

# Optimal Designs for Space-Time Linear Precoders and Decoders

Anna Scaglione, *Member, IEEE*, Petre Stoica, *Fellow, IEEE*, Sergio Barbarossa, *Member, IEEE*,  
Georgios B. Giannakis, *Fellow, IEEE*, and Hemanth Sampath, *Member, IEEE*

**Abstract**—In this paper, we introduce a new paradigm for the design of transmitter space-time coding that we refer to as linear precoding. It leads to simple closed-form solutions for transmission over frequency-selective multiple-input multiple-output (MIMO) channels, which are scalable with respect to the number of antennas, size of the coding block, and transmit average/peak power. The scheme operates as a block transmission system in which vectors of symbols are encoded and modulated through a linear mapping operating jointly in the space and time dimension. The specific designs target minimization of the symbol mean square error and the approximate maximization of the minimum distance between symbol hypotheses, under average and peak power constraints. The solutions are shown to convert the MIMO channel with memory into a set of parallel flat fading subchannels, regardless of the design criterion, while appropriate power/bits loading on the subchannels is the specific signature of the different designs. The proposed designs are compared in terms of various performance measures such as information rate, BER, and symbol mean square error.

## I. INTRODUCTION

**D**IVERSITY techniques have been widely studied in the past, although attention was focused on receive diversity. A proliferation of new diversity techniques and optimal space-time coding has followed the recognition of the possible benefits of transmit diversity [1], [32]. At the same time, the demand for high-speed wireless links and the progressive saturation of the radio frequency (RF) bands contributed to the idea of using multiple transmit and receive antennas to increase channel capacity [8], [11], [22]. However, to achieve the information rate and/or the diversity gain afforded by the increased hardware complexity, appropriate precoding and modulation techniques are necessary. Two main approaches emerged from the effort of defining such effective transmission strategies: One uses appropriate mappings of the information symbols in space and time so that, without channel state information (CSI) at the transmitter and with low complexity at the receiver, full

diversity gains become possible (e.g., [1], [23], [32], [33]); the second one addresses specifically the optimization of the information rate in the case of flat fading [8], [11], [22] and frequency-selective channels [24], [25], assuming that CSI is available at both the transmitter and receiver sides. Optimal designs developed in the past, which were based on multi-input multi-output (MIMO) models such as [4], [17], [18], and [26], gained importance because of the new interest in joint transmit-receive diversity schemes. The optimization of pre and post-filters was considered in [21] for a MIMO system distorted by additive noise only. The design paradigm we adopt in this paper is based on an optimal pair of linear transformations  $\mathbf{F}$  (precoder) and  $\mathbf{G}$  (decoder) of blocks of the transmit symbols and receive samples, respectively, that *operate jointly* and *linearly* on the time and space dimensions. The designs target different criteria of optimality and constraints, assuming the channel is known at the receiver as well as at the transmitter end. CSI can be acquired at the transmitter either if a feedback channel is present or when the transmitter and receiver operate in time division duplex (TDD) so that the time-invariant MIMO channel transfer function is the same in both ways. Our linear optimal solutions can appropriately take advantage of the CSI and utilize resources at best while maintaining a reasonable complexity. They also bring additional flexibility into the design that the alternative approaches in [1], [23], [32], and [33] do not have for the latter impose restrictions on the number of antennas for which the algorithms can be implemented.

In all our designs, the paradigm of linear precoding/decoding exploits the channel eigendecomposition in constructing the optimal  $\mathbf{F}$ ,  $\mathbf{G}$ . The distinct solutions are characterized by how the power is loaded on each channel eigenfunction. However, unlike [14] and [16], instead of assuming vector coding, we will show how optimal linear transceivers naturally result in having their specific structure from the criterion. Some designs considered in this paper extend the results of [29] and [30] to the case where space diversity is present and, as in [29] and [30], we make no assumptions regarding the noise color or the channel matrix size and rank, which are arbitrary. The different solutions will be compared in order to underline the tradeoffs or advantages implied by the various designs. Special emphasis will be put on the design that targets minimization of the MSE of the decoded block through the minimization of the trace or the determinant of the error covariance matrix. The first design corresponds to the MMSE criterion, whereas the second one is proved to be equivalent to the maximization of the mutual information between transmitter and receiver. The MMSE design constitutes a valid alternative to the design that maximizes the information

Manuscript received October 19, 2000; revised December 20, 2001. The associate editor coordinating the review of this paper and approving it for publication was Prof. Michail K. Tsatsanis.

A. Scaglione is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA.

P. Stoica is with the Dept. of Systems and Control, Information Technology, Uppsala University, Uppsala, Sweden.

S. Barbarossa is with the Infocom Department, University of Rome "La Sapienza," Rome, Italy.

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA.

H. Sampath is with Iospan Wireless, Inc., San Jose, CA 95134 USA.

Publisher Item Identifier S 1053-587X(02)03281-6.

rate because the former provides satisfactory bit error rates and does not require bit loading.

The paper first introduces the system model in Section II. The joint transmit and receive optimal linear designs are derived in Section III, and their performances are evaluated in closed form in Section IV, after showing that all optimal designs convert the channel into a set of independent flat fading subchannels. Finally, in Section V, numerical examples are provided that highlight some salient aspects of our space-time linear precoding alternatives.

*Notation:* All boldface letters indicate vectors (lower case) or matrices (upper case). The  $\text{tr}(\mathbf{A})$ ,  $|\mathbf{A}|$ ,  $\lambda(\mathbf{A})$  are the trace, determinant, and eigenvalues of matrix  $\mathbf{A}$ ,  $\mathbf{a} = \text{vec}(\mathbf{A})$  is the column vector formed stacking the columns of  $\mathbf{A}$ , and the inverse operation is denoted by  $\mathbf{A} = \text{vec}^{-1}(\mathbf{a}, N)$ , where  $N$  is the number of columns of  $\mathbf{A}$ . Continuous-time multidimensional signals are written as  $\mathbf{a}(t)$ , discrete time multidimensional sequences as  $\mathbf{a}[n]$ , and sequences of vectors obtained by stacking consecutive blocks are characterized by a suffix  $(\ )_i$ , for example,  $\mathbf{a}_i = \text{vec}([\mathbf{a}[iM], \dots, \mathbf{a}[iM + M - 1]])$ .

## II. SYSTEM MODEL

The system considered has  $K$  transmit and  $R$  receive antennas. The baseband equivalent transmitted signal is the vector  $\mathbf{x}(t) := (x_1(t), \dots, x_K(t))^T$  of complex envelopes emitted by the transmit antennas. We assume a digital link with linear modulation so that the vector  $\mathbf{x}(t)$  is related to the (coded) symbol vector  $\mathbf{x}[n]$  by

$$\mathbf{x}(t) = \sum_{n=-\infty}^{+\infty} \mathbf{x}[n]g_T(t - nT) \quad (1)$$

where  $g_T(t)$  is the transmit pulse, and  $1/T$  is the rate with which the data  $\mathbf{x}[n]$  are transmitted. Correspondingly,  $\mathbf{z}(t) = \mathbf{y}(t) + \mathbf{n}(t)$  is the received  $R \times 1$  vector that contains the channel output  $\mathbf{y}(t)$  and additive noise  $\mathbf{n}(t)$ . For a linear (generally time-varying) channel, the input-output (I/O) relationship can be cast in the form

$$\mathbf{y}(t) = \iint g_R(t - \theta)\mathbf{H}(\theta, \tau)\mathbf{x}(\theta - \tau) d\tau d\theta \quad (2)$$

where  $g_R(t)$  is the impulse response of the lowpass receive filter (usually a square-root raised cosine filter) matched to the transmit filter  $g_T(t)$ , and the  $(k, l)$ th entry of matrix  $\mathbf{H}(\theta, \tau)$  is the impulse response of the channel between the  $l$ th transmit and the  $k$ th receive antennas. The received noise-free signal vector is thus

$$\mathbf{y}(t) = \sum_{n=-\infty}^{+\infty} \left[ \iint \mathbf{H}(\theta, \tau) \times g_R(t - \theta)g_T(\theta - \tau - nT) d\tau d\theta \right] \mathbf{x}[n]. \quad (3)$$

Introducing

$$\mathbf{H}[k, k - n] := \iint \mathbf{H}(\theta, \tau)g_T(\theta - \tau - nT)g_R(kT - \theta) d\tau d\theta \quad (4)$$

the vector of received samples  $\mathbf{y}[k] := \mathbf{y}(kT)$  is

$$\mathbf{y}[k] = \sum_{n=-\infty}^{\infty} \mathbf{H}[k, k - n]\mathbf{x}[n]. \quad (5)$$

If the channel discrete-time time-varying impulse response  $\mathbf{H}[k, n]$  is causal and has finite memory  $L$ , we can write the I/O relationship (5) in block FIR form. Specifically, stacking  $P = M + L$  transmit snapshots in a  $PK \times 1$  vector  $\mathbf{x}_i := \text{vec}([\mathbf{x}[iP], \dots, \mathbf{x}[iP + P - 1]])$  and  $M$  received snapshots in a  $MR \times 1$  vector  $\mathbf{y}_i := \text{vec}([\mathbf{y}[iP + L], \dots, \mathbf{y}[iP + P - 1]])$ , where we eliminated the first  $L$  vectors to cancel the interblock interference (IBI), we have

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}_i \quad (6)$$

where  $\mathbf{H}$  is an  $RM \times KP$  block-banded matrix. Alternatively, defining  $\mathbf{x}_i := \text{vec}([\mathbf{x}[iP], \dots, \mathbf{x}[iP + M - 1]])$  and  $\mathbf{y}_i := \text{vec}([\mathbf{y}[iP], \dots, \mathbf{y}[iP + P - 1]])$  and padding with  $L$  zero samples the tail of every block  $\mathbf{x}_i$ , we could have written an equation analogous to (6) but with  $\mathbf{H}$  of dimension  $RP \times KM$ . For simplicity, we will assume that IBI is removed at the receiver, and thus,  $\mathbf{H}$  is  $RM \times KP$ , but most of the derivations in the following are valid in both cases if one replaces  $M$  by  $P$ , and vice versa.

We will precode  $N \times 1$  vectors of symbols  $\mathbf{s}_i$  as

$$\mathbf{x}_i = \mathbf{F}\mathbf{s}_i \quad (7)$$

where  $N \leq \min(KP, RM)$ . Since  $N$  symbols will be embedded in  $\mathbf{x}_i$  through the precoder  $\mathbf{F}$ , it will take  $P = (M + L)$  snapshots to transmit  $N$  information symbols. If  $T$  is the time necessary to transmit one snapshot, the throughput is

$$\frac{N}{PT} \leq \frac{\min(K(M + L), RM)}{(M + L)T} \xrightarrow{M \gg L} \frac{\min(K, R)}{T}. \quad (8)$$

Therefore, by increasing  $M$ , the upper bound on the information rate is not limited by the overhead of  $L$  snapshots added to  $\mathbf{x}_i$  to avoid IBI.

If the channel is also time invariant (LTI), i.e.,  $\mathbf{H}[n, l] \equiv \mathbf{H}[l]$ , where  $\{\mathbf{H}[l]\}_{r,k}$  is the  $l$ th sample of the impulse response characterizing the channel between the  $k$ th transmit element and the  $r$ th receive element, then  $\mathbf{H}$  in (6) becomes a block Toeplitz matrix.

Although our designs are valid for any  $\mathbf{H}$ , in the case of time-varying channels, the assumption of knowledge of the CSI at the transmitter is not realistic, unless the channel can be considered time invariant for a sufficiently long interval or it can be modeled resorting to a few approximately invariant parameters that can be estimated and used to predict the channel evolution with sufficient accuracy [2].

## III. OPTIMAL LINEAR DESIGNS

We assume that the  $N \times 1$  vectors of symbols  $\mathbf{s}_i$  transmitted every  $PT$  sec. satisfy the following.

**a0)** The size  $N$  of the block  $\mathbf{s}_i$  of encoded symbols satisfies  $N \leq \text{rank}(\mathbf{H})$ . Depending on the adoption of the null guard at the transmitter or at the receiver, we have, respectively, that  $\text{rank}(\mathbf{H}) \leq \min(RM, KP)$  or  $\text{rank}(\mathbf{H}) \leq \min(RP, KM)$ .

The received signal  $\mathbf{z}_i = \mathbf{y}_i + \mathbf{n}_i$  contains noise that we model as additive Gaussian noise (AGN) with covariance  $\mathbf{R}_{nn}$ . We will use the notation  $\mathbf{R}_{ss} := E\{\mathbf{s}_i \mathbf{s}_i^H\}$  and assume the following.

**a1)** The transmit symbols are white, i.e.,  $\mathbf{R}_{ss} = \sigma_{ss}^2 \mathbf{I}$ , the noise  $\mathbf{n}_i$  is Gaussian with covariance  $\mathbf{R}_{nn}$ , the noise covariance matrix  $\mathbf{R}_{nn}$  is positive definite, and  $\mathbf{n}_i$  and  $\mathbf{s}_i$  are uncorrelated.

If  $\mathbf{R}_{ss} \neq \sigma_{ss}^2 \mathbf{I}$ , a prewhitening operation can be performed over the symbol blocks prior to precoding, and the corresponding inverse operation can be performed after the decoder  $\mathbf{G}$ . Assumption **a0)** is necessary to guarantee symbol recovery because it is otherwise impossible to invert the channel with a linear equalizer  $\mathbf{G}$ , which operates on a single block of data only, even in the absence of noise and with perfect knowledge of  $\mathbf{H}$ . Note that if the channel is LTI,  $\text{rank}(\mathbf{H}) < \min(KP, RM)$  occurs whenever the channels between the  $K$  transmitters and the  $R$  receivers have common zeros [13, p. 142]. Compared with the single antenna case of  $R = K = 1$ , we can gain in terms of rate because the  $M$  rows of  $\mathbf{H}$  that correspond to a specific channel are linearly independent (they form a Toeplitz sub-matrix), and hence,  $\text{rank}(\mathbf{H}) \geq M$ . The challenge is to define an appropriate encoding strategy that will lead to satisfactory performance for whatever diversity the channel is able to provide. As already stated, the scheme we consider here uses as a precoder a linear mapping  $\mathbf{F}$ , as in (7). The receiver performs an appropriate inverse mapping  $\mathbf{G}$  on the vector  $\mathbf{z}_i = \text{vec}(z[iP + L], \dots, z[iP + P - 1])$ ,<sup>1</sup> estimating the symbols as  $\hat{\mathbf{s}}_i = \mathbf{G}\mathbf{z}_i$ . From (6), we have

$$\hat{\mathbf{s}}_i = \mathbf{G}\mathbf{z}_i = \mathbf{G}\mathbf{H}\mathbf{F}\mathbf{s}_i + \mathbf{G}\mathbf{n}_i. \quad (9)$$

A reasonable criterion to design a linear receiver  $\mathbf{G}$ , for given  $\mathbf{F}$  and  $\mathbf{H}$ , is to minimize the mean square error (MSE) matrix that is given by

$$E\{(\hat{\mathbf{s}}_i - \mathbf{s}_i)(\hat{\mathbf{s}}_i - \mathbf{s}_i)^H\} = \text{MSE}(\mathbf{F}, \mathbf{G}) \quad (10)$$

where

$$\text{MSE}(\mathbf{F}, \mathbf{G}) := (\mathbf{G}\mathbf{H}\mathbf{F} - \mathbf{I})\mathbf{R}_{ss}(\mathbf{G}\mathbf{H}\mathbf{F} - \mathbf{I})^H + \mathbf{G}\mathbf{R}_{nn}\mathbf{G}^H. \quad (11)$$

The cumulative MSE of the estimate of  $\mathbf{s}_i$  is  $E\{\|\hat{\mathbf{s}}_i - \mathbf{s}_i\|^2\} = \text{tr}(\text{MSE}(\mathbf{F}, \mathbf{G}))$ . The  $\mathbf{G}_{\text{opt}}$  that minimizes the whole  $\text{MSE}(\mathbf{F}, \mathbf{G})$  matrix is the same as the MMSE (Wiener) receiver [15], which is known to minimize the  $\text{tr}(\text{MSE}(\mathbf{F}, \mathbf{G}))$  and is given by

$$\mathbf{G}_{\text{opt}} = \mathbf{R}_{ss}\mathbf{F}^H\mathbf{H}^H(\mathbf{H}\mathbf{F}\mathbf{R}_{ss}\mathbf{F}^H\mathbf{H}^H + \mathbf{R}_{nn})^{-1}. \quad (12)$$

Here,  $\text{MSE}(\mathbf{F}, \mathbf{G}_{\text{opt}})$  is minimum in the sense that

$$\begin{aligned} \overline{\text{MSE}}(\mathbf{F}) &:= \text{MSE}(\mathbf{F}, \mathbf{G}_{\text{opt}}) (\mathbf{R}_{ss}^{-1} + \mathbf{F}^H\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}\mathbf{F})^{-1} \\ &\leq \text{MSE}(\mathbf{F}, \mathbf{G}) \end{aligned} \quad (13)$$

and the last inequality indicates that  $\forall \mathbf{G} \neq \mathbf{G}_{\text{opt}}$ , the difference  $[\text{MSE}(\mathbf{F}, \mathbf{G}) - \overline{\text{MSE}}(\mathbf{F})]$  is a positive semidefinite matrix. Because of **a1)**, we can write

$$\mathbf{G}_{\text{opt}} = \mathbf{F}^H\mathbf{H}^H(\mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H + \mathbf{R}_{nn}\sigma_{ss}^{-2})^{-1} \quad (14)$$

<sup>1</sup>The definition of  $\mathbf{z}_i$  would change consistently with the definition of  $\mathbf{y}_i$ , and zero guards are used at the transmitter side. The uniformity of notation in the two cases allows solving both problems at the same time.

and

$$\overline{\text{MSE}}(\mathbf{F}) = \sigma_{ss}^2 (\mathbf{I} + \sigma_{ss}^2 \mathbf{F}^H\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}\mathbf{F})^{-1}. \quad (15)$$

In the following, we will determine  $\mathbf{F}_{\text{opt}}$  based on different performance measures that depend on  $\overline{\text{MSE}}(\mathbf{F})$ . Without any constraint, minimization of  $\overline{\text{MSE}}(\mathbf{F})$  in (17) will lead to the trivial solution of increasing to infinity the norm of  $\mathbf{F}$ . A reasonable constraint is obtained by bounding the expected norm of the transmit vector  $E\{\|\mathbf{x}_i\|^2\} = \text{tr}(\mathbf{F}\mathbf{F}^H)\sigma_{ss}^2$ , which limits the transmit power, and thus, we will refer to it as the power constraint (PC)

$$\text{tr}(\mathbf{F}\mathbf{F}^H)\sigma_{ss}^2 = \mathcal{P}_0. \quad (16)$$

An alternative is to constrain the maximum eigenvalue of the transmit vector covariance  $\mathbf{F}\mathbf{F}^H\sigma_{ss}^2$ , which also limits the power because  $\text{tr}(\mathbf{F}\mathbf{F}^H)\sigma_{ss}^2 \leq \lambda_{\max}(\mathbf{F}\mathbf{F}^H)N\sigma_{ss}^2$ . This corresponds to

$$\lambda_{\max}(\mathbf{F}\mathbf{F}^H)\sigma_{ss}^2 = \mathcal{L}_0. \quad (17)$$

Besides limiting the transmit power, the maximum eigenvalue constraint (17) imposes a limit on the peak power of the output. In effect, the peak of the transmit signal corresponds to the maximum (in absolute value) entry of  $\mathbf{x}_i = \mathbf{F}\mathbf{s}_i$ . The constraint (17) limits the peak power because for  $\max_{i,k}(|\{\mathbf{x}_i\}_k|^2)$ , the following inequalities apply:

$$\begin{aligned} \max_{i,k}(|\{\mathbf{F}\mathbf{s}_i\}_k|^2) &\leq \max_i(\mathbf{s}_i^H\mathbf{F}^H\mathbf{F}\mathbf{s}_i) \\ &\leq \lambda_{\max}(\mathbf{F}^H\mathbf{F})\max_i(\|\mathbf{s}_i\|^2) \end{aligned} \quad (18)$$

where  $\|\mathbf{s}_i\|^2$  is bounded since  $\mathbf{s}_i$  is formed by symbols that are all bounded in amplitude. The advantage of this constraint is that it limits the signal peak, independent of the specific constellation used. The disadvantage is that the bound may not be tight.

Finally, let us introduce the following EVD:

$$\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H} = \bar{\mathbf{V}}\bar{\mathbf{\Lambda}}\bar{\mathbf{V}}^H \quad (19)$$

where  $\bar{\mathbf{V}}$  may be tall if  $\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}$  is rank deficient and  $\bar{\mathbf{\Lambda}}$  is an  $Q \times Q$  diagonal matrix, where  $Q := \text{rank}(\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}) = \text{rank}(\mathbf{H})$ . We assume the following (which is not a restriction).

**a2)** The elements  $\{\lambda_{qq}\}_{q=1}^Q$  in the diagonal of matrix  $\bar{\mathbf{\Lambda}}$ , which are the non-null eigenvalues of  $\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}$ , are arranged in decreasing order. Note that **a0)** requires  $N \leq Q$ . For convenience, we will denote by  $\mathbf{\Lambda}$  the  $N \times N$  diagonal matrix with diagonal entries  $\{\lambda_{qq}\}_{q=1}^N$  ( $\mathbf{\Lambda}$  is equal to the top left  $N \times N$  block of  $\bar{\mathbf{\Lambda}}$ ) and matrix  $\bar{\mathbf{V}}$  denoting the first  $N$  columns of  $\bar{\mathbf{V}}$ , which are the eigenvectors corresponding to the  $N$  largest eigenvalues  $\{\lambda_{qq}\}_{q=1}^N$  of  $\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}$ .

#### A. MMSE Criterion Under Power Constraint

The MMSE design minimizes the  $\text{tr}(\text{MSE}(\mathbf{G}, \mathbf{F}))$  jointly with respect to  $\mathbf{G}$  and  $\mathbf{F}$  under the transmit-power constraint. Analogous criteria formulated in the frequency domain for joint transmit/receive-filter optimization for the scalar case can be found in [3, p. 333] and for the MIMO case in [18]. The joint transmit and receive design that minimizes  $\text{tr}(\text{MSE}(\mathbf{G}, \mathbf{F}))$  can be obtained by minimizing  $\text{tr}(\overline{\text{MSE}}(\mathbf{F}))$  with respect to  $\mathbf{F}$ . The

solution for  $\mathbf{F}_{\text{opt}}$  is given in the following lemma, and  $\mathbf{G}_{\text{opt}}$  can be obtained by replacing  $\mathbf{F}$  with  $\mathbf{F}_{\text{opt}}$  in (12).

*Lemma 1:* The solution of the optimization problem

$$\mathbf{F}_{\text{opt}} = \arg \min_{\mathbf{F}} \text{tr}(\overline{\text{MSE}}(\mathbf{F})), \quad \text{tr}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H) \sigma_{ss}^2 = \mathcal{P}_0 \quad (20)$$

is given by  $\mathbf{F}_{\text{opt}} = \mathbf{V}\Phi$ , where  $\Phi$  is an  $N \times N$  diagonal matrix with the following  $(i, i)$  entry:<sup>2</sup>

$$|\phi_{ii}|^2 = \left( \frac{\mathcal{P}_0 + \sum_{n=1}^{\bar{N}} \lambda_{nn}^{-1} \lambda_{ii}^{-1/2}}{\sigma_{ss}^2 \sum_{n=1}^{\bar{N}} \lambda_{nn}^{-1/2} \lambda_{ii}^{-1/2}} - \frac{1}{\lambda_{ii} \sigma_{ss}^2} \right)^+ \quad (21)$$

where  $(x)^+ := \max(x, 0)$  and  $\bar{N} \leq N$  is such that  $|\phi_{nn}|^2 > 0$  for  $n \in [1, \bar{N}]$  and  $|\phi_{nn}|^2 = 0$  for all other  $n$ .

*Proof:* See Appendix A. ■

Note that  $\bar{N}$  is a function of the eigenvalues as well: For a given  $\mathbf{\Lambda}$  and  $N$ ,  $\bar{N}$  can be found calculating (21) iteratively starting with  $\bar{N} = N$  and, while  $|\phi_{\bar{N}, \bar{N}}|^2 < 0$ , decreasing  $\bar{N}$  progressively by one, as explained in Appendix A. Interestingly, the minimization of the determinant, in lieu of the trace, of the  $\overline{\text{MSE}}(\mathbf{F})$  matrix with respect to  $\mathbf{F}$  is equivalent to maximizing the information rate. As mentioned in the introduction, the capacity of a MIMO channel was first derived in [4] and, for the multiple-antenna and flat fading case, in [8] and [22]. In [24] and [25], the authors generalized the discrete multitone (DMT) scheme for the MIMO frequency-selective case. The optimal space-time processing for the maximization of the information rate is also derived in [9]. Compared with these works, our approach

- 1) jointly optimizes the precoder and decoder explicitly;
- 2) does not treat the frequency-selective and flat-fading cases separately and includes the time-varying case as well;
- 3) does not rely on the full rank of any of the matrices involved;
- 4) links together the MSE metric with the maximum information rate criterion.

*Lemma 2:* The solution of the optimization problem

$$\mathbf{F}_{\text{opt}} = \arg \min_{\mathbf{F}} |\overline{\text{MSE}}(\mathbf{F})|, \quad \text{tr}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H) \sigma_{ss}^2 = \mathcal{P}_0 \quad (22)$$

is given by  $\mathbf{F}_{\text{opt}} = \mathbf{V}\Phi$ , where  $\Phi$  is an  $N \times N$  diagonal matrix with  $(i, i)$  entry

$$|\phi_{ii}|^2 = \left( \frac{\mathcal{P}_0 + \sum_{k=1}^{\bar{N}} \lambda_{kk}^{-1}}{\bar{N} \sigma_{ss}^2} - \frac{1}{\lambda_{ii} \sigma_{ss}^2} \right)^+ \quad (23)$$

and  $\bar{N} \leq N$  is the number of positive  $|\phi_{ii}|^2$ .

*Proof:* See Appendix B. ■

The power loading on the eigenvectors  $\mathbf{V}$  of  $\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H}$  of Lemma 2 is identical to the so-called “water-filling” obtained from the maximization of the mutual information on parallel Gaussian channels (see, e.g., [6], [8], [12], and [22]), and in the context of linear precoding, it leads exactly to the solution described in [29, Lemma 2]. In particular, in Appendix C, we show the following.

<sup>2</sup>Note that only the amplitude of  $\phi_{ii}$  is fixed, whereas the phase is arbitrary; thus,  $\phi_{ii}$  can be a real number.

*Corollary 1:* For a Gaussian input  $\mathbf{s}_i$ , if  $\mathbf{G}$  has the following structure:

$$\mathbf{G} = \tilde{\mathbf{\Gamma}} \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \quad (24)$$

where  $\tilde{\mathbf{\Gamma}}$  is an arbitrary  $N \times N$  matrix, the mutual information  $I(\hat{\mathbf{s}}, \mathbf{x})$  per block does not depend on  $\mathbf{G}$  (see also [29]) and is

$$I(\hat{\mathbf{s}}, \mathbf{x}) = \log |\sigma_{ss}^2 \mathbf{H} \mathbf{F} \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} + \mathbf{I}|. \quad (25)$$

The  $\mathbf{F}_{\text{opt}}$  in (23) and  $\mathbf{G}_{\text{opt}}$  in (12) also maximize the mutual information between transmit and receive data.

*Proof:* See Appendix C. ■

### B. MMSE Criterion Under Maximum Eigenvalue Constraint

*Lemma 3:* The solution of the optimization problem

$$\begin{aligned} \mathbf{F}_{\text{opt}} &= \arg \min_{\mathbf{F}} \text{tr}(\overline{\text{MSE}}(\mathbf{F})) \\ \lambda_{\max}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H) \sigma_{ss}^2 &= \mathcal{L}_0 \end{aligned} \quad (26)$$

is given by  $\mathbf{F}_{\text{opt}} = \sqrt{\mathcal{L}_0 / \sigma_{ss}^2} \mathbf{V}$ .

*Proof:* See Appendix D. ■

*Lemma 4:* The solution of the optimization problem

$$\begin{aligned} \mathbf{F}_{\text{opt}} &= \arg \min_{\mathbf{F}} |\overline{\text{MSE}}(\mathbf{F})| \\ \lambda_{\max}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H) \sigma_{ss}^2 &= \mathcal{L}_0 \end{aligned} \quad (27)$$

is given by  $\mathbf{F}_{\text{opt}} = \sqrt{\mathcal{L}_0 / \sigma_{ss}^2} \mathbf{V}$ .

*Proof:* See Appendix E. ■

As with Lemma 2, it is worth noting that because of (25), the solution in Lemma 4 also provides the maximum information rate under (17).

### C. Maximum $\lambda_{\min}(\text{SNR}(\mathbf{F}, \mathbf{G}))$ Under Power or Maximum Eigenvalue Constraints

Designs minimizing the probability of error are difficult to deal with because they are rarely solvable in closed form; they depend on the symbol alphabet and on the detection rule. Here, we propose design criteria that can come close to the desired goal, although their optimization is *alphabet independent*. Based on (9), the optimal decision rule is the maximum likelihood (ML) detector, provided that the noise is Gaussian and that the symbols are i.i.d.<sup>3</sup> Specifically, if we let  $\mathbf{s}_i(\mathcal{H}_k)$  denote the symbol vector corresponding to hypothesis  $\mathcal{H}_k$ , and let  $\mathcal{D}_i$  denote the decision on the  $i$ th symbol block, then the ML decision rule is [15]

$$\begin{aligned} \mathcal{D}_i &= \arg \min_{\mathcal{H}_k} [\hat{\mathbf{s}}_i - \mathbf{G} \mathbf{H} \mathbf{F} \mathbf{s}_i(\mathcal{H}_k)]^H \\ &\quad \cdot (\mathbf{G} \mathbf{R}_{nn} \mathbf{G}^H)^{-1} [\hat{\mathbf{s}}_i - \mathbf{G} \mathbf{H} \mathbf{F} \mathbf{s}_i(\mathcal{H}_k)]. \end{aligned} \quad (28)$$

An indirect way of reducing the probability of error is to maximize the minimum distance between hypotheses, and this is usually done through the appropriate selection of the code vectors  $\mathbf{s}_i$ . Here, we want to search for the optimal  $\mathbf{F}$  and  $\mathbf{G}$  without changing  $\mathbf{s}_i$  in order to retain the modularity of the system design that only focuses on the choice of  $\mathbf{F}$  and  $\mathbf{G}$ . Therefore, a

<sup>3</sup>If this assumption is not satisfied, one has to use the maximum *a posteriori* probability (MAP) detector.

meaningful formulation of the problem would be to maximize the minimum distance between the hypotheses, i.e.,

$$\max_{\mathbf{F}, \mathbf{G}} \left\{ \min_{h, k; h \neq k} [\mathbf{s}_i(\mathcal{H}_h) - \mathbf{s}_i(\mathcal{H}_k)]^H \mathbf{F}^H \mathbf{H}^H \mathbf{G}^H (\mathbf{G} \mathbf{R}_{nn} \mathbf{G}^H)^{-1} \right. \\ \left. \times \mathbf{G} \mathbf{H} \mathbf{F} [\mathbf{s}_i(\mathcal{H}_h) - \mathbf{s}_i(\mathcal{H}_k)] \right\} \quad (29)$$

under some constraint on  $\mathbf{F}$ . The solution of (29) depends on the symbol alphabet. On the other hand, observing how (29) depends on  $\mathbf{F}$  and  $\mathbf{G}$ , and we can consider the following SNR-like matrix as a sensible measure related to the probability of error:

$$\text{SNR}(\mathbf{F}, \mathbf{G}) := \mathbf{F}^H \mathbf{H}^H \mathbf{G}^H (\mathbf{G} \mathbf{R}_{nn} \mathbf{G}^H)^{-1} \mathbf{G} \mathbf{H} \mathbf{F} \sigma_{ss}^2. \quad (30)$$

Instead of solving (29), realizing by design some properties of  $\text{SNR}(\mathbf{F}, \mathbf{G})$  may provide suboptimal but more general design solutions that are not tied to a certain symbol alphabet. The criterion that we will adopt is based on the observation that the minimum eigenvalue  $\lambda_{\min}(\text{SNR}(\mathbf{F}, \mathbf{G}))$  provides a lower bound for the minimum distance:

$$\min_{h, k; h \neq k} [\mathbf{s}_i(\mathcal{H}_h) - \mathbf{s}_i(\mathcal{H}_k)]^H \text{SNR}(\mathbf{F}, \mathbf{G}) [\mathbf{s}_i(\mathcal{H}_h) - \mathbf{s}_i(\mathcal{H}_k)] \\ \geq \lambda_{\min}(\text{SNR}(\mathbf{F}, \mathbf{G})) \min_{h, k; h \neq k} \|\mathbf{s}_i(\mathcal{H}_h) - \mathbf{s}_i(\mathcal{H}_k)\|^2. \quad (31)$$

Maximizing the lower bound in (31) will possibly force (31) to higher values. The corresponding solutions are given in the following two lemmas.

*Lemma 5:* The solution of the optimization problem

$$(\mathbf{F}_{\text{opt}}, \mathbf{G}_{\text{opt}}) = \underset{\mathbf{F}, \mathbf{G}}{\text{argmax}} \lambda_{\min}(\text{SNR}(\mathbf{F}, \mathbf{G})) \\ \text{tr}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H) \sigma_{ss}^2 = \mathcal{P}_0 \quad (32)$$

is given by  $\mathbf{F}_{\text{opt}} = \mathbf{V} \Phi$  with  $\Phi$  diagonal  $N \times N$  having diagonal entries

$$|\phi_{ii}|^2 = \frac{\mathcal{P}_0}{\sigma_{ss}^2 \sum_k \lambda_{kk}^{-1}} \lambda_{ii}^{-1} \quad (33)$$

and  $\mathbf{G}_{\text{opt}} = \tilde{\Gamma} \mathbf{V}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1}$  with the  $N \times N$  matrix  $\tilde{\Gamma}$  being invertible.<sup>4</sup>

*Proof:* See Appendix F.

Note that the solution leads to

$$\text{SNR}(\mathbf{F}_{\text{opt}}, \mathbf{G}_{\text{opt}}) = \sigma_{ss}^2 \Phi^H \Lambda \Phi = \frac{\mathcal{P}_0}{\sum_k \lambda_{kk}^{-1}} \mathbf{I} \quad (34)$$

that, replaced in (31), enforces the equality with the lower bound. ■

*Lemma 6:* The solution of the optimization problem

$$(\mathbf{F}_{\text{opt}}, \mathbf{G}_{\text{opt}}) = \underset{\mathbf{F}, \mathbf{G}}{\text{argmax}} \lambda_{\min}(\text{SNR}(\mathbf{F}, \mathbf{G})) \quad (35)$$

$$\lambda_{\max}(\mathbf{F} \mathbf{F}^H) \sigma_{ss}^2 = \mathcal{L}_0 \quad (36)$$

is given by  $\mathbf{F}_{\text{opt}} = \mathbf{V} \Phi$  with  $\Phi$  diagonal  $N \times N$  such that

$$|\phi_{ii}|^2 = \frac{\mathcal{L}_0 \lambda_{NN}}{\sigma_{ss}^2} \lambda_{ii}^{-1} \quad (37)$$

<sup>4</sup>The receiver selection is not completely defined by the optimal design criterion. A similar observation was made in [29] in deriving the solution for the maximum information rate.

and  $\mathbf{G}_{\text{opt}} = \tilde{\Gamma} \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1}$  with the  $N \times N$  matrix  $\tilde{\Gamma}$  being invertible.

*Proof:* See Appendix F. Similar to Lemma 5, we have  $\text{SNR}(\mathbf{F}_{\text{opt}}, \mathbf{G}_{\text{opt}}) = \mathcal{L}_0 \mathbf{I}$ . ■

Interestingly, the solution of Lemma 5 coincides with the MMSE solution under the zero forcing (ZF) constraint in [30, th. 3]. The ZF receiver in [30, th. 3] corresponds to selecting  $\tilde{\Gamma} = \mathbf{I}$ , and in this case, the design of Lemma 5 (as well as the one of Lemma 6) leads to an ML detection scheme that performs separately a low complexity quantization of the components of  $\hat{\mathbf{s}}_i$ . Indeed, as will be extensively discussed in the following section, the selection of  $\mathbf{F} = \mathbf{V} \Phi$  and  $\mathbf{G} = \tilde{\Gamma} \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1}$  with  $(\Phi, \tilde{\Gamma})$  diagonal matrices leads to diagonalizing the overall channel  $\mathbf{G} \mathbf{H} \mathbf{F}$  and the noise covariance  $\mathbf{G} \mathbf{R}_{nn} \mathbf{G}^H$ . Such diagonalization decomposes the system in a set of  $N$  parallel independent AGN subchannels for which symbol-by-symbol decision is optimal [15]. As will be shown in Section IV, the particular feature of  $\mathbf{F}_{\text{opt}}$  in Lemmas 5 and 6 is that the decision on each component of  $\hat{\mathbf{s}}_i$  is characterized by the same SNR. As a last remark, it is interesting to observe that for arbitrary  $\mathbf{F}$  and  $\mathbf{G}$ , we can extend to the Gaussian MIMO case the capacity formula of the SISO AWGN channel as follows:

$$I(\hat{\mathbf{s}}, \mathbf{x}) = \log |\mathbf{I} + \text{SNR}(\mathbf{F}, \mathbf{G})|. \quad (38)$$

#### IV. PERFORMANCE OF THE OPTIMAL DESIGNS

In this section, we will derive expressions for performance measures such as the mutual information, the probability of error, and the mean square error achievable with the optimal precoding/decoding schemes presented so far. As mentioned before, all optimal designs lead invariably to loading the power across the eigenvectors of  $\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H}$ .

##### A. Equivalent Decomposition Into Independent Subchannels

*Lemma 7:* All optimal designs we described so far have solutions of the following form:

$$\mathbf{F}_{\text{opt}} = \mathbf{V} \Phi, \quad \mathbf{G}_{\text{opt}} = \Gamma \Lambda^{-1} \mathbf{V}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \quad (39)$$

where  $\Phi$  and  $\Gamma$  are diagonal matrices.

*Proof:* See Appendix G. ■

The matrices  $\mathbf{F}_{\text{opt}}$  and  $\mathbf{G}_{\text{opt}}$  in (39) cascaded with the channel matrix  $\mathbf{H}$  in between are depicted in Fig. 1. Matrix  $\mathbf{V}$  tunes the transmit filters to the eigenstructure of the propagation channel that depends on  $\mathbf{H}$  and the AGN covariance  $\mathbf{R}_{nn}$ . The matrix equivalent of the cascade inside the box of Fig. 1 is

$$\Lambda^{-1} \mathbf{V}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{V} = \Lambda^{-1} \mathbf{V}^H \bar{\mathbf{V}} \bar{\Lambda} \bar{\mathbf{V}}^H \mathbf{V} = \mathbf{I} \quad (40)$$

and the noise correlation at the output of the box is

$$\Lambda^{-1} \mathbf{V}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{R}_{nn} \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{V} \Lambda^{-1} = \Lambda^{-1}. \quad (41)$$

Thus, the matrix (or block) channel is described by the diagonal transfer matrix  $\Gamma \Phi$  and additive noise with correlation matrix  $\Gamma^H \Gamma \Lambda^{-1}$ . Hence, the  $N$  subchannels are decoupled, and Fig. 1 becomes equivalent to Fig. 2, in which case, the flat fading on

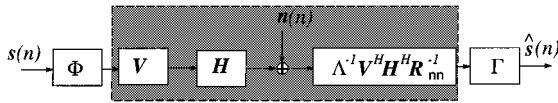


Fig. 1. Optimal transceivers: Matrix model.

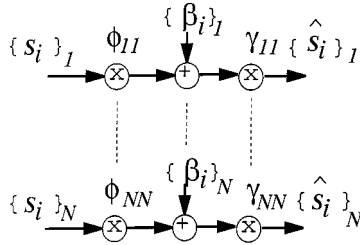


Fig. 2. Equivalent subchannels.

each of the parallel subchannels corresponds to the diagonal elements of  $\mathbf{\Gamma}\mathbf{\Phi}$ , and the noise components  $\{\beta_i\}_k, k = 1, \dots, N$  are uncorrelated with variance  $\lambda_{kk}^{-1}$ .

### B. Performance Measures

The decomposition shown in the previous subsection streamlines the performance analysis, as we summarize in the following corollaries.

*Corollary 2:* With  $\mathbf{\Phi}$  and  $\mathbf{\Gamma}$  diagonal, the transceivers in (39) render the MIMO linear AGN channel with memory equivalent to  $N$  parallel independent ISI-free subchannels, each with flat fading gain  $\phi_{kk}\gamma_{kk}$  and AGN  $\{\beta_i\}_k$ , with variance  $1/\lambda_{kk}$  and  $\{\beta_i\}_k, \{\beta_i\}_j$  uncorrelated for  $k \neq j$ , i.e.,

$$\{\hat{s}_i\}_k = \phi_{kk}\gamma_{kk}\{s_i\}_k + \gamma_{kk}\{\beta_i\}_k, \quad k = 1, \dots, N. \quad (42)$$

The SNR at the output for the  $k$ th subchannel is

$$\text{SNR}_k = \frac{\sigma_{ss}^2 |\phi_{kk}|^2 |\gamma_{kk}|^2}{\lambda_{kk}^{-1} |\gamma_{kk}|^2} = \sigma_{ss}^2 |\phi_{kk}|^2 \lambda_{kk}. \quad (43)$$

The independence of the parallel subchannels implies the following.

*Corollary 3:* For the linear transceivers of (39), the mutual information is given by

$$I(\hat{\mathbf{s}}; \mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \log_2(1 + \text{SNR}_i). \quad (44)$$

Under constrained power,  $I(\hat{\mathbf{s}}; \mathbf{x})$  in (44) achieves its maximum when  $\{\phi_{ii}\}_{i=1}^N$  are given by (23).

According to (42), the set of equivalent parallel subchannels are also ISI-free. Thus, assuming that information is quantized and that  $\{s_i\}_k$  belong to a finite alphabet, the optimal decision scheme based on  $\{\hat{s}_i\}_k$  performs symbol-by-symbol detection. When the transceiver design provides us with the freedom to choose among different constellations on different subchannels, we can choose the order of the constellation  $Q_i$  so that  $\log_2(Q_i) \approx \log_2(1 + \text{SNR}_i)$  in (44). Alternatively, the probability of error expression can be used to determine  $Q_i$ ; for ex-

ample, the symbol probability of error  $P_s(i)$  for a  $Q_i$ -QAM system is such that

$$P_s(i)^{\text{QAM}} < 2 \left(1 - \frac{1}{\sqrt{Q_i}}\right) e^{-\frac{3\text{SNR}_i}{2(Q_i-1)}} \quad (45)$$

and (45) can be used to select the constellation  $Q_i$  that leads to  $P_b(i) \leq P_D$ , where  $P_D$  is the desired upper bound on the BER that is necessary to meet a prescribed QoS. This operation is usually referred to as “bit loading” (see, e.g., [5]). We can infer from (45) that it is easier to enforce an upper bound on  $P_s(i)$  rather than  $P_b(i)$  itself. For example, considering the case of a QAM constellation, to have  $P_s(i) < P_D$ , (45) implies

$$Q_i = \left\lceil -\frac{3\text{SNR}_i}{2\log(P_D/2)} + 1 \right\rceil. \quad (46)$$

It is important to remark that the  $\text{SNR}_i$  for one or more subchannels can be such that  $Q_i \leq 2$ . Since not even one bit per block can be received at the prescribed error rate,<sup>5</sup> all subchannels with  $Q_i \leq 2$  are turned off by setting  $Q_i = 0$ . The power originally allocated to these subchannels is distributed over the remaining ones, and the size of the encoded block  $N$  is decreased correspondingly. The solution is found iteratively by discarding the subchannels with the smallest  $\text{SNR}_i$ . Comparing the expression for  $Q_i$  in (46) with the  $\text{SNR}_i$  in Table I, we can also observe that the ZF designs of Lemmas 5 and 6 that lead to uniform  $\text{SNR}_i$  are the only ones that lead to adopting a uniform constellation size across subchannels for a constant  $P_D$ . In contrast, all other designs can benefit from uneven distribution of bits according to (46) and provide an increased transmission rate, as will be illustrated by the examples of Section V.

Based on the model in Fig. 2, we can also derive the following result.

*Corollary 4:* The MSE for the  $\mathbf{F}$  and  $\mathbf{G}$  in (39) coincides with the cumulative MSE over the  $N$  independent subchannels in Fig. 2 and is given by

$$\text{MSE}(\mathbf{\Gamma}, \mathbf{\Phi}) = \sum_{i=1}^N \left[ \frac{|\gamma_{ii}|^2}{\lambda_{ii}} + |\gamma_{ii}\phi_{ii} - 1|^2 \sigma_{ss}^2 \right]. \quad (47)$$

For the MMSE receiver  $\mathbf{\Gamma} = \sigma_{ss}^2 \mathbf{\Phi}^H (\mathbf{\Lambda}^{-1} + \sigma_{ss}^2 \mathbf{\Phi} \mathbf{\Phi}^H)^{-1}$ , the MSE is

$$\overline{\text{MSE}}(\mathbf{\Phi}) = \sum_{i=1}^N \frac{\sigma_{ss}^2}{1 + |\phi_{ii}|^2 \sigma_{ss}^2 \lambda_{ii}} = \sum_{i=1}^N \frac{\sigma_{ss}^2}{1 + \text{SNR}_i} \quad (48)$$

which reaches its minimum for a constrained power when  $\phi_{ii}$  are given by (23).

For the ZF receiver  $\mathbf{\Gamma} = \mathbf{\Phi}^\dagger$ , the MSE is

$$\overline{\text{MSE}}(\mathbf{\Phi}) = \sum_{i=1}^N \frac{1}{|\phi_{ii}|^2 \lambda_{ii}} = \sum_{i=1}^N \frac{\sigma_{ss}^2}{\text{SNR}_i}. \quad (49)$$

Corollaries 1 through 4 show that performance evaluation of our optimal designs requires only specifying the values of  $\text{SNR}_i$ . Table I gives the  $\text{SNR}_i$  characterizing the optimal

<sup>5</sup>We do not consider here possible error-correcting capabilities due to error correction coding that would allow the selection of any rational value of  $Q_i > 0$ .

TABLE I  
COMPARISON OF OPTIMAL LOADING FOR DIFFERENT CODING STRATEGIES (CP := CONSTRAINT ON THE AVERAGE TRANSMIT POWER;  
 $C\lambda_{\max}$  := CONSTRAINT ON THE MAXIMUM EIGENVALUE)

Lemma	Criterion	Constr.	$\phi_{ii}$	$SNR_i =  \phi_{ii} ^2 \sigma_{ss}^2 \lambda_{ii}$
1	$\min(\text{tr}(MSE))$	CP	$\frac{1}{\sigma_{ss}^2} \left( \frac{P_0 + \sum_k \lambda_{kk}^{-1}}{\sum_k \lambda_{kk}^{-1/2}} \lambda_{ii}^{-1/2} - \frac{1}{\lambda_{ii}} \right)^+$	$(\mathcal{K}_{\text{tr}(MSE)} \lambda_{ii}^{1/2} - 1)^+$
2	$\min( MSE )$	CP	$\frac{1}{\sigma_{ss}^2} \left( \frac{P_0 + \sum_k \lambda_{kk}^{-1}}{N} - \frac{1}{\lambda_{ii}} \right)^+$	$(\mathcal{K}_{ MSE } \lambda_{ii} - 1)^+$
3	$\min( MSE )$	$C\lambda_{\max}$	$\mathcal{L}_0 / \sigma_{ss}^2$	$\mathcal{L}_0 \lambda_{ii}$
4	$\min(\text{tr}(MSE))$	$C\lambda_{\max}$	$\mathcal{L}_0 / \sigma_{ss}^2$	$\mathcal{L}_0 \lambda_{ii}$
5	$\max(\lambda_{\min}(SNR))$	CP	$\frac{1}{\sigma_{ss}^2} \left( \frac{P_0}{\sum_k \lambda_{kk}^{-1}} \right) \lambda_{ii}^{-1}$	$\mathcal{K}_{SNR-CP}$
6	$\max(\lambda_{\min}(SNR))$	$C\lambda_{\max}$	$\frac{1}{\sigma_{ss}^2} \mathcal{L}_0 \lambda_{NN} \lambda_{ii}^{-1}$	$\mathcal{K}_{SNR-C\lambda_{\max}}$

designs of Section III. Comparing the  $SNR_i$  expressions, we observe the similarity between the MMSE and water-filling solutions of Lemmas 1 and 2: Both of these solutions tend to exclude the most noisy subchannels (corresponding to smallest  $\lambda_{ii}$ s), and the  $SNR_i$  grows as  $\lambda_{ii}^{1/2}$  and  $\lambda_{ii}$ , respectively. The other designs do not include this control and, thus, waste power over these subchannels. Consequently, performance may degrade significantly for  $N$  close to its maximum value of  $Q = \text{rank}(\mathbf{H}^H \mathbf{R}_{nn} \mathbf{H})$  when the smallest eigenvalues are close to zero. This effect will be illustrated by the numerical examples in Section V.

### C. Complexity

The most demanding operation in terms of computational complexity is the estimation and the eigendecomposition of  $\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H}$ , which is a  $KM \times KM$  matrix. However, it is worth pointing out the following.

- 1) For LTI channels, the number of parameters to be estimated is actually less than or equal to  $LKR$ , where  $L$  is the maximum channel order.
- 2) OFDM transmission with  $L$  long cyclic prefix or zero guard will diagonalize each multipath channel  $h_{k,r}(l)$  for any block length  $M$ .
- 3) Assuming that the noise is spatially and temporally uncorrelated, the noise correlation matrix is  $\mathbf{R}_{nn} = [\text{diag}[\sigma_{nn}^2(1), \dots, \sigma_{nn}^2(R)] \otimes \mathbf{I}_{M \times M}]$ , where  $\otimes$  denotes the Kronecker product.

Hence, by incorporating OFDM modulation, the equivalent channel matrix  $\tilde{\mathbf{H}}$  is block diagonal, i.e.,  $\tilde{\mathbf{H}} = \text{diag}(\tilde{\mathbf{H}}[0], \dots, \tilde{\mathbf{H}}[M-1])$ , where  $\tilde{\mathbf{H}}[q] = \sum_{l=0}^L \mathbf{H}[l] \exp(-j2\pi ql/M)$  is the  $K \times R$  matrix of the channel's frequency response. Since IFFT/FFT operations do not change the noise color, the noise covariances at the output of the receivers' FFT will still be diagonal. Therefore, instead of the eigendecomposition of the  $KM \times KM$  matrix  $\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H}$ , the design will require  $M$  eigendecompositions of  $K \times K$  matrices of the form  $\tilde{\mathbf{H}}^H[q] \mathbf{\Delta}_{nn}^{-1} \tilde{\mathbf{H}}[q]$ , with  $\mathbf{\Delta}_{nn} := \text{diag}[\sigma_{nn}^2(1), \dots, \sigma_{nn}^2(R)]$  and  $q = 0, \dots, M$  (see also [24]). Subspace tracking techniques can be used to track the eigenvectors corresponding to the

strongest eigenvalues of  $\tilde{\mathbf{H}}^H[q] \mathbf{\Delta}_{nn}^{-1} \tilde{\mathbf{H}}[q]$ , which are the ones relevant to the optimal design.

## V. NUMERICAL EXAMPLES AND DISCUSSION

In this section, we provide some numerical examples that illustrate the performance of our optimal designs, relying on the performance measures presented in Section IV. We adopted the following channel model.

*Channel Model:* The FIR channel taps  $h_{k,r}(l)$  are uncorrelated complex Gaussian random variables (Rayleigh fading). The standard high performance radio LAN (HIPERLAN) provides short distance, high-speed radio links among computer systems using the 5.2– or 17.1-GHz frequency band. The variance of the taps  $\sigma_{k,r}^2(l) = \sigma^2(l)$  follows the channel power-delay profile named ‘‘Channel A,’’ which chosen as a typical indoor multipath scenario for HIPERLAN/2 in [7], operating at 5.2 GHz, with  $B = 100$  MHz ( $T = 10$  ns). The channel order is  $L \approx 19$  for the impulse response samples beyond the 19th (i.e., after 190 ns) are statistically very small. The results are always averaged over 100 random channels. We also simulated white complex stationary additive Gaussian noise having the same variance for each antenna equal to  $N_0$ . The horizontal axis in the plots that follow is the average block SNR (in decibels) defined as  $SNR := \text{tr}(\mathbf{F}\mathbf{F}^H) \sigma_{ss}^2 / N_0$ , which does not include in its definition possible gain/attenuation of the channel realization.

The assumption of uncorrelated scattering is optimistic, especially if the multipath is dominated by a few strong reflectors in the far field [24]. The matrices  $\mathbf{H}$  generated according to our model above will be, in most cases, full rank and well conditioned. However, the real scattering is likely to be correlated and, hence, leads to matrices with high condition numbers. It is also reasonable to expect that by increasing the number of antennas, the diversity gain will tend to saturate, and this aspect is not exhibited by our model. Nevertheless, the model is sufficiently accurate to provide some insight into the proposed designs.

*Example 1:* The first set of curves in Fig. 3 compare the designs in Lemmas 1–6 in terms of MSE, BER, mutual information, and number of bits that can be transmitted with a prescribed

QoS (specified as an upper bound on  $P_s(i)$ ), using the expressions derived in Section IV. Note that Lemma 3 and 4 lead to the same solution, and therefore, only five curves appear on each figure. For Lemmas 3, 4, and 6, we set the same peak eigenvalue  $\mathcal{L}_0 = \mathcal{P}_0/N$ . Because, in Lemma 3 and 4,  $\Phi = \mathcal{L}_0 \mathbf{I}$ , the  $\text{tr}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H) = \mathcal{P}_0$  for every channel. For Lemma 6, instead, if  $\lambda_{\max}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H) = \mathcal{P}_0/N$ , then  $\text{tr}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H) \neq \mathcal{P}_0$ . The curves corresponding to Lemma 6 are given as a function of the corresponding  $\text{tr}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H)/\sigma_n^2$  averaged over the channels.

Lemma 1, which minimizes the MSE subject to the average power constraint, provides perhaps the best compromise between BER and information rate. If, as described in Section IV, the constellation size is allowed to change according to (48), Fig. 3(d) shows that the design of Lemma 2, which minimizes the determinant of the MSE subject to the average power constraint, provides the highest rate for a given QoS. In Figs. 3(c) and (d), the design in Lemma 3 that minimizes the MSE subject to the peak power constraint, which leads to  $\Phi = \mathcal{K} \mathbf{I}$ , performs very similar to that in Lemma 2, which requires  $\Phi \neq \mathbf{I}$ , confirming that the power loading is not strictly necessary when maximizing the information rate because most of the water-filling gain in terms of information rate derives from the bit loading. The designs of Lemma 5 and 6 that maximize a lower bound on the minimum distance under the power and maximum eigenvalue constraints are ZF designs and tend to perform better than the other criteria only in terms of BER. The reason why Lemma 6 (eigenvalue constraint) performs slightly worse than Lemma 5 (average power constraint) is that we fixed for Lemma 6 the same  $\mathcal{L}_0 = \mathcal{P}_0/N$  used for Lemma 3 and 4 instead of fixing the average power. Had we fixed the average power to be the same, the two algorithms of Lemma 5 and 6 would have performed identically.

In Figs. 3(e) and (f), we show the MSE and information rate curves versus the ratio of the maximum eigenvalue (which is an upperbound for the peak power) over the noise variance. The curves are obtained normalizing the value in the abscissa  $\lambda_{\max}(\mathbf{F} \mathbf{F}^H)/\sigma_n^2$  to the same constant  $\mathcal{L}_0$  for all criteria in all iterations and averaging over 100 random channels. The plots clearly show the gain obtained by the designs that use the eigenvalue constraint over the corresponding criteria using the average power constraint [cf. Table I]. Only in the case of Lemma 5 and 6, because of the normalization, do the two criteria perform identically.

*Example 2:* One important aspect of the optimal designs in Lemmas 1–6 is the effect of increasing  $M, K, R$ , and the selection of the block size  $N$  (that has to be smaller than  $\text{rank}(\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H}) \leq \min(MK, MR)$  according to assumption a0). Even though the construction of the optimal transceivers allows  $N = \text{rank}(\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H})$ , various circumstances may suggest choosing a smaller  $N$ . If some  $\lambda_{ii} \ll 1$ , the ZF designs of Lemmas 5 and 6 will drain most of the power on such subchannels. Other designs, such as the ones in Lemmas 1 and 2, will tend automatically to exclude subchannels<sup>6</sup> with very small  $\lambda_{ii}$ s. However, in those cases, the choice of a smaller  $N$  might be dictated by the need of reducing the complexity of

the design. For the other designs, an  $N < \text{rank}(\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H})$  can be chosen to trade off transmission rate versus diversity gain and complexity. In this experiment, we chose to set  $K = R, \mathcal{P}_0 = M/K$  and  $N = 0.75MK$ . In Fig. 4, we show the performances normalized by  $MN$  averaged over 100 random channel for the criterion  $\min(\overline{|\text{MSE}(\mathbf{F})|}) - \text{CP}$ . The interesting observation that can be made is that the normalized performance tends to be invariant with respect to  $M$  or  $K$  and  $R$  for  $K, R \gg 1$ , i.e., they all grow proportionally to  $\min(MK, MR)$ . This applies also to the other designs not shown in Fig. 4. The linear grow of the capacity with the number of antennas was first shown in [8] and [22]. The property can be explained using the theory of random eigenvalues and the asymptotic distribution of random matrices, as was done in [20].

*Example 3:* Another important point to emphasize is the relation between  $N$  and the system parameters  $K, R$ . If we count all the physical links between information source and destination, there are  $KR$  channels in parallel, and we transmit  $P = M + L$  samples per block. However, since the channels are interfering and there is ISI, the rate does not increase proportionally to the product  $KRP$ . The rate grows as the number of equivalent uncorrelated Gaussian channels  $Q = \text{rank}(\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H})$ , which can be utilized upon appropriate precoding. Therefore, the diversity increases as the  $\text{rank}(\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H})$ . Since  $N \leq \text{rank}(\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H}) \leq \min(KP, RM)$ , the first useful observation is that  $K = R$  is preferable to  $K > R$  or  $R > K$  because  $N$  is tied to the minimum between  $K$  and  $R$ . Although it is not shown in the figures, choosing  $K = x$  and  $R = y$  leads to the same average performance as choosing  $K = y$  and  $R = x$ , as one would expect by symmetry. The difference  $|K - R|$  only helps one to increase the norm of  $\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H}$  and, thus, the relative power on the subchannels. However, it will not add diversity, which is the main source of performance improvement, as also emphasized in several papers on transmit diversity (see e.g., [11]). This is confirmed by Fig. 5 where, unlike Fig. 4, the different curves now pertain to different values of the pair  $K, R$ , assuming  $K + R$  is constant. Of course, it all depends on the nature of the scattering environment and on the way it affects the eigenvalues of  $\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H}$ . Often, the channels are modeled as having flat uncorrelated balanced Rayleigh fading coefficients. However, with only a few effective scatterers and if the transmission bandwidth is not sufficient to resolve the delay spread due to the different locations of the array elements, the number of paths, and not  $\min(K, R)$ , will limit the  $\text{rank}(\mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H})$  (see also [24]).

*Example 4:* Compared with schemes that do not make use of the channel state information (CSI) at the transmitter, the advantages in performance, power saving, and flexibility of the optimal linear designs can justify their extra complexity. Designs like the one in [1] cannot be generalized to an arbitrary number of antennas and do not make use or take advantage of the CSI, if available. Methods such as [1], [23], [32], [33] or their extension to frequency selective fading [19] operate as diversity schemes that exploit multiple antennas to increase the symbol SNR but not the symbol rate. Therefore, the number of transmitted symbols does not increase with the number of antennas and remains equal to the size of the block in time, i.e.,

<sup>6</sup>Another option is to combine some of the equivalent subchannels described in Section IV-A by sending the same information symbol on two or more of them.



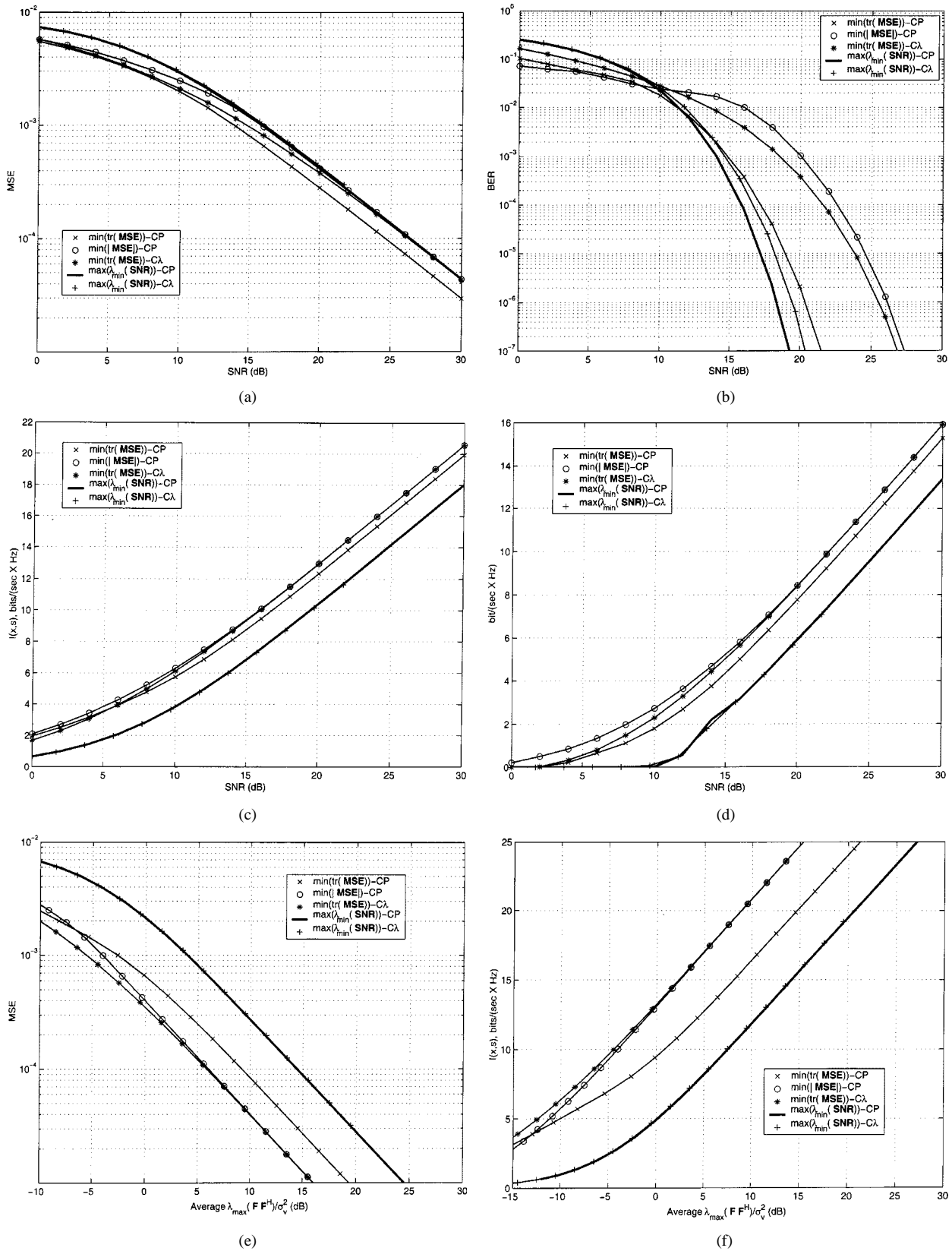


Fig. 3. Comparison between the designs in Lemma 1–6 [cf. Table I]  $M = 32$ ,  $K = R = 4$ ,  $N = 112$ ,  $\mathcal{P}_0 = 32$ . (a)–(e)  $\text{MSE} = \text{tr}(\overline{\text{MSE}}(\mathbf{F}))$ . (b) Average BER. (c)–(f)  $I(x, \hat{s})$ . (d) Number of bits  $= \sum_i \max(\lceil \log_2 Q_i \rceil, 0)$ ,  $Q_i$  given by (48) with  $P_D \leq 10^{-3}$ .

$N = M = 32$ . Setting the parameters to match the values required by the design in [1], for example  $K = 2$ ,  $R = 2$ ,  $N = M$  symbols can be transmitted over  $M$  orthogonal sub-

carriers (see, e.g., [19]). In this case, it can be shown that the  $\text{SNR}_k$  of the symbols received in each frequency bin for the scheme of [1] is  $\text{SNR}_k = \sum_{k,r=1}^2 |H_{k,r}(i)|^2 \mathcal{P}_0 / (2M\sigma_{nn}^2)$ ,

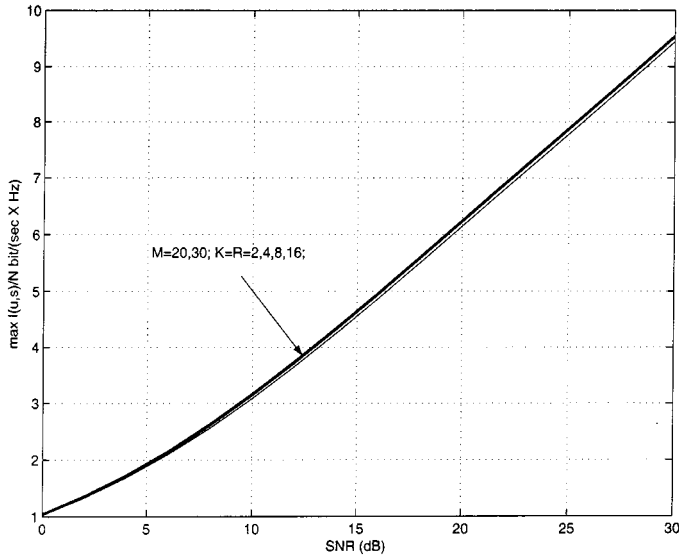


Fig. 4. Performance of the optimal designs for  $N = 0.75MK \mathcal{P}_0 = M/K$   $K = R$  antennas:  $I(\mathbf{x}, \hat{\mathbf{s}})$  Lemma 2— $(\min(|\overline{\text{MSE}}(\mathbf{F})|) - \text{CP})$ .

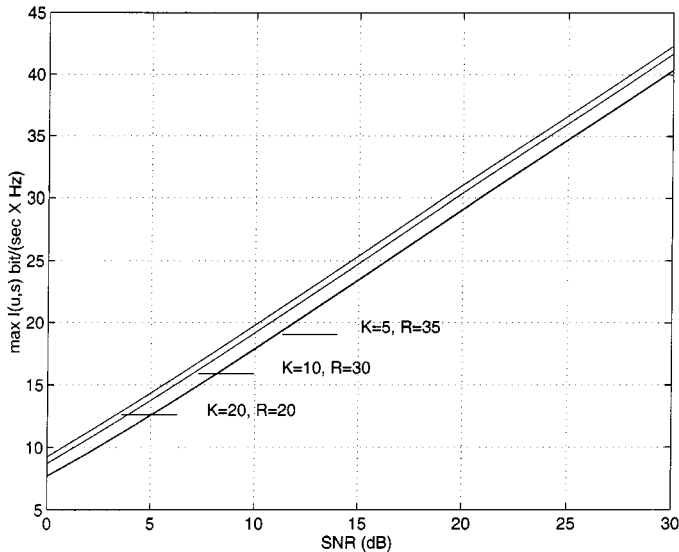


Fig. 5. Performance of the optimal designs over 100 channels according to Model B, block size  $M = 32$ ,  $N = 160$ , and various values of  $K$  and  $R$ :  $I(\mathbf{x}, \hat{\mathbf{s}})$  (Lemma 2).

where  $H_{k,r}(i) = \text{FFT}[h_{k,r}(l)]$  is the  $k, r$  channel transfer function at frequency bin  $i/M$ . We compare the BER of our scheme with that of [1] in Fig. 6.

## VI. CONCLUSION

In this paper, we derived several optimal linear designs for MIMO transmission systems with finite memory, which targeted minimum MSE and BER, under constraints on the transmit power or the maximum output of the transmitter. We also provided closed-form expressions for various performance measures of the communication link that streamline the interpretation of the optimization results. The optimal designs allow us to define space-time modulation designs that take advantage of the channel state information and offer simple closed-form solutions, scalable with respect to the number of antennas, size of the coding block, and transmit average/peak power.

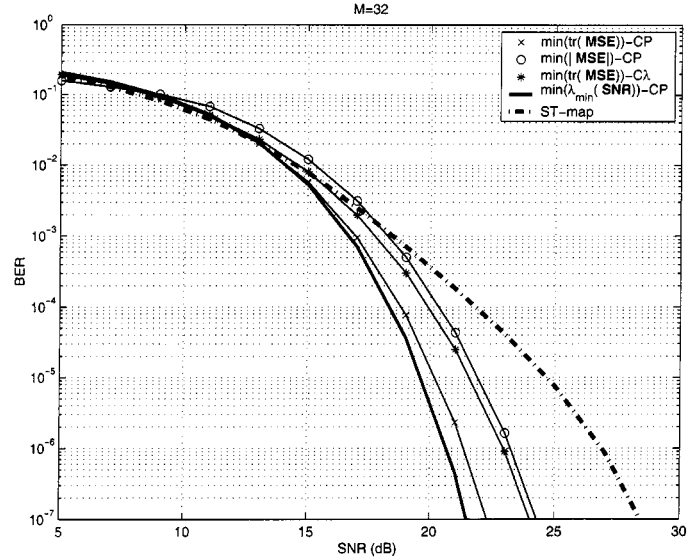


Fig. 6. BER obtained with the designs in Lemma 1–5  $K = R = 2$ ,  $M = 32$ ,  $N = 32$  and the design in [2] (ST-map).

## APPENDIX

### A. Proof of Lemma 1

Optimizing (20) could follow the steps in [18], where IIR frequency-domain designs were optimized using iterative minimization of Lagrange multipliers. However, similar to [30], our discrete time-domain matrix formulation will lead to a *closed-form* selection of the redundant FIR precoder/decoder matrices. An alternative proof to that in [30] can be found in [27] for white symbols and white noise.

The proof in [30] considers the more general case of colored symbols but restricts the attention to the case  $N = Q = \text{rank}(\mathbf{H}^H \mathbf{R}_{ss}^{-1} \mathbf{H})$ , where  $\bar{\mathbf{V}} = \mathbf{V}$  and  $\bar{\mathbf{\Lambda}} = \mathbf{\Lambda}$ . Denoting by  $\mathbf{\Delta}$  the diagonal matrix with entries equal to the eigenvalues  $\{\delta_{ii}\}_{i=1}^Q$  of matrix  $\mathbf{R}_{ss}$  sorted in decreasing order, (21) is nothing but a special case of the solution in [30] obtained for  $\{\delta_{ii}\}_{i=1}^Q = \sigma_{ss}^2$ . In [30], to prevent negative solutions for  $\phi_{ii}^2$ , a constraint was imposed on the minimum value for  $\mathcal{P}$ . Here, we would like to show that the solution can also be found for  $\mathcal{P}$  lower than the minimum value by having some  $|\phi_{ii}|^2 = 0$ . In fact, the solution in [30] was obtained setting to zero the gradient of

$$\begin{aligned} \mathcal{J} &= \min_{\mathbf{F}} [\overline{\text{MSE}}(\mathbf{F}) + \mu(\text{tr}(\mathbf{F}\mathbf{F}^H)\sigma_{ss}^2 - \mathcal{P})] \\ &= \min_{\phi_{ii}} \sum_i \frac{\delta_{ii}}{1 + |\phi_{ii}|^2 \delta_{ii} \lambda_{ii}} + \mu \left( \sum_i |\phi_{ii}|^2 \delta_{ii} - \mathcal{P} \right) \end{aligned} \quad (50)$$

with  $\{\delta_{ii}\}_{i=1}^Q = \sigma_{ss}^2$  in our case. The gradient provides us the following condition:

$$\left[ -\frac{\delta_{ii}^2 \lambda_{ii}}{(1 + |\phi_{ii}|^2 \delta_{ii} \lambda_{ii})^2} + \mu \right] \phi_{ii}^* = 0 \quad (51)$$

implying that either we send to zero the first term or we simply set  $\phi_{ii} = 0$ . Since  $\overline{\text{MSE}}(\mathbf{F})$  is a decreasing function of  $|\phi_{ii}|^2$ , if the power constraint cannot be enforced keeping all  $\phi_{ii} > 0$ , then the negative values have to be set to zero. In this case, where

$\{\delta_{ii}\}_{i=1}^Q = \sigma_{ss}^2$ , this can be done iteratively, starting from the components with the highest indexes because they are associated with the smallest  $\lambda_{ii}$  and, hence, provide the highest contribution to (50). With  $\bar{N}$  indicating the maximum number of  $\phi_{ii} > 0$ , we arrive at the solution in (21).

For  $N < Q$ , we can generalize the proof by expanding with  $Q - N$  zeros the symbol vector and having then the same setting as in [30]. By construction, in this case, the last  $Q - N$  diagonal entries of  $\mathbf{\Delta}$  are zero, i.e.,  $\delta_{ii} = \sigma_{ss}^2$  for  $i = 1, \dots, N$  and  $\delta_{ii} = 0$  for  $i = N + 1, \dots, Q$ . The proof of (50) does not require that  $\delta_{ii}$  are strictly positive and therefore is also valid for  $\delta_{ii} = \epsilon$ . Arguing by continuity and considering the limit in  $\delta_{ii} = \epsilon \rightarrow 0$  in (51) requires  $\mu\phi_{ii}^* = 0$ , and thus,  $\phi_{ii} = 0$ .

### B. Proof of Lemma 2

We will make use of the following corollary.

*Corollary 5:* Without loss of generality, the structure of  $\mathbf{F}_{\text{opt}}$  is  $\mathbf{F}_{\text{opt}} = \mathbf{V}\mathbf{\Phi}$ , where  $\mathbf{\Phi}$  is an arbitrary  $Q \times N$  matrix.

*Proof:* Any matrix  $\mathbf{F}$  and, thus, the optimal solution as well, can be decomposed as

$$\bar{\mathbf{F}}_{\text{opt}} = (\bar{\mathbf{V}}, \mathbf{V}_n) \begin{pmatrix} \mathbf{\Phi} \\ \mathbf{\Phi}_n \end{pmatrix} \quad (52)$$

with  $\mathbf{V}_n$  standing for the orthogonal complement of  $\bar{\mathbf{V}}$ . Let us introduce  $\mathbf{F}_{\text{opt}} = \bar{\mathbf{V}}\mathbf{\Phi}$ , with  $\mathbf{\Phi} Q \times N$ , which is restricted to the subspace spanned by  $\bar{\mathbf{V}}$ . We have that

$$\begin{aligned} \overline{\text{MSE}}^{-1}(\bar{\mathbf{F}}_{\text{opt}}) &= \left( \mathbf{I} + \mathbf{\Phi}^H \mathbf{V}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{V} \mathbf{\Phi} \right. \\ &\quad \left. + \mathbf{\Phi}_n^H \mathbf{V}_n^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{V}_n \mathbf{\Phi}_n \right)^{-1} \\ &\equiv \overline{\text{MSE}}^{-1}(\mathbf{F}_{\text{opt}}) \end{aligned}$$

while at the same time,  $\text{tr}(\bar{\mathbf{F}}_{\text{opt}} \bar{\mathbf{F}}_{\text{opt}}^H) = \text{tr}(\mathbf{\Phi} \mathbf{\Phi}^H) + \text{tr}(\mathbf{\Phi}_n \mathbf{\Phi}_n^H) \geq \text{tr}(\mathbf{F}_{\text{opt}} \mathbf{F}_{\text{opt}}^H)$ , where the equality holds true for  $\mathbf{\Phi}_n \equiv \mathbf{0}$ . Hence, we can focus on matrices such as  $\mathbf{F}_{\text{opt}} = \bar{\mathbf{V}}\mathbf{\Phi}$  without restricting our search for the optimal solution, and replacing it in the objective function, we can search equivalently for the optimal matrix  $\mathbf{\Phi}$ . ■

Let us assume initially that  $N = Q$  and, therefore, that  $\bar{\mathbf{V}} = \mathbf{V}$ . Using  $\mathbf{F} = \mathbf{V}\mathbf{\Phi}$  in (15), it is simple to verify that  $\mathbf{F}_{\text{opt}} = \min_{\mathbf{F}} |\overline{\text{MSE}}(\mathbf{F})| = \mathbf{V}\mathbf{\Phi}_{\text{opt}}$ , where

$$\mathbf{\Phi}_{\text{opt}} = \underset{\mathbf{\Phi}}{\text{argmax}} \left| \mathbf{I} + (\mathbf{\Phi}^H \mathbf{\Lambda} \mathbf{\Phi}) \sigma_{ss}^2 \right| \quad (53)$$

subject to  $\text{tr}(\mathbf{\Phi} \mathbf{\Phi}^H) \sigma_{ss}^2 = \sum_{n=1}^N \sum_{q=1}^Q |\phi_{q,n}|^2 = \mathcal{P}_0$ . We will use the Hadamard inequality [6, th. 16.8.2] to maximize  $|\mathbf{I} + (\mathbf{\Phi}^H \mathbf{\Lambda} \mathbf{\Phi}) \sigma_{ss}^2|$ . The Hadamard inequality dictates that the product of the diagonal entries of a positive definite matrix is always greater or equal than the matrix determinant with equality holding true if and only if the matrix is diagonal

$$\begin{aligned} |\overline{\text{MSE}}(\mathbf{F})|^{-1} &= \left| \mathbf{I} + (\mathbf{\Phi}^H \mathbf{\Lambda} \mathbf{\Phi}) \sigma_{ss}^2 \right| \\ &\leq \prod_{n=1}^N \left( 1 + \sum_{q=1}^Q \lambda_{qq} |\phi_{q,n}|^2 \right). \end{aligned} \quad (54)$$

Furthermore, for any pair of matrices  $\mathbf{A}$  and  $\mathbf{B}$  with compatible dimensions, it holds that the non-null eigenvalues of  $\mathbf{AB}$  and  $\mathbf{BA}$  coincide, i.e.,

$$\lambda(\mathbf{AB}) \equiv \lambda(\mathbf{BA}) \Rightarrow |\mathbf{I} + \mathbf{AB}| \equiv |\mathbf{I} + \mathbf{BA}| \quad (55)$$

and therefore

$$\begin{aligned} |\overline{\text{MSE}}(\mathbf{F})|^{-1} &= \left| \mathbf{I} + \mathbf{\Lambda} \mathbf{\Phi} \mathbf{\Phi}^H \sigma_{ss}^2 \right| \\ &\leq \prod_{q=1}^Q (1 + \lambda_{qq} |\{\mathbf{\Phi} \mathbf{\Phi}^H\}_{q,q}|^2) \end{aligned} \quad (56)$$

where the upperbound can be reached selecting  $\mathbf{\Phi} \mathbf{\Phi}^H$  diagonal. If  $N = Q$ , we can set, without loss of generality,  $\mathbf{\Phi}$  as diagonal, and the gradient with respect to  $\phi_{qq}$  is

$$\frac{\partial \mathcal{J}}{\partial \phi_{q,n}} = \left( \frac{1}{\lambda_{qq}^{-1} + |\phi_{q,q}|^2} - \mu \right) \phi_{q,q}^* = 0 \quad (57)$$

which leads to the water-filling principle:

$$|\phi_{q,q}|^2 = ((\mathcal{K} - \lambda_{qq}^{-1}), 0)^+ \quad q \in [1, N]. \quad (58)$$

Depending on  $\mathcal{P}$ , it may or may not be possible to set all the  $N$  diagonal entries to positive values. If  $\mathcal{P}$  is not sufficient, then other entries of  $\mathbf{\Phi}$  have to be set to zero. Since  $\text{MSE}(\mathbf{F})$  is a decreasing function of  $\lambda_{qq}$ , which are sorted in decreasing order, the minimum  $\text{MSE}(\mathbf{F})$  is obtained selecting those entries that correspond to the first (and bigger) diagonal elements of  $\mathbf{\Lambda}$ . By imposing the power constraint to determine  $\mathcal{K}$ , we arrive at (23).

However, note that if  $N < Q$ , then  $\mathbf{\Phi} \mathbf{\Phi}^H$  is rank deficient; therefore, the enforcing a diagonal structure requires having  $Q - N$  rows of  $\mathbf{\Phi}$  equal to zero. This restricts the possible values of  $|\mathbf{I} + (\mathbf{\Phi}^H \mathbf{\Lambda} \mathbf{\Phi}) \sigma_{ss}^2|$ , but it is not a restrictive assumption for our optimal solution search. In fact, since  $\lambda_{qq}$  are sorted in decreasing order, from (54), it follows that any  $\hat{\mathbf{\Phi}}$  having the last  $Q - N$  rows equal to zero and such that  $\forall n$

$$|\hat{\phi}_{1,n}|^2 = \sum_{n=N+1}^Q |\phi_{q,n}|^2 \quad (59)$$

still satisfying the constraint provides an higher or equal upper bound on  $|\mathbf{I} + (\mathbf{\Phi}^H \mathbf{\Lambda} \mathbf{\Phi}) \sigma_{ss}^2|$ . At the same time, matrices having such structure as  $\hat{\mathbf{\Phi}}$  can reach this upper bound by using the same design as in (58), setting  $|\phi_{q,q}|^2 = 0$  for  $q = N+1, \dots, Q$ . Thus, the cost function is maximized by (58) for  $N \leq Q$ . Denoting by  $\mathbf{\Phi}$  only the top  $N \times N$  nonzero part of the original matrix  $\mathbf{\Phi}$  used above, based on a2), we can write  $\mathbf{F}_{\text{opt}} = \mathbf{V}\mathbf{\Phi}$ , where  $\mathbf{\Phi}$  is  $N \times N$  and diagonal.

### C. Proof of Corollary 1

The mutual information  $I(\hat{\mathbf{s}}, \mathbf{x})$  per block is

$$\begin{aligned} I(\hat{\mathbf{s}}, \mathbf{x}) &= \log \frac{|\mathbf{G}(\sigma_{ss}^2 \mathbf{H} \mathbf{F} \mathbf{F}^H \mathbf{H}^H + \mathbf{R}_{nn}) \mathbf{G}^H|}{|\mathbf{G} \mathbf{R}_{nn} \mathbf{G}^H|} \\ &= \log \left| \sigma_{ss}^2 \mathbf{G} \mathbf{H} \mathbf{F} \mathbf{F}^H \mathbf{H}^H \mathbf{G}^H (\mathbf{G} \mathbf{R}_{nn} \mathbf{G}^H)^{-1} + \mathbf{I} \right|. \end{aligned} \quad (60)$$

Because of (59), (60) can be written as

$$\begin{aligned} I(\hat{\mathbf{s}}, \mathbf{x}) &= \log \left| \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1/2} \Pi_{\mathbf{R}^{1/2} \mathbf{G}^H} \mathbf{R}_{nn}^{-1/2} \mathbf{H} \mathbf{F} \sigma_{ss}^2 + \mathbf{I} \right| \\ &\leq \log \left| \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{F} \sigma_{ss}^2 + \mathbf{I} \right| \end{aligned} \quad (61)$$

where  $\Pi_{\mathbf{R}^{1/2} \mathbf{G}^H} = \mathbf{R}_{nn}^{1/2} \mathbf{G}^H (\mathbf{G} \mathbf{R}_{nn} \mathbf{G}^H)^{-1} \mathbf{G} \mathbf{R}_{nn}^{1/2} \leq \mathbf{I}$  is the orthogonal projector onto the range space of  $\mathbf{R}_{nn}^{1/2} \mathbf{G}^H$ . The upper bound on  $I(\hat{\mathbf{s}}, \mathbf{x})$  is reached if and only if  $\mathbf{R}_{nn}^{1/2} \mathbf{G}^H = \mathbf{R}_{nn}^{-1/2} \mathbf{H} \mathbf{F} \tilde{\mathbf{\Gamma}}^H$ , where  $\tilde{\mathbf{\Gamma}}$  is an  $N \times N$  invertible arbitrary matrix, i.e., if  $\mathbf{G}$  has the structure in (24). Lemma 7 shows that the MMSE receiver in (14) can be written equivalently in the same form as (24) [see (70) in Appendix G]; thus,  $\mathbf{G}_{\text{opt}}$  in (12) is optimal also in the sense of maximizing  $I(\hat{\mathbf{s}}, \mathbf{x})$ .

From (25), one can observe that

$$\begin{aligned} I(\hat{\mathbf{s}}, \mathbf{x}) &= \log \left| \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{F} \sigma_{ss}^2 + \mathbf{I} \right| \\ &= \log \left| \sigma_{ss}^2 \overline{\text{MSE}}^{-1}(\mathbf{F}) \right|. \end{aligned} \quad (62)$$

Because maximizing  $|\overline{\text{MSE}}^{-1}(\mathbf{F})|$  or minimizing  $|\overline{\text{MSE}}(\mathbf{F})|$  is equivalent,  $\mathbf{F}_{\text{opt}}$  in (25) maximizes  $I(\hat{\mathbf{s}}, \mathbf{x})$ .

#### D. Proof of Lemma 3

The difficulty in finding the  $\min_{\mathbf{F}} \text{tr}(\overline{\text{MSE}}(\mathbf{F}))$  under the maximum eigenvalue constraint is in showing that the cost function can be minimized, without loss of generality, by a precoder that loads power separately on each channel eigenvector (as  $\mathbf{F} = \mathbf{V} \Phi$  with  $\Phi$  diagonal). Assuming  $N = Q$ , we can use two facts: 1) For any positive semidefinite matrix,  $\mathbf{A}$ ,  $\{\mathbf{A}^{-1}\}_{qq} \geq \{\mathbf{A}\}_{qq}^{-1}$  (see also [21]), and 2)  $|\phi_{qn}|^2 \leq \lambda_{\max}(\Phi^H \Phi) \equiv \mathcal{L}_0$ . Because of Corollary 5, (55), and because of the two properties stated above, we can write

$$\begin{aligned} \text{tr}(\overline{\text{MSE}}(\mathbf{F})) &= \text{tr}((\mathbf{I} + \Phi \Phi^H \Lambda)^{-1}) \\ &\geq \sum_{n=1}^N (1 + \{\Phi \Phi^H\}_{nn} \lambda_{nn})^{-1} \\ &\geq \sum_{n=1}^N (1 + \mathcal{L}_0 \lambda_{nn})^{-1} \end{aligned} \quad (63)$$

where the minimum is attained when  $\Phi \Phi^H = \mathcal{L}_0 \mathbf{I}$  so that  $\Phi$  allocates on each column of  $\mathbf{V}$  the maximum power allowed, which is equal to  $\mathcal{L}_0$ . For  $N < Q$ , because  $\Phi$  is of size  $Q \times N$ ,  $\Phi \Phi^H$  has only  $N$  nonzero eigenvalues. Arguing as done at the end of Appendix B, the optimum choice is to have only the top right  $N \times N$  submatrix of  $\Phi \Phi^H$  not equal to zero, and all the nonzero eigenvalues equal to the maximum  $\mathcal{L}_0 / \sigma_{ss}^2$ . Therefore,  $\Phi$  can simply denote the  $N \times N$  nonzero matrix such that  $\Phi \Phi^H = \mathcal{L}_0 / \sigma_{ss}^2 \mathbf{I}$  so that  $\mathbf{F} = \mathbf{V} \Phi$  is the optimum solution. In particular,  $\Phi = \sqrt{\mathcal{L}_0 / \sigma_{ss}^2} \mathbf{I}$  is a possible choice for  $\Phi$ .

#### E. Proof of Lemma 4

Normalizing the objective function by  $\sigma_{ss}^2$ , we want to solve the equivalent optimization problem

$$\begin{aligned} \max_{\mathbf{F}} \left| \mathbf{I} + \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{F} \sigma_{ss}^2 \right| \quad \text{and} \\ \lambda_{\max}(\mathbf{F}^H \mathbf{F}) \sigma_{ss}^2 = \mathcal{L}_0. \end{aligned} \quad (64)$$

Using (55), we can write

$$\begin{aligned} \left| \mathbf{I} + \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{F} \sigma_{ss}^2 \right| \\ = \left| \mathbf{I} + \mathbf{F}^H \bar{\mathbf{V}} \bar{\Lambda}^{1/2} \bar{\Lambda}^{1/2} \bar{\mathbf{V}}^H \mathbf{F} \sigma_{ss}^2 \right| \\ = \left| \mathbf{I} + \bar{\Lambda}^{1/2} \bar{\mathbf{V}}^H \mathbf{F} \mathbf{F}^H \bar{\mathbf{V}} \bar{\Lambda}^{1/2} \sigma_{ss}^2 \right| \\ \leq \left| \mathbf{I} + \mathcal{L}_0 \bar{\Lambda} \right| \end{aligned} \quad (65)$$

where the inequality takes into account the fact that  $Q - N$  eigenvalues of  $\mathbf{F} \mathbf{F}^H$  are equal to zero. The upperbound is achieved when  $\mathbf{F}_{\text{opt}} = \sqrt{\mathcal{L}_0 / \sigma_{ss}^2} \bar{\mathbf{V}}$ .

#### F. Proof of Lemmas 5 and 6

Arguing as in Appendix C, we can write

$$\begin{aligned} \text{SNR}(\mathbf{F}, \mathbf{G}) &= \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1/2} \Pi_{\mathbf{R}^{1/2} \mathbf{G}^H} \mathbf{R}_{nn}^{-1/2} \mathbf{H} \mathbf{F} \\ &\leq \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{F} \end{aligned} \quad (66)$$

where  $\Pi_{\mathbf{R}^{1/2} \mathbf{G}^H} = \mathbf{R}_{nn}^{1/2} \mathbf{G}^H (\mathbf{G} \mathbf{R}_{nn} \mathbf{G}^H)^{-1} \mathbf{G} \mathbf{R}_{nn}^{1/2} \leq \mathbf{I}$  is the orthogonal projector onto the range space of  $\mathbf{R}_{nn}^{1/2} \mathbf{G}^H$ . The SNR( $\mathbf{F}, \mathbf{G}$ ) reaches its upper bound if and only if  $\mathbf{R}_{nn}^{1/2} \mathbf{G}^H = \mathbf{R}_{nn}^{-1/2} \mathbf{H} \mathbf{F} \tilde{\mathbf{\Gamma}}^H$ , where  $\tilde{\mathbf{\Gamma}}$  is an  $N \times N$  invertible arbitrary matrix, i.e., if  $\mathbf{G} = \tilde{\mathbf{\Gamma}} \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1}$ . Let us denote by  $\overline{\text{SNR}}(\mathbf{F}) = \mathbf{F}^H \mathbf{H}^H \mathbf{R}_{nn}^{-1} \mathbf{H} \mathbf{F}$ . It is not difficult to extend the result in Corollary 5 to the problem of maximizing the minimum eigenvalue of  $\overline{\text{SNR}}(\mathbf{F})$ , and for brevity, we will not repeat the proof here. Based on Corollary 5, assuming for the moment that  $\Phi$  is  $Q \times N$ , we have

$$\text{SNR}(\mathbf{F}, \mathbf{G}) = \Phi^H \bar{\Lambda} \Phi \sigma_{ss}^2. \quad (67)$$

Hence, instead of (32), we can solve the equivalent problem

$$\begin{aligned} \Phi_{\text{opt}} &= \arg \max_{\Phi} \lambda_{\min}(\bar{\Lambda} \Phi \Phi^H) > 0 \quad \text{and} \\ \text{tr}(\Phi \Phi^H) \sigma_{ss}^2 &= \mathcal{P}_0 \end{aligned} \quad (68)$$

where we also used the fact that for matrices with compatible dimensions, the non-null eigenvalues are such that  $\lambda(AB) = \lambda(BA)$ . The solution to (68) is obtained for  $\bar{\Lambda} \Phi \Phi^H = \alpha \mathbf{I}$  so that the minimum and the maximum non-null eigenvalues of  $\bar{\Lambda} \Phi \Phi^H$  coincide. The power constraint leads to associating these non-null eigenvalues to the corresponding  $N$  diagonal larger entries of  $\bar{\Lambda}$ . This, as well as the constraint, can be enforced by setting to zero the last  $Q - N$  rows and selecting the top  $N \times N$  part of  $\Phi$  that, for simplicity, we will still denote as  $\Phi$ , as  $\Phi = \sqrt{\alpha} \bar{\Lambda}^{-1/2}$  with  $\alpha = \mathcal{P}_0 / (\sigma_{ss}^2 \text{tr}(\bar{\Lambda}^{-1}))$ . This proves Lemma 5. The solution of Lemma 6 is also in the form  $\Phi = \sqrt{\alpha} \bar{\Lambda}^{-1/2}$  with the only

difference being that to meet the eigenvalue constraint (which can be equivalently written as  $\lambda_{\max}(\Phi\Phi^H)\sigma_{ss}^2 = \mathcal{P}_0/N$ ), we need  $\alpha = \mathcal{L}_0/\sigma_{ss}^2 \max_n(\lambda_{nn}^{-1}) = \lambda_{NN}\mathcal{L}_0/\sigma_{ss}^2$ .

### G. Proof of Lemma 7

As far as Lemmas 5 and 6 are concerned, (39) is simply verified from their statements, except that the structure of  $\tilde{\mathbf{\Gamma}}$  is left arbitrary and, in particular

$$\begin{aligned} \mathbf{G}_{\text{opt}} &= \tilde{\mathbf{\Gamma}}\Phi^H\Lambda\Lambda^{-1}\mathbf{V}^H\mathbf{H}^H\mathbf{R}_{nn}^{-1} \\ &= \mathbf{\Gamma}\Lambda^{-1}\mathbf{V}^H\mathbf{H}^H\mathbf{R}_{nn}^{-1} \end{aligned} \quad (69)$$

where  $\mathbf{\Gamma}$  can be chosen to be  $\tilde{\mathbf{\Gamma}} = \text{diagonal}$ . Selecting to be diagonal simplifies the decision rule (28) that instead of operating jointly on all the components of  $\hat{\mathbf{s}}_i$ , it operates separately on every symbol since in this case,  $\text{SNR}(\mathbf{F}, \mathbf{G}) = \Phi\Phi^H\Lambda\sigma_{ss}^2$  is diagonal [cf. (30), (40), and (41)].

In Lemmas 1–4, the structure of  $\mathbf{F}_{\text{opt}}$  is the same as in (39), whereas  $\mathbf{G}_{\text{opt}}$ , selected as in (12), appears to have a different structure. Nevertheless, simple derivations show that the expression of  $\mathbf{G}_{\text{opt}}$  in (12) can be rearranged in the same fashion as in (39). In fact, with  $\mathbf{A} := \sigma_{ss}\mathbf{R}_{nn}^{-1/2}\mathbf{H}\mathbf{F} = \Lambda^{1/2}\mathbf{V}^H\mathbf{V}\Phi = \Lambda^{1/2}\Phi$ , we can write (12) as

$$\begin{aligned} \mathbf{G} &= \sigma_{ss}\mathbf{F}^H\mathbf{H}^H\mathbf{R}_{nn}^{-1/2} \left( \sigma_{ss}^2\mathbf{R}_{nn}^{-1/2}\mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H\mathbf{R}_{nn}^{-1/2} + \mathbf{I} \right)^{-1} \\ &\quad \cdot \sigma_{ss}\mathbf{R}_{nn}^{-1/2} = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \mathbf{I})^{-1}\mathbf{R}_{nn}^{-1/2}\sigma_{ss} \end{aligned}$$

and using the matrix inversion lemma,<sup>7</sup> we have that

$$\mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \mathbf{I})^{-1} = (\mathbf{A}^H\mathbf{A} + \mathbf{I})^{-1}\mathbf{A}^H.$$

Therefore, we can write

$$\begin{aligned} \mathbf{G}_{\text{mmse}} &= (\mathbf{A}^H\mathbf{A} + \mathbf{I})^{-1}\mathbf{A}^H\mathbf{R}_{nn}^{-1/2} \\ &= (\mathbf{A}^H\mathbf{A} + \mathbf{I})^{-1}\mathbf{A}^H\mathbf{R}_{nn}^{-1/2} \\ &= \tilde{\mathbf{\Gamma}}\mathbf{F}^H\mathbf{H}^H\mathbf{R}_{nn}^{-1} \end{aligned} \quad (70)$$

which, after replacing the expression of  $\mathbf{F} = \mathbf{V}\Phi$  and with some extra manipulations, as in (69) and denoting by  $\tilde{\mathbf{\Gamma}} := (\mathbf{A}^H\mathbf{A} + \mathbf{I})^{-1}\mathbf{A}^H\sigma_{ss}^2$  and by  $\mathbf{\Gamma} := \tilde{\mathbf{\Gamma}}\Phi^H\Lambda$ , we arrive at the expression in (39).

### REFERENCES

- [1] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
- [2] S. Barbarossa and A. Scaglione, "Theoretical bounds on the estimation and prediction of multipath time-varying channels," in *Proc. Int. Conf. Acoust. Speech, Signal Process.*, Istanbul, Turkey, June 5–9, 2000.
- [3] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [4] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell Syst. Tech. J.*, vol. 53, pp. 745–778, May/June 1974.
- [5] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL applications," *IEEE J. Select. Areas Commun.*, pp. 895–908, Aug. 1991.

<sup>7</sup>The matrix inversion lemma states the following identity:

$$(\mathbf{A} + \mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{D}\mathbf{A}^{-1}.$$

- [6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [7] *Norme ETSI*, 1998.
- [8] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [9] G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Optimal space-time processing with multiple transmit and receive antennas," *IEEE Commun. Lett.*, vol. 5, no. 3, pp. 85–87, 2001.
- [10] G. D. Forney Jr. and M. V. Eyuboğlu, "Combined equalization and coding using precoding," *IEEE Commun. Mag.*, pp. 25–34, Dec. 1991.
- [11] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading multipath environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [12] R. Gallager, *Elements of Information Theory*. New York: Wiley, 1968, sec. 8.
- [13] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.
- [14] S. Kasturia, J. T. Aslanis, and J. M. Cioffi, "Vector coding for partial response channels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 741–761, July 1990.
- [15] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation/Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993, vol. 1, 2.
- [16] J. W. Lechleider, "The optimum combination of block codes and receivers for arbitrary channels," *IEEE Trans. Commun.*, pp. 615–621, May 1990.
- [17] J. Yang and S. Roy, "Joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) with decision feedback," *IEEE Trans. Inform. Theory*, vol. 42, pp. 3221–3231, Sept. 1994.
- [18] —, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, pp. 3221–3231, Dec. 1994.
- [19] Y. Li, J. Chuang, and N. R. Sollenberger, "Transmitter diversity for ofdm systems and its impacts on high-rate wireless networks," *Proc. IEEE Int. Conf. Commun.*, vol. 1, pp. 534–538, June 1999.
- [20] Ü. Sakoğlu and A. Scaglione, "Asymptotic capacity of space-time coding for arbitrary fading channels: A closed form solution using Girko's law," *Proc. IEEE ICASSP*, May 2001.
- [21] H. S. Malvar and D. H. Staelin, "Optimal pre- and postfilters for multichannel signal processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 287–289, Feb. 1988.
- [22] T. L. Marzetta and B. H. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.
- [23] A. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communication," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1459–1478, Oct. 1998.
- [24] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.
- [25] G. G. Raleigh and V. K. Jones, "Multivariate modulation and coding for wireless communication," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 851–866, May 1999.
- [26] J. Salz, "Digital transmission over cross-coupled linear channels," *AT&T Tech. J.*, vol. 64, pp. 1147–1159, July/Aug. 1985.
- [27] H. Sampath and A. Paulraj, "Joint transmit and receive optimization for high data rate wireless communications using multiple antennas," in *Proc. 33rd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 1999, pp. 215–219.
- [28] H. Sampath, P. Stoica, and A. Paulraj, "Generalized precoder and decoder design for MIMO channels, using the weighted MMSE criterion," *IEEE Trans. Commun.*, to be published.
- [29] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1019–1032, Apr. 1999.
- [30] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2006, July 1999.
- [31] 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.
- [32] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility and multiple paths," *IEEE Trans. Commun.*, vol. 47, pp. 199–207, Feb. 1999.
- [33] —, "Combined array processing and space-time coding," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1121–1128, May 1999.



**Anna Scaglione** (M'99) received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Rome, "La Sapienza," Rome, Italy, in 1995 and 1999, respectively.

In August 2000, she joined the Electrical and Computer Engineering Department, University of New Mexico, Albuquerque, as an Assistant Professor. In 1999, she was with the the University of Minnesota, Minneapolis, as a Postdoctoral Researcher. She is now with the School of Electrical and Computer Engineering, Cornell University,

Ithaca, NY. Her background is in statistical signal processing and digital communications. Her research field is in signal processing for communications, and her main publications are on the optimization of digital modulation and multiplexing techniques for frequency-selective fading channels and broadband transmission.

Dr. Scaglione, with co-authors, received the 2000 IEEE Signal Processing Best Paper Award.



**Petre Stoica** (F'94) received the D.Sc. degree in automatic control from the Bucharest Polytechnic Institute (BPI), 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. Previously, he was a Professor of Signal Processing at BPI. He held longer visiting positions with the Eindhoven University of Technology, Eindhoven, The Netherlands; Chalmers University of Tech-

nology, Göteborg, Sweden (where he held a Jubilee Visiting Professorship); UU; the University of Florida, Gainesville; and Stanford University, Stanford, CA. 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. The most recent book he co-authored is entitled *Introduction to Spectral Analysis* (Englewood Cliffs, NJ: Prentice-Hall, 1997). Recently, he co-edited two books on signal processing advances in wireless and mobile communications (Englewood Cliffs, NJ: Prentice-Hall, 2000). He is on the editorial boards of five journals in the field, including *Signal Processing*; *Journal of Forecasting*; *Circuits, Signals, and Signal Processing*; *Multidimensional Systems and Signal Processing*; and *Digital Signal Processing: A Review Journal*. He was a guest co-editor for several special issues on system identification, signal processing, spectral analysis, and radar for some of the aforementioned journals as well as for *Proceedings of the IEE*.

Dr. Stoica was co-recipient of the 1989 ASSP Society Senior Award for a paper on statistical aspects of array signal processing and recipient of the 1996 Technical Achievement Award of the IEEE Signal Processing Society for fundamental contributions to statistical signal processing with applications in time series analysis, system identification, and array processing. In 1998, he received a Senior Individual Grant Award from the Swedish Foundation for Strategic Research. He was also co-recipient of the 1998 EURASIP Best Paper Award for Signal Processing for a work on parameter estimation of exponential signals with time-varying amplitude, a 1999 Signal Processing Society Best Paper Award for a paper on parameter and rank estimation of reduced-rank regression, a 2000 IEEE Third Millennium Medal, and the 2000 IEEE W. R. G. Baker Paper Prize Award for a work on maximum likelihood methods for radar. He was a member of the international program committees of many topical conferences. From 1981 to 1986, he was a Director of the International Time Series Analysis and Forecasting Society, and he has been a Member of the IFAC Committee on Modeling, Identification, and Signal Processing since 1994. He is also a Member of the Romanian Academy and a Fellow of the Royal Statistical Society.



**Sergio Barbarossa** (M'88) received the B.S. degree (cum laude) in electrical engineering in 1984 and the Ph.D. degree in information and communication engineering in 1989, both from the University of Rome "La Sapienza," Rome, Italy.

He joined the Radar System Division of Selenia in 1985 as a radar system engineer. From November 1987 to August 1988, he was a Research Engineer at the Environmental Research Institute of Michigan (ERIM), Ann Arbor, where he was involved in research activities on synthetic aperture radars. After three years at the University of Perugia, Perugia, Italy, as an Adjunct Professor, in 1991, he joined the University of Rome "La Sapienza," where he is now a Full Professor. He has held positions as Visiting Scientist and Visiting Professor at the University of Virginia, Charlottesville, and at the University of Minnesota, Minneapolis. His current research interests lie in the areas of signal processing for digital communications, multiple access methods, space-time coding, and communications over time-varying channels.

Dr. Barbarossa has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and since 1997, he has been a member of the IEEE Technical Committee for Signal Processing in Communications. He is co-recipient of the 2000 IEEE Best Paper Award in the Signal Processing for Communications area.



**Georgios B. Giannakis** (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1981. From September 1982 to July 1986, he was with the University of Southern California (USC), Los Angeles, where he received the M.Sc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of electrical engineering in 1997. Since 1999, he has been with the University of Minnesota, Minneapolis, as a Professor of electrical and computer engineering. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 120 journal papers, 250 conference papers, and two edited books. Current research topics focus on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, redundant precoding, and space-time coding for block transmissions, multicarrier, and wideband wireless communication systems. He is a frequent consultant for the telecommunications industry.

Dr. Giannakis is the (co-) recipient of Best Paper Awards from the IEEE Signal Processing (SP) Society in 1992, 1998, and 2000. He also received the Society's Technical Achievement Award in 2000. He co-organized three IEEE-SP Workshops (HOS in 1993, SSAP in 1996, and SPAWC in 1997) and guest (co-) edited four special issues. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, a secretary of the SP Conference Board, a member of the SP Publications Board, and a member and vice chair of the Statistical Signal and Array Processing Committee. He is a member of the Editorial Board for the PROCEEDINGS OF THE IEEE, he chairs the SP for Communications Technical Committee, and he serves as the Editor in Chief for the IEEE SIGNAL PROCESSING LETTERS. He is a member of the IEEE Fellows Election Committee and the IEEE-SP Society's Board of Governors.



**Hemanth Sampath** (M'01) received the B.S.E.E. degree (with Honors) in 1996, the M.S. degree in 1998, and the Ph.D. degree in 2001, all in electrical engineering from the University of Maryland, College Park.

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