Minimum BER Linear Transceivers for MIMO Channels via Primal Decomposition

Daniel Pérez Palomar, *Member, IEEE*, Mats Bengtsson, *Member, IEEE*, and Björn Ottersten, *Fellow, IEEE*

***Abstract*—This paper considers the employment of linear transceivers for communication through multiple-input multiple-output (MIMO) channels with channel state information (CSI) at both sides of the link. The design of linear MIMO transceivers has been studied since the 1970s by optimizing simple measures of the quality of the system, such as the trace of the mean-square error matrix, subject to a power constraint. Recent results showed how to solve the problem in an optimal way for the family of Schur-concave and Schur-convex cost functions. In particular, when the constellations used on the different transmit dimensions are equal, the bit-error rate (BER) averaged over these dimensions happens to be a Schur-convex function, and therefore, it can be optimally solved. In a more general case, however, when different constellations are used, the average BER is not a Schur-convex function, and the optimal design in terms of minimum BER is an open problem. This paper solves the minimum BER problem with arbitrary constellations by first reformulating the problem in convex form and then proposing two solutions. One is a heuristic and suboptimal solution, which performs remarkably well in practice. The other one is the optimal solution obtained by decomposing the convex problem into several subproblems controlled by a master problem (a technique borrowed from optimization theory), for which extremely simple algorithms exist. Thus, the minimum BER problem can be optimally solved in practice with very simple algorithms.**

***Index Terms*—BER, convex optimization theory, decomposition techniques, linear precoder, MIMO channel, transceiver, water-filling.**

I. Introduction

MANY different communication channels of diverse physical nature can be treated in a unified way as multiple-input multiple-output (MIMO) channels, which can be conveniently and compactly represented by a channel matrix. Such an abstract representation allows for an elegant, simple, and powerful vector-matrix notation. The two paradigmatic examples of MIMO channels are digital subscriber line (DSL) channels [1] and wireless channels with multiples antennas at both sides of the link [2], [3], which have recently attracted a significant interest because they provide an important increase in capacity over single-input single-output (SISO) channels [4], [5].

When channel state information (CSI) is available at both the transmitter and receiver, the system can adapt to each channel realization to improve the performance and/or the spectral efficiency. From an information-theoretic viewpoint, the best design in terms of capacity is well known and is given by the employment of ideal Gaussian codes [3], [4], [6], [7]. From a more practical point of view, the system may be divided into an uncoded part, which transmits symbols drawn from some constellations, and a coded part that builds upon the uncoded system. Although the ultimate system performance depends on the combination of both parts, it is convenient to consider the uncoded and coded parts independently to simplify the analysis and design. This paper focuses on the uncoded part of the system and, specifically, on the employment of linear transceivers (commonly referred to as linear precoder at the transmitter and equalizer at the receiver) for complexity reasons.

The design of linear MIMO transceivers has been studied since the 1970s by optimizing easily tractable measures of quality of the system, such as the sum of the mean-square error (MSE) of all channel substreams or, equivalently, the trace of the MSE matrix [8]–[12], with a power constraint. In [13], the determinant of the MSE matrix was minimized instead. In [11], a maximum signal-to-interference-plus-noise ratio (SINR) criterion with a zero-forcing (ZF) constraint was also considered. In [14], the results were extended to the case of a peak power constraint. In [15], a general framework was developed to consider a wide range of different design criteria; in particular, the optimal design for Schur-concave and Schur-convex cost functions was obtained.

However, rather than the MSE or the SINR, the ultimate performance of a system is given by the bit-error rate (BER), which is more difficult to handle. In [16], the minimization of the BER (and also of the Chernoff upper bound) averaged over the channel substreams was treated in detail when a diagonal structure is imposed. Recently, the minimum BER design without the diagonal structure assumption was independently obtained in [15] and [17], obtaining an optimal nondiagonal structure. This result, however, only applies when the constellations used in all channel substreams are equal, in which case the cost function happens to be Schur-convex [15]. The solution for the general case of different constellations remains unsolved (in such a case, the cost function is neither Schur-convex nor Schur-concave). Some interesting results along this line were obtained in [18] by combining the suboptimal diagonal structure with the optimal nondiagonal solution for equal constellations.

Manuscript received December 22, 2003; revised September 18, 2004. This work was supported by the Swedish Research Council. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Athina Petropulu.

D. P. Palomar was with the Department of Signals, Sensors and Systems, Royal Institute of Technology (KTH), Stockholm, Sweden. He is now with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544 USA (e-mail: danielp@princeton.edu)

M. Bengtsson and B. Ottersten are with the Department of Signals, Sensors and Systems, Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden (e-mail: mats.bengtsson@s3.kth.se; Bjorn.Ottersten@s3.kth.se).

Digital Object Identifier 10.1109/TSP.2005.850373

1053-587X/$20.00 © 2005 IEEE

This paper considers and solves the minimum BER problem (BER averaged over the channel substreams) with arbitrary constellations, which are assumed given and fixed (obtained, for example, using some kind of bit allocation strategy). The problem can be formulated either as the minimization of the average BER with a power constraint or as the minimization of the power subject to an average BER constraint. First, the problem is reformulated in convex form after some manipulations based on majorization theory [19]. Then, a heuristic and suboptimal solution, which performs remarkably well in practice, is proposed. Finally, the optimal solution of the problem is obtained based on a primal decomposition approach of the convex problem, a technique borrowed from optimization theory [20]–[22]. In a nutshell, the original complicated problem is decomposed into several subproblems that can be independently and easily solved (possibly in a parallel fashion) and a master problem that coordinates the subproblems. Thus, the minimum BER problem can be optimally solved in practice with very simple and efficient algorithms.

The paper is structured as follows. In Section II, the signal model is introduced and the problem is formulated. Section III provides some preliminary results. Section IV contains the main results of the paper: 1) the reformulation of the minimum BER problem in convex form, 2) its decomposition into several subproblems controlled by a master problem, 3) the full characterization of the subproblems, and 4) a simple optimal algorithm for the master problem. A summary of the whole design process is given in Section V. Section VI is devoted to numerical simulations. Finally, in Section VII, the main conclusions of the paper are drawn.

The following notation is used. Boldface uppercase letters denote matrices, boldface lowercase letters denote column vectors, and italics denote scalars. $\mathbb{C}^{m\times n}$ represents the $m\times n$ complex field. The superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ denote transpose, complex conjugate, and Hermitian operations, respectively. $[\mathbf{X}]_{i,j}$ (also $[\mathbf{X}]_{ij}$) denotes the (i,j)th element of matrix $\mathbf{X}$. $\mathbf{d}(\mathbf{X})$ and $\boldsymbol{\lambda}(\mathbf{X})$ are the vectors containing the diagonal elements and eigenvalues of matrix $\mathbf{X}$, respectively, and diag$(\{x_k\})$ is a diagonal matrix with diagonal elements given by the set $\{x_k\}$. $|\cdot|$ denotes the size or cardinality of a set, Tr$(\cdot)$ denotes the trace of a matrix, and $(x)^+ \triangleq \max(0,x)$ is the projection on the non-negative orthant. The derivative of a function f is denoted by f' and the gradient by ∇f.

II. System Model and Problem Formulation

In this section, after introducing the basic signal model for MIMO channels and linear processing, the problem is mathematically formulated.

A. System Model

Consider a general MIMO communication channel with n_T transmit and n_R receive dimensions. The sampled baseband signal model is

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{s} \in \mathbb{C}^{n_T\times 1}$ is the transmitted vector, $\mathbf{H} \in \mathbb{C}^{n_R\times n_T}$ is the channel matrix, $\mathbf{y} \in \mathbb{C}^{n_R\times 1}$ is the received vector, and $\mathbf{n} \in \mathbb{C}^{n_R\times 1}$ is a zero-mean circularly symmetric complex Gaussian interference-plus-noise vector with arbitrary (but known) covariance matrix $\mathbf{R}_n$, i.e., $\mathbf{n} \sim \mathcal{CN}(\mathbf{0},\mathbf{R}_n)$. Note that the model in (1) is indeed very general and can model a wide range of different communication systems of diverse physical nature, ranging from DSL systems to wireless multiantenna systems to convolutional channels[1] (see [23] for an overview of how to model different scenarios as (1)).

In some situations (such as in multicarrier systems), it may be useful to model the system as a set of N parallel and noninterfering MIMO channels, with the signal model at the kth MIMO channel similarly given by

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{s}_k + \mathbf{n}_k, \quad 1\le k\le N \tag{2}$$

where each term is similarly defined as in the case of a single MIMO channel and $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0},\mathbf{R}_{n_k})$ (noise from different carriers is assumed uncorrelated). Since the case of multiple MIMO channels (2) is mathematically more general than (1), it will be considered in the sequel. We will restrict the results to the use of linear processing at the transmitter and at the receiver, obtaining linear transceivers.

It is important to remark that a multicarrier MIMO channel can be modeled in two different ways: as a set of N parallel MIMO channels, for which we will need N pairs of transmit-receive matrices, and as a single big MIMO channel, for which we will need a single pair of transmit-receive matrices. In the former approach, an independent processing per carrier is assumed (termed *carrier-noncooperative*), whereas in the latter approach, a joint processing over all carriers is assumed (termed *carrier-cooperative*).

Considering the use of transmit linear processing matrices (commonly termed *linear precoders*), the transmitted vector at the kth MIMO channel can be written as (see Fig. 1)

$$\mathbf{s}_k = \mathbf{B}_k\mathbf{x}_k \tag{3}$$

where $\mathbf{B}_k \in \mathbb{C}^{n_T\times L_k}$ is the transmit matrix, $\mathbf{x} \in \mathbb{C}^{L_k\times 1}$ is the data vector that contains the L_k symbols to be transmitted (which, in principle, can be larger than the rank of the channel matrix), and $L_T = \sum_{k=1}^{N} L_k$ is the total number of transmitted symbols. In the same way as L_k denotes the number of established links, it is convenient to define $\check{L}_k \triangleq \min(L_k, \text{rank}(\mathbf{H}_k))$ as the number of effective channel eigenvalues used and $L_{0,k} \triangleq L_k - \check{L}_k$.

The total average transmitted power (in units of energy per transmission) is

$$P_T = \sum_k \mathbb{E}\left[\|\mathbf{s}_k\|^2\right] = \sum_k \text{Tr}\left(\mathbf{B}_k\mathbf{B}_k^H\right) \tag{4}$$

where we have assumed zero-mean unit-energy uncorrelated symbols, i.e., $\mathbb{E}[\mathbf{x}_k\mathbf{x}_l^H] = \mathbf{I}\delta_{kl}$.

[1]Convolutional channels can be conveniently modeled, for example: 1) using a filterbank approach [7], [11], 2) defining a convolutional matrix as channel [3], and 3) in a multicarrier fashion [3].

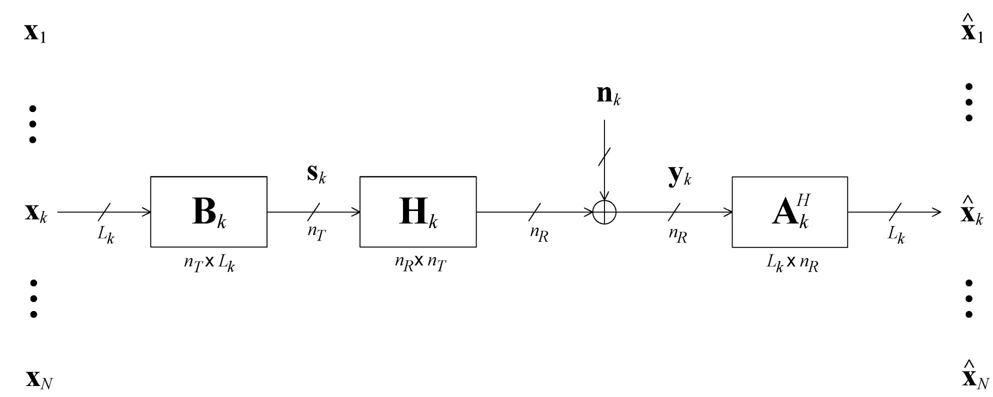

Fig. 1. Block diagram of multiple MIMO channels with linear processing at both the transmitter and the receiver (linear transceivers).

Similarly, assuming linear receive processing matrices (usually referred to as *linear equalizers*), the estimated data vector at the kth MIMO channel is (see Fig. 1)

$$\hat{\mathbf{x}}_k = \mathbf{A}_k^H \mathbf{y}_k \tag{5}$$

where $\mathbf{A}_k^H \in \mathbb{C}^{L_k \times n_R}$ is the receive matrix.

The quality of the ith established link in the kth MIMO channel can be measured, among others, in terms of MSE, SINR, or BER, defined, respectively, as[2]

$$\begin{aligned}\mathrm{MSE}_{k,i} &\triangleq \mathbb{E}\left[|\hat{x}_{k,i} - x_{k,i}|^2\right] \\ &= \left|\mathbf{a}_{k,i}^H \mathbf{H}_k \mathbf{b}_{k,i} - 1\right|^2 + \mathbf{a}_{k,i}^H \mathbf{R}_{in_k} \mathbf{a}_{k,i}\end{aligned} \tag{6}$$

$$\mathrm{SINR}_{k,i} \triangleq \frac{\text{desired component}}{\text{undesired component}} = \frac{\left|\mathbf{a}_{k,i}^H \mathbf{H}_k \mathbf{b}_{k,i}\right|^2}{\mathbf{a}_{k,i}^H \mathbf{R}_{in_{k,i}} \mathbf{a}_{k,i}} \tag{7}$$

$$\mathrm{BER}_{k,i} \triangleq \frac{\#\text{ bits in error}}{\#\text{ transmitted bits}} \approx \tilde{g}_{k,i}(\mathrm{SINR}_{k,i}) \tag{8}$$

where $\mathbf{b}_{k,i}$ and $\mathbf{a}_{k,i}$ are the ith columns of matrices $\mathbf{B}_k$ and $\mathbf{A}_k$, respectively, $x_{k,i}$ and $\hat{x}_{k,i}$ are the ith elements of vectors $\mathbf{x}_k$ and $\hat{\mathbf{x}}_k$, respectively, $\mathbf{R}_{in_{k,i}} = \sum_{j \neq i} \mathbf{H}_k \mathbf{b}_{k,j} \mathbf{b}_{k,j}^H \mathbf{H}_k^H + \mathbf{R}_{n_k}$ is the interference-plus-noise covariance matrix seen by the (k,i)th link, and $\tilde{g}_{k,i}$ is a function that relates the BER to the SINR at the (k,i)th substream, which depends on the constellation used (c.f. Section III-C).

B. Problem Formulation

This paper considers the design of linear transceivers, i.e., of the linear transmitters $\mathbf{B}_k$ and receivers $\mathbf{A}_k$, to minimize the transmitted power while satisfying some global average BER constraint. The constellations used on the different substreams are assumed known and fixed (the choice of the constellations or, equivalently, the bit allocation can range from simply using the same constellations to using a sophisticated bit allocation strategy based on a gap-approximation approach [24, Part II]). The mathematical formulation of the problem is

$$\begin{aligned}\min_{\{\mathbf{B}_k, \mathbf{A}_k\}} \quad & \sum_{k=1}^{N} \mathrm{Tr}\left(\mathbf{B}_k \mathbf{B}_k^H\right) \\ \text{s.t.} \quad & \frac{1}{L_T} \sum_{k=1}^{N} \sum_{i=1}^{L_k} \mathrm{BER}_{k,i} \leq \mathrm{BER}_0\end{aligned} \tag{9}$$

[2]Note that in any real system, we have $0 \leq \mathrm{MSE} \leq 1$, $\mathrm{SINR} \geq 0$, and $0 \leq \mathrm{BER} \leq 0.5$.

where BER_0 is the maximum permitted average BER and each $\mathrm{BER}_{k,i}$ depends on the linear transceiver [see (7) and (8)].

Similarly, the problem could have been formulated as the minimization of the average BER subject to some power constraint (both formulations are, in fact, equivalent). This paper focuses on the BER-constrained problem in (9), although the same methodology used in this paper to solve problem (9) could be followed for the alternative power-constrained formulation. Note also that given an algorithm for the BER-constrained problem, one can readily solve the power-constrained problem simply by using a bisection method [25, Alg. 4.1]. This observation comes straightforwardly by noticing that both problems characterize the same strictly monotonic curve of power versus BER.

The motivation for the problem formulation in (9) hinges on the fact that the ultimate measure of performance of a system is the BER. It makes sense then to design the system by dealing directly with the BER rather than using some other quantity, such as the commonly used sum of the MSEs. The related problem of imposing independent BER constraints on each of the substreams rather than a global average BER was considered in [26] (c.f. Section IV-A). The interested reader is referred to [15] for a comparison of different design criteria and also to the following illustrative example that shows how the classical method of minimizing the sum of the MSEs does not necessarily perform very well since it ignores the knowledge of the constellations.

Illustrative Example: Consider a system with the following characteristics: a diagonal 2×2 MIMO channel $\mathbf{H} = \mathrm{diag}(\{4, 1\})$, a white normalized noise $\mathbf{R}_n = \mathbf{I}$, two quadrature phase shift keying (QPSK) symbols simultaneously transmitted with a ZF receiver (c.f. Section III-B), and a signal-to-noise ratio (SNR) of 10 dB. If the transceiver is designed according to the classical criterion of minimizing the sum of the MSEs with a diagonal transmission, which does not take into account the tradeoff between the constellations and the channel eigenvalues, the transmitter is $\mathbf{B} = \mathrm{diag}(\{\sqrt{2}, \sqrt{8}\})$, the MSEs are 0.031 25 and 0.125, the BERs are $\sim 7.7 \times 10^{-9}$ and $\sim 2.3 \times 10^{-3}$, and the average BER is $\sim 1.2 \times 10^{-3}$. If, instead, the transceiver is properly designed by minimizing the average BER, one possible solution is $\mathbf{B} = \mathrm{diag}(\{\sqrt{2}, \sqrt{8}\})\mathbf{H}_2/\sqrt{2}$ (where $\mathbf{H}_2$ is a 2×2 Hadamard matrix; c.f. [15]), the MSEs are both equal to 0.078 125, and the BERs are also equal with average value of $\sim 1.7 \times 10^{-4}$,

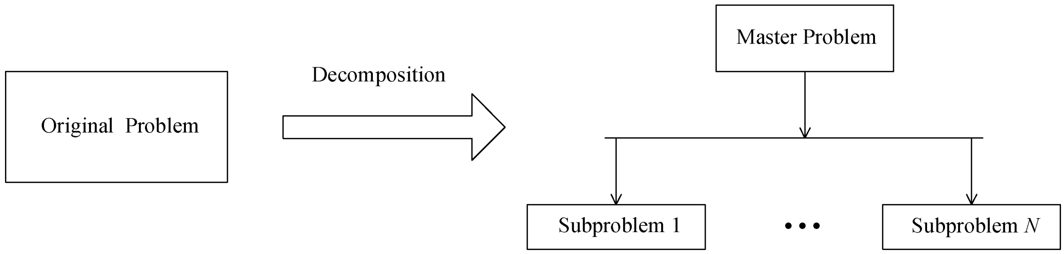

Fig. 2. Decomposition of a large problem into several subproblems controlled by a master problem.

which is one order of magnitude below the classical design based on the sum of the MSEs.

III. PRELIMINARIES

This section contains some preliminary results that are required before attempting to solve the problem in (9). First, we describe the main ingredient to solve it, borrowed from optimization theory. Then, the minimum mean-squared error (MMSE) and ZF receivers are described as optimal linear receivers. Finally, the BER of the system is analytically characterized.

A. Decomposition Techniques in Optimization Theory

Many optimization problems stemming from real applications have a large number of variables and constraints. In principle, the existing general-purpose methods to solve convex problems, e.g., interior-point methods [25] or cutting-plane methods [22], are capable of handling such large problems. In many cases, however, problems have a very particular structure that allows simplification based on decomposing the original problem into smaller and simpler subproblems [20]. For example, the problem may decouple into independent subproblems when some of the optimization variables are fixed. A master problem is then necessary to coordinate the subproblems by means of the coupling variables or constraints (see Fig. 2) [20], [22], [27].

Decomposition approaches are particularly efficient and useful if the structure of each subproblem permits the use of special, simple, and fast solution methods. In addition, such decompositions allow practical parallel implementations very appealing in hardware implementations or parallel-computation platforms. Regarding the master problem, it may, in general, be a nondifferentiable problem for which cutting-plane methods and simple subgradient methods have been developed [21], [22].

Most of the existing techniques to decompose problems can be classified into *primal decomposition* and *dual decomposition* methods. The former (also called decomposition by right-hand side allocation or decomposition with respect to variables) is based on decomposing the original primal problem, whereas the latter (also termed Lagrangian relaxation of the coupling constraints or decomposition with respect to constraints) is based on decomposing the dual of the problem [21], [22]. Primal decomposition methods have the interpretation that the master problem gives each subproblem an amount of resources that it can use; the role of the master problem is then to properly allocate the existing resources. In dual decomposition methods, the master problem sets the price for the resources to each subproblem that has to decide the amount of resources to be used depending on the price; the role of the master problem is then to obtain the best pricing strategy (see [28] for an example of dual decomposition for simultaneous routing and resource allocation in wireless data networks).

This paper uses a primal decomposition approach to simplify and solve the minimum BER problem. It is based on the key and simple fact that a problem can be optimized by first optimizing over some variables and then over the remaining ones [25, Sec. 4.1.3] (see also [22, Sec. 6.4.2])

$$\min_{\mathbf{x},\mathbf{y}} f(\mathbf{x},\mathbf{y}) = \min_{\mathbf{x}} \min_{\mathbf{y}} f(\mathbf{x},\mathbf{y}). \tag{10}$$

This is commonly called *concentration* in the literature of estimation theory [29].

B. Optimal MMSE and ZF Linear Receivers

The linear MMSE receiver, also termed Wiener filter, is obtained as the optimal solution that minimizes simultaneously all MSEs [29], which are given by the diagonal elements of the MSE matrix defined as

$$\begin{aligned}\mathbf{E}_k &\triangleq \mathbb{E}\left[(\hat{\mathbf{x}}_k - \mathbf{x}_k)(\hat{\mathbf{x}}_k - \mathbf{x}_k)^H\right] \\ &= \left(\mathbf{A}_k^H\mathbf{H}_k\mathbf{B}_k - \mathbf{I}\right)\left(\mathbf{B}_k^H\mathbf{H}_k^H\mathbf{A}_k - \mathbf{I}\right) + \mathbf{A}_k^H\mathbf{R}_{n_k}\mathbf{A}_k \end{aligned} \tag{11}$$

i.e., $\mathrm{MSE}_{k,i} = [\mathbf{E}_k]_{ii}$. The ZF receiver is the unbiased version of the MMSE receiver. It can be similarly obtained with the additional ZF constraint $\mathbf{A}_k^H\mathbf{H}_k\mathbf{B}_k = \mathbf{I}$, which is desirable to avoid crosstalk among the different links established through the kth MIMO channel. Due to the ZF constraint, the ZF receiver is only defined for $L_k \leq \mathrm{rank}(\mathbf{H}_k)$ or $L_{0,k} = 0$ (as opposed to the MMSE receiver, which is always well defined).

For a given transmit matrix $\mathbf{B}_k$, the MMSE and ZF receivers can be compactly written as [15], [23], [26], [29]

$$\mathbf{A}_k = \mathbf{R}_{n_k}^{-1}\mathbf{H}_k\mathbf{B}_k\left(\nu\mathbf{I} + \mathbf{B}_k^H\mathbf{R}_{H_k}\mathbf{B}_k\right)^{-1} \tag{12}$$

where ν is a parameter defined as

$$\nu \triangleq \begin{cases} 1, & \text{for the MMSE receiver} \\ 0, & \text{for the ZF receiver} \end{cases}$$

and $\mathbf{R}_{H_k} \triangleq \mathbf{H}_k^H\mathbf{R}_{n_k}^{-1}\mathbf{H}_k$ is the squared whitened channel matrix. The MSE matrix reduces then to $\mathbf{E}_k = (\nu\mathbf{I} + \mathbf{B}_k^H\mathbf{R}_{H_k}\mathbf{B}_k)^{-1}$, and the MSEs are given by

$$\mathrm{MSE}_{k,i} = \left[\left(\nu\mathbf{I} + \mathbf{B}_k^H\mathbf{R}_{H_k}\mathbf{B}_k\right)^{-1}\right]_{ii}. \tag{13}$$

Interestingly, the MMSE and ZF receivers are also optimum in the sense that they maximize simultaneously all SINRs

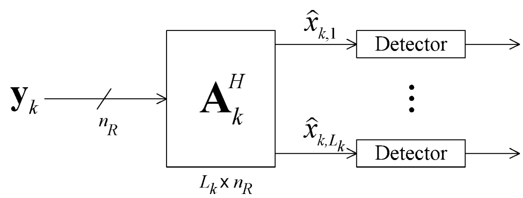

Fig. 3. Independent detection of the substreams.

[15], [23], [30, Prob. 6.5], which are then nicely related to the MSEs by

$$\mathrm{SINR}_{k,i} = \frac{1}{\mathrm{MSE}_{k,i}} - \nu. \tag{14}$$

As a consequence of the relation between the BER and the SINR in (8) through a decreasing function (c.f. Section III-C), the MMSE and ZF receivers are optimum (for linear receiver structures) in the sense that they minimize simultaneously all BERs as well.

C. Characterization of the BER Function

To characterize the BER of the MIMO system under consideration, we assume that the different links are independently detected after the joint linear processing with the receive matrices $\mathbf{A}_k$ (see Fig. 3). This reduces the complexity drastically compared to a joint maximum likelihood (ML) decoding.

For most types of modulations, the BER of a communication link can be analytically expressed as a function of the SINR [as in (8)] when the interference-plus-noise term follows a Gaussian distribution [30], [31] (see, for example, [32] for exact BER expressions of amplitude modulations).[3] Using (14), the BER can also be expressed as a function of the MSE:

$$\mathrm{BER} = g^{(\nu)}(\mathrm{MSE}) \triangleq \tilde{g}(\mathrm{SINR} = \mathrm{MSE}^{-1} - \nu). \tag{15}$$

In the rest of paper, we will just denote the BER as a function of the MSE by g, keeping in mind that the function g depends on ν [recall that the MSE also depends on ν (13)].

For the case of a ZF receiver, each of the established links contains only Gaussian noise and, therefore, the analytical BER characterization can be exact. For the case of an MMSE receiver, however, there is crosstalk among the established links with a non-Gaussian distribution. The computation of the BER involves then a summation over all the possible values of the interfering signals, which is exponential in the size of the constellations. In order to reduce the complexity of evaluating these expressions, it is customary to obtain an approximate statistical model for the crosstalk. In fact, the central limit theorem can be invoked to show that the distribution converges almost surely to a Gaussian distribution as the number of interfering signals increases (c.f. [33], where several asymptotic scenarios are considered). Therefore, the analytical BER characterization of (15) can be used to obtain approximate results with an increasing accuracy in the number of established links (even when the central limit theorem cannot be invoked, it is, in general, possible to obtain some approximate expression for the BER as a function of the SINR [24, Part I, Sec. III.B]).

In the sequel, we will use the following properties of the BER function g for a given constellation $\mathcal{C}$:

P1) g is strictly increasing and $g(0) = 0$.

P2) g is strictly convex on the interval $[0, u]$[4] (for mathematical convenience, g is defined $+\infty$ elsewhere[5]). In addition, $u_1 \geq u_2$ for $|\mathcal{C}_1| \leq |\mathcal{C}_2|$.

P3) $g_1(\rho) < g_2(\rho)$ and $g_1'(\rho) < g_2'(\rho)$ for $|\mathcal{C}_1| < |\mathcal{C}_2|$.

It is important to remark that the properties P1-P3 are indeed very natural for any reasonable family of constellations. The increasingness of g is clear since a higher MSE must always be worse than a lower MSE. The convexity of g is a natural result for the range in which the MSE as a function of the SINR (which is a convex function) is approximately linear (this follows since we expect a system to have a BER at some SINR_0 smaller than (or at least equal to) the average BER that would be achieved by a time-division approach using two different SINRs satisfying $(\mathrm{SINR}_1 + \mathrm{SINR}_2)/2 = \mathrm{SINR}_0$). Clearly, $g_1(\rho) < g_2(\rho)$ must be satisfied for $|\mathcal{C}_1| < |\mathcal{C}_2|$; otherwise, we could transmit more bits at a lower BER, which does not make any sense. In addition, $g_1'(\rho) < g_2'(\rho)$ is a natural result since it simply means that larger constellations (normalized with unit energy) are expected to have a higher sensitivity with respect to changes in the MSE, which is an expected result because higher constellations have a smaller minimum distance.

The analytical characterization of the minimum BER problem in Section IV is valid only for systems that work in the convex region, i.e., with a sufficiently small MSE at each established substream. This is a mild restriction because, if the gain of a substream is too low, it may be better not to use it at all and to decrease the total number of transmitted symbols (in [18], it is shown that if the bit allocation is properly done, then the BER function operates in the convex regime). Nevertheless, it is worth mentioning that, in practice, the method proposed in Section IV also works in the nonconvex region.

The following result will prove extremely useful in the sequel.

Lemma 1: Let g_1 and g_2 be two BER functions corresponding to the constellations $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively, with $|\mathcal{C}_1| \leq |\mathcal{C}_2|$ and satisfying properties P1-P3. Then

$$\begin{aligned} & g_1(\rho_2) + g_2(\rho_1) \\ & \quad \leq g_1(\rho_1) + g_2(\rho_2), \qquad \text{for} \quad \rho_1 < \rho_2 \end{aligned} \tag{16}$$

$$\begin{aligned} & g_1\left(\frac{\rho_1+\rho_2}{2}\right) + g_2\left(\frac{\rho_1+\rho_2}{2}\right) \\ & \quad < g_1(\rho_1) + g_2(\rho_2). \end{aligned} \tag{17}$$

In addition, if $|\mathcal{C}_1| < |\mathcal{C}_2|$, then (16) is satisfied with strict inequality [provided that $g_1(\rho_2) + g_2(\rho_1) < +\infty$].

Proof: See Appendix A. ■

Important Example: QAM Constellations

To illustrate the previous characterization of the BER, we now consider a particular example of a great interest: QAM constellations. As previously mentioned, for QAM constellations under

[3]Note that the BER function is valid for the MMSE receiver only when the decision regions of the detector are scaled to account for the bias in the MMSE receiver [24, Part I].

[4]The value of u can be obtained by analyzing the convexity properties of the function g [see, for example, (20) and (21) for QAM constellations].

[5]Property P1 is then satisfied only on the interval $[0, u]$.

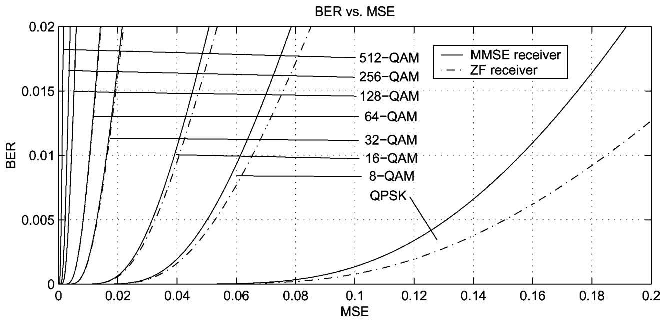

Fig. 4. BER as a function of the MSE for different QAM constellations.

Gaussian noise, the BER can be analytically expressed in an exact way [32]. However, for simplicity of exposition, we approximate the expression with the most significative term.

The BER corresponding to an M-ary QAM constellation (assuming that a Gray encoding is used to map the bits into the constellation points) is [31], [32]

$$\mathrm{BER}(\mathrm{SINR}) \approx \frac{\alpha}{\log_2 M} \mathcal{Q}(\sqrt{\beta\,\mathrm{SINR}}) \tag{18}$$

where $\mathcal{Q}$ is defined as $\mathcal{Q}(x) \triangleq 1/\sqrt{2\pi}\int_x^\infty e^{-\lambda^2/2}d\lambda$ [30],[6] $\alpha = 4(1-1/\sqrt{M})$, and $\beta = 3/(M-1)$ are parameters that depend on the constellation size.[7] It is sometimes convenient to use the Chernoff upper bound of the tail of the Gaussian distribution function $\mathcal{Q}(x) \le (1/2)e^{-x^2/2}$ [30] to approximate the BER (which becomes a reasonable approximation for high values of the SINR) as

$$\mathrm{BER}(\mathrm{SINR}) \approx \frac{\alpha}{2\log_2 M} e^{-\frac{\beta}{2}\mathrm{SINR}} \tag{19}$$

(see [34] for better approximations for M-QAM and M-PSK constellations based on curve fitting and [35] for approximations in the neighborhood of some nominal point).

As can be seen from Fig. 4 (and can also be proved analytically), the BER as a function of the MSE satisfies the expected properties P1-P3. In fact, the range of MSE for which the function is convex for an MMSE receiver, as shown in (20) at the bottom of the page [15], [23], and for a ZF receiver is [23]

$$\mathrm{MSE} \le \begin{cases} \beta/3, & \text{for Q-function approximation} \\ \beta/4, & \text{for Chernoff approximation.} \end{cases} \tag{21}$$

The same results hold for the exact BER expression as given in [32].

As a rule of thumb, the BER function is convex for BER $\le 2\times 10^{-2}$, which is a reasonable and mild assumption for communication systems with CSI at the transmitter.

[6]The $\mathcal{Q}$-function and the commonly used complementary error function "erfc" are related as $\mathrm{erfc}(x) = 2\mathcal{Q}(\sqrt{2}x)$ [30].

[7]For $I\times J$ rectangular constellations, the parameters are $\alpha = 2(((I-1)/I) + ((J-1)/J))$ and $\beta = 6/(I^2+J^2-2)$ [32].

IV. Primal Decomposition for Minimum BER Transceivers

In this section, we propose a simple and efficient way to optimally solve the original problem (9) by means of a primal decomposition approach as described in Section III-A. Using the optimal MMSE and ZF receive matrices obtained in Section III-B and the characterization of the BER function $\mathrm{BER} = g(\mathrm{MSE})$ of Section III-C, problem (9) can be written as

$$\begin{aligned} \min_{\{\mathbf{B}_k,\mathbf{p}_k\}} \quad & \sum_{k=1}^N \mathrm{Tr}\left(\mathbf{B}_k\mathbf{B}_k^H\right) \\ \text{s.t.} \quad & g_{k,i}\left(\left[\left(\nu\mathbf{I} + \mathbf{B}_k^H\mathbf{R}_{H_k}\mathbf{B}_k\right)^{-1}\right]_{ii}\right) \le p_{k,i}, \quad \forall k,i \\ & \frac{1}{L_T}\sum_{k=1}^N\sum_{i=1}^{L_k} p_{k,i} \le \mathrm{BER}_0 \end{aligned} \tag{22}$$

where the $p_{k,i}$'s are additional variables introduced for convenience that represent the BER on each link.

The successful application of a decomposition technique to problem (22) presents the following challenges: 1) problem (22) has to be reformulated in convex form [see (35) in the proof of Theorem 2]; after this step, it is clear that the problem can be optimally solved with some general-purpose method; 2) the problem has to be decomposed in the right way so that the subproblems behave well and the master problem can be easily solved [see (28) in Theorem 2]; 3) the subproblems have to be fully characterized, which includes obtaining a simple method to solve them (Theorem 1 and Algorithm 1) and analyzing the differentiability and convexity (Proposition 1); and 4) the master problem has to be easy to solve, which holds in this case because the problem was properly formulated so that the feasible set of the master problem is essentially a simplex [see (30) and Algorithm 2].

A. Characterization of the Subproblems: Quality-of-Service (QoS)-Constrained Systems

The problem of designing the minimum-power transmit matrices $\mathbf{B}_k$, subject to a set of QoS requirements $\mathrm{MSE}_{k,i} \le \rho_{k,i}\ \forall k,i$, was optimally solved in [26] for the case of an MMSE receiver. In this section, we first extend such a result to more general constraints of the form $g_{k,i}(\mathrm{MSE}_{k,i}) \le p_{k,i}\ \forall k,i$ and to include the case of a ZF receiver as well, obtaining an efficient and optimal way to solve the problem in practice; then, we characterize the optimal objective value as a function of the QoS constraints, showing its convexity and differentiability, which are key properties for the success of the primal decomposition approach.

Theorem 1: Consider the following nonconvex power minimization problem subject to BER QoS constraints p_i [assumed

$$\mathrm{MSE} \le \begin{cases} \left((\beta+3) - \sqrt{\beta^2 - 10\beta + 9}\right)/8, & \text{for Q-function approximation} \\ \beta/4, & \text{for Chernoff approximation} \end{cases} \tag{20}$$

bounded by $0 < p_i < 0.5$ and ordered without loss of generality (w.l.o.g.) such that $1 > g_i^{-1}(p_i) \geq g_{i+1}^{-1}(p_{i+1})$]:

$$\begin{array}{ll} \min\limits_{\mathbf{B}} & \mathrm{Tr}(\mathbf{B}\mathbf{B}^H) \\ \text{s.t.} & g_i\left(\left[\left(\nu\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B}\right)^{-1}\right]_{ii}\right) \leq p_i, \qquad 1 \leq i \leq L \end{array} \tag{23}$$

where each g_i is a BER function satisfying property P1.

If feasible, the optimal solution to problem (23) satisfies all QoS constraints with equality and is given by

$$\mathbf{B} = \mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}\mathbf{Q} \tag{24}$$

where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T \times \check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L}$ largest eigenvalues in increasing order $\{\lambda_i\}_{i=1}^{\check{L}}$, $\mathbf{\Sigma}_{B,1} = [\mathbf{0}\ \mathrm{diag}(\{\sigma_{B,i}\})] \in \mathbb{C}^{\check{L}\times L}$ has zero elements except along the right-most main diagonal (assumed real w.l.o.g.), and $\mathbf{Q}$ is a unitary matrix such that $g_i([(\nu\mathbf{I} + \mathbf{B}^H\mathbf{R}_H\mathbf{B})^{-1}]_{ii}) = p_i$ for $1 \leq i \leq L$ (see [36, Sec. IV-A] for a practical algorithm to obtain $\mathbf{Q}$). The squared diagonal elements of matrix $\mathbf{\Sigma}_{B,1}$, denoted by $z_i = \sigma_{B,i}^2$, can be easily obtained as the solution to the following convex problem:

$$\begin{array}{ll} \min\limits_{\{z_i\}} & \sum\limits_{i=1}^{\check{L}} z_i \\ \text{s.t.} & \sum\limits_{j=i}^{\check{L}} \dfrac{1}{\nu + z_j\lambda_j} \leq \sum\limits_{j=i}^{\check{L}} \tilde{\rho}_j, \qquad 1 \leq i \leq \check{L} \\ & z_i \geq 0 \end{array} \tag{25}$$

where

$$\tilde{\rho}_i \triangleq \begin{cases} \sum_{k=1}^{L_0+1} \rho_k - L_0, & i = 1 \\ \rho_{i+L_0}, & 1 < i \leq \check{L} \end{cases} \tag{26}$$

and $\rho_i = g_i^{-1}(p_i)$.[8] The problem is feasible if and only if $\sum_{i=1}^{L} g_i^{-1}(p_i) > L_0$. If feasible, problem (25) has a unique solution.

Proof: The proof is similar to that of [26, Th. 2] (see also [23, Th. 6.2]) and is sketched in Appendix B. ∎

The following result characterizes the optimal value of the problem considered in Theorem 1.

Proposition 1: The optimal objective value of the problem (25) as a function of the BER QoS requirements $\mathbf{p}$ (for any ordering of the p_i's and λ_i's), denoted by $P(\mathbf{p})$, satisfies the following (it is further assumed that each BER function g_i satisfies property P2 as well as P1):

a) The function $P(\mathbf{p})$ is strictly convex on $\mathcal{P} \triangleq \{\mathbf{p} | p_i \in (0, g_i(u_i)), \sum_{i=1}^{L} g_i^{-1}(p_i) > L_0\}$ ($\mathcal{P}$ is the feasible set such that each g_i operates in the strictly convex regime) and convex everywhere.

b) A subgradient of $P(\mathbf{p})$ at some feasible $\mathbf{p}$ is given by $\mathbf{s}(\mathbf{p})$ with components

$$s_i(\mathbf{p}) = \begin{cases} -\tilde{\mu}_1(\mathbf{p})\left(g_i^{-1}\right)'(p_i), & 1 \leq i \leq 1 + L_0 \\ -\tilde{\mu}_{i-L_0}(\mathbf{p})\left(g_i^{-1}\right)'(p_i), & 1 + L_0 < i \leq L \end{cases}$$

where the $\tilde{\mu}_i(\mathbf{p})$ are the Lagrange multipliers of (25) (which can be readily obtained by squaring the water levels $\tilde{\mu}_i^{1/2}$ obtained by Algorithm 1) and $(g_i^{-1})'(p_i) \triangleq 0$ for $p_i > g_i(u_i)$.[9]

c) The function $P(\mathbf{p})$ is differentiable on $\mathcal{P}$.

Proof: See Appendix C. ∎

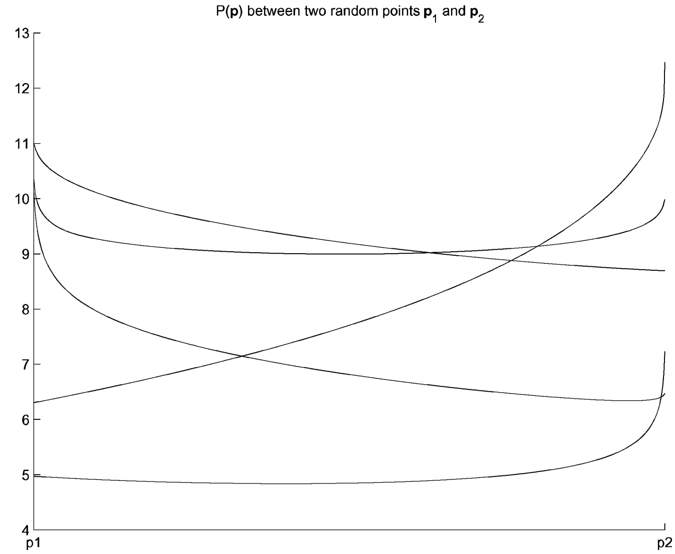

Fig. 5. Some random examples of the function $P(\mathbf{p})$ between two random points $\mathbf{p}_1$ and $\mathbf{p}_2$ (6×6 MIMO channel with $L = 5$ substreams using QPSK and 16-QAM constellations).

In Fig. 5, several examples of the function $P(\mathbf{p})$ are plotted between two random points $\mathbf{p}_1$ and $\mathbf{p}_2$ in the set $\mathcal{P}$, from which the strict convexity and the differentiability can be easily observed.

Interestingly, the convex problem (25) obtained in Theorem 1 can be easily solved in practice with Algorithm 1 (which works for any ordering on $\tilde{\rho}_i$ and λ_i). The algorithm always produces the optimal solution (provided that the problem is feasible) as proved in [26, Prop. 2] [the proof of the optimality of the algorithm is based on showing that the obtained solution satisfies the KKT optimality conditions of the problem (25)].

Algorithm 1 Practical multilevel waterfilling algorithm to solve the convex problem (25).

Input: Dimension of the problem $\check{L}$, set of positive eigenvalues $\{\lambda_i\}_{i=1}^{\check{L}}$, and set of upper bounds $\{\tilde{\rho}_i\}_{i=1}^{\check{L}}$.

Output: Set of allocated powers $\{z_i\}_{i=1}^{\check{L}}$ and set of water-levels $\{\tilde{\mu}_i^{1/2}\}_{i=1}^{\check{L}}$.

[8]The function g_i^{-1} denotes the inverse function of g_i such that $g_i^{-1}(g_i(\rho)) = \rho$.

[9]For completeness, we give a closed-form expression of $g(\rho)$ and $g'(\rho)$ for QAM constellations according to (15) and (18) $g(\rho) = (\alpha/\log_2 M)\mathcal{Q}(\sqrt{\beta(\rho^{-1}-\nu)})$ and $g'(\rho) = (\alpha/\log_2 M)\sqrt{\beta/(8\pi)}e^{-\beta(\rho^{-1}-\nu)/2}(\rho^{-3} - \nu\rho^4)^{-1/2}$. The derivative of the inverse function of $g(\rho)$ on $(0, u)$ can be readily obtained using the relation $(g^{-1})'(p) = 1/g'(\rho)$, where $p = g(\rho)$ [note that $g'(\rho) > 0$ on $(0, u)$] [37, p. 114].

0) Set $i_1 = 1$ and $i_2 = \check{L}$.
1) Solve the waterfilling $z_i = (\mu_{[i_1,i_2]}^{1/2}\lambda_i^{-1/2} - \nu\lambda_i^{-1})^+$ for $i_1 \leq i \leq i_2$ with the constraint $\sum_{i=i_1}^{i_2}(1/\nu + z_i\lambda_i) = \sum_{i=i_1}^{i_2}\tilde{\rho}_i$ [see (27) for $\nu = 0$ and [26, Alg. 1] for $\nu = 1$] and then set $\tilde{\mu}_i = \mu_{[i_1,i_2]}$ for $i_1 \leq i \leq i_2$.
2) If all intermediate constraints are satisfied ($\sum_{j=i}^{i_2}(1/(\nu + z_j\lambda_j)) \leq \sum_{j=i}^{i_2}\tilde{\rho}_j$ for $i_1 < i \leq i_2$) then, if $i_1 = 1$, finish and, if $i_1 > 1$, set $i_2 = i_1 - 1$, $i_1 = 1$, and go to step 1.
   Otherwise, set $i_1 = \min\{i | i_1 < i \leq i_2, \sum_{j=i}^{i_2}(1/(\nu + z_j\lambda_j)) > \sum_{j=i}^{i_2}\tilde{\rho}_j\}$ and go to step 1.

The worst-case number of iterations in Algorithm 1 is $\mathcal{O}(\check{L}^2)$ [26]. However, each iteration requires the computation of the simple waterfilling solution $z_i = (\mu^{1/2}\lambda_i^{-1/2} - \nu\lambda_i^{-1})^+$ with $\sum_i 1/(\nu + z_i\lambda_i) = \rho$. For the ZF receiver ($\nu = 0$), the waterfilling trivially simplifies to

$$z_i = \left(\rho^{-1}\sum_{j=1}^{L}\lambda_j^{-\frac{1}{2}}\right)\lambda_i^{-\frac{1}{2}}. \tag{27}$$

For the MMSE, the waterfilling over n substreams can be easily evaluated in practice with an algorithm with a worst-case number of n basic iterations [26, Alg. 1]; the worst-case total number of basic iterations is then $\mathcal{O}(\check{L}^3)$.

B. Characterization of the Master Problem

In this subsection, the BER-constrained problem in (22) is solved first in a suboptimal way and then in an optimal way using a primal decomposition approach and characterizing the master problem.

1) Heuristic and Suboptimal Solution: A very simple solution to (22) is obtained by imposing the same BER on each of the substreams, i.e., $p_{k,i} = \mathrm{BER}_0\ \forall k, i$ (equal-BER solution), and then using the results of Section IV-A (Theorem 1 and Algorithm 1). In other words, the heuristic solution is obtained by evaluating $P_k(\mathbf{p}_k)$ at $\mathbf{p}_k = \mathrm{BER}_0 \cdot \mathbf{1}$ for $1 \leq k \leq N$, where $\mathbf{1}$ is a vector of ones.

In principle, this solution is not necessarily the optimum. However, for the particular case in which the constellations used at some k are equal, i.e., $\mathcal{C}_{k,i} = \mathcal{C}_k\ \forall i$, then the optimal solution happens to have equal BERs within that k, i.e., $p_{k,i} = p_k\ \forall i$, as was shown in [15] [see also Proposition 2(c)]. As a consequence, if the system is modeled with a single MIMO channel $(N = 1)$ and all the constellations are equal, then the proposed equal-BER solution is the optimum. On the other hand, if the system is modeled with multiple MIMO channels $(N > 1)$ each with equal constellations, then the equal-BER solution will be close to the optimum if the different MIMO channels are sufficiently correlated. Interestingly, for the case of different constellations, the performance of the equal-BER solution happens to be remarkably close to the optimum, as obtained in the numerical results of Section VI.

2) Optimal Solution: In this section, problem (22) is optimally solved by optimizing the $\mathbf{p}_k$'s as well. In principle, it is possible to rewrite problem (22) in convex form and then solve it directly using a general-purpose method. However, the problem can be solved much more efficiently with very simple algorithms by decomposing the original problem into several subproblems controlled by a convex and simple master problem.

Theorem 2: Consider the nonconvex (and complicated) power minimization problem subject to a global average BER constraint in (22), where the constellations $\mathcal{C}_{k,i}$ are assumed w.l.o.g. ordered with increasing size for each k, i.e., $|\mathcal{C}_{k,i}| \leq |\mathcal{C}_{k,i+1}|$, and each $g_{k,i}$ satisfies the properties P1-P3.

Then, the optimal solution satisfies the BER constraint with equality (provided that BER_0 is sufficiently small such that each $g_{k,i}$ operates in the convex regime) and is given by $\mathbf{B}_k = \mathbf{U}_{H_k,1}\mathbf{\Sigma}_{B_k,1}\mathbf{Q}_k$, where all the terms are defined as in Theorem 1 for each k, and the optimal diagonal elements of $\mathbf{\Sigma}_{B_k,1}$ are implicitly obtained by solving the following convex master problem:

$$\begin{aligned} \min_{\{\mathbf{p}_k\}} \quad & \sum_{k=1}^{N} P_k(\mathbf{p}_k) \\ \text{s.t.} \quad & \frac{1}{L_T}\sum_{k=1}^{N}\sum_{i=1}^{L_k} p_{k,i} \leq \mathrm{BER}_0 \\ & 0 \leq p_{k,i} \leq \check{p}_{k,i}, \qquad \forall k, i \end{aligned} \tag{28}$$

where $\check{p}_{k,i} \triangleq g_{k,i}(u_{k,i})$ and $P_k(\mathbf{p}_k)$ is the optimal value of the kth subproblem, which corresponds to the problem considered in Theorem 1. The problem may be unfeasible only if $L_{0,k} > 0$ for some k (this depends on the value of BER_0).

Proof: See Appendix D. ■

Theorem 2 says that the minimum BER problem can be efficiently solved in practice by repeatedly evaluating $P_k(\mathbf{p}_k)$ and adjusting $\mathbf{p}_k\ \forall k$ by the master program [as opposed to the previous heuristic solution based on evaluating $P_k(\mathbf{p}_k)\ \forall k$ just once]. In the following subsection, a simple gradient algorithm guaranteed to converge to the global optimum is proposed to update the variables $\mathbf{p}_k$. We now state some interesting properties of the solution to the minimum BER problem.

Proposition 2: The convex problem (28) satisfies the following properties:

a) If feasible, it has a unique solution.
b) Higher constellations have smaller MSE at an optimal point, or in other words, $\rho_{k,i} \geq \rho_{k,i+1}$.
c) Equal constellations have equal MSE at an optimal point, or in other words, $\rho_{k,i} = \rho_{k,i+1}$ if $|\mathcal{C}_{k,i}| = |\mathcal{C}_{k,i+1}|$.

Proof: See Appendix E. ■

Practical Algorithm for the Master Problem

The master problem obtained in the decomposition of Theorem 2, i.e., (28), is, in general, nondifferentiable and can be optimally solved, for example, with a cutting-plane method or a subgradient method [21], [22], which simply require being able to evaluate $P_k(\mathbf{p}_k)$ and to obtain a subgradient (both available in this case). Subgradient methods are extremely simple to implement, and the update at each iteration is similar to a gradient method, where a subgradient is used in lieu of the gradient (the choice of the setp size follows different guidelines than for gradient methods) [22, Sec. 6.3.1]. However, the objective function in (28) is differentiable, provided that $\sum_{i=1}^{L} g_{k,i}^{-1}(p_{k,i}) > L_{0,k}\ \forall k$ [by Prop. 1(c)] and, under that mild condition (which is always satisfied for the common situation with $L_k \leq \text{rank}(\mathbf{H}_k)\ \forall k$), other methods for differentiable functions can be used instead of the subgradient method. Since our interest is in simple methods that can be readily implemented in any platform, we will focus on a very simple gradient method, called the *conditional gradient method*, which is guaranteed to converge to a global solution [22] and has a very good performance for the problem at hand.

Consider the following general convex minimization problem [which includes (28) as a particular case]:

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in \mathcal{X}. \end{array} \tag{29}$$

To solve the problem (29) and for simplicity, we consider a feasible direction method that obtains a sequence of feasible points $\{\mathbf{x}_k\}$ (here, k denotes the iteration) as [22, Secs. 2.2 and 2.3]

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \tag{30}$$

where $\alpha_k \in (0, 1]$ is a step size, and $\mathbf{d}_k = \bar{\mathbf{x}}_k - \mathbf{x}_k$ is a feasible direction (assuming $\bar{\mathbf{x}}_k \in \mathcal{X}$), which is also a descent direction, i.e., $\nabla f(\mathbf{x}_k)^T \mathbf{d}_k < 0$ (provided that $\mathbf{x}_k$ is not stationary). A very simple way to obtain the feasible direction $\mathbf{d}_k$ or, equivalently, $\bar{\mathbf{x}}_k$ is by the conditional gradient method [22, Sec. 2.2.2]: $\bar{\mathbf{x}}_k = \arg\min_{\mathbf{x}\in\mathcal{X}} \nabla f(\mathbf{x}_k)^T(\mathbf{x} - \mathbf{x}_k)$, which can be efficiently obtained in this case with Algorithm 2 because the set $\mathcal{X}$ is basically a simplex. Interestingly, it is straightforward to define a termination criterion since we can easily compute the following bounds on the optimal value:

$$f(\mathbf{x}_k) \geq \min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \geq f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T(\bar{\mathbf{x}}_k - \mathbf{x}_k). \tag{31}$$

To obtain the step sizes α_k in a simple way, we can opt for a fixed value $\alpha_k = \alpha$ or for a diminishing rule that satisfies $\alpha_k \rightarrow 0$ and $\sum_{k=0}^{\infty} \alpha_k = \infty$ [21], [22], for example, $\alpha_k = \alpha_0((1+m)/(k+m))$, where $\alpha_0 \in (0, 1]$ is the initial step size, and m is a fixed positive integer (see [22, Sec. 2.2] for other choices of step size rules).

Algorithm 2 Practical algorithm to obtain an optimal solution to

$$\begin{array}{ll} \min_{\mathbf{x}} & \sum_{i=1}^{N} w_i x_i \\ \text{s.t.} & 0 \leq x_i \leq x_i^{\max}, \qquad 1 \leq i \leq N \\ & \sum_{i=1}^{N} x_i \leq P. \end{array}$$

Input: Bounds of the feasible set $\mathbf{x}^{\max}$, weights $\mathbf{w}$, and P.
Output: Optimal solution $\mathbf{x}$.

0) Set $\mathbf{x} = \mathbf{0}$.
1) If $\mathbf{1}^T\mathbf{x} = P$ or $w_{\min} \triangleq \min_{i:x_i \neq x_i^{\max}} \{w_i\} \geq 0$, then finish.
2) Set $x_i = \min(x_i + (P - \mathbf{1}^T\mathbf{x})/|\mathcal{I}|, x_i^{\max})$ for $i \in \mathcal{I} \triangleq \{i | x_i \neq x_i^{\max}, w_i = w_{\min}\}$ and go to step 1.

As a final practical note, it is worth pointing out that although the result obtained in Theorem 2 only holds when each function $g_{k,i}$ operates in the convex regime (typically for BERs no higher than 2×10^{-2}), it is possible to relax such constraints in practice without loss of optimality. For example, we can set the upper bounds $\check{p}_{k,i}$ in the master problem of (28) to 10^{-1} as is done in the simulations of Section VI (in such a case, however, the gradient algorithm is only guaranteed to converge to a local minimum due to the nonconvexity of the master problem).

V. Summary of the Transceiver Design

The following is a summary of all the steps necessary to design the transmit and receive matrices:

1) Choose the constellations to be used at each substream (e.g., simply using the same constellation in all substreams or optimizing the constellations by means of a bit allocation based on the gap approximation [24, Part II]).
2) Make a choice between an MMSE receiver ($\nu = 1$) or a ZF receiver ($\nu = 0$).
3) Obtain $\mathbf{U}_{H_k,1}$ from the eigendecomposition of the squared channel matrix $\mathbf{R}_{H_k} = \mathbf{H}_k \mathbf{R}_{n,k}^{-1} \mathbf{H}_k$ for each k (there are efficient methods to obtain eigendecompositions in practice [38]).
4) Obtain $\mathbf{\Sigma}_{B_k,1}$ for each k using
   - the heuristic solution: evaluating $P_k(\mathbf{p}_k)$ (with Algorithm 1) only once at $\mathbf{p}_k = \text{BER}_0 \cdot \mathbf{1}$.
   - the optimal solution: evaluating $P_k(\mathbf{p}_k)$ (with Algorithm 1) several times updating $\mathbf{p}_k$ with the conditional gradient method proposed in Section IV-B-2 [(30) and Algorithm 2].
5) Obtain the optimal rotation $\mathbf{Q}_k$ to satisfy the BERs in $\mathbf{p}_k$ for each k using the practical algorithm in [36, Sec. IV-A] (see also [23, Alg. 3.2]).
6) Finally, obtain the optimal transmitters as $\mathbf{B}_k = \mathbf{U}_{H_k,1}\mathbf{\Sigma}_{B_k,1}\mathbf{Q}_k$ and the optimal receivers as $\mathbf{A}_k = \mathbf{R}_{n,k}^{-1}\mathbf{H}_k\mathbf{B}_k(\nu\mathbf{I} + \mathbf{B}_k^H\mathbf{H}_k^H\mathbf{R}_{n,k}^{-1}\mathbf{H}_k\mathbf{B}_k)^{-1}$ for each k.

VI. Simulation Results

For the numerical simulations, we consider a wireless communication system with multiple antennas at both sides of the link (in particular 4×4) with perfect CSI. The MIMO channels were generated using the parameters of the European standard for Wireless Local Area Networks (WLANs) HIPERLAN/2 [39], which is based on the multicarrier modulation OFDM (64 carriers were used in the simulations), including the frequency selectivity and the spatial correlation as measured in real scenarios. The frequency selectivity of the channel was modeled using the power delay profile type C for HIPERLAN/2, as specified in [40], which corresponds to a typical large open-space indoor environment for nonline-of-sight conditions with 150 ns average r.m.s. delay spread and 1050 ns maximum delay (the sampling period is 50 ns [39]). The spatial correlation of the MIMO channel was modeled according to the *Nokia*

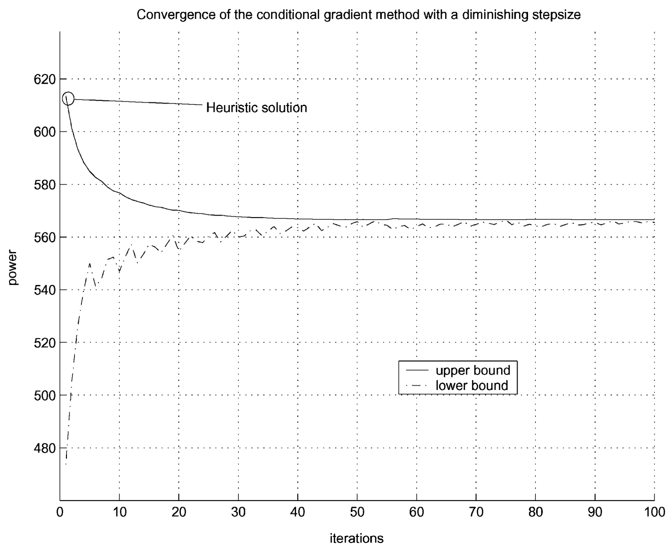

Fig. 6. Convergence of the conditional gradient method for the master problem for a multicarrier 4×4 MIMO channel (carrier-cooperative approach) with $L = 3$ (QPSK, 256-QAM, and 512-QAM constellations).

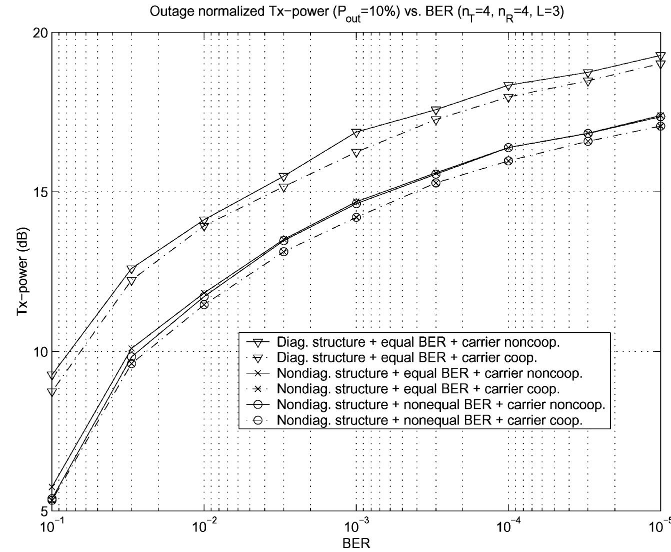

Fig. 7. Outage normalized transmit power versus BER for a multicarrier 4×4 MIMO channel with $L = 3$ (equal QPSK constellations).

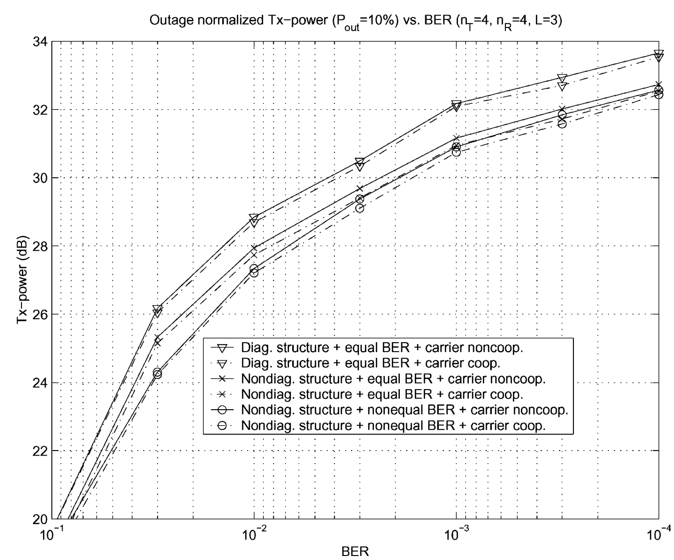

Fig. 8. Outage normalized transmit power versus BER for a multicarrier 4×4 MIMO channel with $L = 3$ (QPSK, 512-QAM, and 512-QAM constellations).

model defined in [41], which corresponds to a conference hall or a shopping galleria scenario, as specified by the correlation matrices of the envelope of the channel fading at the transmit and receive side. The matrix channel generated was normalized so that $\sum_n \mathbb{E}[|[\mathbf{H}(n)]_{ij}|^2] = 1$.

Three methods are compared in the numerical results of this section. First, we consider a benchmark based on imposing a diagonal structure on the MSE matrix $\mathbf{E}$ and the same BER on all links (called diag. structure + equal BER), which basically means that the transmitter is given by $\mathbf{B} = \mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}$ [note that this expression lacks the rotation $\mathbf{Q}$ that appears in the optimal solution (see Theorem 1)]. Then, we consider the two approaches proposed in this paper: the heuristic solution obtained in Section IV-B1 (termed nondiag. structure + equal BER) and the optimal solution obtained in Section IV-B2 (called nondiag. structure + nonequal BER). The multicarrier MIMO channel is modeled in two different ways: as a set of N parallel MIMO channels with an independent processing per carrier (carrier-noncooperative approach) and also as a big MIMO channel with a joint processing over all carriers (carrier-cooperative approach). Thus, the total number of methods evaluated and compared is six. Unless otherwise stated, the MMSE receiver is employed.

The results are given in terms of required transmit power versus BER. To be more specific, we plot the transmitted power normalized with the noise spectral density N_0: $\mathrm{Tr}(\mathbf{B}\mathbf{B}^H)/(N \times N_0)$, where N is the number of carriers.

Convergence of the Conditional Gradient Method: We first consider the convergence of the conditional gradient method with a diminishing step-size rule for the master problem as proposed in Section IV-B2. In Fig. 6, the upper and lower bounds on the transmitted power [see (31)] are plotted as a function of the iterations for a random example with $L = 3$ (QPSK, 256-QAM, and 512-QAM). In all the numerical simulations, the heuristic solution of Section IV-B1 is used as the initial point of the gradient method and the parameters of the step-size rule are $m = 10$ and $\alpha_0 = 0.1$. It can be observed that after 10–20 iterations, most of the gain has already been achieved.

Equal Constellations: In Fig. 7, steady-state results are plotted for a multicarrier 4×4 MIMO channel with $L = 3$ using equal QPSK constellations. It can be observed that the heuristic and optimal solutions have the same performance for the carrier-cooperative approach (as predicted by theory) and an almost indistinguishable performance for the carrier-noncooperative case (c.f. Section IV-B1). For this particular scenario, the gain of the proposed schemes with respect to the benchmark imposing diagonality is of 2–3 dB. This difference is due to the strong suboptimality of the diagonal structure for equal constellations [26, Lem. 6].

Different Constellations: In Fig. 8, steady-state results are shown for a multicarrier 4×4 MIMO channel with $L = 3$ using QPSK, 512-QAM, and 512-QAM constellations. In this

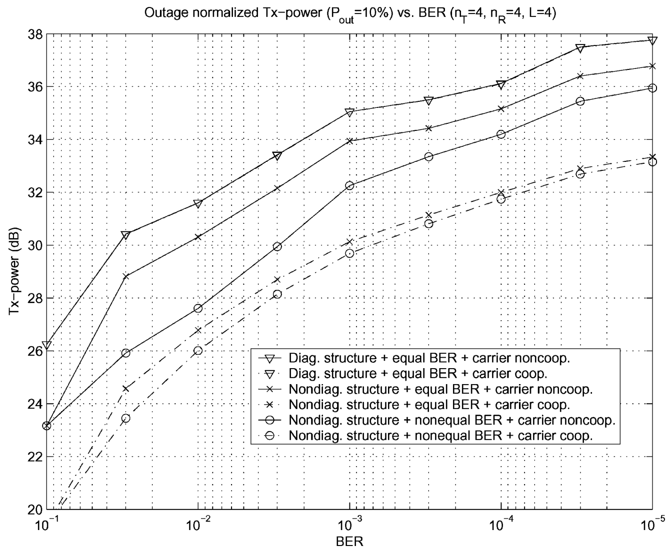

Fig. 9. Outage normalized transmit power versus BER for a multicarrier 4×4 MIMO channel with $L = 4$ (QPSK, 16-QAM, 16-QAM, and 512-QAM constellations).

case, the performance of the heuristic solution is distinguishable from the optimal performance, although it is still very close to it (less than 1 dB). The suboptimality of the benchmark is smaller in this case because with different constellations, the diagonal structure is less suboptimal than with equal constellations. In fact, the more different the constellations are, the less suboptimal the diagonal structure is.

In Fig. 9, similar results are obtained with $L = 4$ using QPSK, 16-QAM, 16-QAM, and 512-QAM constellations. In this case, the loss in performance of the heuristic solution is between 1–2 dB for the carrier-noncooperative approach and less than 1 dB for the carrier-cooperative case. The benchmark works quite poorly in this case because the diagonal structure forces one symbol to be transmitted directly through the worst-channel eigenvalue, which may require a significant amount of power to be equalized.

ZF Receiver: The plots corresponding to the ZF receiver are not included for space limitations and because they can be easily summarized as follows. For $L = 1, 2, 3$, the performance with the ZF receiver is almost indistinguishable with that of the MMSE receiver (especially for BERs lower than 10^{-2}), and for $L = 4$, the difference is still insignificant (less than 0.5 dB).

VII. Conclusion

This paper has dealt with the design of linear transceivers for MIMO channels in terms of minimum BER averaged over the established links. After reformulating the originally complicated and nonconvex problem as a convex problem, a simple heuristic suboptimal solution and the optimal solution have been derived. The simple heuristic solution happens to perform extremely well in practice, becoming a very attractive choice. The optimal solution is also easily computed in practice due to the structure of the problem that allows the decomposition of the original problem into several simple subproblems controlled by a master problem. Such a decomposition is very useful because the subproblems, as well as the master problem, can be easily solved in practice with the simple algorithms that have been proposed.

Appendix A
Proof of Lemma 1

Assume $\rho_1, \rho_2 \in [0, u_2]$; otherwise, (16) is readily verified from property P2. Since $|\mathcal{C}_1| \leq |\mathcal{C}_2|$, it follows from property P3 that $g_1'(\rho) \leq g_2'(\rho)$, and as a consequence, $\int_{\rho_1}^{\rho_2} g_1'(\rho) d\rho \leq \int_{\rho_1}^{\rho_2} g_2'(\rho) d\rho$. Using now $\int_{\rho_1}^{\rho_2} g_i'(\rho) d\rho = g_i(\rho_2) - g_i(\rho_1)$, (16) is obtained. In the case that $|\mathcal{C}_1| < |\mathcal{C}_2|$ [and that $g_1(\rho_2) + g_2(\rho_1) < +\infty$], the result holds with strict inequality since $g_1'(\rho) < g_2'(\rho)$ by property P3.

To prove (17), define the function $f((\rho_1, \rho_2)) \triangleq g_1(\rho_1) + g_2(\rho_2)$. Then, for $0 < \theta < 1$

$$\begin{aligned} & f\left(\theta(\rho_1, \rho_2) + (1-\theta)(\rho_2, \rho_1)\right) \\ & \quad < \theta f\left((\rho_1, \rho_2)\right) + (1-\theta) f\left((\rho_2, \rho_1)\right) \\ & \quad \leq \theta f\left((\rho_1, \rho_2)\right) + (1-\theta) f\left((\rho_1, \rho_2)\right) \\ & \quad = f\left((\rho_1, \rho_2)\right) \end{aligned}$$

where the first inequality follows from the strict convexity of f (property P2) and the second from the fact that $f((\rho_2, \rho_1)) \leq f((\rho_1, \rho_2))$ [this is just a restatement of (16)]. Particularizing for $\theta = 1/2$, (17) is obtained. ■

Appendix B
Proof of Theorem 1

This proof is similar to that of [26, Th. 2] (see also [23, Th. 6.2]), and most details are omitted. We prove the theorem in three steps. First, we eliminate the functions g_i to simplify the notation; then, we show the equivalence between the original complicated problem and a simpler problem; and finally, we solve the simple problem. The two last steps are identical to those in [26, Th. 2].

First of all, since g_i is strictly increasing, we can rewrite each constraint $g_i([(\nu \mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B})^{-1}]_{ii}) \leq p_i$ as $[(\nu \mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B})^{-1}]_{ii} \leq \rho_i \triangleq g_i^{-1}(p_i)$, and problem (23) can be written as

$$\begin{aligned} \min_{\mathbf{B}} \quad & \mathrm{Tr}(\mathbf{B}\mathbf{B}^H) \\ \text{s.t.} \quad & \left[\left(\nu \mathbf{I} + \mathbf{B}^H \mathbf{R}_H \mathbf{B}\right)^{-1}\right]_{ii} \leq \rho_i, \quad 1 \leq i \leq L \end{aligned} \tag{32}$$

which is the problem considered in [26, Th. 2].

It follows from [26, Th. 2] that problem (32) (problem P1) is equivalent to the following problem (problem P2):

$$\begin{aligned} \min_{\tilde{\mathbf{B}}} \quad & \mathrm{Tr}(\tilde{\mathbf{B}}\tilde{\mathbf{B}}^H) \\ \text{s.t.} \quad & \tilde{\mathbf{B}}^H \mathbf{R}_H \tilde{\mathbf{B}} \qquad \text{diagonal (increasing diag. elements)} \\ & \mathbf{d}\left(\left(\nu \mathbf{I} + \tilde{\mathbf{B}}^H \mathbf{R}_H \tilde{\mathbf{B}}\right)^{-1}\right) \succ^w \boldsymbol{\rho} \end{aligned}$$

where $\succ^w$ denotes the weakly majorization relation [19] (the proof of this equivalence hinges on fundamental results of majorization theory [19]). The second constraint guarantees the existence of a unitary matrix $\mathbf{Q}$ such that $\mathbf{d}(\mathbf{Q}^H(\nu\mathbf{I}+\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}})^{-1}\mathbf{Q}) \le \rho$ ($\mathbf{a} \le \mathbf{b}$ refers to the element-wise relation $a_i \le b_i$) [19, 9.B.2 and 5.A.9.a] or, in other words, such that $[(\nu\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B})^{-1}]_{ii} \le \rho_i$ with $\mathbf{B}=\tilde{\mathbf{B}}\mathbf{Q}$ [As can be observed from (23) and is later formally proved, these QoS constraints are satisfied with equality at an optimal point].

Now that problems P1 and P2 have been shown to be equivalent, we focus on solving problem P2, which is much simpler than problem P1. Since in problem P2, the matrix $\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}}$ is diagonal with diagonal elements in increasing order, it follows that $\tilde{\mathbf{B}}$ can be assumed without loss of optimality of the form $\tilde{\mathbf{B}}=\mathbf{U}_{H,1}\mathbf{\Sigma}_{B,1}$ (c.f. [15, Lem. 12] and [23, Lem. 5.11], where $\mathbf{U}_{H,1} \in \mathbb{C}^{n_T\times\check{L}}$ has as columns the eigenvectors of $\mathbf{R}_H$ corresponding to the $\check{L} \triangleq \min(L, \mathrm{rank}(\mathbf{R}_H))$ largest eigenvalues in increasing order, and $\mathbf{\Sigma}_{B,1}=[\mathbf{0}\ \mathrm{diag}(\{\sigma_{B,i}\})] \in \mathbb{C}^{\check{L}\times L}$ has zero elements except along the right-most main diagonal (assumed real w.l.o.g.). Writing the weakly majorization constraint of problem P2 explicitly, defining $z_i \triangleq \sigma_{B,i}^2$, and denoting with the set $\{\lambda_i\}_{i=1}^{\check{L}}$ the $\check{L}$ largest eigenvalues of $\mathbf{R}_H$ in increasing order, the problem reduces to

$$
\begin{aligned}
\min_{\{z_i\}}\quad & \sum_{i=1}^{\check{L}} z_i \\
\text{s.t.}\quad & \sum_{j=i-L_0}^{\check{L}} \frac{1}{\nu+z_j\lambda_j} \le \sum_{j=i}^{L}\rho_j, \qquad L_0 < i \le L \\
& (L_0-i+1)+\sum_{j=1}^{\check{L}} \frac{1}{\nu+z_j\lambda_j} \le \sum_{j=i}^{L}\rho_j, \quad 1\le i\le L_0 \\
& z_i \ge 0.
\end{aligned}
$$

The desired problem formulation of (25) follows by noting that, since $\rho_i < 1$, the constraints for $1 \le i \le L_0$ (in case that $L_0 > 0$) imply and are implied by the constraint for $i=1$: $\sum_{j=1}^{\check{L}}(1/(\nu+z_j\lambda_j)) \le \sum_{j=1}^{L}\rho_j - L_0$.

The constraints can be satisfied for sufficiently large values of z_i (equivalently, the problem is feasible) if and only if $\sum_{j=i}^{L}\rho_j > 0$ for $L_0 < i \le L$ (always satisfied because $\rho_i > 0$) and $\sum_{i=1}^{L}\rho_i > L_0$.

It is straightforward to see that $\sum_{j=1}^{\check{L}}(1/(\nu+z_j\lambda_j)) \le \sum_{j=1}^{L}\rho_j - L_0$ must be satisfied with equality at an optimal point. Otherwise, z_1 could be decreased until it is satisfied with equality or z_1 becomes zero (in which case, the same reasoning applies to z_2 and so forth). This means that an optimal solution to problem P2 must satisfy $\mathbf{d}((\nu\mathbf{I}+\tilde{\mathbf{B}}^H\mathbf{R}_H\tilde{\mathbf{B}})^{-1}) \succ \boldsymbol{\rho}$, which in turn implies that $g_i([(\nu\mathbf{I}+\mathbf{B}^H\mathbf{R}_H\mathbf{B})^{-1}]_{ii}) = p_i$.

The uniqueness of the solution of (25) follows easily from the strict convexity of each function $1/(\nu+z_i\lambda_i)$ (recall that λ_i is positive), since we could otherwise form a convex combination of two different optimal solutions, resulting in $\sum_{j=1}^{\check{L}}(1/(\nu+z_j\lambda_j)) < \sum_{j=1}^{L}\rho_j - L_0$, which means that a strictly better solution could be obtained. ■

Appendix C
Proof of Proposition 1

We now prove the three properties stated in Proposition 1. It is important to realize that we can use $\check{p}_i = \min(p_i, g_i(u_i))$ instead of p_i in the problems (23) and (25) without affecting the solution. For convenience of notation, we define the following set:

$$
\mathcal{Z}(\tilde{\boldsymbol{\rho}}) \triangleq \left\{ \mathbf{z} | z_i \ge 0, \sum_{j=i}^{\check{L}} \frac{1}{\nu+z_j\lambda_j} \le \sum_{j=i}^{\check{L}} \tilde{\rho}_j, \quad 1\le i\le \check{L} \right\}
$$

and with some abuse of notation, we also define $\mathcal{Z}(\boldsymbol{\rho}) \triangleq \mathcal{Z}(\tilde{\boldsymbol{\rho}}(\boldsymbol{\rho}))$ and $\mathcal{Z}(\mathbf{p}) \triangleq \mathcal{Z}(\boldsymbol{\rho}(\mathbf{p}))$, where $\tilde{\rho}_i$ is given by (26) and $\rho_i = g_i^{-1}(\check{p}_i)$. Note that all previous sets are convex in $\mathbf{z}$. Similarly, we will refer to $P(\tilde{\boldsymbol{\rho}})$ and $P(\mathbf{p})$ as the minimum required power $\sum_{i=1}^{\check{L}} z_i$ among $\mathbf{z}$ in the sets $\mathcal{Z}(\tilde{\boldsymbol{\rho}})$, $\mathcal{Z}(\boldsymbol{\rho})$, and $\mathcal{Z}(\mathbf{p})$, respectively [recall that $P(\mathbf{p}) = P(\check{\mathbf{p}})$].

1) The function $P(\mathbf{p})$ is strictly convex on $\mathcal{P}$ and convex everywhere.

Let $P(\mathbf{z},\mathbf{p}) \triangleq \begin{cases} \sum_{i=1}^{\check{L}} z_i & \mathbf{z}\in\mathcal{Z}(\mathbf{p}) \\ +\infty & \text{otherwise} \end{cases}$, and recall that $P(\mathbf{p}) = \inf_{\mathbf{z}} P(\mathbf{z},\mathbf{p})$ is the optimal value of (25) for a given $\mathbf{p}$, which is achieved at a unique point $\mathbf{z}^\star(\mathbf{p})$ by Theorem 1 [if $\mathcal{Z}(\mathbf{p})$ is not empty]. It follows that $P(\mathbf{p})$ is convex in $\mathbf{p}$ because $P(\mathbf{z},\mathbf{p})$ is jointly convex in $\mathbf{p}$ and $\mathbf{z}$ [25, Sec. 3.2.5], [22, Sec. 5.4.4] [in fact, $P(\mathbf{z},\mathbf{p})$ is linear on the feasible set $\mathcal{ZP} \triangleq \{(\mathbf{z},\mathbf{p})|\mathbf{z}\in\mathcal{Z}(\mathbf{p})\}$].

We now show that it is indeed strictly convex on $\mathcal{P}$ due to the strict concavity of each g_i^{-1} on $[0, g_i(u_i)]$ [which follows since g_i is strictly convex and increasing on $[0,u_i]$]. For $\mathbf{p}_1 \ne \mathbf{p}_2$ (both in $\mathcal{P}$) and $\theta\in(0,1)$, we have

$$
\begin{aligned}
&\theta P(\mathbf{p}_1) + (1-\theta)P(\mathbf{p}_2) \\
&\quad = \theta P(\mathbf{z}^\star(\mathbf{p}_1),\mathbf{p}_1) + (1-\theta)P(\mathbf{z}^\star(\mathbf{p}_2),\mathbf{p}_2) \\
&\quad = P(\theta\mathbf{z}^\star(\mathbf{p}_1) + (1-\theta)\mathbf{z}^\star(\mathbf{p}_2), \theta\mathbf{p}_1 + (1-\theta)\mathbf{p}_2) \\
&\quad > P(\mathbf{z}^\star(\theta\mathbf{p}_1 + (1-\theta)\mathbf{p}_2), \theta\mathbf{p}_1 + (1-\theta)\mathbf{p}_2) \\
&\quad = P(\theta\mathbf{p}_1 + (1-\theta)\mathbf{p}_2)
\end{aligned}
$$

where the second equality follows from the linearity of $P(\mathbf{z},\mathbf{p})$ on the convex set $\mathcal{ZP}$, and the strict inequality follows from the strict concavity of each g_i^{-1} (since some constraints would be satisfied with strict inequality and some z_i could be decreased).

2) A subgradient of $P(\mathbf{p})$ at $\mathbf{p}\in\mathcal{P}$.

First, recall that the Lagrangian corresponding to problem (25) is

$$
\begin{aligned}
&L(\mathbf{z},(\boldsymbol{\mu},\boldsymbol{\gamma})) \\
&\quad = \sum_{i=1}^{\check{L}} z_i + \sum_{i=1}^{\check{L}} \mu_i \left(\sum_{j=i}^{\check{L}} \frac{1}{\nu+z_j\lambda_j} - \sum_{j=i}^{\check{L}} \tilde{\rho}_j \right) - \sum_{i=1}^{\check{L}} \gamma_i z_i
\end{aligned}
$$

and the dual function is $g(\boldsymbol{\mu}, \boldsymbol{\gamma}) = \inf_{\mathbf{z}} L(\mathbf{z}, (\boldsymbol{\mu}, \boldsymbol{\gamma}))$. Suppose for the moment that $\mathbf{p}_0 \in \mathcal{P}$ and $\mathbf{z} \in \mathcal{Z}(\mathbf{p})$. Then, by strong duality (Slater's condition is satisfied), we have (see [22, Sec. 5.4.4] for a similar approach)

$$\begin{aligned} P(\mathbf{p}_0) &= P(\tilde{\boldsymbol{\rho}}_0) = g\left(\boldsymbol{\mu}^\star(\tilde{\boldsymbol{\rho}}_0), \boldsymbol{\gamma}^\star(\tilde{\boldsymbol{\rho}}_0)\right) \\ &\leq L\left(\mathbf{z}, \left(\boldsymbol{\mu}^\star(\tilde{\rho}_0), \boldsymbol{\gamma}^\star(\tilde{\boldsymbol{\rho}}_0)\right)\right) \\ &\leq \sum_{i=1}^{\check{L}} z_i + \sum_{i=1}^{\check{L}} \mu_i^\star(\tilde{\rho}_0) \sum_{j=i}^{\check{L}} (\tilde{\rho}_j - \tilde{\rho}_{0,j}) \\ &= \sum_{i=1}^{\check{L}} z_i + \sum_{i=1}^{\check{L}} \mu_i^\star(\tilde{\rho}_0)(\tilde{\rho}_i - \tilde{\rho}_{0,i}) \\ &= \sum_{i=1}^{\check{L}} z_i + \tilde{\mu}_i^\star(\tilde{\boldsymbol{\rho}}_0) \sum_{j=1}^{L_0+1} (\rho_j - \rho_{0,j}) \\ &\quad + \sum_{i=2}^{\check{L}} \tilde{\mu}_i^\star(\tilde{\boldsymbol{\rho}}_0)(\rho_{i+L_0} - \rho_{0,i+L_0}) \\ &\leq \sum_{i=1}^{\check{L}} z_i + \tilde{\mu}_1^\star(\mathbf{p}_0) \sum_{i=1}^{L_0+1} \left(g_i^{-1}\right)'(p_{0,i})(p_i - p_{0,i}) \\ &\quad + \sum_{i=2+L_0}^{L} \tilde{\mu}_{i-L_0}^\star(\mathbf{p}_0) \left(g_i^{-1}\right)'(p_{0,i})(p_i - p_{0,i}) \end{aligned}$$

where $\tilde{\mu}_i^\star(\tilde{\boldsymbol{\rho}}) = \sum_{j=1}^{i} \mu_j^\star(\tilde{\rho})$, and the last inequality follows from the concavity of g_i^{-1} on $[0, g_i(u_i)]$ (note that strict convexity of g_i is not required for this result) and the relation $\check{p}_i \leq p_i$ [note that $(g_i^{-1})'(p_{0,i}) \geq 0$]. Now, if $\mathbf{p}_0$ is feasible but $p_{0,i} > g_i(u_i)$ for some i, then $\rho_i - \rho_{0,i} \leq 0$ and the two last inequalities hold with $(g_i^{-1})'(p_{0,i})$ replaced by 0. Thus, $P(\mathbf{p}) \geq P(\mathbf{p}_0) + \mathbf{s}^T(\mathbf{p}_0)(\mathbf{p} - \mathbf{p}_0)\ \forall \mathbf{p}$, where $\mathbf{s}(\mathbf{p})$ is defined as in Proposition 1.

3) The function $P(\mathbf{p})$ is differentiable on $\mathcal{P}$.

To prove the differentiability of $P(\mathbf{p})$, it suffices to show that $P(\tilde{\boldsymbol{\rho}})$ is differentiable, since function $P(\mathbf{p})$ is a composition of functions $P(\tilde{\boldsymbol{\rho}})$ and $\tilde{\boldsymbol{\rho}}(\mathbf{p})$, and $\tilde{\boldsymbol{\rho}}(\mathbf{p})$ is differentiable on $\{\mathbf{p} | p_i \in (0, g_i(u_i))\}$ [$\tilde{\rho}_i(\boldsymbol{\rho})$ is given in (26) and $\rho_i = g_i^{-1}(p_i)$] [37, Th. 9.15].

We will prove differentiability of $P(\tilde{\boldsymbol{\rho}})$ by showing that it has a unique subgradient at each $\tilde{\boldsymbol{\rho}}$ [22, Prop. B.24]. Instead of proving the uniqueness of the subgradient directly, we will do so by first showing that there is a one-to-one mapping between subgradients and optimal Lagrange multipliers and then showing the uniqueness of the latter (the constraint $\sum_{i=1}^{L} g_i^{-1}(p_i) > L_0$ is needed to avoid $P(\tilde{\boldsymbol{\rho}}) = +\infty$ and ensure the existence of optimal Lagrange multipliers). We have already seen that an optimal Lagrange multiplier determines uniquely a subgradient [see Property (b) with $g_i(\rho) = \rho$]. It remains to show the opposite, i.e., that a subgradient uniquely determines an optimal Lagrange multiplier.

Suppose that $\tilde{\mathbf{s}}(\tilde{\boldsymbol{\rho}}_0)$ is a subgradient of $P(\tilde{\boldsymbol{\rho}})$ at $\tilde{\boldsymbol{\rho}}_0$:

$$P(\tilde{\boldsymbol{\rho}}_0) \leq P(\tilde{\boldsymbol{\rho}}) - \tilde{\mathbf{s}}^T(\tilde{\boldsymbol{\rho}}_0)(\tilde{\boldsymbol{\rho}} - \tilde{\boldsymbol{\rho}}_0), \qquad \forall \tilde{\boldsymbol{\rho}}.$$

First of all, note that since $P(\tilde{\boldsymbol{\rho}})$ is nonincreasing in each component of $\tilde{\boldsymbol{\rho}}$, it follows that $\tilde{s}_i \leq 0\ \forall i$(otherwise, the RHS of the previous inequality would be unbounded below). Since $P(\tilde{\boldsymbol{\rho}}) = \inf_{\mathbf{z} \in \mathcal{Z}(\tilde{\boldsymbol{\rho}})} \sum_{i=1}^{\check{L}} z_i$, we can equivalently write

$$P(\tilde{\boldsymbol{\rho}}_0) \leq \inf_{\mathbf{z} \in \mathcal{Z}(\tilde{\boldsymbol{\rho}})} \left\{ \sum_{i=1}^{\check{L}} z_i - \tilde{\mathbf{s}}^T(\tilde{\boldsymbol{\rho}}_0)(\tilde{\boldsymbol{\rho}} - \tilde{\boldsymbol{\rho}}_0) \right\}, \qquad \forall \tilde{\boldsymbol{\rho}}.$$

Therefore

$$\begin{aligned} P(\tilde{\boldsymbol{\rho}}_0) &\leq \inf_{\tilde{\boldsymbol{\rho}}} \inf_{\mathbf{z} \in \mathcal{Z}(\tilde{\boldsymbol{\rho}})} \left\{ \sum_{i=1}^{\check{L}} z_i - \sum_{i=1}^{\check{L}} \tilde{s}_i(\tilde{\boldsymbol{\rho}}_0)(\tilde{\rho}_i - \tilde{\rho}_{0,i}) \right\} \\ &= \inf_{\tilde{\boldsymbol{\rho}}} \inf_{\mathbf{z} \in \mathcal{Z}(\tilde{\boldsymbol{\rho}})} \left\{ \sum_{i=1}^{\check{L}} z_i - \sum_{i=1}^{\check{L}} s_i(\tilde{\boldsymbol{\rho}}_0) \sum_{j=i}^{\check{L}} (\tilde{\rho}_j - \tilde{\rho}_{0,j}) \right\} \\ &= \inf_{\mathbf{z} \geq 0} \left\{ \sum_{i=1}^{\check{L}} z_i - \sum_{i=1}^{\check{L}} s_i(\tilde{\boldsymbol{\rho}}_0) \right. \\ &\qquad \left. \times \left(\sum_{j=i}^{\check{L}} \frac{1}{\nu + z_j \lambda_j} - \sum_{j=i}^{\check{L}} \tilde{\rho}_{0,j} \right) \right\} \\ &= g\left(-\mathbf{s}(\tilde{\boldsymbol{\rho}}_0)\right) \end{aligned}$$

where $s_i = \tilde{s}_i - \tilde{s}_{i-1}$ ($\tilde{s}_0 \triangleq 0$), and $g(\boldsymbol{\mu})$ is the dual function corresponding to problem (25). However, from weak duality, $P(\tilde{\boldsymbol{\rho}}_0) \geq \sup_{\boldsymbol{\mu} \geq \mathbf{0}} g(\boldsymbol{\mu})$. Thus, $P(\tilde{\boldsymbol{\rho}}_0) = g(-\mathbf{s}(\tilde{\boldsymbol{\rho}}_0))$, and $-\mathbf{s}(\tilde{\boldsymbol{\rho}}_0)$ is an optimal Lagrange multiplier.

We now prove the uniqueness of the optimal Lagrange multipliers of problem (25). From Theorem 1, we know that the optimal value is achieved at a unique point $\mathbf{z}^\star(\tilde{\rho})$. Then, it suffices to show that the Lagrange multipliers that satisfy the following simplified KKT conditions corresponding to (25) (c.f. [26, Prop. 2]]) jointly with $\mathbf{z}^\star(\tilde{\boldsymbol{\rho}})$ are unique:

$$\begin{aligned} & z_i = \left(\tilde{\mu}_i^{\frac{1}{2}} \lambda_i^{-\frac{1}{2}} - \nu \lambda_i^{-1}\right)^+, \qquad 1 \leq i \leq \check{L} \\ & \sum_{j=i}^{\check{L}} \frac{1}{\nu + z_j \lambda_j} \leq \sum_{j=i}^{\check{L}} \tilde{\rho}_j \\ & \sum_{j=1}^{\check{L}} \frac{1}{\nu + z_j \lambda_j} = \sum_{j=1}^{\check{L}} \tilde{\rho}_j \\ & \tilde{\mu}_i \geq \tilde{\mu}_{i-1} \quad (\tilde{\mu}_0 \triangleq 0) \\ & (\tilde{\mu}_i - \tilde{\mu}_{i-1}) \left(\sum_{j=i}^{\check{L}} \frac{1}{\nu + z_j \lambda_j} - \sum_{j=i}^{\check{L}} \tilde{\rho}_j \right) = 0. \end{aligned} \tag{33}$$

This is clearly true for the Lagrange multipliers $\tilde{\mu}_i$ corresponding to $z_i > 0$ since they are uniquely determined by z_i. For $z_i = 0$, it follows from the following lemma that $\tilde{\mu}_i = \tilde{\mu}_{i_0}$, where $\tilde{\mu}_{i_0}$ is uniquely determined by $z_{i_0} > 0$. ■

Lemma 2: The optimal solution of problem (25) satisfies $\begin{cases} z_i = 0, & i < i_0 \\ z_i > 0, & i \geq i_0 \end{cases}$, and the Lagrange multipliers satisfy $\tilde{\mu}_i = \tilde{\mu}_{i_0} > 0$ for $i < i_0$.

Proof: For $\nu = 0$, all z_i must be positive at any feasible point, and the result is trivial. We next focus on the case $\nu = 1$.

From $z_i\lambda_i = (\tilde{\mu}_i^{1/2}\lambda_i^{1/2} - 1)^+$, $\tilde{\mu}_{i+1} \geq \tilde{\mu}_i$, and $\lambda_{i+1} \geq \lambda_i$, it follows that $z_{i+1}\lambda_{i+1} \geq z_i\lambda_i$. In other words, if $z_i > 0$, then $z_{i+1} > 0$, and conversely, if $z_{i+1} = 0$, then $z_i = 0$. We can then find i_0 as $i_0 = \min\{i | z_i > 0\}$, and the first part of the result is proved. To prove the second part, note that the constraints $\sum_{j=i}^{\check{L}}(1/(1+z_j\lambda_j)) \leq \sum_{j=i}^{\check{L}}\tilde{\rho}_j$ can be rewritten as

$$\sum_{j=i_0}^{\check{L}} \frac{1}{1+z_j\lambda_j} \leq \sum_{j=i_0}^{\check{L}} \tilde{\rho}_j - \sum_{j=i}^{i_0-1}(1-\tilde{\rho}_j), \qquad i < i_0.$$

Since $\sum_{j=1}^{\check{L}}(1/(1+z_j\lambda_j)) = \sum_{j=1}^{\check{L}}\tilde{\rho}_j$ at an optimal point and $\tilde{\rho}_i < 1$, it follows that

$$\sum_{j=i}^{\check{L}} \frac{1}{1+z_j\lambda_j} < \sum_{j=i}^{\check{L}} \tilde{\rho}_j, \qquad 1 < i \leq i_0$$

and, from the complementary slackness conditions in (33), that $\tilde{\mu}_i = \tilde{\mu}_{i-1}$ for $1 < i \leq i_0$. ■

Appendix D
Proof of Theorem 2

The proof is straightforward by noting that problem (22) decouples naturally into several subproblems like problem (23) if the $\mathbf{p}_k$ are fixed. We can then invoke (10) to minimize the problem with respect to the $\mathbf{B}_k$ for fixed $\mathbf{p}_k$ and then with respect to the $\mathbf{p}_k$. The original nonconvex problem (22) can be rewritten as

$$\begin{array}{ll} \min\limits_{\{\mathbf{p}_k\}} & \sum\limits_{k=1}^{N} P_k(\mathbf{p}_k) \\ \text{s.t.} & \frac{1}{L_T}\sum\limits_{k=1}^{N}\sum\limits_{i=1}^{L_k} p_{k,i} \leq \text{BER}_0. \end{array}$$

Now, since each of the subproblems $P_k(\mathbf{p}_k)$ is unfeasible if $p_{k,i} \leq 0$ for some i and having $p_{k,i} > \check{p}_{k,i} \triangleq g_{k,i}(u_{k,i})$ does not decrease the required power (from property P2), we can explicitly write such constraints without affecting the problem, and then (28) is obtained, which is strictly convex.

There is, however, a small hidden detail: Problem (23) assumes that the $p_{k,i}$ are ordered such that $\rho_{k,i} \geq \rho_{k,i+1}$ $\forall k, i$, and this may not be the case when evaluating the subproblems $P_k(\mathbf{p}_k)$. In the following, we take care of this detail, which happens to be irrelevant as long as the constellations $\mathcal{C}_{k,i}$ are ordered with increasing size $|\mathcal{C}_{k,i}| \leq |\mathcal{C}_{k,i+1}|$ $\forall k, i$.

The problem (22) can be rewritten as

$$\begin{array}{ll} \min\limits_{\{\mathbf{B}_k\},\{\boldsymbol{\rho}_k\},\{\mathbf{p}_k\}} & \sum\limits_{k=1}^{N} \text{Tr}\left(\mathbf{B}_k\mathbf{B}_k^H\right) \\ \text{s.t.} & \left[\left(\nu\mathbf{I} + \mathbf{B}_{k,i}^H\mathbf{R}_{H_k}\mathbf{B}_{k,i}\right)^{-1}\right]_{ii} \leq \rho_{k,i} \\ & 1 \leq i \leq L_k,\ 1 \leq k \leq N \\ & g_{k,i}(\rho_{k,i}) \leq p_{k,i} \\ & \frac{1}{L_T}\sum\limits_{k=1}^{N}\sum\limits_{i=1}^{L_k} p_{k,i} \leq \text{BER}_0 \end{array} \tag{34}$$

where we have introduced the additional variables $\rho_{k,i}$. We can now make the problem convex exactly as was done in Theorem 1, obtaining

$$\begin{array}{ll} \min\limits_{\{\mathbf{z}_k\},\{\boldsymbol{\rho}_k\},\{\mathbf{p}_k\}} & \sum\limits_{k=1}^{N}\sum\limits_{i=1}^{L_k} z_{k,i} \\ \text{s.t.} & \sum\limits_{j=i}^{\check{L}_k} \frac{1}{\nu + z_{k,j}\lambda_{k,j}} \leq \sum\limits_{j=i+L_{0,k}}^{L_k} \rho_{k,[j]} \\ & 1 \leq i \leq \check{L}_k,\ 1 \leq k \leq N \\ & \sum\limits_{j=1}^{\check{L}_k} \frac{1}{\nu + z_{k,j}\lambda_{k,j}} \leq \sum\limits_{j=1}^{L_k} \rho_{k,[j]} - L_{0,k} \\ & z_{k,i} \geq 0 \\ & g_{k,i}(\rho_{k,i}) \leq p_{k,i} \\ & \frac{1}{L_T}\sum\limits_{k=1}^{N}\sum\limits_{i=1}^{L_k} p_{k,i} \leq \text{BER}_0 \end{array} \tag{35}$$

where $\rho_{k,[j]}$ denotes the elements in decreasing order $\rho_{k,[j]} \geq \rho_{k,[j+1]}$. Note that at an optimal point, it must be that $g_{k,i}(\rho_{k,i}) = p_{k,i}$ and $\sum_{j=1}^{\check{L}_k}(1/\nu + z_{k,j}\lambda_{k,j}) = \sum_{j=1}^{L_k}\rho_{k,[j]} - L_{0,k}$ by Theorem 1.

In the first two sets of constraints in (35), we require $\rho_{k,i}$ in decreasing order, which is denoted by the notation $\rho_{k,[i]}$. In principle, this is not a problem, since such a constraint is still convex, as can be seen from the following relation [25]:

$$\sum_{i=j}^{n} x_{[i]} = \min\left\{x_{i_1} + x_{i_2} + \cdots + x_{i_{n-j+1}} \mid 1 \leq i_1 < i_2 < \cdots < i_{n-j+1} \leq n\right\} \tag{36}$$

where the RHS is clearly a concave function since it is the pointwise minimum of concave (affine) functions. Nevertheless, it is simpler in practice to deal with unordered $\rho_{k,i}$. If we simply remove the ordering constraint, the problem relaxes, and therefore, it may not be equivalent to the original one. However, by Proposition 2(b), we know that any optimal solution of the relaxed problem has $\rho_{k,i}$ in decreasing order for each k (due to the assumed ordering of the constellations). As a consequence, we can safely remove the ordering constraints on the $\rho_{k,i}$, and the resulting problem is equivalent. The decomposition principle of (10) can then be applied to the relaxed version of (35) using (25) as a subproblem, obtaining (28). ■

Appendix E
Proof of Proposition 2

1) If feasible, the convex problem (28) has a unique solution.

   By Proposition 1(a), each $P_k(\mathbf{p}_k)$ is strictly convex on $\mathcal{P}$, and so is the objective function $\sum_{k=1}^{N} P_k(\mathbf{p}_k)$. It follows then that, if feasible, the problem can only have one solution (for, if there were two optimal solutions, a convex combination of them would give a strictly lower objective value, which cannot be).

In fact, the convex problem (35) (and also the relaxed version) prior to the decomposition (this problem not only has $p_{k,i}$ as variables but also $\rho_{k,j}$ and $z_{k,i}$) has a unique solution as well since the solution to each of the subproblems is unique by Theorem 1.

2) Higher constellations have smaller MSE at an optimal point of (28) or, equivalently, of the relaxed version of the problem (35). In other words, $\rho_{k,i} \geq \rho_{k,i+1}$.

Suppose that $\rho_{k,i} < \rho_{k,i+1}$ for some k. We next show that this cannot be since the problem would have a strictly smaller objective value using instead $\tilde{\rho}_{k,i} = \rho_{k,i} + \epsilon$ and $\tilde{\rho}_{k,i+1} = \rho_{k,i+1} - \epsilon$, where $\epsilon = (\rho_{k,i+1} - \rho_{k,i})/2$ is a positive value.

Consider the relaxed version of the problem (35). The only constraints that are affected by using $\tilde{\rho}_{k,i}$ and $\tilde{\rho}_{k,i+1}$ instead of $\rho_{k,i}$ and $\rho_{k,i+1}$ are

$$\sum_{j=i+1-L_{0,k}}^{\check{L}_k} \frac{1}{\nu + z_{k,j}\lambda_{k,j}} \leq \sum_{j=i+1}^{L_k} \rho_{k,j} \quad (\text{for } i \geq L_{0,k}) \tag{37}$$

and

$$\frac{1}{L_T} \sum_{k=1}^{N} \sum_{i=1}^{L_k} g_{k,i}(\rho_{k,i}) \leq \mathrm{BER}_0 \tag{38}$$

where we have made use of the fact that $g_{k,i}(\rho_{k,i}) = p_{k,i}$ at an optimal point. If we consider instead problem (35) with the ordering constraints, the only constraint affected is (38).

By Lemma 1, the term $(1/L_T)\sum_{k,i} g_{k,i}(\rho_{k,i})$ is strictly decreased when using $\tilde{\rho}_{k,i}$ and $\tilde{\rho}_{k,i+1}$. This means that we could increase both $\tilde{\rho}_{k,i}$ and $\tilde{\rho}_{k,i+1}$ until the average BER constraint (38) was satisfied with equality, which would require strictly less power and would lead to a lower objective value.

We now show that the required power to satisfy (37) when using $\tilde{\rho}_{k,i}$ and $\tilde{\rho}_{k,i+1}$ is not increased. If $i < L_{0,k}$, then (37) is not affected. If $i = L_{0,k}$, then (37) is a loose constraint [recall that the problem has the additional tighter constraint $\sum_{j=1}^{\check{L}_k}(1/(\nu + z_{k,j}\lambda_{k,j})) \leq \sum_{j=i+1}^{L_k} \rho_{k,j} - (L_{0,k} - \sum_{j=1}^{i} \rho_{k,j})$], and we can directly use $\tilde{\rho}_{k,i}$ and $\tilde{\rho}_{k,i+1}$ without affecting it (this is because the amount in which $\rho_{k,i+1}$ is decreased satisfies $\epsilon < 1 - \rho_{k,i} < L_{0,k} - \sum_{j=1}^{i} \rho_{k,j}$). We can then consider the case $i > L_{0,k}$. Defining $\alpha_{k,i+L_{0,k}} \triangleq (1/(\nu + z_{k,i}\lambda_{k,i}))$, (37) can be rewritten as

$$\alpha_{k,i+1} \leq \rho_{k,i+1} + c_{k,i}$$

where $c_{k,i} \triangleq \sum_{j=i+2}^{L_k} \rho_{k,j} - \sum_{j=i+2-L_{0,k}}^{\check{L}_k} (1/(\nu + z_{k,j}\lambda_{k,j})) \geq 0$. We can focus on the case in which $\alpha_{k,i+1} = \rho_{k,i+1} + c_{k,i}$; otherwise, we can decrease $\rho_{k,i+1}$ (and increase $\rho_{k,i}$ by the same amount) either until it is satisfied or until $\rho_{k,i+1} = \rho_{k,i}$, in which case, the proof is complete. Then, we can write (37) and the constraint corresponding to i as

$$\alpha_{k,i+1} = \rho_{k,i+1} + c_{k,i}$$
$$\alpha_{k,i} + \alpha_{k,i+1} \leq \rho_{k,i} + \rho_{k,i+1} + c_{k,i}$$

from which we can readily obtain that $\alpha_{k,i} < \alpha_{k,i+1}$ ($\alpha_{k,i} \leq \rho_{k,i} < \rho_{k,i+1} = \alpha_{k,i+1} - c_{k,i} \leq \alpha_{k,i+1}$). Recalling that $\alpha_{k,i}$ and $\alpha_{k,i+1}$ are achieved with some power allocation $z_{k,i-L_{0,k}}$ and $z_{k,i+1-L_{0,k}}$ (simply from the definition of $\alpha_{k,i}$), we just have to show that $\tilde{\alpha}_{k,i+1} = \alpha_{k,i+1} - \epsilon$ and $\tilde{\alpha}_{k,i} = \alpha_{k,i} + \epsilon$ (corresponding to the constraints involving $\tilde{\rho}_{k,i}$ and $\tilde{\rho}_{k,i+1}$) can be achieved with strictly less power $\tilde{z}_{k,i-L_{0,k}}$ and $\tilde{z}_{k,i+1-L_{0,k}}$.

Recalling $\alpha_{k,i} < \alpha_{k,i+1}$ and noting that $\alpha_{k,i} + \epsilon \leq \rho_{k,i} + \epsilon = \rho_{k,i+1} - \epsilon \leq \alpha_{k,i+1} - \epsilon$, it follows that $\alpha_{k,i}(\alpha_{k,i} + \epsilon) < \alpha_{k,i+1}(\alpha_{k,i+1} - \epsilon)$. After some manipulations and using $0 < \lambda_{k,i-L_{0,k}} \leq \lambda_{k,i+1-L_{0,k}}$, it can be shown that

$$z_{k,i-L_{0,k}} + z_{k,i+1-L_{0,k}} > \tilde{z}_{k,i+1-L_{0,k}} + \tilde{z}_{k,i-L_{0,k}}.$$

Thus, we have proved that $\rho_{k,i} \geq \rho_{k,i+1}$ at an optimal point of the relaxed version of the problem (28) and of the problem (35). This shows the equivalence between these two problems in the sense that the set of optimal solutions of one problem equals the set of optimal solutions of the other one.

3) Equal constellations have equal MSE at an optimal point: $\rho_{k,i} = \rho_{k,i+1}$ if $|\mathcal{C}_{k,i}| = |\mathcal{C}_{k,i+1}|$.

From result 2), we know that at an optimal point, we must have $\rho_{k,i} \geq \rho_{k,i+1}$. We now show that if $|\mathcal{C}_{k,i}| = |\mathcal{C}_{k,i+1}|$, then $\rho_{k,i} = \rho_{k,i+1}$. Suppose $\rho_{k,i} > \rho_{k,i+1}$. We could then swap $\rho_{k,i}$ and $\rho_{k,i+1}$ without increasing the objective value [the BER constraint (38) would not be affected, and the MSE constraint (37) would be relaxed, which could lead to an improvement of the objective value]. Then, we would have two solutions, which cannot be true from result 1). ■

REFERENCES

[1] M. L. Honig, K. Steiglitz, and B. Gopinath, "Multichannel signal processing for data communications in the presence of crosstalk," *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 551–558, Apr. 1990.

[2] 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, no. 2, pp. 41–59, Autumn 1996.

[3] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.

[4] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999.

[5] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.

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

[7] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1019–1032, Apr. 1999.

[8] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. COM-24, no. 12, pp. 1283–1290, Dec. 1976.

[9] J. Salz, "Digital transmission over cross-coupled linear channels," *AT&T Tech. J.*, vol. 64, no. 6, pp. 1147–1159, Jul.–Aug. 1985.

[10] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[11] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. Part I: unification and optimal designs," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1988–2006, Jul. 1999.

[12] 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, no. 12, pp. 2198–2206, Dec. 2001.

[13] J. Yang and S. Roy, "Joint transmitter-receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1334–1347, Sep. 1994.

[14] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[15] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[16] E. N. Onggosanusi, A. M. Sayeed, and B. D. V. Veen, "Efficient signaling schemes for wideband space-time wireless channels using channel state information," *IEEE Trans. Veh. Technol.*, vol. 52, no. 1, pp. 1–13, Jan. 2003.

[17] Y. Ding, T. N. Davidson, Z.-Q. Luo, and K. M. Wong, "Minimum BER block precoders for zero-forcing equalization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2410–2423, Sep. 2003.

[18] Y. Ding, T. N. Davidson, and K. M. Wong, "On improving the BER performance of rate-adaptive block-by-block transceivers with applications to DMT," in *Proc. IEEE Global Commun. Conf.*, San Francisco, CA, Dec. 1–5, 2003.

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

[20] L. S. Lasdon, *Optimization Theory for Large Systems*. New York: Macmillian, 1970.

[21] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. Berlin, Germany: Springer-Verlag, 1985.

[22] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[23] D. P. Palomar, "A unified framework for communications through MIMO channels," Ph.D. dissertation, Technical Univ. of Catalonia (UPC), Barcelona, Spain, May 2003.

[24] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding—parts I-II: equalization results and coding results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582–2604, Oct. 1995.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[26] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[27] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[28] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, Jul. 2004.

[29] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[30] S. Verdú, *Multiuser Detection*. New York: Cambridge Univ. Press, 1998.

[31] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. New York: Kluwer, 1999.

[32] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002.

[33] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 858–871, May 1997.

[34] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1561–1571, Sep. 2001.

[35] F. Rey, M. Lamarca, and G. Vázquez, "Transmit filter optimization based on partial CSI knowledge for wireless applications," in *Proc. IEEE Int. Conf. Commun.*, vol. 4, Anchorage, AK, May 11–15, 2003, pp. 2567–2571.

[36] P. Viswanath and V. Anantharam, "Optimal sequences and sum capacity of synchronous CDMA systems," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1984–1991, Sep. 1999.

[37] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.

[38] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[39] ETSI, Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) Layer, pp. 1–41, Feb. 2001.

[40] ETSI, Channel Models for HIPERLAN/2 in Different Indoor Scenarios, pp. 1–8, Mar. 1998.

[41] MIMO Channel Characterization, L. Schumacher, J. P. Kermoal, F. Frederiksen, K. I. Pedersen, A. Algans, and P. E. Mogensen. (2001, Feb.). [Online]. Available: http://www.ist-imetra.org/metra/

Daniel Pérez Palomar (S'99–M'03) was born in Barcelona, Spain, in 1975. He received the Electrical Engineering and the Ph.D. degrees (both with honors) from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

During 1998, he was with the Department of Electronic Engineering, King's College London (KCL), London, U.K. From January 1999 to December 2003, he was a Research Assistant with the Department of Signal Theory and Communications at UPC. From April to November 2001, he held a visiting research appointment at the Department of Electrical Engineering, Stanford University, Stanford, CA. From January to December 2002, he was a Visiting Researcher at the Telecommunications Technological Center of Catalonia (CTTC), Barcelona. From August to November 2003, he was a Guest Researcher at the Department of Signals, Sensors, and Systems, Royal Institute of Technology (KTH), Stockholm, Sweden. From November 2003 to February 2004, he was a Visiting Researcher at the INFOCOM Department, University of Rome "La Sapienza," Rome, Italy. Since March 2004, he has been a Fulbright Research Fellow at Princeton University, Princeton, NJ. His primary research interests include information-theoretic and communication aspects of wireless MIMO channels and array signal processing, with special emphasis on convex optimization theory applied to communication systems.

Dr. Palomar received the 2002/03 Rosina Ribalta first prize for the Best Doctoral Thesis within the areas of information technologies and communications by the Epson Foundation. He also received the 2003 prize for the Best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT. He has been awarded with a Fulbright Research Fellowship.

Mats Bengtsson (S'96–M'00) received the M.S. degree in computer science from Linköping University, Linköping, Sweden, in 1991 and the Tech. Lic. and Ph.D. degrees in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1997 and 2000, respectively.

From 1991 to 1995, he was with Ericsson Telecom AB, Karlstad, Sweden. He currently is a Research Associate at the Royal Institute of Technology. His research interests include statistical signal processing and its applications to antenna array processing and communications, radio resource management, and propagation channel modeling.

Björn Ottersten (S'87–M'89–SM'99–F'04) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986. In 1989, he received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

He has held research positions at the Department of Electrical Engineering at Linköping University, the Information Systems Laboratory at Stanford University, and the Katholieke Universiteit Leuven, Leuven, Sweden. From 1996 to 1997, he was the Director of Research at ArrayComm Inc, San Jose, CA. In 1991, he was appointed Professor of signal processing at the Royal Institute of Technology (KTH), Stockholm, Sweden, and he is currently head of the Department of Signals, Sensors, and Systems at KTH. He is also a Visiting Professor at the University of Luxembourg. His research interests include wireless communications, stochastic signal processing, sensor array processing, and time series analysis.

Dr. Ottersten has authored papers that received the Signal Processing Society Paper Award in 1993 and 2001. He has served as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and is a member of the editorial board of *EURASIP Journal of Applied Signal Processing*. He is currently Coeditor in Chief of *EURASIP Signal Processing Journal* and a member of the editorial board of IEEE SIGNAL PROCESSING MAGAZINE.