}).\end{aligned} \tag{33}$$

From [23, p. 471] it readily follows that

$$(\boldsymbol{A} + \boldsymbol{B})^{-1} \preceq \boldsymbol{B}^{-1}, \qquad \boldsymbol{A} \succeq 0,\ \boldsymbol{B} \succ 0 \tag{34}$$

and the first term in (33) is therefore upper-bounded as

$$|\boldsymbol{m}^*\boldsymbol{R}^{-1}((\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)+\boldsymbol{R}^{-1})^{-1}\boldsymbol{R}^{-1}\boldsymbol{m}|\leq\|\boldsymbol{m}\|^2\|\boldsymbol{R}^{-1}\|.$$

Since the determinant equals the product of the eigenvalues, the second term can be written as

$$\sum_{k=1}^{MN}\log(1+\lambda_k)$$

where λ_k is the kth eigenvalue of

$$(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)^{-1/2}\boldsymbol{R}^{-1}(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)^{-1/2}.$$

This matrix is Hermitian and positive definite which means that its eigenvalues and its singular values are the same. Thus, the eigenvalues can be upper-bounded as

$$\begin{aligned}\lambda_k=\sigma_k\leq\sigma_{\max}&=\|(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)^{-1/2}\boldsymbol{R}^{-1}(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)^{-1/2}\|\\&\leq\frac{\|(\boldsymbol{I}_N\otimes\boldsymbol{Z})^{-1/2}\|^2\|\boldsymbol{R}^{-1}\|}{\eta}\leq\frac{\|\boldsymbol{R}^{-1}\|}{\varepsilon\eta}\end{aligned}$$

where σ_k represents the kth singular value and $\sigma_{\max}$ denotes the largest singular value. The second equality is due to the fact that the spectral norm is equal to the maximum singular value of its argument. An upper bound to the second term in (33) may be formed as

$$\sum_{k=1}^{MN}\log(1+\lambda_k)\leq MN\log\left(1+\frac{\|\boldsymbol{R}^{-1}\|}{\varepsilon\eta}\right). \quad (35)$$

By utilizing the triangle inequality it is now clear that

$$\begin{aligned}&\sup_{\boldsymbol{Z}\in\mathcal{Z}(\varepsilon)}\left|\ell'(\boldsymbol{Z})-\bar{\ell}_1(\boldsymbol{Z})\right|\\&\qquad\leq\|\boldsymbol{m}\|^2\|\boldsymbol{R}^{-1}\|+MN\log\left(1+\frac{\|\boldsymbol{R}^{-1}\|}{\varepsilon\eta}\right).\end{aligned}$$

Since this expression, for a constant $\varepsilon>0$, clearly tends to zero as $\|\boldsymbol{R}^{-1}\|\to 0$, we have shown that $\ell'(\boldsymbol{Z})$ converges uniformly to $\bar{\ell}_1(\boldsymbol{Z})$ within the parameter set defined by $\mathcal{Z}(\varepsilon)$. This completes the derivation for the no channel knowledge case.

B. Case 2: Infinite SNR

In the second case it is assumed that the SNR tends to infinity, i.e., $\eta\to\infty$. Again, we start by assuming $\varepsilon>0$. Similarly to the previous case, an equivalent criterion function can be written as in (30), which also in this case converges uniformly in $\boldsymbol{Z}$ to

$$\bar{\ell}_2(\boldsymbol{Z})=\lim_{\eta\to\infty}\ell'(\boldsymbol{Z})=-N\log\det(\boldsymbol{Z})=\bar{\ell}_1(\boldsymbol{Z}).$$

To see that the convergence is uniform, consider the two terms in (33). Utilizing (34), the first term is now upper-bounded by

$$\|\boldsymbol{m}\|^2\|\boldsymbol{R}^{-1}\|^2\|(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)^{-1}\|\leq\frac{\|\boldsymbol{m}\|^2\|\boldsymbol{R}^{-1}\|^2}{\varepsilon\eta}$$

whereas the upper bound of the second term is again given by (35). Hence, we have

$$\begin{aligned}&\sup_{\boldsymbol{Z}\in\mathcal{Z}(\varepsilon)}|\ell'(\boldsymbol{Z})-\bar{\ell}_2(\boldsymbol{Z})|\\&\qquad\leq\frac{\|\boldsymbol{m}\|^2\|\boldsymbol{R}^{-1}\|^2}{\varepsilon\eta}+MN\log\left(1+\frac{\|\boldsymbol{R}^{-1}\|}{\varepsilon\eta}\right)\end{aligned}$$

which obviously tends to zero as the SNR tends to infinity. The convergence is therefore uniform. The arguments following (32) then show that the asymptotically optimal linear transformation may once more be chosen as $\boldsymbol{W}_{\rm as}=\boldsymbol{I}_M/\sqrt{M}$. This completes the derivation for the case of infinite SNR.

C. Case 3: Perfect Channel Knowledge

The third case concerns perfect channel knowledge. For the present and the next case, we can let $\varepsilon=0$. The original constraints are therefore assumed. Parameter-independent terms and factors in the limit function are removed by considering the equivalent criterion function

$$\begin{aligned}\ell'(\boldsymbol{Z})&=(\ell(\boldsymbol{Z})-\log\det(\boldsymbol{R})-\boldsymbol{m}^*\boldsymbol{R}^{-1}\boldsymbol{m})/\eta\\&=\boldsymbol{m}^*(\boldsymbol{I}_{MN}+(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)\boldsymbol{R})^{-1}\boldsymbol{R}^{-1}\boldsymbol{m}/\eta\\&\quad-\log\det(\boldsymbol{I}_{MN}+\boldsymbol{R}^{1/2}(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)\boldsymbol{R}^{1/2})/\eta\\&\quad-\boldsymbol{m}^*\boldsymbol{R}^{-1}\boldsymbol{m}/\eta.\end{aligned} \quad (36)$$

We start by showing that this function converges uniformly in $\boldsymbol{Z}$ to the obviously continuous limit function

$$\bar{\ell}_3(\boldsymbol{Z})=\lim_{\|\boldsymbol{R}\|\to 0}\ell'(\boldsymbol{Z})=-\boldsymbol{m}^*(\boldsymbol{I}_N\otimes\boldsymbol{Z})\boldsymbol{m}. \quad (37)$$

The Taylor series [23, p. 301]

$$(\boldsymbol{I}-\boldsymbol{X})^{-1}=\sum_{k=0}^{\infty}\boldsymbol{X}^k$$

valid if $\|\boldsymbol{X}\|<1$ is used for writing the first term in (36) as

$$\boldsymbol{m}^*\left(\boldsymbol{R}^{-1}-(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)+\sum_{k=2}^{\infty}(-(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)\boldsymbol{R})^k\boldsymbol{R}^{-1}\right)\boldsymbol{m}/\eta.$$

By exploiting the triangle inequality and the formula for a geometric series an upper bound of the infinite sum, for sufficiently small $\eta\|\boldsymbol{R}\|$, is obtained as

$$\begin{aligned}\left\|\sum_{k=2}^{\infty}(-(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)\boldsymbol{R})^k\boldsymbol{R}^{-1}\right\|&\leq\sum_{k=2}^{\infty}\|\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta\|^k\|\boldsymbol{R}\|^{k-1}\\&=\frac{\|\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta\|^2\|\boldsymbol{R}\|}{1-\|\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta\|\,\|\boldsymbol{R}\|}\\&=\frac{\eta^2\|\boldsymbol{Z}\|^2\|\boldsymbol{R}\|}{1-\eta\|\boldsymbol{Z}\|\,\|\boldsymbol{R}\|}\\&\leq\frac{\eta^2\|\boldsymbol{R}\|}{1-\eta\|\boldsymbol{R}\|}.\end{aligned}$$

For the last inequality, we used the fact that $\|\boldsymbol{Z}\|\leq 1$, which is due to the trace constraint on $\boldsymbol{Z}$. Now, let λ_k represent the kth eigenvalue of

$$\boldsymbol{R}^{1/2}(\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta)\boldsymbol{R}^{1/2}.$$

Since it then holds that

$$\lambda_k\leq\|\boldsymbol{R}^{1/2}\|^2\|\boldsymbol{I}_N\otimes\boldsymbol{Z}\eta\|=\eta\|\boldsymbol{R}\|\|\boldsymbol{Z}\|\leq\eta\|\boldsymbol{R}\|$$

the second term in (36) is upper-bounded by

$$MN\log(1+\eta\|\boldsymbol{R}\|)/\eta.$$

Finally, collecting the results for the first two terms yields

$$\sup_{\boldsymbol{Z}\in\mathcal{Z}(0)} \left|\ell'(\boldsymbol{Z})-\overline{\ell}_3(\boldsymbol{Z})\right| \leq \frac{\eta\|\boldsymbol{R}\|\;\|\boldsymbol{m}\|^2}{1-\eta\|\boldsymbol{R}\|}+MN\log(1+\eta\|\boldsymbol{R}\|)/\eta.$$

The right-hand side clearly tends to zero as $\|\boldsymbol{R}\| \to 0$. Hence, the convergence is uniform.

Changing the sign of the limit function in (37) and reparameterizing using $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{W}^*$ shows that the optimum of the limit function $\overline{\ell}_3(\boldsymbol{Z})$ is given by

$$\boldsymbol{W}_{\text{as}} = \arg \max_{\substack{\boldsymbol{W} \\ \|\boldsymbol{W}\|_{\text{F}}^2=1}} \boldsymbol{m}^*(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{m}. \tag{38}$$

To solve this, let $\boldsymbol{\Omega} = \boldsymbol{m}\boldsymbol{m}^*$ and define

$$\boldsymbol{\Theta} = \sum_{k=1}^{N} \boldsymbol{\Omega}_k \tag{39}$$

where $\boldsymbol{\Omega}_k$ denotes the kth block of size $M \times M$ on the diagonal of $\boldsymbol{\Omega}$. The cost function in the above optimization problem can then be written as

$$\begin{aligned}\boldsymbol{m}^*(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{m} &= \text{tr}((\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{\Omega}) \\ &= \text{tr}(\boldsymbol{W}^*\boldsymbol{\Theta}\boldsymbol{W}) \\ &= (\text{vec}(\boldsymbol{\Theta}^*\boldsymbol{W}))^*\text{vec}(\boldsymbol{W}) \\ &= (\text{vec}(\boldsymbol{W}))^*(\boldsymbol{I}_M \otimes \boldsymbol{\Theta})\text{vec}(\boldsymbol{W}) \end{aligned} \tag{40}$$

where the two last equalities are due to (4) and (5), respectively. The power constraint is written on the form $\|\text{vec}(\boldsymbol{W})\| = 1$. Such an optimization problem is readily solved utilizing the EVD of $\boldsymbol{I}_M \otimes \boldsymbol{\Theta}$. For this purpose, let λ_M denote the largest eigenvalue of $\boldsymbol{\Theta}$ and recall the assumption that it is strictly larger than all the other eigenvalues, i.e., λ_M is unique. It can easily be verified that the eigenvalues of $\boldsymbol{I}_M\otimes\boldsymbol{\Theta}$ are obtained by repeating the eigenvalues of $\boldsymbol{\Theta}$ M times. Hence, λ_M is also the largest eigenvalue of $\boldsymbol{I}_M\otimes\boldsymbol{\Theta}$, with multiplicity M. The set of optimum solutions of (38) is therefore given by the eigenspace associated with λ_M. Introducing the complex-valued scalars $\{\mu_k\}_{k=1}^M$, the solution can be written in the form

$$\text{vec}(\boldsymbol{W}_{\text{as}}) = \mu_1\boldsymbol{u}_1 + \mu_2\boldsymbol{u}_2 + \cdots + \boldsymbol{u}_M \tag{41}$$

where

$$\boldsymbol{u}_1 = \begin{bmatrix}\boldsymbol{v}_M \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0}\end{bmatrix},\ \boldsymbol{u}_2 = \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{v}_M \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0}\end{bmatrix},\ \ldots,\ \boldsymbol{u}_M = \begin{bmatrix}\boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{v}_M\end{bmatrix}$$

and $\boldsymbol{v}_M$ are the eigenvectors of $\boldsymbol{I}_M \otimes \boldsymbol{\Theta}$ and $\boldsymbol{\Theta}$, respectively, corresponding to λ_M. Here, $\boldsymbol{0}$ is an $M \times 1$ vector with all elements equal to zero. Using (41) all the solutions may also be expressed as

$$\boldsymbol{W}_{\text{as}} = [\,\mu_1\boldsymbol{v}_M \quad \mu_2\boldsymbol{v}_M \quad \cdots \quad \mu_M\boldsymbol{v}_M\,]$$

implying that

$$\boldsymbol{Z}_{\text{as}} = \boldsymbol{W}_{\text{as}}\boldsymbol{W}_{\text{as}}^* = \boldsymbol{v}_M\boldsymbol{v}_M^*\sum_{k=1}^{M}|\mu_k|^2.$$

Combining this with the power constraint $\|\boldsymbol{W}\|_{\text{F}}^2 = 1$ means that $\sum_{k=1}^M |\mu_k|^2 = 1$, and hence $\boldsymbol{Z}_{\text{as}} = \boldsymbol{v}_M\boldsymbol{v}_M^*$. Consequently, regardless of the unit norm vector $\text{vec}(\boldsymbol{W}_{\text{as}})$ chosen from the aforementioned eigenspace, it holds that $\boldsymbol{Z}_{\text{as}} = \boldsymbol{v}_M\boldsymbol{v}_M^*$, which is thus a unique minimum point of $\overline{\ell}_3(\boldsymbol{Z})$. Accordingly, the use of $\overline{\ell}_3(\boldsymbol{Z})$ in the asymptotical analysis is justified. Letting, for example, $\mu_1 = 1$, $\mu_2 = \cdots = \mu_M = 0$, and utilizing $\boldsymbol{Z}_{\text{as}} = \boldsymbol{W}_{\text{as}}\boldsymbol{W}_{\text{as}}^*$, an asymptotically optimum solution is given by

$$\boldsymbol{W}_{\text{as}} = [\,\boldsymbol{v}_M \quad \boldsymbol{0} \quad \cdots \quad \boldsymbol{0}\,].$$

As previously indicated, the solution is not unique. For example, permuting the columns gives the same value of the cost function.

D. Case 4: Zero SNR

In the fourth case, the SNR is assumed to tend to zero, i.e., $\eta \to 0$. The derivation is to a large extent similar to the previous case. The Taylor expansion

$$\log\det(\boldsymbol{I}+\boldsymbol{X}) = \text{tr}(\boldsymbol{X}) + \mathcal{O}(\|\boldsymbol{X}\|^2)$$

where $\mathcal{O}(\cdot)$ is the big ordo operator, is used to write the second term of (36) as

$$\text{tr}(\boldsymbol{R}^{1/2}(\boldsymbol{I}_N \otimes \boldsymbol{Z})\boldsymbol{R}^{1/2}) + \mathcal{O}(\eta).$$

Combining this with (37) results in the limit function

$$\begin{aligned}\overline{\ell}_4(\boldsymbol{Z}) &= \lim_{\eta\to 0}\ell'(\boldsymbol{Z}) \\ &= -\boldsymbol{m}^*(\boldsymbol{I}_N \otimes \boldsymbol{Z})\boldsymbol{m} - \text{tr}\big(\boldsymbol{R}^{1/2}(\boldsymbol{I}_N \otimes \boldsymbol{Z})\boldsymbol{R}^{1/2}\big).\end{aligned} \tag{42}$$

It is now evident from

$$\sup_{\boldsymbol{Z}\in\mathcal{Z}(0)} \left|\ell'(\boldsymbol{Z}) - \overline{\ell}_4(\boldsymbol{Z})\right| \leq \frac{\eta\|\boldsymbol{R}\|\|\boldsymbol{m}\|^2}{1-\eta\|\boldsymbol{R}\|} + |\mathcal{O}(\eta)|$$

that the convergence is uniform. Thus, after changing the sign of $\overline{\ell}_4(\boldsymbol{Z})$ and parameterizing in terms of $\boldsymbol{W}$, the cost function can be taken as

$$\boldsymbol{m}^*(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{m} + \text{tr}(\boldsymbol{R}^{1/2}(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{R}^{1/2}).$$

Using the relation $\text{tr}(\boldsymbol{A}\boldsymbol{B}) = \text{tr}(\boldsymbol{B}\boldsymbol{A})$, this expression can be rewritten as

$$\text{tr}((\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)(\boldsymbol{m}\boldsymbol{m}^* + \boldsymbol{R})) = \text{tr}((\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{\Omega}) \tag{43}$$

where now $\boldsymbol{\Omega} = \boldsymbol{m}\boldsymbol{m}^*+\boldsymbol{R}$. Finally, due to the similarity between (43) and (40), the development from the previous case shows that an asymptotically optimum linear transformation is given by

$$\boldsymbol{W}_{\text{as}} = [\,\boldsymbol{v}_M \quad \boldsymbol{0} \quad \cdots \quad \boldsymbol{0}\,]$$

where $\boldsymbol{v}_M$ is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{\Theta}$. Here, $\boldsymbol{\Theta}$ is again defined as in (39).

Appendix II
An Algorithm for a Simplified Scenario

In this appendix, the solution of the optimization problem defined by (20)–(23) is derived. It is easily seen that both the cost function and the feasibility set are convex. Thus, the solution is given by the KKT conditions. In order to simplify the development, the optimization problem is temporarily relaxed by omitting (23). For the remaining problem, the optimum is given by any $\{\lambda_k\}$ that satisfy the KKT conditions

$$\sum_{k=1}^{M} \lambda_k = 1 \tag{44}$$

$$\lambda_i \geq 0 \tag{45}$$

$$-\frac{\alpha\eta\hat{\lambda}_i}{(1+\alpha\eta\lambda_i)^2} - \frac{\alpha\eta N}{1+\alpha\eta\lambda_i} + \mu - \nu_i = 0 \tag{46}$$

$$\nu_i \geq 0 \tag{47}$$

$$\nu_i \lambda_i = 0 \tag{48}$$

where $i = 1, \ldots, M$ and where μ and ν_i are Lagrange multipliers for the power constraint and the inequality constraints, respectively. We start by solving for ν_i in (46) and substituting into (47) and (48). Thus, the last three conditions reduce to

$$\frac{\alpha\eta\hat{\lambda}_i}{(1+\alpha\eta\lambda_i)^2} + \frac{\alpha\eta N}{1+\alpha\eta\lambda_i} \leq \mu \tag{49}$$

$$\lambda_i \left(\mu - \frac{\alpha\eta\hat{\lambda}_i}{(1+\alpha\eta\lambda_i)^2} - \frac{\alpha\eta N}{1+\alpha\eta\lambda_i} \right) = 0. \tag{50}$$

First, assume that $\lambda_i > 0$. It follows that the second factor in (50) must be zero. Rewriting this condition as

$$\mu(1+\alpha\eta\lambda_i)^2 - \alpha\eta N(1+\alpha\eta\lambda_i) - \alpha\eta\hat{\lambda}_i = 0$$

and solving for λ_i gives

$$\lambda_i = \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2 N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta} \tag{51}$$

where the positive root was picked due to (45). Note that (49) is now satisfied by equality. Hence, we have a valid solution as long as (51) gives a positive result and (44) is satisfied. On the other hand, for the case of a nonpositive result, we let $\lambda_i = 0$. That this indeed satisfies the KKT conditions is seen by verifying that (49) is true. Since

$$\frac{\alpha\eta N + \sqrt{\alpha^2\eta^2 N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta} \leq 0$$

and $\lambda_i = 0$ implies that

$$\mu \geq \alpha\eta\hat{\lambda}_i + \alpha\eta N \tag{52}$$

it is obvious that all the KKT conditions, with the possible exception of (44), are satisfied. Finally, also (44) can be handled by writing the optimum eigenvalues as

$$\lambda_i = \max\left\{0, \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2 N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta}\right\} \tag{53}$$

and then solving for μ in (44). It is apparent that (53) gives eigenvalues that are sorted in ascending order. Therefore, the constraint that was initially omitted, i.e., (23), is automatically satisfied. Thus, (53) gives the optimum eigenvalues also for the original problem.

REFERENCES

[1] J. H. Winters, "The diversity gain of transmit diversity in wireless systems with rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 119–123, Feb. 1998.

[2] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 527–537, Apr. 1999.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[4] ——, "Space–time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.

[5] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs Tech. Memo., Mar. 1995.

[6] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[7] T. Hattori and K. Hirade, "Multitransmitter digital signal transmission by using offset frequency strategy in land–mobile radio telephone system," *IEEE Trans. Veh. Technol.*, vol. 27, pp. 231–238, Nov. 1978.

[8] A. Hiroike, F. Adachi, and N. Nakajima, "Combined effects of phase sweeping transmitter diversity and channel coding," in *Proc. IEEE Vehicular Technology Conf.*, May 1992.

[9] V. Weerackody, "Characteristics of simulated fast fading indoor radio channel," in *Proc. IEEE Vehicular Technology Conf.*, 1993, pp. 231–235.

[10] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," in *Proc. IEEE Int. Conf. Communications*, May 1993, pp. 1630–1634.

[11] G. J. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas," *AT&T Bell Labs Tech. J.*, pp. 41–59, 1996.

[12] J.-C. Guey, M. P. Fitz, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over rayleigh fading channels," in *Proc. IEEE Vehicular Technology Conf.*, 1996, pp. 136–140.

[13] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[14] G. G. Raleigh and J. M. Cioffi, "Spatio–temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.

[15] G. Jöngren and B. Ottersten, "Combining transmit antenna weights and orthogonal space–time block codes by utilizing side information," in *Proc. 33th Asilomar Conf. Signals, Systems and Computers*, Oct. 1999.

[16] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining transmit beamforming and orthogonal space–time block codes by utilizing side information," in *Proc. 1st IEEE Sensor Array and Multichannel Signal Processing Workshop*, Mar. 2000.

[17] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.

[18] A. Wittneben, "Optimal predictive TX combining diversity in correlated fading for microcellular mobile radio applications," in *Proc. IEEE Global Telecommunications Conf.*, 1995, pp. 48–54.

[19] R. Negi, A. M. Tehrani, and J. Cioffi, "Adaptive antennas for space–time coding over block-time invariant multi-path fading channels," in *Proc. IEEE Vehicular Technology Conf.*, May 1999, pp. 70–74.

[20] 3rd Generation Partnership Project (3GPP), "Physical channels and mapping of transport channels onto physical channels," report, TS25.211 V2.4.0, 1999.

[21] 3rd Generation Partnership Project (3GPP), "Physical layer procedures," report, 3GPP RAN 25.214 V1.3.0, 1999.

[22] W. C. Jakes, *Microwave Mobile Communications*. Piscataway, NJ: IEEE Press, 1994.

[23] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[24] A. Graham, *Kronecker Products and Matrix Calculus With Applications*. Chichester, U.K.: Ellis Horwood Ltd., 1981.

[25] S. Bäro, G. Bauch, and A. Hansmann, "Improved codes for space–time trellis-coded modulation," *IEEE Commun. Lett.*, vol. 4, pp. 20–22, Jan. 2000.

[26] G. Jöngren, M. Skoglund, and B. Ottersten, "Design of channel estimate dependent space–time codes based on quantized side information," *IEEE Trans. Commun.*, submitted for publication.

[27] S. Boyd and L. Vandenberghe. (1999) Introduction to convex optimization with engineering applications. Course Notes. [Online]. Available: http://www.stanford.edu/class/ee364

[28] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, Mar. 1996.

[29] M. S. Bazaraa and H. D. Sherali, *Nonlinear Programming: Theory and Algorithms*, 2nd ed. New York: Wiley, 1993.

[30] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM Pub., 1994.

[31] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Statist.*, vol. 34, pp. 122–148, Feb. 1963.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[33] E. Visotsky and U. Madhow, "Space–time precoding with imperfect feedback," in *Proc. IEEE Int. Symp. Information Theory*, June 2000, p. 312.

[34] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice Hall, 1989.