Optimal Linear Precoders for MIMO Wireless Correlated Channels With Nonzero Mean in Space–Time Coded Systems

Mai Vu, *Student Member, IEEE,* and Arogyaswami Paulraj, *Fellow, IEEE*

***Abstract*—This paper proposes linear precoder designs exploiting statistical channel knowledge at the transmitter in a multiple-input multiple-output (MIMO) wireless system. The paper focuses on channel statistics, since obtaining real-time channel state information at the transmitter can be difficult due to channel dynamics. The considered channel statistics consist of the channel mean and transmit antenna correlation. The receiver is assumed to know the instantaneous channel precisely. The precoder operates along with a space–time block code (STBC) and aims to minimize the Chernoff bound on the pairwise error probability (PEP) between a pair of block codewords, averaged over channel fading statistics. Two PEP design criteria are studied—minimum distance and average distance. The optimal precoder with an orthogonal STBC is established, using a convex optimization framework. Different relaxations then extend the solution to systems with nonorthogonal STBCs. In both cases, the precoder is a function of both the channel mean and the transmit correlation. A linear precoder acts as a combination of a multimode beamformer and an input shaping matrix, matching each side to the channel and to the input signal structure, respectively. Both the optimal beam direction and the power of each mode, obtained via a dynamic water-filling process, depend on the signal-to-noise ratio (SNR). Asymptotic analyses of the results reveal that, for all STBCs, the precoder approaches a single-mode beamformer on the dominant right singular vector of the channel mean as the channel K factor increases. On the other hand, as the SNR increases, it approaches an equipower multiple-mode beamformer, matched to the eigenvectors of the transmit correlation. Design examples and numerical simulation results for both orthogonal and nonorthogonal STBC precoding solutions are provided, illustrating the precoding array gain.**

***Index Terms*—Channel side information (CSI), correlated fading channels, multiple-input multiple-output (MIMO) wireless, precoding, space–time codes.**

I. Introduction

EXPLOITING channel side information (CSI) at the transmitter in a multiple-input multiple-output (MIMO) wireless system has been an active research area. Transmit CSI (CSIT) can enhance MIMO system performance by increasing the spectral efficiency or reducing the error probability [7], [10], [31]. A precoder is a transmit processing block that exploits the CSIT. The separation between a precoding function, which depends on the CSIT, and channel coding, which assumes no CSIT, has been shown to be capacity-optimal for fading channels, based on information-theoretic analyses for both single-antenna [8] and multiple-antenna systems [9]. These results apply to the CSI case studied in this paper.

Channel information can be obtained at the transmitter by measuring the reverse channel, invoking the reciprocity principle in time-division duplexing systems, or by using feedback from the receiver. The fluctuating nature of a wireless channel often induces unreliable instantaneous CSIT, due to a time delay or a frequency offset between the channel measurement and its use. An alternative is to use statistical information, such as the antenna correlations [29] and/or the channel mean [30]. These statistics vary at a much slower rate than the instantaneous channel and, therefore, can be obtained reliably at the transmitter. Statistical CSIT is especially relevant for fast time-varying channels. Other forms of CSIT include channel parameters, such as the channel K factor, a condition number, or the signal-to-noise ratio (SNR).

Research on exploiting CSIT uses a variety of criteria. One category is to study the optimal input signal characteristics that achieve the channel ergodic capacity with the CSIT condition. The problem reduces to finding the covariance of the optimal Gaussian input signal as a function of the CSIT. The optimal signal with this covariance is equivalent to an independent identically distributed (i.i.d.) signal going through an optimal linear precoder. Research in this category includes studies on exploiting antenna correlations [11]–[13], the channel mean [11], [14], [15], or the K factor and phase statistics [16].

Another category in utilizing the CSIT is to minimize the system error probability. A common setup includes a space–time block code (STBC) exploiting channel diversity and a linear precoder exploiting the CSIT. An STBC is often designed for i.i.d. Rayleigh fading channels, assuming no CSIT. A linear precoder functions as a multimode beamformer matching to the channel, based on the CSIT, with orthogonal beam directions as the left singular vectors, and the beam power-loadings as the squared singular values of the precoding matrix. The STBC and precoder combination, therefore, is robust to channel fading and can concurrently exploit the available CSIT. Moreover, if an STBC is capacity lossless for an i.i.d. MIMO channel, then combining this STBC with a linear

Manuscript received January 26, 2005; revised June 7, 2005. This work was supported in part by the Rambus Stanford Graduate Fellowship and the Intel Foundation PhD Fellowship. This work was also supported in part by NSF Contract DMS-0354674-001 and ONR Contract N00014-02-0088. Parts of this paper were presented at the IEEE Vehicular Technology Conference, Los Angeles, CA, September 26–29, 2004, and the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Montreal, QC, Canada, March 18–23, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Martin Haardt.

The authors are with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: mhv@stanford.edu; apaulraj@stanford.edu).

Digital Object Identifier 10.1109/TSP.2006.871960

precoder is capacity optimal for the channel with CSIT [9]. Such a setup also provides flexibility in designing precoders to adapt to various CSIT conditions, without changing the STBC or the detection scheme in an existing system.

Examples of prior literature using the error criteria include schemes exploiting imperfect channel estimates [18], [22], the channel mean feedback [19], or the transmit antenna correlation [20], [21]. These papers primarily focused on exploiting only one form of channel statistics: either the channel mean or the transmit antenna correlation ([18] formulates the problem for a general mean and covariance model, but only solves for the uncorrelated case). The linear precoder solutions reduce to fixed beam directions, given by the singular vectors or eigenvectors of the mean or correlation matrices, with per-beam power allocation, obtained by a water-filling solution dependent on the SNR.

In this paper, we design a precoder to simultaneously exploit both the channel mean and the transmit correlation to minimize the error probability in MIMO systems, using an approach similar to [18]. Exploiting both mean and transmit correlation allows us to address a wider range of channels. Our precoder designs can also be applied to the CSIT scenario of an imperfect channel estimate with a known error covariance. In our problem, due to the interaction of the mean and the transmit correlation, the precoder solution does not have fixed beam directions. Instead, both the direction and the power loading of each beam are functions of both the mean and correlation matrices and are dependent on the SNR. Thus, we call the process of solving for the optimal precoder a *dynamic* water-filling process. Asymptotic analyses reveal that the precoder depends primarily on the channel mean at high K factors and primarily on the transmit correlation at high SNRs.

The precoder obtains an array gain, attributed to the optimal beam directions and the water-filling type power allocation among these beams. Due to statistical CSIT, no diversity gain can be extracted from the channel information. Therefore, the diversity achieved by the system is controlled by the STBC. Our precoder design works with all STBCs. In addition, our framework focusing on extracting an array gain is fundamentally different from, and complementary to, achieving the diversity-multiplexing frontier at high SNRs as in [27] and [28].

The paper is organized as follows. Section II outlines the channel model and the system setup. Section III discusses the PEP measure and the design criteria. Section IV then proceeds to design the optimal precoder for a system with an orthogonal STBC. Section V extends the precoder result to systems with a general STBC. Asymptotic analyses of the precoder result at a high K factor and a high SNR are provided in Section VI. Special cases of the precoder design are discussed in Section VII. Section VIII then provides specific design examples and performance results. Finally, we conclude in Section IX.

Notations used in this paper are as follows: $(.)^*$ for conjugate transpose, $E[.]$ for expectation, $\mathrm{tr}(.)$ for trace, $\|.\|_F$ for the Frobenius norm, $\lambda(.)$ for eigenvalues, $\sigma(.)$ for singular values, $(.)^\star$ for optimal values, $(.)_+$ for a positive value inside the parenthesis or zero, $\succcurlyeq$ for the matrix positive semidefinite relation, $\succ$ for positive definite, boldface for matrices (except when referring to a generic matrix). When ordering matters, we use the eigenvalue (singular value) index to indicate an increased sorting order, i.e., for a matrix A with N real eigenvalues, $\lambda_1(A) \leq \lambda_2(A) \leq \ldots \leq \lambda_N(A)$.

II. Channel Model and System Setup

A. Channel Model

Consider a MIMO wireless communication system with N transmit and M receive antennas. The channel is frequency-flat and quasi-static block fading, represented by a matrix $\mathbf{H}$ of size $M \times N$. Assuming a non-zero-mean channel with a transmit antenna correlation, the channel matrix can be written in the form

$$\mathbf{H} = \sqrt{\frac{K}{K+1}}\mathbf{H}_0 + \mathbf{H}_w\sqrt{\frac{1}{K+1}}\mathbf{R}_0^{\frac{1}{2}}. \tag{1}$$

Here, K is the ratio of the power in the mean component to the average power in the random components of the channel; $\mathbf{H}_w$ is a complex Gaussian random matrix with independent zero-mean unit-variance entries, i.e., $\mathbf{H}_w \sim \mathcal{N}(0, \mathbf{I})$; $\mathbf{H}_0$ is the normalized channel mean, and $\mathbf{R}_0$ is the normalized transmit correlation matrix, such that

$$\begin{aligned} \mathrm{tr}\,(\mathbf{H}_0^*\mathbf{H}_0) &= MN \\ \mathrm{tr}(\mathbf{R}_0) &= N. \end{aligned} \tag{2}$$

This normalization maintains a constant average power gain of MN in the channel for any combination of mean and correlation matrices: $E[\mathrm{tr}(\mathbf{H}^*\mathbf{H})] = MN$. The transmit signal power is assumed to have been adjusted accordingly. The normalization ensures a fair performance comparison between systems operating on channels with different mean or correlation matrices and, therefore, allows an objective precoding gain measure (otherwise, a stronger channel power gain will likely result in a better precoding gain). For simpler expressions, except when studying the effect of the K factor on the precoder, we will use the following channel model in our analysis:

$$\mathbf{H} = \mathbf{H}_m + \mathbf{H}_w\mathbf{R}_t^{\frac{1}{2}} \tag{3}$$

where $\mathbf{H}_m = \sqrt{K/(K+1)}\mathbf{H}_0$ is the channel mean, and $\mathbf{R}_t = (1/K+1)\mathbf{R}_0$ is the transmit correlation matrix.

We assume that the transmit correlation matrix $\mathbf{R}_t$ is full rank and, hence, is invertible. The rank of the channel mean, on the other hand, can be arbitrary. The mean of the channel can correspond to the Rician line-of-sight component or to clusters of strong paths in the propagation environment. (For a study that explores Rician channels with a non-full-rank correlation, see [17], for example.) The mean and correlation values can also correspond to a channel estimate and its error covariance; although in that case, the statistics are only short term, as they change when a new estimate becomes available. Such a channel estimate with known error covariance model is established in [18].

The receiver is assumed to know the channel perfectly (i.e., it knows the channel realization $\mathbf{H}$), whereas the transmitter only knows the channel mean $\mathbf{H}_m$ and the transmit correlation $\mathbf{R}_t$. The mean and the correlation values are more stable than the instantaneous channel and can be obtained reliably at the transmitter, by either utilizing the reverse channel statistics or

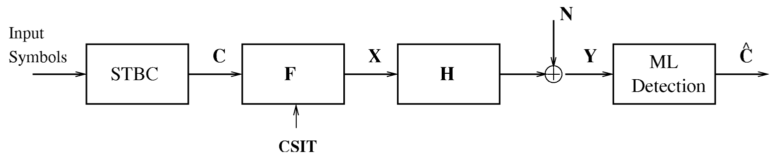

Fig. 1. System architecture with a linear precoder and a STBC.

using feedback from the receiver. These statistics can be obtained by time-averaging operations on channel measurements, over a window of tens of channel coherence times; they remain valid for tens to hundreds of channel coherence times. For example, suppose that we have causal channel measurements of $\mathbf{H}$ (3) up to the current time sample 0, where each measurement is created by a different realization of $\mathbf{H}_w$. Then, the channel mean and transmit correlation matrices are calculated as

$$\mathbf{H}_m = \frac{1}{L}\sum_{k=-L+1}^{0}\mathbf{H}_k$$
$$\mathbf{R}_t = \frac{1}{M}\left(\frac{1}{L}\sum_{k=-L+1}^{0}\mathbf{H}_k^*\mathbf{H}_k - \mathbf{H}_m^*\mathbf{H}_m\right)$$

where L is the averaging window, and $\mathbf{H}_k$ is the instantaneous channel measurement at the sample time k.

B. System Setup

We consider a system using a space–time block code to capture the channel diversity and a precoder $\mathbf{F}$ to capture the CSIT, depicted in Fig. 1. In such a system, the precoder can be viewed as a processing block to enhance performance in addition to an existing STBC, based on the available CSIT. This scenario covers many practical wireless systems, such as those in wireless local-area networks (802.11) and metropolitan-area networks (802.16). We restrict our attention to a linear precoder. The combination of a linear precoder and a space–time block code is capacity-optimal with CSIT, if the STBC is capacity lossless for an i.i.d. MIMO channel without CSIT [9]. The capacity-optimal signal for an i.i.d. MIMO channel without CSIT is zero-mean complex Gaussian with an identity covariance matrix [6]. Therefore, the linear precoder has the effect of shaping the transmit signal such that it has a covariance matrix optimal for the channel with CSIT.

At each time instance, the linear precoder functions as a beamformer with either one or multiple beams. The beam directions are the left singular vectors of the precoder matrix; the beam power loadings are the squared singular values. The right singular vectors of the precoding matrix combine the output symbols of the STBC to feed into each beam. The STBC helps to mitigate the effect of unknown fading in a channel with incomplete CSIT. The combination of a STBC and a precoder makes the system robust against fading, while delivering both diversity and antenna array gains. It also provides the flexibility of adapting to various CSIT conditions, without changing the STBC or the detection algorithm already implemented. Detection in a system with precoding is performed over the effective channel created by the precoder and the actual channel. The receiver can construct the precoder using the same algorithm and parameters as the transmitter, which implicitly implies that the receiver must know all the parameters that the transmitter knows. This assumption is reasonable since the receiver can usually obtain channel measurements more readily than the transmitter through the use of pilots, and both can agree on using the same precoding design algorithm.

To maintain a constant average sum transmit power, the precoding matrix must satisfy the power constraint

$$\mathrm{tr}(\mathbf{F}\mathbf{F}^*) = 1. \tag{4}$$

Let $\mathbf{C}$ of size $N \times T$ be the block codeword in the STBC, the receive signal block is then

$$\mathbf{Y} = \mathbf{H}\mathbf{F}\mathbf{C} + \mathbf{N}$$

where $\mathbf{N} \sim \mathcal{N}(0, \mathbf{I}\sigma^2)$ is the additive complex white Gaussian noise, with σ^2 being the noise power per spatial dimension. The receiver performs maximum-likelihood (ML) detection over a codeword $\mathbf{C}$ to obtain

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}\in\mathbb{C}} \|\mathbf{Y} - \mathbf{H}\mathbf{F}\mathbf{C}\|_F^2 \tag{5}$$

where $\mathbb{C}$ is the STBC codebook, and the subscript F here denotes the Frobenius norm. In this system, we analyze the performance without an outer error-correction code.

III. PEP Measure and Design Criteria

A. Pairwise Error Probability Measure

We consider the pairwise error probability (PEP), which is the probability that a transmitted codeword $\mathbf{C}$ has a worse detection metric than another codeword $\hat{\mathbf{C}}$. With ML detection (5) and applying the Chernoff bound, similar to [2], the PEP can be tightly upper-bounded by

$$P(\mathbf{C} \to \hat{\mathbf{C}}) \le \exp\left(-\frac{\left\|\mathbf{H}\mathbf{F}(\mathbf{C} - \hat{\mathbf{C}})\right\|_F^2}{4\sigma^2}\right). \tag{6}$$

The PEP depends on the specific codeword pair $(\mathbf{C}, \hat{\mathbf{C}})$. We define the following expression as the codeword distance product matrix

$$\mathbf{A} = \frac{1}{P}(\mathbf{C} - \hat{\mathbf{C}})(\mathbf{C} - \hat{\mathbf{C}})^* \tag{7}$$

where P is the average sum transmit power. The Chernoff bound function in (6) can then be written as

$$f(\mathbf{H}, \mathbf{A}, \mathbf{F}) = \exp\left(-\frac{\rho}{4}\mathrm{tr}(\mathbf{H}\mathbf{F}\mathbf{A}\mathbf{F}^*\mathbf{H}^*)\right) \tag{8}$$

where $\rho = P/\sigma^2$ is the SNR.

While the PEP is not the system codeword-error rate, it is a measure strongly related to the system performance. The system average codeword-error probability can be written as:

$$\bar{P}_e = E_{\mathbf{H}}\left[\sum_i p_i \mathrm{Pr}\left(\bigcup_{j\neq i}(\mathbf{C}_i \to \mathbf{C}_j)\right)\right]$$

where p_i is the probability of the codeword $\mathbf{C}_i$, and $(\mathbf{C}_i \to \mathbf{C}_j)$ is the event that $\mathbf{C}_i$ is mis-detected as $\mathbf{C}_j$. The above error expression is not very tractable and, therefore, does not lend itself to the analysis of the precoder design problem. We choose to use the PEP, averaged over the channel fading, as the performance criterion. The exact PEP expression, however, is still complex to analyze. Therefore, we aim to design a precoder that minimizes the Chernoff upper-bound on the PEP. The Chernoff bound tracks the performance well and provides an analytical framework for establishing closed-form precoder solutions. The PEP with the Chernoff bound is a framework commonly used in the literature [2], [18], [20].

B. Worst-Case and Average-Case Designs

Since the PEP is codeword-pair dependent, we need a design rule for picking the codeword distance product matrix $\mathbf{A}$ (7), over which the PEP is optimized. Noting that the Chernoff bound expression (8) is monotonic in $\mathbf{A}$ (specifically, $\mathbf{A}_1 \succcurlyeq \mathbf{A}_2$ implies $f(\mathbf{H}, \mathbf{A}_1, \mathbf{F}) \leq f(\mathbf{H}, \mathbf{A}_2, \mathbf{F})$), we confine the codeword distance selection to two options. The first is the minimum distance, corresponding to the largest PEP. Of interest is the average performance over channel fading; therefore, this criterion can be expressed as

$$\mathbf{F} = \arg\min_{\mathbf{F}}\left\{\max_{\mathbf{A}} E_{\mathbf{H}}\left[f(\mathbf{H}, \mathbf{A}, \mathbf{F})\right]\right\}. \tag{9}$$

If the occurrence probability of the minimum-distance codeword pairs is not small, these pairs will dominate the error performance; thus, the minimum-distance design will lead to a reasonable overall performance gain. This distance criterion guarantees a minimum precoding gain, based on the gain obtained from optimizing the Chernoff bound, and has often been used in the literature [18].

We also examine the average distance over all codeword pairs. Since the precoder only acts on one column of the codeword $\mathbf{C}$ at a time, and detection is performed jointly over the whole code-block of T symbol periods, we propose an average-distance measure as

$$\bar{\mathbf{A}} = \frac{1}{PT}E\left[(\mathbf{C}-\hat{\mathbf{C}})(\mathbf{C}-\hat{\mathbf{C}})^*\right] = \frac{1}{PT}\sum_{i\neq j} p_{ij}\Delta_{ij}\Delta_{ij}^* \tag{10}$$

where $\Delta_{ij} = \mathbf{C}_i - \mathbf{C}_j$, and p_{ij} is the probability of the pair $(\mathbf{C}_i, \mathbf{C}_j)$ among all pairs of distinct codewords. In effect, $\bar{\mathbf{A}}$ is the covariance of the codeword error. The design criterion in this case becomes

$$\mathbf{F} = \arg\min_{\mathbf{F}}\left\{E_{\mathbf{H}}\left[f\left(\mathbf{H}, E[\mathbf{A}], \mathbf{F}\right)\right]\right\}. \tag{11}$$

The average distance leads to a smaller value of the Chernoff bound (8) compared to the minimum distance; therefore, the gain obtained from optimizing this bound may not be guaranteed to be the minimum precoding gain of the system. However, the average distance has an advantage over the minimum distance for nonorthogonal STBCs, in that $\bar{\mathbf{A}}$ is more likely to be a scaled-identity matrix. The implication of a scaled-identity $\mathbf{A}$ will be discussed in Section V.

C. Optimization Problem

Assume that a codeword distance product matrix $\mathbf{A}$ has been chosen, based on the appropriate design choice. The objective is to find a precoder $\mathbf{F}$ to minimize the expression $E_{\mathbf{H}}[f(\mathbf{H}, \mathbf{A}, \mathbf{F})]$, where f is the Chernoff bound in (8). This problem formulation follows closely from [18]. Given the probability density distribution of the channel

$$g(\mathbf{H}) = \frac{\exp\left(-\mathrm{tr}\left[(\mathbf{H}-\mathbf{H}_m)^*\mathbf{R}_t^{-1}(\mathbf{H}-\mathbf{H}_m)\right]\right)}{\pi^{MN}\det(\mathbf{R}_t)^M}$$

after averaging (8) over the channel statistics, we obtain the following bound on the average PEP:

$$\bar{P}_e \leq \frac{\exp\left[\mathrm{tr}\left(\mathbf{H}_m\mathbf{W}^{-1}\mathbf{H}_m^*\right)\right]}{\det(\mathbf{W})^M}\det(\mathbf{R}_t)^M \times \exp\left[-\mathrm{tr}\left(\mathbf{H}_m\mathbf{R}_t^{-1}\mathbf{H}_m^*\right)\right] \tag{12}$$

where

$$\mathbf{W} = \frac{\rho}{4}\mathbf{R}_t\mathbf{F}\mathbf{A}\mathbf{F}^*\mathbf{R}_t + \mathbf{R}_t. \tag{13}$$

Expression (12) is a special case of the general problem setup in [18], in which the upper bound is given for a general mean and covariance CSIT model. [18] provides the precoder solution for a channel mean CSIT with an identity covariance (equivalent to uncorrelated antennas) in systems using orthogonal STBCs. The precoder design problem for a non-Kronecker covariance, or antenna correlation, still remains unsolved. Our formulation and solution apply to the CSIT in the form of an arbitrary mean and separable Kronecker antenna correlation [29], assuming uncorrelated receive antennas.

Minimizing the bound in (12) is equivalent to minimizing the logarithm of this bound; ignoring the constant terms leads to the following objective function:

$$J = \mathrm{tr}\left(\mathbf{H}_m\mathbf{W}^{-1}\mathbf{H}_m^*\right) - M\log\det(\mathbf{W}) \tag{14}$$

which is convex in the matrix variable $\mathbf{W}$. Combining this objective function with the precoding power constraint (4), an optimization problem for designing $\mathbf{F}$ can be posed as

$$\begin{aligned} \min_{\mathbf{F}} \quad & J = \mathrm{tr}\left(\mathbf{H}_m\mathbf{W}^{-1}\mathbf{H}_m^*\right) - M\log\det(\mathbf{W}) \\ \text{subject to} \quad & \mathbf{W} = \frac{\rho}{4}\mathbf{R}_t\mathbf{F}\mathbf{A}\mathbf{F}^*\mathbf{R}_t + \mathbf{R}_t \\ & \mathrm{tr}(\mathbf{F}\mathbf{F}^*) = 1. \end{aligned} \tag{15}$$

Due to the nonlinear power constraint, this problem is not convex in $\mathbf{F}$ and, hence, is not directly tractable in the original variable $\mathbf{F}$. The tractability of this problem depends on the structure of the matrix $\mathbf{A}$. Since $\mathbf{A}$ depends on the STBC, we will first solve the precoder design problem for orthogonal

STBCs, which lend an attractive property to $\mathbf{A}$, and then extend the analysis to general STBCs.

IV. Precoder Design With an Orthogonal STBC

A. The Optimization Problem for Precoding With an OSTBC

We first consider precoding with an orthogonal STBC (OSTBC) [3] specifically. Due to its orthogonality, the distance product matrix of an OSTBC has a special form

$$\mathbf{A} = (\mathbf{C} - \hat{\mathbf{C}})(\mathbf{C} - \hat{\mathbf{C}})^* = \mu \mathbf{I}$$

where μ represents the codeword distance, depending on the specific codeword pair.

Let μ_0 be the value corresponding to the matrix $\mathbf{A}$ used in the precoder optimization problem. Based on the design choice (Section III-B), μ_0 can either be the minimum distance or the average distance over all distinct-codeword pairs. In the numerical simulation section (Section VIII), we will give examples of how μ_0 is calculated in each case.

A scaled-identity matrix $\mathbf{A}$ helps to significantly simplify the optimization problem (15). The variable $\mathbf{W}$ is now a linear function of $\mathbf{FF}^*$ and (15) becomes

$$\begin{aligned} \min_{\mathbf{F}} \quad & J = \mathrm{tr}\left(\mathbf{H}_m \mathbf{W}^{-1} \mathbf{H}_m^*\right) - M \log\det(\mathbf{W}) \\ \text{subject to} \quad & \mathbf{W} = \frac{\mu_0 \rho}{4} \mathbf{R}_t \mathbf{F}\mathbf{F}^* \mathbf{R}_t + \mathbf{R}_t \\ & \mathrm{tr}(\mathbf{FF}^*) = 1. \end{aligned} \tag{16}$$

Denote $\eta_0 = {}^1/_4 \mu_0 \rho$. This problem can be cast in terms of $\mathbf{W}$ as follows:

$$\begin{aligned} \min_{\mathbf{W}} \quad & \mathrm{tr}\left(\mathbf{H}_m \mathbf{W}^{-1} \mathbf{H}_m^*\right) - M \log\det(\mathbf{W}) \\ & \text{subject to} \quad \mathrm{tr}\left(\mathbf{R}_t^{-1} \mathbf{W} \mathbf{R}_t^{-1} - \mathbf{R}_t^{-1}\right) = \eta_0 \\ & \qquad\qquad \mathbf{R}_t^{-1} \mathbf{W} \mathbf{R}_t^{-1} - \mathbf{R}_t^{-1} \succcurlyeq 0 \end{aligned} \tag{17}$$

where the last inequality results from the positive semidefinite (PSD) property that $\mathbf{FF}^* \succcurlyeq 0$. The new formulation (17) is convex in the matrix variable $\mathbf{W}$ and can be solved analytically.

B. Precoder Design Algorithm

Problem Analysis: We first analyze problem (17), using the Lagrangian dual method [32]. Let $\mathbf{\Phi}$ be the following function of $\mathbf{W}$:

$$\mathbf{\Phi}(\mathbf{W}) = \mathbf{R}_t^{-1} \mathbf{W} \mathbf{R}_t^{-1} - \mathbf{R}_t^{-1} \tag{18}$$

then the two constraints in (17) can be rewritten $\mathbf{\Phi}$ as

$$\mathrm{tr}(\mathbf{\Phi}) = \eta_0 \tag{19}$$

$$\mathbf{\Phi} \succcurlyeq 0. \tag{20}$$

Form the Lagrangian of (17) as

$$\begin{aligned} \mathcal{L}(\mathbf{W}, \nu, \mathbf{Z}) = & \ \mathrm{tr}\left(\mathbf{H}_m \mathbf{W}^{-1} \mathbf{H}_m^*\right) - M \log\det(\mathbf{W}) \\ & + \nu \left[\mathrm{tr}(\mathbf{\Phi}) - \eta_0\right] - \mathrm{tr}(\mathbf{Z\Phi}) \end{aligned} \tag{21}$$

where ν is the Lagrange multiplier associated with the equality constraint (19), and $\mathbf{Z} \succcurlyeq 0$ is the Lagrange multiplier in the matrix form associated with the inequality constraint (20). Strong duality holds for problem (17), easily verifiable using Slaters condition [32]. This condition requires the existence of a strictly feasible point: a positive definite matrix $\mathbf{\Phi}_0 \succ 0$ satisfying the equality constraint (19). An example is $\mathbf{\Phi}_0 = \mathbf{I}/N$. Therefore, the primal and dual optimal points of (17) satisfy the Karush–Kuhn–Tucker (KKT) conditions [32]

$$\begin{aligned} \mathrm{tr}(\mathbf{\Phi}^\star) &= \eta_0 \\ \mathbf{\Phi}^\star &\succcurlyeq 0 \\ \mathbf{Z}^\star &\succcurlyeq 0 \\ \mathrm{tr}(\mathbf{Z}^\star \mathbf{\Phi}^\star) &= 0 \\ \left.\frac{\partial \mathcal{L}(\mathbf{W}, \nu, \mathbf{Z})}{\partial \mathbf{W}}\right|_{\mathbf{W}^\star, \nu^\star, \mathbf{Z}^\star} &= 0 \end{aligned}$$

where $(.)^\star$ denotes the optimal value, and $\mathbf{\Phi}^\star$ is the value of $\mathbf{\Phi}$ evaluated at the optimal $\mathbf{W}^\star$.

In particular, the three conditions $\mathbf{\Phi}^\star \succcurlyeq 0$, $\mathbf{Z}^\star \succcurlyeq 0$ and $\mathrm{tr}(\mathbf{Z}^\star \mathbf{\Phi}^\star) = 0$ mean that all eigenvalues of the positive semidefinite product $\mathbf{Z}^\star \mathbf{\Phi}^\star$ are zero. Thus, $\mathbf{\Phi}^\star$ and $\mathbf{Z}^\star$ must have the same eigenvectors, and their eigenvalue patterns are complementary; that is, if $\lambda_i(\mathbf{\Phi}^\star) > 0$, then $\lambda_i(\mathbf{Z}^\star) = 0$, and vice versa. $\mathbf{Z}$ effectively ensures that $\mathbf{\Phi}$ is positive semidefinite: $\mathbf{Z}$ represents the eigenmodes that are dropped in a water-filling solution, whereas $\mathbf{\Phi}$ represents the modes that are active (with nonzero power). Therefore, if we distribute power over the correct number of positive eigenvalues of $\mathbf{\Phi}$ and let the rest of eigenvalues be zero, then in the Lagrangian (21), the term $\mathrm{tr}(\mathbf{Z\Phi})$ is automatically zero at the optimal values and, hence, can be ignored. From this observation, we present a two-step algorithm for solving problem (17) as outlined below.

Design Algorithm: First, we assume that the optimal $\mathbf{\Phi}^\star$ is full rank, effectively ignoring the PSD inequality constraint (20), and solve for $\mathbf{W}$. If the solution of this step produces $\mathbf{\Phi} \succcurlyeq 0$, then it is also the solution for the original problem (17). The matrix $\mathbf{\Phi}$ becomes the scaled precoder product $\eta_0 \mathbf{FF}^*$, which is full-mode in this case. However, if the solution of the first step does not produce a PSD $\mathbf{\Phi}$, we then proceed to the second step. In this step, we force the weakest eigenvalue of $\mathbf{\Phi}$ to be zero, effectively reducing the rank of $\mathbf{\Phi}$, and resolve the problem, iterating this step until $\mathbf{\Phi}$ is PSD. This second step is equivalent to mode-dropping in a water-filling process.

Ignoring $\mathrm{tr}(\mathbf{Z\Phi})$ in the Lagrangian (21), the optimality condition is obtained by differentiating this expression with respect to $\mathbf{W}$ [36] to arrive at

$$-\mathbf{W}^{-1} \mathbf{H}_m^* \mathbf{H}_m \mathbf{W}^{-1} - M \mathbf{W}^{-1} + \nu \mathbf{R}_t^{-2} = 0.$$

Multiply this equation on both the left and right with $\mathbf{W}$ to get a quadratic matrix equation. Solving the equation leads to the solution for $\mathbf{W}$ as [23]

$$\mathbf{W} = \frac{1}{2\nu} \mathbf{R}_t \left(M \mathbf{I}_N + \mathbf{\Psi}^{\frac{1}{2}}\right) \mathbf{R}_t,$$

where

$$\mathbf{\Psi} = M^2 \mathbf{I}_N + 4\nu \mathbf{R}_t^{-1} \mathbf{H}_m^* \mathbf{H}_m \mathbf{R}_t^{-1}. \tag{22}$$

From here, we need to find the Lagrange multiplier ν, based on the transmit power equality constraint (19).

1) Full-Mode Solution: The full-mode solution is obtained by solving for ν, assuming $\mathbf{\Phi}$ is full rank. Equation (19) can then be written as

$$\mathrm{tr}\left(\frac{1}{2\nu}\left(M\mathbf{I}_N + \mathbf{\Psi}^{\frac{1}{2}}\right) - \mathbf{R}_t^{-1}\right) = \eta_0. \quad (23)$$

Let $\lambda_i (i = 1 \ldots N)$ be the eigenvalues of $\mathbf{R}_t^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{R}_t^{-1}$ sorted in increasing order, $\beta_0 = 2[\mathrm{tr}(\mathbf{R}_t^{-1}) + \eta_0]$, and noting that $\lambda(M\mathbf{I} + A) = M + \lambda(A)$, the above equation becomes

$$MN + \sum_{i=1}^{N} \sqrt{M^2 + 4\nu\lambda_i} - \beta_0\nu = 0. \quad (24)$$

In the general case $(N > 1)$, this equation does not appear to have a closed-form solution. However, solving for ν can be done efficiently using a binary search, which we call the *inner algorithm*.

Inner Algorithm: Since the left-hand-side expression in (23) is monotonous in ν, the following lower and upper bounds on ν can be established:

$$\nu_{\text{lower}} = \frac{4N^2\lambda_1}{\beta_0^2} + \frac{2MN}{\beta_0}, \quad \nu_{\text{upper}} = \frac{4N\alpha_0}{\beta_0^2} + \frac{2MN}{\beta_0} \quad (25)$$

where λ_1 is the minimum eigenvalue and α_0 is the trace of $\mathbf{R}_t^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{R}_t^{-1}$. The lower bound is obtained from (24) by replacing all λ_i with λ_1, while the upper bound is obtained by applying the Cauchy–Schwartz inequality [34] to the summation term. A numerical binary search can then be carried out between these bounds to find the solution for (24) up to a desired precision. The number of iterations depends on the problem parameters, but convergence usually occurs rapidly since this is one-dimensional (1-D) binary search.

2) Mode-Dropping Solution: If the full-mode solution does not produce $\mathbf{\Phi} \succcurlyeq 0$, then we drop the weakest eigenmode of $\mathbf{\Phi}$ and resolve for ν. This step is equivalent to a water-filling iteration. The total power will now be distributed over the $N-1$ largest eigenvalues of $\mathbf{\Phi}$, and the power constraint (23) changes to

$$\sum_{i=k+1}^{N} \lambda_i\left(\frac{1}{2\nu}\left(M\mathbf{I}_N + \mathbf{\Psi}^{\frac{1}{2}}\right) - \mathbf{R}_t^{-1}\right) = \eta_0 \quad (26)$$

where $k = 1$, the number of modes dropped, and $\lambda_i(\cdot)$ is the ith eigenvalue of the matrix in the parenthesis, sorted in increasing order $(\lambda_1 \leq \cdots \leq \lambda_N)$. The solution for $\mathbf{\Phi}$ is obtained after solving this equation for ν, forming the right-hand-side expression in (18), and forcing its smallest eigenvalue to be zero, after which, it satisfies (19). If this solution does not satisfy (20), then the number of dropped modes k is increased by 1 and (26) is re-solved. Again, this equation does not have a closed-form solution. Using an efficient binary search, we design an algorithm to numerically solve for ν, which is termed the *outer algorithm*.

Outer Algorithm: There is no explicit function relating the eigenvalues of a general matrix sum to the individual eigenvalues; therefore, each eigenvalue in (26) cannot be written as an explicit function of ν (except in the special cases discussed in Section VII). Fortunately, noting that the sum of eigenvalues in (26) is monotonous in ν, we can derive upper and lower bounds on ν, then use a binary search to find the solution efficiently up to any desired numerical precision. The bounds in the general case with k modes dropped $(1 \leq k \leq N-1)$ are

$$\nu_{\text{upper}} = \frac{\lambda_N}{\beta_k^2} + \frac{M}{\beta_k}, \quad \nu_{\text{lower}} = \frac{\lambda_1}{\beta_k^2} + \frac{M}{\beta_k} \quad (27)$$

where λ_N and λ_1 are the maximum and minimum eigenvalues of $\mathbf{R}_t^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{R}_t^{-1}$, respectively, and

$$\beta_k = \frac{1}{N-k}\left(\eta_0 + \sum_{i=k+1}^{N} \frac{1}{\lambda_i(\mathbf{R}_t)}\right).$$

The derivation of these bounds is given in the Appendix.

3) Dynamic Water-Filling: The mode-dropping process above is similar to the water-filling process, in that at each outer iteration, an eigenmode (the weakest among the active modes) is dropped, and the total transmit power is reallocated over the rest of the modes. However, there is a significant difference between this process and the conventional water-filling process. In conventional water-filling, only the power allocation, or the water level, changes after each iteration, but the mode directions remain the same. In our problem, the mode directions also evolve at each iteration, due to the interaction of the channel mean and transmit correlation matrices. To see this effect more clearly, rewrite the expression for $\mathbf{\Phi}$ in the following form:

$$\mathbf{\Phi} = \frac{M}{2\nu}\mathbf{I}_N + \left(\frac{1}{2\nu}\mathbf{\Psi}(\nu)^{\frac{1}{2}} - \mathbf{R}_t^{-1}\right)$$

where the notation $\mathbf{\Psi}(\nu)$ illustrates the dependence of $\mathbf{\Psi}$ on ν. The "water-level" here is $M/2\nu$, and the mode directions are determined by the eigenvectors of the matrix expression inside the large parenthesis. When ν changes at each outer iteration, both the water-level (hence, the power allocation) and the mode directions change. Moreover, since the ν solution depends on the SNR ρ, both the precoder power allocation and mode (beam) directions are functions of the SNR. For this reason, we call this a *dynamic* water-filling process.

C. Optimal Precoder Solution With an OSTBC

When ν is found satisfying the PSD constraint (20), then the matrix $\mathbf{\Phi}$ provides the solution for $\mathbf{FF}^*$ as

$$\mathbf{\Phi} = \frac{1}{2\nu}\left(M\mathbf{I}_N + \mathbf{\Psi}^{\frac{1}{2}}\right) - \mathbf{R}_t^{-1} = \eta_0\mathbf{FF}^*. \quad (28)$$

From this product expression, an optimal precoder can be derived. The optimal precoder is not unique. Let the eigenvalue decomposition of $\mathbf{\Phi}$ be

$$\mathbf{\Phi} = \mathbf{U}_\Phi\mathbf{\Lambda}_\Phi\mathbf{U}_\Phi^*$$

then, in terms of the singular value decomposition (SVD), the optimal precoder matrix is

$$\mathbf{F} = \frac{1}{\sqrt{\eta_0}}\mathbf{U}_\Phi\mathbf{\Lambda}_\Phi^{\frac{1}{2}}\mathbf{V}^*. \quad (29)$$

The left singular vectors and the singular values of an optimal precoder are the eigenvectors and the scaled square roots of the eigenvalues of $\mathbf{\Phi}$, respectively. The right singular vectors $\mathbf{V}$, however, can be any unitary matrix, attributed to the codeword distance product matrix $\mathbf{A}$ being scaled-identity. For simplicity, we can choose $\mathbf{V} = \mathbf{I}$ for a precoder with an OSTBC.

D. Summary

In this section, by solving a convex optimization problem with matrix variables, we have established an analytical design algorithm for a precoder with an orthogonal STBC. The algorithm resembles the water-filling process. At first, it is assumed that all precoder eigenmodes are used, and power allocation is performed on all modes. If any mode has negative power, the most negative mode is dropped, and the power is reallocated accordingly. During this process, the precoder eigenbeam directions also evolve with the water-filling iterations. Each iteration essentially aims at finding a Lagrange multiplier solution, using an efficient binary search. The algorithm produces the optimal beam directions, which are the precoder left singular vectors, and the optimal beam power allocation, which gives the singular values. The precoder right singular vectors, however, can be arbitrary, due to the isotropic property of the orthogonal STBC, and are usually omitted.

V. Precoder Design With a General STBC

We now return to the original optimization problem (15) and examine the case that the codeword distance product matrix $\mathbf{A}$ (7) results from any STBC. For a nonorthogonal STBC, $\mathbf{A}$ is not always a scaled-identity matrix. Since the scaled-identity $\mathbf{A}$ case was solved in Section IV, we focus on solving for the nonidentity $\mathbf{A}$ next, then discuss the precoder solution with a general STBC in Section V-B.

A. Problem Analysis for a Nonidentity Codeword Distance Product

When $\mathbf{A}$ is not a scaled-identity matrix, solving the optimization problem (15) is more difficult. In particular, due to the nonconvexity in this case, it is not obvious if the problem can be solved exactly. In this section, we will analyze and reformulate the problem, then apply different relaxations to obtain the precoder analytically.

To analyze (15) for a nonidentity $\mathbf{A}$, consider the following more constrained problem:

$$\begin{aligned} \min_{\mathbf{F}} \; & J = \operatorname{tr}\left(\mathbf{H}_m \mathbf{W}^{-1} \mathbf{H}_m^*\right) - M \log\det(\mathbf{W}) \\ \text{subject to } & \mathbf{W} = \frac{\rho}{4}\mathbf{R}_t \mathbf{F}\mathbf{A}\mathbf{F}^* \mathbf{R}_t + \mathbf{R}_t \\ & \operatorname{tr}(\mathbf{F}\mathbf{F}^*) = 1 \\ & \operatorname{tr}(\mathbf{F}\mathbf{A}\mathbf{F}^*) = \gamma \end{aligned} \quad (30)$$

where γ is a positive constant. Since this problem is more constrained than the original problem (15), the optimal J value of this problem will be larger than, or equal to, the optimal J value in (15). The smallest optimal J value in (30) across different γ values, however, will equal the optimal J value in (15), which is the point at which the two problems become equivalent.

Now consider problem (30) with different γ values. In this problem, the condition $\operatorname{tr}(\mathbf{F}\mathbf{A}\mathbf{F}^*) = \gamma$ acts as an additional transmit power constraint. Obviously, with more transmit power, the error probability, or the objective function J equivalently, will be smaller. Thus, we are interested in this problem with the largest feasible γ; that is, the largest γ value such that there exists an $\mathbf{F}$ satisfying both equality constraints $\operatorname{tr}(\mathbf{F}\mathbf{F}^*) = 1$ and $\operatorname{tr}(\mathbf{F}\mathbf{A}\mathbf{F}^*) = \gamma$.

Applying the matrix inequality $\operatorname{tr}(AB) \le \sum_i \lambda_i(A)\lambda_i(B)$ for PSD matrices [35], we have

$$\gamma = \operatorname{tr}(\mathbf{F}\mathbf{A}\mathbf{F}^*) = \operatorname{tr}(\mathbf{F}^*\mathbf{F}\mathbf{A}) \le \sum_i \lambda_i(\mathbf{F}^*\mathbf{F})\lambda_i(\mathbf{A}). \quad (31)$$

The equality occurs when the eigenvectors of $\mathbf{F}^*\mathbf{F}$ are the same as those of $\mathbf{A}$. Let the singular value decomposition of the precoder be $\mathbf{F} = \mathbf{U}\mathbf{D}\mathbf{V}^*$, this equality condition means

$$\mathbf{V} = \mathbf{U}_{\mathbf{A}} \quad (32)$$

where $\mathbf{A} = \mathbf{U}_{\mathbf{A}}\mathbf{\Lambda}_{\mathbf{A}}\mathbf{U}_{\mathbf{A}}^*$ is the eigenvalue decomposition of $\mathbf{A}$. Since the eigenvalues of $\mathbf{F}^*\mathbf{F}$ are the same as those of $\mathbf{F}\mathbf{F}^*$, the equality condition translates to $\lambda_i(\mathbf{F}\mathbf{A}\mathbf{F}^*) = \lambda_i(\mathbf{F}\mathbf{F}^*)\lambda_i(\mathbf{A})$. Condition (32) ensures the largest value for γ, without imposing any constraint on the eigenvalues of $\mathbf{F}\mathbf{A}\mathbf{F}^*$ or $\mathbf{F}\mathbf{F}^*$, except their orders relative to $\lambda_i(\mathbf{A})$. With this condition, problem (30) and the original problem (15) become equivalent when γ is chosen to be the same as that resulted from the optimal solution of (15): $\gamma = \sum_i \lambda_i(\mathbf{A})\lambda_i^\star(\mathbf{F}\mathbf{F}^*)$.

Based on these arguments, condition (32) is optimal for the original problem (15). This problem is then equivalent to the following problem:

$$\begin{aligned} \min_{\mathbf{B}} \quad & J = \operatorname{tr}\left(\mathbf{H}_m \mathbf{W}^{-1} \mathbf{H}_m^*\right) - M \log\det(\mathbf{W}) \\ \text{subject to} \quad & \mathbf{W} = \frac{\rho}{4}\mathbf{R}_t \mathbf{B} \mathbf{R}_t + \mathbf{R}_t \\ & \sum_i \xi_i \lambda_i(\mathbf{B}) = 1 \\ & \mathbf{B} \succcurlyeq 0 \end{aligned} \quad (33)$$

where $\mathbf{B} = \mathbf{F}\mathbf{A}\mathbf{F}^*$, and $\xi_i = [\lambda_i(\mathbf{A})]^{-1}$ are the inverses of the nonzero eigenvalues of $\mathbf{A}$.

By finding the optimal right singular vectors of $\mathbf{F}$ in (32), we have reformulated the original problem (15) into a new problem in terms of the variable $\mathbf{B}$. The new formulation (33), however, is not convex in $\mathbf{B}$, due to the nonlinear equality constraint involving the eigenvalues of $\mathbf{B}$. We proceed by applying some relaxation to this constraint to obtain an analytical precoder solution. Due to relaxation, the solution in this case may not be optimal for the original problem. Two different relaxations follow.

1) Minimum Eigenvalue Relaxation Method: Employing the inequality $\sum_i \xi_i \lambda_i(\mathbf{B}) \le \xi_{\max}\operatorname{tr}(\mathbf{B})$, we relax the problem by replacing all ξ_i with $\xi_{\max}$, which is equivalent to approximating $\mathbf{A}$ in (15) with an identity matrix, scaled by the minimum nonzero eigenvalue of $\mathbf{A}$. This approximation effectively produces a smaller $\mathbf{W}$ (in the positive semidefinite sense), hence loosening the upper bound on the PEP in (12). We then obtain the same problem formulation as in the orthogonal STBC case (16), where the value μ_0 for the minimum-distance design is [24]

$$\mu_0 = \min_{\Delta_{ij}} \lambda_{\min}\left(\Delta_{ij}\Delta_{ij}^*\right)$$

and for the average-distance design is $\mu_0 = \lambda_{\min}(\bar{\mathbf{A}})$ (with $\lambda_{\min} \ne 0$). This relaxation method works well if the condition number of $\mathbf{A}$ is reasonably small.

2) Trace Relaxation Method: Another relaxation method is to replace the equality constraint on the eigenvalues of $\mathbf{B}$ with the linear constraint $\mathrm{tr}(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B}) = 1$, making the relaxed problem convex. Since $\mathrm{tr}(AB) \geq \sum_i \lambda_{N-i+1}(A)\lambda_i(B)$ for ordered eigenvalues [34], this relaxation method will result in a precoder with $\mathrm{tr}(\mathbf{FF}^*) \leq 1$, meaning that the total transmit power may be less than the original constraint (4). A scaling factor can then be applied to the precoder solution of the relaxed problem to increase the power to meet the original constraint.

In order to solve this relaxed problem, reformulate the problem in terms of $\mathbf{W}$ as

$$\begin{aligned} \min_{\mathbf{W}} \quad & \mathrm{tr}\left(\mathbf{H}_m \mathbf{W}^{-1}\mathbf{H}_m^*\right) - M\log\det(\mathbf{W}) \\ \text{subject to} \quad & \mathrm{tr}\left(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\left(\mathbf{R}_t^{-1}\mathbf{W}\mathbf{R}_t^{-1} - \mathbf{R}_t^{-1}\right)\right) = \frac{\rho}{4} \\ & \mathbf{R}_t^{-1}\mathbf{W}\mathbf{R}_t^{-1} - \mathbf{R}_t^{-1} \succcurlyeq 0. \end{aligned} \quad (34)$$

This problem is similar to (17), but with a more general trace constraint; it can be solved using a similar approach. The steps for solving this problem and the solutions for $\mathbf{W}$ and $\mathbf{B}$ are given in the Appendix.

B. Precoder Solution With a General STBC

Returning to the general STBC case, if the codeword distance product matrix $\mathbf{A}$ is a scaled-identify matrix, then the precoder solution is similar to that of the orthogonal STBC case (29). For a nonidentity $\mathbf{A}$, employ one of the relaxation methods outlined above and solve for $\mathbf{B}$. Perform the eigenvalue decomposition of this matrix as $\mathbf{B} = \mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}_{\mathbf{B}}\mathbf{U}_{\mathbf{B}}^*$, then the precoder solution is

$$\mathbf{F} = \mathbf{U}_{\mathbf{B}}\mathbf{\Lambda}_{\mathbf{B}}^{\frac{1}{2}}\mathbf{\Lambda}_{\mathbf{A}}^{-\frac{1}{2}}\mathbf{U}_{\mathbf{A}}^*. \quad (35)$$

For an orthogonal STBC, since $\mathbf{A}$ is always a scaled-identity matrix, $\mathbf{U}_{\mathbf{A}}$ is an arbitrary unitary matrix and, hence, can be omitted. For a nonorthogonal STBC, $\mathbf{U}_{\mathbf{A}}$ depends on the STBC structure. Specifically, when $\mathbf{A}$ is not a scaled-identity matrix, the input signal produced by the STBC has a certain codeword-error shape with directions $\mathbf{U}_{\mathbf{A}}$ and power loadings $\lambda_i(\mathbf{A})$. The precoder then matches both its input signal structures and the channel. It effectively remaps the input signal directions from $\mathbf{U}_{\mathbf{A}}$ into $\mathbf{U}_{\mathbf{B}}$ and redistributes the transmit power according to the CSIT to optimally match the channel. Note that the precoder beam directions (the left singular vectors $\mathbf{U}_{\mathbf{B}}$) depend only on the CSIT; the input shaping matrix (the right singular vectors $\mathbf{U}_{\mathbf{A}}$) depends only on the precoder input signal—the STBC structure; and the power allocation now depends on both sides.

Recall that the matrix $\mathbf{A}$ can be chosen based on the minimum- or average-distance criterion. Between these two criteria, the average-distance criterion usually produces a scaled-identity matrix $\bar{\mathbf{A}}$, assuming equiprobability and independence among all input symbols. The reason is that STBCs commonly assume no CSIT and distribute power equally among all antennas and all symbols, leading to a white error covariance. The minimum-distance criterion, on the other hand, can often produce a nonidentity $\mathbf{A}$.

C. Summary

We have extended the precoder design algorithm in Section IV-B to cover precoding with a nonorthogonal STBC. In contrast to the isotropic property of an orthogonal STBC, a nonorthogonal STBC may preshape the input signal with a nonidentity codeword distance product matrix $\mathbf{A}$. In that case, the precoder optimal right singular vectors are given by the eigenvectors of $\mathbf{A}$. For a nonidentity $\mathbf{A}$, we use relaxations to find the precoder left singular vectors and the singular values. In general, the precoder solution contains matching singular vectors on each side to the STBC structure and to the channel, respectively, while the singular values depend on both the STBC and the CSIT. The precoder essentially remaps the spatial directions of the input signal to match those of the channel, based on the CSIT, and allocates transmit power accordingly in a water-filling fashion.

VI. Asymptotic Analysis of the Precoder Results

A. Effect of a High K Factor on the Precoder

In this section, we investigate the effect on the optimal precoder as the channel K factor increases to infinity. An infinite K can correspond to a nonfading channel, or to perfect instantaneous CSIT. In either case, it is useful to study this limit, so that applicable scenarios can be identified.

When K approaches infinity, the objective function (14) is invalid since it also approaches infinity; hence, we need to analyze the full upper-bound (12). Using the channel model with the K factor in (1), we rewrite this bound as

$$\bar{P}_e \leq \frac{\exp\left[\mathrm{tr}\left(K\mathbf{H}_0\mathbf{W}_0^{-1}\mathbf{H}_0^*\right)\right]}{\det(\mathbf{W}_0)^M}\det(\mathbf{R}_0)^M \times \exp\left[-\mathrm{tr}\left(K\mathbf{H}_0\mathbf{R}_0^{-1}\mathbf{H}_0^*\right)\right], \quad (36)$$

where

$$\mathbf{W}_0 = \frac{\rho}{4(K+1)}\mathbf{R}_0\mathbf{FAF}^*\mathbf{R}_0 + \mathbf{R}_0.$$

Express $\mathbf{W}_0$ in the form $\mathbf{W}_0 = \mathbf{R}_0^{1/2}(\mathbf{Q}+\mathbf{I})\mathbf{R}_0^{1/2}$, where

$$\mathbf{Q} = \frac{\rho}{4(K+1)}\mathbf{R}_0^{\frac{1}{2}}\mathbf{FAF}^*\mathbf{R}_0^{\frac{1}{2}}.$$

With sufficiently large K, the largest eigenvalue of the PSD Hermitian matrix $\mathbf{Q}$ will be less than 1, and the following expansion [36] can be applied:

$$\mathbf{W}_0^{-1} = \mathbf{R}_0^{-\frac{1}{2}}(\mathbf{I} - \mathbf{Q} + \mathbf{Q}^2 - \mathbf{Q}^3 + \ldots)\mathbf{R}_0^{-\frac{1}{2}}.$$

Replacing this expansion into the upper-bound (36), and noting that

$$K\mathbf{W}_0^{-1} - K\mathbf{R}_0^{-1} \to -\frac{\rho}{4}\mathbf{FAF}^* \quad \text{as} \quad K \to \infty$$

the limiting upper bound on the average PEP is

$$P_{\mathrm{K\ limit}} = \exp\left[-\mathrm{tr}\left(\frac{\rho}{4}\mathbf{H}_0\mathbf{FAF}^*\mathbf{H}_0^*\right)\right].$$

Minimizing $P_{\mathrm{K\ limit}}$ is equivalent to maximizing the trace expression. We apply the inequality $\mathrm{tr}(ABC) \leq$

$\sum_i \sigma_i(A)\sigma_i(B)\sigma_i(C)$ [35] with $\sigma_i(.)$ as sorted singular values, where the equality is achieved when the right singular vectors of A (and B) align with the left singular vectors of B (and C, respectively). Taking the power constraint (4) into account, the optimal precoder $\mathbf{F}$ has the form [24]

$$\mathbf{F} = \mathbf{u}\mathbf{v}^* \tag{37}$$

where $\mathbf{u}$ is the dominant eigenvector of $\mathbf{H}_0^*\mathbf{H}_0$ and $\mathbf{v}$ is the dominant eigenvector of $\mathbf{A}$. In other words, the optimal precoder, in the limit of an infinite K factor, is a single-mode beamformer, matching the dominant right singular vector of the channel mean. This result applies regardless of the STBC or the choice of the minimum- or average-distance design, as the left singular vectors of $\mathbf{F}$ are independent of $\mathbf{A}$.

Note that for a multiple-input single-output (MISO) system at the infinite K factor (or perfect channel knowledge), single-mode beamforming is also optimal for achieving the channel capacity [7]. Therefore, the proposed precoding solution is optimal for a MISO system in the PEP criterion and asymptotically optimal in capacity. For a MIMO system, however, the optimal solution in the limit of infinite K differs between the two criteria: PEP and capacity. Whereas the capacity solution calls for water-filling over the eigenmodes of the channel, the PEP solution places all transmit power on the dominant mode. Thus, for a MIMO channel with a high K factor, a precoder based on the PEP criterion is suitable for use along with a STBC with a rate of one or less. However, as K increases, the role of the STBC becomes less important in obtaining diversity. Moreover, the power loading on the precoder dominant mode increases as K increases; therefore, precoding can be combined with adaptive modulation and coding to take the full advantage of high K. At a low K factor, due to the multimode beamforming effect, a higher rate STBC can also be used with the precoder in a MIMO system.

B. High-SNR Precoder Analysis

To analyze the effect of the SNR ρ on the precoder, we examine the objective function (14), noting that ρ affects $\mathbf{W}$ as given in (13). When $\rho \to \infty$ then $\mathbf{W}^{-1} \to \mathbf{0}$; thus in the limit, the objective function becomes

$$J_{\text{SNR limit}} = -M \log\det\left(\frac{\rho}{4}\mathbf{R}_t\mathbf{F}\mathbf{A}\mathbf{F}^*\mathbf{R}_t + \mathbf{R}_t\right). \tag{38}$$

At high SNRs, the precoder asymptotically depends only on the correlation $\mathbf{R}_t$. The precoder design problem is now equivalent to maximizing $\log\det((\rho/4)\mathbf{F}\mathbf{A}\mathbf{F}^* + \mathbf{R}_t^{-1})$, subject to the power constraint (4). If $\mathbf{A}$ is a scaled-identity matrix, this is the standard water-filling problem [33]. For the nonidentity $\mathbf{A}$, using the same analysis in Section V-A for the nonasymptotic case, the optimal precoder right singular vectors are given by the eigenvectors of $\mathbf{A}$; that is, $\mathbf{V} = \mathbf{U}_{\mathbf{A}}$. Again, denoting $\mathbf{B} = \mathbf{F}\mathbf{A}\mathbf{F}^*$, the problem can be recast as follows:

$$\max_{\mathbf{B}} \; J_2 = \log\det\left(\frac{\rho}{4}\mathbf{B} + \mathbf{R}_t^{-1}\right) \quad \text{subject to} \; \sum_i \xi_i\lambda_i(\mathbf{B}) = 1 \tag{39}$$

where $\xi_i = 1/\lambda_i(\mathbf{A})$ for nonzero eigenvalues. Let $\mathbf{R}_t = \mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^*$ be the eigenvalue decomposition of $\mathbf{R}_t$, then $J_2 = \log\det((\rho/4)\mathbf{U}_t^*\mathbf{B}\mathbf{U}_t + \mathbf{\Lambda}_t^{-1})$. Noting that the eigenvalues of $\mathbf{U}_t^*\mathbf{B}\mathbf{U}_t$ are the same as those of $\mathbf{B}$, the transformation does not affect the constraint in (39). Thus, by the Hadamard inequality [34], J_2 is maximum when $\mathbf{U}_t^*\mathbf{B}\mathbf{U}_t$ is diagonal. This maximum implies that the optimal left singular vectors of $\mathbf{F}$ are the eigenvectors of $\mathbf{R}_t$. The optimization problem now becomes

$$\max_{\lambda_i(\mathbf{B})} \sum_i \log\left(\frac{\rho}{4}\lambda_i(\mathbf{B}) + \lambda_i\left(\mathbf{R}_t^{-1}\right)\right) \quad \text{subject to} \; \sum_i \xi_i\lambda_i(\mathbf{B}) = 1.$$

This problem is convex and can be solved exactly, using the standard Lagrange multiplier technique [32], to obtain the solution

$$\lambda_i(\mathbf{B}) = \left(\frac{\lambda_i(\mathbf{A})}{\kappa} - \frac{4\lambda_i\left(\mathbf{R}_t^{-1}\right)}{\rho}\right)_+ \tag{40}$$

where κ is chosen to satisfy the equality constraint $\sum_i \xi_i\lambda_i(\mathbf{B}) = 1$. The plus notation means that the expression takes the value inside the parenthesis if this value is positive, otherwise it is zero.

Thus for the high-SNR limit, the optimal precoder has the form

$$\mathbf{F} = \mathbf{U}_t\mathbf{D}\mathbf{U}_{\mathbf{A}}^* \tag{41}$$

where the singular value matrix $\mathbf{D}$ is obtained from the water-filling solution (40) with diagonal elements as $d_i = (\lambda_i(\mathbf{B})/\lambda_i(\mathbf{A}))^{1/2}$. As SNR $\to \infty$, this solution approaches equal power distribution on all the eigenmodes of $\mathbf{R}_t$ that correspond to the nonzero eigenmodes of $\mathbf{A}$.

C. Discussion

Comparing the two limiting cases of the infinite K (37) and the infinite SNR (41), we see that in each case, the precoder $\mathbf{F}$ converges to a solution that depends on only one of the two statistical channel parameters: either the channel mean $\mathbf{H}_m$ or the transmit correlation $\mathbf{R}_t$. When the channel becomes more static, indicated by a high K factor, the channel mean dominates the precoder solution. At a high SNR, however, the fluctuation in the channel becomes more pronounced; hence, the transmit correlation, equivalent to the channel covariance, tends to dominate the precoder solution. A parallel observation is that, as K increases, the precoder tends to drop modes until it becomes a single-mode beamformer, with an asymptotic direction as the dominant right singular vector of the channel mean. On the other hand, as the SNR increases, the precoder tends toward full-mode beamforming on all eigenvectors of the transmit correlation matrix with equal power allocation. These effects are observed, provided that all other parameters are kept constant when the variable of interest, the K factor or the SNR, varies. If both the K factor and the SNR increase, then there exists a K factor threshold for single-mode beamforming that increases with the SNR. An example of this threshold is given in the numerical simulation Section VIII.

VII. Special Scenarios of the Precoder Design

In this section, we analyze the precoder designs for two special scenarios: when the channel has zero mean but is correlated, and when the channel is uncorrelated with a nonzero mean. Precoding for these two scenarios was previously studied in [20] with a general STBC, and in [18] with an OSTBC, respectively. We will analyze our results and show that both cases are covered. We also extend the latter case, precoding for an uncorrelated nonzero mean channel, to the general STBC and obtain an optimal precoder solution without relaxation.

A. Precoding for Correlated Channels With Zero Mean

Consider a correlated channel with zero mean ($\mathbf{H}_m = 0$), corresponding to a correlated Rayleigh fading channel. When the codeword distance product matrix $\mathbf{A}$ is a scaled-identity matrix, from (22), we have $\mathbf{\Psi} = M^2\mathbf{I}$. Thus, the power constraint equation (23) becomes a standard water-filling problem, as follows:

$$\mathrm{tr}\left(\frac{M}{\nu}\mathbf{I}_N - \mathbf{R}_t^{-1}\right) = \eta_0$$

provided the matrix expression inside the parentheses is positive semidefinite. The optimal precoder $\mathbf{F}$ (28) has its beam directions given by the eigenvectors of $\mathbf{R}_t$ and the power allocation obtained via standard water-filling [33] on the eigenvalues of $\mathbf{R}_t$. No numerical binary search is required for ν in this case, and solving for the optimal precoder is simple.

With a nonidentity $\mathbf{A}$, from the original problem (15), we can obtain a formulation similar to the asymptotic high-SNR case (38). The result also reduces to $\mathbf{F}$ having the left singular vector given by the eigenvectors of $\mathbf{R}_t$, the right singular vectors given by the eigenvectors of $\mathbf{A}$, and the power allocation obtained by closed-form water-filling (40), without the necessity of a binary search. This result agrees with the solution established in [20], which was proven using Hadamard and geometric-arithmetic mean inequalities.

B. Precoding for Uncorrelated Channels With a Nonzero Mean

We now consider an uncorrelated channel ($\mathbf{R}_t = \mathbf{I}$) with a nonzero mean, which can correspond to an uncorrelated Rician channel, or to a channel estimate with uncorrelated error at the transmitter.

1) Scaled-Identity $\mathbf{A}$*:* When $\mathbf{A}$ is a scaled-identity matrix, we have $\mathbf{\Psi} = M^2\mathbf{I} + 4\nu\mathbf{H}_m^*\mathbf{H}_m$, and the power constraint (23) becomes a simpler water-filling problem as

$$\mathrm{tr}\left(\frac{1}{2\nu}\left(M^2\mathbf{I} + 4\nu\mathbf{H}_m^*\mathbf{H}_m\right)^{\frac{1}{2}} - \left(1 - \frac{M}{2\nu}\right)\mathbf{I}\right) = \eta_0.$$

The precoder solution has its left singular vectors given by the eigenvectors of $\mathbf{H}_m^*\mathbf{H}_m$—the right singular vectors of the channel mean. The power loadings can be found using the inner and outer algorithms, in which the computation is simpler in this case as the eigenvalues in the outer equation (26) can now be expressed explicitly in terms of ν. This result is similar to the solution established in [18] for precoding with an OSTBC.

2) Nonidentity $\mathbf{A}$*:* For a nonidentity $\mathbf{A}$, we revisit the original problem formulation (15) to obtain

$$\begin{aligned}\min_{\mathbf{F}} \quad J &= \mathrm{tr}\left(\mathbf{H}_m\mathbf{W}^{-1}\mathbf{H}_m^*\right) - M\log\det(\mathbf{W})\\ \text{subject to} \quad \mathbf{W} &= \frac{\rho}{4}\mathbf{F}\mathbf{A}\mathbf{F}^* + \mathbf{I}\\ \mathrm{tr}(\mathbf{F}\mathbf{F}^*) &= 1.\end{aligned}$$

To solve this problem, we note that the $\log\det$ term in the objective function J depends only on $\mathbf{F}$ and $\mathbf{A}$. Apply the inequality $\det(A+B) \le \prod_i \lambda_i(A)\lambda_i(B)$ [35], then without modifying the eigenvalues of $\mathbf{F}$ or $\mathbf{A}$, the expression $\log\det(\mathbf{W}) = \log\det(\mathbf{A}) + \log\det((\rho/4)\mathbf{F}^*\mathbf{F} + \mathbf{A}^{-1})$ is maximized when the right singular vectors of $\mathbf{F}$ are the same as the eigenvectors of $\mathbf{A}$. Now, examine the trace term in J. Applying the inequality $\mathrm{tr}(AB) \ge \sum \lambda_i(A)\lambda_{n-i+1}(B)$ [34], then $\mathrm{tr}(\mathbf{H}_m\mathbf{W}^{-1}\mathbf{H}_m^*)$ is minimized when $\mathbf{W}$ has the same eigenvectors as those of $\mathbf{H}_m^*\mathbf{H}_m$. Noting that both inequalities place no constraints on the eigenvalues, except for their relative order, both the equality conditions can be simultaneously achieved by forcing the SVD of $\mathbf{F}$ to be

$$\mathbf{F} = \mathbf{U}_m\mathbf{D}\mathbf{U}_{\mathbf{A}}^*$$

where $\mathbf{H}_m^*\mathbf{H}_m = \mathbf{U}_m\mathbf{\Lambda}_m\mathbf{U}_m^*$ is the eigenvalue decomposition of $\mathbf{H}_m^*\mathbf{H}_m$. We now merely need to find the singular values of $\mathbf{F}$. To do this, let $\mathbf{B} = \mathbf{F}\mathbf{A}\mathbf{F}^*$, and the problem can be reformulated in terms of $\lambda_i(\mathbf{B})$ as

$$\min_{\lambda_i(\mathbf{B})} \quad \sum_i \left(1 + \frac{\rho}{4}\lambda_i(\mathbf{B})\right)^{-1}\lambda_{m,i} - M\sum_i \log\left(1 + \frac{\rho}{4}\lambda_i(\mathbf{B})\right) \quad \text{subject to} \quad \sum_i \xi_i\lambda_i(\mathbf{B}) = 1$$

where $\xi_i = 1/\lambda_i(\mathbf{A})$, and $\lambda_{m,i}$ are the eigenvalues of $\mathbf{H}_m^*\mathbf{H}_m$. This problem is convex and can be solved using the standard Lagrange multiplier technique to obtain

$$\lambda_i(\mathbf{B}) = \left[\frac{\lambda_i(\mathbf{A})}{2\nu}\left(M + \sqrt{M^2 + 16\nu\frac{\lambda_{m,i}}{\rho\lambda_i(\mathbf{A})}}\right) - \frac{4}{\rho}\right]_+$$

where ν is the Lagrange multiplier satisfying the constraint $\sum_i \xi_i\lambda_i(\mathbf{B}) = 1$. ν can be found using an 1-D binary numerical search similar to the inner and outer algorithms. The singular values of $\mathbf{F}$ are then $d_i = (\lambda_i(\mathbf{B})/\lambda_i(\mathbf{A}))^{1/2}$. Thus, for this CSIT scenario, we can also obtain a closed-form optimal precoder solution for all STBCs.

C. Discussion

In both special cases when only the channel mean or the transmit correlation is present, the precoder solution is significantly simplified. The precoder beamforming directions are fixed and are given by the singular vectors or eigenvectors of the present channel parameter, and the power loadings are obtained by some form of water filling over its eigenvalues. In these cases, closed-form optimal precoder solutions are available for all STBCs, and no relaxation is necessary.

VIII. Design Examples and Numerical Results

We now present design examples and simulation results for several system configurations. We perform simulations using two antenna setups: 2×1 and 4×1. The channel mean and transmit correlation matrices are generated arbitrarily and normalized according to (2), where the correlation matrix is Hermitian and positive definite. These matrices are given in the Appendix. The correlation matrix for the 2×1 channel has the eigenvalues of [1.96, 0.04] and the condition number of 47.93, representing a strong correlation. The correlation matrix for the 4×1 channel has the eigenvalues of [2.05, 1.48, 0.41, 0.06] and the condition number of 34.2. In all systems, we choose the K factor to be 0.1, except when studying the effect of K on the precoder.

A. Precoders With an Orthogonal STBC

1) Precoder Design Examples With the Alamouti STBC: We first show a precoder design example in a 2×1 system using the Alamouti STBC [1]. The code is given as

$$\mathbf{C} = \begin{pmatrix} c_1 & -c_2^* \\ c_2 & c_1^* \end{pmatrix}.$$

We need to identify the value for μ_0 in (16). For the minimum-distance design, $\mathbf{A} = (d^2/P)\mathbf{I}$, where d is the minimum distance in the signal constellation. Consider a square QAM constellation with M_c points for example, the average symbol power is $P = d^2(M_c - 1)/6$, thus $\mu_0 = 6/(M_c - 1)$.

For the average-distance design, assuming that each codeword contains n distinct symbols that are independent and equally likely, we can rewrite (10) as

$$\begin{aligned}\bar{\mathbf{A}} &= \frac{1}{TP} E_{\mathbf{C} \neq \hat{\mathbf{C}}} \left[(\mathbf{C} - \hat{\mathbf{C}})(\mathbf{C} - \hat{\mathbf{C}})^* \right] \\ &= \frac{1}{TP} \frac{M_c^{2n}}{M_c^{2n} - M_c^n} 2 \left(E[\mathbf{C}\mathbf{C}^*] - E[\mathbf{C}\hat{\mathbf{C}}^*] \right) \end{aligned} \quad (42)$$

where the ratio factor results from averaging over only pairs of distinct codewords. This expression applies to any signal constellation of size M_c. Since $\mathbf{C}$ and $\hat{\mathbf{C}}$ are chosen independently, $E[\mathbf{C}\hat{\mathbf{C}}^*] = 0$. For a STBC that gives equal weight to all symbols, it is plausible that $E[\mathbf{C}\mathbf{C}^*] = \mu P \mathbf{I}$ for some μ. Hence, the average-distance criterion usually produces a scaled-identity matrix $\bar{\mathbf{A}}$. For example, the Alamouti code with quaternary phase shift keying (QPSK) has $\bar{\mathbf{A}} = (64/15)\mathbf{I}$, or $\mu_0 = 4.26$.

2) Numerical Results: Fig. 2 shows the performance of a 2×1 system using the Alamouti code and QPSK modulation. For this system, the minimum-distance and the average-distance precoder designs perform exactly the same. The precoding gain is around 2.2 dB at low and medium SNRs and diminishes at higher SNRs. When the SNR increases, as discussed in Section VI-B, the precoder becomes increasingly dependent on the transmit correlation and approaches equi-power on the nonzero eigenmodes of $\mathbf{A}$. Thus, the value of water-filling power allocation decreases at a high SNR, leading to the diminishing precoding gain with the OSTBC. A similar effect was observed from a mutual information analysis for a full-rank correlation CSIT in [10].

Fig. 3 shows the performance for the same 2×1 system with different QAM orders. The precoding gain of the minimum-distance design is consistent over different constellation sizes, whereas the average-distance precoding gain is reduced with a larger constellation at a high SNR. With a larger constellation, the number of minimum-distance codeword pairs becomes larger, while at a higher SNR, the minimum-distance pairs become more dominant in affecting the error probability. Therefore, the minimum-distance precoder design results in more gain than the average-distance design in this domain.

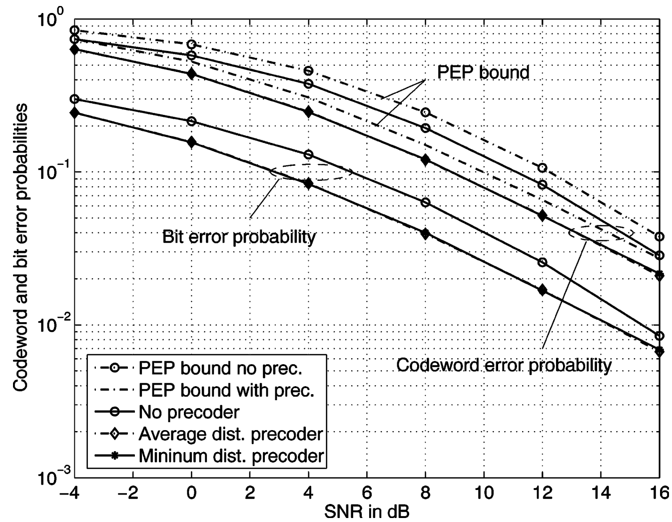

Fig. 2. Performance of a 2×1 system using the Alamouti code and QPSK modulation, with and without precoding.

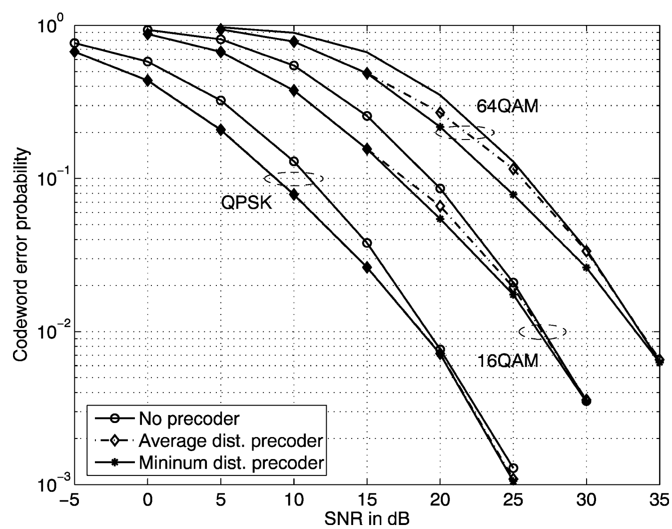

Fig. 3. Performance of a 2×1 system using the Alamouti code with various QAMs, with and without precoding.

B. Precoders With a Nonorthogonal STBC

We continue with an example of a linear precoder design with a nonorthogonal STBC. In general, our precoder design works with all STBCs, including high rate codes which achieve the diversity-multiplexing frontier such as [26], [27]. In this example, we use the quasi-orthogonal STBC (QSTBC) [4], [5] specifically. The name quasi-orthogonal results from the groups of columns of this code being orthogonal to each other, allowing simple ML decoding over pairs of symbols. The code itself is nonorthogonal and provides partial diversity; however, it achieves a higher rate than an OSTBC for more than two transmit antennas. We consider the following form of the QSTBC:

$$\mathbf{C} = \begin{pmatrix} c_1 & c_2 & c_3 & c_4 \\ -c_2^* & c_1^* & -c_4^* & c_3^* \\ c_3 & c_4 & c_1 & c_2 \\ -c_4^* & c_3^* & -c_2^* & c_1^* \end{pmatrix}.$$

In [22], a precoder exploiting channel mean feedback with the QSTBC is derived, using an asymptotic analysis.

1) The Codeword Distance Product Matrix for the QSTBC: For this code, the codeword distance product matrix $\mathbf{A}$ (7) has the form

$$\mathbf{A} = \frac{1}{P}\begin{pmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ b & 0 & a & 0 \\ 0 & b & 0 & a \end{pmatrix}$$

where $a = \sum_{i=1}^{4} |\Delta c_i|^2$ and $b = \Delta c_1 \Delta c_3^* + \Delta c_1^* \Delta c_3 + \Delta c_2 \Delta c_4^* + \Delta c_2^* \Delta c_4$, with $\Delta c_i = c_i - \hat{c}_i$, where $\hat{c}_i$ are symbols in $\hat{\mathbf{C}}$.

Assume that the symbols c_i are part of a constellation $\mathcal{C}$. For the minimum-distance design, $\mathbf{A}$ is given by the case in which there is only one symbol difference between the two codewords, thus $\mathbf{A}_{\min} = \min_{\mathcal{C}}(|\Delta c|^2)\mathbf{I}P$. For the average-distance design, assuming that all symbols c_i are independent and equally likely, and that $E[c_i] = 0$, $\bar{\mathbf{A}}$ (42) is also a scaled-identity matrix with $\mu_0 = 2M_c^8/(M_c^8 - M_c^4)$. In this QSTBC example, $\mathbf{A}$ is a scaled-identity matrix in both designs; hence, no relaxation is needed to solve for the precoder, and the precoder right singular vectors can be omitted.

Note that although the STBC diversity order, or the minimum rank of $\mathbf{A}$, is 2 in this case, the precoder is not limited to rank 2. At each time instance, the precoder acts as a beamformer on a separate column of the space–time code. Since there are 4 different symbols in each column, the precoder can form a maximum of four orthogonal beams, one per symbol, matching the statistically preferred directions in the channel. This beamforming effect causes the precoder rank to be independent of the STBC diversity order. Rather, it depends on the number of distinct symbols in each column of the code. Only when the SNR is not high enough, the precoder reduces its rank by dropping modes.

2) Numerical Results: Fig. 4 shows the performance curves for a 4 × 1 system using the QSTBC and QPSK modulation. The result reveals that both the minimum-distance and the average-distance precoder designs perform similarly for QPSK modulation, with a precoding gain of around 1.7–2 dB. Also shown is the performance of a precoder that has its rank limited to 2. This precoder gain reduces rapidly as the SNR increases; it eventually performs even worse than without precoding at very high SNRs. This example illustrates that the precoder rank should not depend on the STBC diversity order.

Fig. 5 shows similar performance curves for the 16 QAM constellation with Gray bit mapping. In this case, the minimum-distance precoder performs slightly better than the average-distance design, due to the larger constellation at a high-SNR effect, which is also observed in Fig. 3. The difference is small, however, at around 0.5 dB. With this larger constellation, the precoding gain is higher; the minimum-distance precoding gains are around 1.8–2.5 dB.

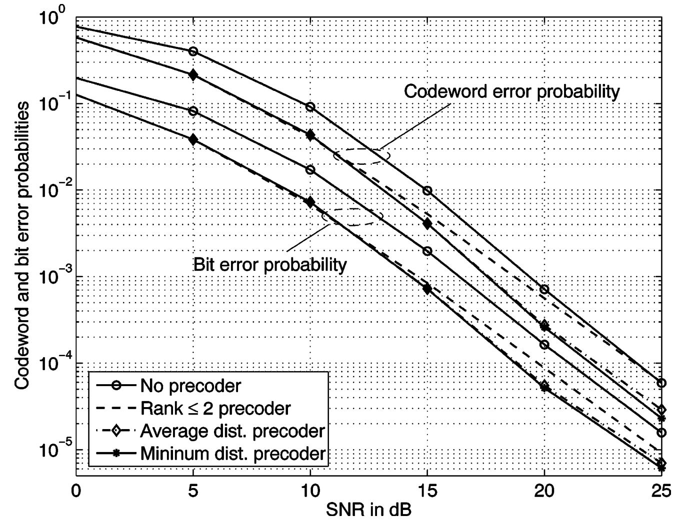

Fig. 4. Performance of a 4 × 1 system using the QSTBC and QPSK modulation, with and without precoding.

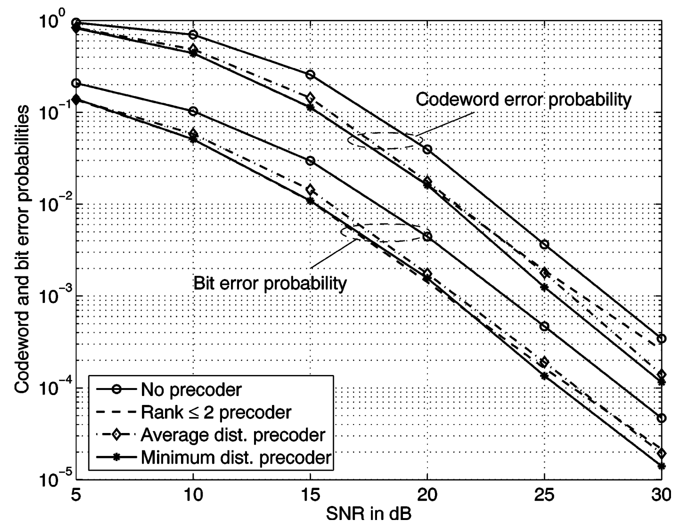

Fig. 5. Performance of a 4 × 1 system using the QSTBC and 16QAM, with and without precoding.

C. Effects of the K Factor and SNR

We now examine effects of the K factor and the SNR on the number of beamforming modes of the optimal precoder. From the asymptotic analyses in Section VI, the precoder approaches a single-mode beamformer as K increases. As the SNR increases, however, it tends to become a multimode beamformer. Hence, there exists a K factor threshold that increases with the SNR, above which the precoder is a single-mode beamformer.

Fig. 6 shows the regions of single- and multimode beamforming, as a function of the K factor and the SNR, for the 2 × 1 system using the Alamouti code and QPSK modulation. There are two noticeably different effects in this figure. In the high-SNR high-K factor domain, single-mode beamforming is because of high K alone. At lower SNRs, however, single-mode

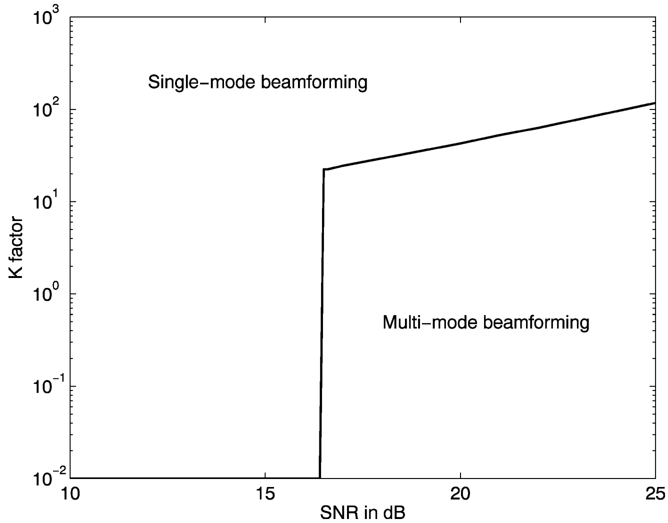

Fig. 6. Single-mode and multimode beamforming regions for a 2×1 system.

beamforming is also the result of the low SNR, when the precoder drops a mode due to insufficient power. The switching point between these two effects depends on the specific channel parameters (mean and correlation). Therefore, the precoder is a multimode beamformer only at a sufficiently high SNR with a sufficiently low K factor.

D. Discussion

In the performance simulations, we use channels with a very low K factor $(K = 0.1)$. The precoding gain will increase with a higher K factor, similar to having a better channel estimate at the transmitter. The number of receive antennas also affects the precoding gain; a channel with the same transmit correlation but with more receive antennas tends to possess a higher gain.

Note that the precoding gain, while dependent on the operating SNR, also depends on the channel parameters—the specific channel mean and transmit correlation matrices. Some channels favor precoding with larger gains, while for others, the precoding gain is less significant. A variety of factors can contribute to this effect: for example, the condition numbers of the mean and correlation matrices, how closely the eigenvectors of these matrices align. This effect is not studied here and can be a subject for further research.

The precoding gain is attributed to two factors: the optimal beam directions, which achieve an array gain, and the water-filling-type power allocation among these beams, which also results in an SNR advantage. Note that precoding on channel statistics obtains only antenna array gain, but not diversity gain. This lack of diversity gain is a property of statistical channel information, with which the precise directions of each channel realization are unknown to the transmitter. The STBC, therefore, plays an important role here in capturing the channel diversity. When the CSIT is perfect instantaneous information, however, the precoder can also deliver a diversity gain.

IX. Conclusion

We have designed a linear precoder to exploit a statistical CSIT—the channel mean and transmit antenna correlation—in a space–time coded MIMO system. We use convex optimization frameworks, based on minimizing the Chernoff bound on the PEP for minimum and average codeword distances, to derive analytical precoder results for both systems with orthogonal and with nonorthogonal STBCs. The precoder right singular vectors act as the input shaping matrix, matched to the eigenvectors of the ST codeword distance product matrix. The left singular vectors are the beam directions matched to the channel, based on the CSIT. The singular values represent the beam power allocation, which depends on both the STBC and the CSIT. We design a dynamic water-filling algorithm to establish the optimal beam directions and the optimal power allocation, both of which evolve with the water-filling iterations and are functions of the SNR. The algorithm involves simple and efficient binary searches to find the Lagrange multiplier. Asymptotic analyses of the precoding solutions reveal that the precoder depends primarily on the channel mean at high K factors, and on the transmit correlation at high SNRs. Simulation results confirm valuable precoding gain, which depends on the channel mean and correlation, the number of antennas, and the SNR, besides the codeword distance of choice. For a large signal constellation at a high SNR, the minimum-distance precoder obtains more gain than the average-distance design. The precoding gain with this statistical CSIT is an array gain, attributed to the optimal beam directions and the water-filling-type beam power allocation. The precoder, therefore, exploits the CSIT by spatially matching the transmit signal to the channel to achieve an SNR advantage.

Although the analysis in this paper has focused on the statistical CSIT, the precoder design can be applied to other CSIT formulations, such as a channel estimate with a known error covariance. In [25], we discuss this relationship and also formulate a framework combining both the channel statistics and potentially outdated channel measurements to create a robust CSIT.

Appendix

A. Derivation of the Bounds on ν in the Outer Algorithm

To derive the bounds in (27), apply Weyl's theorem [34] $\lambda_1(A) + \lambda_k(B) \le \lambda_k(A+B) \le \lambda_N(A) + \lambda_k(B)$ to the left-hand-side expression in (26), we obtain the following general function bounds when k modes $(1 \le k \le N-1)$ are dropped:

$$\begin{aligned} f_{\text{upper}} &= \frac{M(N-k)}{2\nu} + (N-k)\frac{\sqrt{M^2+4\nu\lambda_N}}{2\nu} \\ &\quad - \sum_{i=k+1}^{N} \frac{1}{\lambda_i(\mathbf{R}_t)} \\ f_{\text{lower}} &= \frac{M(N-k)}{2\nu} + (N-k)\frac{\sqrt{M^2+4\nu\lambda_1}}{2\nu} \\ &\quad - \sum_{i=k+1}^{N} \frac{1}{\lambda_i(\mathbf{R}_t)}. \end{aligned}$$

Equating each expression to η_0, we obtain a quadratic equation, from which the corresponding value for ν in (27) is derived.

B. Solving the Trace Relaxation Problem

To solve (34), again form the Lagrangian and differentiate with respect to $\mathbf{W}$ to obtain

$$-\mathbf{W}^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{W}^{-1} - M\mathbf{W}^{-1} + \nu\mathbf{R}_t^{-1}\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1} = 0$$

where ν is the Lagrange multiplier associated with the equality constraint. Let $\mathbf{G}$ be the Hermitian matrix such that $\mathbf{R}_t\mathbf{\Lambda}_{\mathbf{A}}\mathbf{R}_t = \mathbf{G}^2$, then the above quadratic matrix equation has the solution

$$\mathbf{W} = \frac{1}{2\nu}\mathbf{G}\left(M\mathbf{I}_N + \mathbf{\Psi}^{\frac{1}{2}}\right)\mathbf{G},$$

where

$$\mathbf{\Psi} = M^2\mathbf{I}_N + 4\nu\mathbf{G}^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{G}^{-1}.$$

Now $\mathbf{B}$ can be written as

$$\mathbf{B} = \frac{4}{\rho}\left[\frac{1}{2\nu}\mathbf{R}_t^{-1}\mathbf{G}\left(M\mathbf{I}_N + \mathbf{\Psi}(\nu)^{\frac{1}{2}}\right)\mathbf{G}\mathbf{R}_t^{-1} - \mathbf{R}_t^{-1}\right]$$

and the product $\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B}$ becomes

$$\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B} = \frac{4}{\rho}\left(\frac{M}{2\nu}\mathbf{I}_N + \frac{1}{2\nu}\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1}\mathbf{G}\mathbf{\Psi}^{\frac{1}{2}}\mathbf{G}\mathbf{R}_t^{-1} - \mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1}\right).$$

The Lagrange multiplier ν must satisfy the equality constraint in (34). Similar to the orthogonal STBC case, we solve for ν in two steps. We first assume that $\mathbf{B}$ is full rank and solve the equation $\operatorname{tr}(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B}) = 1$. If the resulting $\mathbf{B}$ is not PSD, then we drop the weakest mode of $\mathbf{B}$ and resolve for ν, using the equation $\sum_{i=2}^{N}\lambda_i(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B}) = 1$, and keep dropping modes until $\mathbf{B}$ is PSD. Note that in this case, we cannot write the eigenvalues of $\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B}$ explicitly as functions of ν to obtain an equation similar to (24). However, since the eigenvalue sum is monotonic in ν, similarly, ν can be found by performing an 1-D binary search in each step. The bounds on ν for the equation with k modes dropped ($0 \le k \le N-1$) are given by

$$\nu_{\text{bound}} = \frac{\alpha}{\beta^2} + \frac{M}{\beta} \tag{43}$$

where for the upper bound

$$\alpha = \lambda_{\max}\left(\mathbf{G}^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{G}^{-1}\right)$$
$$\beta = \lambda_{\min}\left(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1}\right) + \frac{\rho}{4(N-k)}$$

and for the lower bound

$$\alpha = \lambda_{\min}\left(\mathbf{G}^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{G}^{-1}\right)$$
$$\beta = \lambda_{\max}\left(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1}\right) + \frac{\rho}{4(N-k)}.$$

To derive the upper bound, we use the inequality $\sum_{k+1}^{N}\lambda_i(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B}) \le (N-k)\lambda_{\max}(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B})$ and further obtain an upper bound on $\lambda_{\max}(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B})$ as

$$\lambda_{\max}\left(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{B}\right) \le \frac{4}{\rho}\left[\frac{M}{2\nu} + \frac{1}{2\nu}\lambda_{\max}\left(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1}\mathbf{G}\mathbf{\Psi}^{\frac{1}{2}}\mathbf{G}\mathbf{R}_t^{-1}\right) - \lambda_{\min}\left(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1}\right)\right]$$

where we have used the inequality $\lambda_{\max}(A+B) \le \lambda_{\max}(A) + \lambda_{\max}(B)$ [35]. Now we have

$$\begin{aligned}
&\lambda_{\max}\left(\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1}\mathbf{G}\mathbf{\Psi}^{\frac{1}{2}}\mathbf{G}\mathbf{R}_t^{-1}\right)\\
&\overset{(a)}{=} \lambda_{\max}\left(\mathbf{G}\mathbf{R}_t^{-1}\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1}\mathbf{G}\mathbf{\Psi}^{\frac{1}{2}}\right)\\
&\overset{(b)}{\le} \lambda_{\max}\left(\mathbf{G}\mathbf{R}_t^{-1}\mathbf{\Lambda}_{\mathbf{A}}^{-1}\mathbf{R}_t^{-1}\mathbf{G}\right)\lambda_{\max}\left(\mathbf{\Psi}^{\frac{1}{2}}\right)\\
&\overset{(c)}{=} \lambda_{\max}\left(\mathbf{\Lambda}_{\mathbf{A}}^{-\frac{1}{2}}\mathbf{R}_t^{-1}\mathbf{G}^2\mathbf{R}_t^{-1}\mathbf{\Lambda}_{\mathbf{A}}^{-\frac{1}{2}}\right)\lambda_{\max}\left(\mathbf{\Psi}^{\frac{1}{2}}\right)\\
&\overset{(d)}{=} \lambda_{\max}\left(\mathbf{\Psi}^{\frac{1}{2}}\right)\\
&\overset{(e)}{=} \sqrt{M^2 + 4\nu\lambda_{\max}\left(\mathbf{G}^{-1}\mathbf{H}_m^*\mathbf{H}_m\mathbf{G}^{-1}\right)},
\end{aligned}$$

where (a) and (c) follow from $\lambda(AB) = \lambda(BA)$ [34], (b) follows from $\lambda_{\max}(AB) \le \lambda_{\max}(A)\lambda_{\max}(B)$ [35], (d) follows from the definition for $\mathbf{G}$, and (e) follows from $\lambda(\mathbf{I}+A) = 1 + \lambda(A)$. The upper bound (43) on ν can then be found by solving

$$\frac{M}{2\nu} + \frac{1}{2\nu}\sqrt{M^2 + 4\nu\alpha} - \beta = 0$$

with α and β specified above. The lower bound is derived similarly.

C. Numerical Simulation Parameters

The channel parameters used in the numerical simulations are listed below.

For the 2×1 channel

$$\mathbf{H}_0 = [+0.3805 + 0.1069i \quad -0.1845 - 0.0985i],$$
$$\mathbf{R}_0 = \begin{bmatrix} 1.1843 & -0.7537 + 0.5638i \\ -0.7537 - 0.5638i & 0.8157 \end{bmatrix}.$$

For the 4×1 channel (numbers are rounded to two digits after the decimal point)

$$\mathbf{H}_0 = [.27 - .17i \quad -.06 + .18i \quad .14 + .25i \quad .33 - .27i],$$
$$\mathbf{R}_0 = \begin{bmatrix} .86 & .26 - .60i & .25 - .14i & .16 + .10i \\ .26 + .60i & 1.03 & .52 + .03i & -.09 - .08i \\ .25 + .14i & .52 - .03i & .72 & -.50 - .25i \\ .16 - .10i & -.09 + .08i & -.50 + .25i & 1.39 \end{bmatrix}.$$

Acknowledgment

The authors would like to thank Dr. R. Mahadevappa, Dr. C. Shabrami, and the four anonymous reviewers for their constructive comments, which helped to improve the quality of the paper.

References

[1] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. , pp. 1451–1458, Oct. 1998.

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

[3] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space–time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[4] H. Jafarkhani, "A quasi-orthogonal space time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, Jan. 2001.

[5] O. Tirkkonen, A. Boariu, and A. Hottinen, "Minimal nonorthogonality rate 1 space–time block code for 3+ Tx antennas," in *Proc. IEEE Int. Symp. Spread Spectrum Techniques Applications*, vol. 2, Sep. 2000, pp. 429–432.

[6] I. E. Telatar, "Capacity of multi-antenna Gaussian channels ," Internal Tech. Memo., AT&T—Bell Laboratories, Oct. 1995.

[7] 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. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.

[8] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

[9] M. Skoglund and G. Jöngren, "On the capacity of a multiple-antenna communication link with channel side information," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 395–405, Apr. 2003.

[10] M. Vu and A. Paulraj, "Some asymptotic capacity results for MIMO wireless with and without channel knowledge at the transmitter," in *Proc. 37th Asilomar Conf. Signals, Systems, Computers*, Nov. 2003, pp. 258–262.

[11] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

[12] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[13] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sep. 2004.

[14] S. Venkatesan, S. H. Simon, and R. A. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," in *Proc. IEEE Veh. Technol. Conf.*, vol. 3, Oct. 2003, pp. 1767–1771.

[15] D. Hösli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," presented at the 5th Int. ITG Conf. Source Channel Coding, Erlangen, Nuremburg, Jan. 14–16, 2004.

[16] M. Vu and A. Paulraj, "Optimum space–time transmission for a high K factor wireless channel with partial channel knowledge," *Wiley J. Wireless Commun. Mobile Comput. (Invited Paper)*, vol. 4, pp. 807–816, Nov. 2004.

[17] R. U. Nabar, H. Bölcskei, and A. J. Paulraj, "Diversity and outage performance in space–time block coded Ricean MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2519–2532, Sep. 2005.

[18] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space–time block coding," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[19] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space–time block coding based on channel mean feedback," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.

[20] H. Sampath and A. Paulraj, "Linear precoding for space–time coded systems with known fading correlations," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 239–241, Jun. 2002.

[21] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space–time block coding based on channel correlations," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1673–1690, Jul. 2003.

[22] L. Liu and H. Jafarkhani, "Application of quasi-orthogonal space–time block codes in beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 54–63, Jan. 2005.

[23] M. Vu and A. Paulraj, "Linear precoding for MIMO channels with nonzero mean and transmit correlation in orthogonal space–time coded systems," in *Proc. IEEE Vehicular Technology Conf.*, Sep. 2004, pp. 2503–2507.

[24] ——, "Linear precoding for MIMO wireless correlated channels with nonzero means: K factor analysis, extension to nonorthogonal STBC," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Mar. 2005, pp. 1113–1116.

[25] M. Vu and A. Paulraj, "A robust transmit CSI framework with applications in MIMO wireless precoding," in *Proc. 39th Asilomar Conf. Signals Systems, Computers*, Nov. 2005, pp. 623–627.

[26] H. Yao and G. W. Wornell, "Structured space–time block codes with optimal diversity-multiplexing tradeoff and minimum delay," in *Proc. IEEE Global Telecommunications Conf. (IEEE GLOBECOM)*, vol. 4, Dec. 2003, pp. 1941–1945.

[27] H. E. Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 968–985, Jun. 2004.

[28] A. Khoshnevis and A. Sabharwal, "On diversity and multiplexing gain of multiple antenna systems with transmitter channel information," presented at the Allerton Conf. Communications, Control, Computing, Montecilo, IL, Sep. 29–Oct. 1, 2004.

[29] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[30] D. S. Baum, D. Gore, R. Nabar, S. Panchanathan, K. V. S. Hari, V. Erceg, and A. J. Paulraj, "Measurement and characterization of broadband MIMO fixed wireless channels at 2.5 GHz," in *Proc. Int. Conf. Personal Wireless Communication*, Dec. 2000, pp. 203–206.

[31] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space–Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

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

[34] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[35] A. M. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.

[36] H. L. Van Trees, *Optimum Array Processing, Part IV of Detection, Estimation and Modulation Theory*. New York: Wiley, 2002.

Mai Vu (S'98) was born in Vietnam. She received the B.E. degree in computer systems engineering from Royal Melbourne Institute of Technology (RMIT), Australia, in 1997 and the M.S.E. degree in electrical engineering from the University of Melbourne, Australia, in 1999. She is currently working toward the Ph.D. degree in the Smart Antenna Research Group at the Department of Electrical Engineering, Stanford University, Stanford, CA.

Her industrial experiences include internships at Telstra Research Laboratory, Synopsys, Compaq, and Beceem Communications. Her research interests span various aspects of multiple-input multiple-output (MIMO) wireless communications, including coding and signal processing techniques, information capacity, and applications of convex optimization and random matrix theory. Her current focus is on exploiting partial channel knowledge at the transmitter.

Ms. Vu has been a recipient of the Intel Foundation Ph.D. Fellowship and the Rambus Corporation Stanford Graduate Fellowship, among others. She was awarded the Australian Institute of Engineers Award for the best final year student in the School of Engineering, RMIT.

Arogyaswami Paulraj (SM'85–F'91) received the Ph.D. degree from the Indian Institute of Technology, New Delhi, in 1973.

Currently, he is currently a Professor with the Department of Electrical Engineering, Stanford University, Stanford, CA, where he supervises the Smart Antennas Research Group, working on applications of space–time techniques for wireless communications. His nonacademic positions have included Head of the Sonar Division, Naval Oceanographic Laboratory, Cochin, India; the Director of the Center for Artificial Intelligence and Robotics, India; the Director of the Center for Development of Advanced Computing, India; the Chief Scientist of Bharat Electronics, India; Chief Technology Officer (CTO) and Founder of Iospan Wireless, Inc.; Co-Founder and CTO of Beceem Communications Inc. He is the author of over 300 research papers and holds 20 patents. His research has spanned several disciplines, emphasizing estimation theory, sensor signal processing, parallel computer architectures/algorithms, and space–time wireless communications. His engineering experience has included development of sonar systems, massively parallel computers, and broadband wireless systems.

Dr. Paulraj is a Member of the Indian National Academy of Engineering. He has won several awards for his research and engineering contributions, including the IEEE Signal Processing Society's Technical Achievement Award.