Generalized Linear Precoder and Decoder Design for MIMO Channels Using the Weighted MMSE Criterion

Hemanth Sampath, *Member, IEEE*, Petre Stoica, *Fellow, IEEE*, and Arogyaswami Paulraj, *Fellow, IEEE*

***Abstract*—We address the problem of designing jointly optimum linear precoder and decoder for a MIMO channel possibly with delay-spread, using a *weighted minimum mean-squared error (MMSE) criterion* subject to a transmit power constraint. We show that the optimum linear precoder and decoder diagonalize the MIMO channel into eigen subchannels, for any set of error weights. Furthermore, we derive the optimum linear precoder and decoder as functions of the error weights and consider specialized designs based on specific choices of error weights. We show how to obtain: 1) the maximum information rate design; 2) QoS-based design (we show how to achieve any set of relative SNRs across the subchannels); and 3) the (unweighted) MMSE and equal-error design for fixed rate systems.**

***Index Terms*—MIMO, MMSE, precoding, space–time coding, spatial multiplexing.**

I. Introduction

SPACE–TIME coding [19]–[21] and spatial multiplexing [5], [9] are prime candidates for achieving high data rates over MIMO channels. Spatial multiplexing involves transmitting independent streams of data across multiple antennas to maximize throughput, whereas space–time coding appropriately maps input symbol streams across space and time for transmit diversity and coding gain at a given data rate. Neither schemes require channel knowledge at the transmitter. In a number of applications, channel knowledge is also available at the transmitter. Channel status information can be fedback to the transmitter in frequency division duplex (FDD) systems or can be estimated in the receiver mode in time division duplex (TDD) systems. In such scenarios, channel-dependent linear transmit and receive processing (linear precoding and decoding respectively) of data streams can improve performance by optimally allocating resources such as power and bits over multiple antennas, depending on the channel. Linear precoding and decoding schemes are scalable to any number of antennasand are simpler to implement when compared to nonlinear schemes.

Paper approved for publication by M. Chiani, the Editor for Transmission Systems of the IEEE Communications Society. Manuscript received May 30, 2000; revised April 15, 2001. This paper was presented in part at the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers 2000.

H. Sampath was with the Information Systems Laboratory, Stanford University, Stanford, CA 94305-9510 USA. He is now with Iospan Wireless Inc., San Jose, CA USA (e-mail: hemanth1@stanfordalumni.org).

P. Stoica is with the Department of Systems and Control, Uppsala University, Uppsala, Sweden (e-mail: ps@syscon.uu.se).

A. Paulraj is with the Information Systems Laboratory, Stanford University, Stanford, CA 94305-9510 USA (e-mail: paulraj@rascals.stanford.edu).

Publisher Item Identifier S 0090-6778(01)10618-5.

We now summarize some well-known jointly optimum linear precoder and decoder designs for MIMO systems. The optimum linear precoder and decoder that maximize information rate or that minimize *sum of output symbol estimation errors*, decouple the MIMO channel into eigen subchannels (eigenmodes[1]) and allocate power on these subchannels according to the water-pouring policy [11], [14] and inverse-water pouring policy [18], [12], [15], respectively. Other designs such as that maximizing the output SNR subject to a zero forcing constraint and a transmit power constraint (see, for example, [15]) have also been considered in recent years.

While the minimum mean-squared error (MMSE) design minimizes the sum of symbol estimation errors across eigenmodes, it does not tell us anything about the mean square error on each eigenmode. Some eigenmodes may have larger errors compared to the others. In order to get a better handle on the errors on each eigenmode, we chose to design a generalized linear precoder and decoder that minimize any weighted sum of symbol estimation errors. The solution to the weighted MMSE design yields solutions to well-known designs (mentioned above) as well as new designs. Our generalized criterion provides a unified framework for designing jointly optimal linear precoders and decoders.

In this paper, we derive the generalized jointly optimum linear precoder and decoder that minimize any *weighted sum of symbol estimation errors*, assuming *total transmit power constraint* across all transmit antennas. We summarize now the main results of the paper.

- We first introduce the system models for single carrier and multicarrier systems operating on MIMO channels with and without delay spread and show that they have the same form. This allows us to apply our generalized design to all such scenarios.
- Then we derive the optimum structure for the linear precoder and decoder, assuming a total power constraint, and show that they diagonalize the MIMO channel into eigen subchannels, for any set of error weights. In other words, eigenmode transmission is optimum for any set of error weights.[2]
- Next, closed-form solutions are derived for the optimum precoder and decoder as functions of error weights,

[1]Eigen subchannels and eigenmodes are equivalent terms and are used interchangeably in this paper.

[2]Interestingly, it is shown in [13] that eigenmode transmission need not be the optimum strategy if we impose a peak power constraint on each transmit antenna.

transmit power, receiver noise variance, and eigenvalues of the MIMO channel. We show how to select appropriate error weights to obtain: 1) the maximum information rate (max-IR) design; 2) QoS-based design (we show how to achieve any set of relative SNRs across the subchannels); and 3) optimum precoder and decoder designs for fixed-rate transmission systems, namely, the (unweighted) MMSE and the equal-error designs.

II. System Model

A generic MIMO communication system model is shown in Fig. 1. The input bit streams are coded and modulated to generate symbol streams. The latter are then passed through the linear precoder which is optimized for a fixed and known channel. The precoder is a matrix with complex elements and can add redundancy to the input symbol streams to improve system performance (as will be explained later). The precoder output is launched into the MIMO channel through M_T transmit antennas. The signal is received by M_R receive antennas and processed by the linear decoder, which is optimized for the fixed and known channel. The linear decoder also operates in the complex field and removes any redundancy that has been introduced by the precoder. In the present paper, we will focus on the *boxed* section of the communication system shown in Fig. 1, i.e., we will not consider coding and modulation design, but instead focus only on the design of the linear precoder and decoder.

A. Single Carrier System in Flat-Fading Channel

For a MIMO channel without any delay-spread, the system equation is

$$\hat{\mathbf{s}} = \mathbf{GHFs} + \mathbf{Gn} \tag{1}$$

where $\mathbf{H}$ is an $M_R \times M_T$ channel matrix, $\hat{\mathbf{s}}$ is the $B \times 1$ received vector, $\mathbf{s}$ is the $B \times 1$ transmitted vector, where $B = \mathrm{rank}(\mathbf{H}) \leq \min(M_R, M_T)$ is the number of parallel streams of data that is to be transmitted, $\mathbf{n}$ is the $M_R \times 1$ noise vector (at the given symbol time), $\mathbf{G}$ is the $B \times M_R$ decoder matrix, and $\mathbf{F}$ is the $M_T \times B$ precoder matrix. The latter adds a redundancy of $M_T - B$ across space, since it has B input symbols and M_T precoded output symbols that are launched simultaneously through M_T transmit antennas. Note that linear precoding and decoding are implemented by simple matrix multiplications. We assume that

$$E(\mathbf{ss}^*) = \mathbf{I}; \quad E(\mathbf{nn}^*) = \mathbf{R}_{\mathbf{nn}}; \quad E(\mathbf{sn}^*) = 0 \tag{2}$$

where the superscript * denotes the conjugate transpose. Note that, for simplicity of analysis, we have assumed uncorrelated input sources, each normalized to unit power, and that $B = \mathrm{rank}(\mathbf{H})$. Based on the structure of the optimum precoder and decoder, we will later show how to handle the case when $B < \mathrm{rank}(\mathbf{H})$. The case when $B > \mathrm{rank}(\mathbf{H})$ is not possible, since the transmitter knows the channel and hence for any practical system will not transmit more than $\mathrm{rank}(\mathbf{H})$ independent data streams.

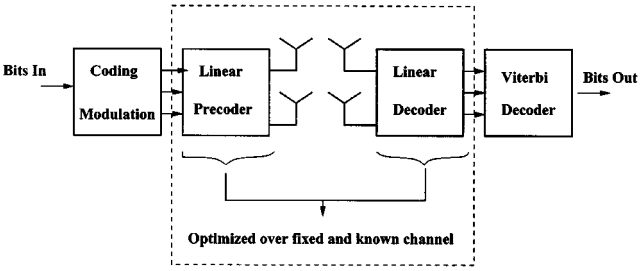

Fig. 1. MIMO communication system.

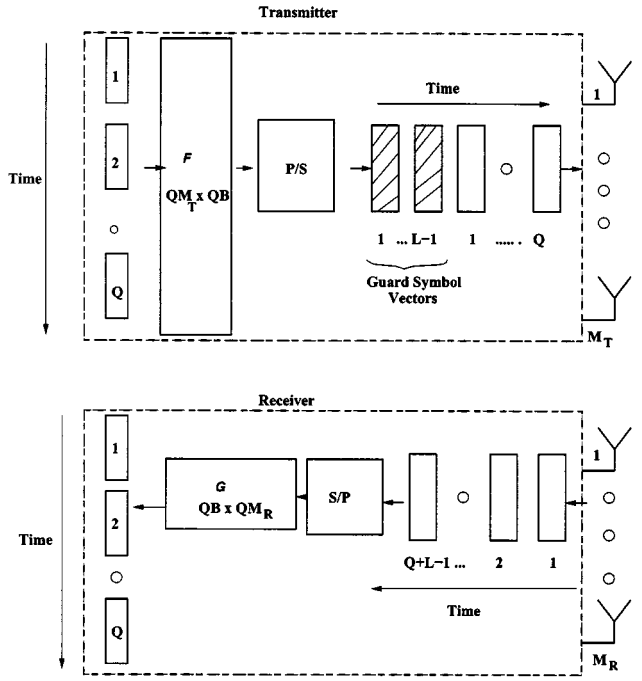

Fig. 2. Linear block precoding and decoding for delay-spread channels.

B. Single Carrier System in Delay-Spread Channel

For a MIMO channel with a finite delay-spread that spans L symbol periods, we can write the system equation (see Fig. 2) for a finite number of samples in block form. In the sequel, we will restrict ourselves to the simpler block processing of data symbols and will not consider dynamic processing where the input blocks are allowed to overlap in time.

Consider a data *block* with Q input symbol vectors (in time succession), denoted as $\mathcal{S}_{ij}$ where i is the symbol time index within a block and j is the block index. We stack Q input symbol vectors, $\mathcal{S}_{1j}, \mathcal{S}_{2j}, \ldots, \mathcal{S}_{Qj}$ in a $(QB \times 1)$ vector $\mathcal{S}_j = \mathrm{vec}([\mathcal{S}_{1j} \ldots \mathcal{S}_{jQ}])$. This vector is processed by a $QM_T \times QB$ precoder matrix $\mathcal{F}$ that outputs QM_T symbols. These symbols are passed through a parallel-to-serial (P/S) converter to get Q vectors of size $M_T \times 1$ each. These vectors are then launched into the MIMO channel in time succession, at the end of which $L-1$ zero guard symbol vectors of size $M_T \times 1$ each are transmitted to prevent inter-block interference.

The system equation for the jth data block has a similar form compared to (1) and is given as

$$\hat{\mathcal{S}}_j = \mathcal{G}\bar{\mathcal{H}}\mathcal{F}\mathcal{S}_j + \mathcal{G}\mathcal{N}_j \tag{3}$$

where $\mathcal{G}$ is a $QB \times (Q+L-1)M_R$ space–time decoder block matrix that outputs an estimate $\hat{\mathcal{S}}_j = \text{vec}([\hat{\mathcal{S}}_{1j} \dots \hat{\mathcal{S}}_{jQ}])$ of the transmitted symbol vectors, and $\bar{\mathcal{H}}$ is given by

$$\bar{\mathcal{H}} = \begin{pmatrix} \mathcal{H}_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \mathcal{H}_L & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \mathcal{H}_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathcal{H}_L \end{pmatrix}_{(Q+L-1)M_R \times QM_T} \tag{4}$$

where $\mathcal{H}_j, 1 \le j \le L$, denotes the jth tap of the MIMO channel impulse response matrix and where we have assumed that the channel is stationary over the input block length. Note that $\mathcal{F}$ and $\mathcal{G}$ need not be Toeplitz matrices, i.e., we are not imposing any structure on $\mathcal{F}$ and $\mathcal{G}$. Also, $\mathcal{F}$ operates across space and time and can be thought of as a *spatio-temporal channel-dependent block code* that operates in the complex field.

C. Multicarrier System

Recently, multicarrier modulation in the form of orthogonal frequency division multiplexing (OFDM) or discrete matrix multi tone (DMMT) modulation has received considerable attention for MIMO transmission systems to combat delay spread [1], [3], [11].

In such a scheme, the total frequency bandwidth is divided into Q frequency bins. Consider a data block with Q input symbol vectors of size $B \times 1$ each, denoted as $\mathcal{S}_{ij}, i = 1, \dots Q$, where i is the frequency bin index and j is the block index. We stack the Q input symbol vectors, $\mathcal{S}_{1j}, \mathcal{S}_{2j}, \dots, \mathcal{S}_{Qj}$ in a $(QB \times 1)$ vector $\mathcal{S}_j = \text{vec}([\mathcal{S}_{1j} \dots \mathcal{S}_{jQ}])$. This vector is processed by a $QM_T \times QB$ precoder matrix $\mathcal{F}$ to get Q vectors of size $M_T \times 1$ each. The OFDM time waveform is then generated by taking the Q-point vector inverse fast Fourier transform (IFFT) of the above data block. A cyclic prefix (CP) of length $L-1$ symbol vectors is then added to the beginning of the time waveform and launched into the MIMO channel. At the receiver, the first and last L symbol vectors in the received time waveform are ignored. A Q-point vector FFT of the resulting time waveform is taken, the output of which is then processed by a $QB \times QM_R$ linear decoder to get the received data block, $\hat{\mathcal{S}}_j$.

Mathematically, the above procedure can be represented as

$$\hat{\mathcal{S}}_j = \mathcal{G}\left(\mathbf{\Theta}_{\mathbf{Q}}^* \otimes \mathbf{I}_{M_R \times M_R}\right) \tilde{\mathcal{H}} (\bar{\mathbf{\Theta}}_{\mathbf{Q}} \otimes \mathbf{I}_{M_T \times M_T}) \mathcal{F} \mathcal{S}_j + \mathcal{G}\left(\mathbf{\Theta}_{\mathbf{Q}}^* \otimes \mathbf{I}_{M_R \times M_R}\right) \mathcal{N}_j \tag{5}$$

where $\mathcal{S}_j = \text{vec}([\mathcal{S}_{1j} \dots \mathcal{S}_{jQ}])$, $\hat{\mathcal{S}}_j = \text{vec}([\hat{\mathcal{S}}_{1j} \dots \hat{\mathcal{S}}_{jQ}])$, $\mathcal{F}$ and $\mathcal{G}$ are the linear precoder and decoder matrices, respectively, $\otimes$ is the Kronecker product [8], and

$$\tilde{\mathcal{H}} = \begin{pmatrix} \mathcal{H}_L & \cdots & \mathcal{H}_1 & 0 & \\ 0 & \mathcal{H}_L & \cdots & \mathcal{H}_1 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots 0 & \mathcal{H}_L & \cdots & \mathcal{H}_1 \end{pmatrix}_{QM_R \times (Q+L-1)M_T} . \tag{6}$$

Note that $\bar{\mathbf{\Theta}}_{\mathbf{Q}}$ is the $(Q+L) \times Q$ matrix obtained by taking a $Q \times Q$ IFFT matrix $\mathbf{\Theta}_{\mathbf{Q}}$ (the (l,m)th entry of $\mathbf{\Theta}_{\mathbf{Q}}$ is given as $\mathbf{\Theta}_{Q_{l,m}} = e^{-2\pi\sqrt{-1}lm/Q}; l, m = 0, 1, \dots Q-1$) and adding the last L rows on top, creating the so-called cyclic prefix. Defining $\tilde{\mathbf{H}} := (\mathbf{\Theta}_{\mathbf{Q}}^* \otimes \mathbf{I}_{M_R \times M_R}) \tilde{\mathcal{H}} (\bar{\mathbf{\Theta}}_{\mathbf{Q}} \otimes \mathbf{I}_{M_T \times M_T}) = \text{diag}(\tilde{\mathbf{H}}_1, \dots, \tilde{\mathbf{H}}_Q)$ to be a $QM_R \times QM_T$ block diagonal matrix made from the MIMO transfer function values at the frequency bins $(1, 2, \dots, Q)$, we can write from (5)

$$\hat{\mathcal{S}}_j = \mathcal{G}\tilde{\mathbf{H}}\mathcal{F}\mathcal{S}_j + \mathcal{G}\tilde{\mathbf{N}}_j$$

which has a similar form to (1). The advantage of using the OFDM approach is that an equivalent channel matrix which is block diagonal enforces a similar structure upon the optimal precoder and decoder, which simplifies the transmitter and receiver structure.

III. Problem Formulation

In the sequel, we will use the system equation that corresponds to the single carrier system in a flat-fading channel. However, as explained before, our analysis is also applicable to the single carrier system in a delay-spread channel, as well as multicarrier systems employing OFDM modulation. In the latter scenarios, only the dimensionality of the system variables (i.e., $\mathbf{G}, \mathbf{F}, \mathbf{s}$ and $\mathbf{n}$) as well as the structure of $\mathbf{H}$ will change as discussed in the previous section.

Our goal is to design the $\mathbf{F}$ and $\mathbf{G}$ matrices to minimize any weighted combination of symbol estimation errors, namely $E[\mathbf{e}^*\mathbf{W}^{*1/2}\mathbf{W}^{1/2}\mathbf{e}]$ where $\mathbf{e} = \mathbf{s} - (\mathbf{GHFs} + \mathbf{Gn})$ is the $B \times 1$ error vector and $\mathbf{W}^{1/2}$ is the $B \times B$ square root of a *diagonal* positive definite weight matrix $\mathbf{W}$. We assume that the channel matrix $\mathbf{H}$ is fixed and known at the transmitter and receiver. Mathematically, the problem statement is as follows:

$$\min_{\mathbf{G},\mathbf{F}} : c(\mathbf{G}, \mathbf{F}) = E\|\mathbf{W}^{1/2}\mathbf{e}\|^2 \quad \text{Subject to: } \text{tr}(\mathbf{FF}^*) \le p_0 \tag{7}$$

where the expectation (E) is with respect to the distributions of $\mathbf{s}$ and $\mathbf{n}$, and p_0 is the total power available. Note that

$$c(\mathbf{G}, \mathbf{F}) = E\|\mathbf{W}^{1/2}\mathbf{e}\|^2 = E(\text{tr}[\mathbf{W}^{1/2}\mathbf{e}\mathbf{e}^*\mathbf{W}^{*1/2}]) = \text{tr}[\mathbf{W} \times \mathbf{R}_{\mathbf{e}}(\mathbf{G}, \mathbf{F})] \tag{8}$$

where $\mathbf{R}_{\mathbf{e}}(\mathbf{G}, \mathbf{F})$ is the error covariance matrix, defined as $\mathbf{R}_{\mathbf{e}}(\mathbf{G}, \mathbf{F}) := E(\mathbf{e}\mathbf{e}^*)$. Using the expression for $\mathbf{e}$ and the assumptions made in (2), we have that

$$\mathbf{R}_{\mathbf{e}}(\mathbf{G}, \mathbf{F}) = (\mathbf{GHF} - \mathbf{I})(\mathbf{GHF} - \mathbf{I})^* + \mathbf{GR}_{\mathbf{nn}}\mathbf{G}^*. \tag{9}$$

We use the method of Lagrange duality and the Karush–Kuhn–Tucker (KKT) conditions (see below) to solve the optimization problem in (7). We first form the Lagrangian (below μ is the Lagrange multiplier)

$$L(\mu, \mathbf{G}, \mathbf{F}) = c(\mathbf{G}, \mathbf{F}) + \mu[\text{tr}(\mathbf{FF}^*) - p_0]. \tag{10}$$

Using (8) and (9) in (10), we can write

$$L(\mu, \mathbf{G}, \mathbf{F}) = \text{tr}[\mathbf{W}(\mathbf{GHF} - \mathbf{I})(\mathbf{GHF} - \mathbf{I})^* + \mathbf{WGR}_{\mathbf{nn}}\mathbf{G}^*] + \mu[\text{tr}(\mathbf{FF}^*) - p_0]. \tag{11}$$

The following KKT conditions are necessary and sufficient for optimality: $\mathbf{G}$ and $\mathbf{F}$ are optimal if and only if there is a μ that together with $\mathbf{G}$ and $\mathbf{F}$ satisfy the conditions

$$\nabla_{\mathbf{G}} L(\mu, \mathbf{G}, \mathbf{F}) = 0 \tag{12}$$

$$\nabla_{\mathbf{F}} L(\mu, \mathbf{G}, \mathbf{F}) = 0 \tag{13}$$

$$\mu \geq 0; \quad \mathrm{tr}(\mathbf{F}\mathbf{F}^*) - p_0 \leq 0 \tag{14}$$

$$\mu[\mathrm{tr}(\mathbf{F}\mathbf{F}^*) - p_0] = 0. \tag{15}$$

Using (11) in (12) and (13), we get the following relations between $\mathbf{G}$ and $\mathbf{F}$, respectively:

$$\mathbf{H}\mathbf{F} = \mathbf{H}\mathbf{F}\mathbf{F}^*\mathbf{H}^*\mathbf{G}^* + \mathbf{R}_{\mathbf{nn}}\mathbf{G}^* \tag{16}$$

$$\mathbf{W}\mathbf{G}\mathbf{H} = \mathbf{F}^*\mathbf{H}^*\mathbf{G}^*\mathbf{W}\mathbf{G}\mathbf{H} + \mu\mathbf{F}^*. \tag{17}$$

To obtain (16) and (17), we have used the fact that $(\partial\,\mathrm{tr}(\mathbf{A}\mathbf{X}\mathbf{B}))/(\partial\mathbf{X}) = \mathbf{B}\mathbf{A}$ and $(\partial\,\mathrm{tr}(\mathbf{A}\mathbf{X}^*\mathbf{B}))/(\partial\mathbf{X}) = 0$ (see, for example, [8]) for a complex-valued[3] $\mathbf{X}$.

IV. Optimum Precoder and Decoder

We now solve (16) and (17) to derive the optimum precoder and decoder. Let us define the eigenvalue decomposition (EVD)

$$\mathbf{H}^*\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H} = (\mathbf{V} \quad \tilde{\mathbf{V}}) \begin{pmatrix} \mathbf{\Lambda} & 0 \\ 0 & \tilde{\mathbf{\Lambda}} \end{pmatrix} (\mathbf{V} \quad \tilde{\mathbf{V}})^* \tag{18}$$

where $\mathbf{V}$ is an $M_T \times B$ orthogonal matrix which forms a basis for the range space of $\mathbf{H}^*\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the B nonzero eigenvalues arranged in a decreasing order from top-left to bottom-right; $\tilde{\mathbf{\Lambda}}$ contains the zero eigenvalues; $\tilde{\mathbf{V}}$ is a $M_T \times (M_T - B)$ orthogonal matrix which constitutes a basis for the null space of $\mathbf{H}^*\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}$. Note that for now we have assumed $\mathrm{rank}(\mathbf{H}) = B$ as explained in the previous section.

Lemma 1: The optimum $\mathbf{F}$ and $\mathbf{G}$ matrices can be assumed to have the following structure without any loss of generality[4]:

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi}_{\mathbf{f}} \tag{19}$$

$$\mathbf{G} = \mathbf{\Phi}_{\mathbf{g}}\mathbf{V}^*\mathbf{H}^*\mathbf{R}_{\mathbf{nn}}^{-1} \tag{20}$$

where $\mathbf{\Phi}_{\mathbf{f}}$ and $\mathbf{\Phi}_{\mathbf{g}}$ are $B \times B$ matrices.

Proof: See Appendix I-A. ∎

Lemma 2: The matrices $\mathbf{\Phi}_{\mathbf{f}}$ and $\mathbf{\Phi}_{\mathbf{g}}$ in Lemma 1 are diagonal with nonnegative elements on the diagonal. Hence, the optimum precoder and decoder diagonalize $\mathbf{H}$ into eigen subchannels for any set of error weights. More exactly, we have $\mathbf{G}\mathbf{H}\mathbf{F} = \mathbf{\Phi}_{\mathbf{g}}\mathbf{\Lambda}\mathbf{\Phi}_{\mathbf{f}}$.

Proof: See Appendix I-B. ∎

Theorem 1: The optimum $\mathbf{\Phi}_{\mathbf{f}}$ and $\mathbf{\Phi}_{\mathbf{g}}$ in Lemma 1 are given by

$$\mathbf{\Phi}_{\mathbf{f}} = (\mu^{-1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{1/2} - \mathbf{\Lambda}^{-1})_+^{1/2} \tag{21}$$

$$\mathbf{\Phi}_{\mathbf{g}} = (\mu^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{-1/2} - \mu\mathbf{\Lambda}^{-1}\mathbf{W}^{-1})_+^{1/2}\mathbf{\Lambda}^{-1/2}. \tag{22}$$

[3]When evaluating the derivatives in (12) and (13), $\mathbf{G}$ and $\mathbf{G}^*$ are treated as independent variables and, similarly, for $\mathbf{F}$ and $\mathbf{F}^*$; the differentiation of (11) w.r.t. $\mathbf{F}^*$ and $\mathbf{G}^*$ just gives (16) and (17), and hence brings no additional equations.

[4]We thank Dr. Anna Scaglione for valuable comments on formulating and proving this lemma.

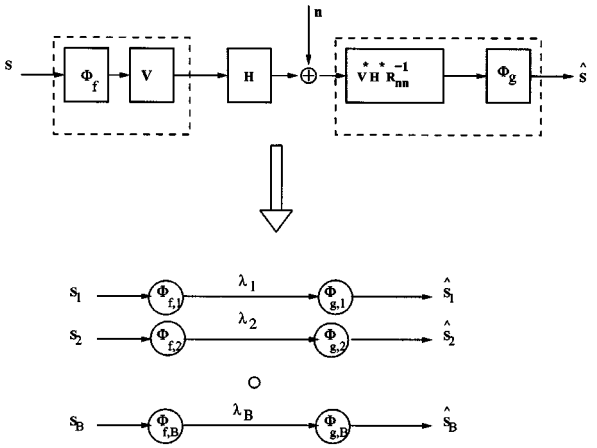

Fig. 3. Optimum transceivers: decomposition into eigen subchannels.

The symbol $(\cdot)_+$ denotes the fact that that the negative elements of the diagonal bracketed matrix are replaced by zero. The μ in (21) and (22) is chosen to satisfy the transmitter power constraint.

Proof: See Appendix I-C. ∎

We now show how to obtain a $\mu > 0$ that satisfies the power constraint and also is such that the the bracketed matrices in both (21) and (22) are positive semi-definite. Let $\mathbf{\Lambda} = \mathrm{diag}([\lambda_1, \lambda_2, \ldots, \lambda_B])$, $\mathbf{W} = \mathrm{diag}([w_1, w_2, \ldots, w_B])$, and $\rho_k = \lambda_k w_k$ and let them be ordered in a decreasing manner: $\rho_1 \geq \cdots \geq \rho_k \geq \cdots \geq \rho_B$. We get the following expression for μ from the trace constraint $\mathrm{tr}(\mathbf{\Phi}_{\mathbf{f}}^2) = p_0$, when $k \leq B$ subchannels are used for transmission:

$$\mu^{1/2} = \frac{\sum_{j=1}^{k} \left(\lambda_j^{-1/2} w_j^{1/2}\right)}{p_0 + \sum_{j=1}^{k} \left(\lambda_j^{-1}\right)} \tag{23}$$

The following iterative procedure initialized with $k = B$ can be used to optimally compute $\mathbf{\Phi}_{\mathbf{f}}$. Let $\mathbf{\Phi}_{\mathbf{f}} = \mathrm{diag}([\phi_{f1}, \phi_{f2}, \ldots, \phi_{fB}])$.

1) Assume $\mu \leq \rho_k$, which from (21) implies $\phi_{fj} \geq 0$, for $j = 1, \ldots, k$. Compute μ according to (23) under this assumption. If $\mu \leq \rho_k$, stop; otherwise go to step 2.
2) Set $\phi_{f,k} = 0$ and $k = k - 1$. Go to step 1.

It can be verified that the above iteration will terminate in $B - 1$ steps at most.

Remark 1: The optimum precoder and decoder were derived for the case when $B = \mathrm{rank}(\mathbf{H})$. For the case when $B < \mathrm{rank}(\mathbf{H})$, it is likely that eigenmode transmission is still optimum. In such a case, the power is allocated to only those B eigenmodes $(\rho_1, \ldots, \rho_B)$, that minimize the total cost function in (8).

Remark 2: Lemmas 1 and 2 and Theorem 1 suggest that the MIMO channel can be decoupled into eigen subchannels for any set of error weights. Only the power allocation across the eigen subchannel need be computed depending on the criterion to be optimized (see Fig. 3).

We now provide expressions for the BER, SNR, and MSE across the subchannels for the weighted MMSE design. We will need these expressions later on. Assuming a square QAM constellation, the probability of uncoded symbol error (and hence

the BER assuming Gray encoding) for the ith subchannel is given by (see, for example, [10])

$$P_{e,i} \simeq N_{e,i} Q\left(\sqrt{\beta_{M_i}\Gamma_i}\right) \tag{24}$$

where for a given constellation size $M_i, \beta_{M_i} = (3/(M_i - 1))$, $N_{e,i}$ is the number of nearest neighbors and Γ_i is the subchannel SNR and is obtained from the (i,i)th element in the SNR matrix defined as $\mathbf{\Gamma} = (\mathbf{F}^*\mathbf{H}^*\mathbf{G}^*)(\mathbf{G}\mathbf{R}_{\mathbf{nn}}\mathbf{G}^*)^{-1}\mathbf{G}\mathbf{H}\mathbf{F}$.

Using the optimum $\mathbf{F}$ and $\mathbf{G}$ matrices from Lemma 1 and Theorem 1, $\mathbf{\Gamma}$ can be simplified to

$$\mathbf{\Gamma} = \mathbf{\Phi}_\mathbf{f}^2\mathbf{\Lambda} = (\mathbf{W}^{1/2}\mu^{-1/2}\mathbf{\Lambda}^{1/2} - \mathbf{I})_+ \tag{25}$$

and the $\mathbf{R_e}$ matrix in (9) can be simplified to

$$\mathbf{R_e} = \mathbf{W}^{-1/2}\mathbf{\Lambda}^{-1/2}\mu^{1/2}. \tag{26}$$

From (25) and (26), we see that $\mathbf{\Gamma} = \mathbf{R_e}^{-1} - \mathbf{I}$. Hence, equal SNRs on each subchannel implies equal MSEs, for any set of error-weights.

V. Applications

In this section, we show how to obtain linear precoder and decoder designs for different applications by appropriately choosing the error weights $\mathbf{W}$.

A. Maximum Information Rate Design

It is well known (see, for example, [11] and [14]) that the optimum precoder and decoder that maximize information rate are given by (19) and (20), with $\mathbf{\Phi_f}$ chosen according to the well-known *water-pouring* solution

$$\mathbf{\Phi_f} = \left(\frac{\mathbf{I}}{\mu^{1/2}} - \mathbf{\Lambda}^{-1}\right)_+^{1/2} \tag{27}$$

where $\mu > 0$ is computed according to the transmit power constraint. Note that in the expression for the decoder given in (20), $\mathbf{\Phi_g}$ can be any arbitrary full-rank diagonal matrix since it does not enter in the expression for the maximum information rate given as (see for example, [4] and [11])

$$\mathcal{R} = \sum_{i=1}^{B}\mathcal{R}_i = \sum_{i=1}^{B}\log_2\left(1 + \phi_{f,i}^2\lambda_i\right) \tag{28}$$

where $\mathcal{R}_i$ is the information rate for the ith subchannel.

Lemma 3: The choice $\mathbf{W} = \mathbf{\Lambda}$ in the expression for $\mathbf{\Phi_f}$ obtained from the weighted MMSE design (see (21)) results in the well-known water-pouring solution given in (27). Hence, the maximum information rate design is just a special case of our generalized design.

For the sake of completeness, we give below the expression for $\mathbf{\Phi_g}$ obtained when we substitute $\mathbf{W} = \mathbf{\Lambda}$ into (22): $\mathbf{\Phi_g} = (\mu^{1/2}\mathbf{\Lambda}^{-1} - \mu\mathbf{\Lambda}^{-2})_+^{1/2}\mathbf{\Lambda}^{-1/2}$.

The choice of error weights $\mathbf{W} = \mathbf{\Lambda}$ tells us that stronger subchannels support higher rates (when compared to the weaker subchannels) and hence need to be more heavily weighed in the cost function. The max-IR design finds applications in *adaptive modulation systems* (see, for example, [6]), where more power and higher order modulation are used on subchannels with higher gains to improve data rates.

B. QoS-Based Design

Consider a multimedia application where different types of information (audio, video, etc.) need to be sent simultaneously on different subchannels. Video typically has a higher SNR requirement than audio, for successful transmission. In such QoS-based applications, it is necessary to have subchannels with different SNRs.

We now show how to pick the $\mathbf{W}$ matrix to achieve any set of relative output SNRs across the subchannels. From (25), this boils down to solving the following equation[5] with respect to $\mathbf{W}$:

$$\mathbf{\Gamma} = \mathbf{W}^{1/2}\frac{1}{\mu^{1/2}}\mathbf{\Lambda}^{1/2} - \mathbf{I} = \gamma\mathbf{D} \tag{29}$$

where $\mathbf{D} = \mathrm{diag}([d_1, d_2, \ldots, d_B])$ is a diagonal matrix of relative SNRs across subchannels and $\gamma > 0$ is a scalar. We assume that $\sum_{i=1}^{B} d_i = 1$. The solution $\mathbf{W}$ is readily computed from (29) as

$$\mathbf{W}^{1/2} = (\mathbf{I} + \gamma\mathbf{D})\mathbf{\Lambda}^{-1/2}\mu^{1/2}. \tag{30}$$

Note that μ is a function of $\mathbf{W}$, given from (23) with $k = B$ as

$$\mu^{1/2} = \frac{\mathrm{Tr}(\mathbf{\Lambda}^{-1/2}\mathbf{W}^{1/2})}{\mathrm{Tr}(\mathbf{\Lambda}^{-1}) + p_0}. \tag{31}$$

Substituting the expression for $\mathbf{W}$ from (30) into (31) and simplifying,we get $\gamma = p_0/\mathrm{Tr}(\mathbf{\Lambda}^{-1}\mathbf{D})$. Inserting this into (30) yields

$$\mathbf{W}^{1/2} = \mu^{1/2}\left(\frac{p_0}{\mathrm{Tr}(\mathbf{\Lambda}^{-1}\mathbf{D})}\mathbf{D} + \mathbf{I}\right)\mathbf{\Lambda}^{-1/2}. \tag{32}$$

Substituting (32) into the expressions for $\mathbf{\Phi_f}$ and $\mathbf{\Phi_g}$ in Theorem 1 and using the optimum precoder and decoder structure from Lemma 1, we get the following design.

Theorem 2: The optimum precoder and decoder that provide any set of relative SNRs ($\mathbf{D}$) across subchannels are given by

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi_f}; \quad \mathbf{G} = \mathbf{\Phi_g}\mathbf{V}^*\mathbf{H}^*\mathbf{R}_{\mathbf{nn}}^{-1}$$

$$\mathbf{\Phi_f} = \gamma^{1/2}\mathbf{D}^{1/2}\mathbf{\Lambda}^{-1/2} \tag{33}$$

$$\mathbf{\Phi_g} = \gamma^{1/2}\mathbf{D}^{1/2}\mathbf{\Lambda}^{-1/2}(\gamma\mathbf{D} + \mathbf{I})^{-1} \tag{34}$$

$$\gamma = \frac{p_0}{\mathrm{tr}(\mathbf{D}\mathbf{\Lambda}^{-1})}. \tag{35}$$

C. (Unweighted) MMSE Design

The jointly optimal linear precoder and decoder that minimize the *sum of symbol estimation errors* is well known in the literature [18], [15]. Such a design can be trivially obtained by choosing $\mathbf{W} = \mathbf{I}$ in Theorem 1. The MMSE design improves MIMO system performance [18], [12], [14], as will be illustrated later via Monte Carlo simulations.

Note that, while the MMSE design minimizes the sum of the symbol estimation errors across all subchannels, it does not

[5]We have assumed for simplicity that $\mathbf{\Phi_f} > 0$ and that $B = \mathrm{rank}(\mathbf{H})$. The case when $B < \mathrm{rank}(\mathbf{H})$ can be handled by using the argument in Remark 1.

guarantee that the MSEs and SNRs on each subchannel is minimized. In fact, the weaker subchannels can have a higher MSE than the stronger subchannel (as observed by substituting $\mathbf{W} = \mathbf{I}$ in (25) and (26)). Furthermore, the MMSE power allocation policy allocates no power to an eigenmode, if its gain is less than a certain threshold, i.e., the weakest eigenmodes are dropped. The power is then redistributed among the remaining eigenmodes, so that more power is allocated to the weaker eigenmodes and vice versa.

D. Equal Error Design

As discussed in the previous subsection, the MMSE design can result in unequal MSE across subchannels. For fixed rate systems that require the reliable transmission of B symbol streams using identical modulation and coding scheme, we need that all B subchannels have equal errors.

This translates into requiring an equal SNR for the given subchannels [see (24)]. This requirement can be satisfied trivially by selecting $\mathbf{D} = (1/B)\mathbf{I}$ in the QoS-based design (see Theorem 2).

Theorem 3: The optimum precoder and decoder that ensure equal errors across subchannels for a fixed-rate system with B data streams is given by[6]

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi_f}; \quad \mathbf{G} = \mathbf{\Phi_g}\mathbf{V}^*\mathbf{H}^*\mathbf{R}_{\mathbf{nn}}^{-1}$$

$$\mathbf{\Phi_f} = \gamma^{1/2}\mathbf{\Lambda}^{-1/2} \tag{36}$$

$$\mathbf{\Phi_g} = \gamma^{1/2}(1+\gamma)^{-1}\mathbf{\Lambda}^{-1/2} \tag{37}$$

$$\gamma = \sqrt{\frac{p_0}{\mathrm{Tr}(\mathbf{\Lambda}^{-1})}}. \tag{38}$$

From the power allocation policy, we see no subchannel is dropped, regardless of the channel realization. Furthermore, more power is alloted to weaker modes and less power to stronger modes so that the subchannel SNRs and MSEs are equal. It can be verified that $\mathbf{GHF} = (\gamma)/(\gamma+1)\mathbf{I}$. Hence, the optimum design transforms the MIMO channel matrix into a scaled identity matrix. We will illustrate later on the differences between MMSE and equal-error designs through numerical examples.

VI. Numerical Examples

In this section, we provide numerical examples to illustrate the specialized designs obtained from the weighted MMSE criterion. For our Monte Carlo simulations, we assume that the total transmit power *across all transmit antennas* is normalized to unity $(p_0 = 1)$. We assume a MIMO channel without delay spread and with uncorrelated noise across receiver antennas $(\mathbf{R}_{\mathbf{nn}} \sim \sigma^2\mathbf{I})$. The elements of the MIMO channel matrix $\mathbf{H}$ are obtained from an i.i.d. complex Gaussian distribution with mean 0 and variance 1. Each realization of $\mathbf{H}$ is assumed known at the transmitter and receiver, and the linear precoder and decoder are optimized for each channel realization.

Finally, all numerical results are plotted as a function of the ratio of total transmit power (summed across all transmit antennas) to the noise variance at each receive antenna.

[6]A similar design was also obtained in [15], albeit using a different optimization criterion and in a different context.

TABLE I
Maximum Information Rate Design

i	λ_i	$\phi_{f,i}^2$	Γ_i (dB)	$\mathcal{R}_i$ (bps/hz)	M_i
1	300	0.22	18.2	6.1	64
2	100	0.21	13.28	4.5	16
3	60	0.21	10.92	3.7	8
4	30	0.19	7.54	2.7	4
5	20	0.17	5.38	2.1	4

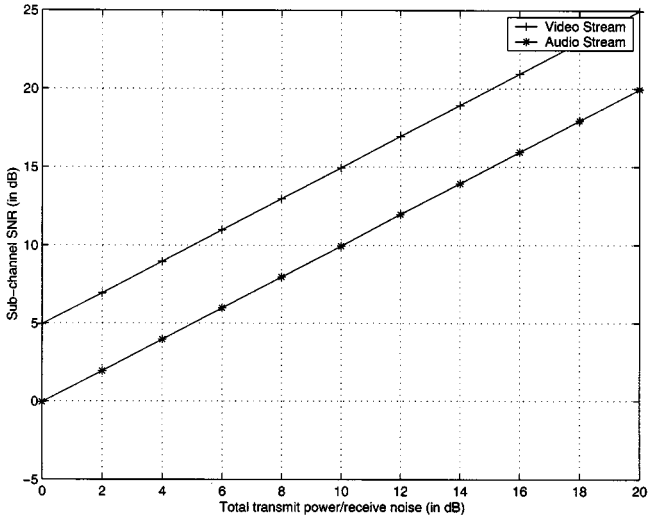

Fig. 4. QoS-based design.

Example 1: Maximum Information Rate Design: Consider the maximum information rate precoder and decoder design for a 5×5 spatial multiplexing system. The precoder allocates power on the 5 eigen subchannels (i) using the water-filling solution. Table I shows an example subchannel gain (λ_i), transmit power $(\phi_{f,i}^2)$, SNR (Γ_i), and the resulting data-rate $(\mathcal{R}_i)$ for a fixed and known channel realization. The signaling QAM constellation on each subchannel is given by $M_i = 2^{\mathrm{floor}(\mathcal{R}_i)}$.

Example 2: QoS Based Design: Consider a 3×3 MIMO spatial multiplexing system over which we want to transmit $B = 2$ independent symbol streams—an audio stream and a video stream. Typically, audio transmission has low data rates and can tolerate a fairly high uncoded BER, while video transmission has high data rates and requires a lower uncoded BER. From (24), it then follows that video requires a much higher SNR than audio, say, for example, 5 dB (for the sake of illustration). Fig. 4 illustrates the performance of the QoS-based design, which indeed guarantees 5-dB higher (receive) SNR for the video stream. Our results are obtained by averaging over 10 000 channel realizations. The optimal linear precoder and decoder are optimized for each channel realization.

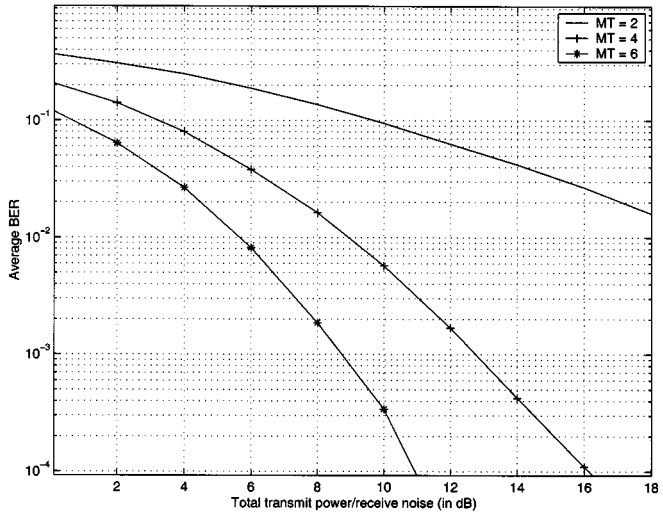

Fig. 5. Performance improvement with MMSE design.

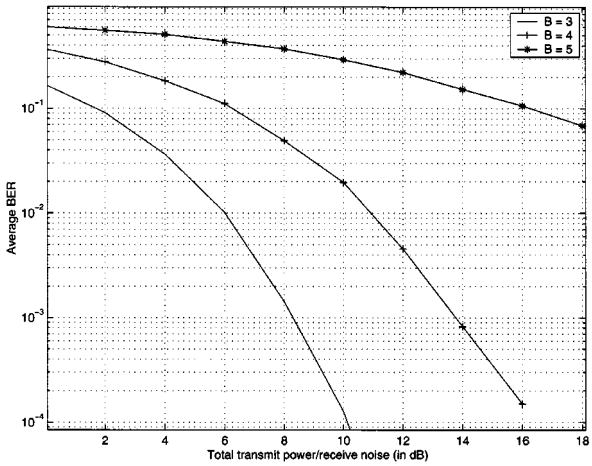

Fig. 6. Equal-error design.

Example 3: (Unweighted) MMSE Design: Consider an application where we want to send $B = 2$ streams of data with 4-QAM modulation, over an $M_T \times 2$ spatial multiplexing system $(M_T \geq 2)$. Fig. 5 illustrates the performance improvement by using an MMSE design, as the number of transmit antennas increase. The performance improvement can be attributed to the increase in transmit diversity and array gain with the increase in number of transmit antennas. The results were obtained by averaging the BERs over 500 000 channel realizations. The optimal linear precoder and decoder are optimized for each channel realization.

Example 4: Equal Error Design: Let us suppose that we need to transmit $B <= 5$ independent streams of data using 4-QAM modulation, reliably over a 5×5 spatial multiplexing system. We adopt the equal-error design that guarantees that each of the B data streams are transmitted with equal errors, for any channel realization. Fig. 6 shows the performance of equal-error design, for different values of B, averaged over 500 000 channel realizations. The optimum choice of B is dictated by the BER requirement. We see that as B increases (higher multiplexing gain), the average BER increases, and vice versa. This can be explained by the decrease in the array gain and transmit diversity gain experienced by each stream, as B is increased.

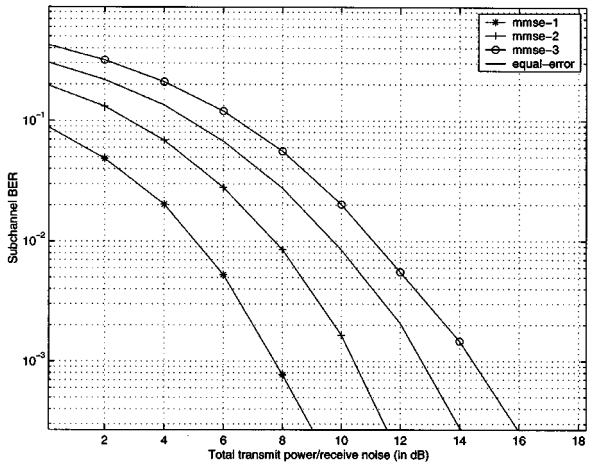

Fig. 7. Comparison of error–error and MMSE designs: subchannel BERs.

This example illustrates the tradeoff between multiplexing and diversity gain.

Example 5: Comparison of Equal-Error and MMSE Design: Consider a 4×4 spatial multiplexing system over which we transmit $B = 3$ independent symbol streams with 4-QAM modulation. We compare the performances of the equal-error and MMSE designs for such a system. It can be verified that the equal-error and MMSE designs have similar total average BERs. However, the main difference lies in the subchannel BER for each of the three streams, as shown in Fig. 7. The equal-error design guarantees equal BER for all three streams, unlike the MMSE design. In the latter case, some of the symbol streams have higher BER than the equal-error design and some have lower BER.

VII. Conclusion

We derived the optimum linear precoder and decoder for MIMO channels with and without delay spread, using the weighted MMSE criterion, subject to a transmitter power constraint. We showed that the optimum precoder and decoder diagonalize the MIMO channel into eigen subchannels for any set of error weights. Only the power alloted to each subchannel need be recomputed for each application.

Suggestions for Future Research: The performance gain possible by using linear precoding in lieu of or in conjunction with space–time codes needs to be quantified. Furthermore, performance degradation in the presence of channel estimation errors needs to be analyzed. Finally, the question of how to modify our designs in the case of partial channel knowledge (such as knowledge of only the first- and second-order statistics of the channel) at the transmitter is an open research topic.

Appendix I

A. Proof of Lemma 1

Consider the most general expressions for $\mathbf{F}$ and $\mathbf{G} = \bar{\mathbf{G}}\mathbf{R}_{\mathbf{nn}}^{-1/2}$

$$\mathbf{F} = \mathbf{V}\mathbf{\Phi}_{\mathbf{f}} + \tilde{\mathbf{V}}\tilde{\mathbf{\Phi}}_{\mathbf{f}} = \mathbf{F}_{\|} + \mathbf{F}_{\perp} \tag{39}$$

$$\bar{\mathbf{G}} = \mathbf{\Phi}_{\mathbf{g}}\mathbf{V}^{*}\mathbf{H}^{*}\mathbf{R}_{\mathbf{nn}}^{-1/2} + \tilde{\mathbf{\Phi}}_{\mathbf{g}}\tilde{\mathbf{V}}^{*}$$

$$\mathbf{H}^{*}\mathbf{R}_{\mathbf{nn}}^{-1/2} = \mathbf{G}_{\|} + \mathbf{G}_{\perp} \tag{40}$$

where $\mathbf{F}_{||}$ and $\mathbf{G}_{||}$ are each in the range space of $\mathbf{R_{nn}^{-1/2}H}$, and $\mathbf{F}_{\perp}$ and $\mathbf{G}_{\perp}$ are each in the null space of $\mathbf{R}_{\mathbf{nn}}^{-1/2}\mathbf{H}$; $\tilde{\mathbf{\Phi}}_{\mathbf{f}}$ is any $(M_T - B) \times B$ matrix, and $\tilde{\mathbf{\Phi}}_{\mathbf{g}}$ is any $B \times (M_T - B)$ matrix. Note that the above decomposition for $\mathbf{G}$ is valid since $\mathbf{R}_{\mathbf{nn}}^{-1/2}$ is square and full-rank, and imposes no restrictions in our search for optimal $\mathbf{G}$. Furthermore, we have

$$\mathbf{F}_{\perp}^{*}\mathbf{F}_{||} = 0, \quad \mathbf{F}_{||}^{*}\mathbf{F}_{\perp} = 0, \quad \mathbf{R}_{\mathbf{nn}}^{-1/2}\mathbf{H}\mathbf{F}_{\perp} = 0 \tag{41}$$

$$\mathbf{G}_{\perp}\mathbf{G}_{||}^{*} = 0, \quad \mathbf{G}_{||}\mathbf{G}_{\perp}^{*} = 0, \quad \mathbf{G}_{\perp}\mathbf{R}_{\mathbf{nn}}^{-1/2}\mathbf{H} = 0. \tag{42}$$

Substituting the pair $(\mathbf{G}\bar{\mathbf{R}}_{\mathbf{nn}}^{-1/2}, \mathbf{F})$ into (16) and (17), respectively, we get

$$\mathbf{HF} = \mathbf{HFF}^{*}\mathbf{H}^{*}\mathbf{R}_{\mathbf{nn}}^{-1/2}\bar{\mathbf{G}}^{*} + \mathbf{R}_{\mathbf{nn}}^{1/2}\bar{\mathbf{G}}^{*} \tag{43}$$

$$\mathbf{WG}\bar{\mathbf{R}}_{\mathbf{nn}}^{-1/2}\mathbf{H} = \mathbf{F}^{*}\mathbf{H}^{*}\mathbf{R}_{\mathbf{nn}}^{-1/2}\bar{\mathbf{G}}^{*}\mathbf{W}\bar{\mathbf{G}}\mathbf{R}_{\mathbf{nn}}^{-1/2}\mathbf{H} + \mu\mathbf{F}^{*}. \tag{44}$$

Premultiply (43) by $\mathbf{G}_{\perp}\mathbf{R}_{\mathbf{nn}}^{-1/2}$ and use (42) to get $\bar{\mathbf{G}}_{\perp}\bar{\mathbf{G}}^{*} = 0 \Rightarrow \bar{\mathbf{G}}_{\perp}\bar{\mathbf{G}}_{\perp}^{*} = 0 \Rightarrow \bar{\mathbf{G}}_{\perp} = 0$. Post-multiply (44) by $\mathbf{F}_{\perp}$ and use (41) to get $\bar{\mathbf{F}}_{\perp}^{*}\mathbf{F} = 0 \quad \Rightarrow \bar{\mathbf{F}}_{\perp}^{*}\mathbf{F}_{\perp} = 0 \quad \Rightarrow \bar{\mathbf{F}}_{\perp} = 0$. Using $\bar{\mathbf{F}}_{\perp} = \bar{\mathbf{G}}_{\perp} = 0$ in (39) and (40), we achieve the desired result.

B. Proof of Lemma 2

Premultiplying (16) by $\mathbf{G}$ and post-multiplying (17) by $\mathbf{F}$, and using the expressions for $\mathbf{F}$ and $\mathbf{G}$ from Lemma 1 in the so-obtained equations, we obtain

$$\begin{aligned}\mathbf{\Phi_g V^{*}H^{*}R_{nn}^{-1}HV\Phi_f} &= \mathbf{\Phi_g V^{*}H^{*}R_{nn}^{-1}HV\Phi_f\Phi_f^{*}V^{*}H^{*}R_{nn}^{-1}HV\Phi_g^{*}} \\ &\quad + \mathbf{\Phi_g V^{*}H^{*}R_{nn}^{-1}HV\Phi_g^{*}}\end{aligned} \tag{45}$$

$$\begin{aligned}\mathbf{W\Phi_g V^{*}H^{*}R_{nn}^{-1}HV\Phi_f} &= \mathbf{\Phi_f^{*}V^{*}H^{*}R_{nn}^{-1}HV\Phi_g^{*}W\Phi_g V^{*}H^{*}R_{nn}^{-1}HV\Phi_f} \\ &\quad + \mu\mathbf{\Phi_f^{*}V^{*}V\Phi_f}.\end{aligned} \tag{46}$$

Using the EVD of $\mathbf{H}^{*}\mathbf{R}_{\mathbf{nn}}^{-1}\mathbf{H}$, and the fact that $\mathbf{V}^{*}\mathbf{V} = \mathbf{I}$, the above equations reduce to

$$\mathbf{\Phi_g\Lambda\Phi_f} = \mathbf{\Phi_g\Lambda\Phi_f\Phi_f^{*}\Lambda\Phi_g^{*}} + \mathbf{\Phi_g\Lambda\Phi_g^{*}} \tag{47}$$

$$\mathbf{W\Phi_g\Lambda\Phi_f} = \mathbf{\Phi_f^{*}\Lambda\Phi_g^{*}W\Phi_g\Lambda\Phi_f} + \mu\mathbf{\Phi_f^{*}\Phi_f}. \tag{48}$$

Note from (47) that $\mathbf{\Phi_g\Lambda\Phi_f}$ must be Hermitian since the other terms in the equation are Hermitian. Using the same argument in (48), we conclude that $\mathbf{W\Phi_g\Lambda\Phi_f}$ must also be Hermitian. But since $\mathbf{W}$ is a diagonal matrix, this implies that $\mathbf{\Phi_g\Lambda\Phi_f}$ must be real diagonal, assuming that $\mathbf{W}$ has distinct diagonal elements. For the case when $\mathbf{W}$ has repeated diagonal elements w_{ii}, we can consider $\tilde{\mathbf{W}} = \mathbf{W} + \mathbf{\Delta_W}$ where $\mathbf{\Delta_W}$ is a perturbation matrix that ensures that w_{ii} are distinct. It can be observed from (16) and (17) that $\mathbf{G}$ and $\mathbf{F}$ are continuous functions[7] of $\mathbf{W}$ and hence $\lim_{\mathbf{\Delta_W}\to 0}\mathbf{G}(\tilde{\mathbf{W}}) = \mathbf{G}(\mathbf{W})$ and similarly for $\mathbf{F}$. This means that we can take $\mathbf{\Phi_g\Lambda\Phi_f}$ to be real diagonal for any $\mathbf{W}$. Note that the case of $\mathbf{W} = \mathbf{I}$ has been well studied in the literature [15], [18] where the fact that $\mathbf{\Phi_g\Lambda\Phi_f}$ is diagonal was shown to be true. Next, we use $\mathbf{D_1}, \mathbf{D_2}$, etc., to denote real-valued diagonal matrices of appropriate dimensions. We showed above that

$$\mathbf{\Phi_g\Lambda\Phi_f} = \mathbf{D_1}. \tag{49}$$

Then, it follows from (47) and (48) that $\mathbf{\Phi_g\Lambda\Phi_g^{*}} = \mathbf{D_2} \geq 0$; $\mathbf{\Phi_f^{*}\Phi_f} = \mathbf{D_3} \geq 0$ or equivalently

$$\mathbf{\Phi_g^{*}} = \mathbf{\Lambda}^{-1/2}\mathbf{U_g}\mathbf{D_2}^{1/2} \tag{50}$$

$$\mathbf{\Phi_f} = \mathbf{U_f}\mathbf{D_3}^{1/2} \tag{51}$$

where $\mathbf{U_g}$ and $\mathbf{U_f}$ are unitary matrices that are arbitrary for now and the symbol ≥ 0 denotes a positive semidefinite matrix. Inserting (50) into (49) yields $\mathbf{D_2}^{1/2}\mathbf{U_g^{*}\Lambda}^{1/2}\mathbf{U_f D_3}^{1/2} = \mathbf{D_1}$ or, equivalently

$$\mathbf{U_g^{*}\Lambda}^{1/2}\mathbf{U_f} = \mathbf{D_4}. \tag{52}$$

In particular, (52) implies that

$$\mathbf{U_g^{*}\Lambda U_g} = \mathbf{D_4^2}; \quad \mathbf{U_f^{*}\Lambda U_f} = \mathbf{D_4^2}. \tag{53}$$

In the case the eigenvalues in $\mathbf{\Lambda}$ are distinct, (53) along with (52) and the fact that $\mathbf{U_g}$ and $\mathbf{U_f}$ are unitary give

$$\mathbf{U_g} = \mathbf{U_f} = \mathrm{diag}([e^{j\Theta_1}, \ldots, e^{j\Theta_b}]) = \mathbf{\Theta} \tag{54}$$

for arbitrary $\Theta_k \epsilon [0, 2\pi]$. The case of repeated eigenvalues in $\mathbf{\Lambda}$ can be handled by means of a perturbation argument similar to the one used in the analysis of (47) and (48) to obtain (49). Hence, we conclude that (54) holds in the general case. Using (54) in (50) yields

$$\mathbf{\Phi_g} = \mathbf{\Theta}^{*}\mathbf{D_2}^{1/2}\mathbf{\Lambda}^{-1/2}; \quad \mathbf{\Phi_f} = \mathbf{D_3}^{1/2}\mathbf{\Theta}. \tag{55}$$

It is readily checked that the $\mathbf{\Theta}$ in (55) does not affect the Lagrangian function in (11) in anyway. Accordingly, setting $\mathbf{\Theta} = \mathbf{I}$ in (55) imposes no restriction in our optimization search. With this observation, the proof is concluded.

C. Proof of Theorem 1

Substituting (55) with $\mathbf{\Theta} = \mathbf{I}$ into (47), (48), and (49), we obtain

$$\mathbf{D_1} = \mathbf{D_1^2} + \mathbf{D_2} \tag{56}$$

$$\mathbf{WD_1} = \mathbf{WD_1^2} + \mu\mathbf{D_3} \tag{57}$$

$$\mathbf{D_2 D_3 \Lambda} = \mathbf{D_1^2}. \tag{58}$$

From (56) and (57), we obtain, respectively,

$$\mathbf{D_2} = \mathbf{D_1}(\mathbf{I} - \mathbf{D_1}) \quad \mu\mathbf{D_3} = \mathbf{WD_1}(\mathbf{I} - \mathbf{D_1}). \tag{59}$$

Inserting (59) into (58), we obtain $\mathbf{WD_1^2}(\mathbf{I} - \mathbf{D_1})^2\mathbf{\Lambda} = \mu\mathbf{D_1^2}$ or, equivalently, the solution (assuming $(\mathbf{D_1} > 0)$

$$\mathbf{D_1} = (\mathbf{I} - \mu^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{-1/2})_{+}. \tag{60}$$

Using (60) in (59) yields, for $\mu > 0$, $\mathbf{D_2} = (\mu^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{-1/2} - \mu\mathbf{\Lambda}^{-1}\mathbf{W}^{-1})_{+}$ and $\mathbf{D_3} = \mathbf{W}(\mu^{-1/2}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{-1/2} - \mathbf{\Lambda}^{-1}\mathbf{W}^{-1})_{+}$. The above expressions for $\mathbf{D_2}$ and $\mathbf{D_3}$ along with (55) (with $\mathbf{\Theta} = \mathbf{I}$) lead to (22) and (21), which concludes the proof for the case when $\mathbf{D_1} > 0$.

[7]Note that (16), (17) are obtained from the KKT conditions only, and do not assume any *a priori* structure on $\mathbf{G}$ and $\mathbf{F}$.

ACKNOWLEDGMENT

The authors would like to thank Dr. A. Scaglione from the University of New Mexico for reviewing the manuscript and providing very valuable comments and Dr. P. Viswanath and Dr. H. Bolcskei from the University of Illinois at Urbana Champaign for stimulating discussions on the topic of weighted MMSE designs.

REFERENCES

[1] D. Agarwal, V. Tarokh, A. Naguib, and N. Seshadri, "Space–time coded OFDM for high data-rate wireless communications over wide band channels," in *Proc. 48th IEEE Vehicular Technology Conf.*, vol. 3, May 1998.

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

[3] H. Bolcskei, D. Gesbert, and A. Paulraj, "On the capacity of OFDM-based multi-antenna systems," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, 2000.

[4] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY: Wiley, 1991.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, pp. 41–59, Autumn 1996.

[6] A. Goldsmith and S. G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 595–602, May 1998.

[7] W. Jang, B. R. Vojcic, and R. L. Pickholtz, "Joint transmitter-receiver optimization in synchronous multiuser communications over multi path channels," *IEEE Trans. Commun.*, vol. 46, pp. 269–278, Feb. 1998.

[8] H. Lutkepohl, *Handbook of Matrices*. New York: Wiley, 1996.

[9] A. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)," U.S. Patent 5 345 599, 1993.

[10] J. G. Proakis, *Digital Communications*, 2nd ed. New York, NY: McGraw-Hill, 1989.

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

[12] H. Sampath and A. Paulraj, "Joint transmit and receive optimization for high data rate wireless communications using multiple antennas," in *Proc. Asilomar Conf. Signals, Systems and Computers*, vol. 1, 1999.

[13] H. Sampath, "Linear precoder and decoder designs for multiple-input multiple-output wireless systems," Ph.D. dissertation, Stanford University, Stanford, CA, Apr. 2001.

[14] H. Sampath, P. Stoica, and A. Paulraj, "A generalized space-time linear precoder and decoder design using the weighted MMSE criterion," in *Conf. Rec. Thirty-Fourth Asilomar Conf. on Signals, Systems and Computers*, vol. 1, 2000, pp. 753–758.

[15] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filter bank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1019–1032, Apr. 1999.

[16] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filter bank precoders and equalizers part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, pp. 1958–2022, July 1999.

[17] A. Scaglione, P. Stoica, G. B. Giannakis, S. Barbarossa, and H. Sampath, "Optimal designs for space-time linear precoders and equalizers," *IEEE Trans. Signal Processing*, to be published.

[18] E. Telatar, "Capacity of multi-antenna Gaussian channels," ATT Bell Laboratories, Tech. Memo., 1995.

[19] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, pp. 3221–3231, Dec. 1994.

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

[21] V. Tarokh, A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1121–1128, May 1999.

[22] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 451–460, Mar. 1999.

Hemanth Sampath (M'01) received the B.S.E.E. degree (honors) from the University of Maryland, College Park, in 1996 and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1998 and 2001, respectively.

Since July 2000, he has been working as a Senior Member of Technical Staff at Iospan Wireless Inc., San Jose, CA. His research interests include wireless communications and signal processing for MIMO systems.

Petre Stoica (SM'91–F'94) received the D.Sc. degree in automatic control from the Bucharest Polytechnic Institute, Bucharest, Romania, in 1979 and an Honorary Doctorate degree in science from Uppsala University (UU), Uppsala, Sweden, in 1993.

He is a Professor of System Modeling with the Department of Systems and Control at UU. His main scientific interests are in the areas of system identification, time series analysis and prediction, statistical signal and array processing, spectral analysis, wireless communications, and radar signal processing. He has published seven books, ten book chapters, and some 450 papers in archival journals and conference records on these topics. He is on the editorial boards of five journals in the field of signal processing.

Dr. Stoica was the co-recipient of the 1989 ASSP Society Senior Award, recipient of the 1996 Technical Achievement Award of the IEEE Signal Processing Society, co-recipient of the 1998 EURASIP Best Paper Award for signal processing, a 1999 IEEE Signal Processing Society Best Paper Award, a 2000 IEEE Third Millennium Medal, and the 2000 W. R. G. Baker Paper Prize Award. He is an honorary member of the Romanian Academy and a fellow of the Royal Statistical Society.

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

He has been a Professor of Electrical Engineering at Stanford University, Stanford, CA, since 1992, where he leads a large group in wireless communications. Prior to that, he has directed several research laboratories in India and won a number of awards for his contributions to the technology development in India. His research has spanned several disciplines, emphasizing signal processing, parallel computer architectures/algorithms, and communication systems. He is the author of over 250 research papers and 10 patents. In 1999, he founded Iospan Wireless Inc., San Jose, CA, to develop broad-band wireless access systems exploiting concepts initially developed at Stanford University. He serves on the board of directors/advisory panels for a few U.S. and Indian companies/venture capital partnerships.