Robust Design of Spatial Tomlinson-Harashima Precoding in the Presence of Errors in the CSI

Miquel Payaró, *Student Member, IEEE*, Antonio Pascual-Iserte, *Member, IEEE*, Ana I. Pérez-Neira, *Senior Member, IEEE*, and Miguel Á. Lagunas, *Fellow, IEEE*

***Abstract*— In this letter, we study spatial Tomlinson-Harashima precoding (STHP) from an information theoretic point of view encompassing both the single and the multiuser scenarios. Since the performance of any communications precoding scheme relies on the quality of the channel state information (CSI) that is made available at the transmitter side, in this work we focus our attention on the analysis of the robustness of STHP in terms of rate loss when the CSI is imperfect. Precisely, we present a robust design of the set of moduli used in STHP and, consequently, of the corresponding power allocation, that maximizes the achievable rates for the worst-case errors in the CSI in the small errors regime.**

***Index Terms*— MIMO robust designs, Tomlinson-Harashima precoding, multiuser communication, MIMO channels, decision-feedback equalization.**

I. Introduction

THE pioneering work by Costa in 1983 [1] and recent theoretical results describing the sum-capacity when using multiple antennas to communicate with multiple users in a known rich scattering channel [2]–[6], have motivated a significant research on multi-input multi-output (MIMO) downlink transmit strategies. One of such downlink strategies [7], [8] was the extension of the non-linear precoding scheme for temporal intersymbol interference mitigation proposed by Tomlinson [9] and Harashima [10], into spatial interference equalization for MIMO systems, leading to the so-called spatial Tomlinson-Harashima precoding (STHP) scheme. The main idea behind STHP is to pre-subtract the interference that is produced by the presence of the MIMO channel before the signal is actually transmitted, which turns out to be computationally very simple to implement.

Manuscript received December 2, 2005; revised July 15, 2006 and October 18, 2006; accepted October 30, 2006. The associate editor coordinating the review of this paper and approving it for publication was S. Zhou. This work was partially supported by the Catalan Government under grants 2003FI-00195, 2005SGR-00690, and 2005SGR-00996; and by the Spanish Government under projects TEC2005-08122-C03 (ULTRA-PROCESS) and 2A103 (MIMOWA) from MEDEA+ program.

This work has been partially presented in IEEE Vehicular Technology Conference Fall (VTC'04), Los Angeles, CA (USA), Sept. 2004 and IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP'05), Philadelphia, PA (USA), Mar. 2005.

M. Payaró conducted his part of this research while he was with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain. He is now with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: miquel.payaro@ust.hk).

A. Pascual-Iserte, A. I. Pérez-Neira, and M. Á. Lagunas are with the Universitat Politècnica de Catalunya (UPC) and Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona, Spain (e-mail: {tonip, anuska}@gps.tsc.upc.es, m.a.lagunas@cttc.es).

Digital Object Identifier 10.1109/TWC.2007.05941.

In addition to STHP, other downlink precoding strategies have been proposed, such as the transmission architecture for the MIMO broadcast channel developed in [11], [12], where a transmission scheme based on a regularized inversion of the channel combined with a vector perturbation was developed and analyzed. A similar scheme as in [11] was proposed in [13], where optimality was sacrificed to reduce the computational load by utilizing Babai's closest-point approximate solution. Other downlink schemes include the STHP optimized with respect to the sum of mean square errors as given in [14], or the non-linear precoding scheme with vector perturbation described in [15] which shows the best performance among all the cited schemes.

A commonality among all these schemes is that, since the precoder has to be matched to the channel, some degree of channel state information (CSI) has to be made available at the transmitter. In realistic scenarios, this CSI is imperfect, due to errors in the estimation of the channel or due to the presence of a noisy feedback channel, implying a performance degradation of the system. The effects of imperfect CSI for STHP in a single user scenario, in terms of signal-to-noise ratio (SNR) loss, were studied in [16]. The authors in [17] proposed a robust design of the transceiver matrices in STHP in a multiuser broadcast scenario taking a Bayesian modeling of the errors in the CSI and looking for the minimization of the sum of the MSE for all the users. An additional zero-forcing constraint but keeping the same design objective and error modeling was given in [18].

In this letter, we address the problem of evaluating the *rate* loss of STHP due to small errors in the available CSI both for the single and multiuser cases from an information theoretic point of view. We present a maximin *robust* design of the set of moduli used in STHP and, consequently, of the corresponding power allocation among the transmitted data streams, that maximizes the worst-case achievable rates. The choice of analyzing the STHP scheme is due to the fact that it is a widely accepted transmission architecture which can be implemented in real systems because it requires a small amount of computational effort and also for reasons of simplicity of analysis.

II. System Model

We consider the downlink of a flat fading wireless communications system with n_T transmitters at the base station (BS) and n_R receivers. This gives rise to a MIMO channel, which is represented by the matrix $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$. The received signal vector, $\mathbf{y} \in \mathbb{C}^{n_R}$, is $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{x} \in \mathbb{C}^{n_T}$ is the transmitted vector and $\mathbf{n} \in \mathbb{C}^{n_R}$ is the noise vector, whose

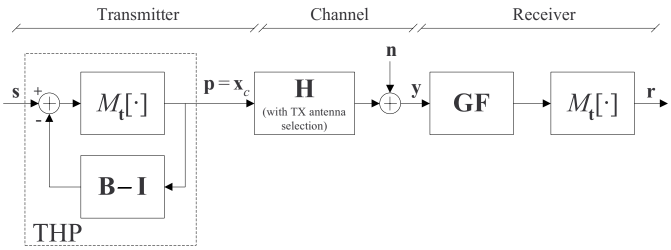

Fig. 1. Single user communication scheme with Spatial Tomlinson-Harashima Precoding.

entries are considered to be i.i.d. circularly symmetric complex Gaussian random variables with zero mean and variance σ^2. In the single user case, see Fig. 1, the k-th element of the received vector, y_k, represents the received signal at the k-th antenna of that single user, and in the multiuser case, see Fig. 2, y_k stands for the received signal by the single antenna of k-th user.

The input of the STHP is a vector, $\mathbf{s} \in \mathbb{C}^{n_S}$, whose entries are the n_S data symbols that are to be transmitted, s_k, $k = 1, \ldots, n_S$. The output of the precoder, $\mathbf{p} \in \mathbb{C}^{n_S}$, is obtained from these data symbols in a similar way to the recursive causal spatial relation described in [7], [8], as

$$p_k = M_{t_k}\left[s_k - \sum_{l=1}^{k-1} b_{kl} p_l\right], \quad \text{for } k = 1, \ldots, n_S\,. \quad (1)$$

In (1), b_{kl} are a set of coefficients to be determined, and $M_{t_k}[z]$ represents a complex modulo reduction of $z \in \mathbb{C}$ into the complex square region $\mathcal{D}_k \equiv [-t_k, t_k) \times [-\mathrm{j}\, t_k, \mathrm{j}\, t_k)$, which is used to limit the dynamic range of the output signal to overcome the problem of the increase of the transmitted power in precoding schemes where the channel response is inverted. Notice that a difference between (1) and the scheme in [7] (and other previous works) is that our scheme allows different moduli, t_k, for each component of the precoder output $\mathbf{p}$, in a similar way as the inflated lattice concept introduced in [19] and [20]. As it will be shown later, this enables rate adaptation per stream, in an analogous way as in [8], in which the rate per stream was proposed to be adapted by the use of discrete constellations with variable number of points. As an additional contribution, in the following section we robustly design the set of moduli against CSI imperfections.

A more compact notation can be used for the output of the precoder in (1) by arranging the coefficients b_{kl} into the elements below the diagonal of a lower triangular matrix $\mathbf{B} \in \mathbb{C}^{n_S \times n_S}$ with ones in the main diagonal as $[\mathbf{B}]_{kl} = b_{kl},\ k > l$. The output of the precoder is then $\mathbf{p} = \mathbf{B}^{-1}(\mathbf{s}+\mathbf{a})$, where the elements of vector $\mathbf{a}$, $a_k = 2t_k u_k + \mathrm{j}2t_k v_k$, with $u_k, v_k \in \mathbb{Z}$, must be chosen so that the elements of $\mathbf{p}$ lie inside the modulo region defined by vector $\mathbf{t} = [t_1 \cdots t_{n_S}]^T$, *i.e.*, $p_k \in \mathcal{D}_k$.

In the following, we will focus on the single user case, *i.e.*, the different receivers are allowed to cooperate by jointly processing the received data (coupled receivers). The multiuser case, where the receiving antennas belong to different users and therefore only individual processing of the received symbols is possible (decoupled receivers), will be considered in Section IV.

For the case where the receivers are coupled (see Fig. 1), the transmitted signal vector, $\mathbf{x}_c$, is

$$\mathbf{x}_c = \mathbf{p} = \mathbf{B}^{-1}(\mathbf{s} + \mathbf{a}). \quad (2)$$

From dimensional analysis, last equation implies that the number of transmit antennas has to be equal to or higher than the number of transmitted signals, *i.e.*, $n_T \geq n_S$. If the number of antennas is higher than n_S then, a subset of n_S antennas is selected, where the selection is made according to an achievable rates maximization criterion as it is commented later. Moreover, since the single user has to recover all the n_S transmitted symbols, its mobile terminal has to be equipped with $n_R \geq n_S$ antennas. Let us assume that the selected antennas are numbered from 1 to n_S, then, the transmitted power for the coupled case, P_T^c, is calculated as $P_T^c = \mathbb{E}\, \mathbf{x}_c^H \mathbf{x}_c = \sum_{k=1}^{n_S} P_k$, where P_k is the power transmitted through k-th antenna. The power P_k is related with the modulus t_k as $P_k = \mathbb{E}\left[\mathbf{x}_c \mathbf{x}_c^H\right]_{kk} = 2t_k^2/3$, which results from assuming that the statistical distribution of s_k is uniform in $\mathcal{D}_k$, which is the one that maximizes the rate (see [21] and [22] for a detailed proof).

The transceiver design consists in specifying the set of moduli $\{t_k\}$ and the matrices $\mathbf{B}$, $\mathbf{G}$, and $\mathbf{F}$, as depicted in Fig. 1, which were found in [8] for a zero-forcing criterion. Here we recall their results: from the ql-factorization $\mathbf{H} = \mathbf{F}^H \mathbf{S}$, where $\mathbf{F}$ is unitary and $\mathbf{S}$ is lower triangular, we get

$$\begin{aligned} \mathbf{G} &= \mathbf{diag}(\{s_{ii}^{-1}\}), \\ \mathbf{B} &= \mathbf{GS}, \\ g_{ii} &= [\mathbf{G}]_{ii} = s_{ii}^{-1}. \end{aligned} \quad (3)$$

With these definitions, we reproduce in (4) the estimate, $\mathbf{r} \in \mathbb{C}^{n_S}$, of the data symbols, $\mathbf{s}$.

$$\begin{aligned} \mathbf{r} = M_{\mathbf{t}}\left[\mathbf{GFy}\right] &= M_{\mathbf{t}}\left[\mathbf{GFHx}_c + \mathbf{GFn}\right] = \\ &= M_{\mathbf{t}}\left[\mathbf{Bx}_c + \mathbf{GFn}\right] = M_{\mathbf{t}}\left[\mathbf{s} + \mathbf{GFn}\right]. \end{aligned} \quad (4)$$

Notice that $M_{\mathbf{t}}[\mathbf{z}]$ performs a modulo-t_k operation for each element z_k. We define a new noise vector $\tilde{\mathbf{n}} = \mathbf{Fn}$, which has the same statistical behavior as $\mathbf{n}$. In addition, since $\mathbf{G}$ is a diagonal matrix, $\mathbf{r}$ can be expressed in the form of n_S parallel data streams as

$$r_k = M_{t_k}\left[s_k + g_{kk}\tilde{n}_k\right], \quad k = 1, \ldots, n_S. \quad (5)$$

This model of parallel data streams clarifies the purpose of utilizing a precoder to presubtract the intersymbol interference that is caused by the channel matrix: we obtain n_S parallel (or independent) data streams between the transmitter and the receiver. Note that the quality of the data streams is dictated by the components of the diagonal of the lower triangular matrix $\mathbf{S}$, which fulfill that $s_{ii} \neq 0$, $\forall i$, as $\prod s_{ii} = \det(\mathbf{S}^H\mathbf{S}) = \det(\mathbf{H}^H\mathbf{H})$, and $\mathbf{H}$, which is assumed to be full rank, is redefined as a matrix containing the n_S columns of the original channel matrix corresponding to the selected antennas.

III. Robust Design with Imperfect CSI in the Single User Case (Coupled Case)

In this section we consider the practical scenario where the CSI at the transmitter side is imperfect and, hence, there is a mismatch between the ideal transmitter design, as given by $\mathbf{B}$,

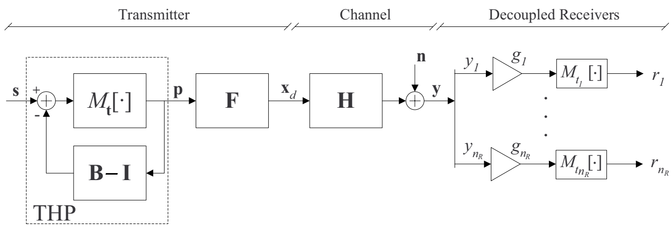

Fig. 2. Multiuser communication scheme with Spatial Tomlinson-Harashima Precoding.

and the actual transmitter which we shall denote by $\widehat{\mathbf{B}}$, which is the erroneous version of its ideal counterpart. Precisely, the main contribution of this section is that we obtain a robust design, whose objective is the maximization of the sum rate taking into account explicitly the presence of errors in the CSI. Since we are interested in high data rate transmission, the following analysis is done assuming that the SNR is high (for practical purposes this implies that it has to be greater than 10 dB) and that the error in the precoding matrix $\widehat{\mathbf{B}}$ is kept small.

We consider that the receiver feeds back the elements below the main diagonal of the matrix $\mathbf{B}$ to the transmitter (note that feeding back these elements requires a lower amount of feedback than transmitting the whole channel matrix $\mathbf{H}$ from the receiver to the transmitter). If some kind of error is considered in the feedback link or in the channel estimation, the transmitter will work with an erroneous precoding matrix, $\widehat{\mathbf{B}} = \mathbf{B}+\mathbf{\Delta_B}$ with $\mathbf{\Delta_B}$ being a strictly lower triangular matrix. In this case, using the notation in (2), the transmitted signal, $\mathbf{x}_{c,\widehat{\mathbf{B}}}$, becomes

$$\mathbf{x}_{c,\widehat{\mathbf{B}}} = \widehat{\mathbf{B}}^{-1}(\mathbf{s}+\mathbf{a}_{\widehat{\mathbf{B}}}) = (\mathbf{B}+\mathbf{\Delta_B})^{-1}(\mathbf{s}+\mathbf{a}_{\widehat{\mathbf{B}}}). \quad (6)$$

Notice that $\mathbf{a}_{\widehat{\mathbf{B}}}$ in (6) must now be chosen so that $\mathbf{x}_{c,\widehat{\mathbf{B}}}$ lies inside the modulo region defined by vector $\mathbf{t}$. In addition, it is important to state that, in general, $\mathbf{a}_{\widehat{\mathbf{B}}} \neq \mathbf{a}$, where $\mathbf{a}$ is defined as in (2). However, if the variance of the elements of the error matrix $\mathbf{\Delta_B}$ is kept small it can be assumed that (**As1**): $\mathbf{a} = \mathbf{a}_{\widehat{\mathbf{B}}}$ (see Section V for a validity evaluation). Under (**As1**), substituting (6) in (4) and using the matrix inversion lemma, $(\mathbf{B}+\mathbf{\Delta_B})^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}(\mathbf{\Delta_B}\mathbf{B}^{-1}+\mathbf{I}_{n_S})^{-1}\mathbf{\Delta_B}\mathbf{B}^{-1}$, the received signal can be expressed as $\mathbf{r}_{\widehat{\mathbf{B}}} = M_{\mathbf{t}}[\mathbf{s}+(\mathbf{\Delta_B}\mathbf{B}^{-1}+\mathbf{I}_{n_S})^{-1}\mathbf{\Delta_B}\mathbf{B}^{-1}(\mathbf{s}+\mathbf{a})+\mathbf{G}\tilde{\mathbf{n}}]$. Notice that $\mathbf{B}^{-1}(\mathbf{s}+\mathbf{a}) = \mathbf{x}_c$, *i.e.* we can reduce $\mathbf{B}^{-1}(\mathbf{s}+\mathbf{a})$ to $\mathbf{x}_c$, which would be the transmitted signal in the coupled case if no errors were present in $\widehat{\mathbf{B}}$ matrix, and whose components are bounded in $[-t_k, t_k)\times[-\mathrm{j}\,t_k, \mathrm{j}\,t_k)$. Defining $\mathbf{L} = (\mathbf{\Delta_B}\mathbf{B}^{-1}+\mathbf{I}_{n_S})^{-1}\mathbf{\Delta_B}$ the received signal reads

$$\mathbf{r}_{\widehat{\mathbf{B}}} = M_{\mathbf{t}}[\mathbf{s}+\mathbf{L}\mathbf{x}_c+\mathbf{G}\tilde{\mathbf{n}}], \quad (7)$$

where $\mathbf{L}$ is a strictly lower triangular matrix.[1] Attention must be paid to the fact that if the elements of $\mathbf{\Delta_B}$ are sufficiently small, the first order approximation of $\mathbf{L} = \mathbf{\Delta_B} + \mathrm{o}(\mathbf{\Delta_B})$, becomes valid.

Once we have obtained an expression for the input-output relation in the presence of feedback or estimation errors we can proceed to maximize the mutual information between $\mathbf{s}$ and $\mathbf{r}_{\widehat{\mathbf{B}}}$ with respect to the set of moduli $\{t_k\}$. As it was described in Section II, the power transmitted through k-th selected antenna, P_k, is controlled by the modulo t_k by the relation $P_k = 2t_k^2/3$. In [21], the design of $\{t_k\}$ in the perfect CSI case was addressed and we found that performing antenna selection and choosing all the $\{t_k\}$ to be equal (uniform power allocation) is a quasi-optimal solution in terms of achievable rates. Note that the optimal set of selected antennas in the imperfect CSI case need not be the same as in the perfect CSI case, but, in practice, it can be checked that the two sets coincide (at least with the small errors assumption). From now on, the antennas of the active set are numbered from 1 to n_S, *i.e.* $P_1,\ldots,P_{n_S} \geq 0$, $P_{n_S+1},\ldots,P_{n_T} = 0$. In the following, the set of moduli $\{t_k\}$ (or, equivalently, the power distribution) will be adapted to maximize the worst-case achievable rates when feedback errors are present, giving thus rise to a robust design. It will be shown that in the robust case the optimal moduli are not equal.

The first order approximation in $\mathbf{\Delta_B}$ of the received signal vector is (**As2**): $\mathbf{r}_{\widehat{\mathbf{B}}} \approx M_{\mathbf{t}}[\mathbf{s}+\mathbf{\Delta_B}\mathbf{x}_c+\mathbf{G}\tilde{\mathbf{n}}]$. If $\mathbf{\Delta_B}\mathbf{x}_c$ is treated as an unknown interference, the mutual information between k-th element of data vector, s_k, and the corresponding element of received signal, $r_k = [\mathbf{r}_{\widehat{\mathbf{B}}}]_k$, is

$$I(s_k;r_k) = \log(6P_k) - h\left(M_{t_k}\left[g_{kk}\tilde{n}_k + \sum\nolimits_{j<k}\delta_{kj}x_j\right]\right), \quad (8)$$

where $\delta_{kj} = [\mathbf{\Delta_B}]_{kj}$ and $h(\cdot)$ denotes differential entropy, analogously as in [21]. Note that, as $x_k = [\mathbf{x}_c]_k$ is the output of the precoder, it is uniformly distributed in $\mathcal{D}_k$ with variance P_k.

Let us define the random variable $z_k = g_{kk}\tilde{n}_k + \sum_{j<k}\delta_{kj}x_j$ with power $\mathbb{E}\{|z_k|^2\} = g_{kk}^2\sigma^2 + \sum_{j<k}|\delta_{kj}|^2P_j$. It can be deduced from [23, chapters 6 and 8] that z_k can be approximately modeled as a complex Gaussian random variable as long as (**As3**): $\max_j |\delta_{kj}|^2P_j \lesssim g_{kk}^2\sigma^2/3$. Under (**As3**), we found in [21] that the mutual information expression (8) is very well approximated by

$$(\mathbf{As4}):\ \ I(s_k;r_k) \approx \log^+\left(\frac{6P_k}{\pi e(g_{kk}^2\sigma^2 + \sum_{j<k}|\delta_{kj}|^2P_j)}\right), \quad (9)$$

where $\log^+(x) = \max(\log(x), 0)$. The achievable rates for the STHP structure will then be the sum of the mutual information of each active stream, $\sum_{k=1}^{n_S} I(s_k;r_k)$.

In order to describe the noise worst-case scenario we consider that the squared moduli of the components of the error matrix $\mathbf{\Delta_B}$ are bounded, *i.e.* $|\delta_{kj}|^2 < \alpha_{kj}$. In addition, for the sake of simplicity we assume that $\alpha_{kj} = \alpha$, $\forall k,j$. Notice that, since we are interested in the worst-case, no generality is lost with last restriction, because, in case that the values of α_{kj} were different for some k,j, it would suffice to let $\alpha = \max_{k,j}\alpha_{kj}$.

From all the considerations above, the power distribution that maximizes the worst-case achievable rates when the CSI

[1] Notice that the inverse of a triangular matrix is also triangular. Additionally, the product of a lower triangular matrix and a strictly lower triangular matrix is a strictly lower triangular matrix.

is imperfect is the solution to the following maximin problem:

$$C_{\text{THP}}^{rob,c} = \max_{\{P_i\}} \min_{\{\delta_{ij}\}} \sum_{k=1}^{n_S} I(s_k; r_k), \quad \text{s.t.} \quad |\delta_{ij}|^2 \le \alpha,\ \forall i,j, \quad \sum_{k=1}^{n_S} P_k = P_T.$$

The solution to the minimization part is trivial, since each term $I(s_k; r_k)$ is a decreasing function of $|\delta_{ij}|^2$ and each $|\delta_{ij}|^2$ is upper bounded independently of the others. Thus, the minimum will be attained when $|\delta_{ij}|^2 = \alpha, \forall i, j$. The resulting maximization problem is a standard constrained optimization problem, and can be solved with the use of the Lagrange method. The Lagrange function is, up to a constant,

$$\mathcal{L}(\{P_k\};\lambda) = \sum_{k=1}^{n_S} \log\left(\frac{P_k}{g_{kk}^2\sigma^2 + \alpha\sum_{j<k} P_j}\right) + \lambda\left(\sum_{k=1}^{n_S} P_k - P_T\right). \quad (10)$$

Now, the optimal power allocation should satisfy

$$\frac{\partial \mathcal{L}(\{P_k\};\lambda)}{\partial P_k} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}(\{P_k\};\lambda)}{\partial \lambda} = 0, \quad \forall k, \quad (11)$$

with the additional constraint that $\{P_k\}$ is non-negative $\forall k$. With some basic manipulations from (10) and (11) a recursive relation of the type $P_{n_S-k} = f(P_{n_S-k+1}, \ldots, P_{n_S})$ for $k = 1, \ldots, n_S - 1$ between the assigned power to each antenna can be found as

$$P_{n_S-k} = P_{n_S-k+1} \frac{w_{n_S-k+1} + \alpha\left(P_T - \sum_{j>n_S-k} P_j\right)}{w_{n_S-k+1} + \alpha\left(P_T - \sum_{j>n_S-k+1} P_j\right)}, \quad (12)$$

where $w_i = g_{ii}^2\sigma^2$ and $k = 1, \ldots, n_S - 1$. Notice that the second factor in last equation is always lower than unity, which implies necessarily that $P_1 \le P_2 \le \ldots \le P_{n_S-1} \le P_{n_S}$, which is a reasonable solution since power interference is progressive in the sense that P_k interferes with streams from $k+1$ to n_S, but not with streams from 1 to $k-1$, see (9). The set of equations in (12) together with $\sum_{k=1}^{n_S} P_k = P_T$ can be solved numerically obtaining the robust power allocation or set of moduli.

IV. Robust Design with Imperfect CSI in the Multiuser Case (Decoupled Case)

In the multiuser scenario (see Fig. 2), since the receivers may be located at geographically separated places, only individual processing of each element of the received signal vector is permitted. In [8], it was found that the design, for a zero-forcing criterion, of the matrices $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{G}$ is based on the lq-factorization of the channel matrix $\mathbf{H}$. The design is given by

$$\mathbf{H} = \mathbf{S}\mathbf{F}^H, \quad \mathbf{G} = \mathbf{diag}(\{s_{ii}^{-1}\}), \quad \mathbf{B} = \mathbf{G}\mathbf{S}, \quad g_{ii} = [\mathbf{G}]_{ii} = s_{ii}^{-1}, \quad (13)$$

where $\mathbf{S}$ is a lower triangular matrix and $\mathbf{F}$ is unitary.

In the decoupled case, two precoding matrices are necessary at the transmitter side: $\mathbf{B}$ and $\mathbf{F}$, which cannot be calculated at the receivers side because each receiver only knows one of the rows of the full matrix $\mathbf{H}$. Consequently, we assume that the transmitter is informed, *e.g.*, through a feedback link or by direct estimation, with an erroneous channel matrix $\widehat{\mathbf{H}} = \mathbf{H} + \mathbf{\Delta_H}$, and that, based upon that estimate, the transmitter calculates $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{F}}$ following (13). Furthermore, we consider that the entries of $\mathbf{\Delta_H}$ are i.i.d. circularly symmetric complex Gaussian random variables. The error present in the estimate $\widehat{\mathbf{H}}$ propagates to the estimates $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{F}}$ as $\widehat{\mathbf{B}} = \mathbf{B} + \mathbf{\Delta_B}$, and $\widehat{\mathbf{F}} = \mathbf{F} + \mathbf{\Delta_F}$. This error propagation has been recently studied in [24] to characterize the BER in STHP systems. Fortunately, the explicit expressions for $\mathbf{\Delta_B}$ and $\mathbf{\Delta_F}$ are not needed here, because the authors of [24] also obtained a very simple expression for the estimate of the data symbols vector in the multiuser case when $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{F}}$ are utilized in the transmitter design:

$$\mathbf{r} = M_{\mathbf{t}}\left[\mathbf{s} + \boldsymbol{\xi}(\mathbf{s}) + \mathbf{G}\mathbf{n}\right]. \quad (14)$$

In last equation, $\boldsymbol{\xi}(\mathbf{s})$ is a random vector that represents the effects of having imperfect versions of $\mathbf{B}$ and $\mathbf{F}$ at the transmitter side, and is independent of $\mathbf{n}$ but depends on the transmitted symbols sequence. In [24], the authors also found that, when the entries of $\mathbf{\Delta_H}$ are i.i.d. Gaussian random variables, $\boldsymbol{\xi}$ is distributed as a zero mean circularly symmetric complex Gaussian random vector with $\mathbb{E}\boldsymbol{\xi}\boldsymbol{\xi}^H = \Xi$. An explicit expression for the calculation of Ξ is given in [24] and is not reproduced here for the sake of space. We simply need to know that Ξ can be calculated for the worst possible sequence of transmitted symbols as a function of $\mathbf{H}$ and the power of the entries of $\mathbf{\Delta_H}$. Notice that, in this case, the structure of the interference is not progressive, as it was found to be in the coupled case (7). This is due to the fact that, in the decoupled case, the output of the precoder $\mathbf{p}_{\widehat{\mathbf{B}}}$ is multiplied by the matrix $\widehat{\mathbf{F}}$ which is unitary and consequently distributes uniformly the interference that is present in $\mathbf{p}_{\widehat{\mathbf{B}}}$ among all the components of the transmitted signal $\mathbf{x}_d = \widehat{\mathbf{F}}\mathbf{p}_{\widehat{\mathbf{B}}}$.

In general, the matrix Ξ is not diagonal, which implies that the entries of the interference term $\boldsymbol{\xi}$ in (14) can be correlated. However, since the receivers can not cooperate, no advantage can be taken from the correlation of the interference term in the decoding process, and thus we only need to consider the diagonal entries of Ξ, which represent the interference power seen by each user.

From what has been said above, the entries of the interference term $\boldsymbol{\xi}$ can be accurately modeled as an additional source of Gaussian noise independent of $\mathbf{Gn}$, and whose power is given by the diagonal elements of the matrix Ξ. Consequently, since the sum of two independent Gaussian variables is another Gaussian variable whose power is given by the sum of the individual powers, the two noise terms can be easily grouped into a single noise term, *i.e.* $\xi_k + g_{kk}n_k \sim \mathcal{CN}(0, g_{kk}^2\sigma^2 + [\Xi]_{kk})$. Now, the received vector in (14) can thus be equivalently expressed as

$$\mathbf{r} = M_{\mathbf{t}}\left[\mathbf{s} + \mathbf{D}\boldsymbol{\nu}\right], \text{ or } r_k = M_{t_k}\left[s_k + d_{kk}\nu_k\right], \forall k,$$

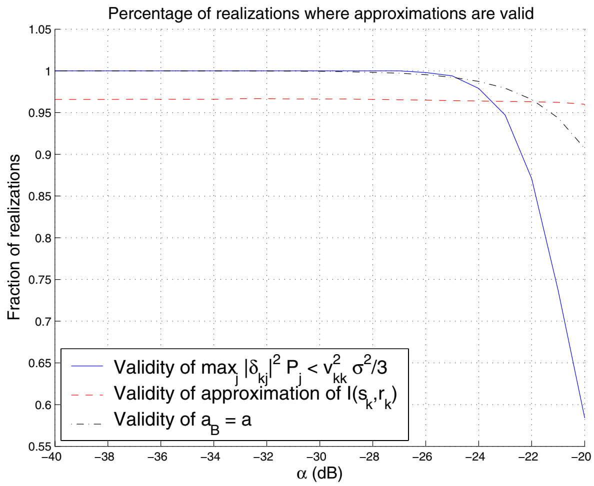

Fig. 3. Validation of the approximations done in the analysis. For a given channel realization, we have computed the ratio where the approximations done in Section III were valid as a function of the α parameter. Similar results have been obtained for other realizations of the channel matrix.

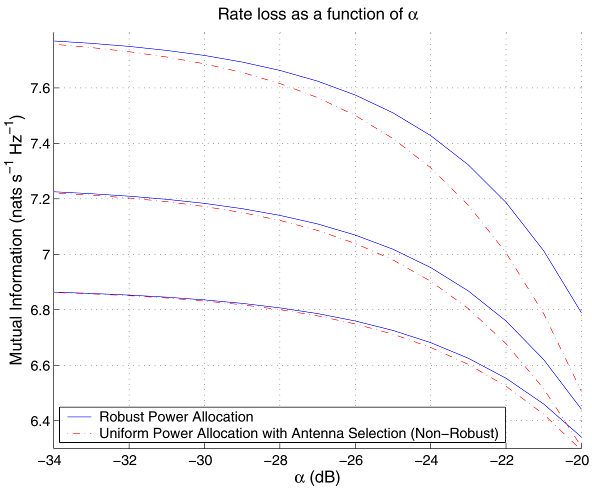

Fig. 4. Capacity for different power allocation strategies as a function of α for the single user scenario for three realizations of the channel matrix.

where $\mathbf{D}$ is a diagonal matrix, with $[\mathbf{D}]_{kk}^2 = d_{kk}^2 = g_{kk}^2\sigma^2 + [\boldsymbol{\Xi}]_{kk}$, and where $\boldsymbol{\nu} \sim \mathcal{CN}(0, \mathbf{I}_{n_S})$.

We can now express the robust achievable rates for the decoupled case as the solution to

$$C_{\text{THP}}^{rob,d} = \max_{\{P_k\}} \quad \sum_{k=1}^{n_S} \log(6P_k) - h\left(M_{\sqrt{3P_k/2}}[d_{kk}\nu_k]\right), \quad \text{s.t.} \quad \sum_{k=1}^{n_S} P_k = P_T. \tag{15}$$

However, one readily sees that the problem in (15) is the same that was solved in [21], but with a set of noise powers according to $d_{kk}^2 = g_{kk}^2\sigma^2 + [\boldsymbol{\Xi}]_{kk}$. Consequently, the problem can be solved quasi-optimally by selecting an active set of streams (*users*) and performing uniform power allocation among them (which is equivalent to saying that t_k has a constant value for all users in the active set). Note that, in the imperfect CSI case, the robust set of active users does not have to be the same as in the perfect CSI case. This is due to the fact that, when the CSI is imperfect, the ordering of the new set of noise powers d_{kk}^2 may be different than the set $g_{kk}^2\sigma^2$. Consequently, the algorithm may deactivate a user whose channel was good with perfect CSI, but whose channel quality has decreased significantly due to the presence of the term $[\boldsymbol{\Xi}]_{kk}$.

V. Simulations and Conclusion

To give graphical representations of our results, we have considered a flat-fading 3×3 MIMO channel matrix, with i.i.d. complex Gaussian entries. The working mean SNR is $P_T/\sigma^2 = 15$ dB.

On the one hand, for the single user case, as the robust capacity analysis has been done using numerous approximations (**As1**), (**As2**), (**As3**), and (**As4**), before presenting the simulations results, the validity of the approximations is shown in Fig. 3 by plotting the fraction of realizations with respect to the noise in which the approximations are valid. Precisely, (**As4**) is considered to hold true when the relative difference between the approximation (9) and the actual value (8), calculated numerically, is lower than 10^{-3}. It can be seen that, for the particular values of the simulation parameters taken in this section, the capacity analysis is valid for values of α up to -21 dB. For values of α higher than -21 dB our analysis is not valid. However, from the slope of the robust achievable rates curves at $\alpha = -21$ dB in Fig. 4, the rates that can be achieved for $\alpha > -21$ dB seem to decrease fast, so even with a robust technique it would be difficult to cope with the problems associated with having an imperfect feedback link. Notice that the validity of (**As 2**) has not been plotted as the differences between $\mathbf{L}$ and $\boldsymbol{\Delta}_{\mathbf{B}}$ have always been found negligible. In Fig. 4, we have plotted the maximum achievable rates for the robust power allocation strategy and also for the non-robust uniform power allocation with antenna selection scheme described in [21] for different realizations of the MIMO channel. In Fig. 5 we have plotted the fraction of the total power that is transmitted through each one of the antennas, which is related to the set of moduli used in STHP, for a particular realization of the channel. It can be seen that, as α gets close to -21 dB, the robust power allocation differs substantially from the uniform power allocation.

On the other hand, in the multiuser case, similar simulations have been conducted but in this case the parameter α represents the noise power of each component of the estimation error in the channel matrix $\widehat{\mathbf{H}} = \mathbf{H} + \boldsymbol{\Delta}_{\mathbf{H}}$, *i.e.*, $\mathbb{E}|[\boldsymbol{\Delta}_{\mathbf{H}}]_{ij}|^2 = \alpha$. For each channel realization the worst-case matrix $\boldsymbol{\Xi}$ has been computed following [24] and then a set of active users has been selected according to the set of noise powers given by $d_{kk}^2 = g_{kk}^2\sigma^2 + [\boldsymbol{\Xi}]_{kk}$. The resulting maximum achievable sum rate has been plotted in Fig. 6 for various channel realizations. In addition, since in the multiuser case we are not so strongly conditioned on the validity of the approximations done in the single user case, we can extend the domain of α. As the value of α increases the maximum achievable sum rate decreases very rapidly, which implies that, as the estimation noise increases, even the robust technique is not able to cope with the presence of errors in the feedback link. Note that this

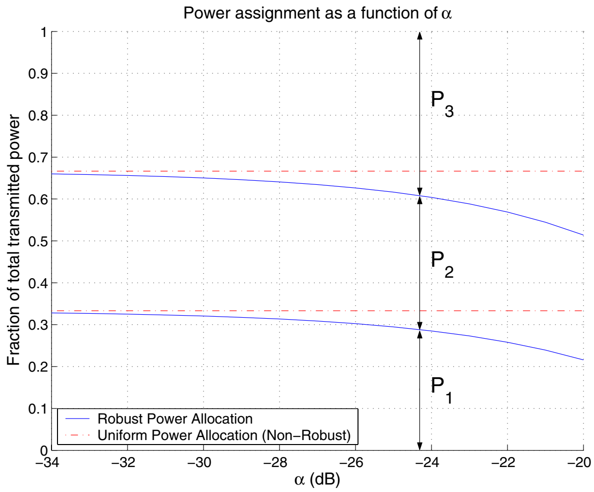

Fig. 5. Graphical representation of the robust power allocation with respect to the uniform power allocation. In the figure we plotted the fraction of the total transmitted power that is assigned to each of the three antennas as a function of α for the single user scenario.

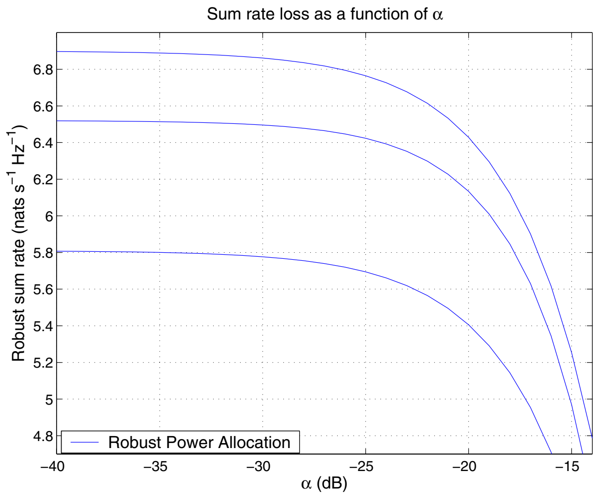

Fig. 6. Sum rate for the robust power allocation technique as a function of α in the multiuser scenario for three realizations of the channel matrix.

conclusion, also valid for the single user case, is not surprising since the robust design presented in this paper has been derived assuming small errors and, therefore, the degradation can be quite high when this assumption does not hold.

To conclude we would like to remark that in this paper we have analyzed some issues concerning the achievable rates of the STHP scheme in the presence of errors in the CSI. Initially, we have added degrees of freedom in the design of the STHP by allowing different modulo operations at the output of the precoder. Next, we have found two robust strategies for the design of the set of moduli or, equivalently, of the power allocation, for the single and multiuser scenarios, that maximize the mutual information when the available CSI at the transmitter side is not perfect.

Finally, we wish to highlight that the proposed algorithm could work in a realistic deployment since the presence of errors in the feedback link has been taken into account in the design process.

REFERENCES

[1] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 439–441, May 1983.

[2] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1691–1706, July 2003.

[3] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1912–1921, Aug. 2003.

[4] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, Sept. 2004.

[5] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2658–2668, Aug. 2003.

[6] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[7] R. Fischer, C. Windpassinger, A. Lampe, and J. Huber, "Space-time transmission using Tomlinson-Harashima precoding," in *Proc. 4th ITG Conference on Source and Channel Coding*, Jan. 2002.

[8] C. Windpassinger, R. Fischer, T. Vencel, and J. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, July 2004.

[9] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, no. 2, Mar. 1971.

[10] H. Miyakawa and H. Harashima, "Information transmission rate in matched transmission systems with peak transmitting power limitation," *Nat. Conf. Rec., Inst. Electron., Inform., Commun. Eng. of Japan*, vol. 7, no. 2, pp. 138–139, Aug. 1972.

[11] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, Jan. 2005.

[12] ——, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar. 2005.

[13] C. Windpassinger, R. Fischer, and J. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2057–2060, Dec. 2004.

[14] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding," in *Proc. ITG Conf. on Source and Channel Coding*, Erlangen (Germany), Jan. 2004, pp. 387–394.

[15] D. Schmidt, M. Joham, and W. Utschick, "Minimum mean square error vector precoding," in *Proc. IEEE Intl. Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'05)*, Berlin (Germany), Sept. 2005.

[16] R. Fischer, C. Windpassinger, A. Lampe, and J. Huber, "Tomlinson-Harashima precoding in space-time transmission for low-rate backward channel," in *Proc. IEEE Intl. Zurich Seminar on Broadband Comm. (IZS '02)*, Feb. 2002.

[17] F. Dietrich, P. Breun, and W. Utschick, "Tomlinson-Harashima precoding: a continuous transition from complete to statistical channel knowledge," in *Proc. IEEE Global Telecommunications Conference (Globecom'05)*, vol. 4, Nov. 2005, pp. 2379–2384.

[18] R. Hunger, F. Dietrich, M. Joham, and W. Utschick, "Robust transmit zero-forcing filters," in *Proc. ITG Workshop on Smart Antennas*, Munich (Germany), Mar. 2004.

[19] R. Fischer, "The modulo-lattice channel: the key feature in precoding schemes," *International J. Electron. and Commun.*, pp. 244–253, June 2005.

[20] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.

[21] M. Payaró, A. Pérez-Neira, and M. Lagunas, "Achievable rates for generalized spatial Tomlinson-Harashima precoding in MIMO systems," in *Proc. IEEE Vehic. Tech. Conf. Fall*, Sept. 2004.

[22] R. Wesel and J. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Trans. Inf. Theory*, vol. 2, no. 2, pp. 824–831, Mar. 1998.

[23] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed. McGraw-Hill, 1991.

[24] S. Bahng, J. Liu, A. Høst-Madsen, and X. Wang, "The effects of channel estimation on Tomlinson-Harashima precoding in TDD MIMO systems," in *Proc. IEEE Sig. Proc. Adv. in Wireless Comm. (SPAWC'05)*, New York, June 2005.