

# Joint Transmit–Receive Space–Time Equalization in Spatially Correlated MIMO Channels: A Beamforming Approach

Daniel Pérez Palomar, *Student Member, IEEE*, and Miguel Angel Lagunas, *Fellow, IEEE*

**Abstract**—Multi-input multi-output (MIMO) channels have been shown in the literature to present a significant capacity increase over single-input single-output ones in some situations. To achieve this theoretical capacity, the constituent parallel subchannels arising from the MIMO channel have to be properly used. Many practical schemes are being currently developed to achieve this goal. In this paper, we first show that, from an information-theoretic point of view, beamforming becomes asymptotically optimal as the spatial correlation of the channel fading increases. In light of this result, wideband beamvectors are jointly derived for both transmission and reception. We allow a controlled partial response and design zero-forcing and minimum mean-squared error transmit–receive filters. Conceptually, the beamforming scheme is shown to decompose into two stages: the first one corresponds to a spatial flattening of the MIMO channel, i.e., choosing the subchannel with the highest gain at each frequency; the second stage depends on the particular design criterion and performs a power distribution at the transmitter and defines the equalizer at the receiver. These methods are further extended to the general case of multiple beamforming, i.e., when more than one subchannel are used. An exact and practical implementation of a modified “waterfilling” solution required for the filter design is proposed. All derived methods are assessed and compared in terms of capacity and bit-error rate.

**Index Terms**—Array signal processing, beamforming, joint transmit–receive equalization, multi-input multi-output (MIMO) systems, space–time filtering, waterfilling.

## I. INTRODUCTION

**M**ULTI-INPUT multi-output (MIMO) channels arise from the use of multiple dimensions for transmission and reception. Recently, MIMO channels arising from the use of spatial diversity at both the transmitter and the receiver have attracted considerable attention [1], [2]. They have been shown to present a significant increase in capacity over single-input single-output (SISO) systems because of the constituent parallel

subchannels (also termed channel eigenmodes) existing within the MIMO channel. The capacity of a MIMO channel depends on a variety of factors such as the number of antennas utilized, the correlation of the fades [3], the power allocation strategy, and the frequency-selectivity of the channel [4].

Multiple high bit rate communication schemes have been recently proposed for MIMO channels. In most situations, the channel is assumed known at the receiver. A significant part of the techniques assume no channel state information at the transmitter (CSIT) [5]–[7]. Another significant part of the methods assume CSIT and, therefore, the transmitter can adapt to each channel realization [8], [9]. Channel knowledge at the transmitter can be achieved either by means of a feedback channel or by estimating the received channel and then applying the channel reciprocity property (when applicable). In the sequel, it is assumed that the channel is known at both ends of the communication link and a joint transmit–receive processing is considered.

One of the first approaches in the design of joint transmit–receive filters was done for frequency-selective SISO channels in [10] (and references therein) to minimize the mean square error (MSE), where an iterative “waterfilling” algorithm was found for optimum energy distribution. In [11], the solution was extended to  $2 \times 2$  MIMO channels. Decision feedback schemes were considered in [12]. A generalization to  $N \times N$  matrix channels was obtained in [13]. In [14], the case was generalized to an arbitrary  $M \times N$  matrix channel with correlated data symbols, colored noise, both near- and far-end crosstalk and excess bandwidth (although a closed-form expression was not provided, an iterative solution was presented). In [8] and [15], joint transmit–receive filters were derived using an elegant notation for a general framework including excess bandwidth and decision feedback systems. Remarkably, the joint transmit–receive design for MIMO systems was already solved in 1976 for flat channels [16]. In [17], the flat MIMO case was considered giving useful insights from the point of view of beamforming. In [18] and [19], an interesting approach based on linear filtering using a block notation was derived under a multirate filterbank framework.

This paper considers space–time filtering for frequency-selective MIMO channels at both sides of the link, commonly referred to linear precoder at the transmitter and equalizer at the receiver. The focus is on simplified processing, i.e., a beamforming approach, which is preferred in terms of complexity because only one data stream needs to be considered and coding and transmission can be done in a much easier

Manuscript received January 10, 2002; revised October 23, 2002. This work was supported in part by the European Commission under Project IST-2000-30148 I-METRA; by the Spanish Government (CICYT) TIC99-0849, TIC2000-1025, FIT-070000-2000-649 (MEDEA + A105 UniLAN); and in part by the Catalan Government (DURSI) 1999FI 00588, 2001SGR 00268. This work was presented in part at the Conference on Information Sciences and Systems (CISS), Baltimore, MD, March 2001 (Palomar, 2001).

D. P. Palomar is with the Department of Signal Theory and Communications, Technical University of Catalonia (UPC), Campus Nord UPC, 08034 Barcelona, Spain (e-mail: daniel@gps.tsc.upc.es).

M. A. Lagunas is with the Department of Signal Theory and Communications, Technical University of Catalonia (UPC), Campus Nord UPC, 08034 Barcelona, Spain and also with the Telecommunications Technological Center of Catalonia (CTTC), 08034 Barcelona, Spain (e-mail: m.a.lagunas@cttc.es).

Digital Object Identifier 10.1109/JSAC.2003.810324

manner (as in single antenna systems) [20]. Since beamforming implies a rank-one transmit covariance matrix, we first prove the asymptotic optimality (in the sense of achieving capacity) of beamforming on MIMO channels as the spatial channel fading correlation increases at least at one end of the link. Then, we jointly derive optimum transmit and receive beamvectors according to different criteria such as zero-forcing (ZF) or minimum mean squared error (MMSE) allowing an arbitrary partial response termed partial response-joint ZF (PR-JZF) and partial response-joint MMSE (PR-JMMSE), respectively. Conceptually, the beamforming scheme is shown to decompose into two stages: the first one corresponds to a spatial flattening of the MIMO channel (i.e., choosing the subchannel with the highest gain at each frequency); the second stage depends on the particular design criterion and performs a power distribution at the transmitter and defines the equalizer at the receiver. The derivation of the filters is then extended to the more general case of using a set of parallel subchannels or eigenmodes (multiple beamforming), using the PR-JZF criterion (which is a generalization of the ZF criterion used in [18]) and the PR-JMMSE criterion (generalizing the results of [8] with a derivation much simpler and shorter). We also propose an exact and practical algorithm for the modified waterfilling solution that is required in the MMSE solution as an alternative to the existing iterative approaches [10], [11], and [8]. This algorithm is based on and generalizes the work in [21, Ch. 4] and [19, Alg. 1], where the classical capacity-achieving waterfilling solution was considered.

The paper is structured as follows. The signal model is introduced in Section II. In Section III, the asymptotic optimality of beamforming is proven. Then, in Section IV, the joint derivation of the transmit–receive beamvectors is carried out and further extended to the case of multiple beamforming in Section V. A practical and exact implementation of the generalized waterfilling algorithm is given in Section VI. Section VII is devoted to the numerical simulations using realistic channel models. The final conclusions of the paper are summarized in Section VIII.

The following notation is used. Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors, and italics denote scalars. The superscripts  $(\cdot)^T$ ,  $(\cdot)^*$ , and  $(\cdot)^H$  denote transpose, complex conjugate, and Hermitian operations, respectively. The determinant and trace of a matrix are denoted by  $\det(\cdot)$  and  $\text{tr}(\cdot)$ , respectively.  $[\mathbf{X}]_{ij}$  denotes the  $(i$ th,  $j$ th) element of matrix  $\mathbf{X}$ .  $\text{Re}(\cdot)$  and  $\text{Im}(\cdot)$  denote real and imaginary parts and  $\mathcal{E}[\cdot]$  and  $\mathcal{F}\{\cdot\}$  denote mathematical expectation and Fourier transform, respectively. We define  $(x)^+ \triangleq \max(0, x)$ .

## II. SIGNAL MODEL

A frequency-selective MIMO  $(n_T, n_R)$  channel corresponding to the generic situation of  $n_T$  transmit and  $n_R$  receive antennas (depicted in Fig. 1) can be represented by a channel matrix  $\mathbf{H}(\tau) \in \mathbb{C}^{n_R \times n_T}$  defined as

$$\mathbf{H}(\tau) = \sum_{l=1}^{L_\tau} \mathbf{A}^{(l)} \delta(\tau - \tau_l) \quad (1)$$

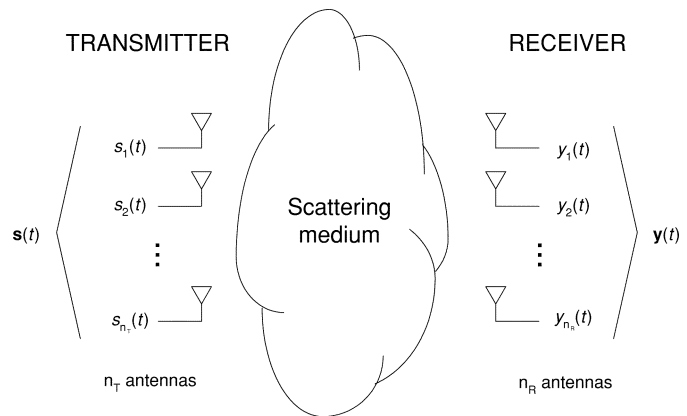


Fig. 1. General scheme of a multiantenna MIMO channel.

where  $L_\tau$  denotes the number of resolvable multipath components and matrix  $\mathbf{A}^{(l)}$  is the channel matrix at delay  $\tau_l$  defined as

$$\mathbf{A}^{(l)} = \begin{bmatrix} \alpha_{11}^{(l)} & \alpha_{12}^{(l)} & \cdots & \alpha_{1n_T}^{(l)} \\ \alpha_{21}^{(l)} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \alpha_{n_R 1}^{(l)} & \cdots & \cdots & \alpha_{n_R n_T}^{(l)} \end{bmatrix} \in \mathbb{C}^{n_R \times n_T}$$

where  $\alpha_{ij}^{(l)}$  is the complex transmission coefficient (fading or gain) from the  $j$ th transmit antenna to the  $i$ th receive antenna. The statistics and correlations of the fades, e.g., power delay profile and spatial correlation, make the channel representative of a specific to a particular environment such as indoor or outdoor.

The received signal can be expressed as

$$\mathbf{y}(t) = \mathbf{H}(t) * \mathbf{s}(t) + \mathbf{n}(t) \quad (2)$$

where  $\mathbf{s}(t) \in \mathbb{C}^{n_T \times 1}$ ,  $\mathbf{y}(t) \in \mathbb{C}^{n_R \times 1}$ , and  $\mathbf{n}(t) \in \mathbb{C}^{n_R \times 1}$  are the vector-valued transmitted signal, received signal, and noise, respectively. The noise  $\mathbf{n}(t)$  (possibly including interferences) is assumed to be a complex Gaussian stationary stochastic vector process with arbitrary spatial and temporal correlation  $\mathbf{R}_n(\tau) = \mathcal{E}[\mathbf{n}(t)\mathbf{n}^H(t-\tau)]$  and the convolution operator  $*$  is defined over matrices similarly to the matrix product, i.e.,  $[\mathbf{A}(t) * \mathbf{B}(t)]_{ij} \triangleq \int [\mathbf{A}(\tau)\mathbf{B}(t-\tau)]_{ij} d\tau$ .

Assuming the utilization of a matrix linear filter at the transmitter (also known as linear precoder), the transmitted signal can be written as (see Fig. 2)

$$\mathbf{s}(t) = \sum_k \mathbf{B}(t - kT)\mathbf{x}(k) \quad (3)$$

where  $\mathbf{x}(k) \in \mathbb{C}^{L \times 1}$  is the vector of symbols to be transmitted,  $\mathbf{B}(t) \in \mathbb{C}^{n_T \times L}$  is the space–time matrix precoder,  $T$  is the symbol period, and  $L$  is the number of simultaneous symbols transmitted or, equivalently, the number of spatial subchannels used. We can think of each column of  $\mathbf{B}(t)$  as a wideband beamvector associated to a symbol. In other words,

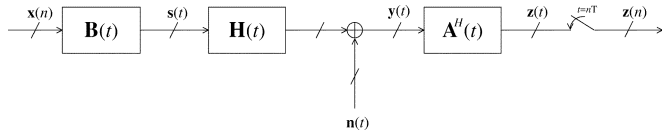


Fig. 2. Scheme of the transmit–receive linear filtering approach.

a space–time matrix filter represents a multiple beamforming scheme.

Similarly, assuming a matrix linear filter at the receiver (also known as equalizer), the processed and sampled signal is (see Fig. 2)

$$\mathbf{z}(n) = \mathbf{A}^H(t) * \mathbf{y}(t)|_{t=nT} \quad (4)$$

where  $\mathbf{A}^H(t) \in \mathbb{C}^{L \times n_R}$  is the space–time matrix equalizer. The detection of the symbols is then performed over  $\mathbf{z}(n)$ , possibly including a decision feedback (DF) equalizer.

In the case of simplified signal processing (i.e., beamforming), as will be seen in the sequel with detail, a single subchannel is used (i.e.,  $L = 1$ ), the transmit covariance matrix becomes rank-one and, therefore, the transmit and receive filters simplify to beamvectors  $\mathbf{b}(t) \in \mathbb{C}^{n_T \times 1}$  and  $\mathbf{a}(t) \in \mathbb{C}^{n_R \times 1}$ , respectively.

### III. ASYMPTOTIC OPTIMALITY OF BEAMFORMING IN CORRELATED MIMO CHANNELS

A frequency-selective MIMO channel with  $n_T$  transmit and  $n_R$  receive antennas has potentially  $K_f = \min(n_R, n_T, L_f^{(\text{rays})})$  parallel subchannels at frequency  $f$  [9], where  $L_f^{(\text{rays})}$  denotes the number of rays between the transmitter and the receiver at frequency  $f$ . In a rich scattering environment,  $L_f^{(\text{rays})} \gg \min(n_T, n_R)$  and, therefore, the number of potential parallel subchannels is  $K = \min(n_R, n_T)$  for the whole utilized bandwidth. The capacity of a MIMO channel is given by [22], [23]

$$C = \max_{\mathbf{Q}(f)} \int_W \log_2 \det(\mathbf{I}_{n_R} + \mathbf{R}_n^{-1}(f) \mathbf{H}(f) \mathbf{Q}(f) \mathbf{H}^H(f)) df$$

$$= \max_{\{P_i(f)\}} \sum_{i=1}^K \int_W \log_2 \left( 1 + \sigma_{\tilde{H},i}^2(f) P_i(f) \right) df \quad \text{bps} \quad (5)$$

where  $W$  is the bandwidth of the transmitted signal,  $\mathbf{Q}(f)$  and  $\mathbf{R}_n(f)$  are the power spectral density matrices of the transmitted signal and the received noise, respectively,  $\sigma_{\tilde{H},i}^2(f)$  denotes the  $i$ th diagonal element of matrix  $\Sigma_{\tilde{H}}(f)$  (with diagonal elements in decreasing order), which is obtained from the singular value decomposition (SVD) of the whitened channel  $\tilde{\mathbf{H}}(f) \triangleq \mathbf{R}_n^{-1/2}(f) \mathbf{H}(f) = \mathbf{U}_{\tilde{H}}(f) \Sigma_{\tilde{H}}(f) \mathbf{V}_{\tilde{H}}^H(f)$ , and  $P_i(f)$  represents the power allocated to the  $i$ th spatial subchannel. The maximization in (5) is over all  $\mathbf{Q}(f)$  that satisfy the average transmit power constraint given by

$$\int_W \text{tr} \mathbf{Q}(f) df = \sum_{i=1}^K \int_W P_i(f) df \leq P_{\text{av}} \quad (6)$$

where  $P_{\text{av}}$  is the maximum average transmit power in units of energy per second.

As can be seen from (5), the global channel capacity is given by the sum of the capacity of the  $K$  spatial subchannels. Therefore, in order to achieve capacity, all these constituent subchannels must be properly used by allocating the power according to the optimum waterfilling solution [23]

$$P_i(f) = \left( \mu - \frac{1}{\sigma_{\tilde{H},i}^2(f)} \right)^+ \quad (7)$$

where  $\mu$  is the “waterlevel” chosen to satisfy the power constraint of (6) with equality.

If the transmitter uses beamforming, then the transmit covariance matrix  $\mathbf{Q}(f)$  becomes rank-one and the maximum bit rate that can be achieved is [24]

$$C_{\text{bf}} = \max_{P(f)} \int_W \log_2 \left( 1 + \sigma_{\tilde{H},\max}^2(f) P(f) \right) df \quad (8)$$

which means that, at each frequency, only the spatial subchannel with highest gain is used. For a multiple beamforming scheme that uses  $L$  beamvectors simultaneously (with  $L \leq K$ ), the maximum achievable bit rate is given by an expression similar to (5) in which the sum is only over the  $L$  spatial subchannels with highest gain [24]. Clearly, beamforming is in principle a sub-optimum approach because it uses only one spatial subchannel at each frequency [compare (8) with (5)]. However, as we shall see, for some particular situations with high spatial correlation or low signal-to-interference-plus noise ratio (SINR), it may be optimum or almost optimum.

To analyze the effect of the spatial fading correlation, we model the random channel at delay  $\tau$  in the delay-tapped channel of (1) as  $\mathbf{H}_\tau = \Phi_R^{1/2} \mathbf{H}_{w,\tau} \Phi_T^{H/2}$  (see [3] for further details), where  $\mathbf{H}_{w,\tau}$  is a circularly symmetric complex Gaussian random matrix with independent and identically distributed (i.i.d.) entries, and  $\Phi_R$  and  $\Phi_T$  are fixed matrices that introduce correlation among the entries of  $\mathbf{H}_{w,\tau}$ . To be specific,  $\Phi_T = \Phi_T^{1/2} \Phi_T^{H/2}$  and  $\Phi_R = \Phi_R^{1/2} \Phi_R^{H/2}$  are the fading correlation matrices at the transmitter and at the receiver, respectively, ( $\Phi_T^{1/2}$  and  $\Phi_R^{1/2}$  represent one of the infinite ways to decompose  $\Phi_R$  and  $\Phi_T$ ). In the frequency domain, we have  $\mathbf{H}(f) = \Phi_R^{1/2} \mathbf{H}_w(f) \Phi_T^{H/2}$ . The rank of  $\Phi_T$  (with respect to  $\Phi_R$ ) decreases as the fading at the transmitter (receiver) becomes more correlated (a fully correlated fading corresponds to a rank-one correlation matrix). It then follows that for asymptotically fully correlated fading at either the transmit or the receive side, the channel matrix becomes asymptotically rank-one

$$\mathbf{H}(f) \longrightarrow \mathbf{h}_R(f) \mathbf{h}_T^H(f) \quad (9)$$

where  $\mathbf{h}_R(f) \in \mathbb{C}^{n_R \times 1}$  and  $\mathbf{h}_T(f) \in \mathbb{C}^{n_T \times 1}$  represent the transmit and receive spatial signatures at frequency  $f$ . In fact, using the more general MIMO channel model described in [25], it is possible to have rank-one channels even with uncorrelated fading at both the transmitter and the receiver (called “pin-hole” channels in [25]). The asymptotic channel given by (9) has a single nonvanishing singular value, in which case (5) and (8) are equivalent and, therefore, beamforming is asymptotically optimum as the spatial correlation increases.

Regardless of the spatial correlation, beamforming also becomes optimum for sufficiently low SINR. To be more specific,

beamforming is optimum if and only if the waterfilling power allocation in (7) presents at most one active spatial subchannel per frequency  $\sigma_{\tilde{H},2}^2(f) \leq \mu^{-1} \forall f$  (assuming eigenvalues in decreasing order), i.e., if

$$P_{\text{av}} \leq \frac{|\Omega|}{\max_{f \in W} \sigma_{\tilde{H},2}^2(f)} - \int_{\Omega} \frac{1}{\sigma_{\tilde{H},1}^2(f)} df \quad (10)$$

where  $\Omega$  denotes the set of active frequencies (with  $P_1(f) > 0$  according to the waterfilling solution in (7) over the highest spatial subchannels) and  $|\Omega| \triangleq \int_{\Omega} df$  is the size of  $\Omega$ .

Thus, for beamforming to become an optimum scheme, the channel matrix does not have to be strictly rank-one as indicated in (9). It suffices to have a fading correlation high enough to make  $\sigma_{\tilde{H},2}^2(f)$  sufficiently small so that (10) is satisfied (which also depends on  $P_{\text{av}}$  or, equivalently, on the SINR). As will be observed from numerical simulations in Section VII, for correlated scenarios with an angle spread (AS) of the order of  $8^\circ$ , (10) is indeed satisfied.

In [26] and [27], the optimality of beamforming was analyzed for different degrees of channel feedback quality. In [28], different transmission strategies are considered (including a beamforming approach) depending on the degree of channel knowledge at the transmitter and the spatial correlation. Interestingly, beamforming happens to be optimum for multiuser scenarios with a sufficiently large number of users regardless of the spatial correlation [29].

#### IV. JOINT TRANSMIT-RECEIVE BEAMFORMING DESIGN

The transmit precoding filter can be designed to achieve capacity and then the receive equalizer according to some criteria such as ZF or MMSE as was done in [19]. Another interesting alternative is to consider the signal constellation fixed (possibly obtained after some kind of optimization process) and perform a joint equalization of the channel, i.e., to jointly design the transmit and receive space-time filters according to some criteria such as ZF or MMSE as in [18] and [8], respectively.

Motivated by the previous result of the asymptotic optimality of beamforming, we focus on the joint design of transmit and receive beamvectors [see Fig. 3(a)]. Beamforming implies that a single symbol is transmitted per symbol period, i.e., that only one spatial subchannel is being used ( $L = 1$ ). For clarity of presentation, we consider Nyquist bandlimited systems, i.e., limited to  $[-1/2T, 1/2T)$ . The generalization to the case of excess-bandwidth systems is straightforward following the elegant approach of [8].

Given the transmitted signal  $\mathbf{s}(t) = \sum_k \mathbf{b}(t - kT)x(k)$ , the transmit power constraint is

$$\int_{-1/2T}^{1/2T} \frac{1}{T} S_{xx}(f) \mathbf{b}^H(f) \mathbf{b}(f) df \leq P_{\text{av}} \quad (11)$$

where  $\mathbf{b}(f) = \mathcal{F}\{\mathbf{b}(t)\}$  and  $S_{xx}(f) \triangleq \sum_{k=-\infty}^{\infty} \mathcal{E}[x(n)x^*(n-k)]e^{-j2\pi f k T}$  is the power spectral density of the data symbols  $x(n)$ . We similarly define  $\mathbf{H}(f) = \mathcal{F}\{\mathbf{H}(t)\}$  and  $\mathbf{a}^*(f) = \mathcal{F}\{\mathbf{a}^*(t)\}$ .

In the following sections, optimum transmit and receive wideband (frequency-selective) beamvectors  $\mathbf{b}(t)$  and  $\mathbf{a}(t)$  are

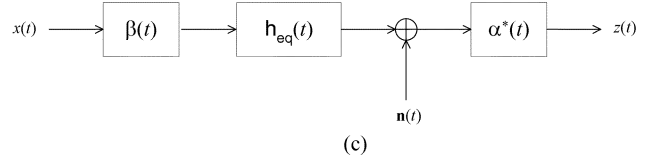
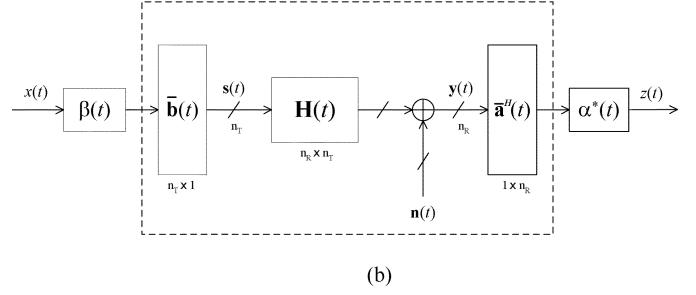
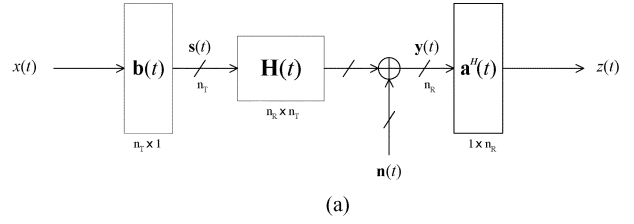


Fig. 3. Schematic of the flattening of the MIMO channel to an equivalent SISO channel. (a) Basic communication scheme. (b) Decomposed communication scheme. (c) Equivalent flattened channel scheme.

jointly designed according to the ZF and MMSE criteria. An arbitrary controlled intersymbol interference (ISI) commonly termed desired impulse response (DIR) or partial response is allowed in the design [30]. To start with, the partial response  $g(n)$  is assumed fixed and given by the design specifications. Section IV-C deals with the optimum choice of  $g(n)$ . Note that the classical criteria without partial response is obtained by choosing  $g(n) = \alpha_0 \delta(n - n_0)$ , where  $\alpha_0$  and  $n_0$  represent an arbitrary attenuation and delay respectively. We will make use of the definitions  $\tilde{S}_{xx}(f) \triangleq (1/T)S_{xx}(f)$  and  $X_f \triangleq X(f)$ .

##### A. Joint Tx-Rx Beamforming Design Under the PR-JZF Criterion

The partial response joint ZF (PR-JZF) criterion generalizes the classical ZF criterion by allowing any desired partial response  $g(n)$ . It is given by imposing  $\mathbf{a}^H(t) * \mathbf{H}(t) * \mathbf{s}(t)|_{t=nT} = g(n) * x(n)$ . In the frequency domain, this can be expressed as

$$\frac{1}{T} \mathbf{a}^H(f) \mathbf{H}(f) \mathbf{b}(f) = g(f), \quad f \in \left[-\frac{1}{2T}, \frac{1}{2T}\right). \quad (12)$$

For simplicity of notation, we define  $\tilde{g}(f) \triangleq T \cdot g(f)$ . It is assumed that  $\mathbf{H}(f) \neq \mathbf{0}$ ,  $f \in [-1/2T, 1/2T)$  (for excess-bandwidth systems this requirement is relaxed) so that (12) can be satisfied. We further impose the minimization of the noise power after filtering at the receiver

$$\min_{\mathbf{a}_f} \xi = \int_{-1/2T}^{1/2T} \mathbf{a}^H(f) \mathbf{R}_n(f) \mathbf{a}(f) df. \quad (13)$$

The Lagrangian containing (11)–(13) can be written as

$$\mathcal{L} = \int_{-1/2T}^{1/2T} \mathbf{a}_f^H \mathbf{R}_n(f) \mathbf{a}_f + \lambda_b \tilde{S}_{xx}(f) \mathbf{b}_f^H \mathbf{b}_f - \lambda_R(f) \cdot \text{Re}(\mathbf{a}_f^H \mathbf{H}_f \mathbf{b}_f) - \lambda_I(f) \text{Im}(\mathbf{a}_f^H \mathbf{H}_f \mathbf{b}_f) df. \quad (14)$$

Setting  $\partial \mathcal{L} / \partial \mathbf{a}_f^* = \mathbf{0}$  and using the ZF constraint of (12) to find the Lagrangian multipliers, we obtain the following expression for the receive beamvector as a function of the transmit beamvector:

$$\mathbf{a}_f = \frac{\tilde{g}_f^*}{\mathbf{b}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{b}_f} \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{b}_f. \quad (15)$$

Introducing (15) into the design equations and decomposing without loss of generality (w.l.o.g.) the transmit beamvector as  $\mathbf{b}_f = \beta_f \bar{\mathbf{b}}_f$ , where  $\|\bar{\mathbf{b}}_f\| = 1$ , we obtain the following constrained minimization:

$$\min_{\beta_f, \bar{\mathbf{b}}_f} \xi = \int \frac{|\tilde{g}_f|^2}{|\beta_f|^2 (\bar{\mathbf{b}}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \bar{\mathbf{b}}_f)} df \quad \text{s.t.} \quad \int |\beta_f|^2 \tilde{S}_{xx}(f) \leq P_{av}. \quad (16)$$

In order to minimize  $\xi$ ,  $\rho_f \triangleq \bar{\mathbf{b}}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \bar{\mathbf{b}}_f$  must equal the maximum eigenvalue of  $(\mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f)$ , being  $\bar{\mathbf{b}}_f$  its associated eigenvector. The error function  $\xi$  becomes then a convex function. Again, forming the Lagrangian and setting  $\partial \mathcal{L} / \partial |\beta_f|^2 = 0$ , the following solution is obtained:

$$|\beta_f|^2 = \mu_{ZF}^{-1/2} \sqrt{\frac{|\tilde{g}_f|^2}{\rho_f \tilde{S}_{xx}(f)}} \quad (17)$$

where  $\mu_{ZF}^{1/2} = (1/P_{av}) \int \sqrt{|\tilde{g}_f|^2 \tilde{S}_{xx}(f) / \rho_f} df$  is found by imposing the power constraint with equality because the error function  $\xi$  is monotonic decreasing in  $P_{av}$ . Note that the phase of  $\beta_f$  does not affect the solution and can be freely chosen.

### B. Joint Tx–Rx Beamforming Design Under the PR-JMMSE Criterion

The partial response-joint (MMSE) criterion generalizes the classical MMSE criterion by allowing any desired partial response  $g(n)$ . It is based on the minimization of

$$\xi = \mathcal{E} \left[ |z(n) - g(n) * x(n)|^2 \right] \quad (18)$$

where  $z(n)$  is given by (4). The Lagrangian containing (18) in the frequency domain along with the average transmit power constraint of (11) is

$$\mathcal{L} = \int_{-1/2T}^{1/2T} \left( |\tilde{g}(f) - \mathbf{a}_f^H \mathbf{H}_f \mathbf{b}_f|^2 \tilde{S}_{xx}(f) + \mathbf{a}_f^H \mathbf{R}_n(f) \mathbf{a}_f + \lambda_b \tilde{S}_{xx}(f) \mathbf{b}_f^H \mathbf{b}_f \right) df. \quad (19)$$

Setting  $\partial \mathcal{L} / \partial \mathbf{a}_f^* = \mathbf{0}$ , we obtain

$$\mathbf{a}_f = \left( \mathbf{H}_f \mathbf{b}_f \mathbf{b}_f^H \mathbf{H}_f^H \tilde{S}_{xx}(f) + \mathbf{R}_n(f) \right)^{-1} \mathbf{H}_f \mathbf{b}_f \tilde{S}_{xx}(f) \tilde{g}_f^* \quad (20)$$

$$= \frac{\tilde{g}_f^*}{\mathbf{b}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{b}_f + \tilde{S}_{xx}^{-1}(f)} \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{b}_f \quad (21)$$

where we have used the matrix inversion lemma.<sup>1</sup> Using (20) into the MSE of (18) along with the power constraint of (11) reduces to the following constrained minimization problem:

$$\min_{\beta_f, \bar{\mathbf{b}}_f} \xi = \int \frac{|\tilde{g}_f|^2}{|\beta_f|^2 (\bar{\mathbf{b}}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \bar{\mathbf{b}}_f) + \tilde{S}_{xx}^{-1}(f)} df \quad \text{s.t.} \quad \int |\beta_f|^2 \tilde{S}_{xx}(f) \leq P_{av}. \quad (22)$$

In this case, again,  $\rho_f \triangleq \bar{\mathbf{b}}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \bar{\mathbf{b}}_f$  must equal the maximum eigenvalue of  $(\mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f)$  and  $\bar{\mathbf{b}}_f$  must be its corresponding eigenvector.

The optimal  $|\beta_f|^2$  that minimizes the convex function  $\xi$  subject to the convex constraint (11) is given by the following modified waterfilling solution [31]:

$$|\beta_f|^2 = \left( \mu_{\text{MMSE}}^{-1/2} \sqrt{\frac{|\tilde{g}_f|^2}{\rho_f \tilde{S}_{xx}(f)}} - \frac{1}{\rho_f \tilde{S}_{xx}(f)} \right)^+ \quad (23)$$

where

$$\mu_{\text{MMSE}}^{1/2} = (1/P_{av} + \int_{\Omega} \rho_f^{-1} df) \int_{\Omega} \sqrt{(|\tilde{g}_f|^2 \tilde{S}_{xx}(f) / \rho_f)} df$$

(chosen so that the power constraint is satisfied with equality since the error function  $\xi$  is monotonically decreasing in  $P_{av}$ ) and  $\Omega$  denotes the set of frequencies for which  $|\beta_f|^2 \neq 0$ . Again, the phase of  $\beta_f$  is not defined and can be freely chosen. The solution corresponding to (23) can be solved as in [10], [11], and [8], where a parametric approach is given by expressing the MSE and the transmit power as a function of the parameter  $\mu$  and then selecting one point of the curve. This method becomes an iterative solution when having a given transmit power. In Section VI, a nonparametric and exact algorithm is given.

*JMMSE Versus JZF:* In classical receive-only filter design, it is well-known that as the noise power goes to zero (in our case, as the SINR goes to infinity), the MMSE solution tends to the ZF one [32]. For the case of joint transmit–receive design, this assertion holds as well, as we show next. As will be further analyzed in Section IV-D,  $\rho_f$  is the channel frequency-dependent effective gain. For  $\rho_f \rightarrow \infty$ , the expression of  $\mathbf{a}_f$  in (21) clearly tends to (15). As for  $\mathbf{b}_f$ , the normalized beamvector  $\bar{\mathbf{b}}_f$  is clearly the same in both criteria [the eigenvector corresponding to the maximum eigenvalue of  $(\mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f)$ ] and the scaling factor  $|\beta_f|^2$  is asymptotically equal. Indeed, from (23)

$$|\beta_f|^2 = \mu_{\text{MMSE}}^{-1/2} \sqrt{\frac{|\tilde{g}_f|^2}{\rho_f \tilde{S}_{xx}(f)}} \times \left( 1 - \mu_{\text{MMSE}}^{1/2} \sqrt{\frac{1}{\rho_f |\tilde{g}_f|^2 \tilde{S}_{xx}(f)}} \right)^+ \rightarrow \mu_{\text{ZF}}^{-1/2} \sqrt{\frac{|\tilde{g}_f|^2}{\rho_f \tilde{S}_{xx}(f)}}$$

where we have used the fact that  $\mu_{\text{MMSE}} \rightarrow \mu_{\text{ZF}}$ .

<sup>1</sup>Matrix Inversion Lemma:  $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{D} \mathbf{A}^{-1} \mathbf{B} + \mathbf{C}^{-1})^{-1} \mathbf{D} \mathbf{A}^{-1}$ .

### C. Design of the Partial Response

Thus far, we have assumed that the partial response filter  $g(f)$  used in (12) and (18) was given and fixed by the design specifications (e.g., we may have a detector matched to a given partial response such as the duobinary channel  $1 + D$ , or the DC-notch channel  $1 - D$  [30]). However, we can also consider  $g(f)$  as additional degrees of freedom in the design procedure to improve performance. In that case, the resulting  $g(f)$  should be taken into account by the subsequent detector. In principle, we are interested in a symbol-by-symbol detector, possibly including a DF block (the DF filter is given by the strictly causal filter  $g(f) - 1$ ). A DF scheme is computationally appealing, although it may present problems of error propagation at low SINR. It is also possible to use a maximum-likelihood sequence estimator (MLSE). In that case, however, the previous derivations of transmit and receive filters are not optimal anymore since the receive filter correlates the noise and (13) or (18) do not correspond to the optimal metric of the MLSE (they correspond instead to a MLSE that ignores the temporal noise correlation).

To optimize  $g(f)$ , we first rewrite the error function of (16) and (22) as

$$\xi = \int \frac{|\tilde{g}(f)|^2}{q(f)} df \quad (24)$$

where  $q(f) \triangleq \mathbf{b}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{b}_f + \kappa \tilde{S}_{xx}^{-1}(f)$  with  $\kappa = 0$  for the ZF case and  $\kappa = 1$  for the MMSE case. Using well-known results on spectral factorization theory [30] (see also [33]),  $q(f)$  can be written as

$$q(f) = s_o |v(f)|^2$$

where  $v(f)$  is canonical, i.e., causal, monic and minimum phase, and  $s_o$  is given by<sup>3</sup>

$$\log s_o = T \int_{-1/2T}^{1/2T} \log q(f) df.$$

Using the Schwarz inequality, it can be shown that the error function in (24) is minimized for  $\tilde{g}(f) = v(f)$  and is given by

$$\begin{aligned} \xi &= \int_{-1/2T}^{1/2T} \frac{1}{s_o} df \\ &= \frac{1}{T} e^{-T \int_{-1/2T}^{1/2T} \log(\mathbf{b}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{b}_f + \kappa \tilde{S}_{xx}^{-1}(f)) df}. \end{aligned}$$

The previous error is minimized when  $\mathbf{b}_f$  lies along the direction of the eigenvector corresponding to the maximum eigenvalue of  $(\mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f)$  and with squared norm given by

$$|\beta_f|^2 = \left( \mu \frac{1}{\tilde{S}_{xx}(f)} - \kappa \frac{1}{\rho_f \tilde{S}_{xx}(f)} \right)^+$$

where  $\mu$  is chosen to satisfy the power constraint of (11) with equality.

<sup>2</sup>This implies that  $g(f)$  has to be a monic filter, i.e., with the first tap equal to one.

<sup>3</sup>We assume that  $q(f)$  is not zero over any measurable interval so that the spectral factorization exists [30].

### D. Conceptual Interpretation of the Joint Beamforming Scheme

To get insight into the solutions obtained for the joint wide-band beamforming design, we decompose the frequency-dependent beamvectors as follows:

$$\begin{cases} \mathbf{a}_f = \alpha_f \bar{\mathbf{a}}_f \\ \mathbf{b}_f = \beta_f \bar{\mathbf{b}}_f \end{cases}$$

where  $\bar{\mathbf{a}}_f$  and  $\bar{\mathbf{b}}_f$  contain the spatial structure of the beamvectors and are normalized w.l.o.g. so that  $\bar{\mathbf{a}}_f^H \mathbf{R}_n(f) \bar{\mathbf{a}}_f = 1$  and  $\bar{\mathbf{b}}_f^H \bar{\mathbf{b}}_f = 1$ , and  $\alpha_f$  and  $\beta_f$  are frequency-dependent scaling factors [see Fig. 3(b) where  $\bar{\mathbf{b}}(t) \triangleq \mathcal{F}^{-1}\{\bar{\mathbf{b}}(f)\}$ ,  $\bar{\mathbf{a}}^*(t) \triangleq \mathcal{F}^{-1}\{\bar{\mathbf{a}}^*(f)\}$ ].

The normalized receive beamvector  $\bar{\mathbf{a}}_f$ , corresponding to both the ZF (15) and the MMSE criterion (21), coincides and is given by

$$\bar{\mathbf{a}}_f = \frac{\mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{b}_f}{\sqrt{\mathbf{b}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{b}_f}}.$$

The normalized transmit beamvector  $\bar{\mathbf{b}}_f$  is also identical for both criteria and corresponds to the eigenvector associated with the maximum eigenvalue of  $(\mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f)$ . Interestingly, the expressions obtained for the normalized beamvectors  $\bar{\mathbf{a}}_f$  and  $\bar{\mathbf{b}}_f$  maximize the SINR and the mutual information at frequency  $f$ .

The frequency-dependent SINR can then be expressed as

$$\text{SINR}(f) = \frac{|\mathbf{a}_f^H \mathbf{H}_f \mathbf{b}_f|^2}{\mathbf{a}_f^H \mathbf{R}_n(f) \mathbf{a}_f} \tilde{S}_{xx}(f) = \rho(f) P(f)$$

where  $\rho(f) \triangleq |\bar{\mathbf{a}}_f^H \mathbf{H}_f \bar{\mathbf{b}}_f|^2 / \bar{\mathbf{a}}_f^H \mathbf{R}_n(f) \bar{\mathbf{a}}_f = \bar{\mathbf{b}}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \bar{\mathbf{b}}_f$  is the effective gain of the spatially flattened channel and  $P(f) = |\beta_f|^2 \tilde{S}_{xx}(f)$  is the allocated power at frequency  $f$ .

We can conclude that the normalized beamvectors are equivalent for all criteria and essentially select the constituent spatial SISO subchannel arising within the MIMO channel with highest gain [see Fig. 3(c)]. Equivalently, we can say that the MIMO channel is spatially flattened in the sense that it is reduced to a SISO one,  $h_{eq}(f) = \bar{\mathbf{a}}_f^H \mathbf{H}_f \bar{\mathbf{b}}_f$ , with maximum effective gain. The scaling factors are regular filters that operate on the equivalent spatially flattened channel [see Fig. 3(c)]. These scaling factors differ for different criteria such as ZF, MMSE, and capacity-achieving solution.

## V. EXTENSION TO MULTIPLE-BEAMFORMING DESIGN

In this section, we generalize the derivation of the transmit-receive filters of the previous section to the case of multiple beamforming, i.e., matrix filters  $\mathbf{B}(t)$  and  $\mathbf{A}(t)$  as opposed to vector filters. The average transmit power constraint is given by

$$\int_{-1/2T}^{1/2T} \text{tr}(\mathbf{B}(f) \tilde{\mathbf{R}}_x(f) \mathbf{B}^H(f)) df \leq P_{av} \quad (25)$$

where  $\tilde{\mathbf{R}}_x(f) \triangleq (1/T) \mathbf{R}_x(f)$  and  $\mathbf{R}_x(f) \triangleq \sum_{k=-\infty}^{\infty} \mathcal{E}[\mathbf{x}(n) \mathbf{x}^H(n-k)] e^{-j2\pi f k T}$  is the power spectral density matrix of the data vector  $\mathbf{x}(n)$ .

We first formulate the problem under the PR-JZF criterion, which is an extension of the result obtained in [18] where a different derivation was done based on the maximization of the SINR subject to the ZF constraint. The ZF criterion with arbitrary partial response  $\mathbf{G}(n)$  can be expressed in the frequency domain as

$$\frac{1}{T} \mathbf{A}^H(f) \mathbf{H}(f) \mathbf{B}(f) = \mathbf{G}(f) \quad f \in \left[-\frac{1}{2T}, \frac{1}{2T}\right) \quad (26)$$

and the noise minimization at the receiver is

$$\min_{\mathbf{A}(f)} \xi_{\text{ZF}} = \int_{-1/2T}^{1/2T} \text{tr} \left( \mathbf{A}^H(f) \mathbf{R}_n(f) \mathbf{A}(f) \right) df. \quad (27)$$

The Lagrangian containing (25)–(27) can be written as in (28), shown at the bottom of the page from which  $\mathbf{A}_f$  can be found as a function of  $\mathbf{B}_f$  by setting  $\partial \mathcal{L} / \partial \mathbf{A}_f^* = \mathbf{0}$  and using the constraints to find the Lagrangian multipliers

$$\mathbf{A}_f = \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{B}_f \left( \mathbf{B}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{B}_f \right)^{-1} \tilde{\mathbf{G}}_f^H \quad (29)$$

where  $\tilde{\mathbf{G}}_f \triangleq T \cdot \mathbf{G}_f$ .

Alternatively, the problem can be formulated under the PR-JMMSE criterion. It is based on the minimization of the MSE  $\mathcal{E} \|\mathbf{z}(n) - \mathbf{G}(n) * \mathbf{x}(n)\|^2$ . The solution to this problem for the particular case of  $\mathbf{G}(n) = \alpha_0 \mathbf{I}_L \delta(n - n_0)$  was obtained in [8] by proving a series of lemmas on the diagonality of some matrix expressions. In [34], instead of minimizing the Euclidean norm of the error vector (equivalently, the trace of the error matrix), a unified framework that includes most reasonable objective functions is developed based on majorization theory. In the following, we make use of a basic result of the theory of majorization [35] and provide a compact and simple original derivation of the solution for a general choice of the partial response  $\mathbf{G}(n)$  (which includes DF schemes).<sup>4</sup> The MSE can be written in the frequency domain as shown in (30), at the bottom of the page, where  $\mathbf{R}_y(f) = \mathbf{H}_f \mathbf{B}_f \tilde{\mathbf{R}}_x(f) \mathbf{B}_f^H \mathbf{H}_f^H + \mathbf{R}_n(f)$ . Again, we can obtain  $\mathbf{A}_f$  as a function of  $\mathbf{B}_f$  as

$$\mathbf{A}_f = \left( \mathbf{H}_f \mathbf{B}_f \tilde{\mathbf{R}}_x(f) \mathbf{B}_f^H \mathbf{H}_f^H + \mathbf{R}_n(f) \right)^{-1} \mathbf{H}_f \mathbf{B}_f \tilde{\mathbf{R}}_x(f) \tilde{\mathbf{G}}_f^H. \quad (31)$$

<sup>4</sup>The considered design accounting for an arbitrary partial response is mathematically equivalent to having as a design criterion a weighted trace of the error matrix. We would like to thank C. Aldana for pointing out the existence of [36] (at that time unpublished), where the weighted trace criterion is considered and solved using the Karush–Kuhn–Tucker conditions arising in convex optimization theory.

To obtain the optimal  $\mathbf{B}_f$ , we consider the PR-JZF and the PR-JMMSE criteria in a unified way. Plugging (29) and (31) into (27) and (30), respectively, (and applying the matrix inversion lemma in the PR-JMMSE method), the noise error in the PR-JZF criterion (27) and the MSE in the PR-JMMSE criterion (30) are obtained setting  $\kappa = 0$  and  $\kappa = 1$ , respectively, in

$$\xi = \int_{-1/2T}^{1/2T} \text{tr} \left( \mathbf{B}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \mathbf{B}_f + \kappa \tilde{\mathbf{R}}_x^{-1}(f) \right)^{-1} \mathbf{R}_{\mathbf{G}_f}(f) df \quad (32)$$

where  $\mathbf{R}_{\mathbf{G}_f} \triangleq \tilde{\mathbf{G}}_f^H \tilde{\mathbf{G}}_f$ .

Defining  $\tilde{\mathbf{B}}_f \triangleq \mathbf{B}_f \tilde{\mathbf{R}}_x^{1/2}(f)$  and  $\mathbf{R}_s(f) \triangleq \tilde{\mathbf{R}}_x^{1/2}(f) \mathbf{R}_{\mathbf{G}_f} \tilde{\mathbf{R}}_x^{1/2}(f)$ , the problem can be written as

$$\min_{\tilde{\mathbf{B}}_f} \int \text{tr} \left( \tilde{\mathbf{B}}_f^H \mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f \tilde{\mathbf{B}}_f + \kappa \mathbf{I} \right)^{-1} \mathbf{R}_s(f) df \quad \text{s.t.} \quad \int \text{tr} \left( \tilde{\mathbf{B}}_f \tilde{\mathbf{B}}_f^H \right) df \leq P_{\text{av}}. \quad (33)$$

In Appendix A, it is shown that the optimal  $\tilde{\mathbf{B}}_f$  is (if  $L > n_T$ , instead of  $L$  use  $\tilde{L} \triangleq \min(L, n_T)$ )

$$\tilde{\mathbf{B}}_f = \mathbf{U}_f \boldsymbol{\Sigma}_{B_f} \mathbf{V}_f^H \quad (34)$$

where  $\mathbf{U}_f$  and  $\mathbf{V}_f$  are unitary matrices that diagonalize  $\mathbf{R}_s(f)$  and  $(\mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f)$ , respectively

$$\mathbf{V}_f^H \mathbf{R}_s(f) \mathbf{V}_f = \mathbf{D}_s(f) \quad (35)$$

$$\mathbf{U}_f^H (\mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f) \mathbf{U}_f = \mathbf{D}_n^{-1}(f) \quad (36)$$

where the eigenvalues of  $\mathbf{R}_s(f)$  and  $(\mathbf{H}_f^H \mathbf{R}_n^{-1}(f) \mathbf{H}_f)$  (equivalently, the diagonal elements of  $\mathbf{D}_s(f)$  and  $\mathbf{D}_n^{-1}(f)$ ) are denoted by  $\lambda_{s,i}(f)$  and  $\lambda_{n,i}^{-1}(f)$ , respectively. The only unknowns now are the diagonal elements of  $\boldsymbol{\Sigma}_{B_f}(f)$  denoted by  $\sigma_{B,i}(f)$ . The constrained minimization problem can be finally written in convex form as

$$\min_{\{\sigma_{B,i}\}} \xi = \int \sum_{i=1}^L \frac{\lambda_{s,i}}{\left( \frac{\sigma_{B,i}^2}{\lambda_{n,i}} + \kappa \right)} df \quad \text{s.t.} \quad \int \sum_{i=1}^L \sigma_{B,i}^2 df \leq P_{\text{av}} \quad (37)$$

where the dependence on the frequency has been omitted for simplicity of notation. Assuming that  $\lambda_{s,i}$  are in decreasing

---


$$\mathcal{L} = \int_{-1/2T}^{1/2T} \text{tr} \left( \mathbf{A}_f^H \mathbf{R}_n(f) \mathbf{A}_f + \lambda_b \mathbf{B}(f) \tilde{\mathbf{R}}_x(f) \mathbf{B}^H(f) - \text{Re} \left( \mathbf{A}_f^H \mathbf{H}_f \mathbf{B}_f \right) \boldsymbol{\Lambda}_R(f) - \text{Im} \left( \mathbf{A}_f^H \mathbf{H}_f \mathbf{B}_f \right) \boldsymbol{\Lambda}_I(f) \right) df \quad (28)$$


---

$$\xi_{\text{MMSE}} = \int_{-1/2T}^{1/2T} \text{tr} \left( \tilde{\mathbf{G}}_f \tilde{\mathbf{R}}_x(f) \tilde{\mathbf{G}}_f^H + \mathbf{A}_f^H \mathbf{R}_y(f) \mathbf{A}_f - \mathbf{A}_f^H \mathbf{H}_f \mathbf{B}_f \tilde{\mathbf{R}}_x(f) \tilde{\mathbf{G}}_f^H - \tilde{\mathbf{G}}_f \tilde{\mathbf{R}}_x(f) \mathbf{B}_f^H \mathbf{H}_f^H \mathbf{A}_f \right) df \quad (30)$$

order w.l.o.g., it follows that the error  $\xi$  is minimized for  $\lambda_{n,i}$  in increasing order (see Appendix B). The optimal  $\{\sigma_{B,i}\}$  that minimizes the convex function  $\xi$  subject to the convex region corresponding to the power constraint (25) is given similarly to (17) and (23)

$$\sigma_{B,i}^2 = \begin{cases} \mu_{\text{ZF}}^{-1/2} \sqrt{\lambda_{s,i} \lambda_{n,i}}, & \text{for } \kappa = 0 \text{ (PR-JZF)} \\ \left( \mu_{\text{MMSE}}^{-1/2} \sqrt{\lambda_{s,i} \lambda_{n,i}} - \lambda_{n,i} \right)^+, & \text{for } \kappa = 1 \text{ (PR-JMMSE)} \end{cases} \quad (38)$$

where  $\mu_{\text{ZF}}^{1/2} = (1/P_{\text{av}}) \sum_{i=1}^L \int \sqrt{\lambda_{s,i} \lambda_{n,i}} df$ ,  $\mu_{\text{MMSE}}^{1/2} = (1/(P_{\text{av}} + \sum_i \int_{\Omega_i} \lambda_{n,i})) \sum_i \int_{\Omega_i} \sqrt{\lambda_{s,i} \lambda_{n,i}} df$  (both chosen to satisfy the power constraint with equality), and  $\Omega_i$  denotes the set of frequencies for which  $\sigma_{B,i}^2 > 0$ . Although  $\sigma_{B,i}(f)$  is assumed real, an arbitrary phase can be absorbed in  $\mathbf{U}_f$  or  $\mathbf{V}_f$ .

The modified waterfilling solution resulting from the PR-JMMSE criterion can be efficiently implemented using the algorithm given in Section VI. The partial response  $\mathbf{G}(n)$  for the multiple beamforming case can also be designed using the theory of spectral factorization on matrices as in Section IV-C. For more details, the reader is referred to [15].

## VI. PRACTICAL WATERFILLING ALGORITHM

In this section, an efficient and exact practical implementation of a generalized modified waterfilling algorithm is derived based on [21, Ch. 4] and [19, Alg. 1].

The design equations for the PR-JMMSE criterion (23) and (38), noting that in a practical implementation the continuous frequency domain is approximated by a finite set of frequency bins, can be cast into the following general form:

$$\begin{cases} E_k = (\mu b_k - a_k)^+ & 1 \leq k \leq N \\ \sum_{k=1}^N E_k = E_T \end{cases} \quad (39)$$

where  $\mu$  is a constant (commonly termed ‘‘waterlevel’’) selected to satisfy the constraint. It is important to remark that the waterfilling solution of (39) is a generalization of the classical capacity-achieving waterfilling solution which is obtained by particularizing  $b_k = 1$  as in [21, Ch. 4] and [19, Alg. 1].

We can proceed in the same way as in [21, Ch. 4] by making the hypothesis that all subchannels are active and expressing (39) in matrix form as

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & -b_1 \\ 0 & 1 & \cdots & 0 & -b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -b_N \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_N \\ \mu \end{bmatrix} = \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_N \\ E_T \end{bmatrix}.$$

Summing the first  $N$  rows and using  $\sum_{k=1}^N E_k = E_T$ , the waterlevel  $\mu$  is obtained as

$$\mu = \frac{E_T + \sum_k a_k}{\sum_k b_k}$$

from which the potential values of the energies  $E_k$  for  $1 \leq k \leq N$  is straightforwardly computed as  $E_k = \mu b_k - a_k$ .

If all the energies are nonnegative, the problem is solved. Otherwise, some of the  $N$  subchannels have to be removed, i.e., their energy has to be set to zero, and the process has to be repeated until no negative energy arises. To know the order in which the channels have to be removed, we first obtain the following lemma.

*Lemma 1:* Given two lists of positive numbers  $\{a_k\}$  and  $\{b_k\}$ , and a set of energies  $E_k = \mu b_k - a_k$ , where  $\mu = (E_T + \sum_k a_k) / \sum_k b_k$  ( $\sum_k E_k = E_T$ ), if  $a_i$  and  $b_i$  for a given  $i$  are removed from the lists, then  $\mu$  increases if  $E_i = \mu b_i - a_i > 0$  and  $\mu$  decreases if  $E_i = \mu b_i - a_i < 0$ .

*Proof:* This result is very intuitive since it is simply saying that if a subchannel with a positive allocated energy is removed, this energy can be reallocated over the rest of subchannel (i.e.,  $\mu$  is increased) and vice-versa.

The value of  $\mu$  after the deletion of  $a_i$  and  $b_i$  is

$$\mu^{\text{new}} = \frac{E_T + \sum_k a_k - a_i}{\sum_k b_k - b_i} = \mu^{\text{old}} \cdot \gamma$$

where

$$\gamma = \frac{1 - \frac{a_i}{E_T + \sum_k a_k}}{1 - \frac{b_i}{\sum_k b_k}}.$$

Clearly,  $\mu^{\text{new}} > \mu^{\text{old}}$  if and only if  $\gamma > 1$ , which in turn implies that  $a_i / (E_T + \sum_k a_k) < b_i / \sum_k b_k$  or, equivalently,  $a_i < b_i \mu^{\text{old}}$  and, therefore,  $0 < \mu^{\text{old}} b_i - a_i = E_i$ . Thus,  $\mu^{\text{new}} > \mu^{\text{old}}$  if and only if  $E_i > 0$  and vice-versa. ■

Using the previous lemma, we can state that a channel  $i$  with positive energy cannot be removed, because  $\mu$  would increase and then  $\mu b_i - a_i$  would still be positive which is a contradiction with  $E_i = (\mu b_i - a_i)^+$ . On the other hand, if a channel  $j$  with negative energy is deleted, then  $\mu$  decreases and, consequently,  $\mu b_j - a_j$  still remains negative which means that  $(\mu b_j - a_j)^+ = 0$  holds and will hold while channels with negative energy are removed. In light of those results, a first version of this modified waterfilling algorithm is summarized in Table I.

It is also possible to derive another version of the algorithm based on [21, Ch. 4] by noting that the channel with highest quotient  $a_k/b_k$  is the first one to become negative when decreasing  $\mu$ . To be more exact, since  $E_i/b_i = \mu - a_i/b_i$ , we can clearly state that for  $a_i/b_i > a_j/b_j$  it follows that if  $E_j$  is negative so is  $E_i$  and, conversely, if  $E_i$  is positive so is  $E_j$ . Therefore, a second version of the modified waterfilling algorithm is summarized in Table II.

Note that both algorithms require  $N$  loops or iterations in the worst-case to obtain the exact solution. This is in contrast to the existing iterative approaches for the modified waterfilling solution [8], [10], [11] that converge to the optimum solution as the number of iterations goes to infinity (they select a value for  $\mu$  and then compute  $\sum_k E_k$ , if it is greater than  $E_T$  then the constant  $\mu$  is reduced, otherwise it is increased and so forth).



TABLE I  
FIRST VERSION OF THE MODIFIED WATERFILLING ALGORITHM

- 
1. Compute list of pairs  $\{(a_k, b_k)\}$  for  $k \in \mathcal{S} = \{1, \dots, N\}$ .
  2. Calculate current value of  $\mu$  according to:  $\mu = \frac{E_T + \sum_{k \in \mathcal{S}} a_k}{\sum_{k \in \mathcal{S}} b_k}$ .
  3. Compute the energies using the current  $\mu$  as:  $E_k = \mu b_k - a_k, \quad k \in \mathcal{S}$ .
  4. If all the energies in  $\mathcal{S}$  are non-negative (*i.e.*,  $E_k \geq 0, \forall k \in \mathcal{S}$ ), then finish.  
Otherwise, remove the channels with negative energy:  $\mathcal{S}^- = \{k \in \mathcal{S} : E_k < 0\}$ ,  

$$E_k = 0 \quad k \in \mathcal{S}^-,$$

$$\mathcal{S} \leftarrow \mathcal{S} - \mathcal{S}^-,$$
and go to step 2.
- 

TABLE II  
SECOND VERSION OF THE MODIFIED WATERFILLING ALGORITHM

- 
1. Compute list of pairs  $\{(a_k, b_k)\}$  for  $1 \leq k \leq N$ .
  2. Reorder the list of pairs  $\{(a_k, b_k)\}$  so that  $a_k/b_k$  are in increasing order.
  3. Set  $\tilde{N} = N$  (number of active subchannels).
  4. Calculate current value of  $\mu$  according to:  $\mu = \frac{E_T + \sum_{k=1}^{\tilde{N}} a_k}{\sum_{k=1}^{\tilde{N}} b_k}$ .
  5. If  $E_{\tilde{N}} = \mu b_{\tilde{N}} - a_{\tilde{N}} < 0$ , then  $\tilde{N} = \tilde{N} - 1$  and go to step 4.  
Otherwise, compute all the energies according to  $E_k = (\mu b_k - a_k)^+$  and finish.
- 

## VII. SIMULATION RESULTS

In this section, we first consider a simple and illustrative example that will allow us to gain insight into the problem and then we deal with a realistic scenario (using a channel model based on field measurements).

### A. Illustrative Example

For this example, a simple MIMO (2,2) channel was artificially generated in order to obtain visual and qualitative results on the two-stage process described in Section IV-D (see Fig. 3). The two stages of the process are shown in Fig. 4. The first stage, common for all criteria, is the flattening of the MIMO channel, whereas the second stage consist on the power distribution which depends on each particular criterion. From Fig. 4, it can be observed that whereas the capacity-achieving solution (or unconstrained beamforming) and the JMMSE method allocate the transmitted power in an intelligent way according to the gain at each frequency by not allocating power to highly attenuated areas, the JZF criterion attempts to perfectly equalize the spatially flattened channel by allocating more power to highly attenuated areas and therefore enhancing the noise. Although the JZF criterion may in general perform similarly to the JMMSE, it fails when strong nulls exists; in other words, it lacks robustness.

### B. Realistic Simulations

We now consider a realistic scenario based on channel models obtained from field measurements. Since the interest of this

paper is on beamforming for spatially correlated MIMO channels, we first analyze the goodness of beamforming for different degrees of spatial correlation and then evaluate and compare the proposed JZF and JMMSE beamforming methods.

Regarding the scenario for the simulations, realistic MIMO (4,4) channels were generated for a typical situation of a communication between a mobile unit (MU) and a base station (BS). The MU is generally immerse in a rich scattering environment due to the presence of multiple objects (buildings, cars, etc.), having therefore an almost uncorrelated fading among the antennas for a half-wavelength separation. On the other hand, the BS is assumed to be on top of a building, receiving the signal transmitted by the MU from a mean direction of arrival (DoA) and within a specific AS, which depends basically on the height of the antenna elements of the BS. The fading correlation at the BS is considered a function of the AS, the antenna spacing and the mean DoA according to [37], where a uniform-shaped angular distribution was assumed to derive closed-form expressions. The mean DoA was drawn from a uniform distribution on  $[-60^\circ, 60^\circ]$  (assuming a three-sectorial cell deployment) and a half-wavelength separation of the antennas was assumed. Given the envelope correlation matrices at both ends of the link [37], the spatially correlated MIMO channel is generated as explained in [38]. No interfering signals were considered in the simulations. The temporal dispersion of the channel was generated following a vehicular power delay profile as specified by ETSI [39]. The matrix channel generated was normalized so that  $\sum_n \mathcal{E} |\mathbf{H}_{ij}(n)|^2 = 1$  and is assumed perfectly known. The transmission block is assumed

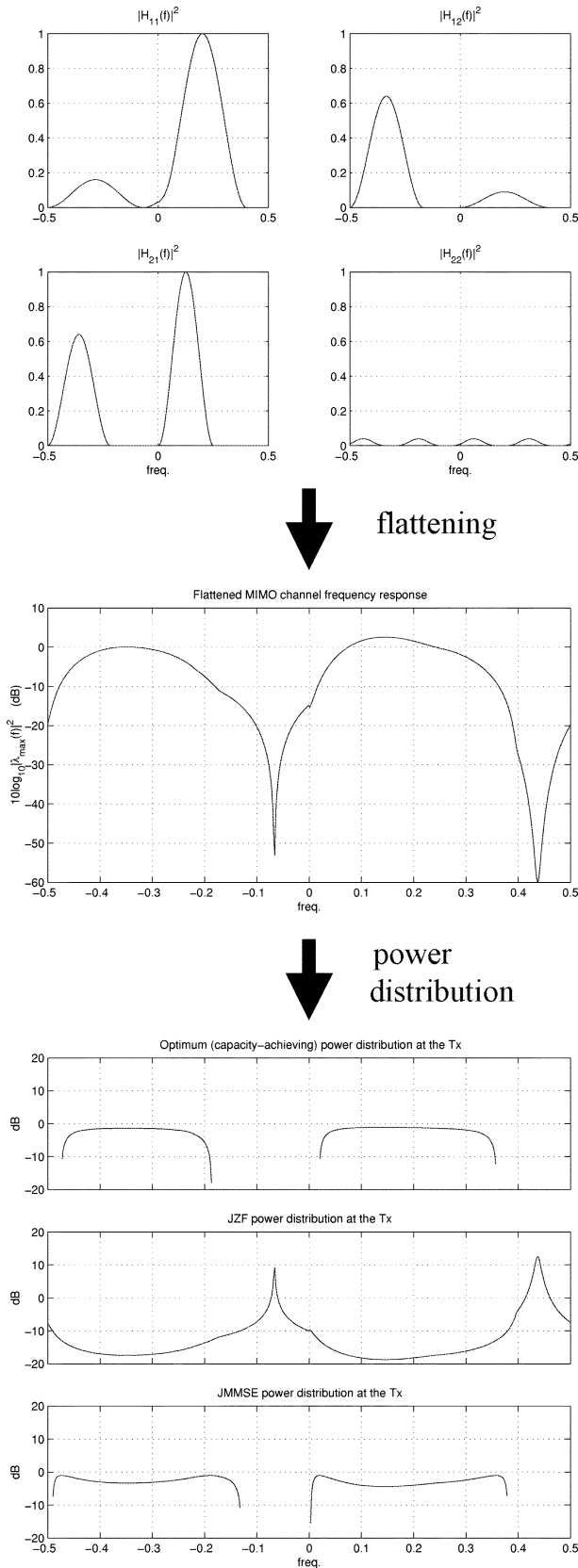


Fig. 4. Illustration of the two-stage process with the simple MIMO (2,2) channel: 1) flattening of the MIMO channel (equal for all methods), and 2) power allocation for SNR = 6 dB (different for each method).

sufficiently short so that the channel remains constant and also sufficiently long so that the standard information-theoretic as-

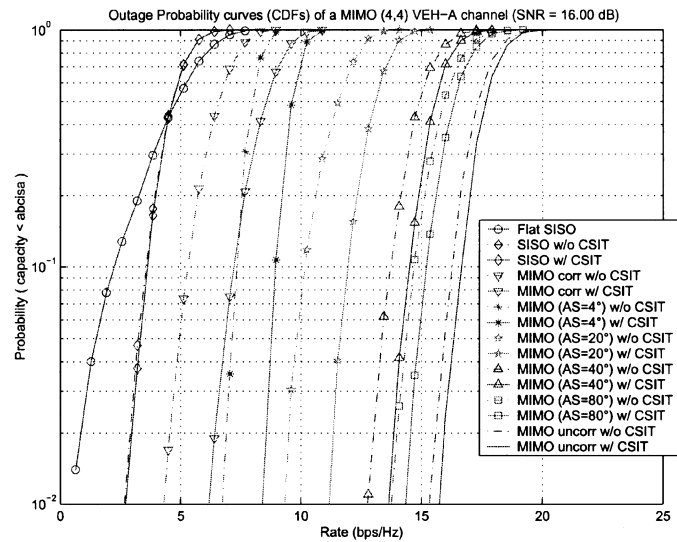


Fig. 5. Outage capacity of a MIMO (4,4) vehicular-A channel (SNR = 16 dB) with different fading correlation and degree of knowledge of the channel at the transmitter.

sumption of infinitely long code block length is a useful idealization. The transmitted signal-to-noise ratio (SNR) is defined as (assuming a Nyquist bandlimited system, i.e., a bandwidth of  $1/T$ , and noise with a flat power spectral density at  $N_0$ )  $SNR \triangleq P_{av}/P_n = P_{av}T/N_0 = E_s/N_0$  where  $E_s$  is the transmitted symbol energy.

Most of the simulation results are presented in terms of outage capacity. For communications without delay constraints in which the transmission duration is so long as to reveal the long-term ergodic properties of the fading process (assuming the channel an ergodic process in time), the ergodic capacity is a useful measure of the average achievable bit rate. The ergodicity assumption, however, is not necessarily satisfied in practical communication system with stringent delay constraints operating on fading channels because no significant channel variability may occur during the whole transmission. In these circumstances, the outage capacity defined as the capacity that cannot be supported for only a small outage probability  $P_{out}$  is the appropriate measure [2], [1]. With no CSIT, an outage means that the transmitter is transmitting at a rate higher than capacity and, therefore, information cannot be reliably transmitted. With CSIT, however, the situation is different since the transmitter knows what is the maximum rate that can be supported by the specific channel realization and, therefore, it can adapt by transmitting at a lower rate (this may or may not be acceptable depending on the specific application). Since the capacity is a function of the random channel realization, it is a random quantity that can be described by its cumulative distribution function (CDF).

*Range of Optimality of Beamforming:* In Fig. 5, the outage capacity curves corresponding to MIMO (4,4) channels with different degrees of spatial correlation (ranging from completely uncorrelated fading to fully correlated) with (w/) and without (w/o) CSIT are plotted. The highest capacity corresponds to the completely uncorrelated case (because the gain of the channel eigenmodes is high), whereas the fully correlated case has the lowest capacity (because it only has one

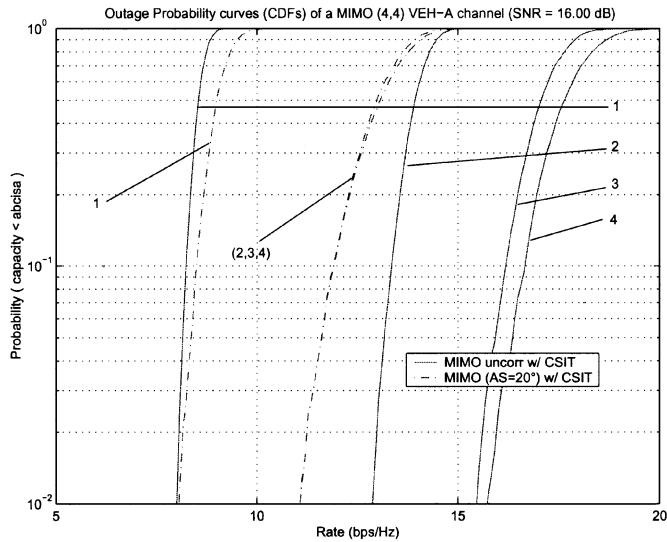


Fig. 6. Outage capacity of a MIMO (4,4) vehicular-A channel as a function of the number of parallel subchannels used (from one to four).

nonvanishing channel eigenmode). The rest of the cases (partial spatial correlation) lay within them and present increasing capacity with AS as expected. For low angle spread (on the order of  $4^\circ$ ), the capacity is close to the fully correlated case. For high angle spread (on the order of  $30^\circ$ ), the capacity is close to the uncorrelated case. CSIT becomes less important with decreasing spatial correlation.

In Fig. 6, we show the effect of using a reduced number of spatial subchannels instead of using them all (four in this case). For the completely uncorrelated case, the decrease in capacity with respect to the use of all four subchannels at 10% outage when using three subchannels is about 0.5 bps/Hz (2.9%), whereas when using two subchannels the difference increases significantly to 4 bps/Hz (a loss of 26%). Therefore, at least three out of the four subchannels should be used. For a partially spatially correlated channel, however, the importance of using all available spatial subchannels diminishes significantly. For  $AS = 20^\circ$ , it is only necessary to use two subchannels, whereas for  $AS = 4^\circ$  (not depicted), using a single subchannel is enough since the four CDF curves collapse into one. The reason why the number of significant parallel subchannels decreases is directly related to the pdf of the channel eigenvalues [24]. It is well-known that a uniform power distribution performs close to the waterfilling distribution provided that the appropriate subchannels are used (basically avoiding zero-gain subchannels on which the allocated power would be wasted). For the uncorrelated case, a uniform power distribution over one, two, or three subchannels is almost optimal<sup>5</sup> whereas for the partially correlated channel with  $AS = 20^\circ$ , it is almost optimal only over one or two subchannels.

In Fig. 7, the capacity at an outage probability of 5% is plotted as a function of the angle spread when using all constituent subchannels and when beamforming is applied (if instead we plot the ergodic capacity, similar results are obtained in the sense that the shape of the curves remains the same but with less conserva-

<sup>5</sup>By “almost optimal,” we really mean that if we plot a CDF curve of the achievable rates corresponding to the waterfilling and uniform distribution, they coincide and cannot be distinguished.

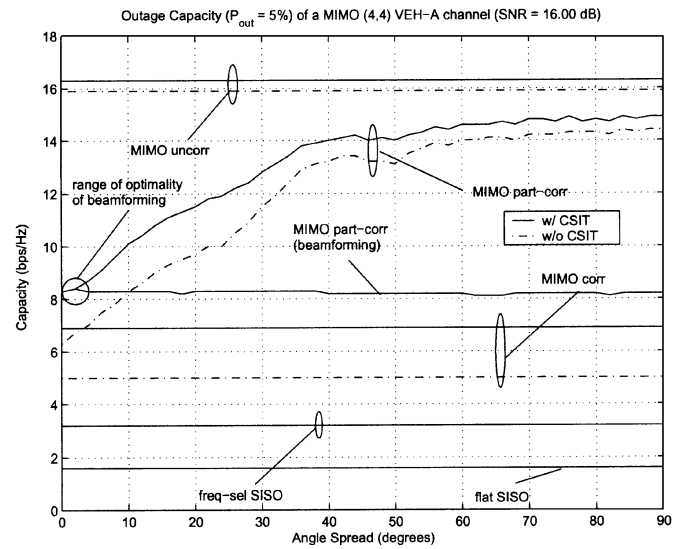


Fig. 7. Outage capacity (at a probability of outage of 5%) of a MIMO (4,4) vehicular-A channel with  $SNR = 16$  dB as a function of the AS.

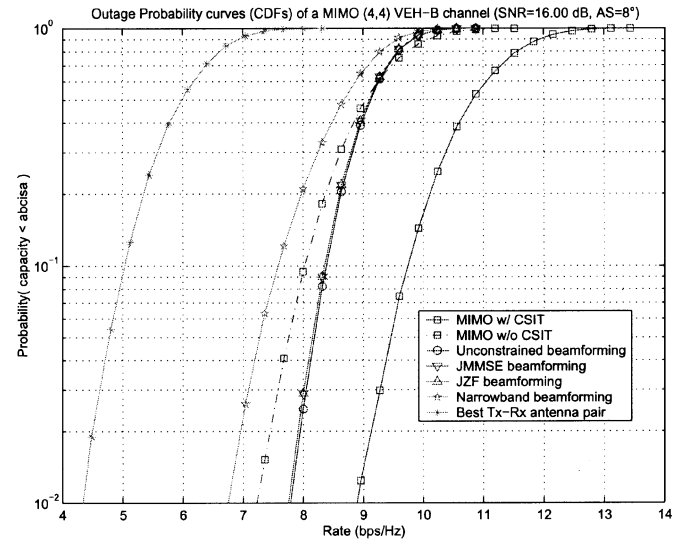


Fig. 8. Outage capacity of a MIMO (4,4) vehicular-B channel corresponding to different transmit-receive methods for  $SNR = 16$  dB.

tive capacity values). From this picture, the range of optimality of beamforming can be seen: for an angle spread on the order of  $4^\circ$ – $8^\circ$ , the loss in capacity of the beamforming scheme is negligible.<sup>6</sup>

*Evaluation and Comparison of the JZF and JMMSE Methods:* We now present a numerical evaluation and comparison of the proposed beamforming methods along with two simple benchmark schemes in terms of outage capacity and bit-error rate (BER). The two benchmark schemes use simple narrowband (i.e., flat in frequency) beamvectors for their simplicity and low complexity. The first one consists on using just the best pair of transmit-receive antennas, i.e.,  $\mathbf{a} = \mathbf{e}_{i_R}$  and  $\mathbf{b} = \sqrt{P_{av}} \mathbf{e}_{j_T}$ , where  $\mathbf{e}_{i_R} \in \mathbb{C}^{n_R \times 1}$  and  $\mathbf{e}_{j_T} \in \mathbb{C}^{n_T \times 1}$

<sup>6</sup>Note that the curve for partial correlation in Fig. 7 is not monotonic on the AS. This is due to the model of the ray distribution (uniform-shaped angular distribution) used to compute the fading correlation [37]. For other distribution models, slightly different curves are obtained but the underlying trend is always the same.

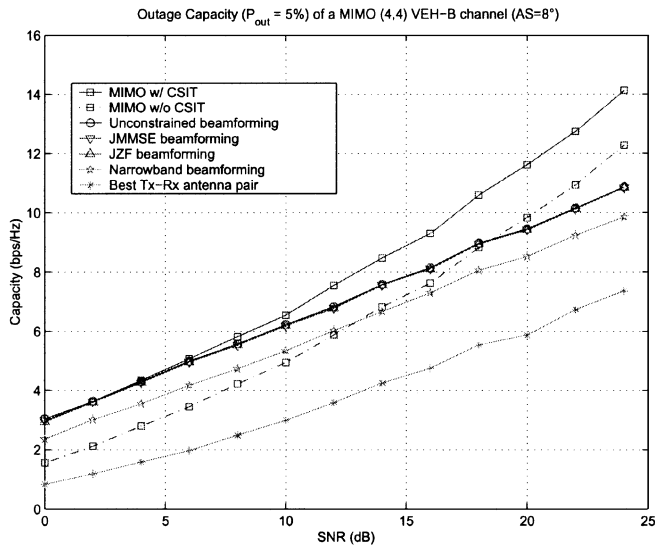


Fig. 9. Outage capacity (at a probability of outage of 5%) of a MIMO (4,4) vehicular-B channel corresponding to different transmit–receive methods as a function of the SNR.

are all-zero vectors with a one in the  $i_R$ th and  $j_T$ th position, respectively. To be more precise, the transmit–receive antenna combination providing the highest gain is selected according to

$$(i_R, j_T) = \arg \max_{(i,j)} \int |[\mathbf{H}(t)]_{i,j}|^2 dt.$$

The second benchmark method designs the transmit beamvector to maximize the energy of the equivalent received channel,  $\mathbf{g}(t) = \mathbf{H}(t)\mathbf{b}$ , subject to the power constraint. The solution to that criterion is given by

$$\left( \int \mathbf{H}^H(t)\mathbf{H}(t)dt \right) \mathbf{b} = \lambda_{\max} \mathbf{b}.$$

This benchmark, further simplified by assuming  $n_R = 1$ , is one of the methods proposed for downlink transmit beamforming in the third generation of digital mobile radio systems Universal Mobile Telecommunication System (UMTS) by the standardization organization Third-Generation Partnership Project (3GPP) [40, Sec. 4.7]. At the receiver, the narrowband beamvector is designed to maximize the SINR as in [41].

Outage capacity curves when using the JZF and JMMSE methods [using  $g(n) = \alpha_0 \delta(n - n_0)$ ] are plotted in Fig. 8 for SNR = 16 dB and AS = 8°. We can observe that the JZF and JMMSE beamforming methods exhibit a capacity virtually identical to that of the unconstrained beamforming case. In Fig. 9, the capacity at a probability of outage of 5% is depicted for all methods as a function of the SNR (as happened with Fig. 7, if the ergodic capacity is plotted, similar results are obtained). Again, it is observed that both the JZF and JMMSE methods perform remarkably well compared with the unconstrained beamforming case. Finally, in Fig. 10, BER results are given for an uncoded system with a quaternary phase-shift keying (QPSK) constellation using an implementation based on finite-impulse response (FIR) filters with 64 taps at the Tx and 128 taps at the Rx. The JMMSE method is clearly more robust and performs better than the JZF.

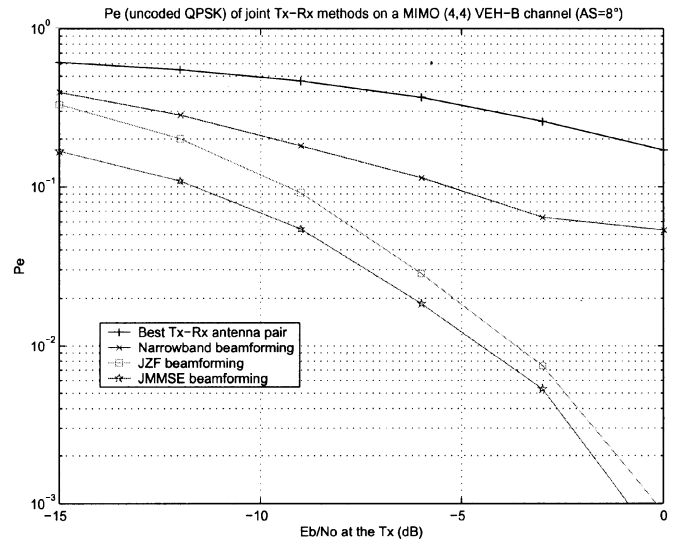


Fig. 10. BER curves of uncoded QPSK for a MIMO(4,4) vehicular-B channel with AS = 8° using FIR filters of 64 taps at the Tx and 128 at the Rx.

## VIII. CONCLUSION

In this paper, we have analyzed communication schemes through wireless MIMO channels with channel knowledge at both sides of the link. We have first shown that, from an information-theoretic point of view, beamforming becomes asymptotically optimum as the spatial correlation of the channel fading increases (typically in outdoor scenarios with BSs placed in high positions). Based on this result, we have then derived practical transmit–receive beamforming schemes according to different design criteria such as ZF and MMSE allowing any desired partial response. Conceptually, these simplified processing schemes have been shown to decompose into two different stages: the first one corresponds to a spatial flattening of the MIMO channel, i.e., choosing the eigenmode with the highest gain at each frequency; the second stage depends on the particular design criterion and performs the power distribution at the transmitter and defines the equalizer at the receiver. An extension of the beamforming approach to the case of using multiple spatial subchannels or eigenmodes (multiple beamforming) has also been obtained in a compact and simple way. We have proposed an exact and practical implementation of the modified waterfilling solution obtained in the MMSE design. The proposed joint ZF and joint MMSE methods have been evaluated and compared with numerical simulations based on realistic scenarios, showing the increased robustness of the joint MMSE approach compared with the joint ZF method as in the classical receive-only equalization.

## APPENDIX A

### OPTIMALITY OF THE DIAGONAL STRUCTURE

We need the following lemma easily obtained from majorization theory [35].

*Lemma 2:* [35, 9.H.1.h] If  $\mathbf{A}$  and  $\mathbf{B}$  are  $n \times n$  positive semidefinite Hermitian matrices, then

$$\text{tr}(\mathbf{AB}) \geq \sum_{i=1}^n \lambda_{A,i} \lambda_{B,n-i+1}$$

where  $\lambda_{A,i}$  and  $\lambda_{B,i}$  are the eigenvalues of  $\mathbf{A}$  and  $\mathbf{B}$ , respectively, in decreasing order.

Focusing on a particular frequency and dropping the frequency dependence for the sake of notation, we have from Lemma 2 (defining  $\mathbf{R}_H \triangleq \mathbf{H}^H \mathbf{R}_n^{-1} \mathbf{H}$ )

$$\text{tr}(\mathbf{B}^H \mathbf{R}_H \mathbf{B} + \kappa \mathbf{I})^{-1} \mathbf{R}_s \geq \sum_{i=1}^L \lambda_i(\mathbf{R}_s) \frac{1}{\lambda_i(\mathbf{B}^H \mathbf{R}_H \mathbf{B}) + \kappa}$$

where  $\lambda_i(\mathbf{X})$  is the  $i$ th eigenvalue of  $\mathbf{X}$  in decreasing order. For any given matrix  $\mathbf{B}$ , the lower bound can always be achieved by post-multiplying  $\mathbf{B}$  by a proper unitary matrix [note that any rotation of  $\mathbf{B}$  does not affect the power  $\text{tr}(\mathbf{B}\mathbf{B}^H)$ ]. To be more specific, given  $\mathbf{B}$ , the lower bound is achieved by  $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{U}_{HB}\mathbf{U}_s^H$ , where  $\mathbf{R}_s = \mathbf{U}_s\mathbf{D}_s\mathbf{U}_s^H$  and  $\mathbf{B}^H\mathbf{R}_H\mathbf{B} = \mathbf{U}_{HB}\mathbf{D}_{HB}\mathbf{U}_{HB}^H$ .

In other words, for  $\mathbf{B}$  to be optimal, it must be that  $\mathbf{U}_s^H\mathbf{B}^H\mathbf{R}_H\mathbf{B}\mathbf{U}_s$  is a diagonal matrix. If the eigenvalues of  $\mathbf{R}_H$  are all different, it follows (see [34] for details) that we can write w.l.o.g.  $\mathbf{B}\mathbf{U}_s = \mathbf{U}_H\mathbf{\Sigma}_B$  or, equivalently

$$\mathbf{B} = \mathbf{U}_H\mathbf{\Sigma}_B\mathbf{U}_s^H \quad (40)$$

where  $\mathbf{U}_H$  is a unitary matrix whose  $L$  first columns are the eigenvectors of  $\mathbf{R}_H$  corresponding to the  $L$  largest eigenvalues and  $\mathbf{\Sigma}_B$  has zero elements except (possibly) along its main diagonal. In case that some of the  $L$  largest eigenvalues of  $\mathbf{R}_H$  are equal, the optimal  $\mathbf{B}$  is not unique since  $\mathbf{U}_H$  allows for arbitrary subrotations. Similarly, in case that some eigenvalues of  $\mathbf{R}_s$  are equal, the optimal  $\mathbf{B}$  is not unique since  $\mathbf{U}_s$  allows for arbitrary subrotations [e.g., if  $\mathbf{R}_s = \sigma_s^2\mathbf{I}$  then any unitary matrix  $\mathbf{U}_s$  can be used in (40)]. In any case, although the set of optimal  $\mathbf{B}$ 's may have a more general decomposition, (40) always describes an optimal  $\mathbf{B}$ .

## APPENDIX B

### OPTIMUM ORDERING OF THE CHANNEL EIGENVALUES

In this appendix, we first show that assuming  $\lambda_{s,i}$  in decreasing order w.l.o.g., the error  $\xi$  in (37) is minimized by the ordering  $\sigma_{B,i}^2/\lambda_{n,i} \geq \sigma_{B,i+1}^2/\lambda_{n,i+1}$  and then we show that the terms  $\sigma_{B,i}^2/\lambda_{n,i}$  can be achieved with a lower power when  $\lambda_{n,i}$  are in increasing order.

*Lemma 3:* [35, 6.A.3] The summation  $\sum_i a_i b_i$  is maximized (minimized) when the sequences  $\{a_i\}$  and  $\{b_i\}$  are in the same (opposite) ordering, i.e., for  $a_i \geq a_j$  then  $b_i \geq (\leq) b_j$ .

The first part of the proof follows by a direct application of Lemma 3 to  $\sum_i \lambda_{s,i}/(\sigma_{B,i}^2/\lambda_{n,i}) + \kappa$ . The second part is similarly proved by defining  $\alpha_i \triangleq \sigma_{B,i}^2/\lambda_{n,i}$  (from which  $\alpha_i$  are in decreasing order) and then writing the utilized power as  $\sum_i \sigma_{B,i}^2 = \sum_i \alpha_i \lambda_{n,i}$ . Again, by a direct application of Lemma 3, it is clear that if  $\lambda_{n,i}$  are in increasing order, the same  $\alpha_i$  can be achieved with less power.

## ACKNOWLEDGMENT

The authors would like to thank W. Rhee, C. Aldana, and J. Loveaux for reviewing the paper and for their helpful comments. We would also like to thank the anonymous reviewers

for their valuable comments on the paper and D. Gesbert for pointing out some related references.

## REFERENCES

- [1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. on Telecommun. (ETT)*, vol. 10, pp. 585–595, Nov.-Dec. 1999.
- [2] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.
- [3] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.
- [4] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.
- [5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, pp. 41–59, 1996.
- [6] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
- [7] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
- [8] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, pp. 3221–3231, Dec. 1994.
- [9] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.
- [10] T. Berger and D. W. Tufts, "Optimum pulse amplitude modulation. Part I: Transmitter-receiver design and bounds from information theory," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 196–208, Apr. 1967.
- [11] N. Amitay and J. Salz, "Linear equalization theory in digital data transmission over dually polarized fading radio channels," *AT&T Bell Labs. Tech. J.*, vol. 63, pp. 2215–2259, Dec. 1984.
- [12] M. Kavehrad and J. Salz, "Cross-polarization cancellation and equalization in digital transmission over dually polarized multipath fading channels," *AT&T Tech. J.*, vol. 64, pp. 2211–2245, Dec. 1985.
- [13] J. Salz, "Digital transmission over cross-coupled linear channels," *AT&T Tech. J.*, vol. 64, pp. 1147–1159, July-Aug. 1985.
- [14] M. L. Honig, P. Crespo, and K. Steiglitz, "Suppression of near- and far-end crosstalk by linear pre- and post-filtering," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 614–629, Apr. 1992.
- [15] J. Yang and S. Roy, "Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1334–1347, Sept. 1994.
- [16] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. COM-24, pp. 1283–1290, Dec. 1976.
- [17] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2172–2178, Nov. 2000.
- [18] 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.
- [19] 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.
- [20] A. Narula, M. D. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2418–2432, Nov. 1999.
- [21] J. M. Cioffi, EE379C Advanced Digital Communications. Stanford Univ., 1998, Stanford, CA. [Online]. Available: <http://www.stanford.edu/class/ee379c>
- [22] 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.
- [23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [24] D. P. Palomar and M. A. Lagunas, "Capacity of spatially flattened frequency-selective MIMO channels using linear processing techniques in transmission," in *Proc. Conf. Information Sciences and Systems (CISS-2001)*, Baltimore, MD, Mar. 21–23, 2001.

- [25] D. Gesbert, H. Bölcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, pp. 1926–1934, Dec. 2002.
- [26] S. A. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems with imperfect feedback," *Proc. IEEE Int. Symp. Information Theory (ISIT 2001)*, p. 321, June 2001.
- [27] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.: Special Issue on MIMO Systems and Applications*, vol. 21, pp. 406–417, Apr. 2003.
- [28] M. T. Ivrlac, W. Utschick, and J. A. Nossek, "Fading correlations in wireless MIMO communication systems," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 819–828, June 2003.
- [29] W. Rhee, W. Yu, and J. M. Cioffi, "The optimality of beam-forming in uplink multiuser wireless systems," *IEEE Trans. Wireless Commun.*, Dec. 2001, submitted for publication.
- [30] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding – Part I: Equalization results," *IEEE Trans. Commun.*, vol. 43, pp. 2582–2594, Oct. 1995.
- [31] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.
- [32] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp. 1349–1387, Sept. 1985.
- [33] S. L. Ariyavisitakul, J. H. Winters, and I. Lee, "Optimum space-time processors with dispersive interference: Unified analysis and required filter span," *IEEE Trans. Commun.*, vol. 47, pp. 1073–1083, July 1999.
- [34] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A convex optimization approach," *IEEE Trans. Signal Processing*, Feb. 2002, submitted for publication.
- [35] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, New York: Academic, 1979.
- [36] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, pp. 2198–2206, Dec. 2001.
- [37] J. Salz and J. H. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 1049–1057, Nov. 1994.
- [38] K. I. Pedersen, J. B. Andersen, J. P. Kermaol, and P. E. Mogensen, "A stochastic multiple-input-multiple-output radio channel model for evaluation of space-time coding algorithms," *Proc. IEEE Vehicular Technology Conf. VTC-2000 Fall*, vol. 2, pp. 893–897, Sept. 2000.
- [39] "Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS (UMTS 30.03)," ETSI, TR101 112, v3.2.0, ETSI, 1998.
- [40] "Physical Layer Procedures (TDD)," 3GPP, 3G TS 25.224, TSGR, 3.1.1 ed., 1999.
- [41] M. A. Lagunas, J. Vidal, and A. Pérez-Neira, "Joint array combining and MLSE for single-user receivers in multipath Gaussian multiuser channels," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2252–2259, Nov. 2000.



**Daniel Pérez Palomar** (S'99) was born in Barcelona, Spain, in 1975. He received the electrical engineering degree from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998. In 1999, he was the recipient of a Ministry of Education Grant to complete his Ph.D. thesis and joined the Department of Signal Theory and Communications at UPC as a Research Assistant, where he is currently working toward the Ph.D. degree in electrical engineering.

During 1998, he was with the Department of Electronic Engineering, King's College London (KCL). From 1999 to 2000, he participated in European projects (ACTS-SUNBEAM and IST-METRA). In 2001, he held a Visiting Research Appointment at the Department of Electrical Engineering, Stanford University, Stanford, CA, with Prof. John M. Cioffi. Since January 2002, he has been a Visiting Researcher at the Telecommunications Technological Center of Catalonia (CTTC), Barcelona, Spain. His primary research interests include: statistical array signal processing, information-theoretic and communication aspects of wireless MIMO channels, convex optimization theory, majorization theory, DMT transmission, multiuser detection, and spread-spectrum techniques.



**Miguel Angel Lagunas** (S'73–M'78–SM'89–F'97) was born in Madrid, Spain, in 1951. He received the telecommunications engineering degree in 1973 from Universidad Politécnica de Madrid (UPM), Madrid, Spain, and the Ph.D. degree in telecommunications from Universidad Politécnica de Barcelona (UPB), Barcelona, Spain.

From 1971 to 1973, he was a Research Assistant at the Semiconductor Lab ETSIT, Madrid; a Teacher Assistant in Network Synthesis and Semiconductor Electronics from 1973 to 1979, an Associate Professor of digital signal processing from 1979 to 1982. Since 1983, he has been a Full Professor, teaching courses in signal processing, array processing, and digital communications. He was Project Leader of high-speed SCMA (1987–1989) and ATM (1994–1995) cable network. He is also Codirector of the first projects for the European Spatial Agency and the European Union, providing engineering demonstration models on smart antennas for satellite communications using DS and FH systems (1986) and mobile communications GSM (Tsunami, 1994). His research activity is devoted to spectral estimation, adaptive systems, and array processing. His technical activities are in advanced front-ends for digital communications combining spatial with frequency-time and coding diversity. He was Vice President for Research of UPC from 1986 to 1989 and Vice-Secretary General for Research, Cicyt, Spain, from 1995 to 1996. Currently, he is Director of the Telecommunications Technological Center of Catalonia (CTTC), Barcelona.

Dr. Lagunas is a member-at-large of Euraspip. He is an Elected Member of the Academy of Engineers of Spain and of the Academy of Science and Art of Barcelona. He was a Fulbright Scholar at the University of Boulder, Boulder, CO.