# MIMO Gaussian Channels with Arbitrary Inputs: Optimal Precoding and Power Allocation

Fernando Pérez-Cruz, Miguel R. D. Rodrigues, and Sergio Verdú

**Abstract**

We investigate the linear precoding and power allocation policies that maximize the mutual information for general multiple-input multiple-output (MIMO) Gaussian channels with arbitrary input distributions, by capitalizing on the relationship between mutual information and minimum mean-square error. The optimal linear precoder satisfies a fixed-point equation as a function of the channel and the input constellation. For nonGaussian inputs, a nondiagonal precoding matrix in general increases the information transmission rate, even for parallel noninteracting channels. Whenever precoding is precluded, the optimal power allocation policy also satisfies a fixed-point equation; we put forth a generalization of the mercury/waterfilling algorithm, previously proposed for parallel noninterfering channels, in which the mercury level accounts not only for the nonGaussian input distributions, but also for the interference among inputs.

**Index Terms:** Optimum Power Allocation, Multiple-Input Multiple-Output Systems, Mutual Information, MMSE, Precoding, Waterfilling, Gaussian Noise Channels.

F. Pérez-Cruz and S. Verdú are with the Electrical Engineering Department in Princeton University, Princeton (NJ). F. Pérez-Cruz is also with Universidad Carlos III de Madrid (Spain). E-mail: {fp,verdu}@princeton.edu

M. R. D. Rodrigues is with Instituto de Telecomunicações and the Department of Computer Science in Porto University (Portugal) and belongs to the Instituto de Telecomunicações. E-mail: mrodrigues@dcc.fc.up.pt

This work was partially supported by NSF Grants NCR-0074277 and CCR-0312879 and collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. F. Pérez-Cruz is supported by Marie Curie Fellowship 040883-AI-COM. M. R. D. Rodrigues is supported by Fundação Luso-Americana and Instituto de Telecomunicações.

Portions of this paper were presented at the 2007 Allerton Conference [13] and 2008 IEEE ICC [12].

## I. Introduction

Linear precoding for mutually interfering channels that maximizes mutual information yields the maximal achievable rates in conjunction with reliable error control codes. For parallel noninteracting channels corrupted with independent Gaussian noise, the mutual information is maximized, under a total power constraint, by independent Gaussian inputs with power selected according to the waterfilling allocation. Maximization of mutual information has also been considered for general multiple-input multiple-output (MIMO) channels, where the channels are mutually interfering and the noise covariance is arbitrary. Mutual information is maximized by imposing a covariance structure on the Gaussian input vector that satisfies two conditions. First, transmitting along the right eigenvectors of the channel, the vector channel reduces to a set of parallel noninterfering subchannels. Second, the power allocation follows the waterfilling policy based on the signal-to-noise ratio in each noninteracting subchannel.

Practical constraints dictate the use of discrete constellations, such as PSK and QAM, which depart significantly from the optimal Gaussian distributions. Thus, it is of interest to revisit the linear precoding problem under the constraint that each input follows a specified discrete constellation. In [8] the optimal power allocation for parallel noninteracting channels with arbitrary inputs is obtained exploiting the relation between the mutual information and the minimum mean-square error (MMSE) found in [5]. The results in [8] demonstrate that capacity-achieving strategies for Gaussian inputs may be quite suboptimal for discrete input constellations. One might expect that since the channels are noninterfering, to maximize mutual information, it is sufficient to take into account the nonGaussianness of the input constellations by means of the mercury/waterfilling power allocation. However, linear precoding techniques that introduce correlation among the channel inputs may achieve higher information transmission rates.

The optimization of a linear precoder and equalizer for MIMO channels has been typically addressed using an MMSE criterion [7], [15], [19], [18], with the optimal solution being the diagonalization of the channel matrix. In [10], a unifying approach using different criteria, such as the MMSE, the signal to interference-plus-noise ratio (SINR) and the bit error rate (BER), leads to identical results. In [3] nonlinear pre-equalization based on Tomlinson-Harashima-Miyakawa precoding is considered, together with a trellis signal shaping algorithm; the nonlinear precoder and the linear equalizer are adjusted following the zero-forcing criterion.

In this paper, we propose to optimize a linear precoder to maximize the input-output mutual information for general (not necessarily diagonal) rectangular deterministic channel matrices. A necessary condition for the solution is expressed as a fixed-point equation in terms of the relation between the MMSE and

mutual information [5], [11]. In contrast to the situation for Gaussian inputs [20], the optimal linear precoder for arbitrary inputs does not diagonalize the channel and, in particular, it does not reduce to a diagonal matrix for parallel noninterfering channels.

In those cases in which no linear processing of the inputs is allowed, it is necessary to restrict the precoder to be diagonal (a power allocation matrix). The optimum power allocation matrix satisfies a fixed-point equation and we provide a graphical interpretation for the optimum solution. In the mercury/waterfilling algorithm in [8], the base level is modified by the addition of mercury to account for the nonGaussian input constellation. In our generalization, the mercury level accounts for two factors: the nonGaussianness of the input distribution and the interference (caused by the other inputs).

The rest of the paper is organized as follows. We derive the optimal precoding policy in Section II-A. We study the low and high signal to noise ratio ($\mathsf{snr}$) regimes, respectively, in Sections II-B and II-C. We illustrate the reported results with representative examples in Section II-D. We derive the optimal power allocation policy in Section III-A. Section III-B particularizes the solution for Gaussian inputs and provides a graphical interpretation in Section III-C. We devote Section III-D to optimum power allocation in the low and high $\mathsf{snr}$ regimes and show some illustrative examples in Section III-E.

## II. Optimum Precoder

### A. General result

Consider the deterministic complex-valued vector channel[1]:

$$\mathbf{y} \;=\; \sqrt{\mathsf{snr}}\,\mathbf{H}\,\mathbf{P}\,\mathbf{x} \;+\; \mathbf{w} \tag{1}$$

where the $n$-dimensional vector $\mathbf{y}$ and the $m$-dimensional vector $\mathbf{x}$ represent, respectively, the received vector and the independent zero-mean unit-variance transmitted information vector. The distributions of the components of $\mathbf{x}$ are fixed, not necessarily Gaussian nor identical. The $n \times m$ complex-valued matrix $\mathbf{H}$ corresponds to the deterministic channel gains (known to both encoder and decoder) and $\mathbf{w}$

[1] *Notation*: Boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors, and italics denote scalars. The superscript $(\cdot)^\top$ and $(\cdot)^\dagger$ denote transpose and hermitian operations. $(\cdot)^*$ denotes optimum, $\mathrm{Tr}\{\cdot\}$ denotes the trace of a matrix and $\Re\{\cdot\}$ denotes real part of a complex number. $(\mathbf{X})_{ij}$ and $x_{ij}$ denotes the $i^{th}$ element of the $j^{th}$ column of the matrix $\mathbf{X}$, $\mathbf{x}_j$ denotes the $j^{th}$ column of $\mathbf{X}$, and $x_i$ denotes the $i^{th}$ element of $\mathbf{x}$. The Frobenius norm of a matrix is denoted by $\|\mathbf{X}\| = \sqrt{\mathrm{Tr}\{\mathbf{X}\mathbf{X}^\dagger\}}$, which reduces to the $\mathrm{L}^2$-norm $\|\mathbf{x}\|$ in the special case of a vector. The Kronecker's delta function is $\delta_{ij}$ and $\mathbf{1}$ is a column-vector of ones with the appropriate dimensions.

is the $n$-dimensional complex Gaussian noise with independent zero-mean unit-variance components[2]. The optimization of the mutual information is carried out over all $m \times m$ precoding matrices $\mathbf{P}$ that do not increase the transmitted power. The precoding problem can be cast as a constrained nonlinear optimization problem:

$$\max_{\mathbf{P}} I(\mathbf{x};\mathbf{y}) \tag{2}$$

subject to:

$$\mathsf{Tr}\left\{\mathbb{E}[\mathbf{P}\mathbf{x}\mathbf{x}^\dagger\mathbf{P}^\dagger]\right\} = \mathsf{Tr}\{\mathbf{P}\mathbf{P}^\dagger\} \leq 1 \tag{3}$$

*Theorem 1:* Let the input distribution be $P_{\mathbf{x}}$. The optimum precoding matrix $\mathbf{P}^*$ that solves (2) subject to (3) satisfies:

$$\mathbf{P}^* = \nu^{-1}\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E} \tag{4}$$

with

$$\nu = \|\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}\| \tag{5}$$

and

$$\mathbf{E} = \mathbb{E}\left[(\mathbf{x}-\mathbb{E}[\mathbf{x}|\mathbf{y}])(\mathbf{x}-\mathbb{E}[\mathbf{x}|\mathbf{y}])^\dagger\right] \tag{6}$$

$$= \mathbb{E}\left[\mathbf{x}\mathbf{x}^\dagger\right] - \mathbb{E}\left[\mathbb{E}[\mathbf{x}|\mathbf{y}]\,\mathbb{E}[\mathbf{x}|\mathbf{y}]^\dagger\right] \tag{7}$$

is known as the MMSE matrix [6], where

$$\mathbb{E}[\mathbf{x}|\mathbf{y}] = \frac{\sum_{\mathbf{x}} \mathbf{x} p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) P_{\mathbf{x}}(\mathbf{x})}{\sum_{\mathbf{x}'} p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}') P_{\mathbf{x}}(\mathbf{x}')} \tag{8}$$

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^n} e^{-\left\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{H}\mathbf{P}\mathbf{x}\right\|^2} \tag{9}$$

*Proof:* Appendix A. ■

The proof of Theorem 1 relies on the KKT conditions [4] and the relation between the mutual information and the MMSE [5], [11]. The KKT conditions are satisfied by any critical point (minimum, maximum or saddle point). There is a unique $\mathbf{P}^*$ that satisfies the KKT conditions when the problem is strictly concave, corresponding to the global maximum. In general, the power allocation in (2) and (3) is nonconcave. It becomes concave in some specific cases, e.g. for Gaussian input distributions with

[2]It is also possible to consider nonidentity noise covariance matrices, since we could add a whitening filter at the receiver that only affects the definition of the channel matrix.

arbitrary channel matrices [2]. Thus, it is important to realize that Theorem 1 gives a necessary condition for the optimum precoder, but in general, it does not identify uniquely the optimum precoding matrix.

For discrete constellations, the computation of the MMSE matrix (6) entails averaging over all possible transmitted vectors. Since the number of transmitted vectors grows exponentially with the input dimension $m$, for large dimensions we typically estimate the MMSE matrix $\mathbf{E}$ using Monte Carlo methods. A way to obtain an approximation of the solution to (4) is the iteration

$$\mathbf{P}_{k+1} = \alpha_k \left( \mathbf{P}_k + \mu \mathbf{H}^\dagger \mathbf{H} \mathbf{P}_k \mathbf{E}_k \right) \tag{10}$$

where $\mu$ is a small constant, the MMSE matrix $\mathbf{E}_k$ depends on $\mathbf{P}_k$ through (6)-(9), and $\alpha_k$ ensures that the norm of (10) is normalized to 1. This procedure ascents in the direction of the gradient and is fairly robust to noise in the estimation of $\mathbf{E}_k$. A similar approach was used in [11] to illustrate the gradient of the mutual information with respect to $\mathbf{P}$. To lessen the chance of getting trapped in local minima or nonglobal maxima, we may run (10) with several different $\mathbf{P}_0$, and keep the stationary point precoding matrix offering the largest mutual information.

### *B. Low* snr*-regime*

We now consider the optimal precoding policy for MIMO Gaussian channels with arbitrary input distributions in the regime of low-snr. We consider the low-snr expansion to the MMSE, and, by virtue of the relationship between mutual information and MMSE, the low-snr expansion to the mutual information. The form of the optimal precoder follows from the low-snr expansions.

We restrict the analysis to the case of zero-mean uncorrelated proper complex inputs, where $\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger] = \mathbf{I}$ and $\mathbb{E}[\mathbf{x}\mathbf{x}^T] = \mathbf{0}$. We note that as $\mathsf{snr} \to 0$ the conditional probability density function $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$ can be expressed as:

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^n} e^{-||\mathbf{y} - \sqrt{\mathsf{snr}}\mathbf{H}\mathbf{P}\mathbf{x}||^2} \tag{11}$$

$$= \frac{1}{\pi^n} e^{-||\mathbf{y}||^2} e^{2\sqrt{\mathsf{snr}}\Re\{y^\dagger \mathbf{H}\mathbf{P}\mathbf{x}\} - \mathsf{snr}||\mathbf{H}\mathbf{P}\mathbf{x}||^2} \tag{12}$$

$$= \frac{1}{\pi^n} e^{-||\mathbf{y}||^2} \left( 1 + 2\sqrt{\mathsf{snr}}\Re\{y^\dagger \mathbf{H}\mathbf{P}\mathbf{x}\} + \mathcal{O}(\mathsf{snr}) \right) \tag{13}$$

Using this expansion, it is possible to express the low-snr expansion of the MMSE matrix (6) as

$$\mathbf{E} = \mathbf{I} - (\mathbf{H}\mathbf{P})^\dagger (\mathbf{H}\mathbf{P}) \cdot \mathsf{snr} + \mathcal{O}(\mathsf{snr}^2) \tag{14}$$

Consequently, the low-snr expansion of the MMSE is given by;

$$\begin{aligned}\mathsf{mmse}(\mathsf{snr}) &= \mathbb{E}\big[(\mathbf{HPx}-\mathbf{HP}\mathbb{E}[\mathbf{x}|\mathbf{y}])^\dagger(\mathbf{HPx}-\mathbf{HP}\mathbb{E}[\mathbf{x}|\mathbf{y}])\big] = \mathsf{tr}\Big\{\mathbf{HPE}(\mathbf{HP})^\dagger\Big\} \\ &= \mathsf{tr}\big\{(\mathbf{HP})(\mathbf{HP})^\dagger\big\} - \mathsf{tr}\Big\{\Big((\mathbf{HP})(\mathbf{HP})^\dagger\Big)^2\Big\}\cdot\mathsf{snr} + \mathcal{O}\big(\mathsf{snr}^2\big) \end{aligned} \tag{15}$$

whereas the low-snr expansion of the mutual information is given by:

$$\mathsf{I}(\mathsf{snr}) = \mathsf{tr}\big\{(\mathbf{HP})(\mathbf{HP})^\dagger\big\}\cdot\mathsf{snr} - \mathsf{tr}\Big\{\Big((\mathbf{HP})(\mathbf{HP})^\dagger\Big)^2\Big\}\cdot\frac{\mathsf{snr}^2}{2} + \mathcal{O}\big(\mathsf{snr}^3\big) \tag{16}$$

Note that it is very important to retain the first two terms in the low-snr expansion of the MMSE, or, equivalently, the first- and the second-order terms in the low-snr expansion of the mutual information. In fact, $\ddot{\mathsf{I}}(0) = \frac{d^2\mathsf{I}(\mathsf{snr})}{d\mathsf{snr}^2}\big|_{\mathsf{snr}=0}$ is a key low-power performance measure since the bandwidth required to sustain a given rate with a given (low) power is proportional to $-\ddot{\mathsf{I}}(0)$ [17].

According to (16), for first-order optimality, $\mathbf{PP}^\dagger$ ought to maximize (subject to $\mathsf{tr}(\mathbf{PP}^\dagger)=1$):

$$\mathsf{tr}\big\{(\mathbf{HP})(\mathbf{HP})^\dagger\big\} = \mathsf{tr}\big\{\mathbf{\Lambda_H}\mathbf{V_H}^\dagger\mathbf{PP}^\dagger\mathbf{V_H}\mathbf{\Lambda_H}\big\} \tag{17}$$

where we have used the singular value decomposition of the channel matrix $\mathbf{H} = \mathbf{U_H}\mathbf{\Lambda_H}\mathbf{V_H}^\dagger$, with $\mathbf{U_H}$ and $\mathbf{V_H}$ unitary $n\times n$ and $m\times m$ matrices respectively and $\mathbf{\Lambda_H}$ is a diagonal matrix with $(\mathbf{\Lambda_H})_{ii}$ equal to the $i^{th}$ singular value of $\mathbf{H}$. Therefore the non-unique first-order optimal $\mathbf{P}$ is equal to $\mathbf{V_H}$ times a diagonal matrix whose diagonal coefficients are zero except at coefficients corresponding to the (generally nonunique) maximum singular value of $\mathbf{H}$. In addition, optimizing (16) (see also [17]) second-order optimality requires that there are as many nonzero coefficients as the multiplicity of the largest singular value, and that they are all equal.

It is instructive to observe that the optimal precoder performs two operations. First, it aligns the transmit directions with the right singular vectors of the channel. This decomposes the vector channel into a set of independent scalar subchannels. Second, it assigns power on an equal basis to the strongest scalar subchannels. As expected in the regime of low-snr the optimal precoding policy for proper complex inputs is entirely equivalent to the optimal one for complex Gaussian inputs [17].

### *C. High* snr*-regime for discrete inputs*

We now consider the optimal precoding policy for MIMO Gaussian channels with arbitrary discrete input distributions in the high snr-regime. We prove upper and lower bounds to the MMSE, and consequently, upper and lower bounds to the mutual information by exploiting the relationship between mutual information and MMSE [5], [11]. Then we use the upper and lower bounds to derive the form of the

optimal policy. These results rest upon [8, Theorem 3], where the MMSE for single-input single-output Gaussian channels with QPSK input distribution is expanded for large $\mathsf{snr}$. We generalize the bound in [8, Theorem 4], to general (nondiagonal, rectangular) channel matrices.

Consider a single-input single-output channel model

$$y \;=\; \sqrt{\mathsf{snr}}\, x \;+\; w \tag{18}$$

where the input $x \in \{\pm 1\}$ is an equiprobable random variable and the noise $w$ is a complex random variable Gaussian with zero-mean and unit-variance. The MMSE obeys the integral equation:

$$\mathsf{mmse}(\mathsf{snr}) = \mathbb{E}\left[|x - \mathbb{E}[x|y]|^2\right] \tag{19}$$

$$= 1 - \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \tanh\left(2\sqrt{\mathsf{snr}}\xi\right) e^{-\left(\xi - \sqrt{\mathsf{snr}}\right)^2} d\xi, \tag{20}$$

*Theorem 2 (Large Series Expansion for binary phase shift keying (BPSK) from [8]):* The $\mathsf{mmse}(\mathsf{snr})$ in (20) admits the large-$\mathsf{snr}$ expansion:

$$\mathsf{mmse}(\mathsf{snr}) = \frac{e^{-\mathsf{snr}}}{2\sqrt{\mathsf{snr}}} \left(\sqrt{\pi} + \sum_{i=1}^{\infty} \frac{\zeta_i}{(4\mathsf{snr})^i}\right) \tag{21}$$

where

$$\zeta_i = \frac{(-1)^i}{\sqrt{\pi} 8^i} \left(\sum_{k=0}^{\infty} (k + 1/4)^{-2i-1} - (k + 3/4)^{-2i-1}\right) \prod_{q=1}^{i} (2q - 1) \tag{22}$$

*Corollary 1:* Consider the channel model:

$$\mathbf{y} \;=\; \sqrt{\mathsf{snr}}\, \mathbf{h}\, x \;+\; \mathbf{w} \tag{23}$$

where the input $x \in \{\pm 1\}$ is an equiprobable random variable and $\mathbf{w}$ is the complex Gaussian noise with independent zero-mean and unit-variance components. The $\mathsf{mmse}(\mathsf{snr})$ admits the representation for those $\mathsf{snr}$ large enough for the series to converge

$$\mathsf{mmse}(\mathsf{snr}) = \mathbb{E}\left[||\mathbf{h}(x - \mathbb{E}[x|\mathbf{y}])||^2\right] \tag{24}$$

$$= ||\mathbf{h}||^2 \mathbb{E}\left[(x - \mathbb{E}[x|\mathbf{y}])\right] \tag{25}$$

$$= ||\mathbf{h}||^2 \frac{e^{-||\mathbf{h}||^2 \mathsf{snr}}}{2||\mathbf{h}||\sqrt{\mathsf{snr}}} \left(\sqrt{\pi} + \sum_{i=1}^{\infty} \frac{\zeta_i}{(4||\mathbf{h}||^2 \mathsf{snr})^i}\right) \tag{26}$$

with $\zeta_i$ as defined in (22).

*Proof:* Replace $\sqrt{\mathsf{snr}}$ by $\sqrt{\mathsf{snr}}\mathbf{h}$ in (21). ■

The proof of Theorem 2 is based on the expansion of $Q(\cdot)$ in [16, (3.53)]. Typically only one or two terms are considered for this expansion. Furthermore, the first two terms of the large series expansion of

$Q(\cdot)$ provide, respectively, an upper and lower bound for this function [14]. Therefore we can construct an upper and lower bound for the MMSE using the first two terms of its expansion in (26).

*Corollary 2:* We can bound the MMSE for the channel model in (23) by:

$$||\mathbf{h}||^2 \frac{e^{-||\mathbf{h}||^2 \mathsf{snr}}}{2||\mathbf{h}||\sqrt{\mathsf{snr}}} \left( \sqrt{\pi} - \frac{4.37}{4||\mathbf{h}||^2 \mathsf{snr}} \right) \leq \mathsf{mmse}(\mathsf{snr}) \tag{27}$$

$$\leq ||\mathbf{h}||^2 \frac{e^{-||\mathbf{h}||^2 \mathsf{snr}}}{2||\mathbf{h}||\sqrt{\mathsf{snr}}} \sqrt{\pi} \tag{28}$$

*Proof:* Appendix B. ■

From Theorem 2 and its corollaries, we see that the large-series expansion of the MMSE depends on the distance between the two mean received points: $2\sqrt{\mathsf{snr}}||\mathbf{h}||$. This property is generalized in the following theorem, in which we upper and lower bound the MMSE for the MIMO channel model in (1).

*Theorem 3:* For the channel model in (1), assuming equiprobable signaling for each input:

$$\frac{d_{min}^2}{2M} \frac{e^{-d_{min}^2 \mathsf{snr}/4}}{d_{min}\sqrt{\mathsf{snr}}} \left( \sqrt{\pi} - \frac{4.37}{d_{min}^2 \mathsf{snr}} \right) \leq \mathsf{mmse}(\mathsf{snr}) \tag{29}$$

$$\leq (M-1) d_{max}^2 \frac{e^{-d_{min}^2 \mathsf{snr}/4}}{d_{min}\sqrt{\mathsf{snr}\pi}} \tag{30}$$

where

$$\mathsf{mmse}(\mathsf{snr}) = \mathbb{E}\left[ ||\mathbf{HP}(\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathbf{y}])||^2 \right] \tag{31}$$

and $M$ is the product of the constellations cardinality, $d_{min}$ is the minimum distance between the $M$ possible realizations of the input vector (which we refer to as *constellation vectors*):

$$d_{min} = \min_{\substack{\overline{\mathbf{x}},\mathbf{x} \\ \mathbf{x} \neq \overline{\mathbf{x}}}} ||\mathbf{HP}(\mathbf{x} - \overline{\mathbf{x}})|| \tag{32}$$

and $d_{max}$ is the maximum distance:

$$d_{max} = \max_{\overline{\mathbf{x}},\mathbf{x}} ||\mathbf{HP}(\mathbf{x} - \overline{\mathbf{x}})|| \tag{33}$$

*Proof:* Appendix C. ■

We can use (29) and (30) and the relation between the MMSE and the mutual information to bound the mutual information for large $\mathsf{snr}$ and discrete input distributions.

*Theorem 4:* The mutual information between $\mathbf{x}$ and $\mathbf{y}$ for the channel model in (1) can be bounded by:

$$\log M - \frac{2(M-1)d_{max}^2}{d_{min}^3 \sqrt{\mathsf{snr}\pi}} e^{-d_{min}^2 \mathsf{snr}/4} \leq I(\mathbf{x};\mathbf{y}) \tag{34}$$

$$\leq \log M - d_{min}^2 \frac{e^{-d_{min}^2 \mathsf{snr}/4}}{M d_{min}^3 \sqrt{\mathsf{snr}}} \left( \sqrt{\pi} - \frac{4.37 + 2\sqrt{\pi}}{d_{min}^2 \mathsf{snr}} \right) \tag{35}$$

*Proof:* Appendix D. ■

These results show that for a MIMO channel model with discrete inputs, the precoding matrix $\mathbf{P}$ that maximizes the information transmission rate converges as the $\mathsf{snr} \to \infty$ to the matrix that maximizes the minimum distance between the constellation vectors. This matrix also ensures the asymptotic minimization of the MMSE and the SER. This result is a direct generalization of a result in [8] for noninterfering channels. This is an appealing result as it tells us that for discrete inputs in high $\mathsf{snr}$ regime, minimizing the symbol error rate is equivalent to both minimizing the mean-square error and maximizing the mutual information, thereby linking three standard criteria typically used for power allocation in the all-important case of discrete input distributions.

We can illustrate with a simple example why linear precoders achieve larger information transmission rates than diagonal power allocation matrices even for diagonal channel matrices as the $\mathsf{snr}$ tends to infinity. Let us assume a simple real-valued communication system with a noninterfering channel matrix

$$\mathbf{H} = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}, \tag{36}$$

in which both inputs are BPSK distributions. We first set $\mathbf{P} = \mathbf{I}$ and the minimum distance in the received constellation is 2. The optimal power allocation matrix corresponding to the mercury/waterfilling [8] solution is:

$$\mathbf{P}_0 = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & \sqrt{3/2} \end{bmatrix} \tag{37}$$

and the minimum distance increases to $\sqrt{6}$. We can further compute the optimal linear precoder, using the procedure described at the end Section II-A:

$$\mathbf{P}^* = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}, \tag{38}$$

and obtain the largest minimum distance of $\sqrt{8}$. In Figure 1 we have plotted the constellations for the optimal precoder and power allocation matrices to highlight the difference between the received constellations. The difference between the transmission rates for precoding and power allocation strategies are quite substantial, as illustrated in the applications section.

### *D. Applications*

In this section we show the benefit of using a full precoding matrix instead of a diagonal power allocation matrix for maximizing the mutual information in a MIMO communication channel with Gaussian noise when the inputs are QPSK.

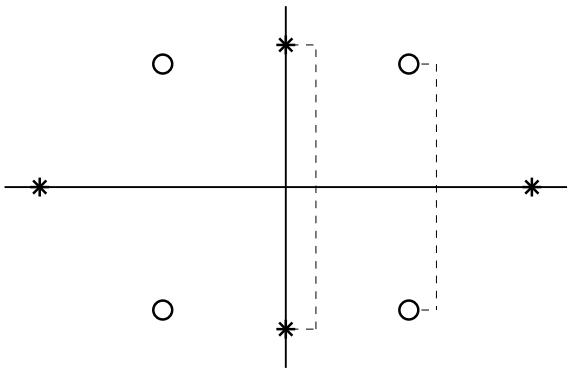


Fig. 1. '$\circ$' represents the received constellation for the diagonal power allocation solution $\mathbf{P}_0$ and '$*$' represents the output constellation achieved with the optimum linear precoder $\mathbf{P}^*$. The dashed lines represent the minimum distance in each constellation.

In the first example we deal with a $5 \times 5$ diagonal channel matrix and compare the information transmission rate obtained using the optimal mercury/waterfilling power allocation proposed in [8], corresponding to a diagonal real-valued matrix $\mathbf{P}$, to the optimal nondiagonal precoding matrix.

In the second example we use a full $4 \times 4$ channel matrix typically encountered in each subchannel of DSL systems [1]. However, rather than performing diagonalization of the channel matrix followed by waterfilling power allocation over the subchannels, as proposed in [1], we directly optimize the precoding matrix [12], significantly increasing the input-output mutual information.

*1) Diagonal channel matrix:* We assume QPSK inputs and

$$\mathbf{H}=\begin{bmatrix} 2e^{-\mathrm{j}2.51} & 0 & 0 & 0 & 0 \\ 0 & 1.5e^{\mathrm{j}1.60} & 0 & 0 & 0 \\ 0 & 0 & e^{-\mathrm{j}0.79} & 0 & 0 \\ 0 & 0 & 0 & 0.8e^{-\mathrm{j}3.05} & 0 \\ 0 & 0 & 0 & 0 & 0.6e^{-\mathrm{j}1.67} \end{bmatrix}. \tag{39}$$

For this diagonal channel model, Figure 2 compares the performance of the optimal power-allocation policy corresponding to a diagonal real-valued $\mathbf{P}$ [8] (denoted in Figure 2 as MWF) and the optimal nondiagonal precoding matrix, together with the waterfilling solution for Gaussian inputs, which serves as an upper bound to the mutual information. It can be seen that the mutual information for the full precoding matrix is substantially higher than that for the diagonal power allocation matrix, yielding gains higher than 2dB. Figure 2 also shows the performance of the precoder that maximizes the minimum

distance and, thus, is asymptotically optimal for large snr, given by:

$$\mathbf{P}_{\max dmin} = \begin{bmatrix} 0.28e^{\mathrm{j}2.5} & 0.26e^{\mathrm{j}1.8} & 0.27e^{-\mathrm{j}0.1} & 0.16e^{\mathrm{j}1.5} & 0.23e^{\mathrm{j}1.8} \\ 0.09e^{\mathrm{j}1.9} & 0.21e^{\mathrm{j}0.9} & 0.16e^{-\mathrm{j}3.1} & 0.34e^{-\mathrm{j}2.5} & 0.25e^{\mathrm{j}2.7} \\ 0.26e^{\mathrm{j}0.7} & 0.09e^{-\mathrm{j}2.6} & 0.15e^{\mathrm{j}1.6} & 0.17e^{\mathrm{j}0.1} & 0.27e^{\mathrm{j}3.0} \\ 0.07e^{\mathrm{j}0.6} & 0.23e^{\mathrm{j}1.4} & 0.21e^{\mathrm{j}1.8} & 0.02e^{-\mathrm{j}0.1} & 0.17e^{-\mathrm{j}0.7} \\ 0.18e^{\mathrm{j}0.8} & 0.17e^{-\mathrm{j}2.9} & 0.15e^{-\mathrm{j}0.5} & 0.21e^{-\mathrm{j}1.0} & 0.12e^{-\mathrm{j}1.0} \end{bmatrix} \tag{40}$$

Note that the maximum minimum distance precoder achieves higher information transmission rates than the optimal diagonal power allocation for $\mathsf{snr} > 2.9\text{dB}$.

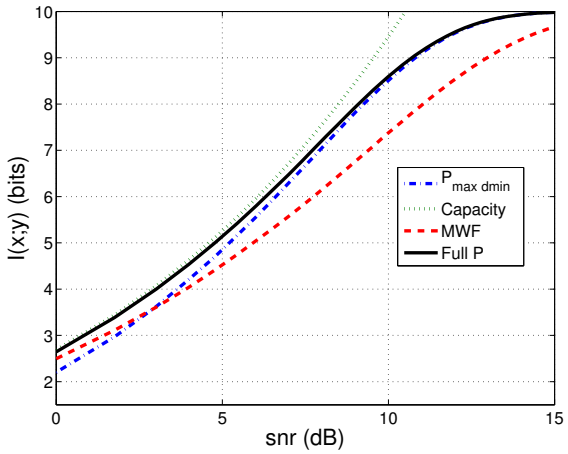



Fig. 2. Mutual Information rates for the $5 \times 5$ matrix in (39).

As the snr increases, the optimal diagonal power-allocation matrix tends to:

$$\mathbf{P}_1 = \begin{bmatrix} 0.20 & 0 & 0 & 0 & 0 \\ 0 & 0.27 & 0 & 0 & 0 \\ 0 & 0 & 0.41 & 0 & 0 \\ 0 & 0 & 0 & 0.51 & 0 \\ 0 & 0 & 0 & 0 & 0.68 \end{bmatrix} \tag{41}$$

while the full precoding matrix tends to (40).

For the diagonal power-allocation matrix, the assigned power is asymptotically inversely proportional to the norm of the subchannel. Hence asymptotically the mercury-waterfilling policy performs power equalization as noted in [8]. The nondiagonal precoding matrix $\mathbf{P}_{\max dmin}$ approximately assigns the same power for all inputs, as the norms of the columns of $\mathbf{P}_{\max dmin}$ are almost identical. However, more power is assigned to the stronger channels, as the norm of the rows of $\mathbf{P}_{\max dmin}$ decreases with the channel gain. Therefore, using a full precoding matrix, we are able significantly increase the mutual information for all but low snr.

*2) Full channel matrix, the DSL scenario:* In this second example, we illustrate for a nondiagonal channel matrix $\mathbf{H}$ that a full precoder provides substantially higher information transmission rates than standard channel diagonalization followed by mercury/waterfilling power allocation. We consider a situation encountered in Gigabit DSL systems with four copper-wire pairs typically available in the last distribution area [1]. In particular, we use the following channel matrix:

$$\mathbf{H} = \begin{bmatrix} e^{\mathrm{j}2.80} & 0.5e^{-\mathrm{j}0.97} & 0.01e^{-\mathrm{j}1.02} & 10^{-3}e^{-\mathrm{j}1.06} \\ 0.5e^{-\mathrm{j}0.43} & e^{-\mathrm{j}2.98} & 0.1e^{\mathrm{j}0.12} & 0.01e^{-\mathrm{j}3.12} \\ 0.01e^{\mathrm{j}2.93} & 0.1e^{\mathrm{j}1.27} & e^{-\mathrm{j}2.00} & 0.5e^{-\mathrm{j}1.83} \\ 10^{-3}e^{\mathrm{j}2.63} & 0.01e^{-\mathrm{j}2.06} & 0.5e^{\mathrm{j}2.38} & e^{\mathrm{j}2.69} \end{bmatrix} \tag{42}$$

in which the diagonal coefficients have unit gain.

In Figure 3, we show the mutual information for the optimal precoding as well as for the optimal power allocation policy using the mercury/waterfilling algorithm [8] performed after diagonalization of the channel matrix (denoted as MWF). As in the previous example, the full precoder significantly increases the information transmission rate with respect to the optimal power allocation strategy.

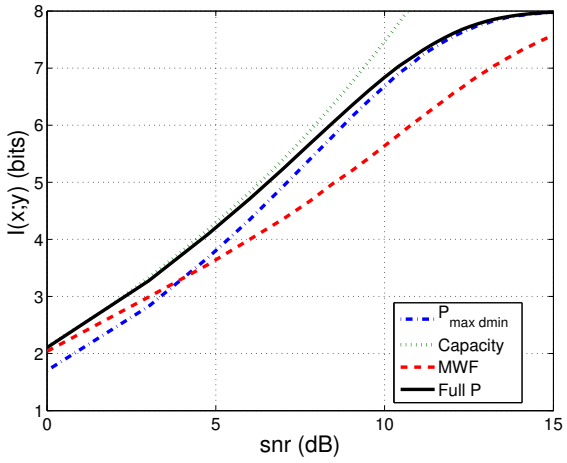



Fig. 3. Mutual Information rates for the $4 \times 4$ matrix in (42).

## III. Optimum Power Allocation

### A. General Result

The mercury/waterfilling algorithm in [8] sets the optimal power for parallel interference-free channels (diagonal $\mathbf{H}$ and $\mathbf{P}$ in (1)) with arbitrary inputs. It involves two phases: a mercury-pouring phase followed by a water-pouring one. The mercury-pouring phase amends the base level to account for the mismatch

between the actual input and the optimal Gaussian distribution, while the water-pouring phase represents the power allocation. Reference [8] capitalized on the relationship between the mutual information and the MMSE [5] to obtain this power-allocation policy, thereby circumventing the lack of mutual information expressions for arbitrary input distributions.

In this section, we determine the optimal power-allocation policy for MIMO interfering channels with arbitrary input distributions, thereby generalizing the mercury/waterfilling solution. We generalize the setting in [8] to any rectangular channel matrix $\mathbf{H}$, limiting ourselves to optimization over a diagonal matrix $\mathbf{P}$. As pointed out in [8], there are communications systems in which precoding policies are precluded, preventing the channel from being diagonalized (or subject to any of linear transformation at the encoder). In this section we extend the water-filling algorithm for MIMO interfering channels and, in this more general setup, we show that the mercury level accounts for both the mismatch between the actual input and an optimal Gaussian distribution, as well as, the interference between inputs [13].

The power-allocation problem can be cast as a linearly-constrained nonlinear optimization problem:

$$\max_{\mathbf{p}} I(\mathbf{x};\mathbf{y}) \tag{43}$$

subject to:

$$(\mathbf{P})_{ij} = \sqrt{p_j}\delta_{ij} \tag{44}$$

$$\sum_{j=1}^{m} p_j \leq 1 \tag{45}$$

$$p_j \geq 0 \qquad \forall j = 1,\ldots,m \tag{46}$$

where $\mathbf{p} = [p_1,\ldots,p_m]^\top$.

*Theorem 5:* The optimal power allocation $\mathbf{p}^*$ that solves (43) subject to (44)-(46) satisfies:

$$p_j^* = \gamma^{-1}\left(\mathbf{P}^*\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}\right)_{jj} \qquad \forall j = 1,\ldots,m \tag{47}$$

where

$$\gamma = \mathrm{Tr}\left\{\mathbf{P}^*\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}\right\} \tag{48}$$

*Proof:* Appendix E. ■

If $\mathbf{H}$ is diagonal, so is $\mathbf{E}$, and (47) can be simplified to:

$$p_j^* = \gamma^{-1}p_j^*|h_{jj}|^2\mathbb{E}\left[|x_j - \mathbb{E}[x_j|y_j]|^2\right] \tag{49}$$

$$= \gamma^{-1}p_j^*|h_{jj}|^2\mathsf{mmse}_j(\mathsf{snr}|h_{jj}|^2p_j^*) \tag{50}$$

Recalling that the mean-square error in (50) cannot be greater than 1, we can express the optimal power allocation as follows:

$$p_j^* = \begin{cases} \frac{1}{\mathsf{snr}|h_{jj}|^2} \mathsf{mmse}_j^{-1}\left(\frac{\gamma}{|h_{jj}|^2}\right), & \gamma \leq |h_{jj}|^2 \\ 0, & \gamma > |h_{jj}|^2 \end{cases} \tag{51}$$

As expected, the optimal power allocation in (51) reduces to Theorem 2 in [8] for parallel noninterfering channels.

The proof of Theorem 5, as that of Theorem 1, relies on the KKT conditions [4] and the relation between the mutual information and the MMSE [5], [11]. There is a unique $\mathbf{p}^*$ that satisfies the KKT conditions when the problem is strictly concave, corresponding to the global maximum. In general, the power allocation as stated in (43)-(46) is nonconcave. It becomes concave in some specific cases, e.g. for Gaussian input distributions [2] or diagonal channel matrices with arbitrary input distributions [8]; it is also concave for low $\mathsf{snr}$. Thus, as for Theorem 1, it is important to realize that Theorem 5 only gives a necessary condition for the optimum power allocation. To compute the optimal power-allocation matrix we proceed as we did in Section II-A to compute the optimal precoder.

Solution (47) does not provide much insight about how the power is allocated. We provide an interpretation for the optimal power-allocation policy, generalizing the mercury/waterfilling algorithm put forward in [8] for noninterfering channels.

*Definition 1*: The *water level* is $\beta = (\mathsf{snr}\gamma)^{-1}$, the inverse of the Lagrange multiplier associated with the inequality constraint in (45).

*Definition 2*: The *base level* associated with the $j^{th}$ input is given by

$$b_j = \frac{1}{\left.\frac{\partial I(\mathbf{x};\mathbf{y})}{\partial p_j}\right|_{\substack{p_j = 1 \\ x_j \sim \mathcal{N}(0,1)}}} - 1 \tag{52}$$

$$= \frac{1}{\mathsf{snr}||\mathbf{h}_j||^2} \tag{53}$$

where $\mathbf{h}_j$ is the $j^{th}$ column of $\mathbf{H}$.

These definitions correspond to the water and base levels for waterfilling [2] and mercury/waterfilling [8] for noninterfering channels. Note that the base level definition ensures that the difference between the water and the base levels is 1, if the $j^{th}$ input uses a Gaussian distribution and is assigned all the available power.

*Definition 3*: The *mercury level* associated with the $j^{th}$ input is given by

$$m_j = \frac{1}{\left.\frac{\partial I(\mathbf{x};\mathbf{y})}{\partial p_j}\right|_{\mathbf{P}=\mathbf{P}^*}} - p_j^* \tag{54}$$

For $p_j^* > 0$

$$\left.\frac{\partial I(\mathbf{x};\mathbf{y})}{\partial p_j}\right|_{\mathbf{P}=\mathbf{P}^*} = \frac{\mathsf{snr}}{\sqrt{p_j^*}}\left(\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}\right)_{jj} = \beta^{-1} \tag{55}$$

therefore

$$p_j^* = (\beta - m_j)^+ \tag{56}$$

where $(\cdot)^+ = \max(\cdot, 0)$.

If $p_j^* = 0$, we do not need to define the mercury level as the $j^{th}$ input is not assigned any power. However even in that case we choose to define a mercury level to ensure it is a continuous function of the $\mathsf{snr}$. For $p_j^* = 0$, the derivative of the mutual information is not given by (55), because it is an indeterminate form $(0/0)$. The value of the mercury level for $p_j^* = 0$ is computed in Appendix F and it is shown to be above the water and base levels.

### *B. Gaussian inputs*

The special case of Gaussian inputs is interesting in those problems where the suboptimality comes from constraints that prevent the input components from beeing correlated. In that case, we have the following results:

*Theorem 6:* For Gaussian independent inputs with any channel matrix $\mathbf{H}$:

$$\frac{\partial(m_j - b_j)}{\partial p_i} \geq 0, \tag{57}$$

and

$$\frac{\partial(m_j - b_j)}{\partial \mathsf{snr}} \leq 0. \tag{58}$$

*Proof:* Appendix G. ■

*Theorem 7:* For Gaussian independent inputs and if $\mathbf{H}\mathbf{H}^\dagger$ is invertible:

$$\lim_{\mathsf{snr}\to\infty} p_i = \frac{1}{m} \tag{59}$$

*Proof:* Appendix H. ■

Theorem 7 indicates that as the $\mathsf{snr}$ increases the power is equally shared between all the subchannels, as in the noninterfering case, as long as $\mathbf{H}\mathbf{H}^\dagger$ is full rank.

### C. Graphical interpretation

In the mercury/waterfilling interpretation for noninterfering channels in [8], the mercury level for the $j^{th}$ channel directly depends on $\mathsf{snr}|h_{jj}|^2$, through the (tabulated) inverse MMSE, and the water level[3]. The water level is numerically calculated ensuring that $\{p_j^*\}_{j=1}^m$ add up to 1. For nondiagonal $\mathbf{H}$ matrices, we solve the optimization problem in (43)-(46), jointly obtaining the power allocation (primal variables) and the water level (Lagrange multiplier).

Figure 4 depicts the generalized mercury/waterfilling reinterpretation for mutually interfering MIMO channels. The mercury level for $p_j^* = 0$ is the main difference between the mercury/waterfilling for mutually interfering MIMO channels and the one in [8] for noninterfering channels. In interference-free mercury waterfilling [8], mercury is never added to a base level that is above the water level. In contrast, for interfering channels the mercury level can be nonzero, even if the base level is above the water level, or if the base level is below the water level, but the mercury level is above the water level and thus, $p_j^* = 0$. This added mercury accounts for the interference suffered by the $j^{th}$ input.

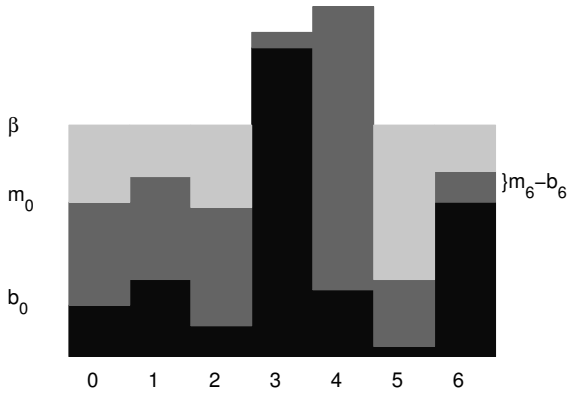



Fig. 4. Illustration of the mercury/waterfilling power allocation for a seven-input MIMO channel. The power for each input is the difference between the water and mercury level. The base, mercury and water are depicted in black, dark grey and light grey, respectively. The mercury in channels 3 and 4 is only due to the interferences caused by the other inputs, because their allocated power is zero.

In the MIMO setup with arbitrary input constellations, we can distinguish three types of mechanisms that curtail the reliable information rate achievable by an individual input:

a) the additive white Gaussian noise.

b) the nonGaussian nature of the input.

[3]The water level depends on every $\mathsf{snr}|h_{ii}|^2$, thus indirectly the mercury level depends on all channel responses.

c) the presence of interference from other inputs.

In the conventional waterfilling solution, b) and c) are absent, and the base level that accounts for the individual signal-to-noise ratio of that channel is sufficient to capture the difficulty to convey information through that input.

In the mercury/waterfilling solution of [8], c) is absent, and the mercury accounts for the suboptimality of the input. The mercury level depends on the available power, on the signal-to-noise ratios seen by the other inputs as well as on all the input distributions.

If a), b) and c) are present, the mercury level also accounts for the interference suffered from those inputs that receive positive power. It is instructive to consider first the special case when all the inputs are Gaussian, and thus only mechanisms a) and c) are present. In that case, the mercury level serves simply to augment the base level with the additive Gaussian noise received from the other inputs. Thus, as in [8], the mercury level depends on the total available power. Furthermore, if the power is sufficiently low that only one input is allocated power, there is no need to add mercury to that input. However, there is indeed mercury added to the other inputs, as their barrier for nonzero power allocation is now increased due to the presence of the interference from the sole input that has been assigned power. As we increase power, at some point the water level exceeds the mercury level of one or several inputs, and thus, a small nonzero mercury level is added to the original input that received all the power. As indicated by Theorem 6, the mercury level of every input is monotonically nondecreasing with the total available power.

Returning to the general case where a), b) and c) are present, the mercury level does not decompose in separate components due to b) and c) since the deleterious effect of an interfering input depends on its distribution. However, as the power increases, discrete-constellation interferers have an asymptotically vanishing effect on the available rate, and therefore in the high snr, the solution behaves as if c) were not present.

### *D. Low and high* snr*-regime*

We now consider the optimal power-allocation policy for MIMO Gaussian channels with arbitrary input distributions in the regime of low-snr. The study of the high-snr regime is identical to Section II-C, exchanging the precoder by a power allocation matrix.

The low-snr analysis of the optimal power allocation policy follows closely the low-snr analysis of the optimal precoding policy. However, it is important to note that the power allocation matrix is constrained to be a diagonal matrix rather than a general matrix.

For first-order optimality, the optimal diagonal power allocation matrix $\mathbf{PP}^\dagger$ has to maximize:

$$\begin{aligned}\mathsf{I}(\mathsf{snr}) &= \mathsf{tr}\big\{(\mathbf{HP})(\mathbf{HP})^\dagger\big\}\cdot\mathsf{snr}+\mathcal{O}\big(\mathsf{snr}^2\big)\\ &= \sum_i \|\mathbf{h}_i\|^2(\mathbf{PP}^\dagger)_{ii}\cdot\mathsf{snr}+\mathcal{O}\big(\mathsf{snr}^2\big)\\ &= \sum_i \|\mathbf{h}_i\|^2 p_i\cdot\mathsf{snr}+\mathcal{O}\big(\mathsf{snr}^2\big)\end{aligned} \tag{60}$$

subject to the power constraint $\mathsf{tr}(\mathbf{PP}^\dagger)=\sum_i p_i=1$, with $p_i\geq 0$, $\forall i$. This shows that power is allocated only to the inputs associated with the columns of the channel matrix $\mathbf{H}$ with the highest norm.

For both first- and second-order optimality, the optimal diagonal power allocation matrix $\mathbf{PP}^\dagger$ has to maximize:

$$\begin{aligned}\mathsf{I}(\mathsf{snr}) &= \mathsf{tr}\big\{(\mathbf{HP})(\mathbf{HP})^\dagger\big\}\cdot\mathsf{snr}-\mathsf{tr}\big\{(\mathbf{HP})(\mathbf{HP})^\dagger(\mathbf{HP})(\mathbf{HP})^\dagger\big\}\cdot\frac{\mathsf{snr}^2}{2}+\mathcal{O}\big(\mathsf{snr}^3\big)\\ &= \sum_i \|\mathbf{h}_i\|^2(\mathbf{PP}^\dagger)_{ii}\cdot\mathsf{snr}-\sum_i\sum_j \mathbf{h}_i^\dagger\mathbf{h}_j\mathbf{h}_j^\dagger\mathbf{h}_i(\mathbf{PP}^\dagger)_{ii}(\mathbf{PP}^\dagger)_{jj}\cdot\frac{\mathsf{snr}^2}{2}+\mathcal{O}\big(\mathsf{snr}^3\big)\\ &= \sum_i \|\mathbf{h}_i\|^2 p_i\cdot\mathsf{snr}-\sum_i\sum_j \mathbf{h}_i^\dagger\mathbf{h}_j\mathbf{h}_j^\dagger\mathbf{h}_i p_i p_j\cdot\frac{\mathsf{snr}^2}{2}+\mathcal{O}\big(\mathsf{snr}^3\big)\end{aligned} \tag{61}$$

subject to the power constraint $\mathsf{tr}(\mathbf{PP}^\dagger)=\sum_i p_i=1$, with $p_i\geq 0$, $\forall i$. This shows that if there is a unique column of the channel matrix with the highest norm, then power is allocated solely to the input associated with that column of the channel matrix. In contrast, if there are multiple columns of the channel matrix with the highest norm, then power is divided equally between the inputs associated with those columns of the channel matrix.

### *E. Examples*

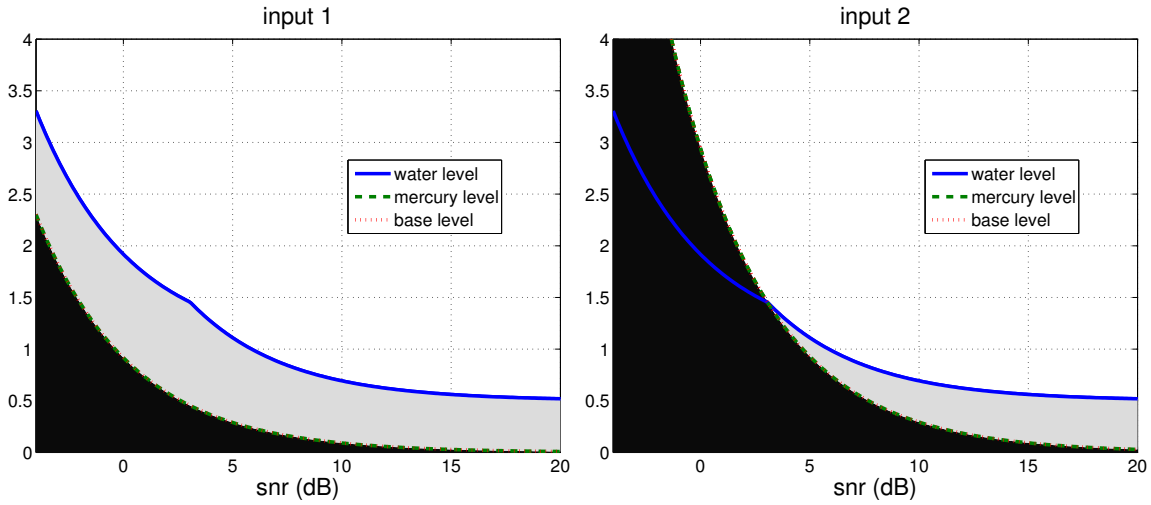
We now cast further insight into the mercury/waterfilling interpretation for MIMO channels by analyzing four cases of interest: (i) diagonal channel matrix with Gaussian inputs; (ii) diagonal channel matrix with BPSK inputs; (iii) nondiagonal channel matrix with Gaussian inputs; and (iv) nondiagonal channel matrix with BPSK inputs. Cases (i) and (ii) are used for comparison purposes and allow us to recover the standard waterfilling or the mercury/waterfilling interpretations. Case (iii) allows drawing conclusions into the operation of the general mercury/waterfilling interpretation for interfering channels. Unlike the conventional graphical illustrations of waterfilling and mercury-waterfilling which illustrate the power allocation for a given power, here we illustrate the effects of varying the power (in dB) on the resulting solution. As in Figure 4, in Figures 5-8, the water is represented by the light grey shaded area and the

mercury by the dark grey shaded area. The allocated power is equal to the height of the light grey shaded area.

(i) *Gaussian inputs and diagonal channel matrix:* In Figure 5, we show the base (=mercury) and water levels for the diagonal channel matrix

$$\mathbf{H} = \begin{bmatrix} \sqrt{1.09} & 0 \\ 0 & \sqrt{0.34} \end{bmatrix}, \tag{62}$$

with Gaussian inputs as a function of the snr. We can verify the standard waterfilling interpretation, in which all the power is assigned to the strongest input (largest $|h_{jj}|$) for low snr, and it is evenly split between the inputs for high snr.

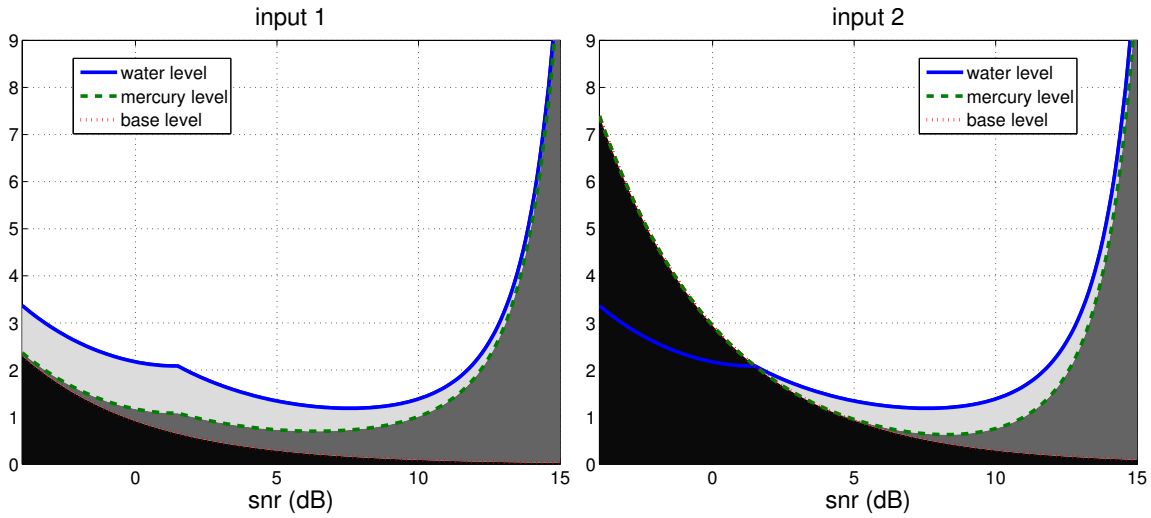



(a) (b)

Fig. 5. Power allocation for input 1 in (a) and input 2 in (b) for Gaussian inputs and diagonal channel matrix in (62).

(ii) *NonGaussian inputs and diagonal channel matrix:* In Figure 6, we show the base, mercury and water levels for the diagonal channel (62) with BPSK inputs. For the input with highest channel value (first input) we always add mercury to the base level. This mercury compensates for the mismatch between the input distribution and the optimal Gaussian distribution. Note that the mercury increases with the snr and vanishes as the snr tends to zero. For the input with the lowest gain (second input), we only add mercury to the base level once this input is allocated some power. In fact, for zero power we could consider the input distribution to be Gaussian, and we should not add mercury to the base level. As shown in [8], for discrete input distributions, the mercury grows exponentially with snr and so does the water level. For high snr, the mercury level (mismatched between Gaussian

and the input distribution) determines the power allocation, because the base level is negligible for every input.




(a) (b)

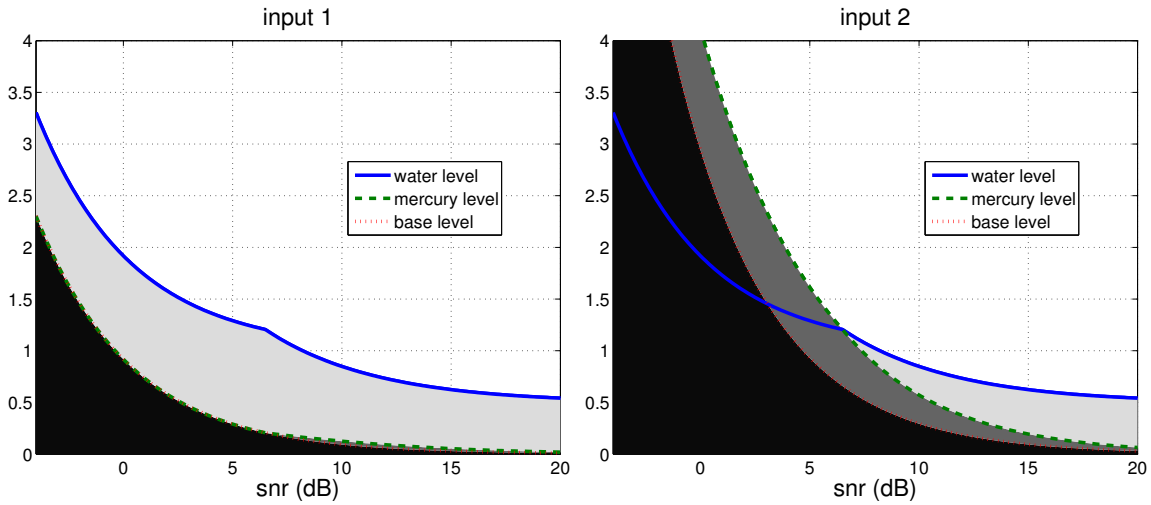
Fig. 6. Power allocation for input 1 in (a) and input 2 in (b) for BPSK inputs and diagonal channel matrix in (62).

(iii) *Gaussian inputs and nondiagonal channel matrix:* In Figure 7, we show the base, mercury and water levels for the interfering channel matrix:

$$\mathbf{H} = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 0.5 \end{bmatrix}, \tag{63}$$

with Gaussian inputs. For low $\mathsf{snr}$ values, the strongest input (input 1) gets all the power and therefore it is not interfered by the second input, consequently, there is no added mercury. The weakest input (input 2) is not assigned any power until the $\mathsf{snr}$ increase is able to compensate not only the noise but also the interference by the first input. For high $\mathsf{snr}$ values, the second input is assigned some power, as it is able to overcome both the noise and the interference. The mercury level for the first input starts rising once the second input is assigned some power, due to the interference it causes on the second input. This is barely noticeable in Figure 7, because the mercury tends to zero when the $\mathsf{snr}$ tends to infinity for a full-rank $\mathbf{H}\mathbf{H}^{\dagger}$, as predicted by Theorem 7. Note that the power is equally divided between inputs as $\mathsf{snr} \to \infty$.

(iv) *NonGaussian inputs and nondiagonal channel matrix:* In Figure 8, we show the base, mercury and water levels for the interfering channel (63) with BPSK inputs. For low $\mathsf{snr}$, the first input is assigned

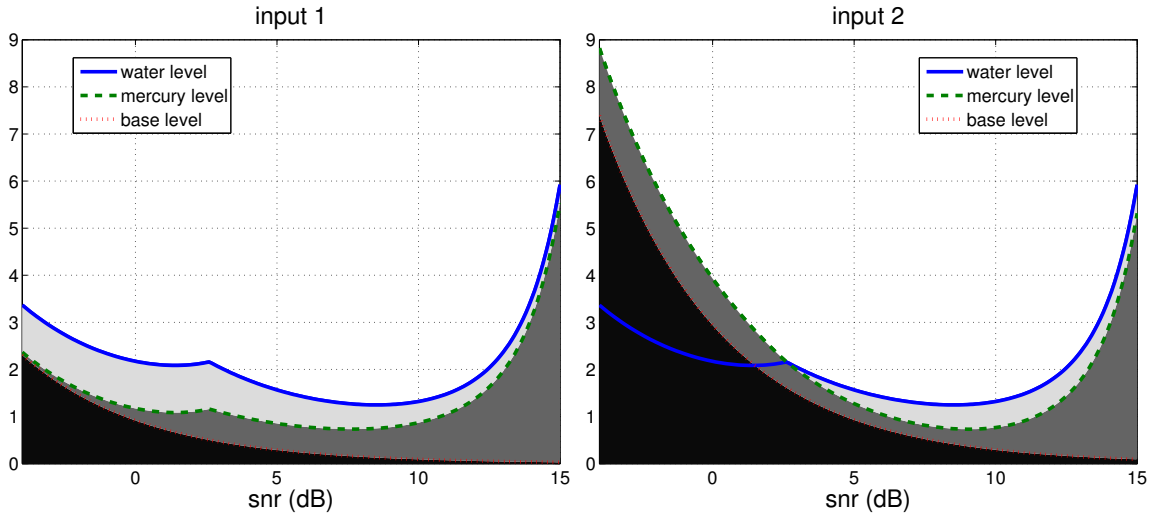



(a) (b)

Fig. 7. Power allocation for input 1 in (a) and input 2 in (b) for Gaussian inputs and nondiagonal channel matrix in (63).

all the power. The mercury level for input 1 depends only on the distribution mismatch, since there is not interference from the second input. For low snr, the base, mercury and water levels of the first input are identical to the levels of input 1 in Example (ii). The mercury level for input 2 is only due to interferences, since input 2 is not assigned any power it cannot be penalized for its distribution. This mercury level is not identical to the mercury level in Example (iii) (Gaussian input with the same nondiagonal channel matrix), because the interference comes from a different input distribution. For high snr, the mercury level is mainly due to the input distribution mismatch, because for discrete input distributions the interferences have an asymptotically negligible effect. Note that the mercury level is not equal to Example (ii) because the mercury level depends not only on the norm of the columns of $\mathbf{H}$.

For high snr, the examples with Gaussian inputs share equally the available power as the base and mercury levels vanishes, because $\mathbf{H}\mathbf{H}^{\dagger}$ is full rank. For Examples (ii) and (iv) the mercury level only depends on the inputs distributions, as the interferences are negligible, but the mercury levels and allocated powers are different in each example, because the channel matrices are different and the mercury level depends on the channel response to every input.




(a) (b)

Fig. 8. Power allocation for input 1 in (a) and input 2 in (b) for BPSK inputs and nondiagonal channel matrix in (63).

## IV. Conclusions

Regardless of whether the channel matrix is diagonal, we have shown that when the inputs are not Gaussian, information transmission rates are maximized when a full precoder is used instead of a power allocation strategy coupled to channel diagonalization. The optimal precoder matrix is expressed using the relation between the MMSE and the mutual information and it is computed by solving a fixed-point equation.

As in [8], we devoted particular attention to the low and high-snr regimes for discrete input constellations. In the low-snr regime, the optimal linear precoder diagonalizes the channel matrix, because the nonGaussianness of the input distribution has little effect on the maximum information transmission rate. We have also shown that for asymptotically high snr the optimal precoder matrix for discrete inputs achieves maximal minimum distance and minimal mean-square error, linking three common criteria for designing digital communication systems.

Whenever precoding is precluded, we optimize a power allocation matrix, and we generalize the mercury/waterfilling algorithm in [8] to interfering MIMO channels. We also put forth a graphical interpretation, in which the mercury accounts not only for the suboptimal constellation, but also for the interferences between the subchannels. For power allocation, the low and high-snr regimes, respectively, follow [8] and the high-snr regime results for the precoder in Section II-C, but using a power allocation

matrix.

## APPENDIX A
## PROOF OF THEOREM 1

The possible solutions to (2) subject to (3) are characterized by the Karush-Kuhn-Tucker theorem (e.g. [4]), which gives necessary conditions for the matrix $\mathbf{P}^*$ to be a critical point, known as the KKT or first order conditions. To compute the KKT conditions, we first build the Lagrangian:

$$\mathfrak{L}(\mathbf{P}, \lambda) = -I(\mathbf{x}; \mathbf{y}) - \lambda \left(1 - \mathrm{Tr}\{\mathbf{P}\mathbf{P}^\dagger\}\right) \tag{64}$$

in which the Lagrange multiplier $\lambda$, accounting for the inequality constraint, has to be nonnegative.

The first order conditions are given by:

$$\nabla_{\mathbf{P}} \mathfrak{L}(\mathbf{P}, \lambda) = -\nabla_{\mathbf{P}} I(\mathbf{x}; \mathbf{y}) + \lambda \mathbf{P} = 0 \tag{65}$$

$$\lambda \left(1 - \mathrm{Tr}\{\mathbf{P}\mathbf{P}^\dagger\}\right) = 0 \tag{66}$$

$$\lambda \geq 0 \tag{67}$$

If $\mathrm{Tr}\{\mathbf{P}\mathbf{P}^\dagger\} = 1$, $\lambda$ is nonzero, and the solution $\mathbf{P}^*$ to (65) satisfies

$$\mathbf{P}^* = \lambda^{-1} \nabla_{\mathbf{P}} I(\mathbf{x}; \mathbf{y}) \tag{68}$$

$$= \lambda^{-1} \mathsf{snr} \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E} \tag{69}$$

$$= \nu^{-1} \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E} \tag{70}$$

where (68) follows from the gradient of the mutual information with respect to $\mathbf{P}$ found in [11]; and in (70) we eliminate $\lambda$ by introducing

$$\nu = \frac{\lambda}{\mathsf{snr}} = \left\| \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E} \right\| \tag{71}$$

For any $\mathbf{P}$ with $\mathrm{Tr}\{\mathbf{P}\mathbf{P}^\dagger\} < 1$, we can define another precoder $\widetilde{\mathbf{P}} = \frac{1}{\sqrt{\mathrm{Tr}\{\mathbf{P}\mathbf{P}^\dagger\}}} \mathbf{P}$, which achieves higher mutual information than $\mathbf{P}$, because:

$$\frac{\partial I(\mathbf{x}; \mathbf{y})}{\mathsf{snr}} > 0. \tag{72}$$

Therefore, we only need to consider precoding matrices that fulfill (3) with equality.

## APPENDIX B
## PROOF OF COROLLARY 2

The MMSE for the channel model in (23) is given by:

$$\mathsf{mmse}(\mathsf{snr}) = 4||\mathbf{h}||^2 \sum_{i=0}^{\infty} (-1)^i e^{(i^2+i)4||\mathbf{h}||^2\mathsf{snr}} Q\left((2i+1)||\mathbf{h}||\sqrt{2\mathsf{snr}}\right), \tag{73}$$

which can be obtained from [8, (79)], replacing $d$ by $2||\mathbf{h}||$.

The bounds in (27) and (28) are obtained using the well-known bounds (e.g. [14]):

$$\frac{e^{-\xi^2/2}}{\sqrt{2\pi}\xi}\left(1-\frac{1}{\xi^2}\right) \leq Q(\xi) \leq \frac{e^{-\xi^2/2}}{\sqrt{2\pi}\xi} \tag{74}$$

## APPENDIX C
## PROOF OF THEOREM 3

The proof of Theorem 3 follows the proof in [8, Theorem 4] with some minor modifications. We label each of the different $M$ constellation vectors by $\{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$, in such a way that

$$d_{min} = ||\mathbf{HP}(\mathbf{x}_i - \mathbf{x}_{(i+1)_M})|| \tag{75}$$

for at least one $i$ and $(i)_M = i - M\lfloor (i-1)/M \rfloor$ is the modulo $M$ operation.

To prove the lower bound, we construct a genie that supplies the estimator with a pair of constellation vectors: if the transmitted constellation vector is $\mathbf{x}_i$, then with probability $1/2$ the genie gives the pair $\{\mathbf{x}_i, \mathbf{x}_{(i+1)_M}\}$ and with probability $1/2$ gives the pair $\{\mathbf{x}_{(i-1)_M}, \mathbf{x}_i\}$. Hence,

$$\mathsf{mmse}(\mathsf{snr}) = \frac{1}{M}\sum_{i=1}^{M} \mathbb{E}\left[\|\mathbf{HP}(\mathbf{x}_i - \mathbb{E}[\mathbf{x}|\mathbf{y}])\|^2 \,|\mathbf{x}=\mathbf{x}_i\right] \tag{76}$$

$$\geq \frac{1}{2M}\sum_{i=1}^{M}\Bigg(\mathbb{E}\left[\left\|\mathbf{HP}(\mathbf{x}_i - \bar{\mathbf{x}}(\mathbf{y}, \{\mathbf{x}_{(i-1)_M}, \mathbf{x}_i\}))\right\|^2 |\mathbf{x}=\mathbf{x}_i\right]$$

$$+ \mathbb{E}\left[\left\|\mathbf{HP}(\mathbf{x}_i - \bar{\mathbf{x}}(\mathbf{y}, \{\mathbf{x}_i, \mathbf{x}_{(i+1)_M}\}))\right\|^2 |\mathbf{x}=\mathbf{x}_i\right]\Bigg) \tag{77}$$

$$\geq \frac{d_{min}^2}{2M}\frac{e^{-d_{min}^2\mathsf{snr}/4}}{d_{min}\sqrt{\mathsf{snr}}}\left(\sqrt{\pi} - \frac{4.37}{d_{min}^2\mathsf{snr}}\right) \tag{78}$$

where

$$\bar{\mathbf{x}}(\mathbf{y}, \{\mathbf{a}, \mathbf{b}\}) = \frac{\mathbf{a}e^{-\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{HPa}\|^2} + \mathbf{b}e^{-\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{HPb}\|^2}}{e^{-\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{HPa}\|^2} + e^{-\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{HPb}\|^2}} \tag{79}$$

The inequality in (78) follows from $\bar{\mathbf{x}}_i(\mathbf{y}, \mathsf{snr}, \{\mathbf{a}, \mathbf{b}\})$ being the mean-square estimate for the channel model in Corollary 1 with $\mathbf{h} = (\mathbf{HP}(\mathbf{a}-\mathbf{b}))/2$, the lower bound in Corollary 2, and taking the largest term in the sum. This completes the proof of the lower bound in (29).

We now prove the upper bound in (30), and as in [8], we replace the optimal MSE estimator by the maximum likelihood estimator:

$$\mathbf{x}_{ML}(\mathbf{y}) = \underset{\mathbf{x}}{\operatorname{argmin}} ||\mathbf{y} - \sqrt{\mathsf{snr}}\mathbf{H}\mathbf{P}\mathbf{x}||^2, \tag{80}$$

thereby

$$\begin{aligned}
\mathsf{mmse}(\mathsf{snr}) &= \mathbb{E}\left[||\mathbf{H}\mathbf{P}(\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathbf{y}])||^2\right] && (81)\\
&\leq \mathbb{E}\left[||\mathbf{H}\mathbf{P}(\mathbf{x} - \mathbf{x}_{ML}(\mathbf{y}))||^2\right] && (82)\\
&= \frac{1}{M}\sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M} \mathbb{E}\left[||\mathbf{H}\mathbf{P}(\mathbf{x}_i - \mathbf{x}_j)||^2 | \mathbf{x} = \mathbf{x}_i, \mathbf{y} \in \mathcal{V}_j\right] \mathbb{P}[\mathbf{y} \in \mathcal{V}_j | \mathbf{x} = \mathbf{x}_i] && (83)\\
&= \frac{1}{M}\sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M} ||\mathbf{H}\mathbf{P}(\mathbf{x}_i - \mathbf{x}_j)||^2 \mathbb{P}[\mathbf{y} \in \mathcal{V}_j | \mathbf{x} = \mathbf{x}_i] && (84)
\end{aligned}$$

where $\mathcal{V}_j$ is the Voronoi region for $\mathbf{H}\mathbf{P}\mathbf{x}_j$ in the received constellation. We have assumed as in (77) that all the constellation vectors are equally likely.

We can now complete the proof for the upper bound in (30):

$$\begin{aligned}
\mathsf{mmse}(\mathsf{snr}) &\leq \frac{1}{M}\sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M} ||\mathbf{H}\mathbf{P}(\mathbf{x}_i - \mathbf{x}_j)||^2 \mathbb{P}[\mathbf{y} \in \mathcal{V}_j | \mathbf{x} = \mathbf{x}_i] && (85)\\
&\leq \frac{d_{max}^2}{M}\sum_{i=1}^{M} \mathbb{P}[\mathbf{y} \notin \mathcal{V}_i | \mathbf{x} = \mathbf{x}_i] && (86)\\
&\leq \frac{d_{max}^2}{M}\sum_{i=1}^{M} (M-1) Q\left(d_{i,min}\sqrt{\frac{\mathsf{snr}}{2}}\right) && (87)\\
&\leq d_{max}^2 (M-1) Q\left(d_{min}\sqrt{\frac{\mathsf{snr}}{2}}\right) && (88)\\
&\leq d_{max}^2 (M-1) \frac{e^{-d_{min}^2 \mathsf{snr}/4}}{d_{min}\sqrt{\mathsf{snr}\pi}} && (89)
\end{aligned}$$

We obtain the inequality in (86) replacing $\mathbb{E}\left[||\mathbf{H}\mathbf{P}(\mathbf{x}_i - \mathbf{x}_j)||^2 | \mathbf{x} = \mathbf{x}_i, \mathbf{y} \in \mathcal{V}_j\right]$ by $d_{max}^2$ (the maximum distance in the received between the constellation vectors (33)). The inequality in (87) is obtained,

computing the union bound [14] for the probability that $\mathbf{y}$ does not lie in the Voronoi Region of $\mathbf{HPx}_i$:

$$\mathbb{P}[\mathbf{y} \notin \mathcal{V}_i | \mathbf{x} = \mathbf{x}_i] \leq \sum_{\substack{j=1 \\ j \neq i}}^{M} \mathbb{P}[\mathbf{y} \notin \mathcal{V}_{ij} | \mathbf{x} = \mathbf{x}_i] \tag{90}$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^{M} Q\left( d_{ij} \sqrt{\frac{\mathsf{snr}}{2}} \right) \tag{91}$$

$$\leq (M-1) Q\left( d_{i,min} \sqrt{\frac{\mathsf{snr}}{2}} \right) \tag{92}$$

where $\mathcal{V}_{ij}$ is the Voronoi region for $\mathbf{HPx}_i$, when we assume that only $\mathbf{x}_i$ and $\mathbf{x}_j$ have been transmitted,

$$d_{ij} = ||\mathbf{HP}(\mathbf{x}_j - \mathbf{x}_i)||, \tag{93}$$

and $d_{i,min} = \min_j d_{ij}$.

We sum $M$ times the largest term in the sum to reach the inequality in (88) and the inequality in (89) uses the upper bound in (74) for the $Q(\cdot)$ function.

This theorem generalizes [8, Theorem 4] to nondiagonal rectangular channel matrices and it follows the same ideas to prove the upper and lower bounds with two minor modifications. For the upper bound (30), we use the union bound for $\mathbb{P}[\mathbf{y} \notin \mathcal{V}_i | \mathbf{x} = \mathbf{x}_i]$. The union bound allows to get a tighter bound for large $\mathsf{snr}$, as it incorporates an extra $1/\sqrt{\mathsf{snr}}$ in the bound. For the lower bound in (78), we use the lower bound in (27), instead of the first order approximation to $Q(\cdot)$ as they do in [8], and a different definition for the genie.

## APPENDIX D
## PROOF OF THEOREM 4

From the relation between the mutual information and the MMSE in [5] for the channel model in (1), we know that we can express the mutual information as:

$$I(\mathbf{x}; \mathbf{y}) = \int_0^{\mathsf{snr}} \mathsf{mmse}(\xi) d\xi \tag{94}$$

$$= \log M - \int_{\mathsf{snr}}^{\infty} \mathsf{mmse}(\xi) d\xi \tag{95}$$

We use (95) to bound the mutual information, because the bounds that we have presented for the MMSE are tight for large $\mathsf{snr}$.

We can upper bound the MMSE as:

$$\mathsf{mmse}(\mathsf{snr}) \leq (M-1) d_{max}^2 \frac{e^{-d_{min}^2 \mathsf{snr}/4}}{d_{min} \sqrt{\mathsf{snr}\pi}} \left( 1 + \frac{2}{d_{min}^2 \mathsf{snr}} \right) \tag{96}$$

The inequality in (96) follows from (89) and from $\frac{2}{d_{min}^2 \mathsf{snr}} \geq 0$. We get the lower bound of the mutual information in (34) when we integrate (96) from $\mathsf{snr}$ to infinity.

We can lower bound the MMSE as:

$$\mathsf{mmse}(\mathsf{snr}) \geq \frac{d_{min}^2}{2M} \frac{e^{-d_{min}^2 \mathsf{snr}/4}}{d_{min}\sqrt{\mathsf{snr}}} \left( \sqrt{\pi} - \frac{4.37}{d_{min}^2 \mathsf{snr}} - \frac{6(4.37 + 2\sqrt{\pi})}{d_{min}^4 \mathsf{snr}^2} \right) \tag{97}$$

The inequality in (97) follows from (78) and from $\frac{6(4.37+2\sqrt{\pi})}{d_{min}^4 \mathsf{snr}^2} \geq 0$. We get the upper bound of the mutual information in (35) from integrating (97) from $\mathsf{snr}$ to infinity.

## APPENDIX E
## PROOF OF THEOREM 5

The Lagrangian for the optimization problem in (43)-(46) is given by:

$$\mathfrak{L}(\mathbf{p}, \lambda, \boldsymbol{\eta}) = -I(\mathbf{x}; \mathbf{y}) - \lambda \left( 1 - \sum_{j=1}^{m} p_j \right) - \sum_{j=1}^{m} \eta_j p_j, \tag{98}$$

in which the Lagrange multipliers $\lambda$ and $\eta_j$ for the inequality constraints have to be nonnegative.

The KKT conditions are given by:

$$\frac{\partial \mathfrak{L}(\mathbf{p}, \lambda, \boldsymbol{\eta})}{\partial p_j} = -\frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} + \lambda - \eta_j = 0 \tag{99}$$

$$\lambda \left( 1 - \sum_{j=1}^{n} p_j \right) = 0 \tag{100}$$

$$p_j \eta_j = 0 \tag{101}$$

$$\lambda, \eta_j \geq 0 \tag{102}$$

The derivative of $I(\mathbf{x}; \mathbf{y})$ with respect to $p_j$ is given by:

$$\frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} = \frac{\mathsf{snr}}{\sqrt{p_j}} \left( \mathbf{H}^\dagger \mathbf{H} \mathbf{P} \mathbf{E} \right)_{jj}, \tag{103}$$

which can be computed from [11, (21)] and the chain rule.

For $p_j^* > 0$:

$$\lambda = \left. \frac{\partial I(\mathbf{x}; \mathbf{y})}{\partial p_j} \right|_{\mathbf{P}=\mathbf{P}^*} \tag{104}$$

$$= \frac{\mathsf{snr}}{\sqrt{p_j}^*} \left( \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \mathbf{E} \right)_{jj} \geq 0, \tag{105}$$

because $\eta_j = 0$ for nonzero $p_j^*$. $\lambda$ is nonzero because $\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}$ is positive semi-definite, thus $\sum_j p_j^* = 1$. For $p_j^* = 0$:

$$\left(\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}\right)_{jj} = 0, \tag{106}$$

The optimal solution can be expressed as:

$$p_j^* = \frac{\mathsf{snr}}{\lambda}\sqrt{p_j^*}\left(\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}\right)_{jj} \tag{107}$$

$$= \gamma^{-1}\left(\mathbf{P}^*\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}\right)_{jj} \tag{108}$$

where we have combined (105) and (106) to obtain a single equation for all $p_j^*$ and $\gamma = \frac{\lambda}{\mathsf{snr}} = \mathrm{Tr}\left\{\mathbf{P}^*\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\mathbf{E}\right\}$ to ensure that $\sum_j p_j^* = 1$.

## APPENDIX F
## MERCURY LEVEL FOR $p_j^* = 0$

Using (54) and for $p_j^* = 0$, we can rewrite the mercury level as follows:

$$m_j^{-1} = \left.\frac{\partial I(\mathbf{x};\mathbf{y})}{\partial p_j}\right|_{\mathbf{P}=\mathbf{P}^*} \tag{109}$$

$$= \mathsf{snr}\|\mathbf{h}_j\|^2 - \mathsf{snr}\sum_{i=1}^{m}\mathbf{h}_j^\dagger\mathbf{h}_i\sqrt{p_i^*}\mathbb{E}\left[\left(\frac{\mathbb{E}\left[x_j|\mathbf{y}\right]}{\sqrt{p_j^*}}\right)^\dagger\mathbb{E}\left[x_i|\mathbf{y}\right]\right] \tag{110}$$

For $p_j^* = 0$, $\mathbb{E}[\mathbf{x}_j|\mathbf{y}] = 0$ and (110) is indeterminate. To derive the mercury level, we compute the limit of (110) as $p_j^* \to 0$:

$$\lim_{p_j^*\to 0}\frac{\mathbb{E}\left[x_j|\mathbf{y}\right]}{\sqrt{p_j^*}} = \lim_{p_j^*\to 0}\int\frac{x_j}{\sqrt{p_j^*}}\frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})p_{\mathbf{x}}(\mathbf{x})}{p_{\mathbf{x}}(\mathbf{y})}d\mathbf{x} \tag{111}$$

$$= 2\frac{\sqrt{\mathsf{snr}}}{\pi^n}\int e^{-\|\mathbf{a}\|^2}\Re\{\mathbf{a}^\dagger\mathbf{h}_j x_j\}x_j\frac{p_{\mathbf{x}}(\mathbf{x})}{p_{\mathbf{y}}(\mathbf{y})}d\mathbf{x} \tag{112}$$

where

$$\mathbf{a} = y - \sqrt{\mathsf{snr}}\mathbf{H}\mathbf{P}^*\mathbf{x} + \sqrt{\mathsf{snr}}\mathbf{h}_j\sqrt{p_j^*}x_j, \tag{113}$$

because of the cancellation in (113), note that $\mathbf{a}$ does not depend on either $x_j$ or $p_j^*$.

To obtain the limit in (112), we have expanded $p(\mathbf{y}|\mathbf{x})$ as follows:

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^n}e^{-\left\|\mathbf{y}-\sqrt{\mathsf{snr}}\mathbf{H}\mathbf{P}^*\mathbf{x}\right\|^2} \tag{114}$$

$$= \frac{1}{\pi^n}e^{-\left\|\mathbf{a}-\sqrt{\mathsf{snr}}\mathbf{h}_j\sqrt{p_j^*}x_j\right\|^2} \tag{115}$$

$$= \frac{e^{-\|\mathbf{a}\|^2}}{\pi^n}\left(1 + 2\sqrt{\mathsf{snr}p_j^*}\Re\{\mathbf{a}^\dagger\mathbf{h}_j x_j\}\right) + \mathcal{O}(p_j^*) \tag{116}$$

and to obtain the last equality in (112), we should notice that $\mathcal{O}(p_j^*)/\sqrt{p_j^*} \to 0$ for $p_j^* \to 0$ and that for any matrix $\mathbf{P}$:

$$\frac{1}{\pi^n}\int e^{-\|\mathbf{a}\|^2} x_j \frac{p_{\mathbf{x}}(\mathbf{x})}{p_{\mathbf{y}}(\mathbf{y})} d\mathbf{x} = 0, \tag{117}$$

because the $j^{th}$ marginal of $p_{\mathbf{x}}(\mathbf{x})$ is symmetric and $\mathbf{a}$ does not depend on $x_j$.

We can show that the mercury level is above the water level if $p_j^* = 0$, using the first order conditions in Appendix E. For $p_j^* = 0$, (99) becomes

$$\eta_j = \lambda - \left.\frac{\partial I(\mathbf{x};\mathbf{y})}{\partial p_j}\right|_{\mathbf{P}=\mathbf{P}^*} \geq 0, \tag{118}$$

thus

$$m_j \geq \beta \tag{119}$$

Finally we can prove that the mercury level is above the base level for $p_j^* = 0$ using (110). $h_j\mathbb{E}[x_j|\mathbf{y}]$ and $h_i\mathbb{E}[x_i|\mathbf{y}]$ are positive correlated, then:

$$m_j^{-1} \leq \mathsf{snr}\|\mathbf{h}_j\|^2 = b_j^{-1} \tag{120}$$

## APPENDIX G
## PROOF OF THEOREM 6

We first prove that the derivative of $m_j - b_j$ with respect to $\mathsf{snr}$ is nonpositive for Gaussian inputs. For Gaussian inputs the derivative of the mutual information with respect to $p_j$ can be computed analytically:

$$\frac{\partial I(\mathbf{x};\mathbf{y})}{\partial p_j} = \mathsf{snr}\mathbf{h}_j^\dagger \left(\mathbf{I} + \mathsf{snr}\mathbf{H}\mathbf{P}^2\mathbf{H}^\dagger\right)^{-1}\mathbf{h}_j \tag{121}$$

Hence, we can rearrange $m_j - b_j$ as follows:

$$m_j - b_j = \frac{1}{\mathsf{snr}\mathbf{h}_j^\dagger\left(\mathbf{I} + \mathsf{snr}\mathbf{H}\mathbf{P}^{*2}\mathbf{H}^\dagger\right)^{-1}\mathbf{h}_j} - p_j^* - \frac{1}{\mathsf{snr}\|\mathbf{h}_j\|^2} \tag{122}$$

$$= \frac{\|\mathbf{h}_j\|^2 - \mathbf{h}_j^\dagger\left(\mathbf{I} + \mathsf{snr}\mathbf{H}\mathbf{P}^{*2}\mathbf{H}^\dagger\right)^{-1}\mathbf{h}_j}{\mathsf{snr}\|\mathbf{h}_j\|^2\mathbf{h}_j^\dagger\left(\mathbf{I} + \mathsf{snr}\mathbf{H}\mathbf{P}^{*2}\mathbf{H}^\dagger\right)^{-1}\mathbf{h}_j} - p_j^* \tag{123}$$

$$= \frac{\mathbf{h}_j^\dagger\mathbf{H}\mathbf{P}^*\left(\mathbf{I} + \mathsf{snr}\mathbf{P}^*\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\right)^{-1}\mathbf{P}^*\mathbf{H}^\dagger\mathbf{h}_j}{\|\mathbf{h}_j\|^2\mathbf{h}_j^\dagger\left(\mathbf{I} + \mathsf{snr}\mathbf{H}\mathbf{P}^{*2}\mathbf{H}^\dagger\right)^{-1}\mathbf{h}_j} - p_j^* \tag{124}$$

$$= \frac{\mathbf{h}_j^\dagger\mathbf{M}_1\mathbf{h}_j}{\|\mathbf{h}_j\|^2\left(\|\mathbf{h}_j\|^2 - \mathsf{snr}\mathbf{h}_j^\dagger\mathbf{M}_1\mathbf{h}_j\right)} - p_j^* \tag{125}$$

where we have used the matrix inversion lemma to obtain the last two equalities and have defined:

$$\mathbf{M}_1 = \mathbf{H}\mathbf{P}^*\left(\mathbf{I} + \mathsf{snr}\mathbf{P}^*\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^*\right)^{-1}\mathbf{P}^*\mathbf{H}^\dagger \tag{126}$$

The derivative of $m_j - b_j$ is given by:

$$\frac{\partial(m_j - b_j)}{\partial \mathsf{snr}} = \frac{\mathbf{h}_j^\dagger \frac{\partial \mathbf{M}_1}{\partial \mathsf{snr}} \mathbf{h}_j \left( \|\mathbf{h}_j\|^2 - \mathsf{snr} \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \right)}{\|\mathbf{h}_j\|^2 \left( \|\mathbf{h}_j\|^2 - \mathsf{snr} \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \right)^2} \tag{127}$$

$$+ \frac{\mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \left[ \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j + \mathsf{snr} \mathbf{h}_j^\dagger \frac{\partial \mathbf{M}_1}{\partial \mathsf{snr}} \mathbf{h}_j \right]}{\|\mathbf{h}_j\|^2 \left( \|\mathbf{h}_j\|^2 - \mathsf{snr} \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \right)^2} \tag{128}$$

$$= \frac{\left( \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \right)^2 - \|\mathbf{h}_j\|^2 \mathbf{h}_j^\dagger \mathbf{M}_1^2 \mathbf{h}_j}{\left( \|\mathbf{h}_j\| \left( \|\mathbf{h}_j\|^2 - \mathsf{snr} \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \right) \right)^2} \tag{129}$$

where the derivative of $\mathbf{M}_1$ with respect to $\mathsf{snr}$ is given by:

$$\frac{\partial \mathbf{M}_1}{\partial \mathsf{snr}} = -\mathbf{H}\mathbf{P}^* \left( \mathbf{I} + \mathsf{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \right)^{-1} \mathbf{P}^* \mathbf{H}^\dagger \tag{130}$$

$$\times \mathbf{H}\mathbf{P}^* \left( \mathbf{I} + \mathsf{snr} \mathbf{P}^* \mathbf{H}^\dagger \mathbf{H} \mathbf{P}^* \right)^{-1} \mathbf{P}^* \mathbf{H}^\dagger \tag{131}$$

$$= - \mathbf{M}_1^2 \tag{132}$$

We can see that

$$\left( \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \right)^2 = \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \tag{133}$$

$$= \mathrm{Tr} \left\{ \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \right\} \tag{134}$$

$$= \mathrm{Tr} \left\{ \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \right\} \tag{135}$$

$$\leq \mathrm{Tr} \left\{ \mathbf{h}_j \mathbf{h}_j^\dagger \right\} \mathrm{Tr} \left\{ \mathbf{M}_1 \mathbf{h}_j \mathbf{h}_j^\dagger \mathbf{M}_1 \right\} \tag{136}$$

$$= \|\mathbf{h}_j\|^2 \mathbf{h}_j^\dagger \mathbf{M}_1^2 \mathbf{h}_j \tag{137}$$

where we have used that for positive semi-definite matrices: $\mathrm{Tr}\{\mathbf{A}\mathbf{B}\} \leq \mathrm{Tr}\{\mathbf{A}\}\mathrm{Tr}\{\mathbf{B}\}$ [9].

We have just proven with (137) that the numerator of (129) is nonpositive and the denominator is always positive, because it is the square of a real number. Hence, the derivative of $m_j - b_j$ with respect to the $\mathsf{snr}$ is nonpositive for any MIMO Gaussian channel with Gaussian inputs.

We use the first equality in (125) to compute its derivative with respect to $p_i$:

$$\frac{\partial(m_j - b_j)}{\partial p_i} = - \frac{\mathsf{snr} \mathbf{h}_j^\dagger \frac{\partial (\mathbf{I} + \mathsf{snr} \mathbf{H} \mathbf{P}^{*2} \mathbf{H}^\dagger)^{-1}}{\partial p_i} \mathbf{h}_j}{\left( \mathsf{snr} \mathbf{h}_j^\dagger \left( \mathbf{I} + \mathsf{snr} \mathbf{H} \mathbf{P}^{*2} \mathbf{H}^\dagger \right)^{-1} \mathbf{h}_j \right)^2} - \delta_{ij} \tag{138}$$

$$= \frac{\left\| \mathbf{h}_j^\dagger \left( \mathbf{I} + \mathsf{snr} \mathbf{H} \mathbf{P}^{*2} \mathbf{H}^\dagger \right)^{-1} \mathbf{h}_i \right\|^2}{\left( \mathbf{h}_j^\dagger \left( \mathbf{I} + \mathsf{snr} \mathbf{H} \mathbf{P}^{*2} \mathbf{H}^\dagger \right)^{-1} \mathbf{h}_j \right)^2} - \delta_{ij} \tag{139}$$

This derivative is nonnegative for $i \neq j$ and zero for $i = j$.

## APPENDIX H
## PROOF OF THEOREM 7

The derivative of the mutual information with respect to $p_j$ for Gaussian inputs is given by:

$$\frac{\partial I(\mathbf{x};\mathbf{y})}{\partial p_j} = \mathbf{h}_j^\dagger \left(\mathbf{H}\mathbf{P}^2\mathbf{H}^\dagger + \mathbf{I}/\mathsf{snr}\right)^{-1} \mathbf{h}_j \tag{140}$$

If all the channels are assigned the same power ($p_j = 1/m$, $\forall p_j$) and $\mathbf{H}\mathbf{H}^\dagger$ is full-rank, we have

$$\lim_{\mathsf{snr}\to\infty} \frac{\partial I(\mathbf{x};\mathbf{y})}{\partial p_j} = m\mathbf{h}_j^\dagger \left(\mathbf{H}\mathbf{H}^\dagger\right)^{-1} \mathbf{h}_j = m \qquad \forall j \tag{141}$$

This means that the water level $\beta$ tends to $1/m$ and the mercury level tends to zero. Therefore, equal power allocation for asymptotically large $\mathsf{snr}$ fulfills the conditions of Theorem 5 and for Gaussian inputs the objective function is asymptotocally concave.

## REFERENCES


[1] J. Cioffi, S. Jagannathan, M. Mohseni, and G. Ginis. CuPON: The copper alternative to PON 100gb/s DSL networks. *IEEE Communications Magazine*, 45(6):132–139, June 2007.

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

[3] R. F. H. Fischer, C. Stierstorfer, and C. Windpassinger. Precoding and signal shaping for transmission over MIMO channels. In *Proceedings of the Canadian Workshop on Information Theory*, pages 83–87, May 2003.

[4] R. Fletcher. *Practical Methods of Optimization*. Wiley, New York, second edition, 1987.

[5] D. Guo, S. Shamai, and S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, April 2005.

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

[7] K. H. Lee and D. P. Petersen. Optimal linear coding for vector channels. *IEEE Transactions on Communications*, 24:1283–1290, december 1976.

[8] A. Lozano, A. M. Tulino, and S. Verdú. Optimum power allocation for parallel Gaussian channels with arbitrary input distributions. *IEEE Transactions on Information Theory*, 52(7):3033–3051, July 2006.

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

[10] 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 Transactions on Signal Processing*, 51(9):2381–2401, September 2003.

[11] D. P. Palomar and S. Verdú. Gradient of mutual information in linear vector Gaussian channels. *IEEE Transactions on Information Theory*, 52(1):141–154, January 2006.

[12] F. Pérez-Cruz, M. Rodrigues, and S. Verdú. Optimal precoding for digital subscriber lines. In *International Conference on Communications (ICC)*, May 2008.

[13] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú. Generalized mercury/waterfilling for multiple-input multiple-output channels. In *45th Allerton Conference on Communication, Control, and Computing*, September 2007.

[14] J. G. Proakis and M. Salehi. *Communication Systems Engineering*. Prentice Hall, New York, 2 edition, 2002.

[15] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath. Optimal designs for space-time linear precoders and decoders. *IEEE Transactions on Signal Processing*, 50(5):1051–1064, May 2002.

[16] S. Verdú. *Multiuser Detection*. Cambridge University Press, New York, 1998.

[17] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Transactions on Information Theory, Special Issue on Shannon Theory: Perspective, Trends and Applications*, 48(6):1319–1343, June 2002.

[18] J. Yang and S. Roy. Joint transmitter-receiver optimization for multiple-input multiple-output systems with decision feedback. *IEEE Transactions on Information Theory*, 40:1334–1347, September 1994.

[19] J. Yang and S. Roy. On joint transmitter and receiver optimization for multiple-input multiple-output (MIMO) transmission systems. *IEEE Transactions on Communications*, 42:3221–3231, december 1994.

[20] W. Yu, W. Rhee, S. Boyd, and J. M.Cioffi. Iterative waterfilling for Gaussian vector multiple-access channels. *IEEE Transactions on Information Theory*, 50(1):145–152, January 2004.